Title: A Method for Detailed, Short-Term Energy Yield Forecasting of Photovoltaic Installations

Corresponding Author: Mr. Dimitrios Anagnostos,

Corresponding Author's Institution: NTUA, KUL

First Author: Dimitrios Anagnostos

Order of Authors: Dimitrios Anagnostos; Thomas Schmidt, PhD; Spiridon Cavadias; Dimitrios Soudris; Jef Poortmans; Francky Catthoor

Abstract: The global shift towards renewable energy production suggests a promising future on reducing carbon emissions and avoiding the detrimental effects of global warming, but also creates additional challenges on all levels of energy production and distribution. The expected penetration of electric cars, increasing energy usage of cloud computing centers and the transformation of the electricity grid itself towards the "Smart Grid" requires novel solutions on all levels of energy production and management. Forecasting of energy production especially will become a major component for design and operation in all temporal and spatial scales, creating opportunities for optimized control of energy storage, local energy exchange etc. To this end, a method for the creation of detailed and accurate energy yield forecasts for PV installations is presented. Based on sky-imager information and using tailored neural networks, highly detailed energy yield forecasts are produced for a monitored test installation, for horizons up to 15 minutes and with a resolution of 1 second. The described method manages to outperform state of the art models by up to 39% in forecast skill, while at the same time retaining temporal resolutions that enable control schemes and energy exchange in a local scale.

Prof. Soteris Kalogirou

Editor-in-Chief

Renewable Energy

February 18th 2018

Dear Prof. Kalogirou,

Please find enclosed our manuscript entitled "A Method for Detailed, Short-Term Energy Yield Forecasting of Photovoltaic Installations", which we would like to submit for publication as a research paper in Renewable Energy.

This study deals with the development of a method to accurately predict real energy yield of PV installations for horizons up to 15 minutes and with resolution down to 1 second. It is based on prior work from the authors, both on sky-imager based irradiation forecasting and detailed dynamic modeling of PV systems, but a different approach is described, where a model is optimized specifically for energy yield prediction. State-of-the-art methodologies usually report a single value of average produced power for the desired horizon, whereas the described method provides a detailed trace, suitable for applications such as storage management and datacenter workload scheduling among others.

A rooftop PV installation is monitored and forecasts are also estimated with a state-of-the-art method, in order to compare the accuracy with the proposed method. Forecast skill is improved up to 39% in relation to typical GHI forecasting techniques and insights are provided for the sources of the error for each case.

The forecasting community will benefit from the results of this study, as it provides a solid reference for the potential gains of including cloud class information and the dynamic effects of PV systems in forecasting models. The methods are described in a way that any researcher can adjust them for future research.

We confirm that this manuscript has not been published elsewhere and is not under consideration by another journal. Parts of this work were included in the master thesis of S.Cavadias, which can be found online on the NTUA repository[1]. All authors have approved the manuscript and agree with submission to Renewable Energy. The authors have no conflicts of interest to declare.

---

[1] http://dspace.lib.ntua.gr/handle/123456789/46487

Please address all correspondence to:

Dimitrios Anagnostos

National Technical University of Athens (NTUA), School of Electrical & Computer Engineering

9 Heroon Polytechneiou, Zographou Campus, 157 80 Athens - Greece

tel: +30 210 772 3653

e-mail: anagnostos.d@microlab.ntua.gr

We look forward to hearing from you at your earliest convenience.

Yours sincerely,

Dimitrios Anagnostos

*Highlights

- Detailed energy yield forecasts are enablers for advanced control schemes towards the smart grid transformation
- Developed models showcase superior accuracy in high resolution energy yield forecasting
- Combining sky-imaging information and classification with NN dynamic modelling of PV systems is explored thoroughly

A Method for Detailed, Short-Term Energy Yield Forecasting of Photovoltaic Installations

D. Anagnostos[1,3], T. Schmidt[4], S. Cavadias[1], D. Soudris[1], J. Poortmans[2,3], F. Catthoor[2,3]

[1]National Technical University of Athens, Heroon Polytechneiou 9, Zographou Campus 157 80 Athens – Greece
[2]imec, Kapeldreef 75, 3001 Heverlee , Belgium
[3]Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, 3001 Heverlee, Belgium
[4]DLR Institute of Networked Energy Systems, Energy Systems Analysis, Carl-von-Ossietzky-Str. 15, 26129 Oldenburg, Germany

1   *Abstract*—**The global shift towards renewable energy production suggests a promising future on reducing carbon**
2   **emissions and avoiding the detrimental effects of global warming, but also creates additional challenges on all levels of**
3   **energy production and distribution. The expected penetration of electric cars, increasing energy usage of cloud**
4   **computing centers and the transformation of the electricity grid itself towards the "Smart Grid" requires novel solutions**
5   **on all levels of energy production and management. Forecasting of energy production especially will become a major**
6   **component for design and operation in all temporal and spatial scales, creating opportunities for optimized control of**
7   **energy storage, local energy exchange etc. To this end, a method for the creation of detailed and accurate energy yield**
8   **forecasts for PV installations is presented. Based on sky-imager information and using tailored neural networks, highly**
9   **detailed energy yield forecasts are produced for a monitored test installation, for horizons up to 15 minutes and with a**
10   **resolution of 1 second. The described method manages to outperform state of the art models by up to 39% in forecast**
11   **skill, while at the same time retaining temporal resolutions that enable control schemes and energy exchange in a local**
12   **scale.**

13
14   *Keywords*— **Energy yield forecasting, Sky-imager, Neural Networks.**

15   1. INTRODUCTION

16   The whole world is witnessing a rapid shift towards renewable energy production. Several countries that were skeptical in
17 adopting these new technologies, relying on cheap, locally produced coal are now becoming huge supporters of solar and wind
18 power, each for their own reasons. Following the Paris Climate Agreement, China and India are adopting their strategies and will
19 be responsible for more than half of any new installation by 2022 (Anon, 2017). These numbers seem promising for a "greener"
20 future, but also spell trouble and challenges for the electrical grid, its operators and the energy producers. High levels of
21 renewable energy sources penetration require proper handling of the grid, so that it stays balanced. Therefore the stochastic
22 nature of renewable energy production needs to be addressed in a way that does not endanger the stability of operations.

23   A change into even more reactive behavior of the electric grid has been researched intensely for the past years, paving the way
24 for a "smart" grid that enables advanced, bidirectional energy exchange schemes, including storage energy capabilities and
25 dynamic pricing. Storage facilities are expected to become necessary for this transition, ensuring stability on all levels of the
26 grid. Nonetheless, batteries impose an extra cost depending on their size and one of the ways to optimize the storage capacity is
27 through coupling with production and load forecasting (Golshannavaz et al. 2014). In fact, forecasting is becoming essential for
28 all energy exchange activities, from grid control (Golshannavaz et al., 2014), including charging policies for electric cars, to
29 scheduling the services of data centers across the world (Aksanli et al., 2011). The latter are the same data centers that support
30 the digital revolution of our age, while at the same time burning increasing amounts of energy which could account for 13% of
31 worldwide consumption by 2030 (Avgerinou et al., 2017). So, even though there is still no global agreement on how such a grid
32 will be organized, it is apparent that the stochastic nature of renewable energy necessitates predictive models that will act as
33 enablers and will aid in better utilizing all available resources.

34   Focusing more on solar energy, there are numerous methods of forecasting the power output of a photovoltaiv (PV)
35 installation, depending on the required spatial and temporal resolution. Whereas forecasts for large areas and time horizons of
36 hours to days can be produced by numerical weather prediction, machine learning methods and satellite imaging among others, it
37 is the short-term local forecasts that enable fine-grained control over grid stability, storage management and task scheduling. On
38 top of that, most methods create a global irradiation forecast which is then transposed to the plane of a specific installation and
39 translated to power production through an extra model, thus injecting additional propagated errors into the final result. Finally,
40 the forecasted values represent average power production over a certain time period and are not in any way optimized to create a
41 real energy yield curve, which is more meaningful when balancing an electrical network with batteries.

42   In this context of fine-grain control, the development of a PV energy forecast model is described, able to produce highly
43 detailed electrical energy forecast time-series with 1s resolution, for the next 15min for any given installation. The model

translates the input directly to an energy yield curve for the installation, minimizing in that way the additional uncertainties imposed by chaining different models together and providing a cumulative yield time series. The proposed model utilizes a low cost sky-imager in order to produce intermediate inputs. These are then passed to a specialized cluster of neural networks that make use of previously published methods (Anagnostos et al., 2017, Goverde et al., 2016) in order to capture the full dynamic behavior of the installation, including its thermal state. As a result, a highly accurate, highly detailed forecast is produced (<20% nRMSE for the whole horizon), which can then be used for applications such as local grid control, energy storage management, energy pricing, data center scheduling etc.

The core innovations of this work can be summarized as the following:

• Description of a novel, multi-stage topology of networks that utilize cloud type information

• Direct energy yield reporting with no error propagation through a chain of models

• Deployment and validation of the proposed model for a real, small PV installation, resulting in superior accuracy over state of the art models for higher time resolutions, enabling fine-grain control schemes including task scheduling

• Demonstration of robustness towards partly missing input data

The rest of this publication is comprised of Section 2, which provides some relevant background, Section 3, which describes the methods and models used, Section 4 where the results are presented and discussed and finishes with the conclusions in Section 5.

2. RELATED WORK

Short-term solar power forecasting has become a popular research topic in the past years, therefore many different methods and results can be traced in publications. An essential differentiator among these methods is the utilization of exogenous inputs in the forms of images, whether from satellite or sky-imagers. When such source is not available, statistical or machine learning methods are applied on a variety of historical data, mainly irradiation and temperature.

A very detailed overview of such models is given in (Pedro and Coimbra, 2012), presenting promising results for the usage of neural network models. Another methodology in this category would be the "peer to peer", or correlation lag approach, as presented in different forms (Elsinga, van Sark, 2017, Yang et al. 2014a, Lorenzo et al., 2015). In this approach, the production profile of one installation is assumed to be almost identical with that of another installation in close proximity, given that the same cloud "shadows" on the ground, pass from both, but with a time difference which can be calculated. This assumes a dense enough network of photovoltaic installations, almost perfect data quality and no "edge" effects for systems without neighbors in all directions.

The issue with relying on historical data in order to predict future events is that when increasing the spatial and temporal resolution requirements the methods lose accuracy due to less averaging on the inputs and the outputs. Furthermore, the cloud formation is also dynamic, so methods that rely on persistence of the cloud state (which translates to correlation between historical and future values) tend to underperform on the days that are most needed, the cloudy ones. A solution to the problem can be the inclusion of exogenous input, usually in the form of images coming from a sky-imager. Such methods attempt to forecast the movement of clouds on the sky by analyzing successive images of the sky dome, therefore calculating the motion vectors of possible sources of shadow. These models are especially successful in capturing future ramp events and retain high accuracies even for more dynamic days (Yang et al., 2014b, Kuhn et al., 2017, Schmidt et al., 2016). The biggest challenge for this class of models is to actually translate the forecasted pixels into accurate irradiance values and then energy yield for the installation. Relevant work has been reported in (Chen, 2011, Tzoumanikas et al., 2016), where the cloud type information is used with similar number (Tzoumanikas et al., 2016) or fewer classes (Chen, 2011).

All the previous methods are optimized for the prediction of "power" and not "energy". As exhibited in (Anagnostos et al., 2017, Goverde et al., 2015), the thermal state of the PV system cannot ignored for time steps below 5min, as it augments inaccuracies in the final estimation. Another important piece of information that is rarely used is the type of weather itself. Formally classifying cloud types can be used as an additional differentiator for training models. Moreover, most of these models require translations between irradiation and power, increasing the error margins even further.

Most importantly, the current state of the art methods report significant inconsistencies for performance between clear-sky days and days with more dynamic conditions, usually doubling their error. This effect does not allow reliable control of the grid/sub-system nor the optimization of the installed storage, which has to be larger in order to accommodate such uncertainties. Most of the research performed on storage optimization (Megel et al., 2015), data-center scheduling based on green energy

(Aksanli et al., 2011, Goiri et al., 2015, Li et al., 2017) and smart grid fine-control (Golshannavaz et al., 2014), is based on the assumption of perfect (or close to) detailed energy yield forecasts in order to enable the rest of the findings. Any increase in error margins reduces cost efficiency and therefore makes the solutions unattractive. The latter problem is exactly what this work is effectively addressing.

3. METHODS DESCRIPTION

A. *Image acquisition and measurement installation description*

The data set used for the studies has been acquired at the University of Oldenburg (53.15232 °N, 8.166022 °E) and spans from the 19th of July until 31st of August 2015. The logged information is diffuse horizontal irradiance (DHI, measured with a ventilated and shaded K&Z CM11), direct normal irradiance (DNI, measured with Eppley NIP) and photovoltaic output parameters (Current and Voltage in MPP, Backsheet Temperature). The investigated PV module is a 180W peak BP Solar panel installed with a tilt angle of 51 degrees and orientated south. All parameters are sampled with 1 Hz frequency and data quality has been ensured through visual inspection and statistical analysis.

Moreover, sky images are retrieved every 10s from sunrise to sunset. The sky imager used in Oldenburg is a commercial Vivotek FE8172V network camera equipped with a fisheye lens (Fig. 1). The typical application of this camera type is surveillance of wide, open areas. Compared to cameras developed specifically for sky and cloud observations, commercial network cameras are less expensive (<1k€) and therefore of interest for solar energy applications. The most important technical specifications of the camera are a full 180° field of view, a circular fisheye frame in a 1920 x 1920 pixels image plane and a dynamic range of more than 57 dB. The camera configurations including color settings, white balance and exposure settings are applied browser-based. A Python-based interface has been developed to control most of the settings fully automatic.

The sun tracker equipped with a pyrheliometer for DNI and a shaded pyranometer for DHI measurements is placed three meters north-west of the camera. All sensors are maintained and cleaned typically once per week. The investigated PV module is located about 19 meters from the camera and suntracker. In this study all sensors are assumed to be at one location.
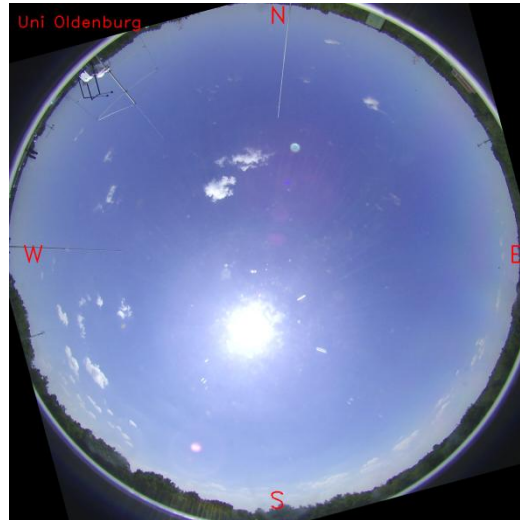

Figure 1. Example image of the used sky imager.

B. *Image feature extraction*

Specific image features are computed for each image, then provided as inputs for the machine learning applications. The features are quantified characteristics like image textures, color values and other metrics. These features are selected to quantify special image characteristics discriminating for example a full cloud cover from mixed cloud conditions. In terms of irradiance modeling, features describing the pixel intensity in the circumsolar area are used as the implicit information about direct solar radiation. For example, in overcast conditions with thick clouds, the circumsolar area pixel intensity value will be much lower than in clear sky conditions. The features are computed on the masked image, so that non-sky parts of the image are masked out (stationary or moving obstacles) before image features are computed. In this study, the extracted image features are used for irradiance modelling (k-neighbors neural network model), cloud classification (support vector classification model) and the energy yield prediction with neural networks.

The features used for the focus of this study, the neural network based energy yield prediction, are presented briefly. First, the image Contrast retrieved from the grey-level co-occurrence matrix (GLCM) (Haralick et al., 1973) is used as input to the neural network. Contrast quantifies the difference in color and brightness in the image. Additionally, image features are derived from

RGB channel-based color statistics. The average blue and red color of the image, the average red to blue color ratio, the standard deviation of the blue color and the average difference between blue and red color are determined. Especially the combination of red and blue is a good indicator of overall cloud coverage (Shields et al., 1988). Finally, a forecast of the (possible) passing clouds in front of the circumsolar area is provided in terms of RGB values. The forecast horizon for this input is 900s for this application and corresponds to the energy yield forecast horizon.

In addition to the image features, the solar position relative to the observer (zenithal and azimuthal angle) and the cloud coverage as the ratio of cloudy pixels to all pixels are used as input for the neural network.

C. Cloud classification and cloud class probabilities

The sky imaging software determines for each image the predominant sky or cloud type as one of seven categories. The algorithm is a modification of (Heinle et al., 2010). The seven categories are meteorologically justified according to the mentioned paper and an example for each is given in Fig. 2:

- Class 1: Cumulus (Cu) - low, puffy clouds with clearly defined edges, white or light-grey
- Class 2: Cirrostratus (Cs), Cirrus (Ci) - high, thin clouds, wisplike or sky covering, whitish
- Class 3: Cirrocumulus (Cc), Altocumulus (Ac) - high patched clouds of small cloudlets, mosaic-like, white
- Class 4: Clear sky (Clear) - no clouds and cloudiness below 10%
- Class 5: Stratocumulus (Sc) - low or mid-level, lumpy layer of clouds, broken to almost overcast, white or grey
- Class 6: Stratus (St), Altostratus (As) - low or mid-level layer of clouds, uniform, usually overcast, grey
- Class 7: Nimbostratus (Ns), Cumulonimbus (Cb) - dark, thick clouds, mostly overcast, grey

The classification algorithm makes use of machine learning based classification schemes from the scikit-learn package (Pedregosa et al., 2011). After testing several classification schemes the Support Vector Classification (SVC) has been chosen with best classification results. Multi-label datasets (seven cloud types) are used for training the models. From Oldenburg image archive about 21500 images were manually pre-selected into the different cloud type categories. For each image, global image features (previous section) characterizing image texture properties and color values are computed and normalized.
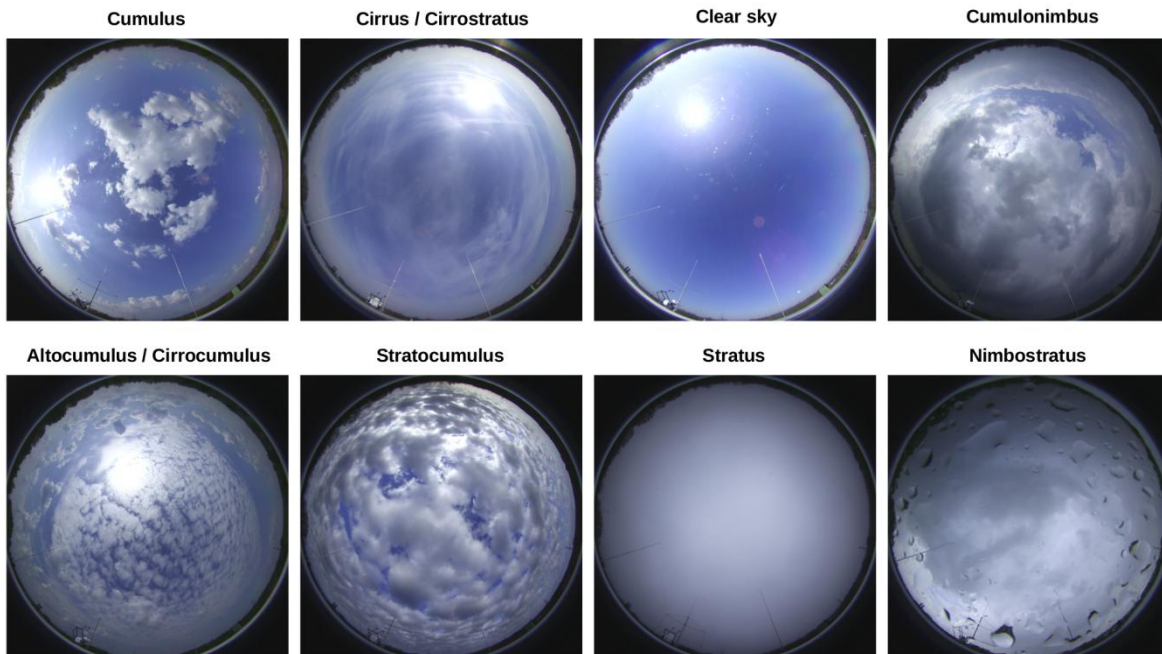

Figure 2. Typical samples of the seven cloud type classes. Nimbostratus and Cumulonimbus are summarized in one class.

It is assumed that the chosen image features represent and discriminate the different categories. The classification performance has been determined by applying a model trained on a subset of pre-classified images (70 %) on the remaining 30% of images. An accuracy of more than 99% correct classification is reached. Note that the image set used for training the classification algorithm is different from the dataset of this publication. Mixed cloud conditions or others than represented by the seven classes can be misclassified or classified with poor probability in operational use.

The frequency of each Cloud Class for the whole utilized dataset can be observed in Fig. 3. Although the data consists of summer days, only 10% of the instances correspond to clear-sky conditions. On the other hand, a total of 72% of the instances are characterized by dynamic conditions (classes 1,2,3,5), with fast transitions of irradiation intensity and significant cloud edge effects. So this poses a real challenge for the error bounds of existing state-of-the-art modeling and forecasting approaches. This will be substantiated for the Power Forecast approach later on.
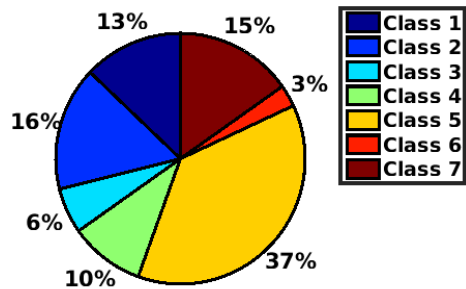

Figure 3. Cloud Class frequencies of the sample.

D. Neural Network modeling

The choice of a neural network topology was based on the requirements of the output as well as the capabilities of the available input. Energy yield calculations inherently include a form of accumulation or regression, since the acquired energy also depends on the short-term history of the system production. Additionally, the RGB forecasted values should be included in the training and testing of the model as they provide a future look into the status of the circumsolar sky. For these reasons a non-linear auto-regressive network with exogenous input was selected (NARX).

An abstract view of such a network is provided in Fig. 4a, where the utilized inputs and feedbacks are also visible. After extensive testing, a two hidden layer structure was chosen. The dataset was segmented into training (70%) and testing/validation (30%) sub-sets. In order to take advantage of each cloud class characteristics separately, a NARX was developed for each class and trained with its classified input, with the purpose of combining all 7 networks with a final layer. The final layer was chosen among a binary decision model, a weighted average model and an extra multi-layer perceptron network, where cloud class probabilities also contributed as inputs.

The binary decision model uses the network output of the class with the highest probability, therefore providing insight on the separate models. This insight will be used for comparisons in the results section. Although the weighted average model was adequately accurate in most cases, it would underperform when the classification was ambiguous, suggesting that a more intelligent approach could compensate for these accuracy losses. Therefore the MLP was chosen. The final system is presented in Fig.4b and consists of the 7 separately trained NARX, combined into a final forecast through a later trained MLP. The output of the network consists of a time-series of aggregated energy output of the system for the next 900s with a 1s resolution, retaining flexibility for all possible grid control and management applications.
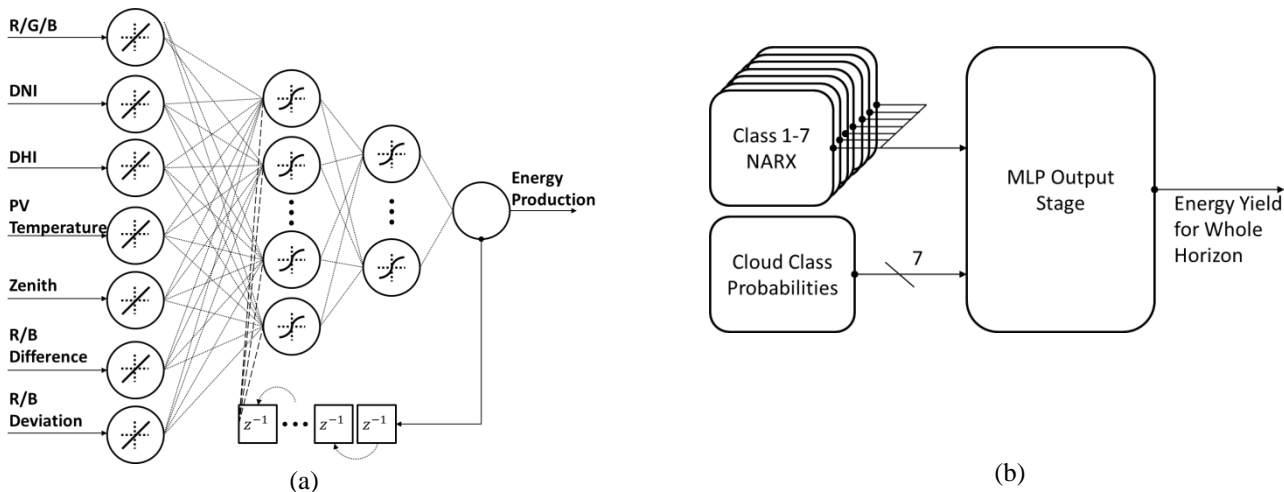

Figure 4. (a) Morphology of the NN for each class, with all inputs present, (b) Topology of the final forecast system

5

Two more distinct models were developed for utilization of the aforementioned data, with the purpose of comparing the accuracy of the proposed methodology. The neural network solution described in this publication is compared to a SoA sky-imager irradiance forecasting methodology developed by University of Oldenburg, which reports equivalent accuracy metrics with other similar models in the bibliography (Chu and Coimbra, 2017, Yang et al., 2014b). Additionally, a persistence model is used as a reference for the calculation of the forecast skill. The reference models are described in the following sections.

E. Irradiance modeling and prediction

The sky imager software developed and used at the University Oldenburg uses the following approach for the modelling and prediction of solar irradiance based on the information of sky images and local pyranometer measurements. The fundamental assumption and simplification for the transformation of binary cloud/sky information is that the actual diffuse irradiance remains constant for the considered forecast horizon. As diffuse irradiance (DHI) is determined by the overall cloud coverage, this assumption holds true, if cloud coverage does not change very fast in the forecast horizon. For this study, current DHI has been estimated by a machine-learning approach using a k-nearest-neighbour (kNN) model to estimate diffuse irradiance using various image features (see Section 3.B) as input. The model has been trained on historical data using measured DHI.

As a consequence of this approach, predicted global irradiance variability is fully determined by variability in the direct irradiance component. Direct irradiance is in turn determined by clouds shading the sensor or, respectively, blocking the sun. In case of optical thick clouds (non-transparent clouds), a binary transformation of cloud information to either 100% or 0% clear sky direct irradiance can be justified. In case of semi-transparent clouds (e.g. cirrus) the binary approach (cloud or sky) fails and either under- or overestimates the direct irradiance component. Summarized, global horizontal irradiance is computed as in Eq.1, where DHI is the kNN-estimated current diffuse irradiance and DNI the clear sky direct normal irradiance. Depending on the future cloud situation, this approach results in binary GHI forecast time series. In order to represent ramps more realistic (transition from shadow to non-shadow regions) cloud edges are smoothed with a gaussian filter applied to the binary cloud/sky image before applying the Equation above and leading to graduations of DNI for cloud edges.

$$GHI = \begin{cases} DHI & , \text{cloudy / shaded} \\ DHI + DNI * \cos(\theta) & , \text{sky/ non-shaded} \end{cases} \quad (1)$$

The translation of irradiance into power and consequently energy is achieved by performing simulation of the characterized installation with the forecasted irradiance as input and the ambient temperature kept persistent with a validated parametric model.

Finally a persistence model is developed in order to serve as a baseline for the accuracy of the proposed methodology. A sliding window with clear-sky index ratio correction is used, as described by Eq.2, where *E(t+FH)* is the forecasted energy value, *FH* is the forecast horizon and *K*(t)* signifies the clear-sky index at time *t*. This calculation is then performed for the whole horizon window, resulting in an energy trace of 1s resolution, comparable to the output of the models.

$$E(t + FH) = E(t - FH) * \frac{K^*(t + FH)}{K^*(t - FH)} \quad (2)$$

F. Error metrics

For most part of the results section, the normalized root-mean-square error is used, between measurements and forecasts from different methods (Eq.3). The errors refer to energy specifically , so for each time step the accumulated production is used. It has to be noted that the normalization is not performed with the maximum expected value of each horizon step, rather than with the mean value across all samples in order to provide a more honest estimation of the expected inaccuracies.
Especially when comparisons are performed versus persistence, the Forecast Skill Score (FSS) is a more widely accepted indicator of the improvement each method achieves (Eq. 4).

$$RMSE_{\%}(FH) = \frac{\sqrt{\sum_i^N (M_i(FH) - F_i(FH))^2 / N}}{\overline{M(FH)}} \quad (3)$$

$$FSS_{\%} = 100 * \left(1 - \frac{RMSE_{forecast}}{RMSE_{persistence}}\right) \quad (4)$$

A.   Accuracy Metrics

The developed network provides forecasts for a 15min horizon with 1s resolution. All the results that follow focus on the whole horizon and only sometimes specific horizons are chosen, in order to provide a qualitative analysis of the error progression.

The regression plots vs. measurements for the whole dataset are presented in Fig.5, for 3 different forecast horizons of 5, 10 and 15 min. There is good agreement between the expected and produced output, with highly linear behavior even for the 15 min horizon. Regression indexes remain above 90% signifying a successful capture of the behavior of the system.
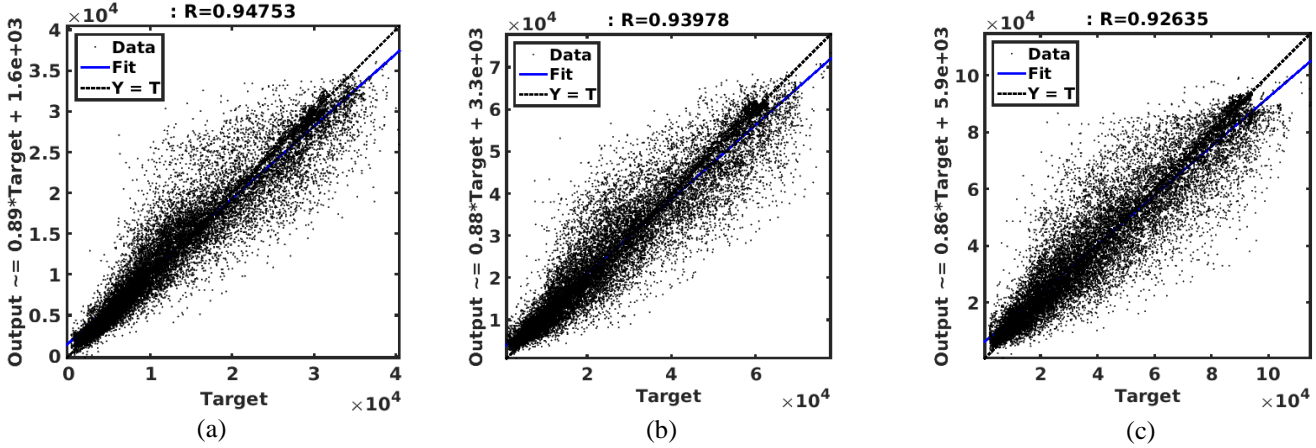

Figure 5. Regression plots of forecast vs. measurements for horizons of (a) 5min, (b) 10min and (c) 15min. Total of 18250 points for each plot.

The accuracy of each separate cloud class NN is illustrated in Fig.6, along with the final accuracy of the combined system. Additionally the errors for the binary decision model are exhibited, in order to assess the improvement enforced by the final MLP layer. Several phenomena can be detected and explained by examination and comparison of the absolute and relative RMS error:

1.   Cloud classes 1 and 3 exhibit the largest of absolute errors for forecasting, since they represent the most dynamic sky conditions, but with high insolation levels between clouds. Therefore their relative RMS is reduced to average levels because the average normalization value is reasonably high. Cloud class 5 also exhibits a high absolute RMS because of the broken clouds it entails, but the actual yield is lower and therefore the relative RMS remains substantial.

2.   Overcast situations, as included in cloud classes 6 and 7, are exhibiting the lowest absolute RMSE errors due to their minimal energy yield. Additionally, when converted to relative values the low denominators also lead to a high relative RMSE.

3.   Clear sky instances are characterized by very low absolute and relative errors, as they are trivial to model, especially in these time-scales where the clear-sky index is not altering significantly due to solar positioning. Nonetheless it is important to stress that the proposed methodology is suitable also for non-dynamic situations, where classical regression approaches would be adequate.

4.   Overall, the final MLP network manages to reduce the error and keep it in the range of 20% for all instances, therefore providing a reliable forecast for all time horizons. As concluded by the comparison of Fig. 6b and Fig.6c, the final MLP layer improves the RMS error by up to 20% for horizons of 15min. The only drawback is bigger error for the first 1 to 2 min for classes 6 and 7, due to the small accumulated energy production. This effect quickly disappears and for horizons of 5min and above leads to superior performance.

Another important improvement of the developed method can be detected if the forecast instances are categorized depending on the amount of total energy produced. For instances that correspond to a production level higher than 50% of the nominal, the forecast RMS error falls below 15% and reaches 12% for instances with more than 75% of nominal energy production. Therefore, it can be safely projected that for cases where the energy production is impactful on the dynamics of the system/grid, the current methodology offers high precision and enables detailed control schemes and task scheduling (e.g. for green servers).
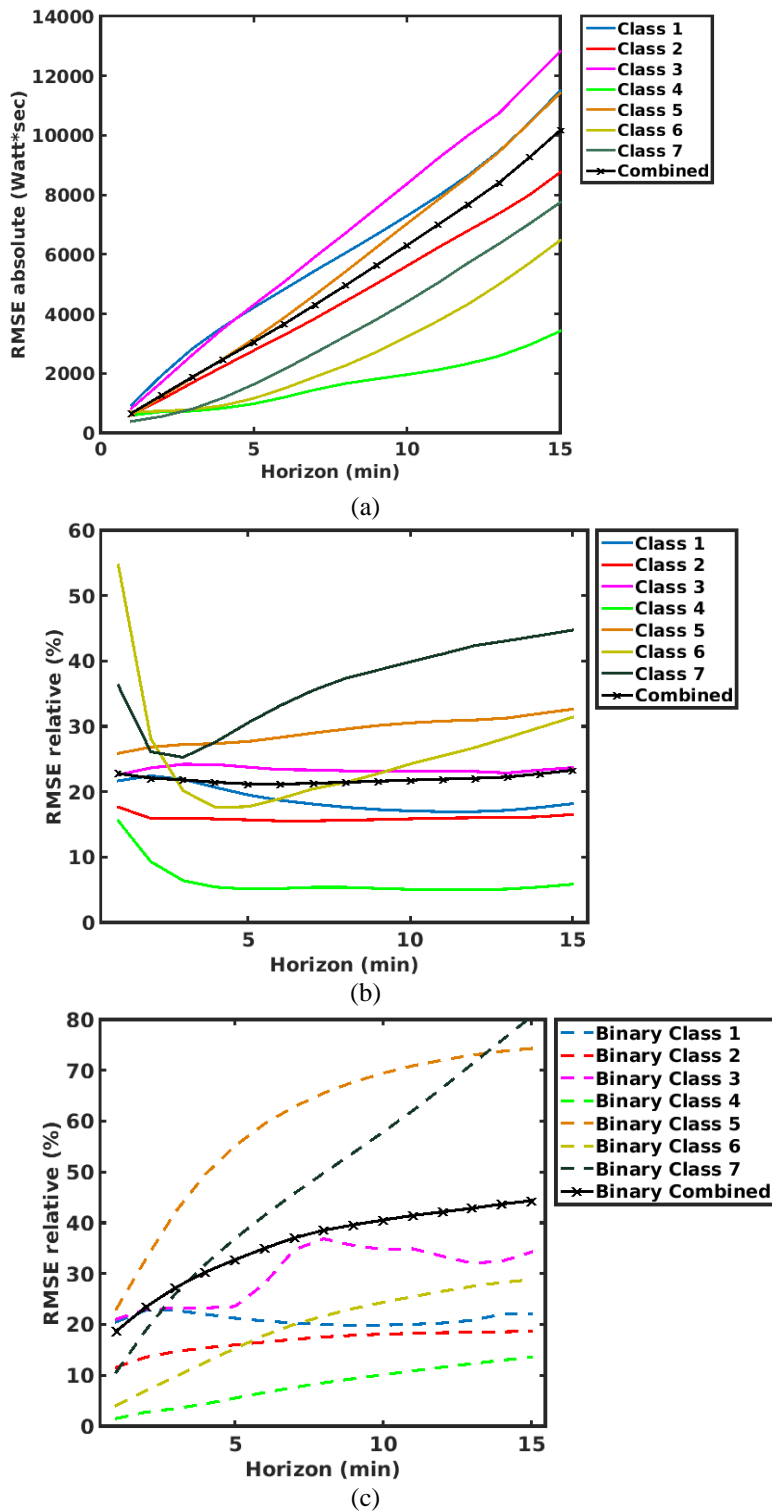
(a)

(b)

(c)

Figure 6. Collection of (a) absolute and (b) normalized RMS errors of the proposed methodology, with breakdown per Cloud Class. Also in (c) the binary decision network relative RMSE for comparison.

1

B. Comparison to Power Forecast

A summary of the errors for the developed model, the power forecast model and the persistence model is presented in Fig. 7. Both models perform better than persistence and should report a positive forecast skill, but several differences can be detected. First, the bias of the power forecast model is increasing with the horizon, reaching values up to 10%. This can be caused by several reasons, including error propagation through the cascaded models that transform irradiation forecasts to energy output. One source of this offset has been recognized in (Schmidt, 2017) , where the bias of the forecasted GHI increases with the horizon in a similar trend. Another contribution to the bias has been described also in (Anagnostos et al., 2017), which is the

inability of the parametric yield models to capture the thermal behavior of a PV system for high time resolutions. This inability usually appears as an offset on the power production estimation, due to the calculated operating temperature and can only be mitigated by a dynamic model, like the proposed one. Moreover, when the power output is accumulated to produce energy this bias can have an additive behavior and increase even further in time.

This is usually not an issue as most forecasts average data in time and space, therefore mitigating the issue, but for local control applications, more detailed time series are required. The developed model tackles the issue by using recursive networks, as suggested in (Anagnostos et al., 2017) and ensuring the correct estimation and retention of the thermal state of the PV system.

Another important observation from Fig.7 is that the power model increases its error up to the levels of persistence for 15min horizons, whereas the developed methodology retains the same levels of error for the whole time segment. Ensuring that the forecast error remains stable and bounded under a specified value (20%) is important for the deployment of control and scheduling algorithms, since there is not a need for multiple tolerances in the design of the methodologies. Additionally, a smaller error margin allows for tight specifications during design of the system, reducing potential costs for energy storage.
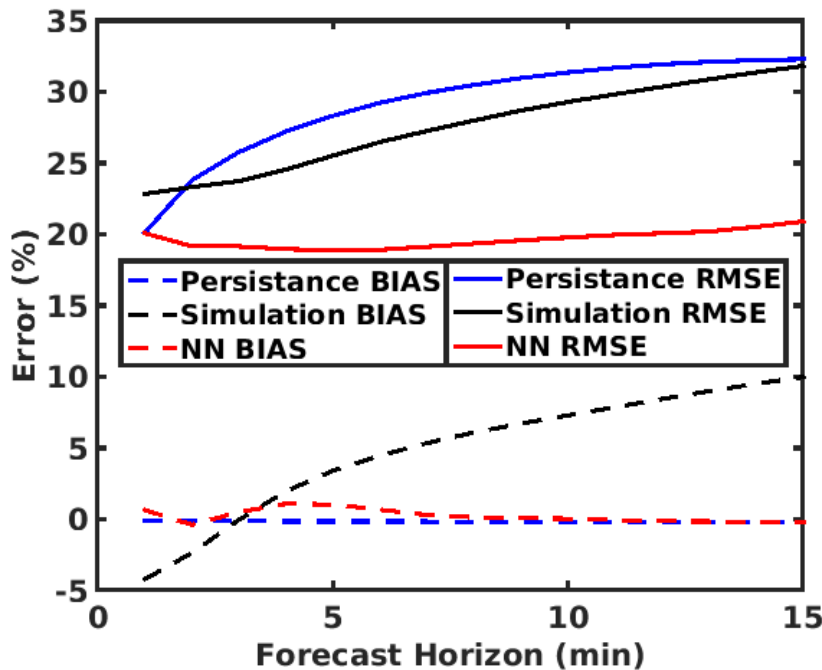

Figure 7. Summary of errors for the used models.

As expected, the low RMS error translates into an improved Forecast Skill Score as illustrated in Fig. 8. Typical values of FSS for published methods range from almost zero to 10% for power forecasts, depending on the aggregation period, the integrated area, the forecast horizon etc. (Yang et al., 2014a, Yang et al., 2014b, Kuhn et al., 2017). In this case, the time resolution is high (1s) in order to enable controlling schemes and the data correspond to one location, meaning that there is almost no gain in performance from averaging effects. Nonetheless, the power model used as a reference achieves skills of up to 10%, being at least at par with similar models. Therefore a comparison can be made with the developed model in order to quantify the improvements such an approach can achieve.

The proposed model performs clearly better on all forecast horizons with regard to persistence vs. the benchmark power model. Skill scores are always positive, as high as 39%, highlighting the inability of persistence to capture all the stochastic effects in such conditions, where averaging is not significant. Of course for longer horizons, bigger area of integration and longer time-steps reported the persistence is expected to improve its performance but this falls outside of the focus of this work, which is detailed, local, energy forecasts.
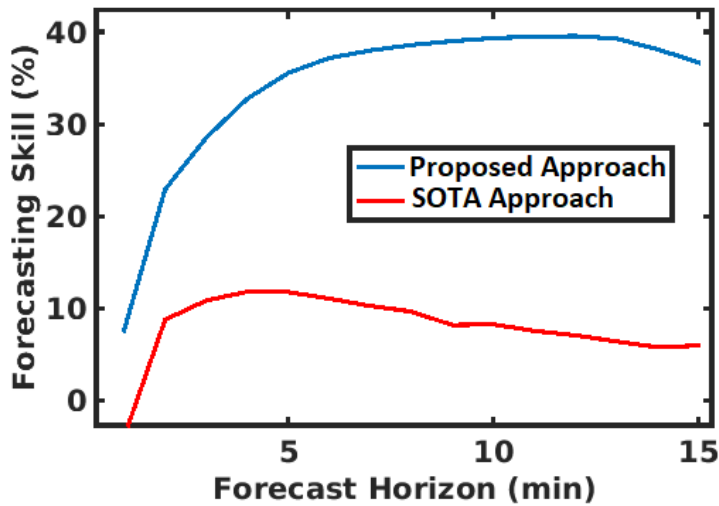
Figure 8. Energy Forecast Skill Scores of both methods for the whole horizon.

C. Missing Input Horizon and Robustness to partly missing Input Data

Another advantage of the proposed methodology is that, being based on neural network topologies, it is more robust to lost or erroneous input data. A major issue with sky-imaging forecast techniques is clipped forecast horizons, which can occur when the pixel information is not enough to apply the motion vector methodology. For example, clouds that are moving from the edge of the camera lens to the center, combined with a relatively high "speed" can lead to losses on the forecast horizon. These gaps in the forecast have to be filled either by a back-up model or by applying a corrected persistence method.

In the case of the proposed model, the persistence can be applied on the RGB inputs instead of the output, leaving the network to generalize with its trained parameters and provide a more accurate result. Indeed, as presented in Fig. 9, the forecast skill of the model suffers minimally for instances that are missing parts of the input horizon (either continuously or in segments), even up to 14min. This result ensures that the proposed forecast method can be robust even in non-optimal situations which will surely occur on the lifetime of an installation.
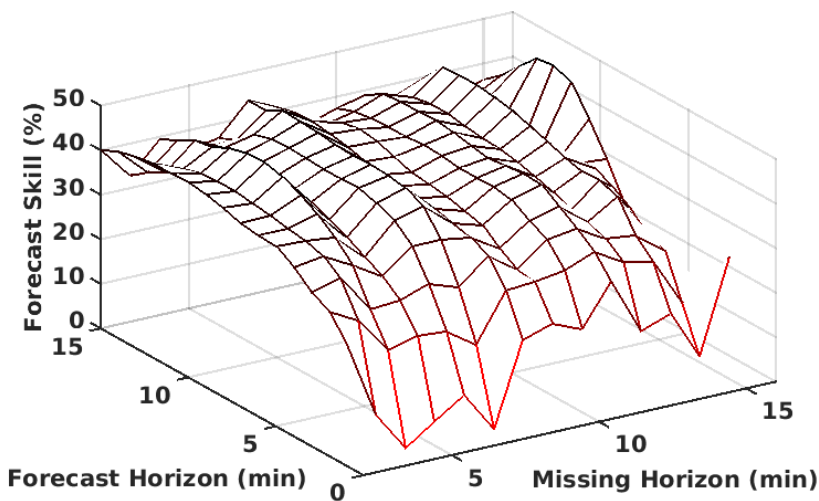

Figure 9. Forecast Skill Score evolution for instances with missing RGB forecasts.

5. CONCLUSION

Upcoming challenges in the production and distribution of renewable energy will require forecast models on all temporal and spatial scales, in order to enable smart methodologies for design and control. This paper presents a method for detailed PV

energy yield forecasting, utilizing a local sky-imager and neural networks. The proposed method eliminates the usual chain of models, from irradiation forecast to energy yield estimation, reducing the propagated errors. Additionally, the thermal dynamics of the PV system are considered, in order to better estimate the output yield based on previous research in dynamic modeling of PV systems. The shift to the energy domain and the utilization of clustering for cloud conditions result in superior performance over a state of the art irradiation forecasting model. Robustness of the forecast output is also enhanced to to the inherent extrapolation capabilities of neural network models. This methodology can prove useful for a multitude of application requiring accurate, detailed PV energy forecasting, from storage control to datacenter workload management.

REFERENCES

Anagnostos D., Goverde H., Catthoor F., Soudris D., Poortmans J., 2017, Systematic cross-validation of photovoltaic energy yield models for dynamic environmental conditions, Solar Energy, Volume 155, Pages 698-705, ISSN 0038-092X, https://doi.org/10.1016/j.solener.2017.07.011.

Anonymous, 2017, Renewables 2017-Analysis and Forecasts to 2022, Executive Summary, International Energy Agency, available Online.

Aksanli B., Venkatesh J., Zhang L., and Rosing T., 2011, Utilizing green energy prediction to schedule mixed batch and service jobs in data centers, In Proceedings of the 4th Workshop on Power-Aware Computing and Systems (HotPower '11), ACM, New York, NY, USA, , Article 5 , 5 pages, DOI=http://dx.doi.org/10.1145/2039252.2039257.

Avgerinou M., Bertoldi P., Castellazzi L., 2017, Trends in Data Centre Energy Consumption under the European Code of Conduct for Data Centre Energy Efficiency, Energies 2017, 10, 1470.

Chen C., Duan S., Cai T., Liu B., 2011, Onine 24-h solar power forecasting based on weather type classification using artificial neural network, Solar Energy,Volume 85, Issue 11, Pages 2856-2870, ISSN 0038-092X, https://doi.org/10.1016/j.solener.2011.08.027.

Chu Y., Coimbra C.F.M., 2017, Short-term probabilistic forecasts for Direct Normal Irradiance, Renewable Energy, Volume 101, Pages 526-536, ISSN 0960-1481, https://doi.org/10.1016/j.renene.2016.09.012.

Elsinga B, van Sark W.G.J.H.M., 2017, Short-term peer-to-peer solar forecasting in a network of photovoltaic systems, Applied Energy, Volume 206, Pages 1464-1483, ISSN 0306-2619, https://doi.org/10.1016/j.apenergy.2017.09.115.

Goiri I., Haque E., Beauchea K.L.R, Nguyen T.D., Guitart J., Torres J., Bianchini R., Matching renewable energy supply and demand in green datacenters, Ad Hoc Networks, Volume 25, Part B, Pages 520-534, ISSN 1570-8705, https://doi.org/10.1016/j.adhoc.2014.11.012.

Golshannavaz S., Afsharnia S. and Aminifar F., 2014, Smart Distribution Grid: Optimal Day-Ahead Scheduling With Reconfigurable Topology, in IEEE Transactions on Smart Grid, vol. 5, no. 5, pp. 2402-2411, doi: 10.1109/TSG.2014.2335815.

Goverde, H., Anagnostos, D., Herteleer, B., Govaerts, J., Baert, K., Aldalali, B., Catthoor, F., Driesen, J., Poortmans, J., 2015, Model requirements for accurate short term energy yield predictions during fast-varying weather conditions, 31st European Photovoltaic Solar Energy Conference (PVSEC).

Haralick R., Shanmugam K., Dinstein I., 1973, Textural Features for Image Classification, Systems, Man and Cybernetics, IEEE Transactions on, SMC-3(6):610–621.

Heinle A., Macke A., Srivastav A., 2010, Automatic cloud classification of whole sky images. Atmospheric Measurement Techniques, 3(3):557–567.

Kuhn P., Wilbert S., Schüler D., Prahl C., Haase T., Ramirez L., Zarzalejo L., Meyer A., Vuilleumier L., Blanc P., Dubrana J., Kazantzidis A., SchroedterHomscheidt M., Hirsch T., and Pitz-Paal R., 2017, Validation of spatially resolved all sky imager derived DNI nowcasts, AIP Conference Proceedings, https://doi.org/10.1063/1.4984522.

Li Y., Orgerie A.C., Menaud J.M., 2017, Balancing the Use of Batteries and Opportunistic Scheduling Policies for Maximizing Renewable Energy Consumption in a Cloud Data Center, 25th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP), St. Petersburg, 2017, pp. 408-415 doi: 10.1109/PDP.2017.24.

Lorenzo A.T., Holmgren W.F., Cronin A.D., 2015, Irradiance forecasts based on an irradiance monitoring network, cloud

motion, and spatial averaging, Solar Energy, Volume 122, Pages 1158-1169, ISSN 0038-092X, https://doi.org/10.1016/j.solener.2015.10.038.

Mégel O., Mathieu J.L., Andersson G., 2015, Scheduling distributed energy storage units to provide multiple services under forecast error, International Journal of Electrical Power & Energy Systems, Volume 72, Pages 48-57, ISSN 0142-0615, https://doi.org/10.1016/j.ijepes.2015.02.010.

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

Pedro H.T.C., Coimbra C.F.M., 2012, Assessment of forecasting techniques for solar power production with no exogenous inputs, Solar Energy, Volume 86, Issue 7, Pages 2017-2028, ISSN 0038-092X, https://doi.org/10.1016/j.solener.2012.04.004.

Shields J.E., Karr M.E., Tooman T.P., Sowle D.H., Moore S.T., 1999, The whole sky imager – a year of progress.

Schmidt T., Kalisch J., Lorenz E., Heinemann D., 2016, Evaluating the spatio-temporal performance of sky-imager-based solar irradiance analysis and forecasts, Atmospheric Chemistry and Physics, 10.5194/acp-16-3399-2016.

Schmidt T., 2017, High resolution solar irradiance forecasts based on sky images, PhD thesis, Universität Oldenburg, available online

Tzoumanikas P., Nikitidou E., Bais A.F., Kazantzidis A., 2016, The effect of clouds on surface solar irradiance, based on data from an all-sky imaging system, Renewable Energy, Volume 95, Pages 314-322, ISSN 0960-1481, https://doi.org/10.1016/j.renene.2016.04.026.

Yang D., Dong Z., Reindl T., Jirutitijaroen P., Walsh W.M., 2014a, Solar irradiance forecasting using spatio-temporal empirical kriging and vector autoregressive models with parameter shrinkage, Solar Energy, Volume 103, Pages 550-562, ISSN 0038-092X, https://doi.org/10.1016/j.solener.2014.01.024.

Yang H., Kurtz B., Nguyen D., Urquhart B., Chow C.W., Ghonima M., Kleissl J., 2014b, Solar irradiance forecasting using a ground-based sky imager developed at UC San Diego, Solar Energy, Volume 103, Pages 502-524, ISSN 0038-092X, https://doi.org/10.1016/j.solener.2014.02.044.