Prediction bands for solar energy: New short-term time series forecasting techniques
Michel Fliess, Cédric Join, Cyril Voyant

To cite this version:

HAL Id: hal-01736518
https://hal-polytechnique.archives-ouvertes.fr/hal-01736518
Submitted on 17 Mar 2018

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.
Abstract

Short-term forecasts and risk management for photovoltaic energy is studied via a new standpoint on time series: a result published by P. Cartier and Y. Perrin in 1995 permits, without any probabilistic and/or statistical assumption, an additive decomposition of a time series into its mean, or trend, and quick fluctuations around it. The forecasts are achieved by applying quite new estimation techniques and some extrapolation procedures where the classic concept of “seasonalities” is fundamental. The quick fluctuations allow to define easily prediction bands around the mean. Several convincing computer simulations via real data, where the Gaussian probability distribution law is not satisfied, are provided and discussed. The concrete implementation of our setting needs neither tedious machine learning nor large historical data, contrarily to many other viewpoints.

Keywords: Solar energy, short-term forecasts, prediction bands, time series, mean, quick fluctuations, persistence, risk, volatility, normality tests.
1. Introduction

Many scientific works and technological issues (see, e.g., Hagenmeyer et al. (2016)) are related to the Energiewende, i.e., the internationally known German word for the “transition to renewable energies.” Among them weather prediction is crucial. Its history is a classic topic (see, e.g., Lynch (2008) and references therein). Reikard (2009) provides an excellent introduction to our more specific subject, i.e., short-term forecasting: “The increasing use of solar power as a source of electricity has led to increased interest in forecasting radiation over short time horizons. Short-term forecasts are needed for operational planning, switching sources, programming backup, and short-term power purchases, as well as for planning for reserve usage, and peak load matching.” Time series analysis (see, e.g., Antonanzas et al. (2016)) is quite popular for investigating such situations: See, e.g., Bacher et al. (2009); Behrang et al. (2010); Boland (1997, 2008, 2015a,b); Diagne et al. (2013); Duchon et al. (2012); Fortuna et al. (2016); Grantham et al. (2016); Hirata et al. (2017); Inman et al. (2013); Lauret et al. (2015); Martin et al. (2010); Ordiano et al. (2016); Paoli et al. (2010); Prema et al. (2015); Reikard (2009); Trapero et al. (2015); Voyant et al. (2011, 2013, 2015); Wu et al. (2011); Yang et al. (2015); Zhang et al. (2015), . . . , and references therein. The developed viewpoints are ranging from the rather classic setting, stemming from econometrics to various techniques from artificial intelligence and machine learning, like artificial neural networks.

No approach will ever rigorously produce accurate predictions, even nowcasting, i.e., short-term forecasting. To the best of our knowledge, this unavoidable uncertainty, which ought to play a crucial rôle in the risk management of solar energy, starts only to be investigated (see, e.g., David et al. (2016); Ordiano et al. (2016); Rana et al. (2015, 2016); Scolari et al. (2016); Trapero (2016)). As noticed by some authors (see, e.g., David et al. (2016); Trapero (2016)), this lack of precision might be related to volatility, i.e., a most popular word in econometrics and financial engineering. Let us stress however the following criticisms, that are borrowed from the financial engineering literature:
1. Wilmott (2006) (chap. 49, p. 813) writes: Quite frankly, we do not know what volatility currently is, never mind what it may be in the future.

2. According to Mandelbrot et al. (2004), the existing mathematical definitions suffer from poor probabilistic assumptions.

3. Goldstein et al. (2007) exhibits therefore multiple ways for computing volatility which are by no means equivalent and might even be contradictory and therefore misleading.

A recent conference announcement (Join et al. (2016)) is developed here. It is based on a new approach to time series that has been introduced for financial engineering purposes (Fliess et al. (2009, 2011, 2015a,b)). A theorem due to Cartier et al. (1995) yields under very weak assumptions on time series an additive decomposition into its mean, or trend, and quick fluctuations around it. Let us emphasize the following points:

- The probabilistic/statistical nature of those fluctuations does not play any rôle.

- No modeling via difference/differential equations is necessary: it is a model-free setting.

- Implementation is possible without arduous machine learning and large historical data.

A clear-cut definition of volatility is moreover provided. It is inspired by the
mean absolute error (MAE) which has been proved already to be more convenient in climatic and environmental studies than the root mean square error (RMSE) (Willmott et al. (2005)). This fact is to a large extent confirmed by Chai et al. (2014) by Section 3.2 which shows that the fluctuations are not Gaussian. See, e.g., (Hyndman (2006)) for further theoretical investigations. Confidence intervals, i.e., a well known notion in statistics (Cox et al. (1974); Willink (2013)), do not make much sense since the probabilistic nature of the uncertainty is unknown. We are therefore replacing them by prediction bands[^3] They mimic to some extent the Bollinger bands (Bollinger (2001)) from technical analysis, i.e., a widespread approach to financial engineering (see, e.g., Béchu et al. (2014); Kirkpatrick et al. (2010)). To pinpoint the efficiency of our tools, numerical experiments via real data stemming from two sites are presented.

Our paper is organized as follows. Time series are the core of Section 2 where algebraic nowcasting and prediction bands are respectively presented in Sections 2.4 and 2.7. The numerical experiments are presented and discussed in Section 3. Considerations on future investigations are presented in Section 4.

2. Time series

2.1. Nonstandard analysis: A short introduction

Robinson (1996) introduced nonstandard analysis in the early 60’s (see, e.g., Dauben (1995)). It is based on mathematical logic and vindicates Leibniz’s ideas on “infinitely small” and “infinitely large” numbers. Its presentation by Nelson (1977) (see also Nelson (1987) and Diener et al. (1995, 1989)), where the logical background is less demanding, has become more widely used. As demonstrated by Harthong (1981), Lobry (2008), Lobry et al. (2008), and several other authors, nonstandard analysis is a marvelous tool for clarifying in a most intuitive way various questions from applied sciences.

[^3]: We might also employ the terminology confidence bands. To the best of our knowledge, it has been already employed elsewhere but with another definitions (see, e.g., Härdle et al. (2004)).
2.2. Time series and nonstandard analysis

2.2.1. A nonstandard definition of time series

Take a time interval $[0, 1]$. Introduce as often in nonstandard analysis the infinitesimal sampling

$$\mathcal{T} = \{0 = t_0 < t_1 < \cdots < t_\nu = 1\}$$

where $t_{i+1} - t_i$, $0 \leq i < \nu$, is infinitesimal, i.e., “very small.” A time series $X$ is a function $\mathcal{T} \to \mathbb{R}$.

Remark 2.1. The normalized time interval $[0, 1]$ is introduced for notational simplicity. It will be replaced here by a time lapse from a few minutes to one hour. Infinitely small or large numbers should be understood as mathematical idealizations. In practice a time lapse of 1 second (resp. hour) should be viewed as quite small when compared to 1 hour (resp. month). Nonstandard analysis may therefore be applied in concrete situations.

2.2.2. The Cartier-Perrin theorem

The Lebesgue measure on $\mathcal{T}$ is the function $\ell$ defined on $\mathcal{T}\setminus\{1\}$ by $\ell(t_i) = t_{i+1} - t_i$. The measure of any interval $[c, d] \subset \mathcal{T}$, $c \leq d$, is its length $d - c$. The integral over $[c, d]$ of the time series $X(t)$ is the sum

$$\int_{[c, d]} X \, d\tau = \sum_{t \in [c, d]} X(t)\ell(t)$$

$X$ is said to be $S$-integrable if, and only if, for any interval $[c, d]$ the integral $\int_{[c, d]} |X| \, d\tau$ is limited, i.e., not infinitely large, and, if $d - c$ is infinitesimal, $\int_{[c, d]} |X| \, d\tau$ is also infinitesimal.

$X$ is $S$-continuous at $t_i \in \mathcal{T}$ if, and only if, $f(t_i) \simeq f(\tau)$ when $t_i \simeq \tau$. $X$ is said to be almost continuous if, and only if, it is $S$-continuous on $\mathcal{T}\setminus R$, where $R$ is a rare subset[4]. $X$ is Lebesgue integrable if, and only if, it is $S$-integrable and almost continuous.

---

[4] $a \simeq b$ means that $a - b$ is infinitesimal.

[5] The set $R$ is said to be rare (Cartier et al. [1995]) if, for any standard real number $\alpha > 0$, there exists an internal set $A \supset R$ such that $m(A) \leq \alpha$. 
A time series $X : \mathbb{S} \to \mathbb{R}$ is said to be *quickly fluctuating*, or *oscillating*, if, and only if, it is $S$-integrable and $\int_A X d\tau$ is infinitesimal for any quadrable subset.\(^6\)

Let $X : \mathbb{S} \to \mathbb{R}$ be a $S$-integrable time series. Then, according to the Cartier-Perrin theorem (Cartier et al. (1995)),\(^7\) the additive decomposition

$$X(t) = E(X)(t) + X_{\text{fluctuat}}(t)$$

(2)

holds where

- $E(X)(t)$, which is called the *mean*, or *trend*\(^8\) is Lebesgue integrable;
- $X_{\text{fluctuat}}(t)$ is quickly fluctuating.

The decomposition (2) is unique up to an additive infinitesimal quantity. Let us stress once again that the above mean is independent of any probabilistic modeling.\(^9\)

### 2.3. Volatility

According to

- our discussion about mean absolute errors (MAE) in Section 1,
- the fact, which follows at once from the Cartier-Perrin theorem, that $|X - E(X)|$ is $S$-integrable,

define the *volatility* $\text{vol}(X)(t)$ of $X(t)$ by

$$\text{vol}(X)(t) = E(|X - E(X)|)(t)$$

(3)

$E(|X - E(X)|)(t)$ in Equation (3) is nothing else than the mean of $|X(t) - E(X)(t)|$.

---

\(^6\)A set is *quadrable* Cartier et al. (1995) if its boundary is rare.

\(^7\)The presentation in the article by Lobry et al. (2008) is less technical. We highly recommend it. Note that it also includes a fruitful discussion on nonstandard analysis.

\(^8\)“Trend” would be the usual terminology in technical analysis (see, e.g., Béchu et al. 2014; Kirkpatrick et al. 2010). It was therefore used by Fliess et al. (2009).

\(^9\)Let us mention that Cartier et al. (1995) also introduced the notion of *martingales* (see, e.g., Williams (1991)) without using any probabilistic tool.
2.4. Forecasting via algebraic estimation techniques

In order to forecast via the above setting, new estimation tools have to be summarized (see, e.g., Fliess et al. (2003, 2008a,b); Mboup et al. (2010); Sira-Ramírez et al. (2014)).

2.4.1. First calculations

Start with a polynomial time function

\[ p_1(t) = a_0 + a_1 t, \quad t \geq 0, \quad a_0, a_1 \in \mathbb{R}, \]

of degree 1. Rewrite thanks to classic operational calculus (see, e.g., Yosida (1984))

\[ P_1 = \frac{a_0}{s} + \frac{a_1}{s^2} \]

Multiply both sides by \( s^2 \):

\[ s^2 P_1 = a_0 s + a_1 \quad (4) \]

Take the derivative of both sides with respect to \( s \), which corresponds in the time domain to the multiplication by \(-t\):

\[ s^2 \frac{dP_1}{ds} + 2s P_1 = a_0 \quad (5) \]

The coefficients \( a_0, a_1 \) are obtained via the triangular system of equations (4)-(5). We get rid of the time derivatives, i.e., of \( sP_1, s^2 P_1, \) and \( s^2 \frac{dP_1}{ds} \), by multiplying both sides of Equations (4)-(5) by \( s^{-n} \), i.e., \( n \geq 3 \) (resp \( n \geq 2 \)) for Equation (4) (resp. (5)). The corresponding iterated time integrals are lowpass filters (see, e.g., Shenoi (2006)): they attenuate the corrupting noises, which are viewed as highly fluctuating phenomena (Fliess (2006)). A quite short time

---

10Those techniques have already been successfully employed in engineering. In signal processing, see, e.g., the recent publications by Beltran-Carbajala et al. (2017) and Morales et al. (2016).

11The computations below are often presented via the classic Laplace transform (see, e.g., Doetsch (1976)). Then \( s \) is called the Laplace variable.
window $[0, t]$ is sufficient for obtaining accurate estimates $\hat{a}_0, \hat{a}_1, \hat{a}_0, a_1$, where $n = 2, 3$:

$$\hat{a}_0 = \frac{2}{t^2} \int_0^t (2t - 3\tau)p(\tau)d\tau$$

and

$$\hat{a}_1 = -\frac{6}{t^3} \int_0^t (t - 2\tau)p(\tau)d\tau$$

This last formula shows that a derivative estimate is obtained via integrals. Lanczos (1956) was perhaps the first author to suggest such an approach. In practice, the above integrals are of course replaced by straightforward linear digital filters (see, e.g., Shenoi (2006)).

2.5. Back to time series and short-term forecasts

Assume that the following rather weak assumption holds true: the mean $E(X(t))$ may be associated with a differentiable real-valued time function. Then, on a short time lapse, $E(X(t))$ is well approximated by a polynomial function of degree 1. The above calculations yield via sliding time windows numerical estimates $E(X)_{\text{estim}}(t)$ and $\frac{d}{dt}E(X)_{\text{estim}}(t)$ of the mean and its derivative. Causality is taken into account via backward calculations with respect to time. As in (Fliess et al. 2009, 2011), forecasting the time series $X(t)$ boils down to an extrapolation of its mean $E(X)(t)$. If $T > 0$ is not ‘too large,” i.e., a few minutes in our context, a first order Taylor expansion yields the following extrapolation for prediction at time $t + T$

$$X_{\text{predict}}(t + T) = E(X)_{\text{estim}}(t) + \left(\frac{d}{dt}E(X)_{\text{estim}}(t)\right) \times T$$

2.6. Forecasting for a larger time horizon

With forecasts for a time horizon equal to 1 hour, Equation (6) would provide poor results. Seasonalities, i.e., a more or less periodic pattern, which is classic in time series analysis (see, e.g., Brockwell et al. 1991; Mélard 2008) will be used here. A single day is an obvious season with respect to photovoltaic energy. Figures 7, 10 show that the corresponding pattern may be reasonably well approximated by a parabola $D(t) = \alpha_2t^2 + \alpha_1t + \alpha_0$. Standard least
square techniques permit to obtain such a suitable parabola, that is the set of parameters \( \{ \alpha_0, \alpha_1, \alpha_2 \} \), only with the data collected during a single day.

Replace Equation (6) by

\[
X_{\text{predict}}(t + T) = E(X_{\text{estim}}(t) + \left(\dot{D}(t - 1.\text{day})\right) \times T
\]

(7)

where

- \( T > 0 \) is the time horizon, here between 30 minutes and 1 hour;
- \( D(t - 1.\text{day}) \) is estimated via the data from the day before;
- \( \dot{D}(t - 1.\text{day}) \) is its derivative.

This formula is useful since the parabola is erasing the bumps and the hollows on the trend. Taking derivatives around such bumps and holes leads obviously to a wrong forecasting for a larger time horizon.

### 2.7. Prediction bands

Equation (3) yields the prediction \( \text{Vol}_{\text{predict}}(X)(t + T) \) of the volatility at time \( t + T \) via the following persistence law (Lauret et al. (2015)):

\[
\text{Vol}_{\text{predict}}(X)(t + T) = \text{Vol}(X)(t) = E(|X - E(X)|)(t)
\]

(8)

Define via Equation (8) the first prediction band

\[
\text{CB}_1(t + T) = X_{\text{predict}}(t + T) - \text{Vol}_{\text{predict}}(X)(t + T)\]

\[
\leq \text{CB}_1(t + T) \leq X_{\text{predict}}(t + T) + \text{Vol}_{\text{predict}}(X)(t + T) = \text{CB}_1(t + T)
\]

(9)

In order to improve it, set

\[
\text{CB}_2(t + T) = X_{\text{predict}}(t + T) - \alpha_{t+T} \text{Vol}_{\text{predict}}(X)(t + T)
\]

\[
\leq \text{CB}_2(t + T) \leq X_{\text{predict}}(t + T) + \alpha_{t+T} \text{Vol}_{\text{predict}}(X)(t + T) = \text{CB}_2(t + T)
\]

(10)
where the coefficient $\alpha_{t+T} > 0$ may be chosen in various ways. If, for instance, $\alpha_{t+T} = 1$, we are back to Equation \[[9]\]. Here we select $\alpha_{t+T}$ such that the band \[[10]\] contains during the 3 previous days 68% of the available data.\[^{12}\]

Taking into account the global and diffuse radiations under clear sky will obviously improve the above band \[[10]\]. Note that clear sky models played already some rôle in solar irradiation and irradiance forecasting via time series (see, e.g., \[Cros et al. (2013)\; Inman et al. (2013)\]). This is achieved here by using the quite famous \textit{solis} model \[Mueller et al. (2004)\; Ineichen (2008)\]. The clear sky global horizontal irradiance $I_{g,\text{clsk}}$ reaching the ground and the clear sky beam radiation $I_{b,\text{clsk}}$ are defined by \[Ineichen (2008)\]:

\begin{align*}
I_{g,\text{clsk}} &= I_0 \exp \left( -\frac{\tau_g}{(\sin h)^g} \right) \sin h \tag{11} \\
I_{b,\text{clsk}} &= I_0 \exp \left( -\frac{\tau_b}{(\sin h)^b} \right) \tag{12}
\end{align*}

where

- $I_0$ is the extraterrestrial radiation (depending of the day of the year),
- $h$ is the solar elevation (depending of the hour of the day),
- $\tau_g$ and $\tau_b$ are respectively the global and beam total atmospheric optical depths,
- $g$ and $b$ are fitting parameters.

Diffuse radiation $I_{d,\text{clsk}}$ is defined by

\[I_{d,\text{clsk}} = I_{g,\text{clsk}} - I_{b,\text{clsk}}.\]

\[^{12}\text{The quantity 68\% is obviously inspired by the theory confidence intervals with respect to Gaussian probability distributions.}\]
It yields

\[
\text{CB}_3(t + T) = \max (I_{d,\text{clsk}}(t), X_{\text{predict}}(t + T) - \alpha_{t+T}\text{Vol}_{\text{predict}}(X)(t + T)) \\
\leq \text{CB}_3(t + T) \leq \min (1.1 \times I_{g,\text{clsk}}(t), X_{\text{predict}}(t + T) + \alpha_{t+T}\text{Vol}_{\text{predict}}(X)(t + T)) = \text{CB}_3(t + T)
\]

(13)

where \(\min(\Box(t), \triangle(t))\) and \(\max(\Box(t), \triangle(t))\) are respectively the minimum and maximum values of the arguments \(\Box(t)\) and \(\triangle(t)\) at time \(t\).

The safety margin corresponding to the multiplicative factor 1.1 takes into account a modeling error on \(I_{g,\text{clsk}}\) \cite{Ineichen2008}, whereas \(I_{d,\text{clsk}}\) does not necessitate such a correction \cite{Ineichen2008}.

3. Computer experiments via real data

Three time horizons are considered: 1, 15 and 60 minutes. The following points should be added:

- no exogenous variable,
- no need of large historical data,
- unsupervised method.

3.1. Data

The full year data were collected from two sites in 2013 by means of CMP11 pyranometer (Kipp & Zonen):

- Nancy in the East of France. It has usually a relatively narrow annual temperature range.
- Ajaccio in Corsica, a French island in the Mediterranean sea. This coastal town has hot and sunny summers and mild winters.
Figure 1: Global irradiation profile for the two sites in February

The time granularity of our solar irradiance measurements is 1 minute. Missing values for the sites are less than 2%. See Figures 1 and 2 for excerpts. The numerical values of the parameters in Equations (11) and (12), are:

- Nancy: $\tau_g = 0.49, g = 0.39, \tau_b = 0.66, b = 0.51$;
- Ajaccio: $\tau_g = 0.43, g = 0.33, \tau_b = 0.64, b = 0.51$.

### 3.2. Normality tests

To better justify our definitions of volatility in Section 2.3 and of prediction bands in Section 2.7, we show that, if the fluctuations around the trends are viewed as random variables, they are not Gaussian. Three classic tests (see, e.g., Jarque et al. (1987), Judge et al. (1988), Thode (2002)) are used:

- Jarque-Bera,
- Kolmogorov-Smirnov.

---

13 The data are cleaned as in David et al. (2016).
14 See, e.g., the WEB site of AERONET Data Synergy Tool.
Figures 3, 4, 5 and 6 show most clearly that the Gaussian property is not satisfied.

3.3. Presentation of some results

Figures 1 and 2 present global irradiation during 1 month. Figure 1 displays an irregular radiation behavior during winter. As shown by Figure 1-(a) this is especially true for Nancy. During summer, Figure 2 exhibits a nice daily seasonality even if some deteriorations show up for Nancy.

The red line in Figures 7, 8, 10, 11 on the one hand, and in Figures 9, 12 on the other hand, show the forecasts according respectively to Equations (6) and (7).

Remark 3.1. We follow a common practice by removing night hours, i.e., hours where the solar elevation $h$ in Equations (11)–(12) is less than 10 degrees.
Figure 3: Nancy, February: Signal distribution (blue) and the Gaussian distribution (red)

Figure 4: Nancy, June: Signal distribution (blue) and the Gaussian distribution (red)

Figure 5: Ajaccio, February: Signal distribution (blue) and the Gaussian distribution (red)
The prediction bands defined in Section 2.7 are also displayed in the previous figures. Note the following points:

1. by construction, $CB_2$ yields larger bands than $CB_1$,
2. the mean interval length is reduced with $CB_3$,
3. the widths of the bands increase with the time horizon,
4. Daily profiles, Figures 7, 8, 9 on the one hand, and Figures 10, 11, 12 on the other hand demonstrate that forecasts are better in June than in February.

In order to quantify comparisons, introduce the following quantities:

- The **Mean Interval Length**, or MIL stems from the Mean Relative Error (MRL) (Rana et al. (2015, 2016); Scolari et al. (2016)). It is given by
  \[
  MIL_i = \frac{\sum_{k=1}^{N} BW_i(t_k)}{\sum_{k=1}^{N} X(t_k)} \quad i = 1, 2, 3
  \]
  where
  - $N$ is the number of measurements,
  - $BW_i = CB_i(t_k + T) - CB_i(t_k + T) \geq 0$, is the band width,
  - $X(t_k) \geq 0$ the irradiance measurement.

- The **Prediction Interval Coverage Probability**, or PICP, (Rana et al. (2015, 2016); Scolari et al. (2016)) is defined by
  \[
  PICP_i = \frac{\sum_{k=1}^{M} c_k}{M} \quad i = 1, 2, 3
  \]
where

- $M$ is the number of predictions,
- $c_k = 1$ if the prediction is inside the bands, i.e., $CB_i(t_k + T) \leq X(t_k + T) \leq CB_i(t_k + T),$
- $c_k = 0$ otherwise.

A quite large MIL with a PICP close to 1 is inefficient for grid management. Our objective is a large PICP and a low MIL. Consequently, a compromise is required. Figures [13, 14, 15] and [16] present MIL vs PICP for all forecasting horizons.

On these Figures, four areas characterise the CB qualities. Thus, if a bound is in the “good” area, the result is more interesting than in the “bad” and even more than in the “very bad” areas but less interesting than in the “very good” zone. According to the clear sky concept and to the ad-hoc computing methodologies [Mueller et al. (2004), Ineichen (2008)], the measured global irradiance is between the computed irradiance under clear sky and under totally cloudy sky. So 100% of the predictions, i.e., $\text{PICP}_i = 1$, should be included between the bounds defined by the global radiation $I_{g,\text{clsk}}$ and the diffuse radiation $I_{d,\text{clsk}}$.

Uncertainties and Solis modeling errors explain why it is not always the case.

The space is divided in four zones. The blue line is the vertical limit. It corresponds to a PICP of 0.5: it means that 50% of predictions are in the band. The green line is the horizontal limit. It defines the limit of relevance: all the intervals with a $\text{MIL} > 1$ are not really interesting since the bands are too large. For case 1 ($CB_1$) and 2 ($CB_2$), we find coherent results because the PICP are respectively close to 70% and 50%. Regarding case 2, the MIL is too important. Regarding case 3 ($CB_3$), the best compromise between low MIL and high PICP is obtained thanks to the clear sky model.

3.4. Some preliminary comments on comparisons

Comparing our results with the huge set of numerical calculations in the whole academic literature is obviously beyond the reach of a single journal pub-
Figure 7: Nancy, February, 5min forecasting: irradiance (blue), its prediction (red) and prediction band (black - -)

Figure 8: Nancy, February, 15min forecasting: irradiance (blue), its prediction (red) and prediction band (black - -)
Figure 9: Nancy, February, 60min forecasting: irradiance (blue), its prediction (red) and prediction band (black - -)

Figure 10: Nancy, June, 5min forecasting: irradiance (blue), its prediction (red) and prediction band (black - -)
Figure 11: Nancy, June, 15min forecasting: irradiance (blue), its prediction (red) and prediction band (black - -)

Figure 12: Nancy, June, 60min forecasting: irradiance (blue), its prediction (red) and prediction band (black - -)

Figure 13: Nancy, February: Performance evaluation
Figure 14: Nancy, June: Performance evaluation

Figure 15: Ajaccio, February: Performance evaluation

Figure 16: Ajaccio, June: Performance evaluation
lication. Let us nevertheless summarize the observations in Join et al. (2014): Voyant et al. (2015) on short-term forecasting with respect to artificial neural networks:

1. the performances of the “algebraic” setting, that is presented here, are perhaps slightly better than those via neural nets. For the irradiance (resp. irradiation), the mean absolute error (MAE) between the forecasts and the true values is 26.6% (resp. 23.9%) vs 35.44% (resp. 22.36%).

2. When looking at big data and machine learning, the behavior of the algebraic setting looks much better. Neural nets need data during 3 years whereas the algebraic viewpoint only 1 day.

With respect, for instance, to Grantham et al. (2016); David et al. (2016); Trapero (2016), note that time-consuming and cumbersome calibrations to obtain a convincing probability law, a time series modeling via a difference equation, and a suitable autoregressive conditional heteroskedasticity (ARCH) or generalized autoregressive conditional heteroskedasticity (GARCH) become quite irrelevant.

4. Conclusion

The positive results obtained in this paper ought of course to be verified by considering more diverse situations and by launching more thorough comparisons than in Section 3.4. For future researches, let us emphasize the two following directions:

- The possibility of extending our techniques to larger time horizons is another key point.
- Asymmetric prediction bands might be useful in practice for energy management.

---

The popular concepts of ARCH and GARCH were respectively introduced by Engle (1982) and Bollerslev (1986). Everyone should read the harsh comments by Mandelbrot et al. (2004).
Is the causality analysis by Fliess et al. (2015) useful to improve our forecasting techniques if other facts are taken into account (see, e.g., Badosa et al. (2015))?

The concrete implementation of our approach should be rather straightforward. Finally, if our standpoint encounters some success, the probabilistic techniques (see, e.g., Appino et al. (2018); Gneiting et al. (2014); Hong et al. (2016); Lauret et al. (2017)), which are a today mainstay in all the fields of energy forecasting, price included, might become clearly less central.

Acknowledgements. The authors thank warmly the anonymous reviewers for their most helpful comments.

References


Chai, T., Draxler, R.R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. Geosci. Model Dev. 7, 1247–1250.


Goldstein, D.G., Taleb, N.N., 2007. We don’t quite know what we are talking about when we talk about volatility. J. Portfolio Manage. 33, 84–86.


Wasserman, L., 2006. All of Nonparametric Statistics. Springer.


