

Radial Basis Functions Versus Geostatistics in Spatial Interpolations

Cristian Rusu¹ and Virginia Rusu²

1 Pontificia Universidad Católica de Valparaíso, Escuela de Ingeniería Informática, Av. Brasil No. 2241, Valparaíso, Chile, cristian.rusu@ucv.cl, WWW home page: <http://www.inf.ucv.cl>

2 North University of Baia Mare, Faculty of Sciences, Victoriei 76, Baia Mare, Romania, crusu@vtr.net, WWW home page: <http://www.ubm.ro>

Abstract. A key problem in environmental monitoring is the spatial interpolation. The main current approach in spatial interpolation is geostatistical. Geostatistics is neither the only nor the best spatial interpolation method. Actually there is no “best” method, universally valid. Choosing a particular method implies to make assumptions. The understanding of initial assumption, of the methods used, and the correct interpretation of the interpolation results are key elements of the spatial interpolation process. A powerful alternative to geostatistics in spatial interpolation is the use of the soft computing methods. They offer the potential for a more flexible, less assumption dependent approach. Artificial Neural Networks are well suited for this kind of problems, due to their ability to handle non-linear, noisy, and inconsistent data. The present paper intends to prove the advantage of using Radial Basis Functions (RBF) instead of geostatistics in spatial interpolations, based on a detailed analyze and modeling of the SIC2004 (Spatial Interpolation Comparison) dataset.

1 Introduction

A key problem in many fields (including environmental monitoring) is spatial interpolation (sometimes referred as “surface modeling”). It consists of estimating the values of z variable at any location, based on set of (x_i, y_i, z_i) samples, which usually have a non-uniform distribution. Input data represent z values samples at given (x, y) locations, usually called control points. The problem occurs in geology, geophysics, meteorology, environmental sciences, agriculture, engineering, economy, medicine, social sciences, etc. [9], [19], [22], [23].

Two classes of methods are generally used in spatial interpolations: (1) triangulation, and (2) gridding. Triangulation requires a tessellation by an optimal network of

triangles, with control points at all apices. The triangles set represents an approximation of the surface. A regular array of data is generated by gridding, z parameter being estimated on the grid nodes, based on a set of control points. Gridding offers at least two major advantages over triangulation: (1) it is not necessary to sample the extreme points of the surface to be estimated, and (2) subsequent operations on grid data are facilitated. Usually gridding is not an aim by itself; it is a preliminary step for further processing.

2 Geostatistics in Spatial Interpolations

The main current approach in spatial interpolation nowadays is geostatistical. Geostatistics was originated by the application of statistical methods to the study of geological phenomenon. A complex theory was later developed, being applied not only to earth sciences, but also to many other areas: natural, economic, social phenomenon, among others. Geostatistics use regionalized variables, which values are not random; neither are exactly describable by a function. A regionalized variable may consist of a drift component and residual. A third error component has to be considered.

Geostatistical interpolation estimates values by kriging. Kriging is an exact interpolator which uses geostatistical techniques to calculate the autocorrelation between data points, and produce a minimum variance unbiased estimate, taking in consideration the spatial configuration of the underlying phenomenon.

Geostatistics is neither the only nor the best spatial interpolation method. Actually there is no “best” method, universally valid [3], [12], [19], [24]. The choice of interpolation method may vary, mainly according to the type and nature of data, and the aim of modeling. Choosing of a particular method implies to make assumptions. The understanding of initial assumptions, of the used methods, and the correct interpretation of the interpolation results are key elements of the spatial interpolation process. Comparison between methods can be made based on criteria as goodness of representation (errors in honoring control points), dependency on data distribution, number of control points that can be handled, ease of implementation, speed of computation.

3 Soft Computing Methods in Spatial Interpolations

3.1 Soft Computing Methods

Soft computing methods offer the potential of a more flexible, less assumption dependent approach in spatial interpolations. Even if their validness as spatial interpolation methods was proved by many authors, their use in practice is still limited [2], [5], [8], [10], [20], [25]. Soft Computing differs from conventional (hard) computing in that, unlike hard computing, it is tolerant of imprecision, uncertainty, partial truth, and approximation [11].

3.2 Artificial Neural Networks in Spatial Interpolations

Artificial Neural Networks (ANNs) are information processors, trained to represent the implicit relationship and processes that are inherent within a data set [1], [6], [7], [15], [16]. Sometimes spatial relationship between inputs has to be found (like in geology, for instance). Other areas require the identification of both spatial and temporal relationships (meteorology, environmental sciences, etc.).

The original inspiration for ANN was biological; so much of the terminology of ANN reflects this biological heritage. The basic structure of an ANN consists of a number of simple processing units, also known as neurons (nodes). The basic role of each node is to take the weighted sum of the inputs and process this through an activation function. A connection joins the output of one node to the input of another. Each link has a weight, which represents the strength of the connection. The values of all the weights in a network represent the current state of learning of the network, in a distributed manner. These weights are altered during the training process to ensure that the inputs produce an output that is close to the desired value.

A learning function or algorithm is used to adjust the weights of the network during the training phase. Training can be supervised or unsupervised. Hybrid training techniques and reinforcement learning are also used. During the learning period both the input and output vector are supplied to the network. The network then generates an error signal based on the difference between the actual output and the target vector. The error is used to adjust the weights of the network adequately. Following training, input data are then passed through the trained network in its non-training (recall) mode, where they are transformed within the hidden layers to provide the modeling output values.

ANNs have emerged as an option for spatial data analysis approximately a decade ago. Training data are the observation samples used to derive the predictive model. The independent (predictor) variables are known as the input variables, and the dependent variables (response) are known as the output variables. In supervised learning, an ANN makes use of the input variables and their corresponding output variables to learn the relationship between them. Once found, the trained ANN is then used to predict values for the output variables given some new input data set. For unsupervised learning, an ANN will only make use of the input variables and attempts to arrange them based on their properties, hopefully in a way that is meaningful to the analyst.

3.3 Radial Basis Functions in Spatial Interpolations

Radial Basis Functions (RBF) have various applications in practice, due to their simplicity, generality and fast learning stage [11], [13], [14]. RBF are unidirectional ANNs, of hybrid learning (incorporating both supervised and unsupervised learning). Usually RBF have a three layers' architecture: (1) input layer – sends the input information to the hidden layer, (2) hidden layer – composed by non-linear neurons (usually gaussian), and (3) output layer – composed by linear neurons.

The hidden layer's neurons work based on the distance between the input vector and the synaptic vector of each neurons (centroid). Therefore they offer a localized

response, which will have a significant intensity only if the input vector will be located near the centroid. Thus, a radial basis neuron acts as a detector that produces 1 whenever the input is identical to its weight vector, meaning that the input pattern was recognized. The output layer's neurons only compute the weighted sum of the output of the hidden layer.

The radial functions are usually symmetric, but asymmetric (ellipsoidal) functions may also be used. They will then have preferential search directions of the control points used in the interpolation, for a specific grid node. The gaussian functions are not the only type of radial functions that can be used. The type of the radial functions and their parameters are chosen based on the specific problem to solve and the characteristics of the input data.

Some of the reasons to use RBF in spatial interpolations are the following:

- depending on the radial functions type, the RBF model may offer a localized response (therefore is able to identify the local characteristics of the surface to be modeled), or a global response (identifying this way the global characteristics of the surface to be modeled),
- RBF are exact interpolators, honoring the control points when the point coincides with the grid node being interpolated,
- smoothing factors can be employed in order to reduce the effects of small-scale variability between neighboring data points.

4 RBF Versus Geostatistics

The progress made in spatial interpolation is usually presented only in journals or scientific meetings dedicated to statistics, mining, environmental etc. Users who have a different technical background often do not have in-depth knowledge of spatial interpolation methods. That is why the use of new techniques is often discouraging for newcomers. When spatial interpolation methods are integrated in software tools, they are often implemented in such a rigid way that users have no real choice in selecting the best possible method, according to the true nature of data to process, and the aim of modeling. Moreover, many required parameters are fixed, without any possible way to modify them.

The following is a comparison between RBF and geostatistics, at theoretical, correctness and efficiency levels, with special emphasis on method's usability.

4.1 Common Characteristics

The basis kernel of RBF is somehow analogous to variogram in geostatistics. The basis kernel functions define the optimal set of weights to apply to the data points when interpolating a grid node.

Both RBF and geostatistics (kriging) can be used as exact interpolators or smoothing interpolators. RBF will act like a smoothing interpolator when a smoothing factor will be incorporated to the basis function. Kriging will be a smoothing interpolator when an error nugget effect will be specified.

Both RBF and geostatistics are powerful and flexible methods, and are useful for gridding almost any type of data set. They generate quite similar results for most data sets. Computing time increases significantly when using large data sets. Precision of estimation is quite similar, excepting for small data sets, when a proper variographic study is difficult or impossible to perform, and therefore RBF give better results.

4.2 Problems with RBF

RBF architecture is actually imposed by the input data set itself. It is natural to use a number of RBF neurons equal to the number of the available control points, and to center the basis functions on the control point's locations. So a challenging problem when using ANN, the choose of the right architecture, is implicitly solved when using RBF in spatial interpolations.

Another problem to solve is the adequate choose of the type of the radial function to be used, as gaussian function is not always the best choice in spatial interpolation. Some alternative function may be multiquadric, multilog, inverse multiquadric, or natural cubic spline, among others. All these options where tested for real data sets (as the section 5 shows).

Once the radial function was chosen, setting the working parameters is by far less challenging then using geostatistics. Basically only smoothing factors have to be specified.

4.3 Problems with Geostatistics

Before actually performing the kriging, a variographic study has to be done. This may be quite a challenge, especially for inexperienced users. Based on the experimental variogram (obtained from the input data set), appropriate variogram model and adequate parameters have to be chosen. Moreover, many times different theoretical models have to be mixed in a complex all-in-one model.

The variogram is a measure of how quickly things change on the average. The underlying principle is that, on the average, two observations closer together are more similar than two observations farther apart. Because the underlying processes of the data often have preferred orientations, values may change more quickly in one direction than another. As such, the variogram is a function of direction. The variogram is a three dimensional function. There are two independent variables (the direction q , the separation distance h) and one dependent variable (the variogram value $g(q,h)$). The experimental variogram is a curve that displays the groups of variogram pairs on a plot of separation distance versus the estimated variogram.

Variogram modeling is not an easy or straightforward task. The development of an appropriate variogram model for a data set requires the understanding and application of advanced statistical concepts and tools. In addition, the development of an appropriate variogram model for a data set requires knowledge of the tricks, traps, pitfalls, and approximations inherent in fitting a theoretical model to real world data. An inappropriate variogram model can lead to completely false gridding results.

The development of an appropriate variogram model requires numerous decisions. These decisions can only be properly addressed with an intimate knowledge of the data at hand, and a competent understanding of the data genesis (i.e. the underlying processes from which the data are drawn).

The variogram model mathematically specifies the spatial variability of the data set and the resulting grid file. The interpolation weights, which are applied to data points during the grid node calculations, are direct functions of the variogram model. When the variogram is specified for kriging, the following parameters have to be set: sill, range, and nugget, but also the anisotropy information.

5 Study Case: SIC 2004

The Radioactivity Environmental Monitoring (REM) Group of the Institute for Environment and Sustainability at the Joint Research Center (JRC) of the European Commission has organized Spatial Interpolation Comparison Exercises (SIC97 and SIC2004). Participants were invited to estimate values of a variable observed at N locations with the help of a subset of n observed measurements. Once the participants have made their estimates, REM disclosed the true values observed at the $N-n$ locations, so that the participants may assess the accuracy of their approach. The main objective of SIC97 and SIC2004 was to present the diversity of approaches taken by participants facing a problem that is identical for everyone, and to present the latest developments in the field of spatial statistics [3], [4], [17]. They offered an excellent occasion to test methods, compare results, and further orient research in the field of spatial interpolations.

The data used in SIC2004 were daily mean values of gamma dose rates measured in South West Germany, in an area of approximately 400 x 700 km, which includes 1008 monitoring stations. Participants were invited to estimate values of gamma dose rates variable at 808 locations, with the help of a subset of 200 observed measurements. Later on, the true 808 values were published. Additionally, 10 smaller data sets (of 200 observed measurements each one) were published, in order to allow the calibration of the methods and parameters [4]. The location of the 200 input data and the output 808 estimations are shown in fig. 1. All available SIC 2004 data sets were processed by the authors of the present paper, using various gridding methods [18], [21]. Only the results obtained by RBF and geostatistics will be shown and discussed here.

The interpolation results were compared with the real 808 values. The following statistics were used: Mean Error - ME, Mean Absolute Error - MAE, Percentage Mean Error - PMAE, Minimum Error - MIN, Maximum Error - MAX, Percentage Minimum Error - PMIN, Percentage Maximum Error - PMAE, Pearson's Coefficient of Correlation between the estimated and true values - PEAR.

The modeling results obtained by RBF are presented in a 3D view in fig. 1. The modeling results obtained by kriging are presented in a 3D view in fig. 2. Examining the two drawings, one could think that kriging brings more details, but the small differences are due, in fact, only to a different level of smoothness.

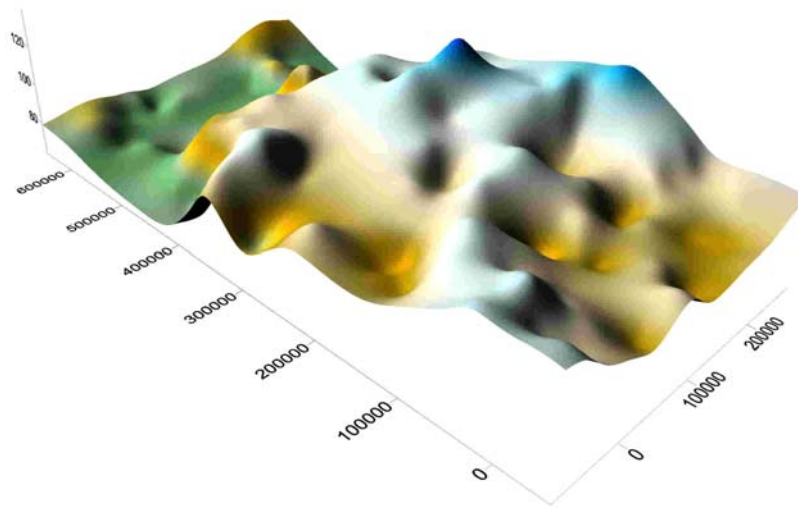


Fig. 1. Modeling results obtained by RBF

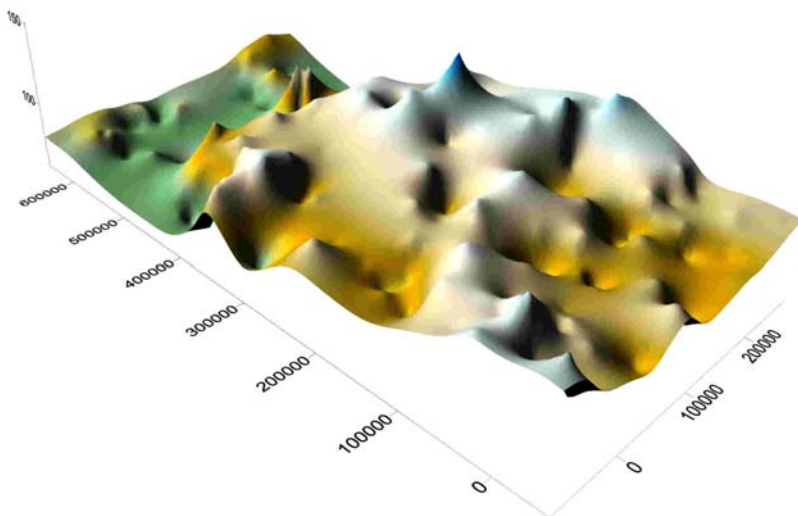


Fig. 2. Modeling results obtained by kriging

Table 1 shows the values of the above-mentioned indicators for RBF and kriging (KRG). RBF and kriging have got similar results, with a slight advantage for RBF over kriging, considering the most significant indicators (MAE, PMAE, PEAR).

Table 1. Statistics of RBF and kriging interpolation results

	ME	MAE	PMAE	MIN	MAX	PMIN	PMAX	PEAR
RBF	-1,41	9,15	9,20	-61,14	44,53	0,00	53,91	0,78
KRG	-1,31	9,28	9,34	-58,39	47,28	0,06	55,82	0,77

Various type of RBF where tested. Table 2 compares the results of applying the following functions: multiquadric - MQ, multilog - MLOG, inverse multiquadric - INVMQ, natural cubic spline - SPLINE.

Table 2. Statistics of interpolation results using various RBF types

	ME	MAE	PMAE	MIN	MAX	PMIN	PMAX	PEAR
MQ	-1,41	9,15	9,20	-61,14	44,53	0,00	53,91	0,78
MLOG	-1,19	9,82	9,94	-71,82	34,15	0,00	45,30	0,77
INVMQ	-12,18	199,40	190,78	-2017	1446	0,42	2004	-0,01
SPLINE	-2,94	53,23	54,59	-635,54	704,15	0,03	706,15	0,17

MQ and MLOG functions give the best results, but INVMQ and SPLINE should not be used in this particular case. The importance of choosing the right type of function is now obvious. Multiquadric-type radial functions offer a more global response than the gaussian-like type, so their use is particularly justified for rather sparse data, like SIC2004 data sets.

The results and the execution time are quite similar for RBF and kriging, but the easy of use of RBF is overwhelming, comparing to the use of kriging. When using RBF, user has to choose only the radial functions type and the smoothing parameter. When using kriging, a complex variogram modeling has to be done.

6 Conclusions

As we saw, spatial interpolation is a key problem in many fields, including environmental monitoring. Even if the main current approach is geostatistical, it is neither the only nor the best spatial interpolation method. There is no “best” method, universally valid. Choosing a particular method implies to make assumptions. The understanding of initial assumption, of the methods used, and the correct interpretation of the interpolation results are key elements of the spatial interpolation process.

A powerful alternative to geostatistics in spatial interpolation is the use of the soft computing methods. They offer the potential for a more flexible, less assumption dependent approach. ANNs are well suited for this kind of problems, due to their

ability to handle non-linear, noisy, and inconsistent data. Particularly useful prove to be RBF.

Both RBF and geostatistics are powerful and flexible methods, and are useful for gridding almost any type of data set. They generate quite similar results for most data sets. Precision of estimation is quite similar, excepting for small data sets, when a proper variographic study is difficult or impossible to perform, and RBF give better results. RBF and geostatistics (kriging) can be used both as exact interpolators and smoothing interpolators.

Using RBF is easier than using geostatistics, even for inexperienced users. As section 4 shows, the geostatistics problems in spatial interpolations are far more complicated than the RBF problems. The development of an appropriate variogram model for a data set requires the understanding and application of advanced statistical concepts and tools. In addition, the development of an appropriate variogram model for a data set requires knowledge of the tricks, traps, pitfalls, and approximations inherent in fitting a theoretical model to real world data. An inappropriate variogram model can lead to completely false gridding results. Variogram modeling is especially difficult for relatively small data sets.

The above-mentioned conclusions were proved based on a detailed analyze and modeling of the SIC2004 (Spatial Interpolation Comparison) dataset, as the 6th section shows. That is way we strongly recommend the use of RBF in spatial interpolations.

7 References

- [1] Aguilar M. and others: *Evaluación de diferentes técnicas de interpolación espacial para la generación de modelos digitales del terreno agrícola*, Mapping Interactivo, Electronic Journal, April, 2002.
- [2] Demyanov V. and others: *Neural Network Residual Kriging Application For Climatic Data*, in Mapping radioactivity in the environment, edited by Dubois G, Malczewski J., and De Cort M., European Commission, Joint Research Center, Luxembourg, 2003, pp. 176-191.
- [3] Dubois G.: *Spatial Interpolation Comparison 97: Foreword and Introduction*, Journal of Geographic Information and Decision Analysis, vol. 2, no. 2, 1998, pp. 1-10.
- [4] Dubois G, Galmarini S.: *Introduction to the Spatial Interpolation Comparison (SIC) 2004 Exercise and Presentation of the Datasets*, Applied GIS, Vol. 1, No. 2, 2005, ISSN 1832-5505 (DOI: 10.2104/ag050009).
- [5] Gilardi N., Bengio S.: *Local Machine Learning Models for Spatial Data*, Journal of Geographic Information and Decision Analysis, vol. 4, no. 1, 2000, pp. 11-28.
- [6] Graubard S.: *El nuevo debate sobre la inteligencia artificial: sistemas simbólicos y redes neuronales*, Gedisa, Barcelona, España, 1999.
- [7] Hiler J., Martínez V.: *Redes Neuronales Artificiales. Fundamentos, modelos y aplicaciones*, Alfaomega, Madrid, Spain, 2000.
- [8] Huang Y., Wong P., Gedeon T.: *Spatial interpolation on overlapping partition surfaces using an optimized dynamic fuzzy-reasoning-based function estimator*, in

- Mapping radioactivity in the environment, edited by Dubois G, Malczewski J., and De Cort M., European Commission, Joint Research Center, Luxembourg, 2003, pp. 201-210.
- [9] Jones Th., Hamilton, D., Johnson C.: *Contouring Geological Surfaces with the Computer*, Van Nostrand Reinhold, New York, 1986.
- [10] Lee S., Cho S., Wong P.: *Rainfall Prediction Using Artificial Neural Networks*, Journal of Geographic Information and Decision Analysis, vol. 2, no. 2, 1998, pp. 253-264.
- [11] Martín B., Sanz A.: *Redes Neuronales y Sistemas Difusos*, Alfaomega, Madrid, Spain, 2002.
- [12] Nagy D. and others: *Comparison of various methods of interpolation and gridding*, XX General Assembly, IUGG, Vienna, Austria, 1991.
- [13] Orr M.: *Introduction to Radial Basis Function Networks*, Center for Cognitive Science, University of Edinburgh, Scotland, 1996.
- [14] Orr M.: *Recent Advances in Radial Basis Function Networks*, Institute for Adaptive and Neural Computation, Edinburgh University, Scotland, 1999.
- [15] Pao Y.: *Adaptive Pattern Recognition and Neural Networks*, Addison-Wesley, Reading, USA, 1989.
- [16] Peter A.: *Rețele neuronale pentru aproximarea și predicția seriilor de timp*, Presa Universitară Clujeană, Cluj-Napoca, Romania, 2000.
- [17] Rusu C.: *Interpolarea spațială a datelor utilizând pachetul de programe ZAZA. Studiu de caz: modelarea cantitatilor de precipitații după un accident nuclear*, Lucrările seminarului de creativitate matematică, vol. 9 (1999-2000), Universitatea de Nord Baia Mare, Romania, pp. 135-146.
- [18] Rusu C.: *Modelado de superficies por métodos basados en redes neuronales artificiales*, Proyecto No. 209.736/2004, Pontificia Universidad Católica de Valparaíso, Chile, 2005.
- [19] Rusu C.: *Modelarea suprafețelor asistată de calculator, cu aplicații în geologie*, PhD Thesis, Universitatea Tehnică Cluj-Napoca, Romania, September 2001.
- [20] Rusu C.: *Neural Network Methods in Surface Modeling. Preliminary Notes*, Creative Mathematics, Vol. 13 (2004), North University of Baia Mare, Romania, pp. 111-120.
- [21] Rusu C., Rusu V.: *Uso de funciones de base radial en monitoreo ambiental*, Actas de la XXXI Conferencia Latinoamericana de Informática (CLEI 2005), Cali, Colombia, 10-14.10.2005.
- [22] Swan A.R.H., Sandilands M.: *Introduction to Geological Data Analysis*, Blackwell Science, Oxford, 1995.
- [23] Watson D.: *nngidr. An Implementation of Natural Neighbor Interpolation*, Vol. I of the Natural Neighbour Series, Claremont, Australia, 1994.
- [24] Wingle W.: *Examining Common Problems Associated with Various Contouring Methods*, Particularly Inverse-Distance Methods, Using Shaded Relief Surfaces, Geotech '92 Conference Proceedings, Lakewood, Colorado, USA, September, 1992.
- [25] Wong K.W. and others: *Rainfall Prediction Using Neural Fuzzy Technique*, in Mapping radioactivity in the environment, edited by Dubois G, Malczewski J., and De Cort M., European Commission, Joint Research Center, Luxembourg, 2003, pp. 213-221.