EGU2020 – 06 May 2020

SHORT-TERM PHOTOVOLTAIC GENERATION FORECASTING USING HETEROGENOUS SOURCES OF DATA BASED ON AN ANALOG APPROACH

Kevin BELLINGUER [(1)]
Robin GIRARD [(1)]
Guillaume BONTRON [(2)]
Georges KARINIOTAKIS [(1)]

(1) MINES ParisTech, PSL University,
PERSEE - Centre for Processes, Renewable Energies and Energy Systems
(2) Compagnie Nationale du Rhône

Research center :

PERSEE

Centre for processes, renewable energies and energy systems

Industrial partner :

Contact: kevin.bellinguer@mines-paristech.fr

Plan

I. Objectives of PV forecasting

II. Proposed approach

   I. A multi-inputs approach

   II. A conditioned learning

III. Models definition / Case study

IV. Outcomes and analysis

V. Conclusions

OBJECTIVES OF PV FORECASTING

Weather dependence

- Production **variability**, limited **controllability**, **uncertainties…**
- … leads to **large-scale integration problems**

New challenges to meet

- Management of the distribution **network** (balance, reserve, …)
- Optimization of **maintenance** scheduling
- Trading on electricity **market**

What we propose:

- **Short-term** forecasting model (i.e. few minutes up to 6h ahead)
- A statistical approach
- A **seamless** model (i.e. a unique model suitable for large range of horizons)
- A **simple** model with **good performances**

A multi-inputs approach – State of the art

State-of-the-art

- Production measurements
- Numerical Weather Prediction (NWP)
- Satellite images
- Sky images

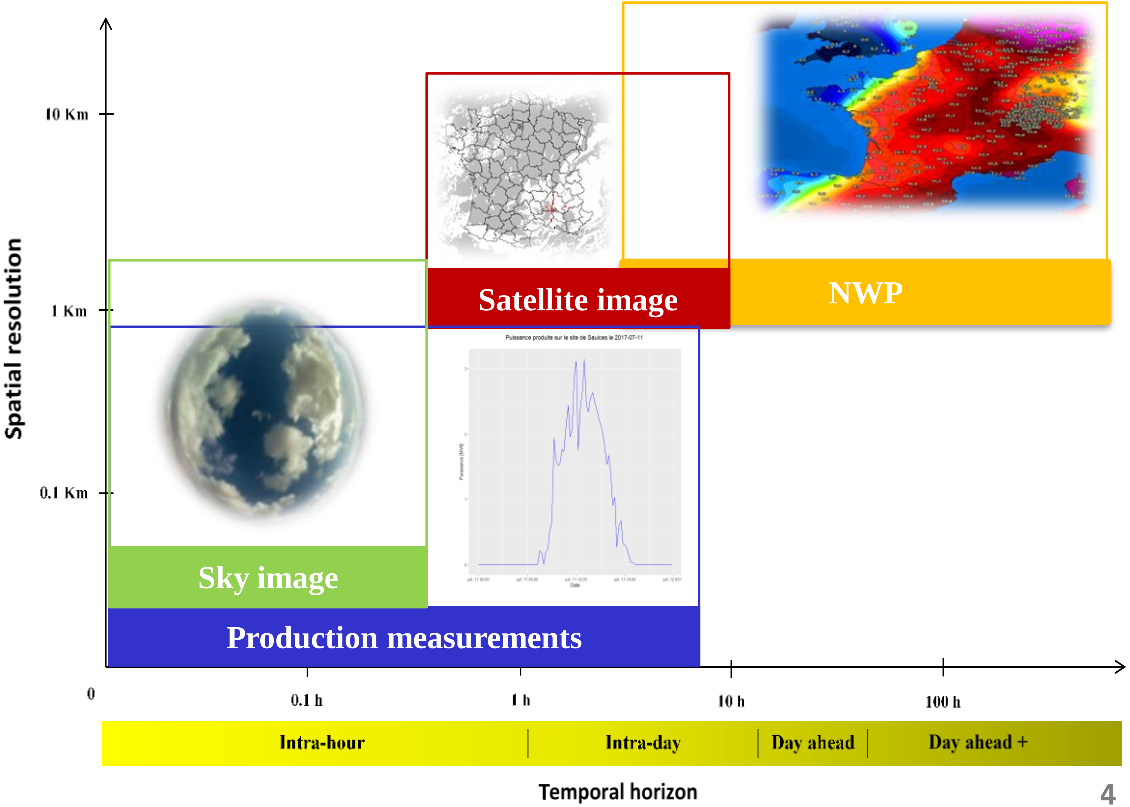
Inputs complementarity

- Due to their temporal & spatial resolutions, each approach observes specific phenomena

Inputs combination

- Improvement of forecasting performance for short-term horizons (Aguiar 2016, Vallence 2018)

Classification of inputs used for PV production forecasting.
Adapted from (*Antonanzas, 2016*) and (*Diagne 2013*).
Photo credits: (Météo-France, meteociel.fr, (*Cañadillas, 2018*), CNR)

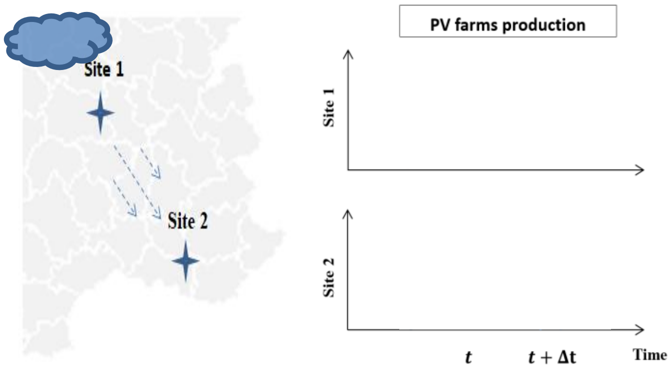


The Spatio-Temporal approach

Description

- Exploits the spatial and temporal correlations between the production data of spatially distributed PV sites
  - Idea: a set of neighbouring PV plants is affected by the same clouds
- Short-term horizons (Bessa 2015)

Spatio-Temporal approach



Interests

- Data easily **available** (provided real time data logging)
- Improve forecasting performance **up to 10%** (Bessa 2015, Agoua 2018)

Limitations

- PV sites **spatial distribution** (low density)
- **Quality** of time series (converter shutdown, …)

A need for adaptivity

Weather variability

- **Variability in cloud cover** leads to variability in PV production
- Models need to operate on a **wide range of weather conditions** (from sunny to overcast skies)
- An **adaptative model** is more **accurate** than a static model (*Bacher, 2009*)

Proposed approach

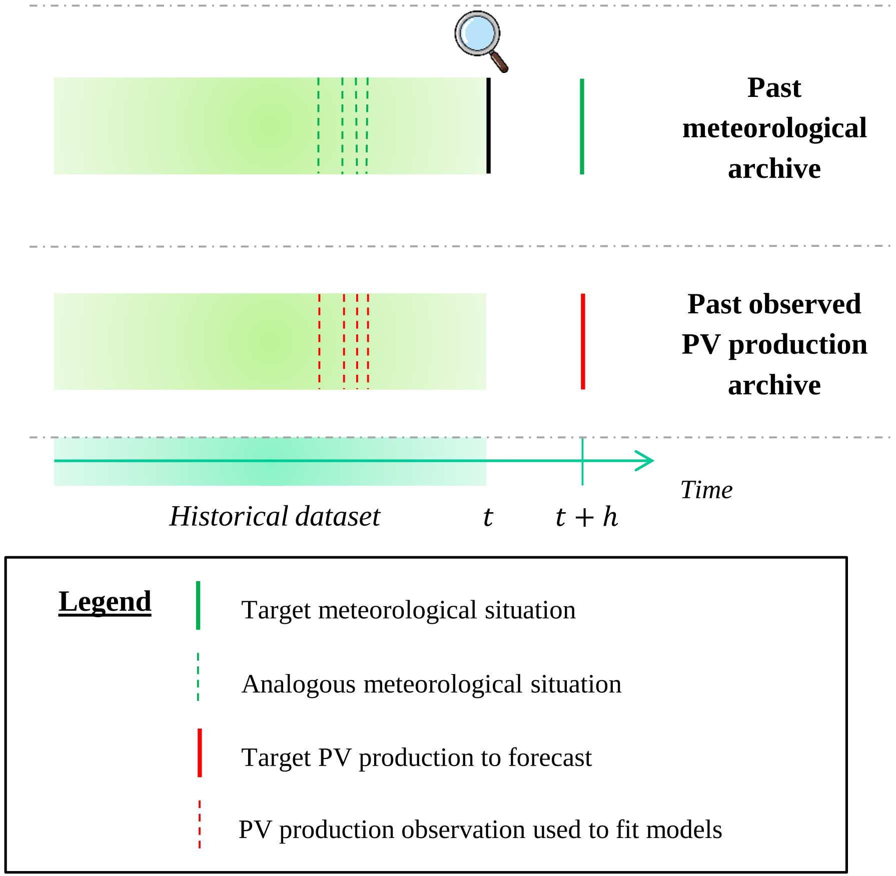
- Hypothesis: **similar PV production dynamics** are observed under **similar weather conditions**
- Thus, fitting a model on all past PV observations can **drown relevant information**
- It seems wise to **sample the learning set** to obtain a **weather coherent** PV production observations **subset** in regard with the future weather state
- To do so, an **analogy based approach** is implemented

A conditioned learning

An analog based approach:

- Construction of **2 past records** for the same period
    - **Meteorological archive**
    - **Observed PV production archive**
- Retrieve **forecast** of analog predictors **at time $t + h$**
- Compute the **analogy score** between the target meteorological situation and all the candidate situations from the meteorological archive.
- Each **meteorological situation** is **ordered** from the most similar to the less one. N **best meteorological situations** are selected.
- The N **associated PV production observations** are selected to **train the forecasting model**.
- **Forecast of PV production at time $t + h$.**

Which leads to an **adaptative learning**



An analog based approach

Analog variable

- **Geopotential field** is chosen as an **analog predictor**
- Geopotential field is considered as a **wind driver**

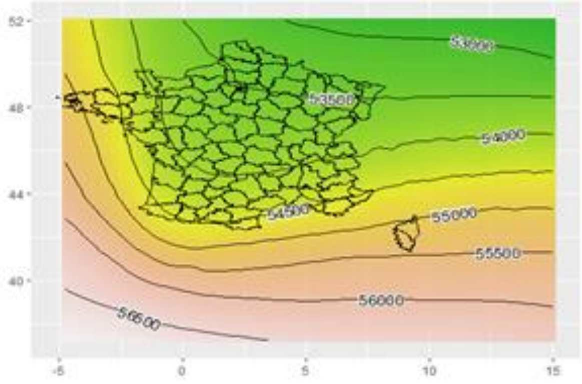
Analog metric

- Need to take into account the predictor **spatial distribution**
- Score **S1** is used to **measure similarity** between the target and the candidate situations
  - (*Teweless and Wobus 1954, Obled, 2002*)

Data integration

- ***Perfect prognosis*** mode
  - Regarding target situation, reanalysis data are used instead of predictions
  - Quantify the interest of using geopotential field as an analog predictor

Geopotential height fields



$$S_1 = 100\,\frac{\sum_{i=1}^{I-1}\sum_{j=1}^{J}\left|\Delta_{i,j}^{i,Target} - \Delta_{i,j}^{i,Candidate}\right| + \sum_{i=1}^{I}\sum_{j=1}^{J-1}\left|\Delta_{i,j}^{j,Target} - \Delta_{i,j}^{j,Candidate}\right|}{\sum_{i=1}^{I-1}\sum_{j=1}^{J}\max\left(\left|\Delta_{i,j}^{i,Target}\right|,\left|\Delta_{i,j}^{i,Candidate}\right|\right) + \sum_{i=1}^{I}\sum_{j=1}^{J-1}\max\left(\left|\Delta_{i,j}^{j,Target}\right|,\left|\Delta_{i,j}^{j,Candidate}\right|\right)}$$

Terms are defined in the annex section

Data Stationarization

Inputs stationarity

- ARIMA process requirement
- Get time series more « easily » forecastable

PV production variability

- Deterministic component: sun's path
- Stochastic component: clouds, aerosol, …

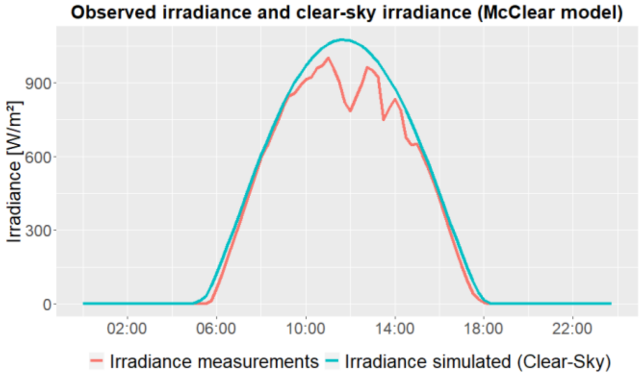
Taken into account by stationarity procedure

Methods

- Clear-sky normalization (*Bacher, 2009*), (*Agoua,2018*)
- Seasonal decomposition (*Cleveland, 1990*), (*Yang, 2012*)
- …

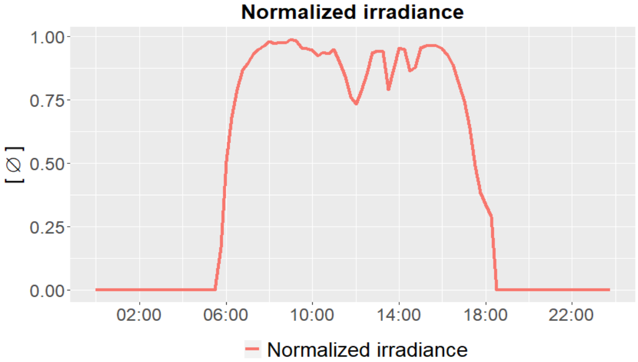
Clear sky model used:

- McClear model (Lefèvre, 2013)



Normalization

$$P^{stat}(t) = \frac{P^{obs}(t)}{P^{Clear-Sky}(t)}$$



Models definition

Performance evaluation

Score	Skill Score
$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(\widehat{P}_t - P_t\right)^2}$	$SS(h) = 1 - \frac{RMSE_{Model}(h)}{RMSE_{Reference}(h)}$

Reference model: smart persistence

$$\widehat{\overline{P}}^x_{t+h|t} = \begin{cases} \overline{P_t}^x & if\ \overline{P_t}^x \neq 0\ (i.e.\ daytime) \\ \overline{P}^x_{t+h-24H} & if\ \overline{P_t}^x = 0\ (i.e.\ nighttime) \end{cases}$$

Auto-Regressive (AR) model

$$\widehat{\overline{P}}^x_{t+h|t} = \hat{\beta}^0_h + \sum_{l=0}^{L} \hat{\beta}^l_h \overline{P}^x_{t-l}$$

Terms are defined in the annex section

Conditioned learning

A two steps conditioned learning approach (CAR model)

- First, the learning set is sample according to the **hour of the day (CAR-T)**
- Then, the previous subset is sample again in respect to the **synoptic situations** using the analogy based method **(CAR-T.An)**

Conditioned Auto-Regressive (CAR-T.An)

$$\hat{\hat{P}}^{x}_{t+h|A_{t+h}} = \hat{\beta}^{0}_{h} + \sum_{l=0}^{L} \hat{\beta}^{l}_{h}\, f_{A_{t+h}}(\bar{P}^{x}_{t-l})$$

On-site observations

CAR-T.An + Spatio-Temporal data (CARST-T.An)

$$+ \sum_{l=0}^{L} \sum_{y \in X} \hat{\beta}^{l,y}_{h} f_{A_{t+h}}(\bar{P}^{y}_{t-l})$$

$$+ \sum_{i=1}^{N} \hat{\beta}^{Sat}_{i,h} f_{A_{t+h}}(\overline{Sat}^{i}_{t})$$

+ nearby sites observations
+ satellite images

CARST-T.An + eXogenous inputs (CARXST)

$$+ \hat{\beta}^{NWP}_{h} f_{A_{t+h}}(\overline{NWP}^{x}_{t+h})$$

+ NWP

Terms are defined in the annex section

Features selection

Least Absolute Shrinkage and Selection Operator (LASSO)

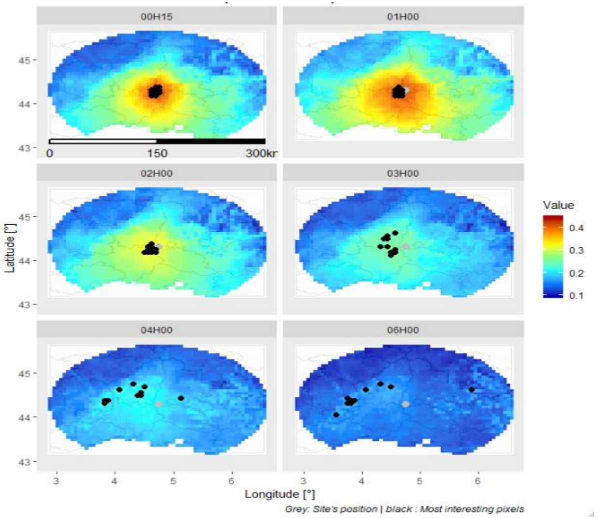
- To provide a **seamless model**, variable selection is carried out on each horizon to keep most **informative features**

$$\hat{\beta}_h^{LASSO} = \underset{\beta_h}{argmin} \left\{ \frac{1}{2} RSS(\beta_h) + \lambda |\beta_h| \right\}$$

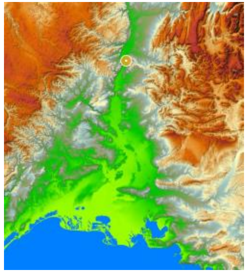
Satellite pixels selection

- To avoid a too large number of variables, we limits our approach to the **10 most informative pixels**
- To quantify the correlation between the stationarized PV production and the stationarized satellite data for various lags, the **Mutual Information Criterion** is used (Carriere, 2020).
- The correlation area is highly influenced by the Rhône valley topography.
- The shorter the horizon, the closer the most relevant pixels.
- For horizon higher than 3H00 ahead, most informative points are located westward.
    - In this region, main wind seams to be **westerly wind** (see annex)

Location of the 10 most informative pixels



Rhône valley topography



Case study

Production measurements

- 9 PV plants
- $D \in [7.3\,;133]$ km
- $\overline{P}_t = \frac{\overline{P_t}}{P_{nom}}$, with $P_{nom} \in [1.2\,;12]$ MWp
- $\Delta t = 15'$

Satellite images

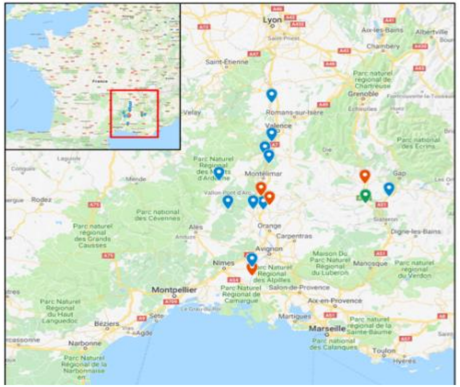
- Helioclim-3 (Blanc, 2011)
- Stationarized GHI
- $\Delta t = 15'$

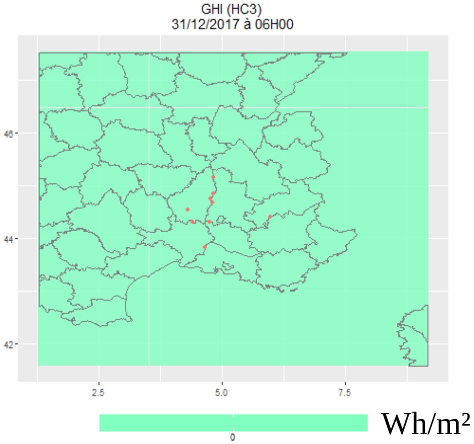
NWP (exogenous inputs)

- ARPEGE – Météo France
- Operational forecasting conditions
  - Dissemination schedule
- Stationarized GHI
- $\Delta t = 1H$ interpolated to $\Delta t = 15'$

Analog predictor

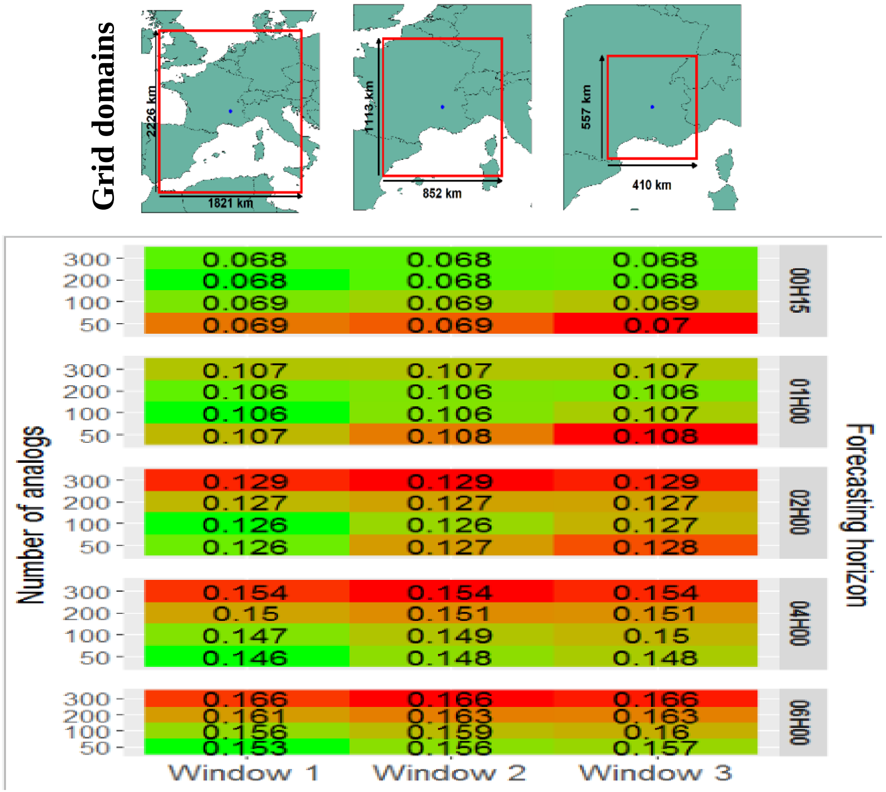
- ERA5 – ECMWF – Reanalysis
- *Perfect prognosis* mode
- Geopotential field at 500 & 925hPa
- $\Delta t = 1H$





Terms are defined in the annex section

Sensibility analysis



Parameters of the sensibility analysis

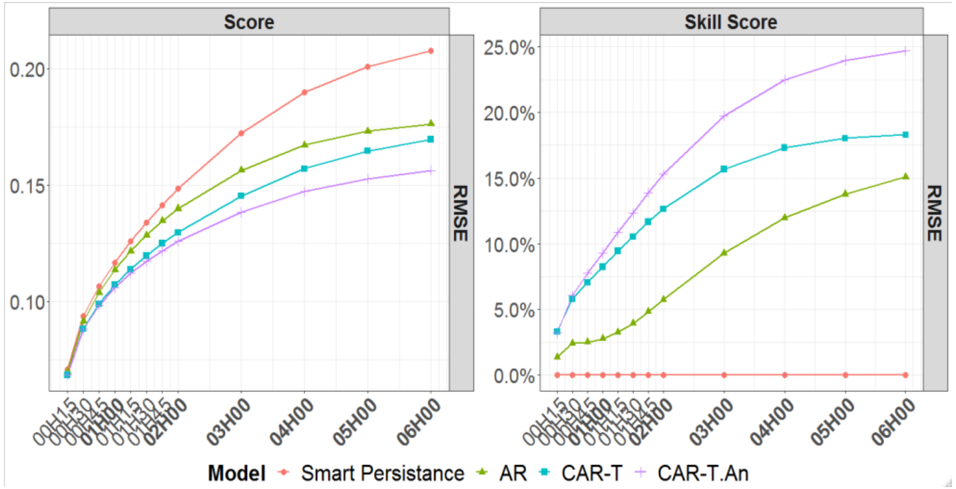
- Grid domains of the analog spatial window
  - 3 regions centered over the Rhône valley region
- Number of analogs situations to train the models
  - From 50 up to 300 analogs situations

Conclusions

- Grid domains
  - **The larger spatial window exhibits better performances.**
- Number of analogs
  - For very short-term horizons (i.e. from 15' up to 1H ahead), the more analogs, the better the performances,
  - For longer horizons (i.e. from 2H up to 6H ahead), better performances are achieved with less analogous situations.
  - **100 analog situations seems a good compromise.**

Influence of the conditioned approach over performances

Performance evaluation of the proposed conditioning approach over the reference model



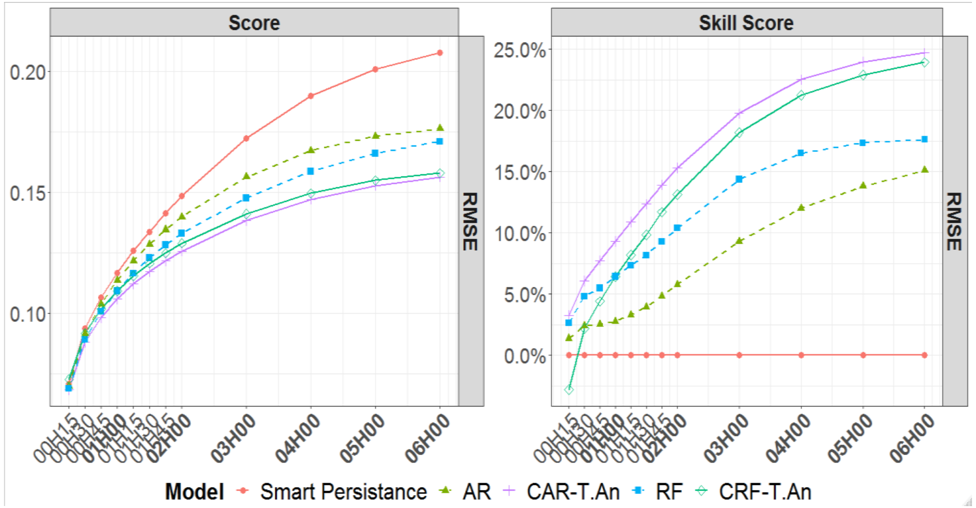
Model names are defined in the annex section

Conclusions

- AR model
  - AR model outperforms the persistence up to ~15% for a 6H horizon.
- CAR-T model
  - Conditioning of the learning set to the hour of day, improves performances in comparison with the AR model
  - This phenomenon can be explained by:
    - The stationnarisation procedure is not perfect, especially for dawn times (Agoua, 2018)
    - The PV production dynamics varies according to the time of the day
- CAR-T.An
  - The CAR-T.An model outperforms both the previous models. Compared to persistence model, improvement can reach ~24% for a 6H lead time
  - **Better performances** are obtained when forecast model **depends on the weather situation**

Comparison with a more advanced model

Performance evaluation of the proposed conditioning approach with the Random Forest model (RF)



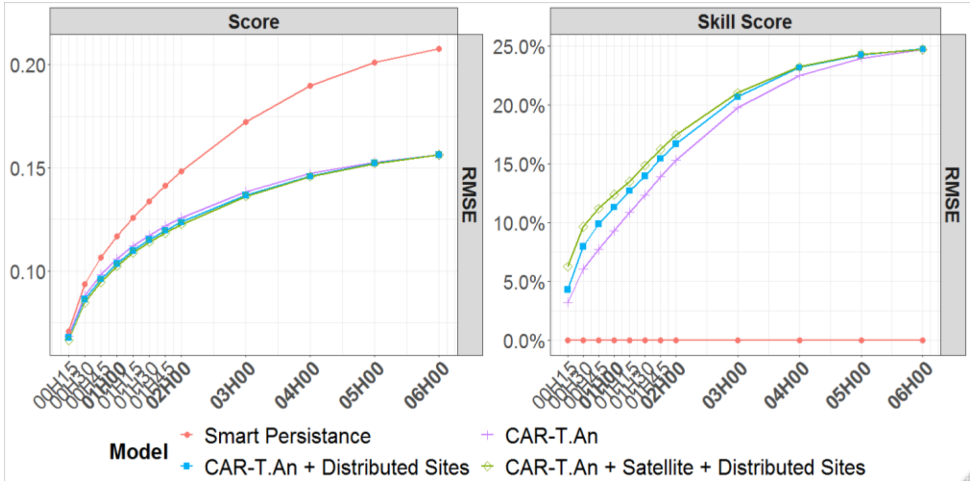
Model names are defined in the annex section

Conclusions

- AR model
  - The RF approach outperforms the AR model regardless of the considered forecasting horizon
- CAR model
  - The proposed conditioning approach outperforms the RF and CRF models.
  - Bad performances from CRF for very short times are supposed to result from over fitting

Spatio temporal inputs – CARST model

Performance evaluation of the CAR model with ST inputs



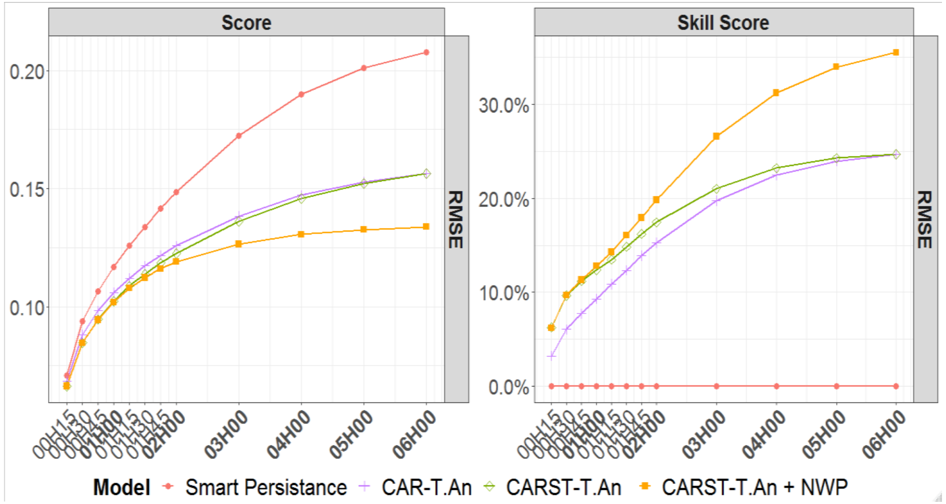
Model names are defined in the annex section

Conclusions

- CARST-T.An Model
  - **ST inputs** (i.e. distributed sites and satellite pixels) **improves performances** for horizons below 6H00 ahead.
  - **Improvement** are higher for 30’ horizon and **decrease with time**
  - At **6H00** horizon, the **influence of ST is neglectable**

NWP inputs – CARXST model

Performance evaluation of the proposed models



Model names are defined in the annex section

Conclusions

- CARST-T.An Model
  - For horizon ranging from **15’ up to 45’**, performances **improvement** result from **ST data**
- CARXST-T.An Model
  - From **1H up to 6H** ahead horizon, the main source of performance **improvement** is due to **NWP**.

Seamless way to integrate capacity of NWP outputs to extend forecasting horizon

Conclusion et perspectives

Conclusion

- The proposed **conditioned learning improve performances up to 25%** in comparison with a persistence model for a **6H ahead horizon**
- **ST data improve performances for horizon below 6H ahead,**
  - By ~4% for a 30 min horizon
  - Improvement decrease progressively to become neglectable for a 6H horizon
- Combining the proposed conditioning approach with ST and NWP inputs, performances reach ~35% for a 6H lead time in regards with the persistence model
- The **LASSO** features selection enable to propose a **seamless approach**

Perspectives

- Operational framework: consider NWP of geopotential fields rather than reanalysis
- Improve ST data integration by considering Cloud Motion Vector (CMV) approach

Thanks for your attention.

Reference

- CHOWDHURY, RAHMAN, **Forecasting sub-hourly solar irradiance for prediction of photovoltaic output**, IEEE Photovoltaic Specialists Conference, 1987
- CLEVELAND, CLEVELAND, MCRAE, STL - **A seasonal-Trend Decomposition Procedure Based on Loess**, Journal of Official Statistics , 1990
- BACHER, MADSEN, NIELSEN, **Online short-term solar power forecasting**, Solar Energy, 2009
- LORENZ, HURKA, HEINEMANN, **Irradiance Forecasting for the Power Prediction of Grid-Connected Photovoltaic Systems**, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2009
- YANG, JIRUTITIJAROEN, WALSH, **Hourly solar irradiance time series forecasting using cloud cover index**, Solar Energy, 2012
- DIAGNE, DAVID, LAURET, **Review of solar irradiance forecasting methods and a proposition for small-scale insular grids**, Renewable and Sustainable Energy Reviews , 2013
- DONG, YANG, REINDL, **Short-term solar irradiance forecasting using exponential smoothing state space model**, Energy, 2013
- HUANG, KOROLKIEWICZ, AGRAWAL, **Forecasting solar radiation on an hourly time scale using a Coupled AutoRegressive and Dynamical System (CARDS) model**, Solar Energy, 2013
- AKYUREK, KLEISSL, ROSING, **TESLA - Taylor expanded solar analog forecasting,** 2014 IEEE International Conference on Smart Grid Communications (SmartGridComm), 2014
- ALESSANDRINI, DELLE MONACHE, SPERATI, **An analog ensemble for short-term probabilistic solar power forecast**, Applied Energy, 2015
- BESSA, TRINDADE, MIRANDA, **Spatial-Temporal Solar Power Forecasting for Smart Grids**, IEEE Transactions on Industrial Informatics, 2015
- ANTONANZAS, OSORIO, ESCOBAR, **Review of photovoltaic power forecasting**, Solar Energy, 2016
- AGUIAR, PEREIRA, LAURET **Combining solar irradiance measurements, satellite-derived data and a numerical weather prediction model to improve intra-day solar forecasting**, Renewable Energy, 2016
- GRADITI, FERLITO, ADINOLFI, **Comparison of Photovoltaic plant power production prediction methods using a large measured dataset**, Renewable Energy, 2016
- AGOUA, GIRARD, KARINIOTAKIS, **Short-Term Spatio-Temporal Forecasting of Photovoltaic Power Production**, IEEE Transactions on Sustainable Energy, 2018
- KAPOURCHALI, SPEHRY, ARAVINTHAN, **Multivariate Spatio-temporal Solar Generation Forecasting: A Unified Approach to Deal With Communication Failure and Invisible Sites**, IEEE Systems Journal, 2018
- CANADILLAS, GONZALEZ-DIAS, RODRIGUEZ, **A low-cost two-camera sky-imager ground-based intra-hour solar forecasting system with cloud base height estimation capabilities working in a smart grid**, 2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC) , 2018
- VAN DER MEER, WIDEN, MUNKHAMMAR, **Review on probabilistic forecasting of photovoltaic power production and electricity consumption**, Renewable and Sustainable Energy Reviews, 2018
- CARRIERE, VERNAY, PITAVAL, KARINIOTAKIS, **A Novel Approach for Seamless Probabilistic Photovoltaic Power Forecasting Covering Multiple Time Frames**, IEEE Transactions on Smart Grid,pp. 1–1, 2019,
- BLANC, GSCHWIND, LEFEVRE, and WALD, **The HelioClimProject: Surface Solar Irradiance Data for Climate Applications**, Remote Sensing, vol. 3, no. 2, pp. 343–361, Feb. 2011
- N. N. W. Service, “Glossary - NOAA’s National Weather Service,”00000 Library Catalog: w1.weather.gov. [Online]. Available: https://w1.weather.gov/glossary/index.php?letter=h
- OBLED, BONTRON, GARCON, **Quantitative precipitation forecasts: a statistical adaptation of model outputs through ananalogues sorting approach**, Atmospheric Research, vol. 63, no.3-4, pp. 303–324, Aug. 2002.

Glossary

Abbreviation	Definition
AR	**A**uto **R**egressive
CAR	**A**uto **R**egressive model with **C**onditioned learning
CARST	**C**onditioned **A**uto **R**egressive model with **S**patio-**T**emporal inputs
CARXST	**C**onditioned **A**uto **R**egressive model with **S**patio-**T**emporal and e**X**ogenous inputs
CMV	**C**loud **M**otion **V**ector
CNR	**C**ompagnie **N**ationale du **R**hône
GHI	**G**lobal **H**orizon **I**rradiance
LASSO	**L**east **A**bsolute **S**hrinkage and **S**election **O**perator
NWP	**N**umerical **W**eather **P**redictions
PV	**P**hoto**V**oltaic
ST	**S**patio **T**emporal
SS	**S**kill **S**core
RF	**R**andom **F**orest

Model name definition

Abbreviation	Definition
AR	**A**uto **R**egressive model considering all available PV production observations
RF	**R**andom **F**orest model considering all available PV production observations
CAR	**A**uto **R**egressive model with **C**onditioned learning
CAR-T	**AR** model **C**onditioned to the **T**ime of the day
CAR-T.An	**AR** model conditioned to the **T**ime of the day and the synoptic state through an **An**alog based method
CRF-T.An	**RF** model conditioned to the **T**ime of the day and the synoptic state through an **An**alog based method
CARST-T.An	**C**onditioned **A**uto **R**egressive model with **S**patio-**T**emporal inputs
CARXST-T.An	**C**onditioned **A**uto **R**egressive model with **S**patio-**T**emporal and e**X**ogenous inputs

S1 score

Definition of the S1 score

$$S_1 = 100\,\frac{\sum_{i=1}^{I-1}\sum_{j=1}^{J}\left|\Delta_{i,j}^{i,Target} - \Delta_{i,j}^{i,Candidate}\right| + \sum_{i=1}^{I}\sum_{j=1}^{J-1}\left|\Delta_{i,j}^{j,Target} - \Delta_{i,j}^{j,Candidate}\right|}{\sum_{i=1}^{I-1}\sum_{j=1}^{J}\max\left(\left|\Delta_{i,j}^{i,Target}\right|,\left|\Delta_{i,j}^{i,Candidate}\right|\right) + \sum_{i=1}^{I}\sum_{j=1}^{J-1}\max\left(\left|\Delta_{i,j}^{j,Target}\right|,\left|\Delta_{i,j}^{j,Candidate}\right|\right)}$$

$$\begin{cases}\Delta_{i,j}^{i,X} = V_{i+1,j}^{X} - V_{i,j}^{X} \\ \Delta_{i,j}^{j,X} = V_{i,j+1}^{X} - V_{i,j}^{X}\end{cases} \quad X \in \{Target, Candidate\}$$

<u>Where:</u>

- $\Delta_{i,j}^{i}$ – The east-west geopotential gradient
- $\Delta_{i,j}^{j}$ – The north-south geopotential gradient
- $V_{i,j}$ – The geopotential field at grid node (i,j)

Models parameters definition

Parameters definition

- $\bar{P}_t^x$ – Observed stationarized PV production at time *t* for plant *x*
- $\hat{\bar{P}}_{t+h}^x$ – Predicted stationarized PV production at time *t+h* for plant *x*
- A_{t+h} – Synoptic weather situation expected at time *t+h*
- $f_{A_{t+h}}$ – Conditioning approach based on synoptic situation and time of the day
- $\hat{\beta}$ – Regression coefficients
- L – Maximum lag of PV production observations (here, 2H)
- X – Neighbouring sites (here, 5 closest sites are considered)
- $\overline{Sat}_t^i$ – Stationarized GHI observed at pixel *i* and time *t*
- $\overline{NWP}_{t+h}^x$ – Stationarized GHI forecast obtained from NWP output at plant *x* and time *t+h*

Skill score definition

$$SS(h) = \frac{RMSE_{Model}(h) - RMSE_{Reference}(h)}{\underbrace{RMSE_{Perfect\ model}(h)}_{=0\ \forall h} - RMSE_{Reference}(h)}$$

Main wind in the Rhône valley

Principal Component Analysis (PCA)

- **PCA** is performed on 1 years of hourly reanalysis (ERA 5) of wind velocity at 750hPa
- The first Principal Component (PC) is represented at the opposite graph
- **Westerly wind** are observed