Spatial Predictions of Extreme Wind Speeds over Switzerland Using Generalized Additive Models

ETIENNE, Christophe, et al.

Abstract

The purpose of this work is to present a methodology aimed at predicting extreme wind speeds over Switzerland. Generalized additive models are used to regionalize wind statistics for Swiss weather stations using a number of variables that describe the main physiographical features of the country. This procedure enables one to present the results for Switzerland in the form of a map that provides the 98th percentiles of daily maximum wind speeds (W98) at a 10-m anemometer height for cells with a 50-m grid interval. This investigation comprises three major steps. First, meteorological data recorded by the weather stations was gathered to build local wind statistics at each station. Then, data describing the topographic and landscape characteristics of the country were prepared using geographic information systems (GIS). Third, appropriate regression models were selected to make spatially explicit predictions of extreme wind speeds in Switzerland. The predictions undertaken in this study provide realistic values of the W98. The effects of topography on the results are particularly conspicuous. Wind speeds increase with [...]
Spatial Predictions of Extreme Wind Speeds over Switzerland Using Generalized Additive Models

CHRISTOPHE ETIENNE, ANTHONY LEHMANN, AND STÉPHANE GOYETTE
Climatic Change and Climate Impacts Group, Institute for Environmental Sciences, University of Geneva, Geneva, Switzerland

JUAN-IGNACIO LOPEZ-MORENO
Instituto Pirenaico de Ecología, CSIC, Campus de Aula Dei, Zaragoza, Spain

MARTIN BENISTON
Climatic Change and Climate Impacts Group, Institute for Environmental Sciences, University of Geneva, Geneva, Switzerland

(Manuscript received 3 March 2009, in final form 14 April 2010)

ABSTRACT

The purpose of this work is to present a methodology aimed at predicting extreme wind speeds over Switzerland. Generalized additive models are used to regionalize wind statistics for Swiss weather stations using a number of variables that describe the main physiographical features of the country. This procedure enables one to present the results for Switzerland in the form of a map that provides the 98th percentiles of daily maximum wind speeds (W98) at a 10-m anemometer height for cells with a 50-m grid interval. This investigation comprises three major steps. First, meteorological data recorded by the weather stations was gathered to build local wind statistics at each station. Then, data describing the topographic and landscape characteristics of the country were prepared using geographic information systems (GIS). Third, appropriate regression models were selected to make spatially explicit predictions of extreme wind speeds in Switzerland. The predictions undertaken in this study provide realistic values of the W98. The effects of topography on the results are particularly conspicuous. Wind speeds increase with altitude and are greatest on mountain peaks in the Alps, as would be intuitively expected. Relative errors between observations and model results calculated for the meteorological stations do not exceed 30%, and only 12 out of 70 stations exhibit errors that exceed 20%. The combination of GIS techniques and statistical models used to predict a highly uncertain variable, such as extreme wind speed, yields interesting results that can be extended to other fields, such as the assessment of storm damage on infrastructures.

1. Introduction

Wind storms can generate severe damage to infrastructure and therefore lead to large economic losses (SwissRe 2007; MunichRe 2008). In the past two decades, two major storms affected Switzerland. The February 1990 “Vivian” storm resulted in severe damage in the Alps (Schüepp et al. 1994) and economic losses related to infrastructure (Schraft et al. 1993). In December 1999, the “Lothar” storm caused even higher damage in France, Germany, and Switzerland (Bresch et al. 2000; MunichRe 2001), with total economic losses estimated at about US$12 billion.

Damage functions have been investigated in the past two decades to assess the damage to buildings resulting from wind storms (Heneka and Ruck 2004). Losses related to storm events can be assessed through empirical formulas that take into account variables describing the characteristics of a windstorm event, such as mean or maximum wind speeds (Dorland et al. 1999; Unanwa et al. 2000; Huang et al. 2001) or storm duration (Schraft et al. 1993). The most important parameter remains...
wind speed at the standard anemometer height of 10 m (maximum wind gusts) that is often the dominant factor leading to damage (Heneka and Ruck 2004). Gust values are frequently normalized as a function of local climate conditions, essentially taking into account relative wind speeds rather than absolute wind speeds to assess storm damage (Klawa and Ulbrich 2003; Pinto et al. 2007). The underlying hypothesis is that in wind-prone areas, buildings (and also forests) are generally well adapted to conditions where strong winds are frequent; as a result, the vulnerability of these regions to strong wind events is generally less than in other regions where these events are rare. The 98th percentile of daily maximum wind speed data (W98) is a key parameter that provides information on local wind conditions; in addition, W98 is often used as a threshold in empirical formulas that attempt to translate the damage caused by wind storms into monetary terms (Klawa and Ulbrich 2003).

In mountain areas such as the Alps, wind flows show high spatial variability. As discussed by Barry (1992) or Mortensen and Petersen (1998), topography has a strong influence on the estimation of wind speeds. Ridge crests, deep valleys, or other irregular landscapes are important orographic features capable of exerting an influence on boundary layer flows. Many local climate phenomena as well as natural channeling effects or thermally induced circulations are found within the alpine domain. Wind flows can be very strong at one location but very weak in a neighboring valley, thus exhibiting major discontinuities within small areas. These discontinuities make the spatial interpolation of wind speed values difficult (Tveito et al. 2008). Working with wind speeds within the atmospheric boundary layer always remains a challenge, since it is a highly fluctuating and nonlinear element of atmospheric flows, where the variability in both time and space can be high. Because of many turbulence effects and the presence of numerous roughness elements (Stull 1988), modeling wind and extreme events, as well as attempting to regionalize wind speed and direction remains a difficult task, even with the help of numerical models (Goyette et al. 2001). However, a number of methods have been developed helping to address the issues related to the regionalization of wind speeds (Porch and Rodriguez 1987; Palomino and Martin 1995; Nielsen 1999; Schaffner et al. 2006). Many correction factors related to topographic features, such as slope angles, altitude, or landform characteristics, have been added to the calculation of wind speeds in order to adjust the results and thus enable these to be in better agreement with observations.

There have been several studies aimed at assessing wind speeds over Switzerland (Jungo et al. 2002; Schaffner et al. 2006), including within the topographically complex Alpine domain (Schaffner et al. 2006), in other regions (Troen and Petersen 1989; Frank and Landberg 1997; Wei et al. 2006) or for specific types of surfaces (Dellwik et al. 2005; Reutter et al. 2005). There has even been an attempt at interpolating wind over the entire globe by Paulo et al. (2000). The resulting maps provide wind velocities at different altitudes (10, 50, 100 m, or more). Past studies of the spatial distribution of wind speeds have used interpolation methods with the help of correction factors related to elevation or slope values (Schaffner et al. 2006), or statistical relationships between gust factors and mean wind speeds (Jungo et al. 2002). Unlike these past studies, this paper focuses on the maximum wind speeds rather than the mean wind speeds, in order to pave the way for future work on damage costs related to severe wind storms.

The method presented in this paper discusses a new approach to regionalize wind speeds measured at weather stations distributed throughout Switzerland. It aims to test the use of spatial prediction derived from state-of-the-art statistical models in the field of meteorological applications. A regression-based model is proposed that is capable of predicting W98 values in the complex terrain characteristic of Switzerland, based on variables that describe topographic characteristics (Barry 1992; Mortensen and Petersen 1998). One of the objectives of the present work is thus to investigate to what extent wind velocities can be explained at the local scale by surrounding landscape characteristics. This research also assesses a regression model for wind analysis, assuming that the local relationship between the predicted pattern (e.g., wind speed) and its direct or neighboring environment is stronger than the relationships with measurements obtained from other, more distant weather stations.

Statistical models such as generalized additive models (GAMs; Hastie and Tibshirani 1990) have been widely used in the past for a range of scientific domains (Lehmann et al. 2002a). GAMs represent a nonparametric extension of generalized linear models (GLMs) that themselves are an extension of multiple linear regression models (MLR). Just as in the case of GLMs, GAMs can fit response data following Gaussian, binomial, Poisson, or gamma statistical distributions. However, smooth functions rather than parametric terms are used to estimate response curves in GAMs (Yee and Mitchell 1991). These smooth functions provide a better fit to environmental fields than linear or parabolic curves, as several other response shapes are commonly observed (e.g., logarithmic, plateau, bimodal). The statistical tool generalized regression analysis and spatial predictions (GRASP) is used for this study, allowing both an investigation of spatial predictions and a link with geographic information systems (GIS; Lehmann et al. 2002b). This paper applies these statistical methods to
regionalize meteorological data over heterogeneous terrain in Switzerland, using fine GIS layers that describe the topographic, geographic, and landscape characteristics of the country.

GIS applications yield valuable information that helps improve statistical models and their derived spatial predictions by considering data layers that are of relevance to explain the meteorological phenomena under study. With the help of digital elevation models (DEMs), GIS tools enable the development of detailed data layers that accurately describe the complex terrain of Switzerland. Slope and curvature layers are often employed, but a combination of these layers through a powerful GIS extension (Weiss 2001; Jenness 2006) leads to a precise classification of the terrain according to topographic characteristics, distinguishing high ridges from hills, deep and shallow valleys or plains. This fine representation of complex terrain is of key importance when dealing with wind characteristics.

Results provide reliable wind data over the country, allowing previous work on past heavy windstorm events or climate extremes (Beniston 2007). In addition, the method can also be adapted to predict mean wind speeds or return periods (Johnson and Watson 1999; Palutikof et al. 1999; Della-Marta et al. 2009), and can be investigated at different resolutions. Besides evaluating windstorm damage, this work can lead to other applications, such as the evaluation of the potential for wind-generated energy production (Schaffner and Gravdahl 2003).

In the following sections, an overview of the different GIS procedures is provided, followed by a description of the meteorological and physiographical data. In addition, a detailed explanation of the important preparation steps is given, with the resulting maps for different regions of Switzerland. Following the compilation of this information, various model selection methods have been tested using the variables by separately including different groups of predictors. The main statistical model selection results are listed, from which daily maximum wind speeds have been predicted and wind maps have been created. Finally, results are discussed and an outlook for future research is presented.

2. Study area

Switzerland is a small country in central Europe covering approximately 41,000 km² that stretches roughly 350 km from east to west and 220 km from north to south (Fig. 1). The landscape is extremely diversified and irregular, with a relatively low and narrow mountain range in the north (the Jura), the very high mountain peaks of the Alps that often exceed 4000 m above mean sea level in the south, and a relatively flat area (the Swiss Plateau) stretching from east to west between the two mountain ranges. South of the Alps begins the Po valley, which is the lowest region of the country. As mountain areas occupy more than 60% of the surface of the country, the altitudinal gradient is very high. The Alps...
act as a climate barrier between northern Switzerland, which is under the influence of both Atlantic and continental air masses, and southern Switzerland, where the Mediterranean Sea exerts a substantial influence. Schüpp (1978) has compiled over 40 different climate types over the Alps. Furthermore, winds can be determined by synoptic situations as discussed by Jungo et al. (2002); most of the important damaging historical storms have occurred during the winter period (Pfister 1999; Alexandersson et al. 2000).

3. Data and methods

a. Weather stations network

Figure 1 shows the distribution of the automatic network of the Anetz (Automatisches Netzwerk) meteorological stations, managed by the Swiss weather service (MeteoSwiss). The network provides reliable and quality checked weather data in digital form (Bantle 1989; Begert et al. 2005). Seventy stations have been measuring daily maximum wind speeds since 1981 and were used for this study. Wind speeds are recorded every 3 s, and mean and maximum wind speeds are archived each 10 min; daily maximum wind speeds were recorded at each station for the 1981–2005 period. The W98 values were then calculated from the listed data for each Anetz station, providing information on the different wind conditions over the country.

b. Topographic data

All physical factors likely to disturb wind flows need to be included in the regression analysis and accurately represented by GIS layers. The main input datasets was the DEM of Switzerland at a resolution of 50 m provided by the Swiss Federal Office of Topography (more information available online at www.swisstopo.ch). Using a GIS function tool (Weiss 2001; Jenness 2006), relevant data layers at different scale levels were generated for the study.

From the original DEM, various elevation-related parameters were calculated, each of them at the same resolution of 50 m. First, surface elevation, which has an obvious influence on wind speeds (Tennekes 1973; Miller and Davenport 1998; Winstral et al. 2009), was estimated at different spatial scales that may influence flow characteristics. To take into account the surrounding topographic characteristics, different maps were created by considering all elevation values in a radius of 100 m, 200 m, 500 m, 1 km, and 2 km respectively. For example, each 50-m cell of the dem500 layer reflects mean elevation values within a radius of 500 m.

Slopes and curvature (including profile and plane curvature) were also calculated to provide further topographic information, and the rescaling steps were carried out for each. Curvature describes the shape of the ground, whether the land is convex or concave, whereas plane and profile curvature describe the curvatures perpendicular and transverse to the mountain, respectively. Wind flows are affected by terrain undulations, which explain why all mountain shapes need to be taken into account to accurately describe wind speeds. Effects of curvature on wind flows have been pointed out by McKee and O’Neal (1989); slope influence was assessed by Petkovsek and Hocevar (1971), McNider and Pielke (1984), and Smith (1985). In addition, the orientation of the mountains was considered, as the main airflow directions are likely to be related to the major morphological characteristics. Orientation angles were transformed using sine and cosine functions, leading to two separate variables—northings and eastings—again assessed at different scale levels.

Topographic characteristics of the surface have a marked influence on the wind velocity (Barry 1992). Canyon shapes, midslope areas as well as high ridges can disrupt wind flows and thus explain channeling effects (Kaufmann and Weber 1998). Studies showing the importance of mountain geometry on wind speeds were made by Nappo (1977), Steinacker (1984), and Vergeiner and Dreiseitl (1987). Based on the DEM and the slope map at 50 m, simple procedures can lead to a classification of the surface into several landform categories which can then be used as inputs for regression models. This procedure creates topographic position indices (TPI) for each cell of the DEM (Weiss 2001; Jenness 2006). TPI compares the cell elevation with the average cell elevation in the neighborhood; positive differences imply that the position is on a hill, whereas negative values would rather describe a valley. Null values represent either flat areas or midslopes.

For this work, two TPIs have been constructed with mean altitudes located within a 250-m and 2-km radius, respectively. These two indices are compared in a first classification step. For instance, a high TPI value in a small neighborhood associated with a low TPI value in a larger neighborhood would correspond to a small hill located within a larger valley. This first step may not be sufficient for the analysis, as two different surfaces may lead to the same TPI. To avoid this confusion, the results are combined with the slope values. This combination produces the final landform map that describes the topographic characteristics of Switzerland according to 10 different categories. Further information as well as a more detailed explanation on the algorithms used in this procedure can be found in Jenness (2006).

The final landform map can be seen in Fig. 2. An overall view of Switzerland is shown, and also an
enhanced view over the Aletsch Glacier in the Alps. In the large-scale view, a clear difference can be observed between the mountain regions and the Swiss Plateau. The classification procedure yields the different shapes of the topography, by distinguishing between the uppermost mountain ridges and the adjoining upper slopes and upland drainages. The tool also highlights the shallow valleys and the open slopes beneath the high peaks, highlighting the shape of the mountain. It is thus seen that, starting from a basic DEM associated with a slope layer, simple algorithms can lead to a fine classification of highly heterogeneous terrain. This landform classification map will be further introduced as a factor type variable in the regression procedure by sampling information around weather station locations.

c. Sampling physiographical conditions at the meteorological stations

Values of all the physiographical variables mentioned above and prepared for the study (Table 1) were sampled by reading the information found for each Anetz station on a 50-m grid. To describe the landform characteristics, 10 categories were originally used. Unfortunately, these 10 categories are not all represented...
d. Generalized regression analysis and spatial predictions

As GAMs have been previously used in different environmental fields (Lehmann et al. 2002a), they were employed in this work to predict wind speeds using the GRASP extension tool. GRASP is a statistical package that has been developed to automate the production of spatial predictions from point observations using GAMs (Lehmann et al. 2002b), and it can be used in the S-Plus software package or as an extension for the statistical software package known as R. This tool is made up of a number of functions designed to facilitate regression analysis aimed at spatial predictions, and is associated with a user-friendly graphical interface. As GRASP was originally created to predict the spatial distribution of vegetation (Zamieszki et al. 2002; Maggini et al. 2006) or animal species (Castella et al. 2001; Ray et al. 2002; Luoto et al. 2006), this tool has been adapted from its original use in spatial predictions in ecology to predictions in meteorology. Indeed, GRASP is a generalized method for predicting any point observations from statistically and spatially related explanatory variables, and thus can be used in a range of environmental fields.

The true challenge in selecting a regression model remains a complex issue (Burnham and Anderson 2002). These authors have listed the selection methods in three categories: 1) optimization of Akaike’s information criterion (AIC) or Bayesian information criterion (BIC), 2) tests of hypotheses (chi-square and F tests), and 3) ad hoc methods. The AIC (Akaike 1973) relies on an empirical log-likelihood function that assesses the relative distance between the fitted model and the observed data (Burnham and Anderson 2002). The log-likelihood value is penalized by the number of parameters in the fitted model, favoring models with fewer selected predictors. As for AIC, BIC is also penalized by the number of observations (Schwarz 1978). Statistical tests such as chi square or F help to understand the consequences of including or excluding a term from the model.

A cross-validation is a frequently used solution for a model selection (Stone 1977; Maggini et al. 2006). The main advantage of this procedure is that no other observation dataset is required for the model validation. Other datasets are usually rare and, if available, are generally used directly for the analyses (Araújo et al. 2005). The cross-validation process creates random subsets of the main dataset to validate the model results (Lehmann et al. 2002b) and includes all the available data in the modeling process. The K-fold cross validation (Franklin et al. 2000) and the bootstrap technique (Efron and Tibshirani 1993; Guisan and Harrell 2000) are considered as the most interesting compromise, since they assess model stability without any loss of information.

e. Model selection steps

For this analysis, the W98 data calculated from the 70 meteorological stations was fitted to a Poisson distribution because of the skewed shape of its strictly positive distribution (also tested by the chi-square goodness-of-fit test). This assumption implies that only positive values of wind speeds can be predicted.

In the context of this analysis, correlation between variables must be taken into account, as two variables might lead to the same result (Lehmann et al. 2002b). To address this issue of orthogonality, maximum correlation between variables was set to 0.8 for all the selection tools (Leathwick et al. 2005, 2006). Beyond this threshold, one of the two correlated variables is automatically removed from the selection.

Several model selection methods were tested using different sets of candidate variables. Available selection methods in the GRASP package (AIC, BIC, and F and chi-square tests, as well as cross validation) were tested on several groups of variables (Table 2), and the results were analyzed for each method through validation statistics: correlation and cross-correlation coefficients. In all cases, stepwise procedures were used for the selection of significant predictors, with a backward direction for the cross-validation method.

Some of the main results are described in this paper and only one “best” model is fully detailed. In fact the notion of best model is fairly subjective with so many potential predictors. It is of course possible to fit a wide range of models with all possible combinations of variables. The general rule in regression analysis is to avoid overfitting the data with too many variables. The best way to avoid overfitting is to cross validate each model.

<table>
<thead>
<tr>
<th>Predictor name</th>
<th>Description</th>
<th>Resolutions (m)</th>
<th>Data class</th>
</tr>
</thead>
<tbody>
<tr>
<td>demXXm</td>
<td>Altitude</td>
<td>XX = 50–2000</td>
<td>Numeric</td>
</tr>
<tr>
<td>cuXXm</td>
<td>Curvature</td>
<td>XX = 50–2000</td>
<td>Numeric</td>
</tr>
<tr>
<td>cuXXpl</td>
<td>Plane curvature</td>
<td>XX = 50–2000</td>
<td>Numeric</td>
</tr>
<tr>
<td>cuXXpr</td>
<td>Profile curvature</td>
<td>XX = 50–2000</td>
<td>Numeric</td>
</tr>
<tr>
<td>sloXXm</td>
<td>Slope</td>
<td>XX = 50–2000</td>
<td>Numeric</td>
</tr>
<tr>
<td>eaXX</td>
<td>Easting</td>
<td>XX = 50–2000</td>
<td>Numeric</td>
</tr>
<tr>
<td>noXX</td>
<td>Northing</td>
<td>XX = 50–2000</td>
<td>Numeric</td>
</tr>
<tr>
<td>ldXXm</td>
<td>Landform</td>
<td>50 Factor</td>
<td></td>
</tr>
</tbody>
</table>
from the simpler one to the more complex one, and to keep the model that exhibits the best cross-validated statistic (Maggini et al. 2006). Cross validation indeed allows testing the stability of a model. In this study, the final model that was chosen was one with a high correlation coefficient and that included variables that were physically coherent.

The first step is related to the variable selection, namely to answer the question as to which of the 43 variables should be included in the GRASP procedure. One can work with all the variables, or focus only on certain categories. In the present analysis, the study was investigated with the five groups of variables shown in Table 2. The idea behind these diverse groups is to investigate the influence of different variables on the wind flows. The first group of variables takes into account those without any previous manual selection (group 1). It allows evaluating the manner in which GRASP handles the entire list and which variables are selected by the model. Working with the local data at 50 m (group 2) provides information on the impact of the ground characteristics on the wind speeds in the immediate neighborhood; whereas the 2-km group analyzes the influence of the surroundings (group 3). Another possibility is to sort the predictors by selecting variables from the list related to both their correlation with the original W98 sample data and their autocorrelation (group 4). At each step, the variable most correlated to the original W98 data is selected, and all other variables that have a correlation coefficient greater than 0.8 are removed. This sorting results in a list of 15 variables. The final group represents a subjective selection from the original list (group 5), obtained from several tests that were made with different groups of variables.

4. Results and discussion

a. Model selection

Table 2 lists the five groups of predictors that were tested with the different selection methods. The first four groups showed no substantial result, as no convincing model was found, either because the selected variables were unsatisfactory, as no clear physical explanation could be derived, or because the correlation coefficients were unsatisfactory. A new attempt was made only with the variables at 1 km. At this scale level, information on the behavior of the variables is detailed in a mid-scale neighborhood, which roughly corresponds to the mean of the spatial scales chosen for the analysis. Because they were rarely included in the models and most of the time were removed from the selection methods, the northing and easting predictors were excluded from final models; these variables clearly fit better when taking into account wind directions in the analysis. Therefore, the work was conducted here with the following predictors: dem1km, cu1km, cu1kmpl and cu1kmpr, slo1km, and ld2km (see below for definitions; group 5 of Table 2). First attempts with these variables gave interesting results, but the landform predictor was consistently removed from the model. As the landform characteristics were considered to be the most influential parameter on the wind flow, that variable was forced to be included in the model selection.

Table 3 lists the results from the different selection tools (AIC, BIC, F, chi square, and cross validation) applied to predictors of group 5. The model arising from the cross-validation selection method was chosen for the prediction of wind speeds. Apart from the landform (ld2km) whose presence was forced in the model, three variables were included in the model: slope (slo1km), profile curvature (cu1kmpr), and altitude (dem1km).

This cross-validation model resulted in a Pearson’s correlation coefficient of 0.89 between observed and predicted values and a cross-correlation coefficient of 0.77 (Table 3). Correlation coefficients between the four selected predictors were all under the chosen threshold value of 0.8.

Figure 3 shows the response curves for each of the four selected predictors included in the model. The range of values of the different predictors is given on the abscissa, while the sensitivity of the model is plotted on the ordinate: this is represented on the scale of the linear predictor before transformation by the inverse response function.

Figure 3d emphasizes the sensitivity of the model to altitude. As expected, wind speed values clearly increase with altitude. The curve decreases slightly toward the end, probably as a result of the lack of high altitude data, since only two weather stations are located above 3000 m. Below 1500 m, wind speeds exhibit a relatively low sensitivity to elevation. A few weather stations at altitudes of roughly 1000 m in the Alps exhibit low W98 values, because they are located in sheltered valleys: these particular stations
Table 3. Results obtained by different selection methods presenting the resulting model formula, the correlation between observed and predicted values (COR), the cross-validated correlation (CVCOR), the percentage of explained deviance* (D2), and the number of DFU.

<table>
<thead>
<tr>
<th>Method</th>
<th>Selected predictors</th>
<th>COR</th>
<th>CVCOR</th>
<th>D2</th>
<th>DFU</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>ld2km + dem1km</td>
<td>0.85</td>
<td>—</td>
<td>0.72</td>
<td>10</td>
</tr>
<tr>
<td>BIC</td>
<td>ld2km</td>
<td>0.79</td>
<td>—</td>
<td>0.63</td>
<td>6</td>
</tr>
<tr>
<td>$F \ (p &lt; 0.05)$</td>
<td>ld2km + dem1km</td>
<td>0.85</td>
<td>—</td>
<td>0.72</td>
<td>10</td>
</tr>
<tr>
<td>Chi square ($p &lt; 0.05$)</td>
<td>ld2km</td>
<td>0.79</td>
<td>—</td>
<td>0.63</td>
<td>6</td>
</tr>
<tr>
<td>Cross validation</td>
<td>ld2km + slo1km + cu1kmpr + dem1km</td>
<td>0.89</td>
<td>0.77</td>
<td>0.77</td>
<td>18</td>
</tr>
</tbody>
</table>

* Deviance is a measure of deviation for GAMs, in a similar way that standard deviation is for linear regression models.

explain the flat response curve below 1500 m seen in Fig. 3d. Above 1500 m, the weather stations are less sheltered, thus explaining the sudden increase of wind speeds. Uncertainties are seen to grow as altitudes increase, since points are more scattered in higher zones.

Surprisingly, slope shows an opposite trend, as wind speeds decline when the slope increases (Fig. 3b). The opposite could have been expected, since steep slopes are more likely to be situated in elevated mountain areas, leading to high wind speeds. However, because shallow slopes represent more open areas, where wind flows are less affected by obstacles, wind speeds can indeed be higher than in hilly zones where many discontinuities are present.

Land shapes also have an influence on the results. Figure 3a shows that categories do not seem to have an important effect on the flows, except the far right category, which corresponds to mountain tops or high ridges. Wind speeds on elevated peaks are higher than elsewhere, which logically corresponds to field observations where strong winds are indeed measured close to mountain tops.

The final predictor selected is more delicate to interpret (Fig. 3c). Profile curvature describes the shape of mountains slopes; that is, whether they are convex or concave. Negative values indicate that the surface is upwardly convex at that cell, while positive values imply that the surface is upwardly concave. Zero values indicate that the side surface of the mountain is flat, therefore that the slope is constant. Convex zones (negative values, on the left) lead to higher wind speeds. Concave shapes are less represented in the dataset, but the trend seems to show increasing or constant values. In the middle, which corresponds to flat (linear) slopes, the effect on flows is close to neutral. Negative values of profile curvature correspond to upward convex forms, which relate to hills or mountain tops. It is therefore coherent to find higher wind speeds where negative curvature values are calculated, because wind speeds are indeed higher over mountain tops. On the contrary, positive values match with concave shapes; that is, valleys. In these areas, wind speeds are more likely to be low, unless there are some channeling effects that are difficult to predict. In convex areas, two situations can be observed: either high wind speeds due to channeling effects, or low wind speeds because the area is sheltered. Although few stations are located in concave zones, the right side of Fig. 3c shows important variability, which may indicate the different effects of concavities. Analysis of values plotted on the ordinate suggests that wind speeds are more sensitive to the curvature and the slope variables than to landform and altitude.

b. Spatial predictions

Spatial predictions of wind speed were produced from the selected model. Bootstrap techniques resulted in 200 predicted maps of W98 values, from which a mean value was calculated to produce the final W98 map.

Bootstrap methods described in Efron and Tibshirani (1993) enable to assess model uncertainty without the need of additional data (Guisan and Harrell 2000). From the original sample, random subsamples are created and models are built from these different samples, leading to as many maps as samples drawn. Finally, mean and standard deviation values can be calculated on the basis of a cell-by-cell analysis.

The W98 results present a bell-shaped curve with a mean around 21 m s$^{-1}$ and a standard deviation of roughly 43 m s$^{-1}$. A first appraisal shows a clear influence of the topography on the results, as the mountain areas are clearly taken into account in the process. Highest values are in the Alpine region, while low results are mostly located in the flat topography of the Swiss Plateau or within deep valleys. In the Jura Mountains, wind speeds are also fairly high, but lower than in the Alps.

Wind speeds derived from the methods discussed above range from 10 to 50 m s$^{-1}$ (Fig. 4). A few values reach spurious wind speeds of up to 150 m s$^{-1}$. Such errors are linked to extreme negative profile curvature values; as can be seen in the response curve, low curvature values have a positive impact on wind speeds, and values have been extrapolated in this case, leading to
unrealistic wind speed results. Furthermore, these high values are all located over high peaks in the Alps. An upper threshold wind speed of 50 m s\(^{-1}\) has been introduced in order to avoid these problems of extrapolation. This value is a reasonable threshold since out of nearly 20 million cells, only two exceeded 100 m s\(^{-1}\), and less than 0.03\% of these cells were above 50 m s\(^{-1}\).

In the canton of Valais, the lowest values of W98 are found in narrow valleys. In the main valley of the canton (the Rhone valley), which is wider but also lower in altitude, the map shows higher values. These results correspond to observations, as strong winds can occur in wide valleys, but flows have more difficulty in entering smaller or more remote valleys.

Similar conditions can be observed to the north in the Jura, which is a lower mountain range with more open areas. In these wide and open zones, high W98 values are modeled. Again, as the valleys become narrower, values decrease because strong winds rarely flow through narrow valleys. On the mountain tops, wind speeds are at their maximum, as for the Alps. When comparing the mean of the W98 values within the 1000–1500-m altitude range in the Jura and the Alps, the Jura exhibit higher values. This difference is explained by the landscape parameter, where the elevation range corresponds mainly to ridge crests in the Jura but to high valleys or U-shaped valleys in the Alps. The landform variable has an evident influence on the wind speed distributions, as the valley shapes are clearly important in the model predictions.

Over all lakes, fairly uniform values on the order of 20 m s\(^{-1}\) were modeled. The very small wind speed differences between the lakes are due to the elevation, but the altitude variable does not seem to contribute sufficiently to make a significant difference in W98 over Swiss lakes.

The standard deviation map was created by analyzing the 200 bootstrap map results, calculating the grid-by-grid standard deviation. Over Switzerland, minimum standard deviation values were found around 0.46 m s\(^{-1}\), while some values exceeded 10 m s\(^{-1}\). Mean values over the country were, however, roughly 2 m s\(^{-1}\). The highest values are found in the Alps, in the high altitudes zone. Low standard deviations are found over the Swiss Plateau, and also over the lakes. As explained above, values for most of the lakes are similar and thus standard deviations are very low.

c. Model accuracy

Figure 5 displays the predicted wind speeds versus the observations at the 70 Anetz weather stations. The graph shows a large group of wind speed values on the left, while only a few are over 30 m s\(^{-1}\). The regression
slope is therefore determined by the values exceeding 30 m s\(^{-1}\). On the left graph, lowest predicted values and the ones over 25 m s\(^{-1}\) seem to follow a positive trend; whereas, no significant tendency can be observed for values between 20 and 25 m s\(^{-1}\).

The validation scattered graph on the right gives a Pearson’s coefficient of \(R^2 = 0.79\). The cross-validation graphic confirms that model results are robust, as a cross correlation coefficient of \(R^2 = 0.59\) has been determined. As the distribution of the data does not fit perfectly to a normal distribution, calculations other than the Pearson coefficient were made for a quantitative validation of the results (Table 4). The Spearman rank correlation coefficient was calculated to assess errors, as well as the RMSE. The reliability index (RI) quantifies the average factor by which model predictions differ from observations (Leggett and Williams 1981), whereas the mean average error (MAE), Willmott’s index of agreement (D; Willmott 1982) and the modeling efficiency (MEF) are good indicators of goodness of fit (Mayer and Butler 1993). Values of the listed indicators confirm that the model functions correctly, but it does not seem to explain all situations. Furthermore, a measure of the bias can be assessed when looking at the slope and intercept of best-of-fit line equations shown on Fig. 5.
The absolute percentage of relative errors was assessed at the 70 station points by calculating the difference between observed and predicted values of W98. Mean error was 10%, maximum was 28% and the standard deviation of these error values was 7.5 m s\(^{-1}\). Most errors exceeding 20% occurred at stations located in the Alps. Considering the relative errors, both negative (i.e., when predicted wind speeds overestimate observations) and positive values were mainly located in alpine areas. Apart from this geographical trend, the model had difficulty approaching extreme W98 wind speeds, as can be seen when looking at slope values of best-fit equations of the scattered plots (Fig. 5). Highest positive relative errors were linked to the uppermost observed W98 values of the weather stations, while high negative errors corresponded to lowest W98 velocities. Overall, the model reproduced the values measured at the weather stations in a realistic manner despite having difficulties reaching extreme W98 wind speeds.

**d. Discussion**

This part of the study investigates to what extent wind speeds can be described by local terrain characteristics. The focus here is thus on the relation between long-term wind statistics at weather stations and the description of their surrounding environment. Valleys produce their own local wind systems as a result of thermal differences, and topographic features interfere with the wind flow by generating turbulent vortices (McNider and Pielke 1984; Oke 1987), vertical air compression, and frictional drag (Barry 1992). Turbulence effects, vorticity, or atmospheric stability and their impact on wind flows are neglected in this study, although they are essential when explaining wind velocities in the boundary layer (Stull 1988; Holton 2004). These physical processes cannot be considered by a GIS approach, but would certainly help explain errors measured at weather stations.

Predictors used for this study have been assessed at different spatial scales, with a maximum scale of 2 km, assuming that local wind speeds are linked to the environmental conditions in close proximity. Nevertheless, wind flows are also influenced by physiographical features within a larger environment, as wind speeds are also governed by regional circulations and larger-scale synoptic situations. Here again, GIS tools have difficulty in describing these large scale or regional scale weather conditions.

The number of observations in GAMs influences the possible complexity of the model [i.e., the numbers of

---

**Fig. 5.** Scatterplots of model predictions and observations at the Anetz stations. (left) The scatter graph shows the correlation for the cross-validation selection, and (right) the result for the validation. On both graphs, the line of best fit and its equation are shown (full line), as well as the 1:1 line (dotted line).

**Table 4.** Various indicators providing a qualitative validation of the results. The Pearson and the Spearman rank correlation coefficients are listed, as well as the RMSE. The last columns show three indices: RI, D, MAE, and MEF index. All of the calculations were made for the validation and the cross-validation results. The last line provides values of the indicators in the case of a perfect agreement.

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Pearson</th>
<th>Spearman</th>
<th>RMSE</th>
<th>D</th>
<th>RI</th>
<th>MAE</th>
<th>MEF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validation</td>
<td>0.89</td>
<td>0.65</td>
<td>2.78</td>
<td>0.94</td>
<td>1.05</td>
<td>2.19</td>
<td>0.80</td>
</tr>
<tr>
<td>Cross validation</td>
<td>0.77</td>
<td>0.42</td>
<td>4.05</td>
<td>0.86</td>
<td>1.07</td>
<td>3.15</td>
<td>0.58</td>
</tr>
<tr>
<td>Perfect values</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
degrees of freedom used (DFU) by the selected variables, which is the case for all regression techniques. The more available data, the more complex is the model. In this case, we used the data from 70 weather stations and selected a final model with four variables with 18 DFU. This is a reasonable ratio between number of observations and DFU. More observations would not necessarily affect the resulting model but rather decrease the uncertainty in the predictions. In general, regression results are, however, not directly influenced by the spatial density of the observations but rather by the representativeness of the environmental gradients being sampled. Spatial density would be much more significant when employing methods based on spatial interpolation. Studies have already shown the impact of the number of observations in similar models in ecology (Stockwell and Peterson 2002). Distance between stations can, however, influence the independence of observations. When stations are located too close to one another, they become spatially autocorrelated, and the assumption of independent, identically distributed (iid) data required by the regression technique, can be violated. In the present case, most stations are separated by tens of kilometers and are therefore considered sufficiently independent.

Information on land cover can also help to assess the wind speeds. In the logarithmic wind profile commonly used to compute wind speeds close to the ground (Tennekes 1973; Oke 1987), the influence of the type of land cover is described by the roughness length parameter $z_0$, which reflects the roughness characteristics of the surface by quantifying its impact on wind flows within the boundary layer. Smooth surfaces such as flat ice fields are associated with very low values of $z_0$, while urban areas and forests are associated with higher values. The $z_0$ parameter is directly linked to the surface cover, and tables relating land use to roughness length values are widely found in the literature (Davenport 1960; Garrat 1992; Wieringa 1993; Graefe 2003). The land-use parameter was initially included in the regression analysis, but was rarely selected, because it gave unsatisfactory results; it was therefore removed from the study. Land use would probably be a much more appropriate explanatory factor at microscale levels; whereas, at the scales used in this study, it appears to be overshadowed by the other physiographical variables.

5. Conclusions and outlook

Fine GIS layers combined to state-of-the-art statistical models have been used to produce a map of extreme wind speeds over Switzerland. The resulting W98 map provides information on local wind conditions, and calculated values may in a future step be used to assess storm damage to infrastructure.

The following lists the most significant results that emerge from this study.

1) GIS tools are appropriate tools to generate accurate layers that describe the physiographical elements of Switzerland. Powerful algorithms helped classify the landscape into 10 categories with different topographic characteristics. The resulting landform map successfully distinguishes between high ridges, valleys, hills, open slopes, plains, etc. The three main geographical zones of Switzerland are highlighted, with clear differences between the Jura Mountains, the Swiss Plateau and the Alps. Within the latter, narrow valleys, wide valleys, ridges, and midslope areas are identified. This 50-m resolution map produces high-resolution data that is used in the multilinear regression model to predict extreme wind speeds over Switzerland.

2) GRASP has shown great potential for a distributed analysis of wind speed and is clearly a promising tool for other climatic variables. GAMs were applied to regionalize W98 values on an elevated heterogeneous landscape.

Variables likely to have an influence on wind flows were included in the regression process. The GRASP tool, which allows the use of GAMs, has been applied to the model selection. A cross-validation model selection method was chosen to select the final model. To assess errors, bootstrap methods were used that enabled a quantification of means and standard deviations of W98. The final model is based on a combination of data at a 1-km resolution. The altitude, the slope, and the profile curvature were included, and the landform map was entered into the model. Pearson’s correlation coefficients were satisfactory, with values of 0.89 and 0.77 for the cross correlation.

Thus, adaptation of regression models and GAMs to predict wind speeds provides convincing results, which can help alleviate some of the problems associated with direct modeling of wind in the complex terrain of Switzerland. A combination of high-resolution GIS data associated with recent model selection methods offers a good alternative to predict historical wind speeds, especially over complex and heterogeneous terrain.

3) Topography plays a major role in the spatial distribution of extreme wind speeds. The wind map shows a clear influence of the topography, as wind velocities for sites located above mountains are higher than sites situated on the Swiss Plateau. In the Alps, effects of valley widths are well reproduced. There is
a clear impact of the landform variable on the predicted wind speeds. Slope, elevation, and profile curvature have also been identified as significant predictors of wind speed intensity.

Future work to improve the performance of the model selection is under consideration. Bootstrap methods can be more thoroughly employed in the procedure by making a random selection within the input variables and selecting models from the chosen variables. As suggested by Thuiller (2007), one could investigate the mean values of wind speeds predicted by different model selection methods. Boosted regression tree methods can be used to better express interactions and nonlinearities in the data. They have been fully described in Elith et al. (2008). The use of multivariate adaptive regression splines could better describe nonlinear relationships between the predictors and the response. Detailed information on this method can be found in Leathwick et al. (2005).

The W98 are defined as threshold values in many formulas that address the issues of damage functions. Using the results discussed in this paper, severe windstorm damage can be determined by taking into account the wind data recorded during past wind storms in Switzerland. Impact of climate change on future storm events can also be estimated as part of future research, and the associated economic losses can be quantified.

This work opens new perspectives for further research. While the research has focused on a fairly large and heterogeneous area, similar research could be envisaged for smaller domains. Similarly, work could also concentrate on specific altitude zones, only focusing on high zones located in mountains, or on open plains and flat areas, in order to investigate the behavior of winds over more homogeneous environments.

Predicting the wind speeds over Switzerland was based on physiographical predictors only. Investigation of wind direction statistics at weather stations, possibly linked to the prevailing climate conditions (Schüpp 1978; Schmidtke and Scherrer 1997; Jungo et al. 2002) expressed by the northing and easting predictors, could help detect regional wind systems that occur in Switzerland such as the Bise (Wanner and Furger 1990) or the foehn (Hoinka 1985). Inclusion of further information such as meteorological variables (temperature, turbulence, stability, etc.) would lead to the refinement of climate models (Goyette et al. 2001; Goyette 2008), which could better represent the numerous atmospheric interactions and complex physical description of wind flows.

REFERENCES


Bandle, H., 1989: Program documentation for the Swiss climate data base at the computing center of ETH Zurich. MeteoSwiss publication, Zurich, Switzerland (in German), 8 pp.


**1970**

**JOURNAL OF APPLIED METEOROLOGY AND CLIMATOLOGY**

**VOLUME 49**


