

¹ Institut Pierre Simon Laplace/Laboratoire de Météorologie Dynamique,
Ecole Polytechnique/ENS/UPMC/CNRS, Palaiseau, France
² Institut Pierre Simon Laplace/Laboratoire des Sciences du Climat et de l'Environnement,
CEA/UVSQ/CNRS, Gif-sur-Yvette, France

Statistical downscaling of near-surface wind over complex terrain in southern France

T. Salameh¹, P. Drobinski¹, M. Vrac², P. Naveau²

With 5 Figures

Received 29 September 2007; Accepted 10 July 2008
Published online 16 December 2008 © Springer-Verlag 2008

Summary

Accurate and rapid determination of near-surface wind fields in a complex area (orography, inhomogeneous surface properties) is a challenge for applications like the evaluation of wind energy production, the prediction of pollution transport and hazardous conditions for aeronautics and ship navigation, or the estimation of damage to farm plantations, among others. This paper presents a statistical downscaling approach based on generalized additive models that provides accurate, rapid and relatively transparent simulations of local-scale near-surface wind field based on a method calibrated on both large-scale upper air and surface atmospheric fields. Our statistical method is used to downscale near-surface wind components to weather surface stations in southern France from ERA-40 reanalyses between 1991 and 2001. The region of interest is characterized by the presence of major mountain ranges which play a major role in redirecting large-scale circulations making difficult the prediction of local wind. This study compares the performance of our statistical approach with different sets of explanatory variables, to explain the near-surface wind field variability. The performances are interpreted by evaluating the contribution of the explanatory variables in the equations of motion. This approach generates accurate depictions of the local surface wind field, and allows to go one

step further in statistical wind speed downscaling. Indeed, it is adapted to explain wind components and not only wind speed and energy in contrast to past studies and it is suited for complex terrain and robust to time averaging in this region.

1. Introduction

Trends in near-surface wind speeds are acknowledged as having particular importance for climate impacts on society (e.g., insurance industry, coastal erosion, forest and infrastructure damage, storm surges, air–sea exchange). They also have relevance for applications such as pollutant diffusion evaluation, wind energy resource estimation and construction issues. Moreover, surface wind speeds exhibit variability at much smaller spatial scales than that resolved by general circulation models (GCM). Hence there is a need to develop tools for downscaling GCM projections to generate finer scale projections of near-surface wind climatologies. Downscaling is the process of deriving regional climate information based on large-scale climate conditions. Both dynamical and statistical downscaling methods have been used extensively in the last

Correspondence: Tamara Salameh, Institut Pierre Simon Laplace/Laboratoire de Météorologie Dynamique, Ecole Polytechnique, 91128 Palaiseau Cédex, France (E-mail: tamara.salameh@lmd.polytechnique.fr)

decade to produce regional climate (see Wilby and Wigley 1997 for a review). Statistical downscaling is a computationally inexpensive method for obtaining high-resolution climate from GCMs by deriving statistical relationships between observed small-scale (often station level) variables and larger (GCM) scale variables, using either circulation typing (or weather regimes) or regression analysis (Wilby and Wigley 1997; Wilby et al. 1998). Statistical downscaling of wind is generally restricted to wind speed and energy (for wind energy application) (e.g., De Rooy and Kok 2004; Gutiérrez et al. 2004; Pryor et al. 2005) and is not designed to produce wind components. Dynamical downscaling consists in driv-

ing a regional climate model (RCM) by a GCM over an area of interest since decreasing grid spacing in meso-scale models generally improves the realism of the results (Mass et al. 2002). It is computationally more expensive but it is suited to complex terrain environment and gives access to the gridded fine-scale wind vector field (Zagar et al. 2006; Frech et al. 2007; Gustafson and Leung 2007).

This paper presents a statistical downscaling approach based on generalized additive models (GAM) that provides accurate, rapid and relatively transparent simulations of local-scale near-surface wind components based on a method calibrated on both large-scale upper air and sur-

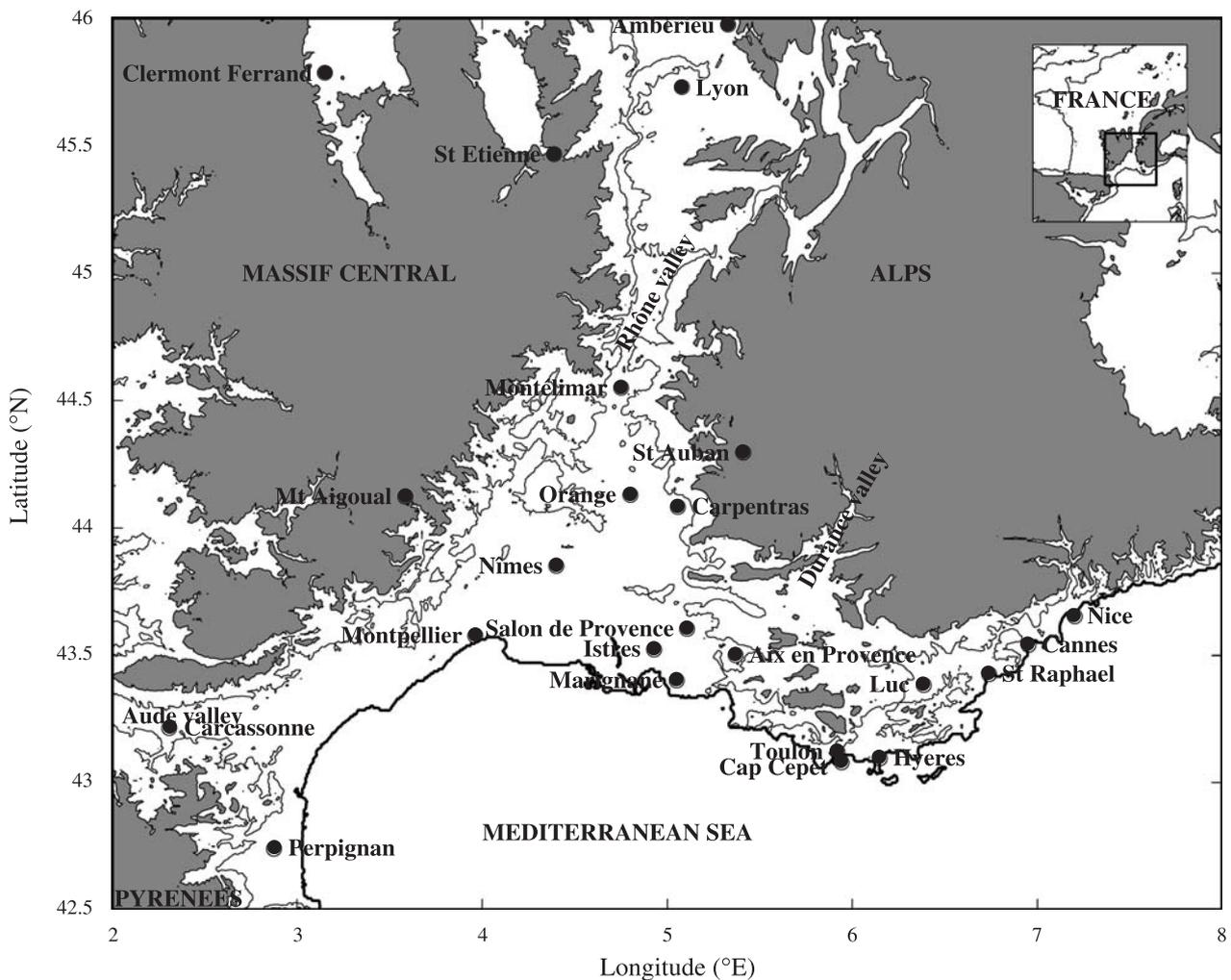


Fig. 1. Map of southern France with the topography shaded in grey when higher than 500 m above sea level. 200 m height isocontour as well as the coastline are shown with solid and thick solid lines, respectively. The large black dots indicate the locations of the operational meteorological surface stations operated by Météo-France. The small subpanel at the upper right corner indicates with a rectangle the region of France shown in the main panel

face atmospheric fields. Our statistical method is used to downscale near-surface wind components to weather surface stations in southern France in the north western Mediterranean basin. While the Mediterranean climate system is relatively isolated during summer with very local breeze-like wind system (e.g., Drobinski et al. 2006, 2007), teleconnexions, i.e. long-range interactions, are more important during other periods relating more tightly large-scale circulation to local surface wind (e.g., Dünkeloh and Jacobeit 2003). In particular, the southern France region features a complex coast shape, high orography (the Alps, the Massif Central and the Pyrénées culminating at 4807 m, 1885 m and 3298 m, respectively; see Fig. 1) which play a crucial role in steering the large-scale air flow so that energetic meso-scale features are present in the atmospheric circulation which can evolve to high-impact weather systems during fall and winter such as wind storms over very urbanized littorals. The ability to model such dramatic events remains weak because of the contribution of very fine-scale processes and their non-linear interactions with the larger scale processes. Offshore wind storms like the Mistral and its companion wind, the Tramontane are frequent (5 to 15 days per month) and develop along the Rhône and Aude valleys (Fig. 1) (e.g., Drobinski et al. 2005; Guénard et al. 2005, 2006; Salameh et al. 2007). These wind storms can cause severe damage to farm plantations, hazardous conditions for aeronautics and ship. Onshore winds are also frequent during fall season causing frequent intense precipitations and flash-flooding in the Cévennes region (e.g., Ducrocq et al. 2002).

Considering the connexion between large-scale patterns and the Mediterranean circulation during the fall-winter seasons, the aims of the present study are thus (1) to identify between November and March, possible winter relationships between dominant fine-scale patterns of near-surface winds and large-scale weather regimes characterizing the typical large-scale atmospheric circulation in the northern and meridional European area and (2) to model the distribution of near surface wind components over complex terrain in southern France, using a statistical approach. Our statistical approach is based on the physical explanation of the wind circulations and of the choice of large scale explanatory variables.

Following this introduction in Sect. 1, possible links between weather regimes and near surface wind fields are investigated in Sect. 2. Section 3 details the statistical method used to downscale the two components of the wind at different weather stations in southern France, and the results are discussed. In Sect. 4, we conclude and suggest some future works.

2. Relationships between weather regimes and near surface wind fields

In order to downscale wind, we check the link between persistent large scale circulations over the North-Atlantic region and local near surface circulations. Indeed, it has been shown in the literature a strong connexion between the regional Mediterranean circulation and the so-called North Atlantic weather regimes which are persistent large-scale atmospheric flow regimes (e.g., Plaut and Simonnet 2001; Simonnet and Plaut 2001; Dünkeloh and Jacobeit 2003). The concept of weather regimes has been developed for obtaining high-resolution climate from GCMs by deriving statistical relationships between observed small-scale variables and larger (GCM) scale variables (e.g., Plaut and Simonnet 2001; Casola and Wallace 2007). The derivation of weather regimes requires clustering techniques consisting in either defining states of the atmosphere with the highest probability of occurrence (e.g., Mo and Ghil 1988; Molteni et al. 1990; Smith et al. 1998), or in searching for patterns associated with geopotential anomalies, that persist more than a given number of days (Mo and Ghil 1988), or that are quasi stationary (Michelangeli et al. 1995). Here, the North-Atlantic weather regimes are estimated by classifying daily geopotential height anomalies at 500 hPa (Z500) from ERA40 re-analyses (horizontal resolution 1.125°) for fall and winter data (November through March), from 1991 to 2001, over the domain ranging between $-60^\circ/30^\circ$ E and $30^\circ/70^\circ$ N (Vautard 1990). Before applying our clustering algorithm, we perform an empirical orthogonal function (EOF) analysis in order to reduce the number of spatial variables (projection on the first 10 eigenvectors explaining 85% of the variability). We apply the k-means algorithm to the corresponding principal components (Michelangeli et al. 1995).

We emphasize that our goal is not to explore new clustering methods or circulation structures but rather to base our study on well-established weather regimes. The clustering step is here limited to the classical Z500 variable over the North Atlantic region. Other variables like Z700 (geopotential height anomalies at 700 hPa) or mean sea level pressure, have not been retained for this study. Indeed, Z500 is a classical variable used for determining North Atlantic recurrent weather regimes influencing Western Europe and Mediterranean climates (e.g., Cheng and Wallace 1991; Michelangeli et al. 1995; Plaut and Simonet 2001). Moreover, Vrac et al. (2007a) and Simonnet and Plaut (2001) showed that weather regimes obtained from mean sea level

pressure and Z700 display patterns similar to those obtained from Z500. They also remarked that surface pressure data present too much variability to be correctly captured by a small number of regimes. From our clustering algorithm, we obtain the classical four North Atlantic circulation regimes (e.g., Vautard 1990; Michelangeli et al. 1995; Plaut and Simonet 2001; Yiou and Nogaj 2004) shown in Fig. 2: the “Atlantic Ridge” (AR) with a north-westerly flux over France and a depression over Greenland, the Scandinavian “Blocking” (BL) with a northerly flux over central and southern France and a depression on the east of Greenland and the two phases of the North Atlantic Oscillation NAO+ and NAO− corresponding respectively to a north-westerly to

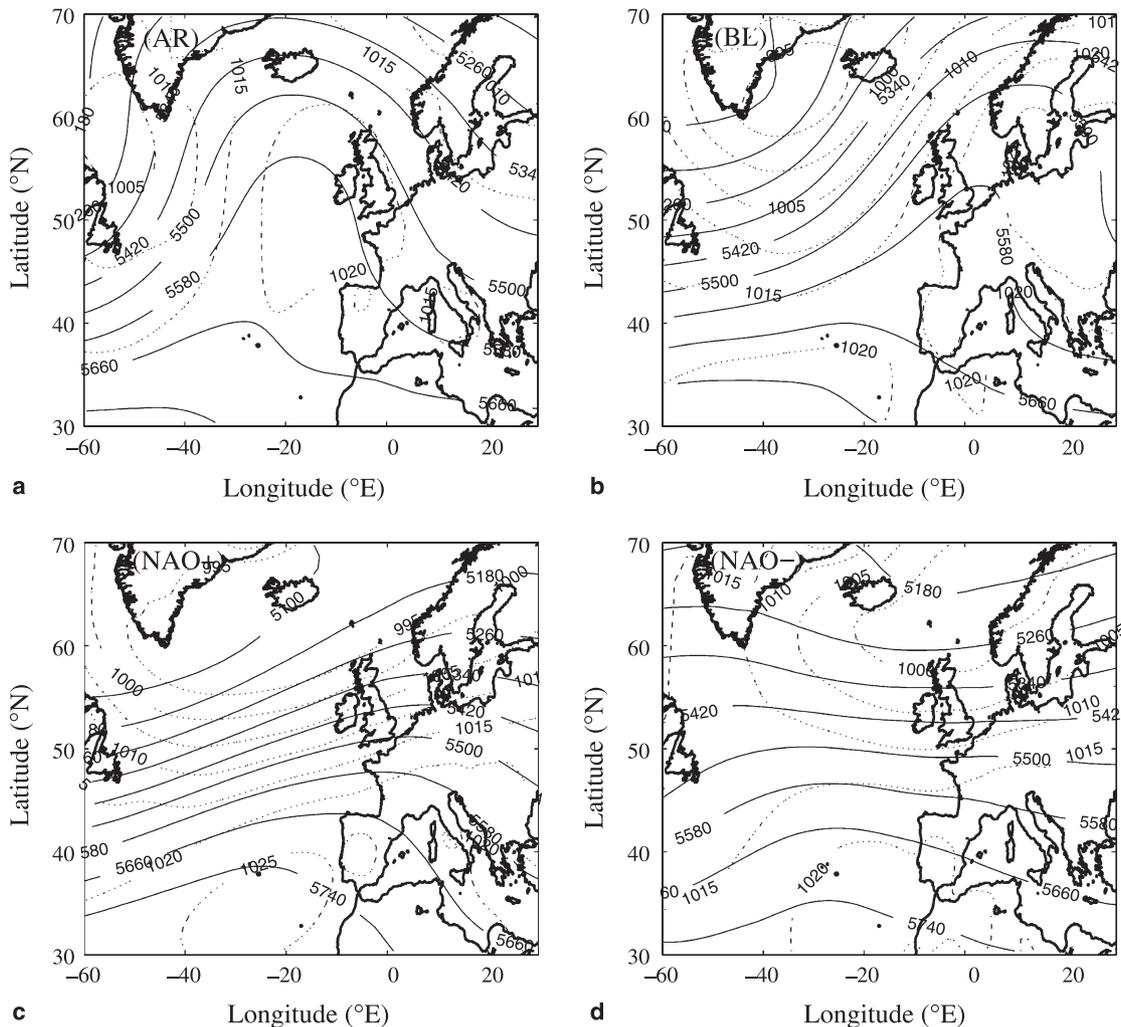


Fig. 2. Weather regimes (full lines) represented by the mean geopotential heights at 500 hPa, taken from ERA-40 re-analyses from 1991 to 2001. The weather regimes found are: (a) the Atlantic Ridge (AR), (b) the Blocking (BL), (c) the NAO+, and (d) the NAO−. Dotted lines represent the mean sea level pressure corresponding to each regime

Statistical downscaling of near-surface wind over complex terrain in southern France

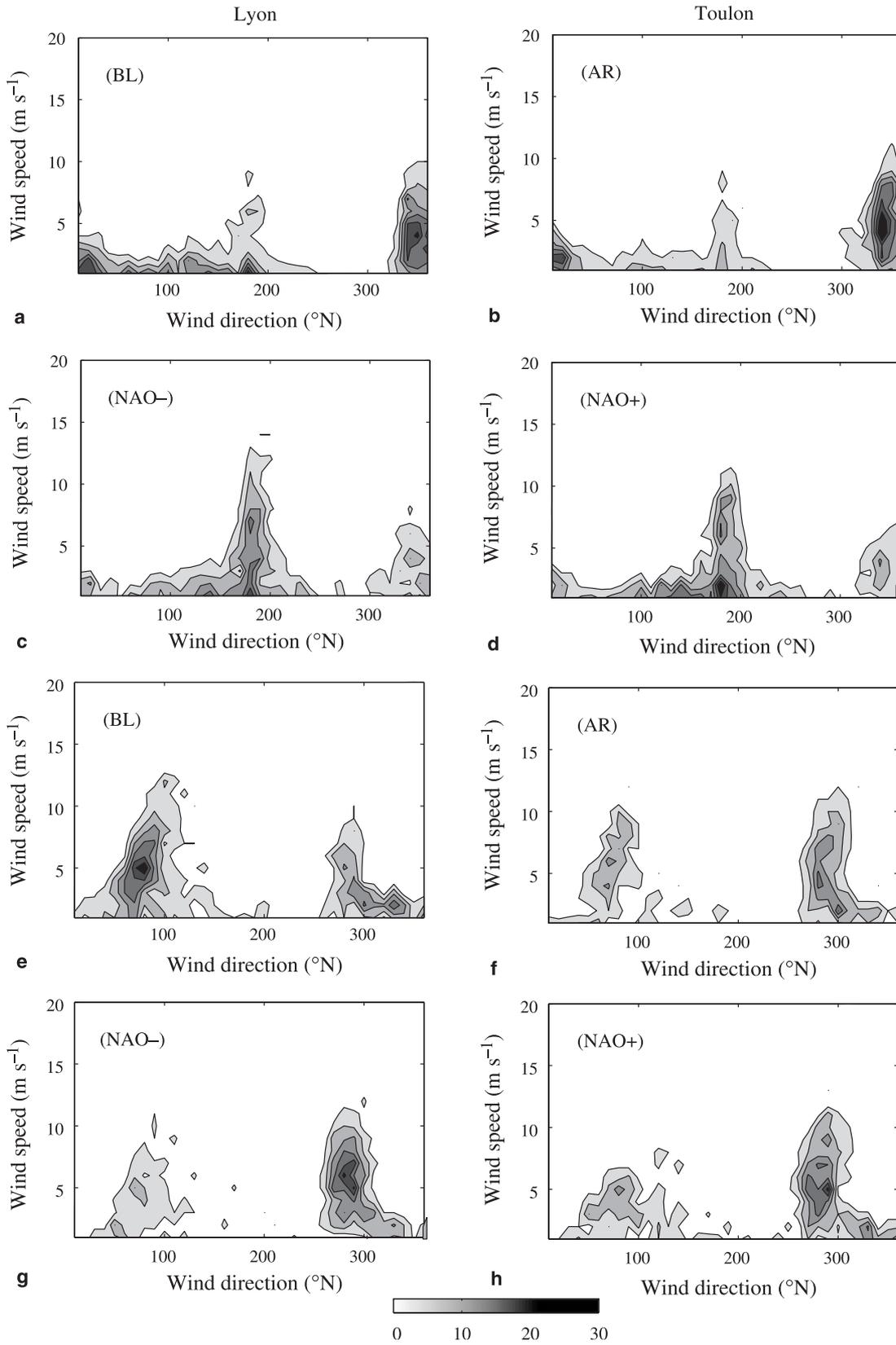


Fig. 3. Frequency of occurrence [%] of wind speed versus wind direction as a function of the weather regime at Lyon (left column) and Toulon (right column). Panels (a) and (b, c) and (d, e) and (f) and (g and h) correspond to weather regimes AR, BL, NAO+ and NAO-, respectively. The darker the colors, the higher the occurrence

westerly flux over France and a westerly to north-westerly flux over France.

At fine scale, the 10 m wind data available from the Météo-France surface stations for the last three decades are used to derive the fine-scale patterns of near-surface winds. Figure 1 displays the locations of the meteorological surface stations in southern France. Figure 3 shows the frequency of occurrence of wind speeds versus wind directions corresponding to each weather regime, at the stations of Lyon and Toulon, located at the Rhône valley entrance and exit, respectively. For Lyon, it shows a large occurrence of strong north-westerly winds for regimes 1 (AR) and 2 (BL) (maximum around 10 m s^{-1} and a mean around 4.4 m s^{-1}), and strong south-westerly winds for regimes 3 (NAO+) and 4 (NAO-) (maximum around 12 m s^{-1} and a mean around 5.5 m s^{-1}). For all regimes we however have bimodal distributions with the strong winds peaking around the south (between 160° and 230°) and north (between 310° and 40°) directions. This allows to concluding that local circulations do not derive univocally from large-scale weather regimes (and conversely). Indeed, local circulations are mainly controlled by the orography (Drobinski et al. 2003) since the main wind directions are north and south which correspond to the valley axis direction in the area of Lyon. The occurrences of northerly and southerly winds represent 87% of the total wind observations at Lyon. Similar bimodal distributions are found at all the meteorological surface stations (e.g., wind data at Toulon in Fig. 3, right panels). The classical North-Atlantic weather regimes are thus not relevant for surface wind downscaling.

3. Statistical downscaling of near-surface winds using a generalized additive model

Our goal is thus to develop a statistical model capable of explaining the values of the surface wind components (u , v) called explained variables – i.e. values to be explained – provided by Météo-France. The explanatory variables X_j – i.e. the variables used to explain u and v – are variables from ERA-40 re-analyses. The statistical approach retained is a Generalized Additive Model (GAM) that models the explained variables u and v as a sum of spline functions applied

to different explanatory variables X_j (Hastie and Tibshirani 1990):

$$u_i = \sum_{j=1}^p f_{i,j}^u(X_j) + \varepsilon_{u_i}, \quad (1a)$$

$$v_i = \sum_{j=1}^p f_{i,j}^v(X_j) + \varepsilon_{v_i}, \quad (1b)$$

where i corresponds to the station indice, j indicates an explanatory variable, p is the number of explanatory variables, u_i and v_i are the wind components at station i , $f_{i,j}^u$ and $f_{i,j}^v$ are the spline functions ε_{u_i} and ε_{v_i} are the errors for u_i and v_i , respectively. This model is applied to each station separately and corresponds to a nonlinear regression between large and small scale. Splines are piecewise parametrical or non-parametrical functions. That means that for each piece, a function (of a given form) is fitted. For example, if the chosen form is a second order polynomial function and the number of pieces (or intervals) selected is three, the associated spline function corresponds to three second-order polynomial functions. Consequently, in the present context, the regressions in Eq. (1) are non-linear. In this work, the chosen splines are piecewise third order polynomial functions (Hastie and Tibshirani 1990). Conditionally on the distribution family of the predictands u and v , the statistical theory imposes some constraints on the distribution family of the model errors ε_{u_i} and ε_{v_i} (see Hastie and Tibshirani 1990). In this study, u and v are assumed to be Gaussian so this implies that ε_{u_i} and ε_{v_i} have a zero-mean normal distribution.

We first apply this model to the explanatory variables used to downscale wind speed similar to those used in Pryor et al. (2005), namely the surface pressure gradients along the Rhône valley axis, between Lyon and Toulon (see Kossmann and Sturman 2002) and along the French Riviera between Nice and Toulon, and the relative vorticity at 500 hPa. This set of large-scale explanatory variables is referred to as set (A). Note that because we apply the GAM model on raw six-hourly wind measurements, we do not use the variance of the 500 hPa relative vorticity as done in Pryor et al. (2005). The best explanation of u_i and v_i consists in determining the best combination of large scale explanatory variables. Table 1 shows the wind component variance explained

Table 1. Explained variance of near surface wind components by the GAM with respect to the measurements collected at different Météo-France weather stations for three sets of large-scale explanatory variables from ERA-40 reanalysis: set (A) includes surface pressure gradient and 500 hPa relative vorticity (similar to Pryor et al. 2005); set (B) includes near-surface pressure gradient, low-level winds at 925 and 850 hPa and geostrophic wind at 700 hPa; set (C) is the optimized set of explanatory variables (i.e. giving the largest explained variance). It includes set (B) and relative vorticity and geopotential at low levels. For each weather station (left column), the first, second and third rows correspond to results for raw six-hourly data, daily and weekly averaged data

Explained variance (%)	Sets of large-scale explanatory variables					
	Set (A)		Set (B)		Set (C)	
	u	v	u	v	u	v
Cannes	58.12	27.82	60.12	36.33	61.54	38.34
	73.72	34.52	76.62	47.23	77.72	49.29
	75.81	45.41	84.12	52.71	89.72	58.09
Lyon	19.29	83.10	25.93	85.25	28.78	85.89
	30.85	88.16	39.32	90.54	46.13	91.37
	25.55	88.22	52.00	90.69	66.17	90.68
Nice	19.15	12.51	22.41	21.79	24.13	24.38
	35.51	18.57	40.24	28.91	43.42	34.61
	34.38	21.86	57.74	38.79	65.46	55.86
Nimes	42.24	67.20	53.02	70.84	55.26	71.46
	57.65	77.88	69.31	82.59	71.40	83.33
	64.68	86.20	77.79	88.92	86.07	90.60
Orange	51.82	78.98	58.51	80.37	60.10	81.12
	60.63	87.75	69.44	89.31	70.99	90.04
	60.84	88.74	64.63	90.03	68.78	92.61
Toulon	80.01	33.82	81.15	38.89	82.01	41.81
	88.55	51.37	89.90	58.76	90.84	62.93
	91.42	53.69	92.68	63.35	95.28	76.65

by GAM. The explained wind components are six-hourly (first row), daily and weekly averaged (second and third rows, respectively). The analysis of different time frames aims at evaluating the sensitivity of the statistical regression coefficients to time averaging, or in other words, at checking whether one single statistical regression can be used when predicting instantaneous surface wind, daily, weekly or monthly averaged wind components. Only six-hourly data, daily and weekly averaged data are shown and discussed in the paper.

One can first notice that generally only one component is accurately modeled, while the explained variance is much lower for the other component. This is due to the fact that in this region, the wind blows along the flanks of the mountains and along the valley axes (Aude, Rhône, Durance valleys). The wind component

perpendicular to the valley and mountain side-walls is generally much weaker than the along-valley wind component, so the unexplained variance is due to small scale turbulence or to very localized source of perturbation, which is not explained by large-scale variables. So for the three weather stations located in the Rhône valley (Lyon, Nimes, and Orange), the meridional wind component (v) is more accurately explained than the zonal wind component (u). Indeed, for these three stations, the explained variance for the six-hourly wind ranges between 67.2% (Nimes) and 83.1% (Lyon) for the v -component and 19.3% (Lyon) and 51.8% (Orange) for the u -component (similar values are found for the other weather stations in the Rhône valley displayed in Fig. 1, not shown). Since the Rhône valley is much narrower at Lyon (about 80 km width) than at Orange (about 120 km width) or Nimes (about

250 km width), the u -component is very weak at Lyon (the wind direction distribution is peaked around 170 and 350° , see Fig. 3) and thus most of its variability is due to local turbulence. For Orange and Nimes, the valley becomes a delta and the wind direction distribution is broader, explaining the increase of the explained variance on the u -component also. Similarly, for the weather stations located along the French Riviera (Cannes, Nice, Toulon), the u -component is more accurately explained than the v -component. The explained variance for the six-hourly wind ranges between 19.1% (Nice) and 80.0% (Toulon) for the u -component and 12.5% (Nice)

and 33.8% (Toulon) for the v -component (similar values are found for the other weather stations along the French Riviera displayed in Fig. 1, not shown). The very poor explanation of the near-surface wind in Nice can be attributed to either the set of explanatory variables which may not be appropriate to this area or to the domination of local elements of wind perturbations. Figure 4 shows the splines for u and v at Lyon. The spline functions are nearly flat for the 500 hPa relative vorticity suggesting a full decoupling between the near-surface circulation and the atmospheric circulation at 500 hPa (indeed, replacing the relative vorticity at 500 hPa by oth-

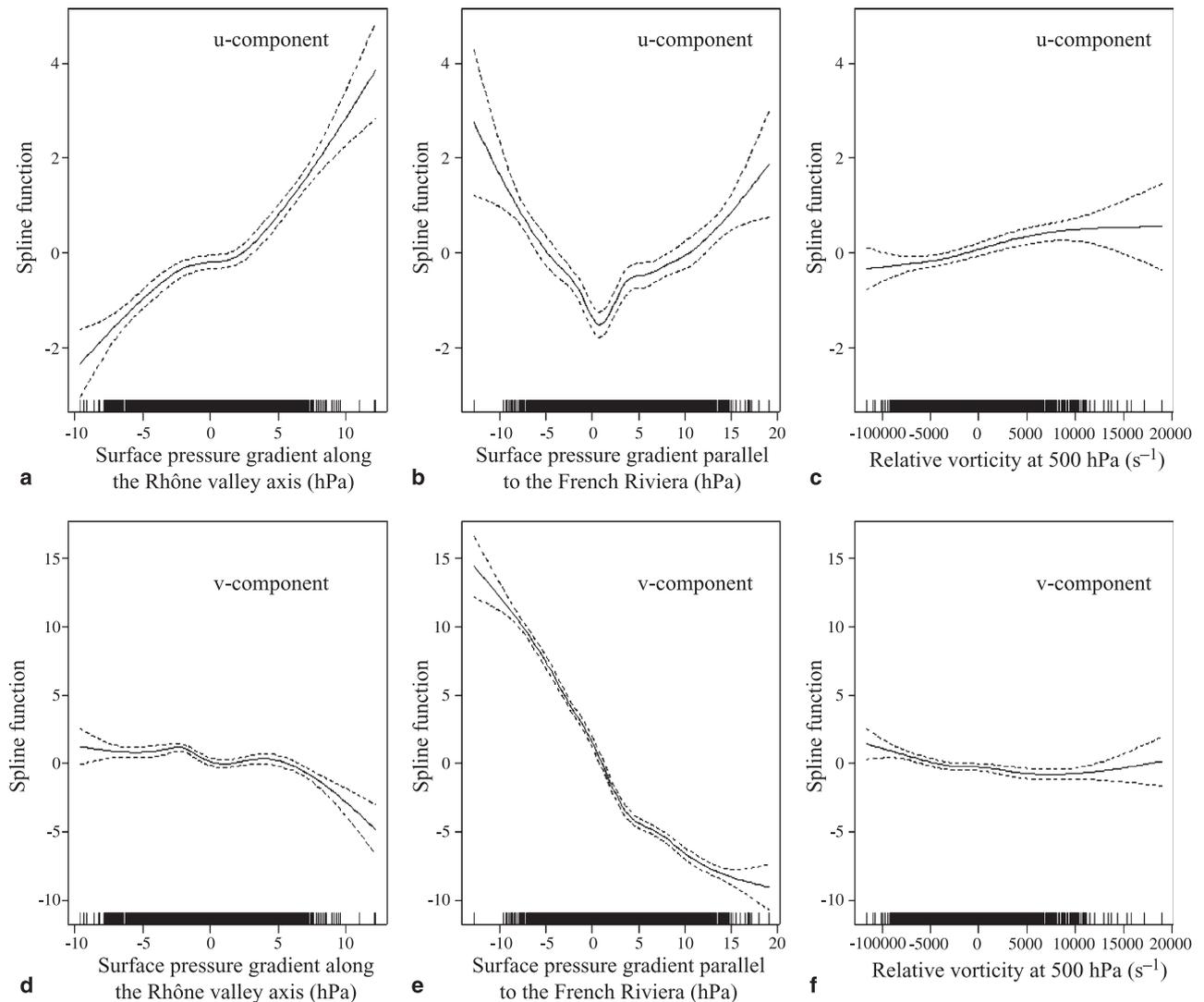


Fig. 4. Spline functions for u and v wind components at Lyon corresponding to set (A), i.e., the pressure gradients along the Rhône valley (a) and (d) and along the French Riviera (b) and (e) and the relative vorticity at 500 Pa (c) and (f), respectively

er variables at that pressure level, such as geopotential, wind, . . . does not improve the modeling). In Pryor et al. (2005), the less complex nature of the terrain allows the 500 hPa vorticity to be a better explanatory variable. Another feature to be underlined in Fig. 4 is the piecewise linear shape of the spline functions relating the wind components to the surface pressure gradient. This is an important feature as will be discussed hereafter.

We thus try to find the best set of explanatory variables for the near-surface wind components in such a complex environment. To do so, we consider the steady equation for u and v (the downscaling technique does not account for the dynamical evolution of the near-surface wind):

$$u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} + w \frac{\partial u}{\partial z} - fv = -\frac{1}{\rho} \frac{\partial p}{\partial x} + F_u, \quad (2a)$$

$$u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} + w \frac{\partial v}{\partial z} + fu = -\frac{1}{\rho} \frac{\partial p}{\partial y} + F_v, \quad (2b)$$

where u , v are the wind velocity components along x (west to east axis) and y (south to north axis), respectively, p is the atmospheric pressure, ρ is the air density, F_u and F_v are the turbulence induced friction terms and f is the Coriolis parameter. The pressure term is the sum of the synoptic pressure gradient, the orography-induced pressure perturbation at the scale of ERA-40 resolution caused by mass adjustment (due to lateral constriction in valleys), the mountain slopes, and the small scale pressure perturbation caused by orographic effects at smaller scale or very localized effects (e.g., thermal gradient due to inhomogeneous land-use, . . .). Splitting the variables in a large scale component (subscript ls) given by the ERA-40 reanalyses and a small-scale component (subscript ss) to be explained by the downscaling technique, the equations of motion are thus:

$$\begin{aligned} & (u_{ls} + u_{ss}) \frac{\partial(u_{ls} + u_{ss})}{\partial x} + (v_{ls} + v_{ss}) \frac{\partial(u_{ls} + u_{ss})}{\partial y} \\ & + w_{ls} \frac{\partial(u_{ls} + u_{ss})}{\partial z} - f(v_{ls} + v_{ss}) \\ & = -\frac{1}{\rho} \frac{\partial p_{ls}}{\partial x} - \underbrace{w_{ss} \frac{\partial(u_{ls} + u_{ss})}{\partial z} - \frac{1}{\rho} \frac{\partial p_{ss}}{\partial x}}_{\text{source of unpredictability}} + F_u \end{aligned} \quad (3a)$$

$$\begin{aligned} & (u_{ls} + u_{ss}) \frac{\partial(v_{ls} + v_{ss})}{\partial x} + (v_{ls} + v_{ss}) \frac{\partial(v_{ls} + v_{ss})}{\partial y} \\ & + w_{ls} \frac{\partial(v_{ls} + v_{ss})}{\partial z} + f(u_{ls} + u_{ss}) \\ & = -\frac{1}{\rho} \frac{\partial p_{ls}}{\partial y} - \underbrace{w_{ss} \frac{\partial(v_{ls} + v_{ss})}{\partial z} - \frac{1}{\rho} \frac{\partial p_{ss}}{\partial y}}_{\text{source of unpredictability}} + F_v \end{aligned} \quad (3b)$$

From Eq. (3), one can see that u_{ss} is a function of u_{ls} , v_{ls} , $\partial u_{ls}/\partial x$, $\partial u_{ls}/\partial y$, $\partial u_{ls}/\partial z$ and $\partial p_{ls}/\partial x$ and v_{ss} is a function of u_{ls} , v_{ls} , $\partial v_{ls}/\partial x$, $\partial v_{ls}/\partial y$, $\partial v_{ls}/\partial z$ and $\partial p_{ls}/\partial y$ with a source of unpredictability due to small-scale turbulence (random properties) and local sources of wind perturbations (i.e., small-scale pressure perturbations which may not have random properties). We thus find that large scale surface wind u_{ls} (here taken at 925, 850, and 700 hPa) and surface pressure gradients are possibly relevant explanatory variables for the small scale component. This set of explanatory variables is referred to as set (B). Set (C) finally includes the geopotential and relative vorticity (in order to account for large-scale wind gradient) at 500 hPa. Set (C) proved to be the best set of explanatory variables.

Table 1 summarizes the explained variance for sets (B) and (C) and shows a significant improvement. The physical dependence between the near-surface wind components and the large-scale explanatory variables of set (B) is outlined in Eq. (3) and makes easier the interpretation of the explained variance. The addition of the low-level relative vorticity and geopotential brings valuable even at 500 hPa. In brief, for the weather stations located in the Rhône valley, the explained variance for the six-hourly wind ranges between 70.8% (set B) and 71.5% (set C) (Nimes) and 85.2% (set B) and 85.9% (set C) (Lyon) for the v -component and 25.9% (set B) and 28.8% (set C) (Lyon) and 58.5% (set B) and 60.1% (set C) (Orange) for the u -component. For the weather stations located along the French Riviera, the explained variance for the six-hourly wind ranges between 22.4% (set B) and 24.1% (set C) (Nice) and 81.1% (set B) and 82.0% (set C) (Toulon) for the u -component and 21.8% (set B) and 24.4% (set C) (Nice) and 38.9% (set B) and 41.8% (set C) (Toulon) for the v -component. The still low performance of GAM in Nice is

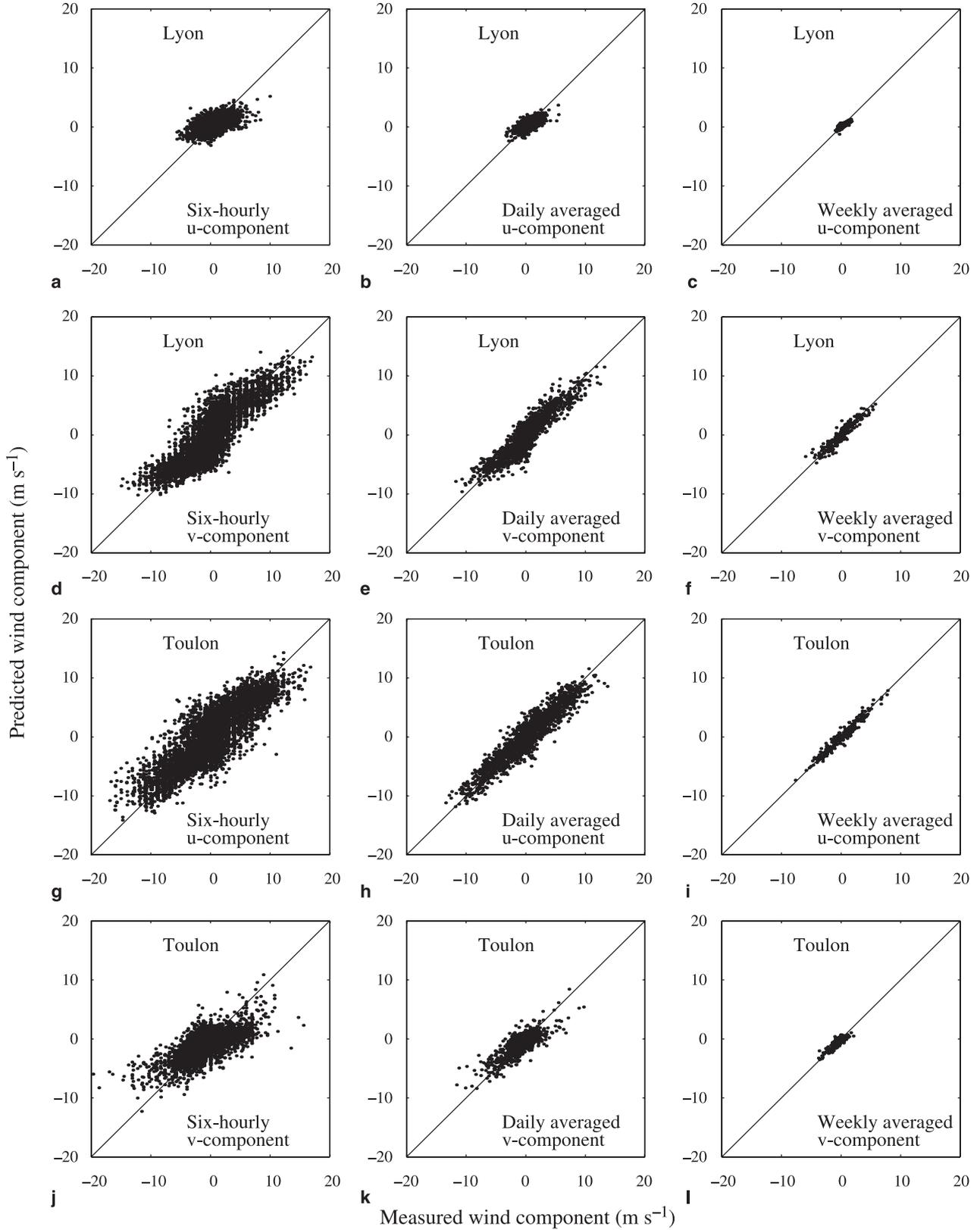


Fig. 5. Scatter plots of the six-hourly data (first column), daily (second column) and weekly (third column) averaged explained u - (first and third rows) and v - (second and fourth rows) wind components at Lyon (first two rows) and at Toulon (last two rows) as a function of the corresponding measured wind components for set (C)

probably due to the location of the weather station nearby a local source of wind perturbations. One key element is that, as for set (A), all spline functions display a piecewise linear trend (not shown).

Part of the unexplained wind is due to small scale turbulence. We thus apply GAM on daily and weekly averaged wind components and large-scale explanatory variables, since this random contribution to the unpredictability of GAM is to be reduced by averaging. The impact of the averaging procedure must however be analyzed with great care. Let us denote by a bar the time averaging operator. We here apply GAM on the time-averaged large-scale explanatory variable and compute spline functions g on the time-averaged near-surface wind components measured at the weather stations. We thus compute:

$$\bar{u}_i = \sum_{j=1}^p g_{i,j}^u(\bar{X}_j) + \varepsilon_{u_i}, \quad (4a)$$

$$\bar{v}_i = \sum_{j=1}^p g_{i,j}^v(\bar{X}_j) + \varepsilon_{v_i}. \quad (4a)$$

If we apply the time-averaging operator on Eq. (1), we have:

$$\bar{u}_i = \sum_{j=1}^p \overline{f_{i,j}^u(X_j)} + \bar{\varepsilon}_{u_i}, \quad (5a)$$

$$\bar{v}_i = \sum_{j=1}^p \overline{f_{i,j}^v(X_j)} + \bar{\varepsilon}_{v_i}. \quad (5b)$$

Since the spline functions are all quasi-linear or piecewise linear, we can write

$$\overline{f_{i,j}^u(X_j)} \approx f_{i,j}^u(\bar{X}_j), \quad (6a)$$

$$\overline{f_{i,j}^v(X_j)} \approx f_{i,j}^v(\bar{X}_j), \quad (6b)$$

so $g_{i,j}^u \approx f_{i,j}^u$ and $g_{i,j}^v \approx f_{i,j}^v$ which means that the spline functions estimated at each weather station with six-hourly measurements are robust to time averaging thus explaining why we obtain very similar spline functions in the absence of averaging or after averaging (not shown). In detail, the explained variance significantly improves with averaging and can reach 170% of relative increase for the lowest explained variance of the six-hourly wind components (e.g., the two wind components at Nice) (Table 1). The lowest relative increase is 6% for the already high explained

variance of six-hourly wind components (e.g., the v -component at Lyon).

Figure 5 displays a scatter plot of the six-hourly, daily and weekly averaged explained wind components at Lyon (first two rows) and at Toulon (last two rows) as a function of the corresponding measured wind components. Whatever the time averaging, the explained wind components are unbiased. Since GAM optimizes the estimation of the expectation of the wind component distributions, it underestimates the extreme winds (especially visible for the six-hourly values). Time averaging significantly contributes to reduce the scatter.

4. Conclusions

The objective of this paper was to downscale the surface wind components at weather stations from synoptic scale to the local-scale, in southern France, a coastal region characterized by high orography and very urbanized littorals.

As the Mediterranean climate is known to be tightly related to the North-Atlantic weather regimes, a first approach was to check the link between the recurrent large-scale circulation patterns and the measured surface wind. The main achievements of this paper can then be summarized as follows:

- (1) there is no discriminating power of the classical North-Atlantic weather regimes to downscale surface wind in a complex region such as southern France.
- (2) using generalized additive models which generally apply on Gaussian variables, accurate statistical downscaling of the surface wind components is achieved at nearly all investigated weather stations even when located at very complex locations (in most cases, the explained variance of the dominant wind component exceed about 60%). The choice of the explanatory large-scale variables is justified physically and the impact of these choices on the downscaling technique performance is discussed.
- (3) whatever the set of large-scale explanatory variables, the coefficients of the statistical regression are not sensitive to time averaging so the same statistical regression can be used when predicting instantaneous surface wind components or averaged wind components.

This method thus allows to go one step further in wind speed downscaling since it is adapted to model the wind components and not only wind speed and energy, it is suited for complex terrain and robust to time averaging in this region.

However, GAM optimizes the estimation of the expectation of the wind component distributions and is thus not suited for extreme event modeling which are thus underestimated. Future work will be dedicated to improve extreme event modeling, to adapt GAM for vectorial variables in order to explain the two wind-components together: at present the two components are modeled separately; in the future, we should model the components altogether conditioned to the other.

Acknowledgments

The authors thank the two anonymous referees and the editor for their contribution to the significant improvement of the quality of the paper. This research has received funding from the Agence de l'Environnement et de la Maîtrise de l'Energie (ADEME) and support from Météo-France that provided the data from the operational meteorological surface stations. This work was also supported by the European E2-C2 grant, the National Science Foundation (grant: NSF-GMC (ATM-0327936)), by The Weather and Climate Impact Assessment Science Initiative at the National Center for Atmospheric Research (NCAR) and the ANR-AssimilEx project. The authors would also like to credit the contributors of the R project.

References

Casola HJ, Wallace MJ (2007) Identifying weather regimes in the wintertime 500-hPa geopotential height field for the Pacific-North American sector using a limited contour clustering technique. *J Appl Meteor Clim* 46: 1619–30

Cheng X, Wallace J (1991) Cluster analysis of the northern hemisphere wintertime 500-hPa height field: spatial patterns. *J Atmos Sci* 50: 2674–96

De Rooy WC, Kok K (2004) A combined physical-statistical approach for the downscaling of model wind speeds. *Wea Forecast* 19: 485–95

Drobinski P, Dabas A, Haerberli C, Flamant P (2003) Statistical characterization of the flow structure in the Rhine valley. *Bound Layer Meteor* 106: 483–505

Drobinski P, Bastin S, Guénard V, Caccia JL, Dabas AM, Delville P, Protat A, Reitebuch O, Werner C (2005) Summer Mistral at the exit of the Rhône valley. *Quart J Roy Meteorol Soc* 131: 353–75

Drobinski P, Bastin S, Dabas AM, Delville P, Reitebuch O (2006) Variability of the three-dimensional sea-breeze structure in southern France: observations and evaluation of empirical scaling laws. *Ann Geophys* 24: 1783–99

Drobinski P, Saïd F, Ancellet G, Arteta J, Augustin P, Bastin S, Brut A, Caccia JL, Campistron B, Cautenet S, Colette A, Cros B, Corsmeier U, Coll I, Dabas A, Delbarre H, Dufour A, Durand P, Guénard V, Hasel M, Kalthoff N, Kottmeier C, Laskri F, Lemonsu A, Lohou F, Masson V, Menut L, Moppert C, Peuch VH, Puygrenier V, Reitebuch O, Vautard R (2007) Regional transport and dilution during high pollution episodes in southern France: summary of findings from the ESCOMPTE experiment. *J Geophys Res* 112: D13105; DOI: 10.1029/2006JD007494

Ducrocq V, Ricard D, Lafore JP, Orain F (2002) Storm-scale numerical rainfall prediction for five precipitating events over France: on the importance of the initial humidity field. *Wea Forecast* 17: 1236–56

Dünkeloh A, Jacobeit J (2003) Circulation dynamics of Mediterranean precipitation variability 1948–1998. *Int J Climatol* 23: 1843–66

Frech M, Holzäpfel F, Tafferner A, Gerz T (2007) High-resolution weather database for the terminal area of Frankfurt airport. *J Appl Meteor Clim* 46: 1913–32

Guénard V, Drobinski P, Caccia JL, Campistron B, Bénéch B (2005) An observational study of the mesoscale mistral dynamics. *Bound Layer Meteor* 115: 263–88

Guénard V, Drobinski P, Caccia JL, Tedeschi G, Currier P (2006) Dynamics of the MAP IOP-15 severe Mistral event: observations and high-resolution numerical simulations. *Quart J Roy Meteorol Soc* 132: 757–78

Gustafson WI Jr, Leung LR (2007) Regional downscaling for air quality assessment. A reasonable proposition? *Bull Am Meteor Soc* 88: 1215–27

Gutiérrez JM, Cofiño AS, Cano R, Rodríguez MA (2004) Clustering methods for statistical downscaling in short-range weather forecasts. *Mon Wea Rev* 132: 2169–83

Hastie T, Tibshirani R (1990) *Generalized additive models*. Chapman and Hall, London

Karl TR, Wang WC, Schlesinger M, Knight RW, Portman D (1990) A method of relating general circulation model simulated climate to the observed local climate. Part 1: Seasonal statistics. *J Climate* 3: 1053–79

Kossmann M, Sturman AP (2002) Pressure driven channeling effects in bent valleys. *J Appl Meteor* 37: 151–8

Mass FC, Ovens D, Westrick K, Colle BA (2002) Does increasing horizontal resolution produce more skillful forecast? *Bull Am Meteor Soc* 83: 407–30

Michelangeli PA, Vautard R, Legras B (1995) Weather regimes: recurrence and quasi-stationarity. *J Atmos Sci* 52: 1237–56

Mo KC, Ghil M (1988) Cluster analysis of multiple planetary flow regimes. *J Geophys Res* 93: 10927–51

Molteni F, Tibaldi S, Palmer TN (1990) Regimes in the wintertime circulation over the northern extratropics. I: Observational evidence. *Quart J Roy Meteorol Soc* 116: 31–67

Plaut G, Simonnet S (2001) Large scale circulation classification, weather regimes, and local climate over France, the Alps and Western Europe. *Clim Res* 17: 303–24

Pryor SC, Schoof JT, Barthelmie RJ (2005) Empirical downscaling of wind speed probability distributions. *J Geophys Res* 110: D19109; DOI: 10.1029/2005JD005899

Statistical downscaling of near-surface wind over complex terrain in southern France

- Salameh T, Drobinski P, Menut L, Bessagnet B, Flamant C, Hodzic A, Vautard R (2007) Aerosol distribution over the western Mediterranean basin during a Tramontane/Mistral event. *Ann Geophys* 11: 2271–91
- Simonnet E, Plaut G (2001) Space-time analysis of geopotential height and SLP, intraseasonal oscillations, weather regimes, and local climates over the North Atlantic and Europe. *Clim Res* 17: 325–42
- Smith P, Ide K, Ghil M (1998) Multiple regimes in northern hemisphere height fields via mixture model clustering. *J Atmos Sci* 56: 3704–23
- Vautard R (1990) Multiple weather regimes over the North Atlantic: analysis of precursors and successors. *Mon Wea Rev* 118: 2056–81
- Vrac M, Hayhoe K, Stein M (2007a) Identification and intermodel comparison of seasonal circulation patterns over North America. *Int J Climatol* 27(5): 603–20
- Vrac M, Marbaix P, Paillard D, Naveau P (2007b) Non-linear statistical downscaling of present and LGM precipitation and temperatures over Europe. *Clim Past Disc* 3: 899–933
- Wilby RL, Wigley TML (1997) Downscaling general circulation model output: a review of methods and limitations. *Prog Phys Geogr* 21: 530–48
- Wilby RL, Wigley TML, Conway D, Jones PD, Hewitson BC, Main J, Wilks DS (1998) Statistical downscaling of general circulation model output: a comparison of methods. *Water Resour Res* 34: 2995–3008
- Yiou P, Nogaj M (2004) Extreme climatic events and weather regimes over the North Atlantic: when and where? *Geophys Res Lett* 31: L07202; DOI: 10.1029/2003GL019119
- Zagar N, Zagar M, Cedilnik J, Gregoric G, Rakovec J (2006) Validation of mesoscale low-level winds obtained by dynamical downscaling of ERA40 over complex terrain. *Tellus* 58A: 445–55