



**HAL**  
open science

# Lightweight hybrid local PV short term forecasting using both physics and data

Benoit Delinchant, Vincent Imard, Yaiza Bermudez, Frédéric Wurtz

## ► To cite this version:

Benoit Delinchant, Vincent Imard, Yaiza Bermudez, Frédéric Wurtz. Lightweight hybrid local PV short term forecasting using both physics and data. APPEEC, 15th Asia-Pacific Power and Energy Engineering Conference, IEEE PES, Dec 2023, Chiang Mai, Thailand. <hal-04339309>

**HAL Id: hal-04339309**

**<https://hal.science/hal-04339309v1>**

Submitted on 13 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Lightweight hybrid local PV short term forecasting using both physics and data

Benoit Delinchant  
Univ. Grenoble Alpes, CNRS,  
Grenoble INP, G2Elab,  
Grenoble, France  
ORCID : 0000-0003-4296-0993  
benoit.delinchant@grenoble-inp.fr

Vincent Imard  
Univ. Grenoble Alpes, CNRS,  
Grenoble INP, G2Elab,  
Grenoble, France

Yaiza Bermudez  
Univ. Grenoble Alpes, CNRS,  
Grenoble INP, G2Elab,  
Grenoble, France

Frederic Wurtz  
Univ. Grenoble Alpes, CNRS,  
Grenoble INP, G2Elab,  
Grenoble, France

**Abstract**— There is an increase of PV generation on many distributed sites, especially rooftop PV on buildings. The emergence of aware end-users, that are doing their best to maximize their auto-consumption in an energy community using energy management systems, leads to a need for new adapted forecasting models. We ask the question in this article of the ability to produce with few data, a robust day ahead PV generation forecasting. To do that, we have applied a hybrid methodology taking advantage of both physical model and data-based model in order to reduce the amount and kind of data. We reach improvements from 5% to 12% depending on the model used, compared to pure data-based approaches.

**Keywords**— PV forecasting, Hybrid-AI, Physic-based model, Data-based model, Clear sky, cloud cover forecast, LSTM, XGBoost.

## I. CONTEXT AND REQUIRMENT SPECIFICATIONS

The ongoing energy transition is seeing the spread of a large number of individual photovoltaic (PV) installations. These installations have a major impact on the electrical distribution. Buildings allow degrees of flexibility and can be integrated into aggregators that seek to optimize energy flows through Demand Response (DR) mechanisms in interaction with energy markets. They can interact directly with flexible assets or via end users, but must in any case receive their consent and can encounter many barriers [1]. In order to optimize this management, the aggregator must have knowledge of the distributed production, but also of consumer behavior, which is an active research field [2] and subject to many uncertainties. More and more individual Renewable Energy Sources (RES) generation are being implemented at the scale of buildings, or at the scale of energy communities. Producers become actors in the energy system, and can benefit from local Energy Management Systems (EMS).

A first challenge for RES forecasting at local scale development is that it is lightweight and accessible. By accessible we mean that it is available but also easily usable, i.e. not requiring too specific characteristics that it would be impossible to provide without a dedicated scientific study. For instance, this is not the case of the Bird clear sky model which has demonstrated high precision [3] but requires too many details. There are many sources of uncertainties during design especially regarding energy demand [4], which justify works on more simplified PV models.

The EMS can be based on Model Predictive Control (MPC) in order to achieve optimal operation on flexible assets, like HVAC systems. Knowing from the design phase

that such an EMS will be available during the operation phase helps to avoid oversizing storage capacities [5].

When installed, the PV system is also able to produce measures that can be very useful during operation. These data are property of the PV owner, then free and accessible for him. Most of existing models is focusing on accuracy, and for that, once can use expensive sensors like solar radiometers, or sky imager. The quality of the model is not only base on its accuracy but on its accuracy over its training cost [6].

Our contribution here is to provide and analyze advantages and cons of day ahead 24h PV forecasting hybrid models which improves simplicity criteria such as few data learning and very accessible physical modeling, while keeping forecasting robustness.

## II. PV GENERATION FORECASTING METHODS

Sobri et al. in [7] has classified solar photovoltaic power forecasting methods into three major categories i.e., *time-series statistical methods*, *physical methods*, and *ensemble methods* :

1 - Most of *time-series statistical methods* are today depicted as Artificial Intelligence (AI) and experienced a spectacular development in recent years especially with Deep Neural Network (DNN) due to their capability in solving the non-linear and complex structure of historical data.

2 - *physical models* are based on the modeling of solar radiation and its dynamic interaction with the atmosphere. Two main equipments are used:

- the Sky Imager or *Sky Cam* (SC) are used for Cloud-Motion-Vector (CMV) methods that has to face issues of complex multilayer cloud displacements, cloud formation and dissipation.
- the *Satellite-based radiometers* used for irradiance forecasts using Surface Solar Irradiance (SSI) and CMV.

For larger temporal and spatial scales (see Fig. 1), Numerical Weather Prediction (NWP) is used. NWP models are based on numerical integration of coupled differential equations that explain the transportation of radiation mechanism and the dynamics of the atmosphere. In this study, we are focusing on microscale forecasting.

3 - The *ensemble method* refers to any combination (or hybridization) of pervious methods, most of time in a statistical approach in order to provide uncertainty quantification. In this paper, we are developing hybrid-AI which is combination of statistical approach and physic-based approach, but for complementarity and not redundancy.

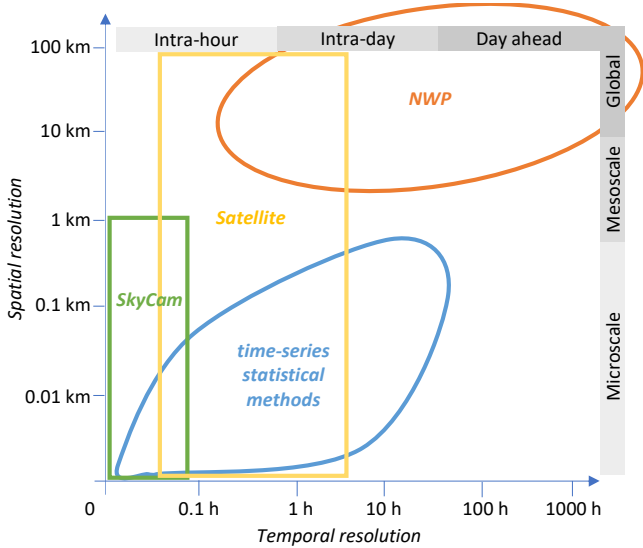


Fig. 1. Model classification based on temporal and spatial resolutions (adapted from [7])

### III. HYBRID MODELING FOR PV FORECASTING

#### A. Hybrid Modeling Methodology

Physical models are approximations due to various factors, such as idealized assumptions, characteristic value approximations, unknown external factors. Then we are assuming that the physical model is incomplete, and unable to represent the whole complexity of the observed data.

It is important to use data to cope with uncertainties, and to use physical laws in order to guide the modeling approach with some physical constraints. Our main concern is to use data and physical characteristic that are easy to access. Then the physical model  $f_p$  can be considered as a coarse model of  $f$  that can be enhanced with purely statistical model  $f_b$  to estimate the bias, as in (1). We can imagine the physical part  $f_p$  taking most of the non-linearities, then correction term  $f_b$  will have to manage less complex regression, requiring then less amount of data.

$$\hat{f} = f_p + f_b \quad (1)$$

If  $f_p$  is known and a ground truth of PV production has been measured, then correction contribution or the model enrichment  $f_b$  is obtain using classical machine learning approach, as in (2).

$$f_b = f_{gt} - f_p \quad (2)$$

Most of time,  $f_p(\theta_p)$  requires to adjust  $\theta_p$  parameters, then the approach is to learn both  $\theta_p$  and  $\theta_b$  at the same time.

$$\hat{f}(\theta_p, \theta_b) = f_p(\theta_p) + f_b(\theta_b) \quad (3)$$

This approach is also used in Physic Informed Neural Network (PINN) or more generally Physical Informed Machine Learning [8] when considering Kernel based regression. The main issue to overcome is about non uniqueness of the decomposition (1).

In [10], authors formulated the learning problem such that “the physical model explains as much of the data as possible, while the data-driven component only describes information that cannot be captured by the physical model; no more, no less”. Then authors guarantee “existence and uniqueness for the decomposition given mild conditions, and show that this formulation ensures interpretability and benefits generalization”.

A third similar approach relative to real time control is called digital twin [8]. Due to real time constraints, it requires to calibrate off-line a Model Order Reduction which implement the coarse model  $f_p(\theta_p)$ . Then  $f_b(\theta_b)$  is update on-line in real time.

In the next parts, we are detailing which models  $f_p$  and  $f_b$  have to be used in our PV forecasting methodology.

#### B. Physic based model

Over the years, numerous physical models have been proposed to accurately model solar irradiance under clear sky, which plays a crucial role in solar power production. Solar irradiance model combined with information on PV panels characteristics is then able to estimate the energy production.

According to the suggestions provided in [11], it is noted that complex models heavily rely on local measurements that can be challenging to obtain. Hence, simpler models seem accurate enough for practical purposes. Considering this, even though the Bird clear sky model has demonstrated high precision [3] ( $\text{RMSPE}^1 \leq 5\%$ ), it is still hard to achieve such accuracy with expert knowledge of the system and its environment. Instead, we will focus use Ineichen-Perez clear sky model ( $\text{RMSPE} \leq 10\%$ ), which requires nine parameters including solar zenith, azimuth, latitude, longitude, altitude.

The pvliv [13] provides access to this model for solar irradiance estimation under clear sky. In particular, it offers the Hay-Davies model as the default for calculating solar irradiance. Additionally, the library includes the isotropic model, which has demonstrated satisfactory results [14].

#### C. Data based model

##### 1) Literature review

Using historical data, a data-driven approach can capture intricate patterns and relationships that may not be easily taken into account by the physics formulas, and can potentially enhance the accuracy of the forecasts. For instance, Zamo presents in his 2014 publication [14] a benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production built without any technical information about panels or irradiation models. Models can be deterministic or probabilistic based on Ensemble Method (EM) while no forecasting models clearly dominates the other whatever the power plant.

Some data-based models do not directly forecast PV generation but solar radiance [16]. A second less complex model can then transformation the solar forecast into PV forecast [17]. But most of small PV plants (such as rooftop PV) have no access to irradiance measures or at the best to the global irradiance only, but rarely diffuse and direct irradiance separation.

<sup>1</sup> Here RMSPE is RMSE divided by the max value of the true values and multiplying by 100 (in percentage)

A literature review summarized in the next table, show that Extreme Gradient Boosting (XGBoost) and Long Short-Term Memory (LSTM) models show better accuracy compared to other models, demonstrating their effectiveness in capturing complex patterns and relationships within the data.

Dimitropoulos [17] has compared several Machine Learning (ML) models with very comparable performances. In the case of XGBoost, the model uses features such as temperature, pressure, humidity, and solar radiance to predict solar power generation. On the other hand, LSTM utilizes historical power generation data to capture temporal dependencies and make forecasts. By considering different sets of input variables, XGBoost focuses on weather-related factors that impact solar power generation, while LSTM emphasizes the historical patterns of power generation itself. Both approaches have demonstrated their efficacy in achieving accurate predictions.

XGBoost and LSTM will be used for ML in our model.

## 2) Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost), is an open source machine learning library designed to perform Gradient Boosting (GB) for decision tree models. GB is close to classical regression tree ensemble Random Forest (RF). Where RF manages the variance problem thanks to bootstrapping, it does not manage at best the bias introduced by the depth of the regression trees. GB helps to better manage the bias. While RF generates random and independent trees in parallel, GB works sequentially to create new trees depending on the performance of the previous ones.

## 3) Long-Short Term Memory

Long short term memory (LSTM) is a recurrent neural network (RNN) dedicated to time series modelling which is able to catch different dynamics of the signal, and bring a solution to the long term gradient vanishing during the back propagation mechanism of classical RNN.

## IV. CASE STUDY: GREEN-ER PV FORECASTING

In order to illustrate the modeling strategy, our case study is a rooftop PV of an engineering school building which owns also solar irradiance instruments for validation purposes.

### A. GreEn-ER instrumentation and PV generation

#### 1) GreEn-ER PV Generation

The 890-square-meter photovoltaic power plant installed on the roof of the GreEn-ER building (Fig. 2) has a capacity of 183kWp. The energy produced is self-consumed, representing 10% of the building's total annual consumption.



Fig. 2. GreEn-ER building with 183kWp PV panels

#### 2) Solar irradiance

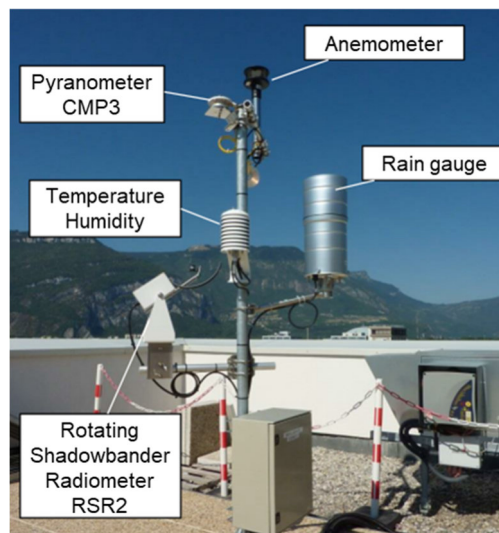


Fig. 3. GreEn-ER weather station

Solar irradiance can be decomposed in Direct Normal Irradiance (DNI, in W/m<sup>2</sup>) and Global Horizontal Irradiance (GHI, in W/m<sup>2</sup>). These two components can be measured using a Rotating Shadowband Radiometer [18] available in our weather station (Fig. 3) which are required in some physical models and are used here for validation purpose only.

#### 3) Sky Imager

Our weather station has also cameras, with sky cam (Fig. 4) which provides images of the surrounding area. These images can be used to produce the cloud cover indicator (also known as cloudiness, cloudage, or cloud amount) which refers to the fraction of the sky obscured by clouds on average when observed from a particular location. The literature mainly exploits Sky Imaging [20] to reach the local scale (Microscale, 1m – 1km).

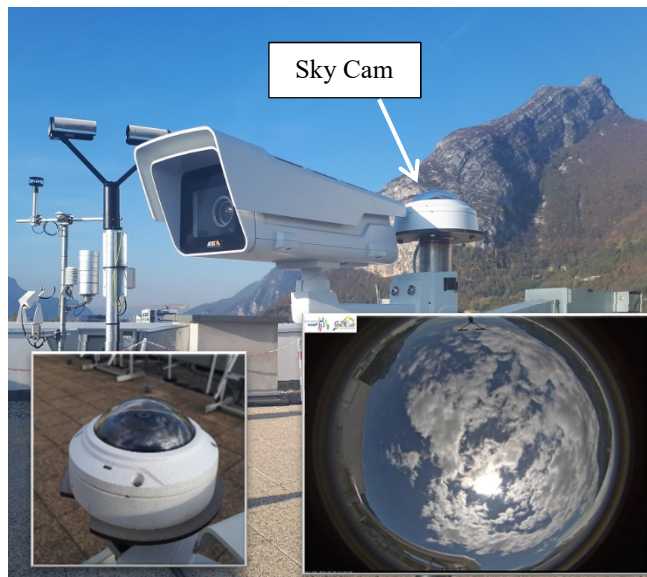


Fig. 4. Sky imager AXIS M3027-PVE, and a sky view with clouds

#### 4) On line web-service weather forecasting

In [19], we showed how cloud cover forecast by on-line weather services can be used to predict PV production from a clear sky and empirical model. Indeed, it is nowadays easy to obtain cloud cover forecast from weather web services such as

AROME model from Meteo-France<sup>2</sup> which is a small scale NWP model, designed to short range forecasts.

Cloud cover history and forecasting is available even for free, from several API providers using forecasting technics based on NWP (see Table 1).

TABLE I. WEATHER FORECASTING HORIZON FOR 4 API PROVIDERS

Time step	Forecasting horizon (free or not)			
	Tomorrow IO	Accu Weather	Weather Bit	Open Weather
Hourly	5 days free	12h free up to 5 days	up to 10 days	up to 4 days
3-hour				5 days free
Daily	5 days free	5 days free up to 15 days	7 day free up to 16 days	up to 16 days

In this work we are considering this kind of NWP forecasting as a good tradeoff between the forecasting scale (local) and the easiness to obtain it (without local sensor).

## V. MODELING PROCEDURE AND RESULTS

### A. Physical model

#### 1) System description

TABLE II. shows parameters for the physical model

TABLE II. PHYSICAL MODEL PARAMETERS

Location :	Photovoltaic System :
<b>Grenoble, France</b>	<b>PV Panels : 452 units</b>
• Latitude : 45.17° N	• DualSun, FLASH 405 Half-Cut:
• Longitude : 5.7° E	• Peak power : 183 kWc
• Time zone : GMT+2	• Efficiency : 15.7%
• Altitude : 212m	• Tilt : 6°
	• Azimuth 1 : 139° for 220 units
	• Azimuth 2 : -41° for 232 units
Ground albedo : 0.2	<b>Inverter : SMA Sunny Tripower STP50-41 : 5 units</b>

#### 2) Theoretical PV production without clouds

The site location is provided to calculate solar positions. Then given the PV surface tilt and azimuth, the angle of incidence is obtained.

The Ineichen and Perez clear sky model parameterizes irradiance in terms of the Linke turbidity, allowing to decompose direct normal irradiance and diffuse horizontal irradiance which are used in Hay and Davies model to determines the diffuse irradiance on the PV. Then Sandia PV Array Performance Model calculates PV power.

This model corresponds to the theoretical 0% cloudcover PV production along a year.

### B. Data sets

#### 1) PV production

The dataset is 1-hour time step recording during 1 year from sept 2022 to sept 2023 to fit with 1-year seasonal variations that can be observed in Fig. 5 for PV production.

#### 2) Learning with few data

Our concern is to analyze the performances of our simple hybrid model (basic physic, and basic machine learning)

which must be able to adapt well even with unseen data thanks to the physical part.

Then, the models will be trained for the first 3 months (25%) and tested on the rest (75%) with a rolling window of 24-hours time step. For the test, we are not considering a continual learning, it means that the models remain the same even if we have access to new data.

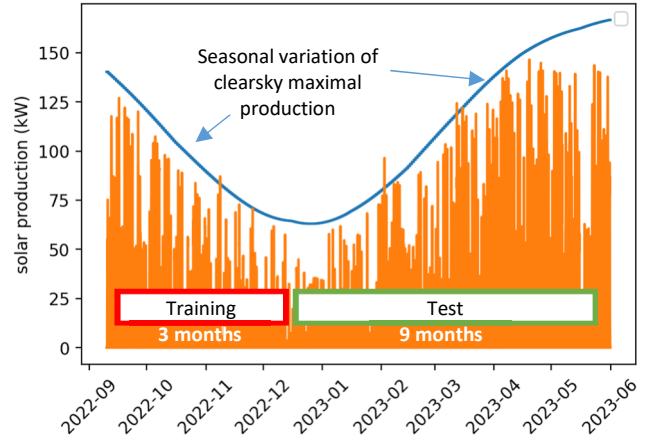


Fig. 5. PV production data training and test set,

#### 3) Cloud cover and PV correlation

The dataset used is the history of day ahead hourly forecasting from Tomorrow IO provider (see Table 1).

A first correlation analysis can be done to see how PV production and cloud cover are linked. As it can be seen in Fig. 6 on the left side, there is no easy to capture trend.

As explained in the Hybrid Modeling Methodology part, the clear sky physical model can be considered as a coarse model of the real PV production, that can be enhanced with purely statistical model to estimate the gap, as in equation (1).

As it can be seen in Fig. 6 on the right side, a linear trend is appearing between cloud cover and the gap.

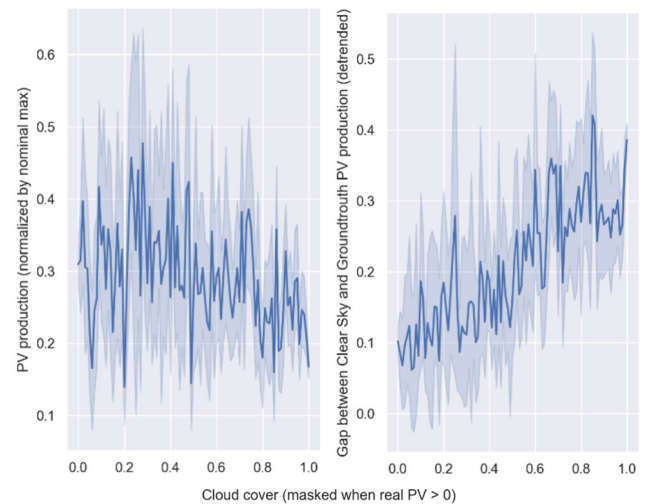


Fig. 6. Correlation analysis between cloud cover and left: PV production, right: Gap between ClearSky and PV production

<sup>2</sup> <https://www.umr-cnrm.fr/spip.php?article120&lang=en>

### C. Fully data based day ahead forecasting models

#### 1) Reference model : fully data-based approach

As described in the literature review, LSTM is well recognized for time series forecasting. It is considered here as our reference model. A Sequence to Sequence LSTM model is then developed with the PV production history observed over the last 24 hours as input and the next 24 hours as output.

The structure is 128 layers LSTM and 1 dense layer with ReLU activation function. Adaptive learning rate stochastic gradient optimization algorithm is used (Adam), with mean square error metric.

The RMSE obtained is 14.55% of the nominal power. This pure data-based model is our reference model ( $PV_{d+1}=LSTM(PV_d)$  ref).

#### D. Including nebulosity with LSTM

Cloud cover is now added as exogenous variables to the LSTM reference model as described in Fig. 7. It leads to an improvement of 23% as summarized in Table 2 for  $PV_{d+1}=LSTM(PV_d,cc_{d+1})$ .

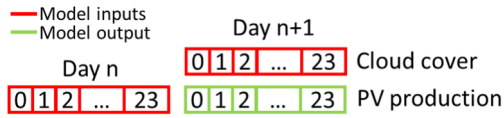


Fig. 7. Seq2Seq LSTM model with cloud cover

The LSTM structure is similar to the reference one, just having one input feature of size 24 by 2. Here we are not analyzing the different technics to add such auxiliary feature with the PV sequence. One can try others structures, for instance one in which cloud cover is appended to the output of the first LSTM and choose to pass both to another learning structure.

#### E. Gradient boosting

As the literature review mentioned both LSTM and gradient boosting, we will check if our conclusions can be confirmed with this different model. XGBoost models are defined with 200 estimators, a maximal depth of 60, and a squared error minimization.

The first model  $PV_{d+1}=XGB(cc_{d+1})$  is using the 24-hours cloud cover forecast in input to predict PV production. It improves the reference model by 16%

The second model  $PV_{d+1}=XGB(PV_d,cc_{d+1})$ , add to the previous one the PV production history, leading to 21% improvement compared to the reference.

All results will be summarized in TABLE III.

#### F. Hybrid day ahead forecasting models

##### 1) Detrend based on daily clearsky maximal production

In the previous methods, PV data were normalized according the linear scaling technic to the range [0;1] based on the nominal power or maximal observation for the whole set. But PV curve can be daily scaled by the trend defined in first approximation by daily clear sky maximal production (Fig. 5). This trend is fully determined by the location on earth and PV tilt, then it is obvious to use the theoretical model..

The reference model is improved by 5% just detrending the data. This improvement is quite significant, and would be less if a full year's data had been used for the learning process,

or if a continual learning approach had been applied. But this detrend is a first step to improve the pure data based approach, especially when few data are available.

##### 2) Correction term learning added to physical model

In order to apply the hybrid methodology detailed earlier, we defined a correction term as in equation (2) to model the gap between clear sky model and PV generation measures.

Then, using the same LSTM structure as previously defined, a first model ( $gap_{d+1}=LSTM(gap_d)$ ) is learnt with this 24-hours gap history, that reach 19% improvement. This result shows the importance of the hybrid approach, the improvement may be explained by a less complex behavior of the error than the PV production itself.

Now we are considering the cloud cover forecasting as a model feature. The new model  $Gap_{d+1}=LSTM(gap_d,cc_{d+1})$  reaches the improvement of 33% compared with the reference (without cloud cover), and 12% compared to  $PV_{d+1}=LSTM(PV_d,cc_{d+1})$ . This last result is mainly due to the fact that cloud cover is explaining the gap more easily than the PV production itself, as we had foreseen with Fig. 6.

We have obtained similar results with gradient boosting for  $Gap_{d+1}=XGB(cc_{d+1})$  and  $Gap_{d+1}=XGB(gap_d,cc_{d+1})$  with respectively 24% and 26% of improvement compared with reference, and respectively 9% and 5% of improvement compared with there equivalent models without hybrid approach.

### G. Results

#### 1) Hybrid detrend methodology performance results

TABLE III. which summarizes results, is providing 2 main columns on left and on the right which list the different models tested to validate the methodology. On the left, 4 models of the PV model based on data, and on the right 4 equivalent models of the gap, based on data and physic.

The feature used are 24h history, and 24h cloud cover forecasting. Two regression technics are used, LSTM and Gradient Boosting.

The two first columns of results are the improvement of each model with regards to the reference model  $PV_{d+1}=LSTM(PV_d)$ . The third column consider only the improvement due to hybrid methodology.

TABLE III. PERFORMANCE COMPARISON WRT REFERENCE MODEL - HYBRID IMPROVEMENT WRT TO PURE DATA MODELS

Pure data Models	Improvement wrt reference	Improvement wrt reference	Hybrid improvement	Hybrid models
$PV_{d+1}=LSTM(PV_d)$	ref	19%	<b>19%</b>	$Gap_{d+1}=LSTM(gap_d)$
$PV_{d+1}=LSTM(PV_d,cc_{d+1})$	23%	33%	<b>12%</b>	$Gap_{d+1}=LSTM(gap_d,cc_{d+1})$
$PV_{d+1}=XGB(cc_{d+1})$	16%	24%	<b>9%</b>	$Gap_{d+1}=XGB(cc_{d+1})$
$PV_{d+1}=XGB(PV_d,cc_{d+1})$	21%	26%	<b>5%</b>	$Gap_{d+1}=XGB(gap_d,cc_{d+1})$

The best performances are obtained by  $Gap_{d+1}=LSTM(gap_d,cc_{d+1})$  with 33% improvement wrt to reference model, and 12% wrt to  $PV_{d+1}=LSTM(PV_d,cc_{d+1})$ . It reached 9.78 RMSE in percentage of the nominal PV production, and a regression score  $R^2=0.78$ .

## H. Modeling procedure

Here we summarize the procedure to obtain such a model:

- Collect PV measurements and Cloud cover forecasting
- Collect PV characteristics and simulate theoretical PV production without clouds using PVLib
- Detrend both measured PV and Clear sky PV
- Define Gap as the difference of both previous time series
- Create training batches of 24 hours data and learn a model of the next 24-hours Gap from Gap history and Cloud cover forecasting features
- Predict the next 24-hours Gap, remove it from detrended clear sky model, and retrend the result which correspond to the next 24-hours PV production.

## VI. CONCLUSIONS AND PERSPECTIVES

In this paper, we proposed a hybrid model for PV production day ahead forecast. It is based on a basic clear sky model which require only few parameters on location (latitude and longitude) and PV installation (tilt, azimuth, peak power and efficiency). Then the bias model is obtained using machine learning technics with Long Short Term Memory (LSTM) or Extreme Gradient Boosting (XGBoost) with 24h history and cloud cover forecasting features. We especially proposed to include cloud cover since it is easy to access from Numerical Weather Prediction (NWP) web-services.

Training often requires a significant amount of historical data, and then need to wait before obtaining and using forecasting models. The hybrid methodology is a way to learn with few data, and improve the robustness with regards to unseen data. Indeed, this work provides results of improvements from 5% to 12% depending on the model used.

3 main factors are explaining the good results:

- Detrend based on daily nominal production.
- The less complex behavior, compared to PV production, of the correction model (the bias), which is here the gap between clear sky physical model and real PV production.
- The better correlation between cloud cover and the gap, than the PV production.

Open-science : Data and models are available at <https://gricad-gitlab.univ-grenoble-alpes.fr/GreEn-ER/pv-production>

## VII. ACKNOWLEDGEMENTS

This research has been partially supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme, and by MIAI@Grenoble Alpes, France (ANR-19-P3IA-0003)

- [1] F. D'Ettoire et al, "Exploiting demand-side flexibility: State-of-the-art, open issues and social perspective", *Renewable and Sustainable Energy Reviews*, Volume 165, 2022, 112605, ISSN 1364-0321
- [2] Daniel Llerena, Beatrice Roussillon, Sabrina Teyssier, Adelaïde Fadhuile, Penelope Buckley, Benoit Delinchant, Jérôme Ferrari, Tiansi Laranjeira, Frédéric Wurtz « Demand response in the workplace: A field experiment », *Journal of Environmental Management*, Volume 341, 2023, 117992, ISSN 0301-4797, <https://doi.org/10.1016/j.jenvman.2023.117992>
- [3] Brighton Mabasa, Meena D. Lysko, Henerica Tazvinga, Nosipho Zwane, and Sabata J. Moloï. "The performance assessment of six global horizontal irradiance clear sky models in six climatological regions in south africa". *Energies*, 2021.
- [4] Benoit Delinchant, Gustavo Felipe Martin Nascimento, Tiansi Laranjeira, Thi-Tuyet-Hong Vu, Muhammad Salman Shahid, Frédéric Wurtz " Machine Learning on Buildings Data for Future Energy Community Services » SGE 2021 - Symposium de Génie Electrique 2021, Nantes, 6-8 juillet 2021
- [5] T T Vu Hong, Benoit Delinchant, Jérôme Ferrari and Quang D Nguyen "Autonomous Electrical System Monitoring and Control Strategies to Avoid Oversized Storage Capacity", 2020 IOP Conf. Ser.: Earth Environ. Sci. 505 012045 <https://doi.org/10.1088/1755-1315/505/1/012045>
- [6] Duc Hoang Tran, Van Linh Nguyen, Huy Nguyen, and Yeong Min Jang. Short-term solar power generation forecasting using edge ai. In *Proceedings of the 13th International Conference on Information and Communication Technology Convergence (ICTC)*, 2022.
- [7] Sobrina Sobri, Sam Koochi-Kamali, Nasrudin Abd. Rahim, "Solar photovoltaic generation forecasting methods: A review", *Energy Conversion and Management*, Volume 156, 2018, Pages 459-497, ISSN 0196-8904,
- [8] Chinesta, F., Cueto, E., Abisset-Chavanne, E. et al. Virtual, Digital and Hybrid Twins: A New Paradigm in Data-Based Engineering and Engineered Data. *Arch Computat Methods Eng* 27, 105–134 (2020). <https://doi-org.sid2nomade-1.grenet.fr/10.1007/s11831-018-9301-4>
- [9] Karniadakis, G.E., Kevrekidis, I.G., Lu, L. et al. Physics-informed machine learning. *Nat Rev Phys* 3, 422–440 (2021). <https://doi.org/10.1038/s42254-021-00314-5>
- [10] Yuan Yin et al « Augmenting physical models with deep networks for complex dynamics forecasting » *J. Stat. Mech.* (2021) 124012
- [11] Stein, Joshua S, Hansen, Clifford W, and Reno, Matthew J. 2012. "Global horizontal irradiance clear sky models : implementation and analysis.". United States. <https://doi.org/10.2172/1039404>. <https://www.osti.gov/servlets/purl/1039404>.
- [12] Pierre Niechen. Validation of models that estimate the clear sky global and beam solar irradiance. *Solar Energy*, 132:332–344 , 2016.
- [13] William F. Holmgren, Clifford W. Hansen, and Mark A. Mikofski. "pvlib python: a python package for modeling solar energy systems." *Journal of Open Source Software*, 3(29), 884, (2018). <https://doi.org/10.21105/joss.00884>
- [14] P.G. Loutzenhiser, H. Manz, C. Felsmann, P.A. Strachan, T. Frank, and G.M. Maxwell. Empirical validation of models to compute solar irradiance on inclined surfaces for building energy simulation. *Solar Energy* , 81:254–267, 2007.
- [15] M. Zamo, O. Mestre, P. Arbogast, O. Pannekoucke, A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production. Part II: Probabilistic forecast of daily production, *Solar Energy*, Volume 105, 2014, Pages 804-816, ISSN 0038-092X, <https://doi.org/10.1016/j.solener.2014.03.026>.
- [16] R. Ahmed, V. Sreeram, Y. Mishra, and M.D. Arif. A review and evaluation of the state-of-the-art in pv solar power forecasting: Techniques and optimization. *Renewable and Sustainable Energy Reviews*, 124(26), February 2020.
- [17] Nikos Dimitropoulos, Nikolaos Sofias, Panagiotis Kapsalis, Zoi Mylona, Vangelis Marinakis, Niccolo Primo, and Haris Doukas. Forecasting of short-term pv production in energy communities through machine learning and deep learning algorithms. In *Proceedings of the 12th International Conference on Information, Intelligence, Systems Applications (IISA)*, 2021.
- [18] S. Wilbert et al., *Task 46: Best Practices for Solar Irradiance Measurements with Rotating Shadowband Irradiometers*. 2015.
- [19] Dung V Nguyen, Benoit Delinchant, Binh V Dinh and Truong X Nguyen, "Irradiance forecast model for PV generation based on cloudiness web service", 2019 IOP Conf. Ser.: *Earth Environ. Sci.*, Vol. 307, conf. 1, 2019, [doi.org/10.1088/1755-1315/307/1/012008](https://doi.org/10.1088/1755-1315/307/1/012008)
- [20] Samuel R. West, Daniel Rowe, Saad Sayeef, Adam Berry, "Short-term irradiance forecasting using skycams: Motivation and development » *Solar Energy*, Volume 110, 2014, Pages 188-207, ISSN 0038-092X