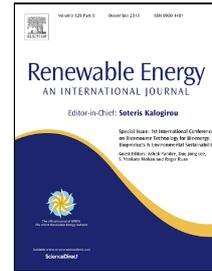


# Accepted Manuscript

Short-term and regionalized photovoltaic power forecasting, enhanced by reference systems, on the example of Luxembourg

Daniel Koster, Frank Minette, Christian Braun, Oliver O’Nagy



PII: S0960-1481(18)30955-8  
DOI: 10.1016/j.renene.2018.08.005  
Reference: RENE 10432  
To appear in: *Renewable Energy*  
Received Date: 10 January 2018  
Accepted Date: 01 August 2018

Please cite this article as: Daniel Koster, Frank Minette, Christian Braun, Oliver O’Nagy, Short-term and regionalized photovoltaic power forecasting, enhanced by reference systems, on the example of Luxembourg, *Renewable Energy* (2018), doi: 10.1016/j.renene.2018.08.005

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# Short-term and regionalized photovoltaic power forecasting, enhanced by reference systems, on the example of Luxembourg

Daniel Koster<sup>1</sup>, Frank Minette<sup>1</sup>, Christian Braun<sup>1</sup>, Oliver O’Nagy<sup>1</sup>

<sup>1</sup>Luxembourg Institute of Science and Technology (LIST) – Environmental Research and Innovation Department (ERIN)

Corresponding author: Daniel Koster; Luxembourg Institute of Science and Technology (LIST); daniel.koster@list.lu; 5, avenue des Hautes-Fourneaux; L-4362 Esch/Alzette

## Abstract

The authors developed a forecasting model for Luxembourg, able to predict the expected regional PV power up to 72 hours ahead. The model works with solar irradiance forecasts, based on numerical weather predictions in hourly resolution. Using a set of physical equations, the algorithm is able to predict the expected hourly power production for PV systems in Luxembourg, as well as for a set of 23 chosen PV-systems which are used as reference systems. Comparing the calculated forecasts for the 23 reference systems to their measured power over a period of 2 years, revealed a comparably high accuracy of the forecast. The mean deviation (bias) of the forecast was 1.1% of the nominal power – a relatively low bias indicating low systemic error. The root mean square error (RMSE), lies around 7.4% - a low value for single site forecasts. Two approaches were tested in order to adapt the short-term forecast, based on the present forecast deviations for the reference systems. Thereby, it was possible to improve the very short term forecast on the time horizon of 1-3 hours ahead, specifically for the remaining bias, but also systemic deviations can be identified and partially corrected (e.g. snow cover).

## Keywords

Photovoltaic forecasting, forecasting performance, rmse, photovoltaic integration, solar forecasting, solar energy integration

## 1. Introduction

The share of decentralized and fluctuating energy sources, such as wind power and photovoltaic (PV), is constantly increasing and will represent a major part of the future energy mix. The reliable management of our electricity supply and grids as well as the containment of increasing price volatility on the electricity market, will depend on the ability to handle these fluctuating renewable sources. The forecasting of the dynamics of PV power production is therefore crucial for the integration of high shares of photovoltaic into our energy system and market.

The different stakeholders involved in the electricity supply and operation of the grids, have their specific needs for load and production forecasting and these needs are changing with the rising shares of fluctuating, distributed generation. Electricity retailers require accurate day-ahead forecasts of PV systems (hourly resolution; updated once or twice a day) for their energy procurement and sales forecast. Since many small scale PV system feed in behind the meter of their customers, they reduce their demand and need to be considered in load forecasting. But also the utility scale PV systems have increased their share in the production portfolios and force the providers to account for them accurately in their production forecasts. The inaccuracies in day-ahead forecasts for production and demand need to be balanced out on the intra-day level, by procurement, respectively sales on the spot market. Hence, forecasting on intra-day (down to 5 minutes resolution and hourly updates) and day-ahead level is of high economic importance for energy retailers. [1] [2]

A second stakeholder is the transmission system operator (TSO), who establishes forecasts one or two days ahead (hourly resolution, daily updates) with the objective to keep demand and supply balanced and to meet the technical constraints of the grid. In order to avoid congestions, TSOs can mobilize reserves, curtail production or set other regulating measures, mainly short-term on the intra-day level. Hence, day-ahead and intra-day (5 min.; hourly updates) are also important forecast horizons for the TSOs. But, in order to run power flow simulations and identify potential congestions, the spatial variation of the PV power forecast is another aspect for the TSOs, although this can be at coarse resolution. [1] [2]

The distribution system operators (DSO), responsible for the electricity transport from the transmission grid to the final customer in mid- or low-voltage grid level, had a much more passive role in the past, as compared to the TSOs. But with the shift to distributed, fluctuating generation in our low-voltage grids, such as PV, their role is changing. Smart distribution grids, decentralized storage and demand response concepts are innovative technologies with the potential to increase the hosting capacity of the distribution grids for decentralized production [3]. But their operation and predictive control will also require accurate PV forecasting in the near future, but at a relatively detailed spatial resolution (e.g. street level).

In the light of above explained developments, the objective of this work was to develop a forecasting approach reaching a high accuracy for regional PV power forecasts on day-ahead, as well as intra-day level, meeting the requirements of the stakeholders and reflecting the availability of the necessary data. Further, the approach should allow for a high spatial differentiation of the regional forecasted PV power. The effort, in terms of necessary computational power or the set up and operation of measurement devices should remain on a manageable level for the concerned stakeholders.

The following paragraphs will give a brief overview on existing methods and how they relate to above described requirements. Existing methods for PV power or solar irradiance forecasting do exist and can be differentiated by several characteristics. It will be explained to which groups our approach belongs to and how it differs from existing methods.

Literature documents direct and indirect methods, where direct methods try to predict directly the expected PV power (mainly for single sites), while indirect methods forecast the solar irradiance and derive the PV power from this most important influence factor [1]. Direct

66 methods use mainly statistical or artificial intelligence (AI) methods and require detailed time series data of the PV site to be forecasted.  
67 This data is, specifically for regional forecasts, in most cases not available [4]. Somehow related to this differentiation, is the  
68 categorisation in methods using endogenous data only (e.g. time series measurements of PV output) and those using (additional)  
69 exogenous data, such as irradiance forecasts, for example [1].  
70

71 In order to reach a detailed spatial differentiation of our regional PV power forecast, we chose a bottom-up approach, representing all  
72 PV systems in our forecast region, without modelling each single system, as explained in section 2.4.. For this reason and due to the  
73 lack of time series data for the regional PV power production, the approach used in this paper can be described indirect (using irradiance  
74 predictions and a PV performance model), using exogenous data and modelling the regional PV power from bottom-up.  
75

76 Currently, to our knowledge, stakeholders in Luxembourg use a top-down approach to forecast the expected PV power which does not  
77 allow regionalisation of the forecasts. Hence, the bottom-up model offers a sufficiently detailed spatial resolution, but needs to be fed  
78 by irradiance forecasts which are accessible by the stakeholders and rely on data and methods which could be handled by them.  
79

80 Several methods have been proposed and are currently used to forecast solar irradiance, the most influential factor in PV power  
81 predictions, including a) statistical and artificial intelligence (AI) methods, working mainly with historical data sets of measured  
82 irradiance, b) remote sensing methods, e.g. basing on satellite images, c) numerical weather prediction models, d) local sensing and  
83 e) hybrid approaches combining the different methods. Each of them requires different input data or measurement devices or has its  
84 strengths and weaknesses depending on the size of the area to be covered and the spatial and temporal resolution to be delivered [1]  
85 [5] [6] [7] [8]. Not all of those approaches can be presented within the state-of-the-art overview of this paper and we will focus on those  
86 directly related to the approach of this study, but many review papers (see references before) do exist.  
87

88 From literature it is known, that numerical weather prediction (NWP) models, such as the European Centre for Medium-Range Weather  
89 Forecasts (ECMWF) data used here, perform best for more than 6h-ahead or day(s)-ahead irradiance forecasting, but on very short  
90 term intra-day forecasts (below 6h-ahead), other approaches might perform better. Hybrid approaches, e.g. NWP combined with cloud-  
91 motion vectors (CMV), demonstrated the potential to benefit from the strengths of various models at different forecast horizons [1] [9].  
92 Since NWP based forecast schemes in combination with a PV performance model can be relatively easy implemented by the  
93 stakeholders and reach high accuracies over a wide range of forecast horizons (above 6h-ahead), the objective was to use this  
94 approach, but to try to enhance the performance on short-term, intra-day time scale. A combination with a satellite imaging-based  
95 forecast would suit this targeted forecast horizon [10] [11], but their application requires access to satellite images and the application  
96 of elaborated methods available to research facilities but not to the stakeholders themselves (energy providers and grid operators). The  
97 hybrid approach described here, therefore alternatively uses smart metering data of PV reference systems in the region to adapt the  
98 NWP based power forecasts.  
99

100 Meanwhile, several studies used data from ground measurements (be it irradiance measurements or PV power) in different manners  
101 to improve their performance. Lorenz et al. [12] used ground measured irradiance data for a post processing of the irradiance forecasts  
102 in order to reduce bias. Also Mathiesen et al. in [13] used post processing of NWP model forecasts by ground measurement, referred  
103 to as Model Output Statistics (MOS), and obtained significant mean bias reductions. Lorenz et al. [14] also uses measured PV reference  
104 systems to upscale the regional power forecasts as well as to upscale the measured actual regional production. In [15] they  
105 demonstrated the potential of a combination of data from a NWP model data, cloud-motion vectors, PV measurements and statistical  
106 learning approaches for regional and single site forecasts, improving the forecast specifically in the short-term time frame. Marquez  
107 [16] combined cloud motion vector data derived from satellite pictures with a statistical learning algorithm (Artificial neural networks, in  
108 this case) and used ground measured global horizontal irradiation data to train the algorithm and validate the results, but not as a (close  
109 to) real-time input parameter to the forecasting model (as tested in this study).  
110

111 The use of power measurement data from nearby PV systems was also tested by some authors, either using it directly in a deterministic  
112 manner to influence the short-term forecast (very few studies do so), or feeding it into a statistical model. Lonij et al. [17] used 80  
113 residential PV systems as irradiance sensors to estimate cloud velocity in a direct manner and outperformed persistence forecasting,  
114 in a forecast horizon between 15 min. and 45 min., normally dominated by persistence forecasts.  
115

116 But most of the studies, that use measurements of nearby PV systems, feed the data into a statistical model and focus on the short-  
117 term forecast horizon (< 6h-ahead). Bessa et al. [18] used a vector auto regression framework to generate probabilistic forecasts for  
118 6h-ahead, out of time series PV measurements. In [4] Fonseca et al. compared support vector regression (SVR), trained by PV  
119 measurement data of three different degrees of details, and a very simple PV performance model, fed by NWP data. They found the  
120 SVR approaches yielding best results, but the simplicity of the PV performance model didn't allow to explore the full potential of the  
121 NWP based model. Vaz et al. [19] used a nonlinear autoregressive model with exogenous inputs (NARX) from nearby PV plants in the  
122 city of Utrecht (NL), but on a quite small geographical scale. Within this forecast horizon (below 6h-ahead), statistical methods seem to  
123 perform very well. Although, on intra-hour level, it is hard to perform better than a simple persistence forecast [1].  
124

125 Some authors try to profit from the strengths of the different approaches by a combination of existing methods. Wolff et al. [10] compared  
126 SVR and physical modelling for PV forecasting, using different input data (CMV, NWP and measurements). The best performance  
127 across a broad forecast horizon was found for a combination of all inputs, using either SVR or statistically enhanced physical modelling.  
128 This demonstrates that the deterministic use of measurement data without advanced statistical methods can still keep up with AI  
129 methods, if the full potential of the physical model is used.  
130

131 The aim of this paper is therefore to contribute to the further development of PV forecasting over a broad forecast horizon by:

- Using widely available and easily accessible irradiance forecasts from NWP models,

- Foregoing the use of MOS, since we needed to work with “original” untreated data in order to reduce errors originating from incorrect predicted cloud movements,
- feeding them into a detailed PV performance model to use physical models as far as possible,
- incorporating measurement data from nearby PV systems in a direct, deterministic manner to enhance the forecast,
- modelling the PV power from bottom-up to achieve a detailed spatial resolution of the regional power forecast,
- assessing the approach over a relatively long evaluation period of 2 years, as compared to other papers.

Unlike the other few papers using measurements from PV stations directly, within this paper we follow the approach to use the deviations of the single site forecasts for the reference systems from their measurements (see chapter 3.). Thereby we're aiming at reducing error originating from imprecise cloud movement predictions in NWP models.

Using smart meter data of PV systems to adapt the power forecasts, as in this paper, has the further advantage that, in the near future, smart meter data of PV systems will be available in high temporal and spatial resolution in many European countries, in contrast to expensive pyranometer measurements. This is due to the aim of the European Union to replace at least 80% of the electricity meters by smart meters by the end of 2020, wherever a replacement is cost-effective [20]. The roll-out of smart meters is ongoing and within the responsibility of the member states, therefore the actual progress in each country and technical details differ. Due to reasons of taxation, billing and reporting, in many EU countries, generation systems such as PV, are often measured individually and independent of the consumption of the PV system owner, which makes the data valuable for forecasting. Data is available to grid operators in many countries, although currently often with a time delay.

The structure of the paper is as follows: Chapter 2 explains the full scheme of the forecasting model, which input data is used and the applied methods for forecasting and evaluation of the performance. Chapter 3 states the idea behind the two tested concepts to adapt the forecast based on the measured deviations for the reference systems. The results are presented and discussed in chapter 4, including the performance on level of different modelling steps, after the adaptation of the forecast, as well as the upscaling on the regional level. In Chapter 5 the paper is concluded and a brief outlook on further development steps is given.

## 2. Forecasting model, data and methods

### 2.1. Description of the forecasting scheme

The approach of the PV power forecast model described here, is building up on geo-referenced irradiance and ambient temperature forecast data, from a NWP model of the European Centre for Medium-Range Weather Forecast (ECMWF) and measurement data of reference PV systems distributed over Luxembourg (Figure 1). The irradiance forecast data, which is being retrieved from the ECMWF web servers once a day, is pre-processed in order to obtain the irradiance in plane of the PV modules. This is done for a number of given PV systems that serve as references and for a matrix of predefined orientations and inclinations, representing the whole portfolio of PV systems in the country. The reference PV systems, of which measured PV power in a temporal resolution of 15 minutes is available, are distributed over the whole country of Luxembourg. A power forecast is being generated based on a set of equations, which describe the behaviour of a whole PV system depending on irradiance and temperature conditions and individual system profiles for each of the 23 reference systems, representing their technical characteristics. The predicted power of the reference systems is compared to their measured generated power, with the aim to set up a feed-back loop that enables the adaptation of the short term forecasts for the whole region, based on prediction errors of previous time steps for the reference systems.

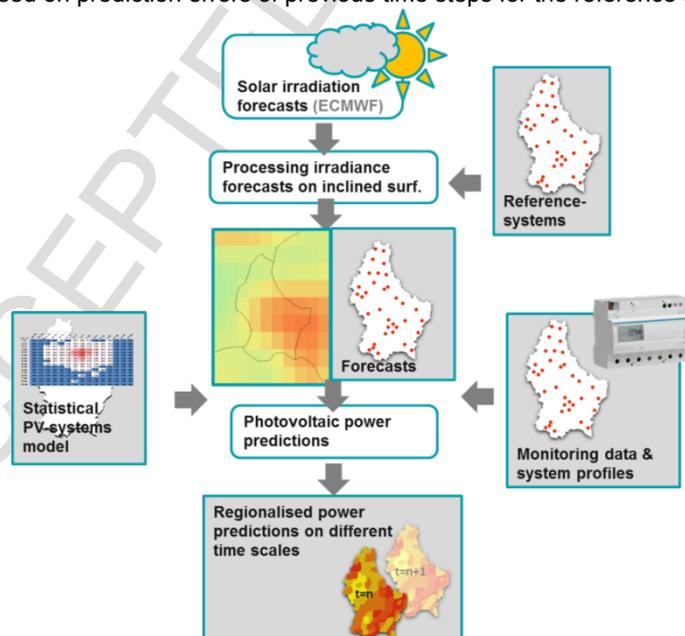


Fig 1 - scheme of the forecasting approach, combining modelling data and statistical information (left hand side) with a feed-back loop from PV reference systems (right hand side)

The methodology is explained in more detail in the following chapters, step-by-step.

## 2.2. Solar irradiance forecasts and processing

Irradiance and ambient temperature forecast data are automatically extracted from the European Centre for Medium Ranged Weather Forecasts (ECMWF), Reading (close to London). Hence, the irradiance forecasts originate from a numerical weather prediction model (NWP). The used parameter is the "surface solar radiation downwards" (ssrd) and can be considered, according to ECMWF, a reasonably good approximation of what would be measured by a global pyranometer at the earth's surface [21]. This value hence represents the global horizontal irradiance ( $G_H$ ) and is available at a spatial resolution of  $0.125^\circ \times 0.125^\circ$  as hourly values. Forecast data cover the time span of 72 hours and are being updated twice a day.

In order to derive the global irradiance on inclined surfaces ( $G_\psi$ ) from  $G_H$ , a simple but proven approach has been used, published by Olmo in [22] and validated against comparable methods in [23]. As explained in more details and providing a validation with field measurements in [24], the approach was found to be adequately reliable, for this application – also since of this model has comparably small impact on the total result, as compared to the irradiance forecasts [25].

$$G_\psi = G_H^{(-k_t(\psi^2 - \psi_H^2))} * F_c \quad \text{Eq. 1}$$

$k_t$ = Clearness index	[ / ]
$\psi$ = angle of incidence	[rad]
$\psi_H$ = elevation angle	[rad]
$F_c$ = ground reflected radiation	[ / ]

Following this approach, our forecast model calculates the  $G_\psi$  in plane of the 23 PV reference systems of known inclination and orientation. The whole portfolio of PV systems installed in Luxembourg has been classified into 57 predefined classes of orientations and inclinations, statistically representing the entirety of PV systems in the country (see 2.4). Also for those planes  $G_\psi$  is calculated.

## 2.3. Modelling of PV-reference systems

In order to calculate the expected power output of a PV system, based on the irradiance and ambient temperature forecast as main parameter and the calculated solar angle of incidence, azimuth and elevation, a model is necessary which represents the behaviour of the PV system and all its components. A set of equations and technical models are chosen which results in a rather detailed representation of the system behaviour. The below described model is used in all its details for the representation of the 23 PV reference systems and is simplified to calculate the up-scaled behaviour of the portfolio of installations in the country.

Angle of incidence reflection losses,  $IAM_B$ , are considered using a physical model published by De Soto et. al [26] and corrected in [27]. After a first implementation of another model for simplicity reasons, the so called "ASHRAE incidence modifier" model, it was found that the known drawbacks (inaccuracies at high angles of incidence), led to unacceptable results. Therefore, the physical IAM model was implemented, which works fine along the full range of possible angles of incidence, but requires assumptions or knowledge of the PV modules glass' main parameter.

$$IAM_B = \left( \frac{\tau(\psi)}{\tau(0)} \right) \quad \text{Eq. 2}$$

$\tau$ = transmittance
$\psi$ = angle of incidence
$\tau(0)$ = transmittance when normal to the sun
$\tau(\psi)$ = transmittance at incidence angle

$$\tau(\psi) = e^{-\left(\frac{KL}{\cos(\psi_r)}\right)} \left[ 1 - \frac{1}{2} \left( \frac{\sin^2(\psi_r - \psi)}{\sin^2(\psi_r + \psi)} \right) + \left( \frac{\tan^2(\psi_r - \psi)}{\tan^2(\psi_r + \psi)} \right) \right] \quad \text{Eq. 3}$$

$K$ = glazing extinction coefficient [ $m^{-1}$ ]
$L$ = glazing thickness [m]
$n$ = index of refraction of the cover glass [ / ]
$\psi_r$ = refraction angle

$$\psi_r = \sin^{-1} \left( \frac{1}{n} \sin(\psi) \right) \quad \text{Eq. 4}$$

$$\tau(0) = \exp(-KL) \left[ 1 - \left( \frac{1-n}{1+n} \right)^2 \right] \quad \text{Eq. 5}$$

Standard values for glass parameter, to be used as assumptions in PV applications, can be found in literature [26] and have been used in the model, if no specific values were known.

The PV modules efficiency, defining the part of the irradiance reaching the PV cell that is actually being transformed into DC current, is depending on the cell temperature ( $T_{\text{module}}$ ) and the irradiance in plane ( $G_\psi$ ), reduced by the reflection losses. In its current state, the used approach is a simple approximation of the modules temperature, based on [14] and own simplifications.

$$T_{module} = T_a + G_{\psi} * \gamma \quad Eq. 6$$

$T_a$  = ambient temperature [°C]  
 $\gamma$  = 0.02 (free standing) or 0.056 (BIPV) [ / ]

$$\eta_{Tm} = \eta_{STC} * (1 - (-\frac{K_T}{P_{MPP}} * (T_{module} - 25^{\circ}C))) \quad Eq. 7$$

$\eta_{Tm}$  = efficiency at operation temperature [%]  
 $\eta_{STC}$  = efficiency at standard test conditions [%]  
 $K_T$  = temp. coefficient for module power [W/°K]  
 $P_{MPP}$  = nominal power of the module [W]

Degradation and mismatch losses can both substantially reduce the yield and actual power of PV systems. Specifically, when working with data on nominal power of the entirety of the PV systems in a region, deductions considering those effects should be made. Even the detailed knowledge on some of the reference systems used in our approach does not allow the system-specific consideration of these effects. Therefore, different lump sum factors have been chosen to take these effects into account. Degradations of the modules performance, due to the light-induced degradation, is known to be higher in its first year and, in most cases, stabilizes in the following years [28]. This is valid for crystalline PV cells, while other cell technology can statistically be neglected for the area of Luxembourg. First year degradation losses have been chosen based on [29], [30], [31]. The long-term degradation losses used in our study are based on an analytical review done by Jordan et. al [28].

$c_{degr\ 1st}$  = degradation losses, 1<sup>st</sup> year = 2.5 [%]  
 $c_{degr\ f}$  = degradation loss, following years = 0.5 [%/a]

Mismatch losses can be caused by different effects and are referred to on module- as well as array level. The relevant mismatch effects for this study are those caused by deviations in the performance characteristics of modules of the same nominal power, operated in series within an array. Based on literature values [32], [33] and own judgement, these losses are also taken into account by a simple lump sum.

$c_{mm}$  = mismatch losses = 2.5 [%]

Wherever the level of detail of information on the reference system allows it, the consideration of wiring losses is system specific: Wiring losses in between the PV module strings and the inverter, hence on direct current level (DC), as well as between the inverter and the point of injection (AC level), are calculated based on cable sections and cable lengths for the PV arrays nominal power (MPP). This value is considered when the PV array is operated at MPP, while part load behaviour is taken into account with this simplified approach, here documented for the example of the DC level – AC level is done accordingly:

$$c_{DCwire} = (m_{part\ load})^2 * c_{DC\ MPP} \quad Eq. 8$$

$c_{DCwire}$  = factor, DC losses at operating cond. [%]  
 $c_{DC\ MPP}$  = factor, DC losses at MPP [%]  
 $m_{part\ load}$  = part load operation mode [%]

$$m_{part\ load} = \frac{P_{mod}}{P_{MPP}} \quad Eq. 9$$

$m_{part\ load}$  = part load operation mode [%]  
 $P_{mod}$  = power on level of the PV module [W]  
 $P_{MPP}$  = nominal power of the PV module [W]

Inverter efficiency is also changing with the current part load mode of operation. The reference systems description of our model contains characteristic points of the efficiency curve of the inverter in part load mode (see Table 1). Depending on the part load operation mode, the part load efficiency of the inverter can be interpolated.

$$\eta_{inv} = f(m_{part\ load}) \quad Eq. 10$$

$\eta_{inv}$  = part load efficiency of the inverter [%]

$m_{part\ load}$ [%]	5	10	20	40	60	80	100
$\eta_{inv}$ [%]	96.7	97.1	97.9	98.3	98.4	98.3	98.0

Table 1 – example: efficiency curve of an inverter with a European inverter efficiency of 98% [34]

This modelling approach, as presented in chapter 2.3, is only briefly described since more details can be found in [24].

A similar approach, as described above for the reference systems, is followed for the modelling of the performance of the entirety of PV systems in the country, but at a reduced level of detail, since some parameter are replaced by standard values (see Table 2).

parameter	value [unit]	description
$\eta_{STC}$	17 [%]	mean module efficiency at STC
$K_T$	-0.5 [W/°K]	temperature coefficient for module power of 1 kW <sub>p</sub>
$A_{kWp}$	6.7 [m <sup>2</sup> /kW <sub>p</sub> ]	mean surface demand for 1 kW <sub>p</sub> module power
$C_{DC\ MPP}$	0.5 [%]	cabling losses (AC as well as DC)
$C_{AC\ MPP}$		

Table 2 - standard values for model parameter representing the entirety of PV systems in the country

#### 2.4. Statistical representation of PV systems on national scale – Luxembourg

In order to be able to estimate PV power forecasts for the whole territory of a country, a region or a city, statistical information on the entirety of PV systems are necessary. As the forecasts should be regionalized (addressing future challenges of integration of high PV shares into our grids) and as irradiance conditions vary over the whole forecasting area, the nominal power and location of the individual PV systems are required. Further, for a time-discrete forecasting, orientation and inclination of the installations are of importance.

In the case of Luxembourg, nominal power and location of the PV systems is known by the grid operators and has been provided for research purposes, in anonymized form. The available data set, provided by two main grid operators, covers 111 [MW<sub>p</sub>] of the 116 [MW<sub>p</sub>] PV installed at the end of 2015 [35].

Specific data on orientation and inclination of individual PV systems is currently not registered, neither by utility companies / energy providers nor by grid operators. This means that precise data on the entirety of installations in the country is not available. Nevertheless, the Administration de l'Environnement (AEV) has a dataset on subsidized photovoltaic installations in Luxembourg which covers, according to the applicable regulation, mainly small scale systems and partially contains erroneous data. After a plausibility check and cleaning of the data, a data set of 37.9 MW<sub>p</sub> has been retained and analysed, representing 32.6% of the installed power in Luxembourg. Thereby, the distribution of orientation and inclination of PV systems is represented and can be statistically applied in the model to a set of PV systems, assuming that this distribution remains constant across the country. As long as the spatial resolution of the model remains relatively rough, the amount of PV systems in a grid cell remains high enough to consider this a valid assumption.

elev/orient	+180°=150°	-150°=120°	-120°=90°	-90°=60°	-60°=45°	-45°=30°	-30°=15°	-15°=15°	15°=30°	30°=45°	45°=60°	60°=90°	90°=120°	120°=150°	150°=180°
0°=5°	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
5°=10°	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.6%	0.3%	0.1%	0.0%	0.1%	0.0%	0.0%	0.0%
10°=15°	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.3%	1.6%	0.1%	0.4%	0.2%	0.0%	0.0%	0.0%	0.0%
15°=20°	0.0%	0.0%	0.0%	0.4%	0.4%	0.9%	1.4%	3.3%	1.7%	1.9%	0.5%	0.4%	0.0%	0.0%	0.0%
20°=25°	0.0%	0.0%	0.0%	0.3%	0.7%	1.4%	5.9%	8.1%	3.7%	2.9%	0.9%	0.5%	0.3%	0.0%	0.0%
25°=30°	0.0%	0.0%	0.0%	0.1%	0.4%	1.4%	4.6%	7.7%	4.5%	2.3%	0.8%	0.6%	0.1%	0.0%	0.0%
30°=35°	0.0%	0.0%	0.0%	0.3%	0.2%	1.4%	3.4%	9.7%	3.6%	2.1%	0.7%	0.5%	0.1%	0.0%	0.0%
35°=40°	0.0%	0.0%	0.0%	0.2%	0.3%	0.7%	1.4%	4.6%	1.0%	0.7%	0.8%	0.6%	0.2%	0.0%	0.0%
40°=45°	0.0%	0.0%	0.0%	0.1%	0.1%	0.1%	0.5%	1.5%	0.4%	0.3%	0.2%	0.1%	0.0%	0.0%	0.0%
45°=50°	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.6%	0.1%	0.2%	0.2%	0.0%	0.0%	0.0%	0.0%
50°=55°	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
55°=60°	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
60°=70°	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
70°=80°	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
80°=90°	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Fig 2 -- PV power distribution [%] by orientation and inclination of PV installations in Luxembourg (based on data from AEV)

#### 2.5. Characterisation of PV-reference systems and introduction of “synthetic system profiles”

The reference PV installations currently implemented in the model, have been chosen from a list of 173 systems, equipped with smart meters, which are able to deliver production data every 15 minutes and were chosen due to different requirements. Owners needed to agree on the collaboration, the PV systems have to be unshaded and of a relative simple design (e.g. not too much different orientations), detailed information on the technical setup is required, the systems need to work seamlessly – just to name a few. Furthermore, the spatial distribution over the forecasting area should be balanced. To that aim, systems were pre-sorted based on the available information by desk-audit. Owners of pre-selected systems were contacted and on-site visits were conducted wherever possible. This assured the high level of quality of information on the individual reference system. On the other hand, these conditions weren't met by a large amount of potential reference systems.

Finally, the lack of information on suitable PV systems led to a two-fold approach: 1) detailed reference systems fulfilling the criteria mentioned above, which are modelled according to the approach described under 2.3, and 2) less detailed reference systems, which are modelled with a similar approach, but using standard values wherever specific data is missing. The secondly described systems are further referred to as “synthetic system profiles”.

#### 2.6. Adaptation and calibration of PV-reference system profiles

After the completion of the model and the choice and characterization of the set of reference systems, the forecasting system went through a first run of the model in order to evaluate the suitability of the model and the chosen or acquired parameter for the individual reference system. The techno-physical model represents the theoretical behaviour of the reference system. Hence, a calibration of the

individual systems parameter might be necessary to better reflect the real performance of the reference system. As irradiance measurements on site of the reference systems were in almost all cases not available, it was not possible to calibrate each reference system independent of the forecast data. But since irradiance forecast are known to be relatively precise on days of clear sky conditions [9] [14], each forecast of the reference systems has been compared with the measurement curve on specific days (in March and July 2014) for which forecast and real measurement showed cloudless conditions. Overlaying both curves (forecast and measurement) revealed the deviations of the forecast. By adapting the reference systems model parameter, the forecast curve can be adapted in three directions, to better fit the measured values. The shape of each curve can be influenced a) in its height (by calibration factor mainly) or b) in its width (by adapting the inclination) and can be c) shifted e.g. towards earlier hours (by turning the orientation angle eastwards). Adaptations in orientation and inclination have been used scarcely. The calibration factors have been chosen after an analysis of the relative monthly error ( $\epsilon_{M dt}$ ), considering day time values only, for the months March and July 2014.

The analysis reveals, as expected, that the model generally overestimates the expected PV power. All calibration factors for the 23 reference systems were below one, ranging from 0.88 minimum up to 0.99 maximum and an average of 0.94.

The effect of applying these calibration factors and adapting the systems parameter to better fit the curves for clear sky days, has been evaluated and documented under 4.1.

## 2.7. Introducing evaluation criteria

The accuracy of the full chain of the forecast model, from the irradiance forecast to the point of injection behind the inverters, can be evaluated by its comparison to the actual measured production for each PV reference system. The fully up-scaled forecast for a full forecasting area cannot be evaluated as such, since no discrete time measurements are available. Hence, the basis for the evaluation are hourly forecasts of the 23 PV preference systems. In order to be comparable to other evaluations of forecasting approaches, the following evaluation criteria have been chosen in analogy to the literature [6] [9] [14].

To evaluate the accuracy of the forecasts for the individual PV reference systems, each hourly value has been compared to the measurement value. The error has been normalized to the nominal power of each reference system and is given as:

$$\epsilon(t) = \frac{P_{pred}(t) - P_{meas}(t)}{P_{nom}(t)} \quad Eq. 11$$

$$\begin{aligned} P_{pred} &= \text{predicted power of the PV system} \quad [\text{kW}] \\ P_{meas} &= \text{measured power of the PV system} \quad [\text{kW}] \\ P_{nom} &= \text{nominal power of the PV system} \quad [\text{kW}_p] \end{aligned}$$

The root mean square error (RMSE) is a common term in the evaluation of forecasting algorithms for solar irradiance [9] as well as for power forecasts in wind and solar. RMSE is considered suitable for power predictions in utility companies, since large errors are disproportionately problematic in those applications, as stated by [14].

$$RMSE = \frac{1}{\sqrt{N}} \sqrt{\sum_{t=1}^N \epsilon(t)^2} \quad Eq. 12$$

The mean value of the error (bias) is further interesting to evaluate the performance and to identify systemic errors in the forecasts:

$$bias = \frac{1}{N} \sum_{t=1}^N \epsilon(t) \quad Eq. 13$$

Another important aspect in the evaluation of solar power forecasting is the handling of night time values. Irradiance forecast and real production are zero during the night. Thus, forecast and measurement completely fit and the error is zero. If night time values (hourly errors) are taken into account when estimating the evaluation criteria (often common practice), the results show better performance obviously, only by trivial night time forecasts. For this reason it has been decided to evaluate both, the performance including night time values and taking only day time values into account. The evaluation criteria are marked with the suffix "dt" if it considers "day time" values only:

$$\begin{aligned} RMSE_{dt} &= \text{root mean square error, day time values only} \\ bias_{dt} &= \text{bias, considering day time values only} \\ \epsilon_{M dt} &= \text{monthly normalized error, day time values only} \end{aligned}$$

The normalized error  $\epsilon$  for each hour, as the base value for the other evaluation criteria, can potentially be larger on days of high solar irradiance and thus high PV power. Hence, the mean power "mean P" within a certain time span is an important reference value. Also "mean P" is normalized to the nominal power for reasons of comparability.

$$\begin{aligned} \text{mean P} &= \text{mean PV power of a system within period} \\ \text{mean P}_{dt} &= \text{mean PV power of a system within period, considering day time values only} \end{aligned}$$

## 3. Theory of feedback loop concepts for error reduction

PV power forecasting is relying on the accuracy of the solar irradiance forecasts and can thus, if solely based on those predictions, never be more accurate than the underlying meteorological forecast. Although, the solar radiation forecasts improved a lot during the last decade, they are still the main source of uncertainty, as the physical description of the PV systems is comparably straightforward.

408 In solar radiation forecasting by numerical weather prediction models (such as the forecasts delivered by the ECMWF), mathematical  
409 equations describe processes in the atmosphere and the models are being fed with measured parameter from the recent past and  
410 current observations. The models predict near future developments in the atmosphere, such as cloud movements and cloud formation  
411 on different heights, which influence strongly the solar radiation reaching the earth's surface. These cloud movements and  
412 transformations on several heights need to be precisely predicted in their speed, direction and thickness, in order to estimate their  
413 regional effect on solar radiation. On larger areas, inaccuracies in cloud movements balance out more strongly and thus the accuracy  
414 of the forecast increases with a coarser spatial resolution [6] [7].

415  
416 If such inaccuracies in the irradiance predictions are not of a purely random nature, but would be, e.g. due to inaccuracies in forecasting  
417 of cloud speed or direction, the error could persist over a short time period – in our case a few hourly time steps. Hence, the forecasting  
418 error could be reduced by the estimation of the current inaccuracy and its projection into the near future. Following this assumption, our  
419 approach tries to adapt the purely model-based forecasts by a feed-back loop from PV-reference systems. Two approaches have been  
420 followed in order to create this feedback loop based on online PV power measurements:

### 422 3.1. Error persistence method

423 First analysis of a comparison of the forecasts for single PV systems to the measured power, did show persisting trends (over few  
424 hours) of over- or underestimation of the real power. Although this is not generally the case, there was enough evidence to test whether  
425 a correction of the forecast, 1h to a few hours ahead, based on the assumption of a persisting error, would increase the accuracy of  
426 the power forecast.

427  
428 In order to adapt the forecast for a specific PV system, the error  $\epsilon$  for each time step was calculated and the deviation from the real  
429 measurement was considered to be persistent over a certain time range:

430 Example: Forecast adaptation by 1h error persistence for the time  $t_0$

$$431 \quad P'_{fc\ t_0} = P_{fc\ t_0} + \epsilon_{abs\ t-1} \quad Eq. 14$$

432  
433  $P'_{fc\ t_0}$  = adapted power forecast for  $t_0$

434  $P_{fc\ t_0}$  = original power forecast for  $t_0$

435  $\epsilon_{abs\ t-1}$  = absolute error for  $t_{-1}$

436  
437 Specifically after noon, when the maximum possible power production is declining hour by hour, this could result in an adaptation of  
438 the forecast above the theoretical possible maximum. To avoid this, the theoretical clear sky irradiance for each time step and system  
439 orientation has been calculated, the resulting maximum power production has been estimated and the forecast adaptation was limited  
440 to this theoretical value (results see 4.4).

### 443 3.2. Error movement vectors

444 The assumption behind this approach is, that a main source of error in solar irradiance forecasting by NWP models arises from  
445 inaccurate forecasting of clouds and cloud movements (direction and/or speed) or thickness [9]. A cloud front moving into the forecast  
446 area over time might thus lead to over- or underestimations along its front-border, depending on whether it is moving faster or slower  
447 than predicted (or inaccurate in direction or less/more opaque). The errors, which can be derived for the single PV systems used as  
448 references, could thus be visualized on maps and might show graphical patterns propagating over the forecast area with time. If clear  
449 movement patterns could be identified, existing methods used to predict cloud movements [17] [36], could be used to forecast the  
450 propagation of error movements. The validation of this hypothesis and the evaluation to which extent a forecasting of these error  
451 patterns is possible and leads to more accurate predictions, was one of the main drivers behind this work (results see 4.5).

## 453 4. Results and Discussion

454 The technical and physical model, reflecting the irradiance data processing and the individual system behaviour of the PV-systems as  
455 described under 2.3, has been assessed by feeding measured irradiance data into the model and comparing the calculated power  
456 production to the smart meter data. The results of this evaluation were very promising and have been published in [24].

### 458 4.1. Efficiency of calibration factors and adaptation of system profiles

459 As described under 2.6, the technical and physical model for the PV systems describes their optimal functionality and assumes the  
460 accuracy of the given data, e.g. on orientation and inclination of the modules. As the PV reference systems are conventional, "real-  
461 world" systems, deviations from the optimal functionality are to be expected and will be calibrated for. The calibration factors and  
462 adaptation of inclination and orientation of the systems parameter, as explained above, has proven to be very effective. The calibration  
463 led to a reduction of the monthly normalized error for all reference systems, with only few exceptions. The improvement ranges from a  
464 0.46% to a 8.5% difference for the examined time period. On average, the calibration reduced the error by 2.66% at a mean deviation  
465 before the calibration of -5.96% - hence, the mean deviation after calibration was -3.3%.

### 467 4.2. Performance evaluation of the forecasts on reference system level

468 Assessing the actual accuracy of the PV power forecast, the focus is on the hourly performance of the forecasts as an important aspect  
469 for the grid operators and utility companies. The accuracy on forecasting the PV production on monthly sums or daily sums is obviously  
470 higher, as compared to hourly values. Furthermore, the accuracy of forecasts for specific single sites (single PV systems, such as the  
471 reference systems in our approach) is lower than for regional forecasts, as in larger forecast areas local phenomena (e.g. cloud  
472 movements) can level out [6] [7] [9] [14].

Although the final output of this PV power forecast model described here are regionalized forecasts, only accuracy on single site forecasts are evaluated. The simple reason for this is the lack of hourly measurement data for the entirety of PV systems in the forecasting region to which we could compare our forecasts. Even though the calculations have been done for three forecast horizons (0-24h, 24-48h, 48-72h), the focus of the assessment is on the intra-day forecast, 0-24 h ahead.

### Performance averaged on monthly basis

In order to evaluate the performance of the forecasting model, the error on hourly forecast values  $\epsilon$  for each reference system, as compared to the measured value, has been calculated and normalized to its nominal power (considering the intra-day forecast, 0-24h ahead). The hourly, normalized error  $\epsilon$  is evaluated on a monthly basis over 2 years, 2014 and 2015.

In Fig 3 the performance evaluation criteria for reference system nr. 0067 are exemplarily depicted. RMSE and bias, related to the left axis, illustrate the accuracy of the forecast for the respective reference system and can be set in relation to the mean power for each month (right axis).

Mean monthly power, normalized to the nominal power of the system, gives the average power the system delivered in the respective month. The value is given in order to set the other evaluation criteria into relation with the mean power, as deviations from the forecast could be larger in months with relatively high irradiance. The system shown here has a typical curve, as compared to the other reference systems. The system reaches its peak production when "mean P" lies around 20% of its nominal power (30% if only day time values are considered). On first sight this seems relatively low, but as the systems reach their nominal power only a few hours a month, even during summer, this is a normal value that can be validated by literature.

The "bias" evaluates the actual mean error of the forecast, without specific weighting. A low bias means that there is low systematic error in the forecast – the system is neither over- nor underestimating the actual PV power constantly. Anyway, there can be large deviations in the single hourly forecasts that might compensate each other and are not visible in bias only. Over the two years, the monthly bias of this exemplary PV system forecast ranges from 2.7% to -1.0% - a representative value for the set of reference systems. Bias shows no clear seasonal deviation over the two years, which can be confirmed by the other systems. If curves of the different systems are compared, similar bias curves can be observed. This hints to a bias originating from the irradiance forecast for the specific months and confirms the suitability of our model throughout the seasons.

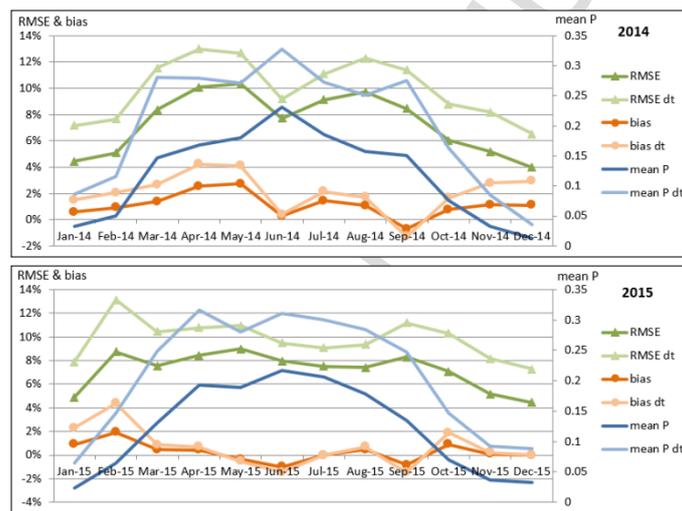


Fig 3 - evaluation criteria of the hourly performance for the two years 2014 and 2015 (here for reference system nr. 0067)

The root mean square error RMSE represents a mean error, weighting larger deviations much stronger than small deviations. The RMSE shown in Fig 3 ranges from values around 4% in January and December, up to 10% in April/May 2014 (for night- and day time values). The graphs show representative curves for the set of reference systems, generally increasing during months of high solar power, as RMSE is specifically sensitive to large deviations which occur more frequent in this period.

Obviously, in February 2015 the RMSE shows an atypical increase for this period, which is not related to a technical problem on this specific reference system, as the same increase can be observed for the other PV systems. This effect is due to snow cover in February 2015 (confirmed by the national meteorological organisation MeteoLux). Hence, the solar irradiance forecast is predicting the irradiance independent from the snow cover, but the PV-systems throughout the country underperform due to snow cover. This is a well-known weak point of PV power predictions based on irradiance forecasts only.

Comparing all reference systems over the two years (Fig 4), the mean performance in terms of RMSE and bias is relatively similar with few exceptions. The mean bias over all systems is 1.1% ( $\text{bias}_{dt} = 2.2\%$ ), while the values might range from -0.12% (nr. 1075) up to 2.45% (nr. 1134). Generally, the bias is positive in the range of 1%, which means an overestimation of the systems expected PV power. Considering bias, there are no extreme exceptions from that trend, but it will be checked if (for some systems, e.g. nr. 1134) a stronger calibration factor might reduce the bias.

On average, the RMSE over all systems lies at 7.4% ( $\text{RMSE}_{dt} = 10.0\%$ ) and ranges from 6.00% (nr. 1159) to 14.09% (nr. 1173). Except for system nr. 0138 and nr. 1173, all systems perform comparably similar and their RMSE lies around 6.9%. The reason for the comparably high RMSE for the two outlying systems is yet unknown but further analysis is ongoing.

526  
527  
528  
529  
530

In comparison to other literature data for the accuracy of single site forecasts (not regional forecasts), the values given above seem reasonable and the model seems to work comparably well (see e.g [1] [9] [8]). Although, a simple direct comparison is not very reasonable, since the performance evaluation throughout the studies is done on different basis for normalisation, using or not considering night values, under versatile climatic conditions and for varying testing periods.

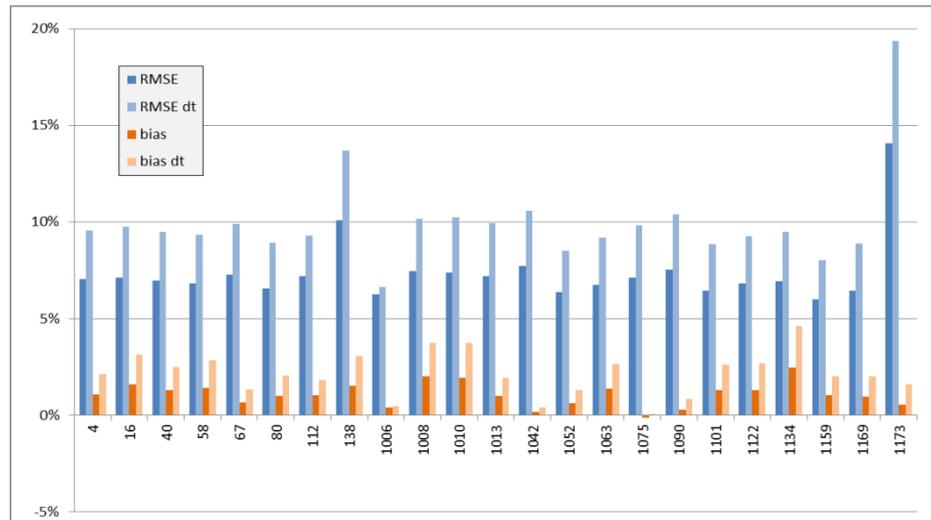


Fig 4 - mean evaluation criteria over 2 years (2014 & 2015) for the reference systems (numbers below is the internal numbering of the reference systems)

531  
532  
533

**Hourly performance & averaged on daily basis**

Although RMSE and bias give already a good impression of the accuracy of the forecast over a larger time scale, only the daily forecast curves give real insight in the daily performance. Therefore, similar plots as Fig 5 have been created for all reference systems and every day in 2014 and 2015.

538

The plots of Fig 5 exemplarily show six days in July 2014 (01.07 – 06.07.) and their respective curves for the three forecast horizons (red/orange/yellow) and the measured production (grey) of system nr. 0067. Obviously, the forecasts fit relatively well the real production on clear days (02.07. & 03.07.). Larger deviations occur on overcast days – this observation correlates with the reported accuracy of the irradiance forecasts from literature.

543

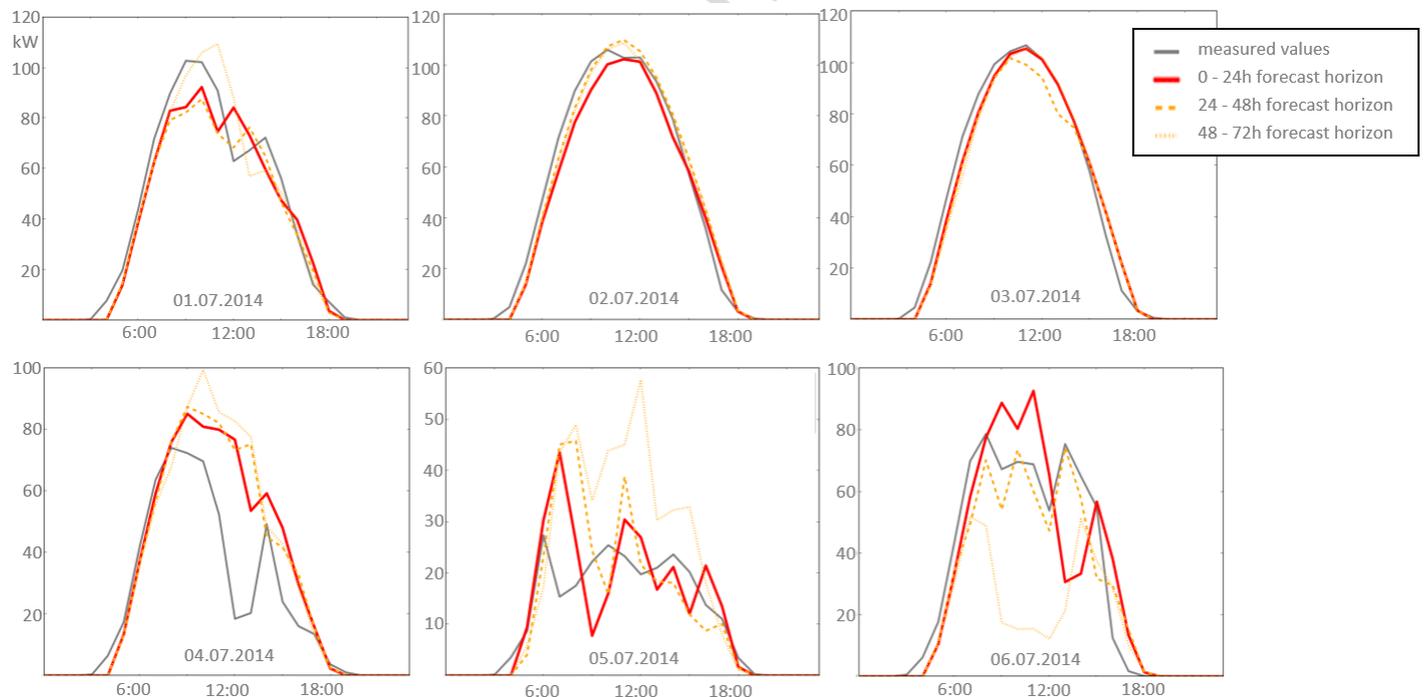


Fig 5 - example for system nr. 0067, six days in July 2014, showing the correlation of the three forecast horizons (0-24h in red solid line / 24-48h in orange dashed line / 48-72h in yellow dotted line) and the measured values (grey line)

544  
545  
546

The plot for 05.07.2014 shows a rather cloudy day, resulting in relatively large hourly deviations from the real power, although the mean production fits well. But as the mean power on such cloudy days is relatively low, the normalized error (see Fig 6) remains in an acceptable range.

550

The boxplot in Fig 6 illustrates the normalized error  $\epsilon_{dt}$  and its variation on the hourly value. The grey boxes for each day depict 50% of the forecast values around the median. The thin lines above and below the box show the upper, respectively lower 25% of the single

552

values. Boxplots allow to give a quick overview on the quality of the hourly forecasts over a full month – they illustrate the bias as well as the scatter of the majority of the values and the extreme outliers. These plots have been evaluated for each month in 2014 and 2015 for each system.

It can be seen that for the large majority of hourly forecasts, the normalized error lies within a range of +/- 10%. But, single hourly forecasts can, in extreme cases, deviate from the real power in a range of more than 50% of the nominal power.

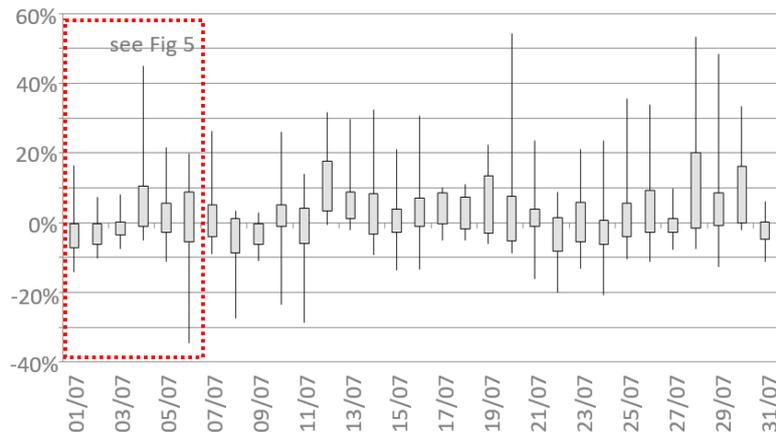


Fig 6 - Boxplot of the normalized error  $\epsilon$  of the hourly forecast for reference system nr.0067 for July '14

4.3. Comparison of the forecast performance for synthetic profiles of reference systems and different forecast horizons  
As explained under 2.5, due to lack of detailed information on some reference systems, the concept of synthetic profiles has been introduced. As these system profiles work with standardized and estimated parameters, the synthetic reference systems were expected to have larger deviations to their actual measured power. But surprisingly, the evaluation of the forecast quality for the reference systems shows no significant difference in their performance: Fig 4 indicates the 2-years performance for all reference systems – the ID numbers above 1000 are synthetic profiles.

Further comparisons have been made between different irradiance forecast horizons. Since the irradiance forecast covers 72 hours and the forecast has been retrieved from the server once a day, the data covered an intra-day forecast, a 1-day-ahead and a 2-days-ahead forecast. Comparing the performance of our PV power forecast over the three forecast horizons resulted in relatively small differences. Over the two years data set, the intra-day forecast fitted best the measured power production, as expected, but the differences to the other two forecast horizons were in a range of 0.7% for the normalized mean error, only.

#### 4.4. Performance evaluation of the history-based forecast adaptation – “error persistence”

The analysis of the error  $\epsilon$  over time for single reference systems, revealed that for a considerable amount of days, the forecasts tended to repeatedly over- or underestimate the real power for a time span of several hours. This led to the development of the approach of error persistence explained under 0. This approach was not expected to deliver appropriate forecast adaptations for longer time spans, but was tested for 1 to 4 hours ahead.

As depicted in Fig 7, the 1-hour ahead forecast adaptation decreases the deviations from the measured value significantly (in this example). The 2-hours-ahead forecast adaptation is already performing considerably worse. Obviously, the approach works in cases of continuous under- or overestimation of the PV production, but can even be counterproductive if the deviations are fluctuating between positive and negative values.

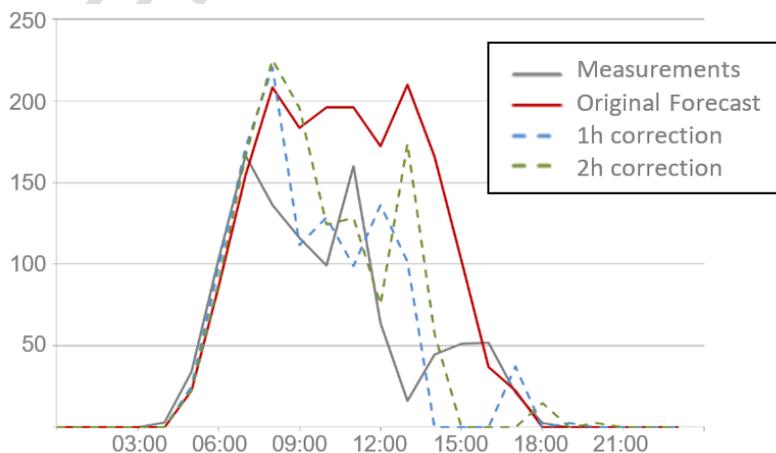


Fig 7 - Example: 1st of May 2014, reference system Nr. 0067, showing the adaptation of the forecasts (red line), based on previously measured deviations (1 hour into the future (blue) / 2 hours into the future (green)) and the deviations from the measured values (grey line)

The performance of the approach needed to be evaluated over longer periods: Shown in Fig 8 are the results for a forecast adaptation based on 1 hour- and 2 hour-error persistence. With this simple approach, it was possible to reduce bias very effectively, which could have been expected. Both, the 2 hours ahead and 1 hour ahead error persistence adaptation reduced the bias considerably. Evaluated over the two years, exemplarily for system 0080, the bias dropped from 1.00% (2.05% for bias<sub>dt</sub>) to 0.14% (0.27% bias<sub>dt</sub>) for the 1 hour ahead error adaptation and to 0.40% (0.74% bias<sub>dt</sub>) for the 2 hours ahead adaptation.

RMSE is less well improved, since this simple approach does reduce systematic error as well as short term persistent over- or underestimations, but does not reduce outliers which influence RMSE to a greater extent. Anyway, for the 1 hour ahead adaptation, mean RMSE over the full two years does decrease from a value of 6.57% (8.95% for RMSE<sub>dt</sub>) to 5.81% (7.62% for RMSE<sub>dt</sub>). The 2 hours ahead forecast adaptation based on error persistence did not generally improve the RMSE. Fig 8 shows months with lower RMSE as well as higher RMSE in other months for the 2 hours ahead adaptation. For longer time periods, 3 hours ahead or 4 hours ahead forecast adaptations, the approach didn't result in any improvement of the forecast.

The very short term error adaptation based on error persistence is hence able to reduce systematic error – which is obvious. This effect is visible in Fig 8 for February 2015. The strong increase in bias and RMSE for Feb.'15 of the original forecast is due to snow cover on the PV modules, which is not represented in the model. This systematic overestimation by the forecast model is being effectively compensated by the error persistence adaptation – which could be a suitable application for this approach.

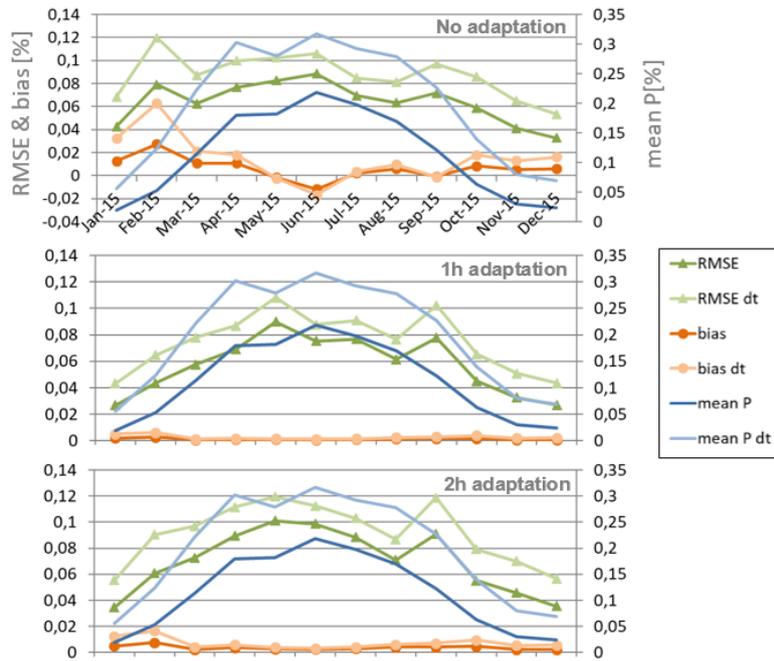
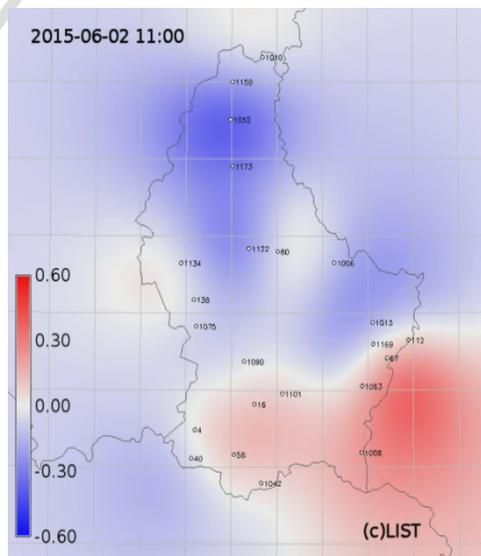


Fig 8 - Evaluation criteria on forecast accuracy for system Nr. 0080 in 2015 without adaptation based on error persistence (top), based on 1h error persistence (middle) and 2h error persistence (bottom)

#### 4.5. Performance evaluation of the history-based forecast adaptation – “error movement vectors”

In order to assess the possibility of identifying error movement vectors (as described under 3.2), the individual normalized error  $\epsilon$  for each time step needed to be estimated and visualized on a map. The data points representing the error at each reference system were calculated and spatially referenced, while the points in between were interpolated. For each hour of the two years under survey, a map similar to Fig 9 has been established.



The error map above (Fig 9) shows the distribution of deviations over the region. The forecast did, for the PV reference systems in the South and South-East of Luxembourg, overestimate the PV power (red), while for the North and middle of the country, the power was underestimated (blue).

The individual hourly error maps are sequentially concatenated to create a video sequence to analyse the changes of the occurring error distribution over time. Monthly video sequences were screened in order to evaluate the possibilities to identify and track error movements on specific days.

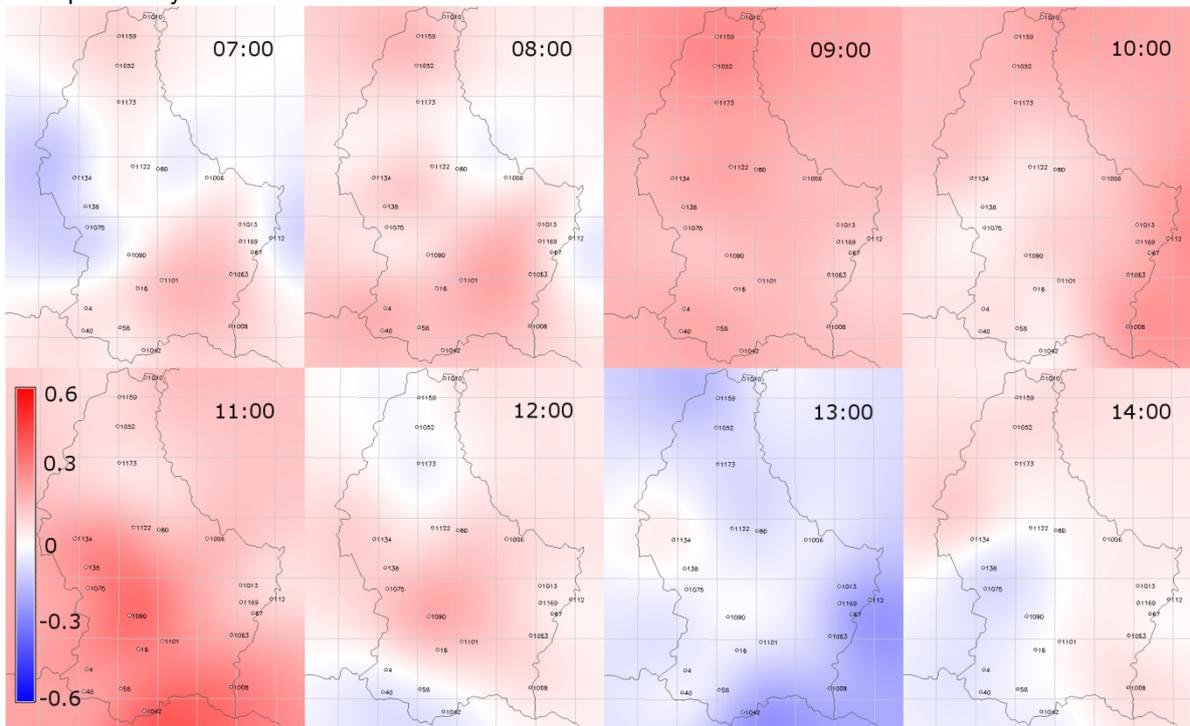


Fig 10 - picture series from an error map video sequence for the 10.08.2014 07:00 (top left) until 14:00 (bottom right)

For only few days, as e.g. shown in Fig 11, error movements are relatively clearly identifiable. On 05.09.2015, the forecasts were relatively well suiting around 08:00 (indicated by pale colours) – the following hours show underestimations of the forecasts in the southern part of Luxembourg and overestimations in the North. The area of overestimation sweeps over the forecast area from North to South within 4-5 hours. Such movements could be identifiable and might be forecasted into the short term future, but even here, an improvement of the forecast might only be possible 1-3 hours ahead.

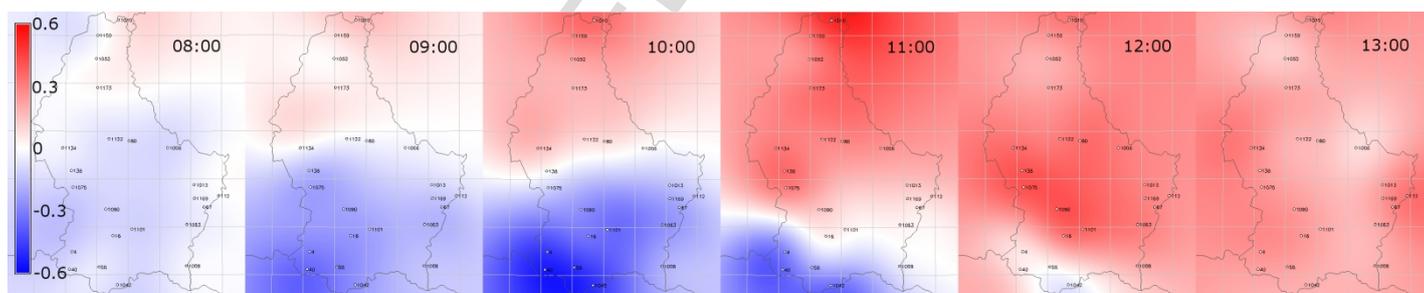


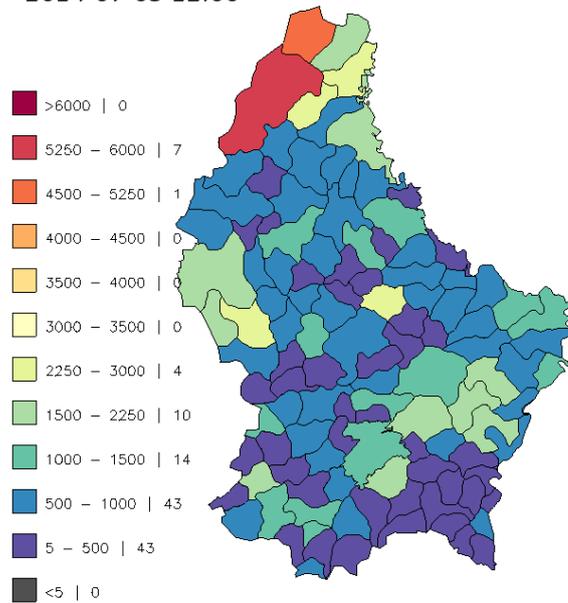
Fig 11 - error maps video sequence showing clear movement of areas of different deviations (05.09.2015)

The conclusion of the analysis of video sequences of error maps is therefore negative, concerning the approach of forecast improvements by “error movement vectors”. Some days show forecast error patterns where the described method might make sense, but most of them not. The single error maps change too drastically from one time step to the next. This leads also to the thesis, that this approach could be more promising at higher time resolutions (e.g. 5 min), which would increase the potential set of pictures that serve as basis to identify movements [17]. The technical possibility to apply this approach would surely increase with the resolution in time, but the absolute forecast horizon for which the method might improve the forecast remains limited (minutes to app. 3-hours-ahead).

#### 4.6. Results of the upscaling to the regional scale (Luxembourg)

The final result of the whole PV power forecasting algorithm is a dynamic and regionalized power forecast. Based on an upscaling procedure (not described here), using a full list of all PV systems installed in the grid of the two largest grid operators, giving their nominal power and location, an artificial power forecast for each system is generated. These powers could be aggregated on different regional or technical scales, e.g. per street, per village, per municipality or if the information would be available, per transformer station. Currently, the expected PV power is aggregated on communal level, as shown in Fig 12.

2014-07-03 12:00



(c)LIST

Fig 12 - forecasted PV power [kW] for 03.07.2014 12:00, aggregated per municipality

The visualization of the predicted PV power on communal level gives a rather varying picture, as the differences in installed power per municipality are relatively high (larger PV capacities installed in the northern, rural municipalities). Therefore, the high differences between the actual PV production of the municipalities are mainly due to the installed nominal power and only to a minor extent caused by different conditions of irradiance. The municipality in the North (Wincrange), illustrated in deep red on the map, has by far the largest installed PV capacity with 6'599 [kW<sub>p</sub>] nominal power and therefore sticks out of nearly every map, relatively independent of the irradiance.

More meaningful is the dynamic dimension of the forecast. Based on these single pictures, video sequences have been produced to illustrate the daily variation in PV power over the two year periods.

## 5. Conclusions and Outlook

Finally, the performance of the individual hourly power forecasts for the 23 reference systems, evaluated over a period of 2 years, is already quite promising. Without any adaptations of the forecast, based on the measurements of the reference systems, the mean deviation (bias) of the forecast was 1.1% of the nominal power (bias<sub>dt</sub> = 2.2%) – indicating low systemic error. Also the overall mean RMSE of 7.4% (RMSE<sub>dt</sub> = 10.0%) indicates a low dispersion of the power forecast. A huge collection of performance indicators for different forecast schemes can be found in recent review papers, such as [1] and [8], but a direct comparison is difficult. As Antonanzas stated [1], besides the large set of different indicators used and lack standardisation in their calculation, there are many factors which hamper a comparison: Climate conditions, day- and night-time values used, base of normalisation, sample aggregation, spatial aggregation level and testing period. Generally, it has been found that, by far, the main uncertainties arise from the irradiance forecast, which is not surprising, but nevertheless the accuracy of the technical part of the model is very satisfactory.

The adaptation of the forecast by the feedback from the reference systems brought ambivalent results. As mentioned above, at least on the time scale of a few hours ahead (1-2h) the error persistence approach did have a positive effect on forecast accuracy. But, the main advantage is rather the reduction of systemic errors, as e.g. in the case of snow cover or soiling, which is a known drawback of purely irradiance forecast depending approaches [1] [4] [14] [37].

The concept of error movement vectors was found to be not applicable for the temporal resolution of hourly forecasts and the given spatial resolution of this study. Also other authors concluded, although using different methods, that cloud movements over rather small regions pass too fast to reach acceptable results, at the respective temporal resolution of their data [17] [18] [19]. But we consider it worth to test the concept again, once the temporal resolution of the irradiance forecasts would increase. Also, the further rollout of smart meters and hence a higher spatial and temporal resolution of reference systems will be beneficial for this approach.

Furthermore, the smart meter rollout would lead to more suitable reference systems than those used in this project, since currently only large scale, complex systems were equipped with adequate meters. A high degree of details on the individual reference systems might even not be necessary, since the concept of synthetic system profile was found to be similarly performant.

The bottom-up structure of the model allows for the free choice of the aggregation level, if the degree of detail on the PV-systems is adequate. It depends on the specific purpose for which the forecast model would be used, if it might make sense to change the aggregation level of the up-scaled power forecast. For an energy provider/retailer the regionalized forecast would even not be necessary, hence the forecast can be aggregated on the level of their customers. For grid operators, if the necessary information on the localisation of individual PV systems is available, the forecast could be aggregated on level of transformer stations or street level, which would enable a very advanced grid management where this model would be applied.

The method described in this paper is easily transferable to other countries or regions, wherever data on the installed PV systems (at least nominal power and location) is available. Numerical weather prediction data can be retrieved from different sources, although in varying quality and resolutions (temporal as well as spatial), for all regions worldwide. Access to smart meter data of PV-systems is not yet state-of-the-art for many regions, but the European smart meter roll-out and the rapid digitalisation of the grid in many other regions of the world (e.g. the U.S.A. and Japan) will hold huge potential to make use of those data to increase accuracy of the NWP based forecasting in the short term time frame.

Targeting on retailers and trading, a further development step will be to change from a deterministic forecasting (point forecasts) to probabilistic forecasts, enabling an improved risk management, which will be tested in a next step.

### Acknowledgements & Conflict of interest

The authors would like to thank Creos, Enovos (and their customers) as well as AEV and ECMWF for data, support and feedback. We are very grateful to Fondation ENOVOS for financing the project. Further gratitude to our former colleagues Jonathan Hervieu and Markus Jonas for their help.

There are no conflicting interests to be declared.

### References

- [1] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. Martinez-de-Pison and F. Antonanzas-Torres, "Review of photovoltaic power forecasting," *Solar Energy*, vol. 136, pp. 78-111, 2016.
- [2] P. Mathiesen, J. Kleissl and C. Collier, "Case studies of solar forecasting with the weather research and forecasting model at GL-Garrad Hassan," in *Solar Energy Forecasting and Resource Assessment*, Waltham, Academic Press, 2013, pp. 358-381.
- [3] H. Ruf, "Limitations for the feed-in power of residential photovoltaic systems in Germany – An overview of the regulatory framework," *Solar Energy*, vol. 159, pp. 588-600, 2018.
- [4] J. da Silva Fonseca, T. Oozeki, H. Ohtake, T. Takashima and K. Ogimoto, "Regional forecasts of photovoltaic power generation according to different data availability scenarios: a study of four methods," *PROGRESS IN PHOTOVOLTAICS: RESEARCH AND APPLICATIONS*, vol. 23, pp. 1203-1218, 2015.
- [5] IEA PVPS Task 14, Subtask 3.1, "Photovoltaic and Solar Forecasting - State of the Art," IEA International Energy Agency, 2013.
- [6] R. H. Inman, H. T. Pedro and C. F. Coimbra, "Solar forecasting methods for renewable energy integration," *Progress in Energy and Combustion Science*, pp. 535-576, 2013.
- [7] E. Lorenz, J. Kühnert, D. Heinemann, K. P. Nielsen, J. Remund and S. C. Müller, "Comparison of global horizontal irradiance forecasts based on numerical weather prediction models with different spatio-temporal resolutions," *Progress in Photovoltaics*, vol. 24, pp. 1626-1640, 2016.
- [8] S. Sobri, S. Koohi-Kamali and N. Rahim, "Solar photovoltaic generation forecasting methods: A review," *Energy Conversion and Management*, pp. 459-497, 2018.
- [9] J. Kleissl, *Solar Energy Forecasting and Resource Assessment*, Oxford: Academic Press, 2013.
- [10] B. Wolff, J. Kühnert, E. Lorenz, O. Kramer and D. Heinemann, "Comparing support vector regression for PV power forecasting to a physical modeling approach using measurement, numerical weather prediction, and cloud motion data," *Solar Energy*, vol. 135, pp. 197-208, 2016.
- [11] R. Perez, S. Kivalov, J. Schlemmer, K. J. Hemker, D. Renné and T. Hoff, "Validation of short and medium term operational solar radiation forecasts in the US," *Solar Energy*, vol. 84, pp. 2161-2172, 2010.
- [12] E. Lorenz, J. Hurka, D. Heinemann and H. G. Beyer, "Irradiance Forecasting for the Power Prediction of Grid-Connected Photovoltaic Systems," *IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING*, vol. 2, no. 1, pp. 2-10, 2009.
- [13] P. Mathiesen and J. Kleissl, "Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States," *Solar Energy*, vol. 85, pp. 967-977, 2011.
- [14] E. Lorenz, T. Scheidsteger, J. Hurka, D. Heinemann and C. Kurz, "Regional PV power prediction for the improved grid integration," *PROGRESS IN PHOTOVOLTAICS: RESEARCH AND APPLICATIONS*, pp. 757-771, 2011.
- [15] E. Lorenz, J. Kühnert, B. Wolff, A. Hammer, O. Kramer and D. Heinemann, "PV Power Predictions on Different Spatial and Temporal Scales Integrating PV Measurements, Satellite Data and Numerical Weather Predictions," in *Proceedings of the EU PVSEC 2014*, Amsterdam, 2014.
- [16] R. Marquez, H. Pedro and C. Coimbra, "Hybrid solar forecasting method uses satellite imaging and ground telemetry as inputs to ANNs," *Solar Energy*, vol. 92, pp. 176-188, 2013.
- [17] V. Lonji, A. Brooks, A. Cronin, M. Leuthold and K. Koch, "Intra-hour forecasts of solar power production using measurements from a network of irradiance sensors," *Solar Energy*, vol. 97, pp. 58 - 66, 2013.
- [18] R. Bessa, A. Trindade, C. S. Silva and V. Miranda, "Probabilistic solar power forecasting in smart grids using distributed information," *Electrical Power and Energy Systems*, vol. 72, pp. 16-23, 2015.
- [19] A. Vaz, B. Elsinga, W. van Sark and M. Brito, "An artificial neural network to assess the impact of neighbouring photovoltaic systems in power forecasting in Utrecht, the Netherlands," *Renewable Energy*, vol. 85, pp. 631-641, 2016.
- [20] EU, "Web site of the European Union," EU , [Online]. Available: <https://ec.europa.eu/energy/en/topics/markets-and-consumers/smart-grids-and-meters>. [Accessed 02 05 2018].
- [21] R. Hogan, "Radiation Quantities in the ECMWF model and MARS," ECMWF, London, 2015.
- [22] F. Olmo, J. Vida, I. Foyo, Y. Castro-Diez and L. Alados-Arboledas, "Prediction of global irradiance on inclined surfaces from horizontal global irradiance," *Energy*, pp. 689-704, 1999.
- [23] S. A. Khalil and A. Shaffie, "Performance of Statistical Comparison Models of Solar Energy on Horizontal and Inclined Surface," *International Journal of Energy and Power (IJEP)*, pp. 8-25, 2013.
- [24] F. Minette, D. Koster and O. O'Nagy, "PV-FORECAST: REGIONALISED FORECASTING OF ENERGY PRODUCTION FROM PHOTOVOLTAIC AND THEIR DYNAMICS BY A COMBINED APPROACH OF MODELLING AND REAL TIME MEASUREMENTS OF REFERENCE SYSTEMS," in *European PV Solar Energy Conference 2014*, Amsterdam, 2014.
- [25] S. Pelland, G. Galanis and G. Kallos, "Solar and photovoltaic forecasting through post-processing of the Global Environmental Multiscale numerical weather prediction model," *PROGRESS IN PHOTOVOLTAICS: RESEARCH AND APPLICATIONS*, vol. 21, pp. 284-296, 2011.

- [26] W. De Soto, S. Klein and W. Beckman, "Improvement and validation of a model for photovoltaic array performance," *Solar Energy*, pp. 78-88, 2006.
- [27] S. N. Laboratories, "PV Performance Modeling Collaborative," [Online]. Available: <https://pvpmc.sandia.gov/>. [Accessed 2016].
- [28] D. Jordan and S. Kurtz, "Photovoltaic Degradation Rates—an Analytical Review," *PROGRESS IN PHOTOVOLTAICS: RESEARCH AND APPLICATIONS*, vol. 21, pp. 12-29, 2013.
- [29] C. Osterwald, A. Anderberg, S. Rummel and L. Ottoson, "Degradation analysis of weathered crystalline-silicon PV modules," in *Conference Record of the IEEE Photovoltaic Specialists Conference, 29th*, New Orleans, 2002.
- [30] S. Sakamoto and T. Oshiro, "Field test results on the stability of crystalline silicon photovoltaic modules manufactured in the 1990's," in *3rd World Conference on Photovoltaic Energy Conversion*, Osaka, 2003.
- [31] E. D. Dunlop, "Lifetime performance of crystalline silicon PV modules," in *3rd World Conference on Photovoltaic Energy Conversion*, Osaka, 2003.
- [32] A. CHOUDER and S. SILVESTRE, "Analysis model of mismatch power losses in pv systems," *Journal of Solar Energy Engineering*, vol. 131, no. 2, 2009.
- [33] D. Picault, B. Raison, S. Bacha, J. de la Casa and J. Aguilera, "Forecasting photovoltaic array power production subject to mismatch losses," *Solar Energy*, vol. 84, pp. 1301-1309, 2010.
- [34] power one, *TRIO-27.6-TL Datasheet*.
- [35] Institut Luxembourgeois de Régulation, "Chiffres clés du marché de l'électricité," ILR, Luxembourg, 2016.
- [36] J. Bosch, Y. Zheng and J. Kleissl, "Deriving cloud velocity from an array of solar radiation measurements," *Solar Energy*, vol. 87, pp. 196-203, 2013.
- [37] E. Lorenz, D. Heinemann and C. Kurz, "Local and regional photovoltaic power prediction for large scale grid integration: Assessment of a new algorithm for snow detection," *PROGRESS IN PHOTOVOLTAICS: RESEARCH AND APPLICATIONS*, vol. 20, pp. 760-769, 2012.

710

711

712

**Glossary**

713

714

AC alternate current

715

AEV Administration de l'Environnement (Environmental Agency)

716

CMV cloud motion vector

717

Creos electricity grid operator [www.creos-net.lu](http://www.creos-net.lu)

718

ASHRAE American Society of Heating, Refrigerating and Air-Conditioning Engineers

719

DC Direct Current

720

Enovos energy supplier in Luxembourg [www.enovos.lu](http://www.enovos.lu)

721

ECMWF European Center for Medium-Range Weather Forecasts [www.ecmwf.int](http://www.ecmwf.int)

722

kWp kilo Watt peak – nominal power of the PV module

723

LIST Luxembourg Institute of Science and Technology

724

MOS Model Output Statistics

725

MPP Maximum Power Point

726

NARX nonlinear autoregression using external data

727

NWP numerical weather prediction

728

PV Photovoltaic

729

RMSE root mean square error

730

ssrd surface solar radiation downwards

731

STC Standard Test Conditions

732

SVR support vector regression

## Short-term and regionalized PV-power forecasting, enhanced by PV-reference systems, on the example of Luxembourg

Daniel Koster<sup>1</sup>, Frank Minette<sup>1</sup>, Christian Braun<sup>1</sup>, Oliver O’Nagy<sup>1</sup>

<sup>1</sup>Luxembourg Institute of Science and Technology (LIST) – Environmental Research and Innovation Department (ERIN)

### Highlights

- A hybrid approach for PV-power forecasting, using metered PV-systems as references
- Demonstrating a comparably accurate forecast performance on our case study over a two years period
- Bottom-up model, able to reach high spatial resolutions if the data is available