Model-based IDF-curves in Norway

Can a high-resolution climate model resolve convective-scale extreme precipitation?

Eirik Nordgård

A thesis presented for the degree of
Master of Meteorology


UiO : University of Oslo

Department of Geosciences
University of Oslo
Norway

June 2021

Acknowledgement

First and foremost, huge thanks to my main supervisor, Malte Müller, for embracing me and my wish to write about extreme precipitation. Many thanks for your time spent with me, advising and guiding me in the right direction. You have been extremely helpful, and with your dedication provided tons of inspiration.

I would also like to thank Anita Verpe Dyrrdal and Julia Lutz for much appreciated expertise on extreme precipitation and help with computing. You are both highly killed in your fields of expertise, and it has been very inspirational for me as a student to listen to your recommendations and reasoning.

And to Jana Sillmann, many thanks for great inspiration in the initial faces of the project. To Andreas Dobler, many thanks for help with data and valuable contribution to discussions.

A special thanks to Oskar Landgren for helping out with my scrips and data. Our discussions and your calm and encouraging approach to my questions and struggles have been very helpful and immensely valuable to my work.

A big thanks goes to my fellow master students, friends and family. Your support has been great and especially valuable through the corona-situation. A special thanks to my life partner Eline for supporting me all the way through this study.

Abstract

High spatial- and temporal-resolution regional climate model output has been used to produce Intensity-Duration-Frequency (IDF) curves for the area surrounding Oslofjorden in Norway. Since operational flood design values are traditionally based on observed data, the scarce precipitation gauge network in Norway poses problems for spatially accurate IDF-values. In this study annual maximum precipitation from a convective-scale climate model simulation is used to infer IDF-values for multiple locations in the Oslofjorden area for a selection of return-periods and durations. A General Extreme Value (GEV) distribution is fitted to the data in an Bayesian inference, allowing derivation of confidence intervals. Different methods are used to compute annual maximum precipitation to highlight the challenges of retrieving single-point statistics from a gridded data-set.

In general the modelled IDF-curves were found to be consistent with observation-based IDF-values, strengthening the claim that the GEV distribution is suitable for short-duration extreme precipitation. However, the consistency of modelled return-values to observation-based return-values was very dependent on the choice of the annual maximum calculation method. The modeled return-values are consistent with with those of the observations, however, the confidence intervals are often too large to judge on the model's quality. This stresses the value of long-duration time-series in IDF calculation.

Furthermore, by analysing a future climate simulation we found the largest expected increase in return-values towards the end of the century for 3-6 hour durations. This is conflicting with multiple studies finding larger increase for shorter durations, and might be a result of still inadequate representation of sub 3 hour duration extreme precipitation in the convective permitting climate model.

Contents

1 Introduction

Flash floods caused by short duration precipitation is one of the costliest and most destructive weather-related hazards on Earth (Willner et al. 2018, Hosseinzadehtalaei et al. 2020b). In Norway roughly 1 billion NOK worth of damage is caused annually from preciptiation (Finansnorge 2020). Fast runoff and little terrestrial absorption of surface water due to impermeable surfaces makes urban areas particularly exposed to extreme precipitation (Mishra et al. 2012). These areas are also those who typically are most densely populated and with largest economic value, thereby posing as the areas with largest potential for overall loss. To cope with surface runoff, water managing structures or other flood control structures in both urban and rural areas are typically built and based upon Intensity-Duration-Frequency (IDF) curves. These are statistical properties of extreme preciptiation, quantifying the frequency and intensity of an event over a range of temporal durations (Hosseinzadehtalaei et al. 2018, Lutz et al. 2020).

There are several ways in which IDF-curves can be calculated. A Generalized Pareto Distribution (GDP) has been used to develop IDF curves for Europe for multiple sub-daily durations for different return periods ranging from 1 to 100 years (Hosseinzadehtalaei et al. 2020a). Another, now frequently used, method was presented to update Norwegian IDF-values. Here a Generalized Extreme Value (GEV) distribution was fitted to block maxima precipitation values (Lutz et al. 2020). In flat regions like Belgium and the Netherlands radar data has been used to estimate spatially continuous IDF-values (Goudenhoofdt et al. 2017), whereas in more mountainous areas like Norway a GEV distribution in an Bayesian hierarchical inference was used to obtain spatially continuous return-level maps (Dyrrdal et al. 2015). Another study derived future IDF-curves from climate change signals from the EURO-CORDEX regional climate models in combination with existing IDF-curves (Hosseinzadehtalaei et al. 2020a).

IDF-curves are usually based on historical data from a methodological stations. As rain gauges have become precise in measurement of liquid preciptiation (rain), they are used to obtain accurate design criteria. However, underestimation due to wind-induced undercatch is potentially huge, especially for snow (Grossi et al. 2017). Induced evaporation due to heating of the tipping bucket is another source of uncertainty and must also be taken into account (Savina et al. 2012). Rain gauge measurements are typically scarce, and those available are rarely long, high quality measurements. Data-availability poses a major problem for reliable, local IDF-values, mainly because spatial

precipitation patterns are poorly captured through a sparse gauge network (Courty et al. 2019).

There are many ways to deal with the data scarcity of observations in IDF computing, one of which is using data from climate models. Global climate models agree reasonably well on the projected changes in future precipitation, but are struggling to capture the intensity of extreme events (Sillmann et al. 2013). Several studies have also found that horizontal grid resolution is an important factor for proper representation of short-duration precipitation extremes (Kopparla et al. 2013, Pope et al. 2002). Recently higher-resolution climate models have performed well in resolving extreme precipitation. Precipitation events forecasted by a 2.5-km resolution weather prediction model was compared to a coarser-resolution model, where the increased resolution added value to forecast parameters such as precipitation magnitudes larger than 15 mm (Müller et al. 2017). Despite providing encouraging results, convective-permitting models are not flawless. Convection is still not properly resolved, sometimes causing artifacts like too intense showers (Hanley et al. 2015). This problem can arise if the typically resolution dependent model parameterization is poorly tuned (Duffy et al. 2003). Nevertheless, convection-permitting models are a big step forward and will be a valuable tool for studies of sub-hourly rainfall extremes in a changing climate in the years to come (Fowler et al. 2021).

As the global temperature increases, so does the potential for precipitation. The rate of change of saturation vapour pressures is described by the Clausius-Clapeyron (CC) relation at approximately $7\%^{\circ}C^{-1}$. Thus, as the average temperature of an area increases, the atmospheric moisture and hence the annual precipitation is expected to increase according to the CC-rate (Fowler et al. 2021). In some studies precipitation is observed to increase at a rate exceeding the CC-rate. Hourly precipitation measurements was found to have this so-called super-CC scaling of approximately $14\%C^{-1}$ (Lenderink et al. 2008), and other studies like (Berg et al. 2013) also found super CC increase for convective precipitation events at hourly time scales. However, sub-daily precipitation extremes has also been found to increase according to the CC-rate, possibly due to summer drying over large parts of the investigated area (Hodnebrog et al. 2019). The CC-relation roughly held for multi-decadal changes in extreme short-time precipitating in Japan (Fujibe 2013). Studies on why the super-CC scaling might occur for shorter-duration events rather than longer-duration annual precipitation has also been done (Park et al. 2017), and rarer sub-daily extreme preciptiation events has been found to intensify more than less rare events (Hosseinzadehtalaei

et al. 2020a). It is evident that to prevent damage and loss of lives in a future, warmer and wetter climate it is important to further develop the tools needed to understand the expected changes in extreme precipitation.

1.1 Structure of thesis

The thesis has the following structure. **Chapter 2** Background, where the purpose and motivation of the thesis is presented alongside background material on convection and precipitation. **Chapter 3** Theory, where detailed information on the underlying theory is presented. **Chapter 4** Data introduces the data and **Chapter 5** Methods explains the method used. **Chapter 6** is covering the Results of the analysis, and **Chapter 7** Summary and Discussion is discussing the method and the findings in the results. Very last comes **Chapter 8** Conclusion & Outlook where a summary and recommendations for further studies are presented.

2 Background

2.1 Purpose & Motivation

Extreme weather events has always fascinated me. Windstorms, landslides and bursts of precipitation are all rather common natural hazard in Norway, thus impacting most Norwegians one way or another. However, in the eastern part of the country where I am situated, convective extreme summer preciptiation is probably the most frequent extreme event. The share force of these summer showers are often stunning, and equally interesting is how challenging the forecasting of these events is. Generally extreme precipitation events are forecasted somewhere in an larger area within a longer period of time or not at all. Since these events often are small-scale precipitation events lasting from minutes to a few hours it is understandable that global climate models cant resolve them explicitly. Even regional forecasting models are struggling to foresee the exact where and when of these events. Once I learned that a high-resolution climate run covering the Scandinavian region recently was completed, I thought it was a great opportunity to gain understanding in how these convective systems are represented in the forecast, but also how they change towards the end of the century. Better understanding of which precipitation events a high-resolution climate model is capable of representing and contributing to understanding of the size and impact of extreme precipitation events in future Norway are both great motivations for this study.

The general purpose of the study is to improve knowledge on extreme precipitation events in Oslo through a high-resolution climate model. Extreme precipitation will in this study be defined as precipitation magnitude for different durations and return-periods derived through an extreme value distribution. Furthermore, data fitted to this extreme value distribution is extracted in different ways to investigate how point-statistics compares to larger-area statistics. This is done to achieve better understating of spatial differences in precipitation return-values and how these are represented through a high-resolution climate model. The official largest measured precipitating values in Norway from 2015 for some durations are presented in Table 1. Seen through Norwegian eyes these values are large, highlighting the potential for extreme precipitation in Norway.

Table 1: Largest observed precipitation values in Norway for different durations at stations operated by the Norwegian Meteorological Institute from 2015. (Førland et al. 2015)

Duration	1min	10min	30min	1h	2h	6h	12h	24h
mm	4.3	25.5	42.0	54.9	64.4	87.4	144.0	229.6

2.2 Precipitation

Precipitation is often categorized by the physical process creating it and may vary on both spatial and temporal scale. One can often relate the precipitation process to the relationship between the vertical and the horizontal extent (Yau et al. 1996). The temporal scale varies from minutes to days, and the the spatial scale varies from hundreds of meters to hundreds of kilometers. Large-scale precipitation events are typically on a synoptic scale, characterized by a horizontal dimension many times greater than the vertical extent. Hence, the vertical velocity is often small in theses systems. Smaller-scale systems have shorter horizontal extent and often vertical extent almost equal to or larger than the vertical extent. In these systems the vertical speed is often large compared to large-scale systems (Yau et al. 1996). Thus, we often classify precipitation as one of two types depending on the dominant mechanism responsible for vertical motion:

- Stratiform. Continuous, widespread precipitation formed by large scale ascent produced by topographic or frontal lifting or large scale convergence. Vertical motion typically on the order of tenths of centimeters (Yau et al. 1996).
- Convective. Showery, localized precipitation associated with smaller cumulus-scale convection in unstable air. Vertical motion typically on the order of meters (Yau et al. 1996).

Although the contrast between convective and stratiform precipitation is not always sharp it is a useful classification. In this study the focus will be oriented around convective-scale precipitation.

2.2.1 Convective Precipitation

Convective precipitation originates from clouds driven by convection. These clouds are cumulus-type clouds forming in an unstable atmosphere as warm

air rises due to buoyancy forces (Ahrens 2014). The vertical extent is partly determined by the depth of the unstable layer and its degree of instability (Yau et al. 1996). The cumulus clouds are distinct in their shape with a flat base and a "cotton-like" appearance. They can also be observed with narrow, towering plumes on top. Horizontally the extent is typically a couple of kilometers, while the initial thermal in which condensation first occurs can be only tens of meters. Typically the lifetime of cumulus clouds are from a few minutes to hours. If the persist for many hours they also tend to grow substantially in size, also horizontally (Yau et al. 1996). The horizontal and vertical extent is comparable, but a cumulus cloud can under right conditions develop to a Cumulonimbus cloud reaching all the way to the top of the stratosphere (Yau et al. 1996). Once the recognizable anvil-shaped Cumulonimbus is formed it is very likely to produce rain, thunder and lightning. Although cumulus clouds are commonly associated with nice weather due to their usual dependence on solar heating from the surface, once the rising parcel of air becomes saturated condensation will occur and precipitation will form. As the warm and moist air rises it cools to the point where the air is saturated on water vapour. Once condensation begins, latent heat is released and precipitation forms.

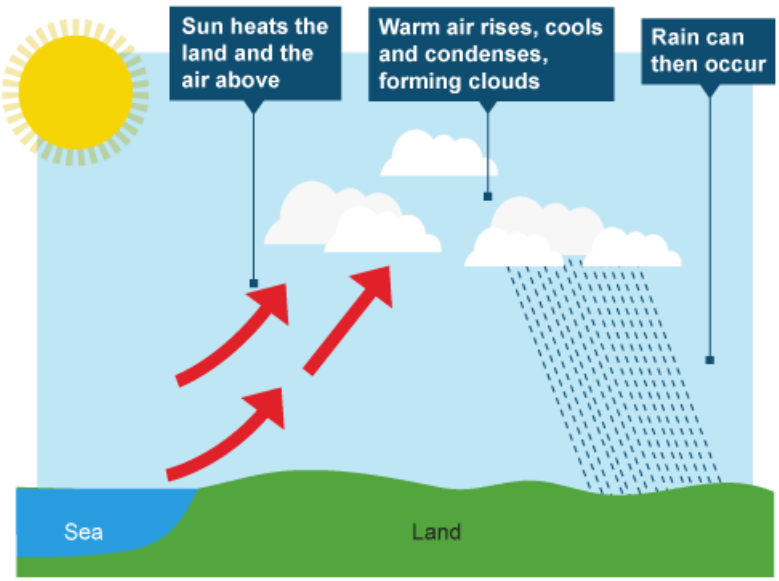

Figure 1: Simple illustration of convection.

Given the short temporal and horizontal scale the convective precipitation, often called showers, they are rather local phenomena you can watch form and propagate past your neighbourhood. Events at such small scales

proves hard to capture by climate models, and even operational weather forecasting models are far from spot on with predicting timing, localization and magnitude of these events.

Figure 2: Formation of typical convective Cumulus clouds with larger vertical than horizontal extent over land in Femundsmarka, Norway. Foto: Eirik Nordgård

2.2.2 Dry convection

When considering convection it is useful to include the concept of an air parcel. Virtually all mixing in the vertical occur due to exchange of macroscale bubbles of air, or "air parcels" with dimensions ranging from millimeters to kilometers (Wallace et al. 2006). Using the concept of an air parcel allows for ease of thought about rising air. It also makes certain assumptions regarding derivation of dynamics and other aspects easier, one of which is an adiabatic process. Convection is often treated as adiabatic, a thermodynamic process in which no heat or mass is transferred between the system in question, here the parcel, and the surroundings (Wallace et al. 2006). However, energy

can be transferred from the system to the surroundings through work. Being a compressible fluid the density of the atmosphere is a function of both temperature and pressure; $\rho = \rho(p, T)$. Furthermore it obeys the ideal gas law, $\rho = p/RT$, where p is pressure, t IS TEMPERATURE AND R is the ideal gas constant. As a parcel rises from an initial height z_1 it moves into an environment of lower pressure at height z_2, adjusting to this pressure by expanding along the way. This expansion exerts work on the environment, making the parcel cool. To determine the buoyancy of the parcel at z_2 one must know the evolution of the parcel temperature between z_1 and z_2 (Marshall et al. 2007), and consider a parcel of ideal gas of unit mass with a volume V. Using the first law of thermodynamics,

$$\delta Q = dU + dW \tag{1}$$

where δQ is an amount of heat exchange with the surroundings, dU is the change in energy and dW is the change in external work. Eq. (1) can be written as

$$\delta Q = c_v dT + pdV \tag{2}$$

where $c_c dT$ is the change in internal energy due to change in temperature dT and pdV is the work done by the parcel on its surroundings by expanding an amount dV. c_v is specific heat at constant volume. Rewriting Eq. (2) with repeated use of $p = \rho RT$ yields

$$\delta Q = c_p dT - \frac{dp}{\rho} \tag{3}$$

where $c_p = R + c_v$ is specific heat at constant pressure.

For an adiabatic processes $\delta Q = 0$, thus

$$c_p dT = \frac{dp}{\rho} \tag{4}$$

Furthermore, the hydrostatic balance,

$$\frac{\partial p}{\partial z} + g\rho = 0 \tag{5}$$

describes how pressure decreases with height in proportion with the weight of the overlying atmosphere. Using the hydrostatic equation and assuming $p \simeq p_e$, for a small upward displacement in Eq.(4) the parcel‘s temperature under an adiabatic displacement will change according to

$$\frac{dT}{dz} = -\frac{g}{c_p} = -\Gamma_d \tag{6}$$

where Γ_d is called the dry adiabatic lapse rate and e denotes *environment*. Now, if the parcel is displaced from z_1 to z_2 it will experience a restoring force depending on the density of the environment. At z_2 the environment has temperature $T_2 \simeq T_1 + (dT/dz)_e \delta z$ where $(dT/dz)_e = \Gamma_e$ is the environmental lapsrate with pressure p_2 and density $\rho_2 = p_2/RT_2$. The parcel has temperature $T_P = T_1 - \Gamma_d \delta z$, pressure p_2 and density $\rho_P = p_2/RT_P$. Depending on whether T_P is greater than, equal to or less than T_2 the parcel will be positively, neutrally or negatively buoyant. Thus, criterion for stability can be written

$$\left.\begin{matrix}\text{Unstable}\\ \text{Neutral}\\ \text{Stable}\end{matrix}\right\} \text{if} \left(\frac{dT}{dZ}\right)_e = \Gamma_e \begin{cases} < -\Gamma_d \\ = -\Gamma_d \\ > -\Gamma_d \end{cases} \tag{7}$$

$$\left.\begin{matrix}\text{Stable}\\ \text{Neutral}\\ \text{Unstable}\end{matrix}\right\} \text{if} \left(-\frac{dT}{dZ}\right)_e \begin{cases} < \Gamma_d \\ = \Gamma_d \\ > \Gamma_d \end{cases} \tag{8}$$

$$\begin{cases} \Gamma = \Gamma_d \\ \gamma = \Gamma_e \end{cases} \tag{9}$$

From Eq. (7) a compressible atmosphere is unstable if temperature decreases with height faster than the adiabatic laps rate. Given $c_p = 1005 Jkg^{-1}K^{-1}$ in Eq. (6), $\Gamma_d \simeq 10 Kkm^{-1}$. Typically the dry laps rate decreases from around 6,5 Kkm^{-1} at low latitudes to around 4.5 Kkm^{-1} at polar latitudes (Mokhov et al. 2006). Thus, no dry convection is expected and the atmosphere can generally be considered stable to dry convection.

2.2.3 Moist convection

Since the atmosphere is normally stable in the absence of condensation, most convection in the atmosphere is moist convection. If a moist air parcel is lifted it cools adiabatically, and if cooled enough to saturate, some water vapor condenses to form a cloud. Once saturated the parcel releases latent heat, adding buoyancy to the parcel and thus increasing the instability. Moisture of air is often expressed as specific humidity

$$q = \frac{\rho_v}{\rho} \tag{10}$$

This is the ratio of mas of water vapor, $\mathbf{q}$, to the mass of dry air per unit volume. q is conserved given no mixing. The saturation-specific humidity, q_*, is the specific humidity at which saturation occurs. q_* can be written:

$$q_* = \frac{\frac{e_s}{R_v T}}{\frac{p}{RT}} = \left(\frac{R}{R_v}\right)\frac{e_s}{p} \tag{11}$$

where e_s is saturated partial pressure of water vapor.

Implicitly this means that q_* is a function of both temperature and pressure. Lifting a parcel would decrease the pressure and cool the parcel. This alone would make q^* increase with altitude, but the exponential dependence of e_s on T counteracts this effect, making the q_* decrease rapidly with altitude. Thus a moist parcel usually do not have to rise a lot before it reaches the condensation level Z_c where $q > q_*$. At this point and above, excess water vapor will condensate creating a visible cloud. The cloud will extend up in the atmosphere until it obtains neutral buoyancy. As the parcel rises latent heat is released, resulting in a slower decrease of temperature with height compared to the dry adiabatic laps rate. Thus, a warmer or moister air parcel will result in a taller cloud compared to drier or cooler air. While the temperature of the parcel changes according to the dry adiabatic laps rate below the cloudbase, it must follow another adiabat upwards from z_c.

When condensation occurs there will be a release of latent heat in the amount $\delta Q = -Ldq$ where L is latent heat of condensation and dq is change in specific humidity q. For an air parcel undergoing moist adiabatic displacement one can insert this into (3) to get

$$c_p dT = \frac{dp}{\rho} - Ldq \tag{12}$$

Assuming hydrostatic balance of the environment, $dp/\rho = -gdz$ gives

$$d(c_p T + gz + Lq) = 0 \tag{13}$$

where $c_p T + gz$ is the dry static energy and Lq is the latent heat content. Combined these two terms are called the moist static energy. Furthermore, since the parcel is always saturated q can be replaced by q_* in Eq.(12), and since $q_* = q_*(p, T)$ now

$$dq_* = \frac{\partial q_*}{\partial p}dp + \frac{\partial q_*}{\partial T}dT \tag{14}$$

Inserting Eq. (11) in Eq. (14) yields

$$\frac{\partial q_*}{\partial p} = -(\frac{R}{R_v})\frac{e_s}{p^2} = -\frac{q_*}{p} \tag{15}$$

and

$$\frac{\partial q_*}{\partial T} = (\frac{R}{R_v})\frac{1}{p}\frac{de_s}{dT} = (\frac{R}{R_v})\frac{\beta e_s}{p} = \beta q_* \tag{16}$$

where $\beta e_s = de_s/dT$. Combining Eq. (15), Eq. (16) and Eq. (12) with some rearranging yields

$$-\frac{dT}{dz} = \Gamma_s = \Gamma_d \left[\frac{1 + Lq_*/RT}{1 + \beta Lq_*/c_p}\right] \tag{17}$$

wher Γ_s is called the saturated adiabatic lapse rate. The term inside the brackets are always less than unity, making $\Gamma_s < \Gamma_d$. However, at high latitudes q_* becomes very small, making them close to equal. The qualitative impact on condensation is evident; release of latent heat within a rising parcel makes it warmer and thus more buoyant, destabilizing the atmosphere, that is if

$$-(\frac{dT}{dz})_e = \gamma < \Gamma_s \tag{18}$$

where $\Gamma_s < \Gamma_d$ and e denotes *environment*.

Thus the stability criteria for moist air (Marshall et al. 2007) is

$$\begin{gathered} \text{Absolutely stable } \gamma < \Gamma_s \\ \text{Saturated neutral } \gamma = \Gamma_s \\ \text{Conditionally unstable } \Gamma_s < \gamma < \Gamma_d \\ \text{Dry neutral } \gamma = \Gamma_d \\ \text{Absolutely unstable } \gamma > \Gamma_d \end{gathered} \tag{19}$$

Once a saturated parcel is displaced upwards its temperature will decrease according to the pseudoadiabatic laps rate. If the environmental laps rate is greater (more negative) than Γ_d the parcel will be warmer than the surroundings and hence accelerate in the direction of the initial displacement. If the environmental laps rate is smaller (less negative) than Γ_s the parcel will be colder than the surroundings and hence accelerate in the direction opposite to the the initial displacement. The stability criterion from Eq. (19) is illustrated in Figure (3). According to (19), an important difference between moist and dry air is that initially stable moist air may be absolutely unstable or conditionally unstable if lifted.

2.3 Area

The area investigated in this thesis is Oslo, the capitol of Norway. This location is selected for a number of reasons. Firstly there are multiple stations in the area witch have reasonably long, high quality data series of precipitation. This is essential because it allows for comparison to the modeled IDF-values. When analysing IDF-curves short time series is a problem in it

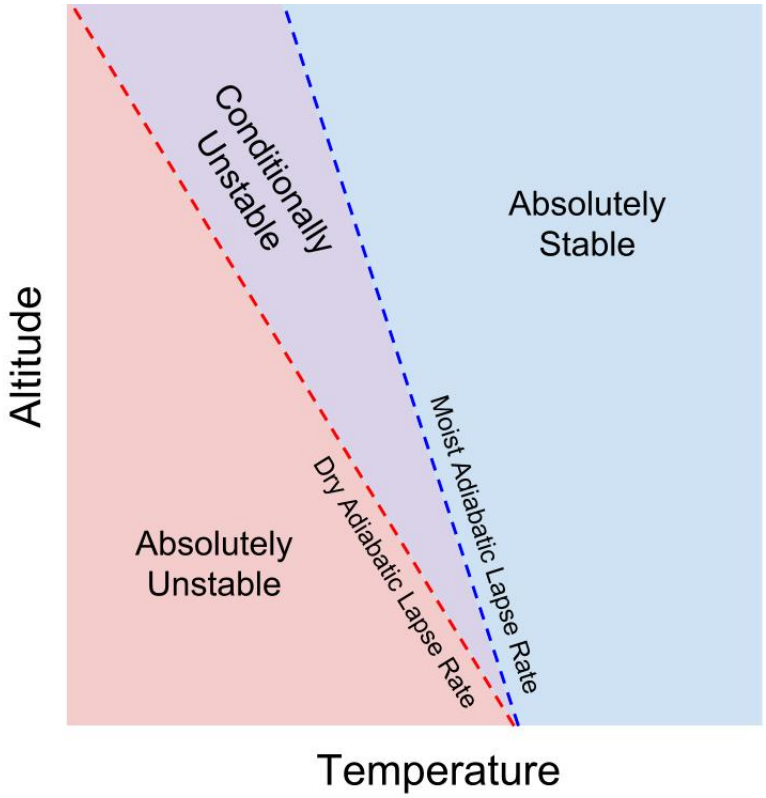

Figure 3: Illustration of stability criteria for moist air.

self, therefore having multiple high quality data series is very valuable to the analysis. Also, multiple stations covering a rather small area like in this case may highlight certain features in the IDF-curves related to topography or other mechanisms influencing precipitation that would otherwise have been hidden. Secondly, the Oslo area is a typical urban area, making it vulnerable to short duration extreme precipitation. Improving the knowledge on extreme precipitation events may potentially save the area and its population for large weather-related costs in the future. Another reason for selecting this area was the availability of processed data. Data from the selected stations had been pre-processed beforehand as part of a study on rainfall design values (Lutz et al. 2020).

2.3.1 Meteorology and climatology of the study area

Due to its complex topography, high latitude and long coastline, Norway experiences a range of different types of weather. Situated at the end of the North Atlantic extratropical cyclone track and in between warm humid air

from the South and cold Arctic air from the North, large frontal systems or storms often makes landfall in Norway (Azad et al. 2017). Alongside the coastline stretches mountains of 1500-2000 meters, enhancing precipitation trough orographic lifting. This makes a very wet coastal region, in contrast to the dry inland region towards the Swedish boarder in the East. In the south and along the Southern coast, the annual precipitation is small compared to locations in the west, but convective summer rainfall is often more intense. Thus, as Oslo is located in the South-East along the coast, the most prevalent cause of flash foods are convective summer precipitation. Although the annual precipitation in the Oslofjorden area is around 700 mm, this area has the largest design values for short-duration precipitation in Norway (Dyrrdal et al. 2019a).

3 Theory

In this section we present the model applied in the study and the theory it is based upon. This includes the general idea of Intensity-Duration-Frequency (IDF)-curves, practices in computation of IDF-values, the Generalized Extreme Value (GEV) distribution, the Bayesian approach and more. Parts of the theory approach here is inspired by the work of researcher Julia Lutz at the Norwegian Meteorological Institute, who's work is directly influencing the operational design-values in Norway.

3.1 The General Idea of IDF values

Intensity-Duration-Frequency (IDF) curves comprise an estimate of rainfall intensities of different durations and recurrence intervals. Durations commonly range from minutes to days and recurrence intervals often range from a year to a couple hundreds of years. The curves can either be calculated for a large region or for a single point. Depending on the topography and governing precipitation processes and patterns, the IDF values may differ quite a lot between locations only kilometres apart. *Duration* refers to the length of the precipitation event. Maximum rainfall intensity for each duration can be related to corresponding *recurrence intervals* or *return-periods*. The return period is defined as $T = 1/P$ where P is the annual exceedance probability. $P = 0.1$ (10%) implies that a precipitation event of a given magnitude has a return period T of ten years. It can also be interpreted as a 10% chance of exceeding the given magnitude in any given year. Thus, the higher return period T of an event, the less likely this event will occur during a given year. The corresponding Cumulative Distribution Frequency (CDF) F will be: $F = 1 - P = 1 - 1/T$. Once F is known the maximum rainfall intensity for each duration and return period is determined through the chosen PDF (e.g GEV, Gumbel) (Nhat et al. 2006). IDF curves can then be presented as precipitation from one duration at all return periods or as precipitation at one return period for all durations.

3.2 IDF calculation packages

R software and packages on extreme value statistics has been used to perform the IDF calculations 8*The R Project for Statistical Computing*). R is a language and environment for statistical computing and graphics, and the R software is available as Free Software under the terms of the Free Software Foundation's GNU General Public License (https://www.r-project.org/about.html). Functionalities in the R package *extRemes*, Weather and Climate Applica-

tions of Extreme Value Analysis, have been used for an extreme value analysis. The package allows for parameter-estimation on extreme-value distributions, implementation of different inference methods and more (Gilleland et al. 2016). The function *fevd* with arguments *type=GEV* and *method=Bayesian* fits the univariate general extreme value distribution to the precipitation data (*fevd: Fit An Extreme Value Distribution (EVD) to Data* 2021). The type-argument states which extreme value distribution to fit to data, while the method-argument states which type of estimation method should be used. The fit is done for desired quantiles, and return-levels for all durations are calculated.

The code is using annual maxima for each duration and station as input. The procedure of extracting annual maximum from raw preciptiation model output is explained in Section 5. A beta distribution is used to describe the prior probability distribution of the shape parameter. This is done to constrain the shape-parameter, as it was found to cause unrealistic IDF-values if not contained within a certain interval (Martins et al. 2000). Then while iterating over stations and durations the GEV estimation with Bayesian inference is called. Here the *fevd* R-package is used. *fevd* is fitting an extreme value distribution, in this case GEV, to data, in this case the annual maximum precipitation of the given duration and station. Within this calculation the quantiles are selected (here the 2.5 (bottom) and 97.5 (top) quantile), in this case making the 95% confidence interval. More detailed information on the GEV distribution and method can be found in Section 3.4 .

3.3 Practices in Computation of IDF values

Short-duration precipitation statistics is often presented with IDF-curves. IDF curves provide information on duration and frequency of pre-defined precipitation events, and they are often used in planning and design of infrastructure and other water-managing structures. There are many ways to calculate IDF values, thus different methods are used in different countries. In Sweden two of these have been applied when (Olsson et al. 2019) fitted the GEV distribution to block maxima and the Generalized Pareto distribution to Peak-Over-Threshold events (Hosking et al. 1987) to calculate regional short-duration rainfall. In Québec in Canada was an Bayesian approach with standard parameter estimation for the GEV distribution used (Huard et al. 2010). In this study the Bayesian approach was recommended for further studies as it implicitly include an estimation of the uncertainty of the IDF values. (Mohymont et al. 2004) used a GEV distribution and a Gumbel distribution to establish IDF-curves for the tropical Central Africa. Until recently the methodology for estimating IDF values in Norway was based on

fitting a Gumbel distribution to the N-highest measurements. (Lutz et al. 2020) explored ways to update the Norwegian IDF routine, fitting a General Extreme Value (GEV) distribution for annual maximum precipitation using a modified Maximum Likelihood estimation and a Bayesian inference. Here the Bayesian method performed best in two goodness-of-fit tests and thus was recommended for further IDF-estimation for Oslo in Norway.

One profound challenge in estimating IDF-values for extreme precipitation is the availability of data. Short data series yields very large uncertainties in precipitation magnitude for large return periods. To properly analyse IDF-curves and the potential impacts of the events they describe it is crucial to capture and understand the uncertainties of the curves. As recommended in (Huard et al. 2010) and practised in (Lutz et al. 2020) fitting an GEV distribution using a Bayesian method will be applied in this study for best estimates of the IDF-values and uncertainties.

3.3.1 Choice of durations and return-periods

In this study the following durations are used in the IDF-calculations: 30, 45, 60, 90, 120, 180, 360, 720 and 1440 minutes. The following return-periods are used in the IDF-calculations: 2, 5, 10, 20, 25, 50, 100 and 200 years.

Since the main focus of the study is on whether a high-resolution climate model can simulate convective-scale extreme precipitation, durations ≤ 6 hours are most important to cover. However, to identify differences in representation of convective precipitation to stratiform precipitation, stratiform-scale durations up to 1440 minutes (24 hours) are also included.

Due to large uncertainties in return-values for large return-periods the main interest for the study is return-periods below 25 years. However, return-periods up to 200 years are included due to their position in the current guidelines in Norway as recommended flood-design values for construction purposes (NVE 2011).

3.4 GEV distribution and extreme value statistics

Extreme value theory provides the statistical framework needed to infer probability of very rare or extreme events. The GEV distribution describes a family of three possible types of distributions of block maxima of a given variable, allowing a continuous range of possible distribution shapes. These distributions are called Type 1, Type 2 or Type 3, also known as Gumbel, Fréchet and Weibull respectively. The block maxima distribution converges to a GEV, G(x), distribution once the record length approaches infinity. Being a three-parameter distribution, G(x) can be written on the form:

$$G(x) = exp\Bigg\{ -\left[1 + \zeta(\frac{x-\mu}{\sigma})\right]^{\frac{-1}{\zeta}} \Bigg\} \tag{20}$$

for

$$1 + \zeta(\frac{x-\mu}{\sigma} > 0 \tag{21}$$

where μ is location-, σ is scale- and ζ is shape-parameter (Dyrrdal et al. 2015). Since the extremes are determined from the tail of the distribution, they are heavily affected by the choice of ζ. ζ determines whether the distribution converges towards Gumbel (ζ=0) with a light upper tail, Fréchet (ζ>0) with a heavy upper tail or Weibull (ζ<0) which is bounded from above (Dyrrdal et al. 2014). As the shape parameter describes the tail of the distribution it will severely affect the estimates for long return periods. The shape parameter can be both positive and negative, depending on the location of interest. Poor choice of this parameter may provide unrealistic precipitation estimates, especially on the longer return periods, thus (Lutz et al. 2020) introduced a prior distribution to constrain the ζ parameter of the GEV distribution.

3.5 Bayesian Approach

3.5.1 Basics

In general the goal is to estimate parameters of a distribution to best fit data. Given a generic parameter β, the likelihood function is the probability density of the observed data as a function of β. βs with high likelihood correspond to models which give high probability to the observed data. Traditionally, a maximum likelihood estimation seeks to adopt the model with greatest likelihood, namely the model that assigns the highest probability to the observed data. Bayesian inference provide an alternative method to draw inference from such a likelihood function.

Here a Bayesian inference is used to estimate the probability distribution for the parameter set α containing the three GEV parameters, $\alpha = (\sigma, \zeta, \mu)$. These parameters are treated as random variables with prior distributions, distributions of the parameter prior to the inclusion of data $x = (x_1, x_2, ..., x_n)$. Bayes‘ Theorem (Joyce 2003) states that the probability of an event is dependent on prior knowledge of conditions that are relevant to the event:

$$P(\alpha|x) = \frac{L(x|\alpha)P(\alpha)}{P(x)}, \tag{22}$$

where $P(\alpha|x)$ is the probability density function of α given the observations x. $L(x|\alpha)$ is the likelihood function and $P(\alpha)$ is the GEV parameters. Since $P(x)$ is constant, Eq. (22) can be written as

$$P(\alpha|x) \propto L(x|\alpha)P(\alpha) \tag{23}$$

or

$$posterior \propto prior * likelihood \tag{24}$$

where the *posterior* is the distribution of α after the inclusion of data. As done in (Lutz et al. 2020), $P(\alpha|x)$ is sampled using the Markov Chain Monte Carlo (MCMC) method with 50 000 iterations, where the last 3000 is used to have stability in the simulated parameters. Furthermore, the posterior distribution allows for direct derivation of quantiles. Here the 95% credible interval is used.

3.5.2 Shortcomings of the GEV model

One major shortcoming of the GEV model used in this study is the assumed stationarity, which is usually not accurate for climate data. Long-term trends are often present in data-series and should always be accounted for. However, since robust trends are not easily detected for short time-series as used in this study the stationary GEV is used.

3.5.3 Advantages of the Bayesian inference

Choosing a Bayesian analysis of extremes over a more traditional likelihood approach can have various advantages. Incorporating additional sources of information to the block maxima, like a prior distribution, to the analysis is considered a large advantage given the scarce nature of the extremes. Another major advantage is that the variance of the posterior distribution can be used to calculate the precision of the inference. This way the resulting IDF-values can be presented with desired confidence levels.

4 Data

4.1 Measurements

Precipitation measurements used in this project is obtained from pluviometers operated by the Norwegian Water Resources and Energy Directorate (NVE) or the respective municipalities in corporation with MET Norway. Since the late 1990s and early 2000s the operating pluviometer has been the Lambrecht 1518H3 tipping bucket pluviometer manufactured by the German company Lambrecht meteo GmbH (Lutz et al. 2020). The Lambrecht pluviometer has a measuring range of 0.1 mm precipitation at time resolution of 1 minute and a given accuracy of $\pm$ 2%. Intensity correction is done to account for loss of rain due to the time required for the bucket to tip. MET Norway supervised the installation of the pluviometers and ensured installation according to the recommendations from the World Meteorological Organization (WMO). Additionally MET Norway performed quality control and storage on all data from the pluviometers. Before the now operational Lambrecht some of the stations operated with the Norwegian produced pluviograph Plumatic, manufactured by Kongsberg Våpenfabrikk A/S. It was replaced partially due to is lack of heating, making it operational only in the extended summer months, from mid-April to mid-October. The Lambrecht also suffers from poorer data quality during winter due to snowcaps or ice-slush obstruction spite being heated.

The proposed method in this study requires annual maxima for each duration as input, thus a requirement for season completeness is necessary. The requirement where here set to at least 80% of the days throughout the season covered and of good quality. Hence the precipitation series extracted where shorter than the total operational period for all stations. The number of years with sufficient data for each station is found in Table 2. (Lutz et al. 2020) analysed monthly precipitation for 1, 2 and 3 hours in two locations in Oslo and found that the highest occurrence of short-duration, high-intensity precipitation was during the summer months. In combination with lack of high-quality data during the winter period, especially from before the 1990s when the Plumatic pluviometer were still in use, makes the extended summer period, 1st of May to 30th of September, best suited for the IDF analysis in this study. Furthermore, time series of 10 years is here considered to be an absolute minimum for calculation of IDF-values. The twelve stations listed in Table 2 are the ones left meeting these criteria in the municipality of Oslo. The location of the stations are pictured in Figure 4.

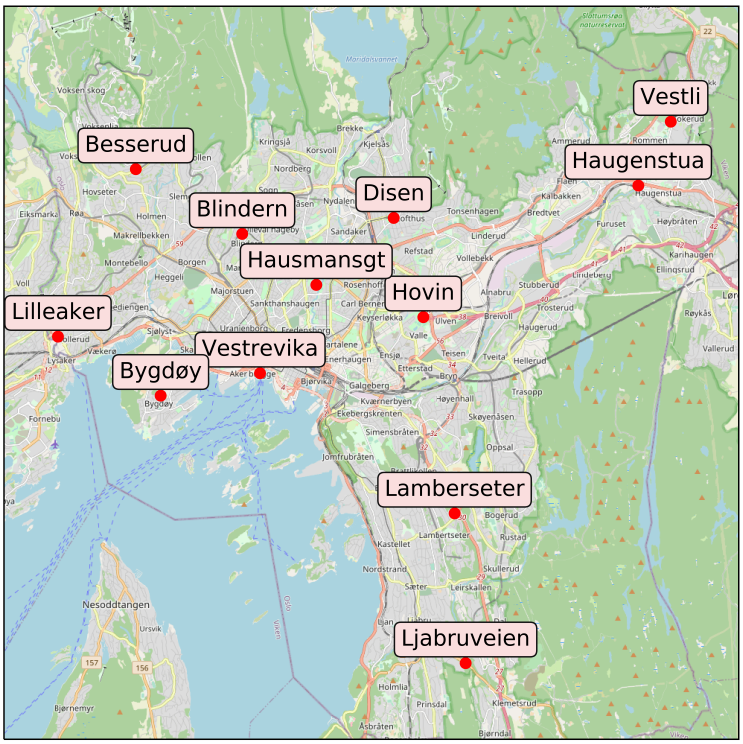

Figure 4: Map of the city of Oslo and the innermost part of Oslofjorden. Red dots are marking all stations in the Oslo area used in the study. The map is oriented North - upwards. The total North-South distance of the figure is 17km, and the total West-East distance is 19km.

4.2 HARMONIE-AROME and data

For this study precipitation output from the HCLIM38 model run has been used. The atmospheric model HARMONIE-AROME (Bengtsson et al. 2017) uses the atmospheric physics from Applications of Research to Operations at Mesoscale (AROME) (Seity et al. 2011) model. This model is developed by Météo-France, and is designed for convection-permitting scales and non-hydrostatic dynamics. The spatial coverage of the model is illustrated as HCLIM38-AROME in Figure 5. HARMONIE-AROME is used for operational high-resolution numerical weather prediction (NWP) in the coop-

Table 2: Station name, station number, available years for annual maximum and operational years of the twelve Oslo stations used in this study.

Station Name	**Station Nr**	**Years AM**	**Operational From-To**
Ljabruveien	17980	17	01.01.1985-d.d.
Lamberseter	18020	25	15.05.1999-d.d.
Hovin	18210	17	15.01.1999-d.d.
Haugenstua	18269	15	01.01.2000-d.d.
Vestli	18270	32	18.04.1974-d.d.
Hausmannsgt	18320	20	21.06.1984-04.11.2013
Disen	18420	20	02.06.1998-d.d.
Vestre Vika	18640	13	22.05.1974-03.10.1998
Blindern PLU	18701	48	16.04.1968-d.d.
Bygdøy	18815	16	01.01.2000-d.d.
Besserud	18920	13	29.09.1998-d.d.
Lilleaker	18980	13	01.01.2000-d.d.

erative effort named Meteorological Cooperation on Operational Numerical Weather Prediction (MetCoOp) between the Norwegian Meteorological Institute (MET-Norway), the Swedish Meteorological and Hydrological Institute (SMHI) and the Finnish Meteorological Institute (FMI) (Müller et al. 2017). Recently HARMONIE-AROME has been used in a regional climate-configuration called HARMONIE-Climate cycle 38 (HCLIM38) for long-term climate simulations (Lind et al. 2020; Lind 2021 in prep.) at 3 km horizontal resolution and with 65 vertical layers. Being one of the first long-term climate simulations on regional convection-permitting (<4km) scales with explicit deep convection for the Scandinavian region (Lindstedt et al. 2015; Lind et al. 2016), this data-set (Lind et al. 2020; Lind 2021 in prep.) provides added opportunities for analysis of extreme precipitation. The temporal resolution of the data used in this study is 15 minutes, covering two periods of time, 1985-2005 and 2080-2099. A comprehensive description of the model system is presented in (Belušić et al. 2020).

The model dynamics of HARMONIE-AROME is based on the fully compressible Euler equations (Roe 1986). At 3km resolution deep convection is expected to be resolved explicitly, hence the model has no deep convection parameterization. Shallow convection on the other hand is parameterized though a mass-flux framework consisting of updrafts and thus transport of heat, momentum and moisture (Bengtsson et al. 2017). Eddy diffusivity is parameterized through a turbulence scheme called HARATU. The surface physics is simulated by the SURFEX surface scheme, which simulates surface and subsurface processes such as energy fluxes and transport of water in the soil, but also how these fluxes interact with the atmosphere (Bengtsson et al. 2017). Surface topography is is based on Global Multi-resolution Terrain Elevation Data (GMTED2010) (Danielson et al. 2011).

The climate setting of HARMONIE-AROME, HCLIM38, proved strong improvement on representation of precipitation compared to similar model-setups with coarser resolution models in (Lind et al. 2020). Most evident was the improvement in reduction of "drizzle" and increased occurrence of high intensity precipitation events in addition to better timing and amplitude of the diurnal cycle. The simulations was conducted within the Nordic Convection Permitting Climate Projections project (NorCP), which is one of the leading projects on increasing knowledge of climate changes and processes over the Fenno-Scandinavian region. The boundary state of the HCLIM38 climate projections is obtained from the global Earth System Model (ESM) EC-Earth (Lind et al. 2020).

The EC-Earth model system is developed based on the seasonal forecast system of the European Centre for Medium-Range Weather Forecasts (ECMWF). It simulates all relevant parts of of the Earth system, including physical, chemical and biological processes. Whereas a typical general circulation model simulates atmospheric and oceanic components, an ESM also includes information on a global carbon cycle, dynamic vegetation, ocean bio-geo-chemistry, atmospheric chemistry and continental ice sheets (*Earth System Modeling, a definition*). Being a coupled model, feedback cycles are also modelled, allowing for gained information on how the climate system is reacting to certain changes like deforestation or reduced surface albedo (Heavens et al. 2013). This improves the complexity and hence the overall representation of physical processes within the system.

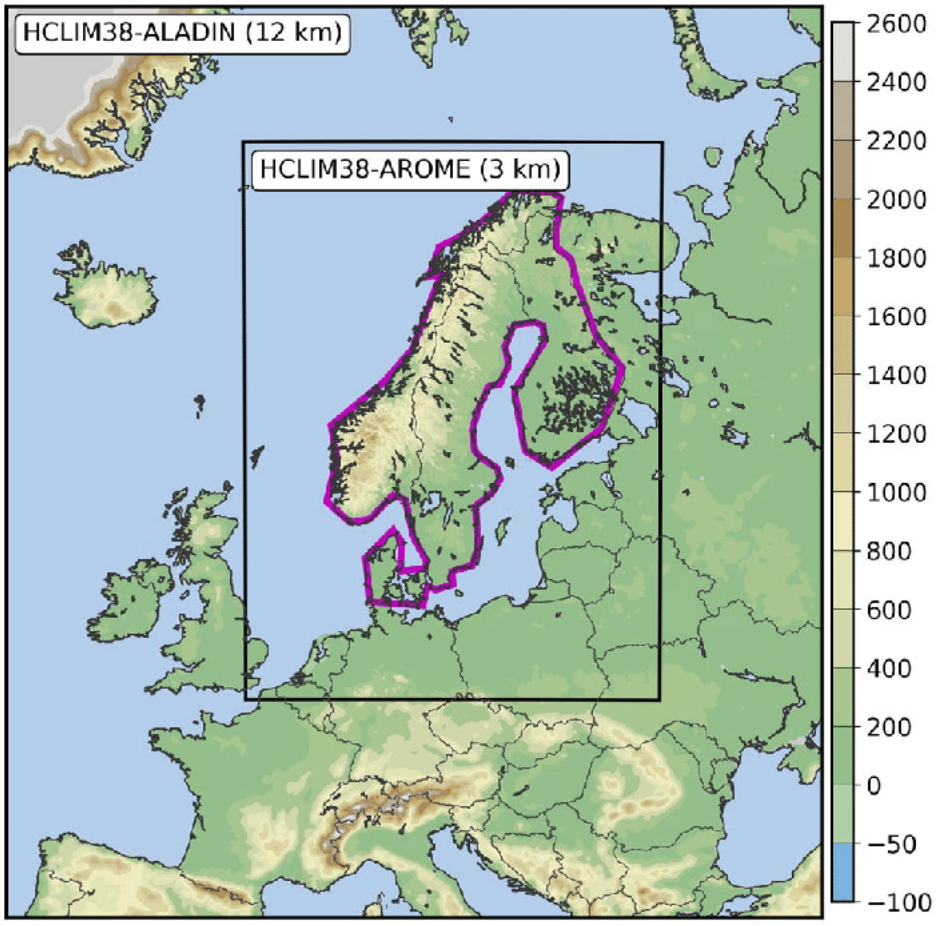

Figure 5: Domain used for HCLIM38-ALADIN (12km) simulation and domain used for HCLIM38-AROME (3km) simulation in the inner black rectangle. The colorscale represents the altitude above mean sea level in meters, and the magenta colored area defines the Fenno-Scandinavian region (Lind et al. 2020).

4.3 AROME: difficulties and added value with convection-permitting forecasting, an example

(Müller et al. 2017) highlights the challenge of precise precipitation forecasting of convective cells. Even though high resolution forecasting systems at times still struggles with convective-scale events, they also highlight the added value of such systems through better representation of advection and hence improved localization of precipitating events compared to lower resolution model systems. This was exemplified through a convective summer event in southern Sweden. Here the event forecast from the high resolution AROME-MetCoOp and the lower, 12km resolution ECMWF-IFS model was

investigated. In Figure 6 the event forecast from the two systems are presented. As visible from panel (a) and (b) the AROME-MetCoOp-forecast predicted 67mm $(24h)^{-1}$ while ECMWF-IFS predicted 55mm $(24h)^{-1}$ precipitation magnitude in the south Sweden area (black square in panel (a)). Indicated by the gray and black boxes in panel (c) and (d) many rain gauges measured 40-100 mm $(24h)^{-1}$ and some measured >100mm $(24h)^{-1}$. In general all maxima of the total domain in panel (a) and (b) are larger for AROME-MetCoOp compared to ECMWF-IFS. Even though the forecast by AROME-MetCoOp was in better agreement with observations on the magnitude compared to ECMWF-IFS, it was still underestimating the observed maxima within the black box of panel (a). The AROME-MetCoOp-forecast was also better collocated with the measured event compared to ECMWF-IFS. As seen in panel (d) the precipitation was mislocated compared to the measurements for the ECMWF-IFS forecast. Here most precipitation was expected along the coastline, whereas the AROME-MetCoOp precipitation was located further inland south of Lake Vänern. As the small boxes in panel (c) and (d) show, the latter was a better fit to the observations. Panel (e) show a radar precipitation product which also collocates very well with the observations and AROME-MetCoOp.

Precipitation parametrization and advection is probably the leading causes for the differences between the ECMWF and AROME forecasts in this case. While deep convection is parameterized in ECMWF, it is resolved explicitly in AROME. In ECMWF this implies that once deep convection is initialized precipitation will form and precipitate immediately. Hence no moisture is advected inland, leading to a mislocated precipitation event. In AROME the moist air is allowed to be advected into another region, precipitating in a different location to that where the convection was initialized. The advantage on spatial accuracy this makes for the high resolution system over the lower resolution system is evident, and in a case like this the AROME-MetCoOp has a large added value compared to ECMWF-IFS when it comes to warning issues and other precautions prior to the event.

However, high resolution forecasts do not always capture convective events like the one in southern Sweden. As with deep convection for the ECMWF-IFS system, shallow convection is not explicitly resolved in HARMONIE-AROME but instead parameterized. Aspects like soil moisture and dew point temperature may also be imperfect, potentially causing a poor forecast for an extreme precipitation event. Here a case study illustrating these challenges is presented.

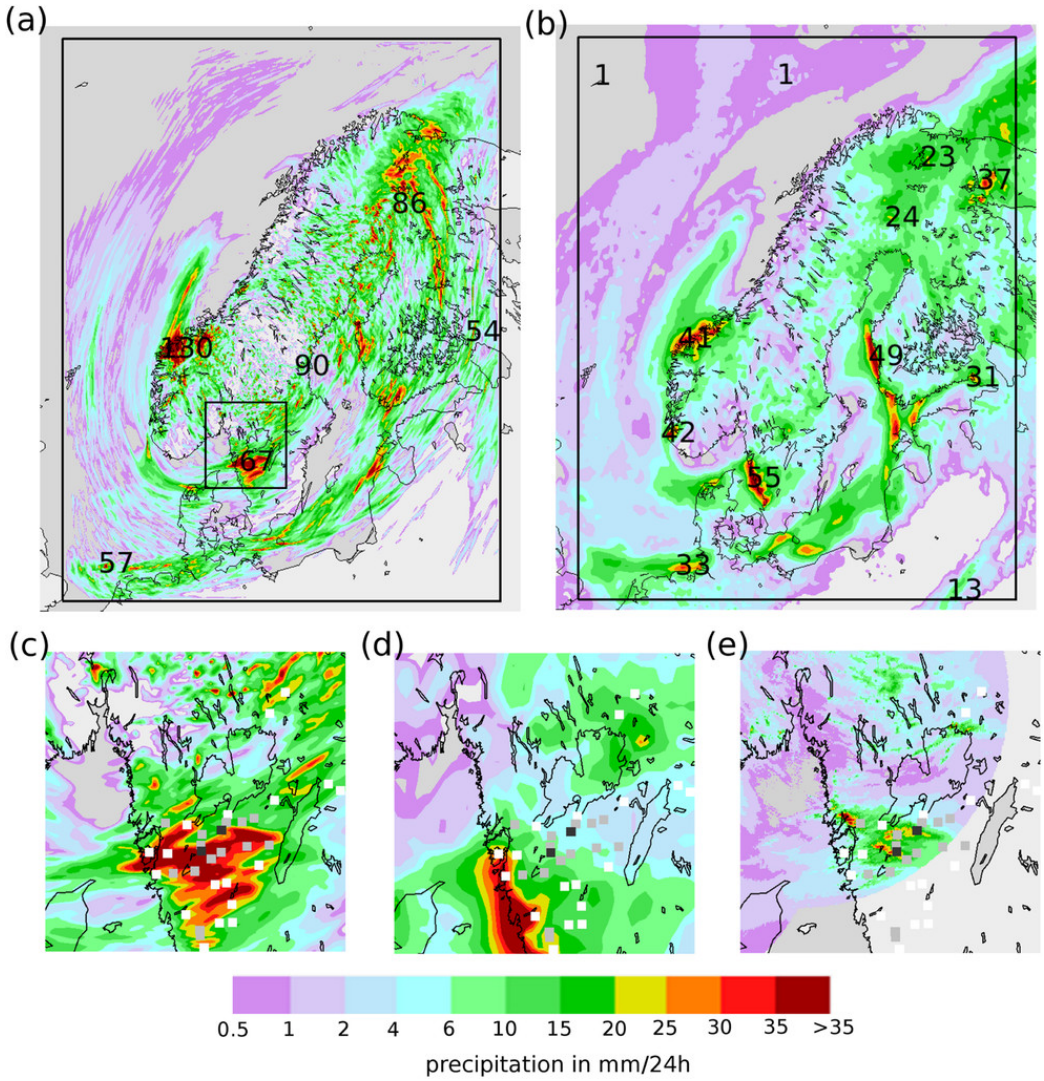

Figure 6: Figure from (Müller et al. 2017). The 24-h accumulated precipitation (mm) at 0600 UTC 20 Aug 2014 from (a) AROME-MetCoOp and (b) ECMWF-IFS forecasts initialized at 0000 UTC 19 Aug 2014. Black numbers indicate the maxima of the simulated precipitation. (c),(d) As in (a),(b), but for a small area in the southern part of the domain (southern Sweden). (e) The radar precipitation product from the Norwegian radar located at Hurum. The squares in (c), (d), and (e) highlight the measurements from the Swedish rain gauge network and are color coded as follows: white, 20–40; gray, 40–100; and black, 100 mm (24 h)1. Citation: Weather and Forecasting 32, 2; 10.1175/WAF-D-16-0099.1

4.3.1 Bærum Case Study

7th of August 2019 15:00-18:00 UTC a heavy rainfall event occurred in Bærum, Norway. According to a METinfo report (Bork et al. 2020) the event caused local damage, clogged drainage systems and numerous flooded basements. The forecast prior to the event revealed unstable air masses with strong vertical movement throughout southern Norway. Besides an elevated CAPE-index north-west of where the event actually occurred there where little indication for such an event in the Bærum-area. During the event several records for short-duration precipitation was set at the Gjettum station in Bærum: 36,5mm($30min^{-1}$), 47,9mm($1h^{-1}$) and 61,3mm($3h^{-1}$). These recordings have return periods of >200 years, >100 years and > 50 years, respectively. The 24 hour accumulated precipitation forecast is produced as a 10-member ensemble mean. The ensemble mean forecast displayed some precipitation in the area west of Bærum with magnitude of around 2 mm/24h.

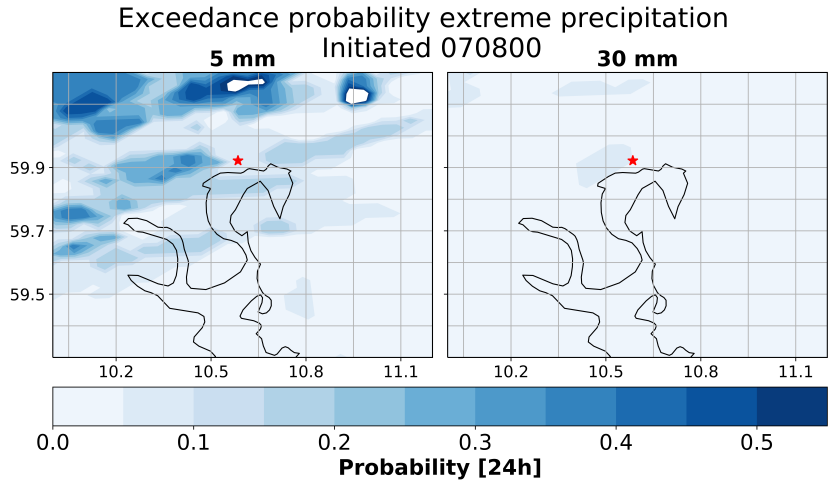

Figure 7: Exceedance probability plot from AROME 24 hour precipitation forecast initiated 7th of August 2019 at 00:00 of the Oslofjord area. On the left panel is the 5 mm exceedance probability, on the right is the 30 mm exceedance probability. The red star marks the Bærum location and the black contour is Oslofjorden.

Even though the ensemble mean forecast does not reveal any major event, the individual members may provide some information on the size and location of an possible event. All members are equally a realization of the model, which together highlight the uncertainty of the precipitation forecasting within AROME. While a single ensemble member does not make a huge impact on the complete forecast, the different members can provide information on the probability of extreme events. This is exemplified in Fig-

ure 7 where the exceedance probability of 5mm($24h^{-1}$) and 30mm($24h^{-1}$) are plotted. As Figure 7 clearly displays the probability of precipitation of 5mm($24h^{-1}$) was above 50% for some areas surrounding Oslofjorden, but in Bærum the 5mm($24h^{-1}$) probability was approximately 10%. The 30mm($24h^{-1}$) probability due west of the Bærum location was also around 5-10%, while at the actual location it was close to 0%. Despite being tiny and probably only suggested by one of the ensemble members, this could imply some probability of an extreme event in the area.

In Figure 8 six of the forecast ensemble members are displayed for the sake of illustration. The immediate impression is that there are little suggestions of an extreme event in the area. Some patches of precipitation magnitudes 15-20mm can be observed some kilometers away from Bærum, others several miles away. In ensemble member two it seems to be hardly any forecasted precipitation close to the Oslofjorden area at all. However, ensemble member four seems to capture the localization quite well, only missing with a few kilometers. The magnitude of the event is also captured quite well in this member with a maximum value of around 45-50mm($24h^{-1}$). While having a very different intensity from the observed at 61.3mm($3h^{-1}$) this is still a fairly large event for the area. Shorter-duration forecasts initiated hours before the observed event did not show intensities anywhere close to the observed event either. Although the event in Bærum was not forecasted, in the sense that no danger warning was issued, investigating the individual ensemble members show that the model system to some extent is capable of capturing these extreme events. That being said, none of the ensemble members are more "true" than the others.

Ensemble member disagreement makes decision making based on this forecast difficult, hence no warning was issued in the area prior to the event. As mentioned in Section 2.2.1 deep convection is expected to be resolved explicitly, while shallow convection is parameterized. In this case either the parameterization failed or the scales in action was to small to be resolved by the model. Even though progress is made within the modelling system to improve convective precipitation forecast, it is evident that new approaches to decision making needs to be developed to better forecast small-scale convective events.

This case study serves as an example of how different "realities" within the model predicts different magnitude and location of convective precipitation. It is not possible to indicate whether one ensemble member is more likely to match observations than others. This behaviour stresses the fact that it is no need for perfect temporal overlap between the observed and the modelled precipitation in a climate perspective.

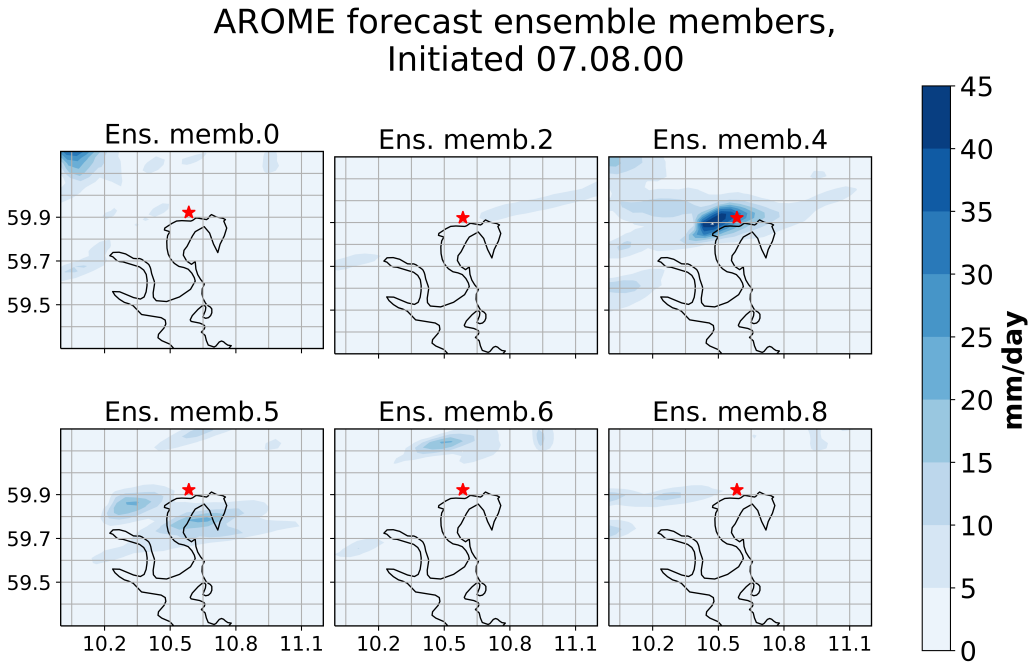

Figure 8: Six ensemble members from AROME forecast initiated 7th of August 2019 at 0:00 for the Oslofjorden area. The colors of the plot correspond to 24 hour precipitation magnitude, where darker colors are larger values and lighter colors are smaller values. The red star marks the Bærum location. The black lines are representing the coastline in Oslofjorden. The numbers on the x- and y-axis is longitude and latitude respectively.

5 Method

There are several ways to extract annual maxima from the precipitation data. In the following paragraphs the different methods used in this study are presented.

5.1 Annual Maxima

The block maxima used in the IDF calculations are annual maximum precipitation for durations ranging from 15 minutes to 24 hours. Since observed data from stations is typically used for the GEV calculation the block maxima represent a single point in space. But in terms of extreme precipitation, what is this single-point measurement really representing? Is it representative for a larger area or only for a specific gridpoint? Convective events during summer like the one presented in Section 4.3.1 are extremely localized. Showers like this might pass only a couple of kilometres away from a station, but still

cause severe damage to the downstream area in which the station is located. Using gridded precipitation values from the HARMONIE-AROME model allows for extraction of annual maximum precipitation not only from the gridpoint closest to the station, but also from multiple gridpoints surrounding the station. The idea is that including data from gridpoints around the one closest to the station may result in an analysis representative for a larger area compared to a single point. As a station potentially misses events by the smallest margin it is also challenging to know whether modelled precipitation at a gridpoint is under- or over-estimated compared to the observed annual maxima. Here follows the four different methods used to compute the annual maximum precipitation from the 20-year data series provided by the climate setting of the HARMONIE-AROME forecasting model. Annual maximum are extracted for each duration, making a data-series with maximum values with length **n** equal to number of years available (20). The raw data has temporal resolution of 15 minutes. The overall goal with these different methods is to find the method that are most consistent with the observed annual maxima and IDF-values, and to determine what the "true" extreme precipitation state of the area actually is.

1. 1GRID, hereafter named "1GRID". Select gridpoint closest to the respective station. A sliding filter for each duration is applied, summing data-points to the given duration length and creating a new dataset. From this time-series the annual maximum is extracted for each duration. Illustrating in Figure 9.

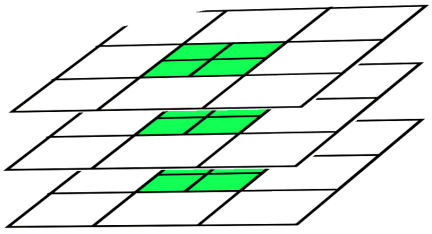

Figure 9: Illustration of 1GRID selection method for 1 hour duration (4*15min). One of the "sheets" represent one timestep for a spatial grid of 3 by 3 grid-cells. The green square represent the gridpoint closest to the station. The smaller squares within the green square represent the 4 timesteps used for the 1 hour precipitation value.

2. 9GRID, hereafter named "9GRID". Select the gridpoint closest to the respective station and the neighbouring gridpoints in a 3x3 matrix with the centre-gridpoint being the one closest to the station. A sliding

filter for each duration is applied, summing data-points in time to the given duration length and creating a new dataset. The filter selects the maximum value out of the 9 gridpoints for each timestep. The resulting annual maximum value for a given year and duration may consist of precipitation values from different gridpoints. Illustration in Figure 10.

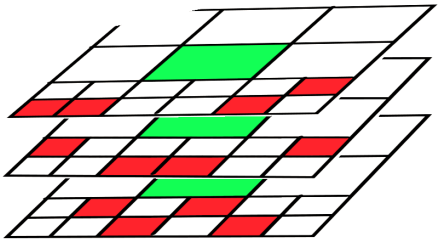

Figure 10: Illustration of 9GRID selection method for 1 hour duration (4*15min). One of the "sheets" represent one timestep for a spatial grid of 3 by 3 grid-cells. The green square represent the gridpoint closest to the station. The smaller red squares represent the 4 data-points used for the 1 hour precipitation value.

3. 9MEAN, hereafter named "9MEAN". Select the gridpoint closest to the respective station and the neighbouring gridpoints in a 3x3 matrix with the centre-gridpoint being the one closest to the station. A sliding filter for each duration is applied, summing data-points to the given duration length and creating a new dataset. The filter selects the maximum value for each of the 9 gridpoints, making 9 time-series with annual maximum values for each duration. The mean annual maximum of the 9 is then calculated for each duration.

4. 9MAX, hereafter named "9MAX". Select the gridpoint closest to the respective station and the neighbouring gridpoints in a 3x3 matrix with the centre-gridpoint being the one closest to the station. A sliding filter for each duration is applied, summing data-points to the given duration length and creating a new dataset. The filter selects the maximum value for each of the 9 gridpoints, making 9 time-series with annual maximum values for each duration. The max annual maximum of the 9 is then calculated. Illustration in Figure 11.

5.1.1 Observations

Observations are hereafter named "OBS". Similarly to the modelled data, the precipitation gauge measurements from the Oslo area presented in Section 4 are represented through annual maximum precipitation for the dif-

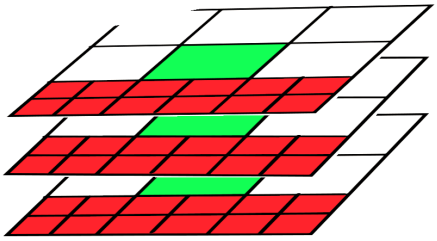

Figure 11: Common for both MAX and MEAN. Illustration of 9GRID selection method for 1 hour duration (4*15min). One of the "sheets" represent one timestep for a spatial grid of 3 by 3 grid-cells. The green square represent the gridpoint closest to the station. The smaller red squares within a large square represent the 4 data-points used for the 1 hour precipitation value. For the sake of readability not all large squares are colored red.

ferent durations. A sliding filter for each duration is applied to the raw data, summing data-points to the given duration length and creating a new time-series. From this time-series the annual maximum is extracted for each duration. The OBS annual maxima is extracted with the same procedure as 1GRID, except that OBS is using the actual observation and 1GRID is using the modelled grid-cell value closest to the observation.

5.2 Annual Maximum Method Intention

1GRID is a plain selection method compared to the other methods. The initial reason for using other methods was that 1GRID was underestimating the resulting IDF-curves compared to the station-based curves for many stations and durations. 9GRID is designed to be an optimal selection method in the sense that the absolute maximum values possible for the area as an entity is chosen for each duration. This was to ensure that any event near the gridpoint in question also was captured. 9GRID is artificial compared to the others because a maximum value for any given duration may consist of precipitation values from different gridpoints, where the next maximum value may consist of entirely different gridpoint values. Whereas in 9GRID the annual maximum value for a given duration and year can originate from several gridpoints, in the 9MAX method it originates from one gridpoint. Thus, 9MAX should be more consistent with OBS, serving like a in-between solution to the OBS IDF-values and the 9GRID IDF-values. 9MEAN is a smoother version of the 9MAX. Since the mean annual maximum value out of the nine gridpoints is used, 9MEAN should ensure an improved area-representation compared to 1GRID. However, as Figure 28 displays, the IDF-value differ-

ence between 1GRID and 9MEAN is very small both for small and large return periods.

5.2.1 Statistics in a single point

It is important to keep in mind that measured preciptiation at a meteorological station or modelled preciptiation at a single grid-cell only represent the preciptiation climate at this single point in space for the measured or modeled period of time. Whether a small-scale precipitation event hit a station or not is arbitrary, partially determined by local conditions. Even though a 200-year event has a small probability of occurring at a single point in space, this event size has a much larger probability of occurring in a larger are. Under this reasoning it is strictly speaking impossible to say for sure what the preciptiation state of a station looks like based on surrounding stations. However, the likelihood of them being similar is very large, and in any given situation it is reasonable to assume they are similar.

5.2.2 Issues with time-series length

As with any other extreme value problem, the data-series length have a major impact on the resulting statistics. The problem arises when a time-series of a given length is used to infer statistics for return-periods longer than the time-series itself. Given a time-series of ten years, it is unlikely that a maximum value for a return-period of 100 years is captured in that time-series. The issue of what return-periods could be represented by your data-series depends a lot on what type of phenomena you are describing. In this case we are concerned with precipitation, and there are some limits to how large and rare an event could be. A time-series of 10-15 years could probably represent a return-period of 20-40 years, while a time-series of 100 years could probably represent a return-period of 500 years or more.

5.3 Metrics and Definitions

Here some metrics and abbreviations used in the results are described.

We define the standard deviation (STD) for the annual maxima (AM) as

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{25}$$

where x is annual maxima.

Station average observed annual maxima is defined as

$$AM^{OBS} = \frac{1}{y}\sum_{i=1}^{y} AM_t \tag{26}$$

where y is the number of stations with annual maximum value for a given year, t denotes the *duration* for which the annual maximum values are extracted. For the different annual maximum methods explained in Section 5.1 the equivalent metrics is written AM^n, where n is the methods *MAX*, *MEAN*, *1GRID* and *9GRID*.

Standardized (z-score) annual maximum values values are calculated using

$$z = \frac{x_i - \mu}{\sigma} \tag{27}$$

where

$$\mu = \frac{1}{N}\sum_{i=1}^{N}(x_i) \tag{28}$$

is the station average annual maximum for a given duration, similar to (26) but for all stations N. σ is standard deviation from 25. x_i in 27 is the individual annual maximum value for a given station.

Furthermore, precipitation intensity is calculated through

$$I = \frac{P}{T} * h \tag{29}$$

where P is precipitation in mm, T is duration in minutes and h is 60 minutes.

6 Results

In the following section the annual maximum time-series and the resulting IDF-values from the observational data and the modelled data are analysed. From Section 5.3 we remember the metrics: Station-average annual maximum from observed time-series: AM^{OBS}. Station-average annual maximum from modelled time-series: AM^n where n are the different methods MAX, MEAN, 9GRID and 1GRID as defined in Section 5.1. AM^n is sometimes written as an umbrella expression for all AM^n. Furthermore, as part of the analysis the precipitation intensity is calculated using using intensity, I, as in Eq. 29. Throughout the analysis the term *return-period* is referring to the recurrence interval of a given event. *Return-value* or *IDF-value* is the expected precipitation magnitude in mm at a given return-period.

6.1 Annual Maximum Precipitation

Investigating how the annual maxima (AM) time-series of the observations and the model behave provides information on the IDF-values and their uncertainties. Figure 12 shows AM for all stations for 15 minutes and 24 hours duration. The black line is the AM^{OBS} precipitation value. As illustrated by Figure 12 the AM^{OBS} have large variability throughout the 1970-2020 period. Some years have large AM^{OBS} while others have very small. Analysis show equivalent variability of AM^{OBS} for the other durations as well (not shown). Despite the close proximity, the range of AM values from individual stations for a given year can be large. This applies for all durations (not shown). As an example, being 7 km apart, in 2019 the 15 minute duration AM value for station 18815 Bygdøy is 5.3mm while for station 18210 Hovin it is 26.2mm. However, for years like 2003 and 2004 the opposite occurs, where the AM value varies little between the stations. This also applies to other years, but often only two to three AM observations exist for those years. From 1968 to 2000 there are mostly two to three stations available each year. In this period these two to three stations makes up the AM^{OBS} for all durations. Thus, the resulting IDF-values for a given station is not necessarily based on the same period of years as the other stations. Station 18640 Vestre Vika was operational from 1977 to 1988, whereas stations like 17980 Ljabruvegen or 18115 Bygdøy was operational from 2000 to d.d. As a result, the OBS return-values of one station might be influenced by natural variability in precipitation more than others due to the OBS AM values' temporal origin. From around year 2000 till 2019 the number of available stations per year increases. Since the AM values for a given year varies between stations, the spread is in generally large these years. Some stations have largest AM

value at coincident years for some or all durations. However, this is more not than often the case. For the model period 1985-2005 the number of available stations varies from three to four in the first half to six to seven in the latter half.

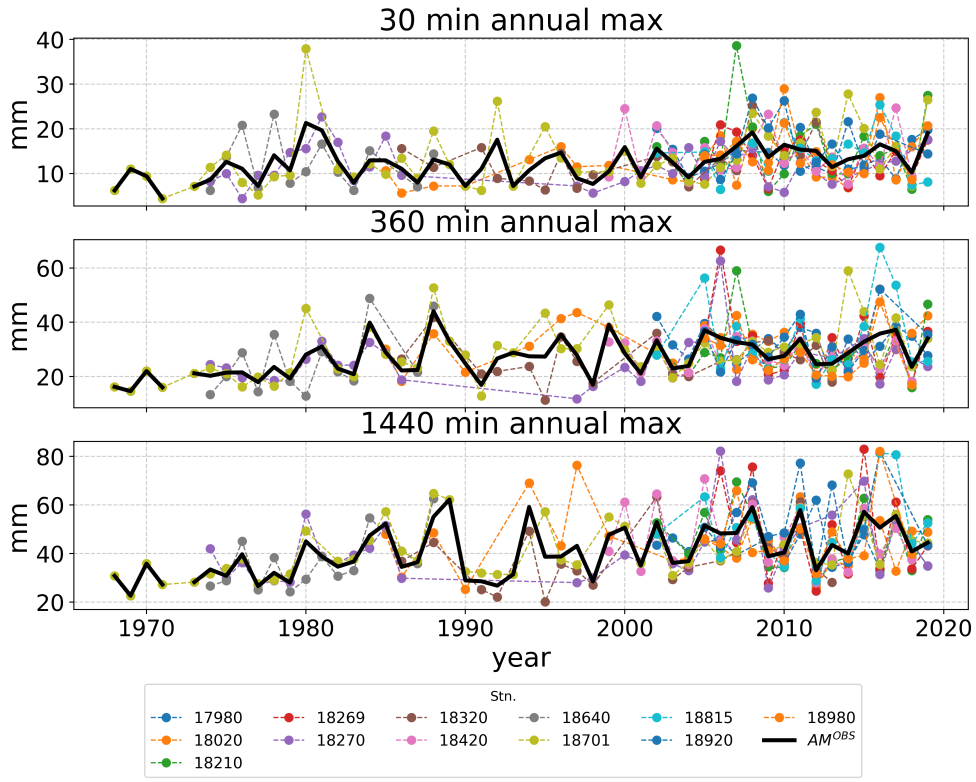

Figure 12: Annual maximum precipitation values for 15min and 24 hours duration. The individual colored dots are the individual stations. The black line are the stations mean values at the given year. Units are mm on the y-axis and years on the x-axis.

Now we investigate how the AM^{MODEL} compares to AM^{OBS}. In Figure 13 the AM^{OBS} from Figure 12 is plotted with AM^{n} from the different AM methods for durations 15, 60, 360 and 1440 minutes from 1985 to 2005. The shaded green area is one AM^{OBS} standard deviation, σ, for the respective durations. In general the four AM methods for modelled precipitation overlap quite well with the OBS range of values. An exception is that with increasing duration 9GRID appears to increasingly overestimate the OBS values. For selected years and durations the AM^{9GRID} is twice the size of AM^{OBS}, and in general it appears severely overestimated for most years on all durations larger than 360 minutes. For all durations, AM^{1GRID} and AM^{9MEAN}

are close to identical throughout the 20 years, while AM^{MAX} falls between AM^{1GRID} or AM^{MEAN} and AM^{9GRID}. In periods like 1988-1992 and 1995-2000 for most durations up to 360 minutes AM^n and AM^{OBS} agree quite well. For almost all years and durations one or more of the AM^n are within the standard deviation of the measurements. However a few exceptions exists, like the year 2001 where all modelled AM^n are outside the AM^{OBS} σ for durations up to 360 minutes. Here we keep in mind that a modelled AM value at year 1995 is not actually representing the observed value in that year. Thus, the overall modelled AM value range seems to be representative for the observed AM values.

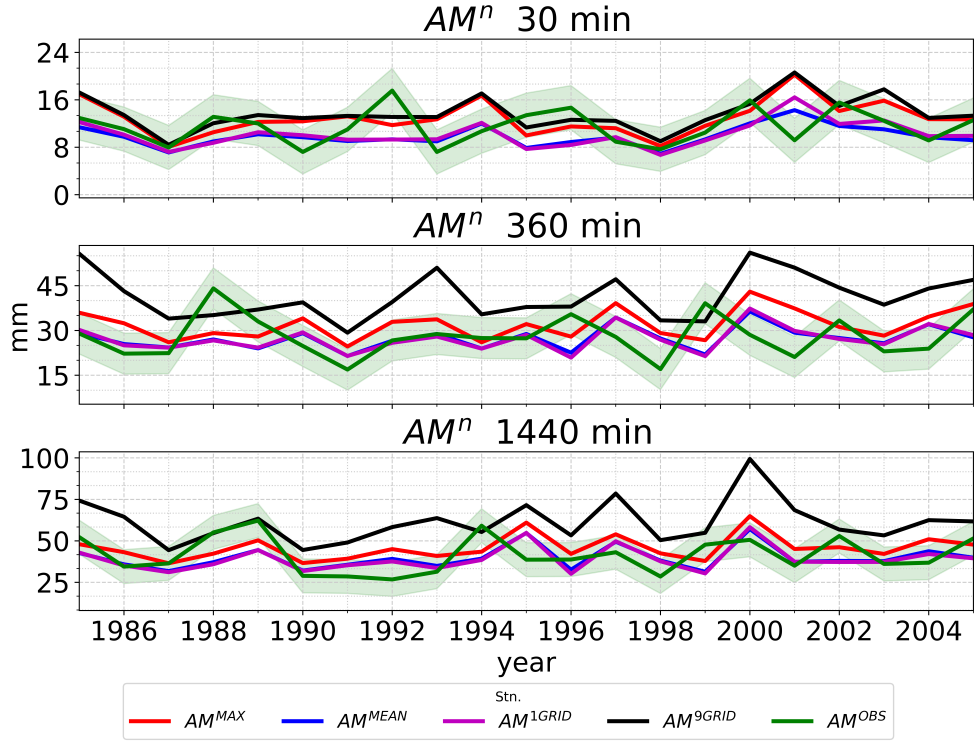

Figure 13: Stations average annual maximum precipitation (AM^n) for 30 min, 6 hours and 24 hours duration from 1985 to 2005. Green line is the OBS series with one STD in shaded green. Black, red, blue and magenta is the modelled 9GRID, MAX, MEAN and 1GRID methods respectively. Units are mm on the y-axis and years on the x-axis.

6.1.1 Standardized AM values for all stations

In Figure 14 the standardized AM values (*z-score*, Eq.27), is presented for all durations and stations. Here the standardization is done for each AM

method to check if any stations have particularly large or small AM values compared to the other stations for the given duration and AM method. A positive value means that the AM value is above the station mean value for the given AM method and duration. Analysing the z-scores will also reveal if the methods agree on the AM value size or not.

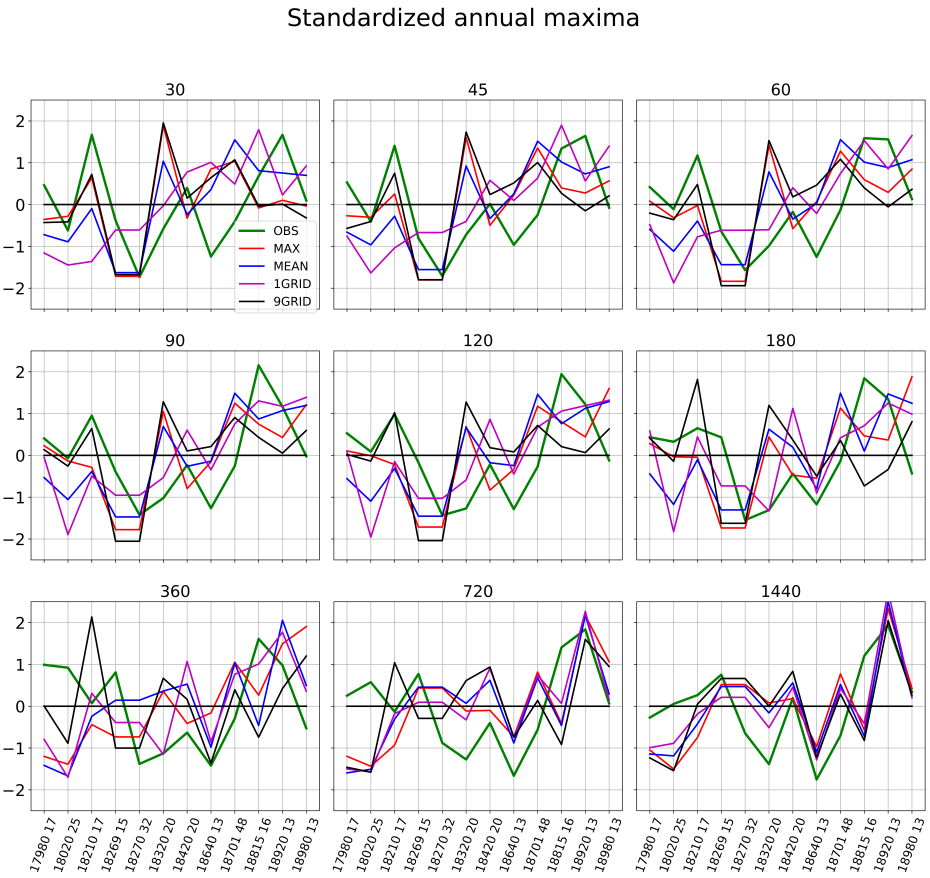

Figure 14: Standardized annual maxima values for all durations and stations. As indicated by the legend, green line is OBS, red line is MAX, blue line is MEAN, magenta line is 1GRID and black line is 9GRID. One the x-axis is the station number followed by the OBS time-series length in years.

Firstly we notice how the different AM methods produces more consistent z-scores for a given station as the duration increases. This means that for the individual station the AM methods agree how large the AM value of this station is compared to the other stations. However, for smaller durations the inter-method agreement on the AM value size is varying. Occurrences of both above and below 0 z-score amongst the AM methods for a given duration is a clear indication that the AM methods do not agree on the AM

value of the respective station compared to the other stations. Station 18320 and 18420 at 45 minute duration serves as an example. In this case, for the 18320 station the 9GRID AM value is almost 2σ above the 9GRID station mean, while the 1GRID AM value is around $1/2\sigma$ below the 1GRID station mean. Hence, it is a poor agreement on the size of the AM value of station 18320 compared to the rest of the stations between 9GRID and 1GRID for 45 min duration.
For small durations like 30 and 45 minutes there are also more occurrences of z-scores around 2 compered to larger durations like 720 and 1440 minutes. For the relevant stations and methods this means that they have large AM values compared to the other stations. It also means that for the larger durations, where the z-score over all is smaller, the AM values of a given AM method is more consistent across the stations.
Furthermore, some stations have consistent, small z-scores for many durations. Examples are stations 18269 (Haugenstua) and 18270 (Vestli) for the MAX, 9GRID and 1GRID method. The two stations have z-score at around -2σ for durations up to (and including) 3 hours. Thus they have small modelled AM values compared to the other stations. However, on 12 and 24 hour duration the z-score becomes slightly positive, implying larger than station-average AM values. The OBS z-scores are very consistent with the modelled z-scores for some stations, and not consistent at all for others. It appears to better consistency between the OBS and the modelled z-scores for larger durations.

6.2 Precipitation Magnitudes: IDF-values

Despite their close proximity in space, the precipitation return-values for a given return-period varies substantially amongst the stations. Figure 28 displays 10 year return-values for all stations and AM methods. The shaded green area indicates the 95 % confidence interval for OBS. As indicated by the number after each station name, the OBS data-series length for the specific station varies. For the two stations with longest data-series, Blindern and Vestli, 9GRID is outside the OBS top percentile, thus overestimating the return-level. The OBS confidence intervals of Blindern and Vestli are also the smallest of the 12 stations. For stations like Lilleaker, Hovin, Lamberseter and Disen 9GRID is right around the OBS top percentile, sometimes within and sometimes outside. For the Haugenstua, Bygdøy and Besserud station 9GRID is by far closest, and at times almost perfectly aligned with the OBS mean value. Common for these three is the relatively short OBS data-series

length, with 15, 16 and 13 years respectively. For the other 9 stations MAX appears to provide the overall best fit at the 10 year return-period. For the 2 year return-period 9GRID is outside the OBS confidence interval for most stations (not shown). At the 5 year return-period 9GRID serves the best fit to stations Bygdøy and Besserud, despite overestimating the return value on every other station by a large margin (not shown). While providing values within the OBS 95% confidence for stations Blindern, Hausmannsgate, Ljabruvegen and Vesti for the 50 year return-period and larger, the 9GRID values appear severely overestimated (not shown). Furthermore, MAX serves a very good fit to OBS for multiple stations, especially at return-periods larger than 10 years. For the smallest return-period MAX is a good fit at most durations on station Haugenstua, Ljabruvegen, Hovin and Besserud, while at 5 year return-period stations Ljabruvegen, Disen, Lilleaker. As the return-period increases beyond 20 years MAX also seems to become an increasingly better fit to longer time-series stations Vestli and Blindern. For corresponding figures to Figure 28 for the other return-periods see Appendix 9.

The MEAN and 1GRID methods provide very similar return-levels overall. For short return-periods like 2 or 5 years the MEAN return-values are close to identical to the 1GRID for all durations and stations, while for the larger returnperiods like 100 or 200 years MEAN yields slightly smaller return-values for durations up to around 3 hours. For 2 year return-period they have the best fit to OBS at Blindern, Vestrevika, Vestli and Hausmannsgate for all durations, but also to Lilleaker and Disen for durations smaller than 12 hours. Haugenstua has a unique feature compared to the other stations, namely that all AM methods underestimate the mean precipitation magnitude for most durations from 50 year return-period and larger. This includes 9GRID, which for all other stations and return-periods are matching or overestimating the precipitation magnitude compared to OBS.

Some stations experience decreasing return-levels with increasing duration for a short duration interval. Examples are Vestrevika from 12 hours to 24 hours or Besserud from 1 hour to 3 hours. This is an issue partially arising when using short time series for the GEV distribution. Each duration is fitted individually, and since the resulting precipitation magnitude is a fit to the available years, some large or low values in a short time-series will possibly have a greater impact on the curves compared to the same values in a longer time-series. This feature is mainly found for some stations at occasional durations for OBS return-values, but it also occurs for modelled return-values. This is a purely statistical phenomena arising when you are representing a population with a small sample. The true maxima of a precipitation event will always increase with increasing duration.

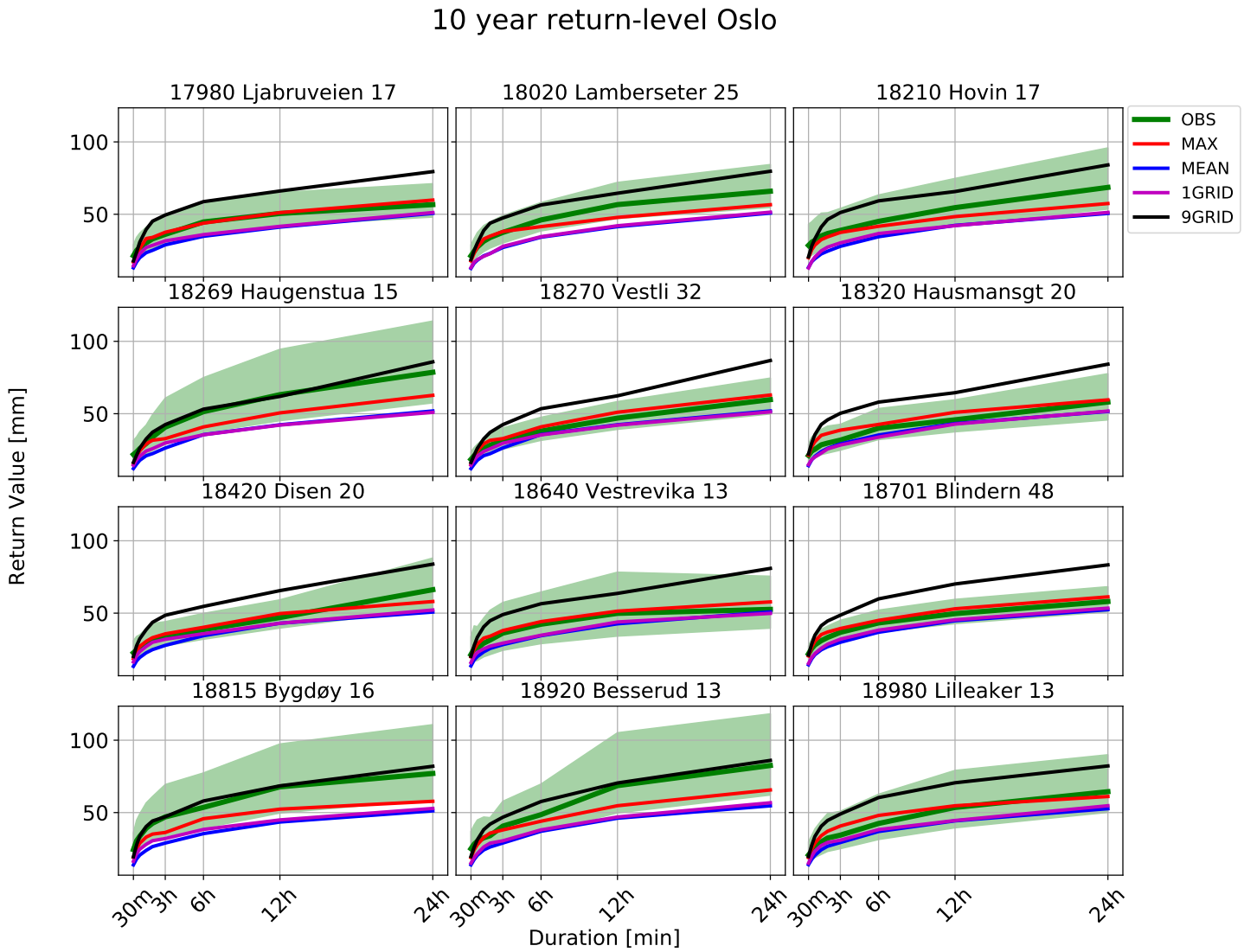

Figure 15: 10 year return-value for durations 30 min to 24 hours for the 12 stations in the Oslo area for the OBS, MAX, MEAN, 9GRID and 1GRID method. Shaded green area is the 95% confidence of OBS. Units on the y-axis is mm and units on the x-axis is duration in minutes or hours. The station name, number and OBS time-series length is written above each subplot.

On average for all stations the MAX method overestimates the return-values slightly for 2 year return-period, and slightly underestimates for all return-periods larger than 5 years. Compared to 1GRID, MAX appear to offer an improved representation of the return-values. In table 3 the relative difference between the respective models and the OBS return-values are listed as station- and duration-averaged percentage. For the larger returnperiods MAX is the overall method most consistent with OBS, especially for the stations with narrow confidence intervals. Providing higher estimates of the return-values than the MEAN and the 1GRID method but smaller estimates compared to the 9GRID, the MAX method serves as a middle ground which according to table 3 is more consistent with the OBS estimates on average for all of the Oslo stations. Only for a 200 year return-period the station and duration average for the 9GRID method is more consistent.

Return Period	1GRID	9GRID	MEAN	MAX
2	88.78	134.92	89.30	107.89
5	82.27	125.45	79.57	100.00
10	79.29	121.10	75.09	96.37
20	77.02	117.78	71.65	93.60
25	76.39	116.86	70.70	92.84
50	74.66	114.33	68.07	90.73
100	73.20	112.20	65.83	88.95
200	71.93	110.36	63.89	87.42

Table 3: Station- and duration-average return-values for each return-period in percentage of the OBS station- and duration-average return-values. Each column is one annual maximum method. Values >100 means average overestimation compared to OBS values and values <100 means underestimation compared to OBS for the given return-period.

6.3 Precipitation Intensity

Precipitation intensity, I are in some cases more applicable than precipitation amount. In Figure 16 the 10 year return-level precipitation intensity is plotted for all durations and all AM methods at the Blindern station. The range of I amongst the AM methods is largest for small durations, decreasing to a very small range between the AM methods at 24 hours. At around 44mmh^{-1} the OBS method has the largest value at 15 minutes duration, while the MEAN method measures around 28 mmh^{-1}. Marked with the dotted lines in Figure 16, the 97.5 percentile value at 15 minutes is almost identical for OBS, MAX9 and 9GRID at approximately 55mmh^{-1}. The

9GRID I quickly turns into the largest intensity for durations above 30 minutes. MAX has slightly higher intensity for all durations except 15 minutes compared to OBS, but are overall very similar to OBS. From 1 hour duration all methods except 9GRID and MEAN are within the I 95% confidence interval of OBS. On smaller durations the 9GRID and MAX I are the only two intensities within the confidence level of OBS.

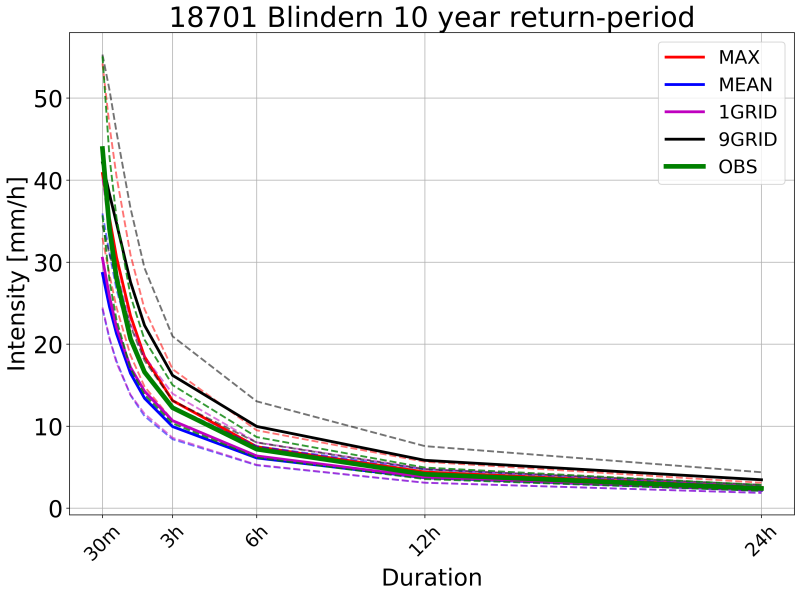

Figure 16: 10 year return-period intensity for OBS, MAX, MEAN, 9GRID and 1GRID at Blindern station. Dotted lines are 95% confidence interval to corresponding method. Duration in minutes and hours on the x-axis and intensity in mm/hour on the y-axis.

Considering the equivalent analysis to Figure 16 for other return-periods a few features appears (not shown). Firstly, the OBS intensity grows larger compared to the other AM methods with increasing return-period. This means that for short durations up to 1 hour and return-period of 100 or 200 year the OBS intensity is higher than the other AM methods compared to the 10 or 20 year return-period for the same duration. For the 2 year return-period MEAN and 1GRID is the most consistent fit to OBS intensity, while for increasing return-period up to 20 years the MAX intensity becomes a increasingly better fit to OBS for most durations. From 20 years onward the MAX intensity is very similar to OBS for durations larger than 1 hour. 9GRID appears to have a gentler slope compared to the other methods as the return-period increases, possibly overestimating I for durations between 1 and 12 hours. In general the preciptiation intensity sees to be well captured

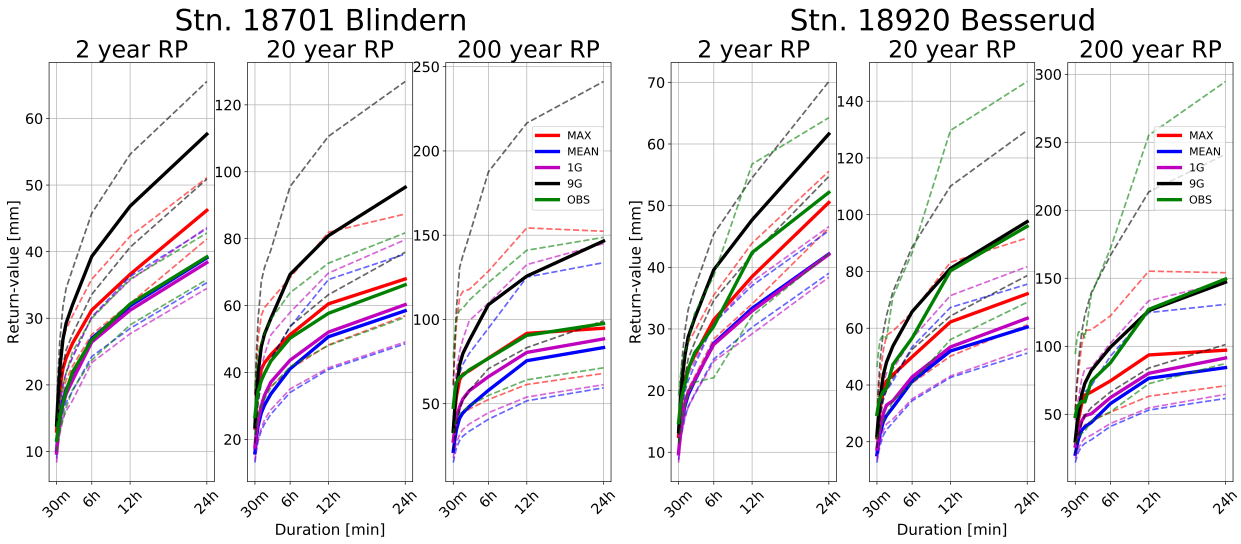

Figure 17: Return-values for 2, 20 and 200 year return-period for the OBS, MAX, MEAN, 9GRID and 1GRID method at station Blindern (a) and station Besserud (b). Dotted lines are 95% confidence intervals for the respective methods. The x-axis is duration in and the y-axis is return-value in mm.

by the HARMONIE-AROME model.

Evaluating the return-levels for the Blindern station in Figure 17 reveals large confidence intervals. From the figure one can observe that MEAN and 1GRID are the only methods inside the OBS confidence interval for the 2 year return-period. For all other return-periods MAX is also well within the OBS confidence interval. At 50 year return-period 9GRID is included in the OBS confidence interval from 30 minutes to 6 hours duration. At 100 years return-period 9GRID is within the confidence interval until around 18 hours duration, while at 200 year return-period it is within the OBS confidence for all durations. However, for all return-periods 9GRID is a round 50 % larger compared to OBS at 24 hours duration. Thus, for small return-period MEAN and 1GRID is very consistent with OBS, while MAX is a better fit as the return-period increases.

Seemingly related to the OBS time-series length, the confidence intervals in Figure 17 are enormous, especially for large return-periods. An example is to look at the 20 year return-period, 6 hour duration OBS return values for Blindern and Besserud (48 and 13 years OBS time-series length respectively). Blindern has a expected return-value of 52mm, bottom percentile at -21% and top percentile at +25%. Besserud has a expected return-value of 57mm, bottom percentile at -28% and top percentile at +58%. Although the bot-

tom percentile is not completely different between the two stations, the top percentile of Besserud is in this case almost twice that of Blindern. Since the confidence interval increases as the duration and the return-period increases, this example illustrates how vulnerable the IDF-statistics is to time-series length. For all return-periods the 9GRID top percentile return-value is sky-high and very large compared to any recordings in the area. The same applies the OBS top percentile at short time-series stations, whereas the MAX top percentile appears to have more realistic values even for the short time-series stations.
For a 2 year return-period the MEAN and 1GRID AM methods has the most consistent fit to long time-series OBS return-values, but at shorter time-series stations MAX has the most consistent fit. For return-period larger than 5 years MAX has the better fit to OBS at longer time-series stations and 9GRID has the most consistent fit for shorter time-series stations.

6.4 2080-2100

A future period from 2080 to 2099, hereafter referred to as "the future" has also been analysed. Here the HARMONIE-AROME climate model output similar to the previous analysis has been used. Overall the precipitation magnitude for the future period is higher compared to the 1985 to 2005 period, hereafter called "historic" period, for all stations, durations and return-periods. However, the increase is not uniform across durations and return-periods. Also between the AM methods there are differences in IDF magnitude change. In Figure 18, 19, 20 and 21 the station average future precipitation magnitude change as percentage are plotted for each duration and a selection of return-periods. The most striking feature is the large peak in increase at around 3 hours, most evident for large return-periods but also present for short return-periods. For 1GRID the increase is fairly similar to that of MAX for 200 year return-period with 88% and 86% increase respectively, while for the 2 year return-period the increase is 35% and 41% respectively. MEAN has the largest increase at 3 hour duration at 103% for 200 year return-period but also the second smallest increase at 36% at the 2 year return-period.

3 hours is also the duration for which the difference in future change between the return-periods is largest. As an example 1GRID precipitation change ranges from 35% to 88% at 3 hours for 2-200 years return-period, while at 30 minutes and 24 hours the range is 24% to 33% and 27% to 41% respectively. Across the methods the smallest increase is found to be around 25 % at 30 minute duration for the 2 year-return-period. For 9GRID the range of change between the 2 and the 200 year return-period at 24 hour

duration is only 5%.

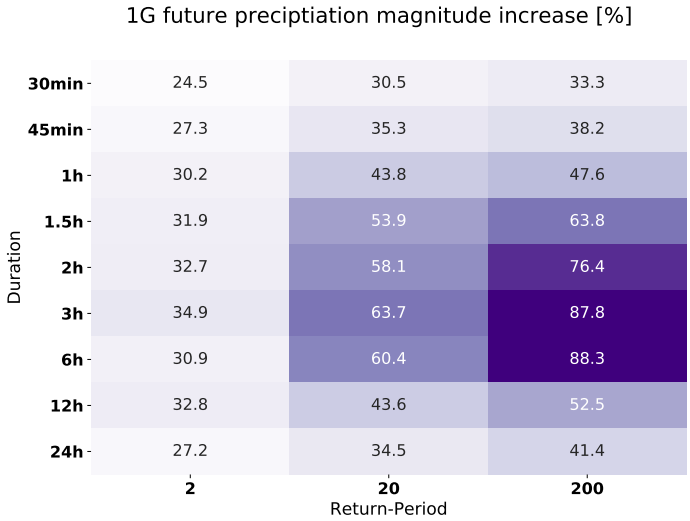

Figure 18

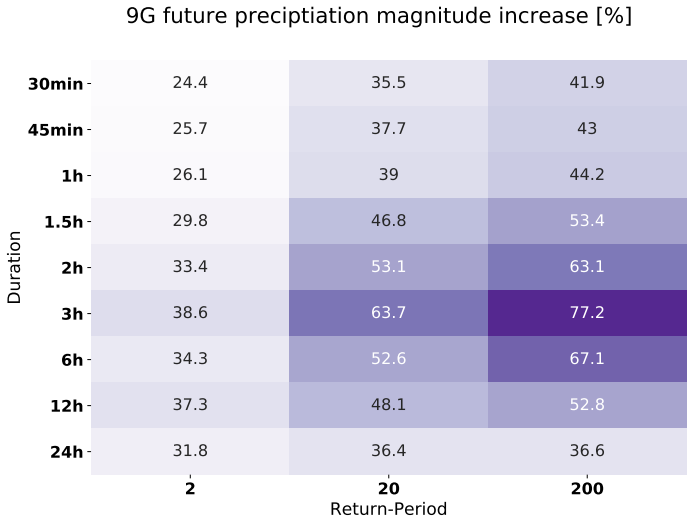

Figure 19

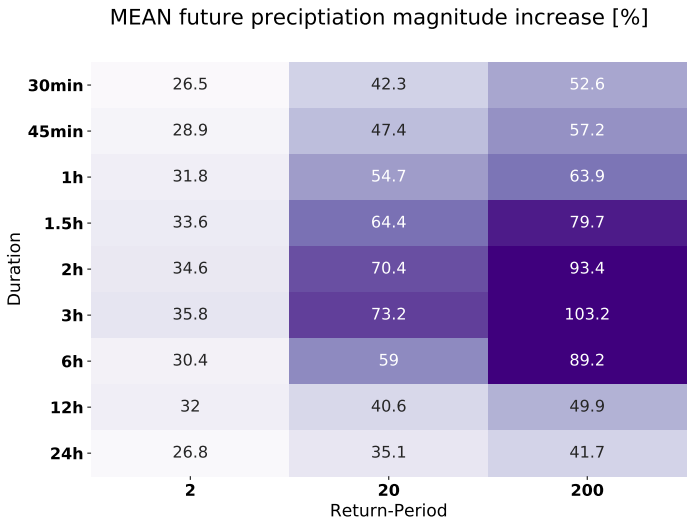

Figure 20

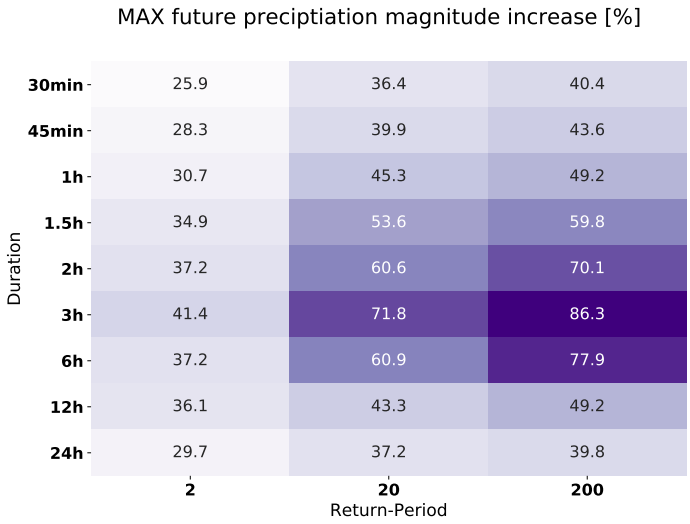

Figure 21: Common caption for Figure 18, 19, 20 and 21. Future precipitation increase in percentage for the different AM methods. Darker color correspond to larger value, lighter color correspond to smaller value.

7 Summary & Discussion

7.1 Model & Method

We have in this analysis used a common approach with some alterations to calculate IDF-values. A GEV distribution was fitted to modelled precipita-

tion data. We here applied a Bayesian framework as this was recommended as the most accurate method for operational IDF-calculation in Norway by the Norwegian Meteorological Institute (Lutz et al. 2020). This framework provided IDF-values with an easily derived confidence interval, allowing for better value and interpretation of the resulting return-values. The data used in the study has not yet been used for deriving IDF-curves in Norway, and it is one of the first 3km resolution climate runs used for this purpose. One major reason for using the dataset was the high resolution, as this is found to add levels of detail and value on extreme preciptiation analysis by multiple studies. The data is a product from a climate setting of an operational weather forecasting model called HARMONIE-AROME. Deep convection is resolved explicitly, making a large improvement over typical parameterization on deep convection of coarser models. Another motivation for using the dataset was exploring currently unexplored data for the purpose of improving knowledge on extreme precipitation in Norway.

The use of different annual maximum extraction methods was essentially inspired by two main ideas. Firstly we wished to infer the GEV statistics and the resulting IDF-curves representative for a larger area compared to single-point or single-gridcell locations. Doing so would make the analysis more suitable for the Oslo area, or at the very least reveal if there are major differences in IDF-values between the various annual maximum methods for the different stations in Oslo. In a report by the Norwegian Resources and Energy Directorate (Førland et al. 2015) it is recommended to check the IDF-values of multiple stations in case of decision making, not only the one closest to the area of operation. This highlights the need of IDF-values representing a larger area compared to single points in space, or the need for improved resolution of IDF-values compared to observations. The other argument for using multiple methods for extraction of annual maximum arose from an early analysis where single grid-cell IDF-values were calculated and proved to be very small compared to the currently used observation-based IDF-values for numerous stations. It would be difficult to verify whether this initial method actually represented the observations or not given the short time-series of most stations in the area, thus making the following analysis unfit to improve knowledge on extreme precipitation in the Oslo-area.

7.2 Results

Choice of method for extraction of annual maxima have proved varying results depending on station, return-period and duration. The question comes down to what a single point in space within the climate model is represent-

ing. Convective systems can be small in both spatial and temporal scale with abrupt borders to the environment. Thus, an event can very well strike one part of Oslo while others remain perfectly dry. In a case like this, despite being only a few kilometers or less apart, this one station might capture the event, while another does not. Yet, the city of Oslo faces potentially large damages due to the precipitation. The same applies for a climate model. Even though a particular gridcell in a climate model experience extraordinary precipitation, the grid cell next to it might experience little or no precipitation.

9GRID, MAX and MEAN all partially solves this issue by including information from multiple gridcells. As discussed in Section 6.2, 9GRID severely overestimates return-values at some stations, while at other stations 9GRID appeared to be a better fit to OBS. The general pattern is that with short time-series stations like Bessserud with 13 years of data, 9GRID is by far the most consistent fit to OBS, although the 95 % confidence interval in such cases is massive. This is especially valid for the 100 and 200 year return-period. Despite being the most consistent method for such cases one might argue that the length of the OBS time-series is too small to properly represent the actual precipitation state of the station. In this case the true precipitation state of the station is unknown, and thus its hard to verify weather one of the AM methods are more consistent with the observations than others. This argument has larger value as the return-period increases due to the increasing OBS confidence interval with increasing return-period.

For longer time-series stations like Blindern or Vestli with 48 and 32 years of data respectively, 9GRID overestimates the return-level severally for all return-periods. For the 2 year return-period MEAN and 1GRID has the most consistent fit, wheres for return-periods of 5 year and larger the MAX methods is very consistent to OBS. These statements also holds for the top and bottom percentile, where MAX is very consistent with OBS for all durations.

The issue of time-series length and its impact on choice of annual maximum method becomes very clear when investigating the return-values of station 18269 Haugenstua and 18270 Vestli. They are located only 2km apart, hence they are very likely to experience the same precipitation events. Haugenstua has 15 years OBS time-series and Vestli has 32 years. For 2-5 years return-periods MAX is the most consistent fit to Haugenstua, whereas 1GRID/MEAN is most consistent fit to Vestli. However, for all return-periods larger that 5 years 9GRID and MAX is the most consistent fit for Haugenstua and Vestli respectively.

Ultimately the predictable return-level for a given return-period is very

limited by the time-series length of the AM data. A data-series of 10-15 years can most likely represent return-periods of around 20 years at most, while a series of 100 years probably could represent around 500 years. This means that the HARMONIE-AROME dataset of 20 years serves best use for precipitation values at return-periods of 2-10 years. Beyond 20 years the confidence interval becomes colossal making the return-value unsuitable for most practical purposes. In the case of Blindern where the OBS time-series is almost 50 years, the representable return-period is probably close to 100 years. Despite the high-resolution dataset used in this study, this feature, being common for most extreme value statistics problems, still poses a problem in return-value estimation.

The fact that some stations have large, either positive or negative, standardized AM value (see Figure 14) for a given duration compared to the other stations indicates high spatial variability in the modelled AM values. Since this variability is also found for the observations it is a indication that HARMONIE-AROME is able to replicate the local AM value range of the area. With exception of station 18320 Hausmannsgate and 18640 Vestre Vika the OBS and modelled standardized AM values co-vary for most stations and durations up to (and including) 2 hours. This indicates that the climate model is resolving the measured small-scale convective processes in the area quite well. Also, the fact that station Haugenstua and Vestli have negative z-score for short durations and positive z-score for larger durations implies different precipitation characteristics between the durations. Also, all AM method agree on this, and that these two are colocated furthest away from the Oslo city center, suggest that HARMONIE-AROME is capable of capturing local precipitation patterns.

Additionally, for longer durations the inter-station agreement between the methods on the AM values is far better compared to shorter durations. This is expected since the larger durations are influenced by stratiform precipitation rather than convective events, making most of the stations experience similar precipitation magnitudes. Rather equal standardized AM values between the different AM methods means that the precipitation magnitude across the entire area is more or less the same for longer duration precipitation. Since the overall consistency between OBS z-score and the modelled z-scores where better for larger durations, it speaks for further need for resolution improvement and small-scale preciptiation representation in the climate model.

The station-average annual maximum values investigated in Figure 13 did not immediately explain why some AM methods were more consistent with

OBS for different return-periods. The modelled AM series for the different durations where generally within the OBS standard deviation, and the AM values were mostly co-varying in size for the modelled series and the OBS series. An exception was the overestimated 9GRID values for durations larger than 3 hours. Thus, it was partially expected that the 9GRID IDF-values generally overestimated the return-levels of OBS. Not so evident from Figure 13 was the fact that MAX return-values was more consistent with the OBS return-values than MEAN or 1GRID for most stations. Additionally, the AM variability of OBS and the modelled methods throughout the 20 years was not synchronized. While some specific years had large AM values for a given duration in the OBS series, this was not true for the same years in the modelled series. This is expected given the reasoning on temporal overlap between climate models and observation in Section 4.3.1. In the case of observations only it is crucial to compare the exact same period of time. In a climate model, however, a given 20 year period is somewhat arbitrary, representing the long-term statistics of the period rather than exact year-to-year measurements.

From Figure 18, 19, 20 and 21 in Section 6.4 it was found that the increase in preciptiation magnitude towards year 2100 peaked at the 3 hour duration for most AM methods and return-periods. The average 3-hour duration increase across the AM methods for 2, 20 and 200 year return-period was 38%, 68% and 89% respectively. This is conflicting with various findings about future change of extreme precipitation, as most other studies find largest increase for the smallest duration studied. In a report about expected future short-term preciptiation in Norway, (Dyrrdal et al. 2019a) the national mean increase at 1 hour duration was found to be 42% ,46% and 54% for the 5, 20 and 200 year return-period respectively, while at 3 hours duration the corresponding increase was found to be 35%, 38% and 43%. (Hodnebrog et al. 2019) found that the expected change in precipitation for Southern Norway was largest for the smallest duration (10 minutes), then decreasing as the duration approached 24 hours. (Hosseinzadehtalaei et al. 2020a) also found that greater change is projected for shorter-duration extreme preciptiation events in Europe. This is conflicting with the findings of this study, both for the 3 hour duration peak increase and for the fact that no definite pattern of largest increase at shortest duration overall was found.

There are some explanations to why the largest increase is found for the 3 hour duration in this study. The sub 3 hour duration change may be underestimated, making the 3 hour duration increase appear very large. There are however little evidence in the initial return-value analysis of this study sug-

gesting that the shortest durations are severely underestimated. Since most findings to date are based on lower-resolution climate models which does not resolve convection explicitly, typical short-duration precipitation projections might be underestimated. This statement can also be supported by the fact that increasing the resolution of the climate model is found to improve the representation of extreme precipitationKopparla et al. 2013. Thus, since deep convection is resolved explicitly in the HARMONIE-AROME climate run, the 3-6 hour duration precipitation might be well resolved. The peak could also be a case of "bad luck" with several large return-period 3-hour events during the modeled 2080-2100 period. Given the short time-series, one or two extremely large events in the 2080-2100 period could severely affect the IDF-values, making the 3 hour duration increase unrealistically large compared to the other durations. Another possible explanation for the peak is an actual change in circulation of the model. There are however little evidence in the literature this is the case.

Regarding the size of the magnitude change in Figure 18, 19, 20 and 21 besides the 3 hour peak it is fairly consistent with the findings on national Norwegian values (Dyrrdal et al. 2019b). They found average preciptiation change in Norway towards year 2100 for duration 1-24 hours in the range of 42-26%, 46-27% and 54-30% for 5, 20 and 200 year return-period respectively. The corresponding changes found in this study was 30-29%, 46-36% and 51-40% for 2, 20, 200 year return-period respectively (not shown, averages of Figure 18, 19, 20 and 21. This result is roughly the same as to Dyrrdal and Førland for the short durations, but a few percent larger for the longer durations.

8 Conclusions & Outlook

8.1 Summary

This study has proved that precipitation from a convective-scale climate model simulation is applicable for IDF-calculations. For the Oslofjorden area, modelled high-resolution precipitation can aid the gauge network and limited quality precipitation data for extreme value calculations. Furthermore, based on this study, the data and methods used here can help derive projected changes in extreme precipitation.

The annual maximum precipitation was found consistent with that of the observations, generally within the OBS standard deviation. Furthermore, modelled return-values were generally consistent with OBS return-levels. However, which annual maximum method was most consistent with OBS was mainly dependent on return-period and the OBS time-series length. 1GRID and MEAN was most consistent with OBS for 2 year return-period and long data series stations, but were otherwise underestimating the OBS return-levels, even being outside the OBS confidence interval for many stations and return-periods. MAX proved suitable for long time-series stations with return-periods over 5 years, or 2 year return-period for short data-series stations. 9GRID generally overestimated the the OBS confidence for long time-series stations on the 2 year return-period, but was very consistent with above 5 years return-period for shorter time-series stations.

Since the AM methods including several gridcells generally are more consistent with OBS than 1GRID, it could point towards single-gridcell representation of convective-scale events are still to coarse to resolve kilometer-scale extreme precipitation. Nevertheless, evidence is found that the model is capable of resolving the range of annual maximum preciptiation values for all durations from 30 min to 24 hours. There are found inter-stations differences in AM values both consistent and inconsistent with the OBS AM values, indication both well resolved convective precipitation and poorer resolved convective precipitation. Therefore, while the climate models further improve on resolution and small-scale convection representation, we here recommend to continue using and further develop multi-gridcell extraction of annual maximum like the MAX method for IDF-calculations.

The study also reveals peak projected extreme precipitation increase at 3-6 hour duration. This is not yet supported by other studies, and it remains to verify whether this is an actual projection or some statistical artifact.

8.2 Future work

An event-based analysis could help further assess the convection-resolving capability of the HARMONIE-AROME climate model. Comparing the modelled annual maximum values or return-values to kilometer-scale resolution data like radar-products would allow better verification of the return-values compared to the sparse in situ station network in Norway. Gauge measurements from regions like the one used in this study with relatively high station-density could be used for this purpose, but it would require verified recordings of strong events covering the entire or parts of the area where these gauges are located. With a growing, much denser observation network, crowd-sourced personal weather stations like NETATMO*Smart Weather Station, Rain Gauge and Anemometer* could also been used for such a purpose. Despite various issues, studies like (Chapman et al. 2017) have proved encouraging results on meteorological studies using NETATMO meteorological data.

In this study the IDF-curves are assumed stationary, achieved through time-independent GEV parameters, $\alpha = \{\zeta, \mu, \sigma\}$. However, under non-stationary conditions like climate change, the parameters and properties of the GEV distribution becomes time-dependent. Here the parameters can be expressed as functions of covariates like a time-dependent trend or climate indices such as for example the Southern Oscillation Index (SOI) in (Sarhadi et al. 2017). Unlike in the analysis of this study, trends throughout the modelled period would then be taken into account.

To further investigate representation of convective-scale systems an additional existing simulation, using the atmospheric reanalysis ERA-Interim as boundary conditions, should be analysed. Although new sources of uncertainty would be introduced, the ERA-I dataset could possibly reveal further strengths or weaknesses to the GCM driven data used in this study. Locations more exposed to stratiform precipitation like in Western Norway could also be included to compare the representation of short-duration return-values of a region heavily influence by convective preciptiation to a region less influenced by convective precipitation.

Furthermore the GEV parameters of the modelled return-values should be examined. They affect the return-values and could highlight some differences between stations or the AM methods that the time-series length alone could not. Analysing the GEV parameters could also be helpful to understand the peak future increase at 3-hour duration. Additionally further analysis on this dataset or similar high-resolution datasets should be done to verify whether this increase is an actual projected increase, and if it is valid for other locations in Norway as well.

References

Ahrens, C. D. (2014). *Essentials of meteorology: an invitation to the atmosphere*. Cengage Learning.

Azad, R. and A. Sorteberg (2017). "Extreme daily precipitation in coastal western Norway and the link to atmospheric rivers". In: *Journal of Geophysical Research: Atmospheres* 122.4, pp. 2080–2095. DOI: https://doi.org/10.1002/2016JD025615.

Belušić D., de Vries H. Dobler A. Landgren O.. Lind P. Lindstedt D. Pedersen R.A. Sánchez-Perrino J.C. Toivonen E. van Ulft B. Wang F. Andrae U. et al. (2020). "HCLIM38: a flexible regional climate model applicable for different climate zones from coarse to convection-permitting scales". In: *Geoscientific Model Development* 13.3, pp. 1311–1333. DOI: 10.5194/gmd-13-1311-2020. URL: https://gmd.copernicus.org/articles/13/1311/2020/.

Bengtsson L., Andrae U. Aspelien T. Batrak Y. Calvo J. Rooy W. Gleeson E. Sass B.-Homleid M. Hortal M. Ivarsson K. Lenderink G. Niemelä S. Nielsen K. Onvlee J. Rontu L. Samuelsson Patrick et al. (Feb. 2017). "The HARMONIE-AROME model configuration in the ALADIN-HIRLAM NWP system". In: *Monthly Weather Review* 145. DOI: 10.1175/MWR-D-16-0417.1.

Berg P., Moseley C. and J. O. Haerter (2013). "Strong increase in convective precipitation in response to higher temperatures". In: *Nature Geoscience* 6.3, pp. 181–185.

Bork, J. and L. Grinde (2020). *MET-info 2019*. Accessed: 2021-23-02. URL: https://www.met.no/publikasjoner/met-info/met-info-2019.

Buonocore, Mauro. *Earth System Modeling, a definition*. Accessed: 2021-26-05. URL: https://www.climateurope.eu/earth-system-modeling-a-definition/.

Chapman L., Bell C. and S. Bell (2017). "Can the crowdsourcing data paradigm take atmospheric science to a new level? A case study of the urban heat island of London quantified using Netatmo weather stations". In: *International Journal of Climatology* 37.9, pp. 3597–3605. DOI: https://doi.org/10.1002/joc.4940.

Courty L.G., Wilby R.L. Hillier J.K. and L.J. Slater (2019). "Intensity-duration-frequency curves at the global scale". In: *Environmental Research Letters* 14.8, p. 084045.

Danielson, J. J. and D. B. Gesch (2011). *Global multi-resolution terrain elevation data 2010 (GMTED2010)*. US Department of the Interior, US Geological Survey.

Duffy P.B., Govindasamy B. Iorio J.P. Milovich J. Sperber K.R. Taylor K.E. Wehner M.F. and S.L. Thompson (2003). "High-resolution simulations of global climate, part 1: present climate". In: *Climate Dynamics* 21.5-6, pp. 371–390.

Dyrrdal, A. V. and E. J. Førland (2019a). *Norsk Klimaservicesenter*. URL: https://klimaservicesenter.no/kss/rapporter/rapporter-og-publikasjoner_2.

Dyrrdal, A.V. and E.J. Førland (2019b). "Klimapåslag for korttidsnedbør-Anbefalte verdier for Norge". In: *NCCS report 5*.

Dyrrdal A. V., Lenkoski A. Thorarinsdottir T. L. and F. Stordal (2015). "Bayesian hierarchical modeling of extreme hourly precipitation in Norway". In: *Environmetrics* 26.2, pp. 89–106. DOI: https://doi.org/10.1002/env.2301.

Dyrrdal A. V., Skaugen T., F. Stordal, and E. Førland (Dec. 2014). "Estimating extreme areal precipitation in Norway from a gridded dataset". In: *Hydrological Sciences Journal* 61, p. 141217125340005. DOI: 10.1080/02626667.2014.947289.

fevd: Fit An Extreme Value Distribution (EVD) to Data (2021). Accessed: 2021-26-05. URL: https://www.rdocumentation.org/packages/extRemes/versions/2.0-12/topics/fevd.

Finansnorge (2020). *Vannskader*. Accessed: 2021-20-05. URL: https://www.finansnorge.no/statistikk/livsforsikringg/skadeforsikring/klimarelaterte-skader/vannskader/.

Førland E.J., Mamen J. Dyrrdal A.V. Grinde L. and S. Myrabø (2015). "Dimensjonerende korttidsnedbør". In: 134.

Fowler H. J., Ali H. Allan R. P Ban N. Barbero R. Berg P. Blenkinsop S., Chan S. Dale M. Cabi N. S., et al. (2021). "Towards advancing scientific knowledge of climate change impacts on short-duration rainfall extremes". In: *Philosophical Transactions of the Royal Society A* 379.2195, p. 20190542. DOI: https://doi.org/10.1098/rsta.2019.0542.

Fujibe, F. (2013). "Clausius–Clapeyron-like relationship in multidecadal changes of extreme short-term precipitation and temperature in Japan". In: *Atmospheric Science Letters* 14.3, pp. 127–132. DOI: https://doi.org/10.1002/asl2.428.

Gilleland, E. and R.W. Katz (2016). "extRemes 2.0: An Extreme Value Analysis Package in R". In: *Journal of Statistical Software* 72.8, pp. 1–39. DOI: 10.18637/jss.v072.i08.

Goudenhoofdt E., Delobbe L. and P. Willems (2017). "Regional frequency analysis of extreme rainfall in Belgium based on radar estimates". In: *Hydrology and Earth System Sciences* 21.10, pp. 5385–5399. DOI: https://doi.org/10.5194/hess-21-5385-2017.

Grossi G., Lendvai A. Peretti G. and R. Ranzi (2017). "Snow precipitation measured by gauges: Systematic error estimation and data series correction in the central Italian Alps". In: *Water* 9.7, p. 461. DOI: https://doi.org/10.3390/w9070461.

Hanley K. E., Plant R. S. Stein T. H. M. Hogan R. J. Nicol J. C.-Lean H W. Halliwell C. and P. A. Clark (2015). "Mixing-length controls on high-resolution simulations of convective storms". In: *Quarterly Journal of the Royal Meteorological Society* 141.686, pp. 272–284. DOI: https://doi.org/10.1002/qj.2356. eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.2356. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2356.

Heavens N. G., Ward D. S. and M.M. Natalie (2013). "Studying and projecting climate change with earth system models". In: *Nature Education Knowledge* 4.5, p. 4.

Hodnebrog Ø., Marelle L. Alterskjær K. Wood R. R. Ludwig R. Fischer E. M. Richardson T.B.-Forster P. M. Sillmann J. and G. Myhre (2019). "Intensification of summer precipitation with shorter time-scales in Europe". In: *Environmental Research Letters* 14.12, p. 124050. DOI: 10.1088/1748-9326/ab549c. URL: https://doi.org/10.1088/1748-9326/ab549c.

Hosking, J.R.M. and J.R. Wallis (1987). "Parameter and quantile estimation for the generalized Pareto distribution". In: *Technometrics* 29.3, pp. 339–349.

Hosseinzadehtalaei P., Tabari H. and P. Willems (2018). "Precipitation intensity–duration–frequency curves for central Belgium with an ensemble of EURO-CORDEX simulations, and associated uncertainties". In: *Atmospheric Research* 200, pp. 1–12. DOI: https://doi.org/10.1016/j.atmosres.2017.09.015.

— (2020a). "Climate change impact on short-duration extreme precipitation and intensity–duration–frequency curves over Europe". In: *Journal of Hydrology* 590, p. 125249. ISSN: 0022-1694. DOI: https://doi.org/10.1016/j.jhydrol.2020.125249. URL: https://www.sciencedirect.com/science/article/pii/S0022169420307095.

— (2020b). "Satellite-based data driven quantification of pluvial floods over Europe under future climatic and socioeconomic changes". In: *Science of The Total Environment* 721, p. 137688. DOI: https://doi.org/10.1016/j.scitotenv.2020.137688.

Huard D., Mailhot A. and S. Duchesne (2010). "Bayesian estimation of intensity–duration–frequency curves and of the return period associated to a given rainfall event". In: *Stochastic Environmental Research and Risk Assessment* 24.3, pp. 337–347.

Joyce, James (2003). "Bayes' theorem". In:

Kopparla P., Fischer E. M. Hannay C. and R. Knutti (2013). "Improved simulation of extreme precipitation in a high-resolution atmosphere model". In: *Geophysical Research Letters* 40.21, pp. 5803–5808. DOI: https://doi.org/10.1002/2013GL057866. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2013GL057866. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2013GL057866.

Lenderink, G. and E. Van Meijgaard (2008). "Increase in hourly precipitation extremes beyond expectations from temperature changes". In: *Nature Geoscience* 1.8, pp. 511–514.

Lind P., Belusic D. Christensen O. B. Dobler A. Kjellström E. Landgren O. Lindstedt D. Matte-D. Pedersen R. A. Toivonen E. and F. Wang (2020). "Benefits and added value of convection-permitting climate modeling over Fenno-Scandinavia". In: *Climate Dynamics*, pp. 1–20.

Lind P., Lindstedt D. Kjellström E. and C. Jones (2016). "Spatial and temporal characteristics of summer precipitation over central Europe in a suite of high-resolution climate models". In: *Journal of Climate* 29.10, pp. 3501–3518. DOI: https://doi.org/10.1175/JCLI-D-15-0463.1.

Lind P., Pedersen R. Kjellström E. Landgren O. Matte D. Dobler A. Belušić D. Toivonen E.-Wang F. Christensen O. B. (2021 in prep.). "High-resolution, convection-permitting, climate simulations of the 21st century over Fenno-Scandinavia". In:

Lindstedt D., Lind P. Jones C. and E. Kjellström (Jan. 2015). "A new regional climate model operating at the meso-gamma scale: Performance over Europe". In: *Tellus* 67. DOI: 10.3402/tellusa.v67.24138.

Lutz, J., L. Grinde, and A. V. Dyrrdal (2020). "Estimating Rainfall Design Values for the City of Oslo, Norway—Comparison of Methods and Quantification of Uncertainty". In: *Water* 12.6. ISSN: 2073-4441. DOI: 10.3390/w12061735. URL: https://www.mdpi.com/2073-4441/12/6/1735.

Marshall, J. and R. A. Plumb (2007). *Atmosphere, Ocean and Climate Dynamics: An Introductory Text (International Geophysics)*. Academic Press, 344pp.

Martins, E. S. and J. R. Stedinger (2000). "Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data". In: *Water Resources Research* 36.3, pp. 737–744. DOI: https://doi.org/10.1029/1999WR900330.

Mishra V., Wallace J. M. and D. P. Lettenmaier (2012). "Relationship between hourly extreme precipitation and local air temperature in the United States". In: *Geophysical Research Letters* 39.16. DOI: https://doi.org/10.1029/2012GL052790.

Mohymont B., Demarée G.R. and D.N. Faka (2004). "Establishment of IDF-curves for precipitation in the tropical area of Central Africa-comparison

of techniques and results". In: *Natural Hazards and Earth System Sciences* 4.3, pp. 375–387. DOI: https://doi.org/10.5194/nhess-4-375-2004.

Mokhov, I. and M. Akperov (July 2006). "Tropospheric lapse rate and its relation to surface temperature from reanalysis data". In: *Izvestiya, Atmospheric and Oceanic Physics* 42, pp. 430–438. DOI: 10.1134/S0001433806040037.

Müller M., Homleid M. Ivarsson K. Køltzow M.AØ. Lindskog M. Midtbø K.H. Andrae U. Aspelien T.-Berggren L. Bjørge D. et al. (2017). "AROME-MetCoOp: A Nordic convective-scale operational weather prediction model". In: *Weather and Forecasting* 32.2, pp. 609–627. DOI: https://doi.org/10.1175/WAF-D-16-0099.1.

Nhat L.M, Tachikawa Y. and K. Takara (2006). "Establishment of intensity-duration-frequency curves for precipitation in the monsoon area of Vietnam". In: *Annuals of Disas. Prev. Res. Inst*, pp. 93–103.

NVE (2011). *Retningslinjer for flomberegninger*. Accessed: 2021-26-05. URL: https://www.nve.no/damsikkerhet-og-kraftforsyningsberedskap/damsikkerhet/regelverk/retningslinjer-for-flomberegninger/.

Olsson J., Södling J. Berg P. Wern L. and A. Eronn (2019). "Short-duration rainfall extremes in Sweden: A regional analysis". In: *Hydrology Research* 50.3, pp. 945–960. DOI: https://doi.org/10.2166/nh.2019.073.

Park, In-Hong and Seung-Ki Min (2017). "Role of convective precipitation in the relationship between subdaily extreme precipitation and temperature". In: *Journal of Climate* 30.23, pp. 9527–9537. DOI: https://doi.org/10.1175/JCLI-D-17-0075.1.

Pope, V and R Stratton (2002). "The processes governing horizontal resolution sensitivity in a climate model". In: *Climate Dynamics* 19.3-4, pp. 211–236.

Roe, P. L. (1986). "Characteristic-based schemes for the Euler equations". In: *Annual review of fluid mechanics* 18.1, pp. 337–365. DOI: 0066-4189/86/0115-0337502.00.

Sarhadi, A. and E. D. Soulis (2017). "Time-varying extreme rainfall intensity-duration-frequency curves in a changing climate". In: *Geophysical Research Letters* 44.5, pp. 2454–2463. DOI: https://doi.org/10.1002/2016GL072201.

Savina M., Schäppi B. Molnar P. Burlando P. and B. Sevruk (2012). "Comparison of a tipping-bucket and electronic weighing precipitation gage for snowfall". In: *Atmospheric Research* 103, pp. 45–51. DOI: https://doi.org/10.1016/j.atmosres.2011.06.010.

Seity Y., Brousseau P. Malardel S. Hello G. Bénard P. Bouttier F. Lac-C. and V. Masson (Mar. 2011). "The AROME-France convective-scale operational model". In: *Monthly Weather Review* 139, pp. 976–991. DOI: 10.1175/2010MWR3425.1.

Sillmann J., Kharin V.V. Zhang X. Zwiers F.W. and D. Bronaugh (2013). "Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate". In: *Journal of Geophysical Research: Atmospheres* 118.4, pp. 1716–1733. DOI: https://doi.org/10.1002/jgrd.50203.

Smart Weather Station, Rain Gauge and Anemometer. Accessed: 2021-26-05. URL: https://www.netatmo.com/en-us/weather.

Team, R Core. *The R Project for Statistical Computing*. Accessed: 2021-26-05. URL: https://www.r-project.org/.

Wallace, J. M. and P. V Hobbs (2006). *Atmospheric science: an introductory survey*. Vol. 92. Elsevier.

Willner S. N., Levermann A. Zhao F. and K. Frieler (2018). "Adaptation required to preserve future high-end river flood risk at present levels". In: *Science advances* 4.1, eaao1914. DOI: 10.1126/sciadv.aao1914.

Yau, M. K. and R. R. Rogers (1996). *A short course in cloud physics*. Elsevier.

9 A Appendix

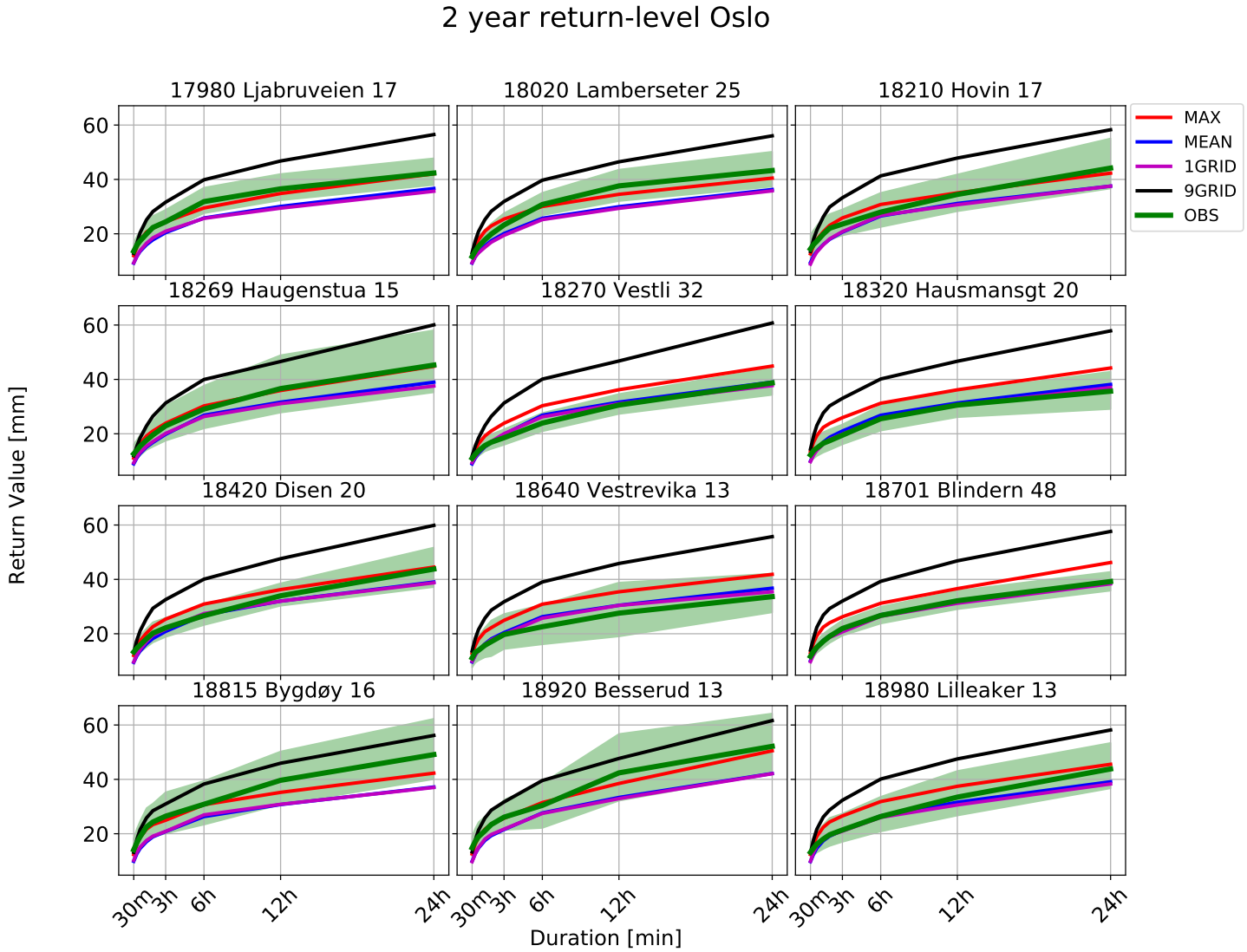

Figure 22: 2 year return-value for durations 30 min to 24 hours for the 12 stations in the Oslo area for the OBS, MAX, MEAN, 9GRID and 1GRID method. Shaded green area is the 95% confidence of OBS. units on the y-axis is mm and units on the x-axis is years. The station name, number and OBS time-series length is written above each subplot.**reload this figure. not supposed to be gray background. Easier to read with white background.**

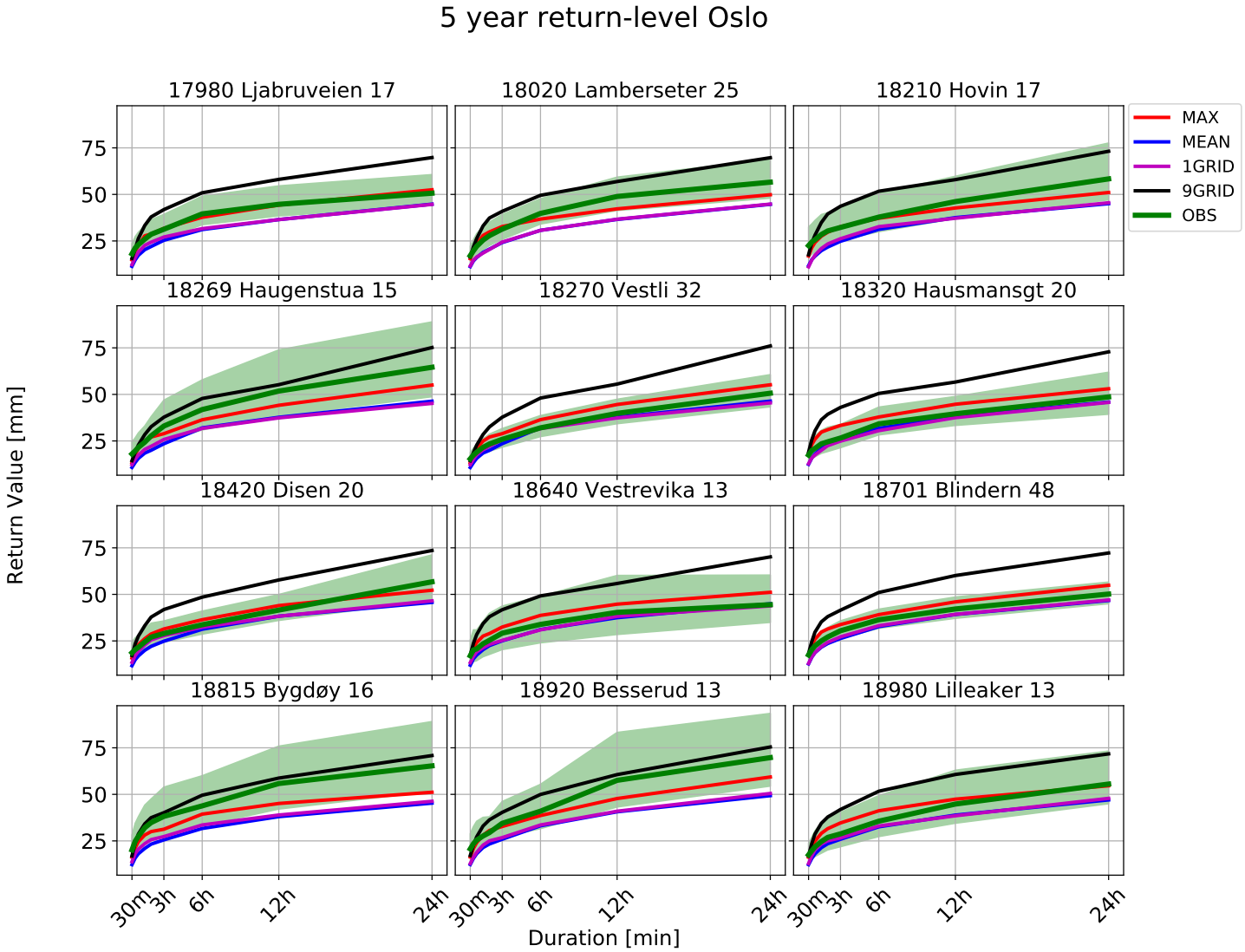

Figure 23: 5 year return-value for durations 30 min to 24 hours for the 12 stations in the Oslo area for the OBS, MAX, MEAN, 9GRID and 1GRID method. Shaded green area is the 95% confidence of OBS. units on the y-axis is mm and units on the x-axis is years. The station name, number and OBS time-series length is written above each subplot.**reload this figure. not supposed to be gray background. Easier to read with white background.**

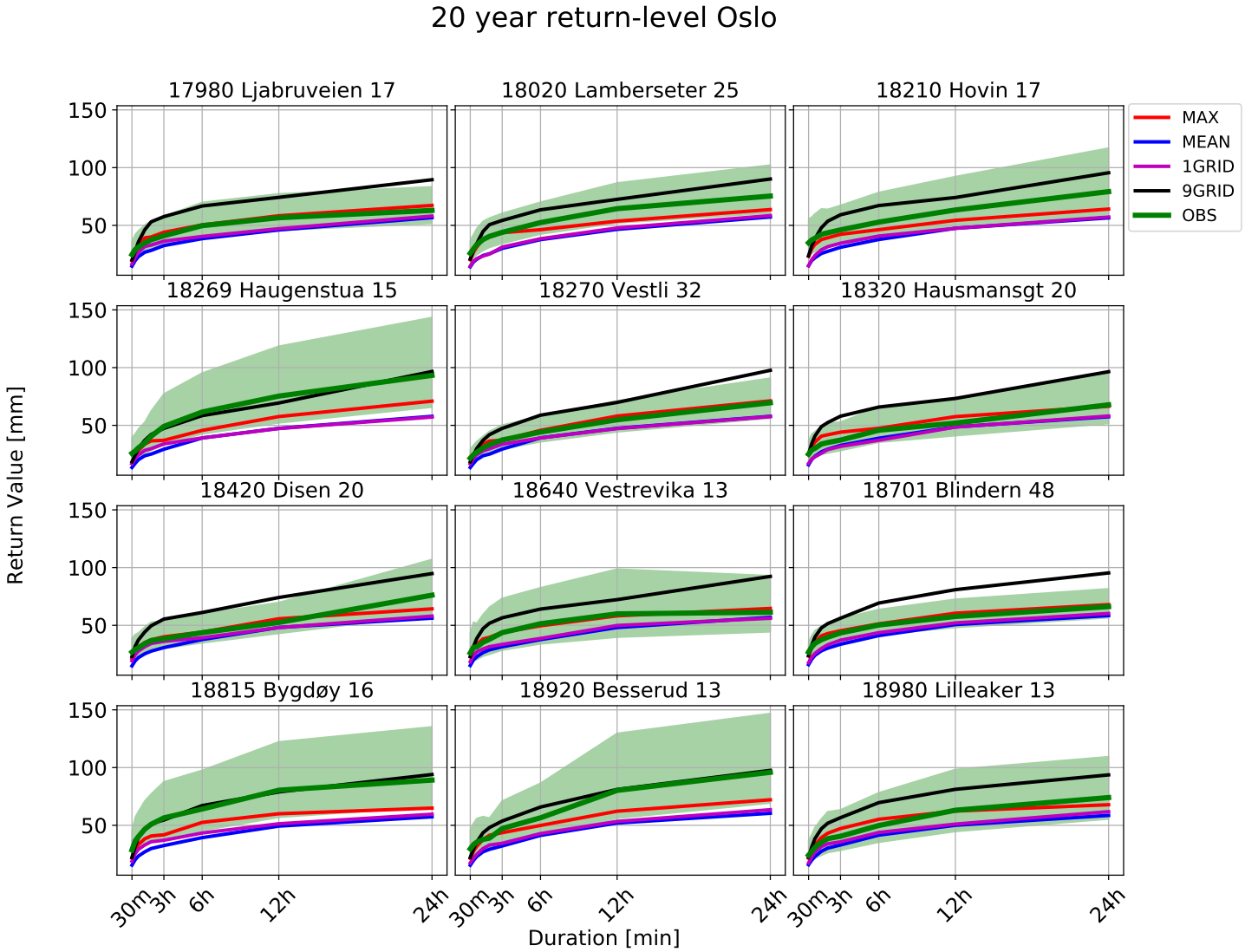

Figure 24: 20 year return-value for durations 30 min to 24 hours for the 12 stations in the Oslo area for the OBS, MAX, MEAN, 9GRID and 1GRID method. Shaded green area is the 95% confidence of OBS. units on the y-axis is mm and units on the x-axis is years. The station name, number and OBS time-series length is written above each subplot.**reload this figure. not supposed to be gray background. Easier to read with white background.**

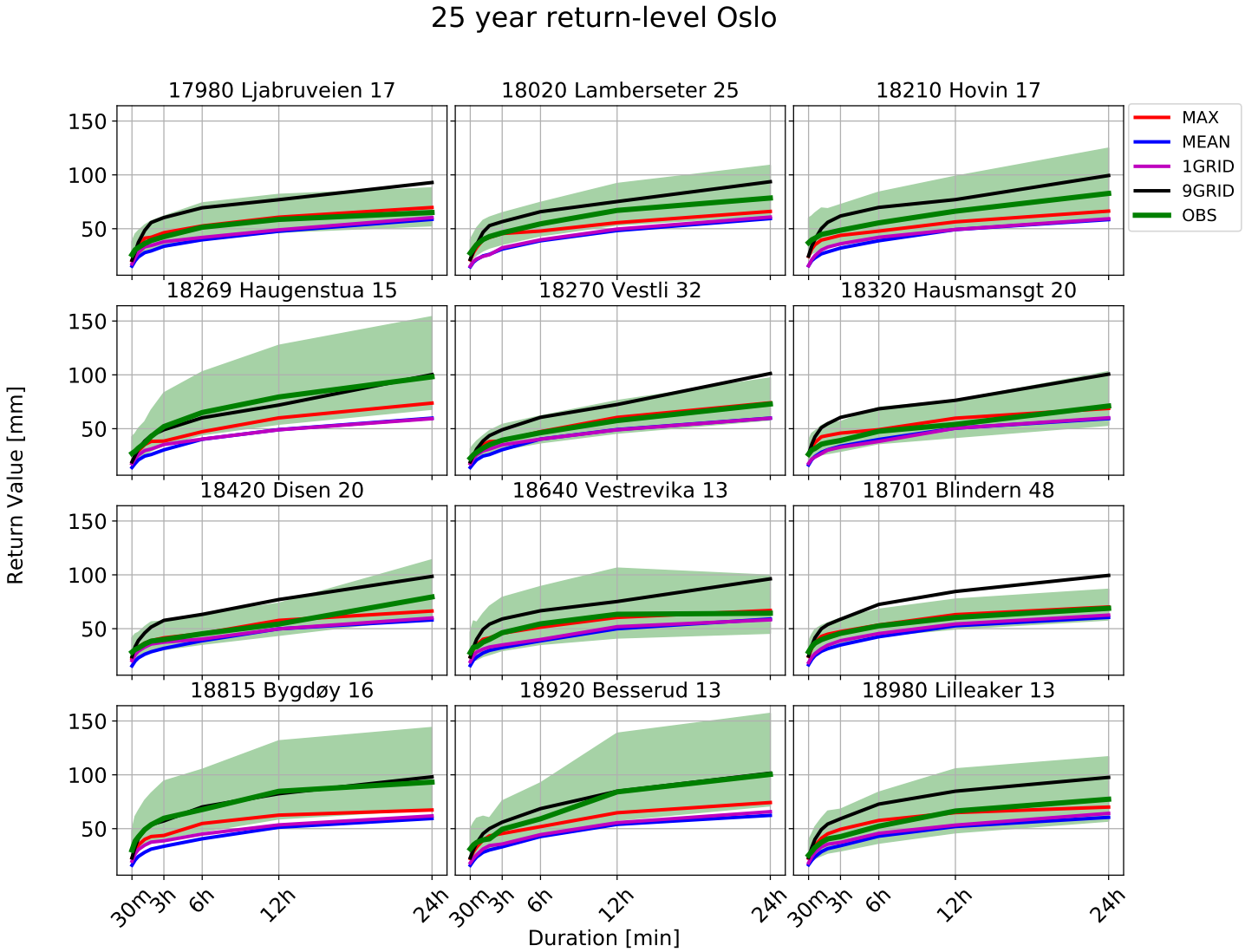

Figure 25: 25 year return-value for durations 30 min to 24 hours for the 12 stations in the Oslo area for the OBS, MAX, MEAN, 9GRID and 1GRID method. Shaded green area is the 95% confidence of OBS. units on the y-axis is mm and units on the x-axis is years. The station name, number and OBS time-series length is written above each subplot.**reload this figure. not supposed to be gray background. Easier to read with white background.**

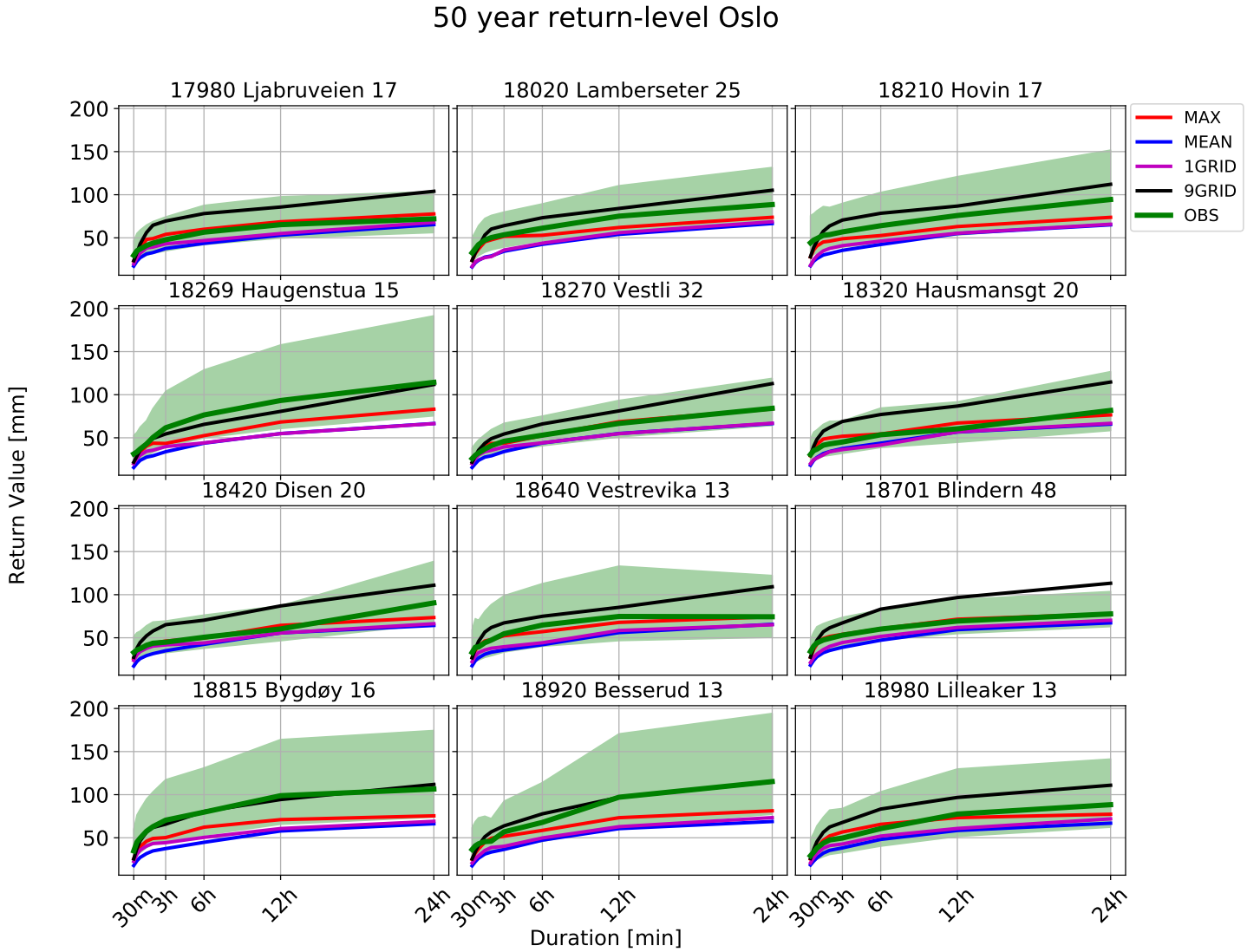

Figure 26: 50 year return-value for durations 30 min to 24 hours for the 12 stations in the Oslo area for the OBS, MAX, MEAN, 9GRID and 1GRID method. Shaded green area is the 95% confidence of OBS. units on the y-axis is mm and units on the x-axis is years. The station name, number and OBS time-series length is written above each subplot.**reload this figure. not supposed to be gray background. Easier to read with white background.**

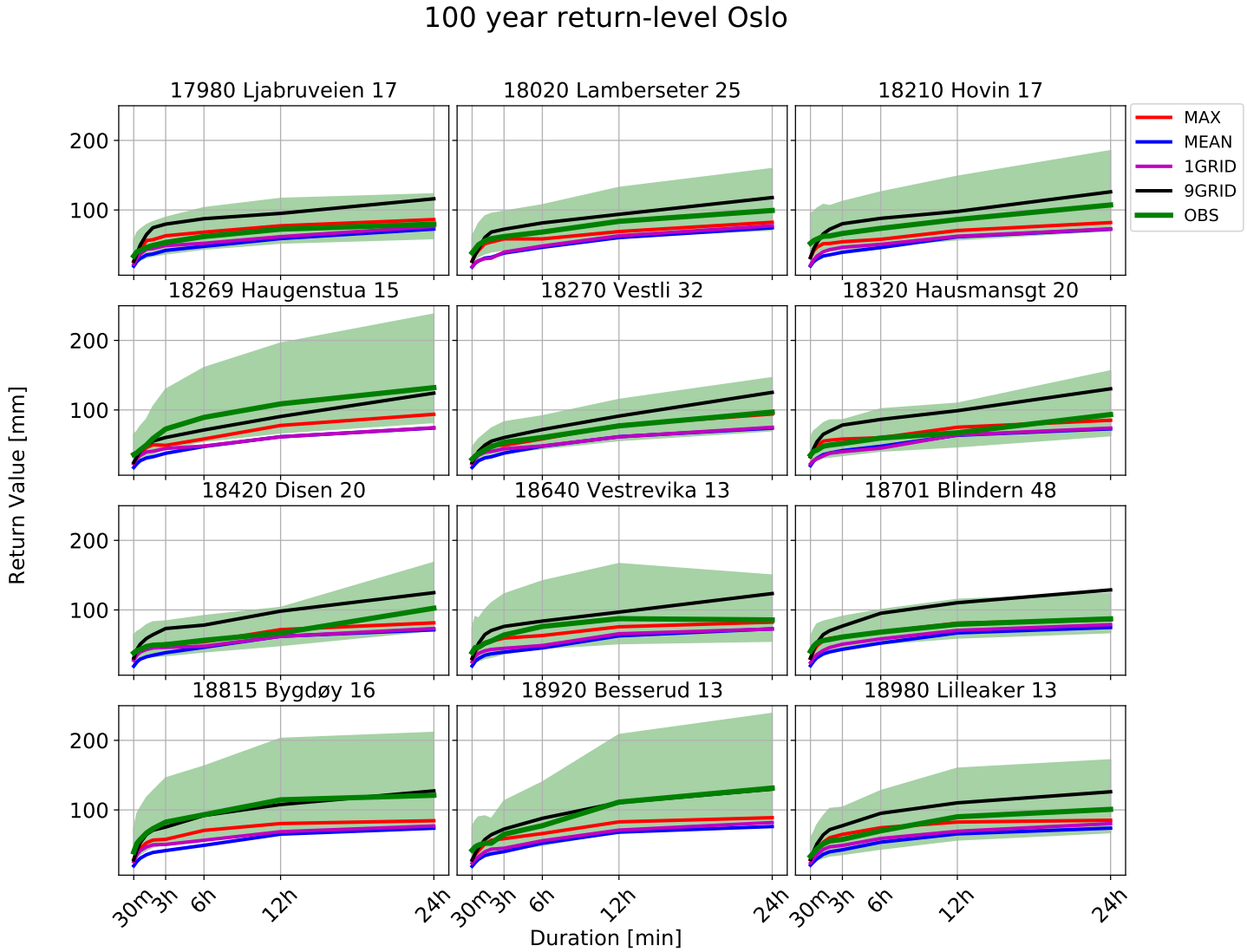

Figure 27: 100 year return-value for durations 30 min to 24 hours for the 12 stations in the Oslo area for the OBS, MAX, MEAN, 9GRID and 1GRID method. Shaded green area is the 95% confidence of OBS. units on the y-axis is mm and units on the x-axis is years. The station name, number and OBS time-series length is written above each subplot.**reload this figure. not supposed to be gray background. Easier to read with white background.**

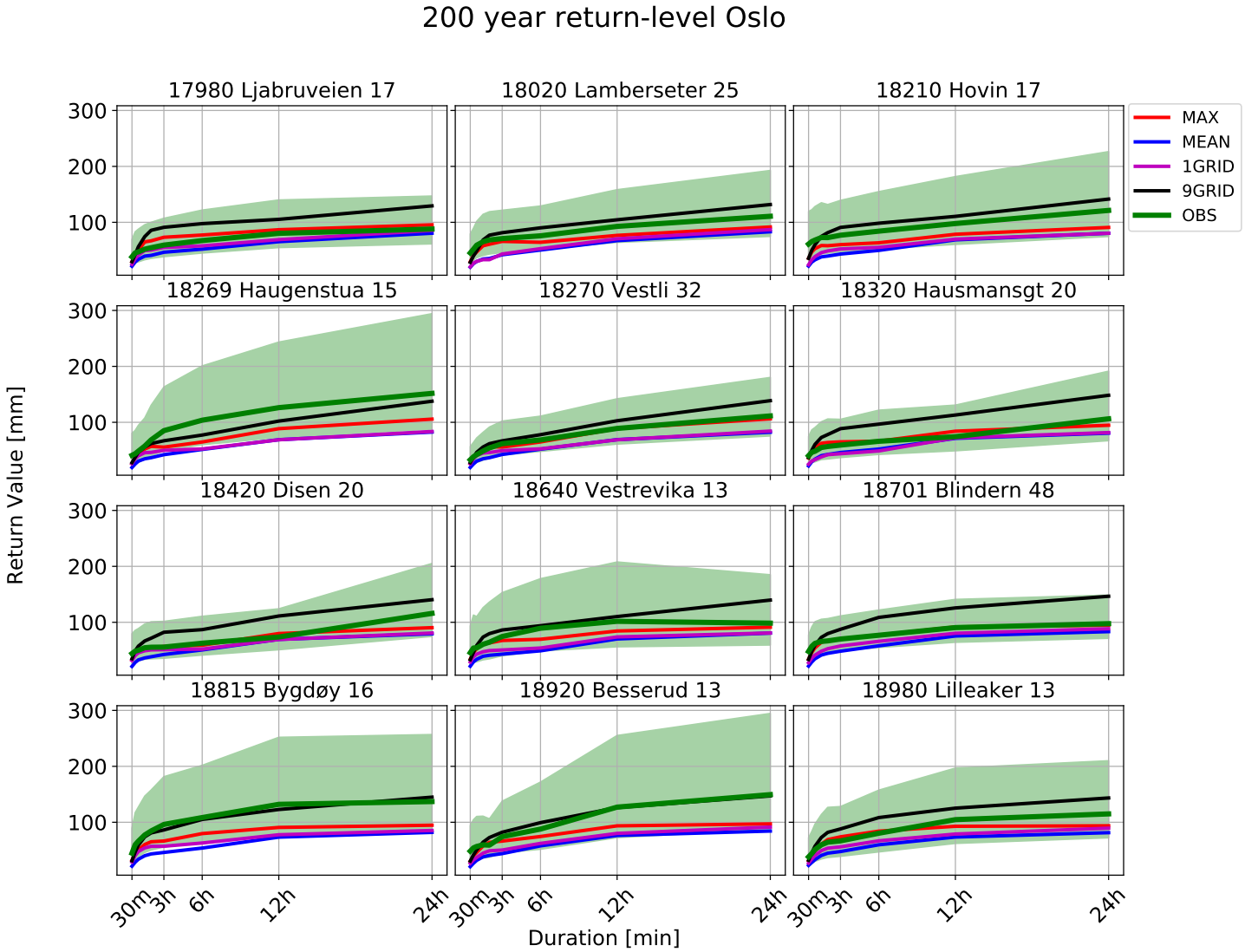

Figure 28: 200 year return-value for durations 30 min to 24 hours for the 12 stations in the Oslo area for the OBS, MAX, MEAN, 9GRID and 1GRID method. Shaded green area is the 95% confidence of OBS. units on the y-axis is mm and units on the x-axis is years. The station name, number and OBS time-series length is written above each subplot.**reload this figure. not supposed to be gray background. Easier to read with white background.**