Using active learning for monitoring networks design: the example of wind power plants sites evaluation

D. Tuia, A. Pozdnoukhov, M. Kanevski

Abstract Locating new wind farms is of crucial importance for energy policies of the next decade. To select the new location, an accurate picture of the wind fields is necessary. However, characterizing wind fields is a difficult task, since the phenomenon is highly nonlinear and related to complex topographical features. In this paper, we propose both a nonparametric model to estimate wind speed at different time instants and a procedure to discover underrepresented topographic conditions, where new measuring stations could be added. Compared to space filling techniques, this last approach privileges optimization of the output space, thus locating new potential measuring sites through the uncertainty of the model itself.

1 Introduction

This paper describes a novel computational approach of monitoring network optimization and its application to a problem of wind power plants sites evaluation. In this complex spatial decision making process it is essential to build a model for the estimation of wind speeds and their constancy over time. Geostatistical models are often used to assess these properties from meteorological data to assemble wind power capacity atlases. However they often rely on incomplete information and insufficient number of measuring stations, especially in mountainous regions

D. Tuia
Laboratoire des Systèmes d'Information Géographique, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland. e-mail: devis.tuia@epfl.ch

A. Pozdnoukhov
National Centre for Geocomputation, National University of Ireland Maynooth (NUIM), Ireland. e-mail: alexei.pozdnoukhov@nuim.ie

M. Kanevski
Institute of Geomatics and Analysis of Risk, University of Lausanne (UNIL), Switzerland. e-mail: mikhail.kanevski@unil.ch

Ninth International Geostatistics Congress, Oslo, Norway, June 11. – 15., 2012

of complex topographies. They also require various corrections to account for the influence of topography on wind speeds.

In this study we apply a methodology based on support vector machines (SVM [1]) to assess the areas of interest for wind power plants construction in Switzerland from the interpolation of monthly average wind speed data. The model incorporates a rich set of topographical indices as explanatory input variables (see Table 1). With this procedure, we produce a spatial map of average wind speeds on the Swiss territory, which helps assessing the suitability for the construction of new plants.

To enhance the accuracy of this map by taking of additional measurements, one has to consider the monitoring network optimization as an exploration of the high-dimensional space of combined geographical and topographical variables. We define an active learning criterion [2, 3, 4] to achieve this goal and target the exploration at uncertain areas close to the decision threshold for power plants construction [5]. Network design can be willingly biased towards the areas at risk of the phenomenon to be modeled, in our case average the wind speeds of about 4.5 [m/s] at a height of 50 [m] over the ground, which is the minimum monthly average wind speed required by the Swiss law for the expediency of a power plant facility construction.

Using this criterion, we study the topographical features of the Swiss territory in terms of interest for a new monitoring station. We pay particular attention to sites capable to improve the accuracy of wind speed models in the areas of maximal uncertainty around the decision threshold of 4.5 [m/s]. With this study, we extract topographical conditions leading to model uncertainty and to extract possible locations of interest for new monitoring stations.

2 Predicting suitable areas for wind farms

We are interested in predicting two quantities: on the one hand, wind speed at a height of 50 [m] over the ground and on the other hand persistence of the wind over time.

For the first problem, we converted the prediction problem into a classification problem: since we were interested only in the meeting of the specific threshold of 4.5 [m/s], we recoded the recorded wind speed into a positive class (+1) when the threshold was met and a negative class (0) otherwise. We extrapolated the average monthly wind speed at 50 [m], $\bar{s}_{50}$, from the average speed measured at 10 [m] by the sensor, $\bar{s}_{10}$, using the following relationship:

$$\bar{s}_{50} = \bar{s}_{10} * \frac{\ln(\frac{50}{0.1})}{\ln(\frac{10}{0.1})} = \bar{s}_{10} * 1.3495 \quad (1)$$

where 0.1 is the average roughness length in Switzerland (see Swiss Bylaw on Energy).

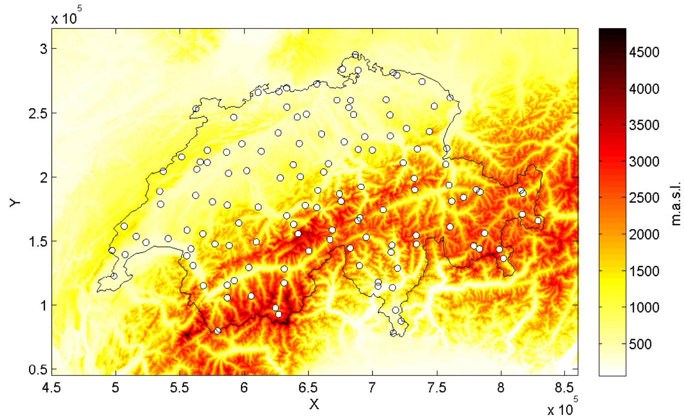

Fig. 1 Swiss sensor network used in this study and corresponding DEM RIMINI at 250 [m] cell resolution.

We then trained a non parametric model based on SVM [1], which is a robust method for data classification. As input space, we used the variables detailed in Table 1. The model was trained on data coming from the monitoring stations of the Swiss sensor network and was used to predict each cell of the RIMINI digital elevation model (see Fig. 1). This way, we predicted for each month the suitability of swiss territory for wind farm development.

For the second problem, we considered prediction maps produced separately for each month of the period 2003-2008 and averaged out the prediction (see Fig. 2) over the 72 models. This way, we track the persistence of winds and retrieve it as a probability to exceed the 4.5 [m/s] threshold. Next section details the selection procedure.

Table 1 Topographic features considered in the study

Number	Symbol	Type	Description	Modality
1 – 2	[X, Y]	Spatial coordinates	Location of the sample[a]	
3	[Z]	Altitude	Altitude of the sample[a]	
4 – 6	[DoG]	Difference of Gaussians	Substraction of two smoothed DEMs, describes convexity of terrain	Small / Medium / Large
7 – 9	[Slope]	Slope	Norm of terrain gradient, describe slopes of terrain	Small / Medium / Large
10 – 13	[DD]	Directional derivatives	highlight natural topographical obstacles that break wind	NorthSouth, EastWest at Small and Medium scales

[a] Monitoring station or pixel extracted by DEM.

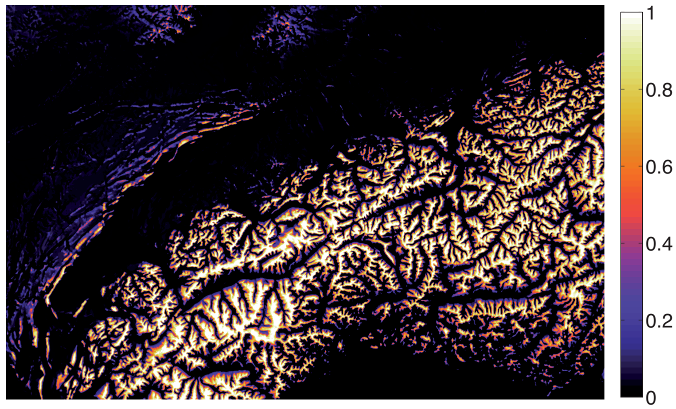

Fig. 2 Average 2003-2008 prediction map. In white, areas always meeting the 4.5 [m/s] threshold. In black, areas unsuitable for wind farming.

3 Identifying critical topographic conditions for new monitoring stations

In order to improve the decision map of Fig. 2, new measurement stations could be added to the current sensor network. A possibility would be to add stations in areas that are not sampled, in a geographical space filling perspective [6, 7]. In this paper, we aim at filling the output space *in areas that are the most uncertain for the current model*. To do so, we assess the uncertainity of the committee of 72 SVM models (one per month) by evaluating the decision functions of the single models on each point of the grid.

A sample predicted with a value of the decision function between 0 and 1 can be considered as uncertain, since it falls within the separating hyperplane of the SVM [1]. In [8], this property was exploited to design a criterion to rank samples by their uncertainty and possible contribution to the model, an *active learning criterion*. Broadly speaking, we are not interested by adding samples that the model can easily classify into windy or calm. On the contrary, if a sample has a decision function between 0 and 1, it means that the current model still classifies it in one of the two classes, but with little confidence. Following this intuition, we can state that adding this sample to the current training set will be beneficial for the improvement of the model, as it will disclose a part of the input space (of the topographic conditions) that is currently not well handled.

Figure 3 illustrates the mode of the decision function for the 72 months: the areas in dark tones are those of reduced confidence in the class assignment, i.e. the areas where the mean speed is often very close to 4.5 [m/s].

With this knowledge we analyzed the specific topographic conditions of the uncertain areas. Due to the high complexity of the features at these uncertain loca-

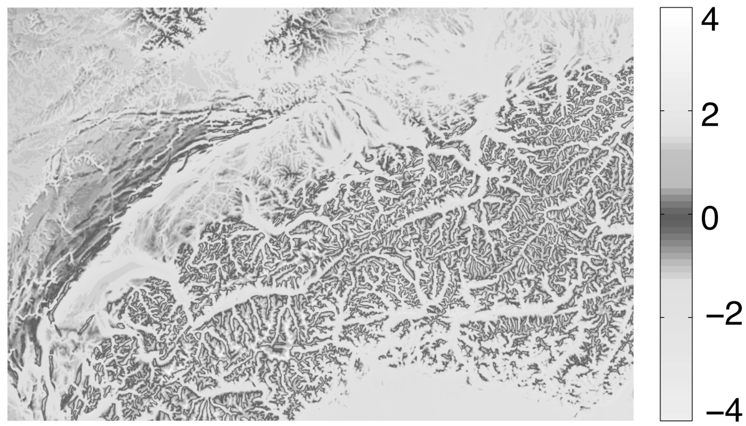

Fig. 3 Mode of the decision functions of the 72 SVM classifiers. In dark tones, ares of persistent uncertainty.

tions we extracted general trends from the data using principal component analysis (PCA). PCA is applied only to the data within the SVM margin ($|f(x)| < 1$) to extract the components of maximal variance and then analyze the behavior of the features in the maximal variance directions. In our setting, a maximal variance direction can be interpreted as a linear combination of terrain features explaining the topographic conditions representative of the SVM uncertain predictions shown in dark tones in Fig. 3. Being a combination of terrain features, it accounts for many of them at the same time, providing an useful visualization tool for uncertain topographical conditions.

Figure 4 illustrates the results of this visualization for the two most informative components. The first, accounting for 23% of the variance, concentrates spatially uncertain topographic conditions related to high altitude and strong slope, while the second, accounting for 18% of the variance, accounts for strong directional derivatives at high elevation. Together, these two maps can be used to plan the localization of new sensors specialized in the estimation of winds.

4 Conclusion

In this study, we assessed the suitability of machine learning algorithms for the assessment of renewable energy potential. We considered wind power, which is one of the most suitable and promising resources for the green-energy policy of the near future in Switzerland. Multitemporal analysis was carried out for the search of potential sites for wind power farms. The integration of topographical information showed desirable properties and led to a multitemporal mapping of the suitability of

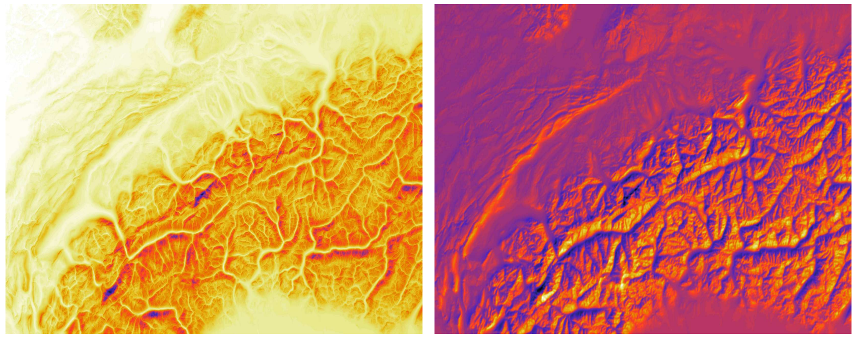

Fig. 4 First two principal components of the uncertain areas in dark tones in Fig. 3.

the territory for wind farming. To increase confidence in the estimation, the design of the weather monitoring network was analyzed and a feedback loop using active learning was designed to detecting potential new measurements sites on the areas of uncertainty of the model. The results showed a clear underrepresentativeness of topographical information related to exposition and slope.

Acknowledgements The authors would like to thank Loris Foresti (Australian Bureau of Meteorology) for the precious help in preparing data and meteorological interpretation of results. The work was supported by the Swiss National Foundation (PZ00P2-136827), by the Strategic Research Cluster grant (07/SRC/I1168) and Stokes Lectureship award by Science Foundation Ireland under the National Development Plan.

References

1. B. Boser, I. Guyon, and V. Vapnik. A training algorithm for optimal margin classifiers. In *5th ACM Workshop on Computational Learning Theory*, pages 144–152, Pittsburgh, USA, 1992.
2. D. Cohn, L. Atlas, and R. Ladner. Improving generalization with active learning. *Mach. Learn.*, 15(2):201–221, 1994.
3. B. Settles. Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin-Madison, 2010.
4. D. Tuia, M. Volpi, L. Copa, M. Kanevski, and J. Muñoz-Marí. A survey of active learning algorithms for supervised remote sensing image classification. *IEEE Journal of Selected Topics in Signal Processing*, 5(3):606–617, 2011.
5. D. Tuia, S. Joost, and A. Pozdnoukhov. Active multiple kernel learning of wind power resources. In *Machine Learning for Sustainability*, NIPS Workshop, Granada, Spain, 2011.
6. W. Müller, L. Pronzato, and H. Waldl. Beyond space-filling: an illustrative case. Technical Report 2011-53, Institut fr Angewandte Statistik (IFAS), 2011.
7. L. Pronzato and W. G. Müller. Design of computer experiments: space filling and beyond. *Statistics and Computing*, in press.
8. G. Schohn and D. Cohn. Less is more: active learning with support vectors machines. In *Intl. Conf. Mach. Learn. ICML*, pages 839–846, Stanford, USA, 2000. Morgan Kaufmann.