Photovoltaic power generation forecasting models analysis

Data driven model VS Theory driven models

BERMUDEZ Yaiza
Grenoble, France
yaiza.bermudez@inp-grenoble.org
Supervised by: MOUNIE Gregory.

Research internship, Juin 2023

Outline

1. Context
2. State of the art
3. The data driven models
4. The theory driven models
5. Evaluation
6. Results and discussion
7. Conclusion
8. References

Outline

1. Context
2. State of the art
3. The data driven models
4. The theory driven models
5. Evaluation
6. Results and discussion
7. Conclusion
8. References

Context

Figure 1: Panels on GreEN-ER site [1]

[1] Image from : https://www.univ-grenoble-alpes.fr/inauguration-de-la-centrale-photovoltaique-de-green-er-1re-experi kjsp

Outline

1. Context
2. State of the art
3. The data driven models
4. The theory driven models
5. Evaluation
6. Results and discussion
7. Conclusion
8. References

Data driven and Theory driven models

Authors	Algorithm	Description	Metrics
[Pedro and Coimbra, 2015]	ANN and KNN	ANN was combined with k-nearest-neighbours (kNN) optimization to forecast global solar irradiance for time horizons from 15 min to 2h.	RMSE(%)<15.00
[Sharma et al. 2016]	WNN	A feed-forward ANN with Wavelet model was developed with Morlet and Mexican hat type activation functions. Its predictions were compared with persistence, ETS, ARIMA and ANN for temporal horizons of 15 min and 24h	nRMSE (%) 9.42 to 15.41
[Qing and Nui, 2018]	LSTM	A LSTM model developed for day ahead prediction of solar irradiance from weather data. Its performance is better than persistence, LLSR and multilayered FFNN-BPNN models.	RMSE 18.34%
[Dimitropoulos et al, 2021]	LSTM	LSTM is an artificial recurrent neural network (RNN) architecture used in the field of deep learning. LSTM has feedback connections and can process entire sequences of data.	RMSE 1.920
	SVM	Support-vector machines are supervised learning models with associated learning algorithms that analyse data for classification and regression analysis.	RMSE 2.014
	MLR	Linear regression is a linear approach to modelling the relationship between a scalar response and one or more explanatory variables.	RMSE 1.938
	XGBoost	Gradient ML technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. When a decision tree is the weak learner, the resulting algorithm is called gradient boosted trees.	RMSE 1.871

Figure 2: Current models and their accuracy

Clear-sky models :

- Bird $RMSE \leq 5\%$
- Simplified Solis $RMSE \leq 10\%$
- Ineichen-Perez $RMSE \leq 10\%$

Irradiance models :

- Hay Davies
- Perez
- Klucher
- Isotropic

Outline

1. Context
2. State of the art
3. The data driven models
4. The theory driven models
5. Evaluation
6. Results and discussion
7. Conclusion
8. References

Extreme Gradient Boosting

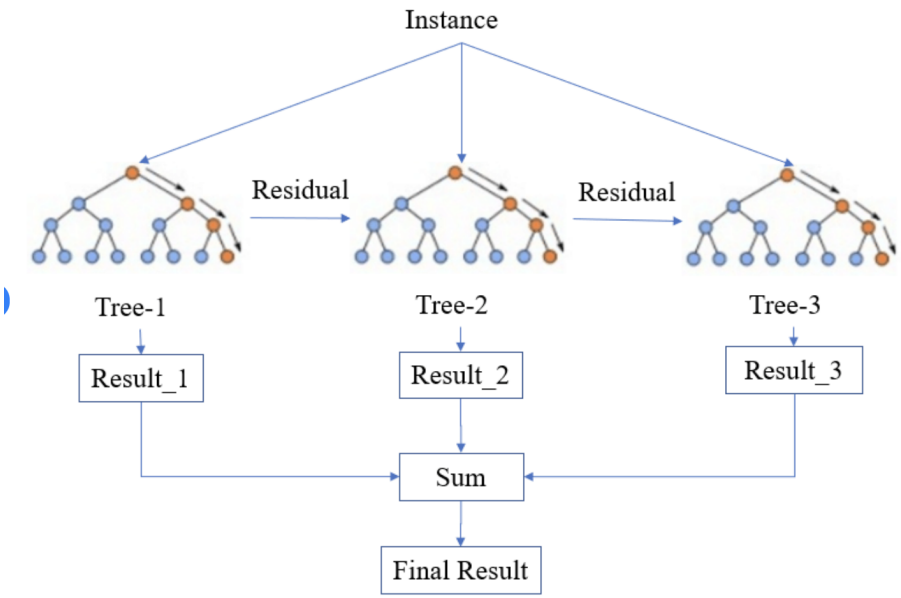

Figure 3: A general architecture of XBoost [2]

Image from:
https://www.researchgate.net/figure/Simplified-structure-of-XGBoost_fig2_348025909

Characteristics:

- n_estimators = 50
- max_depth = 5
- learning_rate = 0.03

Data from 2021-2022:

- Temperature
- Humidity
- Pressure
- Solar irradiation
- Power production

Long-Short Term Memory

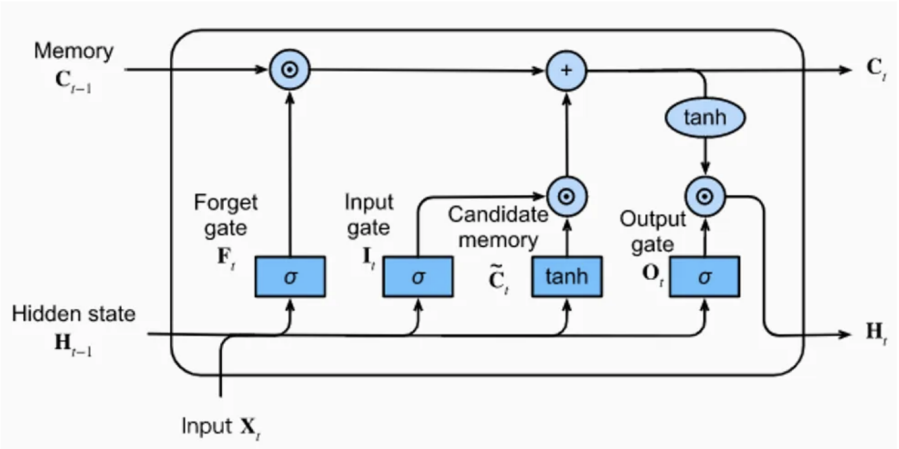

Figure 4: A general architecture of LSTM [3]

The model:

- LSTM layer of 128
- Dense layer with ReLu

Image from: https://medium.com/@ottaviocalzone/an-intuitive-explanation-\of-lstm-a035eb6ab42c

Characteristics

- Earlycallback with patience of 10
- Batch size of 32
- 50 epochs
- loss function : MSE
- optimizer : Adam
- metric RMSE

Data from 2021-2022:

- Historical power production

N-Beats

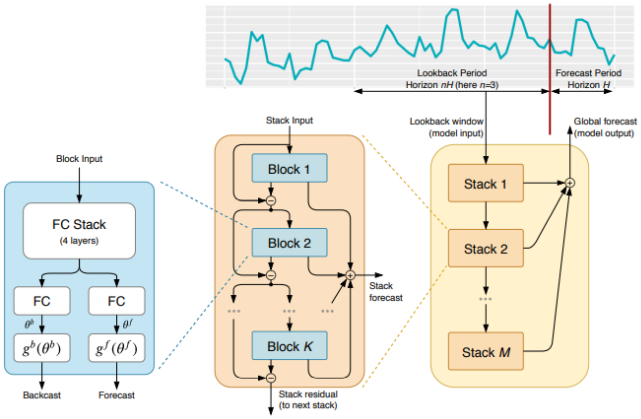

Figure 5: Architecture of the N-BEATS model[4]

The model:

- 5 stacks of 5 block each
- 1 block has 4 dense residual layers of 64 neurons and 2 output dense layers of 24 neurons

Image from : [Oreshkin *et al.*2020]

Characteristics :

- 20 epochs
- batch size = 64
- optimizer = Adam
- loss function = MSE

Data from 2021-2022:

- Historical power production

Outline

1. Context
2. State of the art
3. The data driven models
4. The theory driven models
5. Evaluation
6. Results and discussion
7. Conclusion
8. References

Ineichen-Perez

Global Horizontal Irradiance (GHI), Direct Normal Irradiance (DNI)

$$GHI = a_1 \cdot DNI_{TOA} \cdot \cos(\theta_Z) \cdot e^b \tag{1}$$

Where:

$$b = -a_2 \cdot AM \cdot (fh_1 + fh_2 \cdot (T_L - 1)) \tag{2}$$

DNI_{TOA} : Direct Normal extraterrestrial Irradiance,
AM : absolute atmospheric air mass,
TL : linke turbidity from [Ineichen and Perez2002],
h : altitude above sea level,
a_1, a_2, fh_1, fh_2 : constants based on h,
θ_z : zenith angle.

Simplified Solis

Direct Horizontal Irradiance (DHI)

$$DHI = GHI - DNI \times \cos(\theta_Z) \tag{3}$$

$$GHI = DNI'_{TOA} \cdot e^{\frac{-\tau_g}{\cos(\theta_Z^g)}} \cdot \cos(\theta_Z) \tag{4}$$

Where :

DNI'_{TOA} : modified version of DNI_{TOA},
τ_g : total optical depth,
θ_z : zenith angle.

PVlib to calculate all the needed parameters as well as (1), (4).

PVlib : Theory driven models

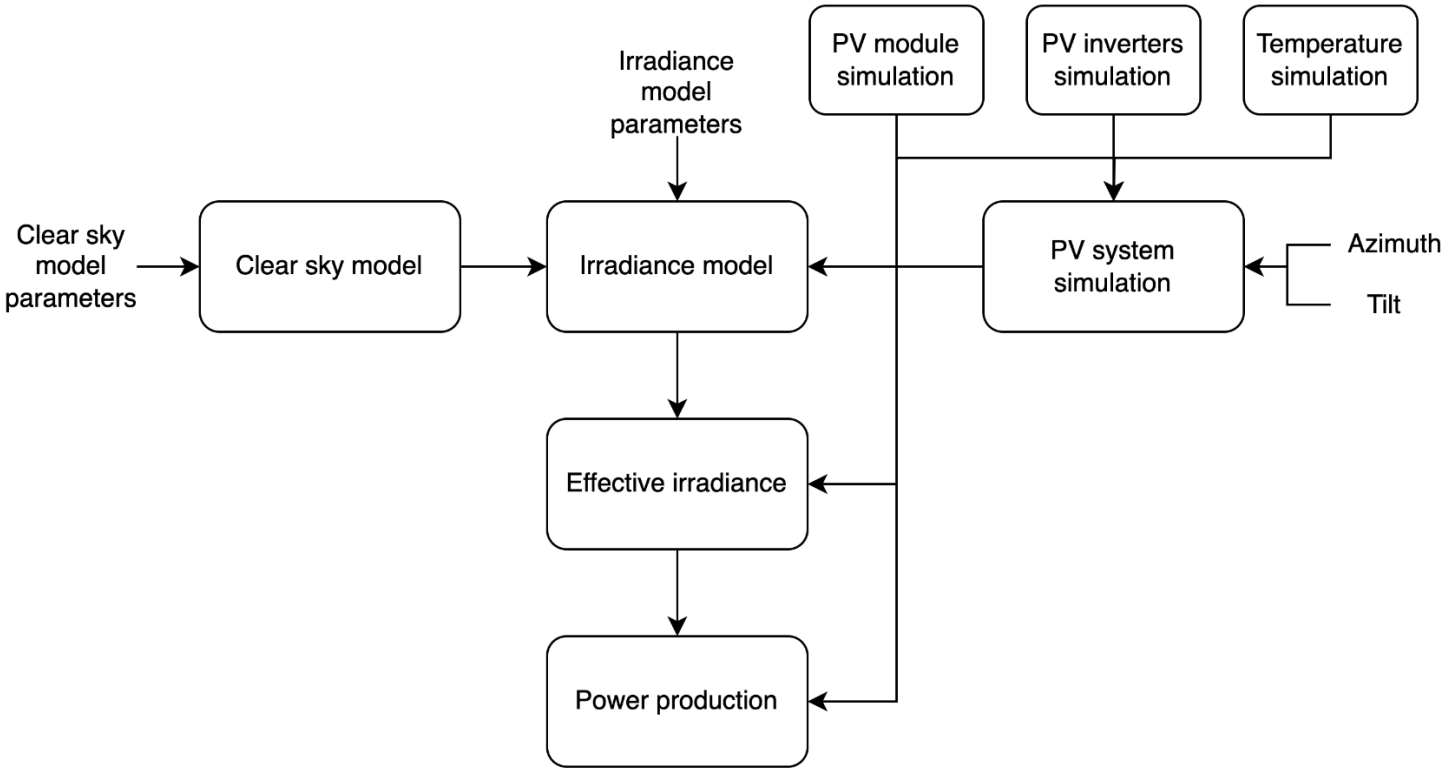

Figure 6: Diagram of the steps to compute solar power production using the physical formulas in pvlib

Outline

1. Context
2. State of the art
3. The data driven models
4. The theory driven models
5. Evaluation
6. Results and discussion
7. Conclusion
8. References

Evaluation

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \qquad (5)$$

Where N is the number of data points, y_i is the true value and $\hat{y}_i$ is the predicted value.

Percentage :

$$\frac{RMSE \times 100}{y_i.max()} \qquad (6)$$

Other criterion :

- Carbon footprint
- Adaptability
- Accessibility

Outline

1. Context
2. State of the art
3. The data driven models
4. The theory driven models
5. Evaluation
6. Results and discussion
7. Conclusion
8. References

Results : XGBoost vs LSTM

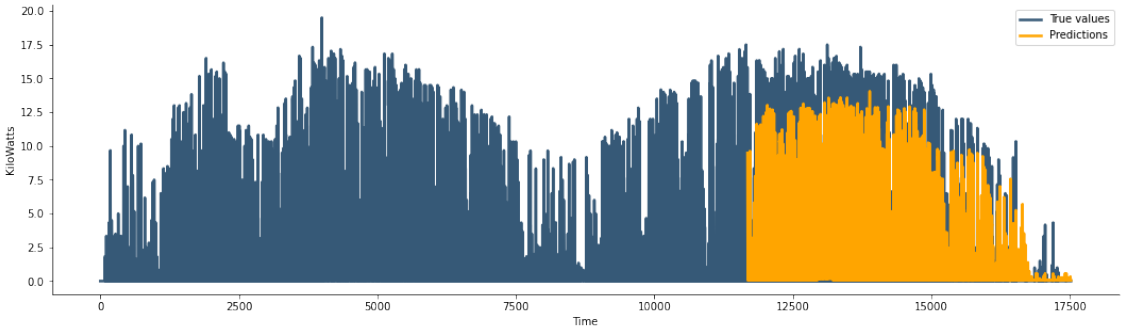

Figure 7: 1h time step predictions of XGBoost trained 2021

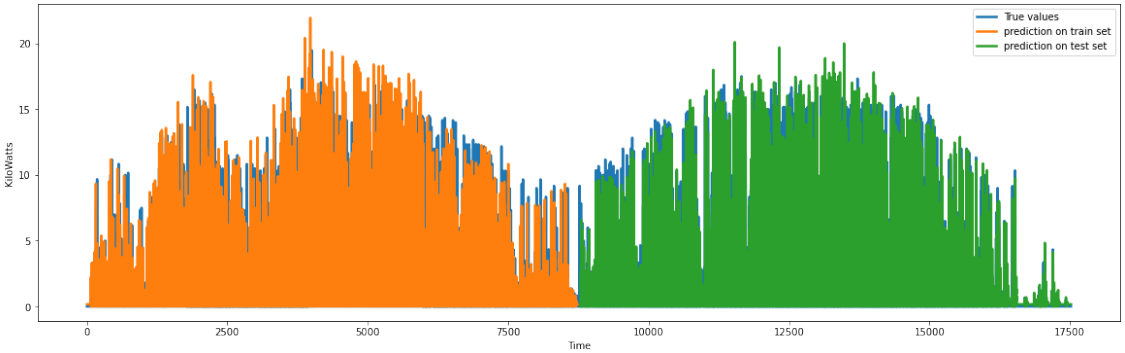

Figure 8: LSTM predictions on 2022 on the smaller panels

Results : LSTM vs N-BEATS

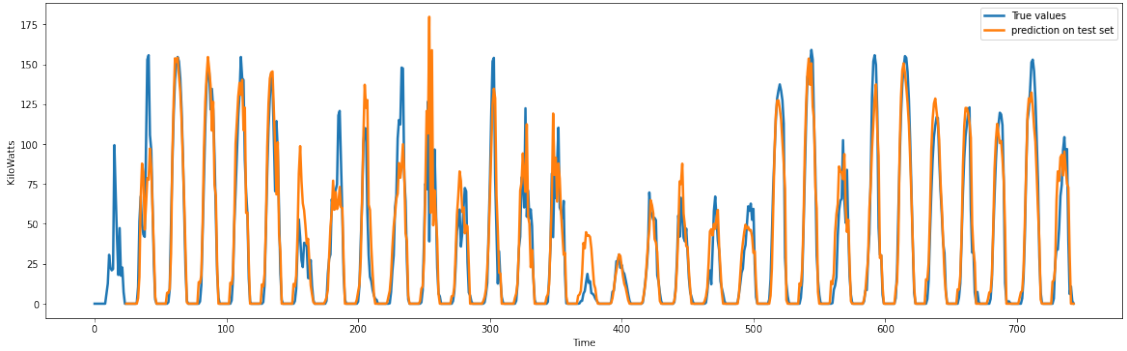

Figure 9: LSTM predictions on 2022-2023 on the roof panels

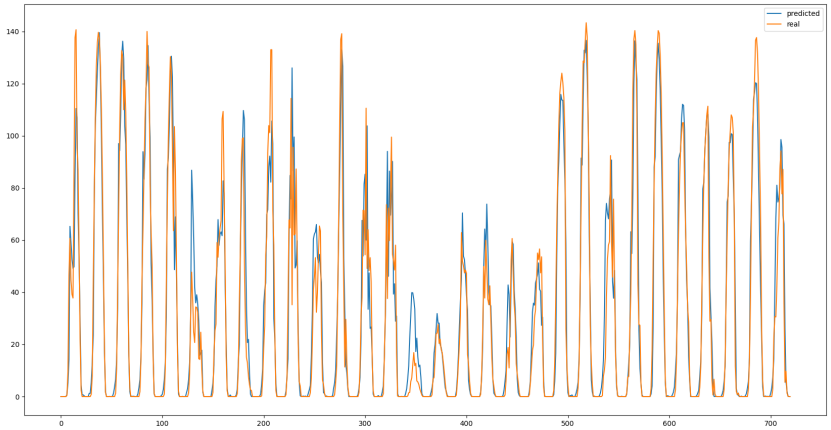

Figure 10: N-BEATS predictions 2022-2023 data for the roof panels

Results : Data vs Theory

Model	Train set on small panels	Test set	RMSE
XGBoost	2020-2021	2022	2.63 15.04%
	2021	2022	2.86 16.35%
LSTM	2021	2022	2.51 14.36%
		roof 2022-2023	16.69 11.63%
N-Beats	2021	2022	1.74 9.94%
		roof 2022-2023	13.49 9.40%

Figure 11: Comparison of the data model based on their RMSE

Clear sky	Get - irradiance	RMSE
Ineichen	haydavies	30.017% RMSE : 43.98
	isotropic	29.658% RMSE : 43.46
Simplified solis	haydavies	33.39% RMSE : 48.93
	isotropic	32.967% RMSE : 48.31

Figure 12: Comparison of the physical model based on their RMSE

Discussion

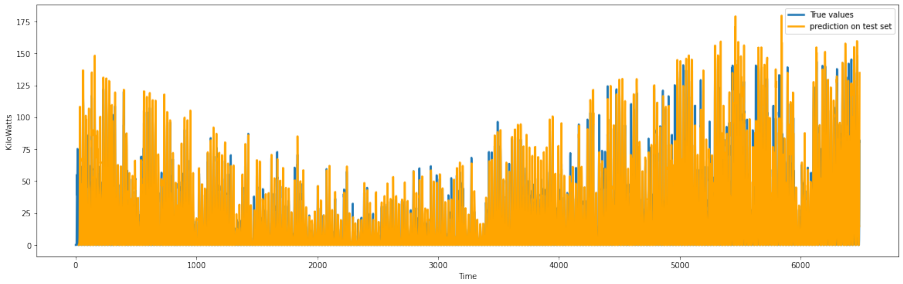

Figure 13: LSTM predictions over 2023 data

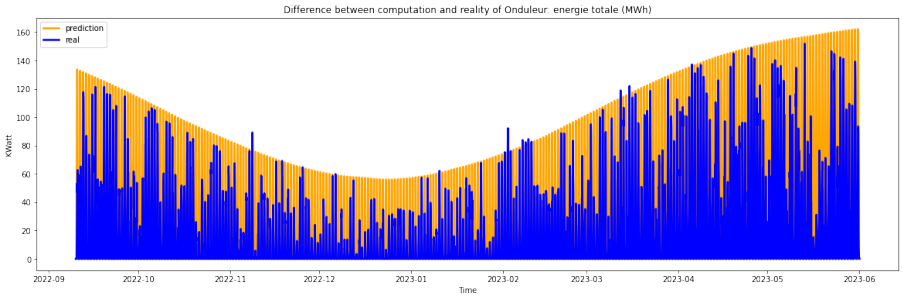

Figure 14: Ineichen-Perez with Isotropic irradiance model over 2023 data

Data driven models :

- better at catching the ungoing trends
- higher carbon footprint
- require available data

Theory driven models :

- better understanding of the physics behind
- less accurate
- require less data

Outline

1. Context
2. State of the art
3. The data driven models
4. The theory driven models
5. Evaluation
6. Results and discussion
7. Conclusion
8. References

Conclusion

- Data driven are more accurate but require data
- Data driven are less adaptable
- Theory driven have a lower carbon footprint
- Hybrid by predicting the next error

Outline

1. Context
2. State of the art
3. The data driven models
4. The theory driven models
5. Evaluation
6. Results and discussion
7. Conclusion
8. References

Bibliography I

Pierre Ineichen and Richard Perez.
A new airmass independent formulation for the linke turbidity coefficient.
Solar Energy, 73(3):151–157, September 2002.

Boris N. Oreshkin, Dmitri Carpov, Nicolas Chapados, and Yoshua Bengio.
N-beats: Neural basis expansion analysis for interpretable time series forecasting.
In *Proceedings of a conference paper at ICLR*, 2020.