



Smart4RES

Toolbox of multi-source data approaches to short term RES forecasting

D3.1 Toolbox of multi-source data approaches to short term RES forecasting

WP3, T3.1

Version V2.0

Authors:

Eva Droste, Johannes Jung, Heidrun Misfeld, Björn Witha, Matthias Lange, *EMSYS*

Jorge Lezaca, Fernanda Norde Santos, *DLR*

Tuhfe Göcmen, Jens Visbeck Madsen, Jennifer Rinker, *DTU*

George Sideratos, Nikos Hatzargyriou, Aris Dimeas, Theodoros Konstantinou, *ICCS*





Disclaimer

The present document reflects only the author's view. The European Climate, Infrastructure and Environment Executive Agency (CINEA) is not responsible for any use that may be made of the information it contains.



**Technical references**

Project Acronym	Smart4RES
Project Title	Next Generation Modelling and Forecasting of Variable Renewable Generation for Large-scale Integration in Energy Systems and Markets
Project Coordinator	ARMINES – MINES ParisTech
Project Duration	November 2019 – April 2023
Deliverable	D3.1: Toolbox of multi-source data approaches to short term RES forecasting
Dissemination level ¹	PU
Nature ²	R
Work Package	WP 3 – Data Science and the Future of RES Forecasting
Task	T 3.1 – Multi-source data approaches to short-term RES forecasting
Lead beneficiary	energy & meteo systems GmbH (EMSYS)
Contributing beneficiary(ies)	EMSYS, DLR, DTU, ICCS
Reviewers	Simon Camal (ARMINES) Christos Vitellas (HEDNO) Theodora Anastopoulou (HEDNO)
Due date of deliverable	30 April 2022

- ¹ PU = Public
PP = Restricted to other program participants (including the Commission Services)
RE = Restricted to a group specified by the consortium (including the Commission Services)
CO = Confidential, only for members of the consortium (including the Commission Services)
- ² R = Report, P = Prototype, D = Demonstrator, O = Other

Document history

V	Date	Description
0.1	01/02/22	Template distributed to partners
0.2	14/04/22	First draft submitted by EMSYS
1.1	20/04/22	Review by Simon Camal, Christos Vitellas and Theodora Anastopoulou
1.2	29/04/22	Revised and final version submitted by EMSYS
2.0	29/04/22	Final version submitted to the European Commission





Executive summary

This report describes the work developed by DLR, DTU, EMSYS and ICCS in the framework of Task 3.1 ("Multi-source data approaches to short term RES forecasting") of the Smart4RES project.

The goal of this task was to develop a toolbox of data-science based methods that combine data collected from multiple sources with RES forecasting models and to demonstrate an improvement of RES power forecasting skill in line with the KPIs of the project.

The following approaches or "tools" have been developed independently by the involved partners as indicated and successfully validated:

- PV power forecasting using the combined Satellite and AllSkylmager Irradiance Forecast (DLR) → up to 18% RMSE improvement compared to only Allskylmager forecast
- Using Skymager data to improve the minute ahead PV power forecasting (EMSYS) → up to 23% RMSE improvement compared to operational EMSYS forecast
- Improving the minute to hour ahead PV forecast during thunderstorms by using lightning data and power measurements (EMSYS) → up to 14% RMSE improvement compared to operational EMSYS forecast
- Obtaining improved power curves through advanced outlier data filtering and machine learning techniques (EMSYS) → up to 8% RMSE improvement compared to operational EMSYS forecast
- LiDAR based second-ahead forecasting of wind power and structural loads (DTU) → up to 35% RMSE improvement compared to persistence forecast
- Investigating the capabilities of AI-based dynamic forecasting approaches and additional SCADA data channels (DTU) → up to 46% RMSE improvement of dynamic GRU networks compared to single GRU networks and persistence
- Blending satellite irradiance maps with NWP data and power time series to improve hour-ahead PV forecasting (ICCS) → up to 5% MAE improvement compared to forecast without satellite irradiance
- Applying AI techniques and combining two NWPs to improve the day-ahead solar power forecasting (ICCS) → up to 17% MAE improvement compared to single NWP forecast

Most of these approaches integrate new sources of data, that have not been used in the context of RES forecasting before, into existing RES forecasting models. It could be demonstrated that all of these approaches yield a significant improvement of RES power forecasting skill in alignment with the Forecasting KPIs of the project, mostly reaching or exceeding the project KPI target values of 9-12% (solar) and 7-9% (wind) RMSE improvement for the up to 30-min ahead forecast.



Table of contents

Executive summary.....	4
Table of contents.....	5
LIST OF TABLES.....	7
LIST OF FIGURES.....	7
1 Introduction.....	11
2 PV power forecasting using the combined Satellite and ASI Irradiance Forecast (DLR).....	12
2.1 Motivation.....	12
2.2 Datasets and case studies.....	12
2.3 Methodology.....	14
2.3.1 PV model description.....	14
2.3.2 PV model Optimization.....	15
2.3.3 PV plant forecast.....	15
2.4 Results and Discussion.....	16
2.4.1 Transposition model comparison.....	16
2.4.2 PV model loss parameter optimization.....	17
2.4.3 High resolution PV forecasting comparison.....	18
3 PV power forecasting with Skylmager data (EMSYS).....	24
3.1 Motivation.....	24
3.2 Case studies.....	24
3.3 Approach.....	25
3.4 Evaluation.....	27
3.5 Results and Discussion.....	28
4 Improved thunderstorm correction for PV forecast (EMSYS).....	30
4.1 Motivation.....	30
4.2 Approach.....	30
4.3 Case studies.....	32
4.4 Evaluation.....	32
4.5 Results and Discussion.....	33
5 Wind power forecasting based on improved power curve modelling (EMSYS) ...	35
5.1 Motivation.....	35
5.2 Outlier data filtering.....	35
5.3 Wind power forecasting based on improved power curve modelling.....	37
5.4 Evaluation.....	38





5.5	Results and Discussion.....	38
6	LiDAR based second-ahead forecasting of wind power (DTU)	41
6.1	Motivation	41
6.2	Overall Approach	41
6.3	Case study for model benchmark.....	41
6.3.1	DTU Risø Test Site	41
6.3.2	Data acquisition, processing and feature selection.....	43
6.4	Forecast models	44
6.4.1	Feed-forward neural network (FFNN)	44
6.4.2	Long-short-term memory (LSTM) network	44
6.4.3	Gated Recurrent Unit (GRU)	45
6.4.4	Random Forest (RF)	45
6.5	Evaluation of the Results & Discussion	45
7	PossPOW SCADA based forecasting (DTU)	48
7.1	Motivation	48
7.2	Case studies.....	48
7.2.1	La Haute Borne onshore wind farm	48
7.2.2	Description of Data, Processing & Model Training	49
7.3	Evaluation of the Results and Discussion.....	49
7.3.1	Single (Generic) GRU Network	50
7.3.2	Dynamic GRU Networks (via Transfer Learning)	50
8	Satellite-based PV forecasting (ICCS)	52
8.1	Motivation	52
8.2	Approach.....	53
8.3	Case studies.....	57
8.4	Evaluation.....	58
8.5	Results and Discussion.....	58
9	Day-ahead solar power forecasting using NWP obtained by different vendors	
	62	
9.1	Motivation	62
9.2	Approach.....	62
9.3	Case studies.....	65
9.4	Results and Discussion.....	66
10	Conclusion	67
11	Appendix A: Summary of Smart4RES KPIs	68
11.1	Project KPIs	68





11.1.1	Forecasting KPIs	68
11.1.2	KPIs on grid management applications	69
11.1.3	KPIs on market applications.....	70

LIST OF TABLES

Table 1: Data configuration for the evaluation of PV power prediction performance.	13
Table 2: PV test sites (with pseudonyms) used on the evaluation of the performance of the irradiance forecasts on PV power prediction.	13
Table 3: Forecasts descriptions.	16
Table 4: nRMSE of PV power forecast method for selected lead-times on the OLD_EAST site. Improvement of nRMSE of Satellite + ASI over only SAT and only ASI..	21
Table 5: nRMSE per forecast methods for selected lead-times on the OLD_CENTER site. Improvement of nRMSE of Satellite + ASI over only SAT and only ASI.	22
Table 6: Details of the DLR's Eye2Sky data set and their location in Northwest Germany.	25
Table 7: Final input features to the power and load forecaster.	44
Table 8: RF hyperparameters.	45
Table 9: Architecture of the Single (generic) GRU network with lag = 2-hours. Hidden layers: GRU_1, GRU_2, GRU_3. Output Layer: Dense_1. The activation functions used are relu, sigmoid and tanh, respectively.	49
Table 10: NMAE from the Rhodes case-study using timeseries data.....	58
Table 11: NMAE from the Rhodes case-study using timeseries and NWP data.....	58
Table 12: NMAE from the Rhodes case-study obtained using all data and the irradiance map pixels are selected by RFVII.	59
Table 13: NMAE from the Rhodes case-study obtained using all data and the irradiance map pixels are selected by gLasso.....	59
Table 14: NMAE from the Rhodes case-study obtained by RF method for each trial.	59
Table 15: NMAE from the Rhodes case-study obtained by the deep learning architectures on each trial.	59
Table 16: Normalised mean absolute error of day-ahead predictions.	66

LIST OF FIGURES

Figure 1: Eye2sky Network overview. Camera stations (blue circles) are camera only stations, Meteo stations (brown circles) are camera + meteorological stations equipped with an RSIs and Reference, the green circles are only meteorological stations equipped with thermal radiometers + solar tracker.....	13
Figure 2: Simplified PV model for a single string.....	14
Figure 3: Monthly bias (left) and RMSE (right) of the POA irradiance estimated with various transposition models for the meteorological station LEEER. (1min average resolution data).....	17





Figure 4: Optimization of the loss and bias parameters for the PV sites OLD_EAST (left) and OLD_CENTER (right). 17

Figure 5: Correlation of measured and modelled AC power for the analysis case (forecast lead-time 0) of all forecast sources on the site OLD_EAST (1 min resolved data for the month of 08.2020). 18

Figure 6: Correlation for measured and modelled AC power for the analysis (forecast lead-time 0) of all forecast sources on the site OLD_CENTER (1 min resolved data for the month of 08.2020). 19

Figure 7: PV forecast performance for the site OLD_EAST. All data used has a temporal resolution of 1min. The Combined ASI + SAT forecast was trained on 07.2020 using 4 Eye2sky stations: OLCLO, OLDON, OLUOL, PVAMM. nMBE, nMAE and nRMSE are normalized to the installed capacity. 20

Figure 8: PV forecast performance for the site OLD_CENTER. All data used has a temporal resolution of 1min. The Combined ASI + SAT forecast was trained on 07.2020 using 4 Eye2sky stations: OLCLO, OLDON, OLUOL, PVAMM. nMBE, nMAE and nRMSE are normalized to the installed capacity. 21

Figure 9: Scatter plot of power derived from Skymager real-time irradiances against power from the PV plant for the three different time periods (left) and time series plot of those data for 2020-06-06 (right). 25

Figure 10: Example of a Skymager irradiance map (left) with the location of the PV plant (red cross) and the timeseries (right) of the extracted power from the real-time irradiances (red) and the nowcast (cyan) together with the measurement from the PV plant (black). The lower right plot is a close-up of the upper right plot. 26

Figure 11: Scheme of the forecasting system Suncast and its components: NWP forecasts (1), measurements from PV plants (2) and Skymager data as additional data source (3). 27

Figure 12: Timeseries of measured power from the PV plant (black) with Suncast forecast (left) and Skymager+Suncast forecast (right) for 2020-06-06. 28

Figure 13: Daily RMSE of Suncast forecast (orange), Skymager nowcast (cyan) and Skymager+Suncast forecast (violet) for the three time periods in March (upper), June (middle) and November (lower). The dashed lines are the averaged RMSEs over each time period. 29

Figure 14: Base forecast (red line) and the different shortest-term corrections (coloured lines) during two days with thunderstorms. Measurements in black. 31

Figure 15: Example of the lightning data by nowcast used for thunderstorm detection. 32

Figure 16: Mean values of RMSE (top), SDE (centre) and Bias (bottom) as a function of forecast horizon for different shortest-term corrections. 33

Figure 17: (a) The highlighted points were detected with the method of Zhao et al. (2018). The outlier plateau is identified. However, the vertical green bar in the middle consists of many false positives. (b) The result of our variant of this method. 35

Figure 18: The left plot shows a heatmap of the raw data in the speed-power-view. The right plot shows a heatmap of the data after we applied the discussed outlier detection method, where the plateau was removed. 36

Figure 19: Example of sliding-window filtering based on outlier candidates. The black dots show valid data points in the time series view. The red dots show outlier



candidates. The result of the sliding-window filter is indicated by the smoothed signal in grey. 37

Figure 20: This plot shows the mean RMSE for various models for six datasets from different wind farms. For the three machine learning methods several models with different feature sets were trained. If no features are given in brackets, all available features were used. Each model was trained on the data from one year and evaluated on the data from the following year. The RMSE values were scaled for each dataset, such that the reference model has an RMSE of 1.0. The error bars show the bootstrap confidence intervals with 95% confidence level. 38

Figure 21: This figure shows the total gain (to be interpreted as "importance") for various features for multiple datasets. The bar plots show the average importance values for the respective feature. The scatter plots (with jitter for visualization purposes) show the raw importance values. The y-axes are broken to get detailed insights into the importance scores. 39

Figure 22: Risø test site, including the location of the Vestas V52 turbine and the met-mast. 42

Figure 23: Configurations of the LiDAR beams. Blue is the 4-beamed LiDAR and red is the 2-beamed LiDAR. 42

Figure 24: Example of a 50 Hz time series of power from SCADA, flap-wise bending moment from strain gauge measurements and LOS wind speeds from the 2- and 4-beamed LiDAR systems. 43

Figure 26: Power forecast benchmark, evaluated on mean absolute error (MAE) and normalised root mean square error (RMSE) using; i) left: 2-beam lidar, ii) right: 4-beam lidar. 46

Figure 27: Structural load forecast benchmark, evaluated on mean absolute error (MAE) and normalised root mean square error (RMSE) using; i) left: 2-beam lidar, ii) right: 4-beam lidar. 47

Figure 28: 20min time series comparison of 5-s ahead RF flapwise bending moment forecasts. 47

Figure 29: La Haute Borne, i) left: Wind rose, ii) right: wind farm layout with turbine IDs. 48

Figure 30: 10-min ahead La Haute Borne wind farm power forecast performance in percentage error and RMSE, i) left: persistence forecasts, ii) right: single (generic) GRU network forecasts. 50

Figure 31: 10-min ahead La Haute Borne wind farm power forecast performance in percentage error and RMSE, dynamically updated GRU networks (via transfer learning). 51

Figure 32: Solar irradiance map of Greece in Wh/m². There is no information above the sea and above small islands. 52

Figure 33: The correlation maps for 1, 2, 3 and 5 hour-ahead horizons with the respective maxima. 53

Figure 34: The structure of the model that uses only timeseries data and it is used as benchmark. 56

Figure 35: The structure of the model that uses timeseries data and NWP's and it is used as second benchmark. 56

Figure 36: The structure of the proposed model that uses timeseries data, NWP's and irradiance maps. 57





Figure 37: Predictions obtained by the RF method and the RFVII pixel selection method for every trial and for 1, 3 and 5 hour-ahead forecasting horizon.	60
Figure 38: Predictions obtained by the different forecasting method the RFVII pixel selection method for 1, 3 and 5 hour-ahead forecasting horizon.....	60
Figure 39: The RBF-CNN regressor structure.	63
Figure 40: The forecasting model architecture (coloured branches signify the activated clusters).....	63
Figure 41: Post-processing solar power forecasting model.	64
Figure 42: Ensemble PDE-NN model architecture.	65



1 Introduction

As stated in the project proposal, the overarching goal of Smart4RES is to aim at developing and validating the next generation tools that jointly enable 1) an increase of at least 15% in RES forecasting performance and 2) enhanced value in applications thanks to a holistic approach that considers the whole model and value chain of RES electricity production forecasting.

This deliverable presents the outcome of Smart4RES Task 3.1 ("Multi-source data approaches to short term RES forecasting") in which a toolbox of different data-science based approaches had to be developed combining data collected from different sources and in different temporal domains, aiming at improving RES power forecasting skill. Several partners (DLR, DTU, EMSYS and ICCS) contributed to the work with independently developed methods (also referred to as "tools").

Besides the main goal of improving RES power forecasting skill, one objective was to develop complementary approaches to blending information from multiple sources of information.

Various types of data from different sources have been used in combination with existing RES forecasting models: satellite images, solar irradiance measurements from satellites, Skycam images, LiDAR remote sensing data, lightning data, SCADA and power data from wind farms and data from numerical weather prediction models.

The following multi-source data approaches have been successfully explored and validated and will be described in detail in the following sections:

- PV power forecasting using the combined Satellite and AllSkylmager Irradiance Forecast (DLR, section 2)
- Using Skyimager data to improve the minute ahead PV power forecasting (EMSYS, section 3)
- Improving the minute to hour ahead PV forecast during thunderstorms by using lightning data and power measurements (EMSYS, section 4)
- Obtaining improved power curves through advanced outlier data filtering and machine learning techniques (EMSYS, section 0)
- LiDAR based second-ahead forecasting of wind power and structural loads (DTU, section 6)
- Investigating the capabilities of AI-based dynamic forecasting approaches and additional SCADA data channels (DTU, section 7)
- Blending satellite irradiance maps with NWP data and power time series to improve hour-ahead PV forecasting (ICCS, section 8)
- Applying AI techniques and combining two NWPs to improve the day-ahead solar power forecasting (ICCS, section 9)

These studies are validation studies which integrate mostly new data sources, that have not been used in the context of RES forecasting before, into existing RES forecasting models. As such, they necessitated ad-hoc processing and developments of RES forecasting models.

2 PV power forecasting using the combined Satellite and ASI Irradiance Forecast (DLR)

2.1 Motivation

Nowadays PV power forecasting is commonly performed using Satellite and NWP irradiance forecasts, typically processed for intra-day horizons, i.e., minutes-ahead to hours-ahead horizons (Agoua et al., 2018). These irradiance forecasts present a low spatial and temporal resolution with the advantage of having a very wide coverage. This results on accurate regional PV forecasts but the accuracy lowers in the case of localized PV forecasts (e.g., individual PV plants). In this context, forecasts based on All Sky Imagers (**ASI**) have been developed. This type of forecast is able to detect clouds with a higher spatial and temporal resolution and therefore present a better localized forecast performance. The DLR has developed in the subtask 2.3.1 of WP2 an ASI network-based forecast which has a spatial resolution of 50m and forecast horizons from 20 to 60 mins (depending on the cloud height and wind conditions).

In order to take advantage of the individual strengths of the different forecasts, the DLR has also developed a highly resolved combined irradiance forecast in the subtask 2.3.2 of WP2. This new forecast is able to combine forecasts with very different spatio-temporal resolutions and achieve an optimized forecasting product that shows a lower statistical error (nRMSE and nMAE) than all the individual input forecasts on the highest spatio-temporal resolution available.

A highly accurate highly-resolve local forecast can be very interesting for individual PV plant owners and grid distributors as this new information can be part of an optimized plant management strategy and daily decision-making strategies. Within this framework, the DLR has decided to perform a comparison case study of the PV power prediction performance on a local level using an operational Satellite forecast and the newly developed combined ASI and Satellite forecasts.

2.2 Datasets and case studies

The evaluation of the PV forecasting performance of the different irradiance forecasts are based on the **dataset #1**, defined in D.1.3 in section 2.2 as the “**NorthWest meteorological measurement network & PV plants**”. This dataset is based on the measurements performed in the DLR’s Eye2Sky network which is shown in Figure 1. The data used on this study is presented in Table 1.

Data type	Parameter type	Locations	Start	End	Duration	filtering
Study on the transposition models	GHI, DHI, DNI, GTI	LEEER	2020-03-01	2021-02-28	1 year	Day data
Training of the combined irradiance forecast	GHI, temperature	OLUOL, OLDON, PVAMM, OLCLO	2020-07-01	2020-07-31	1 month	Solar elevation > 20°

Optimization of the PV model losses	PV Power	OLD_EAST, OLD_CENTER	2020-08-07	-	1 fully clear sky day	Solar elevation > 20°
	GHI	OUOL, OLDON				
Validation of PV power forecast	PV power	OLD_EAST, OLD_CENTER	2020-08-01	2020-08-31	1 month	Solar elevation > 20°

Table 1: Data configuration for the evaluation of PV power prediction performance.

Eye2Sky - Cloud camera and meteorological measurement network in Oldenburg

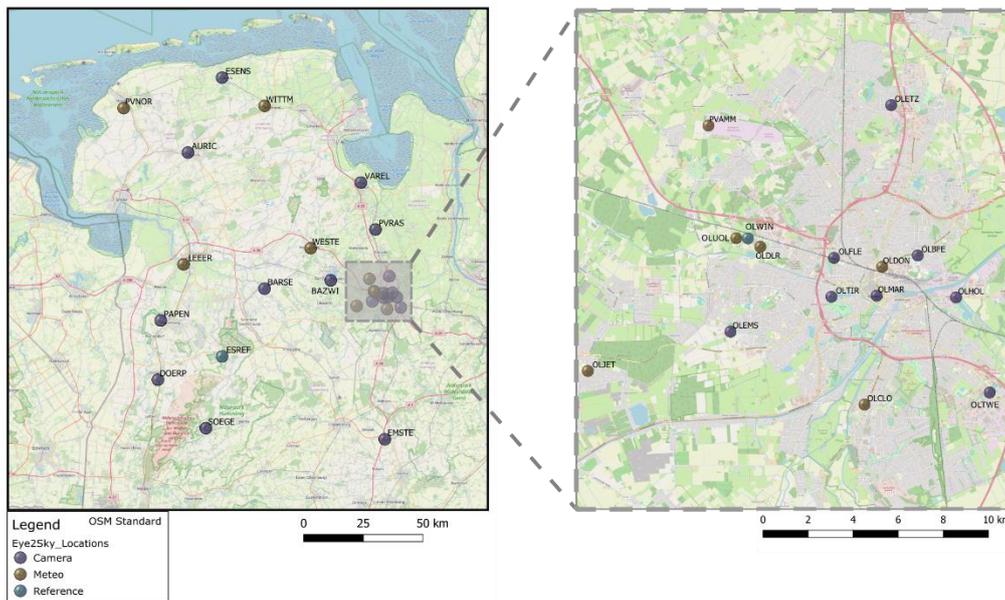


Figure 1: Eye2sky Network overview. Camera stations (blue circles) are camera only stations, Meteo stations (brown circles) are camera + meteorological stations equipped with an RSIs and Reference, the green circles are only meteorological stations equipped with thermal radiometers + solar tracker.

The validation of the power forecast is restrained to one summer month because for this study only 2 full consecutive months of the ASI network forecast were processed: 07.2020 and 08.2020. As the combined forecasts needs 30 historical days for the training, the first available day for the validation using the combined forecast is the 01.08.2020, thus restricting the validation period to the month of 08.2020.

To evaluate the performance of the different forecasts to predict PV power, we have selected 2 PV sites that are inside the Eye2Sky domain. These sites are described in Table 2. They are both rooftop installed PV plants.

Parameter	OLD_EAST	OLD_CENTRE
Install AC power	27 kW	25 kW
Module power	315 W	230 W
Number of modules	S1=48 S2 = 48	110
Modules tilt	S1=10° S2=10°	38°
Modules azimuth	S1 = 60° S2 =240°	253°
Inverter nominal power	S1 = 15kW S2 =15kW	30 kW

Table 2: PV test sites (with pseudonyms) used on the evaluation of the performance of the irradiance forecasts on PV power prediction.

2.3 Methodology

In order to produce a PV power forecast from an irradiance forecast a physical model of the PV plant needs to be developed. Due to the simplicity of the PV plants, a physical model is a good choice as we have all the parameters necessary to fully describe it. We have chosen to use the simplified physical string model shown in Figure 2. The model is based on the work from King in Sandia National Laboratories (King et al., 2004) and implemented using the python PVLIB library (Holmgren et al., 2018).

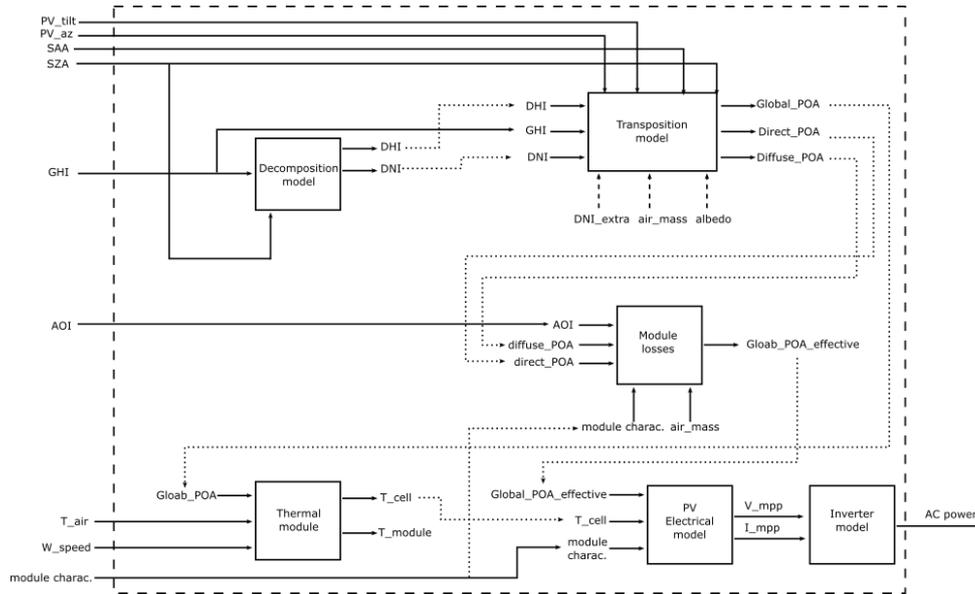


Figure 2: Simplified PV model for a single string.

2.3.1 PV model description

The basic blocks of the model are described here below.

- Decomposition model block:** for this study we have available the Global Horizontal Irradiance (GHI) components of the irradiance forecasts. In order to get the effective available irradiance on the plane of the modules, we need to calculate the Direct Normal Irradiance (DNI) and the Diffuse Horizontal Irradiance (DHI). For the calculation of the DNI, we use the DRINT model (Perez et al., 1992). The DHI is found using the closing equation

$$DHI = GHI - DNI \cos(SZA)$$

where SZA is the Solar Zenith Angle.

- Transposition model block:** Transposition models estimate the global irradiance available on an inclined surface. This is known as Global Tilted Irradiance (GTI) or Plane Of Array (POA) irradiance in the context of PV modules. It is modelled as the sum of the direct, diffuse, and reflected irradiances. The models mostly differ in the way they calculate the diffuse component. The most commonly used transposition models are described in (Garcia et al., 2013).
- Module losses block:** This block models the reflection losses on the module surface and the spectral loss caused by the selective spectral response of the semiconductor material of the PV cells. These losses are modelled using the method developed by King (King et al., 2004).

- **Module Temperature block:** This block estimates the temperature of the PV cell using the temperature model from King (King et al., 2004).
- **PV electrical model block:** In this module the POA effective irradiance is translated into DC power ($V * I$) considering the losses due to the thermal state of the PV cell. Here the PVLIB library has several model implementations. We have opted for the SAPM model (King et al., 2004).
- **Inverter model block:** This module determines the AC power output of the inverter as the result of a DC input voltage and current. Here we have opted for the Sandia Grid-Connected PV inverter model (King et al., 2007).

2.3.2 PV model Optimization

Once the PV model is set up, we proceed with a 2-step optimization of the model

- **Step 1 – Transposition model selection:** Here we used several transposition models to estimate the POA irradiance from the GHI, DHI and DNI measurements. Then we compare these estimations with the measurement of the POA irradiance to find the model that better fits our local conditions. We choose five transposition models to compare: **Liu-Jordan**, **Skartveit**, **Gueymard**, **Reindl** and **Harrison Coombes**. The **Liu-Jordan** isotropic model assumes that the diffuse irradiance is uniformly distributed over the sky globe. The **Skartveit** and **Gueymard** models also consider a circumsolar diffuse component. **Reindl** and **Harrison Coombes** add to that a horizontal brightening diffuse component. A detail of the models can be found in (Garcia et al., 2013) and (Harrison et al., 1988). The implementation has been done using the python PVLIB library for the **Liu-Jordan** and **Reindl** models and our own python implementation for **Skartveit**, **Gueymard** and **Harrison Coombes**.
- **Step 2 - Loss parameter optimization:** in order to consider the losses that are difficult to infer/model from the basic PV plant characteristics (e.g., soiling deposition, DC losses due to cabling, mismatching, shadowing, etc.) an extra loss coefficient and bias are added to the model:

$$AC_{mod_corrected} = L * AC_{mod} + B$$

The coefficients **L** and **B** are found by applying a linear regression on the modelled and measured AC power.

2.3.3 PV plant forecast

Once the PV models are optimized for the individual plants, we proceed to calculate the local PV power forecast produced for each of the irradiance forecasts inputs. The forecasts used on this study are:

- **Satellite based forecast (sat):** This forecast is based on the Heliosat 3 method (Hammer et al., 2003 and Hammer et al., 2015). This method uses the raw images from a satellite (Meteosat Second Generation or MSG) to generate a Cloud Index (CI) image. From the CI images Cloud Motion Vectors (CMVs) are calculated and used to extrapolate the CI into the future. This forecasted CI images are then used to calculate the irradiance value for every pixel. This is done by multiplying the forecasted CI values with irradiances derived from a clearsky model.
- **ASI network forecast (ASInet):** This forecast was developed on the DLR's Eye2Sky network (see section 2.2) and it is fully described in the deliverable D2.3.1. The ASIs on

this network are used to derive a segmented cloud mask, the cloud height and cloud speed. Then CMVs are calculated and used to extrapolate the cloud mask into the future. Finally, ground observations are overlaid on the cloud masks to derive irradiance maps. An intelligent combination of the high spatial density of ASIs (29 in 100k m² area) allows to extend the spatial and temporal coverage compared to a single (or pair) ASI system.

- **Satellite + ASI network combination (*sat+ASInet*):** In the Task 2.3.2 DLR has develop a method to combine these highly heterogenous forecasts inputs. The method is fully described in the Smart4RES Deliverable D2.3. In essence, the forecasts inputs are first homogenized in space and time. Then, historical forecasts (last 30 days) are used to optimize the coefficients on the linear combination of the forecast inputs. Here it is important to note that each forecast lead-time is optimized independently. Once the optimized coefficients are found, the actual forecasts (present) are combined using the optimized weights.
- **Satellite Persistence (*sat_persis*):** This forecast is also based on the Heliosat 3 Method. The difference is that the extrapolation is done using CMVs equal to 0. That is, as if the clouds do not move in time and only the sun does. This forecast is used as minimum reference value to be compared with the other forecasts inputs as all forecasts should at least improve the performance over persistence.

Table 3 summarizes the temporal characteristics of each of the forecasts used.

Name	Temporal resolution	Forecast horizon	Update frequency
Satellite persistence	15 min	6 h	15 min
Satellite	15 min	6h	15 min
ASI network	1 min	Up to 60 min	30 s
Satellite + ASI network combination	1min	30 min	1 min

Table 3: Forecasts descriptions.

2.4 Results and Discussion

2.4.1 Transposition model comparison

For this comparison we used one year of minute average data of the GHI, DHI, DNI and GTI (POA irradiance) measured at the Meteorological station LEEER. In this station, the POA radiometers are installed with an azimuth of 180° and a tilt of 30°. Figure 3 shows the monthly bias (left) and RMSE (right) of the estimated POA irradiance with respect to the measured POA irradiance. We see that all models are capable of describing the seasonal variation. All models show also a tendency to underestimate the POA irradiance. As expected, the **isotropic (Liu-Jordan)** model presents the highest bias and RMSE of all the models. The **Reindl** model (Anisotropic diffuse sky with circumsolar and horizontal brightening) shows the lowest bias and RMSE of all models for all months of the year. This confirms the big influence that the anisotropic nature of the diffuse component can have on the tilted plane irradiance.

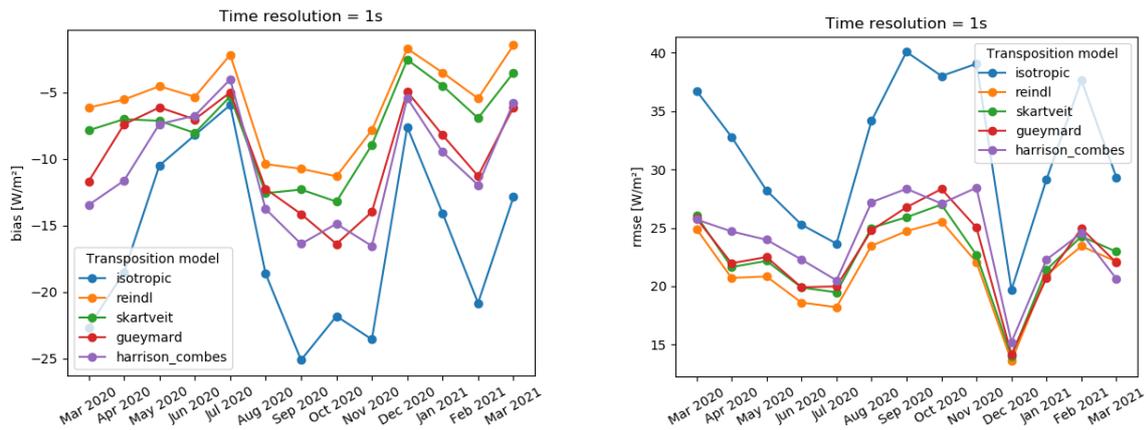


Figure 3: Monthly bias (left) and RMSE (right) of the POA irradiance estimated with various transposition models for the meteorological station LEEER. (1min average resolution data).

The accurate description of the circumsolar and the horizontal brightening diffuse components are thus critical for the improvement of the accuracy of the POA irradiance estimations. After analysing these results, the **Reindl** model was chosen to be used as the transposition function on our PV model.

2.4.2 PV model loss parameter optimization

Once the transposition model is chosen, we proceed to the optimization of the loss and bias coefficients **L** and **B** of the individual PV models (see section 2.3.2). As we do not have GHI measurements on the PV sites, the optimization is performed using the GHI measurement of the nearest available meteorological station in the Eye2Sky network. Due to this, only a clear sky condition will be used, so that there is no significant influence of the clouds between the 2 locations. The day 2020-08-07 was an almost perfect clear sky day and is used for this purpose. As only one month of data is available for the validation, this clear sky day is enough to characterize the loss parameters that try to correct for soiling, cabling losses, shadows, etc. on this time period. For the PV site OLD_EAST the station OLUOL is used (~2.5 km from the PV plant) and for the site OLD_CENTER the station OLDON is used (~1 km from PV plant).

The results for the OLD_EAST PV site are shown in Figure 4 (left).

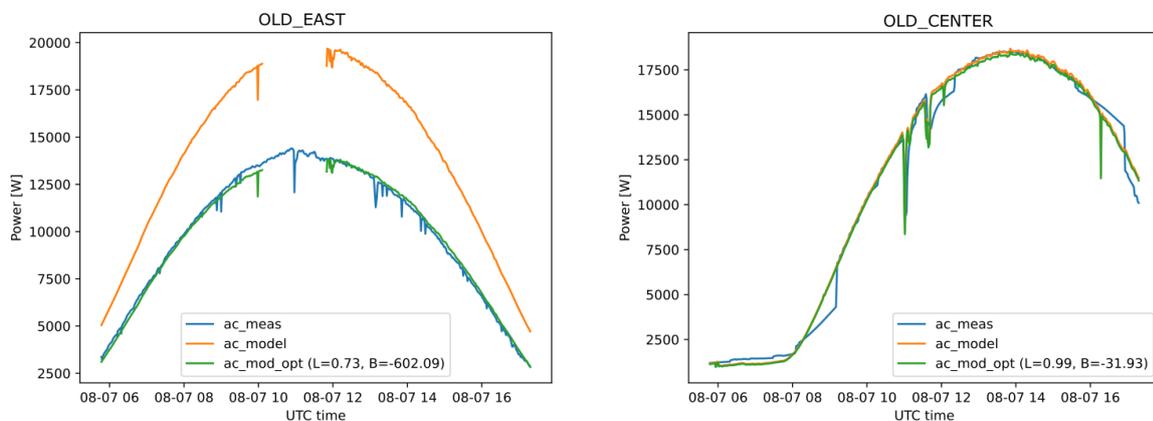


Figure 4: Optimization of the loss and bias parameters for the PV sites OLD_EAST (left) and OLD_CENTER (right).

On the OLUOL station, there are 2 hours of GHI measurements missing. Nevertheless, the data available is sufficient to characterise the losses thanks to the almost clear condition. We see that the AC modelled power has a loss factor of $L=0.73$ and a bias of $B=-602.09\text{ W}$, that is, a normalized bias of $nB=2.23\%$. The 27% of the power loss should be due to DC losses, Soiling, Shadowing, or other unknown condition. After applying the correction factors, we find a very good agreement between the AC powers. Before applying the correction, the $MAE = 4239\text{ W}$ ($nMAE = 15.7\%$) and after correction $MAE_{cor} = 212\text{ W}$ ($nMAE_{cor} = 0.78\%$). The small shift seen on the optimized power is due to the distance between the 2 sites.

The same was done on the PV site OLD_CENTRE and the results are shown in Figure 4 (right). In this case, we see that the model alone does a good characterization of the PV plant. Here a loss factor of $L=0.99$ and a bias of $B=-31.93\text{ W}$ ($nB=0.13\%$) are found. This PV model needs almost no optimization. Nevertheless, we see some strange behaviour of the PV power on this clear day. We see possibly some shadow effects (dip) in the morning hours as some reflection (jump) on the evening. We see also clearly the shift to the afternoon hours on the diurnal pattern that is due to the west facing PV configuration (azimuth=253°).

2.4.3 High resolution PV forecasting comparison

Before showing the PV forecast evaluation, we will evaluate the PV power analysis (forecast lead-time 0) for the different methods.

In Figure 5 we see the scatter density plots between the AC power generated by the system OLD_EAST and the ones generated with the 4 different forecast sources described in section 2.3.3.

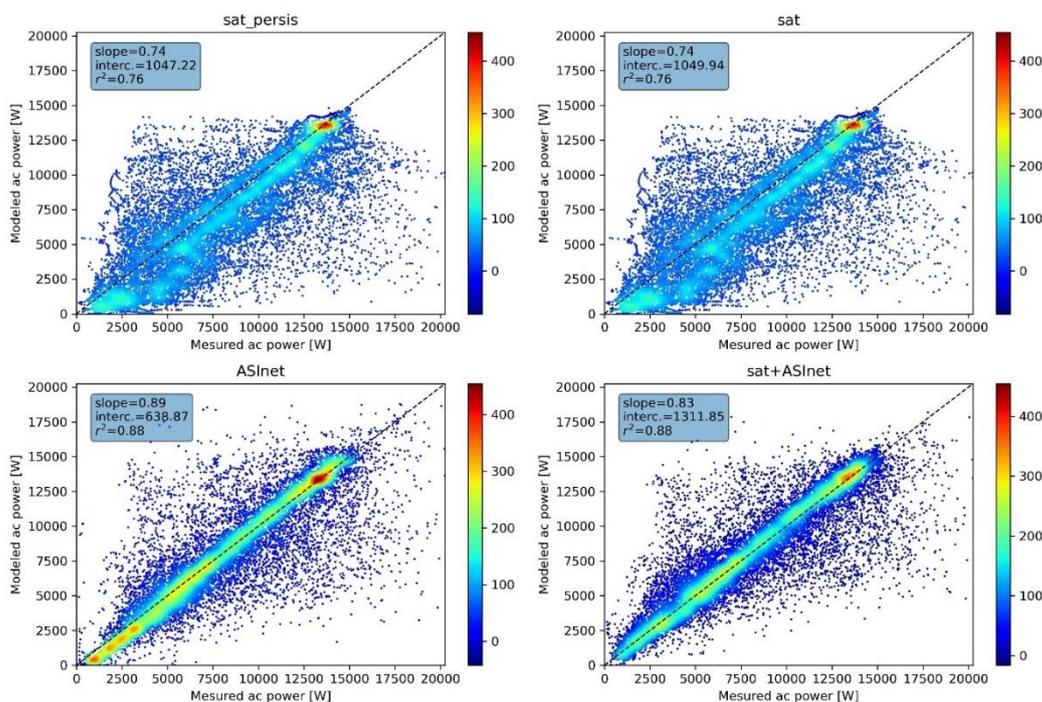


Figure 5: Correlation of measured and modelled AC power for the analysis case (forecast lead-time 0) of all forecast sources on the site OLD_EAST (1 min resolved data for the month of 08.2020).

We see that in general the highest density of the points falls into the high irradiance/power situations. Due to the low spatial and temporal resolution, the satellite-based methods have problems to estimate the intermediate power values / mixed sky conditions. By the contrary,

the ASI method performs much better on these intermediate conditions. As noted by our Partner EMSYS in chapter 3, the measurements of the ASI network correlate well at the up - and downward ramps with the measured power from the PV plant. This is evidenced here by the slope and r^2 values (0.89 and 0.88) which are closer to 1 which is not the case on the satellite-based methods. It is important to note that the PV output at lead-time 0 (known as analysis) for the ASI and the Sat + ASI methods are very similar (also similar slope and r^2 coefficients). This shows that for the analysis (lead-time 0) the combination method learned that the weights for the ASI data are much higher than the ones for the Satellite. There is a clear superiority of the ASI and Satellite + ASI methods over the Satellite only based analysis measurements

Figure 6 shows the same analysis measurements for the PV site OLD_CENTER. Here we see a very different behaviour compared to the PV site OLD_EAST, which is only at about a 3 km distance. In the OLD_CENTER site the higher density of points is found for all sources on the lower power values (1kW to 2kW). This response seems to be linked to the fact that our data is restricted to values with solar elevation $> 20^\circ$. As this is a west facing system with an elevated tilt, the filtering of these data clips the power production on an instant of the day where the generation capacity is still not negligible (see Figure 4). So, our 20° filter artificially moves the distribution peaks to the low power values. At the end, the dataset is still valid, but just has less available data.

Having in mind the apparent distribution shift to low power values, we find in Figure 6 a similar statistical response of the forecast models as for the OLD_EAST site. The best correlation for this analysis measurements (lead-time 0) is found on the ASI model (slope=1.04, intercept=8.4 W, $r^2=0.9$). The higher probability of low power values on the dataset also emphasises the fact that the ASI based method copes very well with lower values of irradiance that can be due to variable cloud conditions. This is seen with the better statistical correlations of the ASI only with respect to the ASI + Sat method (difference that is not seen in the OLD_EAST).

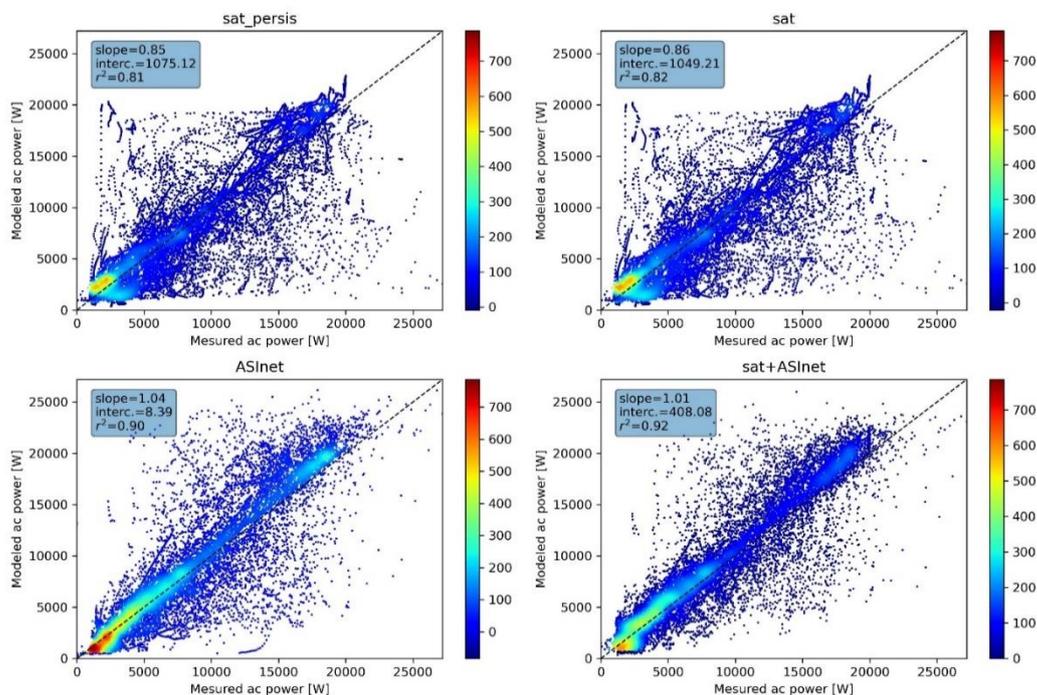


Figure 6: Correlation for measured and modelled AC power for the analysis (forecast lead-time 0) of all forecast sources on the site OLD_CENTER (1 min resolved data for the month of 08.2020).

Now we continue with the analysis of the PV forecasting performances at higher lead-times.



In Figure 7 we find the nMBE, nMAE and nRMSE of the forecasted power estimated with the GHI values from the 4 different forecasting methods. This error scores are normalized to the installed capacity. First, we see that as expected, all forecast methods outperform the nMAE and nRMSE metrics of the satellite persistence method. The Satellite nRMSE and nMAE show a strange U-like behavior for the values between lead-times 0 – 15 and 15 – 30. This comes from the fact that the satellite forecast is only available every 15 minutes (0, 15, and 30 minutes in this case). All other values for this method are interpolations. The interpolation has an averaging effect, that translates in a slight improvement of the error metrics on the interpolated values.

Also as expected, the nRMSE of the ASI forecast outperforms the one of the satellite forecast on the lower lead-times. This is due to the higher resolution data, which translates to a higher quality local information on the cloud situation. As lead-times increase, the clouds move over the field of view of the camera network without other clouds coming into the domain (no information outside the network domain). After 16 min, the nRMSE of the satellite outperforms the ASI. This means that statistically, the ASI network forecast has a noticeable advantage over the satellite until the lead-time 16 where the cloud information is lost or is too degraded to give accurate forecasts.

More interesting is now the combined ASI + satellite forecast. This method shows a bias near 0 for all lead-times. It also outperforms all the other forecasts on nMAE and nRMSE for all lead-times. The improvement on nRMSE on the worst case (lead-time 0) goes from 7.04% (ASI) to 6.73% (ASI + sat) which corresponds to an improvement of 4.4%. In the best case (lead-time 30), the improvement on nRMSE goes from 11.4% to 9.04% which represents an improvement of 14.86%.

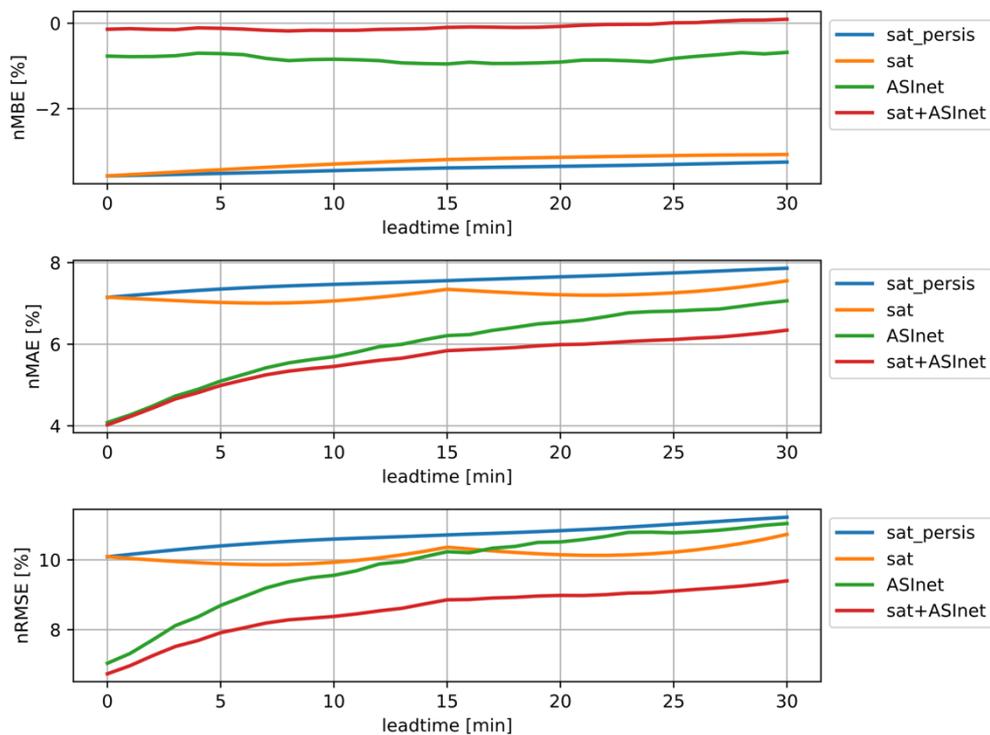


Figure 7: PV forecast performance for the site OLD_EAST. All data used has a temporal resolution of 1min. The Combined ASI + SAT forecast was trained on 07.2020 using 4 Eye2sky stations: OLCLO, OLDON, OLUOL, PVAMM. nMBE, nMAE and nRMSE are normalized to the installed capacity.

Table 4 shows the summary of the nRMSE for the PV forecasting methods based on SAT, ASI and ASI + SAT along with the improvements of ASI + SAT over the other 2 methods.



Lead-time [min]	Satellite [%]	ASI [%]	Satellite+ ASI [%]	Improvement of Sat + ASI over SAT [%]	Improvement Sat + ASI over ASI [%]
0	10.09	7.04	6.73	33.30	4.40
5	9.89	8.70	7.91	20.02	9.08
10	9.93	9.56	8.38	15.61	12.34
15	10.36	10.23	8.86	14.48	13.39
30	10.73	11.04	9.40	12.40	14.86

Table 4: nRMSE of PV power forecast method for selected lead-times on the OLD_EAST site. Improvement of nRMSE of Satellite + ASI over only SAT and only ASI.

Here we demonstrate the potential improvement that the ASI forecast brings to the Satellite forecast when they are combined together. The improvement on RMSE from 5 – 30 min ahead lies between 9.08% and 14.86%. This outperforms the improvements for solar power production RMSE of 9% - 12 % defined on the KPI 1.2.a in the deliverable D1.1.

The same analysis was also performed to the site OLD_CENTER (Figure 8). We first note that this system shows a lower power prediction performance than that of the site OLD_EAST. This could be due to the clipping of the power values on times where still a non-negligible part of the power is being generated (as discussed previously). Nevertheless, this plant also shows the same trends on forecast improvements (Table 5). The improvement on RMSE from 5 – 30 min ahead lies between 15.66 % and 17.78 %. This site also outperforms the improvements for solar power production RMSE of 9% - 12 % defined in the KPI 1.2.a on the deliverable D1.1.

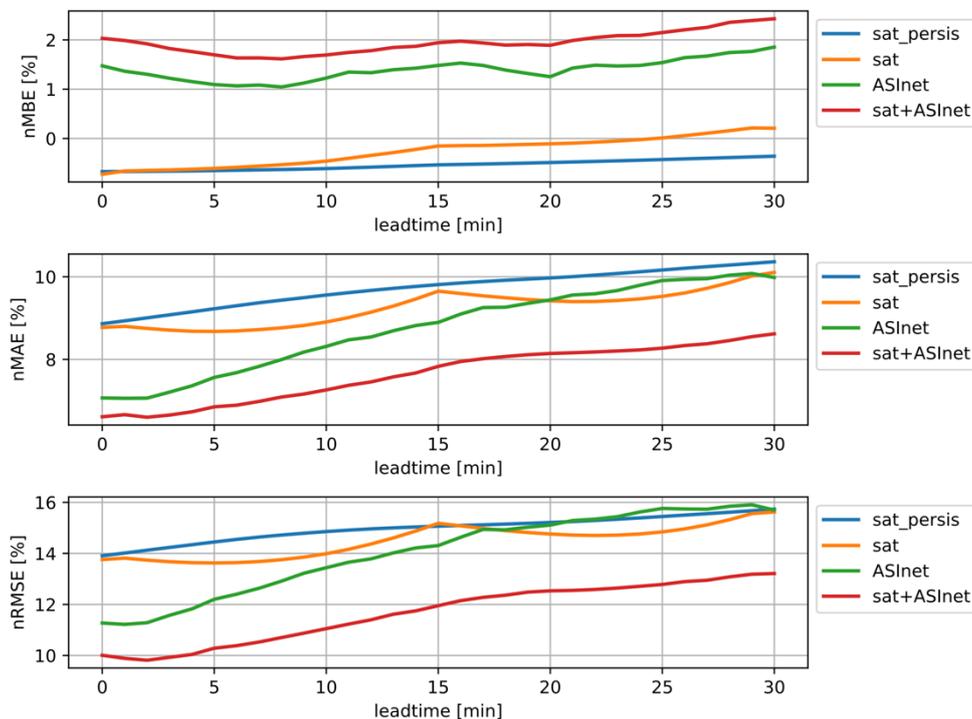


Figure 8: PV forecast performance for the site OLD_CENTER. All data used has a temporal resolution of 1min. The Combined ASI + SAT forecast was trained on 07.2020 using 4 Eye2sky stations: OLCLO, OLDON, OLUOL, PVAMM. nMBE, nMAE and nRMSE are normalized to the installed capacity.



Lead-time [min]	Satellite [%]	ASI [%]	Satellite+ ASI [%]	Improvement of Sat + ASI over SAT [%]	Improvement Sat + ASI over ASI [%]
0	13.76	11.27	10.01	27.25	11.18
5	13.62	12.20	10.29	24.45	15.66
10	13.99	13.44	11.05	21.02	17.78
15	15.18	14.31	11.95	21.28	16.49
30	15.62	15.70	13.21	15.43	15.86

Table 5: nRMSE per forecast methods for selected lead-times on the OLD_CENTER site. Improvement of nRMSE of Satellite + ASI over only SAT and only ASI.

The combination of the ASI forecast with the Satellite forecast clearly brings statistical improvements on the power prediction error. The optimization of the combined forecast was based on 4 reference irradiance points (OLCLO, OLDON, OLUOL, PVAMM). Having denser reference irradiance measurements on the region could still bring more improvements on the error as sites far away from this point will be penalized by the lower quality of the information. Also, the use of an extended ASI grid will allow to extend the forecast horizon of the combined forecast further than the 30 minutes shown here. To further validate this method, ASI forecasts on other months of the year should be processed and combined with satellite in order to assess the seasonal transferability of the results. This was not done here due to unavailability of the data.

References

- Agoua, X. G., Girard R., and Kariniotakis, G., (2018), "Short-Term Spatio-Temporal Forecasting of Photovoltaic Power Production," *IEEE Trans. Sustain. Energy*, vol. 9, no. 2, pp. 538–546, doi: 10.1109/TSTE.2017.2747765
- Gracia, A., and Huld, T. (2013), "Performance comparison of different models for the estimation of global irradiance on inclined surfaces", *EUR--Scientific and Technical Research series*, doi: 10.2790/91554
- Hammer, A. et al. (2003), "Solar energy assessment using remote sensing technologies", *Remote Sensing of Environment* 86, pp. 423–432, doi: 10.1016/S0034-4257(03)00083-X
- Hammer, A., Kühnert, J., Weinreich, K. and Lorenz, E. (2015a), "Short-Term Forecasting of Surface Solar Irradiance Based on Meteosat-SEVIRI Data Using a Nighttime Cloud Index", *Remote Sens.*, 7, doi: 10.3390/rs70709070
- Harrison, A., and Coombes, C. (1988), "Comparison of model and indirectly measured diffuse sky irradiances of tilted surfaces", *Atmosphere-Ocean*, 26(2), 193-202, doi: 10.1080/07055900.1988.9649299
- Holmgren, W., Hansen, C. and Mikofski, M., (2018), "pvlib python: a python package for modeling solar energy systems", *Journal of Open Source Software*, 3(29), 884, doi: 10.21105/joss.00884
- King, D., et al. (2004), "Sandia Photovoltaic Array Performance Model", *SAND Report 3535*, Sandia National Laboratories, Albuquerque, NM
- King, D., Gonzalez, S., Galbraith, G. and Boyson, W. (2007), "Performance Model for Grid-Connected Photovoltaic Inverters", *SAND Report 2007-5036*, Sandia National Laboratories





Perez, R., Ineichen, P., Maxwell, E., Seals, R. and Zelenka, A. (1992), "Dynamic Global-to-Direct Irradiance Conversion Models", *ASHRAE Transactions-Research Series*, pp. 354-369, <https://archive-ouverte.unige.ch/unige:38583>

Reda, I. and Andreas, A. (2004), "Solar position algorithm for solar radiation applications", *Solar Energy*, vol. 76, no. 5, pp. 577-589, doi: 10.1016/j.solener.2003.12.003





3 PV power forecasting with Skymager data (EMSYS)

3.1 Motivation

Energy & meteo systems operates PV power forecasts based on irradiance forecasts from NWP models and uses power measurements from PV plants to improve forecasts in the short-term range of a few hours. The NWP forecasts are updated only every 1-12 hours and have a latency of 1-4 hours due to processing time. Furthermore, they have a coarser spatial resolution of 2-40 km. The PV measurements are available with a latency of only a few minutes, but they are limited to the past and not available for all PV plants.

Therefore, forecast data with high temporal and spatial resolution is needed to better forecast the short-term range. A promising technique is to use the data from the Skymager network Eye2Sky established by DLR in the Northwest of Germany as described in section 2.2 and also in Smart4RES Deliverable 2.3. The Skymager network consists of several distributed sky imaging cameras together with several other measurement instruments (e.g. radiometer, ceilometer) in order to detect clouds and their movement with a high temporal and spatial resolution. The cloud movement is then extrapolated to forecast the up-coming 0-30 minutes. The output of the Skymager method are irradiance forecasts which may be used as an additional data source to NWP forecasts and PV measurements to improve the short-term PV power forecasts on this time scale.

In the presented case study, the forecasts from the DLR Skymager network are used as an additional data source to the state-of-the-art forecast model Suncast of energy & meteo systems based on NWP forecasts and PV measurements.

3.2 Case studies

In this case study, irradiance maps from the DLR's Eye2Sky network in Northwest Germany have been used and evaluated against measurements of the PV plant located in the western part of Oldenburg.

The Eye2Sky network consisted of 27 operational from 39 planned measurement stations at the time of the case study. These stations are equipped with All Sky Imagers (ASIs) and some with additional instruments like radiometers or ceilometers. The images of the ASIs are used to derive cloud masks, cloud heights and cloud speeds. Then, cloud motion vectors are calculated to extrapolate the cloud mask into the future. Finally, the extrapolated cloud masks are transferred into irradiance maps using radiative transfer equations and additionally radiometric measurements.

These irradiance maps have been reprocessed by DLR for 3 periods in 2020 (March, June and November). The periods were chosen because they mainly contain days of broken clouds for which we expect the largest improvements in the short-term forecast. In such situations, the Skymager method can distinguish clouds against the clear sky and extrapolate their movement to the near future. The irradiance maps consisted of real-time data retrieved from the measurements and forecast data with 20 minutes horizon (hereafter called nowcasts). The location of irradiance maps and the details of the data set are described in Table 6.

Spatial and temporal details of the irradiance maps		
spatial resolution	800 x 800 pixel (40 x 40 km)	
time periods	01.03.-15.03.2020 01.06.-30.06.2020 18.11.-29.11.2020	
data types	real-time	nowcast
temporal resolution	30 sec	60 sec
forecast horizon	-	up to 20 minutes

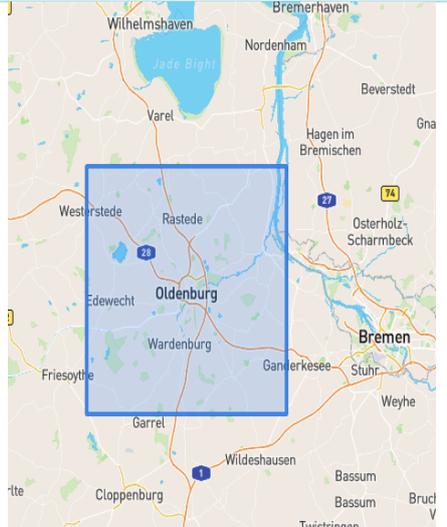


Table 6: Details of the DLR's Eye2Sky data set and their location in Northwest Germany.

3.3 Approach

The DLR Skymager real-time and nowcast irradiance data have been further processed and analysed by energy & meteo systems.

The first step was to extract the data from the irradiance maps for the location of the PV plant. Then the irradiances are converted to power values by using the real-time irradiances and historic PV power measurements for calibration. Figure 9 shows the scatter (left) of the power derived from the Skymager irradiance against the measured power from the PV plant. The majority of data points are close to the diagonal with some outliers for the different time periods indicated by the different colours. A closer look on the timeseries for a single day (right) shows that the power derived from the real-time irradiances (red) correlates well at most of the up - and downward ramps with the measured power from the PV plant (black).

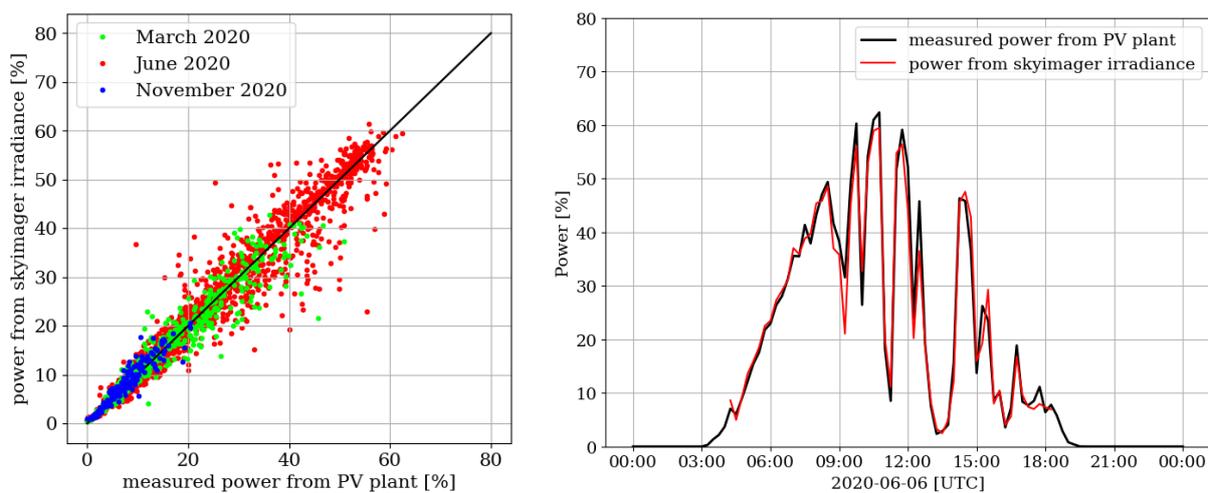


Figure 9: Scatter plot of power derived from Skymager real-time irradiances against power from the PV plant for the three different time periods (left) and time series plot of those data for 2020-06-06 (right).

Figure 10 shows an example of an irradiance map (left side) with the location of the PV plant (red cross) and the time series (right side) of the real-time power from the irradiance map (red line) together with the power measurement (black line) from the PV plant (right side: the lower plot is a closeup of the upper plot). The lower right plot additionally shows the 0–15-minute power nowcast from the irradiance map (cyan line), which mostly predicts the PV power quite accurately, especially at the ramps. The overestimating of the real-time power at 12:30 to 12:35 and the power nowcast at 12:36 to 12:39 comes from the fact that the detected cloud is just at the edge of the PV plant and was slightly more southwards in the irradiance maps.

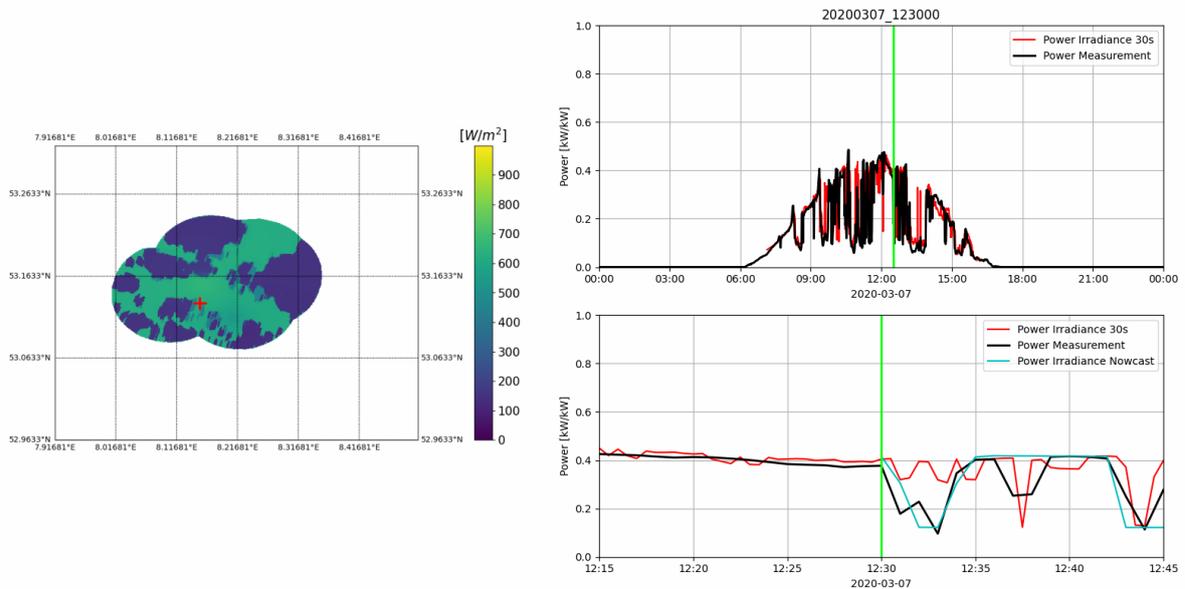


Figure 10: Example of a Skyimager irradiance map (left) with the location of the PV plant (red cross) and the timeseries (right) of the extracted power from the real-time irradiances (red) and the nowcast (cyan) together with the measurement from the PV plant (black). The lower right plot is a close-up of the upper right plot.

The second step was to integrate the Skyimager nowcasts into the state-of-the-art model of the forecasting system Suncast illustrated in Figure 11: it is based on a physical model and uses NWP forecasts (1) and PV measurements (2) as input data. In this case study, the Skyimager data (3) are integrated as additional data source.

For the evaluation three kinds of nowcasts have been processed:

- the Suncast power forecast based on NWP forecasts and PV measurement (1+2)
- the Skyimager nowcast based on the data from the irradiance maps only (3)
- the Skyimager+Suncast forecast using all three data sources (1+2+3)

The Suncast forecast has been processed as a reference in order to evaluate the improvement of the Skyimager nowcasts.

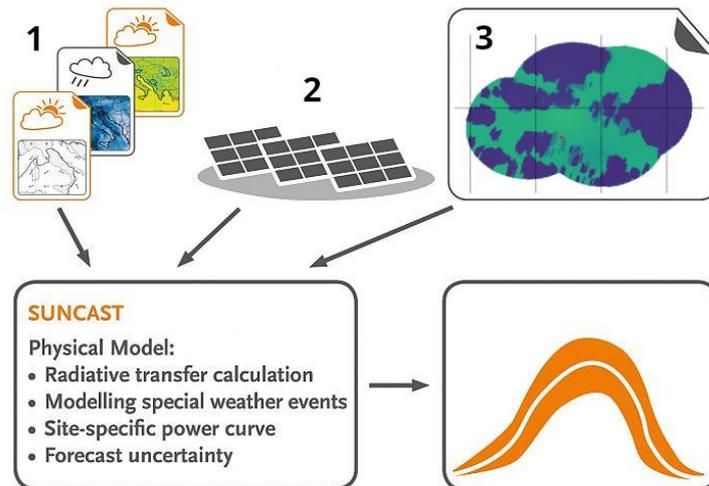


Figure 11: Scheme of the forecasting system Suncast and its components: NWP forecasts (1), measurements from PV plants (2) and Skyimager data as additional data source (3).

3.4 Evaluation

The Skyimager irradiance data have been analysed and evaluated against the measured power from the PV plant. The objective of the Skyimager nowcast is to improve the forecast of the upcoming 15-30 minutes compared to state-of-the-art nowcasts by 9-12% relative improvement of RMSE as defined for KPI 1.2 for solar energy. In this case study, the Suncast forecast is the reference to evaluate the improvement of the Skyimager and Skyimager+Suncast forecasts (described in Section 3.3). The forecasts have been processed for the prediction times every 15 minutes during daytime. The Skyimager nowcasts and the PV measurements have an original temporal resolution of 1 minute. They are aggregated to 15-minute resolution as this is the common resolution for operational evaluations. As the Skyimager nowcasts have a prediction horizon of maximum 20 minutes, only the first 15 minutes are evaluated here, resulting in one averaged prediction value for each prediction time. The Suncast forecast has already been processed on 15 minutes resolution.

Figure 12 shows an example of the nowcasts for 2020-06-06 (left: Suncast forecast, right: Skyimager+Suncast forecast). Each dot in the timeseries represents the forecast for the upcoming 15 minutes. For example, the dot at 12:00 is the forecast at the prediction time 12:00 for the upcoming 15 minutes ranging from 12:00 to 12:15. The Skyimager+Suncast forecast aligns with the up- and downward ramps of the measurement quite well. The Suncast forecast shows a temporal delay to the PV measurement, as it does not have the highly resolved information on moving clouds as the Skyimager forecasts.

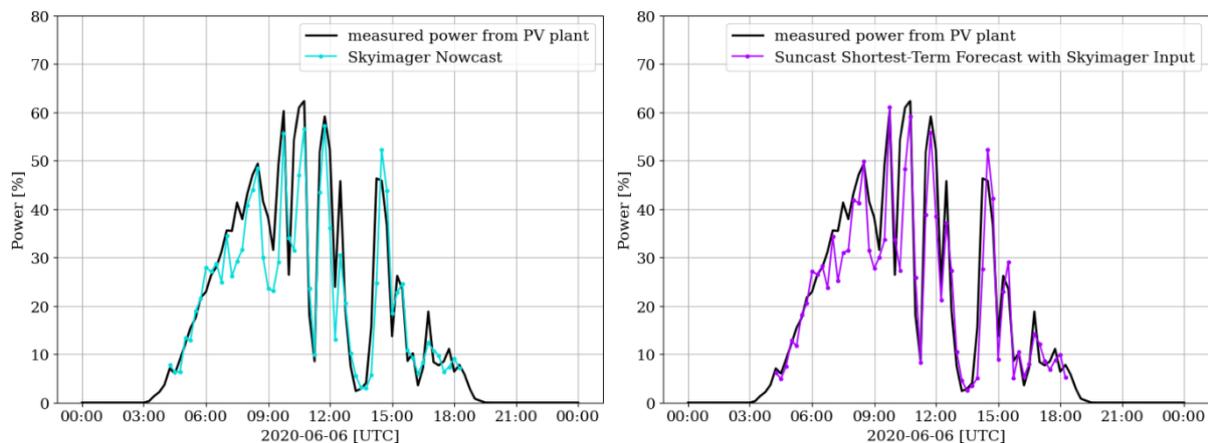


Figure 12: Timeseries of measured power from the PV plant (black) with Suncast forecast (left) and Skyimager+Suncast forecast (right) for 2020-06-06.

3.5 Results and Discussion

The quantitative evaluations for the three kinds of shortest-term forecasts (described in Section 3.3) are performed for the time periods in March, June and November 2020 (described in Section 3.4). The results are given in Figure 13 for each of the three time periods. They show the RMSE for each day (dots) together with the average over each time period (dashed lines).

The average RMSE of the Skyimager nowcast improved compared to the Suncast forecast from 5.7% to 4.7% in March, from 6.7% to 5.8% in June and from 1.9% to 1.7% in November. For the Skyimager+Suncast forecast the average RMSE decreased further to 4.4% in March, 5.3% in June and 1.6% in November. Thus, we achieved an improvement of the RMSE of 21-23% for the evaluated months. As the magnitude of improvement was similar for the three periods from different seasons, we can expect a similar improvement throughout the whole year. At single days with broken clouds, the daily RMSE of the Skyimager forecasts even reduced by almost one half compared to the Suncast forecast, e.g., on 2020-06-07 from 14% to 8%. Furthermore, there were almost no days for which the daily RMSE of the Skyimager forecasts significantly increased compared to the Suncast forecast.

In this case study we demonstrated the potential of the Skyimager data for PV power forecasting in the short-term range. The results show a very good improvement of the shortest-term predictions for the up-coming 15 minutes using Skyimager forecasts as additional data source. The RMSE reduced by 21-23% for the Skyimager forecasts compared to the state-of-the-art reference. This exceeds the objective of an improvement of 9-12% RMSE for 15-30 minutes defined in KPI 1.2. By expanding the network, the forecast horizons could even be extended. As the Skyimager method predicts the cloud movement based on highly resolved measurements, it enables to detect the up - and downward ramps more accurately. This is of special interest for huge PV plants connected to a sparse power grid, as single clouds could cause instabilities in the grid.

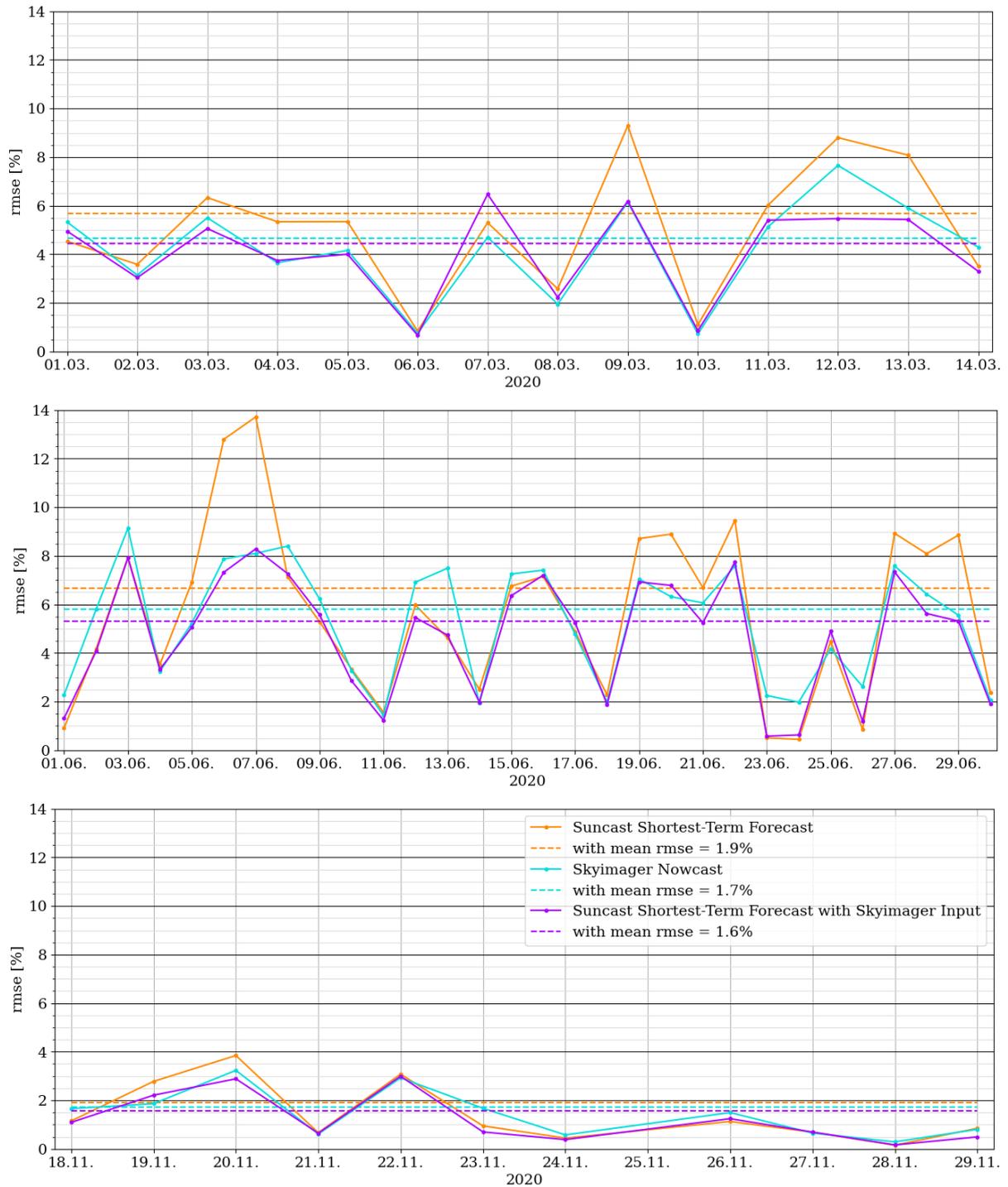


Figure 13: Daily RMSE of Suncast forecast (orange), Skyimager nowcast (cyan) and Skyimager+Suncast forecast (violet) for the three time periods in March (upper), June (middle) and November (lower). The dashed lines are the averaged RMSEs over each time period.



4 Improved thunderstorm correction for PV forecast (EMSYS)

4.1 Motivation

During thunderstorms dense and towering vertical clouds form. The shadows that are casted significantly reduce the performance of PV plants. Typically, there is a sudden reduction in power for individual PV plants. The production remains at a low level for a certain period of time. The sky then clears up and the power increases again.

The NWP models can only inadequately describe such processes due to the coarse grid size (10-50 km for global models) which is much coarser than the processes which lead to thunderstorm formation. In case of isolated thunderstorms, it is impossible to exactly predict when and where a thunderstorm might develop and which track will it follow. Another problem is the fast sideward spreading of the ice clouds at the top of the troposphere, that leads to the typical anvil shape of the thunderstorm. This is often underestimated in NWP models, especially for larger thunderstorms or clusters of thunderstorms. In any case, the sharp drop in performance for individual farms is not included in the power forecast from NWP. In case of thunderstorm clusters that span in larger regions, also the portfolio forecast can be affected by this problem.

With a shortest-term correction, it is possible to react to differences between the base forecast and live measured production values. The correction picks up on the live measured values and converges slowly back towards the base forecast. Such standard corrections are usually too slow to react to the sudden drop and increase in power in case of thunderstorms leading to large errors especially in the transition phases. With the additional information as to whether there is a thunderstorm or not, the correction can be designed differently depending on the situation.

4.2 Approach

A different form of short-term adjustment should take place during thunderstorms. Two criteria are defined to determine whether thunderstorm effects are present: Lightning must be measured within a 10 km radius of the solar farm and there must be dense clouds. The PV Clear Sky Index quantifies cloudiness. An approximation of this index is described as measured power divided by a clear-sky forecast. If the index is less than 0.3, dense clouds are formed. The thunderstorm is considered as ended when the PV Clear Sky Index rises above 0.3 again.

The standardized shortest-term correction ($pred_{stc}$) is based on the live measurement values ($meas$) and the base forecast ($pred_0$). The measured values of the last hour are averaged, so that the correction does not apply directly to the measured values (see blue curve in Figure 14). The difference in error between the mean measurement and the prediction decreases exponentially, so that at sunset the fit and the forecast converge.

In the case of thunderstorms, two different forecasts are combined. The first prediction is based on persistence. It is assumed that the cloudiness will remain the same at the time of calculation. The second prediction is the base prediction ($pred_0$). In this study, three combination approaches are compared to develop adaptations.

- $pred_{nc}$: From historic power data, the average thunderstorm duration has been found to be 60 minutes. At the time of the adjustment, it is calculated how long the thunderstorm has already had an impact on performance. For the following period of

time ($t = 60\text{min} - \text{previous duration}$) the persistence prognosis is assumed (see green line up to 10:45 in Figure 14). A new calculation of the standard shortest-term adjustment then takes place. The error difference decreases over a longer period of several hours.

- pred_{exp} : The first step follows the same principle as pred_{nc} (see green line in Figure 14). The error difference then decays within an hour and the shortest-term forecast converges to the base forecast.
- $\text{pred}_{\text{weight}}$: The persistence forecast and the base forecast are combined with each other. The weighting factors change with each time step. At the beginning, the persistence forecast is assumed to be 100%, at the end of the day it is 0% (see purple line in Figure 14). The factors correspond to the probability that a thunderstorm influence is still present.

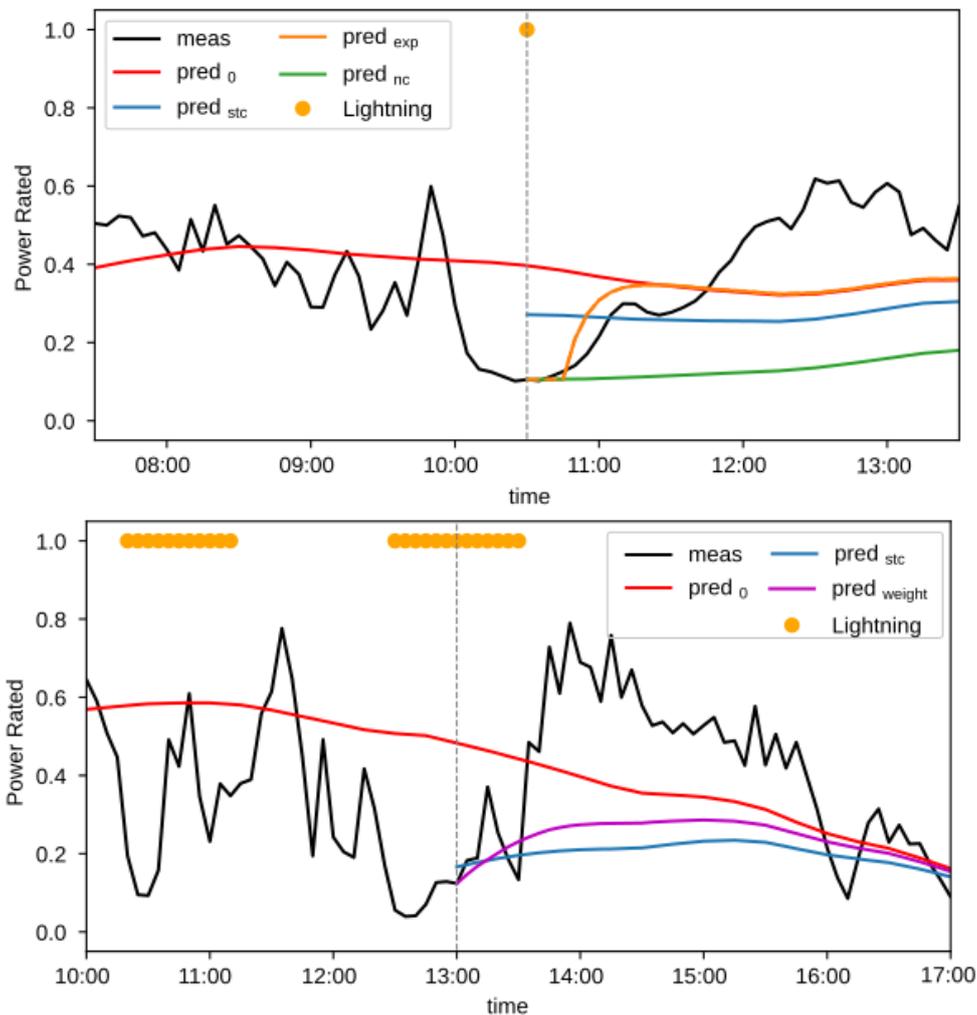


Figure 14: Base forecast (red line) and the different shortest-term corrections (coloured lines) during two days with thunderstorms. Measurements in black.

The main parameters for the adjustment are the temporal lengths of the thunderstorm influences. The average duration and the probability that the thunderstorm influence will subside after a certain period of time have been determined empirically.

4.3 Case studies

In the experimental studies, 500 solar farms in Germany were examined. The period under consideration is from April to August 2020. The base forecast is a meteorological combination forecast from various global and local models. The live measured values and the forecasts are normalized to the installed capacity.

The lightning data are from the service provider nowcast and are available in real time in a spatial grid (see Figure 15). The values in the grid points correspond to the number of detected flashes. However, only the information as to whether or not lightning was measured within a radius of 10 km is used. A logical time series with the values 0 and 1 is created for each solar farm.

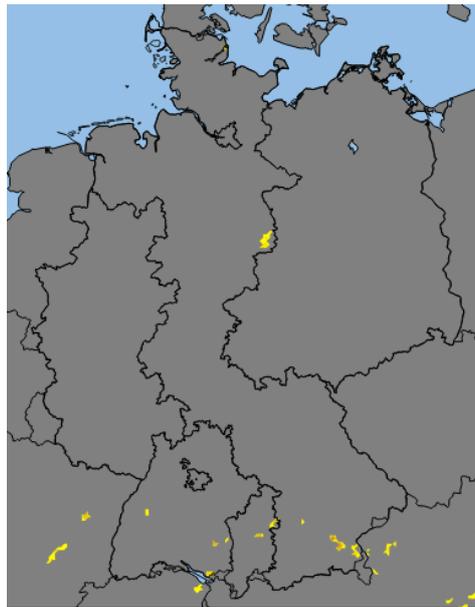


Figure 15: Example of the lightning data by nowcast used for thunderstorm detection.

4.4 Evaluation

The prediction quality is evaluated on farm level. Only calculations with thunderstorms during the day are taken into account. The reference time series are the live measured values. The considered error measurements are RMSE, SDE and Bias. Forecast horizons up to three hours are important.

For each calculation of the shortest-term adjustment, the pairs of forecasted and measured points are taken up to the forecast horizon. The mentioned error values are calculated. There is one error value per calculation and forecast horizon. This is averaged across all farms and presented as a function of the forecast horizon in Figure 16.

The error values are calculated for the base forecast and the various short-term adjustments. The comparison shows whether significant improvements were achieved through the situational adjustments.

4.5 Results and Discussion

The analysis of historical thunderstorms shows and quantifies their influence on the performance of solar farms. Detected thunderstorms according to the defined criteria have an impact of at least 10 minutes to several hours on the performance. The average impact duration is 60 minutes.

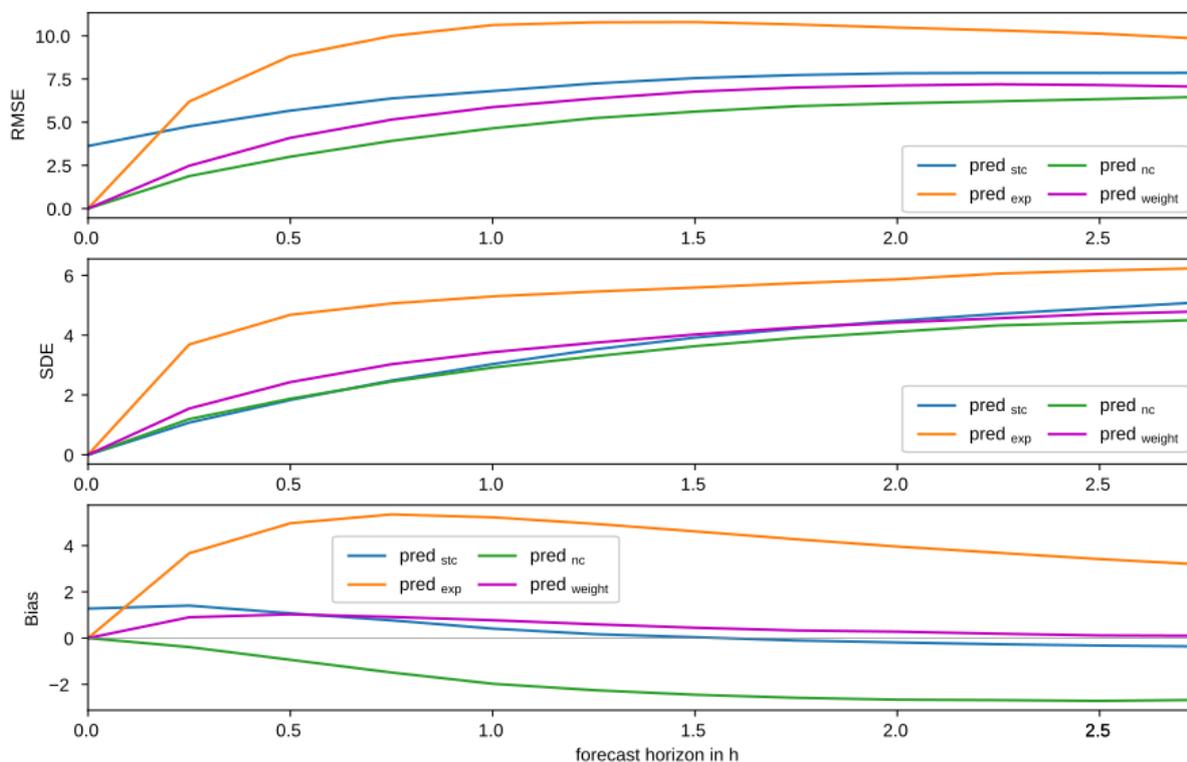


Figure 16: Mean values of RMSE (top), SDE (centre) and Bias (bottom) as a function of forecast horizon for different short-term corrections.

For comparison, the mean values of all calculated error values per adjustment type are shown depending on the forecast horizon (see Figure 16). All adjustments have significantly lower errors than the base forecast $pred_0$. The increased error value of the standard correction at the time of calculation is caused by the averaging of the measured values. For a forecast horizon of 15-30 minutes, this error is not relevant.

- The $pred_{exp}$ fit shows a high bias. Because of the fast exponential approach to $pred_0$, the bias at $pred_{exp}$ is positive. The variation in error, represented by the SDE measure, is more than 74% higher than the standard fit at a one-hour forecast horizon. The RMSE is also significantly higher. This can be explained by the variable length of the thunderstorm influence. The assumption of the mean length is mostly inappropriate.
- The adaptation $pred_{nc}$ also assumes an average thunderstorm length, but due to the slow approach to $pred_0$ there is no abrupt change after the assumed thunderstorm duration. This leads to a low variation of the individual errors as well as a lower RMSE (over 20% less than $pred_{stc}$). However, the bias is negative because the fit is consistently below $pred_0$. The power increase after a thunderstorm is therefore not shown.
- The weighting prediction $pred_{weight}$, on the other hand, is almost unbiased and has a significantly lower RMSE than the standard fit $pred_{stc}$. With a forecast horizon of one hour, the mean improvement is 14%, with two hours it is 10%. The improvement in the



SDE is 9% from a two-hour forecast horizon. The empirical distribution of the weighting factors is not focused on a mean value, but covers the broad spectrum of thunderstorm variations.

The approach of fitting with time-varying weighting factors shows promising results in this test case. This method can help to improve the short-term forecast for solar at the individual farm level. Especially the time horizon of up to 2.5 hours can be significantly improved.



5 Wind power forecasting based on improved power curve modelling (EMSYS)

5.1 Motivation

In EMSYS' short-term power prediction system, the wind power curve plays a crucial role as it relates wind speed forecasts provided by an NWP to the predicted power output of a wind farm. Therefore, accurate power curve modelling is crucial for precise predictions. However, in practice, models are often inaccurate to some degree for various reasons, e.g., due to noisy training data (Zhao et al. 2018).

We propose various improvements for wind data filtering. Furthermore, we demonstrate how flexible machine learning methods provide advanced models for wind power forecasting. In particular, we show that additional features can enhance the predictions and provide a more comprehensive perspective on the actual weather situation on site.

5.2 Outlier data filtering

Outlier detection is the core component of data filtering. With various improvements of an existing method for outlier detection, we create better training data. Our proposed improvements lead to a better detection of power outliers, especially power plateaus in the speed-power view originating from curtailments (see Figure 17). Based on these outliers, we identify latent power outliers in the temporal view (see Figure 19).

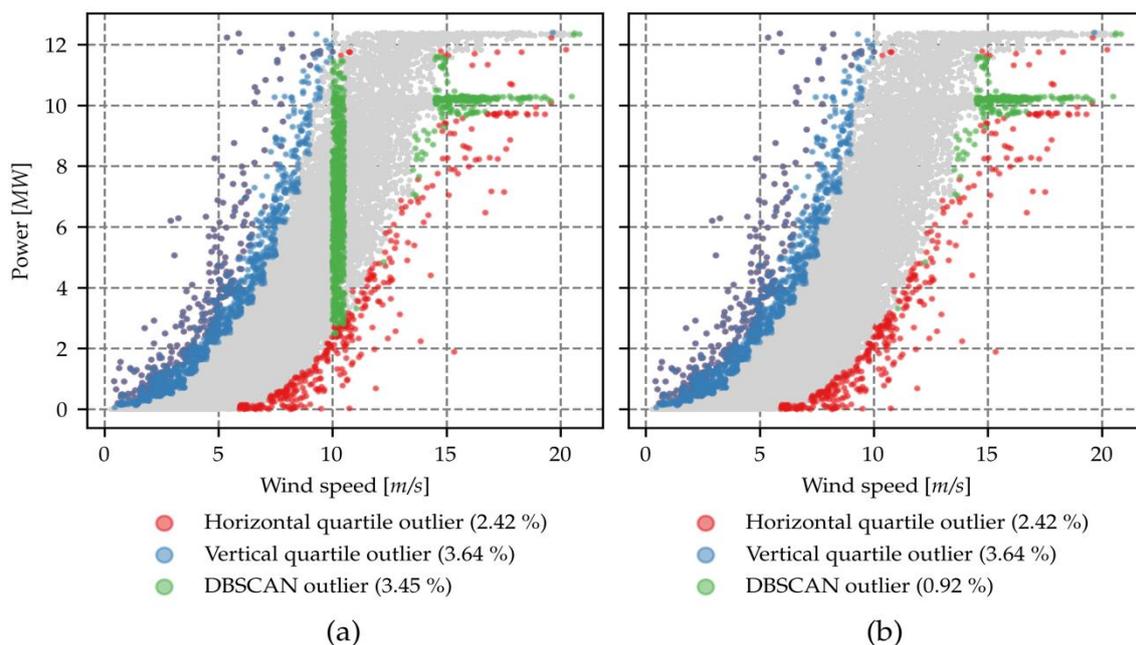


Figure 17: (a) The highlighted points were detected with the method of Zhao et al. (2018). The outlier plateau is identified. However, the vertical green bar in the middle consists of many false positives. (b) The result of our variant of this method.

One approach to remove power plateau outliers is presented by Zhao et al. (2018). We implemented the method of Zhao et al. (2018) according to the description in their paper. Looking at the results of their method for our datasets, we noticed the following shortcomings. On the one hand, the quartile method often seems to remove too many data points, which

may even result in a step pattern in the scatter. On the other hand, in some parts the method still does not remove enough data points, so that the removal using the DBSCAN-based filter in the second step does not work as desired. Sometimes a lot of good data is removed when there are gaps in the data scatter, as only the topmost cluster is kept. Both phenomena are clearly visible in Figure 17(a).

Motivated by the weaknesses of the discussed method, we propose approaches of improvement. Vertical cuts through the data are prevented by allowing outlier clusters to only have a maximum height. The maximum height is set dependent on the installed capacity of the wind farm. We keep all clusters that comprise at least 10% of the power range. The result of this improvement is displayed in Figure 17(b). Another shortcoming is that the assumption on which the above-described horizontal quartile method is based, namely that extreme values with respect to a bin's data median are outliers, is often too strong. Data points at the cut-out wind speeds should not be removed in many cases. If a filter method removed all data points in a wind speed bin, then we would rather keep all data points. This is another simple, yet very effective improvement.

In addition to that, the method does not remove the values that originate from unavailabilities, which are hidden in the dense scatter. We call these latent outliers. This becomes clear when looking at the heatmap in Figure 18. For the filtered data there is still a relatively large number of data points where the plateaus have merged into the dense scatter (approx. at [15m/s, 10 MW]). This indicates that there are power outliers in the data, but hidden between regular data points.

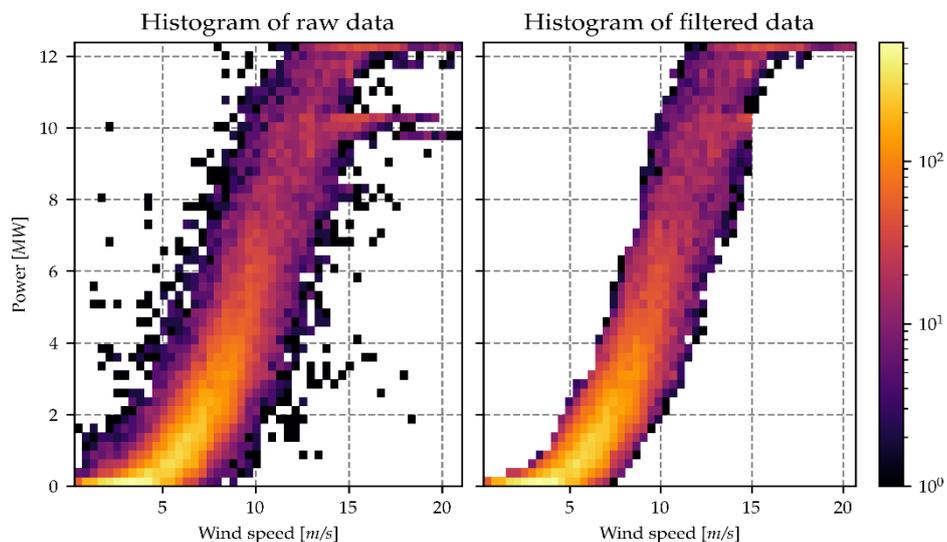


Figure 18: The left plot shows a heatmap of the raw data in the speed-power-view. The right plot shows a heatmap of the data after we applied the discussed outlier detection method, where the plateau was removed.

We use a sliding-window filtering approach to identify outliers which are not detected by the previously described method. The sliding-window filtering uses the temporal information of the data. Here, previously marked samples are treated as outlier candidates. Every data point which is temporally surrounded by a certain amount of outlier candidates is flagged as an outlier. Even if a value was flagged as an outlier (candidate) before, this function does not necessarily flag it as outlier as well, if the fraction of outliers in its surrounding is small enough, see Figure 19.

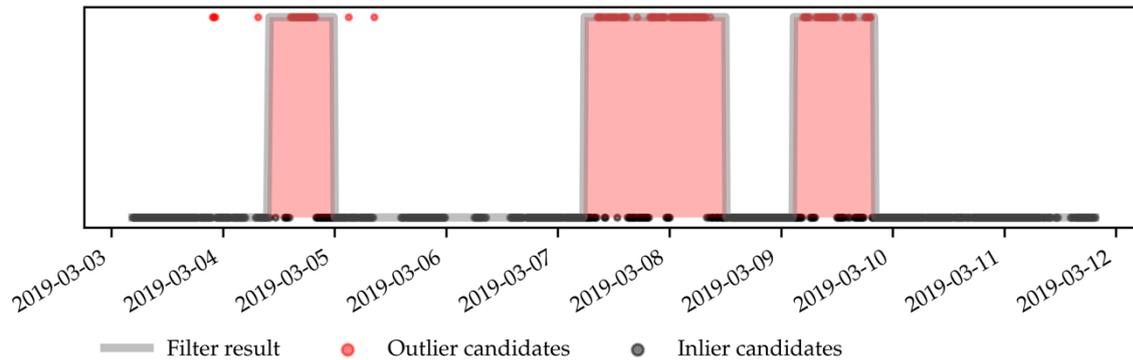


Figure 19: Example of sliding-window filtering based on outlier candidates. The black dots show valid data points in the time series view. The red dots show outlier candidates. The result of the sliding-window filter is indicated by the smoothed signal in grey.

5.3 Wind power forecasting based on improved power curve modelling

We show that the multivariate – using many input features – approach to wind power curve modelling with flexible state-of-the-art machine learning methods, like neural networks and gradient boosted trees (XGBoost), is advantageous.

For the comparison we use different models. The power curves produced by the reference model depend on wind speed and wind direction. For each wind direction sector, a piecewise linear function over the wind speeds is numerically optimized. The “method of bins” is a simple standard procedure for obtaining a power curve based on wind speed. For each wind speed interval, called bin, with a width of 0.5 m/s, the mean of the past observed power values within the interval is calculated and used as prediction. The *method of bins* is a widely used baseline approach.

Tree boosting is a popular and widely used machine learning method based on decision trees. XGBoost is a high-performance implementation of the ensemble model gradient boosted decision trees. It is a good approach for regression with tabular data. The XGBoost model and training algorithm have several hyperparameters that have to be tuned to get a good prediction model. The number of trees K , which is equivalent to the number of boosting iterations, is one important parameter affecting the model complexity. Other important hyperparameters are the maximum depth of the trees, the learning rate, the minimum child weight and minimum loss reduction¹.

Neural networks (NNs) are another popular and widely used machine learning method that is also suitable for this kind of regression problem. Neural networks are able to learn complex non-linear patterns from data. Here we use feed-forward neural networks implemented with Keras². Finding a good neural network architecture and good hyperparameters in general is a challenging task. As there are only around one dozen numerical values as input at most, based on which the power is predicted, it is likely that simple feed-forward architectures with few layers and few neurons in each layer are sufficient. Besides the architectural hyperparameters, for the training of the batch size, weight regularization, learning rate, activation function, optimizer and potential dropout are tuned.

¹ More details on <https://xgboost.readthedocs.io>

² Retrieved from <https://keras.io>



5.4 Evaluation

The outlier data filtering approaches are mainly evaluated qualitatively. Experienced meteorologists, who work with this kind of data on a daily basis can judge the results of outlier data filtering in the speed-power-view. For evaluation of power curve models, we use samples from one year for training and samples from the following year for evaluation. We used a specific variant of cross validation for hyperparameter search. We work with time series data and in order to guarantee that a model is not validated on samples that are older than the training samples, we used an n-fold times series split. This means that the sorted time series is divided into $n + 1$ equally sized parts and each part, except for the first, is used once for validation. The default metric is the root-mean-square error (RMSE). In practice, however, we typically provide additional error scores and perform visual inspections of the prediction versus the measurement time series. Another aspect of evaluation is to investigate the importance of different features, which is also discussed in the next section.

5.5 Results and Discussion

Figure 20 shows a comparison of different methods to create power curve models. For the machine learning methods that can deal with an arbitrary number of input features several models with different feature sets were trained. On average the XGBoost model was the best model here.

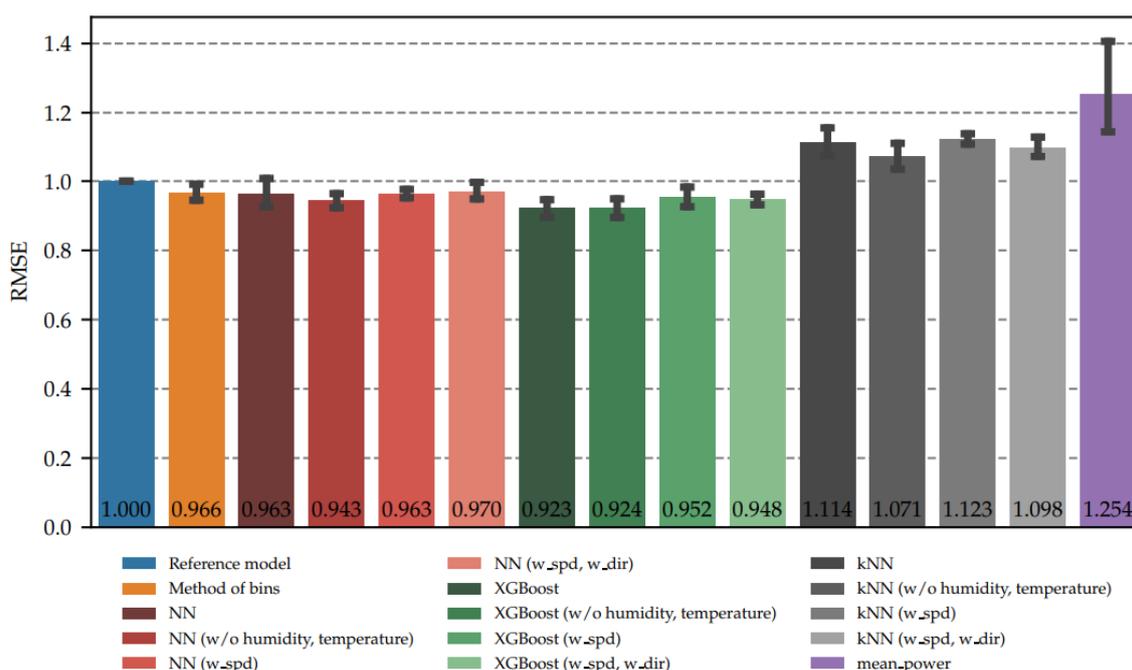


Figure 20: This plot shows the mean RMSE for various models for six datasets from different wind farms. For the three machine learning methods several models with different feature sets were trained. If no features are given in brackets, all available features were used. Each model was trained on the data from one year and evaluated on the data from the following year. The RMSE values were scaled for each dataset, such that the reference model has an RMSE of 1.0. The error bars show the bootstrap confidence intervals with 95% confidence level.

Looking at wind speed and wind direction only, which are the features that the reference model uses, neural networks (NNs) improved the predictions on average by three percent RMSE and XGBoost improved the predictions on average by more than five percent RMSE, compared to the reference model. The XGBoost model with additional features had an almost

eight percent lower RMSE than the reference model. This is a great improvement. For the NN, the results with additional features were not consistently better. The NN with all available features was worse than the one without humidity and temperature, but with all other features. This indicates an overfitting of the NN with all available features. The (k Nearest Neighbour) kNN model was always worse than the reference model. However, we also did not expect that the kNN model produces good results, since it was hardly tuned at all and not likely to generalize. The purple bar depicts the result for the constant prediction of the average power of the training dataset without using any features at all.

Figure 21 gives an overview of the importance of various features for multiple wind farms. It shows the total gain, an XGBoost-specific measure of feature importance, for several datasets. As expected, the wind speed is always the most important feature. However, as the average importance of the wind speed is around 80%, this means that splits on the other features caused 20% of the total gain. On average, the second most important feature was the wind direction. Azimuth and elevation follow as third and fourth most important features with similar importance scores. One possible explanation for their importance is that they indirectly add a temporal component to the model, which helps to characterize the actual weather situation on site.

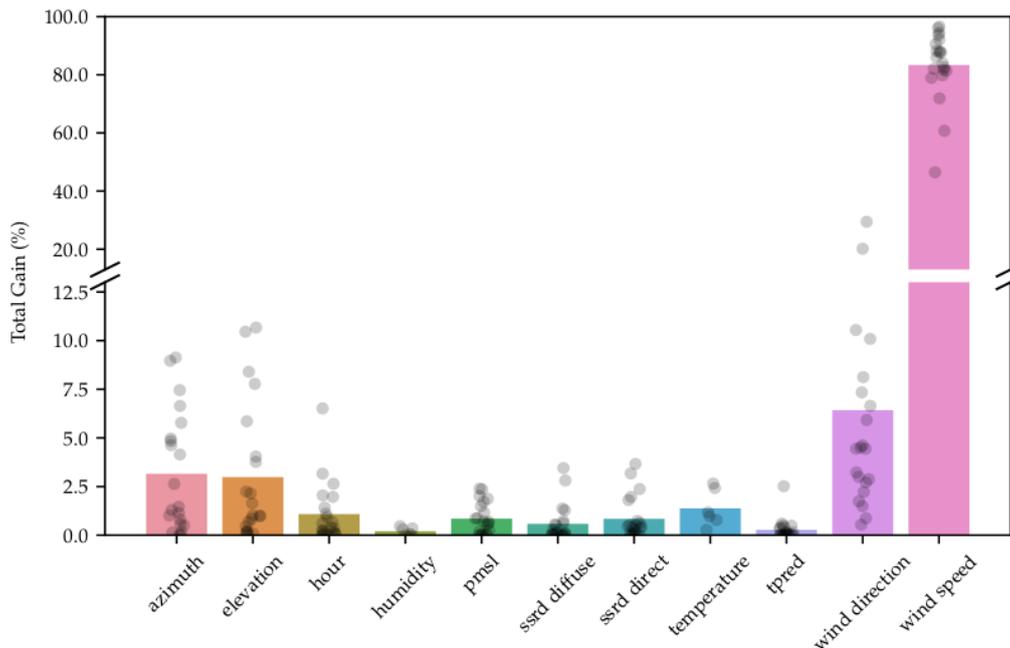


Figure 21: This figure shows the total gain (to be interpreted as “importance”) for various features for multiple datasets. The bar plots show the average importance values for the respective feature. The scatter plots (with jitter for visualization purposes) show the raw importance values. The y-axes are broken to get detailed insights into the importance scores.

Overall, it is important to note that the variability of the importance values is high for some features. Features can be correlated, resulting in the model to learn e.g., only from one of several correlated features, which decreases the apparent importance of the correlated features. This means that there is no absolute importance of the different features, but that these depend on the wind farm, weather model and the model's choice. We always have to understand these importance scores as what was important to the model. The fact that the importance for humidity had a low value here means that it was not important for these models, but not that it is generally unimportant for all possible power curve models. Nevertheless, it is relevant to know the importance scores for a specific model to explain its decisions and gain insights for improvement.



We demonstrate how flexible machine learning methods, like neural networks and especially gradient boosted trees, provide advanced models for wind power forecasting. In particular, we improve power curves through outlier data filtering and additional features.

References

Zhao, Y., Ye, L., Wang, W., Sun, H. and Tang, Y. (2018), "Data-Driven Correction Approach to Refine Power Curve of Wind Farm Under Wind Curtailment", IEEE Trans. Sustain. Energy, vol. 9, no. 1, pp. 95.-105, 2018, doi: 10.1109/TSTE.2017.2717021





6 LiDAR based second-ahead forecasting of wind power (DTU)

6.1 Motivation

Nacelle-mounted lidars have the capability of providing a preview of the wind field in front of the rotor. Such a preview is beneficial when controlling the wind turbine to improve power performance and reduce structural loads. Here in this task, we aim to implement and test novel forecasting methods using lidar observations which adhere to the constraints of real-time usage. As more and more lidars are installed on the turbines globally, the investigation also reflects on the potential benefits and drawbacks of a real-world fulfilment of such a system.

The potential benefits of lidar-based forecasting as analysed here are two-fold: 1) Power forecasting, which can enable increase in power production through an improved representation of the real-time inflow via detection and mitigation of under-performance as well as flow (or wake) control for optimal energy extraction from a wind farm; 2) (Structural) Load forecasting, which can be utilised to reduce the structural (both fatigue and extreme) loads to increase turbine lifetime and reduce costs/LCOE.

6.2 Overall Approach

The study performed in this task investigates the potential of using lidars and machine learning for seconds-ahead forecast of the electrical power and blade root flap-wise bending moment on an 850kW Vestas V52 wind turbine. For that, both a 2-beam and a 4-beam lidar units are analysed and the added value of the latter instrumentation is quantified. It should be noted that the forecasts take lidar measurements as inputs exclusively as the operational condition of the turbine ahead of time is unknown.

6.3 Case study for model benchmark

6.3.1 DTU Risø Test Site

The experimental data used in this analysis have been conducted in the period between 01/08-2020 and 01/12-2020. The test site is DTU Risø Campus which is located next to Roskilde Fjord, Denmark (UTM coordinates 55.68629°N, 12.09759°E). Figure 22 shows the geographical location of the test site and an overview of the test site layout. The wind turbine used in the measurement campaign is Vestas V52 with a rated power of 850 kW, rotor diameter of 52 m and a hub-height of 44 m. It is equipped with strain gauge systems to measure structural loads on the blades, rotor and tower. The calibration and validation of the load measurement system on V52 can be found in Poulsen et al. (2018) and Rinker et al. (2018). The load channel of interest for this study is the flap-wise bending moment and the power output used in this forecasting analysis is extracted from the SCADA system of the turbine. The meteorological mast (met-mast) located at the site is approximately 120 m (2.3 D) west-northwest of the wind turbine, which is also the prevailing wind direction at the site.

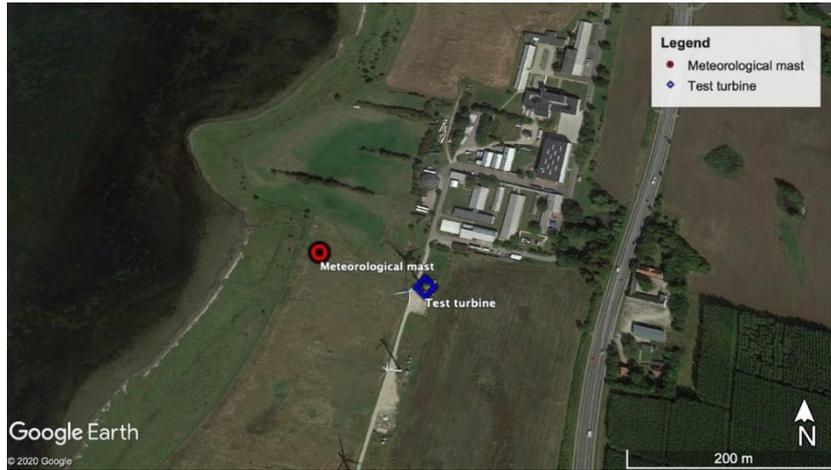


Figure 22: Risø test site, including the location of the Vestas V52 turbine and the met-mast.

The lidars measurements utilised in this study come from nacelle-mounted 2-beamed and a 4-beamed lidars, both manufactured by Windar Photonics A/S. A major advantage of nacelle-mounted lidars is the independence of wind direction as the lidar is fixed on the nacelle and therefore always aligns with yaw position. This gives a much higher availability compared to wind speeds measured with a meteorological mast where only a limited wind section provides coherent wind conditions. The 2-beamed lidar measures at a focus distance of 37 m which results in a measuring plane located 32 m (0.6 D) ahead of the wind turbine rotor. The 4-beamed lidar measures at a focus distance of 62 m resulting in a measuring plane located 59 m (1.1 D) ahead of the rotor. The sampling rate of both lidar systems is 1Hz. The lidars measure only the radial wind speeds along each beam, also called the line-of-sight (LOS) wind speed. However, in many applications it is more useful to express the rotor averaged (also referred as rotor effective) wind speed $\sqrt{U_{eff}U_{eff}}$, which can be estimated as the horizontal wind speed component averaged across the number of beams as:

$$U_{eff} = \frac{1}{B} \sum_{i=1}^B \frac{V_{LOS,i}}{\cos \alpha_i}$$

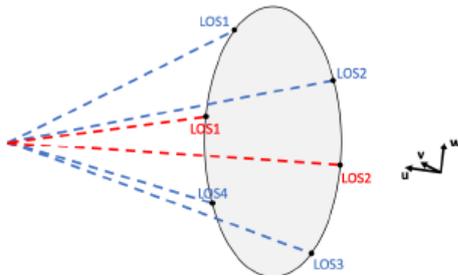


Figure 23: Configurations of the LiDAR beams. Blue is the 4-beamed LiDAR and red is the 2-beamed LiDAR.

where B is the number of beams, α is the horizontal/axial projection angle of the beam i and $V_{LOS,i}$ refers to the LOS wind speed for beam i . Configurations of the lidars beams for the installed setup is illustrated in Figure 23.

The time it takes for the lidar measured plane in front of the nacelle to reach the turbine is considered as the largest forecast horizon in this analysis. Also referred as time delay, it reaches approximately to 40 s for the met-mast (valid only for the aligned wind directions, measuring at 120 m in front), 20 s for the 4-beamed lidar (measuring at 59 m in front) and 10 s (measuring at 32 m in front) for the lowest wind speeds. An example time series of the considered channels from the SCADA, strain gauge and lidar systems is presented in Figure 24.

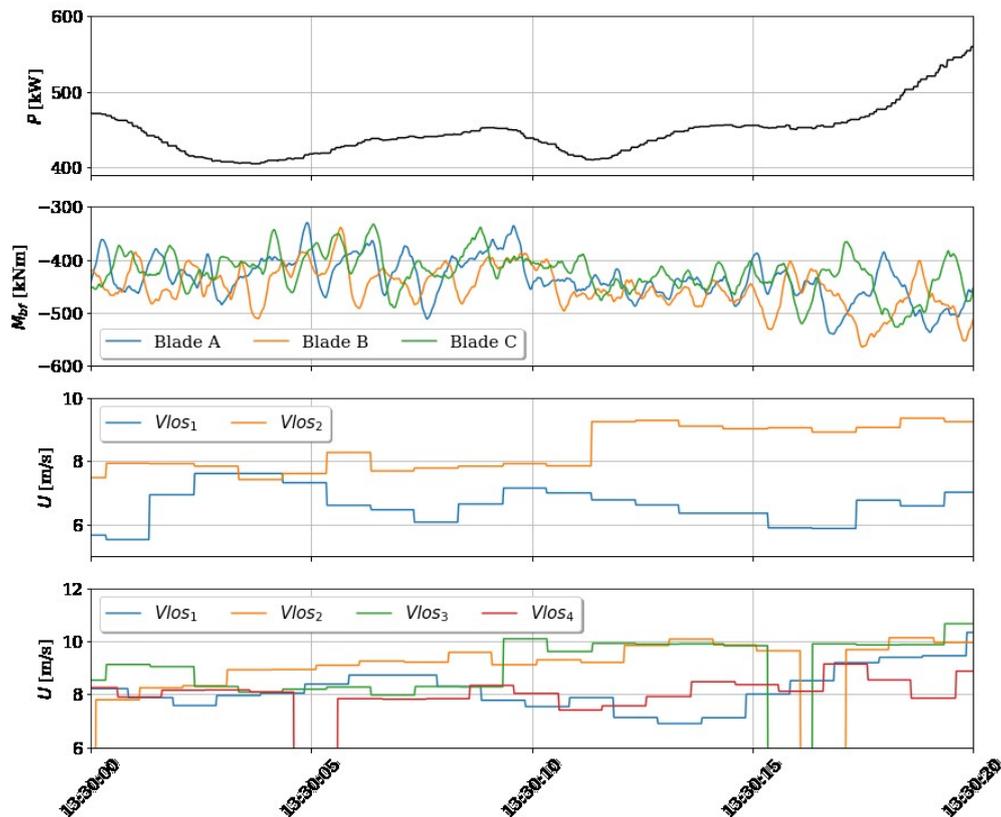


Figure 24: Example of a 50 Hz time series of power from SCADA, flap-wise bending moment from strain gauge measurements and LOS wind speeds from the 2- and 4-beamed LiDAR systems.

6.3.2 Data acquisition, processing and feature selection

The data acquisition flow collects information from three measurement systems (SCADA, strain gauge and lidar systems), which are synchronized to match the high sampling frequency of 50 Hz that is used for the strain gauge system. The 50 Hz synchronized data is then stored together with basic 10-min statistics containing mean, minimum, maximum, and standard deviations. The next step in the process is to apply a base filter using queries on the 10-min statistics. The purpose of this filter is to ensure adequate inflow direction and normal operation of the turbine. In terms of power forecasting, it does not make sense to include data where the turbine is not operational therefore these data are not included. To ensure only operational status of the turbine, 10-min queries based on rotational speed and pitch are introduced. Additionally, only the wind direction sectors with relatively low turbulence levels are included to avoid gusts and additional noise during the initial training and evaluation of the machine learning based ultra-short term forecasters.

For the features to be included into the lidar based power and load forecasting algorithm, a data-driven correlation analysis is performed between the target values and the available signals. However, one of the drawbacks of using data-driven feature selection is the multicollinearity, which is very difficult to avoid in many observational experiments as the channels are typically somewhat correlated by nature, e.g., we expect the LOS wind speeds to be naturally correlated as they measure the same local wind. The final feature selection was based on a trial-and-error process where different combinations of features were used and the performance was compared. The features that were ultimately found to be performing best, and therefore used in the forecasts, are summarized in Table 7.



V_{LOS}	all individual LOS wind speeds
U	horizontal wind speed, scalar
φ	yaw misalignment of the turbine to the incoming wind
σ_U^{60s}	moving standard deviation of horizontal wind speed based on a 60 s window
β	wind shear exponent (only available for the 4-beamed LiDAR)

Table 7: Final input features to the power and load forecaster.

After the final features are selected, the modelling can be initialised. The first step is to divide the dataset into a training and a test set. Following common practices in data analytics, the training/test split was chosen to be an 80% - 20% of the processed database and the splitting was done using random selection which is possible after the sequential transformation has been performed. This ensures equivalent characteristics of the training and test sets which is critical in terms of creating a generic and representative model. After splitting the data, it was scaled to the values between -1 and 1. Practically, the scaler is fitted to the training data only, and then both training and test sets are normalised with the scaler. It is important that the scaling is done after splitting the data into training and test sets. This is done to prevent information about the test set, leaking into the model, thereby creating a biased model.

6.4 Forecast models

For the ultra-short term power and load forecasting using lidar measurements at the DTU Risø test site, 4 different machine learning approaches have been utilised. All the models have the power and flap-wise bending moment as the target features, with the same input features listed in Table 7. Note that only the 4-beam lidar based forecasters utilise the wind shear exponent, β , as the input feature. As mentioned earlier, with the 2-beam lidar inputs the forecast horizon is 10 s with 1 Hz resolution (i.e. 10-steps ahead), where for the 4-beam lidar forecasts reach up to 20 s with 0.5 Hz resolution (i.e. 40-steps ahead).

6.4.1 Feed-forward neural network (FFNN)

There exist several classes of neural networks that process the flow of information through the network differently. The simplest type that is employed within the data-driven forecasting literature is the feedforward neural network (FFNN) which passes information straight through the network without any loops, i.e., it has no concept of memory. In order to address that in a practical manner, the history of the input features is also introduced to the model with lag = 10 s. However, FFNN has structural limitations in identifying the time dependence between these historical inputs and considers them as independent inputs to the network. Using TensorFlow library for the training, grid search for hyperparameter tuning is performed and the final FFNN architecture has 2 hidden layers with 50 neurons each.

6.4.2 Long-short-term memory (LSTM) network

LSTM is a type of recurrent neural networks which is frequently used in sequential and temporal predictive modelling, including forecasting. Due to their 'memory' feature via updating the cell state within the neurons, LSTM networks have the capability to include temporal dependencies. However, they are highly parameterized, requiring large amounts of data for robust training and accurate predictions. After sequential transformation of the input features,

the grid search for the hyperparameter tuning of the LSTM architecture suggested 2 hidden layers with 50 neurons.

6.4.3 Gated Recurrent Unit (GRU)

Another type of recurrent neural networks, GRUs are very similar to LSTMs in the sense of the memory capabilities through the update of the cell state. However, GRUs control the information flow within the past and future states with less parameterization attributed to the “gates” of a neuron. That corresponds to an equivalent performance with a relatively lightweight model hence faster training and relatively less data requirements. Same architecture with 2 hidden layers and 50 neurons each is utilised in this analysis for the GRU forecasters.

6.4.4 Random Forest (RF)

RFs are ensemble learning algorithms that generate a subset of features based on the trends captured in the randomly bootstrapped data (decision trees). It is typically used with tabular and/or categorical data, hence not commonly used in forecasting with high temporal correlation. However, for the ultra-short term forecasting problem defined here where the inflow features are measured ahead of time with lidars, RFs are a suitable choice for wind speed to power or loads conversion (a.k.a. tabular data). They have much less hyperparameters to tune, and the final configuration used for this task are summarised in Table 8.

Parameter	2-beam	4-beam
Number of estimators	300	40
Min samples leaf	3	1
Min samples split	6	2

Table 8: RF hyperparameters.

6.5 Evaluation of the Results & Discussion

As it is still the most commonly used and considered to be the baseline for ultra-short term forecasting, the persistence model is also included in all the results presented in this section. The discussions therefore focus on the improvements based on persistence and elaborate on the behaviour of the best performing model. For the evaluation of the forecasts, the RMSE improvement metric is used as indicated in the KPI index 1.2.a in Annex A (i.e., up to 30min wind forecasts: 7-9% RMSE).

Figure 25 shows the power forecasting performance of the benchmarked models using 2-beam and 4-beam lidar configurations. For the 2-beam, it is seen that the performances of all four data-driven forecasters are equivalent both in terms of the mean absolute error (MAE) and the RMSE, where the baseline persistence is surpassed around 4-s ahead. For the 4-beam, with more information available for the models to train on, the difference in their behaviour is more notable. For the implemented architectures described in Section 6.4, FFNN has the highest RMSE (up to 9% for the longest horizon at 20s); LSTM and GRU has very comparable performance with up to 8% RMSE; and RF is the best performer with less than 7% RMSE at 20s forecasts. All the models surpass persistence forecasts as early as 2-s ahead, and the best performer RF forecaster shows up to 4% decrease in RMSE at 20s in comparison. Although a simpler configuration compared to deep learning architectures, RF has the right capabilities to

capture the conversion of wind speed to power, where the former is known/measured ahead of time with the forward looking lidars. In other words, the benchmark indicates the forecasting problem defined in this task is driven by the ahead of time/space lidar measurements, instead of past observations of the turbine response as seen in conventional forecasting setups.

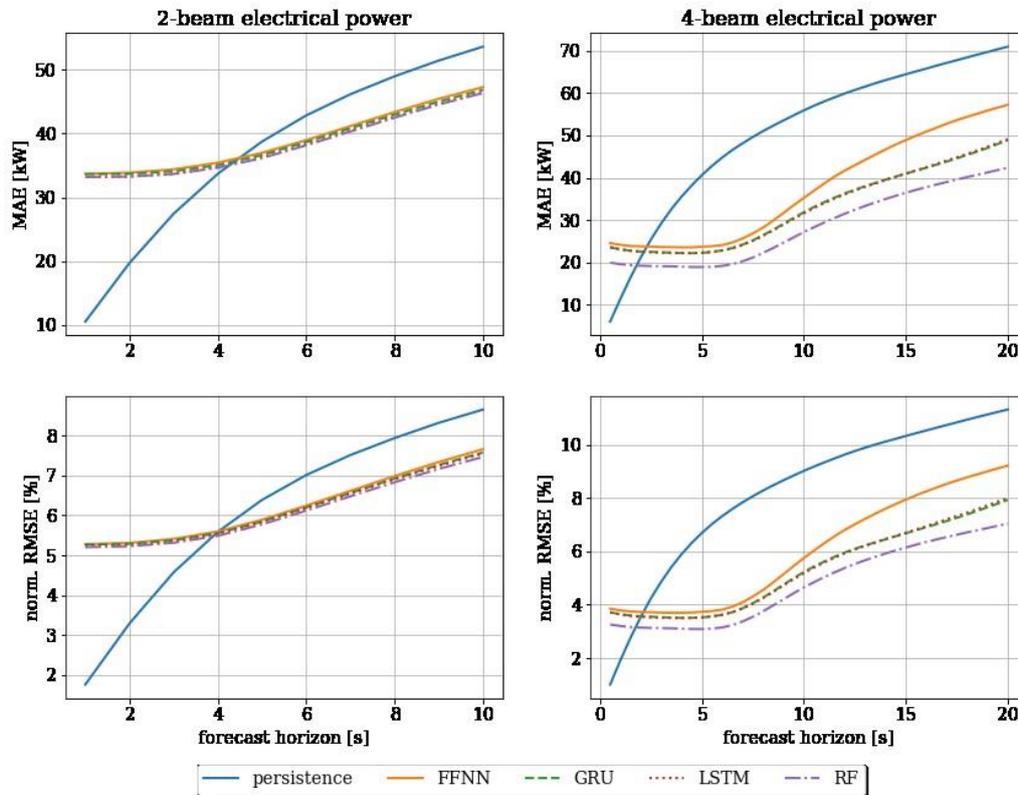


Figure 25: Power forecast benchmark, evaluated on mean absolute error (MAE) and normalised root mean square error (RMSE) using; i) left: 2-beam lidar, ii) right: 4-beam lidar.

Figure 26 presents the performance of the models for their second target value, flapwise bending moment structural load measured at V52 turbine. Due to the cyclic character of the measurements (passing through the tower), the persistence baseline is askew. However, especially for 4-beam lidar configuration, the models produce robust and high-accuracy forecasts, with RF as the best performer (app. 6% RMSE at 20s). This is particularly useful since, once trained, the model forecasts can be used as virtual sensors to estimate the load channels where the measurements are not available. An example 20min time series for the 5-s ahead RF flapwise bending moment forecasts presented in Figure 27, which shows how the involved dynamics are captured and further supports the virtual sensor application.

Further details on the model training, as well as extended discussion of the results are a part of a manuscript under preparation to be submitted to Wind Energy Science journal.

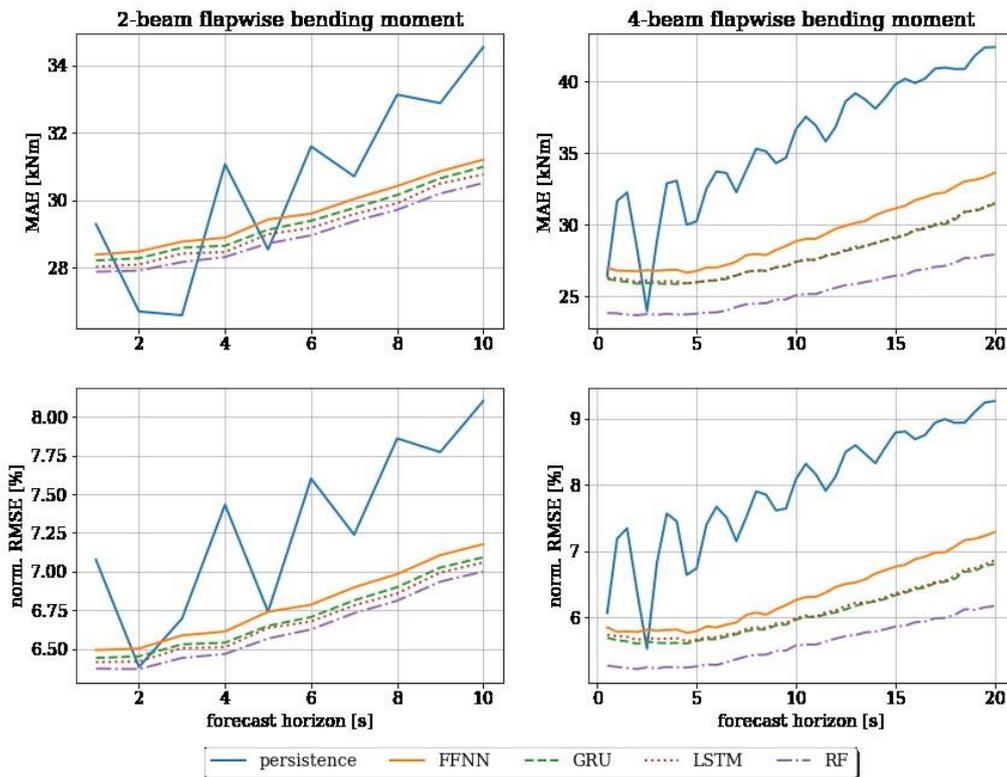


Figure 26: Structural load forecast benchmark, evaluated on mean absolute error (MAE) and normalised root mean square error (RMSE) using; i) left: 2-beam lidar, ii) right: 4-beam lidar.

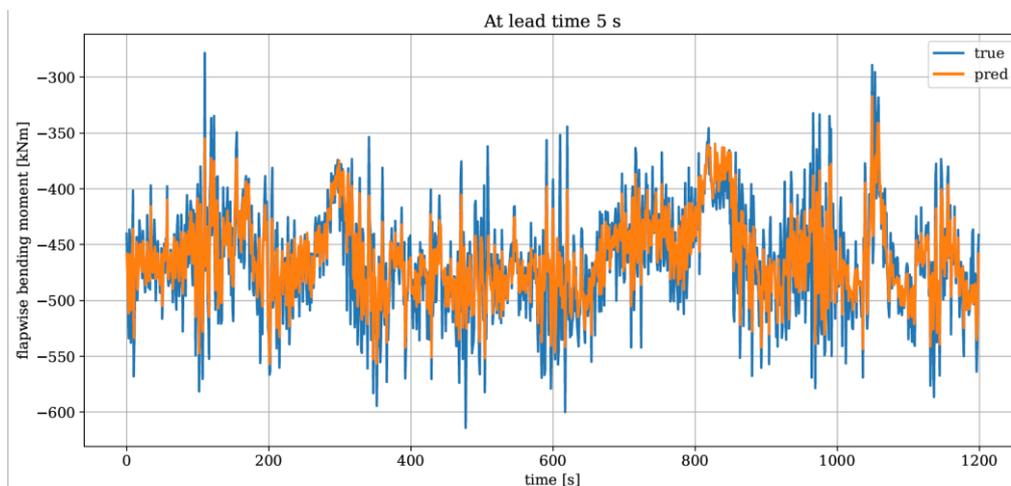


Figure 27: 20min time series comparison of 5-s ahead RF flapwise bending moment forecasts.

References

Paulsen, U. S., Gomiero, M., Larsen, T. J. and Benini, E. (2018), "Load validation of aero-elastic simulations with measurements performed on a 850kW horizontal-axis wind turbine", *Journal of Physics: Conference Series*, 1037(6), [062023], doi: 10.1088/1742-6596/1037/6/062023

Rinker, J. M., Hansen, M. H. and Larsen, T. J. (2018), "Calibrating a wind turbine model using diverse datasets", *Journal of Physics: Conference Series*, 1037(6), [062026], doi: 10.1088/1742-6596/1037/6/062026



7 PossPOW SCADA based forecasting (DTU)

7.1 Motivation

Minute-scale forecasting is prone to operational uncertainties (including availability, down-regulation, etc.) and local atmospheric conditions (including wake effects). Therefore, updated, dynamic forecasts can be very useful for day-to-day operations and optimum control of wind farms. Here in this task, the aim is to investigate the capabilities of the modern AI-based forecasting approaches for such dynamic forecasts, as well as to demonstrate the value of the additional SCADA channels that are not conventionally used in forecasting solutions. Moreover, dynamic models that are continuously updated over time to capture the trends in the most recent observations are implemented and compared with single generic network approach as well as the persistence baseline.

7.2 Case studies

7.2.1 La Haute Borne onshore wind farm

For the SCADA based ultra-short term forecasting in this task, the open dataset from La Haute Borne wind farm³ operated by ENGIE is utilized. It consists of four Repower MM82 turbines with 2MW rated power and 85 m diameter. Within the dataset, approximately 5-years time series of several SCADA channels in their 10-min statistics (minimum, maximum, average and standard deviation) are recorded. Figure 28 shows the layout of the wind farm and the observed wind rose during the available period.

The available sampling rate of the dataset also defines the forecast horizon as 10-min for the investigation in this task.

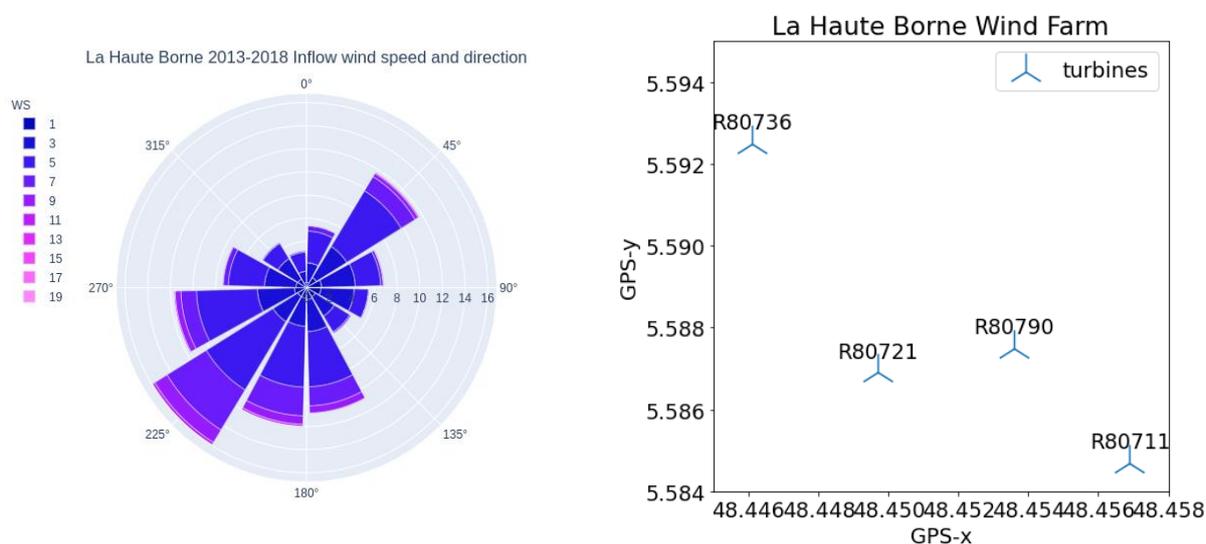


Figure 28: La Haute Borne, i) left: Wind rose, ii) right: wind farm layout with turbine IDs.

³ <https://opendata-renewables.engie.com/>

7.2.2 Description of Data, Processing & Model Training

From the SCADA available, 10-min mean and standard deviation of the channels wind direction, wind speed, pitch, rotational speed and power per turbine are used as input features, together with the month of the year, day of the month and the hour of the day, to the GRU networks described in Section 6.4.3. All the inputs are processed as sequences for the GRUs, where the last 2-hours are used to predict the 10-min ahead mean power production of the individual turbines. Accordingly, the network(s) have 43 input and 4 target features in total.

Within the available period, the first 4-years is used for training the single (generic) network, where the last year is reserved for testing its performance. As the second approach, during this last year, the single network is updated continuously (i.e., at every 10-min step) to generate the dynamic GRU forecaster.

As a standard practice to avoid misinterpretation of the weights and biases during training, the input and target features are normalized by removing the mean and scaling to unit variance. Given the length and variety within the dataset, relatively deeper GRU networks with 3 hidden layers are configured. The grid search for the hyperparameter tuning performed for 32 – 512 neurons per hidden layer as well as relu, tanh or sigmoid activation functions in between resulted in the optimum configuration presented in Table 9. Within TensorFlow, the Glorot uniform initializer is chosen for the kernels and the initial biases are assigned as zeros.

Layer	# of neurons	# of parameters
GRU_1 (frozen layer)	64	14976
GRU_2 (frozen layer)	320	370560
GRU_3 (dynamically updated via TL)	32	33984
Dense_1 (dynamically updated via TL)	4	132
Total parameters: 419 652		
Trainable parameters - Single Network: 491 652, Transfer Learning (TL): 34 116		
Non-trainable parameters - Single Network: 0, Transfer Learning: 385 536		

Table 9: Architecture of the Single (generic) GRU network with lag = 2-hours. Hidden layers: GRU_1, GRU_2, GRU_3. Output Layer: Dense_1. The activation functions used are relu, sigmoid and tanh, respectively.

For the dynamic updates in the GRU network, the parameters of the first two layers (GRU_1 and GRU_2) is 'frozen' where only the parameters in the last hidden layer, GRU_3, as well as the output layer, Dense_1, are re-trained. As seen in Table 9, this transfer learning approach significantly reduces the number of trainable parameters, enabling fast updates in the network providing fine-tuned, dynamic forecasts.

7.3 Evaluation of the Results and Discussion

With persistence as the baseline, for the evaluation of the forecasts generated by both the single and dynamically updated networks, primarily the relative RMSE improvement metric used as indicated in the KPI index 1.2a in Annex A (i.e., up to 30 min ahead wind forecasts: 7-9% RMSE). However, as this task is focused on the dynamic operation of the wind farms, including off-performance conditions, the percentage error of the forecasts is also discussed in

detail. As under or over-prediction of power has a different value e.g., in terms of market scenarios (positive or negative direction in the estimation of the reserves), percentage error, as defined below, is a crucial indicator for the forecasting problem defined in this task.

$$\text{Percentage error [\%]} = \frac{\text{Power}_{\text{predicted}} - \text{Power}_{\text{observed}}}{\text{Power}_{\text{observed}}} \cdot 100$$

Accordingly, the distribution of the percentage error for both of the GRU networks, as well as the persistence baseline is presented and discussed in the rest of the section. The forecasts cover the last year of the available period, starting at the time stamp 2017-01-16, 13:00.

7.3.1 Single (Generic) GRU Network

Evaluation with respect to the KPI 1.2.a yields 5% RMSE in 10-min wind farm power forecasts for both the persistence baseline and the single (generic) GRU network. However, the percentage error distributions presented in Figure 29 for both approaches show significant bias for the persistence towards over-estimation. The generic GRU network on the other hand has less than 2% mean percentage error towards under-estimation with a narrower distribution (4% less standard deviation of the percentage error), indicating lower forecast uncertainties.

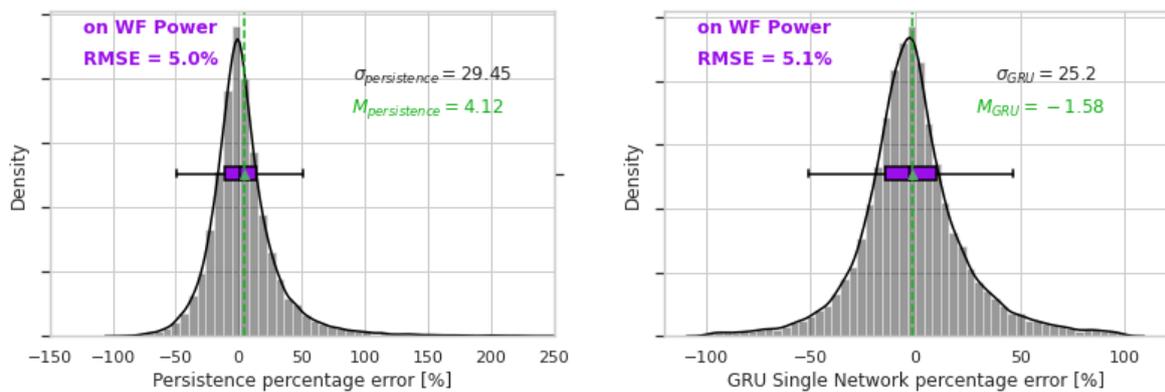


Figure 29: 10-min ahead La Haute Borne wind farm power forecast performance in percentage error and RMSE, i) left: persistence forecasts, ii) right: single (generic) GRU network forecasts.

7.3.2 Dynamic GRU Networks (via Transfer Learning)

Similar to the generic networks and the baseline persistence, the forecasts are evaluated in terms of the RMSE as well as the percentage error distributions. The corresponding performance of the continuously updated GRU networks via transfer learning is presented in Figure 31. Compared to Figure 29, Figure 31 shows significant increase in forecast performance via dynamic GRU networks, with more than 2% improvement in RMSE. Additionally, the bias observed in the forecasts (median of the percentage error, as the distribution is not Gaussian) is less than 1% with more than 2% reduction in uncertainty compared to the single generic network and more than 6% compared to the persistence baseline.

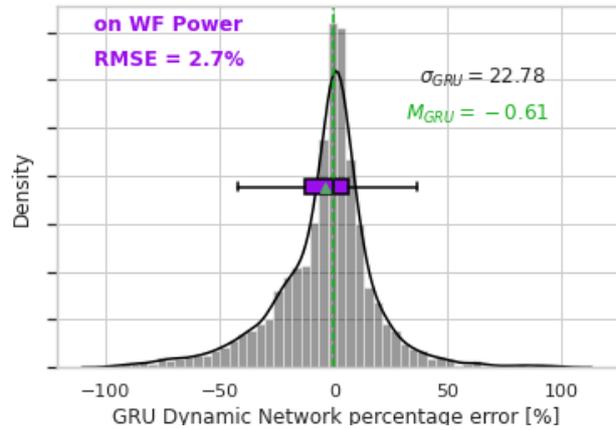


Figure 30: 10-min ahead La Haute Borne wind farm power forecast performance in percentage error and RMSE, dynamically updated GRU networks (via transfer learning).

8 Satellite-based PV forecasting (ICCS)

8.1 Motivation

ICCS investigated the application of satellite images to short-term solar power forecasting as a new input data source, specifically to an island where satellite images do not contain information above the sea. The experiments were performed on Rhodes case study. The total solar power capacity of the island is 18.2 MW and the solar power data cover the period from 01/01/2018 to 31/12/2019. In specific, the 5-hour ahead solar power forecasting of Rhodes's total production that is essential for the economic and secure management of an isolated power system, was performed. In order to evaluate the influence of satellite images to the forecasting performance, ICCS has developed a deep neural network that can handle inputs from different data sources. The data sources used to that investigation are the aggregated solar power production timeseries, ECMWF NWP's and solar irradiance maps from HelioClim-3 solar radiation database (Albarelo et al., 2015, <https://www.soda-pro.com>).

The solar irradiance maps are extracted by the Heliosat-2 model (Albarelo et al., 2015). The Heliosat-2 model processes the images derived by Meteosat Second Generation satellite during a day (72 images) and computes the 15-minute Global Irradiance values over a horizontal plane (GHI). More details about the characteristics of the satellite images database can be found in Blanc et al. (2011) and Vernay et al. (2013). The solar irradiance maps applied here cover the whole Europe area in 5 km spatial resolution, but contain information only above the land. Images for every 15 minutes have been used for the period of the years 2018 and 2019. In Figure 31, solar irradiance maps corresponding to 9 different hours picked randomly from the dataset are illustrated. The figures show that all the values above the sea are missed (blank colour). The values above the small islands are also missed. The only neighbour areas of Rhodes that have available information are the islands of Crete, of Naxos and of Samos.

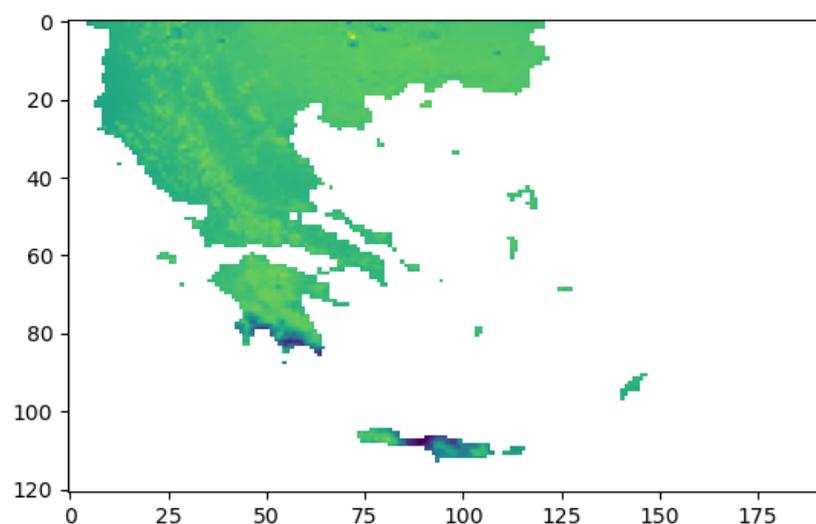


Figure 31: Solar irradiance map of Greece in Wh/m^2 . There is no information above the sea and above small islands.

8.2 Approach

As mentioned before, the solar irradiance maps' dataset has 15-minute timestep. For balancing the dataset with the NWP dataset and the solar power timeseries that are in hourly resolution, only the images corresponding to full hours are kept. Furthermore, in order to remove the influence of the solar position to irradiance maps, each pixel is normalized by the daily maximum value of the clear-sky irradiance. The clear-sky irradiance depends on the solar position and represents the irradiance if the sky was clear.

Figure 31 also shows that there are large areas where the pixel information is highly correlated. In order to reduce the data size and the number of forecasting model parameters, it is necessary to select only the pixels that provide the most useful information (Agoua et al., 2021). To do that, firstly, the map pixels are grouped depending on the correlation between the irradiance timeseries $I_{t,i,j}$ provided by the pixel (i, j) at time t and the solar power timeseries p_{t+h} recorded at time $t+h$, where h is the forecasting horizon in hours, namely between 1..5. Computing the correlation between each pixel of the image $I_{t,i,j}$ and the solar power p_{t+h} , a correlation map for each forecasting horizon h is created. Then, the local maxima of each correlation map are estimated and the pixels are grouped based on the nearest maxima location. As shown in Figure 32, the irradiance maps that correspond to the current hour present positive correlation with the solar power timeseries for 3 hour-ahead, while for the rest forecasting horizon have negative correlation and instead of local maxima, the local minima are computed.

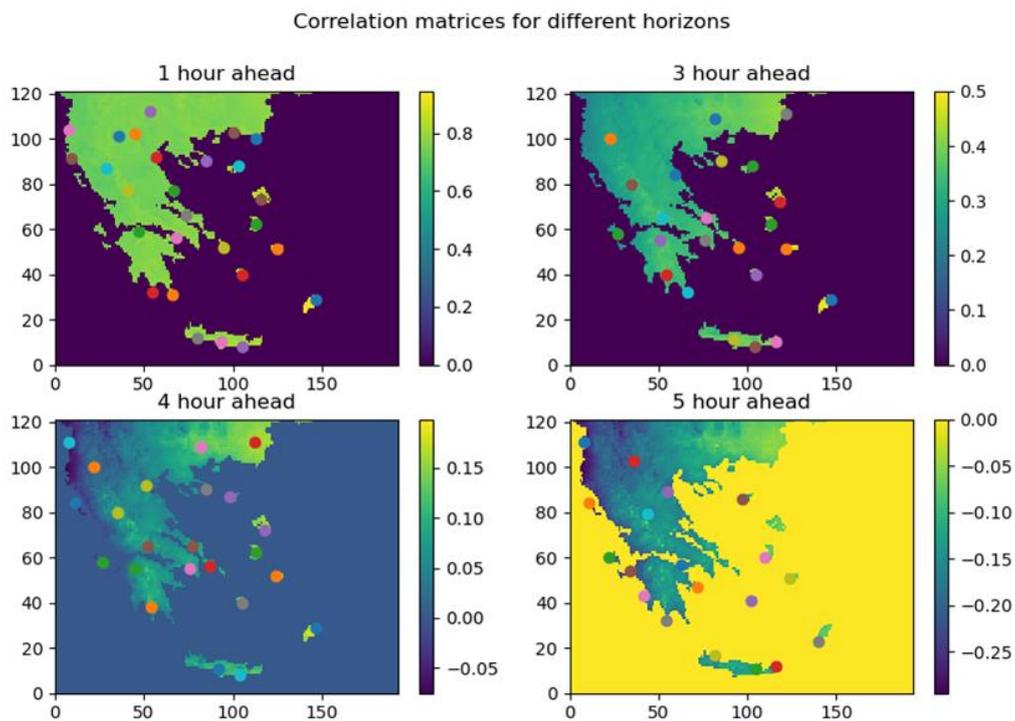


Figure 32: The correlation maps for 1, 2, 3 and 5 hour-ahead horizons with the respective maxima.

After grouping the image pixels, a feature selection method is applied for indicating the pixels with the adequate information. In this work, two different feature selection methods are tested – the random forest variable importance (RF) (Breiman, 2001) and the group lasso method (Yuan & Lin, 2001).

Random forest variable importance

In random forest, each tree has its own out-of-bag sample of data that was not used during construction. This sample is used to calculate importance of a specific variable. First, the prediction accuracy on the out-of-bag sample is measured. Then, the values of the variable in the out-of-bag-sample are randomly shuffled, keeping all other variables the same. Finally, the decrease in prediction accuracy on the shuffled data is measured. The mean decrease in accuracy across all trees is reported. Intuitively, the random shuffling means that, on average, the shuffled variable has no predictive power. This importance is a measure of by how much removing a variable decreases accuracy, and vice versa.

Group lasso method

The group lasso method is an extension of the Lasso (least absolute shrinkage and selection operator) linear regression method that utilizes the L1 regularization, namely an additional parameter in cost function penalizing the large weights. Lasso performs both variable selection and regularization. Group lasso uses predefined groups of input variables to jointly be selected into or out of a model. The objective function of the group lasso method has the following form

$$\min_{\beta \in \mathbb{R}^p} \left\{ \left\| y - \sum_{j=1}^J X_j \beta_j \right\|^2 + \lambda \sum_{j=1}^J \|\beta_j\|_{K_j} \right\}, \quad \|z\|_{K_j} = (z^T K_j z)^{1/2}$$

where the inputs X and covariate vector β have been replaced by a collection of X_j and covariate vectors β_j , for each of the J groups. Additionally, the penalty term is defined by the matrices K_j . If each covariate is in its own group and K_j is an identity matrix, then this reduces to the standard lasso, while if there is only a single group and $K_1=I$, it reduces to ridge regression. At the case of standard lasso, the penalty is piecewise linear, and this gives rise to the piecewise linear solution path. Intuitively, in the group lasso case, the penalty is no longer piecewise linear.

Since the most valuable features are selected, they are placed on a two-dimensional matrix based on their location of the original irradiance maps. Specifically, in order not to lose the spatial interactions of the irradiance maps, the pixel with the higher latitude and the lower longitude is first placed on the (0, 0) matrix position and then the neighbour matrix positions are filled with the pixels that have the lowest spatial distance. That procedure is repeated until all the pixels selected by the feature selection method are placed on the matrix.

The pixel selection and their organization to a two-dimensional matrix are performed for each forecasting step, resulting a four-dimensional tensor with size $N \times T \times M \times M$. Where N is the number of data samples in the training set, T is the forecasting horizon and $M \times M$ is the size of the two-dimensional matrix mentioned above. Actually, M is the maximum of $\{M_1, M_2 \dots M_5\}$ that are the sizes of the matrices constructed at each forecasting step. All the T matrices are resized to $M \times M$ with interpolation. The four-dimensional tensor created is inputted to a deep learning model that consists of one LSTM-Conv and three fully-connected layers. The LSTM-CNN layer is the combination of a long-short memory recurrent layer (LSTM) and T convolutional layers (Conv). Each Conv layer handles one of the T $M \times M$ matrices and their outputs are concatenated to feed the LSTM layer.

A convolutional layer processes the data that have a grid-like topology, such as an image. Typically, a Conv layer has two layers: a convolutional layer and a pooling layer. This layer performs a dot product between two matrices, where one matrix is the set of learnable parameters, known as kernel and the other matrix is the restricted portion of the image. So, the



kernel is spatially much smaller than the input image. During the convolution, the kernel slides across the height and width of the image, producing an image representation of that receptive region. This produces a two-dimensional representation of the image known as an activation map that gives the response of the kernel at each spatial position of the image. The pooling layer replaces the output of the Conv layer at certain locations by deriving a summary statistic of the nearby outputs. This helps in reducing the spatial size of the representation, which decreases the required amount of computation and weights. The pooling operation is processed on every slice of the representation individually.

On the other hand, LSTM networks were designed specifically to overcome the long-term dependency problem faced by recurrent neural networks RNNs (due to the vanishing gradient problem). LSTMs have feedback connections which make them different to more traditional feedforward neural networks. This property enables LSTMs to process entire sequences of data (e.g. time series) without treating each point in the sequence independently, but rather, retaining useful information about previous data in the sequence to help with the processing of new data points. As a result, LSTMs are particularly good at processing time-series. Combining Conv and LSTM layers, the proposed model is able to account for the solar power sequential dependencies in time and the dependencies distributed spatially to the satellite images or to the numerical weather prediction grids.

Figure 35 shows the structure of the proposed model, while Figure 33 and Figure 34 illustrate the benchmark models used for this study. Firstly, the benchmark model of Figure 33, called here as TS, is a multilayer perceptron neural network that receives only data from the solar power timeseries. It has two branches of fully-connected layers that are connected to the output layer, one for calendar data and one for past measurements. The second benchmark model, called here as TS-NWP, has a third branch of layers that consists of a LSTM-Conv layer and two fully-connected layers that process the NWP data. The NWP data are formed to a $N \times T \times 10 \times 10 \times 4$ matrix, where N is the number of training set, T is the forecasting horizon, 4 is the number of the weather variables (cloud coverage, downward short-wave flux, temperature and wind speed) and 10×10 is the frame of NWP grid that surrounds the island of Rhodes. Finally, the proposed model shown in Figure 35 and called here as TS-NWP-SAT, has an additional layer branch similar with the layer branch that processes the NWP data. This layer branch receives as input the four-dimensional tensor with size $N \times T \times M \times M$ that contains the selected irradiance map pixels as described above.

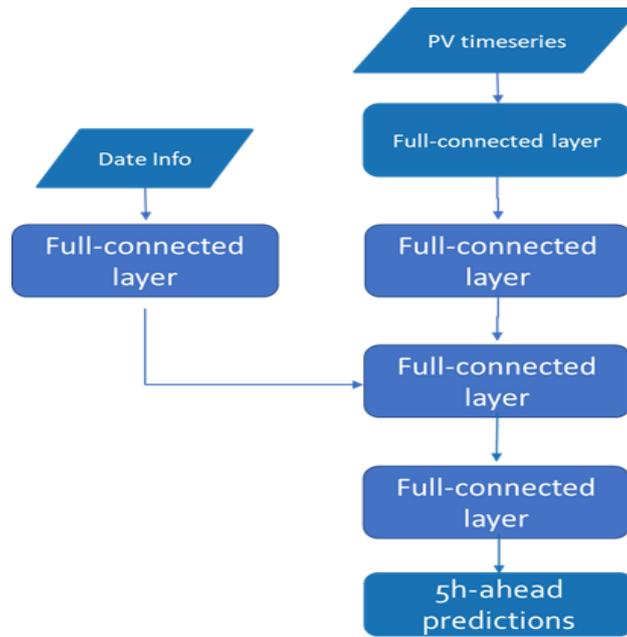


Figure 33: The structure of the model that uses only timeseries data and it is used as benchmark.

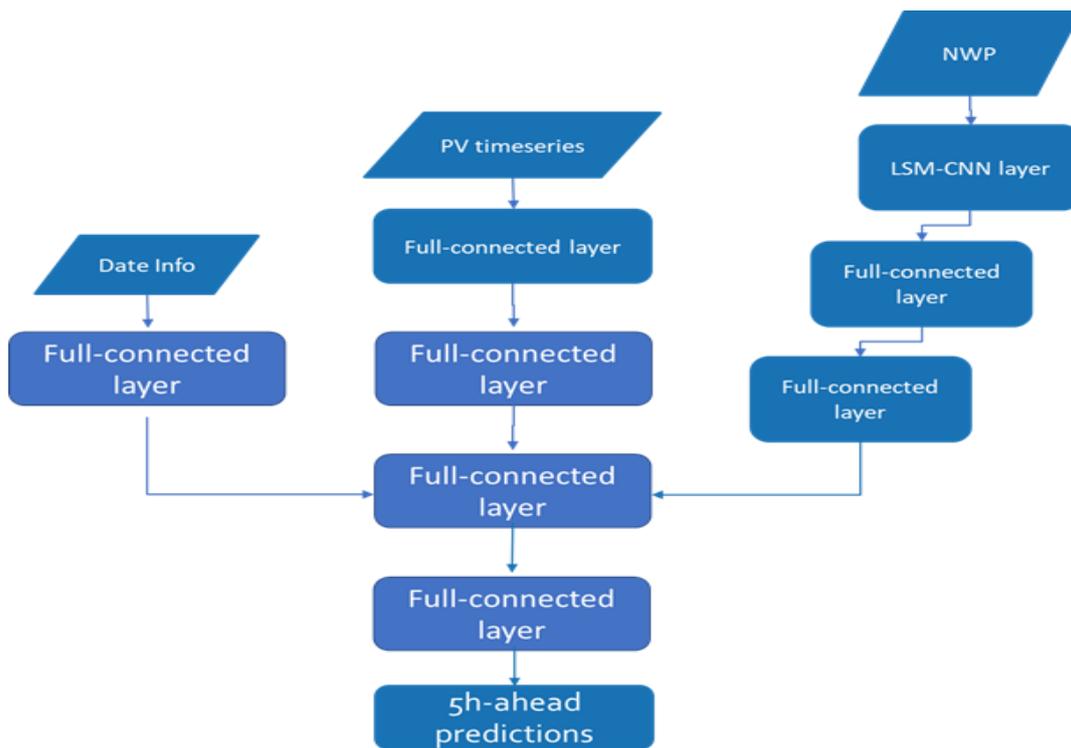


Figure 34: The structure of the model that uses timeseries data and NWP and it is used as second benchmark.

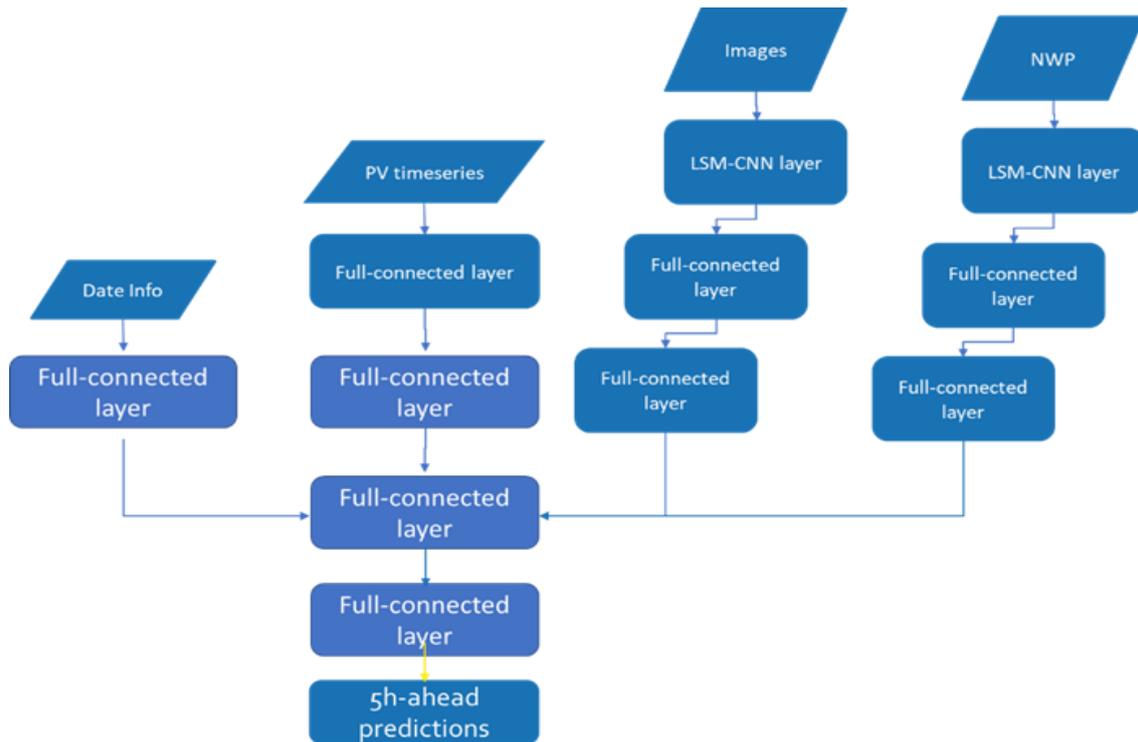


Figure 35: The structure of the proposed model that uses timeseries data, NWP and irradiance maps.

8.3 Case studies

The forecasting models described above are tested on an isolated power system case study: the total solar power production of Rhodes Island. Together with those forecasting models, three prediction methods together with deep learning architectures described above are evaluated. In specific, for each variable selection method, a simple linear regression, the Lasso regression method and the random forest method are applied using the datasets of the three trials. At each trial, the corresponding deep learning architecture dl1 for the first trial, dl2 for the second and dl3 for the third trial are also evaluated.

- **At the first trial, the 5-hour ahead prediction is performed using only timeseries data (solar power, solar position and calendar data. The deep learning architecture dl1 evaluated here.**
- **At the second trial, NWP data is added to the first trial's dataset. The deep learning architecture dl2 evaluated here.**
- **At the third trial, the dataset consists of irradiance data, NWP data and timeseries data. . The deep learning architecture dl3 evaluated here.**

The Rhodes total solar power production timeseries has hourly timestep and cover the period from 01/01/2018 to 31/12/2019. The total PV capacity of the island is 18.2 MW. The data used for training the models comes from the period from 01/01/2018 to 31/08/2019, while the testing period starts at 01/09/2019 and ends at 31/12/2019. In this case, the ECMWF NWP are used, while irradiance maps from SODA covering the area with latitude from 34° to 42° and longitude from 19° to 31° are also used. The irradiance maps covering that area consist of 121x193 pixels.

8.4 Evaluation

As mentioned before, three trials with different datasets are performed in order to prove the effectiveness of satellite images use on short-term solar power forecasting. The data used on the first trial consists of timeseries data, during the second trial timeseries data and NWP data were used and both timeseries data, NWP data and irradiance map data were applied at the third trial. In specific, the included timeseries data are: power measurements of the past three hours, the prediction hour and month and the solar zenith and azimuth. NWPs for each of the five prediction hours are used and the most informative pixels from the irradiance map of the last hour are selected for each forecasting step by the applied feature selection method.

8.5 Results and Discussion

1. Using timeseries data

Four regression methods were implemented on the timeseries data and the results will be used as a benchmark to evaluate the improvement obtained by the use of irradiance maps. **In order to make the comparison fair, only the predictions of daylight hours are considered to the evaluation.** Table 10 contains the normalized mean absolute error (NMAE) of the Linear regression (lr), of the Lasso regression (lasso), of the random forests (RF) and of the deep learning architecture shown in Figure 33 (dl1) obtained on the Rhodes case-study.

	1h-ahead	2h-ahead	3h-ahead	4h-ahead	5h-ahead
lr	6.21%	10.67%	12.42%	13.23%	13.75%
lasso	6.25%	10.69%	12.46%	13.26%	13.76%
RF	3.54%	5.60%	6.62%	6.90%	6.73%
dl1	4.67%	7.03%	8.93%	9.66%	9.68%

Table 10: NMAE from the Rhodes case-study using timeseries data.

The best performance was obtained by the RF method, where the NMAE average over the prediction horizons (1h to 5h ahead) was 5.87% for the Rhodes.

2. Using timeseries and NWP data

Utilizing the same prediction methods with both timeseries and NWP data, but instead of dl1, here the deep learning architecture of Figure 34 (dl2) is applied. The results are presented in Table 11.

	1h-ahead	2h-ahead	3h-ahead	4h-ahead	5h-ahead
lr	4.57%	5.93%	6.03%	5.88%	5.70%
lasso	4.60%	6.30%	6.40%	6.28%	5.73%
RF	3.55%	4.54%	4.74%	4.47%	4.14%
dl2	3.92%	5.26%	5.56%	5.37%	4.94%

Table 11: NMAE from the Rhodes case-study using timeseries and NWP data.

At the Rhodes case-study, the averaged RF NMAE was 4.29% with the NWP data use and the improvement w.r.t the RF performance with only timeseries data was 27%. The deep learning architecture dl2 provided averaged NMAE 5.01% during the second trial, while at the first trial the averaged NMAE of dl1 was 7.99%.

3. Using timeseries data, NWP data and irradiance maps

Finally, collecting the timeseries data, the NWP data and the irradiance maps to one dataset, the dl3 model of Figure 35 and the three benchmark methods lr, lasso and RF were evaluated for each of the two-pixel selection methods, namely the random forest variable importance index (RFVII) and the group Lasso method (gLasso). Firstly, considering Rhodes case study, the results for each pixel selection method are presented in the following two tables.

	1h-ahead	2h-ahead	3h-ahead	4h-ahead	5h-ahead
lr	4.42%	6.00%	6.16%	6.06%	5.86%
lasso	4.89%	6.47%	6.42%	6.24%	5.69%
RF	3.37%	4.38%	4.71%	4.52%	4.20%
dl3	3.98%	4.74%	5.07%	4.76%	4.51%

Table 12: NMAE from the Rhodes case-study obtained using all data and the irradiance map pixels are selected by RFVII.

	1h-ahead	2h-ahead	3h-ahead	4h-ahead	5h-ahead
lr	13.56%	20.01%	23.28%	23.99%	23.52%
lasso	5.92%	10.17%	11.73%	11.79%	11.64%
RF	4.10%	5.73%	6.85%	7.03%	6.61%
dl3	4.45%	5.88%	6.70%	6.82%	6.49%

Table 13: NMAE from the Rhodes case-study obtained using all data and the irradiance map pixels are selected by gLasso.

The best results at the Rhodes case study were obtained by the RFVII method. In specific, the RF method provided averaged NMAE 4.24% when the image pixels were selected by RFVII method. The dl3 model working with RFVII method gave averaged NMAE 4.61%.

Comparing the results obtained with the RF with those without the use of an irradiance map, there is significant improvement at the 1 hour-ahead and 2 hour-ahead predictions (see Table 14). While the deep learning architecture dl3 generally performs not as good as the RF, the improvement using the irradiance map is on average 8%, except for the 1-h ahead predictions (see Table 15). Figure 36 shows the RF performance for every trial and for different horizons. Also, Figure 37 compares lasso, RF, and dl3 performance when they combined with RFVII pixel selection methods.

	1h-ahead	2h-ahead	3h-ahead	4h-ahead	5h-ahead	AVERAGE
RF-TS+NWP+SAT	3.37%	4.38%	4.71%	4.52%	4.20%	4.24%
RF-TS+NWP	3.55%	4.54%	4.74%	4.47%	4.14%	4.29%
RF-TS	3.54%	5.60%	6.62%	6.90%	6.73%	5.88%

Table 14: NMAE from the Rhodes case-study obtained by RF method for each trial.

	1h-ahead	2h-ahead	3h-ahead	4h-ahead	5h-ahead	AVERAGE
dl3-TS+NWP+SAT	3.98%	4.74%	5.07%	4.76%	4.51%	4.61%
dl2-TS+NWP	3.92%	5.26%	5.56%	5.37%	4.94%	5.01%
dl1-TS	4.67%	7.03%	8.93%	9.66%	9.68%	7.99%

Table 15: NMAE from the Rhodes case-study obtained by the deep learning architectures on each trial.

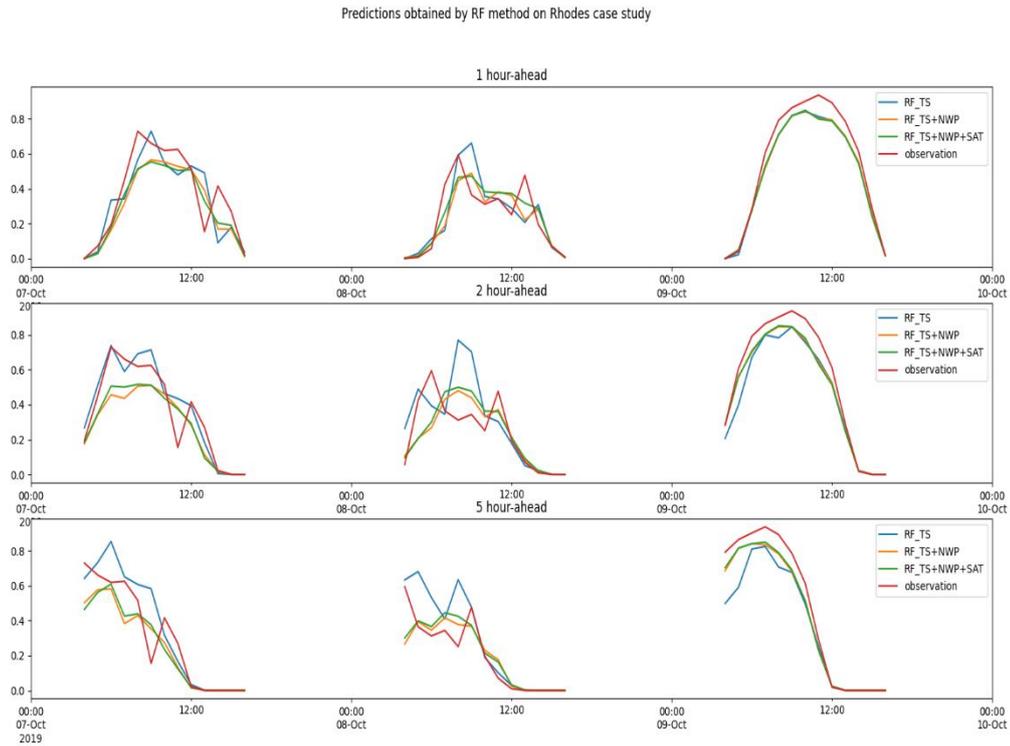


Figure 36: Predictions obtained by the RF method and the RFVII pixel selection method for every trial and for 1, 3 and 5 hour-ahead forecasting horizon.

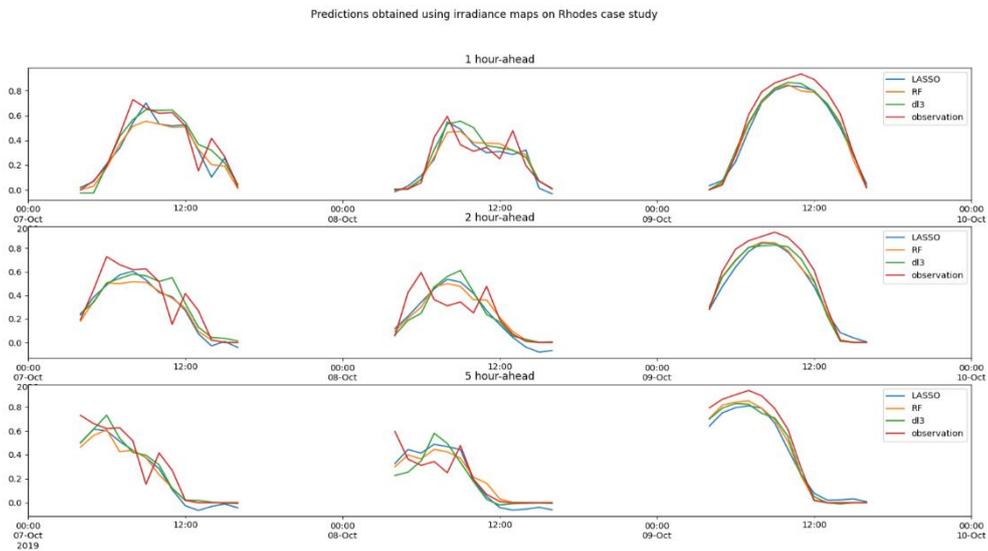


Figure 37: Predictions obtained by the different forecasting method the RFVII pixel selection method for 1, 3 and 5 hour-ahead forecasting horizon.

References

Agoua G. X., Girard R. and Kariniotakis G. (2021), "Photovoltaic Power Forecasting: Assessment of the Impact of Multiple Sources of Spatio-Temporal Data on Forecast Accuracy", *Energies*, 14, 1432



Albarelo T., I. Marie Joseph, A. Primerose, F. Seyler, L. Linguet and Wald, L. (2015), "Optimizing the Heliosat-II method for surface solar irradiation estimation with GOES images", *Canadian Journal of Remote Sensing*, 41, 86-100

Blanc, P., Gschwind, B., Lefèvre, M. and Wald, L. (2011), "The HelioClim Project: Surface Solar Irradiance Data for Climate Applications. *Remote Sens*", 3, 343–361

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

Vernay, C., Blanc, P. and Pitaval, S. (2013), "Characterizing measurements campaigns for an innovative calibration approach of the global horizontal irradiation estimated by HelioClim-3", *Renew. Energy*, 57, 339–347

Yuan M and Lin Y. (2001), "Model selection and estimation in regression with grouped variables", *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68, 49-67

9 Day-ahead solar power forecasting using NWP obtained by different vendors

9.1 Motivation

ICCS investigates if the performance of a forecasting model can be improved when using as input NWP obtained by two different vendors. Specifically, the accuracy of day-ahead predictions of the Rhodes total solar power production is presented here. The total solar power capacity of the island is 18.2 MW. The solar power predictions are provided by a state-of-the-art forecasting model (Sideratos et al., 2020) that uses as input both European Centre for Medium-Range Weather Forecasts (ECMWF) and Global Forecasting System (GFS) NWP that have 0.1° and 0.25° spatial resolution, respectively. The ECMWF NWP come from its HRES model, while GFS NWP have been downloaded by Google datasets⁴. The forecasting model runs once per day and predicts separately the solar power of each hour without considering the time dependence of the power data. The above predictions are compared with those obtained by the same forecasting model that uses NWP from one of the two vendors, namely ECMWF and GFS.

In addition, ICCS tests the application of a post-processing model that consists of a combination of two deep learning neural components – the long short memory recurrent neural layer (LSTM) and the convolutional neural layer (Conv) and performs a fine calibration to daily predictions. The post processing model receives the day-ahead solar power predictions together with NWP in order to capture the time dependencies of the timeseries.

9.2 Approach

The applied forecasting model called Fuzzy-RBF-CNN consists of three layers: a fuzzy clustering, an RBF-CNN regressor and an aggregation layer. At the first layer, a clustering method groups the input data into multiple clusters. The fuzzy clustering method forms clusters that have to share a portion of their space with their neighbouring clusters in order for every input to be able to activate more than one cluster and an ensemble prediction to be created at the second layer. Each cluster created corresponds to a fuzzy rule of the fuzzy clustering layer.

Following the fuzzy clustering method, the inputs are distributed to the second layer where an RBF-CNN regressor is applied to each cluster. As shown in Figure 38, an RBF-CNN regressor is an innovative neural network architecture composed by three RBF, a convolutional, a pooling and two fully connected layers and receives the data subset constructed by the corresponding cluster (fuzzy rule) at the first layer. The convolutional layer creates feature representations of the RBF kernel activations element-wise utilizing the kernel topology and reduces the impact of the non-useful information provided by each kernel. Training of an RBF-CNN regressor requires that the optimal parameters of the RBF kernels, namely the kernel number, centres and widths, are estimated through three different algorithms described in Sideratos and Hatzigargyriou (2020), Sideratos and Hatzigargyriou (2012), Sideratos and Hatzigargyriou (2007). The optimised RBF kernels transform the input data to a higher dimensional space with their activations becoming new data representations. The CNN is trained using the transformed input data at a second stage that analyses in more detail than an RBFNN, the relations of a kernel element with its neighbours; namely with the elements that correspond to a different input variable or kernel. Following the above two stage training procedure, an RBF-

⁴ <https://datasetsearch.research.google.com/>



CNN regressor operates as a compact neural network that consists of three RBF, a convolutional, an averaging pooling and two fully-connected layers.

The final prediction of the proposed model will be provided at the third layer by averaging the ensemble predictions of the RBF-CNN regressors that correspond to the clusters activated in the fuzzy clustering layer. Figure 39 shows the proposed model structure, where three clusters (darker shade) are activated and contribute to the final load prediction. In this example, input x_i activates fuzzy rules 2, 3 and 5 and the RBF-CNNs which are connected to these rules provide an independent prediction. The final prediction is obtained by averaging these predictions.

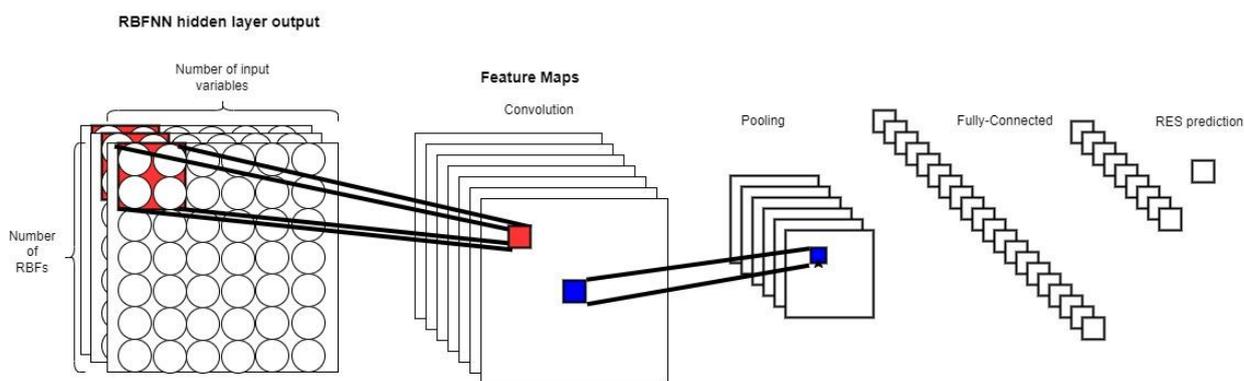


Figure 38: The RBF-CNN regressor structure.

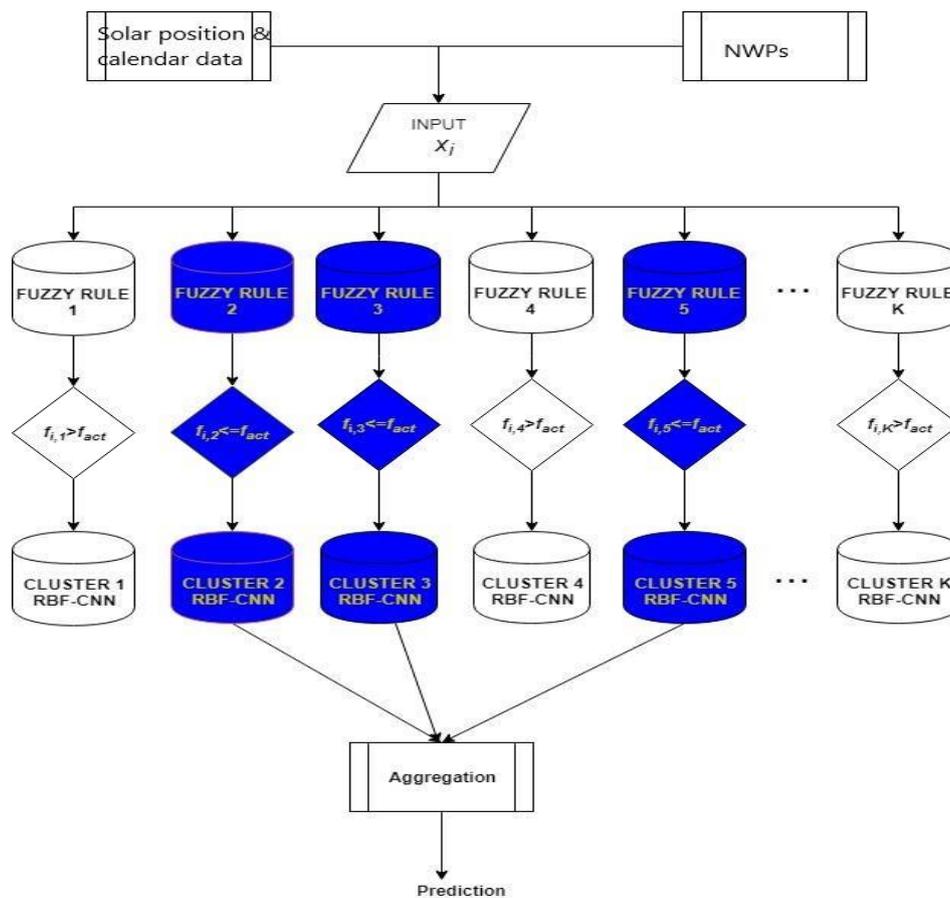


Figure 39: The forecasting model architecture (coloured branches signify the activated clusters).

As mentioned before, the forecasting model runs once per day giving the 24-hour day-ahead prediction. In order to improve the forecasting performance, a post processing model is applied. The post processing model receives the 24-hour predictions from the above forecasting model and the NWP data that cover an area 100x100 km. The NWP data from the two vendors are stacked together forming an $N \times 24 \times 10 \times 10 \times 8$ matrix. The last dimension corresponds to the NWP variables that are 4 from each vendor (downward short-wave flux, cloud coverage, temperature and wind speed). The proposed model consists of a LSTM-Conv neural network. Firstly, 24 Conv layers are applied and each of them receives an $N \times 10 \times 10 \times 8$ matrix that contains the NWPs of a different hour. Then, the corresponding hourly prediction together with their outputs feed a LSTM layer that provides the final 24-hour prediction. The post processing model is illustrated in Figure 40.

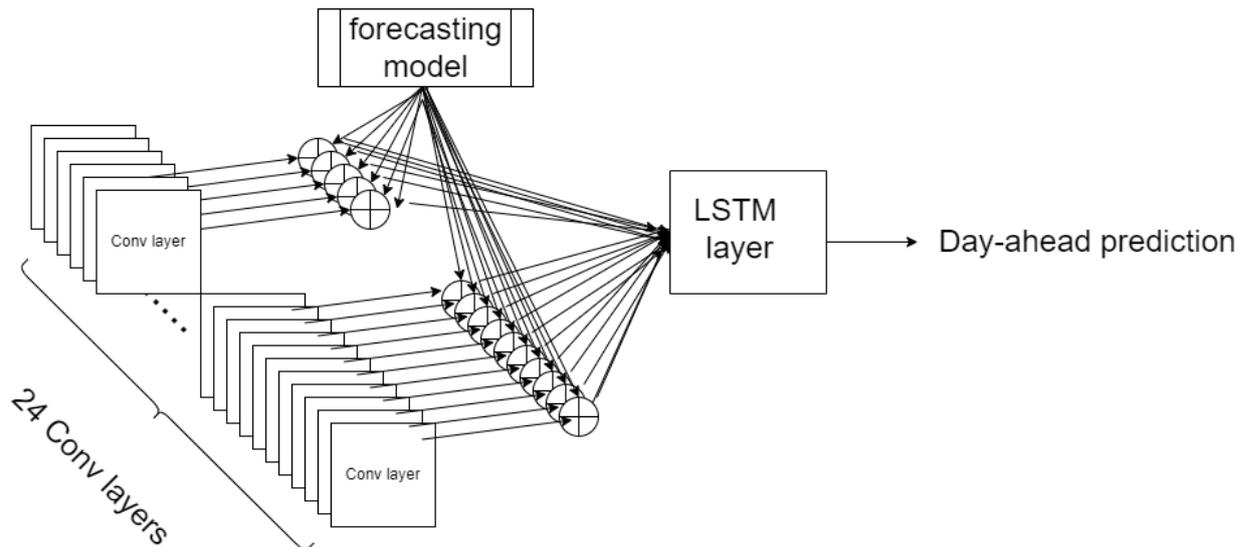


Figure 40: Post-processing solar power forecasting model.

In order to provide a further analysis on the performance of a forecasting model when using input NWPs obtained by two different vendors, an additional forecasting model called "Ensemble PDE-NN", proposed by ICCS as well, is tested. This model includes four components, as presented in Figure 41, each responsible for a different process:

Dimensionality Reduction component: This component is based on an Autoencoder architecture and is responsible for projecting the high-dimensional input data into a latent space of attributes with a lower number of dimensions. Often, input data have a great number of features, which makes the clustering process inefficient, due to the fact that the concept of distance between different vectors becomes less precise as the number of dimensions grows. Therefore, this component is responsible for providing the Clustering layer with the input data in a lower dimension, without any redundant information and noise.

Clustering component: The Clustering component is based on a clustering layer of a neural network. This layer is responsible for dividing the input data in a specific number of clusters. Each input data is assigned to a cluster based on the minimum distance of the input data vector to the centroid vector of each cluster. The centroids of the clusters are randomly initialised and in each training iteration, they are altered based on a loss function. The replacement of the centroids with a centroids' layer, through a supervised training process, gives the flexibility to the clustering process not only to just cluster the data but also choose the centroids that give a lower value to the overall loss function. The centroids' layer through the back-propagation process is updated based on the overall loss function of the model.

Sub-models Component: The Sub-models component consists of a number of neural network-based sub-models. As shown in, each sub-model acts as a different independent branch of the same neural network which produces parametric probabilistic forecasts. The goal of this component is to present to each sub-model module the same input vector and for each sub-model to produce a different output in the form of a probability distribution's parameters, which is scaled by a participation factor calculated by the Meta-learner component. The outputs of the Sub-models component are used as the input data for the final estimation in the Meta-learner component.

Meta-learner component: In each iteration of the training process, the Sub-models and the Clustering component provide the Meta-Learner with a multi-parameters vector and the clustering participation weights of the input vector, respectively. Within the Meta-Learner, there is a two-stage process for estimating the final output of the model. In the first stage, the cluster participation weights are transformed into sub-models participation weights through a fully-connected layer. The optimal sub-models' participation weights are calculated through the training process and are conditional to the cluster participation weights. The multi-parameter vector is element-wise multiplied with the sub-models' participation weights vector. The goal of the first stage is to provide the Meta-learner component the information of which sub-models should participate and to what degree, in the final estimation process. The second stage of the Meta-learner component, uses as input the weighted parameters' vector and using a fully-connected layer estimates the final forecast.

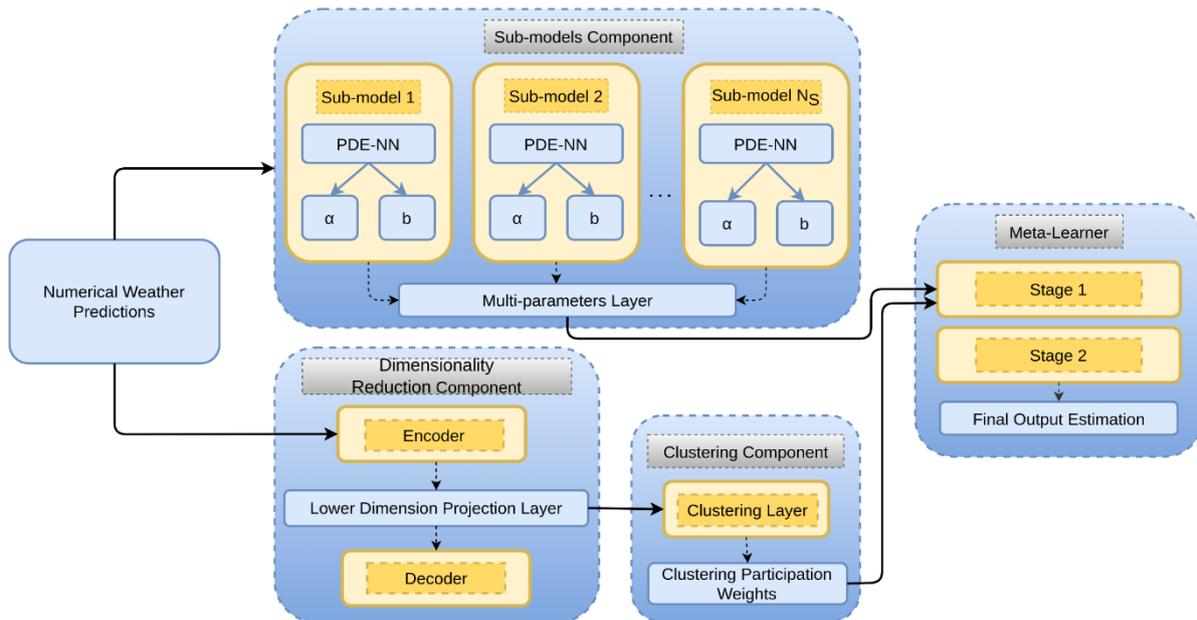


Figure 41: Ensemble PDE-NN model architecture.

9.3 Case studies

The applied forecasting models that provide the initial day-ahead solar power prediction in hourly based and the post processing model that provides the day-ahead prediction at once are evaluated in Rhodes case-study. The timeseries used contains information for the total solar power production of the island and has hourly timestep. The data used for training the models is for the period from 01/01/2018 to 31/08/2019, while the testing period starts at 01/09/2019 and ends at 31/12/2019.

9.4 Results and Discussion

In Table 16, the normalised mean absolute errors of the forecasting model and of the post processing model are illustrated for three cases: The case that only ECMWF NWP are considered, the second case the models use only GFS NWP and the third case that both ECMWF and GFS NWP are used.

Model	ECMWF	GFS	ECMWF+GFS
Fuzzy-RBF-CNN	2.62%	3.12%	2.61%
Fuzzy-RBF-CNN-LSTMConv	2.55%	2.96%	2.58%
Ensemble PDE-NN	2.51%	2.85%	2.38%
Ensemble PDE-NN-LSTMConv	2.48%	2.79%	2.05%

Table 16: Normalised mean absolute error of day-ahead predictions.

The results show that the performance of the models is better when they use the ECMWF NWP. Using GFS NWP, the NMAE is about 20% higher than the other cases. By using NWP from both vendors, the performance of the Fuzzy-RBF-CNN model remains similar to the performance with ECMWF NWP. However, the Ensemble PDE-NN model provides significant improvement when uses data from both vendors due to the dimensionality reduction component. Furthermore, for both cases the LSTMConv post-processing method is tested. As proved by the results, the LSTMConv method has the ability to improve the overall accuracy of a model, but as it displayed in Table 16 we can assume that the performance of the post-processing procedure might not have the ability to further improve the accuracy in certain conditions.

To conclude, the application of more than one vendor of NWP to solar power forecasting can lead to significant improvement and the proposed post processing model that calibrates the day-ahead predictions with their time dependence abilities reduces further the error in all cases.

References

Sideratos, G., Ikononopoulos, A. and Hatziaargyriou, N. (2020), "A novel fuzzy-based ensemble model for load forecasting using hybrid deep neural networks", Electric Power Systems Research, Volume 178

Sideratos, G. and Hatziaargyriou, N. (2020), "A distributed memory RBF-based model for variable generation forecasting", International Journal of Electrical Power & Energy Systems, vol. 120

Sideratos, G. and Hatziaargyriou, N. (2012), "Probabilistic wind power forecasting using radial basis function neural networks", IEEE Trans. on Power System, Vol 27, Issue 4, pp 1788 – 1796

Sideratos, G. and Hatziaargyriou, N. (2007), "An Advanced Statistical Method for Wind Power Forecasting", IEEE Trans. on Power System, Vol. 22, Issue 1, pp. 258-265



10 Conclusion

In this report we have presented several novel approaches developed in the framework of Task 3.1 that use multiple sources of data and improve the skill of RES power forecasts:

- PV power forecasting using the combined Satellite and AllSkyImager Irradiance Forecast (DLR, section 2) → up to 18% RMSE improvement compared to only AllskyImager forecast
- Using Skyimager data to improve the minute ahead PV power forecasting (EMSYS, section 3) → up to 23% RMSE improvement compared to operational EMSYS forecast
- Improving the minute to hour ahead PV forecast during thunderstorms by using lightning data and power measurements (EMSYS, section 4) → up to 14% RMSE improvement compared to operational EMSYS forecast
- Obtaining improved power curves through advanced outlier data filtering and machine learning techniques (EMSYS, section 0) → up to 8% RMSE improvement compared to operational EMSYS forecast
- LiDAR based second-ahead forecasting of wind power and structural loads (DTU, section 6) → up to 35% RMSE improvement compared to persistence forecast
- Investigating the capabilities of AI-based dynamic forecasting approaches and additional SCADA data channels (DTU, section 7) → up to 46% RMSE improvement of dynamic GRU networks compared to single GRU networks and persistence
- Blending satellite irradiance maps with NWP data and power time series to improve hour-ahead PV forecasting (ICCS, section 8) → up to 5% MAE improvement compared to forecast without satellite irradiance
- Applying AI techniques and combining two NWPs to improve the day-ahead solar power forecasting (ICCS, section 9) → up to 17% MAE improvement compared to single NWP forecast

Most of these approaches integrate new sources of data, that have not been used in the context of RES forecasting before, into existing RES forecasting models. It was demonstrated that all of these approaches yield a significant improvement of RES power forecasting skill in alignment with the Forecasting KPIs of the project (see Appendix A, Section 11.1.1), mostly reaching or exceeding the project KPI target values of 9-12% (solar) and 7-9% (wind) RMSE improvement for the up to 30-min ahead forecast.

The approaches contained in this report are validation studies, most of them reaching TRL 4 at this stage. In WP6 some of these will be integrated in the complete model chain of RES power forecasting and evaluated for a number of test cases defined in WP1 to demonstrate and verify the predictability improvement compared to the state of the art in an industrially relevant environment (finally reaching TRL 5-6).

With the approval of this Deliverable, Milestone M3.1 will be achieved: the developed multi-source data approaches to short-term RES forecasting are ready for testing in WP6.



11 Appendix A: Summary of Smart4RES KPIs

This Appendix summarises the KPIs defined in Smart4RES. These KPIs are divided in 2 groups:

- Project KPIs defined in the Description of Action
- Specific KPIs that are needed to evaluate technical results.

Please refer to Smart4RES Deliverable D.1.1 'Use cases, requirements and KPIs for RES forecasting', in particular Section 2.5 for a more detailed description of the different KPIs.

11.1 Project KPIs

11.1.1 Forecasting KPIs

KPI category	KPI index	KPI name	KPI baseline	KPI target
Weather Forecasting				
Project KPI	1.1.a	% absolute improvement Weather Forecasting score: 15 to 30 min ahead	Current operational solutions (AROME/ECMWF/GFS, DLR, EMSYS)	10-15% RMSE
Project KPI	1.1.b	% absolute improvement Weather Forecasting score: Few hours ahead	Whiffle forecast driven with ECMWF boundary conditions and without data-assimilation	10% RMSE
Project KPI	1.1.c	% absolute improvement Weather Forecasting score: From few hours to 96 hours ahead	Current operational solutions of MeteoFrance	10% RMSE, 4-6% CRPS (solar radiative) 5-10% CRPS (wind)
RES Forecasting				
Project KPI	1.2.a	% improvement RES Forecasting score: Up to 30 min ahead	Current operational solutions of EMSYS, EDP-R, errors from public datasets.	Solar: 9-12% RMSE, 3-5% CRPS Wind: 7-9% RMSE, 2-4% CRPS
Project KPI	1.2.b	% improvement RES Forecasting score: Up to 96 h ahead	Conditional evaluation on situations with highest forecasting errors.	Solar: 16-20% RMSE, 4-6% CRPS Wind: 12-15% RMSE, 3-5% CRPS
Specific KPI	1.2.c	% improvement Variogram score for ensemble forecasts	State-of-the-art methods for RES ensemble forecasts	>= 0
Specific KPI	1.2.d	% improvement for seamless generic forecasts	Same as KPI 1.2.a, 1.2.b	Weighted combination of targets in KPI 1.2.a, 1.2.b over lead-times and RES sources

11.1.2 KPIs on grid management applications

KPI category	KPI index	KPI name	KPI baseline	KPI target
Grid management applications				
Project KPI	1.3.a	% decrease of load shedding events in isolated power systems	Method without T5.2 tool and T5.1 storage support functions	>= 80%
Project KPI	1.3.b	% increased RES hosting capacity in MV distribution grids	Method without predictive management of flexibility from T5.3	>= 50%
Project KPI	1.3.c	Number of years in investment deferral in grid reinforcement	Traditional grid reinforcement	> 2 years
Specific KPI	1.3.e	Reduced energy curtailment of RES	Historical time series of RES production under curtailment conditions, without storage and innovative forecasting product	No target from the state-of-the-art. The objective is to minimize the KPI value
Specific KPI	1.3.g	Fulfillment of voltage limits	Traditional grid management without grid state optimization	No target from the state-of-the-art. The objective is to minimize the KPI value, computed in accordance with EN 50160
Specific KPI	1.3.h	Fulfillment of branch current limits		No target from the state-of-the-art. The objective is to minimize the KPI value
Specific KPI	1.5.a	Demonstration of a software-in-the-loop run using an example from the project	N.A.	Successful implementation of at least one Smart4RES use case as code and as a black box on a separate device
Specific KPI	1.5.b	Simulated environment including controls and interaction	N.A.	Successful interaction with the power system for at least one Smart4RES use case
Specific KPI	1.5.c	Test protocol to test for a least one potential risk	N.A.	Test of at least one potential risk of the software solution



11.1.3 KPIs on market applications

KPI category	KPI index	KPI name	KPI baseline	KPI target
Market applications				
Project KPI	1.3.d	% increase in electricity market revenue	Point forecasts and optimal quantile	10-15% decrease in costs stemming from balancing + 10-15% revenue from participation in energy plus ancillary services Up to 20-25% from VPP (RES and storage) in energy and ancillary services
Specific KPI	1.4.b	Analytic forecast evaluation by traders	Same as 1.3.d	Usual error levels observed by traders in similar conditions that chosen evaluation set
Specific KPI	1.4.c	'No-big-change' forecast evaluation by traders		
Specific KPI	1.3.f	Revenue losses per production unit due to curtailment	Historical time series of RES production under curtailment conditions, without innovative forecasting product	No target from the state-of-the-art. The objective is to minimize the KPI value



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 864337