Artificial neural networks for forecasting the 24 hours ahead of global solar irradiance

Mathematical Modeling with Applications (M2A19)

Hamid Ettayyebi & Khalid El Himdi
University of Mohammed V
Department of mathematics
Lab: LMSA

Talk Outlines

- I. Introduction :
- II. Forecast Methodology :
    1. Pre-processing
    2. Time series methods
    3. A machine learning framework
- III. Results and discussions
- IV. Conclusion

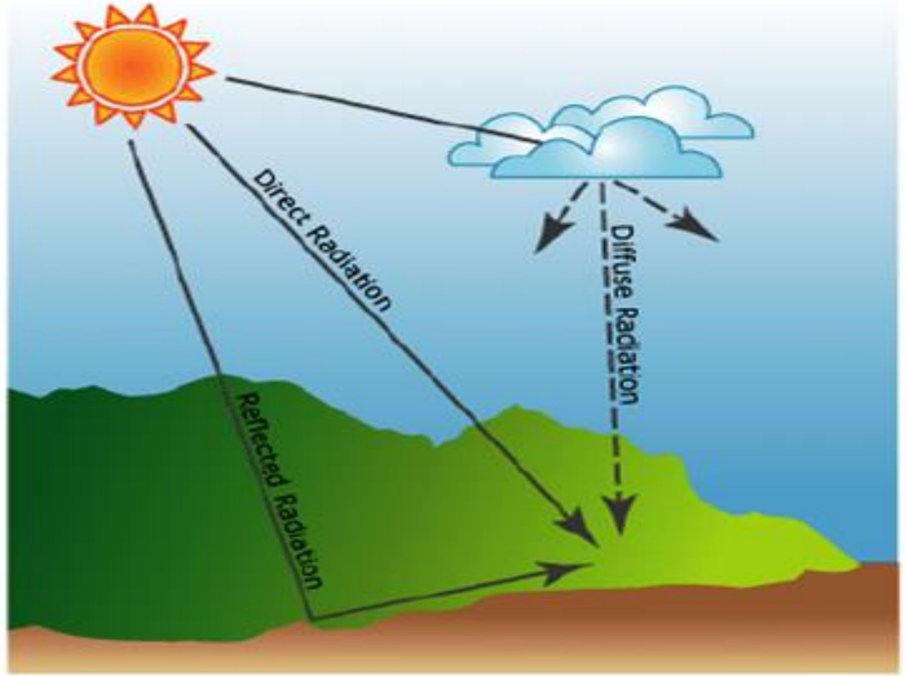
I. Introduction : what is global solar irradiance?

- Direct Normal Irradiance (DNI)

- Diffuse Horizontal Irradiance (DHI)

Source: nrel.gov

- Global Horizontal Irradiance (GHI)



Source: esri.com

Forecasting global horizontal irradiance (GHI) is the first and most essential step in most PV power prediction systems.

I. Introduction : why forecasting global solar irradiance?

- The contribution of photovoltaic systems power production to the electric power supply is constantly increasing.
- Transmission system operators have to deal with the fluctuating input from PV system energy sources.
- This is a new challenge compared with power production from conventional power plants that can be adjusted to the expected load profiles.
- An efficient use of the fluctuating energy output of PV systems requires reliable **forecast information**.
- These **predictions** are used by utility companies, transmission system operators, energy service providers, energy traders, and independent power producers in their scheduling, dispatching and regulation of power.
- An efficient **forecasting method** will help the grid operators to better manage the electrical balance between demand and power generation.

I. Introduction : Forecast Methods

Forecasting methods based on historical data of solar irradiance are two categories:

- statistical methods : Seasonality analysis, Box-Jenkins or Auto Regressive Integrated Moving Average (ARIMA), Multiple Regressions and Exponential Smoothing are examples of statistical models.
- Machine learning : AI algorithms include fuzzy inference systems, genetic algorithm, Neural Networks, Bayesian Networks, etc.

II. Forecast Methodology

Data

- Endogenous : GHI(t)
- Exogenous : T, RH…

Pre-preprocessing

- Filtration
- Missing Data
- Stationarity
- Normalization

Modelisation

- SARIMA
- GARCH
- ANN
- ANN-X

Statistical Model (SARIMA)

Machine learning (ANN)

Prediction

- Forecasting the deseariable hours

Evaluation

- Forecasting Error

II. Forecast Methodology

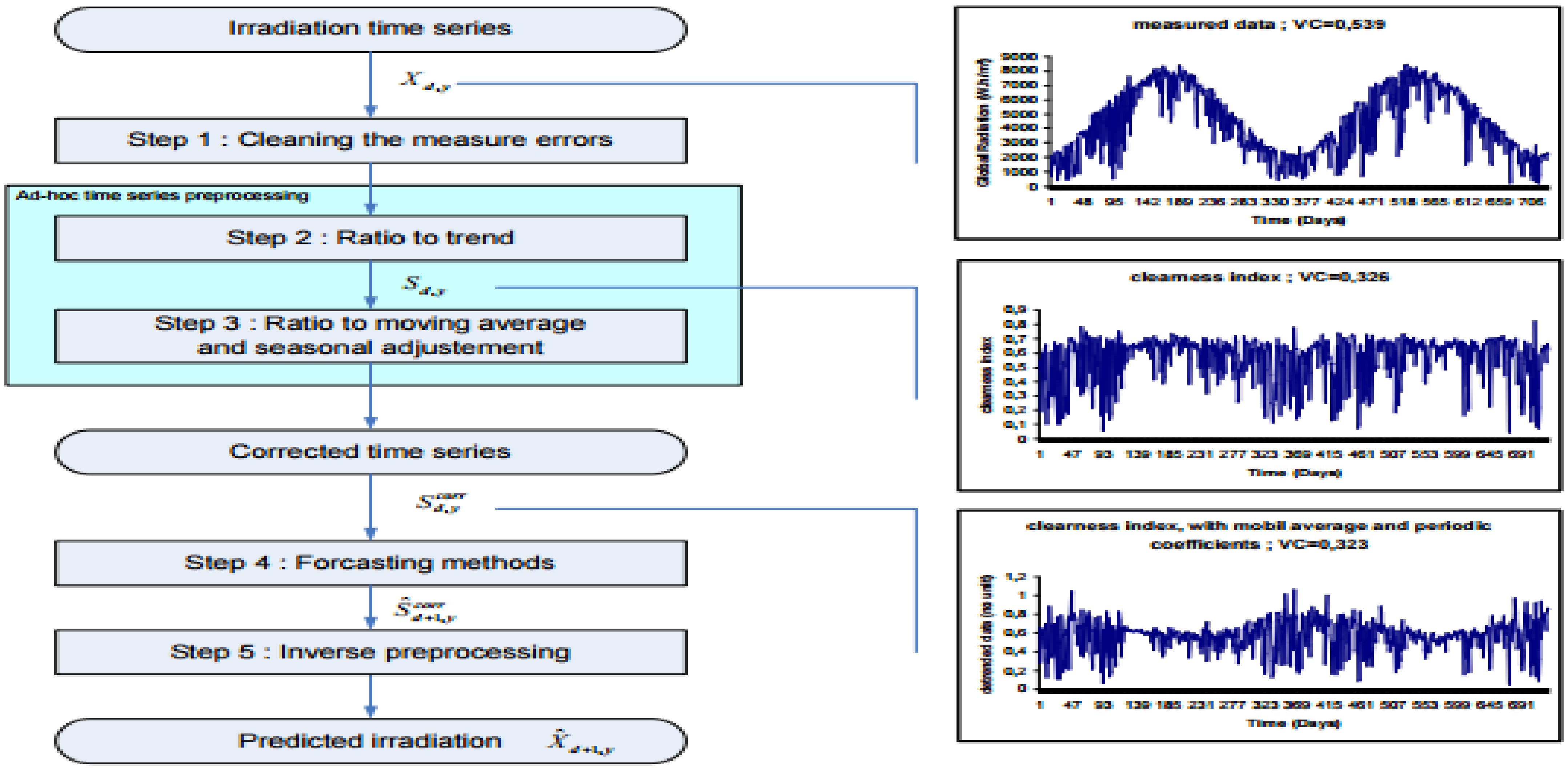


Fig. 5. Summarize of the protocol followed to obtain the predicted irradiation.

II. Forecast Methodology
SARIMA model

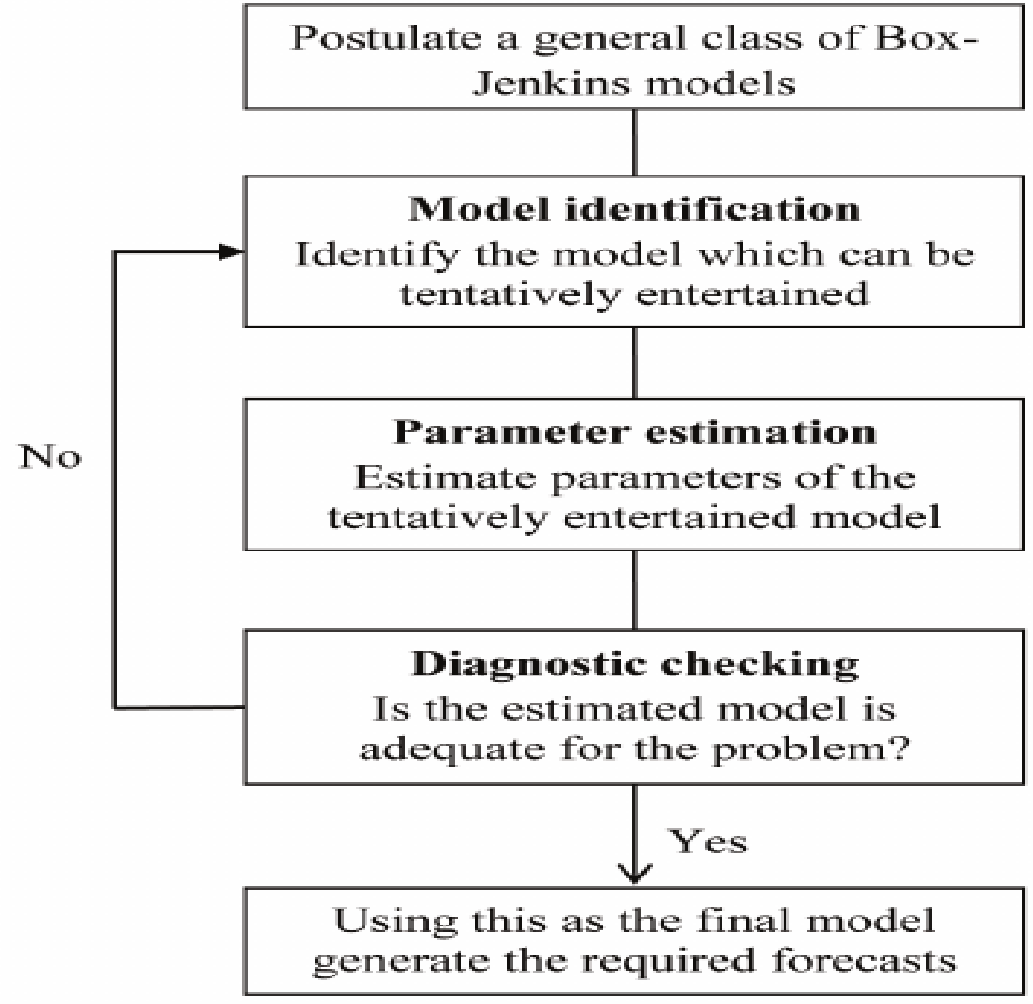
Definition 1

For a time series K_t, a SARIMA$(p,d,q)(P,D,Q)^s$ model is defined as the following :

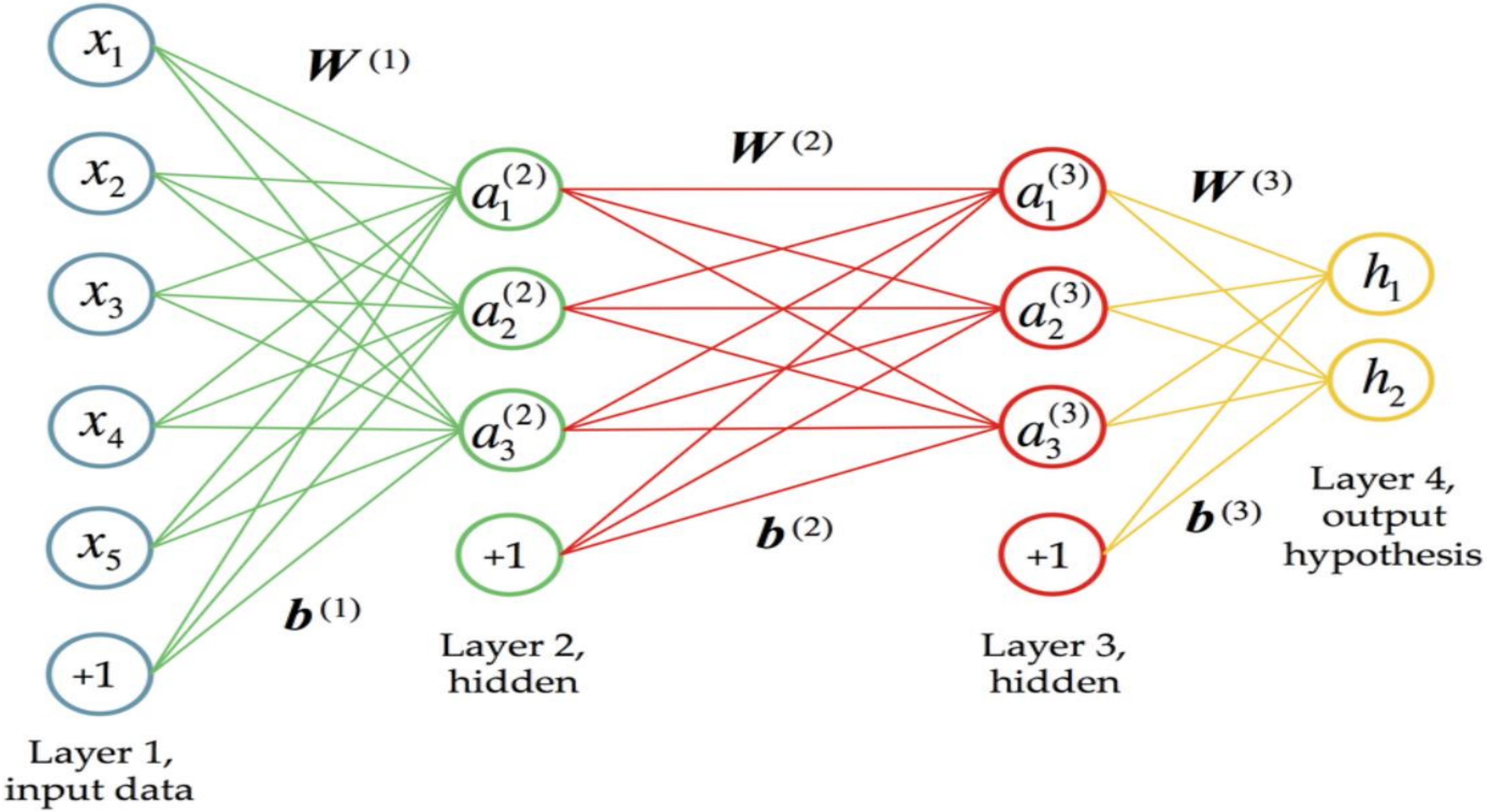
$$\Phi_P(B^s)\phi_p(B)(1-B)^d(1-B^s)^D K_t = \theta_q(B)\Theta_Q(B^s)\epsilon_t, \qquad (1)$$

where,

$$\phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p),$$
$$\theta_q(B) = (1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q),$$
$$\Phi_P(B^s) = (1 + \Phi_1 B^s + \ldots + \Phi_P B^{s+P}),$$
$$\Theta_Q(B^s) = (1 + \Theta_1 B^s + \ldots + \Theta_Q B^{s+Q}).$$



Ⅱ. Forecast Methodology
Artificial neural networks (ANN)



II. Forecast Methodology
MultiLayer Perceprton (MLP)

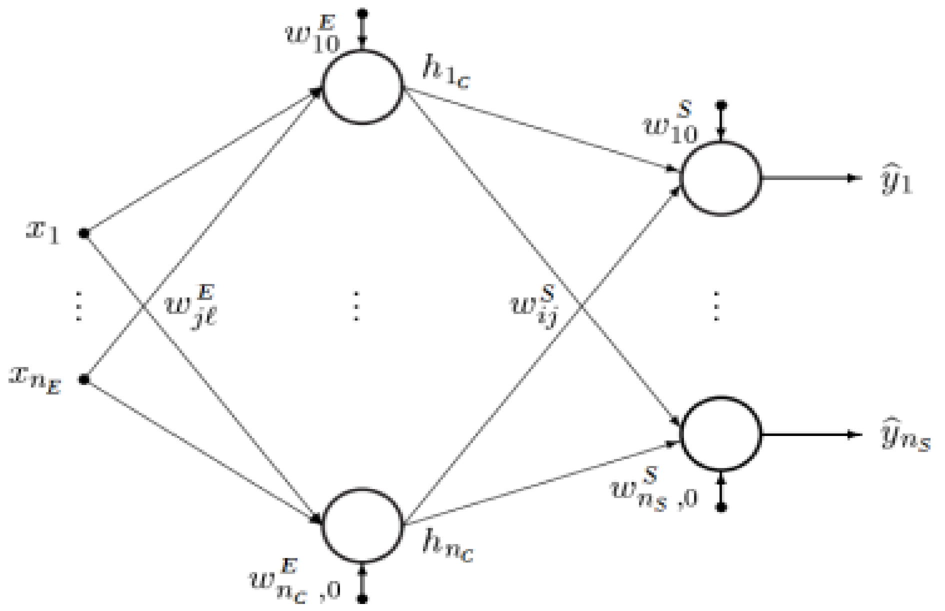


Figure 1 – The structure of a MLP with a single hidden layer.

$$h_j(x) = f\left(\sum_{l=1}^{n_E} \omega_{jl}^E x_l + \omega_{j0}^E\right)$$

$$\hat{y}_i(x) = g\left(\sum^{n_C} \omega_{ij}^S h_j(x) + \omega_{i0}^S\right)$$

The forecast of our time series at the forecast horizon h is given by the equation (4) :

$$\hat{y}_{t+h} = \sum_{j=1}^{n_C} \hat{\omega}_j^S f\left(\sum_{l=1}^{n_E} \hat{\omega}_{jl}^E \hat{y}_{t-l+h} + \hat{\omega}_{j0}^E\right) + \hat{\omega}_0^S, \qquad (4)$$

where $\hat{\omega}_j^S, \hat{\omega}_{jl}^E, \hat{\omega}_{j0}$ and $\hat{\omega}_0^S$ are the network weights after training, and f is the activation function.

II. Forecast Methodology
MultiLayer Perceprton (MLP)

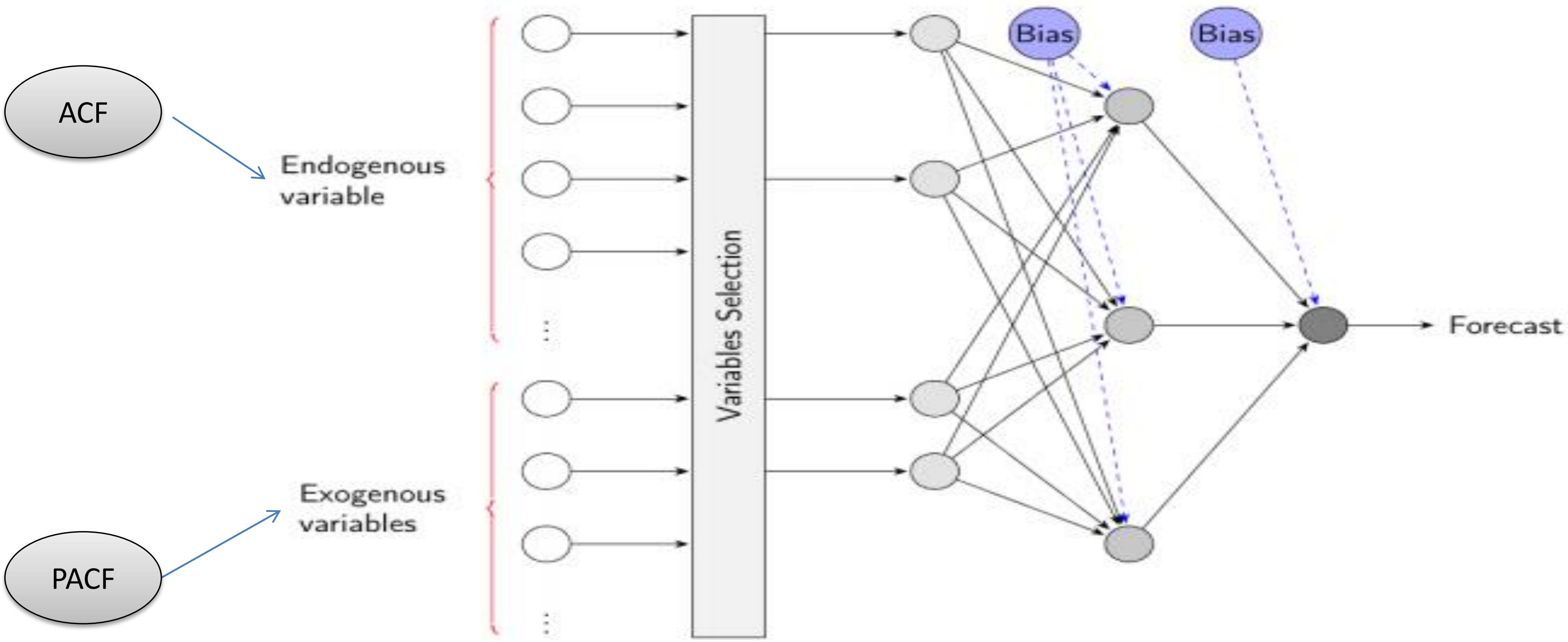


Figure 2 – The forecast methodology based on the MLP.

Performance Metrics

Mean Absolute Percentage Error (MPAE) :

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{e_t}{y_t}\right| \times 100$$

Root Mean Squared Error (RMSE) :

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2}$$

normalized Root Mean Squared Error (nRMSE)

$$\text{nRMSE} = \frac{RMSE}{\bar{y}}$$

III. Results and discussions

The region Rabat

- The dataset of hourly GHI for the region **Rabat** used in this study was provided by The Copernicus Atmosphere Monitoring Service.
- Hourly meteorological dataset was provided by MERRA-2* hosted by NASA.
- The model ARIMA was used to forecast the missing data.

*Modern-Era Retrospective analysis for Research and Applications, version 2.

III. Results and discussions

The region Rabat

Variable	Base	Unity	Period
GHI	Hourly	KWh/m^2	from 2012-01-01 to 2015-12-31
GHI$_{\text{clear-sky}}$	Hourly	KWh/m^2	from 2012-01-01 to 2015-12-31
Temperature (T)	Hourly	C^o	from 2012-01-01 to 2015-12-31
Relative Humidity(RH)	Hourly	%	from 2012-01-01 to 2015-12-31
Pressure (P)	Hourly	hPa	from 2012-01-01 to 2015-12-31
Wind speed (WS)	Hourly	m/s	from 2012-01-01 to 2015-12-31
Wind direction (WD)	Hourly	deg	from 2012-01-01 to 2015-12-31
Rainfall (R)	Hourly	kg/m^2	from 2012-01-01 to 2015-12-31

Tableau 1 – Characteristics of the datasets used in this study.

III. Results and discussions

Time series pre-processing

- Dataset filtration.
- Using the McClear model to product a new time series : the so-called clear-sky index k_t ;

$$k_t = \frac{GHI(t)}{GHI_{clear\text{-}sky}(t)}, \tag{2}$$

-
  - Training : from 01-01-2012 to 29-12-2015.
  - Testing : the last two days.

III. Results and discussions

Time series pre-processing

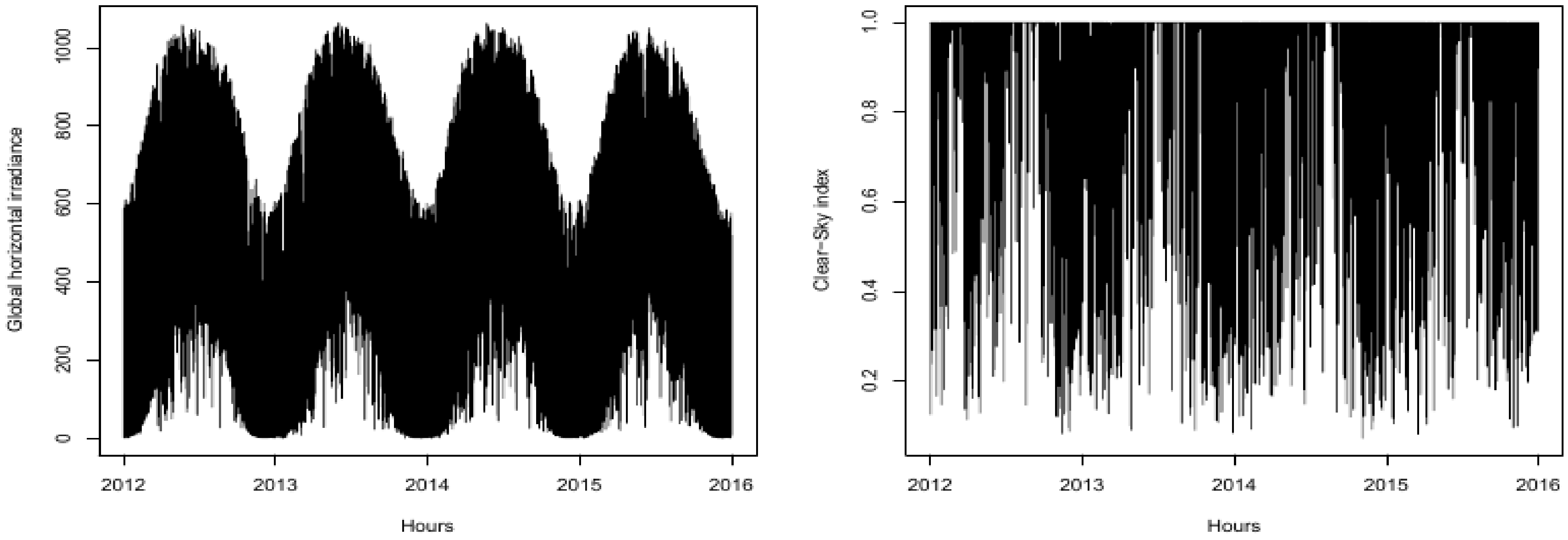


Figure 3 – Hourly global horizontal irradiance before (left) and after the transformation into the clear-sky index (right).

III. Results and discussions

MLP modelization : univariate case

The Partial autoregressive correlation function (PACF) determines the correlation between K_t and K_{t+k} after their linear dependency on the intervening variables $K_{t+1}, K_{t+2}, \ldots, K_{t+k-1}$ has been removed.

Determination of the number of time lags for the endegenous variable

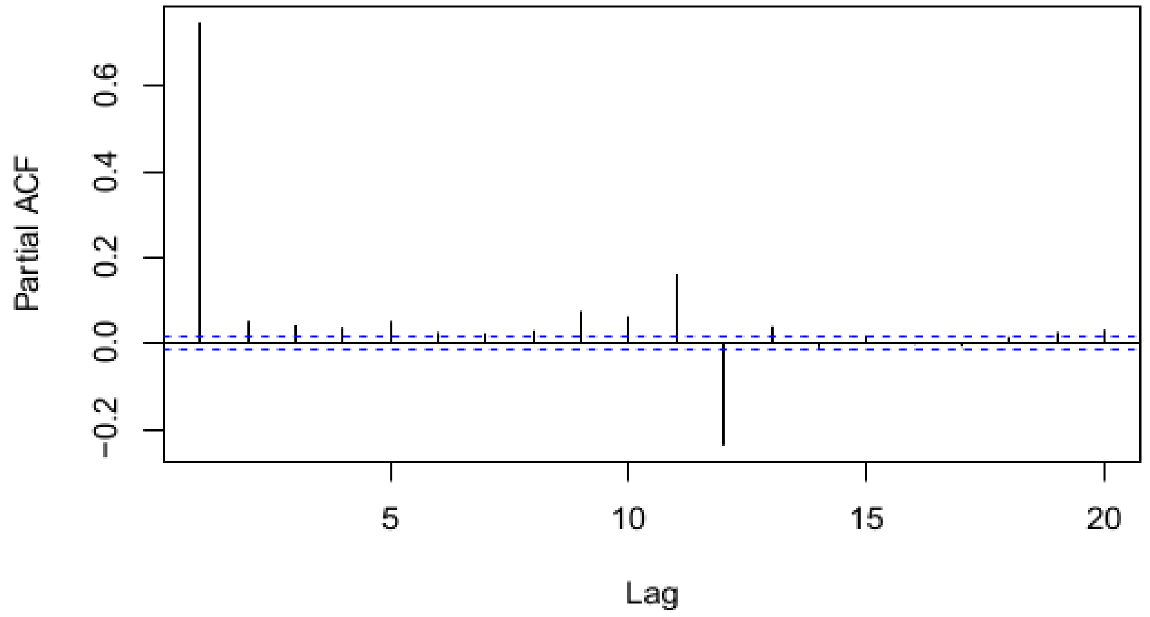


Figure 4 – Sample PACF of the clear-sky index k_t.

III. Results and discussions

MLP modelization : univariate case

Determination of the number of hidden neurons.

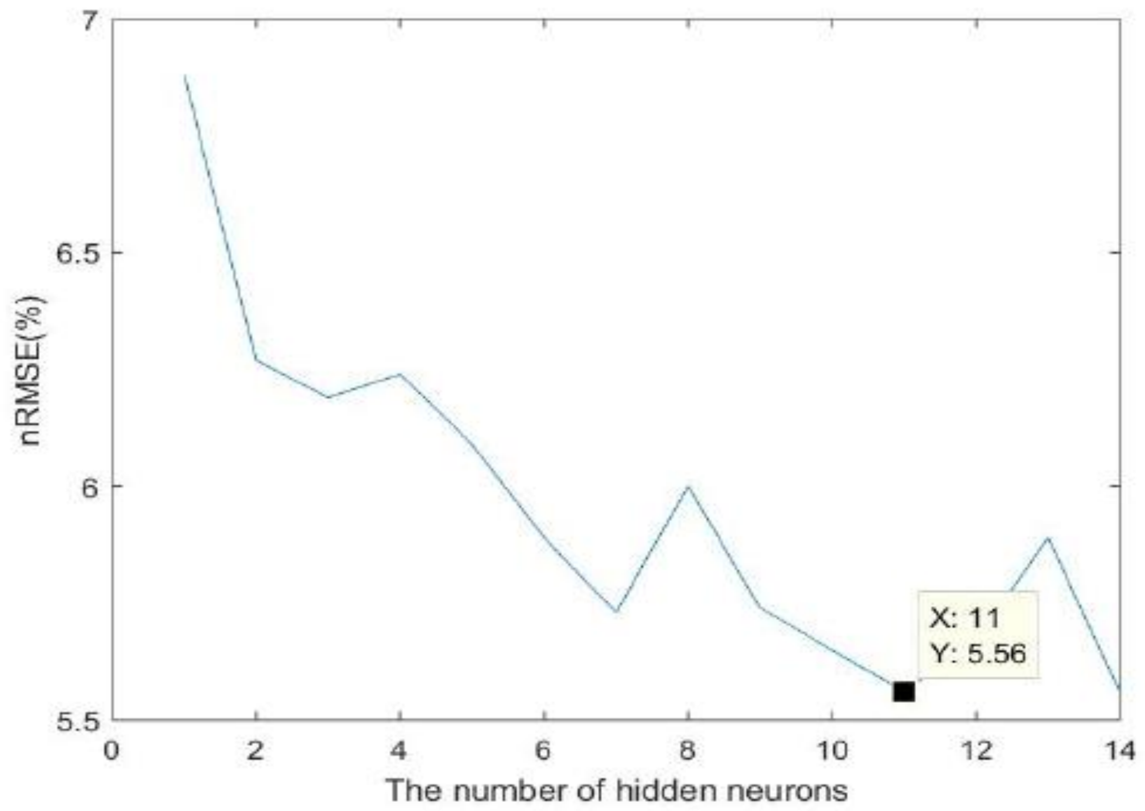


Figure 5 – forecasting error nRMSE of the global horizontal irradiance in function of the number of hidden neurons.

III. Results and discussions

MLP modelization : univariate case

Hence, the choice of the architecture of MLP were fixed on :

- inputs = 13, hidden neurons = 3 and outputs = 1.
- Activation function in the hidden layer : logistic function $\left(f(x) = \dfrac{1}{1 + e^{-2x}}\right)$.
- Activation function in the output layer : linear function $(g(x) = x)$.
- Learning algorithm : Levenberg-Marquardt.

III. Results and discussions
MLP modelization : multivariate case

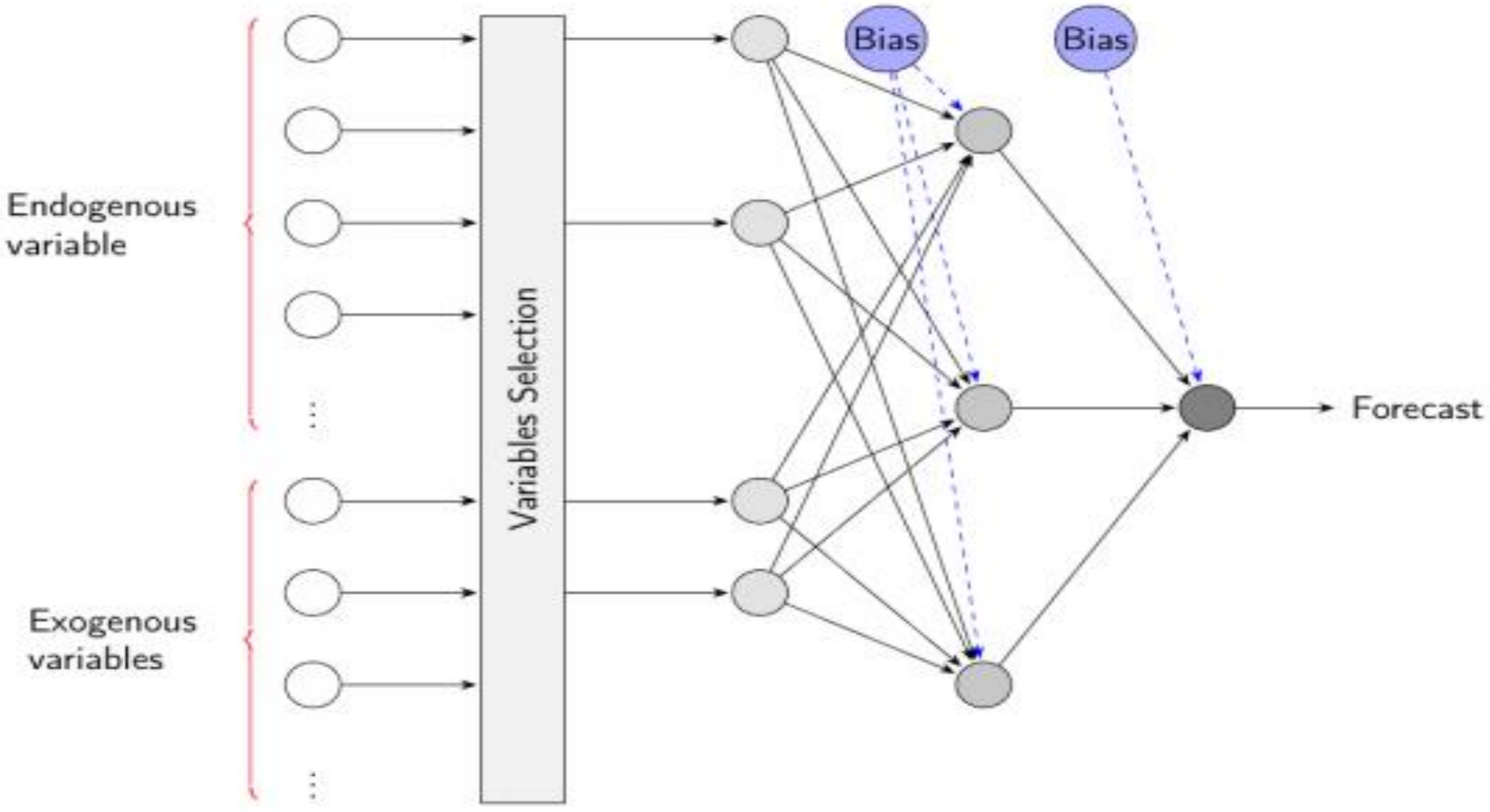


Figure 2 – The forecast methodology based on the MLP.

III. Results and discussions

MLP modelization : multivariate case

Determination of the number of time lags for the endegenous variable.

- The number of endogenous time lags has already identified in the previous section by using the PACF
- The exogenous variables were normalized by the minmax [3] normalization, then they were stationarized in case of no stationary time series.

Determination of the number of time lags for each exogenous variable.

- Cross-correlation study.

III. Results and discussions

MLP modelization : multivariate case

Cross-correlation study

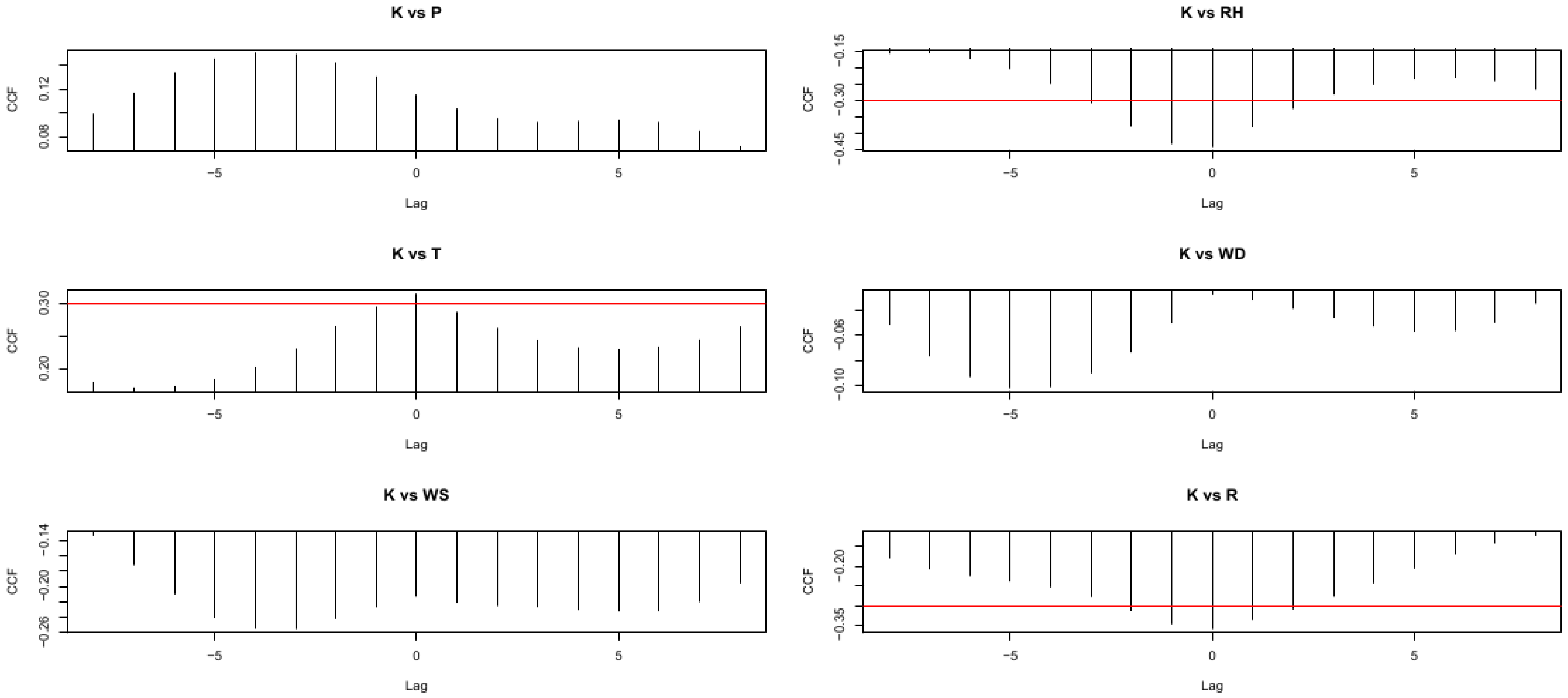


Figure 7 – Sample CCF between the clear-sky index k_t and the meteorological variables (Pressure (P), Temperature (T), Relative Humidity (RH), Wind Speed (WS), Wind direction (WD), Rainfall (R)), the red line represents the chosen confidence band.

III. Results and discussions
MLP modelization : multivariate case

RH time lag	R time lag	T time lag	n_C	nRMSE(%)
2	2	0	11	5.90
2	_	_	11	**5.25**
_	2	_	11	6.53
_	_	0	11	5.72
_	_	_	11	5.56

Tableau 2 – forecast errors of the global irradiance by testing each variable alone from the correlated variables with the clear-sky index

III. Results and discussions

MLP modelization : multivariate case

Hence, the choice of the architecture of ANN-X were fixed on :

- endogenous inputs = 13, exogenous inputs = 2, hidden neurons = 13, outputs = 1.
- Activation function in the hidden layer : logistic function $\left(f(x) = \dfrac{1}{1+e^{-2x}}\right)$.
- Activation function in the output layer : linear function $(g(x) = x)$.
- Learning algorithm : Levenberg-Marquardt.

III. Results and discussions

Forecasting errors for the region Rabat

Methodology	SARIMA	ANN	ANN-X
MAPE (%)	10.15	**8.95**	9.79
nRMSE(%)	8.97	5.56	**5.19**
RMSE (KWh/m^2)	27.72	17.17	**16.03**
Mean of errors	21.66	**-4.41**	-5.89
Variance of errors	0.0051	0.0060	**0.0050**

Tableau 3 – Performance metrics of forecasting the global horizontal irradiance.

III. Results and discussions

The region Hokkaido (Japan)

No.	Type	Contents
1	Total solar power output (each area)	Period: 2016/1/1~2017/12/31 Total solar power output measured in each area (S1, S2)
2	Temperature and global solar radiation values	Period: 2016/1/1~2017/12/31 Temperature and global solar radiation values of 4 measurement locations.

III. Results and discussions

The region Hokkaido (Japan)

Methodology	SARIMA-GARCH	ANN	ANN-X
MAPE (%)	12.87	11.90	**11.5**
nRMSE(%)	10.64	8.55	**8.3**
RMSE (MW)	43.70	35.03	**34.5**

Tableau 4 – Performance metrics of forecasting the total output of solar power.

IV. Conclusion

- The Machine Learning (ANN) approach is more efficient than the conventional forecasting approaches.
- The performance of ANN improves when exogenous variables are added.
- The difficulty of using ANN is to find the optimal structure.

Future work :

- All these results encourage us to study in the future how to adapt at different sites.
- Making the methodology more automatizing.

Let’s work together :

- New family of deep neural network; Deep Neural Ordinary Differential Equations

Refferences

Ettayyebi Hamid and El Himdi Khalid, Artificial neural network for forecasting one day ahead of global solar irradiance. The Proceeding of the International Conference on Smart Applications and Data Analysis for Smart Cities, Elsevier's SSRN's eLibrary, 2018

Ettayyebi Hamid and El Himdi Khalid. Artificial neural networks for forecasting the 24 hours ahead of global solar irradiance. 020010. 10.1063/1.5084983.

Kleissl, Jan. Solar Energy Forecasting and Resource Assessment. Elsevier Academic Press, 9780123971777 . *2013.*

Voyant Cyril and Notton Gilles and Kalogirou Soteris and Nivet Marie-Laure and Paoli Christophe and Motte Fabrice and Fouilloy Alexis, Machine learning methods for solar radiation forecasting: A review. Elsevier Science, Renewable Energy, 0960-1481, 2017. p 569–582

Thank you for your attention