Solar Energy 166 (2018) 519–528

Contents lists available at ScienceDirect

Solar Energy

journal homepage: www.elsevier.com/locate/solener

Prediction bands for solar energy: New short-term time series forecasting techniques

Michel Fliess[a,d,*], Cédric Join[b,d,e], Cyril Voyant[c,f]

[a] LIX (CNRS, UMR 7161), École polytechnique, 91128 Palaiseau, France
[b] CRAN (CNRS, UMR 7039), Université de Lorraine, BP 239, 54506 Vandœuvre-lès-Nancy, France
[c] SPE (CNRS, UMR 6134), Università di Corsica Pasquale Paoli, 20250 Corte, France
[d] AL.I.E.N. (ALgèbre pour Identification & Estimation Numériques), 7 rue Maurice Barrès, 54330 Vézelise, France
[e] Projet Non-A, INRIA Lille – Nord-Europe, France
[f] Hôpital de Castelluccio, Unité de Radiothérapie, BP 85, 20177 Ajaccio, France

ARTICLE INFO

Keywords:
Solar energy
Short-term forecasts
Prediction bands
Time series
Mean
Quick fluctuations
Persistence
Risk
Volatility
Normality tests

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.

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 and Hale (2012), Fortuna et al. (2016), Grantham et al. (2016), Hirata and Aihara (2017), Inman et al. (2013), Lauret et al. (2015), Martín et al. (2010), Ordiano et al. (2016), Paoli et al. (2010), Prema and Rao (2015), Reikard (2009), Trapero et al. (2015), Voyant et al. (2011, 2013, 2015), Wu and Chan (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; Rana and Koprinska, 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.*

* Corresponding author at: LIX (CNRS, UMR 7161), École polytechnique, 91128 Palaiseau, France.
E-mail addresses: Michel.Fliess@polytechnique.edu, michel.fliess@alien-sas.com (M. Fliess), cedric.join@univ-lorraine.fr, cedric.join@alien-sas.com (C. Join), voyant@univ-corse.fr (C. Voyant).

https://doi.org/10.1016/j.solener.2018.03.049
Received 28 May 2017; Received in revised form 11 March 2018; Accepted 17 March 2018
Available online 13 April 2018
0038-092X/

2. According to Mandelbrot and Hudson (2004), the existing mathematical definitions suffer from poor probabilistic assumptions.
3. Goldstein and Taleb (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 and Join, 2009; Fliess et al., 2011; Fliess and Join, 2015a,b). A theorem due to Cartier and Perrin (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.[1]
- No modeling via difference/differential equations is necessary: it is a *model-free* setting.[2]
- 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 and Matsuura, 2005). This fact is to a large extent confirmed by Chai and Draxler (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 and Hinkley, 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 and Dahlquist, 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 and Diener, 2013; Diener and Reeb, 1989), where the logical background is less demanding, has become more widely used. As demonstrated by Harthong (1981),Lobry (2008), Lobry and Sari (2008), and several other authors, nonstandard analysis is a marvelous tool for clarifying in a most intuitive way various questions from applied sciences.

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

$$\mathfrak{T} = \{0 = t_0 < t_1 < \cdots < t_\nu = 1\} \quad (1)$$

where $t_{i+1}-t_i, 0 \leqslant i < \nu$, is *infinitesimal*, *i.e.*, "very small." A time series X is a function $\mathfrak{T} \to \mathbb{R}$.

Remark 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 s (resp. hour) should be viewed as quite small when compared to 1 h (resp. month). Nonstandard analysis may therefore be applied in concrete situations.

2.2.2. The Cartier-Perrin theorem

The *Lebesgue measure* on $\mathfrak{T}$ is the function ℓ defined on $\mathfrak{T}\backslash\{1\}$ by $\ell(t_i) = t_{i+1}-t_i$. The measure of any interval $[c,d] \subset \mathfrak{T}, c \leqslant d$, 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 $t_\iota \in \mathfrak{T}$ if, and only if, $f(t_\iota) \simeq f(\tau)$ when $t_\iota \simeq \tau$.[4] X is said to be *almost continuous* if, and only if, it is S-continuous on $\mathfrak{T}\backslash R$, where R is a *rare* subset.[5] X is *Lebesgue integrable* if, and only if, it is S-integrable and almost continuous.

A time series $\mathscr{X}: \mathfrak{T} \to \mathbb{R}$ is said to be *quickly fluctuating*, or *oscillating*, if, and only if, it is S-integrable and $\int_A \mathscr{X} d\tau$ is infinitesimal for any *quadrable* subset.[6]

Let $X: \mathfrak{T} \to \mathbb{R}$ be a S-integrable time series. Then, according to the Cartier-Perrin theorem (Cartier and Perrin, 1995),[7] the additive decomposition

$$\boxed{X(t) = E(X)(t) + X_{\text{fluctuat}}(t)} \quad (2)$$

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]

[1] This fact should be viewed as fortunate since this nature is rather mysterious if real data are involved.

[2] At least two other wordings, namely "nonparametric" or "data-driven," instead of "model-free" would have been also possible. The first one however is almost exclusively related to the popular field of *nonparametric statistics* (see, *e.g.*, Härdle et al., 2004; Wasserman, 2006), that has been also encountered for photovoltaic systems (see, *e.g.*, Ordiano et al., 2016). The second one has also been recently used, but in a different setting (see, *e.g.*, Ordiano et al., 2017). Let us highlight the numerous accomplishments of *model-free control* (Fliess and Join, 2013) in engineering. See for instance renewable energy Bara et al. (2017),Jama et al. (2015), Join et al. (2016), and agricultural greenhouses Lafont et al. (2015).

[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).

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

[5] The set R is said to be *rare* (Cartier and Perrin, 1995) if, for any standard real number $\alpha > 0$, there exists an internal set $A \supset R$ such that $m(A) \leqslant \alpha$.

[6] A set is *quadrable* Cartier and Perrin (1995) if its boundary is rare.

[7] The presentation in the article by Lobry and Sari (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 and Dahlquist, 2010. It was therefore used by Fliess and Join (2009).

[9] Let us mention that Cartier and Perrin (1995) also introduced the notion of *martingales* (see, *e.g.*, Williams, 1991) without using any probabilistic tool.