

# Regional Wind Speed Evolution Identification and Long-term Correlation Application

Identification de l'évolution régionale de la vitesse du vent et application à la corrélation long terme

B. Buffard, Theolia France, Montpellier



EXTERNAL ARTICLE

ENGLISH - FRANÇAIS

## Summary

Long-term correlation of wind data realized during a wind farm AEP study strongly depends on the data source selected as the reference and can result in high errors. To reduce uncertainties in this aspect of the study, this article proposes to identify regional wind speed evolution taking into account several data sources near the project (weather stations, measurement masts, virtual masts, Reanalysis data and production indexes). It is obtained by a multivariate statistical analysis called Principal Component Analysis and consists in a set of uncorrelated time series. These Principal Components are then used in a multivariate linear regression to reconstruct in the long term the measurement mast of the project and then realize long term correlation study. The result is insensitive to the withdrawal of one of the selected data sources, which confirms its relevance to the site.

## Introduction

A wind farm potential assessment requires a study of wind speeds in the long term to get full life representative production of the park. Without this study, wind speeds measured at the mast could induce an underestimated (and respectively an overestimated) production if the period corre-

## Résumé

La corrélation long terme des données de vent réalisée lors d'une étude de potentiel éolien est fortement dépendante de la source de données choisie comme référence et peut engendrer des erreurs importantes. Afin de réduire les incertitudes sur cet aspect de l'étude de potentiel, cet article propose d'établir l'évolution long terme de la vitesse du vent sur la région du projet en tenant compte des différentes sources de données environnantes (stations météorologiques, mâts de mesure, mâts virtuels, données de Réanalyse, indices de production). Cette dernière est obtenue grâce à une méthode statistique d'analyse de données multivariées appelée Analyse en Composantes Principales et se matérialise sous la forme d'un jeu de séries temporelles non corrélées entre elles. Ces Composantes Principales sont ensuite utilisées dans une régression linéaire multivariée pour reconstituer le mât de mesure du projet sur le long terme puis réaliser la corrélation long terme de l'étude de potentiel. Le résultat obtenu est alors sensible au retrait de l'une des sources de données sélectionnées, ce qui confirme sa pertinence pour le site.

## Introduction

L'évaluation du potentiel d'une ferme éolienne nécessite la

sponds to lower (or respectively higher) wind speeds than in previous years.

Usually, long term wind data of a wind farm is obtained using wind data measured by a weather station close to and representative of the project site [1]. Search for the ideal weather station can be more or less complicated depending on the proximity, the complexity of the field, wind regimes in the region, the availability, consistency and stability of the measurements. The issue of the weather station choice is crucial but depends on the qualitative assessment of the person in charge of the study.

To overcome a large part of this problem, this article proposes to replace the choice of a single weather station as a reference by identifying the evolution of long term wind speeds in the region of the project site from several nearby weather stations and Reanalysis data [2] or wind production indexes. This identification will be carried out through the use of a multivariate statistical analysis called Principal Components Analysis.

Once identified, regional wind speed evolution is directly correlated with wind speeds measured on site to reconstruct a long term representative time series. This data set may then be employed, for example, to establish the long term correlation used in the wind farm AEP study. This article will focus on presenting the statistical method of Principal Component Analysis and its utilization in the identification of regional wind speed evolution. Finally, a direct application will be proposed to realize the long term correlation in the wind farm AEP study.

### **Principal Component Analysis**

Principal Component Analysis (PCA) [3] is a statistical method for analyzing multivariate data and visualizing a set of time series in a new light better explaining data variance. PCA thus diagonalizes the variance-covariance matrix of the data and then projects it in this new orthogonal basis. Time series resulting from this transformation are uncorrelated and called Principal Components. Their construction corresponds to successive projections on orthogonal planes while maximizing the variance explained by each Principal Component. Each initial time series can be decomposed into a linear combination of the Principal Components.

Result of this method is to concentrate in the first Principal Components most of the variance contained in the initial time series. Last Principal Components do not explain much of the data variance and so may be considered as noise and therefore be neglected later in the analysis.

Contributions of initial time series to the construction of Principal Components provide information for their interpretation. They will oppose / gather / exclude certain time series. This information will allow deduction of what each Principal Component represents. Generally, the first one with the largest variance represents the shared evolution of initial time series. This must be confirmed by a significant explained variance with same sign and same level contributions.

### **Regional Long Term Wind Speeds Evolution Identification**

The following paragraphs describe the use of PCA on several wind speed series from different data sources (weather

réalisation d'une étude des vitesses de vent sur le long terme permettant d'obtenir un productible représentatif de la vie complète du parc. Sans cette étude, les vitesses de vent mesurées au mât pourraient induire une sous-estimation (respectivement une sur-estimation) du potentiel si la période de mesures correspond à des vitesses de vent plus faibles (respectivement plus fortes) que les années précédentes.

Habituellement, les données de vent long terme d'un parc éolien sont obtenues à l'aide des données de vent mesurées par une station météorologique proche et représentative de ce site [1]. La recherche de la station météorologique idéale peut alors s'avérer plus ou moins compliquée selon la proximité, la complexité du terrain, les régimes de vent dans la région, la disponibilité, la consistance et la stabilité des mesures. L'enjeu du choix de la station météorologique est majeur mais dépend de l'appréciation qualitative de la personne en charge de l'étude.

Pour remédier en grande partie à cette problématique, cet article propose de remplacer le choix d'une unique station météorologique de référence par l'identification de l'évolution long terme des vitesses du vent dans la région du site à partir de plusieurs stations météorologiques environnantes et éventuellement de données de Réanalyse [2] ou encore d'indices de production éoliens. Cette identification sera menée grâce à l'emploi d'une méthode statistique d'analyse de données multivariées appelée Analyse en Composantes Principales.

Une fois identifiée l'évolution long terme des vitesses du vent dans la région, celle-ci est directement corrélée aux vitesses de vent mesurées sur site afin de reconstituer une série de données représentative du long terme. Cette série de données pourra alors servir, par exemple, à l'établissement de la corrélation long terme utilisée dans le cadre de l'étude de potentiel du parc éolien. Cet article s'attachera à présenter la méthode statistique de l'Analyse en Composantes Principales puis son utilisation dans le cadre de l'identification de l'évolution de la vitesse du vent dans une région. Enfin, une application directe sera proposée pour la réalisation de la corrélation long terme lors de l'étude de potentiel éolien.

### **Analyse en Composantes Principales**

L'Analyse en Composantes Principales (ACP) [3] est une méthode statistique d'analyse de données multivariées permettant de visualiser un jeu de séries temporelles sous un nouvel angle expliquant au mieux la variance des données. L'ACP s'applique ainsi à diagonaliser la matrice de variance-covariance des données pour ensuite les projeter dans cette nouvelle base orthogonale. Les séries temporelles résultant de cette transformation sont alors non corrélées entre elles et appelées Composantes Principales. Leur construction correspond à des projections successives sur des plans orthogonaux tout en maximisant la variance expliquée par chaque Composante Principale. Chaque série temporelle initiale peut alors être décomposée en une combinaison linéaire des Composantes Principales.

Le résultat de cette méthode est de concentrer dans les premières Composantes Principales la majorité de la variance contenue dans les séries temporelles de départ. Les dernières Composantes Principales expliquant très peu la variance des données peuvent être considérées comme du bruit présent dans les données et donc être négligées dans la suite de l'analyse.

stations, measurement masts, virtual masts, Reanalysis data and production indexes), in the search of the evolution of regional long term wind speeds.

### Use of Principal Component Analysis

Using PCA on this type of time series aims to identify the general behavior of wind speed in a particular region and especially amplitudes of wind speed variation on a specific time step. The wind speed level can then be determined by measurements on a site of interest in the region.

PCA realization must lead to the selection of one or several Principal Components explaining wind speed variations on the studied area. As mentioned above, the first Principal Component will often alone match the general evolution of wind speed in the region.

Unselected Principal Components could also be studied to reveal the specificities of each initial time series: for example a change of local environment or configuration of the weather station's mast.

### Screening of data sources

Time series used in PCA should be screened. They must demonstrate a good availability of data over the long term period considered, but also correspond to the region studied in terms of wind climate. The main aspects looked at for this first selection (proximity, topography, roughness, quality measurement...) are identical to those usually regarded for the choice of a single long term reference [1].

Lisbon, Portugal  
13 - 15 Nov. 2012

**11<sup>th</sup> Wind Integration Workshop**

*International Workshop  
on Large-Scale Integration of  
Wind Power into Power Systems  
as well as on Transmission  
Networks for Offshore  
Wind Power Plants*

[www.windintegrationworkshop.org](http://www.windintegrationworkshop.org)

Les contributions des séries temporelles initiales à la formation des Composantes Principales en donnent une interprétation. Ces dernières vont alors opposer / rassembler / exclure certaines séries temporelles, ce qui permettra d'en déduire ce qu'elles représentent. Généralement, la première Composante Principale disposant de la variance la plus importante représente l'évolution générale des séries temporelles initiales. Ceci doit être confirmé par une variance expliquée importante et par des contributions de même niveau et de même signe.

### Identification de l'évolution long terme des vitesses du vent dans une région

Les paragraphes suivants décrivent l'utilisation de l'ACP sur plusieurs séries de vitesses de vent provenant de sources de données différentes (stations météorologiques, mâts de mesure, mâts virtuels, données de Réanalyse, indices de production), dans le cadre de la recherche de l'évolution long terme des vitesses du vent dans une région.

### Emploi de l'Analyse en Composantes principales

L'objectif attendu en employant l'ACP sur ce type de séries temporelles est d'identifier le comportement général de la vitesse du vent dans une région particulière. Il s'agit d'identifier sur un pas de temps déterminé les amplitudes de variation de la vitesse du vent mais pas nécessairement son niveau qui pourra être déterminé ensuite par des mesures effectuées sur un site d'intérêt dans la région.

La réalisation de l'ACP doit conduire à la sélection d'une ou plusieurs Composantes Principales expliquant les variations de la vitesse du vent dans la région étudiée. Comme évoqué ci-dessus, la première Composante Principale correspondra souvent à elle seule à l'évolution générale de la vitesse du vent dans la région.

Les Composantes Principales non retenues pourront éventuellement être étudiées afin de faire apparaître les spécificités de chacune des séries temporelles initiales : par exemple un changement d'environnement local ou de configuration du mât de la station météorologique.

### Présélection des sources de données

Les séries temporelles à utiliser dans l'ACP devront avoir été présélectionnées. Elles doivent en effet faire preuve d'une bonne disponibilité des données sur la période long terme considérée, mais surtout correspondre à la région étudiée en termes de climat de vent. Les principaux aspects regardés pour cette première sélection (proximité, topographie, rugosité, qualité de mesure...) sont donc identiques à ceux regardés habituellement pour le choix d'une unique référence long terme [1].

Pour les autres sources de données pouvant être utilisées ici, telles que les mâts virtuels, les données de Réanalyse et les indices de production, il faudra également veiller à la pertinence des méthodes de calcul utilisées.

Enfin, cette présélection doit correspondre à un échantillon équilibré des sources de données disponibles afin de ne pas accorder plus d'importance à certaines caractéristiques du vent sur la région. Par exemple, les sources de données doivent correspondre à un quadrillage régulier de la région et éviter que certains secteurs soient surreprésentés. En règle générale, les redondances d'information devront être évitées ou bien apparaître dans les mêmes proportions.

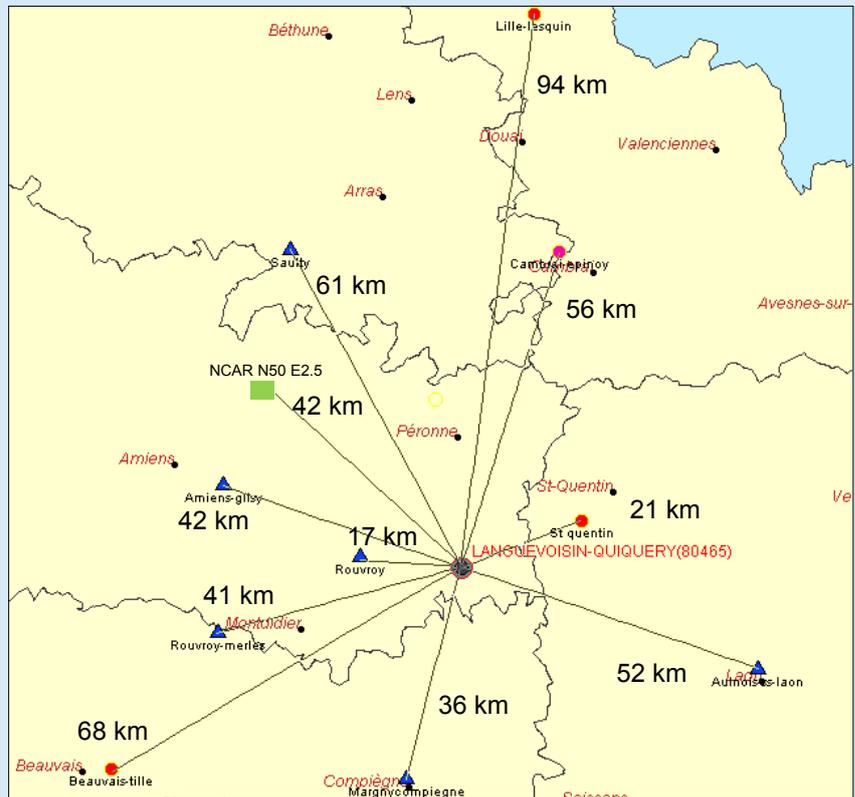


Fig. 1: Data sources selected in the project region

Sources: Météo France, NCEP

Fig. 1: Sources de données sélectionnées dans la région du projet

For other data sources that can be used here, such as virtual masts, Reanalysis data and production indexes, relevance of used calculation methods will have to be looked at carefully.

Finally, this screening must result in a balanced sample of available data sources not to give more importance to certain wind characteristics in the region. For example, data sources must match a regular grid of the area and avoid that certain sectors are over-represented. In general, information redundancy must be avoided or appear in the same proportions.

### Data pre-processing

Once this screening is carried out, data collected must be treated to avoid level effects between sources: it should be avoided that a data source (called  $DS_{i,t}$  below) with very high wind speeds has more weight than another with low wind speeds.

Use of standardized data in the PCA [3] will give the same weight to each of the data sources regardless of its level in absolute value. Each data source will therefore undergo the following operation:

$$\frac{DS_{i,t} - \mu_i}{\sigma_i} \quad (1)$$

$\mu_i$  being the average of  $DS_{i,t}$

$\sigma_i$  being the standard deviation of  $DS_{i,t}$

### Successive PCA achievements

According to the characteristics of the region, the screening of data sources will not necessarily be directly optimal for

### Traitement préalable des données

Une fois cette présélection effectuée, les données recueillies doivent être traitées afin d'éviter des effets de niveau entre les différentes sources : il faut éviter qu'une source de données (nommée  $DS_{i,t}$  ci-après) présentant des vitesses de vent élevées ait plus de poids qu'une autre présentant des vitesses faibles.

L'utilisation de données centrées et réduites pour la réalisation de l'ACP [3] permettra d'accorder la même importance à chacune des sources de données quel que soit son niveau en valeur absolue. Chaque source de données subira donc l'opération suivante :

$$\frac{DS_{i,t} - \mu_i}{\sigma_i} \quad (1)$$

$\mu_i$  étant la moyenne des données  $DS_{i,t}$

$\sigma_i$  étant l'écart-type des données  $DS_{i,t}$

### Réalisations itératives de l'ACP

Selon les caractéristiques de la région, la présélection des sources de données ne sera pas forcément directement optimale pour déterminer l'évolution générale du vent sur la région. L'interprétation des contributions à la formation des Composantes Principales pourra alors s'avérer très utile pour valider la pertinence des sources de données. Ainsi, il conviendra d'affiner la sélection des sources de données par des réalisations successives de l'ACP jusqu'à l'obtention de Composantes Principales satisfaisantes en termes de variance expliquée et de représentation d'une évolution des vitesses de vent partagée par l'ensemble des sources de données retenues. La stabilité des résultats de l'ACP pourra ensuite être vérifiée

Data Source	Saint Quentin	Cambrai Epinoy	Rouvroy en Sarterre	Saulny	Lille-Lesquin	Margny Compiègnes	Beauvais Tillé	Amiens-Glisly	Rouvroy Merles	Aulnoy sous Laon	NCAR N50 E2.5 10m
Distance from project mast [km]	21	56	17	61	94	36	68	42	41	52	42
R <sup>2</sup> with project mast	94%	92%	90%	83%	85%	83%	81%	79%	79%	74%	81%
Mean wind speed 1997-2011 [m/s]	4,2	4,9	4,4	4,7	4,8	4,2	4,3	3,8	3,4	3,9	4,8
Long term correction	7,70%	6,10%	4,10%	3,50%	4,00%	4,60%	5,30%	5,50%	2,40%	5,40%	3,30%

Fig. 2: Long term correction obtained with a single data source  
Sources: Météo France, NCEP, THEOLIA France – Calculations: THEOLIA France  
Fig. 2: Correction long terme obtenue avec une seule source de données

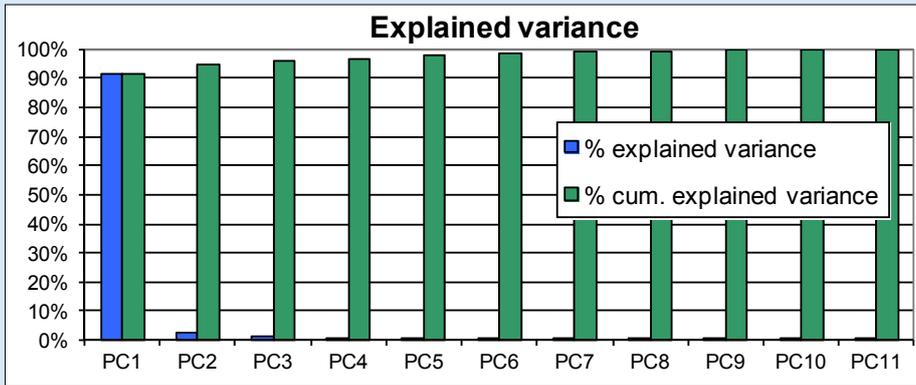


Fig. 3: % of explained variance for each Principal Component  
Sources: Météo France, NCEP – Calculations: THEOLIA France  
Fig. 3: % de variance expliquée par chacune des Composantes Principales

determining the general wind evolution in the region. Interpretation of contributions to the construction of Principal Components can prove very useful for validating the relevance of data sources. Thus, it would be appropriate to refine the data sources selection through successive PCA achievements until Principal Components are satisfactory in terms of explained variance and representation of a wind speed evolution shared by all selected data sources. Stability of PCA results can be verified by sensitivity tests for each selected data source: these tests could involve, for example, comparison of obtained Principal Components omitting in turn one of the data sources.

### Principal Components selection

A high correlation between selected Principal Components and data sources will confirm the results of the estimation procedure.

Choice of the number of Principal Components (called  $PC_{j,t}$  below) will be carried out by looking at  $R^2$  determination coefficient obtained for the estimation of the following multivariate linear regression:

$$DS_{i,t} = \alpha_{i,0} + \alpha_{i,1} \cdot PC_{1,t} + \alpha_{i,2} \cdot PC_{2,t} + \dots + \alpha_{i,n} \cdot PC_{n,t} + \varepsilon_{i,t} \quad (2)$$

$\alpha_{i,j}$  being regression coefficients to be estimated

$\varepsilon_{i,t}$  being residuals of the regression

In case where a single Principal Component is selected, the  $R^2$  indicator directly corresponds to the correlation coefficient squared. Higher the  $R^2$  indicator will be and better data from the source will be represented by the Principal Components.

par des tests de sensibilité à chacune des sources de données retenues : ces tests consisteront par exemple à comparer les Composantes Principales obtenues en omettant tour à tour l'une des sources de données.

### Choix des Composantes Principales à retenir

Une corrélation élevée entre les Composantes Principales retenues et les données sources confirmera les résultats de la procédure d'estimation.

Le choix du nombre de Composantes Principales (nommées  $PC_{j,t}$  ci-après) sera effectué grâce au coefficient de détermination  $R^2$  obtenu pour l'estimation de la régression linéaire multivariée suivante :

$$DS_{i,t} = \alpha_{i,0} + \alpha_{i,1} \cdot PC_{1,t} + \alpha_{i,2} \cdot PC_{2,t} + \dots + \alpha_{i,n} \cdot PC_{n,t} + \varepsilon_{i,t} \quad (2)$$

$\alpha_{i,j}$  étant les coefficients de régression à estimer

$\varepsilon_{i,t}$  étant les résidus de la régression

Dans le cas où une seule Composante Principale est sélectionnée, l'indicateur  $R^2$  correspond directement au coefficient de corrélation élevé au carré. Plus l'indicateur  $R^2$  sera élevé et plus les données de la source seront bien représentées par les Composantes Principales.

### Application à la correction long terme de la vitesse du vent

Les paragraphes suivants détaillent l'une des applications possibles de l'identification de l'évolution long terme de la vitesse du vent dans une région. Il s'agit de l'établissement de la corrélation long terme des vitesses de vent pour un projet situé dans la région étudiée.

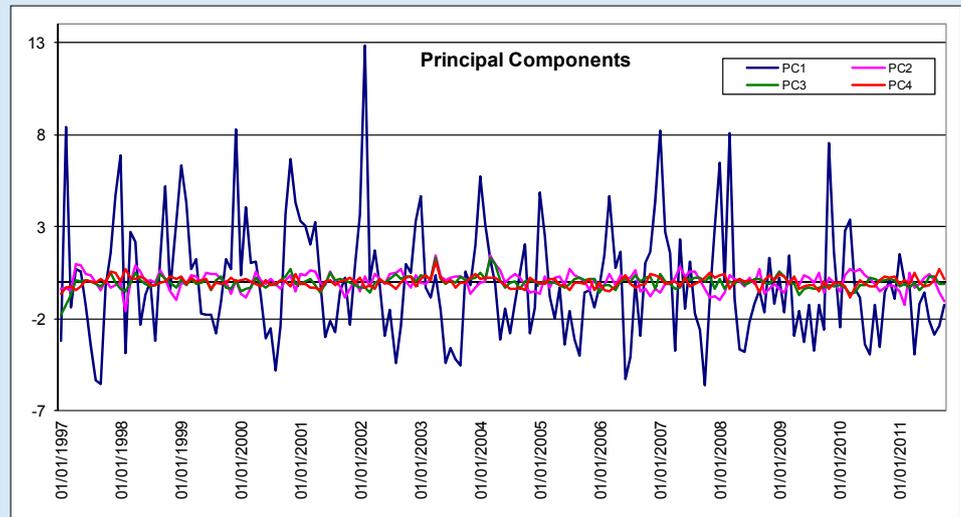


Fig. 4 : Principal Components temporal evolution (standardized values)  
Sources: Météo France, NCEP – Calculations: THEOLIA France

Fig. 4 : Évolution temporelle des Composantes Principales (valeurs centrées réduites)

Data Source	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7	PC 8	PC 9	PC 10	PC 11
Saint Quentin	0,31	0,08	0,10	0,01	-0,41	-0,34	-0,36	0,51	0,04	0,42	-0,17
Cambrai Epinoy	0,31	-0,22	0,07	0,11	-0,40	0,24	-0,36	-0,42	-0,55	-0,11	0,08
Rouvroy en Santerre	0,31	-0,06	0,16	0,00	-0,15	-0,02	-0,21	-0,38	0,77	-0,24	-0,10
Saulty	0,29	-0,44	-0,51	-0,25	0,47	-0,04	-0,37	0,17	0,00	-0,09	0,04
Lille-Lesquin	0,31	-0,13	-0,12	-0,05	-0,02	0,50	0,33	-0,10	0,17	0,67	0,17
Marigny Compiègnes	0,29	0,56	-0,49	0,57	0,11	-0,09	-0,03	-0,11	-0,01	-0,03	0,09
Beauvais Tillé	0,30	0,30	0,17	-0,44	-0,02	-0,17	0,08	0,06	-0,05	-0,16	0,73
Amiens-Glisy	0,31	0,16	-0,03	-0,10	-0,16	0,52	0,22	0,47	-0,02	-0,48	-0,29
Rouvroy Merles	0,30	0,27	0,03	-0,44	0,20	-0,25	0,20	-0,36	-0,23	0,13	-0,55
Aulnoy sous Laon	0,30	-0,04	0,64	0,36	0,57	0,08	-0,08	0,12	-0,09	0,04	0,00
NCAR N50 E2,5 10m	0,30	-0,48	-0,05	0,27	-0,16	-0,45	0,59	0,04	-0,06	-0,16	0,03

Fig. 5 : Contributions to Principal Components construction  
Sources: Météo France, NCEP – Calculations: THEOLIA France

Fig. 5 : Contributions à la formation des Composantes Principales

### Long Term Wind Speed Correction Application

The following paragraphs detail one of the possible applications of this regional long term wind speed evolution identification. It is the establishment of long term wind speed correlation for a project located in the studied area.

#### Long term wind speed reconstruction

The study of long term correlation proposed below goes through reconstructing in the long term wind speeds at the location of the measurement mast installed on the site of the wind farm project.

Once regional long term wind speed evolution has been identified, selected Principal Components will be used in the following multivariate linear regression to reconstruct long term wind speeds (called  $WS_t$  below) at the location of the measurement mast installed on site:

$$WS_t = \beta_0 + \beta_1 \cdot PC_{1,t} + \beta_2 \cdot PC_{2,t} + \dots + \beta_p \cdot PC_{p,t} + \eta_t \quad (3)$$

$\beta_i$  being regression coefficients to be estimated

$\eta_t$  being residuals of the regression

### Reconstitution long terme des vitesses du vent

L'étude de la corrélation long terme proposée ci-après passe par la reconstitution sur le long terme des vitesses de vent à l'emplacement du mât de mesure installé sur le site du projet de centrale éolienne.

Une fois identifiée l'évolution long terme de la vitesse du vent dans la région, les Composantes Principales sélectionnées seront utilisées dans la régression linéaire multivariée suivante pour reconstituer sur le long terme les vitesses du vent (nommées  $WS_t$  ci-après) à l'emplacement du mât de mesure installé sur site :

$$WS_t = \beta_0 + \beta_1 \cdot PC_{1,t} + \beta_2 \cdot PC_{2,t} + \dots + \beta_p \cdot PC_{p,t} + \eta_t \quad (3)$$

$\beta_i$  étant les coefficients de régression à estimer

$\eta_t$  étant les résidus de la régression

L'estimation de cette régression sera alors employée pour prédire les vitesses du vent à l'emplacement du mât de mesure sur site pour les périodes où le mât n'était pas en fonction mais pour lesquelles les autres sources de données sont disponibles. La robustesse de l'estimation dépendra de la lon-

R <sup>2</sup> coefficient	Saint Quentin	Cambrai Epinoy	Rouvroy en Santerre	Saulty	Lille-Lesquin	Marigny Compiègnes	Beauvais Tillé	Amiens-Glisy	Rouvroy Merles	Aulnoy sous Laon	NCAR N50 E2,5 10m	PC 1	Project mast
Saint Quentin	100%												
Cambrai Epinoy	90%	100%											
Rouvroy en Santerre	92%	94%	100%										
Saulty	77%	83%	81%	100%									
Lille-Lesquin	86%	92%	92%	85%	100%								
Marigny Compiègnes	92%	85%	92%	76%	90%	100%							
Beauvais Tillé	90%	83%	90%	72%	85%	94%	100%						
Amiens-Glisy	90%	88%	90%	77%	92%	92%	90%	100%					
Rouvroy Merles	86%	81%	88%	76%	86%	92%	94%	88%	100%				
Aulnoy sous Laon	83%	83%	86%	71%	83%	83%	81%	81%	81%	100%			
NCAR N50 E2,5 10m	83%	86%	86%	83%	86%	79%	74%	79%	74%	77%	100%		
<b>PC 1</b>	<b>94%</b>	<b>94%</b>	<b>98%</b>	<b>86%</b>	<b>96%</b>	<b>96%</b>	<b>92%</b>	<b>94%</b>	<b>92%</b>	<b>88%</b>	<b>88%</b>	<b>100%</b>	
Project mast	94%	92%	90%	83%	85%	83%	81%	79%	79%	74%	81%	<b>94%</b>	100%

Fig. 6: R<sup>2</sup> determination coefficients  
Sources: Météo France, NCEP, THEOLIA France – Calculations: THEOLIA France

Fig. 6 : Coefficients de détermination R<sup>2</sup>

Data source excluded:	Saint Quentin	Cambrai Epinoy	Rouvroy en Santerre	Saulty	Lille-Lesquin	Marigny Compiègnes	Beauvais Tillé	Amiens-Glisy	Rouvroy Merles	Aulnoy sous Laon	NCAR N50 E2,5 10m
Long term correction	4,90%	5,10%	5,30%	5,30%	5,30%	5,30%	5,10%	5,10%	5,50%	5,10%	5,30%

Fig. 7: Long term corrections omitting one data source  
Sources: Météo France, NCEP, THEOLIA France – Calculations: THEOLIA France

Fig. 7 : Corrections long terme en omettant l'une des sources de données

This regression estimation will then be used to predict wind speeds at the location of the measurement mast on site on the periods where the mast was not functioning but for which other data sources are available. Robustness of the estimation will depend on the length of data available at the measurement mast and on the chosen time step.

The reconstructed measurement mast will get the same time step as the Principal Components which can be different from the one of the measurement mast. In this case, the reconstructed measurement mast will not be directly usable as a long term input for the wind farm AEP study. A long term correlation of the measurement mast data will still be needed. It will be based on a long term correction coefficient obtained with the reconstructed measurement mast data.

### Long term correction coefficient estimation

Once the mast is reconstructed in the long term (and if its time step differs from measurement mast's one), it will be used to calculate the long term correction factor to be applied to wind speeds measured at the mast and then conduct wind farm AEP study.

Long term correction factor used here is the ratio of the long term average of wind speeds reconstructed at the measure-

ment mast and the long term average of wind speeds measured at the measurement mast.

Le mâât de mesures reconstruit aura un pas temporel identique à celui des Composantes Principales qui pourra différer de celui du mâât de mesures. Dans ce cas, le mâât de mesures reconstruit ne sera pas directement utilisable comme donnée long terme pour l'étude de potentiel éolien : il sera toujours nécessaire d'effectuer une corrélation long terme des données du mâât de mesures. Celle-ci sera basée sur un coefficient de correction long terme obtenu à partir du mâât de mesures reconstruit.

### Calcul du coefficient de correction long terme

Une fois le mâât reconstitué sur le long terme (et si son pas de temps diffère de celui du mâât de mesures), celui-ci servira à calculer le coefficient de correction long terme à appliquer aux vitesses de vent mesurées au mâât pour réaliser l'étude de potentiel éolien.

Le coefficient de correction long terme utilisé ici est le rapport entre la moyenne long terme des vitesses de vent reconstituées à l'emplacement du mâât de mesure et la moyenne court terme des vitesses de vent mesurées au mâât. Ce rapport est calculé en tenant compte de la saisonnalité annuelle du vent. La sélection de la période long terme utilisée pour calculer ce

ment mast location and the short term average of wind speeds measured at the mast. This ratio is calculated by taking into account the annual seasonality of the wind.

Selection of the long term period to calculate this correction coefficient will be carried out regarding the stability analysis of the reconstructed measurement mast wind data (cf. [1]).

### Application to a site in the north of France

The following example is based on a project located in the commune of Languevoisin-Quiquery, a slightly rugged area in the north of France. The project's mast has 40 months of wind measurements from 2008 to 2011. The long term reference period considered is from January 1997 to December 2011, e.g. 15 full years. The study will be conducted on monthly averaged wind speeds.

Screening of data sources in the region has retained 10 weather stations and 1 set of Reanalysis data (see Fig. 1).

Direct use of one of these data sources would have given the long term correction factors displayed in Fig. 2 that differ strongly from one data source to another. This method for selecting a long term reference shows the risk of a significant error on wind levels and consequently an even higher risk for the wind power potential of the site. PCA realization on all of these data sources shows a first Principal Component gathering 91% of their total variance (Fig. 3).

With less than 3% of explained variance the 10 remaining Principal Components can be considered as noise corresponding to the specificities of each data source. These results are confirmed by the Principal Components temporal graph (Fig. 4) which shows the predominance of the first Principal Component.

Furthermore, analysis of contributions to the Principal Components construction confirms that the first Principal Component fits well with the general evolution of the wind speeds in the region since all data sources contribute to an equivalent level (Fig. 5).

Contributions to other Principal Components construction reveal specificities of the different data sources. Further analysis could provide more information on these features, but the first Principal Component is sufficient to obtain the regional long term wind speed evolution in this example.

As presented in Fig. 6, the first Principal Component is strongly correlated with each data source without exception and the correlation with the project mast is also excellent ( $R^2$  calculated on different periods lengths). It confirms once again that PCA identified the regional long term wind speed evolution.

Using it as long term reference, the long term correction obtained is + 5,2 %. This figure should be compared with those obtained for each data source considered separately as long term reference (Fig. 1). The error could have been very significant in the case of the choice of Saint-Quentin for example while this data source has the best correlation to the measurement mast. Note also that the long term correction with PCA is not just to take the average of long term corrections obtained for each data source independently (+ 4,7 %).

Finally, the sensitivity of the results to the presence of different data sources in the PCA has been tested. Results from PCA remain stable in all cases. And variation of the long term correction is limited to 0,3 % (see Fig. 7). Influence of each data source is relatively low on the final result and

coefficient de correction sera réalisée directement à partir de l'analyse de stabilité des données de vent du mât reconstitué (cf. [1]).

### Application à un site du Nord de la France

L'exemple qui suit est basé sur un projet situé sur la commune de Languevoisin-Quiquery, dans une région faiblement accidentée du Nord de la France. Le mât du projet dispose de 40 mois de mesures du vent de 2008 à 2011. La période de référence long terme considérée court de janvier 1997 à décembre 2011 soit 15 années complètes. L'étude sera réalisée sur les moyennes mensuelles des vitesses de vent.

La présélection des sources de données dans la région a permis de retenir 10 stations météorologiques et 1 jeu de données de Réanalyse (cf Fig. 1).

L'utilisation directe d'une seule de ces sources de données aurait donné les facteurs de correction long terme affichés dans la Fig. 2 qui diffèrent fortement d'une source de données à l'autre. Cette méthode pour la sélection de la référence long terme présente alors le risque d'une erreur importante sur les niveaux de vent et par conséquent encore plus élevée sur le potentiel éolien du site. La réalisation de l'ACP sur l'ensemble de ces sources de données fait ressortir une première Composante Principale rassemblant 91% de leur variance totale (Fig. 3).

Tandis qu'avec moins de 3% de variance expliquée les 10 Composantes Principales restantes peuvent être considérées comme étant du bruit correspondant aux spécificités de chacune des sources de données. Ces résultats sont confortés par le graphique temporel des Composantes Principales (Fig. 4) qui fait apparaître la prédominance de la première Composante Principale.

Par ailleurs, l'analyse des contributions à la formation des Composantes Principales confirme que la première Composante Principale correspond bien à l'évolution générale des vitesses du vent dans la région puisque l'ensemble des sources de données contribue à sa formation à un niveau équivalent (Fig. 5).

Les contributions à la formation des autres Composantes Principales révèlent les spécificités des différentes sources de données. Une analyse plus approfondie pourrait apporter davantage d'informations sur ces particularités, mais la première Composante Principale est suffisante pour obtenir l'évolution long terme de la vitesse du vent dans la région.

Comme présenté dans la Fig. 6, la première Composante Principale est fortement corrélée avec chacune des sources de données sans exception et la corrélation avec le mât du projet est également excellente ( $R^2$  calculés sur des périodes de longueurs différentes). Cela confirme une fois de plus que l'ACP a permis d'identifier l'évolution long terme du vent dans la région.

En utilisant cette dernière comme référence long terme, la correction long terme obtenue est de + 5,2 %. Ce chiffre doit être comparé à celui obtenu pour chacune des sources de données prises séparément comme référence long terme (Fig. 1). L'erreur aurait pu être très importante dans le cas du choix de Saint-Quentin par exemple alors que cette source de données était la mieux corrélée au mât de mesure. A noter également que la correction long terme obtenue avec l'ACP ne se résume pas à retenir la moyenne des corrections long terme obtenues pour chacune des sources de données indépendamment (+ 4,7 %).

therefore reinforces the fact that identification of the regional wind speed evolution was properly conducted.

## Conclusion

The proposed method shows the advantage of defining a regional long term wind speed reference coherent with all surrounding data sources. Unlike the classical study for the choice of a single source of data as long term reference, it preserves from a qualitative judgment that can generate a significant error. Avoided error will be more important when the topography is complex and the nearby data sources became scarce. Obviously the more data sources that corroborate the result, higher the certainty to have identified regional long term wind speed evolution.

Other advantages of the method are the possibility to use various data sources (weather stations, Reanalysis data, measurement masts, virtual masts, production indexes) and to get an estimate free of any temporal instability due to a particular data source. However, this method does not resolve the problem of an error affecting all data sources simultaneously: for example if all studied weather stations were installed over a period with the same type of anemometers overestimating wind speeds.

Use of PCA on wind data allows wider applications than the only study of the long term wind speed correction. Indeed, the method allows identifying and extracting specific features and/or inconsistencies in each data source. It could also be adapted to study long term wind direction frequency evolution or the reconstruction of production indexes. Moreover, the study of cycles in the general evolution of the wind in a region could help defining the long term reference period to be considered to get wind farm AEP estimations independent of the possible wind cycles.

Finally, PCA has some features unused here that could be considered to improve the estimation, such as data weighting. This could allow taking into account the distance from the data source to the studied site or even under-represent some recognized erroneous data over a certain period. Furthermore, the study of contributions of individuals (different dates of the time series) to the construction of Principal Components would provide additional elements of interpretation. Such a process will help to oppose periods with each other and thus to give additional meaning to the different Principal Components.

## References / Bibliographie

- [1] V. Borget, P.A. Monnier, M. Strack: "Long Term Scaling of Site Measurements: Evaluation of Long Term Meteorological Data in France and Comparison of Correlation Methods", DEWI Magazin Nr. 30, February 2007.
- [2] NOAA-CIRES Climate Diagnostics Center: NCEP Reanalysis data, Boulder, Colorado, USA, available at the Web site at <http://www.esrl.noaa.gov/psd/>
- [3] G. Saporta: « Probabilités, analyse des données et statistique », éditions TECHNIP, 2006.

Enfin, la sensibilité des résultats à la présence des différentes sources de données dans l'ACP a été testée. Les résultats de l'ACP restent stables dans tous les cas. Et la variation de la correction long terme est limitée à 0,3 % (cf Fig. 7). L'influence de chaque source de données est donc relativement faible sur le résultat final et conforte donc le fait que l'identification de l'évolution de la vitesse du vent dans la région a été correctement menée.

## Conclusion

La méthode proposée présente l'avantage de définir une référence long terme des vitesses du vent dans une région cohérente avec l'ensemble des sources de données environnantes. Contrairement à l'étude classique menée pour le choix d'une unique source de données comme référence long terme, celle-ci préserve d'un jugement qualitatif pouvant générer une erreur importante. L'erreur ainsi évitée est d'autant plus importante que la topographie est complexe et que les sources de données à proximité se raréfient. Evidemment plus le nombre de sources de données corroborant le résultat est important et plus la certitude est élevée d'avoir obtenu l'évolution long terme de la vitesse du vent dans la région.

Les autres avantages de la méthode sont de pouvoir utiliser des sources de données variées (stations météorologiques, données de Réanalyse, mâts de mesure, mâts virtuels, indices de production) et d'obtenir une estimation exempte de toute instabilité temporelle due à une source de données particulière. En revanche, cette méthode ne résout pas le problème d'une erreur présente dans toutes les sources de données simultanément ; par exemple si toutes les stations météorologiques étudiées ont été installées durant une période avec le même type d'anémomètres surévaluant la vitesse du vent.

L'utilisation de l'ACP sur des données de vent permet des applications plus larges que la seule étude de la correction long terme sur la vitesse du vent. En effet, la méthode permet d'identifier et d'extraire les spécificités et/ou incohérences de chaque source de données. Elle pourrait également être adaptée pour étudier l'évolution long terme des fréquences de direction du vent ou la reconstitution d'indices de production. Par ailleurs, l'étude des cycles présents dans l'évolution générale du vent dans une région permettrait de définir la période de référence long terme à considérer pour disposer d'une estimation de productible indépendante des éventuels cycles de vent.

Enfin, l'ACP présente certaines fonctionnalités non utilisées ici qui pourraient être envisagées pour améliorer l'estimation, comme la pondération des données. Cela permettrait de faire intervenir la distance de la source de données au site étudié ou encore de sous représenter certaines données reconnues erronées sur une certaine période. De plus, l'étude des contributions des individus (les différentes dates des séries temporelles) à la formation des Composantes Principales fournirait des éléments d'interprétation supplémentaires. Un tel procédé permettra d'opposer des périodes les unes aux autres et ainsi de donner davantage de sens aux différentes Composantes Principales.