Mean absolute log-return and solar radiation forecastabity

Summary

In this paper we propose to determinate and to test a set of statistical parameters (20) in order to estimate the short term predictability of the global horizontal irradiation time series and thereby to propose a new prospective tool indicating the expected error regardless the forecasting methods, a modeler can possibly implement. The mean absolute log return, which is a tool usually used in econometry but never in global radiation prediction, proves to be a very good estimator. This study gives a judgment for engineers and researchers on the installation or management of solar plants and could help in minimizing the energy crisis allowing to improve the renewable energy part of the energy mix.

1. Introduction

Several statistical parameters (Perez et al., 2012) aim at assessing the variability and consequently the difficulty to forecast the solar irradiation (Perez et al., 2012, Perez et al., 2014, Gueymard, 2014). The goal of this paper is to find a metric characterizing the time series irradiance that is correlated to the forecasting accuracy (nRMSE; nMAE) and Based on sound numerical experiments (Perez et al., 2012, Marquez and Coimbra, 2012) or even simple metric (variation coefficient, mean, standard deviation, etc.). Conversely, other parameters related to financial econometric community are studied (return, absolute log return, etc.).The paper is organized as follow: Section 2 describes the data and the statistical parameters used. Section 3 exposes the prediction methodology comparing the statistical parameters. In the two following sections, the comparison result is shown concerning 8 different locations and through 3 illustrations, we show that this new metric enables to a priori assess the accuracy of the forecasting methods based on the clear sky index series.

2. Materials and methods

To estimate a time series prediction, a stationary hypothesis is often necessary , for this study, we have chosen the ratio to trend (clear sky index; CSI) related to the simplified “Solis clear sky” model based on radiative transfer calculations and the Lambert-Beer relation.

2.1. Data

To validate this study, we choose 8 cities distributed around the world: 4 insular cities (2 in northern hemisphere, 1 in the northern tropical zone and 1 in the southern tropical zone), 3 continental cities in the north hemisphere and 1 continental city in the southern hemisphere. All these stations are part of a national measurement network and the measurement standards are almost equivalent. The three Islands (4 stations) are Reunion, Guadeloupe and Corsica, the four continental stations are: Marseille, Nice, Montpellier and Melbourne.

2.2. Statistical parameters

In this paper, we want to apply some statistical parameters on different time series and discuss about their impact on the error of prediction generated by different prediction models. In financial modelling or econometrics, a lot parameters were developed (return, volatility, etc.). In the following, we propose to adapt some of these parameters to solar radiation forecasting. All the parameters are listed in the following table

Table1. List and names of the studied statistical parameters (- means that the parameter is not tested here)

	Initial	Ratio	Log-retrun	Absolute log-return
Mean	-	*mean(ratio)*	*mean(logr)*	*Mean(abs_logr)*
Standard deviation	-	*std(ratio)*	*std(logr)*	*std(abs_logr)*
Kurtosis	-	*kurt(ratio)*	*kurt(logr)*	*kurt(abs_logr)*
Skewness	-	*skew(ratio)*	*skew(logr)*	*skew(abs_logr)*
Jarque-Nera stat	-	*JB(ratio)*	*JB(logr)*	*JB(abs_logr)*
Marquez parameter (Marquez, 2012)	**V**	-	-	-
Perez parameter (Perez et al.,2012)	*P*	-	-	-
Coefficient of variation	*CV*	-	-	-
Information dimension (Jiang, 2010)	*ID*	-	-	-
Fractal dimension (Muzy, 2008)	*FD*	-	-	-

Table1. List and names of the studied statistical parameters (- means that the parameter is not tested here)

2.3. Forecast methodology

The forecasting method studied is the persistence model (simplest estimator); the simplest way of producing a forecast. If the clear sky time series is named *CS(t)*, the scaled persistence become $\hat{x}(t+1) \xrightarrow{SP} x(t).\left(\frac{CS(t+1)}{CS(t)}\right)$ [Voyant et al., 2012].

3. Results

In order to estimate the a priori parameters linked to prediction quality, we expose in the table 2 the Spearman and the Pearson correlation factors between nRMSE (and nMAE) versus the 20 statistical parameters mentioned above (a p-value<0.05 meaning a statistical dependence between variables). These correlation factors and the significance are computed over the 8 locations: Ajaccio, Bastia, Saint-Pierre, Melbourne, Marseille, Montpellier, Nice and Le Raizet. In this first study, we decide to show only the scaled persistence predictor.

Table 2. Correlations between 18 statistical parameters computed a priori and the error of prediction done with the scaled persistence (nRMSE/nMAE over 10 repeated random sub-sampling validations). In bold the correlation significantly different from zero (p-value<0.05; i.e. statistical dependence between variables)

	nRMSE				nMAE			
	Pearson		Spearman		Pearson		Spearman	
Parameters	ρ	p-value	ρ	p-value	ρ	p-value	ρ	p-value
mean(ratio	0.305	0.462	0.619	0.115	0.142	0.738	0.548	0.171
mean(logr)	-0.012	0.978	0.190	0.665	0.045	0.916	0.048	0.935
mean(abs_logr)	**0.864**	**0.006**	0.619	0.115	**0.848**	**0.008**	0.548	0.171
std(ratio)	0.302	0.468	0.500	0.216	0.138	0.744	0.286	0.501
std(logr)	0.551	0.156	0.595	0.132	0.433	0.284	0.524	0.197
std(abs_logr)	0.436	0.281	0.571	0.151	0.303	0.466	0.500	0.216
kurt(ratio)	0.478	0.231	0.286	0.501	0.338	0.413	0.000	1.000
kurt(logr)	0.245	0.559	-0.286	0.501	0.077	0.856	-0.548	0.171
kurt(abs_logr)	0.277	0.507	0.190	0.665	0.110	0.795	-0.071	0.882
skew(ratio)	0.419	0.302	0.286	0.501	0.282	0.498	0.000	1.000
skew(logr)	-0.244	0.560	-0.048	0.935	-0.321	0.438	-0.143	0.752
skew(abs_logr)	0.245	0.558	-0.143	0.752	0.078	0.854	-0.405	0.327
JB(ratio)	0.440	0.275	0.286	0.501	0.284	0.495	0.000	1.000
JB(logr)	0.285	0.494	-0.333	0.428	0.120	0.777	-0.571	0.151
JB(abs_logr)	0.295	0.478	0.190	0.665	0.130	0.759	-0.071	0.882
V	0.330	0.425	0.238	0.582	0.426	0.293	0.357	0.389
P	0.330	0.425	0.238	0.582	0.426	0.293	0.357	0.389
CV	0.469	0.241	0.357	0.389	0.447	0.267	0.357	0.389
FD	-0.177	0.676	0.183	0.657	-0.284	0.496	-0.052	0.914
ID	-0.329	0.427	-0.176	0.679	-0.473	0.236	-0.454	0.261

In fact, although parameters related to the distribution trend or Fractal dimension seem interesting and directly linked to the predictability of the time series, the result of this study is that only the mean absolute log-return is linked to the error of prediction concerning the two studied metrics. The relationship is monotonic and linear between the two compared elements. The spearman factor generates any evidence of this link.

4. Conclusion

In this paper, we have shown that the use of a well-chosen statistical parameters could help the modeler in the choice of both a clear sky model and an accurate forecasting model for a given localization. This parameter (i.e. mean absolute log return) never used in global radiation prediction allows the optimization of these models in a fast and simple way. We have shown that the methodology works for the both of scaled persistence and MLP modeling. One way of generalization could be to extend it to others machine learning estimators (SVM, Bayesian neural network, Gaussian process, etc.). Among all the parameter tested, only one has proved efficiency: the mean absolute log-return (never used in global radiation forecasting). However, with others cities or/and more cities the result could have been different. This study has shown that econometrics tools (log-return, volatility, etc.) could help in the estimation of a simple prediction quality indexes inducing that predictions and modelling would be improved and it can be time saving.

5. References

Jiang A-H, Huang X-C, Zhang Z-H, Li J, Zhang Z-Y, Hua H-X.,2010. Mutual information algorithms. Mech Syst Signal Process;24:2947–60. doi:10.1016/j.ymssp.2010.05.015.

Marquez R, Coimbra CFM., 2012. Proposed Metric for Evaluation of Solar Forecasting Models. J Sol Energy Eng;135:011016–011016. doi:10.1115/1.4007496.

Muzy JF, Bacry E, Baile R, Poggi P.,2008. Uncovering latent singularities from multifractal scaling laws in mixed asymptotic regime. Application to turbulence. EPL Europhys Lett;82:60007. doi:10.1209/0295-5075/82/60007.

Perez R, Hoff T, Dise J, Chalmers D, Kivalov S, 2014. The Cost of Mitigating Short-term PV Output Variability. Energy Procedia;57:755–62. doi:10.1016/j.egypro.2014.10.283.

Perez R, Kivalov S, Schlemmer J, Hemker Jr. K, Hoff TE., 2012. Short-term irradiance variability: Preliminary estimation of station pair correlation as a function of distance. Sol Energy;86:2170–6. doi:10.1016/j.solener.2012.02.027.

Voyant C, Muselli M, Paoli C, Nivet M-L, 2012. Numerical weather prediction (NWP) and hybrid ARMA/ANN model to predict global radiation. Energy;39:341–55. doi:10.1016/j.energy.2012.01.0