Abstract Volume
11th Swiss Geoscience Meeting

Lausanne, 15th – 16th November 2013

17. Computational GIScience

sc | nat

Swiss Academy of Sciences
Akademie der Naturwissenschaften
Accademia di scienze naturali
Académie des sciences naturelles

Unil

UNIL | Université de Lausanne
Faculté des géosciences
et de l'environnement

17. Computational GIScience

M. Kanevski, Ch. Kaiser, F. Bavaud

TALKS:

17.1 Bavaud, F.: Spatial autocorrelation and clustering: a minimal yet sufficient formalism

17.2 Emmanouilidis T.: GIS and urban planning: a new clustering algorithm for schools and students location-allocation

17.3 Grangier L., Kaiser C.: Touristic guides on smartphone – The Kalman filter for a smarter GPS localization in mountain area.

17.4 Guex G.: Mixing random walks and shortest paths to create new graph dissimilarities

17.5 Joost S., Stucki S., Leempoel K.: Geocomputational approaches for the analysis of Next-Generation Sequencing (NGS) and multi-scale data in landscape genomics

17.6 Josset L., Demyanov V., Ginsbourger D., Lunati I.: Uncertainty quantification in porous media using stochastic sampling algorithm and Functional Data Analysis

17.7 Kaiser C.: Searching geolocated databases with vague spatial objects

17.8 Kanevski M.: Geospatial data analysis and modelling using geostatistics and machine learning

17.9 Lochbühler T., Vrugt J.A., Linde N.: Realistic model constraints in probabilistic inversion of geophysical data: Summary statistics from training images

17.10 Price B., Kienast F., Seidl I., Verburg P., Ginzler C., Bolliger J.: Spatially explicit modelling of land-use suitability and future land-use pattern for Switzerland

POSTERS:

P 17.1 Calianno M., Gourley J. J., Ruin I.: Forecasting flash flood impacts utilizing anthropogenic exposure factors

P 17.2 Demyanov V., Kanevski M.: Multi-scale spatial modelling of radioactive pollution with a kernel learning algorithm

P 17.3 Golay J., Kanevski M., Vega Orozco C.: The multipoint Morisita index for the analysis of geodemographic data

P 17.4 Leuenberger M., Kanevski M.: Extreme learning of environmental pollution

P 17.5 Leuenberger M., Vega Orozco C., Tonini M., Kanevski M.: Random Forest for susceptibility mapping of natural hazards

P 17.6 Ruggeri P., Irving J., Holliger K.: Evaluation of geostatistical resampling as a proposal mechanism in Bayesian MCMC solutions to hydrogeophysical problems

P 17.7 Tonini M., Vega Orozco C., Kanevski M.: Wildfires spatio-temporal aggregation: from global cluster to local mapping

P 17.8 Vega Orozco C., Tonini M., Kanevski M.: Analysis of the dynamic of urban areas and of their interaction with forest fires

17.1

Spatial autocorrelation and clustering: a minimal yet sufficient formalism

Bavaud François

Institute of Geography and Sustainability, Faculty of Geosciences and Environment, University of Lausanne, CH-1015 Lausanne

One considers n topologically related, irregular regions, with differing multivariate profiles. Measuring and modeling spatial autocorrelation, as well as clustering the regions into m groups requires a minimal yet sufficient set of three non-negative matrices:

- a (n x n) symmetric *exchange matrix* E, whose components, summing to unity, interpret as a the probability of selecting a pair of regions, and whose margin defines the regional weights f
- a (n x n) *dissimilarity* D between regions, chosen as squared Euclidean
- a (n x m) *soft membership matrix* Z giving the probability that region i=1,…,n belongs to group g=1,…,m .

For instance:

- *weighted multidimensional scaling* of (D,f) amounts in the *principal component analyses* routinely used in quantitative Geography
- pointwise Schoenberg transformations of D into new squared Euclidean distances produce the high-dimensional embedding of features encountered in Machine Learning
- the index 1-trace(ED)/(f'Df) provides a multivariate generalization of *Moran's I*, measuring spatial autocorrelation
- *spectral clustering* consists in partitioning the weighted undirected network described by E by considering the eigenvectors of (a standardized version of E) as the regional features.
- minimizing over Z the free energy functional V[D,f,Z]+T H[Z] amounts in the *soft K-means clustering* procedure, where V is the within-groups dispersion, H the regions-groups mutual information and T>0 the temperature, controlling the softness of the resulting clustering.

We present a new algorithm, **landscape clustering**, favoring the grouping of regions with small dissimilarities *and* strong exchanges. It works for any positive semi-definite exchange matrix E, such as resulting from a continuous-time jump process whose infinitesimal generator is given by a binary adjacency matrix, with fixed weights. It contains two free parameters, namely the *temperature* T, as well as the *contiguity contribution* p in [0,1]: p=1 amounts to spectral clustering, p=0 to K-means, and 0<p<1 to a clustering scheme taking into account *both proximity and similarity between regions*. Illustrations are provided by the clustering of French departments, taking into account their spatial contiguity as well as their political configurations in the first round of the French presidential 2012 election.

17.2

GIS and urban planning: a new clustering algorithm for schools and students location-allocation

Emmanouilidis Théophile

Institute of Geography and Sustainability, Faculty of Geosciences and Environment, University of Lausanne, CH-1015 Lausanne

The use of GIS (for data visualisation or simulation) in the public administration has greatly increased in the past years. Since 2009, the school department of Lausanne has been able to answer political and public demands by leading two main GIS-projects. The first one was to give public transport subvention to schoolchildren by considering their age and shortest path distance from home to school. The second one, which will be exposed here, is about primary schools and students (from 4 to 10 years old) location-allocation. Due to major changes in the Swiss educational laws (HarmoS), the department was asked to reduce the population of each school establishment (group of schools) by creating new ones of similar size and taking into account the future growth of the city.

Conceptually the idea is to assign node (student home location) $i=1,\ldots,n$ to the nearest school $g=1,\ldots,m$ considering its location $L^*(g)$ and capacity $\rho^*(g)$. This classical location-allocation problem is solved by a new soft constrained clustering algorithm defined as:

$$z_{ig} = \frac{\rho_g^* \epsilon_g \exp(-\beta F(d_{iL^*(g)}))}{\sum_h \rho_h^* \epsilon_h \exp(-\beta F(d_{iL^*(h)}))}$$

- z_{ig} : probability to attribute node i to school g
- $F(d_{iL^*(g)})$: is an increasing non negative function of distance between i and j. In the present case, it corresponds to the shortest path distance computed through the pedestrian network of the city.
- β: parameter controlling the sharpness of the groups.
- €: is a parameter controlling the capacity constraints, iteratively determined as :

$$\frac{1}{\epsilon_g} = \sum_i \frac{f_i \exp(-\beta F(d_{iL^*(g)}))}{\sum_h \rho_h^* \epsilon_h \exp(-\beta F(d_{iL^*(h)}))}$$

- f_i : proportion of student living in location i (relative weight)

REFERENCES

A. Antunes, D. Peeters (2000) A dynamic optimization model for school network planning, Socio-Economic Planning Sciences 34, pp. 101-120

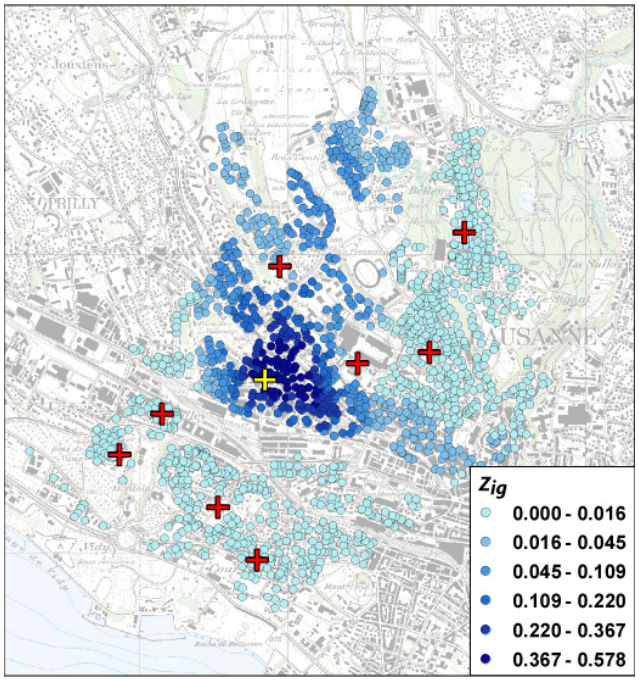
F. Bavaud (2010) Aggregation invariance in general clustering approaches, Advances in Data Analysis and Classification 3, pp. 205-225

T. Emmanouilidis and F. Bavaud (2012) Analysis of the pedestrian network of Lausanne : eccentricity, accessibility and partitionning, to be published

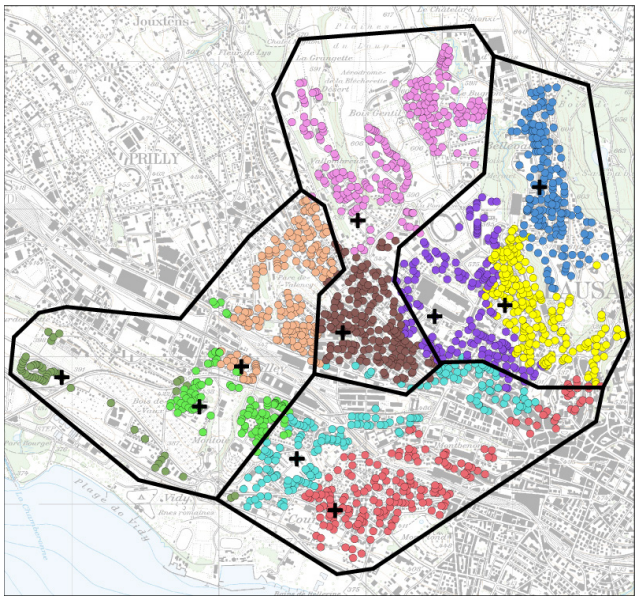
Nock, R. and Nielsen, F. (2006) On weighting clustering in Pattern Analysis and Machine Intelligence Vol.28, n°8

Wilson A G, (1971) A family of spatial interaction models, and associated developments Environment and Planning 3(1), pp. 1 – 32

RESULTS:



Constrained fuzzy-clustering in the West of Lausanne (β=0.005). Each dot represent a schoolchildren(s) building and the colour its probability to be assigned to the school of "Prélaz" (yellow cross) considering the 8 others schools.



"Hardened" soft constrained clustering of the West part of Lausanne: map depicts for *each node the regions of highest membership. Contiguous regions can then be merged to create new school establishments.*

17.3

Touristic guides on smartphone – The Kalman filter for a smarter GPS localization in mountain area.

Grangier Lucien[1], Kaiser Christian[1]

[1]*Institute of geography and sustainability, Mouline – Geopolis, University of Lausanne, 1015 Lausanne (Lucien.grangier@unil.ch)*

With the increasing ubiquity of smartphones and tablet computers, the number of available applications for mobile platforms is growing fast while their quality is highly variable. Nearly every brand, store, and municipality wants a smartphone application to be in the main stream. Many visitor/touristic guides are among this abundance of applications. Generally they permit to guide the user in a city or a museum, to share knowledge and information, to interact with other users and also increasingly with the environment through Augmented Reality (AR).

This kind of applications uses multiple device features such as the camera, Wireless, 3G connection or GPS, which are all voracious in energy. Usually the user is indoors or in an urban environment where it's easy to connect and charge the device. But what happens if we want to use an application in a mountain area? Often, no Internet connection is available, and generally no electricity to charge the battery. Applications designed for use in such an environment need to deal with a lot of constraints, and special care needs to be given to limited battery resources.

The Geographic Information Science Group of the Institute of Geography and Sustainability has developed in collaboration with the bureau "Relief" the application "Geoguide", which is a guide for a didactic walk in Lausanne (available on Android/iOS). Another application has been developed for the Vallon de Nant, which is in a mountain area. The concept of these applications is to offer a didactic trail through the region of interest, and at some precise locations, detailed information around a given theme is given using text, images and potentially videos. These applications provide the opportunity to develop and test new techniques in a real-world scenario.

In order to increase battery lifetime we want to optimise the localisation process. To display the current location on the screen, the device calculates its position every 10 seconds using GPS and cell tower locations. In this paper, we explore a technique to reduce usage of GPS and other location techniques while still providing an optimal user experience. The key idea is to vary the time between two requests to the device's location service. Around points of interest where the applications offers detailed information about the location, the frequency of locational requests is kept at 10 seconds, while in locations between points of interest, the frequency can be as low as one locational request each minute, or even less. In our approach, we use a Kalman filter to predict the user's future location, which is likely to be on the suggested trail. The application calculates the mean walking speed, and makes locational requests only occasionally to confirm and update the location prediction or before arriving at a point of interest.

The approach of using a Kalman filter for locational prediction can also be used for notifying the user when arriving at a point of interest. This is especially useful in a mountain environment where points of interest might be sparse and in some cases considerable physical effort is required between two points of interest. Especially in such an environment, the user of the smartphone application does not want to let the device switched on to permanently checking if detailed information about his current location is available. As a solution, the device calculates the approximate time of arrival at the next point of interest to confirm and update its position more frequently, and finally notifies the user through a sound or vibration notification that information is available. This feature allows for putting the device into standby mode for most of the time, further reducing the energy consumption and increasing battery lifetime.

Link
http://igd.unil.ch/geoguide/

17.4

Mixing random walks and shortest paths to create new graph dissimilarities

Guex Guillaume

Institute of Geography and Sustainability, University of Lausanne, CH-1015 Lausanne

The analysis of geographical networks often requires to compute a dissimilarity between nodes and to use this dissimilarity to perform classifications and visualisation algorithm or to compute derived indices. The *shortest-path distance* is historically the most used dissimilarity, but it has been shown recently that the *commute-time distance* offer an interesting alternative (see e.g. Liu et al. 2013), as it is squared Euclidean for all graphs and gives very different and interesting results.

This contribution presents the properties of known graph dissimilarities in the first part and shows how to build more general *flow-based dissimilarities* in the second part. A flow $X=(x_{ij})$, modelling agents moving on the edges of a graph, is computed to minimize a *free energy functional* $F(X)=U(X)+TG(X)$, where the *energy* $U(X)$ encourages agents to follow routes of least cost, the *entropy* $G(X)$ adds a random component to movements and the *temperature* $T>0$ arbitrates between the conflicting objectives of minimizing the costs and maximizing randomness (Saerens et al. 2009, Bavaud & Guex 2012). The properties of these new dissimilarities will be studied on different networks and various examples of applications on geographical and spatial networks will be exhibited.

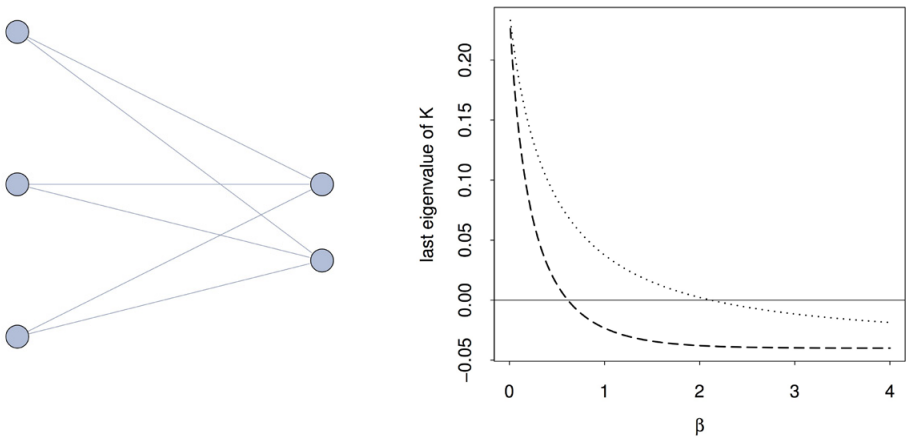


Figure 1. Evolution of the last eigenvalue of K, the matrix of "scalar product" derived from two different flow-based dissimilarities built on the bipartite network shown left, regarding $b=1/T$, the inverse temperature. When b is small, the dissimilarities tend to commute-time distances and are squared Euclidean as K is positive semidefinite. As b increases, they tend to shortest-path distances and negative eigenvalues occur, ruining the squared Euclidean nature of the dissimilarities.

REFERENCES

Bavaud, F. & Guex, G. 2012: Interpolating between random walks and shortest paths: a path functional approach. K. A. et al. (Ed.), SocInfo 2012, Vol. 7710 of Lecture Notes in Computer Science, pp. 68-81. Springer.

Liu, S., Matzavinos, A., Sethuraman, S. 2013: Random walk distances in data clustering and applications. Adv. Data Analysis and Class.7, pp. 83-108.

Saerens, M., Achabany, Y., Fouss, F., Yen, L. 2009: Randomized shortest-path problems: two related models. Neur. Comp. 21, pp. 2363-2404.

17.5

Geocomputational approaches for the analysis of Next-Generation Sequencing (NGS) and multi-scale data in landscape genomics

Joost Stéphane[1], Stucki Sylvie[1], Leempoel Kevin[1]

[1]*Laboratory of Geographic Information Systems (LASIG), School of Architecture, Civil and Environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), Building GC, Station 18, CH-1015 Lausanne (stephane.joost@epfl.ch)*

The application of geocomputation to the field of landscape genomics (Manel et al. 2010) permits to carry out demanding computational tasks that recently emerged because of the advent of large Next-Generation Sequencing data. When investigating the genetic mechanisms of evolution in spatially distributed plants or animals, geocomputation also proves to be useful to process many association models (gene x environment) in a multi-scale context.

One challenge for correlative approaches when detecting genomic regions possibly under natural selection is to distinguish the effect of environmental conditions from effects caused by the demographic history of investigated populations. Several methods have been developed in population genetics, the most promising being the analysis of patterns of linkage disequilibrium (LD) (Jensen et al. 2007). LD is the occurrence of combinations of alleles (variant forms of the same gene) in a population more often or less often than would be expected at random. Correlative approaches also offer a solution by means of the integration of measures of Local Indicators of Spatial Association (LISA). Indeed, Stucki et al. (2012) made it possible to process high-throughput geo-referenced molecular datasets and underlying environmental variables, and to simultaneously provide a list of genomic regions likely to be under selection with a measure of local spatial autocorrelation. The latter constitutes a useful indication as regards the possible kinship relationship between individuals. In the context of a research on local adaptation in 102 Ugandan cattle individuals, more than 2 million of binary markers have been compared to 73 WorldClim and SRTM-derived environmental variables (Stucki et al. 2013). The most interesting model included an interesting genetic marker that maps to a gene (CHST11) involved in cartilage make up (Figure 1).

Another challenge for landscape genomics is to address a fundamental issue often referred as to "at which spatial scale does adaptation act?" Most of the time, research is carried out at one single scale. However, Leempoel et al. (2013) developed a geocomputational framework based on a signal processing generalization technique to produce DEMs at multiple scales allowing for a continuous representation of the landscape. In order to better understand the way an Alpine plant named *Biscutella laevigata* L. adapts to the environment, the approach was used to investigate its adaptive response to environmental variables derived from these DEMs. For this purpose 266 genetic markers from 361 individuals sampled on the ridge of "Les Rochers-de-Naye" in the Swiss Prealps were used. Preliminary results show that resolution matters at a local scale (Figure 2), but that the use of very high resolution variables does not necessarily improve the significance of the results obtained. However, current investigations will show if their use combined with a multi-scale analysis system permits to detect relationships that would have gone unnoticed otherwise.

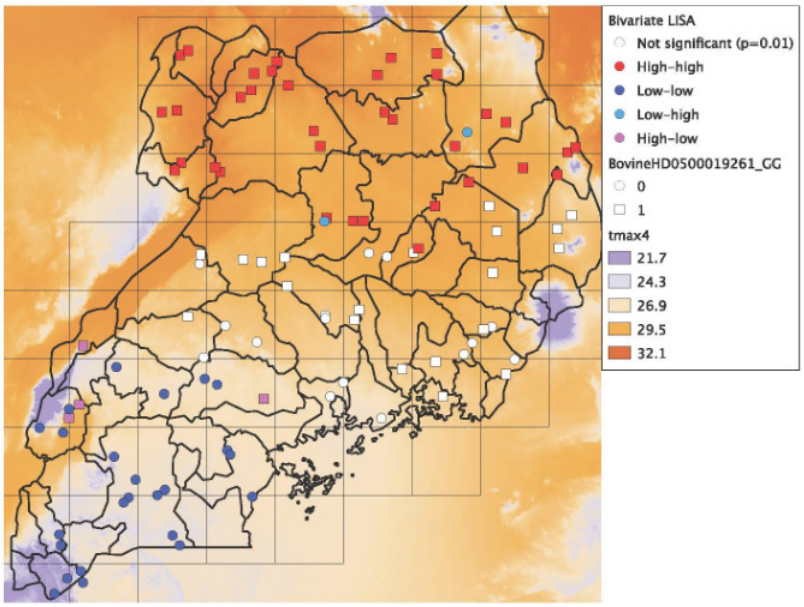


Figure 1.Bivariate LISA between marker BovineHD0500019261_GG and mean temperature in April with corresponding clusters (weighting scheme is K = 20 nearest neighbors). Dots indicate the presence (square) or absence (circle) of markers and their color shows the type of association.

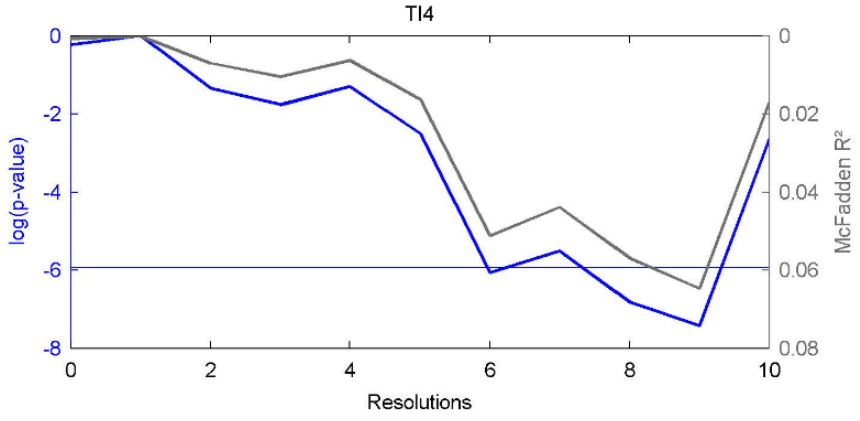


Figure 2. Variation of the log of the p-value in function of the spatial resolution (m) for the model involving genetic marker C1v242 and solar radiation.

REFERENCES

Jensen, J.D., Thornton, et al. 2007: On the utility of linkage disequilibrium as a statistic for identifying targets of positive selection in nonequilibrium populations. Genetics, 176, 2371–2379.

Leempoel, K., Geiser, et al. 2013: A multiscale landscape genomics analytical framework to identify local adaptive variation in the Buckler Mustard, IALE 2013 European Congress, Manchester, September 2013.

Manel, S., Joost, S., et al. 2010: Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. Mol. Ecol., 19, 3760-3772.

Stucki, S., Orozco-terWengel, et al. 2013: Sambada in Uganda: landscape genomics study of traditional cattle breeds with a large SNP dataset, IALE 2013 European Congress, Manchester 9-12 September 2013.

Stucki, S. Agha, S., et al. 2012: Utilization of the Scythe C++ open source library for statistical geocomputation in livestock landscape genomics. Proc. OGRS Symposium, Yverdon-les-Bains, Switzerland, October 2012.

17.6

Uncertainty quantification in porous media using stochastic sampling algorithm and Functional Data Analysis

Josset Laureline[1], Demyanov Vasily[2], Ginsbourger David[3] & Lunati Ivan[1]

[1]*Centre de Recherche en Environnement Terrestre, University of Lausanne, Quartier UNIL-Mouline, Bâtiment Géopolis, CH-1015 Lausanne (laureline.josset@unil.ch)*
[2] *Institute of Petroleum Engineering, Heriot-Watt University, Energy Academy Building, Edinburgh EH14 4AS, UK*
[3]*Institute of Mathematical Statistics and Actuarial Science, University of Bern, Alpeneggstrasse 22, CH-3012 Bern*

Predicting flow in aquifers and reservoirs is difficult due to the lack of information on flow and transport properties of the formation. Generation of multiple model realisations aims at quantifying the uncertainty of flow predictions. A pragmatic approach to uncertainty quantification in flow modelling is to determine the region of the model-parameter space where model results match observed dynamic flow responses (e.g. flow rates, contaminant concentration). However, due to the computational limitation, the systematic evaluation of the flow response for each model realization generated in a Monte Carlo implies prohibitive computational costs.

We propose two strategies to reduce the computational costs: on one hand, we employ an adaptive stochastic sampling algorithm (particle swarm optimization) to accelerate the exploration of the parameters space; on the other hand, to evaluate the flow response of the generated models, we use an approximate model coupled with an error model to predict the corresponding exact response. This error model is constructed using Functional Data Analysis (FDA) on a training set of model solutions for which both exact and approximate responses are known.

The proposed approach is tested on a synthetic test case: a layered aquifer divided by a fault, for which the permeabilities of the layers and the fault throw have to be determined. We evaluate the performance of this approach and present a comparison with concurrent methods that have been previously applied to this test case.

17.7

Searching geolocated databases with vague spatial objects

Kaiser Christian

Institute of geography and sustainability, University of Lausanne, CH-1015 Lausanne

Geographic databases are increasingly available on the Web. For example, Viatimages (www.unil.ch/viatimages) is an online database with geolocated historic images mainly with Alpine landscapes. Many other similar examples exist on the Web. Such a database frequently contains geographic features with vague boundaries or unknown extent. During the geolocalisation process, the exact location of an item might not be known, or the chosen location is wrong. Or the user of the database is interested in all items that are located «around the Lake of Geneva» or «along the Rhône river». The human representation of these locations is usually relatively vague. At the same time, the digital representation of the same concepts is frequently built around vector geometries, which have very precise boundaries. This paper shows a possible solution to this discrepancy, by introducing vague vector geometries and simple spatial query operators.

Geolocated items in a database, such as images with known location, are usually stored as simple points. Regions are mostly represented as polygons. For such vector geometries, efficient spatial operators such as intersection, membership (A within B) or union exist. Due to the precise boundary of vector geometries, they are unsuitable for spatial objects with vague extent or limits. Consequently, vague spatial objects are generally modelled using a raster representation with values varying between 0 (outside of the object) and 1 (inside the object). For big objects with high raster resolution, this approach leads to big datasets while still being limited in the spatial resolution. As an alternative, we use vector geometries enhanced with information on the vagueness of the boundary. Different types of transition from «inside the spatial object» to «outside the spatial object» can be defined. Such a transition is any function going from 1 (inside the object) to 0 (outside

the object). Figure 1 gives some possible transition functions. For each point on the boundary of a vector geometry, a transition function can be defined. A transition function usually needs at least one parameter defining the size of the vague boundary. In the case of a Gaussian function, this parameter is usually referred to as «bandwidth». Figure 2 shows a simple representation of an example polygon with a vague boundary. The small lines orthogonal to the polygon boundary are the definition of the transition function, where the length of the line represents the bandwidth of the Gaussian function. For points between two transition functions, a weighted average based on the proximity to the transition function is calculated. Such a representation also allows for simple editing of vague vector geometries. It is important to note that these vague vector geometries are a simple extension of traditional vector geometries with hard boundaries.

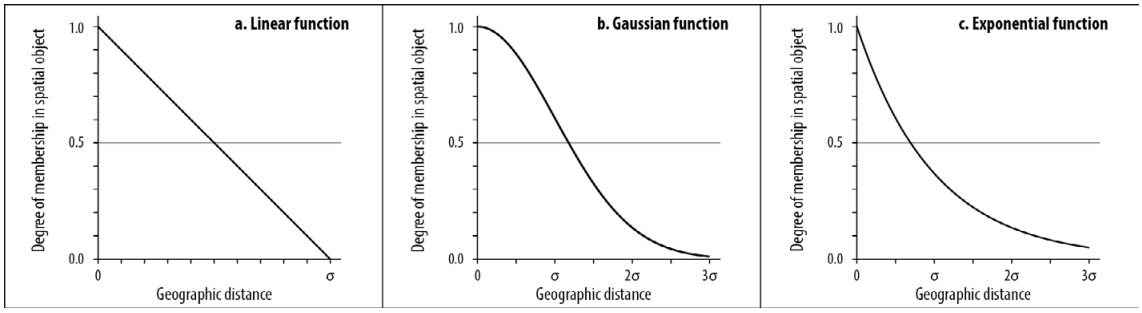


Figure 1. Different possible functions for defining vague boundaries on vector geometries.

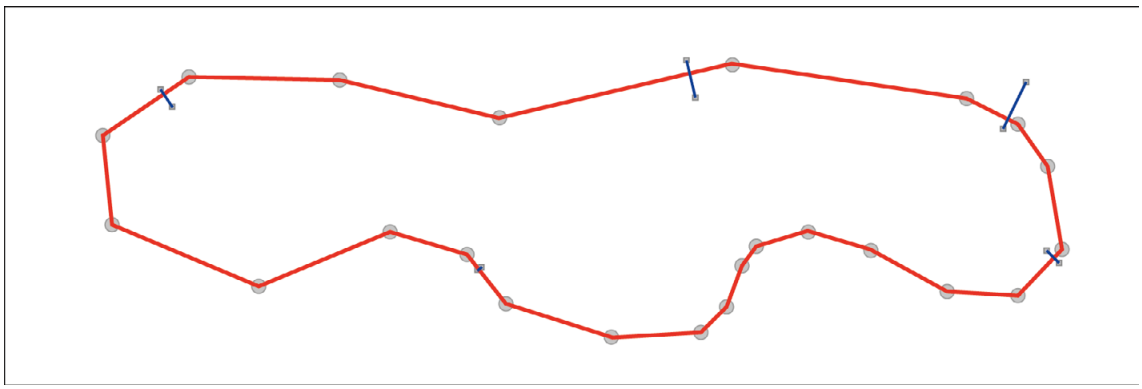


Figure 2. Definition of a polygon with vague boundary. The blue lines with small rectangle handles for manipulation allow for definition of the extent of the vague boundary.

Once the spatial objects with vague boundaries are defined, they can be used for spatial search using membership and intersection operators known from traditional GIS. Due to the imprecise boundary, these operators need to be modified. For example, a query of type «Does polygon A contain point B?» does not necessarily have an answer «true» (=1) or «false» (=0), but 0.9 indicating that there is a high chance that point B is inside polygon A. The search result can be sorted according to these membership operator values.

Membership and intersection operators are only in some special cases easy to define for geometries with vague boundaries. In most cases, a discrete approximation needs to be calculated, which can be done by creating a local raster image at locations with vague boundaries. The operators are well defined for the case of discrete raster images.

The presented approach to define vague spatial objects is a relatively simple extension of existing vector geometries, allowing at the same time for quite easy editing of the geometries. It gives the flexibility to deal with a wider range of spatial objects in a straightforward manner, and enabling spatial databases with more flexible search functionality.

17.8

Geospatial data analysis and modelling using geostatistics and machine learning

Kanevski Mikhail

Centre for Research on Terrestrial Environment, University of Lausanne, CH-1015 Lausanne (Mikhail.Kanevski@unil.ch)

The presentation gives an overview of the research on geospatial data analysis, modeling and visualization carried out at the Faculty of Geosciences and Environment, University of Lausanne. The research covers both elaboration and adaptation of new methods and models, as well as algorithms and software tools developments. The main applications deal with geospatial environmental, natural hazards, socio-economic and demographic data. Some of the achievements were published in books and book chapters, see the references below.

The main approaches widely used for geospatial data include: geostatistics (variography, predictions and simulations, including risk mapping); machine learning (supervised, unsupervised, semi-supervised) algorithms (artificial neural networks of different architectures – multilayer perceptron, general regression neural networks, probabilistic neural networks, radial basis function networks, Gaussian mixtures, mixture density networks, random forest; and statistical learning theory - kernel-based methods, e.g. support vector machines, etc.), and visualization of multivariate and high dimensional data using, between many others, parallel coordinates, Andrews plots, Kohonen self-organizing maps.

It should be noted that, the most relevant and efficient methods and tools are selected depending on the quantity and quality of data, dimensionality of the problem and objectives of the study. In case of application of machine learning, an important problem is testing and justification of the quality of the results by characterizing the uncertainties and confidence intervals. An important part of the methodology is a comprehensive exploratory analysis of raw data and the modelling residuals.

In fact, many environmental, natural hazards and socio-economic phenomena should be considered and modeled in a high dimensional feature space (d ~10-100). Therefore the problem of relevant features/characteristics extraction/selection becomes an important one, both for the understanding of the phenomena and the explanation of the results. For this reason, machine learning algorithms which have inherent capabilities to rank or select features (adaptive general regression neural networks, random forest, multiple kernel learning, and others) are very useful.

Some of the recent and representative results on the application of machine learning algorithms for geo- and environmental data (pollution and natural hazards) are presented and discussed within the framework of the developed self-consistent data modelling methodology. An important part of the methodology concerns the analysis and optimization of monitoring networks. The problem is quite difficult, especially when the phenomena are considered in a high dimensional space. Within the framework of machine learning an active learning approach was successfully applied for monitoring networks design/redesign both for low and high dimensional environmental data.

The most important future developments deal with better characterization of uncertainties, multiscale analysis of geospatial data, automatic feature selection, and better understanding and modelling of space-time phenomena.

The author would like to thank to many colleagues and coauthors for fruitful and interesting collaboration on the topics considered.

REFERENCES

Kanevski, M. & Maignan M. 2004: Analysis and Modelling of Spatial Environmental Data, EPFL Press, Lausanne.

Kanevski M. (Editor). 2008: Advanced Mapping of Environmental Data, iSTE and Wiley.

Kanevski M., Foresti L., Kaiser Ch., Pozdnoukhov A., Timonin V. & Tuia D. 2009: Machine Learning Models for Geospatial Data. In: F. Bavaud and C. Mager (Eds.), Handbook of Quantitative and Theoretical Geography, University of Lausanne, pp. 175-227

Kanevski M., Pozdnoukhov A. & Timonin V. 2009: Machine Learning for Spatial Environmental Data. Theory, Applications and Software. EPFL Press, Lausanne.

Tuia D., Pozdnoukhov A., Foresti L. & Kanevski M. 2013: Active learning for monitoring network optimization. In: Mateu J. & W. Muller, Spatio-temporal design. Wiley.

17.9

Realistic model constraints in probabilistic inversion of geophysical data: Summary statistics from training images

Lochbühler Tobias[1], Vrugt Jasper A.[2], Linde Niklas[1]

[1]*Applied and Environmental Geophysics Group, Faculty of Geosciences and Environment, University of Lausanne, Lausanne, Switzerland (tobias.lochbuehler@unil.ch)*
[2]*Department of Civil and Environmental Engineering, University of California Irvine, Irvine, CA, USA*

Multiple-point statistics training images are conceptual geological models that feature the dominant lithological units and structural patterns of a site of interest. They provide strong *a priori* information on model morphology that is used to constrain models in probabilistic inversion. This is done by extracting summarizing statistical metrics from realizations of a training image and including the probability of each proposal state given the difference between the observed and simulated summary metrics in the calculation of the posterior probability of each proposal model. Realistic constraints are thus imposed on the inverse models avoiding expensive model updating of common geostatistical approaches. The inverse problem is solved in a Bayesian formulation by Markov chain Monte Carlo sampling with the MT-DREAM$_{(ZS)}$ algorithm. Additional to the summary metrics, the training image realizations are used to define case-dependent optimal parameterizations by compressed sensing analysis in the discrete cosine domain, which drastically reduces the dimensionality of the inverse problem but maintains the ability to recover realistic subsurface structures. The methodology is applied to crosshole ground-penetrating radar data for two synthetic case studies. The benefits of two different summary metrics are evaluated: the frequency of occurrence of different geological facies and the total sum of discrete cosine transform coefficients as a global measure of model variability. It is shown that using the model constraints helps to (1) steer the inversion towards geologically more realistic *a posteriori* models, (2) prevent inversion artifacts, and (3) decrease the deviation between the most probable model and the true subsurface, thus mitigating a common pitfall of high-dimensional inverse problems. The proposed approach is general and allows great flexibility in terms of the applied model constraints and the type of geophysical forward problem.

17.10

Spatially explicit modelling of land-use suitability and future land-use pattern for Switzerland

Price Bronwyn[1], Kienast Felix[1], Seidl Irmi[1], Verburg Peter[2], Ginzler Christian[1] & Bolliger Janine [2]

[1] *Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Zürcherstrasse 111, CH-8903 Birmensdorf (bronwyn.price@wsl.ch)*
[2] *Institute for Environmental Studies (IVM), VU University Amsterdam, De Boelelaan 1087, 1081 HV Amsterdam, The Netherlands*

Using three time steps of the high-resolution Swiss Land-Use/Land-Cover Statistics (1985, 1997 and 2009), we map and measure past land-use and land-use type transitions. We model the suitability for a given land-use and the probability of land-use transitions for the whole of Switzerland with respect to a suite of biophysical explanatory variables, comparing the predictive ability of various (geo)statistical and data-mining techniques such as regression kriging, random forests and logistic regression. The models of land-use suitability are a key input to land-use change models for Switzerland.

Socio-economic processes are strong drivers of land-use change across European landscapes. Land abandonment, relating to a decline of agricultural significance, has been a dominant process affecting European landscapes, and in particular mountainous regions such as in Switzerland since the mid 20th century. Urbanisation in Switzerland is increasing at a rapid rate as population increases and, in particular, with increasing demand in living-space per capita. Decrease in public support for nuclear power stations is driving a push towards increased production of renewable energy within Switzerland. These inter-related and sometimes competing processes have significant implications for land-use and patterns of land-use change within Switzerland, yet the extent and location of anticipated land-use changes remain unknown, as does the impact on landscape services.

This project defines 5 scenarios of future land-use demands for Switzerland under different projections for urban-sprawl, land abandonment and land use for renewable energy production.

Using the Dyna-CLUE land-use change modelling framework (Verburg and Overmars, 2009) we applied the 5 future scenarios to determine and visualize (map) future land-use patterns in a spatially explicit manner. Under a 'business-as-usual' scenario we evaluate the ability of the model to produce the 'current' land-use pattern (Swiss land-use statistics 2009), from the 1997 initial state.

The resulting spatially explicit land-use scenarios will be freely available for download by researchers and policy makers. These scenarios will provide key base information for future work including assessing conflicts and synergies in land-use planning or assessing impacts of land-use change on landscape services.

REFERENCES

Verburg, P.H. & Overmars, K.P. 2009: Combining top-down and bottom-up dynamics in land use modeling: exploring the future of abandoned farmlands in Europe with the Dyna-CLUE model. Landscape Ecology, 24: 1167-1181. http://dx.doi.org/10.1007/s10980-009-9355-7

P 17.1

Forecasting flash flood impacts utilizing anthropogenic exposure factors

Calianno Martin[1], Gourley Jonathan J.[2] & Ruin Isabelle[3]

[1]*Institut de Géographie et Durabilité (IGD), University of Lausanne, Géopolis, CH-1015 Lausanne (martin.calianno@unil.ch)*
[2]*NOAA, National Severe Storm Laboratory (NSSL), Norman, USA*
[3]*Laboratoire d'étude des Transferts en Hydrologie et Environnement (LTHE), Université Joseph Fourier, Grenoble, France*

If you've ever attempted to hand contour of map of damage from flash flooding, you will quickly discover the spatial character is strongly influenced by surface controls, which creates discontinuous and untidy contours. Low-water crossings, infrastructure built in close proximity to streams, normally dry riverbeds in populated zones, and roads in low-lying areas are all examples of anthropogenic factors that increase societal susceptibility to flash flooding.

Figure 1. Flash flood in Fort Worth, TX, June 27, 2007. © AP photograph/Fort Worth Star-Telegram, Ron Jenkin

Unfortunately, tools commonly used to monitor and predict flash flooding, including sophisticated distributed hydrologic models, rarely incorporate any information about these factors. In this study, we rely on a detailed impact classification and analysis of flash flood reports from the National Weather Service (NWS) Storm Data and Severe Hazards Analysis and Verification Experiment (SHAVE) conducted at the National Severe Storms Laboratory (NSSL) in Norman, Oklahoma (Gourley et al. 2010). The flash flood impact database is then used to derive static exposure indices using GIS layers of road networks, population density, stream locations, elevation, degree of imperviousness (urbanization), land use, etc (Calianno et al. 2013). The exposure indices are then used in conjunction with distributed hydrologic model simulations with forcing from radar quantitative precipitation estimations (QPE) to improve the specificity and accuracy of forecasts, with a focus on the particular impacts. Two extreme flash flooding events in Oklahoma are studied to demonstrate the new approach, highlighting improvements and shortcomings.

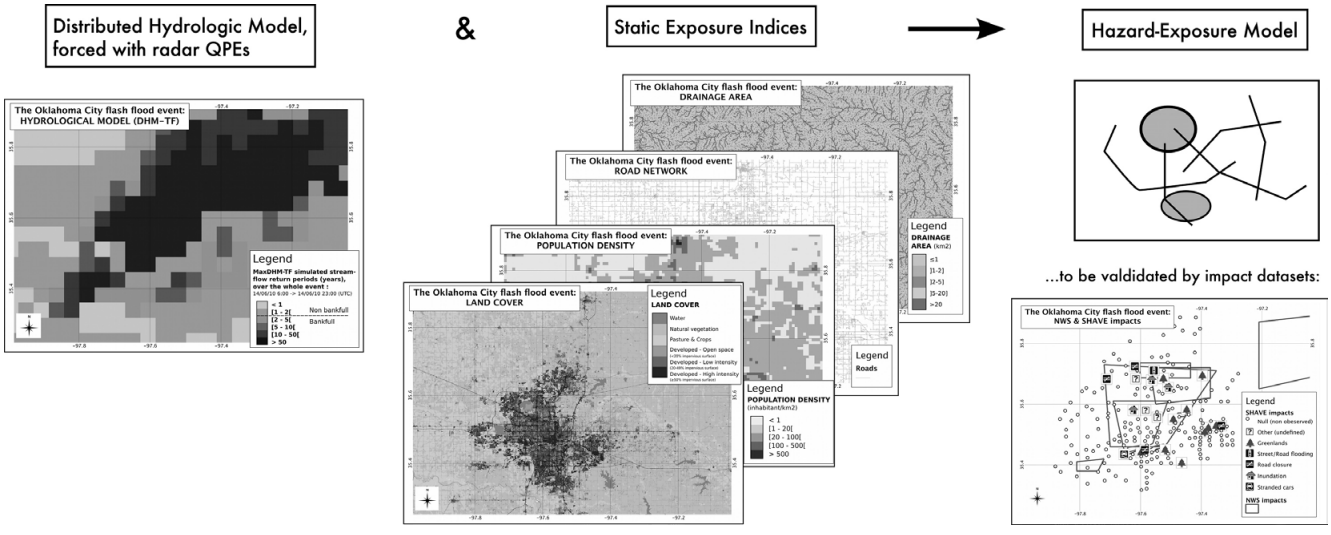


Figure 2. Conceptual scheme of the Hazard-Exposure model for flash flood impact forecasting.

REFERENCES
Calianno M., Ruin I. & Gourley J.J. 2013: Supplementing flash flood reports with impact classifications. J. Hydrol. 477, 1-16.
Gourley, J.J., Erlingis, J.M., Smith, T.M., Ortega, K.L. & Hong, Y. 2010: Remote collection and analysis of witness reports on flash floods. J. Hydrol. 394, 53-62.

P 17.2

Multi-scale spatial modelling of radioactive pollution with a kernel learning algorithm

V. Demyanov[1], M. Kanesvki[2]

*Heriot-Watt University, Erinbugh, UK, vasily.demyanov@pet.hw.ac.uk
**University of Lausanne, Suisse

Modelling multi-scale spatial patterns originated from complex combination of sources and impact factors is a challenging problem. Many conventional spatial modelling algorithms (geostatistics) are limited by stationarity assumptions and are not flexible enough to cope with highly variable patterns correlated at multiple scales. Machine learning approach provides an effective alternative in modelling spatial patterns of different complexity at a range of spatial scales without the limitation of the stationarity assumption.

Multi-scale spatial structure can be modelled using kernel learning methods that fuse together in a non-linear model the information extracted from multiple input features, which represent a range spatial scales. The contribution of the input features is controlled by a set of kernel functions though Multiple Kernel Learning (MKL). The kernel learning model is subject to the kernel width choice as well as to a regularisation constant and the error threshold [Kanevski et.al. 2009]. The choice of these parameters can be routinely selected through cross-validation, training/testing or even ad-hoc trial and error approaches.

The present work proposes to use a modern stochastic optimisation technique to find multiple combinations of the MKL parameters that provide the best predictions. Adaptive stochastic sampling technique is good at searching effectively the multi-dimensional model parameter space to identify the subset of models with better prediction quality. Multi-criteria optimisation provides a way to minimise the objective function based on different components representing different aspects of the prediction quality: matching the test data, matching the target distribution, avoiding predictions outside the realistic range of values.

The proposed approach was applied to the Chernobyl radioactive fallout data, which are well known for their multi-scale character and are difficult to model with conventional stationary algorithms. The problem of multi-scale spatial mapping is tackled by extracting information from spatial features that represent different scales of the phenomenon. The features were derived with Gaussian kernel smoothing from the measurement data using different kernel sizes. This provides a range of input spatial features representing spatial scales from 1 to 50 km. Furthermore, the spatial features were processed to derive gradients and differences, which were also used as MKL inputs.

The methods resulted in multiple optimal prediction maps, which corresponds to a different balance of spatial multi-scale features contributing to the resulting prediction.

REFERENCES
Kanevski M., Pozdnoukhov A., and Timonin V. (2009) Machine Learning Algorithms for Environmental. Spatial Data. Theory, Applications and Software. EPFL Press.

P 17.3

The multipoint Morisita index for the analysis of geodemographic data

Golay Jean[1], Kanevski Mikhail[1] & Vega Orozco Carmen[1]

[1]*Center for Research on Terrestrial Environment, Faculty of Geosciences and Environment, University of Lausanne, CH-1015 Lausanne (jean.golay@unil.ch)*

The present research deals with the application of the functional multipoint Morisita index (Hulbert, 1990; Golay et al., 2013) to geodemographic data. The main objective is to detect and characterize spatial structures in the distribution of the Swiss population at different scales. The population dataset of Year 2000, considered in this study, is based on a high resolution grid (100m x 100m). It can thus be treated as the realization of a point process where each point is associated with a spatial location and the number of inhabitants of the considered hectare.

The classical Morisita index is a global measure of clustering. It measures how many times more likely it is to randomly select two points belonging to the same quadrat (the spatial dataset is covered with a grid of varying cell size) than it would be if the points were distributed at random (i.e. Poisson process). It can be generalized through the multipoint Morisita index (k-Morisita) which takes into account k points with $k \geq 2$. Besides, the k-Morisita index is closely related to the concept of multifractality (Golay et al., 2013), which helps to gain a deeper understanding of its behavior when applied to complex point distributions.

In this study, the fundamental idea was to compare the raw data clustering (i.e. clustering of the considered grid nodes) with the clustering of reference random patterns (produced by shuffling the original one) at different thresholds of the measured function (i.e. the number of inhabitants). The levels of clustering were estimated with the k-Morisita index with $k = 2,3,4,5$; one hundred shuffled datasets were generated and the decile thresholds were applied. It was shown that the k-Morisita index was a powerful tool for studying complex population distributions embedded in geographical spaces characterized by a highly irregular topography.

ACKNOWLEDGEMENTS

The research was partly supported by the Swiss NSF project No. 200021-140658: “Analysis and modeling of space-time patterns in complex regions”.

REFERENCES

Golay, J., Kanevski, M., Vega Orozco, C., & Leuenberger, M. 2013: The multipoint Morisita index for the analysis of spatial patterns, arxiv (http://arxiv.org/pdf/1307.3756.pdf).

Hulbert, S. H. 1990: Spatial distribution of the Montane Unicorn, Oikos, 58, 257-271.

P 17.4

Extreme learning of environmental pollution

Leuenberger Michael[1], Kanevski Mikhail[1]

[1]*Center for Research on Terrestrial Environment, University of Lausanne, Géopolis, CH-1015 Lausanne (michael.leuenberger@unil.ch)*

Extreme learning machine (ELM) is a fast and powerful algorithm being part of the machine learning algorithm category. Developed by G.-B. Huang et al.(2006), it follows the structure of a multilayer perceptron (MLP) with one single-hidden layer feedforward neural networks (SLFNs).

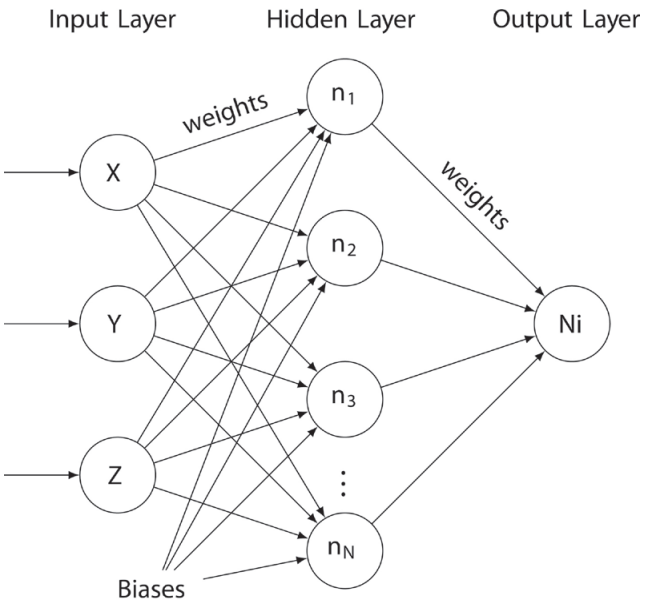


Figure 1. Structure of one single-hidden layer feedforward networks (SLFNs).

The learning step of classical artificial neural networks, like MLP, deals with the optimization of weights and biases by using gradient-based learning algorithm (e.g. back-propagation algorithm). Opposed to this optimization phase, which can fall in local minima, ELM generates randomly the weights between the input layer and the hidden layer and also the biases in the hidden layers. By this initialisation, it optimizes just the weight vector between the hidden layer and the output layer in a single way. The main advantage of this algorithm is the speed of the learning step. In a theoretical context and by growing the number of hidden node, the algorithm can learn any set of training data with zero error. To avoid overfitting, cross-validation method or "true validation" (by randomly splitting data in training, validation and testing subsets) are recommended in order to find the optimal number of neuron. Practically, the optimal number of neuron is catch when the minimum error on the validity set is reached. So with its universal propriety and its theoretical basis, ELM is a good machine learning algorithm which can push the field forward.

The main objective of this study is to: 1) highlight the concept and the theory underlying in this new algorithm, 2) use ELM for environmental pollution data, and 3) compare with geostatistical tools - variography and models – predictors – family of kriging models (Kanevski & Maignan, 2004).

The database used for this study is composed of 200 spatial measurement points in Lake Leman. Each of them has information about the sediment pollution by heavy metals. Focused on the Nickel pollutant we optimize the ELM algorithm in order to draw a spatial prediction map. This process is controlled/verified by geostatistics and by scenario permutation based-model comparison. Although it is a benchmark case study, it allows us to validate the whole methodology from the data validation to the final results via the preprocessing of the data and the data mining.

Other related works in progress use ELM for multidimensional data, multitask learning, features selection and also risk assessment, in particular in natural hazards assessments.

REFERENCES

Huang G.-B., Zhu Q.-Y., & Siew C.-K. 2006: Extreme learning machine: theory and applications. Neurocomputing 70, 489-501.

Kanevski M. & Maignan M. 2004: Analysis and Modelling of Spatial Environmentla Data. EPFL Press.

P 17.5

Random Forest for susceptibility mapping of natural hazards

Leuenberger Michael[1], Vega Orozco Carmen[1], Tonini Marj[1], Kanevski Mikhail[1]

[1]*Center for Research on Terrestrial Environment, University of Lausanne, Géopolis, CH-1015 Lausanne (michael.leuenberger@unil.ch)*

The present research deals with the application of the Random Forest algorithm (RF) to analyze and model natural hazards in a high dimensional input feature space. RF (Breiman 2001) is a type of recursive partitioning methods which involves different classification trees. By a pseudo-random variable selection for each split node, the algorithm grows a variety of classification/regression trees which return different results. A committee system votes (or averages) these results and assigns the predicted values to the unlabeled locations within the validity domain. Furthermore, the algorithm provides the measure of the contribution of each variable. This measure can be used to display the mainly factors affecting the occurrences of the natural hazard under study. In this way, RF allowed assessing the degree of predisposition to the hazard. We applied a binary approach for which the occurrences (i.e. the observations) are displayed against simulated event-locations, randomly distributed in the area where the hazard is highly unlikely to occur. This assumption let us to elaborate susceptibility maps which give an estimation of the probability that a hazard event occurs in a specific area without considering an absolute temporal scale.

This method was successfully applied in the Swiss territory to different study cases, such as landslides, forest fires and permafrost. As results of the application of RF algorithm we obtained an estimation of the importance of different environmental variables for predicting the specific natural hazard occurrence. Moreover, susceptibility maps were elaborated based on the selected variables. The list of the environmental variables influencing the occurrences includes topographic, geological and anthropogenic features, all available at high resolution.

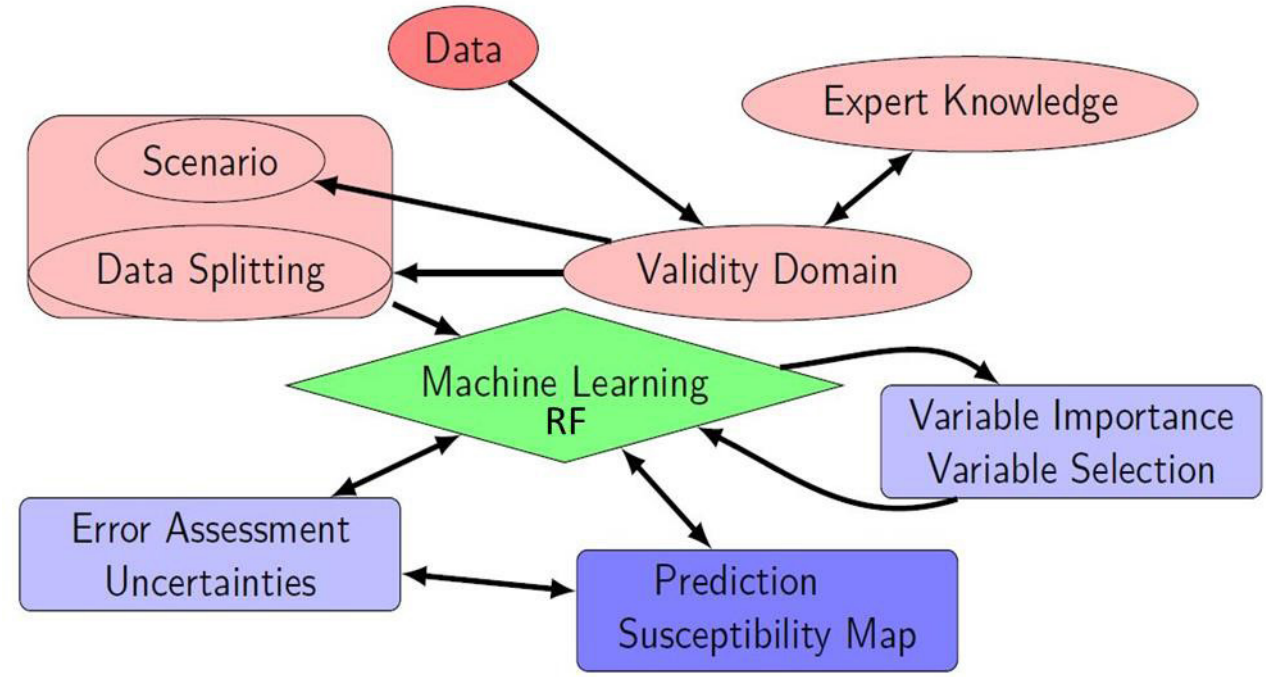


Figure 1. Mind map of the methodology.

REFERENCES

Breiman, L. 2001: Random Forest, Mach. Learn. 45, 5–32.
Liaw, A. & Wiener, M. 2002: Classification and Regression by randomForest. R News 2(3), 18–22.

P 17.6

Evaluation of geostatistical resampling as a proposal mechanism in Bayesian MCMC solutions to hydrogeophysical inverse problems

Ruggeri Paolo[1], James Irving[1] & Klaus Holliger[1]

[1]*Applied Geophysics Group, Center for Research of the Terrestrial Environment, University of Lausanne, Unil-Mouline, CH-1015 Lausanne (paolo.ruggeri@unil.ch)*

Quantification of model parameter uncertainties is essential for hydrological risk assessment and the development of effective groundwater management and/or remediation strategies. While deterministic inverse theory offers a robust and proven framework for estimating spatially distributed subsurface parameters from geophysical and hydrological data, the corresponding parameter uncertainty estimates are well known to significantly underrepresent our lack of knowledge for many problems. To address this limitation, much interest has recently been expressed in the use of stochastic inverse methods for hydrogeophysical parameter estimation and uncertainty analysis. In particular, Markov-chain-Monte-Carlo (MCMC) sampling of the Bayesian posterior distribution has gained significant attention, as it offers the potential for full uncertainty quantification in a relatively straightforward manner. Bayesian-MCMC methods, however, are severely limited by their high computational cost, which results from the typically large size of the parameter space and the need for small model perturbations in order to ensure reasonable rates of proposal acceptance. One key aspect of reducing this computational cost is to incorporate as much prior information as possible into the proposals, such that the number of subsurface configurations tested is limited to a small subset of the total number of possibilities. In this regard, geostatistical simulation methods are attractive, and have gained much recent popularity, because of the inherent flexibility and ease with which they allow us to represent complex prior information as well as the fact that they can be conditioned to a wide variety of measured and previously simulated data. In this work, we evaluate in detail the viability of sequential geostatistical resampling as a proposal mechanism for MCMC methods applied to high-dimensional geophysical and hydrological inverse problems. Focusing on a wide range of realistic crosshole georadar tomographic examples characterized by different numbers of observed data, data error levels, and degrees of model parameter spatial correlation, we investigate the efficiency of different resampling strategies with regard to their ability to generate independent realizations from the Bayesian posterior distribution. We also investigate the potential of a resampling strategy based on a gradual deformation method to generate optimal proposals based on estimates of model parameter sensitivity.

REFERENCES

Cordua, K.S., Hansen, T.M. & Mosegaard, K., 2012: Monte Carlo full-waveform inversion of crosshole GPR data using multiple-point geostatistical a priori information, Geophysics, 77, H19-H31.

Fu, J. & Gomez-Hernandez, J.J., 2009: A BLocking Markov Chain Monte Carlo Method for Inverse Stochastic Hydrogeological Modeling, Mathematical Geoscience, 41, 105-128.

Hansen, T.M., Cordua, K.S. & Mosegaard, K., 2012: Inverse problems with non-trivial priors: Efficient Solution through sequential Gibbs sampling, Computational Geosciences, 1-19.

Hu, L.Y., Blanc, G. & Noetinger, B., 2001: Gradual deformation and iterative calibration of sequential stochastic simulations, Mathematical Geology, 33, 475-489.

Irving, J. & Singha, K., 2010. Stochastic inversion of tracer test and electrical geophysical data to estimate hydraulic conductivities., Water Resources Research, 46, W11514.

Mariethoz, G., Renard, P. & Caers, J., 2010: Bayesian inverse problem and optimization with iterative spatial resampling, Water Resources Research, 46, W11530.

P 17.7

Spatio-temporal aggregation of wildfires: from global cluster to local mapping

Tonini Marj[1], Vega Orozco Carmen[1], Kanevski Mikhail[1]

[1]*Centre for Research on Terrestrial Environment (CRET), University of Lausanne, Quartier UNIL-Mouline, Bâtiment Géopolis, CH-1015 Lausanne (Marj.Tonini@unil.ch)*

Many natural hazards, such as wildfires, can be modelled as stochastic point processes where events are represented as sets of geographical coordinates and time indicating where and when events occurred. Stochastic point processes refer to as sets of random points (events) distributed within a space and/or a time. However, natural events are normally not randomly distributed; instead, they are grouped in clusters over a range of scales. Therefore, the analysis of their spatio-temporal aggregation is of paramount importance to understand predisposing factors, as well as, for prevention and forecasting purposes.

Space-time cluster analysis let us to identify whether groups of observed events are closer in time and in space than expected for a random distribution (absence of structure). The results of this exploratory data analysis allow in turn detecting more vulnerable areas and time-frame periods where hazardous events can more likely occur. Although numerous environmental studies aim at investigating the global spatial clustering behaviour, they often miss a comprehensive analysis of data aggregation both in space and in time.

In the present study authors attempt to detect whether space and time interact to generate clusters and, secondly, to map the results. The case study is represented by wildfire events inventoried in canton Ticino (Switzerland) over the last four decades (data delivered by the WSL Federal Institute). A global cluster indicator, the Ripley's space-time *K*-function, was applied to this purpose.

Computationally, the space-time *K*-function *(K(s,t))* is a bivariate function where space and time represent the two variables of the equation. It is defined as the number of further events occurring within a distance 's' and time 't' of an arbitrary event. The perspective plot of the difference *(D)* between *K(s,t)* and the product of the purely space and purely temporal *K*-function *(K(s)*K(t))* provide a first diagnostic for space-time clustering. Namely, if there is no interaction between space and time, *K*(s,t) equals *K*(s)**K*(t) and the difference is zero, while positive values of *D* indicate space-time interaction at a well detectable scale. Results of our analyses revealed annual clusters with a temporal maximum aggregation over 6 months and a spatial peak clustering around 3 km (Fig.1a). These results helped to setup the optimal bandwidth for the kernel density function which enabled to produce density maps of wildfires for different time periods (Fig.2b).

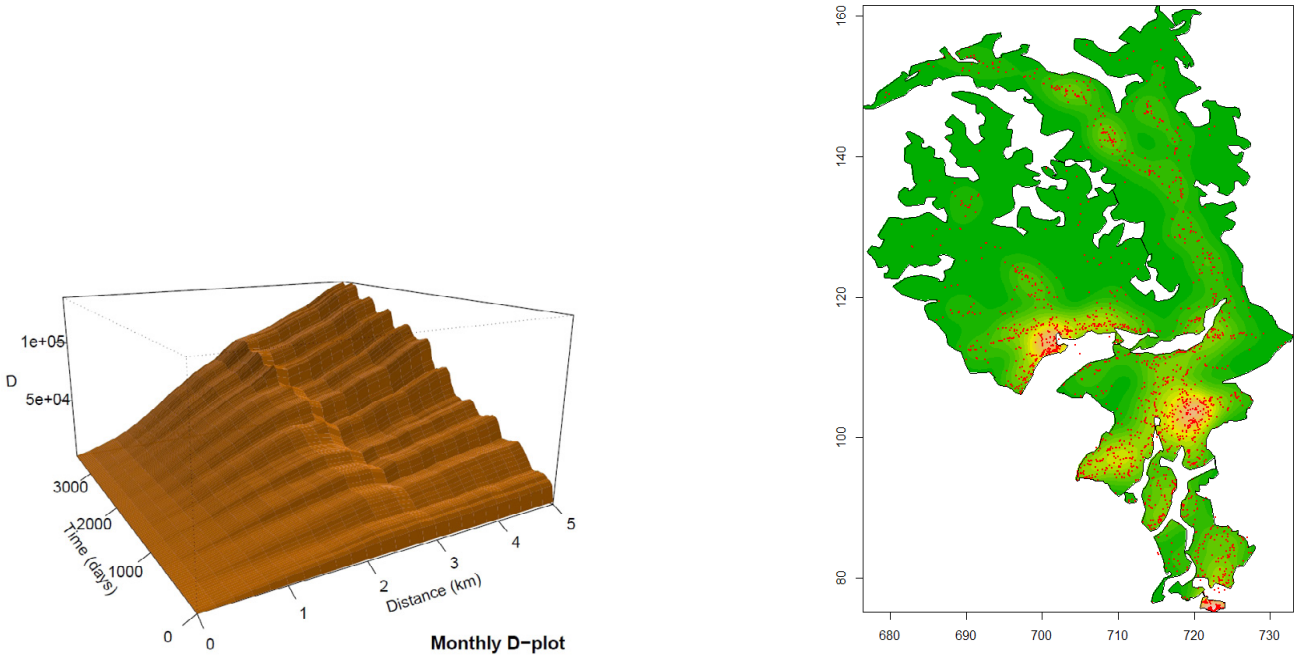


Figure 1. (Left) Perspective plot of the function $D = K(s,t) - K(s)*K(t)$ considering a monthly temporal aggregation. (Right) Density map of forest fires occurrences in Ticino from 1971 to 2010.

REFERENCES

Bivand, R., Rowlingson, B. & Diggle, P. 2012: splancs package in R project.

Diggle, P., Chetwynd, A., Haggkvist, R. & Morris, S. 1995: Second-order analysis of space-time clustering, Statistical Methods in Medical Research, 4(2), 124-136.

Tonini, M., Pedrazzini, A., Penna, I. & Jaboyedoff, M. 2013: Spatial pattern of landslides in Swiss Rhone Valley, Natural Hazard, DOI: 10.1007/s11069-012-0522-9.

Vega Orozco, C., Tonini, M., Conedera, M. & Kanveski, M. 2012: Cluster recognition in spatial-temporal sequences: the case of forest fires, GeoInformatica, 16(4), 653-673.

Acknowledgments

This work was partly supported by the SNFS projects No. 200021-140658: “Analysis and modelling of space-time patterns in complex regions”

P 17.8

Analysis of the dynamic of urban areas and of their interaction with forest fires

Vega Orozco Carmen[1], Tonini Marj[1], Kanevski Mikhail[1]

[1]*Centre for Research on Terrestrial Environment, University of Lausanne, Géopolis, CH-1015 Lausanne (carmendelia.vegaorozco@unil.ch)*

The delimitation of the urban area is an important issue for studies related to urban growth or urban interface. Its definition has to be carefully evaluated since it affects statistical analyses aiming to investigate the distribution of phenomena occurring inside the urban space or the neighbouring areas.

In the present study we applied the "City Clustering Algorithm" (CCA) as proposed by Rozenfeld et al. (2008) to define the urban area. This method is based on the spatial location of the population (e.g. zip code or aggregated grid) and defines the urban areas as the clusters of connected nonzero-populated cells of a defined size (Figure 1). The dynamic of the urban zones was carried out by changing the spatial scales at which the demographic cells were connected. We analysed the Swiss population census data in Canton Ticino for the years 1990, 2000 and 2010 delivered as a hectometric grid (100 x100 m resolution) issued by the Swiss Federal Statistical Office. Changing the cluster's cell sizes (from 100 m up to 500 m) allowed us to aggregate connected clusters and to evaluate the dynamic of the urban space. A first result consisted of an analysis of the urban growth in Canton Ticino in the last three decades.

Additionally, we examined the spatial interaction between the defined urban areas and a natural hazard, namely the forest fires (Figure 2). These events are quite important in Canton Ticino and they are more likely to occur close to the urban spaces (Vega Orozco et al. 2012). Thus, the forest fire events falling inside the urban zones were easily selected and the frequency and spatio-temporal evolution of their occurrences inside the urban clusters were explored.

This method provides an innovative approach to investigate the dynamic of urban areas and the surrounding forest fire occurrences. Furthermore, this can be applied to study and assess complex phenomena delivered or enhanced by the coexistence of both human activities and natural hazards.

Acknowledgments
This work was partly supported by the SNFS Project No. 200021-140658, "Analysis and Modelling of Space-Time Patterns in Complex Regions".

REFERENCES

Rozenfeld, H.D., Rybski, D., Andrade, J.S., Batty, M., Stanley, E. & Makse, H.A. 2008: Laws of population growth. PNAS 105, 18702–18707.

Vega Orozco, C., Tonini, M., Conedera, M. & Kanevski, M. 2012: Cluster recognition in spatial-temporal sequences: the case of forest fires. Geoinformatica 16, 653–673.

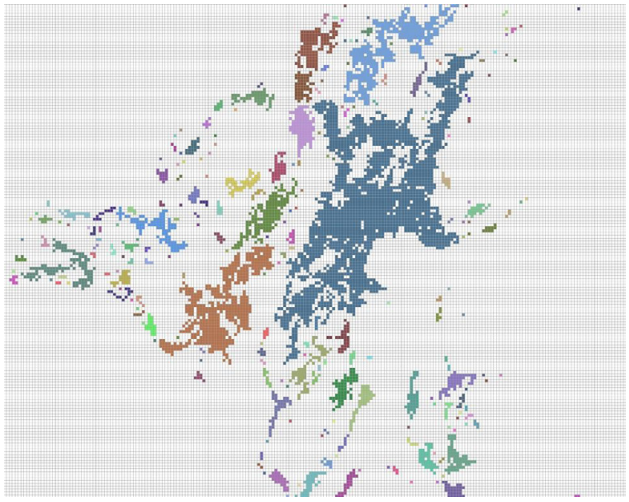


Figure 1. Urban clusters generated by CCA. Same-coloured cells belong to the same cluster.

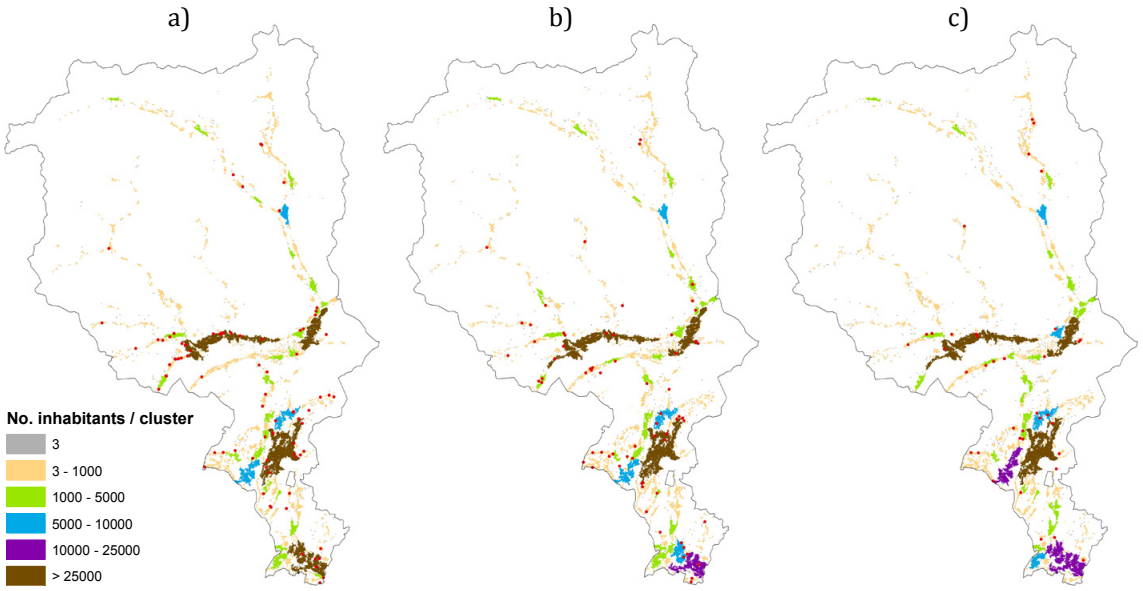


Figure 2. Urban clusters originated from CCA for the three periods in canton Ticino: a) 1981-1990, b) 1991-2000, and c) 2001-2010. Red dots represent the forest fire occurrence inside the urban clusters.