



# High resolution solar irradiance forecasts based on sky images

Von der Fakultät für Mathematik und Naturwissenschaften  
der Carl von Ossietzky Universität Oldenburg  
zur Erlangung des Grades und Titels eines  
Doktors der Naturwissenschaften (Dr. rer. nat.)  
angenommene

## Dissertation

von **Thomas Schmidt**  
geboren am 26.02.1986 in Bad Berleburg

ERSTGUTACHTER: **Prof. Jürgen Parisi**  
ZWEITGUTACHTER: **Prof. Carsten Agert**

TAG DER DISPUTATION: **14. Juli 2017**

# Erklärung der Selbstständigkeit

Hiermit versichere ich, die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie die Zitate deutlich kenntlich gemacht zu haben. Die Dissertation hat weder in ihrer Gesamtheit noch in Teilen einer anderen Hochschule zur Begutachtung in einem Promotionsverfahren vorgelegen.

Oldenburg, den 24.April 2017

Thomas Schmidt



## Kurzfassung

*Kürzestfristvorhersagen der solaren Einstrahlung bis zu 30 Minuten im Voraus mittels bodengebundener Wolkenkameras stellen eine neue Methode der Prognose der solaren Energieerzeugung dar. Hochaufgelöste 180°-Bilder des Himmels, aufgenommen mit Fischaugenkameras, sind im Gegensatz zu traditionellen Datenquellen aus numerischen Wettermodellen und Satellitenbildern in der Lage, kleinskalige Wolkenstrukturen zu erfassen. Die Aufnahmen sind die entscheidende Basis für die präzise und verlässliche Prognose der kurzfristigen Wolkenbedeckung und bodennahen Einstrahlungsbedingungen. Die Prognosen können in unterschiedlichen Anwendungen, z.B. für die Erhöhung der Netzstabilität, optimiertes Speichermanagement, kurzfristigen Energiehandel oder zur Erhöhung des PV-Anteils in hybriden Offgrid-Systemen zum Einsatz kommen. Diese Arbeit stellt ein neues, für verschiedene Anwendungen entwickeltes Wolkenkamera-basiertes Vorhersagemodell vor. Der Schwerpunkt der Arbeit liegt auf der detaillierten Beschreibung der Modellkomponenten, der Diskussion von methoden- und modellspezifischen Eigenschaften sowie der Evaluierung der Vorhersagequalität. Die Kernmodule des Modells sind die Wolkenerkennung, die Analyse der Wolkenbewegung, die Projektion der Wolkenschatten und die Strahlungsmodellierung. Das Modell nutzt Aufnahmen der Kamera, lokale Strahlungsmessungen und Abschätzungen der Wolkenbasishöhe um eine Verteilung der kurzfristigen bodennahen Einstrahlung zu prognostizieren. Daraus können Zeitreihen der Globalstrahlung mit einer Auflösung von bis zu einer Sekunde für jeden Standort in der von der Kamera abgebildeten Umgebung abgeleitet werden. Ohne Information der Wolkenbasishöhe können nur Vorhersagen für den Standort der Kamera mittels Prognosen der Verdeckung der Sonne durch Wolken erstellt werden. Das modulare Design des Modells erlaubt die flexible Einbindung verschiedener Datenquellen und neuer Methoden. Umfangreiche Datensätze von Kamerabildern, Strahlungsmessungen und teilweise Wolkenhöheninformationen und PV Leistungsmessungen bilden die Basis für eine tiefgehende Analyse der Vorhersagequalität. Hervorzuheben ist eine starke Abhängigkeit des Vorhersagefehlers mit dem vorherrschenden Bewölkungstyp sowie dem Tages- und Jahresgang. Zur Ermittlung der Vorhersagegüte werden die Modellprognosen hierbei mit Persistenzprognosen verglichen. Eine positive Vorhersagegüte kann vor allem unter stark variablen Einstrahlungsbedingungen, die häufig mit vereinzelter oder durchbrochener Bewölkung einhergehen, festgestellt werden. In homogenen Situationen, zum Beispiel bei wolkenfreiem oder vollständig bedecktem Himmel ist der Persistenzprognosefehler nur schwer zu unterbieten. Methodenspezifische Prognoseunsicherheiten werden insbesondere durch die Perspektive der Kamera auf Wolken nahe des Horizonts verursacht. Ergebnisse eines Kürzestfristvorhersage-Experiments für ein netzfernes PV-Diesel Hybridsystem zeigen eine gute Verlässlichkeit in der Prognose des Eintretens von Bewölkung nach wolkenfreien Situationen.*

# Abstract

*Very short-term solar forecasts based on sky images of ground-based cameras introduce a new forecasting methodology for solar energy applications which covers forecast horizons up to 30 minutes. High resolution images of the sky, recorded with cameras equipped with fisheye lenses providing a 180°-field of view, are able to resolve small-scale cloud cover and can therefore overcome the lack of spatial and temporal resolution of traditional solar forecasting methodologies like numerical weather prediction models or satellite image-based forecasting. The images are the fundamental information source needed for accurate and reliable short-term forecasts of cloud cover and surface solar irradiance distribution. These forecasts can be used in a variety of applications, e.g. for increasing grid stability, optimizing storage management, short-term trading or enhancing the PV hosting capacity in offgrid PV-Diesel hybrid systems. In this thesis, a newly developed image-based forecasting model for the usage in different applications is presented. The focus is put on the detailed description of the model components, the discussion of the methodology-inherent and model-specific characteristics and the evaluation of its forecast performance. The core components of the model are cloud detection, cloud motion tracking, cloud shadow projection and irradiance modelling. The model takes raw camera images, local solar irradiance measurements and cloud base height estimations as input data and provides estimations of near-future surface solar irradiance distribution. Time series of global horizontal irradiance can be derived with a resolution of up to 1 second for any location covered by the field of view of the camera. Without cloud base height information a local site forecast based on the prediction of the sun-occlusion by clouds can be processed. A modular design of the model allows the flexible handling of different data sources and the implementation of new methods. Large data sets comprising sky images, pyranometer measurements and in some cases cloud base height and PV power measurements are the basis for an in-depth analysis of its forecast performance. The results demonstrate a strong dependency of forecast error on predominant cloud conditions and also season and daytime. Model forecasts are compared with persistence forecasts to evaluate its skill. The model achieves positive forecast skill especially in times of high irradiance variability caused by scattered or broken clouds. In homogeneous sky conditions like clear sky or overcast, the model can hardly outperform persistence. Methodology-inherent uncertainties are especially introduced for clouds seen near the horizon. Results of a short-term forecast experiment for an off-grid PV-Diesel hybrid system show promising reliability in predicting cloud events after clear sky periods.*

# Contents

|       |                                                                           |    |
|-------|---------------------------------------------------------------------------|----|
| 1     | Introduction                                                              | 1  |
| 2     | Background                                                                | 3  |
| 2.1   | Camera based solar forecasting - state of the art                         | 3  |
| 2.2   | Very short-term solar irradiance fluctuations                             | 5  |
| 2.3   | Solar energy related applications                                         | 11 |
| 2.3.1 | Large grid connected systems                                              | 11 |
| 2.3.2 | Off-grid / Remote area electricity systems                                | 12 |
| 3     | Data sets                                                                 | 15 |
| 3.1   | University of Oldenburg, Germany                                          | 15 |
| 3.1.1 | Sky imager                                                                | 16 |
| 3.1.2 | Meteorological and radiation measurements                                 | 17 |
| 3.1.3 | Cloud height measurements                                                 | 18 |
| 3.1.4 | Forecasts                                                                 | 18 |
| 3.2   | HOPE measurement campaign, Juelich, Germany                               | 18 |
| 3.2.1 | Sky imager                                                                | 19 |
| 3.2.2 | Pyranometer network                                                       | 19 |
| 3.2.3 | Cloud height measurements                                                 | 20 |
| 3.2.4 | Forecasts                                                                 | 20 |
| 3.3   | 1-MW PV plant, Munich, Germany                                            | 20 |
| 3.3.1 | Autonomous sky imager                                                     | 20 |
| 3.3.2 | PV-Plant                                                                  | 21 |
| 3.3.3 | Forecasts                                                                 | 22 |
| 3.4   | Murdoch University, Perth, Western Australia                              | 22 |
| 4     | SkySol - Sky imager based solar irradiance analyses and forecasting model | 23 |
| 4.1   | Overview                                                                  | 23 |
| 4.2   | Camera model                                                              | 25 |
| 4.2.1 | Fisheye lenses                                                            | 26 |
| 4.2.2 | Projection model                                                          | 26 |
| 4.3   | Cloud detection                                                           | 31 |
| 4.3.1 | Postprocessing                                                            | 34 |
| 4.3.2 | Alternative methods                                                       | 34 |
| 4.4   | Cloud motion                                                              | 36 |
| 4.4.1 | Optical Flow                                                              | 36 |
| 4.5   | Cloud shadow mapping                                                      | 39 |
| 4.5.1 | Cloud mapping                                                             | 39 |

|       |                                                                       |    |
|-------|-----------------------------------------------------------------------|----|
| 4.5.2 | Shadow mapping . . . . .                                              | 41 |
| 4.5.3 | Cloud height estimations . . . . .                                    | 41 |
| 4.5.4 | Limitations in ground-based sky imaging . . . . .                     | 44 |
| 4.6   | Irradiance modelling . . . . .                                        | 47 |
| 4.6.1 | Transformation . . . . .                                              | 48 |
| 4.6.2 | Histogram method . . . . .                                            | 50 |
| 4.6.3 | Machine learning method . . . . .                                     | 50 |
| 4.6.4 | Measured components . . . . .                                         | 52 |
| 4.6.5 | Method comparison . . . . .                                           | 52 |
| 4.7   | Irradiance forecast . . . . .                                         | 53 |
| 4.7.1 | Forecast resolution . . . . .                                         | 53 |
| 4.7.2 | Post-processing . . . . .                                             | 54 |
| 4.8   | Additional components of SkySol . . . . .                             | 57 |
| 4.8.1 | Image features . . . . .                                              | 57 |
| 4.8.2 | Cloud type classification . . . . .                                   | 59 |
| 4.8.3 | Solar geometry . . . . .                                              | 62 |
| 4.8.4 | Clear sky irradiance modelling . . . . .                              | 62 |
| 4.9   | Spatial coverage and forecast horizon . . . . .                       | 63 |
| 4.9.1 | Modelling the spatial coverage of multiple cameras . . . . .          | 63 |
| 4.9.2 | Modelling maximum forecast horizon for varying cloud motion . . . . . | 64 |
| 5     | Forecast Performance . . . . .                                        | 69 |
| 5.1   | Error metrics . . . . .                                               | 69 |
| 5.2   | Handling of missing values . . . . .                                  | 71 |
| 5.3   | Single site forecast performance . . . . .                            | 72 |
| 5.4   | Seasonal and diurnal performance . . . . .                            | 76 |
| 5.5   | Cloud motion dependency . . . . .                                     | 78 |
| 5.6   | Cloud type dependency . . . . .                                       | 79 |
| 5.7   | Variability and forecast performance . . . . .                        | 81 |
| 5.8   | Performance for a 1-MW PV power plant . . . . .                       | 82 |
| 5.8.1 | Irradiance and PV power nowcast results . . . . .                     | 82 |
| 5.8.2 | Forecast results . . . . .                                            | 83 |
| 5.9   | Forecasting sunny or cloudy sky - a binary approach . . . . .         | 84 |
| 5.9.1 | Clear sky index threshold . . . . .                                   | 84 |
| 5.9.2 | Predicting cloud events for PV-Diesel hybrid systems . . . . .        | 85 |
| 5.10  | Model validation for spatially distributed locations . . . . .        | 91 |
| 6     | Conclusions . . . . .                                                 | 95 |

# 1. Introduction

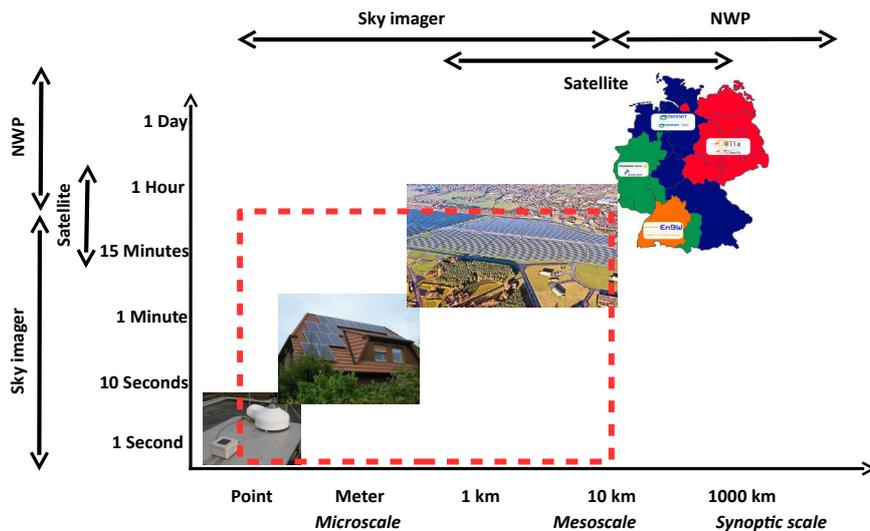
Since the first Intergovernmental Panel on Climate Change (IPCC) reports, the traditional energy supply system is about to change from a system based on fossil fuels to a system based on renewable energy sources. The extraction and burning of crude oil, natural gas, shale gas, hard coal and lignite coal from ground has accelerated in the past decades hand in hand with the global economic growth. As a result, atmospheric CO<sub>2</sub> levels are increasing and fueling climate change [1]. 2016 broke the record in global surface temperature since measurements started in late 19th century [2, 3]. Moreover, the fossil energy system depends on energy resources with limited reserves and limited access. These systems can in most cases be characterized as centralised systems concentrating power and capital. As a result, crude oil is still the key driver in geopolitical conflicts. Various economic systems highly depend on either the import or export of fossil fuels making them susceptible for economic and social crisis. Renewable energies including solar, wind, geothermal and hydro power are the climate neutral energy sources and key technologies to mitigate climate change and socio-economic crisis in the 21th century.

The energy transition introduces solar energy systems characterized as weather dependent and decentral. Especially weather dependency brings up new challenges for almost all stakeholders from producers to consumers of electricity. This thesis focuses on the meteorological small scale (micro- and mesoscale) weather patterns and the feasibility of its (very) short-term prediction for solar energy applications.

The availability of solar energy for applications transforming it into useful electricity or heat depends basically on the position of the sun relative to the plant. As a result, solar energy has a deterministic diurnal variation with peak yield at noon and no yield during night. For higher latitudes also seasonal variations become important. While these variations can be predicted accurately with low effort, the influence of clouds and aerosols on the amount of solar energy reaching the ground is incomparable more difficult to predict. So the history of meteorology and weather forecasting is much longer than the large scale utilization of solar energy for electricity production, the prediction of clouds is still one of the major challenges in weather forecasting. Due to the complex microphysics of clouds and the small scale dynamics, present-day numerical weather prediction (NWP) models are limited in describing the atmosphere accurately. They also lack temporal and spatial resolution to be able to make accurate

predictions of small scale atmospheric phenomena like cumulus clouds. Clouds geometry (shape, size and position) as well as their optical properties (thickness, transparency) influence surface solar irradiance on a scale of a few meters and in a time frame of seconds and even less. If these patterns should be analysed and predicted accurately, very high resolution sensors are needed. While remote sensing from satellites is very important for medium (meso-) and large (synoptic) scale weather monitoring and forecasting, their images will not be able to resolve these small scale structures. As a consequence, only ground based sensors can fulfill these requirements (Fig. 1.1). Digital cameras, equipped with fisheye lenses to capture the whole sky, have a long history in meteorological cloud observations. The sky images provide useful information of the current sky conditions and due to the high update rate also of its development. Based on the analysis of the images, an image of the future sky can be processed and fill the forecasting gap by providing a sub-kilometer resolution of cloud coverage. In the high resolution images, small scale cloud patterns can be detected precisely. Combined with additional information about the clouds height above ground, basic geometrical considerations and supported with machine learning, maps of local surface solar irradiance can be computed [4].

This thesis gives an overview of the state of the art in camera based solar forecasting (Section 2.1). The Section is followed by an introduction in specific characteristics of solar irradiance on the relevant temporal and spatial scales used in sky imager based forecasting (Section 2.2). An overview of typical applications, which potentially benefit from very short-term solar forecasting is given in Sec. 2.3. The main part of this thesis is the introduction in *SkySol*, the “sky imager based solar irradiance analyses and forecasting model” in Chapter 4, which has been developed during the PhD period. This chapter is followed by a comprehensive study on the forecast performance (Chapter 5) of *SkySol* on different data sets (introduced in Chapter 3).



**Figure 1.1:** Schematic overview of different forecast methods used for solar energy applications on different spatial and temporal scales. Sky imager based forecasting covers the area surrounded with red-dotted lines.

## 2. Background

### 2.1 Camera based solar forecasting - state of the art

A brief summary of recent and ongoing research and commercialisation of sky imager based solar irradiance forecasting is given in this section. Early research interest on hemispheric sky images has been reduced in the field of meteorology to classical cloud coverage and other synoptical observation related parameters. The usage for solar irradiance estimations or cloud cover and irradiance predictions for solar energy related applications is the result of the recently upcoming need for very short-term irradiance forecasts.

Strong contributions in the development of camera-based solar forecasting have been made by the *Solar Resource Assessment and Forecasting Lab* of the University of California in San Diego. Chow et al. [5] published first forecast results based on images from a *Total Sky Imager (TSI)* in 2011. They already implemented cloud base height (CBH) to estimate cloud layer position used to estimate distribution of cloud shadows on the ground. A simple irradiance modelling scheme has been used to provide irradiance time series. Later, the group developed the new UCSD Sky Imager *USI* [6], that provides improvements in image quality, especially by using *High Dynamic Range (HDR)* imaging technique as a step to overcome problems with brightness in the circumsolar area. Based on this camera, further articles on methods like cloud detection [7], cloud height estimation based on stereography [8] and cloud shadow speed [9], geometric calibration of the camera based on sun position [10], irradiance retrieval based on image characteristics [11] and forecasting [12] have been published by this group.

Next to this development, further international contributions regarding hardware, image processing techniques for cloud detection, cloud motion, cloud height estimates and irradiance forecasting have been published recently. While several contributions have demonstrated the potential of sky imager based short-term forecasting, forecast uncertainties or errors naturally remain. Regarding the inherent limitations of ground based sky imaging and the complexity of cloud development, further improvements will increase the model complexity by introducing cloud optical properties and more complex cloud motion and development [13]. Multiple cameras will be used to retrieve 3D

information about cloud distribution [14] and to cover whole regions (e.g. distribution grids). The usage of advanced machine learning methods (e.g. [15, 16]) and the integration of additional information (e.g. from satellite images or numerical weather models) are current research topics to enhance forecast accuracy.

First papers on the application of sky imager based forecast models underline the interest in the operational use of sky imagers for solar energy applications. For example, Nguyen et al. [17] has studied the impact of high PV penetration on a distribution network and its hosting capacity. Forecasts for utility scale PV power plants have been published by Chu et al. [16] for a 48 MW and by ourselves for a 1 MW PV plant [18]. Forecast data of the model presented in this thesis have been applied by Peters et al. [19, 20] to simulate the integration of short-term forecasts for the optimized operation of remote off-grid hybrid PV-Diesel networks.

The improvements in forecast reliability and accuracy in the past few years lead to the offer of a few commercial camera systems with basic forecasting software supplying a small but growing market (e.g. EKO “SRF-02”, Fulcrum3D “CloudCam”). Further activities will be driven by increasing shares of PV generation capacity in the electricity grid that goes along with an increased interest in very short-term forecasts for ramp smoothing, short-term trading, demand-side management, backup generation and storage management will increase as well.

In this thesis, a complete sky imager based forecast model for local and areal irradiance forecasts is presented. The model has been developed since the beginning of this thesis in autumn 2013 and improved continuously. Its performance has been evaluated on different data sets with different focus for each. The choice for specific algorithms used for this model bases on a weighting of accuracy, technical constraints and time effort. Several implemented methods like the cloud detection or the cloud motion scheme are based on existing methods and have been developed further. Machine learning-based models complement the model with an improved irradiance modelling and a cloud type classification scheme. The model has been applied to an utility scale PV plant and for a PV-Diesel hybrid system to demonstrate its skill in real applications. To account for the fast development of new methods, the forecast model has been build up as modular as possible to allow the implementation of different algorithms or methods for the same task. The option for the further development by different users should be feasible with a well-structured and clearly documented model algorithm.

Although a number of publications summarize and compare different proposed methods for a specific task, e.g. cloud motion, no comparison or benchmarking of different camera hardware and software developments in a parallel setup or measurement campaign for irradiance forecasting has been carried out yet. As this is the case for the proposed sky imager based forecast model, no quantitative comparison to the performance of hardware and software developed by other groups can be provided in this thesis.

## 2.2 Very short-term solar irradiance fluctuations

Very short-term irradiance forecasts based on sky imagery show high temporal and spatial resolution but are limited in spatial coverage and forecast horizon. Regarding solar irradiance fluctuations on these scales, some important characteristics have to be highlighted.

During the advent of solar energy applications, intra-day, day to day, and seasonal variability was investigated for the purpose of solar resource assessment, energy trading and dispatching of fossil backup capacities. Short-term variability has been attracted the attention of researchers recently, because power quality (e.g. voltage and frequency stability) can be negatively affected by variable PV power output. New data sets from measurement campaigns provide higher station densities of distributed pyranometers that sample at high frequencies of at least 1 Hz (e.g. HOPE campaign [21]). These high resolution measurements allow the research community to study spatio-temporal irradiance variability in more detail. With increasing share of PV generation, ramp rate reduction by spatial smoothing of distributed PV or the implementation of storage capacities becomes more and more important (e.g. [22, 23]). For example, Lohmann et al. [24] state, that “there is a need to understand the biases in representation of temporal variability resulting from temporally coarse-resolution observations [25], as well as how spatial averaging (as would come from having distributed PV over a region) mitigates variability [26]”.

Solar irradiance variability has three main causes:

1. Deterministic diurnal and seasonal variations due to astronomical earth-sun constellations.
2. Variations in cloud cover, cloud motion and cloud optical properties.
3. Variations in atmospheric composition (excluding clouds).

Regarding the scale covered by sky imagery (forecast horizon up to 30 minutes), variations due to apparent motion of the sun are rather small. For larger scales, the deterministic variations of the first mentioned category play a more important role, but can be computed with relatively little effort. Among the atmospheric compounds, aerosols and water vapour are the major contributors to solar irradiance variability. Other atmospheric compounds either show little variability (i.e. their influence is predictable), or the effect is small or local. On short timescales and for small areas, considerable impacts of aerosols on solar irradiance variability will be limited to special locations which are affected for example by local dust events. If the actual atmospheric composition is known, the impact on solar irradiance extinction will likely stay constant and is therefore predictable in the forecast horizon of interest. Therefore, the major contributor to local short-term irradiance variability is variability in cloud cover. To reduce short-term irradiance variability to its cloud-induced fraction, the clear sky

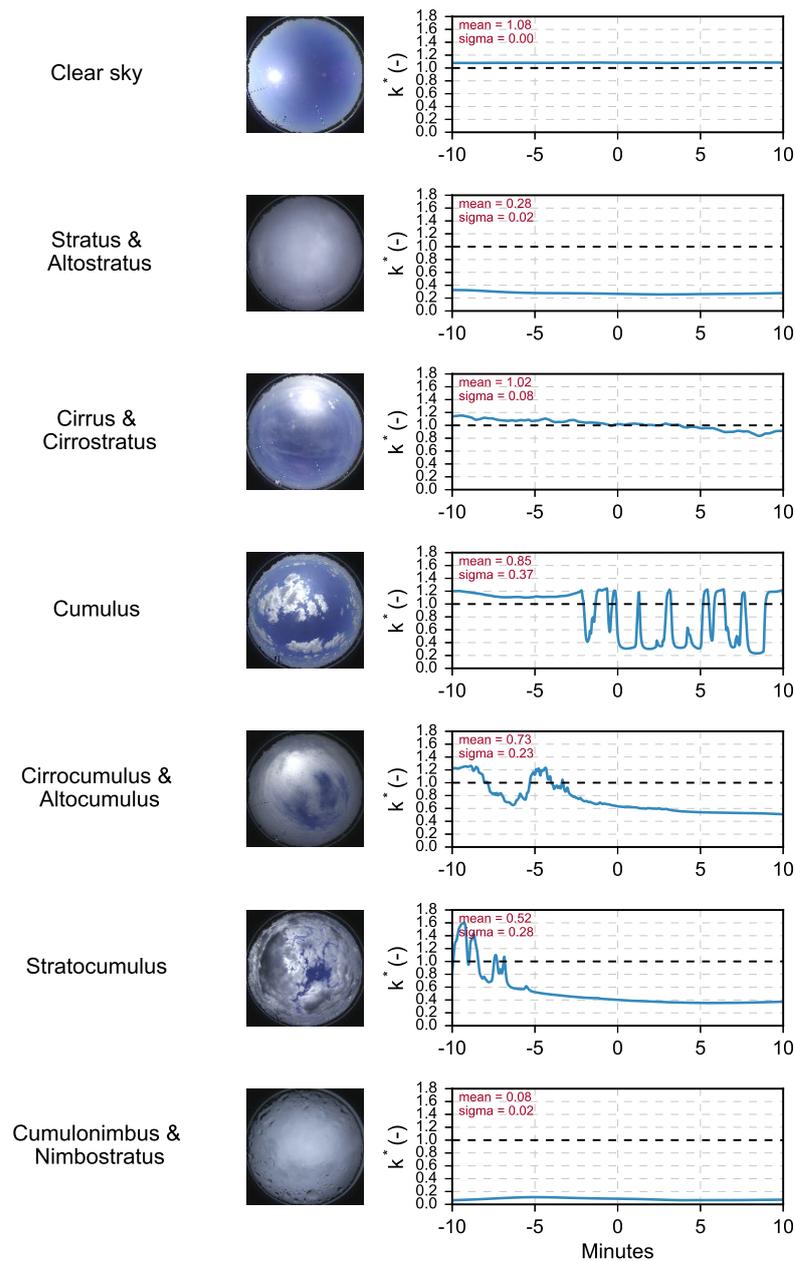
index  $k^*$  (Eq. 2.1) is widely used. The clear sky index is the ratio of the absolute measured irradiance  $G$  by its modelled clear sky (cloud free) irradiance  $G_{clear}$  (see Sec. 4.8.4 on clear sky irradiance models)

$$k^* = \frac{G}{G_{clear}}. \quad (2.1)$$

The complex interactions between land, ocean, atmosphere and the sun as the main energy source lead to the formation, advection and dissolution of clouds. Clouds are the result of water vapour condensation, that occurs when humid air cools down below its so called dew point (i.e. the temperature at which relative humidity reaches 100 %). Clouds consist of water droplets, ice particles or a mixture of both. The scattering of sunlight by clouds is described by the Mie-Debye scattering for cloud particles with a similar diameter to the wavelengths of the scattered light and geometrical optics for larger cloud droplets and ice particles [27]. The amount of solar irradiance reaching the ground under cloudy conditions therefore varies depending on the whole cloud cover and the cloud optical properties (e.g. transparency).

Cloud-induced solar irradiance variability varies with cloud size distribution, transparency and also cloud speed. With cloud sizes ranging from several meters to kilometres [28], and cloud speeds reaching more than 100 km per hour the expected single point solar irradiance variability varies also in frequency and amplitude. Figure 2.1 shows typical 20 minute irradiance time series recorded 10 minutes before and after a corresponding sky images. The seven cloud classes introduced here are meteorologically justified by Heinle et al. [29] and will be used in this thesis to differentiate forecast performance under different cloud conditions (Sec. 4.8.2). Irradiance variability characteristics in Fig. 2.1 differ considerably between cloud types. While clear sky and overcast conditions result in rather homogeneous irradiance time series, the cloud size distribution and the optical properties decide about the repetition rate of fluctuations and their magnitude in mixed cloud conditions. Being aware of this strong dependency of the irradiance variability characteristics on different cloud properties is the basis for the successful development of an advanced sky image-based surface solar irradiance forecasting model. For example, the homogeneous conditions are of minor importance and persistence models, which are based on actual measured conditions, are virtually perfect due to the persistence in clear sky index. However, the accurate prediction of the beginning of a highly variable mixed-cloud time period is of particular importance, if electrical systems are backed with flexible generation or storages that have to be activated to ensure a stable operation of the electrical grid.

A bachelor's thesis on the occurrence and the statistical characteristics of ramp events measured in Oldenburg has been conducted on a one-year data set of 1 Hz irradiance measurements with a photodiode pyranometer [30]. The author found an average ramp duration of 23 seconds (maximum to minimum measured global horizontal irradiance) and a total number of up to 1100 detected significant ramps per week in summer times. The short duration time and the large overall number of ramps depict the need for high pyranometer sampling rates on the one hand, and the relevance of small scale irradiance variability on the other hand [24].

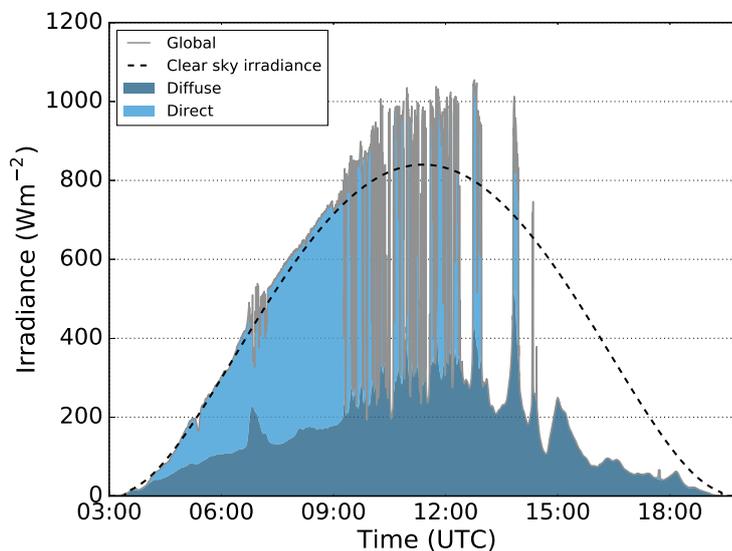


**Figure 2.1:** Typical examples of seven different cloud type classes. Sky images are given with corresponding time series (1 Hz sampling rate) of clear sky index 10 minutes before and after the image was taken. The cloud types are in the order of their occurrence: Clear sky, Stratus and Altostratus, Cirrus and Cirrostratus, Cumulus, Cirrocumulus and Altocumulus, Stratocumulus, Cumulonimbus and Nimbostratus.

There are two main components of broadband (spectral irradiance is not considered in this thesis) irradiance, which are direct (normal) irradiance (DNI) as the irradiance that comes directly from the sunspot (full opening angle of  $5^\circ$  [31]) and diffuse (horizontal) irradiance (DHI), that originates from all directions in the sky except for the sunspot region. The sum of both components (after conversion of DNI to horizontal plane) is

called global irradiance. A common way to measure both components is to use two sensors on a two-axis sun tracker, so that the pyrliometer (measuring DNI) is always pointing directly into the sun, while tracker-mounted shadow balls continually shade the broadband pyranometer (measuring DHI). Figure 2.2 illustrates the share of both components and their variability during a day in May 2015. Under clear skies (in the morning), direct and diffuse irradiance variability is low and changes in absolute numbers follow the diurnal course in clear sky irradiance. In overcast conditions (afternoon) direct irradiance variability diminishes almost entirely. During noon, small scale cumulus clouds formed in the sky and caused considerable fluctuations in direct irradiance variability (frequent reductions to 0% due to shading by opaque clouds). Diffuse irradiance varies also during mixed conditions, but with an appreciably smaller amplitude, especially on the short time scales considered in this thesis.

Figure 2.2 also reveals another characteristic in noon hours only visible in high resolution data. Very short peaks in the time series protrude the diurnal shape of irradiance by several percentage. These enhancements originate from an unobscured solar disk in which situations the diffuse component of irradiance is increased due to surrounding clouds and measured global irradiance temporarily exceeds expected clear sky global irradiance (and sometimes even the solar constant of  $1367 \text{ Wm}^{-2}$ ) [32]. These enhancements are often neglected, because they usually average out when averages of 15 minutes or longer are considered. However, due to fixed inverter sizing in the design phase of a PV system, inverter saturation or “clipping” can lead to energy losses because of capping the maximum power output [32].

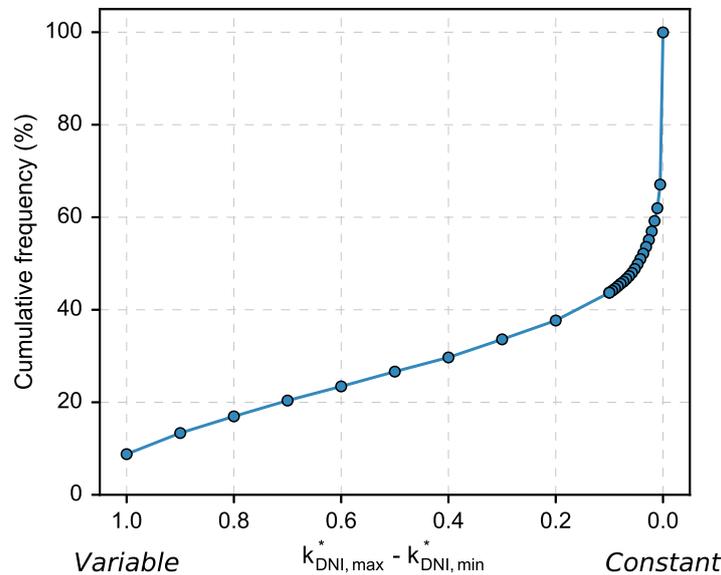


**Figure 2.2:** Time series of measured global, diffuse and direct irradiance on May, 17th 2015 in Oldenburg. The three main sky conditions clear (morning), mixed (noon) and overcast (afternoon) are present on this day. Sample rate: 1 Hz.

The potential for very short-term solar forecasting is not only determined by the amount of solar energy installed, but also by the location-specific probability of very short-term fluctuations (as far as is known no global climatology of very short-term solar variability

exists yet).

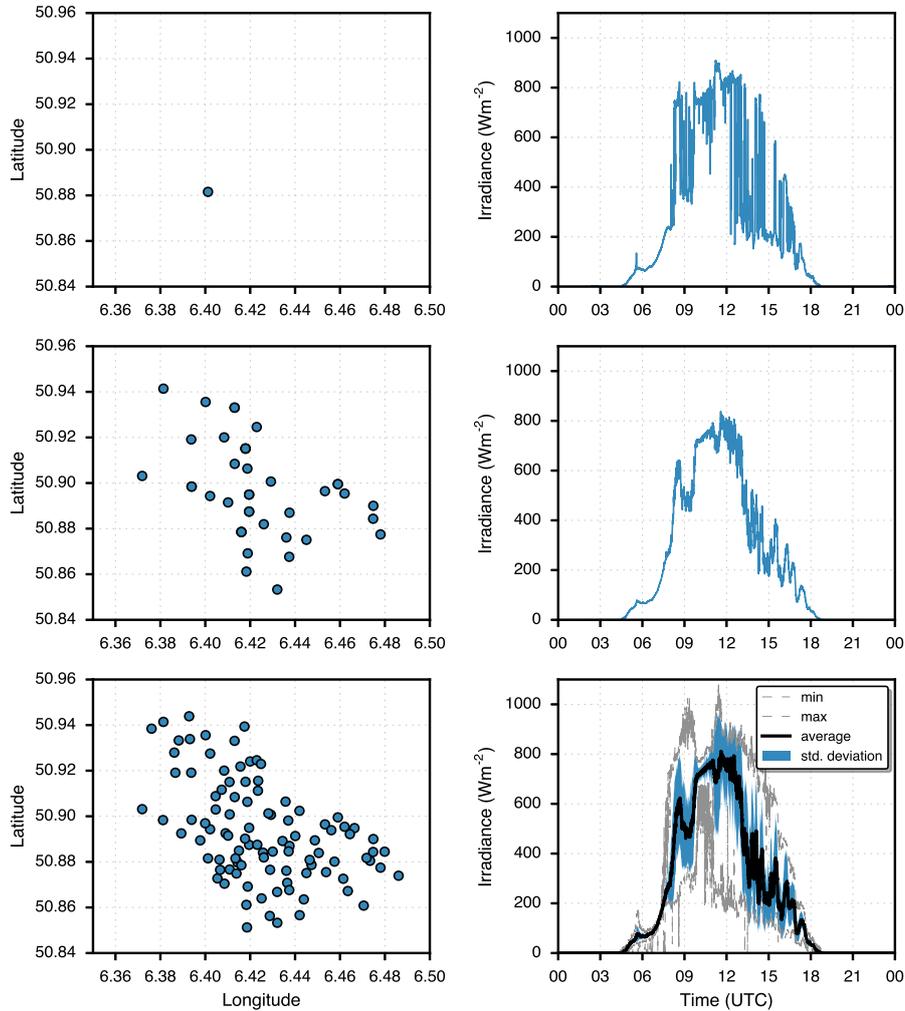
To illustrate the dimension and importance of short-term variability, an analysis of strong and short-lasting shading events is presented here, using direct normal irradiance data measured in Oldenburg in 2015 as an example. The measurements have been first normalized to its clear sky index  $k_{DNI}^*$  using model-derived direct clear sky irradiance (Eq. 2.1). Then, the data has been split into 15 minute time windows. For each 15 minute window, the maximum range  $k_{DNI,max}^* - k_{DNI,min}^*$  was determined. A low range in 15 minute  $k_{DNI}^*$  values indicates low variability, while higher ranges indicate strong variability in direct irradiance. The respective cumulative distribution of all windows is presented in Fig. 2.3. It shows, for example, that about 50 % of all periods feature low variability with a range in  $k_{DNI}^*$  of maximum 0.05, while about 20 % still exhibit ranges of at least 0.7, which indicates significant drops or peaks in DNI.



**Figure 2.3:** The share of 15 minute time periods in which the difference between the maximum and minimum value of clear sky indices of DNI ( $k_{DNI}^*$ ) is above a specific threshold. Modelled clear sky irradiance has been used to normalize DNI. Dataset: 1 Hz measurements of DNI (night time values removed), Oldenburg, 2015.

The spatial smoothing effect of irradiance variability can be demonstrated by evaluating data of spatially distributed high resolution measurements. The measurement campaign HOPE with 99 pyranometers deployed over an area of about 10 km x 12 km provides the necessary data to illustrate the smoothing effect in mixed cloud conditions. Figure 2.4 shows the daily time series of a single sensor during the campaign. Large fluctuations are present especially after noon. If 40 randomly picked sensors are averaged one can see a strong reduction in the short-term fluctuations. Obviously, small sized clouds induced short-lasting fluctuations at single sensors but are averaged to a medium insolation if the size of the averaging region is larger than the cloud size. If even more sensors are averaged, the additional smoothing in the irradiance time series is negligible

in this example. The variations still present are the result of domain-wide changes in cloud cover. In the worst case, clouds of the size of the whole domain can move across the sensor array. Here, the effect of averaging will be small and the resulting ramp of the aggregated domain will remain high (for example around 13 UTC in Fig. 2.4). The duration of such a ramp event strongly depends on cloud speed.



**Figure 2.4:** Daily time series of global horizontal irradiance (right panel) measured during the HOPE campaign [21] for one single pyranometer (top), average of 40 randomly chosen sensors (middle) and statistics of all 99 sensors (bottom). The geographical distribution of the sensors is given in the left panel.

Even if small scale phenomena are smoothed when spatial and temporal averages are used and evaluated, the demand for high resolution irradiance forecasts remains for various solar energy applications. Short-term drops or peaks in power output due to small clouds passing the plant, known as ramps, become important for several plant and grid operators, depending on size of the system, its grid connection, and the overall energy system configuration. Various examples of typical applications for very short term forecasts are discussed in the next chapter.

## 2.3 Solar energy related applications

Filling the gap of solar forecasting for very short-term forecast horizons with ground-based sky imagers fulfils a need for a variety of solar energy related applications. Several stakeholders, i.e. plant operators, grid operators or energy traders, have a potential interest in the short-term solar power output of the individual systems. They might either be interested in energy yield predictions for a trading horizon of the next 0-15 minutes, in ramp forecasting for dispatching backup capacities, in battery storage management, load shifting or grid management. Overall, the information about short-term solar power output will increase the systems' reliability and efficiency [33].

### 2.3.1 Large grid connected systems

The main driver for the development of the sky imager based forecast system presented in this thesis was the research done in the EU-project *PerformancePlus*. Research focused on the detailed modelling of PV systems under the influence of fluctuating solar input. Detailed modelling of PV systems on module or even cell level [34] requires high resolution solar irradiance and module temperature data input. Anagnostos et al. [35] demonstrated 15 minutes energy yield prediction results based on high resolution irradiance forecasts processed with the newly developed sky imager model and a detailed electro-thermal PV model. The authors argued, that capturing the short-term irradiance fluctuations for a more detailed and accurate energy yield prediction can contribute to the mitigation of ramping effects, the regulation of the grid and the usage in smart building applications to balance the energy production and demand. Ramp mitigation, especially for large PV plants and depending on the underlying grid, is an important tool to increase grid stability when large and short-term drops or increases in PV generation are critical for balancing production and demand. One proposed method for ramp mitigation is linear down-curtailment of PV generation before a forecasted down ramp [12].

Nguyen et al. [17] studied the impact of high PV penetration on a distribution network and its hosting capacity by using sky imagers for a high resolution resource assessment with power system simulation. Here, a main advantage of the sky imager compared to a ground-based sensor network or metering installations is the reduction to a single or a few devices, which are able to reconstruct the cloud shadow distribution on a large area. When studying distributed PV or distribution grids respectively, high resolution solar forecasting can also be used for PV inverter control and storage management to optimize the distributed generation while reducing the impacts from high penetration on the grid.

Battery storage management can benefit from very short-term solar forecasting by smart charging and discharging coupled with load- and market-pricing information [33]. Smart algorithms incorporating generation and load forecasts can reduce the number

of charging cycles and increase battery lifetime [36]. Increasing the battery's lifetime will in turn make the storage economically more feasible.

In order to gain benefit through very short-term trading of PV power, market conditions have to be considered. When trading intervals and horizons will be reduced to intervals in the order of a few minutes and if update rates are increased, sky imager based short-term solar forecasting can contribute to less energy yield forecasting errors and therefore reduce overall system costs. Bilateral trading, e.g. between plant owners and consumers (e.g. industries with flexible loads, power to gas/heat facilities) or even between small-scale producers with rooftop PV capacities and neighborhood consumers could contribute to better PV integration, e.g. in smart grids.

### 2.3.2 Off-grid / Remote area electricity systems

The role of very short-term irradiance forecasts in off-grid, respectively remote area electricity systems, is somehow different to the usage in grid connected PV systems. Off-grid systems, that nowadays rely significantly on electricity provided by diesel based generators, benefit even more than grid-connected systems from the integration of solar energy capacities by the substitution of expensive diesel fuel. Diesel generators are widely used in off-grid systems because of their small to medium size along with a relatively high flexibility needed to balance higher load variability compared to large electricity grids with large coal or nuclear power plants. High costs for diesel result mainly from long distances (islands or remote areas) increasing the cost for transportation. As a consequence, the enhanced economic benefit (Levelized Costs of Electricity LCOE) of using PV is the main driver for the demand of significantly more PV installations in these systems. However, the generally small spatial footprint of PV installations in such electricity grids lead to considerably high PV variability. For example, a small island or remote community or industry can be shaded by single fast moving clouds in a timespan of a few seconds to minutes reducing and increasing the PV output dramatically in a short time. In order to keep generation and consumption in these electrical grids balanced to ensure power quality and grid stability, flexibility of either diesel generators or storage have to be provided to deal with the introduced variability. If no or not enough storage is available, grid operators often limit the maximum PV penetration to certain levels often defined by the capability of the diesel generators to cope for the maximum expected PV (and load) fluctuations [37].

The systems flexibility to cope with PV generation variability includes spinning reserve and step load capability. Spinning reserve is the extra generation capacity provided by the system to account for sudden changes in net load (e.g. PV output or consumption). It is usually provided by the diesel generators that do not run at maximum load. In that case, they are able to increase their load in short time. Therefore, higher PV generation will also introduce the need for more spinning reserves. Storage systems can also provide spinning reserve. Another option is to use extra diesel generators that are switched off in times of high PV generation and only used when PV output is predicted to be reduced due to cloud shading. The need for a specific starting time (around

30-120 seconds, numbers provided by different personal communication) motivates the benefit of short-term forecasts that cover this critical time window. We investigated the reliability of short-term forecasts to predict the transition from clear sky to variable conditions in the horizon of 2-5 minutes [38]. Step-load capability describes the ability of the system to react on different ramp rates. In case of high PV shares, battery storage or inverter control can be used to smooth the generation ramps introduced by cloud-induced PV variability. Short-term forecasts can also contribute to a more efficient ramp-rate control, which reduces the number of storage cycles or generator step-loads increasing as a result generator or battery lifetime [39].

Overall, the role of solar forecasting for off-grid systems is to maximise solar energy penetration, to minimise conventional fossil fuel use and corresponding costs while ensuring grid reliability, stability and quality of supply [4].

The final usage and benefit of short-term forecasts for remote hybrid electricity grids can vary with its specific design. Peters et al. [19] integrated short-term solar forecasts in predictive control simulations for a PV-Diesel system. They also investigated the impact of solar forecasting on fuel savings, power quality and PV curtailment while integrating battery storage in the simulation [20]. The results underline the dependency of the benefit of short-term forecasts on the specific design of the system.



## 3. Data sets

Multiple data sets have been investigated during the development and evaluation of the sky imager based forecast system. The most important data basis originates from the measurement station at the University of Oldenburg (Sec. 3.1). This data set comprises multi-year archives of sky images, radiation measurements and auxiliary meteorological and PV-related measurements at a local site. It has been mainly used to develop the forecast algorithms, for the evaluation of multi-year forecast performance (Sec. 5.3) and some complementary studies presented in the section about the model description itself (Chapter 4). Unfortunately, no distributed irradiance measurements or multiple sky imager are located yet in the surrounding area to provide additional data for a more comprehensive study of spatial irradiance distribution.

For this study, sky images, irradiance measurements of distributed pyranometers and cloud height measurements for the geolocation of clouds has been used from a temporarily limited measurement campaign (Sec. 3.2). It provides an excellent data set for the evaluation of spatial effects of camera based analyses and forecasting [40].

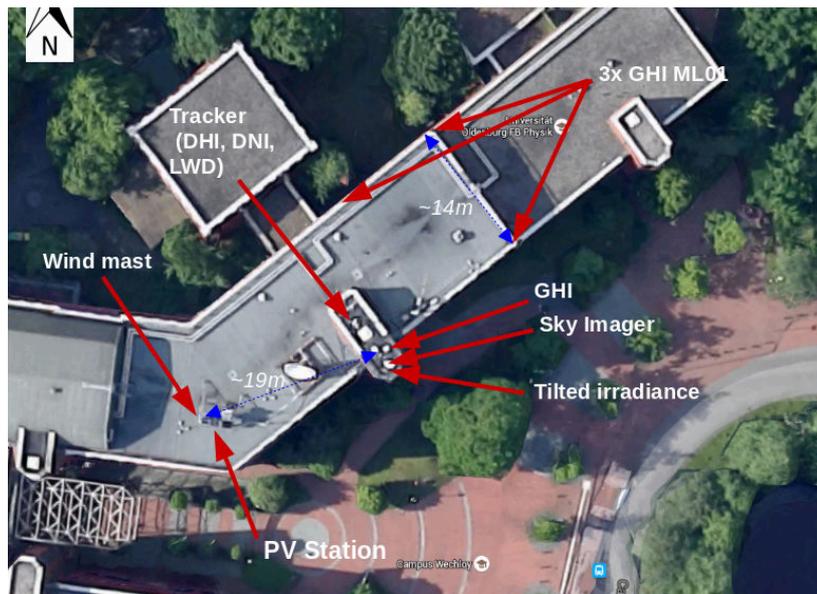
Furthermore, the forecast performance has been demonstrated for a 1 MW PV plant at the rooftop of the trade fair building in Munich (Sec. 3.3). Here, a self-developed autonomous sky imager system has been installed to collect the necessary images.

A local dataset from Western Australia (Sec. 3.4) was investigated for the purpose of the optimal control of fossil fuel based backup capacities (diesel generator in this case) with solar forecasting.

### 3.1 University of Oldenburg, Germany

The most comprehensive data set used for the studies has been acquired at the University of Oldenburg itself. On the rooftop of building W2 on campus Wechloy (53.15232 °N, 8.166022 °E), an energy meteorology station equipped with several sensors measuring radiation, wind speed and direction, air temperature, relative humidity and PV related parameters like PV maximum power point (MPP) current and voltage,

module temperature and 3D wind conditions close to the module has been initiated in autumn 2013. Up to now, all sensors record measurements in a high temporal resolution of 1 to 0.1 second. Moreover, a sky imager takes images every 10 seconds from sunrise to sunset. The station has been set up in the framework of the EU research project PerformancePlus (grant agreement no: 308991). In the framework of this project, the data has been used for the development and validation of a detailed optical-thermal-electrical PV model [41] and a short-term irradiance forecasting model based on sky images [35], the object of this thesis. An overview of the current setup is given in Figure 3.1.



**Figure 3.1:** Top view on energy meteorology station at University of Oldenburg. Three main components of horizontal irradiance (GHI, DHI, DNI) are located together with a sky imager on the staircase of the building. A PV station with wind measurements is located about 19 meters nearby. A triangle of GHI measurements with photodiodes for cloud motion estimation is also located on the rooftop. Background source: Google

### 3.1.1 Sky imager

The sky imager used in Oldenburg is a commercial *Vivotek FE8172V* network camera equipped with a fisheye lens (Fig. 3.2). The typical application of this camera type is surveillance of wide, open areas. Compared to cameras developed specifically for sky and cloud observations, commercial network cameras are less expensive (<1k €) and therefore of interest for solar energy applications. The most important technical specifications of the camera are a full 180° field of view, a circular fisheye frame in a 1920 x 1920 pixels image plane and a dynamic range of more than 57 dB. The camera configurations including colour settings, white balance and exposure settings are applied browser-based. A Python-based interface has been written to control most of the settings fully automatic.

The camera was installed in autumn 2013 and configured to record images every 10 seconds with an automatic exposure time. In summer 2016, the follow-up model *FE8174V* has been mounted next to it to compare the image quality of both cameras (see Fig. 3.2 left). An example image with additional labels marking the hemispheric directions is shown in Figure 3.2. Obviously, no big obstacles are in the field of view of the camera. The main irradiance sensors are located nearby. The sun tracker equipped with a pyrheliometer for DNI and a shaded pyranometer for DHI measurements is placed three meters north-west of the camera. Some lightning rods interfere the field of view in the north-west but do not cover large areas of the image. In summer 2016, the camera dome featured increasing number of scratches and had to be replaced.



**Figure 3.2:** Left: Sky imagers at the University of Oldenburg. Right: Example image of the *Vivotek FE8172V* camera. The image has been rotated to point northwards at the top.

### 3.1.2 Meteorological and radiation measurements

For a comprehensive study of the performance of the sky imager based forecasts, additional ground-based measurements are required. A minimal setup comprises only the sky imager and one pyranometer measuring the global horizontal irradiance (GHI) with a high temporal resolution not larger than the image recording interval. In operational mode, these measurements are needed for the calibration of the irradiance modelling scheme, which is mapping cloud scenes to surface irradiance. Naturally, measurements are also needed for the validation of forecast accuracy. Direct normal irradiance (DNI) and/or diffuse horizontal irradiance (DHI) measurements are useful for gathering information about the part of broadband global irradiance originating directly from the circumsolar area (DNI) and from the overall sky except for the circumsolar area (DHI). In Oldenburg, ventilated *Kipp&Zonen CM11* are used to measure GHI and DHI (shaded on sun tracker). DNI is measured with a pyrheliometer (*Eppley NIP*) attached to the suntracker. The sensors are cleaned once a week to achieve a high data quality. Pyranometers are replaced for calibration every 2-3 years. At this point it has to be noted, that the *CM11* pyranometer is a thermopile pyranometer classified as a high quality sensor fulfilling the secondary standard norm by ISO 9060.

However, thermopile sensors in contrast to photodiode sensors that respond almost instantaneous to signal changes have a specific response time  $< 15$  s. Therefore, it is often stated, that the sensors are less suitable for measuring very rapid changes in radiation [42]. Nevertheless, thermopile pyranometers are still recognized as the most accurate sensor technology for the measurement of broadband irradiance [43]. Since forecasts are evaluated for this dataset on minutely averages, the influence from measurement uncertainties on forecast performance validation is assumed to be small.

A setup of three photodiode pyranometers has been installed in May 2015 for the estimation of cloud shadow motion. A master thesis [44] investigated the possibility of cloud height estimations based on ground-based cloud shadow speed measurements and cloud tracking in sky images. The arrangement and the methodology required higher sample rates of at least 10 Hz compared to the reference measurements sampling with 1 Hz. More details on the method are given in Section 4.5.3 on cloud height estimations.

### 3.1.3 Cloud height measurements

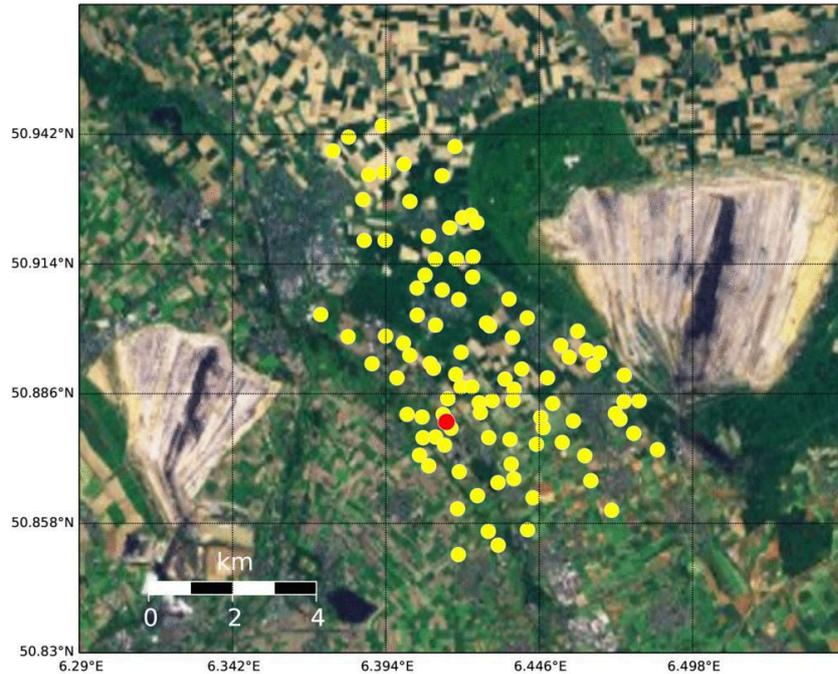
Ceilometer-based cloud height measurements have been provided by the national weather service Deutscher Wetterdienst (DWD) for the station Friesoythe- Altenoythe, located about 30 km south-west of Oldenburg with a temporal resolution of 10 minutes.

### 3.1.4 Forecasts

Image based forecasts have been computed for the period from January 2014 to November 2016. GHI forecasts for the site of the camera have been computed with 1 second resolution and a maximum forecast horizon of 25 minutes. Pyranometer measurements of GHI are used for the validation.

## 3.2 HOPE measurement campaign, Juelich, Germany

The datasets used in this short-term forecast experiment have been collected during the High Definition of Clouds and Precipitation for advancing Climate Prediction (HD(CP)<sup>2</sup>) measurement campaign HOPE in 2013. For the present work, the campaign operator provided data from a network of 99 irradiance sensors, one ceilometer and one sky imager (Fig. 3.3), containing measurements from April 1st to May 31st, 2013. The measurement site is located in Jülich, Germany (50.918798 °N, 6.367952 °E). The area is rather flat and surrounded by two large lignite open-cast minings (Fig. 3.3).



**Figure 3.3:** Experimental setup: Distribution of pyranometers (yellow), sky imager (red) and ceilometer (red) at the measurement site. Map section corresponds to the chosen domain size of 20 km x 20 km. Source: [40].

### 3.2.1 Sky imager

A sky imager developed at the GEOMAR Helmholtz Centre for Ocean Research [45] was used for continuous sky observations. The imager was part of the LACROS supersite within the HOPE measurement campaign, see [21] for the location and details. The digital CCD camera by *Canon* equipped with a fisheye lens by *Raynox* realized a field of view of  $183^\circ$ . The hemispheric sky images with  $2592 \times 1744$  pixels resolution were sampled at a rate of 15 s. In contrast to all other data sets used in this thesis, the images from the sky imager used in HOPE have been recorded and provided by external groups.

### 3.2.2 Pyranometer network

An irradiance measurement network with 99 pyranometer stations was set up around Jülich, Germany on an area of 10 km x 12 km. Each station was equipped with an *EKO ML-020VM* photodiode pyranometer. The 10-bit data logging system was synchronized with the global positioning system (GPS) time. Irradiance was measured with 10 Hz resolution and averaged to 1 Hz. The maintenance as well as cleanliness and tilt control were performed on a weekly basis.

Madhavan et al. [21] give a description of the pyranometer network within the HOPE campaign, details of the hardware and an investigation of measurement uncertainties.

### 3.2.3 Cloud height measurements

Information about cloud base height is retrieved from a *Jenoptik CHM15k-x* ceilometer, that was located next to the sky imager. One measurement was done every 20 s. As a ceilometer provides only point measurements, the median of the last 30 measurements was used in order to smooth the signal and to provide a representative value for the whole scene. Although multi-layer cloud height information is available, only lower level cloud height was used, because the used sky imager algorithm does not yet support multilayer clouds.

### 3.2.4 Forecasts

Forecasts have been computed for all 99 measurement stations with a temporal resolution of 1 second and a maximum forecast horizon of 25 minutes.

## 3.3 1-MW PV plant, Munich, Germany

Another study has been conducted at a 1 MW PV plant on the rooftop of the trade fair building in Munich (48.135968 °N, 11.700247 °E). The objective of the study was to apply the system on a utility-scale PV plant and to validate very short-term forecasts of electrical feed-in power on plant data. The camera setup was designed to be low-cost and autonomous. Measurements of solar irradiance and electrical power are recorded in high-temporal resolution by the plant owner. A summary of the forecast results [18] are presented in this thesis.

### 3.3.1 Autonomous sky imager

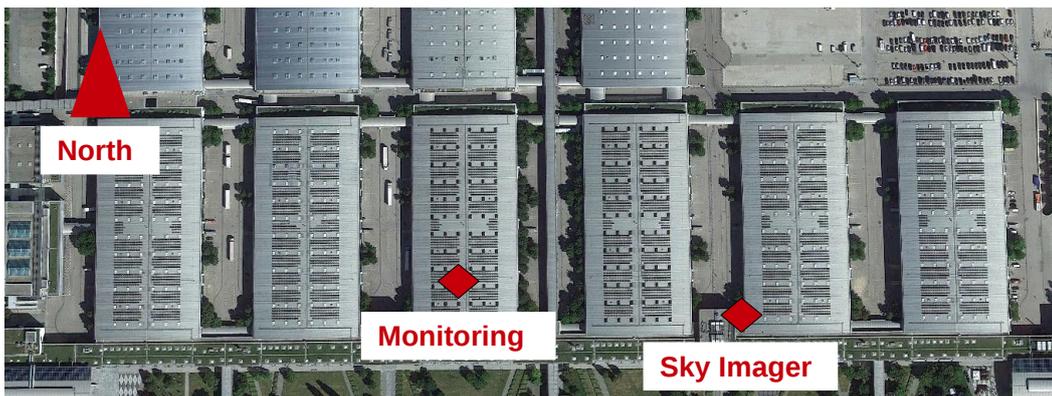
The camera used is a *Vivotek FE8174V*, the follow-up model of the camera used in Oldenburg (Fig. 3.4). It is characterized by a stronger distortion of the lens to its edges leading to a higher pixel-per-meter resolution in the center but higher compression at the edges compared to the former model. The camera was mounted on the rooftop relatively centered in the PV plant. It was tried to preclude objects like buildings in the field of view of the camera. A *Raspberry Pi* was programmed to acquire and archive images in an interval of 10 seconds from sunrise to sunset. Internal time was synchronized with a GPS receiver in the beginning and synchronized with the network time protocol (NTP) after attaching a webstick to achieve remote control. The system was mounted in March 2015 and removed in November 2015. Due to technical problems with image acquisition and memory cards only images from 11th August to 19th November were used for the analyses presented in Sec. 5.8.



**Figure 3.4:** Weather resistant junction box with the Raspberry Pi, the external storage and the power-over-ethernet (POE) adapter used for the experiment (image left). The image in the right shows the camera mounted on the rooftop (PV plant in the background).

### 3.3.2 PV-Plant

The PV plant 'Solardach München-Riem' on the roofs of the trade fair building in Munich (see Fig. 3.5) has an installed capacity of 1016 kW. It covers six roofs of an area of about 600 m x 200 m. The modules are orientated south with a tilt angle of 28 degrees. A monitoring station measures horizontal and plane-of-array global irradiance about 150 m from the camera position. Module temperature is measured and logged at six different modules. The mentioned data will be used for the evaluation of forecasts and to simulate the plants power output. All parameters are sampled with 1 Hz and provided as minutely averages.



**Figure 3.5:** Overview of the PV plant 'Solardach München-Riem', the pre-existing monitoring station and the installation of the camera system. Background source: Google.

### 3.3.3 Forecasts

For a two months dataset of images, forecasts of local global horizontal irradiance have been processed. Forecasts have been averaged to one minute values to account for the spatial extent of the PV plant and the temporal resolution of the measurements. Next, forecasted GHI has been converted to plane of array (POA) irradiance using the Klucher-model [46]. Using predicted POA irradiance and measured module temperature PV power output could be simulated using the model of Beyer et al. [47].

## 3.4 Murdoch University, Perth, Western Australia

The Murdoch University in Perth, Western Australia has been set up two sky imagers (32.068285 °S, 115.833274 °E) in a distance of about 650 m of the type *Vivotek FE8174V* analogue to the sky imager in Oldenburg. Sky images are recorded every 10 seconds since 2016. High-resolution GHI and PV power measurements complete the installation and can be used to calibrate and validate the camera based forecasts. A nearby airport provides cloud base height measurements every 30 minutes. The two camera setup is supposed to be used for the development of a cloud base height triangulation method.

In collaboration with the Photovoltaic Systems group of NEXT ENERGY, EWE-Research Centre for Energy Technology e.V. and the Electrical Engineering group at Murdoch University the usage of sky imager based forecasting for the integration of higher shares of PV in remote electricity grids is under current investigation. Australian locations are supposed to have more clear sky days compared to mid-latitude locations like Oldenburg. A focus is put on the reliability to predict the transition of clear sky periods to cloudy conditions [38], a critical situation for the stability of the electrical grid, if spinning reserves provided by diesel generators are reduced in clear sky periods. Results of a forecast experiment for local site predictions are based on a two month data set of sky images recorded from November to December 2016 and are validated on local pyranometer measurements.

## 4. SkySol - Sky imager based solar irradiance analyses and forecasting model

In this chapter the *sky imager based irradiance analysis and forecasting model - SkySol* is introduced. Details about its core components are given in the following sections. In the end of this chapter, a general study about the spatial coverage and the maximum forecast horizon of the camera based forecast system is presented. The model is under development since the first sky imager has been installed at the University of Oldenburg in September 2013. The model has been applied to a variety of datasets including different sky imagers as well as different locations and climates. Irradiance forecasts, the main product of the model, have been evaluated comprehensively (results presented in Chapter 5).

### 4.1 Overview

*SkySol* has been developed for the purpose of high resolution short-term irradiance forecasts in the field of view of the camera. The heart of the model is the image analysis and forecasting module. Sky images are used as input, then evaluated and processed in different steps to a final forecast product. Multiple configurations allow to apply the model on different data sets, using different optional components and different parameter settings without touching the models internal code. Tools for a comprehensive forecast validation and visualization of results complete the package. The code is written in Python programming language. The model is provided as a Python package built up on several 3rd party packages, i.e. for image processing or data in- and output.

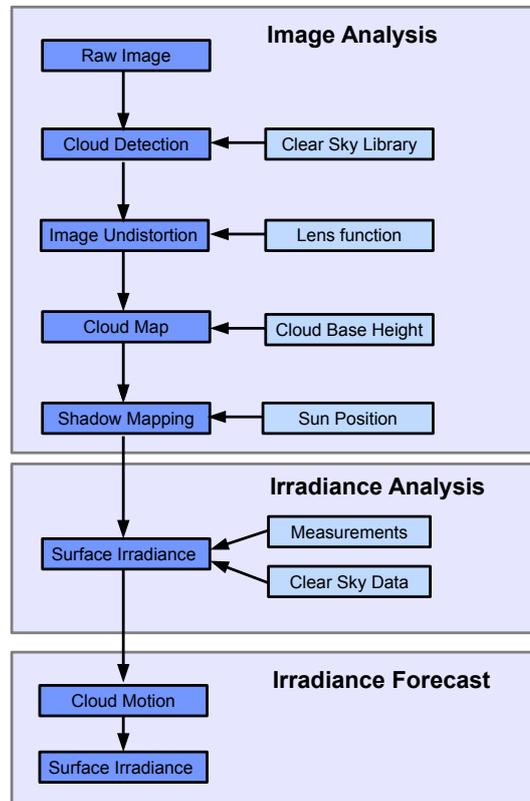
The modular design of *SkySol* allows a comfortable integration of new algorithms, e.g. for cloud detection or cloud motion. The investigation of the algorithms influence on the forecast performance can be analyzed quite easily. This user-friendly design is important for further necessary development of the model and for other researchers to

work with the model and to improve it.

The model has interfaces for the integration of additional required data like online measurements of irradiance and cloud base height. Furthermore, interfaces for solar position algorithms (Section 4.8.3) and clear sky irradiance reference calculations (Section 4.8.4) are implemented. Model products like forecast time series of irradiance are provided in the hierarchical data format 5 (HDF5).

One requirement on the model was its applicability in an operational mode in near-real time. To keep the time between the image acquisition and the provision of a usable product (e.g. irradiance forecast) as short as possible, the computational effort must be reduced while keeping the level of detail and complexity high.

The functionality of *SkySol* can be explained in the order of the processing steps from the raw image to the final product (Fig. 4.1).



**Figure 4.1:** Schematic overview of the principle steps in *SkySol* in image processing, cloud tracking and irradiance modelling for the final forecast product.

Image processing is designed as following: The first step is cloud detection (Sec. 4.3), meaning a segmentation of the image in pixels which are cloudy or not. Shades of cloudiness (transparent, opaque) are possible, but not yet implemented. To determine the spatial distribution of clouds and clouds' shadows, cloud position (distance from camera and height above ground) and sun position relative to the cloud has to be considered. In *SkySol*, a cloud mapping and a cloud shadow mapping scheme is

implemented (Sec. 4.5). The cloud mapping applies the lens function of the camera to undistort or rectify the fisheye image. Clouds' base height is used to derive the 3D position of the cloudy pixel. The cloud shadow mapping computes the position of cloud shadows using sun position and cloud height. Up to now, *SkySol* assumes a single cloud base height for all cloudy pixels in one image. This reduces the complexity of cloud and cloud shadow computation drastically, but loses accuracy in cloud conditions with multiple layer clouds or clouds with large vertical extent. An exception to this procedure is made in case of missing information about the clouds' height above ground. Here, the cloud shadow mapping can be omitted and forecasts can only be processed for the location of the camera. For this specific condition, the time-to-arrival for clouds occluding the sun is predicted.

In a next step, the cloud shadow maps are transformed to surface solar irradiance with the use of local irradiance measurements and clear sky irradiance data. Different implemented and tested methods for the irradiance modelling are introduced in Section 4.6.

To compute forecasts, cloud motion (Sec. 4.4) is determined based on analysing and tracking image features in subsequent images. Several methods exist to approximate either a cloud motion vector for the whole cloud scene or for single pixels or clusters (single clouds). In *SkySol*, a global motion vector method is implemented based on *Optical Flow* technique.

With both cloud detection and cloud motion a simple prediction of further cloud scenes can be computed by extrapolating the current cloud image based on the cloud motion vectors derived beforehand. The surface solar irradiance maps are then used to derive time series for every position of interest in the covered region (Sec. 4.7).

*SkySol* provides additional products like cloud type classification (Sec. 4.8.2), image features (Sec. 4.8.1) and raw data export like forecast time series of red, green and blue pixel intensities in the RGB image, image coordinates and image resolution along the forecast path. This information can be used as input to machine-learning based post-processing of solar irradiance forecasts.

Before a detailed description of the image processing, details about the camera calibration (Sec. 4.2) are presented. Camera calibration has to be performed once in order to find the relationship between image pixel coordinates in the 2D space and the 3D real world coordinate system. It determines the intrinsic and extrinsic camera parameters linked to fisheye lens distortion and camera orientation in the outdoor installation.

## 4.2 Camera model

For very short-term solar irradiance forecasting based on visual sky imaging, digital cameras pointing skywards are used. Equipped with 180°-fisheye lenses, the field of

view covers the whole sky. So called hemispheric images allow a maximum spatial coverage and forecast horizon to a first approximation only limited by the curvature of the Earth (far distant clouds are below horizon) and the ability of the camera to resolve clouds (camera resolution).

Quantitative measurements for cloud tracking or cloud shadow mapping require accurate geometrical camera calibration. The resulting camera model is used for the transformation of 3D real world coordinates to image pixel coordinates in the 2D image plane and vice versa. We call it the forward and backward camera projection model. The camera model described here and used for *SkySol* considers both the fisheye lens projection used by the camera optics and the camera orientation in the real world (site specific). Additional cloud height estimates allow the geolocation of identified clouds and their shadows on the ground. One can imagine, that errors introduced with geometrical calibration leads to errors in positioning and timing of clouds and cloud shadows.

#### 4.2.1 Fisheye lenses

Fisheye lenses are (ultra) wide-angle camera lenses, utilize a heavy barrel distortion to map an infinitely wide object plane into a finite image area. In contrast to pinhole camera models using perspective projection which preserve straight lines in the original scene, fisheye projection results in curved lines to the rim of the image. A huge variety of projections are used by lens manufacturers to accomplish the mapping of the hemispheric or world scene to the image plane. In the later subsection, the method used to determine this mapping is described.

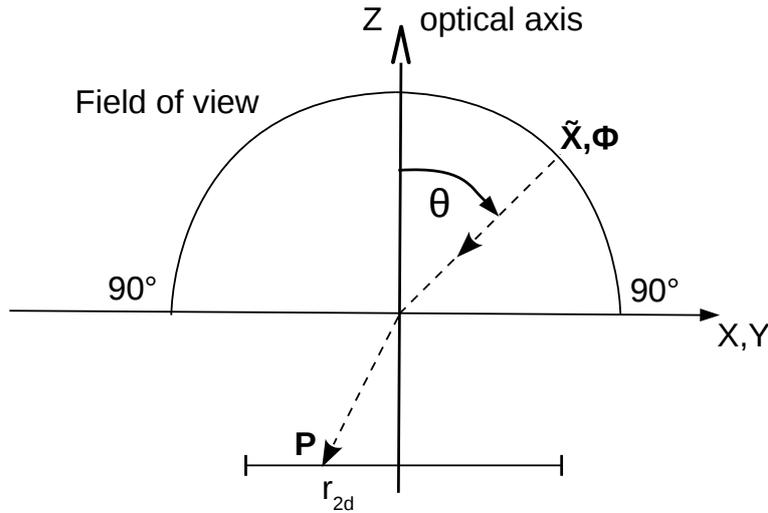
It should be emphasized, that the usage of fisheye lenses has a long history in environmental science. Cloud observations for meteorology (e.g. [48, 49, 50, 51]), space observation for astronomy e.g. [52] and canopy observations in forest sciences e.g. [53, 54] have used hemispheric images from cameras equipped with fisheye lenses.

Usually, circular fisheye lenses are used to project the whole 180°-field of view onto the image plane.

#### 4.2.2 Projection model

In order to project a 3D real world geometric coordinate  $\mathbf{X} = (x, y, z)^T$  to fisheye image pixel  $\mathbf{P} = (u, v)^T$  and vice versa, a two-step conversion with two different parameter sets is used.

First, the fisheye projection function using the intrinsic camera parameters describes the distortion introduced by the optics to map 2D image pixel objects  $\mathbf{P}$  onto a 3D unit sphere. We define a ray emanating from a single viewpoint on the unit sphere



**Figure 4.2:** Illustration of the fisheye projection of a ray emanating from a single viewpoint of the unit sphere onto the image plane. An equidistant projection has been used here for the mapping.

in cartesian coordinates  $\tilde{\mathbf{X}} = (\tilde{x}, \tilde{y}, \tilde{z})^T$  and in polar coordinates  $\Phi = (\phi, \theta, 1)^T$ . The angle  $\theta$  is the zenithal angle between the incoming ray and the optical axis, and  $\phi$  is the azimuthal angle in the horizontal image plane. The projection or mapping of the point  $\tilde{\mathbf{X}}$  to its corresponding image point  $\mathbf{P} = (u, v)^T$  (Eq. 4.1) with radius  $r_{2d}$  to the image center is illustrated in Figure 4.2 in two dimensions.

$$\mathbf{P} = (u, v)^T = D(\Phi) = D(\phi, \theta) \quad \mathbf{P} = (u, v)^T = D(\tilde{\mathbf{X}}) = D(\tilde{x}, \tilde{y}, \tilde{z}) \quad (4.1)$$

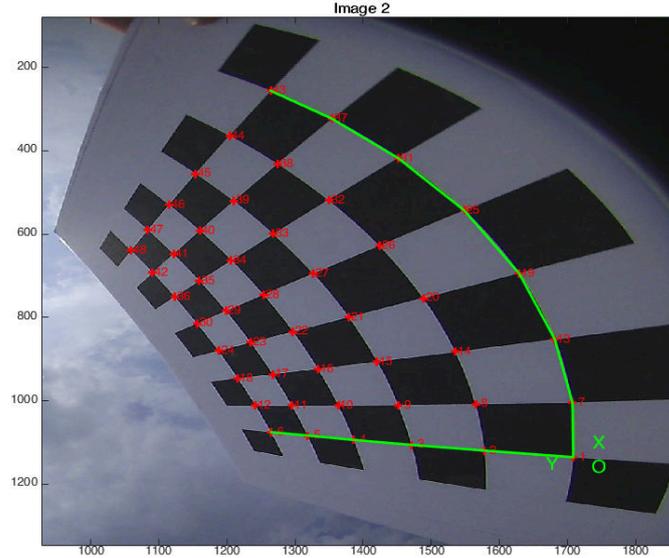
Next, the orientation of the unit sphere in the real world has to be described. The outdoor mounting of the camera and its position relative to the geographical coordinate system is considered and expressed by a rotation matrix  $\mathbf{R}$  accounting for orientation errors (horizontal misleveling and tilt) in all three axes of the image plane (Eq. 4.2).

$$\mathbf{X} = \mathbf{R} * \tilde{\mathbf{X}} \quad (4.2)$$

For quantitative measurements the camera model includes the perspective projection from pixels on the unit sphere to a regular cartesian grid. The perspective projection preserves straight lines and pixels have equal distance. It is therefore used to rectify the image and for the geolocation of image objects. Several image processing steps like cloud (shadow) mapping and forecasting are performed on the perspective projection of the image.

### Fisheye projection and intrinsic parameters

Fisheye lenses are often manufactured to follow a certain projection function. A very common projection is the equidistant projection (Eq. 4.3), which is a linear mapping



**Figure 4.3:** Example image of the *OCamCalib* calibration procedure. Pictures with chessboards at different positions in the image are used to calibrate the lens. The equidistant corners are detected automatically. From the sets of corner coordinates the lens function can be derived.

of the incidence angle  $\theta$  in the interval  $[0, \theta_{max}]$  to the distance  $r$  between the image point and the principal point.  $f$  is the focal length.

$$r = f\theta \text{ (equidistant projection)} \quad (4.3)$$

As real lenses do not exactly follow the designed projection model [55], a more general approach described in the following, has been selected for the lens calibration.

The intrinsic parameters are determined by a calibration of the fisheye lens with *OCamCalib Toolbox* [56]. The *OCamCalib Toolbox* detects straight known lines and edges on photographs of a checkerboard (see Fig. 4.3) for calibrating the parameters needed for the projection model [57]. It is assumed that the image projection function can be described by a Taylor series expansion whose coefficients are estimated by solving a two-step least-squares linear minimization problem. Moreover, a radially symmetric distortion from the center of the distortion (principal point) is assumed. The software determines both the coefficients for the forward projection model  $f(\theta) = r_{2d}$  and the backward projection model  $f(r_{2d}) = z = r_{3d} * \cos(\theta)$  where  $r_{2d}$  is the pixels distance from image center in the image plane and  $r_{3d}$  the radius to the given point on the unit sphere. In our case, a 5th-degree polynomial function with coefficients  $k$  to be estimated is fitted on the detected data points to determine the forward projection model (Eq. 4.4).

$$f(\theta) = r_{2d} = k_0 + k_1\theta + k_2\theta^2 + k_3\theta^3 + k_4\theta^4 + k_5\theta^5 \quad (4.4)$$

The image pixels coordinates  $(u,v)^T$  can be derived from  $r_{2d}$  with Equation 4.5, where  $(u_c,v_c)^T$  is the image distortion center (in most cases not equal to image center that can also be determined by *OcamCalib Toolbox*)

$$\begin{pmatrix} u \\ v \end{pmatrix} = r_{2d} * \begin{pmatrix} \sin\phi \\ \cos\phi \end{pmatrix} + \begin{pmatrix} u_c \\ v_c \end{pmatrix}. \quad (4.5)$$

A 9th-degree polynomial function for the backward projection model (Eq. 4.6) is than computed from the inverse problem

$$r_{3d} * \cos(\theta) = z = f(r_{2d}) = l_0 + l_1 r_{2d} + l_2 r_{2d}^2 + \dots + l_9 r_{2d}^9. \quad (4.6)$$

### Camera orientation and extrinsic parameters

In *SkySol* extrinsic parameters are estimated by a visual comparison of the reprojected computed sun position (azimuth and zenith angle) to image coordinates (forward projection model) and their visual appearance in the image. To simplify this step, we assume a perfectly horizontal mounted camera and define a rotation matrix which rotates the top of the image to geographic north. A probably more accurate method described by several authors uses parameter estimation based on automatic detection of fixed objects with well-known position like sun, stars or moon, e.g. [14, 10, 58]. The quality of this method depends on the automatic detection of these objects. Therefore, exposure settings of the camera and image quality must be appropriate for the detection of sun, stars or moon.

### Forward model

The mapping of 3D real world coordinates on the sphere to image coordinates, e.g. the sun position, is given in Equation 4.7:

$$\begin{pmatrix} u \\ v \end{pmatrix} = \mathbf{R} * f(\theta) * \begin{pmatrix} \sin\phi \\ \cos\phi \end{pmatrix} + \begin{pmatrix} u_c \\ v_c \end{pmatrix}. \quad (4.7)$$

It is the combination of equations 4.1 and 4.2 and includes the extrinsic parameters in the rotation matrix  $\mathbf{R}$ , the intrinsic parameters in the forward projection model  $f(\theta)$  and the image (distortion) center coordinates  $(u_c, v_c)^T$ . The forward model is commonly used to project computed sun position into the image (validating the camera model) and for horizon limits (e.g. only 160° FOV should be considered).

### Backward model

The mapping of image coordinates to 3D real world coordinates, e.g. cloudy pixels for geolocation, can be done by applying the inverse functions.

The projection of a 2D image coordinate  $\mathbf{P} = (u, v)^T$  to its 3D vector  $\mathbf{X} = (x, y, z)^T$  in cartesian and  $\Phi = (\phi, \theta, r_{3d})^T$  in polar coordinates on the unit sphere can be done by applying the inverse projection model  $z = f(r_{2d})$  from Eq. 4.6 with  $r_{2d} = \sqrt{(u')^2 + (v')^2}$  and  $u' = u - u_c, v' = v - v_c$ :

$$\mathbf{X} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathbf{R}^{-1} * \begin{pmatrix} u' \\ v' \\ f(r_{2d}) \end{pmatrix} \quad (4.8)$$

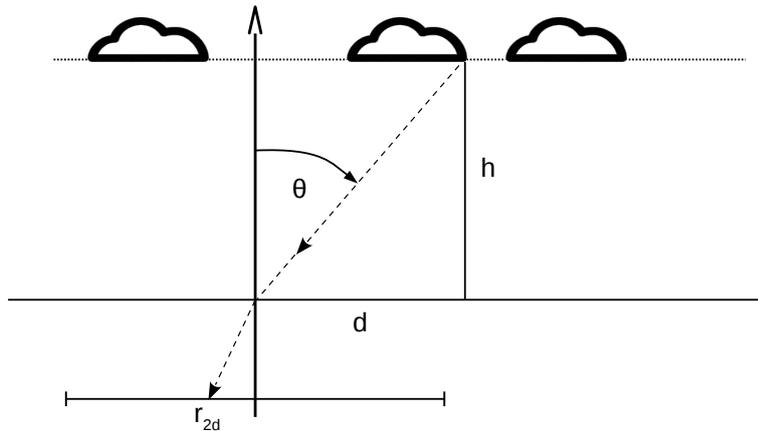
$$\Phi = \begin{pmatrix} \phi \\ \theta \\ r_{3d} \end{pmatrix} = \begin{pmatrix} \arctan \frac{x}{y} \\ \arccos \frac{z}{\sqrt{x^2 + y^2 + z^2}} \\ \sqrt{x^2 + y^2 + z^2} \end{pmatrix}. \quad (4.9)$$

### Perspective projection and cloud height

Up to now, we assume all pixels to be positioned on a unit sphere. In order to project them to a flattened plane (e.g. single atmospheric layer), a perspective projection is applied. Figure 4.4 shows the geometrical relationship between the height of a cloud  $h$ , its viewing angle  $\theta$  and its correspondent distance to the camera  $d$ , which can be expressed with

$$d = h * \tan(\theta). \quad (4.10)$$

This projection scales the distance of a cloud linearly with its height pronouncing the importance of accurate cloud height estimates for precise cloud (shadow) mapping. The projection also maps all pixels to a plane which is defined for a FOV  $< 180^\circ$ . One can imagine, how cloud scenes are stretched at zenithal angles close to the horizon.



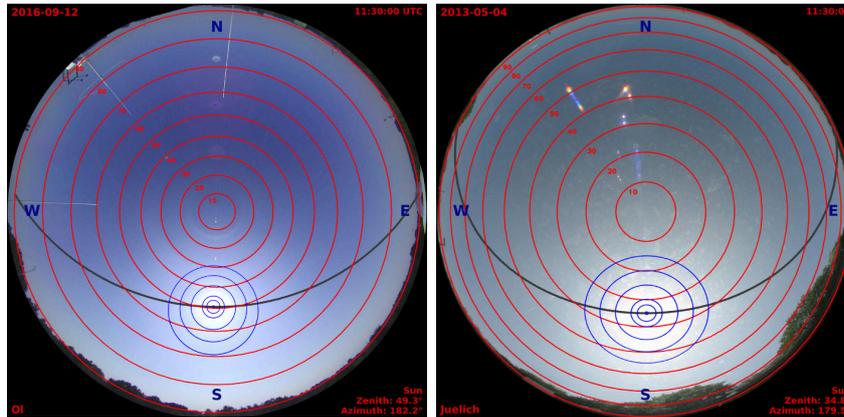
**Figure 4.4:** Illustration of the projection of clouds in a certain height  $h$  seen under a viewing angle  $\theta$  onto the image plane.

With  $\mathbf{X} = (x, y, z)^T$  the vector emanating from a single viewpoint on a unit sphere

in cartesian coordinates (Eq. 4.9), the vector  $\tilde{\mathbf{X}}$  in a geographic cartesian coordinate system can then be computed with the estimated clouds height  $h$  above ground (horizontal plane) with

$$\tilde{\mathbf{X}} = \begin{pmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \end{pmatrix} = h * \begin{pmatrix} \tan(\theta) * \sin(\phi) \\ \tan(\theta) * \cos(\phi) \\ 1 \end{pmatrix}^T. \quad (4.11)$$

Figure 4.5 summarizes the geometrical camera calibration for two example images of the Vivotek FE8172V camera used in Oldenburg and the Canon camera used during the HOPE campaign. It shows the results of the forward projection model on sun position (sun zenith and azimuth angle), isolines of sun angular distances, isolines of zenithal angles and the sun path of the specific day projected onto the image. The rotation to hemispheric north has been applied in order to account for orientation errors from mounting the camera. The different fisheye projections (as well as the colour settings) used in both lenses are apparent.



**Figure 4.5:** Example hemispheric sky images taken with a Vivotek FE8172V camera in Oldenburg (left) and during HOPE campaign (right). The images have been rotated to hemispheric north. The isolines of equal incidence angle derived from the lens function are shown as red circles. The blue lines represent lines of equal angular distance to the sun. The diurnal path of the sun on the specific day is described by the grey line.

### 4.3 Cloud detection

To identify clouds, we apply a binary classification (cloud or sky) of each image pixel. As a consequence, we do not account for varieties in cloud optical thickness (e.g. from thin semi-transparent clouds to thick opaque clouds). In *SkySol*, the concept of the Red-Blue-Ratio (RBR) is used. It was first developed by Scripps Institution of Oceanography [59, 60, 61]. RBR is the ratio between the red colour channel and the blue colour channel of the image (Eq. 4.12). The approach uses the light scattering characteristics in the sky. In sky images, clouds usually appear white because of Mie

scattering of clouds particles with sizes in the order or larger than the wavelengths of visible light. On the other hand, the sky looks blue due to wavelength dependent Rayleigh scattering at small atmospheric molecules. Therefore, the RBR indicates, if the scattered light comes from a cloud (value close to 1) or from the blue sky (value « 1). A threshold of 0.7 (for the camera in Oldenburg) classifies each pixel as cloudy or non-cloudy. The threshold has been determined by a visual comparison on a test data set of 40 images covering different sky conditions. Due to the varying colour settings used for different cameras the threshold has to be adapted for each camera. Figure 4.6 shows examples of the proposed cloud detection approach.

$$RBR = \frac{\text{Red Colour}}{\text{Blue Colour}} \quad (4.12)$$

#### Error sources when using a RBR-threshold for cloud detection

Two main problems remain when using RBR-thresholding. First, the RBR is not homogeneously distributed over the whole field of view for the same sky conditions. RBR has an angular dependency [62] and the area close to the sun (circumsolar region) is affected by the bright sun (i.e. high pixel intensity results in a  $RBR \approx 1$ ). Consequently, misclassifications are likely when a single global threshold is applied to the image (Fig. 4.6i). Secondly, RBR is not always suitable for discriminating clouds from sky. Optically dense clouds for example appear quite dark in the center of their base [4]. Here, the RBR is very low and clouds can be misclassified as sky. The example in Figure 4.6iii shows this phenomena. Thin clouds like cirrus but also fog, haze, mist or dust often have RBR values close to the threshold. In this case, the single threshold is a bad discriminator for these phenomena. An example is given in Figure 4.6iv illustrating the RBR distribution around the threshold. The turquoise and grey colours in the binary cloud image (subfigure (h)) mark RBR values bigger and less than 0.02 from the threshold. As a result, the RBR threshold is very sensitive to these clouds. The observed impact of these semitransparent clouds (or atmospheric phenomena in general) on GHI is relatively small. However, the scattering of direct sunlight leads to a reduction in DNI that is accompanied with enhanced DHI measurements. The classification as no cloud might therefore result in better forecast performance compared to the GHI reduction linked to opaque clouds. Up to now, the model does not handle semitransparent clouds specifically and they remain an error source of the binary discrimination of clouds and sky.

#### Clear Sky Library (CSL)

To address the problems associated with the inhomogeneous distributed brightness and RBR distribution, clear sky reference images are used to correct the current RBR image. As the brightness and RBR distribution of the sky background in the image

strongly depends on the position of the sun in the image, a clear sky library (CSL) for different sun positions is build up of clear sky images in the past. This approach has been proposed first by Shields et al. [63] and been applied in a forecasting study by [5]. In *SkySol* a similar CSL is build up from clear sky images. Images are stored in a database with its sun position (zenith and azimuth angle with  $0.1^\circ$  resolution) and its timestamp. The timestamp labels updates in the database. Images in the database are updated if the apparent cloud cover (percentage of cloudy pixels) is less than 10%. The update is necessary, because not only sun position, but also seasonal and day-to-day variations in atmospheric aerosol composition influence atmospheric scattering strongly and therefore the overall RBR distribution. Obviously, appropriate CSL images should not be too old in order to represent current atmospheric conditions as best as possible [7].

### RBR-CSL method

The approach uses a correction of the directly computed RBR image (computed from the raw RGB image) based on the CSL image, the pixels intensity and the grade of saturation in the circumsolar area.

The modified RBR ( $R_{mod}$ ) is given for each pixel at the image position  $i,j$  by the following equation:

$$R_{mod,i,j} = R_{orig,i,j} - R_{CSL,i,j} * (a * S - b * (I_{i,j} - 200)). \quad (4.13)$$

Eq. 4.13 first accounts for the difficult circumsolar area. Weighted by the grade of saturation ( $S \in [0, 1]$ ) in the disc up to an angular distance of 5 degrees to the center of the sunspot, we subtract the clear sky RBR ( $R_{CSL}$ ) from the original RBR ( $R_{orig}$ ). Moreover, a correction based on the pixel intensity  $I^1$  and clear sky RBR  $R_{CSL}$  is applied, which increases RBR in case of dark clouds/pixels and decreases RBR in case of bright clouds/pixels.

The coefficients a and b as well as the global RBR threshold were determined empirically on a test dataset of several images with different sky conditions. Note that the used CSL introduces errors on days where solar zenith and azimuth angles deviate from the reference day.

The proposed approach aims to reduce the mentioned misclassifications in the circumsolar area and in case of thick and dark clouds. Figure 4.6 summarizes the proposed method for four different cloud conditions. When comparing raw RBR images to the corrected versions, a clear reduction in high RBR values around the sun is visible (Fig. 4.6i, 4.6ii, 4.6iv), while in case of dark and thick clouds (Fig. 4.6iii) misclassifications can be avoided by increasing RBR values. Moreover, the RBR histograms of the corrected images have a clearer bimodal shape and therefore a better discrimination

---

<sup>1</sup>  $I = 0.299 * Red + 0.587 * Green + 0.114 * Blue$

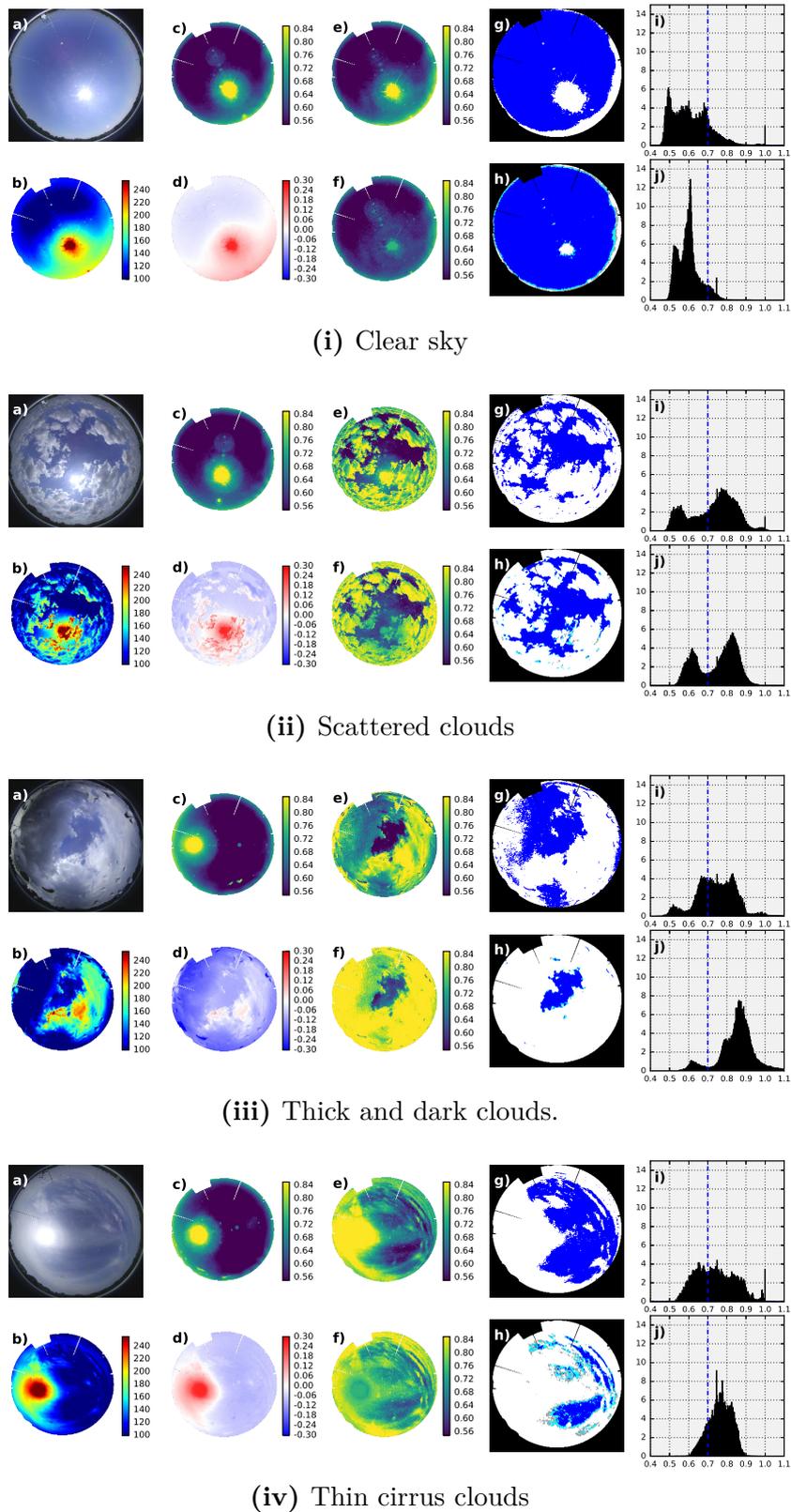
between high (clouds) and low (sky) RBR values. Nevertheless, the cirrus conditions in Figure 4.6iv are still badly interpreted based on the binary discrimination.

### 4.3.1 Postprocessing

A pixel-based thresholding can lead to noisy binary images if pixel RBR values are close to the chosen threshold or clear sky reference images are not perfectly clear. Regarding cloud shadow projection and solar irradiance modelling based on the binary image, a pixel-wise modelling can result in overestimations of variability due to noisiness in the binary image. As the opening angle of the sunspot is not infinitely low (about  $0.5^\circ$ ), very small clouds will not affect surface solar irradiance to a significant amount and can be neglected. A median filter is applied to the binary cloud map (in fisheye space) before the cloud shadow projection. The filters' kernel size has been set to 11 pixels, which is determined empirically but has to be adapted if image size for different cameras changes.

### 4.3.2 Alternative methods

Cloud detection based on RBR was used in several sky imager based forecast applications e.g. [5, 12, 6]. Some researchers used approaches with modelling the clear sky reference instead of using a look up table from historical images [64, 65]. Instead of using the red to blue ratio RBR, Heinle et al. [29] used the red to blue difference. Mantelli Neto et al. [66] used multivariate colour space features to classify clouds and sky pixels by means of a pattern statistical characterization using the Euclidean geometric distance. Li et al. [67] proposes the hybrid thresholding algorithm (HYTA) as a modification of the RBR in the  $(R-B)/(R+B)$  colour space. More approaches using combinations of RGB and HSV (Hue, Saturation, Value) colour channels and machine learning approaches are proposed by Kazantzidis et al. [50], Yang et al. [68], Ghonima et al. [7], Jayadevan et al. [69], Su et al. [70]. Some of the listed methods also include non-binary pixel classification with introducing new classes like thin and thick clouds. The question of which of the proposed methods is the most appropriate, has not been identified yet in a comprehensive methodology comparison study. Nevertheless, identifying clouds is just a part of the larger objective of surface solar irradiance modelling as a result of cloud-irradiance interaction. Therefore, a next step will be to introduce optical properties like transmissivity in the cloud detection model to better estimate resulting surface direct and diffuse components of global irradiance.



**Figure 4.6:** Cloud detection results for four different sky conditions. a) Original image. b) Pixel intensity measured in the range  $[0,255]$ . c) RBR clear sky reference. d) Absolute RBR correction. e) RBR without correction. f) RBR with correction. g) Binary cloud map without correction. h) Binary cloud map with correction. i) RBR histogram for g). j) RBR histogram for h). The RBR threshold is marked in the histogram with a vertical dashed line.

## 4.4 Cloud motion

The fundamental information needed for cloud forecasts are cloud movement and cloud transformation. As the transformation (development and dissolution) of clouds is a very complex task, the proposed algorithm does not account for that yet. As a consequence, predicted cloud scenes are the result of a translation of the current analysed cloud scene assuming persistence of cloud speed, size and shape. Different methods have been developed in recent years to estimate the cloud motion field. Chauvin et al. [71] give an overview of the techniques used and compares their advantages and disadvantages. While most methods estimate a mean cloud motion vector (CMV) for the whole sky, dealing with multiple layer clouds or clouds moving in different directions is still a challenging task. In contrast to other applications like human tracking, clouds differentiation is difficult because of their subtle texture and their tendency to overlap [72]. Nevertheless, Xu et al. [72] published a method that segments clouds before tracking. For single-layer cloud tracking, techniques like cross-correlation [5], particle image velocity [73], non-rigid transformation [74] and variations of optical flow [75, 4, 76] have been used. The selection for a specific method depends on the trade-off between computational efficiency and accuracy. For *SkySol* the optical flow technique has been chosen for cloud tracking by reason of simplicity, processing time and accuracy. To reduce complexity, up to now, also mean cloud motion estimation is used.

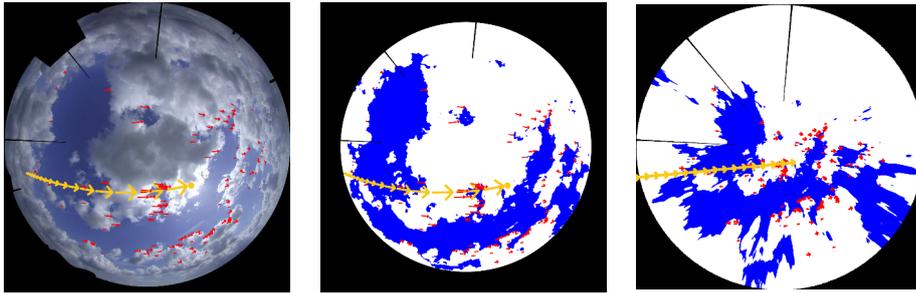
### 4.4.1 Optical Flow

Cloud motion is determined by applying the optical flow algorithm available in OpenCV (Open Source Computer Vision Library<sup>1</sup>). While dense optical flow estimates CMV for each image pixel, sparse optical flow based on a dramatically reduced number of specific salient pixels is used in *SkySol* to reduce processing time. In sparse optical flow, the first step is to determine *good features to track* in the image [77]. These objects - mostly found on strong gradients like cloud edges - serve as input for the Lucas-Kanade tracking algorithm [78, 79]. The algorithm yields cloud motion vectors for each pixel determined as an appropriate feature to track.

In this study, new features are determined every two minutes because old features do change too much or move out of the field of view. The algorithm is applied to the original grey colour image, where artificial objects are masked out. Each single CMV is rectified (from fisheye space to perspective euclidean space) to compute the average vector. In a homogeneous flow, CMVs should have equal length and direction. An example showing the transformation of a cloud scene with marked CMVs from the circular fisheye image to the perspective projection (with a mapping to the defined underlying grid) is shown in Fig. 4.7.

---

<sup>1</sup> <http://opencv.org>



**Figure 4.7:** Example of cloud motion analysis with the optical flow technique. Derived cloud motion vectors (CMV) are drawn in each image. The forecast path resulting from the global motion vector and heading to the position of the sun is drawn as a black line. The rectified cloud map with the CMVs also rectified is shown in the right. The number of detected vectors is reduced for this visualization.

### Image features to track

The fundamental assumption behind optical flow is that a pixel value does not change over consecutive frames, but only shift position. This assumption is often violated in sky images due to changing illumination, occlusion, non-Lambertian reflectance, etc. For example, clouds in proximity to the sun are brighter than in other locations due to forward scattering and this leads to significant deviations from the brightness constancy assumption [76]. Consequently, the Shi-Tomasi algorithm [77], which determines suitable pixels to track, is applied after masking the circumsolar area, the image horizon and artificial objects. A suitable pixel is often found at cloud edges providing traceable textural properties. The maximum number of tracked pixels is limited to 750. The features are detected on the original greyscale image. The interval for feature updates should be not more than a few images. Otherwise, tracks can be lost or objects can move out of the field of view.

### Feature tracking

Subsequent images are used to track features detected in previous images [78]. The image acquisition interval has been chosen to be not higher than 10 s for all Vivotek cameras used in this work. This will keep the cloud displacement from one image to the next small. Moreover, it can be expected that the brightness constancy assumption described in the previous chapter is less violated compared to longer intervals between two images. The feature tracking algorithm determines the resulting position of a specific feature in the next image. As a result a cloud motion vector can be determined. If the algorithm is not able to find a corresponding pixel, it returns a flag marking the feature as not trackable.

## Rectification

Cloud motion vectors derived on original distorted images must be rectified or undistorted by applying the backward camera model to the vectors initial and terminal pixel. The backward camera model rectifies the image leading to vectors stretched at image edges and compressed in the central part analog to cloud projection. In addition to cloud direction, cloud speed in metrical units can be computed when cloud height estimates are available. Again analog to cloud projection, CBH acts like a linear scaling factor leading to higher cloud speeds for higher CBH. If no CBH is available, cloud speed can only be expressed in pixels per second and assumes all clouds scenes in a unit height.

## Quality control

Quality control must be ensured to sort out bad CMVs before averaging to a global vector. Initial CMVs are flagged as invalid, if their speed is less than  $0.2\text{ms}^{-1}$  (or 1 pixel per second if no mapping to a metric grid is applied). This avoids tracking artificial objects or noise in the image. If clouds are moving at a speed slower than the threshold and all CMVs are flagged as invalid, a persistent cloud mask is assumed. For follow-up vectors, sudden changes in direction and speed (changes in cloud speed  $> 2\text{ms}^{-1}$  or  $> 5$  pixels per second when no cloud height is available), which can occur if brightness in the image changes rapidly, the vectors are also sorted out.

## Mean cloud motion

The final rectified CMVs are averaged to one global vector which determines the principal movement of the cloud scene for the forecast. In Figure 4.7 the mean cloud motion vector heading to the position of the sun indicating the forecast path for clouds occluding the sun is drawn as a black line. In order to stabilize the global vector over time, the last four global vectors are also averaged in time. This is justified by the fact, that real changes in cloud motion are rather inert. Furthermore, each change of the average CMV will affect the forecasted cloud distribution and the irradiance forecast. An approach that uses the uncertainty in cloud motion for an estimation of uncertainty irradiance forecasts is in progress.

## No cloud motion / clear sky

In case of clear sky no CMVs can be derived. Moreover, if the number of good quality CMVs is too low ( $<10$ ), no mean cloud motion is computed. In this case, persistence of the current cloud scene is assumed (no cloud motion detected  $\rightarrow$  no cloud motion present). For the irradiance forecast, a persistent clear sky index is predicted for the

whole specified forecast horizon. Note, that this persistence assumption might be violated and lead to erroneous forecasts, if cloud motion is present in regions out of the camera's field of view. Clouds, which "enter" the field of view, might shade the location of interest in a much shorter time than expected. Therefore, using cloud motion information from satellite images or from past CMVs might be a more appropriate approach. See section 4.9 for a detailed discussion on variable maximum forecast horizons.

## 4.5 Cloud shadow mapping

### 4.5.1 Cloud mapping

The determination of the 3D position of a cloudy pixel in a real world coordinate system orientated to hemispheric north with origin at the camera position is computed based on the backward camera projection model described in section 4.2. The backward camera model transforms 2D image coordinates to 3D spherical coordinates on a unit sphere. The pixels incidence angle  $\theta_{i,j}$  ranges from zenith to horizon ( $90^\circ$ ). The azimuth angle  $\phi$  of the cloud around the horizon, ranges from hemispheric north and increasing to east according to meteorological definitions. The third coordinate, the distance  $r$  of the cloud to the camera, can only be determined with the knowledge about its height above ground, approximated by its cloud base height (CBH)  $h$ .

From geometry, the distance on the ground  $d$  from the position of the camera can be computed with

$$d = h * \tan(\theta) \quad (4.14)$$

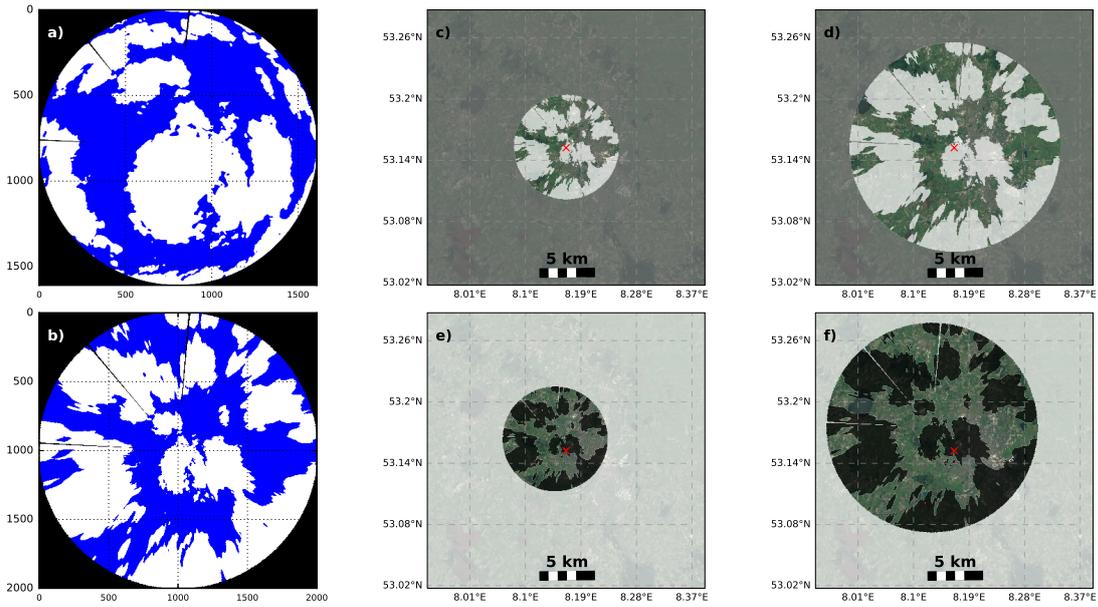
and units according to the unit of the given CBH  $h$ .

If described in cartesian coordinates, the metric coordinates of a cloudy pixel  $\mathbf{X} = (x,y,z)^T$  are

$$\mathbf{X} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} d * \sin(\phi) \\ d * \cos(\phi) \\ h, \end{pmatrix} \quad (4.15)$$

where  $x$ ,  $y$  and  $z$  are the north-south and east-west coordinates with respect to its origin, the camera location according to Eq. 4.11 from the camera calibration Section 4.2.

Obviously, if only one CBH is used for the whole cloud scene, the CBH is scaling the cloud map radially and linearly. Figure 4.8 illustrates the effect if one compares subfigures c) and d), where a CBH of 1500 m and 2000 m respectively are considered. A unit CBH can be used if no information about cloud height is available (shown in



**Figure 4.8:** Illustration of cloud and shadow projection for a mixed cloud scene with a solar zenith angle of  $60^\circ$  and azimuth angle of  $135^\circ$ . The field of view is limited to  $80^\circ$  viewing angle. a) binary cloud image in fisheye space oriented north. b) cloud projection without CBH information on a regular grid with  $2000 \times 2000$  pixels. c) cloud projection for CBH  $h=1500$  m with grid resolution of 15 m and grid size of 30 km in both directions. d) cloud projection for CBH  $h=2000$  m. e) shadow projection for CBH  $h=1500$  m. f) shadow projection for CBH  $h=2000$  m. The red cross indicates the camera location.

subfigure b). In this case, the cloud scene is only transformed to perspective projection. Moreover, if a single CBH is used, the cloud map can be interpreted as a 2D cloud layer. In *SkySol* image pixels are mapped to a regular cartesian grid with its origin at the camera position and the axis heading north and east, respectively. An appropriate inverse mapping function (Eq. 4.16) based on the backward camera projection model is defined for a computational efficient mapping. Inverse mapping is the process of determining the location of a target pixel in the original image.

$$g(x,y) = f(h(x,y)) \quad (4.16)$$

where  $g()$  provides the remapped image,  $f()$  the source image and  $h(x,y)$  is the mapping function that operates on image pixel coordinates  $(x,y)$ .

The scaling of the image pixels with the tangens of its incidence angle  $\theta$  (Eq. 4.14) leads to the non-equidistant projection of cloudy pixels with its distance to image center. As mentioned before in the camera calibration section, pixels to the image border have a larger pixel by pixel distance than pixels in the center of the image. As a consequence, if cloud maps are evaluated on a regular equidistant 2D grid, pixels in the center of the image are averaged while pixels to the border have to be stretched. The image radius for which a single pixel is mapped to a single grid cell (one-to-one

pixel correspondence) depends on the grid size used. Nearest neighbour interpolation is used for non-integer pixel mapping to reduce processing time. In *SkySol* the grid size is user-specified. Subfigure b) in Fig. 4.8 gives an impression of the stretching and compression of pixels in the center and the outer parts of the image. If the grid size has been chosen too low compared to raw size of the image, some information in the center of the image is lost due to compression. On the other hand, with larger grid sizes one can get rid of this disadvantage, but computational costs increase with grid size. If CBH information is available (subfigures c) and d)), additional information about the grid resolution defined in meters is necessary and a fixed size grid will be defined. As a result, cloud maps can be smaller or larger than the chosen grid leading to either unused information or information losses depending on the current CBH and the corresponding scaling of the cloud map.

### 4.5.2 Shadow mapping

Cloud shadow projection on the ground is very similar to cloud mapping. As long as a single CBH is used, cloud shadows are just a displacement of the cloud map depending on CBH and sun position. For the area covered by the sky image, the error introduced by using a single sun position for the location of the camera can be neglected. Subfigure e) and f) in Fig. 4.8 show the displacement of cloud shadows to north-west as the sun is low in a south-east position.

In *SkySol*, the current sun position (azimuth angle  $\phi_{sun}$  and zenith angle  $\theta_{sun}$ ) is computed with the SG2 algorithm [80]. Together with CBH  $h$  a sun ray tracing is applied to map the cloud layer as a shadow layer on the ground. Equation 4.17 gives the basic formula for the computation of the horizontal distance  $d$  from the camera of a cloud's shadow on the ground.

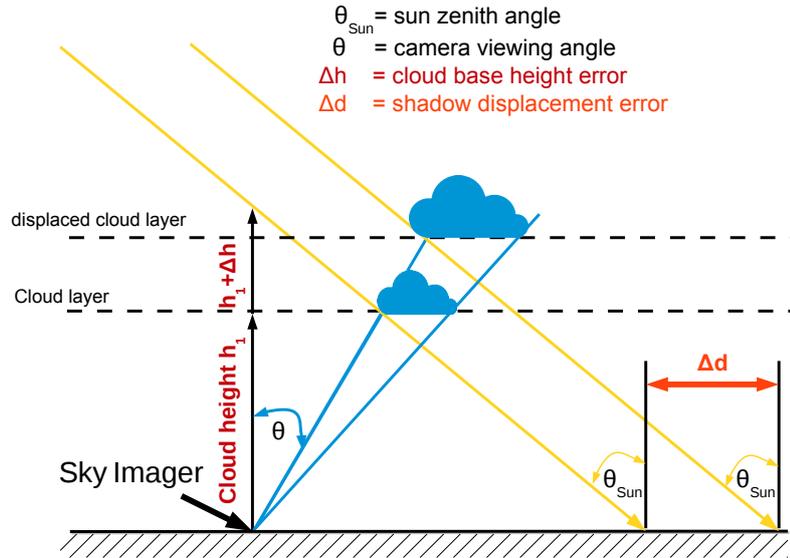
$$\begin{aligned} dx &= h * \tan(\theta) * \sin(\phi) - h * \tan(\theta_{sun}) * \sin(\phi_{sun}) \\ dy &= -h * \tan(\theta) * \cos(\phi) + h * \tan(\theta_{sun}) * \cos(\phi_{sun}) \\ d &= \sqrt{dx^2 + dy^2} \end{aligned} \tag{4.17}$$

Up to now, a topographic flat surface is assumed. Using a realistic topography should be used in hilly areas. For this thesis, the model has only been used in almost flat terrains.

### 4.5.3 Cloud height estimations

As mentioned in the previous section, cloud base height (CBH) estimations are a necessary parameter for cloud and cloud shadow mapping. Uncertainties in CBH estimations lead to projection errors and therefore to wrong positions of clouds and

cloud shadows. Naturally, the error in cloud positioning is enhanced for clouds closer to the image border. The error in cloud shadow positioning depends beyond that on the sun position. For typical cloud heights ranging from 50 m to 10000 m height above surface, uncertainties of several hundred meters are likely for complex sky conditions. Regarding the aimed forecast resolution of a few meters, these considerations underline the need for accurate CBH estimations. Figure 4.9 illustrates the cloud (shadow) positioning error for a vertically displaced cloud seen under the same viewing angle.



**Figure 4.9:** Simplified illustration of a vertically displaced cloud seen under the same viewing angle and its shadow positions on the ground. Note: For an exact calculation of the location error, the azimuthal position of the pixel and of the sun have to be considered.

In meteorology, the reference instrument to measure CBH is the ceilometer. Ceilometers are recognized by the world meteorological organization (WMO) as the most accurate, reliable and efficient means of measuring cloud base from the ground when compared with alternative equipment [81]. Ceilometers are laser instruments pointing vertically and deriving vertical profiles of atmospheric backscatter intensity of which CBH can be computed with a sub-minute resolution. High prizes make ceilometers inappropriate to be installed at each sky imager site. Additionally, in most regions of the world, the density of the meteorological stations equipped with a ceilometer is too low to account for the small-scale differences of clouds and also its heights above ground. Usually, meteorological observatories and airports are equipped with ceilometers and report CBH as 30 minute averages in so called METAR reports.

CBH information derived from numerical weather models (e.g. operational weather forecasting models) and satellite images (cloud top height) have the advantage to be available for large regions (model domain or coverage of the satellite image respectively) and for reasonable costs if access to the data sources is guaranteed. Nevertheless, spatial and temporal resolution is fairly limited for both data sources providing only a

first approximation but is not sufficient for resolving small scale cloud structures.

Due to these limitations and the strong dependency on CBH information great efforts to develop cost-efficient local methods to derive CBH estimations onsite with different uncertainties have been made.

To give an overview about different techniques, a categorization on following aspects should be made:

- hardware requirements
- installation and data processing costs
- accuracy

#### Triangulation / Stereography

The most popular technique is triangulation of clouds with multiple cameras installed at the site of interest. Stereography is used to match cloud images of at least two cameras installed in specific distance. The method has the advantage to gain CBH estimates with a high temporal and spatial resolution with almost no delay in time. Several researchers developed and improved methodologies to derive CBH from multiple camera images [8, 14, 82, 83, 84, 85, 86, 87]. All approaches have in common, that at least two cameras with a distinct distance between each other have to be installed in order to have a best matching of a single cloud scene in both cameras. Depending on CBH, this optimal distance varies [8]. For operational usage, e.g. for a solar power plant, the optimal location regarding the frame conditions of the location has to be considered. A big advantage is the possibility to derive dense maps of CBH for single clouds. The estimations are always up to date as no past information has to be considered. The effort for the synchronization of both cameras and the configuration of the stereographic method is high.

#### Cross-correlation of image derived and sensor-based irradiance time series

Another option is to cross-correlate time series derived from the analyses of sky images for a specific location and varying CBH candidates with measured time series. For this idea one want to find the optimal CBH that produces the best matching time series with reference time series from ground based pyranometer or PV power measurements. The approach requires measurements in some distance to the camera to enable the generation of different time series from sky images for this location. The optimal location for such a sensor depends on the sensitivity of the irradiance analyses on CBH uncertainties and the introduction of errors by cloud detection and projection for distant locations.

### Cross-correlation of image derived and sensor-based cloud motion vectors

The idea of cross-correlating different data sources to obtain the cloud base height can also be extended to cloud motion. Assuming that cloud speed in metrical units derived from sky images depends on the height of the clouds detected in the image one can match the results with a ground based estimation of the clouds shadow speed. Fung et al. [88] proposed a cloud shadow speed sensor using several single photodiode sensors to estimate the cloud shadow speed and direction from the timing differences of signals at each sensor [89, 90]. Based on subsequent sky images, one can also estimate the speed and direction from the same cloud in pixels per time. As mentioned above, the conversion of clouds' speed to metrical units depends linearly on its height above ground. Cross-correlation can be used to determine the cloud height that matches best with the ground-based measured cloud shadow speed. Here, the accuracy of the method strongly depends on the accuracy of both the shadow speed measurement, the cloud motion estimation in the sky image and the matching of detected clouds to detected surface shadows [9]. A masters' thesis on the investigation of using a triplet of photodiode pyranometers in a distance of about 14 m to estimate shadow motion for the matching with sky image based cloud motion estimation has been conducted in Oldenburg in the year 2015 [44].

Due to the modular design of *SkySol* various CBH sources can be implemented. A benchmarking of the single approaches hasn't been performed yet. For this thesis, CBH estimations originate from ceilometer measurements alone.

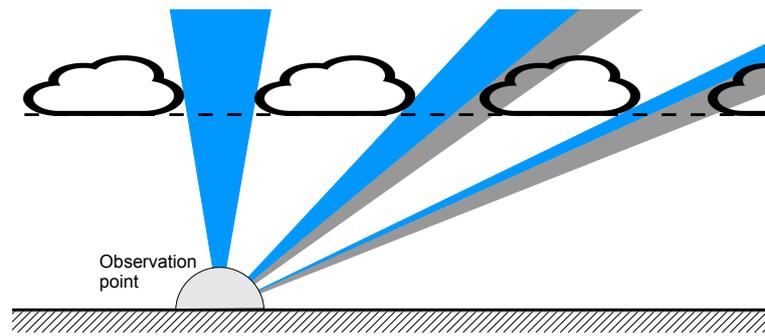
#### 4.5.4 Limitations in ground-based sky imaging

Regarding the main objective of cloud recognition and mapping to 3D real world coordinates from sky images, it has to be considered that in all times only a part of the sky can be observed by only a single ground-based camera. Every cloud or part of a cloud that is behind the optical path of a visible cloud can not be seen in the image due to the passive light detection of the camera. Apart from that and apart from the advantage of having the whole sky view in a fisheye image, the fact that watching clouds under a large viewing angle together with limited resolution of the lens introduces three negative consequences which are

- perspective errors increasing to the image borders
- image resolution decreasing to the image borders
- cloud mislocation mainly due to uncertainties in cloud base height leads to cloud shadow mislocation increasing to the image borders

Summarized, whole-sky imaging introduces errors in cloud and shadow projection with increasing distance to the image center.

Perspective errors are of different nature. First, while clouds in the center of the image



**Figure 4.10:** Illustration of the perspective error introduced when an observer on the ground looks to clouds and cloud gaps under large viewing angles. Depending on the vertical extent of the clouds, the perspective error increases with larger viewing angles while image resolution decreases. The sky that is seen as free of clouds is coloured in blue, while the part of the cloud gap that is covered by the side of the cloud in the back is coloured in grey.

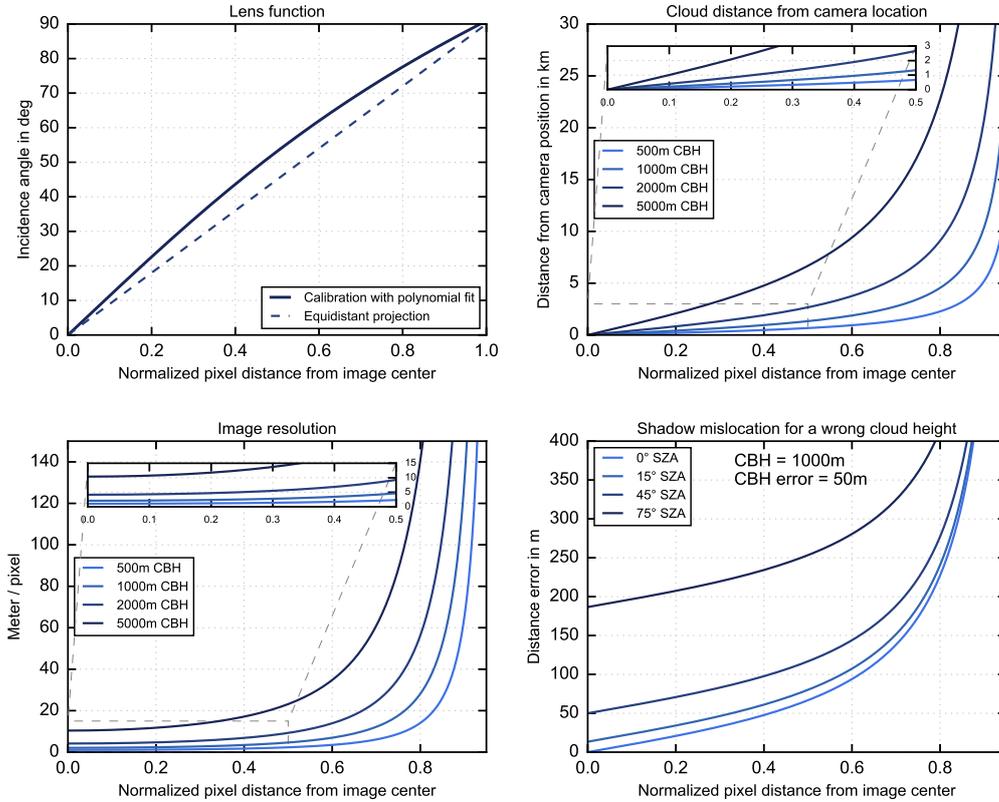
are definitely seen from its bottom, clouds to the border of the image can be seen from its side and it is hard to analyse if the cloudy pixels refer to the bottom or the side of a cloud. From a meteorological point of view, single convective clouds (cumulus type) introduces the largest errors, while flat stratus-type clouds (e.g. stratocumulus) lead to lower errors if clouds are regarded as flat in a single layer. This phenomena leads to overestimations of cloud cover at the image borders as cloud gaps can be covered by cloud sides in a larger distance.

This is even more pronounced when image resolution decreases to the image borders. The reduced image resolution to the image borders has also the effect, that smaller clouds or small cloud gaps can not be resolved anymore and the detected cloud field gets smoothed. This is of importance for variability in irradiance time series derived from the shadow maps.

Both effects (perspective error and reduced resolution) are illustrated in Figure 4.10. In this example, a constant cloud distribution with equal-sized clouds and equal distances between clouds is assumed. When viewing from a single point from below (like an observer does), clouds in a larger distance to the observer appear smaller resulting in a smaller viewing angle (reduced resolution). Imagine what appears to the observer when looking to the cloud gaps. Depending on the clouds' vertical height, the part of the cloud gap that appears as a cloud (actually the side of the cloud in the back) increases with viewing angle. The viewing angle under which cloud gaps disappear totally can be computed, if clouds' height, the gap size and the vertical extension are known.

The latter effect of cloud (shadow) mislocation can be derived from simple geometry. The farther the cloud is from the camera location, a wrong cloud height will at first lead to an incorrect cloud position. The error of the shadow projection varies depending on sun's zenith angle (Fig. 4.9). Depending on the cloud pixels azimuthal position the shadow positioning error can be enhanced or even compensated compared to the cloud

positioning error.



**Figure 4.11:** Results of the camera calibration showing the lens function (a) and corresponding cloud projection results for clouds in different heights (b), the image resolution (c) as the derivative of (b) and shadow projection errors given different sun zenith angles and for pixels in line with the sun azimuth angle (d) as a function of the pixels' normalized distance to the image center.

Figure 4.11 illustrates some of these effects based on the geometrical camera calibration of the Vivotek FE8172V camera used in Oldenburg. It shows the lens function or projection model (top left), the area covered by the sky image depending on different cloud heights (top right), the image resolution computed from the pixel-by-pixel distance for different cloud heights (bottom left) and the shadow mislocation for different sun zenith angles if a wrong cloud height is used (bottom right). The figure points out the strong limitation of pixels at the image border. As clouds distance and therefore the covered area is scaled with tangens function, clouds distance increases and resolution decreases fast for large viewing angles. Therefore, calculation of the image resolution can be used to determine, which field of view (FOV) one wants to use to suppress this effect. For example, for an incidence angle of about  $80^\circ$ , which corresponds to about 0.8 normalized pixel distance to the image center (1.0 is equal to the image radius up to the horizon), the image resolution for clouds of 500 m height is still below 20 m per pixel, while for clouds above 2000 m, resolution is strongly decreasing from 60 m per pixel to more than 140 m for clouds in 5000 m height. To obtain a high consistency in image resolution when projecting pixels to real world, one can consider discarding

outer parts above  $60^\circ$  incidence angle. On the other hand, in this case a large part of the overall forecast horizon will be removed and clouds entering the field of view can be detected much later. This leads to the question if fisheye lenses with  $180^\circ$  FOV are the best choice for sky imaging? Alternatives like cameras with multiple sensors and images with lower opening angles stitched together or multiple cameras at different locations also merging its images could lead to improved results. Figure 4.11d shows the effect of a wrong cloud base height (here: 50 m) on the shadow projection. The results are valid for pixels in line with the sun's azimuth angle. The graph highlights, that a small CBH error is amplified by the projection to more than 300 m for clouds with a CBH=1000 m at the border of the image. If we are interested in spatial and temporal resolutions of meters and seconds, this is already quite a large number.

## 4.6 Irradiance modelling

The fundamental transformation of cloud shadow maps to irradiance maps is simplified in *SkySol* to the transformation of the binary shadow map (shadow and no shadow) to a binary irradiance map. The error introduced by the binary approach is tolerated and the focus is put on the best binary representation of spatial irradiance distribution. In *SkySol* irradiance modelling methods make use of on-site measurements of either GHI only, or GHI and DHI. Even if single point measurements might be not representative for the whole domain or time (when forecasting is considered), they provide useful information to determine representative irradiance maps. The modelling is done for every forecast update considering the current image and irradiance measurements. Therefore the binary mapping dynamically adapts to current conditions by integrating near real-time local irradiance measurements.

To understand the transformation process, a basic understanding of cloud-irradiance interactions is needed. For simplicity, a differentiation of three different cases is useful:

- clear sky,
- overcast,
- mixed sky.

In clear sky conditions no clouds are present and the irradiance field is mostly homogeneous (disturbances by local variations of atmospheric composition of aerosols or water vapour are neglected in the model). In this case, GHI, DHI and DNI can be modeled with clear sky irradiance models. They model the radiative transfer in cloud-free conditions. The accuracy, if compared to measurements, depends on how accurate they represent the atmospheric compounds. For example, high aerosol load in hazy skies lead to a reduction in DNI while DHI is enhanced due to circumsolar scattering at aerosol particles. Online and onsite measurements can be used to correct the estimations for these deviations from modelled clear sky irradiance.

Overcast conditions are characterized by a full coverage of the sky with clouds. Usually, if clouds are non-transparent, no DNI can be measured. In this case, GHI is equal to DHI. The level of DHI may vary in space and time depending on the cloud optical properties. These conditions can be classified similar to clear sky conditions as homogeneous, but on a lower absolute GHI level (typically 10-30% of clear sky irradiance).

Mixed skies are characterized by scattered (single small clouds) or broken clouds (cloud layer with gaps) in a single layer or in multiple heights above ground. Cloud-irradiance interactions are complex due to scattering at cloud edges or due to clouds with varying optical properties. The resulting irradiance distribution is characterized by different shades of GHI intensity depending mainly on the variations in DNI. DNI is directly affected and reduced by clouds in the path of the sun rays. DHI is in most cases enhanced from clear sky levels due to scattering and reflections at cloud edges. Its variability depends mainly on the overall cloud cover and is lower than DNI variability.

A time series of measured GHI, DHI and DNI of a day in May 2015 in Oldenburg is shown in Figure 2.2 illustrating typical patterns in all three described sky types (clear sky in the morning, mixed sky at noon and overcast in the afternoon).

#### 4.6.1 Transformation

The mapping of binary shadow maps to irradiance maps is done as following: For cloudy or respectively shaded areas GHI is equal to DHI. For the non-cloudy, respectively non-shaded areas, GHI is the sum of DHI and the horizontal component of the cloud-free direct irradiance  $DNI_{clear}$  (Eq. 4.18).

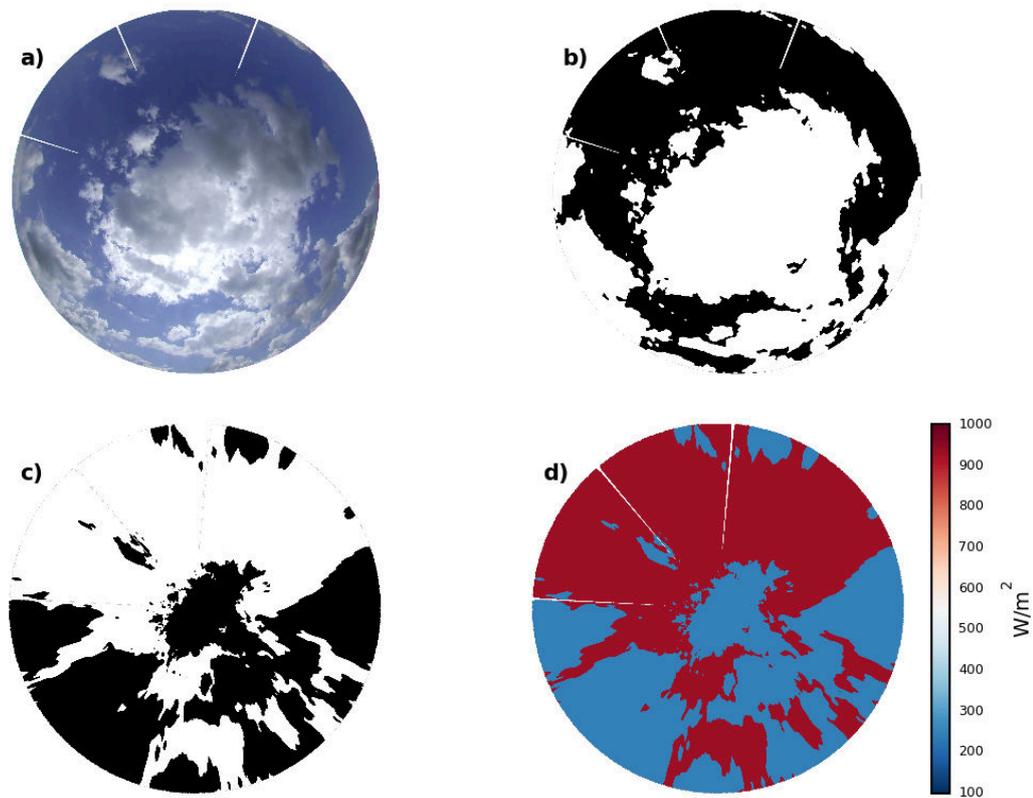
This approach neglects that clouds can be semi-transparent and  $DNI > 0$  could be measured even if the location is shaded. Due to the rather low variability in diffuse radiation, we assume DHI to be constant in time (for the forecast horizon considered) and space (for the chosen domain or covered area respectively).

$$GHI = \begin{cases} DHI & , \text{cloudy / shaded} \\ DHI + DNI_{clear} * \cos(\theta_{sun}) & , \text{sky / non-shaded.} \end{cases} \quad (4.18)$$

Next, the levels of DHI and DNI in the area of interest have to be determined. In *SkySol* three different approaches are implemented:

1. “histogram”: both states are estimated from statistical analysis of past GHI measurements
2. “machine learning”: DHI is estimated from image features and DNI is computed as the difference of measured GHI and estimated DHI
3. “measured components”: using on-site and real-time measurements of DHI and DNI.

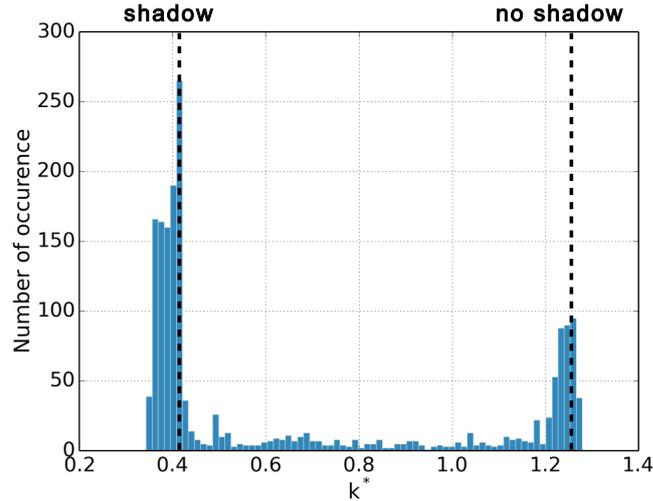
All methods require additional measurements of at least GHI (1,2) or additionally DHI (3). The machine learning approach (2) needs DHI measurements for the training of the model. It can be assumed that in most applications at least a GHI sensor is available and data can be accessed in near real-time. It can be expected, that measuring the irradiance components directly would be the most precise method. The methods have in common, that they provide a single point estimation of DHI and DNI. Both values are used to transform the shadow map to an irradiance map based on Eq. 4.18. As a result, there is no spatial variation in DHI and DNI levels. For a better representation of the spatial variability, a more detailed modeling based on cloud distribution and cloud optical properties is needed to account for transparencies or scattering at cloud edges. An example of the binary transformation is presented in Fig 4.12. Obviously, the binary pattern of the shadow map is conserved in the irradiance map. The GHI values in the shaded regions are equal to the current DHI value, while GHI in the non-shaded regions equals to the sum of DHI and  $DNI * \cos(\theta_{sun})$ .



**Figure 4.12:** Global horizontal irradiance mapping. a) Masked raw image. b) Cloud detection results (clouds in white). c) Cloud shadow projection results (cloud shadows in black). d) GHI mapping.

### 4.6.2 Histogram method

The histogram method is based on past records of GHI measured with a pyranometer at the location of the camera or at the site of interest and normalized to clear sky index  $k_{GHI}^*$  (Eq. 2.1). A typical histogram of measured  $k_{GHI}^*$  has two peaks or modes for shaded and non-shaded conditions (Fig. 4.13). These two modes are identified, converted back to absolute irradiance and serve as input for Eq. 4.18.



**Figure 4.13:** Histogram of measured clear sky indices  $k^*$  in the past 30 minutes in order to determine  $k^*$  for shadow and no shadow state.

The histogram is computed with measurements in the past 30 minutes. The method takes the global distribution peak for  $k^* < 0.5$  for shadow state and  $k^* > 0.9$  for no shadow. We decided to use 100 bins for  $0.2 \leq k^* \leq 1.4$ . If no peaks can be determined (in case of homogeneous irradiance conditions in the past 30 minutes), default values of  $k_{hist,low}^* = 0.4$  and  $k_{hist,high}^* = 1.0$ , respectively, have been assigned for the two states. The corresponding GHI can then be calculated with

$$GHI = \begin{cases} k_{hist,low}^* * GHI_{clear} & , \text{cloudy / shaded} \\ k_{hist,high}^* * GHI_{clear} & , \text{sky / non-shaded.} \end{cases} \quad (4.19)$$

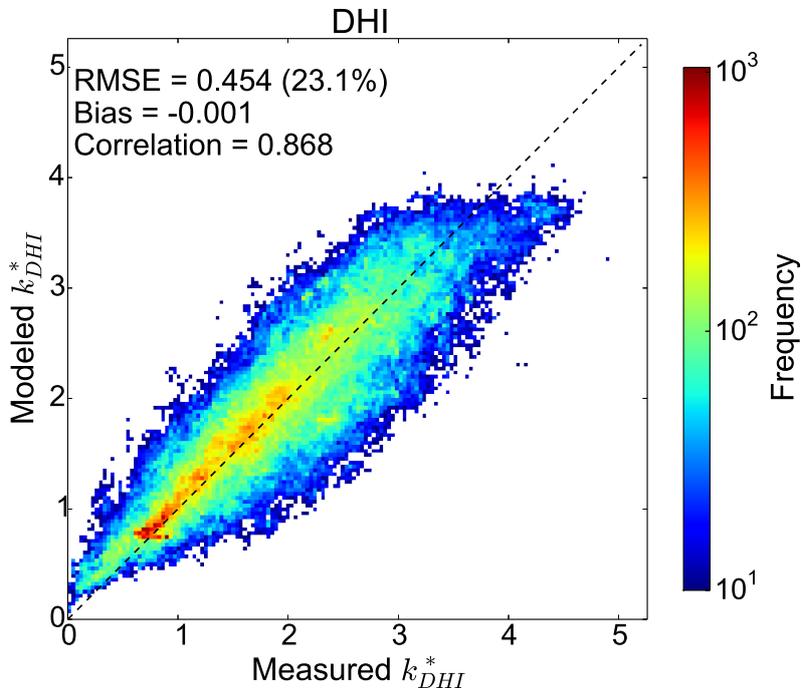
One disadvantage of the histogram method is the lack of currentness of data as its estimations are the result of statistics of 30 minutes of measurements and therefore react on changes with certain delay.

### 4.6.3 Machine learning method

Machine learning algorithms are supposed to learn the statistical relationship between a set of features (in this case describing the image) and a certain target value called

label (in this case irradiance). Once the training of the model has been finished (and the relationship does not change), the model can be applied on new images to get an estimation of irradiance conditions. The approach should be used to retrieve an estimation of current DHI and DNI values based on current image characteristics and real-time on-site GHI measurements. The method requires at least DHI measurements and the image features (described in Section 4.8.1) in its learning phase. In the application phase, current DHI can be estimated by the model (with the current features as input) and DNI can be computed with the use of on-site GHI measurements.

While the histogram-method determines both values from past time series, this method should be based on current conditions to account for short-term variations in DHI and clear sky DNI levels.



**Figure 4.14:** Scatterplot of kNN-modelled and measured  $k_{DHI}^*$ . The results are based on a test data set of 2014 recorded in Oldenburg. The comparison is based on 1 Hz irradiance measurements.

The model trained for the forecast results presented in this thesis is a k-nearest-neighbour (kNN) model, trained on a data set of image features and irradiance measurements recorded in Oldenburg in the year 2014 [91]. The algorithm provided by the scikit-learn package [92] has been implemented in *SkySol*. The data set has been split up in a training data set (70 %) and a test data set (30 %). The kNN-model with 500 neighbours has been trained to predict the clear sky index of DHI with a set of image features. The clear sky index is used to remove diurnal and seasonal variations in the radiation data. Two types of models have been tested. A model with a full set of features and a model with a reduced set of features only using the most important features. For the latter model, feature importances have been estimated first using a *RandomForest*

model. The feature subset has been defined by a forward feature selection that adds incrementally features sorted by their importances to the kNN-model and stops, when the forecast error does not decrease anymore. The model using a reduced subset of features performs has a correlation of 0.87 with high-resolution measurements based on the test data set (Fig. 4.14). Schmidt et al. [91] gives a more detailed description of the model and its performance. Similar approaches has been presented recently by Dev et al. [93] and Kurtz and Kleissl [11].

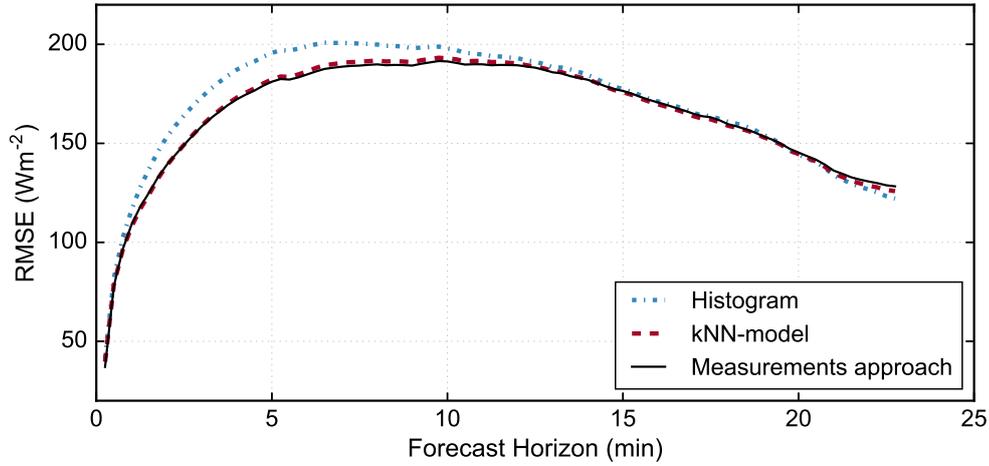
As mentioned above, with the real-time estimation of DHI, DNI can be estimated as the difference of measured GHI and the DHI value estimated by the kNN-model. To select a representative clear sky  $\text{DNI}_{\text{clear}}$  value needed for the non-shaded areas, the real-time  $\text{DNI}_{\text{est}}$  estimation cannot be used directly, as it might be estimated in a non cloud-free situation. Therefore,  $\text{DNI}_{\text{est}}$  estimations of the past 20 minutes are used to determine the maximum  $\text{DNI}_{\text{est}}$  in the timeseries, which is assumed to be the clear sky  $\text{DNI}_{\text{clear}}$ . If this value is below a certain threshold, no clear sky condition could be evaluated and  $\text{DNI}_{\text{clear}}$  is set to the clear sky  $\text{DNI}_{\text{clear,model}}$  from the clear sky reference model.

#### 4.6.4 Measured components

Ideally, DHI is measured at the location of interest. Together with GHI measurements, up-to-date and accurate DHI information can be used for the irradiance estimation. For this thesis, we use measurements as a reference to the former methods.

#### 4.6.5 Method comparison

A comparison of the three proposed methods has been performed on a two-month data set from June to July 2015 in Oldenburg. Single point GHI forecasts for the location of the camera have been computed with equal overall model configuration but varying irradiance modelling approach (see next Section 4.7 on irradiance forecasting). The forecasts include all postprocessing options described in the next section. 15 seconds averages are evaluated to compare the performance of all three methods. Figure 4.15 shows the root mean square error (RMSE, definition in Eq. 5.3) as a function of the forecast horizon. An RMSE reduction of the kNN-model compared to the histogram-method can be stated for the first 15 minutes, while using DHI measurements adds almost no additional benefit. For longer time horizons, all methods show similar RMSE. RMSE has not been computed for time horizons longer than 23 minutes since less than 20% of the forecast data was available for longer horizons. A discussion on the decreasing RMSE for longer time horizons is presented in the forecast performance chapter.



**Figure 4.15:** Forecast RMSE for three different irradiance model configurations. Forecasts have been computed with a maximum forecast horizon of 25 minutes in 1 s resolution for two months from June to July 2015.

## 4.7 Irradiance forecast

In order to predict future irradiance distribution, cloud maps (Sec. 4.5) are extrapolated by applying the determined cloud motion vector (Sec. 4.4). Shadow maps are generated (Sec. 4.5) based on the sun position at the corresponding timestep. With the transformation of the shadow maps to irradiance maps (Sec. 4.6), time series of GHI can be derived for each location in the covered area. Using the estimations of the diffuse irradiance level for the whole area  $DHI_{est}$  and for direct irradiance  $DNI_{clear}$  at non-cloudy locations the forecasted  $GHI_{fcst}$  timeseries can be determined by

$$GHI_{fcst} = \begin{cases} DHI_{est} & , \text{ no sun} \\ DHI_{est} + DNI_{clear} * \cos(\theta) & , \text{ sun} , \end{cases} \quad (4.20)$$

representating a binary timeseries with high  $GHI_{fcst}$  values in non-shaded conditions and low  $GHI_{fcst}$  values in times of cloud shading. Deviations from binary timeseries are a result of averaging effects.

### 4.7.1 Forecast resolution

Temporal and spatial forecast resolution can be chosen from highest 1 s and 1 m resolution to coarser resolutions. The internal temporal resolution of the model is always 1 s but will be averaged to lower resolution if is required. The spatial resolution depends on the selected underlying grid for the cloud and shadow projection (Sec. 4.5). Forecasts for a single point is based on the grid point information in which the point

of interest is located. Due to smoothing effects from averaging, forecasted time series variability depends on the resolution given by the model configuration.

### 4.7.2 Post-processing

In *SkySol* three optional features can be used to modify the direct transformation from shadow maps described in the previous subsection.

- Cloud edge smoothing
- Online bias correction
- Circumsolar area correction

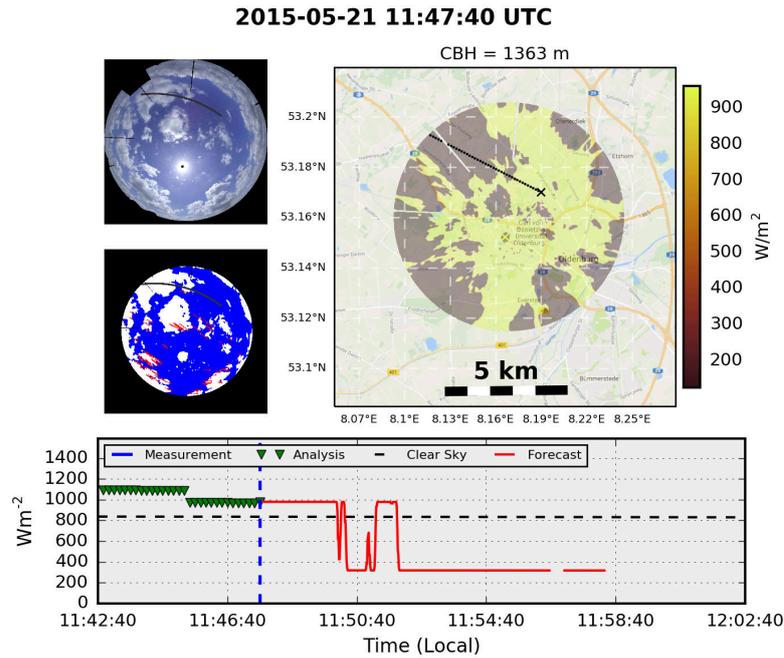
#### Cloud edge smoothing

Smoothing cloud edges also smoothes irradiance ramps that result from the transition from shadow to non-shadow regions. The cloud edge smoothing is closer to reality, because cloud edges are rarely sharp leading to irradiance ramps having variable duration depending on the sharpness itself, cloud speed and height. In order to smooth cloud edges, a gaussian blurring is applied on the binary shadow map. The strength of the blurring is computed based on physical considerations including the sun opening angle and the cloud height. This accounts for the sun angular diameter and represents the penumbra. If no cloud height is available and only local forecasts are computed, a fixed value for the blurring is used.

The forecast example in Figure 4.16 shows a forecast for a selected location (no measurements available) a few kilometers north-east of the camera in the city of Oldenburg illustrating the cloud edge smoothing. From the derived cloud motion vector and the cloud shadow projection based on the provided cloud base height and the sun position the forecasted irradiance time series for the example location is derived. The field of view for the forecast has been limited to  $70^\circ$  viewing angle. Due to fast moving clouds, the forecast horizon is about 10 minutes and a cloud event (shading) is predicted to happen in about one minute. The irradiance levels are modelled with the kNN-model. Obviously, the predicted ramps are smoothed according to the described algorithm. Without the cloud edge smoothing, an immediate drop from from the high to the low irradiance level would be the case.

#### Online bias correction

An online bias correction can be applied to correct the irradiance level estimations by using near real time local GHI measurements. The correction accounts for under- and overestimations of DHI levels in cloudy conditions and is implemented as following:



**Figure 4.16:** Forecast example for a selected point in Oldenburg a few kilometers north-east of the camera. For the given cloud base height (provided by DWD station Altenoythe-Friesoythe 30km SW from Oldenburg) of 1363 m the cloud shadows have been projected on a map and converted to GHI. Clouds were moving from west to east. The predicted cloud path from the maximum field of view ( $70^\circ$  zenith) to the station is drawn as black dotted line in the map. The projection of the location back to the image is drawn in the fisheye images (black solid lines).

1. Estimate  $DHI_{est}$  from past GHI measurements (Option 1) or from the kNN-model (Option 2).
2. compute  $DNI_{est}$  as the difference  $DNI_{est} = \frac{GHI_{meas} - DHI_{est}}{\cos(\theta)}$ .
3. if  $DNI_{est} < 0$  (i.e. overestimation of DHI),  $DHI_{est}$  will be corrected to match  $GHI_{meas}$  and  $DNI_{est}$  is set to 0.
4. if  $DNI_{est} == 0$ , correct  $DHI_{est}$  bias that it matches to  $GHI_{meas}$ .

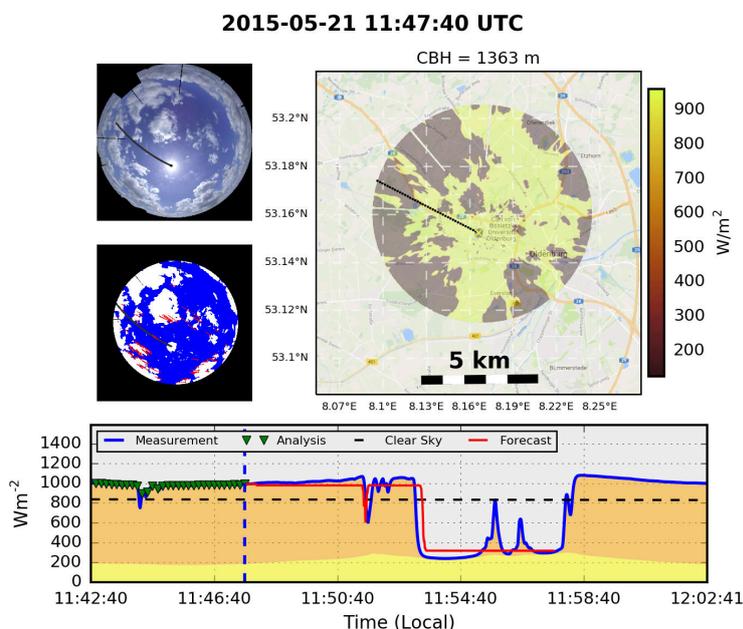
#### Circumsolar area correction

The circumsolar area is often affected by scattered sunlight and misclassifications are likely. Usually, the sunspot is classified as cloudy even if it is not covered by clouds (see Section 4.3 for a more detailed discussion on the circumsolar area). The misclassifications affect the very short-term forecasts for locations close to the camera. As a result, irradiance is likely underestimated in the first one to two minutes. The duration depends on cloud speed and the size of the circumsolar area affected by brightness and misclassification.

A simple correction for the circumsolar area is applied to account for these problems. For local site forecasts (predicting the time until sun is blocked by clouds) the pixel intensity as well as the RBR value along the forecast path is measured. In a clear situation, the first pixels (along the forecast path) starting from the sunspot, should be saturated (pixel intensity = 255) and the RBR should be equal to one. An intensity threshold of 80% saturation is used together with the RBR threshold for cloudy pixels to classify a pixel as part of the affected circumsolar area. As long as these conditions are met, the forecast value is independent of the binary cloud detection. In this case, a simple persistence of the locally measured values is used (Eq. 4.21).

$$GHI(t_0 + \Delta t) = GHI(t_0) \quad (4.21)$$

The forecast example in Figure 4.17, which is showing local single site forecasts (for the camera location) for the same image instance as in Figure 4.16 shows the effect of circumsolar area correction (note: measurements are available for this site). The circumsolar area is misclassified as cloudy (see cloud decision and shadow projection map), but the forecast time series shows a persistence of initial values until the circumsolar area is left (first seconds). Otherwise, first seconds forecasts would predict cloudy conditions leading to large very short-term forecast errors. Compared to the previous figure, the measured time series proves that the forecast match the irradiance levels as well as the down-ramping time at about 11:53 local time. The cloud edge smoothing reduces the ramp rate slightly in accordance to the measurements. Moreover, a short GHI drop is predicted due to a small scale cloud predicted to occlude the sun



**Figure 4.17:** Same graphics as in previous Fig. 4.16, but with an illustrated forecast for the location of the camera. Locally measured irradiance components (diffuse irradiance in yellow, direct horizontal irradiance in orange) are added to the timeseries.

in about five minutes (11:51). The cloud edge smoothing leads to a smoothed and minor reduction in GHI also matching with the measurements.

The circumsolar area correction makes use of the saturation characteristics in this region to reduce forecast errors, but will not overcome the problem of brightness or even saturation in the circumsolar area in case of direct sunlight. Future developments in camera hard- and software improving the dynamical range (e.g. high dynamic range HDR) already demonstrated by other groups (e.g [33, 12]) show promising results, however cannot completely overcome the problem.

## 4.8 Additional components of SkySol

### 4.8.1 Image features

Image features are computed for each image for the usage in machine learning applications. The features are quantified characteristics like image textures, color values and other metrics. As said in the irradiance modelling section (Sec. 4.6), machine learning algorithms are supposed to train the relationship between the features describing the image and a certain target value called *label*. In *SkySol*, irradiance modeling (predicting actual surface solar irradiance from image features, Sec. 4.6) and cloud type classification (predicting a certain dominant cloud type in the image, Sec. 4.8.2) are implemented. Features are selected to quantify special image characteristics discriminating for example a full cloud cover from mixed cloud conditions. In terms of irradiance modeling, features describing the pixel intensity in the circumsolar area are used as the implicit information about direct solar radiation. For example, in overcast conditions with thick clouds, the circumsolar area pixel intensity will be much lower than in clear sky conditions. The features are computed on the masked image, that means non-sky parts of the image are masked out, before image features are computed.

The features can be grouped in global textural features, global color-based features and local features.

#### Global textural features

Global textural features are image texture properties derived from the grey-level co-occurrence matrix (GLCM) and defined by Haralick et al. [94]. The *angular second-moment (ASM)* feature is a further measure for homogeneity, *correlation* is a measure of grey-tone linear dependencies and *dissimilarity* is a measure that defines the variation of grey level pairs in an image [95]. For more details on the textural features interested readers are referred back to Heinle et al. [29]. The features are listed below:

- Homogeneity
- Contrast
- Energy
- ASM
- Correlation
- Dissimilarity

### Global color-based features

Global color-based features are features based on color values of the raw RGB (or transformed HSV color space) image. They include statistics of single color channels or are derived from combinations. Most of them are self-explaining, but described in Heinle et al. [29]. Cloud coverage is the result of the cloud decision process and means the ratio of cloudy pixels to all pixels. The cloud coefficient and the grey coefficient are defined as features averaging the pixel intensity of all pixels classified as clouds weighted with their distance to the center of the image. Sky pixels do not contribute to the final value. Both coefficients differ in weighting pixel intensity. While cloud coefficient is based on a linear weighting of the distance to the center of the undistorted image, the grey coefficient weights pixels based on the cosine of its viewing zenith angle.

- Mean Red
- Mean Blue
- Mean pixel intensity
- Difference Red-Blue
- Difference Red-Green
- Standard deviation Blue
- Skewness Blue
- Ratio of saturated pixels to all pixels in HSV color space
- Ratio of saturated pixels to all pixels in RGB color space
- Mean RBR
- Cloud coverage
- Cloud coefficient
- Grey coefficient

### Local features

Local features are the percentage of saturated pixels in a specific angular distance to the sun. The saturation is computed based on the number of zero values of Hue in the HSV color space:

- saturation in a sun angular distance  $< 0.5^\circ$
- saturation in a sun angular distance  $< 1^\circ$
- saturation in a sun angular distance  $< 5^\circ$
- saturation in a sun angular distance  $< 7^\circ$
- saturation in a sun angular distance  $< 10^\circ$
- saturation in a sun angular distance  $< 15^\circ$
- saturation in a sun angular distance  $< 20^\circ$

The saturation in the circumsolar area can also be used as an indicator for direct sunshine [91, 11].

### 4.8.2 Cloud type classification

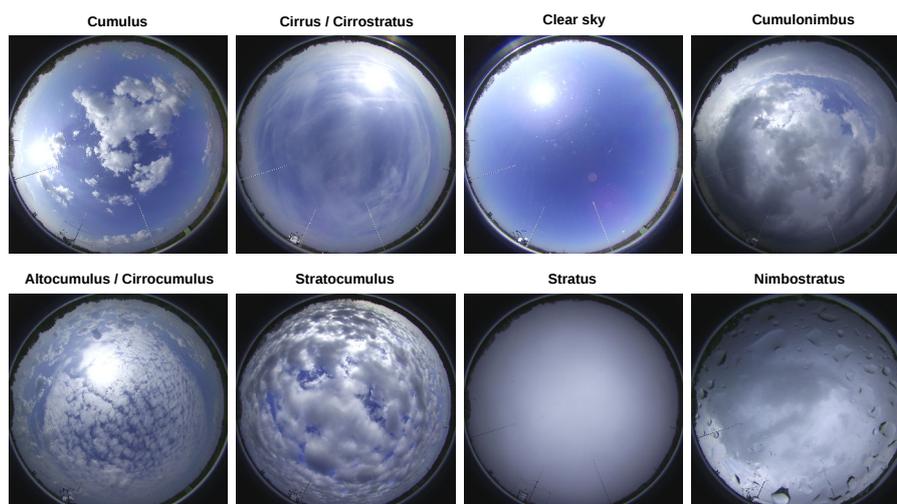
A cloud type classification scheme is part of *SkySol*. The objective is to classify each image in one of seven cloud type categories. We use a modified version of the scheme introduced by Heinle et al. [29]. The seven categories are meteorologically justified according to the mentioned paper and an example for each is given in Fig. 4.18:

- **Cumulus (Cu)** - low, puffy clouds with clearly defined edges, white or light-grey
- **Stratocumulus (Sc)** - low or mid-level, lumpy layer of clouds, broken to almost overcast, white or grey
- **Cirrocumulus (Cc), Altopcumulus (Ac)** - high patched clouds of small cloudlets, mosaic-like, white
- **Nimbostratus (Ns), Cumulonimbus (Cb)** - dark, thick clouds, mostly overcast, grey
- **Stratus (St), Altostratus (As)** - low or mid-level layer of clouds, uniform, usually overcast, grey
- **Cirrostratus (Cs), Cirrus (Ci)** - high, thin clouds, wisplike or sky covering, whitish
- **Clear sky (Clear)** - no clouds and cloudiness below 10

Typical images representing the seven categories and synchronized  $k^*$  readings are presented in Figure 2.1 in the introduction. Obviously, characteristic  $k^*$  timeseries for each category can be figured out showing that this kind of classification is meaningful in terms of solar irradiance forecasting.

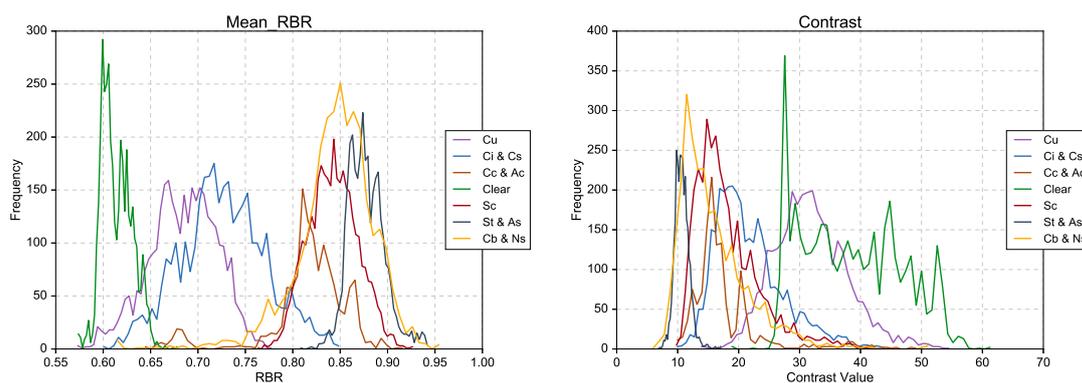
The classification can be used to evaluate forecast performance under different cloud conditions (see Sec. 5.6 for forecast performance results). Interested readers are referred to a review of existing cloud detection and classification methodologies by Tapakis and Charalambides [96].

The cloud type classification module makes use of machine learning based classification



**Figure 4.18:** Typical samples of the seven cloud type classes. Nimbostratus and Cumulonimbus are summarized in one class.

schemes from the scikit-learn package [92]. After testing several classification schemes the *Support Vector Classification (SVC)* has been chosen with best classification results. Multi-label datasets (seven cloud types) are used for training the models. From Oldenburg image archive about 21500 images were manually pre-selected into the different cloud type categories. For each image, global image features (Sec. 4.8.1) characterizing image texture properties and color values are computed and normalized. It is assumed that the chosen image features represent and discriminate the different categories. Figure 4.19 illustrate the probability distribution of each cloud type class for the global mean Red-Blue-Ratio and the Contrast feature. Obviously, single features are able to discriminate some cloud type classes better than others.

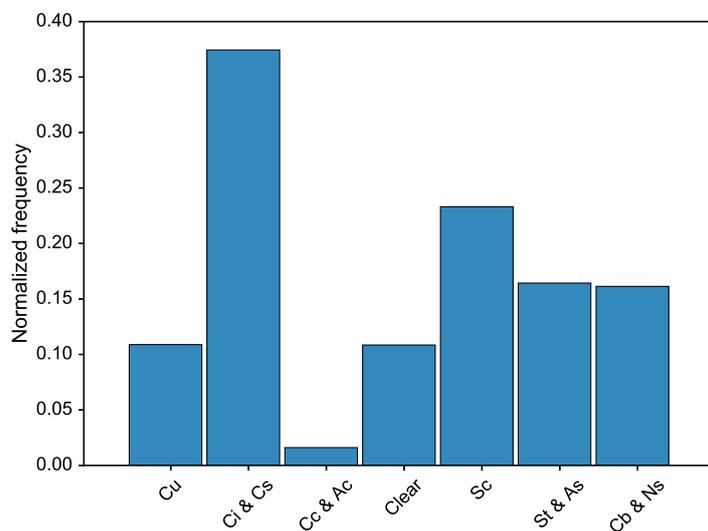


**Figure 4.19:** Probability distribution of each cloud type class for the feature “mean RBR” (left) and “Contrast” (right) computed for the manually classified dataset.

A feature ranking based on a decision tree algorithm has been tested to evaluate the importance of each feature used and also to avoid correlations among the features that might lead to an overfitting of the model. Results have shown no indications of

overfitting when using a high number of features even if high performances ( $>95\%$  accurate classification of independent images) are reached for the five most important features.

For the application on new images the SVC model returns probabilities for each class to be the dominant cloud type in the image. The cloud type class with the highest probability is selected as the final decision. The classification performance has been determined by applying a model trained on a subset of pre-classified images (70%) on the remaining 30% of images. An accuracy of more than 99% correct classification is reached. When interpreting the evaluated performance it has to be stated, that only characteristic images for each cloud type are used. Mixed cloud conditions or others than represented by the seven classes are not considered and can be misclassified or classified with poor probability in operational use.

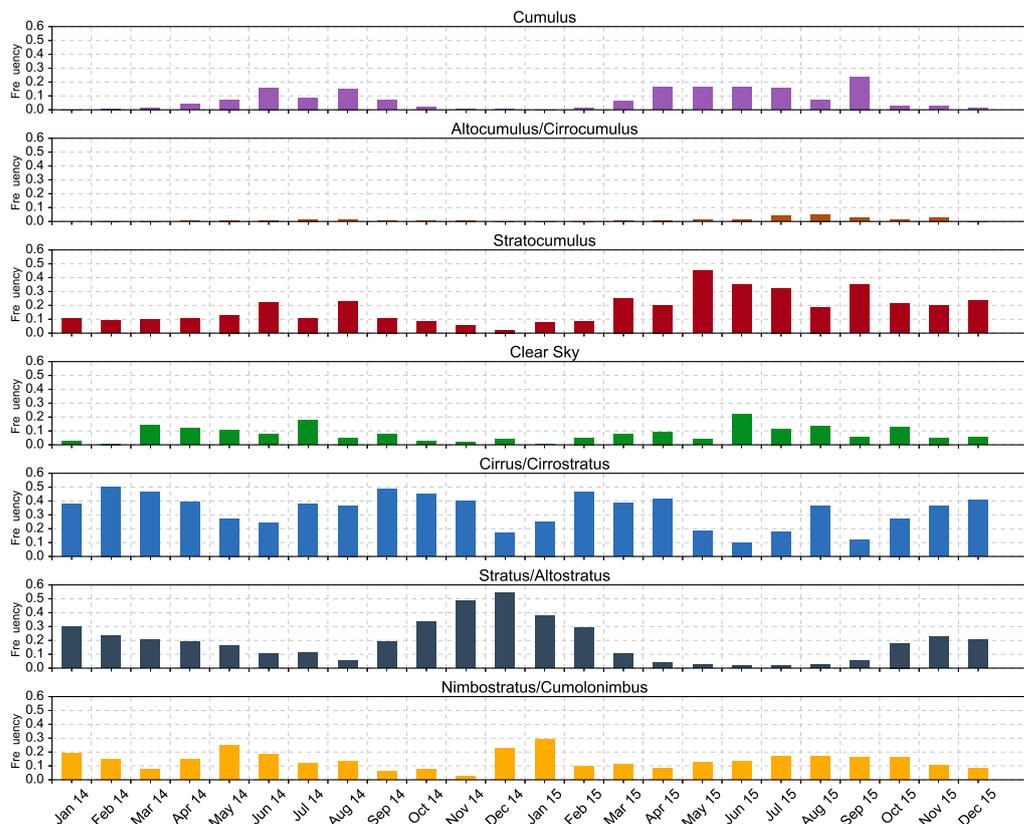


**Figure 4.20:** The share of cloud types classified with a probability of more than 70% for the years 2014 and 2015.

The cloud type classification has been applied and used for the Oldenburg image archive and the images from the HOPE campaign. Different training sets have been used for different camera types, because camera settings (e.g. color balance) differ too much. Misclassifications are the result, if the model trained on images of one camera type is applied to another camera type. A more general approach which can be used independent of camera characteristics and settings has to be developed in future. For the HOPE campaign, a smaller dataset with a reduced number of features has been used by Schmidt et al. [40].

Figure 4.20 show the histogram of cloud type occurrences in Oldenburg that are classified with a probability higher than 70% accounting for about 75% of all images. Cirrus and cirrostratus clouds are the most frequent. Clear sky periods are present in almost 10% of the time. Typical seasonal variations can be seen in the timeseries given in

Figure 4.21. Convective cloud types like cumulus are more common in summer than in winter as expected. On the other hand, stratus dominates in winter times. While altocumulus and cirrocumulus are rare, cirrus type clouds are present often throughout the year.



**Figure 4.21:** Monthly shares of dominant cloud types classified with a probability of more than 70% for the years 2014 and 2015.

### 4.8.3 Solar geometry

The solar zenith and azimuth angle for any location and time are calculated with the solar geometry2 (SG2) algorithm [80].

### 4.8.4 Clear sky irradiance modelling

Clear sky irradiance is estimated with the clear sky model of Dumortier [97] and turbidity values according to [98] and [99]. The clear sky irradiance of all components (GHI, DHI, DNI) is frequently used in *SkySol* for the normalization of absolute irradiance. Especially the irradiance modelling scheme uses clear sky indices due to their independency of diurnal and seasonal patterns in irradiance.

## 4.9 Spatial coverage and forecast horizon

As already mentioned, the forecast horizon that can be reached for a single location in the area covered by the sky image depends on several parameters and is variable in time and space. Here, the most important ones are:

- **Clouds base height (CBH)** above ground influences clouds' distance to the camera and the position of its shadow on the ground. As a result, CBH scales linearly the area covered by the sky image (see Fig. 4.11).
- **Sun position** influences the distance and direction of a cloud shadows projection on the ground. The area covered by the shadow maps varies during day and season.
- **Maximum viewing angle** that is considered for cloud detection, influences the field of view. For large viewing angles perspective errors become dominant. On the other hand, the distance of a cloud to the camera and therefore the field of view is doubled, if the viewing angle is increased from 70 to 80 degrees (see Fig. 4.11).
- **Cloud motion direction** strongly influences the forecast horizon for locations distant of the camera. In combination with sun position, the largest horizon is reached, if clouds are moving from the opposite direction of the sun azimuth.
- **Cloud motion speed** influences the forecast horizon by the duration in which clouds are passing the field of view.

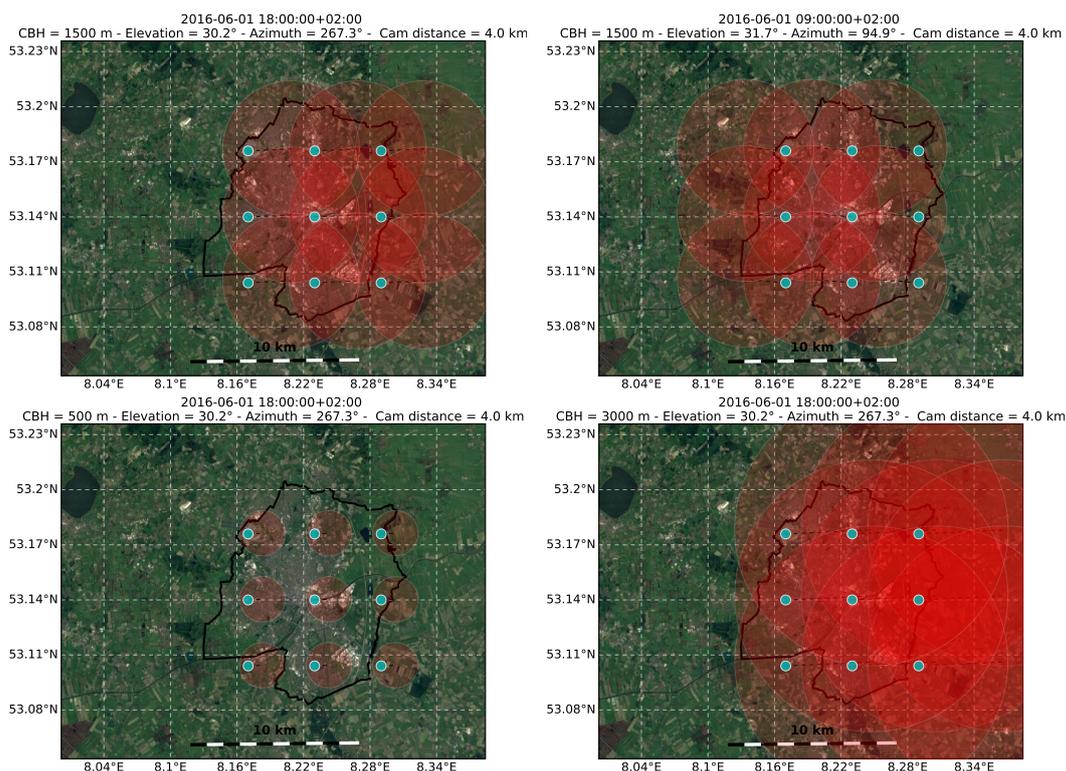
Note, that this discussion on the achievable forecast horizon is based on a single cloud layer assumption and neglects the 3-dimensional arrangement of clouds.

The complex relationship of forecast horizon and its influencing factors has been modeled for the location of Oldenburg and results are presented in the following sections.

### 4.9.1 Modelling the spatial coverage of multiple cameras

First, the objective was to illustrate the coverage of multiple cameras distributed in the city of Oldenburg to cover the whole city independent of the prevailing cloud motion. Figure 4.22 shows the results for nine cameras positioned with four kilometres distance between each other for a day in June. The maximum viewing angle was fixed to 70 degrees zenith angle. A single cloud layer with a height of 1500 metres (top row of the figure) above ground has been defined. This modeled example shows that for the given configuration, the city can be covered with nine cameras. Obviously, the maximum forecast horizon is different for locations in the center and the borders. This phenomena is amplified by the prevailing cloud motion which is discussed in the next subsection. The figures also show that the covered region is moving during the day as the projection of cloud shadows depends on the sun position.

In this configuration, all cameras cover a region of about 15 km x 15 km with overlapping areas. The overlapping can be used for example for CBH estimation with triangulation. The coverage and the areas that overlap are about to change, when clouds have a higher or lower height. The same Figure 4.22 (bottom row) shows the results for cloud heights of 500 and 3000 meters, respectively. Apparently, low clouds reduce the area covered by the image and therefore the forecast horizon dramatically. For a full coverage of the city a much denser distribution would be needed. In contrast, for high clouds a very large covered area with large overlapping regions can be reached.



**Figure 4.22:** Spatial coverage of nine cameras (blue dots) distributed over the city of Oldenburg (black line) as modeled for different configurations. The coverage (red circle) of each camera represents the projected shadow of clouds in a single cloud layer visible in the camera image. The black dotted lines illustrates the length and direction of the shadow. A distance of 4 km between each camera is assumed. Top left: Model results for CBH = 1500 m, late afternoon. Top right: CBH = 1500 m, early morning. Bottom left: CBH = 500 m, late afternoon. Bottom right: CBH = 3000 m, late afternoon. Background image source: *Google Tiles*.

#### 4.9.2 Modelling maximum forecast horizon for varying cloud motion

The influence of cloud motion on the reached forecast horizon has been modeled for different specific cloud directions and speed. It can be expected, that higher cloud speed reduce the forecast horizon. Moreover, the cloud direction influences the forecast horizon for locations apart from the camera position. For example, for prevailing

western cloud directions (clouds moving *from west to east*), locations in the east of the camera can be predicted with a larger horizon. Moreover, the combination of daytime or solar position respectively with cloud motion influences the forecast horizon.

Figure 4.23 shows analysis of maximum forecast horizons modelled for a defined cloud height of 1000 metres. The maximum forecast horizon in minutes has been computed for a whole year to cover all different solar positions also influencing the forecast horizon. It has been assumed, that for locations outside of the covered area, no forecasts can be computed. This assumption must not be true in reality, if a location is not covered at the moment, but for the predicted timestep. Here, these situations are left out. To summarize the statistics of all maximum forecast horizons, the 90th percentile of all computed conditions is shown in Fig. 4.23. Here, the maximum viewing angle is fixed to  $80^\circ$ .

First, the results show an oval shape of the area covered by the forecasts. The oval shape results from the diurnal course of the sun and the resulting shadow positions. The distribution of maximum forecast horizons confirms the expectation, that the area with largest forecast horizons is shifted with the prevailing cloud direction. For clouds moving from south-western  $250^\circ$  with a speed of  $5 \text{ ms}^{-1}$ , areas north-east of the camera location can be predicted more than 30 minutes ahead in 90 % of the time. Meanwhile, the camera location itself has a reduced forecast horizon. Increasing the cloud speed to  $15 \text{ ms}^{-1}$  (Fig. 4.23, top right) will reduce the overall forecast horizon by several minutes, because clouds are passing the area faster.

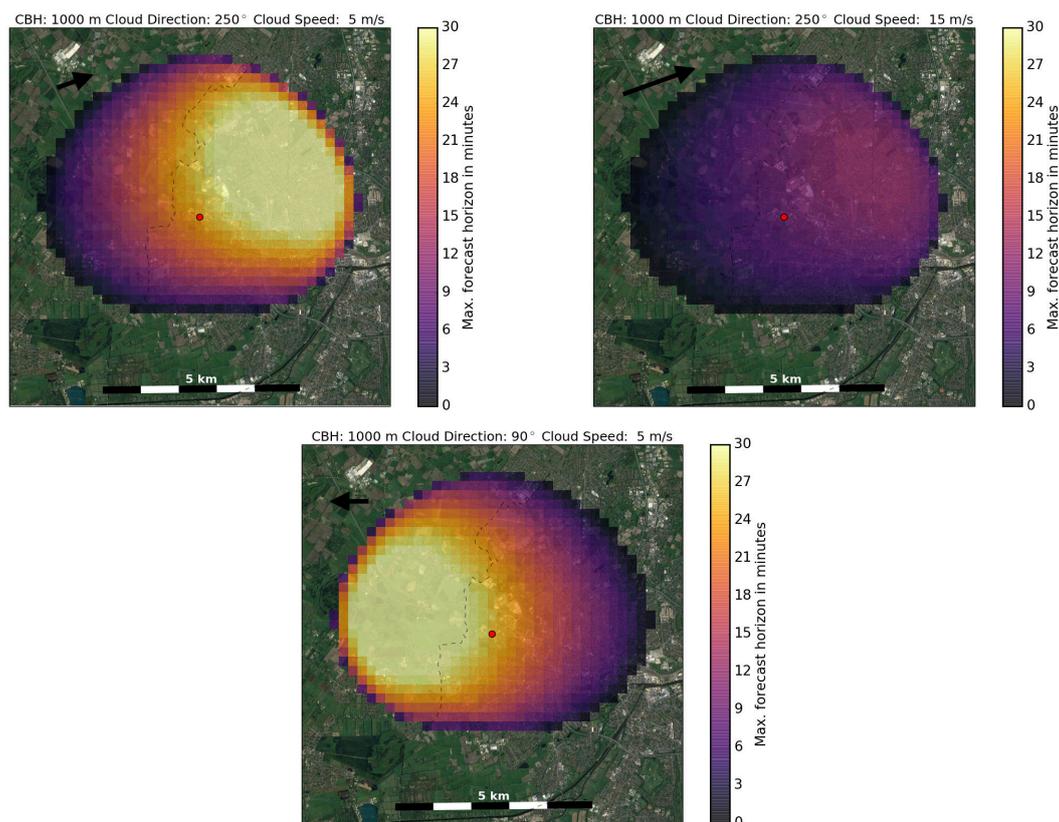
The results of these analysis on spatial coverage and forecast horizons influences considerations about sky imager setups in a certain location with specific climates. The decision where to install the camera and how dense a multiple installation must be if a larger area is about to be covered should be taken based on the results presented in this section.

#### Forecast horizons based on real data

For Oldenburg site, the maximum forecast horizon has been analysed for the year of 2015. Here, only the location of the camera is investigated (local single site forecast). As a result, CBH is not influencing the forecast horizon. That means, that no shadow projection has been applied. Forecasts has been computed based on the time, a cloud will block the sun (respectively shading the camera location). Analysing the maximum forecast horizon for this case means investigating the interaction between sun position and cloud motion.

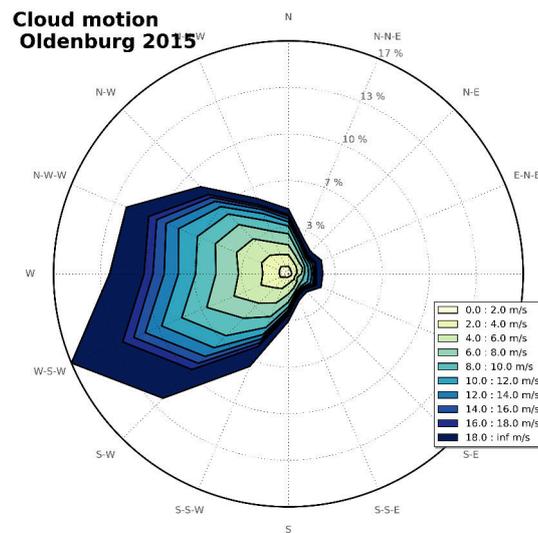
In Oldenburg, clouds are moving predominantly from west to east (Fig. 4.24). Therefore, it can be expected, that the maximum forecast horizon is larger in the morning than in the afternoon hours.

Figure 4.25 summarizes the cumulative distribution of forecast horizons computed for



**Figure 4.23:** Modeled maximum forecast horizons (90th percentile) reached for different cloud motion conditions: direction south-east  $250^\circ$ , speed  $5 \text{ ms}^{-1}$  (top left),  $250^\circ$  and  $15 \text{ ms}^{-1}$  (top right) and for  $90^\circ$  easterly cloud direction and  $5 \text{ ms}^{-1}$  cloud speed (bottom). The red circle indicates the location of the camera. Background source: *GoogleTiles*.

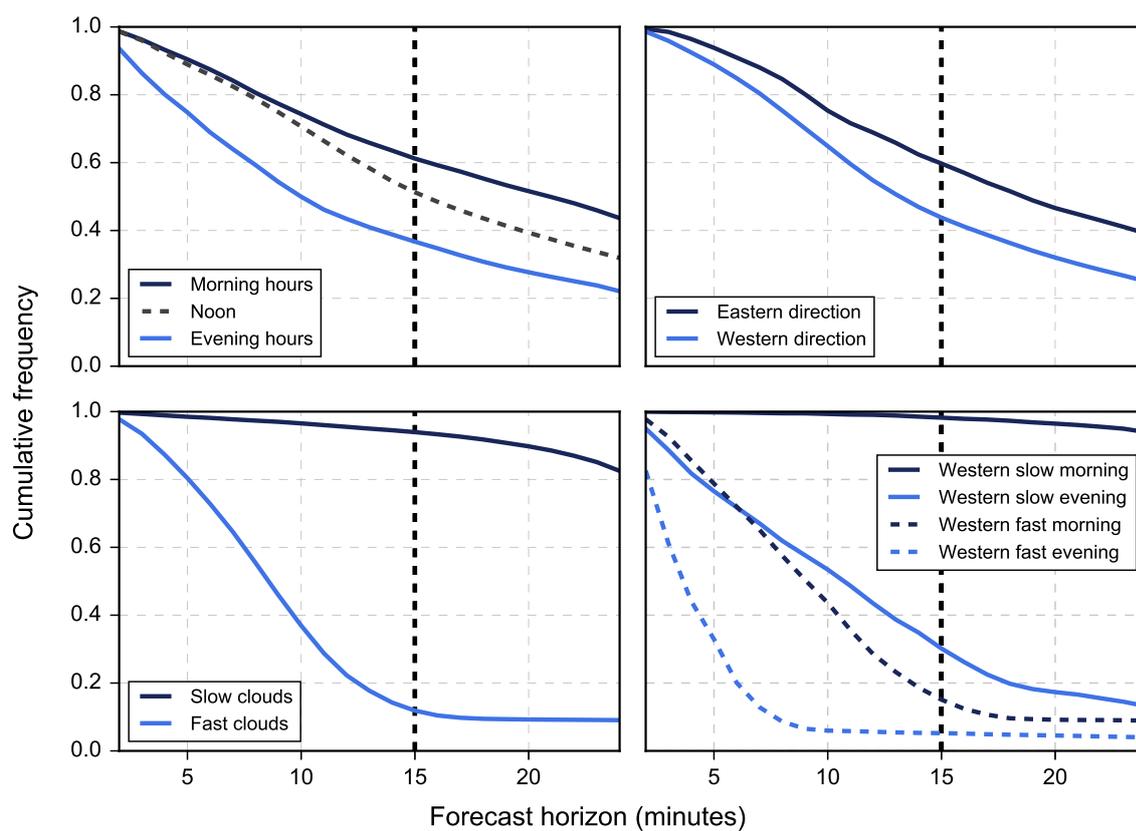
Oldenburg site. Here, forecasts are evaluated for sun elevations higher than  $10^\circ$  for the whole year 2015. With an update rate of 10 seconds more than one million forecasts build the basis of this analysis. First, it can be stated, that for noon hours (2 hours before and after solar transit) about 50 % of all forecasts reach 15 minutes horizon. For morning and evening hours (3 hours after sunrise or before sunset, respectively) slightly different forecast horizons are reached. This difference can be explained by the prevailing western cloud direction associated with higher average cloud speeds. This phenomena is highlighted in the top right figure. While in eastern cloud directions the 15 minute horizons is reached in 60 % of the time, less than 45 % reach 15 minutes for western cloud directions. Even more pronounced differences occur, if only conditions with fast ( $> 10$  pixel per seconds) and slow clouds ( $< 5$  pixel per seconds) are considered. While more than 90 % of all forecasts reach 15 minutes when clouds are moving slowly, the number drops to about 15 % for fast clouds. If we come back to the situation of westerly cloud motion in different day times illustrated before, Fig. 4.25 (bottom right) shows the extreme situations. In evening hours, the fraction of forecasts reaching 15 minutes is reduced to less than 30 % even for slow clouds. If only fast clouds are considered, almost 70 % of the forecasts will not be longer than 5 minutes. Consequently,



**Figure 4.24:** Distribution of cloud speed and direction derived with optical flow on sky images for the year 2015.

in morning hours under same conditions the forecast horizons are much longer because clouds passing the whole field of view before blocking the sun.

Overall, it could be pointed out, that the prevailing cloud conditions influence the forecast horizons that can be reached for a specific location in a specific climate regime. Here, the considerations have been evaluated for a location with dominating westerly cloud motion, leading to largest forecast horizons in morning hours when clouds are moving slowly.



**Figure 4.25:** Cumulative distribution of maximum forecast horizons reached for Oldenburg in the year 2015. The dataset has been evaluated considering different conditions: daytime (top left), cloud direction (top right), cloud speed (bottom left) and combinations (bottom right). The vertical dashed line highlights the 15 minute horizon.

# 5. Forecast Performance

The proposed model is evaluated by its forecast performance in this section. This is done by comparing the predicted irradiance (GHI) or PV output values (DC power) with on site measurements using different error metrics defined below.

The forecast performance will be demonstrated for different data sources (introduced in Chapter 3) and questions. The dataset from Oldenburg is the most comprehensive and used to show general forecast error metrics as a function of forecast lead time. The comparison with a reference forecast (persistence) is used to compute a forecast skill (Sec. 5.3). The data set covers multiple years and can therefore be used to give an overview about seasonal dependencies (Sec. 5.4). It will also be used to show the dependencies of forecast performance on cloud motion (Sec. 5.5) and different cloud conditions using the implemented image based cloud type classification (Sec. 5.6). In Section 5.7, the dependency of forecast skill under different cloud conditions is generalized with a variability measure. Another approach, demonstrated for a virtual PV-Diesel-Hybrid system, is the reduction of forecasted irradiance to binary cases, clear sky and cloudy, with a focus on the reliability to predict the transition from clear sky to a variable state (Sec. 5.9.2). Forecasting PV power output is evaluated based on the dataset from Munich, where sky images and PV data from a 1 MW-plant is available (Sec. 5.8). Unfortunately, no irradiance or PV data from a distant sensor is available for the Oldenburg data set. Therefore, the sky images and the pyranometer data from the HOPE campaign are used to evaluate performance for locations in different distances to the camera (Sec. 5.10).

## 5.1 Error metrics

The error metrics applied here are the “root mean square error” (RMSE), the “mean bias error” (MBE), the “forecast skill” (FS) and “accuracy” (ACC) for forecasted and measured time series comparison. The forecast error of a quantity  $x$ , e.g. global horizontal irradiance GHI is given as:

$$\varepsilon_i = x_i - y_i, \tag{5.1}$$

with  $x_i$  the measured and  $y_i$  the predicted value.

Standard measures for forecast performance are MBE and RMSE. MBE is the average deviation of the forecast  $y$  from the measurement  $x$  indicating systematic over- or underestimations.

$$\text{MBE} = \frac{1}{n} \sum_{i=1}^n \varepsilon_i, \quad (5.2)$$

where subscript  $i$  refers to a single forecast  $y$  or measurement  $x$  in the considered data set of size  $n$ .

By definition, RMSE is given by

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n \varepsilon_i^2}. \quad (5.3)$$

RMSE uses a quadratic weighting of forecast errors resulting in a stronger weighting of larger forecast errors.

Furthermore, the relative RMSE is the normalization of RMSE by the mean of all observations  $\bar{x}$ . It is used to make forecast errors more comparable and is expressed in percentage

$$\text{rel. RMSE} = 100 * \frac{\text{RMSE}}{\bar{x}}. \quad (5.4)$$

The comparison with a reference forecast is expressed in the forecast skill FS. It measures the improvement of the model with respect to the reference forecast:

$$\text{FS} = 1 - \frac{\text{RMSE}_{\text{Model}}}{\text{RMSE}_{\text{Reference}}}. \quad (5.5)$$

A positive FS means that the sky imager based forecast outperforms the reference forecast. As a reference, we use persistence for the whole evaluation. Persistence forecasts, as defined here, account for changing sun angles, but assume no change in cloudiness which is described by a constant clear sky index  $k^*$  respectively:

$$\text{GHI}(t_0 + \Delta t) = k^*(t_0) \cdot \text{GHI}_{\text{clear}}(t_0 + \Delta t). \quad (5.6)$$

For minutely averages,  $k^*(t_0)$  is computed based on the last minute average. For evaluations of data based on a resolution of one second, the measurement (sample) at the time of the forecast initialization is used. Therefore, persistence forecasts have by definition no initial ( $t=0$ ) forecast error.

During the processing chain several assumptions and simplifications are made which contribute to final analysis and forecast errors. One error source is the irradiance retrieval (Sect. 4.6) based on binary cloud maps processed before. Particularly, cloud

irradiance enhancements due to reflections at cloud edges, irradiance reductions due to semi-transparent clouds and changes in diffuse irradiance levels due to a changing cloud distribution cannot be accurately addressed with the proposed methods. Therefore, we evaluate the ability of the forecast to distinguish between the two states (sunny and cloudy) by introducing a threshold of  $k^* = 0.8$ .

Accuracy ACC is used for measuring the ratio of the number of correctly predicted states (sunny and cloudy) by all instances:

$$\text{ACC} = \frac{\text{ts} + \text{tc}}{\text{ts} + \text{tc} + \text{fs} + \text{fc}}, \quad (5.7)$$

where  $\text{ts}$  = true sunny,  $\text{tc}$  = true cloudy,  $\text{fs}$  = false sunny and  $\text{fc}$  = false cloudy. For example, a forecast is true sunny, if measured and predicted  $k^*$  are  $> 0.8$ . A forecast is false sunny, if measured  $k^* > 0.8$  and predicted  $k^* \leq 0.8$ .

## 5.2 Handling of missing values

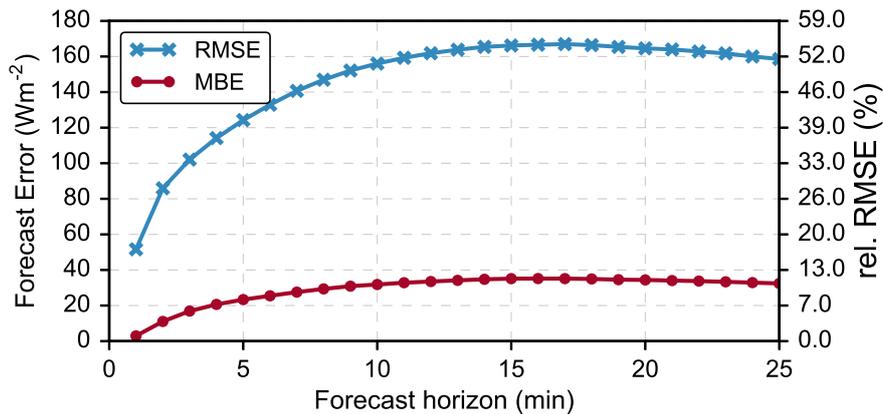
Important to note is the handling of missing values in the computation of error metrics. Due to the limited field of view of sky images, variable cloud height, cloud motion and solar position varying maximum forecast horizons can be reached (see the discussion on maximum forecast horizons in Sec. 4.9). Usually, user-specific maximum forecast horizons are chosen (e.g. 25 minutes) that may be reached only in a fraction (e.g. 50%) of all evaluated forecasts. When comparing error metrics for different forecast horizons, the computations are based on a different data basis (number of instances  $N$ ). As a result, the different availability of forecast instances for the performance evaluation must be considered in the conclusions. The following forecast and forecast evaluation strategy has been used in this study:

- Fixed maximum forecast horizons (temporal) are used in the forecast-generation phase.
- If the limitations in the field of view (image) are reached before the forecast horizon is reached, missing values (NaN) are used to fill up the full forecast instance.
- If no cloud motion is present, a persistence is used (constant forecast for the whole forecast horizon).
- Missing values are not considered in the computation of error metrics.
- No interpolation is used to fill up the forecast instances.
- Reference forecasts (e.g. persistence) are reduced to the same data basis to make both comparable.

- The maximum number of missing values in the evaluated dataset of size  $N$  is set to be 90 %. Otherwise, no error metrics are computed.

### 5.3 Single site forecast performance

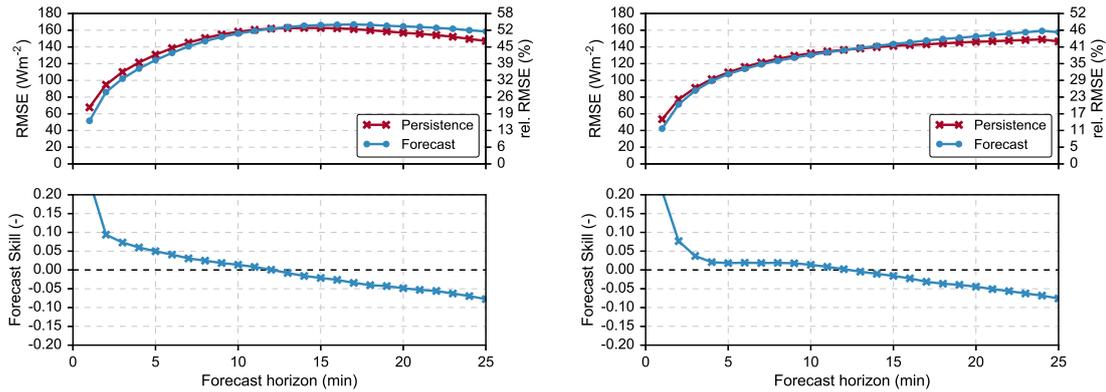
The single site local forecast performance for the Oldenburg data set is based on forecasts from January 2014 to November 2016. The minimum solar elevation has been set to  $10^\circ$  to generate forecasts throughout the year. As a result the maximum field of view has also been set to  $160^\circ$  (zenith angle of  $80^\circ$ ). With an update interval of 10 seconds a total number of 3,400,000 forecast instances are the basis of the following analyses. All forecasts and the corresponding measurements and reference forecasts have been averaged to 1 minute values for reasons of data handling. The maximum forecast horizon used is 25 minutes. The focus is put on forecasts of GHI, which is measured at the location of the camera. Comparisons with persistence are based on persistence of the actual clear sky index as described in the introduction of this section (Eq. 5.6).



**Figure 5.1:** GHI forecast error vs. forecast horizon for Oldenburg data. RMSE (blue) and MBE (red) are given in absolute and relative numbers. The relative errors are based on the mean measured GHI ( $\sim 310 \text{ Wm}^{-2}$ ).

Figure 5.1 shows the computed RMSE and MBE as a function of forecast lead time. RMSE shows an expected increase from  $\sim 50 \text{ Wm}^{-2}$  (first minute) to  $\sim 165 \text{ Wm}^{-2}$  for minute 15. In relative numbers this corresponds to about 53 % rel. RMSE. MBE also increases from the first minute (almost no bias) to minute 15, indicating a slightly underestimated irradiance. The low initial RMSE and MBE illustrate the circumsolar area correction and bias correction (introduced in Sec. 4.7), which are used to handle the circumsolar area problem when predicting the occlusion of the sun. The positive MBE for increased forecast horizons can probably be referred to overestimations of cloud coverage that in turn partly results from perspective errors at large viewing angles (see Sec. 4.5.4 on a discussion on perspective errors).

Figure 5.2 shows the forecast error in comparison to persistence. An overall positive forecast skill can be noted for the first 11 minutes starting with a forecast skill of more than 20 % for forecast horizon  $t=1$  but decreasing with forecast lead time. This reflects the good forecast performance in the first minutes. The persistence forecast error, here computed based on the previous minute average measurements, increases fast in the first minutes (already  $65 \text{ Wm}^{-2}$  at horizon  $t=1$ ). From minute 12 on, the persistence forecast error becomes lower compared to the sky imager based forecast and the FS negative.

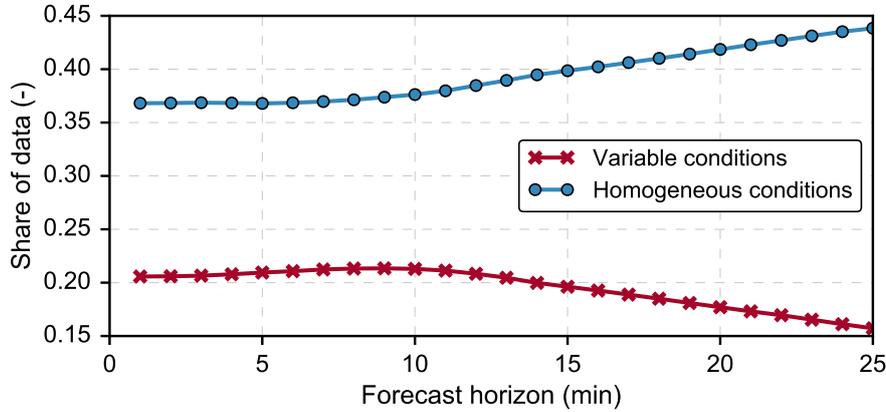


**Figure 5.2:** GHI forecast error (RMSE) vs. forecast horizon for Oldenburg data (top) and corresponding forecast skill (FS) in the bottom. Evaluation on the whole data set (left) and for a reduced dataset including only forecasts reaching a minimum forecast horizon of 24 minutes (right).

Two questions came up when studying these results. First, why is the RMSE of both forecasts decreasing from minute 15 on? Second, why is the forecast skill decreasing with time?

The first question can be answered by considering the data basis used for the computation of the error metrics (see Sec. 5.2 on treatment of missing values). Usually, the forecast error of persistence predicting constant values should increase with time, which is not shown by the validation results in Fig. 5.2 (left). Figure 5.2 (right) shows the forecast validation based on a reduced data set. It only includes forecasts, which reach a minimum forecast horizon of 24 minutes. In this case, the total number of forecasts considered has been reduced by 50 %. Here, RMSE increases as expected until minute 23 and decreases afterwards. This leads to the conclusion, that the reduced number of forecasts reaching the longer forecast horizons typically have lower forecast errors. The lower overall RMSE ( $\sim 130 \text{ Wm}^{-2}$  compared to  $\sim 155 \text{ Wm}^{-2}$  for horizon  $t=10$ ) for the reduced data set strengthens this theory. The physical explanation is that situations with low cloud speed or even no cloud motion (clear sky) leading to large or maximum forecast horizons, is expected to have lower forecast errors. Furthermore, Figure 5.3 shows the share of data compared to the overall number of non-missing data points, in which the conditions are classified as predominantly variable (cloud types Cu, Ac/Cc, Sc) or homogeneous (cloud types clear sky, St/As, Ci/Cs) in dependency of the forecast horizon. It can be seen, that the share of homogeneous conditions (associated with

lower forecast error) increases from minute 10 on, while the share of variable conditions decreases. This findings give an explanation for the decreasing RMSE with forecast lead time. The forecast performance for the single cloud types is presented and discussed more detailed later in Section 5.6.



**Figure 5.3:** Share of data vs. forecast horizon in the evaluated dataset considering conditions classified (>70% probability) only as variable (cloud classes Cu, Ac/Cc and Sc) and homogeneous (clear sky, St/As, Ci/Cs) as a fraction of all available data points (i.e. without missing values due to reached forecast horizon).

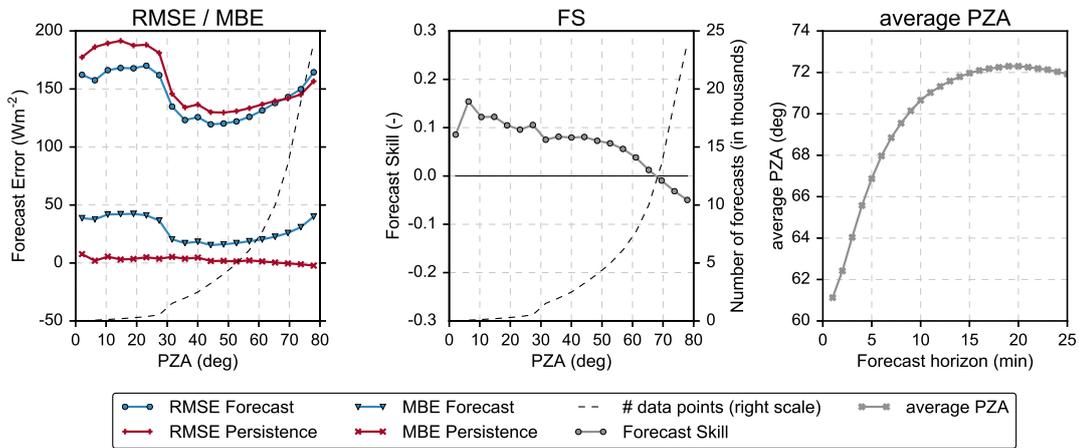
To answer the second question, the characteristics of persistence forecasts and camera-based forecasts have to be considered. Three main influencing factors have been determined:

- The share of homogeneous (low variability) and heterogeneous (high variability) cloud conditions,
- the considered field of view (in the image) and the predicted pixel-zenith-angle (PZA),
- small-scale variability and double penalties.

While for longer forecast horizons the share of less variable situations has a higher share in the database (see discussion above), the forecast skill in these situations also dominates the overall skill. In Section 5.6 and Section 5.7 forecast performance in different sky conditions is discussed. There, the overall negative skill in low variable conditions is demonstrated.

Another effect results from the different viewing angles under which a specific predicted pixel is seen. It can be expected, that generally the pixels of interest at larger forecast horizons are seen under a large viewing angle. The errors introduced with large viewing angles are discussed in Sec. 4.5.4. An evaluation of forecast performance in dependency of the pixel-zenith-angle (PZA), under which the pixel is seen in the camera image, is shown in Figure 5.4. Forecast data has been sorted by bins of PZA and error metrics

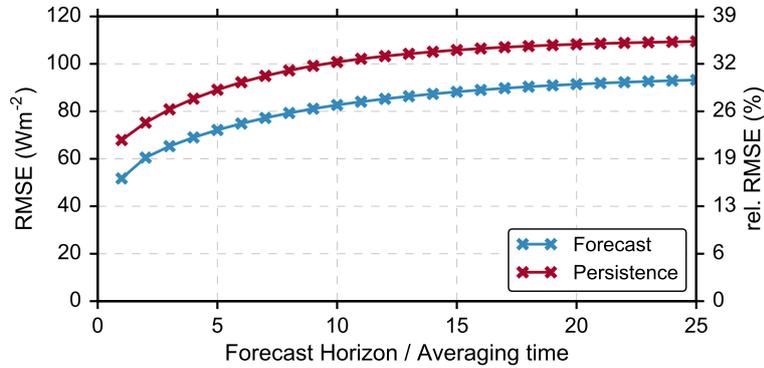
have been averaged for each bin. The graphs demonstrate the increasing forecast error as well as decreasing forecast skill with increasing PZA. The perspective projection of the fisheye image, which is a function of the tangens of the viewing angle (Eq. 4.10), introduces a large perspective error beginning at a PZA or viewing angle of about  $60^\circ$ . This can probably explain the strong reduction in forecast skill for larger viewing angles. Moreover, an increase in mean bias error MBE from that angle on can be observed. This probably indicates an overestimation of cloud cover due to these perspective problems. For the same reason of perspective projection, the number of datapoints in the forecast data set which were seen under high viewing angles are in turn dominating the overall database and therefore influencing the overall forecast error. This finding highlights the strong dependency of forecast error metrics on the chosen maximum image field of view. The graph in the right of Fig. 5.4 confirms the previously mentioned expectation, that for larger forecast horizons the average PZA increases as well.



**Figure 5.4:** Forecast error (RMSE, MBE) and skill (FS) as a function of the pixel zenith angle PZA. The average PZA per forecast horizon is given in the right graph. The number of data points used to compute the error metrics in each bin is given as a dashed line.

The third influencing factor, double penalties on forecasts with slight timing errors or spatial mismatches of small-scale variability have to be considered. Double penalties, well-known from evaluation of precipitation forecasts [100], penalize forecast errors twice. For example, if a single cloud shading event is predicted with a timing or spatial mislocation error, it will count twice in the error metrics. Once for the erroneous forecast of the cloud event and another time for missing the event. While double penalties occur for all forecast horizons, they are assumed to be larger at longer forecast horizons, when particularly uncertainties in cloud motion and cloud development affect the precise timing of small-scale cloud events. The influence of double penalties on the computed forecast error can be estimated if forecasts are averaged in space or time. To investigate the influence, we compute the temporal average GHI from the first minute to an increasing forecast horizon. It can be expected, that the temporal averaging would lead to lower errors compared to persistence, which does not account for variability

in the definition used. Figure 5.5 demonstrates the averaging effect. For averaged forecasts, an overall lower forecast RMSE compared to persistence can be achieved for all averaging times with little relative change at longer forecast horizons indicating the positive influence of averaging on the evolution of forecast error with forecast lead time.



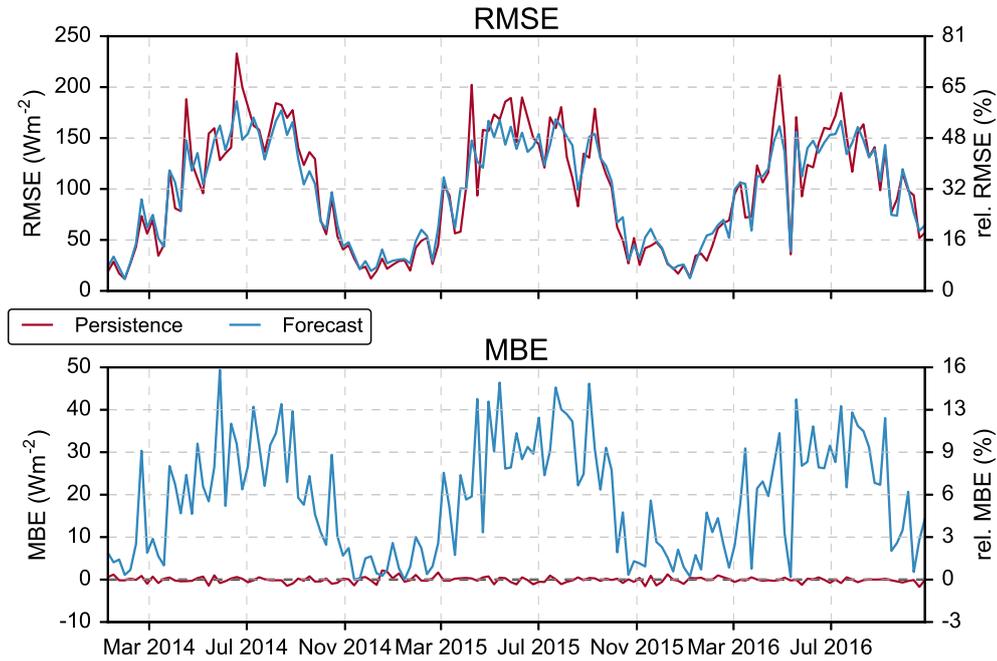
**Figure 5.5:** Forecast error (RMSE) for increasing averaging times.

## 5.4 Seasonal and diurnal performance

The temporal course of forecast errors is shown in Figure 5.6. RMSE and MBE are computed for weekly 5 minute ahead forecasts and presented for both sky imager based forecasts and persistence forecasts. The seasonal shape in absolute error values can clearly be identified. RMSE ranges from  $20 \text{ Wm}^{-2}$  in winter times to  $\sim 150 \text{ Wm}^{-2}$  in summer when highest irradiance values are measured. A seasonal trend of negative forecast skill in winter times and positive skill in summer times can be observed. Especially, in weeks with high persistence forecast RMSE indicating high short-term irradiance variability the forecast RMSE is several  $\text{Wm}^{-2}$  lower than persistence RMSE.

In order to demonstrate the diurnal shape of the forecast performance and to expose dependencies on day time, the forecast performance has been evaluated depending on the solar elevation angle. Figure 5.7 shows the error metrics of sky imager based forecasts and persistence forecasts as a function of solar elevation angle. GHI has been normalized to clear sky index  $k^*$  to make the error metrics independent from the diurnal course. The minimum solar elevation of  $10^\circ$  is the lower limit for which forecasts have been processed. The maximum solar elevation angle of about  $60^\circ$  is the maximum solar elevation angle reached in Oldenburg.

First, an overall increase in forecast errors (RMSE) from 0.1 to 0.2 with solar elevation is present. MBE increases as well for sky imager based forecasts. The increase in RMSE is assumed to be connected to some extent with higher irradiance variability (shown in the right graph) for larger solar elevations. To verify this assumption, variability has



**Figure 5.6:** Forecast errors (RMSE and MBE) as a time series from 2014 to 2016. Error metrics are computed per week for forecasts with a horizon of  $t=5$  minutes. Sky imager based forecasts is drawn in blue and persistence in red. Relative errors are computed based on the average GHI measured in the evaluated time period.

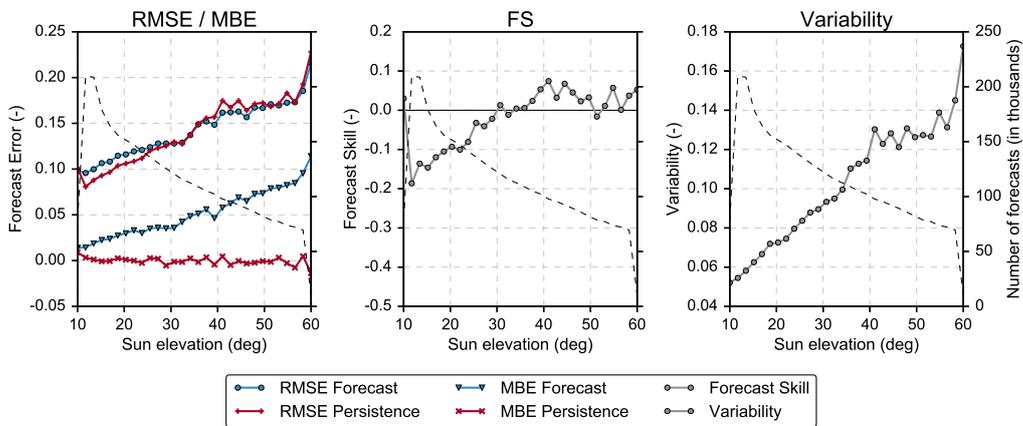
been computed as the standard deviation of the clear sky index in the next 15 minutes according to Eq. 5.8.

$$\sigma_{k^*} = \sqrt{\frac{1}{14} \sum_{i=1}^{15} (k_i^* - \bar{k}^*)^2}. \quad (5.8)$$

Large solar elevations correlate in turn with more formation of convective, scattered clouds. For Oldenburg, the presence of small-scale clouds is much more likely in summer times (Fig. 4.21). Correlations of solar elevation with other factors influencing the forecast performance are likely, but challenging to separate. For example, image quality is affected strongly by dew droplets on the camera dome present in morning hours (and persists longer in winter times). Low solar elevations also influence cloud detection strongly by the bright circumsolar area (due to the long sun path through the atmosphere) when the sun is not occluded by clouds. The larger viewing angles for clouds moving towards the sunspot also influence the performance due to introduced perspective errors and reduced image resolution as mentioned before. However, these effects have negative impact on forecast performance for low solar elevations, which is in contrast to the forecast errors introduced by irradiance variability. The forecast skill FS, shown in the middle graph, shows a negative forecast skill for low solar elevation ( $<30^\circ$ ) and positive forecast skills for larger solar elevations. This demonstrates the

overall better forecast performance compared to persistence at larger solar elevations which are less affected by the impacts listed before and in times that are characterized by more solar variability.

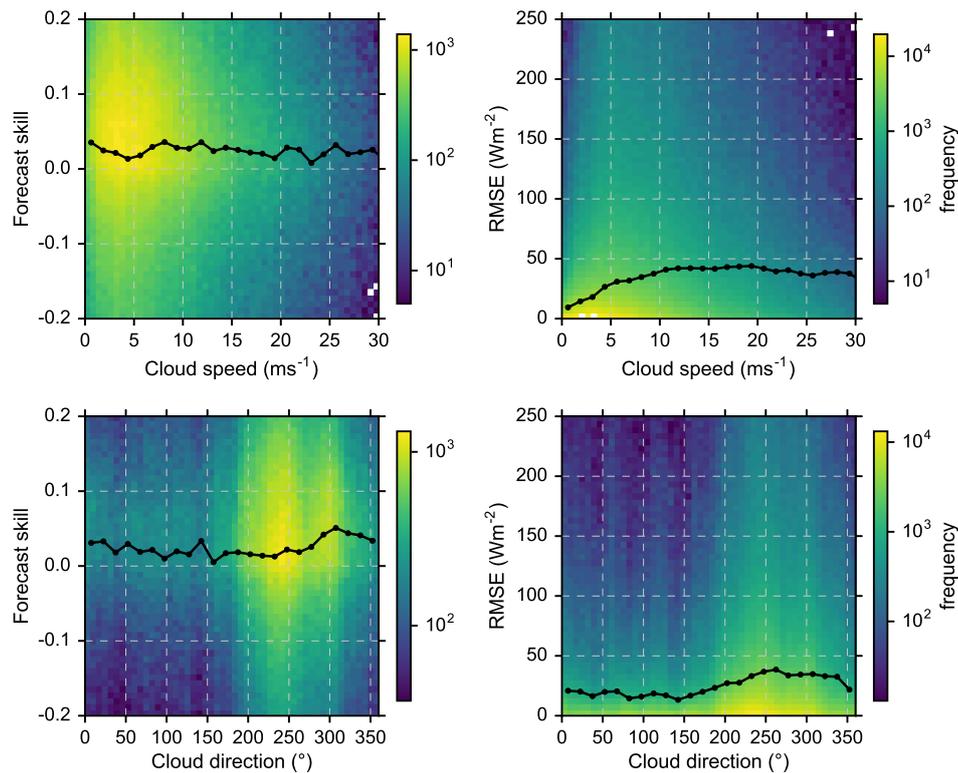
If forecast skill, which here is the comparison with persistence, is used to quantify the performance of a sky imager based forecast model, the dependency on solar elevation angles must be named. The results show, that limiting forecasts to higher solar elevations will increase the overall forecast skill. The much higher share of low solar elevations during the year underlines the influence on the overall forecast skill of the model. In other words and as already stated in the section before, the forecast skill depends on the data basis used for the evaluation.



**Figure 5.7:** Forecast errors (RMSE, MBE), forecast skill (middle graph) and solar variability (right graph) as a function of solar elevation angle. Error metrics are computed for a forecast horizon  $t=5$  minutes and averaged for 30 equidistant bins from the minimum to the maximum solar elevation. All numbers are computed based on clear sky indices to remove the diurnal course of GHI. The number of data points used to compute the error metrics in each bin is given as a dashed line.

## 5.5 Cloud motion dependency

The influence of the prevailing cloud motion on forecast performance is summarized in Fig. 5.8. Cloud speed and direction for each forecast have been obtained from the cloud motion algorithm. Cloud speed is the result of cloud tracking and projection using cloud base height CBH. Considering uncertainties in CBH data recorded in 10 minute intervals at a 40 km nearby weather station, uncertainties in cloud speed, especially in multiple layer cloud conditions are likely. The analysis shows a positive median forecast skill for all cloud speeds up to  $30 \text{ ms}^{-2}$ , but reveals no dependency with cloud speed. The forecast RMSE is increasing with cloud speed probably due to higher variability associated to faster moving clouds. For the cloud direction an increased RMSE, but also slightly enhanced FS can be noticed for the predominant, western to north-western cloud direction.

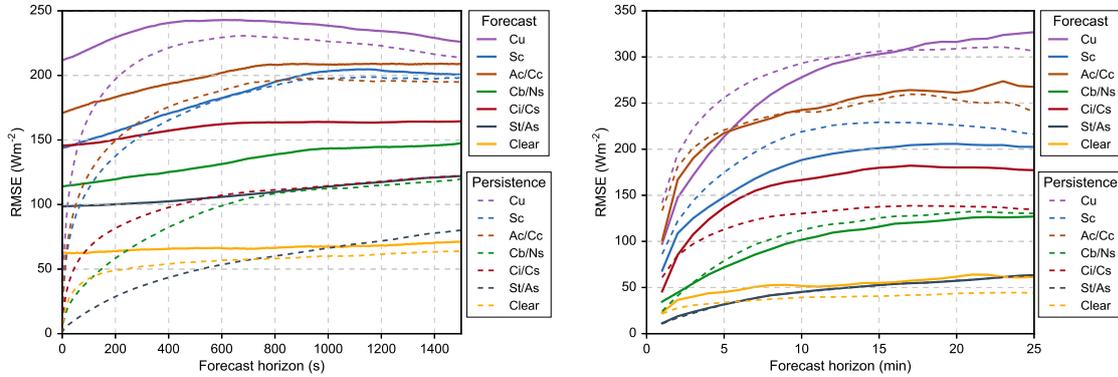


**Figure 5.8:** 2D-histogram (50 bins per axis) showing forecast skill (FS) and error (RMSE) in dependency of cloud speed and direction. The black line represents the median value of each bin. Data basis: 5 minute ahead forecasts and cloud motion processed for Oldenburg for the years 2014 and 2015.

## 5.6 Cloud type dependency

By reason of the varying complexity of present sky or cloud formations the differentiation of forecast performance in dependency of predominant sky or cloud type provide additional knowledge about the model skill. For the analyses, the cloud type classification scheme (Sec. 4.8.2) has been used to determine the predominant cloud type for each image or forecast, respectively. The dataset of HOPE and University of Oldenburg are used to demonstrate the different forecast performance in dependency of cloud type. In Schmidt et al. [40], we found for the HOPE dataset best forecast skill for cumulus-type clouds (stratocumulus, cumulus, altocumulus) and worst skill for stratus-type clouds (stratus, altostratus, cirrostratus) compared to the persistence forecast. It was argued, that stratus-type clouds cause less variability than cumulus-type clouds resulting in lower absolute error, but less forecast skill due to lower persistence errors. In addition, it was argued, that the binary cloud detection of the model does not account for semi-transparent clouds, especially cirrus or cirrostratus clouds. It misclassifies them always as blue sky or opaque cloud (see Sec. 4.3 for the description and discussion of cloud detection). Figure 5.9 shows the results for the HOPE [40] and

for the Oldenburg dataset. While the HOPE dataset only covers two months (April and Mai), two years of forecasts have been evaluated for Oldenburg. Moreover, the analysis comprises forecasts for 99 distributed locations for HOPE and only the camera location for Oldenburg. First, large differences in forecast error for the seven different



**Figure 5.9:** Forecast error (RMSE) for the different cloud types determined by the cloud type classification scheme. Results are shown for HOPE dataset (left) [40] and Oldenburg dataset (right).

cloud types are obvious. The persistence error reflects the variability caused by the predominant cloud type. The cumulus type clouds show the highest forecast errors (RMSE), while less variable situations (stratus-type clouds and clear sky) have lowest forecast errors. For HOPE and for Oldenburg data, the highest FS is achieved for stratocumulus (Sc) clouds. For Oldenburg, in predominant stratocumulus conditions a clear positive forecast skill can be observed for the whole forecast horizon. Here, also the other cumulus-type clouds Cu and Ac/Cc show good forecast skills. For Cu only, the positive FS is most emphasized in the first 15 minutes but is becoming negative for longer horizons. The higher FS for Sc compared to Cu cloud type can be partly explained by higher forecast errors for small-scale scattered cumulus clouds, which typically show more dynamics in the development and have more vertical extension compared to their horizontal dimension that in turn lead to projection errors when assuming a single flat cloud layer. As can be expected, the forecast errors of persistence and the sky imager based forecast differ less in the low variable St/As and Cb/Ns cloud conditions and also in clear sky conditions. Analogue to HOPE, the worst FS for Oldenburg is achieved in conditions with predominately semi-transparent Ci/Cs clouds. Model improvements accounting for the optical properties of these clouds are needed to gain higher forecast skill for these clouds.

Furthermore, the higher initial forecast errors for HOPE are noticeable. These can be referred to the distributed locations of the pyranometers, for which forecasts have been processed and included in the validation. Here, the initial forecast error reflects the error introduced by erroneous cloud detection and cloud shadow mapping. Moreover, a newer version of the model that includes the postprocessing described in the model section has been used for Oldenburg, where local single site sun occlusion forecasts benefit from bias correction and the cirumsolar correction leading to very low initial errors.

Overall, the results for the Oldenburg data set coincide with the findings from Schmidt et al. [40] for HOPE data.

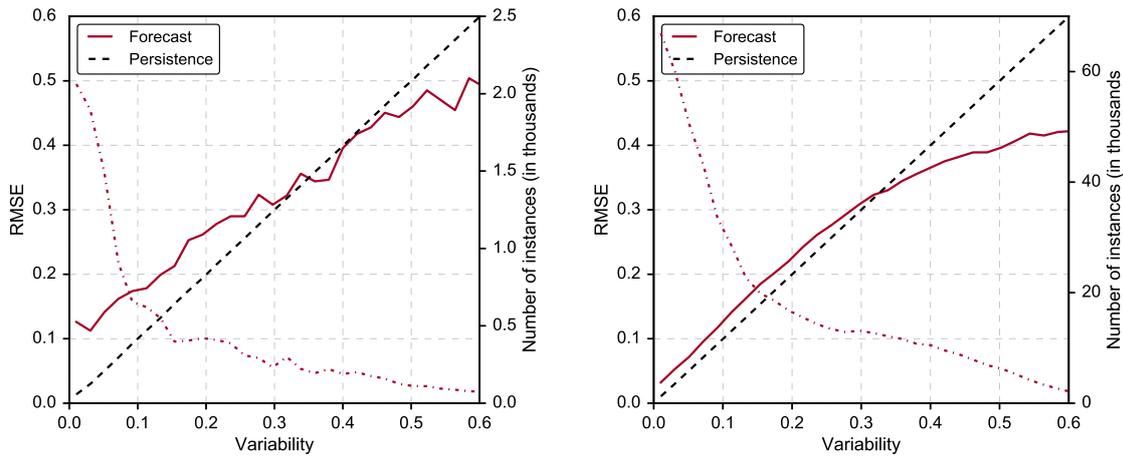
## 5.7 Variability and forecast performance

In the previous sections a strong dependency of forecast performance on prevailing sky conditions has been highlighted. A more general approach for differentiation is the investigation of forecast performance on prevailing irradiance variability based on pyranometer measurements. Marquez and Coimbra [101] proposed a variability measure as the standard deviation of clear sky index increments or step-changes  $\Delta k^*(t_i)$  (increment size  $\Delta t$  equal to a fixed forecast horizon) in a defined period of time ( $N$  timesteps), which is given as

$$V = \sqrt{\frac{1}{N} \sum_{i=1}^N (k^*(t_i + \Delta t) - k^*(t_i))^2} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\Delta k^*(t_i))^2}. \quad (5.9)$$

By this definition, when plotting measured variability against the persistence forecast RMSE (computed based on clear sky index), all points fall on the diagonal line.

Figure 5.10 presents the forecast RMSE in dependency of this definition of measured variability for one pyranometer station of the HOPE dataset [40] (left) and for Oldenburg data (right) for a forecast horizon  $\Delta t=10$  minutes. RMSE and variability are calculated for short moving time windows of 25 minutes each. The time step between two time windows is 1 min, resulting in an overlapping database. The graphs in Fig. 5.10 represent the average values of each bin. Obviously, the forecast error increases with increasing variability. In situations of low  $k^*$ -variability also negative forecast skill (forecast RMSE higher than persistence RMSE) can be found. An overall lower RMSE for Oldenburg can be stated for these low variable situations that can be referred (similar to previous section) to the improved model and real time bias correction that has been used for the local single site forecasts. These results also match well with the differentiation in cloud types, where homogeneous sky conditions like stratus clouds or clear sky show low forecast skill. With increasing variability the forecast skills increases, too. The high irradiance variability is caused by scattered or broken clouds like cumulus or stratocumulus. Here, we find a “break-even” value between 0.3 and 0.4  $k^*$ -variability in both graphs, where average forecast skill becomes positive. Again, this indicates the strength of the model in high variable sky conditions. Nevertheless, like mentioned before for solar elevation and pixel-zenith-angle, these situations are much less frequent (see dash-dotted line in Fig. 5.10) and therefore hiding the skill in the average error statistics.



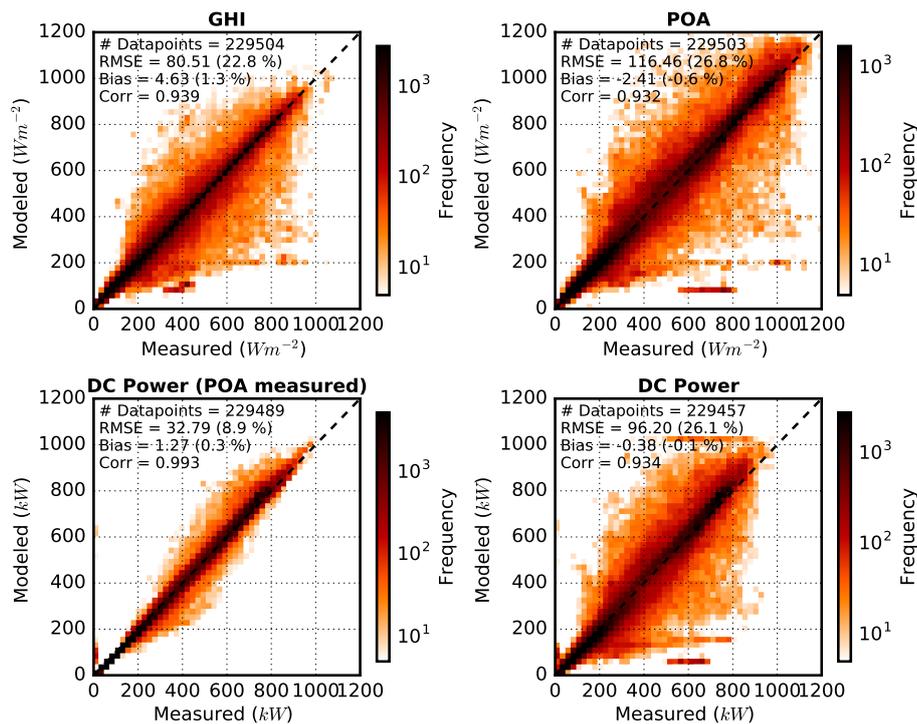
**Figure 5.10:** Forecast error (RMSE) vs. variability for pyranometer 33 located close to the camera (solid line). Persistence error (dashed line) is marked on the diagonal. The number of instances averaged in each bin with size  $k^* = 0.02$  (dash-dotted line) is given on the right y axis to illustrate the robustness [40].

## 5.8 Performance for a 1-MW PV power plant

*SkySol* has been applied to a data set of sky images acquired at a 1 MW PV plant to demonstrate the applicability for utility scale PV plants [18] (see Sec. 3.3 for data description). For this demonstration case, an earlier version of *SkySol* has been used to process the forecasts for the PV-plant. In particular, the cloud detection scheme and the irradiance modelling scheme have been improved afterwards. Since no CBH estimations were available, local point forecasts have been computed to be representative for the whole PV plant. Minutely averages of PV power, module temperatures, GHI and plane-of-array (POA) irradiance have been recorded at the plant and used for the forecast validation. This section presents the results of maximum 15 minute ahead forecasts on a three month data set.

### 5.8.1 Irradiance and PV power nowcast results

The term nowcasting is used here explicitly for the forecast horizon  $t=1$  minute. The scatterplots in Figure 5.11 shows modelled versus observed GHI, plane-of-array (POA) irradiance, DC power using POA irradiance from measurements and DC power using modelled POA irradiance. For all parameters, low systematic errors (MBE) are observable. For the basic product of the forecast model, GHI, a relative RMSE of 22.8% has been computed for the first minute. The conversion to POA irradiance increases rel. RMSE to about 26%. Similar error metrics can be observed for the final product, simulated DC Power. This analyses should give an impression on the error propagation in the PV simulation chain. The scatterplot for the simulated DC



**Figure 5.11:** Comparison of modelled and measured parameters for forecast horizon  $t=1$  minute: Global horizontal irradiance (GHI, top left), conversion to plane of array (POA, top right), simulated DC power from measured POA irradiance and module temperature (bottom left) and simulated DC power using modelled POA (bottom right). Temporal resolution: one minute averages. Relative errors are computed based on average measurements.

power based on measured POA irradiance and module temperature identifies the error introduced by the PV simulation clearly (rel. RMSE  $\sim 9\%$ ).

### 5.8.2 Forecast results

The forecast results are presented in table 5.1 showing RMSE for forecast horizons up to 15 minutes. Additionally, persistence forecasts are added to show the skill compared to a reference forecast. Obviously, persistence error stays below model forecast error for all horizons (32.8% vs. 40.3% for minute 15).

In Schmidt et al. [18], more detailed analyses of forecast performance regarding solar variability and binarized forecasts have been presented. The concept of binarized forecasts is presented in the next section.

**Table 5.1:** Comparison of absolute (top) and relative (bottom) RMSE

|          |                   | $GHI [Wm^2]$ | $POA [Wm^2]$ | $DC [kW]$<br>Persistence | $DC [kW]$<br>Forecast |
|----------|-------------------|--------------|--------------|--------------------------|-----------------------|
| absolute | <b>Nowcast</b>    | 80.5         | 116.5        | 0.0                      | 96.2                  |
|          | <b>5 minutes</b>  | 118.2        | 163.1        | 34.4                     | 145.5                 |
|          | <b>10 minutes</b> | 138.8        | 189.1        | 119.2                    | 155.1                 |
|          | <b>15 minutes</b> | 139.3        | 189.9        | 127.4                    | 156.6                 |
| relative | <b>Nowcast</b>    | 22.8         | 26.88        | 0.0                      | 26.1                  |
|          | <b>5 minutes</b>  | 33.0         | 37.1         | 24.4                     | 36.0                  |
|          | <b>10 minutes</b> | 37.7         | 41.6         | 30.9                     | 40.3                  |
|          | <b>15 minutes</b> | 37.6         | 41.4         | 32.8                     | 40.3                  |

## 5.9 Forecasting sunny or cloudy sky - a binary approach

For some applications, a precise forecast of the irradiance time series for the next minutes is not necessary. Here, predicting if parts of the sky will be cloudy or not can provide sufficient information. One application is the prediction of the end of a clear sky period or the transition to cloudy or variable conditions, respectively. Even if persistence models or more sophisticated statistical models that take into account the measured time series show good performance in clear sky conditions or in the first minutes of the forecast horizon, they will not be able to predict these transitions based on the past time series alone.

Sky imager based forecasts are optimal for this purpose, since they can monitor the clear sky and give a warning if clouds are forming or moving in the field of view. Furthermore, against the background of a simplified forecast model, which lacks differentiation in cloud detection (binary segmentation) and cloud shadow projection (single cloud layer assumption, uncertainties in cloud base height), forecast performance will benefit from simplification in the forecast product. Reducing complexity with binarization (e.g. cloudy or sunny) of predicted irradiance can also be used to estimate the forecast error introduced by uncertainties in the irradiance modelling scheme. The weaknesses of the irradiance model result from the transformation of binary information from cloud (shadow) maps to irradiance values. It is unable to represent deviations from this binary assumption, e.g. cloud enhancements. The binary forecast is therefore the better representation of the binary cloud detection scheme.

Two applications of binary forecasts and their evaluation are presented here.

### 5.9.1 Clear sky index threshold

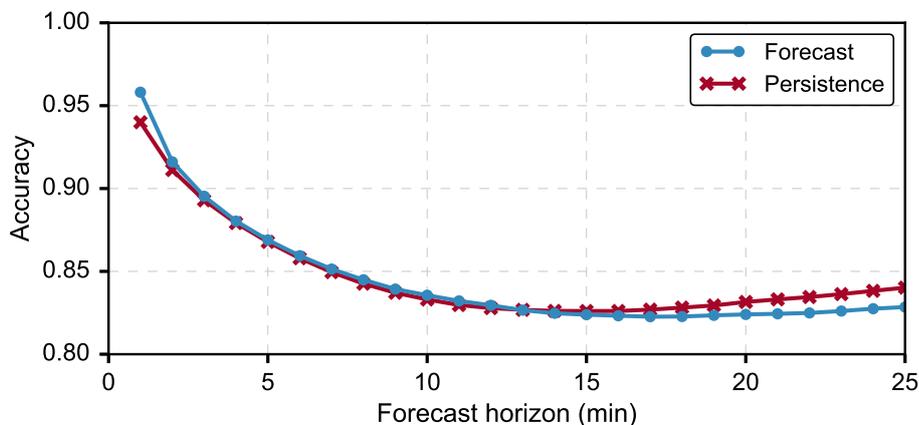
Binarizing irradiance time series is done on clear sky indices to remove diurnal and seasonal variations. A threshold of  $k^* = 0.8$  can be applied to discriminate the time

series into sunny (high clear sky index) and cloudy conditions (low clear sky index). Accuracy (ACC, Eq. 5.7) can be used to determine the forecast performance based on binary values. It measures the number of correct forecasts divided by all forecast instances. In Schmidt et al. [40], the forecast performance has been evaluated for the HOPE dataset, for which binarized forecasts with a resolution of 1 second have been computed. It was found, that an accuracy  $ACC > 90\%$  (more than 90% correct forecasts of binary sky conditions) could be achieved for homogeneous sky conditions like clear sky or overcast. Lowest ACC of 70% are achieved for cumulus (Cu), altocumulus (Ac/Cc) and cirrus (Ci/Cs) clouds. In Schmidt et al. [18], we analysed an ACC of 88% for the whole Munich dataset and for a 10 minute forecast horizon. Figure 5.12 shows the results for Oldenburg (including the whole dataset). Here, the accuracy of predicting the correct sky conditions is above 90% for the first minutes and slightly lower (83%) compared to Munich data for the 10 minute horizon. For the HOPE dataset, also an improvement over persistence forecasts compared to validation based on RMSE could be achieved for some cloud classes (not shown here). For Oldenburg, the improvement to absolute forecast errors (RMSE) is rather small (compare Fig. 5.12 with Fig. 5.2), indicating improved performance of the irradiance modelling scheme. Again, the numbers presented here are difficult to compare because of the differing underlying data sets (especially different seasons considered).

### 5.9.2 Predicting cloud events for PV-Diesel hybrid systems

#### Introduction

The study presented here is a summary of the work that has been done in collaboration with partners from Murdoch University, Perth and NEXT ENERGY, Oldenburg. The results were presented at the *World Renewable Energy Congress 2017* in Perth [38].

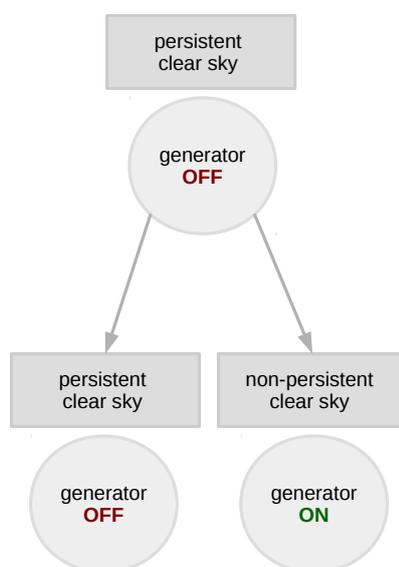


**Figure 5.12:** Accuracy metrics for the binarized forecasts in dependency of the forecast horizon. The clear sky index threshold used is  $k^* = 0.8$ .

We applied *SkySol* (latest version) to the data set of the Murdoch University in Perth, Western Australia (data description in Sec. 3.4). A geometrical lens calibration with several snapshots of chessboards has been applied in advance. The camera orientation has been determined on a clear sky day. A representative set of sky images has been used to test the model and to adjust some of the model parameters. Due to slightly different color settings of the camera compared to Oldenburg, the RBR threshold and the RBR correction parameters (Sec. 4.3) have been adapted to achieve an optimal cloud detection on the test images. The irradiance modelling is based on the histogram-method (Sec. 4.6.2).

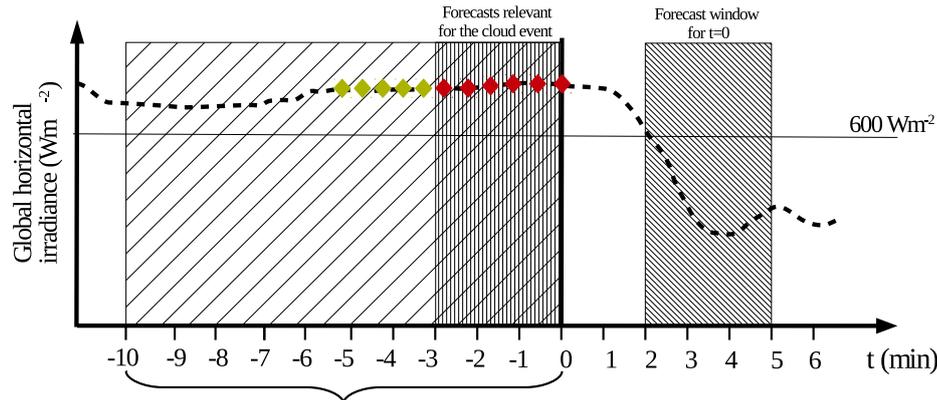
## Methodology

Focusing on PV integration in remote electricity systems with diesel generators, we evaluate the forecast performance regarding the ability of the forecast model to predict solar irradiance in the forthcoming 2-5 minutes, allowing for typical diesel generator starting times. The accurate prediction of changes in solar irradiance in the mentioned critical time window is of importance rather than the accurate prediction of irradiance at a specific point in time and space. Of particular interest is the question of how reliable is a forecast that there will be no cloud event in the time window of interest?



**Figure 5.13:** Diesel generator scheduling as used in the study

A simplified control strategy of a diesel generator with solar irradiance forecasts has been defined to demonstrate a potential application for the reduction of diesel consumption. We assume a diesel generator to remain off during constant clear sky conditions over the last 10 minutes with a minimum (10th percentile) insolation of  $600 \text{ Wm}^{-2}$  (which is about 55 % of local noon clear sky irradiance). The generator has to be switched on



**Figure 5.14:** Illustration of non-persistent clear sky conditions. If the irradiance (black dashed line) drops below the threshold at any time in the forthcoming 2-5 minutes, actual conditions are flagged as non-persistent. While instances at the green diamonds represent persistent conditions, red diamonds indicate non-persistence. The past 10 minutes are considered in the definition of a constant clear sky period.

when the conditions are about to change in the next 2-5 minutes (Fig. 5.13). Otherwise the generator stays off. We call this a persistent clear sky. A non-persistent clear sky means that clouds will shade the pyranometer (cloud event) and therefore end the clear sky period defined above. The sky conditions change to either mixed (fluctuating insolation) or overcast (down-ramp to homogeneous low insolation) in the next 2-5 minutes. The opposite case of an initial overcast or mixed sky is not examined in this study, since it is assumed that the generator will be anyway switched on (standby or under load). Therefore, we only investigate the reliability of the forecast to predict a persistent clear sky correctly or how well an erroneous forecast of a non-persistent clear sky can be avoided. The latter case could lead to grid stability issues if insufficient spinning reserve is available. The total number of constant clear sky conditions is identified by applying the  $600 \text{ Wm}^{-2}$  threshold to sliding windows of 10 minutes of the irradiance time series. Next, we do the same for solar irradiance forecasts and measurements for the horizon from 2 to 5 minutes. Figure 5.14 illustrates the method. Note, that forecasts are updated every 10 seconds resulting in 30 forecasts predicting the same event with different horizons. It must be emphasized, that at least one forecast should predict the irradiance drop two minutes in advance to guarantee the generator to have enough time to start. A confusion matrix lists the correct (true positive  $tp$  and true negative  $tn$ ) and false forecasts (false positive  $fp$  and false negative  $fn$ ) of the binary classifier with the positive and negative conditions of observed “persistent clear sky” and “non-persistent clear sky”, respectively.

- true positive ( $tp$ ): persistent clear sky was observed and predicted correctly.
- false positive ( $fp$ ): persistent clear sky was observed, but a cloud event was predicted.
- true negative ( $tn$ ): non-persistent clear sky was observed / cloud event happened

and predicted correctly.

- false negative (fn): non-persistent clear sky was observed / cloud event happened but persistent clear sky was predicted.

We call the share of correctly predicted conditions accuracy (Eq. 5.10) and the share of incorrectly predicted non-persistent clear sky conditions the miss rate (Eq. 5.11). The share of incorrect forecasts of persistent clear sky conditions is called false alarm rate (Eq. 5.12).

$$\text{accuracy} = \frac{tp + tn}{tp + tn + fp + fn} \quad (5.10)$$

$$\text{miss rate} = \frac{fn}{fn + tn} \quad (5.11)$$

$$\text{false alarm rate} = \frac{fp}{tp + fp} \quad (5.12)$$

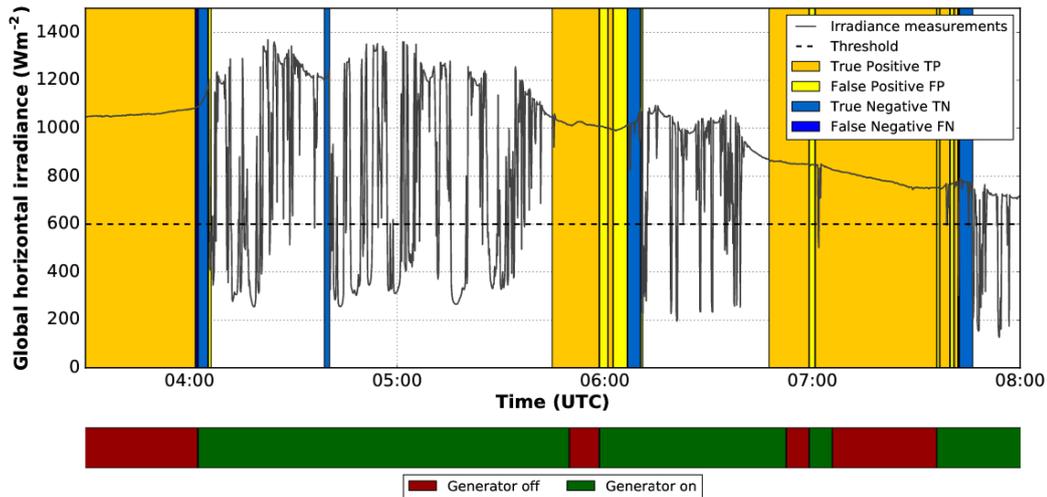
Both miss rate and false alarm rate should be minimized in order to achieve a high potential for reduction of spinning reserves (false alarm rate) and to reduce the probability of grid stability issues due to non-predicted reductions in PV generation (miss rate).

## Results

About half of the days from 10th November to 28th December 2016 were almost completely cloud-free leading to about 61% constant clear sky periods. The other days were dominated by mixed clouds (scattered or broken) moving from different directions with varying speeds. Several situations where constant clear sky conditions are replaced by fluctuating conditions could be identified from irradiance time series and images. This results in a total number of 84 cloud events indicating the ending of a constant clear sky period.

First, Fig. 5.15 shows an example period with five cloud events. The colour shaded area indicates the constant clear sky periods considered for potential generator scheduling and the forecast evaluation. While the yellow colour represents persistent conditions, blue indicates non-persistence due to upcoming clouds after a non-shaded period. The chosen generator scheduling is also given in this figure. In this simplified case, the generator is switched off, when clear sky lasted for at least five minutes and remains off when no cloud event is predicted (tp and fn). It can be noted, that all events in the

shown period could be predicted at least by the latest forecast two minutes in advance (light blue area or tn). The first event was not yet predicted five minutes beforehand (dark blue or fn), but in the subsequent forecasts.



**Figure 5.15:** Example results for the prediction of cloud events on a day with mixed conditions. The yellow area indicates the period defined as persistent clear sky. Blue colours indicate non-persistent conditions with cloud events in the next 2-5 minutes. When the periods end, the forecast system should reliably predict the change. The resulting generator state is shown as the result of past irradiance conditions and forecasts (bottom).

To summarize the results for the whole period, only the constant clear sky periods are evaluated as mentioned before. For these periods Table 5.2 shows the confusion matrix of the binary forecast results. Overall, in 98.2% of the time sky conditions persists clear in the next 2-5 minutes (tp+fp). For only about 1.8% of the time conditions change from clear sky to cloudy (tn+fn). As most of the persistent conditions with a low false alarm rate of only 3% are correctly predicted a high overall accuracy of almost 97% could be achieved. Nevertheless, focusing on the rare, but more important non-persistent clear sky conditions due to cloud events, we assert a miss rate of 28.2% (fn / (fp+tn)). The total number of 1892 forecast instances where the next 2-5 minutes observations are classified as non-persistent clear can be summarized in a total of 84 cloud events as mentioned above. While 76 events were predicted at least by one forecast in advance, still 8 events (~10%) were not foreseen. This is a slightly better result, but not yet satisfying.

Investigating the rare number of falsely predicted cloud events, the lessons to be learnt for improving the forecast model can be exposed. A number of problematic situations for non-predicted cloud events were identified:

- Clouds in the circumsolar area of the image are difficult to detect: Forecasts in the 2-5 minutes horizon are then less reliable, particularly when cloud speed is low.

|      |                      | Forecast            |                      |                 |
|------|----------------------|---------------------|----------------------|-----------------|
|      |                      | Persistent clear    | Non-persistent clear | Sum             |
| Meas | Persistent clear     | tp - 98506 (95.4 %) | fp - 2884 (2.8 %)    | 101390 (98.2 %) |
|      | Non-persistent clear | fn - 533 (0.5 %)    | tn - 1359 (1.3 %)    | 1892 (1.8 %)    |

**Table 5.2:** Confusion matrix with number of forecast instances (absolute and relative) for the evaluation of binary forecasting the persistence of clear sky periods. Note that only the constant clear sky periods (61 % of total) have been considered for the evaluation.

- Low sun elevation and cloud direction along the sun azimuth: When clouds are moving from the horizon towards the sun position forecasts are affected by:
  - Cloud motion is determined from clouds in the centre of the image and not at the border,
  - Large viewing angles lead to perspective errors affecting accuracy.
- Circumsolar area brightness is often enhanced for low sun elevations.
- Clouds move slightly along a different path as predicted by the analysed global cloud motion or clouds form or change size less than 2 minutes before blocking the sun.
- Some very short cloud events were related to single and/or small clouds not detected or not predicted based on a slightly wrong global motion vector.

Based on these findings, improvements for the model can be derived:

1. It is important to note, that local single site forecasts with high temporal resolution have been investigated. Spatial averaging with distributed PV generation will reduce the influence by small, single clouds and average out ramps under scattered cloud conditions.
2. Cloud motion vectors should be derived for single clouds to avoid false cloud tracking due to averaging. Moreover, the uncertainty of the motion vector algorithm should be used to provide confidence intervals rather than one deterministic motion vector.
3. At least for local site forecasting (forecasts for the location of camera), camera hardware with a higher dynamic range should be used to reduce misclassification in the circumsolar area.
4. Problematic conditions (e.g. low solar elevations but also rain drops, birds, dirt) affecting image quality and forecast accuracy should be detected and communicated operationally beforehand so that forecasts can be labelled as unreliable and

appropriate measures can be taken.

5. General improvements in cloud detection and tracking algorithms will enhance forecast performance.

## 5.10 Model validation for spatially distributed locations

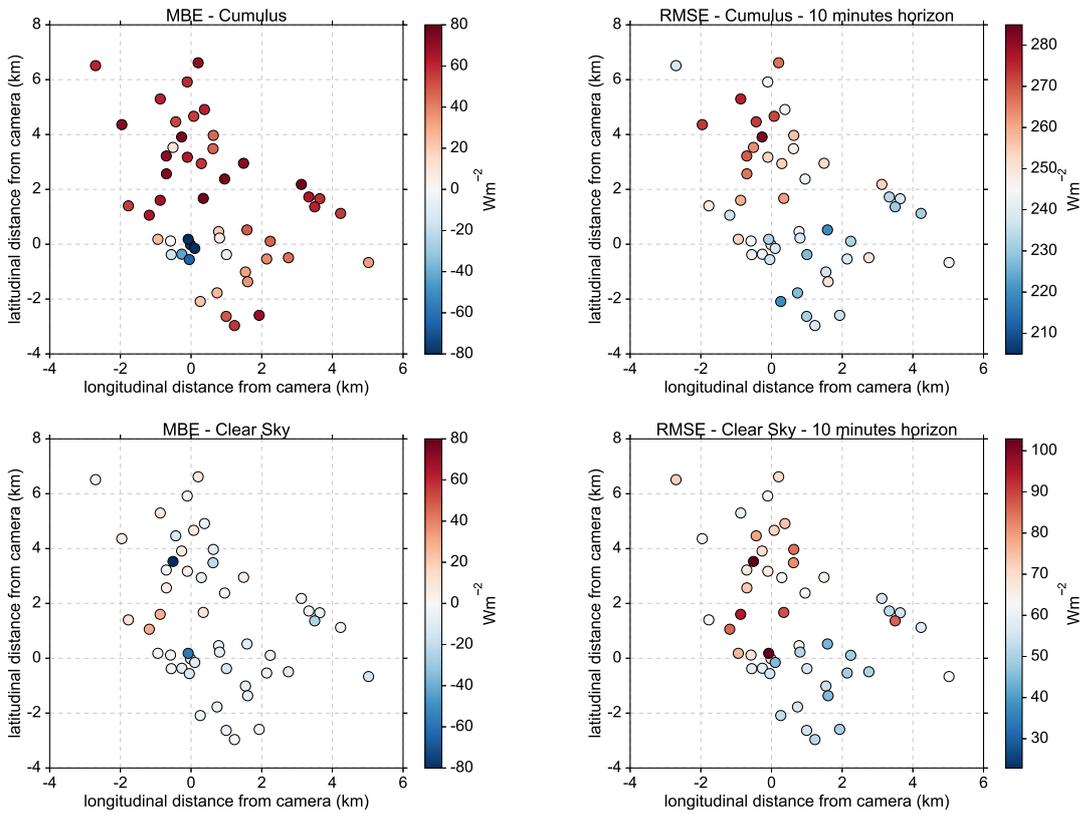
Finally, the model performance should be analysed focusing on locations distant from the camera location. The HOPE dataset (described in Sec. 3.2) provides images from one camera and pyranometer measurements at 99 locations in the field of view of the camera (maximum distance 8 kilometres). Moreover, local ceilometer measurements can be used for cloud base height estimations and a spatial projection for cloud shadows. The dataset is an optimal basis to study the error sources introduced by the viewing angle a distant cloud is seen from the camera.

Irradiance analysis (forecast horizon  $t=0$ ) and forecast performance is evaluated in dependence on the distance of the pyranometer stations from the camera and according to the different cloud type classes determined by the cloud type classification algorithm. Forecast time series are derived from the irradiance maps for each pyranometer station up to 25 minutes with a temporal resolution of one second. Here, the MBE and RMSE for the analysis (forecast horizon  $t=0$ ) and the RMSE for the 600 seconds or 10 minutes forecast is evaluated. A comprehensive analysis of the forecast performance is given by [40]. The data set is described in Sec. 3.2. In order to make the forecast validation for the different pyranometer stations with differing data availability comparable, the following evaluation is based on a reduced subset of 50 pyranometer stations that have a data availability of at least 70%.

The spatial distribution of the mean bias error MBE of the GHI analysis is shown in the left part of Fig. 5.16 for cumulus clouds (Cu) and clear sky situations. The MBE distribution for Cu shows a negative MBE of about  $-80 \text{ Wm}^{-2}$  for stations close to the camera increasing with distance to positive values around  $70 \text{ Wm}^{-2}$ . A similar overestimation for stations close to the camera can also be found for Ac/Cc, Sc and Ci/Cs (not shown here). This phenomenon is not found for St and Ns/Cb situations dominated by (dark) overcast sky not affected by the correction. Moreover, the clear sky MBE distribution in Fig. 5.16 shows, that the correction performs in average well in clear sky situations as no significant MBE for stations surrounding the camera is present. From the analyses of the cloud maps (not shown here) we found for HOPE, that often the whole circumsolar area which is characterized by saturated pixels when the sun is not occluded by clouds, has been corrected by the cloud detection scheme to be non-cloudy. In situations with thin and/or bright clouds the algorithm might misclassify the pixels as non-cloudy. The MBE map for clear sky conditions (no clear bias for stations close to the camera) and the results for Cu (overestimation) reflects this behaviour. For all other datasets (Oldenburg, Munich, Australia) investigated in this thesis, the RBR correction is reduced to avoid these kind of misclassifications

(clouds not detected) while tolerating misclassifications in case of clear sky.

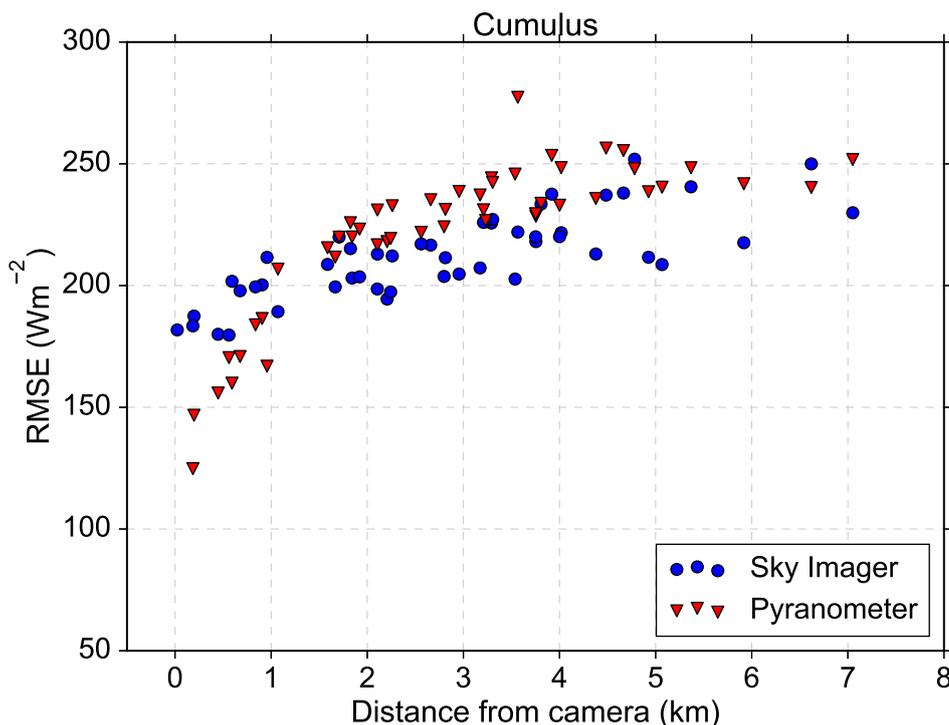
An increasing tendency in MBE with distance from the camera is also found for the aforementioned types Ac/Cc, Sc, Ci/Cs, while during clear sky or overcast stratus clouds (only clear sky is shown here) it is not present.



**Figure 5.16:** Analysis and forecast performance for each pyranometer station for cumulus cloud type (top) and clear sky conditions (bottom). MBE is computed for forecast horizon  $t=0$  minute (left column, from [40]) and RMSE for forecast horizon  $t=10$  minutes (right column).

Several possible explanations for these results can be identified. First, the perspective error increases with distance from the center of the image. As a result, convective clouds with vertical extent (mostly cumulus), which are interpreted as horizontally flat in our scheme, are projected incorrectly if they are seen from their side near the edge of the field of view. Among other things, this leads to an underestimation of gaps in the cloud layer contributing to a positive MBE and RMSE. Furthermore, uncertainties in cloud base height lead to higher errors in the shadow mapping the more distant the clouds are (can be derived from Eq. 4.17). Similar results have been found and discussed for the Oldenburg data set and longer forecast horizons when large distances to the camera location and perspective errors for large viewing angles are contributing to forecast error (Sec. 5.3). Moreover, cloud base height was measured at the position of the camera. Therefore, its representativeness for locations more distant is reduced depending on the cloud situation. This displacement of shadow patterns contributes

mainly to RMSE. As the temporal and spatial resolution of 1 Hz and 20 m, respectively, is quite high, double penalties in case of small cumulus or broken cloud layers are likely [100] and enhance RMSE even more (see discussion on double penalty effects in Sec. 5.3). Furthermore, the image resolution is reduced for larger lens incidence angles. This leads to a reduced spatial resolution for locations distant from the camera which affects the accuracy of the camera based irradiance analysis. This interpretation is in accordance to the considerations discussed in the general section about limitations in projecting cloud shadows (Sec. 4.5.4).



**Figure 5.17:** Root mean square error (RMSE) of sky imager analysis for each station and for cumulus cloud type (blue dots). The red triangles mark the RMSE of each station when compared to the station closest to the camera. This station has no error and therefore is not shown in this figure.

The previous considerations were based on irradiance analysis. Figure 5.16 also presents the spatial distribution of RMSE for a forecast horizon of 10 minutes. For both cumulus dominated and clear sky conditions, a north-west to south-east gradient in RMSE, with a maximum difference of about  $80 \text{ Wm}^{-2}$  can be identified. For the interpretation of the spatial RMSE gradient, shadow projection, image resolution and cloud motion (speed and direction) has to be taken into consideration. For north-western locations, it can be stated, that these locations are always affected by image parts close to the edge of the image. Against the background of a prevailing western cloud motion (typical for this region, data not given here), the forecast horizon is rather short for these locations and clouds predicted to shade the locations will be seen under an even larger viewing angle. Large forecast errors due to problems associated to large viewing angles are likely. For the central locations, the determining part of the image (in terms of cloud shading) is

the center. From the previous explanations and according to Fig. 5.16, these locations have the lowest RMSE in the nowcast. Yet another factor contributing to spatially different RMSE is the sun position. With the course of the sun, the location that is shaded by clouds in the centre of the image varies in time (see Sec. 4.5). Adding the westerly cloud motion, with a forecast horizon of 10 minutes, the locations in the south-east are predicted to be shaded by clouds that have been analysed in the center of the image.

Moreover, Fig. 5.17 shows the RMSE introduced, if a single pyranometer is used representatively for the whole area. It is assumed that the pyranometer closest to the camera is the reference sensor and RMSE of its measurements compared to the remaining pyranometers are calculated. As expected, the error increases very fast with distance as the cross-correlation between the sensor pairs is reduced especially in conditions with high GHI variability. It can be stated that the “break-even” distance where the sky imager based irradiance analysis outperforms a single sensor spatial extrapolation for this highly variable cloud conditions is found at a distance between 1 and 2 km from the camera. For other convective cloud types a distance of 2-3 km for Sc and Ac/Cc and 6 km for Ns/Cb is found. In case of St/As and Ci/Cs clouds and in clear sky conditions, the analysis error is always larger due to the high sensor pair correlation in these less variable situations.

## 6. Conclusions

This thesis presents a new very short-term solar forecasting model based on local sky images called *SkySol*. For the optimal operation and integration of high shares of solar energy generation capacities into electrical grids a need for very short-term local forecasts in the order of seconds to minutes ahead along with high update rates has been arisen. *SkySol* is a complete forecasting model that makes use of locally recorded high resolution sky images, onsite irradiance measurements and optionally cloud base height estimations to provide forecasts for spatially distributed locations in the vicinity of the camera. Using sky imager based forecast models allows for short-term cloud arrival and solar irradiance prediction with a spatial and temporal resolution that cannot be achieved with the traditional methods (numerical weather models, satellite images).

A comprehensive irradiance forecast performance analysis has been performed on the basis of multiple data sets. The forecasts have been validated with onsite pyranometer or PV data. For the purpose of comparison to reference forecasts, clear sky index persistence forecasts have been used to compute the forecast skill. A focus was put on forecast performance analyses in dependency of several influential parameters including the season and daytime (in terms of solar elevation), prevailing cloud motion and sky condition (cloud type), irradiance variability and the distance of a location to the camera. Moreover, forecast performance for a utility scale PV plant has been demonstrated and an approach for the usage in PV-Diesel hybrid systems has been presented and evaluated based on the basis of binarized GHI forecasts.

The models capabilities to predict short-term irradiance variability under different conditions and for different applications could be pointed out. Inherent model characteristics originating from the fisheye image characteristics as the main information source are discussed. The limited field of view and the perspective on the cloud layer seen from a ground-based camera together with varying cloud height, sun position and cloud motion leads to varying spatial coverage and therefore maximum achievable forecast horizon that have to be considered in the design phase of a camera setup. Varying image resolution with highest resolution in the image center and lowest to the border of the image in combination with perspective errors for distant clouds influences forecast performance depending on the image region of interest. Therefore, the understanding of the consequences of sky image characteristics can help to interpretate

forecast results or even provide additional uncertainty information when providing forecasts for stakeholders like plant or grid operators who take decisions based on forecast information.

Model characteristics originating from the algorithms and approximations used are also exposed and discussed. The core components comprise the extraction of cloud information from images, the prediction of its development and its interaction with solar irradiance. Several approximations in cloud detection, cloud motion and irradiance modelling are emphasized and discussed in the evaluation of forecast performance. In summary, a strong dependency of forecast performance on the prevailing sky conditions could be identified. A binary cloud detection scheme, that does not account for detailed optical properties of clouds might be sufficient in sky conditions that fulfil the approximation but leads to misclassifications in times of semi-transparent clouds, for example. Similar conclusions can be drawn for the approximation of all visible clouds being based in a single layer and moving with a uniform speed and direction. A more detailed cloud motion algorithm with detailed cloud height information taking single cloud motion into account or even modelling cloud transformations will certainly enhance the forecast accuracy in times of complex cloud conditions.

*SkySol* comprises two different approaches to determine surface solar irradiance. Image information and ground based measurements are used to convert the cloud or cloud shadow information to maps of irradiance. Since high resolution cloud-irradiance interactions are complex to model, statistical approaches using latest onsite irradiance measurements on the one hand and estimating irradiance from image features based on a beforehand trained machine-learning model on the other hand have been implemented and compared. These approaches convert the binary information from cloud or cloud shadow maps into binary irradiance maps aiming to optimize irradiance conditions in shaded and non-shaded areas. Given that irradiance distribution can deviate distinctly from this binary assumption by reason of different cloud optical properties and scattering in the atmosphere and clouds, forecast errors are introduced by this simplification. Significant error reductions can be expected from a more detailed irradiance modelling which might be based on improvements in cloud detection including cloud optical properties, implementing radiative transfer models or adding advanced statistical learning and post-processing.

Good forecast skill could be identified in highly variable situations. Even if forecast errors are unavoidable, the camera-based system makes use of the visible sky information and can predict cloud events or cloud gaps. In these situations statistical models based on time series analysis will fail due to the non-periodic nature of cloud coverage. On the other hand, traditional satellite image based forecast methods lack of resolution to predict the small scale cloud and irradiance variability. It was discussed that especially cloud event (e.g. ramps) timing errors due to spatial projection errors or uncertainties in cloud development and motion contribute to enhanced forecast errors when validating high resolution forecasts. On the other hand, in applications where timing can be neglected (e.g. only predicting 15 minute averages or variability), the camera-based model outperforms persistence.

In clear sky or overcast homogeneous sky conditions persistence forecasts typically show low forecast errors accompanied by low to negative forecast skill of the camera-based model. It was found, that these conditions often dominate the short-term horizon leading to an overall low model forecast skill when evaluating the forecast performance. In clear sky conditions, the moment of the transition to cloudy conditions is a critical parameter that have to be reliably predicted, if e.g. grid stability is affected in times of sudden drops in PV generation. A demonstration case study for a PV-Diesel hybrid system is presented showing a high accuracy for the prediction of these transitions. A number of suggestions for model improvements to enhance the reliability to predict these occurrences and to avoid missing a cloud event are proposed. These include spatial averaging to reduce timing or mislocating errors, single cloud tracking to reduce uncertainties in cloud motion, using high-dynamic range imaging for an improved cloud detection in the circumsolar area and a forecast uncertainty algorithm that provides information about the forecast reliability especially when image quality is affected by raindrops, birds, dirt or bright sunshine. As mentioned above, uncertainty information can also be retrieved based on the findings in this thesis about the influential parameters for varying forecast performance.

Beyond the scope of the present thesis is the incorporation of other data sources like satellite images, numerical weather prediction models and online data from distributed ground-based irradiance sensors or PV plants to extent the forecast horizon and to enhance forecast accuracy and reliability. Moreover, the usage of multiple sky imagers for the coverage of larger areas, the retrieval of 3D cloud information and the extension of forecast horizon has been discussed and simulated but not yet implemented in a field study.



## 6. Bibliography

- [1] IPCC. Summary for Policymakers. In T. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. Midgley, editors, *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, pages 1–30. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013. ISBN ISBN 978-1-107-66182-0. DOI: 10.1017/CBO9781107415324.004.
- [2] G. Team. GISS Surface Temperature Analysis (GISTEMP), 2017.
- [3] J. Hansen, R. Ruedy, M. Sato, and K. Lo. Global surface temperature change. *Reviews of Geophysics*, 48(4), 2010.
- [4] S. R. West, D. Rowe, S. Sayeef, and A. Berry. Short-term irradiance forecasting using skycams: Motivation and development. *Solar Energy*, 110(0):188 – 207, 2014.
- [5] C. W. Chow, B. Urquhart, M. Lave, A. Dominguez, J. Kleissl, J. Shields, and B. Washom. Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. *Solar Energy*, 85(11):2881 – 2893, 2011.
- [6] B. Urquhart, B. Kurtz, E. Dahlin, M. Ghonima, J. E. Shields, and J. Kleissl. Development of a sky imaging system for short-term solar power forecasting. *Atmospheric Measurement Techniques Discussions*, 7(5):4859–4907, 2014.
- [7] M. S. Ghonima, B. Urquhart, C. W. Chow, J. E. Shields, A. Cazorla, and J. Kleissl. A method for cloud detection and opacity classification based on ground based sky imagery. *Atmospheric Measurement Techniques*, 5(11):2881–2892, 2012.
- [8] D. A. Nguyen and J. Kleissl. Stereographic methods for cloud base height determination using two sky imagers. *Solar Energy*, 107:495–509, 2014.
- [9] G. Wang, B. Kurtz, and J. Kleissl. Cloud base height from sky imager and cloud speed sensor. *Solar Energy*, 131:208–221, 2016.
- [10] B. Urquhart, B. Kurtz, and J. Kleissl. Sky camera geometric calibration using solar observations. *Atmospheric Measurement Techniques*, 9(9):4279–4294, 2016.

- 
- [11] B. Kurtz and J. Kleissl. Measuring diffuse, direct, and global irradiance using a sky imager. *Solar Energy*, 2016.
- [12] H. Yang, B. Kurtz, D. Nguyen, B. Urquhart, C. W. Chow, M. Ghonima, and J. Kleissl. Solar irradiance forecasting using a ground-based sky imager developed at {UC} San Diego. *Solar Energy*, 103(0):502 – 524, 2014.
- [13] M. Li, Y. Chu, H. Pedro, and C. Coimbra. Quantitative evaluation of the impact of cloud transmittance and cloud velocity on the accuracy of short-term DNI forecasts. *Renewable Energy*, 86:1362–1371, 2016. cited By 0.
- [14] C. Beekmans, J. Schneider, T. Läbe, M. Lennefer, C. Stachniss, and C. Simmer. Cloud photogrammetry with dense stereo for fisheye cameras. *Atmospheric Chemistry and Physics*, 16(22):14231–14248, 2016.
- [15] Y. Chu, H. T. C. Pedro, and C. F. M. Coimbra. Hybrid intra-hour DNI forecasts with sky image processing enhanced by stochastic learning. *Solar Energy*, 98, Part C(0):592 – 603, 2013.
- [16] Y. Chu, B. Urquhart, S. M. I. Gohari, H. T. C. Pedro, J. Kleissl, and C. F. M. Coimbra. Short-term reforecasting of power output from a 48 MWe solar PV plant. *Solar Energy*, 112(0):68 – 77, 2015.
- [17] A. Nguyen, M. Velay, J. Schoene, V. Zheglov, B. Kurtz, K. Murray, B. Torre, and J. Kleissl. High PV penetration impacts on five local distribution networks using high resolution solar resource assessment with sky imager and quasi-steady state distribution system simulations. *Solar Energy*, 132:221–235, 2016.
- [18] T. Schmidt, J. Kalisch, E. Lorenz, D. Heinemann, and G. Becker. Kürzestfristprognosen für eine 1-MW PV-Anlage basierend auf Wolkenkamerabildern. In *Tagungsband des 31. Symposiums Photovoltaische Solarenergie*, Staffelstein, 2016.
- [19] D. Peters, R. Völker, T. Kilper, M. Calais, T. Schmidt, C. Carter, K. von Maydell, and C. Agert. Model-Based Design and Simulation of Control Strategies to Maximize the PV Hosting Capacity in Isolated Diesel Networks - Using Solar Short-Term Forecasts for Predictive Control of Diesel Generation. In *Proceedings of the 32nd European Photovoltaic Solar Energy Conference and Exhibition*, 2016.
- [20] D. Peters, T. Jamal, T. Kilper, and M. Calais. Solar short-term forecasts for predictive control of battery storage capacities in remote PV diesel networks. Perth, Western Australia, 2017.
- [21] B. L. Madhavan, J. Kalisch, and A. Macke. Shortwave surface radiation budget network for observing small-scale cloud inhomogeneity fields. *Atmospheric Measurement Techniques Discussions*, 8(3):2555–2589, 2015.

- [22] M. Lave, J. Kleissl, and E. Arias-Castro. High-frequency irradiance fluctuations and geographic smoothing. *Solar Energy*, 86(8):2190 – 2199, 2012.
- [23] J. Marcos, L. Marroyo, E. Lorenzo, and M. García. Smoothing of PV power fluctuations by geographical dispersion: Smoothing of PV power fluctuations. *Progress in Photovoltaics: Research and Applications*, 20(2):226–237, 2012.
- [24] G. M. Lohmann, A. H. Monahan, and D. Heinemann. Local short-term variability in solar irradiance. *Atmospheric Chemistry and Physics Discussions*, 2016:1–23, 2016.
- [25] G. H. Yordanov, T. O. Saetre, and O. M. Midtgård. Optimal temporal resolution for detailed studies of cloud-enhanced sunlight (Overirradiance). In *2013 IEEE 39th Photovoltaic Specialists Conference (PVSC)*, pages 0985–0988, 2013.
- [26] T. E. Hoff and R. Perez. Quantifying PV power Output Variability. *Solar Energy*, 84(10):1782–1793, 2010.
- [27] C. Zender. Radiative transfer in the Earth system. *GNU Free Doc. License*, 2001.
- [28] R. Heinze, A. Dipankar, C. C. Henken, C. Moseley, O. Sourdeval, S. Trömel, X. Xie, P. Adamidis, F. Ament, H. Baars, C. Barthlott, A. Behrendt, U. Blahak, S. Bley, S. Brdar, M. Brueck, S. Crewell, H. Deneke, P. Di Girolamo, R. Evaristo, J. Fischer, C. Frank, P. Friederichs, T. Göcke, K. Gorges, L. Hande, M. Hanke, A. Hansen, H.-C. Hege, C. Hoose, T. Jahns, N. Kalthoff, D. Klocke, S. Kneifel, P. Knippertz, A. Kuhn, T. van Laar, A. Macke, V. Maurer, B. Mayer, C. I. Meyer, S. K. Muppa, R. A. J. Neggers, E. Orlandi, F. Pantillon, B. Pospichal, N. Röber, L. Scheck, A. Seifert, P. Seifert, F. Senf, P. Siligam, C. Simmer, S. Steinke, B. Stevens, K. Wapler, M. Weniger, V. Wulfmeyer, G. Zängl, D. Zhang, and J. Quaas. Large-eddy simulations over Germany using ICON: a comprehensive evaluation: Evaluation of ICON in Realistic LES Configuration. *Quarterly Journal of the Royal Meteorological Society*, 2017.
- [29] A. Heinle, A. Macke, and A. Srivastav. Automatic cloud classification of whole sky images. *Atmospheric Measurement Techniques*, 3(3):557–567, 2010.
- [30] A. Kramer. Untersuchung des Auftretens von schnellen Einstrahlungsänderungen auf Basis hochaufgelöster Messungen mit Pyranometern. 2016.
- [31] P. Blanc, B. Espinar, N. Geuder, C. Gueymard, R. Meyer, R. Pitz-Paal, B. Reinhardt, D. Renné, M. Sengupta, L. Wald, and S. Wilbert. Direct normal irradiance related definitions and applications: The circumsolar issue. *Solar Energy*, 110: 561–577, 2014.
- [32] J. Luoma, J. Kleissl, and K. Murray. Optimal inverter sizing considering cloud enhancement. *Solar Energy*, 86(1):421–429, 2012. WOS:000300459600044.

- [33] B. Urquhart, M. Ghonima, D. A. Nguyen, B. Kurtz, C. W. Chow, and J. Kleissl. Sky-Imaging Systems for Short-Term Forecasting. In *Solar Energy Forecasting and Resource Assessment*, pages 195–232. Elsevier, 2013. ISBN 978-0-12-397177-7.
- [34] H. Goverde, D. Goossens, J. Govaerts, V. Dubey, F. Catthoor, K. Baert, J. Poortmans, and J. Driesen. Spatial and temporal analysis of wind effects on {PV} module temperature and performance. *Sustainable Energy Technologies and Assessments*, 11:36 – 41, 2015.
- [35] D. Anagnostos, T. Schmidt, H. Goverde, J. Kalisch, F. Catthoor, and D. Soudris. PV Energy Yield Nowcasting Combining Sky Imaging with Simulation Models. *Proceedings of 31th European Photovoltaic Solar Energy Conference*, 2015.
- [36] A. Nottrott, J. Kleissl, and B. Washom. Energy dispatch schedule optimization and cost benefit analysis for grid-connected, photovoltaic-battery storage systems. *Renewable Energy*, 55:230–240, 2013.
- [37] APVA/CEEM. Carnarvon: A Case Study of Increasing Levels of PV Penetration in an Isolated Electricity Supply System. A report by the UNSW Centre for Energy and Environmental Markets for the Australian PV Association. Technical report, 2012.
- [38] T. Schmidt, M. Calais, E. Roy, A. Burton, D. Heinemann, T. Kilper, and C. Carter. Short-term solar forecasting based on sky images to enable higher PV generation in remote electricity networks. Perth, Western Australia, 2017.
- [39] Fulcrum3D Pty Ltd. Cloud detection and prediction for maximising solar PV utilisation in off-grid hybrid power systems Final report: project results and lessons learnt. Technical report, Fulcrum3D Pty Ltd, 2016.
- [40] T. Schmidt, J. Kalisch, E. Lorenz, and D. Heinemann. Evaluating the spatio-temporal performance of sky-imager-based solar irradiance analysis and forecasts. *Atmospheric Chemistry and Physics*, 16(5):3399–3412, 2016.
- [41] H. Goverde, B. Herteleer, D. Anagnostos, G. Köse, D. Goossens, B. Aldaladi, J. Govaerts, K. Baert, F. Catthoor, J. Driesen, and others. Energy Yield Prediction Model for PV Modules Including Spatial and Temporal Effects. In *29th European Photovoltaic Solar Energy Conference and Exhibition*, Amsterdam, 2014.
- [42] M. A. Martínez, J. M. Andújar, and J. M. Enrique. A New and Inexpensive Pyranometer for the Visible Spectral Range. *Sensors*, 9(6):4615–4634, 2009.
- [43] M. Sengupta, P. Gotseff, and T. Stoffel. Evaluation of Photodiode and Thermopile Pyranometers for Photovoltaic Applications. *Proceedings of 27th European Photovoltaic Solar Energy Conference*, 2012.

- [44] R. G. López Granados. Cloud height estimations through combination of cloud motion vectors from cloud camera and radiation measurements. 2016.
- [45] J. Kalisch and A. Macke. Estimation of the total cloud cover with high temporal resolution and parametrization of short-term fluctuations of sea surface insolation. *Meteorologische Zeitschrift*, 17(5):603–611, 2008.
- [46] T. M. Klucher. Evaluation of models to predict insolation on tilted surfaces. *Solar energy*, 23(2):111–114, 1979.
- [47] H. G. Beyer, J. Betcke, 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 European Photovoltaic Solar Energy Conference and Exhibition, Paris, France*, volume 7, page 2004, 2004.
- [48] U. Feister and J. Shields. Cloud and radiance measurements with the VIS/NIR Daylight Whole Sky Imager at Lindenberg (Germany). *Meteorologische Zeitschrift*, 14(5):627–639, 2005. WOS:000233751200004.
- [49] A. Cazorla, F. J. Olmo, and L. Alados-Arboledas. Development of a sky imager for cloud cover assessment. *Journal of the Optical Society of America. A, Optics, image science, and vision*, 25(1):29–39, 2008.
- [50] A. Kazantzidis, P. Tzoumanikas, A. F. Bais, S. Fotopoulos, and G. Economou. Cloud detection and classification with the use of whole-sky ground-based images. *Atmospheric Research*, 113(0):80 – 88, 2012.
- [51] S. Wacker, J. Gröbner, C. Zysset, L. Diener, P. Tzoumanikas, A. Kazantzidis, L. Vuilleumier, R. Stöckli, S. Nyeki, and N. Kämpfer. Cloud observations in Switzerland using hemispherical sky cameras. *Journal of Geophysical Research: Atmospheres*, 120(2):695–707, 2015.
- [52] S. B. Mende, R. H. Eather, and E. K. Aamodt. Instrument for the monochromatic observation of all sky auroral images. *Applied Optics*, 16(6):1691, 1977.
- [53] E. Schwalbe, H.-G. Maas, M. Kenter, and S. Wagner. Hemispheric Image Modeling and Analysis Techniques for Solar Radiation Determination in Forest Ecosystems. *Photogrammetric Engineering & Remote Sensing*, 75(4):375–384, 2009.
- [54] Y. Zhang, J. M. Chen, and J. R. Miller. Determining digital hemispherical photograph exposure for leaf area index estimation. *Agricultural and Forest Meteorology*, 133(1-4):166–181, 2005.
- [55] J. Kannala and S. S. Brandt. A generic camera model and calibration method

- for conventional, wide-angle, and fish-eye lenses. *IEEE transactions on pattern analysis and machine intelligence*, 28(8):1335–1340, 2006.
- [56] D. Scaramuzza. *OCamCalib: Omnidirectional Camera Calibration Toolbox for Matlab [WWW Document]*. 2014.
- [57] D. Scaramuzza, A. Martinelli, and R. Siegwart. A Toolbox for Easy Calibrating Omnidirectional Cameras. In *Proceedings to IEEE International Conference on Intelligent Robots and Systems (IROS 2006)*, Beijing, China, 2006.
- [58] J.-F. Lalonde, S. G. Narasimhan, and A. A. Efros. What Do the Sun and the Sky Tell Us About the Camera? *International Journal of Computer Vision*, 88(1):24–51, 2010.
- [59] R. Johnson, W. Hering, and J. Shields. Automated Visibility & Cloud Cover Measurements with a Solid State Imaging System. *SIO Ref. 89- 7, GL- TR-89-0061, NTIS No. ADA216906*, 1989. Final rept. 26 Sep 1984-25 Sep 1988,.
- [60] R. W. Johnson, T. L. Koehler, and J. Shields. Analysis and Interpretation of Simultaneous Multi-Station Whole Sky Imagery. *SIO 91-33, PL- TR-91-2214, NTIS No. ADA253685*, 1991. Final rept. 26 Sep 1984-25 Sep 1988,.
- [61] J. E. Shields, M. E. Karr, T. P. Tooman, D. H. Sowle, and S. T. Moore. *The whole sky imager – a year of progress*. 1999.
- [62] G. Pfister, R. L. McKenzie, J. B. Liley, A. Thomas, B. W. Forgan, and C. N. Long. Cloud Coverage Based on All-Sky Imaging and Its Impact on Surface Solar Irradiance. *Journal of Applied Meteorology*, 42(10):1421–1434, 2003.
- [63] J. E. Shields, M. E. Karr, A. R. Burden, R. W. Johnson, V. W. Mikuls, J. R. Streeter, and W. S. Hodgkiss. Research toward Multi-site Characterization of Sky Obscuration by Clouds. Technical report, DTIC Document, 2009.
- [64] J. Nou, R. Chauvin, S. Thil, J. Eynard, and S. Grieu. Clear-sky Irradiance Model for Real-time Sky Imager Application. *Energy Procedia*, 69:1999–2008, 2015.
- [65] J. Yang, Q. Min, W. Lu, Y. Ma, W. Yao, T. Lu, J. Du, and G. Liu. A total sky cloud detection method using real clear sky background. *Atmospheric Measurement Techniques*, 9(2):587–597, 2016.
- [66] S. L. Mantelli Neto, A. von Wangenheim, E. B. Pereira, and E. Comunello. The Use of Euclidean Geometric Distance on RGB Color Space for the Classification of Sky and Cloud Patterns. *Journal of Atmospheric and Oceanic Technology*, 27(9):1504–1517, 2010.
- [67] Q. Li, W. Lu, and J. Yang. A Hybrid Thresholding Algorithm for Cloud Detection

- on Ground-Based Color Images. *Journal of Atmospheric and Oceanic Technology*, 28(10):1286–1296, 2011.
- [68] J. Yang, W. Lu, Y. Ma, and W. Yao. An Automated Cirrus Cloud Detection Method for a Ground-Based Cloud Image. *Journal of Atmospheric and Oceanic Technology*, 29(4):527–537, 2012.
- [69] V. T. Jayadevan, J. J. Rodriguez, and A. D. Cronin. A New Contrast-Enhancing Feature for Cloud Detection in Ground-Based Sky Images. *Journal of Atmospheric and Oceanic Technology*, 32(2):209–219, 2015.
- [70] F. Su, W. Jiang, J. Zhang, H. Wang, and M. Zhang. A local features-based approach to all-sky image prediction. *IBM Journal of Research and Development*, 59(2-3), 2015.
- [71] R. Chauvin, J. Nou, S. Thil, and S. Grieu. Cloud motion estimation using a sky imager. In *AIP Conference Proceedings 1734*, page 150003, 2016.
- [72] J. Xu, S. Yoo, D. Yu, D. Huang, J. Heiser, and P. Kalb. Solar irradiance forecasting using multi-layer cloud tracking and numerical weather prediction. pages 2225–2230. ACM Press, 2015. ISBN 978-1-4503-3196-8.
- [73] R. Marquez and C. F. M. Coimbra. Intra-hour DNI forecasting based on cloud tracking image analysis. *Solar Energy*, 91(0):327 – 336, 2013.
- [74] D. Bernecker, C. Riess, E. Angelopoulou, and J. Hornegger. Continuous short-term irradiance forecasts using sky images. *Solar Energy*, 110(0):303 – 315, 2014.
- [75] P. Wood-Bradley, J. Zapata, and J. Pye. Cloud tracking with optical flow for short-term solar forecasting. *Solar Thermal Group, Australian National University, Canberra, Australija*, 2012.
- [76] C. W. Chow, S. Belongie, and J. Kleissl. Cloud motion and stability estimation for intra-hour solar forecasting. *Solar Energy*, 115:645–655, 2015.
- [77] J. Shi and C. Tomasi. Good features to track. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on*, pages 593–600, 1994.
- [78] B. D. Lucas and T. Kanade. An Iterative Image Registration Technique with an Application to Stereo Vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2, IJCAI'81*, pages 674–679, San Francisco, CA, USA, 1981. Morgan Kaufmann Publishers Inc.
- [79] J.-Y. Bouguet. Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm. *Intel Corporation*, 5:1–10, 2001.

- [80] P. Blanc and L. Wald. A library for computing the relative position of the sun and the earth. Technical report, GMES, Paris, France, 2011.
- [81] World Meteorological Organization. *Guide to meteorological instruments and methods of observation*. World Meteorological Organization, Geneva, Switzerland, 2008. ISBN 92-63-10008-X 978-92-63-10008-5.
- [82] M. C. Allmen and W. P. Kegelmeyer. The Computation of Cloud-Base Height from Paired Whole-Sky Imaging Cameras. *Journal of Atmospheric and Oceanic Technology*, 13(1):97–113, 1996.
- [83] E. Kassianov, C. N. Long, and J. Christy. Cloud-Base-Height Estimation from Paired Ground-Based Hemispherical Observations. *Journal of Applied Meteorology*, 44(8):1221–1233, 2005.
- [84] G. Seiz, J. Shields, U. Feister, E. P. Baltsavias, and A. Gruen. Cloud mapping with ground-based photogrammetric cameras. *International Journal of Remote Sensing*, 28(9):2001–2032, 2007. WOS:000246903300008.
- [85] F. M. Janeiro, F. Carretas, K. Kandler, F. Wagner, and P. M. Ramos. Advances in cloud base height and wind speed measurement through stereophotogrammetry with low cost consumer cameras. *Measurement*, 51:429–440, 2014.
- [86] Z. Peng, D. Yu, D. Huang, J. Heiser, S. Yoo, and P. Kalb. 3d cloud detection and tracking system for solar forecast using multiple sky imagers. *Solar Energy*, 118:496 – 519, 2015.
- [87] F. M. Savoy, J. C. Lemaitre, S. Dev, Y. H. Lee, and S. Winkler. Cloud base height estimation using high-resolution whole sky imagers. In *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2015.
- [88] V. Fung, J. L. Bosch, S. W. Roberts, and J. Kleissl. Cloud shadow speed sensor. *Atmospheric Measurement Techniques*, 7(6):1693–1700, 2014.
- [89] J. L. Bosch, Y. Zheng, and J. Kleissl. Deriving cloud velocity from an array of solar radiation measurements. *Solar Energy*, 87(0):196 – 203, 2013.
- [90] J. L. Bosch and J. Kleissl. Cloud motion vectors from a network of ground sensors in a solar power plant. *Solar Energy*, 95(0):13 – 20, 2013.
- [91] T. Schmidt, J. Kalisch, E. Lorenz, and D. Heinemann. Retrieving direct and diffuse radiation with the use of sky imager pictures, 2015.
- [92] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

- [93] S. Dev, F. M. Savoy, Y. H. Lee, and S. Winkler. Estimation of solar irradiance using ground-based whole sky imagers. *arXiv preprint arXiv:1606.02546*, 2016.
- [94] R. Haralick, K. Shanmugam, and I. Dinstein. Textural Features for Image Classification. *Systems, Man and Cybernetics, IEEE Transactions on*, SMC-3(6): 610–621, 1973.
- [95] A. Gebejes and R. Huertas. *Texture Characterization based on Grey-Level Co-occurrence Matrix*, volume 2. EDIS - Publishing Institution of the University of Zilina, 2013. ISBN 978-80-554-0648-0.
- [96] R. Tapakis and A. G. Charalambides. Equipment and methodologies for cloud detection and classification: A review. *Solar Energy*, 95(0):392 – 430, 2013.
- [97] M. Fontoynt, D. Dumortier, D. Heinnemann, A. Hammer, J. Olseth, A. Skarveit, P. Ineichen, C. Reise, J. Page, L. Roche, H. G. Beyer, and Wald, Lucien. 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, Paris, 1998. American Meteorological Society Ed., Boston, Massachusetts, USA.
- [98] D. Bourges, B. Yearly variations of the Linke turbidity factor. In *Climatic data handbook of Europe*, pages 61–64. Kluwer Academic Publishing, Dordrecht, 1992.
- [99] D. Dumortier. The Satellight model of turbidity variations in Europe. Technical report, École Nationale des Travaux Publics de l'État, Vaulx-en-Velin, France, 1998.
- [100] E. Gilleland, D. Ahijevych, B. G. Brown, B. Casati, and E. E. Ebert. Intercomparison of Spatial Forecast Verification Methods. *Weather and Forecasting*, 24(5): 1416–1430, 2009.
- [101] R. Marquez and C. F. M. Coimbra. Proposed Metric for Evaluation of Solar Forecasting Models. *Journal of Solar Energy Engineering*, 135(1):011016–011016, 2012.



# Glossary

## Cloud Types

|    |               |
|----|---------------|
| Ac | altocumulus   |
| As | altostratus   |
| Cb | cumulonimbus  |
| Cc | cirrocumulus  |
| Ci | cirrus        |
| Cs | cirrostratus  |
| Cu | cumulus       |
| Ns | nimbostratus  |
| Sc | stratocumulus |
| St | stratus       |

## Image related

|                   |                                                                             |
|-------------------|-----------------------------------------------------------------------------|
| $\theta$          | viewing angle in the camera image from zenith to the horizon [ $^{\circ}$ ] |
| CSL               | clear sky library                                                           |
| FOV               | field of view                                                               |
| HDR               | high dynamic range                                                          |
| HSV               | hue, saturation, value image color space                                    |
| HYTA              | hybrid thresholding algorithm                                               |
| PZA               | pixel zenith angle [ $^{\circ}$ ]                                           |
| $R_{\text{CSL}}$  | clear sky RBR                                                               |
| $R_{\text{mod}}$  | modified RBR                                                                |
| $R_{\text{orig}}$ | original RBR                                                                |
| RBR               | red-blue ratio                                                              |
| RGB               | red, green, blue image color space                                          |
| TSI               | total sky imager                                                            |
| USI               | UCSD sky imager                                                             |

## Error metrics

|                |                              |
|----------------|------------------------------|
| ACC            | accuracy                     |
| FS             | forecast skill               |
| MBE            | mean bias error              |
| RMSE           | root mean square error       |
| tp, tn, fp, fn | true/false positive/negative |
| ts, tc, fs, fc | true/false sunny/cloudy      |

## Irradiance

|                                   |                                                         |
|-----------------------------------|---------------------------------------------------------|
| $\theta_{sun}$                    | sun zenith angle [ $^{\circ}$ ]                         |
| $k^*$ , $k^*_{DHI}$ , $k^*_{DNI}$ | clear sky index of global / diffuse / direct irradiance |
| DHI                               | diffuse horizontal irradiance [ $Wm^{-2}$ ]             |
| DNI                               | direct normal irradiance [ $Wm^{-2}$ ]                  |
| $DNI_{clear}$                     | clear sky direct normal irradiance [ $Wm^{-2}$ ]        |
| $DNI_{est}$                       | estimated direct normal irradiance [ $Wm^{-2}$ ]        |
| G, $G_{clear}$                    | irradiance / clear sky irradiance [ $Wm^{-2}$ ]         |
| GHI                               | global horizontal irradiance [ $Wm^{-2}$ ]              |
| $GHI_{fcst}$                      | forecasted global horizontal irradiance [ $Wm^{-2}$ ]   |
| $GHI_{meas}$                      | measured global horizontal irradiance [ $Wm^{-2}$ ]     |

## Other

|                 |                                                 |
|-----------------|-------------------------------------------------|
| ASM             | angular second moment, texture measure          |
| CBH             | cloud base height                               |
| CMV             | cloud motion vector                             |
| CO <sub>2</sub> | carbon dioxide                                  |
| DC              | direct current [W]                              |
| DWD             | German Weather Service (Deutscher Wetterdienst) |
| GLCM            | grey-level co-occurrence matrix                 |
| GPS             | global positioning system                       |
| HDF5            | hierarchical data format 5                      |
| IPCC            | intergovernmental panel on climate change       |
| kNN             | k-nearest neighbours                            |
| LCOE            | levelized cost of electricity                   |
| NaN             | not a number, missing value                     |
| NTP             | network time protocol                           |
| NWP             | numerical weather prediction                    |
| OpenCV          | open source computer vision library             |
| POA             | plane of array                                  |
| PoE             | power over ethernet                             |
| PV              | photovoltaic                                    |
| SG2             | solar geometry 2 library                        |
| SVC             | support vector classification                   |
| UTC             | coordinated universal time                      |
| WMO             | World Meteorological Organization               |