

Spatial Component Models with Artificial Neural Networks for Spatially Constrained Regionalization

Michael Govorov ^{a, *}, Giedrė Beconytė ^b, Gennady Gienko ^c

^a Vancouver Island University, Nanaimo, BC, Canada; govorovm@viu.ca

^b Vilnius University, Vilnius, Lithuania, giedre.beconyte@gf.vu.lt

^c University of Alaska Anchorage, Anchorage, AK, USA; ggienko@alaska.edu

* Corresponding author

Keywords: Artificial Neural Network, kriging, Generalized Linear Mixed Model, Spatial Component Model, spatial effect, deep learning, regionalization

Abstract:

The authors have investigated into different geostatistical point data modeling approaches for regionalization purposes that employ the Artificial Neural Network (ANN) techniques. Regionalization is a spatially constrained adjacency classification problem. In this study, regionalization is viewed as classification of spatial objects (non-uniformly distributed points) into a smaller number of geographic regions defined by their spatial and attributive characteristics or regionalized variables. For regionalization, we take into consideration the non-stationarity and autocorrelation properties of the spatial data.

The distinctive aspect of geostatistics is the use of the regionalized variables. Regionalized variables are in between the random variables and the completely deterministic variables. We assume that the spatial variation of the regionalized variables consists of the following components:

- structural deterministic component – a constant mean or trend ($\mu(\vec{s}_i)$) that represents large scale variation,
- random, but spatially correlated stochastic component ($\eta(\vec{s}_i)$) that represents a zero-mean stationary process for small scale variation,
- spatially uncorrelated, zero-mean white random noise ($\varepsilon(\vec{s}_i)$) that consists of micro-scale variation and/or measurement errors, uncorrelated with the $\eta(\vec{s}_i)$. [after Oliver and Webster et al., 2015].

Random regionalized variables $z(\vec{s}_i)$ at locations \vec{s}_i are described using the spatial component model such as $z(\vec{s}_i) = \mu(\vec{s}_i) + \delta(\vec{s}_i)$, where $\delta(\vec{s}_i) = \eta(\vec{s}_i) + \varepsilon(\vec{s}_i)$ is a stochastic component, a zero-mean stationary process.

In the classical geostatistics, random regionalized variables are modeled by using kriging techniques such as ordinary, universal, co-kriging, trans-gaussian, indicator and disjunctive kriging, etc. The data $z(\vec{s}_i)$ used in kriging are observations of random variables $Z(\vec{s}_i)$ at locations \vec{s}_i in a study area D that form a stochastic process $\{Z(\vec{s}_i) : \vec{s}_i \in D, i = 1, 2, \dots\}$. The trend can be estimated using different methods such as: use of polynomials (universal kriging); median polish kriging; intrinsic random function (IRF) kriging; the thin-plate spline [Berke, 1999] and the Multilayer Perceptron (MLP) network [Kanevski et al., 2008]. Normally, a random correlated stochastic component is predicted based on a valid semivariogram γ or covariogram function of the distance d (directional dependence also can be introduced) between any two observable random spatial objects at the locations \vec{s}_i and $\vec{s}_i + d$. A covariogram or a semivariogram measures the structure of spatial dependence of a random process.

In the classical model-based approach, generalized linear mixed model (GLMM) for an observable random variable z with the random effects δ is described as

$$y = g(E(z|\eta)) = X\beta + Z\eta + \varepsilon, y|\eta \sim F,$$

where y is the linear predictor; $g(\cdot)$ is the monotonic differentiable link function that states how the mean of z , $E(z) = \mu$, is related to the linear predictor y ; X is a $n \times p$ matrix of the predictor variables (or covariates) for the fixed effects; β is a $p \times 1$ parameter vector of the fixed effects (the first element is the intercept, if there is one); $Z\eta$ is a stochastic process capturing the spatial dependence; Z is a $n \times r$ design matrix for the random effects; η is a $r \times 1$ vector of the random effects that are assumed to be normally distributed with the mean 0 and variance matrix G such that $E(\eta) = 0$; ε is a vector of random noise (measurement errors for each observation) with variance matrix R such that $E(\varepsilon) = 0$, and $cov\left(\begin{pmatrix} \eta \\ \varepsilon \end{pmatrix}\right) = \sigma^2 \begin{pmatrix} G & 0 \\ 0 & R \end{pmatrix}$; σ^2 is a positive constant; F is the conditional target probability distribution; n is the number of complete cases in the dataset; p is the number of the parameters in the model; r is number of the random effects. In this model, η and ε represent two Gaussian random terms with different correlation structures: η values can be correlated among different observations, while ε is the residual error with independent values for each observation. If there are no random effects, the model is reduced to the generalized linear model $y = g(E(z)) = \beta X + \varepsilon, y \sim F$. Inference can be done by likelihood-based methods or Bayesian methods.

In the model-based geostatistics, classical generalized linear mixed model (GLMM) is used to predict a random spatially correlated component $Z\eta$ by use of the variogram [Cressie, 1993, Diggle and Ribeiro, 2007, et al.]. That is the geostatistical innovation in the GLMM field.

However, many applications of non-spatially adapted ANN (e.g., MLP) show that such ANN models substantially outperform the spatial GLMM. This is attributed to their greater capability for accurate approximation of any continuous and non-linear function. *Nonparametric* ANN models have many advantages over all linear models including GLMM. Feedforward networks learn pattern statistics from the training sets. The GLMM performs *parametric* estimation of the covariance structures. In ANN, the relationship between predictors and targets can be solved by using iterative data fitting via neural networks' back-propagation. ANN training can be implemented with global or local-basis functions. In GLMM implementations, the number of monotonic differentiable link functions $g(\cdot)$ is limited; in contrast, ANN may not require prior knowledge about the statistical distributions of the source variables or the assumptions on linearity. ANN works better with missing data that does not apply to the GLMM.

There have been attempts to model the deterministic component with MLP, for example, Neural Network Residual Kriging [Kanevski et al., 2008], MLP for densification of training data points [Jia et al., 2018]. Shallow Radial Basis Function (RBF) network was used to incorporate spatial information into at three different layers of RBF [Hu and Sung, 2005], and to construct Probabilistic Neural Networks (PNN)/General Regression Neural Network (GRNN) [Specht, 1995]. It has been found that shallow learning (MLP, RBF) with only one hidden layer work better than GLMM.

In this paper, we propose modeling the components of spatial variation of regionalized variables by using deep neural network (DNN) techniques. An efficient way to learn a complicated model is to combine a set of the network layers that are taught sequentially. We have investigated several DNN models for encoding the components of the random regionalized variables measured in irregularly distributed points. We applied two sets of models: (a) local filter-wised DNN models – Convolution Neural Networks (CNN) and Capsule Neural Network (CapsNet), and (b) hybrid learning models that combine different types of deep neural networks with Bayesian probabilistic approaches to model uncertainty. We trained the parameters CNN and CapsNet such as the number of filters, filter sizes and the architecture of the network will be capturing and encoding the trend and the local variations of the geostatistical point dataset separately.

We trained the parameters CNN and CapsNet such as the number of filters, filter sizes and the architecture of the network with the purpose of separate capturing and encoding of separately the trend and the local variations of the geostatistical point dataset.

First, we encoded the global trend of the variation of variables that describe a continuous phenomenon. Then the stochastic variations of the variables were encoded, taking into account the fact that the nearby local points are more strongly correlated than the more distant points. We demonstrated how the ANN techniques can be applied to explore the hierarchy in spatial patterns for the classification purposes. Performance of the tested ANNs have been evaluated.

The proposed modeling approaches have been tested with the attribute-rich point dataset of the geological surveys in Ukraine. The study resulted in a series of maps outlining homogeneous regions of the primary environmental variables (factors), based on their relationships with the concentration of natural uranium in the groundwater. The regionalization of Ukraine by the concentration of uranium enables to determine the regions with different concentration of uranium in drinking water that do not meet quality standards, thus allowing for the better control of the problem.