



Introducing the Temporal Distortion Index to perform a bidimensional analysis of renewable energy forecast



Laura Frías-Paredes ^{a,*}, Fermín Mallor ^b, Teresa León ^c, Martín Gastón-Romeo ^a

^a Cener (Nacional Renewable Energy Centre), Ciudad de la Innovación 7, Sarriguren 31621, Navarre, Spain

^b Dept. Statistics and Operations Research and Institute of Smart Cities, Public University of Navarre, Spain

^c Dept. Statistics and Operations Research, University of Valencia, Spain

ARTICLE INFO

Article history:

Received 29 March 2015

Received in revised form

18 October 2015

Accepted 22 October 2015

Available online 21 November 2015

Keywords:

Forecast accuracy

Temporal misalignment

Dynamic time warping

Renewable energy

Temporal distortion index

Bidimensional error

ABSTRACT

Wind has been the largest contributor to the growth of renewal energy during the early 21st century. However, the natural uncertainty that arises in assessing the wind resource implies the occurrence of wind power forecasting errors which perform a considerable role in the impacts and costs in the wind energy integration and its commercialization. The main goal of this paper is to provide a deeper insight in the analysis of timing errors which leads to the proposal of a new methodology for its control and measure. A new methodology, based on Dynamic Time Warping, is proposed to be considered in the estimation of accuracy as attribute of forecast quality. A new dissimilarity measure, the Temporal Distortion Index, among time series is introduced to complement the traditional verification measures found in the literature. Furthermore we provide a bi-criteria perspective to the problem of comparing different forecasts. The methodology is illustrated with several examples including a real case.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

These last years have observed a rapid growth in wind electricity generation and it is expected that wind will cover a larger percentage of the generation mix in next decades [1,2]. As wind is inherently variable, wind power is a fluctuating source of electrical energy. Therefore, having accurate wind power forecasts is essential for grid integration, system planning and electricity trading in certain electricity markets. The analysis of the prediction errors appears as a critical task to be capable of comparing prediction models and decide which one is preferable. Even more, a deeper knowledge of the nature of the prediction errors will be very helpful to improve the quality of a forecasting model and to increase the economic benefits obtained by decision makers through the use of forecast.

Many references in the literature have been devoted to the analysis of prediction errors and several statistical measures of the average inaccuracy associated with a set of model-produced estimates have been traditionally used for comparisons: the MAE (Mean Absolute Error), the RMSE (Root Mean Squared Error), the

MSE (Mean Squared Error), the SDE (Standard Deviation of Error) and the corresponding normalized ones by using the installed capacity of each wind farm [3]. The book edited by Ian T. Jolliffe and David B. Stephenson [4] is an indispensable reference in forecast verification. A review of a wide range of forecast verification methods currently being employed in weather and climate forecasting centres is made. Madsen et al. in Refs. [5,6], introduce a standardized protocol consisting of a set of criteria, for the evaluation of short-term wind power prediction systems. The authors comment some problems with the standard use of some of the usual statistics and introduce a set of reference predictors such as persistence, global mean, and a new reference model. By using this protocol it is possible to derive conclusions on the performance of prediction methods and on which factors may affect this performance (terrain, season, horizons, etc.) This work was developed in the frame of the European R&D project Anemos and the protocol was used to evaluate more than 10 prediction systems in Ref. [7]. Martí et al. in Ref. [8], also in the frame of the Anemos project, present the results of the comparison of nine most used power prediction models. Six wind farms, located in four different countries, were selected to cover a wide range of conditions with respect to climatology and terrain. The authors also develop a distribution-oriented approach for forecast verification. Distribution-oriented approaches are based on the idea that the joint distribution of

* Corresponding author.

E-mail address: lfrias@cener.com (L. Frías-Paredes).

forecasts and observations contains all the non-time dependence information about the quality of a prediction method. Their results showed a dependency of the prediction errors on the complexity of the terrain as well as on the forecast horizon (For more details see Refs. [9–11]). Bielecki et al. in Ref. [12] provide a methodology for the characterization of errors in wind power forecasting. The authors examine errors between the actual and commercially forecasted power production data from a typical wind power plant in the Northwestern United States. The authors include an algorithm for ramp identification (that allows to select significant ramp events from the data) paying particular attention to the error analysis during these events (More detailed information can be found in Ref. [13]). Besides, a beta density function is proposed to fit the distribution of the prediction errors. Different probability distributions have been considered in the literature to model the distribution of wind power forecast errors: Hyperbolic [14,15], Normal [16,17], Weibull [18], Beta [19] and a KDE (kernel density estimation) is considered in Ref. [20].

Roughly speaking, most of the verification methods compare pairs of time series, observations and predictions, and verification measures look at "vertical" distances between the two series. However, when the possibility of one series shows a misalignment in time with respect to the other is considered, a "horizontal" distance or time-distance becomes necessary. We could think of a model that correctly predicts the events although not "on time" but with a certain time lag. It also may happen that the model is capable of identifying correctly the occurrence of an event but not its duration. Fig. 1 intends to illustrate these situations showing the prediction of wind energy generated by LocalPred model currently used by National Renewable Energy Centre (CENER) [21–23], orange (in web version) line, together with the corresponding measured power (black line).

The first and second squares show time lags in the prediction. The model collects the real behavior but a little time ahead of the real series. On the other hand, the two last squares contain events that have been well identified but with less duration than in the real data. Therefore, the model predicts most of the events but not always on time and the estimation of their duration is neither accurate. In these cases it would be desirable a method allowing elastic shifting of the time axis, to accommodate sequences that are similar but out of phase.

An approach to include the timing errors in the assessment of forecast methods has been recently made in Ref. [24] by combining a standard Dynamic Time Warping procedure with Diebold–Mariano test [25]. DTW (Dynamic Time Warping) was introduced in Refs. [26] and [27] as a DP (dynamic programming) based time-normalization algorithm for spoken word recognition. The purpose of DTW was to eliminate timing differences between two speech patterns by warping the time axis of one of them so that the maximum coincidence with the other was attained.

The main goal of this paper is to provide a deeper insight in the analysis of timing errors which leads to the proposal of a new methodology for its control and measure. This methodology is

based on the DTW principles which obtains the optimal alignment of two time series by applying dynamic optimization to a shortest path problem. In this problem nodes represent possible temporal pairings between the two series, while possible transitions and distances are defined by a recursive function. This function manages which temporal leaps are allowed and its associated cost to reach a new coupling. This article shows how these functions allow the user to take into account factors impacting on the management of the renewable energy, i.e. non symmetry considerations in temporal errors, the penalty of temporary advance of one series from the other one or limitation of the time lag, all of them of key importance in storage of renewable energy systems [28]. To be more precise, for different applications suitable adaptations of the DTW basic algorithm have been proposed giving rise to a variety of dynamic-time-warping-based techniques for the alignment of forecast and observed renewable energy data series.

On the other hand, it is well known that forecast quality is a multidimensional concept described by several different scalar attributes such as overall bias, reliability/calibration, uncertainty, sharpness/refinement, accuracy, association, resolution and discrimination [4]. Therefore, in this paper we also present a bi-criteria analysis for the estimation of the accuracy as an attribute of forecast quality. A new dissimilarity measure among time series is presented, the TDI (Temporal Distortion Index), which complements the traditional verification measures found in the literature. This measure minimizes the effects of shifting and distortion in time by allowing an adaptive transformation of the time series. The simultaneous consideration of both error measures, TDI and a traditional verification measure, leads to a bi-criteria perspective when comparing different forecasts.

This paper is organized as follows. Section 2 describes our dynamic-time-warping-based methodology specifically designed to deal with forecasting of renewable energies and the definition of the Temporal Distortion Index. In Section 3 a new bi-criteria forecast quality measurement is detailed together with an example artificially created to illustrate it. A real case is presented in Section 4 with the analysis of the temporal component occurred in a case of wind energy prediction. To finish, the conclusions and future lines of work are summarized in Section 5.

2. Control and measurement of the timing error component

2.1. Classical techniques for time series alignment

Let us introduce and discuss the main ideas of classical DTW which works by warping the time axis iteratively until an optimal match between the two sequences is found. This technique which had been used extensively for speech recognition during the 70s, was introduced to the database community by Berndt and Clifford [29]. It has been applied to the analysis and monitoring of batch processes [30], chromatography [31], gesture recognition [32], surveillance [33], medicine [34], in gene expression studies [35], etc.

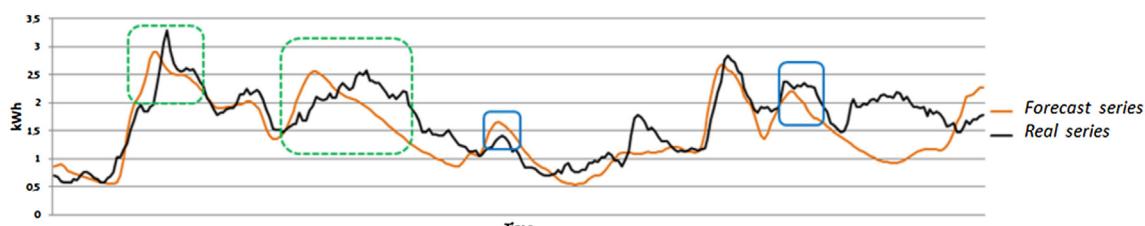


Fig. 1. Examples of temporal events.

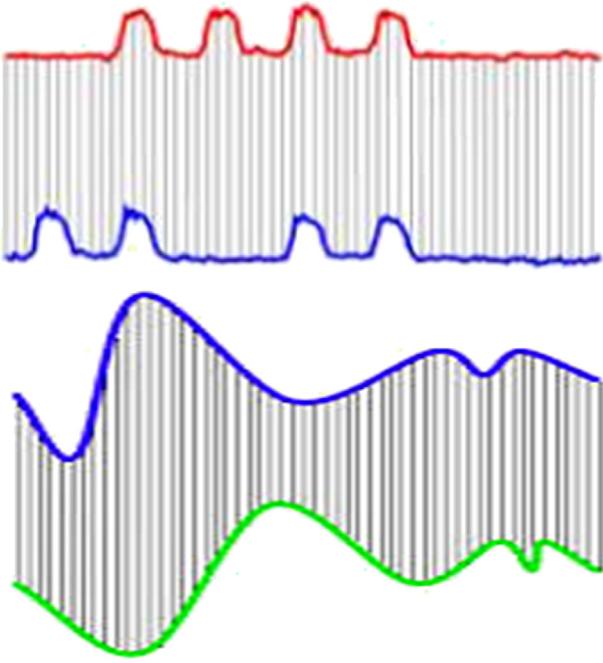


Fig. 2. Sequences are aligned one to one.

Figs. 2 and 3 illustrate the convenience of aligning two time series in order to quantify their similarity. In **Fig. 2** the i -th data point of a sequence is matched to the i -th data point of the other sequence. Any distance (Euclidean, Manhattan, among others) which aligns data in this way would produce a low similarity score. In **Fig. 3** the time axis is warped so that each data point in one sequence is optimally aligned to a point in the other sequence. The similarity score computed after this process gives a more intuitive result.

The process that is used to characterize the temporal component of the prediction error is shown in this section.

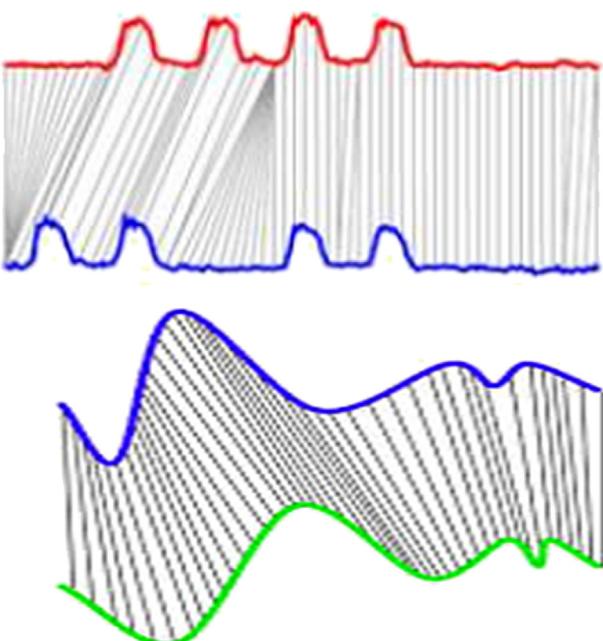


Fig. 3. Nonlinear alignments are possible.

The rationale behind DTW is the following: given two time series the method stretches and/or compresses them locally in order to make one resemble the other as much as possible. So, we work with two time series, the test (or query) series $T = (T_1, T_2, \dots, T_N)$ with $N \in \mathbb{N}$ and the reference series $R = (R_1, R_2, \dots, R_M)$ with $M \in \mathbb{N}$. The DTW processing produces a set of time modifications in the test series to align it into the reference series. Without loss of generality we are considering M equal to N .

As first step, we define the local distance between pairs of elements T_i and R_j as a non-negative function $f : \Phi \times \Phi \rightarrow \mathbb{R}_{\geq 0}$:

$$f(T_i, R_j) = d(T_i, R_j) = d_{ij} = \|T_i - R_j\| \geq 0 \quad (1)$$

The DTW algorithm begins by the setting-up of a matrix of local distances ($d \in \mathbb{R}^{N \times N}$) -called Local Cost Matrix-that contains all pairs of corresponding distances between both series. In **Fig. 4** we can visualize the Local Cost Matrix in the left part and the graphical version of it in the right. This graphical presentation of the matrix is useful in the interpretation of results.

Note that each cell (i,j) represents the local distance between the i -th element in the test series (T_i) and the j -th element in the reference one (R_j).

Once the Local Cost Matrix is defined the concept of path between series is introduced as a sequence of points $w = (w_1 = (i_1, j_1), w_2 = (i_2, j_2), \dots, w_l = (i_l, j_l), \dots, w_k = (i_k, j_k))$ $k \in \mathbb{N}$, where $w_l = (i_l, j_l) \in [1 : N] \times [1 : N]$ for $l \in [1:k]$ under the following restrictions:

- Boundary conditions: $w_1 = (1,1)$ and $w_k = (N,N)$. This condition requires that the path must begin and end on the first and last points, respectively, of the sequences.
- Monotonicity condition: Given $w_l = (i_l, j_l)$ then $w_{l-1} = (i_{l-1}, j_{l-1})$ being $i_l - i_{l-1} \geq 0$ and $j_l - j_{l-1} \geq 0$. This condition ensures that through the path the points are sorted according to time.
- Condition of continuity or step size: Given $w_l = (i_l, j_l)$ then $w_{l-1} = (i_{l-1}, j_{l-1})$ being $i_l - i_{l-1} \leq 1$ and $j_l - j_{l-1} \leq 1$. This condition ensures that the path does not present big jumps, even more, it is restricted to neighboring points.

A path can be plotted on a grid $N \times N$ where the x -axis represents the temporal index in the test series and the y -axis represents the temporal index in the reference series (See **Fig. 5**).

Now, the total cost associated to a path w between series T and R, with respect to its local distance, is identified by $c_w(T, R)$ and is calculated with the next expression:

$$c_w(T, R) := \sum_{l=1}^k d(T_{i_l}, R_{j_l}) \quad \text{where } (i_l, j_l) = w_l$$

Accordingly, an optimal path between the series T and R is a path w^* that presents a minimum total cost taking into account all the possible paths:

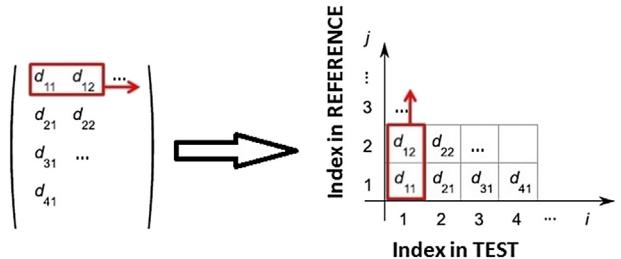


Fig. 4. Local Cost Matrix in matricial and graphic format.

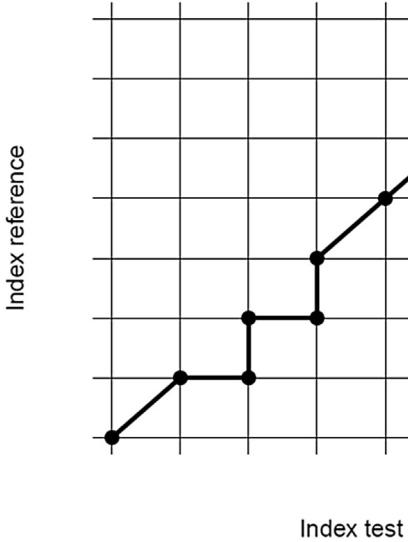


Fig. 5. Example of path.

$$c_{w^*}(T, R) = \min\{c_w(T, R) | w \text{ is a path}\} \quad (2)$$

The optimal path w^* is obtained by using dynamic programming. Dynamic programming (DP) is a procedure that solves optimization problems by breaking them down into simpler problems. The basis of DP is Bellman's principle of optimality [36].

We identify the sequences $T(1:i) := (T_1, T_2, \dots, T_i)$ where $i \in [1:N]$ and $R(1:j) := (R_1, R_2, \dots, R_j)$ where $j \in [1:N]$ and we define:

$$D(i, j) = c_{w^*}(T(1:i), R(1:j)) \quad (3)$$

Then, $D(i, j)$ express the cost associated to the best path matching the series (T_1, T_2, \dots, T_i) and (R_1, R_2, \dots, R_j) . So, the values $D(i, j)$ define a matrix $D \in \mathbb{R}^{N \times N}$ which is refereed by the name of accumulated cost matrix (Cost Matrix). Clearly $D(N, N) = c_{w^*}(T, R)$.

Taking into account the constraints cited in the path definition, the calculation of the Cost Matrix elements can be automatized using dynamic programming by means of the following recursive formula:

$$D(i, j) = \begin{cases} D(i-1, j) + d(i, j) \Rightarrow \text{Branch(1)} \\ D(i, j-1) + d(i, j) \Rightarrow \text{Branch(2)} \\ \underbrace{D(i-1, j-1) +}_{\text{Cumulative-Cos t}} \underbrace{d(i, j)}_{\text{Actual-Cos t}} \Rightarrow \text{Branch(3)} \end{cases} \quad (4)$$

Note that there are three possible transitions from one pair in the path to the next one, all of them verifying the continuity condition. Fig. 6 contains the notational scheme of the recursive formula, also called Step Pattern.

The test series is transformed, using the optimal path, in a new series called aligned series and denoted by $S = (S_1, S_2, \dots, S_N)$ ($N \in \mathbb{N}$), with a smaller vertical difference to the reference series than the test series.

2.2. Interpreting the optimal path and controlling the degree of time coupling between series

Let us provide an interpretation of the movements executed by the optimal path to clarify how they affect to the transformation of the time axis. The different possibilities are analyzed below.

- Horizontal segments:

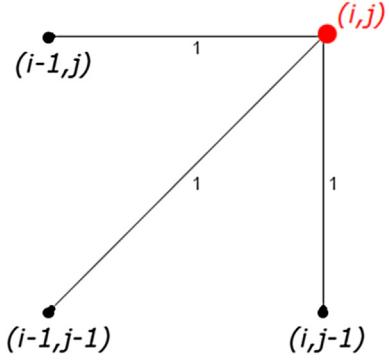


Fig. 6. Notational scheme of recursive formula. Representation of so called Step Pattern Symmetric1.

Let us the optimal path $w^* = (w_1, w_2, \dots, w_l, w_{l+1}, \dots, w_{l+r}, \dots, w_k)$, a horizontal segment is defined by $w_l = (i_l, j_l)$, $w_{l+1} = (i_{l+1}, j_l), \dots, w_{l+r} = (i_{l+r}, j_l)$ (see Fig. 7).

That is, an index j_l in the reference series is associated to more than one consecutive index $(i_l, i_{l+1}, \dots, i_{l+r})$ in the test series. The aligned series is defined as:

$$S_{j_l} = T_{\frac{i_l+i_{l+r}}{2}} = T_{i_l+\frac{r}{2}}$$

If r is an odd number then $T_{i_l+\frac{r}{2}}$ is obtained by interpolating its neighbors $T_{\left[i_l+\frac{r}{2}\right]}$ and $T_{\left[i_l+\frac{r}{2}\right]+1}$.

Therefore, when a horizontal segment, with length r , appears in the optimal path, the new series progresses $i_l + \frac{r}{2}$ units of time with respect to the current temporal position. That is, the effect is like traveling forward in time, i.e. "the events are forced to happen sooner". The events are identified with delay by the test series and they appear before in the aligned series.

- Vertical segments:

Let us the optimal path $w^* = (w_1, w_2, \dots, w_l, w_{l+1}, \dots, w_{l+r}, \dots, w_k)$, a vertical segment is defined by $w_l = (i_l, j_l)$, $w_{l+1} = (i_l, j_{l+1}), \dots, w_{l+r} = (i_l, j_{l+r})$ (see Fig. 8).

That is, an index i_l in the test series is associated to more than one consecutive indices $(j_l, j_{l+1}, \dots, j_{l+r})$ in the reference series. The aligned series is defined as:

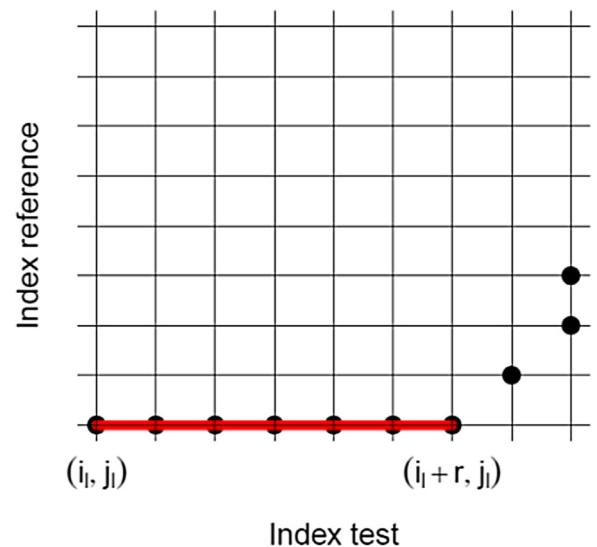


Fig. 7. Horizontal segment in the optimal path.

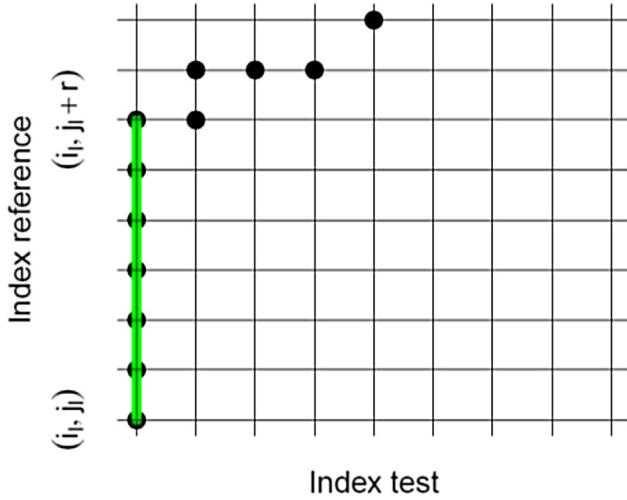


Fig. 8. Vertical segment in the optimal path.

$$S_{j_l+d} = T_{i_l} \quad (0 \leq d \leq r, d \in \mathbb{N})$$

Therefore, when a vertical segment appears, with length r , the new series stops r units of time with respect to the current temporal position taking the value of the original series in $t = i_l$. That is, the effect is like stopping the time, i.e. "the events are forced to occur later". The events are identified early by the test series and they appear later in the aligned series.

- Segments with slope equal to 1:

Let us the optimal path $w^* = (w_1, w_2, \dots, w_l, w_{l+1}, \dots, w_k)$, a segment with slope equal to 1 is defined by $w_l = (i_l, j_l)$, $w_{l+1} = (i_l + 1, j_l + 1), \dots, w_{l+r} = (i_l + r, j_l + r)$ (see Fig. 9).

The aligned series is defined as:

$$S_{j_l+d} = T_{i_l+d} \quad (0 \leq d \leq r, d \in \mathbb{N})$$

Note that segments with slope equal to 1 suppose that no time modification is made.

However, the condition of continuity included in the former path definition can be relaxed by introducing additional possible transition between consecutive elements in the path.

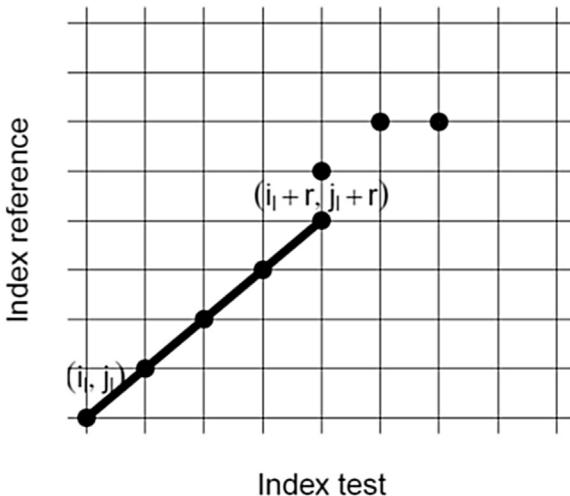


Fig. 9. Segment with slope equal to 1 in the optimal path.

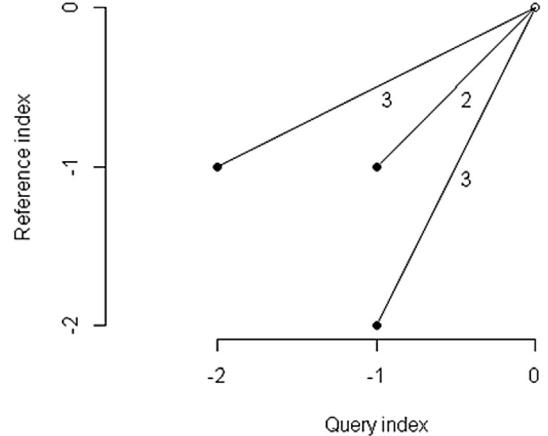


Fig. 10. Example of additional possible transition.

Also by weighing local distances it is possible to penalize certain movements as can be seen in the example shown in Fig. 10 and Equation (5).

$$D(i, j) = \begin{cases} D(i-2, j-1) + 3d(i, j) \\ D(i-1, j-1) + 2d(i, j) \\ D(i-1, j-2) + 3d(i, j) \end{cases} \quad (5)$$

Thus new movements could appear in the optimal path which are interpreted below:

- Segments with slope less than 1:

Let us the optimal path $w^* = (w_1, w_2, \dots, w_l, w_{l+1}, \dots, w_k)$, a segment with slope less than 1 is defined by $w_l = (i_l, j_l)$, $w_{l+1} = (i_l + a, j_l + b), \dots, w_{l+r} = (i_l + ar, j_l + br)$, with $a > b$ (see Fig. 11). In this case the aligned series is defined as:

$$S_{j_l+d} = T_{i_l+\frac{da}{b}} \quad (0 \leq d \leq rb, d \in \mathbb{N})$$

If $i_l + \frac{da}{b} \notin \mathbb{N}$ then $T_{i_l+\frac{da}{b}}$ is obtained by interpolating its neighbors $T_{\left[i_l+\frac{da}{b}\right]}$ and $T_{\left[i_l+\frac{da}{b}\right]+1}$.

In this case, the time is being compressed, that is, time is continuously accelerated. An event that took a time period equal to a , now it presents a time duration of b where $b < a$.

- Segments with slope greater than 1:

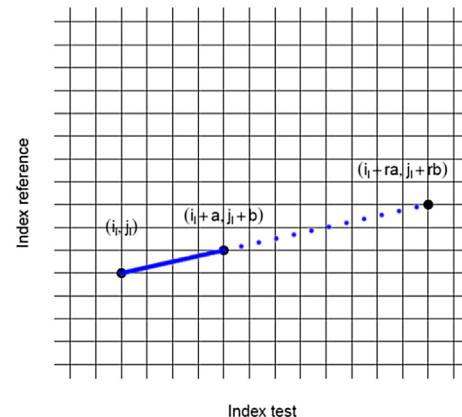


Fig. 11. Segment with slope equal to 1 in the optimal path.

In this case, the time is being expanded, that is, time is continuously decelerated. An event that took a time period equal to a , now it presents a time duration of b where $b > a$.

Consequently the aligned series can be calculated following the next expressions,

$$S_j = T_{f(j)} \quad (j = 1, \dots, n) \quad (6)$$

where f is the interpolation function that fulfills $f(j_l) = i_l^1$ ($l = 1, \dots, k$) and

$$S_j = g(f(j)) \quad (7)$$

whenever $f(j) \notin \mathbb{N}$ and where g represents the interpolation function that verify $g([f(j)]) = T_{[f(j)]}$ and $g([f(j)]+1) = T_{[f(j)]+1}$.

Now some modifications to the set of constraints are included in order to obtain a higher control of the possible paths. In this way, the paths can be controlled according to the practical needs of the analyzed case. For instance, it can be necessary to check that paths never cross the main diagonal and therefore always stay ahead or behind in time with respect the reference series. This can be key in the forecast of variables related to the sources of energy capable of storing, since prediction errors will not be considered symmetrical. This condition would be collected by the recursion presented in Equations (8) and (9):

a) Test series always stays behind of reference series:

$$D(i,j) = \begin{cases} D(i-1,j) + d(i,j) \\ D(i,j-1) + d(i,j) \quad s.t. \quad i \leq j \\ D(i-1,j-1) + d(i,j) \end{cases} \quad (8)$$

b) Test series always stays ahead of reference series:

$$D(i,j) = \begin{cases} D(i-1,j) + d(i,j) \\ D(i,j-1) + d(i,j) \quad s.t. \quad i \geq j \\ D(i-1,j-1) + d(i,j) \end{cases} \quad (9)$$

In practice, it would be interesting that the time lag with respect to the reference series is limited by a constant value C (see Equation (10)). This condition implies that the optimal path is contained by a band like the one showed in Fig. 12.

c) Limiting the maximum time lag between series:

$$D(i,j) = \begin{cases} D(i-1,j) + d(i,j) \\ D(i,j-1) + d(i,j) \quad s.t. \quad |i-j| < C \\ D(i-1,j-1) + d(i,j) \end{cases} \quad (10)$$

Accordingly, it is needed to define and to control the recursive formula (RF) that will be used before the calculus of the Cost Matrix and the optimal path. The recursive formula collects the kind of allowed movements and the respective weights associated to each direction.

2.3. Measurement of the timing error component. Temporal Distortion Index (TDI)

An identity path is understood as the one that is formed by the following points $\{(1,1), (2,2), \dots, (N-1, N-1), (N, N)\}$ and it will be denoted by w_l . Note that when the test series and the reference

¹ When the points are vertically positioned we will work with the midpoint.

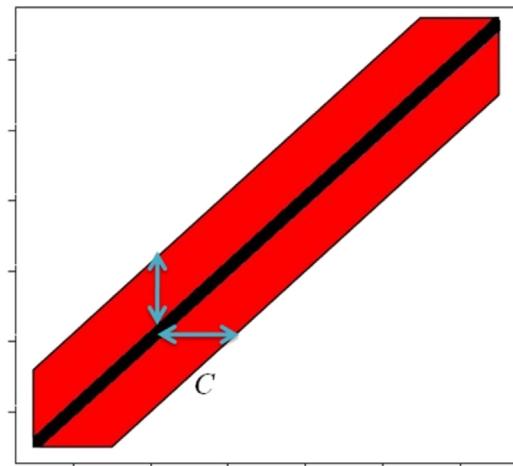


Fig. 12. Bands.

series coincide it is fulfilled that $w^* = w_l$. Thus the obtained aligned series does not make any temporal correction within the test series, therefore it is deduced that $T = S$ ($T_i = S_i \forall i$).

If the optimal path w^* passes through the point $w_l = (i_l, j_l)$ it can be inferred that there has been a temporal movement of $|i_l - j_l|^2$ units of time between the test series and the aligned series. Thus, the greater $|i_l - j_l|$, the greater temporal misalignment in the test series in order to match the reference series. Therefore, this difference should be taken into account to obtain the element $D(i,j)$ of the Cost Matrix, especially when the minimum of the recursive formula is reached in two or more different branches of the recursive formula.

For instance, if two possible predecessors to the element (i,j) , (i^*, j^*) and (i^{**}, j^{**}) , provide the same value for $D(i,j)$, the pair producing the minor temporal modification within the test series should be selected, $\min\{|i^* - j^*|, |i^{**} - j^{**}|\}$. In this way, the test series and the aligned series are as much synchronized as possible.

Consequently, the resulting aligned series is the one, among all those series associated to an optimal path, providing a minor temporal misalignment of the test series. Equivalently, the aligned series is the one whose associated optimal path is the closest to the identity path.

Hence, a global measure of the temporal distortion carried out in the test series, in order to obtain the aligned series, is provided by the area between the resulting optimal path and the identity path, which is denoted by TDI (Temporal Distortion Index). So, this parameter will serve to describe the temporal component of the error. The expression of this measure is collected in Equation (11).

$$P_l = \int_{i_l}^{i_{l+1}} \left(x - \frac{(x - i_l)(j_{l+1} - j_l)}{(i_{l+1} - i_l)} + j_l \right) dx$$

$$TDI = \frac{2 \sum_{l=1}^{k-1} \left| P_l \right|}{N^2} \quad (11)$$

TDI is a dimensionless number varying in the interval [0,1], where 0 corresponds with the null temporal distortion and 1 with the maximum temporal distortion, which occurs when the optimal path follow the bounds of the index plot.

² In case of $w_l = (i_l, j_l)$ be included in a horizontal segment, with a length r , we will have a forward leap of $|i_l + \frac{r}{2} - j_l|$.

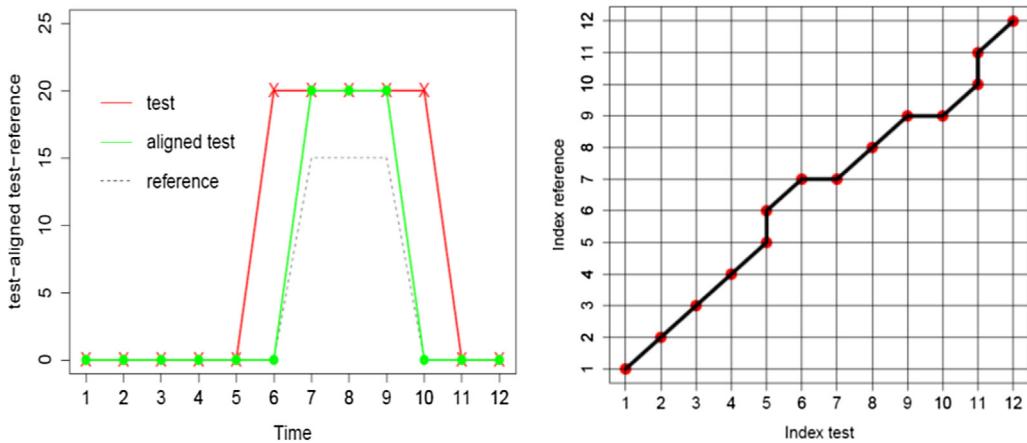


Fig. 13. Homothecy and associated optimal path.

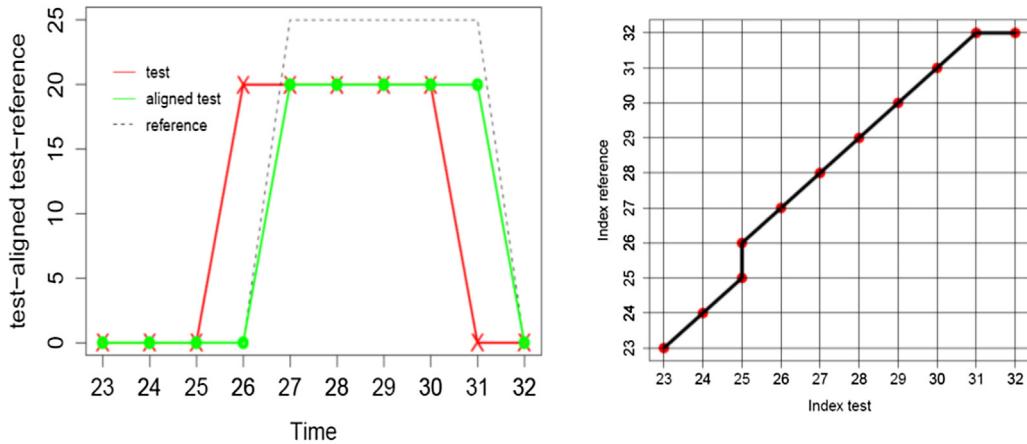


Fig. 14. Translation and associated optimal path.

Therefore, the interpretation of the deviation from the identity path of the optimal path plays a key role in the measurement of the timing errors. It is shown below how two typical effects observed in a prediction (homothecy and translation) are captured by the TDI and reflected in the aligned series. Fig. 13 contains an example of a homothecy and the associated optimal path.

To correct a homothecy, from the temporal point of view, the optimal path needs two opposite separations from the identity

path. In each one, time is dragged and accelerated in the right way so that the new aligned series is time fitted to the reference one.

Fig. 14 contains an example of a translation and the optimal associated path.

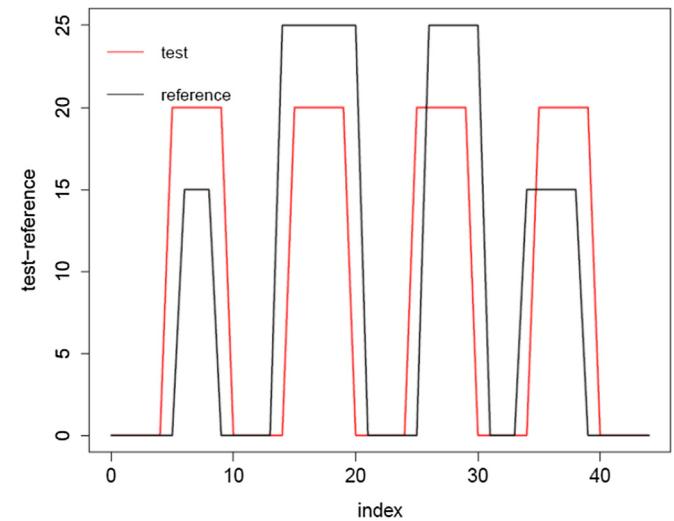


Fig. 15. Time series values. Reference and test.

Table 1
Time series values. Reference and test.

	Test	Reference		Test	Reference		Test	Reference	
1	0	0		16	20	25	31	0	25
2	0	0		17	20	25	32	0	0
3	0	0		18	20	25	33	0	0
4	0	0		19	20	25	34	0	0
5	0	0		20	20	25	35	0	15
6	20	0		21	0	25	36	20	15
7	20	15		22	0	0	37	20	15
8	20	15		23	0	0	38	20	15
9	20	15		24	0	0	39	20	15
10	20	0		25	0	0	40	20	0
11	0	0		26	20	0	41	0	0
12	0	0		27	20	25	42	0	0
13	0	0		28	20	25	43	0	0
14	0	0		29	20	25	44	0	0
15	0	25		30	20	25	45	0	0

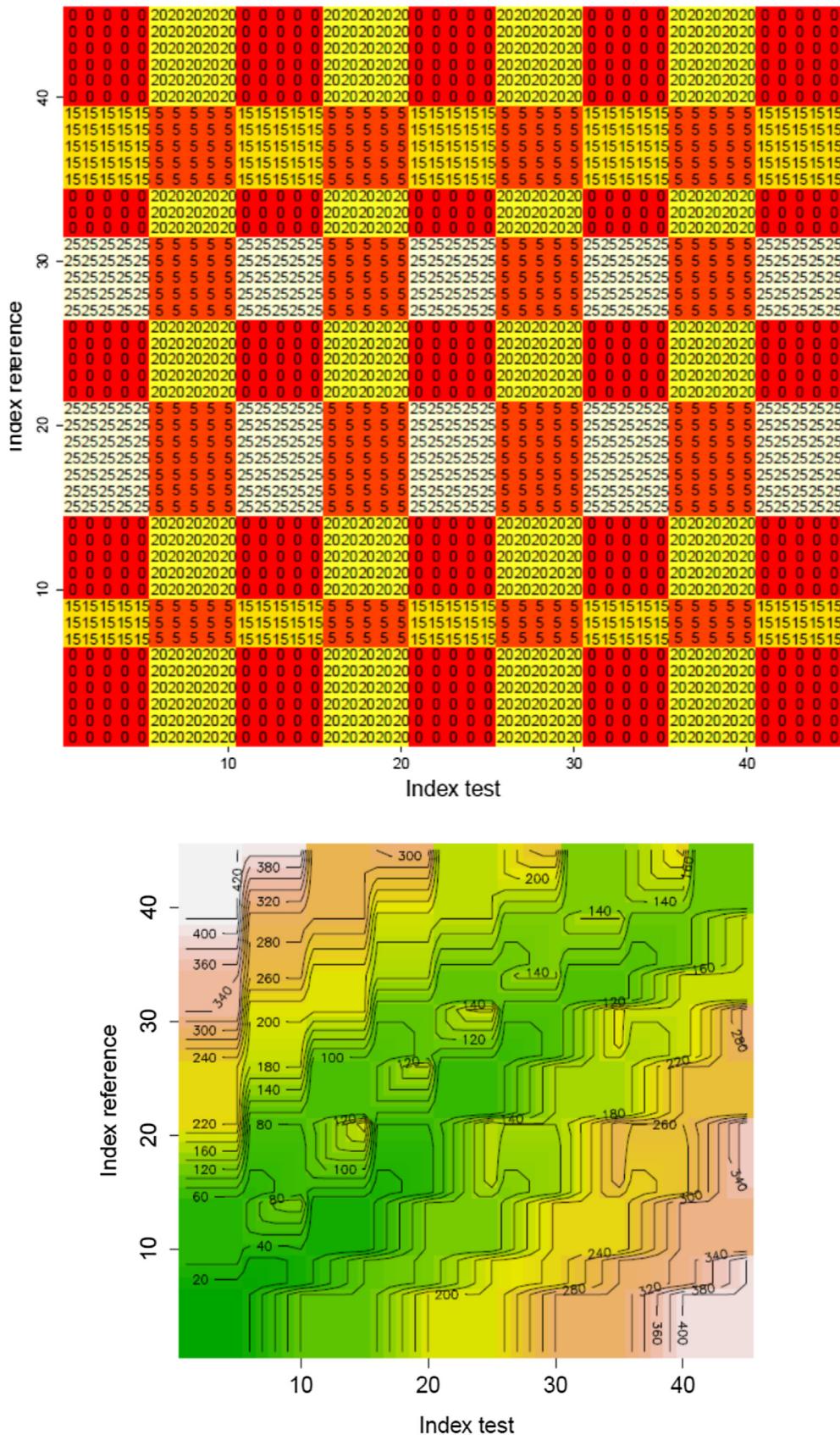


Fig. 16. Local Cost Matrix and density plot of Cost Matrix.

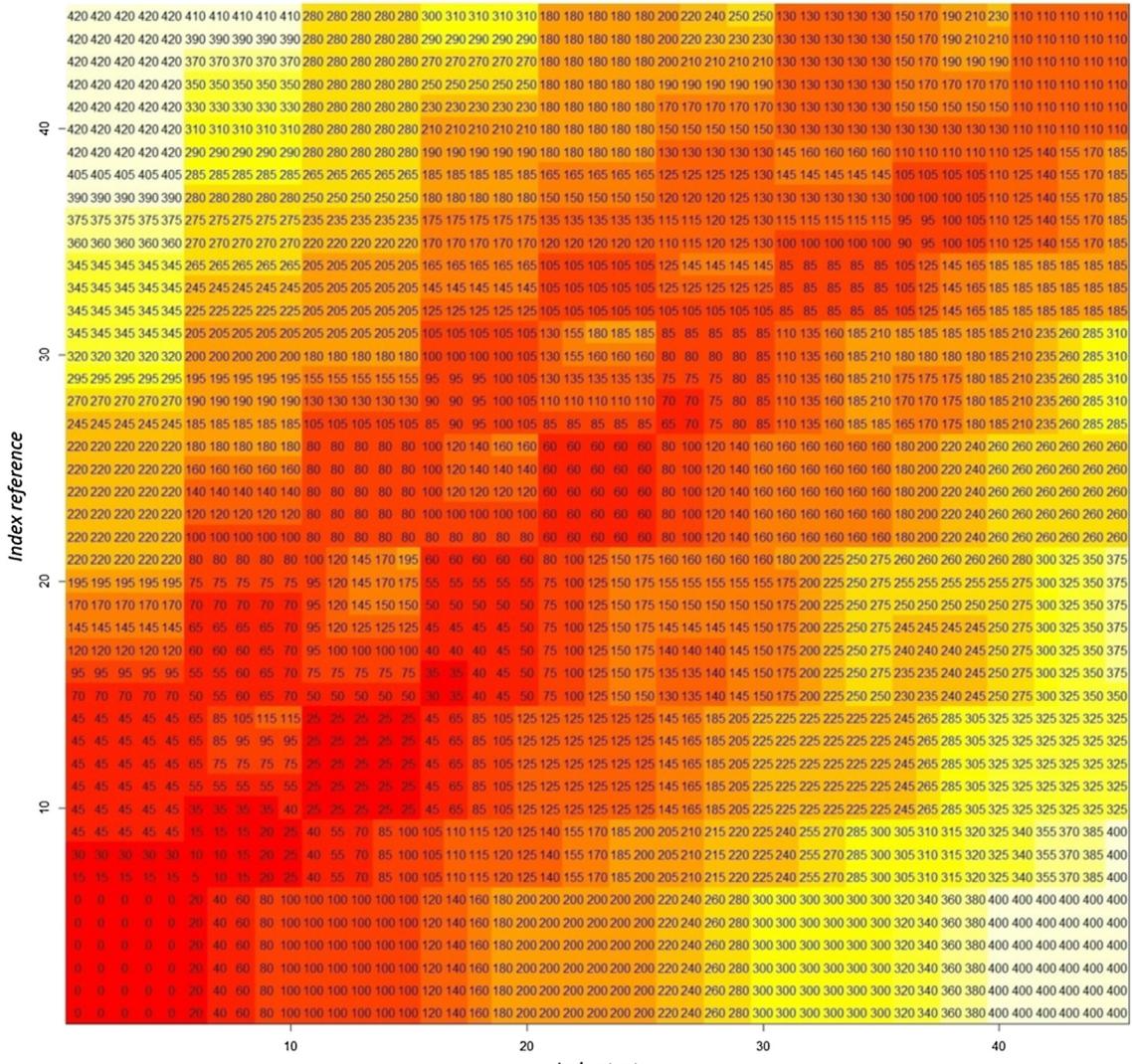


Fig. 17. Cost matrix.

Note that to correct a translation, from the temporal point of view, the optimal path has needed a continuous separation from the identity path related with the magnitude and direction of the translation.

3. Methodology to obtain the bi-dimensional measure for the forecast error assessment

The TDI measure, defined in previous section, is complemented by a typical error statistic of accuracy (static error) between the reference series and the aligned series, the MAE:

$$MAE(S) = \sum_{i=1}^N \frac{|S_i - R_i|}{N} \quad (12)$$

These pair of measures make up the bidimensional error vector (with static and temporal components) which is denoted by BE_{RF} :

$$BE_{RF}(T, R) = (TDI, MAE(S))_{RF} \quad (13)$$

The subscript RF indicates the recursive function used. It is important to note that different alignments (obtained from

different recursive functions) produce different bidimensional error vectors.

Note that the bidimensional error between the test series and the reference series when no alignment is made, i.e. when $RF = 0$, fulfills that the first component is equal to 0 ($TDI = 0$) meanwhile the second one contains the $MAE(T)$ and is denoted by:

$$BE_0(T, R) = (0, MAE(T)) \quad (14)$$

An illustrative example. An example using synthetic data is presented containing basic situations to illustrate the whole methodology proposed in this work to assess the forecast error. First, a time series playing the role of the reference series (the observed of the renewable energy) is considered, and then, from this one, another time series is generated by mean of a set of time modifications, translations and homotheties in both directions. This modified series plays the role of the test series (the forecast of the renewable energy). Table 1 contains the time series values and Fig. 15 plots both series. The test series is plotted in red (in web version) and the reference one in black. Next steps describe the methodology to characterize the temporal component of the prediction error and create the BE_{RF} vector.

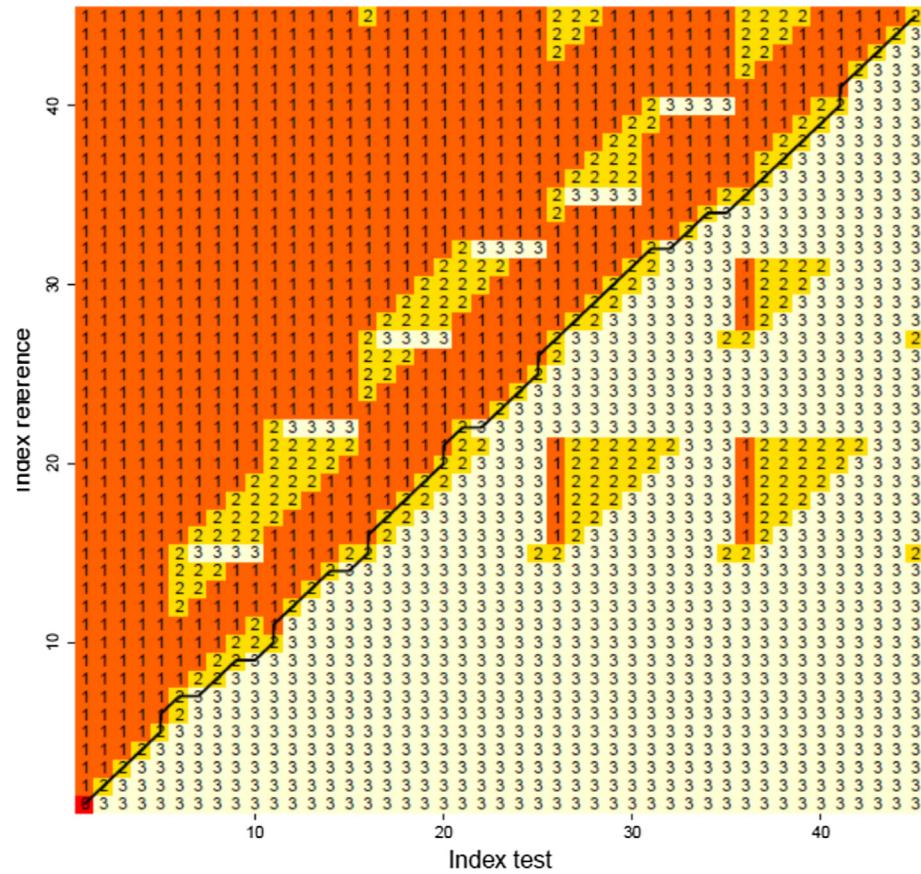


Fig. 18. Direction cost matrix.

Step 1. Selection of the recursive formula:

In this demonstrative example the simple recursive formula represented by the Step Pattern named symmetric 1 is used (see Equation (4)).

Step 2. Calculation of the Local Cost Matrix and the Cost Matrix: The Local Cost Matrix is calculated following Equation (1). The result is displayed at the top of Fig. 16. The associated Cost Matrix (calculated according to Equation (3)) is displayed in

Fig. 17. Different colors are used to represent the different levels of values in the matrix, grading from red (in web version), for the lowest values, to white, for the highest. An alternative graphical representation of the Cost Matrix is used to display the results, as it can be seen in Fig. 16. This density plot of the Cost Matrix is very useful when the visualization of the typical Cost Matrix plot is not possible or difficult due to the complex structure of its value distribution.

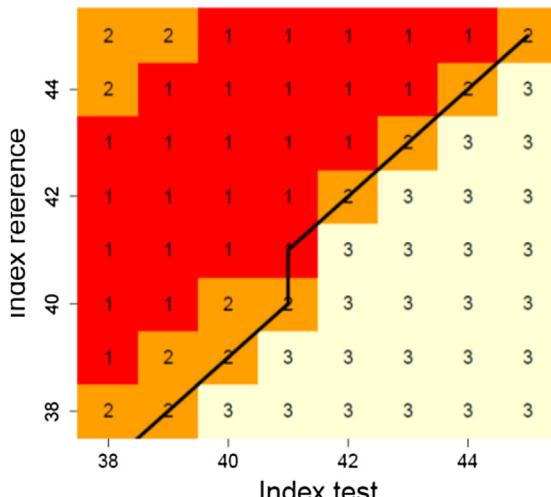


Fig. 19. Zoom of the upper right corner of the Direction Cost Matrix.

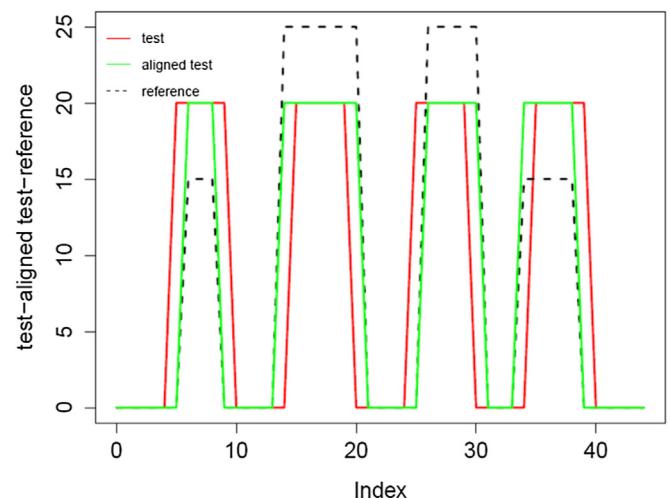


Fig. 20. Time series values. Reference, test and aligned.

Table 2

Time series values. Reference, test and aligned.

	Test	Reference	Aligned		Test	Reference	Aligned		Test	Reference	Aligned
1	0	0	0		16	20	25		31	0	25
2	0	0	0		17	20	25		32	0	0
3	0	0	0		18	20	25		33	0	0
4	0	0	0		19	20	25		34	0	0
5	0	0	0		20	20	25		35	0	15
6	20	0	0		21	0	25		36	20	20
7	20	15	20		22	0	0		37	20	15
8	20	15	20		23	0	0		38	20	15
9	20	15	20		24	0	0		39	20	15
10	20	0	0		25	0	0		40	20	0
11	0	0	0		26	20	0		41	0	0
12	0	0	0		27	20	25		42	0	0
13	0	0	0		28	20	25		43	0	0
14	0	0	0		29	20	25		44	0	0
15	0	25	20		30	20	25		45	0	0

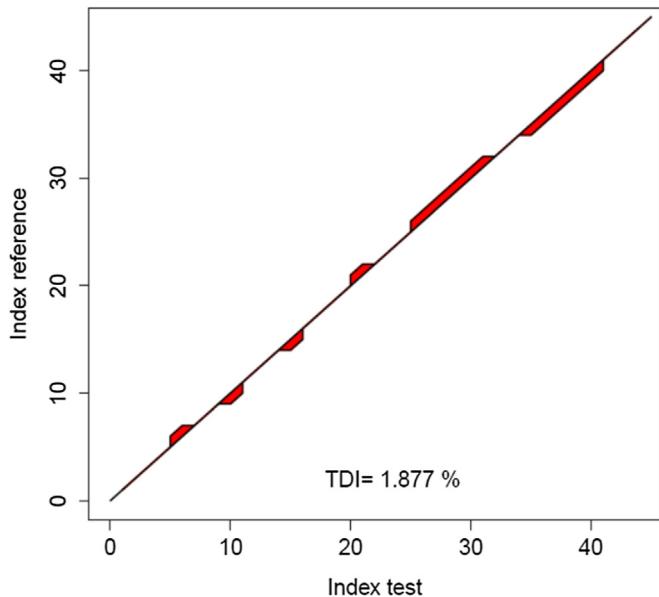


Fig. 21. Area bounded between the optimal path and the identity path.

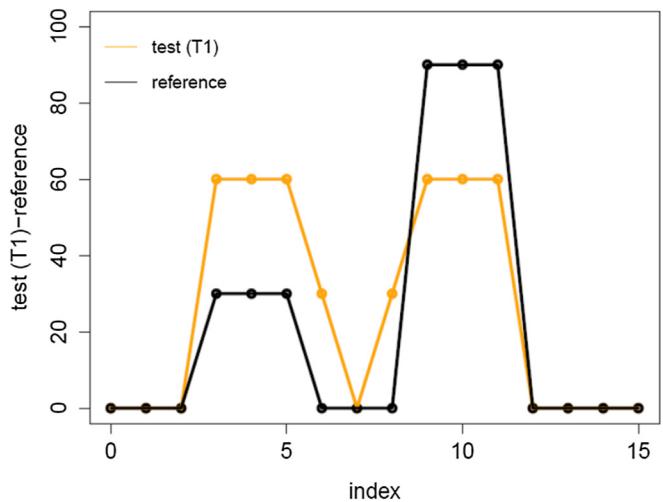


Fig. 23. Test T1 and reference.

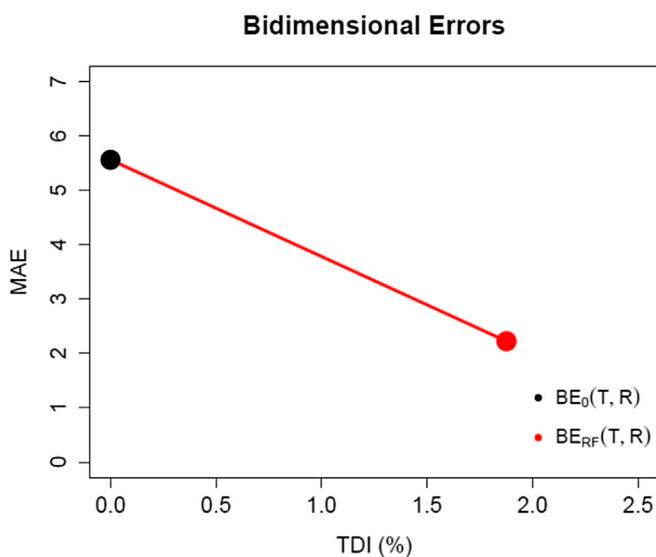


Fig. 22. Bidimensional errors.

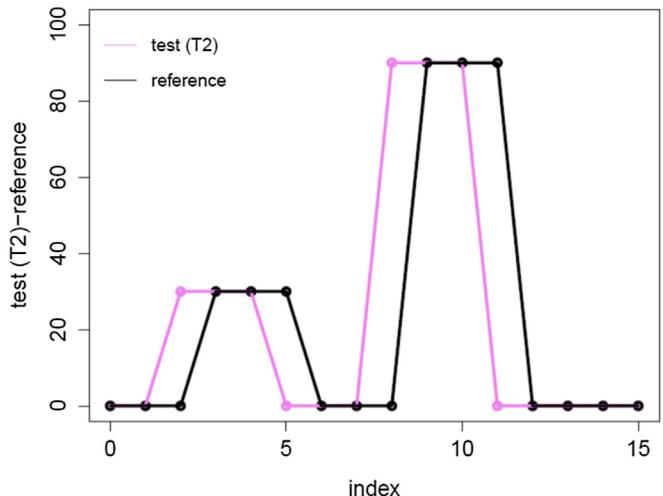


Fig. 24. Test T2 and reference.

Step 3. Determination of the optimal path:

The calculation of the optimal path w^* by using Equation (2) is based on a matrix that collects the information about the branch of the recursive formula chosen in the calculation of each element $D(i,j)$ of the Cost Matrix. This matrix is called Direction Cost Matrix (DCM). Fig. 18 displays the DCM for the illustrative example. The optimal path is easily obtained starting its construction from its last element $w_k = (N,N)$. Let us consider the illustrative example and the zoom of the upper right corner of the DCM in Fig. 19. The value of the element (N,N) in the matrix is 2, which indicates that the pair $w_k = (N,N)$ was reached from $w_{k-1} = (N-1,N-1)$ by using the branch 2 of the recursive function. Then, the value 2 in the position $(N-1,N-1)$ of the DCM indicates that $w_k = (N-1,N-1)$ was reached from $w_{k-2} = (N-2,N-2)$ by using again the branch 2. These backwards movements to recover the optimal path are followed until the path's origin $w_1 = (1,1)$ is reached:

$$\begin{aligned} w_k &= (N, N) \xrightarrow{(2)} w_{k-1} = (N-1, N-1) \xrightarrow{(2)} w_{k-2} \\ &= (N-2, N-2) \xrightarrow{(2)} w_{k-3} \\ &= (N-3, N-3) \xrightarrow{(2)} w_{k-4} \\ &= (N-4, N-4) \xrightarrow{(1)} w_{k-5} = (N-4, N-5) \dots \end{aligned}$$

Step 4. Construction of the aligned series:

The aligned series S is built up by using Equations (6) and (7). Fig. 20 and Table 2 show the aligned series associated to the optimal path determined in the Step 3 for the illustrative example. The green line corresponds to the aligned series and the red (in web version) one corresponds to the original test series. The reference series is plotted by the black dots. The aligned series is capable to capture the different time distortions introduced in the reference series of the synthetic data: both, translations and homothethies, are detected and perfectly corrected by the methodology.

Step 5. Calculation of the Temporal Distortion Index, TDI, and the bi-dimensional error vector, BE_{RF} :

The Temporal Distortion Index, TDI, is calculated according to Equation (11), measuring the area between the optimal path and the identity path. Fig. 21 displays both paths and highlights the area between them. The bidimensional error vector, introduced in expression (13) is graphically represented in Fig. 22 by the red (in web version) point. Note that the black point represents the bidimensional error between the test series and the reference series when no alignment is made ($RF = 0$).

This Section ends with another example artificially created to highlight the advantage of including the temporal error in the assessment of the forecast accuracy. A reference series and two different test (forecast) series are considered. They are displayed in Fig. 23 and Fig. 24. Test series T2 completely catches the shape, size and duration of the events in the reference series but it is translated a few time units. Test series T1 provides a worst prediction of the shape, size and duration of those events although their prediction is better aligned. The static measurement of the error via MAE results in the same value for both forecast series. However, when the TDI is calculated and the bi-dimensional error vector considered (see Fig. 25), the difference in both predictions comes out. All lack of accuracy in T2 is due to the temporal error as it is expressed by $BE_{RF}(T2,R) = (0.4688,0)$ while in T1 most of the error is of static nature, $BE_{RF}(T1,R) = (0.1172,12.1875)$.

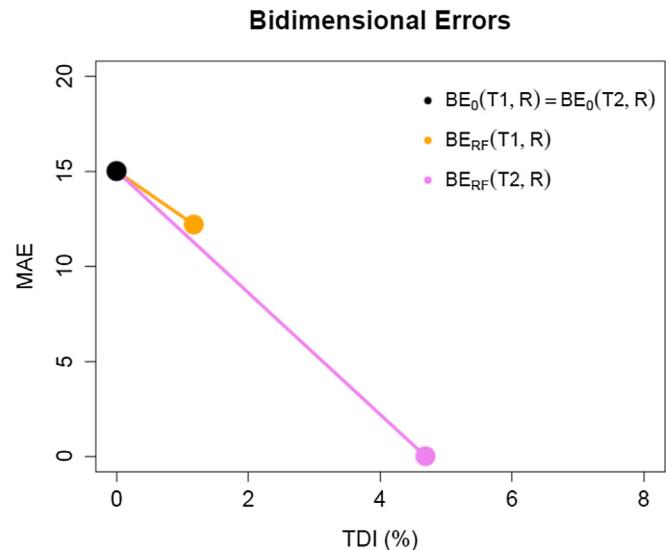


Fig. 25. Bidimensional errors.

4. Bidimensional error measurement: a real case

In previous sections we have presented the methodology for assessing the temporal dimension of the error forecast and have illustrated its main features by using synthetic data. In this section we test the methodology by using real data series: a real wind energy series is compared with a wind energy forecast. Furthermore, we provide a deeper insight in the bi-dimensional measure of the error by considering different recursive formulas.

The time series consist of, on one hand, three days of wind energy production in a wind farm located in the north of Spain and, on the other hand, the respective energy prediction made 72 h in advance. The wind energy prediction has been obtained by the LocalPred model currently used by the National Renewable Energy Center (CENER, www.cener.com). The measured energy series will play the role of reference series while the predicted energy series plays the role of test series. Both series are plotted in Fig. 26.

Following the methodology exposed in Section 3, the temporal error is measured by the area between the identity path and the optimum path associated to the selected recursive formula (see TDI

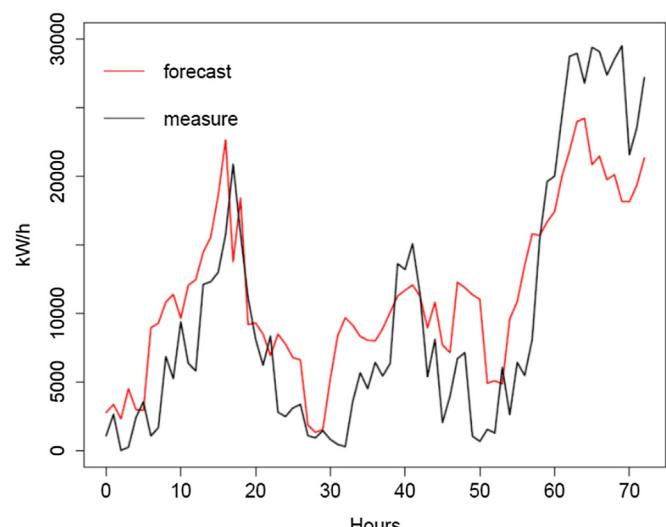
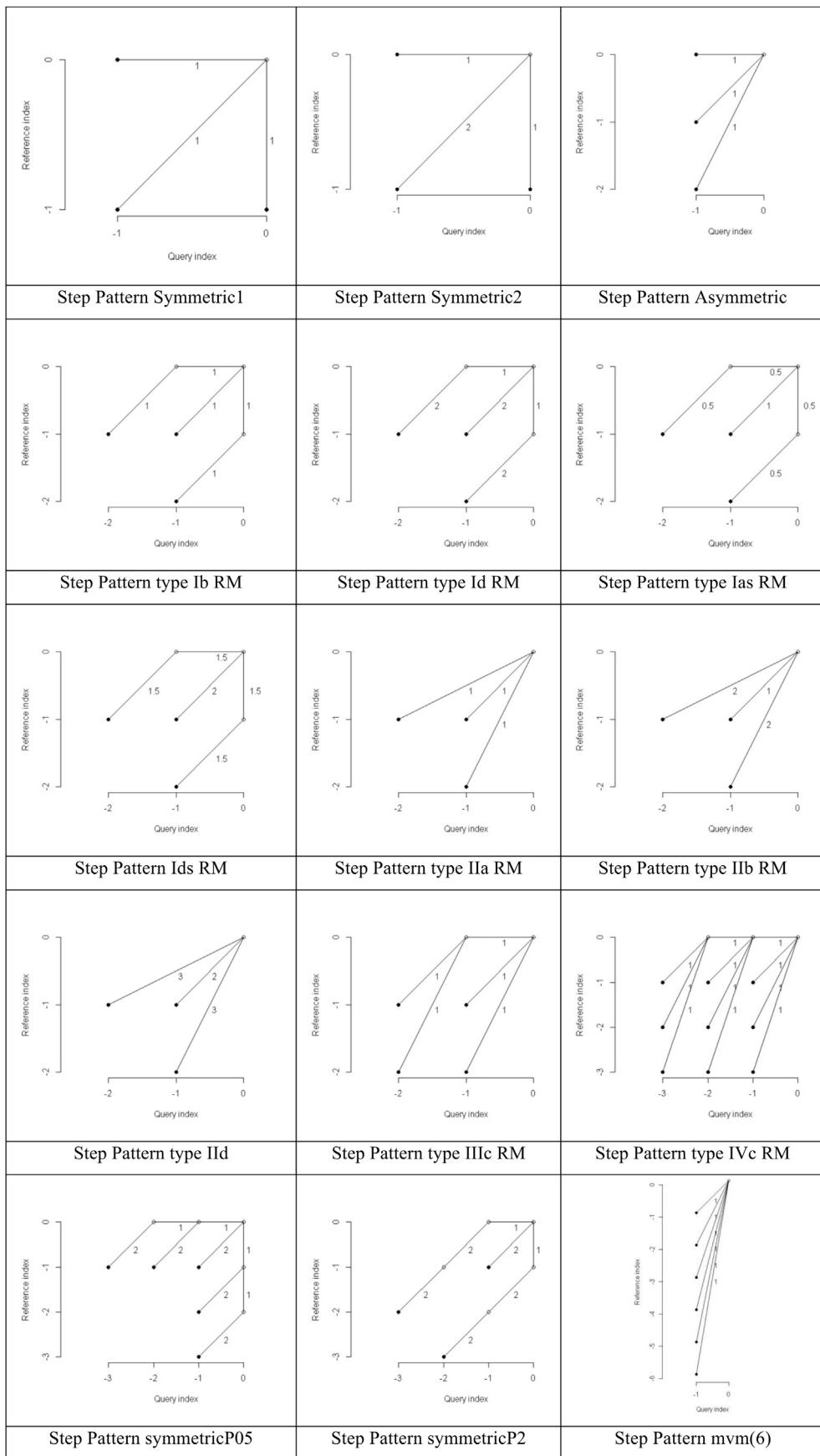


Fig. 26. Three days of predicted and measured wind energy.

Table 3
Step Patterns.



in Equation (11)) and the static component of the error by the mean absolute error between the real time series and the aligned series (see MAE in Equation (12)). As we commented in Section 2, the recursive formula defines the degree of freedom left to the optimal path to match both series. In general, the more freedom is left to the path matching both series the greater the TDI is, and also, the smaller the static component error is. To illustrate this behavior we have considered a set of recursive formula that differ among them in amplitude of the allowed steps as well as in the weights assigned to different movements. The name of the step patterns associated to these recursive formula and their graphical representations are included in Table 3. A study of their properties is included in Refs. [27,37]. The set of the associated bidimensional error vectors (see Equation (13)) is represented in Fig. 27. In general, it is observed that an increase in the temporal error is accompanied by a reduction in the static error, although it is not always the case: there is some recursive formula providing a better value for both components of the error than other ones. In these cases we can discard the dominated recursive formula and only keep to the error analysis the non-dominated ones. In this way a Pareto Frontier [38] is obtained (line black in Fig. 27).

This Pareto Frontier better characterizes the error in the forecast than only one value or pair of values. Nevertheless, in order to simplify the assessment of forecast errors and for ease the comparison between different methods one or few representative points should be selected. These representative points could be identified through bounds on the allowed temporal error or analyzing the trade-off between both error components. One should choose a bidimensional error assessment with a temporal error so that increasing it doesn't give much better value for the static error. More precisely, in the plot of the Pareto Frontier the first error assessment will reduce much static error but at some point the marginal gain will drop, giving an angle in the graph. A representative bidimensional error assessment should be chosen at this zone, and hence the "elbow criterion". Fig. 28 contains an example of this situation, the analysis correspond to comparison between wind speed forecasts and the related measurements.

5. Conclusions and future lines

A key issue when analyzing the behavior of a prediction model is the study of mismatches observed in the time axis. This paper

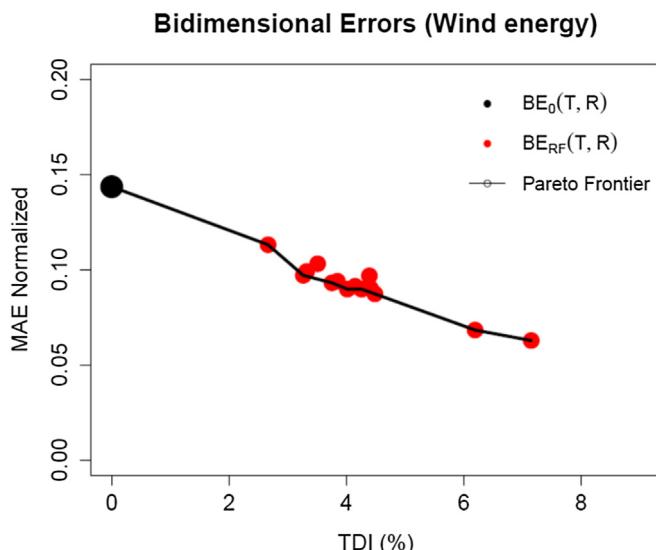


Fig. 27. Temporal Distortion Index obtained by different Step Pattern (Wind energy).

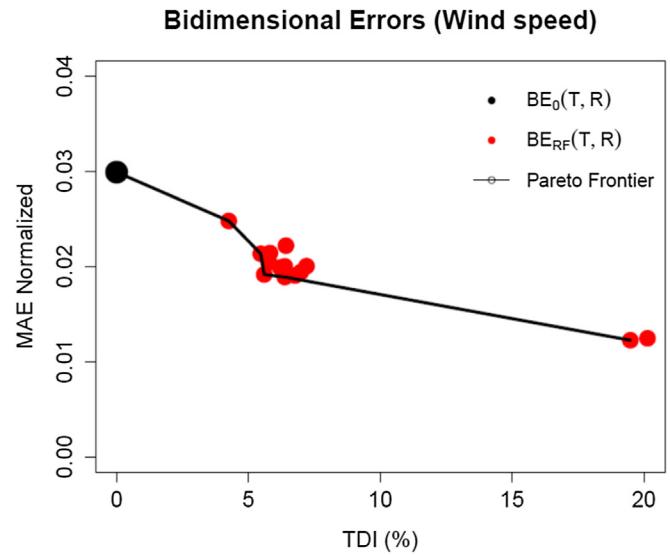


Fig. 28. Temporal Distortion Index obtained by different Step Pattern (Wind speed).

shows that the information on the lags in the identification of events and also on the difference between predicted and measured durations of events enable a deeper understanding of the prediction models. This paper proposes a comprehensive methodology to analyze and evaluate these temporary mismatches. A new index of temporal error, named Temporal Distortion Index, TDI, has been defined. It is used as first component of a two-dimensional error vector, BE, whose second component is a static error measurement, as the MAE. The information provided by this vector allows for a joint analysis of errors in time and scale. Broadly speaking, the algorithm performs transformations in the time axis of the predicted series in order that it closely resembles the real series. This process is based on Dynamic Time Warping techniques and uses recursion formulas which control the types of changes that can be made in the time axis. In addition, the availability of different formulas of recurrence leads to the use of Pareto Frontiers to assess the behavior of the prediction models. The goodness of the new methodology is shown by using both synthetic and real data that reflect the typical temporal effects in the prediction of renewable energy. The new temporal error index TDI is able to capture the temporal distortion when applied to laboratory examples but also when it is tested with real examples.

The methodology proposed here can be applied to compare different prediction models in both wind energy and solar radiation. The comparison can be done in terms of temporal distortion and in terms of scale distortion. When extended data are available, it could be possible to analyze what model outperforms others and in what circumstances. Comprehensive studies to characterize prediction errors of forecast models by using the new methodology should consider different energy production scenarios defined by the prediction horizon, topography, season, special wind regimes (or cloud formation in the case of solar energy), wind (solar) farm capacity, etc. In this way, it could emerge the dependency of the forecast method performance with those factors defining the energy production scenario. From a technical point of view, it is necessary a deeper analysis of the recurrence functions to obtain parametrized families of functions whose parameters allow the control of the temporal distortion magnitude. By varying such parameters it would be possible to create Pareto Frontiers more dense which would be used as characteristic error curves of the methods.

Acknowledgments

This paper has been supported under Grants MTM 2012–36025 and DPI 2013–47279-C2-1-R. The authors are grateful to the research staff of the National Renewable Energy Center of Spain (CENER) for their help in the development of this new methodology of analysis of errors contributing with their prediction model LocalPred.

References

- [1] E.W.E.A. European-Wind-Energy-Association. Wind energy—the facts: a guide to the technology, economics and future of wind power. Routledge; 2012.
- [2] Kaldellis JK, Zafirakis D. The wind energy (r) evolution: a short review of a long history. *Renew Energy* 2011;36(7):1887–901.
- [3] Tina G, De Foire S, Ventura C. Analysis of forecast errors for irradiance on the horizontal plane. *Energy Convers Manag* 2012;64:533–40.
- [4] Jolliffe IT, Stephenson DB. Forecast verification: a practitioner's guide in atmospheric science. John Wiley & Sons; 2012.
- [5] Madsem H, Karinotakis G, Nielsen G, Aalborg H, Nielsen T, Pinson P. A protocol for standardizing the performance evaluation of short-term wind power prediction models. Lyngby, Denmark: Technical University of Denmark, IMM; 2004. Deliverable ENK5-CT-2002-00665.
- [6] Madsen H, Karinotakis G, Nielsen G, Nielsen T, Pinson P, Aalborg H. Standardizing the performance evaluation of shortterm wind power prediction models. *Wind Eng* 2005;29(6):475–89.
- [7] Karinotakis G, Martí I, Casas D, Pinson P, Nielsen TS, Madsen H, et al. What performance can be expected by short-term wind power prediction models depending on site characteristics. In: Proceedings of the European Wind Energy Conference, Vol. 2004.. London: EWEC; 2004.
- [8] Martí I, Karinotakis G, Pinson P, Sánchez I, Nielsen T, Madsen H, et al. Evaluation of advanced wind power forecasting models—Results of the Anemos project. In: Proceedings of the European Wind Energy Conference. Greece, Athens: EWEC Athenes; 2006.
- [9] Murphy A, Winkler R. A general framework for forecast verification. *Monthly Weather Rev* 1985;115:1330–8.
- [10] Murphy A, Winkler RL. What is a good forecast? an essay on the nature of goodness in weather forecasting. *Weather Forecast* 1993;8:281–93.
- [11] Pinson P. Estimation of the uncertainty in wind power forecasting. Ecole des Mines de Paris. Centre for Energy and Proceses; 2006.
- [12] Bielecki MF, Kemper JJ, Acker TL. A methodology for comprehensive characterization of errors in wind power forecasting. In: ASME 2010 4th International Conference on Energy Sustainability; 2010.
- [13] Bielecki MF. Statistical characterization of errors in wind power forecasting. Northern Arizona University; 2010.
- [14] Hodge B, Milligan M. Wind power forecasting error distributions over multiple timescales. In: Power and Energy Society General Meeting, 2011. IEEE; 2011.
- [15] Hodge B-M, Lew D, Milligan M, Holttinen H, Sillanpaa S, Gómez-Lázaro E, et al. Wind power forecasting error distributions: an international comparison. 2012.
- [16] Castronovo ED, Peas Lopes J. On the optimization of the daily operation of a wind-hydro power plant. *Power Syst IEEE Trans* 2004;19(3):1599–606.
- [17] Methaprayoon K, Yingvivatanapong C, Lee W-J, Liao JR. An integration of ANN wind power estimation into unit commitment considering the forecasting uncertainty. *Industry Appl IEEE Trans* 2007;43(6):1441–8.
- [18] Dietrich K, Latorre J, Olmos L, Ramos A, Perez-Arriaga I. Stochastic unit commitment considering uncertain wind production in an isolated system. In: 4th Conference on Energy Economics and Technology; 2009. Dresden, Germany.
- [19] Bludszuweit H, Domínguez-Navarro JA, Llombart A. "Statistical analysis of wind power forecast error. *Power Syst IEEE Trans* 2008;23(3):983–91.
- [20] Zhang J, Hodge B-M, Florita A. Investigating the correlation between wind and solar power forecast errors in the western interconnection. In: ASME 2013 7th International Conference on Energy Sustainability collocated with the ASME 2013 Heat Transfer Summer Conference and the ASME 2013 11th International Conference on Fuel Cell Science, Engineering and Technology; 2013.
- [21] Martí I, Usaola J, Sánchez I, Lobo M, Navarro J, Roldán A, et al. Wind power prediction in complex terrain. LocalPred and SIPREÓLICO. In: Proceedings of the European Wind Energy Conference. Spain: EWEC Madrid; 2003.
- [22] Martí I, San Isidro M, Cabezón D, Loureiro Y, Villanueva J, Cantero E, et al. Wind power prediction in complex terrain: from the synoptic scale to the local scale. In: Proceedings of the science of making torque from wind; 2004. p. 326–7. Delft (NL).
- [23] Foley AM, Leahy PG, Marvuglia A, McKeogh EJ. Current methods and advances in forecasting of wind power generation. *Renew Energy* 2012;37(1):1–8.
- [24] Gilletland E, Roux G. A new approach to testing forecast predictive accuracy. *Meteorol Appl* 17 October 2014;22(3):534–43.
- [25] Diebold FX, Mariano RS. Comparing predictive accuracy. *J Bus Econ Statistics* 1995;13:253–63.
- [26] Sakoe H, Chiba S. A dynamic programming approach to continuous speech recognition. In: Proceedings of the Seventh International Congress on Acoustics; 1971.
- [27] Sakoe H, Chiba S. Dynamic programming algorithm optimization for spoken word recognition. *Acoust Speech Signal Process IEEE Trans* 1978;26(1):43–9.
- [28] Azcárate C, Blanco R, Mallor F, Garde R, Aguado M. Peaking strategies for the management of wind-H₂ energy systems. *Renew Energy* 2012;47:103–11.
- [29] Berndt DJ, Clifford J. Using dynamic time warping to find patterns in time series. In: KDD workshop, Seattle; 1994.
- [30] Kassidas A, MacGregor JF, Taylor PA. Synchronization of batch trajectories using dynamic time warping. *AIChE J* 1998;44(4):864–75.
- [31] Tomasi G, Van den Berg F, Andersson C. Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *J Chemom* 2004;18(5):231–41.
- [32] Gavrila D, Davis L. Towards 3-d model-based tracking and recognition of human movement: a multi-view approach. In: International workshop on automatic face-and gesture-recognition; 1995. p. 272–7. Citeseer.
- [33] Zhang Z, Huang K, Tan T. Comparison of similarity measures for trajectory clustering in outdoor surveillance scenes. *Pattern Recognition*, 2006. ICPR 2006, 18th International Conference on; 2006.
- [34] Caiani E, Porta A, Baselli G, Turiel M, Muzzupappa S, Pieruzzi F, et al. Warped-average template technique to track on a cycle-by-cycle basis the cardiac filling phases on left ventricular volume. *Comput Cardiol* 1998;1998.
- [35] Aach J, Church GM. Aligning gene expression time series with time warping algorithms. *Bioinformatics* 2001;17(6):495–508.
- [36] Bellman R. Dynamic programming. 1st edn. , Princeton: Princeton University Press; 1957.
- [37] Rabiner L, Juang BH. Fundamentals of speech recognition. Prentice Hall; 1993.
- [38] Miettinen K. Nonlinear multiobjective optimization, vol. 12. Springer Science & Business Media; 1999.