

Mapping and simulations of Geneva soils, using geostatistics and ANN

M. Maignan (1), M. Kanevski, (1), F. Celardin (2), A. Besson, (2)

1. University of Lausanne, Faculty of Geosciences & Environment

2. DIP, Ecole d'Ingénieurs de Lullier, GE

Lausanne, 13th March 2004

Abstract

Over 2500 physical analysis of three components, clays, silt, sand were at disposal by the Engineering School; soil classes had been defined according to swiss and french classifications at the samples location, and had not been valorized since many years. The topology of the sample location, taking into account the lake, the french and Vaud borders, as well as the urban areas without sampling, did not enable a simple and straight forward approach for mapping of soils types, and consequently its suitability for diverse uses. The redundancy of physical analysis, closing to 100% allowed highly valuable checks. In addition to multivariate statistics, methods of Geostatistical estimations and simulations, as well as ANN MLP Multilayer perceptrons algorithms were used in order to establish coherent estimation and simulation maps of the components and of the class types.

Keywords

Soils types, Classification, Geostatistics, ANN, Geostat Office Software, Geneva

1. Introduction

The target was to provide a reliable mapping of the soils classification in a dozen of classes with an correct understanding of the distribution of the 3 components: shales, silt, sand. The maps should be both at scale of the whole canton, and at the scale of zooming at small distances on areas with specified higher interest. The localisation inferred an distinction of several zones a priori: the north side on the right bank of the Lake Lemman, the east side, and different zones in large south west part of the Canton. This study has focused on the mapping of all zones at once, for further consistent representation on GIS maps. The definition of classes show the primacy of shales content, with the complementary importance of silt and sand content.

2. Statistical analysis of Data

The shales are the main discriminant component, and the sand has a behaviour of extreme values, some histograms show clearly a dichotomy. A target of these present investigations was to clearly establish if the dichotomy in 2 subpopulations was a real dichotomy in soil populations spatially also. The histograms of classes according to the french classification and the swiss classification diverge.

3. Variography on shales, silt, sand

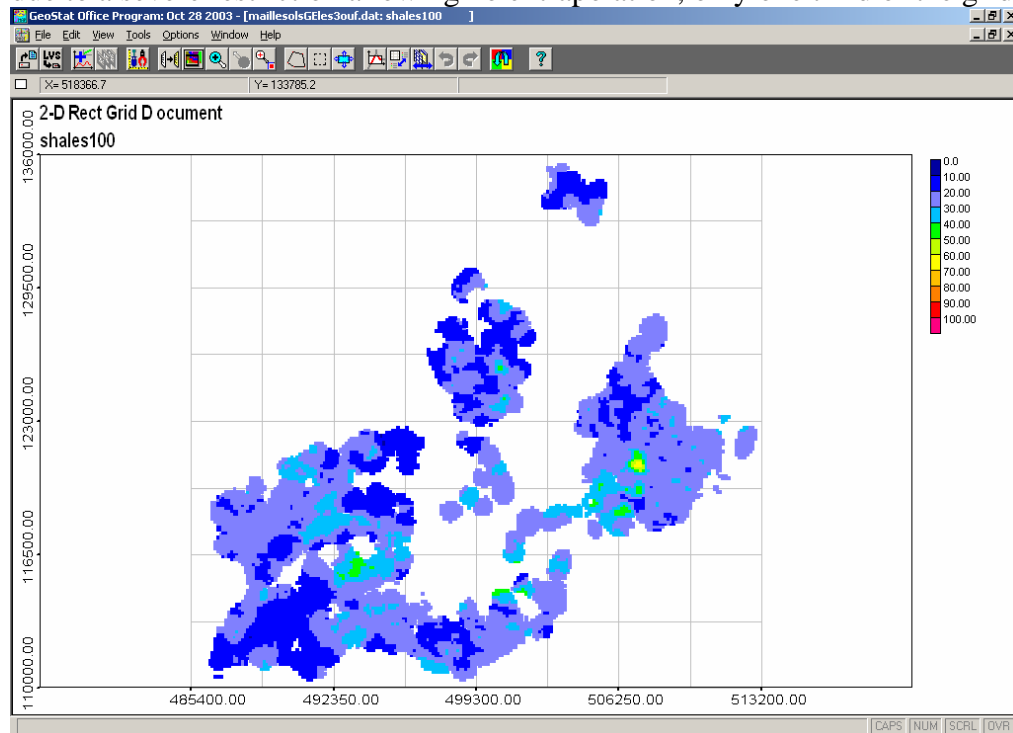
Omnidirectional and directional variography were extensively performed. The directional effects were established as partially actual ones, and partially as artefacts related to the topology of sampling locations. Variography performed on subzones showed distinct behaviours. The nugget effects of some variograms remain in the order of magnitude of

1/3 of the sil, variograms for silts show higher nugget effect. The ranges around 400m-600m are fully compatible with the dense grid of sampling, but do not allow for extension over large zones.

4. Estimations

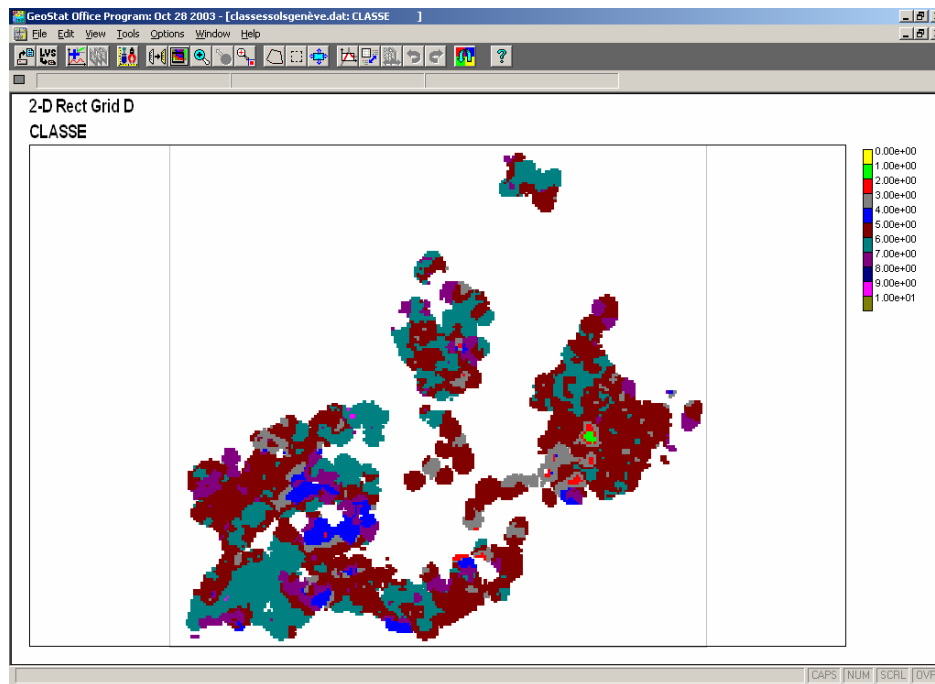
Maps of the different component were established, the map of clays hereafter, established by GSO Geostat Office Software is shown for illustrative purposes. The results are fully in accordance with the expectations of the pedologists. Two zones of restricted areas show contents of shales over 50%.

Coestimation of the 3 chemical-physical components were established and compared to the maps of each independent variable. A main step consists in fact in comparing, and enabling consistency between the mapping of each three variables, and calibrating to 100%, and the co-mapping. The limited zones where discrepancies arise could be related to extreme values zones of anomalies with some spatial structures. The Map of addition of the 3 components verified the overall adjustment close to 100% as of higher quality. With such accuracy close to 100%, the normalized estimation maps were established for the 42'000 grid points (in fact, due to a severe restriction allowing no extrapolation, only one third of the grid was computed)



5. Mapping of Classes

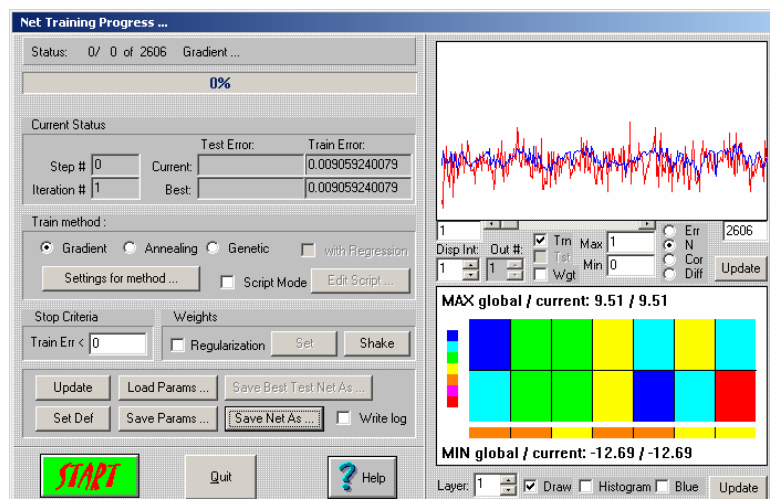
A map of classification has been successfully established; the present Simulations works attempt to propose a probability of occurrence of a class considering the neighbouring classes..



6. Mapping via ANN Artificial Neural Networks

Since the Classification map established with geostatistical methods shows a rather detailed zoning, we chose to extract large scale structure, the ANN MLP Multilayerperceptron approach is our usual method for this purpose.

The first step of training show on the graph below that of course the high isolated values cannot be reconstructed; even if in this case, we use a very large number of iterations, and if we overwnt the step of validation. The graph below show the model learned from the data for the shales.



As illustration, 2 maps of Geostatistics and of ANN MLP compared the detailed mapping and the mapping of larger structures. This is to be completed by indications of probabilities. Maps were consequently established for the residual between localized geostatistical estimations and the ANN MLP large structures estimations, in search for spatial correlations of those residuals.

7. Conclusion

The problem posed by the soil type cartography, when the soil type is defined arbitrarily in classes in imbricated zoning was solved.

This studies has allowed to establish several main results:

- the spatial correlations of the three physico-chemical components
- the kriging mapping with consistent maps, in line with the expectation of pedologists
- the cokriging mapping
- As expected, with NNRK Neural Network Residual Kriging (IAMG95, Osaka), the MLP MultilayerPerceptron mapping delivers a smooth map for the large tendencies which allow to avoid some problems of non-stationnarity.
- Simulation should provide some indication of probability of misclassifications, to put in relationship with the multivariate factorial analysis and the statistical taxonomy.
- The export of the Geostat Office results on a GIS Map, allows to visualize the administrative borders of the communes

8. References

Analysis and Mapping of Environmental Spatial Data, M. Kanevski, M. Maignan, 300p. With software, EPFL Press, Marcel Dekker Press, 2004

M. Kanevski, Pozdnukhov A., McKenna S., Murray Ch., Maignan M. Statistical Learning Theory for Spatial Data. In proceedings of GeoENV2002 conference. Barcelona, 2002.

M. Kanevski et al. Environmental data mining and modelling based on machine learning algorithms and geostatistics. Journal of Environmental Modelling and Software, 2004

M. Kanevski, M. Maignan et al. Advanced geostatistical and machine learning models for spatial data analysis of radioactively contaminated territories. Journal of Environmental Sciences and Pollution Research, pp.137-149, 2003.