Ensemble-based retrospective analysis of the seasonal snowpack

Kristoffer Aalstad

Thesis submitted for the degree of Philosophiae Doctor (PhD)

Department of Geosciences
Faculty of Mathematics and Natural Sciences
University of Oslo
Oslo, Norway
2019

Series of dissertations submitted to the
Faculty of Mathematics and Natural Sciences, University of Oslo
No. 2186

ISSN 1501-7710

Cover: Hanne Baadsgaard Utigard.
Print production: Reprosentralen, University of Oslo.

Abstract

The seasonal snowpack, with its high albedo, low thermal conductivity and large water storing capacity, is a key component of the terrestrial energy, water, and carbon balance. At the same time, it is widely acknowledged that accurate estimation of the distribution of snow water equivalent (SWE) remains an outstanding challenge, especially in complex terrain. This is unfortunate because, as a modulator of the terrestrial balances, SWE is an essential climate variable.

Since the dawn of the satellite era, the depletion of snow-cover retrieved from optical satellite sensors has been used to reconstruct SWE in a deterministic manner. The basic idea is to use the remotely sensed date of disappearance of the snowpack to perform a backwards calculation to reconstruct SWE using modeled snowmelt rates. Despite some success, this approach is limited given that it completely ignores uncertainties in the forcing, model, and retrievals. Recently, ensemble-based snow retrospective analysis (reanalysis) that accounts for these uncertainties has emerged as a robust alternative to traditional reconstruction. In this approach, the assimilation of the remotely sensed depletion of fractional snow-covered area (fSCA) is used to constrain an ensemble of modeled realizations of the seasonal snowpack. Coincidentally with this emergence, a new generation of optical satellite sensors, such as Sentinel-2 MSI, have been launched into orbit to complement the already existing satellite climate data record.

In this work, multiple satellite sensors are leveraged in the pursuit of ways to further improve the ensemble-based reanalysis approach to SWE reconstruction. A modular reanalysis framework is presented with four components: a topographic downscaling routine, a simple snow model, fSCA retrieval algorithms, and ensemble-based data assimilation schemes. This framework can in principle be applied anywhere on Earth. This work focuses on three disparate study sites: the Brøgger peninsula in high-Arctic Svalbard archipelago, the Mammoth Lakes Basin in the Californian Sierra Nevada, and the Swiss Alps. Spectral unmixing is shown to provide the means of retrieving unbiased hill-slope scale fSCA from optical satellite sensors. It is also shown that an iterative ensemble smoother algorithm can outperform the data assimilation schemes that have previously been proposed for snow reanalysis. Through the experimental framework, a taste is provided of what is possible when combining robust ensemble-based data assimilation with emerging remotely-sensed data streams in the reanalysis exercise.

Acknowledgements

For me, the PhD adventure has been a voyage with both peaks and troughs. It would not have been possible to complete this journey alone. As such, there are many people in many places that I would like to express my sincere thanks to. Most of all I would like to thank my supervisor, Sebastian Westermann, for your endless supply of support, encouragement, and ideas. Your contagious optimism has been a great source of inspiration, particularly during the more difficult periods of the PhD. My work would not have been possible without your invaluable efforts, help, and guidance, which I will always be grateful for. I would also like to extend my gratitude to my co-supervisor, Laurent Bertino, for all the crucial help and feedback that you have provided during the course of my PhD work. Again, my work would not have been possible without your contributions. Thanks as well for hosting several fruitful visits to NERSC. I also want to acknowledge all the other co-authors on the scientific papers in this thesis. Thanks to Thomas Vikhamar Schuler and Julia Boike for your help on the first paper. A big thank you also to Joel Fiddes for your contributions to the third and fourth papers as well as our many fruitful conversations. Thanks as well to my PhD committee: Francois Renard (administrator), Gabriëlle de Lannoy (first opponent), and Stefan Wunderle (second opponent) for your invaluable comments and time spent on evaluating my PhD thesis.

In Oslo, I would like to thank all my colleagues, past and present, at the Geography and Hydrology (GeoHyd) Section at the University for providing a great working environment. Thanks in particular to the organizers and attendees of GeoFootball and GeoPils which always added an extra social dimension. I would like to especially thank a few people that I have worked more closely with at GeoHyd. Thanks to Paul Leclercq for including me in in your work on glacier data assimilation, to my fellow uncertainty disciple Aynom Teweldebrhan for many productive discussions, and to Bas Altena for your many suggestions regarding satellite imagery. In addition, I would like to thank my two office mates, Thorben Dunse and Jaroslav Obu, for making the office amicable.

In Bergen, I would like to especially thank colleagues associated with NERSC and the Nordic Centre for Excellence EmblA for their help during my visits. In particular, I would like to thank both Alberto Carrassi and Patrick Raanes for their comments during the beginning of my PhD, as well as Hans Wackernagel for his

geostatistical wisdom. In addition, I would like to thank Geir Evensen for his suggestions regarding ensemble smoothers. Thanks also to the organizers of CHESS for leading a great research school that provided my curriculum with engaging courses.

In Boulder, I would like to thank Logan Karsten at NCAR for hosting my long term visit and Roy Rasmussen for helping to make the visit possible. Thanks as well to Ethan Guttman and James McCreight for several valuable exchanges and for being interested in my work. In addition, I want to thank Danica Lombardozzi, Katie Dagon, Dave Lawrence, and Alessandro Fanfarillo for being so welcoming and making the stay in Boulder enjoyable.

Elsewhere, I would like to thank several people that I have learned from. Thanks to Alexandre Emerick for your suggestions on ensemble smoothing, to Sjur Kolberg for your comments on snow inversion, to Keith Beven for both confusing me and teaching me valuables lessons about uncertainty, and also to both Thomas Painter and Kat Bormann for providing data from the Airborne Snow Observatory. In addition, I would like to thank the agencies, centers, institutes, and companies that have provided me with data during the course of my PhD: NASA, ESA, USGS, JAXA, NSIDC, ECMWF, NPI, and Planet.

Finally, I would like to thank my family and friends. Without your never ending support I would never have been able to complete my PhD. I apologize for all the odd times that I have been immersed in work when I should have been available to you. Thanks for being there for me despite my occasional absences and forays into the academic jungle. A special thanks to my sister, Elin Marie Aalstad, for reading through the thesis. Finally, I want to extend a heartfelt thanks to my wife, Yeliz Yılmaz, for being my partner during the PhD adventure.

Kristoffer Aalstad Oslo, 14 June 2019

Funding

The primary source of funding for this work has been the "Satellite-based Permafrost Modeling across a Range of Scales" (SatPerm) project #239918 supported by the Research Council of Norway. A 1 year research visit to the National Center for Atmospheric Research in Boulder, Colorado (USA) was additionally supported by a Personal Overseas Research grant that was provided by the Research Council of Norway. The CHESS research school fully funded my participation in several courses and two national meetings. In addition, ResClim/CHESS provided me with a travel grant associated with a best poster presentation award which supported my attendance to the 2016 AGU fall meeting. The LATICE research group fully supported my attendance to two annual meetings. The CryoJaNo partnership and exchange program fully funded my participation in the GEO9411 field course that took place on Hokkaido. Evaluation of Ocean Synthesis COST Action ES1402 fully funded my participation in the 2016 School on Data Assimilation and Data Analysis Methods in Lecce, Italy. The industrial liaison (IL) fund at the Department of Geosciences of the University of Oslo supported my participation to the 2nd Asian Conference on Permafrost in 2017 and the ESA Living Planet Symposium in 2019.

List of papers

Paper I

Aalstad, K., Westermann, S., Schuler, T. V., Boike, J., and Bertino, L. (2018): Ensemble-based assimilation of fractional snow-covered area satellite retrievals to estimate the snow distribution at Arctic sites. *The Cryosphere* 12: 247-270, doi: 10.5194/tc-12-247-2018.

Paper II

Aalstad, K., Westermann, S., and Bertino, L. (2019): Evaluating satellite retrieved fractional snow-covered area at a high-Arctic site using terrestrial photography. *Manuscript submitted to Remote Sensing of Enviornment*.

Paper III

Aalstad, K., Westermann, S., Fiddes, J., and Bertino, L. (2019): Ensemble-based snow reanalysis using dense time stacks of multisensor multispectral satellite imagery. *Manuscript to be submitted*.

Paper IV

Fiddes, J., **Aalstad, K.**, and Westermann, S. (2019): Hyper-resolution ensemble-based snow reanalysis in mountain regions using clustering. *Manuscript submitted to Hydrology and Earth System Sciences*, doi: 10.5194/hess-2019-37.

Contribution

For Papers I, II, and III, as first author I obtained most of the data, wrote the code, performed the analyses, and wrote the manuscripts which were subsequently edited by all the co-authors. For paper IV the first author, J. Fiddes, did the majority of the work. As second author, I designed the data assimilation framework and helped to implement it. In addition, I wrote the data assimilation section and edited the manuscript substantially.

List of Acronyms

This list contains the most important and widely used acronyms in this thesis and the accompanying papers.

ASO	Airborne Snow Observatory
DA	Data Assimilation
DEM	Digital Elevation Model
ECV	Essential Climate Variable
EnKF	Ensemble Kalman Filter
ES	Ensemble Smoother
ES-MDA	Ensemble Smoother with Multiple Data Assimilation
ETM+	Enhanced Thematic Mapper Plus
fSCA	Fractional Snow-Covered Area
fCCA	Fractional Canopy-Covered Area
GRACE	Gravity Recovery and Climate Experiment
LIDAR	LIght Detection And Ranging
MCMC	Markov Chain Monte Carlo
MODIS	MODerate resolution Imaging Spectroradiometer
MSI	MultiSpectral Instrument
NDSI	Normalized Difference Snow Index
NDVI	Normalized Difference Vegetation Index
OLI	Operational Land Imager
PDF	Probability Density Function
PF	Particle Filter
SDC	Snow Depletion Curve
SNICAR	SNow, ICe, and Aerosol Radiative transfer model
SSD	Subgrid Snow water equivalent Distribution
SSM	Simple Snow Model
SU	Spectral Unmixing
SWE	Snow Water Equivalent
TWSA	Terrestrial Water Storage Anomaly
VNIR	Visible Near InfraRed
VSWIR	Visible ShortWave InfraRed

List of symbols

This list contains most of the symbols used in the thesis and the associated papers. The name and units associated with each symbol is provided. For non-scalars, expressed in bold-font, the dimensions are included. Variables are mainly defined with respect to a point or grid cell on Earth's surface.

α : Observation error inflation coefficient
α_s : Snow albedo [-]
$\mathcal{A}$: Net accumulation rate [m s^{-1}]
$\mathbf{a}$: Fractional abundance vector ($N_b \times 1$)
b_M : Snowmelt bias parameter
b_P : Precipitation bias parameter
c_i : Specific heat of ice = 2100 [J kg^{-1} K^{-1}]
c_p : Specific heat of air at constant pressure = 1005 [J kg^{-1} K^{-1}]
c_w : Specific heat of liquid water = 4180 [J kg^{-1} K^{-1}]
$\mathcal{C}_{\widetilde{\mathbf{\Theta}}\widehat{\mathbf{Y}}}$: Transformed parameter-predicted observation error covariance matrix
$\mathcal{C}_{\widehat{\mathbf{Y}}\widehat{\mathbf{Y}}}$: Predicted observation error covariance matrix
χ : Peak subgrid coefficient of variation [-]
d : Snow depth (height of snow) [m]
D : Snow water equivalent (SWE) depth [m]
$\overline{D}$: Subgrid mean SWE depth [m]
D_m : Accumulated melt depth [m]
δ : Binary shadow mask [-]
Δt : Time step [s]
e : Water vapor pressure [-]
ε : Emissivity
$\boldsymbol{\epsilon}$: Residual vector ($N_o \times 1$)
f_{diff} : Diffuse fraction [-]
f_P : Peak subgrid SWE distribution [m^{-1}]
f_R : Rain fraction [-]
g : Acceleration of gravity at Earth's surface = 9.81 [m s^{-2}]
γ : Horizon angle [rad]
h : Measurement height [m]
H : Scale height [m]
$\mathbf{H}$: Linear observation operator ($N_o \times [N_s \times N_t]$)
$\mathcal{H}$: Observation operator
k_t : Clearness index [-]
κ : von Kármán constant = 0.4

$L^{\downarrow}$: Incoming longwave radiation [W m^{-2}]
L_f : Latent heat of fusion = 3.34×10^5 [J kg^{-1}]
L_s : Latent heat of sublimation = 2.835×10^6 [J kg^{-1}]
L_v : Latent heat of vaporization = 2.5×10^6 [J kg^{-1}]
L_* : Obukhov length [m]
$\mathbf{M}$: Theoretical reflectance matrix ($N_b \times N_m$)
$\mathcal{M}$: Forward model operator
μ : Peak mean SWE [m]
N_a : Number of assimilation cycles
N_b : Number of spectral bands
N_e : Number of particles/ensemble members
N_m : Number of endmembers
N_o : Number of observations
N_p : Number of parameters
N_s : Number of state variables
N_t : Number of time steps
p : Air pressure [Pa]
$p(\cdot)$: Probability density function [argument units^{-1}]
P : Precipitation rate [m s^{-1}]
P_S : Snowfall rate [m s^{-1}]
P_R : Rainfall rate [m s^{-1}]
ϕ_z : Aspect [rad]
ϕ_0 : Solar azimuth angle [rad]
$\phi_{M,H,W}$: Universal stability functions for momentum, heat, and moisture [-]
$\Psi_{M,H,W}$: Integrated stability functions for momentum, heat, and moisture [-]
$\boldsymbol{\psi}$: Gaussian anamorphosis function
q : Specific humidity [-]
Q_E : Latent heat flux [W m^{-2}]
Q_G : Ground heat flux [W m^{-2}]
Q_H : Sensible heat flux [W m^{-2}]
Q_M : Snowmelt flux [W m^{-2}]
Q_P : Heat advected by precipitation [W m^{-2}]
Q_R^* : Global radiation [W m^{-2}]
Q_0 : Initial ground heat flux [W m^{-2}]
r : Reflectance [-]
$r_a^{H,W}$: Aerodynamic resistance for sensible and latent heat [s m^{-1}]
r_s : Surface resistance against evapotranspiration [s m^{-1}]
$\mathbf{r}$: Band reflectance vector
RH : Relative humidity [%]
R : Specific gas constant for dry air [287 J kg^{-1} K^{-1}]

$\mathbf{R}$: Observation error covariance matrix ($N_o \times N_o$)
ρ_a : Air density [kg m^{-3}]
ρ_s : Bulk density of the snowpack [kg m^{-3}]
ρ_w : Density of fresh liquid water [kg m^{-3}]
$S^{\downarrow}$: Incoming shortwave radiation [Wm^{-2}]
$S^{\downarrow}_{\infty}$: Incoming exo-atmospheric shortwave radiation [Wm^{-2}]
σ_{sb} : Stefan-Boltzmann constant = 5.67×10^{-8} [W m^{-2} K^{-4}]
T : Air temperature [K]
T_d : Dewpoint temperature [K]
T_0 : Melting point = 273.15 [K]
θ_i : Illumination angle [rad]
θ_z : Slope [rad]
θ_0 : Solar zenith angle [rad]
$\boldsymbol{\theta}$: perturbed parameter vector ($N_p \times 1$)
$\boldsymbol{\Theta}$: perturbed parameter matrix ($N_p \times N_e$)
u : Meridional wind component [m s^{-1}]
u_* : Friction velocity [m s^{-1}]
U : Wind speed [m s^{-1}]
v : Zonal wind component [m s^{-1}]
V_z : Sky-view factor [-]
w : Water vapor mixing ratio [-]
w : Posterior particle weight
$\mathbf{x}$: State vector ($N_s \times 1$)
$\mathbf{y}$: Observation vector ($N_o \times 1$)
$\widehat{\mathbf{y}}$: Predicted observation vector ($N_o \times 1$)
$\widehat{\mathbf{Y}}$: Predicted observation matrix ($N_o \times N_e$)
z : Elevation or altitude [m]
z_0 : Roughness length of the snowpack = 0.001 [m]
Z : Geopotential height [m^2 s^{-2}]

Contents

Part I

Overview

Chapter 1

Introduction

1.1 Snow: An essential climate variable

1.1.1 Background

Snow, which on a seasonal basis can cover anywhere between less than 10% to over 40% of the Northern Hemisphere's land area, is one of the most rapidly varying natural surface features on Earth (Hall, 1988). At its maximal extent of approximately 4×10^7 km^2, around three times the extent of the Antarctic continent, this snow-cover is the largest single element of Earth's cryosphere in terms of area. While some of this snow is ephemeral, lasting only a few days or weeks, in cold regions much of the snow is seasonal, typically forming in autumn or winter and persisting for several months into spring or summer.

The primary controls on the distribution and variability of the snow-cover are topography, vegetation, precipitation, wind, radiation, humidity, temperature, and avalanching (Clark et al., 2011). Since these controls may vary strongly over a range of spatial and temporal scales, snow is often unevenly distributed across the terrain. This often becomes apparent during the snowmelt season when, even in relatively homogeneous terrain, some areas may quickly become bare while others remain snow-covered for weeks or even months longer. An example of this phenomenon is shown in Figure 1.1. This means that one can use the snow-cover melt out date, or the related snow-cover duration, as a proxy for the peak snow water equivalent at a given point in space. If the ablation rates are the same, then an area that melts out later will have had a deeper snowpack at the peak of the snow season than an area that melts out earlier. This is the simple but ingenious

idea behind snow reconstruction (Martinec and Rango, 1981). I will return to this idea in the problem formulation. Often the spatial patterns of the snow-cover duration are quite similar from year to year, and there lies much untapped potential in fully exploiting these patterns in snow modeling (Sturm and Wagner, 2010). In the rest of this section, I will briefly review some of the key properties of the snowpack.

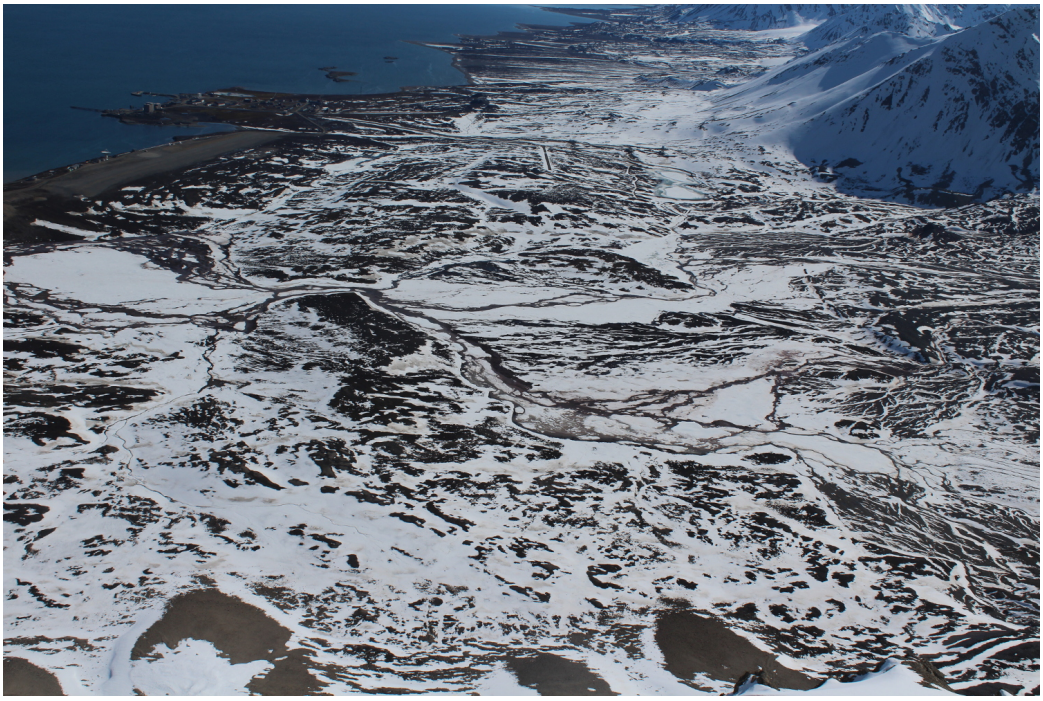


Figure 1.1: A patchy snow-cover in the Bayelva catchment near Ny-Ålesund, Svalbard, Norway, photographed by an automatic camera system on the 05/06/2016.

Primarily a mixture of ice crystals and air, snow is endowed with a unique combination of physical properties. Pure snow has a high albedo: it is a white surface in the visible part of the electromagnetic spectrum, and hence reflects the majority of any incoming insolation. Beyond the visible part of the electromagnetic spectrum, moving into the near and shortwave infrared wavelengths, snow is a surprisingly colorful medium (Dozier, 2011). The spectral reflectance of snow drops off rapidly from the near infrared (0.8 μm) to the shortwave infrared ($1.6 - 2.1$ μm) where snow is almost black (absorbs most insolation). This sets snow apart from most other naturally occurring surfaces and makes it possible to identify the snow-cover and many of its properties using multispectral satellite imagery.

Since the thermal conductivity of snow is low, being closer to that of air than ice (Sturm et al., 1997), snow has a strong insulating effect on the underlying ground. For comparison, the thermal conductivity of snow is several times lower than that of typical mineral soils (Zhang, 2005). The fact that snow is an excellent insulator implies that it modulates the ground thermal regime substantially. Effectively, the snowpack thermally decouples the ground from the atmosphere. The decoupling leads to strong differences between the near-surface air temperature (which will usually be close to the snow surface temperature) and the ground surface temperature. This difference is known as the nival offset (Trofaier et al., 2017). Whether or not the insulating effect of snow is warming or cooling for the ground depends on the duration, timing (i.e. onset and melt-out), and thickness of the seasonal snowpack (Zhang, 2005). A common effect is that snow will insulate the ground in winter such that ground surface temperatures are kept considerably higher than that of the cold winter air (Trofaier et al., 2017).

The seasonal snowpack has the capacity to store water in its frozen and to a lesser extent liquid form for months on end. At a typical end of season bulk density of 330 kg m^{-3}, a snowpack that is 3 meters deep has the same mass as around 1 meter of liquid water. Hydrologists will often refer to the snow water equivalent, usually abbreviated SWE, of a snowpack. This is simply the liquid water equivalent (depth or mass) that would result from melting the snowpack instantaneously. Thereby, the SWE depth (which I denote symbolically as D) is simply related to the snow depth (d) through

$$\rho_w D = \rho_s d \,, \tag{1.1}$$

where ρ_w is the density of fresh liquid water and ρ_s is the bulk density of the snowpack in question. Snow density is thus frequently expressed as the density relative to liquid water, i.e. $\widehat{\rho}_s = \rho_s / \rho_w$, and this is typically less than 0.5 (corresponding to 500 kg m^{-3}), being lowest for fresh snow. For perennial snow-cover, as the snow ages and compacts and the relative density exceeds 0.5, we move into firn and eventually almost pure ice. Herein we are primarily concerned with seasonal snow.

The high latent heat of fusion of water implies that it takes a lot of energy to melt snow. In fact, melting 1 meter of SWE takes around 335×10^6 J m^{-2}, which is enough to power a modern 10 W light bulb for a year. Moreover, it takes much more energy to melt snow than it does to heat it up. Snow temperature is bounded

from above at the melting point: $0°$C. Once it reaches that temperature, the snowpack is said to be 'ripe' and starts melting. It takes around 160 times more energy to melt a given volume of ripe snow than it does to raise the temperature of the same volume of snow (were it not ripe) by 1°C. An effect of the latent heat of fusion is that the air near the snow surface is rarely above freezing. As pointed out by Mott et al. (2013), this confinement effect due to the latent heat sink may have two conflicting effects. On the one hand, the confinement can lead to the development of stable internal boundary layers that inhibit the snowmelt. On the other hand, it may promote horizontal sensible heat advection over patchy snow-cover which enhances snowmelt.

Snowfall tends to smooth out microtopograhic relief in calm conditions (Filhol and Sturm, 2019). In addition, the surface of the snowpack usually has a low aerodynamic roughness length, on the order of 10^{-3} m (Andreas et al., 2010) which is around 100 times smaller than that of grass, whereby the snowpack typically damps turbulent exchange at the surface. Radiatively, snow exhibits a near constant high thermal emissivity of $\varepsilon_s = 0.99$ (Dozier and Warren, 1982). This implies not only that longwave radiative cooling is effective over snow, but also that the snowpack has an equally high longwave absorptivity. Thereby, downwelling longwave radiation from clouds can be a notable energy source for the snowpack even in winter (Zhang et al., 1996).

1.1.2 Snow in the climate system

The effect of this unique combination of physical properties is that the snowpack strongly modulates the near-surface energy, water, and carbon balance. Through land-atmosphere coupling these local modulations regulate the global radiation balance and hydrological cycle, making the snow-cover a driver of the atmospheric circulation and the associated climate (Henderson et al., 2018). The cooling effect of the high latitude snow-cover helps to maintain the equator to pole temperature gradient which powers the extratropical circulation through its influence on the jet stream (Hall, 1988). Moreover, it is known that anomalies in the Eurasian snow-cover exert a strong influence on the Indian summer monsoon, with years of high snow accumulation weakening the monsoon rainfall (Halder and Dirmeyer, 2017).

The snow-albedo feedback is a textbook example of a positive feedback in Earth's

climate system. An increase in greenhouse forcing leads to a retreat of the global snow-cover, reducing the planetary albedo, increasing shortwave absorptivity, and thus further enhancing global warming (e.g. Hall and Qu, 2006). The snow-albedo feedback is positive because it amplifies, rather than dampens, the original change (be it an increase or decrease in radiative forcing). It is closely related to the sea-ice albedo feedback and the two are often lumped together in the snow-ice albedo feedback mechanism. The snow-albedo feedback may be exacerbated by the radiative forcing from light absorbing particles (e.g. soot, dust, or algae) present in the snowpack. These particles directly darken the snow, enhancing the snowmelt, which in turn indirectly darkens the snow through accelerated snow grain size growth (Skiles et al., 2018). It is worth emphasizing that large inter-model spread currently exists in estimates of the strength of the snow-albedo feedback, indicating that this feedback remains a considerable source of uncertainty in future climate projections (Qu and Hall, 2014). On a planetary scale, simulated climate is so sensitive to snow albedo that changing the parametrized snow grain shape from spherical to irregular, which results in an increase in albedo of around 0.025, yields a decrease in modeled global mean surface temperature of over $1°$C (Räisänen et al., 2017). More regionally, the snow-albedo feedback is also one of the main contributors to polar amplification (Stuecker et al., 2018).

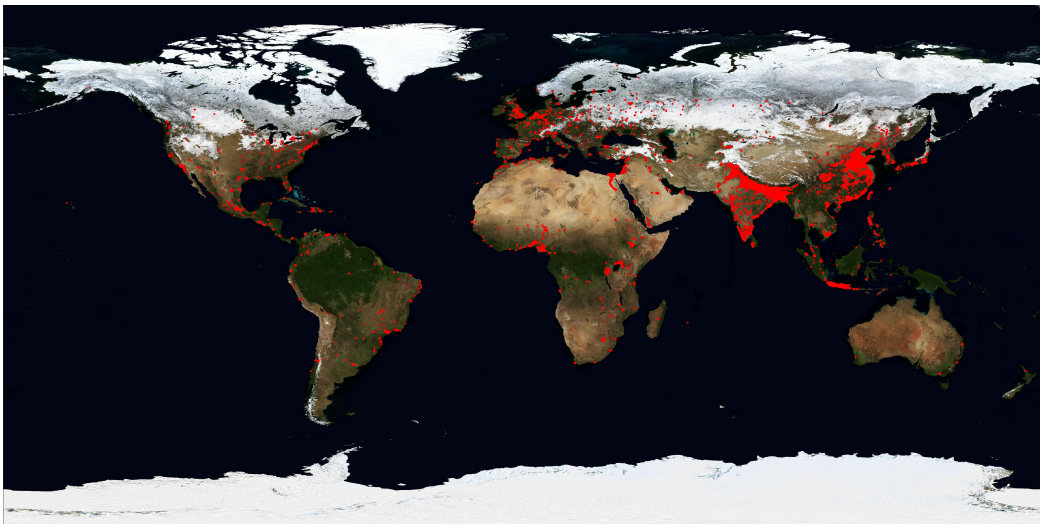


Figure 1.2: MODIS true color composite for January (northern hemisphere) and August (southern hemisphere) from Stöckli et al. (2004). Areas with human population estimates for 2020 exceeding 500 people per square km (based on CIESIN, 2017) are indicated by red dots.

Seasonally the snowpack can act as an important natural fresh water reservoir, typically storing water in the fall and winter and then releasing it again in spring and summer. Globally, in excess of one billion people currently rely on snowmelt runoff as an integral component of their freshwater supply (Barnett et al., 2005). An overview of the planet's main human population centers and the typical maximum extent of the seasonal snow-cover is provided in Figure 1.2. Much of this population lives in the snow sensitive basins around $25 - 45°$N, near the subtropical high pressure centers, such as the Indus River basin, the Euphrates and Tigris basin, and the Colorado River basin (Mankin et al., 2015). Mountains in particular can play an important role as natural 'water towers' (Viviroli et al., 2007), as they capture large amounts of snowfall and release this as runoff to downstream lowland areas in the snowmelt season. In North America, for example, mountains only cover 24% of the land area, but it has recently been estimated that they contain as much as 60% of the continent's SWE (Wrzesien et al., 2018). Concerningly, many of the snow-sensitive basins have been projected to have a decline in snow supply in the coming century under a business as usual scenario (Mankin et al., 2015). Climatic warming can shift the phase of precipitation from snow to rain and lead to earlier and slower snowmelt (Musselman et al., 2017). Such changes can impede runoff efficiency and shift peak streamflow from summer and autumn, when demand is the highest, to early spring and winter (Stewart et al., 2005). An ongoing long-term snowpack depletion can already be observed in several mountain ranges, such as those of western North America (Mote et al., 2018) and the Near East (Yılmaz et al., 2019).

As mentioned, through its physical properties, snow exerts a strong influence on the surface energy balance and the thermal regime of the underlying ground. Snow is thus a key control on the state and extent of permafrost (Zhang, 2005). In cold regions permafrost is a commonly occurring phenomenon which is defined thermally as ground that is permanently (i.e. for at least two consecutive years) below $0°$C. A recent estimate indicates that the permafrost region underlies around a quarter of the Northern hemisphere's exposed land area (Obu et al., 2019). A map of the permafrost region, on which the aforementioned recent estimate is based, is provided in Figure 1.3. It is estimated that the permafrost region holds approximately 1300 Pg of soil organic carbon (Hugelius et al., 2014). For comparison, the atmosphere currently holds just under 900 Pg of carbon (equivalent to 414 ppm CO_2), an increase of around 300 Pg since 1850 (when CO_2 concentrations were around 280 ppm). Around 1000 Pg of this permafrost carbon is stored in the

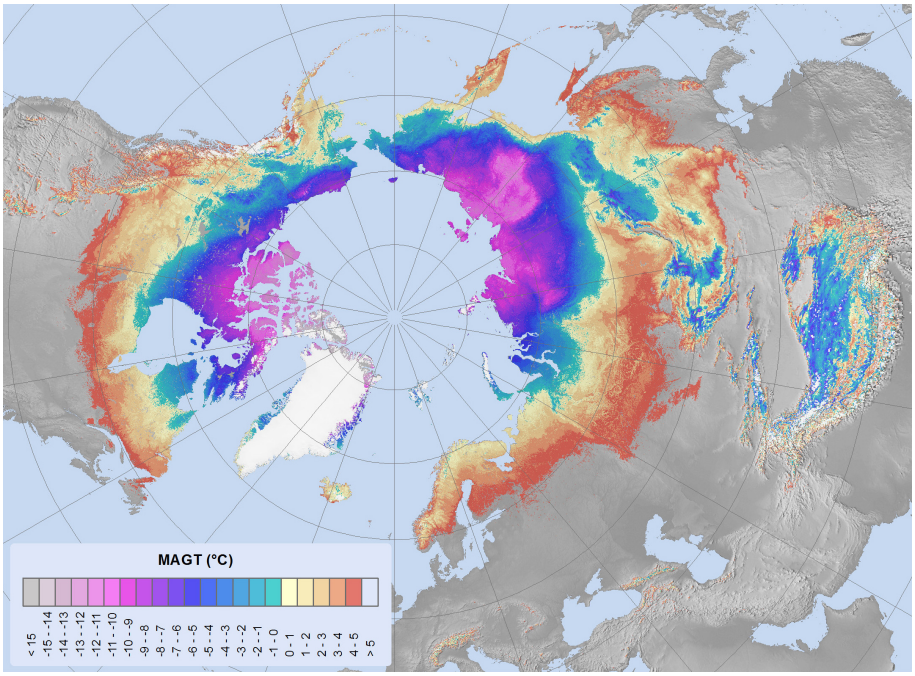


Figure 1.3: Estimated average mean annual ground temperature at the top of the permafrost in the Northern Hemisphere. Adapted with permission from Obu et al. (2019).

upper 3 meters of soil. It is believed that the rest of Earth's biomes (excluding the Arctic and the boreal zone) contain 2050 Pg of carbon in the upper 3 meters of soil (Jobbággy and Jackson, 2000). As such, the areal density of soil organic carbon is around twice as high in the permafrost region. Current estimates indicate that we can safely emit another -30 to 140 Pg of carbon (Peters, 2018), depending on (among other things) developments in carbon capture and storage capabilities, to stay below $1.5°$C of global warming (IPCC, 2018).

Worryingly, due to amplified warming at high latitudes, vast tracts of this permafrost soil are quite vulnerable to thaw (Chadburn et al., 2017). These developments are concerning given that such a thaw would mobilize previously frozen carbon to the atmosphere which would further enhance global warming (Schuur et al., 2015). There is large uncertainty as to the magnitude of this so-called permafrost-carbon feedback tied to how much carbon will be mobilized. This, in turn, depends on the spatial distribution of soil organic carbon as well as the location specific rate of thawing. For the latter, snow plays an important role given its aforementioned effect on the surface energy balance and the thermal regime of

the underlying ground (Zhang, 2005). In particular, the spatiotemporal distribution of SWE exerts a strong influence on the spatial extent and thermal state of permafrost, and, in turn, the permafrost-carbon feedback. Better SWE estimation techniques would directly improve ongoing satellite-based permafrost modeling efforts (e.g. Westermann et al., 2017) since SWE is an essential variable in the exercise of simulating snow's insulating effects on ground surface temperatures (Trofaier et al., 2017).

1.1.3 The socioeconomic value of snow

For those concerned with economics, a monetary valuation could also be placed on the global seasonal snowpack as a resource. Such valuations would likely lie in the range of trillions of US dollars (henceforth $), especially when the risk of potential losses are factored in (Sturm et al., 2017). Snow currently helps to sustain a thriving multi-billion dollar outdoor winter tourism industry around the world (e.g. Burakowski and Magnusson, 2012). Many of the tourist resorts scattered around the world's mountains that rely on winter sports are under threat from climate change. For example, resort skiing on natural snow will likely no longer be possible by the end of the century in both the French Alps and the Pyrenees (Spandre et al., 2019). Snowmelt also helps to support freshwater required to sustain agricultural activity. In the western US, for example, snowmelt runoff provides more than three quarters of the total freshwater supply (Bales et al., 2006). In California alone, the Sierra Nevada snowpack helps support agricultural production worth more than $50 billion per year (CDFA, 2019). This agriculture is highly sensitive to the extent of the seasonal snowpack which may vary considerably as shown in Figure 1.4. Snowmelt runoff can also play an important role in hydropower generation. In Norway, around 50% of the precipitation falls as snow and almost 100% of the electricity supply is generated through hydropower (Winther and Hall, 1999).

The snowpack can also represent a hazard in the cases where it melts quickly, potentially leading to flooding. Such rapid melt may for example be triggered by rain on snow events (Musselman et al., 2018). Snowmelt flooding can not only disrupt the economy but may also put lives in danger, as was the case in the recent Oroville dam crisis where almost 200 000 people had to be evacuated (Vahedifard et al., 2017). As emphasized by Sturm et al. (2017), the valuation of snow should

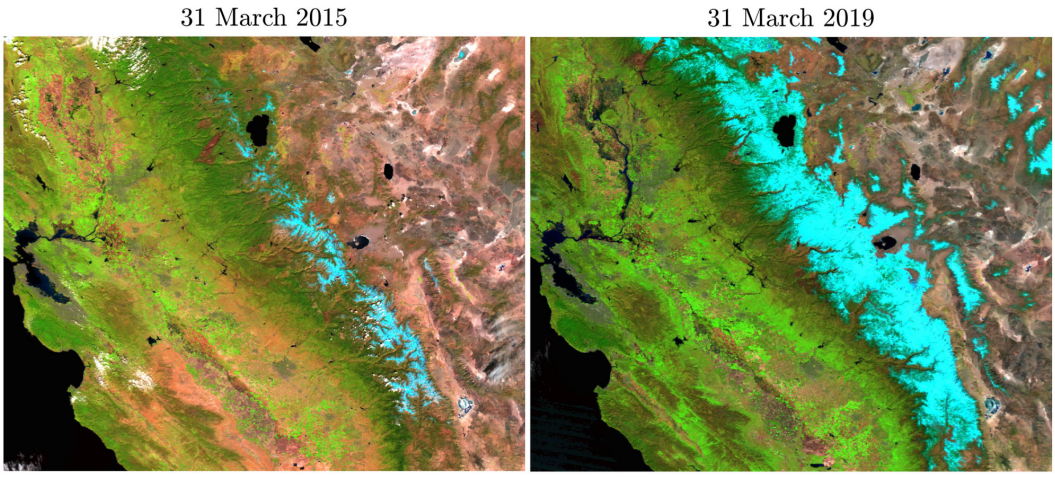


Figure 1.4: False color images of the Californian Sierra Nevada taken four years apart, 31 March 2015 (left panel) and 31 March 2019 (right panel), by the Terra MODIS sensor (MOD09GA v06 product; Vermote and Wolfe, 2015). Snow appears turquoise in this band combination.

also include the climate service that snow provides by making our planet habitable primarily through its effect on the planetary albedo. The value that snow and the cryosphere as a whole provides in effectively being the planet's air conditioner dwarfs the remaining contributions to the valuation (Euskirchen et al., 2013).

1.1.4 Snow in the context of global warming

The cold regions of our planet, situated at high latitudes and elevations, are experiencing an amplified anthropogenic global warming (Chapin et al., 2005; Serreze et al., 2009; Pepin et al., 2015). This amplification is clearly seen in the observational record, where temperatures in the Arctic have risen by around 1°C more than the globe as a whole during the last three decades (see Figure 1.5). Cold region amplification can be tied to a location specific concoction of climate feedback mechanisms, where the lapse-rate feedback is the most important and the Planck and snow-ice albedo feedbacks play secondary roles (Pithan and Mauritsen, 2014; Stuecker et al., 2018). These feedbacks are closely tied to the fact that these regions are cold and exhibit a seasonally or perennially snow-covered surface. Under amplification, the seasonal snow-cover is undergoing changes with large socioeconomic implications (e.g. Mote et al., 2018; Yılmaz et al., 2019; Spandre et al., 2019). Snowpack depletion can also help to amplify these changes through various feedback mechanisms (Duan et al., 2019).

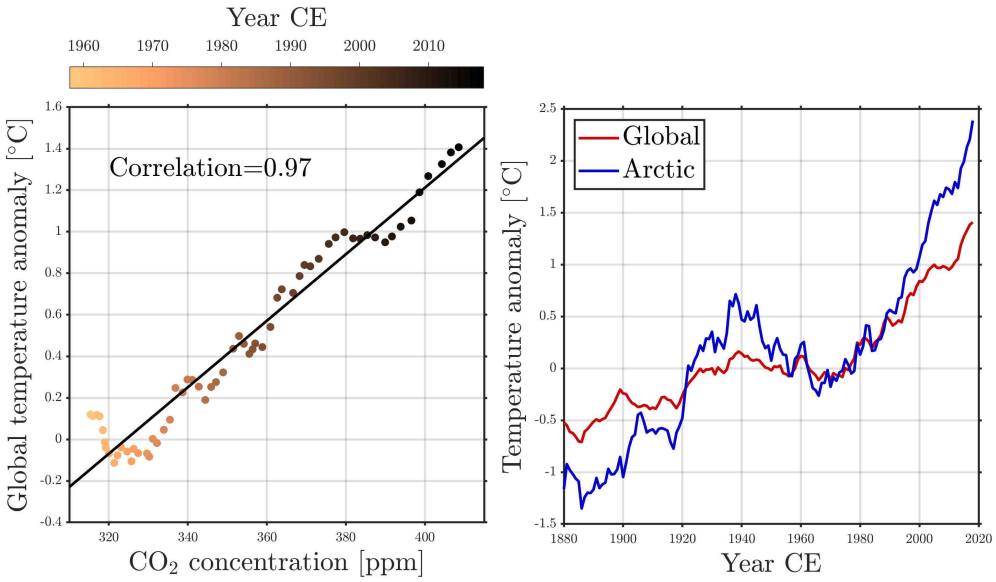


Figure 1.5: **Left panel:** Scatterplot showing the (5 year moving average) global mean surface temperature anomaly (GISSTEMP Team, 2019) versus the CO_2 concentration at the Mauna Loa Observatory in Hawaii (Keeling and Keeling, 2017) from 1958 until 2018. In the face of a causal physical mechanism, here the greenhouse effect, correlation can imply causation. **Right panel:** Time series of the (5 year moving average) global and Arctic ($> 64°$N) mean surface temperature anomaly (GISSTEMP Team, 2019) from 1880 until today.

On a broader scale, our planet as a whole is currently undergoing historically unprecedented changes related to the ongoing anthropogenic global warming (Diffenbaugh et al., 2017). At present, the average temperature at the surface of our planet has likely risen by over 1°C since the beginning of the industrial era. As predicted by Broecker (1975), this global warming has gone hand in hand (see Figure 1.5) with a growth in the concentration of carbon dioxide in the atmosphere, which has risen by more than 100 parts per million (ppm) since the 1800s to now sit over 400 ppm. Based on simple accounting we know that this is in large part due to human emissions of greenhouse gases, primarily through the burning of fossil fuels. We have understood since the slow beginnings of this growth that the interaction between such gases and terrestrial radiation cause the well known greenhouse effect (Fleming, 1999; Benestad, 2017). Historically this heat trapping effect is what has kept our planet habitable, allowing for the evolution and development of humans. Now we are pushing the boundaries and moving into terra incognita with our continued alteration of the constituents of the atmosphere. The fact that we are driving global warming through greenhouse forcing leading to a plethora of climatic changes is well known (c.f. IPCC, 2013, and references

therein). The main source of uncertainty lies in how far future emissions will push global warming, and that uncertainty primarily depends upon which emission pathway we follow.

There exists certain positive feedback mechanisms in the Earth system, such as the permafrost-carbon feedback, that are not fully understood and may change the magnitude of global warming. Together, these self-reinforcing mechanisms can help to push the system beyond a threshold that may lock us onto a pathway towards much warmer conditions, a so-called hothouse Earth. A rapid, on the scale of a few centuries, transition to such a regime would undoubtedly be disruptive to human civilization as a whole. Steffen et al. (2018) suggest that the risk of a tipping cascade towards such a hothouse might be considerable already at a temperature rise of 2°C above pre-industrial levels. As such, there is currently great interest surrounding pathways to limit warming below 1.5°C to reduce such risks (IPCC, 2018). To be able to more accurately assign such pathways it is important that we continue the scientific effort to improve our understanding of the various climate feedback mechanisms that influence the warming response to human emissions. It is nonetheless vital that we begin to act and cut emissions now if we are to limit warming to safe levels, as emphasized by IPCC (2018). Concerningly, due to uncertainty surrounding various positive feedback mechanisms, it is just as possible that we are underestimating, rather than overestimating, our planet's climate sensitivity (e.g. Schneider et al., 2019).

1.2 Estimating SWE: A grand challenge

Following the preceding discussion it is clear that snow is a vital component of Earth's climate system. As such, several snow-related variables have been designated the status of essential climate variables (ECV) by the Global Climate Observing System (GCOS; Bojinski et al., 2014). The snow-related ECVs are the: albedo, fractional snow-covered area (fSCA), snow depth, and SWE. Given their designation as ECVs, the ability to accurately estimate these four variables remains a 'grand challenge' for weather and climate prediction (Hock et al., 2017). In particular, the problem of estimating the distribution of SWE in mountainous terrain has been coined the most important unsolved problem in snow hydrology (Dozier et al., 2016). In this thesis, I will present advances in my efforts towards tackling this SWE estimation problem. In this section, I will outline this problem

and the challenges surrounding it.

1.2.1 Modeling snow

The bulk state of a snowpack, typically characterized through the SWE, is the result of the sequence of accumulation and ablation events that have occurred since the onset of that snowpack. The estimation of SWE through mechanistic modeling thus requires accurately accounting for the spatiotemporal distribution of these events. As noted by Sturm (2015), the blueprints for spatially distributed hydrological modeling, which could easily be translated to the more specific case of snow modeling, have existed for half a century thanks to the work of Freeze and Harlan (1969). Interestingly the emergence of this blueprint closely coincided with one of the earliest published descriptions of a snow energy-balance model (Anderson, 1968). Since then, several detailed and potentially distributed snow models, such as SNOWPACK (Bartelt and Lehning, 2002) and Crocus (Vionnet et al., 2012), were developed over time as computing power increased. Many snow models now have a large user base and are widely employed in research as well as to address more operational questions.

Despite their wide range of uses, distributed hydrological models are typically fraught with issues. Beven (2001) categorizes these issues into five different problems. As pointed out by Sturm (2015), all of these problems extend to the more specific domain of snow modeling as well. The first of Beven's problems is nonlinearity: that the governing equations of a model, regardless of whether or not they are physically-based or conceptual, typically contain nonlinear terms. This is a problem because it can impose a strong sensitivity to initial and boundary conditions which may be poorly known, just as in numerical weather prediction (c.f. Bauer et al., 2015). For snow modeling the initial condition is not so much of a problem so long as the snowpack is seasonal given that then by definition the ground should be bare at the start of the hydrological year. The boundary conditions are more problematic: a snow model is very sensitive to errors in the input data, especially the hydrometeorological forcing (Raleigh et al., 2015). Nonlinearity may also imply strong heterogeneity at the subgrid scale, making it difficult to correctly sample subgrid scale variability.

The second, quite related, problem is that of scale. In the context of snow model-

ing the most important scale issues revolve around whether or not the model scale, measurement scale, and dominant process scale overlap sufficiently (Blöschl, 1999). Scale issues are also commonly referred to as problems of representativeness or commensurability. For example, for a model to adequately resolve the snow drifts (i.e., the areas with the longest snow-cover duration) in Figure 1.6, it would have to be run at a resolution of around 100 m. This scale is known as the hillslope-scale in hydrology and it corresponds to that required to generally resolve the dominant modes of snow variability (Clark et al., 2011). This scale typically resolves the main horizontal gradients in the surface energy balance, due for example, to terrain configuration. At the same time, at finer scales than this wind redistribution becomes increasingly important and this is a process that is difficult to resolve explicitly without coupling to expensive turbulence resolving large eddy simulations (e.g. Vionnet et al., 2017). The importance of representing hillslope-scale structures, which seem to organize stores and fluxes of mass and energy in the terrain, in Earth System Models was recently stressed by Fan et al. (2019).

The third problem raised in Beven (2001) is that of uniqueness or transferability. This relates to the fact that each basin under study has unique characteristics, such that transferring a model from one basin to another may be quite challenging. Ultimately this problem stems from the fact that there is rarely a closed-form unified physically-based set of governing equations describing the behaviour of the land surface system under study due to its intrinsic heterogeneity. This sets land surface and hydrological modeling, of which snow modeling is a subfield, apart from for example meteorology where a complete set of (albeit far from trivial) conservation equations exist (see e.g. Vallis, 2017). Recent efforts to address this uniqueness issue in hydrology have been undertaken by Clark et al. (2015) where a unified approach to hydrological modeling is presented. For snow modeling, an example of the transferability problem is that the ground heat flux is often ignored since in many environments it is deemed to be unimportant (e.g Cline et al., 1998). A strong exception is raised in the high-Arctic study of Aalstad et al. (2018) where permafrost underlying the snow entails that the ground heat flux is a strong energy sink for the snowpack during the snowmelt, implying that the model of Cline et al. (1998) would not be directly transferable to this domain. Conversely, the model of Aalstad et al. (2018) did not account for canopy effects (since it did not need to), and so it would not transfer well to the partially forested environment studied in Cline et al. (1998). Non-uniqueness or transferability is an important goal to strive for as it could improve the transfer of calibrated models to ungauged catchments.

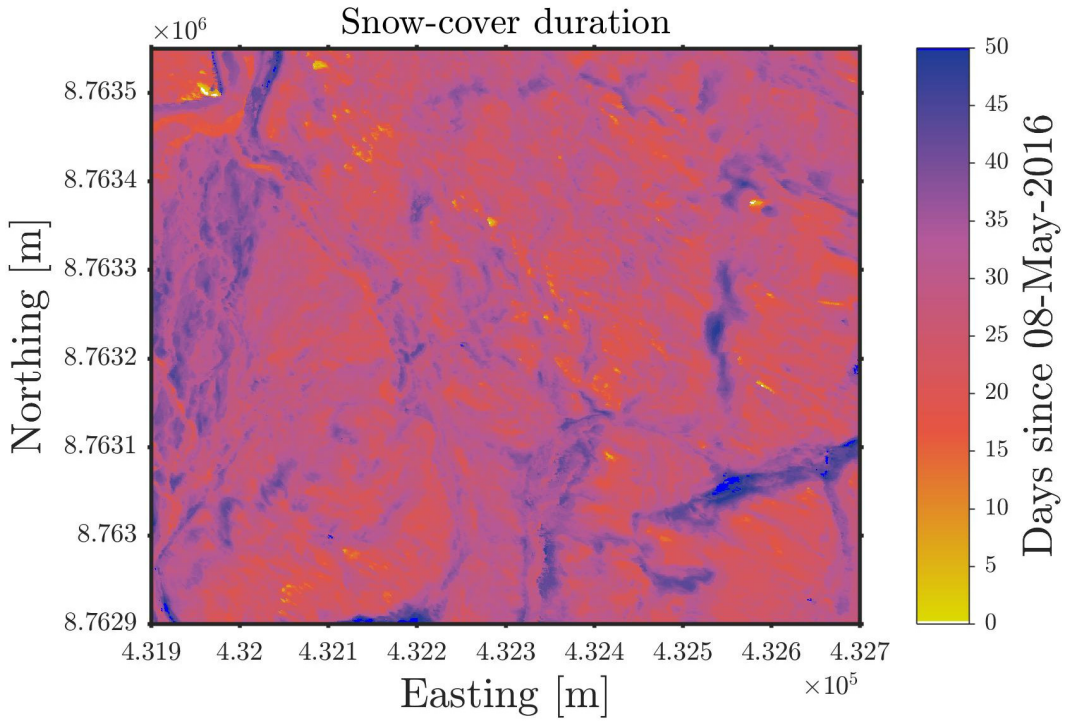


Figure 1.6: Snow-cover duration expressed as days since 08.05.2016 for the 2016 ablation season estimated using 0.5 m resolution time-lapse photography over a small $\simeq$1 km^2 area in the Bayelva catchment near Ny-Ålesund, Svalbard, Norway. Just within this small and relatively flat area, the snow-cover duration varies by almost 2 months.

The fourth problem discussed in Beven (2001) is that of equifinality. Briefly stated, equifinality pertains to the fact that typically several different model configurations can yield equally acceptable results. A much more comprehensive treatment of this issue is provided in Beven (2006). As a result of equifinality, the quest for an optimal or "best" model solution should abandoned. Instead, this notion should be replaced with a representation of the multiple model configurations that are behavioural, i.e. acceptable (with respect to validation or calibration data) to some predefined degree. In Aalstad et al. (2018) we provide a classical example of an equifinality problem in snow modeling when noting that an earlier melt-out date can be due to both a decrease in snowfall and an increase in snowmelt (or more generally ablation). This ambiguity can not be resolved with observations of the snow-cover depletion alone. A similar example is provided and discussed extensively in the study of Clark and Vrugt (2006).

The final problem discussed by Beven (2001), which is closely related to the issue of equifinality, is that of uncertainty. In its broadest sense, I take uncertainty to mean *unknown* error, rather than trivial (i.e. known and thus easily correctable) error. Following the discussion in Beven (2016), I distinguish between four types of uncertainty. The first is aleatory or random uncertainty, this is uncertainty with stationary statistical characteristics. An example would be the noise present in a measurement system. The second is that of epistemic uncertainty. This is related to lack of knowledge about the modeled system (model structural error) and forcing data as well as disinformation in observations. The third type of uncertainty is language uncertainty, which is related to what a statement or quantity (such as uncertainty itself) actually means. The final type of uncertainty is ontological uncertainty which is related with different systems of belief. An example of this could be different interpretations of probability in statistics (see e.g. Lyons and Wardle, 2018). There may be considerable overlap between the latter two types of uncertainty, since personal definitions may be tied to beliefs held by a practitioner. If we can agree on a common scientific framework, then the two last types of uncertainty can probably be ignored, such that we are left with aleatory and epistemic uncertainty. The dominant uncertainties in distributed snow modeling are typically epistemic and related to the forcing data (Slater et al., 2013; Raleigh et al., 2015).

1.2.2 Measuring snow

Given all of these problems with distributed hydrological models in general and snow models in particular, observations could be a helpful alternative. Unfortunately, observations are themselves fraught with some of these very same problems. Observations suffer from sampling issues due to both non-linear phenomena and scaling issues. For example, manual in-situ sampling of the typically highly variable distribution of SWE (Clark et al., 2011) is both challenging and impractical. This is due to the limited spatial extent, spacing, and support (Blöschl, 1999) of snow probe surveys typically employed for that purpose. An example of such a snow survey conducted on Svalbard is provided in Figure 1.7. This survey involved sampling SWE along a ski transect of around 44 points which took several hours to complete. For that particular survey, the weather was quite pleasant and forgiving. This was by no means a guarantee. In addition, the survey involved

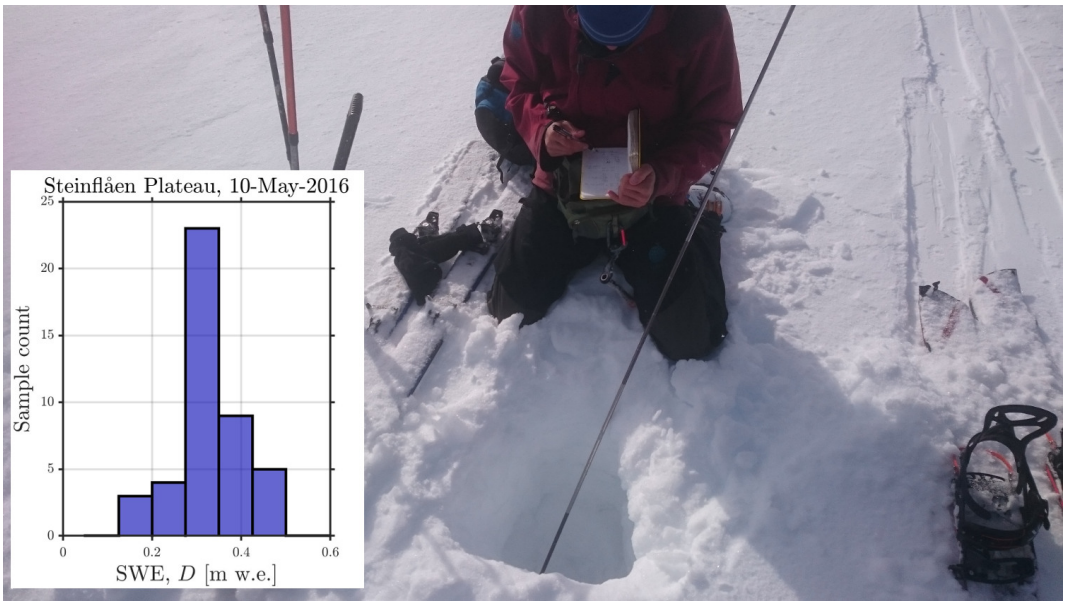


Figure 1.7: Measuring SWE using snow probes and a density sampler along a ski transect at the Steinfåen plateau near Ny-Ålesund, Svalbard 10.05.2016 (Photograph: K. Aalstad). Inset figure: Histogram of the sampled SWE distribution along the entire transect based on 44 samples.

passing through terrain with some avalanche hazard and potentially polar bears. The point here is that manually sampling SWE in the terrain is not only tedious and time consuming but also potentially dangerous, especially in steep terrain. Manual SWE surveys are also uncertain, due to the issues with snow probing as well as difficulties in accurately sampling the bulk snow density. These problems can both be exacerbated by the presence of thick ice layers that are difficult to penetrate.

Other terrestrial measurement techniques can overcome some of the issues with manual snow surveys. Terrestrial laser scanning, for example, allows one to estimate snapshots of spatially distributed snow depth (e.g. Revuelto et al., 2017). The technique works by taking the difference between a snow-covered digital elevation model (DEM) and a reference snow free DEM, both estimated from scans over the same area. Examples of such scans are provided in Figure 1.8. These scans can often have gaps due to terrain obstructions in the field of view of the sensors. Moreover, they usually provide estimates with a limited (typically only up to a few km^2) spatial extent and require quite some post processing. Most importantly, terrestrial laser scanning only provides an estimate of snow depth,

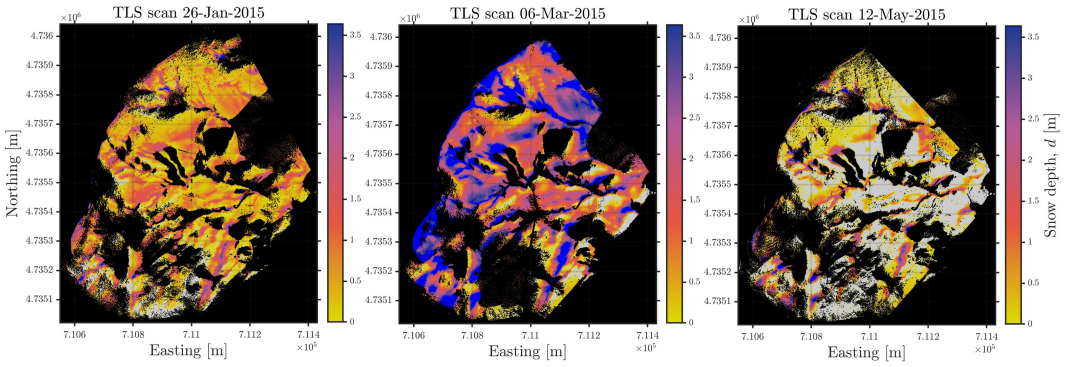


Figure 1.8: Spatially distributed snow depth fields (1 meter resolution) estimated using terrestrial laser scans at the Izas experimental catchment in the Spanish Pyrenees during the 2015 snow season. Gray: bare ground, black: missing data. Data courtesy of Revuelto et al. (2017).

manual snow density measurements are still needed to obtain SWE. Ground penetrating radar is also a valuable tool for estimating spatially distributed snow depth (e.g. Gisnås et al., 2014), but it too suffers from the inability to retrieve SWE on its own. Spatially distributed snow depth can also be estimated by combining time-lapse photography with fixed snow depth measuring stakes planted in the terrain (e.g. Parajka et al., 2012), however this technique is difficult to automate. Snow depth can also be measured continuously using sonic ranging devices (e.g. Boike et al., 2018). Using other sensors such as snow pillows and gamma ray sensors SWE can be retrieved directly (Smith et al., 2017), but these sensors have the limitation of being stationary and thereby only measure a single fixed support. Less conventional techniques such as global positioning system (GPS) interferometric reflectometry (McCreight et al., 2014) have also emerged as potentially promising ways to estimate SWE over a larger (1000 m^2) area. Crucially, all of these terrestrial techniques can only monitor a limited spatial extent. In addition, each measurement technique has a specific set of uncertainties associated with it.

SWE-related variables can also be retrieved using airborne remote sensing. Airborne LIght Detection And Ranging (LIDAR), for example, allows for the generation of DEMs multiple times during the snow season, allowing for the estimation of high-resolution spatially distributed snow depth by DEM differencing. For example, the Airborne Snow Observatory (ASO; Painter et al., 2016), a coupled LIDAR and imaging spectrometer flown on-board an airplane, is able to retrieve spatially distributed SWE at unprecedented spatial resolution. This is achieved by

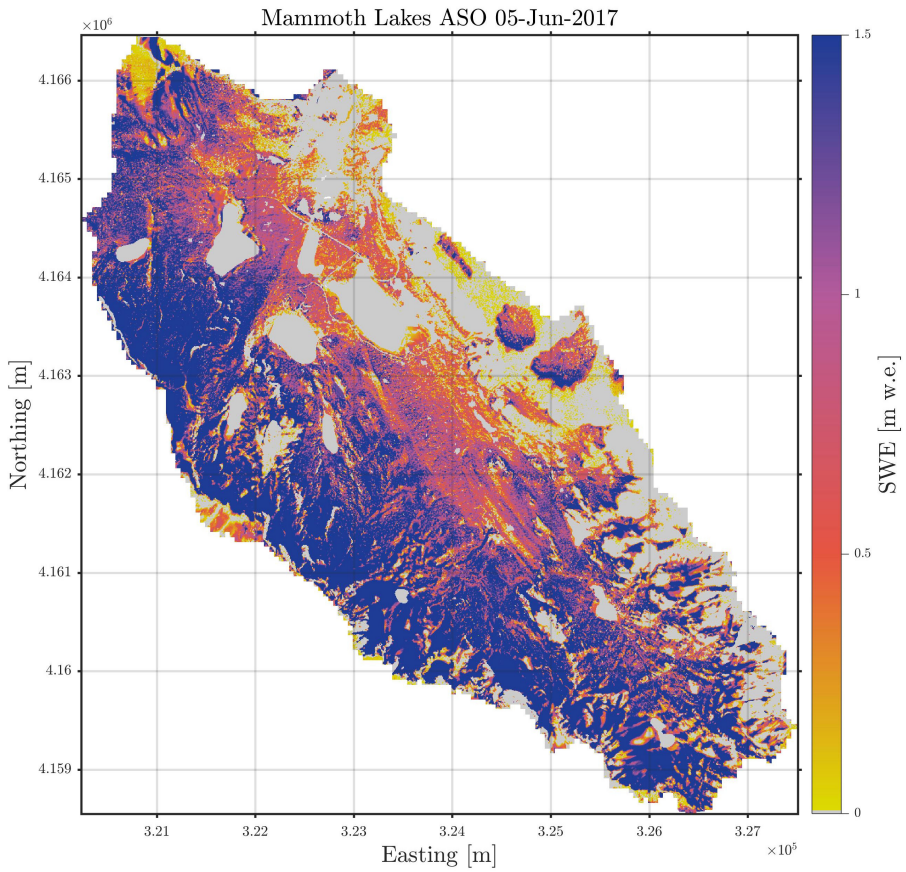


Figure 1.9: SWE distribution (resampled to 3 m resolution) over the Mammoth Lakes Basin in the Californian Sierra Nevada retrieved from the ASO on the 05.06.2017. Gray: zero SWE, blue: maximum (1.5 m w.e.) SWE. Raw data courtesy of NASA-JPL.

combining the LIDAR derived snow depth with constrained model estimates of snow density. An example of a detailed SWE map retrieved from the ASO is provided in Figure 1.9. Currently the ASO is mainly flown in the western US, and it is the first system capable of providing detailed maps of SWE across entire mountain basins. Airborne interferometric Synthetic Aperture Radar (SAR) systems such as the Glacier and Ice Surface Topography Interferometer (GLISTIN) have also successfully been employed to retrieve snow depth at high-resolution (Moller et al., 2017). Another example of airborne SWE retrieval is the use of gamma ray remote sensing, as is done operationally by the National Operational Hydrologic Remote Sensing Center (NOHRSC) in the US using the Gamma Radiation Detection System (Cline et al., 2009). Other examples of airborne remote sensing

of SWE, including passive and active microwave, are also provided in Cline et al. (2009). Recently, structure from motion technology using unmanned aerial vehicles has emerged as another promising approach to retrieve snow depth at high spatial resolution (Bühler et al., 2016). The use of airborne remote sensing greatly extends the capabilities of SWE and snow depth measurements in terms of spatial extent when compared to terrestrial approaches. There will nonetheless always be uncertainties inherent in the respective retrieval algorithms that are employed. No measurement is error free. As these techniques require flying some form of aircraft, spatiotemporal coverage remains limited. Still, given the combined high spatial resolution and considerable spatial extent of many of the airborne remote sensing SWE or snow depth products these can be invaluable for model validation, calibration, and data assimilation at the basin scale.

To further improve spatial coverage it is necessary to observe Earth from a different perspective, namely from space-borne satellites. We will restrict our attention to polar orbiting, rather than geostationary, satellites as these provide coverage at high latitudes. Most techniques for satellite remote sensing of snow rely on electromagnetic radiation and its interaction with the snowpack. A notable exception is gravimetric remote sensing, such as the twin satellites that make up the Gravity Recovery and Climate Experiment (GRACE). GRACE allows us to estimate changes in Earth's gravitational field which can be used to, among other things, retrieve terrestrial water storage anomalies (TWSA) at pixel resolutions on the order of 100 km. Given the gravimetric nature of the retrieval algorithm, these retrievals are available independent of cloud cover. An example of global monthly TWSA retrieved from GRACE is shown in Figure 1.10. Provided that changes in the other components of this storage (glaciers, surface water, and groundwater) are known, retrievals of TWSA can be used to help constrain modeled estimates of SWE (Niu et al., 2007). Retrievals of TWSA have also been used to help monitor climatological trends in the state of the seasonal snowpack (e.g. Yılmaz et al., 2019). Still it turns out that accurately decomposing the TWSA signal is quite challenging since, in addition to the SWE, the other components are generally also poorly known. Moreover, the spatial resolution is quite coarse and precludes hillslope-scale analysis.

Space-borne passive microwave satellite sensors, such as AMSR-E and AMSR2, also have the ability to retrieve SWE at resolutions on the order of 25 km. Passive

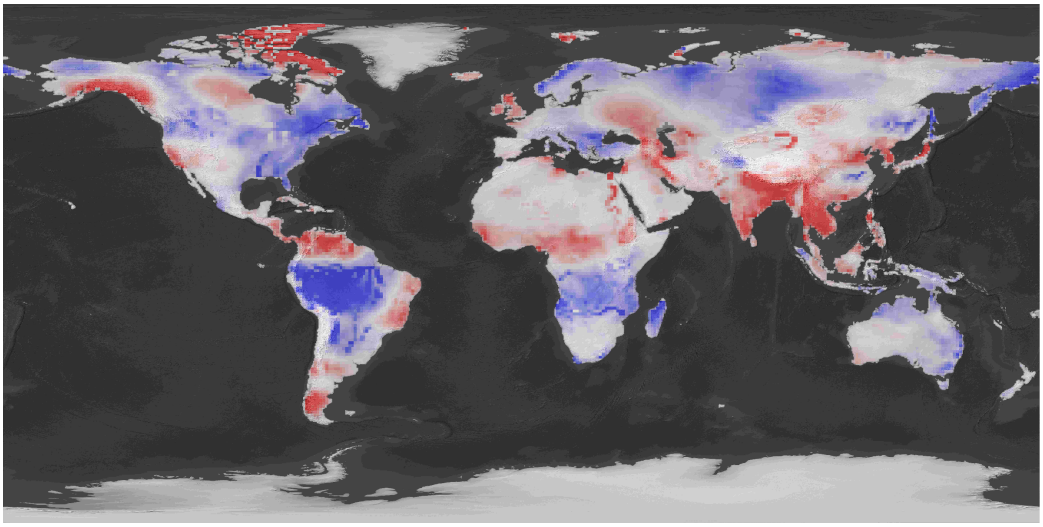


Figure 1.10: Monthly average gravimetric estimate of the terrestrial water storage anomaly from GRACE at 1° resolution for April 2015. Color scale: blue/red=±25 mm water equivalent.

microwave SWE retrieval algorithms are based either on relatively simple empirical relationships (Chang et al., 1987) or considerably more complex physically-based inversions (Takala et al., 2011). All the retrieval algorithms exploit the fact that the attenuation of microwave radiation emitted by the ground underlying the snow increases as a function of frequency and snow grain size due to volume scattering. Passive microwave remote sensing retrievals have been used extensively to constrain modeled estimates of SWE (De Lannoy et al., 2012; Li et al., 2017). They have also been used in regionally surveying the climatological state of the seasonal snowpack (Yılmaz et al., 2019). Unfortunately the signal that passive microwave relies on saturates in wet and deep snowpacks as well as in areas with considerable relief or dense forests (Foster et al., 2005). The retrieval algorithms are thus quite unreliable in mountainous terrain (Bormann et al., 2018). Note for example how the retrieved SWE is unexpectedly low for several mountainous regions (e.g. the European Alps and the Scandes) in Figure 1.11, this is because of the breakdown in the performance of the retrieval algorithm. Instead the retrieved SWE is at a maximum in the flatter terrain of the tundra in north east Siberia. Passive microwave sensors do have the advantage that they are able to sense the surface independent of cloud cover. Summarizing, although able to to function in all weather conditions, passive microwave remote sensing is not only too coarse to be employed at the hillslope scale but it is also plagued by uncertainties in many environments. Active microwave (i.e. radar) satellite sensors have mainly been

used to retrieve wet snow status at the hillslope scale (Baghdadi et al., 1996), but they undoubtedly have more potential (Moller et al., 2017; Lievens et al., 2019).

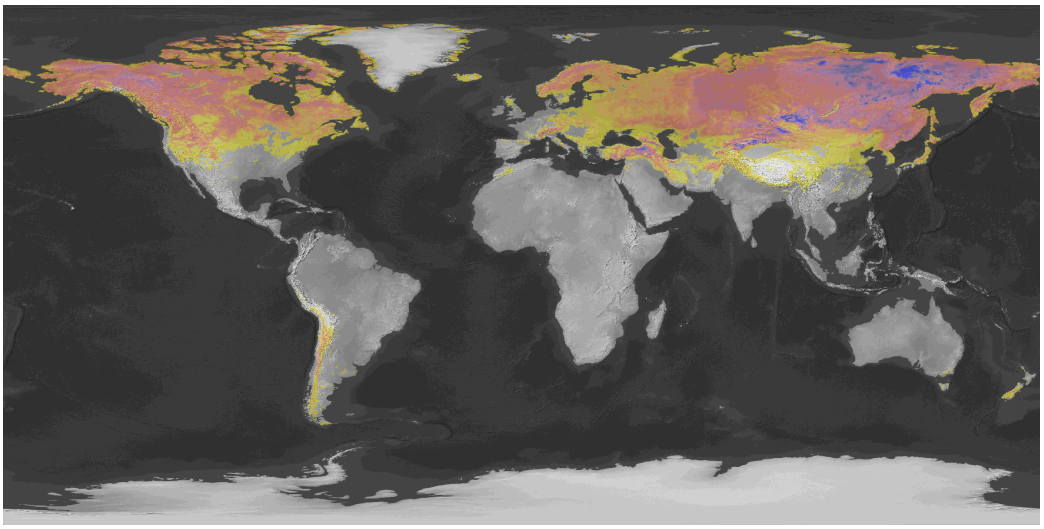


Figure 1.11: Monthly averaged passive microwave estimate of SWE from AMSR2 at 0.25° resolution for February 2017 (Northern Hemisphere) and August 2017 (Southern Hemisphere). Color scale: yellow-blue=1-20 mm water equivalent.

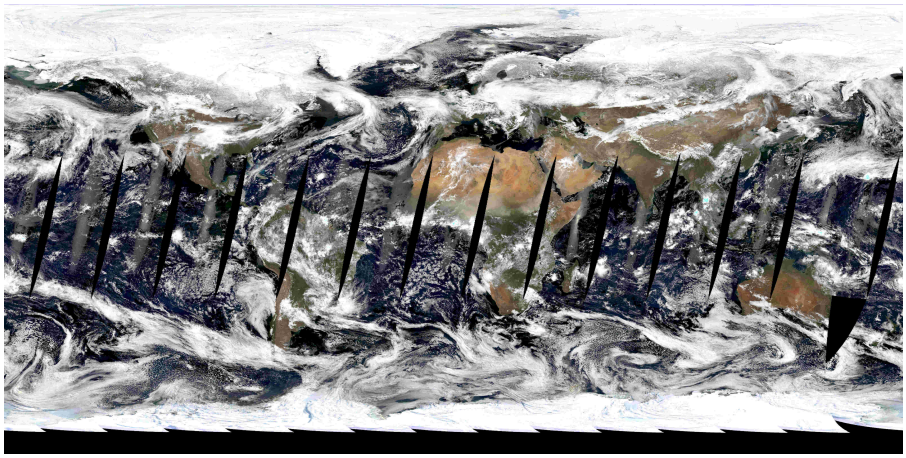
Retrievals from optical satellite sensors are based on sensing shortwave radiation that is reflected from the surface of the Earth. This radiation may either have originated as natural insolation, i.e. shortwave radiation emitted from the sun, or from the sensor itself. Passive optical sensors, which measure reflected insolation in various spectral bands, are the most common. These sensors exploit the spectral signature of snow and other natural surfaces to retrieve fSCA, albedo, snow grain size, and impurity concentration (Dozier et al., 2009). Snow is highly reflective in the visible but nearly opaque in the shortwave infrared which sets it apart from most other naturally occurring surfaces. This can be appreciated in Figure 1.12 where snow in the visible bands is quite difficult to differentiate from clouds, whereas when the same global image is plotted in a band combination that includes a shortwave infrared band, snow-covered mountain ranges such as the entire Californian Sierra Nevada and parts of the European Alps stick out quite clearly. Examples of passive optical sensors that are commonly used to retrieve primarily fSCA include the moderate resolution imaging spectroradiometer (MODIS) on-board the Terra and Aqua satellites as well as the higher resolution

Landsat and Sentinel-2 constellations (see e.g. Aalstad et al., 2019a).

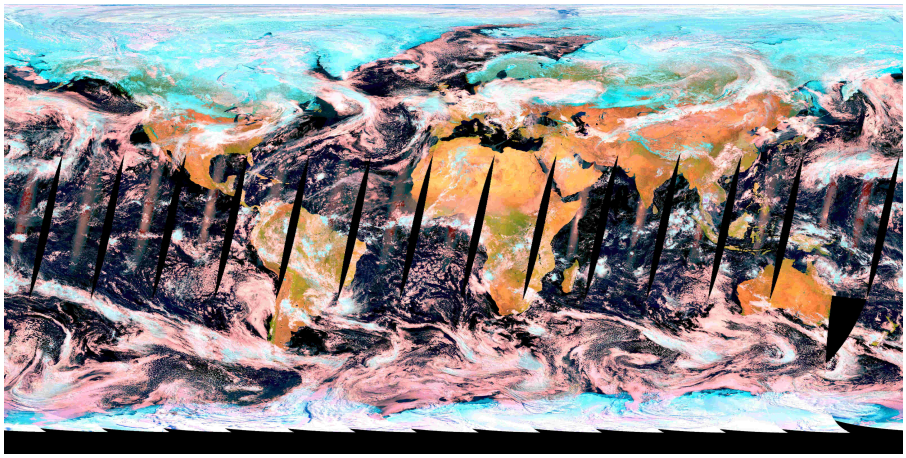
If passive optical sensors have the ability to observe the same area with multiple different view angles at relatively short time intervals, i.e. so-called stereo imaging capability, then they can also be used to construct high-resolution DEMs. By differencing summer and winter DEMs high-resolution snow depth maps can then be derived as was done in the study of Marti et al. (2016) using the Pléiades satellite. Unfortunately the freely available stereo capable satellite sensors (such as ASTER) currently do not provide the resolution and accuracy required for stereo snow depth reconstruction. Active optical satellite sensors, such as space-borne LIDAR, can also be used to reconstruct snow depths at relatively high-resolution if a sufficiently accurate reference DEM is available. An example of this technique is provided in the study of Treichler and Kääb (2017) using the ICESat satellite. A downside with space-borne LIDAR is the relatively short data record provided by ICESat which is no longer in orbit. This may improve with the recent launch of the follow on mission ICESat2. All optical sensors suffer from the fact that they can not observe the surface in the cases where an obscuring cloud cover is present. In addition, passive optical sensors can not sense at night (including polar night) in the absence of isolation. Moreover, as always, errors are propagated through the retrieval algorithms. None of the optical retrieval algorithms provide direct estimates of SWE. They do, however, allow us to retrieve information concerning a wide breadth of other important snow properties at the hillslope-scale. This information could serve as a valuable constraint for snow models.

1.3 Data assimilation: Fusing models & observations

It should be clear by now that neither snow models nor observations on their own are enough to solve the problem of accurately estimating SWE at the hillslope-scale for an arbitrary location. As discussed, snow models suffer from several problems which were categorized into five categories following Beven (2001): nonlinearity, scale, uniqueness, equifinality, and uncertainty. We recognize that perhaps the most serious problem for snow models is their sensitivity to uncertainty in the input hydrometeorological forcing data that is used to drive them (Slater et al., 2013; Raleigh et al., 2015). From the discussion in the preceding section it is clear that observations suffer from many of the same problems that models do. Moreover, terrestrial and airborne observations do not provide the



(a)



(b)

Figure 1.12: Terra MODIS enhanced true color [R,G,B] (panel a) and false color [SWIR1,NIR,R] (panel b) images sensed 01.04.2018 at $0.05°$ resolution. Based on the MOD09CMG product. Snow and ice appears turquoise in the seocnd image (panel b).

global coverage necessary to estimate SWE, or related variables, everywhere on Earth. This leaves us with satellite remote sensing of snow. Here many techniques are available, and these can be split into three categories: gravimetric, microwave, and optical remote sensing. Out of these it is primarily the passive optical sensors that offer the spatiotemporal resolution and breadth of information needed to quantify essential snowpack properties at the hillslope scale (Dozier et al., 2016; Bormann et al., 2018). Unfortunately these sensors also have their downsides as they can not retrieve SWE directly and are prone to data gaps in bad weather due to clouds.

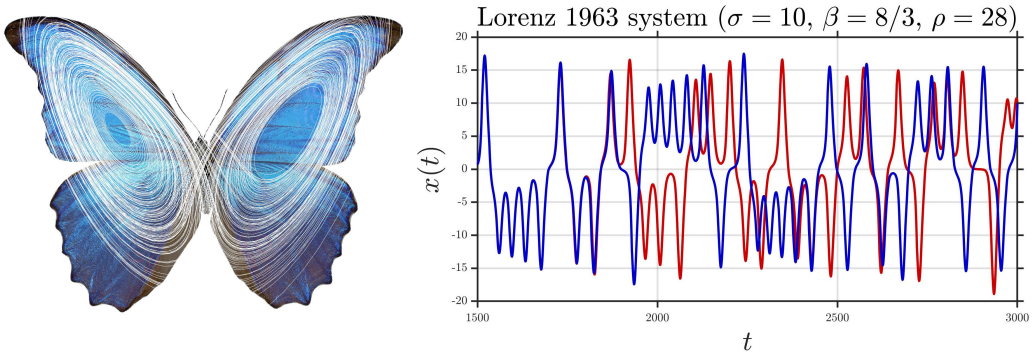


Figure 1.13: An illustration of the 'butterfly effect'. **Left:** Trajectory of the solution to the Lorenz (1963) system of equations in the $x-z$ plane (white) overlaid on a *Morpho didius* butterfly. **Right:** The x-component of two solutions to the same system that vary by only 10^{-9} in the initial condition, they start to diverge after around 2000 time steps.

The preferred course of action is clearly to try and get the best of both worlds by merging unconstrained, continuous, and direct model estimates with relatively accurate, discontinuous, and typically indirect satellite remote sensing data. It is imperative to note that in this merger both the observational data and the model estimates are uncertain, neither is perfect and we do not know exactly how wrong they are. Fortunately, an entire field of study exists that is devoted to fusing uncertain model estimates with uncertain observational data and that field is known as data assimilation (DA; see e.g. Carrassi et al., 2018, for a recent overview). DA emerged out of operational needs in meteorology, most notably the need to accurately initialize numerical weather prediction models (c.f. Bauer et al., 2015). The extreme sensitivity to initial conditions in chaotic systems such as the weather has come to be known as the 'butterfly effect', an example of this is provided in Figure 1.13. Nowadays DA techniques are still employed for this purpose, but also

for a wide range of other applications such as parameter estimation in models and performing retrospective analyses (reanalyses) of the climate system. Most forms of modern data assimilation can be viewed as exercises in updating mechanistic (typically dynamical) models with observational data using Bayesian inference. I will return to the Bayesian link in the Methods chapter. For now, data assimilation should be viewed simply as a framework for combining models with data.

1.3.1 Snow reconstruction

Snow data assimilation is already a field in itself, and I by no means claim to be the first to have ventured into its path. The recent emergence of snow data assimilation review articles are clearly testament to this effect (e.g. Helmert et al., 2018). The pioneering work of Martinec and Rango (1981) was arguably one of the earliest application of satellite data assimilation in snow science. By then the Landsat satellite program had begun and the early Landsats (1 & 2) had already been used to retrieve snow-covered area from space (Salomonson and Rango, 1974). The potential for using these retrievals in making improved snowmelt runoff forecasts had also already been demonstrated (Ødegaard et al., 1980). The ingenious idea of Martinec and Rango (1981) was to use the fact that snow-cover duration is closely tied to peak snow water equivalent. Most simply stated, the idea is that if the ablation rates in two areas are the same and one area becomes snow-free later than the other, then the area where the snow lasts longer must have had more SWE at the onset of the ablation. Generalizing this idea, they realized that they could reconstruct the peak SWE through backward calculation of the accumulated snowmelt from the day of snow disappearance to the day of peak SWE (the onset of the ablation season). In doing so, they could reconstruct spatially distributed peak SWE by accounting for the variation in snowmelt rates across the terrain. The operation could also be reversed and carried out in a forwards manner by summing the daily ablation from the day of peak SWE to the day of disappearance of the snow-cover. Moreover, the approach could even be used to estimate SWE at any day after peak SWE by starting the summation at a later day. The day of snow disappearance could be inferred from the optical satellite imagery and the snowmelt rates could be estimated using a snowmelt model, in their case a simple degree day (temperature index) model (c.f. Rango and Martinec, 1995). This approach where no errors are assumed in the observations (the snow disappearance date) is an application of DA through direct insertion.

The approach of Martinec and Rango (1981) has come to be known simply as snow reconstruction. It has since its inception been adapted and applied successfully on a number of occasions. Cline et al. (1998) combined distributed snow energy balance modeling with Landsat imagery to reconstruct SWE. They performed an important modification to the original approach by also weighting the potential ablation in each grid cell by the fractional snow-covered area (inferred from Landsat) for each day during the snow-cover depletion. In that way the actual grid-cell scale ablation in low fSCA conditions towards the end of the melt season, where potential ablation can be high, was not overestimated. The original approach of Martinec and Rango (1981) was strictly only valid for an instantaneous step (i.e. Heaviside-like) snow-cover depletion, going directly from 100% to 0% fSCA in one day. This crucial modification of Cline et al. (1998) would remain in subsequent applications of reconstruction. As such, snow reconstruction henceforth refers to this fSCA weighted approach. A schematic overview of the snow reconstruction approach is provided in Figure 1.14.

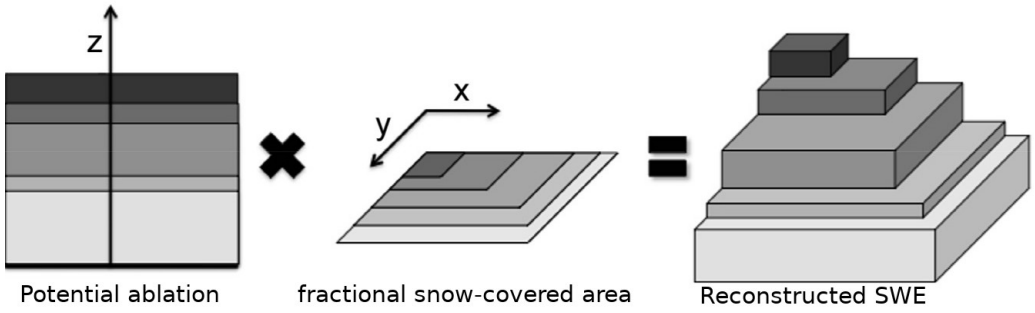


Figure 1.14: A conceptual schematic illustrating the snow reconstruction method. Each potential ablation increment (going sequentially in time from light to dark gray) is weighted by the corresponding fSCA for that day to arrive at the actual ablation for that day. The actual ablation for each day is then summed to reconstruct the peak SWE. Adapted from Kahl et al. (2013).

Snow reconstruction was subsequently implemented at multiple resolutions using multiple satellite sensors in the study of Molotch and Margulis (2008). Their study indicated that in the context of reconstruction, in the trade off between the spatial and temporal (i.e. repeat frequency) resolution of satellite sensors, it was the spatial resolution that was more important. This triggered a comment by Slater et al. (2009), who were surprised by this "un-intuitive" finding given that a key component of reconstruction is estimation of the snow disappearance date which would

be more accurate with a higher repeat frequency. Slater et al. (2009) emphasized and demonstrated three major shortcomings in the reconstructions implemented by Molotch and Margulis (2008): (i) not accounting for model uncertainties, (ii) assuming that melt season precipitation is insignificant, and (iii) extreme sensitivity to satellite image availability. This of course triggered a response, Molotch et al. (2010), where a point by point refutation of the issues raised by Slater et al. (2009) was provided. This then all culminated in the study of Slater et al. (2013) which showed rather systematically that snow reconstruction is plagued by uncertainties in both the satellite retrieved snow-cover depletion and the forcing data. Slater et al. (2013) concluded that for snow reconstruction errors in the forcing data were at least as important as satellite image availability. Despite the controversy and the undeniable legitimacy of the issues raised by Slater et al. (2013), several successful implementations of snow reconstruction were undertaken since the study of Molotch and Margulis (2008). Notable examples include the study of Rittger et al. (2016), where the entire Californian Sierra Nevada snowpack was reconstructed for 12 snow seasons at 100 m resolution, and the study of Bair et al. (2016) where unbiased SWE reconstructions were demonstrated by using the ASO for validation.

1.3.2 Bayesian inversion

Assimilating fSCA in particular is so appealing because accurate retrievals of fSCA are readily obtained at the hillslope using multispectral imagery from optical satellite sensors (Dozier et al., 2009). An example of a high-resolution snow-cover map obtained from satellite imagery is provided in Figure 1.15. There thus exists a strong motivation to resolve the issues related to traditional snow reconstruction outlined in Slater et al. (2013), since despite its flaws, it has been shown to be a promising means of (retrospectively) estimating SWE (Dozier et al., 2016). More or less at the same time as the debate concerning the efficacy of snow reconstruction was ongoing, different approaches to fSCA assimilation were being explored. Kolberg and Gottschalk (2006) developed a Bayesian approach to reconstruction that hinged upon updating the parameters and state variables that defined a probabilistic snow depletion curve (SDC). Probabilistic SDCs parametrize the relationship between SWE and fSCA by assuming a certain distribution for

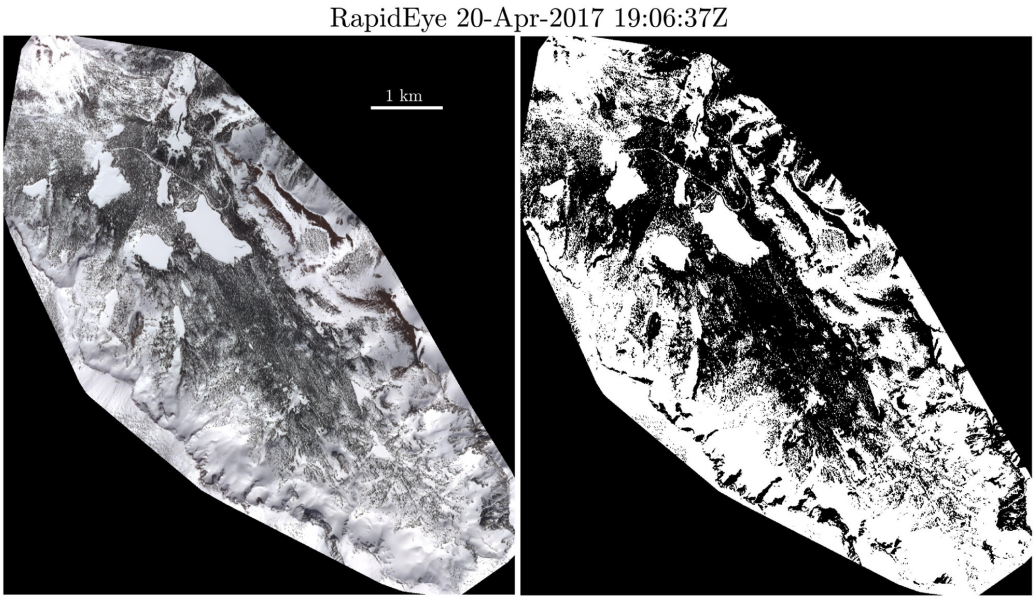


Figure 1.15: **Left:** True color image of the Mammoth Lakes basin sensed 20.04.2017 by a RapidEye satellite at 5 m resolution. **Right:** A binary snow-cover map (white: snow, black: no snow) retrieved from the same image using a segmentation approach.

subgrid peak SWE and assuming homogeneous subgrid melt (Liston, 2004). Kolberg and Gottschalk (2006) were the first to realize that by making use of such an SDC, reconstruction could be recast as a Bayesian inverse problem (c.f. Tarantola, 2005; Stuart, 2010). Bayesian inversion can be viewed as a special case of data assimilation that is primarily concerned with batch parameter estimation. In their inversion, the distributions of uncertain parameters related to the input to the SDC were updated based on the fSCA depletion that was observed by Landsat 7. The inversion was subsequently refined by using a spatial model for their prior parameter distribution (Kolberg et al., 2006). As with reconstruction, the inversion approach of Kolberg and Gottschalk (2006) could only be carried out retrospectively after the snow-cover had melted out thus limiting its uses for forecasting purposes. Still, Kolberg and Gottschalk (2010) showed how the results of the inversion in other years could be used to calibrate their model and significantly improve predictions of snowmelt runoff. A downside of the Bayesian inversion applied in Kolberg and Gottschalk (2006) and the subsequent studies was the use of Markov Chain Monte Carlo (MCMC) to sample from the posterior distribution (see Neal, 1993, for a review). This is problematic because MCMC sampling is computationally expensive, requiring tens of thousands of model runs. This precludes the use of such an inversion framework with more complex snow models

running at high (i.e. hillslope scale) resolution for large domains.

1.3.3 Sequential ensemble-based data assimilation

Coincidentally with the work of Kolberg et al. (2006) a new class of sequential ensemble-based DA algorithms were being applied to snow-related problems. Here the term 'ensemble' simply refers to the fact that a collection of model realizations (so-called ensemble members) are performed, it is analogous to Monte Carlo simulation. These algorithms emerged out of the work of Gordon et al. (1993) and Evensen (1994) a decade earlier where what are now known as particle filters (PF) and ensemble Kalman filters (EnKF) were introduced. These algorithms are called sequential because they assimilate the data 'online' as it becomes available, making a Markovian ('memoryless') assumption that the future state depends only on the present state (Bertino et al., 2003; Carrassi et al., 2018). This assumption is justified only if the temporal auto-correlation of errors is low. This is typically the case in numerical weather prediction which is mostly governed by hyperbolic partial differential equations with limited memory, at least in the Eulerian (fixed rather than flow-following Lagrangian) frame. It is not as clear if this should be the case for land surface models, of which snow models surely are a subset, where model errors often have considerable memory (e.g. Koster and Suarez, 2001).

Slater and Clark (2006) were the first to apply ensemble-based snow data assimilation when they employed the EnKF to assimilate spatially interpolated SWE observations into the SNOW-17 conceptual model. Despite improvements in both the simulated accumulation and ablation seasons after the updates from the EnKF, they noted that the increase in skill was marginal, likely due to the large temporal auto-correlation exhibited by the snowpack. Subsequently, Clark et al. (2006) performed a so-called twin experiment, where synthetic observations are generated by adding noise to a 'truth' run and the results are compared to this truth run, to assimilate fSCA using the EnKF. They found only minor improvements in the accuracy of simulated streamflow through the assimilation. Andreadis and Lettenmaier (2006) assimilated real MODIS fSCA and AMSR-E SWE retrievals into the Variable Infiltration Capacity (VIC) hydrological model. Large errors were identified in the AMSR-E data owing to the previously discussed issues with passive microwave remote sensing of SWE (Foster et al., 2005). For the fSCA assimila-

tion, they noted a modest improvement over the open-loop (no data assimilation) run using independent SWE observations for validation. So overall, all three of these early implementations of the EnKF for snow data assimilation noted only modest improvements over the open-loop.

Despite these findings, the EnKF has since found widespread use in snow data assimilation. For example, Su et al. (2010) performed a multisensor assimilation of both GRACE TWSA and fSCA from MODIS to estimate SWE over North America using the EnKF and found a significant improvement compared to assimilating MODIS fSCA alone. De Lannoy et al. (2010) explored the potential of assimilating coarse-scale passive microwave SWE into a high-resolution (1 km^2) land surface model in a twin experiment using the EnKF. They found greatest improvements when simultaneously assimilating multiple neighboring passive microwave SWE observations, reducing the root mean squared error (RMSE) by as much as 60% compared to the open-loop. Subsequently, De Lannoy et al. (2012) performed a real multisensor experiment where AMSR-E SWE and MODIS fSCA retrievals were assimilated jointly using the EnKF, providing significant improvements over the open-loop both in terms of RMSE and correlation. Magnusson et al. (2014) demonstrated how flux (i.e. snowfall and snowmelt) data derived from point scale SWE measurements could be assimilated into a calibrated distributed model using the EnKF and improve the model performance. Huang et al. (2017) showed how the data from SWE measurement stations could be assimilated into a calibrated coupled SNOW-17-Sacramento Soil Moisture Accounting (SAC-SMA) model to improve streamflow prediction using the EnKF. The improvement was greatest for basins that had a poorer historical performance prior to assimilation. Stigter et al. (2017) jointly assimilated fSCA retrieved from MODIS and in-situ snow depth observations from several locations within a Himalayan catchment using the EnKF to calibrate parameters in their modified SeNorge model. The EnKF-calibrated model outperformed the open loop.

It is only more recently that the PF has started to be applied to snow data assimilation studies. Unlike the EnKF, the PF requires no assumptions about the linearity of the model or the Gaussianity of error statistics (see van Leeuwen, 2009, for a review). It is, however, prone to degeneracy (filter collapse) when the dimensionality (number of state variables and parameters) in the problem at hand is large. Probably the first such application was in the work of Leisenring and Moradkhani (2011), where the performance of several versions of the PF was compared to

several versions of the EnKF in the assimilation of synthetic SWE observations into a calibrated SNOW-17 model. In their calibrated synthetic twin experiments they used 'only' 2000 particles (i.e. ensemble members, model realizations) and found that the PF outperformed the EnKF. Charrois et al. (2016) used the PF to assimilate synthetic optical reflectance retrievals and snow depth measurements into the Crocus model at the Col du Lautaret in the French Alps. In their point-scale twin experiments they found that the PF considerably outperformed the open loop, reducing the RMSE by around 70%. Magnusson et al. (2017) assimilated snow depth observations into the Jules Investigation Model (JIM) using the PF. The assimilation reduced errors in SWE, snowmelt runoff, and ground temperature even when forcing the model with coarse meteorological reanalysis data at Col de Porte in the French Alps. At several sites in the Swiss Alps, the assimilation was found to reduce the RMSE compared to the open loop by 64%. Piazzi et al. (2018) performed a joint assimilation of several in-situ measurements (snow temperature, SWE, albedo, snow depth, and snow density) at three sites. They found, among other things, that both the forcing and the parameters had to be perturbed to avoid filter degeneracy. Smyth et al. (2019) also recently demonstrated how the assimilation of snow depth with the PF can be used to improve both SWE and snow density estimates, with implications for observing systems such as the ASO.

1.3.4 Ensemble-based reanalysis

Durand et al. (2008) made the key realization that the SWE estimation problem could be approached in a novel way by combining the strengths of the snow reconstruction approach of Cline et al. (1998) and the new snow DA ideas that were emerging at that time (e.g. Kolberg and Gottschalk, 2006; Slater and Clark, 2006; Clark et al., 2006; Andreadis and Lettenmaier, 2006). In their approach, they made use of the ensemble smoother (ES; van Leeuwen and Evensen, 1996) which is a fixed-interval smoother version of the EnKF (see Cosme et al., 2012). The advantage with using a smoother over a filter is that the state of the model could at any time be updated using (relative to the state) future as opposed to just past and present observations. In the context of the seasonal snowpack this is advantageous because then the remotely sensed snow-cover depletion can be used to also update the model state during the accumulation period. This would not be possible with filters such as the EnKF or the PF since in sequential data assimilation (i.e. filtering) there is no backwards propagation of information (Carrassi et al., 2018).

In their implementation Durand et al. (2008) actually used the ES to update a fixed parameter controlling the bias in the precipitation, based on the residual between the observed and predicted fSCA depletion. By assimilating synthetic fSCA observations using the ES, Durand et al. (2008) improved the mean peak SWE error relative to that of the prior by 86%. They used a log-normal probabilistic SDC (following Liston, 2004) to model fSCA. In their approach the ES performed its update in a single batch, using all the observations available for the snow season that was being analyzed. An advantage with this is that when the model is rerun with the updated parameters the states will be consistent with the model physics. This is not the case with a filter. The idea of using a batch-updating ES to perform retrospective analyses (i.e., reanalyses) in land data assimilation can be traced back to the work of Dunne and Entekhabi (2005). To differentiate traditional snow reconstruction from the approach of Durand et al. (2008) I will refer to approaches following the latter as snow reanalyses. This is also meant to acknowledge the use of smoothers which can account for the memory in the errors of the model and observations and use the future to update the past. Note that the approach of Kolberg and Gottschalk (2006) was quite similar to that of Durand et al. (2008), except that costlier MCMC simulations were used.

To get a better idea of what is meant by synthetic (or twin) experiments, I have included results from some of my own initial synthetic ensemble-based reanalyses. These were carried out at 7 km model resolution over the southern Swiss Alps using the GEOS-5 Nature Run (Putman et al., 2014) as forcing. For the data assimilation, I employed an iterative version of the ES (Emerick and Reynolds, 2013) that will be described in the methodology. The setup is very similar to the study of Durand et al. (2008), I merely added a few more uncertain parameters. Such synthetic experiments are useful for initial testing of the feasibility of a data assimilation system. We can check if we can recover the synthetic truth (unknown in reality) using ensemble-based DA in a realistic scenario, i.e. with biased forcing, noisy observations, and uncertain parameters. Figure 1.16 shows the results for an arbitrarily selected grid cell in the domain. Note how the synthetic truth is more or less exactly recovered by the posterior (after data assimilation) median (solid blue line) for both the fSCA and the SWE.

Following the study of Durand et al. (2008), Girotto et al. (2014b) demonstrated how the probabilistic approach to reconstruction developed by Durand et al. (2008)

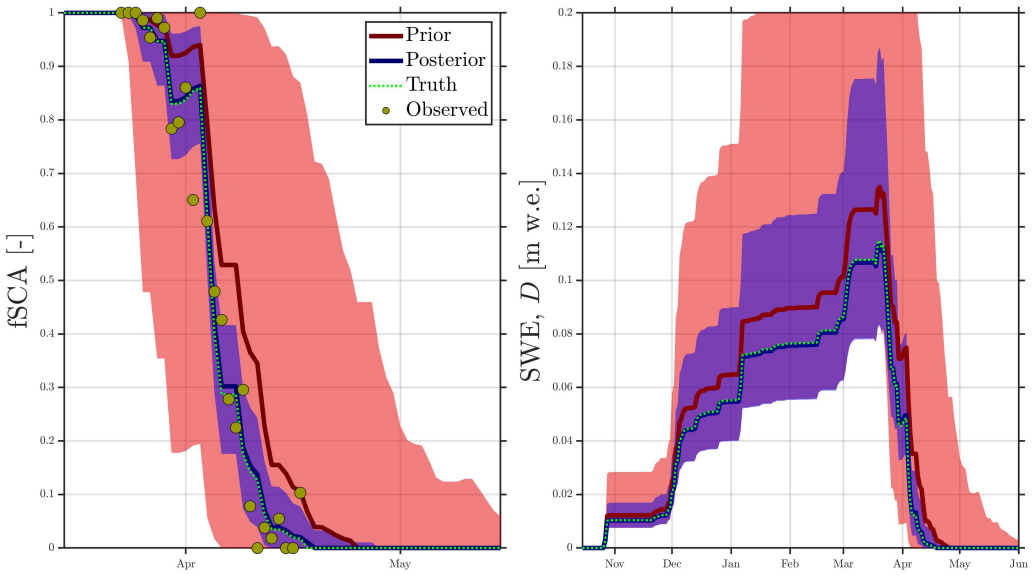


Figure 1.16: Results for an example grid cell from a synthetic experiment over the southern Swiss Alps using the GEOS-5 Nature Run (Putman et al., 2014) as forcing. Evolution of the fSCA (left) and average SWE depth (right) with the prior (before DA) in red,the posterior (after DA) in blue, observations as yellow circles, and the synthetic truth as the dashed green line. Shading represents the 90-th percentile range of the ensemble while solid lines show the ensemble median.

generalized and outperformed the traditional deterministic reconstruction approach (e.g. Cline et al., 1998). This demonstration was performed through a real (as opposed to synthetic) experiment where actual fSCA retrievals from Landsat 5 were used for two water years in the Tokopah watershed in the Californian Sierra Nevada. Validation was performed using independent in-situ SWE observations. The advantage with the framework of Durand et al. (2008) (and Kolberg and Gottschalk (2006) too for that matter) was that it enabled the experimenter to account for the uncertainties in traditional snow reconstruction (Slater et al., 2013) and constrain these through DA. Girotto et al. (2014a) subsequently applied the same framework to conduct a 27 year reanalysis over the Kern River watershed in the Californian Sierra Nevada with promising results compared to in-situ SWE observations.

Margulis et al. (2015) developed a so-called particle batch smoother (PBS) for ensemble-based snow reanalysis. This is essentially a particle filter without resampling. As such it has the same apparent advantages over the ES that the PF has over the EnKF, i.e. no need for the Gaussian linear assumptions. They demonstrated that the PBS markedly outperformed the ES for snow reanalysis. Subse-

quently, Margulis et al. (2016) employed the PBS and Landsat fSCA retrievals to perform a 30 year reanalysis (1985-2015) over the entire Sierra Nevada at a daily temporal and 90 m spatial resolution. The agreement with 9000 station years of independent observations was remarkable. For snow pillows, the posterior (after DA) peak SWE estimate had an RMSE of 0.11 m w.e. and a correlation of 0.97. Compared to snow courses, posterior SWE estimates had a RMSE of 0.13 and a correlation of 0.95. This Sierra Nevada snow reanalysis has, among other things, provided valuable new insights about controls on the snowpack (Hunning and Margulis, 2017) and serves as a valuable benchmark for other mountain range-scale SWE estimation techniques (Wrzesien et al., 2017).

Cortés et al. (2016) subsequently applied the PBS reanalysis approach to estimate SWE for the same Landsat era (1985-2015) for several basins in the extra-tropical Andes. The reanalysis provided significant improvements over the open loop when using snow surveys for validation: the correlation increased from 0.49 to 0.87 and the RMSE decreased by 60%. Cortés and Margulis (2017) further extended the reanalysis by considering the entire extra-tropical Andes at 180 m resolution. They found a significant positive correlation between peak SWE and El Niño, although La Niña was not found to have a corresponding significant negative correlation with peak SWE. Baldo and Margulis (2018) implemented a multi-resolution approach to ensemble-based reanalysis using the PBS in the upper Yampa basin in Colorado. They demonstrated that their multi-resolution approach, where the resolution of individual grid cells is dictated by a physiographic complexity metric, can save orders of magnitude in simulation time while not adversely affecting results relative to a full high-resolution reanalysis.

From the above exposition it is clear that the reanalysis approach to SWE reconstruction has now grown into a mature method for SWE estimation after the pioneering work of Kolberg and Gottschalk (2006) and Durand et al. (2008). These techniques can account for and constrain the uncertainties that plague traditional reconstruction (Slater et al., 2013). Limited computing power precludes the use of accurate MCMC sampling strategies for full Bayesian inversion (as in Kolberg and Gottschalk, 2006) at least with complex models at high-resolution. This is where the ensemble-based data assimilation techniques, smoothers in particular, shine as they provide a relatively accurate approximate framework for Bayesian snow reconstruction. At present, ensemble-based reanalysis probably provides the best means of tackling the SWE estimation problem, at least retrospectively.

Still, much work remains to be done as previous studies (e.g. Margulis et al., 2016; Cortés and Margulis, 2017) have restricted themselves to the use of the Landsat constellation. Several new satellite constellations, such as Sentinel-2, have recently been launched. These could provide an ideal complement to the Landsat constellation. In addition, it is well known that particle filters can be subject to problems with degeneracy (van Leeuwen, 2009). As such, the quest for improved ensemble-based snow reanalysis frameworks is still worth pursuing.

1.4 Aim and research objectives

The overarching aim of this thesis is to contribute towards a solution to the unsolved problem of accurately estimating SWE anywhere on Earth (Dozier et al., 2016). In this effort, I will restrict myself to what is arguably the most feasible current means of solving this problem for a seasonal snowpack, namely the ensemble-based snow reanalysis approach. This is the aforementioned probabilistic extension to snow reconstruction that was pioneered by Durand et al. (2008) and subsequently refined by Girotto et al. (2014b) and Margulis et al. (2015). In this approach, ensemble-based batch smoother schemes are used to assimilate fSCA satellite retrievals into a snow model so as to estimate SWE retrospectively. The ensemble-based snow reanalysis approach is actually a framework that is composed of multiple modular components, namely: a downscaling routine for atmospheric forcing, a snow model, fSCA satellite retrieval algorithms, and finally ensemble-based data assimilation schemes. In this thesis, my aim is to build such a framework and subsequently improve upon components of previously proposed frameworks. As is typical in modern science, my contribution is incremental and by no means some kind of paradigm shift. It also remains to be seen how contributing to solving a strictly retrospective SWE estimation problem can aid in snow-related predictions.

Having established my aim, contributing to a retrospective solution to the SWE estimation problem, I will outline the objectives that together will enable us to address this lofty target. The research objectives of this thesis are:

1. To design a modular ensemble-based snow reanalysis framework that can be employed anywhere on Earth where the snowpack is seasonal.

2. To implement a means of downscaling coarse-scale meteorological reanalysis data to the desired hillslope-scale resolution of a snow model.

3. To develop a simple snow model that can be run efficiently in ensemble-mode while resolving the key physical processes that are relevant to most seasonal snowpacks.

4. To probe our current capabilities of accurately retrieving fSCA from polar-orbiting satellites with near-global coverage.

5. To put into practice various ensemble-based data assimilation algorithms, evaluate their performance in terms of SWE reconstruction, and select the most robust algorithm for the reanalysis framework.

When putting these objectives into practice my work will be focused on a few key study areas where I have access to high quality independent validation data. The main study area in this thesis is the Brøgger peninsula which is situated in the north-western part of Spitsbergen island in the high-Arctic Norwegian archipelago of Svalbard. Here, I have access to data from extensive manual snow surveys that were conducted near peak SWE for several snow seasons (c.f. Figure 1.7). In addition, I employ time-lapse photography from an automatic camera system which I use to create accurate high-resolution snow-cover maps that are employed to validate satellite retrievals as well as results from the snow reanalysis framework. I also benefited from the availability of snow depth data from the Bayelva climate station (Boike et al., 2018). The reanalysis framework was also implemented in the Mammoth Lakes Basin (MLB) in the Californian Sierra Nevada. Here, I had access to LIDAR-derived snow depth and SWE retrievals from 11 ASO flights spanning three snow seasons (2016, 2017, and 2018). In addition, the framework was applied to several larger domains in the Swiss alps where we used station data and retrievals from an airborne digital sensor for validation. We present results from the Brøgger peninsula in paper I through III (Aalstad et al., 2018, 2019a,b), results from the MLB in paper III (Aalstad et al., 2019b), and results from the Swiss Alps in paper IV (Fiddes et al., 2019). The reanalysis framework has also been tested out in the Spanish Pyrenees at the Izas catchment, and the Daisetsu mountains (see Figure 1.17) in Hokkaido, Japan, however, these results are not included in the papers.

This thesis is structured in a similar fashion to a scientific paper. First the methodology in Chapter 2 covers the different components of the reanalysis framework in detail. Next, the results of the thesis are presented in the form of four papers, the first of which (Aalstad et al., 2018) is published, the second (Aalstad et al.,

2019a) is under review, the third (Aalstad et al., 2019b) is a manuscript that is to be submitted, and the fourth, Fiddes et al. (2019), is also under review. The thesis continues with Chapter 4 which offers a discussion and outlook where the implications of the papers are discussed. This includes a presentation of thoughts on what work remains to be done to improve snow reanalyses and extend their applicability to other domains such as hydrological prediction. The thesis ends with a conclusion in Chapter 5.

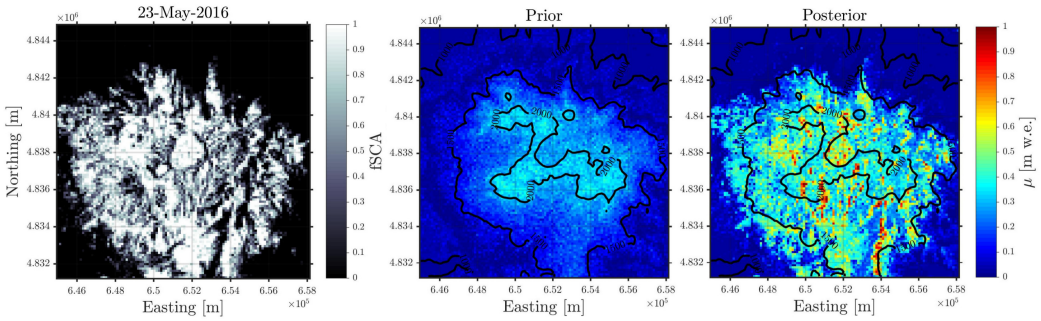


Figure 1.17: A 120 m resolution ensemble-based reanalysis of the 2016 snowpack over the Daisetsu mountains in Hokkaido, Japan. **Left**: fSCA retrieved from Sentinel-2A on the 23.05.2016. **Middle**: Prior peak SWE estimate. **Right**: Posterior peak SWE estimate. Note the similar patterns in the fSCA image and the posterior peak SWE image.

Chapter 2

Methods

In this chapter, the methodology in the ensemble-based snow reanalysis framework that was developed in Aalstad et al. (2018), Aalstad et al. (2019a), Aalstad et al. (2019b), and Fiddes et al. (2019) is described in detail. This is a modular framework consisting of a routine to downscale forcing, a simple snow model, techniques to retrieve fSCA from optical sensors, and ensemble-based data assimilation schemes. The idea is that each component can be switched out and refined without adversely affecting the remaining components. Herein, each component that is currently implemented in the framework will be reviewed.

2.1 Forcing

Forcing fields are required to drive the surface energy and mass balance in snow models. The fields required are typically the near-surface air temperature (T), wind speed (U), precipitation (P), specific humidity (q), incoming longwave radiation ($L^{\downarrow}$), and incoming shortwave radiation ($S^{\downarrow}$). In this section, I will outline a methodology to downscale these fields from coarse-scale meteorological reanalysis data to the resolution of the model grid. This methodology is based on the TopoSCALE approach of Fiddes and Gruber (2014) which is a topographic downscaling technique based on interpolation and various terrain corrections. It does not require running any dynamical meteorological model (e.g. Gutmann et al., 2016) and is as such an extremely computationally efficient downscaling technique. Some of my implementation of TopoSCALE varies from Fiddes and Gruber (2014) and so I have chosen to describe it here. The advantage with TopoSCALE over other approaches (e.g. Girotto et al., 2014b; Rittger et al., 2016)

is that it uses the vertical structure in meteorological reanalyses and thus avoids the need to prescribe fixed lapse rates.

2.1.1 Input meteorological reanalysis data

As inputs to TopoSCALE we require a coarse meteorological reanalysis data to downscale. In our work, we have primarily used the European reanalysis (ERA) products from the European Center for Medium-Range Weather Forecasts (ECMWF). In Aalstad et al. (2018), we used the ERA-Interim reanalysis (Dee et al., 2011) which was downscaled based on TopoSCALE (Fiddes and Gruber, 2014) following the approach outlined in Østby et al. (2017). ERA-Interim is available globally from 1979 up until today (2019, but it will be discontinued) at 0.75° resolution and 6 hourly temporal resolution on 60 atmospheric pressure levels from 1000 [hPa] to 0.1 [hPa]. In Aalstad et al. (2019b) and Fiddes et al. (2019) (as well as in Yılmaz et al. (2019)) we used the ERA5 reanalysis product, which is the successor of ERA-Interim. This fifth generation of global climate reanalyses from ECMWF is available from 1950 up until today (2019) in near real time at 0.25° spatial resolution and hourly temporal resolution on 137 atmospheric pressure levels from the surface up to a height of 80 km. ERA5 was downscaled following TopoSCALE as described briefly in Aalstad et al. (2019b), and it is this procedure that I will describe in more detail herein.

Both ERA5 and ERA-Interim are available on surface levels (i.e. at the surface of the reanalysis) and pressure levels (i.e. at various pressure levels up in the atmosphere). These two level types must be downloaded separately. ERA5 is available for download form the from the Copernicus Climate Change Service (C3S), while ERA-Interim is available from ECMWF. The pressure levels must be selected such that they encompass the topography of the area of interest (AOI). In this selection, the barometric equation (derived assuming hydrostatic balance), is quite useful for estimating the height z of a pressure level:

$$z = H \ln(\mathrm{p}_0/\mathrm{p}) \tag{2.1}$$

where $H \simeq 8$ [km] is a typical scale height in the troposphere, p is the known pressure at the level of interest and $\mathrm{p}_0 \simeq 1000$ [hPa] is the reference surface pressure. The barometric equation can be used, with some wiggle room to account for variations in meteorological conditions, to quite easily select the pressure level

range of interest.

When downloading the coarse-scale reanalysis data, a bounding box is selected that encompasses the AOI such that the centroids of the edge coarse-scale grid cells lie outside the AOI. This is to avoid subsequent extrapolation. I will use the subscript $(\cdot)_c$ to denote coarse-scale variables from the reanalysis. The surface file is three dimensional (horizontal and time) while the pressure level file is four dimensional (horizontal, vertical, and time). From the three dimensional surface level file, the following variables are extracted: the incoming shortwave radiation at the surface ($S_c^{\downarrow}$), the incoming longwave radiation at the surface ($L_c^{\downarrow}$), the near-surface air temperature ($T_{a,c}$), the near-surface dew point temperature ($T_{d,c}$), the total precipitation (P_c), the geopotential height of the ERA topography ($Z_{s,c}$), the surface pressure ($p_{s,c}$), and the exo-atmospheric shortwave radiation that is incoming at the top of the atmosphere ($S_{\infty,c}^{\downarrow}$). From the four dimensional pressure level file, the following variables are selected at the pressure levels neighboring that of the AOI: the meridional wind component (v_c), the zonal wind component (u_c), the air temperature (T_c), the relative humidity (RH_c), and geopotential height (Z_c).

To complement ERA5 at one of our study sites, Mammoth Lakes Basin in the Californian Sierra Nevada, we employed precipitation fields from phase 2 of the North American Land Data Assimilation System (NLDAS-2; Xia et al., 2012). NLDAS-2 provides hourly precipitation at a resolution of 0.125° over the continental United States based on a gauge-based analysis that is corrected for orographic effects. This is the same precipitation forcing that was employed to great effect for the Sierra Nevada snow reanalysis conducted by Margulis et al. (2016). We employed this product because it has been shown that global reanalysis products tend to underestimate snow accumulation in the complex terrain of the Californian Sierra Nevada mountains (Wrzesien et al., 2017) whereas regional reanalyses such as NLDAS-2 that can better account for orographically enhanced precipitation should at least partly resolve this deficiency.

2.1.2 Deriving topographic parameters

To perform the topographic downscaling we need a series of topographic parameters that determine the influence of the terrain on the near-surface meteorological fields. The first parameter required to derive the remaining parameters was the spatially distributed elevation field $z = z(x, y)$ where x is the easting and y is the

northing coordinate in a given UTM zone. Such a field is readily obtained from digital elevation models (DEMs). These are of varying quality, resolution, and spatial coverage. In our work we used the DEM generated by the Shuttle Radar Topography Mission (SRTM; Farr et al., 2007). The SRTM DEM is available globally at around 30 m resolution except for (absolute) latitudes exceeding $60°$. For latitudes greater than $60°$N, the ArcitcDEM is an option worth considering (Poster et al., 2018). Nonetheless, for the high-Arctic Brøgger peninsula site we instead opted for a high-resolution (5 m) DEM from the Norwegian Polar Institute (NPI, 2014). Before computing topographic parameters the elevation field from the DEMs were aggregated (upscaled) to the coarser model resolution (either 100 m or 1 km) using spatial averaging.

The majority of the topographic parameters were subsequently derived based on the work of Dozier and Frew (1990). The slope, θ_z, (in radians) was obtained through

$$\theta_z = \arctan\left(|\nabla_H z|\right) \,, \tag{2.2}$$

where the modulus of the horizontal gradient is

$$|\nabla_H z| = \sqrt{(\partial_x z)^2 + (\partial_y z)^2} \,, \tag{2.3}$$

in which the partial derivatives $\partial_x z$ and $\partial_y z$ are the directional gradients of the elevation. The aspect, ϕ_z, (also in radians) was obtained as follows

$$\phi_z = \arctan\left(\frac{-\partial_y z}{-\partial_x z}\right) \,. \tag{2.4}$$

Numerically the double negative is important since the aspect is computed by calling `atan2(`$-\partial_y z$`,`$-\partial_x z$`)`. Note that (2.4) defines the aspect anticlockwise from south in the range $[-\pi, \pi]$ where south is 0. This is simply for the mathematical convenience of having a conventional right handed coordinate system. For visualization purposes, we would typically convert the aspect back to the traditional bearing (clockwise from north) orientation. For both the slope and the aspect, the partial derivatives representing the respective directional gradients were approximated through centered finite differences, i.e. for a grid (x_i, y_j) with centroids

$$x_i = x_1 + (i-1)\Delta x \;\text{ and }\; y_j = y_1 + (j-1)\Delta y \,, \tag{2.5}$$

then

$$\partial_y z|_{(x_i,y_j)} \simeq \frac{z_{i,j+1} - z_{i,j-1}}{2\Delta y} \text{ and } \partial_x z|_{(x_i,y_j)} \simeq \frac{z_{i+1,j} - z_{i-1,j}}{2\Delta x}\,, \tag{2.6}$$

where typically the DEM grid is regular such that $\Delta x = \Delta y$.

Horizon angles, $\gamma_z(\phi)$, were determined as a function of binned azimuth angles (ϕ) independently for each grid cell. I set the binned horizon angles γ_n to the maximum positive elevation angle inside each bin. If no positive elevation angle was found γ_n was set to zero. Maximum elevation angles were determined for each azimuth bin and grid cell using a naive brute force search algorithm. This is quite an expensive search. In future implementations for larger domains it would be worth investing in a more efficient algorithm such as the one presented in Dozier et al. (1981). For economy, only $N_h = 36$ azimuthal horizon bins (i.e. 10° increments) were used. Consequently, the spatially varying sky-view factor, V_z, was estimated following Žaksek et al. (2011) through

$$V_z = 1 - \frac{1}{N_h}\sum_{n=1}^{N_h} \sin\left(\gamma_n\right)\,. \tag{2.7}$$

The sky-view factor, as the name suggests, represents the fraction of the overlaying sky hemisphere that is visible at a point on the surface. Dynamic (i.e., time varying) solar geometry in the form of the solar zenith angle θ_0 and solar azimuth angle ϕ_0 (defined in the same coordinate system as ϕ_z) was obtained from the ephemeris routine in the PV_LIB toolbox (SNL, 2014). These were used to calculate the dynamic local illumination angle (θ_i) through (Dozier and Frew, 1990)

$$\theta_i = \arccos\left(\cos(\theta_0)\cos(\theta_z) + \sin(\theta_0)\sin(\theta_z)\cos(\phi_0 - \phi_z)\right)\,, \tag{2.8}$$

the local illumination angle is the angle that an incident solar beam makes with the local surface normal (c.f. Fiddes and Gruber, 2014). The coarse-scale illumination angle is calculated analogously by accounting for the slope and the aspect of the coarse-scale reanalysis grid. A dynamic local binary shadow mask δ is also derived based on the horizon angles and the solar zenith angle as follows

$$\delta = \begin{cases} 1, & \text{if } \gamma^* < (\pi/2 - \theta_0) \\ 0, & \text{otherwise} \end{cases}\,, \tag{2.9}$$

where γ^* is the horizon angle for the bin corresponding to the current solar az-

imuth angle. Essentially, equation (2.9) checks if the sun is above the horizon or not, setting $\delta = 1$ if it is and $\delta = 0$ if it is not. An example visualization of some of the derived topographic parameters is provided in Figure 2.1.

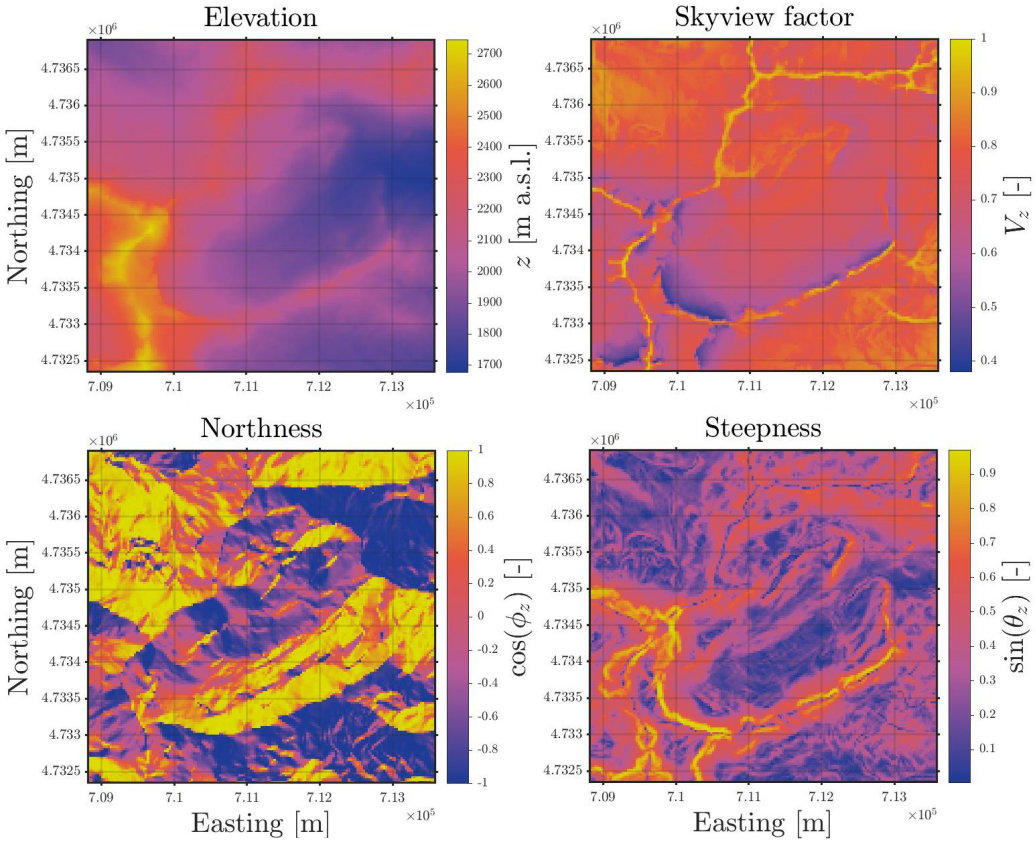


Figure 2.1: Topographic parameters derived at 30 m resolution for a 22 km^2 area surrounding the Izas catchment in the Spanish Pyrenees (UTM 30N). Northness (bottom left) is the cosine of the aspect (as a bearing) which is 1 for north facing slopes and -1 for south facing slopes. Steepness (bottom right) is the sine of the slope which is 0 for flat ground and 1 for completely vertical slopes. The base LIDAR-derived DEM was obtained from the Spanish National Geographic Institute.

2.1.3 Topographic downscaling

Having access to the coarse-scale reanalysis fields and the high-resolution topographic parameters acquired by manipulating a DEM, we are ready to proceed with the topographic downscaling algorithm TopoSCALE (Fiddes and Gruber,

2014). In a first step, the altitude, z_c, of the pressure levels at each time step is computed using the relationship between geopotential height and altitude, i.e.

$$z_c = Z_c/g\,, \tag{2.10}$$

where $g \simeq 9.81$ ms^{-2} is the acceleration of gravity at the Earth's surface which is (justifiably) assumed to be invariant in the troposphere (the lower $\simeq$10 km of the atmosphere) which we are concerned with. The coarse-scale elevation of the topography in ERA is obtained analogously, i.e. $z_{s,c} = Z_{s,c}/g$.

Next all the variables defined on surface levels are horizontally interpolated onto the higher resolution grid using bi-linear interpolation. These variables will nonetheless keep the subscript $(\cdot)_c$ because most of them will still be subject to further downscaling. Subsequently, the variables that are defined on pressure levels ($u_c, v_c, T_c, \mathrm{RH}_c$) are downscaled using three dimensional linear interpolation. That is to say, the pressure level variables defined on the three dimensional coarse grid x_c, y_c, z_c are interpolated onto the two dimensional surface $z(x, y)$ which is the elevation of the resampled high-resolution DEM. It is worth emphasizing that the elevations z_c will vary in both space and time since we are employing the dynamic geopotential in the meteorological reanalyses. As such, this interpolation is dynamic with both vertical and horizontal gradients varying from timestep to timestep. This is in contrast to typical approaches where a fixed or possibly seasonally varying lapse rate is applied to adjust temperature with elevation (e.g. Girotto et al., 2014b; Rittger et al., 2016). Based on the downscaled wind field (u, v) the near-surface wind speed U is computed in a standard fashion, i.e. as the magnitude of the horizontal wind vector

$$U = \sqrt{u^2 + v^2}\,. \tag{2.11}$$

Similarly, the downscaled specific humidity is obtained from the downscaled near-surface air temperature and near-surface relative humidity (defined with respect to liquid water in meteorological reanalyses) using a long chain of standard hygrometric relationships (see Lawrence, 2005) as follows

$$q = \mathrm{w}/(1 + \mathrm{w})\,, \tag{2.12}$$

where the water vapor mixing ratio w is obtained through

$$\mathrm{w} = \varepsilon_0 e/(\mathrm{p}_s - e)\,, \tag{2.13}$$

where $\varepsilon_0 = 0.622$ is the the ratio of the molecular weights of water and dry air, p_s is the surface pressure obtained from Equation 2.28, e is the water vapor pressure obtained through

$$e = \mathrm{RH}e_s/100\,, \tag{2.14}$$

where the near-surface saturation water vapor pressure is obtained through the Magnus formula

$$e_s = C_1 \exp\left(A_1 T'/(B_1 + T')\right)\,, \tag{2.15}$$

where $A_1 = 7.625$, $B_1 = 243.04°\mathrm{C}$, $C_1 = 610.94\,\mathrm{Pa}$ and $T' = T - T_0$ (where $T_0 = 273.15$ K) is the downscaled temperature in °C. The coarse-scale horizontally interpolated near-surface dewpoint temperature is converted to relative humidity as follows:

$$\mathrm{RH}_c = 100 \exp\left(\frac{A_1 T'_{d,c} - C_2 T'_{d,c} - B_1 C_2}{B_1 + T'_{d,c}}\right)\,, \tag{2.16}$$

where again $T'_{d,c} = T_{d,c} - T_0$ and

$$C_2 = A_1 T'_{a,c} / \left(B_1 + T'_{a,c}\right) \tag{2.17}$$

where again $T'_{a,c} = T_{a,c} - T_0$. RH_c is then used in (2.14) to obtain e_c which is the coarse-scale (but effectively horizontally interpolated) near-surface water vapor pressure needed for the longwave calculations.

For the downwelling longwave radiation, the clear sky emissivity is diagnosed using an empirical relationship after Konzelmann et al. (1994)

$$\varepsilon^{(\mathrm{clear})} = 0.23 + x_1 \left(e/T\right)^{x_2}\,, \tag{2.18}$$

where $x_1 = 0.43$ and $x_2 = 5.7$ following Gubler et al. (2012). The coarse-scale clear sky emissivity $\varepsilon_c^{(\mathrm{clear})}$ is computed analogously. In addition, the coarse-scale all sky emissivity is diagnosed through Stefan-Boltzmann's equation, i.e.

$$\varepsilon_c^{(\text{all sky})} = L_c^{\downarrow}/\sigma_{\mathrm{sb}} T_c^4\,, \tag{2.19}$$

where $\sigma_{\mathrm{sb}} = 5.67 \times 10^{-8}\ \mathrm{W\ m^{-2}\ K^{-4}}$ is the Stefan-Boltzmann constant. Assuming

that the coarse and downscaled cloud emissivity are the same, then the downscaled downwelling longwave radiation is given by

$$L^{\downarrow} = \varepsilon^{(\text{all sky})} \sigma_{\text{sb}} T^4 \,, \tag{2.20}$$

where

$$\varepsilon^{(\text{all sky})} = \varepsilon^{(\text{clear})} + \left(\varepsilon_c^{(\text{all sky})} - \varepsilon_c^{(\text{clear})} \right) \,. \tag{2.21}$$

Finally, the incoming longwave radiation is scaled to take into account terrain effects as follows

$$L^{\downarrow} \longleftarrow V_z L^{\downarrow} + 0.5 \left(1 + \cos(\theta_z)\right) \left(1 - V_z\right) \varepsilon_s \sigma_{\text{sb}} T_0 \,, \tag{2.22}$$

where $\varepsilon_s = 0.99$ is the surface emissivity (assumed to be equal to the snow emissivity) and $T_0 = 273.15$ K is the surface temperature which is for simplicity assumed to be equal to the melting point. The first term in (2.22) accounts for terrain occlusion of downwelling atmospheric longwave radiation while the second term accounts for longwave radiation emitted by the surrounding terrain. The term involving terrain slope that accounts for the terrain configuration is based on the work of Dozier and Frew (1990). The notation $\longleftarrow$ is algorithmic and sets the left hand side of the arrow to the right hand side. The effects of a canopy on radiative fluxes may also need to be accounted for (c.f. Link and Marks, 1999; Garren and Marks, 2005). Here the fractional canopy covered area (fCCA) can be obtained from ancillary sources such as the national land cover data base (NLCD; Yang et al., 2018) for the US. A correction is then applied for the canopy covered part of each grid cell, and the final corrected radiative flux is the weighted combination of the canopy-covered and open sky radiative flux as in Bair et al. (2016). Thereby, the longwave radiation can be corrected for canopy absorption and emission as follows

$$L^{\downarrow} \longleftarrow \text{fCCA} \left(\tau_c L^{\downarrow} + (1 - \tau_c) \varepsilon_c \sigma_{\text{sb}} T^4 \right) + (1 - \text{fCCA}) L^{\downarrow} \,, \tag{2.23}$$

where τ_c is the canopy transmissivity (see Link and Marks, 1999), $\varepsilon_c \simeq 0.96$ is the canopy emissivity (Garren and Marks, 2005), and T is the downscaled near-surface air temperature which, for simplicity, is assumed to be equal to the temperature of the canopy.

For the downwelling shortwave radiation, I present a modification to the approach

in Fiddes and Gruber (2014) that corrects errors originating in the study of Ruiz-Arias et al. (2010b) where Beer's law is formulated incorrectly. To clarify, Ruiz-Arias et al. (2010b) present the path length as being proportional to the cosine of the solar zenith angle when in fact it is inversely proportional to this cosine. As a first step, the coarse-scale (but already bi-linearly horizontally interpolated) incoming shortwave radiation is split into a diffuse and a direct beam component. This partitioning is performed based on the clearness index, k_t, which is defined as the ratio of surface to exo-atmospheric incoming shortwave radiation, i.e.

$$k_t = S_c^{\downarrow} / S_{\infty,c}^{\downarrow} \,. \tag{2.24}$$

This clearness index is then used to compute the diffuse fraction, f_{diff}, based on the following empirical double exponential relationship obtained by Ruiz-Arias et al. (2010a)

$$f_{\text{diff}} = 0.952 - 1.041 \text{exp}\left(-\text{exp}(2.3 - 4.702 k_t)\right) \tag{2.25}$$

this is then used to diagnose the diffuse ($S_{\text{diff},c}^{\downarrow}$) and direct ($S_{\text{dir},c}^{\downarrow}$) components as follows

$$S_{\text{diff},c}^{\downarrow} = f_{\text{diff}} S_c^{\downarrow} \;\text{ and }\; S_{\text{dir},c}^{\downarrow} = (1 - f_{\text{diff}}) S_c^{\downarrow} \,. \tag{2.26}$$

Subsequently, $S_{\text{dir},c}^{\downarrow}$ is vertically interpolated to the model grid using the following relationship

$$S_{\text{dir}}^{\downarrow}(z) = S_{\infty,c}^{\downarrow} \left(\frac{S_{\text{dir},c}^{\downarrow}(z_{s,c})}{S_{\infty,c}^{\downarrow}} \right)^{\text{p}(z)/\text{p}(z_{s,c})} \tag{2.27}$$

which I derive from Beer's law in Appendix A. In (2.27), $\text{p}(z_{s,c})$ is known and $\text{p}(z)$ can be obtained through the hypsometric equation (Wallace and Hobbs, 2006) as follows

$$\text{p}(z) = \text{p}(z_{s,c}) \text{exp}\left(-\frac{(z - z_{s,c})}{H} \right) \tag{2.28}$$

where

$$H = R\overline{T}/g \tag{2.29}$$

is the scale height, in which $R = 287$ J kg^{-1} K^{-1} is the specific gas constant for dry air, g is the already defined acceleration of gravity, and $\overline{T}$ is the mean temperature of the layer between z and $z_{s,c}$ which I approximate through

$$\overline{T} \simeq \left(T + T_{a,c}\right)/2 \tag{2.30}$$

where $T_{a,c}$ is the horizontally interpolated coarse-scale near-surface air temperature. Next, the local illumination angle and cast shadows are accounted for through the following expression

$$S^{\downarrow}_{\text{dir}} \longleftarrow S^{\downarrow}_{\text{dir}} \frac{\cos(\theta_i)}{\cos(\theta_{i,c})} \delta \tag{2.31}$$

where θ_i is the dynamic local illumination angle, $\theta_{i,c}$ is the dynamic coarse-scale illumination angle, and δ is the dynamic local binary shadow mask. Details on how to compute the topographic parameters were already provided. For simplicity, TopoSCALE presently ignores incoming shortwave radiation that has been reflected by the surrounding terrain, even though this can be a notable contribution when the surroundings are snow-covered. Accurately accounting for this effect would, however, require storing the dynamic albedo of the surroundings somehow, which would slow down computations considerably. Instead, the effect could simply be added as a correction term in a distributed snow model. The direct shortwave radiation is also corrected for canopy extinction following Garren and Marks (2005) through

$$S^{\downarrow}_{\text{dir}} \longleftarrow \text{fCCA} \left[S^{\downarrow}_{\text{dir}} \text{exp} \left(-\frac{\mu_c h_c}{\cos(\theta_0)} \right) \right] + (1 - \text{fCCA}) S^{\downarrow}_{\text{dir}} \tag{2.32}$$

where h_c is the height of the canopy and μ_c is the canopy extinction coefficient (see Link and Marks, 1999).

The diffuse radiation is obtained by assuming negligible variation with elevation and correcting for terrain obscuring effects as follows

$$S^{\downarrow}_{\text{diff}} = V_z S^{\downarrow}_{\text{diff},c} \,. \tag{2.33}$$

The diffuse radiation is then corrected for canopy effects through (Garren and Marks, 2005)

$$S^{\downarrow}_{\text{diff}} \longleftarrow \text{fCCA} \left[S^{\downarrow}_{\text{diff}} \tau_c \right] + (1 - \text{fCCA}) S^{\downarrow}_{\text{diff}} \,. \tag{2.34}$$

where τ_c is the canopy transmissivity (see Link and Marks, 1999). Finally, the downscaled incoming shortwave radiation is obtained by recombining the downscaled diffuse and direct components, i.e.

$$S^{\downarrow} = S^{\downarrow}_{\text{diff}} + S^{\downarrow}_{\text{dir}} \,. \tag{2.35}$$

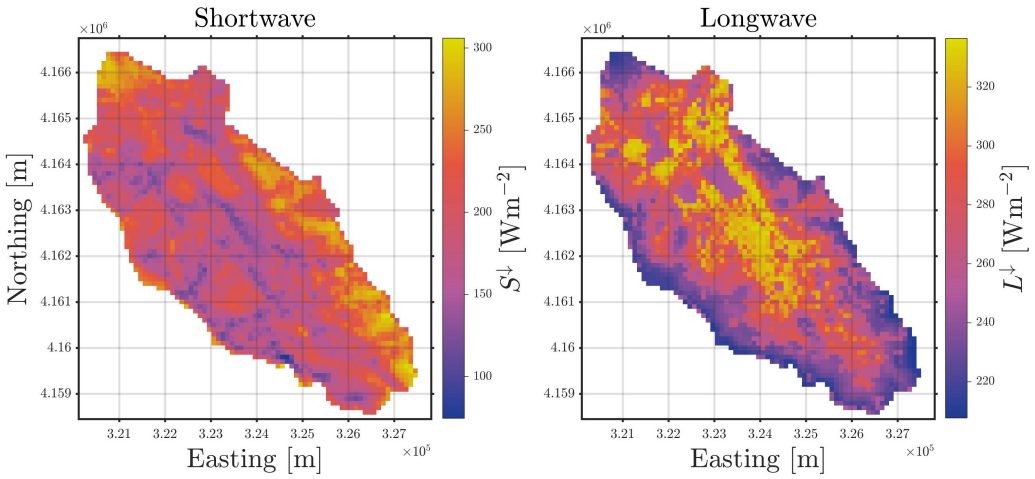


Figure 2.2: Topographically downscaled incoming shortwave radiation (left) and longwave radiation (right) averaged over for the 2018 snow season (October 2017 - September 2018) at the Mammoth Lakes basin in the Californian Sierra Nevada at 100 m resolution. The original coarse-scale fields were obtained from the ERA5 meteorological reanalysis. This high elevation basin is flanked by steep slopes on almost all sides and contains a relatively dense canopy in its center.

In the original TopoSCALE, precipitation is scaled according to monthly lapse rates (following Liston and Elder, 2006) using the elevation difference between the coarse-grid and the high-resolution model grid. I chose not to implement this in my version of TopoSCALE as it is arguably a little too much of an ad-hoc approach. In Aalstad et al. (2018), the precipitation was instead downscaled as in Østby et al. (2017) where a linear model for orographically enhanced precipitation (c.f. Smith and Barstad, 2004) was employed. This approach has been shown to be very promising for downscaling precipitation in complex terrain (Schuler et al., 2008) and it would complement the ensemble-based reanalysis approach nicely. The orographic downscaling of precipitation is currently not implemented in the version of TopoSCALE described here that was used in Aalstad et al. (2019b) and Fiddes et al. (2019). We do, however, plan to implement it as soon as possible. Right now the only downscaling of precipitation is through the horizontal regridding to the higher resolution model grid using bi-linear interpolation.

TopoSCALE generates gridded hourly outputs of all the fields needed to drive a snow model. So if the spatiotemporal domain is large and the spatial resolution is high then the amount of data is quite considerable. This can make storing the outputs of TopoSCALE quite demanding. It is nonetheless desirable to store the outputs to avoid having to rerun TopoSCALE each time the model is called. To alleviate this issue, I store all the TopoSACLE output variables as 16-bit integers

by scaling them according to their dynamic range. Scale factors are provided in the metadata that permit simple rescaling of the outputs to double precision once they are to be used as inputs in a model. The idea is that one only needs to keep in memory at double precision the forcing for the grid cells that one is currently simulating. A brief evaluation of the outlined TopoSCALE routine is shown in Figure 2.3. Note in particular how the scheme does a good job estimating variables such as near-surface air temperature, but overestimates the wind speed since boundary layer and terrain effects are ignored.

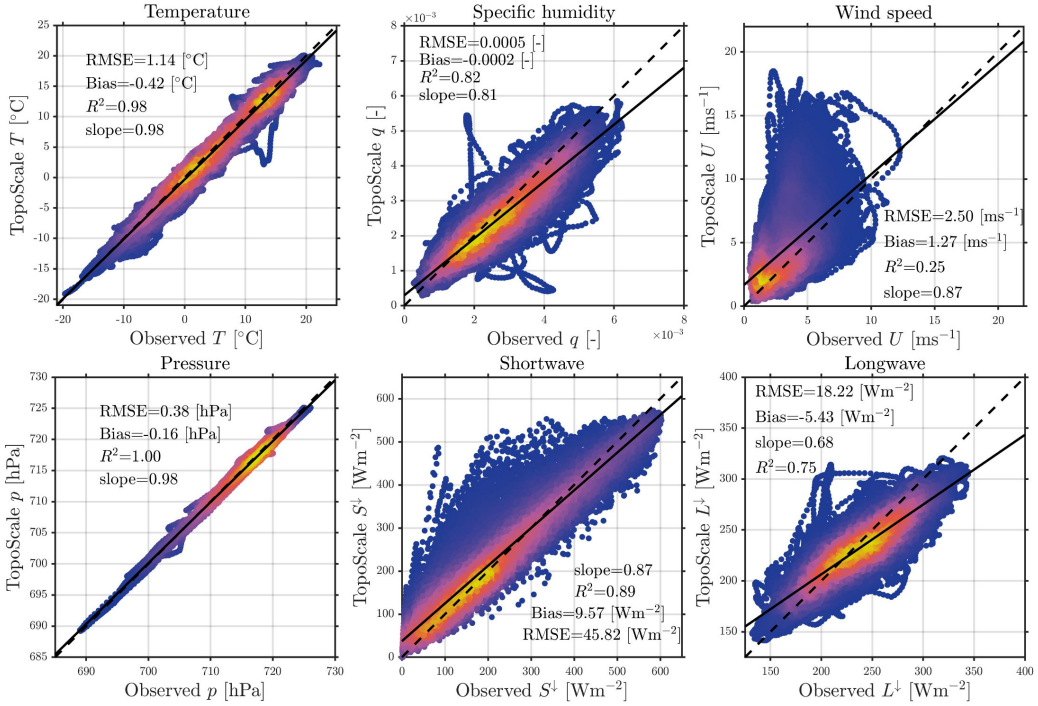


Figure 2.3: Evaluation of 30 m resolution topographically downscaled ERA5 data using 7 years of in-situ data from the CUES site near the Mammoth Lakes basin (see Bair et al., 2018). All panels show the TopoSCALE estimates versus observations. Marker color: scatter density (blue=low, yellow=high); solid black line: linear-best fit line; dashed black line: 1-1 line. All fields are hourly values to which a 1 day moving average has been applied.

2.2 Simple snow model

Here, I present the simple snow model (SSM) that we apply in Aalstad et al. (2018) and Aalstad et al. (2019b). This model is simple in the sense that it was built

to be parsimonious by only taking into account the dominant physical processes (the mass and energy balance) that dictate the bulk behaviour of the snowpack. This, in turn, allows the code to be completely vectorized permitting parallel runs for a large ensemble (as many as 10^5 ensemble members were used in Aalstad et al. (2018)) of model integrations. The model is largely based on the single layer Utah Energy Balance (UEB; Tarboton and Luce, 1996) snow model. The main simplification over UEB is that the simple snow model does not account for the cold content of the snow (i.e. no prognostic snow temperature), and instead assumes that the snowpack is always isothermal at $0°$C. This greatly simplifies the computation of the surface energy balance and possible melt rates. It also means that there are less concerns about numerical stability as no complex time stepping schemes or iterations are needed. The presentation herein largely follows Aalstad et al. (2018), but adds some new terms after Aalstad et al. (2019b) and is more explicit in the formulation of the respective fluxes of mass and energy.

2.2.1 Energy balance

With the assumption of an isothermal snowpack fixed at $0°$C, i.e. ignoring the cold content, the snowpack energy balance becomes

$$Q_M = Q_R^* + Q_P - Q_E - Q_H - Q_G \,, \tag{2.36}$$

where Q_M is the snowmelt flux, Q_R^* is the global radiation, Q_E is the latent heat flux, Q_H is the sensible heat flux, Q_G is the ground heat flux, and Q_P is the heat advected by precipitation (solid or liquid). Here, the first two fluxes on the right hand side of (2.36) are defined as positive when directed downwards towards the snowpack, whereas the last three are defined as positive when directed away from the snowpack. I will go through the formulation of the fluxes on the right hand side of (2.36) term by term. The global radiation is approximated via

$$Q_R^* = (1 - \alpha_s)S^{\downarrow} + L^{\downarrow} - \varepsilon_s \sigma_{\text{sb}} T_0^4 \tag{2.37}$$

in which the first term is the net incoming shortwave radiation (incoming$-$outgoing) where α_s is the snow albedo, the second term is the incoming longwave radiation, and the final term is the outgoing longwave radiation (assuming a snowpack at $T_0 = 273.15$ K). The snow albedo is parametrized prognostically using the continuous reset formulation of Dutra et al. (2010) as presented in Aalstad et al. (2018).

The heat advected by precipitation is diagnosed following Tarboton and Luce (1996) through

$$Q_P = P_S c_i \rho_w \text{min}\left(T - T_0, 0\right) + P_R \left[L_f \rho_w + c_w \rho_w \text{max}\left(T - T_0, 0\right)\right] \tag{2.38}$$

where $c_i = 2100$ Jkg^{-1}K^{-1} is the specific heat of ice, $L_f = 3.34 \times 10^5$ Jkg^{-1} is the latent heat of fusion, and $c_w = 4180$ Jkg^{-1}K^{-1} is the specific heat of liquid water. Total precipitation, P, is separated into rainfall, P_R, and snowfall, P_S, using a linear temperature transition after You et al. (2014) as follows

$$P_R = f_R P \text{ and } P_S = (1 - f_R) P \tag{2.39}$$

where the rain fraction, f_R, is given by

$$f_R = \begin{cases} 1, & \text{if } T > T_R \\ \frac{T - T_S}{T_R - T_S}, & \text{if } T_S < T < T_R \\ 0, & \text{otherwise} \end{cases} \tag{2.40}$$

where $T_S = 272.15$ K and $T_R = 276.15$ K are the snow and rain thresholds which are assumed to be constant for simplicity. Note that the heat advected by precipitation is a relatively minor term, it is for example dwarfed by the latent heat flux during rain on snow events.

Both of the turbulent heat fluxes are parametrized through a resistance approach (see Foken, 2008) using Monin-Obukhov Similarity Theory (MOST) following Westermann et al. (2016). Thereby, the latent heat flux is diagnosed through

$$Q_E = \frac{\rho_a L_v}{r_a^W + r_s} \left(q - q_{si}\right) \tag{2.41}$$

where ρ_a is the air density (diagnosed with the ideal gas law), $L_v = 2.5 \times 10^6$ Jkg^{-1} is the latent heat of vaporization, r_s is the adjustable surface resistance against evapotranspiration, r_a^W is the aerodynamic resistance for latent heat, q is the near-surface specific humidity, and q_{si} is the specific humidity at saturation with respect to ice at $T_0 = 273.15$ K which is obtained as a function of surface pressure using

$$q_{si} = \text{w}_{si} / (1 + \text{w}_{si}) \tag{2.42}$$

where the saturation mixing ratio is

$$\mathrm{w}_{si} = \varepsilon_0 C_1 / (\mathrm{p}_s - C_1)\,, \tag{2.43}$$

which equals 0.004 for standard sea level pressure. The sensible heat flux is diagnosed analogously through

$$Q_H = \frac{\rho_a c_p}{r_a^H} \left(T - T_0\right) \tag{2.44}$$

where $c_p = 1005$ Jkg^{-1}K^{-1} is the specific heat of air at constant pressure. Recall that $T_0 = 273.15$ K is the assumed skin (and bulk) temperature of the snowpack, whereas T is the downscaled near-surface air temperature. The aerodynamic resistance can be obtained via

$$r_a^{H,W} = \left(\kappa u_*^{-1}\right) \left[\ln\left(\frac{h}{z_0}\right) - \Psi_{H,W}\left(\frac{h}{L_*}, \frac{z_0}{L_*}\right)\right]^{-1} \tag{2.45}$$

where $\kappa = 0.4$ is the Von Kármán constant, $h = 2$ m is the assumed hypothetical measurement height, $z_0 = 0.001$ m is the assumed roughness length for snow, and

$$u_* = U\kappa \left[\ln\left(\frac{h}{z_0}\right) - \Psi_M\left(\frac{h}{L_*}, \frac{z_0}{L_*}\right)\right]^{-1} \tag{2.46}$$

is the friction velocity whereas

$$L_* = \frac{\rho_a c_p T u_*^3}{\kappa g \left[Q_h + \frac{\varepsilon_0 c_p}{L_v} T Q_e\right]} \tag{2.47}$$

is the Obukhov length. The integrated stability functions are given by (e.g. Foken, 2008)

$$\Psi_{M,H,W} = \int_0^\zeta \left[1 - \varphi_{M,H,W}(\zeta')\right] \, d\ln\zeta' \tag{2.48}$$

where $\zeta = h/L_*$ and $\varphi_{M,H,W}$ are universal stability functions for momentum, heat, and moisture given by MOST. I use the following stability functions (see Westermann et al., 2016, and references therein) for unstable ($\zeta < 0$) and neutral ($\zeta = 0$) conditions

$$\varphi_M = (1 - 19\zeta)^{-1/4} \quad \text{and} \quad \varphi_{H,W} = 0.95(1 - 11.6\zeta)^{-1/2}\,. \tag{2.49}$$

While for stable conditions ($\zeta > 0$) I use the following functions

$$\varphi_M = 1 + \frac{6.5\zeta(1+\zeta)^{1/3}}{1-3+\zeta} \text{ and } \varphi_{H,W} = 1 + \frac{5\zeta(1+\zeta)}{1+3\zeta+\zeta^2}\,, \tag{2.50}$$

to subsequently arrive at numerical look up table approximations for the integral in (2.48) using these stability-dependent universal functions. Note that together Equations (2.41), (2.44), (2.46), and (2.47) constitute a closed set of nonlinear equations that I solve numerically to diagnose the turbulent heat fluxes. There are limitations to the applicability of MOST tied to the underlying assumptions. This is particularly the case under stable stratification where assumptions of flow stationarity and homogeneity are prone to breakdown (Mahrt, 1999). Still, MOST presents the best currently available parametrizations for diagnosing turbulent heat fluxes in land surface models. Simpler alternative parametrizations based on the bulk Richardson number are often applied in snow models (e.g. Essery, 2015).

The final term that needs to be diagnosed to be able to solve for the snowmelt flux is the ground heat flux (Q_G). For areas where this flux is important, such as permafrost regions, this is flux is parametrized during the ablation season as in Aalstad et al. (2018) using an exponential decay formulation as follows

$$Q_G = Q_0 \text{exp}\left(-d_H t_m / z_E^2\right) \tag{2.51}$$

where Q_0 is the initial ground heat flux before the onset of the ablation, d_H is the thermal diffusivity of the ground, and z_E is the effective depth of heat transfer below the base of the snowpack, whereas t_m is the current number of days with melting conditions after peak accumulation. With the possible exception of d_H, the parameters in (2.51) are uncertain and should be estimated using e.g. data assimilation or calibration techniques. With the remaining surface energy balance terms diagnosed, we may now solve for the snowmelt flux which enters into the mass balance. We typically diagnose the surface energy balance terms at an hourly timescale which is the temporal resolution of the downscaled forcing from TopoSCALE. The snowmelt flux may subsequently be aggregated to a coarser (e.g. daily) temporal resolution. Note that this would not be possible if we modeled the cold content (i.e. energy storage) of the snowpack.

2.2.2 Mass balance

The mass balance is diagnosed through the net accumulation rate $\mathcal{A}$ (units ms^{-1})

$$\mathcal{A} = P_S + P_R - M - E\,, \tag{2.52}$$

were P_S is the snowfall rate, P_R is the rainfall rate, M is the snowmelt rate, and E is the sublimation rate. The snowfall and rainfall rates have already been obtained from (2.39). I set the rainfall rate to zero in cases where the accumulated melt depth (to be defined) is non-zero, assuming that in such cases the snowpack has reached its liquid water holding capacity. The snowmelt rate is obtained from the snowmelt flux through

$$M = \max\left(\frac{Q_M}{\rho_w L_f}, 0\right)\,, \tag{2.53}$$

whereas the sublimation rate is obtained following Tarboton and Luce (1996) through

$$E = \frac{Q_E}{\rho_w L_s}\,, \tag{2.54}$$

where $L_s = 2.835 \times 10^6$ Jkg^{-1} is the latent heat of sublimation.

The net accumulation rate is subsequently used to update the accumulated melt depth, D_m, at time step n (timestamp t_n) through

$$D_{m,n+1} = \max\left(D_{m,n} - \mathcal{A}_n \Delta t, 0\right) H\left(\mu_n\right)\,, \tag{2.55}$$

where Δt is the model time step (typically daily), $H(x)$ is the alternative Heaviside function which equals 0 if $x \leq 0$ and 1 otherwise, and μ is the peak mean (i.e. spatial mean of the model grid cell) SWE so far in the snow season which is diagnosed through

$$\mu_{n+1} = \mu_n + \max\left(\mathcal{A}_n \Delta t - D_{m,n+1}, 0\right)\,. \tag{2.56}$$

Given that we are only dealing with seasonal snowpacks, both the peak SWE and accumulated melt depth are initialized as being zero at the start of the water year. They are also reset to zero in cases that the fSCA becomes zero.

2.2.3 Snow depletion curve

The fSCA is diagnosed as a function of peak mean SWE and the accumulated melt depth using the probabilistic snow depletion curve (SDC) formulation of Liston (2004). This formulation assumes that the peak subgrid SWE distribution, f_P, is a log-normal probability density function (PDF), i.e.,

$$f_P(D|\mu,\chi) = \frac{1}{D\widetilde{\sigma}\sqrt{2\pi}}\exp\left(-\frac{(\ln(D)-\widetilde{\mu})^2}{2\widetilde{\sigma}^2}\right), \tag{2.57}$$

where $\widetilde{\mu}$ and $\widetilde{\sigma}$ are the mean and standard deviation of the normally distributed (log transformed) variable $\ln(D)$ which are related to the peak mean SWE and the peak subgrid coefficient of variation $\chi = \sigma/\mu$ (σ is the standard deviation of f_P) through

$$\widetilde{\sigma} = \sqrt{\ln(1+\chi^2)} \text{ and } \widetilde{\mu} = \ln(\mu) - 0.5\widetilde{\sigma}^2\,. \tag{2.58}$$

For clarification, f_P dictates the spatial probability distribution of peak SWE within a grid cell. Examples of distributions with different parameters are shown in Figure 2.4. Important features of the log-normal distribution are that it has positive support (i.e. it is not defined for negative values) and positive skewness (i.e. it has a tail to the right). This makes it well suited for modeling the peak subgrid SWE distribution. Another common choice for this is the gamma distribution (e.g. Kolberg and Gottschalk, 2006). As emphasized by Luce and Tarboton (2004) the value selected for χ is much more important than the choice of parametric distribution, provided that the distribution is skew with positive support. Note that as $\chi \to 0$ then f_P is less skew and tends towards a Dirac delta function.

The probabilistic SDC formulation also assumes that subgrid melt rates are homogeneous, i.e. that the snowmelt flux is spatially uniform within each grid cell. In such a case, the cumulative distribution function evaluated at $D_{m,n}$ will give the fractional snow-free area at time t_n. Thereby, the fSCA is given by the complementary cumulative distribution function, i.e.

$$\text{fSCA}_n = 1 - \int_0^{D_{m,n}} f_P(D|\mu,\chi)\,dD = \int_{D_{m,n}}^{\infty} f_P(D|\mu,\chi)\,dD\,. \tag{2.59}$$

Since f_P is log-normal this has a simple analytical solution given by

$$\text{fSCA}_n = \frac{1}{2}\text{erfc}\left(\mathcal{Z}_{m,n}\right) \tag{2.60}$$

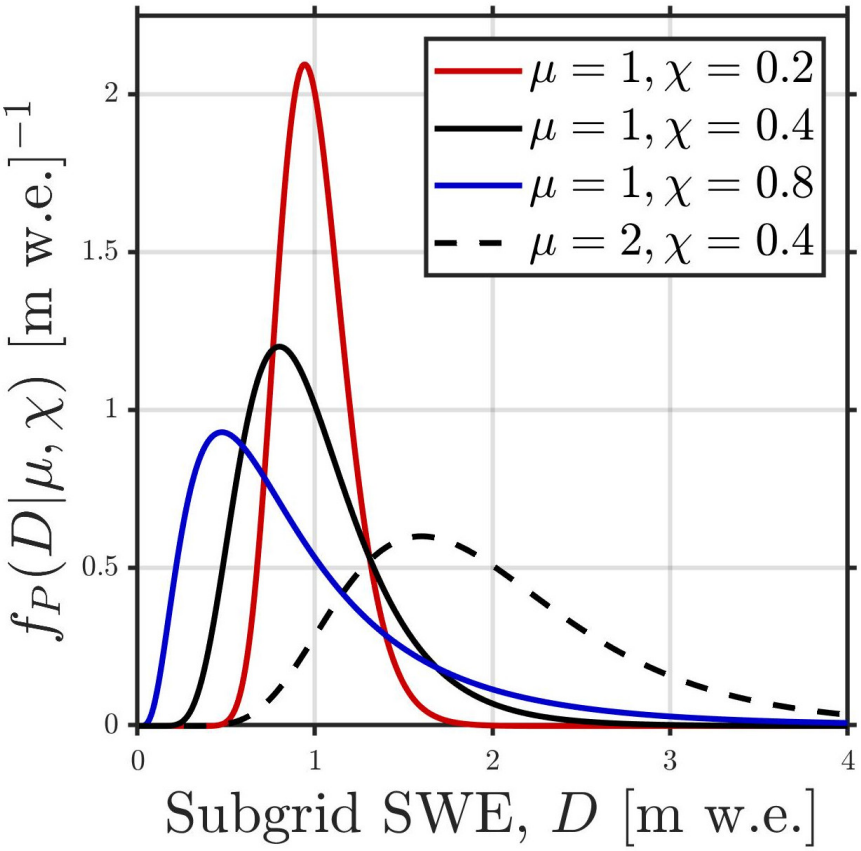


Figure 2.4: log-normal peak subgrid SWE distributions, f_P, as a function of subgrid SWE depth D with different values for the peak mean SWE, μ, and the coefficient of variation χ.

where erfc is the complementary error function (available as a native function in many programming environments) and

$$\mathcal{Z}_{m,n} = \frac{\ln(D_{m,n}) - \widetilde{\mu}}{\sqrt{2}\widetilde{\sigma}} \,. \tag{2.61}$$

The SDC formulation is also used to diagnose the subgrid mean SWE at time t_n, $\overline{D}_n$, through

$$\overline{D}_n = \int_{D_{m,n}}^{\infty} (D' - D_{m,n})\, f_P(D'|\mu_n, \chi)\, dD' \tag{2.62}$$

which has the following analytical solution

$$\overline{D}_n = \frac{1}{2}\exp\left(\widetilde{\mu} + \frac{1}{2}\widetilde{\sigma}^2\right) \text{erfc}\left(\frac{\mathcal{Z}_{m,n} - \widetilde{\sigma}}{\sqrt{2}}\right) - \text{fSCA}_n D_{m,n} \,. \tag{2.63}$$

To account for the infinite tail in f_P, I define the disappearance of the snowpack as when fSCA< 0.01 in which case fSCA, μ, D_m, and $\overline{D}$ are set to zero.

2.2.4 Model evaluation

Given the many simplifying assumptions made in the model, particularly not accounting for compaction or heat conduction within the snowpack, I performed a brief evaluation to gauge the effect of these simplifications. In the evaluation I compared the SWE simulated by SSM to results from the more detailed multi-layer (3 layers) factorial snowpack model (FSM; Essery, 2015) and observations at the Col de Porte site (c.f. Lejeune et al., 2019) in France for the 2006 snow season. In the evaluation, both FSM and SSM were forced by meteorology that was observed in-situ as opposed to downscaled reanalysis data to reduce the effect of forcing error. As can be seen in Figure 2.5, at least for this site, when the forcing is near perfect SSM is able to simulate the evolution of the SWE just as well as the FSM. In fact, it even outperforms or at least matches FSM for all the evaluation metrics considered. I do not expect this to be the case in general, given that FSM accounts for more physical processes, but it is promising that SSM can perform so well despite its simplicity.

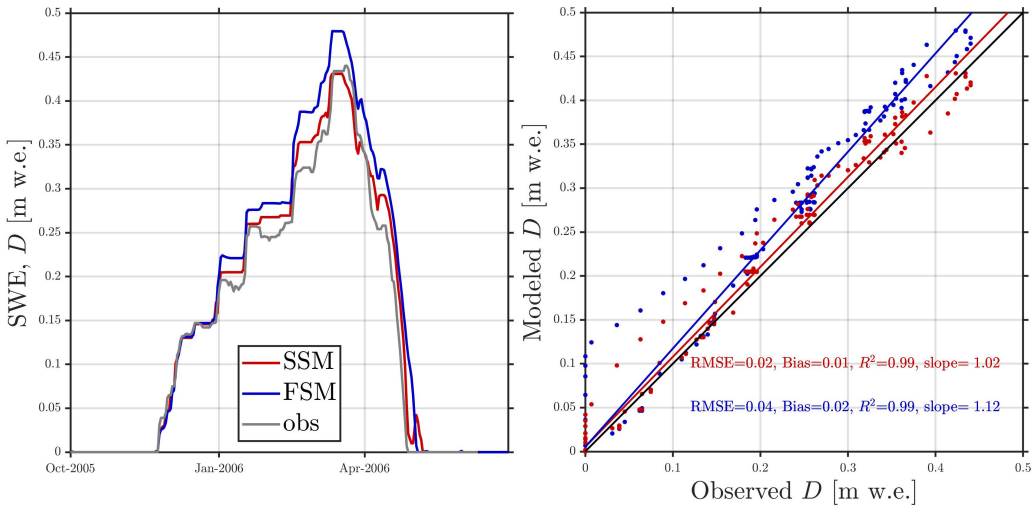


Figure 2.5: Performance of the simple snow model (SSM) relative to the more detailed factorial snowpack model (FSM; Essery, 2015) and observations at the Col de Porte site for the 2006 snow season. **Left:** Time series of SWE for the 2006 snow season. **Right:** Scatter plots (FSM in blue, SSM in red) of modeled vs. observed SWE, the 1:1 line is shown in black and the linear best fit lines are shown in blue (FSM) and red (SSM).

Recall that the major source of uncertainty in snow modeling is typically the hydrometeorological forcing (Raleigh et al., 2015). As such, the philosophy behind the SSM is rather to spend computational resources on the ability to run a large ensemble to allow for data assimilation that can constrain this uncertainty. This is as opposed to running a deterministic run with many physical processes but with erroneous forcing. We know that some physical processes such as the cold content are not always that important. Recall that it takes 160 times more energy to melt a given volume of ripe snow than it does to raise the snow temperature by 1 degree. So the latent heat sink during the snowmelt is typically much more important than the cold content. Of course, accounting for the cold content is important when simulating heat conduction in the snowpack to explicitly couple it to the underlying ground or for avalanche prediction. In the future, I envisage that SSM will adopt many of the features present in intermediate complexity snow models such as FSM. At the same time, I do not see SSM morphing into a fully fledged detailed snow model with tens or even hundreds of layers such as Crocus (Vionnet et al., 2012) or SNOWPACK (Bartelt and Lehning, 2002).

2.3 Optical remote sensing of snow

The next ingredient in ensemble-based snow reanalysis is the optical satellite remote sensing of snow where our main objective is the retrieval of fSCA. Presently, there exists many optical satellite sensors that can obtain multispectral imagery from which fSCA can be retrieved. We are primarily concerned with polar orbiting, rather than geostationary, satellites as these can provide near global coverage.

Optical sensors sense the upwelling solar radiance that has been reflected at some point in its passage from the top of the atmosphere down towards Earth's surface. Using knowledge of the incoming exo-atmospheric solar radiance this can be converted to a top of the atmosphere reflectance. Through atmospheric corrections, this can in turn be used to estimate the surface reflectance. For the purposes of this thesis, I will primarily be concerned with the two types of satellite derived reflectances, namely the the top of the atmosphere reflectance and the surface reflectance. According to Schaepman-Strub et al. (2006) satellite measurements of reflectance closely correspond to Hemispheric-Directional Reflectance Factors (HDRFs), although strictly speaking they are Hemispheric-Conical Reflectance

Factors (HCRFs).

Traditionally it is mainly MODIS and Landsat that have been employed for snow reconstructions. In this thesis I build on this by complementing the multispectral images from these with more imagery from recently launched satellite constellations such as Sentinel-2, PlanetScope, and RapidEye. In this section, I will first review the key spectral properties of snow, then I outline the characteristics of the optical satellite sensors that I have employed, and finally I present the techniques used to retrieve fSCA from these sensors.

2.3.1 Spectral signature of snow

A key aspect that is exploited in the fSCA retrieval algorithms is the unique spectral signature of snow: it is white in the visible part of the electromagnetic spectrum, but surprisingly colorful in the infrared (Dozier et al., 2009). Provided that the snowpack can be considered semi-infite, i.e. provided that it is deep enough (5-10 cm), then the underlying surface will have no effect on the reflectance and the reflectance will provide a 'pure' snow signal (Dozier, 1989). Impurities in the snowpack may nonetheless adversely effect the snow spectrum. In this thesis, for simplicity, these impurities are ignored even though they may be very important radiative forcing contributions due to their darkening effect (e.g. Flanner et al., 2007; Skiles et al., 2018). Techniques do exist to retrieve impurity concentration in the surface layer of the snowpack (Painter et al., 2012).

To derive snow reflectance spectra I employ the Snow, Ice, and Aerosol Radiative (SNICAR) model originally developed by Flanner et al. (2007) and recently refined by He et al. (2018). For simplicity, I use the default version of SNICAR which assumes spherical snow grains. SNICAR models a multilayer two stream radiative transfer within the snowpack using optical properties derived from Mie theory (see He et al., 2018, and references therein). In this way, SNICAR simulates the spectral reflectance of snow as a function of both effective snow grain radius and solar zenith angle (see Figure 2.6). The spectral reflectance of snow increases as the solar zenith angle increases (solar elevation angle decreases) due to the strong forwards scattering properties of snow (Dozier et al., 2009). In addition, the reflectance decreases as the snow metamorphoses and the grain radius increases.

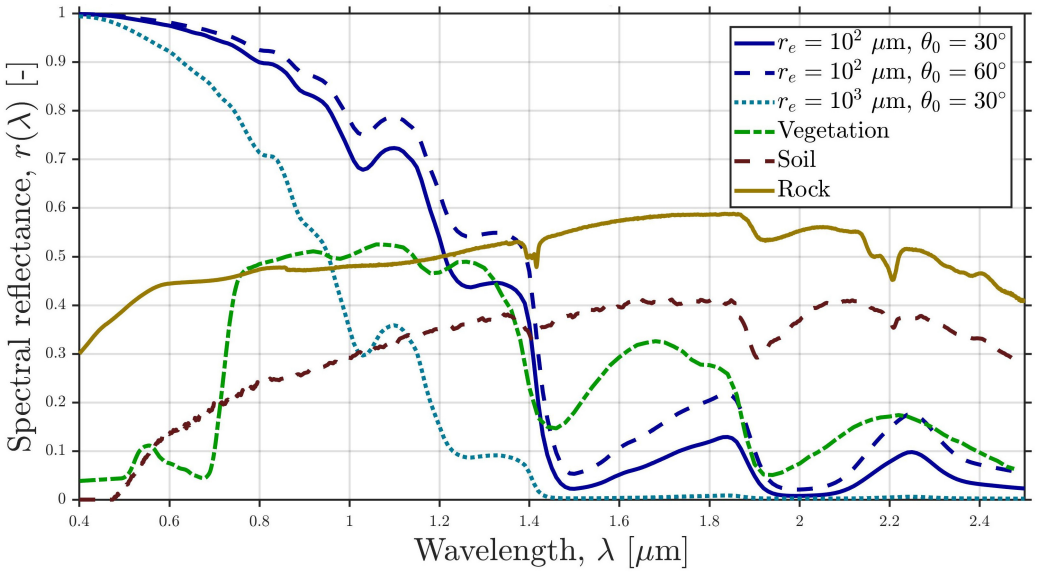


Figure 2.6: Examples of reflectance spectra for snow (in blue) obtained from the SNICAR radiative transfer model (He et al., 2018) for different effective snow grain radii (r_e) and solar zenith angles (θ_0) as well as reflectance spectra for vegetation (grass), soil (pale brown silty clay loam) and rock (gray sandstone) obtained from the JPL spectral library (Baldridge et al., 2009).

I also obtained theoretical reflectance spectra for hundreds of other natural surfaces from the JPL spectral library (Baldridge et al., 2009) which is based on a combination of modeling, laboratory, and field measurements. A representative subset of these is also shown in Figure 2.6. We note how both soil and rock spectral reflectances tend to increase with wavelength from the visible in to the near infrared and shortwave infrared. Vegetation spectra also increase from the visible to the near infrared, but they drop off again towards the shortwave infrared. The snow spectra are quite different as they peak near unity in the visible and drop off quite steadily into the near infrared to near zero in the shortwave infrared. In this sense, the spectral signature of snow is different from most other naturally occurring surfaces. The fact that snow is nearly opaque in the shortwave infrared can also help discriminate snow from clouds, which are otherwise spectrally similar in the visible and near infrared (Painter et al., 2009). Still, certain ice clouds remain spectrally quite similar to snow even in the shortwave infrared which makes automated cloud/snow discrimination challenging (Musial et al., 2014; Selkowitz and Forster, 2016). Note that it is the spectral signature of snow that I exploit in false color imagery in this thesis when I visualize the shortwave infrared band as red, the near infrared band as green, and the red band as blue. With such a band combination snow will stand out as a relatively unique turquoise (due to

the low shortwave infrared reflectance, but relatively high red and near infrared reflectance) with most other surfaces appearing in different colors.

2.3.2 Optical sensors

In my work I have employed multispectral satellite imagery from a variety of optical sensors: the Moderate Resolution Imaging Spectroradiometer (MODIS) on-board the Aqua and Terra satellites, the Enhanced Thematic Mapper Plus (ETM+) on-board the Landsat 7 satellite, the Operational Land Imager (OLI) on-board the Landsat 8 satellite, and the MultiSpectral Instrument (MSI) on-board the Sentinel-2A and Sentinel-2B satellites. These sensors all have bands in the shortwave infrared, making the discrimination of snow from other natural land surfaces relatively simple provided that a satellite pixel is either snow-free or completely snow-covered. The spectral properties of these sensors, i.e. the spectral response functions for each of the bands that I employed, are displayed in Figure 2.11. For each of these sensors I used the following six bands: blue (~ 0.49 μm), green (~ 0.56 μm), red (~ 0.66 μm), near infrared (~ 0.83 μm), shortwave infrared 1 (~ 1.6 μm), and shortwave infrared 2 (~ 2.2 μm). I also complemented imagery from these sensors with higher resolution imagery from the PlanetScope and RapidEye constellations. These latter constellations only have bands in the visible and the near infrared, but not in the shortwave infrared, thus requiring different techniques for the retrieval of snow-cover. Here, I will briefly review the key characteristics of each of these sensors.

MODIS, operated by the National Aeronautic and Space Administration (NASA), is currently in polar orbit on-board the Aqua and Terra satellites and delivers multispectral imagery in 36 bands from 0.4 μm through to 14.4 μm in a wide swath. Terra was launched in December 1999 whereas Aqua was in May 2002. Thereby, MODIS has been delivering an invaluable climate data record for almost 20 years. Thus, this record may now be exploited to perform investigations of the climatological evolution of the snowpack over the last two decades (e.g. Bormann et al., 2018; Yılmaz et al., 2019). As previously noted, we are primarily interested in the bands that span the visible through to the shortwave infrared (up to ~ 2.2 μm) part of the electromagnetic spectrum (henceforth VSWIR). Given its wide swath the equatorial revisit period of the MODIS sensor at the equator is quasi daily for both Terra and Aqua, such that when these two satellites are combined the revisit period is sub-daily particularly at higher latitudes. In the VSWIR MODIS

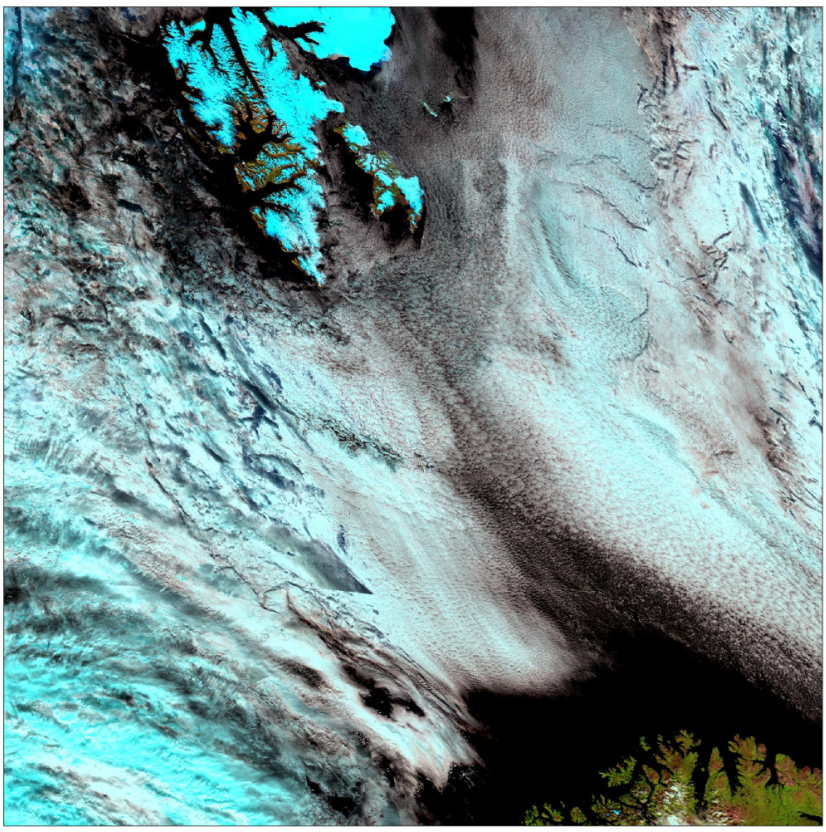


Figure 2.7: MODIS false color image showing the entire h18v01 tile (roughly 1.4 million km^2) that covers most of the Svalbard archipelago (top center) and the northern tip of mainland Norway (bottom left). This image is based on the MOD09GA (version 6) surface reflectance product.

delivers imagery at an effective ground sample distance (GSD) of approximately 500 m for all bands (250 m for the red and near infrared band). Of the sensors employed in this study, MODIS has the coarsest GSD. Nonetheless, the daily revisit period and the high number of bands (resulting in quite reliable cloud masks) have made MODIS an ideal sensor for developing operational snow-cover products (Painter et al., 2009; Riggs et al., 2017). An example MODIS image has been included in Figure 2.7. I employ versions 5 and 6 of the MODIS snow-cover product (MOD10A1; Riggs et al., 2017), version 6 of the MODIS surface reflectance product (MOD09GA; Vermote and Wolfe, 2015), and the MODIS Snow Covered-Area and Grain size (MODSCAG; Painter et al., 2009) product.

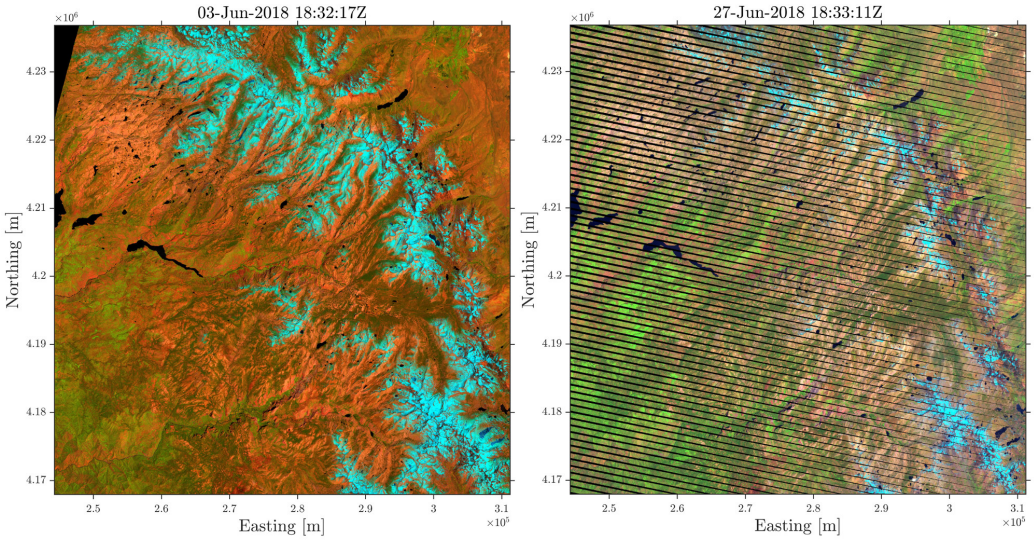


Figure 2.8: Example false color imagery from Landsat 7 ETM+ (right) and Landsat 8 OLI (left) over the Tuolumne basin in the Californian Sierra Nevada. The stripes in the ETM+ image show the scan line corrector failure. These images are based on the level 1 top of the atmosphere reflectance products from OLI and ETM+. For information, the axis ticks are 10 km apart.

The Landsat program, whose satellite's have been in orbit since 1972, is a joint venture between the NASA and the U.S. Geological Survey (USGS) to survey Earth's land areas. The ETM+ and OLI sensors on-board Landsat 7 and Landsat 8, respectively, share similar characteristics to the earlier Landsat instruments such as the Thematic Mapper (TM) on-board Landsat 4 and Landsat 5 which are no longer in orbit. All these sensors provide VSWIR imagery at an effective GSD of 30 m with a 16 day repeat period at the equator. Landsat 7 was launched in April 1999 such that ETM+ has been in orbit for over 20 years. Unfortunately, the scan line corrector of the ETM+ sensor has not been working since 2003 such that much of the ETM+ is corrupted with gaps. Only the central part of each ETM+ tile is gap free, and the gaps (stripes) become more severe towards the edges of each tile. Nonetheless, ETM+ still provides a valuable data record that has been successfully employed to map snow-cover (e.g. Selkowitz and Forster, 2016). Landsat 8 was launched in February 2013 meaning that OLI has been in orbit for over 6 years, complementing ETM+ after the failure of its scan line corrector. OLI has considerably higher radiometric resolution (i.e. it has narrower spectral response functions) than ETM+, making OLI better able to detect fine scale spectral features. The Landsat constellation has a long history of use in snow reconstruction (e.g. Cline et al., 1998) forming an integral component of the recently conducted Sierra Nevada snow reanalysis (Margulis et al., 2016). Com-

pared to MODIS, these Landsat sensors provide imagery at a significantly higher spatial resolution at the cost of a coarser temporal resolution (lower acquisition frequency). I employ the level 1 top of the atmosphere reflectance product and the level 2 surface reflectance product from both ETM+ and OLI (c.f. Vermote et al., 2016). Example false color imagery using top of the atmosphere reflectances in the shortwave infrared (1.6 μm band), near infrared, and red band from ETM+ and OLI is shown in Figure 2.8.

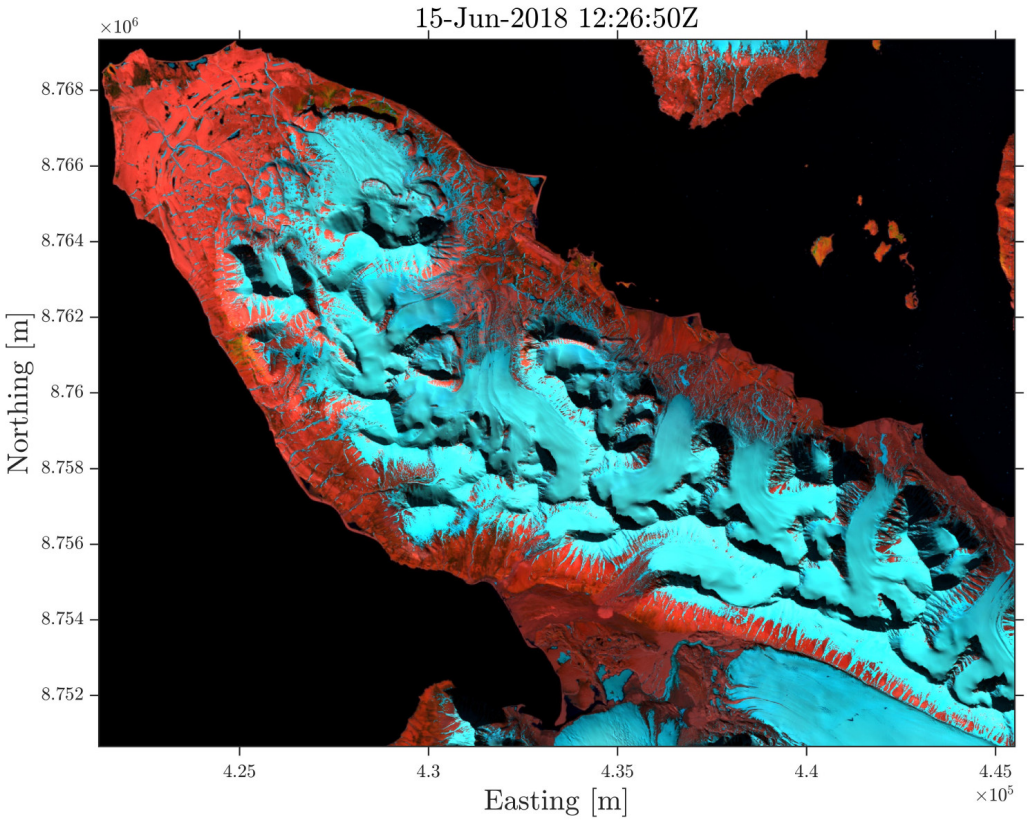


Figure 2.9: Example false color imagery from Sentinel-2B MSI over the Brøgger peninsula in the Svalbard archipelago. This image is based on the level 2 surface reflectance product. For scale, the x-axis tick marks are 5 km apart while the y-axis tick marks are 2 km apart.

The Sentinel-2 satellites are operated by the European Space Agency (ESA) as part of the European Union's Copernicus Program to monitor land areas and coastal waters. The MSI sensor on-board the Sentinel-2 satellites is quite similar

to the Landsat sensors in that it provides higher resolution imagery than MODIS but at a somewhat coarser temporal resolution (Drusch et al., 2012). In particular, MSI provides multispectral imagery at an effective GSD of 20 meters (10 meters for the visible and near infrared bands) and the Sentinel-2A and Sentinel-2B twin satellites both have an equatorial revisit period of 10 days. So, combining Sentinel-2A and Sentinel-2B, MSI has a revisit period of only 5 days over the equator which is a marked improvement over the Landsat constellation. In polar regions, where the orbits converge, this revisit period is considerably lower. The data record is, however, considerably shorter than that of the Landsat constellation given that Sentinel-2A was launched in June of 2015 and Sentinel-2B was launched in March of 2017. Nonetheless, the Sentinel-2 satellites are remarkably complementary to the Landsat constellation and are intended to continue to provide temporally dense high-resolution imagery for several years to come. I employ the level 1 top of the atmosphere reflectance product and the level 2 surface reflectance product from Sentinel-2 MSI (c.f. Richter et al., 2012; Müller-Wilm, 2018). An example Sentinel-2 MSI false color image is shown in Figure 2.9.

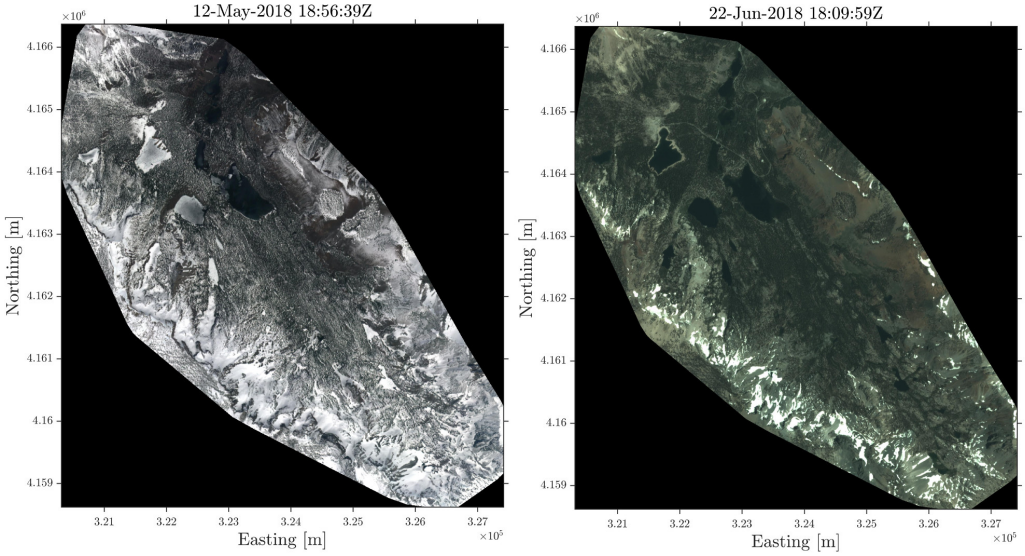


Figure 2.10: Example true color imagery from the RapidEye (left) and PlanetScope (right) constellation over the Mammoth Lakes basin in the Californian Sierra Nevada. These images are based on the level 1 top of the atmosphere reflectance products from RapidEye and PlanetScope. For information, the axis ticks are 1 km apart.

I also employ imagery from the RapidEye and PlanetScope satellite constellations that are currently operated by the private company Planet (Planet Team, 2019). The PlanetScope constellation is a collection of around 150 optical miniaturized satellites (c.f. Kääb et al., 2019) whereas the RapidEye constellation is made up of 5 optical satellites. The PlanetScope and RapidEye constellations are both in near-polar orbit and deliver imagery in visible (RGB) and near infrared bands at a pixel resolution of 3 m and 5 m, respectively. As of 2017, PlanetScope delivers daily imagery for most of our planet's land areas, whereas RapidEye has been delivering imagery at irregular intervals since its launch in 2008. The sensors in both the RapidEye and PlanetScope constellation are currently not able to sense in the shortwave infrared part of the spectrum, making snow/cloud discrimination even more challenging than with the previously discussed constellations. Example true color imagery from RapidEye and PlanetScope is shown in Figure 2.10.

2.3.3 Retrieving fSCA

Here I briefly summarize several techniques that can be used to retrieve fSCA from VSWIR and visible near infrared (VNIR) satellite imagery. Note that in Aalstad et al. (2019a) we provide a thorough evaluation of several different techniques used to retrieve fSCA from multiple optical satellite sensors. Some of these techniques were also employed in the earlier study of Aalstad et al. (2018). This section provides a summary of the material covered in more depth in Aalstad et al. (2019a), and also lends some more detail as to how fSCA was obtained from the VNIR imagery from PlanetScope and RapidEye used in Aalstad et al. (2019b).

For VSWIR imagery the fSCA retrieval techniques exploit the aforementioned spectral signature of snow, i.e. the strong decrease in reflectance from the visible to the shortwave infrared part of the electromagnetic spectrum over snow-covered surfaces. The most straightforward way to characterize this signature is to employ the normalized difference snow index (NDSI) defined as (Dozier, 1989)

$$\mathrm{NDSI} = \frac{r_{\mathrm{Green}} - r_{\mathrm{SWIR1}}}{r_{\mathrm{Green}} + r_{\mathrm{SWIR1}}}, \tag{2.64}$$

where r_{Green} is the reflectance in the green band and r_{SWIR1} is the reflectance in the shortwave infrared 1 band. In general these may either be top of the atmosphere or surface reflectances. The work of Aalstad et al. (2019a) indicates that the type of reflectance used should not make much of a difference, although correcting for

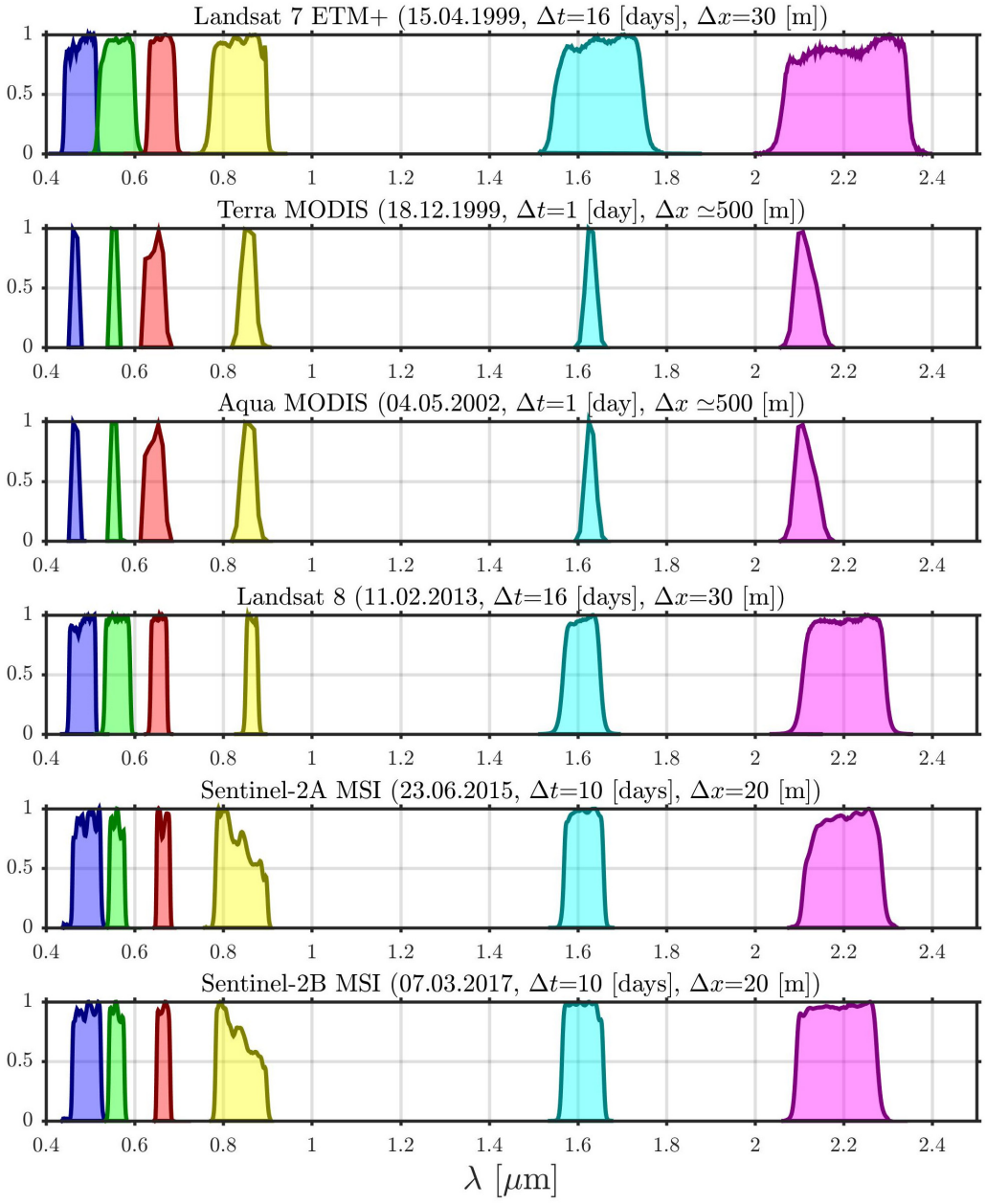


Figure 2.11: Spectral response functions for the bands of most of the sensors employed in this work. The six bands used are the blue (~ 0.49 μm, blue curve), green (~ 0.56 μm, green curve), red (~ 0.66 μm, red curve), near infrared (~ 0.83 μm, yellow curve), shortwave infrared 1 (~ 1.6 μm, turquoise curve), and shortwave infrared 2 (~ 2.2 μm, purple curve). The launch date, equatorial repeat period, and (maximum) ground sampling distance is shown in the title of each panel. Note that the spectral response functions here have been normalized by their maximum values (i.e. they are not probability densities). The response functions were obtained from EUMETSAT.

topographic effects on the reflectance can be important in complex terrain (see Meyer et al., 1993). The NDSI will be large (closer to unity) for snow-covered pixels and will tend to zero or even negative values for other surfaces (see Figure 2.6). Note that the NDSI is defined relatively analogously to the normalized difference vegetation index

$$\text{NDVI} = \frac{r_{\text{NIR}} - r_{\text{Red}}}{r_{\text{NIR}} + r_{\text{Red}}}, \tag{2.65}$$

where r_{Red} is the reflectance in the red band and r_{NIR} is the reflectance in the near infrared band. The NDVI is used to map vegetation by exploiting the fact that vegetation is relatively opaque in the red part of the spectrum but highly reflective in the near infrared. I will return to the NDVI when I set out how to map snow from VNIR imagery.

The simplest way to obtain fSCA using the NDSI is just to set a pixel-specific threshold. Typically a threshold of $\tau_{\text{NDSI}} = 0.4$ is used (Hall et al., 2002; Gascoin et al., 2015; Wunderle et al., 2016), such that pixel-wise binary 'fSCA' is given a binary value as follows

$$\text{fSCA} = \begin{cases} 1 & \text{if NDSI} > \tau_{\text{NDSI}} \\ 0 & \text{otherwise} \end{cases}. \tag{2.66}$$

Although a threshold of 0.4 is widely employed it may not necessarily be the optimal threshold for a specific scene, as emphasized by Yin et al. (2013) and Härer et al. (2018). Adaptive segmentation methods such as the technique of Otsu (1979), which is based on minimizing the intra-class variance, may provide automatic means of improving the threshold selection. Nonetheless, Aalstad et al. (2019a) did not find any significant improvement when using an automatic adaptive as opposed to fixed NDSI threshold. Additional thresholds may also be set on bands to account for clouds or shadows (see Aalstad et al., 2019a). Once the fSCA has been set at the pixel level using thresholding rules such as (2.66) then the fSCA is upscaled to the snow model resolution (typically in the range 100 m - 1 km) through spatial averaging of the binary fSCA pixels. Obviously such simple thresholding techniques would not be suitable when using MODIS where the pixel resolution is already 500 m, since this would entail resolving the fSCA in increments of 0.25 at a model resolution of 1 km. Instead, this binary thresholding is more suitable for use with higher resolution satellite imagery such as that

from the Landsat and Sentinel-2 constellations. A generic problem with binary thresholding is the fact that it is binary and so it will not accurately account for mixed pixels, that is to say pixels that are partly snow-covered.

A simple way to alleviate the mixed pixel problem is to employ a linear regression on the NDSI (see Salomonson and Appel, 2004). Here, the fSCA is estimated for each pixel as a linear function of the NDSI as follows

$$\text{fSCA} = \beta_1 \text{NDSI} + \beta_0 \,, \tag{2.67}$$

where the β_i are regression coefficients. These coefficients can be estimated by regressing MODIS NDSI against fSCA derived by a more accurate technique, such as using higher resolution satellite imagery, and then inverting the relationship. The work of Salomonson and Appel (2004) suggests $\beta_1 = 1.45$ and $\beta_0 = -0.01$ as appropriate values when using MODIS. It is this kind of relationship that was employed in version 5 of the MOD10A1 MODIS snow-cover product. The relationship can also be used to obtain fSCA from the newer version (collection) 6 of the MOD10A1 product which provides NDSI values (see Riggs et al., 2017). Moreover, this simple linear regression approach can be applied to higher resolution imagery as a simple way to deal with the mixed pixel problem, given that we would expect pixels with a higher fSCA to also have a higher NDSI regardless of the spatial resolution.

Arguably the most elegant and physically-based way to deal with the mixed pixel problem when retrieving fSCA is the use of spectral unmixing (SU; Keshava and Mustard, 2002). The use of unmixing for fSCA retrieval was pioneered by Nolin et al. (1993) and has since been applied on a number of occasions (e.g Vikhamar and Solberg, 2003; Sirguey et al., 2009; Painter et al., 2009; Cortés et al., 2014). It is worth highlighting that products such as binary snow-extent are arguably not that meaningful given that areas with seasonal snow regularly present mixed pixels as opposed to fully snow-covered pixels even during winter. Reporting a pixel as either fully snow-covered or not snow-covered is not only unrealistic, but it can also lead to bias upon aggregation. The SU approach accounts for the fact that a pixel can contain multiple surfaces, or endmembers as they are called in the unmixing approach. This is done by a priori stipulating a series of possible endmembers with known reflectance spectra and hypothesizing that the observed reflectance in each band is the result of a (linear) combination of

these endmembers. Mathematically we may express this as follows (Aalstad et al., 2019a)

$$\mathbf{a}^{\star} = \underset{\mathbf{a}}{\operatorname{argmin}} \left| \mathbf{M} \cdot \mathbf{a} - \mathbf{r} \right|^2 \text{ subject to } \mathbf{a} \geq \mathbf{0} \text{ and } \mathbf{1}^{\mathrm{T}} \cdot \mathbf{a} = 1 \,, \tag{2.68}$$

where $\left| \cdot \right|$ is the ℓ^2-norm, the $N_b \times N_m$ (number of bands times number of endmembers) matrix $\mathbf{M}$ contains the theoretical reflectances for each endmember (columns) averaged to the bands of the sensor in question (rows), the $N_m \times 1$ vector $\mathbf{a}$ contains the fractional abundances of each endmember, the $N_b \times 1$ vector $\mathbf{r}$ contains the retrieved surface reflectances in each band, and $\mathbf{1}^{\mathrm{T}}$ is a $1 \times N_b$ vector of ones. The physical constraints on the fractional abundances of the endmembers are the abundance non-negativity constraint and the abundance sum-to-one constraint. Note that (2.68) is a constrained least squares problem and not an exact inversion of the form $\mathbf{a} = \mathbf{M}^{-1}\mathbf{r}$ because the matrix $\mathbf{M}$ is typically rank deficient (i.e. the columns are not linearly independent) and so it is not invertible even when it is square ($N_b = N_m$).

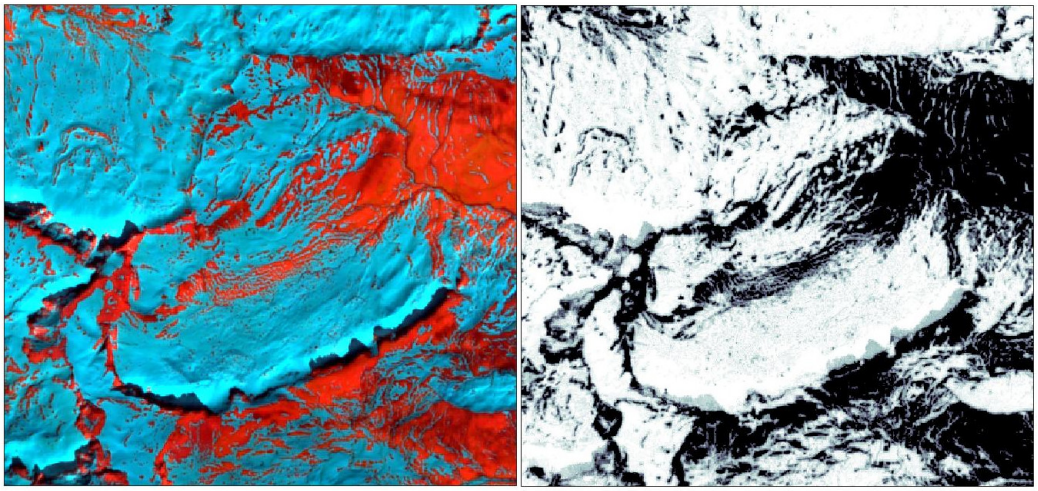


Figure 2.12: **Left**: Sentinel-2A MSI false color image over a 22 km^2 area surrounding the Izas catchment sensed on the 16.04.2017. **Right**: fSCA map (black is bare, gray is partially snow-covered, and white is fully snow-covered) retrieved using spectral unmixing on the same image.

A widely used technique to solve the fully constrained least squares linear SU problem in (2.68) in practice is provided by Heinz and Chang (2001). The solution will nonetheless depend on how the matrix $\mathbf{M}$ is designed, i.e. which endmember

spectra are used. In my case, I specify the non-snow endmember spectra based on knowledge of the study area in question, typically assigning one rock, one soil, and one vegetation endmember. It is of course possible to implement more sophisticated automated techniques for endmember selection, but this will increase the computational cost. For the snow endmembers, I use the previously introduced SNICAR radiative transfer model (Flanner et al., 2007; He et al., 2018). Snow spectral reflectance depends on the solar geometry, but this can be set a priori based on the sensing time and location. As in Cortés et al. (2014) I ignore azimuthal effects, since these are not expected to be large for near-nadir looking sensors such as those on-board Landsat and Sentinel-2. Thus, the snow spectra to be considered are set by the local solar zenith angle at the scene sensing time. The snow grain radius remains a free parameter, and as is in Painter et al. (2009) I loop over all the radii in my look up table and find the snow endmember spectrum that provides the lowest misfit. That is to say, I minimize the solution $|\mathbf{M} \cdot \mathbf{a}^{\star} - \mathbf{r}|$ over all snow grain radii. Thus, the constrained least squares problem in (2.68) has to be solved for each possible grain radius in the look-up table. Nonetheless, the unmixing technique is still relatively computationally efficient and it is easy to envisage means to speed it up further. When employing SU it is important to use surface reflectance products from satellites since the theoretical endmember spectra in the matrix $\mathbf{M}$ are typically surface reflectances. Once an optimal abundance solution has been found, the fSCA is simply the snow endmember abundance. An example of a spectrally unmixed image is provided in Figure 2.12.

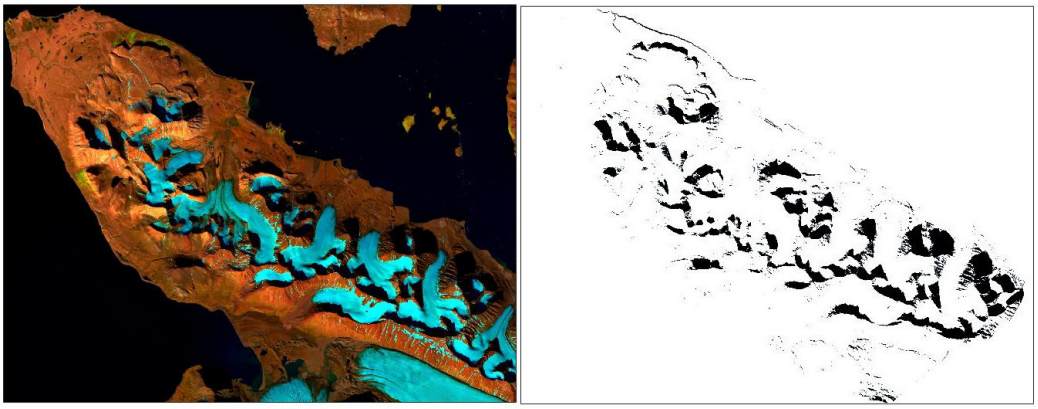


Figure 2.13: **Left**: Sentinel-2A MSI false color image over the Brøgger peninsula sensed on the 31.07.2017 at 12:47:03Z. **Right**: A shadow mask (black areas are shadows) derived independently for the same scene based on a DEM and solar geometry obtained from an ephemeris.

For PlanetScope and RapidEye imagery, which lack a shortwave infrared band, SU is not feasible. Instead I employ a two-step approach based on image segmentation and the NDVI inspired by the Fmask algorithm of Zhu and Woodcock (2012). In the first step, I start by computing the whiteness which I define as the mean visible top of the atmosphere reflectance. I then derive a scene specific whiteness threshold (bounded below by 0.3) using the segmentation technique of Otsu (1979). Any pixels with a whiteness exceeding this threshold are flagged as being potentially snow-covered. Because bright surfaces other than snow, such as bright rocks or sand, may exist in a scene I perform a second step where only the potential snow pixels with an NDVI less than 0.2 are finally flagged as snow-covered. The reason I use a low positive threshold on the NDVI is that rocks and sand typically have a positive spectral slope in the red to near-infrared range whereas snow has a negative spectral slope in this range.

Shadows can be a problem for the fSCA retrieval algorithms, since their darkening effect may hide the spectral signature of snow even over fully snow-covered pixels. As such, for deeply shadowed pixels it may be difficult to determine if the pixel is fully (or partially) snow-covered or completely bare. Fortunately, it is quite straightforward to diagnose the extent of cast shadows at the satellite scene sensing time by combining a DEM with a solar geometry from an ephemeris. For this purpose, I employ the TopoToolbox (Schwanghart and Scherler, 2014) using solar geometry obtained from the ephemeris in the SNL (2014) toolbox. The fractional shadow-covered area for each pixel is then stored, and upon aggregation the fSCA retrieval for a model grid cell is ignored if the aggregated fractional shadow-covered area exceeds a certain threshold (typically 25%). An example of one of the derived shadow masks is shown in Figure 2.13.

As mentioned, it is necessary to aggregate from the retrievals from the finer scale of the satellite pixels to the coarser model resolution (typically 100 m - 1 km). For this aggregation I employ a weighted spatial averaging approach called area weighted averaging. As the name suggests, this is a simple weighted average where each pixel is weighted by the fraction of its area that overlaps with the model grid cell in question. These weights are normalized to sum to one to ensure a proper averaging procedure. The number of pixels contributing to a model grid cell will depend on the model resolution as well as the satellite sensor employed. Running the snow model at 1 km resolution and using MODIS fSCA retrievals, there would be between 4 and 9 MODIS pixels contributing to each model grid

cell. If, instead, Sentinel-2 fSCA retrievals were employed at the same resolution there would be roughly 10^4 pixels contributing to the retrieved fSCA for a 1 km model grid cell. One obvious reason for preferring higher resolution imagery is that random errors cancel out when aggregated in larger numbers. Systematic error (i.e. bias), on the other hand, does not cancel out upon aggregation. In this section, I have briefly shown how fSCA can be retrieved using different algorithms for different sensors, and how these retrievals can be aggregated (upscaled) to the desired model resolution.

2.4 Data assimilation

The final step in the ensemble-based snow reanalysis framework is data assimilation (DA) which brings the forcing, the simple snow model, and the satellite retrievals together. Figure 2.14 shows how DA fits into the workflow of the general snow reanalysis framework. I noted already in the introduction how data assimilation should be viewed as a framework for combining models with data (i.e. some form of observation). The term 'data assimilation' comes from the fact that it is an approach to assimilate (i.e. ingest) observational data into a mechanistic model of some sort. In their sublime tutorial, Wikle and Berliner (2007) define DA informally as "*work in progress*" and more formally as "*an approach for fusing data (observations) with prior knowledge (e.g. mathematical representations of physical laws; model output) to obtain an estimate of the distribution of the true state of a process*." A recent overview of the current state of DA in the geosciences, including an extensive literature review, has been conducted by Carrassi et al. (2018). Recently, DA has slowly started to move beyond the confines of the (already trans-disciplinary) field of geosciences and has found use in other fields such as medicine (Albers et al., 2018).

It is the use of a mechanistic (typically dynamic) modeling that sets DA apart from approaches such as traditional statistical inference or the currently rather voguish field of machine learning (e.g. MacKay, 2005; Goodfellow et al., 2016) which are typically purely data driven. A downside with purely data driven approaches is that the resulting trained models (such as neural networks) are basically black boxes that struggle to generalize, this is not the case when one uses mechanistic models based on prior scientific knowledge. Another downside with purely data-driven approaches is that they can not be used on their own to perform causal

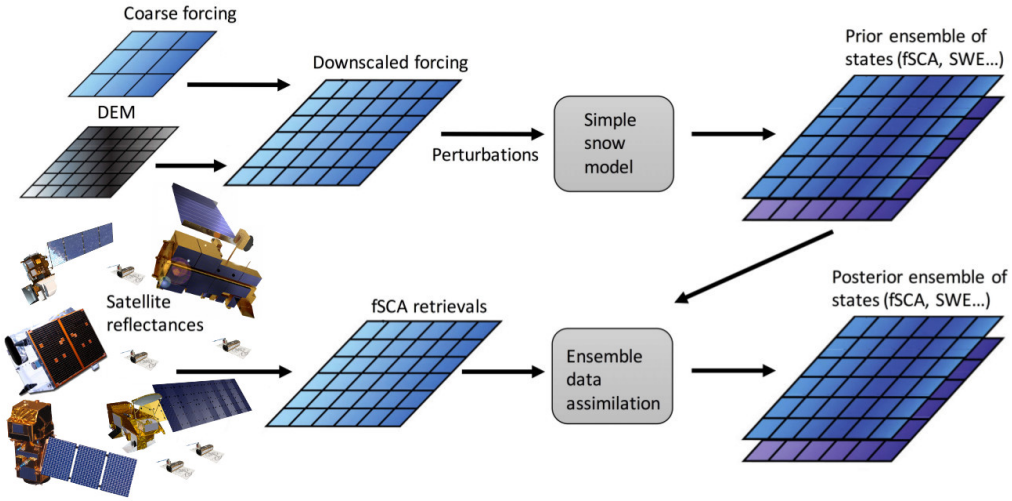


Figure 2.14: A schematic overview of the workflow in the reanalysis framework. Adapted from Durand et al. (2008).

inference (Pearl, 2009), such as event attribution, fortunately DA does not have this limitation (Carrassi et al., 2016). That is not to say that machine learning is not valuable. For example, machine learning approaches such as Gaussian process models have shown promise in emulating more complex mechanistic models, allowing for otherwise prohibitively expensive MCMC parameter estimation (Fer et al., 2018). Both DA and machine learning have in common that they can be cast in the wider frame of Bayesian inference (Wikle and Berliner, 2007; MacKay, 2005). This is the route that I will follow.

In this section I will introduce data assimilation in the Bayesian context, discuss the generation of the prior ensemble, and finally outline the DA schemes used in my work. These schemes are the ensemble smoother, the ensemble smoother with multiple data assimilation, and the particle batch smoother. All of the schemes that I will present are batch smoothers that are used to update uncertain time-invariant parameters. Being batch smoothers, the schemes are sensitive to the entire (annual) time trajectory of the model and observations which allows the future to update the past, a feat that is not achievable using filters. In this sense, I am solving a special static parameter estimation case of the general data assimilation problem (i.e. combining models and data) which is more traditionally known as inversion (see Tarantola, 2005; Stuart, 2010) or Bayesian calibration (Kennedy and O'Hagan, 2001). In the following, I will as closely as possible follow conven-

tional notation in the DA literature (Ide et al., 1999).

2.4.1 Bayesian inference

In statistics, there are two major interpretations of probability (see Lyons and Wardle, 2018, for a physicists perspective). The first is the frequentist or objective interpretation which stipulates that probability is the relative frequency of an event in an infinite, or at least very long, series of repeated identical and independent trials. The other is the Bayesian or subjective interpretation in which probability is a measure of the degree of belief in something. There is, or at least was, some controversy in the statistics literature surrounding which of these interpretations is superior or which one should be followed (e.g. Efron, 1986). In adopting the Bayesian interpretation I will not dwell on these issues, but note in passing that it is hard to see how inference is ever not subjective, in that it always requires assumptions (MacKay, 2005). Moreover these assumptions need not be problematic, so long as they are made explicit then unique, reproducible, and objective inferences can be made. Crucially, under the Bayesian interpretation probability becomes an expression of uncertainty.

The axioms of probability amount to (Jazwinski, 1970): the probability of an event (denoted Pr) is non-negative, the probability of all events in an event space is equal to 1, and the joint probability of mutually exclusive events is zero. This is true regardless of which probability interpretation is followed. A quite generalized form of Bayes theorem (or rule) then amounts to

$$\Pr(\mathrm{H}|\mathrm{E},\mathrm{A}) = \frac{\Pr(\mathrm{E}|\mathrm{H},\mathrm{A})\Pr(\mathrm{H}|\mathrm{A})}{\Pr(\mathrm{E}|\mathrm{A})}\,, \tag{2.69}$$

where H is the hypothesis, E is the evidence, and A is the assumptions, '|' signifies 'given that', and ',' (as usual) signifies 'and'. This theorem is in itself quite uncontroversial and follows directly from the axioms of probability and the definition of conditional probability as shown in Appendix B. The term of interest, $\Pr(\mathrm{H}|\mathrm{E},\mathrm{A})$, i.e. the probability of the hypothesis being true given the evidence and the assumptions, is known as the posterior probability, or sometimes the inverse probability. It is called the inverse probability because it typically maps from effect to cause. For example, it could be the probability that we have a disease given that we test positive. It is typically difficult to quantify independently. The term $\Pr(\mathrm{E}|\mathrm{H},\mathrm{A})$ is known as the likelihood, or sometimes the forwards probability. It is

called forwards probability because it maps from cause to effect. Staying with the same example, this would be the probability of testing positive given that one has a disease. The denominator, the probability of the evidence, is often just referred to as 'the evidence'. This would be the probability of testing positive regardless of whether or not you have the disease. The term $\mathrm{Pr}(\mathrm{H}|\mathrm{A})$ is known as the prior. This would be the probability of having the disease prior to testing. This term can be difficult to specify 'objectively', but fortunately, following MacKay (2005), we have reminded ourselves that the entire inference is conditional on the assumptions. From now on, we will just bear this in mind and keep the assumptions implicit for economy.

In the context of DA, I would instead reformulate Bayes theorem as follows

$$\mathrm{Pr}(\mathcal{M}|\mathcal{D}) = \frac{\mathrm{Pr}(\mathcal{D}|\mathcal{M})\mathrm{Pr}(\mathcal{M})}{\mathrm{Pr}(\mathcal{D})}\,, \tag{2.70}$$

where $\mathcal{M}$ is my model and $\mathcal{D}$ is the observed data. In geophysics we typically deal with continuous variables and parameters, so I would further write this as

$$p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})p(\mathbf{x}, \boldsymbol{\theta})}{p(\mathbf{y})}\,, \tag{2.71}$$

where $p(\cdot)$ denotes a PDF, $\mathbf{x}$ is a $N_s \times 1$ vector containing the model state, $\boldsymbol{\theta}$ is a $N_p \times 1$ vector containing the model parameters, and $\mathbf{y}$ is a $N_o \times 1$ vector of observations possibly at multiple time steps. All vectors are column vectors unless otherwise stated. The posterior PDF $p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y})$ encapsulates all the relevant information concerning the models state and parameter space. It can, for example, be used to compute the mean or median model state as well as measures of uncertainty such as the standard deviation or percentile ranges. It is important to note that I am not trying to find a best optimal estimate, instead I am trying to estimate an entire distribution. This differentiates Bayesian inference from optimization approaches such as least squares. Note that the evidence can be expressed as

$$p(\mathbf{y}) = \int_{\boldsymbol{\theta}} \int_{\mathbf{x}} p(\mathbf{x}, \mathbf{y}, \boldsymbol{\theta})\, d\mathbf{x}\, d\boldsymbol{\theta} = \int_{\boldsymbol{\theta}} \int_{\mathbf{x}} p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})p(\mathbf{x}, \boldsymbol{\theta})\, d\mathbf{x}\, d\boldsymbol{\theta}\,, \tag{2.72}$$

such that it can be viewed as a normalizing constant in (2.71), whereby I may write

$$p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})p(\mathbf{x}, \boldsymbol{\theta})\,, \tag{2.73}$$

where $\propto$ denotes 'proportional to'. This reads as follows: the posterior is proportional to the product of the prior and likelihood. The equation shows how Bayes theorem can be used as a means to update the uncertain state and parameter space of a model using uncertain observational data.

2.4.2 Strong constraint batch smoothing

In this thesis I am dealing with the strong constraint, i.e. a perfect model structure, data assimilation problem where I assume (for simplicity) that all the uncertainty is in the model parameters. This parameter uncertainty of course extends to the forcing which I suspect is typically the main source of uncertainty in snow modeling (Raleigh et al., 2015). Moreover, I am solving the batch smoothing problem where I estimate the parameters based on the entire trajectory of observations within a model integration (in my case one snow season). In petroleum reservoir geophysics, solving this kind of inverse problem is known as history matching (c.f. Evensen, 2018). Note that recently Evensen (2019) has presented weak constraint (with model error) approaches to solving the same problem. I mention history matching because one of the cornerstone algorithms in this thesis is borrowed from this community (Emerick and Reynolds, 2013).

I define the entire modeled time trajectory of states (SWE, fSCA, etc...) within a snow season as

$$\mathbf{x} = \mathcal{M}(\boldsymbol{\theta})\,, \tag{2.74}$$

where $\mathcal{M}$ is the model operator that performs the annual model integration as a function of the uncertain parameters $\boldsymbol{\theta}$. The model makes predictions that correspond to the observations (fSCA) at N_o observation times, I will denote these predicted observations as $\widehat{\mathbf{y}}$ and they are tied to the state and parameters through

$$\widehat{\mathbf{y}} = \mathcal{H}\left(\mathbf{x}\right) = \mathcal{H}\left(\mathcal{M}(\boldsymbol{\theta})\right)\,, \tag{2.75}$$

where $\mathcal{H}$ is the observation operator that picks out the relevant predicted observations from the state. By relevant, I mean the predicted observations that are coincident with the actual observations. The actual observations, $\mathbf{y}$, are related to the true state of the system, $\mathbf{x}_t$, through

$$\mathbf{y} = \mathcal{H}\left(\mathbf{x}_t\right) + \boldsymbol{\epsilon}\,, \tag{2.76}$$

where the residuals $\boldsymbol{\epsilon}$, are a realization of observation error that may also contain representation error (van Leeuwen, 2015). I assume that the residuals $\boldsymbol{\epsilon}$ are Gaussian distributed with zero mean and covariance $\mathbf{R}$. From (2.73) through (2.76), under the assumption of Gaussian additive error and *given* that $\mathbf{x}$ and $\boldsymbol{\theta}$ (and hence $\widehat{\mathbf{y}}$) are true, the likelihood is given by

$$p(\mathbf{y}|\mathbf{x},\boldsymbol{\theta}) = p(\mathbf{y}|\widehat{\mathbf{y}}) = p(\boldsymbol{\epsilon}) \propto \exp\left(-0.5\,(\mathbf{y}-\widehat{\mathbf{y}})^{\mathrm{T}}\,\mathbf{R}^{-1}\,(\mathbf{y}-\widehat{\mathbf{y}})\right) \tag{2.77}$$

where I have used that $\boldsymbol{\epsilon} = \mathbf{y} - \widehat{\mathbf{y}}$ conditional on $\widehat{\mathbf{y}}$ being true. The uknown truth can itself be thought of as a realization of a (possibly stochastic) process but not a random variable (van Leeuwen et al., 2019). The assumption of additive Gaussian noise is typically made in DA for convenience and because of the central limit theorem which basically states that the sum of independent random variables tends to a Gaussian distribution. Note that it is not always necessary to make this assumption, more complex likelihoods are possible (e.g. Schoups and Vrugt, 2010).

Now, if I expand the Bayesian update equation, i.e. (2.73), using the rules of conditional probability then

$$p(\mathbf{x},\boldsymbol{\theta}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})\,. \tag{2.78}$$

In my batch smoothing parameter estimation problem, I am interested in the marginal posterior PDF $p(\boldsymbol{\theta}|\mathbf{y})$. If I insert for the likelihood and using that in the strong constraint formulation the transition density $p(\mathbf{x}|\boldsymbol{\theta}) = \delta\left(\mathbf{x} - \mathcal{M}(\boldsymbol{\theta})\right)$ where $\delta(\cdot)$ is the Dirac delta function (Evensen, 2018), I can integrate over $\mathbf{x}$ to obtain the marginalized posterior

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\boldsymbol{\epsilon})p(\boldsymbol{\theta})\,, \tag{2.79}$$

where $p(\boldsymbol{\epsilon})$ is given by (2.77). Now (2.79) gives us a Bayesian updating rule for the uncertain parameters given the observations under the strong constraint assumption. Recall that what makes this a batch smoothing problem is that the likelihood contains all the available fSCA residuals during the snow season. Given that I am effectively inverting a probabilistic snow depletion curve, I only use observations in the ablation season as these are the observations that are likely to contain information on the uncertain parameters. The updated states are obtained by simply rerunning the model with the updated parameters. A benefit with this is that the states will be dynamically consistent which is not the case when filtering.

2.4.3 Gaussian anamorphosis

Now that we are equipped with an updating rule for the parameters we would of course like to apply it in practice. The ensemble smoother methods that I will use for this purpose are based on the Kalman filter and are as such optimally implemented when the prior parameter distribution is Gaussian (van Leeuwen and Evensen, 1996). This sounds like a severe restriction, given that many parameters may be physically bounded. For example, many variables such as precipitation physically can not be negative. It turns out that there exists a clever technique called Gaussian anamorphosis from geostatistics that I can use to alleviate this issue (Bertino et al., 2003).

The idea with this technique is to come up with a Gaussian anamorphosis transform such that, given a non-Gaussian distributed uncertain parameter $\boldsymbol{\theta}$, I arrive at an anamorphosed parameter $\widetilde{\boldsymbol{\theta}}$ through

$$\widetilde{\boldsymbol{\theta}} = \boldsymbol{\psi}(\boldsymbol{\theta}) \,, \tag{2.80}$$

which is normally (i.e. Gaussian) distributed. Here $\boldsymbol{\psi}(\cdot)$ is the Gaussian anamorphosis function. These analytical anamorphosis transform functions can be viewed as lenses through which the non-Gaussian parameters appear Gaussian. This kind of transformation is also known as a normal score transform. Each anamorphosis function has an associated inverse anamorphosis transform function through which the transformed Gaussian parameter again becomes non-Gaussian. I may express this as follows

$$\boldsymbol{\theta} = \boldsymbol{\psi}^{-1}\left(\widetilde{\boldsymbol{\theta}}\right) \tag{2.81}$$

where $\boldsymbol{\psi}^{-1}$ is the inverse anamorphosis function. It is possible to construct the anamorphosis function empirically (e.g. Simon and Bertino, 2009). I instead adopt a simpler approach by applying analytical anamorphosis functions. In this effort, I use the log transform for lower bounded parameters and the logit transform for double bounded parameters (i.e. bounded both from below and above). Effectively, this amounts to assuming that lower bounded parameters are log-normally distributed and that double bounded parameters are logit-normally distributed. Visual examples of logit-normal distributions and their associated normal distributions are provided in Figure 2.15

As presented in Aalstad et al. (2018), the logit transform for a parameter θ bounded

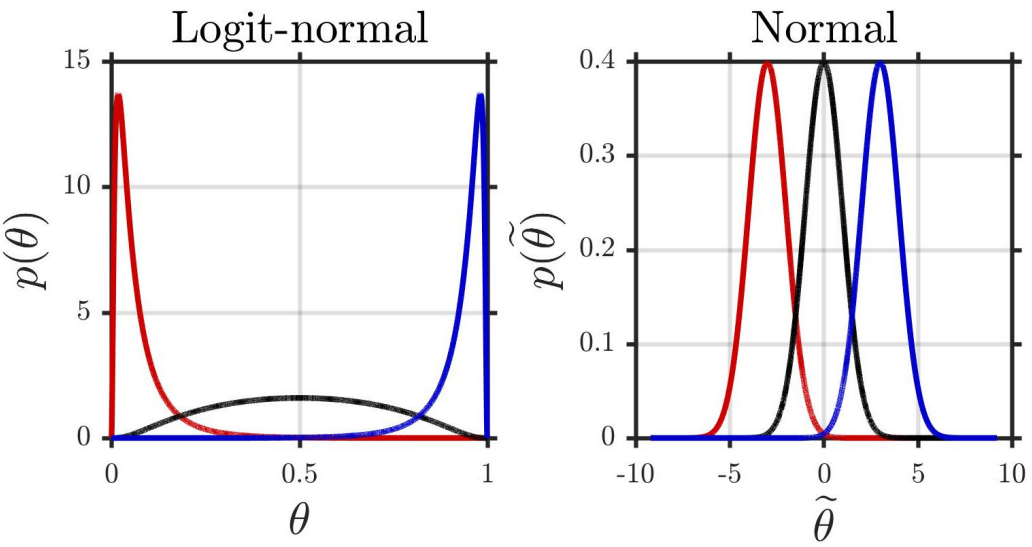


Figure 2.15: Example of how certain non-Gaussian parameters can be seen as Gaussian when properly transformed using an anamorphosis function. **Left**: Three different logit-normal (non-Gaussian) probability density functions. **Right**: The three associated anamorphosed normal (Gaussian) probability density functions in corresponding colors. Here $\widetilde{\theta} = \text{logit}(\theta)$.

between a and b is given by

$$\widetilde{\theta} = \text{logit}_{(a,b)}(\theta) = \ln\left(\frac{\theta - a}{b - a}\right) - \ln\left(1 - \frac{\theta - a}{b - a}\right) \tag{2.82}$$

whereas the inverse transform is given by

$$\theta = \text{logit}^{-1}_{(a,b)}(\widetilde{\theta}) = a + (b - a)/(1 + e^{-\widetilde{\theta}})\,, \tag{2.83}$$

The scaled natural log transform for a parameter that is lower bounded by a is given by

$$\widetilde{\theta} = \ln_{(a)}(\theta) = \ln(\theta - a)\,, \tag{2.84}$$

whereas the inverse transform is given by

$$\theta = \ln^{-1}_{(a)}(\widetilde{\theta}) = a + e^{\widetilde{\theta}}\,. \tag{2.85}$$

These are the two analytical anamorphosis functions (and their inverses) that I will use in this thesis, it is worth noting that there exists infinitely many more such transforms (e.g. Box and Cox, 1964). Later on, I will use these two transforma-

tions to generate samples from logit-normal and log-normal distributions. This is done as follows: first apply the transform to the desired mean of the distributions, then add realizations of normally distributed noise with the desired variance, and then apply the inverse transform. Once an ensemble (collection of samples) has been generated in this fashion, the transformed ensemble is generated by applying the relevant transformations to this ensemble.

Using the analytical anamorphosis functions, it is now possible to obtain samples of the transformed parameters, $\widetilde{\boldsymbol{\theta}}$, that are normally distributed. Note that I may now write the expression for the predicted observations, (2.75), in terms of the transformed parameters as

$$\widehat{\mathbf{y}} = \mathcal{H}\left(\mathbf{x}\right) = \mathcal{H}\left(\mathcal{M}(\boldsymbol{\theta})\right) = \mathcal{H}\left(\mathcal{M}(\boldsymbol{\psi}^{-1}(\widetilde{\boldsymbol{\theta}}))\right) , \tag{2.86}$$

such that it is possible to express the residual, $\boldsymbol{\epsilon} = \mathbf{y} - \widehat{\mathbf{y}}$, in terms of the transformed parameters. Thereby, I may write my update equation in terms of the transformed parameters as well

$$p(\widetilde{\boldsymbol{\theta}}|\mathbf{y}) \propto p(\boldsymbol{\epsilon})p(\widetilde{\boldsymbol{\theta}}) , \tag{2.87}$$

where $p(\widetilde{\boldsymbol{\theta}})$ is the Gaussian prior of the transformed parameters. The posterior for the (untransformed) parameters is simply obtained by taking the inverse transform of this transformed posterior.

2.4.4 Gaussian analysis

The transformed prior parameter distribution is a Gaussian of the form

$$p(\widetilde{\boldsymbol{\theta}}) \propto \exp\left(-0.5(\widetilde{\boldsymbol{\theta}} - \overline{\widetilde{\boldsymbol{\theta}}})^{\mathrm{T}}\mathcal{C}_{\widetilde{\boldsymbol{\theta}}\widetilde{\boldsymbol{\theta}}}^{-1}(\widetilde{\boldsymbol{\theta}} - \overline{\widetilde{\boldsymbol{\theta}}})\right) \tag{2.88}$$

where $\overline{\widetilde{\boldsymbol{\theta}}}$ is a vector containing the means of the transformed parameters whereas $\mathcal{C}_{\widetilde{\boldsymbol{\theta}}\widetilde{\boldsymbol{\theta}}}$ is the transformed parameter covariance matrix. Inserting (2.88) and (2.77) in (2.87) we have an expression for transformed posterior

$$p(\widetilde{\boldsymbol{\theta}}|\mathbf{y}) \propto \exp\left(-\frac{1}{2}\mathcal{J}\right) , \tag{2.89}$$

where the cost function

$$\mathcal{J} = (\mathbf{y} - \widehat{\mathbf{y}})^{\mathrm{T}} \mathbf{R}^{-1} (\mathbf{y} - \widehat{\mathbf{y}}) + (\widetilde{\boldsymbol{\theta}} - \overline{\widetilde{\boldsymbol{\theta}}})^{\mathrm{T}} \mathcal{C}_{\widetilde{\boldsymbol{\theta}}\widetilde{\boldsymbol{\theta}}}^{-1} (\widetilde{\boldsymbol{\theta}} - \overline{\widetilde{\boldsymbol{\theta}}}) \,. \tag{2.90}$$

Since the exponential is a monotonous function, minimizing this cost function is tantamount to maximizing the posterior, i.e. finding the maximum a posteriori (MAP) estimate. Since $\widehat{\mathbf{y}}$ is generally a non-linear function of the (transformed) parameters, finding this minimum, or more generally solving for the posterior, is quite challenging and can not be done analytically. Instead, we can resort to variational or ensemble-based data assimilation techniques (c.f. Carrassi et al., 2018).

At this point I will switch to a lower gear and consider a simple example where an analytical solution for the posterior actually exists. Consider the simple case where I only have one directly observed state variable, i.e.

$$y = x + \epsilon \,, \tag{2.91}$$

where ϵ is normally distributed. Then the cost function becomes (by inserting x for $\widetilde{\boldsymbol{\theta}}$) simply

$$\mathcal{J}(x) = \frac{(y - x)^2}{\sigma_o^2} + \frac{(x - \overline{x}^f)^2}{\sigma_f^2} \,. \tag{2.92}$$

where σ_o^2 and σ_f^2 are the observation and forecast error variances and $\overline{x}^f$ is the prior mean (the so-called 'forecast') for the state variable. Since the model is linear and both the prior and posterior are Gaussians, the posterior will also be a Gaussian. Recall that for Gaussians the mode and mean are the same. So the minimum of the cost function will be at the posterior mean. Letting $\overline{x}^a$ denote the mean of the posterior, also called the analysis, this implies that $\partial_x \mathcal{J}|_{x=\overline{x}^a} = 0$ whereby

$$0 = -\frac{(y - \overline{x}^a)}{\sigma_o^2} + \frac{(\overline{x}^a - \overline{x}^f)}{\sigma_f^2} \tag{2.93}$$

Solving for $\overline{x}^a$ yields

$$\overline{x}^a = \left(\frac{\sigma_f^2}{\sigma_f^2 + \sigma_o^2} \right) y + \left(\frac{\sigma_o^2}{\sigma_f^2 + \sigma_o^2} \right) \overline{x}^f \tag{2.94}$$

which can be written as

$$\overline{x}^a = \overline{x}^f + K \left(y - \overline{x}^f \right) \tag{2.95}$$

where

$$K = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_o^2} \tag{2.96}$$

is the (univariate, directly observed) Kalman gain which satisfies $K \in (0, 1)$ provided both error variances are non-zero. It can also be shown (c.f. Jazwinski, 1970) that the analysis (posterior) error variance becomes

$$\sigma_a^2 = (1 - K)\sigma_f^2 = K\sigma_o^2 \,. \tag{2.97}$$

This variance will be lower than both the observation and forecast error variances. The resulting posterior distribution will simply be $p(x|y) = \mathcal{N}(\overline{x}^a, \sigma_a)$ where $\mathcal{N}$ is the normal distribution. An example of this simple analysis is visualized in Figure 2.16. Although a simple example, this analysis demonstrates some of the key ideas behind the celebrated Kalman filter, in particular how to update a univariate state estimate with a single direct observation. In the 'filtering' aspect of the Kalman filter the analysis is performed sequentially in time as data becomes available, with the posterior of today becoming the prior of tomorrow.

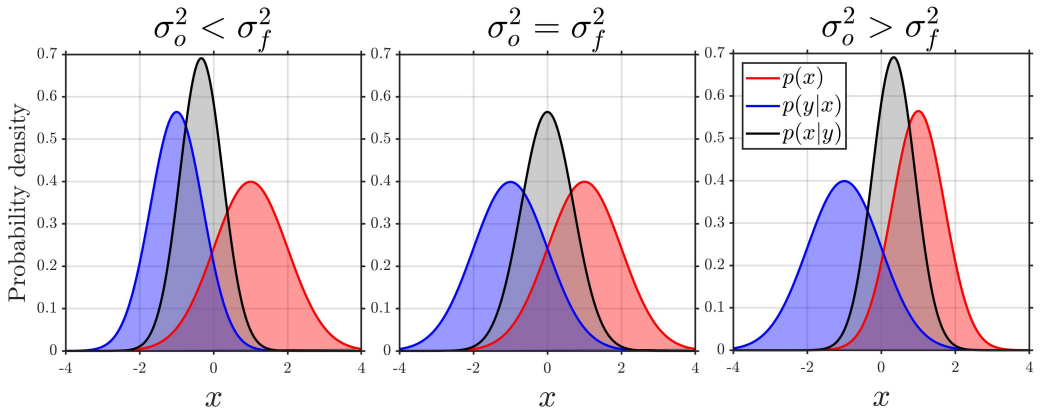


Figure 2.16: Simple analysis with $y = -1$ and $\overline{x}^f = 1$. The posterior ($p(x|y)$ in black) is closer to the prior ($p(x)$ in red) than the likelihood ($p(y|x)$ in blue) if the forecast variance (σ_f^2) is lower than the observation variance (σ_y^2) as in the right panel, and vice versa in the reverse case as in the left panel. The two variances are equal in the middle panel, hence the update merely constrains the spread. The posterior variance is always lower than that of both the prior and the likelihood.

2.4.5 Generating the prior ensemble

As we discovered in the last section, solving for the posterior is generally not possible analytically. In this thesis, I resort to ensemble-based data assimilation techniques as my means of solving for the Bayesian update problem in Equation (2.88). An advantage with ensemble-based techniques is that, unlike the variational techniques, they are able to estimate the full form of the posterior not just the MAP, such that they also provide robust uncertainty estimates (Carrassi et al., 2018). In addition, the ensemble-based techniques are considerably more computationally efficient than full MCMC inversion which typically requires tens of thousands of model integrations (Neal, 1993).

Ensemble-based DA techniques are based on Monte Carlo sampling of probability distributions. In the context of strong constraint batch smoothing for parameter estimation, the prior ensemble is a collection of parameter set combinations that are randomly drawn from specified prior probability distributions. This prior ensemble of perturbed parameters is used to drive multiple realizations of the simple snow model. The ensemble of snow model realizations is then used to update the prior parameter ensemble by comparing the ensemble of predicted observations to the actual observations of the fSCA depletion during the ablation season.

Essentially we could chose to perturb any uncertain parameter in the simple snow model. For simplicity and to constrain the parameter space, I limit myself to a few parameters that I deem to be the most important parameters in the model. These include parameters that are used to perturb the forcing. The model parameters that I perturb are: the peak subgrid coefficient of variation (χ) of the snow depletion curve, the minimum snow albedo (α_{min}), and the initial ground heat flux (Q_0). These parameters are all drawn from double bounded logit-normal prior probability distributions. In addition, I also update two forcing perturbed parameters: the precipitation bias b_P and the snowmelt bias b_M. Both of these are time-invariant parameters that are used to perturb the net accumulation rate through the mass balance forcing at each timestep n as follows (c.f. Equation 2.52)

$$\mathcal{A}_{n,j} = b_{P,j}\left(P_{S,n} + P_{R,n}\right) - b_{M,j}\left(M_n + E_n\right) \tag{2.98}$$

for the $j \in 1 : N_e$ ensemble members. These perturbed parameters are both drawn form lower-bounded log-normal prior probability distributions. All of the perturbed parameters are for simplicity assumed to be independently distributed

(i.e. uncorrelated). The form of the prior perturbed parameter distributions is as given in Table 4 of Aalstad et al. (2018). Therein, we consider $N_p = 5$ perturbed parameters. Note that the prior mean and variance of each of the parameters, as well as the correlation between the parameters, can be changed as desired based on prior knowledge or beliefs. The important point is that the priors used should always be *specified* and picked independently of the observations to avoid circularity. In the future, it would be worthwhile to explore the impact of adding yet more perturbed parameters to the ensemble-based snow reanalysis framework. It is also desirable to explore the performance of many different schemes. So far in my work, I have used the particle batch smoother, the ensemble smoother, and the ensemble smoother with multiple data assimilation (ES-MDA). In the next sections, I will review these three schemes.

2.4.6 Particle batch smoother

The particle batch smoother (PBS) was introduced in the context of snow reanalysis by Margulis et al. (2015). It is essentially a particle filter (Gordon et al., 1993) without re-sampling that uses the prior as the importance density (see van Leeuwen, 2009). The PBS is presented as the "direct ensemble method" in van Leeuwen and Evensen (1996). As noted by Aalstad et al. (2018), it is also the same as the generalized likelihood uncertainty estimation technique (GLUE; Beven and Binley, 1992, 2014), which is widely used in hydrology, with a formal Gaussian likelihood function. The PBS has since its inception been widely used for snow reanalyses (e.g. Margulis et al., 2016; Cortés and Margulis, 2017; Baldo and Margulis, 2018; Teweldebrhan et al., 2019; Fiddes et al., 2019).

Since the PBS is relatively straightforward to derive from the Bayesian update equation I will do so here. Note that this technique does not make any assumptions on the linearity of the model or the Gaussianity of the prior (though typically a Gaussian likelihood is assumed) so I will work with the untransformed parameters. Under the particle approximation, the prior PDF is given by a sum of Dirac-delta functions, i.e. for N_e particles (ensemble members) we have

$$p(\boldsymbol{\theta}) = \frac{1}{N_e} \sum_{j=1}^{N_e} \delta \left(\boldsymbol{\theta} - \boldsymbol{\theta}_j \right) \ . \tag{2.99}$$

This approximation has a straightforward interpretation: the prior essentially be-

comes a probability mass function with equal prior probability $1/N_e$ for each of the prior particles. Plugging this into the Bayesian update equation we have that

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta}) \sum_{j=1}^{N_e} \delta\left(\boldsymbol{\theta} - \boldsymbol{\theta}_j\right)}{\int_{\boldsymbol{\theta}} p(\mathbf{y}|\boldsymbol{\theta}) \sum_{k=1}^{N_e} \delta\left(\boldsymbol{\theta} - \boldsymbol{\theta}_k\right) \, d\boldsymbol{\theta}} \,, \tag{2.100}$$

which effectively becomes

$$p(\boldsymbol{\theta}|\mathbf{y}) = \sum_{j=1}^{N_e} \frac{p(\mathbf{y}|\boldsymbol{\theta}_j)}{\sum_{k=1}^{N_e} p(\mathbf{y}|\boldsymbol{\theta}_k)} \delta\left(\boldsymbol{\theta} - \boldsymbol{\theta}_j\right) = \sum_{j=1}^{N_e} w_j \delta\left(\boldsymbol{\theta} - \boldsymbol{\theta}_j\right) \,, \tag{2.101}$$

where the posterior weights w_j (probability masses) are simply given by a ratio of likelihoods (Aalstad et al., 2018)

$$w_j = \frac{\exp\left(-0.5 \left(\mathbf{y} - \widehat{\mathbf{Y}}_j\right)^{\mathrm{T}} \mathbf{R}^{-1} \left(\mathbf{y} - \widehat{\mathbf{Y}}_j\right)\right)}{\sum_{k=1}^{N_e} \exp\left(-0.5 \left(\mathbf{y} - \widehat{\mathbf{Y}}_k\right)^{\mathrm{T}} \mathbf{R}^{-1} \left(\mathbf{y} - \widehat{\mathbf{Y}}_k\right)\right)} \tag{2.102}$$

where $\widehat{\mathbf{Y}}_j$ are the predicted observations for each of the particles.

2.4.7 Ensemble smoother

What is now known as the ensemble smoother (ES), which is a particularly widely used scheme in the history matching community, was originally proposed by van Leeuwen and Evensen (1996). The starting point for the ES with anamorphosis is the expression for the transformed posterior in Equation 2.89 where a Gaussian transformed prior is assumed. The subsequent derivation of the ES is more involved than that of the PBS and so it will not be undertaken here. It largely follows the approach I took for the simple Gaussian analysis, but for a multidimensional case. As such, its derivation is equivalent to that of the Kalman filter, but with an ensemble representation of the covariance matrices and a batch smoother update. The analysis step for the Kalman filter is derived assuming a linear model as well as a Gaussian prior and likelihood in which case it provides the variance minimizing solution (i.e. best linear unbiased estimate). The reader is referred to page 197 in Jazwinski (1970) and Chapter 3 in Evensen (2009) for more details. The ES has been applied for snow reanalysis in Durand et al. (2008) and Girotto

et al. (2014b,a).

The anamorphosed ES analysis step is given by (c.f. Aalstad et al., 2018)

$$\widetilde{\boldsymbol{\Theta}}^{(a)} = \widetilde{\boldsymbol{\Theta}}^{(f)} + \mathcal{C}_{\widetilde{\boldsymbol{\Theta}}\widehat{\mathbf{Y}}} \left(\mathcal{C}_{\widehat{\mathbf{Y}}\widehat{\mathbf{Y}}} + \mathbf{R} \right)^{-1} \left(\mathbf{Y} - \widehat{\mathbf{Y}}^{(f)} \right) \tag{2.103}$$

where $\widetilde{\boldsymbol{\Theta}}^{(f)} = \left[\widetilde{\boldsymbol{\theta}}_1^{(f)} \ \dots \ \widetilde{\boldsymbol{\theta}}_j^{(f)} \ \dots \ \widetilde{\boldsymbol{\theta}}_{N_e}^{(f)} \right]$ is the $N_p \times N_e$ matrix containing the ensemble of prior (forecast) transformed perturbed parameters, while

$$\mathcal{C}_{\widetilde{\boldsymbol{\Theta}}\widehat{\mathbf{Y}}} = \frac{1}{N_e} \widetilde{\boldsymbol{\Theta}}^{(f)'} \widehat{\mathbf{Y}}^{(f)'\mathrm{T}}, \tag{2.104}$$

is a $N_p \times N_o$ (ensemble) transformed perturbed parameter-predicted observation error covariance matrix where primes ($'$) denote anomalies (i.e. deviations from the ensemble mean), and

$$\mathcal{C}_{\widehat{\mathbf{Y}}\widehat{\mathbf{Y}}} = \frac{1}{N_e} \widehat{\mathbf{Y}}^{(f)'} \widehat{\mathbf{Y}}^{(f)'\mathrm{T}}, \tag{2.105}$$

is the $N_o \times N_o$ (ensemble) predicted observation error covariance matrix, $\mathbf{R}$ is the observation error covariance matrix, $\mathbf{Y} = [\mathbf{y}_1 \ \dots \ \mathbf{y}_j \ \dots \ \mathbf{y}_{N_e}]$ is a $N_o \times N_e$ matrix of perturbed observations generated by independently adding realizations of Gaussian noise with covariance $\mathbf{R}$ to the observation vector, and the $N_o \times N_e$ matrix $\widehat{\mathbf{Y}}^{(f)} = \left[\widehat{\mathbf{y}}_1^{(f)} \ \dots \ \widehat{\mathbf{y}}_j^{(f)} \ \dots \ \widehat{\mathbf{y}}_{N_e}^{(f)} \right]$ contains the predicted prior (forecast) observations. The use of perturbed observations can be attributed to the work of Burgers et al. (1998). These perturbations make this a so-called stochastic version of the ES, it would also possible to implement deterministic or square root versions of the ES (c.f. Sakov and Oke, 2008).

Algorithmically, the ES is implemented as follows: first perform an ensemble model integration, then transform the prior parameters, then update these using the analysis step in Equation 2.103, then perform the inverse transform to obtain the updated parameters ($\boldsymbol{\theta}^{(a)}$) and rerun an ensemble integration with these parameters to obtain the updated model states. An advantage with the ES is that it does not require stopping the model time stepping, as in the EnKF, instead the model is merely rerun once the parameters are updated (Stordal and Elsheikh, 2015). This helps to ensure dynamical consistency. Effectively with the ES the DA is performed 'offline' being wrapped on the outside, as opposed to inside, the model

making implementation quite simple. Still, the ES is variance minimizing only in the case where the model is linear and error statistics are Gaussian. Gaussianity can be implemented in quite a flexible manner through the Gaussian anamorphosis approach that I previously outlined. The linear assumption, however, is more troublesome and can cause biases in the update (relative to the true Bayesian posterior) when the ES is used with non-linear models. In this sense, the ES can be viewed as a form of approximate Bayesian computation (ABC) (see e.g. Sadegh and Vrugt, 2013). The PBS is also a form of ABC since the posterior will always be biased by the prior sampling error, i.e. regions of high likelihood may not be sampled.

2.4.8 Ensemble smoother with multiple data assimilation

The ensemble-smoother with multiple data assimilation is an iterative version of the ES that was proposed by Emerick and Reynolds (2013), based on earlier work in Emerick and Reynolds (2012), in order to mitigate the impact of the linear assumption in the ES. The idea with the iterations is to perform multiple smaller analysis steps which is expected to yield a better approximation of the true Bayesian posterior than a large single step (see Evensen, 2018). The iterations are formulated in such a way that they mimic simulated annealing and yield the same result as the ES for a linear model (Stordal and Elsheikh, 2015). Effectively, the posterior is annealed through the likelihood by inflating the observation error covariance matrix at each iteration. Several studies have shown that the ES-MDA outperforms the ES in the case of a non-linear model (e.g. Emerick and Reynolds, 2013; Aalstad et al., 2018; Evensen, 2018). Being an iterative algorithm, the ES-MDA is however considerably more expensive than both the ES or PBS in that it requires multiple ensemble model integrations and analysis steps.

The analysis step for the ℓ-th iteration of the ES-MDA is as follows

$$\widetilde{\mathbf{\Theta}}^{(\ell+1)} = \widetilde{\mathbf{\Theta}}^{(\ell)} + \mathcal{C}^{(\ell)}_{\widetilde{\mathbf{\Theta}}\widehat{\mathbf{Y}}} \left(\mathcal{C}^{(\ell)}_{\widehat{\mathbf{Y}}\widehat{\mathbf{Y}}} + \alpha^{(\ell)} \mathbf{R} \right)^{-1} \left(\mathbf{Y}^{(\ell)} - \widehat{\mathbf{Y}}^{(\ell)} \right) , \tag{2.106}$$

where $\alpha^{(\ell)}$ is the observation error inflation coefficient and the covariances are now given by

$$\mathcal{C}^{(\ell)}_{\widetilde{\mathbf{\Theta}}\widehat{\mathbf{Y}}} = \frac{1}{N_e} \widetilde{\mathbf{\Theta}}^{(\ell)\prime} \widehat{\mathbf{Y}}^{(\ell)\prime \mathrm{T}} \text{ and } \mathcal{C}^{(\ell)}_{\widehat{\mathbf{Y}}\widehat{\mathbf{Y}}} = \frac{1}{N_e} \widehat{\mathbf{Y}}^{(\ell)\prime} \widehat{\mathbf{Y}}^{(\ell)\prime \mathrm{T}} , \tag{2.107}$$

and the perturbed observations are obtained through

$$\widehat{\mathbf{Y}}^{(\ell)} = \mathbf{y} \otimes \mathbf{1}^{\mathrm{T}} + \sqrt{\alpha^{(\ell)}}\mathbf{R}^{1/2}\boldsymbol{\epsilon}^{(\ell)} \tag{2.108}$$

where $\boldsymbol{\epsilon}$ is an $N_o \times N_e$ matrix containing zero mean Gaussian white noise with a variance of 1. By comparing (2.103) and (2.106) it is clear that, other than the iterations, the ES-MDA differs from the ES through the inclusion of the inflation coefficient $\alpha^{(\ell)}$. In my implementations of the ES-MDA, the inflation coefficient is a constant equal to N_a (the number of assimilation cycle iterations) which I set to 4 based on the sensitivity study in Aalstad et al. (2018). This ensures that the inverse inflation coefficients sum to one, i.e. $\sum_{\ell=1}^{N_a} 1/\alpha^{(\ell)} = 1$, as required for consistency in the ES-MDA (Emerick and Reynolds, 2013).

Algorithmically the ES-MDA is implemented as an iterative version of the ES. Starting with $\ell = 0$ a prior distribution of parameters is generated and an ensemble integration of the model is carried out. Next, the prior parameters are transformed and updated using the ES-MDA analysis step (2.106), after which the updated transformed parameters are inverse transformed and ℓ is increased by one. Subsequently, a new ensemble model integration is carried out. This process is repeated for a total of N_a analysis steps and $N_a + 1$ ensemble model integrations. The flow diagram in Figure 2.17 shows schematically how the ES-MDA is implemented and how it fits into the reanalysis framework. The advantage with the ES-MDA over the ES is that it lessens the impact of the linear model assumption, leading to a better approximation of the Bayesian posterior even for non-linear forward models. This is important because the simple snow model that I am employing is non-linear, for example the snow depletion curve is a non-linear function of several unknowns. As shown in Aalstad et al. (2018) and Aalstad et al. (2019b) this means that the ES-MDA typically outperforms the ES for snow reanalysis applications. In addition, the same studies have shown that the ES-MDA also outperforms the PBS which shows signs of degeneracy when more than just a couple of parameters in the reanalysis are deemed uncertain. The downside is the computational cost. The ES-MDA requires $N_a + 1$ (so 5 in my case) sequential ensemble model integrations compared to 2 for the ES and 1 for the PBS. I use the ES-MDA as the algorithm of choice due to its robustness, with the ES and PBS primarily used as benchmarks to gauge the performance of the ES-MDA.

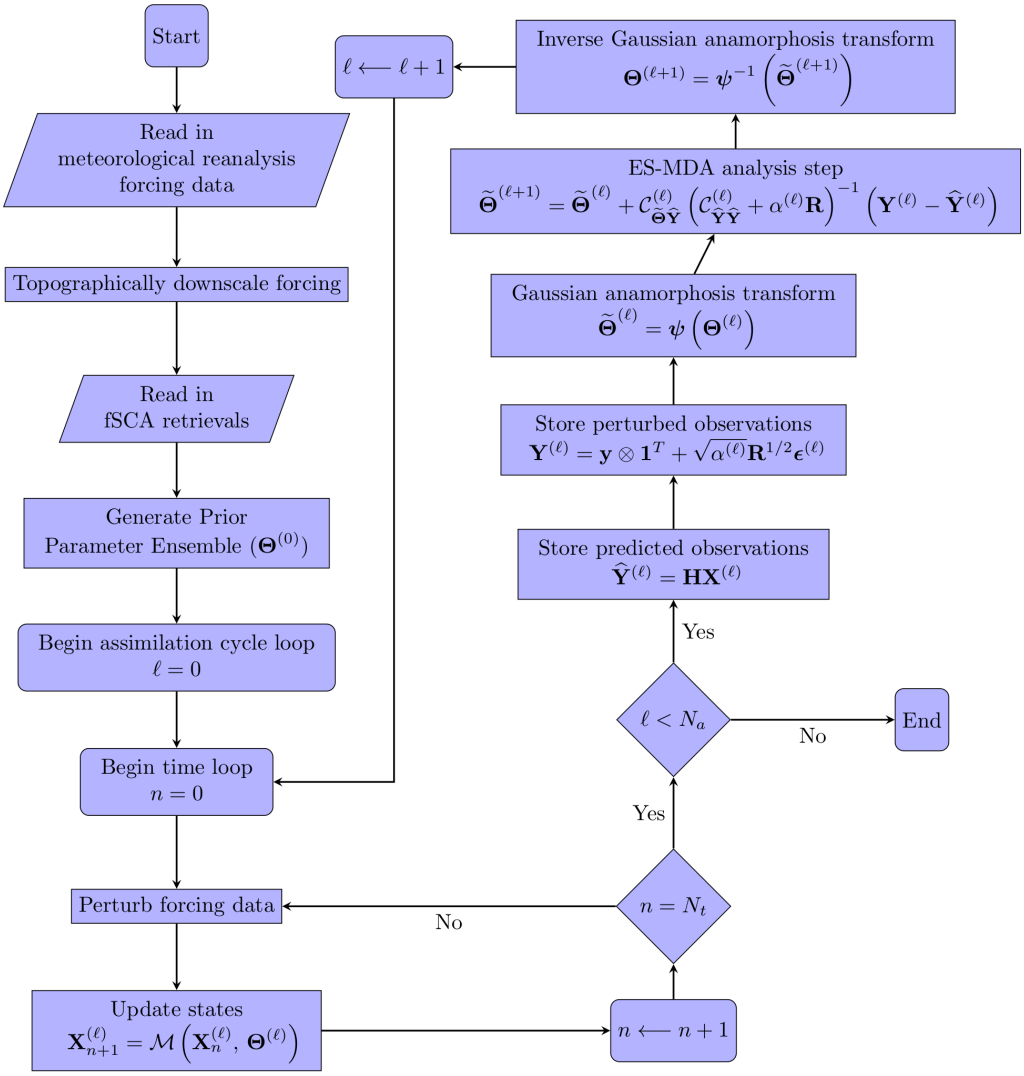


Figure 2.17: The reanalysis framework using the ES-MDA. Adapted from Aalstad et al. (2018).

Chapter 3

Summary of research

The results of my work are presented through four papers where we address the aim of the thesis and the five research objectives that were laid out in Section 1.4. All of these papers contribute to research objective 1, namely designing a modular ensemble-based snow reanalysis framework. This is again tied to the overarching aim of the thesis: trying to solve the SWE estimation problem (Dozier et al., 2016) retrospectively. As such, research objective 1 is also a superset of all the other research objectives.

Through the first paper, Aalstad et al. (2018), we build and apply a snow reanalysis framework to the task of estimating SWE at the kilometer scale at three sites on the Brøgger peninsula. This paper is the first step towards addressing research objective 1. It is in this paper that the simple snow model is first presented, which also ticks of research objective 3. In addition, this work addresses objective 5 through an extensive evaluation of the three different batch smoother data assimilation schemes. In the second paper, Aalstad et al. (2019a), we focus exclusively on research objective 2 by evaluating fSCA retrieval algorithms from different sensors in great detail. This study gives us an idea of what kind of errors to expect from different fSCA satellite retrievals. In the third paper, Aalstad et al. (2019b), we use higher resolution SU to retrieve fSCA at the hill-slope scale at both the Brøgger peninsula and the Mammoth Lakes Basin. The retrievals are subsequently assimilated into the reanalysis framework. In this study, all the research objectives are satisfied and we see the contours of an operational framework. In the final paper, Fiddes et al. (2019), the framework is employed with a clustering approach that enables scaling up the reanalysis to larger areas. In this chapter, the four papers that make up the results of this thesis are summarized.

3.1 Paper I

Summary

Paper I, Aalstad et al. (2018), is a proof of concept study where we construct an ensemble-based snow reanalysis framework and apply it in the estimation of the peak subgrid SWE distribution (SSD) at the kilometer scale at three sites on the Brøgger peninsula. We present the simple snow model by describing in detail its governing equations and tabulating model constants. The model is forced by topographically downscaled ERA-Interim reanalysis data. We then outline how ensemble integrations are carried out using prior distributions of perturbed parameters. Next, we describe the implementation of three different data assimilation schemes: the ES, the PBS, and the ES-MDA. These are all used to update the prior ensemble through the assimilation of fSCA retrievals.

For the most part, we assimilate MODIS fSCA into the simple snow model. In the last of the studied snow seasons, 2016, we also retrieve fSCA from the recently launched Sentinel-2A satellite and assimilate these. To the best of my knowledge, this is the first study to assimilate fSCA retrieved from the Sentinel-2 constellation. Prior and posterior SSD estimates obtained from the respective DA schemes are compared to independent observations of the SSD obtained from extensive snow surveys conducted near peak SWE at the three study sites: Bayelva, the Steinflåen plateau, and Kvadehuksletta. Through this comparison, we find that the ES-MDA performs most robustly, outperforming or at least nearly matching the performance of the other schemes across all the evaluation metrics that were considered.

Figure 3.1: Kvadehuksletta as seen from the Steinflåen plateau in the north western part of the Brøgger peninsula in the high-Arctic archipelago of Svalbard. Photo: K. Aalstad (10.05.2016).

3.2 Paper II

Summary

Paper II, Aalstad et al. (2019a), is a study in which we evaluate our current capabilities of retrieving fSCA in the Bayelva catchment from polar-orbiting satellites. Hundreds of time-lapse photographs obtained from an automatic camera system mounted on Scheteligfjellet (719 m a.s.l.), overlooking the catchment, during 6 ablation seasons are georeferenced and orthorectified. These orthophotos are subsequently classified into high-resolution (0.5 m) binary snow-cover maps using manually selected image specific thresholds. An example of an orthophoto and the associated binary snow-cover map is provided in Figure 3.2. The binary snow-cover maps are then aggregated to obtain ground-truth fSCA at 100 m and 500 m spatial resolution that we used to evaluate coincident satellite retrieved fSCA.

Satellite retrievals of fSCA are obtained from several sensors: MODIS on-board the Terra and Aqua satellites, MSI on-board the Sentinel-2 satellites, and OLI on-board the Landsat 8 satellite. The retrievals are obtained using three different algorithms: thresholding of the normalized difference snow-index (NDSI) and subsequent aggregation, regression on the NDSI, and fully constrained linear spectral unmixing (SU). These retrievals were then compared to the ground truth using a variety of fractional and binary evaluation metrics. For MODIS, the coarser (500 m) off-nadir looking sensor, we find that retrievals based on regressing the NDSI outperform SU. Overall, the best fSCA retrievals were obtained from the higher resolution near-nadir looking sensors, MSI and OLI, using SU. Such retrievals need to become operational.

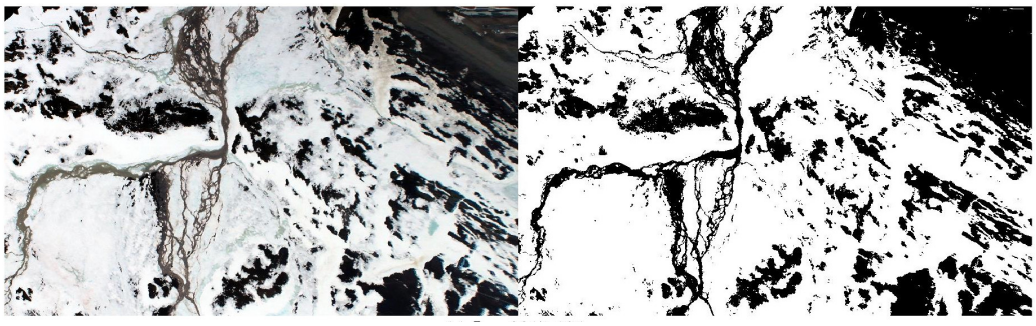


Figure 3.2: **Left**: Orthorectified true color time-lapse photograph over the Bayelva catchment on the 14.06.2017. **Right**: Resulting reference snow-cover map (black=bare, white=snow).

3.3 Paper III

Summary

In paper III, Aalstad et al. (2019b), the ensemble-based snow reanalysis framework is applied at the hillslope-scale (100 m) for two disparate study areas, the Brøgger peninsula (165 km^2) and the Mammoth Lakes Basin (29 km^2), for the 2016, 2017, and 2018 snow seasons. A view over the Mammoth Lakes Basin is provided in Figure 3.3. We use snow surveys and the ASO LIDAR for validation at Brøgger and Mammoth, respectively. TopoSCALE is used to downscale meteorological reanalysis data from ERA5 and NLDAS-2 to force the simple snow model. We assimilate dense time stacks of hillslope-scale fSCA retrieved from Landsat 7&8, Sentinel-2, PlanetScope, and RapidEye.

To the best of my knowledge, this is the first study to assimilate such a dense time stack of hillslope-scale fSCA retrievals. We demonstrate once again that the ES-MDA systematically outperforms the PBS and the ES also at Mammoth. In addition, we show how the use of multiple sensors leads to an improved reanalysis compared to using a single sensor. Finally, at Brøgger we showcase how the higher resolution satellite retrievals can be used to further downscale the reanalysis to the pixel-scale using deterministic reconstruction.

Figure 3.3: Mammoth Lakes Basin as seen from near the top of Mammoth Mountain (3371 m a.s.l.) in the Californian Sierra Nevada. Photo: F. Schneider (14.09.2018).

3.4 Paper IV

Summary

In paper IV, Fiddes et al. (2019), the ensemble-based snow reanalysis framework outlined in Aalstad et al. (2018) is coupled to the approach for large-area land surface simulations in heterogeneous terrain presented by Fiddes et al. (2015). This approach uses TopoSCALE (Fiddes and Gruber, 2014) to downscale ERA5 reanalysis forcing, TopoSUB (Fiddes and Gruber, 2012) as a clustering algorithm to speed up simulations, the GEOtop land surface model (Endrizzi et al., 2014), and the PBS (Margulis et al., 2015) to assimilate MODIS fSCA retrievals. Simulations are performed at hyper-resolution (30 m) for nine $0.25°$ ERA5 grid boxes in the Swiss Alps (see Figure 3.4). The novelty of this work is the combined use of clustering and snow data assimilation. The clustering technique uses k-means clustering to represent the subgrid heterogeneity of the terrain. As such, instead of simulating each 30 m DEM pixel within an ERA5 grid box explicitly, we simulate k clusters of DEM pixels. Effectively, each DEM pixel is associated to a cluster with similar terrain characteristics (e.g. slope, aspect, and elevation). In this way, the number of GEOtop simulations required to represent an ERA5 grid box is reduced by several orders of magnitude from around 10^6 DEM pixels to 10^2 clusters. This allows us to afford ensemble simulation and data assimilation even when performing simulations across large areas. Validation against independent observations reveals marked improvements through the PBS data assimilation.

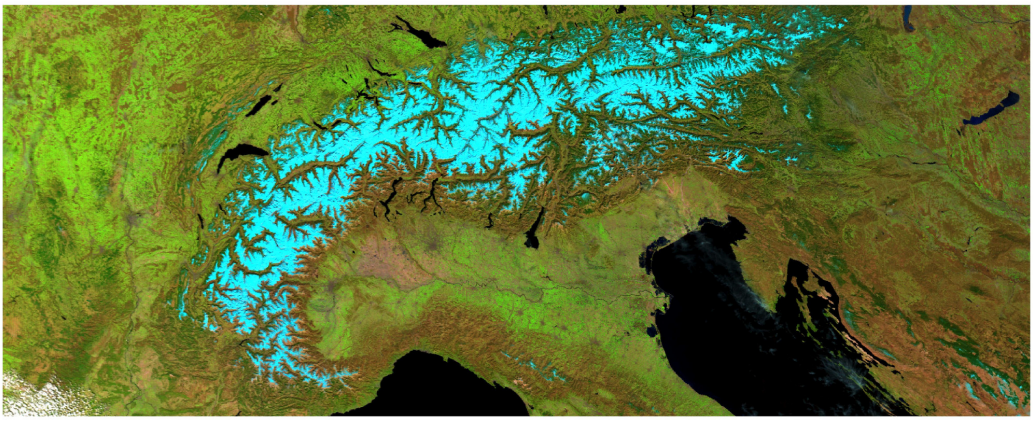


Figure 3.4: False color image of the European Alps taken by Terra MODIS on the 30.03.2019.

Chapter 4

Discussion

Through the papers, we have arrived at a modular ensemble-based snow reanalysis framework that can be applied almost anywhere on Earth. These papers include an extensive discussion of the results which addresses the limitations of the methodology and envisages possible extensions through an outlook. As noted in Aalstad et al. (2019b), the framework is still very much in the research and development stage and is not yet ready to become operational. Several improvements to the respective components of the framework are readily apparent and should be undertaken in due course. Here, I continue the discussion, with a focus on the limitations of each of the components of the framework and new avenues that may be explored in the future.

4.1 Forcing

Our reanalysis framework currently leans heavily upon the use of the TopoSCALE routine to downscale coarse-scale meteorological reanalysis data to the higher resolution model scale (100 m - 1 km). This is a relatively robust and computationally efficient topographic downscaling scheme that has been shown to perform well in a variety of environments (e.g. Fiddes and Gruber, 2014; Fiddes et al., 2019). A particular advantage of TopoSCALE over other downscaling routines that have been employed for reconstruction (c.f. Girotto et al., 2014b; Rittger et al., 2016) is that TopoSCALE does not apply fixed lapse rates (vertical gradients), but instead interpolates between the pressure (vertical) levels in the meteorological reanalysis. This is likely to be much more accurate than assigning static lapse rates for temperature for example in that it is possible to capture meteorological phenom-

ena such as inversions which are likely to be present in the vertical structure of the reanalysis data.

The performance of the scheme nonetheless varies considerably, with some variables being easier to downscale than others. This can be appreciated in Figure 2.3 where the errors in the different fields vary from negligible (e.g. air pressure and temperature) to considerable (wind speed). In that particular case I could explain the discrepancy in the wind speed by the fact that currently TopoSCALE does not account for boundary layer or terrain effects on the windfield. Accounting for these effects would probably make the errors considerably less severe, but it would require knowledge of the dynamic atmospheric stability and is thus best implemented as part of a snow model rather than a downscaling routine.

The computational frugality of the TopoSCALE approach clearly comes at a cost of reduced accuracy. In particular, it is optimistic to hope that all of the relatively simple empirically based relationships, which are calibrated to certain locations, used in the routine readily extend to arbitrary locations. Note that this is not a problem that plagues TopoSCALE alone, indeed it is a problem with many other empirically based topographic downscaling techniques such as those presented in Havens et al. (2017) and Karger et al. (2017). Currently, the main empirical steps in TopoSCALE involve diagnosing the clear sky emissivity (Konzelmann et al., 1994) and the partitioning of shortwave radiation into its diffuse and direct components (Ruiz-Arias et al., 2010a). Here, many other choices of empirical functions are available (e.g. Girotto et al., 2014b; Rittger et al., 2016) and it would be instructive to perform an extensive validation of the possible approaches.

As mentioned, the treatment of wind in TopoSCALE is also currently rather simplistic, being based purely on interpolation. It could be advantageous to also take into account terrain effects which is possible using relatively straightforward approaches outlined in Liston and Elder (2006). In addition, it would be worthwhile to account for boundary and surface layer effects by applying flux-profile relationships in the vertical interpolation of the wind field (see Foken, 2008). Precipitation is also treated in a relatively simple manner in TopoSCALE. As mentioned, it would be desirable to apply a correction for the effects of orography on precipitation following the theory in Smith et al. (2017) that was applied in Schuler et al. (2008) and Østby et al. (2017).

A more fundamental limitation with TopoSCALE is that it is not a dynamical downscaling routine. This makes it difficult to capture the typically non-linear gradients that tend to exist in the real four dimensional atmosphere. For simplicity in TopoSCALE these gradients are just approximated through linear interpolation between nodes in the 'parent' atmospheric reanalysis that is to be downscaled. The reason that a full blown regional atmospheric model is not employed to dynamically downscale reanalysis data, as in e.g. Rasmussen et al. (2011), is that it is currently prohibitively expensive for multi-decadal reanalyses, at least with my resources. Hybrid approaches such as the Intermediate Complexity Atmospheric Research (ICAR) model (Gutmann et al., 2016) present promising avenues for quasi-dynamical downscaling.

4.2 Snow modeling

In our work, particularly Aalstad et al. (2018) and Aalstad et al. (2019b), we have employed a custom built parsimonious snow model, namely the simple snow model (SSM). The model was specifically designed to enable runs with a large ensemble. This model solves a simplified mass and energy balance while trying to capture what are typically the most important physical processes involved in the accumulation and ablation of the seasonal snowpack. As such, all of the fluxes of mass and energy entering the snowpack are represented in some manner in the model.

Many of the internal snowpack processes such as heat conduction and compaction are not considered in the model. Properly accounting for these processes would require the use of a multi-layer snow model (e.g. Essery, 2015; Vionnet et al., 2012; Bartelt and Lehning, 2002). The model also currently does not consider the spatially variable effects of light absorbing impurities on the snow albedo which may constitute an important contribution to the radiative forcing (Skiles et al., 2018). Wind redistribution of snow is also only treated implicitly at the subgrid level through the probabilistic snow depletion curve (Liston, 2004).

Despite its simplicity, I have shown (see Figure 2.5) that the SSM is able to match the performance of more complex models such as the FSM of Essery (2015). It is worth noting that this comparison was only conducted for a single site, namely Col de Porte in France (Lejeune et al., 2019). As such, the similar performance

may just come down to the ignored processes not being important at that particular site. A more comprehensive comparative validation is required before I can claim that SSM outperforms FSM in general. In fact, I expect FSM to typically outperform SSM since, as a layered model, it resolves more physical processes that are known to be quite important in several environments.

In the future, I hope to borrow parametrizations and routines from the more complex snow models when upgrading SSM to a model of moderate complexity. It will be important to keep the vectorized nature of the SSM code which is what permits the use of a large parallel ensemble run on single cores. As such, I would stay away from routines that require a varying number of iterations or adaptive time stepping. Nonetheless, some of the properties in moderate complexity models such as FSM or iSnobal (Marks et al., 1999), particularly the use of a few layers, would be relatively easy to implement in the SSM. This would allow us to consider the cold content and compaction which would be important especially when coupling SSM to a model for the underlying ground (e.g. Westermann et al., 2016).

A more detailed treatment of the canopy would also be necessary when applying the model to densely forested areas where currently unaccounted for processes such as interception and canopy sublimation are important. Currently, the effects of the canopy are only considered through TopoSCALE where the canopy merely modifies the radiative fluxes following the approach outlined in Garren and Marks (2005) and Link and Marks (1999). To be able to potentially assimilate different satellite retrievals, it would also be important to include these as state variables. This would require a proper forwards operator for these variables. An example is the snow grain size which is currently a by-product of the SU retrieval algorithm. Snow grain size could for example be modeled using the approach in the Crocus model (Vionnet et al., 2012) and serve as an additional constraint in the ensemble-based data assimilation. Considering passive or active microwave remote sensing as an additional constraint would require coupling SSM to a microwave emissions model (e.g. Proksch et al., 2015).

4.3 Retrieving fSCA

Through the extensive evaluation study conducted in Aalstad et al. (2019a), we have probed our current capabilities of retrieving fSCA using polar orbiting optical satellite sensors. On the one hand, for the moderate resolution off-nadir looking MODIS instrument, we showed that SU performed relatively poorly partly due to directional effects that were not accounted for by the snow endmembers. On the other hand, for the higher resolution near-nadir looking instruments OLI (on Landsat 8) and MSI (on Sentinel-2), these directional effects were negligible and SU was by a long way the best performing retrieval algorithm. Importantly, we highlighted that simple thresholding of the NDSI and subsequent aggregation leads to a positive bias in fSCA retrievals from higher resolution sensors. This has important implications given that these kinds of retrievals that are typically used to validate coarser resolution fSCA retrievals such as those from MODIS over larger areas. It is, of course, highly undesirable to have biased validation data. Note in particular that biases, as opposed to random errors, do not disappear upon spatial aggregation so this effect is still present at the scale of a MODIS pixel (500 m).

The results from Aalstad et al. (2019a) are based on the use of high quality validation data in the form of snow-cover maps obtained from time-lapse photography taken by an automatic camera system installed on a mountain overlooking the Bayelva catchment in Svalbard. Despite the data quality, the area of interest where this data was available was quite small (order 2 km^2) and exhibits relatively simple topography with gently undulating unforested terrain and no steep slopes (Aalstad et al., 2018). As such, the error estimates that we obtain are relatively conservative and would likely be higher in the presence of steep slopes and forests. Indeed, through a comparison of spectrally unmixed fSCA to reference ASO retrieved fSCA in Aalstad et al. (2019b) we found slightly higher errors in the complex terrain of the Mammoth Lakes Basin. The conservative nature of the error estimates implies that the typical use of thresholding and aggregation to generate validation data from higher resolution sensors is quite error-prone.

The obvious way to deal with the directional effects in SU, that turn out to be important for off-nadir looking sensors (Cortés et al., 2014), would be to include them in the radiative transfer model used to generate the snow endmembers. SNICAR (Flanner et al., 2007; He et al., 2018) generates snow spectral albedo, i.e. bihemspheric reflectance, as a function of solar zenith angle and snow grain size.

It does not produce the hemispheric directional reflectance factors that are actually measured by a satellite, which are also a function of the satellite and solar azimuth (Schaepman-Strub et al., 2006). This is what we mean by directional (strictly azimuthal) effects. In particular, since it is a strongly forwards-scattering medium, when snow is viewed into the direction of illumination (i.e. towards the forwards scattering direction) the reflectance can easily exceed unity. The case is the converse in the backwards-scattering direction. This forwards-scattering is also why the reflectance increases with illumination angle (Dozier, 1989). To be able to account for these directional effects, however, much more complex and computationally demanding radiative transfer modeling is required (c.f. Schaepman-Strub et al., 2006).

For simplicity, our SU approach did not account for the presence of light absorbing impurities. This can be easily amended since SNICAR can accommodate such impurities (c.f. He et al., 2018), but it would increase the free-parameter space by one dimension. Thus, the computational cost of the unmixing would increase considerably by effectively adding another dimension to the 'look up table' of generated snow endmember spectra. It is not likely that this effect plays a major role in the retrieval of fSCA. On the other hand, if we were more interested in the accurate retrieval of snow grain size and albedo as additional constraints then the concentration of impurities would likely be more important (c.f. Painter et al., 2009, 2012).

A major caveat in the fSCA retrieval process is clouds. As noted in Aalstad et al. (2019b), the problem with clouds is twofold; they both obscure the land surface and have quite a similar spectral signature to snow in many bands. As such, not only do clouds result in data gaps, but snow can be falsely labeled as cloud (or vice versa). The way to deal with clouds is to accurately mask them out. With a sensor such as MODIS, which has quite a high number of spectral bands, it has proven to be possible to generate relatively accurate cloud masks even in snow-covered areas (Riggs et al., 2017). For sensors such as Landsat 8 OLI and Sentinel-2 MSI with fewer spectral bands, the cloud masking exercise is considerably more difficult, especially in snow-covered regions. In Aalstad et al. (2019a), we demonstrated how the native cloud masks from both of these sensors (c.f Zhu and Woodcock, 2012; Richter et al., 2012) contained considerable cloud commission errors. In our work, we have thus abandoned the native cloud masks from these sensors and resulted to manual scene selection. This of course leads to a

considerable data loss since partly cloudy scenes are not employed. Manual scene selection is also time consuming if the reanalysis framework is to be scaled up to larger areas over larger time slices. Fortunately, it seems that there has recently been progress made in addressing the cloud-masking issue over snow-covered areas (Musial et al., 2014; Gascoin et al., 2019).

There exists many available optical satellite sensors that we did not employ in our work. These are all worthwhile considering in future reanalyses. For the current era, at the moderate resolution end, we could have also considered retrieving fSCA from the AVHRR sensor (Hüsler et al., 2014), the VIIRS sensor (Riggs et al., 2017), the PROBA-V Vegetation instrument, and the Sentinel-3 OLCI sensor to name but a few examples. At the higher resolution end, we could have also retrieved fSCA from the SPOT and WorldView satellites, although much of this data is unfortunately not freely available even for scientific use. If we went back to the more nascent stages of the satellite-era we could have leveraged the earlier Landsat satellites (4 and 5) as has been done by Margulis et al. (2016) and Cortés and Margulis (2017). In the future, I look forwards to working with new platforms such as the OLI-2 sensor that will be mounted on Landsat 9 as well as the EnMAP hyperspectral satellite mission which may yield improved spectral unmixing.

4.4 Data assimilation schemes

In Aalstad et al. (2018) and Aalstad et al. (2019b), we have shown that the ES-MDA scheme systematically outperforms or at least nearly matches the performance of both the PBS and the ES with respect to various evaluation metrics. The fact that the ES-MDA outperforms the ES is perhaps not surprising, given that the former is an iterative version of the latter which can better handle model non-linearity (c.f. Emerick and Reynolds, 2013; Stordal and Elsheikh, 2015; Evensen, 2018). Both the ES and the ES-MDA are optimal, i.e. variance minimizing, for Gaussian linear models. We have shown how the Gaussian prior assumption does not have to be a major restriction provided that Gaussian anamorphosis is used to accommodate for bounded parameters (Bertino et al., 2003; Aalstad et al., 2018). The ES-MDA performs better than the ES for non-linear models because it performs iterations of smaller analysis steps as opposed to one abrupt analysis step. This is akin to the idea behind simulated annealing in optimization (see Stordal and Elsheikh, 2015). One problem that we noted with the implementation of the

ES-MDA in Aalstad et al. (2018) was that it tended to underestimate uncertainty; i.e. the ensemble standard deviation was typically lower than the actual RMSE. This is a well known issue with the stochastic (perturbed observation) ES-MDA that could be alleviated by employing the recently proposed deterministic ES-MDA (Emerick, 2018).

As for the PBS, we noted in both Aalstad et al. (2018) and Aalstad et al. (2019b) how this scheme is prone to degeneracy (ensemble collapse) even in the moderately dimensioned parameter space that we were addressing. This is a well known downside with importance sampling, which is essentially what the PBS amounts to, that is well documented in the literature (van Leeuwen and Evensen, 1996; Evensen, 2009; van Leeuwen, 2009; Morzfeld et al., 2018). I was nonetheless a bit surprised to see the PBS become degenerate even in this 5-dimensional parameter space. It is worth emphasizing that it is the often criticized Gaussian assumption in ensemble Kalman-based schemes, such as the ES-MDA, that makes these robust also in large dimensional spaces where importance sampling fails. As I have noted, the PBS is also quite sensitive to Monte Carlo error and the results may vary considerably from one ensemble integration to the next. Despite the problems with degeneracy and Monte Carlo sampling the PBS has still proven itself to be a useful scheme for ensemble-based snow reanalysis, as the many successful implementations of the scheme can testify to (e.g. Margulis et al., 2015; Cortés and Margulis, 2017; Fiddes et al., 2019).

The advantages with the PBS, and importance sampling in general, is that it does not require the Gaussian linear assumption and will thus tend to work well for non-linear models with non-Gaussian error statistics provided that the parameter space is small enough. In Aalstad et al. (2019a) we noted, for example, how fSCA retrieval errors tended to be heteroscedastic and increase for intermediate fSCA. This would encourage the use of a more complex likelihood such as that presented in Schoups and Vrugt (2010) rather than the Gaussian. Such a likelihood could be readily implemented into a PBS-like scheme. Another advantage with the PBS is that it only requires a single ensemble model integration. As an alternative to circumvent the degeneracy problem, it could be worth considering iterative importance sampling algorithms, i.e. an iterative PBS, such as those presented in Morzfeld et al. (2018). This would hopefully lead to performance gains and would also allow for a more fair comparison with the ES-MDA which currently uses considerably more computational resources through its iterations.

In our work, we have mainly used the PBS and ES to benchmark the performance of the ES-MDA. Given that both the PBS and ES both have problems associated with them, it would perhaps have been more instructive to use a more robust algorithm for benchmarking. In other parameter estimation studies (e.g. Stordal and Elsheikh, 2015) much costlier but very accurate MCMC simulation is typically used to generate the benchmark. For this purpose, I could, for example, employ the robust differential evolution adaptive metropolis (DREAM; Vrugt et al., 2009) algorithm. Arguably, an entire study could be devoted to identifying the most robust ensemble-based snow reanalysis scheme by using DREAM as a benchmark along with independent validation data. The ideal scheme would be the one that best balances accuracy (in terms of estimating the posterior) and computational cost. Computational cost (per grid cell) is an especially important criterion when considering that the overall aim of snow reanalyses should not only be to attain high accuracy, but also to become global while maintaining hillslope-scale resolution. It is worth noting that currently both the model and the data assimilation are executed on a per-grid cell basis, such that the entire framework is embarrassingly parallel and easily scaled up using high performance computing.

4.5 Emerging constraints

As is typical in the snow reanalysis exercise we have restricted ourselves to using retrievals of fSCA from optical sensors to constrain the ensemble. There are good reasons for this, namely that it is primarily the optical sensors that currently offer the potential to constrain the evolution of the snowpack at the hillslope-scale. At the same time, these sensors can offer other retrievals that could be worth considering in a reanalysis of the seasonal snowpack. In addition to fSCA, it is also possible to retrieve snow grain size, snow wetness, and albedo (c.f. Dozier et al., 2009). All of these could serve as additional constraints in the reanalysis.

As mentioned, there exists several optical sensors that are currently in polar orbit, many of which we have not yet employed to retrieve fSCA or other snow variables. At the hillslope-scale, the following polar orbiting optical sensors (and possibly more) are currently available: Landsat 7 ETM+, Landsat 8 OLI, Sentinel-2 MSI, PlanetScope, RapidEye, WorldView, SPOT5-7, and Pléiades. Landsat 9 OLI-2 and EnMAP will be launched in the near future as well. Together, these

satellites enable the retrieval of fSCA at hillslope-scale resolution at daily time steps, weather permitting. Unfortunately, data from many of these satellites (all but the Landsat and Sentinel-2 constellations) is currently not freely available in unlimited volumes. This, along with the lack of high quality cloud masks, currently precludes the use of these satellites in the generation of a global high-resolution snow reanalysis. Hopefully this will change in the future. At a slightly coarser scale (order 500 m), several sensors are also available including: MODIS, AVHRR, VIIRS, Sentinel-3, and PROBA-V. Despite the coarser resolution, these sensors typically feature more spectral bands than their hillslope-scale counterparts, which can make cloud masks more reliable. This is advantageous for operational reanalyses. In addition, these coarser sensors typically feature a wider swath which results in near daily repeat frequencies for each sensor. In Aalstad et al. (2019b), we have shown that it is advantageous to make use of as many sensors as possible in ensemble-based snow reanalysis. Effectively, the better the retrieval of the depletion of fSCA, both in terms of the accuracy of the individual retrievals and the temporal density, the better the reanalysis.

Other types of sensors could also serve as valuable constraints on the state of the snowpack at different scales. For example, it is well known that wet snow status (i.e. whether or not the snowpack is ripe) can be retrieved at the hillslope-scale from synthetic-aperture radar (SAR) sensors independently of cloud cover (Baghdadi et al., 1996). Sentinel-1 and TerraSAR-X are both satellites that could be readily employed in the retrieval of wet snow status. Knowing whether or not the snow is ripe and melting would be invaluable in a reanalysis. In addition, this kind of retrieval would be useful for filling gaps in fSCA time series that occur due to clouds. For example, if the snow-cover disappears over an extended cloudy period, a wet snow SAR retrieval would be useful in determining the earliest possible disappearance of the snowpack. Recent work also suggests that SAR can even be used to accurately estimate snow depth (Moller et al., 2017).

As for passive microwave retrievals of SWE, despite saturation problems for wet and deep snow and issues in complex terrain (Foster et al., 2005), these too could serve as valuable coarse-scale ($0.25°$) constraints. In their synthetic experiment, De Lannoy et al. (2010) have already demonstrated how coarse-scale passive microwave SWE can be assimilated into high-resolution models. Furthermore, De Lannoy et al. (2012) have shown that real coarse-scale passive microwave retrievals from AMSR-E can be assimilated jointly with MODIS fSCA into a higher

resolution snow model. Using passive microwave SWE retrievals as coarse-scale constraints would be particularly appealing over flat tundra areas where the retrievals are more likely to be accurate. The reason for this is that in such environments the limiting factors laid out in Foster et al. (2005) are less prone to manifest themselves. In the assimilation of passive microwave SWE, I would probably end up employing the multiple coarse-scale observation approach presented and evaluated in De Lannoy et al. (2010).

In the same vein, gravimetric retrievals of terrestrial water storage anomalies (TWSA) could also serve as a valuable constraint on the state of the snowpack. Such retrievals are available from two sets of twin satellites: GRACE (2002-2017) and GRACE-FO (2018-present) both at around 1° spatial resolution. As such, TWSA retrievals would also serve as coarse-scale constraints. Usage of TWSA retrievals would also require knowledge of the other components of the terrestrial water storage such as ground water, soil moisture, surface water, and ice (Yılmaz et al., 2019). A decomposition of the TWSA signal into its constituent parts so as to retrieve SWE would likely require a full land surface model that can represent all the components. This kind of retrieval approach is possible, as was shown in the study of Niu et al. (2007). Moreover, Su et al. (2010) have demonstrated that it is also possible to assimilate GRACE TWSA into a land surface model so as to improve the estimation of snowpack variables. To assimilate GRACE data, I would again most likely employ the multiple coarse-scale observation approach laid out in De Lannoy et al. (2010).

ICESat (2003-2010) was a satellite in polar orbit that was equipped with a spaceborne-LIDAR. It was subsequently able to accurately retrieve snow depth at relatively high (order 100 m) resolution using DEM differencing as shown in Treichler and Kääb (2017). Given that the successor mission, ICESat-2, was recently launched space-borne LIDAR also presents a viable constraint on the state of the seasonal snowpack. The advantage with these satellites is that they can be used to retrieve spatially representative snow depth at the modeled hillslope-scale which will serve as a strong constraint in the assimilation. It has already been shown that the assimilation of snow depth observations substantially improves the estimation of SWE in snow models (Magnusson et al., 2017), so assimilating spaceborne-LIDAR derived snow depth into our framework will undoubtedly be a fruitful addition that is worth exploring.

4.6 Smoothing, filtering, and prediction

In this thesis, I have exclusively considered the reanalysis or smoothing problem as its known in the DA literature. We can clearly differentiate between the three related problems of prediction, filtering, and smoothing (Jazwinski, 1970; Carrassi et al., 2018). In the prediction problem, the model trajectory is only constrained by past observations relative to the current timestep. Prediction can also be referred to as forecasting. Note in particular that prediction refers to unconstrained estimation of the future relative to the last assimilated observation. So by future I mean a relative future, not necessarily the actual future. In the filtering problem, the model trajectory is constrained both by past and present observations relative to the current time step. Filtering is also known as sequential data assimilation. In the smoothing problem that I have been concerned with, the model trajectory is constrained by past, current, and future observations. In particular, I have considered the batch smoothing problem where all observations are assimilated at once to update uncertain model (and forcing) parameters. A graphical overview of the difference between the three problems is provided in Figure 4.1.

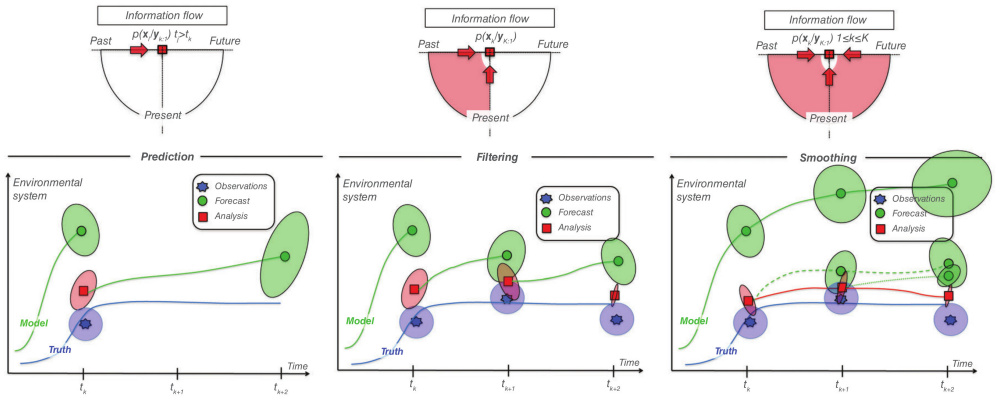


Figure 4.1: Schematic illustration of the difference between the prediction, filtering, and smoothing problems. Adapted with permission from Carrassi et al. (2018).

The smoothing approach can be superior to filtering in the context of land surface data assimilation in general and snow data assimilation in particular. Smoothing can take into account the long memory (auto-correlation) of model errors (Koster and Suarez, 2001; Dunne and Entekhabi, 2005) while filtering ignores this due to the Markov assumption. Moreover, in the snow reconstruction exercise we need to be able to use future observations to update the past model behaviour. This is

necessary because information concerning biases in the accumulation season will be contained in the observed snow-cover depletion that occurs during the ablation season. Smoothers allow for this backwards propagation of information while filters do not. Another advantage with batch smoothing is that when the model is rerun with updated parameters its states will be dynamically consistent, which is not the case when sequentially updating states and parameters with a filter. In addition, when batch smoothing entire model integrations are carried out before any updating is performed, such that the DA can be easily applied even to models where the source code is not openly available or where it is difficult to stop time stepping. Other than fixed-interval batch smoothers, there exists a plethora of different smoothers as reviewed in Cosme et al. (2012).

When DA is performed online, i.e. as observations become available, it is of course often more practical to use a filter than a smoother. This is the typical case in weather forecasting where the posterior of today becomes the prior of tomorrow. Another advantage with filtering is that one does not need to store the entire vector of observations and predicted observations, as one does when smoothing, which can be prohibitively expensive in certain cases. When modeling systems that are governed by largely hyperbolic (as opposed to parabolic or elliptic) systems of governing equations it thus often makes more sense to apply a filter. This explains why the PF and EnKF enjoy such widespread use in the meteorological and oceanographic communities (see van Leeuwen, 2009; Evensen, 2009, and references therein). In particular, filters are especially nicely tailored for initializing prediction problems. It is not as clear, however, if this is a good reason to employ such filters in the land surface modeling community where the dynamics are quite different and there is often more uncertainty tied to boundary conditions than to initial conditions. This will often lead to model biases which are more challenging to handle with a filter.

In the fields where snow is an important component, such as hydrology, weather, and climate science there is a great deal of interest in the prediction problem. That is to say, we often want to make predictions about a future that is unconstrained by observations. It may not be immediately clear how snow reanalysis, which is retrospective by definition, can help solve the prediction problem. Upon some reflection, one way in which snow reanalyses can be helpful is for validation and model benchmarking. For example, Wrzesien et al. (2017) used the Sierra Nevada snow reanalysis conducted by Margulis et al. (2016) as a reference to evaluate

estimates of SWE at the mountain range scale from global products and a regional model. The regional model that was used, namely the Weather Research and Forecasting (WRF) model, is frequently used to make both weather and climate predictions (e.g. Rasmussen et al., 2011). In addition to validating models, the snow reanalysis could be used to develop and tune parametrizations in the snow-component of models. For example, given the hillslope-scale of the snow reanalysis, it could be used to help further improve parametrizations of subgrid snow variability in coupled land-atmosphere models (Aas et al., 2017). Similarly, the reanalysis, which also represents a physically-based interpolation of remotely sensed fSCA time series, can be used to tune the parameters in the necessarily relatively simple snow-depletion curves that are employed in climate models (e.g. Swenson and Lawrence, 2012). This would then hopefully lead to improved climate predictions at both seasonal and decadal timescales (Henderson et al., 2018).

Recently, SWE results from deterministic snow reconstructions have also been used as targets to train machine learning algorithms which can subsequently be used to make predictions (Bair et al., 2018). There is no reason that the (probably improved) results from an ensemble-based snow reanalysis could not be used for the same purpose. For example, the snow reanalysis of Margulis et al. (2016) would provide an extensive library of target SWE for training machine learning algorithms that spans back to the beginnings of the satellite era. Trained machine learning algorithms, such as neural networks or random forests, could then be used to make real-time estimates using predictors such as various satellite retrievals and terrain characteristics as in Bair et al. (2018).

Another important potential application of data assimilation is in the context of model selection. Since the framework is modular, it is possible to employ the same model in the reanalysis and later prediction. That is, given sufficient computational resources, it would be straightforward to switch out the simple snow model with a more complex snow model that could be coupled to a meteorological or climate forecast and be used to make predictions. The reanalysis framework could then help in selecting the appropriate model for this task even in the absence of independent validation data, i.e. using the fSCA data that is assimilated. By employing a series of different models (or model configurations) the model selection could then be performed by considering the model evidence (the denominator in Bayes theorem) as an alternative to the RMSE (Metref et al., 2019).

Chapter 5

Conclusion

In this thesis, my aim was to help tackle a longstanding unsolved problem in snow hydrology (Dozier et al., 2016), namely our current inability to accurately estimate the snow water equivalent (SWE) anywhere on Earth at the hillslope scale. On my quest to contribute a solution to this problem, I took a purely retrospective approach by developing a framework for ensemble-based reanalysis of the seasonal snowpack. The framework was inspired by pioneering applications of snow reconstruction (Martinec and Rango, 1981), Bayesian inversion of snow depletion curves (Kolberg and Gottschalk, 2006), and ensemble-based snow reanalysis (Durand et al., 2008). The modular framework was built up of four components:

(i) A routine to topographically downscale forcing from coarse-scale meteorological reanalyses which largely follows the TopoSCALE approach of Fiddes and Gruber (2014).

(ii) A simple snow model based largely on the Utah Energy Balance model (Tarboton and Luce, 1996), with some simplifying assumptions, and the probabilistic snow depletion curve of Liston (2004).

(iii) Algorithms to retrieve fractional snow-covered area (fSCA) using multispectral imagery obtained from polar orbiting optical satellites. The algorithms that we used ranged in complexity from simple thresholding of the NDSI (Dozier, 1989) to spectral unmixing (SU) (Painter et al., 2009).

(iv) An ensemble-based batch smoother data assimilation (DA) scheme. Here we primarily used the ensemble smoother with multiple data assimilation (ES-MDA; Emerick and Reynolds, 2013), altough the ensemble smoother (ES;

van Leeuwen and Evensen, 1996) and the particle batch smoother (PBS; Margulis et al., 2015) were also used.

I laid out five research objectives that were tied to constructing these components and tying them together into an ensemble-based snow reanalysis framework. Through the papers Aalstad et al. (2018), Aalstad et al. (2019a), Aalstad et al. (2019b), and Fiddes et al. (2019), we targeted these objectives at three disparate study areas, the Brøgger peninsula, the Mammoth Lakes Basin, and the Swiss Alps. In Aalstad et al. (2018), the framework was introduced and we showed how it could be used to accurately reconstruct the subgrid SWE distribution at several sites on the Brøgger peninusla using fSCA retrievals from MODIS and Sentinel-2. We found that the ES-MDA scheme exceeded or at least nearly matched the performance of previously proposed DA schemes with respect to various evaluation metrics. Subsequently, in Aalstad et al. (2019a), we gauged our current capabilities of retrieving fSCA from polar orbiting optical satellites in the Brøgger peinsula. Here, we used time-lapse photography to help establish that only the SU technique provided unbiased satellite retrievals of fSCA at the hillslope-scale. In Aalstad et al. (2019b), we used multiple satellite constellations (Sentinel-2, Landsat 7&8, PlanetScope, and RapidEye) to retrieve and assimilate fSCA into the framework at the Brøgger peninsula and the Mammoth Lakes Basin. Using independent validation data, we showed how using fSCA retrievals from multiple sensors as opposed to a single sensor provided marked improvements in performance. We demonstrated once again that the ES-MDA was more robust than both the PBS and the ES in the snow reanalysis exercise. In Fiddes et al. (2019), we showed how the framework can be scaled up and applied to larger areas in the Swiss Alps with the help of a clustering technique.

Given that our reanalysis framework is still in the research and development stage, there is work to be done before it can become operational. Currently the biggest bottleneck is the processing of the satellite imagery, particularly regarding the cloud masking. At the same time, being a modular framework that can be applied anywhere on Earth it can readily be adapted to other uses. For example, the snow reanalysis framework could be efficiently coupled to a permafrost model such as that of Westermann et al. (2016) to help deliver a transient permafrost reanalysis where snow-induced uncertainties have been constrained. More generally, parts of the framework could be implemented into the next generation of global fully coupled land-atmosphere reanalysis systems. Overall, we have shown that, given the inherent uncertainties, models and data work best together, not in isolation.

Part II

Journal Publications

Paper I

I

Ensemble-based assimilation of fractional snow-covered area satellite retrievals to estimate the snow distribution at Arctic sites

K. Aalstad, S. Westermann, T. V. Schuler, J. Boike, L. Bertino

Published in *The Cryosphere*, January 2018, volume 12, pp. 247–270,
doi: 10.5194/tc-12-247-2018.

The Cryosphere, 12, 247–270, 2018
https://doi.org/10.5194/tc-12-247-2018

The Cryosphere Open Access EGU

Ensemble-based assimilation of fractional snow-covered area satellite retrievals to estimate the snow distribution at Arctic sites

Kristoffer Aalstad[1], **Sebastian Westermann**[1], **Thomas Vikhamar Schuler**[1], **Julia Boike**[2], **and Laurent Bertino**[3]

[1]Department of Geosciences, University of Oslo, P.O. Box 1047, Blindern, 0316 Oslo, Norway
[2]Alfred Wegener Institute Helmholtz Center for Polar and Marine Research, Telegrafenberg A43, 14473 Potsdam, Germany
[3]Nansen Environmental and Remote Sensing Center, Thormøhlens gate 47, Bergen 5006, Norway

Correspondence: Kristoffer Aalstad (kristoffer.aalstad@geo.uio.no)

Received: 13 June 2017 – Discussion started: 4 July 2017
Revised: 1 December 2017 – Accepted: 12 December 2017 – Published: 23 January 2018

Abstract. With its high albedo, low thermal conductivity and large water storing capacity, snow strongly modulates the surface energy and water balance, which makes it a critical factor in mid- to high-latitude and mountain environments. However, estimating the snow water equivalent (SWE) is challenging in remote-sensing applications already at medium spatial resolutions of 1 km. We present an ensemble-based data assimilation framework that estimates the peak subgrid SWE distribution (SSD) at the 1 km scale by assimilating fractional snow-covered area (fSCA) satellite retrievals in a simple snow model forced by downscaled reanalysis data. The basic idea is to relate the timing of the snow cover depletion (accessible from satellite products) to the peak SSD. Peak subgrid SWE is assumed to be log-normally distributed, which can be translated to a modeled time series of fSCA through the snow model. Assimilation of satellite-derived fSCA facilitates the estimation of the peak SSD, while taking into account uncertainties in both the model and the assimilated data sets. As an extension to previous studies, our method makes use of the novel (to snow data assimilation) ensemble smoother with multiple data assimilation (ES-MDA) scheme combined with analytical Gaussian anamorphosis to assimilate time series of Moderate Resolution Imaging Spectroradiometer (MODIS) and Sentinel-2 fSCA retrievals. The scheme is applied to Arctic sites near Ny-Ålesund (79° N, Svalbard, Norway) where field measurements of fSCA and SWE distributions are available. The method is able to successfully recover accurate estimates of peak SSD on most of the occasions considered. Through the ES-MDA assimilation, the root-mean-square error (RMSE) for the fSCA, peak mean SWE and peak subgrid coefficient of variation is improved by around 75, 60 and 20 %, respectively, when compared to the prior, yielding RMSEs of 0.01, 0.09 m water equivalent (w.e.) and 0.13, respectively. The ES-MDA either outperforms or at least nearly matches the performance of other ensemble-based batch smoother schemes with regards to various evaluation metrics. Given the modularity of the method, it could prove valuable for a range of satellite-era hydrometeorological reanalyses.

1 Introduction

The spatiotemporal distribution of seasonal snow cover is a key control on the terrestrial surface energy and water balance in mid- to high-latitude regions and mountainous areas (Boike et al., 2003; Barnett et al., 2005). With its high albedo and large water-holding capacity, snow is a modulator of the global radiation balance and hydrological cycle, making it one of the drivers of the atmospheric circulation and the associated climate (Andreadis and Lettenmaier, 2006; Liston, 1999). Since the snow water equivalent (SWE) can exhibit considerable variability over small distances (Clark et al., 2011), mapping the SWE distribution remains a difficult task (Dozier et al., 2016).

The primary controls on the distribution and variability of SWE are topography, vegetation, precipitation, wind, radiation and avalanching (Sturm and Wagner, 2010; Clark et al., 2011). While topography and vegetation are relatively fixed in time, the other controls vary strongly over a range of spatiotemporal scales. In unforested regions, snow tends to be affected by wind drift (e.g., Gisnås et al., 2014), leading to

accumulation in areas with preferential deposition, such as topographic depressions or the lee side of a ridge. The scale of such features can vary dramatically across the landscape. Nonetheless, the processes occurring at a given site are often consistent from year to year, and so the SWE distribution is often quite similar to the climatological snow distribution pattern (Sturm and Wagner, 2010; Kępski et al., 2017). Manual measurement surveys are usually impractical for mapping SWE over large areas given their limited support, large spacing and small extent (Blöschl, 1999). Instead, modeling and remote sensing can be employed to map SWE.

Snow models range in complexity from relatively simple single-layer models, such as the Utah Energy Balance model (UEB; Tarboton and Luce, 1996; You et al., 2014), to detailed multilayer snowpack models, such as Crocus (Vionnet et al., 2012) and SNOWPACK (Bartelt and Lehning, 2002). Some snow models (e.g., ALPINE3D; Lehning et al., 2006) can also be run in distributed mode to simulate the snow distribution over large areas. The accuracy of the model results is limited by the hydrometeorological forcing data, be it from reanalyses or local measurements, whose errors are typically the major source of uncertainty in snow modeling (De Lannoy et al., 2010; Raleigh et al., 2015). In addition, snow models are generally developed as point-scale models; even if they are run as distributed models, the grid-scale values predicted by the model may not be representative of the corresponding process scale (Blöschl, 1999). For example, if a snow model is forced by near-point-scale hydrometeorological measurements, the model results will only be representative for a grid cell if that particular point is representative of the mean conditions within the grid cell. To circumvent this problem, probabilistic snow depletion curve (SDC) parametrizations have been developed (Liston, 1999; Luce and Tarboton, 2004; Liston, 2004) in which a probability distribution function is assigned to the SWE within a grid cell at peak accumulation. Assuming uniform melt across the grid cell, this allows for a direct relationship between the mean SWE, melt depth and fractional snow-covered area (fSCA) of the grid cell. Liston (2004) used such a SDC parametrization in conjunction with land-cover-specific subgrid coefficients of variation of SWE with the ClimRAMS model to map the fSCA over North America. As a result, the total snow-covered area increased considerably compared to the control run. Aas et al. (2017) used a tiling approach to represent subgrid snow variability in the WRF model coupled to the Noah land surface scheme over southern Norway. The tiling reduced the cold bias in the modeled near-surface air temperatures and greatly improved the match to the observed fSCA evolution. Nevertheless, due to the inherently large uncertainties in the forcing, modeling alone is usually not a sufficiently accurate tool for mapping SWE. Instead, models need to be combined with relevant data from remote sensing.

Snow-related data sets can be acquired from a variety of remote-sensing platforms with near-global coverage. The Gravity Recovery and Climate Experiment (GRACE) twin satellites allow for the retrieval of terrestrial water storage (TWS), from which SWE can be recovered at around 100 km spatial resolution (e.g., Niu et al., 2007). Passive microwave (PM) satellite sensors can retrieve SWE based on brightness temperature at a resolution of around 25 km. However, PM SWE retrievals have problems over forested areas and complex topography, as well as for wet and deep snowpacks (Foster et al., 2005). Both gravimetric and PM sensors are able to retrieve SWE independent of cloud coverage, resulting in gap-free time series. While not capable of measuring SWE, moderate-resolution optical sensors such as the Moderate Resolution Imaging Spectroradiometer (MODIS) can retrieve binary information on snow cover (i.e., snow or no snow), fSCA and snow grain size (Hall et al., 2002; Salomonson and Appel, 2004; Painter et al., 2009) at approximately 500 m resolution with a daily revisit frequency. In addition, higher-resolution optical sensors, such as those on board the Landsat and Sentinel-2 satellites, can map fSCA at around 30 m resolution (e.g., Cortés et al., 2014). Optical sensors can not see through clouds, which results in data gaps over most snow-covered regions. To obtain gap-free time series, it is thus necessary to either interpolate optical remote-sensing data in time and space or ingest them in models.

Data assimilation (DA) methods can objectively fuse uncertain information from observations and models. Deterministic SWE reconstruction techniques (Girotto et al., 2014b) that directly insert remotely sensed fSCA data in models represent the simplest form of snow data assimilation. Such schemes back-calculate peak SWE from the disappearance date of the snow cover (as determined from fSCA retrievals) using snowmelt models. Martinec and Rango (1981) used Landsat fSCA retrievals during the melt season in conjunction with a simple degree day snowmelt model to estimate the peak mean SWE. Similarly, Cline et al. (1998) used Landsat fSCA retrievals combined with a distributed energy balance model to reconstruct the SWE distribution. More recently, Molotch and Margulis (2008) used fSCA information from multiple sensors for deterministic SWE reconstruction. Durand et al. (2008) introduced a probabilistic framework for SWE reconstruction. This was based on assimilating synthetic fSCA retrievals during the ablation into the Simplified Simple Biosphere version 3 (SSiB3) land surface model coupled to the SDC of Liston (2004) using the ensemble smoother (ES; Van Leeuwen and Evensen, 1996) in batch mode (cf. Dunne and Entekhabi, 2005). The assimilation of synthetic fSCA in this twin experiment was used to correct annual biases in the snowfall and facilitated the recovery of the SWE distribution. Using the Durand et al. (2008) framework, Girotto et al. (2014b) assimilated Landsat fSCA retrievals to recover the SWE distribution, yielding a significant reduction in root-mean-square error (RMSE) relative to deterministic SWE reconstruction. Subsequently, Girotto et al. (2014a) used the same framework to perform a 27-year reanalysis of SWE distributions. Recently, Margulis et al. (2015) modified this probabilistic approach by adopt-

ing a particle batch smoother (PBS) as opposed to the ES for the assimilation of fSCA retrievals to estimate the SWE distribution. The PBS was found to outperform the ES, considerably reducing the RMSE. Based on this work, Margulis et al. (2016) adopted the PBS framework to conduct a 30-year reanalysis of SWE over the Sierra Nevada (USA) using Landsat fSCA retrievals. Cortés et al. (2016) applied the same PBS framework to construct a 30-year reanalysis of SWE over six instrumented basins in the Andes. Cortés and Margulis (2017) subsequently adopted this approach to perform a 31-year SWE reanalysis over the entire extratropical Andes.

Several other snow DA techniques have recently been employed. Andreadis and Lettenmaier (2006) assimilated MODIS fSCA retrievals into the VIC model through the ensemble Kalman filter (EnKF; Evensen, 2009) using a simple SDC for the SWE-fSCA inversion. However, the improvement compared to the open loop (OL; i.e., no DA) run was only modest, which was also found in similar studies (Clark et al., 2006; Slater and Clark, 2006). A Bayesian technique was used by Kolberg and Gottschalk (2006) to assimilate Landsat fSCA retrievals into a snow model with a probabilistic SDC to estimate the peak SWE distribution. They found a significant reduction in uncertainty when retrievals were assimilated simultaneously as opposed to sequentially. At the continental scale, a multisensor assimilation of both GRACE TWS and MODIS fSCA using the ES and EnKF for TWS and fSCA, respectively, yielded significant improvements relative to the OL (Su et al., 2010). De Lannoy et al. (2010) used the EnKF in a twin experiment to assimilate synthetic PM SWE retrievals and greatly outperformed the OL. This was extended to a real multisensor experiment by jointly assimilating PM SWE and MODIS fSCA retrievals (De Lannoy et al., 2012). Li et al. (2017) used the ES to assimilate PM SWE retrievals and estimate the SWE distribution, markedly outperforming the OL. Of late, particle filter (PF; see Van Leeuwen, 2009) schemes have been gaining popularity in snow DA studies (Charrois et al., 2016; Magnusson et al., 2017). For example, Charrois et al. (2016) assimilated synthetic optical reflectance retrievals into Crocus using the sequential importance re-sampling PF at a point scale and considerably outperformed the OL.

It is worth emphasizing that the most popular schemes in the snow DA community, both the EnKF and the PF, are filters (i.e., sequential techniques). As such, they are Markovian of order 1 (memoryless): the future state at a given point in time depends only on the present state. Furthermore, observations are assimilated sequentially with only the current observation affecting the current state. Batch smoothers (Dunne and Entekhabi, 2005), on the other hand, take into account the entire history of a model trajectory within a batch (observation window) and as such have memory (non-Markovian) so that they are better suited for reanalysis problems.

In this study, we build on the probabilistic SWE reconstruction technique outlined in Girotto et al. (2014b) to recover subgrid SWE distributions (SSDs) for a study area in the Arctic based on fSCA retrievals from MODIS and Sentinel-2. The novelty of our study lies in the use of an iterative batch smoother scheme, namely the ensemble smoother with multiple data assimilation (ES-MDA; Emerick and Reynolds, 2013). To update physically bounded parameters, we make use of analytical Gaussian anamorphosis (Bertino et al., 2003). We investigate the performance of the ES-MDA in terms of SWE reconstruction and compare it to the ES and the PBS employed by Girotto et al. (2014b) and Margulis et al. (2015), respectively. The results are evaluated against independent field measurements of fSCA and snow surveys conducted over six snow seasons.

2 Study area

2.1 Physical characteristics and climate

The study area is located in NW Svalbard close to the research town of Ny-Ålesund (78°55′ N, 11°50′ E) on the Brøgger Peninsula. Field measurements are available from three sites (Fig. 1). "Bayelva", about 2 km west of Ny-Ålesund, is the main study site where multiyear in situ records on, for instance, the surface energy balance, permafrost thermal regime and snow distribution are available (Westermann et al., 2009; Gisnås et al., 2014; Boike et al., 2017). In addition, snow surveys for a single season (2016) are available from "Steinflåen plateau" and "Kvadehuksletta". All sites feature gently undulating topography with small hills and surfaces characterized by patterned ground features, leading to strong differences in snow cover due to wind drift. Bayelva and Kvadehuksletta are located between 10 and 50 m a.s.l., while the Steinflåen plateau is at a higher elevation of around 200 m a.s.l. Kvadehuksletta is exposed to most wind directions, whereas Bayelva and Steinflåen plateau are partly sheltered by mountains. The sites are located within the continuous permafrost zone (Boike et al., 2003) with a maximum active layer depth of around 1.5 m at the Bayelva site (Westermann et al., 2009).

The Bayelva site is located around the heavily instrumented Bayelva climate and soil monitoring station (Boike et al., 2017). This area has been the subject of extensive studies spanning permafrost (Roth and Boike, 2001; Boike et al., 2008; Westermann et al., 2011a), the surface energy balance (Boike et al., 2003; Westermann et al., 2009), CO_2 exchange (Lüers et al., 2014; Cannone et al., 2016), ecology (Kohler and Aanes, 2004), snow (Bruland et al., 2001; Gisnås et al., 2014; López-Moreno et al., 2016), hydrology (Nowak and Hodson, 2013) and satellite retrieval validation (Westermann et al., 2011b, 2012). The surface cover at Bayelva and Kvadehuksletta alternates between bare soil, rocks and sparse low vegetation (Westermann et al., 2009), while the more elevated Steinflåen plateau is predominantly covered by loose rocks.

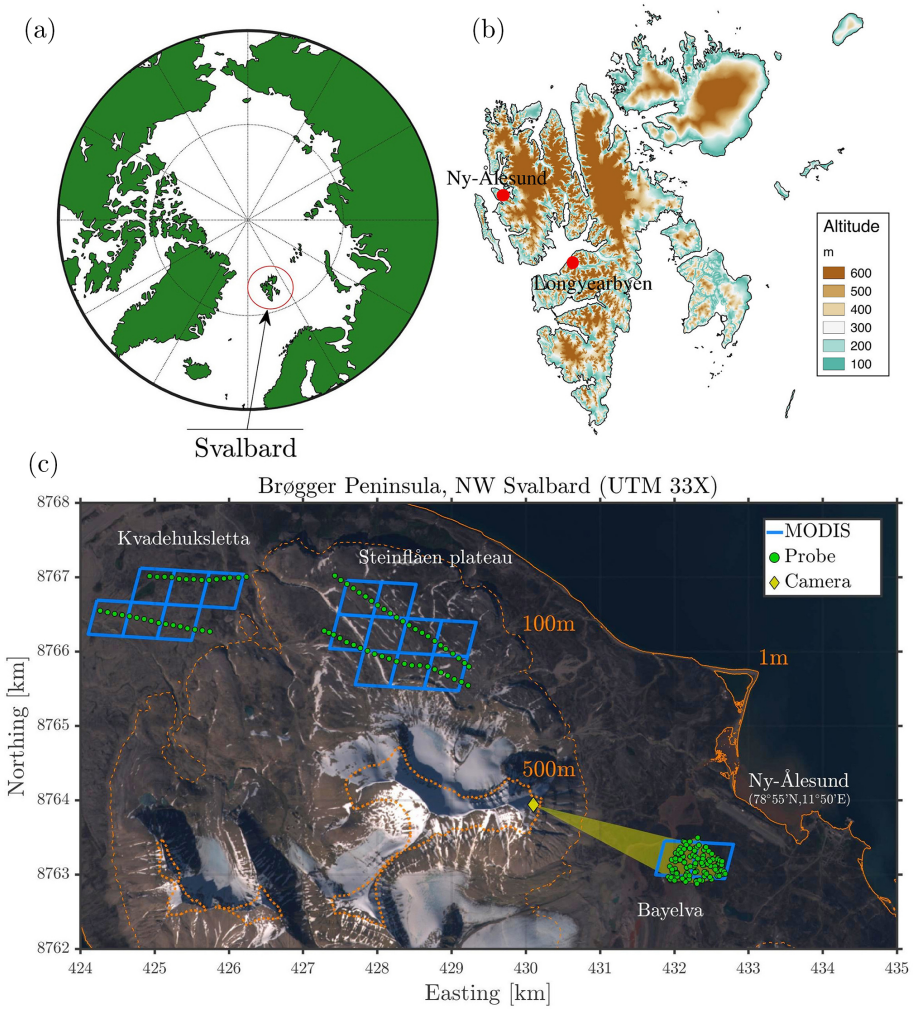


Figure 1. (a) location of Svalbard (in red) in the Arctic. **(b)** Map of Svalbard; the study area is close to Ny-Ålesund. Bottom: Sentinel-2A true-color image (2 July 2016) of the western Brøgger Peninsula with the three study sites Kvadehuksletta, Steinflåen plateau and Bayelva; green dots: snow survey probe locations; blue polygons: MODIS pixels; yellow diamond: automatic camera system on Scheteligfjellet; yellow shading: field of view of the camera; contour lines courtesy of the NPI (2014) DEM.

The climate of western Svalbard is influenced by the relatively warm West Spitsbergen Current causing a maritime climate with mild winters and cool summers for this latitude (Esau et al., 2012). At Ny-Ålesund the winter, summer and annual (1981–2010) average air temperatures were −12.0, 3.8 and −5.2 °C, respectively, while the average annual precipitation was 427 mm (Førland et al., 2011). Between September/October and May the precipitation mainly falls as snow, although rain-on-snow events have become more frequent due to the warming of the local climate (Nowak and Hodson, 2013; López-Moreno et al., 2016). The seasonal snow cover usually forms in late September or early October and lasts until mid-June to early July, with a melt season of around 1 month (Winther et al., 2002). The dominant energy source during the snowmelt is radiation (longwave and shortwave), while the heat flux required to warm the frozen ground underlying the snow is an important energy sink (Boike et al., 2003; Westermann et al., 2009).

2.2 Field measurements

Manual surveys of snow depth and density were carried out in April/May for 6 years at the Bayelva site and for 1 year (2016) at the two other sites (Table 1). At this time, the snow depth is near its maximum but the snowpack is still dry. The snow density was sampled in vertical layers at every fifth point. As no systematic stratification of the snow density was found, SWE was finally calculated from snow depth and the average snow density at each site in a given year. At Bayelva, the snow density was generally confined to a range of $350 \pm 50\,\mathrm{kg\,m^{-3}}$ for all the surveys, while the snow density was found to be around $450\,\mathrm{kg\,m^{-3}}$ at Steinflåen plateau in 2016. At Kvadehuksletta and Steinflåen plateau, the sur-

Table 1. Overview of the study sites and snow surveys. $\bar{z}$ is the mean elevation, and σ_z is the standard deviation in the elevation, both based on the NPI (2014) DEM. ^{t}: transect; ^{r}; randomized array.

Location	$\bar{z}$ [m a.s.l.]	σ_z [m]	Area [km^2]	Survey years	Samples per survey
Bayelva	23	9	0.5	2008, 2009, 2013–2016	853^{t}, 617^{t}, 105^{r}
Steinflåen plateau	210	11	1.1	2016	45^{t}
Kvadehuksletta	55	6	0.9	2016	30^{t}

veys were conducted along transects with regular sample intervals (see Fig. 1). A randomized array of sample points was employed for Bayelva in most years, except for the first 2 years where transects were used.

Basal ice layers resulting from rain-on-snow events (Kohler and Aanes, 2004; Westermann et al., 2011a) occur in the area and can constitute a major source of uncertainty for SWE measurements. In 2016, the depth of basal ice layers was measured using ice screws, and their contribution to the SWE was accounted for. In addition, internal ice layers and the spatial variability of average snow densities (see above) contribute to the uncertainty of the measurements. Furthermore, only a limited number of sampling points are available, so that the obtained snow distributions are expected to deviate to a certain extent from the true snow distributions in the area. Although the snow surveys coincide closely with peak SWE, some accumulation (ablation) may occur after (before) the surveys. To assess the magnitude of this error source, we used snow depth measurements at the Bayelva station (Boike et al., 2017) to compare the snow depth at the survey dates to the maximum snow depth for each snow season. We found an average relative difference of 8 % (maximum: 17 %; minimum: 0.3 %).

In 2012, 2013 and 2016, an automatic time-lapse camera was deployed near the summit of Scheteligfjellet (694 m a.s.l.; cf. Fig. 1), overlooking the Bayelva site. The camera was a standard digital camera triggered by a Harbortronics time-lapse system, delivering daily images except for prolonged periods with low cloud cover. The raw camera images were orthorectified at a 1 m resolution, and snow was mapped for each pixel using a threshold on the intensity, so that fSCA could be determined for each image. The orthorectified images for two of the years are freely available in Westermann et al. (2015a).

In 2008, aerial images were obtained for the Bayelva site for five dates in June during the beginning of the snowmelt period. This was accomplished by mounting a digital camera to an unmanned aerial vehicle (UAV) flying at an altitude of 100 to 250 m above ground which took between 700 and 1000 images per mission at nadir angles. As the images were taken in a near-random fashion over the entire area, fSCA was calculated by averaging over the fSCA determined for each image using a simple threshold criterion. GPS-based surveys of the remaining snow patches were available for five additional dates, so that a complete fSCA time series is available for the snowmelt period in 2008.

3 Method

3.1 Simple snow model

To efficiently run a large number of ensemble members, a simple snow model (SSM) is employed, which computes snowmelt rates according to surface energy balance formulations (as in the CryoGrid 3 ground thermal model; Westermann et al., 2016). The model is a blend of a single-layer mass balance scheme, based on the UEB model (Tarboton and Luce, 1996; You et al., 2014), and the Liston (2004) SDC. Many internal snow processes (occurring inside the snowpack), including heat conduction and meltwater percolation, are omitted. In addition, several external processes such as sublimation and deposition are ignored. In the following sections, we describe the governing equations of the SSM (see Table 2 for the model constants).

3.1.1 Snow depletion curve

We use the SDC parametrizations discussed in Liston (1999), Luce and Tarboton (2004) and Liston (2004) which parametrize the relationship between fSCA, melt depth and SWE by using a probability density function (pdf) to represent the peak SSD. A key assumption is that the melt rate is spatially uniform within each grid cell. The relationship between the accumulated melt depth (D_m), the peak SSD pdf (f_P) and the fSCA within the grid cell at time t is given by

$$\mathrm{fSCA}(t) = \int_{D_\mathrm{m}(t)}^{\infty} f_\mathrm{P}(D)\,\mathrm{d}D\,. \tag{1}$$

Similarly, the mean SWE depth is given by

$$\overline{D}(t) = \int_{D_\mathrm{m}(t)}^{\infty} (D - D_\mathrm{m}(t))\, f_\mathrm{P}(D)\,\mathrm{d}D\,. \tag{2}$$

Following Liston (2004), we parametrize the peak SSD using a two-parameter lognormal pdf $f_\mathrm{P} = f_\mathrm{P}(D|\mu, \chi)$, where μ is the peak mean SWE and $\chi = \sigma/\mu$ is the peak subgrid coefficient of variation (σ is the standard deviation). χ is a

Table 2. List of the model constants used in the simple snow model runs along with the corresponding reference studies.

Symbol	Name	Value	Units (SI)	Reference
$\alpha_{\max}$	Maximum snow albedo	0.85	–	Dutra et al. (2010)
τ_S	Threshold snowfall	0.01	m (w.e.)	Dutra et al. (2010)
τ_F	Aging constant for melting snow	2.78×10^{-8}	s^{-1}	Dutra et al. (2010)
τ_A	Aging constant for non-melting snow	9.26×10^{-8}	s^{-1}	Dutra et al. (2010)
T_R	Threshold temperature for rain	276.15	K	You et al. (2014)
T_S	Threshold temperature for snow	272.15	K	You et al. (2014)
ε_S	Emissivity of snow	0.99	–	Westermann et al. (2016)
d_H	Thermal diffusivity of the ground	6×10^{-7}	$m^2\,s^{-1}$	Westermann et al. (2009)
z_E	Effective transfer depth	1	m	–
Δt	Daily time step	86 400	s	–
L_f	Specific latent heat of fusion	3.35×10^5	$J\,kg^{-1}$	Tarboton and Luce (1996)
ρ_w	Density of fresh liquid water	10^3	$kg\,m^{-3}$	Tarboton and Luce (1996)
σ_{SB}	Stefan–Boltzmann constant	5.67×10^{-8}	$W\,m^{-2}\,K^{-4}$	Tarboton and Luce (1996)

perturbation parameter (see Table 4) that is updated in the assimilation. Our choice of parametric distribution was motivated by independent measurements of the SSD which fit reasonably well to a lognormal distribution (Bruland et al., 2001). Equations (1) and (2) can both be solved analytically as presented in Liston (2004).

3.1.2 Mass and energy balance

To obtain the instantaneous net accumulation rate, $\mathcal{A}(t)$, we follow the UEB model through (You et al., 2014)

$$\mathcal{A}(t) = P(t) - M(t)\,, \tag{3}$$

where $P(t)$ is the precipitation rate and $M(t)$ is the melt rate. Sublimation is not considered as it is a relatively small contribution to the mass balance at our study area (Westermann et al., 2009). We use a linear transition to delineate between snowfall and rainfall (You et al., 2014), with thresholds given in Table 2. We only consider rainfall as a positive contribution to the mass balance during non-melting conditions when the rainwater generally refreezes in the snowpack (Westermann et al., 2011a). For melting conditions (where $D_m > 0$), we assume that rainfall directly becomes runoff.

The melt rate, M, is calculated based on a simplified snow energy balance defined by

$$Q_M(t) = Q_R^*(t) + Q_P(t) - Q_H(t) - Q_E(t) - Q_G(t)\,, \tag{4}$$

where Q_M is the snowmelt flux, Q_R^* is the global radiation, Q_P is the heat advected by precipitation, Q_H is the sensible heat flux, Q_E is the latent heat flux and Q_G is the ground heat flux. The last three fluxes are defined as positive when directed away from the surface and vice versa for the first two on the right-hand side of Eq. (4). The SSM differs from UEB in that we calculate the surface energy balance for a melting snowpack, i.e., isothermal at 0 °C, at all times. In this case, the global radiation is

$$Q_R^* = (1 - \alpha_S)\, S^{\downarrow} + L^{\downarrow} - \varepsilon_S \sigma_{SB} T_0^4\,, \tag{5}$$

in which $S^{\downarrow}$ and $L^{\downarrow}$ are the downwelling shortwave and longwave irradiances, and the last term is the upwelling longwave radiation for the assumed isothermal snowpack at $T_0 = 273.15$ K. The snow albedo (α_S) is parametrized prognostically through the continuous reset formulation following Dutra et al. (2010), which computes the albedo for time increments Δt by distinguishing between accumulating, steady and ablating conditions:

$$\alpha_S(t + \Delta t) = \tag{6}$$
$$\begin{cases} \alpha_S(t) + \min\left(1, \mathcal{A}(t)\Delta t/\tau_S\right)\left(\alpha_{\max} - \alpha_S(t)\right)\,, & \mathcal{A}(t) > 0\,, \\ \max\left(\alpha_S(t) - \tau_A \Delta t, \alpha_{\min}\right)\,, & \mathcal{A}(t) = 0\,, \\ \left(\alpha_S(t) - \alpha_{\min}\right)\exp\left(-\tau_F \Delta t\right) + \alpha_{\min}\,, & \mathcal{A}(t) < 0\,. \end{cases}$$

Here, $\alpha_{\min}$ and $\alpha_{\max}$ are the minimum and maximum snow albedo values, respectively, while τ_A and τ_F are aging (decay) rates for non-melting and melting snow, respectively. τ_S is a threshold for daily snowfall which, if exceeded, leads to a reset of the snow albedo to its maximum value. $\alpha_{\min}$ is a perturbation parameter (see Table 4) that is updated in the assimilation. This simple decay and reset type of snow albedo parametrization has been shown to perform reasonably well at Bayelva (Pedersen and Winther, 2005). The heat advected by rainfall (Q_P) is computed as in Tarboton and Luce (1996), while the turbulent fluxes of sensible (Q_H) and latent (Q_E) heat are evaluated following Westermann et al. (2016). The ground heat flux (Q_G) is parametrized through a simple e-folding relationship during the melting period, i.e.,

$$Q_G = Q_0 \exp\left(-d_H t_m / z_E^2\right), \tag{7}$$

where Q_0 is the initial ground heat flux, d_H is the thermal diffusivity of the ground, z_E is the effective depth of the heat transfer below the base of the snowpack and t_m is the number of days with melting conditions after peak accumulation. Q_0 is a perturbation parameter (see Table 4) that is updated in the

assimilation, d_H is selected according to field measurements (Westermann et al., 2009) and z_E is set so that the ground heat flux decays to near zero a month into the melt season.

The snowmelt flux Q_M can now be evaluated through Eq. (4). We recall that an isothermal snowpack at 0 °C is assumed for Eq. (4), which is only justified for a melting snowpack. In this case, positive Q_M values correspond to melting and SWE reduction, while negative values correspond to refreezing of meltwater and thus SWE increase. For a dry snowpack (as is generally the case before the snowmelt), negative Q_M values would lead to a cooling of the snowpack, which is not considered in this simple snowmelt scheme. To discard unphysical values (negative melt rates), we only consider days with net melting conditions, i.e., positive daily average snowmelt fluxes. Thus, the daily averaged melt rate M_n at day n (lasting from t_n to t_{n+1}) is given by

$$M_n = \max\left(\frac{1}{\rho_w L_f \Delta t}\int_{t_n}^{t_{n+1}} Q_M(t)\,\mathrm{d}t,\, 0\right), \tag{8}$$

where ρ_w is the density of freshwater, L_f is the latent heat of fusion and Δt is the daily time step. We emphasize that the effects of refreezing are still considered at a subdaily time resolution in Eq. (8). Similarly, the daily averaged precipitation rate is

$$P_n = \frac{1}{\Delta t}\int_{t_n}^{t_{n+1}} P(t)\,\mathrm{d}t\,. \tag{9}$$

Now the daily averaged net accumulation rate can be obtained through

$$\mathcal{A}_n = P_n - M_n\,, \tag{10}$$

and the accumulated melt depth D_m is accounted for through

$$D_{m,n+1} = \max\left(D_{m,n} - \mathcal{A}_n \Delta t, 0\right) H(\mu)\,. \tag{11}$$

The peak mean SWE μ is updated via

$$\Lambda = \mu_n + \max\left(\mathcal{A}_n \Delta t - D_{m,n+1}, 0\right) \tag{12}$$

through

$$\mu_{n+1} = \Lambda H(\Lambda - \tau_S)\,, \tag{13}$$

where the alternative Heaviside function is defined through $H(x) = 0$ if $x \leq 0$ and $H(x) = 1$ otherwise. Consequently, in Eq. (13) the peak mean SWE μ is only nonzero if Λ exceeds the threshold τ_S. Note that the formulation in Eq. (11) gradually resets the melt depth towards zero in the case of snowfall after the onset of melt, following Liston (2004). This means that fSCA is not reset to unity in the case of new snowfall after a melting period unless the new snowfall leads to an increase in the peak SWE. In the study area, snowfall events occurred rarely during the snowmelt period, and the new snow cover lasted only a short time. At sites where such events are more frequent, Durand et al. (2008) presents an alternative solution, albeit at an increased computational cost. The annual model integrations start in the beginning of September, when the surface is assumed to be snow free, so that both μ and D_m are initialized as zero. Both μ and D_m are reset to zero following the complete disappearance of the snowpack, defined as when the fSCA decreases below 0.01 to account for the infinite tail of f_P. The model resolution is defined by the footprint of (area encompassed by) the snow surveys for each site (see Table 1 and Fig. 1).

3.1.3 Forcing

Forcing terms in the form of precipitation, air temperature, relative humidity and wind speed, as well as downwelling longwave and shortwave radiation, are required to diagnose the mass and energy balance in Eq. (3) and Eq. (4). These terms are obtained by downscaling ERA-Interim reanalysis data (Dee et al., 2011) at 0.75° resolution following Østby et al. (2017). This method uses the linear theory of orographic precipitation in Smith and Barstad (2004) to downscale precipitation and a modification of the TopoSCALE approach (Fiddes and Gruber, 2014) for the remaining fields. The reanalysis forcing is downscaled onto 1 km resolution digital elevation model (DEM) grid cells centered on each of the study sites. The downscaling is performed based on the mean physiographic conditions (elevation, slope and aspect) within each of these grid cells. The resulting values at 1 km spatial and 6-hourly temporal resolution are linearly interpolated in time to facilitate a stable computation of the time evolution of turbulent energy fluxes following Westermann et al. (2016). From these fluxes, and the remaining surface energy balance fields, diurnally averaged melt rates are calculated. Similarly, diurnally averaged rainfall and snowfall rates are computed by delineating between rain and snow in the time-interpolated precipitation rate (You et al., 2014) and then taking diurnal averages. While the resolution of the downscaled forcing data does not exactly match the model resolution (i.e., the footprint of the snow surveys, Sect. 3.1.2), the mismatch is small considering the gentle topography of the study sites (Sect. 2.1).

3.2 Satellite retrievals

We make use of satellite retrievals between May and September, which contain the snowmelt period for all the investigated years. Only retrievals that fall inside the melt season are assimilated as these contain information about the snow cover depletion. Due to frequent cloud cover, the effective revisit frequency of fSCA retrievals is irregular, with prolonged data gaps occurring regularly. An overview of the number of available scenes is given in Table 3.

Table 3. Number of MODIS and Sentinel-2 scenes per melt season with field measurements available for the three study sites.

Location	Melt season	No. of MODIS scenes	No. of Sentinel-2 scenes
Bayelva	2008, 2009, 2012, 2013, 2014, 2015, 2016	8, 9, 8, 9, 6, 14, 11	–, –, –, –, –, –, 7
Steinflåen plateau	2016	5	8
Kvadehuksletta	2016	11	7

Table 4. Overview of the distributions from which the prior ensemble of perturbation parameters are independently drawn.

Symbol	Name	Distribution	Support	Mean	Variance	Units
χ	Coefficient of variation	Logit-normal	$(0, 0.8)$	0.4	0.01	–
Q_0	Initial ground heat flux	Logit-normal	$(0, 40)$	20	20	$\mathrm{W\,m^{-2}}$
$\alpha_{\min}$	Minimum snow albedo	Logit-normal	$(0.45, 0.55)$	0.5	0.02	–
b_{P}	Precipitation bias	Lognormal	$(0, \infty)$	1	0.04	–
b_{M}	Melt bias	Lognormal	$(0, \infty)$	1	0.01	–

3.2.1 MODIS

We employ version 6 of the Level 3 daily 500 m resolution fSCA retrievals from MODIS on board the satellites Terra (MOD10A1 product; Hall and Riggs, 2016a) and Aqua (MYD10A1 product; Hall and Riggs, 2016b). The retrieval algorithm is based on a linear fit of the normalized difference snow index (NDSI) measured by MODIS to fSCA retrievals from ground truth Landsat scenes as described in Salomonson and Appel (2004). The NDSI exploits the fact that snow is highly reflective in the visible but a good absorber in the shortwave infrared, which sets it apart from other natural surfaces such as clouds, vegetation and soil (Painter et al., 2009).

We average over all the pixels for each day and study site (see Fig. 1). This average is only taken if cloud-free (as determined by the MODIS cloud mask) retrievals are available for each of these pixels. If both Terra and Aqua retrievals are available for a given day, only the former are used. Despite small deviations in the measurement footprint (see Fig. 1), we compare MODIS fSCA retrievals to the field measurements of fSCA obtained from the automatic camera system, UAV and GPS surveys (Sect. 2.2). From this comparison, we estimate a RMSE of $\sigma_{\mathrm{MOD}} = 0.13$ for the MODIS fSCA. We use σ^2_{MOD} as the observation error variance in the corresponding diagonal entries of the observation error covariance matrix (Sect. 3.3.2).

3.2.2 Sentinel-2

For the year 2016, we complement the MODIS fSCA retrievals with aggregated 20 m resolution retrievals from the Sentinel-2A mission (Drusch et al., 2012). fSCA estimates are derived from the Level 1C orthorectified top of the atmosphere reflectance product, with cloud-free scenes manually selected. For this purpose, NDSI is computed from reflectances (r) from a visible (b3, centered on 0.56 µm) and a shortwave infrared band (b11, centered on 1.61 µm) through

$$\mathrm{NDSI}_{\mathrm{S2}} = \frac{r_{\mathrm{b3}} - r_{\mathrm{b11}}}{r_{\mathrm{b3}} + r_{\mathrm{b11}}}. \tag{14}$$

Each pixel is then classified as either snow covered ($\mathrm{NDSI} \geq 0.4$) or snow free ($\mathrm{NDSI} < 0.4$), where the NDSI threshold was chosen in line with Hall et al. (2002). The binary (snow/no snow) pixels are then aggregated to the approximate footprint of the independent snow surveys conducted at each site (Fig. 1) to obtain Sentinel-2-derived fSCA estimates. Therefore, the areal extent of the Sentinel-2 fSCA retrievals closely matches the areas of the corresponding study sites given in Table 1. The retrieval process is illustrated schematically in Fig. 2. By comparing the Sentinel-2 retrievals to the field measurements of fSCA from the automatic camera system in 2016, we estimate a RMSE of $\sigma_{\mathrm{S2}} = 0.09$. We use σ^2_{S2} as the observation error variance in the corresponding diagonal entries of the observation error covariance matrix (Sect. 3.3.2).

3.3 Ensemble data assimilation

In this section we outline how the prior ensemble of model realizations is set up and how it is updated to a posterior ensemble through the assimilation of fSCA satellite retrievals using ensemble-based batch smoother schemes.

3.3.1 Ensemble generation

The prior ensemble of model realizations is generated by independently drawing perturbation parameter values from the distributions listed in Table 4. These perturbation parameters are held constant throughout the annual integration of the model. Two of these are multiplicative bias parameters that perturb the mass balance through the net accumulation rate,

$$\mathcal{A}_{n,j} = b_{P,j} P_n - b_{M,j} M_{n,j},$$

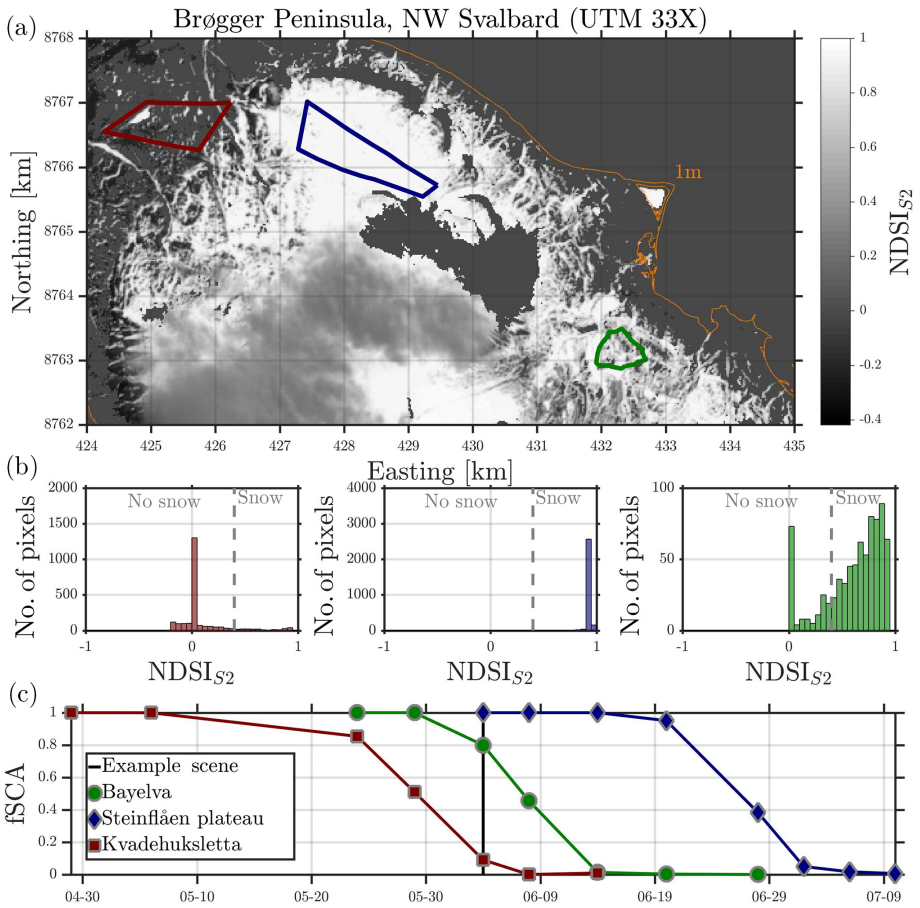


Figure 2. (a) Sentinel-2 NDSI estimates from an example scene (4 June 2016) over the Brøgger Peninsula with Kvadehuksletta, Steinflåen plateau and Bayelva marked with red, green and blue polygons, respectively; coastline in orange. **(b)** NDSI histograms of the same example scene (Kvadehuksletta: left; Steinflåen plateau: middle; Bayelva: right) with the threshold at NDSI= 0.4 marked. **(c)** Time series of Sentinel-2 NDSI-based fSCA retrievals for the 2016 melt season.

for $j \in 1 : N_e$, where N_e is the number of ensemble members. We inherently assume the model forcing to be the major source of uncertainty (De Lannoy et al., 2010; Raleigh et al., 2015). Furthermore, we assume that the error in the forcing can be modeled through constant multiplicative biases (fixed throughout the annual integration) in the mass balance. Consequently, the bias parameters are modeled as positive definite lognormal random variables. This is in line with the perturbations in Girotto et al. (2014b) on the precipitation rate, but we also perturb the melt rate. Moreover, we assume that the ensemble of net accumulation rates is on average unbiased due to the applied downscaling method (Østby et al., 2017) and thus assign the two bias parameters a mean of unity. The precipitation rates are also perturbed by the same bias parameter in the computation of the heat advected by precipitation (Q_P) in the surface energy balance that contributes to the melt rate M_n.

In addition to the mass balance forcing, the peak subgrid coefficient of variation χ (Sect. 3.1.1) is a source of uncertainty. We assume a prior mean of 0.4 for χ, which corresponds to the value provided by Liston (2004) for "Arctic tundra". Moreover, χ is assumed to be double bounded between 0 and 0.8, with negative values being unphysical and the upper bound close to the maximum value in Liston (2004). Furthermore, both the initial ground heat flux at the onset of melt (Q_0) and the minimum snow albedo ($\alpha_{\min}$) are uncertain, and we also assume that these are double bounded.

The probability distributions of double-bounded random variables are modeled as logit-normal distributions, with the logit transform for a variable x bounded between a and b given by

$$\widetilde{x} = \mathrm{logit}_{(a,b)}(x) = \ln\left(\frac{x-a}{b-a}\right) - \ln\left(1 - \frac{x-a}{b-a}\right), \quad (15)$$

while the inverse transform is given by

$$x = \mathrm{logit}^{-1}_{(a,b)}(\widetilde{x}) = a + (b-a)\,/\left(1 + e^{-\widetilde{x}}\right). \quad (16)$$

To generate a prior ensemble of a logit-normally distributed random variable, we first apply the logit transform to the

mean. Then, we add N_e realizations of Gaussian white noise with a consistent variance to the transformed mean and subsequently apply the inverse transform. We emphasize that through the perturbation parameters we effectively perturb the melt rate, precipitation rate and coefficient of variation. By performing a subsequent ensemble integration of the SSM, we also get an ensemble of state variables that are consistent with the prior perturbation parameter ensemble.

3.3.2 Batch smoothers

Here, we describe our implementation of three batch smoother schemes: the ES-MDA, the ES and the PBS. The ES-MDA is our focus, while the two latter schemes are used for comparison. In a batch smoother all the observations, in this case all fSCA retrievals from the snow cover depletion during one melt season, are assimilated at once in a single batch (Dunne and Entekhabi, 2005), as opposed to sequentially as in a filter (Bertino et al., 2003). We follow the conventional notation in the DA literature, as laid out in Ide et al. (1997). Let N_e, N_o, N_a, N_s, N_p and N_t denote the number of ensemble members, observations, assimilation cycles, state variables, perturbed parameters and time steps, respectively, during an annual (September–August) model integration. $\mathbf{X}$ is the $(N_s \times N_t) \times N_e$ matrix containing the ensemble of states ($\mathrm{fSCA}_{n,j}$, $D_{m,n,j}$ and $\mu_{n,j}$), and $\mathbf{\Theta}$ is the $N_p \times N_e$ matrix containing the ensemble of perturbation parameters listed in Table 4. The $N_o \times 1$ observation vector $\boldsymbol{y}$ contains all the fSCA satellite retrievals during the ablation season (Sect. 3.2), $\mathbf{Y}$ is the $N_o \times N_e$ matrix containing the ensemble of perturbed fSCA satellite retrievals and $\widehat{\mathbf{Y}}$ is the $N_o \times N_e$ matrix containing the ensemble of predicted fSCA observations. $\mathbf{H}$ is the linear observation operator, which is a binary $N_o \times (N_s \times N_t)$ matrix that picks out the predicted fSCA observations from the ensemble of states, and $\mathbf{R}$ is the $N_o \times N_o$ diagonal observation error covariance matrix containing the observation error variances (Sect. 3.2).

The ES-MDA (Emerick and Reynolds, 2013) is an iterative scheme, requiring multiple ensemble model integrations and analysis steps. Collecting the perturbed and predicted observations during the ensemble integration into a batch and performing the analysis step is referred to as one assimilation cycle, and we will let the current iteration number be denoted as ℓ. In such a case, the ES-MDA scheme is set up as follows, for $\ell \in 0 : N_a$ iterations:

1. Run an ensemble model integration; i.e., for $n \in 0 : (N_t - 1)$ time steps compute

$$\mathbf{X}_{n+1}^{(\ell)} = \mathcal{M}\left(\mathbf{X}_n^{(\ell)}, \mathbf{\Theta}^{(\ell)}\right), \tag{17}$$

   where $\mathcal{M}$ is the SSM operator defined through equations Eq. (1), Eq. (11) and Eq. (13).

2. If $\ell < N_a$ (*otherwise stop the algorithm here*), collect the batch of predicted observations,

$$\widehat{\mathbf{Y}}^{(\ell)} = \mathbf{H}\mathbf{X}^{(\ell)}, \tag{18}$$

   and perturbed observations,

$$\mathbf{Y}^{(\ell)} = \mathbf{y} \otimes \mathbf{1}^T + \sqrt{\alpha^{(\ell)}}\mathbf{R}^{1/2}\boldsymbol{\epsilon}^{(\ell)}, \tag{19}$$

   where $\otimes$ is the outer product, 1 is an $N_o \times 1$ vector of ones, the $^{\mathrm{T}}$ superscript denotes the transpose, $\alpha^{(\ell)}$ is the observation error inflation coefficient and $\boldsymbol{\epsilon}^{(\ell)}$ is a $N_o \times N_e$ matrix containing zero mean Gaussian white noise with a variance of 1.

3. Transform the perturbation parameters using analytical Gaussian anamorphosis functions ψ (Bertino et al., 2003):

$$\widetilde{\mathbf{\Theta}}^{(\ell)} = \psi\left(\mathbf{\Theta}^{(\ell)}\right). \tag{20}$$

   ψ is the natural logarithm and the logit for the biases and the remaining perturbation parameters, respectively.

4. Perform the Kalman-like analysis step in the transformed space:

$$\widetilde{\mathbf{\Theta}}^{(\ell+1)} = \widetilde{\mathbf{\Theta}}^{(\ell)} + \mathcal{C}_{\widetilde{\mathbf{\Theta}}\widehat{\mathbf{Y}}}^{(\ell)}\left(\mathcal{C}_{\widehat{\mathbf{Y}}\widehat{\mathbf{Y}}}^{(\ell)} + \alpha^{(\ell)}\mathbf{R}\right)^{-1}\left(\mathbf{Y}^{(\ell)} - \widehat{\mathbf{Y}}^{(\ell)}\right). \tag{21}$$

   The transformed perturbation parameter-predicted observation and predicted observation error covariance matrices are

$$\mathcal{C}_{\widetilde{\mathbf{\Theta}}\widehat{\mathbf{Y}}}^{(\ell)} = \frac{1}{N_e}\widetilde{\mathbf{\Theta}}^{(\ell)'}\widehat{\mathbf{Y}}^{(\ell)'\mathrm{T}} \tag{22}$$

   and

$$\mathcal{C}_{\widehat{\mathbf{Y}}\widehat{\mathbf{Y}}}^{(\ell)} = \frac{1}{N_e}\widehat{\mathbf{Y}}^{(\ell)'}\widehat{\mathbf{Y}}^{(\ell)'\mathrm{T}}, \tag{23}$$

   respectively, in which primes ($'$) denote anomalies (deviations from the ensemble mean).

5. Apply the appropriate inverse transforms to recover the updated perturbation parameters:

$$\mathbf{\Theta}^{(\ell+1)} = \psi^{-1}\left(\widetilde{\mathbf{\Theta}}^{(\ell+1)}\right). \tag{24}$$

   ψ^{-1} is the exponential and the inverse logit for the biases and the remaining perturbation parameters, respectively.

The observation error inflation coefficient $\alpha^{(\ell)}$ in Eq. (21) together with the iterations sets the ES-MDA apart from the traditional ES (Van Leeuwen and Evensen, 1996). For $N_a = \alpha^{(\ell)} = 1$, the ES scheme, which was used in the probabilistic SWE reconstruction of Durand et al. (2008) and

Girotto et al. (2014b), is recovered. The idea behind the ES-MDA is to perform multiple smaller analysis steps as opposed to one abrupt analysis step. In the case of a nonlinear model, this is expected to yield a better approximation of the true posterior (Emerick and Reynolds, 2013). A requirement for the ES-MDA to give a nearly unbiased estimate (cf. Stordal and Elsheikh, 2015) is that the coefficients satisfy $\sum_{\ell=0}^{N_a-1} 1/\alpha^{(\ell)} = 1$. In our case this is accomplished by setting all the coefficients to $\alpha^{(\ell)} = N_a$ and specifying N_a before any assimilation cycles are carried out. We emphasize that the analysis step (Eq. 21) only updates the perturbation parameters and a consistent ensemble of states is found from the subsequent ensemble model integration. The model constants listed in Table 2 remain unchanged by the analysis and the integration. As mentioned, the perturbation parameter matrix $\widetilde{\boldsymbol{\Theta}}$ in Eq. (21) is transformed through analytical Gaussian anamorphosis (Bertino et al., 2003) to ensure that the priors are Gaussian. In this case, the Kalman-like analysis step (Eq. 21) is variance minimizing for a linear model (Van Leeuwen and Evensen, 1996). The entire methodology, with the ES-MDA as the DA scheme, is depicted in Fig. 3.

Margulis et al. (2015) introduced the PBS for snow data assimilation. In this scheme, each particle (i.e., ensemble member; Van Leeuwen, 2009) is given an equal prior weight of $1/N_e$. Then, after an ensemble model integration, the normalized posterior importance weights $w_j \in [0, 1]$ are diagnosed through the analysis step

$$w_j = p\left(\boldsymbol{y}|\widehat{\mathbf{X}}_j\right) p\left(\widehat{\mathbf{X}}_j\right) / \sum_{j=1}^{N_e} \left(p\left(\boldsymbol{y}|\widehat{\mathbf{X}}_j\right) p\left(\widehat{\mathbf{X}}_j\right)\right), \tag{25}$$

where $\widehat{\mathbf{X}}_j = [\mathbf{X}_j\ ;\ \boldsymbol{\Theta}_j]$ is the augmented state vector for the j-th particle and the Gaussian likelihoods are given by

$$p\left(\boldsymbol{y}|\widehat{\mathbf{X}}_j\right) = c_0 \exp\left[-0.5\left(\boldsymbol{y} - \widehat{\mathbf{Y}}_j\right)^{\mathrm{T}} \mathbf{R}^{-1} \left(\boldsymbol{y} - \widehat{\mathbf{Y}}_j\right)\right]. \tag{26}$$

This is a direct application of Bayes' rule in which the normalizing denominator has two important consequences. Firstly, $c_0 = 1/\sqrt{(2\pi)^{N_o}|\mathbf{R}|}$ cancels out, thus avoiding errors introduced through floating-point arithmetic ($(2\pi)^{N_o}$ is generally large). Secondly, the prior weights $p(\widehat{\mathbf{X}}_j)$ also cancel as they are equal for all particles. With Gaussian likelihoods, Eq. (25) becomes

$$w_j = \frac{\exp\left[-0.5\left(\boldsymbol{y} - \widehat{\mathbf{Y}}_j\right)^{\mathrm{T}} \mathbf{R}^{-1} \left(\boldsymbol{y} - \widehat{\mathbf{Y}}_j\right)\right]}{\sum_{j=1}^{N_e} \exp\left[-0.5\left(\boldsymbol{y} - \widehat{\mathbf{Y}}_j\right)^{\mathrm{T}} \mathbf{R}^{-1} \left(\boldsymbol{y} - \widehat{\mathbf{Y}}_j\right)\right]}, \tag{27}$$

where the posterior weights w_j sum to unity. The posterior ensemble still spans the range of the prior ensemble, as the analysis step only changes the relative weights of the ensemble members and not their position within the state and perturbation parameter space. Marginal cumulative distributions are recovered through the individual ranking of the ensembles of state variables and perturbation parameters followed by a cumulative summation of the correspondingly sorted weights. These distributions allow for the estimation of quantile values. Note that the PBS is equivalent to running a particle filter without re-sampling and using the prior as the importance density (see Van Leeuwen, 2009). As such, the PBS corresponds to the generalized likelihood uncertainty estimation method (GLUE; Beven and Binley, 1992) with a Gaussian likelihood function. Due to the absence of re-sampling, even for medium-dimensional systems with a large number of observations to be assimilated, the PBS can become degenerate with very few particles carrying the majority of the importance weights (Van Leeuwen, 2009). Nevertheless, a major advantage of the PBS is its computational efficiency, requiring only one ensemble model integration and one efficient analysis step (Eq. 27). In this study, the PBS and the ES are used to benchmark the ES-MDA.

4 Results

4.1 Interannual variability and comparison to field measurements

In this section, we present results of the ES-MDA scheme with 100 ensemble members and four assimilation cycles (Sect. 3.3.2) for all the years and sites where snow surveys were conducted. Figure 4 shows the time series of the prior and posterior fSCA (panel a) and mean SWE (panel b) estimates, as well as the assimilated fSCA satellite retrievals (panel a) and the independently observed peak mean SWE (from the snow surveys, panel b). The assimilation generally brings the posterior estimates closer to the observed fSCA and considerably constrains the spread of the ensemble compared to the prior. For some occasions – such as Bayelva in 2008, Bayelva in 2009 and Kvadehuksletta in 2016 – the timing of the snowmelt in the prior is significantly biased by as much as 3 weeks compared to the assimilated fSCA retrievals. Even if the prior ensemble does not encompass the retrievals, the iterative ES-MDA scheme allows the posterior to converge towards the fSCA observations (panel a), leading to much improved SWE estimates (panel b). On other occasions, such as 2015 at Bayelva and 2016 at Steinflåen plateau, the prior ensemble is a reasonable estimate and the assimilation merely constrains the spread of the ensemble and adjusts the median slightly. Both for Bayelva in 2015 and Kvadehuksletta in 2016, some of the early fSCA retrievals, which indicate a slight ablation, are completely ignored by the assimilation, as this early onset of melt is inconsistent with the model – even when biases are accounted for. However, this ablation could be real and due to processes not accounted for in the model, such as wind erosion.

Field measurements of peak mean SWE are available for years with low (2008, 2016), medium (2013, 2015) and high (2009, 2014) values of peak mean SWE, ranging from 0.08 m water equivalent (w.e.) (Kvadehuksletta in 2016) to

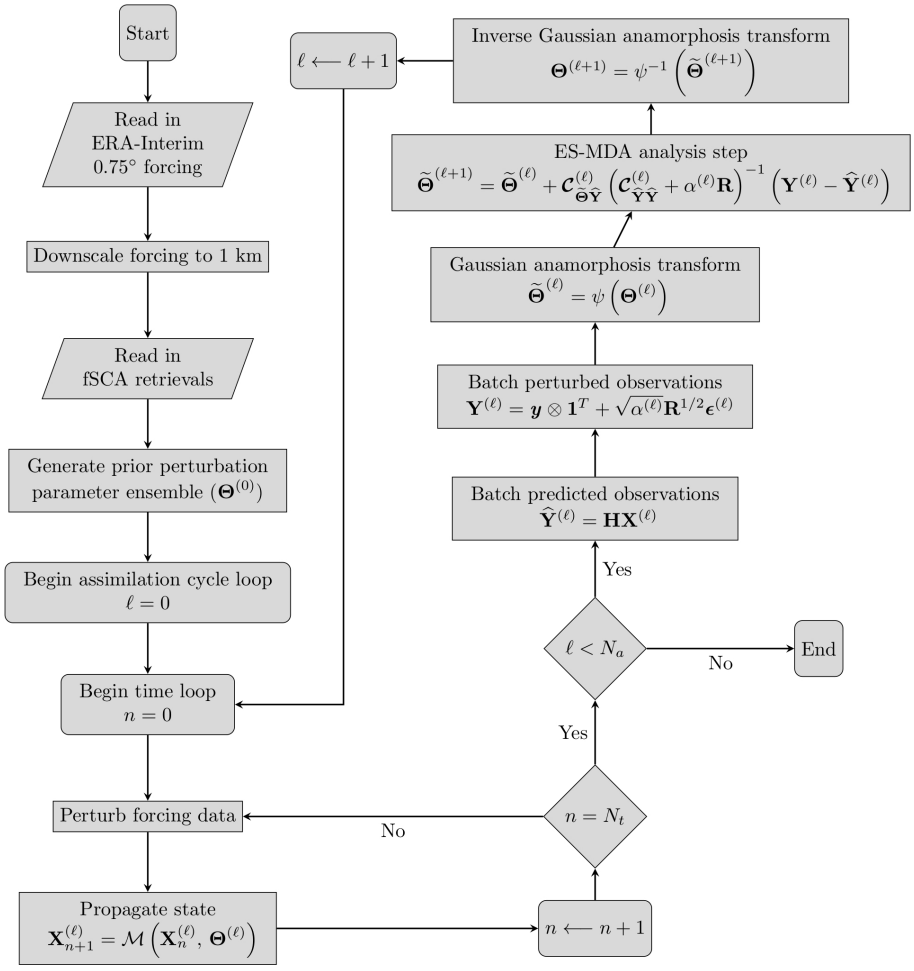


Figure 3. Flowchart depicting the methodology with the ES-MDA as the DA scheme. Symbols are defined in the text.

0.48 m w.e. (Bayelva in 2014). With the exception of two cases (Bayelva in 2013 and Steinflåen plateau in 2016), the assimilation brings the ensemble median closer to the observed peak mean SWE, while at the same time constraining the spread of the ensemble. We emphasize that the assimilation performs a global bias correction for peak SWE. This is especially evident for Kvadehuksletta in 2016, for which the assimilation unrealistically truncates the duration of the snow season as a result of a strong correction for the positive bias. Both in 2008 and 2009 for Bayelva, the ES-MDA shifts the estimates to better match field measurements of SWE (which were not assimilated), despite the prior range being far from the observations. The posterior ensemble median peak mean SWE is generally close to the independently observed peak mean SWE, but absolute relative differences up to 40 % (minimum: 0.5 %; mean: 19 %) occur.

Figure 5 displays the prior, posterior and observed peak SSDs for the years and sites with field measurements. Again, with the exception of Bayelva in 2013 and Steinflåen plateau in 2016, the assimilation brings the mean of the peak SSD closer to the observations. The agreement between the posterior and observed mean value is striking for a number of years and sites, such as Bayelva in 2009 and 2014 as well as Kvadehuksletta in 2016. Furthermore, the shapes of the observed and posterior distributions agree well, for example, for Bayelva in 2008, 2013 and 2016. Once more, the correction from prior to posterior is largest for Bayelva in 2008 and 2009, for which the prior fSCA was furthest from the satellite retrievals. The prior ensemble SSD, apart from Bayelva in 2013, is generally too positively skewed (i.e., has a long tail) compared to the observed SSD. On some occasions the match between the posterior and observed SSDs is poor, such as Steinflåen plateau in 2016 and Bayelva in 2015. We conclude that the analysis typically improves the fit between modeled and observed snow distributions. Some of the observed distributions, such as that for Kvadehuksletta in 2016, are hard to match as they do not conform well to a lognormal distribution, possibly due to the limited number of sample points (Sect. 2.2).

The posterior bias parameters can be directly evaluated by comparing the bias-corrected forcing to field measurements. Due to a lack of snowfall observations (see Boike et al.,

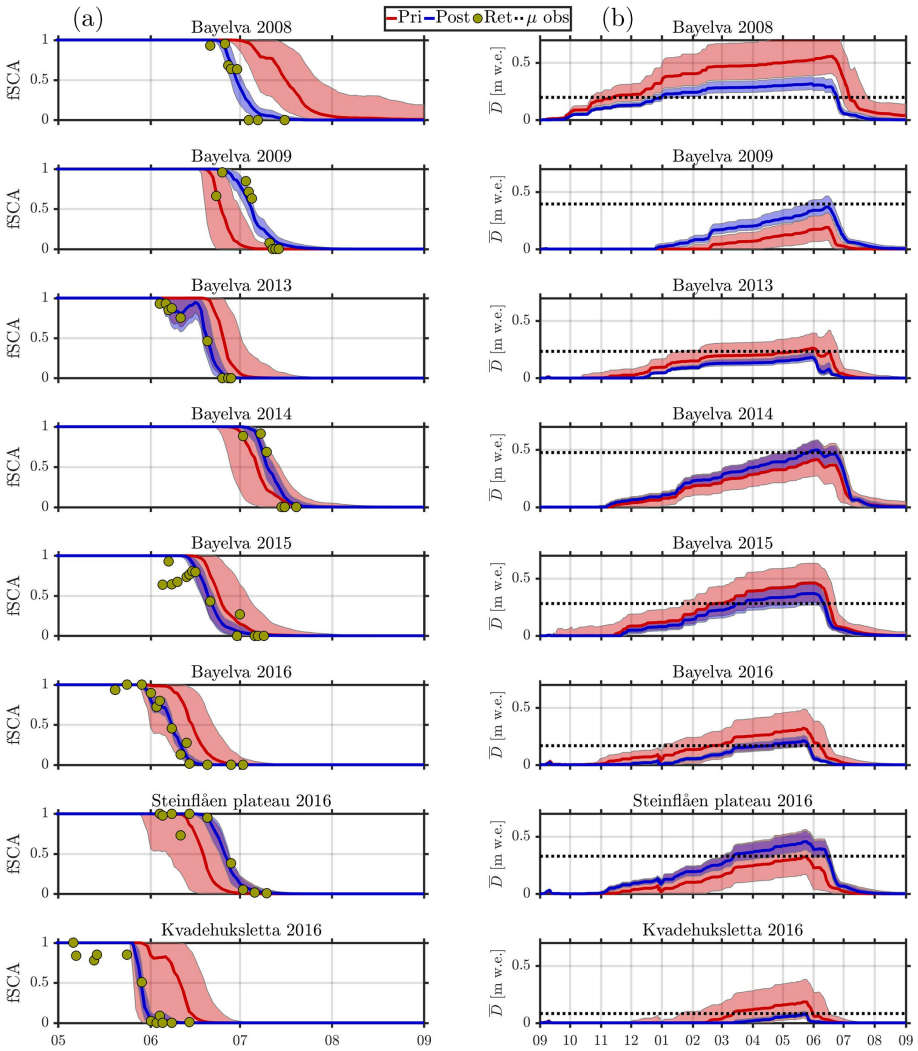


Figure 4. Time evolution of the prior (red) and ES-MDA ($N_a = 4$, $N_e = 10^2$) posterior (blue) fSCA (column **a**) and mean SWE ($\overline{D}$; column **b**); shading: 90th-percentile range; solid lines: ensemble median; yellow dots: assimilated MODIS and Sentinel-2 fSCA retrievals; dotted black line: independently observed peak mean SWE (μ) from snow surveys (Sect. 2.2); x axis: months. These results are from a single run.

2017), a direct evaluation of the precipitation bias parameter is not possible. However, the melt bias parameter can be evaluated by comparing the estimated snowmelt flux (which is directly proportional to the perturbed melt depth) to field-based values. For June 2008, Westermann et al. (2009) estimate an average snowmelt flux of $27\,\mathrm{Wm}^{-2}$, which compares well to the ES-MDA posterior median (averaged for the same period) of $29\,\mathrm{Wm}^{-2}$, while the prior median is too low at $19\,\mathrm{Wm}^{-2}$.

In ensemble-based data assimilation the spread of the posterior ensemble should represent the uncertainty. To verify this, one can compare two metrics: the residual, i.e., the instantaneous posterior RMSE of the ensemble relative to the corresponding independent field measurement, and the ensemble standard deviation (e.g., Evensen, 2009). For this comparison we define the relative residual as the ratio of the residual to the standard deviation. Ideally this ratio should have a value of 1, which indicates that the two metrics are equal, so that the posterior ensemble spread accurately captures the estimation uncertainty. For the fSCA, peak mean SWE and peak subgrid coefficient of variation, the average (over all available field measurements) relative residuals were 2.22, 1.53 and 1.66, respectively, so the posterior ensemble underestimates the uncertainty. This effect has been extensively described by Evensen (2009); it arises in part because of model structural errors related to neglected physical processes (Sect. 3.1). Still, the assimilation is generally able to simultaneously (but not to the same extent) reduce the spread and the error in the ensemble (Fig. 4).

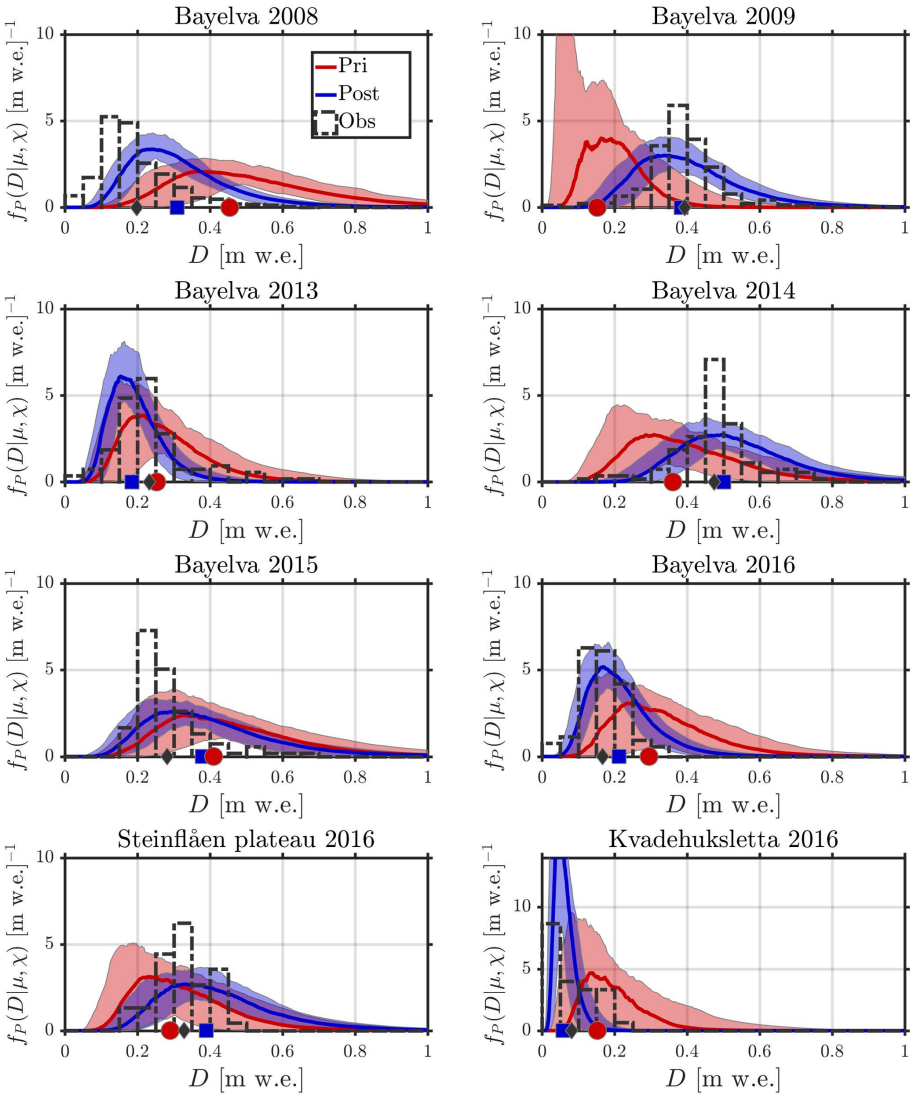


Figure 5. Prior (red), ES-MDA ($N_a = 4$, $N_e = 10^2$) posterior (blue) and the corresponding independently observed (from snow surveys; dashed black) peak subgrid SWE distributions; shaded areas: 90th-percentile range; solid lines: ensemble median; markers: mean value. These results are from a single run.

4.2 Evaluation of data assimilation schemes

In addition to the ES-MDA scheme, we evaluate the PBS and ES (Sect. 3.3.2) with regards to field measurements, using an ensemble size of 100 members for all schemes. Three error metrics are summarized in Table 5: the bias (mean error), RMSE and square correlation coefficient. For the fSCA, all the schemes achieve a major improvement relative to the prior, with an almost 10-fold reduction in bias, a halving of RMSE and an almost perfect correlation to the field measurements of fSCA (Sect. 2.2). For the peak mean SWE (μ), the PBS performs best in terms of RMSE and bias, followed closely by the ES-MDA, which, in turn, has the highest correlation coefficient. With regards to the peak subgrid coefficient of variation (χ), the ES-MDA performs best across all the metrics, tying with ES for (absolute) bias and the PBS for RMSE. As considerably more field measurements are available for fSCA than for μ and χ, the evaluation for fSCA must be considered more robust. The scatterplots in Fig. 6 visualize the performance of the prior and all the considered DA schemes relative to the field measurements.

Observed, prior and posterior peak mean SWE and peak subgrid coefficient of variation for different years/sites are shown in Fig. 7. As discussed in Sect. 1, the assimilation moves the posterior peak mean SWE estimates closer to the observed peak mean SWE in most cases when compared to the prior. However, clear performance differences are found between the different schemes for a number of situations. In 2008, the PBS is not able to correct for as much of the bias in the peak mean SWE as the ES-MDA and the ES. For the

Table 5. Summary of evaluation metrics – i.e., bias, RMSE and square correlation coefficient (R^2) – for the fSCA, peak SWE (μ) and peak subgrid coefficient of variation (χ). These metrics are based on comparisons to all the field measurements presented in Sect. 2.2 with the number of observations for the comparisons in brackets next to the corresponding symbols. All the metrics are averaged over 10^2 independent runs, each with 100 ensemble members. The ES-MDA was run with $N_a = 4$ assimilation cycles.

	Prior			ES-MDA			ES			PBS		
Symbol	Bias	RMSE	R^2	Bias	RMSE	R^2	Bias	RMSE	R^2	Bias	RMSE	R^2
fSCA (106)	0.21	0.02	0.80	0.03	0.01	0.97	0.03	0.01	0.97	0.03	0.01	0.97
μ (8) [m w.e.]	0.13	0.21	0.01	0.06	0.09	0.77	0.06	0.12	0.76	0.06	0.08	0.76
χ (8)	0.01	0.16	0.00	−0.02	0.13	0.37	0.02	0.14	0.33	−0.03	0.13	0.06

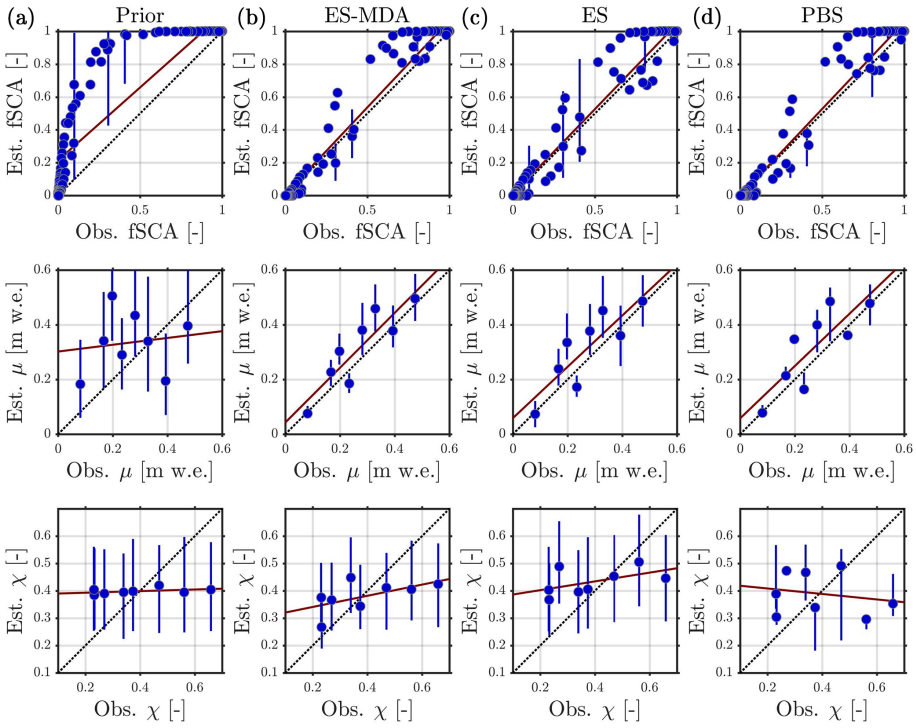


Figure 6. Scatterplots of the prior median (column **a**) as well as ES-MDA ($N_a = 4$; column **b**), ES (column **c**) and PBS (column **d**) posterior median estimates of fSCA (top row), peak mean SWE μ (middle row) and peak subgrid coefficient of variation χ (bottom row) versus the observations (field measurements); error bars: 90th-percentile range; all DA schemes were run with 10^2 ensemble members; one to one: dotted black line; linear best fit: solid red line. These results are from a single run.

remaining years, the performance of the schemes in terms of estimating peak mean SWE is similar, but the spread of the ES is by far the largest, followed by the PBS and the ES-MDA. The PBS ensemble shows indications of degeneracy for some years (e.g., 2008 and 2009) where the median coincides with either the 5th- or 95th-percentile value. This indicates that the majority of the weight in the PBS is carried by just a few ensemble members. For the coefficient of variation, the 90th-percentile range of the ES-MDA posterior ensemble typically encompassed the observed value (with two exceptions), while this was not true to the same extent for the ES (three exceptions) and the PBS (five exceptions). These performance differences explain the higher correlation coefficient for the coefficient of variation for the ES-MDA scheme (Table 5). The PBS also shows signs of degeneracy (e.g., Bayelva in 2009) for the coefficient-of-variation estimation. On some occasions (e.g., Bayelva in 2008, 2009 and 2014), the posterior ensemble median is effectively pulled closer to the observed coefficient of variation when compared to the prior. On the same occasions the ensemble spread is slightly constrained. Compared to the peak mean SWE, it is much harder to constrain estimates of the coefficient of variation regardless of scheme, although it is possible to shift the ensemble in the right direction.

We gauged the sensitivity of the three batch smoother schemes with respect to ensemble size and the number of as-

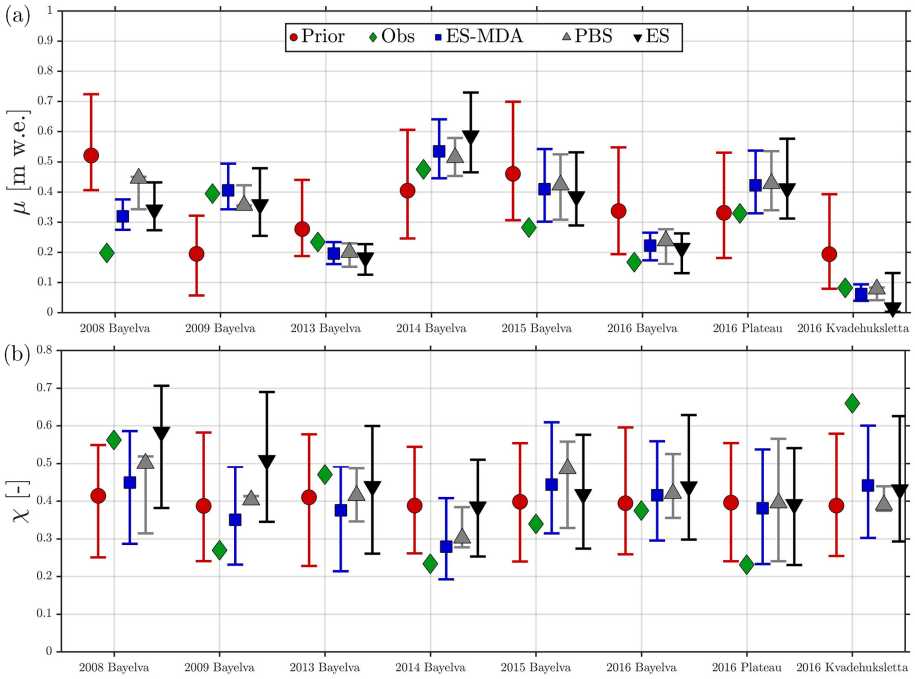


Figure 7. Prior median, observed, ES-MDA (with $N_a = 4$), PBS and ES posterior median peak mean SWE μ **(a)** and peak subgrid coefficient of variation χ **(b)**; error bars: 90th-percentile range; all DA schemes were run with 10^2 ensemble members. These results are from a single run.

similation cycles by considering the fractional improvement (FI) in RMSE that was achieved through the analysis step based on all available field measurements (Fig. 8). On the one hand, the PBS requires an ensemble size of 1000 to converge to a stable FI of around 75, 20 and 60 % for the fSCA, peak subgrid coefficient of variation and peak mean SWE, respectively. On the other hand, the ES-MDA with four assimilation cycles converges with just 100 ensemble members at similar FIs to the PBS. The ES performs worst regardless of ensemble size, with FIs of around 70, 10 and 55 % even with 10^5 ensemble members, requiring 100 ensemble members for convergent results. For all schemes the available validation data suggest that the greatest improvements are achieved for fSCA, followed by peak mean SWE, while by far the lowest improvements are found for the peak subgrid coefficient of variation. With 100 ensemble members, the ES-MDA converges to a stable performance at four assimilation cycles; i.e., there is no marked increase in FI for more cycles (Fig. 8, bottom right).

4.3 Effects of observation error and assimilation frequency

The effects of observation error and assimilation frequency are studied by running the ES-MDA ($N_e = 10^2$, $N_a = 4$) and assimilating first only MODIS and then both MODIS and Sentinel-2 retrievals for the 2016 snow season at all study sites. As discussed in Sect. 3.2, the Sentinel-2 fSCA retrievals are based on higher-resolution optical reflectance retrievals. As such they are expected to be less prone to representativeness error and thus observation error since the area in which the snow surveys were conducted is more accurately covered by the retrievals. Furthermore, the Sentinel-2 scenes used for fSCA retrievals were manually checked to be cloud free, which was not the case for the MODIS scenes. Table 6 summarizes various performance metrics for the two different runs. For the peak mean SWE depth (μ), there is no difference when including Sentinel-2 fSCA retrievals in the assimilation. For the coefficient of variation (χ), however, there is an increase in FI for both the bias and the RMSE, as well as an increase in the square correlation coefficient. Sentinel-2 fSCA retrievals with lower observation error help to further constrain the shape of the snow depletion curve which explains the improvement in the χ estimation. We emphasize that this evaluation is based on the only three available field measurements of μ and χ in 2016 (from the snow surveys), so that these preliminary results need to be consolidated by future studies with more validation data.

5 Discussion

5.1 Interannual variability and comparison to field measurements

For all considered years and sites, the ES-MDA scheme both brings the ensemble median fSCA closer to the observed

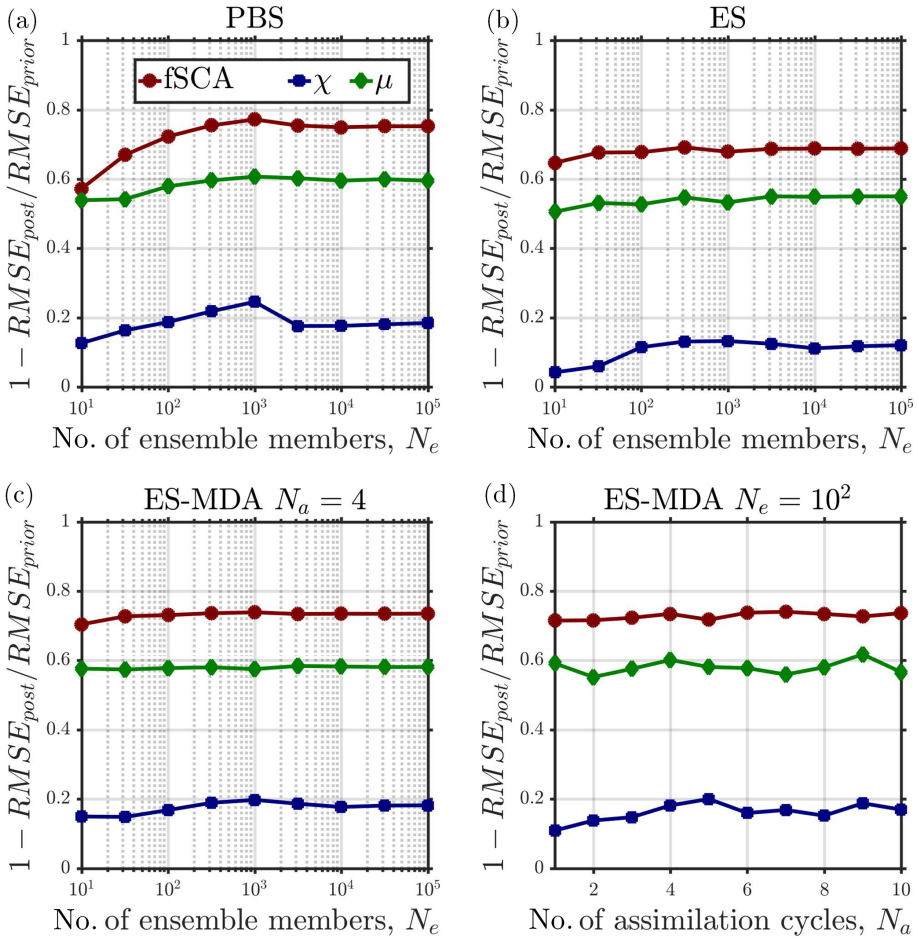


Figure 8. Fractional improvement in RMSE through the analysis step (1 being perfect and 0 no effect) as a function of the number of ensemble members for the fSCA, peak mean SWE μ and coefficient of variation χ; **(a)** particle batch smoother, PBS; **(b)** ensemble smoother, ES; **(c)** ensemble smoother with multiple data assimilation, ES-MDA; **(d)** FI as a function of assimilation cycles in the ES-MDA. The FIs for $N_e \leq 100$ are averaged over 10^2 independent ensemble model integrations. Errors were computed based on comparisons to all the corresponding field measurements presented in Sect. 2.2.

fSCA and significantly constrains the spread of the ensemble (Fig. 4). Thus, the posterior effectively fills the gaps in the remotely sensed fSCA time series using a physically based snow model which is bias-corrected through the assimilation, while at the same time accounting for uncertainties in the retrievals. In addition, the ES-MDA is generally able to correct the prior estimates of the peak mean SWE towards the independently observed values, which is essentially achieved through a bias correction of the model forcing. Although the downscaled forcing is biased, it is a more reliable input than forcing data obtained directly from coarse-scale reanalyses (Østby et al., 2017). For example, the lapse rate correction on temperature in the downscaling (cf. Fiddes and Gruber, 2014) influences the snowfall and melt rates at the more elevated Steinflåen plateau. This effect is not captured in the reanalysis product (Dee et al., 2011), for which the elevation of the nearest grid point is near sea level.

An inherent equifinality problem (see Beven, 2006) exists in SWE reconstruction since different perturbation parameter sets can provide similar results. For example, if the prior fSCA melts out earlier than the observations, this could be due to the prior precipitation having a negative bias, the prior melt having a positive bias or a combination of these two. The opposite would be true if the prior fSCA melts out too late. It is not possible to resolve this equifinality problem with observations of fSCA alone. A key assumption in deterministic SWE reconstruction is that the melt flux is more constrained than the precipitation so that uncertainty in the melt is ignored (Slater et al., 2013). We perturb both the precipitation and the melt, although the latter is assigned a lower uncertainty (Table 4). Through the assimilation we obtain snowmelts that are consistent with the observed snow cover depletion. The close match of the posterior peak mean SWE estimates to the independent field measurements (Fig. 7) suggests that the assimilation yields consistent accumulations

Table 6. Summary of evaluation metrics, i.e., fractional improvement in bias and RMSE, as well as prior and posterior square correlation coefficient (R^2), using the ES-MDA ($N_e = 10^2$, $N_a = 4$) for peak mean SWE (μ) and coefficient of variation (χ) when assimilating only MODIS as well as assimilating both MODIS and Sentinel-2 observations. These metrics are based on a comparison to all the snow surveys conducted in 2016 (see Table 1) and are averaged over 10^2 independent runs, each with 100 ensemble members.

		MODIS				MODIS+S2			
Symbol	No. of obs	FI Bias	FI RMSE	R^2_{prior}	R^2_{post}	FI Bias	FI RMSE	R^2_{prior}	R^2_{post}
μ	3	0.61	0.62	0.67	1.00	0.61	0.62	0.69	1.00
χ	3	−0.77	0.11	0.00	0.41	−0.60	0.15	0.00	0.48

and that the inherent equifinality problem is of minor consequence.

Figure 5 shows that for most years the prior median is a poor estimate of the observed peak mean SWE. This indicates that a deterministic (no assimilation and unperturbed) run is not a good representation of the true state. In addition to biases in the precipitation and melt forcing, crucial processes for peak SWE, such as deposition, are not included in the simple snow model. Furthermore, the subgrid variability of the SWE is typically overestimated in the prior, with the prior distributions typically being too skewed. To circumvent these issues, a more sophisticated model (e.g., ALPINE3D; Lehning et al., 2006) accounting for wind drift could be employed, and the climatological snow distribution pattern (Sturm and Wagner, 2010) could help formulate the prior peak subgrid coefficient-of-variation distribution.

The posterior distributions are on the other hand much closer to the observed distributions for most of the years and sites considered. This suggests that there is sufficient information contained in the remotely sensed snow cover depletion to constrain the peak SSD estimates. On some occasions, especially for Bayelva in 2015, the posterior SSD is far from the observed SSD both in shape and in mean. However, the posterior estimate is still slightly better than the prior, indicating that the assimilation has a positive effect on the outcome. A similar marginal performance is found for Steinåen plateau in 2016, but the number of SWE point observations (see Table 1) is not sufficient to reliably constrain the shape of the observed distribution in this case.

5.2 Evaluation of data assimilation schemes

The ES-MDA exceeds or at least nearly matches the performance of the ES and the PBS, which were used in previous studies (e.g., Durand et al., 2008; Girotto et al., 2014b; Margulis et al., 2015), for all the evaluation metrics considered: bias, RMSE and correlation coefficient for fSCA, peak mean SWE and peak subgrid coefficient of variation. The performance gain over the ES is explained by the iterative nature of the ES-MDA, performing a sequence of smaller corrections in the analysis steps as opposed to one abrupt correction (Emerick and Reynolds, 2013; Stordal and Elsheikh, 2015). Particularly in the case of a nonlinear model, as is the case for the SSM, this process of simulated annealing in the ES-MDA leads to a better approximation of the posterior than a single analysis step.

At least with a low number of ensemble members the ES-MDA also outperforms the PBS. A possible reason for this is that the PBS posterior ensemble spans the same range as the prior ensemble and only changes the relative weights of the ensemble members in the analysis. Thus, if the prior ensemble is so biased that it does not encompass the observations, the PBS is incapable of correcting the posterior towards the observations outside the bounds of the prior. In such a case, the region with high likelihood is very small and not necessarily close to the observations. A good example is the 2008 season at Bayelva (cf. Fig. 4 and Fig. 7), for which the prior is far away from the observed fSCA. Consequently, the PBS is unable to shift the ensemble outside the prior range as opposed to both the ES and the ES-MDA. In several years, the PBS also shows signs of degeneracy; i.e., a large part of the weight is carried by a very small number of particles. As the PBS is essentially a particle filter without re-sampling (Van Leeuwen, 2009), the weights can quickly converge on just a few particles in high-likelihood regions. Therefore, the remaining particles become redundant even for low-dimensional systems with a relatively large number of observations such as the one considered here.

The sensitivity analysis for the ensemble size is consistent with higher-dimensional models. The ES-MDA requires relatively few ensemble members for convergence, similar to the EnKF (Evensen, 2009), while the PBS requires a larger ensemble for convergence as with the PF (Van Leeuwen, 2009). The number of assimilation cycles required for convergence of the ES-MDA (four cycles) is also in line with previous studies (Emerick and Reynolds, 2013). While the PBS and the ES-MDA have not yet been compared with respect to improvement in RMSE, our findings are in agreement with previous studies for both the PBS versus the ES (Margulis et al., 2015) and the ES-MDA versus the ES (Emerick and Reynolds, 2013).

A major downside of the ES-MDA is the computational cost. The ES-MDA requires $N_a + 1$ ensemble model integrations and N_a analysis steps, where N_a (typically ≥ 2) is

the number of assimilation cycles (Sect. 3.3.2). On the other hand, the ES requires only two ensemble model integrations and a single analysis step, while the PBS only needs one ensemble model integration and a single analysis step. Based on a sensitivity analysis (Sect. 4.2), we set $N_a = 4$, so the computational cost of the ES-MDA is higher than for the other schemes. For more complex models, such as Crocus (Vionnet et al., 2012) and SNOWPACK (Bartelt and Lehning, 2002), the ES-MDA could prove to be prohibitively expensive. However, an adaptive version of the ES-MDA (Le et al., 2016) could be employed instead in which the inflation coefficients are calculated on the fly based on a cost function and the iterations stop once the algorithm converges. This scheme could significantly reduce the computational costs for applications of the ES-MDA, as it is equivalent to the ES in years for which the prior encompasses the fSCA retrievals, requiring only one iteration. In years where the prior is far from the observations to be assimilated, on the other hand, multiple data assimilation steps are performed. Furthermore, both the snow model (which has no interaction between neighboring grid cells) and the ES-MDA algorithm can be parallelized using high-performance computing.

5.3 Effects of observation error and assimilation frequency

All the DA methods have problems constraining the spread of the peak subgrid coefficient of variation (χ; see Fig. 7), although they can pull the median in the right direction. A likely reason is the limited information available in the remotely sensed snow cover depletion, with either too-sparse or too-uncertain fSCA retrievals. It is worth considering fSCA retrievals from even more satellites, such as Landsat and PROBA-V, which increases the chances of obtaining more cloud-free scenes. With more scenes available, it may be possible to better constrain the posterior χ ensemble: even with just a few additional retrievals from Sentinel-2 the performance was improved with respect to χ estimation across all evaluation metrics. This also points towards the benefits of including higher-resolution fSCA retrievals from the Landsat and Sentinel-2 satellites, which will be more representative and thus accurate. The effective MODIS footprint is inhomogeneous and differs markedly from the nominal 500 m pixel resolution when the view angle deviates from nadir (Peng et al., 2015). So, even for gridded applications, there is a considerable representativeness error in MODIS fSCA, although this is reduced when several pixels are aggregated.

5.4 Outlook

Several extensions to the presented ensemble-based data assimilation framework could be considered. The first is to change the grid scale of the framework from the order of 1 km to larger or smaller scales. For the latter, it would be possible to assimilate only Landsat- and Sentinel-2-based fSCA retrievals and operate at a grid scale on the order of 100 m, in line with the work of, for example, Girotto et al. (2014a). For the former, one would aggregate the satellite retrievals even further and perform the assimilation at a grid scale on the order of 10 km or larger. This implementation could be problematic as the uniform snowmelt assumption in the SDC (Liston, 2004) may no longer be justified across such large grid cells.

Furthermore, the method could be applied to a larger domain in spatially distributed mode (i.e., multiple grid cells). In this case, fSCA assimilation could be complemented by the assimilation of GRACE TWS and/or PM SWE retrievals, which can also improve SWE estimates during the entire buildup, not only at peak SWE. Both TWS and PM retrievals could constrain the large-scale areal-average SWE estimate within the domain and thus further bias-correct the multiple grid-scale peak mean SWE estimates. However, GRACE TWS retrievals feature a very coarse resolution (around 100 km) so that they would only be useful in conjunction with fSCA retrievals for very large scale applications. On the other hand, the use of higher-resolution PM SWE retrievals (order 25 km) in the assimilation has shown particular promise (e.g., De Lannoy et al., 2012; Li et al., 2017). At the same time, PM SWE retrievals are not accurate in complex topography and forested areas, nor for wet and deep snowpacks (Foster et al., 2005), which might limit the applicability of such multisensor assimilation approaches.

The major problem in the assimilation of fSCA retrievals is the occurrence of clouds, which causes extended gaps in time series obtained from optical sensors. As discussed, using fSCA retrievals from even more sensors could help to fill in the gaps in the remotely sensed snow cover depletion and further constrain the peak subgrid coefficient of variation (χ). The use of additional higher-resolution fSCA retrievals with lower representativeness error (and thus observation error) could also prove especially beneficial for constraining χ.

To reduce the computational costs of the ES-MDA, the adaptive ES-MDA (Le et al., 2016) should be considered. Furthermore, the bias-corrected ES-MDA outlined in Stordal and Elsheikh (2015) may be worth pursuing for future applications, especially when applied to bigger domains with possibly even larger misfits between the prior and the observations. Using a more complex snow model such as Crocus (Vionnet et al., 2012) or SNOWPACK (Bartelt and Lehning, 2002) may not only improve the modeled melt rates but also offer the possibility to assimilate snow grain size retrievals (cf. Painter et al., 2009), as noted by Durand et al. (2008). In addition, the method could be applied in a fully coupled land–atmosphere model. The Intermediate Complexity Atmospheric Research Model (ICAR; Gutmann et al., 2016) shows particular promise in terms of an atmospheric model that can efficiently and iteratively be run in ensemble mode, as required for applications of ES-MDA. In principle, one could run ICAR in ensemble mode coupled to a land surface

model with an adequately complex snow scheme and assimilate fSCA (and possibly PM and TWS) retrievals with the ES-MDA to deliver a snow reanalysis.

As snow is a crucial driver for many terrestrial and atmospheric processes, the presented framework has the potential to improve process modeling in a range of disciplines, especially since the spatial resolution is considerably higher than in passive-microwave-derived SWE data sets. For example, the subgrid variability of permafrost temperatures is closely tied to that of SWE depth (e.g., Gisnås et al., 2016), which has major implications for permafrost mapping (e.g., Westermann et al., 2015b, 2017). Similarly, snow cover information is an important component of many ecological models (e.g., Kohler and Aanes, 2004), and peak SWE is intimately linked to streamflow, which is crucial for hydrology and water resource management (Andreadis and Lettenmaier, 2006; Barnett et al., 2005). Finally, knowledge of the snow distribution and snowmelt is of interest for tourism given its importance for, for example, skiing, hiking and backcountry travel.

6 Conclusions

In this study, we use the ensemble smoother with multiple data assimilation (ES-MDA) scheme to estimate peak SWE distributions at the kilometer scale from time series of remotely sensed fractional snow-covered area (fSCA) from MODIS and Sentinel-2. The ES-MDA is combined with analytical Gaussian anamorphosis to update perturbation parameters that are either lower or double bounded in physical space. The data assimilation is applied to a simple snow model based on the surface energy balance coupled to a probabilistic snow depletion curve. The scheme is driven by downscaled ERA-Interim reanalysis data. As such, both the model forcing and the satellite retrievals are globally available.

The results are compared to field measurements of fSCA and peak SWE distributions from Arctic sites near Ny-Ålesund (79° N, Svalbard, Norway) so that the performance can be evaluated with respect to the estimated fSCA, peak mean SWE and peak subgrid coefficient of variation. From this study, the following conclusions can be drawn:

- At the kilometer scale, the ES-MDA is able to successfully assimilate fSCA retrievals into the simple snow model and estimate the peak subgrid SWE distribution prior to the snowmelt.
- A physically based interpolation of the remotely sensed fSCA time series is obtained that takes into account uncertainties in both the model and the retrievals.
- For the peak mean SWE, the ES-MDA features an average RMSE of 0.09 m w.e. compared to field measurements.
- For the peak subgrid coefficient of variation that controls the width and skewness of the distribution, the ES-MDA usually manages to pull the posterior median in the right direction, but the spread of the ensemble is difficult to constrain.
- By including higher-resolution fSCA retrievals from Sentinel-2, the posterior peak subgrid coefficient-of-variation ensemble can be better constrained. This highlights the potential benefits of assimilating additional higher-resolution fSCA retrievals from sensors on board the Landsat and Sentinel-2 satellites in future work.
- In line with previous studies, the ES-MDA converges with as few as 100 ensemble members and four assimilation cycles.
- With this ES-MDA configuration, the fractional improvement in RMSE from prior to posterior is around 75, 60 and 20 % for the fSCA, peak mean SWE and peak subgrid coefficient of variation.
- The ES-MDA exceeds or at least nearly matches the performance of the particle batch smoother and the ensemble smoother for all evaluation metrics considered.

As the scheme exploits high- and medium-resolution satellite images from optical sensors, it is capable of estimating snow distribution at considerably higher spatial resolutions than traditional SWE products, for example based on passive microwave retrievals. On the other hand, the scheme can only recover the peak subgrid SWE distribution prior to the onset of melt, as opposed to providing information on the seasonal evolution of the snow distribution, so that it can rather complement than replace existing SWE retrieval algorithms. However, the method could become a part of satellite-era hydrometeorological reanalysis schemes with a wide range of applications.

Code and data availability. Data and code are made available upon request from the corresponding author.

Author contributions. KA performed this study as part of his PhD project, co-supervised by SW and LB. KA wrote the code and performed the analysis. SW conceived the original idea and helped download the satellite data. SW performed the field measurements, with KA assisting in the final year. LB assisted with the implementation of the data assimilation. TVS performed the statistical downscaling. JB provided data from and information about the Bayelva station. The manuscript was first written by KA and then edited by all the co-authors.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work was funded by SatPerm (239918; Research Council of Norway), in collaboration with EmblA (56801; Nordforsk) and ESA GlobPermafrost (www.globpermafrost.info). Sebastian Westermann acknowledges additional support by COUP (2449037/E10; JPI Climate; Research Council of Norway) and PermaNor (255331/E10; Research Council of Norway). Thomas Vikhamar Schuler acknowledges funding from ESCYMO (224024; Research Council of Norway). This study is a contribution to the strategic research area LATICE at the University of Oslo. We would like to thank two anonymous reviewers and Marie Dumont, all of whom contributed to significantly improving this paper.

Edited by: Marie Dumont
Reviewed by: two anonymous referees

References

Aas, K. S., Gisnås, K., Westermann, S., and Berntsen, T. K.: A Tiling Approach to Represent Subgrid Snow Variability in Coupled Land Surface-Atmosphere Models, J. Hydrometeorol., 18, 49–63, https://doi.org/10.1175/JHM-D-16-0026.1, 2017.

Andreadis, K. M. and Lettenmaier, D. P.: Assimilating remotely sensed snow observations into a macroscale hydrology model, Adv. Water Res., 29, 872–886, https://doi.org/10.1016/j.advwatres.2005.08.004, 2006.

Barnett, T. P., Adam, J. C., and Lettenmaier, D. P.: Potential Impacts of a warming climate on water availability in snow-dominated regions, Nature, 438, 303–309, https://doi.org/10.1038/nature04141, 2005.

Bartelt, P. and Lehning, M.: A physical SNOWPACK model for the Swiss avalance warning: Part I: numerical model, Cold Reg. Sci. Technol., 35, 123–145, https://doi.org/10.1016/S0165-232X(02)00074-5, 2002.

Bertino, L., Evensen, G., and Wackernagel, H.: Sequential Data Assimilation Techniques in Oceanography, Int. Stat. Rev., 71, 223–241, https://doi.org/10.1111/j.1751-5823.2003.tb00194.x, 2003.

Beven, K.: A manifesto for the equifinality thesis, J. Hydrol., 320, 18–36, https://doi.org/10.1016/j.jhydrol.2005.07.007, 2006.

Beven, K. and Binley, A.: The Future of Distributed Models: Model Calibration and Uncertainty Prediction, Hydrol. Proc., 6, 279–298, https://doi.org/10.1002/hyp.3360060305, 1992.

Blöschl, G.: Scaling issues in snow hydrology, Hydrological Processes, 13, 2149–2175, https://doi.org/10.1002/(SICI)1099-1085(199910)13:14/15<2149::AID-HYP847>3.0.CO;2-8, 1999.

Boike, J., Roth, K., and Ippisch, O.: Seasonal snow cover on frozen ground: Energy balance calculations of a permafrost site near Ny-Ålesund, Spitsbergen, J. Geophys. Res., 108, 1–11, https://doi.org/10.1029/2001JD000939, 2003.

Boike, J., Ippisch, O., Overduin, P. P., Hagedorn, B., and Roth, K.: Water, heat and solute dynamics of a mud boil, Spitsbergen, Geomorphology, 95, 61–73, https://doi.org/10.1016/j.geomorph.2006.07.033, 2008.

Boike, J., Juszak, I., Lange, S., Chadburn, S., Burke, E., Overduin, P. P., Roth, K., Ippisch, O., Bornemann, N., Stern, L., Gouttevin, I., Hauber, E., and Westermann, S.: A 20-year record (1998–2017) of permafrost, active layer, and meteorological conditions at a High Arctic permafrost research site (Bayelva, Spitsbergen): an opportunity to validate remote sensing data and land surface, snow, and permafrost models, Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-100, in review, 2017.

Bruland, O., Sand, K., and Killingtveit, Å.: Snow Distribution at a High Arctic Site at Svalbard, Hydrol. Res., 32, 1–12, 2001.

Cannone, N., Augusti, A., Malfasi, F., Pallozzi, E., Calfapietra, C., and Brugnoli, E.: The interaction of biotic and abiotic factors at multiple spatial scales affects the variability of CO_2 fluxes in polar environments, Polar Biol., 39, 1581–1596, https://doi.org/10.1007/s00300-015-1883-9, 2016.

Charrois, L., Cosme, E., Dumont, M., Lafaysse, M., Morin, S., Libois, Q., and Picard, G.: On the assimilation of optical reflectances and snow depth observations into a detailed snowpack model, The Cryosphere, 10, 1021–1038, https://doi.org/10.5194/tc-10-1021-2016, 2016.

Clark, M. P., Slater, A. G., Barret, A. P., Hay, L. E., McCabe, G. J., Rajagopalan, B., and Leavesley, G. H.: Assimilation of snow covered area information into hydrologic and land-surface models, Adv. Water Res., 29, 1209–1221, https://doi.org/10.1016/j.advwatres.2005.10.001, 2006.

Clark, M. P., Hendrikx, J., Slater, A. G., Kavetski, D., Anderson, B., Cullen, N. J., Kerr, T., Hreinsson, E. O., and Woods, R. A.: Representing spatial variability of snow water equivalent in hydrologic and land-surface models: A review, Water Resour. Res., 47, 1–23, https://doi.org/10.1029/2011WR010745, 2011.

Cline, D. W., Bales, R. C., and Dozier, J.: Estimating the spatial distribution of snow in mountain basins using remote sensing and energy balance modeling, Water Resour. Res., 34, 1275–1285, https://doi.org/10.1029/97WR03755, 1998.

Cortés, G. and Margulis, S.: Impacts of El Niño and La Niña on interannual snow accumulation in the Andes: Results from a high-resolution 31 year reanalysis, Geophys. Res. Lett., 44, 6859–6867, https://doi.org/10.1002/2017GL073826, 2017.

Cortés, G., Girotto, M., and Margulis, S. A.: Analysis of sub-pixel snow and ice extent over the extratropical Andes using spectral unmixing and historical Landsat imagery, Remote Sens. Environ., 141, 64–78, https://doi.org/10.1016/j.rse.2013.10.023, 2014.

Cortés, G., Girotto, M., and Margulis, S.: Snow process estimation over the extratropical Andes using a data assimilation framework integrating MERRA data and Landsat imagery, Water Resour. Res., 52, 2582–2600, https://doi.org/10.1002/2015WR018376, 2016.

De Lannoy, G. J. M., Rechile, R. H., Houser, P. R., Arsenault, K. R., Verhoest, N. E. C., and Pauwels, V. R. N.: Satellite-scale snow water equivalent assimilation into a high-resolution land surface model, J. Hydrometeorol., 11, 352–369, https://doi.org/10.1175/2009JHM1192.1, 2010.

De Lannoy, G. J. M., Reichle, R. H., Arsenault, K. R., Houser, P. R., Kumar, S., Verhoest, N. E. C., and Pauwels, V. R. N.: Multiscale assimilation of Advance Microwave Scanning Radiometer-EOS snow water equivalent and Moderate Resolution Imaging Spectroradiometer snow cover fraction observations in northern Colorado, Water Resour. Res., 48, 1–17, https://doi.org/10.1029/2011WR010588, 2012.

Dee, D., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V.,

Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Q. J. Roy. Meteorol. Soc., 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Dozier, J., Bair, E. H., and Davis, R. E.: Estimating the spatial distribution of snow water equivalent in the world's mountains, Wiley Interdisciplinary Reviews: Water, https://doi.org/10.1002/wat2.1140, 2016.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., and Bargellini, P.: Sentinel-2:ESA's Optical High-Resolution Mission for GMES Operational Services, Remote Sens. Environ., 120, 25–36, https://doi.org/10.1016/j.rse.2011.11.026, 2012.

Dunne, S. and Entekhabi, D.: An ensemble-based reanalysis approach to land data assimilation, Water Resour. Res., 41, W02013, https://doi.org/10.1029/2004WR003449, 2005.

Durand, M., Molotoch, N. P., and Margulis, S. A.: A Bayesian approach to snow water equivalent reconstruction, J. Geophys. Res., 113, D20117, https://doi.org/10.1029/2008JD009894, 2008.

Dutra, E., Balsamo, B., Viterbo, P., Miranda, P., Beljaars, A., Schär, C., and Elder, K.: An Improved Snow Scheme for the ECMWF Land Surface Model: Description and Offline Validation, J. Hydrometeorol., 11, 899–916, https://doi.org/10.1175/2010JHM1249.1, 2010.

Emerick, A. A. and Reynolds, A. C.: Ensemble smoother with multiple data assimilation, Comput. Geosci., 55, 3–15, https://doi.org/10.1016/j.cageo.2012.03.011, 2013.

Esau, I., Argentini, S., Przybylak, R., Repina, I., and Sjöblom, A.: Svalbard Meteorology, Adv. Meteorol., 2012, 1–4, https://doi.org/10.1155/2012/818473, 2012.

Evensen, G.: Data Assimilation: The Ensemble Kalman Filter, Springer-Verlag Berlin Heidelberg, https://doi.org/10.1007/978-3-642-03711-5, 2009.

Fiddes, J. and Gruber, S.: TopoSCALE v.1.0: downscaling gridded climate data in complex terrain, Geosci. Model Dev., 7, 387–405, https://doi.org/10.5194/gmd-7-387-2014, 2014.

Førland, J. E., Benestad, R., Hanssen-Bauer, I., Haugen, J. E., and Skaugen, T. E.: Temperature and Precipitation Development at Svalbard 1900–2100, Adv. Meteorol., 2011, 1–14, https://doi.org/10.1155/2011/893790, 2011.

Foster, J. L., Sun, C., Walker, J. P., Kelly, R., Chang, A., Dong, J., and Powell, H.: Quantifying the uncertainty in passive microwave snow water equivalent observations, Remote Sens. Environ., 94, 187–203, https://doi.org/10.1016/j.rse.2004.09.012, 2005.

Girotto, M., Cortés, G., Margulis, S. A., and Durand, M.: Examining spatial and temporal variability in snow water equivalent using a 27 year reanalysis: Kern River watershed, Sierra Nevada, Water Resour. Res., 50, 6713–6734, https://doi.org/10.1002/2014WR015346, 2014a.

Girotto, M., Margulis, S. A., and Durand, M.: Probabilistic SWE reanalysis as a generalization of deterministic SWE reconstruction techniques, Hydrol. Proc., 28, 3875–3895, https://doi.org/10.1002/hyp.9887, 2014b.

Gisnås, K., Westermann, S., Schuler, T. V., Litherland, T., Isaksen, K., Boike, J., and Etzelmüller, B.: A statistical approach to represent small-scale variability of permafrost temperatures due to snow cover, The Cryosphere, 8, 2063–2074, https://doi.org/10.5194/tc-8-2063-2014, 2014.

Gisnås, K., Westermann, S., Schuler, T. V., Melvold, K., and Etzelmüller, B.: Small-scale variation of snow in a regional permafrost model, The Cryosphere, 10, 1201–1215, https://doi.org/10.5194/tc-10-1201-2016, 2016.

Gutmann, E., Barstad, I., Clark, M., Arnold, J., and Rasmussen, R.: The Intermediate Complexity Atmospheric Research Model (ICAR), J. Hydrometeorol., 17, 957–973, https://doi.org/10.1175/JHM-D-15-0155.1, 2016.

Hall, D. K. and Riggs, G. A.: MODIS/Terra Snow Cover Daily L3 Global 500m Grid, Version 6, Tile h18v01, https://doi.org/10.5067/MODIS/MOD10A1.006 (last access: 5 May 2017), Boulder, Colorado USA, NASA National Snow and Ice Data Center Distributed Active Archive Center, 2016a.

Hall, D. K. and Riggs, G. A.: MODIS/Aqua Snow Cover Daily L3 Global 500m Grid, Version 6, Tile h18v01, https://doi.org/10.5067/MODIS/MYD10A1.006 (last access: 5 May 2017), Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center, 2016b.

Hall, D. K., Riggs, G. A., Salomonson, V. V., DiGirolamo, N. E., and Bayr, K. J.: MODIS snow-cover products, Remote Sens. Environ., 83, 181–194, https://doi.org/10.1016/S0034-4257(02)00095-0, 2002.

Ide, K., Courtier, P., Ghil, M., and Lorenc, A.: Unified Notation for Data Assimilation: Operational, Sequential and Variational, J. Meteorol. Soc. Japan, 75, 181–189, 1997.

Kępski, D., Luks, B., Migała, K., Wawrzyniak, T., Westermann, S., and Wojtuń, B.: Terrestrial Remote Sensing of Snowmelt in a Diverse High-Arctic Tundra Environment Using Time-Lapse Imagery, Remote Sens., 9, 1–22, https://doi.org/10.3390/rs9070733, 2017.

Kohler, J. and Aanes, R.: Effect of Winter Snow and Ground-Icing on a Svalbard Reindeer Population: Results of a Simple Snowpack Model, Arctic, Antarctic and Alpine Research, 36, 333–341, https://doi.org/10.1657/1523-0430(2004)036[0333:EOWSAG]2.0.CO;2, 2004.

Kolberg, S. A. and Gottschalk, L.: Updating of snow depletion curve with remote sensing data, Hydrol. Proc., 20, 2363–2380, https://doi.org/10.1002/hyp.6060, 2006.

Le, D. H., Emerick, A. A., and Reynolds, A. C.: An Adaptive Ensemble Smoother With Multiple Data Assimilation for Assisted History Matching, Society of Petroleum Engineers Journal, 21, 2195–2207, https://doi.org/10.2118/173214-PA, 2016.

Lehning, M., Völksch, I., Gustafsson, D., Nguyen, T. A., Stähli, M., and Zappa, M.: ALPINE3D: a detailed model of mountain surface processes and its application to snow hydrology, Hydrol. Proc., 20, 2111–2128, https://doi.org/10.1002/hyp.6204, 2006.

Li, D., Durand, M., and Margulis, S. A.: Estimating snow water equivalent in a Sierra Nevada watershed via spaceborne radiance data assimilation, Water Resour. Res., 53, 647–671, https://doi.org/10.1002/2016WR018878, 2017.

Liston, G. E.: Interrelationships among Snow Distribution, Snowmelt, and Snow Cover Depletion: Implications for Atmospheric, Hydrologic, and Ecologic Modeling, J. Appl. Meteorol., 38, 1474–1487, 1999.

Liston, G. E.: Representing Subgrid Snow Cover Heterogeneities in Regional and Global Models, J. Climate, 17, 1381–1397, https://doi.org/10.1175/1520-0442(2004)017<1381:RSSCHI>2.0.CO;2, 2004.

López-Moreno, J. I., Boike, J., Sanchez-Lorenzo, A., and Pomeroy, J. W.: Impact of climate warming on snow processes in Ny-Ålesund, a polar maritime site at Svalbard, Global Planet. Change, 146, 10–21, https://doi.org/10.1016/j.gloplacha.2016.09.006, 2016.

Luce, C. H. and Tarboton, D. G.: The application of depletion curves for parameterization of subgrid variability of snow, Hydrol. Proc., 18, 1409–1422, https://doi.org/10.1002/hyp.1420, 2004.

Lüers, J., Westermann, S., Piel, K., and Boike, J.: Annual CO_2 budget and seasonal CO_2 exchange signals at a high Arctic permafrost site on Spitsbergen, Svalbard archipelago, Biogeosciences, 11, 6307–6322, https://doi.org/10.5194/bg-11-6307-2014, 2014.

Magnusson, J., Winstral, A., Stordal, A. S., Essery, R., and Jonas, T.: Improving physically based snow simulations by assimilating snow depths using the particle filter, Water Resour. Res., 53, 1–19, https://doi.org/10.1002/2016WR019092, 2017.

Margulis, S. A., Girotto, M., Cortés, G., and Durand, M.: A Particle Batch Smoother Approach to Snow Water Equivalent Estimation, J. Hydrometeorol., 16, 1752–1772, https://doi.org/10.1175/JHM-D-14-0177.1, 2015.

Margulis, S. A., Cortés, G., Girotto, M., and Durand, M.: A Landsat-Era Sierra Nevada Snow Reanalysis (1985-2015), J. Hydrometeorol., 17, 1203–1221, https://doi.org/10.1175/JHM-D-15-0177.1, 2016.

Martinec, J. and Rango, A.: Areal Distribution of Snow Water Equivalent Evaluated by Snow Cover Monitoring, Water Resour. Res., 17, 1480–1488, https://doi.org/10.1029/WR017i005p01480, 1981.

Molotch, N. P. and Margulis, S. A.: Estimating the distribution of snow water equivalent using remotely sensed snow cover data and a spatially distributed snowmelt model: A multi-resolution, multi-sensor comparison, Adv. Water Res., 31, 1503–1514, https://doi.org/10.1016/j.advwatres.2008.07.017, 2008.

Niu, G. Y., Seo, K. W., Yang, Z. L., Wilson, C., Su, H., Chen, J., and Rodell, M.: Retrieving snow mass from GRACE terrestrial water storage change with a land surface model, Geophys. Res. Lett., 34, 1–5, https://doi.org/10.1029/2007GL030413, 2007.

Nowak, A. and Hodson, A.: Hydrological response of a High-Arctic catchment to changing climate over the past 35 years: a case study of Bayelva watershed, Svalbard, Polar Res., 32, 1–17, https://doi.org/10.3402/polar.v32i0.19691, 2013.

NPI: Norwegian Polar Intsitute DEM: Terrengmodell Svalbard (S0 Terrengmodell), Delmodell 5m 2010_13828_33, available at: https://doi.org/10.21334/npolar.2014.dce53a47 (last access: 3 December 2016), 2014.

Østby, T. I., Schuler, T. V., Hagen, J. O., Hock, R., Kohler, J., and Reijmer, C. H.: Diagnosing the decline in climatic mass balance of glaciers in Svalbard over 1957–2014, The Cryosphere, 11, 191–215, https://doi.org/10.5194/tc-11-191-2017, 2017.

Painter, T. H., Rittger, K., McKenzie, C., Slaughter, P., Davis, R. E., and Dozier, J.: Retrieval of subpixel snow covered area, grain size and, and albedo from MODIS, Remote Sens. Environ., 113, 868–879, https://doi.org/10.1016/j.rse.2009.01.001, 2009.

Pedersen, C. and Winther, J.: Intercomparison and validation of snow albedo parameterization schemes in climate models, Clim. Dynam., 25, 351–362, https://doi.org/10.1007/s00382-005-0037-0, 2005.

Peng, J., Liu, Q., Wang, L., Liu, Q., Fan, W., Lu, M., and Wen, J.: Characterizing the Pixel Footprint of Satellite Albedo Products Derived from MODIS Reflectance in the Heihe River Basin, China, Remote Sens., 7, 6886–6907, https://doi.org/10.3390/rs70606886, 2015.

Raleigh, M. S., Lundquist, J. D., and Clark, M. P.: Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework, Hydrol. Earth Syst. Sci., 19, 3153–3179, https://doi.org/10.5194/hess-19-3153-2015, 2015.

Roth, K. and Boike, J.: Quantifying the thermal dynamics of a permafrost site near Ny-Ålesund, Svalbard, Water Resour. Res., 37, 2901–2914, https://doi.org/10.1029/2000WR000163, 2001.

Salomonson, V. V. and Appel, I.: Estimating fractional snow cover from MODIS using the normalized difference snow index, Remote Sens. Environ., 89, 351–360, https://doi.org/10.1016/j.rse.2003.10.016, 2004.

Slater, A. G. and Clark, M. P.: Snow Data Assimilation via an Ensemble Kalman Filter, J. Hydrometeorol., 7, 478–493, 2006.

Slater, A. G., Barrett, A. P., Clark, M. P., Lundquist, J. D., and Raleigh, M. S.: Uncertainty in seasonal snow reconstruction: Relative impacts of model forcing and image availability, Adv. Water Res., 55, 165–177, https://doi.org/10.1016/j.advwatres.2012.07.006, 2013.

Smith, R. B. and Barstad, I.: A Linear Theory of Orographic Precipitation, Journal of the Atmospheric Sciences, 61, 1377–1391, https://doi.org/10.1175/1520-0469(2004)061<1377:ALTOOP>2.0.CO;2, 2004.

Stordal, A. S. and Elsheikh, A. H.: Iterative ensemble smoothers in the annealed importance sampling framework, Adv. Water Res., 86, 231–239, https://doi.org/10.1016/j.advwatres.2015.09.030, 2015.

Sturm, M. and Wagner, A. M.: Using repeated patterns in snow distribution modeling: An Arctic Example, Water Resour. Res., 46, 1–15, https://doi.org/10.1029/2010WR009434, 2010.

Su, H., Yang, Z. L., Dickinson, R. E., Wilson, C. R., and Niu, G. Y.: Multisensor snow data assimilation at the continental scale: The value of Gravity Recovery and Climate Experiment terrestrial water storage information, J. Geophys. Res., 115, 1–14, https://doi.org/10.1029/2009JD013035, 2010.

Tarboton, D. G. and Luce, C. H.: Utah Energy Balance Snow Accumulation and Melt Model, Computer model techincal description and users guide, Utah Water Research Laboratory and USDA Forest Service Intermountain Research Station, Logan, Utah, 1996.

Van Leeuwen, P. J.: Particle Filtering in Geophysical Systems, Mon. Weather Rev., 137, 4089–4114, https://doi.org/10.1175/2009MWR2835.1, 2009.

Van Leeuwen, P. J. and Evensen, G.: Data assimilation and inverse methods in terms of a probabilistic formulation, Mon. Weather Rev., 124, 2898–2913, https://doi.org/10.1175/1520-0493(1996)124<2898:DAAIMI>2.0.CO;2, 1996.

Vionnet, V., Brun, E., Morin, S., Boone, A., Faroux, S., Le Moigne, P., Martin, E., and Willemet, J.-M.: The detailed snowpack scheme Crocus and its implementation in SURFEX v7.2,

Geosci. Model Dev., 5, 773–791, https://doi.org/10.5194/gmd-5-773-2012, 2012.

Westermann, S., Lüers, J., Langer, M., Piel, K., and Boike, J.: The annual surface energy budget of a high-arctic permafrost site on Svalbard, Norway, The Cryosphere, 3, 245–263, https://doi.org/10.5194/tc-3-245-2009, 2009.

Westermann, S., Boike, J., Langer, M., Schuler, T. V., and Etzelmüller, B.: Modeling the impact of wintertime rain events on the thermal regime of permafrost, The Cryosphere, 5, 945–959, https://doi.org/10.5194/tc-5-945-2011, 2011a.

Westermann, S., Langer, M., and Boike, J.: Spatial and temporal variations of summer surface temperatures of high-arctic tundra on Svalbard - Implications for MODIS LST based permafrost monitoring, Remote Sens. Environ., 115, 908–922, https://doi.org/10.1016/j.rse.2010.11.018, 2011b.

Westermann, S., Langer, M., and Boike, J.: Systematic bias of winter-time land surface temperatures inferred from MODIS at a site on Svalbard, Norway, Remote Sens. Environ., 118, 162–167, https://doi.org/10.1016/j.rse.2011.10.025, 2012.

Westermann, S., Boike, J., Guglielmin, M., Gisnås, K., and Etzmüller, B.: Snow melt monitoring near Ny-Ålesund, Svalbard, using Automatic Camera Systems, available at: https://doi.org/10.1594/PANGAEA.846617 (last access: 5 May 2017), Department of Geosciences, University of Oslo, 2015a.

Westermann, S., Østby, T. I., Gisnås, K., Schuler, T. V., and Etzelmüller, B.: A ground temperature map of the North Atlantic permafrost region based on remote sensing and reanalysis data, The Cryosphere, 9, 1303–1319, https://doi.org/10.5194/tc-9-1303-2015, 2015b.

Westermann, S., Langer, M., Boike, J., Heikenfeld, M., Peter, M., Etzelmüller, B., and Krinner, G.: Simulating the thermal regime and thaw processes of ice-rich permafrost ground with the land-surface model CryoGrid 3, Geosci. Model Dev., 9, 523–546, https://doi.org/10.5194/gmd-9-523-2016, 2016.

Westermann, S., Peter, M., Langer, M., Schwamborn, G., Schirrmeister, L., Etzelmüller, B., and Boike, J.: Transient modeling of the ground thermal conditions using satellite data in the Lena River delta, Siberia, The Cryosphere, 11, 1441–1463, https://doi.org/10.5194/tc-11-1441-2017, 2017.

Winther, J. G., Godtliebsen, F., Gerland, S., and Isachsen, P. E.: Surface albedo in Ny-Ålesund, Svalbard: variability and trends during 1981–1997, Global Planet. Change, 32, 127–139, https://doi.org/10.1016/S0921-8181(01)00103-5, 2002.

You, J., Tarboton, D. G., and Luce, C. H.: Modeling the snow surface temperature with a one-layer energy balance snowmelt model, Hydrol. Earth Syst. Sci., 18, 5061–5076, https://doi.org/10.5194/hess-18-5061-2014, 2014.

Paper II

Evaluating satellite retrieved fractional snow-covered area at a high-Arctic site using terrestrial photography

K. Aalstad, S. Westermann, L. Bertino

Submitted to *Remote Sensing of Environment*, April 2019.

II

Evaluating satellite retrieved fractional snow-covered area at a high-Arctic site using terrestrial photography

Kristoffer Aalstad[a,*], Sebastian Westermann[a], Laurent Bertino[b]

[a]*Department of Geosciences, University of Oslo, Oslo, Norway*
[b]*Nansen Environmental and Remote Sensing Center, Bergen, Norway*

Abstract

The seasonal snow-cover is one of the most rapidly varying natural surface features on Earth. It strongly modulates the terrestrial water, energy, and carbon balance. Fractional snow-covered area (fSCA) is an essential snow variable that can be retrieved from multispectral satellite imagery. In this study, we evaluate fSCA retrievals from multiple sensors that are currently in polar orbit: the operational land imager (OLI) on-board Landsat 8, the multispectral instrument (MSI) on-board the Sentinel-2 satellites, and the moderate resolution imaging spectroradiometer (MODIS) on-board Terra and Aqua. We consider several retrieval algorithms that fall into three classes: thresholding of the normalized difference snow index (NDSI), regression on the NDSI, and spectral unmixing. We conduct the evaluation at a high-Arctic site in Svalbard, Norway, by comparing satellite retrieved fSCA to coincident high-resolution snow-cover maps obtained from a terrestrial automatic camera system. For the lower resolution MODIS retrievals, the regression-based retrievals outperformed the unmixing-based retrievals for all metrics but the bias. For the higher resolution sensors (OLI and MSI), retrievals based on NDSI thresholding overestimated the fSCA due to the mixed pixel problem whereas spectral unmixing retrievals provided the most reliable estimates across the board. We therefore encourage the operationalization of spectral unmixing retrievals of fSCA from both OLI and MSI.

Keywords: Fractional snow-cover, Optical, Photography, Unmixing, NDSI, Mixed pixel, MODIS, Landsat 8, Sentinel-2

*Corresponding author.
E-mail address: kristoffer.aalstad@geo.uio.no

1. Introduction

Seasonally, snow covers between less than 10% to over 40% of the land area of the northern hemisphere, making it one of the most rapidly varying natural surface features on Earth (Hall, 1988). Given its unique combination of physical properties: high albedo, low thermal conductivity, and large water storing capacity, this seasonal snowpack strongly modulates the terrestrial surface energy and water balance in cold regions (Zhang, 2005; Dozier et al., 2016). These high latitudes and elevations are experiencing anthropogenically induced climatic warming that is amplified by positive climate feedbacks, including the snow-albedo feedback (Chapin et al., 2005; Pepin et al., 2015). A warmer climate can shift the phase of precipitation from snow to rain and lead to an earlier and slower snowmelt (Stewart et al., 2005; Musselman et al., 2017). Thereby, warming may reduce runoff efficiency and shift peak streamflow from summer and autumn, when demand is highest, to early spring and even winter (Barnett et al., 2005; Mankin et al., 2015). This is concerning given that more than 25% of the worlds population relies on mountain snowmelt for fresh water supply (Mankin et al., 2015). A shortening of the snow season due to climatic warming may also have a detrimental effect on Arctic and alpine vegetation, with implications for the albedo and carbon balance in these regions (Niittynen et al., 2018). At high latitudes, amplification accelerates the thawing of vast tracts of permafrost soils which mobilizes previously frozen carbon to the atmosphere and may further enhance global warming (Schuur et al., 2015; Chadburn et al., 2017).

Satellite remote sensing is an invaluable tool for monitoring the state of the seasonal snowpack (Bormann et al., 2018; Yılmaz et al., 2019). We are using polar-orbiting satellite sensors in this study as they provide measurements in polar regions where geostationary satellite sensors do not provide coverage. Most snow remote sensing techniques rely on the spectral spectral signature of electromagnetic radiation induced by interactions with the snowpack. Optical sensors measuring reflected shortwave radiation in multiple bands in the visible-shortwave infrared (VSWIR) can be combined with the spectral signature of snow to retrieve snow-covered area, albedo, snow grain size, and impurity concentration (Nolin et al., 1993; Dozier et al., 2009; Painter et al., 2013) at scales from tens to hundreds of meters. Furthermore, high horizontal resolution (1-100 m) snow depth estimates can be retrieved

by taking the difference of snow-covered and snow-free digital surface models (DSMs) obtained from stereo satellite imagery (Marti et al., 2016) or laser altimetry (Treichler and Kääb, 2017). As a caveat, cloud cover frequently prevents remote sensing of snowpack information with space-borne sensors operating in the VSWIR.

At this stage, primarily VSWIR sensors offer the spatiotemporal resolution and breadth of information needed to quantify essential snowpack properties at fine spatial scales in complex terrain (Dozier et al., 2016; Seidel et al., 2016; Bormann et al., 2018). VSWIR sensors can detect reflectance from the surface layer of the snowpack (upper 5-10 cm Dozier et al., 2009), such that individual images can be used to retrieve snow-covered area (binary or fractional), but not snow water equivalent (SWE) or snow depth. However, time series of satellite images have long been used to ingest the remotely sensed depletion of snow-covered area into snowmelt models to reconstruct SWE (Martinec and Rango, 1981; Girotto et al., 2014; Rittger et al., 2016; Dozier et al., 2016; Aalstad et al., 2018). Furthermore, satellite retrievals of snow-covered area have a wide range of uses, for example: automatic mapping of persistent ice and snow-cover (Winsvold et al., 2016; Selkowitz and Forster, 2016), investigating the snow-cover climatology (Gascoin et al., 2015; Bormann et al., 2018; Yılmaz et al., 2019), generating species distribution models (Niittynen et al., 2018), constraining permafrost models (Trofaier et al., 2017), and improving parametrizations used in Earth system models (Swenson and Lawrence, 2012).

The various applications of remotely sensed snow-cover products highlight the need to understand and characterize their uncertainty. Extensive evaluations of snow-covered area retrievals have been undertaken (Hall et al., 2002; Salomonson and Appel, 2004; Painter et al., 2009; Rittger et al., 2013; Cortés et al., 2014; Arsenault et al., 2014; Gascoin et al., 2015; Masson et al., 2018), but usually using other satellite retrievals as a reference. In this study, we retrieve reference fractional snow-covered area (fSCA) maps for a 1.77 km^2 area of interest in the high-Arctic, using imagery taken by an automatic camera system (ACS) deployed on a mountain top. These maps are retrieved at a high spatiotemporal resolution (0.5 m, daily during the snowmelt season) and are used to evaluate fSCA retrieved from three optical satellite sensors that are currently in polar orbit: the Moderate Resolution Imaging Spectroradiometer (MODIS) on-board the Terra and Aqua satellites, the Operational Land Imager (OLI) on-board Landsat 8, and the MultiSpectral Instrument (MSI) on-board Sentinel-2A and Sentinel-2B. In addition to evaluating exist-

ing products, such as MOD10A1 (Riggs et al., 2017) and MODSCAG (Painter et al., 2009), we also evaluate implementations of retrieval algorithms ranging from simple thresholding to fully constrained linear spectral unmixing.

2. Data and methods

2.1. Study area

The study area (see Figure 1) is located in the Bayelva catchment on the Brøgger peninsula in north western Svalbard. The catchment is situated about 3 km west of the village of Ny-Ålesund (78°55′N, 11°50′E) which is the northernmost permanent civilian settlement in the world and serves as a major hub for polar research. The relatively warm West Spitsbergen (ocean) current ensures that this part of Svalbard has a maritime climate with mild winters and cool summers for this latitude (Boike et al., 2018). For the period 1981-2010, Ny-Ålesund received 427 mm annual average precipitation with winter, summer, and annual average air temperatures of -12.0, 3.5, and -5.2°C (Førland et al., 2011). At Bayelva, the climatological range of the catchment averaged peak SWE is from around 0.1 to 0.5 m w.e, the end of season snow density is typically 350±50 kg m^{-3}, and the snow-cover usually disappears between early June and mid-July (Aalstad et al., 2018; Boike et al., 2018). Given the high latitude, during solar noon on the summer solstice, when the sun is highest in the sky, the solar zenith angle reaches a minimum of 56° (i.e., a solar elevation angle as low as 34°). Between the 21st of October and the 20th of February, the sun is below the true horizon during polar night.

Our area of interest (AOI) is situated around the Bayelva climate station (BCS), where continuous measurements related to permafrost, snow, and hydrometeorology have been undertaken for the last two decades (e.g. Westermann et al., 2011, 2012; Aalstad et al., 2018; Boike et al., 2018). The area features elevations between 5 and 55 m a.s.l., and gently undulating topography with small hills (see Figure 2), causing large differences in snow-cover due to wind drift. The surface cover alternates between bare soil (silty loam and silty clay), rocks (shale, sandstone, and coal), and low lying vegetation (mosses, lichens, and grasses). The AOI is located between two mountains, Zeppelinfjellet (556 m a.s.l.) to the south east and Scheteligfjellet (719 m a.s.l.) to the west on which the ACS is mounted. To the south the AOI is bordered by the two branches of the Brøggerbreen glacier that feeds the Bayelva river. The AOI highlighted in Figure 1 spans most of the Bayelva

catchment, covering an area of around 1.77 km^2 between Scheteligfjellet and Ny-Ålesund.

2.2. Terrestrial Photography

Here, we describe the processing chain for the reference images obtained from terrestrial photography. The entire chain is shown schematically in Figure 3. For all snowmelt seasons (May to August) from 2012 to 2017, an automatic time lapse camera system (ACS) was deployed near the summit of Scheteligfjellet (c.f. Figure 1), overlooking the Bayelva site. A Canon EOS 1100D digital single-lens reflex camera was triggered by a Harbotronics time-lapse system, delivering high-quality daily images over the study area except for days with prolonged low-cloud cover or system malfunction. For each day, the photograph with the most favorable illumination conditions, i.e. close to solar noon and lack of cloud or other shadows, was selected. After filtering out low quality images, in total 305 terrestrial images from different days were available for further analysis.

A reference image and DEM were employed to georeference and othorectify the images. We used an orthophoto and DEM derived from the airborne high resolution stereo camera (HRSC-AX) mission flown over the Brøgger peninsula around noon on the 17$^{\text{th}}$ of July 2008 (see Boike et al., 2018, and references therein) with a ground sampling distance (GSD) of 0.5 m. The DEM and associated true color image, cropped to our AOI, is displayed in Figure 2, whereas the full HRSC true color image used for georeferencing is shown in Figure 3.

For this study, we employ orthorectified images similar to the ones available through Westermann et al. (2015) which were already used to evaluate satellite retrievals of fSCA in a data assimilation experiment (Aalstad et al., 2018). The orthorectification procedure relies on a range of natural ground control points visible in both the camera images and the HRSC-AX orthophoto. We then fit the parameters of a simple camera model (Bouguet, 2015), while the position of the ACS was derived from Differential Global Positioning System (DGPS) measurements. For this study, we applied this standardized orthorectification process to the images from the snowmelt seasons in all the years 2012-2017. The AOI was chosen to exclude the outer edges of the orthorectified images where significant distortions occur.

We conducted an independent evaluation of the georeferencing, comparing the location of 19 landmark features (different to the ones used as ground control points in the orthorectification) in the reference (HRSC-AX) image

to those in the orthorectified terrestrial photographs. For each snowmelt season, we randomly selected 5 images, such that a subset consisting of 6×5=30 images was used to diagnose the georeferencing error. Due to the snow-cover, it was not always possible to identify all the landmarks in an image. In such cases the obscured landmarks were ignored. Based on this exercise, we obtained a root mean squared error (RMSE) of 2.25 m and a bias of 1.85 m for the georeferencing in the orthorectified terrestrial photographs. An example of the georeferencing error present in the images is shown in Figure 3(c). By separating the georeferencing error into its directional components, we found that the georeferencing biases in the northing and easting direction were quite small (0.16 m and 0.59 m, respectively) indicating no systematic shifts. Using the easting coordinate as a proxy for the distance from the camera, given that the camera is situated due west of the AOI, we found a negligible correlation (-0.11) with the georeferencing error.

Next, the orthorectified and georeferenced images were cropped to the AOI (c.f. Figure 3), which is not exactly rectangular to exclude the runway at Ny-Ålesund airport. Due to the frequent plowing of snow on this runway it was excluded from the AOI as it could lead to undesirable situations where a satellite image was captured before and the terrestrial image after the plowing, or vice versa. Other than the runway, there are no significant anthropogenic disturbances within the AOI.

We manually classified the $\sim 7 \times 10^6$ pixels in each cropped image in two categories, snow-covered and snow-free, using a simple image specific threshold in the blue band of each image. This method has been shown to perform reasonably well for much larger areas than our AOI (Fedorov et al., 2016). The threshold was chosen iteratively for each image until a good visual agreement was found between the snow-cover in the true color image and the resulting classification. In most cases, (see Figure 3) the blue band histogram exhibited clear bimodality, and the threshold could easily be chosen as the local minimum between the snow peak and snow-free peak in the histogram. We gauged the sensitivity of the reference snow-cover maps to the choice of thresholds by varying the selected thresholds by ±5% for each of the aforementioned 30 randomly selected images. For an increase of 5%, we found a mean increase in fSCA of 0.01 and a maximum increase of 0.02. For a decrease of 5%, we also found a mean decrease in fSCA of 0.01 with a maximum decrease of 0.04. This suggests that the reference snow-cover maps are relatively insensitive to small variations in the selected thresholds. Three images contained sunglint over areas with surface water, which made

it impossible to pick a valid threshold for these images. Consequently, these three sunglinted images were excluded from further analysis, resulting in a total of 302 classified images that we use as a ground truth for evaluating various satellite retrievals of fSCA.

As a final step, to gauge the uncertainty in our ground truth data, we translated the georeferencing error to a potential error in the reference fSCA maps using geostatistical analysis (see e.g. Wackernagel, 2003). For each reference snow-cover image, we calculated the fSCA for 100×100 meter moving windows using a 2D digital filter. We then computed the empirical variogram as a function of separation distance. Based on these variograms, assuming isotropic and stationary conditions, we could diagnose the estimation standard deviation (σ_E) for the reference fSCA as a function of separation distance for each image. Considering all the 302 images, for a separation distance equal to the georeferencing RMSE we obtained a worst case $\sigma_E \simeq 0.01$, and for a separation distance equal to three times the georeferencing RMSE we obtained a worst case $\sigma_E \simeq 0.03$.

2.3. Satellite retrievals

In this study, we employed fSCA retrievals from multiple satellite sensors using several different retrieval algorithms. The retrievals were both in the form of already processed fSCA products and our own retrievals obtained through algorithms of varying complexity. Here, we outline both the satellite platforms, sensors, and products as well as the retrievals algorithms. An overview of all the retrievals evaluated in this study is provided in Table 1.

2.3.1. MODIS

The MODIS instrument that is currently in orbit on the Terra (2000-present) and Aqua (2002-present) satellite platforms provides multispectral imagery in the VSWIR wavelength range at a daily revisit period and a GSD of approximately 500 meters (for all the VSWIR bands). In this study, we used several MODIS products, namely: both collection/version 5 and 6 of the NASA snow-cover products MOD10A1 (Terra) and MYD10A1 (Aqua) distributed by NSIDC (Hall et al., 2006a,b; Hall and Riggs, 2016a,b), the MODIS Snow Covered-Area And Grain size (MODSCAG) product from Caltech/Jet Propulsion Laboratory (JPL) (Painter et al., 2009), and the NASA surface reflectance products MOD09GA (Terra) and MYD09GA (Aqua) distributed by LP DAAC (Vermote and Wolfe, 2015a,b). The MOD10A1, MYD10A1, and MODSCAG products are used as is, while the MOD09GA

and MYD09GA products are used as inputs for the unmixing described in Section 2.3.5. For all the MODIS products, we use the five pixels in tile h18v01 that fall completely within the AOI (see Figure 1).

2.3.2. Landsat 8 OLI

The OLI sensor on-board Landsat-8 has been in orbit since February 2013 and delivers multispectral imagery in the VSWIR at a GSD of 30 m with a revisit period of 16 days at the equator. For this study, we used Landsat-8 OLI level 1 top of the atmosphere (TOA) reflectances and level 2 surface reflectances (Vermote et al., 2016; USGS, 2018), both distributed by USGS EROS through the EarthExplorer online user interface. Both the TOA and surface reflectances were used in NDSI thresholding and regression algorithms, while only the latter surface reflectances were used for spectral unmixing. Some scenes were not available in the correct UTM zone for our AOI (33X). In these cases, we reprojected the scenes to zone 33X and performed a nearest neighbor interpolation to a regular grid over our AOI. The level 2 product includes the pixel land-cover classification from the Fmask algorithm (Zhu and Woodcock, 2012; USGS, 2018) which we also evaluate in terms of snow-cover mapping performance. Fmask is based on a thresholding of the NDSI, in addition to many other tests to diagnose the pixel land cover and cloud status. We simplified the Fmask classification to be more relevant for snow-cover mapping as follows:

- Pixels classified as cloud shadows, terrain occluded, and/or saturated were reclassified as 'missing'.
- Pixels classified as high confidence cloud (i.e. all cloud bits equal to 1) were reclassified to 'cloudy'.
- Pixels that were not 'missing' or 'cloudy' but with snow bit equal to 1 were reclassified as 'snow' pixels.
- All remaining pixels were reclassified as 'snow-free' pixels.

For the Landsat 8 OLI imagery we manually selected cloud free scenes over our AOI. This resulted in a total of 26 scenes with accompanying reference snow-cover maps for the ablation seasons (May-July) from 2013 to 2017.

2.3.3. Sentinel-2 MSI

The MSI sensor is currently in orbit on-board Sentinel-2A (2015-present) and Sentinel-2B (2017-present) delivering VSWIR multispectral imagery at a GSD of 10-20 m with a revisit frequency of 5 days at the equator (Drusch et al., 2012). In our study, we used two products from Sentinel-2 MSI: level 1 (L1C) TOA reflectances and level 2 (L2A) surface reflectances distributed by the Copernicus Open Access Hub. Since not all L2A scenes had been processed for Svalbard yet, particularly in 2016, we performed conversions from L1C to L2A using the Sen2Cor python script (Müller-Wilm, 2018). As with Landsat 8 OLI, both the TOA and surface reflectances were used for NDSI thresholding and regression retrieval algorithms, while only the surface reflectances were used in spectral unmixing. Scenes that were not available in the correct UTM zone were reprojected and resampled in the same way as for the Landsat 8 OLI products. We also resampled the SWIR bands at 20 m GSD to be in line with the other bands at 10 m GSD using nearest neighbor interpolation. In addition to the L1C and L2A reflectances, we employed the scene land cover classification from the L2A product ("L2A_SceneClass", henceforth SLCC-MSI) described in Richter et al. (2012) in our evaluation. The SLCC-MSI is also based on a thresholding of the NDSI, in addition to many other tests to diagnose the pixel land cover and cloud status. We simplified the SLCC-MSI classification to be more relevant for snow-cover mapping in the following way:

- Shadow, saturated, dark area, cloud shadows, and unclassified pixels were reclassified as 'missing' pixels.
- Cloud medium probability, cloud high probability, and thin cirrus pixels were all reclassified as 'cloudy' pixels.
- Snow pixels remained 'snow' pixels.
- Vegetation, non-vegetation, and water pixels were reclassified as 'snow-free' pixels.

For the Sentinel-2 MSI imagery we manually selected cloud free scenes over our AOI. This resulted in a total of 24 Sentinel-2 scenes with accompanying reference snow-cover maps for the 2016 and 2017 ablation seasons (May-July).

2.3.4. Normalized difference snow index

The NDSI is the basis for the simpler snow-cover retrievals evaluated in this study. It exploits the fact that snow is highly reflective in the visible but a strong absorber in the shortwave infrared, which differentiates snow from most other natural surfaces (Dozier et al., 2009). The NDSI is defined as (Dozier, 1989)

$$\mathrm{NDSI} = (r_{\mathrm{Green}} - r_{\mathrm{SWIR1}})/(r_{\mathrm{Green}} + r_{\mathrm{SWIR1}})\,, \tag{1}$$

where r_{Green} is reflectance in the green band and r_{SWIR1} is the first shortwave infrared reflectance band (typically around 1.6 μm). In general, these may be either TOA or surface reflectances.

For version 6 of the MOD10A1 & MYD10A1 products the NDSI (based on TOA reflectances) is used when assigning a confidence on within-pixel snow presence through the 'NDSI snow-cover' field (Riggs et al., 2017). For pixels for which this NDSI snow-cover field is defined, the fSCA can be obtained through a simple linear relationship of the form

$$\mathrm{fSCA} = \beta_1 \cdot \mathrm{NDSI} + \beta_0\,, \tag{2}$$

where the β_i are regression coefficients. Equation (2) is then subject to the physical constraints $\mathrm{fSCA} \in [0, 1]$. In our study, we use the 'universal FRA6T' relationship from Salomonson and Appel (2004), where $\beta_1 = 1.45$ and $\beta_0 = -0.01$. This relationship was developed by correlating MODIS NDSI with fSCA obtained from coincident Landsat 7 Enhanced Thematic Mapper-Plus scenes (Salomonson and Appel, 2004). Note that employing separate relationships for Aqua MODIS is not necessary for version 6, since Aqua band 6 has been restored to scientific quality through the quantitative image restoration technique (Riggs et al., 2017). For version 5 of the MOD10A1 & MYD10A1 products, fSCA is provided directly as a field based on separate linear relationships for Aqua and Terra (Salomonson and Appel, 2004). In addition to the MODIS products, we also apply the linear regression in (2) to Sentinel-2 MSI and Landsat 8 OLI derived NDSI to gauge the applicability of this relationship at higher resolution.

The NDSI can also be used directly to obtain binary snow cover maps. This is especially appealing for the higher resolution sensors on board Landsat 8 and the Sentinel-2 satellites, since the structure of individual snow patches is more readily resolved which moderates the problem of mixed pixels. To binarize pixels by classifying them as either bare or snow-covered,

a threshold for the NDSI is required. Based on experience, several studies suggest a fixed threshold of 0.4 (e.g. Hall et al., 2002; Gascoin et al., 2015; Aalstad et al., 2018) which is often sufficiently accurate to discriminate snow and non-snow covered pixels. On the other hand, unsupervised adaptive detection algorithms can determine scene-specific thresholds that may yield superior results (Yin et al., 2013; Härer et al., 2018). In this study, we compare both techniques, using a fixed (0.4) and Otsu's segmentation algorithm (Otsu, 1979) as an adaptive thresholding technique which performed best in the intercomparison of Yin et al. (2013). For an (adaptive or fixed) NDSI threshold τ, we define the binary snow-cover status of a pixel for the Sentinel-2 and Landsat 8 reflectance products through the following rules:

$$\mathrm{fSCA} = \begin{cases} 1 & \text{if NDSI} > \tau \text{ and } r_{\mathrm{Red}} > 0.12 \text{ and } r_{\mathrm{SWIR1}} < 0.16\,, \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where we again use either TOA or surface reflectances. The additional thresholds in (3), applied to the reflectances in the red (r_{Red}) and SWIR1 (r_{SWIR1}, around 1.6 μm) bands are based on Dozier (1989) and Gascoin et al. (2015). The condition on the red band accounts for self, cast, or cloud shadows, while the threshold on the SWIR1 band helps to distinguish clouds (that may have a high NDSI) from snow. In agreement with Dozier (1989), the exact values of these individual band thresholds is not critical and we found little sensitivity to small (± 0.05) variations. Once (3) has been evaluated on a pixel level, the results are aggregated through spatial averaging to a desired coarser resolution to obtain the fSCA.

2.3.5. Spectral unmixing

fSCA retrieved using spectral unmixing (SU) of multispectral satellite imagery, reviewed in Keshava and Mustard (2002), was also employed in this study. The MODSCAG product detailed in Painter et al. (2009) applies unmixing to version 5 of the MOD09GA (Vermote and Wolfe, 2015a) product, and we base our unmixing procedure loosely on the MODSCAG algorithm. Other implementations of unmixing for snow-cover mapping include Vikhamar and Solberg (2003), Sirguey et al. (2009), and Cortés et al. (2014). In our implementation of spectral unmixing we solve the fully constrained least squares problem, i.e. we seek an optimal abundance $\mathbf{a}^\star$, which satisfies

$$\mathbf{a}^\star = \underset{\mathbf{a}}{\operatorname{argmin}} \left| \mathbf{M} \cdot \mathbf{a} - \mathbf{r} \right|^2 \text{ subject to } \mathbf{a} \geq \mathbf{0} \text{ and } \mathbf{1}^{\mathrm{T}} \cdot \mathbf{a} = 1\,, \tag{4}$$

where $\left| \cdot \right|$ is the ℓ^2-norm, the $N_b \times N_m$ (number of bands times number of endmembers) matrix $\mathbf{M}$ contains the theoretical reflectances for each endmember (columns) averaged to the bands of the sensor in question (rows), the $N_m \times 1$ vector $\mathbf{a}$ contains the fractional abundances of each endmember, the $N_b \times 1$ vector $\mathbf{r}$ contains the retrieved surface reflectances in each band, and $\mathbf{1}^{\mathrm{T}}$ is a $1 \times N_m$ vector of ones. The physical constraints on the fractional abundances of the endmembers are the abundance non-negativity constraint (ANC: $\mathbf{a} \geq \mathbf{0}$) and the abundance sum-to-one constraint (ASC: $\mathbf{1}^{\mathrm{T}} \cdot \mathbf{a} = 1$). The conversion from spectral reflectance, $r(\lambda)$, to band reflectance, r_b, is carried out through the following averaging operation:

$$r_b = \int_0^\infty r(\lambda)\phi_b(\lambda)\, d\lambda\,, \tag{5}$$

where $\phi_b(\lambda)$ is the sensor specific spectral response function (normalized to a probability density) for band b. In the SU experiments, we use the following six bands $b \in \{\text{Blue, Green, Red, NIR, SWIR1, SWIR2}\}$ for all the sensors. As an example, the spectral response functions for the bands employed for MSI on-board Sentinel-2A are shown in Figure 4(a). In case of ambiguity, we use the band with the higher spatial (as opposed to radiometric) resolution, e.g. for MSI we employ near infrared band 8 and not band 8a (see Müller-Wilm, 2018).

In practice, (4) can be solved using the fully constrained least squares unmixing approach described in Heinz and Chang (2001). This approach first augments both $\mathbf{M}$ and $\mathbf{r}$ with an extra row of ones to deal with the ASC and subsequently employs the non-negative least squares algorithm of Lawson and Hanson (1995) to enforce the ANC. In our study, theoretical spectra for the non-snow endmembers were obtained from the JPL spectral library (Baldridge et al., 2009). In particular, the non-snow endmembers employed were: vegetation ('grass'), soil ('pale brown dry silty clay loam'), and rock ('gray sandstone'). These non-snow endmembers were chosen based on knowledge of the typical land cover of the study area (see Boike et al., 2018) and are visualized in Figure 4(b). We also added a shade endmember (zero reflectance in all bands) to account for shadows in the unmixing procedure. The snow endmember spectra were obtained from a look up table generated by running the SNICAR radiative transfer model (He et al., 2018) with varying solar zenith angles (bins of 5°) and 7 different effective snow grain radii in the range 100-5000 μm. For simplicity, light absorbing impurities in snow are ignored in this study. After fixing the solar zenith angle to the nearest value

in the look up table, we looped over all possible grain radii and selected the one that minimized the root mean square residual error in (4). Some example snow endmember spectra are shown in Figure 4(b). As in Cortés et al. (2014), we ignore azimuthal effects since, despite the strong forward scattering properties of snow, these should not be significant for near nadir looking sensors on-board the Landsat 8 and Sentinel-2 satellites. For the off-nadir looking MODIS sensor this assumption may be more problematic. After obtaining the optimal $\mathbf{a}^\star$ across all radii, the fSCA was obtained by dividing the snow endmember abundance by the total unshaded abundance. The assumption in this normalization is that the relative abundances are the same in the shaded and non-shaded regions. All the non-snow endmembers and a subset of the snow endmembers are visualized in Figure 4. Independent SU experiments were undertaken for all the sensors employed in this study. To speed up the unmixing, an initial NDSI screening was applied: the algorithm was only executed for pixels with a NDSI> 0 as only these pixels were deemed to possibly have fSCA> 0.

2.3.6. Shadow masking

The fSCA retrieval algorithms do not work well if a considerable fraction of a satellite pixel is covered in shadow (see Figure 5(b)). For the higher resolution (OLI & MSI) sensors we diagnosed cast and self shadowing for each scene using a dynamic shadow mask. This was achieved in a two stage process. First the solar geometry (solar azimuth and elevation angle) was diagnosed for the AOI at the scene sensing time using the ephemeris routine in the PV_LIB toolbox (SNL, 2014). Our AOI is small enough that the solar geometry is invariant to a good approximation. The geometry was then used in conjunction with the TopoToolbox (Schwanghart and Scherler, 2014) and a 5 m GSD DEM of the entire Brøgger peninsula produced by the Norwegian Polar Institute (NPI, 2014) to calculate the self and cast shadows within the AOI. To account for uncertainties in the ephemeris and the employed DEM, we dilated the resulting shadow mask by one DEM pixel. The shadow mask is thus conservative in that it purposely slightly overestimates the shadowed area. For each satellite scene the binary shadow mask at the DEM resolution was aggregated to obtain fractional shadows at the resolution of the satellite image. For some satellite scenes, Scheteligfjellet casts a shadow over the AOI, so it was important to mask this out in an automatic fashion without having to discard the entire scene. An example of such a shadow and independently derived shadow mask is shown in Figure 5. Cast shadows did not affect any

of the reference fSCA maps derived from terrestrial imagery.

2.3.7. Aggregation

All the higher resolution retrievals from MSI and OLI are aggregated to coarser pixels at 100 m and 500 m resolution. The aggregation is performed in a conservative manner using a form of inverse distance weighting. For each coarse pixel, subpixels that either partially or fully fall within this pixel are weighted according to the fraction of their area that lies inside the coarse pixel. If more than 25% of a coarse pixel consists of missing data (e.g. if it has a shade fraction $> 25\%$), the pixel is excluded from further analysis. A scale of 100 m was chosen in line with typical reanalysis experiments (e.g. Girotto et al., 2014; Cortés et al., 2016; Aalstad et al., 2018), since this scale has been shown to resolve the key modes of snow spatial variability. The scale of 500 m was chosen in order to compare the performance of the higher resolution retrievals with the MODIS retrievals. As an example, Figure 6 visualizes how well some of the satellite retrieval algorithms for Sentinel-2 MSI can replicate the coincident reference snow cover prior to aggregation. For the regression and SU experiments that (unlike thresholding) provide fSCA without the need for spatial aggregation, we also performed the evalation at the native pixel scale of the sensors.

Abbreviation	Description	Sensor	Product	Reflectance
FT-MSI-TOA	Fixed threshold	S2A/B MSI	No	TOA
AT-MSI-TOA	Adaptive threshold	S2A/B MSI	No	TOA
FT-MSI-SFC	Fixed threshold	S2A/B MSI	No	Surface
AT-MSI-SFC	Adaptive threshold	S2A/B MSI	No	Surface
SLCC-MSI	Scene classification	S2A/B MSI	Yes	Surface
SU-MSI	Spectral unmixing	S2A/B MSI	No	Surface
RG-MSI-TOA	Regression	S2A/B MSI	No	TOA
RG-MSI-SFC	Regression	S2A/B MSI	No	Surface
FT-OLI-TOA	Fixed threshold	LS8 OLI	No	TOA
AT-OLI-TOA	Adaptive threshold	LS8 OLI	No	TOA
FT-OLI-SFC	Fixed threshold	LS8 OLI	No	Surface
AT-OLI-SFC	Adaptive threshold	LS8 OLI	No	Surface
Fmask-OLI	Scene classification	LS8 OLI	Yes	Surface
SU-OLI	Spectral unmixing	LS8 OLI	No	Surface
RG-OLI-TOA	Regression	LS8 OLI	No	TOA
RG-OLI-SFC	Regression	LS8 OLI	No	Surface
MOD10A1-v5	MOD10A1 version 5	Terra MODIS	Yes	TOA
MOD10A1-v6	MOD10A1 version 6	Terra MODIS	Yes	TOA
MYD10A1-v5	MYD10A1 version 5	Aqua MODIS	Yes	TOA
MYD10A1-v6	MYD10A1 version 6	Aqua MODIS	Yes	TOA
MODSCAG	MOD09GA v5 unmixing	Terra MODIS	Yes	Surface
SU-MOD09GA	MOD09GA v6 unmixing	Terra MODIS	No	Surface
SU-MYD09GA	MYD09GA v6 unmixing	Terra MODIS	No	Surface

Table 1: Overview of the fSCA retrievals that we evaluated. fSCA retrievals that are not operational products were performed on reflectances only as a part of this study.

2.4. Evaluation metrics

We perform the evaluation by pairing coincident (i.e. same day and spatial area) pixels from the satellite fSCA retrievals (the estimate) and the reference fSCA retrievals (the 'truth'). The number of sample pairs is different depending on the type of satellite retrieval. For example, there are many more MODIS scenes than there are Sentinel-2 scenes, so the number of MODIS samples at 500 m resolution is considerably larger. At this resolution, the number of sample pairs for Sentinel-2 MSI (144) is nonetheless

more than enough for a robust evaluation. The estimate-truth pairs are used to compute various evaluation metrics, most of which are typical of earlier fSCA evaluation studies (e.g. Rittger et al., 2013; Gascoin et al., 2015; Masson et al., 2018). The evaluation metrics can either be fractional, being derived directly from the fSCA, or binary, derived from binary snow cover (snow-covered or bare) obtained through thresholding the fSCA.

For the fractional metrics, the fSCA error of each sample pair is computed by subtracting the true fSCA from the estimated fSCA. Based on this error we compute the bias (mean error) as a measure of systematic differences and the root mean squared error (RMSE) as a measure of spread. In addition, we also compute the correlation coefficient to measure the strength of the linear correlation (R) between the truth and the estimate. In this study, the truth and the estimate are strongly and positively correlated. We therefore report R^2 instead to better distinguish between high correlations. Moreover, we perform a standard scatter plot analysis where best fit lines are computed using simple linear regression through ordinary least squares estimation.

For the binary metrics, we first convert fSCA to a binary snow-covered or bare pixel status, for which a threshold in the range $[0, 1]$ must be defined. Previous studies have used various thresholds, for example 0.15, 0.5 (Rittger et al., 2013) or 0 (Masson et al., 2018). Since fSCA can be interpreted as the probability of a randomly selected point in a pixel being snow-covered, it is difficult to justify a specific threshold. As the binary evaluation is nonetheless highly sensitive to this choice, we employ three thresholds representing various degrees of fSCA in this study, specifically: $0.1, 0.5$, and 0.9.

Given a specific threshold, which we apply to both the true and estimated fSCA, we may diagnose the number of true positives (TP, both are snow-covered), true negative (TN, both are bare), as well as false positives (FP, estimate is snow-covered while truth is bare) and false negatives (FN, estimate is bare while truth is snow-covered). Based on these counts we can compute various binary metrics. First we diagnose the positive predictive value (also called precision) which is the ratio of true positives to the total number of estimated positives (TP+FP). Then we compute the true positive rate (TPR, also known as recall and hit rate) which is the ratio of the number of true positives to the number of actual positives in the true data (TP+FN). We also diagnose the false positive rate (FPR) which is the ratio of the number of false positives to the actual number of negatives (TN+FP).

By varying the estimate binarization threshold from 0 to 1 and keeping the true binarization threshold fixed, TPR and FPR can be plotted against

one another to construct receiver operating characteristics (ROC) curves as described in Fawcett (2006). Briefly, the ROC curve starts in the top right corner of ROC space (TPR vs FPR) for an estimator binarization threshold of $-\infty$ where FPR=TPR=1 and ends in the bottom left corner for a binarization threshold of ∞ where FPR=TPR=0. In the ROC curves we emphasize the points where the two binarization thresholds are equal since for a perfect classifier these should fall in the top left corner of ROC space. We employ the area under the ROC curve (denoted $\mathcal{A}$) to summarize the the ROC performance in a single number. For an ideal classifier $\mathcal{A} = 1$. Conversely, $\mathcal{A} = 0.5$ indicates that the classifier is no better than a random guess. The extreme case $\mathcal{A} = 0$ indicates that the classifier has its labels completely reversed (i.e. positive is negative and vice versa). In addition to the area under the curve, in line with Rittger et al. (2013) and Masson et al. (2018), we also calculate the F-score which is the harmonic mean of the positive predictive value and the true positive rate, where a score of F= 1 is perfect and F= 0 is the worst possible score. The F-score penalizes both errors of commission and omission. For both $\mathcal{A}$ and F we employ the three previously mentioned binarization thresholds. These two metrics are fundamentally different: F is a mean measure of how well an algorithm detects positives (true negatives are not considered), whereas $\mathcal{A}$ is the probability that the algorithm ranks a positive higher than a negative.

3. Results

3.1. Annual snow-cover depletion

We present results for the remotely sensed snow-cover depletion, aggregated to the scale of the AOI, for the six ablation seasons considered in this study (2012-2017). Figure 7 shows time series of the fSCA depletion as sensed by the ACS and a subset of the satellite retrieval algorithms. The years 2012, 2013, 2015, and 2017 can all be characterized as normal snow years with the snowpack completely disappearing towards the end of June/early July (see black dots in Figure 7), around a month long melt season, and a peak snow depth at the BCS around 1 m (Boike et al., 2018). In 2014 the snowpack was unusually high, persisting until mid July (Aalstad et al., 2018) with a peak depth at the BCS of 1.6 m. However, in 2014, the ACS malfunctioned, so that ground truth imagery was not available in most of the melt season (c.f. Figure 7). In 2016, the snowpack had completely disappeared already

in mid June, with a peak depth of 0.8 m at the BCS and a melt period of only half a month.

Using the ACS as a reference, Figure 7 shows that the plotted MODIS products (MOD10A1 & MYD10A1 version 6) tend to slightly overestimate the fSCA during the snowmelt. While there are a few gaps in the snowmelt season due to clouds, these products are nonetheless still able to identify the end of the snowmelt season fairly well. The SU retrievals from Sentinel-2 MSI and Landsat 8 OLI do not suffer from the same positive bias as the MODIS products during the snowmelt period. The unmixing retrievals are also remarkably accurate throughout the entire ablation season. Still, given the lower revisit frequency, there are slightly fewer retrievals than for MODIS, at least prior to 2017 and the launch of Sentinel-2B. Due to orbital convergence, the revisit frequency of the higher resolution sensors (OLI and MSI) is considerably higher at high latitudes than at the equator. In fact, Sentinel-2 MSI provides a near daily (as opposed to 8 day) revisit frequency in the high-Arctic. Using all the satellite sensors together (see 2017 in Figure 7), it is possible to get quite a robust reconstruction of the true fSCA depletion curve at the scale of our AOI.

3.2. Fractional evaluation

Scatter plots of all the available sample pairs for the various fSCA satellite retrievals at 100 m and 500 m resolution are shown in Figure 8 and Figure 9, respectively. The fractional metrics that summarize the performance visualized in these scatter plots, namely bias, RMSE, and R^2, are listed in Table 2, Table 3, and Table 4 for the 100 m, 500 m, and pixel-scale retrievals, respectively. We begin by describing the evaluation of the higher resolution retrievals.

Figure 8 shows that all thresholding based algorithms at 100 m resolution have a tendency to overestimate the fSCA for both the OLI and MSI sensors. The overestimation is visible through the clustering of scatter points above the 1:1 line (blue) especially when the reference fSCA is higher than 0.4. This positive bias results in linear best fit lines with positive intercepts that lie above but run nearly parallel to the 1:1 line. For all thresholding retrievals, the biases are on the order of $0.03 - 0.05$ for MSI and $0.05 - 0.07$ for OLI. Of the thresholding experiments, the built in scene classifications, i.e. Fmask-OLI and SLCC-MSI, perform worst in terms of the fractional evaluation metrics, showing the highest combined bias and RMSE and the lowest correlation coefficient. The performance of SLCC-MSI is particularly

poor with a 34% reduction in the number of available samples compared to the other MSI experiments. For Fmask-OLI the reduction in the number of samples, compared to the other OLI retrievals, is only 5%. For both scene land cover classifications, the reduction in the number of samples is due to cloud commission errors.

Name	N	Bias	RMSE	R^2	$\mathcal{A}_{0.1}$	$\mathcal{A}_{0.5}$	$\mathcal{A}_{0.9}$	$\mathrm{F}_{0.1}$	$\mathrm{F}_{0.5}$	$\mathrm{F}_{0.9}$
FT-MSI-TOA	3960	0.03	0.10	0.96	0.99	1.00	0.94	0.98	0.98	0.92
AT-MSI-TOA	3960	0.05	0.12	0.94	0.99	1.00	0.96	0.95	0.97	0.93
FT-MSI-SFC	3960	0.03	0.08	0.97	1.00	1.00	0.96	0.99	0.98	0.93
AT-MSI-SFC	3960	0.04	0.10	0.96	1.00	1.00	0.95	0.97	0.97	0.92
SLCC-MSI	2628	0.05	0.15	0.85	1.00	0.92	0.81	1.00	0.98	0.90
SU-MSI	3960	-0.00	0.07	0.98	1.00	1.00	0.98	0.98	0.99	0.90
RG-MSI-TOA	3960	0.02	0.08	0.97	0.99	1.00	0.98	0.97	0.98	0.95
RG-MSI-SFC	3960	0.02	0.07	0.98	1.00	1.00	0.98	0.98	0.99	0.95
FT-OLI-TOA	4205	0.05	0.13	0.92	0.99	0.99	0.87	0.99	0.97	0.90
AT-OLI-TOA	4205	0.06	0.14	0.90	0.99	0.99	0.88	0.98	0.97	0.90
FT-OLI-SFC	4205	0.05	0.13	0.92	0.99	0.99	0.87	0.98	0.97	0.91
AT-OLI-SFC	4205	0.06	0.14	0.90	0.99	0.99	0.86	0.98	0.97	0.90
Fmask-OLI	3980	0.07	0.17	0.83	0.99	0.95	0.80	0.98	0.96	0.88
SU-OLI	4205	0.02	0.09	0.96	1.00	1.00	0.96	0.98	0.98	0.91
RG-OLI-TOA	4205	0.04	0.11	0.93	0.99	1.00	0.91	0.98	0.98	0.90
RG-OLI-SFC	4205	0.04	0.10	0.94	0.99	1.00	0.92	0.99	0.98	0.92

Table 2: Summary of evaluation metrics, i.e., number of samples (N), Bias, RMSE, square correlation coefficient (R^2), area under the ROC curve ($\mathcal{A}$), and F-score (F) for the satellite retrievals at 100 m resolution. The subscript for the binary metrics ($\mathcal{A}$ and F) indicates the fSCA binarization threshold employed. For an overview of the various satellite retrievals see Table 1.

For both OLI and MSI, there is no noticeable performance gain when using an adaptive (AT) instead of a fixed (FT) threshold on the NDSI. In fact, for all the experiments at 100 m resolution, the adaptive threshold performs slightly worse than the corresponding fixed threshold, exhibiting a higher bias and RMSE. Based on the scatter plots, the most visible discrepancy between the adaptive and fixed thresholds are at low reference fSCA, with the adaptive threshold leading to a larger overestimation as the true fSCA

approaches zero. Using surface instead of TOA reflectances for MSI (see e.g. FT-MSI-TOA vs FT-MSI-SFC in Table 2), i.e. applying an atmospheric correction, only results in a marginal improvement in performance, while there is no difference for OLI (Table 2, Figure 8). Using the regression relationship in (2), quite precise retrievals were obtained with considerably lower RMSE than the thresholding based retrievals for OLI and MSI, but a positive (albeit lower) positive bias remained. For these regression-based retrievals there was a marginal improvement in RMSE and correlation when using surface rather than TOA reflectances. For both OLI and MSI, the SU technique shows the best performance. For the unmixing, the positive bias is almost eliminated (0.02 for OLI and 0.00 for MSI) and the linear best fit line closely tracks the 1:1 line. In addition, the RMSE is around half of the worst performing thresholding experiments. While the unmixing experiments still show a slight tendency towards overestimating fSCA at intermediate $(0.25 - 0.75)$ fSCA, it is smaller than for the thresholding retrievals and not visible at low and high fSCA values.

At 500 m resolution, fSCA retrievals from both the moderate (MODIS) and higher resolution (OLI and MSI) sensors are available (Figure 9). For all higher resolution sensors, the RMSE is reduced by around 0.02 (see Table 3) due to averaging of random errors in the spatial aggregation procedure. The reduction in random error also results in a higher linear correlation for these retrievals at 500 m than at 100 m resolution. Note that the bias is not reduced accordingly as bias is a systematic error that does not cancel out upon spatial aggregation. As for the 100 m resolution retrievals, the overestimation of fSCA in the thresholding experiments is clearly visible, particularly at intermediate fSCA. This effect is still present, but less pronounced for the OLI and MSI SU retrievals. The low number of samples for the SLCC-MSI stands out in Figure 9 and complicates interpretation.

For the $\sim$ 500 m resolution MODIS retrievals, products based on the regression in (2) (version 5 and 6 of MOD10A1 and MYD10A1), as well as unmixing retrievals from the MODSCAG product and experiments with location specific endmembers are available (SU-MYD09GA and SU-MOD09GA). On the one hand, there is typically considerably more scatter than for the higher resolution sensors, which leads to a generally higher RMSE for the MODIS retrievals. On the other hand, the biases are typically around the same magnitude for the MODIS retrievals $(0 - 0.09)$ and the higher resolution sensors $(0.01 - 0.08)$. We emphasize that the 500 m evaluation is at the scale of individual pixels for MODIS, but combines approximately 275

aggregated (30 m GSD) OLI pixels and 2500 (10 m GSD) MSI pixels for the higher resolution sensors.

Name	N	Bias	RMSE	R^2	$\mathcal{A}_{0.1}$	$\mathcal{A}_{0.5}$	$\mathcal{A}_{0.9}$	$\mathrm{F}_{0.1}$	$\mathrm{F}_{0.5}$	$\mathrm{F}_{0.9}$
FT-MSI-TOA	144	0.03	0.08	0.98	0.99	1.00	0.94	0.98	0.97	0.92
AT-MSI-TOA	144	0.05	0.09	0.97	0.98	1.00	0.98	0.90	0.97	0.93
FT-MSI-SFC	144	0.03	0.06	0.98	1.00	1.00	0.96	0.99	0.97	0.93
AT-MSI-SFC	144	0.04	0.08	0.98	1.00	1.00	0.96	0.95	0.97	0.93
SLCC-MSI	80	0.04	0.11	0.67	1.00	1.00	0.85	1.00	1.00	0.91
SU-MSI	144	-0.01	0.05	0.99	1.00	1.00	0.98	0.98	0.98	0.93
RG-MSI-TOA	144	0.02	0.06	0.98	0.99	1.00	0.98	0.97	0.98	0.94
RG-MSI-SFC	144	0.02	0.05	0.99	1.00	1.00	0.98	0.98	0.98	0.94
FT-OLI-TOA	152	0.05	0.10	0.95	1.00	1.00	0.93	0.98	0.99	0.88
AT-OLI-TOA	152	0.06	0.11	0.94	0.99	1.00	0.94	0.97	0.98	0.88
FT-OLI-SFC	152	0.05	0.10	0.95	1.00	1.00	0.93	0.99	0.99	0.88
AT-OLI-SFC	152	0.06	0.11	0.94	1.00	1.00	0.92	0.97	0.98	0.88
Fmask-OLI	143	0.08	0.15	0.88	0.99	1.00	0.84	0.98	0.97	0.85
SU-OLI	152	0.03	0.07	0.97	1.00	1.00	0.96	0.99	0.99	0.88
RG-OLI-TOA	152	0.04	0.09	0.95	0.99	1.00	0.93	0.96	0.99	0.88
RG-OLI-SFC	152	0.04	0.09	0.96	1.00	1.00	0.96	0.99	0.99	0.90
MOD10A1-v5	409	0.05	0.15	0.91	1.00	0.97	0.83	0.99	0.96	0.83
MOD10A1-v6	489	0.05	0.14	0.91	0.99	0.98	0.86	0.98	0.96	0.84
MYD10A1-v5	361	0.03	0.11	0.95	0.99	0.99	0.85	0.99	0.99	0.85
MYD10A1-v6	447	0.04	0.11	0.95	0.98	0.99	0.86	0.96	0.99	0.87
MODSCAG	383	0.05	0.26	0.67	0.94	0.96	0.85	0.85	0.87	0.56
SU-MYD09GA	481	0.03	0.18	0.85	0.99	0.99	0.85	0.90	0.97	0.75
SU-MOD09GA	638	-0.00	0.16	0.87	0.99	0.98	0.85	0.97	0.96	0.65

Table 3: Same as Table 2, but for the satellite retrievals at 500 m resolution.

The regression relationship (2) of Salomonson and Appel (2004) with the 'universal FRA6T' coefficients was applied to the to the NDSI obtained from the 'NDSI snow-cover' field in version 6 of the MOD10A1 & MYD10A1 products. This brought the version 6 product in line with version 5, with a negligible difference in performance. For the bias and RMSE the largest changes had a value of 0.01 going form version 5 to version 6, whereas the

square correlation coefficient remained the same. Note that there are considerably more (23% for Aqua and 20% for Terra) retrievals available in the newer version 6 given that version 5 was discontinued after 2016.

Name	N	Bias	RMSE	R^2	$\mathcal{A}_{0.1}$	$\mathcal{A}_{0.5}$	$\mathcal{A}_{0.9}$	$\mathrm{F}_{0.1}$	$\mathrm{F}_{0.5}$	$\mathrm{F}_{0.9}$
SU-MSI	417945	-0.00	0.14	0.92	0.99	0.99	0.98	0.96	0.96	0.87
RG-MSI-TOA	417945	0.02	0.15	0.90	0.99	0.98	0.96	0.95	0.96	0.94
RG-MSI-SFC	417945	0.02	0.14	0.91	0.99	0.99	0.97	0.97	0.96	0.94
SU-OLI	48628	0.02	0.14	0.89	0.99	0.99	0.95	0.97	0.96	0.91
RG-OLI-TOA	48628	0.04	0.14	0.90	0.99	0.98	0.91	0.97	0.97	0.91
RG-OLI-SFC	48628	0.04	0.16	0.87	0.99	0.97	0.90	0.98	0.96	0.93

Table 4: Same as Table 2, but for the higher-resolution fSCA satellite retrievals at the pixel-scale.

Surprisingly, the empirically based MOD10A1 & MYD10A1 for the majority of the fractional evaluation metrics (see Table 3) outperform the unmixing based MODIS retrievals. Of the latter, MODSCAG performs worst with the algorithm struggling to recover the true fSCA in the low (< 0.25) and high (> 0.75) snow-cover cases, and with considerable scatter for intermediate fSCA. To further investigate this unfavorable performance, we performed two custom experiments: SU-MOD09GA and SU-MYD09GA. Unlike the MODSCAG products for the AOI, these custom unmixing experiments are based on version 6 (as opposed to version 5) of the MODIS surface reflectance products (MOD09GA, MYD09GA). Moreover, we considered AOI specific endmembers for these custom MODIS unmixing experiments, whereas the endmembers used in MODSCAG are unknown to us. The custom MODIS unmixing experiments performed slightly better than MODSCAG with a marked reduction in RMSE and increase in correlation. Notably, the custom experiments performed better in high fSCA situations, but still had problems in the low fSCA case. Despite the slight improvements relative to MODSCAG, the performance of these custom unmixing experiments was still worse than the simpler MOD10A1 & MYD10A1 products. At the same time, it is worth emphasizing that the SU-based MODIS retrievals feature an equal or lower bias than the regression-based MOD10A1 & MYD10A1 products.

The results for the higher-resolution sensors (OLI and MSI) at the pixel-scale largely mirrors the results at the 100 m scale. That is, the regression-based retrievals obtain a similar RMSE to the SU-based retrievals but with a somewhat higher positive bias. As expected, due to the lack of spatial averaging, the RMSE is somewhat larger at this higher resolution while the bias remains the same. Note that the thresholding-based retrievals can not be evaluated at this pixel scale as they can only provide binary (snow or bare) estimates prior to spatial aggregation.

3.3. Binary evaluation

Receiver operating characteristics (ROC) curves for the binarized fSCA satellite retrievals at 100 m and 500 m resolution are shown in Figure 10 and Figure 11, using different binarization thresholds (fSCA=$0.1, 0.5$ and 0.9, see Section 2.4). The binary metrics that summarize the performance of the binarized retrievals, namely the area under the ROC curve ($\mathcal{A}$) and the F-score (F) are listed in Table 2, Table 3, and Table 4.

For all retrievals at 100 m resolution (Figure 10), the area under the curve is always lowest for a reference binarization threshold of fSCA= 0.9 (blue curve). This indicates that the algorithms perform worse for high snow-cover situations compared to intermediate (fSCA= 0.5) and low (fSCA= 0.1) snow-cover situations. For the lower binarization thresholds, all algorithms perform well ($\mathcal{A} \geq 0.99$), with the exception of the native scene classifications. Considering all binarization thresholds, the regression and unmixing retrievals from MSI and OLI perform best with respect to $\mathcal{A}$, which is clearly visible in the ROC curves especially for the higher threshold. The native scene classifications (Fmask-OLI and SLCC-MSI) again perform worst. For the F-score, all algorithms perform favorably for the lower thresholds with F$>$ 0.95, indicating a very good performance both in terms of the positive predictive value (precision: ratio of true positives to predicted positives) and the true positive rate (recall: ratio of true positives to the number of actual positives). This implies that both the number of errors of commission (false positives) and omission (false negatives) are low. As for $\mathcal{A}$, the F-score is worst (lowest) for the higher binarization threshold. It is worth highlighting that the true positive rate (recall) is near unity (i.e. a perfect score) in the ROC curves for most of the retrievals when the retrieval binarization threshold equals to the reference binarization threshold (cases marked by circles in Figure 10). This suggests that low positive predictive values (precision) reduce the F-score from an ideal value of 1, indicating more errors of

commission than omission.

For the retrievals at 500 m resolution, we see a similar picture. Once more, the performance for the lower reference binarization thresholds is noticeably (see Figure 11) better than for the highest reference binarization threshold. For the lower thresholds, $\mathcal{A}$ exceeds 0.98 for all MSI and OLI retrievals, with the exception of the native scene classifications. Furthermore, the F-score shows a similar pattern as for the high resolution retrievals, with markedly worse scores for the highest binarization threshold. Once more, this is primarily due to a low positive predictive value (see the circular markers in Figure 11), indicating that the algorithms produce more false positives than false negatives. This finding is in line with the fractional evaluation metrics for which a positive bias occurred for most of the products. As previously noted, the bias does not disappear upon spatial aggregation, showing the same patterns in the binary evaluation at 100 m and 500 m. For the MODIS retrievals, we note that the high binarization threshold yields the lowest area under the ROC curves (Figure 11). As for the the fractional evaluation, MODSCAG performs considerably worse than both generations of the regression-based MODIS snow products with an area under the ROC curve well below unity for all the reference binarization thresholds. The performance is improved for the customized SU-MOD09GA and SU-MYD09GA experiments with both a higher $\mathcal{A}$ and F-score closer to those of the regression-based retrievals.

4. Discussion

4.1. Error analysis

In this study, we used high (0.5 m) resolution reference snow-cover maps obtained from accurately georeferenced (bias and RMSE on the order of 2 meters) orthorectified photographs. The high spatial resolution of these reference binary snow-cover maps means that they are less susceptible to the mixed pixel problem in that the majority of 0.5 m pixels are either fully snow-covered or fully snow-free. The estimated error in our reference fSCA retrievals due to the georeferencing error is on the order $\sigma_E = 0.01$ at 100 m (see Sect. 2.2), which is considerably smaller than the RMSE of the best performing satellite retrievals (e.g. SU-MSI with RMSE=0.07, see Table 2). Even in the infrequent cases of a 3σ georeferencing error, inaccurate georeferencing translates to an error on the reference fSCA of $\sigma_E = 0.03$, which is less than half the RMSE (based on the reference fSCA) of the best performing retrieval. Due to the effects of spatial averaging, the estimation standard

deviation for the reference fSCA at 500 m will be even lower. Thus, our estimates of the satellite retrieval error can closely represent the true error of the retrievals for our site. Although we had a relatively small (1.77 km^2) AOI, 302 high resolution reference images were available for evaluation, so that we consider our reference dataset to be sufficiently extensive for the evaluation of higher resolution imagery form OLI and MSI. For the moderate (500 m) resolution MODIS retrievals, our study area only contains 5 pixels which is rather low. Nonetheless, the fact that we had hundreds of reference images translated into a total of on the order 400 fSCA retrievals at 500 m resolution that we could evaluate for each of the MODIS products. Combined with the fact that our study area is relatively flat and representative of large swathes of Arctic tundra, the validation of MODIS fSCA retrievals at our study site becomes a valuable exercise. It should be noted, however, that we do not expect our evaluation to necessarily reflect the general performance of the global MODIS snow-cover products given the limited size of the evaluation area and the lack of global representativeness.

Earlier work (e.g. Cortés et al., 2014; Masson et al., 2018) has emphasized the need for an evenly sampled fSCA so that situations with no-snow or full snow-cover are not over-represented. These edge cases should be the easiest for any algorithm to capture, with intermediate partial snow-cover being more challenging. In our case, we only employ reference fSCA maps from the entire ablation season (see Figure 7), which ensures that the entire range from 0 to 1 is represented. However, the reference data set does not feature an even distribution of fSCA values, as many almost fully snow-covered and bare scenes were included. Our study shows that it is important to test the algorithms in these simple cases as well. As can be seen in Figure 10 and Figure 11, many of the algorithms did not perform well in situations with high snow-cover (fSCA$>$0.75). When computing fractional metrics, such as RMSE, we did not see the need (c.f. Masson et al., 2018) to remove situations with zero snow-cover since the MODIS retrievals in particular had some large excursions from the truth in situations with zero fSCA. Such excursions are clear signs of cloud contamination and errors of omission in the the MODIS cloud mask. Cloud masking is a challenging task that we did not investigate explicitly (see Zhu and Woodcock, 2012); instead we manually selected the higher resolution imagery to ensure that it was cloud free. Cloud commission or omission errors are a large source of uncertainty in any automatic snow-cover mapping algorithm as discussed in Gascoin et al. (2015) and Stillinger et al. (2019).

For the binary evaluation, we considered more than just a single fSCA binarization threshold. With the ROC analysis, all the algorithms performed least favorably for high snow-cover, i.e. for a reference binarization threshold of 0.9. This would not have been clear if only a single low threshold (e.g. 0) would have been employed (e.g. Masson et al., 2018). The ROC analysis showed that low F-scores were due to low precision (errors of commission) rather than recall (errors of omission). This is connected to the positive bias in the thresholding algorithms due to the mixed-pixel problem. One benefit of a ROC analysis over other binary evaluations is that it is quite insensitive to a skewed data distribution, i.e. to the distribution of positives and negatives in the reference data. Moreover, ROC evaluation is natural when one considers fSCA defined as the probability of occurrence of snow-cover within a pixel since the ROC analysis is specifically designed to deal with probabilistic classification (Fawcett, 2006).

The evaluation showed that the errors for most of the retrievals were dependent on the fSCA itself (i.e. heteroscedastic). For most of the retrievals, this heteroscedasticity manifests itself (see Figure 12) through errors that are higher for intermediate fSCA and lower for high and low fSCA. Our results suggest that the error variance in satellite retrieved fSCA can be parametrized through the following expression (after Sakov et al., 2012)

$$\sigma^2_{\text{obs}} = \sigma^2_0 + \left(\frac{\sigma^2_1 - \sigma^2_0}{\alpha} \right) (\alpha - |\alpha - \text{fSCA}|) \tag{6}$$

where σ^2_0 is the minimum error variance, σ^2_1 is the maximum error variance, and we propose $\alpha = 0.5$ for symmetry. Thereby, the error variance is at its minimum for full or zero fSCA and at its maximum for intermediate fSCA. For spectral unmixing at 100 m resolution, appropriate values for the error standard deviations are $\sigma_0 = 0.05$ and $\sigma_1 = 0.1$. Expression (6) can be a useful addition for fSCA data assimilation studies which so far have typically employed a fixed observation error variance (e.g. Girotto et al., 2014; Cortés et al., 2016; Aalstad et al., 2018).

4.2. Thresholding algorithms

The native scene land cover classification algorithms Fmask-OLI and SLCC-MSI are designed to conservatively map changes in land use and land cover by masking out undesired artifacts, such as clouds or snow (see Zhu and Woodcock, 2012). Perhaps not surprisingly, they perform worst of all

the thresholding algorithms with respect to snow cover mapping. Both the Fmask (Zhu and Woodcock, 2012) and SLCC-MSI (Richter et al., 2012) algorithms perform an expansion of the snow mask through a dilation operation, which likely explains the high bias. While all the selected scenes were cloud free, SLCC flagged a considerable amount of snow-covered pixels as cloudy, which led to an omission of 34% of the valid data. The cloud commission errors were not as severe in Fmask, with only 5% data loss. The fact that OLI (in contrast to MSI) has a thermal band may explain some of the differences in cloud masking, although temperature based cloud masking can be challenging over snow (e.g. Westermann et al., 2012; Østby et al., 2014). For larger scale snow-mapping, there is potential to improve the cloud cover masking in the SLCC-MSI algorithm, for example following Gascoin et al. (2019) and Stillinger et al. (2019).

For the remaining thresholding retrievals, no appreciable difference was found between using top of the atmosphere or surface reflectances. This most likely can be explained with the relative insensitivity to atmospheric scattering and absorption of the green and SWIR band, which are employed for the NDSI (Dozier, 1989; Riggs et al., 2017). Accordingly, Yin et al. (2013) found little performance gain when employing atmospheric correction for NDSI snow-cover mapping, and the MODIS NDSI-based snow-cover product does not employ any atmospheric correction (Riggs et al., 2017). As accurate atmospheric correction (see e.g. Vermote et al., 2016) can be difficult and often introduces additional uncertainty in the retrieval process, it is encouraging that NDSI-based fSCA detection performs well when employing TOA reflectances directly. Note that for MODIS Masson et al. (2018) demonstrated an improvement in NDSI snow-cover mapping when using atmospheric and topographically corrected reflectances in mountainous terrain. Our results may be an indication that the improvement in their results is largely due to the topographic correction.

Previous studies (e.g. Yin et al., 2013; Härer et al., 2018) suggest a performance gain when employing an adaptive (scene-dependent) NDSI threshold instead of a fixed threshold of 0.4. We tested the best performing algorithm in Yin et al. (2013), namely the binarization of Otsu (1979), only constraining the threshold to lie in the range [0.1,0.7] (which is reasonable for snow-covered pixels, Riggs et al., 2017). Our evaluation did not show an improvement in performance when using this algorithm (see Figure 8), and the performance was even slightly worse for low fSCA for which there was a noticeable increase in the positive bias. The algorithm tended to pick

a threshold near the lower limit of 0.1 for low fSCA scenes, which resulted in an overestimation of the snow-covered area, including several errors of commission over the Bayelva-river floodplain. While it is worth optimizing the location and scene-specific choice of the NDSI threshold, our results indicate that using an automatic thresholding algorithm may not be ideal in all situations. This is in line with findings of Härer et al. (2018) who only found small differences when implementing the segmentation method of Otsu (1979). This segmentation method does not work well when the foreground (snow) and background (bare) clusters are poorly defined, especially in the case that one cluster covers a much smaller portion of the scene than the other. In agreement with Härer et al. (2018), we do not recommend applying automatic NDSI threshold selection techniques without ground truth data to support their validity.

All thresholding algorithms exhibited a considerable positive fSCA bias (see Tables 2, 3), which compromises the use of aggregated fSCA maps produced by thresholding of higher resolution NDSI maps as a reference for moderate resolution retrievals from sensors, such as MODIS (e.g. Salomonson and Appel, 2004; Gascoin et al., 2015; Masson et al., 2018). In Rittger et al. (2013), the potential limitations of using a thresholding based retrieval as reference were pointed out. At least for our AOI, the errors in the aggregated threshold-based retrievals are on the same order of magnitude as the errors of some of the coarser-scale MODIS retrievals. Therefore, if the threshold-based retrievals were used as a reference, many of the MODIS retrievals would appear to perform much worse than in reality. In fact, the retrieval algorithms used by all MODIS products partly overcome the mixed pixel problem that affects all thresholding-based retrievals. For the validation exercise, we strongly recommend exclusively making use of fSCA retrievals that have accounted for the mixed pixel problem, such as SU or regression, for the reference data set, as in the study of Rittger et al. (2013). An alternative is to employ thresholding retrievals from very high resolution satellite imagery (Cortés et al., 2014), as this strongly moderates the mixed pixel problem.

4.3. Mixed pixel problem

In natural landscapes, spatial heterogeneity often leads to mixed pixel problems: a single satellite pixel is typically composed of more than one endmember (Painter et al., 2003). This is often the case in the middle of

winter, not just the ablation season, due to rocks and vegetation that protrude through the snow. As such, classifying a pixel into a single land-cover is inherently problematic, although it is implicitly understood that the land-cover assigned to a pixel represents the most abundant endmember in that pixel. The mixed pixel problem becomes less severe with higher spatial resolution, but does not fully disappear until the scale of a single point.

As a result, any retrieval algorithm that uses thresholds to binarize a pixel to a single endmember is inherently flawed, but one assumes that the error introduced by thresholding is not systematic and cancels out in the process of aggregation. Our results suggest that this is not necessarily the case for snow cover, with thresholding algorithms upon aggregation systematically overestimating fSCA. This is even the case at the higher resolution of 100 m, for which all thresholding based retrievals exhibit a positive bias and lower scores for the binary evaluation metrics, especially for high binarization thresholds. The positive bias in these retrievals is explained by the aggregation of pixels that are only partly snow-covered, but are classified as snow-covered by the binary thresholding. If as many low-snow cover pixels were mapped as snow-free as high snow-cover pixels were mapped as snow-covered, thresholding would give an on average unbiased response, but this is clearly not the case in our study. The overestimation can be identified in the ROC space (see Figure 10) where the false positive rate is generally high (≥ 0.1) for the larger binarization thresholds (blue and gray circles). This effect is also seen in Figure 6 in which most of the Bayelva floodplain is mapped as fully snow-covered by the thresholding algorithms, while it is only partially snow-covered in reality.

Spectral unmixing algorithms, pioneered for fSCA retrieval by Nolin et al. (1993) and operationalized by Painter et al. (2009), explicitly take the mixed pixel problem into account (see equation (4)). In Figure 6, we note that much of the Bayelva river floodplain is correctly retrieved as partially snow-covered in the unmixing experiment with Sentinel-2 MSI. For the higher resolution sensors, the unmixing algorithms clearly outperform all thresholding algorithms. For SU, applying an atmospheric correction is vital, given that the employed endmember spectra are surface reflectances, not TOA reflectances (see discussion in Schaepman-Strub et al., 2006). As discussed, this atmospheric correction procedure may introduce additional uncertainty, but the unmixing-based experiments in our study nonetheless performed best for the higher resolution sensors (OLI and MSI), with almost zero bias and relatively low RMSE values. The favorable performance can in large parts be

attributed to the physically-based nature of the SU retrieval algorithm and the fact that the mixed-pixel problem is explicitly accounted for (Keshava and Mustard, 2002).

In contrast to the favorable performance for the higher resolution sensors and unlike previous studies (e.g. Rittger et al., 2013), the unmixing-based MODSCAG algorithm (Painter et al., 2009) did not perform well in our study (see Figure 9). We initially suspected that the (unknown) choice of endmembers in MODSCAG could explain the poor performance, and by picking AOI specific endmembers we could slightly improve the performance in the custom experiments SU-MOD09GA and SU-MYD09GA (see Table 3). These custom experiments also benefited from the updated (version 6) MODIS surface reflectance products, as opposed to the version 5 products that were still used in the MODSCAG retrievals available for our AOI. Still, even the custom experiments were outperformed by the simpler MOD10A1 & MYD10A1 products. We highlight that, to the best of our knowledge, MODSCAG has not previously been evaluated in the Arctic. It is therefore possible that some of the assumptions in MODSCAG, such as ignoring directional effects (Painter et al., 2009), are more problematic in the Arctic where solar zenith angles are much higher during the ablation season than at lower latitudes (minimum $\theta_0 = 56°$ in the study area). This can be especially problematic for MODIS which, unlike OLI or MSI, is not a nadir looking sensor, since high solar zenith angles make forward scattering from snow in the view direction more likely (Painter et al., 2009). The combination of off-nadir view angle and high solar zenith angle may also make situations with sunglint possible (which is not considered in the unmixing). Nonetheless, we found negligible correlation between the retrieval errors and sensor zenith angle as well as solar zenith angle for MODSCAG and the custom SU experiments, respectively. This indicates that directional effects are not likely to be the source of error in line with the findings of Painter et al. (2009).

Dozier et al. (2008) show that the effective pixel dimension can increase by a factor of 10 when the view angle of MODIS is greatly off nadir, which could also be a significant source of error in unmixing with MODIS (Painter et al., 2009). In our study, this is probably not the case as this error source should then also manifest itself in the regression-based MODIS products MOD10A1 & MYD10A1 which it does not seem to do. Based on the formulae provided by Dozier et al. (2008), the pixel stretching factor (change in area) is 1.5 for a sensor view zenith angle of 30°. For the MODSCAG retrievals analyzed herein we found that the view zenith angle is lower than 30° in the

overwhelming majority of cases. Due to this geometric stretching and the triangular response function of the whiskbroom MODIS sensor (see Nishihama et al., 1997) admixing of signals emanating from neighboring pixels is a potential issue. To check if this was the case, we diagnosed the correlation in the fSCA time series between the different MODIS pixels. For all of the MODIS retrievals, we found high (> 0.9) between pixel fSCA correlations, indicating that our AOI is homogeneous in terms of fSCA at the MODIS pixel-scale. Moreover, strongly different surfaces, in particular the ocean, are in general one pixel or more away from the MODIS pixels considered in our AOI (see Figure 1). Thus admixing of neighboring pixels does not appear to be a major issue for our AOI and does not change the conclusions provided.

Both versions of the MOD10A1 & MYD10A1 products performed reasonably well for both the binary and fractional evaluation metrics (see Table 3), performing similarly to the thresholding experiments. At first glance, this is surprising given that the thresholding experiments at 500 m are based on the aggregation of a large number of smaller pixels, which considerably reduces random error. At the same time, the thresholding retrievals are biased due to the mixed pixel problem, and this bias does not disappear upon aggregation. The MOD10A1 & MYD10A1 products are based on the empirical regressions in (2) with coefficients presented in (Salomonson and Appel, 2004). As noted by Riggs et al. (2017), the NDSI in (fractionally) snow-covered landscapes typically ranges from 0 to 1, which the linear regression can account for. As such, the NDSI regression implicitly considers the mixed pixel problem. This implicit accounting also explains why the regression-based retrievals for MSI and OLI markedly outperformed the thresholding-based retrievals, although a slight bias remained. Both versions 5 and 6 of MOD10A1 & MYD10A1 performed reasonably well, with for the most part superior evaluation metric scores compared to the SU-based MODIS fSCA retrievals. The biases in these regression-based retrievals were, however, comparable or even larger than those in the SU-based retrievals. Thus, at a certain level of aggregation (upscaling), due to the cancellation of random errors, the SU-based MODIS retrievals will match or even outperform the regression-based MODIS retrievals. It is possible that a better performance can be obtained by tuning the coefficients of the regressions for this particular site (Riggs et al., 2017). However, as noted by Masson et al. (2018), employing the 'universal' coefficients is advantageous for validation purposes, since the results can be generalized beyond the study area.

Based on our results, we make the general recommendation of employing fSCA retrieval algorithms that take into account the mixed pixel problem. It is only in this way that systematic error can be eliminated from the retrieval process. Such algorithms could go beyond the linear mixed pixel accounting approaches considered here (i.e., SU and NDSI regression), and include nonlinear regression approaches from the machine learning community such as artificial neural networks (e.g. Czyzowska-Wisniewski et al., 2015). To extend the applicability of our results for the higher level MODIS products, which were based on a low number of pixels, we envisage increasing the size of the validation area and making use of higher resolution satellite retrievals (from e.g. Sentinel-2 MSI or Landsat 8 OLI) as a reference. These reference satellite retrievals would have to account for the mixed pixel problem to avoid a biased validation procedure. Such an extension is, however, beyond the scope of the present study which focuses on the use of a dense time series of very high resolution retrievals from a terrestrial automatic camera system that are only available for a small limited area.

5. Conclusion

In this study, we evaluated satellite retrievals of fractional snow-covered area (fSCA) using several hundred high resolution snow-cover maps retrieved from orthorectified terrestrial images. These images were taken over the course of six snowmelt seasons (2012-2017) by an automatic camera system installed on a mountain peak near Ny-Ålesund in the high-Arctic archipelago of Svalbard. We considered fSCA retrievals from three different optical satellite sensors, namely: the operational land imager (OLI) on-board Landsat-8, the multispectral instrument (MSI) on-board the Sentinel-2 satellites, and the moderate resolution imaging spectroradiometer (MODIS) on-board Terra and Aqua. The algorithms employed to retrieve fSCA from multispectral satellite imagery ranged from simple thresholding of the normalized difference snow index (NDSI) to fully constrained spectral unmixing (SU). The satellite retrievals were extensively evaluated with respect to coincident reference snow-cover maps from the terrestrial photographs, considering both fractional and binary evaluation metrics. From the results of our study, we draw the following main conclusions:

- fSCA retrieved by spatially aggregating NDSI-derived binary snow-cover maps from Sentinel-2 MSI and Landsat 8 OLI is inherently biased due to the mixed pixel problem.

- SU with the same sensors can explicitly account for this problem, leading to near unbiased estimates with lower error variance.
- NDSI-regression based fSCA retrievals implicitly account for the mixed pixel problem and can thus provide satisfactory results.
- The Fmask-OLI and SLCC-MSI native scene classifications overestimate both cloud and snow cover.
- Unsupervised adaptive NDSI threshold selection does not necessarily outperform a fixed threshold of 0.4.
- Atmospheric correction has little impact on fSCA retrieved using NDSI thresholding.

The conservative uncertainty estimates established in this study are useful for reanalyses that ingest satellite-retrieved fSCA in models using data assimilation. Given that the SU retrievals for the higher resolution OLI and MSI sensors performed considerably better than all the other retrievals, we strongly recommend efforts towards operationalizing merged SU products from these sensors. Spectral unmixing is only slightly more computationally expensive than the other retrievals, and it appears that the benefit of moderating the mixed pixel problem can outweigh the costs. Although the 500 m operational MOD10A1 and MYD10A1 products performed reasonably well, a higher resolution merged operational unmixing product would yield marked reductions in error without significantly compromising the revisit time, especially in the Arctic.

Acknowledgements

Our work was funded by SatPerm (239918; Research Council of Norway), in collaboration with EmblA (56801; Nordforsk) and ESA Permafrost CCI (http://cci.esa.int/Permafrost). This study is a contribution to the strategic research area LATICE at the University of Oslo. We gratefully acknowledge all the data providers for making their data open and freely available. Landsat 8 OLI scenes were downloaded from USGS Earth Explorer (http://earthexplorer.usgs.gov). Sentinel-2A/2B MSI scenes were downloaded from the Copernicus Open Access Hub (https://scihub.copernicus.eu). MODIS products were downloaded from NSIDC (MOD10A/MYD10A1), LP DAAC

(MOD09GA/MYD09GA), and JPL (MODSCAG). The spectral response functions of the different sensors were obtained from EUMETSAT and ESA. The HRSC-AX imagery and DEM was obtained from the data set associated with the study of Boike et al. (2018). The DEM for the entire Brøgger peninsula was obtained from the Norwegian Polar Institute (NPI, 2014). The vast majority of the work conducted in this study was carried out in the MATLAB computing environment using native functions and toolboxes. In addition, solar geometry was diagnosed using the ephemeris routine in the PV_LIB toolbox developed at Sandia National Labs. Based on the ephemeris, shadows were obtained using the TopoToolbox (Schwanghart and Scherler, 2014) thanks to a suggestion from Bas Altena. We would like to thank He et al. (2018) for making the SNICAR MATLAB routines available. Thanks also to George A. Riggs and Kat Bormann for their helpful comments about the MODIS products. We gratefully acknowledge field support from Kjersti Gisnås, Jaroslav Obu, and Håvard Kristiansen. We also want to thank the editor, Menghua Wang, and three anonymous reviewers, all of whom contributed to significantly improving this paper.

References

Aalstad, K., Westermann, S., Schuler, T.V., Boike, J., Bertino, L., 2018. Ensemble-based assimilation of fractional snow-covered area satellite retrievals to estimate the snow distribution at Arctic sites. The Cryosphere 12, 247–270. doi:10.5194/tc-12-247-2018.

Arsenault, K.R., Houser, P.R., De Lannoy, G.J.M., 2014. Evaluation of the MODIS snow cover fraction product. Hydrological Processes doi:10.1002/hyp.9636.

Baldridge, A.M., Hook, S.J., Grove, C.I., Rivera, G., 2009. The ASTER spectral library version 2.0. Remote Sensing of Environment 113, 711–715. doi:10.1016/j.rse.2008.11.007.

Barnett, T.P., Adam, J.C., Lettenmaier, D.P., 2005. Potential impacts of a warming climate on water availability in snow-dominated regions. Nature 438. doi:10.1038/nature04141.

Boike, J., Juszak, I., Lange, S., Chadburn, S., Burke, E., Overduin, P.P., Roth, K., Ippisch, O., Bornemann, N., Stern, L., Gouttevin, I., Hauber,

E., Westermann, S., 2018. A 20-year record (1998-2017) of permafrost, active layer and meteorological conditions at high Arctic permafrost research site (Bayelva, Spitsbergen). Earth System Science Data 10, 355–390. doi:10.5194/essd-10-355-2018.

Bormann, K., Brown, R.D., Derksen, C., Painter, T.H., 2018. Estimating snow-cover trends from space. Nature Climate Change doi:10.1038/s41558-018-0318-3.

Bouguet, J.Y., 2015. Camera Calibration Toolbox for Matlab. Available at: http://www.vision.caltech.edu/bouguetj/calib_doc/ (last access: December 2018).

Chadburn, S.E., Burke, E.J., Cox, P.M., Friedlingstein, P., Hugelius, G., Westermann, S., 2017. An observation-based constraint on permafrost loss as a function of global warming. Nature Climate Change 7, 340–344. doi:10.1038/nclimate3262.

Chapin, F.S., Sturm, M., Serreze, M.C., McFadden, J.P., Key, J.R., Lloyd, A.H., McGuire, A.D., Rupp, T.S., Lynch, A.H., Schimel, J.P., Beringer, J., Chapman, W.L., Epstein, H.E., Euskirchen, E.S., Hinzman, L.D., Jia, G., Ping, C.L., Tape, K.D., Thompson, C.D.C., Walker, D.A., Welker, J.M., 2005. Role of Land-Surface Changes in Arctic Summer Warming. Science 310, 657–660. doi:10.1126/science.1117368.

Cortés, G., Girotto, M., Margulis, S.A., 2014. Analysis of sub-pixel snow and ice extent over the extratropical Andes using spectral unmixing and historical Landsat imagery. Remote Sensing of Environment 141, 64–78. doi:10.1016/j.rse.2013.10.023.

Cortés, G., Girotto, M., Margulis, S.A., 2016. Snow process estimation over the extratropical Andes using a data assimilation framework integrating MERRA data and Landsat imagery. Water Resources Research 52. doi:10.1002/2015WR018376.

Czyzowska-Wisniewski, E.H., van Leeuwen, J.D., Hirschboeck, K.H., Marsh, S.E., Wisniewski, W.T., 2015. Fractional snow cover estimation in complex alpine-forested environments using an artificial neural network. Remote Sensing of Environment doi:10.1016/j.rse.2014.09.026.

Dozier, J., 1989. Spectral Signature of Alpine Snow Cover from the Landsat Thematic Mapper. Remote Sensing of Environment 28, 9–22. doi:10.1016/0034-4257(89)90101-6.

Dozier, J., Bair, E.H., Davis, R.E., 2016. Estimating the spatial distribution of snow water equivalent in the world's mountains. WIREs Water 3, 461–474. doi:10.1002/wat2.1140.

Dozier, J., Green, R.O., Nolin, A.W., Painter, T.H., 2009. Interpretation of snow properties from imaging spectrometry. Remote Sensing of Environment 113, S25–S37. doi:10.1016/j.rse.2007.07.029.

Dozier, J., Painter, T.H., Rittger, K., Frew, J.E., 2008. Time-space continuity of daily maps of fractional snow cover and albedo from MODIS. Advances in Water Resources doi:10.1016/j.advwatres.2008.08.011.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. Remote Sensing of Environment doi:10.1016/j.rse.2011.11.026.

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recognition Letters 26, 861–874. doi:10.1016/j.patrec.2005.10.010.

Fedorov, R., Camerada, A., Fraternali, P., Tagliasacchi, M., 2016. Estimating Snow Cover From Publicly Available Images. IEEE Transactions on Multimedia 18, 1187–1200. doi:10.1109/TMM.2016.2535356.

Førland, J.E., Benestad, R., Hanssen-Bauer, I., Haugen, J.E., Skaugen, T.E., 2011. Temperature and Precipitation Development at Svalbard 1900-2100. Advances in Meteorology 2011, 1–14. doi:10.1155/2011/893790.

Gascoin, S., Grizonnet, M., Bouchet, M., Salgues, G., Hagolle, O., 2019. Theia snow collection: high-resolution operational snow cover maps from sentinel-2 and landsat-8 data. Earth System Science Data doi:10.5194/essd-11-493-2019.

Gascoin, S., Hagolle, O., Huc, M., Jarlan, L., Dejoux, J.F., Szczypta, C., Marti, R., Sánchez, R., 2015. A snow cover climatology for the pyrenees

from modis snow products. Hydrology and Earth System Sciences 19, 2337–2351. doi:10.5194/hess-19-2337-2015.

Girotto, M., Margulis, S.A., Durand, M., 2014. Probabilistic SWE reanalysis as a generalization of deterministic SWE reconstruction techniques. Hydrological Processes 28. doi:10.1002/hyp.9887.

Hall, D.K., 1988. Assessment of Polar Climate Change Using Satellite Technology. Reviews of Geophysics 26, 26–39. doi:10.1029/RG026i001p00026.

Hall, D.K., Riggs, G.A., 2016a. MODIS/Aqua Snow Cover Daily L3 Global 500m Grid, Version 6, h18v01, Boulder, CO, USA, NASA NSIDC DAAC. doi:10.5067/MODIS/MYD10A1.006. Accessed: 06-2018.

Hall, D.K., Riggs, G.A., 2016b. MODIS/Terra Snow Cover Daily L3 Global 500m Grid, Version 6, h18v01, Boulder, CO, USA, NASA NSIDC DAAC. doi:10.5067/MODIS/MOD10A1.006. Accessed: 06-2018.

Hall, D.K., Riggs, G.A., Salmonson, V.V., DiGirolamo, N.E., Bayr, K.J., 2002. MODIS snow-cover products. Remote Sensing of Environment 83, 181–194. doi:10.1016/S0034-4257(02)00095-0.

Hall, D.K., Salomonson, V.V., Riggs, G.A., 2006a. MODIS/Aqua Snow Cover Daily L3 Global 500m Grid, Version 5, h18v01, Boulder, CO, USA, NASA NSIDC DAAC. doi:10.5067/ZFAEMQGSR4XD. Accessed: 06-2018.

Hall, D.K., Salomonson, V.V., Riggs, G.A., 2006b. MODIS/Terra Snow Cover Daily L3 Global 500m Grid, Version 5, h18v01, Boulder, CO, USA, NASA NSIDC DAAC. doi:10.5067/63NQASRDPDB0. Accessed: 06-2018.

Härer, S., Bernhardt, M., Siebers, M., Schulz, K., 2018. On the need for a time- and location-dependent estimation of the NDSI threshold value for reducing existing uncertainties in snow cover maps at different scales. The Cryosphere 12, 1629–1642. doi:10.5194/tc-12-1629-2018.

He, C., Flanner, M.G., Chen, F., Barlage, M., Liou, K.N., Kang, S., Ming, J., Qian, Y., 2018. Black carbon-induced snow albedo reduction over the Tibetan Plateau: uncertainties from snow grain shape and aerosol-snow mixing state based on an updated SNICAR model. Atmospheric Chemistry and Physics 18. doi:10.5194/acp-18-11507-2018.

Heinz, D.C., Chang, C.I., 2001. Fully Constrained Least Squares Linear Spectral Mixture Analysis for Material Quantification in Hyperspectral Imagery. IEEE Transactions on Geoscience and Remote Sensing doi:10.1109/36.911111.

Keshava, N., Mustard, J.F., 2002. Spectral unmixing. IEEE Signal Processing Magazine doi:10.1109/79.974727.

Lawson, C.L., Hanson, R.J., 1995. Linear Least Squares with Linear Inequality Constraints, in: Solving Least Squares Problems. SIAM. chapter 23, pp. 160–165. doi:10.1137/1.9781611971217.ch23.

Mankin, J.S., Viviroli, D., Singh, D., Hoekstra, A.Y., Diffenbaugh, N.S., 2015. The potential for snow to supply human water demand in the present and future. Environmental Research Letters 10. doi:10.1088/1748-9326/10/11/114016.

Marti, R., Gascoin, S., Berthier, E., de Pinel, M., Houet, T., Laffly, D., 2016. Mapping snow depth in open alpine terrain from stereo satellite imagery. The Cryosphere 10, 1361–1380. doi:10.5194/tc-10-1361-2016.

Martinec, J., Rango, A., 1981. Areal Distribution of Snow Water Equivalent Evaluated by Snow Cover Monitoring. Water Resources Research 17, 1480–1488. doi:10.1029/WR017i005p01480.

Masson, T., Dumont, M., Dalla Mura, M., Sirguey, P., Gascoin, S., Dedieu, J.P., Chanussot, J., 2018. An Assessment of Existing Methodologies to Retrieve Snow Cover Fraction from MODIS Data. Remote Sensing 10. doi:10.3390/rs10040619.

Müller-Wilm, U., 2018. Sen2Cor Configuration and User Manual. Ref. S2-PDGS-MPC-L2A-SUM-V2.5.5. Available at: http://step.esa.int/main/third-party-plugins-2/sen2cor/ (last access: December 2018).

Musselman, K.N., Clark, M.P., Liu, C., Ikeda, K., Rasmussen, R., 2017. Slower snowmelt in a warmer world. Nature Climate Change 7. doi:10.1038/nclimate3225.

Niittynen, P., Heikkinen, R.K., Luoto, M., 2018. Snow cover is a neglected driver of Arctic biodiversity loss. Nature Climate Change doi:10.1038/s41558-018-0311-x.

Nishihama, M., Wolfe, R., Solomon, D., Patt, F., Blanchette, J., Fleig, A., Masuoka, E., 1997. MODIS Level 1A Earth Location: Algorithm Theoretical Basis Document Version 3.0.

Nolin, A.W., Dozier, J., Mertes, L.A.K., 1993. Mapping alpine snow using a spectral mixture modeling technique. Annals of Glaciology 17, 121–124. doi:10.3189/S0260305500012702.

NPI, 2014. Terrengmodell Svalbard (S0 Terrengmodell), Delmodell 5m 2010_13828_33. doi:10.21334/npolar.2014.dce53a47. norwegian Polar Institute (NPI) DEM. Last access: December 2018.

Østby, T.I., Schuler, T.V., Westermann, S., 2014. Severe cloud contamination of MODIS Land Surface Temperatures over an Arctic ice cap, Svalbard. Remote Sensing of Environment doi:10.1016/j.rse.2013.11.005.

Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics doi:https://doi.org/10.1109/TSMC.1979.4310076.

Painter, T.H., Dozier, J., Roberts, D.A., Davis, R.E., Green, R.O., 2003. Retrieval of subpixel snow-covered area and grain size from imaging spectrometer data. Remote Sensing of Environment 85, 64–77. doi:10.1016/S0034-4257(02)00187-6.

Painter, T.H., Rittger, K., McKenzie, C., Slaughter, P., Davis, R.E., Dozier, J., 2009. Retrieval of subpixel snow covered area, grain size, and albedo from MODIS. Remote Sensing of Environment 113, 868–879. doi:10.1016/j.rse.2009.01.001.

Painter, T.H., Seidel, F.C., Bryant, A.C., McKenzie Skiles, S., Rittger, K., 2013. Imaging spectroscopy of albedo and radiative forcing by light-absorbing impurities in mountain snow. Journal of Geophysical Research: Atmospheres 118. doi:10.1002/jgrd.50520.

Pepin, N., Bradley, R.S., Diaz, H.F., Baraer, M., Caceres, E.B., Forsythe, N., Fowler, H., Greenwood, G., Hashmi, M.Z., Liu, X.D., Miller, J.R.,

Ning, L., Ohmura, A., Palazzi, E., Rangwala, I., Schöner, W., Severskiy, I., Shahgedanova, M., Wang, M.B., Williamson, S.N., Yang, D.Q., 2015. Elevation-dependent warming in mountain regions of the world. Nature Climate Change 5. doi:10.1038/nclimate2563.

Richter, R., Louis, J., Müller-Wilm, U., 2012. Sentinel-2 MSI - Level 2A Products Algorithm Theoretical Basis Document. Issue 2.0.

Riggs, G.A., Hall, D.K., Román, M.O., 2017. Overview of NASA's MODIS and Visibile Infrared Imaging Radiometer Suite (VIIRS) snow-cover Earth System Data Records. Earth System Science Data 9, 765–777. doi:10.5194/essd-9-765-2017.

Rittger, K., Bair, E.H., Kahl, A., Dozier, J., 2016. Spatial estimates of snow water equivalent from reconstruction. Advances in Water Resources 94, 345–363. doi:10.1016/j.advwatres.2016.05.015.

Rittger, K., Painter, T.H., Dozier, J., 2013. Assessment of methods for mapping snow cover from MODIS. Advances in Water Resources 51, 367–380. doi:10.1016/j.advwatres.2012.03.002.

Sakov, P., Counillon, F., Bertino, L., Lisæter, K.A., Oke, P.R., Korablev, A., 2012. TOPAZ4: an ocean-sea ice data assimilation system for the North Atlantic and Arctic. Ocean Science doi:10.5194/os-8-633-2012.

Salomonson, V.V., Appel, I., 2004. Development of the Aqua MODIS NDSI Fractional Snow Cover Algorithm and Validation Results. IEEE Transactions on Geoscience and Remote Sensing 44, 1747–1756. doi:10.1109/TGRS.2006.876029.

Schaepman-Strub, G., Schaepman, C., Painter, T.H., Dangel, S., Martonchik, J.V., 2006. Reflectance quantities in optical remote sensingdefinitions and case studies. Remote Sensing of Environment doi:10.1016/j.rse.2006.03.002.

Schuur, E.A.G., McGuire, A.D., Schädel, C., Grosse, G., Harden, J.W., Hayes, D.J., Hugelius, G., D., K.C., Kuhry, P., Lawrence, D.M., Natali, S.M., Olefeldt, D., Romanovsky, V.E., Schaefer, K., Turetsky, M.R., Treat, C.C., Vonk, J.E., 2015. Climate change and the permafrost carbon feedback. Nature doi:10.1038/nature14338.

Schwanghart, W., Scherler, D., 2014. Short Communication: TopoToolbox 2 - MATLAB-based software for topographic analysis and modeling in Earth surface sciences. Earth Surface Dynamics doi:10.5194/esurf-2-1-2014.

Seidel, F.C., Rittger, K., Skiles, S.M., Molotch, N.P., Painter, T.H., 2016. Case study of spatial and temporal variability of snow cover, grain size, albedo and radiative forcing in the Sierra Nevada and Rocky Mountain snowpack derived from imaging spectroscopy. The Cryosphere 10, 1229–1244. doi:10.5194/tc-10-1229-2016.

Selkowitz, D.J., Forster, R.R., 2016. Automated mapping of persistent ice and snow cover accross the western U.S. with Landsat. ISPRS Journal of Photogrammetry and Remote Sensing 117, 126–140. doi:10.1016/j.isprsjprs.2016.04.001.

Sirguey, P., Mathieu, R., Arnaud, Y., 2009. Subpixel monitoring of the seasonal snow cover with MODIS at 250 m spatial resolution in the Southern Alps of New Zealand: Methodology and accuracy assessment. Remote Sensing of Environment 113, 160–181. doi:10.1016/j.rse.2008.09.008.

SNL, 2014. Sandia National Labs PV_LIB Toolbox. Available at: https://pvpmc.sandia.gov/applications/pv_lib-toolbox (last access: December 2018).

Stewart, I.T., Cayan, D.R., Dettinger, M.D., 2005. Changes toward Earlier Streamflow Timing across Western North America. Journal of Climate 18, 1136–1155. doi:10.1175/JCLI3321.1.

Stillinger, T., Roberts, D.A., Collar, N.M., Dozier, J., 2019. Cloud Masking for Landsat 8 and MODIS Terra Over SnowCovered Terrain: Error Analysis and Spectral Similarity Between Snow and Cloud. Water Resources Research doi:10.1029/2019WR024932.

Swenson, S.C., Lawrence, D.M., 2012. A new fractional snow-covered area parameterization for the Community Land Model and its effect on the surface energy balance. Journal of Geophysical Research: Atmospheres doi:10.1029/2012JD018178.

Treichler, D., Kääb, A., 2017. Snow depth from ICESat laser altimetry - A test study in southern Norway. Remote Sensing of Environment 191, 389–401. doi:10.1016/j.rse.2017.01.022.

Trofaier, A.M., Westermann, S., Bartsch, A., 2017. Progress in space-borne studies of permafrost for climate science: Towards a multi-ECV approach. Remote Sensing of Environment 203, 55–70. doi:10.1016/j.rse.2017.05.021.

USGS, 2018. Product Guide Landsat 8 Surface Reflectance Code (LASRC) Product. Available at: https://www.usgs.gov/land-resources/nli/landsat/files (last access: December 2018).

Vermote, E., Justice, C., Claverie, M. Franch, B., 2016. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. Remote Sensing of Environment doi:10.1016/j.rse.2016.04.008.

Vermote, E., Wolfe, R., 2015a. MOD09GA MODIS/Terra Surface Reflectance Daily L2G Global 1km and 500m SIN Grid V006, NASA EOSDIS LP DAAC. doi:10.5067/MODIS/MYD09GA.006. Accessed: 06-2018.

Vermote, E., Wolfe, R., 2015b. MYD09GA MODIS/Aqua Surface Reflectance Daily L2G Global 1km and 500m SIN Grid V006, NASA EOSDIS LP DAAC. doi:10.5067/MODIS/MYD09GA.006. Accessed: 06-2018.

Vikhamar, D., Solberg, R., 2003. Snow-cover mapping in forests by constrained linear spectral unmixing of MODIS data. Remote Sensing of Environment 88, 309–323. doi:10.1016/j.rse.2003.06.004.

Wackernagel, H., 2003. Multivariate Geostatistics. doi:10.1007/978-3-662-05294-5.

Westermann, S., Boike, J., Guglielmin, M., Gisnås, K., Etzelmüller, B., 2015. Snow melt monitoring near Ny-Ålesund, Svalbard, using Automatic Camera Systems. Department of Geosciences, University of Oslo, PANGAEA. doi:10.1594/PANGAEA.846617. last access: 5 May 2017.

Westermann, S., Langer, M., Boike, J., 2011. Spatial and temporal variations of summer surface temperatures of high arctic tundra on Svalbard - Implications for MODIS LST based permafrost monitoring. Remote Sensing of Environment doi:10.1016/j.rse.2010.11.018.

Westermann, S., Langer, M., Boike, J., 2012. Systematic bias of average winter-time land surface temperatures inferred from MODIS at a site on Svalbard, Norway. Remote Sensing of Environment 118, 162–167. doi:10.1016/j.rse.2011.10.025.

Winsvold, S., Kääb, A., Nuth, C., 2016. Regional glacier mapping using optical satellite data time series. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing doi:https://doi.org/10.1109/JSTARS.2016.2527063.

Yılmaz, Y., Aalstad, K., Sen, O.L., 2019. Multiple Remotely Sensed Lines of Evidence for a Depleting Seasonal Snowpack in the Near East. Remote Sensing doi:10.3390/rs11050483.

Yin, D., Cao, X., Chen, X., Shao, Y., Chen, J., 2013. Comparison of automatic thresholding methods for snow-cover mapping using Landsat TM imagery. International Journal of Remote Sensing 34, 6529–6538. doi:10.1080/01431161.2013.803631.

Zhang, T., 2005. Influence of the Seasonal Snow Cover on the Ground Thermal Regime: An Overview. Reviews of Geophysics 43. doi:10.1029/2004RG000157.

Zhu, Z., Woodcock, C.E., 2012. Object-based cloud and cloud shadow detection in Landsat imagery. Remote Sensing of Environment 118, 83–94. doi:10.1016/j.rse.2011.10.028.

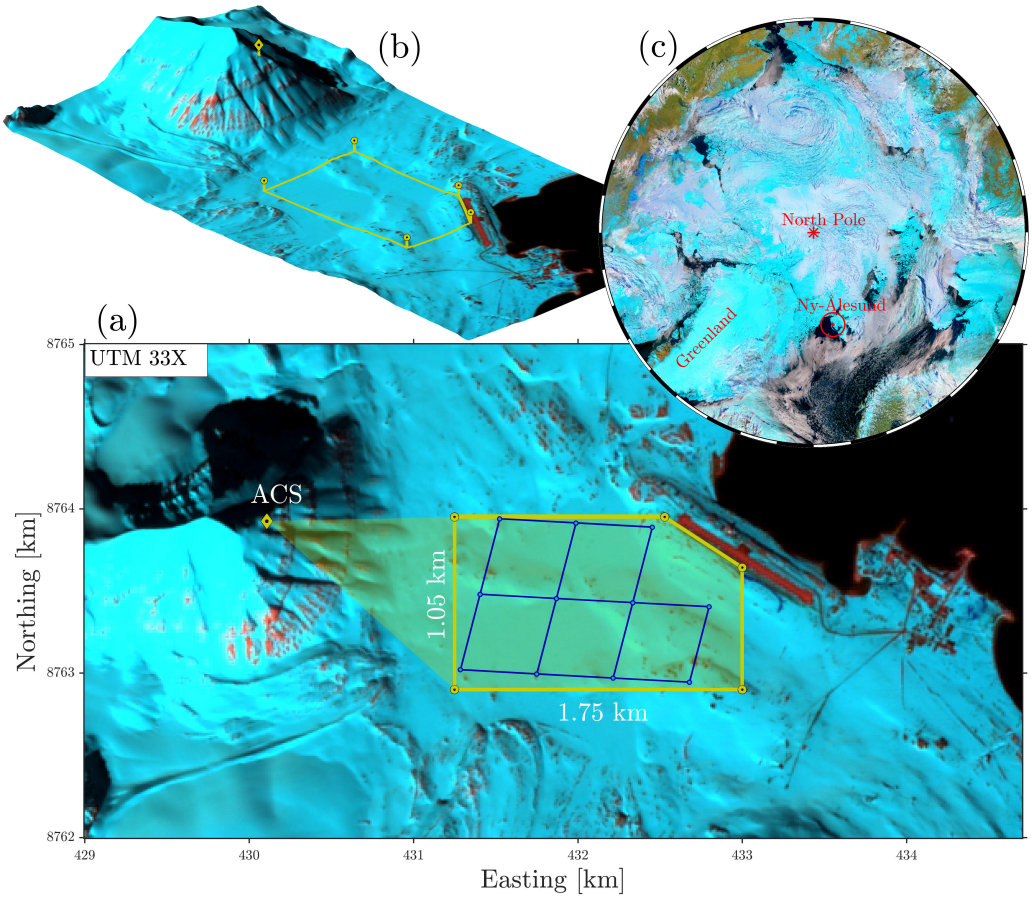


Figure 1: (a): Sentinel-2A MSI L2A enhanced false color (FC) image, sensed 01-Jun-2017 12:47Z, showing the location of the ACS (yellow diamond), the associated AOI (yellow polygon), and the footprint of the MODIS pixels considered (blue polygons). (b): The same image draped over a DEM (NPI, 2014) viewed from the south east at an elevation angle of 25°. (c): Terra MODIS composite (MOD09CMG) FC image (>65°N) from the same day showing the location of Ny-Ålesund in the Arctic.

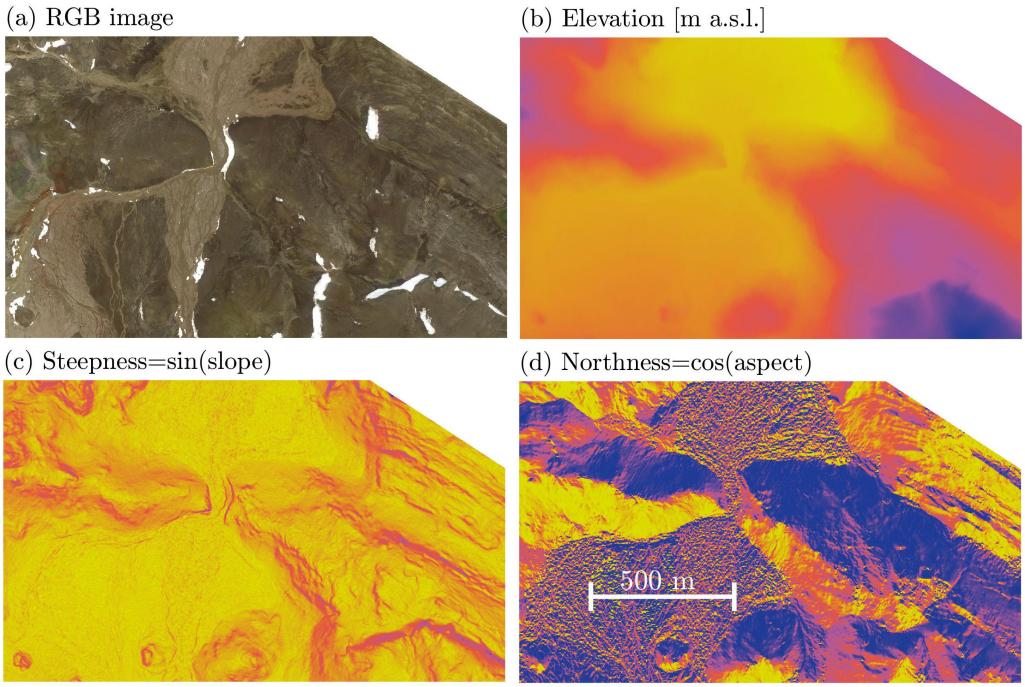


Figure 2: Overview of the AOI and the HRSC-AX products (Boike et al., 2018) that were used as a reference. All the panels are cropped to the AOI; (a) nadir looking true color image, (b) elevation (yellow=5 m a.s.l, blue=55 m a.s.l.), (c) 'steepness' (sine of the slope: yellow=0, blue=60° slope), (d) 'northness' (cosine of the aspect: yellow=south facing, blue=north facing). All these images have a one to one pixel aspect ratio with a spatial resolution of 0.5 m.

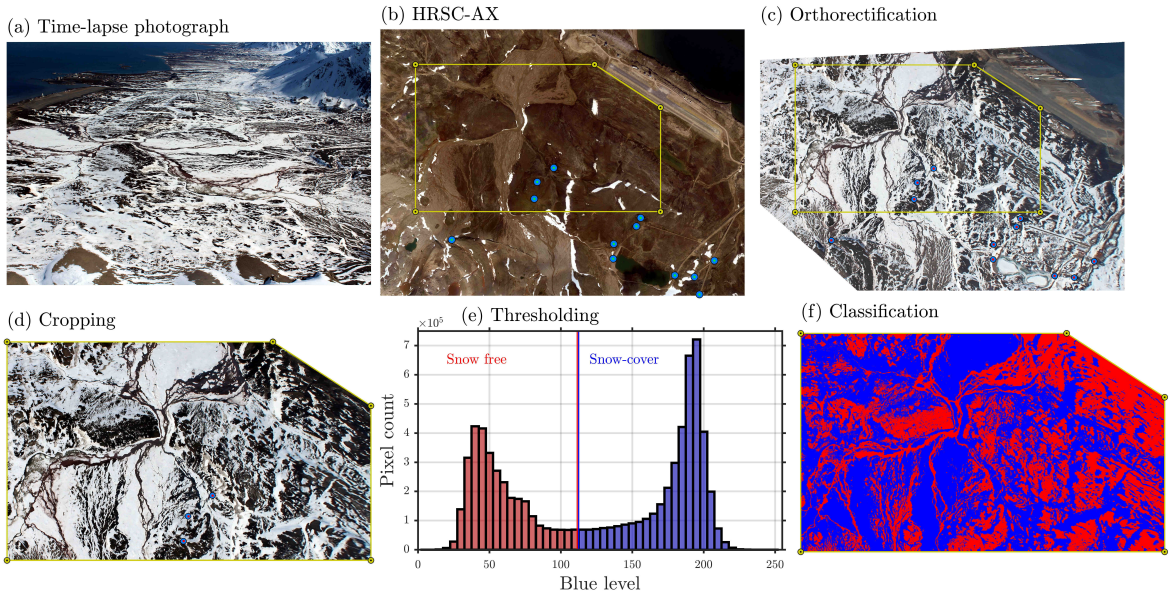


Figure 3: The processing chain for the terrestrial photography; **(a)** an example terrestrial photograph taken from Scheteligfjellet by the ACS at 10:31 UTC on the 03.06.2016; (b) HRSC-AX reference image used to evaluate the georeferencing error (blue dots show a subset of the available landmarks) with the AOI enclosed by the yellow polygon; **(c)** Orthorectified and georeferenced terrestrial photograph (the distance between the red dots and the center of the blue dots is the georeferencing error: RMSE=2.05 m and bias = 1.84 m for this image); **(d)** the same photograph cropped to the AOI; **(e)** the thresholding procedure, for this image a threshold blue level of 112 was selected; **(d)** The final classification (blue=snow, red=snow free) based on the selected binary threshold.

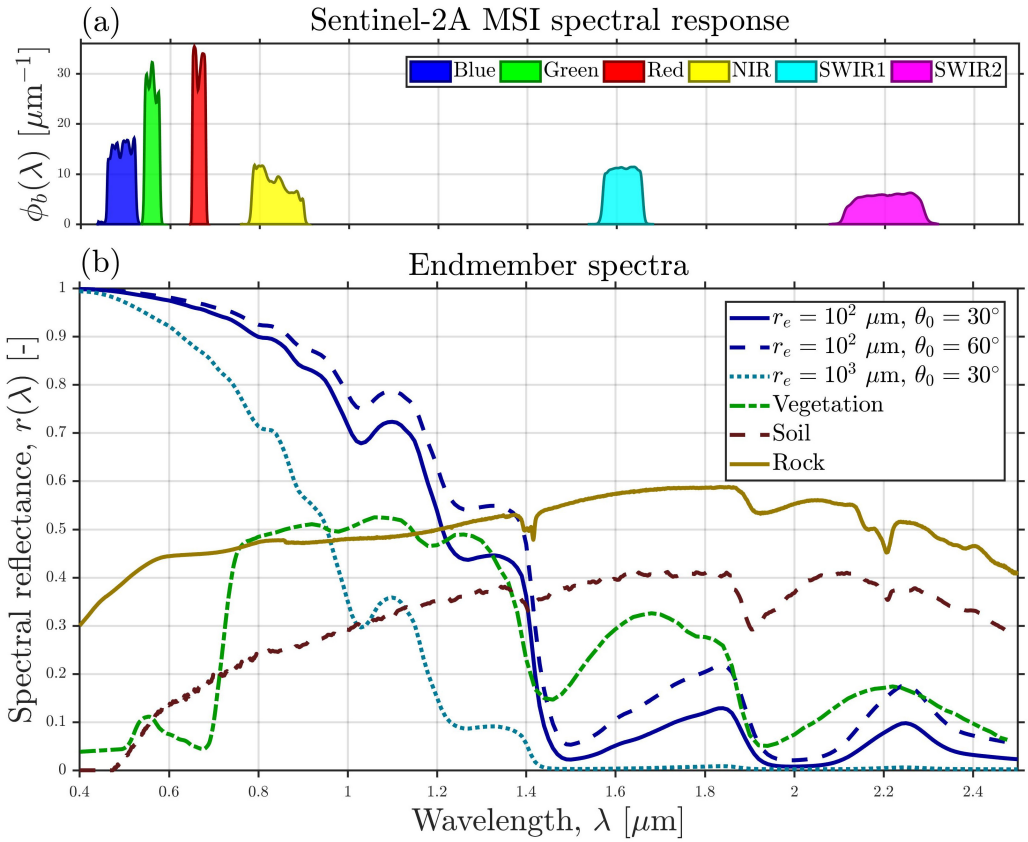


Figure 4: Examples of spectral response functions and endmembers used in the SU procedure; **(a)** Spectral response functions ϕ_b (normalized to probability density) for the Sentinel-2A MSI sensors for the $N_b = 6$ bands considered in the spectral unmixing; **(b)** The spectral reflectances of the non-snow endmembers from the JPL spectral library and a subset of the snow endmembers modeled through SNICAR for different effective snow grain radii r_e and solar zenith angles θ_0.

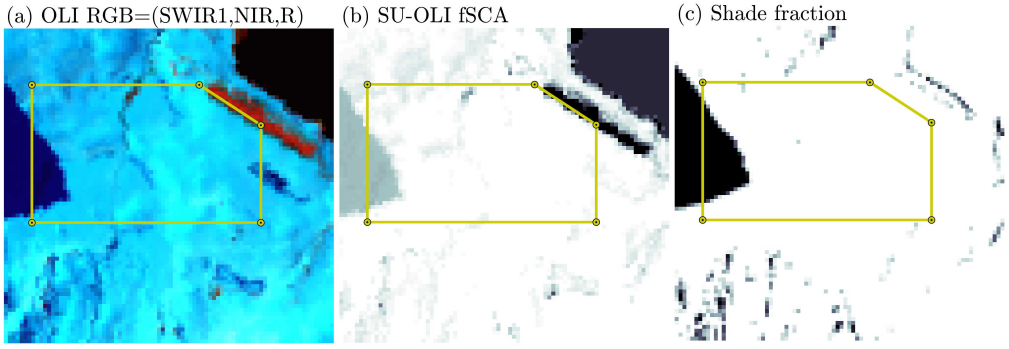


Figure 5: The effect of Scheteligfjellet's cast shadow on a Landsat 8 OLI level 2 scene sensed at 18:23Z on the 15.06.2014; **(a)** False color RGB=(SWIR1,NIR,R) image (note the shadow to the left in the image) with the yellow polygon dileneating the AOI; **(b)** The fSCA determined from SU (white=100%, black=0%); **(c)** The independently modeled shade fraction (white=0%, black=100%) from the DEM based shadow mask.

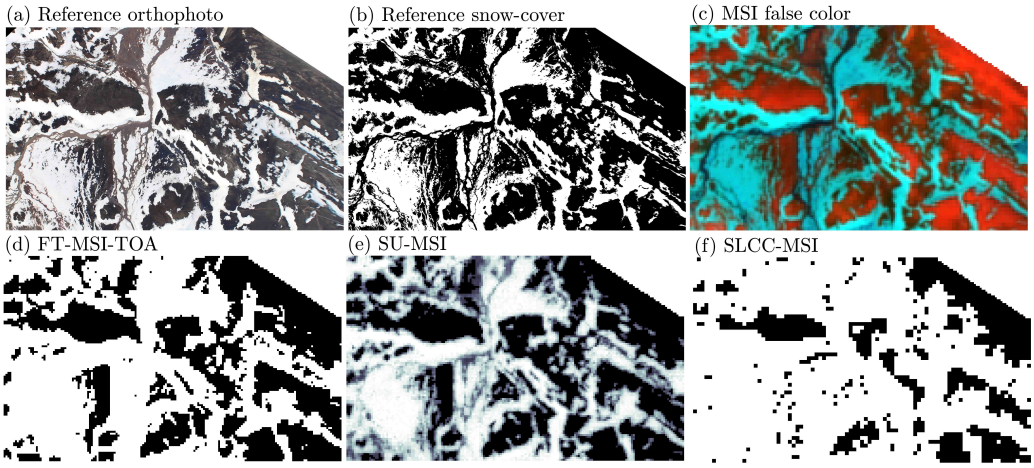


Figure 6: Sentinel-2 snow-cover mapping example: **(a)** Reference orthophoto taken around noon on the 21.06.2017; **(b)** Reference snow-cover map (white=snow, black=no snow); **(c)** S2A MSI L2A FC RGB=(SWIR1,NIR,R) image sensed 12:47:00Z on the same day; **(d)** Snow-cover map derived using a fixed NDSI threshold on TOA reflectances (FT-MSI-TOA) from the same S2A scene; **(e)** fSCA derived using spectral unmixing (SU-MSI) on the same S2A scene; **(f)** Snow-cover map derived from the L2A scene classification (SLCC-MSI) applied to the same S2A scene.

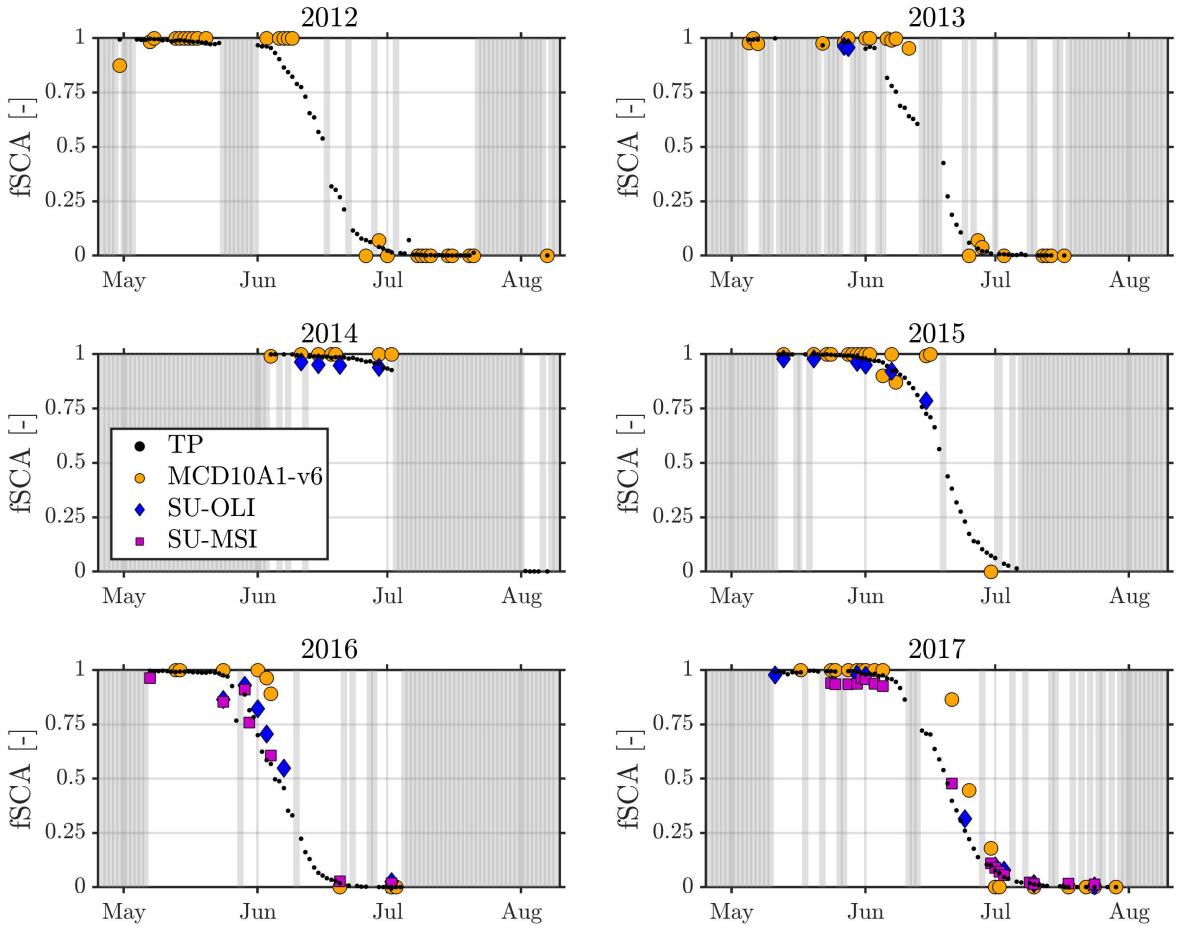


Figure 7: Time series of the reference fSCA (black dots) and a subset of the satellite retrieved fSCA (blue diamonds: OLI spectral unmixing, purple squares: MSI spectral unmixing, orange circles: merged MOD10A1 and MYD10A1 version 6) aggregated to the scale of the entire AOI for the 6 ablation seasons (2012-2017) considered in this study. For the gray columns there was no reference data and no retrievals are shown for these dates.

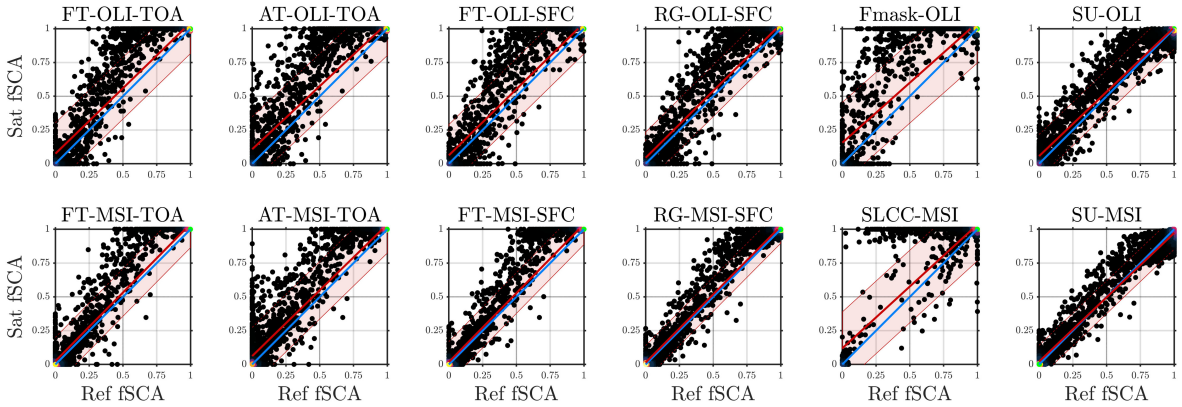


Figure 8: Scatter plots showing the performance of some of the satellite fSCA retrievals (y-axis) versus the fSCA retrieved from the terrestrial photography (x-axis) at 100 m spatial resolution. The blue line shows the 1:1 line, while the red line is the linear best fit with 95% prediction intervals shown in red shading. For an overview of the various satellite retrievals (indicated by the title of each panel) see Table 1. Results from several retrievals are not shown for visibility (see Table 2 metrics).

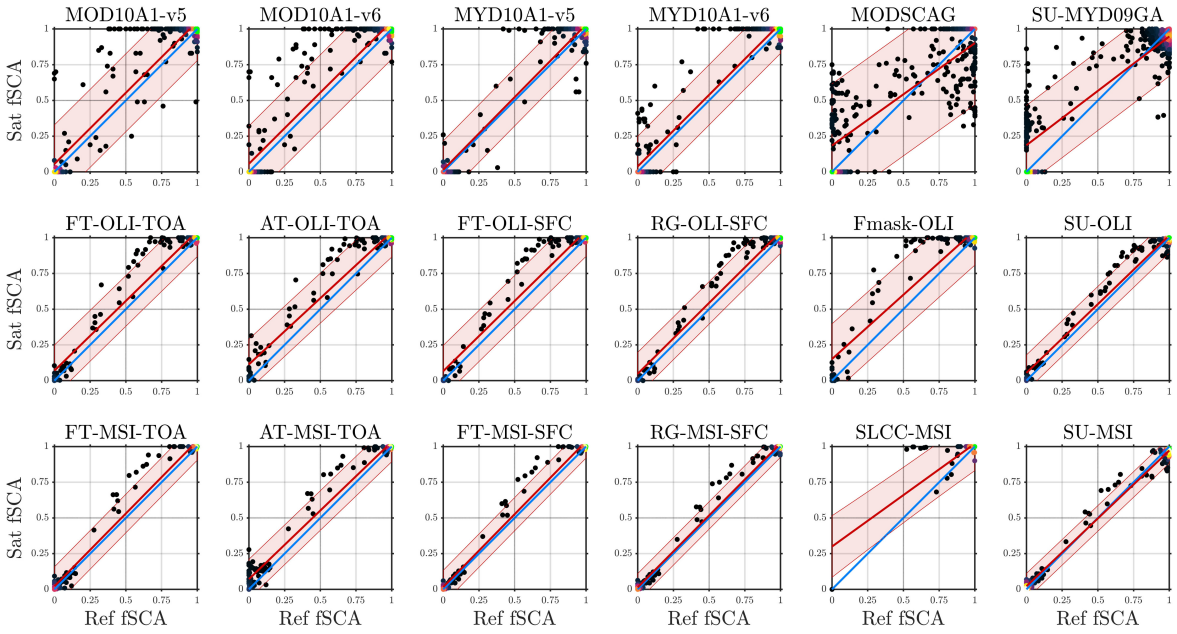


Figure 9: Same as Figure 8, but at 500 m spatial resolution. For an overview of the various satellite retrievals (indicated by the title of each panel) see Table 1. Results from several retrievals are not shown for visibility (see Table 3 metrics).

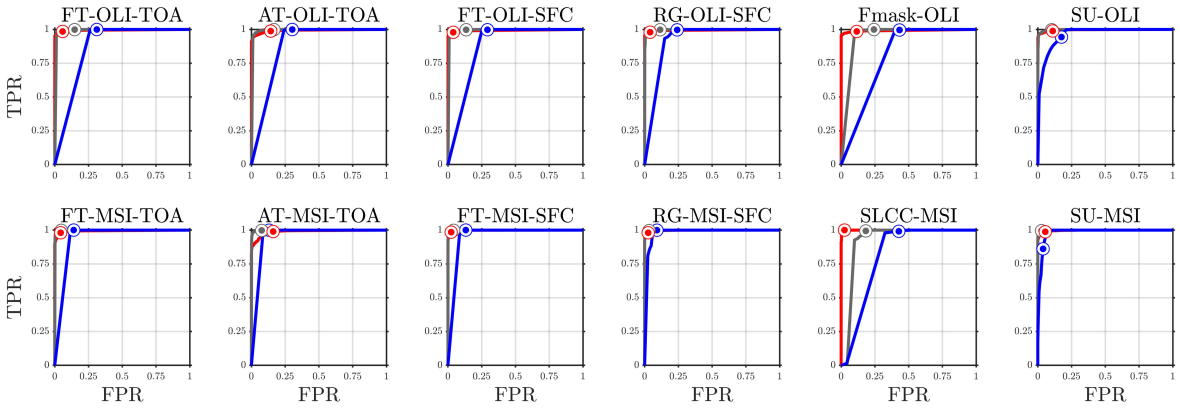


Figure 10: Receiver operating characteristics curves for the satellite retrievals using fSCA binarization thresholds on the terrestrial photography retrievals of 0.1 (red) 0.5 (gray) and 0.9 (blue) at 100 m spatial resolution. The satellite retrieval binarization threshold applied to the curves starts at 0 in the top right corner and ends at 1 in the bottom left corner of each panel. The markers show the points where the satellite retrieval binarization threshold matches that of the terrestrial retrievals, ideally these points should be in the top left corner of each panel. For an overview of the various satellite retrievals (indicated by the title of each panel) see Table 1. Results from several retrievals are not shown for visibility (see Table 2 metrics).

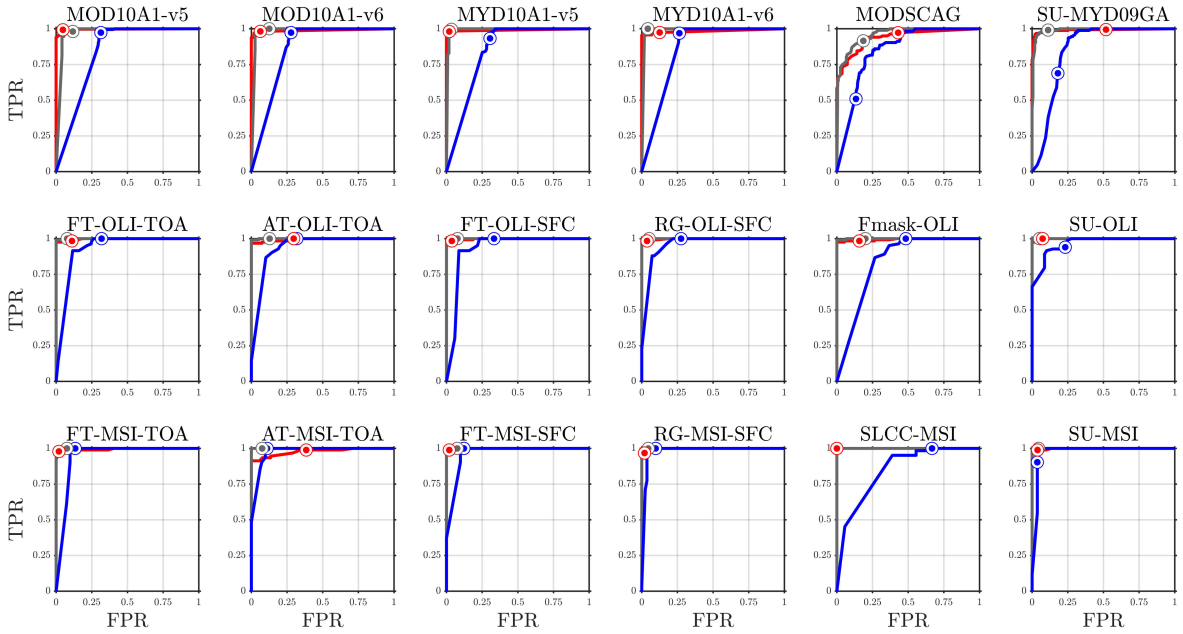


Figure 11: Same as Figure 10, but at 500 m spatial resolution. For an overview of the various satellite retrievals (indicated by the title of each panel) see Table 1. Results from several retrievals are not shown for visibility (see Table 3 metrics).

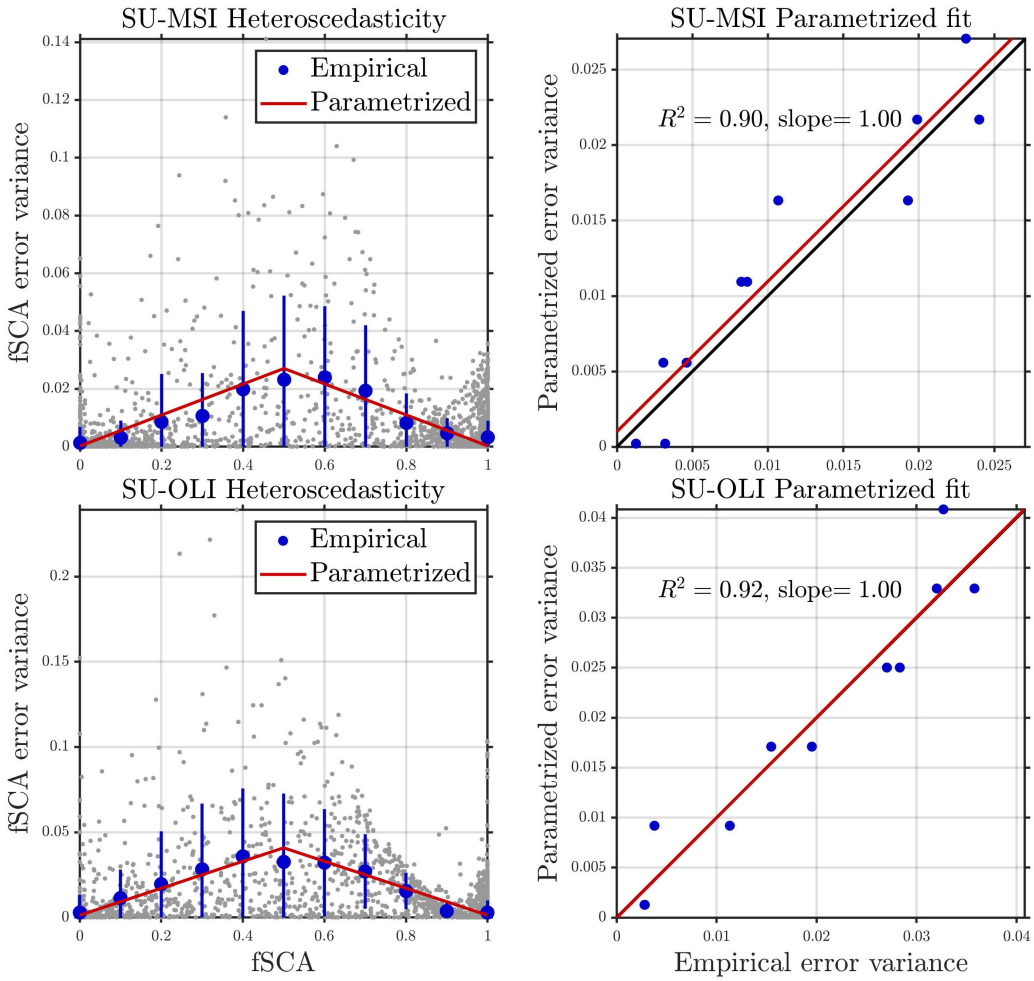


Figure 12: Heteroscedasticity in the error variance of the SU retrievals for Sentinel-2 MSI (top row) and Landsat 8 OLI (bottom row) at the 100 m scale. The left panels show the error variance as a function of fSCA with sample squared errors in gray, binned empirical error variances shown as blue circles (with vertical lines indicating ± 1 standard deviation), and the proposed parametrization shown by the red line. The right panels show the result of a linear regression of the parametrized variances on the (binned) empirical error variances with the red line showing the linear best fit line and the black line the 1:1 line. The slope of the linear best fit line and the square correlation coefficient (R^2) is also included.

Paper IV

Hyper-resolution ensemble-based snow reanalysis in mountain regions using clustering

J. Fiddes, K. Aalstad, S. Westermann

Submitted to *Hydrology and Earth System Sciences*, May 2019,
doi: 10.5194/hess-2019-37.

IV

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-37
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 10 May 2019

Hydrology and Earth System Sciences
Discussions

Hyper-resolution ensemble-based snow reanalysis in mountain regions using clustering

Joel Fiddes[1,2], Kristoffer Aalstad[1], and Sebastian Westermann[1]

[1]University of Oslo
[2]WSL Institute for Snow and Avalanche Research SLF

Correspondence to: Joel Fiddes (joel.fiddes@slf.ch)

Abstract. Spatial variability in high-relief landscapes is immense, and grid-based models cannot be run at spatial resolutions to explicitly represent important physical processes. This hampers the assessment of the current and future evolution of important issues such as water availability or mass movement hazards. Here, we present a new processing chain that couples an efficient subgrid method with a downscaling tool and data assimilation method with the purpose to improve numerical simulation of surface processes at multiple spatial and temporal scales in ungauged basins. The novelty of the approach is that while we add 1-2 orders of magnitude of computational cost by ensemble simulations, we save 4-5 orders of magnitude over explicitly simulating a high resolution grid. This approach makes data assimilation at large spatio-temporal scales feasible. In addition, this approach utilises only freely available global datasets and is therefore able to run globally. We demonstrate marked improvements in estimating snow height and snow water equivalent at various experimental scales using this approach. We propose this as a suitable method for a wide variety of operational and research applications where surfacecheck models need to be run at large scales with sparse to non-existent ground observations and with the flexibility to assimilate diverse variables retrieved by EO missions.

1 Introduction

Accurate simulation of energy and water cycles in high mountain environments is critical for a wide range of operational and research applications related to water resources and natural hazards, particularly in the current era of dramatic changes in mountain regions worldwide (Mankin et al., 2015). However, basic surface variables in many remote mountain areas remain poorly quantified despite large increases in the capacity of in-situ observations, remote sensing platforms and atmospheric model products. Spatial resolutions of 100 m are commonly recommended for modelling of land surface variables such as snow cover or surface temperature in complex terrain (Bierkens et al., 2015; Wood et al., 2011; Baldo and Margulis, 2018) and has come to be known as hyper-resolution (Wood et al., 2011). This is due to the fact that energy and mass fluxes exhibit strong lateral variation due to the effects of topography (Gruber S. and Haeberli W., 2007), and surface/subsurface properties such as vegetation cover (Shur and Jorgenson, 2007), ground material (Gubler et al., 2012) or snow distribution (Zhang, 2005; Liston, 2004) further compound these effects.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-37
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 10 May 2019

Hydrology and Earth System Sciences
Discussions

Most continental to global modelling studies operate on regular grids which has placed limitations on model resolutions despite advances in computing power. However, previous efforts using Hydrological Response Units, HRUs (Beven and Kirkby, 1979; Durand et al., 1993; Fiddes and Gruber, 2012), triangular irregular networks (Mascaro et al., 2015; Tucker et al., 2001), or multi-resolution approaches (Baldo and Margulis, 2018) suggest that regular grids are not only expensive but sub-optimal as often only subsets of watersheds require detailed model descriptions in order to characterise the system adequately. In addition, deterministic modelling schemes have limitations even at hyper-resolution due to errors in forcing data, particularly with fields such as precipitation which suffer from both measurement and modelling biases. The numerical weather prediction community has been addressing this problem for several decades using various data assimilation (DA) approaches. DA methods, often with Bayes' rule as a starting point, attempt to ingest uncertain observations into uncertain model simulations (Lahoz and Schneider, 2014; Carrassi et al., 2018). It is a class of methods that are implicitly Bayesian in that uncertainty in both simulation and observation are accounted for. These methods are diverse in design and application and the reader is directed to Liu et al. (2012) for a review relevant to the land surface community or Carrassi et al. (2018) for a timely overview. Only relatively recently has data assimilation started to be utilised in land surface modelling schemes (Liu et al., 2012), but it has already shown much promise in the current era of plentiful remote sensing data. Recently, ensemble-based DA has been successfully applied to the problem of improving snowpack estimates at various spatial scales (Margulis et al., 2015; Aalstad et al., 2018; Magnusson et al., 2017; Griessinger et al., 2016), this is particularly pertinent as it is widely recognised that estimating the spatial distribution of snow water equivalent (SWE) in mountain regions is currently one of the most important unsolved problems in snow hydrology (Dozier et al., 2016) and in understanding spatial distribution of other processes dependent on the snowpack mass balance, such as the surface energy balance.

Ensemble-based data assimilation revolves around the use of an ensemble (i.e. a collection) of model trajectories. Each trajectory is referred to as an ensemble-member or particle, for economy we will use the latter. An ensemble allows for the quantification of uncertainty through the prior (before assimilation) and posterior (after assimilation) distribution of particles. The use of an ensemble increases the computational burden, often adding orders of magnitude to computation times. Given that computation time is practically limited, in ensemble-simulation there is always a trade-off between a model's spatio-temporal resolution and the number of particles. Both are desirable, given that higher spatio-temporal resolution (is expected to) increases model realism whereas a higher number of particles allows for improved uncertainty estimation. This is why the dual quest for efficiency in models and DA is important. We argue that sometimes some of the resources that are spent on explicit high resolution spatial modelling could be better spent on the ensemble. When discussing computational expense it's worth noting that the intended application is important to consider. Given a large HPC infrastructure and enough time, today, we have the ability to use brute force deterministic numerical simulations to solve many resource intensive problems. However, the question is (a) what better purposes could that computation time be used for (e.g. uncertainty quantification) and (b) are we producing a final product (where one off large simulations are tolerable) or as is more commonly the case, at least in research (but also operational centres), are we part way through a development cycle where we expect to make many iterations in order to gain knowledge of the system. In this second case there is a strong motivation for methods that allow quick development cycles and knowledge gain. The previously published TopoSUB and TopoSCALE models (Fiddes and Gruber,

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-37
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 10 May 2019

Hydrology and Earth System Sciences
Discussions

2012, 2014) are hyper-efficient approaches which may provide a solution for this problem, particularly in data sparse regions. TopoSUB is a subgrid method that permits order of magnitude efficiency gains in applying numerical models over large areas. It achieves this by using a multivariate clustering of input predictors (normally topographical parameters) to reduce the number of simulations required to accurately represent surface heterogeneity by orders of magnitude. TopoSCALE provides point scale meteorological forcing at any given point on the earths surface by downscaling gridded reanalysis (or other atmospheric model data) using pressure levels to account for gradients with elevation and topographic correction for surface energy balance terms. The computational resources saved by not simulating domains explicitly in 2D can then be redirected to ensemble simulation for the purpose of data assimilation or uncertainty analysis in general. This approach has successfully been used to generate a regional scale permafrost map at 30 m resolution (Fiddes et al., 2015).

In this paper we present a new processing chain that couples an efficient subgrid method (TopoSUB), a downscaling tool (TopoSCALE) and data assimilation method with the purpose to improve numerical simulation of ground surface processes at multiple spatial and temporal scales in ungauged basins. The novelty of the approach is that while we add 2 orders of magnitude of computational cost by ensemble simulations, we save 4-5 orders of magnitude over explicitly simulating a high resolution grid. This approach makes data assimilation at large spatio-temporal scales feasible. In addition, this approach utilises only freely available global datasets and is therefore able to run globally.

Applications of this approach are numerous and diverse as it addresses 3 common bottlenecks: (a) availability of an appropriately downscaled forcing (b) ability to apply complex models at high resolution over large areas and (c) addressing uncertainty in the model chain. Applications could for example include large scale assessments of mass movements, glacier mass balance, or snowpack water availability. By translating GCM/RCM results to local slope scale impacts with appropriate surface models, climate change impacts can be estimated at appropriate scales.

2 Methods

The modelling pipeline used in this study employs two previously described methods (1) TopoSUB (Fiddes and Gruber, 2012) and (2) TopoSCALE (Fiddes and Gruber, 2014). These tools are briefly described here for clarity, however the reader is directed to the original publications for full details. An overview of the full tool chain is given in Figure 1.

2.1 Surface model

The surface model used in this study, GEOtop, is a physically-based model originally developed for hydrological research (Endrizzi et al., 2014). It couples energy and water budgets, represents the energy exchange with the atmosphere and has a multilayer snow pack. Further information is given by Bertoldi et al. (2006); Rigon et al. (2006); Endrizzi (2007); Dall'Amico et al. (2011). A description of model uncertainty and sensitivity is given by Gubler et al. (2012). Model parameters and soil stratigraphy are setup as defined in Fiddes et al. (2015).

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-37
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 10 May 2019

Hydrology and Earth System Sciences
Discussions

2.2 Downscaling forcing

TopoSCALE is a scheme which generates point-scale model forcing using gridded atmospheric model datasets. It achieves this as follows: (1) interpolate data available on pressure levels: air temperature (T_a), relative humidity (RH), wind speed (U), wind direction (φ_U) to point of interest in order to provide a dynamic scaling at each timestep, (2) incoming longwave radiation ($L^{\downarrow}$) is scaled by accounting for downscaled T_a, RH and sky emissivity; (3) we apply a topographic correction to both radiation fields ($S^{\downarrow}$/$L^{\downarrow}$); (4) an elevation based lapse-rate is applied to precipitation, P. The output is a full set of scaled meteorological fields required to drive a numerical model at hourly timesteps.

2.3 Subgrid scheme

TopoSUB is a scheme which samples land surface heterogeneity at high resolution based on a DEM and other surface data (here SRTM-3, 30 m). Input predictors describing dimensions of variability are clustered with a K-means algorithm to reduce computational units in a given simulation domain to a set of clusters. A 1-D surface model is then applied to each cluster using its mean physiographic properties. This approach allows multiple orders of magnitude savings in computational effort over distributed approaches. For example, a simulation domain represented by an ERA5 grid cell (25 km $\times$ 25 km) contains approximately 10^6 SRTM-3 pixels. This domain can be simulated using 100 TopoSUB clusters, which represents a 10^4 reduction in computational load during simulation.

2.4 Data assimilation

We build on previous efforts (e.g. Girotto et al., 2014; Margulis et al., 2015; Aalstad et al., 2018) that focus on the reanalysis of snowpack characteristics (particularly SWE and HS) through ensemble-based assimilation of fractional snow covered area (fSCA) retrievals from optical satellite sensors. We choose to use fSCA retrievals because currently only optical satellite sensors can offer the resolution, coverage, accuracy and breadth of information needed to constrain snowpack simulations in complex terrain (see Dozier et al., 2016). We use fSCA retrieved from the MODIS sensors onboard the Aqua and Terra satellites. These retrievals have a sub-kilometric spatial resolution and a near daily equatorial revisit frequency (in the absence of clouds), so the reanalysis we perform could be applied to any mountain range on Earth. By assimilating fSCA observations we exploit the dynamic information content contained in the depletion of the fractional snow-cover. The idea is that if one grid-cell melts out later than another, there must either have been more snow there to begin with, a slower ablation, or a combination of the two and vice-versa for an earlier melt out (Aalstad et al., 2018). This is the essence of traditional snow reconstruction where the snowpack is built up in reverse from the observed date of disappearance of the snow-cover to the day of peak SWE using modelled snowmelt rates (Martinec and Rango, 1981; Dozier et al., 2016). By using ensemble-based DA we can account for uncertainties in the remotely-sensed fSCA depletion, the meteorological forcing and the snow model that are ignored in traditional reconstruction (Slater et al., 2013) and arrive at an improved reanalysis (Girotto et al., 2014). Snow reanalysis problems are best approached using batch smoother DA algorithms rather than the more commonly used filters since

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-37
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 10 May 2019

the snowpack has a long memory (i.e. high temporal autocorrelation) relative to (e.g.) synoptic-scale weather (Margulis et al., 2015; Aalstad et al., 2018). By using a smoother that assimilates all the fSCA retrievals during the ablation season at once to constrain the ensemble of annual snowpack trajectories, we are able to use the observed ablation to inform the accumulation season which would not be possible with a particle filter.

2.4.1 Generating the prior ensemble

In line with previous studies (e.g. Raleigh et al., 2015), we assume that the main source of uncertainty in modelling the snowpack is in the meteorological forcing and specifically the main variables that control the mass and energy balance, namely air temperature (T_a), precipitation (P), incoming shortwave ($S^{\downarrow}$) and longwave ($L^{\downarrow}$) radiation. To generate the prior ensemble we perturb the forcing time series using normally (T_a, $S^{\downarrow}$, $L^{\downarrow}$) and log-normally (P) distributed multiplicative perturbation parameters that are fixed throughout the annual integration. Following Navari et al. (2016) we generate a correlated ensemble of perturbation parameters for the different forcing variables. This is to avoid unrealistic perturbations such as a large increase in both precipitation and shortwave radiation. We do this in two steps. First, generate independent perturbation parameters for each of the forcing variables using normal and lognormal random draws. Secondly, we account for the correlation between the different perturbation parameters by performing a Cholesky decomposition of the covariance matrix. All hyper-parameters used in generating the prior ensemble are given in Table 1.

2.4.2 Particle batch smoother

When performing DA we are usually interested in approximating the Bayesian posterior: the probability of model trajectories given the observations. The DA method employed in this study is the particle batch smoother (PBS) presented in the context of snow reanalysis in Margulis et al. (2015). The PBS is a basic importance sampling particle filter where no resampling takes place (see Van Leeuwen, 2009). This means that it is equivalent to the generalized likelihood uncertainty estimation (GLUE) with a formal likelihood function (Beven and Binley, 1992). The apparent advantage of this smoother is that, unlike the ensemble smoother (ES), it makes no assumptions about the linearity of the model or the Gaussianity of the error statistics (Van Leeuwen and Evensen, 1996). This can also be a disadvantage in higher dimensional problems where the method is prone to degeneracy and large sampling error unless a very large number of particles is used (Van Leeuwen and Evensen, 1996; Van Leeuwen, 2009). Nonetheless, for snow reconstruction problems where the dimensionality of the parameter space is relatively low, the PBS has been shown to outperform the ES even with a moderate number of particles (Margulis et al., 2015; Aalstad et al., 2018). Crucially, using the PBS instead of the ES (or its iterative variants) avoids the need for running more than one ensemble model integration, which would be more costly and difficult to reconcile with the clustering (TopoSUB) framework. Since the PBS is derived elsewhere (Van Leeuwen and Evensen, 1996; Van Leeuwen, 2009; Margulis et al., 2015), here we are content with presenting the analysis equation for the posterior and how to implement it for the snow reconstruction problem. Each particle represents a different annual integration of the snow model and will have a unique forcing history associated with it as dictated by the perturbation parameters described in Section 2.An overview of the tool chain is given in Figure 1.3.1. A priori, each of these histories is assumed to be equally likely. The observed fSCA depletion for the given water

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-37
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 10 May 2019

Hydrology and Earth System Sciences
Discussions

year and its assumed error structure is then used to constrain the ensemble of particles through the PBS analysis. In the PBS, the Bayesian posterior is approximated by a discrete probability mass function consisting of the posterior weights of each of the particles (model trajectories). As shown in Aalstad et al. (2018), when each particle is given an equal prior weight ($1/N_e$) and a Gaussian likelihood is used, the posterior weight for the j-th particle is given by

$$w_j = \frac{\exp\left(-||\mathbf{d}_j||^2_{\mathbf{R}}/2\right)}{\sum_{k=1}^{N_e}\exp\left(-||\mathbf{d}_k||^2_{\mathbf{R}}/2\right)}, \tag{1}$$

where the square norm of the innovations (residuals) for an arbitrary particle k is given by

$$||\mathbf{d}_k||^2_{\mathbf{R}} = \left(\mathbf{y}-\widehat{\mathbf{Y}}_k\right)^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{y}-\widehat{\mathbf{Y}}_k\right) \tag{2}$$

in which $^{\mathrm{T}}$ denotes the matrix transpose, $\mathbf{R}$ is the observation error covariance matrix, $\mathbf{y}$ is the observation vector containing the remotely sensed fSCA depletion for a given snow season, and $\widehat{\mathbf{Y}}_k$ is the predicted observation vector containing the corresponding modelled fSCA for particle k. The particle approximation of the Bayesian posterior represented by (1) improves as the number of particles increases. It should be clear from the analysis step (1) that by definition the posterior weights sum to one. Furthermore, unlike the ES, the PBS only changes the relative weights of the particles and not their position within the model space. This makes the PBS particularly attractive in a clustering framework as we do not need to rerun the ensemble after the analysis.

An important component of DA is the prescribed error covariance structure of the observations. Since the MODIS fSCA retrievals that we are assimilating are affected by various error sources that vary from day to day, such as atmospheric conditions and viewing angle, we assume that the observation errors are uncorrelated in time. Moreover, we assume a fixed observation error variance σ_y^2. Thereby, we use a simple scalar diagonal observation error covariance matrix $\mathbf{R} = \sigma_y^2\mathbf{I}$ where $\mathbf{I}$ is the identity matrix in line with similar studies (e.g. Margulis et al., 2015; Aalstad et al., 2018). This simplifies (2) which reduces to a simple square sum of innovations normalized by a constant (σ_y^2). We prescribe an observation error standard deviation of $\sigma_y = 0.13$ based on the estimate in Aalstad et al. (2018) (see Section 3.3). In order to make the model trajectories comparable to the fSCA retrievals during the analysis step, i.e. to generate the predicted observations $\widehat{\mathbf{Y}}$, an observation operator is required. We use a simple threshold on the SWE to determine the binary (snow/no-snow) snow-cover of each modelled grid cell based on values from Thirel et al. (2013) while also accounting for possible surface roughness. Due to the scale difference between the MODIS pixels (500 m) and the model grid cells (30 m), the modelled fSCA within a MODIS pixel is then simply the average of the binary snow cover in all model grid cells that fall within that pixel.

3 Data

3.1 Meteorological forcing

Driving climate data are obtained from the ERA5 reanalysis from ECMWF. This is the latest reanalysis from ECMWF that updates the ERA-Interim reanalysis. The main improvements are an increase of spatial resolution to 31 km, hourly temporal

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-37
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 10 May 2019

resolution, and increase in vertical model levels to 137. Accumulated values are now from the last time step and not last forecast as in ERA-Interim. This means that we can easily obtain the mean rates required to drive our numerical model by simply dividing these accumulations by the hourly time step (Fiddes and Gruber, 2014). Forcing data is detailed in Table 2. For each TopoSUB cluster, defined by the mean physiographic characteristics of a cluster, (Fiddes and Gruber, 2012) the ERA5 meteorological fields are downscaled using TopoSCALE (Fiddes and Gruber, 2014).

3.2 Surface properties

TopoSUB requires topographical parameters as input predictors to the clustering algorithm. We derive the following topographic parameters from the SRTM-3 digital elevation model: elevation, slope, aspect and sky view factor (proportion of visible sky). Surface cover is characterized in a simple 3 mode classification in order to approximate sub-surface stratigraphies: first a threshold on MODIS NDVI is used to classify vegetated surfaces, then a simple model further differentiates between steep bedrock and debris slopes. Further details are available in Fiddes et al. (2015).

3.3 Assimilated fSCA observations

We assimilate fSCA retrievals obtained from version 6 of the level 3 daily MODIS snow-cover product from the Terra (MOD10A1 product; Hall and Riggs, 2016a) and Aqua (MYD10A1 product; Hall and Riggs, 2016b) satellites. The retrieval algorithm is based on the inversion of a linear regression of MODIS normalized difference snow index (NDSI) on reference fSCA estimated from coincident Landsat imagery and it is given by the 'FRA6T' relationship in Salomonson and Appel (2006). The normalized difference snow index exploits the fact that snow is highly reflective in the visible but a good absorber in the shortwave infrared which differentiates it from most other natural surfaces (Painter et al., 2009). If cloud free retrievals are available from both Terra and Aqua retrievals for a given day then the Terra retrievals are used. Aalstad et al. (2018) compared MODIS fSCA retrievals to reference fSCA estimates obtained from a time-lapse photography, imagery from an unmanned aerial vehicle, as well as snow surveys at a site on Svalbard and obtained an RMSE of $\sigma_y = 0.13$ for the MODIS retrievals. This estimate is in reasonable agreement with those found at other sites (e.g. Mason et al., 2018), and so we use this as the as the observation error variance (σ_y^2) in the assimilation (Section 2.3.2).

3.4 Evaluation

3.4.1 Station data

SWE (mm) measurements obtained manually by observers are available at approximately biweekly intervals from snow profiles across Switzerland. Here we use the GCOS dataset which consists of 11 sites (Figure 2). We call these sites 'stations' throughout the paper. The dataset is openly available (Marty, 2017). Automatic HS (cm) measurements performed by sonic ranger (Campbell Scientific SR50) are available from the Intercantonal Measurement and Information System (IMIS) station network at 30 minute intervals. This is a high elevation station network that forms the backbone of the national avalanche service in Switzerland.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-37
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 10 May 2019

3.4.2 Airborne snow height retrievals

The Airborne Digital Sensor (ADS) opto-electric line scanners ADS80 and ADS100 from Leica Geosystems were used to acquire summer and winter stereo images which were processed into high resolution digital terrain models (DTM) using photogrammetry (Bühler et al., 2015). HS is then retrieved by subtracting summer from winter DTM and available for two footprints in the Davos region covering the Wannengrat area (3.5 × 7.5 km) and the Dischma area (7 × 17 km)) of high alpine terrain. The footprint of this survey is shown in Figure 2. These data are used for spatial evaluation of the scheme. Acquisition dates are 20 March 2012, 15 April 2013 and 17 April 2014. All snow depth maps were calculated using a summer DSM from 3 September 2013. The resolution of this dataset is 2m with a vertical RMSE of around +/- 30 cm (Vögeli et al., 2016; Bühler et al., 2015). The datasets are resampled to 100m (Vögeli et al., 2016) and used here to evaluate the methods. Snow depth in areas covered with forest, scrub, buildings and water bodies can not be determined using the ADS (Bühler et al., 2015) and are therefore masked out from the datasets. This dataset is openly available (Vögeli et al., 2016). Additionally as only a single summer (2013) DTM was used, all glacier areas were masked out to avoid errors associated with changing glacier surfaces. Glacier outlines were obtained from the GLIMS repository (Raup et al., 2007).

4 Experimental setup

In this study we conduct experiments at various spatio-temporal scales in order to comprehensively test the framework and assess its suitability for various applications. The experimental setup is shown in Figure 2. Simulations are run in 9 ERA5 grid boxes spanning the Swiss Alps. Each grid box contains at least 1 SWE measurement location and additionally several IMIS stations that are used to evaluate HS results. In addition we perform large area simulations on the entire Swiss Alps domain to explore how seasonal extremes are represented at large scale.

A prerequisite to the first two experiments (Section 4.1-4.2) is the PBS analysis step as described in Section 2 which generates the posterior weights matrix $\mathbf{W}_p$ based on PBS analysis units of MODIS cells. This then has dimensions $N_e \times N_p$ where N_e is the number of MODIS pixels and N_p is the number of particles (ensemble members). The following describes how $\mathbf{W}_p$ is used to generate posterior estimates of a given state variable (SWE or HS). The third experiment (Section 4.3) differs in that the analysis unit is the ERA5 grid cell itself and aims to correct aggregated grid level bias in forcing.

4.1 Point DA

Point scale DA is accomplished by simply mapping the DEM cell corresponding to point of interest (e.g. a validation station) to the corresponding MODIS pixel for that location. The $\mathbf{W}_p$ derived from the PBS analysis for that MODIS pixel is then used directly to generate the posterior estimate for that point. Model state results are obtained from the TopoSUB cluster that the DEM cell is a member of. Cumulative distributions are computed through the ranking of the ensemble of state variables followed by a cumulative summation of the correspondingly sorted weights. These distributions allow for the estimation of quantile values of the posterior model state.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-37
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 10 May 2019

Hydrology and Earth System Sciences
Discussions

4.2 Spatially distributed DA

The particle batch smoother has typically been applied at point scale or on regular grids. Here we generalise the method so as to fit the TopoSUB approach. The basic aim is to generate posterior weights for each TopoSUB cluster (which has no location, only physiographic attributes) so that the weights can be used in a highly scalable and powerful manner to generate different products such as a timeseries of the posterior median aggregated to give basin level statistics. The key challenge in this aim is how to map the spatial unit of the PBS algorithm, the MODIS pixel, which has a location in space, to a TopoSUB cluster which does not. We achieve this through the following

$$\mathbf{w}_c = \mathbf{W}_p \cdot \mathbf{a} \tag{3}$$

where $\mathbf{W}_p$ is the $N_e \times N_p$ weights matrix and $\mathbf{a}$ is a $N_p \times 1$ vector containing the fractional abundance (cover) of cluster c represented in each of the MODIS pixels. Here $\mathbf{w}_c$ contains the weight of each forcing history for cluster c and is computed for each of the clusters. This yields the weights matrix $\mathbf{w}_c$ that contains the weights of the forcing histories for each cluster. As a second step the weights $\mathbf{w}_c$ are renormalized to sum to one since that is not guaranteed in Eq. 3.

4.3 Coarse grid DA

The third DA method addresses bias in forcing at grid level only, it is the most efficient and lightweight of the three approaches. It also differs from the previous methods in that the PBS analysis step is computed at ERA5 grid unit not MODIS pixel unit. This makes the analysis step highly efficient and scalable over large areas. It is emphasised that while the two previous methods address both aggregated bias in forcing at grid level they also correct errors in the subgrid method (such as physiographic description) and downscaling (such as precipitation distribution), this method only corrects the bias at grid-level. However it is of interest if we seek a simple and robust way to feedback subgrid information to large scale atmospheric grid cells, in this case using ERA5:

1. Compute MODIS fSCA aggregated to the large scale atmospheric grid cell (ERA5) while accounting for clouds (max 10 % cloudiness tolerated). Cloud pixels are filled with mean fSCA value of the cluster to which the pixel belongs.
2. Compute the predicted observations, i.e. the modelled fSCA, for each cluster and aggregate these to the ERA5 grid cell scale by multiplying by cluster members.
3. Run PBS at ERA5 grid level to generate a single weight vector for the ensemble.

4.4 Run Configurations

All runs are performed using 100 particles, 150 TopoSUB clusters and cover the period 1 September 2011 - 1 September 2017. The specific temporal period covered by a given result is defined in the text. Throughout the paper a single year refers to the year in which melt occurs, e.g., "2012" refers to the period 1 September 2011 - 1 September 2012. These "water years" are prefixed with WY (e.g. WY2012) to avoid ambiguity. We measure computational effort through the number of GEOtop

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-37
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 10 May 2019

Hydrology and Earth System Sciences
Discussions

model runs (N_r) required per year in an ERA5 grid cell. Recall that the ERA5 grid cell is the fundamental unit on which the downscaling and clustering is performed. In terms of the number of clusters (N_s) and particles (N_e), this effort becomes

$$N_r = N_e \times N_s \,. \tag{4}$$

In the case of the configuration used in this study ($N_e = 100$, $N_s = 150$) this amounts to 1.5×10^4 individual model runs. At 30 m resolution, there are 10^6 model grid cells within a single $0.25°$ ERA5 grid cell. So, an explicit fully distributed simulation with 100 particles would require $N_r = 10^8$, a four order of magnitude increase in computational effort relative to the setup used in this study.

5 Results

5.1 Evaluating the forward model

Figure 3 shows performance of the forward model at the Weissfluhjoch (WFJ) research site (see Figure 2) assessed over the period WY2012-2017. It illustrates the performance of the downscaling routine in providing an adequate forcing to the model (forcing bias) and performance of the model in simulating the target variables SWE and HS when driven by downscaled ERA5 reanalysis (revealing model and forcing errors) and station observations (revealing model and observation errors). It shows that the TopoSCALE downscaling routine does a reasonable job of providing forcing to the forward model (top row) with the 0.71 °C RMSE for 2 m air temperatures being particularly low. Conversely, high wind velocities tend to be positively biased, most likely as wind fields representing the free atmosphere on pressure levels have no surface drag that would be present in surface observations. Modelled HS and SWE (bottom row) are captured fairly well capturing both the onset and melt of the snowpack. However, peak values are generally negatively biased with respect to observations and station driven model runs. WY2012 is an obvious outlier with large snowfalls not captured by ERA5 precipitation. This can be seen by cumulative precipitation totals computed with and without WY2012 totals (Figure 3). This is reflected in simulated HS and SWE totals. The performance of the forward model can be analysed by driving with station measurements to remove most uncertainty associated with driving reanalysis data (but with residual observation errors). ERA5 driven simulations are comparable or even outperform station runs in WY2013 and WY2014.

5.2 Point DA

In this experiment we compare the prior and (single pixel) posterior HS and SWE for WY2016 to the measured values at the respective stations. An example at Truebsee GCOS station (Engelberg) is shown in Figure 4. This figure demonstrates the effect the assimilation has not only on the fSCA (which is assimilated), but also on estimates of the other state variables (in this case SWE) which get closer to independent observations. Here you see clearly how the posterior estimate of SWE (blue shading) is constrained by the assimilation and the posterior median (blue line) is much closer to validation SWE observations than the prior (red line). We then scaled this up to 9 ERA5 grid boxes that span the Swiss Alps and contain 11 GCOS SWE stations. Additionally each box contains multiple IMIS stations measuring HS, which we also looked at in the interest of obtaining more

validation data. Significant improvements in the posterior were seen in the estimation of both variables (Figure 5). We found improvement in SWE was greater than that of HS. We hypothesize that this is due to representation of snow densities in the snow model. However, the improved representation of the snowpack mass balance as shown by improved SWE estimates is the main variable of interest in our approach. Stations where DA performed worse than the prior (supplement) can be attributed to poorly characterised melt seasons, lack of MODIS retrievals, and/or the MODIS retrievals not being representative at the scale of the observations (c.f. Section 6).

5.3 Spatially distributed DA

We evaluated the performance of the method in improving the spatial patterns and absolute quantities over large areas using data from an airborne digital sensor which has been used to generate high resolution surfaces of HS in WY2012, WY2013 and WY2014 (Figure 6, WY2014 only). Both WY2012 and WY2014 show marked improvement in all spatial statistical measures including the mean value, standard deviation (indicating increased variation) and error statistics such as RMSE and bias. WY2013 shows little improvement. We would expect a better performance for SWE than for HS due to the previously mentioned issues with the modelled snow density (see Section 5.2 and Figure 5). Figure 6 shows how the 90th percentile range is constrained by the analysis going from the prior to the posterior. Figure 7 shows probability density distributions for observations, prior and posterior in WY2014. The shape and moments such as the mean more closely match the observations in the posterior distribution. However, the method fails to capture the very highest accumulations in the distribution (> 2.5 m), possibly due to averaging effects of generalising weights to TopoSUB clusters.

5.3.1 Interannual validity of weights

We tested the ability of weights obtained in a given hydrological year to improve results in a different year. We did this by looking at statistics on the Dischma basin through a cross validation exercise where each year was forced with results from the two other years (WY2012, WY2013, WY2014). Posteriors forced by weights of other years improved performance over priors in all cases (Figure 9). This suggests that the DA method here also works to correct errors that are consistent from year to year. This could be related to spatial patterns of melt, a consistent bias in the forcing or errors in the model itself. This is an interesting result that suggests that while this method is primarily a post processing method it could be used to improve now/forecasts by using previous year weights. Additionally an analogue approach could be used to find years of best fit to current season in order to select weight sets (Kolberg and Gottschalk, 2010).

5.3.2 Large-scale application: Seasonal variability

December 2016 was an extremely snow poor month and start to the winter season. Many ski areas throughout the Alps could not open until late January due to lack of snow. We compare this to December 2011 which was relatively snow rich with above average precipitation and average temperatures for the month (cf. www.meteoswiss.admin.ch). Specifically, we investigated how open-loop runs in two contrasting seasons compared to observed spatial patterns of fSCA from MODIS and SLF reports

dated 22 December of respective year (Figure 10). The model was found to compare well with both spatial patterns of fSCA and SLF snow depth maps, which are an operational product created by interpolating station data constrained by AVHHR observed snow extent. Both fSCA and HS show snow free zones deep into alpine valleys during December 2016, and absence of snow over the northern regions and Jura Mountains. The red box indicates domain of DA runs for these seasons shown in Figure 11.

Next we zoomed in and compared spatial patterns of HS between the deterministic open loop and Posterior run which demonstrated that DA has increased elevation gradients of variability by reducing HS in valley bottoms and increasing it on higher slopes. DA as mentioned previously, therefore has the effect of increasing variability in the snow cover distribution. Snow cover extent estimation is also improved by DA with increased snow free area in valley bottoms showing improved fit to MODIS observations.

5.4 Coarse scale DA

Aggregated series of observed and modelled fSCA are computed at ERA-grid level. This could equally be a hydrological unit such as a basin. The main idea is to correct grid level biases in the forcing only. If we assume this is the main source of uncertainty, especially with a view to correct large scale biases, this is an effective method to apply at the scale of meteorological reanalysis. Data from WFJ is used to illustrate this point. Figure 8 shows two contrasting snow season WY2012 (high) and WY2014 (low) where mean snow depths and SWE differed by a factor of 2, as recorded at WFJ (Figure 3). We compare the total ERA5 precipitation (PSUM) over the winter period Jan-April 2014 (400 mm) and compare to totals recorded over the same period at WFJ station (350 mm), ERA5 captures WFJ totals well. It should be added that there is some elevation difference between the ERA5 grid (2024 m asl) and the WFJ station (2560m asl). However, in WY2012 we see quite a different story. The ERA5 grid gives us slightly higher PSUM values of 440 mm whereas the measured PSUM was almost double this at 826 mm. Figure 8 shows how grid level biases in the driving forcing from ERA5 have been successfully decreased in WY2012 resulting in increased SWE totals, whereas in WY2014 where ERA5 performance was much better (cf. Figure 3), DA has had a negligible effect. This simple approach is an extremely cost effective method of assimilating slope scale subgrid information (in this case fSCA) to correct coarse grid scale forcings (ERA5). It is additionally generic enough that it could be used with various other subgrid observations such as soil moisture, to improve grid level responses.

6 Discussion

In the following discussion some emphasis is placed on sources of uncertainty arising due to generally unknown errors in both the model, observations and forcing. These errors can be systematic (bias) or random as well as errors of representativeness (e.g. Lahoz and Schneider, 2014). Accounting for the uncertainty that results from these errors is an important component of any DA framework.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-37
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 10 May 2019

6.1 Sources of error

6.1.1 Forcing bias

In this work we encounter two forms of bias in the forcing, firstly (i) grid level inputs or bias in the forcing ERA5 reanalysis. This may exist due to e.g. a bias in assimilated observations (Synop stations tend to be in valley bottoms) or errors, omissions, parameterisations in the atmospheric model itself. How different is the forcing from observed grid averaged conditions? Of course this is a very difficult question to answer. Although with products such as precipitation radars such comparison of model and measured grid integrated precipitation may be possible. The second (ii) is error (random or bias) in the downscaling routines or disaggregation of the forcing at subgrid level. Do we get gradients along topographic correct? These sources of error could well be reinforcing or indeed cancelling, as they can be independent sources of error. In the approach of spatial DA we address both systematic and random error in the forcing but with an emphasis on the former. There is no 2D redistribution in terms of longitude and latitude position of a grid box. All members of a cluster are equally perturbed and clusters do not have x,y coordinates. In point DA, again both sources are addressed but with a stronger focus on (ii) as the data assimilation is done at MODIS pixel level and therefore redistributes precipitation not only with topographic parameters but also in a spatial x,y sense. In grid DA we only address (i) which could be a useful approach in differentiating and quantifying sources of bias as well as simply and robustly addressing the question of grid level bias.

6.1.2 Model error

We do not focus on structural errors in the forward model as this was not the subject of this study, and further the methods are designed to be quite independent of model type. However it is worth commenting that the majority of the results in this paper have focused on HS due to higher data availability. However Figure 5 shows that results are significantly better for SWE, possibly due to errors in the model densification parameterisations. This is however reassuring as HS results can be interpreted as conservative and therefore if we were able to validate more extensively against spatial distributed SWE measurements, we would likely see improved results.

6.1.3 Melt period definition

An important feature to mention and not often addressed by DA studies (Morzfeld et al., 2018, is a nice exception), is sensitivity of data assimilation methods to the observations chosen for assimilation. In the case of fSCA assimilations a melt period is defined as this is when the observations provide information about the snow depletion curve (e.g. Aalstad et al., 2018). We identify the end of the snowpack as the first day the fSCA values reach zero. There may be short increases in fSCA after this date but these will generally be late spring/summer snowfalls that are transient and melt rapidly. However, this date is the first available zero fSCA observation which does not necessarily equate to the exact date the snowpack melts-out as there can be a lack of observations due to cloud cover. Therefore this should be considered a source of potential bias in the system. We then found that a fixed window of 30 days prior to this date was a simple and robust way of defining the melt period. We trialled

other methods of automatically defining the end of the "complete snowcover" period but we did not find a way to do this that could work robustly over several hundred thousand MODIS pixels. Additionally, the MODIS products are quite noisy as they are generated from an empirical relationship with higher resolution Landsat data (Salomonson and Appel, 2006). This adds to the difficulty in defining a robust, general algorithm that defines the start of the melt period. As mentioned above this is a little discussed topic but due to the sensitivity of final results to the chosen method, would certainly benefit from further research efforts.

6.1.4 Scale issues in assimilation

The scale difference between validation data (station or snow profile) and fSCA retrievals from MODIS creates several issues. In this study many of the sites are in valley bottoms so they are accessible on a regular basis. However, this creates a bias caused by how representative is the point measurement of the larger MODIS pixel footprint. In Alpine valley bottoms there tends to be a lot of infrastructure, housing and rivers, which will tend to be snow free earlier (or never snow covered) as compared to the station site that will be well protected from interference allowing natural accumulation of snowfall. Therefore the MODIS footprint will tend to be observed to be "snow free" earlier than the validation point. In addition features such as a rivers, road clearance, urban heat islands are not considered in the modelling and will generate bias in the data assimilation. The most reliable sites for data assimilation, or actually we should say for validating the method, are therefore at high elevations away from effects due to human activity or infrastructure that are not considered in the model. Figure 12 gives an example of a DA failure that is not due to the DA algorithm, this has worked well, but the representativeness of the fSCA retrievals. As you can see the posterior is pulled in the direction of what has been detected to be the main melt period (red dots). Erroneous snowfalls during late spring/ early summer are ignored as expected. End of the winter snowpack as detected by the fSCA retrievals has been correctly identified around the beginning of April. However this site is in the middle of Zermatt town and the MODIS pixel will likely contain signal from urban effects unaccounted for by the model.

6.1.5 Observational errors

In addition to the scale issues, there are actual errors and cloud-induced data gaps in the MODIS retrievals. This could be incorrectly classified clouds (as snow or vice-versa) or uncertainty in the empirical fSCA algorithm. In addition the method can also suffer from a lack of observations due to persistent cloudiness at key points in the melt-period which will create uncertainties during DA. It may be worthwhile to consider fSCA retrievals from different higher resolution satellite constellations such as Landsat (30 m resolution) and Sentinel-2 (20 m resolution). This would increase the chance of obtaining cloud free scenes as well as reduce representativeness errors even at resolutions as high as 100 m. Furthermore, the aggregation of higher resolution retrievals would lead to a reduction of random error. The effective MODIS footprint of individual pixels can be quite variable and differs markedly from the nominal 500 m pixel resolution when the view angle deviates from nadir (Dozier et al., 2008). So, even for gridded applications, there is a considerable representativeness error in MODIS fSCA.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-37
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 10 May 2019

6.2 Applications

With the methods described in this paper a range of processing pipelines can be built to address a wide array of both research and operational problems. Specific strengths of the approach are:

- Slope scale forcing (climate, reanalysis, forecast) globally.
- Explicitly include the effects of high resolution topography on surface-atmosphere interactions.
- Efficient method to make large ensemble simulations feasible.
- Data assimilation to correct bias in forcing and quantify uncertainty.

Perhaps most importantly this approach allows applications to be built in remote regions where dense observation networks do not exist, such as High Mountain Asia or parts of North America. These capabilities allow for operational applications such as large area mass movement assessments related to dynamics of surface and subsurface processes. Driving the system with NWP (forecast) data would allow nowcasting/ forecasting applications to be setup with a suitable assimilation framework such as the EnKF. While the assimilation of fSCA would be less informative in a sequential method (such as the EnKF), ensemble simulations would still provide a useful quantification of uncertainty.

Transient climate change studies using a combination of reanalysis and climate model data (e.g. CMIP5) would be a valuable research application based on this approach, for example quantifying dynamics of permafrost extent over large areas according to a range of scenarios and models or generate a regional snowpack reanalysis product with projected future changes.

An important operational application and currently a great humanitarian need in many remote regions in Asia (e.g. Afghanistan/ Tajikistan) could be an operational avalanche forecast based on a snowpack model (e.g. SNOWPACK, CROCUS), driven by an NWP ensemble to generate a large area probabilistic forecast where few ground stations exist. This would be a relatively cost-effective system to deploy and give first order hazard assessment where none currently exists.

6.3 Further work

For the moderate (MODIS-like) resolution satellites we hope that products will emerge from Sentinel-3 and VIIRS to prolong and expand the MODIS record. For high resolution sensors there is a strong need for operational products that ideally combine available and emerging sources such as Landsat8, Sentinel-2. The French inter-agency initiative THEIA Land Data Centre is starting to produce Sentinel-2 based snow cover products. Additionally, improved cloud masks are needed as misclassified clouds are potential and significant sources of error in the framework. Both too strict and too relaxed cloud masking is problematic, the former leads to throwing out valid and potentially important retrievals while the latter corrupts the signal that we are trying to assimilate (the actual snow cover depletion).

For additional datasets (other than fSCA) land surface temperature can be retrieved from both MODIS and Landsat and provide a means to constrain uncertainty in the surface energy balance. However, the current MODIS products are coarse at 1 km and therefore not ideal for mountain regions. Snowmelt status (i.e. binary melting/not melting) from synthetic aperture

radar (SAR) e.g. Sentinel-1 has potential to constrain uncertainty in fSCA during cloudy periods. There is also potential from ICESAT2 which could provide a way to constrain snow depth directly.

Assimilation of sparse point data could be an important extension of this work to provide means to assimilate data sources such as ICESAT2 but also be used to improve TopoSCALE by assimilating point data (stations) to improve the downscaling of reanalysis data. This could be interesting as where TopoSCALE performs most poorly is in valleys, where surface effects are poorly represented by the atmospheric model and this is precisely where stations tend to be most abundant globally. For real-time applications in remote regions extending the method to assimilate sparse observations is important as fSCA is known to have limited value in sequential (i.e real-time) data assimilation. We have shown there is some interannual validity of results in our limited test-case suggesting that systematic biases relevant to real-time applications could be addressed through reanalysis. Additionally, by creating a long term library of "best possible" reconstructions/reanalyses then training an "analogue ensemble" or even a more machine learning type approach like neural nets, could be promising.

7 Conclusions

In this study we have demonstrated a processing pipeline capable of producing improved land surface simulations at scale in ungauged regions. It consists of downscaling, subgrid and data assimilation components and uses only globally available datasets for both the model forcing and assimilated observations, it is therefore suitable for global applications. Specifically we have shown:

- Use of PBS data assimilation can significantly improve estimates of snowpack at various spatial scales.
- TopoSUB clustering efficiency gains make large ensemble simulations feasible.
- The methods can be used to reduce biases both at coarse atmospheric grid scale and also those related to the downscaling routines.
- The approach is suitable for regional to global applications due to efficiency and data requirements.
- A flexible set of tools allow various research and operational problems to be addressed where high resolution surface models are needed in heterogeneous terrain.

data was obtained from We propose this as a suitable method for complex model ensemble runs at scale i.e. large numbers of particles, large spatial areas or long temporal periods. Application areas include any problem where accurate slope scale forcings are required and surface atmosphere interactions need to be simulated at slope scale e.g. large area avalanche warning where the snowpack is explicitly simulated or regional-scale hazard assessment of mass movements where changing ground thermal regime is a risk factor. The toolchain can be flexibly driven by a range of forcings e.g. climate scenario data, reanalysis of past climate or real-time NWP and drive impact models for a range of domains e.g., hydrology, snowcover, soil stability or permafrost. New developments in multi-platform processing pipelines of high resolutions products from Sentinel-2 and Landsat will further improve the method in terms of representativity and availability of observations.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-37
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 10 May 2019

Hydrology and Earth System Sciences
Discussions

Code availability. All code used in this publication is available at https://github.com/joelfiddes/topoMAPP.

Author contributions. JF wrote the grant for postDoc SNF project: TopoSAT. JF and SW designed the study, KA contributed to implementation of the DA module and multiple aspects of the analysis. JF wrote the manuscript which has been contributed to and edited by all authors.

Competing interests. To the authors knowledge there are no competing interests.

Acknowledgements. We acknowledge the Swiss National Science Foundation project number P2ZHP2_165435 "TopoSAT: High resolution surface modelling of the Himalayan cryosphere with satellite data assimilation" and SatPerm (239918; Research Council of Norway) for providing funding for this study. We thank all data providers for making their data freely available.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-37
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 10 May 2019

Hydrology and Earth System Sciences
Discussions

References

Aalstad, K., Westermann, S., Schuler, T. V., Boike, J., and Bertino, L.: Ensemble-based assimilation of fractional snow-covered area satellite retrievals to estimate the snow distribution at Arctic sites, The Cryosphere, 12, 247–270, 2018.

Baldo, E. and Margulis, S. A.: Assessment of a multiresolution snow reanalysis framework: a multidecadal reanalysis case over the upper Yampa River basin, Colorado, Hydrology and Earth System Sciences, 22, 3575–3587, https://doi.org/10.5194/hess-22-3575-2018, https://www.hydrol-earth-syst-sci.net/22/3575/2018/, 2018.

Bertoldi, G., Rigon, R., and Over, T. M.: Impact of watershed geomorphic characteristics on the energy and water budgets, J. Hydrometeorol., 7, 389–403, https://doi.org/10.1175/JHM500.1, 2006.

Beven, K. and Binley, A.: The Future of Distributed Models: Model Calibration and Uncertainty Prediction, Hydrological Processes, 6, 279–298, https://doi.org/10.1002/hyp.3360060305, 1992.

Beven, K. J. and Kirkby, M. J.: A physically based, variable contributing area model of basin hydrology / Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant, Hydrol. Sci. Bull., 24, 43–69, 1979.

Bierkens, M. F. P., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., de Roo, A., Döll, P., Drost, N., Famiglietti, J. S., Flörke, M., Gochis, D. J., Houser, P., Hut, R., Keune, J., Kollet, S., Maxwell, R. M., Reager, J. T., Samaniego, L., Sudicky, E., Sutanudjaja, E. H., van de Giesen, N., Winsemius, H., and Wood, E. F.: Hyper-resolution global hydrological modelling: what is next?, Hydrol. Process., 29, 310–320, 2015.

Bühler, Y., Marty, M., Egli, L., Veitinger, J., Jonas, T., Thee, P., and Ginzler, C.: Snow depth mapping in high-alpine catchments using digital photogrammetry, The Cryosphere, 9, 229–243, 2015.

Carrassi, A., Bocquet, M., Bertino, L., and Evensen, G.: Data assimilation in the geosciences: An overview of methods, issues, and perspectives, WIREs Climate Change, doi:10.1002/wcc.535, 2018.

Dall'Amico, M., Endrizzi, S., Gruber, S., and Rigon, R.: A robust and energy-conserving model of freezing variably-saturated soil, The Cryosphere, 5, 469–484, https://doi.org/10.5194/tc-5-469-2011, 2011.

Dozier, J., Painter, T. H., Rittger, K., and Frew, J. E.: , Advances in Water Resources, 31, 1515–1526, https://doi.org/10.1016/j.advwatres.2008.08.011, 2008.

Dozier, J., Bair, E. H., and Davis, R. E.: Estimating the spatial distribution of snow water equivalent in the world's mountains, Wiley Interdisciplinary Reviews: Water, 3, 461–474, 2016.

Durand, Y., Brun, E., Merindol, L., Guyomarc'h, G., Lesaffre, B., and Martin, E.: A meteorological estimation of relevant parameters for snow models, Ann. Glaciol., 18, 65–71, 1993.

Endrizzi, , Gruber, S., , Dall'Amico, S., , Rigon, M., and R.: GEOtop 2.0: simulating the combined energy and water balance at and below the land surface accounting for soil freezing, snow cover and terrain effects, Geoscientific Model Development Discussions, 6, 6279–6341, 2014.

Endrizzi, S.: Snow cover modelling at a local and distributed scale over complex terrain, Environ. Eng., 2007.

Fiddes, J. and Gruber, S.: TopoSUB: a tool for efficient large area numerical modelling in complex topography at sub-grid scales, Geoscientific Model Development, 5, 1245–1257, 2012.

Fiddes, J. and Gruber, S.: TopoSCALE v.1.0: downscaling gridded climate data in complex terrain, Geoscientific Model Development, 7, 387–405, 2014.

Fiddes, J., Endrizzi, S., and Gruber, S.: Large-area land surface simulations in heterogeneous terrain driven by global data sets: application to mountain permafrost, The Cryosphere, 9, 411–426, 2015.

Girotto, M., Margulis, S. A., and Durand, M.: Probabilistic SWE reanalysis as a generalization of deterministic SWE reconstruction techniques, Hydrol. Process., 28, 3875–3895, 2014.

Griessinger, N., Seibert, J., Magnusson, J., and Jonas, T.: Assessing the benefit of snow data assimilation for runoff modeling in Alpine catchments, Hydrol. Earth Syst. Sci., 20, 3895–3905, 2016.

Gruber S. and Haeberli W.: Permafrost in steep bedrock slopes and its temperature-related destabilization following climate change, Journal of Geophysical Research: Earth Surface, 112, F02S18, 2007.

Gubler, S., Gruber, S., and Purves, R. S.: Uncertainties of parameterized surface downward clear-sky shortwave and all-sky longwave radiation., Atmos. Chem. Phys., 12, 5077–5098, 2012.

Hall, D. K. and Riggs, G. A.: MODIS/Terra Snow Cover Daily L3 Global 500m Grid, Version 6, Tile h18v01, https://doi.org/10.5067/MODIS/MOD10A1.006, Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center, last access: June 2018, 2016a.

Hall, D. K. and Riggs, G. A.: MODIS/Aqua Snow Cover Daily L3 Global 500m Grid, Version 6, Tile h18v01, https://doi.org/10.5067/MODIS/MYD10A1.006, Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center, last access: June 2018, 2016b.

Kolberg, S. and Gottschalk, L.: Interannual stability of grid cell snow depletion curves as estimated from MODIS images, Water Resour. Res., 46, 2010.

Lahoz, W. A. and Schneider, P.: Data assimilation: making sense of Earth Observation, Frontiers in Environmental Science, 2, 16, https://doi.org/10.3389/fenvs.2014.00016, 2014.

Liston, G. E.: Representing subgrid snow cover heterogeneities in regional and global models, J. Clim., 17, 1381–1397, 2004.

Liu, Y., Weerts, A. H., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., Seo, D.-J., Schwanenberg, D., Smith, P., van Dijk, A. I. J. M., van Velzen, N., He, M., Lee, H., Noh, S. J., Rakovec, O., and Restrepo, P.: Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities, Hydrol. Earth Syst. Sci., 16, 3863–3887, 2012.

Magnusson, J., Winstral, A., Stordal, A. S., Essery, R., and Jonas, T.: Improving physically based snow simulations by assimilating snow depths using the particle filter, Water Resour. Res., 53, 1125–1143, 2017.

Mankin, J. S., Viviroli, D., Singh, D., Hoekstra, A. Y., and Diffenbaugh, N. S.: The potential for snow to supply human water demand in the present and future, Environmental Research Letters, 10, doi:10.1088/1748-9326/10/11/114016, 2015.

Margulis, S. A., Girotto, M., Cortés, G., and Durand, M.: A Particle Batch Smoother Approach to Snow Water Equivalent Estimation, J. Hydrometeorol., 16, 1752–1772, 2015.

Martinec, J. and Rango, A.: Areal Distribution of Snow Water Equivalent Evaluated by Snow Cover Monitoring, Water Resources Research, 17, 1480–1488, doi:10.1029/WR017i005p01480, 1981.

Marty, C.: GCOS SWE data from 11 stations in Switzerland, https://doi.org/10.16904/15, 2017.

Mascaro, G., Vivoni, E. R., and Méndez-Barroso, L. A.: Hyperresolution hydrologic modeling in a regional watershed and its interpretation using empirical orthogonal functions, Adv. Water Resour., 83, 190–206, 2015.

Mason, T., Dumont, M., Dalla Mura, M., Sirguey, P., Gascoin, S., Dedieu, J. P., and Chanussot, J.: An Assessment of Existing Methodologies to Retrieve Snow Cover Fraction from MODIS Data, Remote Sensing, 10, doi:10.3390/rs10040619, 2018.

Morzfeld, M., Adams, J., Lunderman, S., and Orozco, R.: Feature-based data assimilation in geophysics, Nonlinear Processes in Geophysics, 25, 355–374, https://doi.org/10.5194/npg-25-355-2018, 2018.

Navari, M., Margulis, S. A., Bateni, S. M., Tedesco, M., Alexander, P., and Fettweis, X.: Feasibility of improving a priori regional climate model estimates of Greenland ice sheet surface mass loss through assimilation of measured ice surface temperatures, The Cryosphere, 10, 103–120, https://www.the-cryosphere.net/10/103/2016/, doi:10.5194/tc-10-103-2016, 2016.

Painter, T. H., Rittger, K., McKenzie, C., Slaughter, P., Davis, R. E., and Dozier, J.: Retrieval of subpixel snow covered area, grain size, and albedo from MODIS, Remote Sensing of Environment, 113, 868–879, https://doi.org/10.1016/j.rse.2009.01.001, 2009.

Raleigh, M. S., Lundquist, J. D., and Clark, M. P.: Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework, Hydrology and Earth System Sciences, 19, 3153–3179, doi:10.5194/hess-19-3153-2015, 2015.

Raup, B., Racoviteanu, A., Khalsa, S. J. S., Helm, C., Armstrong, R., and Arnaud, Y.: The GLIMS geospatial glacier database: A new tool for studying glacier change, Glob. Planet. Change, 56, 101–110, 2007.

Rigon, R., Bertoldi, G., and Over, T. M.: GEOtop: A Distributed Hydrological Model with Coupled Water and Energy Budgets, J. Hydrometeorol., 7, 371–388, https://doi.org/10.1175/JHM497.1, 2006.

Salomonson, V. V. and Appel, I.: Development of the Aqua MODIS NDSI Fractional Snow Cover Algorithm and Validation Results, IEEE Transactions on Geoscience and Remote Sensing, 44, 1747–1756, https://doi.org/10.1109/TGRS.2006.876029, 2006.

Shur, Y. L. and Jorgenson, M. T.: Patterns of permafrost formation and degradation in relation to climate and ecosystems, Permafrost Periglacial Processes, 18, 7–19, 2007.

Slater, A. G., Barrett, A. P., Clark, M. P., Lundquist, J. D., and Raleigh, M. S.: Uncertainty in seasonal snow reconstruction: Relative impacts of model forcing and image availability, Advances in Water Resources, 55, 165–177, doi:10.1016/j.advwatres.2012.07.006, 2013.

Thirel, G., Salamon, P., Burek, P., and Kalas, M.: Assimilation of MODIS Snow Cover Area Data in a Distributed Hydrological Model Using the Particle Filter, Remote Sensing, 5, 5825–5850, 2013.

Tucker, G. E., Lancaster, S. T., Gasparini, N. M., Bras, R. L., and Rybarczyk, S. M.: An object-oriented framework for distributed hydrologic and geomorphic modeling using triangulated irregular networks, Comput. Geosci., 27, 959–973, 2001.

Van Leeuwen, P. J.: Particle Filtering in Geophysical Systems, Monthly Weather Review, 137, 4089–4114, doi:10.1175/2009MWR2835.1, 2009.

Van Leeuwen, P. J. and Evensen, G.: Data assimilation and inverse methods in terms of a probabilistic formulation, Monthly Weather Review, 124, 2898–2913, doi:10.1175/1520-0493(1996)124<2898:DAAIMI>2.0.CO;2, 1996.

Vögeli, C., Lehning, M., Wever, N., and Bavay, M.: Scaling Precipitation Input to Spatially Distributed Hydrological Models by Measured Snow Distribution, Front Earth Sci. Chin., 4, 108, 2016.

Wood, E. F., Roundy, J. K., Troy, T. J., van Beek, L. P. H., Bierkens, M. F. P., Blyth, E., de Roo, A., Döll, P., Ek, M., Famiglietti, J., Gochis, D., van de Giesen, N., Houser, P., Jaffé, P. R., Kollet, S., Lehner, B., Lettenmaier, D. P., Peters-Lidard, C., Sivapalan, M., Sheffield, J., Wade, A., and Whitehead, P.: Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water, Water Resour. Res., 47, 1–10, 2011.

Zhang, T.: Influence of the seasonal snow cover on the ground thermal regime: An overview, Rev. Geophys., 43, RG4002, 2005.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-37
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 10 May 2019

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

Table 1. Hyperparameters (means, variances and correlations) defining the joint probability distribution from which the ensemble of multiplicative perturbation parameters are drawn. Based on Navari et al. (2016).

Perturbed variable	Marginal	Mean	Variance	Corr(T_a)	Corr(P)	Corr($S^{\downarrow}$)	Corr($L^{\downarrow}$)
Air temperature (T_a)	Normal	1	2.5e-5	1	-0.1	0.3	0.6
Precipitation (P)	Log-normal	1	0.25	-0.1	1	-0.1	0.5
Shortwave ($S^{\downarrow}$)	Normal	1	0.04	0.3	-0.1	1	-0.3
Longwave ($L^{\downarrow}$)	Normal	1	0.01	0.6	0.5	-0.3	1

Table 2. Description of the hourly fields obtained from the ERA5 reanalysis. All the columns headers are terms defined by ECMWF. 'levtype' refers to the level type: surf=surface, pl=pressure level. The 'type' is either: fc=forecast, an=analysis, or inv=invariant.

name	shortName	levtype	type	units
2 metre dewpoint temperature	d2m	surf	fc	K
Surface thermal radiation downwards	strd	surf	fc	Jm^{-2}
Surface solar radiation downwards	ssrd	surf	fc	Jm^{-2}
Total precipitation	tp	surf	fc	m
TOA incident solar radiation	tisr	surf	fc	Jm^{-2}
2m temperature	2t	surf	fc	K
Temperature	t	pl	an	K
Relative humidity	r	pl	an	%
U component of wind	u	pl	an	ms^{-1}
V component of wind	v	pl	an	ms^{-1}
Geopotential	z	surf	inv	$\mathrm{m}^2\mathrm{s}^{-2}$

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-37
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 10 May 2019

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

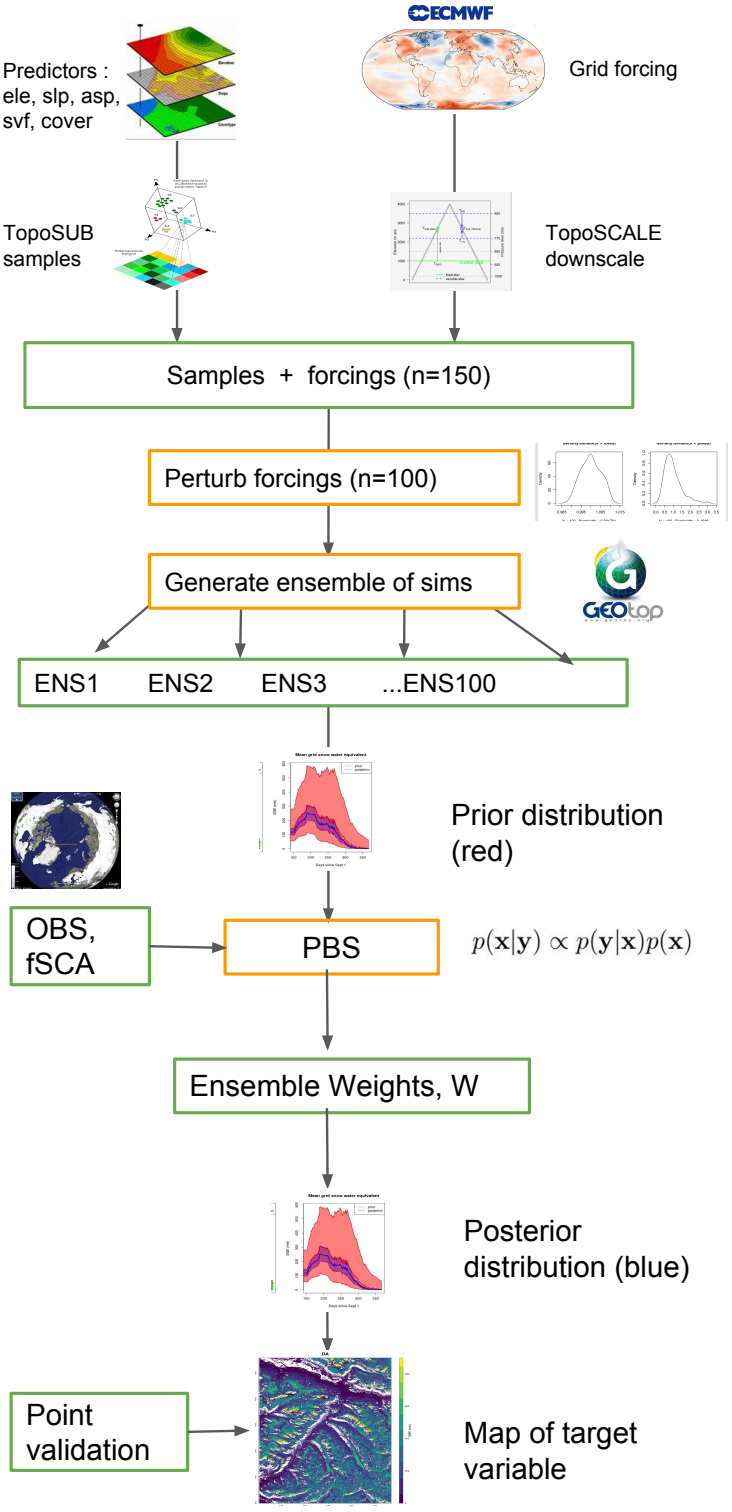


Figure 1. Schematic of the modelling setup.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-37
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 10 May 2019

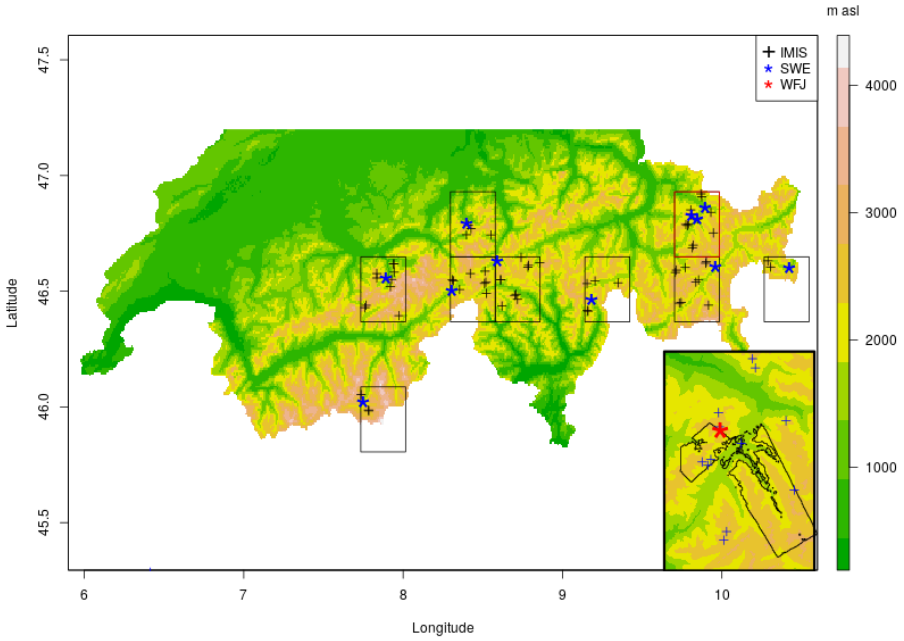


Figure 2. Experimental setup: 9 ERA5 grid boxes are simulated chosen for existence of GCOS SWE monitoring sites (11 stations). All IMIS stations in each box are used for evaluation (39 stations). Box in red is located the Weissfluhjoch research station as well as the flightpath of ADS data (inset).

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-37
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 10 May 2019

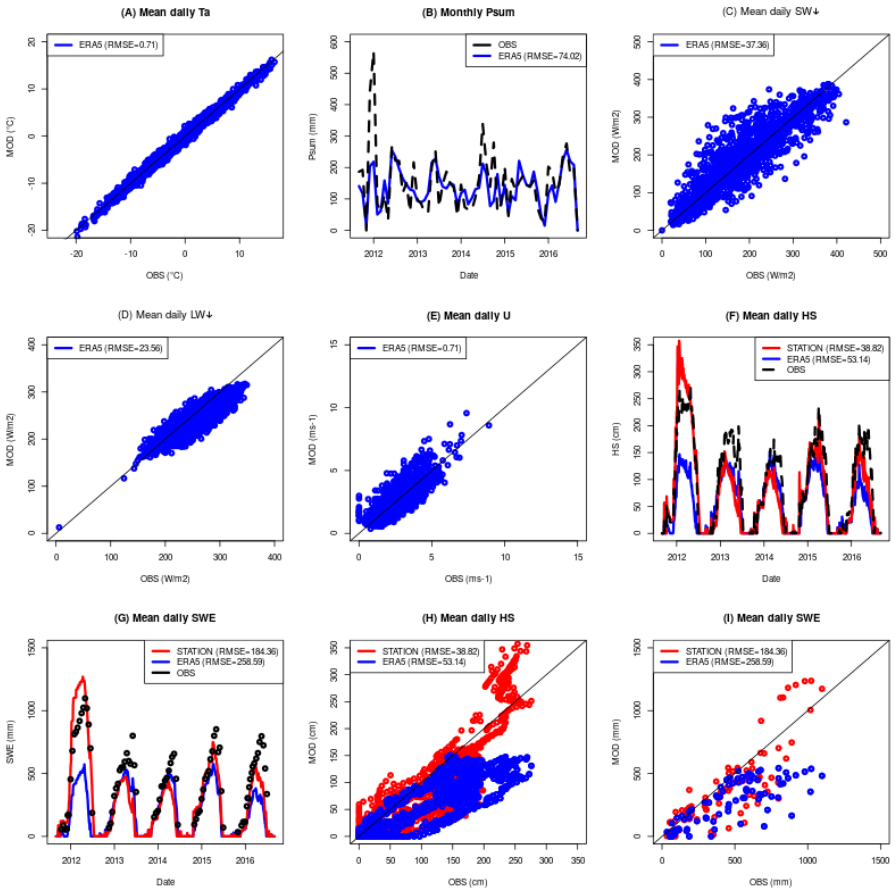


Figure 3. Multiyear simulations at station WFJ (WY2012-2017) in order to show baseline results for the modelling scheme. (A-E) assesses the downscaling scheme by showing downscaled ERA5 data (MOD) compared to station measurements (OBS). (F-I) assesses the simulation of target variables SWE and HS in both timeseries and scatterplots. Here, MOD is a simulation driven either by downscaled ERA5 or directly by station measurements. OBS are SWE and HS measurements made at the station. WY2012 is a clear outlier in poor performing ERA5 as shown by cumulative precipitation errors and in HS and SWE time series. HS and SWE scatter plots also show this low performance in high values attributed to WY2012. Additionally, ERA5 simulated HS is increasingly biased with depth as errors accumulate over the season to max depths. The same pattern is evident with SWE. It is worth noting that in differentiating sources of error these plots are useful. OBS - STATION approximates model error whereas STATION -ERA5 approximates the forcing error.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-37
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 10 May 2019

Hydrology and Earth System Sciences
Discussions

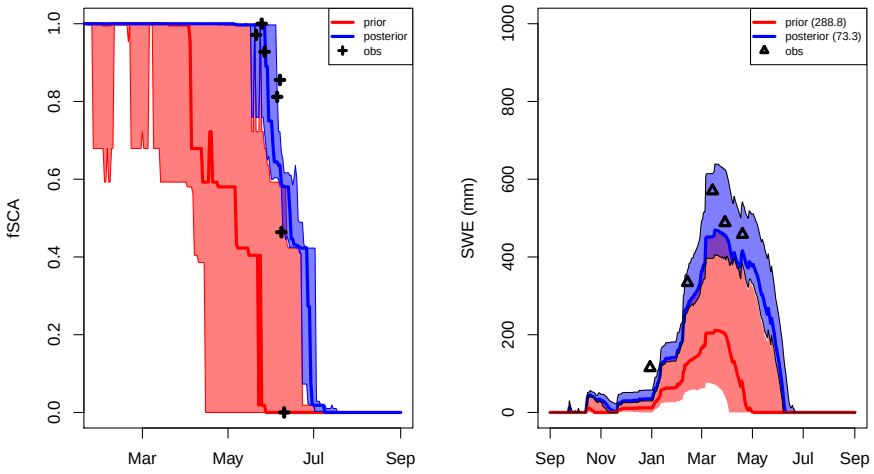


Figure 4. DA run at the Truebsee GCOS station, Engelberg. In the left panel is the assimilated observation, in this case fSCA represented by dots. Green dots show all observations from the combined MODIS products that are available for this stations location. Red dots indicate the observations that have been assimilated per the melt season definition. The shading and solid lines show the 90th percentile range and median of the prior (red) and posterior (blue) estimates. The right panel shows the target variable, SWE in this case. Posterior/prior are denoted in the same way. Black triangles indicate measurements used for validation.

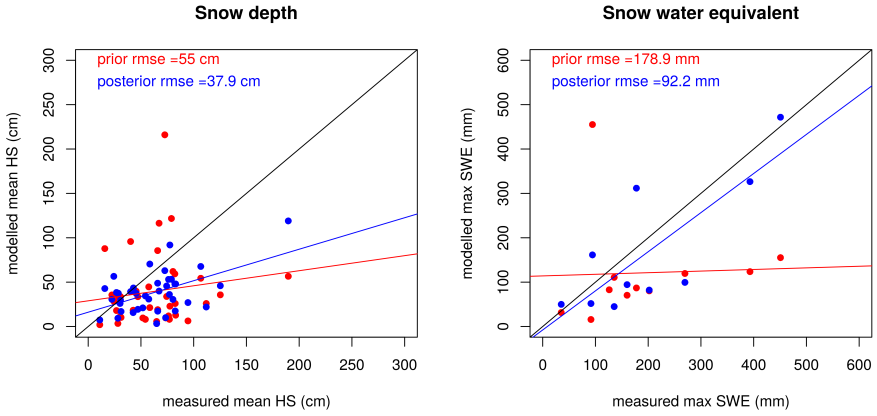


Figure 5. Simulated snow depth at IMIS stations (HS) and snow water equivalent at GCOS stations (SWE) for both the prior (red) and posterior median (blue) compared to observation mean. The mean is computed from all values over the entire WY2016. Posterior estimate is markedly improved in both variables. Regression lines compare the fit of posterior and prior estimates with respect to observations against the 1:1 line.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-37
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 10 May 2019

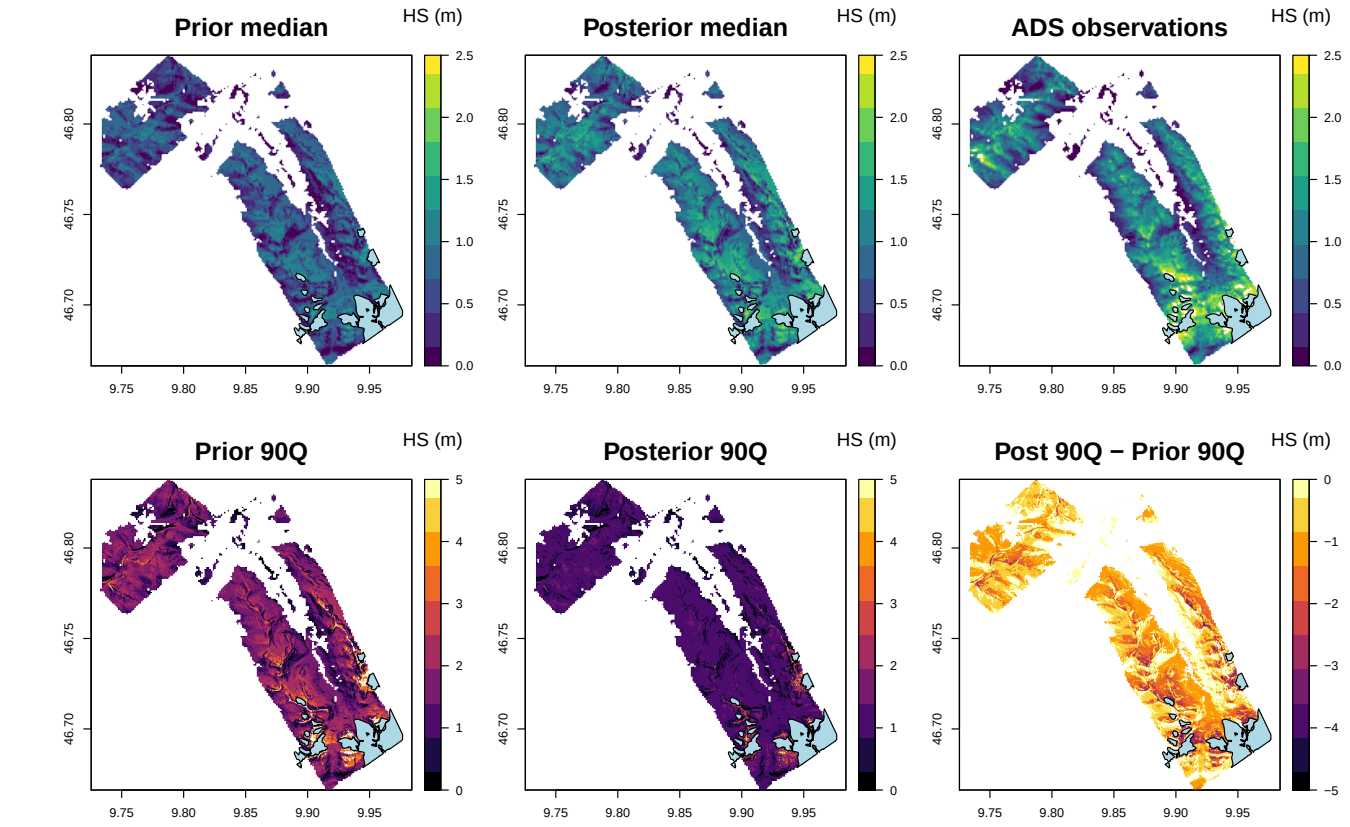


Figure 6. Topo row: Prior/posterior median and observations from ADS sensor flights in Davos region of 14 April 2014 (see Fig 2 for location). Bottom row: uncertainty represented by the 90th percentile range of the ensemble and reduction in uncertainty in the posterior. Glacier mask is shown in blue. Posterior median is improved with respect to observations and uncertainty is reduced by the DA scheme.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-37
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 10 May 2019

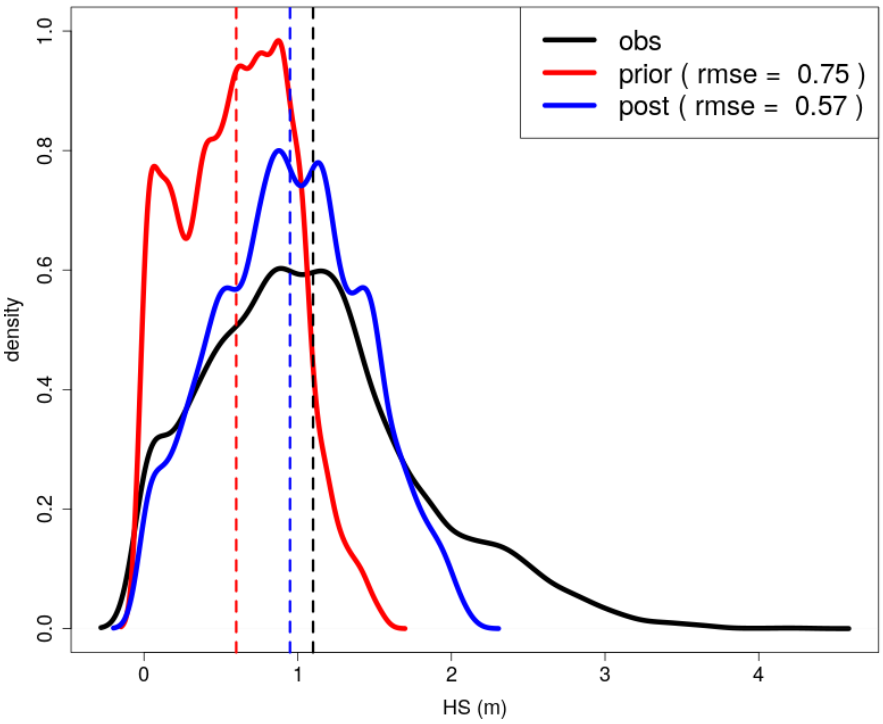


Figure 7. Density distribution plots for HS obs, prior and posterior within the ADS footprint for 14 April 2014 (see Figure 6). The observed distribution is better captured by the prior. Dashed lines give the respective mean values.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-37
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 10 May 2019

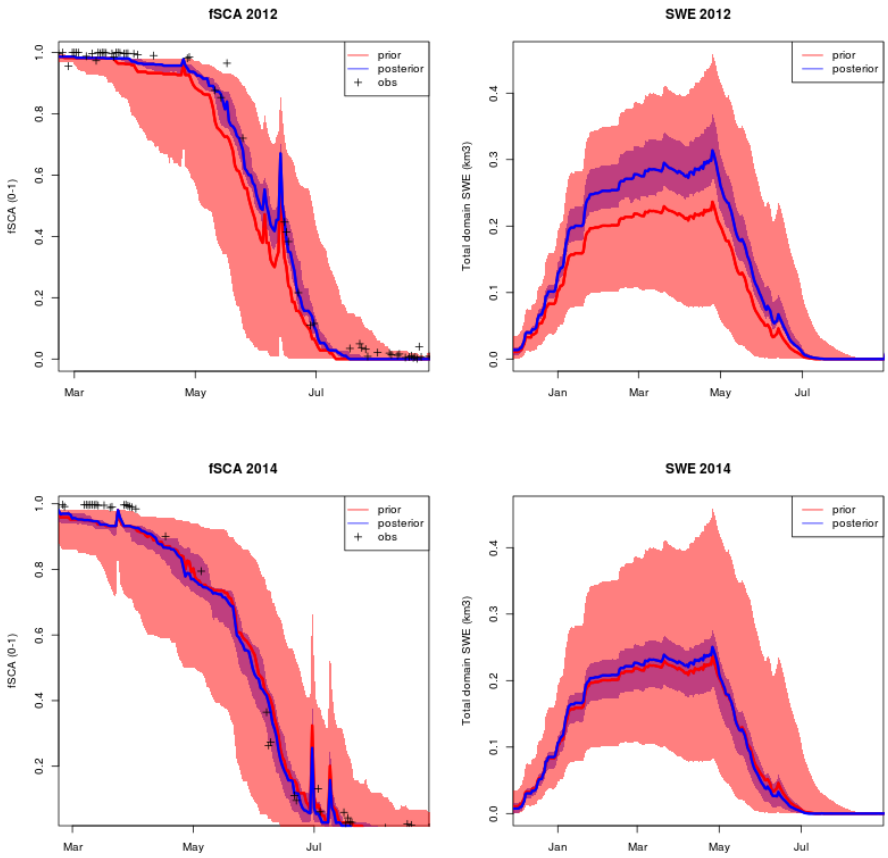


Figure 8. Assimilation of fSCA at grid level which targets bias in grid level forcing. Two contrasting seasons, WY2012 (top row) and WY2014 (bottom row) are shown. Vertical dashed lines give the assimilation window. Grid level biases in WY2012 are compensated for by DA. Grid level forcing was much more accurate in WY2014 and resulting effect of DA was negligible.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-37
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 10 May 2019

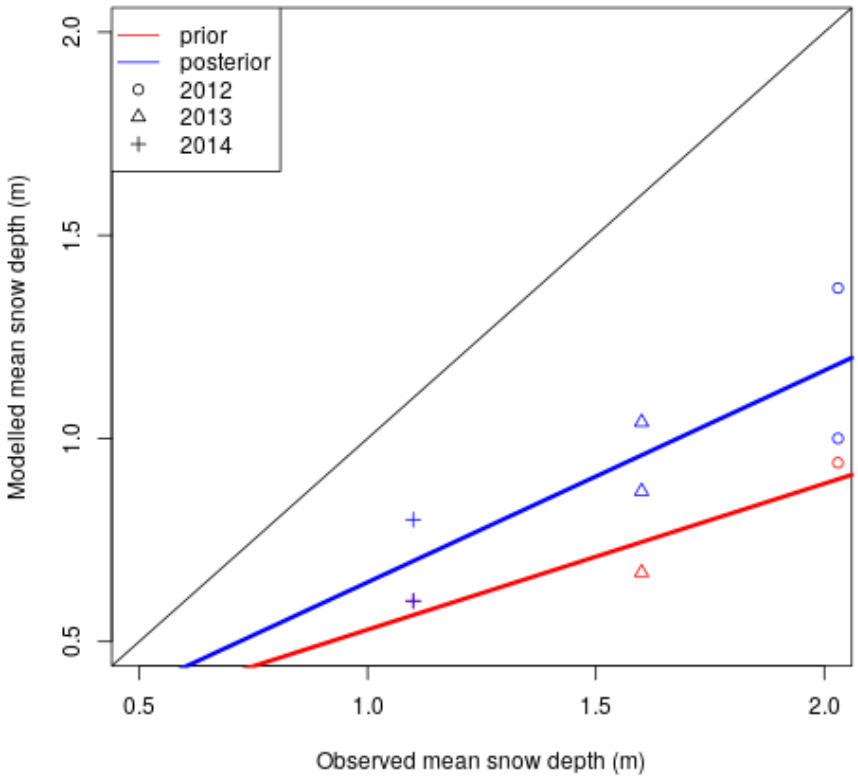


Figure 9. Interannual validity of weights generated by the DA scheme. Modelled versus observed mean snow depth averaged over the entire ADS zone on ADS acquisition dates WY2012, WY2013, WY2014 are shown. Posteriors (blue) are generated using weights of other two years and compared to the prior (red). For example posteriors of WY2012 are generated using weights of WY2013 and WY2014.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-37
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 10 May 2019

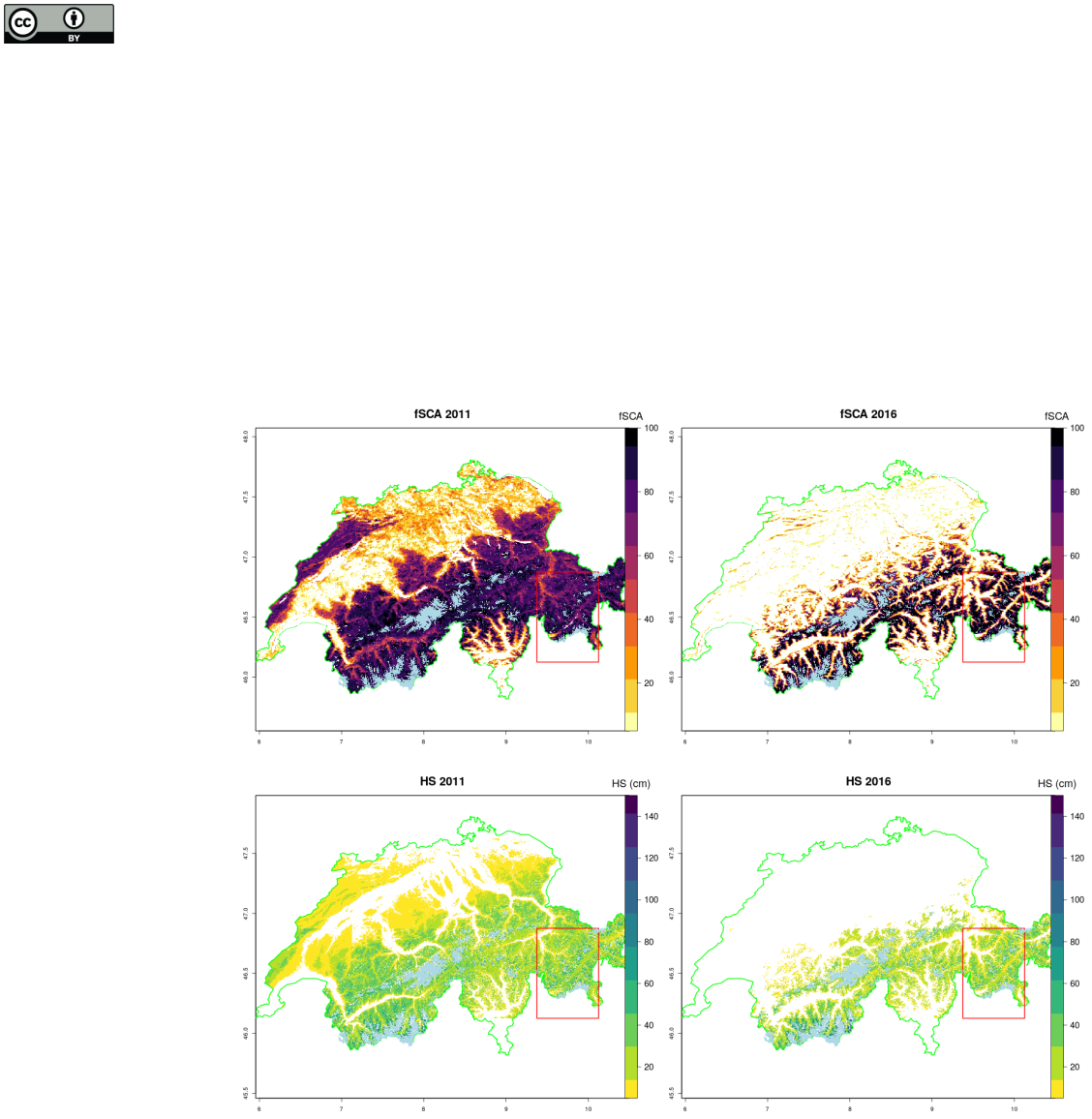


Figure 10. Mean December high resolution (30 m) large area HS simulations (open-loop) in two contrasting seasons (bottom) compared to observed spatial patterns of fSCA from mean December MODIS retrievals (top). Modelled HS compares well to spatial pattern of fSCA. December 2016 was an extremely dry start to the season with many ski resorts unable to open until late January. Both observed fSCA and modelled HS show snow reflect this fact with snow free zones deep into alpine valleys. Red box indicates domain of DA runs for theses seasons shown in Figure 11. Glacier mask given in light blue.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-37
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 10 May 2019

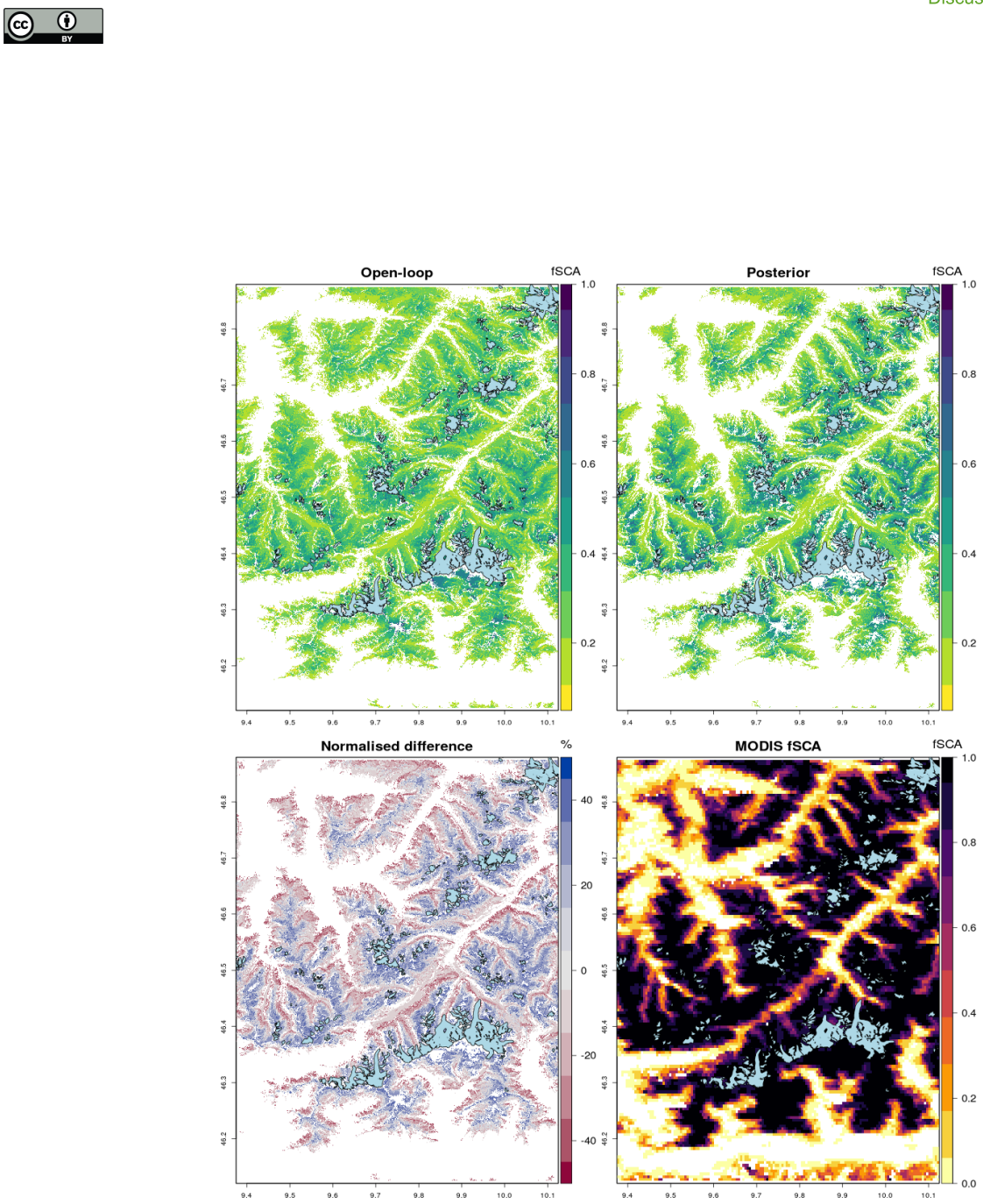


Figure 11. Mean HS December 2011 DA is compared to open-loop. Difference plot shows how DA has reduced low elevation snow height and increased high elevation snow height. Variability has been increased. Snow free valley bottoms show improved match to MODIS OBS. However both open–loop and posterior capture spatial patterns of snow cover reasonably well.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-37
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 10 May 2019

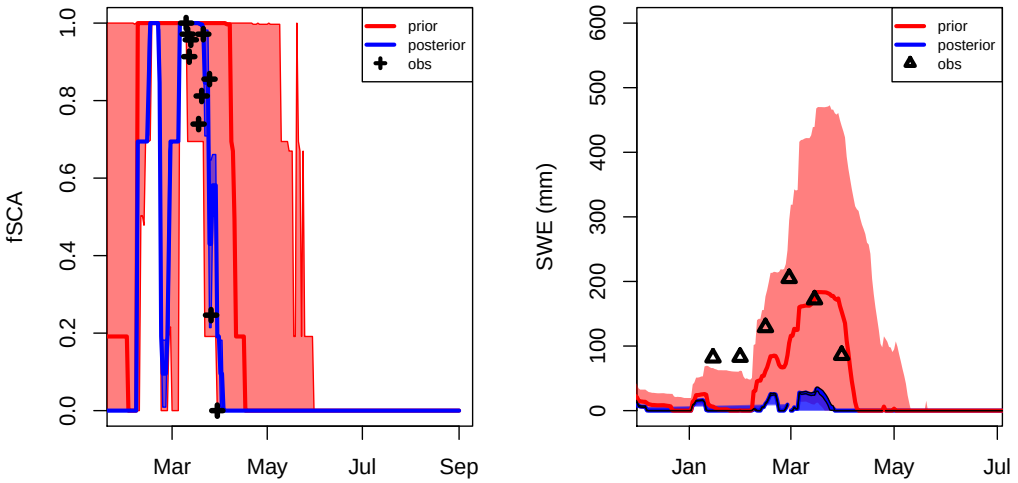


Figure 12. An example of poor performance due to non-representative fSCA retrievals. The posterior has been correctly pulled back to the observed depletion curve. However it is likely that the depletion curve does not well represent the validation station due to urban effects.

Part III

Appendix

Appendix A

Beer's law downscaling

The attenuation of a beam of direct shortwave radiation from the top of the atmosphere (TOA), denoted $S^{\downarrow}_{\infty}$, to some altitude z is to a good approximation given by Beer's law for a plane parallel atmosphere (e.g. Wallace and Hobbs, 2006)

$$S^{\downarrow}_{\text{dir}}(z) = S^{\downarrow}_{\infty}\text{exp}\left(-\frac{1}{\mu_0}\int_z^{\infty} k(z')\rho(z')dz'\right)\,, \tag{A.1}$$

where $\mu_0 = \cos(\theta_0)$ is the cosine of the solar zenith angle, k is the broadband attenuation coefficient and ρ is the air density. Note that as expected, when the sun sets below the horizon and $\theta_0 \to 90^{\circ}$ then $S^{\downarrow}_{\text{dir}}(z) \to 0$.

The hydrostatic equation reads

$$\frac{\partial \text{p}}{\partial z} \simeq -\rho g\,, \tag{A.2}$$

where g is the acceleration of gravity and p is air pressure. Inserting the change of variables $d\text{p}/g = -\rho\, dz$ in (A.1) and assuming g to be independent of height then

$$S^{\downarrow}_{\text{dir}}(z) = S^{\downarrow}_{\infty}\text{exp}\left(\frac{1}{g\mu_0}\int_{\text{p}(z)}^{\text{p}_{\infty}} k(\text{p}')d\text{p}'\right)\,, \tag{A.3}$$

Defining the average attenuation coefficient between some pressure level $\text{p}(z_1)$ and the TOA using that by definition $\text{p}_{\infty} = 0$

$$\overline{k}^{z_1} = -\frac{1}{\text{p}(z_1)}\int_{\text{p}(z_1)}^{\text{p}_{\infty}} k(\text{p}')d\text{p}'\,, \tag{A.4}$$

then (A.3) in terms of $\overline{k}^{z_1}$ reads

$$S^{\downarrow}_{\text{dir}}(z_1) = S^{\downarrow}_{\infty}\text{exp}\left(-\frac{\text{p}(z_1)\overline{k}^{z_1}}{g\mu_0}\right) , \tag{A.5}$$

solving for $\overline{k}^{z_1}$

$$\overline{k}^{z_1} = \frac{g\mu_0}{p(z_1)}\text{ln}\left(\frac{S^{\downarrow}_{\infty}}{S^{\downarrow}_{\text{dir}}(z_1)}\right) . \tag{A.6}$$

Now consider the case where there are two nearby altitudes z_1 and z_2 where $S^{\downarrow}_{\text{dir}}(z_1)$ and $S^{\downarrow}_{\infty}$ are known and we are interested in approximating $S^{\downarrow}_{\text{dir}}(z_2)$. Assuming $\delta z = z_2 - z_1$ is small, then we may approximate $\overline{k}^{z_2} \simeq \overline{k}^{z_1}$ such that

$$S^{\downarrow}_{\text{dir}}(z_2) \simeq S^{\downarrow}_{\infty}\text{exp}\left(-\frac{\text{p}(z_2)\overline{k}^{z_1}}{g\mu_0}\right) , \tag{A.7}$$

which, inserting for $\overline{k}^{z_1}$ and simplifying, reads

$$S^{\downarrow}_{\text{dir}}(z_2) \simeq S^{\downarrow}_{\infty}\left(\frac{S^{\downarrow}(z_1)}{S^{\downarrow}_{\infty}}\right)^{\frac{\text{p}(z_2)}{\text{p}(z_1)}} . \tag{A.8}$$

As verification, note that at the TOA when $\text{p}(z_2) = 0$ then $S^{\downarrow}(z_2) = S^{\downarrow}_{\infty}$ and when $\text{p}(z_2) = \text{p}(z_1)$ then $S^{\downarrow}(z_2) = S^{\downarrow}(z_1)$. This approximation may be quite useful when interpolating direct shortwave radiation between different pressure levels, so long as they are quite close to one another and the $\overline{k}^{z_2} \simeq \overline{k}^{z_1}$ assumption is not strongly violated.

Appendix B

Bayes theorem and rain in Bergen

In the spirit of keeping things as simple as possible, but hopefully not too simple, I will demonstrate Bayesian inference for a toy example. Consider the following: you observe that it is raining, how probable is it that you are in Bergen (a notoriously rainy Norwegian city)?

In this case the event space is: $\Omega = \{\mathcal{B}, \mathcal{R}, \mathcal{B}', \mathcal{R}'\}$ where $\mathcal{B}$ is the event that you are in Bergen, $\mathcal{B}'$ is the complementary event that you are elsewhere, $\mathcal{R}$ is the event that it is raining, and $\mathcal{R}'$ is the event that it is not raining. Remember that the axioms of probability amount to (Jazwinski, 1970): the probability of an event (denoted Pr) is non-negative, the probability of all events in an event space is equal to 1, and the joint probability of mutually exclusive events is zero.

The conditional probability that we seek is defined as (see Tarantola, 2005)

$$\Pr(\mathcal{B}|\mathcal{R}) = \Pr(\mathcal{B}, \mathcal{R})/\Pr(\mathcal{R})\,, \tag{B.1}$$

where | signifies 'given', $\Pr(\mathcal{B}, \mathcal{R})$ is the joint probability that we are in Bergen and it is raining, and $\Pr(\mathcal{R})$ is the probability that it is raining. Both the joint probability and the probability of rain are defined for the entire event space, whereas the conditional probability is defined for the subset of the event space where it is raining. Similarly, using that $\Pr(\mathcal{B}, \mathcal{R}) = \Pr(\mathcal{R}, \mathcal{B})$, the related conditional probability that it is raining given that we are in Bergen is

$$\Pr(\mathcal{R}|\mathcal{B}) = \Pr(\mathcal{B}, \mathcal{R})/\Pr(\mathcal{B})\,, \tag{B.2}$$

where $\Pr(\mathcal{B})$ is the probability that we are in Bergen. By comparing (B.1) and

(B.2) it is clear that the two conditional probabilities $\Pr(\mathcal{B}|\mathcal{R})$ and $\Pr(\mathcal{R}|\mathcal{B})$ are *not* equal unless $\Pr(\mathcal{R}) = \Pr(\mathcal{B})$ which is improbable. Falsely equating the two, known as the "error of the transposed conditional", is a major source of misunderstanding in science, particularly when testing for significance (Ambaum, 2010).

Solving for the joint probability in (B.1) and (B.2), equating the two expressions, and solving for the conditional probability we are after, we obtain Bayes theorem

$$\Pr(\mathcal{B}|\mathcal{R}) = \Pr(\mathcal{R}|\mathcal{B})\Pr(\mathcal{B})/\Pr(\mathcal{R})\,, \tag{B.3}$$

where, using the definition of conditional probability, the probability of rain is

$$\Pr(\mathcal{R}) = \Pr(\mathcal{R}|\mathcal{B})\Pr(\mathcal{B}) + \Pr(\mathcal{R}|\mathcal{B}')\Pr(\mathcal{B}')\,, \tag{B.4}$$

whereby

$$\Pr(\mathcal{B}|\mathcal{R}) = \frac{\Pr(\mathcal{R}|\mathcal{B})\Pr(\mathcal{B})}{\Pr(\mathcal{R}|\mathcal{B})\Pr(\mathcal{B}) + \Pr(\mathcal{R}|\mathcal{B}')\Pr(\mathcal{B}')}\,. \tag{B.5}$$

Which is a complete expression for the probability of being in Bergen given that it's raining, based entirely on the axioms of probability.

Before plugging in some numbers to make the example more concrete, let's constrain the problem slightly by saying that we must either be in Bergen or Oslo (the slightly less rainy Norwegian capital) and that either option is equally likely to be true, so that $\Pr(\mathcal{B}) = \Pr(\mathcal{B}') = 0.5$. In that case, roughly based on climatology, we also assume that $\Pr(\mathcal{R}|\mathcal{B}) = 0.8$ and $\Pr(\mathcal{R}|\mathcal{B}') = 0.4$. Therefore,

$$\Pr(\mathcal{B}|\mathcal{R}) = \frac{0.8 \times 0.5}{0.8 \times 0.5 + 0.4 \times 0.5} = 0.67\,. \tag{B.6}$$

This implies that $\Pr(\mathcal{B}'|\mathcal{R}) = 0.33$. So, based on the evidence that it is raining, it is $100 \times (0.67/0.33 - 1) \simeq 100\%$ more probable that you are in Bergen than Oslo. This result of course depends upon your prior. It shows us how we can update our knowledge in the face of new evidence (i.e. learn). Of course $\Pr(\mathcal{R}|\mathcal{B})$ for a specific day (as opposed to the climatology) is the more interesting quantity. As noted in Lyons and Wardle (2018), the 'objective' frequentist notion of probability does not permit assigning $\Pr(\mathcal{R}|\mathcal{B})$ for a specific day, to do so we must employ the 'subjective' Bayesian degree of belief interpretation.

Appendix C

Summary of remaining research

C.1 Co-authored publications

- Yılmaz, Y., **Aalstad, K.**, and Sen, O. L (2019): Multiple Remotely Sensed Lines of Evidence for a Depleting Seasonal Snowpack in the Near East, *Remote Sensing*, 11(5), 483, doi: 10.3390/rs11050483
  Summary: A marked depletion of the seasonal snowpack over the last decades contributed to the decline of terrestrial water storage in the snow-fed river basins of the Near East. Multiple satellite retrievals (MODIS, GRACE, AMSR-E, AMSR2) and reanalysis data (ERA5) were used in the analysis.

C.2 Presentations

Herein follows a list of talks and poster presentations held as 'first author' during the course of the PhD at various conferences, meetings, workshops, and seminars, in Norway and abroad in chronological order.

- **Aalstad, K.** and Westermann, S. and Bertino, L.: Applying Ensemble Data Assimilation to the Cryosphere. EmblA meeting, 10 December 2015, Fontainebleau, France. (**Talk**)
- **Aalstad, K.** and Westermann, S. and Bertino, L.: Developing an ensemble-based subgrid snow data assimilation framework. Monthly LATICE meeting, 6 January 2016, Oslo, Norway. (**Talk**)
- **Aalstad, K.** and Westermann, S. and Boike, J. and Bertino, L. and Aas, K. S.: An ensemble-based snow data assimilation framework with applications

to permafrost modeling*. ResClim All Staff meeting, 2 March 2016, Sotra, Norway. *Won best poster award and an associated travel grant. (**Poster**)

- **Aalstad, K.** and Westermann, S. and Boike, J. and Bertino, L. and Aas, K. S.: An ensemble-based snow data assimilation framework with applications to permafrost modeling. LATICE annual meeting, 9 March 2016, Sundvolden, Norway. (**Poster**)

- **Aalstad, K.** and Westermann, S. and Boike, J. and Bertino, L. and Aas, K. S.: An ensemble-based snow data assimilation framework with applications to permafrost modeling. 11th International Conference on Permafrost, 20 June 2016, Potsdam, Germany. (**Poster**)

- **Aalstad, K.** and Westermann, S. and Bertino, L.: An ensemble-based subgrid snow data assimilation framework. 5th Annual International Symposium on Data Assimilation, 18 July 2016, Reading, England. (**Poster**)

- **Aalstad, K.** and Westermann, S. and Bertino, L.: An ensemble-based subgrid snow data assimilation framework. Workshop on Data Assimilation in Terrestrial Systems, 20 September 2016, Bonn, Germany. (**Poster**)

- **Aalstad, K.** and Westermann, S. and Bertino, L.: An ensemble-based subgrid snow data assimilation framework applied to the southern Swiss alps. MedCLIVAR 2016 Conference, 28 September 2016, Athens, Greece. (**Talk**)

- **Aalstad, K.** and Westermann, S. and Laurent, Bertino: An ensemble-based subgrid snow data assimilation framework. American Geophysical Union Fall Meeting, 15 December 2016, San Francisco, California, USA. (**Talk**)

- **Aalstad, K.** and Westermann, S. and Schuler, T. V. and Boike, J. and Bertino, L.: Towards High-Resolution SWE Mapping in Permafrost Regions. The 2nd Asian Conference on Permafrost, 3 July 2017, Sapporo, Japan. (**Poster**)

- **Aalstad, K.** and Westermann, S. and Schuler, T. V. and Boike, J. and Bertino, L.: Towards High-Resolution SWE Mapping in Permafrost Regions. The 2nd Asian Conference on Permafrost, 4 July 2017, Sapporo, Japan. (**Talk**)

- **Aalstad, K.** and Westermann, S. and Schuler, T. V. and Boike, J. and Bertino, L.: Ensemble-based subgrid snow data assimilation. 4th Conference on Modelling Hydrology, Climate and Land Surface Processes, 13 September 2017, Lillehammer, Norway. (**Talk**)

- **Aalstad, K.** and Westermann, S. and Bertino, L. and Schuler, T. V. and Boike, J. and Karsten, L.: Towards high-resolution Bayesian snow reconstruction in permafrost regions. CESM Land Model and Biogeochemistry Working Group Meeting, 6 February 2018, Boulder, Colorado, USA. (**Talk**)

- **Aalstad, K.** and Westermann, S. and Karsten, L. and Fiddes, J. and Bertino, L.: Snow history matching in mountainous terrain. 8th GEWEX Open Science Conference: Exteremes and Water on the edge, 7 May 2018, Canmore, Alberta, Canada. (**Poster**)

- **Aalstad, K.** and Westermann, S. and Karsten, L. and Gutmann, E. and McCreight, J. and Fiddes, J. and Bertino, L.: Ensemble-based reanalysis of the seasonal montane snowpack: Lessons from the ASO. 3rd Annual NASA Airborne Snow Observatory Workshop, 11 of September 2018, Mammoth Lakes, California, USA. (**Talk**)

- **Aalstad, K.** and Westermann, S. and Bertino, L.: Ensemble-based reanalysis of the seasonal snowpack using multispectral satellite imagery. NCAR-RAL-HAP seminar, 20 September 2018, Boulder, Colorado, USA. (**Talk**)

- **Aalstad, K.** and Westermann, S. and Bertino, L.: Data Assimilation: Dicing with Uncertainty. GeoHyd lunch seminar, 8 February 2019, Oslo, Norway. (**Talk**)

- **Aalstad, K.** and Westermann, S. and Bertino, L.: Ensemble-based reconstruction of remote seasonal snowpacks. CHESS annual meeting, 15 March 2019, Oslo, Norway. (**Talk**)

- **Aalstad, K.** and Westermann, S. and Fiddes, J. and Bertino, L.: Ensemble-based reanalysis of remote seasonal snowpacks. LATICE annual meeting, 20 March 2019, Sundvolden, Norway. (**Poster**)

- **Aalstad, K.** and Westermann, S. and Fiddes, J. and Bertino, L.: Retrieving the depletion of snow-covered area from multiple optical satellite sensors with applications for SWE reanalysis. ESA Living Planet Symposium, 16 May 2019, Milan, Italy. (**Talk**)

C.3 Courses

C.3.1 Curriculum

Herein follows a list of the academic courses that were completed to meet the required formal theoretical background for the PhD degree at the Faculty of Mathematics and Natural Sciences at the University of Oslo. The theoretical background amounts to course work that corresponds to 30 credits in the European Credit Transfer System (ECTS) at the PhD level. Courses were attended both at the University of Oslo, elsewhere in Norway, and abroad. The attended courses are listed in chronological order below.

- Cryospheric Modelling (GEO9440). Department of Geosciences, University of Oslo. Spring 2016. 5 credits.
- Atmosphere-Vegetation-Soil Interaction: From Diurnal to Climate Scales. Research school on changing climates in the coupled earth system (CHESS), University of Bergen, Norway. Spring 2016, 1 credit.
- Ocean/Atmosphere Time Series Analysis: Theory and Practice I. Research school on changing climates in the coupled earth system (CHESS), University of Oslo, Norway. Spring 2016. 3 credits.
- Uncertainty in Environmental Modelling. Department of Earth Sciences, Uppsala University, Sweden. Spring 2016. 5 credits.
- Science, Ethics and Society (MNSES9100). Faculty of Mathematics and Natural Sciences, University of Oslo, Norway. Fall 2016. 5 credits.
- Ocean/Atmosphere Time Series Analysis: Theory and Practice II. Research school on changing climates in the coupled earth system (CHESS), University of Oslo, Norway. Spring 2017. 2 credits.
- Advanced Statistics Training for Climate Resarch. Research school on changing climates in the coupled earth system (CHESS), University of Bergen, Norway. Fall 2017. 2 credits.
- Land Surface Modelling Course. Research school on changing climates in the coupled earth system (CHESS), University of Bergen, Norway. Fall 2017. 2 credits.

- Field course in glacial and periglacial geomorphology/geocryology (GEO9411). Department of Geosciences, University of Oslo, Norway. Fall 2017. 5 credits.

C.3.2 Other

In addition to the courses attended to meet the formal requirements of the PhD degree, some other courses, workshops, and research schools were attended. These are listed in chronological order below.

- School on Data Assimilation and Data Analysis Methods, Evaluation of Ocean Synthesis COST Action ES1402, Lecce, Italy. April 2016.
- Geostatistics course, Nansen Environmental and Remote Sensing Center, Bergen, Norway. January 2017.
- Towards Regional Information to Improve Our Understanding on Weather, Water, and Climate Extreme Events, Joint YESS-YHS Early Career Researcher (ECR) Workshop 2018, Canmore, Ablerta, Canada. May 2018.

Part IV

References

Bibliography

Aalstad, K., Westermann, S., and Bertino, L. (2019a). Evaluating satellite retrieved snow-covered area at a high-Arctic site using terrestrial photography. Submitted to Remote Sensing of Environment.

Aalstad, K., Westermann, S., Fiddes, J., and Bertino, L. (2019b). Ensemble-based snow reanalysis using dense time stacks of multisensor multispectral satellite imagery. Manuscript to be submitted.

Aalstad, K., Westermann, S., Schuler, T. V., Boike, J., and Bertino, L. (2018). Ensemble-based assimilation of fractional snow-covered area satellite retrievals to estimate the snow distribution at Arctic sites. *The Cryosphere*. doi: 10.5194/tc-12-247-2018.

Aas, K., Gisnås, K., Westermann, S., and Berntsen, T. K. (2017). A Tiling Approach to Represent Subgrid Snow Variability in Coupled Land Surface–Atmosphere Models. *Journal of Hydrometeorology*. doi: 10.1175/JHM-D-16-0026.1.

Albers, D. J., Levine, M. E., Stuart, A., Mamykina, L., Gluckman, B., and Hripcsak, G. (2018). Mechanistic machine learning: how data assimilation leverages physiologic knowledge using Bayesian inference to forecast the future, infer the present, and phenotype. *Journal of the American Medical Informatics Association*. doi: 10.1093/jamia/ocy106.

Ambaum, M. H. P. (2010). Significance Tests in Climate Science. *Journal of Climate*. doi:10.1175/2010JCLI3746.1.

Anderson, E. A. (1968). Development and testing of snow pack energy balance equations. *Water Resources Research*. doi: 10.1029/WR004i001p00019.

Andreadis, K. M. and Lettenmaier, D. P. (2006). Assimilating remotely sensed snow observations into a macroscale hydrology model. *Advances in Water Resources*. doi: 10.1016/j.advwatres.2005.08.004.

Andreas, E. L., Persson, P. O. G., Jordan, R. E., Horst, T. W., Guest, P. S., Grachev, A. A., and Fairall, C. W. (2010). Parameterizing Turbulent Exchange over Sea Ice in Winter. *Journal of Hydrometeorology*. doi: 10.1175/2009JHM1102.1.

Baghdadi, N., Gauthier, Y., and Bernier, M. (1996). Capability of multitemporal ERS-1 SAR data for wet-snow mapping. *Remote Sensing of Environment*. doi: 10.1016/S0034-4257(96)00180-0.

Bair, E. H., Rittger, K., Davis, R. E., Painter, T. H., and Dozier, J. (2016). Validating reconstruction of snow water equivalent in California's Sierra Nevada using measurements from the NASA Airborne Snow Observatory. *Water Resources Research*. doi: 10.1002/2016WR018704.

Bair, E. N., Calfa, A. A., Rittger, K., and Dozier, J. (2018). Using machine learning for real-time estimates of snow water equivalent in the watersheds of Afghanistan. *The Cryosphere*. doi: 10.5194/tc-12-1579-2018.

Baldo, E. and Margulis, S. (2018). Assessment of a multiresolution snow reanalysis framework: a multidecadal reanalysis case over the upper Yampa River basin, Colorado. *Hydrology and Earth System Sciences*. doi: 10.5194/hess-22-3575-201.

Baldridge, A. M., Hook, S. J., Grove, C. I., and Rivera, G. (2009). The ASTER spectral library version 2.0. *Remote Sensing of Environment*, 113:711–715. doi: 10.1016/j.rse.2008.11.007.

Bales, R. C., Molotch, N. P., Painter, T. H., Dettinger, M. D., Rice, R., and Dozier, J. (2006). Mountain hydrology of the western united states. *Water Resources Research*. doi: 10.1029/2005WR004387.

Barnett, T. P., Adam, J. C., and Lettenmaier, D. P. (2005). Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature*. doi: 10.1038/nature04141.

Bartelt, P. and Lehning, M. (2002). A physical SNOWPACK model for the Swiss avalanche warning: Part I: numerical model. *Cold Regions Science and Technology*. doi: 10.1016/S0165-232X(02)00074-5.

Bauer, P., Thorpe, A., and Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*. doi: 10.1038/nature14956.

Benestad, R. E. (2017). A mental picture of the greenhouse effect. *Theoretical and Applied Climatology*. doi: 10.1007/s00704-016-1732-y.

Bertino, L., Evensen, G., and Wackernagel, H. (2003). Sequential Data Assimilation Techniques in Oceanography. *International Statistical Review*. doi: 10.1111/j.1751-5823.2003.tb00194.x.

Beven, K. (2001). How far can we go in distributed hydrological modelling? *Hydrology and Earth System Sciences*. doi: 10.5194/hess-5-1-2001.

Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*. doi: 10.1016/j.jhydrol.2005.07.007.

Beven, K. (2016). Facets of uncertainty: epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication. *Hydrological Sciences Journal*. doi: 10.1080/02626667.2015.1031761.

Beven, K. and Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*. doi: 10.1002/hyp.3360060305.

Beven, K. and Binley, A. (2014). GLUE: 20 years on. *Hydrological Processes*. doi: 10.1002/hyp.10082.

Blöschl, G. (1999). Scaling issues in snow hydrology. *Hydrological Processes*. doi: 10.1002/(SICI)1099-1085(199910)13:14/15<2149::AID-HYP847>3.0.CO;2-8.

Boike, J. et al. (2018). A 20-year record (1998-2017) of permafrost, active layer and meteorological conditions at a high Arctic permafrost research site (Bayelva, Spitsbergen). *Earth System Science Data*, 10:355–390. doi: 10.5194/essd-10-355-2018.

Bojinski, S., Verstraete, M., Peterson, T. C., Richter, C., Simmons, A., and Zemp, M. (2014). The Concept of Essential Climate Variables in Support of Climate Research, Applications, and Policy. *Bulletin of the American Meteorological Society*. doi: 10.1175/BAMS-D-13-00047.1.

Bormann, K., Brown, R. D., Derksen, C., and Painter, T. H. (2018). Estimating snow-cover trends from space. *Nature Climate Change*. doi: 10.1038/s41558-018-0318-3.

Box, G. E. P. and Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society: Series B*. doi: 10.1111/j.2517-6161.1964.tb00553.x.

Broecker, W. S. (1975). Climatic Change: Are We on the Brink of a Pronounced Global Warming? *Science*. doi: 10.1126/science.189.4201.460.

Bühler, S., Adams, M. S., Bösch, R., and Stoffel, A. (2016). Mapping snow depth in alpine terrain with unmanned aerial systems (UASs): potential and limitations. *The Cryosphere*. doi: 10.5194/tc-10-1075-2016.

Burakowski, E. and Magnusson, M. (2012). Climate Impacts on the Winter Tourism Economy in the United States. Prepared for Protect Our Winters (POW) and Natural Resources Defense Council (NRDC).

Burgers, G., van Leeuwen, P. J., and Evensen, G. (1998). Analysis Scheme in the Ensemble Kalman Filter. *Monthly Weather Review*. doi: 10.1175/1520-0493(1998)126%3C1719:ASITEK%3E2.0.CO;2.

Carrassi, A., Bocquet, M., Bertino, L., and Evensen, G. (2018). Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *WIREs Climate Change*. doi: 10.1002/wcc.535.

Carrassi, A., Bocquet, M., Hannart, A., and Ghil, M. (2016). Estimating model evidence using data assimilation. *Quarterly Journal of the Royal Meteorological Society*. doi: 10.1002/qj.2972.

CDFA (2019). California Department of Food and Agriculture, California Agricultural Production Statistics. Available at: https://www.cdfa.ca.gov/statistics/.

Chadburn, S. E., Burke, E. J., Cox, P. M., Friedlingstein, P., Hugelius, G., and Westermann, S. (2017). An observation-based constraint on permafrost loss as a function of global warming. *Nature Climate Change*. doi: 10.1038/nclimate3262.

Chang, A. T. C., Foster, J. L., and Hall, D. K. (1987). Nimbus-7 SMMR Derived Global Snow Cover Parameters. *Annals of Glaciology*. doi: 10.3189/S0260305500200736.

Chapin, F. S. et al. (2005). Role of Land-Surface Changes in Arctic Summer Warming. *Science*. doi: 10.1126/science.1117368.

Charrois, L., Cosme, E., Dumont, M., Lafaysse, M., Morin, S., Libois, Q., and Picard, G. (2016). On the assimilation of optical reflectances and snow depth observations into a detailed snowpack model. *The Cryosphere*. doi: 10.5194/tc-10-1021-2016.

CIESIN (2017). Center for International Earth Science Information Network - CIESIN - Columbia University. 2017. Gridded Population of the World, Version 4 (GPWv4): Population Count Adjusted to Match 2015 Revision of UN WPP Country Totals, Revision 10. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). doi: 10.7927/H4JQ0XZW, last accessed February 2019.

Clark, M. P. et al. (2011). Representing spatial variability of snow water equivalent in hydrologic and land-surface models: A review. *Water Resources Research*. doi: 10.1029/2011WR010745.

Clark, M. P. et al. (2015). A unified approach for process-based hydrologic modeling: 1. Modeling concept. *Water Resources Research*. doi: 10.1002/2015WR017200.

Clark, M. P., Slater, A. G., Barrett, A. P., Hay, L. E., McCabe, G. J., Rajagopalan, B., and Leavesley, G. H. (2006). Assimilation of snow covered area information into hydrologic and land-surface models. *Advances in Water Resources*. doi: 10.1016/j.advwatres.2005.10.001.

Clark, M. P. and Vrugt, J. A. (2006). Unraveling uncertainties in hydrological model calibration: Addressing the problem of compensatory parameters. *Geophysical Research Letters*. doi: 10.1029/2005GL025604.

Cline, D. et al. (2009). NASA Cold Land Processes Experiment (CLPX 2002/03): Airborne Remote Sensing. *Journal of Hydrometeorology*. doi: 10.1175/2008JHM883.1.

Cline, D. W., Bales, R. C., and Dozier, J. (1998). Estimating the spatial distribution of snow in mountain basins using remote sensing and energy balance modeling. *Water Resources Research*. doi: 10.1029/97WR03755.

Cortés, G., Girotto, M., and Margulis, S. (2016). Snow process estimation over the extratropical Andes using a data assimilation framework integrating MERRA data and Landsat imagery. *Water Resources Research*. doi: 10.1002/2015WR018376.

Cortés, G., Girotto, M., and Margulis, S. A. (2014). Analysis of sub-pixel snow and ice extent over the extratropical Andes using spectral unmixing and historical Landsat imagery. *Remote Sensing of Environment*. doi: 10.1016/j.rse.2013.10.023.

Cortés, G. and Margulis, S. (2017). Impacts of El Niño and La Niña on interannual snow accumulation in the Andes: Results from a high-resolution 31 year reanalysis. *Geophysical Research Letters*. doi: 10.1002/2017GL073826.

Cosme, E., Verron, J., Brasseur, P., Blum, J., and Auroux, D. (2012). Smoothing Problems in a Bayesian Framework and Their Linear Gaussian Solutions. *Monthly Weather Review*. doi: 10.1175/MWR-D-10-05025.1.

De Lannoy, G. J. M., Reichle, R. H., Arsenault, K. R., Houser, P. R., Kumar, S., Verhoest, N. E. C., and Pauwels, V. R. N. (2012). Multiscale assimilation of Advanced Microwave Scanning Radiometer-EOS snow water equivalent and Moderate Resolution Imaging Spectroradiometer snow cover fraction observations in northern Colorado. *Water Resources Research*. doi: 10.1029/2011WR010588.

De Lannoy, G. J. M., Reichle, R. H., Houser, P. R., Arsenault, K. R., Verhoest, N. E. C., and Pauwels, V. R. N. (2010). Satellite-scale snow water equivalent assimilation into a high resolution land surface model. *Journal of Hydrometeorology*. doi: 10.1175/2009JHM1192.1.

Dee, D. P. et al. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*. doi: 10.1002/qj.828.

Diffenbaugh, N. S., Singh, D., Mankin, J. S., Horton, D. E., Swain, D. L., Touma, D., Charland, A., Liu, Y., Haugen, M., Tsiang, M., and Rajaratnam, B. (2017). Quantifying the influence of global warming on unprecedented extreme climate events. *Proceedings of the National Academy of Science*. doi: 10.1073/pnas.1618082114.

Dozier, J. (1989). Spectral Signature of Alpine Snow Cover from the Landsat Thematic Mapper. *Remote Sensing of Environment*, 28:9–22. doi: 10.1016/0034-4257(89)90101-6.

Dozier, J. (2011). Mountain hydrology, snow color, and the fourth paradigm. *EOS, Transactions, American Geophysical Union*. doi: 10.1029/2011EO430001.

Dozier, J., Bair, E. H., and Davis, R. E. (2016). Estimating the spatial distribution of snow water equivalent in the world's mountains. *WIREs Water*. doi: 10.1002/wat2.1140.

Dozier, J., Bruno, J., and Downey, P. (1981). A faster solution to the horizon problem. *Computers & Geosciences*. doi: 10.1016/0098-3004(81)90026-1.

Dozier, J. and Frew, J. (1990). Rapid Calculation of Terrain Parameters For Radiation Modeling From Digital Elevation Data. *IEEE Transactions on Geoscience and Remote Sensing*. doi: 10.1109/36.58986.

Dozier, J., Green, R. O., Nolin, A. W., and Painter, T. H. (2009). Interpretation of snow properties from imaging spectrometry. *Remote Sensing of Environment*. doi: 10.1016/j.rse.2007.07.029.

Dozier, J. and Warren, S. G. (1982). Effect of Viewing Angle on the Infrared Brightness Temperature of Snow. *Water Resources Research*. doi: 10.1029/WR018i005p01424.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., and Bargellini, P. (2012). Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*. doi: 10.1016/j.rse.2011.11.026.

Duan, L., Cao, L., and Caldeira, K. (2019). Estimating Contributions of Sea Ice and Land Snow to Climate Feedback. *Journal of Geophysical Research: Atmospheres*. doi: 10.1029/2018JD029093.

Dunne, S. and Entekhabi, D. (2005). An ensemble-based reanalysis approach to land data assimilation. *Water Resources Research*. doi: 10.1029/2004WR003449.

Durand, M., Molotch, N. P., and Margulis, S. A. (2008). A Bayesian approach to snow water equivalent reconstruction. *Journal of Geophsyical Research: Atmospheres*. doi: 10.1029/2008JD009894.

Dutra, E., Balsamo, B., Viterbo, P., Miranda, P., Beljaars, A., Schär, C., and Elder, K. (2010). An Improved Snow Scheme for the ECMWF Land Surface Model: Description and Offline Validation. *Journal of Hydrometeorology*. doi: 10.1175/2010JHM1249.1.

Efron, B. (1986). Why Isn't Everyone a Bayesian? *The American Statistician*. doi: 10.2307/2683105.

Emerick, A. A. and Reynolds, A. C. (2012). History matching time-lapse seismic data using the ensemble Kalman filter with multiple data assimilations. *Computational Geosciences*. doi: 10.1007/s10596-012-9275-5.

Emerick, A. E. (2018). Deterministic ensemble smoother with multiple data assimilation as an alternative for history-matching seismic data. *Computational Geosciences*. doi: 10.1007/s10596-018-9745-5.

Emerick, A. E. and Reynolds, A. C. (2013). Ensemble smoother with multiple data assimilation. *Computers & Geosciences*. doi: 10.1016/j.cageo.2012.03.011.

Endrizzi, S., Gruber, S., Dall'Amico, M., and Rigon, R. (2014). GEOtop 2.0: simulating the combined energy and water balance at and below the land surface accounting for soil freezing, snow cover and terrain effects. *Geoscientific Model Development*. doi: 10.5194/gmd-7-2831-2014.

Essery, R. (2015). A factorial snowpack model (FSM 1.0). *Geoscientific Model Development*. doi: 10.5194/gmd-8-3867-2015.

Euskirchen, E. S., Goodstein, E. S., and Huntington, H. P. (2013). An estimated cost of lost climate regulation services caused by thawing of the Arctic cryosphere. *Ecological Applications*. doi: 10.1890/11-0858.1.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*. doi: 10.1029/94JC00572.

Evensen, G. (2009). *Data Assimilation: The Ensemble Kalman Filter*. Springer. doi: 10.1007/978-3-642-03711-5.

Evensen, G. (2018). Analysis of iterative ensemble smoothers for solving inverse problems. *Computational Geosciecnes*. doi: 10.1007/s10596-018-9731-y.

Evensen, G. (2019). Accounting for model errors in iterative ensemble smoothers. *Computational Geosciences*. doi: 10.1007/s10596-019-9819-z.

Fan, Y. et al. (2019). Hillslope Hydrology in Global Change Research and Earth System Modeling. *Water Resources Research*. doi: 10.1029/2018WR023903.

Farr, T. G. et al. (2007). The shuttle radar topography mission. *Rev. Geophys.* doi: 10.1029/2005rg000183.

Fer, I., Kelly, R., Moorcroft, P. R., Richardson, A. D., Cowdery, E. M., and Dietze, M. C. (2018). Linking big models to big data: efficient ecosystem model calibration through Bayesian model emulation. *Biogeosciences*.

Fiddes, J., Aalstad, K., and Westermann, S. (2019). Hyper-resolution ensemble-based snow reanalysis in mountain regions using clustering. *Hydrology and Earth System Sciences Discussions*. doi: 10.5194/hess-2019-37.

Fiddes, J., Endrizzi, S., and Gruber, S. (2015). Large-area land surface simulations in heterogeneous terrain driven by global data sets: application to mountain permafrost. *The Cryosphere*. doi: 10.5194/tc-9-411-2015.

Fiddes, J. and Gruber, S. (2012). TopoSUB: a tool for efficient large area numerical modeling in complex topography at sub-grid scales. *Geoscientific Model Development*. doi: 10.5194/gmd-5-1245-2012.

Fiddes, J. and Gruber, S. (2014). Toposcale v.1.0: downscaling gridded climate data in complex terrain. *Geoscientific Model Development*. doi: 10.5194/gmd-7-387-2014.

Filhol, S. and Sturm, M. (2019). The smoothing of landscapes during snowfall with no wind. *Journal of Glaciology*. doi: 10.1017/jog.2018.104.

Flanner, M. G., Zender, C. S., Randerson, J. T., and Rasch, P. J. (2007). Present-day climate forcing and response from black carbon in snow. *Journal of Geophysical Research: Atmospheres*. doi: 10.1029/2006JD008003.

Fleming, J. R. (1999). Joseph Fourier, the 'greenhouse effect', and the quest for a universal theory of terrestrial temperatures. *Endeavour*. doi: 10.1016/S0160-9327(99)01210-7.

Foken, T. (2008). Micrometeorology. doi: 10.1007/978-3-540-74666-9.

Foster, J. L., Sun, C., Walker, J. P., Kelly, R., Chang, A., Dong, J., and Powell, H. (2005). Quantifying the uncertainty in passive microwave snow water equivalent observations. *Remote Sensing of Environment*. doi: 10.1016/j.rse.2004.09.012.

Freeze, A. R. and Harlan, R. L. (1969). Blueprint for a physically-based, digitally-simulated hydrologic response model. *Journal of Hydrology*. doi: 10.1016/0022-1694(69)90020-1.

Garren, D. C. and Marks, D. (2005). Spatially distributed energy balance snowmelt modelling in a mountainous river basin: estimation of meteorological inputs and verification of model results. *Journal of Hydrology*. doi: 10.1016/j.jhydrol.2005.03.026.

Gascoin, S., Grizonnet, M., Bouchet, M., Salgues, G., and Hagolle, O. (2019). Theia Snow collection: high-resolution operational snow cover maps from Sentinel-2 and Landsat-8 data. *Earth System Science Data*. doi: 10.5194/essd-11-493-2019.

Gascoin, S., Hagolle, O., Huc, M., Jarlan, L., Dejoux, J.-F., Szczypta, C., Marti, R., and Sánchez, R. (2015). A snow cover climatology for the Pyrenees from MODIS snow products. *Hydrology and Earth System Sciences*. doi: 10.5194/hess-19-2337-2015.

Girotto, M., Cortés, G., Margulis, S. A., and Durand, M. (2014a). Examining spatial and temporal variability in snow water equivalent using a 27 year reanalysis: Kern River watershed, Sierra Nevada. *Water Resources Research*, 50:6713–6734. doi: 10.1002/2014WR015346.

Girotto, M., Margulis, S. A., and Durand, M. (2014b). Probabilistic SWE reanalysis as a generalization of deterministic SWE reconstruction techniques. *Hydrological Processes*. doi: 10.1002/hyp.988.

Gisnås, K., Westermann, S., Schuler, T. V., Litherland, T., Isaksen, K., Boike, J., and Etzerlmüller, B. (2014). A statistical approach to represent small-scale

variability of permafrost temperatures due to snow cover. *The Cryosphere*. doi: 10.5194/tc-8-2063-2014.

GISSTEMP Team (2019). GISS Surface Temperature Analysis (GISTEMP). NASA Goddard Institute for Space Studies. Dataset accessed 2019-02-22 at https://data.giss.nasa.gov/gistemp/.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press.

Gordon, N. J., Salmond, D. J., and Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F - Radar and Signal Processing*. doi: 10.1049/ip-f-2.1993.0015.

Gubler, S., Gruber, S., and Purves, R. S. (2012). Uncertainties of parameterized surface downward clear-sky shortwave and all-sky longwave radiation. *Atmospheric Chemistry and Physics*. doi: 10.5194/acp-12-5077-2012.

Gutmann, E., Barstad, I., Clark, M., Arnold, J., and Rasmussen, R. (2016). The Intermediate Complexity Atmospheric Research Model. *Journal of Hydrometeorology*. doi: 10.1175/JHM-D-15-0155.1.

Halder, S. and Dirmeyer, P. A. (2017). Relation of Eurasian Snow Cover and Indian Summer Monsoon Rainfall: Importance of the Delayed Hydrological Effect. *Journal of Climate*. doi: 10.1175/JCLI-D-16-0033.1.

Hall, A. and Qu, X. (2006). Using the current seasonal cycle to constrain snow albedo feedback in future climate change. *Geophysical Research Letters*. doi: 10.1029/2005GL025127.

Hall, D. K. (1988). Assessment of Polar Climate Change Using Satellite Technology. *Reviews of Geophysics*, 26:26–39. doi: 10.1029/RG026i001p00026.

Hall, D. K., Riggs, G. A., Salmonson, V. V., DiGirolamo, N. E., and Bayr, K. J. (2002). MODIS snow-cover products. *Remote Sensing of Environment*. doi: 10.1016/S0034-4257(02)00095-0.

Härer, S., Bernhardt, M., Siebers, M., and Schulz, K. (2018). On the need for a time- and location-dependent estimation of the NDSI threshold value for reducing existing uncertainties in snow cover maps at different scales. *The Cryosphere*. doi: 10.5194/tc-12-1629-2018.

Havens, S., Marks, D., Kormos, P., and Hedrick, A. (2017). Spatial Modeling for Resources Framework (SMRF): A modular framework for developing spatial forcing data for snow modeling in mountain basins. *Computers & Geosciences*. doi: 10.1016/j.cageo.2017.08.016.

He, C., Flanner, M. G., Chen, F., Barlage, M., Liou, K.-N., Kang, S., Ming, J., and Qian, Y. (2018). Black carbon-induced snow albedo reduction over the Tibetan Plateau: uncertainties from snow grain shape and aerosol-snow mixing state based on an updated SNICAR model. *Atmospheric Chemistry and Physics*, 18. doi: 10.5194/acp-18-11507-2018.

Heinz, D. C. and Chang, C. I. (2001). Fully Constrained Least Squares Linear Spectral Mixture Analysis for Material Quantification in Hyperspectral Imagery. *IEEE Transactions on Geoscience and Remote Sensing*. doi: 10.1109/36.911111.

Helmert, J. et al. (2018). Review of Snow Data Assimilation Methods for Hydrological, Land Surface, Meteorological and Climate Models: Results from a COST HarmoSnow Survey. *Geosciences*. doi: 10.3390/geosciences8120489.

Henderson, G. R., Peings, Y., Furtado, J. C., and Kushner, P. J. (2018). Snow–atmosphere coupling in the Northern Hemisphere. *Nature Climate Change*. doi: 10.1038/s41558-018-0295-6.

Hock, R., Hutchings, J. K., and Lehning, M. (2017). Grand Challenges in Cryospheric Sciences: Toward Better Predictability of Glaciers, Snow and Sea Ice. *Frontiers in Earth Science*. doi: 10.3389/feart.2017.00064.

Huang, C., Newman, A. J., Clark, M. P., Wood, A. W., and Zheng, X. (2017). Evaluation of snow data assimilation using the ensemble Kalman filter for seasonal streamflow prediction in the western United States. *Hydrology and Earth System Sciences*. doi: 10.5194/hess-21-635-2017.

Hugelius, G., Strauss, J., Zubrzycki, S., Harden, J. W., Schuur, E. A. G., C.-L., P., Schirrmeister, L., Grosse, G., Michaelson, G. J., Koven, C., O'Donnel, J. A., Elberling, B., Mishra, U., Camill, P., Yu, Z., Palmtag, J., and Kuhry, P. (2014). Estimated stocks of circumpolar permafrost carbon with quantified uncertainty ranges and identified data gaps. *Biogeosciences*. doi: 10.5194/bg-11-6573-2014.

Hunning, L. and Margulis, S. A. (2017). Investigating the Variability of High-Elevation Seasonal Orographic Snowfall Enhancement and Its Drivers across Sierra Nevada, California. *Journal of Hydrometeorology*. doi: 10.1175/JHM-D-16-0254.1.

Hüsler, F., Jonas, T., Riffler, M., Musial, J. P., and Wunderle, S. (2014). A satellite-based snow cover climatology (1985-2011) for the European Alps derived from AVHRR data. *The Cryosphere*. doi: 10.5194/tc-8-73-2014.

Ide, K., Ghil, M., and Lorenc, A. (1999). Unified Notation for Data Assimilation: Operational, Sequential and Variational. *Journal of the Meteorological Society of Japan*. doi: 10.2151/jmsj1965.75.1B_181.

IPCC (2013). Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. [Stocker et al. (eds)] Cambridge University Press, Cambridge, United Kongdom and New York, NY, USA, 1535 pp.

IPCC (2018). Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. [Masson-Delmotte et al. (eds)] World Meteorological Organization, Geneva, Switzerland.

Jazwinski, A. H. (1970). *Stochastic Processes and Filtering Theory*. Academic Press.

Jobbággy, E. G. and Jackson, R. B. (2000). The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecological Applications*. doi: 10.1890/1051-0761(2000)010[0423:TVDOSO]2.0.CO;2.

Kääb, A., Altena, B., and Mascaro, J. (2019). River ice and water velocities using the Planet optical cubesat constellation. *Hydrology and Earth System Sciences Discussions*. doi: 10.5194/hess-2019-62.

Kahl, A., Winstral, A., Marks, D., and Dozier, J. (2013). Using Satellite Imagery and the Distributed Isnobal Energy Balance Model to Derive SWE Heterogeneity in Mountainous Basins. In Pomeroy, J. W., Whitfield, P. H., and Spence, C., editors, *Putting Prediction in Ungauged Basins into Practice*, chapter 18. Canadian Water Resources Association.

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., and Kessler, M. (2017). Data Descriptor: Climatologies at high resolution for the earth's land surface areas. *Scientific Data*. doi: 10.1038/sdata.2017.122.

Keeling, R. F. and Keeling, C. D. (2017). Monthly In Situ CO2 Data - Mauna Loa Observatory, Hawaii. In Scripps CO2 Program Data. UC San Diego Library Digital Collections. doi: 10.6075/J08W3BHW, last accessed February 2019.

Kennedy, M. C. and O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. doi: 10.1111/1467-9868.00294.

Keshava, N. and Mustard, J. F. (2002). Spectral unmixing. *IEEE Signal Processing Magazine*. doi: 10.1109/79.974727.

Kolberg, S. and Gottschalk, L. (2006). Updating of snow depletion curve with remote sensing data. *Hydrological Processes*. doi: 10.1002/hyp.6060.

Kolberg, S. and Gottschalk, L. (2010). Interannual stability of grid cell snow depletion curves as estimated from MODIS images. *Water Resources Research*.

Kolberg, S., Rue, H., and Gottschalk, L. (2006). A Bayesian spatial assimilation scheme for snow coverage observations in a gridded snow model. *Hydrology and Earth System Sciences*. doi: 10.5194/hess-10-369-2006.

Konzelmann, T., van de Wal, R. S. W., Greuell, W., Bintanja, R., Henneken, E. A. C., and Abe-Ouchi, A. (1994). Parameterization of global and longwave incoming radiation for the Greenland Ice Sheet. *Global and Planetary Change*. doi: 10.1016/0921-8181(94)90013-2.

Koster, R. D. and Suarez, M. J. (2001). Soil Moisture Memory in Climate Models. *Journal of Hydrometeorology*. doi: 10.1175/1525-7541(2001)002%3C0558:SMMICM%3E2.0.CO;2.

Lawrence, M. G. (2005). The Relationship between Relative Humidity and the Dewpoint Temperature in Moist Air. *Bulletin of the American Meteorological Society*. doi: 10.1175/BAMS-86-2-225.

Leisenring, M. and Moradkhani, H. (2011). Snow water equivalent prediction using Bayesian data assimilation methods. *Stochastic Environmental Research and Risk Assessment*. doi: 10.1007/s00477-010-0445-5.

Lejeune, Y., Dumont, M., Panel, J.-M., Lafaysse, M., Lapalus, P., Le Gac, E., Lesaffre, B., and Morin, S. (2019). 57 years (1960–2017) of snow and meteorological observations from a mid-altitude mountain site (Col de Porte, France, 1325 m of altitude). *Earth System Science Data*.

Li, D., Durand, M., and Margulis, S. A. (2017). Estimating snow water equivalent in a sierra nevada watershed via spaceborne radiance data assimilation. *Water Resources Research*. doi: 10.1002/2016WR018878.

Lievens, H., Demuzere, M., Marshall, H. P., Reichle, R. H., Brucker, L., Brangers, I., de Rosnay, P., Dumont, M., Girotto, M., Immerzeel, W. W., Jonas, T., Kim, E. J., Koch, I., Marty, C., Saloranta, T., Schöber, J., and De Lannoy, G. J. M. (2019). Snow depth variability in the Northern Hemisphere mountains observed from space. Manuscript submitted to Nature Communications. Currently in review.

Link, T. and Marks, D. (1999). Distributed simulation of snowcover mass- and energy-balance in the Boreal forest. *Hydrological Processes*. doi: https://doi.org/10.1002/(SICI)1099-1085(199910)13:14/15%3C2439::AID-HYP866%3E3.0.CO;2-1.

Liston, G. E. (2004). Representing subgrid snow cover heterogeneities in regional and global models. *Journal of Climate*. doi: 10.1175/1520-0442(2004)017<1381:RSSCHI>2.0.CO;2.

Liston, G. E. and Elder, K. (2006). A Meteorological Distribution System for High-Resolution Terrestrial Modeling (MicroMet). *Journal of Hydrometeorology*. doi: 10.1175/JHM486.1.

Lorenz, E. N. (1963). Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences*. doi: 10.1175/1520-0469(1963)020%3C0130:DNF%3E2.0.CO;2.

Luce, C. H. and Tarboton, D. G. (2004). The application of depletion curves for parameterization of subgrid variability of snow. *Hydrological Processes*. doi: 10.1002/hyp.1420.

Lyons, L. and Wardle, N. (2018). Statistical issues in searches for new phenomena in High Energy Physics. *Journal of Physics G: Nuclear and Particle Physics*. doi:10.1088/1361-6471/aa9408.

MacKay, D. J. C. (2005). *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press.

Magnusson, J., Gustafsson, D., Hüsler, F., and Jonas, T. (2014). Assimilation of point SWE data into a distributed snow cover model comparing two contrasting methods. *Water Resources Research*. doi: 10.1002/2014WR015302.

Magnusson, J., Winstral, A., Stordal, A. S., Essery, R., and Jonas, T. (2017). Improving physically based snow simulations by assimilating snow depths using the particle filter. *Water Resources Research*. doi: 10.1002/2016WR019092.

Mahrt, L. (1999). Stratified atmospheric boundary layers. *Boundary-Layer Meteorology*. doi: 10.1023/A:1001765727956.

Mankin, J. S., Viviroli, D., Singh, D., Hoekstra, A. J., and Diffenbaugh, N. S. (2015). The potential for snow to supply human water demand in the present and future. *Environmental Research Letters*. doi: 10.1088/1748-9326/10/11/114016.

Margulis, S. A., Cortés, G., Girotto, M., and Durand, M. (2016). A Landsat-Era Sierra Nevada Snow Reanalysis (1985-2015). *Journal of Hydrometeorology*, 17:1203–1221. doi: 10.1175/JHM-D-15-0177.1.

Margulis, S. A., Girotto, M., Cortés, G., and Durand, M. (2015). A Particle Batch Smoother Approach to Snow Water Equivalent Estimation. *Journal of Hydrometeorology*, 16:1752–1772. doi: 10.1175/JHM-D-14-0177.1.

Marks, D., Domingo, J., Susong, D., Link, T. E., and Garen, D. C. (1999). A spatially distributed energy balance snowmelt model for application in mountain basins. *Hydrological Processes*. doi: 10.1002/(SICI)1099-1085(199909)13:12/13<1935::AIDHYP868>3.0.CO;2-C.

Marti, R., Gascoin, S., Berthier, E., de Pinel, M., Houet, T., and Laffly, D. (2016). Mapping snow depth in open alpine terrain from stereo satellite imagery. *The Cryosphere*. doi: 10.5194/tc-10-1361-2016.

Martinec, J. and Rango, A. (1981). Areal Distribution of Snow Water Equivalent Evaluated by Snow Cover Monitoring. *Water Resources Research*. doi: 10.1029/WR017i005p01480.

McCreight, J. L., Small, E. E., and Larson, K. M. (2014). Snow depth, density, and SWE estimates derived from GPS reflection data: Validation in the western U.S. *Water Resources Research*. doi: 10.1002/2014WR015561.

Metref, S., Hannart, A., Ruiz, J., Bocquet, M., Carrassi, A., and Ghil, M. (2019). Estimating model evidence using ensemble-based data assimilation with localization - The model selection problem. *Quarterly Journal of the Royal Meteorological Society*. doi: 10.1002/qj.3513.

Meyer, P., Itten, K. I., Kellenberger, T., Sandmeier, S., and Sandmeier, R. (1993). Radiometric corrections of topographically induced effects on Landsat TM data in an alpine environment. *ISPRS Journal of Photogrammetry and Remote Sensing*. doi: 10.1016/0924-2716(93)90028-L.

Moller, D., Andreadis, A. M., Bormann, K. J., Hensley, S., and Painter, T. H. (2017). Mapping Snow Depth From Ka-Band Interferometry: Proof of Concept and Comparison With Scanning Lidar Retrievals. *IEEE Geoscience and Remote Sensing Letters*. doi: 10.1109/LGRS.2017.2686398.

Molotch, N. P. and Margulis, S. A. (2008). Estimating the distribution of snow water equivalent using remotely sensed snow cover data and a spatially distributed snowmelt model: A multi-resolution, multi-sensor comparison. *Advances in Water Resources*. doi: 10.1016/j.advwatres.2008.07.017.

Molotch, N. P., Margulis, S. A., and Jespen, S. M. (2010). Response to comment by A.G. Slater, M.P. Clark, and A.P. Barrett on 'Estimating the distribution of snow water equivalent using remotely sensed snow cover data and a spatially distributed snowmelt model: A multi-resolution, multi-sensor comparison' [[Adv. Water Resour. 31 (2008) 1503–1514]. Adv Water Resour 2009;32(11):1680–4]. *Advances in Water Resources*. doi: 10.1016/j.advwatres.2009.11.008.

Morzfeld, M., Day, M. S., Grout, R. W., Pau, G. S. H., Finsterle, S. A., and Bell, J. B. (2018). Iterative Importance Sampling Algorithms for Parameter Estimation. *SIAM Journal on Scientific Computing*. doi: 10.1137/16M1088417.

Mote, P. W., Li, S., Lettenmaier, D. P., Xiao, M., and Engel, R. (2018). Dramatic declines in snowpack in the western US. *npj Climate and Atmospheric Science*. doi: 10.1038/s41612-018-0012-1.

Mott, R., Gromke, C., Grünewald, T., and Lehning, M. (2013). Relative importance of advective heat transport and boundary layer decoupling in the melt dynamics of a patchy snow cover. *Advances in Water Resources*. doi: 10.1016/j.advwatres.2012.03.001.

Müller-Wilm, U. (2018). Sen2Cor Configuration and User Manual. Ref. S2-PDGS-MPC-L2A-SUM-V2.5.5. Available at: http://step.esa.int/main/third-party-plugins-2/sen2cor/ (last access: December 2018).

Musial, J. P., Hüsler, F., Sütterlin, M., Neuhaus, C., and Wunderle, S. (2014). Probabilistic approach to cloud and snow detection on Advanced Very High Resolution Radiometer (AVHRR) imagery. *Atmospheric Measurement Techniques*.

Musselman, K. N., Clark, M. P., Liu, C., Ikeda, K., and Rasmussen, R. (2017). Slower snowmelt in a warmer world. *Nature Climate Change*, 7. doi: 10.1038/nclimate3225.

Musselman, K. N., Lehner, F., Ikeda, K., Clark, M. P., Prein, A. F., Liu, C., Barlage, M., and Rasmussen, R. (2018). Projected increases and shifts in rain-on-snow flood risk over western North America. *Nature Climate Change*. doi: 10.1038/s41558-018-0236-4.

Neal, R. M. (1993). Probabilistic Inference Using Markov Chain Monte Carlo Methods. Technical Report CRG-TR-91-1, Department of Computer Science, University of Toronto.

Niu, G. Y., Seo, K. W., Yang, Z. L., Wilson, C., Su, H., Chen, J., and Rodell, M. (2007). Retrieving snow mass from GRACE terrestrial water storage change with a land surface model. *Geophysical Research Letters*. doi: 10.1029/2007GL030413.

Nolin, A. W., Dozier, J., and Mertes, L. A. K. (1993). Mapping alpine snow using a spectral mixture modeling technique. *Annals of Glaciology*. doi: 10.3189/S0260305500012702.

NPI (2014). Terrengmodell Svalbard (S0 Terrengmodell), Delmodell 5m 2010_13828_33. Norwegian Polar Institute (NPI) DEM. Last access: December 2018. doi: 10.21334/npolar.2014.dce53a47.

Obu, J., Westermann, S., Bartsch, A., Berdnikov, N., Christiansen, H. H., Dashtseren, A., Delaloye, R., Elberling, B., Etzelmüller, B., Kholodov, A., Khomutov, A., Kääb, A., Leibman, M. O., Lewkowicz, A. G., Panda, S. K., Romanovsky, V., Way, R. G., Westergaard-Nielsen, A., Wu, T., Yamkhin, J., and Zou, D. (2019). Northern Hemisphere permafrost map based on TTOP modelling for 2000-2016 at 1 km^2 scale. *Earth-Science Reviews*. doi: 10.1016/j.earscirev.2019.04.023.

Ødegaard, H. A., Andersen, T., and Østerm, G. (1980). Application of Satellite Data for Snow Mapping in Norway. Part of NASA Conference Publication 2116: Operational Applications of Satellite Snowcover Observations. Avalable at: https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19800017251.pdf [last accessed May 2019].

Østby, T. I., Schuler, T. V., Hagen, J. O., Hock, R., Kohler, J., and Reijmer, C. H. (2017). Diagnosing the decline in climatic mass balance of glaciers in Svalbard over 1957-2014. *The Cryosphere*, 40:191–215. doi: 10.5194/tc-11-191-2017.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*. doi: https://doi.org/10.1109/TSMC.1979.4310076.

Painter, T. H., Bryant, A. C., and Skiles, S. M. (2012). Radiative forcing by light absorbing impurities in snow from MODIS surface reflectance data. *Geophysical Research Letters*. doi: 10.1029/2012GL052457.

Painter, T. H. et al. (2016). The Airborne Snow Observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo. *Remote Sensing of Environment*. doi: 10.1016/j.rse.2016.06.018.

Painter, T. H., Rittger, K., McKenzie, C., Slaughter, P., Davis, R. E., and Dozier, J. (2009). Retrieval of subpixel snow covered area, grain size, and albedo from MODIS. *Remote Sensing of Environment*, 113:868–879. doi: 10.1016/j.rse.2009.01.001.

Parajka, J., Haas, P., Kirnbauer, R., Jansa, J., and Blöschl, G. (2012). Potential of time-lapse photography of snow for hydrological purposes at the small catchment scale. *Hydrological Processes*. doi: 10.1002/hyp.8389.

Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics Surveys*. doi: 10.1214/09-SS057.

Pepin, N. et al. (2015). Elevation-dependent warming in mountain regions of the world. *Nature Climate Change*. doi: 10.1038/nclimate2563.

Peters, G. P. (2018). Beyond Carbon Budgets. *Nature Geoscience*. doi: 10.1038/s41561-018-0142-4.

Piazzi, G., Thirel, G., Campo, L., and Gabellani, S. (2018). A particle filter scheme for multivariate data assimilation into a point-scale snowpack model in an alpine environment. *The Cryosphere*.

Pithan, F. and Mauritsen, T. (2014). Arctic amplification dominated by temperature feedbacks in contemporary climate models. *Nature Geoscience*. doi: 10.1038/ngeo2071.

Planet Team (2019). Planet Application Program Interface: In Space for Life on Earth. San Francisco, CA. https//api.planet.com/ and https//www.planet.com/. last visited 01.04.2018.

Poster, C. et al. (2018). ArcticDEM. doi :10.7910/DVN/OHHUKH, Harvard Dataverse, V1, [Accessed: May 2019].

Proksch, M., Mätzler, C., Wiesmann, A., Lemmetyinen, J., Schwank, M., Löwe, H., and Schneebeli, M. (2015). MEMLS3&a: Microwave Emission Model of Layered Snowpacks adapted to include backscattering, journal = Geoscientific Model Development. doi: 10.5194/gmd-8-2611-2015.

Putman, W. M., Da Silva, A. M., Ott, L. E., and Darmenov, A. (2014). Model Configuration for the 7-km GEOS-5 Nature Run, Ganymed Release. [Non-hydrostatic 7 km Global Mesoscale Simulation]. Published in the the GMAO Publications - Office Notes No. 5 (Version 1.0), 18 pp available from http://ntrs.nasa.gov.

Qu, X. and Hall, A. (2014). On the persistent spread in snow-albedo feedback. *Climate Dynamics*. doi: 10.1007/s00382-013-1774-0.

Räisänen, P., Makkonen, R., Kirkevåg, A., and Debernard, J. B. (2017). Effects of snow grain shape on climate simulations: sensitivity tests with the Norwegian Earth System Model. *The Cryosphere*. doi: 10.5194/tc-11-2919-2017.

Raleigh, M. S., Lundquist, J. D., and Clark, M. P. (2015). Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework. *Hydrology and Earth System Sciences*. doi: 10.5194/hess-19-3153-2015.

Rango, A. and Martinec, J. (1995). Revisiting the Degree-Day Method for Snowmelt Computations. *Journal of the American Water Resources Association*. doi: 10.1111/j.1752-1688.1995.tb03392.x.

Rasmussen, R., Liu, C., Ikeda, K., Gochis, D., Yates, D., Chen, F., Tewari, M., Barlage, M., Dudhia, J., Yu, W., and Miller, K. (2011). High-Resolution Coupled Climate Runoff Simulations of Seasonal Snowfall over Colorado: A Process Study of Current and Warmer Climate. *Journal of Climate*. doi: 10.1175/2010JCLI3985.1.

Revuelto, J., Azorin-Molina, C., Alonso-González, E., Sanmiguel-Vallelado, A., Navarro-Serrano, F., Rico, I., and López-Moreno, I. (2017). Meteorological and snow distribution data in the Izas Experimental Catchment (Spanish Pyrenees) from 2011 to 2017. *Earth System Science Data*. doi: 10.5194/essd-9-993-2017.

Richter, R., Louis, J., and Müller-Wilm, U. (2012). Sentinel-2 MSI - Level 2A Products Algorithm Theoretical Basis Document. Issue 2.0.

Riggs, G. A., Hall, D. K., and Román, M. O. (2017). Overview of NASA's MODIS and Visibile Infrared Imaging Radiometer Suite (VIIRS) snow-cover Earth System Data Records. *Earth System Science Data*, 9:765–777. doi: 10.5194/essd-9-765-2017.

Rittger, K., Bair, E. H., Kahl, A., and Dozier, J. (2016). Spatial estimates of snow water equivalent from reconstruction. *Advances in Water Resources*. doi: 10.1016/j.advwatres.2016.05.015.

Ruiz-Arias, J. A., Alsamamra, H., Tovar-Pescador, J., and Pozo-Vázquez, D. (2010a). Proposal of a regressive model for the hourly diffuse solar radiation under all sky conditions. *Energy Conversion and Management*. doi: 10.1016/j.enconman.2009.11.024.

Ruiz-Arias, J. A., Cebecauer, T., Tovar-Pescador, J., and Šúri, M. (2010b). Spatial disaggregation of satellite-derived irradiance using a high-resolution digital elevation model. *Solar Energy*. doi: 10.1016/j.solener.2010.06.002.

Sadegh, M. and Vrugt, J. A. (2013). Bridging the gap between GLUE and formal statistical approaches: Approximate Bayesian Computation. *Hydrology and Earth System Sciences*. doi: 10.5194/hess-17-4831-2013.

Sakov, P. and Oke, P. R. (2008). A deterministic formulation of the ensemble Kalman filter: an alternative to ensemble square root filters. *Tellus A*. doi: 10.1111/j.1600-0870.2007.00299.x.

Salomonson, V. V. and Appel, I. (2004). Development of the Aqua MODIS NDSI Fractional Snow Cover Algorithm and Validation Results. *IEEE Transactions on Geoscience and Remote Sensing*. doi: 10.1109/TGRS.2006.876029.

Salomonson, V. V. and Rango, A. (1974). ERTS-1 Applications in Hydrology and Water Resources. *Journal - American Water Works Association*. doi: 10.1002/j.1551-8833.1974.tb01996.x.

Schaepman-Strub, G., Schaepman, C., Painter, T. H., Dangel, S., and Martonchik, J. V. (2006). Reflectance quantities in optical remote sensing—definitions and case studies. *Remote Sensing of Environment*. doi: 10.1016/j.rse.2006.03.002.

Schneider, T., Kaul, C. M., and Pressel, K. G. (2019). Possible climate transitions from breakup of stratocumulus decks under greehouse warming. *Nature Geoscience*. doi: 10.1038/s41561-019-0310-1.

Schoups, G. and Vrugt, J. (2010). A Formal Likelihood Function for Parameter and Predictive Inference of Hydrologic Models With Correlated, Heteroscedastic, and Non-Gaussian Errors. *Water Resources Research*. doi: 10.1029/2009WR008933.

Schuler, T. V., Crochet, P., Hock, R., Jackson, M., Barstad, I., and Jóhannesson, T. (2008). Distribution of snow accumulation on the Svartisen ice cap, Norway, assessed by a model of orographic precipitation. *Hydrological Processes*. doi: 10.1002/hyp.7073.

Schuur, E. A. G. et al. (2015). Climate change and the permafrost carbon feedback. *Nature*. doi: 10.1038/nature14338.

Schwanghart, W. and Scherler, D. (2014). Short Communication: TopoToolbox 2 - MATLAB-based software for topographic analysis and modeling in Earth surface sciences. *Earth Surface Dynamics*. doi: 10.5194/esurf-2-1-2014.

Selkowitz, D. J. and Forster, R. (2016). Automated mapping of persistent ice and snow cover across the western U.S. with Landsat. *ISPRS Journal of Photogrammetry and Remote Sensing*. doi: 10.1016/j.isprsjprs.2016.04.001.

Serreze, M. C., Barrett, A. P., Stroeve, J. C., Kindig, D. N., and Holland, M. M. (2009). The emergence of surface-based arctic amplification. *The Cryosphere*. doi: 10.5194/tc-3-11-2009.

Simon, E. and Bertino, L. (2009). Application of the Gaussian anamorphosis to assimilation in a 3-D coupled physical-ecosystem model of the North Atlantic with the EnKF: a twin experiment. *Ocean Science*. doi: 10.5194/os-5-495-2009.

Sirguey, P., Mathieu, R., and Arnaud, Y. (2009). Subpixel monitoring of the seasonal snow cover with MODIS at 250 m spatial resolution in the Southern Alps of New Zealand: Methodology and accuracy assessment. *Remote Sensing of Environment*. doi: 10.1016/j.rse.2008.09.008.

Skiles, S. M., Flanner, M., Cook, J. M., Dumont, M., and Painter, T. H. (2018). Radiative forcing by light-absorbing particles in snow. *Nature Climate Change*. doi: 10.1038/s41558-018-0296-5.

Slater, A. G., Barrett, A. P., Clark, M. P., Lundquist, J. D., and Raleigh, M. S. (2013). Uncertainty in seasonal snow reconstruction: Relative impacts of model forcing and image availability. *Advances in Water Resources*. doi: 10.1016/j.advwatres.2012.07.006.

Slater, A. G. and Clark, M. P. (2006). Snow Data Assimilation via an Ensemble Kalman Filter. *Journal of Hydrometeorology*. doi: 10.1175/JHM505.1.

Slater, A. G., Clark, M. P., and Barrett, A. P. (2009). Comment on "Estimating the distribution of snow water equivalent using remotely sensed snow cover data and a spatially distributed snowmelt model: A multi-resolution multi-sensor comparison" by Noah P. Molotch and Steven A. Margulis [Adv. Water Resour. 31 (2008) 1503-1514]. *Advances in Water Resources*. doi: 10.1016/j.advwatres.2009.09.001.

Smith, C. D., Kontu, A., Laffin, R., and Pomeroy, J. W. (2017). An assessment of two automated snow water equivalent instruments during the WMO Solid Precipitation Intercomparison Experiment. *The Cryosphere*. doi: 10.5194/tc-11-101-2017.

Smith, R. B. and Barstad, I. (2004). A Linear Theory of Orographic Precipitation. *Journal of the Atmospheric Sciences*. doi: 10.1175/1520-0469(2004)061<1377:ALTOOP>2.0.CO;2.

Smyth, E. J., Raleigh, M. S., and Small, E. E. (2019). Particle Filter Data Assimilation of Monthly Snow Depth Observations Improves Estimation of Snow Density and SWE. *Water Resources Research*. doi: 10.1029/2018WR023400.

SNL (2014). Sandia National Labs PV_LIB Toolbox. Available at: https://pvpmc.sandia.gov/applications/pv_lib-toolbox (last access: December 2018).

Spandre, P., François, H., Verfaillie, D., Pons, M., Vernay, M., Lafaysse, M., George, E., and Morin, S. (2019). Winter tourism under climate change in the Pyrenees and the French Alps: relevance of snowmaking as a technical adaptation. *The Cryosphere*.

Steffen, W. et al. (2018). Trajectories of the Earth System in the Anthropocene. *PNAS*. doi: 10.1073/pnas.1810141115.

Stewart, I., Cayan, D. R., and Dettinger, M. D. (2005). Changes toward Earlier Streamflow Timing across Western North America. *Journal of Climate*. doi: 10.1175/JCLI3321.1.

Stigter, E. E., Wanders, N., Saloranta, T. M., Shea, J. M., Bierkens, M. F., and Immerzeel, W. W. (2017). Assimilation of snow cover and snow depth into a snow model to estimate snow water equivalent and snowmelt runoff in a Himalayan catchment. *The Cryosphere*. doi: 10.5194/tc-11-1647-2017.

Stöckli, R., Vermote, E., Saleous, N., Simmon, R., and Herring, D. (2004). The Blue Marble Next Generation - A true color earth dataset including seasonal dynamics from MODIS. Available at: https://visibleearth.nasa.gov/, last accessed February 2019.

Stordal, A. S. and Elsheikh, A. H. (2015). Iterative ensemble smoothers in the annealed importance sampling framework. *Advances in Water Resources*. doi: 10.1016/j.advwatres.2015.09.030.

Stuart, A. M. (2010). Inverse problems: A Bayesian perspective. *Acta Numerica*. doi: 10.1017/S0962492910000061.

Stuecker, M. F. et al. (2018). Polar amplification dominated by local forcing and feedbacks. *Nature Climate Change*. doi: 10.1038/s41558-018-0339-y.

Sturm, M. (2015). White water: Fifty years of snow research in WRR and the outlook for the future. *Water Resources Research*. doi: 10.1002/2015WR017242.

Sturm, M., Goldstein, M. A., and Parr, C. (2017). Water and life from snow: A trillion dollar science question. *Water Resources Research*. doi: 10.1002/2017WR020840.

Sturm, M., Holmgren, J., König, M., and Morris, K. (1997). The thermal conductivity of seasonal snow. *Journal of Glaciology*. doi: 10.3189/S0022143000002781.

Sturm, M. and Wagner, A. M. (2010). Using repeated patterns in snow distribution modeling: An Arctic example. *Water Resources Research*. doi: 10.1029/2010WR009434.

Su, H., Yang, Z. L., Dickinson, R. E., Wilson, C. R., and Niu, G. Y. (2010). Multisensor snow data assimilation at the continental scale: The value of Gravity Recovery and Climate Experiment terrestrial water storage information. *Journal of Geophysical Research: Atmospheres*. doi: 10.1029/2009JD013035.

Swenson, S. and Lawrence, D. M. (2012). A new fractional snow-covered area parameterizationfor the Community Land Model and its effect on the surface energy balance. *Journal of Geophysical Research : Atmospheres*. doi: 10.1029/2012JD018178.

Takala, M., Luojus, K., Pulliainen, J., Derksen, C., Lemmetyinen, J., Kärnä, J.-P., Koskinen, J., and Bojkov, B. (2011). Estimating northern hemisphere snow water equivalent for climate research through assimilation of space-borne radiometer data and ground-based measurements. *Remote Sensing of Environment*. doi: 10.1016/j.rse.2011.08.014.

Tarantola, A. (2005). *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM. doi: 10.1137/1.9780898717921.

Tarboton, D. G. and Luce, C. H. (1996). Utah Energy Balance Snow Accumulation and Melt Model. Computer model techincal description and users guide, Utah Water Research Laboratory and USDA Forest Service Intermountain Research Station, Logan, Utah.

Teweldebrhan, A. T., Burkhart, J. F., Schuler, T. V., and Xu, C. Y. (2019). Improving the Informational Value of MODIS Fractional Snow Cover Area Using Fuzzy Logic Based Ensemble Smoother Data Assimilation Frameworks. *Remote Sensing*. doi: 10.3390/rs11010028.

Treichler, D. and Kääb, A. (2017). Snow depth from ICESat laser altimetry - A test study in southern Norway. *Remote Sensing of Environment*. doi: 10.1016/j.rse.2017.01.022.

Trofaier, A. M., Westermann, S., and Bartsch, A. (2017). Progress in space-borne studies of permafrost for climate science: Towards a multi-ECV approach. *Remote Sensing of Environment*. doi: 10.1016/j.rse.2017.05.021.

Vahedifard, F., AghaKouchak, A., Ragno, E., Shahrokhabadi, S., and Mallakpour, I. (2017). Lessons from the Oroville dam. *Science*. doi: 10.1126/science.aan0171.

Vallis, G. K. (2017). *Atmopsheric and Oceanographic Fluid Dynamics*. Cambridge University Press. doi: 10.1017/9781107588417.

van Leeuwen, P. J. (2009). Particle Filtering in Geophysical Systems. *Monthly Weather Review*. doi: 10.1175/2009MWR2835.1.

van Leeuwen, P. J. (2015). Representation errors and retrievalsin linear and nonlinear data assimilation. *Quarterly Journal of the Royal Meteorological Society*. doi: 10.1002/qj.2464.

van Leeuwen, P. J. and Evensen, G. (1996). Data Assimilation and Inverse Methods in Terms of a Probabilistic Formulation. *Monthly Weather Review*. doi: 10.1175/1520-0493(1996)124%3C2898:DAAIMI%3E2.0.CO;2.

van Leeuwen, P. J., Künsch, H. R., Nerger, L., Potthast, R., and Reich, S. (2019). Particle filters for high-dimensional geoscience applications: A review. *Quarterly Journal of the Royal Meteorological Society*. doi: 10.1002/qj.3551.

Vermote, E., Justice, C., and Claverie, M. Franch, B. (2016). Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sensing of Environment*. doi: 10.1016/j.rse.2016.04.008.

Vermote, E. and Wolfe, R. (2015). MOD09GA MODIS/Terra Surface Reflectance Daily L2G Global 1km and 500m SIN Grid V006, NASA EOSDIS LP DAAC. Accessed: 06-2018, doi: 10.5067/MODIS/MYD09GA.006.

Vikhamar, D. and Solberg, R. (2003). Snow-cover mapping in forests by constrained linear spectral unmixing of MODIS data. *Remote Sensing of Environment*. doi: 10.1016/j.rse.2003.06.004.

Vionnet, V., Brun, E., Morin, S., Boone, A., Faroux, S., Le Moigne, P., Martin, E., and Willemet, J.-M. (2012). The detailed snowpack scheme Crocus and its implementation in SURFEX v7.2. *Geoscientific Model Development*, 5(3):773–791. doi: 10.5194/gmd-5-773-2012.

Vionnet, V., Martin, E., Masson, V., Lac, C., Bouvet, F., and Guyomarc'h, G. (2017). High-Resolution Large Eddy Simulation of Snow Accumulation in Alpine Terrain. *Journal of Geophysical Research: Atmospheres*. doi: 10.1002/2017JD026947.

Viviroli, D., Dürr, H. H., Messerli, B., Meybeck, M., and Weingartner, R. (2007). Mountains of the world, water towers for humanity: Typology, mapping, and global significance. *Water Resources Research*. doi: 10.1029/2006WR005653.

Vrugt, J. A., ter Braak, C. J. F., Diks, C. G. H., Robinson, B. A., Hyman, J. M., and Higdon, D. (2009). Accelerating Markov Chain Monte Carlo Simulation by Differential Evolution with Self-Adaptive Randomized Subspace Sampling. *International Journal of Nonlinear Sciences and Numerical Simulation*.

Žaksek, K., Oštir, K., and Kolaj, v. (2011). Sky-View Factor as a Relief Visualization Technique. *Remote Sensing*. doi: 10.3390/rs3020398.

Wallace, J. W. and Hobbs, P. V. (2006). *Atmospheric Science: An Introductory Survey*. Elsevier.

Westermann, S., Langer, M., Boike, J., Heikenfeld, M., Peter, M., Etzerlmüller, B., and Krinner, G. (2016). Simulating the thermal regime and thaw processes of ice-rich permafrost ground with the land-surface model CryoGrid 3. *Geoscientific Model Development*. doi: 10.5194/gmd-9-523-2016.

Westermann, S., Peter, M., Langer, M., Schwamborn, G., Schirrmeister, L., Etzelmüller, B., and Boike, J. (2017). Transient modeling of the ground thermal conditions using satellite data in the Lena River delta, Siberia. *The Cryosphere*. doi: 10.5194/tc-11-1441-2017.

Wikle, C. K. and Berliner, L. M. (2007). A Bayesian tutorial for data assimilation. *Physica D*. doi:10.1016/j.physd.2006.09.017.

Winther, J.-G. and Hall, D. K. (1999). Satellite-derived snow coverage related to hydropower production in Norway: present and future. *International Journal of Remote Sensing*. doi: 10.1080/014311699211570.

Wrzesien, M. L., Durand, M. T., Pavelsky, T. M., Howat, I. M., Margulis, S. A., and Huning, L. S. (2017). Comparison of Methods to Estimate Snow Water Equivalent at the Mountain Range Scale: A Case Study of the California Sierra Nevada. *Journal of Hydrometeorology*. doi: 10.1175/JHM-D-16-0246.1.

Wrzesien, M. L., Durand, M. T., Pavelsky, T. M., Kapnick, S. B., Zhang, Y., Guo, J., and Shum, C. K. (2018). A New Estimate of North American Mountain Snow Accumulation From Regional Climate Model Simulations. *Geophysical Research Letters*. doi: 10.1002/2017GL076664.

Wunderle, S., Gross, T., and Hüsler, F. (2016). Snow Extent Variability in Lesotho Derived from MODIS Data (2000-2014). *Remote Sensing*. doi: 10.3390/rs8060448.

Xia, Y. et al. (2012). Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *Journal of Geophysical Research: Atmospheres*. doi: 10.1029/2011JD016048.

Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Bender, S. M., Case, A., Costello, C., Dewitz, J., Fry, J., Funk, M., Granneman, B., Liknes, G. C., Rigge, M., and Xian, G. (2018). A new generation of the United States National Land Cover Database: Requirements, research priorities, design, and implementation strategies. *ISPRS Journal of Photogrammetry and Remote Sensing*. doi: 10.1016/j.isprsjprs.2018.09.006.

Yılmaz, Y., Aalstad, K., and Sen, O. L. (2019). Multiple Remotely Sensed Lines of Evidence for a Depleting Seasonal Snowpack in the Near East. *Remote Sensing*. doi: 10.3390/rs11050483.

Yin, D., Cao, X., Chen, X., Shao, Y., and Chen, J. (2013). Comparison of automatic thresholding methods for snow-cover mapping using Landsat TM imagery. *International Journal of Remote Sensing*. doi: 10.1080/01431161.2013.803631.

You, J., Tarboton, D. G., and Luce, C. H. (2014). Modeling the snow surface temperature with a one-layer energy balance snowmelt model. *Hydrology and Earth System Sciences*. doi: 10.5194/hess-18-5061-2014.

Zhang, T. (2005). Influence of the seasonal snow cover on the ground thermal regime: An overview. *Reviews of Geophysics*. doi: 10.1029/2004RG000157.

Zhang, T., Stamnes, K., and Bowling, S. A. (1996). Impact of Clouds on Surface Radiative Fluxes and Snowmelt in the Arctic and Subarctic. *Journal of Climate*. doi: 10.1175/1520-0442(1996)009<2110:IOCOSR>2.0.CO;2.

Zhu, Z. and Woodcock, C. E. (2012). Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment*. doi: 10.1016/j.rse.2011.10.028.