

Improving DTR assessment by means of PCA applied to wind data

Alessandro Bosisio^{a,*}, Alberto Berizzi^a, Dinh-Duong Le^b, Fabio Bassi^c, Giorgio Giannuzzi^c

^a*Politecnico di Milano, Department of Energy, Via la Masa 34, 20156 Milano, Italy*

^b*Faculty of Electrical Engineering, The University of Danang-University of Science and Technology, Danang, Vietnam*

^c*TERNA S.p.a., Via dei della Robbia 41, 50132 Firenze, Italy*

Abstract

Traditionally, the rating of an overhead transmission line is determined under a set of specified and standardized conditions. However, weather conditions along the line change during operation. Therefore, the standard rating of the line might be either underestimated, leading to inefficient utilization of the line, or overestimated, leading to unsecure operation. This is the major drawback of the traditional approach: the so-called Dynamic Thermal Rating (DTR), that takes into account the actual operating conditions along the line to determine the rating, is today a critical need. In this paper, we develop a comprehensive methodology for exploring all necessary information about stochastic processes of environmental variables surrounding and along the line using available data. The results can be used as input to determine the actual rating of the considered transmission line to enhance the determination of the rating for transmission lines.

Keywords: Dynamic Thermal Rating; Principal Component Analysis; Forecast technique; Transmission line; Power system management

*Corresponding author

Email address: alessandro.bosisio@polimi.it (Alessandro Bosisio)

1. Introduction and motivation

The rating of an overhead transmission line is the maximum current that the line is allowed to carry continuously without exceeding the maximum temperature of the conductor. Thus, temperature affects sags between towers and tensile strength; it depends on the magnitude of current through the line, its time duration, and on environmental conditions, such as wind speed, air temperature, solar radiation, and so on [1, 2].

Traditionally, the rating is determined under a set of standard conditions [1, 2, 3]. However, the actual operating conditions, especially weather conditions along the line, vary during operation. Therefore, the rating of the line might be either underestimated, leading to inefficient utilization of the line, or overestimated, leading to unsecure operation. This is why the so-called Dynamic Thermal Rating (DTR), that takes into account the actual operating conditions to determine the rating, is considered today a critical need.

DTR bring us several benefits [4]. It allows the use of existing transmission assets efficiently, thus improving economic benefits, allows deferral of new transmission lines, it makes it possible to manage more easily contingency conditions. The better exploitation of available infrastructure is very useful in the presence of non programmable Renewable Energy Sources (typically, wind power), that often cause congestions in the system and wind power curtailment. DTR has been much cheaper than storage in solving this kind of problems in Italy. That is why, in the present paper, DTR is considered with reference to windy areas, where it can be adopted most profitably.

In order to assess the dynamic rating of the line and to schedule for the best operation for the next 5-15 minutes, necessary information could be provided by Weather Forecast Providers (WFPs); however, the forecast required for an efficient power system operation are not available from WFPs, especially for mountainous areas, where conditions might change very significantly and quickly and for each span of the transmission line. For this reason, Terna investment plans include the installation of several sensors along many electrical lines to

integrate information coming from WFPs with data coming in real-time by its own sensors.

Environmental variables such as wind, temperature, solar radiation, and so on can be considered as random variables that cover a wide geographic area and evolve over time, determining the stochastic process. Hence, in order to obtain future information about it, a suitable space-time modeling method which can explore and capture all necessary information in the processes needs to be developed.

Future information on stochastic processes can be provided by forecast techniques such as point forecast and probabilistic forecast [5]. However, such techniques do not capture temporal correlation between forecast errors (i.e., uncertainty) of different time-steps. In a different way, scenario forecasts [6] can capture such information, especially temporal correlation, and give forecasting results in the form of a set of discrete scenarios. So far, several techniques have been used for scenario generation. Reference [7] discusses on using various techniques including conditional sampling, property-matching methods, optimal discretization, and step-wise growing and cutting methods for generating scenarios for stochastic programming.

Scenario forecasts have been applied to several research areas in practice. In [8], the authors generate arrival capacity scenarios for each airport in United States based on non-parametric methods (in [9]) using only historical data. Clustering techniques are also used in that study, so that the scenarios generated are the averages of clusters of similar profiles.

The technique for scenario construction using both historical data and forecasted data is discussed in [10]: it shows a better performance compared to [8]. In another field, reference [11] uses the daily pattern of change from Global Climate Models to generate rainfall scenarios. However, those methods are only suitable for generating scenarios for a single stochastic process (e.g., wind speed process at a single location).

In order to produce scenario forecasts for multiple stochastic processes which are cross-correlated, e.g., wind speed processes at multiple locations, space-time

modeling techniques are needed including non-stationarity, non-Gaussianity, temporal and spatial correlation.

A careful analysis on the existing methods in this field is provided in [12],
65 where the Authors use orthogonal Markov chain; however, in the paper the
Authors assume spatio-temporal covariance stationarity, i.e., that the auto-
covariance function of each time series at each location and the spatial cor-
relations among time series of different locations do not change.

In the field of wind power studies for generating space-time wind scenarios,
70 in order to characterize interdependence structure of multivariate stochastic
processes, Gaussian copula method [13] is widely used. In [14], the Authors
build the model for multi-site wind speed using a noise vector deriving a vec-
tor auto-regressive process. This model accounts for both spatial and temporal
correlations of wind speed and assumes joint Gaussian distribution and station-
75 arity. In [14] and [15], the problem is simplified by the assumption that the
matrix of auto-regressive coefficients is diagonal, which implies that the multi-
variate time series model can be decoupled into different univariate time series
model for each wind site.

This paper presents a new approach based on a space-time model that, using
80 both weather historical data and real-time measurement data with different
temporal resolutions, provides good forecasting for DTR of transmission lines.
Specifically, the space-time model uses the analysis of current wind situation
provided by a WPF (available every 12 hours with space resolution of 2km x
3km) integrated with real-time measurements available every 5 minutes at a
85 significant number of different specific locations.

The paper is organised as follows: Section 2 motivates the proposed method-
ology. In Section 3 the Space-Time Model is presented, while Section 4 describes
the real-time operation of the proposed procedure. Section 5 reports the results
of the procedure developed, comparing wind forecasting data and real measure-
90 ment data to check the accuracy of the methodology. Concluding remarks are
given in Section 6.

2. Integration of weather data for the DTR of transmission line

In Italy, windy areas show a high penetration of wind power generation that, in windy conditions, are currently limited by the presence of transmission congestions. However, depending on the wind speed and direction, wind can itself contribute to increase the actual rating of the line, thus alleviating congestions. The goal of the present paper is to estimate the wind speed (magnitude and direction) along the spans of an electrical line: the estimated values are then transferred to a line rating model in order to complete the DTR of the line and possibly avoid wind generation curtailment.

The proposed approach aims at integrating the available information sources, which are very different in terms of time and space resolution:

- Weather Forecast Provider (WFP) information: they are high level weather data analysis based on a mathematical-physical model computed according to a very refined spatial mesh for the whole continent. Data are available every 12 hours (at 00:00 and 12:00 every day) for many layers in height;
- Real-time (RT) weather stations data: the Italian TSO, Terna, has been installing for some years many sensors that can measure weather data (temperature, wind, etc.) and transfer such information to the control centre. Typically, such devices are installed on top of quite a few towers along some transmission lines (TL). They provide real-time measurements at their location only at a 5 minute rate; of course, they are scattered on the territory, not on a regular basis, unlike WFP information.

The problem is how to integrate such two set of data: WFP data, given for a very large area with a refined mesh, available with a very poor time resolution, with the RT measurements immediately available for scattered locations along some electrical lines with very good time resolution. The goal is to make it possible to determine reliably the wind conditions along a chosen electrical line for the next 5 minutes to 2 hours, in order to make it possible to keep the rating

of the lines updated and redispach in real-time, if possible, some power that should be curtailed if the DTR were not available.

2.1. WFP data

WFP data are supposed to provide a general picture of how the wind streams
125 are moving in a large scale. As shown in Fig. 1a, where an area in the Centre-
South of Italy is shown, weather data are provided for a selected rectangular
area with a resolution of 2 km for the longitude and 3 km for the latitude.
Data on all the points of the mesh, marked with yellow triangles, are computed
by the WFP in two snapshots are at 00:00 and 12:00 UTC. Hence, WFP data
130 have a high spatial resolution but a poor temporal resolution, and are related
to a mathematical-physical model, not to measured data. Therefore, they are
useful to extract a time-space model, but not to be used immediately for the
DTR. In the proposed approach, the WFP data are used to create and char-
acterize a database of past long-term scenarios, to be used to predict future
135 long-term scenarios. A scenario derived by WFP data includes weather features
at every point in the mesh, with time resolution equal to 12 hours. It should
be considered as useful to learn weather behavior on a larger-scale area. Such
characteristics will be integrated with data coming from RT stations.

As it is clear, WFP dataset is very huge. In order to reduce its size, the
140 following criteria are adopted:

- for WFP mesh points surrounding the considered TL, the highest available resolution (3 km x 2 km) is kept;
- for WFP mesh points far from the transmission line, the resolution is decreased (6 km x 4 km);
- 145 • at any rate, the four closest WFP mesh points around each RT stations are always kept, independent of their distance from the TL, with the goal of identifying and exploiting as much as possible the correlation between WFP and RT data.

In Fig. 1a, areas with a higher space resolution can be observed around both
 150 RTs and the considered TL.

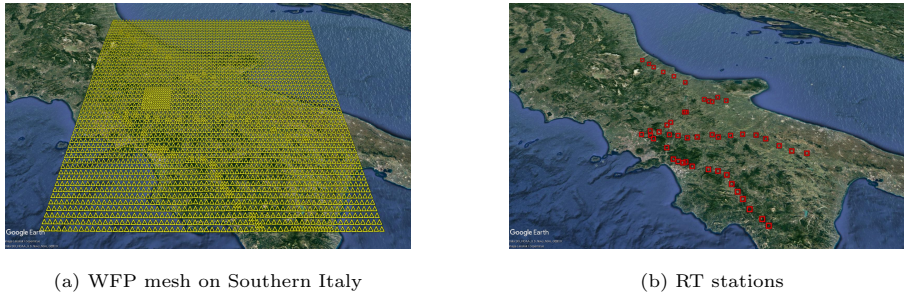


Figure 1: WFP mesh and RT station locations

The vector of wind speed data can be expressed in terms of orthogonal speed
 components, where:

- u is the Zonal speed, the component of the horizontal wind towards East;
- v is the Meridional speed, the component of the horizontal wind towards
 155 North.

Wind data are given for each point of the mesh and for 6 different heights
 above the ground (layers 10m, 30m, 50m, 70m, 100m). Based of the specific TL
 architecture, suitable height is considered.

WFP data are given by the WFP in GRIB format (GRIdded Binary) a data
 160 format standardized by the World Meteorological Organization's Commission.

2.2. Real-time data

Real-time (RT) measurements recording is carried out, with a temporal res-
 olution of 5 minutes, at 43 weather stations installed on a TL tower or in a
 substation. Geographical locations of the RT measurement stations considered
 165 in this paper, marked with red squares, are shown in Fig. 1b.

RT data including, for each sensor, time-stamped wind magnitude and di-
 rection, are transferred every 5 minutes to the TSO control centre. Like any
 real-time data flow, RT data can be affected by missing or erroneous values

in the RT dataset. This issue must be taken into account in the time-space
170 modelling.

2.3. TL considered for DTR

In order to select data for the time-space model, it is necessary to define
the TL subject to DTR. Fig. 2 shows the TL located in the Southern Italy
considered for DTR with its 124 towers. This TL is also equipped with 4 RT
175 stations marked with red squares in Fig. 2. TL has average height above ground
of 25 m (height of the lowest conductor and sensors, when installed).

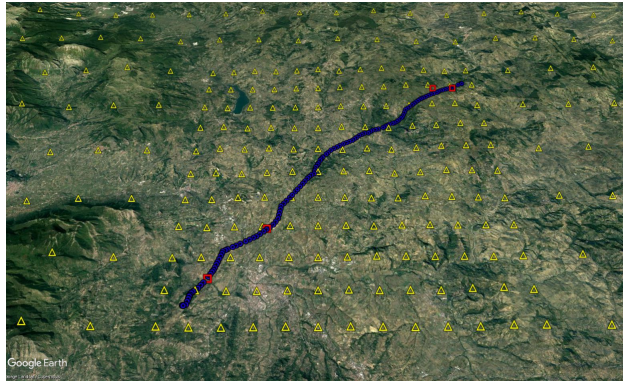


Figure 2: TL path, in the WFP mesh

3. Space-Time Model

The core of the forecasting approach to enhance DTR assessment is an ex-
tension of the space-time modeling presented in [16]. It combines Principal
180 Component Analysis (PCA) with clustering/selection, extrapolation, interpola-
tion techniques. It can capture all salient features of wind speed from multiple
locations: non-Gaussianity, non-stationarity with temporal and spatial correla-
tions thanks to PCA, which also provides an excellent tool for approximating a
large data set by reducing its size.

185 *3.1. Principal Component Analysis*

PCA is an orthogonal transformation of a generic data set into an set of uncorrelated data [17, 18]. It is based on the eigenanalysis of the covariance matrix (or the correlation matrix) of a set of data, that determines a transformation matrix such that transformed data \mathbf{Z} show diagonal covariance matrix. The new variables are called Principal Components (PCs) and the technique is used to reduce as much as possible redundancy in the data set, while keeping only the first PCs.

Wind data are stored in a matrix \mathbf{W} whose elements w_s^h are the wind speed at site $s \in \{1, 2, \dots, S\}$ and at time $h \in \{1, 2, \dots, N\}$. We assume that all considered wind sites have the same number of observations N and that they are synchronized and equally spaced in time.

$$\mathbf{W} = \begin{bmatrix} w_1^1 & w_1^2 & \cdots & w_1^N \\ w_2^1 & w_2^2 & \cdots & w_2^N \\ \vdots & \vdots & \ddots & \vdots \\ w_S^1 & w_S^2 & \cdots & w_S^N \end{bmatrix} \quad (1)$$

At first, data are centered [17] in matrix \mathbf{W}_c by subtracting the mean μ_s of each time series at each site. In the following, the covariance matrix $\mathbf{\Sigma}$ is considered, for the sake of simplicity [17], whose elements are, in the diagonal, the variance σ_i^2 and, off-diagonal, the covariance between the time series at site i and the time series at site j $\sigma_{i,j}^2$.

Its eigenvalues λ_i , $i = 1, 2, \dots, S$, real positive, and the associated eigenvectors \mathbf{u}_i are considered and form the matrix \mathbf{U} :

$$\mathbf{U} = \left[\mathbf{u}_1 \mid \mathbf{u}_2 \mid \cdots \mid \mathbf{u}_S \right] \quad (2)$$

Its elements are also known as PC *coefficients* or *loadings*. Finally, PCs are derived [17] as

$$\mathbf{Z} = \mathbf{U}^T \mathbf{W}_c \quad (3)$$

where \mathbf{Z} is a $S \times N$ matrix. The i -th row of matrix \mathbf{Z} , \mathbf{z}_i , is the i -th PC, that is, a time series univariate and uncorrelated with other PCs [19].

The reconstruction of wind data from PCs is implemented inversely:

$$\mathbf{W} = \boldsymbol{\mu} + \mathbf{UZ} = \boldsymbol{\mu} + \begin{bmatrix} \mathbf{u}_1 & | & \mathbf{u}_2 & | & \cdots & | & \mathbf{u}_S \end{bmatrix} \cdot \begin{bmatrix} \mathbf{z}_1^T & | & \mathbf{z}_2^T & | & \cdots & | & \mathbf{z}_S^T \end{bmatrix}^T \quad (4)$$

It is worth noting that if the distribution considered is multivariate Gaussian, resulting PCs will be independent. Otherwise, PCs will be uncorrelated but still dependent (the diagonal covariance matrix of PCs only implies that they are uncorrelated). In the present paper, we adopt pre-processing and transformation techniques to obtain approximately stationary and Gaussian data sets to improve PCA.

PCA is an excellent tool to approximate a large data set by reducing its dimension [18]. This function makes PCA a powerful tool for high-dimensional data analysis. Each PC \mathbf{z}_l holds an amount of information which is represented by its variance, i.e., by λ_l . Therefore, the contribution of the l -th PC to total variance of the data [18] can be computed as:

$$\gamma_l = \frac{\lambda_l}{\sum_{i=1}^S \lambda_i} \times 100\% \quad (5)$$

Hence, the first row vector \mathbf{z}_1 , corresponding to the largest eigenvalue λ_1 , is the most dominant component, which contains most of the variance in the data set, followed by the second component \mathbf{z}_2 , and so on. This makes it possible to approximate the initial data set by a few PCs, based on the total amount of information. This makes it possible to describe most of the features of the data set by a reduced number of variables. PCA is used in this paper to reduce the number of variables without losing important information and to carry out time series analysis and scenario generation on a few PCs instead of trying to do the same working on longer data sets.

3.2. Time series analysis

A time series is a sequence of observations ordered in time, usually at equally-spaced intervals. There are several methods for fitting to a time series, if it is stationary. For a stationary stochastic process, the joint probability distribution do not change over time. The typical linear model for a stationary time series is ARMA [19], which can be used to characterize a stationary process and for prediction as well. To build a time series model, we follow the procedure proposed by Box-Jenkins.

It should be noted that stationarity is a necessary condition in building an ARMA model. However, this condition may not always hold with real time series data. In such a case, data must be pre-processed. In this paper, we carry out various pre-processing and transformation techniques. At first, input data (usually non-stationary and non-Gaussian) are pre-processed and transformed to obtain stationary and Gaussian data as required. The input for space-time model is matrix \mathbf{W} ; all time series must span the same time horizon with equal intervals (12 hours for WFP mesh and 5 minutes for RT measurements). After that, PCA is carried out and a few (depending on their variance) PCs are fitted by a time series model (to include variability content), and used to generate an adequate number of time series for future time. The generated time series in terms of PCs are back transformed into non-Gaussian data [14, 15] to obtain scenarios obeying all the characteristics of the observed wind data for each site.

Hence, the space-time modeling method can capture most of the main characteristics of the input data and gives the output results in terms of *scenarios of future values* (with a certain future time frame, e.g., 12 hours ahead) for each input quantity with the same temporal resolution as the input data. The novelty of the proposed approach is that it can explicitly capture the main features of stochastic processes of multi-site wind data: marginal distribution, spatial correlation, temporal correlation, diurnal and seasonal non-stationarity and non-Gaussianity.

It is also worth noting that PCA allows a significant reduction of the size of the input data set without losing significant information. This feature is very

useful because the size of data set used in this research is very large, due to a
260 large number of input data.

An important issue is how to deal with missing or erroneous input data. The procedure set up includes the possibility to adopt Probabilistic PCA (PPCA) [20]. However, PPCA needs much more computation time than PCA, especially for a large input data set. In such case, PCA combined with an effective
265 technique for missing data and outliers treatment is preferred.

Based on the distribution curve of historical wind speed data and once set up a significance level considered (5%, 2.5%, 1%, etc.), the procedure detects gross errors, thus filtering wind speed data [21]. Values that deviate significantly are considered as outliers. To avoid the risk of eliminating extreme but possible
270 conditions, the only way is choosing suitable threshold for significant levels (in this paper, significance level of 0.5% is used).

A particular case of gross errors is when values remain constant or nearly constant (for long time). This can occur when an RT station measurement device is out of service. This issue needs to be detected and the relevant RT
275 station data should not be considered.

Missing data can be long consecutive or randomly distributed. Randomly missing values is well fixed by both missing data procedure and PPCA. A different story is for long consecutive missing data. In case long consecutive data are missing, the RT station or the WFP mesh data are not used at all, because
280 it is too difficult to correct data in such case.

4. The proposed methodology

As mentioned, the most challenging issue faced is the integration of data coming from different sources, to identify the most probable values of wind speed along the TL to be used for the DTR. In this Section, we present a
285 comprehensive framework proposed for exploiting available wind data along a TL to provide useful information for improving DTR assessment. The goal is to get, for each point along the electrical line, reliable forecast of the wind (speed

and direction) for the next couple of hours with a time resolution of 5 minutes. This is obtained by generating a number of scenarios with different space and time resolution making use of the PCA approach. The whole structure of the procedure implementation is depicted in Fig. 3.

Two main databases have to be created, with different time resolution:

- DB1 contains wind data with 12 hours time resolution, coming from both WFP model, relevant to each point considered in the mesh, and additional data synchronized in time with WSP data (i.e., every 12 hours), coming from all RT sensors available. PCA working on data from DB1 generates 12-hour resolution scenarios.
- DB2 contains data coming every 5 minute from the RT measurements. They are used to generate scenarios with 5-minute time resolution at RT locations using PCA.

Both databases are created initially by InputWFP and InputRT and updated by blocks Update WFP and UpdateRT as soon as new data are available, i.e., either every 12 hours or every 5 minutes, respectively. Actually, the same procedure is carried out for two components, namely u and v , of wind separately.

The generation of 12-hour scenarios is carried out within Block1, every 12 hours, when new analysis data from the WFP are available in DB1. It includes the Space-Time Modeling (STM1) presented in Section 3 which aims at capturing all main features of the wind stochastic processes at points on both the WFP mesh and the RT stations and then at generating scenario forecasts at all these points, but in particular around the TL under DTR, over a predefined future time frame (for example 12 or 24 hours) and with time resolution of 12 hours. In Block1, due to high computation times that would be needed by the PPCA (PSTM1), the choice has been made for an efficient technique for missing data and outliers treatment combined with PCA. Moreover, since WFP data come from a model and it is assumed to be reliable, this issue is less significant than for the RT data treatment. N1 scenarios are generated by the space-time model and stored in database SDB1.

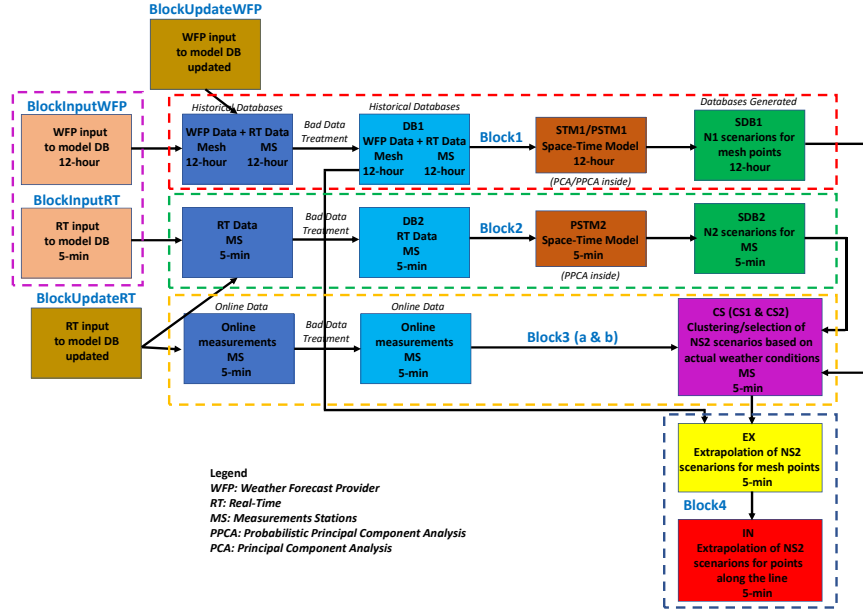


Figure 3: Block diagram of the proposed methodology

In the same way, in Block2, the same is done by STM2 for RT data in DB2, in order to generate scenarios to be stored in SDB2. Data are those relevant to RT measurement locations, with 5-minute time resolution. This block is run every hour, when a set a new data from the RT measurements is newly available and generates N2 scenarios over a predefined future time frame of 2 hours with time resolution of 5 minutes. Since data come from real-time measurements, thus being affected by missing values and outliers, and since the RT stations are fewer than mesh nodes, the procedure uses PPCA (PSTM2) to deal with missing or bad data.

The goal of Block3 is to pick up a few scenarios closest to the actual weather conditions. This is done by a clustering/selection procedure (i.e., block *CS* in Fig. 3). Block3 is in turn divided into two blocks: Block3a, which works with *SDB1* and, using *CS1*, selects *NS1* scenarios among *N1* 12-hour scenarios (it

is run every 12 hours), and Block3b, which works with $SDB2$ and using $CS2$ selects $NS2$ scenarios among $N2$ 5-minute scenarios (it is run every hour). In fact, during the real-time operation, online measurement data at RT stations (*online data* in Fig. 3) are collected. These data are combined with $SDB1$ and $SDB2$ to select $NS2$ most probable scenarios (5-min resolution) for each quantity at RT stations.

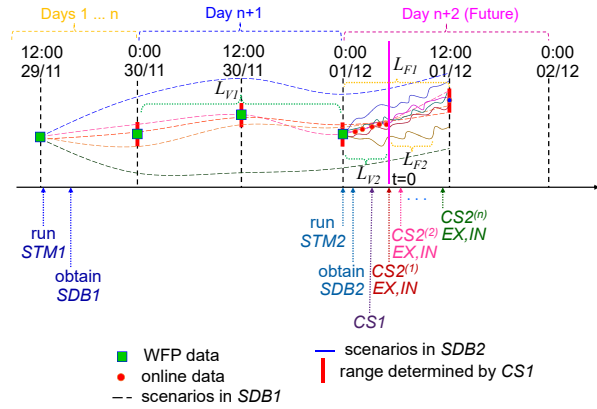


Figure 4: Example of choosing suitable time instances to start running each block

For the sake of clarity, Fig. 4 describes an example of choosing suitable time instances to start running each block. The schedule is made based on both the running time needed to perform each block and the availability of updated input data. Focusing on block CS of Fig. 3, two stages of selection (corresponding to two sub-blocks $CS1$ and $CS2$) are needed:

- $CS1$: selection of $NS1$. In the example of Fig. 4, $STM1$ uses data collected up to 29/11 to generate $N1$ scenarios for four steps ahead (i.e., 0:00 and 12:00 on 30/11, 0:00 and 12:00 on 01/12). With $CS1$, the procedure selects $NS1$ scenarios whose *paths* are most similar to WFP data at 0:00 and 12:00 on 30/11 and 0:00 on 01/12 (i.e., latest known data used as validation set - $L_{V1} = 3$). Once $NS1$ scenarios are identified, they also provide information on most likely range of values for each variable at

12:00 01/12 (unknown/forecasted data, range in red in Fig. 4);

- *CS2*: selection of *NS2* scenarios among *N2* scenarios in *SDB2*. Following the same idea of *CS1*, *CS2* selects *NS2* scenarios which fulfill both following constraints: their *paths* are most similar to the latest and current RT measurements, and their values at 12:00 on 01/12 are within the range determined by *CS1*. Based on the need of forecast for the next 2 hours, *CS2* is implemented in real-time: every hour, when a set of new twelve 5-minute data at RT measurement stations are available, *CS2* is run to select the best *NS2* scenarios in *SDB2* for the next 2 hours (2 hours intervals correspond to $L_{F2} = 24$ time-steps of 5 minutes in the future). In Fig. 4, L_{F2} and L_{V2} are number of forecasted values and number of values in validation set. The process is repeated (called $CS2^{(2)}, \dots$) until the next *PSTM2* is carried out.

To select the best *NS2* scenarios (time series) among *N2* scenarios in *SDB2*, the similarity between each time series i in *N2* $\{w_i^k\}$ and time series of the validation set $\{y_v^k\}$ is computed:

$$D_{w_i, y_v} = \frac{1}{K} \sqrt{\sum_{k=1}^K (w_i^k - y_v^k)^2} \quad (6)$$

where:

$\{y_v^k\}$ ($k = 1, 2, \dots, K$) includes L_{V2} latest online measurement data and one point determined by *CS1* (i.e., $K = L_{V2} + 1$);

$\{w_i^k\}$ ($k = 1, 2, \dots, K$; $i = 1, 2, \dots, N2$) is i in *SDB2* where $\{w_i^k\}$ is synchronized with $\{y_v^k\}$.

NS2 scenarios are selected, corresponding to the *NS2* smallest values of D_{w_i, y_v} .

Similarly, as shown in Fig. 4, for *CS1* we considers $K = L_{V1} = 3$ (latest three set of data in *DB1*).

At this point we have, for each RT station, a 5-minute resolution forecast included in the *NS2* scenarios selected. The last step is then to move to 5-

minute resolution scenarios at points along the line; this is done in Block 4 which includes Extrapolation and Interpolation.

Extrapolation aims at computing, on the grounds of both the current RT measurements and *NS1* and *NS2* scenarios selected, 5-minute forecasted scenarios at points (called *query point*), on the mesh, surrounding the TL. First the relationship between data associated with the wind at this query point and the wind at RT stations needs to be characterized; then, the obtained relationship makes it possible to determine data at any query point for future instances from forecasts at RT locations included in *NS2* scenarios. The method of bins [16, 22] is used; the range of data of the wind speed at a known point (RT stations) is divided into different bins and, for each bin, the average values at known point and query point are computed. Consequently, after obtaining relationship between data at known point and query point, if any value at known point is newly available, the corresponding value at query points can be deduced accordingly.

Finally, Interpolation is used to compute wind speed trajectories over time with 5-minute resolution along the line, which is the final goal for DTR. For obtaining future data for points along the line from future data of surrounding points on the mesh, spatial interpolation techniques are applied. Two main groups of spatial interpolation can be used: deterministic and geostatistical techniques [23]. While deterministic interpolation is based on creating surfaces from surrounding measured points (e.g., a popular method in this group is Inverse Distance Weighted-IDW), geostatistical interpolation is based on the statistical properties of the measured points. For this work, IDW has been used to obtain future data for points along the line from forecasted data of surrounding points on the mesh, which have been already obtained as output of block *EX*.

5. Computational experiments and results

In this section, some selected results related to the space-time model applied to the TL shown in Fig. 2 are reported. As WFP analysis, we considered wind

405 speed data (12-hour resolution) at the 3325 points on the mesh shown in Fig. 1a for the month of November 2016. Regarding RT measurements, we considered wind speed data at the 43 weather stations shown in Fig. 1b.

To check the accuracy of the model forecast, we took out from the input dataset the last day (i.e., 2 samples at 00:00 and 12:00 for each point of the mesh and 288 samples - 12 samples/h x 24h - for each RT) comparing estimated and measured values. In Fig. 5, estimated and measured values (2 hours ahead, 24 x 5 minutes) for the RT station 4 are shown. The pictures depict the 20 best scenarios identified (light curves) and the final estimation carried out, as the mean values of the 20 best scenarios, in dashed blue line. For this RT station, 415 both estimated values (i.e., u and v components of wind) are close to measured data.

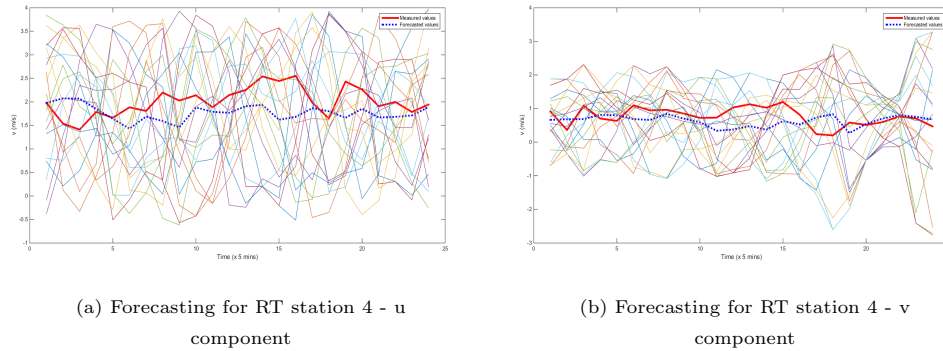
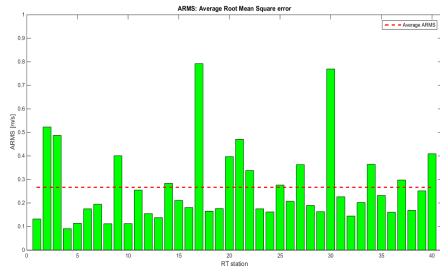


Figure 5: Output of the forecasting process for RTs

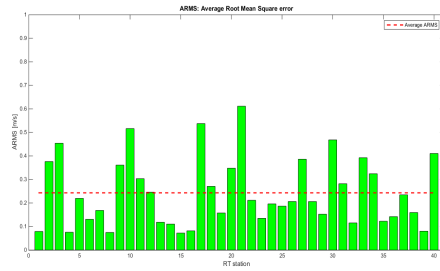
To access quantitatively the capabilities of the method, the Average Root Mean Square error (ARMS) and Mean absolute error (MAE) are calculated, for all the RT stations, as the distance between actual and estimated data.

420 As shown in Fig. 6 and 7, the average ARMS errors is around 0.3 m/s while the average MAE is around 1 m/s.

An example of the output of the methodology is shown in Fig. 8. It depicts the 20 best scenarios interpolated from the 4 closest WFP mesh points surrounding TL tower 1 and, again in dashed blue line, the mean values of those

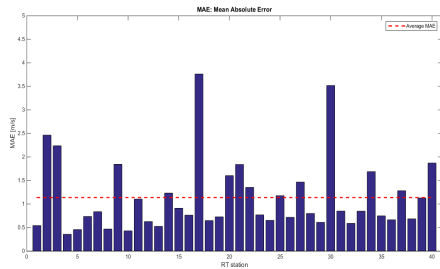


(a) ARMS - u component

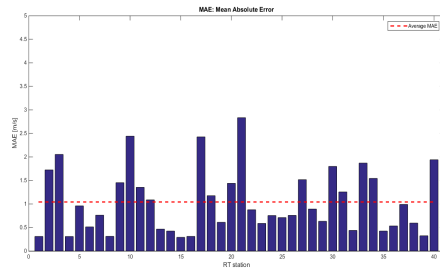


(b) ARMS - v component

Figure 6: RT stations ARMS



(a) MAE - u component



(b) MAE - v component

Figure 7: RT stations MAE

425 scenarios (which is considered as the wind forecast for TL tower 1). The accuracy of the wind forecasting at tower 1 cannot be assessed because no measured data are available at tower 1. The methodology gives a similar wind forecasting for all the other 123 TL towers as well as for any point of the TL.

6. Concluding remarks

430 In this paper, a methodology for exploring main characteristics of stochastic processes (e.g., non-Gaussianity, non-stationarity with distinct diurnal and seasonal patterns, temporal and spatial correlations) of wind speed along a TL is presented. All available data (both historical and RT data from a WFP and RT stations) are used to provide as much information about wind speed as possible
 435 for a desired future time horizon for dynamic rating assessment. In particular,

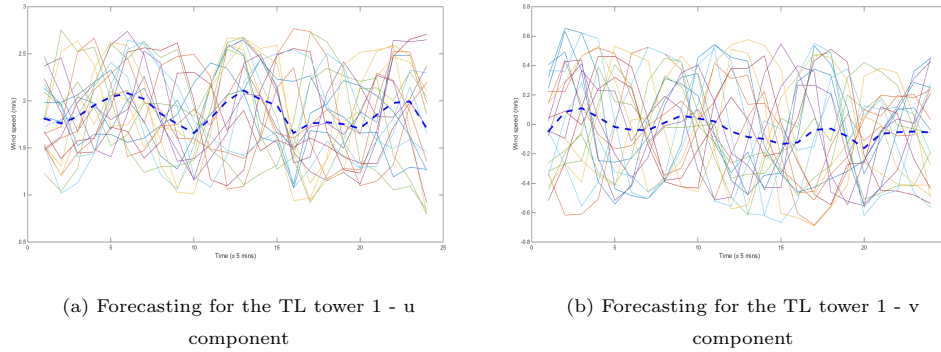


Figure 8: Online forecast for TL towers

no simplifying assumptions are used for building the proposed model and the proposed procedure is very powerful for dealing with a high-dimensional data set due to a large number of variables associated with numerous points on the mesh.

440 For practically testing the proposed methodology, real data related to a TL in the province of Benevento (Italy), the WFP mesh surrounding it and RT stations included in the selected mesh have been considered. Simulation results are presented to show that the methodology proposed can improve the wind speed forecast and consequently the DTR of TL.

445 References

- [1] IEC - International Electrotechnical Commission, Overhead electrical conductors - Calculation methods for stranded bare conductors, IEC, standard 61597 (1995).
- [2] IEEE Standard, IEEE Standard for Calculating the Current-Temperature of Bare Overhead Conductors, IEEE, standard 738 (2006).
450
- [3] S. D. Kim, M. M. Morcos, An application of dynamic thermal line rating control system to up-rate the ampacity of overhead transmission lines, IEEE Transactions on Power Delivery 28 (2) (Apr. 2013) 1231–1232.

- [4] F. Teng, R. Dupin, A. Michiorri, G. Kariniotakis, Y. Chen, G. Strbac,
455 Understanding the benefits of dynamic line rating under multiple sources
of uncertainty, *IEEE Trans. Power Syst.* 33 (3) (May 2018) 3306–3314.
- [5] A. Botterud, Z. Zhou, J. Wang, R. J. Bessa, H. Keko, J. Mendes, J. Sumaili,
V. Miranda, Use of wind power forecasting in operations decisions, Argonne
National Laboratory, Technical report (Sep. 2011).
- 460 [6] P. Pinson, H. Madsen, H. A. Nielsen, G. Papaefthymiou, B. Klockl, From
probabilistic forecaststo statistical scenarios of short-term wind power pro-
duction, *Wind Energy* 12 (1) (2009) 51–62.
- [7] M. Kaut, Scenario Generation for Stochastic Programming: A Practical In-
troduction, Tutorial on Scenario Generation, Håholmen (June 10-12, 2006).
- 465 [8] P. B. Liu, M. Hansen, A. Mukherjee, Scenario-based air traffic flow manage-
ment: From theory to practice, *Transportation Research Part B: Method-
ological* 42 (7) (2008) 685–702.
- [9] J. Dupacova, G. Consigli, S. W. Wallace, Scenarios for multistage stochastic
programs, *Annals of Operations Research* 100 (1) (2000) 25–53.
- 470 [10] G. Buxi, M. Hansen, Generating day-of-operation probabilistic capacity
scenarios from weather forecasts, *Transportation Research Part C: Emerg-
ing Technologies* 33 (2013) 153–166.
- [11] T. I. Harrold, R. N. Jones, Generation of rainfall scenarios using daily pat-
terns of change from gcms, *Water Resources Systems—Water Availabil-
ity and Global Change (Proceedings of symposium IUGG2003)*, Sapporo,
475 Japan (July 2003) 165–172.
- [12] G. A. Baigorria, J. W. Jones, Gist: A stochastic model for generating
spatially and temporally correlated daily rainfall data, *Journal of Climate*
23 (2010) 5990–6008.

- 480 [13] J. Tastu, P. Pinson, H. Madsen, Space-time scenarios of wind power generation produced using a Gaussian copula with parametrized precision matrix, Technical University of Denmark, Technical report (2013).
- [14] J. M. Morales, R. Minguez, A. J. Conejo, A methodology to generate statistically dependent wind speed scenarios, *Applied Energy* 87 (2010) 843–855.
- 485 [15] A. Papavasiliou, S. S. Oren, Stochastic modeling of multi-area wind power production, 12th International Conference on Probabilistic Methods Applied to Power Systems, Istanbul, Turkey.
- [16] D. D. Le, G. Gross, A. Berizzi, Probabilistic modeling of multisite wind farm production for scenario-based applications, *IEEE Trans. Sustainable Energy* 6 (3) (Jul. 2015) 748–758.
- 490 [17] J. E. Jackson, *A Users Guide to Principal Component Analysis*, New York: Wiley, 1991.
- [18] I. T. Jolliffe, *Principal Component Analysis*, New York: Springer, 2002.
- [19] G. E. P. Box, G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, San Francisco, CA: Holden-Day, 1976.
- 495 [20] M. E. Tipping, C. M. Bishop, Probabilistic principal component analysis, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61 (3) (1999) 611–622.
- [21] F. Grubbs, Procedures for detecting outlying observations in samples, *Technometrics* 11 (1) (Feb. 1969) 1–21.
- 500 [22] IEC - International Electrotechnical Commission, *Wind Turbines - Part 12-1: Power Performance Measurements of Electricity Producing Wind Turbines*, IEC, iEC-61400-12 (2005).
- [23] J. Li, A. D. Heap, *A review of spatial interpolation methods for environmental scientists*, Geoscience Australia, 2008.
- 505