Support Vector Regression for Solar Power Prediction

Dissertation zur Erlangung des Grades
eines Doktors der Naturwissenschaften (Dr. rer. nat.)

vorgelegt von
Herrn Björn Wolff, M.Sc.
geboren am 30.12.1985 in Nordhorn

Oldenburg, den 12. April 2017

Abstract

In recent years, renewable energies have been covering an increasing part of the worldwide electrical power demand. The additional volatility introduced to power grids by weather dependent renewable energy sources, i.e., wind and solar, makes it necessary to improve the accuracy of energy forecasts, so that the underlying electrical grid can be operated in a cost efficient way. Governmental funding of renewable energy plants led to a vast increase of photovoltaic (PV) power plants, that now hold the second highest share of installed power capacity in Germany after onshore wind farms.

The focus of this thesis lies on improving forecast accuracy current prediction models for short-term PV power predictions with horizons ranging from 15 minutes to five hours. For this, recent power measurements of PV systems throughout Germany are combined with data from parametric weather forecast approaches. Apart from commonly used numerical weather prediction (NWP) model output, satellite-based cloud motion vector predictions are applied. The data sets of these sources are preprocessed and combined with machine learning techniques. In particular, support vector regression (SVR) is optimized to work as the main regression model. SVR is an extension of the support vector machine classification method and is capable of learning regression functions in continuous space by identifying structures in the mapping of input to output data. One advantage over classical statistical models is its ability to learn non-linear dependencies between the applied data sets.

In an extensive comparison, the SVR-based predictions are able to challenge a physical PV power forecasting model that is used for operational services. Instead of forecasting the PV systems output through physical irradiance to power modeling, the SVR approach does not require detailed information about a PV system. Predictions of the same quality can be achieved by SVR with access to PV measurements and weather forecasts only. To research further possible improvements in prediction quality, a study on NWP weather parameters, that are not yet included in the physical modeling system, is performed to assess their suitability as input features for PV power forecasting. The results confirm the importance of high quality irradiance forecasts. In a subsequent single PV system optimization, PV measurements and irradiance forecasts of neighboring systems are added to the input of the SVR model to make use of spatio-temporal correlations between PV systems. Furthermore, different parameter tuning techniques are evaluated with the goal to improve forecasts on a single system level, and thus develop an approach that dynamically optimizes the forecasts of all single PV systems.

iv

Zusammenfassung

Erneuerbare Energien haben einen immer größer werdenden Anteil an der Deckung des weltweiten Stromverbrauchs. Jedoch wird durch die wetterabhängigen, erneuerbaren Energiequellen wie Wind- und Solarenergie zusätzliche Unsicherheit eingeführt, welche die Verbesserung von bestehenden Vorhersageverfahren notwendig macht, um die effiziente Betriebsführung des Stromnetzes zu garantieren. Durch die u.a. starke Förderung von erneuerbaren Energien durch das Erneuerbare-Energien-Gesetz (EEG) ist die Anzahl an Photovoltaik (PV)-Anlagen deutlich gestiegen, wodurch diese nach Onshore-Windenergieanlagen nun den zweithöchsten Anteil an installierter Leistung in Deutschland zur Verfügung stellen.

Der Fokus dieser Arbeit liegt in der Verbesserung von aktuellen Vorhersagemodellen für PV-Leistung im Kurzzeitbereich von 15 Minuten bis 5 Stunden. Hierzu werden aktuelle Leistungsmessungen mit Vorhersagen aus verschiedenen Wettervorhersagemodellen kombiniert. Neben vielseits genutzten Prognosedaten von numerischen Wettermodellen (NWP) werden auch satelliten-basierte Cloud Motion Vector-Vorhersagen eingesetzt. Diese unterschiedlichen Datensätze werden mit Hilfe von maschinellen Lernverfahren vorverarbeitet und kombiniert. Vor allem die Support Vector Regression (SVR) wird für den Einsatz zur Vorhersage von PV-Leistung optimiert und dient im gesamten Verlauf der Arbeit als Hauptmethode zur Erstellung von Prognosen. SVR ist in der Lage Funktionen in einem kontinuierlichen Wertebereich, durch Erkennung von Strukturen bei der Abbildung von Eingabe- auf Ausgabewerte, zu erlernen. Eine für diesen Anwendungszweck interessante Eigenschaft der SVR ist die Möglichkeit, auch nicht lineare Zusammenhänge der Eingabedaten zu erkennen und entsprechend abzubilden.

In einer umfangreichen Gegenüberstellung wird gezeigt, dass SVR-basierte Ansätze in der Lage sind ebenbürtige Vorhersagen im Vergleich zu physikalisch-parametrischen Modellen, wie sie in aktuellen Anwendungen eingesetzt werden, zu erzeugen. Im Gegensatz zur physikalischen Simulation von PV-Anlagen, um aus Strahlungsprognosen die zukünftige Ausgangsleistung zu erhalten, benötigt die SVR lediglich Messwerte und Strahlungsvorhersagen, um qualitativ gleichwertige Leistungsvorhersagen zu bestimmen. Zusätzlich wurde in dieser Arbeit eine Studie zur Nutzung von weiteren NWP Vorhersageparametern, welche aktuell noch nicht operativ miteinbezogen werden, durchgeführt und ihr Potential als Eingabemerkmale für die PV-Leistungsprognose evaluiert. Hierbei bestätigt sich der große Einfluss von hochqualitativen Strahlungsprognosen für die Vorhersage von PV-Leistung. Daran anschließend wird eine Einzelsystemoptimierung durchgeführt, in der PV-Leistungsmessungen und Strahlungsprognosen benachbarter PV-Anlagen dem SVR-basierten Modell zur Verfügung gestellt werden, wodurch mögliche

raumzeitliche Abhängigkeiten zwischen PV-Systemen genutzt werden können. Des Weiteren werden unterschiedliche Parameteroptimierungsverfahren evaluiert, mit dem Ziel die Vorhersagen auf Einzelstationsniveau zu verbessern und daraus resultierend ein Verfahren zu entwickeln, welches in der Lage ist Einzelstationsprognosen dynamisch zu optimieren.

Contents

Contents

1. Introduction

There are multiple reasons for a transition from conventional fossil fuel-based energy production to sustainable renewable sources. One leading aspect is to contain the causes of climate change by reducing carbon dioxide emissions and using the produced energy more efficiently. Even though nuclear power is not emitting carbon dioxide in its energy production process, there are many risks and unsolved waste issues surrounding this topic, so that it is not suitable as a solution for the energy supply of the future. In contrast to this, wind and solar power are two of the most promising renewable energy sources, where there have been large improvements in the efficiency of their technologies in recent years. Not only efficiency is increasing but also the costs of these technologies are decreasing, e.g., the production of photovoltaic power modules is declining rapidly which makes renewable energy more affordable for the public.

Due to the volatility of these two technologies, the integration of their produced energy into the electricity grid is complex and requires reliable short-term forecasting methods. One way to aquire high quality predictions is to learn from historic data. Hence, the application of machine learning algorithms is getting increasingly common and shows encouraging results.

The aim of this thesis is to analyze the capability of machine learning algorithms to create solar power forecasts. In particular, support vector regression (SVR) is tested and optimized due to its good performance in previous works [55].

This chapter is structured as follows: In Section 1.1, the topic of research is further motivated, where the role of power forecasts in energy systems and the benefit of using machine learning are discussed. Section 1.2 gives an overview of related works in this field. An overview on this thesis is given in Section 1.3.

1.1. Motivation

In recent years, renewable energies have been providing an increasing part of the worldwide electrical power consumption. In 2016, renewables covered around 37% of the net electricity consumption in Germany. Nearly 6.5% of this energy, which corresponds to 38.3 TWh, was produced by decentralized photovoltaic (PV) systems across Germany.

PV is able to reach power output rates of almost 26 *GW* at midday with a total nominal power of around 41 GW_{peak} [52]. To put this into perspective, to generate the same amount of energy one would need about 20 of the bigger nuclear power plants in Germany, whereas today only eight plants with varying capacities are still active. The installed capacity of these nuclear power plants added up to about 11 GW at the end of 2016[1]. In Figure 1.1, the most recent share of installed capacities of all energy sources in Germany is shown, which illustrates the high importance of PV power in the German grid.
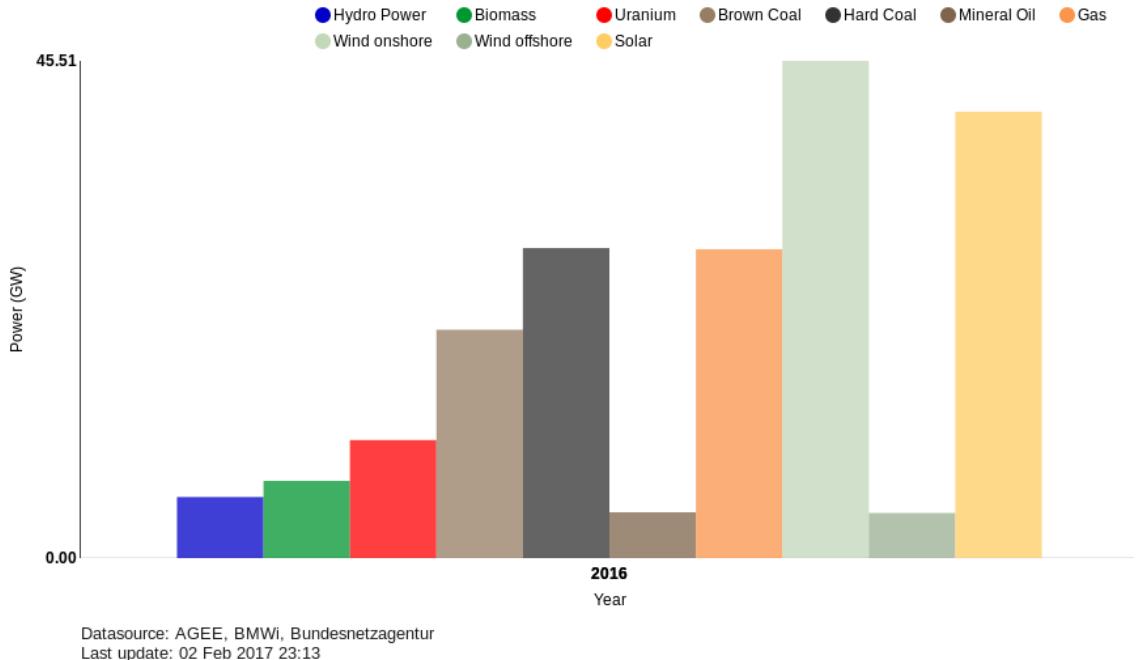
Figure 1.1.: *Share of net installed electricity generation capacities in Germany in 2016 (from Fraunhofer ISE [17]).*

Power Forecasts in Energy Systems   A major challenge with renewable energy sources is their volatile nature. Unlike conventional power plants, one can not access the resource on demand. Changing weather conditions have a big influence on wind and solar power production. The power output of PV systems is highly dependent on the non-deterministic cloud coverage as well as the deterministic incidence angle of the sun's radiation. Accurate forecasts are vital for the integration of renewable energy into the grid.

---

1 Global Energy Observatory, Current List of Nuclear Power Plants: http://globalenergyobservatory.org/list.php?db=PowerPlants&type=Nuclear, retrieved: 2017-04-11

Electrical power is traded at exchanges, where energy supply and demand forecasts determine the prices. There are three submarkets in electricity trading that apply for German traders: the forward market at the European Energy Exchange (EEX)[2] and the spot market at the European Power Exchange (EPEX SPOT)[3], which includes the day-ahead and intraday markets. The trading at the forward market deals with energy deliveries that are due in the next months and up to several years. Traders at the day-ahead market at EPEX SPOT negotiate hourly contingents of energy deliveries that take place at the following day. Day-ahead trading closes at 12 am of the day prior to the delivery day. Starting three hours after the closing of the day-ahead market, the intraday trading starts. Now, hourly quantities of electricity can be traded until 45 minutes before their fulfillment. Especially for day-ahead and intraday trading, precise power forecasts of renewable energies are valuable as their accuracy is directly correlated to the risks with which bids can be placed. The higher the accuracy of renewable energy forecasts, the earlier and cheaper energy can be traded without relying on too much expensive balancing power. In Figure 1.2, the energy marketing process with the mentioned dates is illustrated.
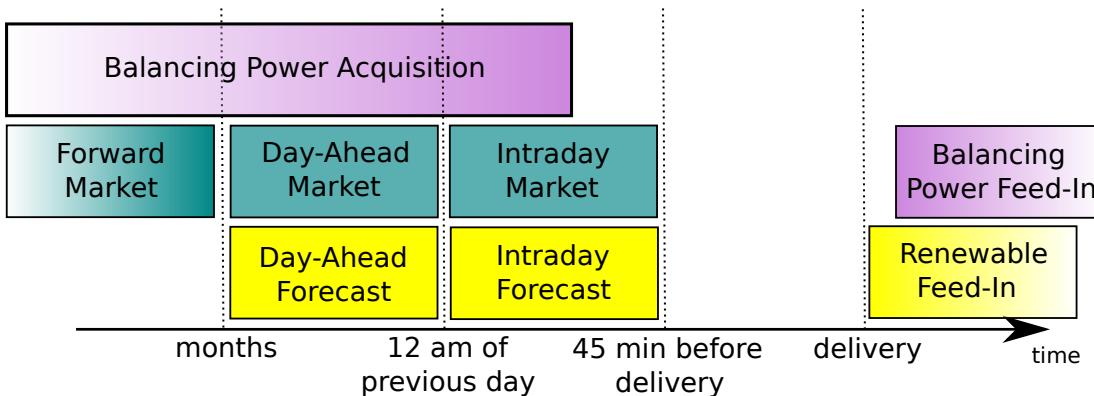
Figure 1.2.: *Overview of the energy marketing process and relevant dates in relation to the delivery date of energy contingents. Green boxes indicate market affiliation, purple boxes show respective balancing power processes, and yellow boxes indicate renewable energy forecasts and feed-in (based on Kühnert [30]).*

Another important part of ensuring the stability of an electricity grid is keeping a fixed utility frequency, which rises with production and lowers with the consumption of energy. Until about a decade ago, transmission system operators solved this problem by simply matching the energy output and dispatch of power plants to the demand of its

---

consumers. Today, the increasing share of weather dependent renewable energy sources on the consumer's side has been introducing additional volatility to the grid, thus keeping stability intact is becoming a more complex task. Here, power forecasts are essential for the (cost) efficient scheduling of fossil-fuel based power plants to stabilize the grid. Especially, as long as direct energy storage in form of batteries is too expensive for large scale application and indirect storage, e.g., pumped hydroelectrical and power-to-gas storage, are dealing with high conversion losses.

Benefits of Machine Learning   With the steady increase in computing power, the application of statistical or machine learning techniques in many different fields is enabled. By detecting structures in datasets, these algorithms are able to create reasonable predictions when new data is presented to them.

In this work, machine learning is applied for forecasting PV power outputs. There are many parametric models for PV power forecasting that use the laws of physics and empirical modeling to assess future power output by converting irradiance forecasts. In contrast to purely data-driven approaches, i.e., machine learning methods, these physical or parametric models require some details about the underlying PV systems to generate accurate forecasts. This includes knowledge about the PV material and technology, installed capacities, the location, orientation and tilt of the system as well as the ambient temperature to determine a PV module's efficiency.

Furthermore, there are always certain limitations of empirical modeling as not all possibilities can be parameterized and accounted for, e.g., in most cases the actual state of the atmosphere at a certain location can only be assumed. Additionally, empirical models may not be suitable for a specific PV system because its model parameters are optimized at a different location under different conditions, e.g., a model that is developed for Europe would not work properly in Australia without readjustment.

Machine learning algorithms only need measurement and forecast data of PV systems to learn the behavior of a PV system and calculate predictions. Apart from that, specifications of a system are not necessary. Thus, the whole system can be seen as a black box with irradiance as input and PV power as output. Most of the classical models assume that there are only linear relationships between the input features of the model and the actual power output, whereas an advantage of SVR is its ability to learn possible non-linear dependencies from the data.

1.2. Related Work

A variety of works exist that implement physical modeling approaches applying PV measurements and/or numerical weather predictions (NWP) on different spatial and temporal scales, see Pelland et al. [47] for a comparison of different models. For short-term forecasts up to a few hours ahead, satellite-based approaches have been investigated by Lorenz et al. [36]. A more detailed description of PV power forecasting with physical parametric modeling can be found in Kühnert [30].

Machine learning or statistical learning algorithms are used increasingly for wind and solar energy predictions. In first approaches, the focus laid on forecasts implementing artificial neural networks (ANN); an overview on ANN forecast models for solar irradiance was done by Mellit [43] in 2008.

More recent works that are using ANN as a main prediction model are extending their process chain with preprocessing steps such as weather type classifications. In Chen et al. [9], a self-organized map is trained to classify site-specific weather types before a forecast is calculated. One way to improve the forecasts from data-driven models is to extend the amount of data that is available to the algorithm. In Marquez et al. [42], additional data from the numerical weather prediction system of the NWS[4] are employed. To keep the computational complexity feasible, an input feature selection is applied as well. A similar model is used by Pedro et al. [46], where genetic algorithms determine the input data of ANN models.

Apart from ANN realizations, there are many works successfully applying and comparing different statistical modeling methods for PV energy production forecasting. In Brabec et al. [5], the energy production of PV system farms in the Czech Republic are modeled with statistical nonparametric regression approaches. These models are trained using irradiance forecasts from the Global Forecasting System (GFS) NWP model of NOAA[5]. A hybrid model that applies ANN and wavelet transformation is presented in Mandal et al. [41]; a model that is forecasting short-term PV power based on similarity in historic data with genetic algorithms is proposed in Monteiro et al [44].

The usage of SVR is relatively new for PV power and irradiance forecasts. Nonetheless, some comparisons to other statistical learning methods (see Zeng and Qiao [56]) with numerical weather prediction features as input, e.g., in Kroemer et al. [29] on single-site, and regional solar irradiance values – as in Gala et al. [18] – are already available. Li et

---

[4] National Weather Service (NWS): http://www.weather.gov/, retrieved: 2017-04-11
[5] National Oceanic and Atmospheric Administration (NOAA): http://www.noaa.gov/, retrieved: 2017-04-11

al. [33] evaluated different models, i.a., SVR and Hidden Markov, for solar irradiance forecasting in Australia with a temporal resolution of five to ten minutes.

In later chapters of this thesis, the machine learning algorithm random forest is employed as another regression model. There are only few works that use random forest as a predictor in the field of renewable energy forecasts. One example where random forest is successfully used for electric load forecasting can be found in Jurado et al. [26].

1.3. Overview of this Thesis

In this work, the focus lies on single-site and regional short-term PV power forecasts in the range of 15 minutes to five hours ahead. Especially, regional forecasts are important for the transmission system operators to manage the electricity grid in a cost-efficient way. These analyzed forecast horizons are most interesting for the intra-day market as they are still useful for traders to negotiate energy contingents. Apart from physical PV power forecasting techniques, i.e., PV simulation models, support vector regression is used and optimized throughout the thesis to evaluate the ability of machine learning to serve as an alternative and complement to physical forecasting models. In contrast to the mentioned works, in a first analysis the SVR model input is restricted to inputs that are used for the physical modeling. In later chapters, the SVR model is further optimized and additional data is tested to improve the quality of SVR PV power forecasts. Moreover, the amount of the employed datasets is significantly larger than in the mentioned works and the application of satellite-based cloud motion vectors in combination with SVR has not been tested yet. In summary, there are three major questions in this work:

- Is machine learning a feasible approach for PV power forecasting? (Chapter 4)

- Are more site-specific input features from numerical weather predictions increasing PV power forecast quality? (Chapter 5)

- Are there spatial-temporal correlations to measurement and forecast data of surrounding PV systems that improve the accuracy of PV power forecasts? (Chapter 6)

In the following, an overview on the chapters of this work is given. The single chapters are grouped into three parts: Foundations, SVR-based prediction, and Ending. Each chapter is concluded with a short summary. The chapters of the SVR-based prediction part always start with a description of the respective experimental settings. As most of

the experiments are highly computationally expensive, the cluster systems of the HPC Facilities[6] of the University of Oldenburg are used.

Foundations In this part of the thesis, the basics of PV power forecasting and machine learning are covered. **Chapter 2** starts with the introduction of different types of irradiance (see Section 2.1) and, in Section 2.2, gives an overview on the data sources that are available for PV power forecasting. The main sources of data are PV power measurements, weather forecasts from numerical weather predictions, and satellite-based irradiance forecasts from cloud motion vectors. Afterwards in Section 2.3, a physical parametric PV forecasting approach is introduced that is using the mentioned datasets and serves as a reference to the applied machine learning models. To improve PV power forecasting, two methods, i.e., persistence (Section 2.3.2) and the simulation of single PV systems by employing irradiance forecasts (Section 2.3.1), are combined in this PV modeling approach.

In **Chapter 3**, a basis for the understanding of machine learning algorithms used throughout the thesis is established. For this, Section 3.1 discusses the basic concepts for supervised and unsupervised machine learning. In Section 3.2, the concept of support vector machine (SVM) is derived. Starting with linear SVMs for linear separable spaces in Section 3.2.1, the steps to a non-linear SVM model for non-linear separable spaces is explained (see Section 3.2.2). The extension from the classification model SVM to the regression modeling technique support vector regression is explained in Section 3.3.

SVR-based Prediction This part discusses different approaches for modeling and optimizing PV power forecasts with SVR. In **Chapter 4**, an initial approach to assess the suitability of support vector regression for PV power forecasting is conducted. To tune the SVR predictor for this task, a testing subset of three different datasets is generated by applying a clustering method in Section 4.2. Afterwards in Section 4.4, the parameters of the SVR are optimized on this subset and SVR models using the different input datasets are evaluated. In this comparison, a SVR model that combines all input datasets is trained as well. In Section 4.5, all implemented and tuned SVR models are compared to the physical parametric PV power forecasting model.

After analyzing SVR-based forecasts and physical modeling, one method to improve SVR forecasts is evaluated in **Chapter 5**. Until this point, only irradiance and temperature forecasts are employed from the NWP models. In this chapter, all available features of a NWP model are considered for PV power forecasting. To assess the information gain

---

6 High Performance Computing (HPC) Facilities, HERO cluster system: https://www.uni-oldenburg.de/fk5/wr/hochleistungsrechnen/hpc-facilities/historie/hero/, retrieved: 2017-04-11

from new features, random forest's ability to measure feature importance is applied. For this, random forest and its feature importance calculation is introduced in Section 5.2. The feature selection is performed in Section 5.3 and resulting models are evaluated in Section 5.4. Additionally, a feature selection for special weather conditions, i.e., snowfall and fog, is done in Section 5.5.

With the findings of the previous chapter, **Chapter 6** concentrates on a single PV system and its optimization. Accordingly, a single system is selected from the PV measurement dataset with a high data availability to enable thorough optimization experiments. Furthermore, data from surrounding PV systems is used for forecasting the PV power output of the single system. A short study on the usefulness of PV measurements of neighboring systems is done in Section 6.2.1. In Section 6.2.2, a feature selection on all possible input features of the single PV system and its neighboring systems is performed to determine how many features are necessary for accurate forecasting. For the resulting subset of input features, three different SVR parameter optimization methods are tested in Section 6.2.3. The optimized models are compared in Section 6.3, where they are also evaluated against the best model of the previous chapters.

Ending This thesis ends with a summary and an outlook of possible next evaluation and experiment steps in **Chapter 7**.

Part I.

Foundations

2. Solar Power Prediction

The most common parametric models for PV power forecasting use past PV system measurements to predict future outputs and/or simulate a PV system by applying an irradiance-to-power conversion model based on existing irradiance and temperature forecasts as well as specifications of single PV systems. To ensure accurate forecasts, these models are in need of high quality input data. Apart from irradiance forecasts from numerical weather prediction systems, data from a satellite-based cloud motion vector approach are applied in this work.

In Section 2.1, global horizontal and clear sky irradiance are introduced. The data sources used for PV power forecasting throughout the thesis are listed and explained in Section 2.2. Afterwards, a parametric PV power and persistence forecasting model that utilizes this input data are described. These models serve as a reference for comparisons in later chapters.

2.1. Irradiance and Clear Sky Index

Irradiance describes the radiant flux received by the earth's surface that is emitted by the sun. The unit of irradiance is specified in power per area, typically in W/m^2. In this work, the term irradiance is freely used for what is actually known as global horizontal irradiance (I), which quantifies the sum of direct (I_{dir}) and diffuse horizontal irradiance (I_{diff}):

$$I = I_{dir} + I_{diff} \tag{2.1}$$

While I_{dir} describes the irradiance that is directly received at the surface, I_{diff} is the part of irradiance that is scattered on its way through the atmosphere. Apart from scattering, there are extinction processes occuring, so that the intensity of the initial extraterrestial irradiance I_{ext} is decreased by the earth's atmosphere.

If only the atmospheric extinction and deterministic aspects of solar irradiance, i.e., seasonal and diurnal variations, are considered, the term clear sky irradiance I_{clear} is used. There are many different models, that are able to estimate the clear sky irradiance at any given location for any given timestep. In this work, the clear sky model of Dumortier

is used – as it is described in Fontoynont et al. [16] – with turbidity values according to Bourges [4] and Dumortier [12]. The clear sky index k^* defines the ratio between I_{clear} and I as

$$k^* = \frac{I}{I_{clear}}. \tag{2.2}$$

This permits a quantification of the transmissivity of clouds and simultaneously takes the absorption, that is not included in the clear sky irradiance, into account. For situations with no clouds, a value of $k^* = 1$ is expected, whereas on overcast conditions the clear sky index approaches 0. A value of $k^* = 0$ is only possible at night, as there is always a fraction of I_{diff} present at daytime.

2.2. Data Sources

In the following sections, different data sources are described that are used to implement PV power forecasts for different forecast horizons. Besides the output of the aforementioned clear sky model, there are three main sources of data: on-site power output measurements of PV systems, weather forecasts of numerical weather prediction models, and satellite-based cloud motion vector irradiance forecasts. The following sections give an overview of these datasets which are used in all of the chapters of this thesis.

2.2.1. PV Power Measurements

This work aims at creating high quality PV power forecasts, so that appropriate measurements are needed for model training purposes. For this, measurement data of PV systems in Germany from 2012 to 2016 with a temporal resolution of 15 minutes is used. These PV power measurements are made available from the monitoring database of meteocontrol GmbH[1].

Chapter	Type	Time Frame	PV Systems
Chapter 4	testing	2012-01-01 – 2012-12-31	1135
	evaluation	2013-01-01 – 2013-12-31	921
Chapter 5	testing	2012-01-01 – 2012-12-31	92
	evaluation	2013-01-01 – 2013-12-31	92
Chapter 6	testing	2014-07-01 – 2015-06-30	1
	evaluation	2015-07-01 – 2016-06-30	1

Table 2.1.: *PV power measurement data used for testing and evaluation in different chapters of this thesis.*

1 meteocontrol GmbH – Energy and Weather Services: `www.meteocontrol.com`, retrieved: 2017-04-11
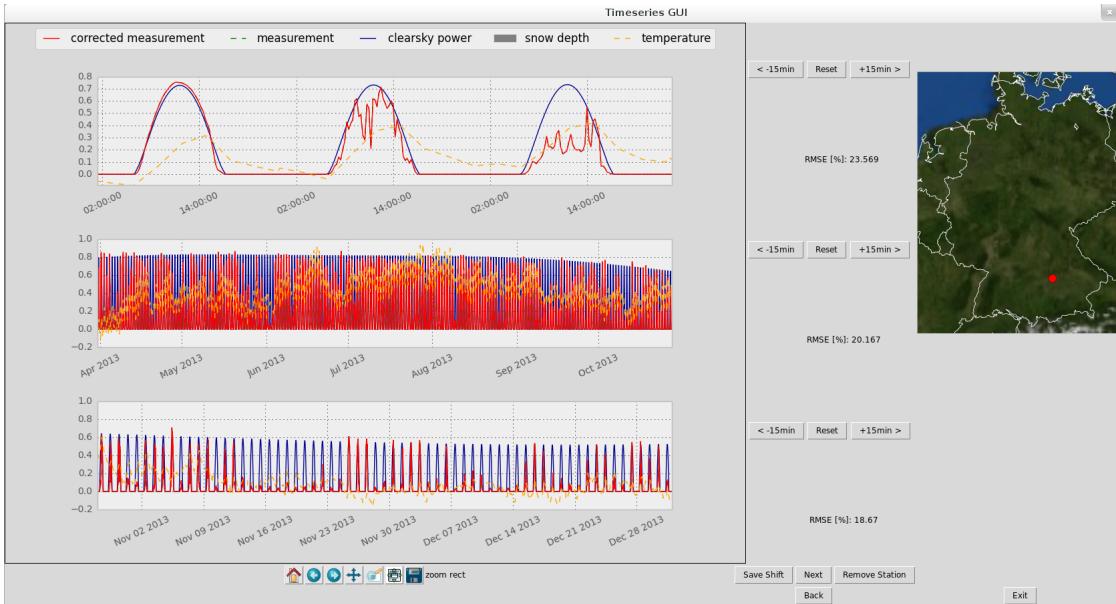
Figure 2.1.: *GUI tool for manual PV system assessment and time series selection.*

From a large PV measurement dataset, PV system subsets are selected that hold a high data availability and passed several quality control tests, i.e., detection and removing of possible time shifts and physically unreasonable values in the data. Additionally to the automatic quality control, suitable systems are selected manually to detect unusual behaviour of the PV modules, e.g., long periods of snow with no power output due to snow covering the modules. A screenshot of a GUI tool, that was implemented especially for this purpose, is shown in Figure 2.1.

An overview on the exact time frames and number of PV systems used in later chapters is listed in Table 2.1. Besides the location, information on installed capacity, orientation, and tilt of each of these PV systems is available. As there is a wide variety in sizes and nominal power, the output of each system is scaled to its installed power capacity to consider all systems equally.

To get an impression of the distribution of the PV systems in Germany, the locations of 921 PV systems from the evaluation dataset, that is used in Section 4, are marked in Figure 2.2(a).

2.2.2. Numerical Weather Predictions

Weather Forecasts from numerical weather prediction (NWP) systems are offered by various providers for prediction horizons from a few hours up to about 15 days ahead. NWP models are based on dynamic equations that use the laws of physics, fluid dynamics,

(a)   (b)
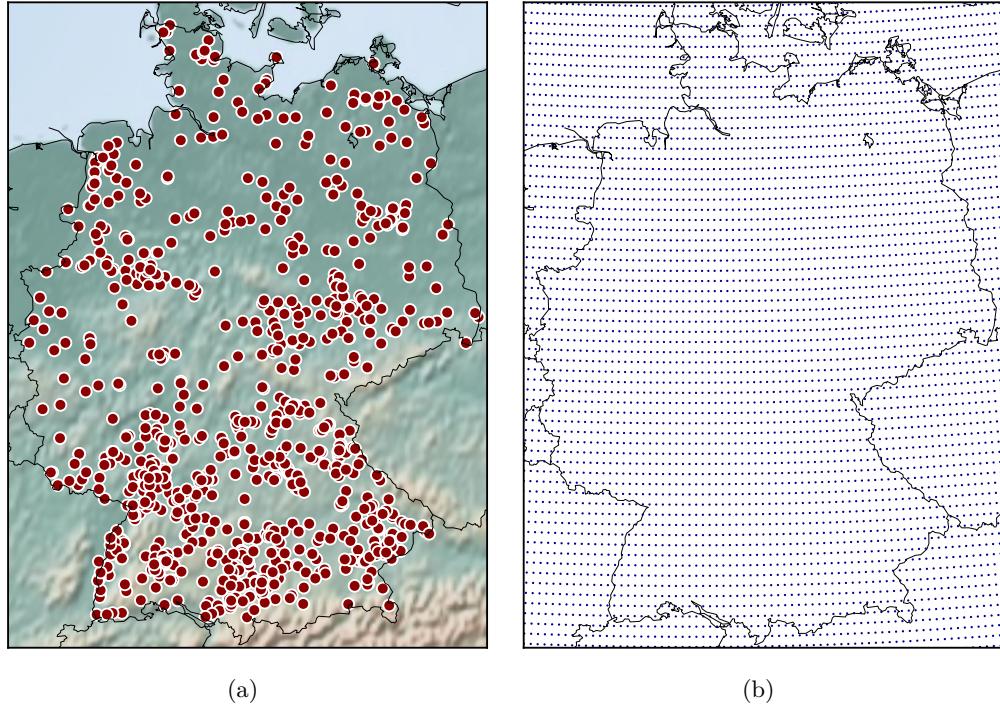
Figure 2.2.: *(a) Locations of 921 PV systems in Germany from the evaluation dataset of Section 4. (b) Spatial distribution of ECMWF's NWP model grid points.*

and chemistry as well as weather observations and measurements. Starting from a recent initial state, changes of the atmosphere are simulated for the following timesteps. As the computational expenses are high, the calculations are restricted to specific horizontal and vertical grid points arranged with a fixed resolution.

In this work, two different NWP models are used: the IFS model of the ECMWF[2] and the COSMO-EU model of the DWD[3]. The ECMWF's IFS model yields a temporal resolution of three hours and a spatial resolution of about 12.5 km × 12.5 km, while the COSMO-EU model generates hourly forecasts with approximately 7 km × 7 km resolution.

In this work, the COSMO-EU model only serves as a second source for irradiance forecasts. All other weather forecasts are extracted from the IFS model, which provides single level forecasts covering over 120 parameters. Single level forecasts are forecasts that are averaged over different height/pressure levels or forecasts for specific heights (e.g., the temperature in 2 meter height). The IFS model has two major forecasting runs

---

2 European Centre for Medium-Range Weather Forecasts (ECMWF) [14]
3 Deutscher Wetterdienst (DWD) (engl.: German Weather Service) [11]

starting at 00 UTC and 12 UTC. These runs are not instantly available at these times, so that forecasts from the 12 UTC run of the previous day are used here to ensure their availability at the point of time of the prediction. In Figure 2.2(b), the grid points of the ECMWF IFS model is shown. As only single level forecasts are employed, there is no need for the vertical expansion of the grid in this case.

2.2.3. Satellite-based Cloud Motion Vectors

The movement of clouds has a major influence on surface irradiance and its modeling is the main subject of intraday forecasts like the cloud motion vector (CMV) approach. Images of a geostationary satellite are used to detect cloud structures in near-real time. Cloud index images, which contain information on the cloudiness of each image's pixels, are calculated by applying the semi-empirical Heliosat method.

To forecast cloud index images as a basis for surface solar irradiance forecasting, two consecutive cloud index images are compared and a cloud motion vector field is computed, which describes the velocity and direction of cloud areas. These are applied to the latest available cloud index image, thus extrapolating the clouds' structures to the next minutes and hours. Details about the methods used for creating irradiance forecasts from satellite images are given in the following paragraphs.

Irradiance from Satellite Images   The Meteosat Second Generation (MSG) satellite of EUMETSAT[4] is positioned in the earth's orbit above the zero degree latitude and longitude mark and is equipped with eleven spectral channels. The spatial resolution of its images amounts to about 3 km $\times$ 3 km; a high resolution visible broadband irradiance channel is providing an even higher spatial resolution of 1 km $\times$ 1 km. Images of Germany are delivered with a resolution of 1.2 km in east-west and 1.8 km in north-south orientation at a temporal resolution of 15 minutes [13].

To assess the irradiance from satellite images, the Heliosat method is used, which was originally proposed in 1986 by Cano et al. [7]. For the data that is used in this work, an improved version for solar energy applications of Beyer et al. [2] and Hammer et al. [21] is applied. The general idea is to determine the extinction of irradiance induced by clouds from the satellite's image and use it to adjust the modeling of clear sky conditions to the actual cloud situation. In a first step, the relative reflectance ρ is calculated with

$$\rho = \frac{C - C_0}{I_{ext}}, \tag{2.3}$$

---

4 European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT): http://www.eumetsat.int, retrieved: 2017-04-11

where C denotes a pixel's intensity received from the satellite's sensors and C_0 results from a known instrumental offset. To assess the cloudiness of a pixel, the dimensionless cloud index n determined as

$$n = \frac{\rho - \rho_{min}}{\rho_{max} - \rho_{min}}. \tag{2.4}$$

is calculated. The lower bound ρ_{min} is set by the ground's reflectivity (ground albedo), which is usually lower than the reflectivity of clouds, and the upper bound ρ_{max} is defined by an empirical cloud reference value. With a cloud index image, where each pixel's value represents a cloud index value, the corresponding satellite-based clear sky index k_{sat}^* can be derived as

$$k_{sat}^* = \begin{cases} 1.2 & \text{if } n \leq -0.2 \\ 1 - n & \text{if } -0.2 < n \leq 0.8 \\ 1.661 - 1.7814n + 0.7250n^2 & \text{if } 0.8 < n \leq 1.05 \\ 0.09 & \text{if } 1.05 < n, \end{cases} \tag{2.5}$$

according to Hammer et al. [22]. By using I_{clear} from the clear sky model in Section 2.1 and applying Equation (2.2), a satellite-based global horizontal irradiance I_{sat} is calculated with

$$I_{sat} = k_{sat}^* \cdot I_{clear}. \tag{2.6}$$

Cloud Motion Vectors Because clouds are the main reason for the power output decline of a PV system, forecasting the cloud-induced shadowing of a system is of high importance. For satellite-based approaches, a prediction of the cloud index value n of a image's pixel at a PV system's location is needed. In this work, the cloud motion vector approach is applied for this task. The CMV method used here is introduced in Lorenz et al. [36] and further evaluations for the application in context of an operational forecasting service are provided in Kühnert et al. [32]. Figure 2.3 gives an illustration of how the cloud motion vector approach works. There are four main steps to the method:

1. The pixels of two consecutive cloud index images $n(t - \Delta t)$ and $n(t)$ are compared. Movement in the images is tracked by matching the correlation between pixels from $n(t - \Delta t)$ to surrounding pixels in the next image $n(t)$.

2. Step 1 is repeated for all pixels and a cloud motion vector field is generated. This field is afterwards applied to $n(t)$, which extrapolates the current movements into the future and creates the cloud motion image $n(t + \Delta t)$.

3. Additionally, a smoothing of $n(t+\Delta t)$ is implemented to include a level of uncertainty to the forecast.
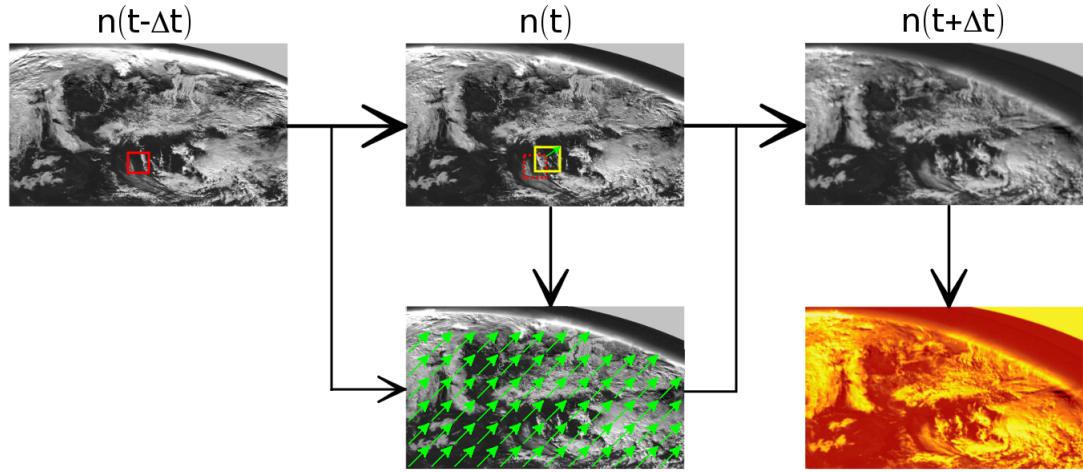
Figure 2.3.: *Retrieving irradiance forecasts from satellite images using the cloud motion vector approach.*

4. The irradiance values for the cloud motion image $n(t + \Delta t)$ are now calculated with the Heliosat method as described earlier in this section.

The resolution of CMV forecasts corresponds to the temporal and spatial resolution of the utilized satellite images described earlier. Studies show advantages of the CMV approach compared to persistence and NWP approaches from 30 minutes to four or five hours ahead [35, 32].

2.3. Physical PV Power Modeling

The reference in the comparisons of Section 4 is a parametric irradiance-to-power conversion model based on explicit physical description of a PV system applied to irradiance forecasts. In Section 2.3.1, the PV simulation model is introduced followed by the description of a persistence approach in Section 2.3.2. The power conversion is applied on CMV and NWP irradiance forecasts seperately. A combination approach of persistence and simulated power outputs from CMV and NWP irradiance forecasts to achieve an optimized PV power forecast is detailed in Section 2.3.3.

2.3.1. PV Simulation

The power output of a PV system can be modeled based on global horizontal irradiance and ambient temperature using a physical modeling approach. In Figure 2.4, an overview of the applied PV simulation model, with references to the individual models used in each

step, is given. This PV power simulation approach utilizes irradiance forecasts received from CMV (I_{CMV}) and NWP (I_{NWP}) as well as the temperature forecast from NWP (T_{NWP}) to obtain PV power forecasts $P_{Sim,CMV}$ and $P_{Sim,NWP}$. The basic steps of the PV simulation are:

- Conversion of global horizontal irradiance **I** – in this case I_{CMV} or I_{NWP} – to irradiance received on the tilted surface of the PV module $\mathbf{I_{tilt}}$.

- Calculating the irradiance dependent efficiency $\eta(\mathbf{I_{tilt}})$ of the PV module with a parametric efficiency model.

- Further adjustment of η considering PV module temperature $\mathbf{T_{mod}}$ derived from NWP ambient temperature forecasts $\mathbf{T_a}$ (here T_{NWP}) and $\mathbf{I_{tilt}}$.

- Calculating the expected power output of the PV system.
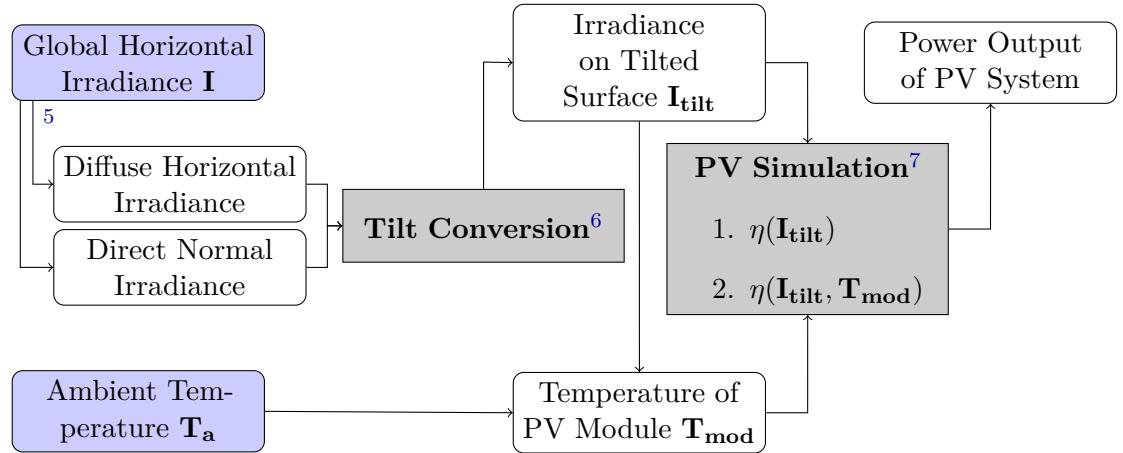
Figure 2.4.: *PV simulation scheme.*

Tilt Conversion The amount of irradiance received by a PV module strongly depends on the tilt and orientation of the system. Irradiance on a tilted surface I_{tilt} is calculated by applying the tilt conversion model on diffuse horizontal I_{diff} and direct irradiance I_{dir}. In this thesis, the splitting of the global horizontal irradiance I into its components I_{diff} and I_{dir} is implemented with the model of Skartveit and Olseth [49].

---

5 Skartveit and Olseth [49]
6 Klucher [27]
7 Beyer et al. [3]

To determine I_{tilt} each of the irradiance components is converted separately into its respective tilted form and later recombined adding irradiance that is reflected by the ground (I_{ref}):

$$I_{tilt} = I_{tilt,diff} + I_{tilt,dir} + I_{ref} \qquad (2.7)$$

In this step, the model of Klucher [27] is applied for modeling $I_{tilt,diff}$. The second component $I_{tilt,diff}$ can be derived by using geometric equations that include the solar zenith angle and the angle of incidence. The reflected irradiance I_{ref} is determined by the ground albedo, i.e., the ground reflectivity of a PV module's surroundings, as well as the landscape of the surface.

Modeling of PV Efficiency The PV efficiency η is simulated with the model described in Beyer et al. [3]. In this model, both I_{tilt} and the ambient temperature T_a at a PV systems' site are considered. The model works in two steps: First, the efficiency of a constant module temperature $T_{const} = 25°C$ is assumed:

$$\eta(I_{tilt}, T_{const}) = a_1 + a_2 \ I_t + a_3 \ log \ I_t, \qquad (2.8)$$

with empirically determined and device-specific parameters a_i, $i \in \{1, 2, 3\}$. In a next step, the actual estimated temperature of a PV module T_{mod} is added:

$$\eta(I_{tilt}, T_{mod}) = \eta(I_{tilt}, T_{const})(1 + \alpha(T_{mod} - T_{const})), \qquad (2.9)$$

with another module type dependent temperature coefficient α. The temperature of the PV module is assumed as $T_{mod} = T_a + \zeta \ I_{tilt}$, where the ambient temperature is increased by a ratio ζ of the irradiance's influence. The DC power output of a PV system is calculated as

$$P_{DC} = \frac{\eta(I_{tilt}, T_{mod})}{\eta_{STC}} \frac{I_{tilt}}{I_{STC}} P_{nom}. \qquad (2.10)$$

The efficiency of a PV module at standard test conditions η_{STC} is calculated by applying Equation (2.8) with $I_{STC} = 1000 \frac{W}{m^2}$. Power losses (e.g., DC to AC conversion), are combined into a single loss parameter ζ_{loss}, so that the final PV simulation model is defined as

$$P_{SIM} = P_{DC} \ \zeta_{loss}. \qquad (2.11)$$

In this method, the parameters a_i, α, ζ, and ζ_{loss} are fitted by applying a least-square algorithm for each PV system using its respective power measurements. A more detailed description of the fitting and the PV simulation model in general can be found in Kühnert [30].

2.3.2. Persistence Forecasts

To include information about the current situation at each PV system's site, a persistence model is applied on recent PV power measurements. PV power measurements P_{meas} and clear sky PV power $P_{Sim,clear}$ are used to calculate the PV power clear sky index

$$k_{P,meas}^*(t) = \frac{P_{meas}(t)}{P_{Sim,clear}(t)}. \qquad (2.12)$$

The clear sky PV power $P_{Sim,clear}$ is achieved by applying the PV simulation model on clear sky irradiance I_{clear} from the clear sky model as well as using NWP temperature forecasts. This PV power clear sky index $k_{P,meas}^*$ is defined in analogy to the irradiance clear sky index k^* from Equation (2.2) and provides a quantification of cloud cover at a PV system's site. Based on measurements, this quantity enables a description of a system's recent state without being limited by resolution and quality of NWP or satellite data. For persistence, the relation

$$k_{P,pers}^*(t + \Delta t) \equiv k_{P,meas}^*(t) \qquad (2.13)$$

is assumed, i.e., $k_{P,meas}^*(t)$ is presumed to be constant for the next time step $t + \Delta t$. The PV power at time step $t + \Delta t$ is calculated by applying the inverse of Equation (2.12):

$$P_{pers,meas}(t + \Delta t) = k_{P,meas}^*(t) \cdot P_{Sim,clear}(t + \Delta t), \qquad (2.14)$$

in order to model the daily course of solar irradiance. The quality of persistence forecasts is strongly influenced by the weather situation. They perform best in stable weather conditions, when the assumption of constant $k_{P,meas}^*(t)$ is well matched for the next hours. In situations with high fluctuations due to a rapid change of cloudiness, the quality of persistence forecasts strongly decreases with the forecast horizon [31].

2.3.3. Combination of Forecast Approaches

While persistence forecasts generate good forecasts for very short forecast horizons, the accuracy decreases quickly with an increased forecast horizon. As PV power forecasts based on NWP models are updated only once per day, these forecasts are producing more general predictions compared to the other approaches. Therefore, they are mainly suited for longer forecast horizons. Satellite-based CMV forecasts provide good forecast inbetween these two. By combining these three approaches, a forecasting model is created that is able to benefit from each of its components and provide accurate forecasts on all

employed forecast horizons. This is discussed in more detail in the evaluation section of Chapter 4.

The individual PV power forecasts, i.e., $P_{pers,meas}$, $P_{Sim,CMV}$, and $P_{Sim,NWP}$ are combined by fitting them linearly to historical measurements. A combined forecast $P_{Sim,comb}$ is achieved by weighting the single forecasts with the coefficents a, b, and c using a least-square linear regression:

$$P_{Sim,comb}(t) = a \cdot P_{pers,meas}(t) + b \cdot P_{Sim,CMV}(t) + c \cdot P_{Sim,NWP}(t) \tag{2.15}$$

These weighting factors are determined for each forecast horizon individually using measured PV power from the preceding days. For single sites and regional averages, individual training and training periods are used. A training period of 50 days provided a sufficient amount of data to learn seasonal trends for single system forecasts. When combining regionally averaged forecasts, a training period of 30 days showed the best results. More details about the fitting of the linear regression can be found in Kühnert et al.[31]. An overview of the physical PV power forecasting process is shown in Figure 2.5.

For the comparison of $P_{Sim,comb}$ to the single model forecasts $P_{pers,meas}$, $P_{Sim,CMV}$, and $P_{Sim,NWP}$, a linear regression approach (lr) is applied to the single forecasts as well. This enables a distinction between statistical improvements of single forecasts due to fitting them to PV measurements and beneficial effects from the combination of these forecasts. The adapted forecasts are calculated as

$$P_{pers,meas}^{lr} = a'_{pers}P_{pers,meas} + b'_{pers} \tag{2.16}$$

$$P_{Sim,NWP}^{lr} = a'_{NWP}P_{Sim,NWP} + b'_{NWP}, \tag{2.17}$$

$$P_{Sim,CMV}^{lr} = a'_{CMV}P_{Sim,CMV} + b'_{CMV}, \tag{2.18}$$

with individual weighting factors a' and b', that are again fitted to PV measurements of the previous days.

2.4. Evaluation of Forecasts

In this section, the basic definitions concerning the evaluation of forecasts are summarized. Comparing different forecasts always entails analyzing time series. In this work, PV measurements are used as reference time series and the quality of each forecast is measured by the deviations between measured and predicted values at each timestep. A common
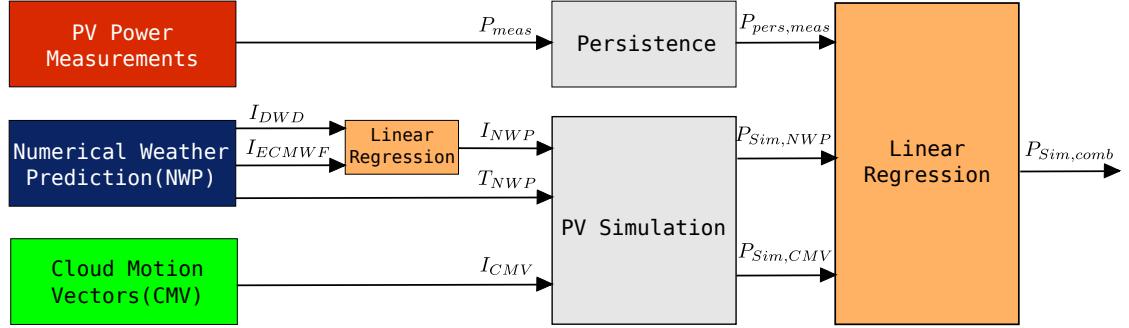
Figure 2.5.: *Physical parametric PV power forecasting process.*

measure of deviation between two time series z and z' is called mean square error (MSE):

$$E_{MSE} = \frac{1}{N} \sum_{i=1}^{N} (z - z')^2, \tag{2.19}$$

where N denotes the length of a time series. The MSE is a good real life measure as it penalizes large deviations, that would have a bigger impact on grid stabilizing actions than small ones. In model comparisons, the root mean square error (RMSE)

$$E_{RMSE} = \sqrt{E_{MSE}} \tag{2.20}$$

is used more frequently, as the values are returned to their original range. While the RMSE is good for assessing the overall performance of a prediction, the mean bias error (BIAS) provides information about over- and underestimations, and therefore is able to point out systematical errors. The BIAS is defined as the averaged deviation of two time series:

$$E_{BIAS} = \frac{1}{N} \sum_{i=1}^{N} (z - z') \tag{2.21}$$

In this thesis, all forecast evaluation is done only on daytime values, where the solar elevation exceeds an angle of ten degrees. This is mainly done because of the unavailability of CMV forecasts for the earlier and later hours of the day, as the usability of satellite images is restricted to these conditions [32]. If different forecast models are compared, the time series are filtered so that only timesteps are considered where all models generate usable output.

The two terms single system and regional average are often used throughout this thesis. A single system forecast is considering the forecast of each single PV system. But instead of calculating the error values of each single system, all time series of the considered

PV systems are concatenated before the performance is evaluated. To some extent, this removes the influence of outliers and improves the assessment of forecast models. In contrast to the evaluation of single system forecasts, the time series for regional forecasts are averaged before error values are calculated.

2.5. Concluding Remarks

In this chapter, PV power forecasting is addressed and different methods are explained. In the first sections, different types of irradiance are discussed and all available datasets are introduced.

There are three main sources of data in this work: PV power measurements, numerical weather prediction (NWP) models, and satellite-based cloud motion vectors (CMV). The PV power measurement datasets are provided by meteocontrol GmbH. From large sets of measurements, smaller subsets are created for the evaluations of this thesis. The temporal resolution of these PV measurements is 15 minutes.

In this work, data of the NWP models of the ECMWF and DWD is applied. These models provide multiple different weather parameters, especially their irradiance forecasts are important for the PV power forecast models of this section. While ECMWF provides three hourly data, DWD's forecasts are hourly. The spatial distribution of both models depends on the underlying calculation grid and are around 12.5 km × 12.5 km (ECMWF) and 7 km × 7 km (DWD).

The third source of data is irradiance forecasts from satellite images. The movement of clouds is detected by comparing two subsequent satellite images and identifying changes on pixel level. Afterwards, the resulting cloud motion vectors are applied to the latest image to extrapolate clouds to their (possible) future position. An irradiance forecast from this new image is derived with the Heliosat method. Due to the image resolution of the satellite, the spatial resolution of the forecasts is restricted to 1.2 km (east-west) and 1.8 km (north-south).

In the second part of this chapter, a physical PV power forecast model is introduced. This model is used as a reference in the later chapters of this thesis. The model combines two approaches for forecasting the power output of PV systems.

The first approach is using a PV simulation model to forecast a PV systems future output by simulating the system with existing irradiance and temperature forecasts. As irradiance forecasts, both NWP and CMV forecasts are applied.

In the second approach called persistence, only PV measurements are utilized. In persistence forecasting, it is assumed that the current cloud conditions last for the next

timesteps, thus the current ratio between PV measurement and the highest physically possible power output is applied at the following timesteps.

Counting persistence and both power forecasts from PV simulations there are three resulting PV power forecasts. These forecasts are combined with linear regression, where the weighting factors of the regression model are tuned on historic PV power measurements of days prior to the actual forecasting timestep.

The last section of this chapter introduces the error measures and experimental settings that apply for all following evaluations of this thesis.

3. Support Vector Regression

In recent years, the support vector machine (SVM) was successfully employed on classification tasks in very different areas of application. As the focus of this work lies on forecasting PV power, and therefore values from continuous space, an extension of SVM called support vector regression (SVR) is applied. The main concept of these methods stays the same: Find a function that represents a given set of training data the best and apply this function to previously unseen data points to acquire a prediction based on the training data.

Starting in Section 3.1 with machine learning basics, this chapter establishes the foundation to understand the notations and algorithms used in later chapters. To ease the transition to regression functions, linear and non-linear SVMs for the classification task are clarified in Section 3.2.1 and Section 3.2.2. Finally, the extension of SVM to regression modeling is discussed in Section 3.3.

3.1. Machine Learning Basics

Today, there are many disciplines utilizing learning algorithms on so-called big data problems. Whether one is using the term machine learning, statistical learning, or data mining to describe these methods, the core principle is learning from data. Learning in this setting describes the identification of basic structures in datasets and applying these on new data of the same form.

To apply machine learning algorithms on specific data, distinguishing the data in an input space $\mathcal{X}$ and an output space $\mathcal{Y}$ is obligatory. The pattern $\mathbf{x} \in \mathcal{X}$ containing features $(x_1, x_2, \ldots, x_d)$, with dimensionality d of $\mathcal{X}$, specify a data point in the input space. Each of these pattern holds a representation $y \in \mathcal{Y}$ in the output space that is called label. Throughout this work, patterns consist mostly of irradiance forecasts and previous PV power measurements and labels are corresponding measurements from PV systems.

In all learning scenarios, there are three important datasets: the training set, the test set, and the evaluation set. While the training set is used to train the learning algorithm, the parameter and model optimization is done on the test set. Afterwards, the optimized

model is analyzed on an independent evaluation dataset. In this case, the training set consists of measurements and forecasts for a predefined number of days prior to the actual timestep that is forecasted.

The aim of learning algorithms is to find a prediction function $f : \mathcal{X} \to \mathcal{Y}$ that assigns a label y to a before unseen pattern $\mathbf{x}$. Hence, the function f is learned on a training set

$$\mathcal{T} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_N, y_N)\} \subset \mathcal{X} \times \mathcal{Y}, \tag{3.1}$$

where N denotes the training set size that contains known relations between patterns and labels. After introducing the training set, it is now possible to distinguish between supervised and unsupervised learning.

Supervised Learning   In case of supervised learning the association of patterns and labels in $\mathcal{T}$ are available to the learning algorithm as seen in Equation (3.1). The most prominent groups of supervised learning methods are classification and regression. For classification tasks, the output space is given as a finite set of discrete values called classes, e.g., for the binary setting used in the following sections $\mathcal{Y} = \{1, -1\}$. In contrast to this, in most cases where a regression function is learned, the output space consists of continuous real-valued labels. In this work, mainly regression functions are learned, thus $\mathbf{x} \in \mathcal{X} = \mathbb{R}^d$ for patterns and $y \in \mathcal{Y} = \mathbb{R}$ for labels apply.

To facilitate the understanding and usage of classification and regression models, k-nearest neighbors (k-NN) [10], is often refered as a simple example for a supervised machine learning algorithm. This approach uses the labels of neighboring patterns to decide which label should be assigned to a new pattern $\mathbf{x}'$. The algorithm considers the k closest data points in input space and uses the label that the majority of patterns hold. The classification decision function f_{knn} with output space $\mathcal{Y} = \{1, -1\}$ can be defined as

$$f_{kNN}(\mathbf{x}') = \begin{cases} 1 & \text{if } \sum_{i \in \mathcal{N}_k(\mathbf{x}')} y_i > 0 \\ -1 & \text{if } \sum_{i \in \mathcal{N}_k(\mathbf{x}')} y_i \leq 0, \end{cases} \tag{3.2}$$

where $\mathcal{N}_k$ denotes the set of the k-nearest neighbors of a new pattern $\mathbf{x}'$ in a training set $\mathcal{T}$. To measure distances in input space, the Euclidian distance

$$\delta(\mathbf{x}, \mathbf{x'}) = (\sum_{j=1}^{d}(x_j - x'_j)^2)^{\frac{1}{2}} \tag{3.3}$$

is often applied. In Figure 3.1, an example illustrates the decision making process of k-NN when a new pattern is introduced into a set of already labeled data points.
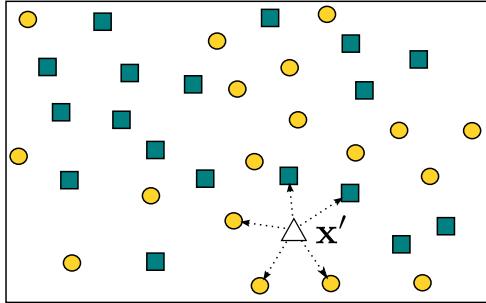
Figure 3.1.: *An example of classifying with k-nearest neighbors. Two classes (yellow circles and green squares) are given and a new data point* $\mathbf{x}'$ *is introduced. For* $k = 5$ *the label of* $\mathbf{x}'$ *would be a yellow circle as three of the five nearest neighbors are labeled this way.*

To demonstrate the connection between classification and regression, the k-NN model for regression is now introduced. The k-NN algorithm is easily extended to be used as a regression function by replacing f_{knn} from Equation (3.2) with the arithmetic mean and calculating the average of the k-nearest neighbors' labels. In this work, a distance weighted version by Bailey and Jain [1]

$$f_{wkNN}(\mathbf{x}') = \sum_{i \in \mathcal{N}_K(\mathbf{x}')} \frac{\delta(\mathbf{x}_i, \mathbf{x}')^{-1}}{\sum_{j \in \mathcal{N}_K(\mathbf{x}')} \delta(\mathbf{x}_j, \mathbf{x}')^{-1}} \; y_i, \tag{3.4}$$

where the inverse of the Euclidian distance, is employed to weight the importance of a neighbor's label depending on its distance to the new pattern $\mathbf{x}'$. In Chapter 5 and 6, k-NN is applied as an easy-to-use model for spatial interpolation of forecast data.

Unsupervised Learning If there is no information about the labels available or prespecified, unsupervised learning methods are applied. With clustering, classifying patterns without preexisting labels y is possible. Patterns that share a similarity will be put (clustered) into the same classes and labeled accordingly. The goal is to create clusters, where the similarity of patterns within one cluster exceeds the similarity of data points of another cluster regarding a preselected similarity measure.

One of the most popular and simple clustering algorithms is k-means clustering, which was first proposed by Lloyd [34]. The main goal of k-means clustering is to divide a set of data points into k distinct clusters and label the members accordingly. For this algorithm, the number k of clusters must be defined in advance. The most common implementation of k-means is iteratively improving the location of cluster center (centroids) by applying each of the following steps:

1. Initialize centroids by randomly selecting (without replacement) k data points $\mathbf{c_1}, \ldots, \mathbf{c_k}$.

2. Calculate the Euclidian distance (see Equation 3.3) between each data point $\mathbf{x}_i$ and each centroid. Assign a label to each $\mathbf{x}_i$ corresponding to the nearest centroid's:

$$y_i = \operatorname*{argmin}_{1 \leq j \leq k} \delta(\mathbf{x}_i, \mathbf{c_j}) \tag{3.5}$$

3. Recalculate the centroids $\mathbf{c_j}$ by averaging the values of all h data points belonging to their respective cluster C_j:

$$\mathbf{c_j} = \frac{1}{h} \sum_{l \in C_j}^{h} \mathbf{x}_l \tag{3.6}$$

4. Repeat steps 2 and 3 until a maximum number of iterations is reached or there are no changes at the centroid position recalculations in step 3.

When this algorithm is finished, all patterns are assigned to one of the k clusters and received a corresponding label. In Chapter 4, k-means clustering is applied to group PV systems according to their locations and specifications.

3.2. Support Vector Machine

The support vector machine classification algorithm was first introduced by Vapnik [51] who described the idea of finding a hyperplane in input space that groups patterns with different labels into classes while keeping the distance across classes at a maximum. The key is, that by maximizing the distance there are only finite possibilities to place a hyperplane in input space. Due to the finiteness, there has to be an optimal placement. Here, the prediction function is build by solving the optimization problem of finding the best pattern-separating hyperplane.

Section 3.2.1 starts by discussing the linear SVM for linear separable points. Afterwards, slack variables are introduced that allow a classification of patterns even if there is no strictly linear separation possible. This is followed by the description of non-linear SVMs in Section 3.2.2. The notations and definitions used throughout the next sections are based on Hastie et al. [23].

3.2.1. Linear Support Vector Machines

The main element of a SVM is a hyperplane which can be defined generally as

$$\{\mathbf{x} \in \mathbb{R}^d : f(\mathbf{x}) = \mathbf{x}^T \boldsymbol{\beta} + \beta_0 = 0\}, \tag{3.7}$$

where $\boldsymbol{\beta} \in \mathbb{R}^d$ denotes an unit vector (i.e., $||\boldsymbol{\beta}|| = 1$), with the distance $\beta_0 \in \mathbb{R}$ to the space's origin. $\boldsymbol{\beta}$ being an unit vector simplifies further definitions and is achieved by scaling $\mathbf{x}$ and β_0 accordingly.

The label y_i of a pattern $\mathbf{x}_i$ is assigned corresponding to its relative position to the hyperplane. For a binary decision with $y_i \in \{1, -1\}$, the decision function is specified as

$$y_i = \operatorname{sgn}(f(\mathbf{x}_i)) = \operatorname{sgn}(\mathbf{x}_i^T \boldsymbol{\beta} + \beta_0). \tag{3.8}$$

The decision function $f(\mathbf{x})$ calculates the signed distance of a data point $\mathbf{x}$ to the hyperplane. Thus, the label is assigned depending on which side of the hyperplane the pattern is located.

Linear Separable Input Space For now, only situations where a linear separation is possible are discussed. Hence, there exists at least one function $f(\mathbf{x}) = \mathbf{x}^T \boldsymbol{\beta} + \beta_0$, with $\forall i : y_i f(x_i) > 0$, that is splitting the labeled data points. From the set of valid decision functions, the SVM algorithm is searching for a function that creates the biggest margin between the training data points of the two different classes. In Figure 3.2 (a), the concept of a margin that surrounds the hyperplane is illustrated. The search for a optimal hyperplane is characterized by the optimization problem

$$\max_{\boldsymbol{\beta}, \beta_0, ||\boldsymbol{\beta}||=1} M, \\ \text{s.t.} \quad y_i(\mathbf{x}_i^T \boldsymbol{\beta} + \beta_0) \geq M, \ i = 1, \dots, N \tag{3.9}$$

M denotes the signed distance from the decision function to the patterns of one class. As stated before, this M is maximized, so that an optimal hyperplane separating the two classes can be found. The constraint $||\boldsymbol{\beta}|| = 1$ is removed by adding it to the side constraint:

$$\begin{aligned} &\frac{1}{||\boldsymbol{\beta}||} y_i(\mathbf{x}_i^T \boldsymbol{\beta} + \beta_0) \geq M \\ \Leftrightarrow\ & y_i(\mathbf{x}_i^T \boldsymbol{\beta} + \beta_0) \geq M||\boldsymbol{\beta}|| \\ \Leftrightarrow\ & y_i(\mathbf{x}_i^T \boldsymbol{\beta} + \beta_0) \geq 1 \end{aligned} \tag{3.10}$$

The last step is possible because for all $\boldsymbol{\beta}$ and β_0 that satisfy these inequalities, any scaled positively multiple is satifying them as well. Thus, $||\boldsymbol{\beta}||$ can be set to $1/M$ without any problems. This leads to a minimization problem that is equivalent to Equation (3.9):

$$
\min_{\boldsymbol{\beta},\beta_0} ||\boldsymbol{\beta}||,
$$
$$
\text{s.t. } y_i(\mathbf{x}_i^T\boldsymbol{\beta} + \beta_0) \geq 1, \ i = 1, \dots, N \tag{3.11}
$$

Now, by finding a feasible vector $\boldsymbol{\beta}$ and value for β_0 it is ensured, that all data points are assigned a correct label by the decision function of Equation (3.8).
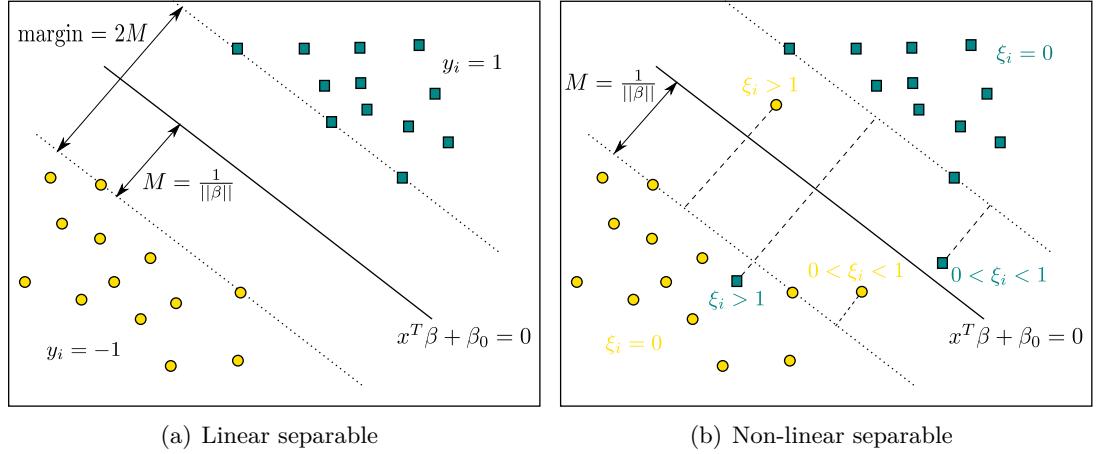
Figure 3.2.: *Illustration of a linear support vector machine for (a) linear separable and (b) non-linear separable data points.*

Non-linear Separable Input Space   With the above definition, it is only possible to split two classes if they are strictly separable. In reality, there are often situations where a sharp partitioning is not possible, i.e., the presence of outliers in the data. To deal with these kind of situations, a softening of the constraints is needed, so that patterns on the wrong side are allowed. This is achieved by adding so-called slack variables $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_N)$ to the side constraint:

$$
y_i(\mathbf{x}_i^T\boldsymbol{\beta} + \beta_0) \geq 1 - \xi_i, \ \forall i : \xi_i \geq 0, \sum_{i=1}^{N} \xi_i \leq c \tag{3.12}
$$

The value of every ξ_i indicates the depositioning of the labeled pattern $\mathbf{x}_i$ in reference to the hyperplane: $\xi_i = 0$ for patterns on the right side of the function, $0 < \xi_i < 1$ for data points on the margin, and $\xi_i > 1$ for patterns on the wrong side of the hyperplane

(see Figure 3.2 (b)). The user-defined parameter c sets an upper bound for positioning violations.

By using the steps from Equations (3.10) and adding the slack variables to the optimization problem of Equation (3.11)

$$\min_{\boldsymbol{\beta}, \beta_0} ||\boldsymbol{\beta}||,$$
$$\text{s.t. } y_i(\mathbf{x}_i^T \boldsymbol{\beta} + \beta_0) \geq 1 - \xi_i, \forall i, \ \xi_i \geq 0, \sum \xi_i \leq c. \tag{3.13}$$

is derived. Nonetheless, for convenience, the following form of the optimization problem

$$\min_{\boldsymbol{\beta}, \beta_0} \frac{1}{2} ||\boldsymbol{\beta}||^2 + C \sum_{i=1}^{N} \xi_i,$$
$$\text{s.d. } \xi_i \geq 0, \ y_i(\mathbf{x}_i^T \boldsymbol{\beta} + \beta_0) \geq 1 - \xi_i \ \forall i \tag{3.14}$$

is more common in literature. This is valid, as it is possible to scale $\boldsymbol{\beta}$ and β_0 as desired. Now, the slack is a fixed part of the minimization and the term $C \sum_{i=1}^{N} \xi_i$ controls how much patterns are misplaced in reference to their respectable classes. The cost parameter $C > 0$ regulates these misplaced data points: the greater C is choosen, the smaller the margin gets, so that less points are on the wrong side of the hyperplane.

Equation (3.14) is a classic formalization of a convex optimization problem or more detailed: a linearly constrained optimization problem with a quadratic objective function. One benefit of convex optimization problems is that local optima must be global minima, which eases the search. A basic way to handle problems of this form is to redefine them using Lagrange multipliers and solve the new optimization problem with quadratic programming. Applying Lagrange multipliers to Equation (3.14) leads to

$$L = \frac{1}{2} ||\boldsymbol{\beta}||^2 + C \sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \alpha_i \left[y_i(\mathbf{x}_i^T \boldsymbol{\beta} + \beta_0) - (1 - \xi_i) \right] - \sum_{i=1}^{N} \mu_i \xi_i, \tag{3.15}$$

where $\boldsymbol{\beta}, \beta_0$, and ξ_i are minimized while the newly introduced Lagrange multipliers α_i and μ_i need to be maximized. Partially deriving the former variables results in

$$\boldsymbol{\beta} = \sum_{i=1}^{N} \alpha_i y_i \mathbf{x_i}, \tag{3.16}$$

$$0 = \sum_{i=1}^{N} \alpha_i y_i, \tag{3.17}$$

$$\alpha_i = C - \mu_i, \forall i, \tag{3.18}$$

31

with α_i, μ_i, and $\xi_i \geq 0 \; \forall i$. Here, the convenient notation of Equation (3.14) helps to derive Equation (3.15) more easily. By applying Equations (3.16) and (3.18) on Equation (3.15) the dual optimization problem

$$\max_{0 \leq \alpha_i \leq C} \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j$$

$$\text{s.t.} \sum_{i=1}^{N} \alpha_i y_i = 0. \tag{3.19}$$

results. To find optimal solutions α_i*, Karush-Kuhn-Tucker conditions are introduced, which complement the Conditions (3.16) – (3.18) into

$$\alpha_i [y_i (\mathbf{x}_i^T \boldsymbol{\beta} + \beta_0) - (1 - \xi_i)] = 0, \tag{3.20}$$

$$\mu_i \xi_i = 0, \tag{3.21}$$

$$y_i (\mathbf{x}_i^T \boldsymbol{\beta} + \beta_0) - (1 - \xi_i) \geq 0, \tag{3.22}$$

with $i = 1, \ldots, N$. The solutions of the Karush-Kuhn-Tucker conditions are sufficiently optimal for quadratic programming, and furthermore equal to the global optimum, as the original optimization problem is convex with only linear constraints. All solutions $\alpha_i^* > 0$ of patterns $\mathbf{x}_i$ that satisfy Condition (3.20) while respecting Condition (3.22) are called support vectors. Hence, support vectors are training patterns that are located on the margin or on the opposite side of the hyperplane (see Equation (3.22)). These vectors enable a solving of the initial problem (3.14) as an optimal solution of $\boldsymbol{\beta}$ is given by

$$\boldsymbol{\beta}^* = \sum_{i=1}^{N} \alpha_i^* y_i \mathbf{x_i} \tag{3.23}$$

which is directly derived from Equation (3.16). The only thing left to find is a solution β_0^*. For this, all training patterns with $\xi_i^* = 0$ that as well comply to $0 < \alpha_i^* < C$ (see Constraints (3.21) and (3.18)) are applicable. Thus, these are data points that are located exactly on the edge of the margin. After finding all solutions, the class of a new pattern $\mathbf{x}'$ is set by

$$y_i = \text{sgn}[f^*(\mathbf{x}')] = \text{sgn}[\mathbf{x}'^T \boldsymbol{\beta}^* + \beta_0^*]. \tag{3.24}$$

3.2.2. Non-linear Support Vector Machines

Until now, the input space was always assumend to be separable by a linear function, but in many fields this is not the case. To achieve reasonable classifications in non-linear

spaces, the decision function needs to be changed. The SVM is able to adjust its decision function by introducing a so-called feature mapping function Φ. This Φ function maps pattern from the non-linear separable space to a higher dimensional feature space $\mathcal{H}$ where a linear separation is possible.

Going back to the dual optimization problem in Equation (3.19), the pattern $\mathbf{x}'$ are only present in form of a dot product. This helps at integrating the feature map Φ as it is only necessary to replace this part with a dot product containing the mapping to a higher dimensional space. This leads to the dual optimization problem for non-linear SVMs:

$$\max_{0 \leq \alpha_i \leq C} \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle$$
$$\text{s.t. } \sum_{i=1}^{N} \alpha_i y_i = 0. \tag{3.25}$$

To avoid the explicit calculation of $\Phi : \mathbb{R}^d \to \mathcal{H}$, the mapping is induced by a kernel function $\mathcal{K}$ with $\mathcal{K}(\mathbf{x}, \mathbf{x}') = \langle \Phi(\mathbf{x}), \Phi(\mathbf{x}') \rangle$. This technique is known as the kernel trick [25].

Hence, it is only necessary to apply the kernel function on the patterns and calculate the dot product afterwards instead of calculating a dot product in a higher dimension. One restriction is that the kernel matrix $\mathbf{K} \in \mathbb{R}^{n \times n}$ used in the quadratic programming is positive-semidefinite, meaning that the condition $\mathbf{x}^T \mathbf{K} \mathbf{x} \geq 0$ must apply. With this, it is ensured that it is possible to separate the data points in the higher dimensional space.

Three common kernel function that satisfy this constraint are:

$$\text{Linear kernel: } \mathcal{K}_{lin}(\mathbf{x}, \mathbf{x}') = \langle \mathbf{x}, \mathbf{x}' \rangle \tag{3.26}$$

$$\text{Polynomial kernel: } \mathcal{K}_{poly}(\mathbf{x}, \mathbf{x}') = \langle \mathbf{x}, \mathbf{x}' \rangle^p, p \in \mathbb{N} \tag{3.27}$$

$$\text{Radial basis function (RBF): } \mathcal{K}_{rbf}(\mathbf{x}, \mathbf{x}') = \exp(-\gamma ||\mathbf{x} - \mathbf{x}'||^2), \ \gamma \in \mathbb{R}, \ \gamma > 0 \tag{3.28}$$

Hereby, the linear kernel is equivalent to the linear SVM explained above. The increase of the computational complexity is now completely dependent on the complexity of the kernel function and the dimensionality of the original input space. More detailed information on kernel functions can be found in Hastie et al. [23].

Adding the kernel function to the dual optimization problem of Equation (3.25) results in

$$
\max_{0 \leq \alpha_i \leq C} \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \mathcal{K}(\mathbf{x}_i, \mathbf{x}_j)
$$
$$
\text{s.t. } \sum_{i=1}^{N} \alpha_i y_i = 0. \tag{3.29}
$$

and the label of a new pattern $\mathbf{x}'$ is determined with the associated decision function as follows:

$$
\begin{aligned}
y_i = \text{sgn}[f^*(\mathbf{x}')] &= \text{sgn}[\Phi(\mathbf{x}')^T \boldsymbol{\beta}^* + \beta_0^*] \\
&= \text{sgn}[\sum_{i=1}^{N} \alpha_i^* y_i \langle \Phi(\mathbf{x}'), \Phi(\mathbf{x}_i) \rangle + \beta_0^*] \\
&= \text{sgn}[\sum_{i=1}^{N} \alpha_i^* y_i \mathcal{K}(\mathbf{x}', \mathbf{x}_i) + \beta_0^*].
\end{aligned} \tag{3.30}
$$

3.3. Support Vector Regression

In this section, the changes that are made to the SVM to use it for regression modeling are explained. The support vector regression was first introduced by Smola and Schölkopf [50] in 2004. Again, the key idea of this algorithm is to find an optimal hyperplane in the input space. But instead of dividing different classes, the SVR searches for a hyperplane that is the best representation of all points in the training dataset while submitting to user-defined constraints.

To get from the classification to the regression model, Equation (3.14) is adapted to the non-linear case by including the feature mapping function Φ from Section 3.2.2:

$$
\min_{\boldsymbol{\beta}, \beta_0} \frac{1}{2} ||\boldsymbol{\beta}||^2 + C \sum_{i=1}^{N} \xi_i,
$$
$$
\text{s.t. } \xi_i \geq 0, \ y_i(\Phi(\mathbf{x}_i)^T \boldsymbol{\beta} + \beta_0) \geq 1 - \xi_i \ \forall i \tag{3.31}
$$

While searching for solutions to this optimization problem, there are only two decisive states for the slack variables ξ_i that are interesting in the optimization process. These two cases can be directly derived from the side constraints: Either $\xi_i = 0$, which relates to no violations, or the highest possible value for violations by even crossing the margin on the other side is decisive, in which case $\xi_i = 1 - y_i(\Phi(\mathbf{x}_i)^T \boldsymbol{\beta} + \beta_0)$ holds. Thus, the side constraints can be simplified to $\xi_i = \max(0, 1 - y_i(\Phi(\mathbf{x}_i)^T \boldsymbol{\beta} + \beta_0))$ and added to the

optimization problem:

$$\min_{\boldsymbol{\beta},\beta_0} \frac{1}{N} \sum_{i=1}^{N} \max(0, 1 - y_i(\Phi(\mathbf{x}_i)^T \boldsymbol{\beta} + \beta_0)) + \lambda ||\boldsymbol{\beta}||^2, \quad (3.32)$$

with $\lambda = 1/2NC$. With this definition, for all data points that are on their right side of the hyperplane, the maximum function returns 0. Otherwise there will be a punishment for constaint violations. A more common definition of the SVM classification problem is utilizing the hinge loss $\mathcal{L}_h(\omega, \tau) = \max(0, 1 - \omega\tau)$. Applying $\omega = y_i$ and $\tau = \Phi(\mathbf{x}_i)^T \boldsymbol{\beta} + \beta_0$ to Equation (3.32) leads to:

$$\min_{\boldsymbol{\beta},\beta_0} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}_h(y_i, \Phi(\mathbf{x}_i)^T \boldsymbol{\beta} + \beta_0) + \lambda ||\boldsymbol{\beta}||^2 \quad (3.33)$$

Now, the introduction of SVM to a continuous input space, i.e., $y \in \mathbb{R}$, can be done. For the support vector regression, Smola and Schölkopf [50] use yet another loss function called ϵ-insensitive loss

$$\mathcal{L}_\epsilon(\omega, \tau) = \max(0, |\omega - \tau| - \epsilon), \ \epsilon \in \mathbb{R}, \ \epsilon > 0. \quad (3.34)$$

With the ϵ-insensitive loss, a tube around the hyperplane is created where no punishment for constraint violations is applied (see Figure 3.3). This minimizes the possibility that the SVR function is overfitted to the training dataset, given an appropriate choice for the value of ϵ.

Finally, the support vector regression optimization problem is defined as

$$\min_{\boldsymbol{\beta},\beta_0} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}_\epsilon(y_i, \Phi(\mathbf{x}_i)^T \boldsymbol{\beta} + \beta_0) + \lambda ||\boldsymbol{\beta}||^2. \quad (3.35)$$

By using the solutions $\boldsymbol{\beta}^*$ and β_0^* of this problem, labels for new patterns $\mathbf{x}'$ are assigned analogous to the classification task. For a more detailed description of the support vector regression, the original work of Smola and Schölkopf [50] or Hastie et al. [23] is recommended. In this thesis, the SVR implementation of the Python package scikit-learn [48], which is based on LIBSVM [8], is utilized.

3.4. Concluding Remarks

At the beginning of this chapter, machine learning basics, including the automatic learning of functions that map values from input to output spaces, is introduced. Values
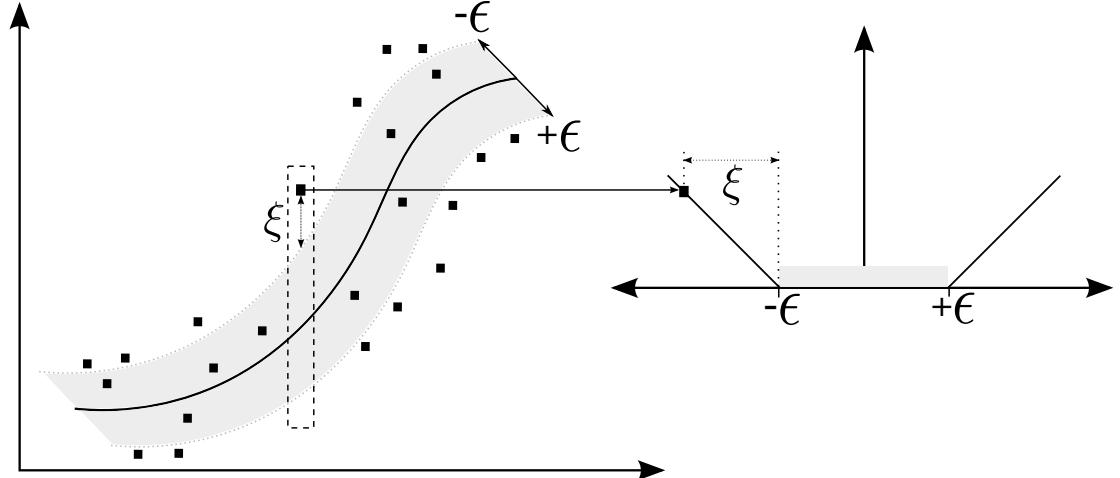
Figure 3.3.: ***Left:*** *Linear SVR with ϵ tube.* ***Right:*** *ϵ-insensitive loss. Data points within the ϵ tube (grey area) are ignored for error punishment calculations. If a data point is more than ϵ away from the function, the punishment is applied linearly (compare Smola and Schölkopf [50]).*

of the input space are called patterns and their opposites in output space are labels. Throughout this work, the term features is used to describe single components of a pattern. For training, testing, and evaluation of classification and regression models, there are corresponding datasets that are introduced as well. With the knowledge about training datasets, a distinction between supervised and unsupervised learning is possible.

Supervised methods use known associations between patterns and labels of a training dataset to learn mapping functions. These functions are later on used to assign appropriate labels to new patterns. To give a simple example of a classification and regression model, the supervised learning algorithm k-nearest neighbors is explained. As this work mainly uses supervised learning methods, there is only a short example of one unsupervised learning method given: k-means clustering. With k-means clustering, it is possible to classify patterns into k groups (cluster) and assign corresponding labels to them, even though there is no training set with known associations to learn from.

On this basis, the second part of this chapter focuses on deriving the fundamentals of support vector machines (SVM). Starting at the linear SVM model's optimization problem for linear separable input spaces, the idea of constructing a hyperplane for the classification of patterns is explained. Based on this, the classification of non-linear separable spaces by applying slack variables is described.

In many cases, a linear SVM is not applicable as the structure of the input space is too complex, therefore a non-linear extension of the SVM is needed. By adding so-

called kernel functions to the model, the optimization problem is projected into a higher dimensional input space, in which a linear separation is possible. Now, the SVM model is able to classify patterns of even highly complex data spaces and assign appropriate labels, complying to the two user-defined parameters C and γ, that regulate the complexity of the learned function.

The last section of this chapter is explaining the changes that are necessary to create regression functions using the SVM. Instead of discrete values, the data spaces are now continuous real-valued. The extension of SVM to regression modeling is called support vector regression (SVR). In this transition process, the ϵ-insensitive loss is integrated into the optimization problem, and with it a new user-defined parameter ϵ is introduced.

Throughout this work, SVR is used as the main modeling technique for creating regression functions for PV power forecasting. One advantage of using SVR is its ability to learn even non-linear functions, which are assumed to emerge in this scenario. In the following chapter, the SVR model is optimized for its PV power forecasting task and compared to the physical modeling described in Section 2.3. In this process, the three user-defined parameters of the SVR are tuned and appropriate training datasets are specified.

Part II.

SVR-based Prediction

4. Comparison to a Parametric Model

One goal of this thesis lies in assessing the general usefulness of machine learning, and especially support vector regression (SVR) as a modeling technique for PV power forecasting. To get an impression of its capabilities, at the end of this chapter, SVR is compared to a physical parametric modeling approach which is detailed in Chapter 2. This model uses irradiance predictions of different sources, i.e., numerical weather prediction (NWP) models and satellite-based cloud motion vectord (CMV), and converts them to power outputs of PV systems. Finally, all single outputs of the physical modeling methods are combined, which also includes a persistence model using historic PV measurements, to receive a forecast for future power outputs.

This evaluation aims at giving a first expressive comparison of machine learning and phyical modeling, so that as many PV systems as possible are featured. Before SVR settings can be concluded, the test set size needs to be reduced to get results in a feasible time frame. This is done in Section 4.2, where a dataset for the optimization process is chosen. Due to the high number of PV systems, a general setting for SVR parameters is applied, which will be determined in Section 4.3. Afterwards in Section 4.4, different SVR forecasts that represent single input forecasts of the parametric model are analyzed. In Section 4.5, the SVR-based models are finally compared to the physical modeling approaches. The results of this chapter are published in Wolff et al. [54].

4.1. Experimental Setup

At the time of this comparison, the data of the years 2012 and 2013 of 921 PV systems is available, which is quality checked as described in Section 2.2.1. To enable a reasonable comparison to the physical PV power modeling, the SVR models strictly use the same input data as the physical forecast models (see Section 2.2) in this chapter. Thus, the

4. Comparison to a Parametric Model

following training pattern-label mappings are applied:

$$P_{SV,meas} := (P_{meas}(t - \Delta t), H(t)) \rightarrow P_{meas}(t), \tag{4.1}$$

$$P_{SV,NWP} := (I_{NWP}(t), T_{NWP}(t), H(t)) \rightarrow P_{meas}(t), \tag{4.2}$$

$$P_{SV,CMV} := (I_{CMV}(t), T_{NWP}(t), H(t)) \rightarrow P_{meas}(t), \tag{4.3}$$

for timestep t with forecast horizon Δt, PV power measurements P_{meas}, CMV-based irradiance forecasts I_{CMV}, NWP-based temperature (T_{NWP}) and irradiance (I_{NWP}) forecasts. Furthermore, including the hour of the day (H) helps to capture the deterministic diurnal course of irradiance without applying any additional models, like, e.g., a clear sky irradiance or PV power model. In addition to the single input forecast models, another training mapping

$$P_{SV,comb} := (P_{meas}(t - \Delta t), I_{NWP}(t), I_{CMV}(t), T_{NWP}(t), H(t)) \rightarrow P_{meas}(t) \tag{4.4}$$

is applied that combines all of the above features. The training process is illustrated in Figure 4.1. The models are retrained every day with the data of the previous N
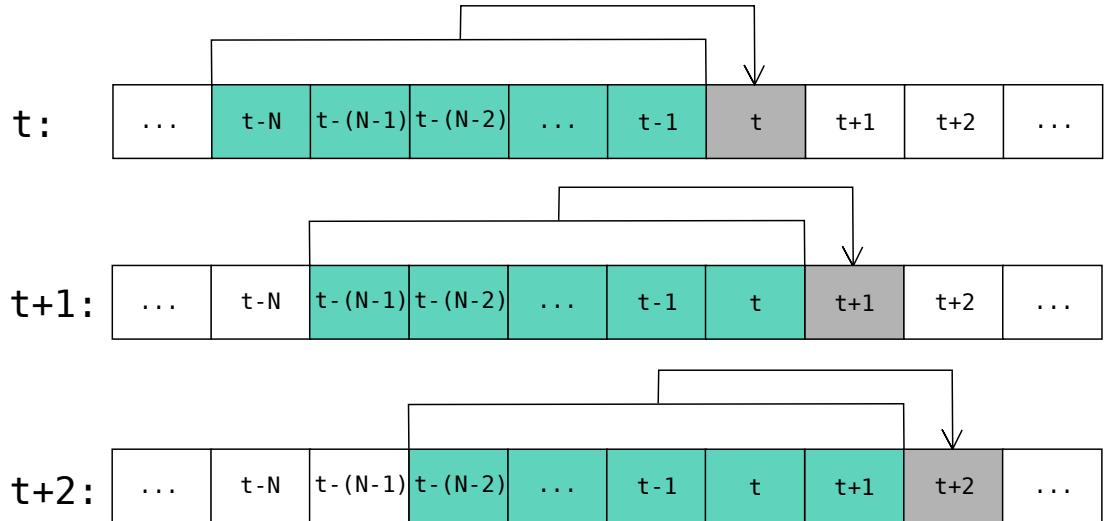
Figure 4.1.: *Sliding window training of support vector regression. The training set containing N days of data is denoted in green; the grey boxes show the days t, $t + 1$, and $t + 2$ that are forecasted.*

days. Each of the green boxes in Figure 4.1 represents pattern and label matchings from Equations (4.1) – (4.6) of the corresponding day for every 15 minutes.

Data Preprocessing To match the temporal and spatial resolution of the PV power measurements, the data is interpolated to 15 minute timesteps, as described in Lorenz and Heinemann [35] or Lorenz et al. [37]. Further preprocessing steps are applied to the data, which are detailed in Lorenz et al. [39].

As there are varying units and ranges for the different datasets, that could encourage an unwelcome weighting of input features, the data is normalized in a way that only real-valued numbers ranging from 0 to 1 are handled by the regression models. As the PV measurements are already normalized to the nominal power of the respective PV systems, only irradiance and temperature forecasts need to be considered. All irradiance forecasts are normalized by $I_{forec}(t) = \frac{I_{forec}(t)}{I_{const}}$, with $forec \in \{CMV, NWP\}$ and the solar constant I_{const} being $1361\frac{W}{m^2}$ [28]. The temperature forecast of the ECMWF's NWP model is normalized by

$$T_{NWP}(t) = \frac{T_{NWP}(t) - T_{min}}{T_{max} - T_{min}}, \tag{4.5}$$

where $T_{min} = -26.71°C$ and $T_{max} = 35.83°C$ are the lowest and highest temperature values found in the 2012 dataset.

4.2. Selection of an Optimization Dataset

To increase the quality of PV predictions, the SVR model and the applied RBF kernel function has to be adjusted to this specific task by optimizing the parameters C, ϵ, and γ as well as the number of days used for training. The optimization process is done on the 2012 data, so that the data from 2013 stays solely for evaluation purposes. With respect to computation time, it is not possible (in a reasonable time frame) to perform an optimization on all 921 PV systems for all timesteps. Hence, a subset of systems and test days, that represent different variability scenarios of the complete dataset, are selected. Here, the aim is to find one set of parameters which results in high quality predictions for all PV systems, seasons, and times of the day.

Selection of Test Systems First, 92 PV systems (10% of all systems) are selected by clustering with *k*-means introduced in Section 3.1. The selection is separated in two clustering steps:

1. Twelve location clusters are formed, so that a spatial distribution is achieved. For this, only latitude and longitude information of each system is used (see Figure 4.2(a)).

(a) (b)
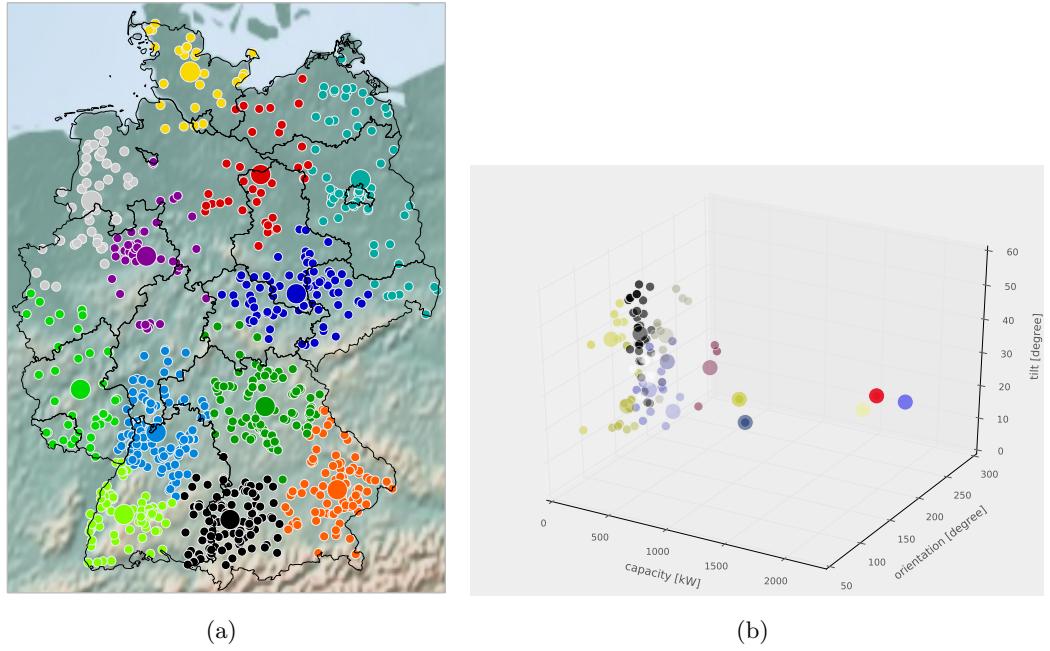
Figure 4.2.: *(a) PV systems after applying k-means clustering (k = 12) on locations. Small dots show the location of a single PV system. (b) Example of one of the 12 subcluster for PV system specifications, where small dots indicate a single system's specification (orientation: 180 degree = south, tilt: 0 degree = horizontal). In both figures, large dots point to the center of the respective cluster.*

2. Each of these twelve clusters is divided into a number of subclusters referring to the PV systems' specifications, namely capacity, orientation, and tilt (see Figure 4.2(b)). The number of subclusters depends on the size of the location cluster and corresponds to 10% of its included PV systems.

In the end, 10% of the PV systems in each subcluster are chosen randomly, resulting in a test set of 92 PV systems. In Figure 4.3, the distribution of PV systems in respect to the specification features power capacity, orientation, and tilt are shown. Additionally, the distributions of 100 randomly sampled PV system sets are illustrated. While only some of the random sampled test sets are accidentally fitting the specification distribution of the complete dataset, the controlled clustering approach selects a list of PV systems that fits the original's density functions for all employed specifications.

Selection of Test Days   After the test systems are selected, a set of optimization test days from these PV systems is constructed. Only days from the second half of the 2012

data are considered, thus there are enough days left for SVR model training purposes and for the analysis of suitable training set sizes.

(a) power capacity

(b) orientation
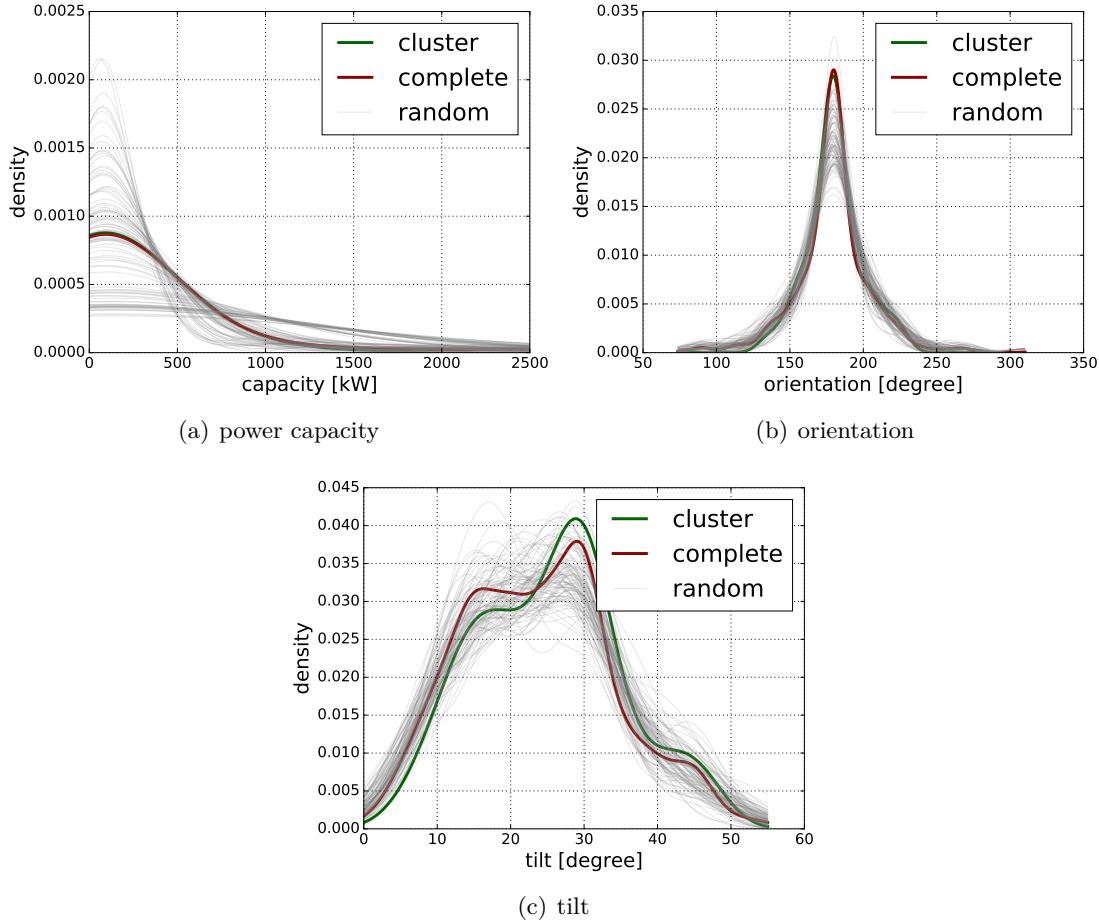
(c) tilt

Figure 4.3.: *Distribution of PV system datasets with respect to the three PV system specifications (a) power capacity, (b) orientation, and (c) tilt. Besides the complete PV system dataset's distributions (complete), the density functions of PV systems selected by the clustering approach (cluster) are shown. Furthermore, the distribution of 100 randomly selected sets (random) is illustrated.*

From July to December 2012, one representative day per month is chosen for each of the following five classes: clear sky, broken cloud, overcast and two classes in between, which results in a test dataset of 30 days. For this, the clear sky irradiance is estimated with a clear sky model (see Section 2.1). Afterwards, the PV simulation model introduced in Section 2.3.1 converts clear sky irradiance to PV power $P_{Sim,clear}$, which is then normalized to the nominal power of the respective PV system. The classification of a certain day is decided by its RMSE value, comparing the measurement and clear sky

power time series. With more clouds blocking the irradiance reception of the PV modules,
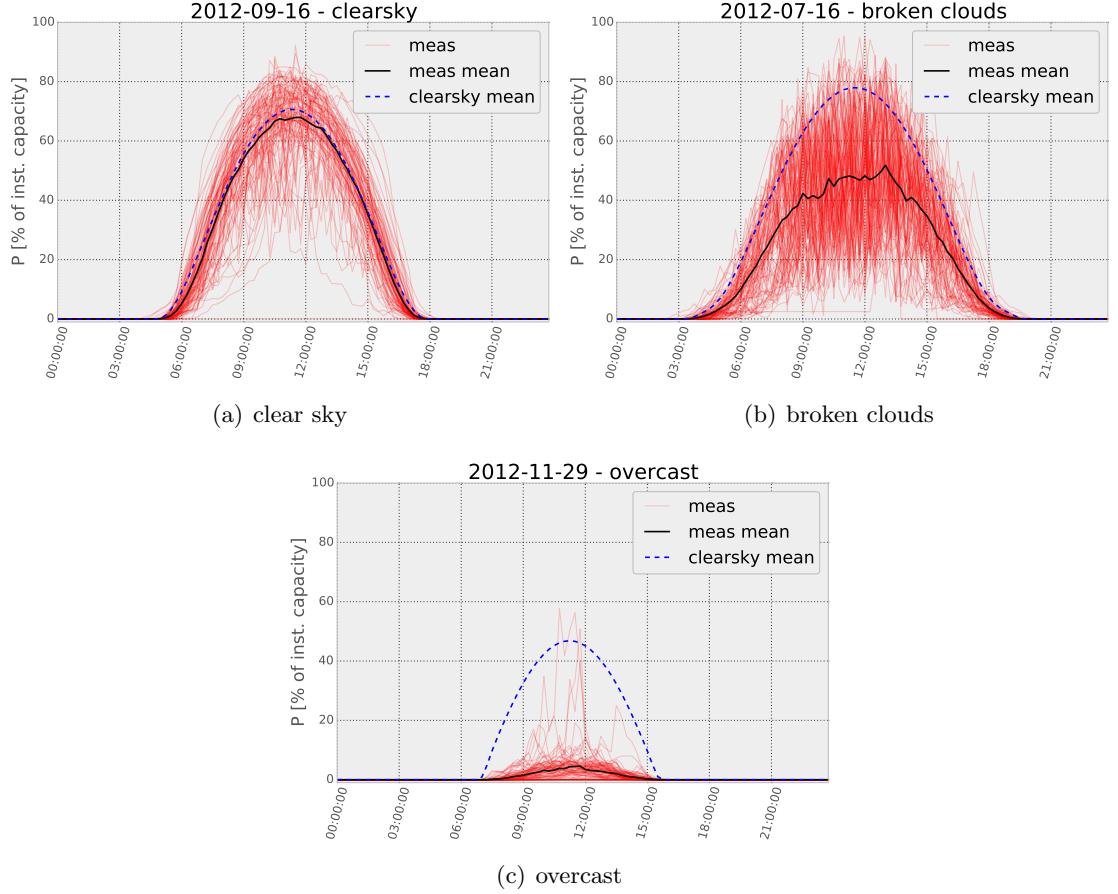
Figure 4.4.: *Example days from the test dataset with (a) clear sky, (b) broken clouds and (c) overcast conditions. The power output P is stated in percentage of the corresponding installed capacity of each PV system.*

the RMSE values increase, so that high RMSE values indicate days with an overcast situation, while days with a low RMSE can be classified as clear sky days. After sorting the days according to their resulting error values and splitting the days in five equally sized groups, one day from each group is randomly picked. A set of example days of the three main scenarios, i.e., clear sky, broken clouds and overcast situations is shown in Figure 4.4. In some cases the measured power is higher than the expected mean clear sky power, which can be caused by an overestimation of the turbidity or inaccurate PV system specifications.

In summary, 30 days are selected – five days from each of the six months – with different weather conditions from the 2012 dataset to be used as a testing set for the SVR parameter optimization.

4.3. Optimization of Model Parameters

Now, after the PV systems and 30 test days are selected, the SVR is optimized for PV power forecasting. Even though the computational time is already reduced due to applying only 10% of the PV systems for the tuning of SVR parameters, the optimization process is divided in two steps, as a simultaneous search is not feasible:

1. Selection of an appropriate training dataset size with standard settings for user-defined SVR parameters C, γ, and ϵ.

2. Optimization of C, γ, and ϵ with the training dataset size selected in Step 1.

For the selection of training set sizes, training sets including data of one to 150 days are tested. The results of this optimzation step are listed in Table 4.1. The optimal training set size, w.r.t. RMSE values, for $P_{SV,CMV}$ are not changing for the different forecast horzions, as the same forecast run is used throughout the whole day. For all other models, there are new patterns every 15 minutes. The exact numbers of days of the best forecasts are more or less random, whereas a small increase of training days for longer forecast horizons is noted (see $P_{SV,CMV}$ and $P_{SV,comb}$). At least one month of data should be available for all models. There is no model that needs more than two and a half months of data, even though a maximum of 150 days is available. After finding

	Forecast Horizon							
	15min	30min	45min	1h	2h	3h	4h	5h
$P_{SV,meas}$	61	38	38	38	38	38	38	38
$P_{SV,NWP}$	52							
$P_{SV,CMV}$	39	39	39	45	50	46	50	50
$P_{SV,comb}$	65	53	48	46	50	50	52	52

Table 4.1.: *Optimal training set sizes in days, w.r.t. RMSE values, for every employed SVR-based forecast in days.*

the optimal training set sizes for each model, the SVR parameters are adjusted. Thus, a grid-search is applied covering 150 parameter combinations in the range of $[10^0, 10^5]$ for C, and $[10^{-5}, 10^{-1}]$ for γ and ϵ. These limits are choosen based on experience from previous evaluations in Wolff et al. [55]. The parameter optimization process is repeated

for all implemented patterns defined in Equations (4.1) - (4.6), and for all considered forecast horizons. To illustrate the parameter search, the RMSE results of the grid-search for $P_{SV,comb}$ are presented in Figure 4.5. For both shown forecast horizons of (a) 15 minutes and (b) five hours, there are wide ranges of parameters that achieve good results, whereas combinations with low values for C and concurrently low values for γ are not advisable.

(a) 15 minutes ahead   (b) 5 hours ahead
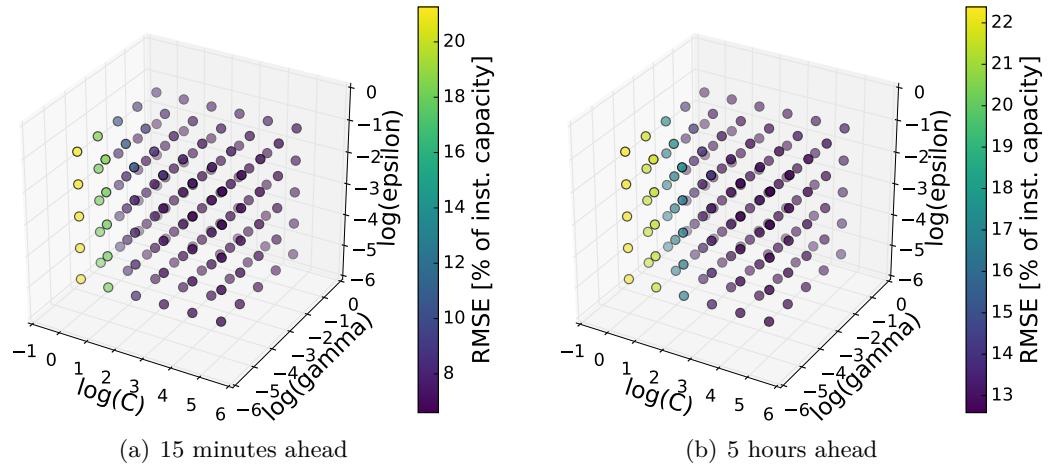
Figure 4.5.: *Grid-search results of C, γ, and ϵ of the combined model $P_{SV,comb}$ for forecast horizons of (a) 15 minutes and (b) five hours ahead.*

As seen in these figures, the influence of ϵ on the forecast quality is neglectable, so that in Figure 4.6, $\epsilon = 0.01$ is set to show the behavior of C and γ in more detail. This value is selected because it appears in all of the best parameter combinations for $P_{SV,comb}$. Parameter combinations resulting in forecasts with the lowest RMSE are marked with red circles in Figure 4.6. In both figures, there is an indication of a linear trend in the interaction of C and γ visible. The lower the values for C, the bigger γ needs to be and vice versa. This interaction is more distinct when looking at the raw RMSE values. Another trend, that appears in the parameter search of all models, is the reduction of C for longer forecast horizons. In this example, the only parameter of the optimal combination, that is changing from a forecast horizon of 15 minutes to five hours, is C. This indicates, that for longer forecast horizons the models need to be more tolerant and less adapted to the training data than for short horizons. This corresponds, at least to some extent, to the increase of the underlying uncertainty these models are trained for.

While the training set sizes vary only between 50 to 65 days for the combined SVR model, its optimal SVR parameters are sensitive to the individual patterns and differ with the forecast horizon. The forecasts in the following sections are performed using the

optimal training set size and SVR parameter set for each individual pattern and forecast horizon configuration.

(a) 15 minutes ahead   (b) 5 hours ahead
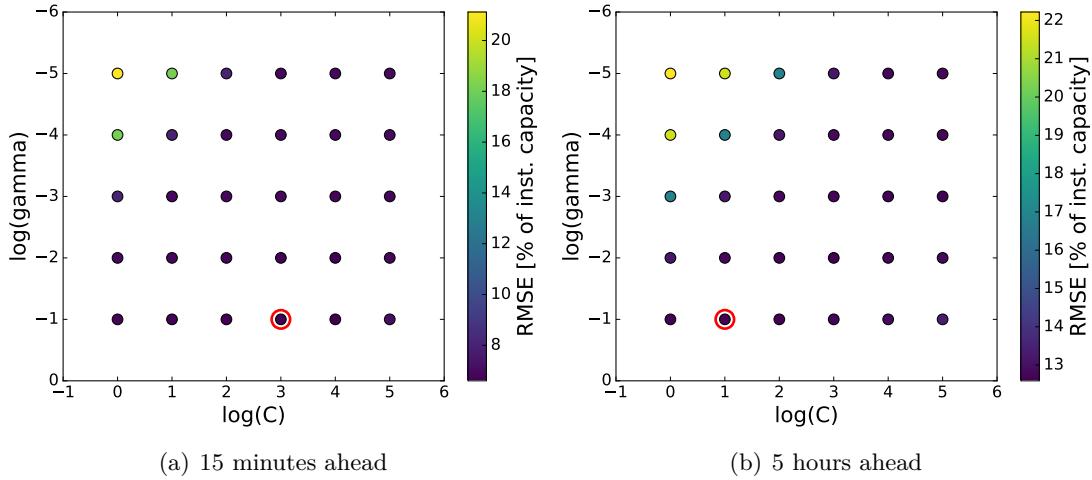
Figure 4.6.: *Grid-search results of C and γ with ε = 0.01 for the combined model $P_{SV,comb}$ on forecast horizons of (a) 15 minutes and (b) five hours ahead. Red circles mark the optimal parameter combination, w.r.t., RMSE values.*

4.4. Comparison of Different SVR-based Models

In this section, the performance of the four SVR-based forecasts introduced in Equations (4.1) - (4.6) are compared against each other. Apart from single system analysis, regional forecasts are analyzed. These regional forecasts are modeled by first calculating the mean of all of the employed single PV systems' patterns and train the SVR models on the averaged input features. For this, parameter optimization and model training is also performed on these averaged time series. Here, only timesteps are used where at least one third of the PV systems' data is available. The PV simulation approach needs information about single PV systems, so that single systems are simulated and the average power output of all single PV simulation forecasts is averaged into a regional forecast.

All of the following comparisons are solely done on the data from March to November 2013 with measurements of 921 PV systems. The other months are neglected due to long periods of snow fall in January, February, and December, causing problems for all models.

(a) Prediction horizon: 15 minutes

(b) Prediction horizon: 15 minutes

(c) Prediction horizon: 2 hours

(d) Prediction horizon: 2 hours

(e) Prediction horizon: 4 hours

(f) Prediction horizon: 4 hours
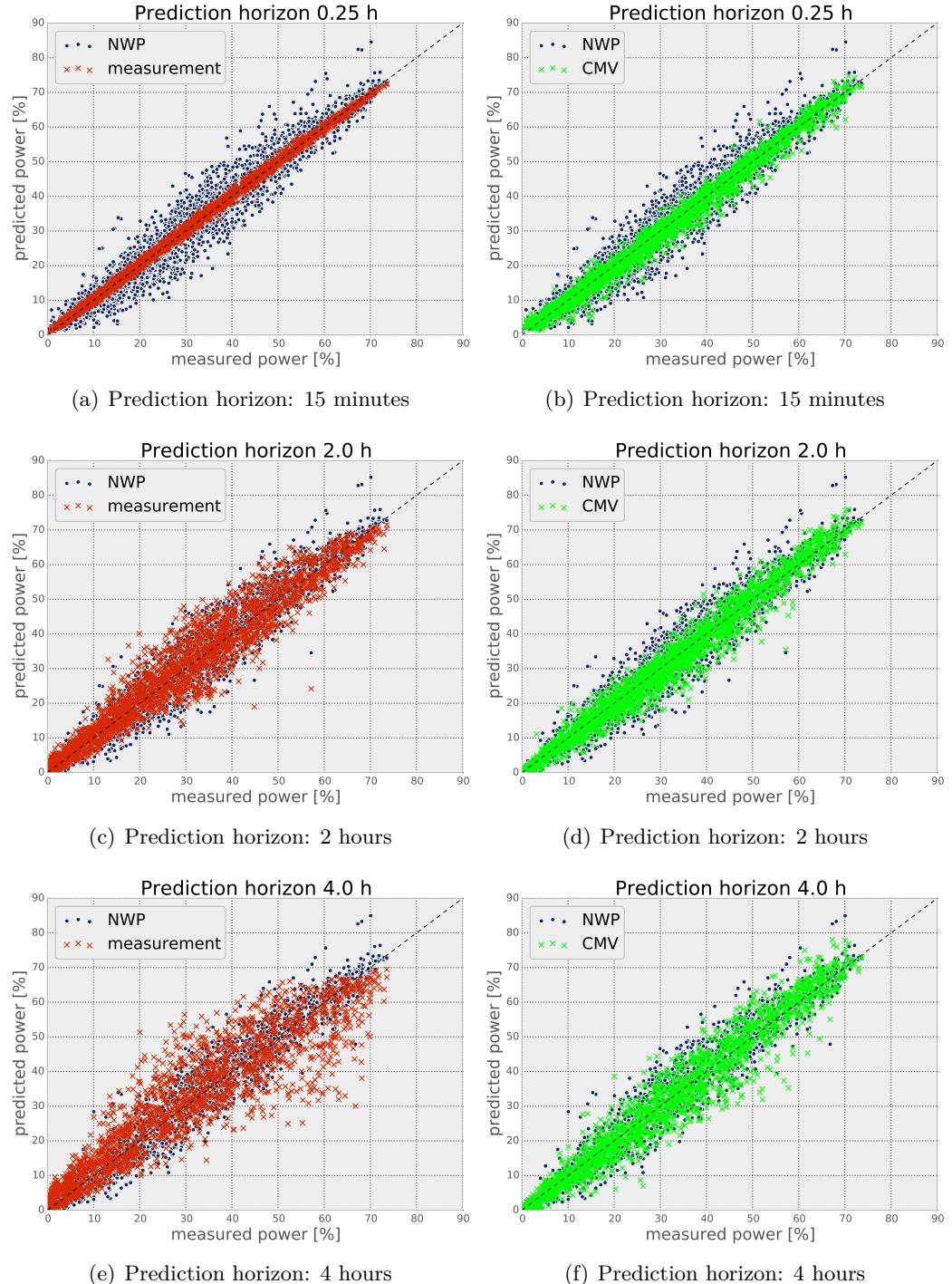
Figure 4.7.: *Comparison of regional SVR forecasts $P_{SV,meas}$ (measurement), $P_{SV,NWP}$ (NWP), and $P_{SV,CMV}$ (CMV) for different prediction horizons on the test set data.*

Analysis of Input Data To get an impression of the general behavior of the different input sources, Figure 4.7 shows the performance of each SVR-based single input forecast model on the regional level. Because NWP forecasts are updated only once a day, the NWP-based predictions stay the same for all forecast horizons, apart from slight differences due to filtering, that is applied to ensure that the same timesteps are used for all models. On the left side of the figure, NWP- ($P_{SV,NWP}$) and measurement-based ($P_{SV,meas}$) forecasts are compared. While $P_{SV,meas}$ is doing well for a prediction horizon of 15 minutes, the accuracy decreases fast with an increasing horizons. On the right hand side, $P_{SV,CMV}$ is not as accurate as $P_{SV,meas}$ for 15 minutes ahead, but for increasing forecast horizons it is more stable. Starting at a forecast horizon of four hours ahead, both forecast models are inferior to $P_{SV,NWP}$.

(a) single systems   (b) regional average
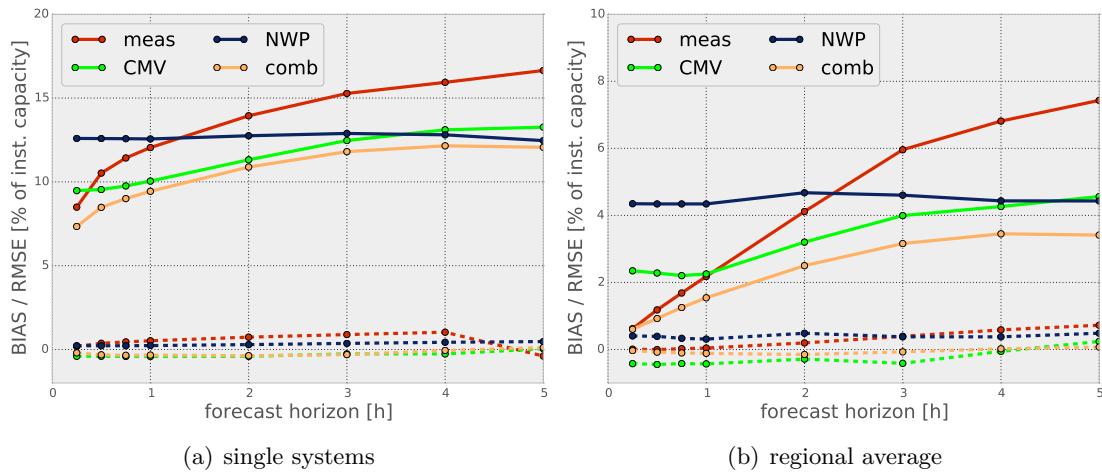
Figure 4.8.: *Comparison of SVR forecasts $P_{SV,meas}$ (meas), $P_{SV,CMV}$ (CMV), $P_{SV,NWP}$ (NWP), and $P_{SV,comb}$ (comb) depending on forecast horizons with RMSE (solid lines) and BIAS (dashed lines) for (a) single systems and (b) regional averages.*

In Figure 4.8, the results for $P_{SV,meas}$, $P_{SV,CMV}$, and $P_{SV,NWP}$ on all forecast horizons are shown for (a) single systems and (b) regional averaged forecasts. The most noticeable difference between single and regional forecasts is the magnitude of error. The errors of regional forecasts are reduced to around one third of the single site forecasts due to averaging effects. While forecasts based solely on measurement data are best for shortest-term forecasts, they rapidly decrease in accuracy with increasing forecast horizons. At a horizon of about 30 minutes (single systems) and one hour (regional), CMV-based SVR forecasts achieve a lower RMSE. The usage of NWP irradiance forecasts as input data becomes more favorable with longer forecast horizons starting at about 3.5 (single systems) and 4.5 hours (regional). The BIAS of all three forecasts is very low, because

systematic errors are learned by the SVR, so that over- and underestimations are reduced [23]. Furthermore, learning on averaged time series, instead of averaging the single site forecast model output, results in slightly better RMSE values for the SVR in these experiments.

(a) single system 15 minutes ahead

(b) single system 2 hours ahead

(c) regional average 15 minutes ahead

(d) regional average 2 hours ahead
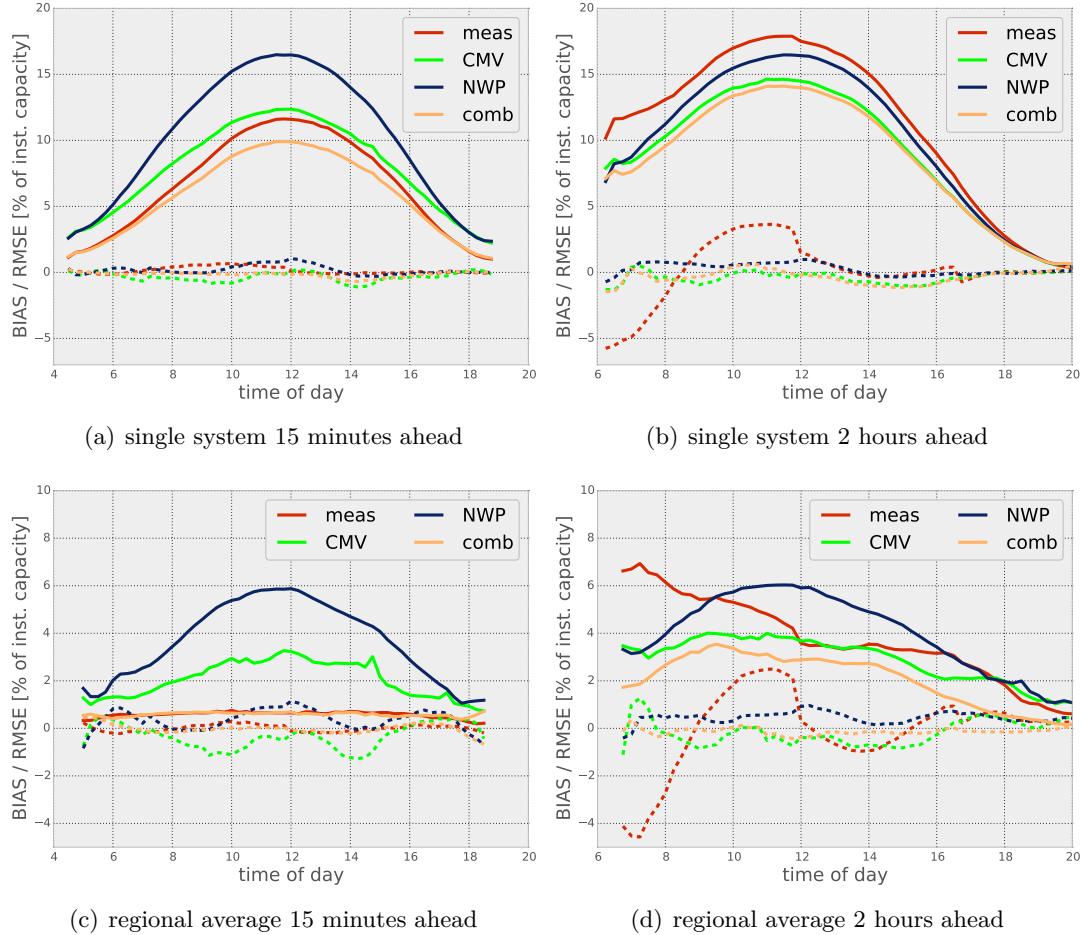
Figure 4.9.: *Single systems (top) and regional average (bottom) prediction quality of different SVR forecasts for each quarter-hour of the day with a forecast horizon of 15 minutes (left) and two hours (right). Prediction quality is measured by RMSE (solid lines) and BIAS (dashed lines).*

Combination of Input Data   After discussing the forecasts of the single input SVR models, this section concentrates on the advantages of a combined model that uses all available features. The combined model $P_{SV,comb}$ introduced in Equation (4.6) is able to benefit from its single data components and their individual performance on different forecast horizons. The quality of forecasts from this model, in comparison to the single

forecasts, is demonstrated in Figure 4.8, where it outperforms the single models on all considered horizons.

(a) single system 15 minutes ahead

(b) single system 2 hours ahead

(c) regional average 15 minutes ahead

(d) regional average 2 hours ahead
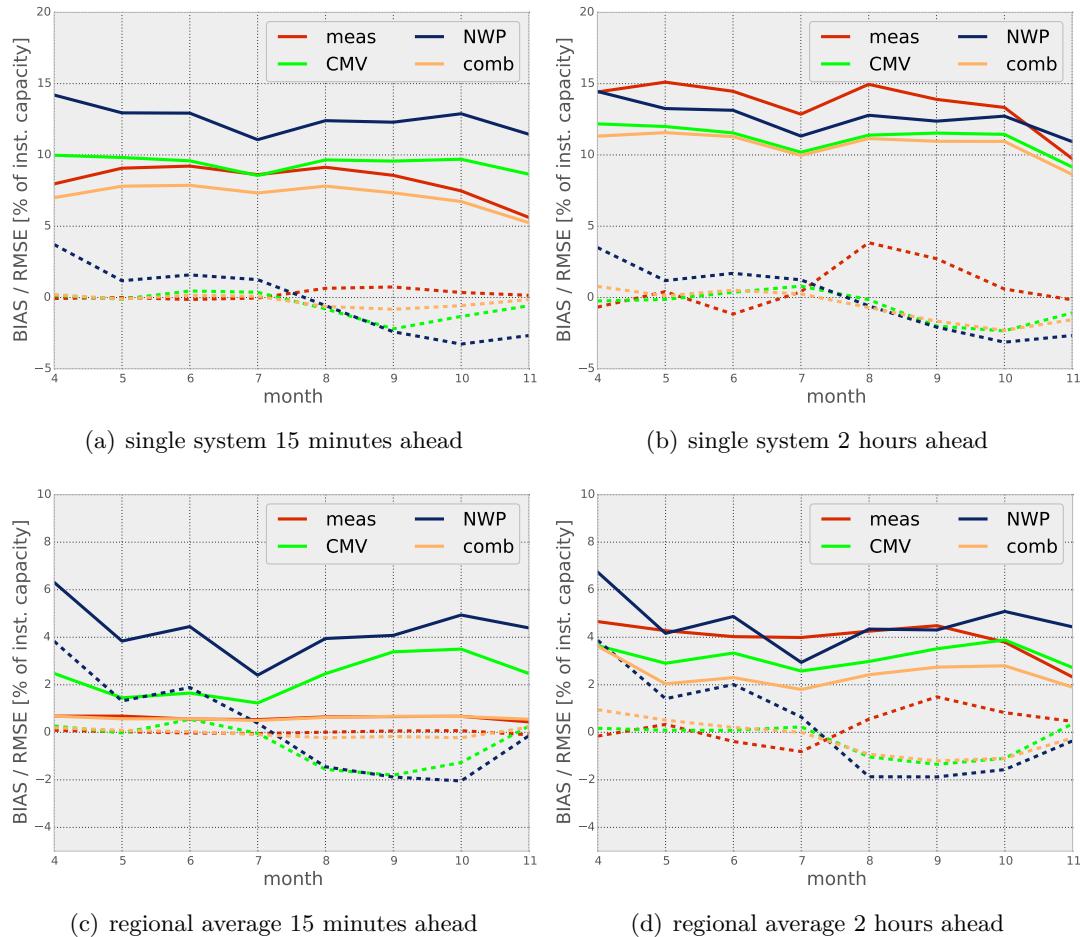
**Figure 4.10.:** *Single systems (top) and regional average (bottom) prediction quality of different SVR forecasts for each of the tested months with forecast horizons of 15 minutes (left) and two hours (right). Prediction quality is measured by the RMSE (solid lines) and BIAS (dashed lines).*

In the following, different timescales and weather situations are analyzed to discover possible trends in the error distribution of the SVR models.

First, Figure 4.9 shows the RMSE and BIAS in dependence of the time of day to look for daily biases in forecast quality. All single system forecasts exhibit a typical rise of the RMSE around the middle of the day, where the produced PV power, and consequently the extent of deviations is the highest. In case of a forecast horizon of 15 minutes (see Figure 4.9(a)), the combined forecast is able to increase the accuracy at midday the most,

while there is almost no improvement at sunrise and sunset in comparison to $P_{SV,meas}$. With increasing forecast horizon, the ranking of the single input forecasts inverts, but $P_{SV,comb}$ is still able to create the best forecasts throughout the day (see Figure 4.9(b)). Most apparent is the problem of $P_{SV,meas}$ at the beginning of the day; the model highly underestimates the power production for the first few hours of the day. This may be caused by the lack of data in the morning because the error increases with the forecast horizon. As with a longer forecast horizon, less data is available at the beginning of the day, $P_{SV,meas}$ is not able to capture the fast increase in PV power at the first forecasting timesteps. In Figures 4.9(c) - (d), the regional forecasts show a similar behavior with a less distinct diurnal error variation. For the 15 minute case, the diurnal variation of $P_{SV,meas}$ and $P_{SV,comb}$ is only visible on a scale with a higher resolution. Considering two hour ahead regional forecasts, the $P_{SV,meas}$ model still underestimates the power output in the early morning, while generating better forecasts in the afternoon (see Figure 4.9(d)).

Next, the error distribution for different months is analyzed to investigate if there are seasonal trends visible. Figure 4.10(a) shows the BIAS and RMSE of all the evaluation months for single site forecasts with a forecast horizon of 15 minutes. While $P_{SV,CMV}$ and $P_{SV,NWP}$ present no apparent monthly dependencies, predictions based on measurements, and consequently $P_{SV,comb}$, perform better during spring and autumn months. In July, there is almost no difference between the usage of CMV or measurements. With increasing forecast horizons the quality ranking changes and at two hours $P_{SV,CMV}$ performs better than the other single input forecasts. Even though the overall performance of $P_{SV,meas}$ at two hours is worse than the others', the prediction quality for different months is comparable to $P_{SV,NWP}$, and occasionally better (see Figure 4.10(b)). The regional forecasts show smaller errors (Figures 4.10(c) - (d)) with larger relative differences between models and higher forecast errors in April.

To investigate forecast performance in dependence of cloud conditions, the measurement-based clear sky index $k^*_{P,meas}$ (see Equation 2.12) is used to categorize the cloud situation at every timestep in one of ten groups ranging from 0, denoting complete overcast, to one, which relates to clear sky conditions (see Figure 4.11). For single sites and 15 minutes ahead (Figure 4.11(a)), $P_{SV,meas}$ performs best for complete overcast situations, whereas changing cloud conditions, often corresponding to intermediate values of $k^*_{P,meas}$ in the range of 0.3 to 0.7, are challenging for all models. In this range, there is only a small difference between $P_{SV,CMV}$ and $P_{SV,meas}$. For clear sky conditions, RMSE values decrease slightly. Here, it needs to be mentioned that values of $k^*_{P,meas}$ close to one do not always correspond to stable clear sky situations, but may as well occur at broken cloud conditions. When evaluating the BIAS in dependence of measured clear sky indices

(a) single system 15 minutes ahead

(b) single system 2 hours ahead

(c) regional average 15 minutes ahead

(d) regional average 2 hours ahead
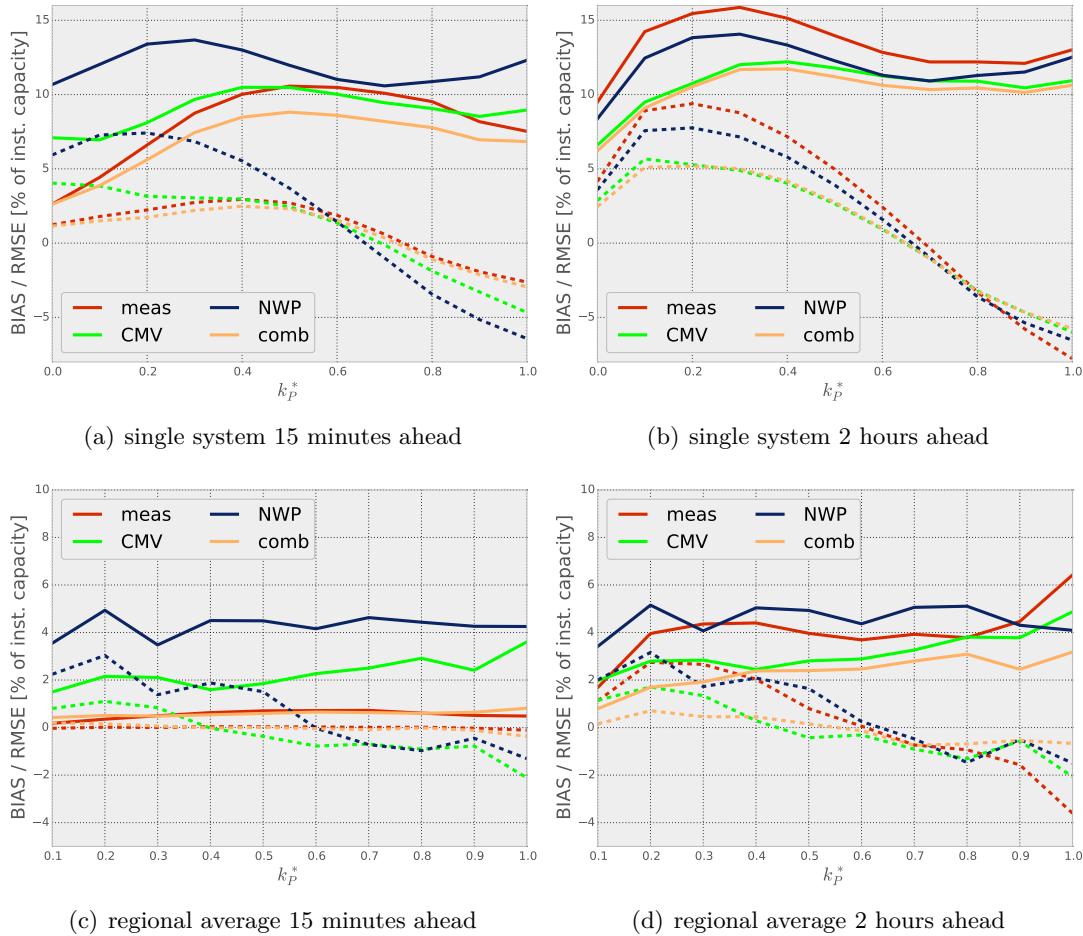
Figure 4.11.: *Single systems (top) and regional average (bottom) prediction quality of different SVR forecasts in relation to the clear sky index $k^*_{P,meas}$ with forecast horizons of 15 minutes (left) and two hours (right). Prediction quality is measured by the RMSE (solid lines) and BIAS (dashed lines).*

$k^*_{P,meas}$, a negative BIAS is observed at clear sky conditions ($k^*_{P,meas}$ close to one). Since forecast errors in these situations – at the very most – correspond to erroneous prediction of clouds, resulting in an underestimation of the PV power output. In contrast to this, the BIAS is positive for overcast situations, where forecast errors are related to predicting less cloud cover than actually present. With increasing RMSE, this trend gets more pronounced, as shown here for $P_{SV,NWP}$ for 15 minutes ahead forecasts and for all models for two hours ahead (Figure 4.11(b)). At two hours ahead, the largest RMSE values for all models are found with intermediate values of $k^*_{P,meas}$.

For the comparison of regional forecasts (Figure 4.11(c) and (d)), the $k_{P,meas}^*$ dependency is not that obvious anymore. Here, predictions for clear sky indices ranging from 0.2 to 0.8 highly benefit from averaging effects. Figure 4.11(d) shows problems for $P_{SV,meas}$ and - to a lesser extent - also for $P_{SV,CMV}$ for predicting regional clear sky situations. These are probably related to the comparatively large errors in the morning hours for these models.

In each of the above comparisons, the SVR predictions with all features combined ($P_{SV,comb}$) resulted in the lowest RMSE values for all analyzed situations, demonstrating the stability and quality of the combined forecast model. This encourages the idea of increasing the feature count in later chapters, so that the SVR model can benefit from even more information about the PV system.

4.5. Comparison to Parametric Model Approaches

In this section, the SVR forecasts are compared to persistence forecasts and PV power predictions based on PV simulations of irradiance forecasts, as described in Section 2.3.

(a) Without linear regression   (b) With linear regression
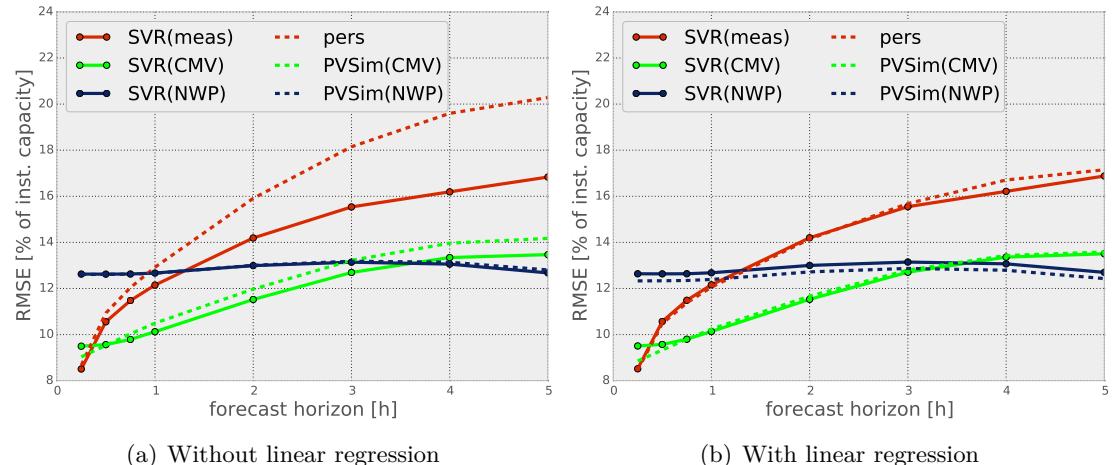
Figure 4.12.: *Comparison of SVR, persistence and PV simulation forecasting using different input data on single systems (a) without and (b) with linear regression post-processing of the physical models.*

In this work, there is also a distinction between purely physical modeling and statistically improved physical modeling using simple linear regression (see Section 2.3.3). Both approaches are compared to the SVR single model forecasts to investigate the difference between simple and advanced statistical methods.

In Figure 4.12, the results of the single input comparison (a) without and (b) with statistical correction are presented. In its initial state (Figure 4.12(a)), the RMSE values of the physical approaches are – in most cases – slightly higher than the respective SVR forecasts. The most apparent difference can be seen between $P_{pers,meas}$ (pers) and $P_{SV,meas}$ (SVR(pers)). Especially for longer forecast horizons, learning on past events works better than extrapolating the current ratio between measured and clear sky PV power.

After applying linear regression from Section 2.3.3, all parametric PV simulation model forecasts are improved. Now, in Figure 4.12(b), $P_{pers,meas}^{lr}$ (pers), $P_{Sim,CMV}^{lr}$ (PVSim(CMV)), and $P_{Sim,NWP}^{lr}$ (PVSim(NWP)) perform slightly better than the respective SVR models. In the following analysis, the post-processed models' outputs are used exclusively.

(a) single systems (b) regional average
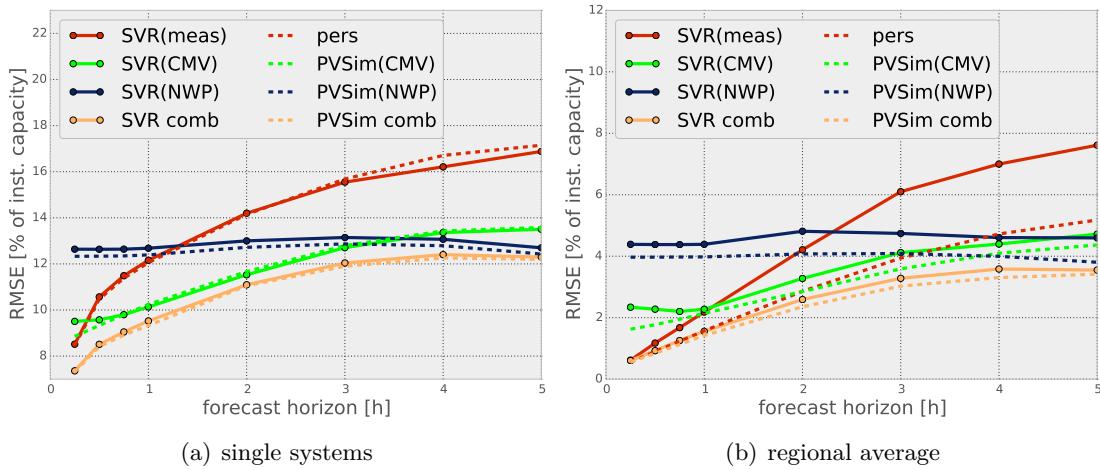
Figure 4.13.: *Comparison of SVR, persistence and PV simulation forecasting using single and combined input data on increasing forecast horizons for (a) single system and (b) regional averaged forecasts.*

First, all models are analyzed in dependence of the forecast horizon. Here, the results of the combined forecasts ($P_{Sim,comb}$ and $P_{SV,comb}$) of both approaches are included as well, showing almost the same results with a small advantage for the physical model in Figure 4.13(a). For regional forecasts the differences between SVR and physical models increase (see Figure 4.13(b)). The most prominent difference is found between $P_{pers,meas}^{lr}$ and $P_{SV,meas}$, while the combined forecasts show nearly equal results.

If one looks at the error throughout the day in Figure 4.14(a), the RMSE values at a forecast horizon of 15 minutes show a similar behavior, and additionally there is almost no difference between $P_{Sim,comb}$ and $P_{SV,comb}$ in case of single system settings. These

differences do not change much with an increasing forecast horizon for single systems. For regional forecasts, the most apparent difference is again between $P_{pers,meas}^{lr}$ and $P_{SV,meas}$ (Figure 4.14(b)). Here, the large error of $P_{SV,meas}$ in the early hours of the day, as described in Section 4.4, is not found for $P_{pers,meas}^{lr}$.

(a) single systems 15 minutes ahead   (b) regional average 2 hours ahead
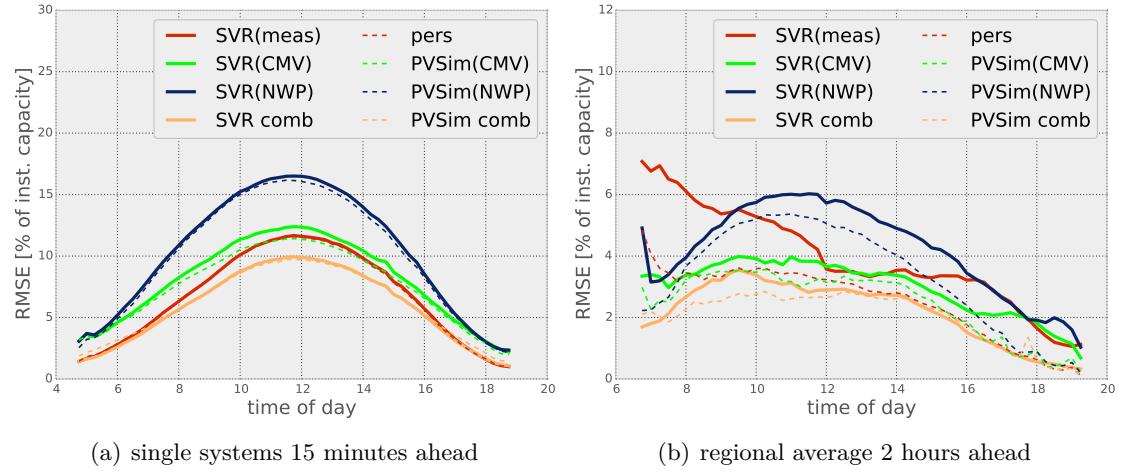
Figure 4.14.: *Single systems (a) and regional average (b) prediction quality of SVR, persistence and PV simulation forecasting (with linear regression) using single and combined input data for each time of day for (a) single systems (15 minutes ahead) and (b) regional averages (two hours ahead).*

Another interesting difference between the SVR and the physical model forecast is shown in Figure 4.15. Again, this figure shows the RMSE depending on the clear sky index k_P^*, that describes the cloud conditions at the predicted time step (see Section 4.4). While the SVR forecasts are able to generate slightly better predictions at overcast and clear sky conditions, situations with broken cloud conditions are better handled by PV simulations combined with linear regression for single sites (see Figure 4.15(a)). This changes when regional forecasts are considered.

In Figure 4.15(b), the physical models generate better forecasts in almost all cases, especially in clear sky scenarios. This probably can be explained by the explicit usage of tilt and orientation information by the physical model, in contrast to the SVR, which has to learn these features implicitly on historic data. The learning seems to work well on single systems, but on average time series this benefit is lost, particularly in cases with very high irradiance input. Still, either $P_{Sim,comb}$ or $P_{SV,comb}$ are superior to all other models, and therefore the best choice for forecasting PV power.

(a) single systems 2 hours ahead   (b) regional average 2 hours ahead
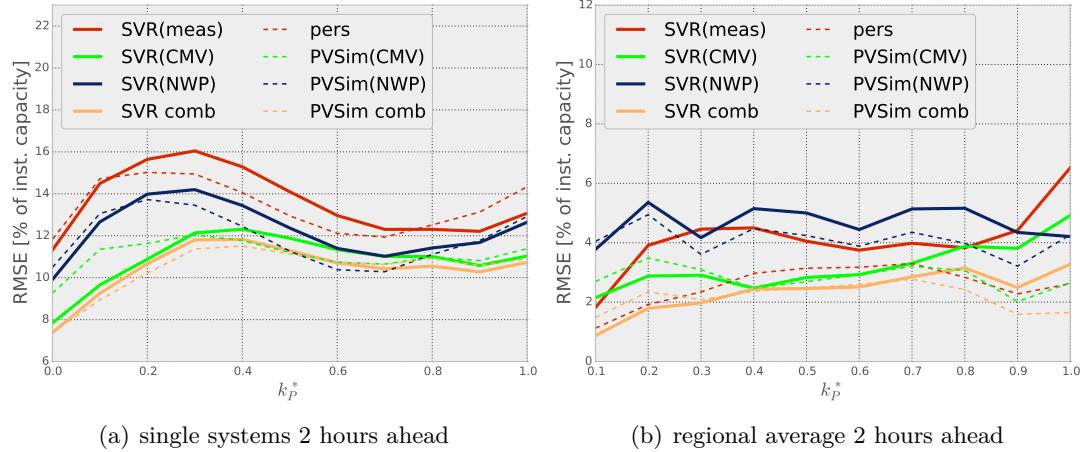
Figure 4.15.: *Single systems (a) and regional average (b) prediction quality of SVR, persistence and PV simulation forecasting (with linear regression) using single and combined input data, w.r.t. $k_{P,meas}^*$, for (a) single systems and (b) regional averages, both with a forecast horizon of two hours.*

Further Combination Approaches   In addition to the previous $P_{SV,comb}$ and $P_{Sim,comb}$ forecasts, another approach to include SVR into the modeling process chain is investigated in Figure 4.16. Here, the linear combination according to Equation (2.15) is substituted with a SVR model, that is optimized to use the output of $P_{pers,meas}$ and both PV simulations ($P_{Sim,CMV}$ and $P_{Sim,NWP}$) to generate a new combined forecast

$$P_{SV,Sim} := (P_{pers,meas}(t), P_{Sim,CMV}(t), P_{Sim,NWP}(t)) \rightarrow P_{meas}(t). \tag{4.6}$$

The new forecast model is able to generate the best predictions for longer forecast horizons of three to five hours ahead, and hence qualifies as an useful alternative to linear regression.

But as the difference in forecast quality is limited, there are no further investigations on this topic in this work. Instead, the following chapters concentrate on analyzing the potential of SVR-based forecasts without the knowledge of physical processes, i.e., the conversion of irradiance forecasts to create PV power output predictions. Nonetheless, the results of these combination models can be seen – to a small degree – as an indicator of non-linear dependencies between the combination of persistence and PV simulation approaches and the actual PV power output, which supports the choice of SVR.
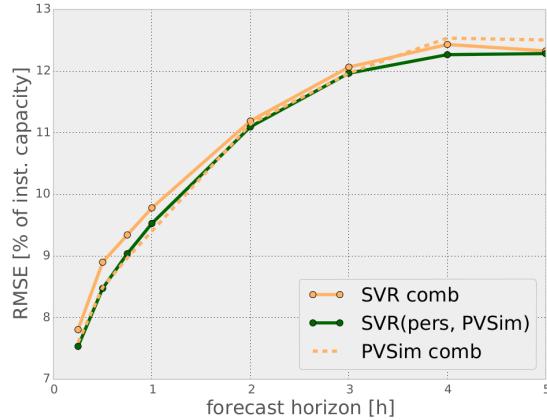
Figure 4.16.: *Comparison of SVR and physical PV power forecast combinations.*

4.6. Concluding Remarks

In this chapter, support vector regression (SVR) using PV power measurements, numerical weather prediction (NWP) and cloud motion vector (CMV) irradiance forecasts is applied for PV power forecasting. To achieve high quality forecasts, the SVR parameters as well as training set sizes are optimized. The parameter optimization is performed on a preselected test set consisting of days with different cloud conditions from half a year of data. This is done on 10% of all 921 PV systems available, which are selected by k-means clustering and the optimization is repeated for all evaluated forecast horizons as well as input feature combinations.

In the evaluation, a test of the single input data sources is done first. While SVR forecasts based on PV measurements show good results for short forecast horizons up to one hour ahead, NWP-based predictions prove to be better on horizons starting at three hours. SVR predictions using irradiance forecasts from CMVs perform best in between. Furthermore, a SVR forecast model is trained on a combination of all these input sources. The combined model outperforms the individual models on all implemented forecast horizons.

In a second step, these results are compared to forecasts created with physical modeling approaches: a PV simulation model is used to convert irradiance forecasts to PV power predictions and a persistence approach is applied to extrapolate recent PV power measurements into possible future outputs. Without statistical improvements, i.e., linear regression, the single and combined SVR forecasts are able to generate slightly better results than the respective physical modeling approaches. After linear regression is applied, the physical models are slightly better than the SVR models. The best results

are again achieved with a combination of all three input sources. For single PV system predictions, there are almost no differences between the combined SVR and the combined physical model forecast, whereas regional forecasts are slightly better handled by the combined physical model. One other important aspect is the error reduction between single system and regional average forecasts of over 75% for the combined model forecasts, and furthermore the high quality of regional PV power predictions for short-term forecast horizons up to one hour ahead.

Even if the SVR is not able to outperform the combined physical model, it is still interesting for further considerations, especially if there is no information concerning a PV system's specification. Thus, a machine learning approach is in fact recommended if there is no information about a PV system given, except its measurements and location. Moreover, the required knowledge about the underlying physical processes for PV power forecasting is reduced as a system's behavior can be learned from its data.

In this comparison, the potential of SVR is restricted by limiting the data input to a level that is comparable to the parametric forecasting model. In theory, the more reasonable information is available to a statistical learning algorithm, the better the results should be. Therefore, the use of additional features will be investigated in the next chapters.

In the following Chapter 5, multiple weather parameters from the NWP system of the ECMWF are analyzed to find features that are able to improve the SVR forecasts. These features may add information that could help to estimate the recent weather conditions at a system's site more accurately. Until now, all NWP data is preprocessed in a way, that an optimal combined irradiance model of DWD and ECMWF is achieved for the operational use. Instead of applying the output of this combined NWP model, the raw data is employed to assess the possiblities of a purely data-driven SVR approach.

4. Comparison to a Parametric Model

5. Feature Analysis and Optimization

After showing the potential of support vector regression (SVR)-based PV forecasting modeling in Chapter 4, different approaches that could lead to an enhancement of prediction quality are tested in the following two chapters. One possibility is the consulting of more forecast parameters of numerical weather prediction (NWP) models. The idea is to give machine learning models access to more information about a PV system's predicted weather situation, so that this information can be included in the estimation of future PV power output.

For this, the importance of NWP features for PV power forecasting needs to be evaluated. The applied datasets in this evaluation and the preprocessing of this data is described in Section 5.1. In Section 5.2, a description of the statistical learning method, i.e., random forest (RF), that is utilized for feature importance estimations is given. The selection of NWP features is done in Section 5.3 and the most important features are evaluated by comparing their application on different regression models in Section 5.4. The results of this chapter are summarized in Wolff et al. [53].

5.1. Experimental Setup and Data Preprocessing

In this work, the data of 92 PV systems spread across Germany (see Figure 5.1), with varying installed capacities from rooftop installations to big PV parks, is selected. The PV measurements of these systems are the same as before (see Section 2.2.1) and the selection is done in the same way as described in Section 4.2 via clustering.

The focus of this chapter lies on evaluating the usefulness of different weather parameters of a numerical weather prediction system for PV power forecasting. Here, the data from the ECMWF, which is introduced in Section 2.2.2, is applied. The ECMWF's IFS model provides over 120 single level forecast weather parameters with a temporal resolution of three hours. In Figure 5.1, the grid points of the ECMWF IFS model are shown with each grid field covering an area of about 12.5 km × 12.5 km.

Instead of using the preprocessed data of Section 4, the raw NWP model output is now processed in the following way to match the data of the PV systems' measurements.

Spatial Interpolation In previous works (see Lorenz et al. [38]), a positive effect of averaging NWP irradiance forecasts of multiple surrounding grid points instead of using the geographically nearest forecast was observed. The best results were achieved by averaging a 4×4 grid around a PV system's location. Due to the easier implementation and faster calculations, the spatial interpolation in this chapter is done with a distance weighted k-nearest neighbor regression model (see Equation (3.4)) with $k = 16$ in regard to the 4×4 grid. For convenience, this method of spatial interpolation is applied for all other NWP forecasts employed in this work.
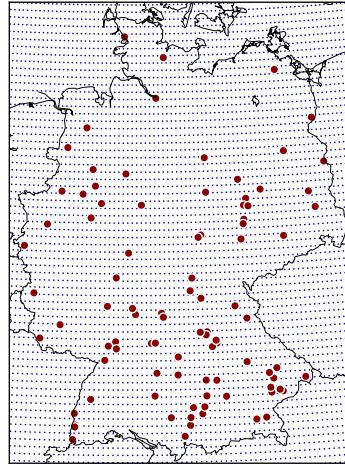
Figure 5.1.: *Locations of 92 PV systems in Germany (red) and grid points of ECMWF's NWP model (blue).*

Temporal Interpolation In case of irradiance forecast parameters, the clear sky model of Dumortier (see Section 2.1) is applied to interpolate the data from three hours to 15 minutes. For all other weather parameters, the values are interpolated with a linear model. The clear sky interpolation is working in three steps:

1. Calculating clear sky index k_{3h}^* for 3 hour values:

$$k_{3h}^* = \frac{I_{forec,3h}}{I_{clearsky,3h}}$$

2. Linear interpolation of k_{3h}^* to 15 minute values results in k_{15min}^*.

3. Using k_{15min}^* as a factor for $I_{clearsky,15min}$ with 15 minute resolution:

$$I_{forec,15min} = k_{15min}^* \cdot I_{clearsky,15min}.$$

By applying this method, a more realistic behavior of the irradiance is estimated that respects the deterministic diurnal trajectory throughout the day, instead of just interpolating in a linear way. In Figure 5.2, an example shows the interpolation of three-hourly to 15 minute data – while respecting the dirunal course of irradiance – via clear sky interpolation.
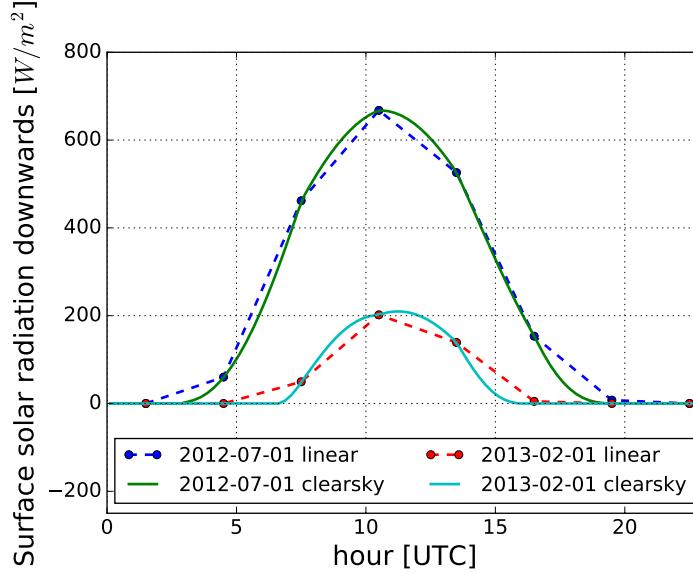
Figure 5.2.: *Illustration of clear sky interpolation vs. linear interpolation of three-hourly ECMWF's NWP irradiance to 15 minute data for two different days.*

Data Normalization   Instead of using fixed values for data normalization (as before in Section 4.1), which is not feasible for many different values and units, a more general standardization is applied. Each feature $x_{i,j}$ of a pattern $\mathbf{x}_i \in \mathbb{R}^d$ of the training set $\mathcal{T}$ with a size of N, is standardized to $z_{i,j}$ as

$$z_{i,j} = \frac{x_{i,j} - \bar{x}_j}{\sigma_j}, \tag{5.1}$$

with $i \in \{1, \ldots, N\}$ and $j \in \{1, \ldots, d\}$. Where the mean of each feature vector $\bar{x}_j$ is calculated with

$$\bar{x}_j = \frac{1}{N} \sum_{i=1}^{N} x_{i,j} \tag{5.2}$$

65

and its standard deviation σ_j as

$$\sigma_j = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_{i,j} - \bar{x}_j)^2}. \tag{5.3}$$

This removes the mean from a feature's value and scales it to the unit variance of a feature vector, so that ideally normalized values from -1 to 1 are presented to the SVR model. The mean and standard deviation are computed on the training dataset and used as reference to normalize all data passed into the model. The normalization ensures that there is no weighting due to different units and data ranges of the input features.

5.2. Random Forest

In this section, the statistical learning method random forest is introduced, which is mainly applied for feature selection. The concept behind the construction of a random forest [6] regressor is bagging (bootstrap aggregation). Bagging makes it possible to "average many noisy but approximately unbiased models, and hence reduce the variance" [23] by combining multiple models. The RF algorithm reduces variance by decorrelating the trees as it randomly selects input features in each node, which describes a splitting point in the decision tree.

To train a random forest model, bootstrapping (random sampling with replacement) is applied on the training set $\mathcal{T}$ to retrieve B, the number of trees in the forest, subsets $\mathcal{T}^{*b} = \{(\mathbf{x}_1^*, y_1^*), (\mathbf{x}_2^*, y_2^*), \ldots, (\mathbf{x}_N^*, y_N^*)\} \in \mathbb{R}^d \times \mathbb{R}$, $b = 1, 2, \ldots, B$. On each of these subsets a RF tree T_b, i.e., a single RF regression model, is grown by applying the following steps until a stop criterion is reached (e.g., a minimum number of samples belonging to a newly created node), according to Hastie et al. [23]:

1. Randomly select a number $m \leq d$ of features from the input pattern.

2. Calculate the best feature for splitting, i.e., the feature that "maximizes the decrease of some impurity measure" [40].

3. Split the node according to the selected feature into two new nodes.

After finishing the tree-growing process, the algorithm outputs an ensemble or set of trees $\{T_b \mid b = 1, 2, \ldots, B\}$ and the random forest regression function is

$$f_{rf}^{B}(\mathbf{x}) = \frac{1}{B} \sum_{b=1}^{B} T_b(x). \tag{5.4}$$

Here, the RF regressor implementation of the scikit-learn [48] Python package, with its standard settings except for the number of trees, is used. The number of trees is increased from ten to 64 to improve tree diversity, while still keeping the computation time at a minimum. This is as well the minimal recommendation of Oshiro et al. [45], even though

No.	Feature name	Unit
1	100 metre U wind component	ms^{-1}
2	100 metre V wind component	ms^{-1}
3	10 metre U wind component	ms^{-1}
4	10 metre V wind component	ms^{-1}
5	2 metre dewpoint temperature	K
6	2 metre temperature	K
7	Clear-sky direct solar radiation at surface*	Jm^{-2}
8	Cloud base height	m
9	Evaporation	m of water equivalent
10	High cloud cover	$(0-1)$
11	Large-scale precipitation	m
12	Low cloud cover	$(0-1)$
13	Medium cloud cover	$(0-1)$
14	Snow density	kgm^{-3}
15	Snow depth	m of water equivalent
16	Snow evaporation	m of water equivalent
17	Snowfall	m
18	Surface net solar radiation, clear sky*	Jm^{-2}
19	Surface solar radiation*	Jm^{-2}
20	Surface solar radiation downwards*	Jm^{-2}
21	Total cloud cover	$(0-1)$
22	Total column ice water	kgm^{-2}
23	Total column liquid water	kgm^{-2}
24	Total column rain water	kgm^{-2}
25	Total column snow water	kgm^{-2}
26	Total column water	kgm^{-2}
27	Total column water vapour	kgm^{-2}
28	Total precipitation	m
29	Total sky direct solar radiation at surface*	Jm^{-2}
30	Zero degree level	m

Table 5.1.: *Selection of ECMWF weather forecast feature list. Marked features (*) are interpolated with a clear sky interpolation method instead of a linear interpolation.*

they have a different field of use. As an impurity measure (Step 2 of the tree-growing algorithm) to decide on the best splitting criterion/feature, this implementation uses the

mean square error (MSE) shown in Equation (2.19). The MSE for each possible feature split is calculated and the feature with the highest MSE decrease is selected.

5.3. Feature Selection

The main goal of this chapter is to assess the ability of additional NWP weather parameters to improve PV power forecasting. Using all available weather parameters of the ECMWF's model would result in feature spaces with over 120 dimensions. This is not feasible for short-term PV power forecasting with forecast horizons of 15 minutes as the time to calculate a forecast would simply take too much time. Moreover, not all features contribute relevant information to a forecast. As a result, less features are selected that still hold the highest possible additional information for the models.

In a first step, with the help of experts in the field of meteorology, all NWP parameters that are physically irrelevant for PV power forecasting are removed. Thus, from the about 120 available IFS single level forecast parameters, 30 features remain for the evaluation. These 30 features are listed in Table 5.1.

In literature (e.g., in Guyon and Elisseeff [20]), there are two major classes of feature selection methods: wrapper and filter. While wrapper methods embed a regression model into another optimization algorithm, e.g., genetic algorithms, that try different feature combinations until a stop-criterion is reached, and select the best solution that is found along the search, filter methods rank single features by some measure that describes their usefulness for the given task.

As one iteration of a wrapper algorithm would take a lot of time in this scenario with 92 PV systems, a filter model is selected to achieve a fast method for evaluating the benefit of single weather parameters.

Feature Selection with Random Forest   To rank the preselected features, random forest's feature importance capabilities are utilized. The feature importance is achieved by traversing all trees with the training dataset, sum the impurity criterion (in this case the MSE value) of each node and weight it with the number of samples that are routed to a specific node. By doing this, features that are choosen often and early in a tree's hierarchy will also receive high importance values. The output of this feature importance algorithm is normalized to one. Details on random forest's feature importance are well discussed in Genuer et al. [19].

Now, all preselected features are presented to the RF algorithm with the goal of calculating a regression function to predict power output of each single PV system.

For this, the daytime[1] feature values of the previous 65 days are used for training as a similar configuration showed good results in the previous chapter (see Section 4.3). Each of the 92 PV systems is trained individually for each day of the year 2012 starting at 2012-03-10. Thus, the feature importance of every single day is assessable and an evaluation of seasonal or daily changes in feature importances is possible. In Figure 5.3,
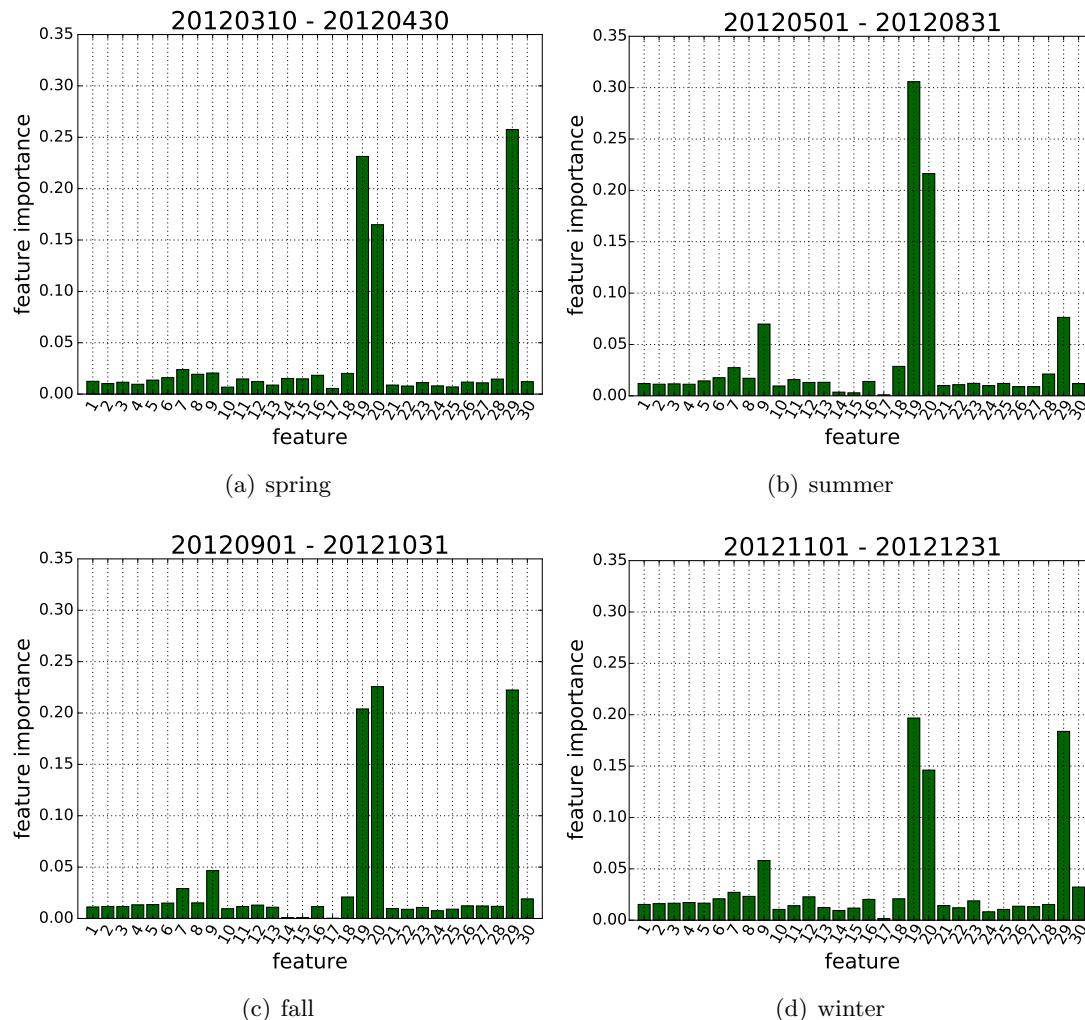
Figure 5.3.: *Random forest feature importances of 30 NWP weather parameters averaged for (a) spring, (b) summer, (c) fall, and (d) winter months for all 92 PV systems.*

the feature importances for all days are averaged over all PV systems for spring, summer, fall, and winter months. In summer, all features are dominated by the forecast of surface solar radiation (19) and surface solar radiation downwards (20) with over 50% feature

---

1 Cosine of the solar zenith angle above 80 degree

importance. Only the parameters total sky direct solar radiation at surface (29) and evaporation (9) seem to hold some information for the resulting PV power output with an importance of about 7.5% each. This changes when looking at the spring, fall, and winter months. The feature importances of 19 and 20 decrease while the importance of other features increase. Especially in spring and fall, the direct component of solar radiation (29) is more important than in summer. Noticeably for winter months, the zero degree level (30) and low cloud cover (12) as well as snow depth (14) and snow evaporation (15) get more important for the forecasting of PV power generation, while in fall these parameters are clearly less important. In all plots, evaporation (9) seems to be one of the more important features except for spring months, where it is only slightly more important than all other features.

Rank	No.	Feature name	Feature importance
1	19	Surface solar radiation	0.244
2	20	Surface solar radiation downwards	0.178
3	29	Total sky direct solar radiation at surface	0.171
4	9	Evaporation	0.050
5	7	Clear-sky direct solar radiation at surface	0.026
6	18	Surface net solar radiation, clear sky	0.023
7	8	Cloud base height	0.020
8	30	Zero degree level	0.019
9	6	2 metre temperature	0.018
10	16	Snow evaporation	0.017

Table 5.2.: *Top ten most important ECMWF weather forecast features ranked by random forest feature importance.*

In the following experiments, it is evaluated if including more weather parameters to the prediction pattern is improving the quality of PV power forecasts. At this point, there is no distinction between seasons, so that only the ten highest ranked features, with regards to their average feature importance of 2012, are evaluated. These features are listed, ranked by their RF feature importance values, in Table 5.2.

5.4. Prediction Comparison

Now, the most important features obtained in Section 5.3 are applied for PV power forecasts. In this evaluation, the models are trained for each day of the year 2013 and the training set consists of the previous 65 days as before. The training and forecasting is done for each PV system separately. Aside from the RF predictor used to determine the feature importance, a linear regression and support vector regression approach is tested,

and thus the applicability of random forest's feature importance for different learning algorithms is evaluated as well. The quality of the different predictors is again measured with the RMSE and the SVR-based parameters are tuned by performing grid-searches employing the respective patterns like in Section 4.3.

Instead of testing all possible combinations, the top ten features are added iteratively to the pattern. The base pattern (Equation (5.5)) consists of the latest available measurement values P_{meas} according to the considered forecast horizon Δt. The iterative pattern/label matchings containing the top ten features have the following structure:

$$(P_{meas}(t - \Delta t)) \rightarrow P_{meas}(t) \tag{5.5}$$

$$(P_{meas}(t - \Delta t), Feature_1(t)) \rightarrow P_{meas}(t) \tag{5.6}$$

$$\dots$$

$$(P_{meas}(t - \Delta t), Feature_1(t), Feature_2(t), \dots, Feature_{10}(t)) \rightarrow P_{meas}(t) \tag{5.7}$$

Figure 5.4 shows the results of forecasts with linear regression, SVR, and RF. First, the shortest forecast horizon of 15 minutes is shown in Figure 5.4(a). The RMSE values of the forecast models using only measurement data (features = 0) are already good and can not profit much from additional features. Especially after adding the NWP forecast

(a) 15 minutes ahead   (b) 2 hours ahead
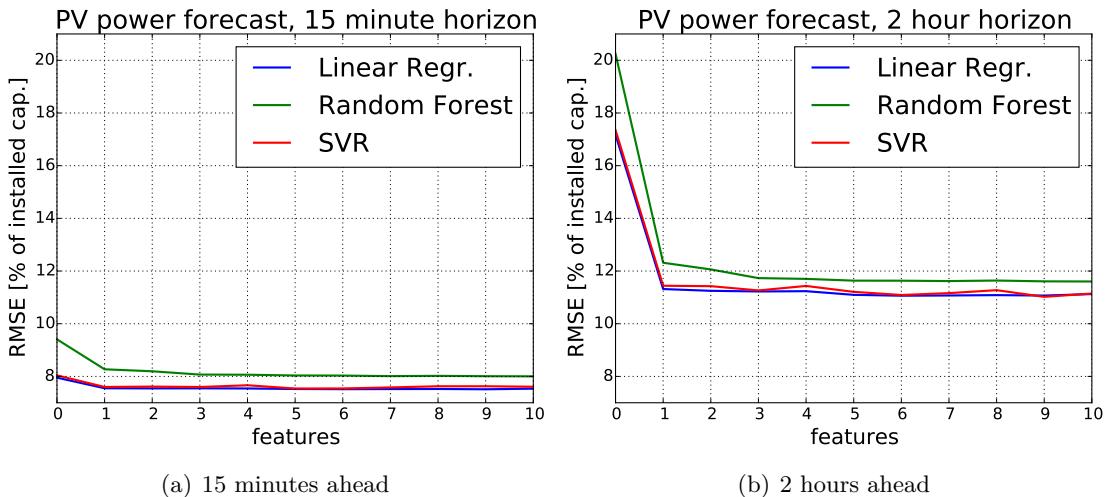
Figure 5.4.: *Comparison of linear regression, random forest, and support vector regression with increasing input feature count for a prediction horizon of (a) 15 minutes and (b) two hours ahead.*

of surface solar radiation, which is deemed the most important feature in the importance

tests. Starting at that point, there is almost no further improvement for the linear and SVR model anymore. Only RF forecasts acquire their best results with the top five features, but the quality is still worse than that of the other two models.

When the temporal difference between the most recent measurement and the forecasted timestep gets bigger, the less important the measurements get. This is demonstrated in Figure 5.4(b) for a two hour forecast horizon. The initial measurement-based forecasts are far worse for a forecast horizon of two hours than in the 15 minutes case. As before, adding only the most important feature is already enough to reach a high forecast accuracy for the linear regression and SVR. There are only insignificant improvements afterwards. Again, the RF model needs more information, i.e., more features to reach its optimum.
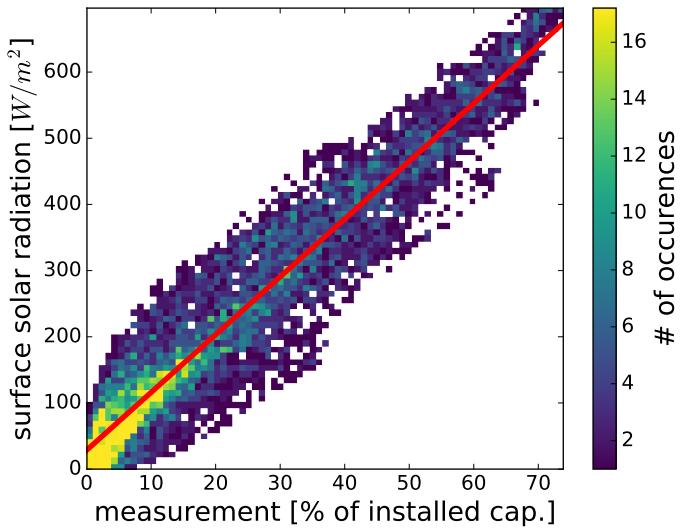
Figure 5.5.: *Scatterplot of measurements and corresponding NWP forecast of surface solar radiation. The red line indicates the linear regression curve of these two variables.*

In both scenarios, the difference between a simple linear regression and the more sophisticated support vector regression with optimized user-defined parameters is small. This indicates that the SVR is not able to benefit from its ability to model non-linear relations, and therefore is uncapable of achieving better results than a linear regression. To confirm this, a scatterplot of the most important feature, i.e., surface solar radiation, is shown in Figure 5.5. In this figure, the average of all stations' measurements and the respective average of the surface solar radiation forecasts is compared. The distribution shows that these parameters hold a strong linear dependency, which is highlighted by a linear regression fit in this figure.

Testing Different Input Availability In the following, different input scenarios are tested to see how the forecast quality changes if the best features are missing. Because there was not much of a difference between the linear and the SVR model, and to account for shorter computation times, the following three combinations are tested exclusively with the linear model:

- *Model 0*: Linear regression with measurements and all top ranked features as before (Linear Regr.)

- *Model 1*: Linear regression with measurements and without the top three ranked (irradiance) features (Linear Regr. 4-10)

- *Model 2*: Linear regression without measurements and with all top ranked features (Linear Regr. no meas)

(a) 15 minutes ahead
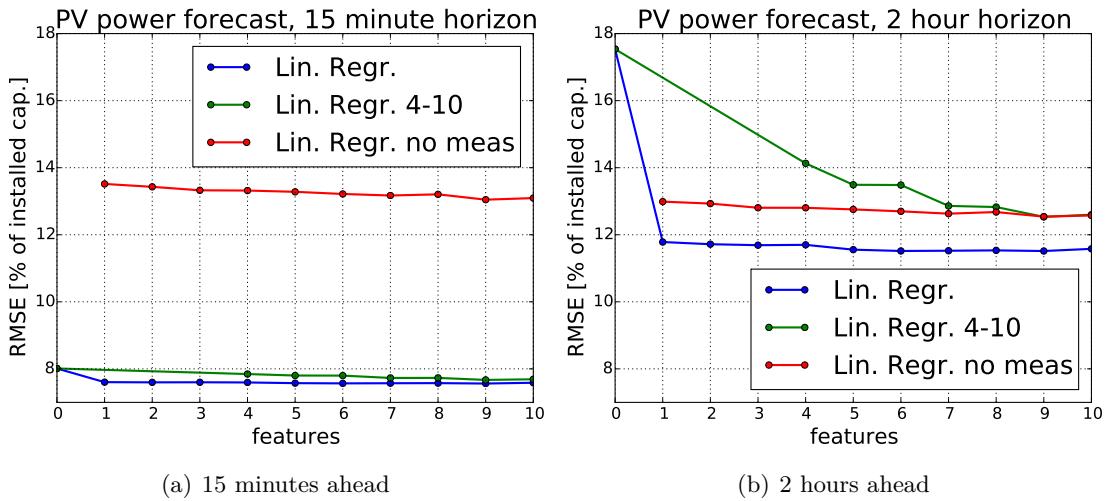
(b) 2 hours ahead

Figure 5.6.: *Comparison of linear regression forecasts with different input feature combinations for a prediction horizon of (a) 15 minutes and (b) two hours ahead.*

In case of the shortest forecast horizon of 15 minutes (Figure 5.6(a)), the models using measurements are substantially better than the model without measurement values. While there are no improvements after adding feature 1 for Model 0, the information of the three missing features of Model 1 can be restored by adding more NWP features. Although the forecasts of Model 2 are inferior to measurement-based ones, an improvement is achieved by adding more of the top ten features. For a forecast horizon of two hours ahead, measurements are not that important anymore (see Figure 5.6(b)). Now, Model 2 is competitive to Model 0 and even better than Model 1, despite the lack of measurements.

Model 1 is profiting from additional NWP features and able to compensate less information about the predicted radiation with increasing feature count. Utilizing all information (Model 0) is still generating the best forecasts, but with an increasing forecast horizon the difference to the other models is increasingly vanishing.

5.5. Feature Selection for Special Weather Situations

As the forecast quality is only unsignificantly improving with more features in a yearly averaged setting, this section aims at forecast improvements at special weather conditions, i.e., snow and fog. The idea is that in these cases NWP irradiance forecasts are not suitable as the sunlight can not be received by the modules, e.g., due to remaining snow covering the module. In these situations, features like snow depth, snow density, and temperature may increase the accuracy of forecasts.

Rank	No.	Feature name	Feature importance
1	19	Surface solar radiation	0.142
2	29	Total sky direct solar radiation at surface	0.093
3	9	Evaporation	0.078
4	15	Snow depth	0.057
5	20	Surface solar radiation downwards	0.056
6	8	Cloud base height	0.046
7	21	Total cloud cover	0.037
8	14	Snow density	0.035
9	11	Large-scale precipitation	0.033
10	23	Total column liquid water	0.032

Table 5.3.: *Top ten most important ECMWF weather forecast features ranked by random forest feature importance for snow and fog situations.*

In this approach, three months with a high snowfall rate and an overall low visibility caused by fog are chosen, i.e., 2013-01-01 until 2013-03-30. Again, the RF feature importance of these months is calculated and the top ten features are applied for PV power forecasting. In Figure 5.7(a), the results of the feature importance study are shown. In comparison to Figure 5.3, the feature importances of snow depth (15) and snow density (14) are significantly increased, whereas surface solar radiation(19), total sky direct solar radiation at surface (29), and especially surface solar radiation downwards (20) are decreased highly. The only difference between (19) and (20) is the ground reflectance component that is removed from (20) [15]. With high snow and fog occurrences, this feature shows a higher inaccuracy, and therefore is less important for the PV power

forecast. Overall, there are indicators that snow features have a higher importance at situations with lots of snow and fog.

In Table 5.3, the top ten features for snow situations are listed. These features are now used to forecast PV power output for the same three months (2013-01-01 – 2013-03-30). As models, linear regression and SVR are applied. To assess the quality of forecasts, two respective models, using solar surface radiation and measurements as input, are compared to the snow models. Furthermore, the models are trained and optimized in the same way as before using grid-search.

(a)

(b)
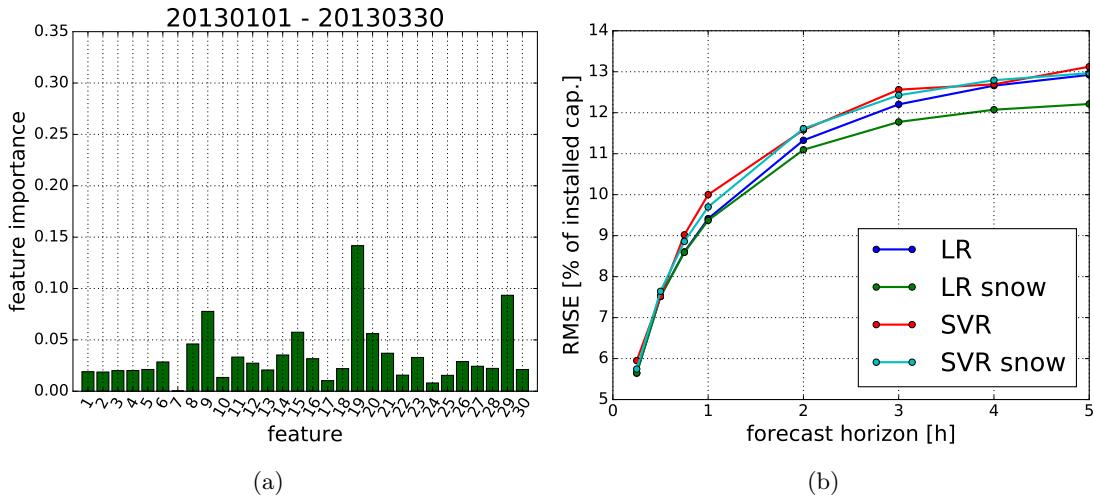
Figure 5.7.: *(a) Random forest feature importances of 30 NWP weather parameters averaged for three months (2013-01-01 – 2013-03-30) with high snowfall and long periods of fog. (b) Comparison between LR and SVR snow models and corresponding models with only one added feature for increasing forecast horizons.*

The comparison of those four models is shown in Figure 5.7(b). The difference between all models is relatively small at forecast horizons of 15 minutes ahead. Only the SVR model without additional features (SVR) is slightly worse than the others. The overall difference between the SVR prediction with and without snow features is neglectable. For linear regression forecasts, the addition of more features at snow situations has a higher impact, in particular, for longer forecast horizons. At five hours ahead, the difference is about 1% RMSE. Still the differences are not as big as the RF feature importance suggested.

5.6. Concluding Remarks

In this chapter, the benefit of using additional weather forecast parameters of the ECMWF's numerical weather prediction (NWP) model for PV power predictions is evaluated. The importance of all features that are related to the output of a PV system are assessed via random forest's feature importance algorithm. Furthermore, the application of this type of feature selection is tested with different regression models.

While there are differences between seasons and weather situations, the averaged results of the feature importance evaluation show that over 50% of importance are shared by irradiance weather parameters. This is later on confirmed by analyzing PV power forecasts, that are using the highest ranked features. The results show that the quality of power output forecasts for more than a few minutes ahead is mostly depending on irradiance forecasts. Otherwise, for shorter prediction horizons (15 to 30 minutes ahead), PV power measurements are essential for generating high quality forecasts. These results are seen with all of the applied regression models, i.e., support vector regression, random forest, and linear regression. As irradiance and a PV module's power output are highly linear correlated, SVR is not able to create better forecasts than the linear regression approach.

This is again demonstrated in the special weather scenario. Here, weather situations with high snowfall and fog occurences are tested. While the RF feature importance changed considerably, the SVR is not able to use additional information about, i.a., snow and can not produce better forecasts than the linear regression in this case.

In the next chapter, the goal is to create an optimized single system forecast. Therefore, considering the results of this chapter, instead of using many weather parameters of the NWP models, only irradiance is added in these experiments. This enables the usage of data from neighboring PV systems, that introduces a new approach to detect incoming clouds via the spatial and temporal relationship of PV systems.

6. Single System Model

In Chapter 4, a general configuration of support vector regression (SVR)-based predictions is constructed and evaluated. The resulting configuration is used on all PV systems of the evaluation dataset and is not aiming to be the best configuration for each single PV system. In this chapter, the focus lies on optimizing the PV power forecasts of one particular system and evaluate the improvement of prediction quality. By using just one PV system, the optimization search spaces are highly reduced, so that further tests with more input datasets, e.g., PV measurements of neighboring systems, become feasible. With respect to the results of the previous chapter, only irradiance forecasts are extracted from DWD's and ECMWF's numerical weather prediction (NWP) models. Instead of using preprocessed datasets like in Chapter 4, the data is again employed as it is provided.

This chapter is structured as follows: In Section 6.1, the experimental setup of this evaluation is described, where at first the selection process of the single PV system for this study is addressed. Afterwards, in Section 6.2, the SVR PV power forecast model is optimized for this single system by applying features from neighboring sytems (Section 6.2.1), selecting most important features (Section 6.2.2), and optimizing SVR parameters in Section 6.2.3. Apart from a standard grid-search for optimizing SVR parameters, a moving average parameter selection approach and an ensemble forecast, that combines forecasts with multiple parameter combinations, are evaluated. In Section 6.3, all optimized models of this chapter are compared to an oracle forecast, which always knows the best SVR parameter settings for each forecasted day, and the previously best SVR model of this thesis, i.e., $P_{SV,comb}$.

6.1. Experimental Setup

After getting an overview of how good the SVR works on a larger and more generalized scale in Chapter 4, the focus of this chapter lies on improving forecast quality of the SVR model on single PV systems. The idea is, that the findings of this chapter – in terms of methods that are used – can be applied to all PV systems. Thus, an increase of the overall performance of SVR-based forecast could be achieved.

For the experiments performed in this chapter, a PV system is selected that has a high data availability starting in January of 2014 until July of 2016. Furthermore, in this setting, the data of surrounding PV systems is utilized as input data for prediction models, too. These neighboring stations comply to the same data availability constraints, which include an availability of day-time values (sun elevation above ten degress) over 80%, without consecutively missing values over the range of more than two days. For this first approach, the data of a single PV system and nine neighboring systems with a spatial extension of approximately 10 km × 10 km is selected. The locations of the chosen PV systems is shown in Figure 6.1[1].
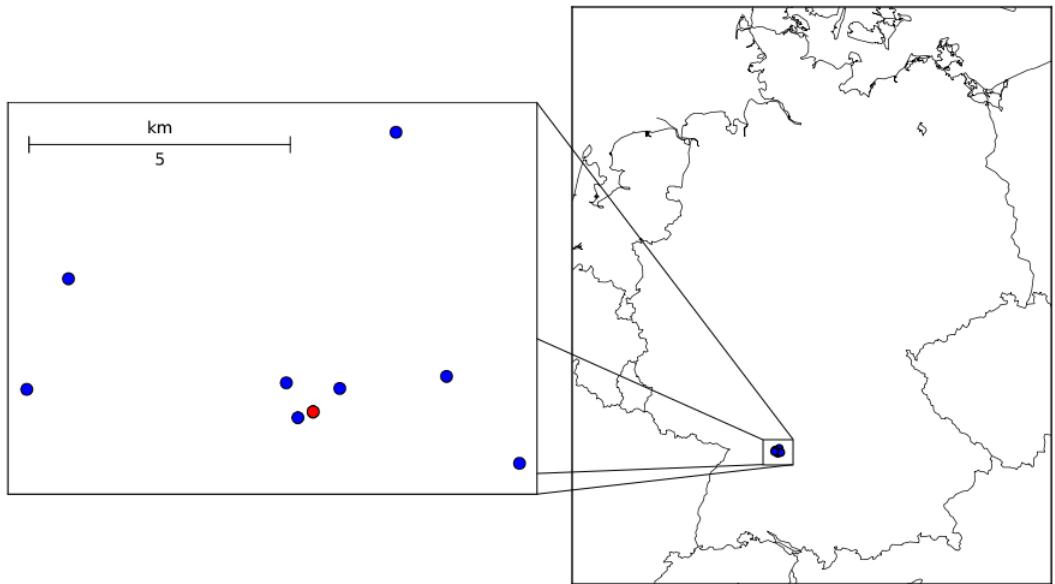
Figure 6.1.: *Location of the selected single PV system (red) and its neighboring systems (blue) in Germany.*

From each of the ten PV systems all available system-site data is considered: measurements, irradiance forecasts from the NWP models of ECMWF and DWD, cloud motion vector (CMV)-based irradiance forecasts, and clear sky irradiance. In contrast to the data used in Section 4, all data is processed as the respective provider or the applied model supplies it. The data normalization is analogous to the normalization of Section 5.1, where PV measurements are normalized to the installed capacity of the corresponding PV system. Irradiance forecasts from NWP models are again interpolated spatially with *k*-NN and temporally with a clear sky interpolation approach as explained in Section 5.1.

---

1 The ninth neighboring PV system is not visible in Figure 6.1 as it is in direct vicinity to the main system.

6.2. Model Optimization

In this section, the SVR is optimized on a single PV system level without the previous constraints of comparability to the operational parametric model as in Chapter 4. Here, the size of the training set is fixed to the data of the previous 50 days, as this is the standard setting of the physical model and the evaluation in Section 4.2 showed only minor differences in using between 50 to 70 days for training, whereas less and more days decreased the quality of predictions more clearly.

By looking at only one particular PV system, the amount of model input data can be increased significantly. Because of this, measurement and forecast data of neighboring systems is added to the input pattern of the prediction models. The usefulness of adding neighboring measurement data is analyzed in Section 6.2.1.

Afterwards, in Section 6.2.2, a reasonable number of features is selected, so that the added complexity of computations stays within a feasible range. The biggest impact on model performance in earlier chapters is caused by SVR parameters choices, thus in Section 6.2.3 different approaches for parameter optimization are tested.

6.2.1. Measurements of Neighboring PV Systems

First, the benefit of adding data of neighboring PV systems is examined. To assess the importance of a neighboring systems' measurements, random forest's (RF) feature importance, described in Section 5.2, is applied.

For each day from 2014-07-01 until 2015-06-30, a RF model is trained with measurement data of all ten PV systems of the previous 50 days with forecast horizons from 15 minutes to five hours. The training mapping is defined as

$$(P_{meas}(t - \Delta t), P_{meas,1}(t - \Delta t), \ldots, P_{meas,9}(t - \Delta t)) \to P_{meas}(t), \qquad (6.1)$$

where training set size $N = 50$ days, P_{meas} denotes the measurement of the single PV system and $P_{meas,j}$, with $j \in \{1, \ldots, 9\}$, describes the measurements of the nine neighboring systems. Again, Δt indicates the respective forecast horizon.

In theory, an increasing forecast horizon should shift the importance of PV measurements of a single system to observations made at surrounding systems. These spatio-temporal correlations in solar power assessment are caused by clouds moving across the observed area.

To some extent, this can be seen in the results of this study illustrated in Figure 6.2. At a forecast horizon of 15 minutes, the single system's measurements consitute to over 40% of feature importance, while the importances of the other PV systems' measurements

are neglectable. The surrounding PV measurements gain more importance starting at a forecast horizon of half an hour, and with two hours ahead the PV system at the most western side holds the highest importance. This could be attributed to the circumstance that the general wind direction in Germany is western, so that – in average – the clouds' influence on the power output occurs earlier on this PV system. To confirm this, additional enviromental and geographical factors need to be considered and evaluations on different PV systems should be conducted. This is not covered here, as for this thesis, a connection between the measurements of the PV systems is sufficient.

(a) 15 minutes ahead

(b) 30 minutes ahead

(c) 2 hours ahead
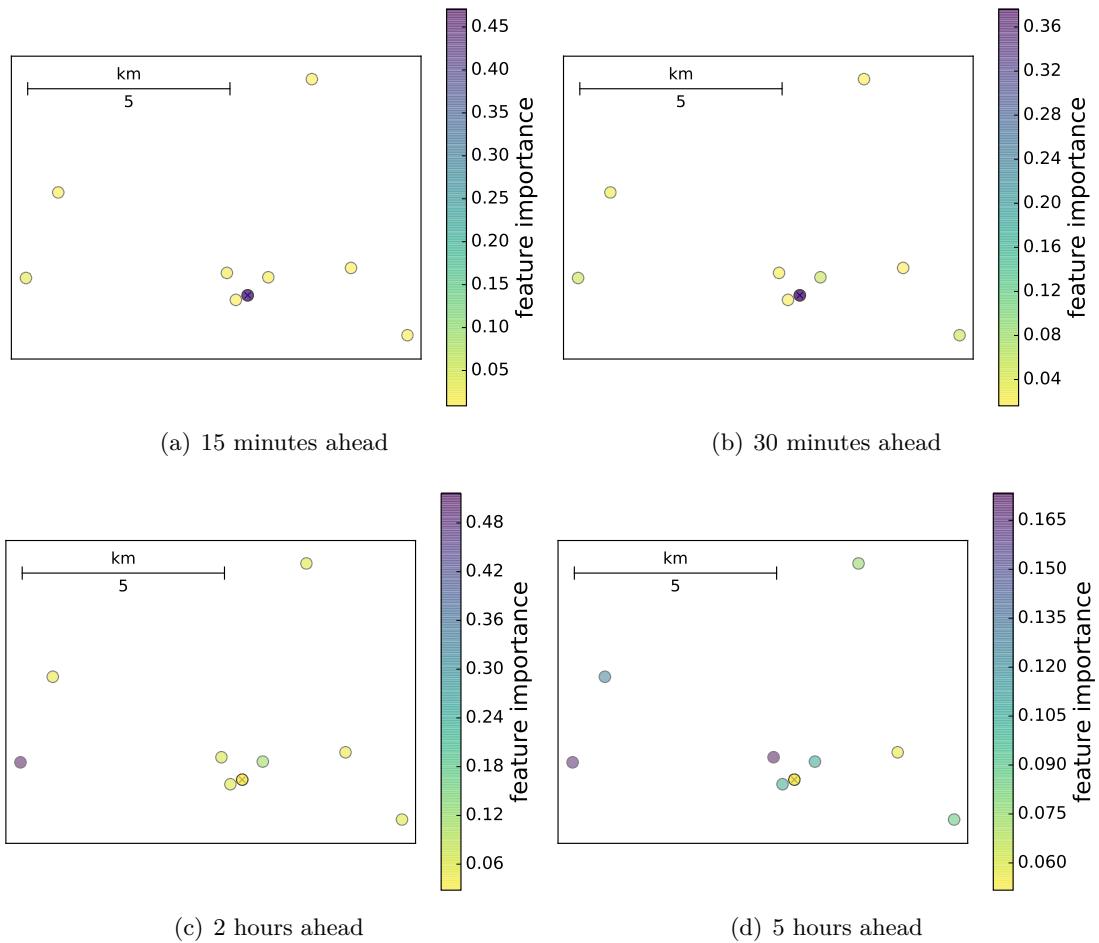
(d) 5 hours ahead

Figure 6.2.: *Random forest feature importance of neighboring PV systems' measurements for forecasting PV power at the single system for different forecast horizons.*

With a forecast horizon of five hours ahead, the shares of feature importances are distributed more evenly with values between 6% and 16.5%. At a forecast horizon of five hours, a radius of five kilometer is already too small to capture possible correlations in

the data due to the speed of cloud movement. Now, a spatial averaging effect applies, so that the relationship of the forecasted value and historic measurements is best described by using all available data. In summary, it can be stated that there exists a connection between the measurement data of surrounding systems and the forecasted power output, especially for larger forecast horizons. Under the assumption that additional data improves the PV power forecasting models, even more data of surrounding PV systems is added to the input data in the following experiments.

6.2.2. Feature Selection

After estimating a possible information gain by adding PV measurements of neighboring systems, this section is focusing on selecting features from the complete input space that add the most value to the PV power forecasts. Thus, the set of features is reduced and the optimization task of the SVR is getting less complex computationally. All available data of each PV system is now included, i.e., measurements, irradiance forecasts from NWP models (ECMWF and DWD) and from the satellite-based CMV approach as well as irradiance estimations from the clear sky model. So that, with the data of all ten PV systems combined, 50 features are applied for the input pattern. Again, the test set starting at 2014-07-01 until 2015-06-30 is evaluated and the training set size contains 50 days of data prior to each of the 365 forecasted days. Furthermore, all eight previously evaluated forecast horizons are considered. To get an impression of how the SVR model performs in comparison to other methods, linear (LR) and random forest (RF) regression are tested as well, whereas all models are trained with their respective standard settings according to the Python package scikit-learn [48].

In a first approach, all features are added iteratively to the pattern to determine if there is a saturation of information gain resulting in a stagnating or decreasing forecast quality. For this, two methods for deciding the order of added features are tested:

1. Adding the features of the main PV system first by using the order: P_{meas}, I_{ECMWF}, I_{DWD}, I_{CMV}, and I_{clear}. The data of the neighboring PV systems is added in the same order beginning with the data of the closest system and ending with the system that has the longest (Euclidian) distance to the main system.

2. Training a RF model on the training set, containing the data of the prior 50 days, and calculating the importances of all features. Afterwards, the features are iteratively added according to their RF feature importance values starting with the most important feature.

(a) 15 minutes ahead   (b) 2 hours ahead
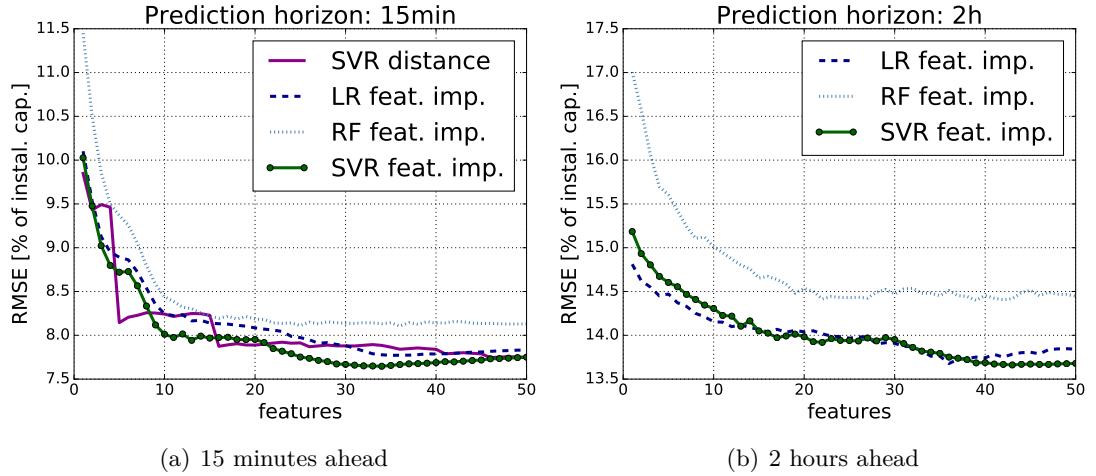
Figure 6.3.: *Comparison of linear regression (LR feat. imp.), random forest (RF feat. imp.), and SVR (SVR feat. imp.) with increasing feature count according to RF feature importance. In (a), the SVR forecast of the distance-based method (SVR distance) is added for comparison for a forecast horizon of 15 minutes, while in (b), at a forecast horizon of two hours, only feature importance-based models are shown.*

In Figure 6.3(a), the RMSE values are plotted for an increasing number of added features for a forecast horizon of 15 minutes ahead. The behavior of the first adding method, i.e., the distance-based appending, is similar for all three regression models, so that for clarity only the best performing SVR-based forecast is shown here (SVR distance). For all other curves, the feature importance-based approach is applied (feat. imp.).

In contrast to the ramps and stagnating parts of the distance-based feature appending method, the quality of predictions with the RF feature importance method increases smoothly and almost constant with adding more features. Even though the distance-based SVR predictor reaches its minimal RMSE value faster at around 16 features, the feature importance-based SVR forecasts reach an even lower RMSE value at its optimum with 33 features included. This trend is found for all forecasts horizons, so that the self-adapting feature importance method is used in the following sections exclusively. At a forecast

Forecast Horizon	15 min	30 min	45 min	1 h	2 h	3 h	4 h	5 h
Feature Count	33	32	32	34	42	42	43	42
RMSE [%]	7.65	9.21	10.29	11.38	13.66	15.13	16.43	17.30

Table 6.1.: *Number of features resulting in the lowest averaged RMSE values for SVR-based predictions for each implemented forecast horizon after applying feature importance ordering. Additionally, the RMSE values of the best feature combination in percentage of installed capacity is listed.*

horizon of 15 minutes ahead, the linear regression and SVR show a similar behaviour. After both models reach their minimum, the accuracy of forecasts decreases with more features. Whereas the quality of RF predictions is stagnating at one point and the model can not benefit from the additional information. At a forecast horizon of two hours ahead (see Figure 6.3(b)), LR and SVR are mostly equal, except for the lower RMSE of the SVR at a higher feature count.

Even as the differences are less significant, the SVR-based forecast still achieves the better prediction quality at a forecast horizon of two hours in Figure 6.3(b). For an optimal prediction at longer forecast horizons, there are increasingly more features needed. The averaged best feature counts of the SVR predictor for all implemented forecast horizons are listed in Table 6.1 with their corresponding RMSE values.

(a) RMSE in percentage of installed capacity

(b) Normalized RMSE of each day
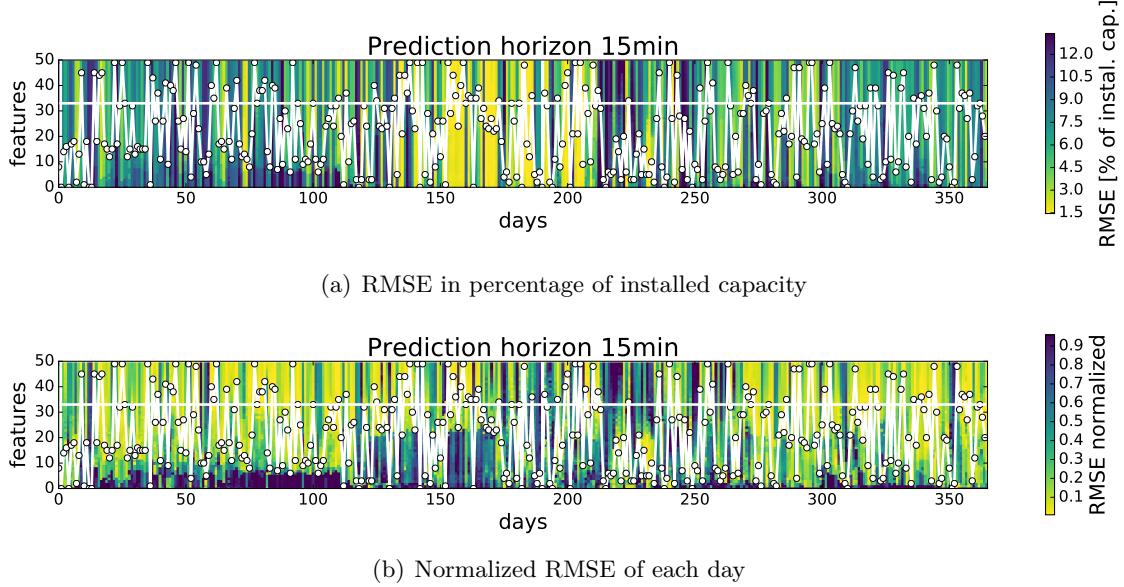
Figure 6.4.: *RMSE values of the SVR forecasts of all tested days with importance-based appending of features with (a) values in percentage of installed capacity and (b) normalized RMSE values for each day for a prediction horizon of 15 minutes ahead.*

Figure 6.4 gives an impression on how much the optimal number of features is varying from day to day. Here, the color-coded RMSE values of the SVR-based prediction for every employed feature combination (vertical axis) for every day of the year (horizontal axis) is plotted. The white dots indicate the numbers of features that achieve the lowest RMSE and the straight white line shows the best yearly-averaged number of features (compare Figure 6.3). There are examples where the best feature count switches from below five features to almost 50 features from one day to the other. This makes it

difficult to determine any number of features in advance. Because of this, for further evaluations, the yearly-averaged best performing number of features is selected. As Figure 6.4(b) illustrates, the quality of forecasts with the best average number of features is not significantly worse than the forecasts with an optimal count from day to day in most cases, but the feature count is still reduced by one third, and hence the computational complexity of the models is lowered. Additionally, by using the feature importance, the model is already training adaptively with the more promising features of the prior days, encouraging a more robust model to address seasonal and diurnal changes.

Forecast Horizon	SVR Parameters			RMSE
[min]	C	γ	ϵ	[% of inst. cap.]
15	12.9155	0.00464	0.02154	7.15
30	1.0000	0.01000	0.02154	8.55
45	3.59381	0.00215	0.10000	9.46
60	12.9155	0.00100	0.00464	10.75
120	12.9155	0.00100	0.10000	13.17
180	12.9155	0.00215	0.02154	14.61
240	3.59381	0.00215	0.10000	16.30
300	1.0000	0.00215	0.10000	17.57

Table 6.2.: *Optimal SVR parameters after conducting a grid-search with 1000 parameter combinations on each day from 2014-07-01 to 2015-06-30 for all employed forecast horizons w.r.t. RMSE values.*

6.2.3. Parameter Optimization

Now, after selecting the importance-based feature selection with fixed numbers of features, the SVR model parameters need to be optimized. As the results of the previous section indicate, the performance of a SVR model with certain parameters varies unpredictably from day to day and deciding on a parameter setting is as complicated as predicting the weather conditions at all. Thus, the following three approaches to find suitable SVR models are implemented and evaluated:

1. A preceding parameter grid-search on data of the past year.

2. A moving average parameter selection that searches for the average best parameter combination by evaluating forecasts of the previous days.

3. A (weighted) ensemble method combining SVR-based forecasts that use different model parameters.

The latter two models aim at improving the SVR without having the benefit of big sets of historic data, and restrict the optimization process on the training set. The goal is to achieve models that are able to dynamically react to changes in forecast conditions within a shorter period of time, compared to a yearly grid-search, that mostly results in more general settings.
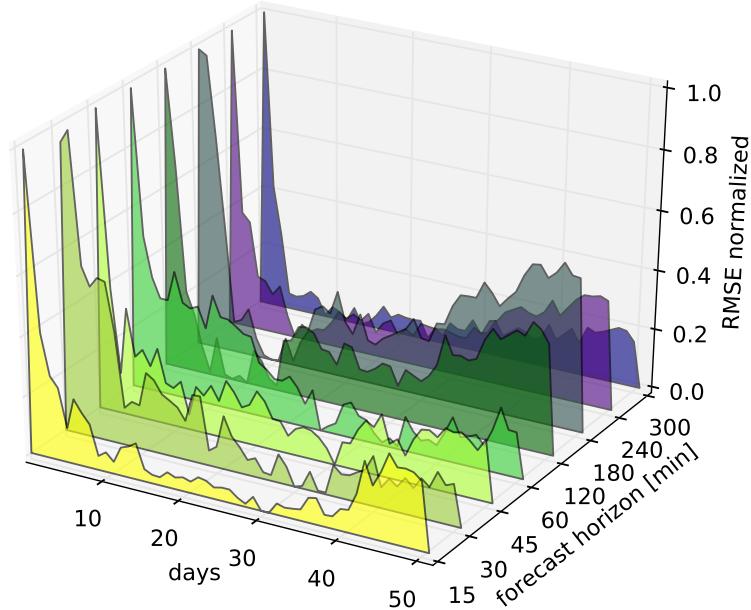
Figure 6.5.: *Comparing the normalized RMSE on an increasing number of days used for the moving average parameter selection for all considered forecast horizons.*

Parameter Grid-Search For each day from 2014-07-01 to 2015-06-30, SVR-based forecasts with different parameter combinations are calculated and the overall best parameter combination is used for the evaluation. This approach is similar to the optimization in Chapter 4, but as it is only applied on one PV system, the test set and the number of tested parameter combinations can be increased. Moreover, the resulting grid-search parameters are now specifically tailored towards the employed single PV system. For each parameter of the SVR model, i.e., C, ϵ, and γ, ten settings are chosen from a logarithmic grid ranging from 10^0 to 10^5 for C and 10^{-1} to 10^{-4} for ϵ and γ, summing up to a thousand parameter combinations. The resulting optimal settings are listed in Table 6.2. One disadvantage of this method is its need for long periods of historic data. If forecasts for a new PV system are requested, the data of one year prior to the first forecasting day is needed.

Moving Average Parameter Selection   For this approach, forecasts with different parameter combinations are calculated and the quality is measured by their RMSE value. The parameter combination that is performing the best in average on the previous days is applied to forecast the following day. In this implementation, all parameter combinations from the previous grid-search are used. In Listing 6.1, a Python pseudocode is illustrating the calculation of a forecast that selects a SVR model depending on the performance at preceding days. Hence, in contrast to the recent grid-search approach, the model is dynamically adapting to the recent situation at a PV system's site.

In Figure 6.5, the numbers of days, that are used for the moving average method for each forecast horizon, are analyzed. An upper bound of 50 days is defined, which is equal to the training set size, so that no additional data is required for the parameter optimization. Because the RMSE values are not changing much after adding the first few days, the normalized RMSE of each forecast horizon is used in the figure to capture even small changes.

```python
def mov_avg_forecast(forecast_days, n_params=m, n_days=n):
    # Moving average forecasts with m parameters
    # of the last n days:
    all_params = list_params(n_params)
    mov_avg_days = n_days

    all_forec = {} # hash map of all forecasts
    mov_avg_forec = {} # hash map of mov. avg. forecasts
    ctr = 0
    for day in forecast_days:
        all_forec[day] = get_forec(day, all_params)
        if ctr >= mov_av_days:
            # Finding index of best average
            # parameter combination
            best_idx = get_best_param_idx(all_forec,
                                            mov_avg_days)
            mov_avg_forec[day] = all_forec[day][best_idx]
        ctr += 1
    return mov_avg_forec
```

Listing 6.1: *Illustration of a SVR forecast function that is applying the best parameter combination of the previous days (moving average).*

Now, it is noticeable that the amount of days needed for an optimal forecast is reduced with increasing forecast horizon. For all forecast horizons, taking less than around five days to calculate an averaged best parameter combination would result in inferior PV power forecasts. Using the whole training set, i.e., 50 days, to select a parameter combinations is not the best choice either. The optimal numbers of days for each forecast horizon are listed in Table 6.3.
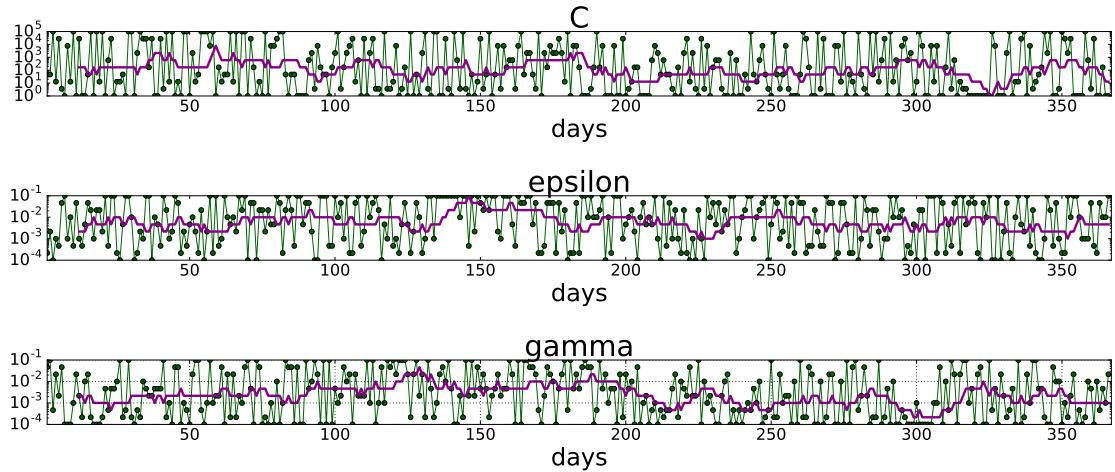
Figure 6.6.: *Best SVR parameter combinations for each day of the evaluation dataset (green dots) for a forecast horizon of two hours ahead. The purple line indicates the best parameter combination of the previous eleven days obtained by the moving average method.*

In Figure 6.6, the moving average method is applied to the evaluation dataset for a forecast horizon of two hours ahead and with an averaging period of eleven days. The green dots show the best option for each of the SVR parameters for every day of the test set.

Forecast Horizon	15 min	30 min	45 min	1 h	2 h	3 h	4 h	5 h
Days (optimal)	30	30	31	25	11	10	9	14
RMSE [%]	7.37	9.63	10.81	11.54	13.71	14.80	15.28	15.65
Days	50							
RMSE [%]	7.55	9.71	10.98	11.67	13.99	15.20	15.73	15.90

Table 6.3.: *Optimal number of days to use for the moving average parameter selection for each forecast horizon resulting in the lowest RMSE values on the evaluation dataset. For comparison reasons, the results of a moving average selection with 50 days, i.e., all days of the training set, is shown.*

There are no indications of deterministic SVR parameter settings, that work especially well on a longer interval. Instead, for almost every day, there is a specific combination that results in the best forecast in regards to the RMSE. To deal with this variability, the moving average of the previous days is used (purple line). With eleven days averaging, the fluctuations of the curve is reduced but parameter trends are still accounted for. These parameter trends are seen for all forecast horizons in more or less characteristic forms, depending on the applied optimal number of days.

(a) 15 minutes ahead   (b) 2 hours ahead
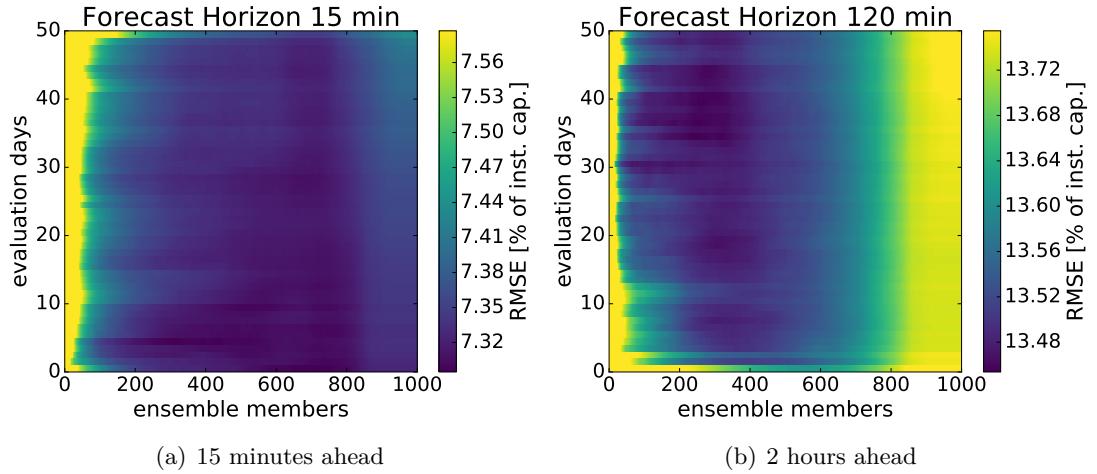
Figure 6.7.: *Comparison of the performance of ensemble forecasts with an increasing number of evaluation days and ensemble members for (a) 15 minutes ahead and (b) two hours ahead PV power forecasts.*

Ensemble Forecast   Instead of using only the best model, the idea of an ensemble forecast is to calculate multiple predictions and average the results into a single forecast. This approach is called simple ensemble in the following sections. Like before, all SVR-based forecast with a thousand different parameter combinations are employed.

In a second approach, the performance of each SVR parameter combination is considered as well. Depending on a parameter combination's performance on previous days, its contribution to the ensemble of the next day is weighted (weighted ensemble). In the following, two parameters are optimized before using the weighted ensemble forecasts: the number of forecasts included in an ensemble (ensemble members) and the number of previous days, that are considered for the performance evaluation of an ensemble member.

The results of the optimization process of these two parameters are shown in Figure 6.7 for 15 minutes and two hour ahead forecast horizons. While forecasts with only a small

number of ensemble members result in forecasts with higher RMSE values, the PV power forecasts improve with increasing ensemble members until about 800 members for 15 minutes ahead. For two hours ahead forecasts, a visible optimum is reached earlier at around 200 to 400 ensemble members. At one point, the RMSE values are stagnating (15 minutes ahead) or increase (two hours ahead) when a certain number of ensemble members is reached. In both cases, the number of evaluation days has almost no influence on the forecast quality.

| Forecast Horizon | Simple Ensemble | Weighted Ensemble (optimal) | | |
| | RMSE | | | RMSE |
[min]	[% of inst. cap.]	Members	Days	[% of inst. cap.]
15	7.40	316	5	7.29
30	9.55	703	2	9.46
45	10.73	645	1	10.53
60	11.65	439	35	11.32
120	13.90	246	35	13.45
180	15.12	368	6	14.63
240	15.75	223	22	15.25
300	15.99	435	19	15.53

Table 6.4.: *Results of the ensemble forecast optimization with optimal settings for all considered forecast horizons. For comparison, the RMSE values of the simple ensemble are listed as well.*

The optimal settings for ensemble members and evaluation days for each forecast horizon are listed in Table 6.4. Here, the optimal weighted ensemble is also compared to the simple ensemble, i.e., an unweighted averaged forecast of all ensemble members. For all forecast horizons, the weighted ensemble is able to create better PV power forecasts, but the improvements are only ranging from 0.6% to 3.2 %.

In the following, the results of all parameter optimization approaches are compared to the combined SVR forecast $P_{SV,comb}$ from Chapter 4 and an oracle model, that uses the best grid-search SVR parameter combination for every day.

6.3. Model Evaluation

After applying the optimization steps on the parameter selection and forecast combination methods, the resulting forecasts are compared to the best previous SVR model $P_{SV,comb}$ from Chapter 4.

Additionally, to assess the best possible SVR forecast, an oracle forecast is built, that uses the best parameter combination for each day. This oracle is not applicable in real-

world conditions as the best parameter combinations are not known in advance. Moreover, deciding on the optimal SVR parameter combination is as difficult as forecasting the PV power output conditions at all. The moving average parameter selection and the (weighted) ensemble method of Section 6.2.3 can be seen as approaches to create PV power forecasts that approximate the oracle predictor.

In Figure 6.8(a), $P_{SV,comb}$ (SVR comb), the oracle SVR forecast (SVR oracle), a SVR model with optimized parameters from grid-search (SVR GS) on the 2014 to 2015 test dataset, and both moving average approaches with 50 days of evaluation (SVR 50) and the optimal number of evaluation days for each forecast horizon (SVR MA) are compared. The quality of the general model $P_{SV,comb}$ (SVR comb) from Section 4 is worse for this single PV system than the average results of the previous comparison suggested (see Figure 4.13). Apart the relatively high RMSE values, the curvation is not as steadily increasing as before, especially the peak at a forecast horizon of two hours ahead is highly irregular. This is most likely the result of unsuitable SVR parameters for this specific PV system.

(a) Moving Average Parameter Selection   (b) (Weighted) Ensemble
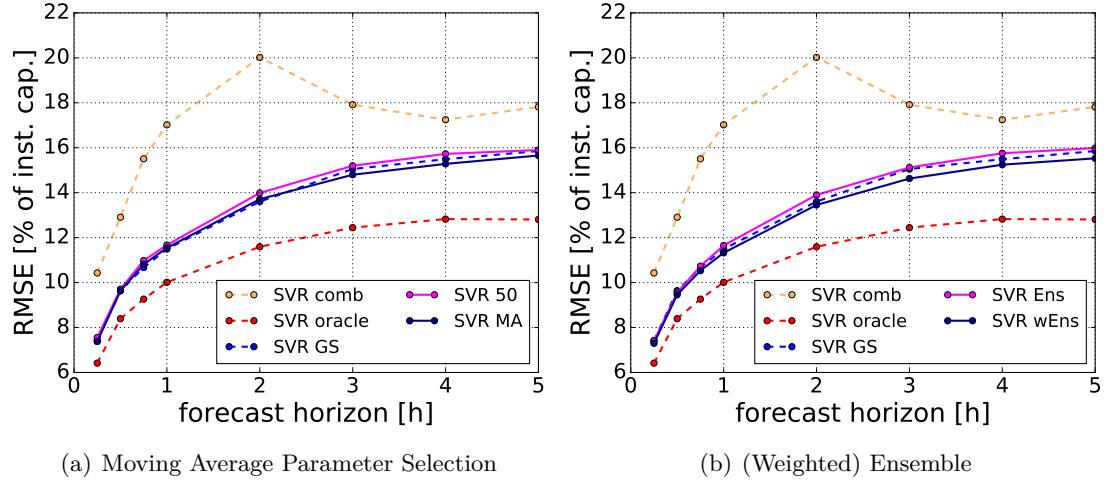
Figure 6.8.: *Comparison of SVR PV power forecasting models. In both figures, the baseline models $P_{SV,comb}$ (SVR comb), SVR oracle, and the grid-search SVR model (SVR GS) are used. In (a) these models are compared to moving average parameter selection approaches with a fixed size of 50 evaluation days (SVR 50) and the optimized numbers of days (SVR MA). In (b), the comparison is done with both employed ensemble methods: simple (SVR Ens) and weighted ensemble (SVR wEns).*

In contrast to this, the oracle predictor (SVR oracle) is – as expected – delivering the best forecasts for all employed forecast horizons.

All other applied methods are operating in the middle of these two extremes. The grid-searched SVR predictor (SVR GS), that was optimized on the previous year, is following the trend of SVR oracle with significant higher RMSE values. But, nonetheless, generates more stable forecasts throughout all forecast horizons in contrast to $P_{SV,comb}$. The moving average method applied on 50 days (SVR 50) is slightly worse than a grid-search on a whole year, whereas the version with optimized numbers of training days (SVR MA) is achieving almost the same results as SVR GS and performs even better starting at a forecast horizon of three hours.

The results of the ensemble approaches are shown in Figure 6.8(b). A simple ensemble (SVR Ens), where the average forecast of all ensemble members is calculated, is achieving similar results than the moving average parameter selection with a fixed evaluation time frame of 50 days (SVR 50), and therefore only slightly worse errors than SVR GS. Out of all applied optimization methods, the SVR weighted ensemble approach is able to generate the lowest RMSE values for all forecast horizons. But still, the results are only slightly better than the RMSE values of SVR GS.

Considering its ability to self-adjust, the weighted ensemble method should be the first choice for a robust SVR PV power predictor. Furthermore, it is applicable after only a few days instead of one year, like the grid-search model would be. Still, the optimal number of ensemble members as well as evaluation days is not directly applicable for every PV system and setting but, as Figure 6.7 showed, trade-off or rule of thumb settings can be found, e.g., applying half of the members and half of the days in the training set, which achieve nearly the same results as the optimized version.

If the computational power required for calculating this many ensemble members is not given, at least a grid-search on historic data is recommended on a single PV system level, as a general model is not always appropriate to achieve high quality forecasts, seen by the performance of $P_{SV,comb}$ in this chapter.

6.4. Concluding Remarks

In this chapter, support vector regression (SVR) PV power forecasts and settings are optimized and specifically tailored for a single PV system instead of relying on general settings as in Chapter 4. For this analysis, a PV system with high data availability throughout the years 2014 to 2016 is selected. Additionally, the PV measurements of nine surrounding PV systems are applied as input data as well. The goal of this chapter is to create a modeling routine applicable for all single systems that is dynamically adapting to the recent conditions at a PV system's site.

In a first approach, the addition of neighboring systems' data to the input pattern is analyzed and tested with random forest's (RF) feature importance for multiple forecast horizons. The results show that the feature importances of neighboring measurements increase with the forecast horizon. Especially at longer forecast horizons, the importance of neighboring system's measurements is even higher than the single system's own output. Furthermore, the analysis shows that PV measurements of certain systems tend to be more important than others.

On this basis, all available data of all employed systems, including PV measurements, irradiance forecasts, and modeled clearsky irradiance is applied. To reduce the set of possible features from 50 to a more feasible size, while keeping the loss of information to a minimum, feature selection is used. Here, a method that uses the RF feature importance to order the features showed better results than using an incremental method based on Euclidian distances of the neighboring systems. Especially SVR and linear regression models profit from a preceding feature selection. In this analysis, the optimal number of features is highly versatile throughout the tested days. So that, for each forecast horizon the averaged best number of features is selected for further experiments. This average varies from 33 to 42 features and is increasing with the forecast horizon. The best features of all following modeling approaches for each forecast horizon are selected with the RF feature importance method. This ensures that the most important features are always available to the model, while the dimensionality of the input data space is reduced.

In the next step, the parameters of the SVR need to be adjusted. Therefore, different approaches for parameter optimzation are considered: parameter grid-search, moving average parameter selection, and ensemble forecasts. While the parameter grid-search is done on the data of the previous year, both other methods are only using the training dataset consisting of 50 days. All models are optimized using a thousand SVR parameter combinations. The optimal number of evaluation days and ensemble members of both approaches are optimized to assess the best possible PV power forecasts with these models.

The comparison of all models shows that the previous general model of Chapter 4 ($P_{SV,comb}$) is not a good choice for the tested single PV system. All other applied models are superior to $P_{SV,comb}$ on all employed forecast horizons. In the comparison, an oracle forecast, that always knows the best possible SVR parameter combination – which is naturally not applicable in real-life – creates the best forecasts by far. But all of the other tested methods for improving PV power forecast, through combining and optimizing SVR parameters, are creating high quality and robust predictions for all prediction horizons. If there is not enough computational power to calculate the amount of ensemble members

or parameter combinations in real-time at hand, at least a grid-search on historic data is recommended. Summarizing the results, optimizing the SVR models on single PV system level is vital for high quality forecasts and could even improve the overall performance of regional PV power forecasts considerably. For this, the developed methods need to be applied and evaluated on further PV systems, so that the effect on a regional level can be examined.

Part III.

Ending

7. Conclusion

An efficient integration of renewable energy sources, i.e., wind and solar, depends on accurate forecasts. These forecasts are needed to keep the electricity grid stable and simultaneously enable a profitable trading of energy. In this thesis, the focus lies on improving short-term photovoltaic (PV) power forecasting with the machine learning algorithm support vector regression (SVR). For this, the suitability of SVR is assessed and different methods for improving its forecasts are evaluated.

In Section 7.1, the thesis is summarized and the most important findings are discussed. Subsequently, potential future experiments, extensions, and evaluations are outlined in Section 7.2.

7.1. Summary

To achieve a quality improvement in solar forecasting by applying machine learning techniques, three questions are answered throughout this thesis. The first question is directed at the suitability of SVR for PV power forecasting. A general approach to assess SVR's capabilities is done in Chapter 4, where a comparison of SVR and physical parametric modeling – that combines PV simulation and persistence modeling – is presented.

To create a reasonable experimental setup for all methods, the input features of the parametric model are used exclusively during the evaluation. The tuning of SVR parameters is performed on a subset of all available data to lower the runtime of a grid-search with multiple parameter combinations. This subset is created by grouping PV systems and weather situations from the complete dataset with k-means clustering and selecting representative systems and days.

To assess the importance of the different input data sources, a single input comparison is performed. This analysis illustrates the distinct forecast horizons at which the input of each data source performs best. While forecasts that only use measurement data create good predictions up to an hour, the satellite-based cloud motion vectors (CMV) are more suited for forecasts with a horizon of one to about three hours ahead. For even longer horizons, forecasts using numerical weather predictions (NWP) perform best. A

combined predictor employing all mentioned input sources creates the best predictions for all forecast horizons.

Compared to a purely physical approach, i.e., without any statictical postprocessing, the SVR-based forecasts are able to create PV power forecasts with a higher accuracy. If postprocessing is applied to the output of the parametric model, the results shift and the physical approach gains the advantage. Nonetheless, SVR-based forecasts are still competitive and recommended if there are no detailed information about the underlying PV systems, i.e., installed capacities, tilt, orientation, and technology, given. These specifications are mandatory for modeling PV power output with an irradiance-to-power conversion method.

In Chapter 5, the usefulness of additional NWP weather parameters as input features for statistical learning methods is evaluated. Thus, addressing the question if more information from NWP models is increasing PV power forecast quality. The idea is, that more data could help to assess the current and future weather conditions at a PV system's site, and therefore enables SVR to create better forecasts.

Instead of using the preprocessed data of Chapter 4, the raw data (as delivered by the provider) is used. The spatial interpolation is done with k-NN; irradiance forecasts of the NWP model are temporally interpolated with a clear sky interpolation method, whereas the interpolation of all other features is linear.

The significance of single weather parameters of the NWP model is measured by using random forest's (RF) feature importance. The results of this study show the dominant importance of irradiance forecasts throughout the year. A comparison of PV power forecasts with iteratively addition of the ten most important features confirmed this fact, as after adding the single most important irradiance weather parameter, no considerable improvement in prediction quality is achieved.

Even though an analysis of special weather conditions, i.e., situations with long periods of fog and snowfall, shows a higher importance for corresponding NWP features, an advantage of using these features for PV power forecasting in the same situations is not established.

On this basis, for Chapter 6, NWP models only provide irradiance forecasts. To investigate further possible improvements by presenting more data to the SVR-based forecasts, measurement and forecast data of surrounding PV systems are added to the input pattern of the models in this chapter. The aim is to give a predictor the ability to learn spatio-temporal correlations between PV systems, e.g., the detection of a cloud moving towards the main system by noticing a decline of PV power at a neighboring PV system. To check this possible correlation, a single system with nine surrounding PV stations is selected and, in a first step, the relationship between the measurements of

these systems is analyzed. To some extent, the results confirm an increase of RF feature importance for measurements from neighboring PV systems with an increasing forecast horizons.

Afterwards, all available data of neighboring systems is included in the forecasting pattern. To reduce the complexity of forecast calculations, a selection of features utilizing RF feature importance performed on the training set showed better results than adding features iteratively based on distance.

Next, this selection method is applied for SVR-based PV power forecasting. To tune the parameters of the SVR, three different methods are evaluated: grid-search on the previous year of data, a moving average parameter selection to assess the best parameter combination of the previous days, and an (weighted) ensemble forecast using the averaged output of multiple SVR-based forecasts, that combines different parameter settings. Instead of using the previous year of data, the latter two methods solely need data from the training set for their optimization. The results show minimal differences between all three approaches, whereas the (weighted) ensemble method is recommended, especially if there is not much historic data available. This analysis shows that finding an optimal parameter combination for SVR is as complicated as predicting the weather conditions of a day at all, as almost each day has its own optimal setting. Nonetheless, all of the mentioned models performed better than the former best model of Chapter 4 applied to this specific single PV system. This indicates the potential improvements for PV power forecasting accuracy, that can be achieved by optimizing on a single PV system level. The introduced selection and optimization process is more flexible as it is able to adapt its forecasts dynamically to the conditions of the previous days, instead of using the same predefined settings for a longer period of time.

Overall, machine learning, mainly support vector regression, proves to be a serious alternative for providing accurate PV power forecasts. In particular, in situations where there is not much information about PV systems available, the advantages of machine learning become apparent. With specialized optimization and adjustment of SVR, it is possible to create self-adapting forecasting systems, that are able to adjust to the recent weather conditions at a PV system's site.

7.2. Outlook

In this section, further research topics, remaining evaluations, and open questions of this thesis are outlined.

Combination with Parametric Modeling and Heterogeneous Ensembles At the end of Section 4.5, a short analysis about the performance of SVR as a substitute for the linear combination approach of the physical model is done. As the focus of this thesis lies on purely data-driven methods, this is not further investigated at this point. The results of Chapter 4 demonstrated the quality of the physical modeling approach, especially for regional forecasts, so that using these forecasts as input for machine learning algorithms or as members of a forecast ensemble could lead to PV power forecasts with an even higher accuracy. Additionally, the construction of heterogeneous ensembles combining multiple machine learning models and/or physical modeling techniques is another logical step to take and already showed encouraging results in applications for wind power forecasting (see Heinermann et al. [24]).

Integration of Different NWP Models and Features In Chapter 5, single level weather forecasts of the ECMWF's NWP model are analyzed to assess their suitability as input features for PV power forecasting. Because all parameters of the NWP model are correlated (as the processes and calculations are the same), there is almost no information gain when looking at further weather parameters. More interesting would be the usage of data from more and different NWP models with higher temporal resolutions, because these additional forecasts could contribute new information to the PV power forecasting models. Furthermore, using dimensionality reduction algorithms, e.g., a principal component analysis, on all relevant weather parameters of a NWP model could lead to combined input values with a higher information gain.

Further Evaluations of Single System Optimization The results of Chapter 6 displayed the increase of accuracy of SVR-based power forecasting when models are optimized for each PV system. In this thesis, only one system with its neighboring PV stations is evaluated. To confirm the overall improvement, further evaluations on more PV systems are needed, so that the influence on regional forecasts can be analyzed as well. Moreover, applying neighboring systems' data from a larger area around a PV system could help at predictions with a longer forecast horizons, so that clouds are detected earlier.

Bibliography

[1] T. Bailey and A. K. Jain. A note on distance-weighted k-nearest neighbor rules. *IEEE Transactions on Systems, Man and Cybernetics*, 8:311–313, 1978.

[2] H. Beyer, C. Costanzo, and D. Heinemann. Modifications of the heliosat procedure for irradiance estimates from satellite images. *Solar Energy*, 56(3):207 – 212, 1996.

[3] H. G. Beyer, J. Bethke, A. Drews, D. Heinemann, E. Lorenz, G. Heilscher, and S. Bofinger. Identification of a general model for the MPP performance of PV-modules for the application in a procedure for the performance check of grid connected systems. In *19th EC Photovoltaic Solar Energy Conference*, pages 3073–3076, Paris, France, 2004.

[4] D. Bourges, B. Yearly variations of the Linke turbidity factor. In *Climatic data handbook of Europe*, pages 61–64. 1992.

[5] M. Brabec, E. Pelikán, P. Krc, K. Eben, and P. Musilek. Statistical modeling of energy production by photovoltaic farms. In *Electric Power and Energy Conference (EPEC), 2010 IEEE*, pages 1–6, Aug 2010.

[6] L. Breiman. Random forests. *Mach. Learn.*, 45(1):5–32, Oct. 2001.

[7] D. Cano, J.-M. Monget, M. Albuisson, H. Guillard, N. Regas, and L. Wald. A method for the determination of the global solar radiation from meteorological satellites data. *Solar Energy*, 37(1):31–39, 1986.

[8] C.-C. Chang and C.-J. Lin. LIBSVM – A library for support vector machines. Website, 2015. http://www.csie.ntu.edu.tw/~cjlin/libsvm/, retrieved: 2016-02-11.

[9] C. Chen, S. Duan, T. Cai, and B. Liu. Online 24-h solar power forecasting based on weather type classification using artificial neural network. *Solar Energy*, 85(11):2856 – 2870, 2011.

[10] T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13:21 –27, 1967.

[11] Deutscher Wetterdienst (DWD). COSMO-EU. Website, 2016. `http://www.cosmo-model.org/content/model/documentation/core/default.htm`, retrieved: 2016-02-11.

[12] D. Dumortier. The Satellite model of turbidity variations in Europe. Technical report, 1998.

[13] EUMETSAT. METEOSAT Second Generation Satellite. Website, 2016. `http://www.eumetsat.int/`, retrieved: 2016-02-11.

[14] European Centre for Medium-Range Weather Forecasts (ECMWF). Integrated forecasting system (IFS). Website, 2015. `https://software.ecmwf.int/wiki/display/IFS/Official+IFS+Documentation`, retrieved: 2016-02-11.

[15] European Centre for Medium-Range Weather Forecasts (ECMWF). Radiation quantaties in the ECMWF model and MARS. Website, 2017. `http://www.ecmwf.int/sites/default/files/radiation_in_mars.pdf`, retrieved: 2017-02-10.

[16] M. Fontoynont, D. Dumortier, D. Heinnemann, A. Hammer, J. Olseth, A. Skarveit, P. Ineichen, C. Reise, J. Page, L. Roche, and others. Satellight: a WWW server which provides high quality daylight and solar radiation data for Western and Central Europe. In *9th Conference on Satellite Meteorology and Oceanography*, pages 434–437. American Meteorological Society Ed., Boston, Massachusetts, USA, 1998.

[17] Fraunhofer Institute for Solar Energy Systems (ISE). Energy charts: Net installed electricity generation capacity in germany in 2016. Website, 2017. `https://www.energy-charts.de/power_inst.htm`, retrieved: 2017-03-08.

[18] Y. Gala, A. Fernández, J. Díaz, and J. R. Dorronsoro. Support vector forecasting of solar radiation values. In *Hybrid Artificial Intelligent Systems*, volume 8073 of *Lecture Notes in Computer Science*, pages 51–60. Springer Berlin Heidelberg, 2013.

[19] R. Genuer, J.-M. Poggi, and C. Tuleau-Malot. Variable selection using random forests. *Pattern Recognition Letters*, 31(14):2225 – 2236, 2010.

[20] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *J. Mach. Learn. Res.*, 3:1157–1182, Mar. 2003.

[21] A. Hammer, D. Heinemann, C. Hoyer, R. Kuhlemann, E. Lorenz, R. Müller, and H. Beyer. Solar energy assessment using remote sensing technologies. *Remote Sensing of Environment*, 86(3):423–432, 2003.

[22] A. Hammer, J. Kühnert, K. Weinreich, and E. Lorenz. Short-term forecasting of surface solar irradiance based on Meteosat-SEVIRI data using a nighttime cloud index. *Remote Sensing*, 7(7):9070–9090, 2015.

[23] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning.* Springer, 2001.

[24] J. Heinermann, J. Lässig, and O. Kramer. *Evolutionary Multi-objective Ensembles for Wind Power Prediction*, pages 92–101. Springer International Publishing, Cham, 2017.

[25] T. Hofmann, B. Schölkopf, and A. J. Smola. Kernel methods in machine learning. *Ann. Statist.*, 36(3):1171–1220, 06 2008.

[26] S. Jurado, J. Peralta, A. Nebot, F. Mugica, and P. Cortez. Short-term electric load forecasting using computational intelligence methods. In *Fuzzy Systems (FUZZ), 2013 IEEE International Conference on*, pages 1–8, July 2013.

[27] T. Klucher. Evaluation of models to predict insolation on tilted surfaces. *Solar Energy*, 23(2):111–114, Jan. 1979.

[28] G. Kopp and J. L. Lean. A new, lower value of total solar irradiance: Evidence and climate significance. *Geophysical Research Letters*, 38(1), 2011.

[29] P. Krömer, P. Musílek, E. Pelikán, P. Krc, P. Jurus, and K. Eben. Support vector regression of multiple predictive models of downward short-wave radiation. In *2014 International Joint Conference on Neural Networks, IJCNN 2014, Beijing, China, July 6-11, 2014*, pages 651–657, 2014.

[30] J. Kühnert. *Development of a photovoltaic power prediction system for forecast horizons of several hours.* PhD thesis, University of Oldenburg, 2016.

[31] J. Kühnert, E. Lorenz, J. Betcke, A. Hammer, and D. Heinemann. Regionale PV-Leistungsvorhersagen für den Kurzzeit-Bereich auf Basis von Satellitendaten, numerischen Wetterprognosen und PV-Leistungsmessungen. In *Tagungsband des 29. Symposiums Photovoltaische Solarenergie*, pages 162–163, 2014.

[32] J. Kühnert, E. Lorenz, and D. Heinemann. Satellite-based irradiance and power forecasting for the german energy market. In *Solar Energy Forecasting and Resource Assessment*, pages 267–297. Academic Press, 2013.

[33] J. Li, J. K. Ward, J. Tong, L. Collins, and G. Platt. Machine learning for solar irradiance forecasting of photovoltaic system. *Renewable Energy*, 90:542 – 553, 2016.

[34] S. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, Mar 1982.

[35] E. Lorenz and D. Heinemann. Prediction of solar irradiance and photovoltaic power. In *Comprehensive Renewable Energy*, volume 1, pages 239–292. Elsevier Ltd., 2012.

[36] E. Lorenz, D. Heinemann, and A. Hammer. Short-term forecasting of solar radiation based on satellite data. In *Proceedings of EuroSun 2004, Freiburg, Germany*, pages 841–848, 2004.

[37] E. Lorenz, J. Hurka, D. Heinemann, and H. Beyer. 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, 2009.

[38] E. Lorenz, J. Hurka, D. Heinemann, and H. G. Beyer. 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, March 2009.

[39] E. Lorenz, J. Kühnert, and D. Heinemann. Short term forecasting of solar irradiance by combining satellite data and numerical weather predictions. In *Proceedings of 27th EUPVSEC*, pages 4401–4405, Frankfurt, Germany., 2012.

[40] G. Louppe, L. Wehenkel, A. Sutera, and P. Geurts. Understanding variable importances in forests of randomized trees. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 431–439. Curran Associates, Inc., 2013.

[41] P. Mandal, S. T. S. Madhira, A. U. haque, J. Meng, and R. L. Pineda. Forecasting power output of solar photovoltaic system using wavelet transform and artificial intelligence techniques. *Procedia Computer Science*, 12:332–337, 2012.

[42] R. Marquez and C. F. Coimbra. Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the NWS database. *Solar Energy*, 85(5):746 – 756, 2011.

[43] A. Mellit. Artificial intelligence technique for modelling and forecasting of solar radiation data: A review. *International Journal of Artificial Intelligence and Soft Computing*, 1:52, 2008.

[44] C. Monteiro, T. Santos, L. Fernandez-Jimenez, I. Ramirez-Rosado, and M. Terreros-Olarte. Short-term power forecasting model for photovoltaic plants based on historical similarity. *Energies*, 6(5):2624–2643, May 2013.

[45] T. M. Oshiro, P. S. Perez, and J. A. Baranauskas. How many trees in a random forest? *Proceedings of Machine Learning and Data Mining in Pattern Recognition: 8th International Conference, MLDM 2012, Berlin, Germany, July 13-20, 2012.*, pages 154–168, 2012.

[46] H. T. Pedro and C. F. Coimbra. Assessment of forecasting techniques for solar power production with no exogenous inputs. *Solar Energy*, 86(7):2017 – 2028, 2012.

[47] S. Pelland, J. Remund, J. Kleissl, T. Oozeki, and K. De Brabandere. Photovoltaic and solar forecasting: State of the art. http://www.iea-pvps.org/index.php?id=1&eID=dam_frontend_push&docID=1690, 2013.

[48] scikit learn. Machine learning in python. Website, 2016. http://scikit-learn.org/, retrieved: 2016-02-11.

[49] A. Skartveit and J. A. Olseth. A model for the diffuse fraction of hourly global radiation. *Solar Energy*, 38(4):271–274, Jan. 1987.

[50] A. J. Smola and B. Schölkopf. A tutorial on support vector regression. *Statistics and Computing*, 14:199–222, 2004.

[51] V. N. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag New York, Inc., NY, USA, 1995.

[52] H. Wirth and K. Schneider. Recent facts about photovoltaics in germany. Website, 2017. https://www.ise.fraunhofer.de/content/dam/ise/en/documents/publications/studies/recent-facts-about-photovoltaics-in-germany.pdf, retrieved: 2017-03-24.

[53] B. Wolff, O. Kramer, and D. Heinemann. *Selection of numerical weather forecast features for PV power predictions with random forests*, pages 78–91. Springer International Publishing, Cham, 2017.

[54] B. Wolff, J. Kühnert, E. Lorenz, O. Kramer, and D. Heinemann. Comparing support vector regression for PV power forecasting to a physical modeling approach using measurement, numerical weather prediction, and cloud motion data. *Solar Energy*, 135:197 – 208, 2016.

[55] B. Wolff, E. Lorenz, and O. Kramer. *Statistical learning for short-term photovoltaic power predictions*, pages 31–45. Computational Sustainability, Springer, 2016.

[56] J. Zeng and W. Qiao. Short-term solar power prediction using a support vector machine. *Renewable Energy*, 52(0):118 – 127, 2013.

List of Figures

List of Figures

112

List of Tables

List of Tables

Erklärung

Hiermit versichere ich, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Außerdem versichere ich, dass ich die allgemeinen Prinzipien wissenschaftlicher Arbeit und Veröffentlichung, wie sie in den Leitlinien guter wissenschaftlicher Praxis an der Carl von Ossietzky Universität Oldenburg und den DFG-Richtlinien festgelegt sind, befolgt habe. Des Weiteren habe ich im Zusammenhang mit dem Promotionsvorhaben keine kommerziellen Vermittlungs- oder Beratungsdienste in Anspruch genommen.

Oldenburg, den 12. April 2017

—————————————————

Björn Wolff

List of Tables

Acknowledgements