We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter 18

Radon Risk Analysis Through Geostatistical Tools Implemented in a WebGIS

Veronica Distefano, Sandra De Iaco,
Monica Palma and A. Spennato

Additional information is available at the end of the chapter

1. Introduction

Radon (Rn) is a colorless, odorless, tasteless, inert radioactive gas, and derives from the decay of uranium, which is a radioactive element that is found in small quantities in all sediments and rocks.

The International Agency for Research on Cancer (IARC) and the World Health Organization (WHO) classify Rn pollution as the second leading cause of lung cancer after smoking.

Since Rn is present, in the depths of the Earth, in gaseous phase, it reaches the surface because it interacts with other natural elements, such as uranium, thorium and radio (precursor elements); moreover other geo-lithological features, such as the mineralogical composition of the rocks, the underground permeability levels, the presence of faults, fractures and cavities, affect the transport of the Rn on the surface.

In this paper, the spatial distribution of the Rn concentrations in soil gas over a survey area located in the South of Apulian Region (Italy) and its prediction at unsampled points have been discussed. In particular, Ordinary Kriging (OK), Log-Normal Kriging (LK), Cokriging with indicator variable (ICK) and Kriging with Varying Means (KVM) [11, 21, 26, 27] have been used to predict Rn concentrations over the study area.

In this context, the integration of a Geographical Information System (GIS) and geostatistical tools can certainly support the evaluation of alternative scenarios, possible strategies for a sustainable development [1].

In addition, the geoprocessing tools (e.g., buffering, overlay, union/intersection, interpolation) can be used to combine and interpret data obtained from different sources. The construction of a GIS project facilitates the data sharing and the integration of environmental and demographic data, as well as the results of spatial analysis.

INTECH
open science | open minds

© 2015 The Author(s). Licensee InTech. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Thus, the usefulness of geostatistical techniques for monitoring Rn concentrations in soil gas and the integration of these results in an *ad hoc* WebGIS, named RnWebGIS, have been described. This innovative tool offers dynamic scenarios for monitoring and analyzing different environmental variables [7, 39]. The proposed RnWebGIS is based on an Open Source Software, which enables users to get online spatial and environmental information.

Hence, after a discussion on the integration of GIS and geostatistical tools (Section 2), a brief review of some geostatistical techniques for modeling and prediction purposes has been presented (Section 3). A thorough case study concerning the Rn concentration in soil gas measured in several sites over Lecce district has been developed (Section 4). Finally, the RnWebGIS, where all the available geo-lithological information of the area under study, as well as the Rn prediction results have been integrated, has been proposed and the steps for its implementation have been detailed.

2. GIS and geostatistical tools

Environmental risk management involves the integrated use of geostatistical techniques and several tools provided in a GIS.

A GIS is a repository of tools for capturing, storing, updating, manipulating, analyzing, and displaying all geographic information [9]. The data are stored as collections of relational tables logically associated with each other by shared attributes. On the other hand, Geostatistics provides advanced techniques to predict a variable of interest at unsampled points and to support decisions concerning monitoring, sampling, planning and re-qualification of a territory [10].

Fig. 1 shows a flowchart regarding the steps which characterize the connection between GIS and geostatistical tools.

The first step is based on the creation of a geodatabase, that is a relational database which stores a series of tables holding feature classes, raster datasets, attributes and geographic information represented by shapefiles (i.e. geospatial vector data format which can spatially describe vector features, such as polygon, line or point, with defined shape and geometric location). Then, the next step concerns the georeferencing, that is the process of assigning real-world coordinates to each pixel of the raster data. In this case, these coordinates are often obtained through field surveys, by collecting coordinates with a GPS device, or by using Google Earth software. On the other hand, the shapefile data do not need to be georeferenced, but simply loaded in a geodatabase by specifying the coordinate system associated to the datasets; hence, the vector features in the shapefile will be represented by using the given coordinate system.

The third step involves the implementation of the geostatistical analysis which is organized in various phases, such as the exploratory spatial data analysis (i.e. mapping the spatial distribution using thematic maps), the estimation and modeling the spatial correlation structure exhibited by the data, predicting the variable of interest at unsampled locations by using spatial interpolation techniques (such as Simple, Ordinary and Universal Kriging, Indicator Kriging, Cokriging) and computing the corresponding prediction errors. Finally, the last step consists in the visualization of geostatistical results by producing reliable prediction maps, probability maps, and other useful cartographic representations.

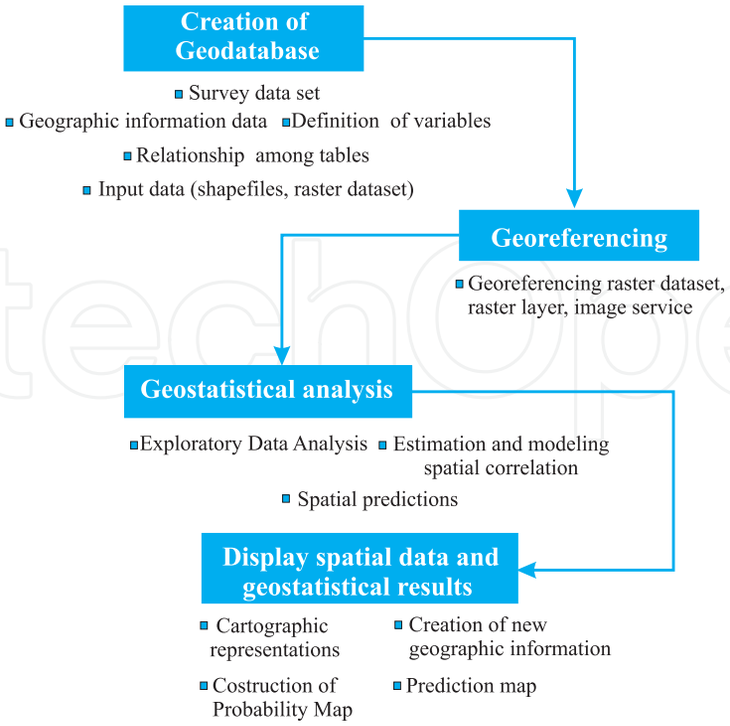


Figure 1. Integration between GIS and geostatistical tools.

The integration between GIS and Geostatistics has been of interest since the early 1990s, when [20] and [8] illustrated the potential benefits of a close link between information systems and Spatial Analysis. Many contributions, concerning the integration of spatial statistical methods and a GIS environment, have been proposed in the literature [2, 3, 6, 9, 14, 37] over the years.

In order to improve the integration between GIS and Geostatistics, different methods of spatial analysis have been implemented in several kinds of GIS softwares [1, 3]. For example, among the most widely used commercial GIS packages, it is worth citing *ArcGIS*, *Geomedia* and *Mapinfo*, which offer extensions for geostatistical analysis; on the other hand, among the *Open Source* packages it is worth mentioning *GRASS* and *Quantum GIS* which use R, Java and Python in order to implement the geostatistical tools. But, the collection of spatial data analysis softwares now available does not completely include all the computational aspects. Therefore, in order to satisfy the specific requirements of the user it is often necessary to develop *ad hoc routines* for sharing GIS data, implementing macros and customized libraries of geostatistical functions in GIS environment, creating suitable script [36].

3. Geostatistical basic notions

Geostatistical methods [15, 23, 28] are useful applied in many areas of environmental sciences (Geography, Geology, Environmental Sciences, Ecology, Meteorology, Agronomy), for studying phenomena which vary continuously over the survey domain. Usually, the final aim of geostatistical analysis consists of obtaining predictions of the variable of interest at unsampled points of the study area. In this case, the use of stochastic models to describe the spatial behaviour of the phenomenon under study, as well as its spatial uncertainty, is reasonable.

In the following sections, the most applied geostatistical techniques, developed for modeling and prediction purposes, both in the univariate and multivariate case, will be briefly described.

3.1. Modeling and prediction in the univariate case

In Geostatistics, the available spatial data concerning the measurements of a single variable, at n points located over the area of interest, are considered as a particular realization of the random variable $Z(\mathbf{u}_\alpha), \alpha = 1,2,\ldots,n$, where $\mathbf{u} \in D \subseteq R^d$, $d \in N_+$. This random variable is called in [25] *regionalized variable*, in order to highlight the feature of the spatial dependency for the variable itself. The spatial distribution of the random variable Z is usually interpreted as the result of two overlapping structures: a systematic structure at the macroscopic level, which describes the global features of the same natural process, and another random structure at the microscopic level, which represents the local irregularity of Z.

In particular, by assuming the existence of the first and the second-order moments of the random variable Z, the basic stochastic model for Z is defined as follows:

$$Z(\mathbf{u}) = m(\mathbf{u}) + Y(\mathbf{u}), \qquad \mathbf{u} \in D \subseteq R^d,\ d \in N_+, \tag{1}$$

where $m(\mathbf{u})$, usually called *trend*, describes the large-scale variations and $Y(\mathbf{u})$ is a second-order stationary random field which represents the local fluctuations of Z.

Under the second-order stationary hypotheses, the expected value of Z exists and does not depend on $\mathbf{u} \in D$; moreover the variogram exists and can be defined as follows:

$$\gamma(\mathbf{h}) = 0.5\, E[Z(\mathbf{u}) - Z(\mathbf{u}+\mathbf{h})]^2, \tag{2}$$

where $\mathbf{h} = \mathbf{u} - \mathbf{u}'$ is the separation vector between two support points $\mathbf{u}$ and $\mathbf{u}'$ of the spatial domain D.

The variogram, by definition, is a measure of dissimilarity, in the sense that it increases when the difference between $Z(\mathbf{u})$ and $Z(\mathbf{u}+\mathbf{h})$ increases; thus, it is usually characterized by small values for short spatial distances.

After computing the sample variogram [32], as a measure of the spatial correlation of the variable of interest, a theoretical admissible model has to be fitted to the estimated variogram.

It is worth noting that the condition for a function to be a variogram is that it be conditionally negative definite [11]. However, given the difficulties to verify this condition, it is advisable for users to look for the best model among the wide parametric families whose members are known to be conditionally negative definite functions. This condition ensures a unique solution in the kriging system (system of linear equations based on the variogram matrix of observations in different spatial points of the domain, as it will be discussed hereafter) used to tackle the general problem of predicting the process of interest, starting from the observed values.

Various classes of variogram models have been proposed in the literature; however the selection of an appropriate class of models is in general based on its geometric features and theoretical properties [10, 12, 13].

Once the spatial correlation of the variable under study has been properly modelled, some well-known based-variogram prediction techniques can be used in order to estimate the analyzed variable at unsampled points over the domain of interest.

The most applied spatial predictor is the Ordinary Kriging (OK) [10, 23] which is based on the following estimator:

$$\widehat{Z}^{OK}(\mathbf{u}) = \sum_{\alpha=1}^{n} \lambda_{\alpha}^{OK}(\mathbf{u}) Z(\mathbf{u}_{\alpha}), \tag{3}$$

where the kriging weights λ_{α}^{OK}, $\alpha = 1, 2, \ldots, n$, depend on the variogram model and the sampling configuration.

The predictor $\widehat{Z}^{OK}(\mathbf{u})$ satisfies both the unbiasedness and the efficiency condition.

In particular, the first property is satisfied if the weights λ_{α}^{OK}, $\alpha = 1, 2, \ldots, n$, of the linear combination are such that

$$\sum_{i=1}^{n} \lambda_{\alpha}^{OK}(\mathbf{u}) = 1. \tag{4}$$

On the other hand, $\widehat{Z}^{OK}(\mathbf{u})$ is efficient, if the weights are chosen in order to minimize the error variance $\sigma_E^2 = E[R(\mathbf{u})]^2 = E[\widehat{Z}(\mathbf{u}) - Z(\mathbf{u})]^2$, under the unbiasedness condition.

Finally, it is important to point out that, if the variable under study is characterized by a high coefficient of variation and/or by different levels of variability, traditional prediction methods, such as OK, might not be appropriate. In particular, when the experimental data shown a skewed distribution, it is possible to apply a non-linear transformation at the observed data, such that the distribution of the transformed data is symmetric. In many cases, the logarithmic transformation is useful for this aim. Hence, structural analysis and kriging are performed by considering the log-transformed data.

3.2. Modeling and prediction in the multivariate case

In many environmental applications, the phenomenon under study is the result of the simultaneous behavior of several variables, which are correlated each other and measured at some points of the area of interest.

In this case, a multivariate spatial structure [34] characterizes the observed data which can be considered as a realization of a multivariate random field (*MRF*) $\{\mathbf{Z}(\mathbf{u}),\ \mathbf{u} \in D \subseteq \mathbb{R}^d\}$, with

$$\mathbf{Z}(\mathbf{u}) = [Z_1(\mathbf{u}), Z_2(\mathbf{u}), \ldots, Z_p(\mathbf{u})]^T, \tag{5}$$

where $p \geq 2$.

Under second-order stationarity, the mean vector of $\mathbf{Z}$ exists and does not depend on $\mathbf{u}$, i.e.:

$$E[\mathbf{Z}(\mathbf{u})] = \mathbf{M} = [m_1, m_2, \ldots, m_p]^T, \qquad \mathbf{u} \in D, \tag{6}$$

and the $(p \times p)$ variogram matrix $\mathbf{\Gamma}$ exists and depends on the spatial separation vector $\mathbf{h}$ between two support points ($\mathbf{u}$ and $\mathbf{u}'$), i.e.:

$$\mathbf{\Gamma}[\mathbf{Z}(\mathbf{u}), \mathbf{Z}(\mathbf{u}')] = E\left\{[\mathbf{Z}(\mathbf{u}) - \mathbf{Z}(\mathbf{u}')][\mathbf{Z}(\mathbf{u}) - \mathbf{Z}(\mathbf{u}')]^T\right\} =$$

$$= \mathbf{\Gamma}(\mathbf{h}) = [\gamma_{ij}(\mathbf{h})], \tag{7}$$

where $\gamma_{ij}(\mathbf{h}),\ i,j = 1,2,\ldots,p,$ is the cross-variogram between $Z_i(\mathbf{u})$ and $Z_j(\mathbf{u}+\mathbf{h})$, when $i \neq j$, and the direct variogram of the $Z_i(\mathbf{u})$, when $i = j$.

In multivariate Geostatistics, *Ordinary Cokriging* (OCK) [34] is a well-known technique used to obtain prediction of one or all the variables under study, at unsampled points of the spatial domain. This technique considers the following linear predictor of the spatial random vector defined in (5):

$$\widehat{\mathbf{Z}}^{OC}(\mathbf{u}) = \sum_{\alpha=1}^{n} \mathbf{\Lambda}_{\alpha}^{OC}(\mathbf{u})\mathbf{Z}(\mathbf{u}_{\alpha}), \tag{8}$$

where $\mathbf{u} \in D$ is an unsampled point in the domain D, $\mathbf{u}_{\alpha} \in D,\ \alpha = 1,2,\ldots,n$, are the sampled points in the domain D and $\mathbf{\Lambda}_{\alpha}^{OC}(\mathbf{u})$ are the $(p \times p)$ matrices of weights, which are determined by ensuring unbiasedness and efficiency conditions.

The cokriging predictor requires a model for the variogram matrix (7). The *Linear Model of Coregionalization* (*LMC*), widely discussed in [23] and [34], is the most commonly model used in the applications. In particular, the *LMC* can be written as

$$\mathbf{\Gamma}(\mathbf{h}) = \sum_{l=1}^{L} \mathbf{B}_l\, g_l(\mathbf{h}), \tag{9}$$

where $\mathbf{B}_l$, $l = 1,2,\ldots,L$, are $(p \times p)$ positive definite matrices, called *coregionalization matrices*, while $g_l(\mathbf{h}),\ l = 1,2,\ldots,L$, are basic variograms which correspond to different scales of spatial variability.

A particular case of cokriging, which allows non-parametric, non-linear predictions [22], is the *Cokriging with Indicator Variable* (ICK), where the values observed for one and/or all the variables under study, are replaced by indicator values (that is 1 or 0, depending on a condition specified by the analyst which is of interest for the phenomenon under study).

Another very useful prediction technique, developed in the context of multivariate Geostatistics, is the *Kriging with Varying Means* (KVM), which corresponds to an extension of the *simple kriging* used in the univariate case. In the context of two or more variables, the

stationary mean of the primary variable is replaced with a known varying mean according to the secondary information [21]. Hence, the KVM predictor is computed as follows:

$$\widehat{Z}^{KVM}(\mathbf{u}) = \sum_{\alpha=1}^{n} \lambda_{\alpha}^{KVM}(\mathbf{u})[Z(\mathbf{u}_{\alpha}) - m^{KVM}(\mathbf{u}_{\alpha})] \tag{10}$$

where $\mathbf{u} \in D$ is the unsampled point, while λ_{α}^{KVM}, $\alpha = 1, 2, \ldots, n$, are the weights of the predictor. Note that if the secondary variable is categorical, the primary local mean is the mean of the primary variable within a specific category of the secondary variable.

In the proposed case study, the above mentioned prediction techniques will be performed and the main prediction results will be discussed.

4. Case study: the survey area and the data set

The study area is located in the Southern part of Italy (Lecce district in the Apulian region), at 199 meters above the sea level (Fig. 2).

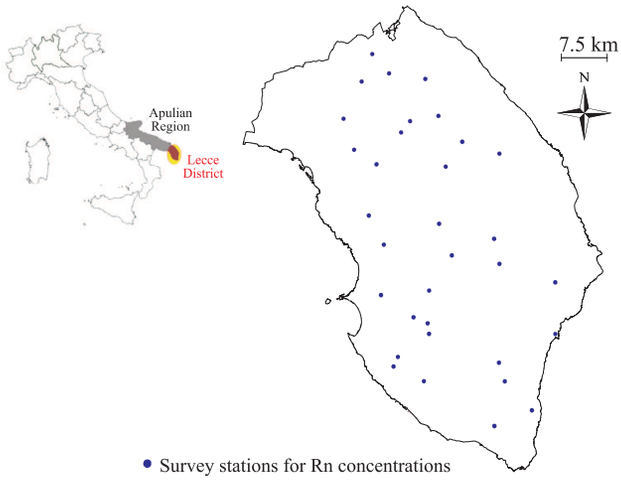


Figure 2. Location map of survey stations in Lecce district.

It is characterized by a limestone substrate and the presence of sink-holes, caves and underground drainages. Moreover, the area is characterized by 6 different lithotypes, with predominance of sandstone (45.7%), limestone and dolomitic limestone (29.4%), as well as siltstone and sandstone (23.9%). Because of this particular lithology, a medium-high level of underground permeability marks the area out. Fig. 3 illustrates the maps of geo-lithological features and permeability levels.

The available geo-lithological information about the survey area represents the *soft* information which can better explain the outdoor Rn concentrations and the corresponding exhalation rate from soil. In order to use the soft information in the modeling and prediction stages of the geostatistical analysis, a regular georeferenced grid covering the area under study (Fig. 4-a)) has been defined. The node separation distance (along the horizontal and vertical directions) has been set equal to 2 km. Thus, the soft data

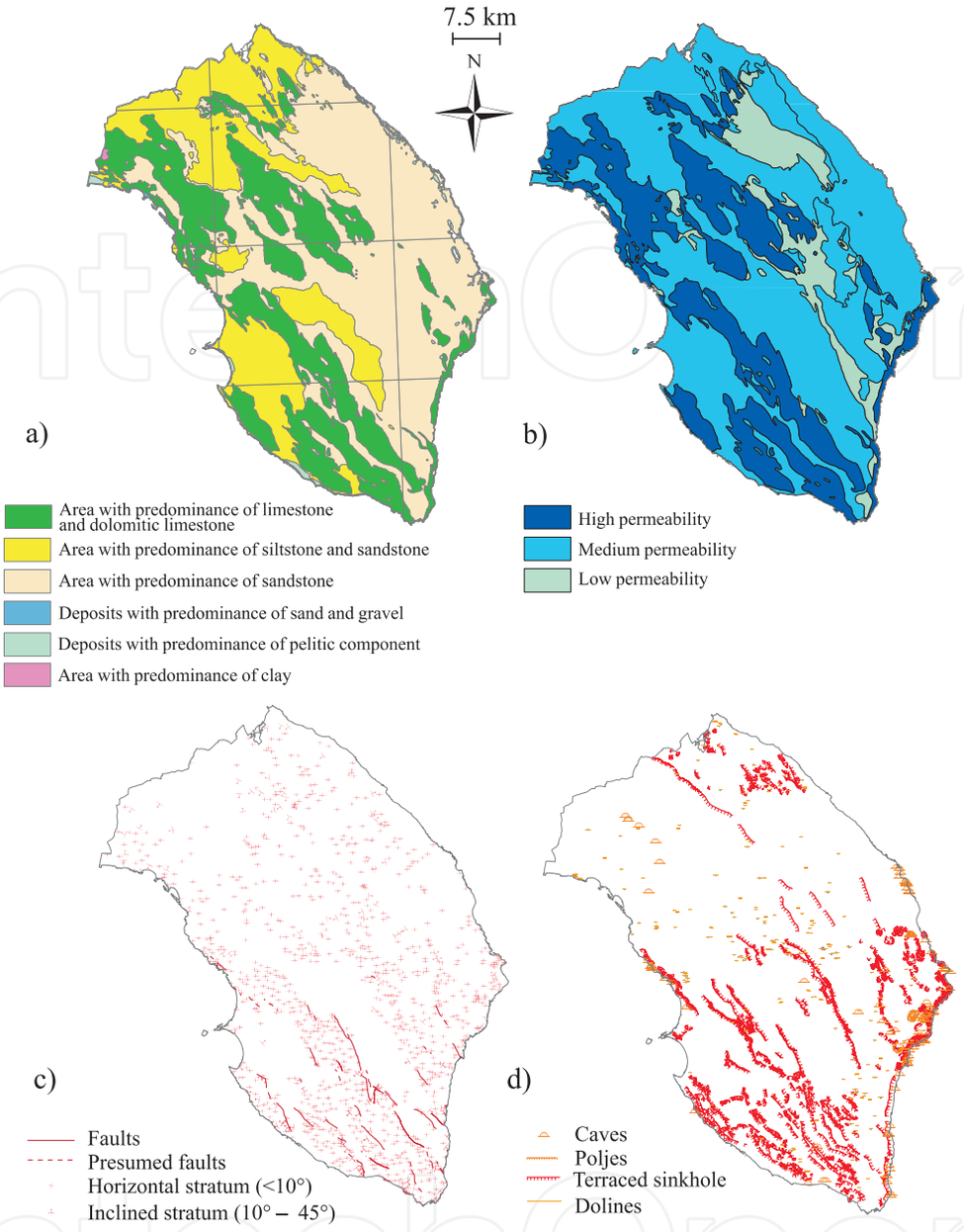


Figure 3. Maps of geo-lithological features in Lecce district. a) Map of geo-lithological features. b) Map of underground permeability levels. c) Map of tectonics features. d) Map of karst forms.

have been stored in a GIS project by using Arcmap of the ArcGIS software and the shapefile (data sources: http://www.sit.puglia.it) concerning the geo-lithological features, the underground permeability level, the tectonic (such as distance from tectonic elements) and karst features (such as distance from karst forms). In particular, by using the *spatial join* tool (a geoprocessing tool of Arcmap), all the soft data have been assigned to each grid point (for example, for each grid point it is possible to identify the type of geo-lithology composition, the distance grid point-faults, the distance grid point-karst forms).

On the other hand, the available Rn concentrations in soil gas have been considered as *hard* information. In particular, the in soil gas Rn levels have been measured at 32 locations over

Lecce district (Fig. 2), during a campaign conducted in June-July 2012 on the basis of a suitable spatial sampling.

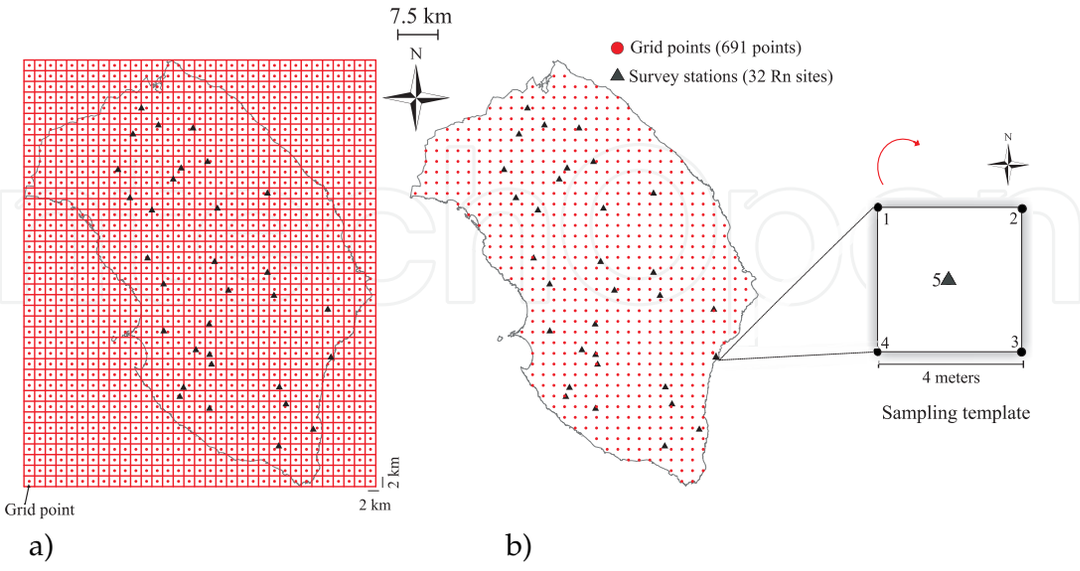


Figure 4. a) Regular grid of cell size 2×2 km and grid points with node separation distance equal to 2 km. b) Survey stations for Rn concentrations in soil gas, together with the sampling template.

For each selected site, the Rn concentrations have been detected by using an active sampling-based system, in which the Rn and its decay products are conveyed in proximity of the radiation detector through a mechanical pumping. A special instrument (AlphaGUARD radon monitor by Saphymo GmbH), equipped with an ionization chamber has been used. The chamber has been connected to a probe, introduced into the ground up to a maximum of 1 meter deep, and to a pump for the aspiration of Rn gas at fixed intervals of time. In particular, at each sample location, the measurements of the Rn levels in soil gas have been performed, every minute for 20 minutes, by aspirating the air through the probe introduced into the ground at 5 points according to a regular template (sampling template in Fig. 4). The template consists of a square with a side of 4 meters; thus the Rn measurements have been made both in the center and the corresponding vertices. Hence, both the center and the vertices of the square have been numbered, starting from the top of the North-West and proceeding clockwise, as considered in the sampling protocol.

Finally, the average of the observed Rn concentrations for each site has been assigned and georeferenced at the central point of the sampling template (number 5 in Fig. 4).

All the 32 sample values (hard data) have been georeferenced and stored in a GIS project; then they have been integrated with the soft information.

4.1. Geostatistical analysis: main results

In the following the spatial concentrations of the Rn in soil gas have been analyzed and predicted at unsampled points of the study area by applying different techniques, such as OK, LK, ICK and KVM.

Note that, the above mentioned multivariate interpolation methods utilize both image-derived (soft) data which are more exhaustive and higher accurate than hard data, typically sparse over the domain. Using ICK and KVM is justified when the measurements of environmental variables can be characterized by a high coefficient of variation and/or by different level of variability; indeed in such cases traditional prediction methods, such as OK and LK, are not appropriate because they do not account for these features.

Ordinary kriging predictions

In order to predict Rn concentrations in soil gas by using OK method, the structural analysis has been performed on the measured values. After computing the sample variogram, the following Gaussian variogram model has been fitted:

$$\gamma(\mathbf{h})^{OK} = 80 + 263 \left[1 - \exp\left(-3||\mathbf{h}||/20\right)^2\right] \tag{11}$$

where nugget and sill correspond to 80 and 263, respectively, while range is equal to 20 km.

Fig. 5-b) shows the estimated Rn concentrations in soil gas obtained by using OK.

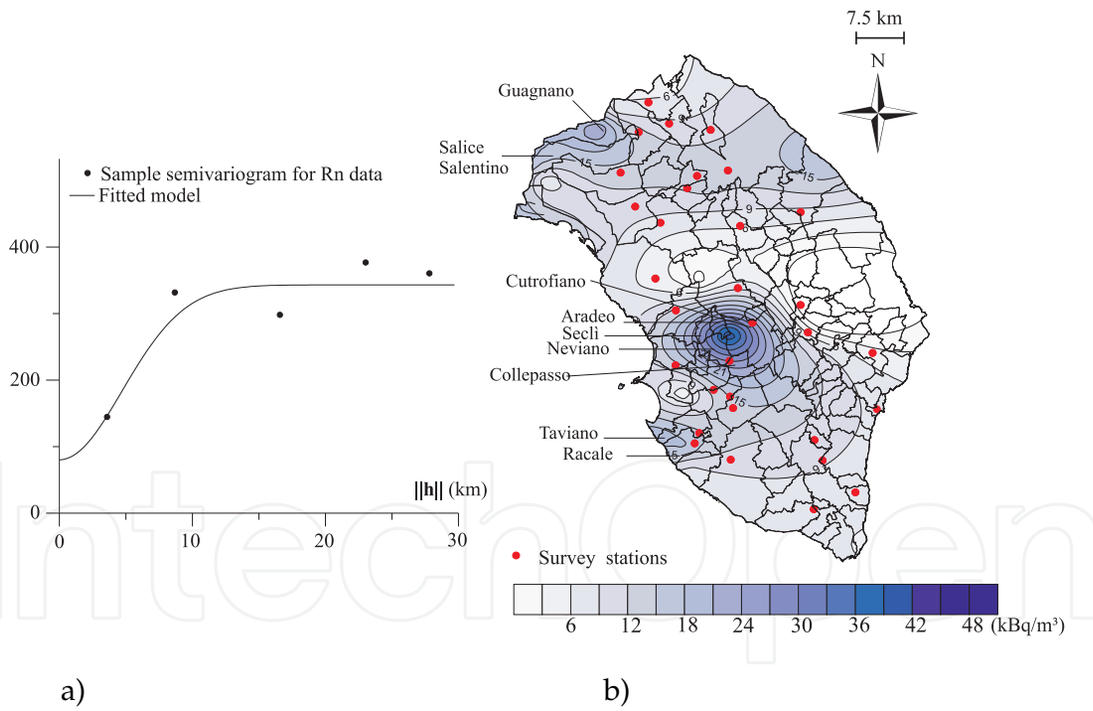


Figure 5. a) Sample variogram and fitted model for Rn concentrations. b) Contour map of the estimated Rn concentrations using Ordinary kriging.

In particular, the prediction map in Fig. 5-b) highlights three Rn prone area, characterized by high Rn concentration values, in the North-Western, South-West and South-South-West of the domain of interest.

Log-Normal kriging predictions

In this case, the Rn measured values have been first processed by applying the logarithmic transformation. Hence, the structural analysis has been performed on the log-transformed data. In Fig. 6-a) the sample variogram computed for the log-transformed data is illustrated together with the fitted model. In particular, the variogram model considered for the LK prediction technique is given below:

$$\gamma(\mathbf{h})^{LK} = 0.1 + 1.2\left[1 - \exp\left(-3||\mathbf{h}||/20\right)^2\right] \tag{12}$$

where nugget and sill correspond to 0.1 and 1.2, respectively, while range is equal to 20 km.

Fig. 6-b) shows the predictions for Rn concentrations obtaining by using LK, after a back-transformation on the data has been applied.

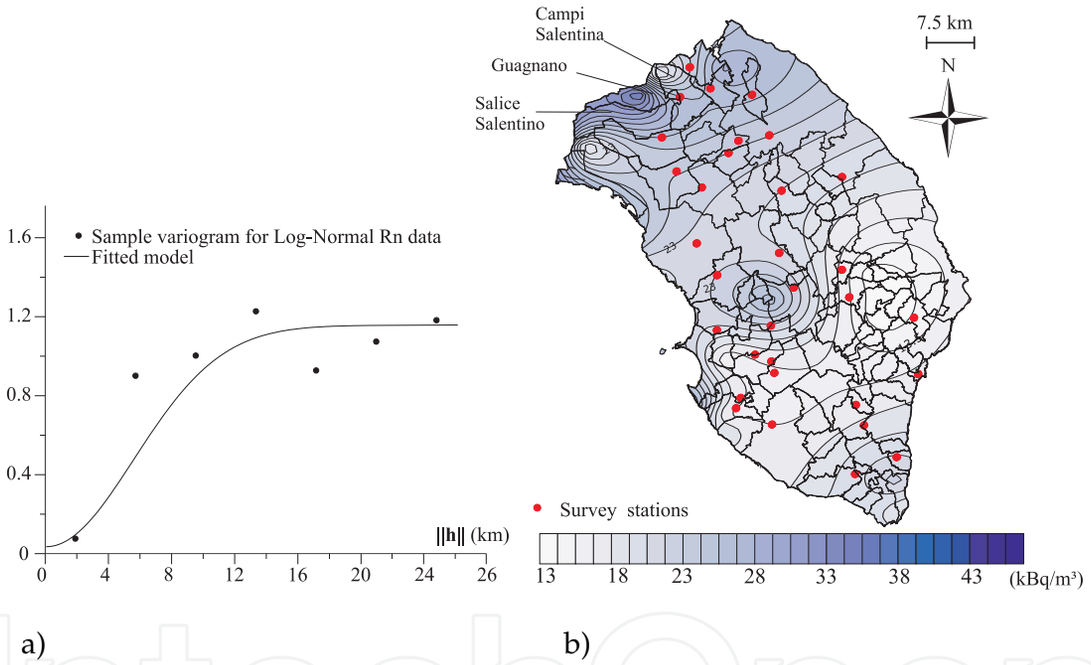


Figure 6. a) Sample variogram and fitted model for Log-Normal Rn concentrations. b) Contour map of estimated Rn concentrations using LK, after a back-transformation on the data.

Despite the previous case (Fig. 5-b)), the prediction map illustrated in Fig. 6-b) highlights only one area with high Rn concentrations, located in the North-Western part of the domain of interest. From the prediction LK results, it is evident the significant *smoothing effect* produced by the interpolation process based on LK.

Cokriging predictions with indicator variable

This method has been applied by using a double source of information, i.e. the sampled Rn concentrations (primary variable) and the underground permeability levels (secondary or auxiliary variable). Evidently the auxiliary information is richer over the study area than the

primary variable, hence it has been used to improve predictions of the Rn concentrations in soil gas. In particular, the auxiliary data, which represent the three permeability classes for any node of the grid, have been derived by the software Arcmap of ArcGIS and the shapefile concerning the underground permeability level in Lecce district. A value has been assigned to each point of the grid and to each survey station, according to the different permeability levels (the value 1 for low permeability, the value 2 for medium permeability and the value 3 for high level of permeability).

In this case, the following *LMC* [33, 34] has been fitted to the sample variogram matrix:

$$\mathbf{\Gamma} = \mathbf{B}_1 \, g_1(\mathbf{h}) + \mathbf{B}_2 \, g_2(\mathbf{h}) \tag{13}$$

where g_1 is a nugget model and g_2 is a gaussian model [23] with unit sill and range equal to 20 km; while $\mathbf{B}_l, l = 1, 2$, are the following coregionalization matrices:

$$\mathbf{B}_1 = \begin{bmatrix} 80 & 2 \\ 2 & 0.125 \end{bmatrix}, \qquad \mathbf{B}_2 = \begin{bmatrix} 263 & 4.8 \\ 4.8 & 0.088 \end{bmatrix}. \tag{14}$$

The sample direct and cross variograms, shown in Fig. 7-a), have been modelled as follows:

$$\gamma_{11}(\mathbf{h}) = 80 g_1(\mathbf{h}) + 263 g_2(\mathbf{h}),$$

$$\gamma_{22}(\mathbf{h}) = 0.125 g_1(\mathbf{h}) + 0.088 g_2(\mathbf{h}),$$

$$\gamma_{12}(\mathbf{h}) = 2 g_1(\mathbf{h}) + 4.8 g_2(\mathbf{h}).$$

Finally, cokriging predictions of Rn concentrations in soil gas have been obtained as shown in Fig. 7-b).

In particular, the prediction map shown in Fig. 7-b) highlights three Rn prone areas, characterized by high Rn concentration values, in the North-Western, South-West and South-South-West of the domain of interest.

Kriging predictions with Varying Means

In order to apply this interpolation technique to the available data set, the Rn concentrations have been grouped by the three permeability classes, then the hypothesis of the presence of different mean levels of Rn concentrations in soil gas has been assumed.

After the structural analysis, the following variogram model (Fig. 8-a)) has been considered:

$$\gamma(\mathbf{h})^{KVM} = 170 \, [1 - \exp(-3||\mathbf{h}||/15)] \tag{15}$$

where sill is equal to 170 and range equals to 15 km.

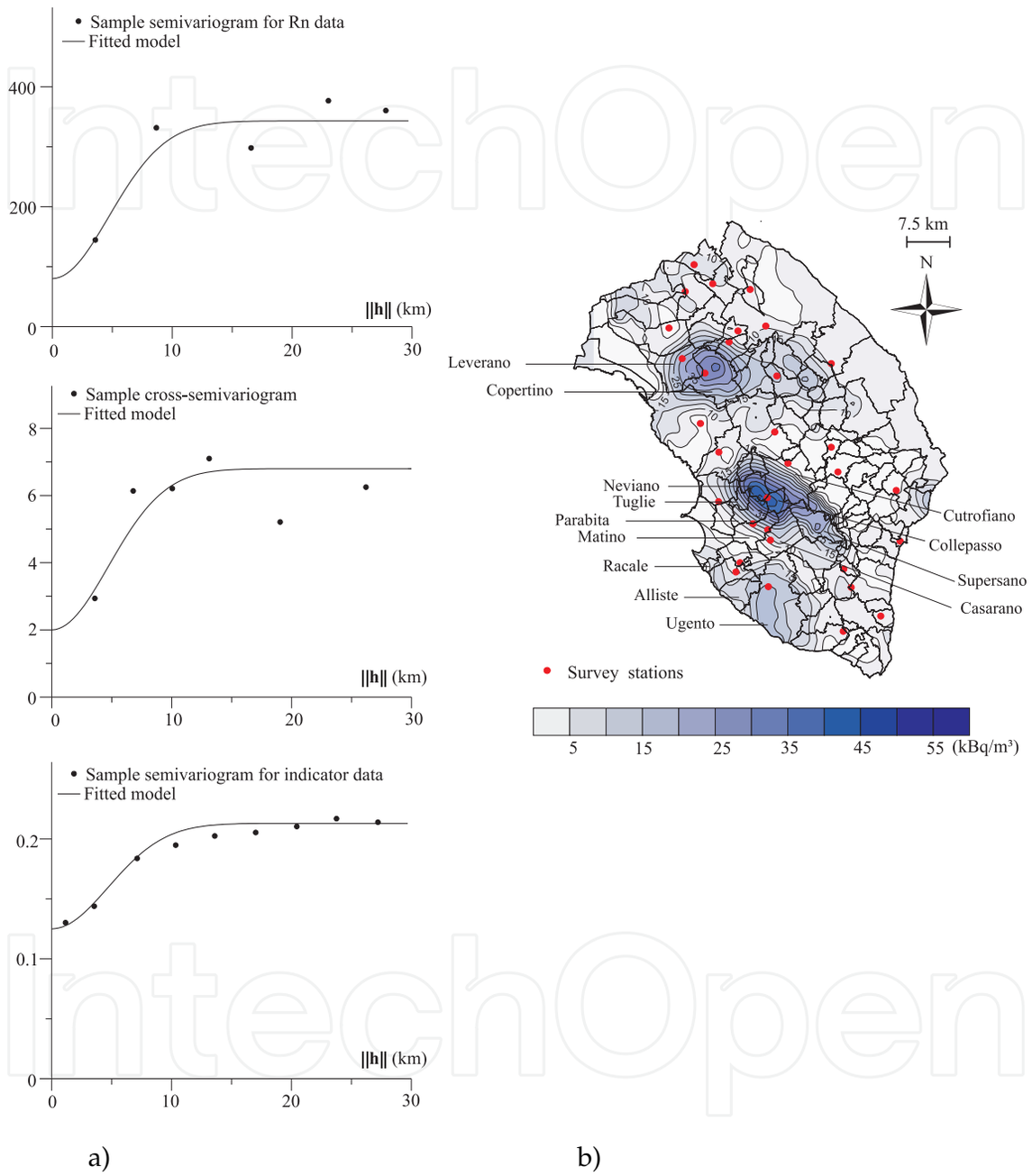


Figure 7. a) Sample direct and cross variograms and fitted model for Rn and indicator data. b) Contour map of estimated Rn concentrations using cokriging with indicator variable.

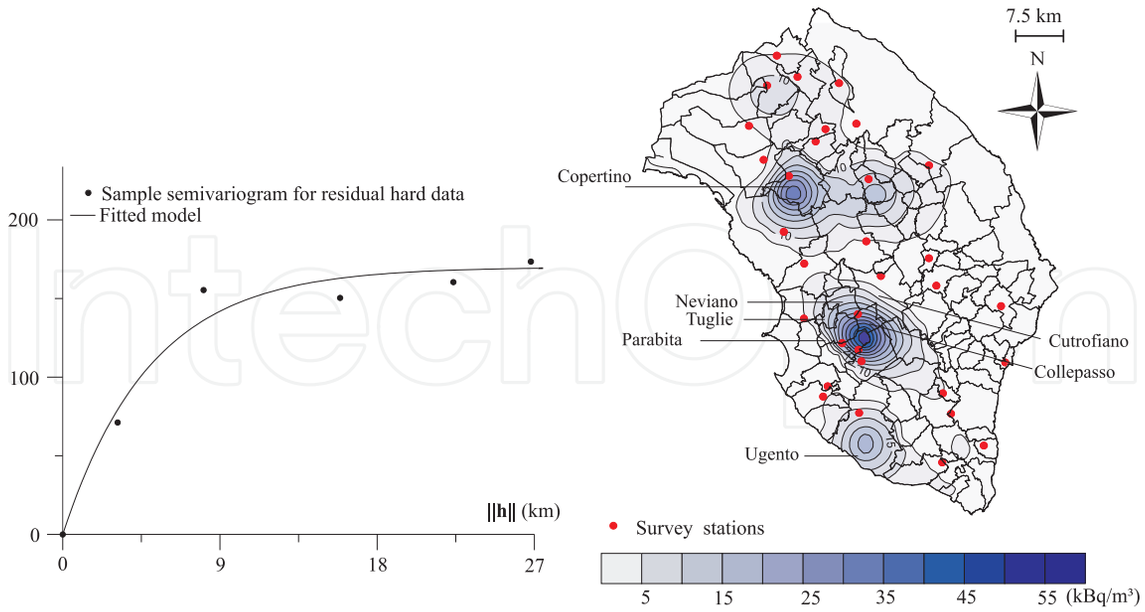


Figure 8. a) Sample variogram and fitted model for residual hard data. b) Contour map of estimated Rn concentrations using KVM.

Hence, predictions for Rn concentrations in soil gas have been obtained for the study area (Fig. 8-b)) by using KVM.

In particular, the prediction map in Fig. 8-b) confirms that high Rn concentration values are located in the North-Western, South-West and South-South-West of the domain of interest.

4.2. A comparison among the prediction results

All the four contour maps illustrated in Figg. 5-b)-8-b) reproduce almost a similar behaviour, with high levels of Rn concentrations located in the Western part of the domain of interest. However, the predictions maps obtained from the KVM and ICK techniques, exploit the auxiliary information and identify the Rn prone areas more accurately than the ones computed by using the OK and LK techniques.

Moreover, the four interpolation techniques have highlighted the existence of a direction (135^o from North to South, close to the Western coast of the study area), along with the Rn concentrations in soil gas present high levels.

The accuracy of the predictions has been measured through the *Root Mean Square Error* (RMSE) index [35, 38], defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{\alpha=1}^{n}[z(\mathbf{u}_\alpha) - \widehat{z}(\mathbf{u}_\alpha)]^2}, \tag{16}$$

where $z(\mathbf{u}_\alpha)$ represent the measured values, $\widehat{z}(\mathbf{u}_\alpha)$ correspond to the estimated values and n is the number of sample data.
By analyzing the RMSE shown in Fig. 9 it is evident that the accuracy of predictions is lower for OK and LK than the other two interpolation techniques.

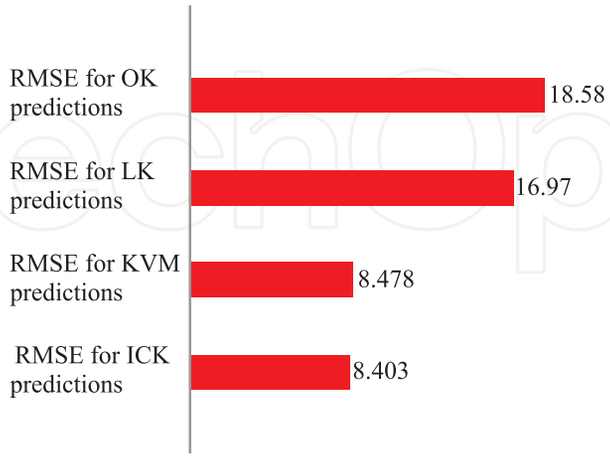


Figure 9. RMSE computed for the OK, LK, KVM and ICK predictions.

However, if the comparison is expressed in terms of the *relative percentage changes* of RMSE index:

- the use of LK method with respect to the OK, improves the predictions reliability of 9.5%;
- the use of ICK method with respect to the KVM, doesn't produce significant improvements for the predictions accuracy (just only 1%);
- the use of multivariate methods (such as KVM and ICK), improve the predictions reliability by about 51%, with respect to LK, and about 55% with respect to OK.

The better performance of the multivariate prediction techniques, compared with the univariate ones (OK and LK), is due to the appropriate use of the information given by the auxiliary variables (soft data). Moreover, among the univariate prediction techniques, the LK has shown a better performance since the variable under study has a skewed distribution.

5. Towards the implementation of a RnWebGIS

The evolution of information technology and GIS allow a potentially unlimited number of users to access and manage geospatial information online [18, 19].

A WebGIS offers an efficient way to provide geospatial information and Web Services to concurrent users without installing a GIS. For this reason a WebGIS system is defined as a web information system that provides geographic information on the web through the Hypertext Transfer Protocol (HTTP) and the HyperText Markup Language (HTML) [17].

Therefore, a WebGIS is different from other web maps, such as Google Earth or Google Maps, since it integrates Internet/web technologies with functionalities of a GIS structure [7].

Fig. 10 shows a schematic diagram regarding the WebGIS architecture, with the following essential steps:

- first step: a user sends a request to the *web browser*;
- second step: such a request is sent via Internet and processed by the server. Note that the WebGIS systems are equipped with a server mapping, such as *MapServer* [24] that obtains information from the geographic databases;
- third step: the request is displayed to the user through the *web browser*.

In the above-mentioned architecture the responsibility is on the server side, while the user (or client) needs only a browser in order to utilize the WebGIS and WebGIS applications. Therefore, the client sends requests to the GIS server which processes it and sends it to the client. Hence, the server performs the task of a *file server*, which contains the mapping data included in the WebGIS, while the client visualizes and consults the cartographic map. Therefore, the main elements of a WebGIS are the following: a) a web server, b) a scripting language, c) a database, d) a server mapping and e) an interface for the server mapping.

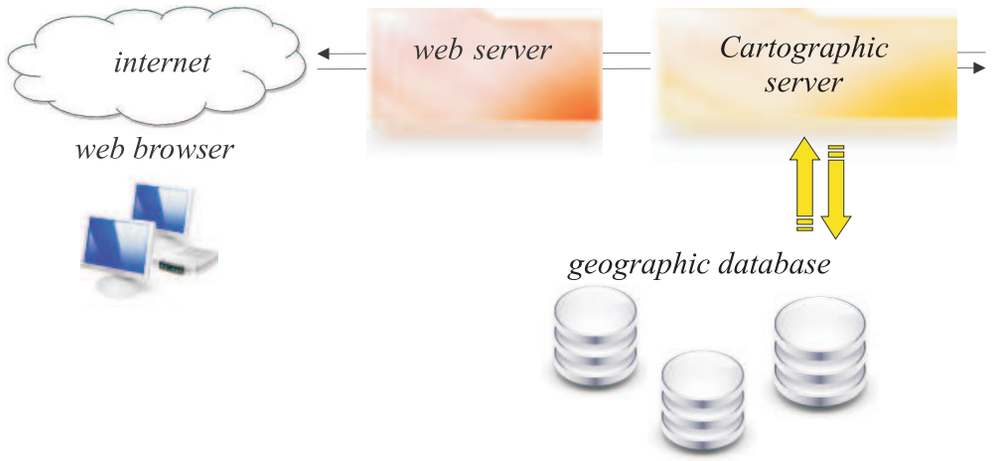


Figure 10. Schematic diagram of the WebGIS architecture.

In this paper, the RnWebGIS has been realized by using Open-source technologies, such as the *web server* Apache, a *MapServer* Project and the client interface *Pmapper*.

In the proposed RnWebGIS, the following databases have been included, such as:

- administrative boundaries:
  - the provincial borders;
  - the municipal boundaries;
  - the urban center boundaries;
- Rn concentrations and the related geographical information (Rn sites);
- the cartographic maps of the permeability and the faults in Lecce district;
- the Web Map Service (WMS), which allows to generate the satellite map of the Lecce district;
- the geostatistical results, which includes the prediction map of ICK and the level curves for the Rn concentrations.

It is important to highlight that the obtained RnWebGIS allows the visualization of alphanumeric information associated with geographic features of interest, through the use of appropriate tables, or to implement the connection to other *web* pages or *database*, by using suitable *hyperlinks* functions.

Through specific queries on thematic maps, the user can display, in tabular form, the required data. For instance, by using the *identify* tool in the navigation bar, the user can obtain detailed information about the "Rn sites " layer as shown in Fig. 11.

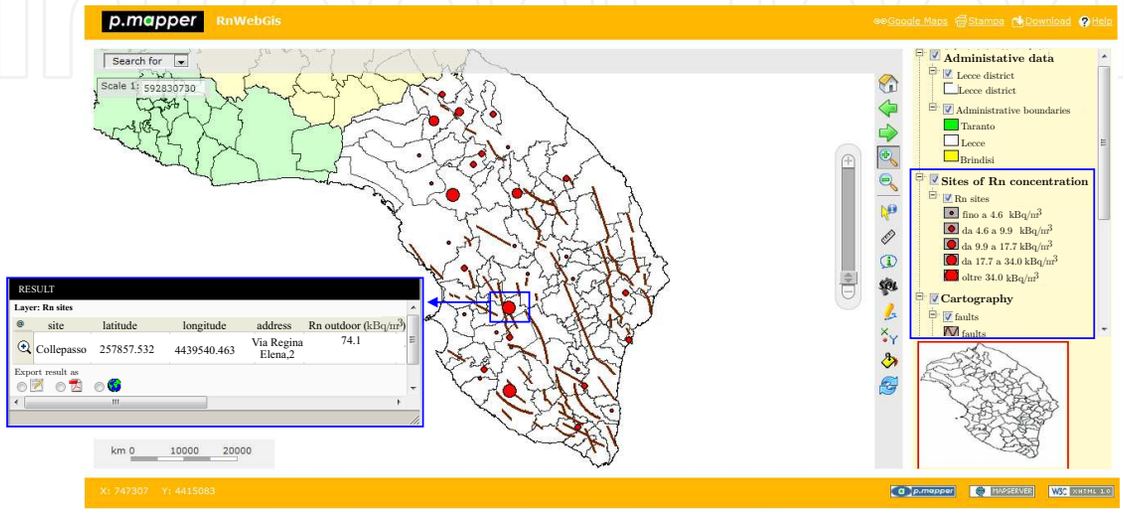


Figure 11. Map Overlay functions: combining layers referred to administrative data, Rn concentrations sites and tectonic faults map. The identify tool highlights an example of Rn site (Collepasso municipality) with high Rn concentrations. Note that the red points indicate the sample locations and their size is proportional to the measured Rn concentration.

Another example of query is obtained by using the *search for* tool, in top left of the *p.mapper* Framework (version 4.1.1.) as shown in Fig. 12.

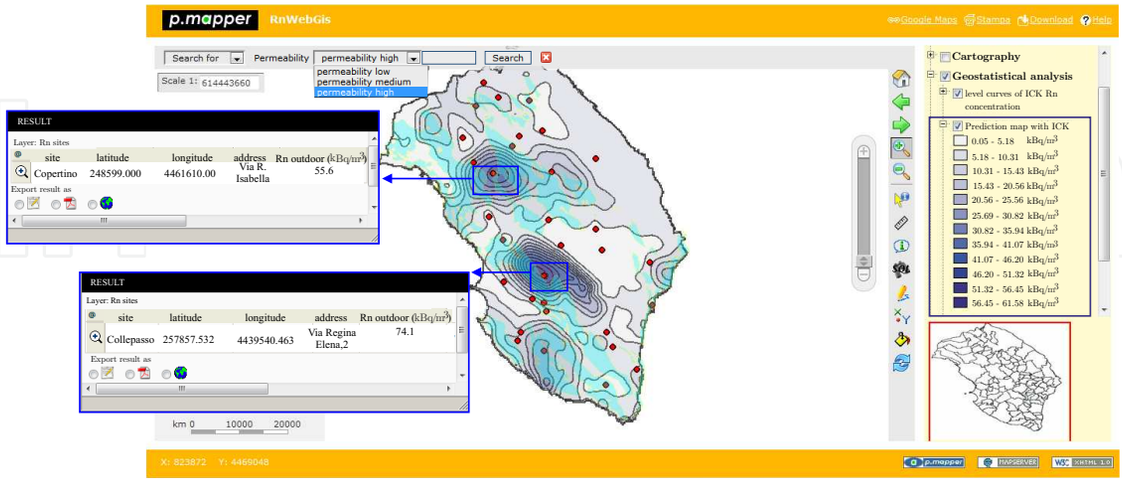


Figure 12. "Search for" tool with the visualization of the layer referred to "high" permeability overlapped on the "Rn site", "level curves of ICK" and "Prediction map of ICK" layers. The identify tool highlights two Rn sites (Collepasso and Copertino municipalities) with high Rn concentrations.

This function has been used to query the database associated with the layer "Permeability" and visualize, by using the thematic map, the study area with high level of permeability overlapped to the layers "level curves of ICK" and "Prediction map of ICK". In this case, the user can identify areas of high level of permeability close to area with high Rn concentrations.

Interactive navigation of maps is possible by using the tools implemented in the RnWebGIS, such as the use of *zooming/panning* available by selecting the object on the map.

On the other hand, in order to ensure the interoperability (i.e. exchange and/or sharing) of geographic data, it has been relevant to build a RnWebGIS according to the specifications of the services, defined by the *Open Geospatial Consortium* (OGC). Indeed, by using the features implemented in the RnWebGIS, the user can also integrate data located on different servers as well as data of different formats. This is possible by recalling the WMS given by the OGC that avoids duplication of data and, at the same time, provides updated geographical data which are certified by shared *standards*.

For example, it is worth highlighting that, by using the appropriate WMS, an orthophoto of Lecce district can be integrated with the layer "Prection map of OCK " as shown in Fig. 13.



Figure 13. Orthophoto of Lecce district, obtained by the WMS and integrated with the "Rn sites", "level curves of ICK" and "Prediction map of ICK" layers.

6. Conclusions

In this paper, after introducing the usefulness of a GIS supported by geostatistical results, and reviewing some geostatistical modeling and predictions techniques used in the univariate and multivariate cases, a case study concerning a thorough geostatistical analysis of the Rn concentrations in soil gas over Lecce district has been discussed. In particular, four different spatial interpolation techniques, that is Ordinary Kriging, Log-normal Kriging, Cokriging with Indicator variable and Kriging with Varying Local Means, have been used in order to predict the Rn concentration levels over the study area. Then, an *ad hoc* WebGIS, called RnWebGIS, has been proposed. The map of the Rn predictions obtained by using Cokriging with Indicator variable, which performed more accurate predictions than the ones obtained by using the other methods, has been stored into the WebGIS, together with several

information about the geo-lithological features (such as permeability and tectonics maps), as well as administrative characteristics (such as demographic data and municipal boundaries) of the study area.
The proposed RnWebGIS highlights the usefulness of the integration between GIS and advanced geostatistical tools, in order to support the management of a sustainable development by taking into account even the risk of exposure to Rn pollution.

Author details

Veronica Distefano, Sandra De Iaco*, Monica Palma, Donato Posa, Alessandra Spennato

*Adress all corresspodence to: s.deiaco@economia.unile.it

University of Salento, Department of Management, Economics, Mathematics and Statistics, Italy

References

[1] Anselin, L. (1992). *Spatial Data Analysis with GIS*, National Center for Geographic Information and Analysis, University of California, USA.

[2] Anselin, L. (2000). Computing Environments for Spatial Data Analysis, *Journal of Geographical Systems* **2**: 201–220.

[3] Bivand, R.S. & Gebhardt, A. (2000). Implementing Functions for Spatial Statistical Analysis using the R Language, *Journal of Geographical Systems* **2**: 307–311.

[4] Beaubien, S.E., Ciotoli, G. & Lombardi, S. (2003). Carbon dioxide and radon gas hazard in the Alban Hills area (Central Italy), *J. Volcanol. Geotherm. Res.* **123**: 63–80.

[5] Beck, H.L. & Gogolak, C.V. (1979). Time-dependent calculations of the vertical distribution of 222Rn and its decay products in the atmosphere, *J. Geophys. Res.* **84**, doi: 10.1029/JC084iC06p03139.

[6] Boots, B. (2000). Using GIS to Promote Spatial Analysis, *Journal of Geographical Systems*, **2**: 17–21.

[7] Bottai, L., De Filippis, T., Mari, R. & Rocchi, L. (2010). *Servizi WebGIS sviluppo sostenibile: la condivisione dell'informazione meteorologica ed ambientale in Toscana*, Proceedings of the 14th National Conference ASITA, Brescia, Italy.

[8] Burrough, P.A. (2001). GIS and Geostatistics: Essential Partners for Spatial Analysis, *Environmental and Ecological Statistics* **8**: 361–377.

[9] Burrough, P.A. & McDonnell, R.A. (1998). *Principles of Geographical Information System*, Oxford University Press.

[10] Chilés, J. & Delfiner, P. (1999). *Geostatistics - Modeling spatial uncertainty*, Wiley.

[11] Cressie, N. (1993). *Statistics for Spatial Data*, Wiley, New York.

[12] De Iaco, S. & Posa, D. (2013). Positive and negative non-separability for space-time covariance models, *J. Stat. Plan. Infer.* **143**: 378–391.

[13] De Iaco, S., Posa, D. & Myers, D.E. (2013). Characteristics of some classes of space-time covariance functions, *J. Stat. Plan. Infer.* **143**: 2002–2015.

[14] De Iaco, S., Distefano, V. & Posa, D. (2013). Geostatistica e GIS, Tecniche e strumenti per il monitoraggio ambientale, in: *Lo sviluppo sostenibile: Ambiente, Risorse, Innovazione, Qualità*, Franco Angeli Editore, Italy.

[15] Delhomme, J.P. (1978). Kriging in the hydrosciences, *Adv. Water Resources* **1**: 251–266.

[16] Durrani, S.A. & Badr, I. (1995). Geostatistically controlled field study of radon levels and the analysis of their spatial variation, *Radiat. Meas.* **25**, 565–572.

[17] Gillavry, E.M. (2000). *Carthographic aspects of WebGIS*, Software Department of Cartography, Utrecht University.

[18] Goodchild, M.F. (1987). A spatial analytical perspective on geographical information systems, *Int. J. Geogr. Inf. Sci.* **1**: 327–334.

[19] Goodchild M.F. (1992). Geographical data modeling, *Comput. Geosci.* **18**: 401–408.

[20] Goodchild, M.F., Haining, R., Wise, S., Arbia, G., Anselin, L., Bossard, E.G., Brunsdon, C., Diggle, P., Flowerdew, R., Green, M., Griffith, D.A., Hepple, L., Krug, T., Martin, R.J. & Openshaw, S. (1992). Integrating GIS and Spatial Data Analysis: Problems and Possibilities, *International Journal of Geographical Information Science* **6**: 407–423.

[21] Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*, Oxford University Press, New York.

[22] Journel, A.G. (1983). Non-parametric estimation of spatial distribution, *Mathematical Geology* **15**: 445–468.

[23] Journel, A.G. & Huijbregts, C.J. (1981). *Mining Geostatistics*, Academic Press, London.

[24] Kropla, B. (2006). *Beginning MapServer: Open Source GIS development*, Berkeley Apress Ed.

[25] Matheron, G. (1962). *Traité de Géostatistique appliquée*. Mémoires du Bureau de Recherches Geologiques et Minieres, Editions Technip, Paris, Tome I, n.14.

[26] Matheron, G. (1965). *Les variables régionalisées et leur estimation: une application de la théorie des fonctions aléatoires aux sciences de la nature*, Masson, Paris.

[27] Matheron, G. (1976). A simple substitute for conditional expectation: the disjunctive kriging. In: *Advanced Geostatistics in the mining industry*, Guarascio M. et al. (eds): 221:236.

[28] Monestiez, P., Courault, D., Allard, D. & Ruget, F. (2001). Spatial interpolation of air temperature using environmental context: Application to a crop model. *Environmental and Ecological Statistics* **8**: 297–309.

[29] O'Dea, J. & Dowdall, M. (1999). Spatial analysis of natural radionuclides in peat overlying a lithological contact in Co. Donegal Ireland, *J. Environ. Radioactiv* **44**: 107–117.

[30] Oliver, M.A. & Khayrat, A.L. (2001). A geostatistical investigation of the spatial variation of radon in soil, *Comput. Geosci.* doi:10.1016/S0098-3004(00)00133-3.

[31] Porstendorfer, J. (1994). Properties and behavior of radon and thoron and their decay products in the air, *J. Aerosol Sci.* **25**: 219–263.

[32] Posa, D. & De Iaco, S. (2009). *Geostatistica. Teoria e Applicazioni*, Giappichelli Ed., Turin, Italy.

[33] Wackernagel, H. (1988). *Geostatistical techniques for interpreting multivariate spatial information*, in: Quantitative Analysis of Mineral and Energy Resources, Chung, C.F. et al. (Eds), 349–409.

[34] Wackernagel, H. (2003). *Multivariate Geostatistics: an introduction with applications*, Springer, Berlin.

[35] Wang, J., Krudy, G., Hou, T., Zhang, W., Holland, G., Xu, X., (2007). *Development of reliable aqueous solubility models and their application in druglike analysis*, Journal of chemical information and modeling. 47(4): 1395-1404.

[36] Wesseling, C.G., Karssenberg, D., Burrough, P.A. & Van Deursen, W.P.A. (1996). Integrating Dynamic Environmental Models in GIS, *Transactions in GIS* **1**: 40–48.

[37] Wise, S., Haining, R. & Ma, J. (2001). Providing Spatial Statistical Data Analysis Functionality for the GIS User: The SAGE Project, *International Journal of Geographical Information Science* **15** (3): 239–254.

[38] Yang, Y., Zhu, J., Tong, X., Wang, D., (2009). The spatial pattern characteristics of soil nutrients at the field scale, Computer and Computing Technologies in Agriculture II **Vol. 1**, Springer US: 125–134.

[39] Zhu, H.C., Charlet, J.M. & Poffijn, A. (2001). Radon risk mapping in southern Belgium: an application of geostatistical and GIS techniques, *Sci. Total Environ.* **272**: 203–210.

IntechOpen

IntechOpen