Predicting Short-Term Solar Irradiance Variability combining Numerical Weather Prediction and Machine Learning

Felix Simon Gödde

München 2023

Predicting Short-Term Solar Irradiance Variability combining Numerical Weather Prediction and Machine Learning

Felix Simon Gödde

Dissertation
an der Fakultät für Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Felix Simon Gödde
aus Berchtesgaden

München, April 2023

Erstgutachter: Prof. Dr. Bernhard Mayer
Zweitgutachter: Prof. Dr. Martin Weißmann
Datum der Abgabe: 24.03.2023
Datum der mündlichen Prüfung: 17.05.2023

Kurzzusammenfassung

In einer digitalisierten Welt zählt eine sichere und stabile Energieversorgung zu den wichtigsten Anliegen von Gesellschaften und Regierungen. Photovoltaikanlagen (PV) sind eine Schlüsseltechnologie, um dieses Anliegen mit der durch den menschengemachten Klimawandel zwingend notwendig gewordenen Transformation hin zu erneuerbaren Energiequellen in Einklang zu bringen. Die Energieerzeugung mittels PV unterliegt jedoch großen wetterbedingten Schwankungen, welche insbesondere durch den Einfluss von Wolken auf die solare Einstrahlung hervorgerufen werden. Aufgrund ihrer groben räumlichen (ca. 2 km) und zeitlichen (ca. 1 h) Auflösung können moderne Wettermodelle diese Fluktuationen nur bis zu einem gewissen Grad prognostizieren. Auf räumlichen und zeitlichen Skalen unterhalb der Modellauflösung treten jedoch große Schwankungen in der solaren Einstrahlung und damit auch in der erzeugten PV Leistung auf, welche für einen sicheren Betrieb des Stromnetzes von großer Bedeutung sind. Daher wurden im Rahmen dieser Dissertation zwei Methoden entwickelt, um diese kleinskaligen Fluktuationen der solaren Einstrahlung basierend auf operationellen Wettervorhersagen des COSMO-D2 Modells des Deutschen Wetterdienstes, trotz deren geringer Auflösung vorhersagen zu können. Als Basis für die Modellentwicklung und Validierung wurden Messungen der horizontalen Globalstrahlung (GHI) mit einer zeitlichen Auflösung von einer Sekunde verwendet, welche im Rahmen zweier Messkampagnen des Forschungsprojekts MetPVNet im Raum Kempten gesammelt wurden. Die erste Methode basiert auf räumlich hochaufgelösten Feldern der GHI, welche mithilfe des 3D Strahlungstransportmodells MYSTIC und hochaufgelösten Wolkenfeldern aus Large Eddy Simulationen (LES) berechnet wurden. Basierend auf diesen GHI Feldern wurde zunächst der Variabilitätsindex V in Abhängigkeit des Sonnenzenitwinkels (SZA), der Wolkenzuggeschwindigkeit sowie des Wolkenbedeckungsgrades in einer Lookup-Tabelle (LUT) abgespeichert. Daraufhin wurden die Wolkenzuggeschwindigkeit sowie der Bedeckungsgrad aus den Vorhersagen des COSMO-D2 Modells extrahiert. Im letzten Schritt wurde das LUT Modell mithilfe eines Trainingsdatensatzes optimiert, wobei sowohl der RMSE (Root Mean Square Error) um 36% der Bias um 83% reduziert werden konnten. Als zweite Methode wurde ein Random-Forest (RF) Modell gewählt, welches zur Klasse der maschinellen Lernmethoden zählt. Ein entscheidender Vorteil von RF Modellen ist, dass eine beliebige Anzahl von Parametern verwendet werden kann, von denen die wichtigsten durch den RF selbst bestimmt werden. Der Datensatz mit einer Größe von anfänglich 121 Parametern konnte mithilfe des RF, unter Verwendung einer Recursive Feature Elimination in Kombination mit der Permutation Feature Importance Metrik, auf 13 Parameter reduziert werden. Weiterhin wurden die Hyperparameter des RF auf einem Trainingsdatensatz mittels eines Grid-Search Algorithmus optimiert. Um den Einfluss von Vorhersagefehlern des COSMO-D2 Modells auf die Vorhersage der Variabilität abzuschätzen, wurden beide Methoden sowohl auf dem originalen als auch auf einem qualitätsgeprüften (gefilterten) Datensatz optimiert und evaluiert. Die Evaluierung der Modelle ergab dabei RMSE Werte von 0,78 (gefiltert) und 0,96 (original) für das LUT Modell sowie 0,039 (gefiltert) und 0,0423 (original) für das RF Modell, welches somit deutlich bessere Ergebnisse erzielte. Ein zusätzlicher Validierungsschritt, bei dem das RF Modell, welches auf dem originalen Datensatz basiert,

auf den gefilterten Datensatz angewendet wurde (und umgekehrt) zeigte, dass das RF Modell in der Lage ist, Fehler in der Wettervorhersage zu kompensieren. Daher ist es aus operationeller Sicht vorteilhaft, das RF Modell auf allen verfügbaren Daten zu trainieren, um die bestmögliche Variabilitätsvorhersage zu erzielen.
Nach bestem Wissen des Autors, sind die hier entwickelten Modelle für die Vorhersage der kleinskaligen und kurzfristigen Variabilität der GHI die ersten, welche ausschließlich modernste operationelle Wettervorhersagen als Eingangsdaten nutzen und stellen daher einen wichtigen Fortschritt im Vergleich zur Literatur dar.

Abstract

Secure and stable power supply is one of the major concerns of societies and governments in a digitalized world. To reconcile this concern with the need of a transformation towards renewable energy sources in the light of climate change, photovoltaic (PV) systems are a key technology. However, photovoltaic energy production is highly dependent on weather conditions and exhibits large and rapid fluctuations in the presence of clouds which shield the PV systems from direct sunlight. State-of-the-art numerical weather prediction (NWP) models are able to predict these fluctuations only to a certain extent as they are operated on coarse resolution of about 2 km in space and typically 1 hour in time. On smaller temporal and spatial scales, which are not resolved by NWP models, the incoming solar radiation and the according PV power show very large fluctuations which are particularly relevant for a save operation of the power grid. To be able to predict these small scale fluctuations despite the low resolution of NWP models, two downscaling methods were developed based on operational forecasts of the German Weather Service COSMO-D2 model. Therefore, measurements of the global horizontal irradiance (GHI) with a temporal resolution of one second gathered in the area of Kempten (Allgäu) in the scope of the MetPVNet project were used as a ground truth for model development and validation. The first modelling approach is based high spatial resolution GHI fields calculated using the 3D radiative transfer model MYSTIC in combination with high resolved cloud fields from large eddy simulations (LES). From the GHI fields the variability index V was derived as a function of solar zenith angle (SZA), cloud advection speed and cloud fraction and stored within a lookup table (LUT). Cloud speed and cloud fraction were then extracted from the COSMO-D2 forecasts to predict the variability index. In a final step the LUT model was tuned on a training dataset which reduced the RMSE (Root Mean Square Error) by 36% and the bias by 83% . The second approach makes use of the random forest (RF) machine learning technique. A major advantage of the RF is that arbitrarily many input parameters can be used from which the most predictive ones can be determined by means of the RF. Therefore, a small set of 13 features was extracted from initially 121 COSMO-D2 forecast parameters by applying a recursive feature elimination in combination with the permutation feature importance metric. Additionally, the hyperparameters of the RF model were tuned on a training dataset by using a grid-search algorithm. To assess the impact of the COSMO-D2 forecast error on variability prediction, both models were built on the original dataset as well as on a quality filtered subset of the NWP forecasts and evaluated on a separated validation dataset. With an RMSE of 0.039 (filtered) and 0.0423 (unfiltered) the RF model outperforms the LUT approach with an RMSE of 0.078 for the filtered and 0.096 for the unfiltered dataset. An additional cross-over prediction – applying the unfiltered RF to the filtered validation dataset and vice versa – revealed that the RF model is able to compensate for errors in the NWP forecast. Consequently, from an operational perspective, it is beneficial to train the RF with all available data to obtain the best performance.
To the knowledge of the author, the developed models for the prediction of small-scale and short-term GHI variability are the first to use exclusively operational state-of-the-art NWP forecasts as input and hence are a major advance compared to existing literature.

Contents

1. Introduction

The increasing energy demand of humanity, accompanied by the severe consequences of climate change, require exploitation of renewable and sustainable energy sources. Solar power technologies like photovoltaics (PV) play a crucial role in this transformation process. The amount of solar energy reaching Earth's surface within less than two hours is equal to World's primary energy consumption in 2021 (572 Exajoules) which illustrates the enormous potential of solar energy harvesting (Trenberth et al., 2009; Ritchie et al., 2022). Theoretically, the global primary energy demand could be covered by placing solar power plants at a few ideal locations like Northern Africa, Australia or the Western US (Kabir et al., 2018). However, solar power generation has drawbacks due to the spatial and temporal variability of solar irradiance. In desert regions with mostly clear skies solar power production is almost exclusively determined by the diurnal and seasonal cycle of the Sun's position which allows for precise estimates of future yields. In contrast, when considering regions like Central Europe, the frequent occurrence of clouds is responsible for a high variability of solar irradiance on a variety of spatial and temporal scales and hence leads to significant fluctuations in the production of solar power. An example for cloud induced high frequency fluctuations of irradiance and PV power output is shown in Figure 1.1. Both, the irradiance measured in the plane of the PV modules and the power output of the system show sudden ramps of more than 70% between the respective minimum and maximum within only 2-3 minutes. In some cases the relative ramps in PV power are even larger than in the irradiance which can for instance be caused by the influence of module temperature on its efficiency (Kreuwel et al., 2020). Mills et al. (2011) report even larger fluctuations of more than 80% within 1 minute which lead to power ramps of around 50% for a large PV power plant with a nominal capacity of 13.2 MWp. These large fluctuations on short timescales are one of the fundamental differences between solar and conventional power production.

The existence of these short term irradiance fluctuations and the possible implications for solar power production were first studied roughly 30 years ago. Suehrcke and McCormick (1988) presented one of the early investigations which showed that the statistics of instantaneous radiation values differ significantly from hourly or daily average values which were mostly available back then. Instantaneous radiation measurements show a bimodal distribution with one peak at small radiation levels, representing purely diffuse radiation within cloud shadows and the second peak at large radiation levels which occur in the sunny regions outside cloud shadows. The findings by Suehrcke and McCormick (1988) were confirmed by Skartveit and Olseth (1992) who analyzed the statistics of 5-minute average irradiance measurements in the US using the clearness index (measured irradiance normalized by simulated clear-sky irradiance). Based on the

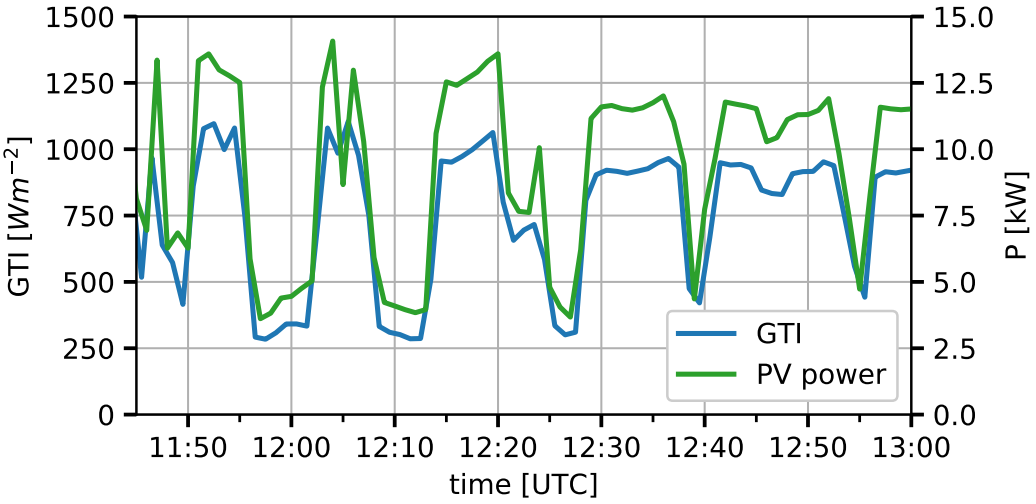

Figure 1.1.: Timeseries of the power output of a rooftop PV system (green) and measured global tilted irradiance (GTI) in the plane of the PV modules with a temporal resolution of one minute. The measurements were taken at location PV_11 on 15 July 2019 in the scope of the MetPVNet project (Meilinger et al., 2021). The PV system has a nominal installed capacity of 18.4 kWp and is tilted by 27° towards south-east.

findings by Suehrcke and McCormick (1988) and Skartveit and Olseth (1992), Gansler et al. (1995) investigated the impact of 1-minute irradiance data for the performance simulation of photovoltaic (PV) systems compared to hourly averages. They found that the calculated hourly power output of a simple PV system differs significantly when using hourly averaged irradiance data instead of 1-minute data and concluded that this is caused by the variability of solar irradiance and the non-linear relation between irradiance and the power output of the system.

With a raised awareness for the possible implications of high frequency fluctuations of solar irradiance for photovoltaic power production, several studies were conducted to investigate their impact on the aggregated output of distributed radiation sensors and photovoltaic systems. An early study on this topic was carried out by Otani et al. (1997) who analyzed 1-minute irradiance measurements of nine sensors which were uniformly distributed over an area of 4x4 km^2. For one month of measurements in October 1996 they found that the aggregated variability over all stations reduces on average by 40% compared to the fluctuations encountered at a single instrument. Thereby the values were ranging between extreme reductions of up to 87% and reductions as low as 20% which occurred under 'highly correlative irradiance fields'. Wiemken et al. (2001) also investigated variability of aggregated sensors, but they used the 5-minute power output of 100 photovoltaic systems distributed over entire Germany covering an area of 600x750 km^2. For a time period in June 1995 they found that the temporal standard deviation of normalized power output over all 100 systems reduces by 39%

during noon hours compared to the one of an individual system. This reduction could be predicted with good accuracy using a model developed by Steinberger-Willms (1993) which fits an exponential relation between the spatial cross-correlations coefficients for all pairs of stations and the distance between them. Furthermore, they conclude that the reduction in standard deviation for their case is mainly due to the large distance between the stations and adding more stations would not affect the aggregated variability.

Based on the analysis of the spatial correlations of distributed PV systems, models were developed to estimate the variability of the power output depending on the specific configuration of a fleet of PV systems. An example for such a model based on fundamental statistical considerations was derived by Hoff and Perez (2010). They show that the standard deviation of the relative power output for a fleet of identically configured PV systems can be quantified based on the number of systems (N) and a dispersion factor (D). Thereby, the dispersion factor is defined as the ratio between the spatial extension of the PV fleet along the wind direction and the cloud speed multiplied by the considered temporal resolution. For crowded regions where PV systems are very close to each other variability decreases proportional to $1/D$ reaching an optimum point where the dispersion factor equals the number of systems which maximizes the smoothing effect. In their follow-up study Hoff and Perez (2012) provided an estimate for the maximum power output variability of a PV fleet which is the total installed capacity of the fleet divided by $\sqrt{2N}$ with N being the number of systems. Additionally, they generalize their model developed two years earlier (Hoff and Perez, 2010) by explicitly considering the correlation between different PV systems when calculating the aggregated output variability. Another model for the estimation of aggregated PV power output was proposed by Marcos et al. (2016). They estimate power output of a dispersed PV fleet considering only the number of systems and their average size which makes this method considerably simpler than the approaches described before. Their model is fitted to power spectra derived from measurements of power output of a dispersed fleet of six PV systems in northern Spain. Thereby they empirically confirm that the standard deviation in power output of a widely distributed PV fleet is $1/\sqrt{N}$ times the standard deviation of a single system which was already derived by Hoff and Perez (2010) based on statistical considerations. They also showed that their model which was calibrated for climatic conditions in northern Spain also applies to dispersed radiation measurements in Colorado, USA.

The described models for the simulation of PV fleet power output variability theoretically enable grid operators to estimate power reserves necessary to compensate for future power fluctuations and ensure a save and stable power grid. However, all of these models rely on the knowledge of the current and in particular the future irradiance variability. Besides its relevance for power reserve planning, high variability in the PV power output can also cause fluctuations in the grid voltage which are a risk for grid stability (Tan and Kirschen, 2007). In this regard PV can be problem and solution at the same time as inverters of PV systems can be used to generate reactive power (Kraiczy et al., 2013) which is crucial to stabilize the voltage within the power grid (Sarkar et al.,

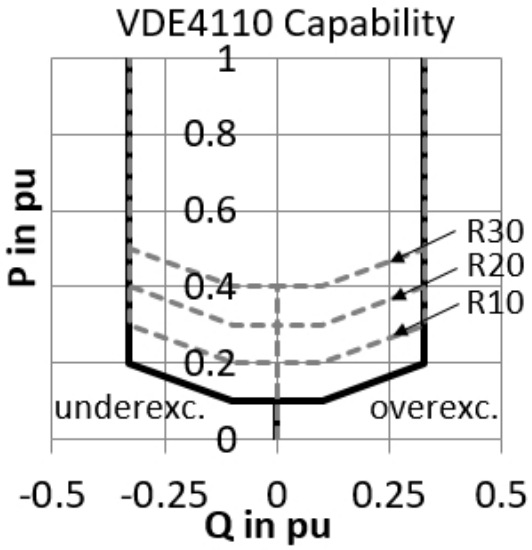

Figure 1.2.: Relation between power output (P) of a power plant and the reactive power (Q) which has to be provided on demand (black line) according to the grid code VDE AR-N 4110 of the German medium voltage grid. Both axes are normalized relative to the nominal capacity of the system (pu = per unit). For simplicity the left branch of the graph as well as the gray dashed lines can be ignored. The maximum reactive power (33% of nominal capacity) has to be provided as soon as the power plant operates at 20% of its nominal capacity or above. Below 10% no reactive power needs to be provided. The graph was taken from Figure 8 of Kraiczy et al. (2021).

2018). In the context of power grids reactive power is referred to as the power which is necessary to build up magnetic or electrical fields at different components of the grid like power lines or transformers. As reactive power is generated (or consumed) by a phase shift between current and voltage it can not be used to perform actual work, but it also contributes to the load on power lines and transformers and hence reduces the capacity for the transmission of active power. Currently, conventional power plants are mostly used to provide reactive power on the level of the high voltage transmission grid (Brückl, 2019), but with the transformation towards renewable energies these power plants are replaced by wind turbines and PV systems. Therefore, an increasing amount of reactive power has to be provided by renewable energy sources which is already considered in current grid codes. For instance, in the German medium voltage grid all connected distributed energy resources (DER) like PV (typical capacity 0.5-10 MWp) have to provide reactive power depending on their current power production (grid code VDE AR-N 4110, see Kraiczy et al., 2021). This means that not only the production of (active) power but also the availability of reactive power from PV systems for grid stabilization is highly weather dependent. In particular, the step like shape of the VDE4110 grid code shown in Figure 1.2 can lead to significant fluctuations in the availability of reactive power in case of high irradiance variability.

As a consequence of the described challenges associated with the integration of large PV capacities into the power grid, accurate forecasts on several timescales are necessary to ensure a save and stable power supply.

Over the last two decades many models were developed based on various techniques to forecast irradiance and its variability on time horizons ranging from seconds to days in advance. As not all techniques can be discussed in detail here, the reader is referred to the reviews on solar irradiance and PV power forecasting by Inman et al. (2013), Chu et al. (2021) and Benavides Cesar et al. (2022) as well as to the report by Pelland et al. (2013). For the shortest time scales of several seconds up to about 20 minutes in advance, total sky imagers (TSI) are frequently used to derive forecast of irradiance and its variability for limited areas of a few kilometers in diameter (Pelland et al., 2013). With their high temporal (seconds) and spatial resolution (10-100 m) they allow for precise forecasts of irradiance ramps and can therefore severe as input for aggregated PV power variability estimations on very short time horizons. Generally, forecast algorithms based on TSI extract cloud masks from individual images and cloud motion information from consecutive images to project the future distribution clouds which is then used to derive the irradiance pattern at the surface. A rather standard algorithm for TSI forecasts was published by Chow et al. (2011), but there are also more advanced methods making use of data assimilation and machine learning techniques as described by Gregor et al. (2023).

As the forecasting horizon of methods based on TSI are limited due to the limited field of view, images from geostationary satellites are used to reach even longer forecast horizons. The methodology for the derivation of solar irradiance forecasts from satellite images is very similar to the one used for TSI, but the larger spatial coverage allows to predict the advection of cloud fields from further away towards the point of interest. However, with the lower spatial resolution of typically 1.5 to 5 km (in the mid-latitudes) and temporal sampling rates of 5 to 15 minutes, satellite images currently contain less detailed information on irradiance variability. An example for a 'classical' satellite image based irradiance forecasting algorithm can be found in Lorenz et al. (2004). Just like for the TSI forecasts, statistical learning methods are also used to improve the performance satellite based forecasts. For instance Marquez et al. (2013) use several cloud properties derived from visible and infrared satellite images as input for an artificial neural network (ANN) to forecast the 30-minute average solar irradiance. To obtain information about the variability of irradiance on temporal and spatial scales smaller than the resolution of satellite data Watanabe et al. (2016) developed a downscaling method based on satellite cloud products and 1-minute resolution measurements of solar surface irradiance in Japan. However, the cloud products are based on observations from polar orbiting satellites with high spatial resolution which makes it necessary to adapt this method for the use with geostationary satellite data which currently have a lower spatial resolution.

The use of satellite imagery has been shown to provide good forecasts up to a time horizon of 6 hours. For predictions beyond 6 hours though, numerical weather prediction (NWP) models provide better estimates of solar irradiance at the surface

(Perez et al., 2010). In recent years there also have been publications using statistical learning methods to reach forecast horizons up to 24 hours like the algorithm published by Nam and Hur (2019) who use a Naive Bayes Classifier method to forecasts the aggregated power output of PV fleets in South Korea. However, when considering forecast horizons of one to several days ahead, NWP forecasts are currently the only tool which is able to predict solar irradiance and its variability.
As outlined earlier forecasts of irradiance and its variability over longer time horizons (>6 h) are important for active and reactive power reserve planning to ensure grid stability. Furthermore, day-ahead forecasts also play an important role energy trading (Visser et al., 2022). Despite these facts only few studies dealing with these time horizons have been published in the last ten years compared to publications on short-term irradiance forecasting. In recent years though, the number of publications is steadily increasing which shows an enhanced awareness for the importance of longer term irradiance forecasts in the scientific community (Benavides Cesar et al., 2022). To improve the accuracy of irradiance predictions from NWP models several techniques have been developed based on the smoothing of predicted irradiance fields over varying length scales to account for the statistical distribution of clouds or by correcting irradiance biases using model output statistics (Pelland et al., 2013; Lorenz et al., 2009). Improvements of the NWP model systems can also lead to better irradiance forecasts as shown by Geiß (2021) who demonstrated that the assimilation of visible satellite images improves irradiance forecasts in the ICON-D2 model.
In recent years there has been a trend towards the use of machine learning approaches to improve the irradiance and PV power forecasts based on NWP data. For instance Andrade and Bessa (2017) use forecasts of the WRF NWP model in combination with the gradient boosting technique to predict the power output of a particular PV system in Portugal and were able to improve the power forecast up to 72 hours ahead. Another approach in this regard was published by Kim and Lee (2021) who use an ensemble of different statistical techniques (e.g. random forest, gaussian process) which are optimally combined to obtain better accuracy. Additionally, they use the prediction error as input for a probabilistic graphical model to determine the range of uncertainty of irradiance and power forecasts.

However, all approaches are confined by the spatial (>2 km) and in particular the temporal resolution of the NWP models of typically 1 hour. Hence, they do not provide information about irradiance variability on smaller scales which is crucial for the operation and security of power grids with high penetration of PV. So far this issue has rarely been addressed in existing literature. One study by Kang and Tam (2015) uses day-ahead forecast of cloud cover from the US National Weather Service (NWS) in combination with 1-minute resolved irradiance measurements to predict daily averaged variability based on the probability of persistence of the clearness index ($\mathrm{POP} - \mathrm{K_D}$). They show that this variable correlates with the magnitude of irradiance ramp rates but as there is no information on the actual time of occurrence this quantity is of limited operational use. In contrast, Gristey et al. (2020) predict the PDF of instantaneous irradiance at the surface from high resolved Large Eddy Simulations (LES) using information like

cloud fraction and the dispersion of the liquid water path as input for machine learning algorithms. Despite the good results, this study is not feasible for operational use as LES simulations are computationally expensive and have a very limited domain size.
The most recent study on this topic by Riihimaki et al. (2021) aims to predict two measures of variability based on the effective transmissivity (ET), which is defined as the measured solar surface irradiance divided by the simulated clear sky irradiance. Similar to Gristey et al. (2020), they use different cloud properties to predict irradiance variability using machine learning techniques like random forest (RF), gradient boosting (GB), k-nearest-neighbour (KNN) and artificial neural networks (ANN). The cloud information is taken from measurements of cloud cover and cloud type which is derived from remote sensing instruments located in Oklahoma, USA. Based on these data they predict the standard deviation of ET ($\sigma(ET)$) calculated for 1-minute averages of ET within a 15-minute time window and the standard deviation of minute to minute differences of ET ($\sigma(\Delta ET)$). The results of the RF predictions which are analyzed in detail show a range dependent bias for both variability measures where small values are mostly overestimated, and large values are underestimated. Although Riihimaki et al. (2021) state that this parametrization is planed to be used in combination with NWP output, there are no information on the performance on NWP data so far.

Based on this detailed literature review it could be shown that the integration of large PV capacities into power grids brings a variety of challenges which make it necessary to provide accurate predictions of irradiance and in particular irradiance variability on a variety of temporal and spatial scales. For predictions on timescales below 6 hours a large amount of methods exists which deliver accurate predictions based on ground and satellite imagery. However, for longer timescales which are particularly important for grid operators, no method was published so far which can be used operationally to predict the variability of solar irradiance on spatial scales of single PV systems and temporal scales of minutes where the magnitude of irradiance fluctuations are largest. Therefore, this study aims to fill this gap by deploying machine learning techniques in combination with operational NWP forecasts to be able to infer the small scale variability of solar irradiance at the surface on time horizons beyond 6 hours.
This study was carried out within the scope of the MetPVNet project in cooperation with partners from research and economic institutions (see Meilinger et al., 2021).

This study is organized as follows: chapter 2 provides background information on clouds and their influence on radiative transfer, the representation of radiation and clouds in NWP models and the basic principle of the random forest machine learning technique. Chapter 3 gives an overview of the datasets used in the scope of this thesis and presents the methods developed to predict GHI variability based on NWP forecasts. In Chapter 4 the models are applied to a validation dataset to evaluate their performance and the results are discussed briefly. Chapter 5 and 6 discuss the results in the light of existing literature, summarize the findings and give an outlook with ideas for further research on GHI variability forecasting. Finally, additional information is provided in the appendix.

2. Background

The goal of this study is to predict the short term and small scale variability of the global horizontal irradiance (GHI). This information is of particular interest for power grid operation as it provides the possibility to estimate the magnitude of fluctuations in photovoltaic (PV) power production. Therefore, this chapter provides background information to gain a basic understanding of clouds and their radiative properties, reasons why NWP models alone are not capable of providing useful forecasts of GHI variability and methods to generate small scale GHI variability forecasts from NWP data. Furthermore, a literature review will show which methods for the prediction of GHI variability have been published recently. As a general introduction to the topic the following section provides some insights into political goals and the current status of photovoltaics in Germany.

2.1. Photovoltaics in Germany - current status and future development

This section aims to motivate the current and in particular the future importance of GHI variability predictions focussing on the situation of photovoltaics (PV) in Germany which might also apply to other countries.

According to the climate protection law passed in June 2021, Germany should be greenhouse gas neutral by 2045. The transformation of energy production towards renewable energies is the key to reach this goal and in particular PV plays an important role in the future and already today. In 2021 Germany had a total installed PV capacity of 59 GWp distributed over more than two million individual systems. These systems produced 51 TWh of electrical energy contributing 9.1% to the country's gross electrical power consumption (Wirth, 2022). Following the results of a study by Sterchele et al. (2020) an installed PV capacity of about 300-450 GWp is necessary to reach the goal of greenhouse gas neutrality. This means that PV-systems with a total nominal power between 10 and 16.25 GWp need to be installed each year from present day until 2045. Comparing this requirement to the current growth rate of about 5 GWp per year in 2021 shows a huge gap to aforementioned numbers, but actually meets goals formulated in laws currently in place. The law on renewable energies (EEG2021) passed in 2020 formulates an increase of PV capacity between 5-7.5 GWp per year and is therefore way below actual needs to meet the goal of the climate protection law. In Figure 2.1 taken from Wirth (2022) the discussed numbers are visualized, but growth rates for the scenarios by Sterchele et al. (2020) are larger as the goals are assumed to be already reached by 2040.

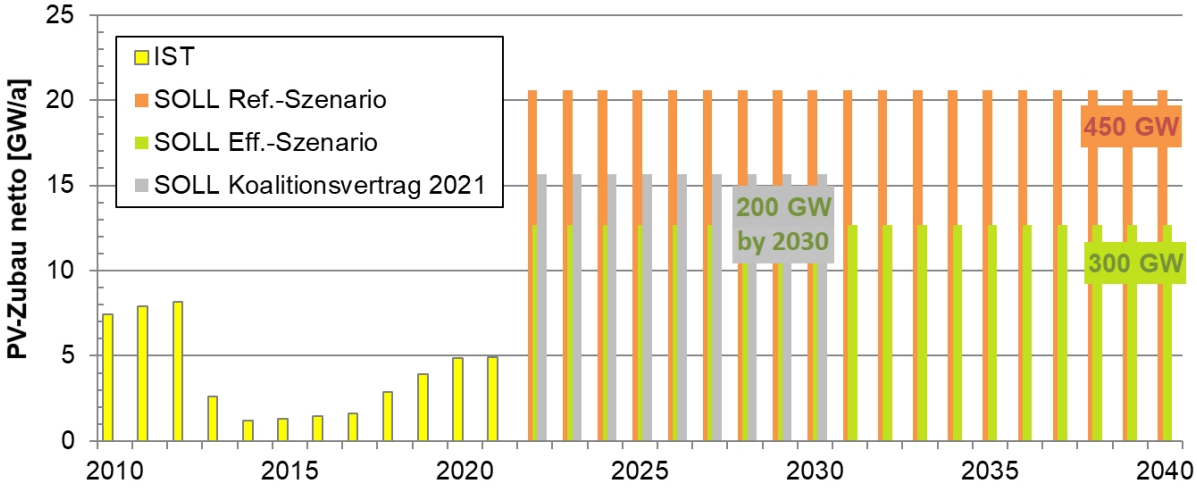

Figure 2.1.: Annual net growth of installed PV-power in GWp. The actual growth until 2021 is depicted in yellow. The green and orange bars indicate the required annual growth to achieve at total installed PV capacity of 300 and 450 GWp by 2040. These numbers originate from two scenarios defined in the study by Sterchele et al. (2020). The gray bars show the growth necessary to meet the 200 GWp goal of the coalition agreement until 2030. The figure was adapted from Wirth (2022).

The discrepancy between climate protection goals and the implementation of measures to reach them slows down the advance of renewable energies. However, the new German government elected in 2021 aims to increase PV capacity to 200 GWp by 2030 according to the coalition agreement. Although the coalition agreement is a statement of intent rather than law, the formulated goal implies a constant growth rate of 15.5 GWp per year and is therefore in line with the findings of Sterchele et al. (2020) (see also Figure 2.1). Furthermore, the original political will for the strong expansion of PV power production has recently increased in the light of the energy crisis caused by the attack of the Russian Federation on Ukraine, especially since the production of electrical power by PV is already economical. Electrical generation costs by PV-systems are 3.12-11.01 euro cents per kWh and large PV power plants with more than one MWp capacity produce at lowest rates, between 3.12 and 5.70 euro cents per kWh (Cost et al., 2021). In comparison, the costs of power production by conventional power plants based on coal and lignite range from 10.28 to 20.04 euro cents per kWh without including environmental costs caused by air pollution and the emission of greenhouse gases. According to calculations of the German Environment Agency from 2013 the environmental costs add another 8.04 to 10.75 euro cents per kWh for oil, coal and lignite (German Environment Agency, 2013) making energy production by fossil energy carriers in total up to ten times more expensive than PV.

In summary the fight against climate change, the current geopolitical circumstances, and the economical advantages of PV-systems, make it very likely that the amount of

installed PV capacity will increase drastically in the future. Therefore, it is necessary to provide reliable forecasts of PV power generation and its variability to be able to guarantee a secure, stable and sustainable power supply.

2.2. Radiative transfer

One of the main foundations of thesis is the transfer of electromagnetic radiation through the Earth's atmosphere. In particular radiative transfer explains how much of the sun's radiation entering the atmosphere at the top finally arrives at Earth's surface and can then be converted to electrical power by PV cells. Therefore, this section defines fundamental quantities used in radiative transfer and describes the radiative transfer equation (RTE) focusing on the properties of radiation in the solar spectral range.
The information in this chapter are mainly taken from Liou (2002) and Zdunkowski et al. (2007), other sources are cited when appropriate.

2.2.1. Important quantities

As any other scientific field, there are some fundamentally important quantities in radiative transfer to describe the propagation of electromagnetic radiation and its interaction with different media. In the following a selection of basic parameters is defined and described:

- **Solid angle** Ω: One can imagine the solid angle as the opening angle of a cone with its tip located at the center of a sphere of radius r. Depending on the value of Ω the cone covers a certain area A at its intersection with the surface of the sphere. On a sphere of radius 1 m a cone with solid angle 1 covers exactly an area of 1 m^2. This property is described in the following formula:

$$\Omega = \frac{A}{r^2} \tag{2.1}$$

  A denotes the area covered by the cone on the sphere's surface and r is the radius of the sphere. In case the whole surface of the sphere is covered, the solid angle takes a value of 4π. Even if Ω is dimensionless according to Equation 2.1, it has been given the unit steradian (sr).

- **Radiance** L: This quantity describes the amount of radiant energy (dQ) per time increment (dt), received from or emitted into a certain solid angle ($d\Omega$) by a surface element (dA) which is oriented perpendicular to the vector of the solid angle. Therefore, the radiance can formally be written like:

$$L = \frac{dQ}{d\Omega\, dt \cos\theta\, dA} \tag{2.2}$$

  where θ denotes the zenith angle of incoming or outgoing radiation and $\cos\theta$ accounts for the proper orientation of the surface element. The unit of radiance is $Wm^{-2}sr^{-1}$.

- **Irradiance** E: Integrating the radiance over a half sphere (2π) yields an energy flux density in Wm^2. In other words, irradiance is the flux of radiant energy (dQ) per time (dt) through an arbitrarily oriented surface (dA):

$$E = \frac{dQ}{dt\, dA} \tag{2.3}$$

  Irradiance as well as radiance are often used as monochromatic quantities. In these cases their units are extended by the wavelength increment ($d\lambda^{-1}$).

- **GHI and GTI**: The global horizontal irradiance (GHI) and the global tilted irradiance (GTI) denote the spectrally and hemispherically integrated solar irradiance at Earth's surface for a horizontally oriented and an arbitrarily tilted surface respectively. GHI is the standard quantity for radiation measurements by pyranometers, GTI is an important quantity when it comes to photovoltaics which are usually tilted with respect to the surface.

- **Absorption and scattering cross-section** $\sigma_{abs}/\sigma_{sca}$: Individual molecules or particles in the atmosphere can absorb and scatter electromagnetic radiation. The magnitude of absorption by a single instance is given by its absorption cross-section generally given in m^2 but cm^2 is also used. Similarly, the scattering cross-section describes the ability of a molecule or particle to scatter radiation.

- **Extinction cross-section** σ_{ext}: The sum of absorption and scattering cross-section is referred to as extinction cross-section and accordingly defined as $\sigma_{ext} = \sigma_{abs} + \sigma_{sca}$.

- **Absorption and scattering coefficient** β_{abs}/β_{sca}: Zooming out from the microscopic view on single objects to absorption and scattering properties of a volume containing many particles or molecules, the absorption/scattering coefficient can be defined as $\beta_{abs} = \sigma_{abs} \cdot n_{abs}$ and $\beta_{sca} = \sigma_{sca} \cdot n_{sca}$ respectively. Here, n denotes the number density of molecules/particles in m^{-3} which yields the unit m^{-1} for β.

- **Extinction coefficient** β_{ext}: Just like σ_{ext}, the extinction coefficient is the sum of absorption and scattering: $\beta_{ext} = \beta_{abs} + \beta_{sca}$. Furthermore, the inverse of β_{ext} can be interpreted as the mean free path of a photon until it interacts with a medium.

- **Optical thickness** τ: Integrating the extinction coefficient along the path of a photon between two points s_1 and s_2 like:

$$\tau = \int_{s_1}^{s_2} \beta_{ext}(s)\, \mathrm{d}s \tag{2.4}$$

  yields the dimensionless optical thickness which can also be separated into scattering and absorption by integrating β_{sca} and β_{abs} independently.

- **Scattering angle** Θ: This quantity is of special interest in the context of solar radiative transfer as it defines the angle of scattered light with respect to its incident

direction. A scattering angle of 0° means that a photon does not change its traveling direction after the scattering event whereas an angle of 180° means that the photon is scattered reversely to its initial direction.

- **Scattering phase function** $p(\Omega', \Omega)$: The scattering phase function gives the probability for a photon approaching from direction Ω' to be scattered into the direction Ω.

Having these general definitions in mind allows a closer look and a better understanding of the radiative transfer equation presented in the following section.

2.2.2. Radiative transfer equation

The radiative transfer equation (RTE) in its general form was proposed by Chandrasekhar (1950). With a different notation and slight rearrangement of the terms the RTE can be written like

$$\underbrace{\frac{dL(\Omega)}{ds}}_{(1)} = \underbrace{-\beta_{ext} L(\Omega)}_{(2)} + \underbrace{\frac{\beta_{sca}}{4\pi} \int_{4\pi} p(\Omega', \Omega) L(\Omega') \, \mathrm{d}\Omega'}_{(3)} + \underbrace{\beta_{abs} B(T)}_{(4)} \tag{2.5}$$

Starting with the left side of the equation it can be described as follows:

1. The change in radiance of a beam of light dL heading into a certain direction Ω with traveled distance ds equals the sum of
2. light removed from the beam by extinction (scattering and absorption),
3. light added to the beam by radiation being scattered into the direction Ω and
4. light added by the thermal emission according to Planck's law (Planck, 1901)

In general the RTE is formulated for a single wavelength because in particular the absorption, scattering and emission properties of media are highly wavelength dependent.

Solar radiative transfer

In the solar spectral range (≈300-4000 nm), which is relevant for photovoltaic power production, the Planck Function is essentially zero for atmospheric temperatures and thermal emission can be neglected in the RTE (eq. 2.5 term 4). Therefore, the RTE reduces to

$$\frac{dL(\Omega)}{ds} = -\beta_{ext} L(\Omega) + \frac{\beta_{sca}}{4\pi} \int_{4\pi} p(\Omega', \Omega) L(\Omega') \, \mathrm{d}\Omega'. \tag{2.6}$$

To calculate the amount of direct solar radiation, i.e. radiation which is not scattered or absorbed on its way to Earth's surface, the scattering integral (last term of eq. 2.6) doesn't need to be considered as it deals with diffuse radiation. Therefore, the RTE can

be solved analytically for the case of direct radiation resulting in the well-known Bouguer-Lambert-Beer law which describes the exponential decrease of light intensity in absorbing or scattering media:

$$L = L_0\, e^{-\tau}. \tag{2.7}$$

L_0 denotes the initial radiance and the optical thickness τ was derived according to Equation 2.4. In clear sky conditions without the influence of clouds, direct solar radiation makes up for about 80-90% of the global horizontal irradiance (GHI, see Section 2.2.1) during most of the daytime. Therefore, scattering could be neglected as a zero order approximation in clear skies. However, as soon as clouds come into play the direct radiation is attenuated and may reaches zero when the clouds are optically thick. This means that the GHI mostly consists of diffuse radiation and the scattering integral in Equation 2.6 becomes highly important. Unfortunately this integral is quite complicated to solve because one has to know the radiances from all directions $L(\Omega')$ which in turn rely on the knowledge of L from all other directions including the direction currently considered. This fact leads to a coupling of the radiances from all directions and hence a large system of equations which all have to be solved at the same time. Such an integro-differential equation can not be solved analytically but introducing a series of approximations makes a solution possible. An exemplary solution of this problem will be shown in Section 2.4.

The role of clouds in atmospheric radiative transfer is of high interest in the context of GHI variability and the resulting fluctuations in photovoltaic power production. To gain a better understanding of this topic some general insights on cloud radiative properties are given in the next section.

2.3. Clouds

As already described in the introduction of this thesis, clouds cause strong fluctuations in global horizontal irradiance (GHI) and make complex numerical predictions of solar power necessary. Therefore, this section tries to provide some basic information about the development of clouds, their global impact on Earth's climate as well as their radiative properties from microscopic to macroscopic scales.

2.3.1. Cloud development

The development of a cloud begins with warm and moist air rising due to a lower density than the surroundings. During ascent the air cools adiabatically at a rate of 9.8 Kelvin per kilometer until the relative humidity exceeds 100%. At this point water molecules start to condense on the surface of so-called cloud condensation nuclei (CCN). CCN are a subclass of atmospheric aerosol particles which are crucial for the formation of cloud droplets. In the absence of CCN air would require a relative humidity of several hundred percent to form droplets directly from the gas phase. The reason for this is the Kelvin effect which states that the saturation vapor pressure above a curved water surface is

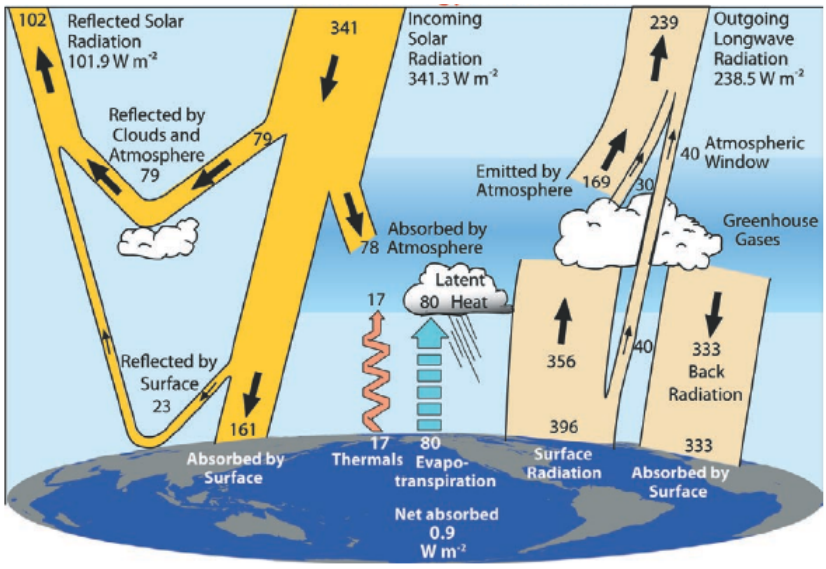

Figure 2.2.: Global annual mean state Earth's energy budget from March 2000 to May 2004. Fluxes are given in Wm^{-2}. Figure from Trenberth et al. (2009).

enhanced compared to a plain one. Therefore, CCN help to overcome this curvature effect and can additionally reduce the vapor pressure by dissolving into the condensed water (Raoult effect). Once the CCN were successfully activated the droplets keep on growing by condensation up to a radius of around 10-20 μm. To form even larger droplets which are finally able to produce precipitation, other growth processes like the coagulation and coalescence are necessary because the efficiency of condensational growth decreases with increasing radius.

It can be summarized that even though clouds appear to be macroscopic objects with an extent of hundreds of meters to several kilometers they are initially formed by tiny aerosol particles smaller than 1 μm, making them a highly complex phenomenon spanning spatial scales of more than eight orders of magnitude.

2.3.2. The role of clouds in Earth's energy budget

If one reduces the zoom level from the physics of individual clouds and has a look from a more global perspective their outstanding role for Earth's climate becomes visible. Beside their large importance for the global water cycle, clouds alter the solar and thermal radiative fluxes in the atmosphere and at Earth's surface. This influence is illustrated in the well known schematic by Trenberth et al. (2009) shown in Figure 2.2. In the solar part of the spectrum clouds reflect a considerable amount of radiation which leads to a cooling effect of around -48 Wm^{-2}. However, in the thermal spectral range they absorb radiation emitted by the surface and re-emit radiation at lower temperatures resulting in a net thermal heating of about 28 Wm^{-2}. Adding up both effects, clouds have a net radiative cooling effect of approximately -20 Wm^{-2} on global average (Allan, 2011).

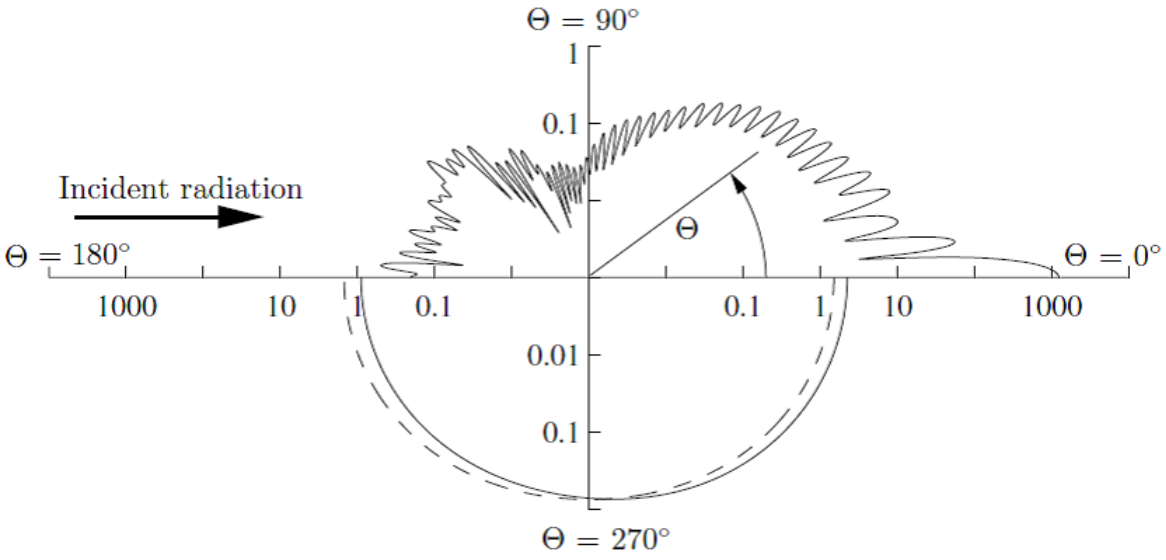

Figure 2.3.: Polar plot of the scattering phase function $\mathcal{P}(\cos\Theta)$ for three different size parameters calculated with Mie theory. Dashed curve: $x = 0.01$, lower solid curve: $x = 1$, upper solid curve: $x = 50$. A complex refractive index of $\mathcal{N} = 1.33 + 0i$ was used. Figure form Zdunkowski et al. (2007).

Comparing this value to the data of the latest IPCC report illustrates the enormous importance of clouds for Earth's climate: the current radiative forcing of 2.72 Wm^{-2} caused by human emissions of greenhouse gases is only 14% of the overall cloud radiative effect, but already led to a transient temperature increase of 1.07 Kelvin, though being still some way from equilibrium (Arias et al., 2021). Although there are no specific numbers for the equilibrium temperature at nowadays CO_2 levels (414 *ppm* in 2020) one can get an estimate by looking at the SSP1-1.6 scenario of the sixth IPCC report which shows a temperature increase of 1.5 K until 2100 at a CO_2 concentration of roughly 400 *ppm* (Lee et al., 2021).

Having discussed these rather general aspects of clouds it is time to go into some detail about their optical properties to understand their behaviour in the interaction with solar radiation.

2.3.3. Radiative properties of clouds

To assess the optical properties of a cloud as a whole one should start at the smallest instance this cloud is made of - a single cloud droplet. As described in Section 2.3.1, those spherical cloud droplets have a radius of around 10 μm. Therefore, the droplets are much larger than the wavelength of electromagnetic solar radiation which has its highest intensity at about 0.5 μm. In this case the scattering phase function (see Section 2.2.1) can be determined by Mie theory (Mie, 1908). Mie theory is based on the solution of Maxwell's equations for homogeneous spheres and is valid for all values of the so-called

size parameter x which is defined as:

$$x = \frac{2\pi r}{\lambda}. \tag{2.8}$$

λ denotes the wavelength of the radiation and r the radius of the particle. For cases where $x \ll 1$ which means that particles are much smaller than the wavelength, Mie theory simplifies to the well-known Rayleigh scattering approximation. Figure 2.3 shows the scattering phase functions for three different size parameters. In this case the phase function depends on the scattering angle Θ. For very small values of the size parameter ($x = 0.01$) the Rayleigh approximation is valid, and the scattering phase function is smooth and nearly isotropic with slightly higher scattering probability in the forward and backward direction (dashed line). As soon as the size parameter becomes considerably larger than one ($x = 50$), which describes cases like the scattering of solar radiation on cloud droplets, the phase function is way more complex containing many distinct peaks. Leaving aside the complex structure of the phase function, there are two important features to be noted:

1. the probability for scattering in the forward direction ($90° < \Theta < 270°$) is several orders of magnitude larger than in the backward direction.

2. There is a very sharp peak around the forward scattering direction ($\Theta = 0°$) which is about two orders of magnitude larger than any other peak (note the logarithmic axis).

These facts can be easily observed in nature which connects the optical properties of single droplets to the ones of an entire cloud. First, the overall strong forward scattering makes clouds appear very bright from below as long as their optical thickness does not become very large - huge thunderstorm clouds look rather dark from below. Second, the sun's disk can often be observed even in a completely overcast sky which is due to the very strong scattering peak in the forward direction.

Beside the aforementioned implications of the scattering phase function of individual droplets for the appearance of a macroscopic cloud, there is still one step in between. Clouds usually do not contain only droplets of a single size, but they exhibit a droplet size distribution. As the exact angular position of individual peaks in the scattering phase function depends on the size of the droplets, many features in the phase function cancel when droplets of different sizes are present. In some cases though, when the size distribution is very narrow, optical phenomena like iridescent clouds can be observed (Reichardt et al., 2015). It should be noted here that one can interpret the scattering phase function as the amount of radiation which is scattered into a certain direction considering a beam of incoming radiation or as the probability of a single photon to be scattered into a certain direction. These two interpretations are used interchangeably in this dissertation.

Going back to the topic of GHI variability, the described scattering properties of

clouds provide indications for the influence of different cloud types on the GHI at the surface. Optically thin clouds with $\tau < 1$, like for example cirrus clouds, do not considerably impact GHI, as most of the scattering inside the cloud is directed forward towards the surface. However, if the cloud is optically thick with values of $\tau \approx 10$ or so, like it is the case for shallow cumulus clouds, radiation will be scattered multiple times. As a consequence the amount of radiation being reflected to space drastically increases which on the other hand reduces the amount of solar radiation reaching the surface.

Based on the presented knowledge about cloud optical properties the next section will deal with the treatment of clouds and radiation in numerical weather prediction (NWP) models. This part is essential to understand why NWP models alone are not capable of providing accurate predictions of small scale and high frequency fluctuations of GHI.

2.4. Representation of clouds and radiation in NWP

Numerical weather prediction (NWP) plays a key role in the transformation towards renewable energy sources. As mentioned before, solar power generation relies on the prediction of global horizontal irradiance (GHI) is strongly dependent on the amount of clouds in the atmosphere. Therefore, this section presents information about the representation of clouds and solar radiation in NWP models with a particular focus on the former operational COSMO-D2 model (replaced by ICON-D2 in 2021) of the German Weather Service (DWD). The information in this section concerning radiative transfer are taken from Liou (2002) and Zdunkowski et al. (2007), details about NWP and in particular the COSMO-D2 model originate from Baldauf et al. (2011), Baldauf et al. (2018) and Coiffier (2011).

2.4.1. Representation of clouds

Sophisticated NWP models like COSMO-D2 are currently operated with a horizontal resolution of around 2 km and a vertical grid spacing starting at about 10 m close to the surface, increasing to several hundred meters at higher altitudes. These models are often referred to as convection permitting models as large convective systems like thunderstorms can be explicitly resolved. To predict the development and evolution of clouds on the resolved scales, complex microphysical parametrizations are necessary as microphysical processes happen on scales far below model resolution (see section 2.3.1). COSMO-D2 uses a one-moment bulk scheme which is quite common in NWP models as it is computationally very efficient. This type of scheme predicts the mass of water for different hydrometeor classes, like cloud droplets or rain, using a prescribed size distribution for each class (Khain et al., 2015).

Although large scale convection can be resolved by the model, clouds formed by shallow convection are often smaller than 2 km in diameter and can therefore not

be treated with the previously described microphysics schemes. Therefore, additional parametrizations are used to calculate subgrid mass fluxes and diagnose cloud cover for subgrid-scale clouds. COSMO-D2 for example uses a shallow convection parametrization developed by Tiedtke (1989) and diagnoses subgrid cloud cover according to Sommeria and Deardorff (1977). A qualitative assessment of the impact of the treatment of clouds and spatial resolution in NWP on the representation of GHI variability reveals two important aspects: First, the simplifications which are necessary to develop computationally efficient parametrizations for cloud microphysics introduce inherent errors in the prediction of clouds. An example for such an error may be a too late onset of precipitation which could cause larger size and longer lifetime of clouds. Such errors could be reduced for example by using more sophisticated parametrizations. Second, subgrid-scale clouds in NWP models are only represented by a subgrid cloud cover valid for the entire grid box which completely ignores the fact that this cloud cover may be made up by a number of individual clouds in reality. Consequently, neglecting horizontal variations of cloud cover leads to less variability in the GHI field at the surface.

The described problems of cloud representation are closely related to the treatment of solar radiation in NWP which will be discussed in the next section.

2.4.2. Representation of solar radiation

As mentioned in Section 2.2.2 the solution of the radiative transfer equation (RTE) is complicated and hence computationally expensive when scattering processes have to be considered which is the case in the solar spectral range. Therefore, the representation of solar radiation is challenging for NWP as computationally efficient and sufficiently accurate solutions are required. One class of solutions which is very commonly used in NWP and climate models are the so-called two-stream-methods (TSM). The term two-stream refers to separation of the RTE into upward and downward directed radiation fluxes. Basic assumptions of TSM are a plan-parallel and horizontally-homogeneous atmosphere as well as an azimuthally averaged radiance field. Furthermore, the transport of radiation in horizontal direction is omitted which makes the radiative transfer a purely one dimensional problem. From a computational point of view this assumption leads to a strong simplification of the calculation because there is no coupling between individual columns of the NWP model.

A very common TSM is the δ-Eddington two-stream (Joseph et al., 1976) which is also used in COSMO-D2. This method uses three important approximations:

1. δ-scaling: The pronounced forward-scattering peak in the phase function of aerosols and cloud droplets (see Section 2.3.3) is cut off and represented by a δ-function. This means that a part of the radiation which would have been scattered is treated as unscattered radiation.

2. Two-term expansion of the scattering phase-function: It is common to represent the scattering phase function as a series expansion of Legendre polynomials. Joseph

et al. (1976) use an expansion up to the second term to represent the δ-scaled phase function.

3. Linear approximation of the radiance distribution: The dependence of the radiance on angle is in particular important for the scattering integral (last term of eq. 2.6). Following Eddington (1916) the radiance is assumed to be $L(\mu) = L_0 + L_1\mu$ with μ being the cosine of the zenith angle θ.

While the δ-scaling improves the accuracy of the method, the approximations in 2 and 3 make it possible to solve the scattering integral and hence the entire RTE analytically for a vertically discretized atmosphere.

In the context of GHI variability the use of a one dimensional radiation scheme is a considerable limitation. For instance cloud shadows are always placed below the cloud which is wrong for all solar zenith angles except for overhead sun. Additionally, the horizontal transport of radiation is neglected which ignores the scattering on and the transmission through cloud sides. These "3D-effects" in cloud-radiation interaction can lead to enhancements of GHI above clear sky level Pecenak et al. (2016) and therefore impact GHI variability. Another source of error concerning variability originates from the treatment of subgrid-scale cloud cover. For simplicity one can assume a model column containing a single partially cloudy layer with a cloud cover of 50%. In this case the radiative fluxes would be calculated for the cloudy as well as for the clear part separately and finally be averaged using the cloud cover as weight. The final output by the model is an average value of GHI for the entire model grid box which however should be divided into a shaded part with low GHI and a clear part with high values of GHI. In reality, also the distinction of a clear and cloudy part is still an approximation as the subgrid-scale cloud cover might be subdivided into even smaller clouds distributed throughout the model gridbox. As a consequence, ignoring subgrid-scale variability of clouds and radiation leads to smoothed GHI fields with spatial gradients being much smaller than in reality. The small spatial gradients are in fact closely related to small temporal gradients via the advection of clouds. This lack in temporal variability of GHI is additionally enhanced by the low calling frequency of the radiation scheme in operational NWP. In COSMO-D2 radiation is only called every 15 minutes and written into the model output only every hour.

In summary, state-of-the-art NWP models like COSMO-D2 drastically underestimate the temporal as well as the spatial variability of GHI due to their coarse resolution, usage of infrequently called one dimensional radiative transfer schemes and the insufficient treatment of subgrid-scale cloud variability. These aspects will be analyzed and discussed in a more quantitative way in Chapter 3. To overcome the deficits of NWP models the implementation of machine learning techniques poses a promising opportunity. Therefore, the following section provides some insights into the structure and basic working principle of random forest models.

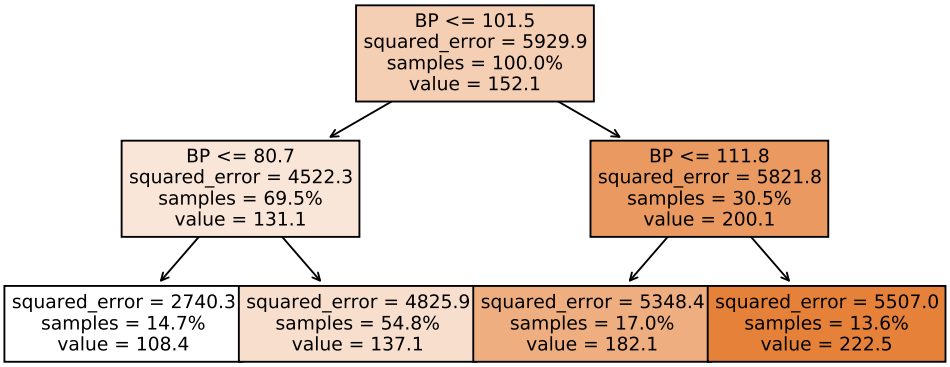

Figure 2.4.: Example of a decision tree using the diabetes dataset by Efron et al. (2004). The depth of the tree was limited to two. The first line of the nodes provides information about the splitting criterion while the mean squared error, the fraction of samples within each node and the average value of the variable used for splitting are listed below. The shading of the nodes indicates the magnitude of the mean value of the splitting variable.

2.5. Machine learning – random forest model

Machine learning (ML) models are capable of inferring complex nonlinear relationships between variables without a priori knowledge about the actual form of the relation. Furthermore, ML is able to handle large amounts of data with high dimensionality where classical regression techniques would not be applicable. Therefore, ML is well suited to solve problems in the field of atmospheric science where many problems exhibit both of the aforementioned attributes. For example Forster et al. (2017) use a random Forest model (RF) to detect halo displays in camera images or Scheck (2021) introduce a neural network in their forward operator for visible satellite images to name only a few.

In this Dissertation an RF is used to predict the variability of GHI based on numerical weather predictions. Hence, this section gives a brief overview of the working principle and basic properties of RF. General information is mainly taken from Frochte (2019) and Biau and Scornet (2016), other sources are cited when appropriate.

2.5.1. Random forest basic principle

Random Forests belong to the class of supervised machine learning techniques. This means the algorithm is designed to learn the connection between input variables x_i - also called features - and a known target variable y of a given dataset. The target variable can either be a certain category (classification problem), for instance different

types of precipitation, or a numeric value, such as the amount of precipitation per time, which would be a regression problem. Unsupervised learning methods are different in the way that no target variable is used and the algorithm detects patterns which can be attributed to certain categories later on.

RFs consist of a variable number of individual decision trees which is the origin of the term "Forest" in RF. In general decision trees are built on a dataset by splitting the data into smaller and smaller instances (nodes) with the intention of reducing a given error metric in the child-nodes created by each new split. Common error metrics are the Gini impurity (GI) for classification and the mean-squared-error (MSE) for regression problems. As the RF is applied to a regression problem in this dissertation, only regression is discussed in more detail. Each split in the tree is created by asking if the value of a feature is larger than a certain threshold. All data points for which this is true are put in one node and the remaining data into a different one. However, each split is not performed on an arbitrary feature with a random threshold but on a specific feature with a certain threshold which leads to the largest reduction in the error metric (e.g. MSE) of all possible splits of all available features. The splitting procedure is repeated for every newly created node until only a single element is left in the final leaf node or a predefined limit of splits is reached. This algorithm is referred to as CART (Classification and Regression Trees) and was developed by Breiman et al. (1984).

An example for a decision tree created with the CART algorithm is shown in Figure 2.4. Here a dataset on diabetes patients published by Efron et al. (2004) was used together with the DecisionTreeRegressor of scikit-learn (Pedregosa et al., 2011). For simplicity only 2 features of the dataset, age and blood pressure (BP), were used and the depth of the tree was limited. In this case the target variable is a measure of disease progression one year after baseline. As can bee seen in Figure 2.4, patients with a BP lower than 101.5 have a considerably slower progression in comparison to the ones with higher BP. Furthermore, also the second split is performed on BP which shows that BP may be a more important feature than age.

Although an individual decision tree could be used as a predictor on its own their predictions are known to have high variance as they are prone to slight changes in the training dataset used to build them. Additionally, decision trees tend to overfit the data when grown too deep which means they may not generalize well on a test dataset especially in cases of noisy data. Therefore, Breiman (1996) developed the so-called **bagging** (from **b**ootstrap **agg**regat**ing**) method. In a first step bootstrap samples are created from the original training dataset which have the same number of elements but differ from the original dataset as they are drawn with replacement (bootstrapping). From each bootstrap sample a decision tree is grown which is different from all other trees as no bootstrap sample contains exactly the same data. Finally, the predictions of all trees on a test dataset are averaged by weighting the outcome of each tree equally (aggregating). This procedure provides better results than a single tree because even though changes in training data may affect some trees, averaging over a large ensemble of

trees reduces their influence on the result and therefore also the variance. Furthermore, the problem of overfitting is addressed as the individual trees overfit on different instances of the training dataset which is also compensated by averaging.

The last step one the way from a single decision tree to an RF as it is formulated in Breiman (2001) is the increase of randomness in the construction of each tree. This is achieved by selecting a random subsample of features at each node which are considered for the split. The advantage of more randomness in the decision trees is the reduction of their mutual correlation which in turn reduces the variance of the prediction (Segal, 2004).

2.5.2. Variable importance

One useful property of RFs is their ability to measure the contribution of individual features to the final result also referred to as feature importance (FI). FI is in particular relevant in science when not only the performance of the model is relevant but rather the insights provided by features which are responsible for its performance. As there are different ways to generate feature importance measures with different properties, two of the most common ones are briefly described here.

A first way to calculate FI by means of RF is to have a look at the error reduction due to all splits of an individual feature in the entire forest. If a feature is important there will be more splits and an overall higher reduction in the error metric (MSE for regression) compared to other features. This behaviour was already described for the simplified setup shown in Figure 2.4 where blood pressure is obviously more relevant than age regarding the progression of diabetes disease. An advantage of this method is that it does not require an extra test dataset but can directly be evaluated from the RF. However, this so-called Gini importance (GI) is biased towards continuous features or features with many categories (Strobl et al., 2007). Features with few categories can only be used a very limited number of times for splitting in each tree and in the special case of binary features only once. Therefore, their overall reduction in the error metric may be relatively low even though they are very important.

Another measure for FI is referred to as permutation feature importance (PFI). To derive PFI the order of a feature is randomly permuted which breaks the correlation between feature and target variable. If no test dataset is available PFI can be estimated using the out-of-bag (OOB) technique (Breiman, 2001). OOB PFI is calculated by means of the data which are not used for an individual tree because of bootstrapping. The left out dataset is propagated through the tree twice, one time in the original state and one time with the considered feature being randomly permuted. Applying this procedure to all trees in the RF provides a solid estimate of FI. Other ways to calculate PFI are to use a separate test dataset or to use cross validation on the training dataset which might be the most reliable approach. The advantage of PFI over GI is that it is less biased towards continuous features and features with many categories as the impact on the final

result rather than on individual splits is investigated.

2.5.3. Quantifying prediction error

An important step in the development of any kind of model is the quantification of its prediction skill. There are a bunch of approaches which can be applied to a variety of machine learning techniques. A standard method is to validate the model against a test dataset which usually contains about 10% of the whole dataset. This can be problematic if the test dataset does not reflect the general nature of the problem but is only valid for a specific case. Optimizing the model to perform well on this test dataset will lead to degraded performance on a different dataset. Furthermore, in cases where the dataset is sparse it might not be possible to exclude a considerable amount of data from the training dataset and save it for testing. In this case RFs offer the unique possibility to estimate the error out-of-bag. This OOB error estimation is performed by using the data which are not contained in a certain bootstrap sample as test data for the respective tree (Breiman, 2001) as it was already described for the feature importance in the previous section. Therefore, no separate test dataset is necessary, and no additional test runs need to be made as the OOB error is evaluated during training. Although the OOB error is a reliable measure in most cases, Janitza and Hornung (2018) and Mitchell (2011) show that the OOB error overestimates the error of the model under certain conditions.

Another alternative for error estimation is referred to as cross-validation (CV). In this procedure the training dataset is divided into k equally sized parts. One of the k parts is used as a test dataset in each step while the remaining $k-1$ parts represent the training dataset (k-fold CV). This way the entire training dataset is used for training and validation which gives a good estimate of the prediction error (Refaeilzadeh et al., 2016) and does not require a separate test dataset. Furthermore, this approach is well suited for tuning the hyperparameters of an RF.

2.5.4. Hyperparameters

One important way to improve the performance of RFs is the adjustment of its hyperparameters. Hyperparameters determine for instance how many trees are used (n_estimators), how deep the trees are grown (max_depth) or which fraction of features is considered for a split at each node (max_features). In this section a selection of hyperparameters will be discussed in more detail. Most of the information is taken from the review paper by Probst et al. (2019), additional sources are cited when appropriate.

Each hyperparameter changes the properties and hence the performance of an RF by modifying the individual trees. Some hyperparameters available in the RandomForestRegressor of the python library scikit-learn (Pedregosa et al., 2011) and their impact are described in the following:

- Number of trees (n_estimators):
  More trees generally improve the performance of the RF. However, there is usually a

saturation threshold above which additional trees have only little effect on accuracy. To keep the balance between accuracy and efficiency of the algorithm it is useful to estimate a reasonable threshold for instance by using the OOB error.

- Number of features used for splitting at each node (max_features):
  This variable controls the influence of less important features especially in the presence of very predictive features. Lower values of max_features yield less correlated trees which makes the predictions of RF more stable but reduces the accuracy on average. Typically, fractions like the square-root or the logarithm of the total number of features are used for max_features.

- Maximum number of data points used to build each tree (max_samples):
  The max_samples parameter determines how many samples are drawn via bootstrapping from the dataset. In the default setting of the RandomForestRegressor of scikit-learn as many samples are drawn (with replacement) as contained in the dataset. However, it can be beneficial to draw smaller fractions as the correlation between individual trees is reduced when they are trained on different subsets of the dataset. Therefore, reducing max_samples has similar influence as the reduction of max_features.

- Minimum number of data points in the last node of each tree (min_samples_leaf):
  Larger numbers reduce the computational time considerably by only slightly decreased prediction accuracy. Furthermore, larger terminal nodes can increase prediction performance due to reduced overfitting especially in the presence of many noise variables (Segal, 2004).

- Minimum number of data points in a node required for splitting (min_samples_split):
  This parameter has similar properties like min_samples_leaf but allows for more uneven splits and can therefore leads to different results.

To achieve optimal performance of the RF the best combination of hyperparameters has to be found. For this purpose it is useful to use algorithms like GridSearch which test all combinations of a given set of parameters and return the best combination for a given error metric.

In summary, it can be stated that RF is a well explored machine learning technique which shows good performance for a large variety of applications. Furthermore, the actual working principle and the behaviour of RFs is more intuitive and explainable than it is the case for techniques like artificial neural networks. Finally, RFs offer the useful possibility to estimate the importance of each feature in the dataset which can provide fundamental insights into the nature of the problem.

2.6. Literature review

As the presented research topic on forecasting of GHI variability is of interest all over the world, there are existing studies which developed models based on machine learning

for this purpose. To be able to explain the differences between this study and existing literature later on, a short summary of the two most relevant papers recently published by Gristey et al. (2020) and Riihimaki et al. (2021) is given in this section. Those have already been mentioned in the introduction but here they will be described in some more detail. As the studies use surface solar irradiance (SSI) as radiation quantity which is essentially the same as GHI, the term SSI will be used for consistency.

In their model based study Gristey et al. (2020) predict the probability density function of the surface solar irradiance (SSI). As a basis they use high resolution large eddy simulations (LES) to simulate a variety of shallow cumulus cloud fields and calculate the SSI using a 3D radiative transfer solver. To describe the PDF of SSI a lognormal distribution is fitted to the larger irradiance mode (regions outside cloud shadow) while a normal distribution is chosen for the smaller irradiance mode (region inside cloud shadow). The parameters of these distributions are then predicted by two different machine learning techniques, a random forest (RF) and an artificial neural network (ANN). Gristey et al. (2020) use six input parameters, namely:

- domain mean cloud fraction
- relative dispersion of liquid water path
- mean in-cloud drop number concentration
- mean horizontally projected cloud area
- mean distance from cloud center to the cloud center of the nearest neighbor
- cosine of the solar zenith angle

The predictions of RF and ANN yield equally good results with an accuracy larger than 80% in all cases. Concerning feature importance both models show that the mean cloud fraction is by far the most predictive feature followed by the dispersion of the liquid water path.

The second study was carried out by Riihimaki et al. (2021) who predict two measures of variability based on the effective transmissivity (ET) which is defined as the measured SSI divided by the simulated clear-sky SSI. There idea is, similar to Gristey et al. (2020), to use different cloud properties to predict SSI variability. However, Riihimaki et al. (2021) base their investigations on measurements gathered at Southern Great Plains (SGP) site of the US Department of Energy (DOE) in Oklahoma rather than on LES model data like Gristey et al. (2020).

Beside radiation measurements, hemispheric cloud fraction is derived from a Total Sky Imager (TSI) and seven cloud classes are defined based on cloud data products derived from active remote sensing instruments. Using the cloud fraction and cloud class as input parameters, five distinct model approaches are implemented including linear

regression, k-nearest-neighbour (KNN), random forest (RF), gradient boosting machine (GBM) and artificial neural network (ANN). The models are fitted/trained to predict the standard deviation of ET ($\sigma(ET)$) calculated for 1-minute averages of ET within a 15-minute time window and the standard deviation of minute to minute differences of ET ($\sigma(\Delta ET)$). $\sigma(ET)$ can be interpreted as a measure for the magnitude of fluctuations in normalized SSI whereas $\sigma(\Delta ET)$ measures the magnitude of fluctuations of the derivative of ET which is the variability of the variability in some sense. To obtain optimal results the hyperparameters of the machine learning models are tuned, before their performance is evaluated using four-fold cross validation.

Although table 2 in the paper indicates that KNN, RF and GBM perform best with only small differences (eg. $r2(\sigma(ET)) = 0.386/0.386/0.390$ for KNN/RF/GBM), Riihimaki et al. (2021) only state that RF and GBM yield similar results and decide to further analyze the results of RF. RF predictions for both variability measures show an overestimation of small variability values which turns into an underestimation for large values. This pattern appears throughout all figures presented in the paper. For future work Riihimaki et al. (2021) plan to apply the same methodology to numerical weather prediction (NWP) models where the input parameters should be retrieved from the NWP output.

In summary both studies apply machine learning techniques to predict different measures of SSI variability. However, the models are currently either based on high resolution LES data (Gristey et al., 2020) or measurement data (Riihimaki et al., 2021) but have not been applied to operational NWP data. It furthermore should be stated here that both studies were published in parallel to the research presented in this dissertation and did therefore not serve as any kind of template for the developed methods.

In the next chapter all methods used and developed to achieve the goal of predicting small scale and short term irradiance variability based on numerical weather prediction models will be described in detail.

3. Methods

This chapter focuses on the development of methods for predicting the variability of the global horizontal irradiance (GHI) as well as on the underlying input and validation data. On the way to this core topic different aspects of variability like its dependency on horizontal resolution, the choice of the radiative transfer scheme or cloud situation are investigated to get a clear picture of the problem to be solved. An import step in this regard is the definition of a sensible measure for variability which will also be provided in this chapter.
Overall this chapter can be divided into two main parts. The first part, covering sections 3.1-3.3 and 3.5, describes the datasets which serve as a main basis for this study as well as existing models and tools used in the scope of the study. In the second part investigations and model development for the prediction of small scale GHI variability are presented in detail which builds the core of this study and can be found in sections 3.4 as well as 3.6-3.8.

As a first step a detailed overview of radiative transfer models, their properties and necessary input parameters for the conducted simulations will be given.

3.1. Radiative transfer with libRadtran

Radiative transfer models are an important tool to investigate the variability of GHI as they are able to simulate the propagation of solar radiation through Earth's atmosphere under various conditions. In this dissertation the libRadtran software package (Mayer and Kylling, 2005; Emde et al., 2016) was used to carry out most of the radiative transfer simulations. LibRatran is a collection of models which solve the radiative transfer equation (RTE) at different levels of complexity. This section will provide an overview of the applied RTE solvers and necessary input parameters for RT simulations with libRadtran.

3.1.1. RTE solvers

For reasons of computing efficiency as well as for sensitivity studies several RTE solvers of libradtran were used to simulate GHI. Their most important properties and fields of application are shortly described in this section.

DISORT

The DIScrete Ordinate Radiative Transfer technique DISORT included in libRadtran provides 1D-solutions of the RTE and relies on horizontally homogeneous atmospheric layers.

For all DISORT simulations in this thesis the default implementation based on the DISORT code of Stamnes et al. (1988) with by improvements by Buras et al. (2011) was used. DISORT solves the RTE for a user defined number of discrete angles and is therefore capable of computing radiance distributions unlike the two-stream approximation discussed in Section 2.4.2 which only provides irradiances. The solver is well suited for simulations of clear sky and totally overcast situations due to its high computational speed and good accuracy. However, for cases with intermediate cloud fractions DISORT can only be used to approximate the actual radiation field by calculating the weighted average of clear and totally cloudy simulations which neglects 3D cloud effects. This method is also referred to as the independent column approximation (ICA). Therefore, DISORT was mainly used to simulate GHI in clear sky situations including aerosol.

MYSTIC – 1D IPA and 3D

The Monte carlo code for the pYSically correct Tracing of photons In Cloudy atmospheres (MYSTIC) developed by Mayer (2009) solves the RTE without approximations by tracing individual photons on their way through the atmosphere. MYSTIC provides the possibility to take the 3D nature of radiation-cloud interaction into account and produces realistic patterns of GHI even in complex cloud situations. This solver can be selected by using the keyword `mystic` for the solver option in libRadtran.

In forward mode, MYSTIC starts photons at the upper boundary of the user defined atmosphere – this is also called the top of the atmosphere (TOA) – which then travel into a certain direction determined by solar zenith and azimuth angle. The probability of a scattering or absorption event along the path is determined by the probability density function (PDF) based on the Bouguer-Lambert-Beer law (see 2.2.2). Using the cumulative PDF, the location of the event can be determined by generating a random number between 0 and 1 and choosing the corresponding value of optical thickness. In the next step it is determined whether a photon is scattered or absorbed which depends on the single scattering albedo (ω_0) defined as $\omega_0 = \beta_{sca}/\beta_{ext}$. Consequently, the probability for scattering is ω_0 and for absorption $1 - \omega_0$. The fate of the photon is again determined by a random number between 0 and 1 where values smaller ω_0 lead to scattering and larger ones to absorption which would mark the end of the photon's path. If scattering occurs the new direction of the photon is assessed based on the cumulative probability of the scattering phase function (see also Section 2.2). As soon as the new travelling direction of the photon has been determined, the described procedure is repeated over and over until the photon is either absorbed or leaves the atmosphere towards space.

However, the MYSTIC simulation in forward mode is inefficient when the GHI at a certain location has to be calculated. Most of the photons started at TOA will not reach the surface at the required location and therefore do not contribute to the result. This problem can be avoided by tracing photons in backward mode which can be done by setting the option `mc_backward` in libRadtran. In this case, photons are started from the location of interest into all directions of the hemisphere. To make sure that the photon

exits the atmosphere in the direction of the sun the so-called local estimate technique is used: for each scattering event during the tracing process, the probability of scattering into the direction of the sun is calculated and multiplied by the probability of the photon traveling the path between the current position and TOA without being scattered or absorbed according to Bouguer-Lambert-Beer law. This probability is then added as a weight to the photon. Talking about photon weights it should be mentioned that the Monte Carlo approach of MYSTIC is improved by introducing them in case of absorption events. It is more efficient in a computational sense to reduce the weight of a photon - which is set to 1 in the beginning - in case of an absorption event by the probability of absorption and continue tracing rather than eliminating it and starting from scratch. To further increase the accuracy of MYSTIC simulations the variance reduction option `VROOM` (Buras and Mayer, 2011) can be used which will be explained im more detail later on.

In this study MYSTIC was used to simulate spatially high resolved GHI using cloud fields from Large Eddy Simulations (LES) as input. To investigate the influence of 3D radiative transfer against 1D solvers with respect to GHI variability MYSTIC was used in 3D as well as in the 1D Independent Pixel Approximation (IPA) mode. In 1D IPA mode each column of the model domain is treated separately by applying periodic boundary conditions when a photon is leaving the considered column. The resulting GHI field is identical to the two-stream approximation described in Section 2.4.2 which is used in numerical weather prediction models like COSMO-D2. To perform the 1D IPA simulation the option `mc_ipa` has to be set in addition to the MYSTIC solver.

Beside the solver for the RTE a variety of input parameters have to be passed to libRadtran to properly define the optical properties of the atmosphere and the Earth's surface. Some of the most important options and parameters are described in the next section.

3.1.2. Input parameters

This section will give a short overview over important parameters which can be selected for radiative transfer simulations in libRadtran with a focus on options which are relevant for the presented study.

Atmosphere and trace gases

One of the most fundamental inputs for radiative transfer are the atmospheric profiles of pressure, temperature, and trace gas concentrations. LibRadtran contains six predefined standard atmosphere files published by Anderson et al. (1986) like for instance the mid-latitude summer atmosphere (afglms) or the U.S. Standard atmosphere (afglus) which are included by the `atmosphere_file` option. Although the available standard atmosphere files yield reasonable results for idealized simulations, varying pressure, temperature and in particular water vapor concentration can have considerable impact on the

result. Therefore, NWP data of COSMO-D2 were used to modify the atmosphere file to achieve accurate results for a specific time and location.

Aerosol

Another parameter particularly relevant for radiative transfer under cloudless conditions is the atmospheric aerosol. Using the `aerosol_default` option libRadtran sets up a standard aerosol configuration according to Shettle (1990) with a vertically integrated optical thickness of about 0.2 at 500 nm wavelength. This default option can be modified or overwritten by several options.

For the sensitivity studies of GHI variability the default aerosol was used, but for simulations of clear sky GHI at measurement sites the more sophisticated OPAC library (Hess et al., 1998) was considered. OPAC contains optical properties of several species of aerosols like sulfate or soot as well as predefined mixtures of species typical for different regions. For the Central European area the mixtures of continental aerosols (clean, average, polluted) and urban type aerosols are the most relevant ones. They can be included by setting the option `aerosol_species_library OPAC` and selecting the mixture using `aerosol_species_file` together with the name of the desired mixture. When measurements of the aerosol optical thickness (AOD) are available, the AOD of previously chosen aerosol options can be modified by specifying AOD at an arbitrary wavelength using `aerosol_set_tau_at_wvl`. In case of measurements at multiple wavelengths an exponential fit according to Angström (1929) can be performed and the parameters passed to libRadtran by `aerosol_angstrom` option. This option was chosen to simulate accurate values of clear sky GHI at measurement locations in the area of Kempten where spectral AOD data acquired by an AERONET sun photometer (Holben et al., 1998; Giles et al., 2019; Sinyuk et al., 2020) were available. Both of the described options (`aerosol_set_tau_at_wvl` and `aerosol_angstrom`) only change the optical thickness or its wavelength dependence, properties like the single scattering albedo or the asymmetry parameter of the aerosol mixture stay the same.

Clouds

The most important ingredient for fluctuations of GHI is the presence of clouds. Water clouds can be included in libRadtran using the `wc_file` option. As only water clouds were targeted in this study, options for ice clouds will not be discussed here. For a 1D cloud a single file has to be provided which defines the altitude, effective droplet radius (r_{eff}) and liquid water content (LWC). However, it was already discussed in Chapter 2 that 1D plane parallel treatment of clouds considerably underestimates variability compared to reality. Therefore, 3D cloud fields are necessary which can then be used for simulations with MYSTIC. To define a 3D cloud file the following parameters have to be specified:

- Number of gridboxes in all three dimensions (Nx, Ny, Nz)
- Horizontal resolution (dx, dy)

- Altitude of vertical levels (z[i])
- LWC and r_{eff} for each cloudy pixel at location ix, iy, iz

For this study, 3D cloud information was taken from Large Eddy Simulations (LES) of shallow cumulus clouds. These LES data contain information about r_{eff} and LWC for each gridbox. To obtain accurate results for GHI the microphysical properties of the LES clouds (LWC, r_{eff}) were converted to optical properties (β_{abs}, β_{sca}, etc.) according to Mie theory (see also 3.1.2). Therefore, the `wc_properties` option in libRadtran was set together with the keyword `mie`. The LES data as well as their use for 3D MYSTIC simulations will be described in more detail in Section 3.2 and Section 3.6 respectively.

In the presence of clouds droplets or aerosols which both have a strong forward peak in the scattering phase function (see Section 2.3.3) MYSTIC produces so-called 'spikes' when using the local estimate method (see Section 3.1.1). These 'spikes' originate from photons with exceptionally high weights which reduce the convergence of the result. To avoid these spikes the VROOM option (Variance Reduction Optimal Options Method, Buras and Mayer (2011)) is implemented in MYSTIC. VROOM is a collection of variance reduction techniques which improve the convergence behaviour by changing the path of individual photons to eliminate the spikes. Consequently, the use of VROOM helps to reduce the number of photons by preserving the accuracy when clouds or aerosols are present.

Albedo

The albedo parameter describes the reflection properties of the Earth's surface and can be passed to libRadtran using the `albedo` option together with a value between zero and one. With this option the surface is treated as a lambertian reflector meaning that the incoming radiation is scattered isotropically in all directions. Furthermore, the surface reflection is assumed to be independent of wavelength and the provided value determines the fraction of reflected radiation. Consequently, an albedo of zero means that no reflection takes place because all photons are absorbed and an albedo of one that all the incoming radiation is reflected.

As the albedo of Earth's surface actually varies with wavelength – otherwise the surface wouldn't show any color – libRadtran offers the possibility to provide a file containing a spectral albedo which can be included using the `albedo_file` option. Going further towards the actual scattering properties of the Earth's surface, one also has to consider the angular dependence of the reflectivity. This means that the reflection is not isotropic but has a directional dependence regarding the incoming and outgoing radiation. This property can be described by a BRDF (Bidirectional Reflectance Distribution Function, Nicodemus et al. (1977)). One example for a BRDF model is the AMBRALS algorithm (Algorithm for Model Bidirectional Reflectance Anisotropies of the Land Surface) developed by Wanner et al. (1995). The model uses three parameters (f_{iso}, f_{vol}, f_{geo})) to represent the BRDF and can also be used for simulations with libRad-

tran. As BRDF is not used explicitly in this thesis it will not be discussed in further detail.

However, the AMBRALS model offers the opportunity to derive a blacksky- and whitesky-albedo via polynomial functions when the three parameters are known. The blacksky-albedo describes the (lambertian) albedo of a surface for incoming radiation from a specific direction as it is the case for direct solar radiation while the whitesky-albedo describes the albedo in case of isotropic radiation from all directions which is approximately true for overcast skies. The actual albedo can be calculated as a weighted average of both contributions depending on the amount of diffuse radiation. In this thesis satellite measurements of the AMBRALS BRDF were used to derive a representative albedo for radiative transfer simulations which will be described in more detail in Section 3.5.2.

3.1.3. Simulating GHI

GHI consists of diffuse and direct irradiance which are calculated separately in MYSTIC. The solution of the RTE for direct irradiance is provided by Bouguer-Lambert-Beer law which gives the probability for a photon to reach the ground without being scattered or absorbed. For diffuse irradiance photons from all directions of the hemisphere contribute to the result which makes it necessary to trace many of them to obtain accurate results with a low level of noise. In backward mode the output quantity is specified using the `mc_backward_output` command followed by `edn` for diffuse and `edir` for direct irradiance.

When performing MYSTIC simulations within a 3D-domain a sample grid has to be defined which determines the spatial resolution of the output quantity. Therefore, the option `mc_sample_grid` has to be passed together with the number of sampling points in x- and y-direction (`Nx, Ny`) and the according resolution (`dx, dy`). If the computations are run in backward mode, the `mc_backward` option has to be set which takes the indices of the start (`ix_start, iy_start`) and end pixels (`ix_end, iy_end`) in x- and y-direction as arguments. With this option it is possible to simulate a specific section instead of the entire domain. The output within the defined sample grid can further be refined by the `mc_backward_increment` option where the step width in x- and y-direction (`ix_step iy_step`) can be defined. This allows to calculate only every nth pixel in a certain dimension which is particularly useful when a high resolution is required but not all pixels are needed which could be the case for statistical analysis or to reduce computing time. For computations with DISORT the output quantity can be specified by setting the `output_user` option and adding the desired quantities like `edn` or `edir`. As these quantities are usually provided as function of wavelength the integrated irradiance can be obtained by using the option `output_process sum` which integrates/summarizes irradiances for all spectral bands.

Regardless of the type of radiation an appropriate parametrization of the spectral absorption by molecules needs to be chosen for simulating spectrally integrated irradiance (diffuse and direct) with libRadtran. The desired parametrization can be chosen

by setting the option `mol_abs_param` followed by the according keyword. Spectrally high resolved simulations can be performed using the Reptran parametrization by Gasteiger et al. (2014) which is very accurate but generally too expensive for simulating spectrally integrated irradiance. Therefore, so-called correlated-k methods which group absorption lines into wider spectral bands are more suited as they are still accurate – errors are typically smaller than 1% – and very fast with respect to computing time. libRadtran contains a variety of correlated-k parametrizations for the solar spectral range among which are the one by Kato et al. (1999) (keyword: `Kato`) and `Kato2` which is an unpublished improvement of `Kato` from 2003 with less wavelength bands but nearly same accuracy. The `Kato2` parametrization has been used for all simulations presented in this study as it makes simulations with high spatial resolution or high temporal resolution feasible.

The general overview about radiative transfer simulations with libRadtran and some more specific insights into GHI simulations presented in this section serve as a basis for more detailed descriptions of simulations carried out in the scope of this thesis. However, before more details are given on radiative transfer simulations, the datasets building the foundation of this thesis are presented and discussed in the next section.

3.2. Data basis for predicting GHI variability

This section aims to provide a detailed description of all datasets which were used to gain a better understanding of GHI variability and to develop models to predict it.

3.2.1. MetPVNet GHI data

As a main basis for this study serves a unique dataset of GHI measurements collected during two field campaigns in autumn 2018 (04.09.2018-15.10.2018) and summer 2019 (25.06.2019-07.08.2019) in the area of Kempten (Allgäu) Germany. Both campaigns were part of the research project MetPVNet which was funded by the German Federal Ministry for Economic Affairs and Energy and aimed to improve the power forecast for photovoltaic systems on small temporal and spacial scales (Meilinger et al., 2021). The GHI was measured at 24 different locations next to photovoltaic systems with a temporal resolution of one second using non-ventilated silicon photodiode pyranometers. From these measurement sites four were selected for this study (MS02, PV01, PV07, PV11). The sites were chosen because of similar terrain characteristics and a distance of at least 10 km between the individual locations to avoid strong correlations of the measurements. A general map of the measurement area and the location of the selected measurement sites is shown in Figure 3.1.

The described pyranometer network was operated by the institute for tropospheric research (TROPOS) and had already been operated successfully in previous measurement campaigns like HOPE (Madhavan et al., 2016) which was part of the HD(CP)^2

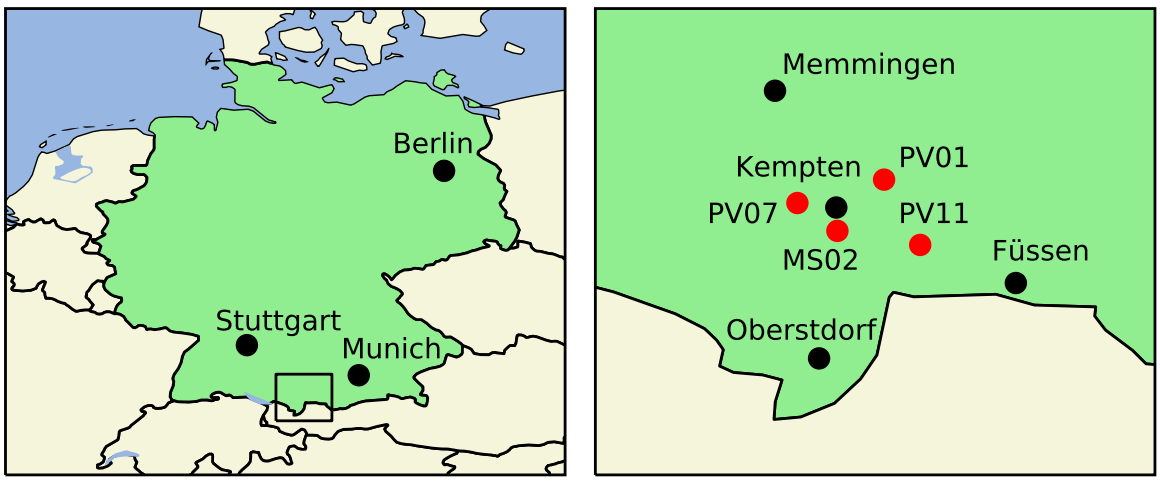

Figure 3.1.: Left: map of Germany (green) and neighboring countries (beige). The black rectangle shows the area around Kempten where the measurements took place. Right: Detail of section enclosed by the black rectangle with cities (black dots) and selected measurement sites (red dots).

research project. The silicon pyranometers have a very fast response to changes in irradiance (e.g. due to clouds) which makes them well suited for investigations of irradiance variability. However, there are certain drawbacks to this type of pyranometers. Two of them are related to the missing ventilation: first, strong heating of the pyranometer under warm and sunny conditions can cause a temperature related bias in measured GHI. The second effect is the formation of dew on the glass dome of the pyranometer during nighttime which has regularly been observed during the measurement campaigns. In the first hours after sunrise the dew droplets cause a strong refraction of the direct solar beam and hence lead to large errors in measured GHI. Another source of error is introduced by the pyranometer type itself. Silicon pyranometers have a smaller spectral bandwidth (400-1200 nm) compared to more sophisticated thermopile pyranometers (280-3000 nm). This gap can be compensated to a large extent by calibration, but the strong absorption bands of water vapor between 1200 and 3000 nm introduce an uncertainty. This uncertainty can not be entirely removed as the water vapor concentration varies significantly with location, season, and even within the course of the day while the calibration can only be valid for a specific water vapor concentration. The last error to be mentioned here is due to the deposition of aerosol on the glass dome. However, this effect is considered to be small as the pyranometers were inspected and cleaned regularly.

The described sources of error will be addressed in Section 3.3 where the process of data filtering will be described in detail.

3.2.2. COSMO-D2 forecasts

For this study operational forecasts of the German Weather Service (DWD) COSMO-D2 model (Baldauf et al., 2018) were used as input for the prediction of GHI variability. This model has been run operationally until February 2021 when it was replaced by the ICON-D2 model. COSMO-D2 is a convection permitting model with a horizontal resolution of 2.2 km. The forecasts considered in this study cover the timeframes of the two measurement campaigns of the MetPVNet project mentioned in the previous subsection. All forecasts were initialized at 0 UTC and provide hourly output with a lead time of 27 hours. From all available parameters 60 were selected which were seen relevant in terms of GHI variability. Among these parameters are wind speed, temperature at different levels as well as cloud properties like cloud fraction and liquid water mixing ratio. A complete list of the parameters can be found in Table A.2 in the appendix.

Two obviously important parameters though, the direct and diffuse solar irradiance calculated by the NWP model are not explicitly contained in the dataset as COSMO-D2 does not write them to the standard output. Instead, COSMO-D2 provides the temporally integrated irradiances since the initialization of the forecast normalized by the forecast lead time (Baldauf et al., 2018). These parameters, namely the average shortwave direct irradiance (ASWDIR_S) and the average shortwave diffuse downward irradiance (ASWDIFD_S) can be converted to the actual hourly averages of irradiance using the following equation:

$$I_n = \frac{I_n^* \cdot t_n - I_{n-1}^* \cdot t_{n-1}}{t_n - t_{n-1}} \tag{3.9}$$

I_n denotes the irradiance (direct/diffuse) at forecast step n which depends on the previously described COSMO-D2 output variables represented by I^* and the time t. Beside the calculation of irradiance a number of other parameters were derived from the existing COSMO-D2 variables which will be described in detail later on. These parameters can also be found in Table A.2 in the appendix.

3.2.3. AERONET AOD data

The AErosol RObotic NETwork (AERONET, Holben et al. (1998); Giles et al. (2019); Sinyuk et al. (2020)) is a global network of ground-based filter sun photometers which are used to derive optical properties of atmospheric aerosol from direct and diffuse sunlight. Measurements are routinely performed for seven wavelength channels between 340 and 1020 nm from which AOD, single scattering albedo (SSA) and other aerosol properties are derived.

During the MetPVNet measurement campaigns in 2018 and 2019 an AERONET sun photometer was operated by TROPOS on the rooftop of the Kempten University of Applied Science. In total 45 days of measurements are available during the measurement campaigns, 30 for the campaign in 2018 and 15 for the one in 2019. For the remaining

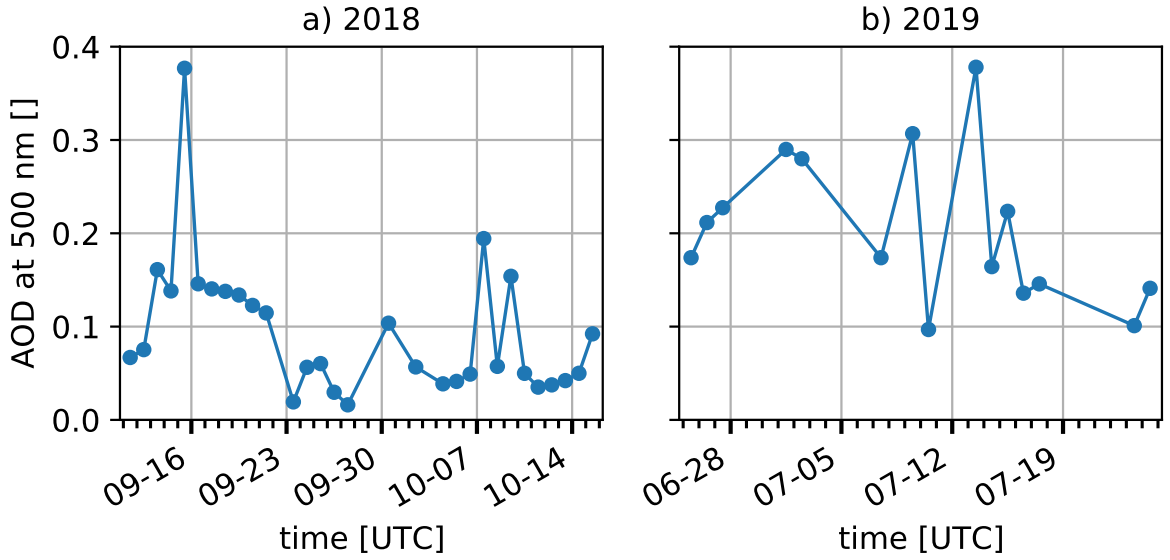

Figure 3.2.: AERONET daily average AOD at 500 *nm* wavelength from the measurement station at Kempten (Germany) for the MetPVNet measurement campaigns in 2018 (a) and 2019 (b). Days without data were filled by the average over all data available for the respective campaign.

days of the campaigns no data are available which is due to a too large cloud fraction or because the measurements have been rejected by the quality control algorithm. The sun photometer was also operated with some interruptions in between the two campaigns where another 86 days of data could be collected. For the radiative transfer simulations in this thesis level 2 (quality-assured) daily average AOD data based on the Version 3 AERONET algorithm were used. To have reasonable AOD data also for days without measurements the mean AOD for each spectral channel was calculated over all available data for the respective campaign period and assigned to the missing days. The daily averaged and processed AOD data for a wavelength of 500 *nm* are shown in Figure 3.2. At this point a short comment should be made on the reason for the use of daily average AOD. Under optimal (cloud free) conditions AERONET provides data at a temporal resolution of several minutes which may leads to the question of why this high resolution was not used here. In this study the AERONET AOD data from Kempten were used for the simulation of clear sky GHI which served as input for the derivation of GHI variability. However, under clear sky conditions were many data are available it is less important to have very accurate clear sky GHI as the variability is zero. On the contrary in broken cloud conditions where the variability is high, but only few measurements are available, it is important to get a good estimate of the clear sky GHI. Consequently, it is sensible to compute the daily averaged AOD to obtain consistent results for the variability computation.

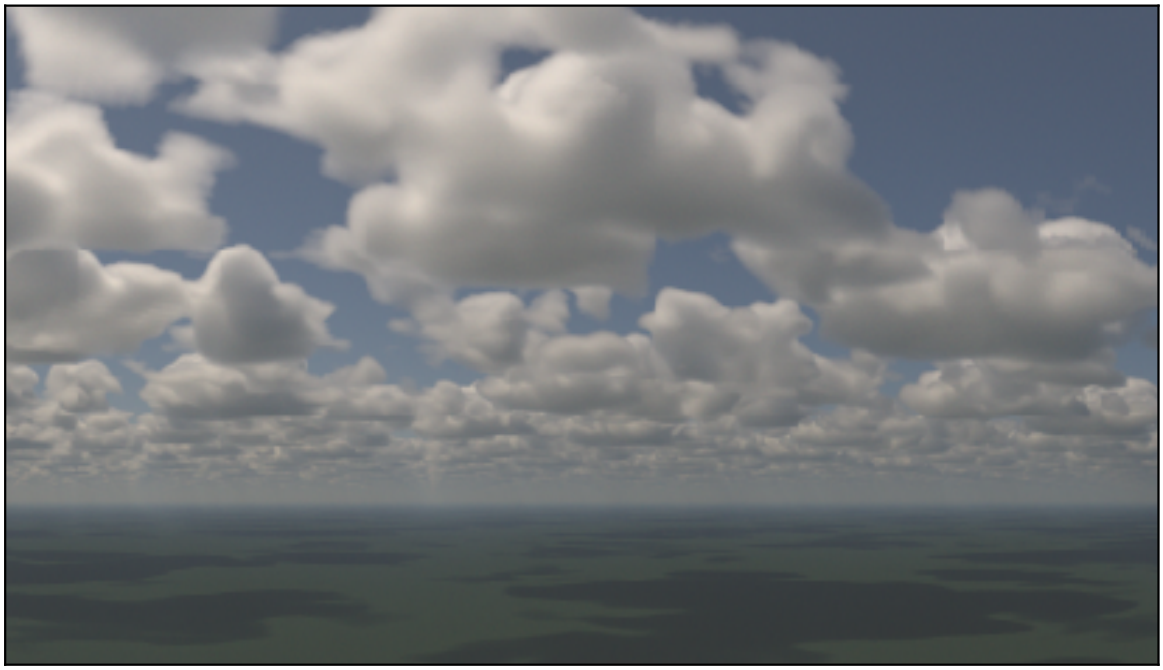

Figure 3.3.: RGB image of a LES cloud scene with a cloud fraction of 43% simulated with the panorama option of MYSTIC.

3.2.4. LES cloud data

To investigate general physical relationships between GHI variability and cloud properties like for instance cloud fraction, shallow cumulus cloud fields simulated with the University of California Los Angeles large eddy simulation (UCLA-LES) model (Stevens et al., 1999, 2005; Stevens and Seifert, 2008) were used as input for radiative transfer simulations. The LES runs were conducted at a horizontal resolution of 25 m with a domain size of 6.4x6.4 km^2 and a vertical domain extent of 4 km. The three-dimensional cloud microphysical properties of the cloud fields required for simulations with libRadtran were obtained as described in the study by Črnivec and Mayer (2019) who used the same LES cloud data. This dataset contains in total 15 cloud scenes from which seven where selected containing cloud fractions between 12 and 100%. These cloud scenes served as input to simulate 2D surface fields of GHI with MYSTIC in 3D and 1D mode. The detailed setup of these simulations will be described at a later point. To get a better impression of the appearance of the cloud fields Figure 3.3 shows a panorama image of a LES scene simulated with MYSTIC. The shallow cumulus clouds with complex shapes in the upper part of the image are casting dark shadows on the surface which are responsible for the strong irradiance fluctuations investigated in this study.

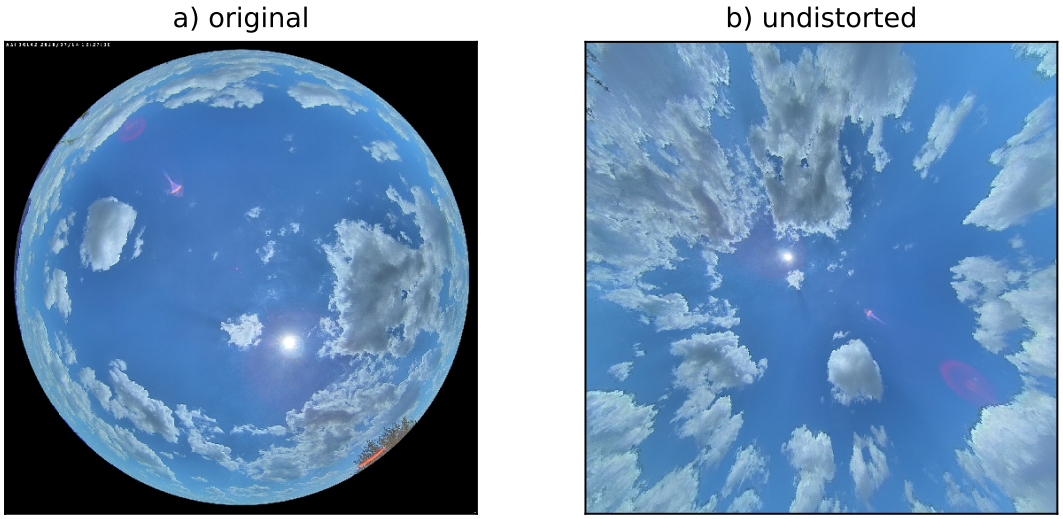

Figure 3.4.: a): original image taken by the All-Sky Imager operated during NETFLEX measurement campaign on 14th of July 2020 at 12:27 UTC. b): undistorted and rotated version of the image shown in a). Images were provided by Philipp Gregor (see also Gregor et al., 2023).

3.2.5. NETFLEX GHI and all-sky camera data

As the previously described LES simulations of shallow cumulus clouds are not related to MetPVNet measurement campaigns and hence do not reflect the orographic and atmospheric conditions it has to be shown first that these clouds are representative for investigating GHI variability in southern Germany. Furthermore, it is generally sensible to be careful when using model data to explain observations. Therefore, the LES clouds were used as input for 3D GHI simulations with MYSTIC which were then compared to measurement data. The goal of this procedure is to validate that the pattern of solar irradiance at the surface caused by the LES clouds is similar to irradiance patterns in the real world. This step is important as the irradiance fields derived from LES clouds with 3D radiative transfer are used in this study to analyze fundamental properties of irradiance variability as well as to infer the small scale variability based numerical weather prediction output. In the following the dataset used for the validation process will be described while the actual comparison will be shown in detail in Section 3.6.2.

To accurately validate irradiance distributions from radiative transfer simulations based on LES clouds against irradiance measurements it is necessary to make sure that the cloud conditions of both data sets are comparable. A very important quantity in this regard is the cloud fraction which was used as a main variable to decide weather measurements can be compared to a LES scene with a certain cloud fraction. While deriving cloud fraction from LES data is straight forward, good estimates for cloud fraction in observations can be obtained from All-Sky Imager (ASI) photographs. Although an ASI was operated during MetPVNet, there was no operational cloud fraction product derived

from this data. Furthermore, this ASI was not geometrically calibrated and there was no ventilation and heating system which reduces the amount of available data in the presence of rain or morning dew. Consequently, data from another measurement campaign which was part of the NETFLEX research project were used for the validation of the LES cloud scenes. The measurements were conducted by Philipp Gregor from the Meteorological Institute of the LMU Munich at a large PV power plant near Egling an der Paar about 50 km west of Munich. This location shows similar surface and terrain characteristics like the Kempten area with small hills in the surroundings and the surface covered by forests and fields which makes the NETFLEX measurement data a good alternative to the MetPVNet measurements. The NETFLEX project is funded by the Federal Ministry of Food and Agriculture with the goal of achieving an optimal operation of joint biogas and photovoltaic power plants using short term PV power forecasts. The measurement equipment consists of a ventilated EKO MS-80 class-A thermopile pyranometer which recorded GHI at a temporal resolution of 2 seconds and of a CMS-Schreder All-Sky Imager (ASI) system which provides RGB images every 10 seconds. An original RGB image captured by the ASI is shown in Figure 3.4a). The dataset contains 8 days of measurements in June and July 2020 (June: 2nd, 13th, 18th, 21st, 23rd, 30th; July: 2nd, 7th) with predominantly shallow convection and different levels of cloud fraction. To use the measurements for LES cloud validation a cloud fraction was derived from cloud masks based on ASI images to compare measurements and cloud scenes with similar cloud fraction. The cloud masks were derived from undistorted ASI images (see Figure 3.4b) with a temporal resolution of 2 minutes. Thereby, cloudy and clear pixels were distinguished by a combination of the normalized red/blue ratio (Li et al., 2011) and the normalized saturation/value (Jayadevan et al., 2015) technique. The cloud mask was already contained in the NETFLEX dataset processed and provided by Philipp Gregor.

In the following sections the presented datasets will be used for analyzing and predicting GHI variability. But as a first step a bunch of filtering criteria are applied to ensure the best possible quality of the data which will be described in the following.

3.3. Data filtering - GHI measurements

Measurements as well as model simulations can have certain types of errors which have to be considered when using and interpreting them. In this study GHI measurements recorded in the scope of the MetPVNet project (see Section 3.2.1) serve as a ground truth for the development and validation of methods for the prediction of GHI variability. To ensure reliability and accuracy of the GHI data within the limits of the expected measurement uncertainty (Madhavan et al., 2016) a manual quality control was performed on the GHI measurements. To detect anomalies in the GHI measurements which go beyond the expected measurement error, clear sky days were compared to a clear sky irradiance simulated with libRadtran. Thereby, a systematic deviation between measurement and simulation was regularly found in the late morning hours. With the help of an All-Sky camera operated at measurement location MS02 the cause for the

deviations in measured GHI could be identified. Just as the silicon pyranometers, the camera has an unventilated and unheated glass dome. During nighttime, the formation of dew droplets on the glass dome could be observed which led to a refraction of the direct solar beam as soon as the sun rises and hence caused an error in the measured GHI. An example for a dew contaminated measurement aside with an image taken from the All-Sky camera is shown in Figure 3.5. The selected day was cloudless from sunrise

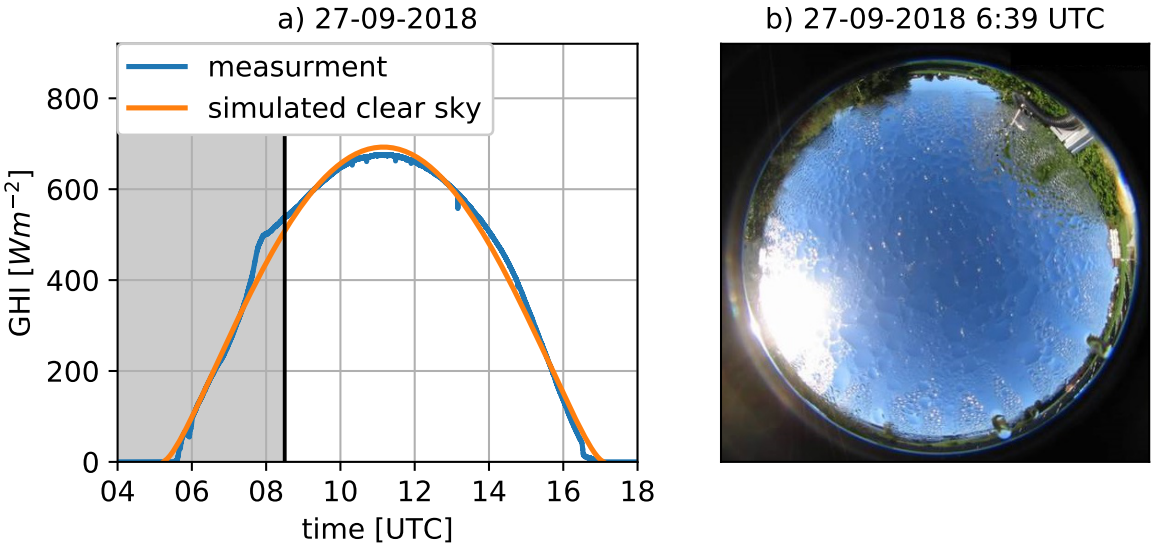

Figure 3.5.: a) Measurement of GHI at MS02 (blue) during a clear sky day in September 2018 with heavy dew formation on the dome of the pyranometer during nighttime compared to the clear sky GHI simulated with libRadtran (orange). The gray shading indicates the data rejected from the dataset because of the impact of dew on measurement quality. b) Image of an All-Sky camera at MS02 just after sunrise with glass dome covered with water drops due to dew formation during night.

until sunset so the measured GHI and the clear sky GHI simulated with libRadtran should be in close agreement. In the first hour after sunrise the dew does not seem to have a large negative effect on the measurement, but at around 7:30 UTC the measured GHI starts to rise well above the simulated clear sky GHI before it slowly converges towards the expected level. This deviation can clearly be attributed to the dew on the glass dome whose complete evaporation observed on the images of the All-Sky camera coincides with measurement reaching the expected GHI at around 8:30 UTC.

Based on these observations, the glass dome of the All-Sky camera at MS02 was used as an indicator for the existence and duration of dew on the glass domes of the pyranometers during both measurement campaigns. For practical reasons the dew was not classified for each individual day, but a general pattern was derived by inspecting selected days with clear skies which provide an upper limit for the duration of dew contamination. From these inspections, time frames for each month of the measurement

campaigns were defined and GHI data within these were excluded from all further processing steps. The time frames to be excluded for a certain month were determined as follows:

- 04.09.2018 - 30.09.2018: sunrise - 8:30 UTC
- 01.10.2019 - 15.10.2018: sunrise - 8:30 UTC
- 25.06.2019 - 30.06.2019: sunrise - 5:00 UTC
- 01.07.2019 - 31.07.2019: sunrise - 5:00 UTC
- 01.08.2019 - 07.08.2019: sunrise - 6:00 UTC

Even though the filter criterion was derived from the ASI located at MS02 it was also applied to the measurements at other stations. This was done as there were no other ASI or different devices from which a dew mask could have been derived at other measurement sites. However, it is likely that the filter criterion also applies to other locations as MS02 was located in a river valley which causes stronger and more persistent dew formation than at other locations due to cold and moist air accumulating in the valley.

Despite the loss of a considerable amount of data, the presented filter criterion for the GHI measurements helps to make the dataset and all conclusions drawn from it more robust and reliable. Transferring the excluded timeframes to the available hourly COSMO-D2 forecasts leaves a total number of 3459 individual forecasts steps in the dataset. The next section tries to assess the prediction error of the COSMO-D2 forecasts based on the filtered GHI measurements.

3.4. Data filtering - COSMO-D2 forecasts

The NWP forecasts of the COSMO-D2 model serve as input for the prediction of GHI variability. These NWP forecasts exhibit a certain error which affects the quality of the variability forecast. For instance there might be errors in prediction of the cloud fraction, wind speed, or other variables which could be related to GHI variability. Therefore, the error of the predicted GHI variability will be a mixture of the COSMO-D2 forecast error and the error introduced by the model used for the variability prediction. To assess the performance of the variability prediction it is desirable to distinguish between errors of the variability prediction itself and errors originating from the COSMO-D2 data. This can be achieved by applying a filter criterion to the COSMO-D2 forecasts which separates “accurate” from “inaccurate” forecasts. Such a filtering procedure not only allows to separate errors in the input data from errors of the variability model but also provides the possibility to assess the ability of the variability model to compensate for errors in the COSMO-D2 forecast.

As a first step of the filtering procedure a measure for the COSMO-D2 forecast quality was defined, which will be shown in the following section.

3.4.1. Quality measure

The first requirement for a quality measure which separates accurate from inaccurate COSMO-D2 forecasts is the availability of the measure in model and measurement data. Furthermore, the measure should be linked to predicted quantity which is the variability of GHI in this case. It would therefore not be a useful choice to filter the NWP data based on 2-meter temperature as the relation between temperature and variability is unknown. A variable which is probably more closely related to variability is the cloud fraction as the consecutive change between sunny and cloudy patches is the main cause for short term changes in GHI. However, cloud fraction is a very problematic measure in particular because there is no clear and quantitative definition of what a cloud actually is. Additionally, there are large differences in cloud fraction estimated from different instruments depending on their sensitivity and field of view which was for instance shown by Stevens et al. (2019) for remote sensing instruments mounted on the HALO research aircraft. Besides, there are no appropriate instruments like All-Sky cameras at all measurement locations. Therefore, it was decided to use GHI as a quality measure because measurements at all sites are available and just like the cloud fraction GHI is directly linked to the prevailing cloud conditions.

As the GHI extracted from the COSMO-D2 forecasts (see Section 3.2.2) reflects the average of the last hour before the timestamp of the forecast, the one-second GHI measurements were averaged over the same time period to achieve a fair comparison between model and measurement. The hourly average for pyranometer measurements at the selected sites (MS02, PV01, PV07, PV11) was only calculated when the considered time interval contained at least 80% of valid data. The according GHI from COSMO-D2 was retrieved for the grid point closest to the respective measurement site.

In the next step a threshold had to be determined which separates between accurate and inaccurate forecasts. The first important factor which has to be considered in this regard is the uncertainty of the GHI pyranometer measurements which will be discussed in the next section.

3.4.2. Assessment of ground truth error

As a first step in the procedure of determining a filter threshold for COSMO-D2 forecasts based on GHI it is important to determine the magnitude of the measurement uncertainty for GHI. A good estimate of this uncertainty is mandatory to avoid excluding accurate COSMO-D2 forecasts because of errors in the GHI measurement but also to make sure that inaccurate forecasts are reliably rejected.

In Subsection 3.2.1 it has already been mentioned that there are several deficits of unventilated silicon pyranometers can cause errors in the GHI measurement. The measurement uncertainty of the silicon pyranometers operated during MetPVNet measurement campaigns has been characterized in detail by Madhavan et al. (2016).

In Table 3 of their paper the irradiance uncertainty is shown for two different solar zenith angles (30° and 80°) and two different values of GHI (50 and 1000 Wm^{-2}). Additionally, they provide values for a 'Standard uncertainty' and an 'Expanded uncertainty'. Focusing on the 'Standard uncertainty', the measurement uncertainty for an insulation of 1000 Wm^{-2} at a SZA of 30° including soiling and possible errors in the tilt of the instruments, is on the order of 50 Wm^{-2} or 5%. Considering the uncertainty for an SZA of 80° in combination with the 'small signal' of 50 Wm^{-2}, an uncertainty of about 16 Wm^{-2} (32%) can be assumed. In fact, simulations with the radiative transfer model showed that an insulation value of 120 Wm^{-2} for this SZA would be more realistic which would lead to an absolute uncertainty of nearly 40 Wm^{-2} when assuming 32% relative uncertainty. Based on these estimations, it was decided to attribute an absolute uncertainty of $\pm 50\ Wm^{-2}$ to the GHI measurements, independent of the SZA.

Besides the uncertainty of the GHI measurements a second factor related to the COSMO-D2 forecasts was considered for estimating the filter threshold. The idea behind these considerations and a quantitative assessment thereof will be described in the following sections.

3.4.3. COSMO-D2 forecast error

In the context of GHI variability a deviation between measured and predicted GHI does not necessarily mean that the NWP forecast has to be rejected. This can be the case if the overall atmospheric state is predicted fairly well, but errors in atmospheric optical properties lead to an error in the GHI predicted by the NWP which may be irrelevant for GHI variability. The first example for such an error are deviations in the Aerosol Optical Depth (AOD) under clear sky conditions which are discussed in the following.

COSMO-D2 AOD error

COSMO-D2 uses climatological AOD values from Tegen et al. (1997) which are a good estimate on average but can deviate considerably from the actual AOD on a daily basis. Under clear sky conditions deviations in AOD can cause significant errors in the predicted GHI. For the data filtering this means that forecasts may be classified inaccurate in cases where COSMO-D2 correctly predicts clear sky but exhibits errors in the AOD when the resulting error in GHI exceeds the defined threshold. From the perspective of GHI variability, correctly forecasted clear sky should be kept in the dataset as clear sky doesn't cause temporal fluctuations in GHI and hence the variability could correctly be inferred based on the COSMO-D2 forecast independently of the error in AOD.

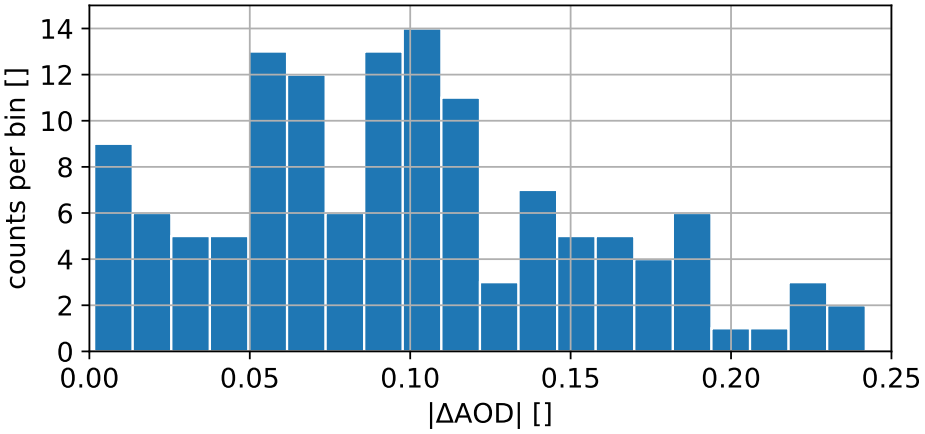

Figure 3.6.: Histogram of absolute deviation of AOD at 550 nm between AERONET daily average AOD and the climatology by Tegen et al. (1997) for all available data between autumn 2018 and summer 2019 at Kempten. AERONET data were spectrally interpolated to 550 nm wavelength and the climatological AOD was interpolated in space to match the measurement location.

To avoid the rejection of correctly predicted clear sky due to errors in the climatological AOD of COSMO-D2 a typical AOD error was derived in a first step. As a ground truth for the ambient AOD at Kempten all available AERONET daily averaged level 2 Version 3 data were used which are in total 131 days of measurements (for details see 3.2.3). These data were compared to the AOD of the climatology by Tegen et al. (1997) which contains monthly average AOD for five different aerosol species (e.g. black carbon, sea salt) at a wavelength of 550 nm. The resolution of the dataset is 4x5° (≈480x600 km^2) and it covers the entire globe. To retrieve the climatological AOD for each measurement the nearest neighbour grid point to the location of the AERONET measurement was selected. Subsequently, the AOD of all five species was summed up for each month to receive the total AOD for each month. For the comparison between measurement and climatology, the difference between each measurement of AOD and the climatological AOD of the corresponding month was calculated. As AERONET sun photometers do not measure at 550 nm the AOD was interpolated using an exponential fit according to Angström (1929) to be able to compare measurements and climatology. Figure 3.6 shows the resulting histogram of the absolute deviation between measurement and climatology for the 131 available days between September 2018 and August 2019. The absolute deviation was used here as possible biases in AOD which lead to systematic deviations in GHI will be considered later on. Most of the time the deviation in AOD is around 0.1 but reaches more than 0.2 in several cases. Based on these findings a deviation in AOD of 0.2, which roughly corresponds to the 95th percentile, was chosen to derive an upper limit for the expected error in GHI due to the use of climatological AOD data.

To translate the deviation in AOD into an error in GHI, radiative transfer simula-

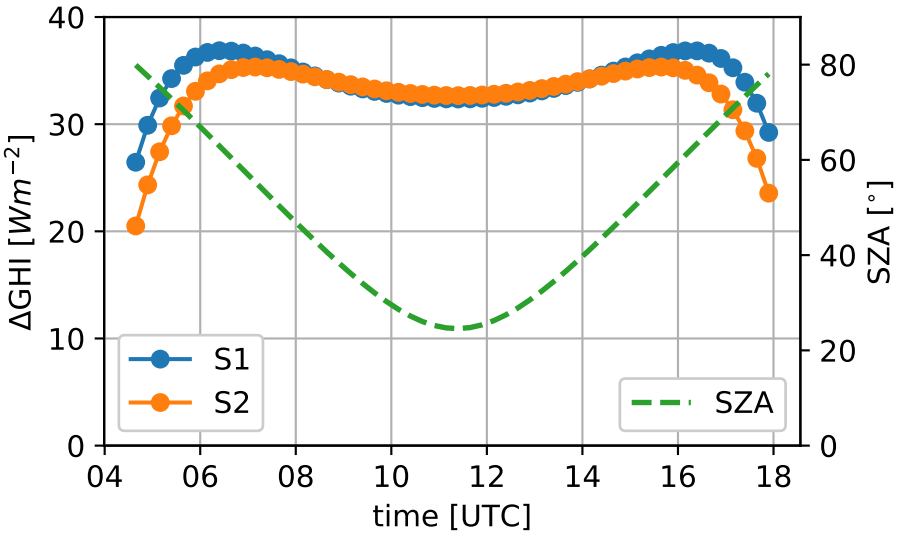

Figure 3.7.: Diurnal cycle of GHI error caused by differences in AOD at 550 *nm* for two different cases. S1 depicts the difference between an AOD of 0.0 and 0.2 (AOD0.0 - AOD0.2) and S2 the difference between AOD 0.1 and 0.3 for an exemplary diurnal cycle of the 1st of July at MS02. The dashed green line shows the course of the solar zenith angle (SZA).

tions with libRadtran (Mayer and Kylling, 2005; Emde et al., 2016) were conducted covering two scenarios for the AOD deviation. In the first scenario (S1) the difference in GHI between simulations with AOD 0.0 and 0.2 at 550 *nm* wavelength was investigated and in the second one (S2) AOD was set to 0.1 and 0.3. These two scenarios were defined to get an estimate for the influence of the initial level of AOD compared to an atmosphere with an 0.2 lower/higher AOD level. The simulations were performed with the DISORT solver (Stamnes et al., 1988; Buras et al., 2011) of libRadtran using OPAC continental average species file for the AOD optical properties. Furthermore, the mid-latitude summer atmosphere was used and AOD was adjusted using the `aerosol_set_tau_at_wavelentgth` option. A detailed list of the libRadtran settings is shown in Table 3.1. For each AOD level a diurnal cycle of GHI was simulated using the suns course on the 1st of July for the location of MS02 (MetPVNet). The resulting curves of GHI were subtracted from each other for both scenarios, which is shown in Figure 3.7. Both curves (blue and orange) are in close agreement with quickly increasing error up to an SZA of around 70° were the maximum error of 37 Wm^{-2} for S1 and 35 Wm^{-2} for S2 is reached. Afterwards the absolute error for both scenarios decreases slowly reaching the minimum of around 33 Wm^{-2} at noon for a SZA of 25°. As the error is larger than 30 Wm^{-2} most of the time in both scenarios but hardly exceeds 35 Wm^{-2}, a value of ± 35 Wm^{-2} was chosen as upper estimate for the error in GHI caused by the use of the aerosol climatology by Tegen et al. (1997) in COSMO-D2 simulations.

Parameter	Value
`solver`	DISORT
`atmosphere_file`	midlatitude summer (afglms)
`albedo`	0.1
`aerosol_species_library`	OPAC
`aerosol_species_file`	continental average
`mole_abs_param`	kato2
`aerosol_set_tau_at_wvl`	550 0.0/0.1/0.2/0.3

Table 3.1.: LibRadtran settings for simulating different values of AOD to derive an upper limit for the GHI error in COSMO-D2 under clear sky conditions

COSMO-D2 COD error

Analogous to the previously described case with correctly predicted clear sky conditions one also can think of situations where a cloudy sky is predicted correctly by COSMO-D2 but errors in the cloud optical depth (COD) result in a significant GHI error. Small deviations in COD may have small effects on variability but significant influence on GHI. Therefore, it would be desirable to also keep these cases in the dataset which makes it necessary to estimate a characteristic COD error for COSMO-D2 which could then be translated to an error in GHI like it was done for the AOD. To the knowledge of the author, there is no study to this date which explicitly examines the deviation of COSMO-D2 COD in comparison with satellite or ground based methods or the resulting effect on GHI. However, the study by Frank et al. (2018a) investigates GHI deviations in COSMO-REA6 reanalysis and finds a mean bias error (MBE) on the order of 15 Wm^{-2} and a mean absolute error (MAE) of 70 Wm^{-2} under cloudy sky conditions. As this study uses reanalysis data instead of operational forecast which additionally have a lower resolution and partly different atmospheric setup – e.g. the AOD was taken from Tanré et al. (1984) in REA6 instead of Tegen et al. (1997) in COSMO-D2 – it can only give a hint on the magnitude of errors to be expected. Furthermore, the errors derived by Frank et al. (2018b) are valid for 15 minute averages of GHI, but in the presented study hourly averages of GHI between model and measurements are compared. For statistical reasons, averaging over longer timescales significantly reduces the difference between model and measurement as short term fluctuations as well as displacement errors of cloud shadows are smoothed out. Consequently, taking into account the higher spatial resolution of COSMO-D2 and the four times longer averaging interval for GHI, the error in GHI is likely to be considerably smaller than 70 Wm^{-2} for the COSMO-D2 forecasts. To obtain a filtering threshold which is consistently applicable in all cases an error of 35 Wm^{-2} was assumed for the GHI error caused by deviations in COD.

To validate that this assumption of the COD induced GHI error translates to sensible errors in COD, radiative transfer simulations with libRadtran were performed. As before the DISORT solver of libRadtran (Stamnes et al., 1988; Buras et al., 2011) was used together with a single cloud layer whose optical thickness was scaled using

`wc_modify tau550 set` option. Simulations were performed for three baseline CODs (10, 20, 30) at fixed SZA of 30°. A list of the libRadtran settings can be found in Table 3.2. Starting from the baseline, the COD was successively increased to receive a relation between the difference in COD and the resulting GHI error. The result of these

Parameter	Value
`solver`	DISORT
`atmosphere_file`	US Standard (afglus) (Anderson et al., 1986)
`sza`	30°
`albedo`	0.1
`mole_abs_param`	kato2
`wc_modify`	tau550 set 'COD'

Table 3.2.: LibRadtran settings to translate deviations in cloud optical depth (COD) to deviations in GHI.

simulations are shown in Figure 3.8. For the baseline COD of 10 an increase by 1.5 to a COD of 11.5 is necessary to change GHI by 35 Wm^{-2} and for the baseline COD of 30 an increase to 36 is needed while the increment for COD 20 lies in between. These COD increments are equal to relative deviations between 15% (COD 10) and 20% (COD 30).

At this point the arising question is whether these numbers are realistic in some way and if they can be classified as a sensible error estimation. To answer this question it is useful to have a look at the microphysical parameters determining the COD of a cloud. Therefore, a simple equation can be considered which describes this relation in good approximation and can for instance be found in Liou (2002):

$$\tau = \frac{3\,LWC}{2\,r_{\text{eff}}\,\rho_w}\Delta z \tag{3.10}$$

The COD (τ) is a function of the Liquid Water Content (LWC) in gm^{-3}, the effective droplet radius (r_{eff}) and the geometrical thickness of the cloud Δz. For r_{eff} and LWC typical values of 10 μm and 0.1 gm^{-3} were assumed respectively which results in a geometrical thickness of 667 m for a COD of 10 and 2000 m for a COD of 30. In the following only the smallest and largest COD (10 and 30) will be discussed as the results for the intermediate COD of 20 are always in between these upper and lower limits. The change in COD of 1.5 (COD 10) and 6 (COD 30) can be converted to a change in the cloud parameters by means of Equation 3.4.3. For Δz the COD changes would result in a deviation of 100 m for a COD of 10 and 400 m for a COD of 30. As the vertical gridspacing of COSMO-D2 above the boundary layer ($\approx$ 1km) is about 100 m increasing to already 200 m at 2 km the estimated changes in Δz correspond to a difference of roughly one or two model layers which appears to be a realistic error for the vertical extent of clouds in COSMO-D2 forecasts. Similarly, the COD increments can be converted to changes in the effective droplet radius. Keeping the other parameters

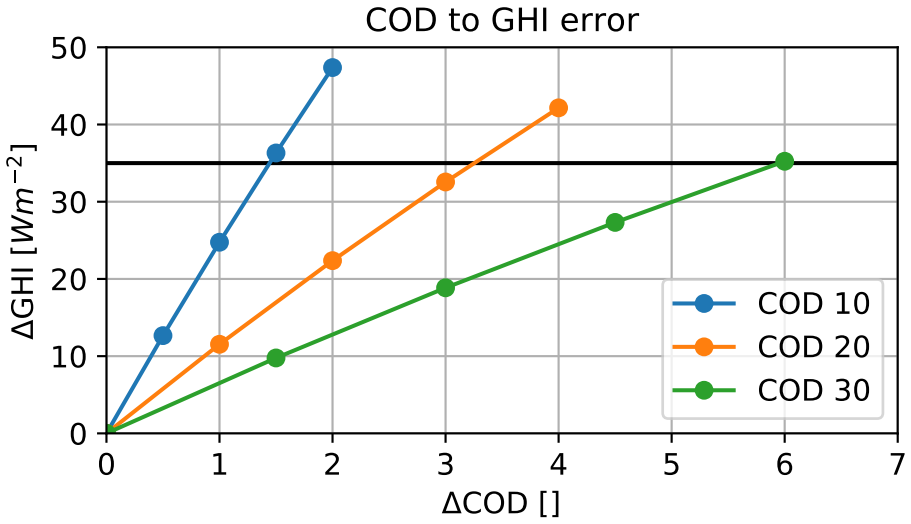

Figure 3.8.: Absolute GHI error as a function of COD deviation for different initial values of COD. COD deviations are always positive starting from the baseline COD of 10 (blue), 20 (orange) and 30 (green).

fixed, changes of 1.3 μm and 1.7 μm are enough to cause deviations to change the COD by 1.5 and 6 for the respective baseline COD. However, remembering that COSMO-D2 operates a one-moment bulk scheme for simulating cloud microphysics (see Section 2.4.1), which only predicts the mass of water for cloud droplets, makes clear that a certain r_{eff} has to be assumed to translate the predicted mass of water to COD. The LWC of the cloud is the last parameter which can be considered as a source of error for deviations in COD according to Equation 3.4.3. Having a look into the LWC of high resolved (25 m) realistic shallow convective clouds from the LES simulations described in Section 3.2.4 shows that the LWC can deviate by more than 100% within a single cloud. As mentioned before, COSMO-D2 operates a relatively simple one-moment bulk scheme and has to use further parametrizations to diagnose cloud water at unresolved scales below 2.2 km. Relating these facts about COSMO-D2 to the large fluctuations of LWC encountered in the LES simulations with almost 100 times higher resolution makes it plausible that errors in the LWC on the order 15 to 20% are likely to occur in COSMO-D2 forecasts. These relative errors correspond to an absolute error of 1.5 (COD 10) and 6 (COD 30) which lead to the assumed deviation in GHI of 35 Wm^{-2}.

Beside the fact that each parameter could cause the estimated deviations in COD on its own, one has to be aware that even smaller changes in each of the parameters at the same time can lead to the same result. However, one has to be aware that errors in the individual parameters may not be independent as for instance the effective droplet radius might be parametrized as a function of the LWC or LWP (LWC integrated over height), but information on the actual algorithms implemented in NWP models like COSMO are hardly available. In summary the radiative transfer simulations showed that an GHI

error of 35 Wm^{-2} in overcast cloud situations is caused by COD deviations of 15-20% for clouds with COD between 10 and 30. These deviations in COD can be provoked by small deviations in macro- and microphysical cloud parameters like geometrical cloud thickness or cloud liquid water content (LWC). It was made plausible that due to deficits in the horizontal and vertical resolution in combination with the simple one-monent bulk microphysics, errors in the cloud optical properties of COSMO-D2 forecasts on the order of 35 Wm^{-2} are likely to occur which motivates that it is a sensible choice to assign this error in cloudy situations.

Further uncertainties

Apart from deviations in the AOD and COD causing errors in GHI predicted by COSMO-D2, there are some more sources of error which are of smaller magnitude. One type of such error sources is the concentration of absorbing trace gases. In the solar spectral range the most important radiatively active gas is water vapor which has strong absorption bands in the visible and infrared solar spectrum. Therefore, errors in the prediction of water vapor by COSMO-D2 also result in errors of simulated GHI. A comparison between the AERONET precipitable water product (Holben et al., 1998; Giles et al., 2019; Sinyuk et al., 2020) and COSMO-D2 forecasts during MetPVNet showed deviations up to 25% for selected days which cause GHI errors of about 8 Wm^{-2} when used for radiative transfer simulations. Here it has to mentioned that the AERONET precipitable water product has a dry bias of 5-6% according to the study of Pérez-Ramírez et al. (2014) which offsets the observed difference between model and measurement.

Another atmospheric gas which can have at least some impact on GHI is ozone. However, the overall influence of ozone on GHI as well as its fluctuations in time are smaller than for water vapor. During MetPVNet measurement campaigns one event with very low ozone levels of 250 DU was observed which is about 30% less than the climatological average concentration of the US Standard Atmosphere (Anderson et al., 1986) (345 DU) used in libRadtran. This special case causes a change in GHI of 3 Wm^{-2} compared to the climatological mean. However, a value of 1-2 Wm^{-2} for usual fluctuations appears to be more realistic. From the presented considerations on the influence of radiatively active gases on GHI, an error of 10 Wm^{-2} was chosen as reasonable estimate.

As a last source of error 3D cloud radiative effects were considered which have already been partly discussed in Section 2.4.2. In partly cloudy situations solar radiation is scattered from cloud sides towards the surface which leads to enhancements of the GHI above clear sky level. Not only locally but also on spatial and temporal average these enhancements lead to a difference in GHI between 3D and 1D radiative transfer solvers like the one used in COSMO-D2. Simulations with MYSTIC in 1D and 3D mode with LES clouds as input suggest that the effect is on the order of 4 to 15 Wm^{-2} depending on cloud fraction and SZA. However, this is only valid in partly cloudy situations and does not play a role in clear and completely overcast situations. Therefore, it was decided

to weight this effect to a smaller extend by attributing an error of 5 Wm^{-2} to the 3D cloud effect. Choosing the 3D error on the lower range of the expected values can furthermore be justified by the fact that the error threshold is estimated by adding up all error components which makes it unlikely that a specific COSMO-D2 forecast would accidentally be rejected due to the assumption of a too small error for 3D effects.

Based on all previous investigations of the GHI error from measurements and COSMO-D2 forecasts a combined error was determined. To obtain an upper limit for the filter threshold the errors were assumed to be additive instead of assuming a Gaussian error propagation. This was done for two reasons: first, there are errors which were not considered in the error estimation as they hardly can be determined accurately. One can think for instance of the effect of the surroundings of the measurement location which could contain houses or trees which might shade parts of the diffuse radiation from the horizon. Secondly, there is a tradeoff between the size of the dataset and its purity achieved through the filtering process. When performing a too restrictive filtering with small error margins the size of the already small dataset might be reduced to an extent where it is no longer possible to derive reliable models from which solid conclusions can be drawn. Therefore, using the sum of all errors as an upper estimate makes sure that actually bad COSMO-D2 forecasts are filtered out by maintaining an acceptable amount of data within the filtered dataset. As a short recap the contributions of the individual error sources are listed below:

1. Measurement error: $\pm 50\ Wm^{-2}$
2. AOD/COD error: $\pm 35\ Wm^{-2}$
3. Trace gas concentration error: $\pm 10\ Wm^{-2}$
4. 3D cloud error: $\pm 5\ Wm^{-2}$

In summary these for contributions add up to a combined error of $\pm 100\ Wm^{-2}$. At this point it has to be made clear that the AOD/COD error only contributes once as the AOD error is assumed to be negligible in the presence of clouds and the COD error does naturally not occur under clear sky conditions.

3.4.4. Qualitative threshold validation

To show that the derived filtering threshold is a reasonable choice, three cases with different levels of GHI error between COSMO-D2 and pyranometer measurements were selected from the MetPVNet dataset. For each of these days the reflectance was derived from images of the SEVIRI (Spinning Enhanced Visible and InfRared Imager) instrument operated on the MSG satellite (Meteosat Second Generation) for the measurement area around Kempten. To compare the satellite observations with the NWP forecasts for the same area, reflectance images were calculated based on the COSMO-D2 output using the MFASIS operator (Scheck et al., 2016). Both, the reflectance from COSMO-D2 and MSG SEVIRI were processed and provided by Leonhard Scheck.

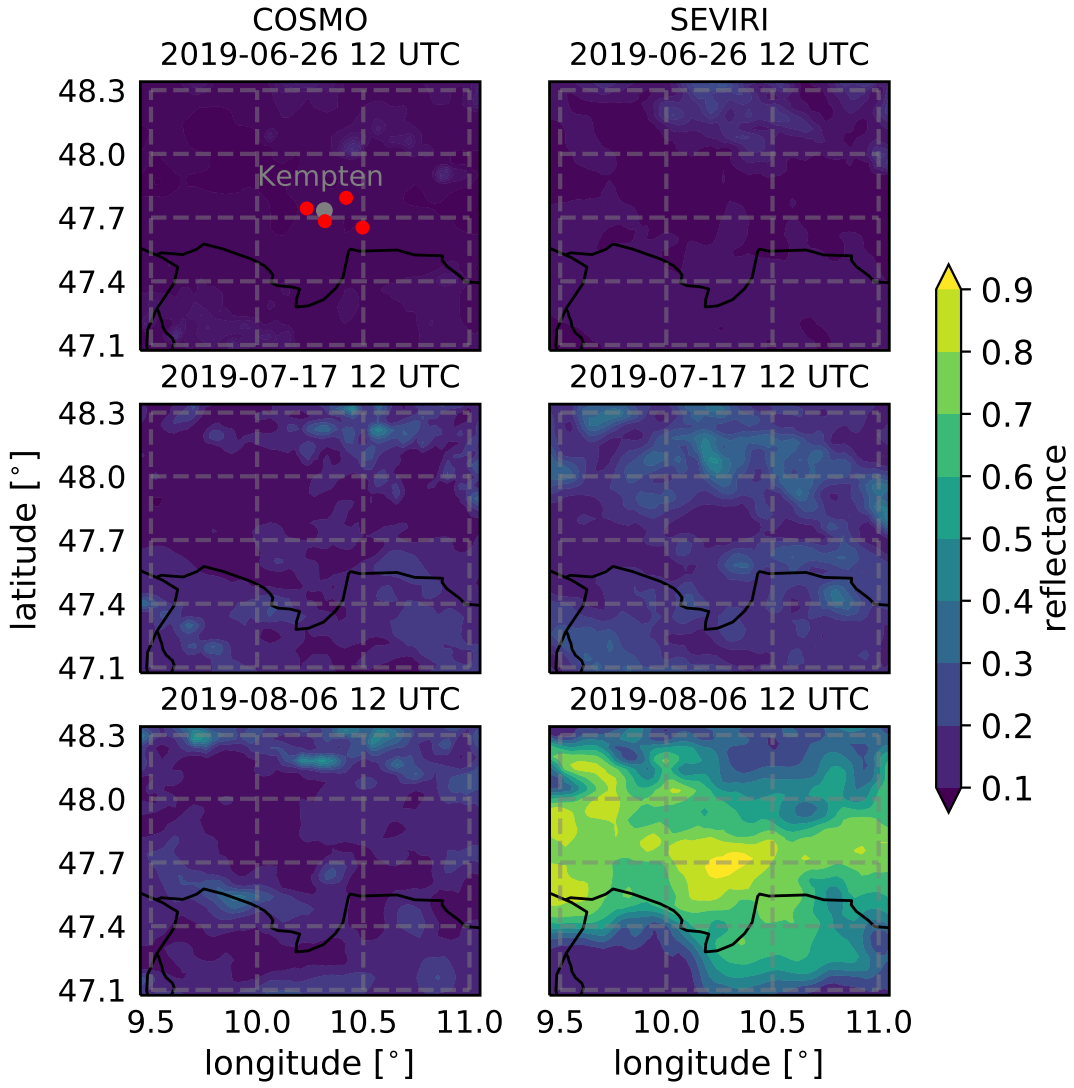

Figure 3.9.: Measured reflectance from MSG SEVIRI for the 0.6 μm channel (left column) interpolated to the COSMO-D2 grid and the corresponding simulated reflectance from COSMO-D2 model output using the MFASIS operator (Scheck et al., 2016) (right column) for the area of Kempten. The first row shows a case with a deviation in GHI between measurement and COSMO-D2 of less than 50 Wm^{-2} for the four considered measurements stations. The second and third row show cases with a deviation between 100 and 150 Wm^{-2} and more than 200 Wm^{-2} respectively. The upper left plot depicts the location Kempten (gray) as well as the 4 measurement locations (red dots).

The comparison of the resulting reflectance images is shown in Figure 3.9. For the first row, where the deviation in GHI between measurement and forecast is lower than 50 Wm^{-2}, model and measurement are in good agreement. The slightly lower reflectance in the southern part of the COSMO-D2 image is due to the use of climatological albedo which did not take the actual snow cover in the mountains into account. For deviations of 100 to 150 Wm^{-2} which is shown in the second row, there is a small but clear underestimation of reflectance due to a lower cloud cover in the COSMO-D2 forecast. The last case depicts a situation where CSOMO-D2 forecast fails heavily by predicting a nearly cloud free sky whereas there was actually a very high cloud cover with large COD resulting in reflectances larger than 0.8. In this case the deviations between modeled and measured GHI were larger than 200 Wm^{-2}.

The presented comparison between satellite observations and COSMO-D2 forecasts suggests that reflectance values start to deviate significantly at GHI errors above 100 Wm^{-2}. For the third panel of Figure 3.9 COSMO-D2 predicts only a very small cloud cover, but the observations show that there actually was a very high cloud cover of optically thick clouds causing GHI deviations of more than 200 Wm^{-2}. Considering this situation from the perspective of GHI variability the COSMO-D2 forecast could result in significant fluctuations of GHI when the small scattered cloud fields are advected over the measurement locations. In contrast, the actual cloud conditions result in very small fluctuations of GHI and hence low variability. Consequently, this case should definitely be removed by the data filtering procedure. For the second panel the situation is different as the deviation between measurement and forecast ranges from 100 to 150 Wm^{-2}. Here COSMO-D2 correctly predicts small and scattered cloud fields, but the cloud cover is smaller than observed. From a variability perspective the COSMO-D2 clouds would lead to smaller and less frequent fluctuations compared to the observations, however the difference is way smaller than for the case in the third panel. Reversing the statement made at the beginning of the paragraph, the described convergence between model and observation with decreasing GHI error suggests that at errors below 100 Wm^{-2} good agreement can be expected concerning GHI variability. These findings support that the derived threshold of ± 100 Wm^{-2} is a sensible choice for data filtering.

3.4.5. Applying filter threshold

Having derived and motivated a threshold of ± 100 Wm^{-2} the data filtering would have usually been applied centered around the 1:1 line between the GHI of COSMO-D2 and the measurement. However, the scatter plot shown in Figure 3.10 indicates a systematic bias between simulation and measurement. Comparisons between the used silicon pyranometers and more precise class A thermopile pyranometers showed that this bias can be explained to a large extent by a bias in the cosine response of the silicon pyranometers. To remove this systematic bias a quadratic regression curve was calculated using the `optimize.least_squares` method of the python library SciPy (Virtanen et al., 2020). The weight of outliers was reduced by choosing the option *loss=cauchy* where the

logarithmic rather than the linear distance is used for optimization. Subsequently, the threshold of $\pm 100\ Wm^{-2}$ was applied centered around the regression line which is also shown in Figure 3.10 (orange, solid). The dashed orange lines indicate the upper and lower boundary defined by the filter threshold. They enclose the region of the highest point density and separate it from more scattered data points which can also be seen as an indicator that the derived threshold is a justified choice.

Applying the filter criterion removed 29% of the data which were available after the dew filtering had been performed (see Section 3.3) yielding a total amount of 2468 samples in the dataset. This means that the remaining two thirds were considered for training and evaluation of the models described in the following.

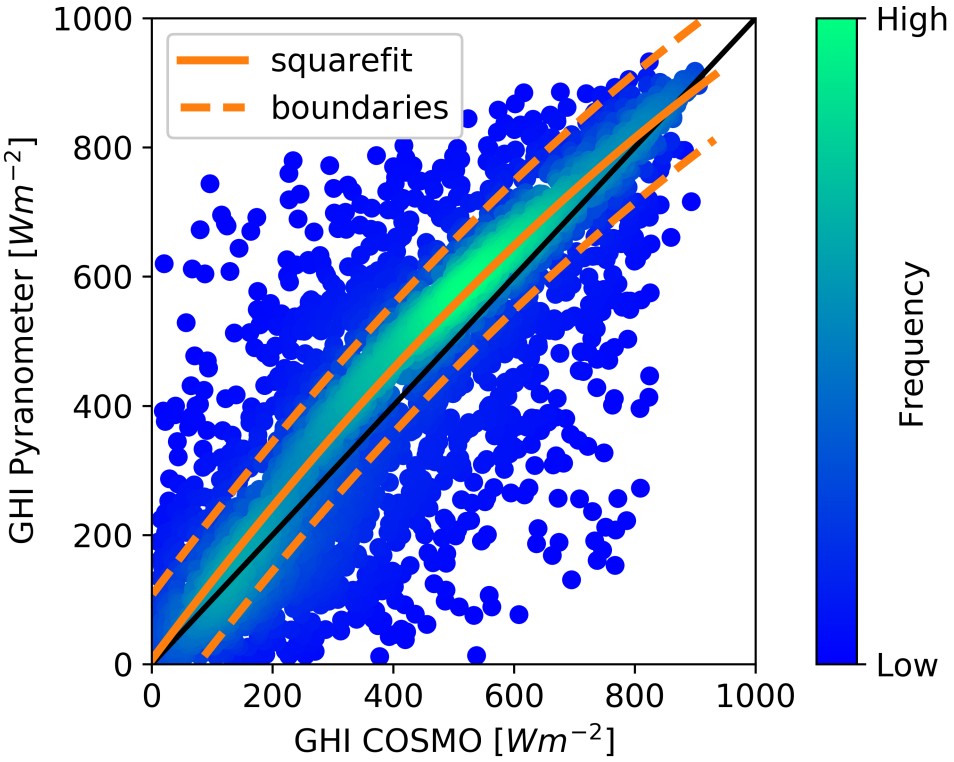

Figure 3.10.: Scatter plot comparing the hourly averaged GHI from silicon pyranometers at 4 selected sites (y-axis) with the GHI predicted by COSMO-D2 (x-axis). The 1:1 line is shown in black, the solid orange line indicates the quadratic fit function and the dashed orange lines mark the lower and upper boundary of the filter threshold ($\pm 100\ Wm^{-2}$).

3.4.6. Terminology: filtered and unfiltered dataset

In the previous sections, quality filtering procedures for the available GHI measurements during MetPVNet (see Section 3.3) and the COSMO-D2 forecasts (see Section 3.4) were presented. Based on these two filtering steps the dataset containing COSMO-D2 forecast

parameters and other quantities will be used in two different versions throughout this study. To avoid any kinds of misunderstanding the terminology for these different dataset versions will be described in the following.

1. The first dataset version is called the 'unfiltered' or 'not quality filtered' dataset and refers to the dataset which was not filtered for inaccurate COSMO-D2 forecasts as described in Section 3.4. However, this dataset was filtered for incorrect GHI measurements due to dew contamination (see Section 3.3).

2. The second dataset is mostly referred to as the 'filtered' or 'quality filtered' dataset. In this case, the dataset was filtered for dew on the pyranometer and also for inaccurate COSMO-D2 forecasts.

In summary the 'unfiltered' as well as the 'filtered' dataset were both filtered for dew contamination of the pyranometers. Consequently, the term 'filtered' or 'unfiltered' only refers to the filtering of inaccurate COSMO-D2 forecasts.

3.5. Quantifying variability of GHI time series

So far the term GHI variability was used to describe the parameter investigated in the scope of this thesis. However, 'GHI variability' is a very general term as it is not obvious if it refers to temporal or spatial variability and it also does not provide information on the involved scales. Therefore, this section provides a quantitative definition of GHI variability and describes the derivation of a ground truth variability from GHI measurements during MetPVNet campaigns (see Section 3.2.1).

3.5.1. Defining a measure of variability

This study investigates the temporal variability of GHI at a single location like a rooftop PV system. For the temporal resolution one-minute averages of GHI were chosen which also has been used in previous studies on GHI variability like for instance in Schroedter-Homscheidt et al. (2018). On the one hand this resolution is sufficiently high to resolve the strong high frequency fluctuations and on the other hand it is reasonably smaller than the temporal resolution of the NWP data of one hour.

As the goal of this study is to predict the GHI variability based on hourly NWP forecasts it is sensible to use a metric which condenses information about variability within a time frame of one hour into a single number. Schroedter-Homscheidt et al. (2018) reviewed a selection of such 'variability indices' which are in particular interesting for the application in energy meteorology. In this study the variability index V defined by Coimbra et al. (2013) was chosen as metric which is calculated as follows:

$$V = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(\frac{I(t)}{I_{clear(t)}} - \frac{I(t-1)}{I_{clear}(t-1)}\right)^2} \tag{3.11}$$

I represents the actual irradiance which is normalized by the simulated irradiance under clear sky conditions I_{clear}. The arguments t and $t-1$ describe the temporal index of two consecutive data points, the number of calculated differences N is determined by the size of the time interval and the temporal resolution of the considered data. The variability

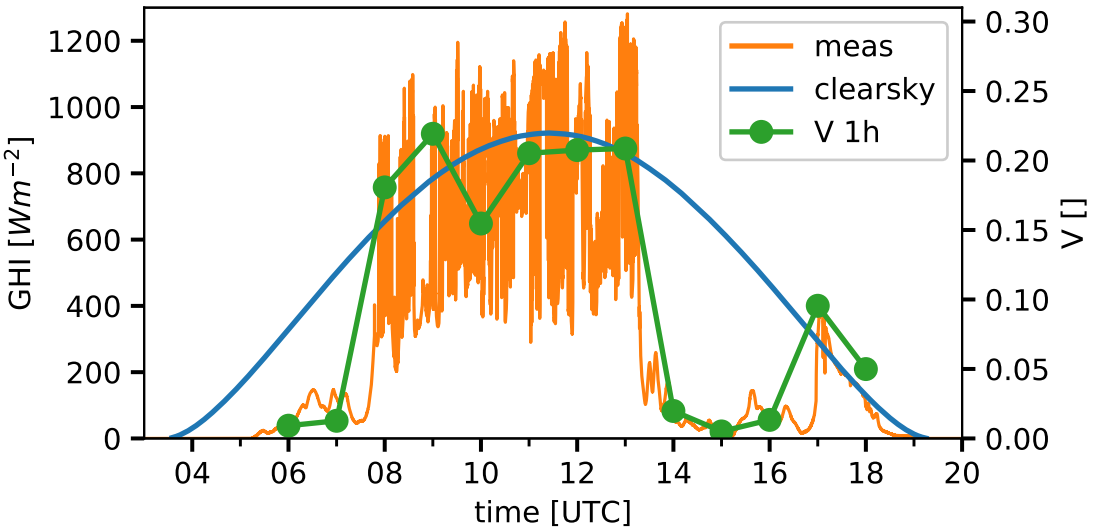

Figure 3.11.: Time series of measured GHI (orange) and simulated clear sky GHI (blue) at MS02 for the 7th July 2019. The hourly variability index V for one-minute averages of measured GHI is shown in green.

index V was selected because it fulfills two important requirements:

1. It provides numbers mostly in the range between 0 and 0.4 when calculated for hourly intervals of GHI measurements with a temporal resolution of one minute which allows for a straight forward interpretation.

2. The strong geometrical influence of the solar zenith angle (SZA) on the absolute variability is eliminated by normalizing the measured (simulated) GHI by the clear sky GHI

Thereby, the second requirement is of special importance when predictive variables for GHI variability need to be identified. By eliminating the strong and obvious geometrical influence of the SZA it is possible to discover predictive variables which would otherwise be considered less important.

To illustrate the behavior of V, Figure 3.11 shows a time series of measured GHI, simulated clear sky GHI and the corresponding variability V. V was calculated for one-minute averaged GHI values for time intervals of one hour centered around every full hour. The day starts with low values of GHI and V because of a closed cloud deck which breaks up in the late morning hours and leads to large GHI fluctuations and hence high variability. Having a closer look at the GHI curve between 6 UTC and 7 UTC, the

fluctuations relative to actual GHI are quite large, but as the overall GHI level is way below clear sky GHI the variability index turns out to be close to zero. At around 13 UTC V and GHI drop again followed by a small increase towards the end of the day. Throughout this study the variability index V is always calculated for hourly intervals. The underlying time series data of GHI, may it be measurements or simulations, always have a native resolution of one second and are averaged over one-minute intervals before V is calculated.

3.5.2. Ground truth variability for MetPVNet data

To obtain a ground truth variability dataset for all further steps of the model development the previously described variability index V was calculated for the filtered MetPVNet GHI measurements (see Section 3.3) for the four selected measurement locations MS02, PV01, PV07 and PV11 (see Section 3.2.1). As the calculation of V requires clear sky GHI as input, the DISORT solver of libRadtran (Stamnes et al., 1988; Buras et al., 2011) was used to simulate clear sky GHI for each measurement site over the respective time period of the campaigns. As detailed information on the DISORT solver and options in libRadtran can be found in Section 3.1 the setup for the clear sky simulation is described in less detail.

To correctly represent the atmospheric state, profiles of pressure, temperature and water vapor concentration from the COSMO-D2 forecasts (see Section 3.2.2) were taken as input for the simulations. Thereby, the data of the model grip point closest to the respective measurement site were used. Additionally, the daily average AOD from AERONET measurements at Kempten as described in Section 3.2.3 served as input. The AOD was integrated by applying an exponential fit according to Angström (1929) to the spectral AOD and using the `aerosol_angstrom` option together with OPAC continental average aerosol mixture. To obtain a representative spectral albedo for each measurement site the MCD43A1 version 6 product Schaaf et al. (2002) of the MODIS (MODerate resolution Imaging Spectrometer) instrument on board the Terra and Aqua satellites was used. The MCD43A1 product contains the three BRDF parameters of the AMBRALS model (see Section 3.1.2) for seven spectral channels between 470 and 2130 nm and has a resolution of 500x500 m^2.

As a first step the product was downloaded from the LP DAAC (Land Processes Distributed Active Archive Center) for the center day of both MetPVNet campaigns to provide a representative surface reflectivity for the respective time period. Then BRDF data of an area with 10 km radius around each measurement site were extracted from the dataset and a spectral interpolation based on the method described by Vidot and Borbás (2014) was applied to each 500x500 m^2 pixel. Subsequently, the spectrally interpolated BRDF parameters were averaged over the entire domain and converted to a spectral black-sky albedo applying the polynomial function described in Schaaf et al. (2002) with a solar zenith angle of 45°. For the simulation of clear sky conditions using the blacksky-albedo is a good assumption as usually about 90% of the radiation reaches the surface from the direction of the sun. The spectral black-sky albedo was integrated

into the radiative transfer simulations using the `albedo_file` option.

Using the previously described input data which are summarized in Table 3.3, the clear sky GHI was simulated for each site and each day of measurements with a temporal resolution of five minutes and were then interpolated to the one-second resolution of the pyranometer measurements. In the next step, simulation and measurement were averaged to a resolution of one minute by applying a centered average. Finally, the index V was calculated for hourly intervals centered around every full hour by means of Equation 3.11. The resulting dataset contains 3459 samples for the variability index distributed over the four measurement locations and the two measurement campaigns. This ground truth dataset of the variability index V was used for developing, training and evaluating the models aiming to predict GHI variability based on NWP forecasts. Before these models will be described, the next section will deal with the splitting of the dataset into a training and validation part which is necessary for model development.

DISORT Setup	
atmospheric profiles (T, p, water vapor)	COSMO-D2 forecast combined with afglus (Anderson et al., 1986) for altitudes above 22 km and other traces gases
molecular absorption	Kato2 (correlated-k, 32 bands), modified version of Kato et al. (1999)
aerosol library	OPAC (Hess et al., 1998)
aerosol species file spectral AOD adaptation	continental average aersol_angstrom with daily average AOD from AERONET
albedo	representative spectral black-sky albedo from MODIS MCD43 product

Table 3.3.: Setup for simulating clear sky GHI with DISORT to calculate the variability index V according to Coimbra et al. (2013) from the MetPVNet GHI measurements.

3.5.3. Training and validation dataset

For the optimization of each kind of model it is important to separate the dataset into at least two parts: one part on which tuning and optimization procedures are performed and a second one which can then be used to show if the implemented optimizations also generalize to data which were not involved in the optimization process. Therefore, 15% of all data points consisting of COSMO-D2 forecast variables and the variability index V

derived from GHI measurements were chosen randomly for the validation dataset which was excluded before the model development was started. This dataset was held out during all processing steps until the final model setup was determined. The remaining data were used to optimize and tune the models developed in the scope of this dissertation.

The separation between training and validation dataset was performed on the entire unfiltered dataset (see Section 3.4.6) which provides the opportunity to assess the impact of the data filtering performed in Section 3.4. As the validation dataset was chosen randomly every further filtering or data selection steps should preserve the ratio of 85% to 15% between training and validation dataset except for statistical noise. And in fact, for the unfiltered dataset with in total 3459 samples the validation dataset contains 518 samples and for the filtered dataset with in total 2468 samples, 370 thereof belong to the validation dataset. For both dataset versions these numbers fulfill the ratio of 85% to 15% for training and validation dataset.

3.6. Investigating GHI variability in idealized simulations

Profound knowledge about the properties of the target quantity is essential to build accurate and reliable models. Therefore, fundamental properties of GHI variability will be investigated in this section based on synthetic GHI fields generated using LES cloud data as input for radiative transfer simulations with MYSTIC.

3.6.1. MYSTIC simulations with LES clouds

As already described in Section 3.1.1, MYSTIC (Mayer, 2009) is capable of simulating GHI accounting for 3D effects in the radiative transfer through the atmosphere. To investigate the properties of GHI variability quantified by the variability index V (see Section 3.5) the LES cloud fields described in Section 3.2.4 were used to simulate spatially high resolved (25 m) surface fields of GHI with MYSTIC in 1D and 3D mode. The cloud fractions chosen from the LES are ranging from 12% to 100% and hence cover a variety of cases. Figure 3.12 shows an exemplary GHI field simulated with 3D MYSTIC and LES clouds. In this case a cloud field with a cloud fraction of 43% was used at an SZA of 30°. Because of the high spatial resolution of the LES cloud data of 25 m, the GHI has to be simulated for 64,000 individual pixels in each MYSTIC run which is very expensive concerning computational demand. As a result, simulating the entire domain on a single CPU would take several days or even weeks. To overcome this issue, the number of photons traced in an individual MYSTIC run was reduced significantly which was compensated by running many simulations in parallel to end up with the same number of traced photons as before. At this point a large advantage of the Monte Carlo technique comes into play: Each traced photon is an individual experiment with a certain outcome, and statistically it doesn't make a difference if 1000 simulations with a single photon or a single simulation with 1000 photons is performed. Therefore, the final result for the GHI can be calculated from the ensemble of simulations by averaging

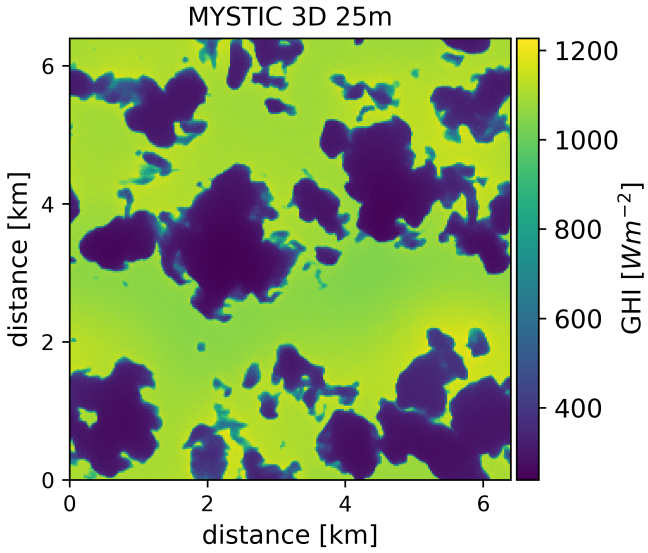

Figure 3.12.: Spatial field of GHI with a resolution of 25 m calculated with the 3D radiative transfer model MYSTIC using LES clouds with a cloud fraction of 43%. The solar zenith angle for the simulation was set to 30°.

over all simulations with reduced photon number to obtain the final result. In cases where simulations contain different number of photons the average has to be calculated by weighting each simulation by the number of photons and dividing by the sum of all photons.

The above procedure is in particular relevant for the simulation of the diffuse part of the GHI as each photon can experience many scattering events which increases the time needed for tracing. For simulations presented in this chapter either 100,000 or 200,000 photons per domain pixel were traced in backward mode for simulating the diffuse irradiance, depending on the specific investigation. These simulations were distributed over 50-400 simulations run in parallel on a computing cluster to reduce computing time. Furthermore, the VROOM option of MYSTIC was used to improve the accuracy of the simulations (see Section 3.1.2). In cases where many parameters were varied the mc_backward_increment option was used additionally to reduce the computational demand (see Section 3.1.3).

To get an idea about the accuracy of the presented simulations some numbers are provided in the following: for a medium cloud fraction of 43% using 200,000 photons a standard deviation averaged over the entire domain of 4.8 Wm^{-2} (1.5%) was achieved for the diffuse irradiance. Reducing the number of photons to 100,000 and using a cloud fraction of 53% yields a domain averaged standard deviation of 7.8 Wm^{-2} (2.4%). In contrast to the diffuse irradiance, the direct irradiance can be calculated with less effort according to Bouguer-Lambert-Beer law (see Section 3.1.3). Therefore, only 1000 photons in total were necessary to achieve an absolute accuracy of 0.6 Wm^{-2} (0.1%).

MYSTIC Setup	
atmospheric profiles	afglus/afglms (Anderson et al., 1986)
molecular absorption	Kato2 (correlated-k, 32 bands), modified version of Kato et al. (1999)
aerosol type	default (Shettle, 1990)
solar zenith angles (SZA)	$20°, 30°, 40°, 60°$
optical cloud properties	mie_interpolate
number of photons	direct: 1,000 diffuse: 100,000-200,000
VROOM	on
sample grid	256 256 0.025 0.025
backward	0 0 255 255
mc_backward_increment	3 3 (only for simulations with 100,000 diffuse photons)
albedo	0.1

Table 3.4.: Simulation setup for 3D radiative transfer simulations with MYSTIC and LES clouds as input.

Adding up diffuse and direct irradiance, the standard deviation of GHI turned out to be 5.0 Wm^{-2} or 0.6% (200,000 photons, cloud fraction 43%) and 7.9 Wm^{-2} or 1.1% (100,000 photons, cloud fraction 53%) on domain average. It has to be noted that the accuracy varies throughout the domain depending on the presence of clouds and their optical thickness in the vicinity of the considered pixel, but the provided values give a good idea of the accuracy of the performed simulations.

To ensure reproducible results, all important options used for the GHI simulations with LES clouds are listed in Table 3.4. Options specific to the respective case studies in this section will be described when appropriate.

3.6.2. Validation of LES clouds against measurements

To make sure that the LES cloud fields used for this study are realistic in the sense that the GHI fields based on these clouds are similar to observations, the histograms of simulated GHI were compared against measurements of GHI. For a fair comparison one needs to make sure that cloud conditions during measurement are similar to the LES scenes. Therefore, the NETFLEX dataset (see Section 3.2.5) was used which not only provides high quality and high resolution GHI measurements but also cloud masks from undistorted camera images. From these cloud masks the cloud fraction during the measurements was calculated and compared to the cloud fraction of the LES scenes. All further processing steps which were necessary to compare histograms of GHI for

measurements and simulations are described in the following.

As mentioned before the cloud fraction for individual camera images needs to be calculated. This can in general be done by calculating the ratio of the number of pixels classified as cloudy divided by the number of all pixels. However, some effects need to be considered to avoid errors in the calculated cloud fraction. The primary source of error originates from the viewing geometry of the ASI. The larger the Viewing Zenith Angle (VZA) gets the larger the fraction of the sky obscured by a cloud which leads to an overestimation of the actual cloud fraction (Henderson-Sellers and McGuffie, 1990). Furthermore, cloud sides are observed rather than the base area of the cloud which introduces additional errors. To get a good estimate of cloud fraction from ASI cloud masks Kassianov et al. (2005) provide guidelines which they derived from studies based on LES simulations of divers cloud fields. They found that the best estimate of the nadir (looking exactly vertical from above) cloud fraction can be achieved by using 15-minute averages of frequently sampled ASI images ($\approx$ every 30 s) using a Field Of View (FOV) of 100° which translates to a VZA of 50°. As the cloud mask of the NETFLEX data is available every two minutes the temporal averaged cloud fraction was calculated for an extended timeframe of 20 minutes centered around each available cloud mask by using the recommended FOV. Finally, the cloud fraction with a resolution of two minutes was interpolated in time to match the resolution of the GHI measurements of two seconds.

In the next processing step the GHI measurements were sorted into bins according to their SZA and cloud fraction. Thereby a bin size of 4° for SZA and a bin size of 20% for cloud fraction was chosen which is a compromise between accuracy and data availability. The cloud fraction bins were centered around 10%, 20% and 30% corresponding to the available cloud fractions from the LES cloud fields of 12%, 21% and 30%. In order to accurately compare the binned measurement data with simulated GHI based on LES cloud fields, the atmospheric conditions during the measurements were taken into account for the radiative transfer simulations. The simulations were conducted with MYSTIC in 3D mode (see Section 3.1) at the native resolution of the LES cloud data of 25 m and 200,000 photons for diffuse irradiance. For the optical properties of the aerosol the default aerosol of libRadtran based on (Shettle, 1990) was used which is a rural type aerosol suitable for summer and spring. As the AOD measurements of the closest AERONET site at Hohenpeissenberg show very low levels of AOD during large fraction of the eight available days of measurements the AOD in libRadtran was set to 0.05 for a wavelength of 500 nm. To account for the concentration of water vapor and the typical atmospheric temperature profile, the mid-latitude summer atmosphere (afglms) (Anderson et al., 1986) was chosen as input for the radiative transfer simulations. Furthermore, the incoming solar irradiance at top of atmosphere was adapted for Earth-Sun distance by choosing 180 for the `doy` (day of year) option in libRadtran. All these corrections were necessary to closely match the absolute values of the GHI between simulations and measurements which otherwise would be shifted relative to each other. More details on the radiative transfer simulation setup and further options used in the input files can be found in the previous section and Table 3.4.

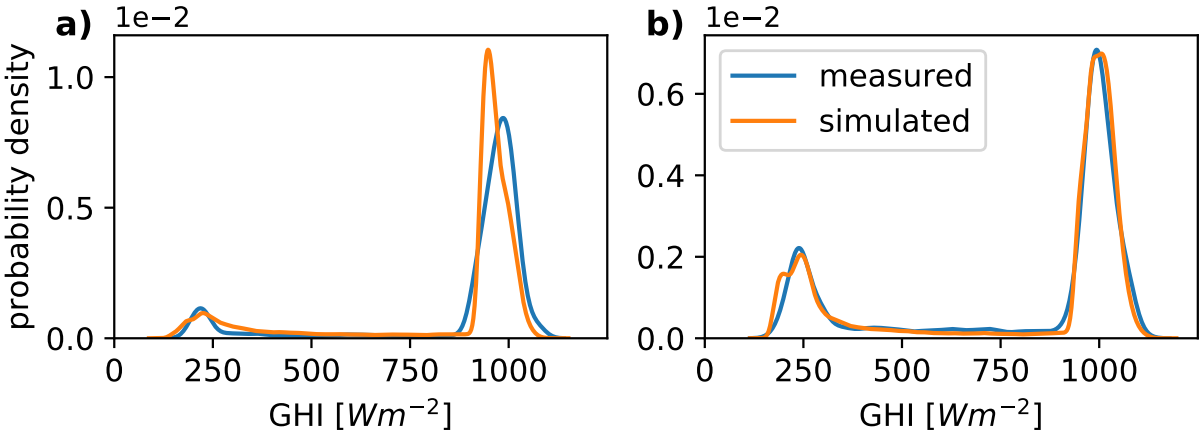

Figure 3.13.: Probability density function for measured (blue) and simulated (orange) GHI using LES cloud fields as input. Graph a) (left) shows the results for a cloud fraction of 21% and graph b) (right) for a cloud fraction of 30%. The SZA for a) and b) is 30°.

To finally compare measured GHI with the one from MYSTIC simulations the Probability Density Function (PDF) for both cases was inferred by means of Kernel Density Estimation (KDE) with a Gaussian kernel. KDE is a non-parametric method for approximating the actual PDF of a distribution from a finite number of sampling points by smoothing the histogram with a kernel of a certain bandwidth (Ghosh, 2017). For the binned measurements with ≈1000 data points a bandwidth of 20 Wm^{-2} was chosen and for the simulations with in total 64,000 data points (pixel) a smaller bandwidth of 10 Wm^{-2} was used. Figure 3.13 shows the resulting PDFs for cloud fractions of 21% and 30% at a SZA of 30°. In Graph 3.13a) (21% cloud fraction) both PDFs show a distribution with a small peak of purely diffuse radiation at around 200 Wm^{-2} and a larger peak with direct radiation slightly below 1000 Wm^{-2}. The peak at large GHI values from the measurements is slightly shifted to higher GHI values which is caused by a higher density of points at the upper limit of the cloud fraction bin. The PDFs in Figure 3.13b) for a cloud fraction of 30% show even better agreement between measurement and simulation. In this case both peaks match in their location and magnitude, which is due to a more even distribution of values within the chosen SZA and cloud fraction bins of the measurements. The discrepancy in the peak of diffuse radiation is probably a statistical effect of the relatively small size of the measurement dataset.

Based on the over all good agreement of the PDFs of measurements and simulations, it can be stated that the LES cloud fields produce realistic patterns of GHI when used as input for 3D radiative transfer simulations and are therefore well suited for further sensitivity studies on GHI variability.

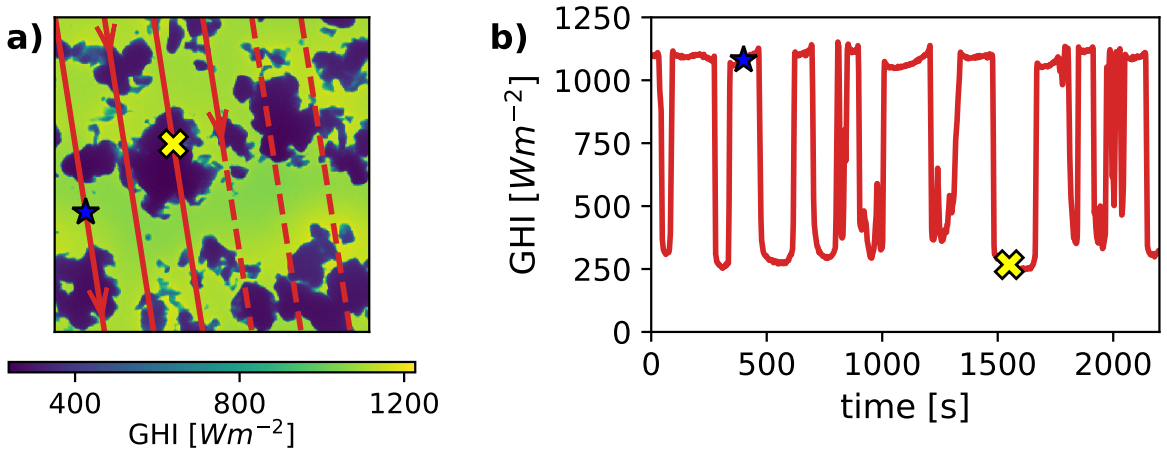

Figure 3.14.: a): Spatial field of GHI derived from 3D radiative transfer simulations with MYSTIC using LES clouds with a cloud fraction of 43% (see also fig. 3.12). The red line shows the path through the domain starting in the upper left corner with periodic boundary conditions. The direction is indicated by the red arrows. The time series derived from the path of the solid red line in a) assuming a cloud speed of 10 ms^{-1}depicted in b) (right). The red dashed line indicated how the path would continue.

3.6.3. Deriving GHI time series through spatial sampling

The goal of this section is to gain deeper insights into the properties and dependencies of the temporal GHI variability. However, the presented MYSTIC simulations only provide spatially resolved GHI for static cloud fields. So how can these data help to understand temporal GHI variability? The key to answer this question lies in the link between spatial and temporal variability via advection. In the atmosphere clouds are moved by the horizontal background wind and generate temporal fluctuations of GHI at a fixed location on the ground. Therefore, time-series of GHI can be generated by sampling along a spatial path through the domain assuming a certain advection speed. As the LES and MYSTIC simulations have periodic boundary conditions the path length is not restricted by the domain size and continuous time series can be derived over longer timescales.

Figure 3.14 shows an example of a path with periodic boundary conditions through the MYSTIC domain (a) and the corresponding time series of GHI (b). The path indicated in a) (solid red) starts in the upper left corner heading towards the lower boundary of the domain (see arrows) and continues at the upper boundary after the lower boundary was reached. To highlight the relation between the spatial path in a) and the corresponding time series in b) the red solid line turns into a dashed one as soon as the end of the time series in Figure b) is reached. The blue star and the yellow cross indicate locations with high GHI outside cloud shadows and low GHI within the shadow respectively. The time series shown in Figure 3.14 as well as all time series

to be shown in the further course of this section were derived assuming a constant advection speed and taking sampling points at a resolution of one second. Furthermore, the GHI for each sampling point was derived by linear interpolation between the grid points and the shape of the clouds stays the same for the entire time series (advection without convection). Based on these time series the variability index V (see eq. 3.11) was calculated for hourly intervals and one-minute average GHI as described in Section 3.5.2. Additionally, it is possible to derive a measure for the variation of V by calculating its standard deviation over an ensemble of one-hour time series from the same cloud scene.

Even if the derivation of GHI time series applying the described procedure is straight forward, there are some limitations which should be mentioned at this point. One of these limitations is the assumption of cloud advection by a constant background while the shape of the clouds doesn't change. Clouds, and in particular convective clouds, in fact change their shape over time in all spatial dimensions which has an influence on the effective advection speed of the cloud. In this context the effective advection speed could be defined as the time it takes until an advected cloud travels a certain distance and shades a considered point on the surface. If the cloud is for instance shrinking during the advection process it takes longer until its shadow reaches the location on the surface and hence the effective advection speed is smaller than the background wind which might changes the magnitude of GHI variability (V). However, one could also imagine that the effect cancels when one proportion of the clouds are shrinking and the other part is expanding during advection. Therefore, it is difficult to say how large the effect might be, but it should nevertheless be kept in mind. Another limitation is due to the LES dataset which only contains shallow convective clouds with a single cloud layer. In the real atmosphere also multiple cloud layers at different altitudes can be observed. However, introducing multiple distinct cloud layers increases level of complexity which makes it hard to draw general conclusions about properties of GHI variability. Therefore, the investigated LES cloud fields provide a good basis to gain general rather than encompassing insights.

3.6.4. 1D versus 3D radiative transfer

In the first case study the differences in GHI variability between 1D and 3D radiative transfer were assessed. Therefore, radiative transfer simulations were carried out for LES cloud scenes with different cloud fractions (21%, 43%, 83%) at a fixed solar zenith angle (SZA) of 30° as well as for SZAs of 30° and 60° for the scene with 43% cloud fraction. The simulations were run at a resolution of 25 m with MYSTIC in 3D and 1D independent pixel mode (see Section 3.1.1) using 200,000 photons for diffuse and 1000 photons for direct radiation. To get a first idea of the differences between 1D and 3D radiative transfer, Figure 3.15 shows the GHI field for 3D on the left (a) and 1D on the right (b) for an LES cloud scene with 43% cloud fraction at a SZA of 30°.

Generally, the 3D simulation yields higher values of GHI in cloud free parts than the 1D simulation which is due to horizontal transport of photons scattered on cloud

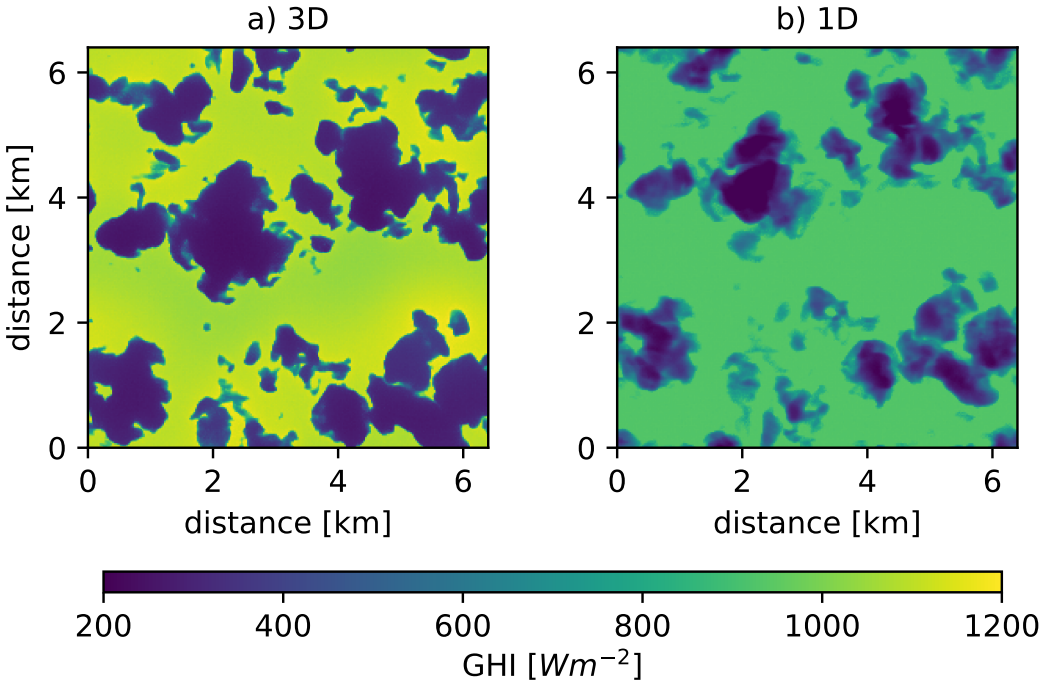

Figure 3.15.: Comparison of GHI fields with a spatial resolution of 25 m simulated with a) 3D MYSTIC and b) 1D MYSTIC independent pixel approximation (IPA). As before a LES cloud field with a cloud fraction of 43% was used and the solar zenith angle of the simulation was 30°.

sides and photons entering the clouds at the top but exiting to the sides which leads to an enhancement of GHI above clear sky levels. Another difference can be observed in the variation of GHI within the shadows of individual clouds. In 1D there are variations of several hundred Wm^{-2} between the optically thicker parts and the optically thinner parts of the cloud whereas in 3D there is only minor variations at the edges of the shadows are visible. The higher values of GHI for the thinner parts of the clouds in 1D can be explained by the restriction to vertical photon transport. In an optically thin cloud a small part of the direct radiation is transmitted according to Bouguer-Lamber-Beer law (see eq. 2.7) whereas the larger part is scattered on cloud droplets. However, most of the scattered photons keep on traveling into the forward direction because of the strong forward peak in the scattering phase function (see Section 2.2.1) and are therefore leaving the cloud at their lower boundary heading downward which leads to high levels of diffuse radiation at the surface. In contrast to 1D, 3D radiative transfer allows for horizontal photon transport which offers the opportunity that photons being scattered slightly different from the exact forward direction either exit the clouds to the sides or travel towards cloud regions with higher Cloud Optical Depth (COD) where more scattering events can occur which completely changes the path of the photon. As a consequence significantly fewer photons reach the surface within the clouds shadow and hence lead to lower GHI below cloud parts with low COD. A last remark can be made on the displacement of the cloud shadows in 3D radiative transfer compared to 1D: in

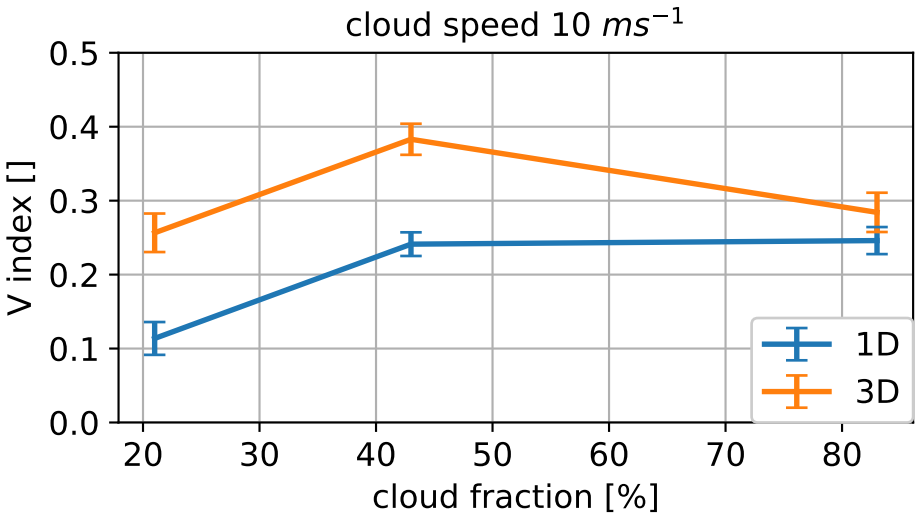

Figure 3.16.: Influence of cloud fraction on the variability index V for GHI fields simulated with MYSTIC in 3D mode (orange) and 1D mode (blue) at a solar zenith angle of 30°. The error bars indicate the standard deviation over ten hourly values of V for a fixed cloud advection speed of 10 ms^{-1}.

1D the cloud shadow is always directly below the cloud as each column is considered independently whereas in 3D the shadows are displaced horizontally depending mainly on the SZA and the altitude of the cloud base. In Figure 3.15 the sun is located in the direction of the upper boundary (positive y-direction) at an SZA of 30° displacing the shadows towards the lower boundary (negative y-direction). This property is especially relevant for high solar zenith angles and/or large cloud fractions as will be shown later on.

To show the difference in GHI variability between 1D and 3D radiative transfer, an ensemble of ten individual time series was derived from each simulated GHI field whereby the average value of V and its standard deviation were calculated over all ensemble members assuming an advection speed of 10 ms^{-1} as described in the previous section. As a result, Figure 3.16 shows the magnitude of V for three selected cloud fractions and for 1D and 3D radiative transfer with MYSTIC. For the cases of low (21%) and medium (43%) cloud fraction the 1D simulations exhibit a nearly constant absolute bias in V of -0.14. An explanation for this behaviour can be found by looking at graph a) in Figure 3.17 which depicts an exemplary time series of normalized GHI at one second resolution for a cloud fraction of 43%. To be able to compare 1D (blue line) and 3D (orange line) simulations the shadows of the 1D simulation were shifted to match the shadows of the 3D simulation which allows for deriving consistent time series. Due to the previously described enhancements of GHI in clear regions caused by horizontal photon transport the difference in (normalized) GHI between cloud shadows and clear regions is significantly larger for 3D compared to 1D. Furthermore, lower GHI values in the cloud shadows which are particularly pronounced for optically thin clouds further

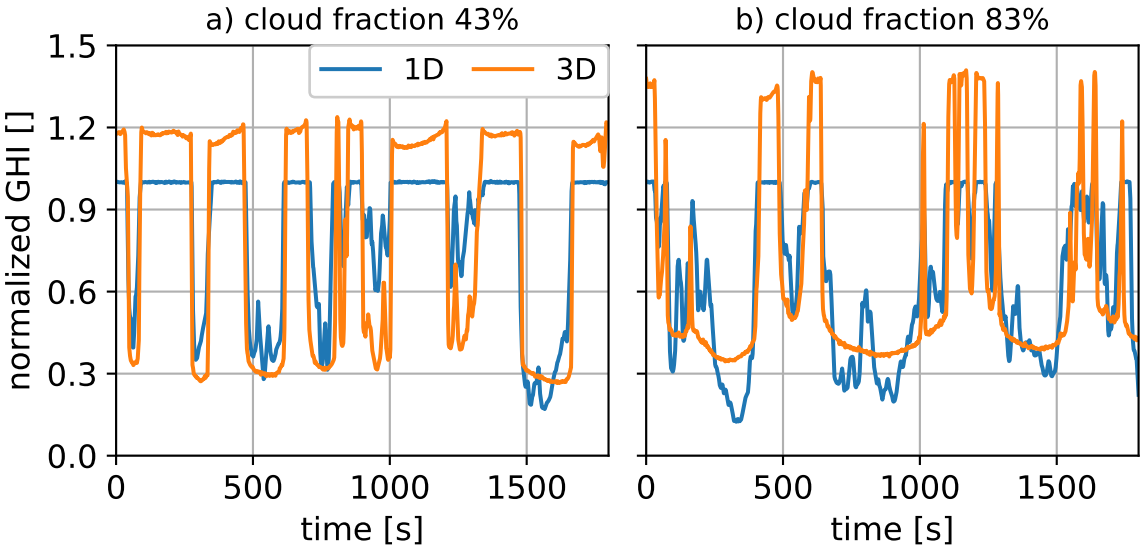

Figure 3.17.: Comparison of time series showing GHI normalized to clear sky GHI, simulated with MYSTIC in 3D mode (orange) and 1D mode (blue). a) depicts the time series for a cloud fraction of 43% and b) for 83%. For better comparability the cloud shadows from the 1D simulations were shifted to match the ones of the 3D simulations.

increase the magnitude of GHI ramps. As a result the variability V derived from 3D simulations is significantly larger than the one derived from 1D simulations considering low and medium cloud fractions.

However, the picture changes when inspecting the variability index for the high cloud fraction scenario of 83% where V is of nearly equal magnitude for 1D and 3D simulations. Compared to the medium cloud fraction (43%), V decreases for 3D by about 0.1 (26%) whereas V is nearly equal for 83% cloud fraction. Again it is helpful to explain this behaviour by looking at the time series of normalized GHI depicted in Figure 3.17b). Analogous to Figure 3.17a) the time series have a resolution of one second and the cloud shadows of the 1D simulation were shifted to match the shadows in the 3D simulation. In the times series derived from the 3D simulation (orange) the magnitude of major ramps (see $t \approx 500s$) at cloud edges is about the same as in the 43% cloud fraction case, but there are less major ramps and more short term peaks – for instance at $t = 1000s$ – caused by small wholes in the cloud deck. These short term peaks of GHI occur at one second resolution but are smoothed considerably when the time series is averaged over one-minute intervals which is done to calculate V. Therefore, their contribution to V is much smaller than the one of major ramps with GHI peaks of longer temporal extend and hence V is smaller for 83% cloud fraction compared to 43%.

As already described above, the variability for the 1D simulation does not decrease when the cloud fraction increases from medium to high levels even if the number

of major ramps is reduced as it was the case for the 3D simulation. One reason for this behaviour can be found in the magnitude of the ramps between cloud shadows and clear regions. In the 83% cloud fraction case larger values of COD in some clouds cause smaller values of GHI in the shadow regions (see $t \approx 300s$) which increases the magnitude of the GHI ramps compared to the medium cloud fraction case. Furthermore, a significant variability of GHI occurs within the shadow regions in the 1D simulation (see $t \approx 800s$), caused by horizontal variations of the COD within the cloud. Because of the larger extent of the shadow regions compared to 43% cloud fraction these fluctuations have a larger contribution to V. A last factor which prevents a reduction of V for high cloud fractions in the 1D simulation is the spatial extent of small clear regions in the cloud field. In 1D the clear and sunny area at the surface is always equal to the cloudless area at the altitude of the cloud. In 3D however, a slant incident angle of the photons (SZA $> 0°$) in combination with the vertical extent of the cloud can cause a reduction of the clear area at the surface compared to its extent at cloud level. Therefore, the temporal duration, especially of small peaks in GHI, is shorter for 3D than for 1D. Such a case can for instance be observed at the right edge of Figure 3.17b) ($t > 1500s$). The broader the peak along the time axis the smaller the effect of the temporal smoothing to a resolution of one minute. Together all three described effects compensate for the lower number of major ramps and therefore reduce the bias of 1D radiative transfer compared to 3D.

The last parameter which will be discussed in this section is the influence of SZA on the bias of V between 1D and 3D radiative transfer simulations. Therefore, additional simulations were carried out for a SZA of 60° using the LES cloud field with a cloud fraction of 43%. For the 3D simulation a small decrease in the mean value of V from 0.383 for a SZA of 30° to 0.358 for 60° SZA can be observed, for the 1D simulations V increases with SZA from 0.241 (30°) to 0.276 (60°). To identify the causes for the observed changes in V, Figure 3.18 shows again a time series of normalized GHI for 1D and 3D radiative transfer with a SZA of 30° (a) and 60° (b).

The main reason for the decrease of V for 3D lies in the reduction of the difference in normalized GHI between cloud shadows and clear regions. For a SZA of 30° the normalized GHI oscillates roughly between 0.3 and 1.2 whereas for 60° the values mostly range from 0.4 to 1.2. For the 1D case the opposite behaviour can be observed: for an SZA of 30° the minima of normalized GHI in the cloud shadow are reaching 0.4 most of the time but only rarely drop below 0.3 which happens more regularly for 60° SZA. This can be explained by the increase in effective COD due to the higher SZA which leads to lower transmission of radiation through the cloud and hence lower values of GHI. It can also be seen from the time series that the extent of the cloud shadows for higher SZA becomes larger in 3D compared to 1D which has no significant effect for medium cloud fractions but becomes important for high cloud fractions. In the high cloud fraction case the increase of cloud shadows due to geometrical effects further reduces the occurrence of sunny patches on the surface and hence reduces variability which is not the case for 1D.

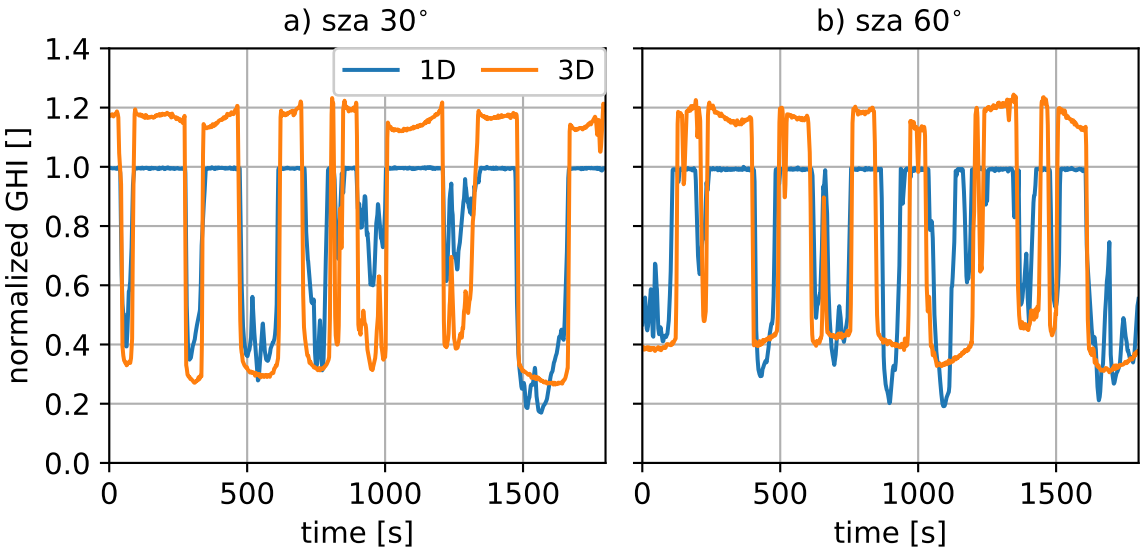

Figure 3.18.: Comparison of time series showing GHI normalized to clear sky GHI, simulated with MYSTIC in 3D mode (orange) and 1D mode (blue). a) depicts the time series for a solar zenith angle (SZA) of 30° and b) for an SZA of 60°. For better comparability the cloud shadows from the 1D simulations were shifted to match the ones of the 3D simulations.

Therefore, it can be expected that V would be larger in 1D than in 3D simulations with high cloud fraction and large SZA.

In summary, it could be shown that there is a significant bias concerning the GHI variability (V) derived from 1D radiative transfer simulations compared to 3D with a negative sign for most cases. However, for cases of high cloud fraction this bias is compensated by the insufficient assumption of independent columns in 1D. For a combination of high cloud fraction and large SZA the bias can even be positive. As a consequence 1D simulations, even though they have high spatial resolution, are inappropriate for studying the properties of spatial as well as temporal GHI variability.

3.6.5. Influence of resolution

Besides the bias introduced by using 1D solvers, the spatial resolution of the radiative transfer simulations plays a crucial role for the representation of GHI variability. Therefore, the effect of model resolution on the variability index V was investigated based on the 3D MYSTIC simulations for LES cloud fields with a resolution of 25 m. To estimate the GHI for resolutions coarser than 25 m, the spatial average GHI was computed for a grid spacing of 100, 400, 800, 1275 and 2125 m based on the high resolution simulation. Additionally, the effect of cloud fraction on GHI variability at certain resolutions was determined using the three different levels of cloud fraction already considered in the previous section (21%, 43%, 83%).

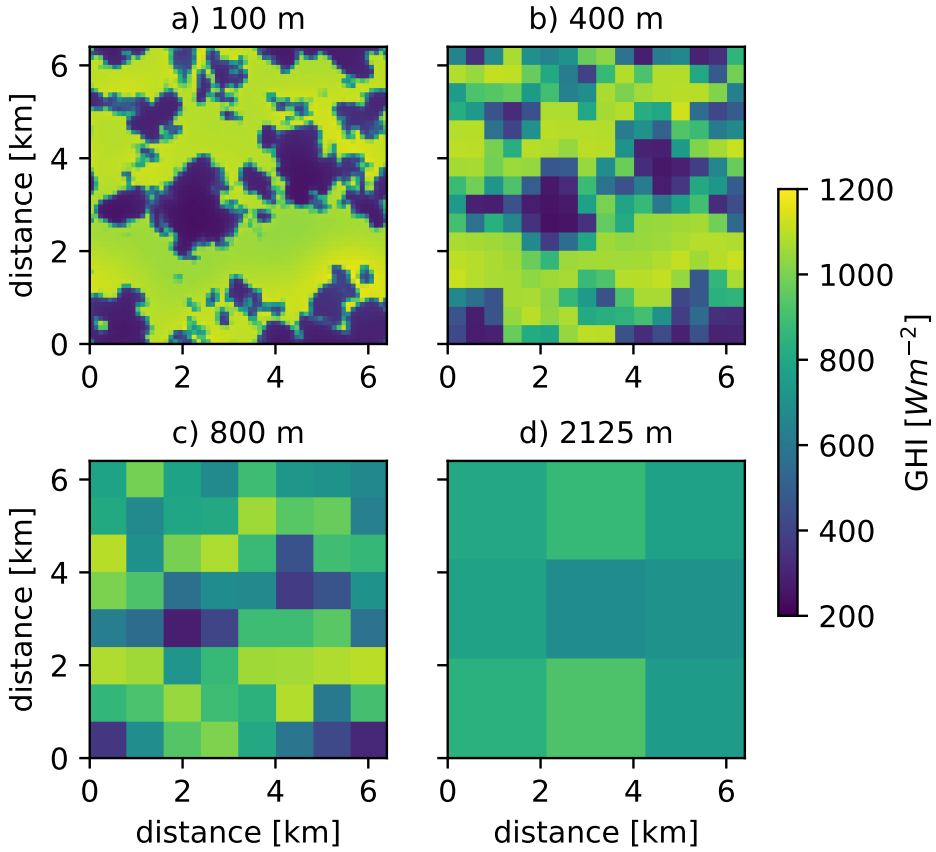

Figure 3.19.: Comparison of spatial GHI fields for the same cloud scene but varying spatial resolution of a) 100 m, b) 400 m, c) 800 m and d) 2125 m which is a typical resolution of NWP models. The original simulation was performed at 25 m resolution with 3D MYSTIC for a cloud fraction of 43% and an SZA of 30° (see Figure 3.12)

To get a first impression, the effect of spatial averaging is depicted in Figure 3.19 which shows 2D fields of GHI of four selected resolutions at a cloud fraction of 43%. For a resolution of 100 m, no significant difference to the native resolution of 25 m (see Figure 3.15a) can be observed. However, decreasing the resolution to 400 m results in smeared out edges of the cloud shadows which significantly reduces the gradient between cloudy and clear parts. Doubling the grid spacing to 800 m further decreases the gradients and the cloud shadows can only hardly be recognized as distinct objects. Finally, considering the resolution of state-of-the-art NWP models like COSMO-D2 with a grid spacing of 2.1 km the initial structure of high GHI in clear regions and low GHI in the cloud shadows is entirely smoothed away and GHI does hardly differ between individual pixels.

Based on the GHI fields with reduced horizontal resolution, time series of GHI were derived as described in Section 3.6.3. From these time series V was calculated

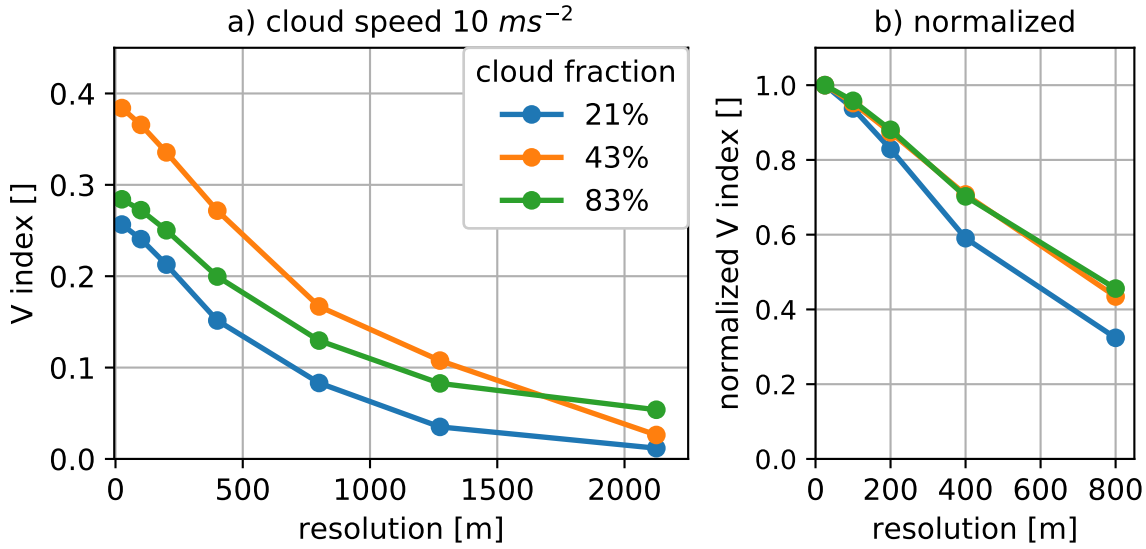

Figure 3.20.: a): Variability index V as a function of spatial resolution for three different cloud fractions at a fixed cloud advection speed of 10 ms^{-1}. b) Detail of a) with V normalized by the maximum value of the respective cloud fraction to visualized differences in the slope.

for one-minute averages within ten separate hourly intervals and subsequently averaged over all intervals as before. Figure 3.20a) shows the resulting curves of V as a function of the spatial resolution for three different cloud fractions. All curves exhibit a similar behaviour with a small slope at the very beginning (25-100 m) which is in accordance with the observations of the GHI field in Figure 3.19a). From there on the change of V with decreasing resolution becomes faster until a resolution of 800 m is reached where the slope starts to decrease again.

One interesting feature of the curve for 43% cloud fraction appears in the resolution step from 1275 m to 2125 m. Compared to the two other curves, the reduction of V is much larger and nearly reaches the lowest curve (21%) at 2125 m resolution. This behaviour could be explained by the horizontal cloud size – more precisely the size of the cloud shadow on the surface – for this specific case. Estimating the size of the cloud shadow from Figure 3.19a) reveals a value of approximately 2000 m in diameter. Therefore, it can be assumed that GHI variability is drastically reduced when the size of an individual pixel reaches or slightly exceeds the cloud shadow size in the scene. This explanation is supported by two facts: first a sensitivity study was conducted where the GHI field was shifted upward and downward before V was calculated. Thereby, the possibility of an artifact due to a special location of the cloud shadows relative to the spatial pixels could be excluded as all tests yielded the same result. A second piece of evidence is contributed by the similar behaviour of the case with a cloud fraction of 21%. Figure 3.20a) shows the same curves as in a) but normalized to the maximum of V which provides the opportunity to compare the relative rather than the absolute changes

in V. Here it can be clearly seen that the relative slope of the 21% cloud fraction case is larger than for the two other cases especially in the range between 200 and 800 m which can again be attributed to the cloud shadow size which is in fact on the order of 400 to 800 m. Finally, the case of 83% cloud fraction does not show a region of steeper slope compared to the other cases. This can be explained by the fact that the cloud shadow size in this case is already close to domain size and therefore beyond the investigated resolutions, which again supports the presented explanation.

Summarizing the findings, spatial resolution is an important factor when studying GHI variability on small temporal and spatial scales. It was shown that the variability index V quickly decreases at spatial resolutions lower than 100 m leading to a significant underestimation if the resolution is chosen too low. Furthermore, the influence of resolution on V also depends on cloud fraction as the cloud sizes in the LES data are increasing with larger cloud fraction. To keep the error of V well below 20% compared to the 3D benchmark simulation with a resolution of 25 m, resolutions higher than 200 m are required. Finally, a comparison of the effect of 1D versus 3D radiative transfer and the impact of resolution on V for the 43% cloud fraction case and a SZA of 30° shows that the difference in V between 1D and 3D (0.15) is comparable to reduction in resolution from 25 m to about 500 m (factor of 20). As a result it can be concluded that it is necessary to use 3D radiative transfer simulations in combination with spatial resolutions higher than 200 m to obtain reliable estimates of GHI variability when considering spatial scales of a single rooftop and temporal scales of one minute.

3.6.6. GHI variability - sensitivity study

After the impact of 1D and 3D radiative transfer as well as the effect of the spatial resolution on the magnitude of the variability index V was shown in the previous sections, this section investigates the dependence of V on parameters like wind speed and cloud fraction in detail. As it has been shown that high resolution and 3D radiative transfer is mandatory to capture small scale GHI variability, all simulations were conducted with MYSTIC in 3D mode. The resolution of the simulations was set to 25 m as before but due to the high computational demand necessary for investigating many parameters, subsampling was applied by setting the `mc_backward_increment` option in MYSTIC to three. Thereby, only every 3rd pixel in x- and y-direction is calculated (see Section 3.1.3) reducing the computation time by almost a factor of ten. Furthermore, the number of photons for diffuse irradiance was reduced from 200,000 to 100,000 which additionally reduces computational effort by only small losses in accuracy (see Section 3.6.1). Sensitivity studies showed that time series derived from the subsampled GHI field with reduced photon number are nearly identical to the one derived from the full simulation which can be seen in Figure 3.21. The reduced photon number leads to a slightly enhanced noise level in GHI and the subsampling causes occasional mismatches at cloud edges or in the presence of small cloud patches. The magnitude of the gradients is thereby almost unaffected because the simulated pixels are at the original high resolution and do not smooth away the extreme values of sunny

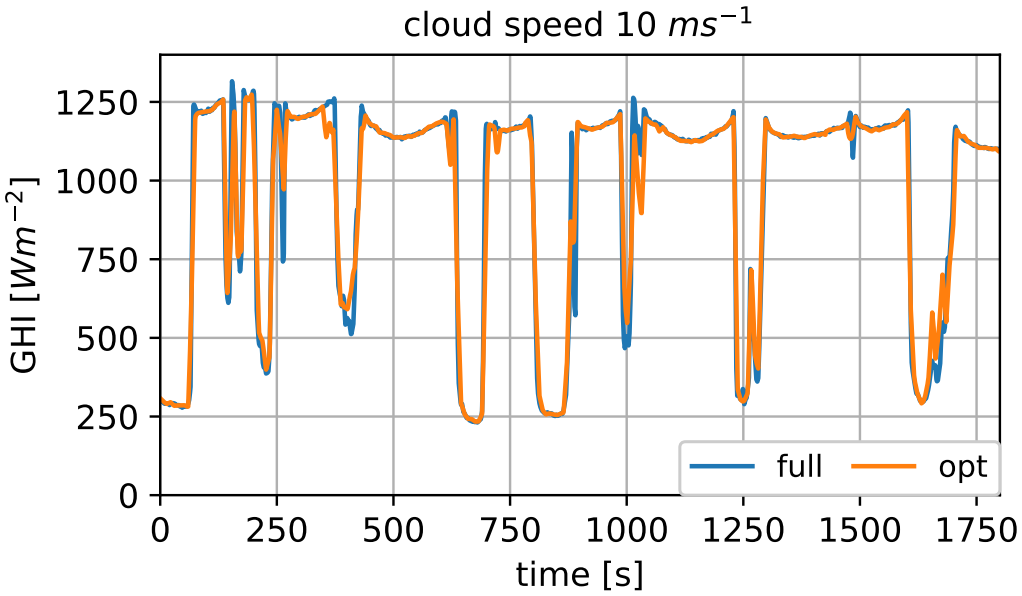

Figure 3.21.: Comparison of GHI time series derived from spatial GHI fields assuming a cloud advection speed of 10 ms^{-1}. In the full simulation (blue) GHI was simulated for all pixels within the domain using 200,000 photons for diffuse radiation. In the optimized simulation only every third pixel was sampled in x- and y-direction and the number of photons was reduced to 100,000. The solar zenith angle (SZA) for the MYSTIC simulations was set to 20° and a LES cloud scene with 30% cloud fraction was used.

and shaded regions. In summary these effects manifest in a deviation of the variability index V of less than 2% which justifies the subsampling as well as the reduction of photons.

For the sensitivity study three different parameters were investigated which are listed below:

- Solar Zenith Angle (SZA): 20°, 40°,60°
- Cloud fraction: 12%, 30%, 53%, 83%, 100%
- Cloud speed: 1, 2, 5, 10 ms^{-1}

The time series necessary for the calculation of V and its standard deviation for each case were derived as described in Section 3.6.3, whereby ten hourly intervals were used to calculate the mean and standard deviation of V. Figure 3.22a) shows the dependency of V on SZA for different cloud fractions and a fixed cloud speed of 10 ms^{-1}. The highest variability throughout all SZAs can be observed for a cloud fraction of 53%. Additionally, the standard deviation of V for this cloud fraction is the smallest of all considered cases. The high variability for this cloud fraction is due

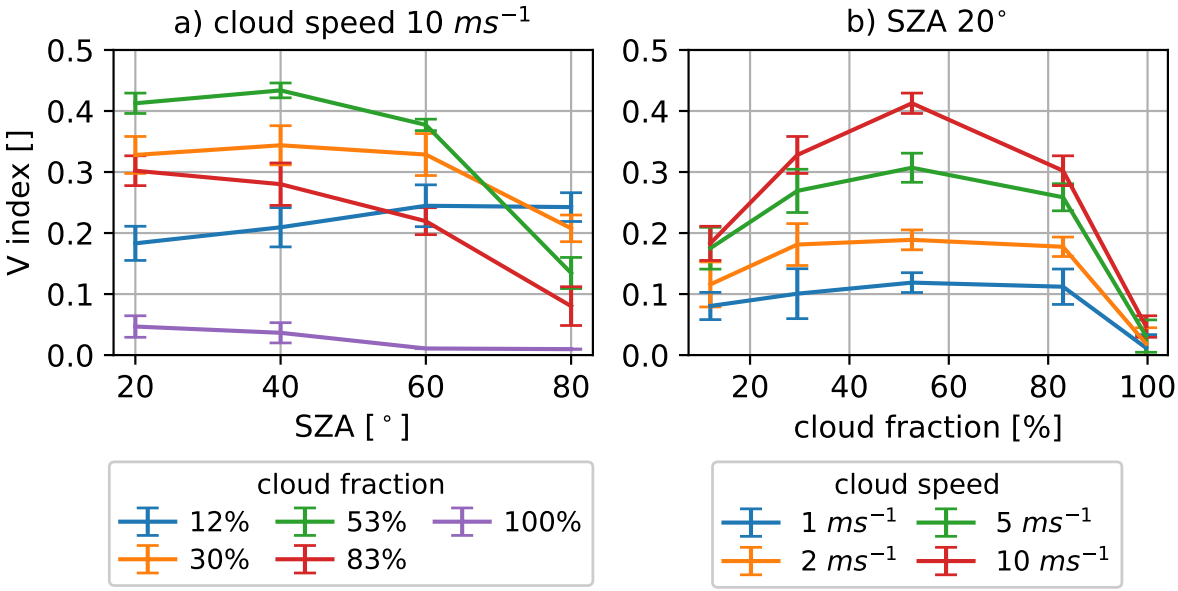

Figure 3.22.: a): Dependency of variability index V on SZA for different cloud fractions with fixed cloud speed. b): Dependency of V on cloud fraction for varying cloud speeds and a fixed SZA.

to the very high frequency of GHI fluctuations caused by alternating periods of cloud shadows and sunny areas. For lower and higher cloud fractions fluctuations are less frequent as either periods of clear skies or cloud shadows are prevailing. Additionally, the 53% case shows very large GHI ramps due to high levels of diffuse radiation in clear regions associated with strong enhancements of GHI above clear sky (>25%). The small standard deviation of V is due to equal amount of clear sky and cloud shadows which are furthermore well distributed within the domain. Therefore, all time series derived from the cloud field are similar, regardless of the location within the domain.

As expected, an overcast sky (100% cloud fraction) has the lowest variability even though it is larger than zero for low SZAs. This can be explained by two effects: first it can be stated that the cloud fraction of the 100% case is actually 99.71% which means that there are small cloud free areas in the cloud deck leading to GHI peaks at the surface. The second effect are spatial variations of the cloud optical depth (COD) which alter the transmission of radiation and hence the GHI field. Both of the described effects can be observed in Figure 3.23. It has to be noted here that the sinusoidal variation of the normalized GHI is a coincidence caused by the distribution of the optical thickness in the LES scene which has a optically thick convective core surrounded by thinner clouds which leads to this alternating pattern when sampling along a certain path through the domain. Looking now at the variability for SZAs greater than 40° in Figure 3.22a) reveals that V almost reaches zero which is mainly due to geometrical effects. The small patches of clear sky in the cloud deck are shaded due to the vertical extend of the cloud which prevents direct radiation from reaching the ground which is also illustrated by Figure

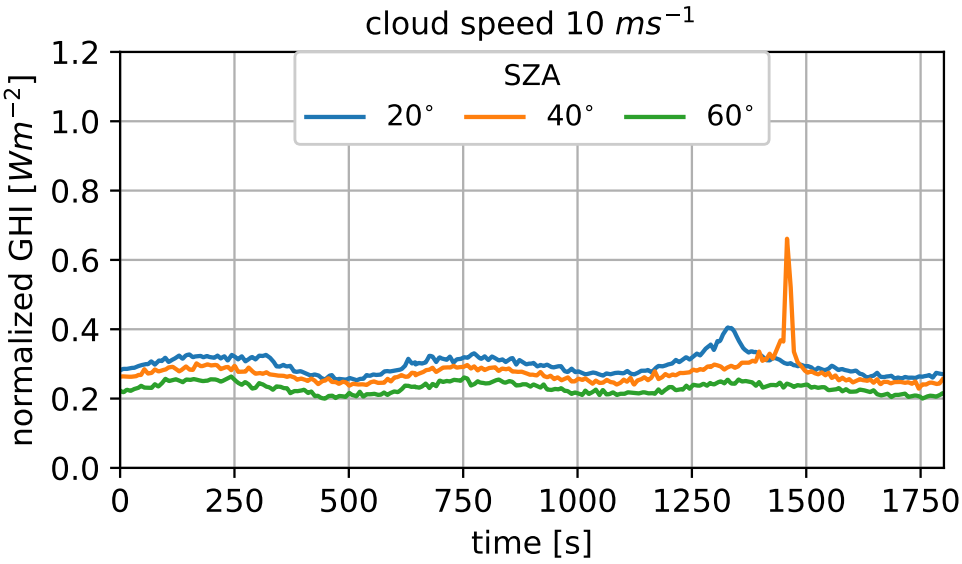

Figure 3.23.: Time series of GHI normalized by GHI under clear sky conditions for a cloud fraction of 100% (99.71%) and varying solar zenith angle (SZA). The time series were derived from spatial GHI fields simulated with 3D MYSTIC assuming a constant cloud advection speed of 10 ms^{-1}.

3.23. Additionally, the effective COD of the cloud increases due to the slant incident angle of the photons which in particular reduces the transmission of thinner parts of the cloud with low COD values. This effect further reduces patches of larger GHI and hence the variability of GHI at the surface. Generally it can be stated the variability of COD within a closed cloud field is highly dependent on cloud type which implicates that the investigated convective clouds may be not representative for more homogeneous stratiform clouds in case of cloud fractions close to 100%. However, in any overcast cloud case GHI variability can be assumed to be small compared to scenes with cloud fractions considerably lower than 100% which means that the investigated scene provides at least a rough idea of the GHI variability in overcast cloud situations.

In contrast to the curves for 53% and 30%, the variability for cases with high (83%) and low (12%) cloud fraction show a more pronounced dependence on SZA between 20° and 60°. For 12% cloud fraction V increases by 34% when increasing SZA from 20° to 60° while the opposite is true for 83% with a decrease of 21%. This behavior can on the one hand be attributed to the size of the cloud shadows which increases with increasing SZA due to the vertical extent of the clouds. In the low cloud fraction case (12%) this means that it becomes more likely to encounter a cloud shadow along path through the domain leading to an increase in the number of large GHI ramps and hence an increase in V. The increase of cloud shadow extent also leads to a longer duration of GHI minima in the time series which are therefore less affected by the temporal averaging of GHI for the calculation of V (see Section 3.6.3). In the high cloud fraction case (83%)

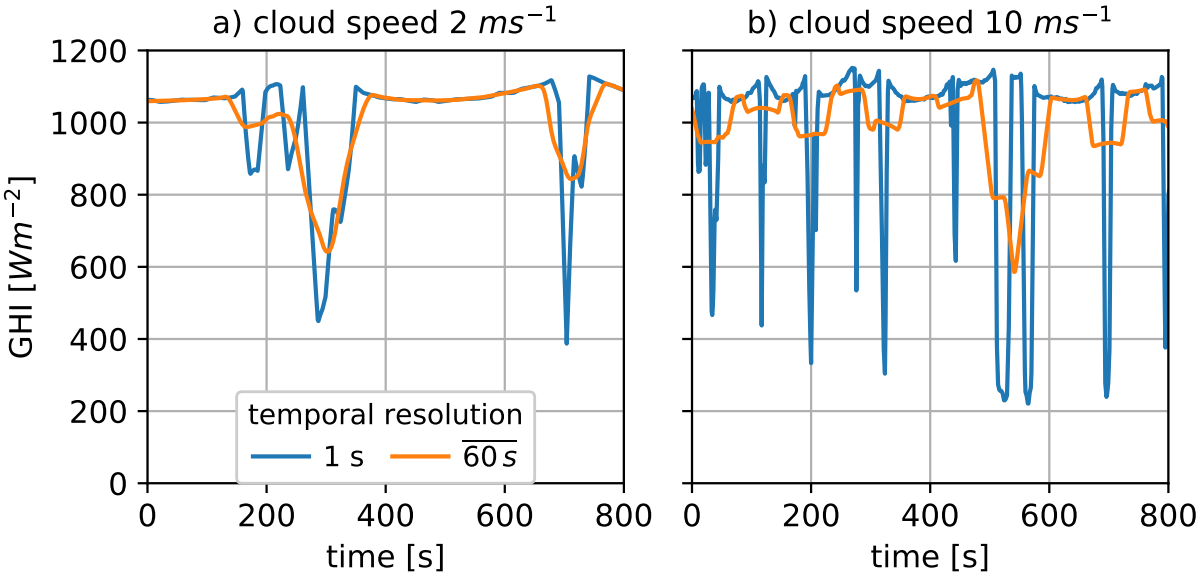

Figure 3.24.: Time series of GHI for a temporal resolution of one second (blue) compared to a 60-second average (orange). The time series were derived for a cloud advection speed of a) 2 ms^{-1} and b) 10 ms^{-1}.

however, the probability to encounter clear sky regions with high GHI reduces when the extent of cloud shadows becomes larger. As a consequence, the number of GHI ramps is reduced which leads to a decrease in V. In addition to the size of the shadows, the magnitude of GHI ramps also increases with SZA for low cloud fraction. The COD along the light path increases with larger SZA leading to a decrease in GHI within the cloud shadows relative to clear sky GHI. Based on these explanations one would expect a further increase of V for an SZA of 80° but the magnitude of GHI ramps relative to clear sky GHI decreases again which compensates for the effect of larger cloud shadows.

In the next step the sensitivity of V on cloud fraction for different cloud speeds and a fixed SZA was investigated. Figure 3.22b) shows that V increases for all cloud fractions when the cloud speed increases, although the amount of increase is highly dependent on cloud fraction. The general increase of V with cloud speed can be explained by a higher frequency of GHI ramps between cloud shadows and clear sky within the time frame of one hour.

To explain the influence of cloud fraction on the increase of V it is helpful to analyze time series of GHI for a specific cloud fraction. Therefore, Figure 3.24 shows time series of GHI for a cloud fraction of 12% with cloud speeds of 2 ms^{-1} (a) and 10 ms^{-1} (b) as well as for the original time resolution of one second and the one-minute average ($\overline{60}\,s$) which is used to calculate V.

In the case of a small cloud advection speed in Figure 3.24a) minima of GHI are relatively broad due to the long time within the cloud shadow, but only few shadows are encountered in the depicted time frame. Averaging the time series over intervals of 60 s

decreases the magnitude of GHI ramps, but remaining fluctuations are still on the order of 200 to 400 Wm^{-2}. When inspecting Figure 3.24b) with a cloud speed of 10 ms^{-1} the picture changes significantly. The number of clear to shadow transitions is about five times higher than for the lower cloud speed, however their temporal extent reduces drastically. Therefore, most of the GHI ramps are nearly smoothed away when applying the one-minute averaging. As a result V only increases by a factor of 1.6 when increasing the cloud speed from 2 ms^{-1} to 10 ms^{-1}, although the number of GHI ramps increases by a factor of five at the same time. For a comparison on can consider the case with a cloud fraction of 53%. Here V increases by a factor of 2.2 for the same change in cloud speed which is 37.5% more than for the 12% cloud fraction case. The temporal extent of minima in GHI time series is directly linked to the cloud size. In the LES dataset larger cloud fractions are associated with larger cloud sizes which cause GHI minima of longer temporal extent in the GHI time series. These are less affected by temporal averaging and hence the increase of GHI ramp frequency with cloud speed yields a larger increase in variability than for a cloud fraction of 12%. At large cloud fraction the size of clear sky patches rather than the size of the cloud is the limiting factor which explains that the increase of V with cloud speed declines again at larger cloud fractions.

In summary, it could be shown in this section that there is a complex nonlinear interplay between cloud fraction, SZA and cloud advection speed. However, the results of the sensitivity study can be explained by analyzing the effect of each parameter separately and the information to get the complete picture. Based on the information gained in this and the previous section a first model for the prediction of GHI variability will be developed in the next section.

3.7. Predicting variability - lookup-table approach

After the properties of GHI variability have been analyzed in detail, it is an obvious choice to use this information to develop a model to predict the variability index V using numerical weather prediction (NWP) data as input. Therefore, this section describes how the information from 3D radiative transfer simulations can be incorporated into a variability model and how mandatory input parameters can be extracted from COSMO-D2 forecasts. Finally, the model will be tuned to optimize its accuracy.

3.7.1. Model description

In the previous section it has been shown that V in particular depends on parameters like cloud advection speed (in the latter also just termed cloud speed), solar zenith angle (SZA) and cloud fraction. The SZA is a purely geometrical quantity and cloud fraction as well as cloud speed can be derived from the NWP data. Therefore, it is possible to develop a model based on these three parameters using NWP data as input.

The core of the model is formed by a three-dimensional lookup table (LUT) which

contains the variability index V as well as its standard deviation as a function of SZA, cloud speed and cloud fraction. For the calculation of V (see eq. 3.11) and its standard deviation (V_{std}) the GHI fields simulated for Section 3.6.6 were used. Those were calculated with 100,000 photons for diffuse and 1000 photons for direct irradiance per pixel and subsampling was applied setting `mc_backward_increment` to three in x- and y-direction (see Table 3.4). Based on the GHI fields V was calculated for ten distinct time series of one hour (see Section 3.6.3). As entries for the LUT the mean value of V and V_{std} were calculated from the ensemble of hourly V values. This procedure was applied to

LUT sampling points	
Parameter	Value
solar zenith angle (SZA)	$20°, 40°, 60°, 80°$
cloud advection speed	0, 1, 2, 5, 10, 15, 20, 30, 50 $[ms^{-1}]$
cloud fraction	0, 12.01, 29.51, 52.59, 82.99, 100 (99.71) [%]

Table 3.5.: Sampling points of the Lookup Table (LUT) model.

all combinations of cloud fraction cloud speed and SZA for the according sampling points which are listed in Table 3.5. For the cloud fraction dimension the clear sky (0% cloud fraction) represents the lower boundary whereby V is zero per definition (see eq. 3.11) and hence V_{std} is zero as well. The upper boundary is provided by the LES cloud field with a cloud fraction of 99.71% which was set to 100% in the LUT to cover the entire range of possible values being aware of the issue of representativeness of this cloud case which was already discussed in the previous section. At this point it should be mentioned that the LUT model is designed to predict V for time intervals of one hour which is due to the temporal resolution of the COSMO-D2 forecasts which serve as input for the model.

An advantage of this LUT model lies in its simple and fast application. After cloud fraction and cloud advection speed have been derived from the NWP and the SZA was determined using an astronomical program or database, the values for the variability index V and its standard deviation can be retrieved from the LUT. In case the input values are in between the sampling points shown in Table 3.5 V and V_{std} are determined applying multilinear interpolation using the `interpn` function of the Python library SciPy (Virtanen et al., 2020) However, there are also limitations to this model approach. A general limitation is the small number of input parameters (3) which may not be able to cover the entire complexity of the problem as other parameters might also play a role. Furthermore, the LES cloud fields may not be representative in each cloud situation which may occurs in the real atmosphere even though it has been shown that they produce GHI patterns found in measurement data (see Section 3.6.2). In particular the overcast cloud scene (cloud fraction 99.71%) has limited representativeness because of its convective origin which was already discussed in Section 3.6.6.

Finally, there are two general issues concerning the derivation of input parameters from NWP forecasts which apply also to other approaches of GHI variability prediction and hence are no limitation of the LUT model itself. The first one deals with the derivation of the cloud advection speed which at first glance could simply be determined using the horizontal wind speed at cloud level. However, clouds usually extend over multiple layers and also have different cloud fractions at different layers which makes it necessary to determine an average steering wind speed for the cloud. Another effect which may be even more important is that clouds not necessarily move at the same speed as the background wind. As already discussed in Section 3.6.3 convective clouds change their shape during the advection which causes a positive (expanding) or negative (shrinking) offset relative to the background wind. Furthermore, depending on surface and terrain properties convective clouds can form over specific locations like flanks of hills tilted towards the sun or dark forest areas which heat up faster than the surroundings. In such cases cumulus clouds are hardly advected as they quickly evaporate on the downstream side.

The second issue concerns the cloud fraction parameter. Optically thin cirrus clouds at high levels can increase the total cloud fraction significantly by having only a very small effect on GHI variability as most of the radiation entering the cirrus cloud is directly transmitted or scattered towards the ground. One can imagine situations with a shallow cumulus cloud field at low levels combined with cirrus clouds at high levels. If the cumulus clouds had a low cloud fraction of for instance 10% which causes moderate variability the additional cloud fraction of the cirrus clouds would increase the total cloud fraction leading to the prediction of significantly higher variability by the LUT. Therefore, the total cloud fraction provided by the NWP forecast might not be the optimal parameter to be used for the prediction of V.

Both of the describes issues are addressed in the following sections where solutions are developed to reduce their influence on model predictions.

3.7.2. Deriving input parameters from COSMO-D2

As previously discussed it is necessary to extract cloud fraction and cloud speed from NWP forecasts to predict GHI variability with the LUT model. In the following it will be described how these parameters were derived from the COSMO-D2 forecasts characterized in Section 3.2.2. For the model developed in this section the quality filtered dataset (see Section 3.4.6) with the training and validation dataset selected in Section 3.5.3 was used.

Cloud fraction

In COSMO-D2 the total cloud fraction of an individual grid box is provided by the parameter CLC_T which is determined by applying the maximum random overlap assumption (Geleyn and Hollingsworth, 1979) to the cloud fraction of the individual vertical layers of a grid box (clc). However, as explained in the previous section, the total cloud fraction

is not the optimal parameter for estimating GHI variability due to the influence of high level cirrus clouds. To remove the negative effect of high clouds, the cloud fraction for an individual model column was calculated up to four different altitude thresholds (4,5,6,7 km) by applying the maximum random overlap assumption to the layerwise cloud fraction (clc). The threshold providing the best result for the LUT prediction will be determined in the next section. After the cloud fraction for a single column had been calculated, a spatial average cloud fraction over 5x5 grid boxes around the measurement location was computed. As COSMO-D2 has a resolution of 2.2 km the average was performed over an area of 11x11 km^2. This spatial average was calculated for two reasons. First, the LUT model predicts the variability index for time intervals of one hour which means that during this time period clouds may be advected from the surroundings. Therefore, it is necessary to take the neighboring gridboxes into account. A second and maybe even more important argument for the spatial averaging is based on the limited representativeness of a single model column for the actual cloud situation. In cases with shallow convection and low cloud fractions the considered grid box might be entirely cloud free as the shallow convection parametrization of the NWP was not activated. Hence, the prediction of V according to this single grid box would yield zero even if there are actually clouds causing GHI variability to be larger than zero.

Cloud advection speed

To estimate the advection speed of the clouds – also termed cloud speed in this thesis – the absolute horizontal wind speed for each layer of a COSMO-D2 model column was calculated from the u- and v-components of the wind vector in a first step. As the cloud fraction was calculated for 5x5 grid boxes around a considered location the horizontal wind was derived for the same area. For each of the 25 columns which contains at least one cloudy layer, an average cloud speed was calculated by weighting the horizontal wind speed at each layer with the cloud fraction and the vertical extent of the cloud. These weighting parameters were introduced as layers with higher cloud fraction have larger impact on variability and cloudy layers with larger vertical extent have larger contribution to the cloud movement. In a final step the cloud speed was averaged over all cloudy columns to obtain one representative value for cloud speed at the considered measurement location. To calculate the cloud speed consistent with the cloud fraction, the respective altitude threshold was also applied in the derivation of cloud speed. The described procedure can also be written in form of an equation:

$$\mathrm{v}_{\mathrm{cloud}} = \frac{1}{N_{\mathrm{cldcol}}} \sum_{i=1}^{N_{\mathrm{lyr}}} \frac{\mathrm{v}_i \cdot cf_i \cdot \Delta z_i}{cf_i \cdot \Delta z_i}. \tag{3.12}$$

Here N_{cldcol} denotes the number of cloudy columns, v_i, cf_i and δz_i represent the horizontal wind speed, cloud fraction and vertical thickness of model layer i respectively. N_{lyr} is the number of model layers which depends on the selected altitude threshold.

As described in the previous section the influence of convection can alter the cloud advection speed which is not taken into account by Equation 3.7.2. This issue will be

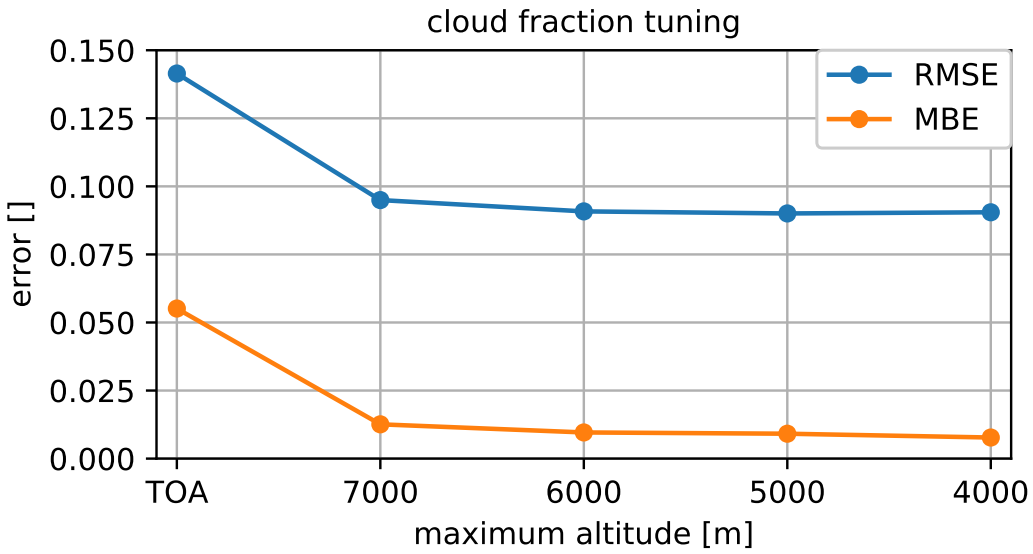

Figure 3.25.: RMSE (blue) and MBE (orange) of the LUT model applied to the training dataset as a function of the altitude threshold for the calculation of the total cloud fraction from the layerwise cloud fraction of COSMO-D2 (clc).

addressed in the next section which aims to improve the predictions of the LUT by tuning it based on available measurements.

3.7.3. LUT model tuning

In the previous section two phenomena were described which provide the opportunity to considerably improve the LUT model. The first one is the unfavorable effect of cirrus clouds on the prediction of the GHI variability. To exclude the contribution of cirrus clouds the cloud fraction as well as the cloud advection speed were calculated up to different altitude thresholds (see Section 3.7.2) whereof the best one has to be determined in this section. Therefore, the LUT model was applied to the training dataset selected in Section 3.5.3 using the cloud fraction and cloud speed for different altitude thresholds as input. Figure 3.25 shows the resulting curves of the root mean square error (RMSE) and mean bias error (MBE) as a function of the altitude threshold. TOA (top of atmosphere) represents the combination of cloud fraction and speed up to the upper boundary of the COSMO-D2 model domain of 22,000 m which means that in this case the cloud fraction is given by the model parameter CLC_T averaged over 5x5 grid boxes. The RMSE and MBE for the TOA case are taken as a reference as it represents cloud speed and cloud fraction derived over the entire vertical extent of COSMO-D2 domain including clouds at all altitudes.

Both, RMSE and MBE show a considerable reduction between TOA and the first altitude threshold of 7000 m of 33% and 77% respectively. Afterwards MBE keeps on

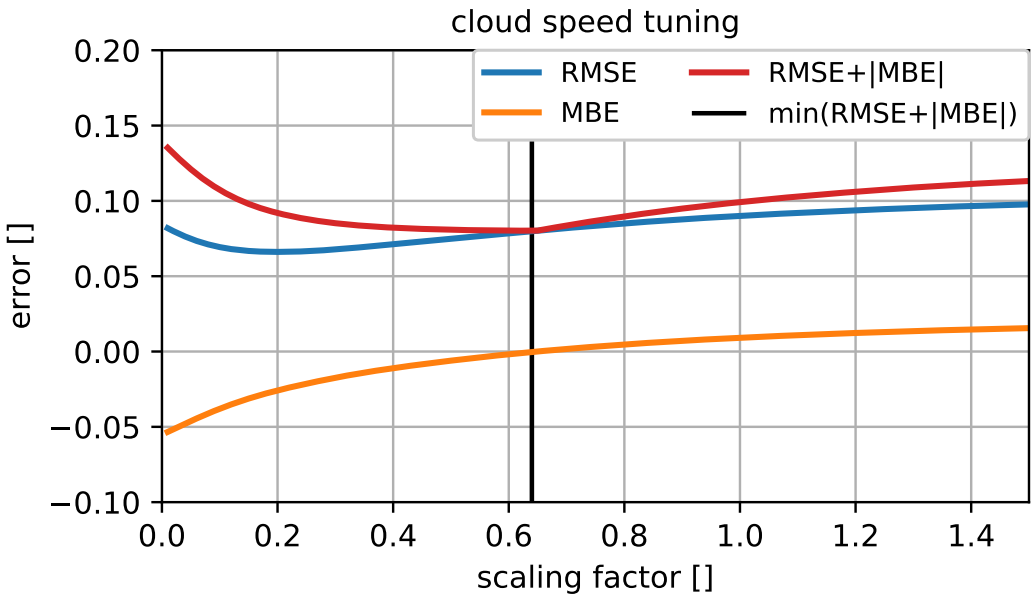

Figure 3.26.: RMSE (blue) and MBE (orange) as a function of the scaling factor for cloud advection speed. The vertical black line indicates the scaling factor which minimizes the sum of RMSE and the absolute value of MBE (red).

decreasing with a smaller slope down to the lowest threshold of 4000 m whereas the RMSE reaches its minimum at 5000 m and then increases slightly towards 4000 m. The large reduction in both error measures indicates that the cirrus clouds actually have a negative effect on the prediction of V based on the LUT model which can be attributed to their contribution to cloud fraction and their impact on the cloud advection speed as the wind speed at high altitudes is generally higher than at lower altitudes. From these results the altitude threshold of 5000 m was selected due to the largest reduction in RMSE of 36% (0.1415 to 0.0901) compared to the TOA reference case. The MBE for this threshold is reduced from 0.0551 to 0.0091 which corresponds to a relative reduction of 83%.

The second phenomenon which introduces error into the prediction of V is the effect of convection on the cloud advection speed. As described in Section 3.7.1 convective clouds can shrink or expand during advection which could lead to an increase or decrease in GHI variability depending on the cloud situation. They might also initiate over certain favorable locations like slopes or regions of low albedo and stay rather stationary there. The Allgäu region investigated in this study is located right in front of the Alps and is characterized by a hilly topography as well as frequent changes between forests and grasslands. Studying time-lapse videos of an All-Sky Imager operated during the measurement campaigns showed that convective clouds are often initiated at the same locations and are hardly advected by the background wind but slowly dissolve on the downstream side. To account for the discussed effects of convection in the LUT model a scaling factor

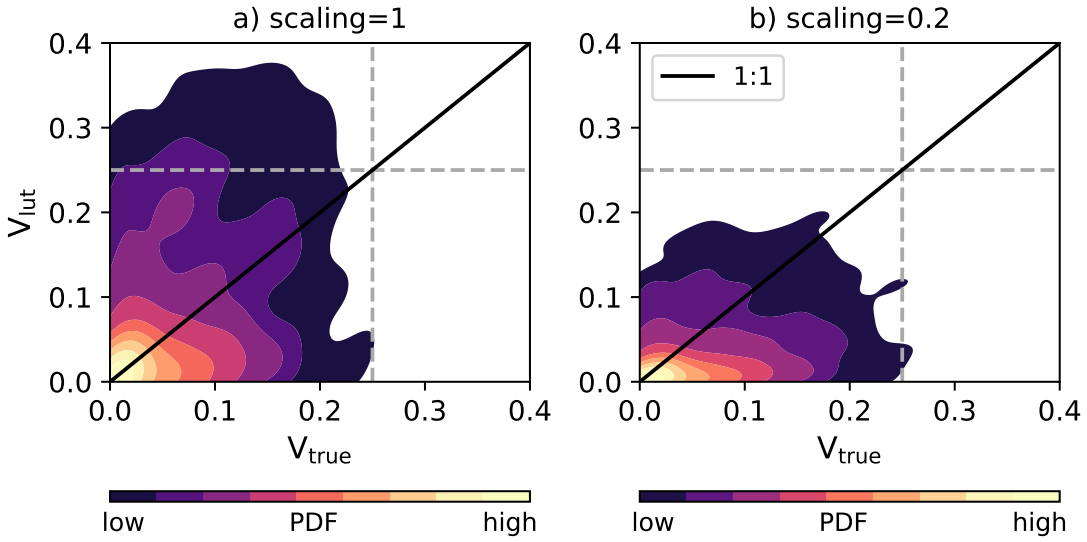

Figure 3.27.: 2D probability density function (PDF) of the predicted variability index V from the LUT model on the y-axis and the true variability on the x-axis. a) depicts the case for a cloud speed scaling factor of 1 and b) the for a scaling factor of 0.2. The color scale changes logarithmically with the probability density which was calculated using 2D kernel density estimation (KDE).

was introduced which is multiplied with the estimated cloud advection speed derived from the COSMO-D2 data (eq. 3.7.2). The optimal scaling factor was determined by applying the LUT model to the training data with the cloud speed scaled by values between 0.0 and 1.5 in steps of 0.01 and finally selecting the factor which minimizes the error.

Figure 3.26 shows the resulting RMSE (blue) and MBE (orange) as a function of the scaling factor. Here it was not possible to only use the RMSE as criterion to decide on the optimal scaling factor. Considering RMSE alone would suggest a scaling factor of less than 0.2 meaning that the cloud speed derived from COSMO-D2 would be reduced by more than 80%. However, the small scaling factor leads to a strong negative bias of the prediction. The reason for this behavior can be explained with the help of Figure 3.27 which depicts the 2D PDF of V predicted by the LUT model using the training data as input plotted against the true value of V for scaling factors of 1.0 (a) and 0.2 (b). Bright colors indicate high point density whereas darker colors are associated with lower point densities. Additionally, it should be noted that the color scale changes logarithmically with point density.

For a scaling factor of 1.0, meaning that the initially derived cloud speed is used, the LUT model tends to strongly overestimate V for a rather small number of cases which however have a large contribution to RMSE as deviations are weighted with the square of the distance. Applying a very small scaling factor of 0.2 as shown in b) reduces the overestimation as expected but pushes the other data points, which had already been closer to the 1:1 line, close to zero. Consequently, the RMSE is considerably reduced on

the expense of a large negative bias as many predictions essentially yield zero. A model which has low RMSE because it predicts zero most of the time with a large negative bias is hardly valuable from a scientific point of view. Therefore, it was decided to use not only the RMSE but the sum of RMSE and the absolute value of the MBE as criterion to determine the best scaling factor. The sum of both error measures is shown by the red curve in Figure 3.26 which has its minimum at a scaling factor of 0.64. Choosing this scaling factor result in a reduction of the RMSE reduces to 0.0799 (-11%) and the MBE turns out to be almost zero (-0.0003) on the training dataset.

In summary the RMSE and MBE of the LUT model prediction of GHI variability index V on the training dataset was reduced significantly by applying the physically motivated optimizations to cloud fraction and cloud advection speed derived from COSMO-D2 forecasts. Furthermore, it should be mentioned that the optimizations and in particular the scaling of the cloud advection speed are to some extent specific to the location of the measurements and may not yield equally good improvements at other locations or for different seasons. Nevertheless, the tuned LUT model serves as a solid reference for the second model which will be described in the next section.

3.8. Predicting GHI variability using a random forest model

As a second approach for the prediction of GHI variability a method based on a Random Forest (RF) model (Breiman, 2001) was developed. RF models belong to the class of supervised machine learning techniques and are widely used due to their simple application and robust results. More details on the RF technique can be found in Section 2.5. The RF model was chosen because no a priori knowledge of the best input variables is necessary. Furthermore, it allows to extract the most important variables from the input parameters by calculating their so-called Feature Importance (FI). The idea behind this approach and the methodology are described in the following. For the model development presented in this section the quality filtered COSMO-D2 dataset was used (see Section 3.4.6). This dataset is split into a training (85%) and validation (15%) dataset as described in Section 3.5.3.

3.8.1. Concept

For the LUT based approach presented in Section 3.7 the idea was to make use of the most important physical relations between GHI variability and atmospheric conditions and apply these relations to COSMO-D2 forecasts to predict variability. With the RF model the problem is tackled the opposite way as the COSMO-D2 variables which are most predictive for GHI variability can be extracted using the model itself. This allows not only to build a good model for predicting GHI variability but also to gain more insights about the correlation of COSMO-D2 variables and GHI variability. Before more details on the structure of the RF model will be provided a comment should be made

on the nomenclature used in the following. The term "feature" is only loosely defined in the context of machine learning as sometimes all variables used as input for the model are termed this way and sometimes only the variables which turn out to actually have prediction skill are called features. In this study all variables are also called features and the terms are used interchangeably.

As a basis for the model development serves the RandomForestRegressor of the Python library scikit-learn (Pedregosa et al., 2011). It is an efficient and easy to use implementation containing a variety of tunable hyperparameters and further useful options which makes it well suited for this study. In the first step of the model development additional features are generated from the original COSMO-D2 variables. This process is called feature engineering and aims to increase the information content of the dataset by deriving new features from existing variables of the dataset. For instance spatial or temporal averages of variables can be computed, or several variables can be combined in a certain way to obtain a new feature. Some examples for engineered features are spatial averages and standard deviations of cloud fraction and radiation parameters as well as the derivation of absolute horizontal wind speed from the u (zonal) and v (meridional) components of the wind vector. Further details on the derived features are given in the next section.

After the final dataset was created, a recursive feature elimination (RFE) procedure is applied in combination with permutation feature importance (PFI) to obtain an importance ranking of all available features. Based on this ranking the error convergence of the RF predictions can be investigated as a function of number of features in the dataset. Therefore, the prediction error of the RF is evaluated starting with a dataset containing only the most important feature and successively adding one feature after the other according to the FI ranking. This procedure allows to estimate the error reduction caused by adding a certain feature and hence gives the opportunity to determine an optimal set of features by searching a threshold according to error convergence. As highly correlated features in dataset do not provide additional information, they are identified by computing the correlation matrix of the remaining features and removed by applying a correlation threshold. Limiting the number of features by only using the most predictive ones has several advantages: first, an RF model with fewer features consumes less time for training and optimization and is therefore easier to apply. Second, using fewer features reduces the overfitting of the algorithm with respect to the training data which leads to better generalization on new data. And third, if there are only features left which actual have a beneficial impact on the prediction, physical connections between features and their role for GHI variability can be derived.

After the optimal set of features was determined the final model setup is obtained by tuning the hyperparameters of the RF. This is done by defining sampling points for selected hyperparameters and testing all combinations to find the combination providing minimal error. This procedure is called grid-search and is also available in the scikit-learn library. In the end the final model is validated by applying it to the validation dataset

which was not considered during the optimization process to check if the model is able to perform well on unseen data. The next sections will give a more detailed description of the steps described above, starting with the derivation of new features from the COSMO-D2 dataset.

3.8.2. Features and feature engineering

This section provides an overview over potentially important features from the COSMO-D2 output and new features which were derived from the original COSMO-D2 variables.

In Section 3.6.6 it has been shown that cloud fraction, cloud advection speed and solar zenith angle (SZA) are important parameters to predict GHI variability. Therefore, the SZA and several cloud parameters available in the COSMO-D2 output were used as input for the RF. First, the cloud fraction of low (CLCL), medium (CLCM) and high level clouds (CLCH) as well as the total cloud fraction (CLCT) were extracted for the grid point closest to each of the selected measurement locations (see fig. 3.1). Unlike in the LUT model all possible features which might have some prediction skill can be added to the dataset as the RF model is able to extract the best features. In addition to cloud fraction also the liquid water path (LWP) and the ice water path (IWP) were derived from the layerwise liquid water content (LWC) and ice water content (IWC) by integrating over the vertical extent of the column. LWP and IWP are related to the cloud optical depth (COD) and therefore contain information about the optical properties of the clouds. However, LWC and IWC only represent water and ice concentration for grid-scale clouds (layers with cloud fraction of 100%) which means that they only contain information about the optical properties of grid-scale clouds. To get the full information about the water content of clouds LWC and IWC for subgrid-scale clouds had to be re-calculated from COSMO-D2 standard output variables. Therefore, a Python version of the original COSMO-D2 subgrid cloud parametrization of COSMO-D2 (provided by Leonhard Scheck) was used as subgrid-scale cloud parameters are only computed internally by COSMO-D2 during runtime. In addition to the subgrid-scale LWC and IWC also the cloud fraction of convective subgrid-scale clouds was of interest as it provides indication for enhanced GHI variability caused by shallow cumulus clouds. As the subgrid convective clouds can extend over several layers with different cloud fractions, the maximum cloud fraction in a considered column was used as a new variable for the dataset. Furthermore, the average vertical thickness and the fraction of grid points containing subgrid-scale convection was computed in a region of 5x5 grid points around the location of measurement and added to the dataset.

For nearly all cloud related variables described above, the spatial average as well as the standard deviation over 5x5 grid points were calculated additional to their value for the grid point closest to the measurement site. This spatial averaging gives a more representative value as a single model column might have a very different cloud fraction or optical thickness than the surroundings and hence does not reflect the overall situation (see also Section 3.7.2). Choosing an extent of 5 grid boxes also avoids

strong overlap between the domains of the individual measurement locations which are at least 10 km apart from each other. Besides the cloud related variables, also the diffuse (DIF) and direct horizontal irradiance (DIR) as well as their spatial average and standard deviation (5x5 grid points) were used as input for the RF. As described in Section 3.2.2 DIF and DIR had to be computed from the original COSMO-D2 parameters ASWDIR_S and ASDIFD_S. Even though it had been shown previously that 1D radiation schemes and very low resolution like in COSMO-D2 considerably underestimate small scale GHI variability (see Sections 3.6.4 and 3.6.5) DIF and DIR might still have some prediction skill the RF is able to discover complex non-linear relationships.

As mentioned previously, the horizontal wind speed is important for GHI variability as it determines the cloud advection speed. But also beyond this direct relation wind speed and wind direction can be an indicator for the large scale weather situation which can be associated with certain magnitudes of GHI variability. Therefore, the absolute wind speed at COSMO-D2 pressure levels was calculated from u (zonal wind) and v (meridional wind) components of the wind vector which gives an indication of the cloud advection speed. Additionally, the u and v components for each pressure level were used as individual features. To obtain the actual u and v components for an ordinary geographic coordinate system the wind vector had to be rotated (see Section 3.1.4 in Baldauf et al. (2018)) because COSMO-D2 operates on a rotated pole coordinate system. Two other variables which are associated with cloud formation are the convective inhibition (CIN) and convective available potential energy (CAPE) which give an indication of the chance for the occurrence of convective clouds and the strength of convection once it occurs. The set of wind and cloud related variables was completed by the cloud advection speed and cloud fraction computed for different altitude thresholds from the LUT model (see Section 3.7.2) were included in the dataset for the RF model.

Beyond the previously described variables which obviously can be connected to GHI variability, other potentially relevant variables were considered. For instance the relative humidity (RH) and the temperature (T) two meter above the surface and for all COMSO-D2 pressure levels were used as additional parameters. RH is also a cloud related variable as the formation of clouds requires a saturated atmosphere with an RH of 100%. Together with the temperature profile RH provides more information about the vertical structure of the atmosphere and the potential for cloud formation. Finally, a bunch of variables which are mainly related to the course of the sun were added to the dataset: solar azimuth angle (AZI), time of the day and hours since sunrise. The intention of using these variables was that they may contain information about general patterns of GHI variability which are related to the duration of solar insulation. An example for such a relation would be that convective clouds in a high pressure regime with substantial CIN often start to form later in the day when the temperature near the surface is large enough to produce convection which overcomes the CIN. Therefore, parameters like hours after sunrise could provide useful information in this regard.

A complete list of the 121 parameters which served as input for the RF model

abbreviation	description
X_mean / X_std	The mean (_mean) and standard deviation (_std) were calculated for 5x5 grid boxes around the location of measurement for a certain COSMO-D2 variable (X)
variable	**description**
CLCL, CLCM, CLCH + _mean/_std	mean/standard deviation of cloud fraction for low (CLCL), medium (CLCM) and high (CLCH) clouds
SWDIR_S, SWDIFD_S + _mean/_std	mean/standard deviation of shortwave direct (SWDIR) and diffuse (SWDIFD) irradiance
lwp_tot, iwp_tot + _mean/_std	mean/standard deviation of total (_tot) liquid (lwp) and ice water path (iwp)
cf_X , with X = 4000, 5000, 6000, 7000	maximium-random-overlap cloud fraction derived from layer cloud fraction (clc) from surface up to and altitude of X meters averaged over 5x5 grid boxes

Table 3.6.: Examples for variables computed from standard COSMO-D2 parameters.

can be found in Table A.2 in the appendix and a selection of engineered features is listed in Table 3.6.

3.8.3. Feature selection

As described in Section 3.8.1 there are several reasons to reduce the number of features – 121 in this case – until only the once with actual prediction skill remain in the dataset. This process of feature selection for the RF model is described in detail in this section.

RFE

As a first step of the optimization procedure for the RF model the most important features were extracted from the variables of the dataset based on COSMO-D2 forecasts derived in the previous section. For this purpose a Recursive Feature Elimination (RFE) based on the Permutation Feature Importance (PFI) proposed by Breiman (2001) was used. PFI was chosen because it is less biased towards continuous features and features with many categories than the so-called Gini Importance (GI) (Strobl et al., 2007) which is the default feature importance measure of the scikit-learn RandomForestRegressor (see also Section 2.5.2). PFI measures the decrease in prediction skill of the RF when the values of a certain feature are randomly permuted which breaks the correlation between the feature and the predicted quantity. The larger the decrease in prediction skill, the higher the importance of the considered feature. As PFI and GI are both unreliable in

the presence of correlated features (Strobl et al., 2008) the RFE was used additionally. In the process of RFE the least important features are removed step by step from the original dataset which is a well suited approach in presence of correlated features (Gregorutti et al., 2017). For the calculation of PFI the sklearn.PermutationImportance of the Python library eli5 in combination with k-fold cross validation (CV) of scikit-learn (Pedregosa et al., 2011) was used.

In the cross validation process the training dataset itself is split into training and validation datasets according to the defined number of folds (k) (see Section 2.5.3). This procedure allows training and validating the model on the entire training dataset which is beneficial to select predictive features with good generalization properties even if the dataset is small like in this study. Additionally, tuning the model based on CV leads to better generalization of the model to unseen data as possible peculiarities of training dataset can be reduced. For CV the number of folds was set to ten which results in the same number of individual train and test datasets. As the k-fold algorithm splits the data linearly along their index dimension, the dataset was shuffled randomly before the splitting to generate more homogeneous data subsets. Furthermore, the Root Mean Square Error (RMSE) was used as criterion to quantify the PFI.

Based on the described technique the RFE was performed in three consecutive steps to keep the computational effort in reasonable limits. In the first step 60 features from the total number of 121 were selected using 500 trees for the RF which was otherwise run with the default settings. At each elimination step the last two features were removed according to the mean PFI of all ten CV results. In the next step 30 of these 60 features were selected with an increased number of 1000 trees and by removing only the least important feature in each elimination step. The final selection step from 30 down to the most important feature was then performed with 2000 trees, removing the least important feature as before. Increasing the number of trees makes the PFI results more stable and hence more reliable, whereby Probst et al. (2019) recommend to use on the order of 1000 or more trees. Therefore, the number of trees was successively increased as the computational effort reduced due to the smaller number of features left in the dataset.

Final set of features

After the RFE had been completed the final set of features was determined. Thereby the focus was to select as few features as possible in the final dataset by maintaining a high prediction accuracy of the RF. Having only a limited number of features is desirable as it makes the model faster and the remaining variables can help to gain a deeper understanding than it would be possible with 100 features or more (see also Section 3.8.1). In a first step the dependence of the RMSE of the RF model on the number of selected features was investigated. Therefore, one feature after the other was successively added to the dataset according to their ranking determined via RFE while the RMSE of the model was evaluated each time a new feature was added. The RMSE was determined

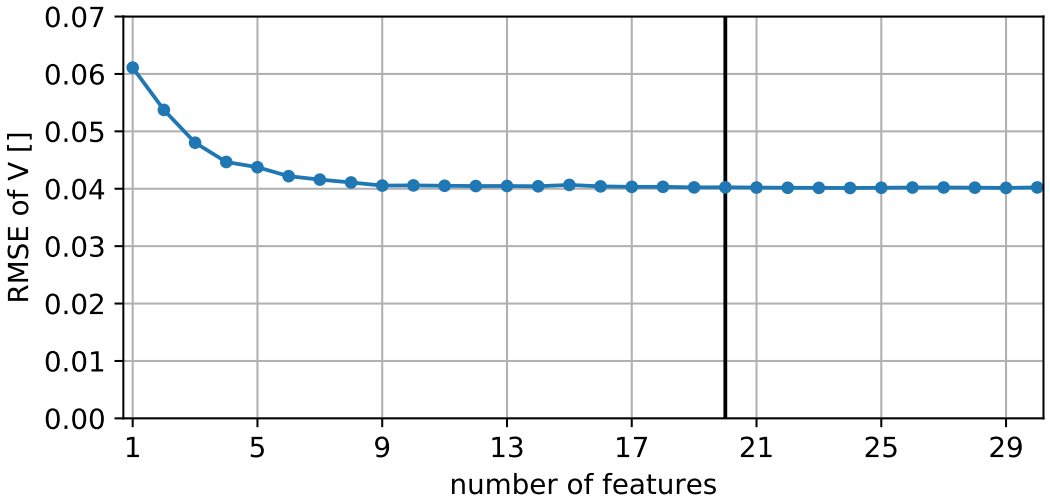

Figure 3.28.: RMSE of the variability index V calculated via tenfold cross-validation on the training dataset as a function of the number of features added to the dataset. The features were added according to their ranking determined by the recursive feature elimination. The vertical black line indicates the 20th feature which is the last one included in the preliminary set of features.

by averaging over all experiments of a tenfold CV on the training dataset with previous random permutation of the dataset as it was done before in the RFE. From the resulting curve shown in Figure 3.28 it could be inferred how many features were necessary until the RMSE converges and adding more features did not yield further improvements. The RMSE quickly reduces for the first five features and is close to 0.04 after ten features were added to the dataset. Afterwards only small changes in RMSE can be observed and even small increases in RMSE appear for instance when feature number 15 is added. Based on the convergence behaviour of the RMSE it was decided take the first 20 features according to the RFE ranking as a first preliminary feature set. The RMSE for these 20 features is 0.0402 and thus 1.5% smaller in comparison to a dataset containing all 121 features. This lager error in case of more features can be attributed to overfitting of the dataset because of too many available degrees of freedom. Overfitting has a negative effect on the generalization ability of the RF model and is therefore another argument for reducing the number of features to a minimum.

Even if RFE in combination with PFI and CV performs well in the presence of correlated features one can not be sure that the selected feature set does not contain highly correlated variables. As correlated variables contain similar and hence redundant information only one of them is needed for the RF model. Therefore, the correlation matrix for all features in the preliminary feature set was calculated based on the Pearson correlation coefficient which is shown in Figure 3.29. The most pronounced characteristic of the correlation matrix is the region of high correlations among the zonal wind variables (u) at different pressure levels. As all pressure levels of u are at 600 hPa or lower –

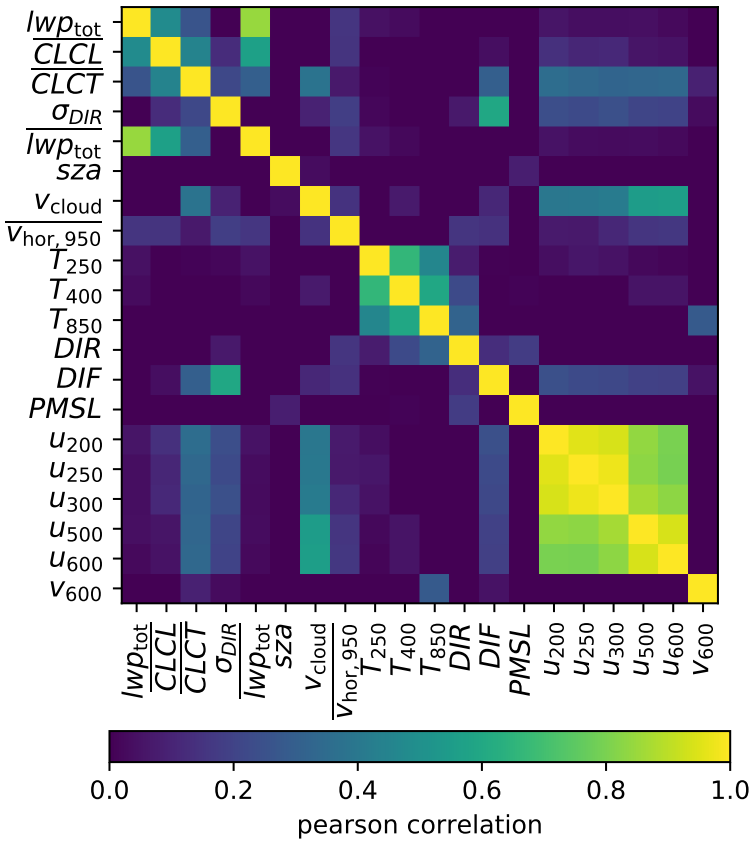

Figure 3.29.: Pearson correlation matrix for the top 20 features of the RF model calculated on the training dataset.

meaning altitudes larger than about 4 km – the correlation is due to the fact that they represent the wind in the free troposphere. In this region the flow is almost in geostrophic balance which means that the sign of u is usually the same for all levels while only the magnitude changes with height.

Another pronounced correlation can be found between the total liquid water path for the grid point closest to the measurement $\mathrm{lwp_{tot}}$ and its average over 5x5 grid boxes $\overline{\mathrm{lwp_{tot}}}$. One can easily imagine that the amount of clouds at a certain grid point is somehow linked to the surroundings. At this point it should be mentioned that $\mathrm{lwp_{tot}}$ does not necessarily represent the actual simulated LWP of COSMO-D2 as the parametrization for subgrid-clouds applies some corrections to the grid scale lwp in order to optimize it for the simulation of radiative transfer. Nevertheless, it obviously contains valuable information concerning GHI variability. To decide which of the correlated features should be excluded a hierarchical clustering was applied based on the correlation matrix using Wards method as a distance metric. The clustering helps to visualize the correlation among different variables as can be seen in Figure 3.30 which shows

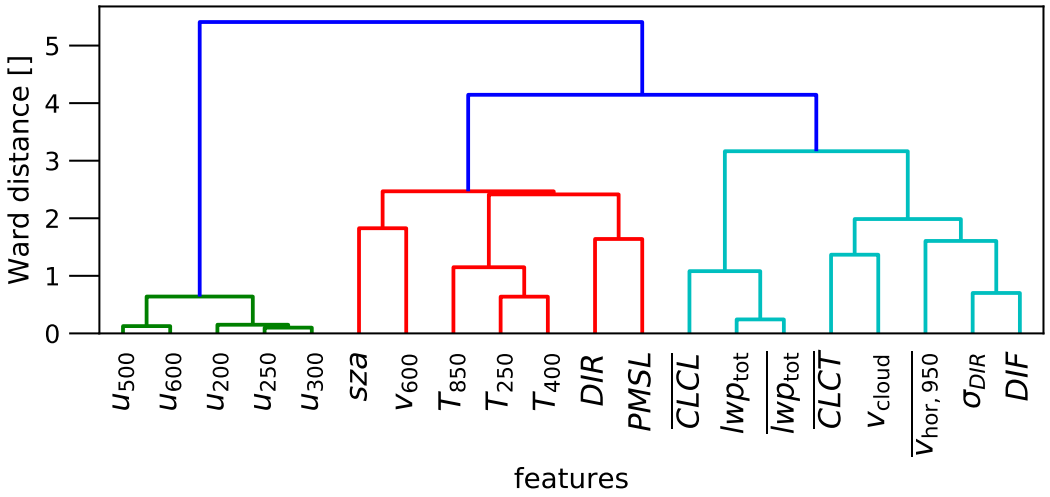

Figure 3.30.: Dendrogram visualizing the result of the hierarchical clustering applied to the correlation matrix shown in figure 3.29. As distance metric Ward's method was used. The smaller the distance between individual features or clusters of multiple features the higher their correlation.

a dendrogram of the identified clusters. On the left-hand side of the dendrogram the previously mentioned zonal wind speed u at different pressure levels are summarized in the green cluster. Thereby the wind speeds at lower pressure levels (higher altitudes, 250 $hPa \approx 10\ km$) on the right branch of the cluster are very closely correlated because they are all located in the tropopause region where the wind direction is essentially the same at all levels and the wind speed scales with altitude depending on the exact location of the jet stream. Therefore, u_{250} and u_{300} were excluded from the dataset keeping only u_{200} as this variable was ranked best of all three by the RFE. The two zonal wind variables at higher pressure levels (lower altitudes, 500 $hPa \approx 5\ km$) on the left branch are also highly correlated and thus u_{600} was excluded as u_{500} is ranked better by the RFE.

Finally, the already discussed correlation between the lwp variables can be seen on the left branch of the cyan cluster on the right-hand side. Even though their Pearson correlation is not as high as the one between the wind variables, when calculating the Spearman rank correlation they almost correlate perfectly. This means that there is a monotonic correlation between both variables which is nonlinear to some degree. As an RF is capable of mapping nonlinear relationships it is sensible to reject one of these features to avoid duplicate information in the dataset. In fact, adding lwp_{tot} to the dataset led to an increase in forecast error which can be observed in Figure 3.28 at feature position 15. Consequently, lwp_{tot} was rejected from the feature set due to its lower ranking in the RFE.

After four of the 20 features from the preliminary feature set were removed due to high correlation, the remaining 16 were again used to derive the RMSE as a function of the added features as already shown in Figure 3.28. The result of this procedure

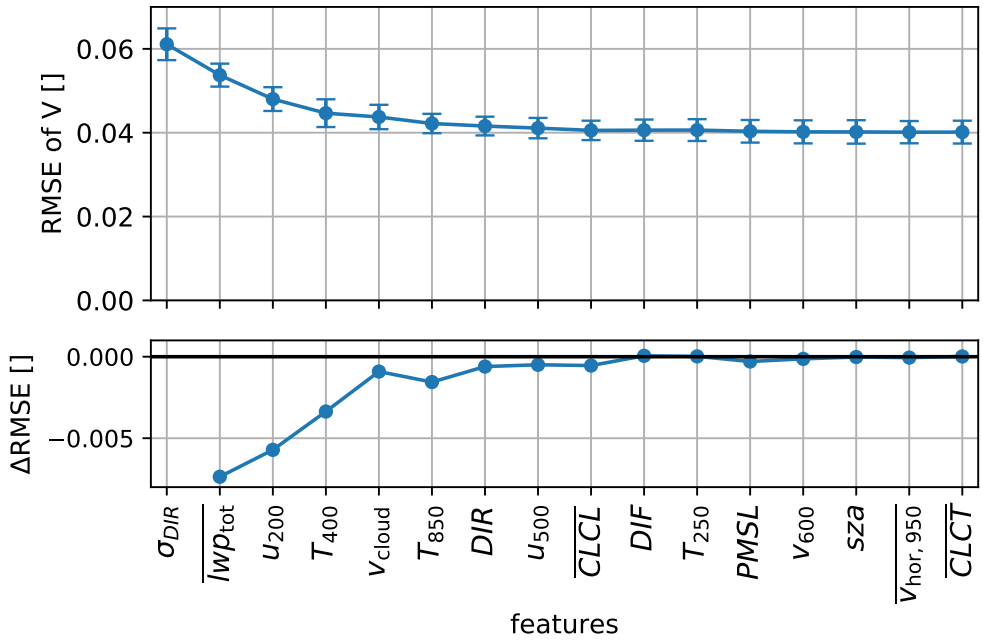

Figure 3.31.: Upper panel: RMSE of the RF model prediction on the (filtered) training dataset calculated via tenfold cross validation as a function of the remaining top 16 features after removal of four correlated features. The errorbars indicate the standard deviation over the ten cross validation experiments. Lower panel: change in RMSE by adding an individual feature to the dataset. Consequently, no change in RMSE can be calculated for the first feature.

is shown in Figure 3.31 where the upper panel depicts the mean RMSE derived via tenfold CV on the training dataset. Furthermore, the standard deviation of the RMSE calculated over the ten individual experiments is illustrated by the vertical bars of each data point. As small changes in RMSE can hardly be noticed in the upper panel, the lower panel depicts the change in RMSE for each added feature. Consequently, the first data point in the lower panel represents the change in RMSE caused by adding the averaged total lwp ($\overline{\text{lwp}_{\text{tot}}}$) to the dataset. The RMSE curve in the upper panel converges quickly towards a value of 0.04 where a plateau is reached after adding the spatially averaged cloud fraction of low clouds ($\overline{\text{CLCL}}$) which persists until the mean sea-level pressure ($PMSL$) is added.$PMSL$ leads to a reduction of RMSE which continues by adding the meridional wind at 600 hPa (v_{600}) while the remaining three features do not lead to further improvements. Therefore, it was decided to keep the first 13 of the 16 correlation-filtered features in the final set of features. The abbreviations of these 13 features together with a short description are listed in Table 3.7.

To be able to assess the benefit of the data filtering performed in Section 3.4 the procedure of feature elimination with subsequent selection of a final set of features was also performed based on the unfiltered dataset. As the procedure is identical it will be

Final feature set	
Parameter	description
σ_{DIR}	standard deviation of direct horizontal solar irradiance within 5x5 grid boxes around location of measurement
$\overline{\mathrm{lwp}_{\mathrm{tot}}}$	total (grid scale + subgrid scale) liquid water path averaged over 5x5 grid boxes around location of measurements
u_{200} u_{500}	zonal windspeed at 200 and 500 hPa pressure level
$\mathrm{v}_{\mathrm{cloud}}$	cloud advection speed estimated as described in equation 3.7.2
v_{600}	meridional windspeed at 600 hPa pressure level
T_{250} T_{400} T_{850}	Temperature at 250, 400 and 850 hPa pressure level
DIR DIF	direct and diffuse horizontal irradiance at the closest grid point
$\overline{\mathrm{CLCL}}$	cloud fraction of low clouds ($p < 800\ hPa$) averaged over 5x5 grid boxes around the location of measurement
PMSL	atmospheric pressure at mean sea level

Table 3.7.: Description of the 13 features contained in the final dataset of the RF based on the quality filtered dataset.

not described in detail here but a short version of it containing the according results can be found in Section A.1 of the appendix.

3.8.4. Discussion of final feature set

One of the ideas of using a random forest model was to extract the most predictive features from COSMO-D2 forecast data which can provide a deeper understanding of the processes controlling GHI variability. As the final feature set has been selected, this section will discuss possible reasons and mechanisms which make the selected variables important for GHI variability. It was tried to group features with similar properties and describe their possible influences as a whole to get a better overview and avoid repetition.

1. Radiation related features - σ_{DIR}, DIR, DIF:
   The spatial variability of the direct horizontal irradiance σ_{DIR} turned out to be the most predictive variable for GHI variability. This might be not a surprising result as the reduction of direct radiation by clouds is the main cause of variability. How-

ever, it was shown in Sections 3.6.4 and 3.6.5 that the variability in cases of 1D radiative transfer and low spatial resolution does not exhibit a simple relationship to the actual small scale variability with 3D effects taken into account. This means that the RF model was able to detect a complex nonlinear relation which provides a better prediction of V with only one parameter than the LUT model which uses three parameters.
Less important but still among the best features are the diffuse and direct irradiance at the grid point closest to the measurement location (DIF and DIR). Both contain information about the cloud situation (e.g. higher DIF relates to larger cloud fraction) and DIR could also serve as a proxy for the solar zenith angle which has been shown in Section 3.6.6 to be a relevant parameter.

2. Cloud related features - $\overline{\mathrm{lwp_{tot}}}$, $\overline{\mathrm{CLCL}}$, $\mathrm{v_{cloud}}$:
The second most important variable is represented by the average total liquid water path $\overline{\mathrm{lwp_{tot}}}$ which includes subgrid-scale cloud water. This feature mainly contains information about the optical thickness of the clouds but also about the cloud fraction. It also shows that it is important to consider the subgrid-scale convective clouds as the grid scale lwp was not ranked within the top 30 features.
Additionally, the average cloud fraction of low level clouds $\overline{\mathrm{CLCL}}$ is a predictive feature for GHI variability. This is in good agreement with the arguments used for the exclusion of high level clouds for the LUT model (see Section 3.7.3). Optically thick and scattered convective clouds at low levels are more relevant for GHI variability than medium or high level clouds which are usually optically thinner which reduces their impact on GI at the surface. Furthermore, they often span larger areas which leads to less GHI ramps at cloud edges.
The third variable assigned to the cloud related variables is $\mathrm{v_{cloud}}$ which contains cloud as well as wind speed information. $\mathrm{v_{cloud}}$ was derived according to Equation 3.7.2 and intended to provide an estimate of the cloud advection speed which is particularly relevant for GHI variability. As this parameter was ranked as the fifth most important feature by the RFE it obviously contains relevant information as expected. In contrast to the LUT approach where $\mathrm{v_{cloud}}$ with an altitude threshold of 5000 m provided the best result, the RF model does not rely on the correct value of the advection speed but only uses the correlation between feature and target variable.

3. Weather regime related features - $\mathrm{u_{200}}$, $\mathrm{u_{500}}$, $\mathrm{v_{600}}$, $\mathrm{T_{850}}$, $\mathrm{T_{400}}$, $\mathrm{T_{250}}$, PMSL:
These seven features mainly describe the overall weather situation rather than the specific conditions at a certain location and form the largest class within the selected dataset. The wind variables provide information about the zonal (u) and meridional (v) flow in the mid and upper troposphere which is associated with the transport of air masses from different regions. A strong zonal flow for instance indicates advection of moist air from the Atlantic Ocean. More precise information about the location within an upper level ridge or trough and the associated lower level high- and low-pressure systems is provided by the combination of u and v: positive u and v for example are occurring on the front side of a trough which is associated with

low pressure near the surface and hence rising air which favors the formation of clouds. In the context of pressure systems the mean sea-level pressure (PMSL) of course provides the most direct measure but as not each low has the same pressure structure the location within the system can be assessed more precisely by the wind components. As a last possible connection the impact of wind speed on cloud advection speed should be mentioned which presumably can be considered less relevant as the v_{cloud} parameter is also part of the feature set.
Focussing on the temperature at different levels, T_{850} is a particular relevant variable in weather forecasting as it is located above the planetary boundary layer and is not influenced by the variable solar heating at the surface. Therefore, it contains important information about the air masses present at a certain location. In combination with the temperatures at higher altitudes these variables also provide information about the stability of the atmosphere and can give an indication of the probability of cloud formation.

Based on the selected final set of 13 features the hyperparameters of the RF were tuned in the last step of the optimization process which is described in the next section.

3.8.5. Hyperparameter tuning

To further improve the performance of the RF predictions the so-called hyperparameters of the RF were adjusted to reduce the RMSE of the RF model. Hyperparameters control the way individual trees of the RF are built and hence influence the outcome of the RF's prediction. For example the parameter min_samples_leaf of the scikit-learn RandomForestRegressor controls the number of elements in the final leaf node of each tree which means that the depth of a tree is reduced by increasing min_samples_leaf. Parameters with similar influence are min_samples_split or max_depth which directly determines the depth of each tree. More details on RF hyperparameters can be found in Section 2.5.4 of the previous chapter or in the review article by Probst et al. (2019).

For the hyperparameter tuning the GridSearchCV function of scikit-learn was used which evaluates the performance of the RF for all possible combinations of a previously defined set of hyperparameters using CV. Just like in the previous section a tenfold CV was performed using the RMSE as performance measure on the training dataset. For all tuning procedures the scikit-learn RandomForestRegressor was run with the default settings except for the parameters which had already been tuned in a previous step and of course the parameter to be tuned in the current step. As a first step in the tuning process the number of trees (n_trees) of the RF was determined. In contrast to other hyperparameters it is generally beneficial to use more trees, however a sensible number can be determined by finding the threshold above which adding more trees does not improve the prediction (Probst et al., 2019). Using the RF with default settings a tenfold CV was performed on the training data yielding the minimum of the RMSE for 800 trees as shown in Figure 3.32. The upper panel depicts the overall behaviour of the RMSE as a function of n_trees which reveals convergence at about 600 trees. To be able to visually identify the optimal number of trees the lower panel shows a detail of the same plot

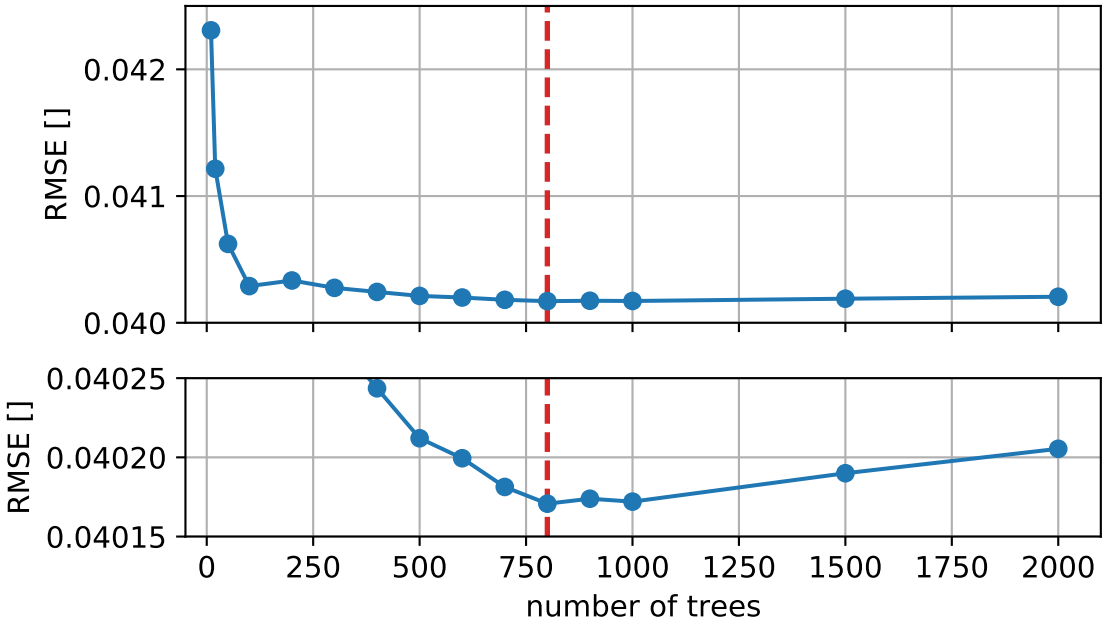

Figure 3.32.: Upper panel: RMSE of the RF model prediction on the (filtered) training dataset calculated via tenfold cross validation as a function of the number of trees. The dataset holds the selected 13 features of the final dataset. The dashed red line indicates the number of trees with the smallest RMSE. Lower panel: detail of the upper panel to visualize the small differences in RMSE of adjacent points.

which shows that the RMSE is smallest for 800 trees (dashed red line). Increasing the number of trees actually increased the RMSE which makes the choice of 800 trees more efficient and more accurate at the same time.

After the number of trees was determined, the following hyperparameters were optimized in the grid search procedure.

- max_features: number of features used to build each tree
- max_samples: number of samples drawn from the dataset to build a tree
- min_samples_leaf: minimum number of data points in the last node (leaf) of each tree
- min_samples_split: minimum number of data points required to split a node

Following the findings of Probst et al. (2019) the first two hyperparameters (max_features, max_samples) were selected as they exhibit the largest potential for considerable improvemnts. The other two parameters (min_samples_leaf, min_samples_split) both affect the depth to which the tree is grown in different ways but according to Probst et al. (2019)

their influence on improvements is rather small. Although the RandomForestRegressor has some more tunable hyperparameters, they mostly have similar properties as the hyperparameters listed above. Furthermore, their potential for further improvements is expected to be negligibly small and hence they were not included in the GridSearch procedure. More detailed information on the hyperparameters and their impact on the RF performance can be fond in Section 2.5 or in Probst et al. (2019).

The values tested for the individual hyperparameters applied in the GridSearch procedure are listed in Table 3.8. For the max_features parameter three widely used values were applied. The first one, *auto*, means that all features are used to build each tree, whereas *log2* and *sqrt* cause the RF to use only a fraction of all features defined by the base-2 logarithm and the square-root of the number of all features respectively. The max_samples parameter was varied between 0.2 and 0.9 which means that each

Parameter	**Values**
max_features	auto (all features), log2, sqrt
max_samples	None (number of all data points), 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2 (fraction of all data points)
min_samples_leaf	1, 2, 5, 10, 15
min_samples_split	2, 5, 10, 15

Table 3.8.: Values of individual hyperparameters for the grid search procedure.

ensemble of data points retrieved via bootstrapping for an individual tree contains between $0.2 \cdot n$ and $0.9 \cdot n$ data points with n being the total number of samples in the dataset. Additionally, max_features was set to *None* causing each bootstrap sample to contain as many data as the entire dataset holds (max_samples $= n$). The values of the other two hyperparameters which mostly control the depth of the trees were varied between 1 (min_samples_leaf)/2 (min_samples_split) which leads to fully grown trees and 15 which reduces the depth of the trees to some extent. All other hyperparameters were set to the default values which are listed in Table 3.9.

According to the GridSearch procedure the following combination of values results for the investigated hyperparameters leads to the best performance of the RF:

- max_features = *sqrt*
- max_samples = *None*
- min_samples_leaf = 1
- min_samples_split = 2

The reduction of max_features to *sqrt* instead of *auto* (all features) gives an indication that the most predictive features like σ_{DIR} dominate the splitting and hence obscure

information contained in less important features. As the square-root of the selected 13 features equals three (rounded to integer), only a quarter of the features is used to build each tree. For the other three parameters the default values were found to yield the best prediction which means that trees are built of bootstrap samples as large as the dataset and trees are grown to maximum depth. One possible reason for this outcome is the relatively small size of the dataset which makes it necessary to explore the data as detailed as possible to extract as much as information as possible. Overall the tuning of the hyperparameters caused a slight reduction of the RMSE from 0.0402 to 0.0401 (0.25%) evaluated via tenfold cross-validation on the training dataset.

With this hyperparameter tuning step the final setup for the RF model was determined. However, there are still tools left tu further improve the prediction of the RF as will be shown in the next section.

3.8.6. Bias correction

Random Forests are known to produce a range dependent bias by overestimating small values and underestimating high values of the predicted variable (Zhang and Lu, 2012; Song, 2015). This property originates from the averaging over several hundred uncorrelated trees in an RF which makes the predictions reliable but also prevents to predict values beyond the extremes contained in the training data. Therefore, in case of very low values of the target variable, each tree will yield an equally small or larger result and an equally large or lower result for very high values which on average leed to the described range dependent bias. An example for the prediction of the variability index V is shown in Figure 3.33a). Here the filtered training dataset selected in Section 3.5.3 was randomly split into two parts whereby the larger one (90% of the data) was used for training an RF with the setup derived in the previous sections. This RF was then applied to the remaining 10% of the dataset to predict the variability index depicted in Figure 3.33. The cyan linear regression line shows the described range dependent bias, which appears to be relatively close to the 1:1 line at the first glance. However, this can be explained by the high point density close to zero and the overall small number of samples in the test set.

To correct this bias the BC1 method described by Zhang and Lu (2012) was used which is the first and best performing approach of the five methods described in their paper. The BC1 method uses a second RF model which is trained to predict the deviation of the original prediction from the true value. As Zhang and Lu (2012) formulate this method in a mathematically rigorous way it is helpful to have a look at Song (2015) who summarizes the algorithm in a cooking recipe like fashion. The described steps were applied to the RF model developed in this thesis as follows: In the first step an RF is trained to predict the target variable (V) on the training dataset based on the selected input features. Then the trained RF is used to predict V for the training dataset itself which also leads to a similar bias as the prediction on unseen data which can be seen by comparing graphs a) and b) in Figure 3.33. In fact the linear regression shows a smaller slope for the prediction of the training data than for the test dataset which may be due

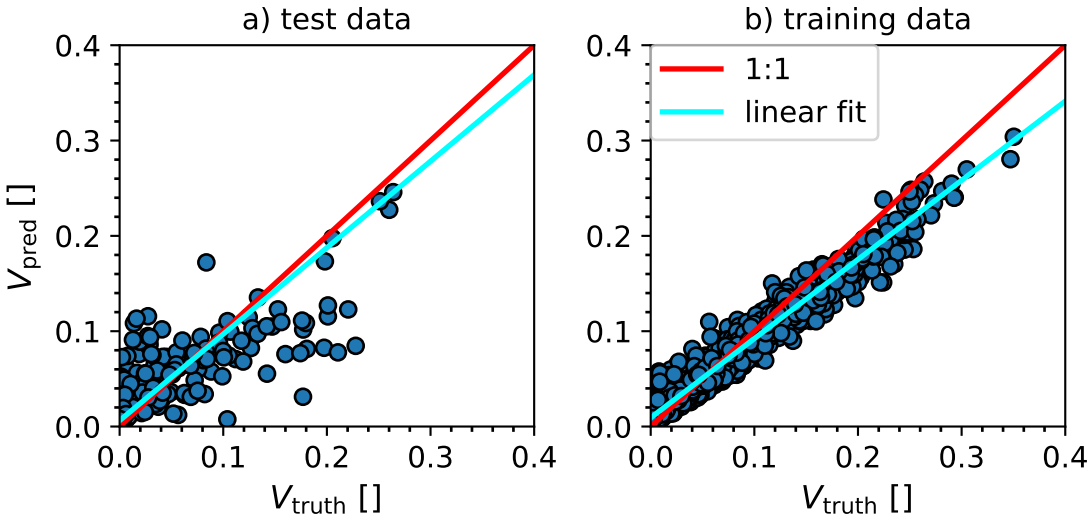

Figure 3.33.: a) variability index predicted by an RF trained on 90% of the filtered training dataset and evaluated on the remaining 10% (V_{pred}) plotted against the measured variability (V_{truth}). The 1:1 line is shown in red and a linear fit in cyan. b) same RF model as in a) but applied to the training dataset used to train the RF.

to the small number of samples in the test dataset depicted in Figure 3.33a). From the prediction of V on the training dataset the residuals to the true value are calculated in the next step. Now, the second RF comes into play (called RF_{res} in Song (2015)) which is intended to predict the residuals based on the prediction of V by the original RF. Therefore, a second training dataset is created containing the same features as the original dataset but with V_{pred} added as additional feature. The second RF (RF_{res}) is trained on this extended dataset to predict the residuals of V. For RF_{res} the same features and hyperparameters were used as for the main RF. This was done as it is very likely that features which are predictive for variability in general are also suited to predict the deviation of the prediction itself. Furthermore, the hyperparameter tuning for the RF improved the result by 0.25% which means that a separate tuning of RF_{res} would give at most similar improvements for the prediction of residiuals. Overall, such a small improvement of the bias correction would hardly have a measuarbale impact on the accuray of the RF predictions as a whole. Consequently, the effort of creating a new RF model for the bias correction from scratch is disproportionate to the expected improvements.

In summary the BC1 method is a two-step step process where the initially predicted value of V is corrected by a second RF prediction which depends on the inital prediction of V. For a qualitative assesment of the impact of the bias correction, the BC1 method was applied to the example shown in Figure 3.33. Figure 3.34 shows the according prediction with applied bias coorection. To visualize the changes due to the

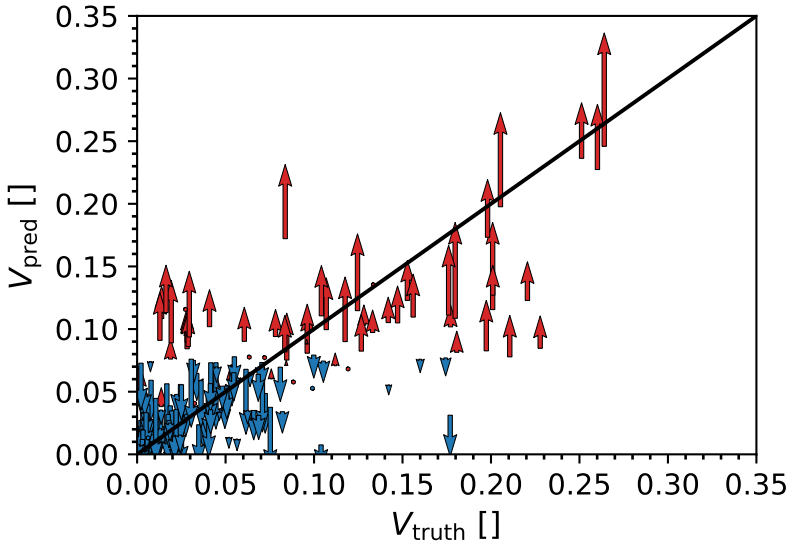

Figure 3.34.: Change in predicted variability (V_{pred}) when the bias correction according to the BC1 method by Zhang and Lu (2012) is applied to the RF prediction shown in Figure 3.33a). Blue arrows indicated a correction to smaller values and red arrows a corection to larger values. The 1:1 line is depicted in black.

bias correction the magnitude of the residuals predicted by RF_{res} is depicted, whereby red arrows pointing upward represent positive and blue arrows pointing downward represent negative corrections in V. Each arrow starts at the initially predicted value of V and points to value of the corrected variability index ($V_{new} = V_{old} + \Delta V$). As expected from the observations in Figure 3.33 small predicted values of V ($V_{pred} < 0.07$), which tend to have a positive bias, are mostly corrected in negative direction and larger values are corrected in positve direction. Consequently, most of the predictions are improved by shifting V closer to the 1:1 line. However, in cases where the initally predicted variability has a large error, the correction further increases the error instead of reducing it. This effect can in particular be observed in the upper left and lower center part of Figure 3.34. Another influence of the bias correction wich has to be taken care of are negative values of V which occur when predicted values close to zero are corrected with negative residuals. These negative values are handled by setting them to zero.

To assess the impact of the bias correction procedure in a quantitative way, the RF model including all previous optimizations was evluated on the filtered training dataset using tenfold cross validation (CV) with and without applying the BC1 bias correction method. Figure 3.35 shows a 2D histograms for the described scenarios where the historam 3.35a) depicts the case without and 3.35b) the case with applied bias correction. Overall, the bias corrected histogram shows better alignment along the 1:1 line compared to the more skewed distribution without bias correction. In particular the impact of the correction is visible for small variability levels where the predictions are

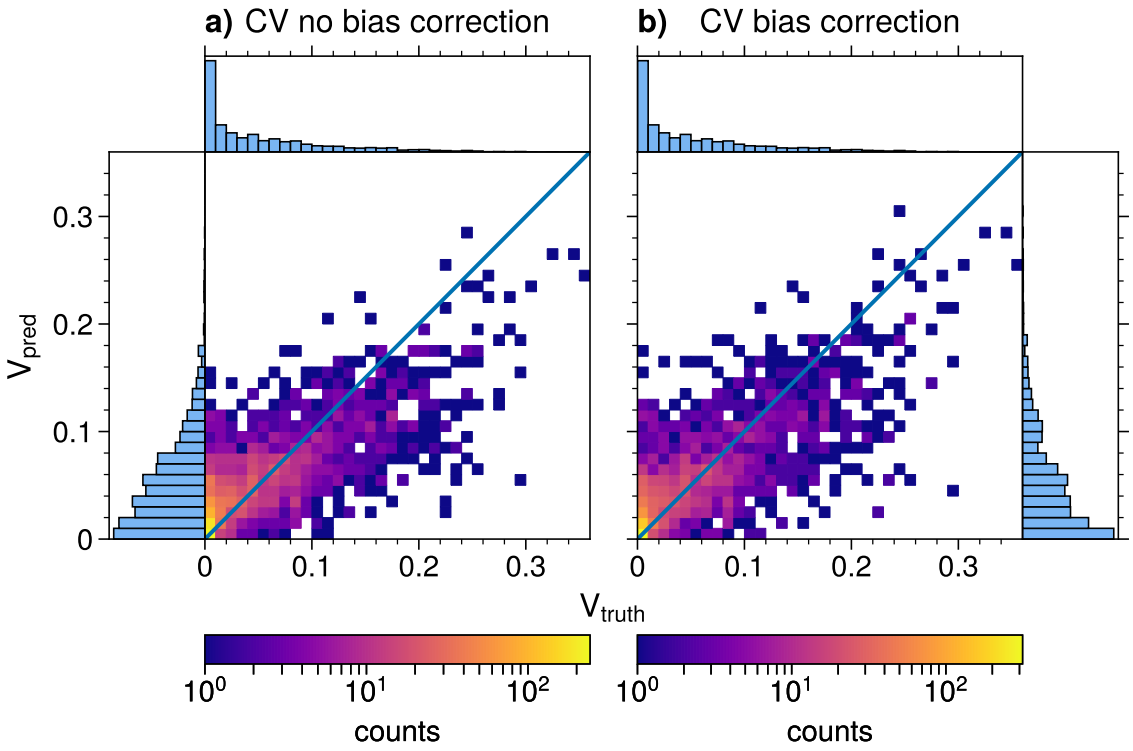

Figure 3.35.: 2D histograms of the true variability index (V_{truth}) plotted against the variability predicted by the RF model (V_{pred}). Graph a) shows the result for the prediction of the training dataset using tenfold cross-validation (CV) without bias correction and graph b) the result for CV with applied bias correction. The color scale changes logarithmically with the number counts in each pixel. 1D histograms on x- and y-axes show the relative frequency within each bin for V_{truth} and V_{pred} respectively. The blue lines in both histograms indicate the 1:1 line.

concentrated closer to zero and also large values of V are improved as expected from the analyisis of Figure 3.34. Furthermore, the previously described wrong correction of outliers becomes also visible in the upper left and lower center part of the graph and leads to an overall broadening of the distribution perpendicular to the 1:1 line. The effect of the bias correction can also be seen in the 1D histograms attached to the x- and y-axes which show the relative frequency of predictions in each variability bin. For the measured variability the highest relative frequency can be found in the first bin followed by a strong decrease in the second bin. Without bias correction the highest relative frequency is also found in the first bin but the relative frequency only decreases slowly in the next few bins. Applying the bias correction increases this gradient in relative frequency which makes the distribution more similar to the measurement.

The described visual improvments achieved by the bias correction are also supported by a reduction of the RMSE by 0.8% (from 0.0401 to 0.0398) and a reduction

of almost 4.2% (from 0.0280 to 0.0269) in MAE. This observation can be attributed to negative impact of the bias correction on outliers which are shifted even further away from the 1:1 line and therefore have especially high impact on RMSE due to the squared distance weighting. As the distance to the true value only contributes linearly in the MAE metric, outliers have a smaller contribution which gives a higher weight to the improved predictions. Overall, it can be stated that the bias correction with the BC1 method improves the skewnes of the RF prediction from a visual point of view and reduces RMSE as well as MAE. Therefore, it was decided to apply this correction also to the predictions of the final validation dataset.

3.8.7. Summary of the final RF model

To give a brief overview over the final setup of the RF model for the prediction of GHI variability index V, Table 3.9 summarizes the options chosen for the underlying RF model of the scikit-learn library (RandomForestRegressor) as well as the 13 selected features. The configuration of the RF model is devided into three parts: the first one lists the hyperparameters which were optimized in the previous section, the second one contains the default values chosen for the remaining hyperparameters which were not included in the optimization process. The third part shows all other parameters of the RandomForestRegressor like the bootstrap parameter which can be switched on or off using True or False as option or the random_state variable where a seed value for the random number generator of the RF can be chosen. This parameter was set to a fixed value for all RF experiments and tuning steps to ensure reproducibility and comparability. Finally the 13 selected features are listed in descending order starting with the most predictive feature σ_{DIR}.

Optimized RF Hyperparameters	
max_features	sqrt
max_samples	None (number of all data points)
min_samples_leaf	1
min_samples_split	2
Default RF Hyperparameters	
min_weight_fraction_leaf	0.0
max_leaf_nodes	None
min_impurity_decrease	0.0
ccp_alpha	0.0
Other RF Parameters	
criterion (splitting)	squared_error
bootstrap	True
oob_score	False
n_jobs	-1
random_state	42
warm_start	False
Bias Correction	
BC1 method according to Zhang and Lu (2012)	
Selected Features	
σ_{DIR}, $\overline{\mathrm{lwp}_{\mathrm{tot}}}$, u_{200}, T_{400}, $\mathrm{v}_{\mathrm{cloud}}$, T_{850}, DIR, u_{500}, $\overline{\mathrm{CLCL}}$, DIF, T_{250}, PMSL, v_{600}	

Table 3.9.: Configuration of the final RF model developed based on the quality filtered dataset. The first three sections of the table list the options chosen for the scikit-learn RandomForestRegressor and the last section contains the 13 selected features in descending order of their importance.

4. Results

In this chapter the models developed to predict GHI variability based on operational COSMO-D2 forecasts, presented in the previous chapter, are evaluated. Therefore, the LUT based model as well as the RF model were applied to the validation dataset which was held out during the optimization and tuning processes of both models. Furthermore, the impact of the data filtering procedure (see Section 3.4) on the prediction error is investigated.

4.1. LUT model evaluation

This section presents the results for the lookup-table (LUT) based model for the prediction of the variability index V described in Section 3.7. As the LUT model is based on the investigation of parameters relevant for GHI variability in realistic GHI simulations, this model serves as a reference for the machine learning approach using a random forest model.

4.1.1. 2D histogram

In the first evaluation step the LUT model was applied to the validation dataset held out during the optimization process. To be able to detect possible (systematic) differences between the training dataset used to tune the model the LUT model was also applied to the training dataset. The 2D histograms for both cases are shown in Figure 4.1, whereby the result for the training dataset is depicted in graph a) and the prediction of the validation dataset in graph b). The true variability index (V_{truth}) derived from measurements is plotted on the x-axis while the predicted variability (V_{pred}) forms the y-axis. Additionally, 1D-histograms along both axes show the relative distribution of data points for the predicted and the observed variability.

Having a look at the histogram of the predicted training dataset (4.1a) one notices the strikingly high density of data points for low values of V. In fact 68% of all data points are located at a variability smaller than 0.1 for V_{truth} and V_{pred}. Looking even closer one can see that the very first pixel in the lower left corner (partly hidden by the 1:1 line) is the one with the highest number count of all pixels holding 27% of all data in the training set. Another feature of the histogram is the high density of points along the x-axis up to $V_{\mathrm{truth}} \approx 0.1$. This systematic underestimation is a result of the tuning of the cloud advection speed by linear scaling with a value of 0.64 (see Section 3.7.3). Thereby the variability of all predictions is reduced and as a consequence points with already

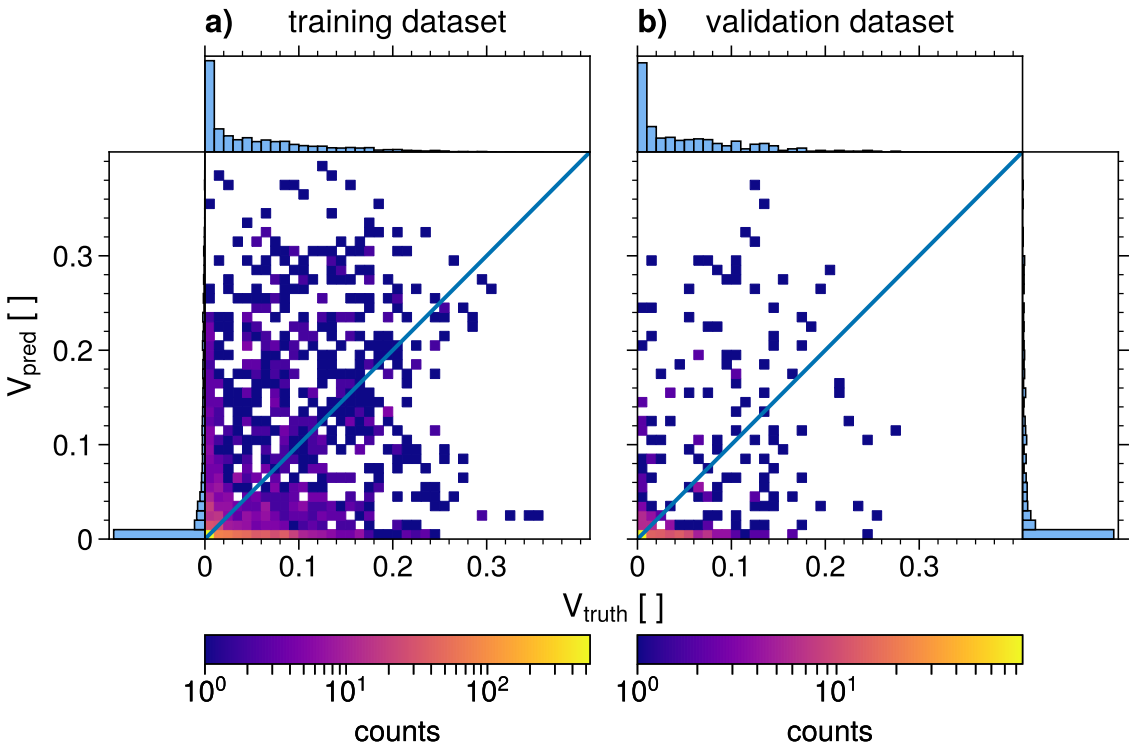

Figure 4.1.: 2D histograms of the true variability index (V_{truth}) plotted against the variability predicted by the LUT model (V_{pred}). Graph a) shows the result for the prediction of the training dataset and graph b) the result for the validation dataset which was not used for model optimization (hold-out validation). The color scale changes logarithmically with the number counts in each pixel. 1D histograms on x- and y-axes show the relative frequency within each bin for V_{truth} and V_{pred} respectively. The blue lines in both histograms indicate the 1:1 line.

small V prior to the scaling reach values close to zero once the scaling is applied. This behaviour had already been shown for a more extreme scaling of 0.2 in Figure 3.27b). A last notable property of Figure 4.1a) is the overestimation of V throughout the entire range of values. It is important to note that this occurs only for a very limited number of cases compared to the total number of data points. However, these outliers have a significant contribution to the overall error especially when considering the RMSE where the deviation from the true value is weighted quadratically. The value of the RMSE for the LUT prediction on the training dataset amounts 0.08.

Comparing the histogram for the prediction of the training dataset (4.1a) to the prediction of the validation dataset (4.1b) shows that all previously described properties of the training dataset also apply to the validation dataset. Despite the low amount of data points, ($\sim$ 300) the high point density around zero as well as the strong outliers above the 1:1 line are clearly visible in the histogram. Consequently, the LUT model

prediction for the validation dataset has an RMSE of 0.078 which is very similar to the prediction of the training dataset. Usually one would expect the error to be smaller for the training dataset as the model was tuned on this data but due to the small amount of data in the validation dataset the small difference in RMSE might be attributed to statistical noise rather than being an actual feature of the LUT model.

4.1.2. LUT model performance on unfiltered data

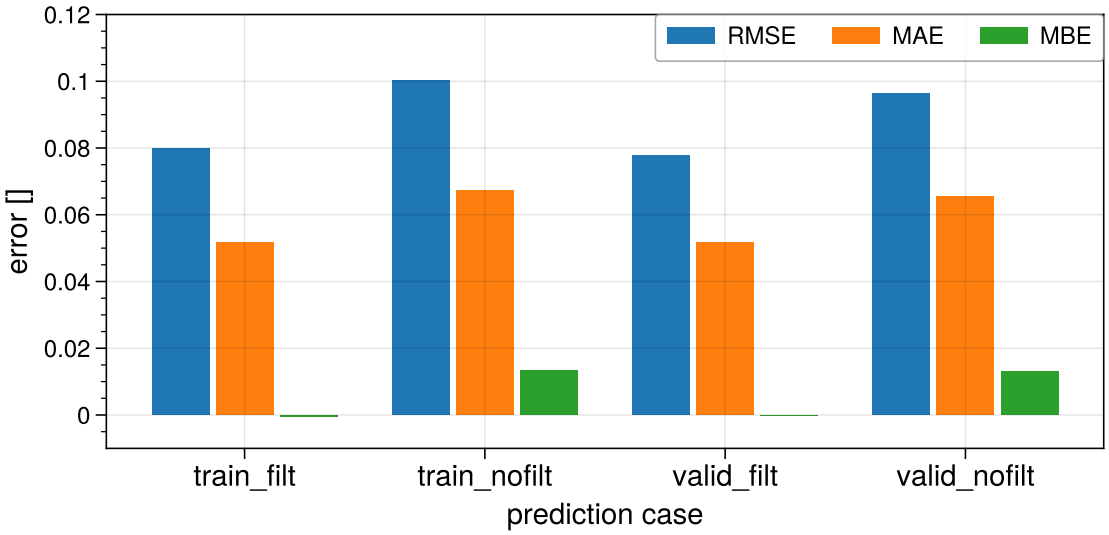

Figure 4.2.: Error measures RMSE, MAE and MBE for different predictions of the LUT model. Starting with the left most case the LUT was used to predict the training dataset with quality filtered data (train_filt), the training dataset without quality filtering (train_nofilt), the quality filtered validation dataset (valid_filt) and the unfiltered validation dataset (valid_nofilt).

To assess the effect of the data filtering described in Section 3.4 on the performance of the LUT model, the LUT was applied to the unfiltered training and validation dataset. For an explanation of the difference between filtered and unfiltered dataset it is referred to Section 3.4.6 of Chapter 3. The results for three different error metrics (RMSE, MAE, MBE) are depicted in Figure 4.2 which also shows the error on the filtered dataset for comparison. Starting on the left side, the bar chart shows the error for the LUT prediction on the filtered training dataset (train_filt) as first case and on the unfiltered dataset (train_nofilt) as the second case. With an RMSE of 0.1 the prediction on the unfiltered dataset has a 25% larger error than the prediction on the filtered dataset. Together with the positive bias and the also increased MAE of the train_nofilt case it can be concluded that the additional input in the unfiltered data caused predictions with large positive errors. Here it should be mentioned that the MBE of almost zero for the train_filt case is a result of the tuning for the sum of RMSE and MBE (see Section 3.7.3).

A similar behaviour can be observed for the prediction of the filtered (valid_filt) and unfiltered validation dataset (valid_nofilt). In this case the RMSE of the predicted variability index increases by 23% from 0.078 to 0.096. Additionally, the MBE is also very close to zero which shows that the LUT model performs consistently on training and validation data. Finally, a general statement can be made about the fact that MAE is considerably lower than RMSE for all cases. This is in particular due to outliers with exceptionally large error which increase RMSE compared to MAE due to the quadratic weighting of the deviation.

4.1.3. Discussion of LUT model results

Based on the results of the LUT model shown in the previous section it can be stated that there is at least some prediction skill of the model for small GHI variability ($V < 0.1$) whereas larger values of V can hardly be reproduced. However, also for smaller values of V the LUT systematically predicts too small variability for a part of the data which is mainly due to the scaling of the cloud advection speed. Additionally, heavy positive outliers across the entire range of V increase the RMSE to about 0.08 which is actually larger than the mean value of V for the training data of 0.055. This small mean value of the data shows an important property of the dataset which could already be seen qualitatively in the histograms of Figure 4.1: about 75% of all data points have a variability index smaller than 0.1. The issue of imbalance between small and high variability in the dataset arises the question if forecast systems for GHI variability should be optimized to predict variability in all cases or if the focus should be on high variability situations as these might be of higher interest for grid operation. This question will be addressed in more detail in the outlook section.

Analyzing the impact of the data filtering procedure it could be shown that the removed data actually lead to a considerable increase of all error metrics. Considering the LUT prediction only for the data rejected by the quality filtering for the training dataset the RMSE almost reaches 0.136 which is nearly twice the RMSE of the prediction on the filtered training dataset. It could be argued at this point that the LUT model was tuned to the filtered dataset and does therefore perform worse on the unfiltered data. And indeed, tuning on the unfiltered dataset reduces the RMSE for the unfiltered training dataset to 0.117 and also the RMSE for the filtered training data to 0.073 (similar results for the validation dataset). However, to compensate for the positive bias caused by the additional outliers in the unfiltered LUT prediction the optimal scaling factor for the cloud speed turns out to be 0.43 whereas it has been 0.64 for the filtered data. As a consequence, more predictions are pushed towards zero which increases systematic errors. This behaviour had already been described in the methods' chapter in Section 3.26. Therefore, it can be concluded that the data filtering removed cases which actually would cause larger errors in the prediction and hence helps to separate errors of the LUT model and errors caused by poor forecasts of COSMO-D2 to some extent.

Finally, even if the LUT model is able to deliver useful predictions for a some

part of the data and in particular for cases with very low variability, the large positive outliers and the missing skill to predict larger values of V cause the Pearson correlation between observation and prediction to be only on the order of 0.45. Consequently, the LUT model obviously is too simple to be able to derive small scale irradiance variability from coarse COSMO-D2 predictions and would therefore need to be significantly improved for scientific as well as operational use.

4.2. Random forest model

In this section the performance of the Random Forest (RF) model presented in Section 2.5 will be evaluated and discussed. Just as for the LUT model, the influence of the data quality filtering will be discussed and the impact on the prediction accuracy will be assessed.

4.2.1. RF validation - Histogram

As a first step of the evaluation the RF model built on the filtered dataset (see Section 3.4.6) in its final configuration (see Table 3.9) was validated on the training dataset using tenfold cross-validation (CV) and on the held out validation dataset. In the CV procedure the RF is trained and evaluated on ten data subsets whereby 90% of the data are used for training and 10% for validation in each step. Before the ten data subsets were selected the dataset was shuffled randomly as it was done in all CV cases during the model development (see 2.5). For the prediction of the validation dataset the RF was trained on the entire training dataset. The results of both validation procedures are shown in Figure 4.3 which is structured analogous to Figure 4.1 for the evaluation of the LUT model.

Graph 4.1a) depicts the result for the tenfold CV on the training dataset. A first feature of the graph is the high point density at low values of V which is due to the large fraction of low variability cases as already described in Section 4.1.3. Due to the bias correction described in Section 3.8.6 the skewness of the predictions relative to the 1:1 line could be reduced significantly (see fig. 3.35). However, there is still a clear overestimation of cases with very small variability. In cases of very large V, predictions are still mostly on the lower side of the 1:1 line. In particular for the overestimation of small variability values it has to be stressed that the color scale changes logarithmically with the number of predictions in a certain bin. Consequently, the vast majority of cases with small variability is predicted accurately with only a small fraction of predictions significantly overestimating V. It finally can be stated that the bias correction was able to correct the range dependent bias to a large extent but was not able to entirely remove it, resulting in an RMSE of 0.0398 for the cross-validation.

Looking at the prediction of the validation dataset in graph 4.3b) it is in close agreement with the results for the CV on the training dataset. It shows the same

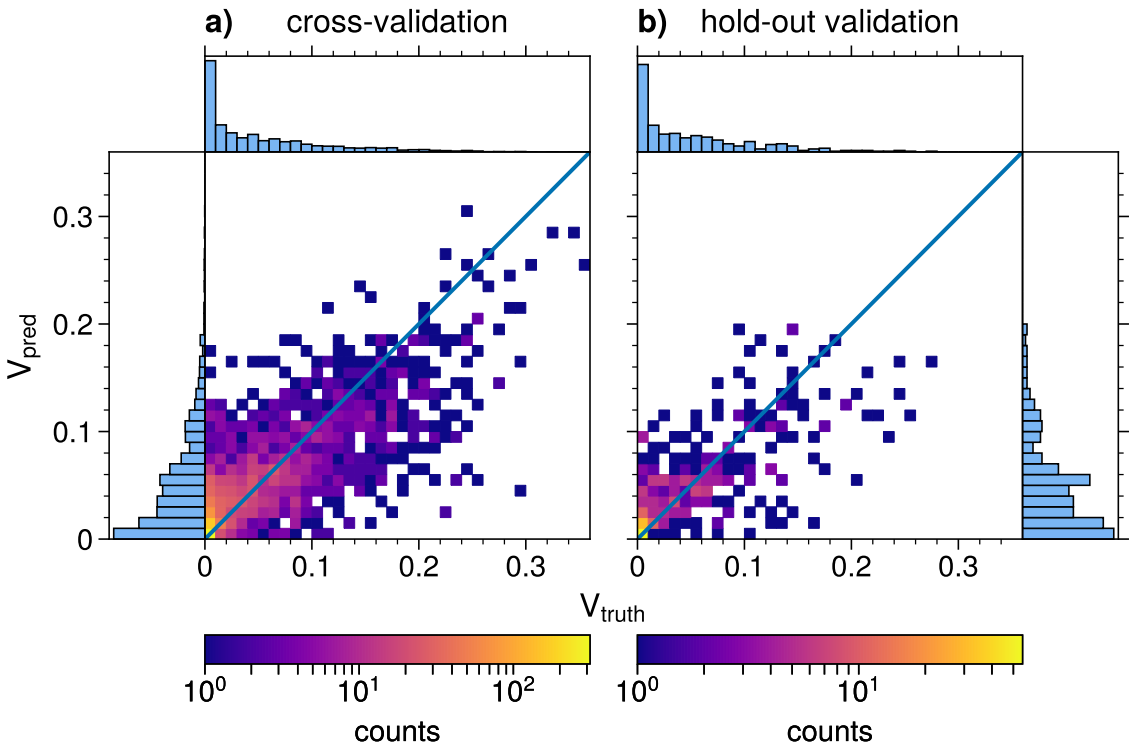

Figure 4.3.: 2D histograms of the true variability index (V_{truth}) plotted against the variability predicted by the RF model (V_{pred}). Graph a) shows the result for the prediction of the training dataset using tenfold cross-validation (CV) and graph b) the result for the validation dataset which was not used for model optimization (hold-out validation). The color scale changes logarithmically with the number counts in each pixel. 1D histograms on x- and y-axes show the relative frequency within each bin for V_{truth} and V_{pred} respectively. The blue lines in both histograms indicate the 1:1 line.

distribution of predictions with a tendency of overestimating small and underestimating large variability cases yielding a similar RMSE of 0.0390. Thereby the error reduction due to the application of the bias correction caused the same relative improvements as for the training dataset (see Section 3.8.6) with 0.8% for RMSE and 4.2% for MAE. As already mentioned for the LUT model, one would also expect the RF model to perform better on the training data as the model was optimized on this dataset. Because the overall size of the dataset is small (2468 samples), drawing 15% of the dataset randomly does not guarantee to yield a validation dataset which perfectly reflects the properties all data. Therefore, it can happen that the model performs better on the validation data if it holds cases with better predictability. However, the good performance is not a result of coincidence but shows that the model was build properly to generalize well on new data.

A comparison of the results for the LUT model and the RF model is shown in Figure 4.4. For both, RMSE and MAE, the RF model has a roughly 50% smaller error

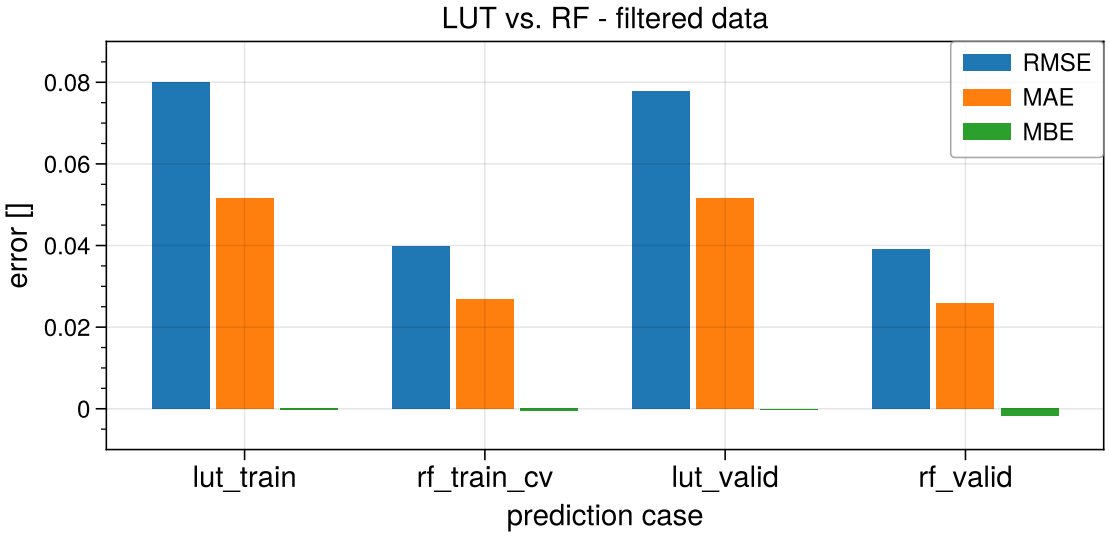

Figure 4.4.: RMSE (blue), MAE (orange) and MBE (green) for the LUT and RF model for the prediction of the training dataset (lut_train, rf_train_cv) and the validation dataset (lut_valid, rf_valid). The prediction of the RF on the training data was performed via cross validation (CV)

than the LUT model for the training as well as for the validation dataset. Concerning the MBE both models appear to be bias free on average, but as it has been shown for the RF model there can be a range dependent bias even if the overall bias is zero. Therefore, one has to be careful to draw major conclusions based on this metric. Another metric showing the advantage of the RF model is the Pearson correlation coefficient of 0.76 between measurement and prediction which is significantly better than for the LUT model with a correlation of 0.45. Finally, it can be noted that both models perform slightly better on the validation dataset which is probably a statistical effect due to the small size of the validation dataset as already mentioned before.

4.2.2. Influence of data filtering

In this thesis a lot of effort has been put into the separation of 'inaccurate' and 'accurate' COSMO-D2 forecasts to be able to reduce the impact of forecast errors from the NWP on the accuracy of the predicted GHI variability (see Section 3.4). To validate that the data filtering actual has an effect on the prediction accuracy of the RF model, the results of the RF based on the filtered dataset (RF_{filt}) were compared to the ones of a second RF. The second RF (RF_{nofilt}) was optimized, trained and evaluated on the unfiltered dataset (see Sections 3.4.6 and A.1). The results for the predictions of the respective training and validation dataset for RMSE, MAE and MBE are shown in Figure 4.5. As expected RF_{filt} performs significantly better in RMSE and MAE than RF_{nofilt} whereby the MBE is very close to zero in all cases. For the CV on the training dataset RMSE is reduced by 7.3% (0.0429 to 0.0398) and MAE by 11.6% (0.0304 to 0.269), for the

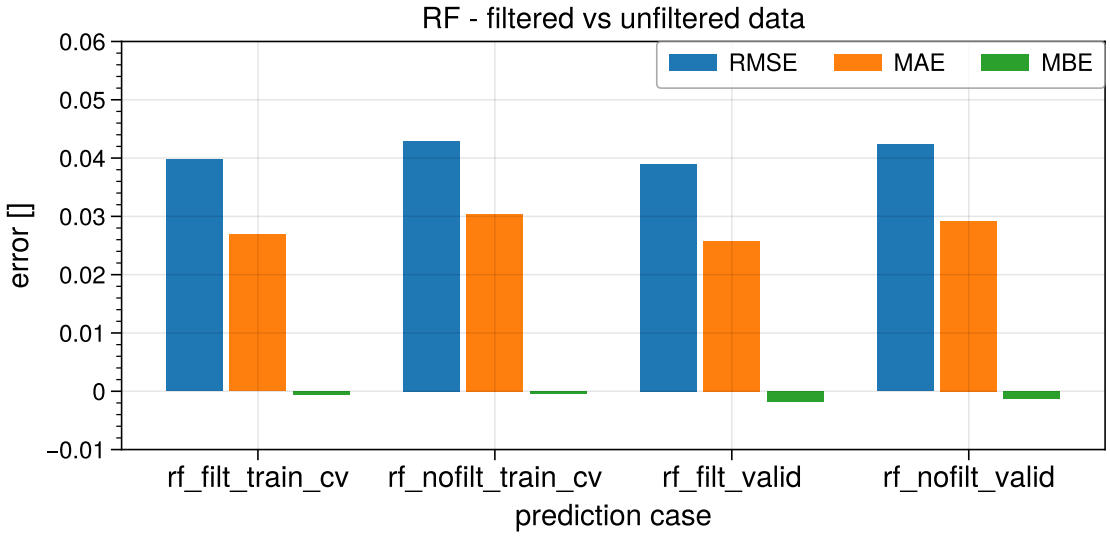

Figure 4.5.: RMSE (blue), MAE (orange) and MBE (green) for two different RF models. The first one was optimized, trained and evaluated on the filtered dataset (rf_filt) and the second one on the unfiltered dataset (rf_nofilt). Predictions are shown for the training dataset via cross validation (rf_filt_train_cv, rf_nofilt_train_cv) and for the validation dataset held out during optimization (rf_filt_valid, rf_nofilt_valid).

validation dataset the improvements are similar with 8% (0.0423 to 0.0390) for RMSE and 11.5% (0.0292 to 0.0258) for MAE. This outcome supports the idea behind the data filtering as the predictions based on all data are much less accurate as the ones only including filtered data. Furthermore, it shows that the approach for the filtering actually removes cases where the NWP provides inaccurate forecasts which are responsible for a degraded accuracy in the prediction of GHI variability.

From a scientific perspective it is a sensible approach to exclude inaccurate NWP forecasts to be able to separate errors in the input data (NWP data) from errors caused by the model which predicts GHI variability (e.g. RF). However, for operational use the variability prediction has to be available and as accurate as possible independent of the quality of the NWP forecast which is not known beforehand. To find out which of the RF models (RF_{filt} or RF_{nofilt}) would be the best choice for operational use, the RF_{filt} model was applied to the unfiltered validation dataset and the RF_{nofilt} to the filtered validation dataset. In the following this procedure will be called crossover-prediction. The first two cases of Figure 4.6 show the prediction error (RMSE, MAE, MBE) of RF_{filt} and RF_{nofilt} on the unfiltered validation dataset. As the RMSE for the prediction of RF_{filt} on the unfiltered validation dataset has an 11.9% larger RMSE (0.0474) and an even 13.1% larger MAE (0.0330) than RF_{nofilt} it is clear that RF_{nofilt} would be the model of choice in operational use.

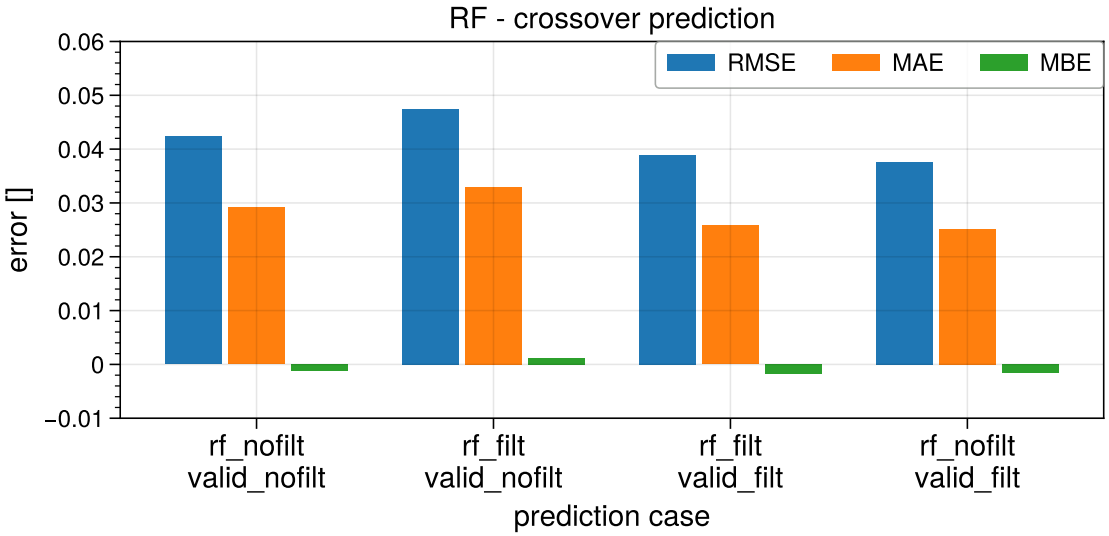

Figure 4.6.: RMSE (blue), MAE (orange) and MBE (green) for crossover-predictions of the two RF models rf_filt (RF_{filt}) and rf_nofilt (RF_{nofilt}) are shown. Thereby the model either predicted the unfiltered validation dataset (valid_nofilt) or the filtered validation dataset (valid_filt).

In addition to the application of RF_{filt} on the unfiltered validation dataset, a second experiment can be made in the scope of the crossover-prediction by applying RF_{nofilt} to the filtered validation dataset. This experiment is of particular interest as another reason for the data filtering was to avoid ambiguities in the dataset which could lead to an overall degraded precision of the variability prediction. An example for such an ambiguity would be if the NWP correctly predicts clear sky for one day but for a second clear sky day it predicts clouds which would cause significant GHI variability in reality. From these two cases the RF would learn that an actual variability index of zero can be either be associated with clear sky or with a cloudy sky which is simply wrong. Following this hypothesis, the RF trained on the unfiltered data may perform worse on the filtered data than the RF trained on the filtered data which contain fewer ambiguities. However, by looking at the last two cases in Figure 4.6, RF_{nofilt} actually performs better on the filtered data than RF_{filt} with the RMSE being reduced by 3.4% and the MAE by 2.7%. The implications of this result will be discussed in the next section.

4.2.3. Discussion of RF model results

The developed approach based on the RF technique shows a good performance for the prediction of the variability index V. In particular the range dependent bias which is a well known issue of RF could be corrected to a large degree. Furthermore, the model generalizes well to unseen data as the prediction accuracy on the validation dataset is almost the same as for cross-validation on the training dataset the RF was optimized on.

The comparison of the RF model based on the quality filtered dataset (RF_{filt}) with the RF built on the unfiltered dataset (RF_{nofilt}) showed that the filtering approach applied in Section 3.4 actually removes cases with bad predictability due to inaccurate COSMO-D2 forecasts as the RMSE for RF_{filt} is up to 8% smaller than for RF_{nofilt}. However, RF_{filt} performs worse than RF_{nofilt} when applied to the unfiltered validation dataset which makes the RF_{nofilt} model more suited for operational use. Even if this result may not be surprising as RF_{nofilt} was trained on a dataset which is 40% larger than the filtered dataset, it shows that the RF model is able to partly compensate for the errors in the COSMO-D2 forecast which is an important finding and a valuable property of the RF model. The second important finding of the crossover-prediction is that RF_{nofilt} performs better on the filtered validation dataset than RF_{filt}. From this result two important conclusions can be drawn: firstly, the presence of ambiguities in the unfiltered dataset does not reduce the prediction accuracy and secondly, more data, even if they are of low quality concerning the NWP forecast, lead to an improvement of the predictions. This means that the entire modelling approach would profit from a larger dataset independent of NWP forecast quality.

Now, after the results of the RF as well as the LUT model (see 4.1) were discussed in detail, both approaches can be compared. A general advantage of the LUT model is that it is mainly based on simulations which could be helpful when only few measurements are available. However, at the same time it is very expensive from a computational point of view as extensive 3D radiative transfer simulations based on high resolved LES simulations with a variety of cloud fractions are necessary. In contrast, the RF model only needs the input data from the NWP together with the measurements to fit a nonlinear relation between them. Here the disadvantage is that a large set of measurements is necessary to build a proper RF model. In the end the comparison of the prediction error clearly shows that the RF model (RF_{filt}) with an RMSE of 0.039 on the validation dataset outperforms the LUT model with an RMSE of 0.078. This difference in accuracy can be attributed to a large extent to the fact that the LUT model only relies on two parameters of COSMO-D2 (cloud speed and cloud fraction) and the purely geometric SZA which is obviously not enough to describe the complex relation between NWP forecast and GHI variability. This becomes particularly clear when comparing it to the RF model which needs 13 features (RF_{filt}) for an optimal prediction. Taking into account the effort as well as the performance of both presented models the RF approach is without doubt the best choice for the prediction of small scale GHI variability based on COSMO-D2 forecasts. Additionally, to the good accuracy and straight forward setup, the RF model delivers insights into the connection between variables derived from COSMO-D2 (or any other NWP model) and the small scale GHI variability by extracting the most predictive features during the optimization procedure (see Section 3.8.4).

Finally, a general issue concerning the composition of the measurement dataset needs to be discussed. As already described in the previous sections, the vast majority (75%) of data points is associated with very small GHI variability ($V < 0.1$). It is questionable if the distribution within the measurement data is a general feature or if this

is specific to the observed periods in autumn 2018 and summer 2019. The distribution of data points in between the maxima and minima in the dataset has influence on the ability of the model to predict situations with a certain variability. When only few high variability cases are contained in the training data the model is less good in prediction such cases for new data. Whether this behaviour needs to be corrected in a certain way or not depends on the specific application. In this dissertation the models were optimized to minimize the prediction error on the entire dataset, but there might be situations where predictions need to be optimized for a certain range of variability cases.

5. Discussion - comparison with literature

To put the results presented in this dissertation into a larger context, the findings will be discussed in the light of existing literature on GHI variability prediction. In Section 2.6 a brief overview of the two most relevant publications on this topic by Gristey et al. (2020) and Riihimaki et al. (2021) was given. The work by Gristey et al. (2020) deals with the prediction of the shape of the GHI histogram which is directly linked to GHI variability However, they use variables like cloud distance or cloud size from high resolved LES data which are not available from the NWP and hence a direct comparison of the results is not possible. The publication by Riihimaki et al. (2021) uses a similar variability metric derived from GHI measurements which is also predicted by means of a random forest (RF) model which makes it possible to compare their results to the findings of this dissertation.

The idea of the work by Riihimaki et al. (2021) is to develop a model to predict GHI variability using cloud information which is taken from in-situ measurements but could also be derived from NWP forecasts. Thereby they specify the variables to be used beforehand (cloud cover and cloud type) which is similar to the LUT model approach but fundamentally different from the RF model developed in this thesis where more than 120 variables from the NWP model (COSMO-D2) were used in the first step. This was done to allow RF model to learn which variables are the most predictive ones. It is very likely that the approach presented in this study leads to better results for most applications as defining the input parameters based on a priori knowledge removes the possibility to find predictive variables one was not aware of initially. A second important difference can be found in the optimization procedure used by Riihimaki et al. (2021). Similarly to this thesis they use a cross-validation for hyperparameter tuning but no bias correction method was applied even though their results show a very pronounced range dependent bias (see Figure 7 of their paper) which was also encountered here (see figs. 3.33, 3.35).

Finally, a comparison of the resulting prediction error can be made. Even if the variability metric $\sigma(\Delta ET)$ in Riihimaki et al. (2021) is similar to the variability index V with a nearly identical range of values (between 0 and 0.35) the comparison has to be interpreted carefully. For their RF model Riihimaki et al. (2021) state a mean squared error MSE 0.04 (tab. 2) which can be translated to the RMSE by taking the square root which yields a value of 0.063. With an RMSE on the validation dataset of 0.0390 for the RF model based on the quality filtered data and an RMSE of 0.0423 for the RF model using the unfiltered dataset, the performance seems to be better than

for the model developed by Riihimaki et al. (2021). As mentioned before, this should be interpreted with care as not the exact same quantities are compared. However, there are solid arguments why the model developed in this dissertation should perform better. Probably the most important reason for the better performance is that the most predictive variables were extracted from a large ensemble by the RF itself rather than providing predefined variables. Thus, it can be ensured that enough variables are available to capture the complexity of the problem and that all variables in the final set of features actually have a positive impact on the accuracy of the RF predictions. It is likely that the RF model of Riihimaki et al. (2021) has not enough variables to deliver better predictions as there is only one continuous variable (cloud cover) and seven binary cloud class variables in the dataset. Furthermore, variables like wind speed which has been shown to have large influence on GHI variability are not considered. A last argument is connected to the optimization procedure of the RF. The strong range dependent bias in the results of Riihimaki et al. (2021) which can for instance be seen in Figure 7 of their paper, could likely be reduced by applying a bias correction technique like the one by Zhang and Lu (2012) used in this dissertation.

The presented comparison to state-of-the-art techniques for the prediction of GHI variability clearly showed that advances in the prediction accuracy were achieved by using the RF model not only for prediction but also for the selection of the most predictive variables. Furthermore, the elaborated optimization procedure including a multistage feature selection, hyperparameter tuning and bias correction further reduced the prediction error. Finally, it could be shown that it is possible to make accurate forecasts of GHI variability for time horizons of more than 6 hours ahead based on operational NWP forecasts of COSMO-D2 which is a major advance compared to existing literature.

6. Summary and outlook

In this dissertation two types of models were developed to predict the variability of the global horizontal irradiance (GHI) on small temporal (one minute) and spatial (25 m) scales for time horizons beyond 6 hours based on numerical weather prediction (NWP) forecasts of the German Weather Service COSMO-D2 model. For model development and validation a unique dataset of GHI measurements was used from which the ground truth GHI variability according to the variability index V (Coimbra et al., 2013) was derived. The measurements were recorded in the region of Kempten (Allgäu) within the MetPVNet research project (Meilinger et al., 2021). To be able to separate errors originating from the NWP forecast from errors in the variability predictions, COSMO-D2 data were quality filtered based on the comparison of hourly average GHI between model and measurement. Furthermore, the dataset was separated into a training dataset for model development and tuning and a validation dataset which was exclusively used for the final model evaluation.

The first model for the variability prediction is based on a lookup-table (LUT) approach which uses information from high resolved spatial fields of simulated GHI. The GHI fields were derived using the 3D radiative transfer model MYSTIC (Mayer, 2009) for solar zenith angles (SZA) between 20 and 80° in combination with cloud fields from high resolved ($\Delta x = \Delta y = \Delta z = 25\ m$) large eddy simulations (LES) of the UCLA-LES model, which exhibit cloud fractions between 12% and 100%. From these GHI fields the variability index was derived via spatial sampling through the domain assuming cloud advection with different speeds yielding a three-dimensional table with SZA, cloud fraction and cloud advection speed as dimensions. To predict V based on the COSMO-D2 forecast, cloud fraction and cloud advection speed were derived from the NWP output. Both parameters were tuned to reduce the prediction error on the previously defined training dataset. The cloud fraction was optimized by excluding the contribution of clouds above an altitude threshold of 5000 m to avoid the impact of cirrus clouds and the advection speed was linearly scaled to take the effect of convection during the advection process into account.

The second model uses the random forest (RF) machine learning technique initially proposed by Breiman (2001). In the first step 121 variables which potentially could contain useful information about GHI variability were used as input for the RF. These variables were either taken directly from the COSMO-D2 output or they were derived from the NWP variables, for instance by calculating spatial averages or standard deviations (feature engineering). Subsequently, a recursive feature elimination (RFE) was performed based on the permutation feature importance (PFI) metric using a tenfold

cross-validation (CV) on the training dataset. The RFE provided an importance ranking of all input variables which was used to eliminate features without prediction skill from the dataset. To determine a final set of features for the RF model, a preliminary set was selected by successively adding features according to the ranking of the RFE and cutting of as soon as the root-mean-square error converged. This preliminary set was refined in a second step by removing highly correlated features according to the Pearson correlation coefficient. Using the derived set of features the model was further optimized by tuning the hyperparameters of the RF with a grid-search procedure, again in combination with tenfold CV on the training dataset. In the final step a bias correction according to the BC1 method by Zhang and Lu (2012) was applied to reduce the well known range dependent bias of random forests. To be able to assess the impact of the data filtering on the prediction accuracy and to test the ability of the RF model to compensate for errors in the NWP forecast the described procedure was also applied to an RF model based on the unfiltered dataset.

Both of the developed models, the LUT and the RF model, were evaluated on the validation dataset in the final step. The evaluation of the LUT model on the filtered validation dataset showed an acceptable prediction skill for small levels of GHI variability. However, high variability cases were predicted with poor accuracy. Additionally, V was overestimated in some cases which leads to an overall large RMSE of 0.078 and a Pearson correlation coefficient of 0.45. One part of the reason for this outcome can be attributed to LES cloud fields which are likely not representative enough to describe GHI variability for different cloud types or more complex multi layer cloud situations. However, the main cause of the large prediction error is that the full complexity of the relation between low resolution NWP forecasts and rapid small scale GHI fluctuations can not be captured by a model with only three parameters as it is the case for the LUT model.

In contrast to the empirical LUT model with predefined parameters, the RF model is able to use an arbitrary number of variables and extract the most predictive ones from them. To reach optimal accuracy 13 features were selected for the RF based on the filtered dataset ($\mathrm{RF_{filt}}$) with the top three being the standard deviation of direct horizontal irradiance, the average total liquid water path (both calculated over 5x5 grid boxes) and the zonal wind speed at 200 hPa. For the second RF model based on the unfiltered dataset ($\mathrm{RF_{nofilt}}$) the best 19 features were included into the final dataset. Both RF models perform significantly better on the respective validation dataset than the LUT model with an RMSE of 0.0390 (filtered) and 0.0423 (unfiltered). This outcome shows to things: firstly, actually more than three parameters are necessary to obtain a good accuracy for the prediction. Secondly, the data filtering procedure obviously removed COSMO-D2 forecasts with large errors as $\mathrm{RF_{nofilt}}$ exhibits an 8.5% larger error on the validation dataset than $\mathrm{RF_{filt}}$.

To test which model would perform better in operational use, where a prediction has to be made regardless of the quality of the NWP forecast $\mathrm{RF_{filt}}$ was applied to the unfiltered validation dataset which yielded an RMSE of 0.0474. As this error is

12% larger than the one of RF_{nofilt}it can be concluded that RF_{nofilt} performs better in operational use and that the RF is able to compensate for larger errors in the COSMO-D2 forecasts of the unfiltered dataset. Performing the reverse experiment by using RF_{nofilt} to predict the filtered validation dataset it can be investigated whether errors and resulting ambiguities in the unfiltered COSMO-D2 predictions degrade the accuracy of RF_{nofilt} for the filtered data with accurate COSMO-D2 predictions. It turned out that the RMSE for RF_{nofilt} on the filtered validation dataset is actually 3.4% smaller compared to the prediction of RF_{filt} on the same data. From this behaviour it can be concluded that possible ambiguities caused by errors in the COSMO-D2 forecast have no negative impact on the prediction performance of (RF_{nofilt}). Furthermore, it can be stated that a larger dataset would have a beneficial impact on forecast accuracy even if those data contain COSMO-D2 forecasts of low quality. Consequently, for an operational use of the variability forecast the RF model should always be trained on all available data to obtain the best possible prediction independent of the quality of the NWP forecast.

Of course, there is still room for improvements and optimizations in the prediction of GHI variability which should be investigated in future studies. Starting at the very basis of the model, there may be additional features with good predictability which could be derived from the NWP forecast and further reduce the prediction error. For instance one could calculate the cloud fraction not only around the location of the measurement but rather in the direction of the sun covering an area up- and downstream of the atmospheric flow because clouds in this area are responsible for fluctuations in GHI at the considered location. Furthermore, it would be desirable, at least from a scientific point of view, to get an estimate for the uncertainty of the variability forecast which could be achieved by applying the RF model to NWP ensemble forecasts.

Going away from these specific issues to a more general view it is a very important step to generalize the prediction of GHI variability from the specific region considered in this dissertation (Kempten) to a national or even continental domain. Therefore, high temporal resolution measurements of irradiance with good spatial coverage are necessary to capture the peculiarities of individual regions. Currently, GHI measurements of the German Weather Service (DWD) have a good spatial coverage with about 120 stations distributed over entire Germany, however only 10-minute average values of GHI are recorded. This resolution does not cover the important short term fluctuations on timescales of one minute or less which were considered in this thesis and can therefore not be used for the according variability forecast. Nevertheless, there is a possibility to obtain the required information by using rooftop PV systems as radiation sensor which would have major advantages compared to measurements from weather stations: firstly PV power can be sampled at the high temporal resolutions needed, and secondly the spatial coverage is much better than the one of available GHI measurements. Additionally, coverage is proportional to the installed PV capacity of a region which means that more data at higher resolutions are available where accurate predictions are particularly important Although this idea sounds straight forward as the power output of PV cells is somehow proportional to the incoming solar radiation, there

are many effects to be considered like the influence of module temperature on efficiency, the module degradation with time or the exact orientation of the module to name only a few. However, there have been promising methods developed by Buchmann (2018) which were refined in the scope of the MetPVNet project (Meilinger et al., 2021).

In the context of the extension of domain size in combination with high temporal and spatial resolution datasets the amount of data drastically increases. Consequently, it might be necessary to switch from the RF model to artificial neural networks (ANN) which are able to efficiently handle very large amounts of data and also provide additional tuning opportunities. Finally, there is one more important step to refine forecast of GHI variability for operational use: The adaption of the forecast to the requirements of the customer. For instance one could imagine that the accurate prediction of very high variability situations is of much higher interest then medium or low variability. In this case the model could be optimized by weighting errors according to their importance during training which should cause better predictions for situations of higher importance.

As a final remark it can be stated that the research goal formulated in the beginning of this dissertation was successfully accomplished as an accurate method was developed to predict the GHI variability on small spatial and temporal scales beyond time horizons of 6 hours using forecasts of the COSMO-D2 NWP model. Although more work is required to develop the method towards an operational product, the presented work is an important tool for the future prediction of PV power fluctuations and the availability of reactive power to ensure an efficient operation of large PV capacities within the power grid and hence supports the transformation into a 100% renewable energy future.

Appendices

A. Appendix

A.1. RF optimization based on unfiltered data

Just as it was done in sections 3.8.3 and 3.8.5 of Chapter 3, a final feature set as well as the optimal hyperparameters will be determined for the RF based on all available data in this section. Furthermore, the bias correction according to the BC1 method of Zhang and Lu (2012) will be applied as described in Section 3.8.6. The term 'all available data' refers to the dataset which was already filtered for data gaps and dew contamination as described in Section 3.4 but not filtered for errors in the COSMO-D2 forecast. As the procedure is analogous to Chapter 3 the individual steps will be described less detailed.

After the recursive feature elimination (RFE) was performed with the same setup as described in Section 3.8.3 the top 30 features were successively added to the dataset and evaluated via tenfold cross-validation (CV) on the training dataset. Figure A.1 shows the RMSE as a function of the number of features added in the top row and the according reduction in RMSE achieved by adding a feature to the dataset. To have a closer look on the impact of the less important features at the low ranks the y-axis of the lower panel does not cover the full range of values. After feature number 12 was added to the dataset additional features mostly yield either small or no improvements at all. As these two cases are occurring in an alternating way it was decided to take all 30 features in the preliminary dataset.

In the next step the correlation matrix for all 30 features was calculated using the Pearson correlation coefficient to exclude highly correlated features from the dataset. The correlation matrix depicted in Figure A.2 in particular shows high correlation of the zonal wind component (u) at different pressure levels and between temperature at different altitudes. For better visualization of the correlations between individual variables a hierarchical clustering was performed like it was done in Chapter 3 for the RF based on the filtered dataset. The according dendrogram is shown in Figure A.3. From the highly correlated zonal wind variables, u_{500} and u_{200} were kept in the dataset due to their better ranks in the RFE while u_{250}, u_{300} and u_{600} were removed. The high correlation between two-meter temperature (T_{2M}) and the temperature at 950 hPa (T_{950}) is due the fact that T_{950} is extrapolated from T_{2M} as the elevation in the region around Kempten is typically larger than 600 m which is associated with a surface pressure of less than 950 hPa. Furthermore, T_{2M} shows a close correlation with T_{850} which is often used by meteorologists to infer the temperature close to the surface because T_{850} is an indicator for the Temperature of the advected air masses rather and is not influenced by radiative heating or cooling of the surface. Therefore, it was decided to keep only T_{850}

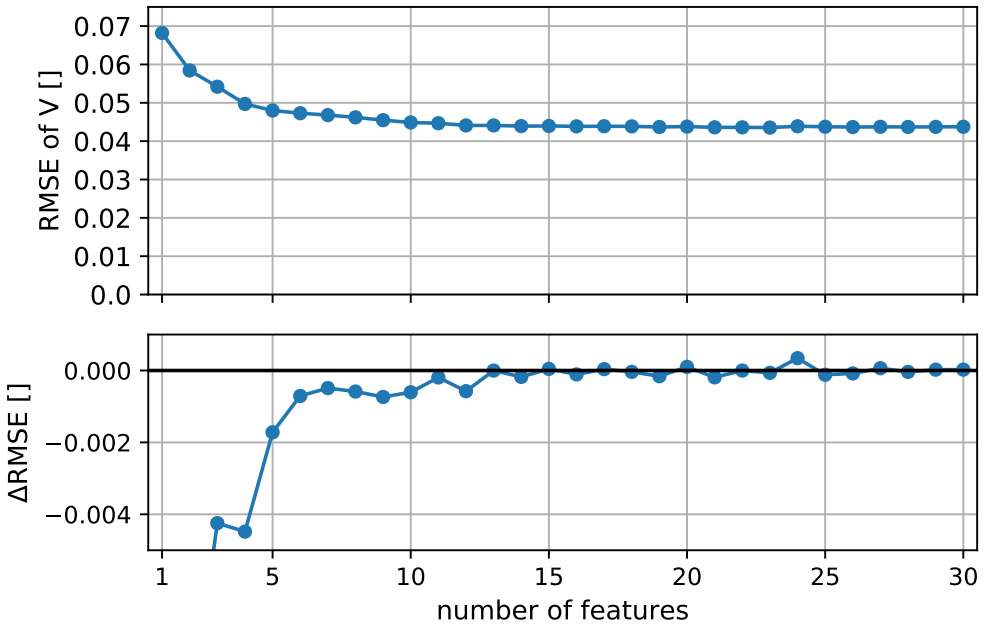

Figure A.1.: Upper panel: RMSE of the variability index V calculated via tenfold cross-validation on the training dataset using all data (unfiltered dataset) as a function of the number of features added to the dataset. Lower panel: change in RMSE caused by adding a certain feature to the dataset starting with the second feature.

and remove T_{950} and T_{2M} from the dataset. The last high correlation is found between the average vertical extent of convective subgrid-scale clouds ($\overline{\Delta z_{\mathrm{sg,conv}}}$) and their average liquid water path ($\overline{lwp_{\mathrm{sg,conv}}}$). This high correlation is not surprising as the liquid water path is the vertical integral of the liquid water content (LWC) and hence directly related to the vertical thickness of the clouds. According to the better ranking $\overline{lwp_{\mathrm{sg,conv}}}$ is kept in the dataset while $\overline{\Delta z_{\mathrm{sg,conv}}}$ is removed from it. In total six of the top 30 features were removed due to high correlation reducing the remaining dataset to 24 features.

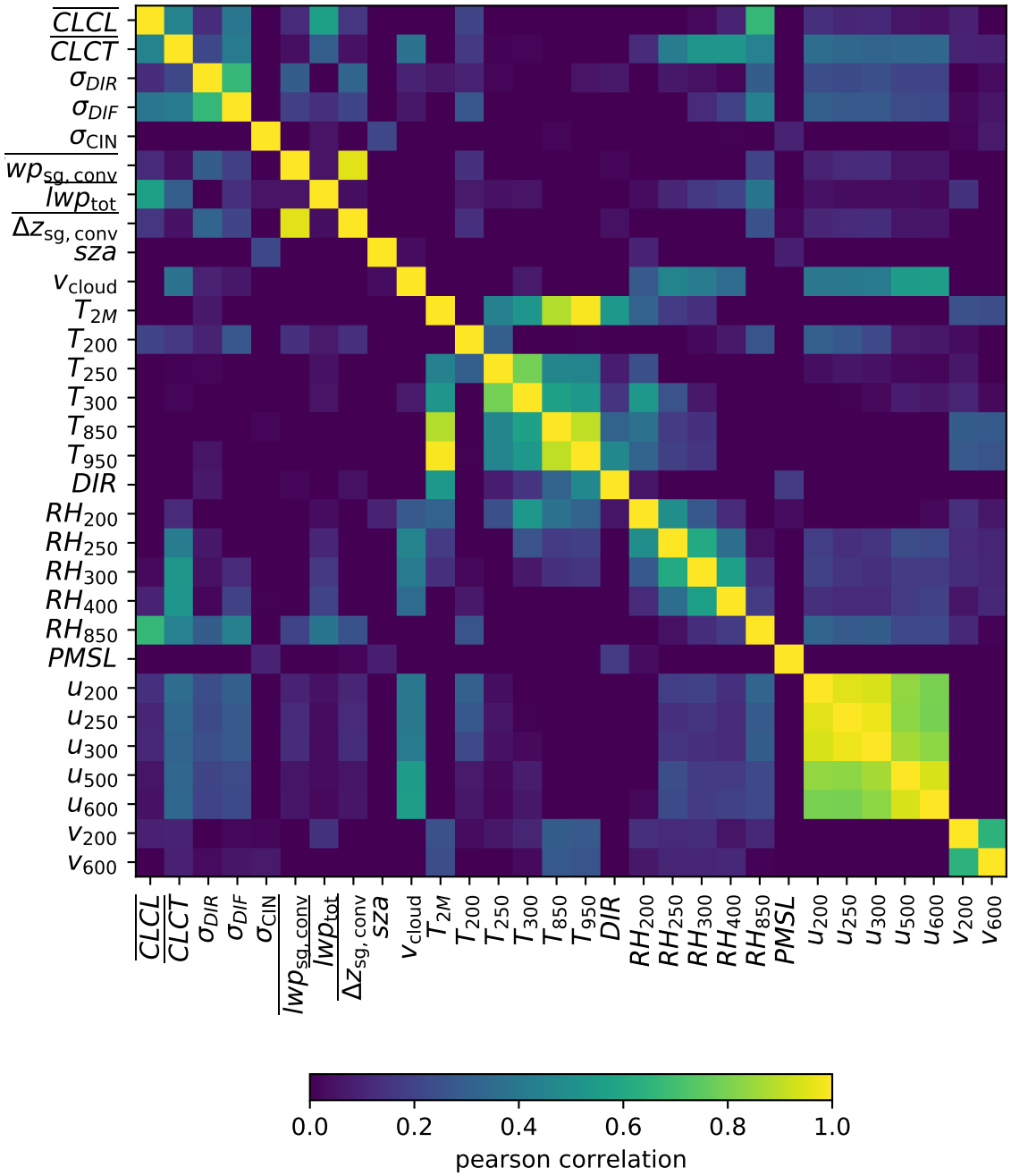

Figure A.2.: Pearson correlation matrix for the top 30 features of the RF model based on the unfiltered data calculated on the training dataset.

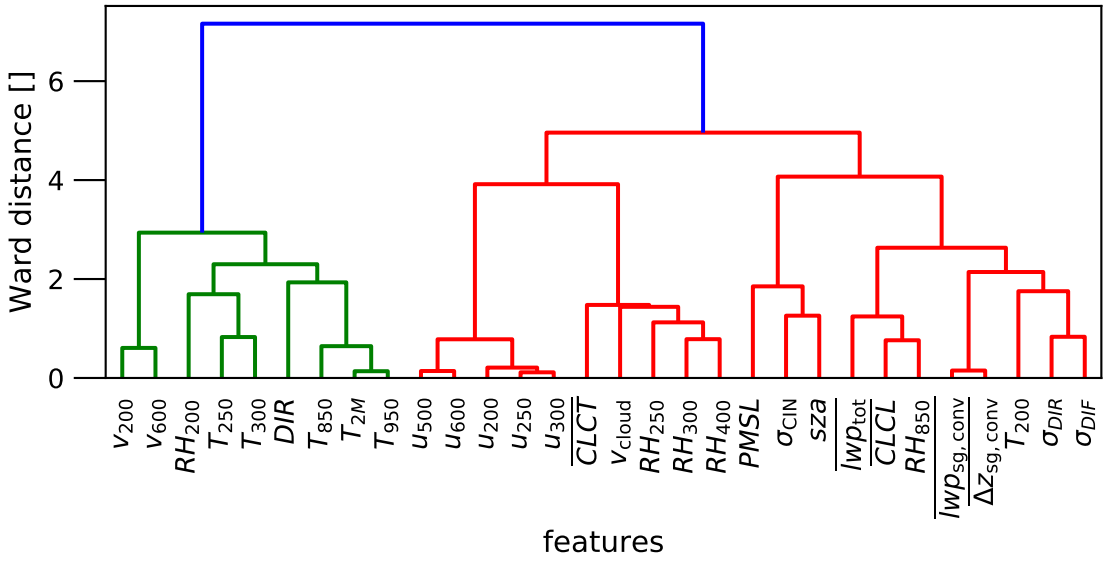

Figure A.3.: Dendrogram visualizing the result of the hierarchical clustering applied to the correlation matrix shown in Figure A.2. As distance metric Ward's method was used. The smaller the distance between individual features or clusters of multiple features the higher their correlation.

In the last step the performance of the RF was again evaluated as a function of the remaining 24 features which were successively added to the dataset. The results of the feature adding experiment are shown in Figure A.4 where the upper panel shows the RMSE of the RF prediction and the lower panel the reduction in RMSE by caused by adding a certain feature. The first twelve features (up to *sza*) all lead to an error reduction while the following features either have no effect or only slightly reduce the RMSE. However, feature number 20 ($\overline{lwp_{\mathrm{sg,conv}}}$) leads to a significant increase of the RMSE. Therefore, it was decided to keep the first 19 features up to v_{200} in the final set of features and remove the remaining five features from the dataset. A list of the final 19 features including a description can be found in Table A.1. The final dataset for the RF based on all data contains almost all 13 features from the RF based on the filtered dataset and additionally the relative humidity at different pressure levels, the spatial variability of the convective inhibition and the solar zenith angle. It can be assumed that these additional variables help to improve the variability prediction in cases where the COSMO-D2 forecasts has a large error.

After the final set of features had been determined, the hyperparameters of the RF were optimized starting with the number of trees as it was done in Section 3.8.5. The optimal number of trees was determined by evaluating the RF as a function of the number of trees using tenfold cross-validation on the training dataset leaving all other hyperparameters unchanged. According to the graph in the upper panel of Figure A.5 the RMSE stabilizes at around 500 trees with only very small improvements towards

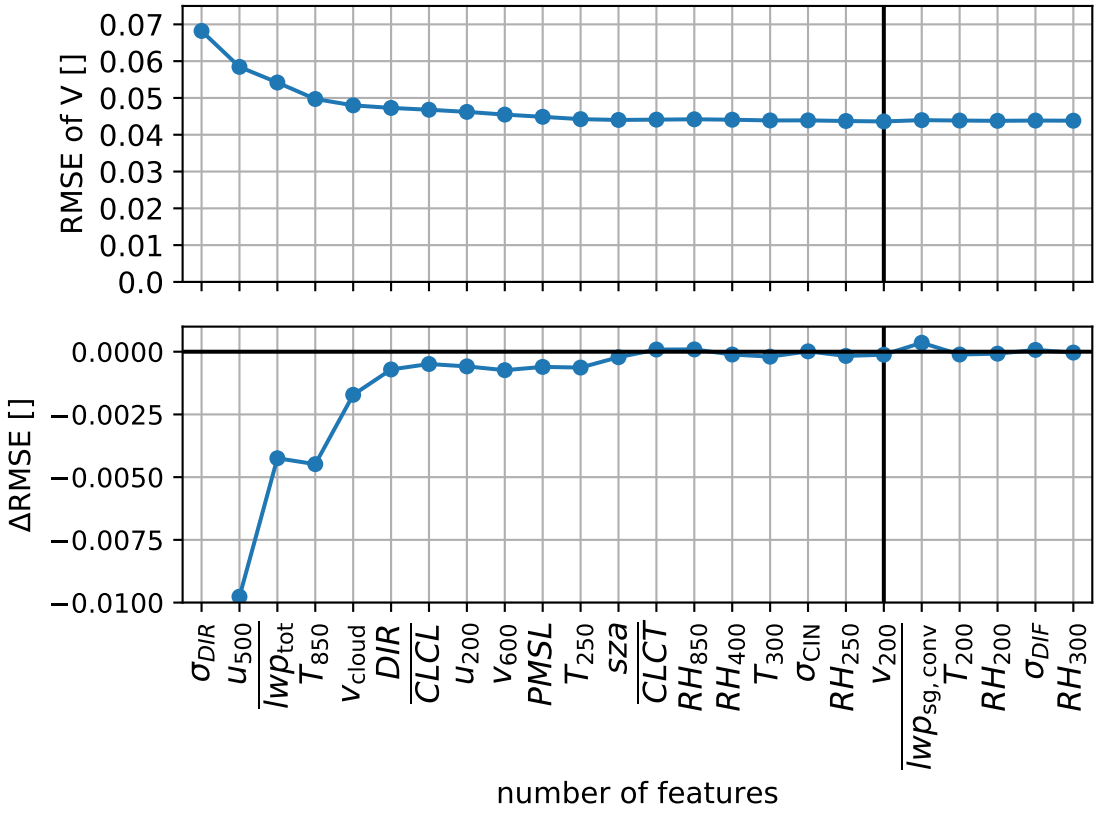

Figure A.4.: Upper panel: RMSE of the RF model prediction on the (unfiltered) training dataset, calculated via tenfold cross validation, as a function of the remaining top 24 features after removal of six correlated features. The features were added according to their ranking determined by the recursive feature elimination. Lower panel: change in RMSE caused by adding a certain feature to the dataset starting with the second feature. The vertical black line indicates the last feature included in the final dataset.

larger numbers of trees. To have a closer look at the behaviour of the RMSE the lower panel of Figure A.5 shows a detail of the upper panel curve with higher resolved y-axis. A local minimum of the RMSE can be identified at 700 trees before the curve starts slowly decreasing again at about 1000 trees. As decision on the number of trees is always a trade-off between accuracy and performance, 700 trees (red dotted line) were chosen for the RF as the RMSE is only 0.15% larger than for 2000 trees which are almost three times more trees.

Finally, the hyperparameter tuning was performed applying the GridSearch procedure with cross-validation on the training dataset fixing the number of trees to 700 and varying the hyperparameters identically to the RF based on the filtered dataset. See Table 3.8 for the tuned hyperparameters and the tested values. In Table A.2 the results

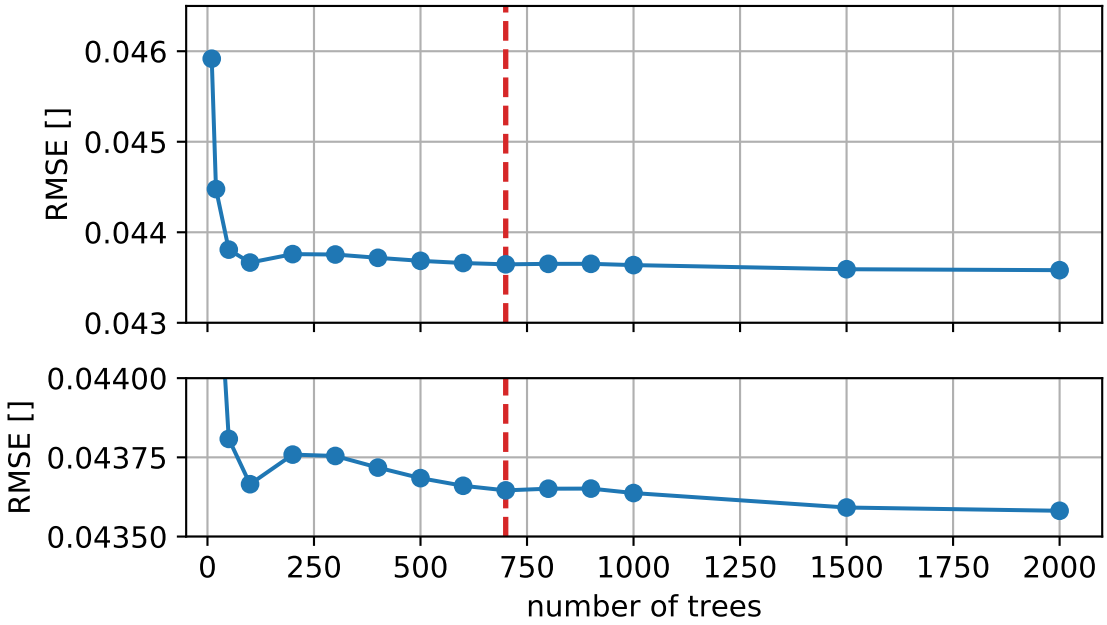

Figure A.5.: Upper panel: RMSE of the RF model prediction on the (unfiltered) training dataset, calculated via tenfold cross validation, as a function of the number of trees. The dataset holds the selected 19 features of the final dataset. The dashed red line indicates the number of trees chosen for the final model setup. Lower panel: detail of the upper panel to visualize the small differences in RMSE of adjacent points.

of the tuning can be found below the heading 'Optimized RF Hyperparameters'. The hyperparameter tuning reduced the RMSE from 0.04365 to 0.04349 which is a relative improvement of 0.36%.

As the last step of the optimization procedure the bias correction according to the BC1 method of Zhang and Lu (2012) was applied. For a detailed description of the method see Section 3.8.6 in Chapter 3. The bias correction improved the RMSE by 1.34% from 0.04349 to 0.04291. A summary of the final configuration of the RF based on all available data can be found in Table A.2.

Final feature set	
Parameter	Description
σ_{DIR}	standard deviation of direct horizontal solar irradiance within 5x5 grid boxes around location of measurement
$\overline{\mathrm{lwp_{tot}}}$	total (grid scale + subgrid scale) liquid water path averaged over 5x5 grid boxes around location of measurements
$\mathrm{u_{200}}$ $\mathrm{u_{500}}$	zonal windspeed at 200 and 500 hPa pressure level
$\mathrm{v_{cloud}}$	cloud advection speed estimated as described in equation 3.7.2
$\mathrm{v_{200}}$ $\mathrm{v_{600}}$	meridional windspeed at 200 and 600 hPa pressure level
$\mathrm{T_{300}}$ $\mathrm{T_{400}}$ $\mathrm{T_{850}}$	temperature at 300, 400 and 850 hPa pressure level
DIR	direct horizontal irradiance at the closest gridpoint
$\overline{\mathrm{CLCL}}$	cloud fraction of low clouds ($p < 800\ hPa$) averaged over 5x5 grid boxes around the location of measurement
PMSL	atmospheric pressure at mean sea level
$\mathrm{RH_{250}}$ $\mathrm{RH_{400}}$ $\mathrm{RH_{850}}$	relative humidity at 250, 400 and 850 hPa pressure level
σ_{CIN}	standard deviation of the Convective INhibition within 5x5 grid boxes around the location of measurement
sza	solar zenith angle

Table A.1.: Description of the 19 features contained in the final dataset of the RF based on all available data.

Optimized RF Hyperparameters	
max_features	sqrt
max_samples	None (number of all data points)
min_samples_leaf	1
min_samples_split	5
Default RF Hyperparameters	
min_weight_fraction_leaf	0.0
max_leaf_nodes	None
min_impurity_decrease	0.0
ccp_alpha	0.0
Other RF Parameters	
criterion (splitting)	squared_error
bootstrap	True
oob_score	False
n_jobs	-1
random_state	42
warm_start	False
Bias Correction	
BC1 method according to Zhang and Lu (2012)	
Selected Features	
σ_{DIR}, u_{500}, $\overline{\mathrm{lwp}_{\mathrm{tot}}}$, T_{850}, $\mathrm{v}_{\mathrm{cloud}}$, DIR $\overline{\mathrm{CLCL}}$, u_{200}, v_{600}, PMSL, T_{250}, sza, $\overline{\mathrm{CLCT}}$, RH_{850} RH_{400}, T_{300}, σ_{CIN}, RH_{250}, v_{200}	

Table A.2.: Configuration of the final RF model developed based on entire unfiltered dataset. The first three sections of the table list the options chosen for the scikit-learn RandomForestRegressor and the last section contains the 19 selected features in descending order of their importance.

A.2. List of all variables from COSMO-D2

The following table lists all 121 variables which were initially used for the development of the random forest model. The table is structured by different headlines, starting with the explanation of important abbreviations. Furthermore, it is divided into 'original COSMO-D2 variables' which were taken directly from the model without further processing and 'engineered features' which were calculated from the original COSMO-D2 variables. Finally, variables which are not related to COSMO-D2 are summarized under the heading 'further variables'. The parameter names are written in the way they were retrieved from the COSMO-D2 dataset and engineered features are written in the same fashion. For better readability some parameter names were rewritten in the text of this thesis using the options provided by LATEX (subscripts etc.). The description of these differently written abbreviations can be found for instance in tables 3.7 and A.1.

Parameter Symbol	Description	Unit
Abbreviations		
_gp	Value of variable at gridpoint closest to measurement location	-
plev	Variable defined on the following pressure levels: 1000, 975, 950, 850, 700, 600, 500, 400, 300, 250, 200	hPa
_mean	Mean value of the variable calculated over 5x5 gridpoints around the measurement location	-
_std	Standard deviation of the variable calculated over 5x5 gridpoints around the measurement location	-
_gs_	Value of the variable on resolved scales (grid-scale)	-
_sg_	Value of the variable on unresolved scales (subgrid-scale)	-
tot	Sum of grid-scale and subgrid-scale contributions for the considered variable	-
COSMO-D2 original variables		
2t_gp	Temperature two meters above surface	K
2r_gp	Relative humidity two meters above surface	%
prmsl_gp	Atmospheric pressure at mean sea level	hPa
CLCL_gp	Cloud fraction of low clouds (1000-800 hpa)	%
CLCM_gp	Cloud fraction of medium high clouds (800-400 hpa)	%
CLCH_gp	Cloud fraction of high clouds (<400 hPa)	%
CLCT_gp	Total cloud fraction	%
CAPE_ML_gp	Mixed Layer Convective Available Potential Energy	JK^{-1}
CIN_ML_gp	Mixed Layer Convective Inhibition	JK^{-1}
u_plev_gp	Zonal wind component on pressure levels	ms^{-1}
v_plev_gp	Meridional wind component on pressure levels	ms^{-1}
t_plev_gp	Temperature on pressure levels	K
r_plev_gp	Relative humidity on pressure levels	%

COSMO-D2 engineered variables		
SWDIR_S_gp	Hourly average of short wave direct irradiance at the surface	Wm^{-2}
SWDIR_S_mean	Spatial mean of hourly averaged short wave direct irradiance at the surface	Wm^{-2}
SWDIR_S_std	Spatial standard deviation of hourly average short wave direct irradiance at the surface	Wm^{-2}
SWDIFD_S_gp	Hourly average of short wave diffuse downward irradiance at the surface	Wm^{-2}
SWDIFD_S_mean	Spatial mean of hourly averaged short wave diffuse downward irradiance at the surface	Wm^{-2}
SWDIFD_S_std	Spatial standard deviation of hourly averaged short wave diffuse downward irradiance at the surface	Wm^{-2}
v_hor_plev_mean	Mean horizontal wind speed on pressure levels	ms^{-1}
CLCL_mean	Spatial mean cloud fraction of low clouds (1000-800 hpa)	%
CLCL_std	Spatial standard deviation for cloud fraction of low clouds (1000-800 hpa)	%
CLCM_mean	Spatial mean cloud fraction of medium high clouds (800-400 hpa)	%
CLCM_std	Spatial standard deviation for cloud fraction of medium high clouds (800-400 hpa)	%
CLCH_mean	Spatial mean cloud fraction of high clouds (<400 hPa)	%
CLCH_std	Spatial standard deviation for cloud fraction of high clouds (<400 hPa)	%
CLCT_mean	Spatial mean of total cloud fraction	%
CLCT_std	Spatial standard deviation of total cloud fraction	%
clc_con_gp	convective subgrid-scale cloud fraction	%
clc_con_mean	Spatial mean of convective subgrid-scale cloud fraction	%
clc_con_std	Spatial standard deviation of convective subgrid-scale cloud fraction	%
cld_depth_con_gp	Vertical thickness of convective subgrid-scale clouds	m
cld_depth_con_mean	Spatial mean of vertical thickness of convective subgrid-scale clouds	m
cld_depth_con_std	Spatial standard deviation of vertical thickness of convective subgrid-scale clouds	m
tcolr_gp	Total column integrated rain	kgm^{-2}
tcolr_std	Spatial standard deviation of total column integrated rain	kgm^{-2}
lwp_gs_gp	Grid-scale liquid water path	kgm^{-2}
lwp_gs_mean	Spatial mean grid-scale liquid water path	kgm^{-2}
lwp_gs_std	Spatial standard deviation grid-scale liquid water path	kgm^{-2}
lwp_sg_con_gp	Liquid water path of convective subgrid-scale clouds	kgm^{-2}

lwp_sg_con_mean	Spatial mean of liquid water path of convective subgrid-scale clouds	kgm^{-2}
lwp_sg_con_std	Spatial Standard deviation of liquid water path of convective subgrid-scale clouds	kgm^{-2}
lwp_tot_gp	Total liquid water path	kgm^{-2}
lwp_tot_mean	Spatial mean of total liquid water path	kgm^{-2}
lwp_tot_std	Spatial standard deviation of total liquid water path	kgm^{-2}
iwp_gs_gp	Grid-scale ice water path	kgm^{-2}
iwp_gs_mean	Spatial mean grid-scale ice water path	kgm^{-2}
iwp_gs_std	Spatial standard deviation grid-scale ice water path	kgm^{-2}
iwp_sg_con_gp	Ice water path of convective subgrid-scale clouds	kgm^{-2}
iwp_sg_con_mean	Spatial mean of ice water path of convective subgrid-scale clouds	kgm^{-2}
iwp_sg_con_std	Spatial Standard deviation of ice water path of convective subgrid-scale clouds	kgm^{-2}
iwp_tot_gp	Total ice water path	kgm^{-2}
iwp_tot_mean	Spatial mean of total ice water path	kgm^{-2}
iwp_tot_std	Spatial standard deviation of total ice water path	kgm^{-2}
sc_flag	Binary flag indicating subgrid-scale shallow convection for a gridbox (0: no , 1: yes)	-
box_sc_flag_mean	Spatial mean of binary flag for subgrid-scale shallow convection for a gridbox (or in other words: fraction of gridboxes containing subgrid-scale convective clouds)	-
cf_4000	Total cloud fraction up to an altitude of 4000 m averaged over 5x5 gridboxes	%
cf_5000	Total cloud fraction up to an altitude of 5000 m averaged over 5x5 gridboxes	%
cf_6000	Total cloud fraction up to an altitude of 6000 m averaged over 5x5 gridboxes	%
cf_7000	Total cloud fraction up to an altitude of 7000 m averaged over 5x5 gridboxes	%
v_cloud	Average cloud velocity calculated according to equation 3.7.2 up to upper boundary of COSMO-D2 (22 km)	ms^{-1}
v_cloud_4000	Average cloud velocity calculated according to equation 3.7.2 up to an altitude of 4000 m	ms^{-1}
v_cloud_5000	Average cloud velocity calculated according to equation 3.7.2 up to an altitude of 5000 m	ms^{-1}
v_cloud_6000	Average cloud velocity calculated according to equation 3.7.2 up to an altitude of 6000 m	ms^{-1}
v_cloud_7000	Average cloud velocity calculated according to equation 3.7.2 up to an altitude of 7000 m	ms^{-1}

Further variables		
sza	Solar zenith angle	*rad*
azi	Solar azimuth angle	*rad*
time_of_day	time	*UTC*
hours_since_sunrise	hours passed since sunrise	*h*

Table A.3.: Parameters selected for use in GHI variability prediction

A.3. Software versions

To ensure the reproducibility of the results presented in this study Table A.4 provides the package versions of the most relevant libraries.

Software package	**Version**
python	3.7 (October 2019)
scikit-learn	0.22
eli5	0.10.1
numpy	1.17.4
netCDF4	1.5.1.2
xarray	0.14.1
pandas	0.25.1

Table A.4.: Software versions of the most relevant Python packages used in this study.

Acknowledgements

First I want to thank my supervisor Prof. Bernhard Mayer for the opportunity to write my dissertation at the Meteorological Institute about this interesting and important topic, and of course for reading through my thesis and providing helpful comments. Furthermore, I want to thank him for all the support during the last 5 years and especially during the Corona pandemic where he wrote at least one email each day to make us almost feel like we are together at university. I also want to thank Stefan Geiß and Leonhard Scheck for providing the COSMO-D2 forecasts, satellite images and explanations about the peculiarities of COSMO-D2 output parameters. Thanks also to Tobias Zinner who was also willing to read parts of the thesis and who provided many helpful comments and ideas, and also thanks to Prof. Martin Weißmann for being the second assessor for my thesis. Furthermore, thanks to Markus Garhammer for being a great help during the preparation of the measurement campaigns, to Markus Kraiczy who spent a lot of time explaining me the German power grid, to Pascal Polonik who was a big help improving my skills in English text writing and of course to my roommates Mihail, Philipp and Paul at university, it was always a great pleasure being with you! I also simply want to thank all friends and colleagues at the institute: Barbara, Veronika, Fabian, Richard, Gregor, Claudia and all the ones I forgot to mention it has been a great time with you. My research would also not have been possible without the MetPVNet project funded by the Federal Ministry for Economic affairs (funding ID: 0350009B) where I also want to thank all partners for the collaborative work and the great achievements. Finally, I want to thank my family for being this patient with me as I was working 5 years on my Phd thesis.

Bibliography

Allan, R. P. (2011). Combining satellite data and models to estimate cloud radiative effect at the surface and in the atmosphere. *Meteorological Applications*, 18(3):324–333.

Anderson, G. P., Clough, S. A., Kneizys, F. X., Chetwynd, J. H., and Shettle, E. P. (1986). AFGL atmospheric constituent profiles (0.120 km). Technical Report AFGL-TR-86-0110, Air Force Geophysics Laboratory Hanscombe Massachusetts.

Andrade, J. R. and Bessa, R. J. (2017). Improving Renewable Energy Forecasting With a Grid of Numerical Weather Predictions. *IEEE Transactions on Sustainable Energy*, 8(4):1571–1580.

Angström, A. (1929). On the Atmospheric Transmission of Sun Radiation and on dust in the air. *Geografiska Annaler*, 11:156–166.

Arias, P. A., Bellouin, N., Coppola, E., Jones, R. G., Krinner, G., Marotzke, J., Naik, V., Palmer, M. D., Plattner, G.-K., Rogelj, J., Rojas, M., Sillmann, J., Storelvmo, T., Thorne, P. W., Trewin, B., Achuta Rao, K., Adhikary, B., Allan, R. P., Armour, K., Bala, G., Barimalala, R., Berger, S., Canadell, J. G., Cassou, C., Cherchi, A., Collins, W., Collins, W. D., Connors, S. L., Corti, S., Cruz, F., Dentener, F. J., Dereczynski, C., Di Luca, A., Diongue Niang, A., Doblas-Reyes, F. J., Dosio, A., Douville, H., Engelbrecht, F., Eyring, V., Fischer, E., Forster, P., Fox-Kemper, B., Fuglestvedt, J. S., Fyfe, J. C., Gillett, N. P., Goldfarb, L., Gorodetskaya, I., Gutierrez, J. M., Hamdi, R., Hawkins, E., Hewitt, H. T., Hope, P., Islam, A. S., Jones, C., Kaufman, D. S., Kopp, R. E., Kosaka, Y., Kossin, J., Krakovska, S., Lee, J.-Y., Li, J., Mauritsen, T., Maycock, T. K., Meinshausen, M., Min, S.-K., Monteiro, P. M. S., Ngo-Duc, T., Otto, F., Pinto, I., Pirani, A., Raghavan, K., Ranasinghe, R., Ruane, A. C., Ruiz, L., Sallée, J.-B., Samset, B. H., Sathyendranath, S., Seneviratne, S. I., Sörensson, A. A., Szopa, S., Takayabu, I., Tréguier, A.-M., van den Hurk, B., Vautard, R., von Schuckmann, K., Zaehle, S., Zhang, X., and Zickfeld, K. (2021). Technical Summary. In Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., editors, *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, pages 33—-144. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Baldauf, M., Gebhardt, C., Theis, S., Ritter, B., and Schraff, C. (2018). Beschreibung des operationellen Kürzestfristvorhersagemodells COSMO-D2 und COSMO-D2-EPS und

seiner Ausgabe in die Datenbank des DWD. Technical report, Deutscher Wetterdienst (DWD), Offenbach.

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M., and Reinhardt, T. (2011). Operational Convective-Scale Numerical Weather Prediction with the COSMO Model: Description and Sensitivities. *Monthly Weather Review*, 139(12):3887–3905.

Benavides Cesar, L., Amaro e Silva, R., Manso Callejo, M. Á., and Cira, C.-I. (2022). Review on Spatio-Temporal Solar Forecasting Methods Driven by In Situ Measurements or Their Combination with Satellite and Numerical Weather Prediction (NWP) Estimates.

Biau, G. and Scornet, E. (2016). A random forest guided tour. *TEST*, 25(2):197–227.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2):123–140.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45:5–32.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). *Classification and regression trees*. Chapman & Hall/CRC, Boca Raton, 1st editio edition.

Brückl, O. (2019). Endbericht der Kommission zur Zukünftigen Beschaffung von Blindleistung. Technical report, Federal Ministry for Economic Affairs and Climate Action.

Buchmann, T. T. (2018). *Potenzial von Photovoltaikanlagen zur Ableitung raum-zeitlich hoch aufgelöster Globalstrahlungsdaten – Entwicklung eines Verfahrens zur Ableitung von Anlagencharakteristika und atmosphärischen Parametern aus Photovoltaikleistungsdaten*. PhD thesis, Universität Heidelberg.

Buras, R., Dowling, T., and Emde, C. (2011). New secondary-scattering correction in DISORT with increased efficiency for forward scattering. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 112(12):2028–2034.

Buras, R. and Mayer, B. (2011). Efficient unbiased variance reduction techniques for Monte Carlo simulations of radiative transfer in cloudy atmospheres: The solution. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 112(3):434–447.

Chandrasekhar, S. (1950). *Radiative transfer*. The international series of monographs on physics. Clarendon Press, Oxford.

Chow, C. W., Urquhart, B., Lave, M., Dominguez, A., Kleissl, J., Shields, J., and Washom, B. (2011). Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. *Solar Energy*, 85(11):2881–2893.

Chu, Y., Li, M., Coimbra, C. F. M., Feng, D., and Wang, H. (2021). Intra-hour irradiance forecasting techniques for solar power integration: A review. *iScience*, 24(10):103136.

Coiffier, J. (2011). *Fundamentals of Numerical Weather Prediction*. Cambridge University Press, Cambridge, UNITED KINGDOM.

Coimbra, C. F. M., Kleissl, J., and Marquez, R. (2013). Chapter 8 - Overview of Solar-Forecasting Methods and a Metric for Accuracy Evaluation. In Kleissl, J. B. T. S. E. F. and Assessment, R., editors, *Solar Energy Forecasting and Resource Assessment*, pages 171–194. Academic Press, Boston.

Cost, C., Shivenes, S., Fluri, V., Peper, D., Memar, A. D., and Schlegl, T. (2021). Stromgestehungskosten Erneuerbare Energien. Technical report, Fraunhofer ISE, Freiburg.

Črnivec, N. and Mayer, B. (2019). Quantifying the bias of radiative heating rates in numerical weather prediction models for shallow cumulus clouds. *Atmospheric Chemistry and Physics*, 19(12):8083–8100.

Eddington, A. S. (1916). On the radiative equilibrium of the stars. *Monthly Notices of the Royal Astronomical Society*, 77:16–35.

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, 32(2):407–499.

Emde, C., Buras-Schnell, R., Kylling, A., Mayer, B., Gasteiger, J., Hamann, U., Kylling, J., Richter, B., Pause, C., Dowling, T., and Bugliaro, L. (2016). The libRadtran software package for radiative transfer calculations (version 2.0.1). *Geoscientific Model Development*, 9(5):1647–1672.

Forster, L., Seefeldner, M., Wiegner, M., and Mayer, B. (2017). Ice crystal characterization in cirrus clouds: a∼sun-tracking camera system and automated detection algorithm for halo displays. *Atmospheric Measurement Techniques*, 10(7):2499–2516.

Frank, C. W., Wahl, S., Keller, J. D., Pospichal, B., Hense, A., and Crewell, S. (2018a). Bias correction of a novel European reanalysis data set for solar energy applications. *Solar Energy*, 164:12–24.

Frank, C. W., Wahl, S., Keller, J. D., Pospichal, B., Hense, A., and Crewell, S. (2018b). Bias correction of a novel European reanalysis data set for solar energy applications. *Solar Energy*, 164:12–24.

Frochte, J. (2019). Entscheidungsbäume. In *Maschinelles Lernen*, chapter 6, pages 117–160. Carl Hanser Verlag GmbH & Co. KG.

Gansler, R. A., Klein, S. A., and Beckman, W. A. (1995). Investigation of minute solar radiation data. *Solar Energy*, 55(1):21–27.

Gasteiger, J., Emde, C., Mayer, B., Buras, R., Buehler, S. A., and Lemke, O. (2014). Representative wavelengths absorption parameterization applied to satellite channels and spectral bands. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 148:99–115.

Geiß, S. (2021). *New approaches for using satellite observations in numerical weather prediction*. PhD thesis, Ludwig-Maximilians-Universität München.

Geleyn, J. and Hollingsworth, A. (1979). An economical analytical method for the computation of the interaction between scattering and line absorption of radiation. *Contributions to Atmosheric Physics*, 52:1–16.

German Environment Agency (2013). Schätzung der Umweltkosten in den Bereichen Energie und Verkehr: Empfehlungen des Umweltbundesamtes. Technical report, Umweltbundesamt, Dessau-Roßlau.

Ghosh, S. (2017). *Kernel Smoothing: Principles, Methods and Applications*. wiley.

Giles, D. M., Sinyuk, A., Sorokin, M. G., Schafer, J. S., Smirnov, A., Slutsker, I., Eck, T. F., Holben, B. N., Lewis, J. R., Campbell, J. R., Welton, E. J., Korkin, S. V., and Lyapustin, A. I. (2019). Advancements in the Aerosol Robotic Network (AERONET) Version∼3 database – automated near-real-time quality control algorithm with improved cloud screening for Sun photometer aerosol optical depth (AOD) measurements. *Atmospheric Measurement Techniques*, 12(1):169–209.

Gregor, P., Zinner, T., Jakub, F., and Mayer, B. (2023). Validation of a camera-based intra-hour irradiance nowcasting model using synthetic cloud data. *Atmospheric Measurement Techniques Discussions [preprint]*, 2023:1–24.

Gregorutti, B., Michel, B., and Saint-Pierre, P. (2017). Correlation and variable importance in random forests. *Statistics and Computing*, 27:659–678.

Gristey, J. J., Feingold, G., Glenn, I. B., Schmidt, K. S., and Chen, H. (2020). On the Relationship Between Shallow Cumulus Cloud Field Properties and Surface Solar Irradiance. *Geophysical Research Letters*, 47(22):e2020GL090152.

Henderson-Sellers, A. and McGuffie, K. (1990). Are cloud amounts estimated from satellite sensor and conventional surface-based observations related? *Remote Sensing*, 11(3):543–550.

Hess, M., Koepke, P., and Schult, I. (1998). Optical Properties of Aerosols and Clouds: The Software Package OPAC. *Bulletin of the American Meteorological Society*, 79(5):831–844.

Hoff, T. E. and Perez, R. (2010). Quantifying PV power Output Variability. *Solar Energy*, 84(10):1782–1793.

Hoff, T. E. and Perez, R. (2012). Modeling PV fleet output variability. *Solar Energy*, 86(8):2177–2189.

Holben, B. N., Eck, T. F., Slutsker, I., Tanré, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A. (1998). AERONET—A Federated Instrument Network and Data Archive for Aerosol Characterization. *Remote Sensing of Environment*, 66(1):1–16.

Inman, R. H., Pedro, H. T. C., and Coimbra, C. F. M. (2013). Solar forecasting methods for renewable energy integration. *Progress in Energy and Combustion Science*, 39(6):535–576.

Janitza, S. and Hornung, R. (2018). On the overestimation of random forest's out-of-bag error. *PloS one*, 13(8):e0201904–e0201904.

Jayadevan, V. T., Rodriguez, J. J., and Cronin, A. D. (2015). A New Contrast-Enhancing Feature for Cloud Detection in Ground-Based Sky Images. *Journal of Atmospheric and Oceanic Technology*, 32(2):209–219.

Joseph, J. H., Wiscombe, W. J., and Weinman, J. A. (1976). The Delta-Eddington Approximation for Radiative Flux Transfer. *Journal of Atmospheric Sciences*, 33(12):2452–2459.

Kabir, E., Kumar, P., Kumar, S., Adelodun, A. A., and Kim, K.-H. (2018). Solar energy: Potential and future prospects. *Renewable and Sustainable Energy Reviews*, 82:894–900.

Kang, B. O. and Tam, K.-S. (2015). New and improved methods to estimate day-ahead quantity and quality of solar irradiance. *Applied Energy*, 137:240–249.

Kassianov, E., Long, C. N., and Ovtchinnikov, M. (2005). Cloud sky cover versus cloud fraction: Whole-sky simulations and observations. *Journal of Applied Meteorology and Climatology*, 44(1):86–98.

Kato, S., Ackerman, T. P., Mather, J. H., and Clothiaux, E. E. (1999). The k-distribution method and correlated-k approximation for a shortwave radiative transfer model. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 62(1):109–121.

Khain, A. P., Beheng, K. D., Heymsfield, A., Korolev, A., Krichak, S. O., Levin, Z., Pinsky, M., Phillips, V., Prabhakaran, T., Teller, A., van den Heever, S. C., and Yano, J.-I. (2015). Representation of microphysical processes in cloud-resolving models: Spectral (bin) microphysics versus bulk parameterization. *Reviews of Geophysics*, 53(2):247–322.

Kim, H. and Lee, D. (2021). Probabilistic Solar Power Forecasting Based on Bivariate Conditional Solar Irradiation Distributions. *IEEE Transactions on Sustainable Energy*, 12(4):2031–2041.

Kraiczy, M., Altayara, A., Wenderoth, F., Winter, K., Hofbauer, P., Meilinger, S., and Braun, M. (2021). Benefits of advanced PV power forecasts for congestion management and reactive power management at the distribution level. In *ETG Congress 2021*, pages 1–6.

Kraiczy, M., Braun, M., Wirth, G., Schmidt, S., and Brantl, J. (2013). Interferences between local voltage control strategies of a hv/mv-transformer and distributed generators. In *European PV Solar Energy Conference*.

Kreuwel, F. P. M., Knap, W. H., Visser, L. R., van Sark, W. G., Vilà-Guerau de Arellano, J., and van Heerwaarden, C. C. (2020). Analysis of high frequency photovoltaic solar energy fluctuations. *Solar Energy*, 206:381–389.

Lee, J.-Y., Marotzke, J., Bala, G., Cao, L., Corti, S., Dunne, J., Engelbrecht, F., Fischer, E., Fyfe, J., Jones, C., Maycock, A., Mutemi, J., Ndiaye, O., Panickal, S., and Zhou, T. (2021). Future Global Climate: Scenario-Based Projections and Near-Term Information. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, chapter 4, pages 553–672. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Li, Q., Lu, W., and Yang, J. (2011). A Hybrid Thresholding Algorithm for Cloud Detection on Ground-Based Color Images. *Journal of Atmospheric and Oceanic Technology*, 28(10):1286–1296.

Liou, K.-N. (2002). *An introduction to atmospheric radiation*, volume 84 of *International Geophysics Series*. Academic press, second edition.

Lorenz, E., Hammer, A., and Heinemann, D. (2004). Short term forecasting of solar radiation based on satellite data. *EUROSUN2004 (ISES Europe Solar Congress)*.

Lorenz, E., Hurka, J., Heinemann, D., and Beyer, H. G. (2009). Irradiance Forecasting for the Power Prediction of Grid-Connected Photovoltaic Systems. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2(1):2–10.

Madhavan, B. L., Kalisch, J., and Macke, A. (2016). Shortwave surface radiation network for observing small-scale cloud inhomogeneity fields. *Atmospheric Measurement Techniques*, 9(3):1153–1166.

Marcos, J., de la Parra, Í., García, M., and Marroyo, L. (2016). Simulating the variability of dispersed large PV plants. *Progress in Photovoltaics: Research and Applications*, 24(5):680–691.

Marquez, R., Pedro, H. T. C., and Coimbra, C. F. M. (2013). Hybrid solar forecasting method uses satellite imaging and ground telemetry as inputs to ANNs. *Solar Energy*, 92:176–188.

Mayer, B. (2009). Radiative transfer in the cloudy atmosphere. *The European Physical Journal Conferences*, 1:75–99.

Mayer, B. and Kylling, A. (2005). Technical note: The libRadtran software package for radiative transfer calculations - description and examples of use. *Atmospheric Chemistry and Physics*, 5(7):1855–1877.

Meilinger, S., Herman-Czezuch, A., Kimiaie, N., Schirrmeister, C., Yousif, R., Geiss, S., Scheck, L., Weissmann, M., Gödde, F., Mayer, B., Zinner, T., Barry, J., Pfeilsticker, K., Kraiczy, M., Winter, K., Altayara, A., Reise, C., Rivera, M., Deneke, H.,

Witthuhn, J., Betcke, J., Schroedter-Homscheidt, M., Hofbauer, P., and Rindt, B. (2021). Entwicklung innovativer satellitengestützter Methoden zur verbesserten PV-Ertragsvorhersage auf verschiedenen Zeitskalen für Anwendungen auf Verteilnetzebene BT - Schlussbericht. Technical report, Hochschule-Bonn-Rhein-Sieg.

Mie, G. (1908). Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen. *Annalen der physik*, 330(3):377–445.

Mills, A., Ahlstrom, M., Brower, M., Ellis, A., George, R., Hoff, T., Kroposki, B., Lenox, C., Miller, N., Milligan, M., Stein, J., and Wan, Y. h. (2011). Dark Shadows. *IEEE Power and Energy Magazine*, 9(3):33–41.

Mitchell, M. W. (2011). Bias of the Random Forest Out-of-Bag (OOB) Error for Certain Input Parameters. *Open Journal of Statistics*, 01(03):205–211.

Nam, S. and Hur, J. (2019). A hybrid spatio-temporal forecasting of solar generating resources for grid integration. *Energy*, 177:503–510.

Nicodemus, F. E., Richmond, J. C., Hsia, J. J., Ginsberg, I. W., and Limperis, T. (1977). Geometrical Considerations and Nomenclature for Reflectance. Technical Report NBS MN_160, National Bureau of Standards - Department of Commerce, Washington D.C. 202334.

Otani, K., Minowa, J., and Kurokawa, K. (1997). Study on areal solar irradiance for analyzing areally-totalized PV systems. *Solar Energy Materials and Solar Cells*, 47(1):281–288.

Pecenak, Z. K., Mejia, F. A., Kurtz, B., Evan, A., and Kleissl, J. (2016). Simulating irradiance enhancement dependence on cloud optical depth and solar zenith angle. *Solar Energy*, 136:675–681.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Pelland, S., Remund, J., Kleissl, J., Oozeki, T., and De Brabandere, K. (2013). Photovoltaic and Solar Forecasting: State of the Art. Technical report, IEA.

Perez, R., Kivalov, S., Schlemmer, J., Hemker, K., Renné, D., and Hoff, T. E. (2010). Validation of short and medium term operational solar radiation forecasts in the US. *Solar Energy*, 84(12):2161–2172.

Pérez-Ramírez, D., Whiteman, D. N., Smirnov, A., Lyamani, H., Holben, B. N., Pinker, R., Andrade, M., and Alados-Arboledas, L. (2014). Evaluation of AERONET precipitable water vapor versus microwave radiometry, GPS, and radiosondes at ARM sites. *Journal of Geophysical Research: Atmospheres*, 119(15):9596–9613.

Planck, M. (1901). Ueber das Gesetz der Energieverteilung im Normalspectrum. *Annalen der physik*, 309(3):553–563.

Probst, P., Wright, M. N., and Boulesteix, A. L. (2019). Hyperparameters and tuning strategies for random forest.

Refaeilzadeh, P., Tang, L., and Liu, H. (2016). Cross-Validation. In Liu, L. and Özsu, M. T., editors, *Encyclopedia of Database Systems*, pages 1–7. Springer New York, New York, NY.

Reichardt, J., Reichardt, S., Hostetler, C. A., Lucker, P. L., McGee, T. J., Twigg, L. W., Dörnbrack, A., Schoeberl, M. R., and Yang, P. (2015). Mother-of-pearl cloud particle size and composition from aircraft-based photography of coloration and lidar measurements. *Applied Optics*, 54(4):B140–B153.

Riihimaki, L. D., Li, X., Hou, Z., and Berg, L. K. (2021). Improving prediction of surface solar irradiance variability by integrating observed cloud characteristics and machine learning. *Solar Energy*, 225:275–285.

Ritchie, H., Roser, M., and Rosado, P. (2022). Global direct primary energy consumption. https://ourworldindata.org/grapher/global-primary-energy.

Sarkar, M. N. I., Meegahapola, L. G., and Datta, M. (2018). Reactive Power Management in Renewable Rich Power Grids: A Review of Grid-Codes, Renewable Generators, Support Devices, Control Strategies and Optimization Algorithms. *IEEE Access*, 6:41458–41489.

Schaaf, C. B., Gao, F., Strahler, A. H., Lucht, W., Li, X., Tsang, T., Strugnell, N. C., Zhang, X., Jin, Y., Muller, J.-P., Lewis, P., Barnsley, M., Hobson, P., Disney, M., Roberts, G., Dunderdale, M., Doll, C., D'Entremont, R. P., Hu, B., Liang, S., Privette, J. L., and Roy, D. (2002). First operational BRDF, albedo nadir reflectance products from MODIS. *Remote Sensing of Environment*, 83(1):135–148.

Scheck, L. (2021). A neural network based forward operator for visible satellite images and its adjoint. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 274:107841.

Scheck, L., Frèrebeau, P., Buras-Schnell, R., and Mayer, B. (2016). A fast radiative transfer method for the simulation of visible satellite imagery. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 175:54–67.

Schroedter-Homscheidt, M., Kosmale, M., Jung, S., and Kleissl, J. (2018). Classifying ground-measured 1 minute temporal variability within hourly intervals for direct normal irradiances. *Meteorologische Zeitschrift*, 27(2):161–179.

Segal, M. R. (2004). Machine Learning Benchmarks and Random Forest Regression. *UCSF: Center for Bioinformatics and Molecular Biostatistics*.

Shettle, E. P. (1990). Models of aerosols, clouds, and precipitation for atmospheric propagation studies. In *AGARD Conference Proceedings*, AA(B, MA.).

Sinyuk, A., Holben, B. N., Eck, T. F., Giles, D. M., Slutsker, I., Korkin, S., Schafer, J. S., Smirnov, A., Sorokin, M., and Lyapustin, A. (2020). The AERONET Version 3 aerosol retrieval algorithm, associated uncertainties and comparisons to Version 2. *Atmospheric Measurement Techniques*, 13(6):3375–3411.

Skartveit, A. and Olseth, J. A. (1992). The probability density and autocorrelation of short-term global and beam irradiance. *Solar Energy*, 49(6):477–487.

Sommeria, G. and Deardorff, J. W. (1977). Subgrid-Scale Condensation in Models of Nonprecipitating Clouds. *Journal of Atmospheric Sciences*, 34(2):344–355.

Song, J. (2015). Bias corrections for Random Forest in regression using residual rotation. *Journal of the Korean Statistical Society*, 44(2):321–326.

Stamnes, K., Tsay, S.-C., Wiscombe, W., and Jayaweera, K. (1988). Numerically stable algorithm for discrete-ordinate-method radiative transfer in multiple scattering and emitting layered media. *Applied Optics*, 27(12):2502–2509.

Steinberger-Willms, R. (1993). *Untersuchung der Fluktuationen der Leistungsabgabe von räumlich ausgedehnten Wind- und Solarenergie-Konvertersystemen im Hinblick auf deren Einbindung in elektrische Versorgungsnetze*. Ph.d. thesis, Carl von Ossietzky Universitat Oldenburg.

Sterchele, P., Brandes, J., Heilig, J., Wrede, D., Kost, C., Schlegl, T., Bett, A., and Henning, H.-M. (2020). Paths to a Climate-Neutral Energy System: The German Energy Transition in Its Social Context. Technical report, Fraunhofer ISE, Freiburg.

Stevens, B., Ament, F., Bony, S., Crewell, S., Ewald, F., Gross, S., Hansen, A., Hirsch, L., Jacob, M., Kölling, T., Konow, H., Mayer, B., Wendisch, M., Wirth, M., Wolf, K., Bakan, S., Bauer-Pfundstein, M., Brueck, M., Delanoë, J., Ehrlich, A., Farrell, D., Forde, M., Gödde, F., Grob, H., Hagen, M., Jäkel, E., Jansen, F., Klepp, C., Klingebiel, M., Mech, M., Peters, G., Rapp, M., Wing, A. A., and Zinner, T. (2019). A High-Altitude Long-Range Aircraft Configured as a Cloud Observatory: The NARVAL Expeditions. *Bulletin of the American Meteorological Society*, 100(6):1061–1077.

Stevens, B., Moeng, C.-H., Ackerman, A. S., Bretherton, C. S., Chlond, A., de Roode, S., Edwards, J., Golaz, J.-C., Jiang, H., Khairoutdinov, M., Kirkpatrick, M. P., Lewellen, D. C., Lock, A., Müller, F., Stevens, D. E., Whelan, E., and Zhu, P. (2005). Evaluation of Large-Eddy Simulations via Observations of Nocturnal Marine Stratocumulus. *Monthly Weather Review*, 133(6):1443–1462.

Stevens, B., Moeng, C.-H., and Sullivan, P. P. (1999). Large-Eddy Simulations of Radiatively Driven Convection: Sensitivities to the Representation of Small Scales. *Journal of the Atmospheric Sciences*, 56(23):3963–3984.

Stevens, B. and Seifert, A. (2008). Understanding macrophysical outcomes of microphysical choices in simulations of shallow cumulus convection. *Journal of the Meteorological Society of Japan. Ser. II*, 86A:143–162.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., and Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, 9(1):307.

Strobl, C., Boulesteix, A.-L., Zeileis, A., and Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8(1):25.

Suehrcke, H. and McCormick, P. G. (1988). The frequency distribution of instantaneous insolation values. *Solar Energy*, 40(5):413–422.

Tan, Y. T. and Kirschen, D. S. (2007). Impact on the Power System of a Large Penetration of Photovoltaic Generation. In *2007 IEEE Power Engineering Society General Meeting*, pages 1–8.

Tanré, D., Geleyn, J. F., and Slingo, J. (1984). *First results of the introduction of an advanced aerosol-radiation interaction in the ECMWF low resolution global model. Aerosols and Their Climatic Effects*. Deepak Publishing, Hampton, Va.

Tegen, I., Hollrig, P., Chin, M., Fung, I., Jacob, D., and Penner, J. (1997). Contribution of different aerosol species to the global aerosol extinction optical thickness: Estimates from model results. *Journal of Geophysical Research: Atmospheres*, 102(D20):23895–23915.

Tiedtke, M. (1989). A Comprehensive Mass Flux Scheme for Cumulus Parameterization in Large-Scale Models. *Monthly Weather Review*, 117(8):1779–1800.

Trenberth, K. E., Fasullo, J. T., and Kiehl, J. (2009). Earth's Global Energy Budget. *Bulletin of the American Meteorological Society*, 90(3):311–324.

Vidot, J. and Borbás, É. (2014). Land surface VIS/NIR BRDF atlas for RTTOV-11: model and validation against SEVIRI land SAF albedo product. *Quarterly Journal of the Royal Meteorological Society*, pages 2186–2196.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., and Bright, J. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods*, 17(3):261–272.

Visser, L., AlSkaif, T., and van Sark, W. (2022). Operational day-ahead solar power forecasting for aggregated PV systems with a varying spatial distribution. *Renewable Energy*, 183:267–282.

Wanner, W., Li, X., and Strahler, A. H. (1995). On the derivation of kernels for kernel-driven models of bidirectional reflectance. *Journal of Geophysical Research: Atmospheres*, pages 21077–21089.

Watanabe, T., Oishi, Y., and Nakajima, T. Y. (2016). Characterization of surface solar-irradiance variability using cloud properties based on satellite observations. *Solar Energy*, 140:83–92.

Wiemken, E., Beyer, H. G., Heydenreich, W., and Kiefer, K. (2001). Power characteristics of PV ensembles: experiences from the combined power production of 100 grid connected PV systems distributed over the area of Germany. *Solar Energy*, 70(6):513–518.

Wirth, H. (2022). Aktuelle Fakten zur Photovoltaik in Deutschland. Technical report, version date 04.02.2022, URL: www.pv-fakten.de, Fraunhofer ISE, Freiburg.

Zdunkowski, W., Trautmann, T., and Bott, A. (2007). *Radiation in the atmosphere: a course in theoretical meteorology*. Cambridge University Press.

Zhang, G. and Lu, Y. (2012). Bias-corrected random forests in regression. *Journal of Applied Statistics*, 39(1):151–160.

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, dass ich meine Dissertation mit dem Titel

Predicting Short-Term Solar Irradiance Variability combining Numerical Weather Prediction and Machine Learning

selbstständig verfasst, sowie keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Marktschellenberg, den 21. März 2023

(Felix Simon Gödde)