

Data-driven topo-climatic mapping with machine learning methods

A. Pozdnoukhov · L. Foresti · M. Kanevski

Received: 31 October 2007 / Accepted: 19 December 2008 / Published online: 16 January 2009
© Springer Science+Business Media B.V. 2009

Abstract Automatic environmental monitoring networks enforced by wireless communication technologies provide large and ever increasing volumes of data nowadays. The use of this information in natural hazard research is an important issue. Particularly useful for risk assessment and decision making are the spatial maps of hazard-related parameters produced from point observations and available auxiliary information. The purpose of this article is to present and explore the appropriate tools to process large amounts of available data and produce predictions at fine spatial scales. These are the algorithms of machine learning, which are aimed at non-parametric robust modelling of non-linear dependencies from empirical data. The computational efficiency of the data-driven methods allows producing the prediction maps in real time which makes them superior to physical models for the operational use in risk assessment and mitigation. Particularly, this situation encounters in spatial prediction of climatic variables (topo-climatic mapping). In complex topographies of the mountainous regions, the meteorological processes are highly influenced by the relief. The article shows how these relations, possibly regionalized and non-linear, can be modelled from data using the information from digital elevation models. The particular illustration of the developed methodology concerns the mapping of temperatures (including the situations of Föhn and temperature inversion) given the measurements taken from the Swiss meteorological monitoring network. The range of the methods used in the study includes data-driven feature selection, support vector algorithms and artificial neural networks.

Keywords Machine learning · Support vector machines · Topo-climatic mapping · Feature selection · Environmental modelling · Downscaling · Decision support systems · Natural hazards

A. Pozdnoukhov (✉) · L. Foresti · M. Kanevski
Institute of Geomatics and Analysis of Risk, University of Lausanne, Lausanne, Switzerland
e-mail: Alexei.Pozdnoukhov@unil.ch
URL: <http://www.unil.ch/igar>; <http://www.geokernels.org>

L. Foresti
e-mail: Loris.Foresti@unil.ch

1 Introduction

The recent technological progress in automatic environmental monitoring and sensor networks provides large and ever increasing volumes of data. The data coming from these networks become one of the main sources of information in environmental sciences. The amounts of data often exceed the abilities of the physical models to incorporate all the available information. Both at the levels of global climatic models, regional forecasting and downscaling, data assimilation becomes one of the most important aspects of environmental and climate modelling (Evensen 2006). In this context, statistical data-driven modelling becomes an important research issue for the problems of topo-climatic mapping (downscaling, gridding, nowcasting). Related to natural hazard research, the increasing amount and accessibility of meteorological observations including the recordings of temperatures, precipitation, wind speed open numerous possibilities. In the operational setting of natural hazard assessment and mitigation in the extreme weather conditions and storms, this information can be used to produce the spatial meteorological maps in real time, which, in turn, would be helpful in decision making. The computational resources required to run a physical model in a region of complex topography at the fine spatial resolution required for decision making in the real time are hard to provide. At regional scales, statistical and data-driven methods are superior to physical models (Daly et al. 1994). Particularly, the task of topo-climatic mapping—spatial predictions of climatic and meteorological variables at local scales—can be efficiently approached in data-driven manner (Dobech et al. 2007). However, the data come noisy and contain outliers, missing values and gaps. Besides the computational efficiency to produce the prediction maps for real-time decision making, one would require these algorithms to be robust to noise and be able to model the non-linear dependencies in non-parametric way without intensive user interaction.

The complex topographies of the mountainous regions highly influence the atmospheric processes (Geiger 1965; Atkinson 1981; Whiteman 2000). These relations, possibly regionalized and non-linear, can be useful for data-driven modelling with the help of data available from digital elevation models (DEMs). Numerous terrain features such as slope, aspect, curvatures can be computed from modern DEMs with sufficient precision. The analysis of the correlation, linear or non-linear, between DEM features and the variable under study (for example, temperature) can help to analyse these relations and use them for predictions. Traditionally, such approaches are based on the use of regression methods which incorporate some topographic indices (Gallant and Wilson 2000). In machine learning (ML), it is approached from a non-parametric perspective by using the feature selection methods. They are aimed at the data-driven automatic selection of the set of relevant features for a particular prediction problem to enhance the interpretability and quality of the predictions.

This article presents an integrated data-driven methodology of topo-climatic mapping with modern ML methods. It is illustrated using a series of real case studies, providing the detailed description of the modelling steps, comparisons to the traditional spatial interpolation methods and discussions.

In the following sections of this article, the reader will first (Sect. 2) find an introduction to “learning from data” paradigm and the statistical learning theory (SLT), which provides a solid breakthrough away from the “black-box” style modelling. Section 2.1 explains the support vector machine (SVM) learning models followed by a short description of neural network (NN) models in Sect. 2.2. Section 2.3 provides an introduction to the ideas of feature selection. Section 3 presents the real case studies. It deals with the topo-climatic

mapping of temperatures in Switzerland following the data from the MeteoSwiss network of meteorological stations. In this case study, the ML approach appears to be a powerful tool for fast predictive modelling of temperatures. The considered cases include the modelling of long-term mean air temperatures, the anomalously high winter temperatures during the extreme Föhn effect caused by Kyrill storm in Europe in January 2007 and severe temperature inversion often observed in mountainous regions. Though the presented case studies are mainly aimed at illustrating the methodology of data-driven modelling and the potential of ML methods, the modelling of these particular situations finds important applications in natural hazard assessment. Spatial prediction of long time scale temperatures provides useful information to analyse anomalies like heat waves. The modelling of the Föhn process is of particular importance due to the related heavy winds and orographic precipitation. Finally, mapping of the temperature inversion phenomenon is essential to classify areas where frost can happen during the morning. On the other hand, the study of this phenomenon, which occurs in stable atmospheric conditions, could be helpful for air pollution modelling when pollutants accumulate in the ground layer of the atmosphere. The final remarks and discussions conclude the article.

2 Learning from environmental data

The growing amount of multi-dimensional information coming from contemporary environmental monitoring networks asks for corresponding tools. The domain of the spatial processes, usually considered as 2D or 3D space and approached with geostatistical models to estimate a relation $\mathbf{x} \rightarrow y$ ($\mathbf{x} = \{x^1, x^2\} \in R^2$ or $\mathbf{x} = \{x^1, x^2, x^3\} \in R^3$, $y \in R^1$), is now extended with, for example, terrain and geomorphological features available from DEMs. Geographical Information Systems can be very useful to handle large datasets, which include the available data on, for example, land use and land cover, and to provide the derived and pre-processed information such as terrain slope and aspect. The desire to incorporate this information into data-driven prediction increases the dimensionality of the data space and forces for estimating the $R^M \rightarrow R$ dependencies, $\mathbf{x} = \{x^1, x^2, \dots, x^M\} \rightarrow y$. From one hand, it facilitates the modelling of complex patterns by bringing useful information. At the other hand, statistical assessment in high-dimensional spaces necessarily encounters with the “curse of dimensionality” (Hastie et al. 2001; Breiman 2001). More clearly, this situation is found while attempting to incorporate the data from multi-band remote sensing images that bring potentially another few hundreds of input dimensions to the analysed information.

The contemporary approaches devoted to extracting knowledge from the finite set of high-dimensional empirical data samples are based on the “learning from data” philosophy (Vapnik 1995; Cherkassky and Mullier 1998). The challenges in learning from data in the fields of signal processing, pattern recognition, computer vision and data mining have led to a revolution in the statistical sciences during the last decade (Breiman 2001). Given considerable advances in the developments in environmental monitoring systems, distributed sensor networks and remote sensing, the use of the ML techniques reveals remarkable potential to incorporate these data sources in environmental predictive models.

A number of advantageous features of ML methods can be outlined when comparing them to more traditional approaches. Mainly developed for high-dimensional data, the ML methods aim at being independent of the dimensionality of the input space in order to avoid the over-fitting to high-dimensional data. Particularly tailored to overcome the curse

of dimensionality are support vector algorithms, which were found to behave well in numerous applied problems (Meyer et al. 2003). They are furthermore intended to deal with non-linear problems in a robust and non-parametric way.

Machine learning methods provide a way of directly incorporating additional information as an input for a learning algorithm. In geostatistics, one of the most well-established approaches for working with spatially distributed data in recent times (Cressie 1993; Chiles and Delfiner 1999), the increasing dimensionality of the input space endows the researcher with the need for higher-dimensional variogram (covariance) models. Another way to incorporate additional data in the model-based approach of geostatistics is to increase the number of the co-variables y (or “outputs” in the ML terminology). Then, the co-kriging style modelling would require modelling numerous co-variograms, which is hardly feasible in practice. The closest link between geostatistics and ML is established in the field of Gaussian processes (Rasmussen and Williams 2006).

Machine learning methods open promising perspectives for environmental data modelling and have a wide and exciting field of applications. A number of applications of data-driven and model-free approaches to solve environmental problems can be found in Cherkassky et al. (2006), See et al. (2007) and Kanevski (2008).

2.1 Support vector learning

2.1.1 Statistical learning theory

Statistical learning theory is devoted to the problem of extracting knowledge from a finite number of empirical observations (Vapnik 1998). Similar to non-parametric statistics, SLT aims to develop non-parametric data-driven models. It formalizes the notion of complexity of the learning functions and provides principles for constructing practical methods. These methods both describe the relationships in available empirical data (called training data below) and have good generalization abilities for prediction. Obviously, the successful predictive algorithms are those that provide accurate estimations at the new (validation) points, where the desired quantity is unknown (Hastie et al. 2001; Cherkassky and Mullier 1998).

How well an algorithm can generalize from a given training data set to predict values of the previously unseen (validation) samples can be measured with the expectation of the loss over the ensemble of the validation data. This value is called the *risk* in terms of SLT. This term should not be confused with the one used in environmental risk assessment. The following bounds on the generalization error or risk R are derived by SLT:

$$R(h) \leq R_{\text{emp}}(h) + R_{\text{conf}}(h), \quad (1)$$

where R_{emp} is an empirical risk, or simply the misfit to the training data, and R_{conf} is a confidence term which penalizes the excessively complex models.

Both terms in the bound (1) depend on the “complexity” h of the learning algorithm. This notion of complexity is an important one and is explained hereafter in more details. The process of learning can be seen as the choice of the most appropriate function $f(x, \lambda)$ from the available set $F(\Lambda) = \{f(x, \lambda), \lambda \in \Lambda\}$. The complexity of the algorithm $f(x, \lambda)$ can be controlled by the choice of the vector of hyper-parameters λ of the modelling functions in the available set, defined by the set Λ of their admissible values. To allow a comparison of the functions in the set, these need to be characterized by a single parameter defined here as the Vapnik–Chervonenkis dimension (VC dimension) of the modelling functions (Vapnik 1998). The VC dimension is plotted on horizontal axis in Fig. 1, while the vertical

axis corresponds to the value of risk. Let us consider the case where the complex model (h is large) can fit any given dataset, a situation that is typically defined as over-fitting. There is no evidence that such a model can generalize well the problem at hand, and the confidence term will remain very large here. On the other hand, a model that is overly simple cannot fit the given data and capture the dependencies of the modelled process: although the confidence term of such models is low, the empirical risk is too high.

The strategy for constructing a learning algorithm is thus to find a trade-off between the model complexity and its fit to the data. This can be achieved by minimizing the training error while maintaining h small (see Fig. 1).

This idea is called structural risk minimization (Vapnik 1998). It provides a general approach to learning from data. Particularly, it was implemented with great success in a family of support vector algorithms.

2.1.2 Support vector machines

A new learning method called support vector machine (SVM) was introduced in the early 1990s (Boser et al. 1992; Cortes and Vapnik 1995). It is based on the structural risk minimization principle of SLT. SVM handles the complexity of discriminating functions used to solve the binary classification problems. Initially proposed to solve two-class classification problems, it was generalized later on to deal with multi-class classification, regression tasks (Smola and Scholkopf 2004), as well as estimations of probability densities. For what concerns their application to spatial data, learning methods based on SVMs were applied to various tasks such as the classification of soil-types, the estimation of contamination levels, the prediction of medium porosity, the predictive mapping of contaminant concentrations, etc. (Kanevski et al. 1996; Kanevski and Canu 2000; Kanevski et al. 2002; Pozdnoukhov 2005). More recently, the use of SVMs also included applications in natural hazard research by approaching snow avalanche danger prediction (Pozdnoukhov et al. 2008), landslide susceptibility modelling (Brenning 2005), hydrological discharge and runoff models (Asefa et al. 2006), downscaling of precipitation (Tripathi et al. 2006). SV-based regression models also showed promising results when used in conjunction with geostatistics (Kanevski and Maignan 2004).

Support vector machines provide non-linear and robust solutions by mapping the input space into a higher-dimensional feature space using kernel functions. This method has the

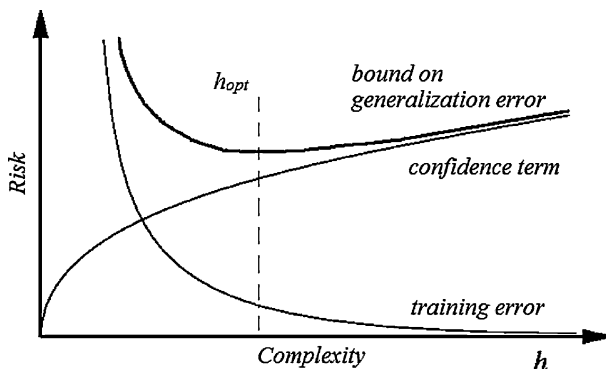


Fig. 1 Bound on generalization error derived in SLT (following Vapnik 1995). The minimum corresponds to the optimal complexity h_{opt} of the model for a given prediction problem

advantage of placing into the same framework, some of the most widely used models such as linear and polynomial discriminating surfaces, feed-forward NNs and networks composed of radial basis functions. When solving classification problems, SVMs provide the classification directly, without solving a more general task of modelling class densities at an intermediate step. Focusing on the marginal and most discriminative data samples, SVMs provide sparse models, i.e. only a (small) subset of data contribute to the classification model. These data samples, called support vectors, usually lie close to the decision surface.

The SVM classification algorithm was initially derived for the linear discriminating surfaces—hyper-planes. It was shown that in order to minimize the complexity of the linear discriminative model one has to maximize the margin it produces between the samples of different classes. More details on support vector for classification can be found in the tutorial (Burges 1998).

2.1.3 Support vector classification

Based on the training set $\{(x_1, y_1), (x_2, y_2), \dots (x_N, y_N)\}$ of high-dimensional i.i.d. input vectors x_i and output measurements y_i , the basic model in support vector methods is a kernel expansion:

$$f(x, \alpha) = \sum_{i=1}^N \alpha_i K(x, x_i) + b, \quad (2)$$

where b is a constant threshold and α_i the weights to be optimized using the training data. For the sake of writing simplicity, α denotes the whole set of the weights $\{\alpha_i, i = 1, \dots, N\}$. $K(x, x_i)$ is a *kernel function*. The model (2) corresponds to a linear model $f(x, w) = wx + b$, given that w is expressed as a linear combination of training samples $w = \sum_{i=1}^N \alpha_i x_i$, and the dot products are substituted with the kernel function: $(x, x_i) \rightarrow K(x, x_i)$. Kernel functions and the kernel “trick” are as much important in support vector learning as the idea of complexity control. Kernel functions are the symmetric positive-definite functions that satisfy the Mercer conditions (Aronszajn 1950). They provide a way for computing dot products in possibly infinite-dimensional spaces (reproducing kernel Hilbert spaces, RKHS). The kernel trick consists in the substitution of dot products between the samples in the input space with the kernel function. A linear algorithm which is formulated in terms of dot products between the samples can therefore be directly turned into its non-linear extension (Scholkopf and Smola 2002).

Consequently, the linear model in some high-dimensional feature space corresponds to the non-linear model in the input space. This duality is a remarkable property of the support vector algorithms.

Because the parameter(s) of the kernel are the hyper-parameter(s) of the SVM, these should be tuned using the available knowledge and data. The usual criterion to tune the parameters of the kernel function is the cross-validation or m -fold cross-validation error, or the testing error if there is enough data to split it into training and testing subsets.

Gaussian radial basis function,

$$K(x, x') = e^{-\frac{(x-x')^2}{2\sigma^2}}, \quad (3)$$

is used as a kernel in many practical problems. It operates with the pair-wise distances between samples. Its bandwidth σ , which is acting here as a hyper-parameter, is

proportional to some characteristic length scale implied by the data. The properties of the model with RBF kernel will follow the behaviour shown in Fig. 1, since the model complexity increases as the value of σ decreases and vice versa. Being a scalar product in RKHS, the kernel function may incorporate prior physical knowledge on the similarities between data samples, though further research is required to elaborate this issue. Another parameter to tune is an upper bound C of the weights α in the expansion (1). It defines a trade-off between fit to data and model complexity.

The output of the classifier can be post-processed with a soft-max transformation to achieve the probabilistic interpretation (Platt 1999).

2.2 Neural networks

Neural networks were the first models developed in the field of ML. Though they do not account for the principle of model complexity control later introduced by SLT, these are widely used for data-driven modelling. Being particularly adapted for non-linear regression, they are used as universal interpolators in this study. A large corpus of literature exist which provide its comprehensive description (Haykin 1998). A brief introduction into these models which proved to be useful for topo-climatic mapping is given below.

2.2.1 Multi-layer perceptron

Multi-Layer Perceptron (MLP) is one of the most popular and traditional models of NNs (for example, Haykin 1998). The main component of MLP is the neuron, a unit which takes a linear combination of the inputs and performs a transformation via the activation function, which is responsible for non-linearity. The number of input neurons is equal to the dimension of the input space. The particular MLP structure used for spatial prediction with no additional information would contain an input neuron for every spatial coordinate; several hidden neurons are arranged in one or two layers and responsible for modelling non-linearity and the output neuron(s) representing the target variable(s).

The number of hidden neurons, which is related to the complexity of the NN model, is subject to optimal configuration for a particular case study. Choosing too many hidden neurons will then lead to over-fitting when MLP loses its ability to generalize the information from the samples. The application of MLP proceeds through a training phase, when the model learns the pattern from the data. This training employs the minimization of the quadratic mean square error (MSE) on the training data. Error back-propagation algorithm is applied to calculate gradient of MSE and adapt the connection weights between neurons.

2.2.2 General regression neural network

General regression neural network (GRNN) is a NN interpretation of the statistical non-parametric Nadaraya–Watson regression estimator (Specht 1991). It is based on Parzen–Rosenblatt kernel density estimation (Parzen 1962). GRNN estimate can be written as a weighted sum of all available measurements. The final formula for regression estimation is

$$f(x) = \frac{\sum_{i=1}^N f_i K_{\sigma}(x, x_i)}{\sum_{i=1}^N K_{\sigma}(x, x_i)},$$

where K is a kernel function and the smoothing parameter σ is a positive number called bandwidth or simply width. It controls the size of the kernel and corresponds to some characteristic scale implied by the data. The most often used type of the kernel is the Gaussian type.

The estimate depends on the spatial distribution of the initial data through pair-wise distances and the smoothing parameter σ . Generally, the smoothing parameter can be considered as a vector of $\sigma = (\sigma_x, \sigma_y, \dots)$, allowing to take into the account possible data anisotropy.

GRNN training consists in tuning the value of smoothing parameter in a way to minimize the cross-validation MSE on the training data. This is a reasonably fast procedure which can be done automatically (Timonin and Savelieva 2005).

2.3 Feature selection

Large number of input variables may bring both information and noise. By removing the irrelevant and redundant features from the data, the performance of data-driven models can be improved. An important benefit is also the chance to improve model interpretability: a better understanding about the data and importance of the features can be achieved. The most obvious variables to consider in a spatial context are the coordinates. The elevation level can be a relevant feature in a number of studies. The selection of relevant features among the available ones in order to improve the model is the task of feature selection.

In statistics, forward and backward stepwise selection, Akaike information criterion, Bayesian information criterion and several other methods were proposed. In ML, feature selection attracts considerable attention nowadays (Guyon and Elisseeff 2003). Feature selection is closely linked with dimensionality reduction. The well-known statistical method of principal component analysis (PCA) can be used to reduce the dimensionality of the input space. It finds statistically uncorrelated features by building a linear combination of the initial inputs. Such approaches are referred to as feature extraction, contrary to the feature selection, since they exploit the transformation of the original features instead of eliminating the redundant ones. The whole set of initial features is required to obtain the PCA representation for every new data sample.

Feature selection methods are often algorithm-specific. The method which will be used in the case study below is a recursive feature elimination developed specifically for SVM (Guyon et al. 2002).

3 Topo-climatic mapping

This section is devoted to the application of the ML techniques for the problem of topo-climatic mapping. The detailed case studies presented here are devoted to air temperature mapping. Being related to natural hazards as a basic meteorological parameter of weather conditions and related hazardous events, it is also an important cause of hazard by itself. It concerns extreme events such as heat waves, temperature anomalies, unexpected frost. Besides, it provides a very illustrative study to highlight the methodology of data-driven modelling with ML.

Depending on the specific meteorological conditions, surface air temperature is related to topography in a different way. Machine learning methods were here applied to detect these relations and produce predictive models by capturing this information from empirical

data without applying the physical knowledge. In order to take into account some particular temperature situations that may cause hazards, three time aggregation periods (monthly, daily and hourly) were considered.

The presentation of applications below is organized in a series of case studies which illustrate the methodology by first starting with a simple example followed by more complicated studies. The classical spatial interpolation methods, such as deterministic inverse distance weighting and geostatistical kriging models, are compared to ML methods in Sect. 3.2.1. Models which use 2D inputs (spatial coordinates) and 3D (spatial coordinates and elevation) are considered in this context. Section 3.2.2 describes the results of mapping the air temperature with MLP when the data reveal linear relations with elevation, which significantly varies locally due to the specific meteorological conditions (Föhn).

In Sect. 3.2.3, SVM classifier is used to model the temperature inversion phenomenon. The phenomenon can only be modelled using the information about the land forms that are computed from DEM. SVM is applied to predict the probability of temperature inversion using the high-dimensional data set of only 107 measurements. This problem setting easily leads most models to over-fit the data. Being prone to over-fitting, SVM is one the most adequate technique to use for this task. This case study also illustrates the use of feature selection techniques for binary and probabilistic predictions and the descriptive analysis of the phenomenon at hand. The final temperature mapping is obtained with GRNN using the temperature inversion probability, the elevation and the spatial coordinates as inputs.

3.1 Data description

The meteorological observation networks of MeteoSwiss (Swiss Federal Office of Meteorology and Climatology) allow observing meteorological phenomena in Switzerland including the region of Swiss Alps. A large number of meteorological parameters (air temperatures, humidity, pressure, wind speed, insolation, etc.) are regularly recorded at the maximum rate of a measure every 10 min. The data may include gaps, noise and outliers. The measurement stations of the network are not unified, and the monitoring networks for different variables may appear to be quite different. The meteorological network measuring air temperature (at 2 m above the ground) has 107 stations situated in the altitude range of 190–3,600 m above sea level (Fig. 3).

3.1.1 Air temperature and elevation

Air temperature in the lower atmosphere is known to decrease with altitude. Usually, this decrease is linear and the gradient is roughly $0.65^{\circ}\text{C}/100$ m of elevation. However, this is a simplification which assumes the adiabatic atmosphere conditions (Atkinson 1981). Such factors as terrain type, exposition to the sun, local winds, latent heat of precipitation and the ground influence the air temperatures bringing the complex non-linear relations into the observed data. So, the situations when the linear correlation is not observed encounter often. Nevertheless, when aggregating air temperature at long time scales, the mean air temperature over a time period shows linear gradients since the influence of the local effects diminishes due to averaging.

Let us consider the mean monthly, maximum daily and mean hourly temperature measurements. Figure 2 clearly illustrates the described situation. While the mean monthly temperature is clearly linearly correlated with altitude (Fig. 2 on the left shows the mean temperatures in August 2005 in Switzerland), the relation is much more complex for the daily and hourly observations. Figure 2 (centre) shows the maximum temperatures for the

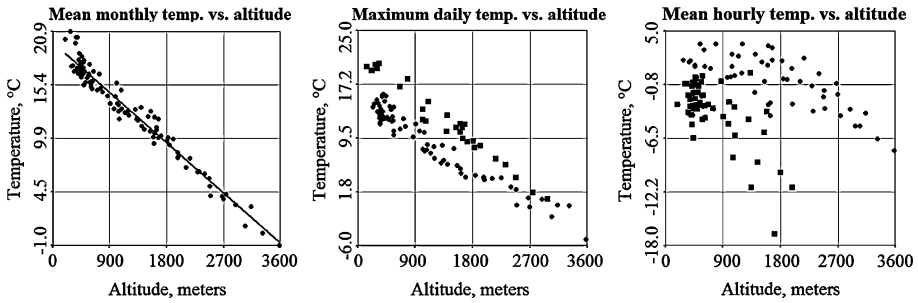


Fig. 2 An example of mean monthly (*left*), daily (*centre*) and hourly (*right*) temperatures versus elevation measured by 107 meteorological stations in Switzerland. For longer observation periods, the influence of local effects vanishes, resulting in a well-known linear correlation with a temperature gradient of roughly $0.65^{\circ}\text{C}/100\text{ m}$. The complex non-linear dependencies observed at shorter time scales are more difficult to model

period of 9 a.m. to 6 p.m., 19.01.2007, the day when the warm north wind (Föhn) on the southern side of the Alps substantially increased the temperature, which reached at 2 p.m. the record value of 24°C for winter months in Ticino. The observations influenced by Föhn are marked with square marks in the figure.

On the right of the Fig. 2, the hourly average temperatures between 6:30 and 7:30 a.m. on 5.02.2007 reveal almost no correlation with altitude. The process which determines this complex relation of temperatures and altitude is a temperature inversion phenomenon. It is regularly observed in mountainous regions under stable atmospheric conditions and appears to be one of the most challenging situations for temperature modelling. The observations which were made under the inversion layer (as qualified by an expert) are shown with square marks.

For producing the spatial maps of air temperature, the use of ML is motivated by the fact that when the data aggregation time is reduced (monthly to hourly), the relationships to topography increase in complexity and require appropriate tools to be modelled.

3.1.2 Digital elevation model

The DEM of Switzerland used in this study is available through the Swiss Federal Office of Topography. It has the resolution of 250 m. The DEM and the locations of temperature measurements (meteorological stations) are shown in Fig. 3. While the exact locations of the stations are known, they obviously do not necessarily match with the DEM grid cells. It was taken into account when computing the geomorphologic features in their vicinities.

3.2 Air temperature mapping

In this section, data-driven air temperature prediction maps will be obtained using ML methods. A simple illustrative case study of mean monthly temperature faces a linear correlation with elevation (Sect. 3.2.1). The precise and reliable modelling of mean air temperatures could help to analyse the anomalies from the 30 years reference period actually used in Switzerland (1961–1990) and especially for particular situations as it was the heat wave of 2003. The more complicated case concerns modelling the non-linear

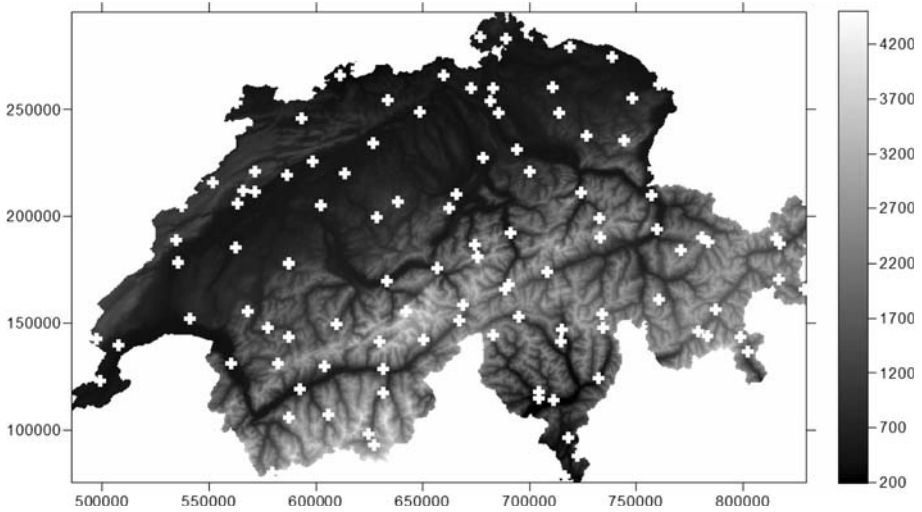


Fig. 3 DEM of Switzerland of 250 × 250 m resolution. The locations of meteorological stations are shown with *white cross marks*. There are a total of 107 stations located at the altitudes between 200 and 3,600 m

regionalized relations between instant air temperature and geomorphologic features caused by temperature inversion (Sect. 3.2.3). The models were compared by using a root mean squared error computed on the reserved validation data subset.

3.2.1 Mean monthly temperatures: global linear gradients

A variety of deterministic, geostatistical and ML tools were applied for the task. The models which use 2D inputs (*X* and *Y* spatial coordinates only) were not able to model the temperatures precisely (Fig. 4a). This is the case for both traditional and ML methods. High linear correlation (Fig. 2, left) between air temperature and elevation (*Z*) facilitates the modelling. An example of MLP mapping performed using *X*, *Y* and *Z* coordinates as the inputs is presented in Fig. 4b. The optimal structure of MLP is the one with three neurons in the hidden layer (3-3-1). The linear output was used to force the linear decrease of temperature with altitude in extrapolation areas (>3,600 m).

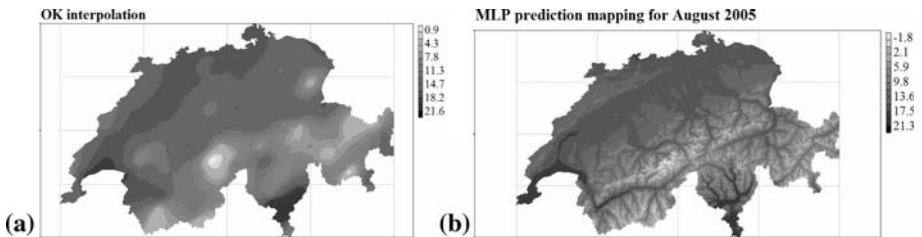


Fig. 4 **a** Ordinary kriging using spatial coordinates. **b** MLP prediction mapping of temperature using both spatial coordinates and elevation level

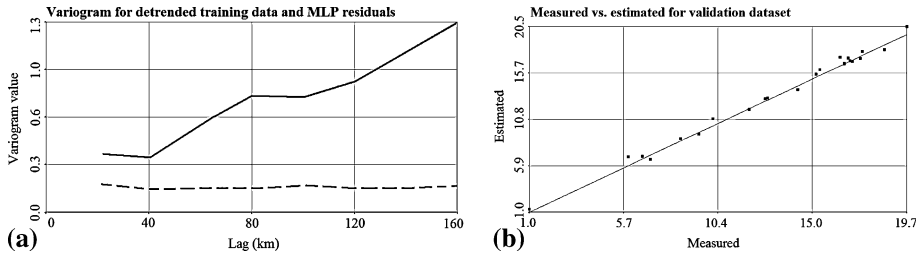


Fig. 5 **a** Variogram of the de-trended training data and the validation residuals (the difference between measured and predicted values) of MLP model. The validation residuals demonstrate pure nugget effect indicating that all spatially structured information was extracted by MLP from the data. **b** Measured values versus the MLP estimates of the validation dataset

Table 1 Models under comparison in terms of training and validation root mean square errors

Method	Temperature, August 2005						
	OK (2D)	IDW (2D)	RegrOK	KrigED	SVR	GRNN	MLP
Train RMSE	4.27	4.12	0.57	0.51	0.45	0.53	0.41
Valid RMSE	4.55	4.57	0.48	0.39	0.47	0.65	0.46

The models under comparison are: OK (2D), ordinary kriging done with 2D input data; IDW, inverse distance weighting; RegrOK, ordinary kriging done on the residuals of the linear regression with elevation; KrigED, kriging with external drift; SVR, support vector regression; GRNN, general regression neural network in fully automatic mode. The training root mean square error ($^{\circ}\text{C}$) of kriging models and IDW is obtained as cross-validation over training data. The results of the comparison are summarized in Table 1. Validation predictions obtained with MLP are illustrated in Fig. 5.

These results suggest the following conclusions. The validation errors of IDW and ordinary kriging using 2D inputs are higher than 4°C , i.e. eight times the errors of ML models with 3D inputs, kriging with an external drift (KrigED) and regression kriging (RegrOK). As expected, 2D input data provide insufficient explanatory power to model temperatures in mountainous regions. The models which use elevation provide comparable performance, with KrigED being the best of the rivals. GRNN provided slightly worse results; however, it was applied in fully automatic mode without any user interaction.

Conventional deterministic and geostatistical methods are useful to interpolate temperatures only when the relationship with the elevation is linear or can be parametrically modelled by user as an external drift. Other situations, where the topography-related relationships become non-linear, will be approached below by means of ML methods.

3.2.2 Regionalized linear trends

In this part of the study, the aggregation time was reduced and hence the relationship to the elevation became more complex (Fig. 2, centre). One of the effects of the mountain relief is that the temperature-elevation gradients may vary in space. A frequently occurring example is a Föhn situation, which happens when a humid air mass converges to a considerable relief. The topography forces the air mass to ascend and to condense, creating clouds. The condensation liberates the latent heat which is gained during the vaporization process over the sea. The release of the latter heats the atmosphere causing a linear

temperature gradient nearby $0.5^{\circ}\text{C}/100\text{ m}$ in the up-wind versant. In the down-wind versant, the dry air descends and heats with an adiabatic linear temperature gradient of about $0.98^{\circ}\text{C}/100\text{ m}$. The repercussion of these physical processes is a different temperature between the two versants at the same elevation. The difference is accentuated if there is an orographic precipitation in the windward side of the mountains.

In order to apply the conventional methods of geostatistics for predicting the temperature in this situation, one would need to avoid the non-stationarity in data. It would be required to extract the temperature trend induced by Föhn effect. Since the trend is not global over the area, it appears to be a complicated problem. The data-driven ML methods would be more suitable for this modelling.

To investigate the data-driven modelling of a Föhn situation (Ambrosetti et al. 2005), the temperatures in Switzerland during the afternoon hours on 19.01.2007 were modelled with MLP. The meteorological characteristics of this day are very particular, because there was a heavy storm over Europe (Kyrill) with north-westward winds over Switzerland exceeding 120 km/h . The low humidity of the air caused a passive Föhn situation without precipitation in the up-wind versant. However, the condensation of the water vapour occurred in the crest of the Alps; this area is supposed to face low gradients of temperature. On the contrary, the south alpine part of Switzerland should have high gradients and this causes an increase of the temperatures until 24°C which is a record for the month of January.

The structure of the NN included three inputs (X, Y, Z), three hidden neurons and one output. The relationship between elevation and temperature was presented in Fig. 2. Two major patterns can be distinguished: the pattern of the north part of the Alps and the pattern of the south part where the gradient is the highest due to the Föhn. They were shown with different marks in the figure.

The MLP prediction map in Fig. 6 confirms the presence of Föhn in the southern part, as well as the local effects of Föhn in the north alpine valleys due to the heavy winds. The map of the temperature lapse rates (Fig. 7, lighter zones correspond to higher gradients) of the model has to be analysed to validate the results considering where the condensation or an adiabatic heat of the air mass takes place. This map was calculated using 278 squared moving windows at the scale of $20 \times 20\text{ km}$. The high gradients in the south part of Switzerland (Ticino) are caused by the Föhn. The local effects of Föhn in Valais (Rhône’s valley), Grisons (Rhine’s valley) and behind the Jura chain can be noticed as well. The Alps in general presents low gradients due to the condensation. To conclude, let us note that all this information was modelled by MLP in a pure data-driven way taking into account only the relationships between DEM (X, Y, Z) and the temperature. In this case, no other terrain-related inputs were required besides the elevation level. It may appear to be insufficient in a more complicated situation, such as the temperature inversion considered below.

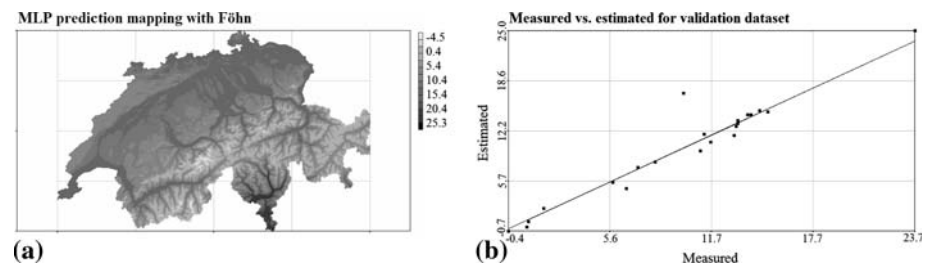


Fig. 6 a MLP prediction map for Föhn. b Measured versus MLP estimated values of the validation dataset

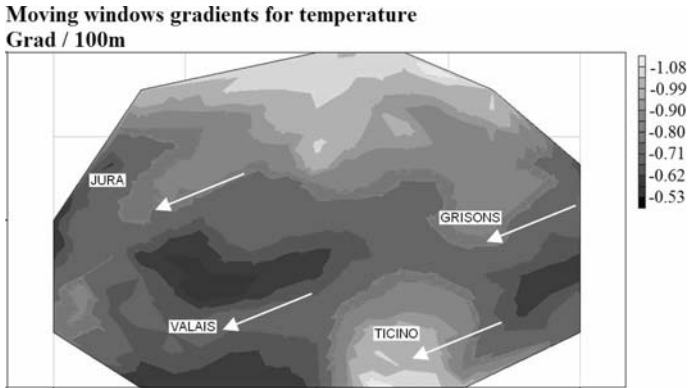


Fig. 7 Temperature lapse rates computed in a moving window over Switzerland

3.2.3 Temperature inversion

In mountainous regions, temperature inversion often occurs in the morning in winter months, during the anticyclones with high air pressure and absence of clouds and precipitation. The simplified scheme of the process is as follows. The land surface, heated by the solar radiation during the day, releases heat during the night (Andre and Mahrt 1982). It cools down causing the cooling of the boundary layer of the near-surface air. In the conditions of stable atmosphere, the cold air masses tend to go downward from the mountain slopes to the adjacent plains forming the cold air “lakes” in the basal layers of the valley atmosphere (Clements et al. 2003). These nocturnal inversions in the boundary layer are quite varied, with depths usually between 100 and 500 m. The inversion layer of anomalously warm air lies above the cold boundary layer. With the increasing altitude, the air gets colder again. In this high (free) atmosphere layers, the conventional linear decrease in air temperature is observed. The scheme of the phenomenon is shown in Fig. 8.

The inversion phenomenon happens locally and at different altitudes. The elevation of the inversion layer is not constant over the whole area, since it depends on the elevation of the bottom of the valley and the topographic characteristics of the region, such as its curvature and the form of the surface (convex/concave). The temperature measurements at

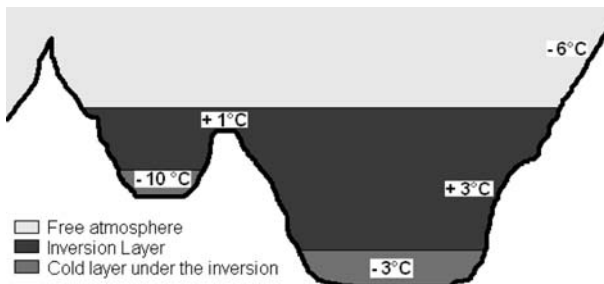


Fig. 8 The scheme of temperature inversion situation. The temperature measured in the valley may be much lower than the temperature measured at the mountain summit, despite the significant difference in elevation level

the surface of the ground are influenced by the inversion layers. The observed temperature–altitude relation becomes as complicated as it is shown on the right of Fig. 2.

The temperature measurements at 7 a.m. in the morning of 05.02.2007 are presented in Fig. 9a. Temperature inversion is observed in the major part of the region, causing the temperatures in some elevated Alpine valleys to drop down to -16°C , while the air temperatures in the free atmosphere even in high mountains do not go below -7°C . Variogram of this temperature data reveals a pure nugget effect, making it impossible to use geostatistical predictors.

To model the influence of inversion on the air temperature model, the useful information provided by the meteorologists would be an indication whether the station which provided a measurement was in the cold near-surface air below the inversion layer or in the free atmosphere. The classification of the stations into two classes is shown in Fig. 9b. By convention, some stations at Swiss Plateau were labelled as the ones influenced by inversion. While this choice may seem to be doubtful, it should be stressed that ML methods can deal with uncertain, imprecise and mislabelled data. The problem which is faced at this point is to provide a map with a probability of a spatial location to be in the cold air layer. It was approached with SVM classifier.

A qualitative classification of the stations was performed manually by exploiting physical consideration. For example, if two neighbouring stations showed temperatures having an inverse relation with elevation (ex: -3°C at 400 m a.s.l. and $+3^{\circ}\text{C}$ at 1,600 m a.s.l.), the station with the lowest temperature was classified in the class “cold air layer under the inversion” and the other one was classified as belonging to the “warm inversion layer”.

The physical conditions for the formation of an inversion layer were described above, and clearly the form of the land surface is one of the main factors to consider. The information about the form of the land surface over all the area of the study including the vicinities of the measurement locations can be computed from the DEM.

3.2.3.1 Feature extraction and selection from digital elevation model The geomorphological features that describe the form of the land surface, its curvature, convexity, inclination can be computed for every location of the considered region, as well as for every measurement location. A total of 21 features were computed, including the slope, different forms of curvatures (plan, profile, tangential), Laplacian, the standard deviation calculated with a moving window of 500 m and the difference of Gaussians. The last feature was computed as a difference of DEMs, derived by applying the smoothing with

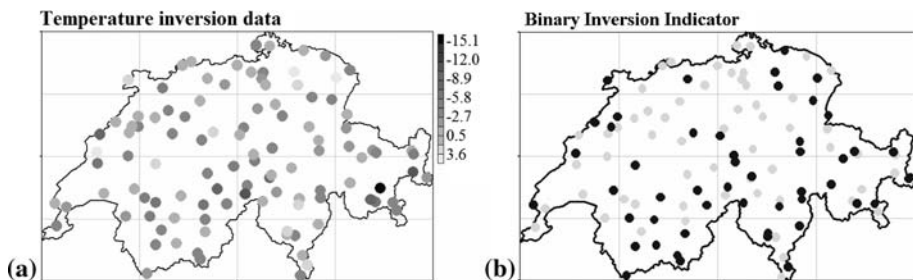


Fig. 9 **a** The measurements of the temperature at 2 m above ground, 7 a.m. on 05.02.2007. Temperature inversion is observed in the major part of the region, causing the temperatures in elevated alpine valleys to drop down to -16°C . **b** The binary classification of meteorological stations was provided by an expert

Gaussian filter of different widths (500, 1000, 3000 and 6000 m). It allows to identify different forms of the relief such as narrow and wide valleys, canyons, mountain chains and ridges and relatively flat surfaces such as lakes and parts of the plateau. An example of the features is shown in Fig. 10.

Some of the computed features are clearly correlated as, for example, slope and local standard deviation. All the differences of Gaussians are pair-wise correlated as well. SVMs do not require the inputs to be uncorrelated, and the computed features can be directly used as inputs. Recursive feature elimination method (Guyon et al. 2002) was applied to select the most relevant six features which provide most information on the inversion phenomenon. The method iterates through the available set of features trying to find the relevant set of features using the properties of SVM and cross-validation estimates on the training data. This is a computationally intensive and quite time-consuming modelling step.

Due to the possible errors in DEM, the derived features may contain incorrect values and noise. To enhance the numerical stability of the computations, the PCA was applied to transform the computed features. Another useful observation was made when visualizing the data in PCA projections (Fig. 11). Note that data are more structured and easily discriminated in PCA projections, compared to the data representation in spatial coordinates. Let us note that PCA step is not compulsory and can be excluded from modelling.

3.2.3.2 SVM classification of temperature inversion regions The input dimension of the data for classification of the temperature inversion regions is equal to eight, which are spatial coordinates X and Y , and six main PCA components. Linear SVM classification

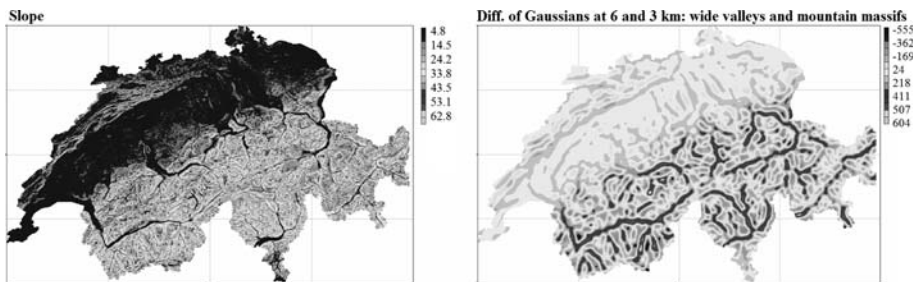


Fig. 10 The slope (in degrees) calculated from DEM. The difference of Gaussians of the widths of 6,000 and 3,000 m provides information on the locations of wide valleys and mountain massifs

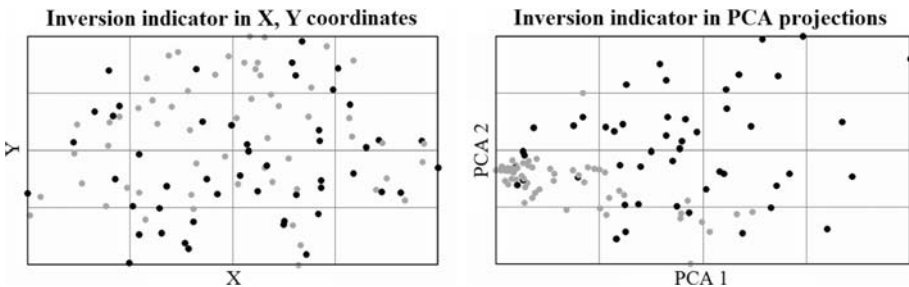


Fig. 11 The X – Y and PCA1–PCA2 projections of the training data. Note that data are more structured and easily discriminated in PCA projections of the selected morphological features than in spatial coordinates

with $C = 100$ provides 15% of training misclassification error, 22% of cross-validation error and 44% of support vectors. Due to the low number of support vectors, an acceptable generalization performance even with this linear model can be expected. However, the relatively high cross-validation error suggests to search for the non-linear classification boundary.

The conventional methodology described in Kanevski et al. (2002) was applied. Gaussian RBF kernel was used, and the parameters of SVM were tuned using the grid search in training error, cross-validation error and the number of support vectors. The error surfaces are presented in Fig. 12.

To avoid over-fitting, the choice of the model parameters has to be made from the region of low cross-validation error. At the same time, a low training error and low number of support vectors are desirable. The considerations lead to the choice of $C = 10$ and $\sigma = 11$. With these parameters, both SVM training and cross-validation error is 15%, and the number of support vectors is 46%.

The prediction mapping of the inversion indicator variable with SVM model and its probabilistic treatment is presented in Fig. 13.

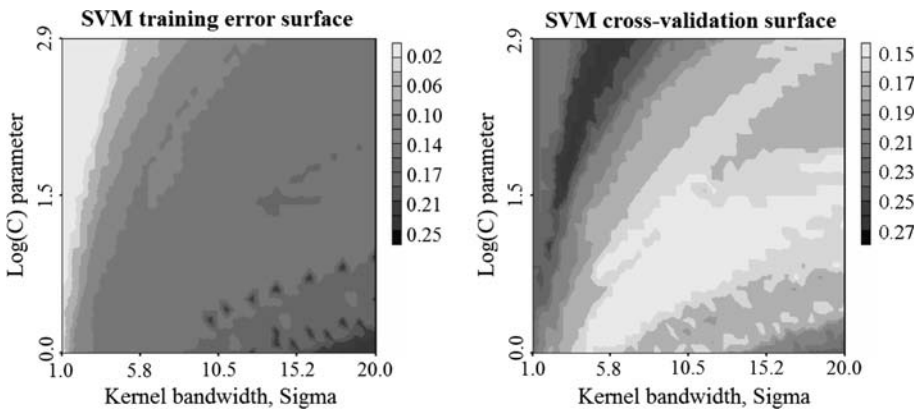


Fig. 12 Training error and cross-validation error surfaces of the SVM with Gaussian RBF kernel. The appropriate choice of the parameters is kernel bandwidth $\sigma = 11$, $C = 10$

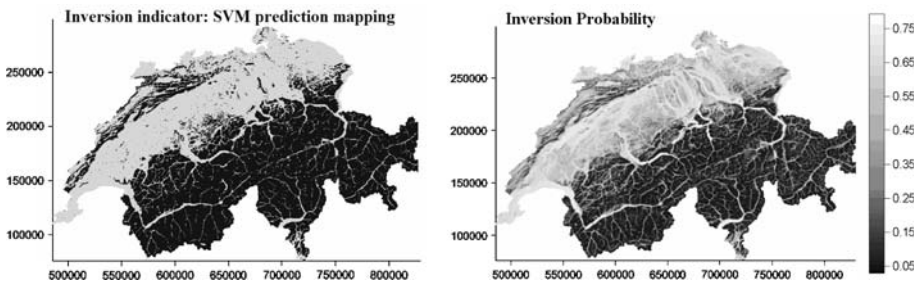


Fig. 13 Left: Prediction mapping of inversion indicator variable produced by SVM with Gaussian RBF kernel. Right: The probabilistic treatment of SVM outputs: the possibility of favourable conditions for cold air accumulation below the inversion in Switzerland at 7 a.m. on 05.02.2007

Due to the uncertain nature of the phenomenon, the probabilistic interpretation of SVM (Platt 1999) provides more adequate interpretation of the developed predictive model. The probability map is presented in Fig. 13 (right). Since the initial data were noisy and partially mislabelled, the probability never reaches the level of 1.

This prediction map can further be used as an input of some spatial mapping method (MLP or GRNN described in Sect. 2.2) to produce a prediction map of the actual temperature.

3.2.3.3 Temperature mapping Three GRNNs were used for temperature mapping: 2D (which only uses the spatial coordinates for temperature modelling), 3D (spatial coordinates and altitude are used) and 4D (spatial coordinates, altitude and the SVM model of inversion probability). The results were compared in using both the prediction maps and validation data of nine samples, which were selected to be representative for the whole area and conditions of the study. All GRNN models were tuned according to the minimum of the cross-validation error on the remaining training set of 98 samples.

The validation results obtained on nine validation samples are presented in Fig. 14 and summarized in Table 2.

The prediction mappings of 2D and 3D GRNNs are presented in Fig. 15.

The prediction mapping of 4D GRNN, which uses spatial coordinates, altitude and probabilistic SVM model of the inversion indicator, is presented in Fig. 16. Temperature inversion was reproduced in a data-driven way from measurements and DEM.

4 Discussion

Statistical data-driven approach to micro- and mesoscale spatial meteorological predictions is indispensable for producing topo-climatic maps. It has a substantial potential to be used in the operational setting for natural hazard assessment and mitigation in the real time,

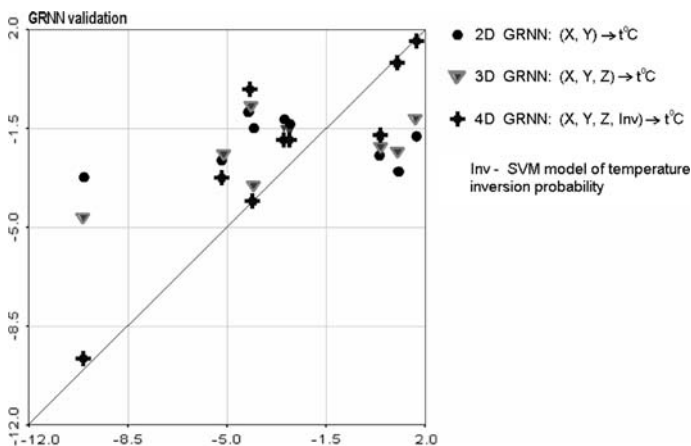


Fig. 14 The validation results of the compared GRNN models. 3D model provides better estimate than the 2D one. 4D GRNN is significantly better than its competitors which do not use the SVM model of the temperature inversion probability. It provides surprisingly good predictions of low and high temperatures both from measurements in free atmosphere and influenced by inversion phenomenon

Table 2 Validation results of the compared GRNN models

Model	Training RMSE	Validation RMSE	Validation Ro
2D GRNN	2.5	3.55	0.23
3D GRNN	1.9	2.7	0.44
4D GRNN	0.8	1.75	0.88

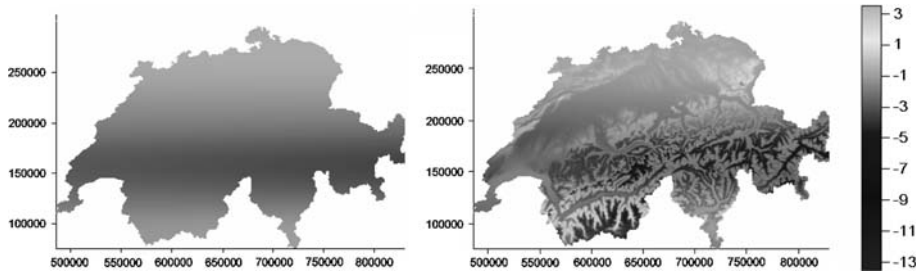


Fig. 15 Prediction mapping results of 2D (*left*) and 3D (*right*) GRNN models. 2D model simply provides the mean temperature, since 2D input data are not sufficient to model the temperature. 3D model provides better results; however, the inversion is not well reproduced and the results are smoothed

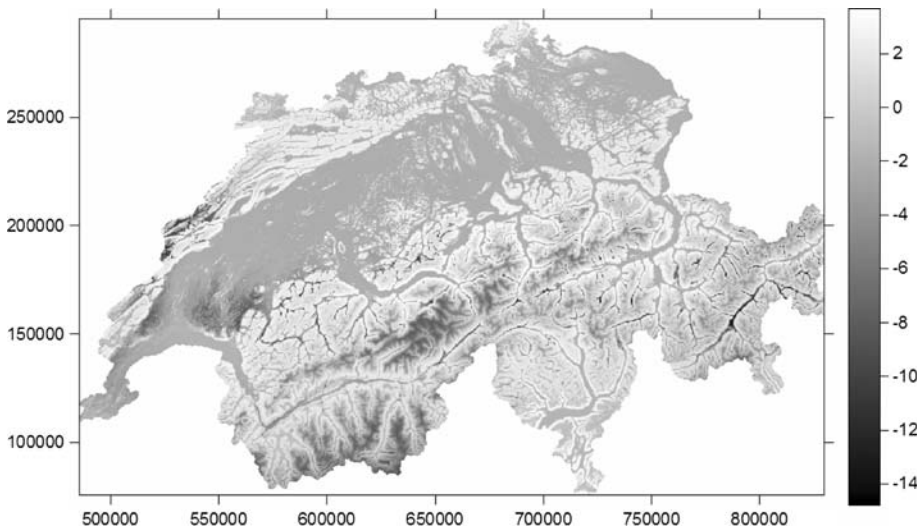


Fig. 16 GRNN model which uses the inversion probability model of SVM (Fig. 14) provides promising mapping results. Temperature inversion is well reproduced

especially in regions characterized by complex topographies. It becomes increasingly important with the growth of the amounts of available data.

A variety of methods can be applied for the task. Geostatistical predictors provide reasonable performance when the information on the phenomenon at hand is spatially

structured. It also applies when the spatial structures can be found after parametric de-trending based, for example, on prior physical knowledge. In the considered case of temperature mapping, the lapse rate based de-trending reveals evident spatial structures (identified by a variogram, Fig. 17) in mean monthly temperature data.

Given a dataset with very descriptive inputs and well-defined patterns, all models, either statistical, geostatistical or ML, would provide reasonable results. The choice of the particular model has to take into consideration practical aspects. GRNN is preferred if automatic mapping is required (Timonin and Savelieva 2005). MLP is the most universal regression tool, but most difficult to control and less appropriate for the interpretation of the results. Due to this model, many other ML methods earned an unjustified “black-box” treatment.

In the shorter time periods, which are more important for approaching the real-time operational use of the methods, the situation complicates. Particularly, the terrain-related dependencies cannot be easily explained. In the considered study on temperature inversion, the data de-trended using conventional lapse rate still demonstrate pure nugget effect (Fig. 17). The solution to this problem is conventionally based on an attempt to build an adequate trend model (Daly et al. 1994). The reasoning at this step ranges from physical laws to common sense heuristics. Though, there are methods of ML especially designed to extract knowledge from empirical data, which can be used as we illustrated in the above series of case studies. It is noticeable that with these techniques, the relevant information on the terrain forms which are favourable for the inversion was extracted from the number of terrain features and the current temperature measurements without relying on a parametric model but based on the mathematically founded principles of learning from data.

The non-parametric data-driven model provided reasonable quantitative accuracy and qualitatively reproduced the phenomenon without taking much computational time as would require the physical model. Though, the incorporation of many data brings both information and noise. The improvement in predictive performance is not guaranteed. Particularly, the high-dimensional GRNN model provided less accurate prediction at one of the nine validation samples, then the baseline approaches (Fig. 14).

It is the non-parametric modelling paradigm which brings both the strength and the drawbacks to the method. Though conventionally accepted in statistical sciences (Breiman 2001), its use is still an open issue in geo- and environmental research, traditionally approached with physical modelling and statistical assessment based on physical heuristics. The critics of non-parametric models for being a “black-box” rather concern the way

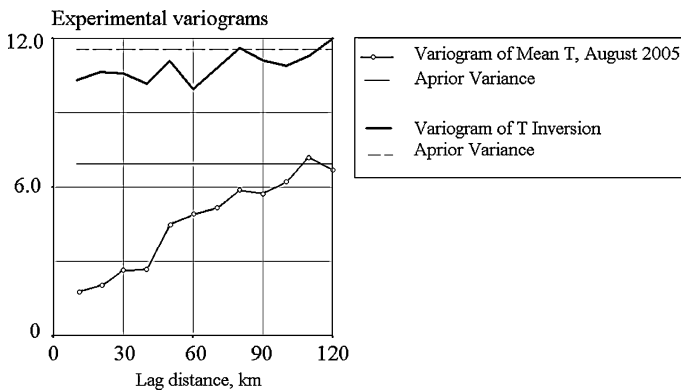


Fig. 17 Experimental omnidirectional variograms for air temperature de-trended at constant lapse rate. Temperature inversion data demonstrate pure nugget effect

they are used and not the essence of the methods. While it remains to be a disputable question, the need for the up-to-date methods to process and extract knowledge from the increasing volumes of meteorological and environmental data is certain.

5 Conclusions

The methodology of applying the ML methods to the task of topo-climatic mapping was considered. It was shown that ML methods are able to extract the non-linear regionalized relations between modelled variables. It makes them suitable for use in the complex topographies of mountainous regions, where physical models find a limited range of applications in real-time decision support mapping and natural hazard assessment. This research extends the use of statistical data-driven modelling, which is conventionally approached with geostatistical and deterministic interpolators that incorporate numerous heuristics to model topography-related dependencies.

The real case study considered in this article was devoted to spatial predictive modelling of temperatures based on the measurements provided by Swiss meteorological monitoring networks. The complex patterns of the temperature data related to the extreme Föhn effect caused by Kyrill storm in January 2007 and severe temperature inversion which is often observed in mountainous regions were successfully modelled.

From a theoretical side, much remains to be done in data-driven environmental modelling. The assimilation of physical knowledge and data-driven approaches is one major and extremely important direction. From a practical side, these approaches need to be done fully automatically (including the feature selection step), robust to noise and interpretable while being fast enough and easy to apply in operational use as a tool for real-time decision support.

Acknowledgements The research was supported by Swiss National Science Foundation Project “Geo-Kernels: Kernel-Based Methods for Geo- and Environmental Sciences” (project no. 200021-113944). The support from Russian Foundation for Fundamental Research through the grant RFFI 07-08-00257 is also acknowledged.

References

- Ambrosetti P, Mariani L, Scioli P (2005) Climatology of North Föhn in Canton Ticino and Western Lombardy. *Riv Ital Agrometeorol* 2:24–30
- Andre JC, Mahrt L (1982) The nocturnal surface inversion and influence of clear air radiative cooling. *J Atmos Sci* 39:864–878
- Aronszajn N (1950) Theory of reproducing kernels. *Trans Am Math Soc* 68:337, 404
- Asefa T, Kembrowski M, McKee M, Khalil A (2006) Multi-time scale stream flow predictions: the support vector machines approach. *J Hydrol* 318(1–4):7–16
- Atkinson BW (1981) Meso-scale atmospheric circulations. Academic Press, London, p 445
- Boser B, Guyon I, Vapnik V (1992) A training algorithm for optimal margin classifiers. In: Fifth annual workshop on computational learning theory. ACM Press, Pittsburgh
- Breiman L (2001) Statistical modelling: the two cultures. *Stat Sci* 16(3):199–231. doi:10.1214/ss/1009213726
- Brenning A (2005) Spatial prediction models for landslide hazards: review, comparison and evaluation. *Nat Hazards Earth Syst Sci* 5(6):853–862
- Burges C (1998) A tutorial on support vector machines for pattern recognition. *Data Min Knowl Discov* 2(2):121–167. doi:10.1023/A:1009715923555
- Cherkassky V, Mullier F (1998) Learning from data. Wiley, New York
- Cherkassky V, Krasnopolsky M, Solomatine D, Valdés J (2006) Introduction to special issue: earth sciences and environmental applications of computational intelligence. *Neural Netw* 19(2):111. doi:10.1016/j.neunet.2006.01.013

- Chiles J-P, Delfiner P (1999) *Geostatistics: modelling spatial uncertainty*. Wiley Series in Probability and Statistics, New York
- Clements CB, Whiteman CD, Horel JD (2003) Cold-air-pool structure and evolution in a Mountain Basin: Peter Sinks, Utah. *J Appl Meteorol* 42:752–768. doi:10.1175/1520-0450(2003)042<0752:CSAEIA>2.0.CO;2
- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20:273–297
- Cressie N (1993) *Statistics for spatial data*, revised edition. Wiley, New York
- Daly C, Neilson RP, Phillips DL (1994) Statistical-topographic model for mapping climatological precipitation over mountainous terrain. *J Appl Meteorol* 33(2):140–158
- Dobech H, Dumolard P, Dyras I (eds) (2007) *Spatial interpolation of climate data. The use of GIS in climatology and meteorology*. ISTE Press, London, ISBN 978-1-905209-70-5
- Evensen G (2006) *Data assimilation: the ensemble Kalman filter*. Springer, Berlin
- Gallant J, Wilson J (2000) Primary topographic attributes. In: Wilson J, Gallant J (eds) *Terrain analysis: principles and applications*. Wiley, Chichester, pp 51–85
- Geiger R (1965) *The climate near the ground*. Harvard University Press, Cambridge, 482 pp
- Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. *J Mach Learn Res* 3:1157–1182. doi:10.1162/153244303322753616
- Guyon I, Weston J, Barnhill S, Vapnik V (2002) Gene selection for cancer classification using support vector machines. *Mach Learn* 46(1–3):389–422. doi:10.1023/A:1012487302797
- Hastie T, Tibshirani R, Friedman J (2001) *The elements of statistical learning*. Springer, New York
- Haykin S (1998) *Neural networks: a comprehensive foundation*. Prentice Hall, Upper Saddle River
- Kanevski M (ed) (2008) *Advanced mapping of environmental data. Geostatistics, machine learning and Bayesian maximum entropy*. ISTE Press, UK, 352 pp, ISBN 9781847040268
- Kanevski M, Canu S (2000) *Spatial data mapping with support vector regression*. IDIAP Research Report RR-00-09
- Kanevski M, Maignan M (2004) *Analysis and modelling of spatial environmental data*. EPFL Press, Lausanne
- Kanevski M, Arutyunyan R, Bolshov L, Demyanov V, Maignan M (1996) Artificial neural networks and spatial estimations of Chernobyl fallout. *Geoinformatics* 7(1):5–11
- Kanevski M, Pozdnoukhov A, Canu S, Maignan M (2002) Advanced spatial data analysis and modelling with support vector machines. *Int J Fuzzy Syst* 4(1):606–616
- Meyer D, Leisch F, Hornik K (2003) The support vector machine under test. *Neurocomputing* 55:169–186. doi:10.1016/S0925-2312(03)00431-4
- Parzen E (1962) On estimation of a probability density function and mode. *Ann Math Stat* 33:1065–1076. doi:10.1214/aoms/1177704472
- Platt J (1999) Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. In: Smola AJ, Bartlett P, Scholkopf B, Schuurmans D (eds) *Advances in large margin classifiers*. MIT Press, Cambridge, MA
- Pozdnoukhov A (2005) Support vector regression for automated robust spatial mapping of natural radioactivity. *Appl GIS* 1(2):21-01–21-10
- Pozdnoukhov A, Purves RS, Kanevski M (2008) Applying machine learning methods to avalanche forecasting. *Ann Glaciol* 49:107–113
- Rasmussen C, Williams C (2006) *Gaussian processes for machine learning*. MIT Press, Cambridge, 266 pp
- Scholkopf B, Smola A (2002) *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT Press, Cambridge
- See L, Solomatine DP, Abrahart R, Toth E (2007) Hydroinformatics: computational intelligence and technological developments in water science applications. *J Hydrol Sci* 52(3):391–396 (Special Issue: Hydroinformatics)
- Smola A, Scholkopf B (2004) A tutorial on support vector regression. *Stat Comput* 14:199–222. doi:10.1023/B:STCO.0000035301.49549.88
- Specht DF (1991) A generalized regression neural network. *IEEE Trans Neural Netw* 2:568–576. doi:10.1109/72.97934
- Timonin V, Savelieva E (2005) Spatial prediction of radioactivity using general regression neural network. *Appl GIS* 1(2):19-01–19-14
- Tripathi S, Srinivas VV, Nanjundiah RS (2006) Downscaling of precipitation for climate change scenarios: a support vector machine approach. *J Hydrol* 330(3–4):621–640
- Vapnik V (1995) *The nature of statistical learning theory*. Springer-Verlag, New York
- Vapnik V (1998) *Statistical learning theory*. Wiley, New York
- Whiteman CD (2000) *Mountain meteorology: fundamentals and applications*. Oxford University Press, New York, 376 pp