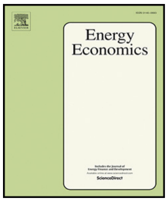
Energy Economics 117 (2023) 106421

Contents lists available at ScienceDirect

Energy Economics

journal homepage: www.elsevier.com/locate/eneeco


Energy Economics

A stochastic time-series model for solar irradiation

Karl Larsson [a,*], Rikard Green [b], Fred Espen Benth [c]

[a] School of Business, Örebro University, Fakultetsgatan 1, S-701 82 Örebro, Sweden
[b] Energy Quant Solutions Sweden AB, Ekelundsvägen 25, S-22472 Lund, Sweden
[c] Department of Mathematics, University of Oslo, P.O. Box 1053 Blindern, N-0316 Oslo, Norway

ARTICLE INFO

JEL classification:
C5
Q4

Keywords:
Solar
Time-series
Photovoltaic
Renewable
Energy
Risk

ABSTRACT

We propose a novel stochastic time series model able to explain the stylized features of daily irradiation level data in 5 cities in Germany. The model is suitable for applications to risk management of photovoltaic power production in renewable energy markets. The suggested dynamics is a low-order autoregressive time series with seasonal level given by an atmospheric clear-sky model. Moreover, we detect a skewness property in the residuals which we explain by a winter–summer regime switch. The stochastic variance is modeled by a seasonally varying GARCH-dynamics. The winter and summer standardized residuals are proposed to be a Gaussian mixture model to capture the bimodal distributions. We estimate the model on the observed data, and perform a validation study. An application to energy markets studying the production at risk for a PV-producer is presented.

1. Introduction

The rapid growth of solar energy production from photovoltaic technology brings new types of risk to the energy market. With significant volumes of electricity coming from solar power many energy systems are now exposed to the stochastic changes in irradiation levels primarily driven by fluctuating cloud cover. The situation is similar with other renewable energy sources most importantly power production from wind. Stochastic factors such as irradiation levels and wind speeds are intermittent, hard to predict and highly variable, and they bring weather related risk to the forefront of financial risk management in many energy markets. New financial instruments have been introduced to allow market participants to manage the related risks more efficiently. The growing share of renewables in the energy mix has also changed the way markets operate. The intraday market is much more active today with participants trading very short term contracts, for delivery within the hour, in order to manage sudden large changes in the volumes produced from renewable sources.

There is an emerging academic literature devoted to the challenges brought about by the increased dependency on renewable energy sources. Of particular interest for risk management purposes are time series models that are able to accurately represent the underlying processes as well as being useful for calculating prices and risks. The literature on solar energy modeling from this perspective is still rather scarce. Two recent studies that model solar energy production directly using data on production volumes from transmission system operator (TSO) areas in Germany are (Benth and Ibrahim, 2017; Lingohr and Müller, 2019). Aggregated production for large areas have very different dynamics compared to single locations. Production at single locations or smaller areas are however of critical interest in many applications. In addition, production data is only available for large areas. In order to assess the time series dynamics of solar energy production at specific coordinates a different approach is needed. A way forward is to instead model the irradiation level at a given coordinate directly and map it to energy output using a production function. This approach is also advocated by many market professionals, see e.g. De Jong (2020), and it is the one we pursue. Data on various measures of solar irradiation is available for arbitrary coordinates from several different data providers. We use data from CAMS (Copernicus Atmosphere Monitoring Service) which is based on satellite imagery and can be downloaded for any coordinate included in the spatial coverage of the service.[1] Europe is e.g. fully covered and we choose to work with data for locations in Germany. Production functions for power production using different photovoltaic technologies have been extensively studied in the associated engineering literature, see e.g. Huld et al. (2011) and Kaldellis et al. (2014).

The main contribution of this paper is a stochastic time series model for solar irradiation. Having a stochastic, and dynamically consistent, time series model is paramount for many applications and we therefore

[*] Corresponding author.
E-mail addresses: karl.larsson@oru.se (K. Larsson), rikard.green73@gmail.com (R. Green), fredb@math.uio.no (F.E. Benth).

[1] https://atmosphere.copernicus.eu/data.

https://doi.org/10.1016/j.eneco.2022.106421
Received 30 August 2021; Received in revised form 9 November 2022; Accepted 18 November 2022
Available online 25 November 2022

focus our attention on developing such a model. The model we propose is demonstrated to capture the most salient features of the time series dynamics of solar irradiation. Among these features are seasonal and auto-regressive effects in both the level and variance of irradiation. There are also seasonal dependencies in the distributional properties of data. Residuals are shown to follow distinctly different bimodal distributions in summer and winter. These distributions are fitted very closely by bimodal Gaussian mixture distributions. Our model is consistent with all these characteristics using only standard time series tools. The simplicity of the model makes it easy to estimate and evaluate which is appealing in both academic and applied work.

The paper (Casula et al., 2020) is similar to ours in some respects. The authors propose a stochastic time series model for solar irradiation and also considers financial applications. However, their model does not acknowledge the very strong seasonality in the conditional volatility and disregards the bimodality and seasonal dependency of the distribution for the residuals. We find that these features constitute important aspects of irradiation data. Accurate description of volatility and distributional properties are particularly important for applications in finance and should not be ignored.

We end the paper with an example application to the management of volume risk for a solar energy producer. There are many other applications that could be addressed with our model. In Cuppari et al. (2021) the authors bring attention to the prospect for landowners to reduce financial risk by co-locating agriculture and solar power production technology. Their analysis is built on an interesting application of diversification and takes a stochastic irradiation level as one model input. The articles (Benth and Ibrahim, 2017; Casula et al., 2020) both employ their models to the valuation of different types of options on the revenue stream of a solar park. The authors in Lingohr and Müller (2019) use their model to analyze a futures contract written on solar energy production. They envision a contract similar in spirit to the wind power production futures contracts traded on the European Energy Exchange (EEX). Another important application is investment decisions for the construction, and location, of new solar parks. Our model is perfectly suited for all these applications. It would also be of interest to develop multivariate time series models e.g. for wind speed, temperature, solar irradiation and electricity prices considered jointly. Such models should ideally build on, and be consistent with, the univariate time series dynamics of each component. Our model provides a natural benchmark for the irradiation part of such models. Detailed investigations of these interesting topics are left for future research.

2. The stochastic model and irradiation data

2.1. Description of data

We use data obtained from CAMS (Copernicus Atmosphere Monitoring Service) which is a part of the European Union's Earth observation programme and implemented by ECMWF (European Centre for Medium-Range Forecasts).[2] The data from CAMS is based on satellite imagery and model input to account for the impact of clouds and other atmospheric conditions. We refer to the CAMS user guide (CAMS, 2019; Gschwind et al., 2019; Lefevre et al., 2017; Qu et al., 2017) for detailed accounts of the methods used. The data is free of charge and can be downloaded for arbitrary coordinates given the spatial coverage of the service. One of the stated goals of the service is to provide radiation data accurate enough for scientific and commercial use in applications to solar energy production. The data is quality controlled and validated on a quarterly basis, see CAMS (2019). Data from CAMS has also been evaluated in several academic studies and fares well in comparison to other data sources, see e.g. Marchand et al. (2020) and Yang and Bright (2020). In Marchand et al. (2020) CAMS data is compared to the HelioClim-3 database, which is also based on satellite imagery, for a number of locations in Germany. Both data sources are found to be able to accurately represent the temporal and spatial variation in irradiation data. The authors point out that satellite based irradiation data has been shown to have a higher accuracy compared to data based on reanalysis (as e.g. the ERA5 and MERRA-2 databases) and that CAMS is a widely used and reliable data source.[3]

Table 1
Latitude and longitude for the locations.

Location	Latitude	Longitude
Hamburg	53.4361	9.6311
Berlin	52.6306	13.6263
Nürnberg	49.4073	10.9265
Stuttgart	48.9296	9.2896
München	48.4437	11.3660

While Germany has many weather stations that measure solar irradiation, only a few of them can provide high quality data, see e.g. Marchand et al. (2020) for a discussion of this point. In most practical cases for energy production the irradiation must be evaluated at coordinates without available high quality measurements. Hence we prefer to model satellite based data directly since it will be the most relevant data scenario for applications envisioned for our model. Our view is that working with this data provides the most information for potential users.

In this study we choose to model global horizontal irradiance (GHI) which is the radiation received on a horizontal plane from all directions. It is the most relevant measure for photovoltaic energy production and it is measured in Wh/m^2. We use data on GHI collected at 11:00 (UCT), around the peak time for solar intensity, for different coordinates in Germany. The locations of our selected coordinates are situated near Hamburg, Berlin, Stuttgart, Nürnberg and München. Our choice of coordinates gives a reasonable coverage of different parts of Germany and makes it possible to address differences between locations that are both close and far apart.

Our final dataset consists of 11 years of daily observations sampled at 11:00 pm (UCT) for each location during the period 2010-01-01 to 2020-12-31. Data from the 10 year period 2010-01-01 to 2019-12-31 is used as in-sample for estimation and analysis, and the last year (2020-01-01 to 2020-12-31) is used in an out-of-sample model prediction exercise as part of the model validation. Leap year days in 2012, 2016 and 2020 were removed. There are very few missing data points with a maximum of 13 out of 4015 observations for Hamburg and Stuttgart. For the out-of-sample period there are no missing observations. We replace missing values using linear interpolation between the nearest observed hours.

The coordinates for the selected locations are given in Table 1 and descriptive statistics are given in Table 2. The mean radiation increases with decreasing latitude which not unexpectedly tells that the more southern locations display higher average radiation levels. We also observe that the standard deviation is higher for the southernmost locations. Additionally, we notice that the locations situated more to the west has a slightly lower standard deviation than the eastern locations. The maximum and minimum values for the irradiation data are confirming higher irradiation for lower latitude. The skewness is positive for all locations but with the southern locations displaying lower values.

[2] Data from the CAMS service is provided by their CAMS data store found at https://atmosphere.copernicus.eu/data.

[3] The ERA5 and MERRA-2 databases are operated by ECMWF and NASA respectively and are both based on reanalysis. We refer to their respective websites at https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5 and https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/ for more information.

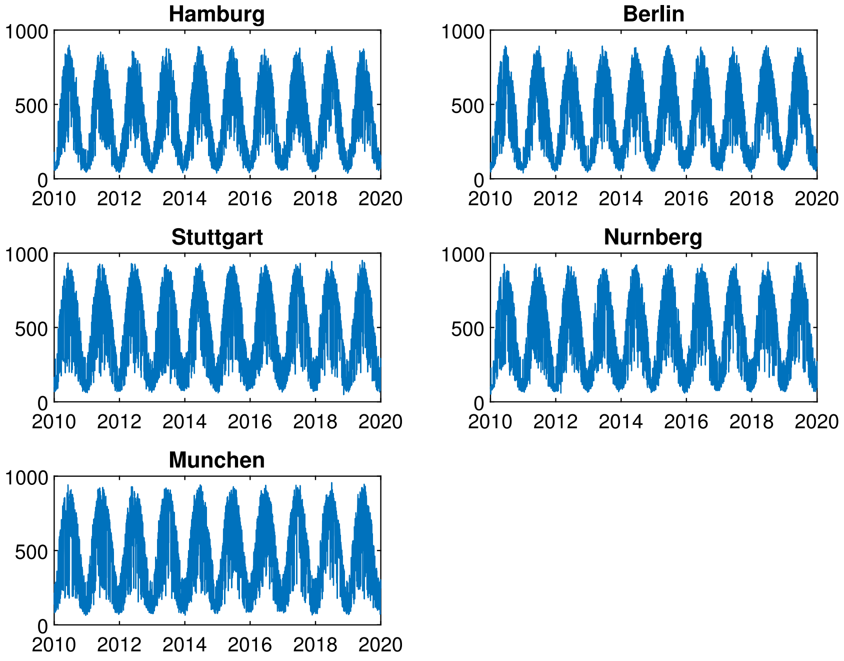


Fig. 1. Time series of irradiation $G(t)$.

Table 2
Descriptive statistics for the locations.

Location	Mean	Std.dev.	Max	Min	Skewness
Hamburg	382.3726	244.3257	897.8311	36.8759	0.3111
Berlin	393.1779	247.1781	896.1819	38.8503	0.2673
Nürnberg	450.5042	262.8204	951.5443	47.8573	0.1819
Stuttgart	434.7473	257.1569	940.1480	51.1439	0.2072
München	456.2307	264.8277	956.9816	63.8053	0.1753

Solar irradiation is influenced by astronomical and atmospherical conditions. Extraterrestrial radiation is the amount of irradiation before it reaches the atmosphere. It can be regarded as a deterministic function and is determined from the relative positions of the earth and the sun. For all our locations the extraterrestrial radiation peaks on June 20 and takes its lowest value on December 20. Extraterrestrial radiation is reduced through absorption and scattering by particles in the atmosphere. The irradiation levels reaching the ground is therefore substantially lower even under clear sky conditions. Due to random interference of clouds and particles irradiation levels at the ground are also stochastic. The cloud cover is a very important stochastic component of irradiation data and it may change drastically even over short periods of time. Such rapid changes in the cloud cover occur more frequently in the summer causing highly variable irradiation levels. The variations in summer are amplified by the seasonally higher levels; the appearance of sudden clouds may cause the irradiation to drop sharply from a very high level thus creating large swings. During the winter on the other hand, the variation is smaller due to periods of more persistent overcast skies and lower overall irradiation levels.

Fig. 1 plots the time series of irradiation values for our 10 year in-sample period and for all locations. The graphs show the strong seasonal pattern present in solar irradiation data. Much of the seasonality is determined from the deterministic extraterrestrial radiation but may also be influenced by seasonal factors in the atmosphere. It can also be noted from Fig. 1 that the summer periods display larger variations compared to the winter periods.

2.2. Time series dynamics

In a given location, let $G(t)$ be the GHI at times $t = 0, 1, \ldots$. We model $G(t)$ as the sum of a deterministic seasonal component $S(t)$ and a stochastic process $Z(t)$ as

$$G(t) = S(t) + Z(t). \tag{1}$$

It is natural to base the seasonal component $S(t)$ on the extraterrestrial radiation which is the amount of radiation reaching the outer boundary of the earth's atmosphere. The extraterrestrial radiation is determined by the earth's movement around the sun and can for all practical purposes be regarded as a deterministic function of time. At a given coordinate and time t the extraterrestrial radiation is given by

$$\Lambda(t) = K\left(1 + 0.033\cos\left(\frac{360t}{365}\right)\right)\cos(\theta(t)) \tag{2}$$

In Eq. (2), K is the solar constant which we set to $K = 1367$ (W/m^2) and $\cos(\theta(t))$ is the zenith angle at the selected coordinate at time t.[4] The function $\theta(t)$ is implicitly determined from the solar zenith angle given by

$$\cos(\theta(t)) = \cos(lat)\cos(w(t))\cos(\delta(t)) + \sin(lat)\sin(\delta(t)) \tag{3}$$

where $w(t)$ and $\delta(t)$ are the solar hour angle and declination, resp., and lat denotes the latitude of the selected coordinate. We refer to Appendix A on how to determine the solar hour angle and declination.

The seasonality in GHI is closely linked to that of the extraterrestrial radiation. Actual GHI may also be affected by seasonal components in the atmosphere and most importantly in the cloud cover. Since the extraterrestrial radiation is not affected by clouds and other atmospheric conditions it attains much larger values than actual GHI received at the ground level. This means that extraterrestrial radiation must be calibrated to GHI data $G(t)$ in order to achieve a proper seasonal adjustment. We propose to use the following seasonality function based on $\Lambda(t)$,

$$S(t) = \Lambda(t)\left(a_0 + a_1 e^{-a_2/\cos(\theta(t))}\right) \tag{4}$$

where a_i, $i = 0, 1, 2$ are constants and $\cos(\theta(t))$ is given in (3). The function (4) can be motivated e.g. by the simple parametric clear sky

[4] The solar constant actually varies to some degree, e.g. due to variations in the earth–sun distance, sun spot activity, etc. Different proposed values of the solar constant can be found in the literature. The differences are small and the specific value does not matter for our purposes since we will calibrate the seasonality function to data.

model presented in Hottel (1976). A parametric clear sky model is a deterministic model for radiation under the assumption of cloud free conditions but taking other atmospheric conditions into account, and should give an upper bound for the actual GHI. The clear sky model in Hottel (1976) has additional equations for determining the parameters a_0, a_1, a_2 but can be written on the simplified form (4). A clear sky model is calibrated to radiation data under clear sky conditions. However, our aim is not to recreate a clear sky model, we are instead using the functional form for $S(t)$ in order to achieve a viable seasonal adjustment of $G(t)$. This means that we can settle for the simpler specification in (4).

We model the seasonally adjusted series $Z(t) = G(t) - S(t)$ as an AR(p) process according to

$$Z(t) = \sum_{i=1}^{p} \beta_i Z(t-i) + u(t) \tag{5}$$

with constant coefficients β_i, $i = 1, \dots, p$. The modeling approach in Eqs. (1) and (5) is strongly supported by the theoretical and empirical analysis in Benth and Šaltytė Benth (2012). With this specification $S(t)$ becomes the seasonal mean function that the irradiation level $G(t)$ mean reverts to which is an appealing property. The residual process $u(t)$ in (5) is specified as

$$u(t) = \sigma(t)e(t) \tag{6}$$

where $\sigma(t)$ is a time varying conditional volatility. We find that the conditional variance of $u(t)$ is heteroscedastic, containing both seasonal and ARCH effects. There are physical reasons for expecting volatility clustering in irradiation data. Scattered and rapidly moving clouds will cause more variation in solar irradiation and can persist for some days. A thick cloud cover on the other hand will lead to less variation that can persist for a few days. These effects could be sources of volatility clustering. We account for seasonality and volatility clustering by specifying the conditional variance as

$$\sigma^2(t) = \sigma_S^2(t)\sigma_G^2(t) \tag{7}$$

with a deterministic seasonal component $\sigma_S^2(t)$ multiplied by a GARCH-type component $\sigma_G^2(t)$. We are using a multiplicative structure similar to Benth and Šaltytė Benth (2012). The seasonal component $\sigma_S^2(t)$ is given by a truncated Fourier series containing three terms,

$$\sigma_S^2(t) = c_0 + c_1 \cos\left(\frac{2\pi t}{365}\right) + c_2 \sin\left(\frac{2\pi t}{365}\right) \tag{8}$$

This formulation is widely used for seasonal volatility modeling, see e.g. Campbell and Diebold (2005), Härdle and Lopez Cabrera (2012) and Benth and Šaltytė Benth (2012, 2013) for applications to temperature and wind speed modeling, and Benth et al. (2008) for applications to energy prices. The parameter restriction $c_0 > \sqrt{c_1^2 + c_2^2}$ ensures that the seasonal variance $\sigma_S^2(t)$ is positive. See Appendix B for details. It is possible, and straightforward, to include more trigonometric terms in (8) but we find that three terms are sufficient. Extending the Fourier series beyond three terms did not result in a substantially different seasonality curve for the variance. In order to keep the number of parameters low we settled for the three term model presented here. The auto-regressive GARCH component $\sigma_G^2(t)$ is assumed to take the form of a standard GARCH(1,1)-process based on the residual series $v(t) = u(t)/\sigma_S(t)$ standardized by the seasonal component. The specification becomes

$$\begin{aligned} \sigma_G^2(t) &= \omega_0 + \omega_1\sigma_G^2(t-1) + \omega_2 v^2(t-1) \\ &= 1 - \omega_1 - \omega_2 + \omega_1\sigma_G^2(t-1) + \omega_2 v^2(t-1) \end{aligned} \tag{9}$$

The unconditional variance of the GARCH part is given by $\omega_0(1 - \omega_1 - \omega_2)^{-1}$ and should be equal to one since it operates on the partly standardized residuals $v(t)$. We therefore employ the parametrization $\omega_0 = 1 - \omega - \omega_2$ which ensures this property and in addition avoids the issue of identifying two separate intercepts in $\sigma_S^2(t)$ and $\sigma_G^2(t)$.

The final model component left to specify is the driving noise process $e(t)$. We find that the irradiation time series display seasonal behavior also in the third moment. The skewness of irradiation data has a clear seasonal pattern turning negative during the summer period from being positive in winter. Our empirical results show that this effect is important for both the understanding of irradiation data and for model fit. The actual definitions of the summer and winter spans are determined from empirical considerations. All our locations display similar patterns and it is a feature that we have not seen documented in other studies. It is common to find bimodality in the distribution of high frequency radiation data. The bimodal property is commonly explained as representing states of either cloudy or clear sky conditions. To account for both the effects of seasonal skewness and bimodality we model the process $e(t)$ as

$$e(t) = d(t)\epsilon_S(t) + (1 - d(t))\epsilon_W(t) \tag{10}$$

where $d(t)$ is a simple dummy-process that takes the value 1 in summer, and the value 0 in winter, and the processes ϵ_k, $k \in \{S, W\}$, are assumed to be independent *iid* sequences of Gaussian mixture distributed random variables. The probability densities for ϵ_k, $k \in \{S, W\}$ are given by,

$$f_k(x) = q_k f_{1,k}(x) + (1 - q_k) f_{2,k}(x), \quad x \in \mathbb{R}, \quad k \in \{S, W\} \tag{11}$$

where $f_{m,k}(x)$ are the densities of Gaussian random variables with means $\mu_{m,k}$ and variances $v_{m,k}^2$ for $m = 1, 2$ and $k \in \{S, W\}$. We use the notation $\mathrm{GM}(\mu_1, \mu_2, v_1^2, v_2^2, q)$ for such a Gaussian mixture distributions with two components. The parameter q is the prior probability of being in the first state $m = 1$. With this notation we thus have

$$\epsilon_k(t) \sim \mathrm{GM}(\mu_{1,k}, \mu_{2,k}, v_{1,k}^2, v_{2,k}^2, q_k), \quad k \in \{S, W\} \tag{12}$$

The moments of Gaussian mixture distributions for $\epsilon_S(t)$ and $\epsilon_W(t)$ are easily calculated given the probability densities (11). We introduce the following notation for the expected value, variance and the centralized third moment of $\epsilon_k(t)$, $k \in \{S, W\}$,

$$\begin{aligned} \mu_k &= \mathbb{E}\left[\epsilon_k(t)\right] \\ \Sigma_k^2 &= \mathrm{Var}\left(\epsilon_k(t)\right) \\ \Omega_k &= \mathbb{E}\left[\left(\epsilon_k(t) - \mathbb{E}\left[\epsilon_k(t)\right]\right)^3\right] \end{aligned}$$

and state, without proof, the following explicit expressions,

$$\begin{aligned} \mu_k &= q_k\mu_{1,k} + (1 - q_k)\mu_{2,k} \\ \Sigma_k^2 &= q_k\left(v_{1,k}^2 + \lambda_{1,k}^2\right) + (1 - q_k)\left(v_{2,k}^2 + \lambda_{2,k}^2\right) \\ \Omega_k &= q_k\left(3\lambda_{1,k}v_{1,k}^2 + \lambda_{1,k}^3\right) + (1 - q_k)\left(3\lambda_{2,k}v_{2,k}^2 + \lambda_{2,k}^3\right) \end{aligned}$$

where $\lambda_{m,k} = \mu_{m,k} - \mu_k$, for $m = 1, 2$. The skewness of $\epsilon_k(t)$ is given by Ω_k/Σ_k^3.

Given the independence between $\epsilon_S(t)$ and $\epsilon_W(t)$ it follows that the corresponding moments of $e(t)$ are given by

$$\begin{aligned} \mathbb{E}\left[e(t)\right] &= d(t)\mu_S + (1 - d(t))\mu_W \\ \mathrm{Var}\left(e(t)\right) &= d(t)\Sigma_S^2 + (1 - d(t))\Sigma_W^2 \\ \mathbb{E}\left[(e(t) - \mathbb{E}\left[e(t)\right])^3\right] &= d(t)\Omega_S + (1 - d(t))\Omega_W \end{aligned}$$

Note that with $\mu_S = \mu_W = 0$ and $\Sigma_S^2 = \Sigma_W^2 = 1$ we also have $\mathbb{E}\left[e(t)\right] = 0$ and $\mathrm{Var}\left(e(t)\right) = 1$ for all t. With $\mathrm{Var}\left(e(t)\right) = 1$ the skewness of $e(t)$ is

$$\mathrm{Skew}\left(e(t)\right) = d(t)\Omega_S + (1 - d(t))\Omega_W \tag{13}$$

which is time-varying and provides a natural channel for explaining the skewness patterns observed over the year. This completes our description of the model. In the next section we demonstrate the models ability to account for the many particular features of irradiation time series data.

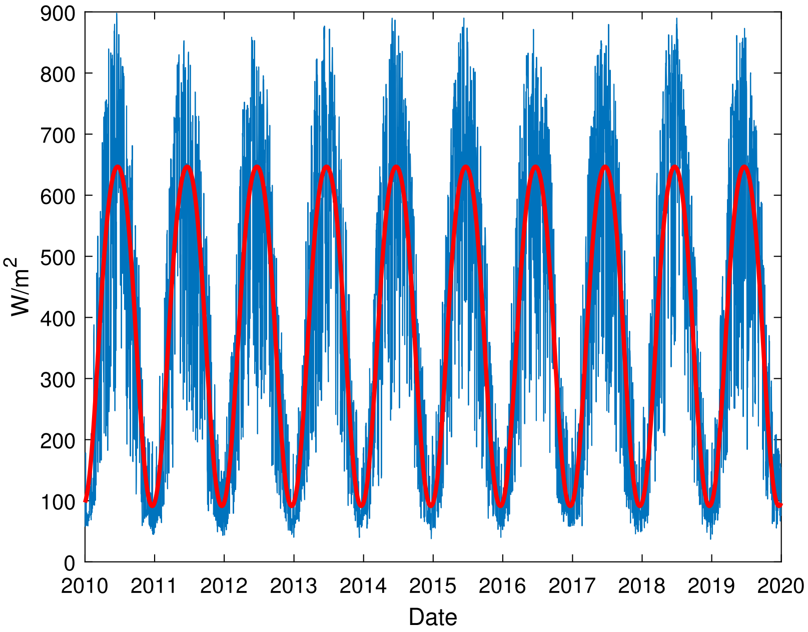


Fig. 2. Irradiation $G(t)$ (blue) and fitted seasonal function $S(t)$ (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

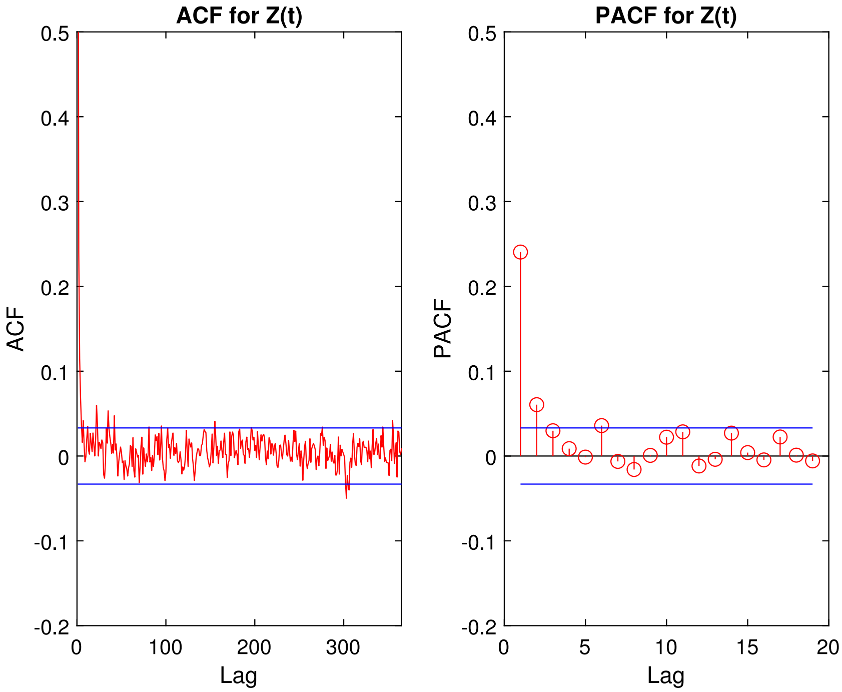


Fig. 3. Autocorrelation function (ACF) (left panel) and partial autocorrelation function (PACF) (right panel), for the seasonal adjusted series $Z(t) = G(t) - S(t)$.

3. Estimation and validation

3.1. Estimation

We propose a stepwise estimation strategy that allows for verification of crucial model properties at each stage. Stepwise estimation strategies have been used by many other authors in similar models, see e.g. Alaton et al. (2002), Härdle and Lopez Cabrera (2012), Campbell and Diebold (2005) and Benth and Šaltytė Benth (2012, 2013). Using stepwise estimation has many advantages, being fast and easy to perform. Since standard errors are not obtained from a single objective function confidence intervals must be treated with some caution. However, checking crucial model properties at each step shows that the estimated model performs well in capturing stylized features of the data. For ease of exposition we illustrate the estimation using data only for the Hamburg location. All parameter estimates for Hamburg are presented in Table 3. The corresponding results for the other locations are presented in Table 4 in Section 3.2. The features observed at the Hamburg location are similar to the other locations.

The first step is to calibrate the seasonal function $S(t)$ given in (4) to data $G(t)$. From (4) we have

$$\frac{S(t)}{\Lambda(t)} = a_0 + a_1 \exp\left(-a_2/\cos(\theta(t))\right)$$

Table 3
Parameter estimates and standard errors for all model parameters for the Hamburg location. The parameters c_0, c_1, c_2, and their standard errors, have been divided by 100.

Parameter	Estimate	Std.Err.	Parameter	Estimate	Std.Err.
a_0	0.2003	0.0374	$\mu_{1,S}$	−1.0407	0.0350
a_1	0.5993	0.0172	$\mu_{2,S}$	0.6688	0.0192
a_2	0.4270	0.0779	$v^2_{1,S}$	0.3703	0.0279
β_1	0.2259	0.0143	$v^2_{2,S}$	0.2606	0.0132
β_2	0.0605	0.0141	q_S	0.3912	0.0149
c_0	191.0228	4.4450	$\mu_{1,W}$	−0.7188	0.0115
c_1	−173.1130	4.6303	$\mu_{2,W}$	0.9479	0.0370
c_2	36.5643	2.6998	$v^2_{1,W}$	0.1341	0.0062
ω_1	0.6165	0.1277	$v^2_{2,W}$	0.5601	0.0376
ω_2	0.0798	0.0164	q_W	0.5687	0.0135

Table 4
Parameter estimates and standard errors for all locations except Hamburg. The corresponding results for Hamburg are given in Table 3. In this table the parameters c_0, c_1 and c_2 have been divided by 100.

Parameter	Berlin	Stuttgart	Nürnberg	Nürnberg
a_0	0.0927	−0.2316	−0.2878	0.0019
	(0.0771)	(0.4741)	(0.4525)	(0.2100)
a_1	0.6851	1.0084	1.0565	0.8149
	(0.0492)	(0.4383)	(0.4195)	(0.1678)
a_2	0.3168	0.1777	0.1725	0.2809
	(0.0771)	(0.1216)	(0.1069)	(0.1287)
β_1	0.2686	0.2467	0.3234	0.2728
	(0.0140)	(0.0143)	(0.0143)	(0.0140)
β_2	0.0532	0.0641	–	–
	(0.0139)	(0.0144)	–	–
c_0	190.1884	261.4136	231.7968	273.2423
	(4.8957)	(6.3013)	(5.9295)	(7.3385)
c_1	−172.1055	−210.5509	−192.4294	−217.5883
	(5.1640)	(7.1180)	(6.6297)	(8.0930)
c_2	32.7133	37.2395	35.7742	50.6053
	(2.9577)	(4.6673)	(4.2530)	(5.2746)
ω_1	0.6829	0.6103	0.5823	0.5103
	(0.1188)	(0.0838)	(0.1522)	(0.0697)
ω_2	0.0794	0.0963	0.0921	0.1477
	(0.0183)	(0.0155)	(0.0200)	(0.0173)
$\mu_{1,S}$	−0.9814	−1.1085	−1.0022	−1.0530
	(0.0443)	(0.0363)	(0.0499)	(0.0442)
$\mu_{2,S}$	0.6785	0.6423	0.6576	0.6465
	(0.0222)	(0.0125)	(0.0199)	(0.0140)
$v^2_{1,S}$	0.4399	0.4863	0.4980	0.5428
	(0.0359)	(0.0353)	(0.0423)	(0.0421)
$v^2_{2,S}$	0.2603	0.1724	0.2372	0.1814
	(0.0142)	(0.0074)	(0.0124)	(0.0083)
q_S	0.4087	0.3669	0.3962	0.3804
	(0.0187)	(0.0124)	(0.0188)	(0.0149)
$\mu_{1,W}$	−0.6868	−0.6945	−0.6190	−0.7008
	(0.0138)	(0.0128)	(0.0163)	(0.0125)
$\mu_{2,W}$	0.8987	1.0679	1.1079	1.0210
	(0.0522)	(0.0267)	(0.0432)	(0.0319)
$v^2_{1,W}$	0.1549	0.1830	0.2242	0.1615
	(0.0082)	(0.0080)	(0.0104)	(0.0071)
$v^2_{2,W}$	0.6790	0.3722	0.4731	0.4616
	(0.0521)	(0.0236)	(0.0384)	(0.0306)
q_W	0.5668	0.6059	0.6416	0.5930
	(0.0184)	(0.0111)	(0.0152)	(0.0123)

where $\Lambda(t)$ is the extraterrestrial radiation in (2). We use this form and calibrate the parameters a_0, a_1, a_2 by minimizing the function

$$Q(a_0, a_1, a_2) = \sum_{t=1}^{T} \left(\frac{G(t)}{\Lambda(t)} - \frac{S(t)}{\Lambda(t)} \right)^2.$$

with non-linear least squares. Fitting the parameters to $G(t)/\Lambda(t)$ instead of directly to $G(t)$ helps produce more precise estimates. The quantity $G(t)/\Lambda(t)$ is sometimes referred to as the clearness index and in some studies it is used as the only seasonal adjustment which may be sufficient for very short periods of time or when observations are aggregated and studied on lower time frequencies. This approach is not sufficient to remove the seasonality from our data which is on a high frequency and not in aggregated form. The irradiation time series $G(t)$ together with the fitted seasonal curve $S(t)$ are plotted in Fig. 2. The function $S(t)$ is capturing the seasonality in the level of $G(t)$ very well. Estimates of a_0, a_1, a_2 are presented in Table 3. The mean of the irradiation $G(t)$ is 382.37 and the mean of the calibrated function $S(t)$ is 382.44. Fig. 3 shows the autocorrelation function (ACF) and partial autocorrelation function (PACF) for the seasonally adjusted series $Z(t) = G(t) - S(t)$. There are no signs of seasonal dependency in the ACF for $Z(t)$ showing that the seasonal level adjustment performs well.

The ACF for $Z(t)$ suggests that a low order AR-structure is a suitable model structure. This is confirmed by the PACF which indicates that an order of $p = 2$ is sufficient to describe the autocorrelation structure. Before estimating the AR-parameters in (5) we subtract the mean of the seasonally adjusted series $Z(t)$. Any remaining mean is very small but may vary between locations depending on the seasonal fit. With an order of $p = 2$ for the zero mean AR-model for $Z(t)$ we need to estimate the parameters β_1 and β_2. This is done using a standard least squares fit. The estimates for these parameters can be found in Table 3 together with their estimated standard errors. The AR-parameters β_1 and β_2 are significant and positive.

From the estimated AR-model part $Z(t)$ we can now obtain observations of the process $u(t)$ from (5). The ACFs for $u(t)$ and $u^2(t)$ are plotted in Fig. 4. There are no seasonal and serial dependencies in the ACF for $u(t)$. The residuals $u(t)$ also passes the Ljung–Box test with a p-value of 0.4001. The ACF for $u^2(t)$ on the other hand shows strong seasonal variation. There are also low order ARCH effects but these are hard to glance from Fig. 4. Both these effects are accounted for by the volatility structure specified by the conditional variance in (7). We estimate the parameters c_i, $i = 0, 1, 2$ of the seasonal part $\sigma^2_S(t)$, and the parameters ω_1 and ω_2 of the GARCH part $\sigma^2_G(t)$ jointly using quasi-maximum-likelihood based on a Gaussian underlying distribution for $u(t)$. Then we fit the Gaussian mixture distributions to the fully standardized residuals $e(t) = u(t)/\sigma(t)$. In a final step we re-estimate the variance parameters from maximum likelihood based on the fitted Gaussian mixture distributions as recommended e.g. by Engle and Gonzlaez-Rivera (1991). This step does not change the parameter estimates for the dominating seasonal variance much and overall model performance is unaffected. Before discussing the estimation of the distributional part of the model in more detail we first examine the observed residuals $e(t)$ and the variance parameters.

From the estimated conditional volatility $\sigma(t)$ we can now study the observed residual series $e(t)$. Fig. 4 shows the ACFs of $e(t)$ and $e^2(t)$. As can be seen, both the seasonality and ARCH-effects have been removed. Judging from these graphs our model adequately describes seasonal and auto-regressive components present both in the conditional mean and variance of radiation data. The estimates of c_i, $i = 0, 1, 2$, are presented in Table 3. The estimate of c_0 corresponds well to the empirical standard deviation of $Z(t)$; $\sqrt{c_0}$ is estimated equal to 138.21 which can be compared to a standard deviation of 138.9 for $Z(t)$. The estimates of c_1 and c_2 multiplying the trigonometric terms in (8) are of different signs and magnitude producing a slight shift away from a symmetric seasonal pattern for the year. The seasonal variance attains its maximum in late June and its minimum in late December. The parameters ω_1 and ω_2 are both significant and fulfill the standard conditions for positivity and stationarity for a GARCH(1,1) process. We remark that the impact of the GARCH-part is low compared to the seasonal variance. As mentioned there are physical reasons to expect clustered variance, and we do detect some large autocorrelations in squared residuals and receive significant parameter estimates, which motivates inclusion of the GARCH-part.

To motivate our model for $e(t)$ in (10) we have plotted the observed residual series $e(t)$ in Fig. 5. The left panel shows the actual series for $e(t)$ and the right panel shows a simulated series from our estimated model. The time series plot of $e(t)$ in the left panel indicate that there

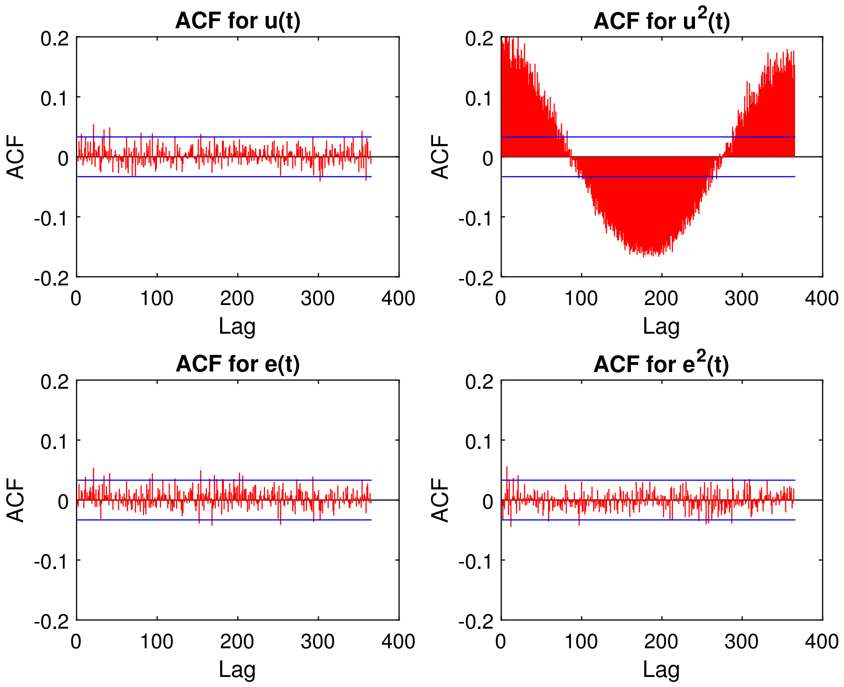


Fig. 4. Autocorrelation functions (ACF) for $u(t)$ (upper left), $u^2(t)$ (upper right), $e(t)$ (lower left), and $e^2(t)$ (lower right).

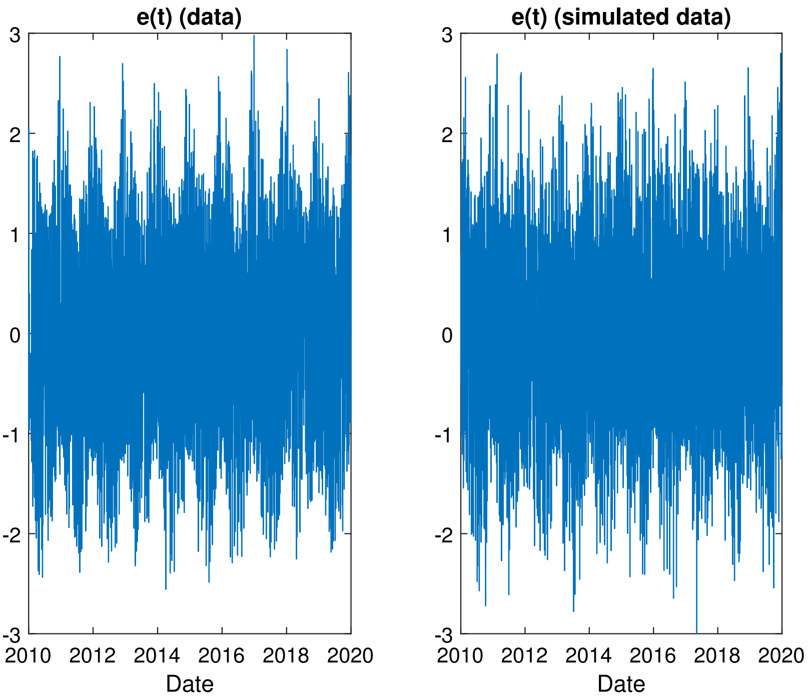


Fig. 5. Time series plots of observed $e(t)$ (left panel) and simulated $e(t)$ (right panel).

are still seasonal effects left, despite having successfully adjusted for seasonality both in the level and conditional variance of data. The right panel, with simulated data, shows that our model is consistent with this behavior.

In Fig. 6 we have plotted the ACF for $e^3(t)$ for both the actual series (left panel), and for simulated data from the estimated model (right panel). The left panel indicates a weak but clear seasonal pattern in the ACF over the year. Again, the right panel, based on simulated data, demonstrates our models ability to capture also this feature.

The remaining seasonality in $e(t)$, demonstrated in Figs. 5 and 6, is related to time variation in the skewness of irradiation data. As explained below the seasonal variation in skewness is due to the fact that the distributions for the residuals are distinctly different in the summer and winter periods. Fig. 7 plots the monthly empirical estimates of the skewness for $G(t)$ and $e(t)$ (left panel) and the same estimate for $e(t)$ together with the skewness function from the estimated model (right panel). The monthly skewness estimates are calculated from grouping data according to month for all observations during

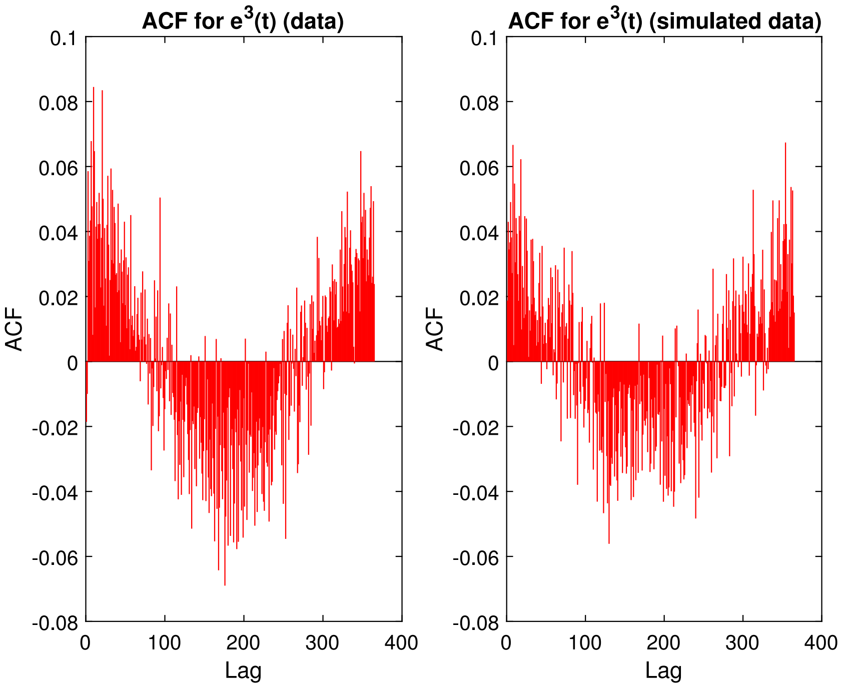


Fig. 6. Autocorrelation functions (ACF) for $e^3(t)$ based on observed (left panel) and simulated (right panel) data.

our 10 year sample. We included skewness estimated for both the original radiation time series $G(t)$ and the residuals $e(t)$ in the left panel to show that the time variation in skewness is not an artifact of the model steps preceding the construction of $e(t)$. The small differences in skewness estimates between $G(t)$ and $e(t)$ are due to the seasonal adjustment function $S(t)$. The estimates based on $e(t)$ show that the skewness changes in a periodical pattern over the year. It is positive in the months November–February and turns negative in the months March–October. This seasonal time variation in skewness is a distinct feature of irradiation data. It is observed at all our locations and we have not seen it documented in the related literature before. The reason for these differences are likely to be related to different cloud conditions during the year with more sustained periods of overcast skies in winter. This explanation is very much in line with the results we find below when we proceed to fit different Gaussian mixture distributions to the summer and winter residuals.

Since the months with positive/negative skewness are the same across locations we label them as summer (March-October) and winter (November-February). The skewness estimates displayed in Fig. 7 show that the shift from positive to negative, and back again, occur quite rapidly during the spring and autumn seasons. In between the shifts the skewness varies much less. This behavior motivates our simple definition of the summer and winter without sacrificing any crucial information. We have tested basing the summer definition instead on the period where the extraterrestrial radiation exceeds its mean level. Since extraterrestrial radiation is deterministic this gives the same summer period for any year and does not depend on data. This led to a summer definition that again is the same across locations (up to a maximum difference of one day) and gave similar empirical results. We prefer to use our empirically motivated definition since it is the same for all locations, transparent, and yielding a slightly better model fit.

In the right panel of Fig. 7 we show the skewness function (13) implied by our model compared to the skewness estimates from residual data $e(t)$. Our model generates a skewness function that is constant in the summer and winter periods, and explains rather well the overall skewness pattern of the data. Comparing the model implied skewness with the empirical estimates for the summer and winter periods we regard our simple skewness model to reach a satisfactory level of fit. The empirical skewness for summer is −0.3946 and our model gives a skewness of −0.3929. For the winter period the empirical estimate is 0.6785 and the model gives 0.6704. Moreover, the comparisons between the observed and simulated $e(t)$ in Figs. 5 and 6 show that our simple model structure is in close alignment with data.

The time variation in skewness is explained by the fact that the distributions for irradiation are very different in the winter and summer periods. A typical feature of high frequency irradiation data is that the distribution is bimodal. In Fig. 8 we plot the histogram of the residual series $e(t)$ where we have superimposed a standard kernel density estimate using a Gaussian kernel with correspondingly optimized bandwidth. The bimodal feature is clearly displayed. All our data series show similar behavior. The shape of the histogram of $e(t)$ reveals an important characteristics of the data and it is one we would like to be able to recreate with our model. However, as observed in our previous discussion the residual series $e(t)$ is not an *iid* sequence and hence the histogram is not a valid representation of the actual distribution of $e(t)$. Fitting a Gaussian mixture distribution directly to $e(t)$ would provide an accurate recreation of the histogram, however, due to the temporal dependence in the data series, it would not be consistent. By instead separating $e(t)$ into summer and winter residuals our model can accomplish this is in a fully consistent way.

We assume Gaussian mixture distributions with two components for both the summer and winter residuals. The parameters of these distributions are estimated using the EM-algorithm where the k-means algorithm is employed to initialize a grouping of data into the different components. This is a standard method of estimating Gaussian mixture distributions, see e.g. McLachlan and Peel (2000), and we have used the Matlab routine *fitgmdist*. The estimates are presented in Table 3. In our model setup we fit different Gaussian mixture distributions for the summer and winter residuals. At the model level these distributions should have zero expected value and unit variance to avoid introducing a small time-variation in the mean and variance. In empirical studies there is nearly always some mismatch between the empirical moments of a residual series and the corresponding moments implied by the fitted distribution. Such deviations usually have no empirical consequences

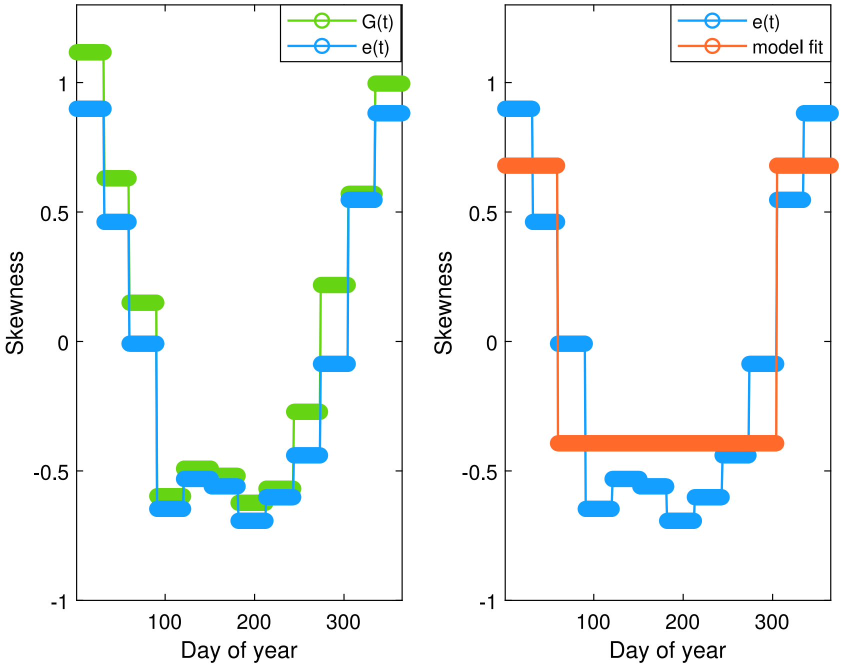


Fig. 7. Empirical skewness per month plotted against the day of year. Left panel plots skewness estimated for radiation data $G(t)$ (green) and for residuals $e(t)$ (blue). Right panel show estimated skewness for $e(t)$ (blue) and the skewness function obtained from the estimated model (orange). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

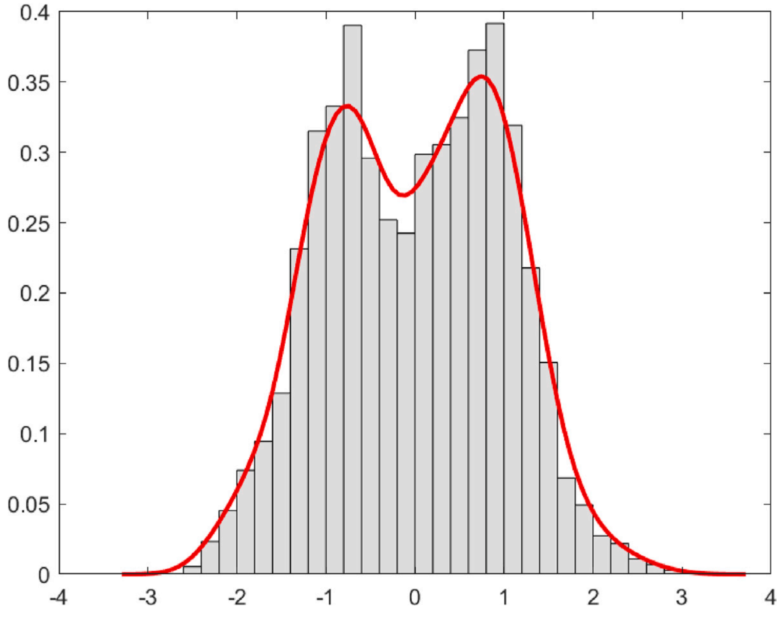


Fig. 8. Histogram (normalized) for residuals $e(t)$ and corresponding kernel density estimate (red curve).

since they typically are very small. We nevertheless employ a simple practical procedure to ensure that the estimated distributions have expected value of zero and unit variance. The EM-algorithm generates parameter estimates that closely matches the first three moments of the residuals. In order to cancel out any small deviations from zero mean and unit variance in the data, we estimate the Gaussian mixtures on residuals modified by an extra standardization using the empirical mean and standard deviation. By virtue of the EM-algorithm this minor adjustment gives us parameter estimates that imply theoretical zero expected values and unit variances to very close approximation. This procedure have negligible effects on data since the mean and variance are already close to zero and one, and it does not affect the skewness. We have compared estimation results obtained with and without this extra standardization and the differences in parameter estimates are very small and does not affect the distributional fit to any significant degree. We still keep it since it gives consistency with regard to the model and its components. The procedure ensures that the estimated distributions for $\epsilon_S(t)$ and $\epsilon_W(t)$ have expected values of zero and unit variances, and hence by implication that also $\mathbb{E}[e(t)] = 0$ and $\mathrm{Var}\,(e(t)) = 1$.

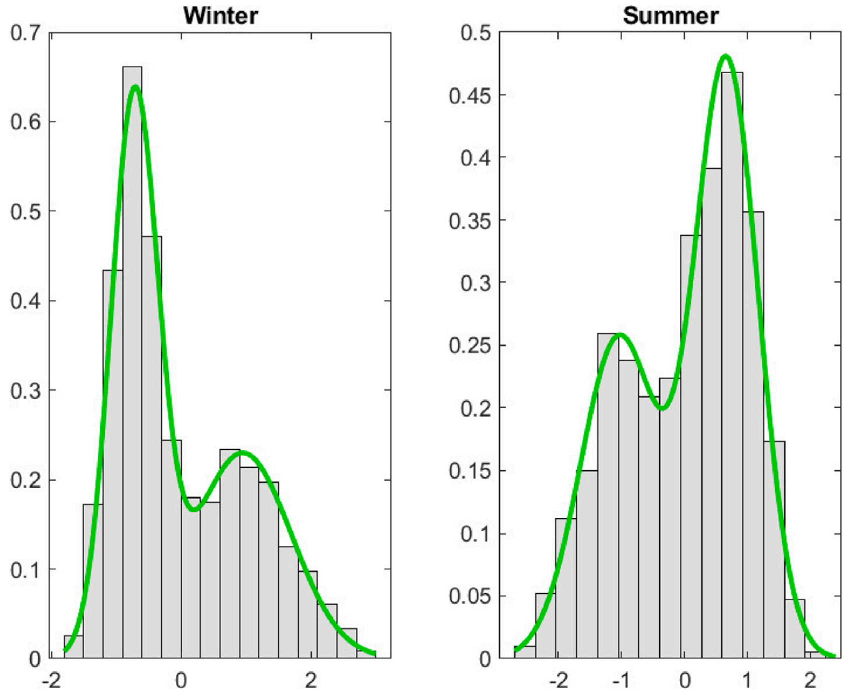


Fig. 9. Histogram (normalized) for winter (left panel) and summer (right panel) residuals $\epsilon_W(t)$ and $\epsilon_S(t)$ respectively.

The random variables $\epsilon_S(t)$ and $\epsilon_W(t)$ follow Gaussian mixture distributions with two components and thus have the stochastic representations,

$$\epsilon_k = B_k Y_{1,k} + (1 - B_k) Y_{2,k}, \quad k \in \{S, W\}$$

where B_k is a Bernoulli random variable with parameter q_k (i.e. $\mathbb{P}(B_k = 1) = q_k$), and the components $Y_{1,k}$ are Gaussian random variables. It is natural to interpret the Gaussian components Y_1 and Y_2 as states of either cloudy or clear sky conditions respectively. We consistently estimate one Gaussian component with a negative mean and one with a positive mean. The interpretation is that the component with negative mean corresponds to the state of cloudy conditions and we refer this state to component 1 in both summer and winter. Component 2 is then representing the state for clear sky conditions and the parameter q_k is the a priori probability of being in the cloudy state.

Fig. 9 shows the normalized histograms of the summer and winter residuals together with the Gaussian mixture densities implied by the parameter estimates. The fit is excellent in both cases and the graphs display interesting differences between the summer and winter periods. Both densities are bimodal but with shoulder-like appearances having one peak significantly higher than the other. For the winter period the mixture distribution has a positive skewness and a significant left peak indicating that the state of cloudy conditions is dominant and carries more probability mass. The opposite situation is seen for the summer period which has a negative skewness and a significant right peak with the probability mass shifted towards the clear sky state. The estimate of the a priori probability for the cloudy state is $q_S = 0.3912$ for the summer, which corresponds to a 0.6088 probability for the clear sky state. In the winter the probability for the cloudy state is higher and estimated to $q_W = 0.5687$. We note that the underlying Gaussian distributions for the cloudy state has a higher variance in summer while the opposite is true in the winter. This is also a natural result consistent with irradiation levels being more dispersed and variable in cloudy conditions during summer.

In Fig. 10 we show that our model is capable of reproducing the histogram of the observed series $e(t)$. In Fig. 10 we have again plotted the normalized histogram for $e(t)$ together with the kernel density estimate, but also added the kernel density estimate obtained from a simulated 10-year sample of $e(t)$ generated by the estimated model. From Fig. 9 we know that our model is highly consistent with the empirical distributions of summer and winter residuals, and Fig. 10 shows that the model is also consistent with the histogram of $e(t)$. The fit in Fig. 10 is rather remarkable given that the model is fitted to summer and winter residuals separately and not on the observed $e(t)$, and the fact that the kernel estimate representing the model is obtained from a single run of simulated data. Different simulations will produce different levels of agreement but we find the fit to be quite stable across simulations.

As a final remark on the estimation results we emphasize that in addition to providing a convincing representation of the irradiation time series our model is highly tractable. The model structure is both simple and transparent. It is based on standard time series model components and it is very easy to estimate and simulate.

3.2. Validation

We validate our proposed model along two different dimensions. First, we investigate the model's ability to represent irradiation data at other locations. This part is done by estimating the model on the remaining four locations in our dataset. We present and discuss the overall model performance at these locations. With residuals obtained for the summer and winter periods at each location we also calculate empirical correlations between the different sites. Second, we perform a standard validation exercise where we study day-ahead predictions and associated confidence intervals.

Parameter estimates for the four remaining locations are presented in Table 4. We use the PACF of $Z(t) = G(t) - S(t)$ to determine the order p in the AR-part (5) of the model. The AR-order is $p = 2$ for Hamburg, Berlin and Stuttgart but for Nürnberg and München an order of $p = 1$ is found to be sufficient. Overall, the model fit is similar across all locations and with rather small differences in parameter estimates. We observe the same basic features as presented for the Hamburg location only with slight variations.

Our definition of the summer and winter periods is based on the months where we observe negative empirical skewness. These periods are the same for all locations. Fig. 11 plots the empirical monthly

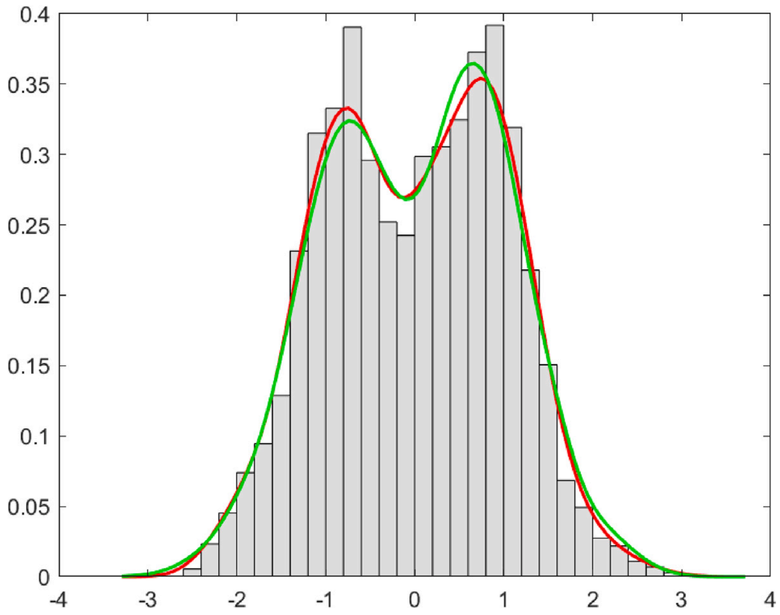


Fig. 10. Histogram (normalized) for residuals $e(t)$ and corresponding kernel density estimate (red curve). The green curve is the kernel density estimate obtained from simulated $e(t)$ from the estimated model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

skewness together with the skewness function implied by the parameter estimates for all locations. The skewness patterns are very similar across locations.

In Fig. 12 we plot the histogram of the residual series $e(t)$ together with superimposed kernel density estimates both from data and simulations constructed from the estimated models. The bimodal mixture distributions provide an excellent fit in all cases. There are some notable differences in the histograms. Hamburg and Berlin display two peaks of comparable height while the southern locations Nürnberg and München exhibit a significantly lower left peak.

The estimation results from the different locations show that the model works well for all the selected coordinates thus providing further validation. Indeed, irradiation data at all locations display the same features and are in fact very similar with regard to underlying stochastic and seasonal properties. The small differences in the AR-order, ranging between $p = 1$ and $p = 2$, is also a sign of the model specification being stable. We conclude that the model structure we have proposed is stable across locations and provides a satisfactory fit at each coordinate. These results indicate that the model structure should be relevant for Germany and most likely for large parts of Europe that share a similar climate. However, other climate types may certainly display quite different properties of solar irradiation and require modifications to the model structure.

The estimated models allow us to identify the summer and winter residuals $\epsilon_S(t)$ and $\epsilon_W(t)$ for each coordinate. Since these are assumed *iid* we can calculate empirical (sample) correlations for the summer and winter periods between the different locations. These correlations are reported in Tables 5 and 6. As can be expected the correlations decrease with physical distance. The distances (in km) are reported in Table 7. Taking Hamburg as a reference point we see that e.g. the summer correlations are steadily decreasing from a correlation of 0.4165 with Berlin to 0.0809 with München. The southern locations Stuttgart, Nürnberg and München are more closely situated and display higher correlations. We find the highest correlation in summer to be 0.6581, found between Nürnberg and Stuttgart, while the highest winter correlation is 0.5874 found between Nürnberg and München.

We also study one-day-ahead predictions of the irradiation level using as out-of-sample the year 2020. The purpose of our prediction study is to confirm that the model produces sensible predictions and associated confidence intervals. The predicted irradiation at time $t + 1$ based on the information $\mathcal{F}_t$ at time t, is given by the conditional expectation $\widehat{G}(t+1) = \mathbb{E}\left[G(t+1)|\mathcal{F}_t\right]$. From (1) and (5) we have

$$G(t+1) = S(t+1) + \sum_{i=1}^{p} \beta_i Z(t+1-i) + \sigma(t+1)e(t+1)$$

from which it follows that

$$\widehat{G}(t+1) = S(t+1) + \sum_{i=1}^{p} \beta_i Z(t+1-i)$$

using that $\mathbb{E}\left[\sigma(t+1)e(t+1)|\mathcal{F}_t\right] = 0$ since $\sigma(t+1)$ is known at t and $e(t+1)$ is independent of $\mathcal{F}_t$. From (5) we have that $Z(t+1-i)$ is known at time t for $i > 0$ and can be expressed in terms of $S(t+1-i)$ and $Z(t+1-i)$. The one-day predictions are thus easily calculated. Confidence intervals

Table 5
Correlation estimates calculated from summer residuals.

	Hamburg	Berlin	Stuttgart	Nürnberg	Nürnberg
Hamburg	1.0000	0.4165	0.1626	0.1786	0.0809
Berlin		1.0000	0.1794	0.2833	0.1473
Stuttgart			1.0000	0.6581	0.5488
Nürnberg				1.0000	0.6276
München					1.0000

Table 6
Correlation estimates calculated from winter residuals.

	Hamburg	Berlin	Stuttgart	Nürnberg	Nürnberg
Hamburg	1.0000	0.3458	0.1167	0.1299	0.0195
Berlin		1.0000	0.1977	0.2463	0.1697
Stuttgart			1.0000	0.5579	0.5262
Nürnberg				1.0000	0.5874
München					1.0000

Table 7
Distances (in km) between locations.

	Hamburg	Berlin	Stuttgart	Nürnberg	Nürnberg
Hamburg	0.0	281.7	501.7	456.9	568.2
Berlin		0.0	512.0	405.1	492.1
Stuttgart			0.0	130.3	161.7
Nürnberg				0.0	111.9
München					0.0

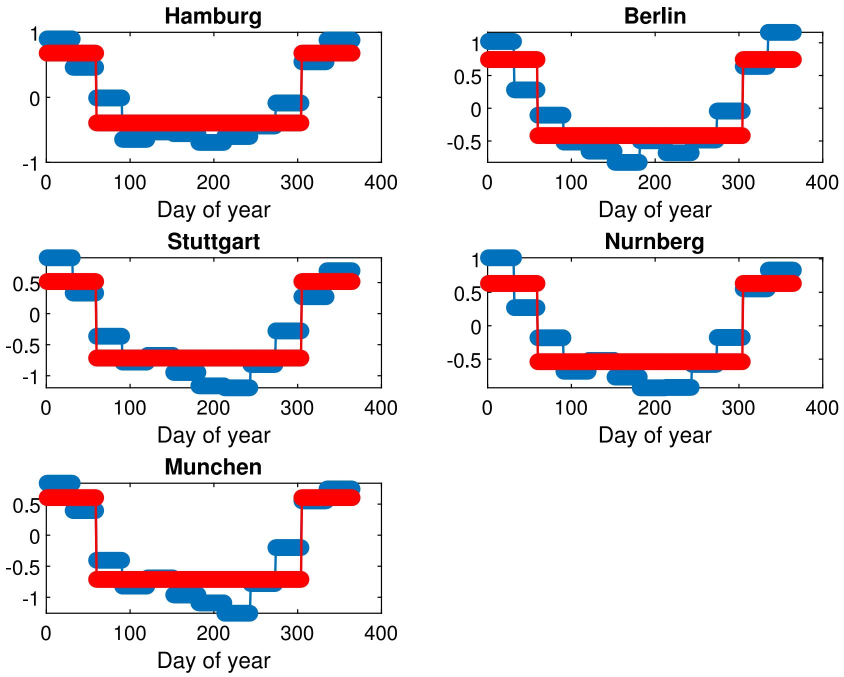


Fig. 11. Empirical monthly skewness estimates (blue) and model fitted skewness functions (red) for all locations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

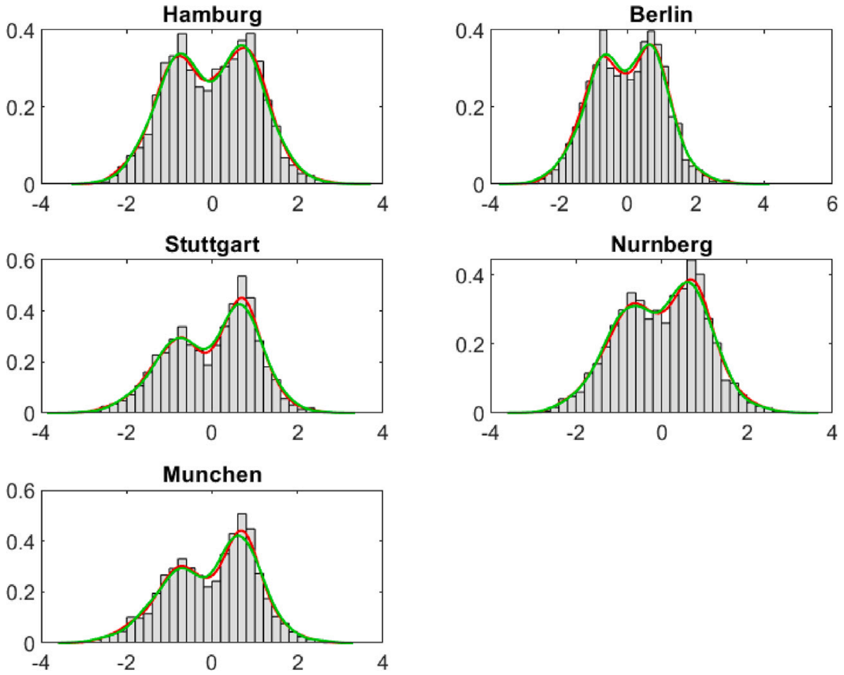


Fig. 12. Histogram (normalized) for residuals $e(t)$ and corresponding kernel density estimate (red) for all locations. The green curves are the kernel density estimates obtained from simulated $e(t)$ from the estimated models. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

for $\widehat{G}(t+1)$ can be calculated from simulations. To simulate from the conditional distribution of $G(t+1)$ at t we only need to draw random numbers from the distribution of $e(t+1)$. This amounts to simulating from a Gaussian mixture distribution, either from $\epsilon_S(t+1)$ or $\epsilon_W(t+1)$ depending on the value of $d(t+1)$, and can be done very efficiently using standard methods. From a simulated sample of M observations from $e(t+1)$ we calculate the empirical quantiles $\lambda_{\alpha/2}$ and $\lambda_{1-\alpha/2}$ to form $100(1-\alpha)\%$ confidence intervals for $\widehat{G}(t+1)$.

We illustrate the predictions for the location near Hamburg. In Fig. 13 we have plotted data and one-day predictions for each day in 2020 together with the estimated 95% confidence bounds, calculated from using $M = 50{,}000$ simulations each day. The prediction errors and the length of the confidence intervals varies with the seasonal conditional volatility. In the Hamburg case the confidence bounds are exceeded 3.84% of the 365 days consistent with the assumed confidence level of 5%. Predicting irradiation levels from one day to the next

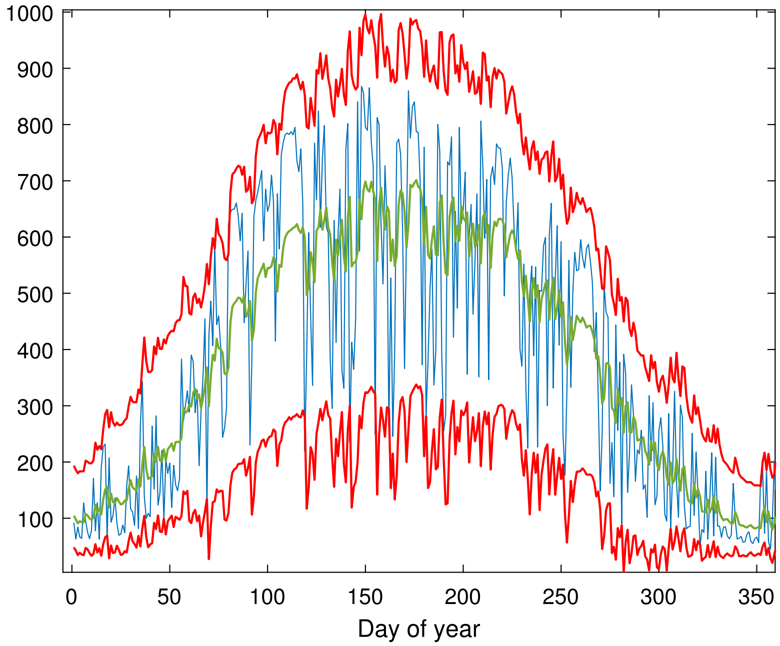


Fig. 13. Time series of irradiation data $G(t)$ (blue), and day-ahead predictions $\widehat{G}(t)$ (green) with 95% confidence bounds (red), for the out-of-sample period of 2020. Hamburg location. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

is a challenging task given the large day-to-day variations present in data. Given our statistical model setup, based only on irradiation data itself, the predictions are reasonably accurate. The mean prediction error in summer is -12.70 and in winter it is 8.06. These numbers are small compared to the mean irradiation levels of 517.53 and 131.88 and correspond to -2.45% and 6.11% in summer and winter respectively.

4. Application to PV risk management in power market

Production volumes from renewable generation assets cannot be perfectly forecasted due to the intermittent and stochastic fluctuations in solar irradiation. The deviation between the forecasted day-ahead production and the actual production typically gives rise to an imbalance cost component, which impacts the producer in the imbalance market, see De Jong and Kovaleva (2021) for a detailed discussion. For reasons of operational risk management and for investment decisions, it is therefore of practical interest to properly understand the distribution of day-ahead production volume as this is the major determinant of future income streams. In analogy with the commonly used Value-at-Risk measure used in financial markets we define the Production-at-Risk (P@R) as a quantile of the distribution for day-ahead production volume.

We apply the proposed model to estimate the P@R for a hypothetical PV solar park by simulating the future solar irradiation distribution and transforming it to production volumes by using a production function. Several mathematical models of PV production functions have been developed and tested for different solar cell technologies, see e.g. Huld et al. (2011) and Kaldellis et al. (2014). We employ a simple linear model which is also used in De Jong (2020) and in Kaldellis et al. (2014). The PV power production per square meter, $h(G,T)$, is defined as a function of the solar irradiation G and the operating temperature T according to the expression

$$h(G,T) = \kappa_1 G\left[1-\kappa_2\left(T-T_{ref}\right)\right] \tag{14}$$

where the parameter κ_1 is the efficiency of transforming solar energy into electrical power at the reference temperature T_{ref}, and κ_2 is the temperature coefficient, which is determined by the cell technology and the reference temperature. In this application we assume an efficiency of 20%, a reference temperature of $T_{ref} = 25$ degrees Celsius, and a temperature coefficient of 0.5%. This corresponds to $\kappa_1 = 0.2$ and $\kappa_2 = 0.005$. Under the given assumptions, and a solar irradiation of 1000 W per square meter, a 50 000 m^2 solar park produces 10 MWh of electricity per hour for a constant operating temperature at 25 degrees Celsius. This situation is a slight variation of the example studied in De Jong (2020).

At time t the produced volume (in MWh) from this solar park is

$$V(t) = 0.05 \times h(G(t), T(t))$$

We treat the temperature as a deterministic external input. The impact of temperature on produced volumes is small compared to irradiation and its contribution to the variance of production even smaller. Conditional on time t information the distribution function of $V(t+1)$ is

$$F_t(x) = \mathbb{P}_t\left(V(t+1) \le x\right) = \mathbb{P}_t\left(G(t+1)H(t+1) \le x\right)$$

where we denote $H(t) = 0.05\kappa_1\left[1-\kappa_2\left(T(t)-T_{ref}\right)\right]$. It is straightforward to show that in our proposed model

$$F_t(x) = q_k F_{1,k}\left(y_t(x)\right) + (1-q_k)F_{2,k}\left(y_t(x)\right) \tag{15}$$

where $k = S$ if $d(t+1) = 1$ and $k = W$ otherwise, $y_t(x)$ is the transformed point

$$y_t(x) = \frac{1}{\sigma(t+1)}\left(\frac{x}{H(t+1)} - S(t+1) - \sum_{j=1}^{p}\beta_j Z(t+1-j)\right) \tag{16}$$

and $F_{m,k}(x)$ are the corresponding Gaussian distribution functions for components $m = 1$ and $m = 2$. Note that at time t all the terms in (15) are known; they are either deterministic or can be computed from historic values in our model. The conditional distribution function in (15) is itself useful for risk management allowing probabilities for different scenarios to be easily calculated.

For a given value of $\alpha \in (0,1)$ the day-ahead P@R at time t, denoted $\lambda(t,\alpha)$, fulfills the relation

$$F_t\left(\lambda(t,\alpha)\right) = \mathbb{P}_t\left(V(t+1) \le \lambda(t,\alpha)\right) = \alpha \tag{17}$$

The P@R value $\lambda(t,\alpha)$ must be solved for numerically from (17). This can be done using simulation or a standard root-finding algorithm

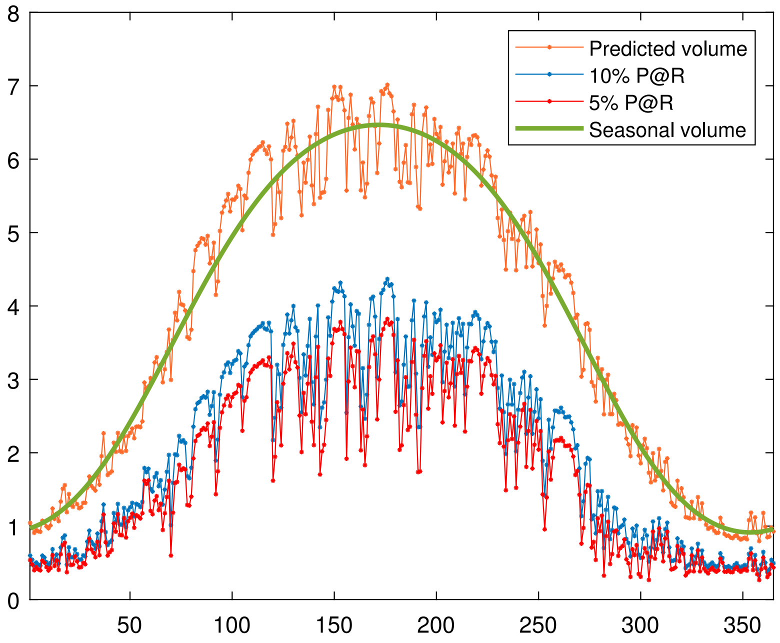


Fig. 14. Production-at-Risk at 10% (blue) and 5% (red) for a 50 000 m^2 solar park at the Hamburg location for year 2020. Included are also predicted day-ahead production volume (orange) and seasonal volume (green). All numbers expressed in MWh. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where (15) and (16) are used to express $F_t\left(\lambda(t,\alpha)\right)$ in (17). In our example we focus on the left tail of the distribution but a corresponding P@R for the right tail can be defined similarly with obvious modifications.

The P@R has a clear dependence on the volatility and shape of the distribution of solar irradiation data. As demonstrated in this paper the solar irradiation distribution exhibits a complicated time structure, which makes it a non-trivial task to dynamically estimate the P@R. We illustrate this by calculating P@R for the hypothetical solar park described above on each day in our out-of-sample data for the year 2020. The solar park is assumed to be located at the Hamburg coordinate and we use the parameter estimates in Table 3 in the calculations. The impact of the temperature on the variation of produced volume is minor and for simplicity we cancel it from our calculations by assuming $T = T_{ref} = 25$. This is obviously unrealistic, especially for the winter period, but the point of this exercise is to illustrate the temporal behavior of the P@R on which the temperature can only have a small influence compared to irradiation. Fig. 14 show the P@R calculated at the 5% and 10% levels for each day in 2020. The predicted and seasonal volumes are also plotted for reference. These are obtained from plugging the predicted, $\widehat{G}(t+1)$, and the seasonal $S(t+1)$ irradiation, into the production function. The P@R-levels in Fig. 14 display significant seasonal variations both in the level and variance. The P@R values naturally follows the seasonality in the irradiation level. It is also clearly visible that the P@R is more variable in the summer. It can also be noted that the 5% and 10% P@R values are much closer during winter compared to summer. This is an effect of the seasonal dependencies present in the distribution of irradiation. The P@R values plotted in Fig. 14 were calculated using simulation. The actual production volumes exceeded the 10% and 5% P@R during 2020 at 36 and 14 instances respectively. This corresponds to 9.86% and 3.84% of the days which is in close agreement with the selected probability levels.

To give some numerical insight we compare the P@R on three different dates. On March 16 the actual produced volume is 4.44 MWh and the 10% and 5% P@R are 2.22 and 1.84. This means that the probabilities that the next day production falls below 1.84 and 2.22 MWh are 5% and 10% respectively. The 10% P@R corresponds to a 50% drop in production from the 16th to the 17th. The maximum production in 2020 occurs on May 28 at 8.68 MWh. The P@R values calculated for May 28 were 3.59, at 10%, and 3.05 at 5%. In contrast, the lowest production in 2020 is 0.45 MWh recorded on Dec 23 with P@R values calculated to 0.41, at 10%, and 0.35, at 5%. The difference in P@R for the 10% and 5% levels on the day of maximum production is 0.55 which is substantial in terms of MWH. On the day of minimum production the difference is significantly smaller at 0.07 MWh. These differences are in fact rather typical for the summer and winter periods and are caused by the different stochastic and seasonal properties of solar irradiation over the year.

5. Concluding remarks

We have proposed a stochastic time series model for the daily (around noon) irradiance in 5 different cities in Germany. Our model is based on a careful analysis of the stylized facts of irradiance levels observed in a long time series of data recordings. Based on astronomical knowledge of the sun's position to the earth, we have explained the mean variation in irradiance by a clear sky model, while the first order stochastic effects are captured by an autoregressive time series of order 2. The residuals show seasonality in both their variance and skewness, where the latter effect has to the best of our knowledge never been observed nor modeled effectively. We suggest a summer and winter parameter changing between two different regimes. The variance is modeled as the product of a seasonally varying function and a GARCH(1,1)-model. The standardized residuals turn out to be uncorrelated Gaussian mixture models, where the bimodality of the distributions in winter and summer can be attributed to the clouds' interference with the sun.

Our proposed model is estimated to daily data, and validated in a prediction study. It is demonstrated that the model captures well the stylized features of irradiance in the studied locations. We presented an application to energy markets, where we show how PV-producers may use the model to assess their risk based on the measure Production at Risk.

Our model and analysis can be extended in several directions. A natural first step is to move towards a higher time resolution and consider an hourly stochastic time series model. Such a model will be interesting for applications in the intraday power market, where PV producers say, can hedge their exposure in the day ahead market. We expect an hourly model to lead to further challenges in capturing a daily seasonality profile, in particular around the hours of sunrise and sunset. Most likely, the memory effect in the time series model will show a higher order of autoregression.

In our study we have considered the correlation across 5 locations in Germany. Several interesting applications call for a spatial stochastic model, which is another interesting extension of our proposed irradiation dynamics. Our analysis indicates a declining correlation of irradiance residuals with distance. A spatial model must detect the spatial correlation structure in terms of distance to define a random field model (see Cressie and Wikle (2011)). Moreover, one must also be able to describe the other parameters of the model as functions of their geographic coordinates. Such a model will be very useful in studies of finding the optimal location of a PV installation, or in combining PV sites in a risk hedging study analogous to Benth et al. (2021).

The irradiation in a location is sensitive to variations in the cloud cover. Our model does not explicitly take into account the effect of the shifting cloud cover over a location. We remark that including cloud cover data for specific coordinates would likely have to use different data sources and that it then may prove difficult to reach adequate consistency between the series. Furthermore, variation in cloud cover is closely connected with the wind field at cloud level, which also correlates with temperature. The power curve of a PV-panel depends on ground temperature. It is a challenge to pin down realistic stochastic models combining all these factors.

Intraday solar production depends on the movements of the solar irradiation over the course of the day, reaching its peak around midday. Unfortunately, this pattern does not coincide well with the demand profile for electricity, which typically peaks in the morning/evening due to the increased human activities at these times of the day. This supply/demand discrepancy leads to a situation where the net demand (demand minus renewable production) falls into a slump during the midday hours, with relatively low demand and high solar generation, and in the morning/evening it sharply peaks due to high demand and low solar generation. The intraday net demand visually resembles the shape of a duck and is commonly known as the “duck curve”. The duck curve points to a problematic situation since it originates from a supply/demand mismatch, which needs to be compensated by conventional production technologies, such as gas or coal, and their flexibility comes at a high economic cost. With an increasing share of renewable production in the power system the duck curve is, ceteris paribus, likely to be even more pronounced in the future. This makes a challenge for the renewable expansion and several solutions to the problem are currently being explored. One promising alternative, which is presently investigated and implemented, is to store the oversupply from the midday solar production in short-term batteries in order to flatten the duck curve. In fact, the stochastic control problem for battery optimization requires a stochastic solar irradiation model as input data, which is an important motivation for the model proposed in this paper.

We leave these interesting and challenging questions of potential model extensions and applications to future studies.

CRediT authorship contribution statement

Karl Larsson: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Rikard Green:** Conceptualization, Investigation, Resources, Methodology, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Fred Espen Benth:** Conceptualization, Formal analysis, Investigation, Resources, Methodology, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank two anonymous referees, Pascal Heider, Kamil Kladivko and seminar participants at EnBW Germany for helpful comments on the paper. F. E. Benth acknowledges financial support from the thematic research group SPATUS funded by UiO:Energy, University of Oslo.

Appendix A

In this appendix we describe how the solar hour angle w and declination δ are determined. We give only a brief description and we refer to Duffie and Beckmann (2013) where more details and explanations can be found. The solar hour angle is given by

$$w = 15^\circ(LST - 12)$$

where LST is the local solar time. The local solar time LST is obtained from the local time LT and a time correction TC as

$$LST = LT + TC/60.$$

The time correction TC is calculated from

$$TC = 4(LSTM - lon) + E$$

where $LSTM = 15^\circ U$ where U is the offset to UTC time (U = +1 in Central European Time (CET)), lon is the longitude, and E denotes the so called Equation of Time given by

$$E = 229.2\,(0.000075 + 0.001868\cos(B) - 0.032077\sin(B) - 0.014615\cos(2B) - 0.04089\sin(2B))$$

with $B = (n-1)360/365$ where n is the yearday which may be expressed as a continuous variable.

The declination δ is directly given by

$$\delta = 23.45^\circ \sin\left(\frac{360(284+n)}{365}\right)$$

There are more accurate, and complicated, expressions available to determine the Equation of Time E and the declination δ, see Duffie and Beckmann (2013). The formulas presented here are sufficient for our purposes since they are only used to construct the seasonal function that we calibrate to data.

Appendix B

In this appendix we provide details on the parameter restriction for the seasonal variance. The seasonal variance part can be written

$$\sigma_S^2(t) = c_0 + c_1\cos(kt) + c_2\sin(kt)$$

where $k = 2\pi/365$. Using the trigonometric relation

$$\sin(kt+z) = \cos(kt)\sin(z) + \sin(kt)\cos(z)$$

we can write

$$\sigma_S^2(t) = c_0 + c\sin(kt+z)$$

where $c = \sqrt{c_1^2 + c_2^2}$ and z is the number such that $\tan(z) = c_1/c_2$, $c_2 \neq 0$. The smallest possible value of $\sigma_S^2(t)$ is therefore $c_0 - c$ and the restriction $c_0 > \sqrt{c_1^2 + c_2^2}$ thus ensures positivity.

Appendix C. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.eneco.2022.106421.

References

Alaton, P., Djehiche, B., Stillberger, D., 2002. On modelling and pricing weather derivatives. Appl. Math. Finance 9 (1), 1–20.

Benth, F.E., Christensen, T.S., Rohde, V., 2021. Multivariate continuous-time modeling of wind indexes and hedging of wind risk. Quant. Finance 21 (1), 165–183.

Benth, F.E., Ibrahim, N.A., 2017. Stochastic modeling of photovoltaic power generation and electricity prices. J. Energy Mark. 10 (3), 1–33. http://dx.doi.org/10.1080/17442508.2016.1177057.

Benth, F.E., Šaltytė Benth, J., 2012. A critical view on temperature modelling for application in weather derivatives markets. Energy Econ. 34, 592–602.

Benth, F.E., Šaltytė Benth, J., 2013. Modeling and Pricing in Financial Markets for Weather Derivatives. World Scientific.

Benth, F.E., Šaltytė Benth, J., Koekebakker, S., 2008. Stochastic Modelling of Electricity and Related Markets. World Scientific.

Campbell, S.D., Diebold, F.X., 2005. Weather forecasting for weather derivatives. J. Amer. Statist. Assoc. 100 (469), 6–16.

CAMS, 2019. User Guide To the CAMS Radiation Service. ECMWF Copernicus report, available at https://atmosphere.copernicus.eu, last accessed 2021-04-19.

Casula, L., D'Amico, G., Masala, G., Petroni, F., 2020. Performance estimation of photovoltaic energy production. Lett. Spatial Resour. Sci. 13, 267–285.

Cressie, N., Wikle, C.K., 2011. Statistics for Spatio-Temporal Data. John Wiley & Sons.

Cuppari, R.I., Higgins, C.E., Characklis, G.W., 2021. Agrivoltaics and weather risk: a diversification strategy for landowners. Appl. Energy 291, 1–16. http://dx.doi.org/10.1016/j.apenergy.2021.116809.

De Jong, C., 2020. Production Patterns of Wind and Solar. Kyos Energy Consulting, available at https://www.kyos.com/wp-content/uploads/2020/05/Production-patterns-of-wind-and-solar-the-financials-of-renewable-power-and-PPA-contracts.pdf.

De Jong, C., Kovaleva, S., 2021. Short Term Forecasting and Imbalance Costs. Kyos Energy Consulting, available at https://www.kyos.com/wp-content/uploads/2021/02/Imbalance-the-financials-of-renewable-power-and-PPA-contracts.pdf.

Duffie, J.A., Beckmann, W.A., 2013. Solar Engineering of Thermal Processes, fourth ed. John Wiley & Sons..

Engle, R.F., Gonzlaez-Rivera, G., 1991. Semiparametric ARCH models. J. Bus. Econom. Statist. 9, 345–359.

Gschwind, B., Wald, L., Blanc, P., Lefevre, M., Schroedter-Homscheidt, M., Arola, A., 2019. Improving the McClear model estimating the downwelling solar radiation at ground level in cloud free conditions – McClear-v3. Meteorol. Z. 28 (2), 147–163. http://dx.doi.org/10.1127/metz/2019/0946.

Härdle, W., Lopez Cabrera, B., 2012. The implied market price of weather risk. Appl. Math. Finance 19 (1), 59–95.

Hottel, H.C., 1976. A simple model for estimating the transmittance of direct solar radiation through clear atmospheres. Sol. Energy 18 (2), 129–134. http://dx.doi.org/10.1016/0038-092X(76)90045-1.

Huld, T., Friesen, G., Skoczek, A., Kenny, R.P., Sample, T., Field, M., Dunlop, E.D., 2011. A power-rating model for crystalline silicon PV modules. Sol. Energy Mater. Sol. Cells 95 (2011), 3359–3369.

Kaldellis, J.K., Kapsali, M., Kavadias, K.A., 2014. Temperature and wind speed impact on the efficiency of PV installations. Experience obtained from outdoor measurements in Greece. Renew. Energy 66, 612–624, 2014.

Lefevre, M., Oumbe, A., Blanc, P., Espinar, B., 2017. Mcclear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions. Atmos Meas Tech 6, 2403–2418. http://dx.doi.org/10.5194/amt-6-2403-2013.

Lingohr, D., Müller, G., 2019. Stochastic modelling of intraday photovoltaic power generation. Energy Econ. 81, 175–186.

Marchand, M., Saint-Drenan, Y.-M., Saboret, L., Wey, E., Wald, L., 2020. Performance of CAMS radiation service and HelioClim-3 databases of solar radiation at surface: evaluating the spatial variation in Germany. Adv. Sci. Res. 17, 143–152. http://dx.doi.org/10.5194/asr-17-143-2020.

McLachlan, G., Peel, D., 2000. Finite Mixture Models. Wiley.

Qu, Z., Oumbe, A., Blanc, P., Espinar, B., Gesell, G., Gschwind, B., Klüser, L., Lefevre, M., Saboret, L., Schroedter-Homscheidt, M., Wald, L., 2017. Fast radiative transfer parameterisation for assessing the surface solar irradiance: The Heliosat-4 method. Meteorol. Z. 26, 33–57. http://dx.doi.org/10.1127/metz/2016/0781.

Yang, D., Bright, J.M., 2020. Worldwide validation of 8 satellite-derived and reanalysis solar radiation products: a preliminary evaluation and overall metrics for hourly data over 27 years. Sol. Energy 210, 3–19. http://dx.doi.org/10.1016/j.solener.2020.04.016.