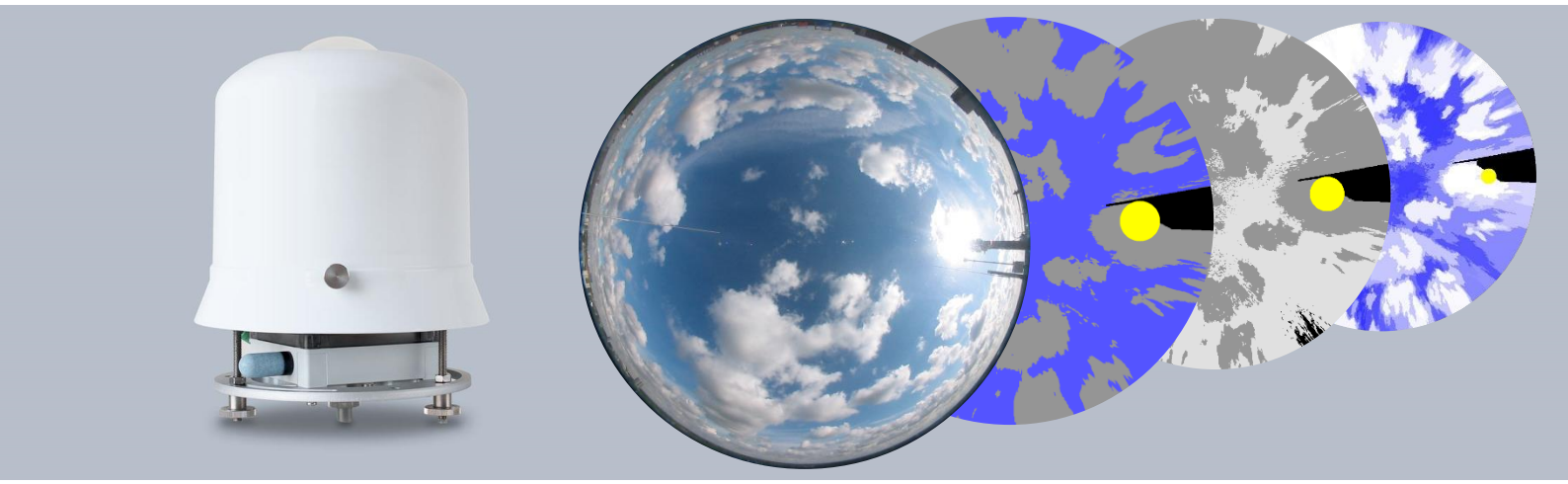
Master's Thesis – master Energy Science

March 2019

Validation of a cloud detection algorithm with an All-Sky Imager



Author:

Bas Wessels

Student number 5518105

b.wessels@students.uu.nl

Energy Science

Supervisors:

MSc. Lennard Visser

Copernicus Institute of Sustainable Development

Prof. Dr. Wilfried van Sark

Copernicus Institute of Sustainable Development

Dr. Mário Pó

EKO Instruments Europe B.V.

po@eko-eu.com

Abstract

With the increase in renewables in the future power mix, fluctuations in the power production become larger due to the intermittency of these energy sources. Accurate forecasting of solar energy production can help to improve unit-commitment decisions and reduce ancillary costs. Various methods can be used for solar forecasting. Most methods, e.g. satellite techniques, often overlook small and thin clouds. The high temporal as well as spatial resolution of All Sky Imagers enables the opportunity to take these small and thin clouds in to account, even at fast changing cloud conditions. Recent efforts at EKO Instruments let to the development of a new cloud detecting algorithm called TRINITY. This study aims at comparing the performance of the new TRINITY algorithm in combination with an All Sky Imager to the existing BRBG and CDOC cloud detecting algorithms. The new algorithm is validated using two approaches. The first method uses shortwave irradiance by determining the clearness index and diffuse fraction as proxies for Cloud Cover Fraction. The other method calculated the Cloud Cover Fraction by using downward longwave irradiation. Data is provided by two cases studies, where data is collected in Utrecht (NL) and Denver (US). Results of the shortwave irradiance method show that lowest errors where achieved by using the diffuse fraction as a proxy. Overall, the mean absolute error of the new TRINITY algorithm was 12%, whereas the BRBG and CDOC algorithms had errors of 17% and 14%, respectively. When differentiating for different sky conditions the TRINITY algorithm outperforms BRBG and CDOC at clear sky conditions, whereas in overcast conditions it outperforms the BRBG algorithm. Furthermore, the unreliable sunrise and sunset periods affect the accuracy of the algorithms and radiation measurements. Excluding the sunrise and sunset improves the accuracies with 11%, 15% and 2% for the BRBG, CDOC and TRINITY, respectively. Testing the effect of the solar position on the performance of the algorithms showed that the BRBG algorithm is most sensitive to low elevation angles, leading to higher errors. The TRINITY algorithm achieved similar performance for all elevation angles and is more stable than the other algorithms. For elevation angles of 35° and higher, all algorithms perform similarly. Preliminary results for using longwave downward radiation show that the accuracies of all algorithms are comparable (53%, 54% and 56% for BRBG, CDOC and TRINITY, respectively) with lowest errors for the BRBG algorithm. Overall, TRINITY is found to perform best followed by the CDOC and BRBG algorithm. Accurate cloud detection by All Sky Imagers will improve the accuracy of short-term solar forecasting.

Acknowledgements

The author gratefully acknowledges the support of Dr. Mário Pó from EKO Instruments Europe B.V for suggestions and reviewing the work and for providing necessary data.

Table of Contents

List of tables

List of figures

List of Abbreviations/symbols

Abbreviations

ASI	All Sky Imager
LW	Longwave irradiance
SW	Shortwave irradiance
GHI	Global Horizontal Irradiance
DNI	Direct Normal Irradiance
DHI	Direct Horizontal Irradiance
CSL	Clear Sky Library
CSI_{SW}	Clear Sky Index for shortwave radiation data (i.e. K_t and K_d)
CSI_{LW}	Clear Sky Index for longwave radiation data (calculated with eq. 10)
CSM	Clear Sky Model
BRBG	Blue/Red + Blue/Green ratio algorithm
CDOC	Cloud Detection and Opacity Classification algorithm
LT	Linke Turbidity
RMSE	Root Mean Square Error
MAE	Mean Absolute Error
CCF	Cloud Cover Fraction, i.e. fraction of sky covered with clouds
CCF_{LW}	Cloud Cover Fraction calculated with longwave radiation data (eq. 5-11)
CCF_{BRBG}	Cloud Cover Fraction calculated with the BRBG algorithm
CCF_{CDOC}	Cloud Cover Fraction calculated with the CDOC algorithm
$CCF_{TRINITY}$	Cloud Cover Fraction calculated with the TRINITY algorithm
NWP	Numerical Weather Prediction

Symbols

K_t	Clearness Index (Calculated with eq. 1)
K_d	Diffuse fraction (Calculated with eq. 2)
LW	Longwave irradiance in W/m^2
$LW_{Cloudless}$	Longwave irradiance in W/m^2 for clear sky days
ϵ_c	Clear sky emittance (calculated with eq. 6)
ϵ_a	Emittance of the sky (calculated with eq. 8)
σ	Stefan Boltzmann constant ($5.67*10^{-8}\ Wm^{-2}K^{-4}$)
T_c	The cloud base temperature in K
T_a	The mean dry bulb temperature in K
e_a	Vapor pressure at the earth's surface
$CCF_{LW}^{normalized}$	Cloud cover fraction after normalization (named CCF_{LW} in text)

1. Introduction

The ongoing concerns of climate change and energy independency pushes global power systems in the direction of sustainable sources for energy production. Within Europe, wind and solar energy will most likely be the prominent sources, since these technologies have the highest potential (Bacher, 2008; Jacobson & Delucchi, 2011). According to the Sustainable Development Scenario of the World Energy Outlook 2017 the installed capacity of wind and solar energy will be higher than the traditional fossil fuel-based technologies (International Energy Agency, 2017) in 2040. Breyer et al. (2017) found that PV electricity generation can contribute to about 69% of the global energy demand in 2050, which exceeds the targets of the Paris Agreement.

With this increase in renewables in the future power mix, fluctuations in the power production become larger due to the intermittency of these energy sources. Furthermore, as power production moves from a centralized to a more decentralized system, grid stability is likely to decrease (Schäfer, Beck, Aihara, Witthaut & Timme, 2017). This increases uncertainty on the power market and in the worst case can lead to blackouts. Accurate forecasting of solar energy production can help to improve not only electric power quality but can also help to reduce ancillary costs (Wan et al., 2015). Furthermore, short term prediction of electricity production improves unit-commitment decisions, optimization of total power production, management of the electricity grid and solar energy trading (Bacher, 2008; Chaturvedi & Isha, 2016; DeMeo, Grant, Milligan & Schuerger, 2005).

Solar power forecasting is relatively new and is not as widely used as wind energy forecasting, but methods are rapidly evolving. Tuohy et al. (2015) distinguishes various methods of radiation forecasting. Climatology methods have the longest time horizon for which radiation values can be forecasted. These methods use average weather statistics of previous years and use these values to make a forecast of over 10 weeks. The Numerical Weather Prediction (NWP) methods have been used for many years, but only recently these methods are improved and specialized for solar irradiance forecasting. These methods consist of computer models which use current weather variables and extrapolate these over a defined time horizon, hereby statistical learning methods are used to improve the accuracy of NWP methods (Chaturvedi & Isha, 2016). Six hours to two weeks is the preferred time horizon for this method. Statistical learning methods use historical site-specific irradiance data to train the methods. By combining this with direct observation of irradiance, solar radiation can be forecasted up to 6 hours. Satellite Imagery methods use satellite images with high spatial resolutions to forecast solar irradiance. However, temporal resolution is low, which cause these methods to function ineffectively at fast changing cloud conditions. Satellite imagery is used best on a time horizon of 1 minute to 6 hours. Finally, All Sky Imaging (ASI) is a method which uses high resolution ground-based cameras which can detect clouds, estimate cloud height and determine cloud motion (Tuohy et al., 2015). By classifying pixels into clear sky, thin clouds or thick clouds the surface solar irradiance can be estimated for the very short term; up to 30 minutes.

Climatology method, NWP models and Satellite imagery have a forecasting horizon focusing on the longer term. Due to this low temporal, and often low spatial, resolution, they are inadequate for forecasting solar irradiance on the very short term (Chow et al., 2011). A high temporal as well as spatial resolution are required to take fast changing cloud conditions in to account. By using an ASI in combination with a cloud detecting algorithm the path of cloud vectors can be extrapolating through which irradiance values can be estimated. Therefore, this method is highly suitable for accurate irradiance prediction for the very short term. Moreover, satellite techniques often overlook small clouds or are confused by thin clouds and the earth surface due to a similar brightness and temperature (Heinle, Macke, & Srivastav, 2010). The high resolution and more equally coloured background makes ASIs a more accurate method in forecasting solar irradiance for the short term.

The images taken by an All Sky Imager (ASI) have to be processed to determine and classify clouds. For this, algorithms are used. The algorithms use the images as input and evaluate them by using different techniques. Before evaluating, the image needs to be corrected for obstructions such as high buildings, the camera arm and the scattering of light close to the sun. All these steps are included in the algorithm. A widely used evaluating technique is to use the red-blue ratio as a proxy to determine the presence of clouds. In a clear sky day, the shorter wavelengths of the light spectrum (primarily blue) are scattered more heavily, which results in a blue colour. Whereas the bigger particles in clouds create a more uniformly scattering, resulting in a grey colour. By using the difference or ratio of both ends of the visible light spectrum, Red-Blue Ratio (RBR) algorithms determine the cloudiness (Kleissl, 2013). Furthermore, other algorithms are developed that use a Clear Sky Library (CSL) in addition to the red-blue ratio. The Cloud Detection and Opacity Classification (CDOC) algorithm, for instance. To run these types of algorithms a CSL has to be (manually) selected. This is a selection of clear sky images with variating elevation angles. Then an image can be evaluated by comparing each pixel with the pixel of the clear sky image.

Recent efforts at EKO Instruments Europe B.V. let to the development of a new algorithm called TRINITY. This new algorithm is based on RBR algorithms, but different to previous RBR algorithms the TRINITY algorithm is able to classify the cloudiness in multiple cloud thicknesses. Besides, the TRINITY algorithm does not need a CSL. Using this new algorithm in combination with an ASI, short term irradiance prediction should be done with a higher accuracy compared to previous algorithms. However, since this algorithm is newly developed, no research has been conducted regarding the performance compared with other cloud detection methods. This has led to the following research question:

How does the TRINITY algorithm in combination with an All Sky Imager perform compared to existing cloud detecting algorithms?

To answer this research question, the research uses two methods in which the newly defined algorithm is validated and evaluated compared to already existing algorithms used for cloud detection. Both methods use data originating from two case studies: Utrecht (the Netherlands) and Denver (United States). By studying the performance of this new algorithm and comparing this to existing literature, the research introduces a new way of short-term solar radiation prediction. By doing so, this research adds theoretical knowledge about the performance of the TRINITY algorithm with respect to already existing algorithms to the field of solar forecasting, or more general to the field of Energy Science.

This report continues with a literature review, which is used as background information to clarify the position of All Sky Imagers as cloud forecasting method with respect to other solar and cloud forecasting methods. Then the methodology section describes the methods used, along with a description of both case studies. Then, results are given. A discussion interpreters the results and denotes research limitations. Finally, a conclusion is given.

2. Literature review

To better understand how ASI and their algorithms work, this literature review clarifies the position of ASI within the field of solar and cloud detecting methods. This section aims to map the most important technologies and methods of forecasting radiation and identifying cloudiness. Although this is a very broad field the goal is to provide a comprehensive overview of solar forecasting methods. Tuohy et al. (2015) created an overview of these forecasting methods. In this part each category is described and assessed in more detail.

Climatology methods

Climatology methods for predicting future solar radiation and other weather conditions is one of the oldest and simplest forecasting approaches. Descriptions and research of this method go back more than 50 years (Inman, Pedro, & Coimbra, 2013; Isenson, 1966). The method is also referred to as 'persistence forecasting', this is a forecast where prediction of the present case is based on outcomes of similar past cases (Riordin & Hansen, 2002). This means that cyclical, or seasonal, trends are observed and extrapolated, which makes the methods static. For instance, if the radiation of a particular day is to be known, one averages the radiation measured on that day for as far back as the dataset allows (Yankeelov, Quaranta, Evans, & Rericha, 2015). Since climatology methods use irradiance data from previous years, the short-time variability of solar radiation is not taken into account. So, by not taking into account these stochastic characteristics of solar radiation the method becomes inferior to other forecasting methods (Inman et al., 2013). This is supported by Isenson (1966) who criticizes the accuracy of climatology methods. However, accuracy can be improved when historical data is not leading, but used as input by an experienced meteorological forecaster to provide his forecast (Hyvärinen, Julkunen, & Nietosvaara, 2007).

For climatology methods there is not a lot of input data required, the largest problem could be the availability of the data. To forecast radiation values for a specific location as precisely as possible one needs, ideally, as much historical data as possible measured at the same location. However, since meteorological instruments which measure irradiation are costly to acquire and this data is usually only collected at meteorological stations or research institutes, availability of data is a common limitation. Climatological forecasts provide prediction for the longer time horizon. However, these forecasts are mostly used as a first estimation or reference (Campbell & Diebold, 2005). Then, other forecasting methods can be used to be more specific. Since climatological methods are often used in a very early stage and complemented by more specific statistical learning methods or NWPs, the spatial resolution is low.

Statistical learning methods

Direct observations of radiation can be used for statistical learning methods to forecast future solar conditions. These statistical learning methods use historical data of site irradiance to train the learning methods. Then, real time data or observations can be used to predict solar irradiance by using the trained learning methods (Tuohy et al., 2015). Machine learning, which is part of statistical learning methods, can be used in several domains and it enables one to solve complex problems which cannot be solved by explicit algorithms. In short, a statistical learning method finds relations between inputs and outputs even if the representation is impossible (Voyant et al., 2017). The use of statistical learning methods is one of the most common approaches used in solar radiation forecasting, Antonanzas et al. (2016) found that 72% of their analysed papers use a statistical learning method.

Statistical learning methods do not need any internal information from the system to forecast radiation. Instead, the methods use historical data to determine relations which can be used to forecast. The quality and amount of data used for this method are very important (Antonanzas et al.,

2016). Since historical data is the only input, the quality of this data determines the quality of the output data. Furthermore, having a large data set is required. In this way a proper selection can be made for the training of the learning methods. Establishing a proper training set can highly affect the accuracy. Statistical methods usually provide reliable forecasts for an intra hour time horizon (Diagne, Lauret, & David, 2012; Tuohy et al., 2015), however by combining this method with NWP a time horizon of 2-4 hours can be achieved. The spatial resolution of statistical learning methods varies a bit, but it doesn't exceed the 10 km (Diagne et al., 2012; Diagne, David, Lauret, Boland, & Schmutz, 2013).

Numerical Weather Prediction

Current weather conditions are used as input to mathematical models to predict the future weather conditions. The mathematical models simulate the processes occurring in the atmosphere. These models are called NWP models (Paulescu, Paulescu, Gravila, & Badescu, 2012). Since the processes occurring in the atmosphere are too uncorrelated with past data to use statistical learning methods, NWP models are used. These types of models do not need historical data and are better suited for day-ahead predictions (Lorenz et al., 2009). However, since statistical learning methods as well as NWP has its own strengths and weaknesses, both methods can be combined to improve the accuracy of the forecast. Verbois, Huva, Rusydi, & Walsh (2018) combined a NWP model with their proposed statistical learning method and found that it outperformed a climatology forecast and a NWP model forecast for the day-ahead forecasting of radiation. For solar radiation forecasts, NWP models are used to determine the probability of cloud formation in a defined area and then use this probability to indirectly estimate the radiation values by using a dynamic atmosphere model (Voyant et al., 2017).

NWP models only need current weather conditions as input variables. This is particular interesting for forecasting irradiance at newly developed solar parks or sites which do not have historical data available. The data is usually obtained from a global network of observations and measurements, but when enough local data is available the spatial resolution of NWP models can be increased. The main input variables for NWP models are: wind, humidity and surface pressure. Besides variables like snow cover or sea surface, temperature can, if available, increase the accuracy of the forecast as well (Diagne et al., 2013). The time horizon of NWP models is widely discussed and varies between 6 hours and 15 days ahead. Diagne et al. (2012) and Voyant et al. (2017) state that forecasts beyond 6 hours and up to several days ahead are generally most accurate when NWP models are used. At shorter time scales, NWP is often combined with post-processing technologies, e.g. Model Output Statistics. This is a statistical learning method which can relate the NWP model output with prior observations or other data. Using NWP in combination with a statistical learning method can thus increase the accuracy for the shorter time horizon. However, NWP models can also be used for the longer time horizon, as Tuohy et al. (2015) and Diagne et al. (2013) shows. They describe that NWP models can predict accurate weather conditions up to 15 days ahead. Next to the time horizon, the spatial resolution is also very dependent on the type of NWP and complementary methods used. Diagne et al. (2013) mention a spatial resolution of 16-50 km, where Antonanzas et al. (2016) and Diagne et al. (2012) describe the spatial resolution to be ranging from 1 to 100 km. Although this is a large range, the spatial resolution of NWP models is often too low to make accurate radiation forecasts which take cloud enhancement into account (Mathiesen & Kleissl, 2011).

Satellite imaging

Another method for predicting the surface solar irradiance at specific locations is the use of geostationary meteorological satellites. These satellites follow the rotation of the earth and are thus always above the same geographical location. This enables these satellites to scan large areas up to several times per hour. In this way, irradiance images can be acquired which provide also information about cloud enhancement and motion vectors (Blanc, Remund, & Vallance, 2017). By using at least two satellite images taken after one another it is possible to derive cloud motion vectors. These vectors can then be extrapolated to forecast the Global Horizontal Irradiance (GHI) for a defined period.

Although this method provides better resolutions than NWP models it is still challenging to accurately predict solar irradiance when fast changing cloud conditions occur. When using satellite imaging to forecast solar radiation in Europe or Africa, the MeteoSat network of satellites can be used. This satellite network is part of a much bigger network worldwide which can be used to cover almost all areas on earth and provide information of cloud properties and movements, except for the North and South Polar extended areas. This worldwide network further exists of the GOES (North and South America), FENG YUN (China) and HIMA-WARI (Asia, Australia and New Zealand) families of satellites (Blanc et al., 2017).

Before predicting solar radiation from satellite images, the images have to be calibrated with a specific location. For this, a Clear Sky Model (CSM) can be used. A CSM calculates irradiance values at clear sky conditions based on geographical coordinates. These models usually require data input like aerosol content, water vapor, elevation and ozone (Tuohy et al., 2015). The modelled clear sky radiation can then be combined with the satellite images, which form the most important inputs. As described, these images can be obtained from a worldwide network of geostationary meteorological satellites. The time horizon for which this method can be used is clearly shorter than the previous methods described. According to Inman et al. (2013) satellite forecasting is ideal for a 30 min to 6 hours horizon. However, when satellite images can be taken with a much shorter time interval, the time horizon could even be lowered to minutes (Tuohy et al., 2015). This is supported by Kleissl (2013) who mentions a time horizon up to 6 hours ahead as well. The spatial resolution of satellite imaging is also better than previous methods described. The METEOSAT satellites used to map central Europe are able to provide a spatial resolution of 2.5 km, where the GOES-13 satellite used in the United States can even reach a spatial resolution of 1 km (Bilionis, Constantinescu, & Anitescu, 2014; Lorenz, Hammer, & Heinemann, 2004). Most satellites have a spatial resolution somewhere between 1 - 10 km (Inman et al., 2013).

Sky imaging

Where NWP models and satellite methods used for solar forecasting provide prediction for a longer time horizon with appropriate resolution, AISs are used to forecast clouds and irradiance values for the very short term. They establish a sub-kilometer view of clouds over e.g. a PV power plant or urban area (Chow et al., 2011). ASIs are cameras aimed at the sky which take short interval images by using a 180° field of view camera system. In this way, high quality images can be acquired which capture the complete horizon. A weatherproof housing enables the cameras to operate continuously and under any weather conditions. The cameras ability to take high resolution pictures as well as the lower spatial resolution compared to other solar forecasting methods, makes this method more expensive with respect to previously described methods. Though, the images taken by these cameras are much more accurate in assessing and classifying cloudiness. Consecutive images taken can be used to estimate cloud velocity to forecast irradiance values on the very short term. The processing and evaluation of the images taken can be done by various algorithms.

This method of solar radiation forecasting does not require a lot of input data. The most important source of input data is the ASI itself. This needs to be installed somewhere free of obstructions which could block the sky. Then after the enclosed software has been installed and the ASI is calibrated, the ASI is ready for use. During this installation the geographical location of the ASI is one of the few necessary input data. Longitude, latitude and altitude of the camera has to be determined. The time horizon for which this method can be used is clearly lower than previous methods described. With the time horizon not exceeding half an hour, literature explains that sky imaging is highly suitable for cloud detection and solar forecasting which takes into account the variability of cloud enhancement (Antonanzas et al., 2016; Diagne et al., 2013; Kuhn et al., 2017; Tuohy et al., 2015). Just as the time horizon used for sky imaging, literature is also very clearly about the spatial resolution of sky imagers. This resolution does not exceed 1 km and Antonanzas et al. (2016) mention a minimum value of 2.5 m. Kuhn et al. (2017) validated an ASI which used a pixel base of 5 m.

ASIs are used in combination with an algorithm. These algorithms are the connecting step between the image taken and the solar forecast. The accuracy of the forecast is thus strongly dependent on the performance of the algorithm. Chauvin, Nou, Thil, Traore, & Grieu (2015) identify three different categories for the existing cloud detection algorithms: thresholding techniques, neural network models and more sophisticated approaches. From these categories thresholding techniques are the most widely used, because of their simplicity and thus the low computational time. Since this research uses a lot of data and the time is limited, the scope will be reduced to algorithms using thresholding techniques. Some recently used and researched threshold algorithms are discussed below. Note that there are a lot of algorithms developed over the years and some of them will have overlap between them.

BRBG

The BRBG algorithm uses the Blue/Red + Blue/Green ratio to differentiate between clouds and clear skies. This method uses the difference in light scattering by clouds versus a clear sky day: in a clear sky, light with shorter wavelengths are scattered more heavily, which explains why we observe a blue sky (Kleissl, 2013). The output of the algorithm is a factor between 0 and 1 which describes the cloudiness of an image (CMS, 2016). In the early stages of sky imaging, a lot of authors have developed cloud detecting algorithms based on the red-green-blue colour space. Yang et al. (2015) describes alternatives of these algorithms that consider different ratios like R/B, R-B and (R-B)/(R+B) which have been used throughout the years. However, the B/R + B/G ratio developed by CMS Schreder is one of the most recently studied ratios and commonly used in combination with EKO Instruments ASIs.

CDOC

The Cloud Detection and Opacity Classification algorithm builds upon the BRBG algorithm and is able to classify the cloudiness of an image in three categories: clear sky, thin clouds or thick clouds. Prior to running the algorithm, a CSL has to be defined. The algorithm then uses the difference, rather than the ratio, between the red-blue ratios of the image and the CSL to determine thick clouds. A Haze Correction Factor is used to distinguish clear sky and thin clouds (CMS, 2016; Ghonima et al., 2012). These calculations are done for each pixel separately and for the overall evaluation the calculations are combined in one figure. The CDOC algorithm provided significantly improved results when compared to the original ASI software. The classification of clear sky and thick clouds was correct for 99% of the cases. For thin clouds the accuracy was 60% (Ghonima et al., 2012). These results were calculated by using manual validation of 30 images originating from one case study.

GBSAT & CSBD

The Green channel Background Subtraction Adaptive Threshold (GBSAT) algorithm developed by Yang et al. (2015) is focused on evaluating partly cloudy images. First the algorithm determines whether the sun is obscured by clouds. Then clouds are detected based on a background subtraction adaptive threshold method. This means that the algorithm detects the solar position, then simulates a clear sky picture including circumsolar region. This simulated background image is then subtracted from the original image to obtain an image more suitable for cloud detection. The proposed method is compared with the R/B, R-B and BSAT methods by human examination. Although these results cannot be quantified, the authors conclude that the GBSAT algorithm obtains more satisfactory results, especially in the circumsolar and near-horizon regions. However, the algorithm is still sensitive for thin clouds in the circumsolar region. The reason was that simulated background could not always represent real sky background. Thus, especially for thin clouds this could cause errors. Therefore, the algorithm was revised such that a real clear sky background was adopted instead of a simulated background. The new Clear Sky Background Differencing (CSBD) algorithm was, as the GBSAT algorithm, visually validated and showed a better performance, especially for the thin clouds where the GBSAT was vulnerable (Jun Yang et al., 2016).

TRINITY

The newly developed TRINITY algorithm is based on the BRBG algorithm. The blue/red + blue/green ratio of the BRBG algorithm outclassed a variety of other ratios in the RGB colour channels (CMS, 2018). TRINITY classifies the sky by recognition of cloud contours (objects). The software then assigns ratio values to these contours. This creates a hierarchy of cloud objects. This enables the software to evaluate a sky image and classify the clouds in multiple layers of thicknesses. A great advantage of the new TRINITY algorithm is that the software only needs the underexposed images to provide the same quality of evaluation (CMS, 2018). Since the algorithm has recently been developed, no research has been done regarding the accuracy nor the relation to other cloud detecting algorithms.

Overview

All the indicators of the different solar forecasting methods are summarized, combined and shown in Table 1. The time horizon and spatial resolution are visualized in Figure 1 as well. Note that the boundaries between different methods are not explicit, and therefore other authors may draw the boundaries elsewhere. The purpose of the visualization is to clarify the differences with respect to the other methods and to create a broad understanding of the various methods for solar forecasting. Figure 1 is based on the data from Table 1. As can be seen from Table 1 and Figure 1, there is some overlap between numerous methods. However, each of the methods described serves its own purpose. It is therefore important to understand that the methods for cloud and solar forecasting are rather complementary than substitutes. Statistical learning methods and satellite imaging for example each use different input data. When solar irradiance for a specific site needs to be estimated statistical learning methods can provide forecast for a shorter time horizon. However, when the site is remote and no historical data is available, satellite imaging may be a more convenient way. Furthermore, part of the overlap between methods occurs due to the ability to combine various methods. Statistical learning methods can be combined with sky imaging, for example. Finally, the climatology methods are shown with a dashed line, since it is hard to quantify the spatial resolution. Literature has shown that the spatial resolution is low and the time horizon is long, but these indicators are not and cannot be quantified in a reliable way because these methods are used to make first approximations of irradiances (e.g. a solar resource assessment before installing PV panels) instead of actual forecasts. More reliable and specific forecasts are made by using other methods.

Table 1: Overview of solar forecasting methods and their indicators

Method	Input data	Time horizon	Spatial resolution
Climatology methods	Historical irradiance data	> 1 day	low
Statistical learning methods	Historical irradiance data (preferably multiple years)	< 4 hours	< 10 km
Numerical Weather Prediction	Current weather conditions: • Wind • Temperature • Humidity • Surface pressure • Site specific data (snow cover, etc.)	6 hours – 15 days	1 – 100 km
Satellite imaging	Satellite images Clear Sky Model variables: • Longitude • Latitude • altitude	1 min – 6 hours	1 – 10 km
Sky imaging	All Sky Images Clear Sky Model variables: • Longitude • Latitude • altitude	< 30 min	< 1 km

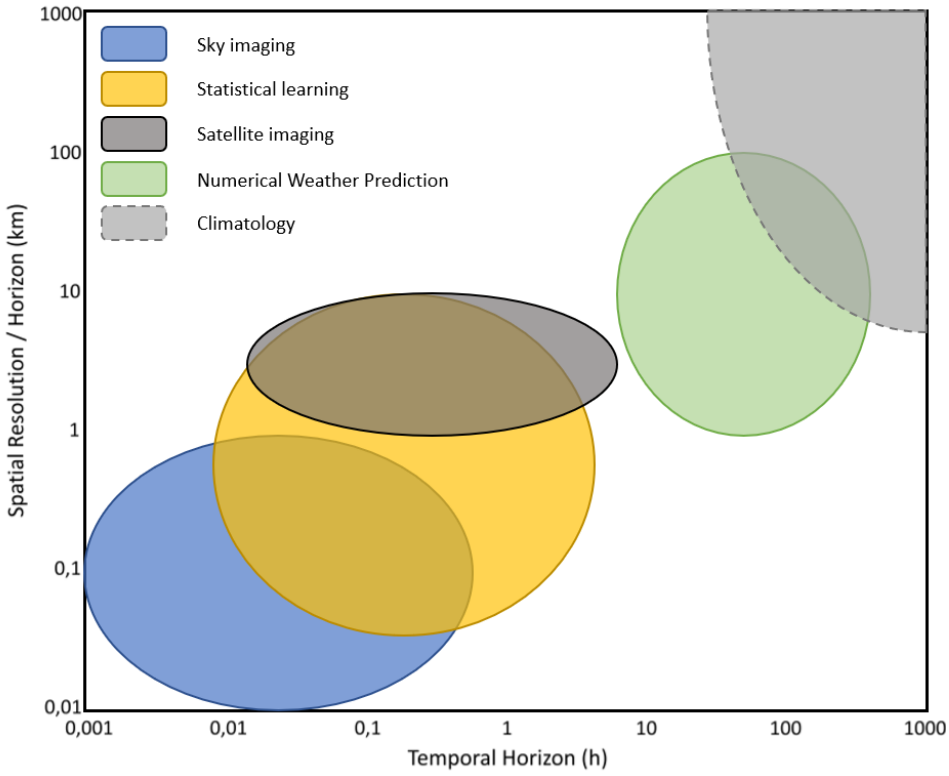


Figure 1: Overview of spatial resolution and temporal horizon for reviewed solar forecasting methods

3. Methodology

In this study the application of a new algorithm is explored and assessed. The study validates the new TRINITY algorithm as well as the existing BRBG and CDOC algorithms. The TRINITY algorithm is compared to the BRBG and CDOC algorithms, since these algorithms are currently being used in combination with ASIs provided by EKO Instruments. The research is conducted by analysing two case studies. These case studies are performed to validate the new TRINITY algorithm as well as the existing BRBG and CDOC algorithms.

Case studies are intensive analyses and descriptions of a single unit or system (Hancock & Algozzine, 2006). They are mostly used by researchers to understand a specific phenomenon or situation and can rarely be generalized (Brown, 1998; Thomas, 2015). Furthermore, case studies are very useful in answering 'How?' and 'Why?' questions and are thus used for exploratory, descriptive or explanatory research (Rowley, 2002). Supported by the case study literature, the case study methodology is well suited for the validation of the algorithms in this research, since the algorithms can be evaluated and validated by using the data collected at the two sites. By running all three algorithms for the same sky images and validating this with the irradiance measured on site, a fair and reliable comparison is made between all algorithms and thus the performance of the TRINITY algorithm with respect to the BRBG and CDOC algorithms is assessed. The algorithms were provided by EKO Instruments, and sky images as well as irradiance data were collected at two different sites, which formed the case studies in this research.

Since for case study I only shortwave irradiation was measured and for case study II both shortwave- and longwave irradiation was measured, the methodology section is divided in two sections. After the case studies are described in the data collection part, the methodologies are explained. First the methodology for shortwave irradiance data is explained, then the method for using longwave irradiance to evaluate the algorithms is explained. Although both methods are very different, it is inevitable that there exists some overlap. Concepts and processes are only explained once, even though some will be used in both case studies.

Data collection

Case study I: Utrecht, the Netherlands

Within this research a collaboration between Utrecht University and EKO Instruments enabled the opportunity to perform the case study. Sky Images were collected at Utrecht University (N 52.08^o, E 5.17^o at 32m) with the Sky Imager EKO SRF-02. Images were taken with an interval of 10 minutes, with each interval 2 images: one normal exposed and one underexposed. Sky images were collected for the period June 2013 until December 2016. However, the period from November 2013 till June 2014 was not usable, since a crane was obstructing the Sky Imagers view. Next to sky images, three years of irradiance data was collected at the same location of the Sky Imager, this data has an interval of 5 minutes and, next to GHI, existed of Direct Normal Irradiance (DNI) and Diffuse Horizontal Irradiance (DHI). The data of case study I is summarized in Table 2.

Case study II: Denver, United States of America

The second case study was enabled by data collected by the National Renewable Energy Laboratory (Andreas & Stoffel, 1981). EKO Instruments installed the new ASI-16 All Sky Imager in Denver, Colorado (N 39.74^o, W 105.18^o at 1829m). The Images were acquired with a 10-minute interval. As in the first case study, a normal exposed and an under exposed image are taken at each timestamp. Since the timestamps of the images were saved in local time and the TRINITY software had some minor bugs, only 17 days of the data available could be used for the case study. All of these days are from January 2018. Although all cloud conditions (clear sky, thin clouds/haze, partly cloudy, overcast) occurred during this period, clear sky conditions were strongest represented. Since the sky imager was assembled close to a sun tracker, the arm is appearing in the west part of the image and is changing throughout the day. Because of this movement, the arm of the sun tracker could not be excluded from the evaluations and thus errors the results obtained by the algorithms. However, since the arm is quite thin, the estimated error occurring is less than 1% and is therefore considered to have a very limited impact on the results.

Next to the sky images acquired on this site, a pyrgeometer was installed to collect downward Longwave (LW) irradiance data. Other data collected on site was GHI, DNI, DHI, sky temperature and ambient temperature. The sky imager was synchronized with the other meteorological instruments on site. Therefore, the sky imagers and meteorological data have the exact same timestamps. The data of case study II is summarized in Table 2.

Table 2: Summary of the case studies

	Case Study I	Case study II
Coordinates	N 52.08^o, E 5.17^o	N 39.74^o, W 105.18^o
Altitude	32m	1829m
Sky Imager	EKO SRF-02	EKO ASI-16
Time period	01-07-2014/01-07-2016	01-01-2018/17-01-2018
Time interval	10 minutes	10 minutes
GHI	x	x
DNI	x	x
DHI	x	x
LW		x
Sky temperature		x
Ambient temperature		x

Data analysis

Shortwave irradiance

The first analysis uses Shortwave (SW) radiation. This part of the light's spectrum is visible to the human eye and has a wavelength ranging from 380 to 750 nm. This range is called the visible light spectrum. Most of the energy emitted by the sun is in the form of visible light, which carries a lot of energy (Young & Freedman, 2016). These shortwave radiation in the form of GHI, DNI and DHI is used in the analysis, which consists of four consecutive steps. Finally, a decomposition analysis, time-of-day analysis and sensitivity analysis are done.

Step 1: Clear Sky Irradiance

In the first step of the analysis the irradiance values at the location where the Sky Imager is placed is calculated for clear sky conditions. This defines maximum radiation values which is used later in step 2. The radiation values were calculated by using a Clear Sky Model (CSM). This is a model which uses meteorological parameters such as solar elevation angle, site altitude or aerosol concentration to calculate the clear sky irradiance, i.e. the radiation reaching the ground at zero cloudiness.

Over the years a wide range of CSMs have been developed. Gueymard (2012) has collected and analysed 18 different CSMs, from which the REST2, Ineichen/Perez and Hoyt models form the top 3. From these models, the Ineichen/Perez model achieves similar performance as the REST2 model but does not need site-specific data which the REST2 model does use. Furthermore, Reno, Hansen, & Stein (2012) recommend the Ineichen/Perez model for most locations, due to its ease of application and high degree of performance. For these reasons, this research uses the Ineichen/Perez model. The required inputs for this model are time (year, month, day, hour, minute and seconds), location (latitude, longitude and altitude) and Linke Turbidity (LT). The LT factor is a good approximation to describe the optical thickness of the atmosphere under clear sky conditions and is dependent on latitude, longitude and month of the year. It takes the absorption and scattering of water vapor and aerosols into account. The Ineichen/Perez model automatically incorporates the right LT into the calculations from a large lookup matrix. This lookup matrix, created by (Remund, Wald, Lefèvre, Ranchin, & Page, 2003), is based on specific geographical locations where accurate measurements are made. Since the KNMI station in De Bilt (less than 2 km away from Utrecht) is one of these sites, the LT is assumed to be accurate. The outputs of the model are GHI, DNI and DHI in Watt per square meter (W/m^2) at clear sky conditions. The model is run in MATLAB and the output was stored with a time interval of 5 minutes.

Step 2: Clear Sky Indices

After the theoretically estimated radiation values have been calculated by the CSM, these were compared with the actual measured radiation values on site. This principle forms the basis of the Clear Sky Index (CSI) approach (Marty & Philipona, 2000). This approach is increasingly being used for modern solar radiation modelling and forecasting (Engerer & Mills, 2014). Since this case study uses shortwave radiation, these CSIs will be denoted as CSI_{SW}. The first CSI_{SW} which is used is the clearness index (K_t) adapted from Cros, Liandrat, Sébastien, Schmutz, & Voyant (2013):

$$K_t = \frac{GHI_{measured}}{GHI_{CSM}} \tag{1}$$

Where K_t is the GHI CSI_{SW}, $GHI_{measured}$ is the global horizontal irradiance measured on site and GHI_{CSM} is the irradiance calculated by the Clear Sky Model. When K_t is 1, there are no clouds and measured radiation equals estimated radiation. As cloudiness increases, $GHI_{measured}$ will decrease and thus the

CSI_{SW} will mainly vary between 0 and 1. However, due to super irradiance the CSI_{SW} can sometimes become higher than 1. Super irradiance occurs on days where clear sky conditions rapidly change to cloudy conditions and vice versa. When the irradiance bends around the edge of a cloud, a temporary peak in irradiance occurs which could exceed maximum clear sky irradiance as calculated by step 1.

Next to the clearness index the Diffuse Horizontal Irradiance CSI_{SW}, i.e. the diffuse fraction, is used as a proxy for Cloud Cover Fraction (CCF) as was done by Butt et al. (2010). They found a very good linear relationship between the CCF and the diffuse fraction. The diffuse fraction K_d is the measured DHI ($DHI_{measured}$) divided by the measured GHI ($GHI_{measured}$) and represents the cloud cover (Okogbue, Adedokun, & Holmgren, 2009). Contrary to the clearness index the diffuse fraction reaches a value of 1 at very cloudy conditions and a value approaching zero at clear sky conditions. The diffuse fraction is shown in the equation below.

$$K_d = \frac{DHI_{measured}}{GHI_{measured}} \tag{2}$$

Step 3: Calculating CCF with the algorithms

This step involves the creation of the CCFs by running the algorithms. The FINDCLOUDS software (version 3.2.0.1) was used to run the BRBG and CDOC. The new algorithm was run by the newly developed FINDCLOUDS TRINITY software (version 4.0). The images taken by the ASI were used as input for the software. Since a lot of data is available, processing this data required substantial computing power. Therefore, small test samples were used till the software was calibrated and an appropriate output had been established. The calibration of the software consists of adjusting the offset angle, the zenith angle, centring of the image and determining the horizon and CSL. After calibration, the output for the BRBG algorithm is a CCF on a scale from 0 to 1, whereby zero is no clouds and one stands for complete overcast. The CCF of the BRBG algorithm is denoted as CCF_{BRBG}. The CDOC algorithm exists of a cloudiness fraction (CCF_{CDOC}) as well, which co-exists of a thin cloud fraction and thick cloud fraction. The output format of the algorithms (BRBG and CDOC) by using the FINDCLOUDS software is shown in Table 3.

Table 3: Example of the output of the BRBG and CDOC algorithms

Date	Time	BRBG	CDOC	Thick	Thin
12-01-2018	12:25:00	0.04	0.40	0.13	0.27
12-01-2018	12:30:00	0.10	0.59	0.38	0.21
12-01-2018	12:35:00	0.09	0.87	0.73	0.14
12-01-2018	12:40:00	0.02	0.73	0.71	0.02

The new FINDCLOUDS TRINITY software evaluates sky images in a different way than the BRBG and CDOC algorithm. Although the new algorithm incorporates the BRBG method, it is mainly based on object detection. This software only needs the under exposed images to calculate cloudiness. The output format, however, is the same as the previous algorithms. A $CCF_{TRINITY}$ of 0 represents clear sky conditions, where a $CCF_{TRINITY}$ of 1 corresponds with overcast conditions. An example of the new TRINITY algorithm output is shown in Table 4. Where 'Ext' represents the extension of the image, an extension of 01 denotes a normal exposed image and an extension of 02 denotes an underexposed image.

Table 4: Example of the output of the TRINITY algorithm

Date	Time	Ext	Cloudiness
01-07-2014	07:00:00	01	0.71
01-07-2014	07:00:00	02	0.76
01-07-2014	07:10:00	01	0.69
01-07-2014	07:10:00	02	0.72
01-07-2014	07:20:00	01	0.71
01-07-2014	07:20:00	02	0.76
01-07-2014	07:30:00	01	0.74
01-07-2014	07:30:00	02	0.78

Step 4: Calculating metrics

In the final step of the case study the CCFs of the algorithms was related with the estimated CSIs. In this step the accuracy of the algorithms was determined and thus the performances of all algorithms were compared. The CSIs are calculated with an interval of 5 minutes, where the CCFs have an interval of 10 minutes. Therefore, the measured irradiance values were reduced to a 10-minute time interval. Besides, a linear interpolation method was used so that the CSIs match the exact timestamps of the sky images taken. To avoid misalignment of sky images and irradiance data during summer, there is accounted for daylight savings by adjusting the UTC Offset from 2 to 1. Furthermore, the CCFs from the FINDCLOUDS software contradicts to the clearness index: a high CCF represents cloudy conditions, where a high value of the clearness index represents clear sky conditions. Therefore, $1 - K_t$ was used to evaluate the algorithms.

Since the quality of this research is highly dependable on the quality of the data used, the irradiance data obtained has been examined. For this, the method of Reindl & Beckman (1990) is used. They state that extreme data as well as data which violate physical limits, needs to be excluded. Therefore, four thresholds were defined which are used to filter out unreliable data. Data which satisfied the following criteria were excluded from the analysis:

1. $K_t > 1$
   When the global solar radiation is exceeding the extra-terrestrial radiation, i.e. when measured GHI is exceeding the theoretical maximum GHI calculated by the CSM.
2. $K_d > 1$
   When the diffuse fraction is higher than 1, the measured DHI exceeds measured GHI, which violates physical laws since $GHI = DNI*\cos(\theta) + DHI$, whereas $DNI \geq 0$.
3. $K_d > 0.8$ and $K_t > 0.6$
   A diffuse fraction higher than 0.8 represents cloudy conditions, whereas a clearness index higher than 0.6 represents clear sky conditions. This criterion is used to eliminate this contradicting and unreliable data. The same limits were used as in Reindl & Beckman (1990).
4. $K_d < 0.9$ and $K_t < 0.2$
   As the previous criterion, this statement excludes contradicting data. A diffuse fraction lower than 0.9 represents clear sky conditions, whereas a clearness index lower than 0.2 represents cloudy conditions. The limits are adapted from Reindl & Beckman (1990).

Finally, the database of the sky images used was not 100% complete as well. For unknown reasons sky images were missing. All calculations for these timestamps were deleted and not used for the evaluation of the algorithms. Sky images taken during sunrise and sunset are evaluated by the software as well. However, especially in winter, these images with low elevation angles can be very dark and the algorithms classify the pictures as cloudy while in fact the images may display a clear sky. Therefore, the software excludes images from the evaluation when it is classified as a dark image. They are not used in determining the accuracy of the algorithms and the evaluation by using statistical criteria.

Throughout the irradiance forecasting literature different evaluation metrics are used to determine errors between predicted and measured values. However, one metric, the Root Mean Square Error (RMSE), is used in much of the evaluations (Feng et al., 2017; Reno et al., 2012; Richardson et al., 2017). This metric squares the difference at each timestep before averaging, which gives a relatively high weight to large errors. This metric was used in the research and is shown in the equation below. Furthermore, the Mean Absolute Error (MAE), adopted and adjusted from Feng et al. (2017) was used. This metric measures the average magnitude of errors in a set of forecasts and is shown below.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(CCF_i - CSI_{SW(i)})^2}{N}} \tag{3}$$

$$MAE = \frac{\sum_{i=1}^{N}|CCF_i - CSI_{SW(i)}|}{N} \tag{4}$$

With CCF_i the cloud cover fractions of BRBG, CDOC and TRINITY. $CSI_{SW(i)}$ is the clear sky index (K_t and K_d) at time step i and N is the total number of timesteps. For both metrics a result of zero indicates a perfect evaluation of the cloudiness. As the value for RMSE or MAE increases, the error increases and thus the cloudiness is not evaluated correctly. The metrics were calculated for each algorithm and for each CSI_{SW}. Results are shown in a matrix to give a clear overview of the performances with respect to each other.

Decomposition analysis

After the overall performance of the algorithms was evaluated, a more specific analysis with respect to the sky conditions was done. Hereby it was possible to test the algorithms at specific conditions, e.g. one algorithm might be more accurate at clear sky conditions while another algorithm may perform better when detecting thin clouds/haze conditions. Therefore, the data obtained after step 4 described above, is divided in 3 parts: clear sky, partly cloudy sky and overcast sky. Although a more specific division could be made, e.g. dividing the data in 9 classes of cloudiness, this simple approach is chosen and adapted from (Luiz, Martins, Costa, & Pereira, 2018) since it is hard to determine the boundary between different cloud thicknesses.

To distinguish between different sky conditions the clearness index or diffuse fraction is used. The CCF of the proxy yielding the lowest MAE and RMSE at step 4 of the main analysis is used to classify the data as clear sky, partly cloudy sky or overcast sky. Therefore, the CCF ($1 – K_t$ or K_d) was first converted in to octas. Octas are a method to determine the cloud cover by estimating the fraction of the sky covered with clouds to the nearest eight (Jones, 1992). In this research the CCFs are converted in to octas by using the thresholds described in Luiz et al. (2018). These thresholds, with the corresponding sky conditions, are shown in Table 5. As can be seen from this table the clear sky, partly cloudy and overcast conditions are defined as $x \leq 2$ octas, 3 octas $\leq x \leq 5$ octas and $x \geq 6$ octas, respectively, where x represents $1 – K_t$ or K_d.

Table 5: Octas to Cloud Cover Fraction conversion and the corresponding sky conditions

Octas	Fraction	Sky condition
0	0 ≤ cloud cover < 0.11	Clear sky
1	0.11 ≤ cloud cover < 0.22	Clear sky
2	0.22 ≤ cloud cover < 0.33	Clear sky
3	0.33 ≤ cloud cover < 0.44	Partly cloudy sky
4	0.44 ≤ cloud cover < 0.56	Partly cloudy sky
5	0.56 ≤ cloud cover < 0.67	Partly cloudy sky
6	0.67 ≤ cloud cover < 0.78	Overcast sky
7	0.78 ≤ cloud cover < 0.89	Overcast sky
8	0.89 ≤ cloud cover ≤ 1	Overcast sky

Time-of-day analysis

The results can further be specified when accounting for the time of the day. During sunrise and sunset, the sky imager makes images as well. However, images with low elevation angles are found to have lower blue/red pixel ratios than the ones at higher elevation angles (Huo & Lu, 2009). This increases the chance of errors when detecting clouds during sunrise and sunset. The images in the dataset used which are taken during sunrise and sunset have a lower brightness than images taken during the day. Some of the images are too dark for evaluation and thus the algorithms evaluate this as cloudy conditions, while in fact these are clear sky conditions quite often. Although the software uses an [exclude by Ratio] option, which should exclude dark images from the evaluation, dark or ambiguous images are still used. Moreover, the shortwave irradiance data measured during sunrise and sunset are less reliable than data measured during the day. The high zenith angle affects the irradiance measured which could cause the clearness index and diffuse fraction to fluctuate very strongly. To account for these unreliable periods the sunrises and sunsets are filtered out of the evaluation. The dataset will be reduced to timestamps between 9 am and 3 pm and the accuracies of the algorithms are specified for hour of the day.

Solar position analysis

In addition to the time-of-day analysis the effect of the solar position on the performance of the algorithms is specified as well. Since the sky images are made with a fish-eye lens, visibility at the horizon is lower compared to the centre of the image. Thus, the algorithms responsible for detecting clouds have more trouble accurately detecting clouds in the near-horizon area than clouds which are situated right above the sky imager (CMS, 2016). By plotting the performance of the algorithms (MAE and RMSE) against the elevation angles, the effect of the solar position on the accuracies of the algorithms is assessed.

Sensitivity analysis

Since there was a difference of 19 seconds between the sky images which have been made and the irradiance which has been measured, the irradiance data has been interpolated to account for this error. At every 10 minutes, the irradiance data was interpolated. If an image is partly cloudy, the algorithms will assess it with a CCF of e.g. 0.5. If, then the irradiance is measured through a gap in the cloud cover the CSI_{SW} can assess this image with a CCF of 0. To account for these errors while evaluating the algorithms, the irradiance data was averaged, instead of interpolated. By taking one-hour averages of the irradiance data as well as the CCFs of the algorithms, the results can be recalculated while excluding the short-term variability of the irradiance data.

Longwave irradiance

The downward longwave irradiance data which was measured on site, enables the opportunity to estimate the CCF based on the sky emissivity (CCF_{LW}). The shortwave irradiance emitted by the sun is absorbed by clouds and the earth's surface, these bodies then re-emit the energy in the form of longwave radiation with a wavelength of 4-100 µm and thus reaches the earth as infrared radiation. The amount of longwave radiation reaching the earth's surface thus mainly depends on the presence of water vapor and aerosols in the earth's atmosphere(Cheng & Nnadi, 2014). Therefore, the LW irradiance measured at the earth surface will be highly influenced by clouds. Since this method uses the infrared part of the light's spectrum, it can be used to estimate CCFs even when there is no daylight. In this case study the CCF_{LW} was estimated by comparing the measured LW irradiance with the estimated clear sky emittance. The CCF_{LW} was then used to evaluate the three algorithms.

The method which was used in this case study is based on the determination of CCF using LW radiation data by Luiz, Martins, Costa, & Pereira (2018). It calculates the CCF_{LW} using meteorological parameters like air temperature, relative humidity and atmospheric pressure. The equation that is given for calculating the CCF by using the LW irradiance data is shown below.

$$CCF_{LW} = (LW - LW_{cloudless})/(1 - \in_C) * \sigma * T_C^4 \tag{5}$$

Where LW is the LW irradiance for each timestep. $LW_{cloudless}$ is the lowest hourly mean LW irradiance, i.e. the LW for clear sky days. σ is the Stefan Boltzmann constant ($5.67*10^{-8}$ $Wm^{-2}K^{-4}$), T_C the cloud base temperature and $\in_c$ is the clear sky emittance. The $\in_c$ was calculated with the next equation.

$$\in_C = \frac{LW_{cloudless}}{\sigma * T_a^4} \tag{6}$$

Where T_a is the mean dry bulb temperature for the same period as $LW_{cloudless}$. Since the LW irradiance, dry bulb temperature and cloud base temperature are all measurements, the CCF_{LW} can in some cases exceed the range of 0-1. Therefore, the CSI approach of Marty & Philipona (2000) is used to adjust and normalize the data.

$$e_a = 6.1094 * e^{17.625*T_a/(243.04+T_a)} \tag{7}$$

$$\in_a = \frac{LW}{\sigma * T_a^4} \tag{8}$$

$$\in_C = 1.24 * \left(\frac{e_a}{T_a}\right)^{1/7} \tag{9}$$

$$CSI_{LW} = \in_a / \in_C \tag{10}$$

First the CSI_{LW} is calculated with the equations above. Note that the CSI_{LW} can become higher than 1 and thus is not the same as CSI_{SW}. $\in_a$ is the emittance of the sky and e_a is the vapor pressure at the earth's surface adapted from (Alduchov & Eskridge, 1996). The dry bulb temperature T_a used in equation 7 is in Celsius, whereas in other equations T_a is in Kelvin. Then for every timestep with $CSI_{LW} > 1.05$ the CCF_{LW} was set to 1 and with $CSI_{LW} < 0.89$ the CCF_{LW} was set to 0. These thresholds were adapted from Luiz et al. (2018). If there were still CCF_{LW} values higher than 1, a normalization was applied according to the equation below (Luiz et al., 2018).

$$CCF_{LW}^{normalized(i)} = (CCF_{LW(i)} - \min(CCF_{LW}))/(\max(CCF_{LW}) - \min(CCF_{LW})) \tag{11}$$

The normalized CCF_{LW} values now range from 0 to 1. To calculate the fit of the CCF_{LW} and cloudiness values, the normalized CCF_{LW} values were used to calculate the RMSE and MAE again according to equations 3 & 4 shown , where CCF_i is the $CCF_{normalized(i)}$.

These CCF values, as well as the CCFs of the algorithms, are then converted into octas according to the method explained at the decomposition analysis and the thresholds shown in Table 5. Then the accuracy is determined for the situations where the CCF_{LW} and CCFs of the algorithms have the same octas. This is also done for a difference of 1 and 2 octas.

4. Results

The results section consists of two analysis. The first analysis is based on the SW irradiance method and used the SW data of the first case study. After the data analysis a decomposition analysis, time-of-day analysis and sensitivity analysis are shown. The sensitivity analysis uses the SW irradiance of the second case study as well. Then, the results of the LW irradiance method are shown. This part only uses the LW irradiance obtained from case study II.

Shortwave irradiance

Data analysis

The first step of the case study analysis is calculating the clear sky irradiance at the location where the irradiance measurements are made. Therefore, the Ineichen/Perez CSM is run in MATLAB for the geographical coordinates of Utrecht University. In Figure 2 and Figure 3 the results of the CSM are plotted against the measured irradiance values for a clear sky day and partly cloudy day, respectively.

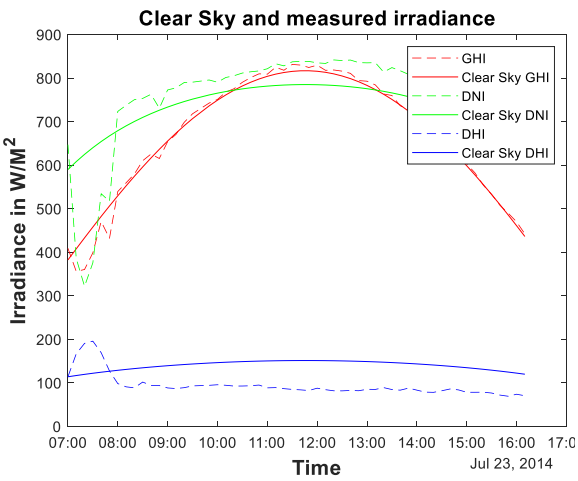


Figure 2: Irradiance for a clear sky day (2014-07-23)

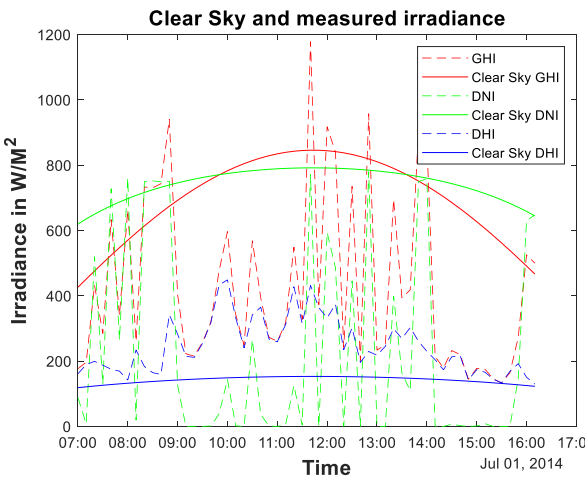


Figure 3: Irradiance for a partly cloudy day (2014-07-01)

As can be seen from the figures above, irradiance values align quite good with the clear sky model at a clear sky day and irradiance values fluctuate very strongly on a partly cloudy day. In Figure 2 measured GHI forms a good fit with clear sky GHI (RMSE = 19.77), where DNI and DHI values deviate slightly from the clear sky model (RMSE = 85.46 and RMSE = 56.17, respectively). This could be due to measurements errors, but more likely the clear sky model is not perfectly accurate. Kato et al. (1997) found that a difference of 5% relative to the measured irradiance is typical, whereas clear sky diffuse irradiance can be overestimated with more than 40%. Although the CSM used is not perfectly accurate small deviations between measured and modelled values are not uncommon. In Figure 3 fast changing cloud conditions cause the irradiance values to fluctuate very strongly, sometimes even leading to super irradiance events.

After the clear sky irradiance values were calculated, the CSIs are determined for each timestamp for which a sky image is available as discussed in step 2. The distribution of the clearness index and diffuse fraction for two years of irradiance data are shown in the histograms of Figure 4 and Figure 5, respectively.

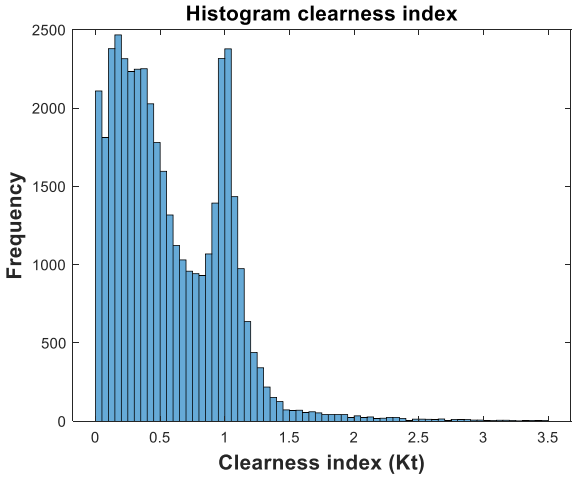


Figure 4: Histogram for the clearness indices

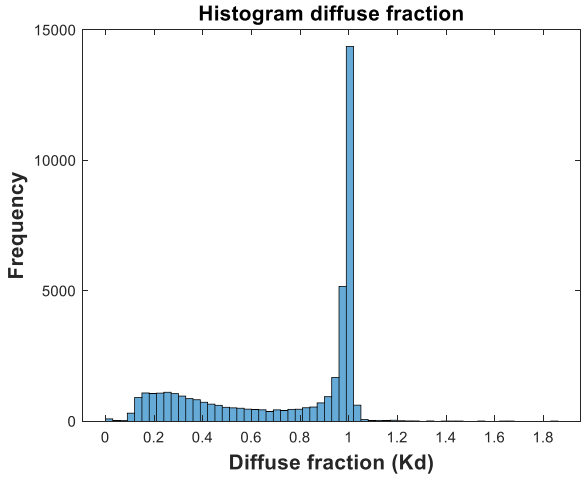


Figure 5: Histogram for the diffuse fractions

From these figures it can be seen that the clearness index mostly varies between 0 and 0.5 and that most K_d values are near a value of one. This makes sense, since the Netherlands has 44 clear sky days on average according to the clear sky days defined in Table 5, whereas the average number of overcast days is 205 per year. The rest of the days are partly cloudy (KNMI, 2019). From Figure 4 one can see that super irradiance events occur often, especially K_t values higher than 1.5 cannot be due to the clear sky model errors and most certainly result from super irradiance events.

Then, the FINDCLOUDS and TRINITY software are used to evaluate the sky images. For each timestamp the cloudiness is given by the BRBG, CDOC and TRINITY algorithm. In Figure 6 until Figure 9 sky images are shown, including the evaluations of the three algorithms. This is done for a clear sky, thin clouds/haze conditions, partly cloudy and complete overcasts conditions, respectively. All images are manually selected and are representative for the particular cloud conditions. The Figures show that the BRBG algorithm only distinguishes clear sky (blue) or clouds (grey), the CDOC algorithm distinguishes clear sky (blue), thick clouds (grey) and thin clouds (white) and the TRINITY algorithm distinguishes multiple layers of cloud cover.

From Figure 6 one can see that all algorithms perform good at clear sky conditions. The blue sky is clearly visible for the software and the BRBG and CDOC algorithms show almost zero cloudiness. The TRINITY algorithm has difficulties with classifying the near sun area. The refraction of the light causes a slightly lighter colour at the bottom part of the image. The algorithm classifies this as (thin) clouds. This is not only an error occurring at the TRINITY algorithm. The BRBG and CDOC algorithms have this as well, though this mostly occurs at sky images with low elevation angles, i.e. especially in the winter months. Figure 7 displays thin clouds or hazy conditions which is found to be one of the most challenging events;. Since this is a mix up of clear sky and cloudy conditions, the results of the algorithm vary widely. As can be seen the BRBG algorithm classifies the image with 0.64, where the CDOC and TRINITY algorithms value the image at 1.00 and 0.95, respectively. Based on manual observation, the TRINITY algorithm appears to be evaluating the image most accurately, which is due to the multiple layer classification. At Figure 8 a partly cloudy image is shown. All algorithms perform good in this case as can be seen from the cloudiness values which don't deviate a lot from each other. The clear contrast of sky and clouds in these conditions enable the algorithms to evaluate partly cloudy images accurately. Figure 9 shows the image and evaluation of a complete overcast event. The BRBG and TRINITY algorithms classify the image accurately, both with a CCF of 1.00. the CDOC algorithm, however, misclassifies a part of the near sun area. As a result, the CDOC algorithm falsely classifies 7% of the image as clear sky. This is an error which occurred often when using the CDOC algorithm and can be seen as typical for this algorithm, the BRBG and TRINITY algorithm did not have this kind of error.

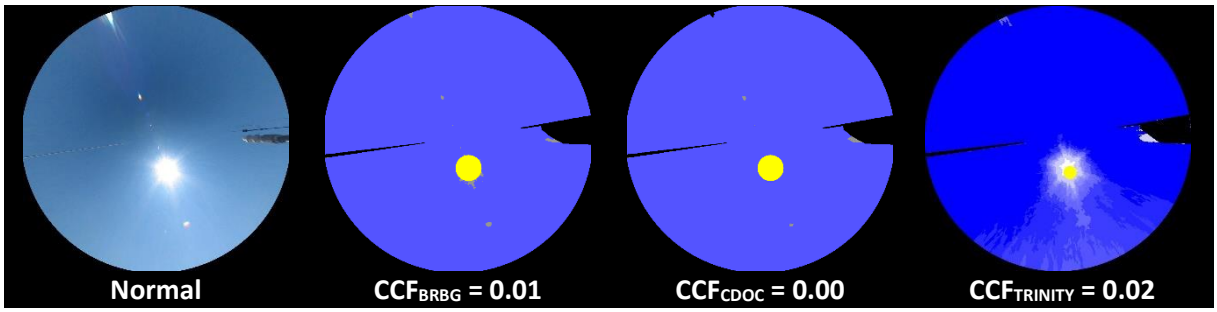


Figure 6: Evaluations of a clear sky image (2014-07-23 11:00:00)

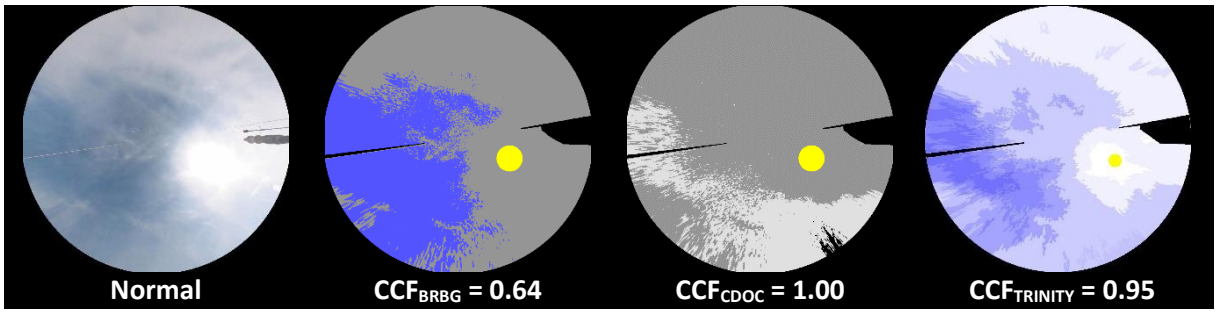


Figure 7: Evaluations of a thin cloud/hazy sky image (2014-07-19 08:30:00)

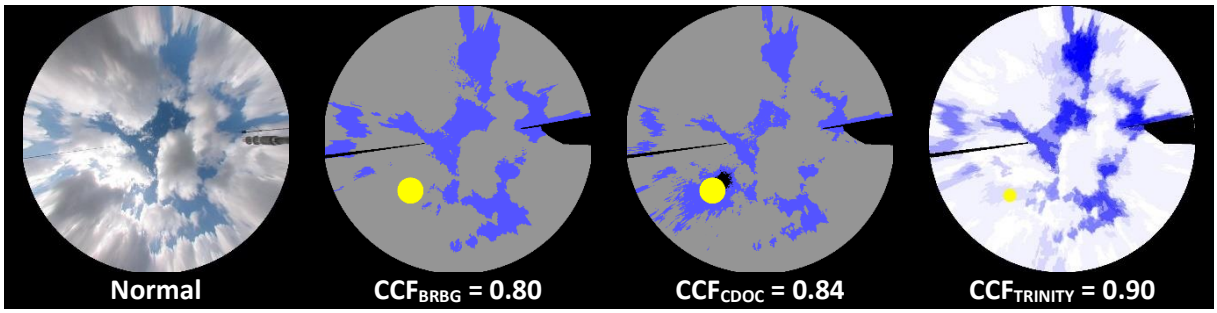


Figure 8: Evaluations of a partly cloudy sky image (2015-04-01 14:00:00)

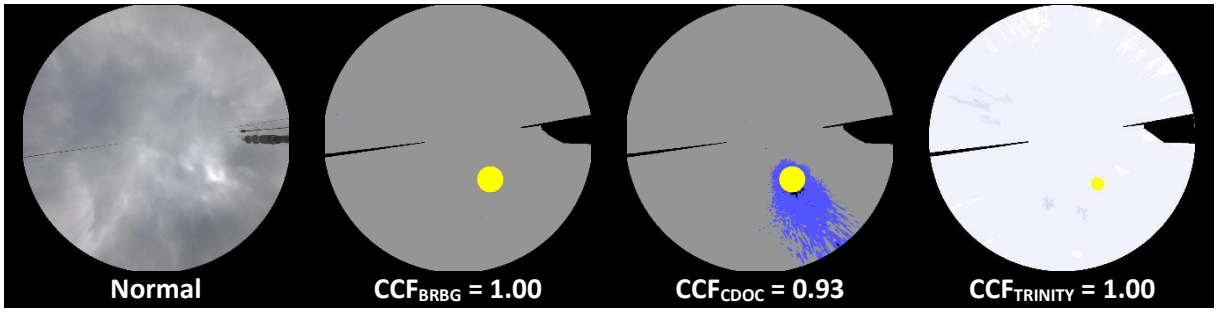


Figure 9: Evaluations of a complete overcast sky image (2015-04-17 09:50:00)

Next, the results of all the evaluations are combined and shown in Figure 10. The boxplots in this Figure summarize the CCFs of the algorithms. The histograms provide additional information regarding the distribution of the data. All three algorithms reach the highest frequency for a cloudiness level near a value of one. This is in line with the distributions of the clearness index and diffuse fraction as shown in Figure 4 and Figure 5, which also show that overcast conditions are more usual than clear sky conditions. As the boxplots and histograms show, the CDOC and TRINITY algorithms differ from the BRBG algorithm in the sense that the data is more concentrated near the value of one. For the BRBG algorithm the middle 50% of the data has a lower bound of 0.3, where for the CDOC and TRINITY algorithm this is around 0.7 and 0.6, respectively. To quantify the differences between the algorithms, the MAE and RMSE are calculated for all combinations of the algorithms. The results are shown in Table 6. As can be seen, both metrics yields the lowest error for the comparison CDOC-TRINITY, while

the highest errors are the combinations of CDOC and TRINITY with BRBG. Thus, the evaluations of the CDOC and TRINITY algorithm are much closer related with each other, than when each of these algorithms is compared with the BRBG algorithm. These results align with the results from Figure 10 in which the CDOC and TRINITY algorithms assign higher cloudiness values to the evaluated images.

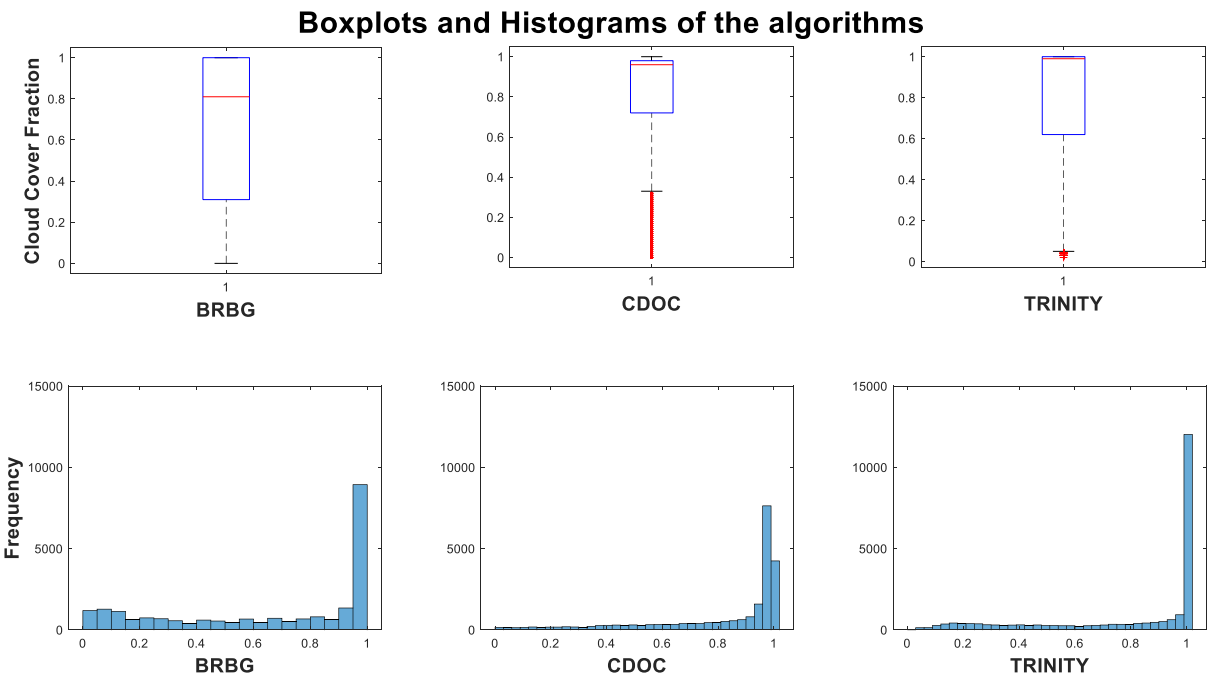


Figure 10: Data distribution of the algorithm's evaluations for the entire dataset

Table 6: Differences of the three algorithms with respect to each other (MAE and RMSE)

	Mean Absolute Error			Root Mean Square Error		
	BRBG	**CDOC**	**TRINITY**	**BRBG**	**CDOC**	**TRINITY**
BRBG	x	0.1680	0.1326	x	0.2667	0.2123
CDOC	0.1680	x	0.0814	0.2667	x	0.1577
TRINITY	0.1326	0.0814	x	0.2123	0.1577	x

Before the CSIs and the CCFs of the algorithms can be compared, extreme data has to be filtered out according to the thresholds described in the methodology. Then, all the remaining timestamps are used for further calculations. For each algorithm two scatter plots are shown in Figure 11, one using the clearness index and one using the diffuse fraction. From these figures one can see that overcast conditions (top right corners) occur much more than clear sky conditions (bottom left corners) in Utrecht. This makes sense, since the Netherlands have an oceanic climate and cloudy conditions are more common than clear skies. On average, only 14% of the days are clear sky days in the Netherlands (KNMI, 2019).

In Figure 11 the difference between the clearness index and diffuse fraction for validating the algorithms are shown. All Figures on the left side (clearness index) are scattered more widely than the Figures on the right side (diffuse fraction). The top left corners, and to a lesser extent the bottom right corners, have a higher density of data points for the clearness index than for the diffuse fraction. This means that the clearness index is less correlated with the CCFs of the algorithms than the diffuse fraction. This is best visible for clear sky conditions. The Figures on the left seem to tend to a random distribution, whereas the Figures on the right seem to have a funnel structure. So, when looking at a

proxy for the comparison of the algorithms with a CSI_{SW}, the diffuse fraction is more suitable than the clearness index.

When looking at the scatter plots of the diffuse fraction one can see that all algorithms perform good at clear sky conditions. In the bottom left corners all data points are concentrated and form the beginning of a funnel. As the cloudiness increases the funnel structure, which ideally should be a linear line from [0,0] to [1,1], becomes wider. This holds especially for the BRBG algorithm. So, conditions with a cloudiness higher than around 0.3 are prominently responsible for the scattering in the Figures. Finally, all algorithms have a high density of data points in the top right corners. This are the overcast conditions and all algorithms appear to evaluate most of these situations correctly. These findings settle with the results of the visual inspection, where the thin clouds/haze conditions are seen as most challenging for the algorithms.

When taking a closer look at the differences between the algorithms, it can be seen that all algorithms have a similar shape. However, the BRBG algorithms has a funnel structure which is less abrupt, or steep, than the CDOC and TRINITY algorithms. This means that a higher diffuse fraction results in a higher cloudiness for the CDOC and TRINITY algorithms compared to the BRBG algorithm. Thus, as the amount of clouds increase the effect on the cloudiness of the CDOC and TRINITY is higher than on the BRBG. This could be due to, e.g. the fact that the CDOC and TRINITY algorithms are able to classify multiple cloud thicknesses. Furthermore, the BRBG has a lot of data points in the bottom right corner which decreases the accuracy of this algorithm. The CDOC on the other hand, has for the full range of diffuse fractions a lot of data points where the cloudiness is 1. These problems occur less when using the TRINITY algorithm. However, for a more accurate comparison one should look at the results of the metrics explained in the methodology.

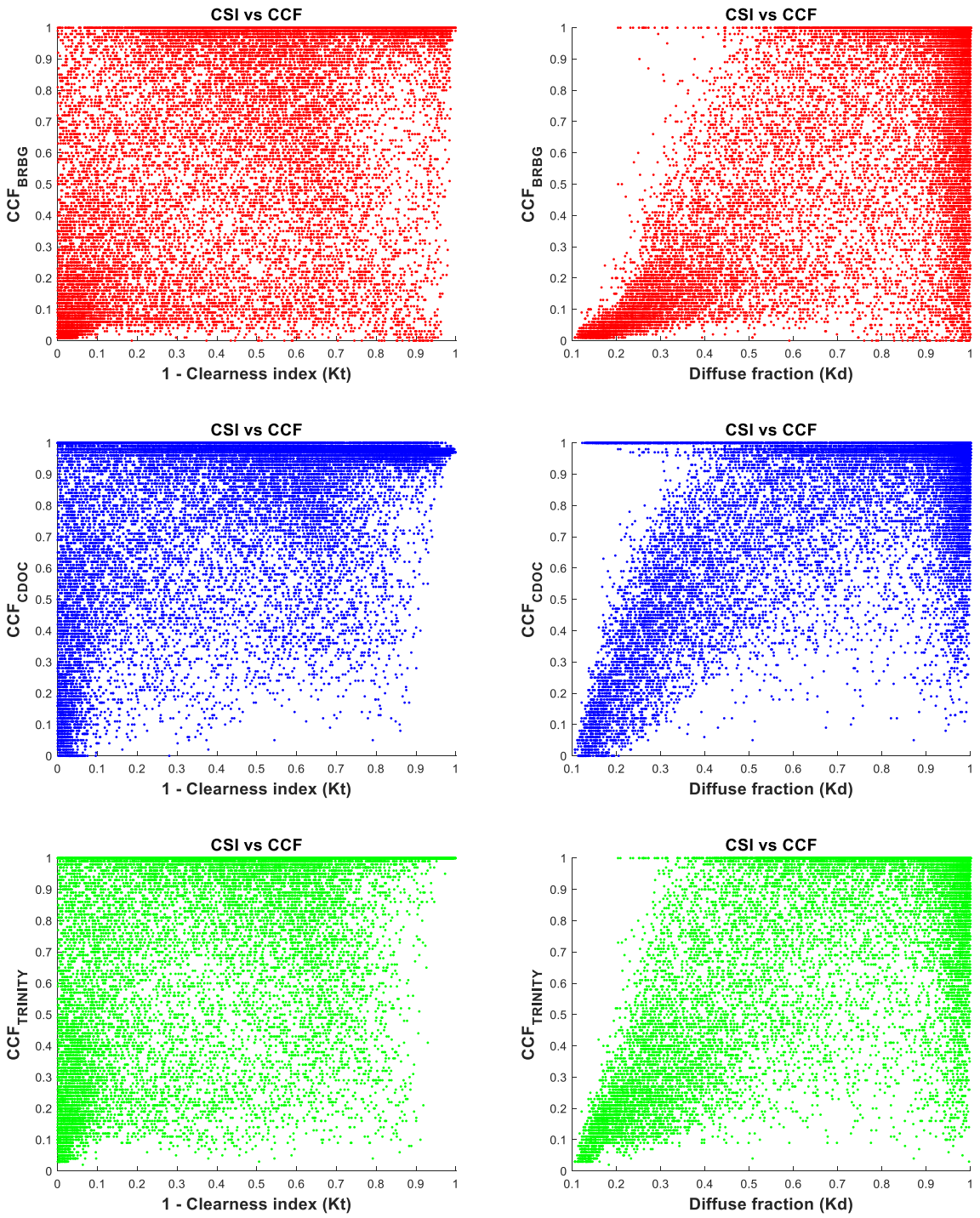


Figure 11: Scatter plots of CSIs against CCFs of the algorithms

The data from the scatter plots in Figure 11 is then used to calculate the MAE and RMSE. Results are shown in Table 7. As can be seen from this table the values for the RMSE are higher than the MAE, which is due to the squaring of the error in the RMSE metric. Differences between the two datasets are thus enhanced in the RMSE metric. According to the RMSE, the BRBG algorithm performs best when compared with the clearness index and the TRINITY algorithm performs best when compared with the diffuse fraction. The MAE is lowest for the BRBG algorithm compared with the clearness index and TRINITY performs best compared with the diffuse fraction. Furthermore, the results in the table support the statement that the diffuse fraction is a better proxy than the clearness index. For each metric calculated, the use of the diffuse fraction results in a lower value and thus a smaller error

between the two datasets. Besides, the differences between the three algorithms is smaller when using the diffuse fraction compared with the clearness index. Thus, the diffuse fraction is a better proxy for evaluation the algorithms in this analysis than the clearness index.

When taking a look at the MAE with the diffuse fraction, one can see that the TRINITY algorithm performs best with an average error of 12.18%. Although this is considerably better than the BRBG algorithm, the CDOC algorithm also performs similar with an average error of 13.74%. The performances of the algorithms compared to each other does not change according to the RMSE. Only the differences between them are increased. The MAE values calculated with the clearness index are much higher than the values calculated with the diffuse fraction. Since these values vary between 26.74% and 33.31% it can be stated that using the clearness index as CSI_{SW} is far less accurate than using the diffuse fraction.

Table 7: RMSE and MAE of the three different algorithms (Case Study I)

MAE	CLEARNESS INDEX	DIFFUSE FRACTION
BRBG	0.2674	0.1675
CDOC	0.3331	0.1374
TRINITY	0.3218	0.1218
RMSE		
BRBG	0.3387	0.2579
CDOC	0.4150	0.2189
TRINITY	0.3940	0.1979

Decomposition analysis

Since the results of Table 7 show that the diffuse fraction yields lowest MAE and RMSE, the K_d is used to classify the data into clear sky, partly cloudy sky and overcast sky as described in the methodology. The results of the decomposition analysis are show in Table 8. From this table one can see that the partly cloudy situations are the most challenging to evaluate by the algorithms. This category includes the thin clouds/hazy conditions as shown in Figure 7. As described earlier these sky conditions are hard to be recognized by the algorithms, since it is a mix of clear sky and cloudiness. This is supported by the fact that all algorithms have the highest MAE at the partly cloudy sky. The error values under these conditions are similar for the different algorithms. Furthermore, the RMSEs at partly cloudy sky are all remarkably high, compared to other sky conditions. The BRBG algorithm performs average at clear sky and partly cloudy conditions, but the performance at an overcast sky is worse than the other algorithms, where the MAE is 76% higher than the other algorithms. The RMSE is also significantly higher at these conditions (66%). The CDOC algorithm, however, performs best at overcast conditions, while performing bad at clear skies. Where the MAE is 54% higher and the RMSE is 95% higher than the other algorithms, other algorithms perform better at clear sky conditions. Finally, the TRINITY algorithm outperforms BRBG and CDOC at clear sky conditions and in overcast situations outperforms BRBG, while the difference with the CDOC remain very low. Overall, the TRINITY algorithm achieves the best results for clear sky situations, all algorithms perform similarly at partly cloudy conditions, while the CDOC and TRINITY achieve the lowest error at overcast conditions.

Table 8: Results of the decomposition analysis: algorithm performance at different sky conditions

MAE	Clear sky	Partly cloudy sky	Overcast sky
BRBG	0.1430	0.2201	0.1625
CDOC	0.2200	0.2679	0.0837
TRINITY	0.1091	0.2541	0.0922
RMSE			
BRBG	0.1611	0.2587	0.2771
CDOC	0.3235	0.3148	0.1425
TRINITY	0.1657	0.3010	0.1672

Time-of-day analysis

Next to the interpolation error, the effect of dark images in the evaluation was also examined. Table 9 shows the results of the analysis when excluding the sunrise and sunset. In this analysis only data which was measured between 9 am and 3 pm was used. As can be seen from this table, the values of the clearness index do not significantly differ when compared to the original evaluation. When looking at the diffuse fraction, however, the MAEs have decreased with 11%, 15% and 2% for the BRBG, CDOC and TRINITY, respectively. All values for the MAE have thus become lower. This means that excluding the periods with low elevation angles improves the accuracy of the algorithms, where the effect is largest for the BRBG and CDOC. This is in line with what was expected, namely that a low elevation angle causes unreliable shortwave irradiance data and that the algorithms would perform worse when the image is darker. To elaborate more on the effect of the time of day and to look closer at the effect on the different algorithms, the MAEs for all algorithms are specified for the hour of the day in Figure 12. The diffuse fraction is used since the MAEs are much lower compared with the clearness index. In the figure, the value of the hour of the day e.g. 7 represents all MAEs calculated for data obtained between 7 pm and 8 pm throughout the entire dataset.

Table 9: Results of sensitivity analysis (including/excluding sunrise and sunset)

	Including Sunrise and Sunset		Excluding Sunrise and Sunset	
MAE	**Clearness index**	**Diffuse fraction**	**Clearness index**	**Diffuse fraction**
BRBG	0.2674	0.1675	0.2898	0.1485
CDOC	0.3331	0.1374	0.3315	0.1173
TRINITY	0.3218	0.1218	0.3440	0.1190
RMSE				
BRBG	0.3387	0.2579	0.3617	0.2334
CDOC	0.4150	0.2189	0.4044	0.1796
TRINITY	0.3940	0.1979	0.4166	0.1942

From Figure 12 one can see that all algorithms obtain the lowest errors somewhere between 9 am and 3 pm. After 3 pm the MAEs of all algorithms increase, especially the error of the BRBG algorithm is prominent. After a small drop at 7 pm, for which there does not seem to be a logical explanation, the errors are very high at 8 pm. When looking at the algorithms, one can see that the BRBG algorithm has the highest errors for most of the hours per day, except for the early morning and late evening. During these times the BRBG algorithm performs better than the CDOC algorithm. This CDOC algorithm has high errors at sunrise and sunset, however during the day it achieves the lowest errors compared to the other algorithms. The TRINITY algorithm seems to perform good in the morning as well as during the day. Only evaluations obtained after 8 pm acquire a high MAE, however, the low amount of data points for this hour is decreasing the reliability.

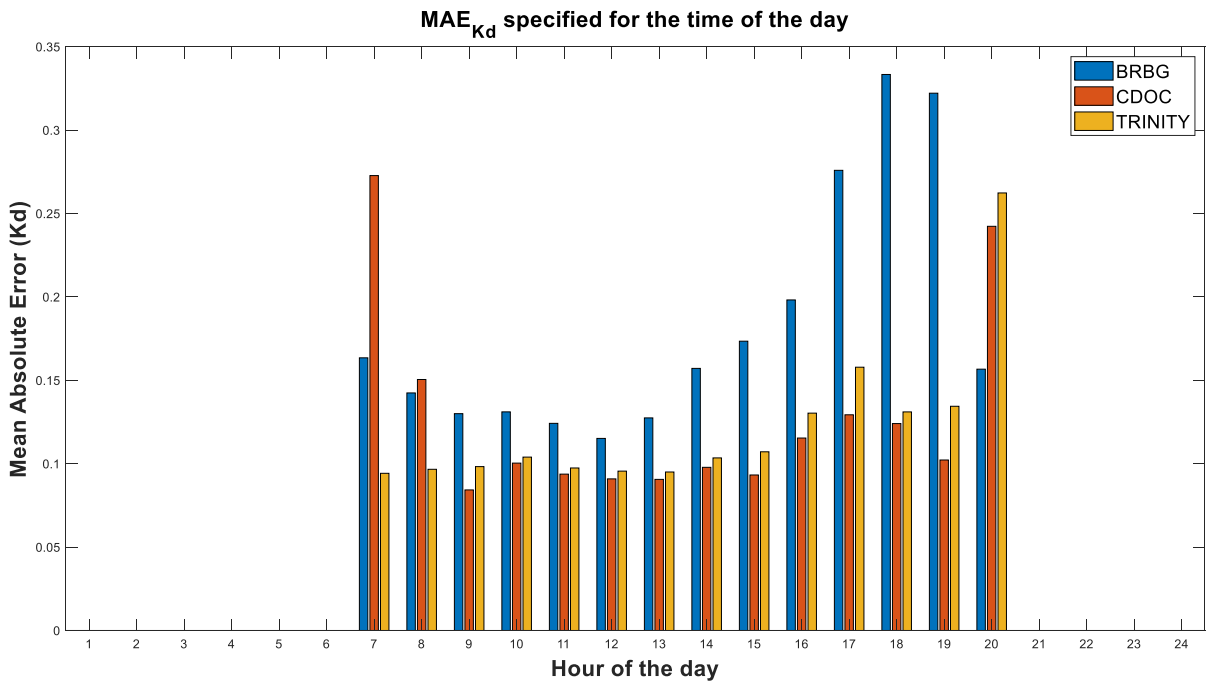


Figure 12: Performance of the algorithms specified for each hour of the day, throughout the entire dataset

Solar position analysis

Since the solar position changes for each hour of the day throughout the year, the MAEs and RMSEs are plotted against the elevation angles in Figure 13 and Figure 14, respectively. From these figures one can see that the BRBG has the highest errors at low elevation angles. The CDOC algorithm outperforms the TRINITY algorithm for the very low elevation angles, whereas it yields higher errors than the TRINITY algorithm for elevation angles of 10° to 35°. For elevation angles of 35° and higher, all algorithms have a similar performance and the effect of the solar position on the performance of the different algorithms seems to be strongly reduced.

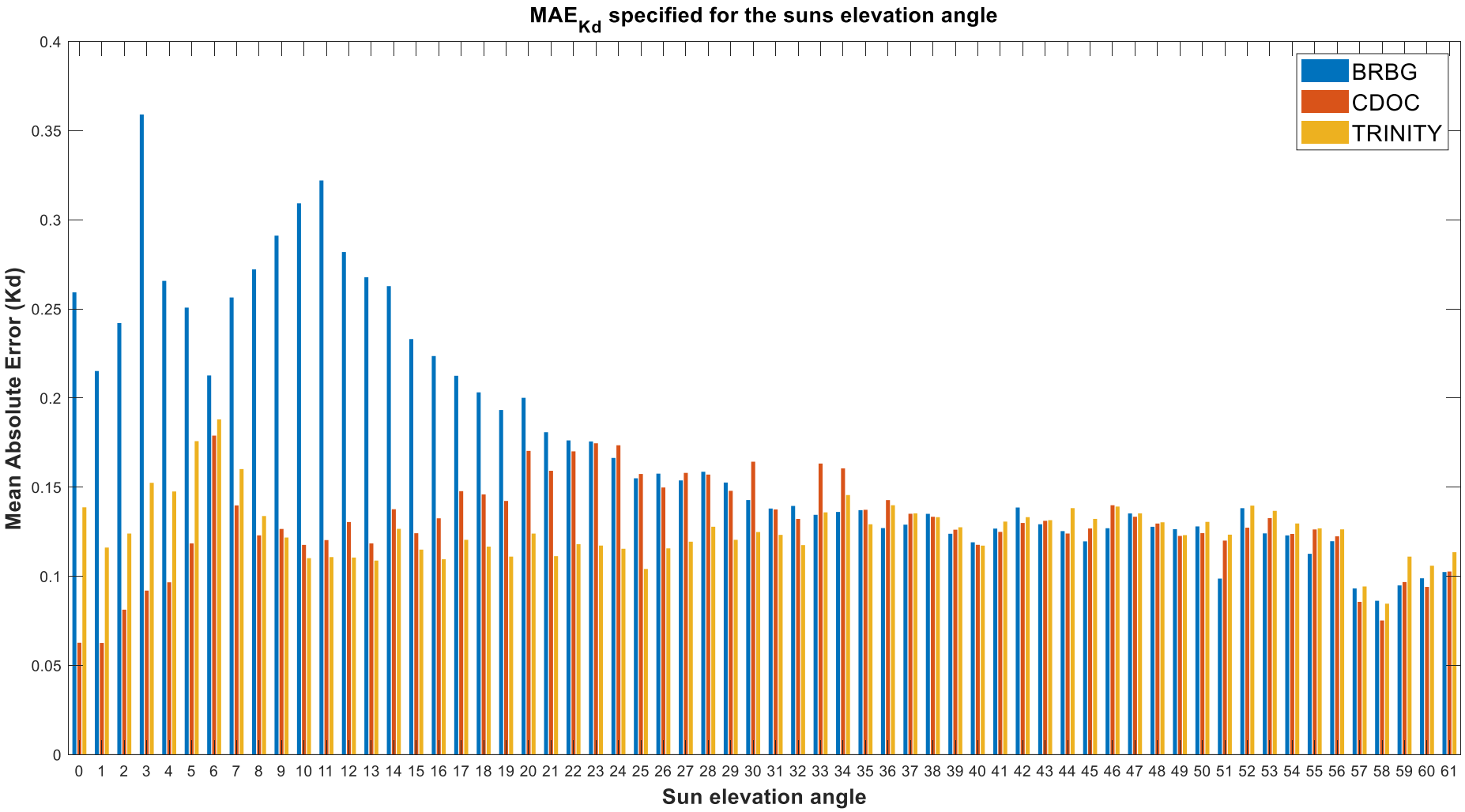


Figure 13: MAEs of the algorithms specified for each elevation angle, throughout the entire dataset

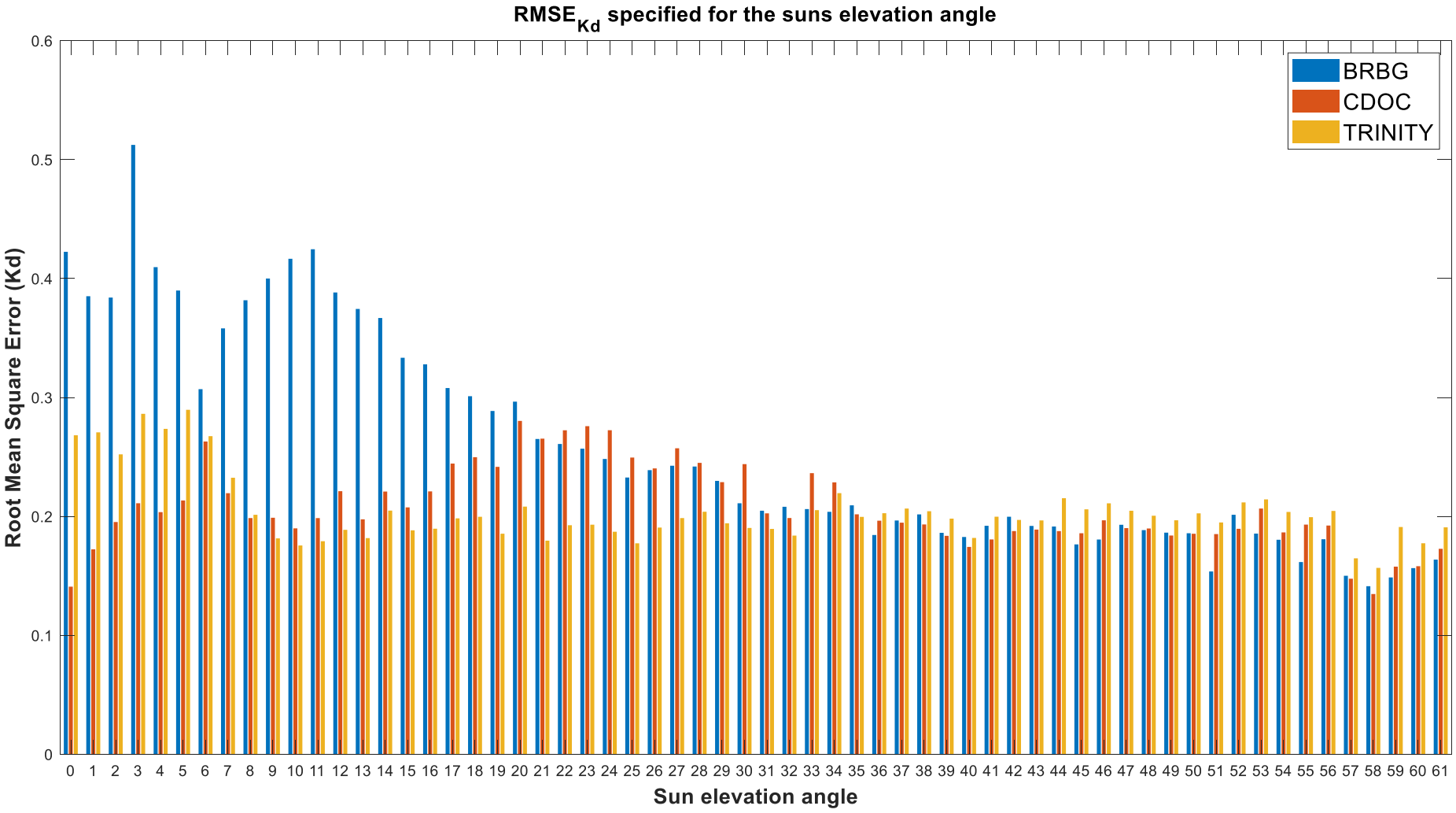


Figure 14: RMSEs of the algorithms specified for each elevation angle, throughout the entire dataset

Sensitivity analysis

To account for the error which could occur when interpolating the irradiance data, the results of the analysis when using hourly means is shown in Table 10. In the analysis the CCFs of the algorithms as well as irradiance data is averaged over the hour. The table shows that both CSI_{SW} do not change a lot. When looking at the clearness index the BRBG algorithm still has the smallest errors. When comparing the diffuse fractions of both analyses one can see that the results have been levelled. The use of the hourly means for all of the data results in a difference of MAE less than 0.03 per algorithm. Though CDOC has a slightly lower error than the TRINITY and BRBG algorithms, all algorithms have a similar performance. The BRBG algorithm has endured the highest decline of the errors (-21% and -24%), which could indicate that these algorithm outputs fluctuate more strongly than the other algorithms. For the CDOC and TRINITY algorithms the differences are smaller.

The SW irradiance measurements of case study II are measured 19 seconds after the sky images have been made. Therefore, the SW irradiance values are interpolated to make the analysis possible. To further examine the effect of this interpolation on the results, the SW irradiance method is applied to the data of case study II as well, since for this case study all data collection is synchronized no interpolation is needed. The results are shown in Table 11, along with the results of case study I. The results in the table show that the errors are higher for case study II than the first case study, though the differences are small. When looking at the diffuse fraction the influence of interpolation in case study I is neglectable, since the results are more accurate than the results of case study II. When looking at the clearness index the results are more accurate in case study II than in case study I, however the errors are still higher as when using the diffuse fraction. Furthermore, the results of both case studies are similar with respect to the different algorithms. When using the clearness index as a proxy, the BRBG algorithm performs best in both case studies. When using the diffuse fraction as a proxy the ranking of the algorithm is identical for both case studies. Although the values for the MAE are a bit higher, the TRINITY algorithm still outperforms the BRBG and CDOC. Since the values of the metrics shown in Table 10 do not change a lot when averaging the data for each hour and the results in Table 11 show that the errors of case study I (interpolation) are lower than the results of case study II (no interpolation), it becomes clear that the interpolation error is small. The variability of the irradiance data thus only has a minor influence on the result of the analysis.

Table 10: Results of sensitivity analysis (10-minute interpolation vs hourly mean)

	10-minute interpolation		**Hourly mean**		**Difference**	
MAE	**Clearness index**	**Diffuse fraction**	**Clearness index**	**Diffuse fraction**	**Clearness index**	**Diffuse fraction**
BRBG	0.2674	0.1675	0.2972	0.1330	11%	-21%
CDOC	0.3331	0.1374	0.3399	0.1281	2%	-7%
TRINITY	0.3218	0.1218	0.3483	0.1300	8%	7%
RMSE						
BRBG	0.3387	0.2579	0.3601	0.1956	6%	-24%
CDOC	0.4150	0.2189	0.4116	0.1789	-1%	-18%
TRINITY	0.3940	0.1979	0.4165	0.1843	6%	-7%

Table 11: Results of both case studies for the SW irradiance method

	Case study I		Case study II		Difference	
MAE	**Clearness index**	**Diffuse fraction**	**Clearness index**	**Diffuse fraction**	**Clearness index**	**Diffuse fraction**
BRBG	0.2674	0.1675	0.2166	0.2105	-19%	26%
CDOC	0.3331	0.1374	0.2944	0.1714	-12%	25%
TRINITY	0.3218	0.1218	0.3038	0.1637	-6%	34%
RMSE						
BRBG	0.3387	0.2579	0.3023	0.3142	-11%	22%
CDOC	0.4150	0.2189	0.3747	0.2540	-10%	16%
TRINITY	0.3940	0.1979	0.3967	0.2514	1%	27%

Since in case study I the difference in the MAE and the RMSE calculated with the clearness index is less significant when using the hourly mean of all data when compared to the diffuse fraction, it can be stated that this index is less correlated with the CCFs of the algorithms than the diffuse fraction is. Figure 15 shows the scatter plots of CCFs and clearness index and diffuse fraction for hourly means, which are similar to Figure 11. As can be seen from this figure, the scatter plots on the left (clearness index) still not correlate with the CCFs as does the diffuse fraction (plots on the right). The plots using the clearness index tend towards a random distribution. Since taking the hourly mean of the data yields similar results as 10-minute interpolation it can be stated that the clearness index is less correlated with the algorithms output. Therefore, results obtained by using the diffuse fraction should mainly be used to determine meaningful conclusions.

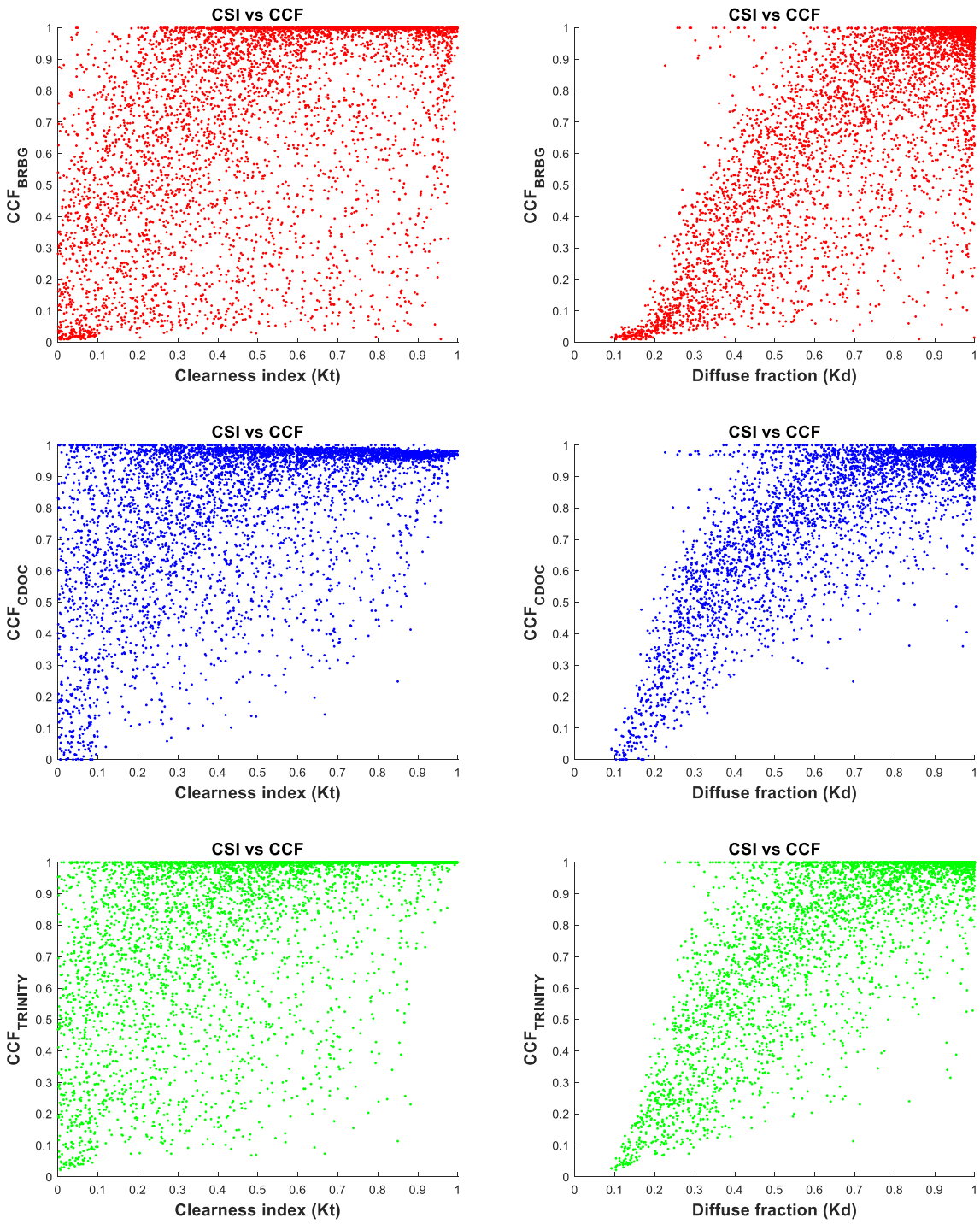


Figure 15: Scatter plots of CSIs against CCFs of the algorithms (by using hourly means of all the data)

Longwave irradiance

Data analysis

The CCFs and CSIs (LW) are calculated and shown in Figure 16 below. As can be seen there is a clear distribution between clear sky situations ($CCF_{LW} = 0$) and cloudy situations ($CCF_{LW} = 1$). The CSI used for this longwave irradiance data distinguishes between clear sky ($CSI_{LW} < 1$) and cloudy sky ($CSI_{LW} > 1$) as was discussed by Marty & Philipona (2000). From this figure one can see that clear sky conditions and complete overcast conditions dominate the dataset used.

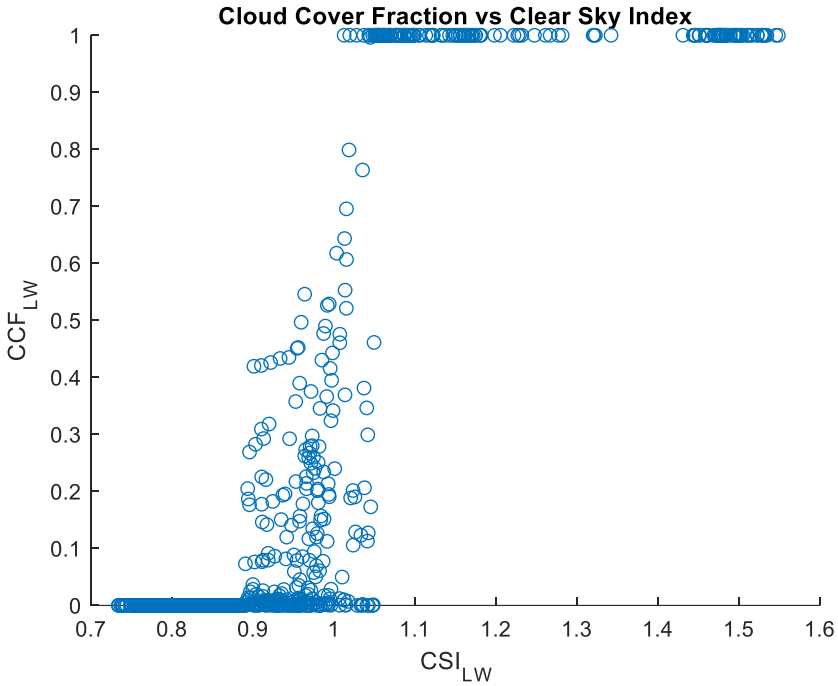


Figure 16: Relationship between the normalized Cloud Cover Fraction CCF_{LW} and the Clear Sky Index CSI_{LW}

After the CCFs have been calculated, the sky images of the new location are evaluated by the algorithms. To assess the evaluation results of the algorithms, but also the accuracy of the LW irradiance method used, Figure 17 till Figure 20 display four situations (clear sky, thin clouds/haze, partly cloudy and overcast conditions) as was done in case study I. Figure 17 shows the results for a clear sky situation. As can be seen all algorithms assign a value higher than 0. The near sun area is hard to evaluate, especially at images with high zenith angles. In Figure 18 a thin clouds/hazy situation is displayed. Again, the scattering of the light in the near sun area is affecting the output of the algorithms. Where the BRBG and CDOC classify the rest of the sky as clear sky, the TRINITY algorithm classifies this as hazy. Based on the visual inspection of this image, the TRINITY algorithms seem to perform better at these challenging conditions, whereas according to the LW method the BRBG algorithm seems to perform best. However, the LW methods seems not to be correct here, since clouds can be observed visually while the CCF_{LW} is zero. In Figure 19 the sun is already out of the scope of the image. All algorithms seem to evaluate this image correctly, however the CDOC and TRINITY have a higher cloudiness than BRBG. Especially at the cloud edges and thinner cloud sections the CDOC and TRINITY look better. Finally, the BRBG and TRINITY algorithm evaluate the overcast condition in Figure 20 perfectly. The CDOC, however has a small error. The area near the sun is evaluated as clear sky, which is probably due to an incomplete CSL. This has also affected the CDOC evaluation in the previous images.

For each situation in Figure 17 till Figure 20 the CCF_{LW} is shown. For the clear sky and overcast images, the CCF_{LW} was 0 and 1, respectively. Despite these values being perfectly accurate for this conditions, the thin clouds and partly cloudy situations are less accurate. The thin cloud situation has a CCF_{LW} of 0. Since thin clouds occur on this image, the CCF_{LW} value is not right. However, the range of the algorithm evaluations (0.22-0.67) already show that results can vary widely in these conditions. Next, the calculated CCF_{LW} is 0 for the partly cloudy situation as well. There are clearly thick clouds here and the algorithms evaluations are high (0.58-0.78). Therefore, there seems to be an error occurring in the data measured or in the method used to calculate the CCF_{LW}.

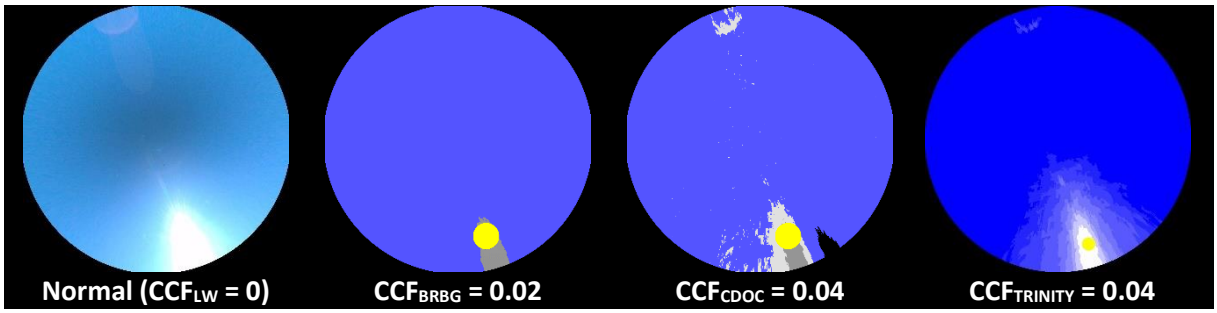


Figure 17: Evaluations of a clear sky image (2018-01-02 11:00:00)

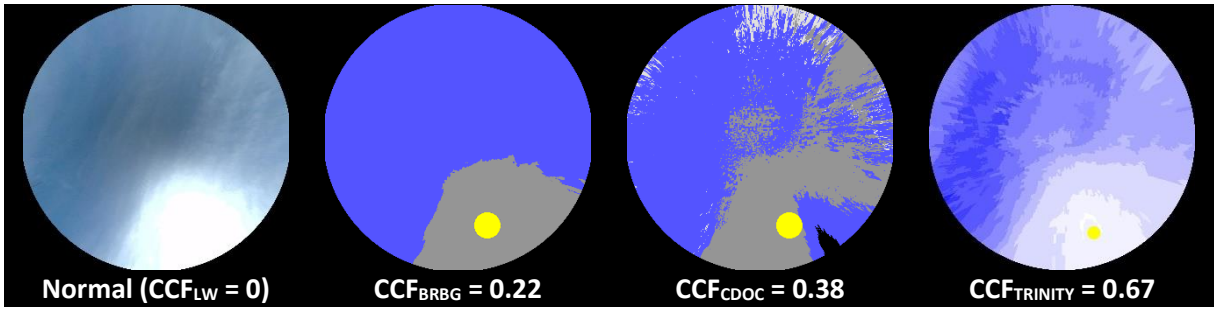


Figure 18: Evaluations of a thin cloud/hazy sky image (2018-01-18 11:00:00)

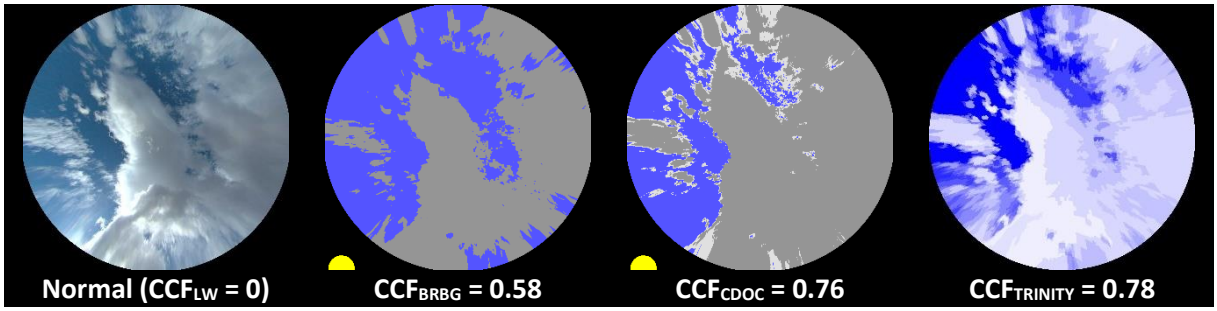


Figure 19: Evaluations of a partly cloudy sky image (2018-01-05 15:40:00)

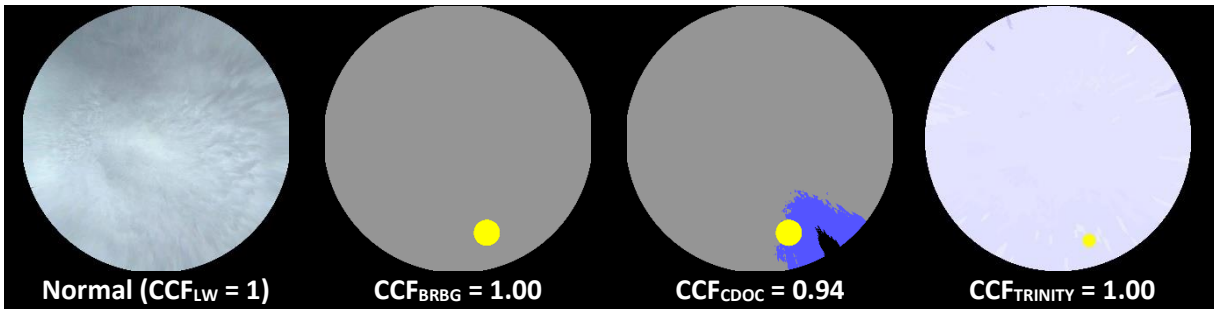


Figure 20: Evaluations of a complete overcast sky image (2018-01-08 11:00:00)

After the algorithms have evaluated the sky images the output is used to calculate the accuracy. Therefore, the CCF_{LW} and CCFs of the algorithms are first categorized in octas. Then the accuracy is calculated by counting the amount of times the CCF_{LW} and algorithms displayed the same cloudiness in octas, this is also done for ± 1 octas and ± 2 octas. Results are shown in Table 12. This table shows that the TRINITY algorithms is most accurate, although the differences are only 3.25% and 1.47% compared to BRBG and CDOC, respectively. When the accuracies are measured with a difference of 1 or 2 octas the performance increases substantially. However, differences between the algorithms stay constant or decrease.

Table 12: Accuracies of the algorithms according to the CCF_{LW}

	Accuracy (%)		
	(± 0 octas)	**(± 1 octas)**	**(± 2 octas)**
BRBG	52.67	65.65	73.51
CDOC	54.45	68.17	76.02
TRINITY	55.92	69.21	77.07

To assess the quality of the accuracies calculated above, the MAE and the RMSE are also calculated for this data set. Therefore, the actual CCF_{LW} and CCFs of the algorithms are used, i.e. not the octas. In this way, the error between CCF_{LW} and algorithm evaluation can be determined most accurately. The results are shown in Table 13. The MAE of the BRBG algorithm is lowest, followed by the TRINITY algorithm. The MAE of the CDOC algorithm is substantially higher than the other algorithms, which is probably due to the incomplete CSL as discussed earlier. When looking at the RMSE however, the CDOC and TRINITY algorithm yield comparable results.

Table 13: RMSE and MAE of the three different algorithms (Case Study II)

MAE	**CCF_{LW}**
BRBG	0.2283
CDOC	0.3299
TRINITY	0.2596
RMSE	
BRBG	0.3717
CDOC	0.4606
TRINITY	0.4750

5. Discussion

With an increasing amount of intermittent renewable electricity generation in to the power mix, the importance of accurate and reliable energy production forecasts becomes more important. For the short term (< 30 min) sky imaging is one of the most important methods of forecasting PV generation. A method which is widely used in PV-farms. This research aimed at studying the performance of various cloud detecting algorithms in combination with an ASI provided by EKO Instruments. By assessing the performance of the new algorithm with respect to previously defined algorithms, new insight is gained in the field of sky imaging and the knowledge base of cloud detection algorithms is enlarged.

The SW irradiance method is based on two years of measurements with a total of 22.973 images used after the data was filtered. The visual inspection at the SW irradiance method showed that the BRBG and TRINITY algorithms performed best, while the CDOC algorithm had trouble with evaluating thin clouds and overcast conditions. The MAE and RMSE turned out to be substantially lower when using the diffuse fraction compared to the clearness index, furthermore the scatter plots as shown in Figure 11 and Figure 15 show that the output of the algorithms correlate much better with the diffuse fraction than with the clearness index. The results when using the diffuse fraction show that the TRINITY algorithm performs best.

The additional decomposition analysis shows that the BRBG algorithm performs worse at overcast conditions, with a 76% and 66% higher MAE and RMSE compared to the other algorithms, respectively. The CDOC algorithm as trouble identifying clear sky conditions. With a MAE 54% higher and a RMSE 95% higher than the other algorithms, the BRBG and TRINITY algorithms have a higher accuracy at clear sky conditions than the CDOC algorithm. The TRINITY algorithm is more stable than other algorithms and shows good performance at all sky conditions (see Table 8). The time-of-day analysis shows that when looking at the diffuse fraction, excluding the sunrise and sunset improves the accuracies with 11%, 15% and 2% for the BRBG, CDOC and TRINITY, respectively. The solar position analysis shows that the BRBG algorithm is most sensitive to low elevation angles, leading to higher errors. The TRINITY algorithm achieves similar performance for all elevation angles and is more constant than the other algorithms. For elevation angles of 35° and higher, all algorithms perform similarly as shown in Figure 13 and Figure 14.

Since the dataset of the second case study has SW irradiance as well, the SW irradiance method is also applied to case study II. For the diffuse fraction the errors were 16-34% higher in case study II, whereas for the clearness index the errors were up to 19% lower in case study II. No explanation could be found for these contradicting indices. Therefore, the differences are most likely due to the small dataset of case study II. Where case study II uses only 17 days of data, case study I uses 2 years of data. Besides, all data of case study II was measured in January and thus only data from the winter months (with low elevation angles) could be included. This data with low elevation angles caused higher errors as was seen in the decomposition analysis. It is expected that using a larger dataset will improve the quality of the results and better align the clearness index and diffuse fraction. However, comparing the results of both case studies when using the diffuse fraction yields the same ranking of the algorithms, besides using the clearness index yields a comparable result as well. Therefore, considering the performance of the algorithms and the decomposition-, time-of-day-, solar- and sensitivity analysis, overall TRINITY is found to perform best followed by CDOC algorithm and the BRBG algorithm.

The LW irradiance method is based on 17 days of measurements with a total of 955 images used in the analysis. Although all cloud conditions (clear sky, thin clouds/haze, partly cloudy, overcast) occurred during this period, clear sky conditions were strongest represented. Therefore, the dataset used is not ideal and the results are preliminary. The visual inspection shows that the TRINITY algorithm performs best, especially at thin cloud/hazy conditions. The CDOC algorithm performs worst, which is partly due

to the incompleteness of the CSL. The MAEs and RMSEs of all algorithms are significantly higher than results obtained by the SW irradiance method, especially for the CDOC and TRINITY algorithms which errors more than double (Table 7 and Table 13). However, according to the LW method the BRBG performs best, followed by the TRINITY algorithm. Since the RMSE is much higher than the MAE for the TRINITY algorithm compared to the other algorithms, one can conclude that the TRINITY output has more extreme outliers than the other algorithms. So, because the accuracy and RMSE of TRINITY is highest, the outliers of the evaluated timestamps must have a larger error. In other words, when the TRINITY algorithm does not evaluate an image correctly, the error is larger than when the other two algorithms wrongly evaluate the image. The accuracies of all algorithms are shown in Table 12. With a difference of zero octas the accuracy of the BRBG algorithm is 52.67%, whereas Luiz et al., (2018) found an accuracy of 64% when using the same methodology. With a difference in accuracy lower than 12% and a much smaller dataset, the results obtained are of comparable reliability.

Based on the visual inspection and the algorithms performances it can be stated that using LW irradiance as a proxy for cloud cover is less accurate than using SW irradiance. The visual inspection of the algorithms at the SW irradiance method shows that the CCF_{SW} better represents the outputs of the algorithms compared to the CCF_{LW} calculated with the LW irradiance method. Especially for the thin cloud/partly cloudy situations the CCF_{SW} appears to be more accurate than the CCF_{LW}. Furthermore, comparing the results of Table 7 and Table 13 shows that the MAEs when using the diffuse fraction are 27%, 58% and 53% lower than when using the LW irradiance method for BRBG, CDOC and TRINITY, respectively.

Limitations of the research

Since the irradiance values used in the research are very dependent on the weather variability, several limitations are inevitable. GHI measurements made on partly cloudy days do not always represent the actual cloud conditions. When, for example, the cloud cover is 80% and the GHI is measured at the other 20%, which forms a gap in the clouds an error occurs. The cloud detecting algorithms will classify the cloud cover as 80%, whereas the clearness index would classify the cloud cover as 0%. The same occurs when there is an almost clear sky, but a few clouds are blocking the path between the sun and the pyranometer. This under- and over estimation of the cloudiness affects the results of the analysis. However, the effect on the differences between the algorithms is expected to remain small, since the same clearness index is used for all algorithms. Besides, part of this unreliable irradiance data is filtered out by using the thresholds described in step 4 of the SW irradiance method.

For determining the clearness index, the clear sky irradiance values are calculated by using a CSM. As can be seen from Figure 2, especially the measured DNI and DHI deviate a bit from the clear sky DNI and DHI, respectively. The CSM GHI has a better fit with the measured GHI. The fit between measured and calculated irradiance values thus only appears to affect the calculation of the diffuse fraction. However, this fraction is subject to another error. At clear sky days the DHI does not reach a value of zero. Since diffuse irradiance is caused by water vapor, ozone and other particles in the sky there will always be DHI at clear sky days (as shown in Figure 2). Consequently, the diffuse fraction at clear sky days will always be higher than 0. This can be seen from Figure 11 in which there are no data points with an x-value smaller than 0.1. nonetheless, the diffuse fraction has a good fit with the algorithms evaluations and the effect of this error on the results are foreseen to be small, since the proxy is used for all algorithms. Besides, cloudy conditions occur much more in the Netherlands than clear sky conditions, which reduces the effect on the results.

The shortwave irradiance was measured with a pyranometer, while longwave irradiance was measured with a pyrgeometer. Both instruments have a field of view of 180°. The sky images are made with an EKO SRF-02 and ASI-16 sky imager for case study I & II, respectively. The view of these sky imagers is, however, restricted to a zenith angle of 70°. This leaves an effective field of view of 140° for the sky imagers. Thus, the irradiance is measured over a larger area than the cloudiness of the

algorithms covers. Setting the zenith angle of the cloud detection algorithms to 90° would decrease the quality of the analysis, since detecting clouds near the horizon is very hard and is more prone to errors than detecting clouds at low zenith angles is. Reducing the field of view of the irradiance measurements equipment would be best. However, for both case studies historical datasets were used which contained irradiance measurements in standard formats.

Theoretical implications

The research conducted introduced and validated a new cloud detecting algorithm. By comparing the newly developed algorithm with existing algorithms, the strengths and weaknesses of the new algorithm as well as the overall performance was added to the theoretical knowledge base of sky imaging, or more general to the literature of cloud detection and irradiance prediction methodologies. By forecasting the short-term irradiance more accurately the stability on the power market can be increased and e.g. unit-commitment decisions can be optimized.

Furthermore, the research introduced, next to the LW irradiance method, a new method of validating a cloud detection algorithm. By using the diffuse fraction, instead of the clearness index, more accurate results were obtained. Were Calbó, González, & Pagès (2001), Cros et al. (2013), Pagès, Calbó, & González (2003) use the clearness index to distinguish between clear and cloudy skies, this research adapted the diffuse fraction approach (Butt et al., 2010) to classify the cloudiness and applied it to validate a cloud detection algorithm in combination with an all sky imager. In this way, the new algorithm is validated quantifiably, whereas other authors (Ghonima et al., 2012; Kazantzidis, Tzoumanikas, Bais, Fotopoulos, & Economou, 2012; J Yang et al., 2015; Jun Yang et al., 2016) used human observations as a reference point to validate cloud detection algorithms.

Recommendations

To better evaluate the accuracies of cloud detecting algorithms it is recommended to synchronize data measurements better for future work. By synchronizing the time at which sky images are made with the time at which irradiance (SW & LW) is measured, the interpolation error can be excluded entirely. Furthermore, data collection at a shorter time interval (e.g. every 60 seconds) increases accuracy of the analysis by taking weather variability into account more effectively. The only condition is, however, that enough storage capacity is present.

To improve the quality of new work validating the TRINITY algorithm or other new cloud detection algorithms, the dataset used can be enlarged. This only holds for the second case study. Due to a software bug this research could only use data collected for 17 days, whereas more data was available. The results obtained by the LW irradiance method are thus seen as preliminary and by expanding the dataset in future work statements and conclusions can be enhanced.

Further, it is recommended to include more geographical locations in future studies. The lack of sufficient data in the second case study limits an accurate comparison between the different locations. By using multiple sites with different latitudes and longitudes different climates can be taken into account and the performance of all algorithms can be assessed more comprehensively. When comparing different geographical locations, it is recommended to use the same sky imagers and measurements equipment at all sites. Consequently, it is important to reduce the field of view of pyranometers and pyrgeometers to 140°. In this way all data collected represents the same area in the sky.

Finally, it is recommended to create new insights in the proxies used to estimate cloud cover. Where some research uses human observations to validate cloud detecting methods (Ghonima et al., 2012; Kazantzidis et al., 2012; J Yang et al., 2015; Jun Yang et al., 2016) others (Calbó et al., 2001; Cros et al., 2013; Luiz et al., 2018; Marty & Philipona, 2000; Pagès et al., 2003) use a more quantified approach. A detailed and comprehensive review of cloud cover proxies such as the clearness index, diffuse fraction

or downward longwave irradiance is missing. Not only would this shed light on the accuracies and differences between the proxies, it could also provide a uniform approach for evaluating cloud cover or solar irradiance forecasting methods.

6. Conclusion

The work in this study focusses on answering the following research question: *How does the TRINITY algorithm in combination with an All Sky Imager perform compared to existing cloud detecting algorithms?* To provide a broader understanding of existing cloud detection and to clarify the position of sky imaging within the field of solar and cloud forecasting methods a literature review maps and identifies similarities and differences between these methods. Then three cloud detecting algorithms (BRBG, CDOC and TRINITY) are evaluated by two different methodologies. Two years of shortwave irradiance data from Utrecht University in combination with a clear sky model is used to calculate the clearness index and diffuse fraction. The FINDCLOUDS software evaluated the sky images and each algorithm calculated the Cloud Cover Fractions for each image. The correlation between these fractions and the indices is then determined by calculating the Mean Absolute Error and Root Mean Square Error. This analysis was based on 2 years of collected data. Data from Denver (US) is used to calculate the Cloud Cover Fraction based on downward longwave radiation considering 17 days of data. The methodology is adapted from Luiz et al. (2018).

Results show that the diffuse fraction has the closest fit with the data and leads to the most accurate results. For the shortwave irradiance method, the TRINITY algorithm performs best with a MAE of 0.12, whereas the BRBG and CDOC have a value of 0.17 and 0.14, respectively. The decomposition analysis shows that the CDOC algorithm has difficulties evaluating clear sky images. At partly cloudy conditions all algorithms perform similarly, whereas at overcast conditions the BRBG algorithm has highest errors. The TRINITY algorithm performs good at all sky conditions and is more stable than other algorithms. The time-of-day analysis shows that excluding the sunrise and sunset periods improves the accuracies of all algorithms. The results of the solar position analysis show that the BRBG algorithm turns out to be most sensitive to low elevation angles, leading to higher errors. The CDOC algorithm is more accurate than the BRBG algorithm, but is still fluctuating a lot with a changing solar position. The TRINITY algorithm is more accurate than the BRBG algorithm as well, besides it is the most stable for the range of elevation angles. The preliminary results for using longwave downward radiation show that the accuracies of all algorithms are comparable (53%, 54% and 56% for BRBG, CDOC and TRINITY, respectively) with lowest errors for the BRBG algorithm. MAEs are, however, significantly higher for the LW method compared with the SW method. Therefore, it can be concluded that the SW irradiance method is most accurate. The newly developed TRINITY algorithm has the best overall performance. Furthermore, it turns out to be good at all sky conditions and more stable at low elevation angles. Therefore, the implementation of the TRINITY algorithm in combination with an ASI will improve the accuracy of short-term solar forecasting.

7. References

Alduchov, O. A., & Eskridge, R. E. (1996). Improved Magnus Form Approximation of Saturation Vapor Pressure. *Journal of Applied Meteorology*, *35*(4), 601–609. https://doi.org/10.1175/1520-0450(1996)035<0601:IMFAOS>2.0.CO;2

Andreas, A.; Stoffel, T.; (1981). NREL Solar Radiation Research Laboratory (SRRL): Baseline Measurement System (BMS); Golden, Colorado (Data); NREL Report No. DA-5500-56488. http://dx.doi.org/10.5439/1052221

Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., Martinez-de-Pison, F. J., & Antonanzas-Torres, F. (2016). Review of photovoltaic power forecasting. *Solar Energy*, *136*, 78–111. https://doi.org/10.1016/J.SOLENER.2016.06.069

Bacher, P. (2008). *Short-term Solar Power Forecasting*. Technical University of Denmark, Kongens Lyngby. Retrieved from www.imm.dtu.dk

Bilionis, I., Constantinescu, E. M., & Anitescu, M. (2014). Data-driven model for solar irradiation based on satellite observations. https://doi.org/10.1016/j.solener.2014.09.009

Blanc, P., Remund, J., & Vallance, L. (2017). Short-term solar power forecasting based on satellite images. *Renewable Energy Forecasting*, 179–198. https://doi.org/10.1016/B978-0-08-100504-0.00006-8

Breyer, C., Bogdanov, D., Aghahosseini, A., Gulagi, A., Child, | Michael, Ayobami, |, … Pasi Vainikka, |. (2017). Solar photovoltaics demand for the global energy transition in the power sector. https://doi.org/10.1002/pip.2950

Brown, P. A. (1998). *A Review of the Literature on Case Study Research*. *Canadian Journal for New Scholars in Education* (Vol. 1). Retrieved from https://journalhosting.ucalgary.ca/index.php/cjnse/article/viewFile/30395/24837

Butt, N., New, M., Malhi, Y., da Costa, A. C. L., Oliveira, P., & Silva-Espejo, J. E. (2010). Diffuse radiation and cloud fraction relationships in two contrasting Amazonian rainforest sites. *Agricultural and Forest Meteorology*, *150*(3), 361–368. https://doi.org/10.1016/J.AGRFORMET.2009.12.004

Calbó, J., González, J.-A., & Pagès, D. (2001). A Method for Sky-Condition Classification from Ground-Based Solar Radiation Measurements. *Journal of Applied Meteorology*, *40*(12), 2193–2199. https://doi.org/10.1175/1520-0450(2001)040<2193:AMFSCC>2.0.CO;2

Campbell, S. D., & Diebold, F. X. (2005). Weather Forecasting for Weather Derivatives. *Journal of the American Statistical Association*, *100*, 6–16. https://doi.org/10.1198/016214504000001051

Chaturvedi, D. K., & Isha. (2016). *Solar Power Forecasting: A Review*. *International Journal of Computer Applications* (Vol. 145). Retrieved from https://pdfs.semanticscholar.org/a679/0993e18db1b5d488b161194fa07a3d5c139f.pdf

Chauvin, R., Nou, J., Thil, S., Traore, A., & Grieu, S. (2015). Cloud detection methodology based on a sky-imaging system ScienceDirect, *69*, 1970–1980. https://doi.org/10.1016/j.egypro.2015.03.198

Cheng, C.-H., & Nnadi, F. (2014). Predicting Downward Longwave Radiation for Various Land Use in All-Sky Condition: Northeast Florida. *Advances in Meteorology*, *2014*, 1–12. https://doi.org/10.1155/2014/525148

Chow, C. W., Urquhart, B., Lave, M., Dominquez, A., Kleissl, J., Shields, J., & Washom, B. (2011). Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. *Elsevier*, 13.

CMS. (2016). *Find Clouds Software (Instruction Manual)*. Kirchbichl. Retrieved from http://romiotto.com.br/uploads/arquivos/literatura-tecnica/SCHREDER/Manual_Findclouds.pdf

CMS. (2018). *Findclouds Trinity (Instruction Manual)*. Kirchbichl.

Cros, S., Liandrat, O., Sébastien, N., Schmutz, N., & Voyant, C. (2013). CLEAR SKY MODELS ASSESSMENT FOR AN OPERATIONAL PV PRODUCTION FORECASTING SOLUTION, 5BV.4.69. Retrieved from https://hal.archives-ouvertes.fr/hal-00870092/

DeMeo, E. A., Grant, W., Milligan, M. R., & Schuerger, M. J. (2005). *Wind Plant Integration*. Retrieved

from http://www.science.smith.edu/~jcardell/Readings/Wind/DeMeo - P&E.pdf
Diagne, H., Lauret, P., & David, M. (2012). *Solar irradiation forecasting: state-of-the-art and proposition for future developments for small-scale insular grids*. Denver. Retrieved from https://hal.archives-ouvertes.fr/hal-00918150
Diagne, M., David, M., Lauret, P., Boland, J., & Schmutz, N. (2013). Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. *Renewable and Sustainable Energy Reviews*, *27*, 65–76. https://doi.org/10.1016/J.RSER.2013.06.042
Engerer, N. A., & Mills, F. P. (2014). KPV: A clear-sky index for photovoltaics. *Solar Energy*, *105*, 679–693. https://doi.org/10.1016/J.SOLENER.2014.04.019
Feng, C., Cui, M., Lee, M., Zhang, J., Hodge, B.-M., Lu, S., & Hamann, H. F. (2017). Short-term global horizontal irradiance forecasting based on sky imaging and pattern recognition. In *2017 IEEE Power & Energy Society General Meeting* (pp. 1–5). IEEE. https://doi.org/10.1109/PESGM.2017.8274480
Ghonima, M. S., Urquhart, B., Chow, C. W., Shields, J. E., Cazorla, A., & Kleissl, J. (2012). A method for cloud detection and opacity classification based on ground based sky imagery. *Atmospheric Measurement Techniques*, *5*(11), 2881–2892. https://doi.org/10.5194/amt-5-2881-2012
Gueymard, C. A. (2012). Clear-sky irradiance predictions for solar resource mapping and large-scale applications: Improved validation methodology and detailed performance analysis of 18 broadband radiative models. *Solar Energy*, *86*(8), 2145–2169. https://doi.org/10.1016/J.SOLENER.2011.11.011
Hancock, D. R., & Algozzine, B. (2006). *Doing Case Study Research*. New York: Teachers College Press.
Heinle, A., Macke, A., & Srivastav, A. (2010). Automatic cloud classification of whole sky images. *Atmospheric Measurement Techniques*, *3*(3), 557–567. https://doi.org/10.5194/amt-3-557-2010
Huo, J., & Lu, D. (2009). Cloud Determination of All-Sky Images under Low-Visibility Conditions. *Journal of Atmospheric and Oceanic Technology*, *26*(10), 2172–2181. https://doi.org/10.1175/2009JTECHA1324.1
Hyvärinen, O., Julkunen, J., & Nietosvaara, V. (2007). Climatological Tools for Low Visibility Forecasting. *Pure and Applied Geophysics*, 14. https://doi.org/10.1007/s00024-007-0224-5
Inman, R. H., Pedro, H. T. C., & Coimbra, C. F. M. (2013). Solar forecasting methods for renewable energy integration. *Progress in Energy and Combustion Science*, *39*, 535–576. https://doi.org/10.1016/j.pecs.2013.06.002
International Energy Agency. (2017). *World Energy Outlook iea.org/weo/ Chapter 1: Introduction and scope*. Retrieved from www.iea.org/t&c/
Isenson, R. S. (1966). TECHNOLOGICAL FORECASTING IN PERSPECTIVE. *Management Science*, *13*(2), 14. Retrieved from https://www.jstor.org/stable/pdf/2628073.pdf?casa_token=Lh9t1H3AKAwAAAAA:wku7WgwvCfI0e1XtdQXxWggfHTMBjlIJ-59lADVQpLbqc0gzM8MlAQODtUXM3VuJj1fXAXHJBK4a3uLxhznbMjU1wO_DKJenBCnuTFGz1BOR5na08eAu
Jacobson, M. Z., & Delucchi, M. A. (2011). Providing all global energy with wind, water, and solar power, Part I: Technologies, energy resources, quantities and areas of infrastructure, and materials. https://doi.org/10.1016/j.enpol.2010.11.040
Jones, P. A. (1992). Cloud-Cover Distributions and Correlations. *Journal of Applied Meteorology*, *31*(7), 732–741. https://doi.org/10.1175/1520-0450(1992)031<0732:CCDAC>2.0.CO;2
Kato, S., Ackerman, T. P., Clothiaux, E. E., Mather, J. H., Mace, G. G., Wesely, M. L., … Michalsky, J. (1997). Uncertainties in modeled and measured clear-sky surface shortwave irradiances. *Journal of Geophysical Research: Atmospheres*, *102*(D22), 25881–25898. https://doi.org/10.1029/97JD01841
Kazantzidis, A., Tzoumanikas, P., Bais, A. F., Fotopoulos, S., & Economou, G. (2012). Cloud detection and classification with the use of whole-sky ground-based images. *Atmospheric Research*, *113*, 80–88. https://doi.org/10.1016/J.ATMOSRES.2012.05.005
Kleissl, J. (2013). *Solar energy forecasting and resource assessment*. Retrieved from

https://books.google.nl/books?hl=en&lr=&id=94KI0_SPwW8C&oi=fnd&pg=PP1&dq=Solar+Energy+Forecasting+and+Resource+Assessment&ots=HcVnRDU7Sy&sig=1meLK-V8KvmU22xk32XdU-w6d6M

KNMI. (2019). KNMI - Daggegevens van het weer in Nederland - De Bilt. Retrieved March 22, 2019, from https://www.knmi.nl/nederland-nu/klimatologie/daggegevens

Kuhn, P., Nouri, B., Wilbert, S., Prahl, C., Kozonek, N., Schmidt, T., … Pitz-Paal, R. (2017). Validation of an all-sky imager-based nowcasting system for industrial PV plants. *Wiley Photovoltaics*, 14. https://doi.org/10.1002/pip.2968

Lorenz, E., Hammer, A., & Heinemann, D. (2004). *Short term forecasting of solar radiation based on satellite data*. Oldenburg. Retrieved from https://www.researchgate.net/profile/Annette_Hammer/publication/267971482_Short_term_forecasting_of_solar_radiation_based_on_satellite_data/links/5583ecb708ae4738295bb661.pdf

Lorenz, E., Remund, J., Müller, S. C., Traunmüller, W., Steinmaurer, G., Pozo, D., … Guerrero, C. G. (2009). *BENCHMARKING OF DIFFERENT APPROACHES TO FORECAST SOLAR IRRADIANCE*. Retrieved from http://www.mesor.net

Luiz, E. W., Martins, F. R., Costa, R. S., & Pereira, E. B. (2018). Comparison of methodologies for cloud cover estimation in Brazil - A case study. *Energy for Sustainable Development*, *43*, 15–22. https://doi.org/10.1016/J.ESD.2017.12.001

Marty, C., & Philipona, R. (2000). The Clear-Sky Index to separate clear-sky from cloudy-sky situations in climate research Snowpack and alpine plants View project pluSnow-Using snow depth data to minimize the error in winter precipitation measurements View project The Clear-Sky Index to separate Clear-Sky from Cloudy-Sky Situations in Climate Research. *Article in Geophysical Research Letters*, *27*(17). https://doi.org/10.1029/2000GL011743

Mathiesen, P., & Kleissl, J. (2011). Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States. *Solar Energy*, *85*(5), 967–977. https://doi.org/10.1016/J.SOLENER.2011.02.013

Okogbue, E. C., Adedokun, J. A., & Holmgren, B. (2009). Review Hourly and daily clearness index and diffuse fraction at a tropical station, Ile-Ife, Nigeria. *INTERNATIONAL JOURNAL OF CLIMATOLOGY Int. J. Climatol*, *29*, 1035–1047. https://doi.org/10.1002/joc.1849

Pagès, D., Calbó, J., & González, J. A. (2003). Using routine meteorological data to derive sky conditions. *Annales Geophysicae*, *21*(3), 649–654. Retrieved from https://hal.archives-ouvertes.fr/hal-00317006/

Paulescu, M., Paulescu, E., Gravila, P., & Badescu, V. (2012). *Weather Modeling and Forecasting of PV Systems Operation.* Springer.

Reindl, D., & Beckman, W. A. (1990). Diffuse fraction correlations. *Solar Energy*, *45*(1), 1–7. https://doi.org/10.1016/0038-092X(90)90060-P

Remund, J., Wald, L., Lefèvre, M., Ranchin, T., & Page, J. (2003). Worldwide Linke turbidity information, *CD-ROM*, 13 p. Retrieved from https://hal.archives-ouvertes.fr/hal-00465791/

Reno, M. J., Hansen, C. W., & Stein, J. S. (2012). *Global Horizontal Irradiance Clear Sky Models: Implementation and Analysis*. Albuquerque. Retrieved from http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

Richardson, W., Krishnaswami, H., Vega, R., Cervantes, M., Richardson, W., Krishnaswami, H., … Cervantes, M. (2017). A Low Cost, Edge Computing, All-Sky Imager for Cloud Tracking and Intra-Hour Irradiance Forecasting. *Sustainability*, *9*(4), 482. https://doi.org/10.3390/su9040482

Riordin, D., & Hansen, B. K. (2002). A fuzzy case-based system for weather prediction. *Engineering Intelligent Systems*, *10*(3), 8. Retrieved from https://pdfs.semanticscholar.org/0719/a63450701017c89bfe3d60800485ae72c76b.pdf

Rowley, J. (2002). *Using Case Studies in Research*. Retrieved from https://pdfs.semanticscholar.org/aab8/c9514b473c4ec9c47d780b7c79112add9008.pdf

Schäfer, B. ;, Beck, C. ;, Aihara, K. ;, Witthaut, D. ;, & Timme. (2017). *Non-Gaussian power grid frequency fluctuations characterized by Lévy-stable laws and superstatistics*. Retrieved from

http://qmro.qmul.ac.uk/xmlui/handle/123456789/29104
Thomas, G. (2015). *How to do your case study*.
Tuohy, A., Zack, J., Haupt, S. E., Sharp, J., Ahlstrom, M., Dise, S., … Collier, C. (2015). Solar Forecasting: Methods, Challenges, and Performance. *IEEE Power and Energy Magazine*. https://doi.org/10.1109/MPE.2015.2461351
Verbois, H., Huva, R., Rusydi, A., & Walsh, W. (2018). Solar irradiance forecasting in the tropics using numerical weather prediction and statistical learning. *Solar Energy*, *162*, 265–277. https://doi.org/10.1016/J.SOLENER.2018.01.007
Voyant, C., Notton, G., Kalogirou, S., Nivet, M.-L., Paoli, C., Motte, F., & Fouilloy, A. (2017). Machine learning methods for solar radiation forecasting: A review. *Renewable Energy*, *105*, 569–582. https://doi.org/10.1016/j.renene.2016.12.095
Wan, C., Zhao, J., Member, S., Song, Y., Xu, Z., Member, S., … Hu, Z. (2015). Photovoltaic and Solar Power Forecasting for Smart Grid Energy Management. *CSEE JOURNAL OF POWER AND ENERGY SYSTEMS*, *1*(4). https://doi.org/10.17775/CSEEJPES.2015.00046
Yang, J., Min, Q., Lu, W., Ma, Y., Yao, W., Lu, T., … Liu, G. (2016). A total sky cloud detection method using real clear sky background. *Atmos. Meas. Tech*, *9*, 587–597. https://doi.org/10.5194/amt-9-587-2016
Yang, J., Min, Q., Lu, W., Yao, W., Ma, Y., Du, J., … Liu, G. (2015). An automated cloud detection method based on the green channel of total-sky visible images. *Atmos. Meas. Tech*, *8*, 4671–4679. https://doi.org/10.5194/amt-8-4671-2015
Yankeelov, T. E., Quaranta, V., Evans, K. J., & Rericha, E. C. (2015). Toward a Science of Tumor Forecasting for Clinical Oncology. https://doi.org/10.1158/0008-5472.CAN-14-2233
Young, H. D., & Freedman, R. A. (2016). *University Physics with Modern Physics, Global Edition* (14th ed.). Harlow: Pearson. Retrieved from https://books.google.nl/books?id=2vhECgAAQBAJ&dq=university+physics+with+modern+physics+14th+edition&hl=en&sa=X&ved=0ahUKEwjkh9bNiLvgAhWEbFAKHbRFBeQQ6AEIRjAF