A Geostatistical Approach to Define Guidelines for Radon Prone Area Identification

Riccardo Borgoni[1], Piero Quatto[1], Giorgio Somà[1]
Daniela de Bartolo[2]

[1]Dipartimento di Statistica, Università degli Studi di Milano - Bicocca
[2]ARPA della Lombardia, Settore Aria e Agenti Fisici, Milano

Abstract. Radon is a natural radioactive gas known to be the main contributor to natural background radiation exposure and the major leading cause of lung cancer second to smoking. Indoor radon concentration levels of 200 and 400 Bq/m^3 are reference values suggested by the 90/143/Euratom recommendation, above which mitigation measures should be taken in new and old buildings, respectively, to reduce exposure to radon. Despite this international recommendation, Italy still does not have mandatory regulations or guidelines to deal with radon in dwellings. Monitoring surveys have been undertaken in a number of western European countries in order to assess the exposure of people to this radioactive gas and to identify radon prone areas. However, such campaigns provide concentration values in each single dwelling included in the sample, while it is often necessary to provide measures of the pollutant concentration which refer to sub-areas of the region under study. This requires a realignment of the spatial data from the level at which they are collected (points) to the level at which they are necessary (areas). This is known as *change of support problem.* In this paper, we propose a methodology based on geostatistical simulations in order to solve this problem and to identify radon prone areas which may be suggested for national guidelines.

Keywords: Radon Prone Areas, kriging, geostatistical conditional simulation, change of support problem

1. Introduction

Radon (the ^{222}Rn isotope with a half-life of 3.8 days) is a naturally occurring decay product of uranium commonly found in rocks and soils. It is an odorless, colorless and tasteless gas that drifts upward through the ground to the Earth's surface. It can reach high levels of indoor concentrations, entering through cracks or holes in foundations and concrete floors however building materials and the water supply can also be sources. Moreover, it is known to be the main contributor to natural background radiation exposure. From miner studies, it was established that the health risk related to radon and radon progeny exposure is lung cancer. In fact radon is also considered to be the major leading cause of lung cancer second to smoking. The American Environmental Protection Agency (EPA) estimates that 7000 to 30000 annual lung cancer deaths in the USA are caused by exposure to residential radon.

Indoor radon concentration levels of 200 and 400 Bq/m^3 are reference values suggested by the 90/143/Euratom recommendation, above which mitigation measures should be taken in new and old buildings, respectively, to reduce exposure to radon. Despite this international recommendation, Italy still does not have mandatory regulations or guidelines to deal with radon in dwellings.

Monitoring surveys have been promoted in a number of western European countries in order to assess the exposure of people to this radioactive gas and to identify *radon prone areas* (Dubois and Bossew, 2006)

A radon prone area according to Italian legislation is defined as an area in which "there is a high probability of finding high radon (indoor) concentrations" (art. 10-ter, comma 2, D.L.vo 241/00). Implicitly many authors refer to Radon prone areas as areas where there is a high probability of exceeding reference values, and, in the absence of national values, the 90/143/Euratom recommendation ones are often used.
Surveys provide concentration values in a single dwelling included in the sample. Geostatistical techniques have been used in recent years to map radon indoor concentration and generate probability maps of exceeding a given threshold. The usual approach consists of mapping radon concentration through punctual predictions using kriging-like procedures (Zhu et al. 2001, Franco-Marina et al. 2003, Dubois et al., 2007). Some authors suggested to use hierarchical modeling for punctual predictions of radon concentration (Apte et al. 1999 and Price et al. 1996). Smith and Cowles (2007) present a different methodology based on multiprocesses spatial modeling in a bayesian perspective. Indeed it is often necessary to provide measures of the pollutant concentration which refer to sub-areas of the region under study in order to have a global measurement assessing the extent to which a whole area is prone to high radon concentration. This can be relevant when remediation actions are to be evaluated or promoted by administrative districts or when one wants to investigate the correlations existing between radon exposure and the spatial dynamic of a disease for which only measurements aggregated at the district levels are available. Moving towards area levels, almost nothing has been done so far to assess how a radon prone area should be defined and identified. Indeed this requires a realignment of the spatial data from the level at which they are collected (points) to the level at which they are necessary (areas). This is known as a "*change of support problem* (COSP)" in geostatistical terminology.
In this paper, we aim to propose a methodology based on geostatistical simulation in order to solve the COSP and identify radon prone areas. In order to do so, we refer to a survey implemented by the agency of environmental protection (ARPA) of Lombardy Region, in 2003 and focus on the province of Bergamo, an area of particularly high radon concentration. The proposed methodology can be extended to larger regions. A previous paper on radon prone areas in the Lombard plain was published in 2005 by Besana et al. Besana and colleagues, however, focused on a different approach based on geological arguments and used a different and quite smaller dataset (411 measurement points collected in seven villages located in the southern area of Bergamo and Brescia provinces).
The paper is organized as follows. In the next section, details of ARPA's monitoring campaign are reported. In section 3, a brief description of the geostatistical techniques employed are provided. Section 4 shows the approach proposed in this paper for the identification of radon prone areas. Some final remarks are presented in section 5.

2. The data

Lombardy is the most populated region of Italy and one of the most exposed to high radon concentration. The Lombardy Region has recently carried out a radon gas monitoring survey on its territory to localize radon prone areas as art. 10-ter, comma 2, D.L.vo 241/00 enacted. To plan the survey (de Bartolo et al. 2005), the region was divided into two different types according to the morphology as well as the presence of a substratum of rock. The area with hills and mountains was investigated more intensively compared to the plain since a higher variability in radon concentration distribution can be expected due to the geological and morphological characteristics. The territory subdivision was based on the standard grid of the technical regional cartography (8 km × 5 km): in the plain the grid consists of 4 units (16 km × 10 km) and in some mountain areas monitored in previous campaigns, a finer subdivision was chosen (8 km × 2.5 km or 4 km × 5 km) in order to increase the number of measurement points. From 5 to 10 measurement points were selected in each mesh. The grid used for the sampling survey is reported in figure 1.

To perform radon indoor concentration measurements, 3,646 measuring points located in 541 municipalities were selected. Locations were ground floor buildings with the necessary characteristics to ensure the tests are representative and comparable (rooms of volume less than 300 m^3, and with cellar or loose stone foundation).
Measurements were carried out using a CR 39 trace detector technique. The detectors were contained in closed plastic canisters, positioned in situ for one year and measured each semester. The average value is used in the present paper.
In this paper, we focus on the province of Bergamo, also depicted in figure 1, a region which is likely to be exposed to high radon indoor concentration. The survey monitored 593 dwellings located in 159 municipalities out of the 244 in the province, four points were discarded because radon concentration was missing. Thirty-two locations were geo-referenced again because of inconsistencies in the original coordinates.

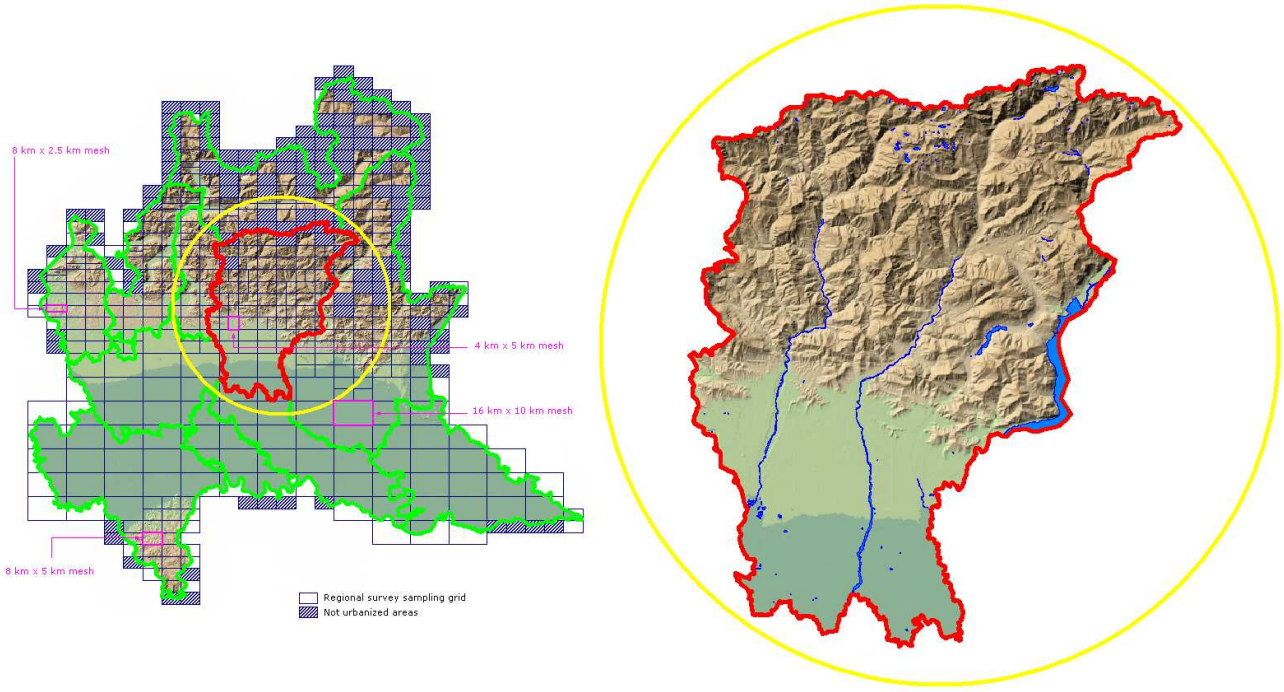

Fig. 1. Spatial sampling design of the radon indoor campaign implemented by the agency of environmental protection (ARPA) of Lombardy Region and the province of Bergamo

Only 13, collocated measurements were performed due to very close dwellings, 7 of which were shifted by one meter.
Table 1 shows the distribution of radon concentration in the province of Bergamo while figure 2 shows the cumulative distribution function and the histogram of radon concentration.

Table 1. Summary statistics of radon concentration in the province of Bergamo

Statistic	value
Mean	163.6 Bq/m^3
Median	101 Bq/m^3
Std Dev	177.8 Bq/m^3
Min	8.6 Bq/m^3
Max	1,182 Bq/m^3
% above 200 Bq/m3	24.1 %
% above 400 Bq/m3	8.1 %

Radon concentration in Bergamo appears to be quite higher than the national mean value of 70 Bq/m^3 (Bochicchio et al., 1996).

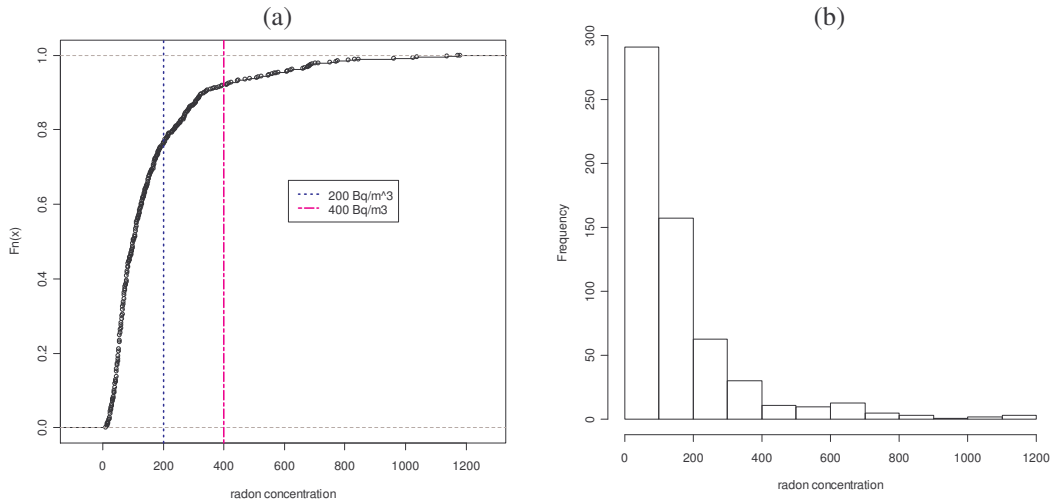

Fig. 2. Cumulative distribution function (a) and histogram (b) of radon indoor concentration in the province of Bergamo

Figure 3-a shows the 589 measurement points considered in this paper. Locations of dwellings with a radon concentration above 200Bq/m^3 and 400Bq/m^3 are displayed in figure 3-b and 3-c respectively.
These maps show that higher radon concentrations are found in the Northern part of the area under study.

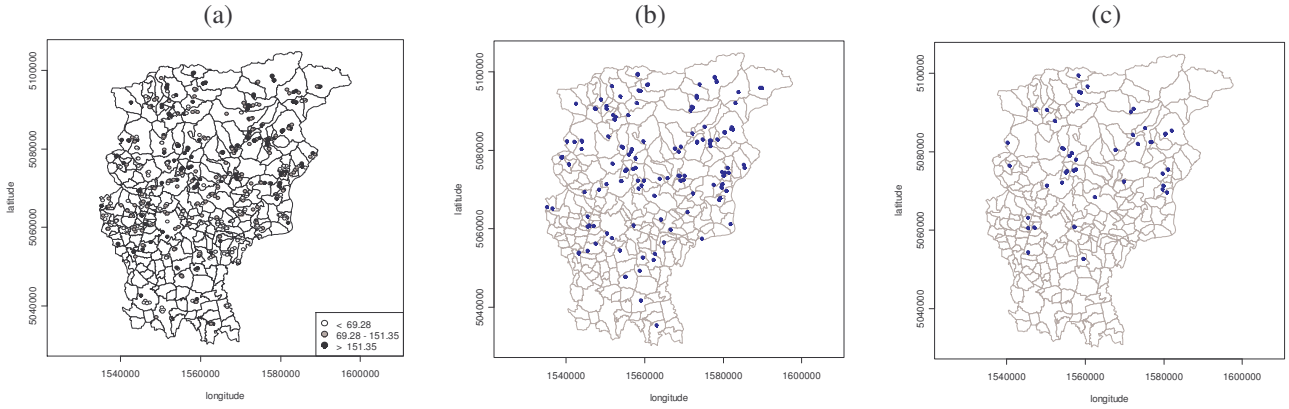

Fig. 3. Location of dwellings included in the survey (a), locations of dwellings with an indoor radon concentration above 200Bq/m^3 (b) and 400Bq/m^3 (c)

Figure 4 shows marginal trends of radon concentration with respect to latitude and longitude obtained via local polynomial regression fitting (LOESS) (Cleveland, 1979). Cartographic coordinates are in the Gauss-Boaga projection therefore the abscissa in the graph is measured in meters. A striking South-North trend is pointed out by the marginal analysis too, whereas the longitude component appears to be less clear.

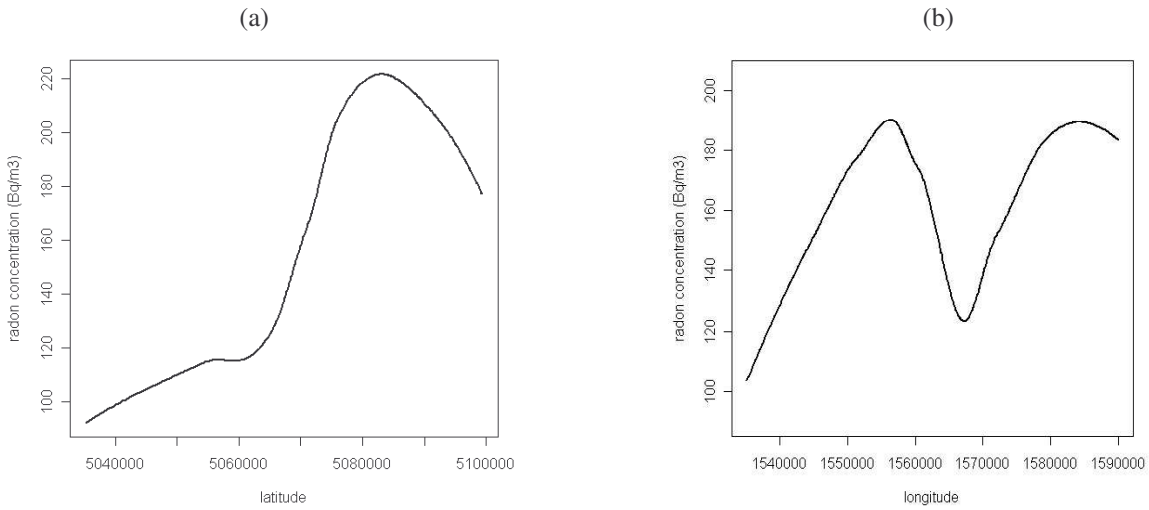

Fig. 4. South-North (a) and (b) East-West trend of radon indoor concentration

3. The Multi-Gaussian approach for facing change of support problems

3.1 The change of support problem

The different types of spatial data (point, line, area, surface), occurring naturally or as a result of the measurement process, often increase the need of integrating them in the statistical analysis. The Change of Support Problem (COSP) refers to the problem of valid inference with data of different supports.

The COSP includes problems associated with changes of spatial scales, aggregation and disaggregation of spatial data, inference with data on different spatial units (points, census tracts, districts etc.). The meaning of "support" is not simply the size associated with each data value, but also the geometrical shape and spatial orientation of the units associated with the measurements. Changing the support of a variable creates a new variable which is related to the original one, but may have different statistical and spatial properties. A particular case, firstly addressed by Arbia (1989), occurs when the spatial process of interest is inherently of one form but the data observed are of another, resulting from a "transformation" of the original process of interest. For example, sometimes the data are simply not available at the desired scale of interest because of confidentiality reasons or simply because they can only be measured at a given resolution: concentration of pollutants, for instance, can be measured only on a finite number of points while the process of interest may be conceived as a continuum. Indeed, a variety of incompatible spatial data are possible and a number of approaches can be adopted in order to sort them out. These issues are thoroughly addressed in a paper by Gotway and Young (2002).

An aspect sometimes neglected when one wants to identify radon prone areas is the operative consequences of such an identification. The final goal is, at least in principle, to help authorities in the planning or promoting of interventions. This requires providing information at a suitable level, such as an administrative district, which is small enough to grasp local needs and emergencies in terms of radiations and, at the same time, has the capability of implementing remediation actions. In the Italian case, local administrative units, which can promote or suggest remediation programs are, for instance, municipalities. From this perspective, the COSP consists of a realignment of the spatial radon measurements from the level at which they are collected, the measurement points, to the level at which they are necessary, the municipality, a sort of COSP called a point-to-block COSP.

In addition to the above mentioned aspect, there is another reason for which radon prone area identification based on block criteria appears to be appropriate. What is of interest in radon prone

area identification is the radon concentration *inside* houses. Punctual predictions based on kriging algorithms that are usually proposed in this context have the intrinsic characteristic that strictly rely on the spatial structure of the pollutant process. The human component, that is to say things such as building characteristics or human habits in terms of restricted ventilation or air conditioning can contribute a great deal to the spatial dynamic of this process. However, this component is mainly spatially unstructured and acts at a very local scale. The big nugget effect of the variogram, representing the spatial continuity of the radon concentration inside buildings, found by some authors (Dubois and Bossew, 2007) reflects this unstructured heterogeneity, even if it should be taken into consideration that living habits can also have, at least in part, a regionalized character as people may tend to live differently in various environments. Procedures used for COSP solving based on block averages, which are described in detail in the next section, integrate out the irregularities due to the anthropic component, since the variability of the pollutant concentration related to it simply compensates on average at the block level. Indeed, one of the main reasons of building a risk map is to point out critical areas in view of further actions and not assess individual cases or profile behavior. This clearly does not mean that dissecting potential determinants of high radon concentration is not a fundamental problem in environmental radioprotection, but simply that it is different, possibly complementary to the one addressed here (see Apte et al. 1999, Price et al. 1996, Smith and Field 2007).
The approach proposed in this paper for the municipality case can be adopted to different territorial aggregation.

3.2 **Determination of the block conditional distribution function: the Multi-Gaussian approach**

The point-to-block COSP has quite a natural solution in the block kriging paradigm (Cressie, 1992 Chilès and Delfiner, 1999) the aim being to predict the average of the random field $\{ Z(\mathbf{x}), \quad \mathbf{x} \in D \subset R^d \}$ on a block $B \subseteq D$:

$$Z(B) = \frac{1}{|B|}\int_B Z(\mathbf{x})\ d\mathbf{x}$$

where D is the region under study and B a block of size |B|, called the support, that one wants to predict.
The block kriging provides a linear predictor which minimises the mean square prediction error.
In radon prone area identification, as in many other problems of environmental remediation, the key quantity one wants to predict is:

$$1-F_B(z|\ Z(\mathbf{x}_1),...,Z(\mathbf{x}_n))=Pr(Z(B)>z|Z(\mathbf{x}_1),...,Z(\mathbf{x}_n)) \qquad (1)$$

where $F_B(z|\ Z(\mathbf{x}_1),...,Z(\mathbf{x}_n))$ is the conditional cumulative distribution function (ccdf) computed in z. Incidentally, the probability of exceeding a cut-off quantity perceived as dangerous has been suggested as a general measure of risk (Cramér, 1930) and has been largely adopted in environmental contexts (Arbia, 2002). Therefore (1) can be considered an appropriate measurement of the contamination risk in a given volume once evaluated on a given threshold. However, such a measure is not a linear function of the data. To face this problem, non linear geostatistics provides a range of tools such as the indicator approach (Goovaerts, 1997) or disjunctive and isofactorial kriging (Chilès and Delfiner, 1999).
In this paper, we use the Multi-Gaussian approach (Verly, 1983).
The multi-Gaussian approach assumes that $Z(\mathbf{x}_1),...,Z(\mathbf{x}_n)$ are n random variables normally jointly distributed possibly obtained from non Gaussian random variables $Y(\mathbf{x}_i)$ i=1,…, n by a monotonic

increasing transformation $Z(\mathbf{x}) = \psi(Y(\mathbf{x}))$. The block B is then discretized into N_B points $\mathbf{u}_1,\ldots,\mathbf{u}_{N_B}$ and Z(B) is approximated as

$$Z(B) \approx \frac{1}{N_B}\sum_{j=1}^{N_B} Z(\mathbf{u}_j). \tag{2}$$

Therefore one obtains:

$$\begin{aligned} F_B(z \mid Z(\mathbf{x}_1),...,Z(\mathbf{x}_n)) &\approx \Pr\left\{\frac{1}{N_B}\sum_{j=1}^{N_B} Z(\mathbf{u}_j) \le z \mid Z(\mathbf{x}_1),...,Z(\mathbf{x}_n)\right\} \\ &= \Pr\left\{\frac{1}{N_B}\sum_{j=1}^{N_B} \psi\big(Y(\mathbf{u}_j)\big) \le \psi(z) \mid Y(\mathbf{x}_1),...,Y(\mathbf{x}_n)\right\}. \end{aligned} \tag{3}$$

The problem is then to estimate the probability on the right-hand side of the previous formula which can be solved via simulation as discussed in the following section. However, before one proceeds with this analysis it is important to check whether the Multi-Gaussian assumption is consistent with the data at hand (Goovaerts, 1997). This will be discussed in greater detail in section 4.1.

3.3 **Gaussian sequential simulation for ccdf estimation**

As mentioned above, stochastic simulations can be used to approximate the ccdf in (3). The basic idea consists in simulating the random function $Z(\mathbf{x})$ on the N_B-point grid superimposed on block B and averaged the N_B generated values obtaining the block value $\hat{Z}(B) = \frac{1}{N_B}\sum_{j=1}^{N_B} Z(\mathbf{u}_j)$ which estimates Z(B). Repeating the procedure a large number of times, say L, produced L block values $\hat{Z}_1(B),\cdots,\hat{Z}_L(B)$. The desired probability in (1) can be estimated as the proportion of times that one gets a value $\hat{Z}_l(B)$ l=1,…,L, larger than z out of the L simulated values. If the multi-Gaussian hypothesis is not supported by the original data, simulations should be performed on transformed data and simulated values are mapped back to the original scale through the inverse transform before averaging them.

The simulation procedure should have two basic properties. Firstly the N values should be drawn jointly from a multi-normal distribution in order to preserve the spatial correlation structure and secondly they must honour the data in the sense that observed values should be reproduced by it. Procedures which have the latter property are called conditional. A range of algorithms have been proposed for conditional simulations. Some of them, such as simulations conditioned to the kriging, produces the desired vector of values in one go as a unique draw from a multi-normal law, others generate the values sequentially. In this paper, we use the sequential approach. The idea is quite simple in principle and relies on the factorization of the cumulative distribution function in a chain of products of ccdf: $F(Z_1,\cdots,Z_N) = F(Z_{\pi_1})F(Z_{\pi_2} \mid Z_{\pi_1})F(Z_{\pi_3} \mid Z_{\pi_2}Z_{\pi_1})\cdots F(Z_{\pi_N} \mid Z_{\pi_{N-1}}\cdots Z_{\pi_2}Z_{\pi_1}))$ where $(\pi_1,\ldots,\pi_N)$ is any permutation of the indexes 1,…,N.

In particular, in the Multi-Gaussian case, the joint and the conditional distribution functions are Normal and the index represents the location of the grid points where the random field is to be simulated. Indeed every conditional distribution is also conditioned to the data values $\mathbf{z}=(z(\mathbf{x}_1),\ldots, z(\mathbf{x}_n))$ observed at the n locations included in the experimental design. In other words, what one obtains is a vector of realizations from the conditional distribution

$$\Phi(Z_1,\cdots,Z_N \mid \mathbf{z}) = \Phi(Z_{\pi_1} \mid \mathbf{z})\Phi(Z_{\pi_2} \mid Z_{\pi_1}\mathbf{z})\Phi(Z_{\pi_3} \mid Z_{\pi_2}Z_{\pi_1}\mathbf{z})\cdots\Phi(Z_{\pi_N} \mid Z_{\pi_{N-1}}\cdots Z_{\pi_2}Z_{\pi_1}\mathbf{z}))$$

where $\Phi(Z)$ stands for the generic Gaussian cdf. The moments of the normal distribution can be derived from the simple kriging applied to the set of data obtained by pooling together the observed

data and the simulated values up to the point reached along the path chosen to visit the nodes of the grid. In other words, when one simulates Z_{π_l}, this is obtained by making a random draw from the univariate (conditional) normal distribution whose mean is the simple kriging prediction at site π_l and whose variance is the simple kriging variance, the kriging being applied to the set of data $Z(\mathbf{x}_1),\cdots.Z(\mathbf{x}_l),Z(\mathbf{u}_{\pi_1}),\cdots,Z(\mathbf{u}_{\pi_{l-1}})$. The simulation process is not affected by the order in which the grid points are scanned. As suggested by Goovaerts (1997, pg. 381), non-stationary behaviour in the mean can be accomplished by other types of kriging algorithms, for instance, local ordinary or universal kriging.

4 Radon prone area identification

In this section, we adopt the methodology discussed in section 3 to identify radon prone areas in the province of Bergamo; in particular we aim to identify radon prone areas in terms of municipalities.

4.1 Preliminary spatial analysis

In order to guarantee, at least marginally, the normality assumption, we transform the data to the *normal score* scale. The normal score is quantile-to-quantile monotone non-decreasing transformation, sometimes called Gaussian transformation or anamorphosis, that turns the original data into a set of standard gaussian values (Goovaerts, 1997, pp 266-271). A new algorithm for normal score transform has been recently introduced by Emery and Ortiz (2005).
Figure 5-a shows the function $\psi(y)$ used in our case to transform radon indoor concentration to the normal score scale.
In principle, marginal gaussianity of the data is not a sufficient condition to guarantee the multi-Gaussian hypothesis underlying the kriging analysis and one should check whether such a hypothesis is consistent with the data. In practice, validating a multi-Gaussian hypothesis is not straightforward because the inference of multiple-point statistics is made difficult by data sparseness and usually only a bivariate normality is checked.
Usual practice is to apply the multi-Gaussian model to the normal scores if the two-point normality is not invalidated by data evidence, deriving the ccdf at any un-sampled location **u** and, as it is of interest in the present case, deriving the ccdf of the block after the calculation of (2) in each iteration.
A few methods have been proposed to check two-point normality (Emery 2005) mostly based on graphical tools. Here, we consider an approach recently introduced by Emery (2005, 2006). Under the assumption that the bivariate distribution of (Y(**x**) ,Y(**x**+**h**)) is Gaussian, one has:

$$\gamma_\omega(\mathbf{h}) = \frac{2^{\omega-1}}{\pi}\Gamma\left(\frac{\omega+1}{2}\right)[\gamma(\mathbf{h})]^{\frac{\omega}{2}}$$

where $\omega\in\,]0,2]$, $\Gamma(x)$ is the Euler gamma function and $\gamma_\omega(\mathbf{h})$ is the semi-variogram of order ω defined as $\gamma_\omega(\mathbf{h}) = \frac{1}{2}E\left\{\left|Y(\mathbf{x}+\mathbf{h}) - Y(\mathbf{x})\right|^\omega\right\}$, $\omega\in\,]0,2]$. When ω=2 one gets the usual semi-variogram. Other popular members of this family are the madogram (ω=1) and the rodogram (ω=0.5).

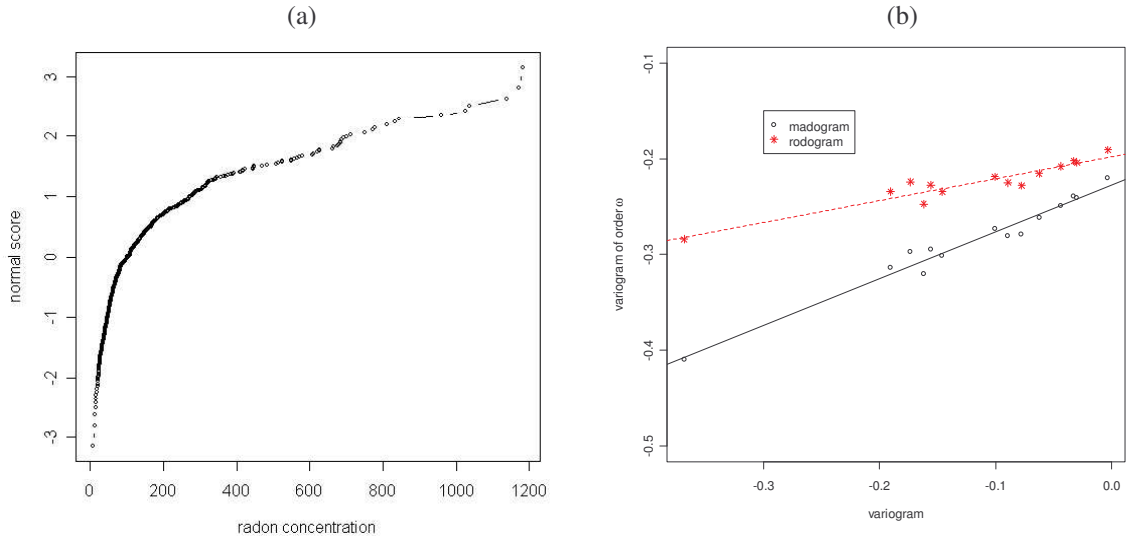

Fig. 5. Normal score transformation for radon indoor concentration (a); graphical check for bi-normality based on a semi-variogram versus semi-variogram of order ω plot (b)

Therefore, plotting on a log-log scale the semi-variogram values towards the semi-variogram of order ω values for a given ω, should produce a line with slope equal to ω/2. A graphical check for two-point normality can be obtained by computing the empirical semi-variogram and the empirical semi-variogram of order ω for a range of lags **h** and plot them on a log-log scale graph.
This is done for our data under the hypothesis of isotropy (discussed later in this section). The results are reported in figure 5-b for ω=1 and ω=0.5. The two lines in the figure are OLS fits. The two slopes are equal to 0.49 and to 0.23 for the madogram and the rodogram respectively suggesting that the hypothesis of two-point normality seems well supported.
We then check the hypothesis of isotropy by drawing the variograms in the four directions associated to an angle of 0°, 45°, 90° and 135° calculated anticlockwise from the azimuthal direction.
The directional semi-variograms on the normal score scale are reported in figure 6-a along with the omni-directional empirical variogram. Non significant anisotropy seems to be present on the data, hence the omni-directional variogram is considered herein.
In order to perform the kriging interpolation used in sequential simulation, we estimate an exponential semi-variogram via weighted least squares (Cressie, 1992). Given the radon nature, it does not seem reasonable to assume that the correlation has a long effect in space.
We prefer to estimate the variogram using pairs of locations less than 19 km apart aiming at better correlation estimates between closer sites. This value was suggested by cross validating the variogram over a large grid of possible maximum distances. Figure 6-b shows the estimated variogram suggesting a satisfactory fit of the proposed model.
In order to produce a map of predicted radon concentrations, a point grid for prediction is preliminarily made. As the same grid will be used in the following section for identifying radon prone municipalities, this grid is defined to account for differences in the municipalities size and shape. Points are regularly spaced inside each municipality. In the case of municipalities consisting of separated areas, coded via disjoint polygons, the grid is built separately for each of them. The grid consists of 2,853 points, the minimum number of points for each municipality being 4 and the maximum 17. (See figure 7-a).
In order to account for the spatial trend in the data, we used a local ordinary kriging. The local kriging allows one to account for non-stationary local means reducing, at the same time, the

computational burden due to solving the kriging system as only the points included in the kriging neighbour are actually used for prediction.

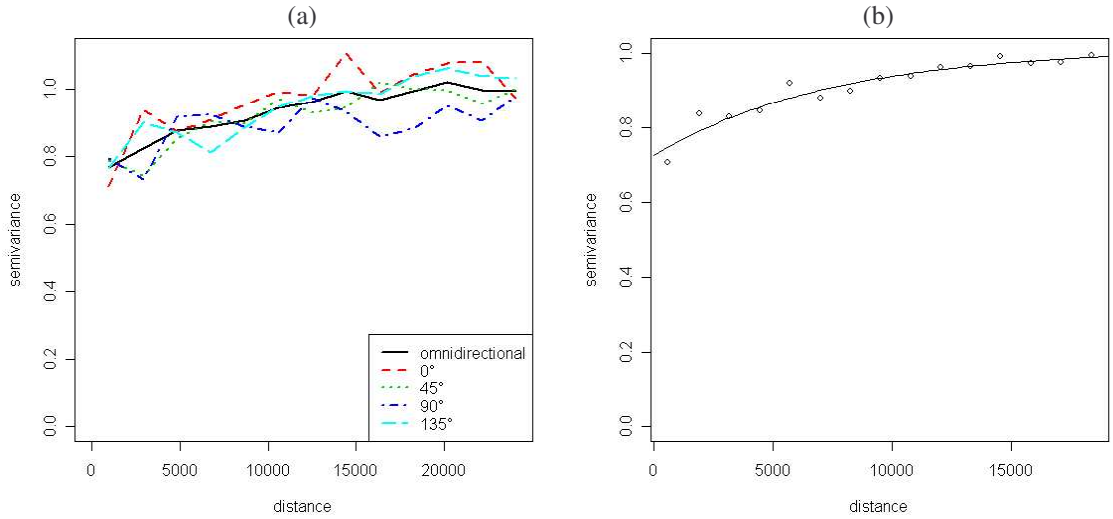

Fig. 6. Directional and omni-directional variograms for radon concentration (a). WLS estimate of the exponential variogram (b)

The latter point is quite relevant for the analysis described in the next section as a number of kriging systems must to be solved in the Monte Carlo procedure. The kriging radius was set to 7500 m. This is the minimum distance which guarantees at least two observation points in the kriging neighbour of each grid point used for the interpolation.

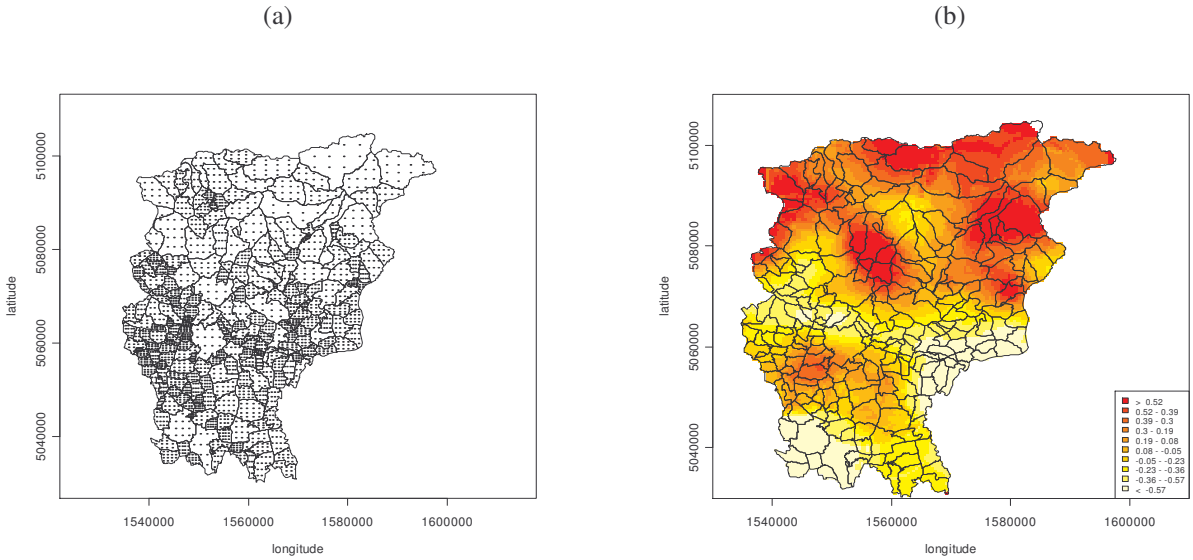

Fig. 7. The grid points for prediction (a) and kriging map of radon concentration (b)

4.2 Radon prone area identification

The local ordinary kriging is used in the multi-gaussian sequential simulation paradigm to obtain a vector of simulated radon concentrations one for each point of the grid.
The simulated data are carefully checked for consistency with the required assumptions. Although a complete diagnostic cannot be reported here, figure 8 gives an idea of the quality of the simulation performed by showing the distributions of eight randomly selected points of the grid which appear to be close to normality. The characteristics of the simulated data match quite well with those of the observed data. For instance, the average concentration obtained from the simulated data is equal to 149 Bq/m^3 while the mean obtained from the declustered data[1] is 157 Bq/m^3 for the whole region. The simulated percentages of concentration above the thresholds of 200 Bq/m^3 and 400 Bq/m^3 are 21.1% and 6.6% whereas the correspondent values obtained from the data after declustering are 24.1% and 6.9% respectively. Figure 9 shows the empirical variograms for five randomly chosen simulated maps which again look quite similar to the empirical variogram observed on the data suggesting that the correlation structure is preserved by the simulation process.
For each simulated set the average concentration for municipality B is then estimated as $\hat{Z}(B) = \frac{1}{N_B}\sum_{j=1}^{N_B} Z(\mathbf{u}_j)$.

Let $\hat{Z}_q^B$ be the q-th quantile obtained by the simulated block values $\hat{Z}_1(B),\cdots,\hat{Z}_L(B)$. An area B could be defined as radon prone if a high order quantile is not below a given high threshold. In particular we refer to the quantile of order 0.9. The 200 Bq/m^3 and 400 Bq/m^3 thresholds we refer to in this paper, are suggested by the 90/143/Euratom recommendation as reference values above which mitigation measures should be taken in new and existing dwellings respectively.
It can be noted that taking q=0.9 turns out to be equivalent to what one of the Italian agencies for environment protection adopted (Regione Veneto, ARPA, 2000). This agency, in fact, proposed to identify an area as radon prone if the probability in (1), calculated in correspondence with the values suggested by the Euratom recommendation, is greater than 0.1.
In order to account for the extra-variability in the estimate $\hat{Z}_{0.9}^B$ due to the simulation procedure, we calculate a right unilateral 95% confidence interval by bootstrapping a 100 times the 1000 independent simulations.

[1] A cell declustering technique is adopted (Goovaerts, 1997, pg. 81).

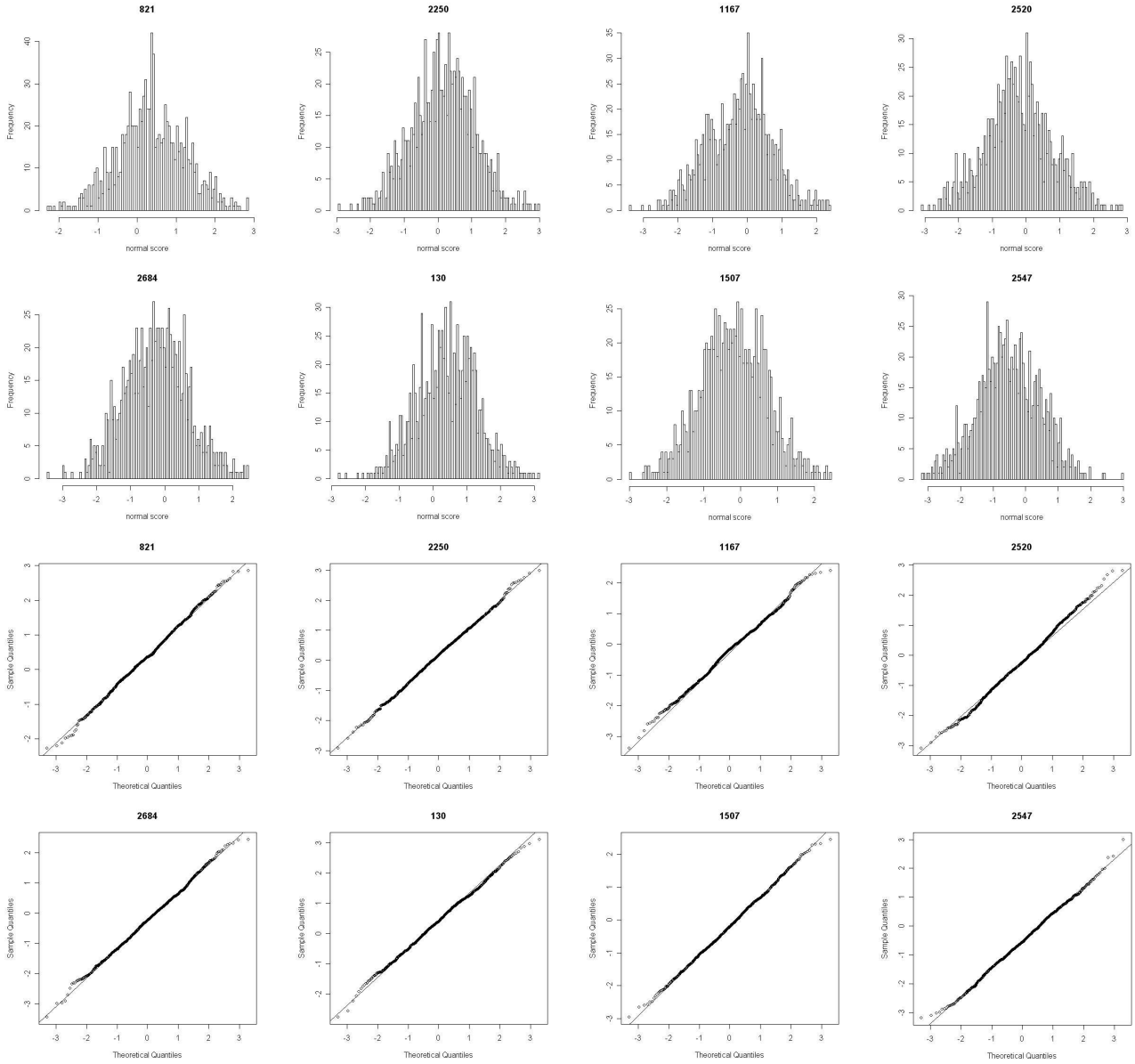

Fig. 8. Distribution of simulated values for eight randomly chosen points in the map. Numbers in the title of each graph represent the grid point identification number

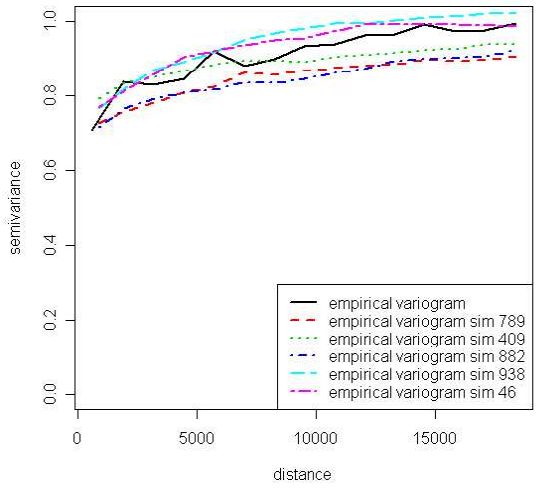

Fig. 9. The empirical variograms for five randomly chosen simulated maps. Numbers in the legend represent the iteration of the Monte Carlo simulation

Eventually a municipality is classified as *radon prone* at either 400 Bq/m^3 or 200 Bq/m^3 if the corresponding threshold is below the 95% CI and as not *radon prone* otherwise. Results representing a possible atlas of areas at risk of high radon concentration in the province of Bergamo are reported in figure 10. For such areas, further investigations should be undertaken in order to plan possible remediation campaigns.

Finally, we observed that the proposed procedure could be modified to accomplish those cases where only a small number of simulations can be performed. This can happen, for instance, if one wants to build a map based on a very fine grid. In this case, the computational burden, due to the simulations, can increase a great deal making it unfeasible even having a moderate number of simulated maps. In this case, in order to obtain a confidence interval for the quantile, one could bootstrap from a smoothed version of the empirical ccdf (Hall et al. 1989) of each municipality.

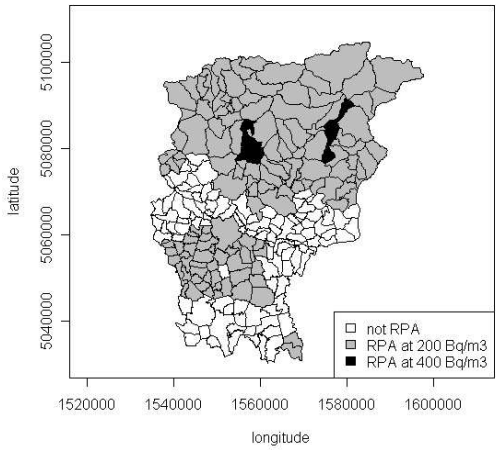

Fig. 10. Radon prone areas (RPA) in the province of Bergamo

4.3 Different kriging approach for Radon prone area identification

The presence of a non-stationary mean can be accounted for in other ways than the local ordinary kriging as universal kriging or kriging on detrendised data. The latter can be easily employed by practitioners as the spatial prediction is based on simple kriging (the trend being removed externally), a technique often implemented in many commercial software for spatial data, as geographical information systems, which have some tools for geostatistical data analysis, while local analysis is not always available.
In this section we compare the results obtained by this approach with those obtained in section 4.2.
The data are firstly detrendised using an additive model, AM, (Hastie, Tibshirani, 1990) on data transformed to the normal score scale. In other words, a regression model of the form $Z(\mathbf{x})=\alpha + f_1(x_1) + f_2(x_2) + \varepsilon(\mathbf{x})$ is estimated where $\mathbf{x} = (x_1,x_2)$ are the cartographic coordinates of the measurements points, $f_i(x_i)$ i=1,2 are LOESS transformations of latitude and longitude and $\varepsilon(\mathbf{x})$ a Gaussian random noise. This model has been compared with a range of large scale models (complete polynomial linear models of order one and two, log-normal and gamma generalised additive model) via cross validation and proved to be the best performing.
Residuals are then computed as $e(x) = Z(x) - \hat{\alpha} - \hat{f}_1(x_1) - \hat{f}_2(x_2)$, where the *hat* indicates estimates of the model counterpart. LOESS components on latitude and longitude strictly resembled the graphs in figure 4, therefore, we decided not to display them again.
Normality of residuals is deemed "good" after testing via the methodologies described in section 4.1. A simple kriging algorithm is then applied for predictions of the residual process assuming a 0 mean for the process. Once the process is simulated on the grid displayed in figure 7-a the estimated trend is added back and predictions are mapped on the original scale. The transformed values are averaged for each municipality. The procedure is repeated one thousand times. In this case too we built a unilateral 95% CI by bootstrapping the 1000 independent simulations a 100 times to account for the extra-variability due to simulations. A municipality is classified as *radon prone* at either 400 Bq/m^3 or 200 Bq/m^3 if the corresponding threshold is below the 95% CI and as not *radon prone* otherwise. Results are reported in figure 11.
As it can be seen, by comparing figure 11 and figure 10, we obtain very similar results. The areas classified differently are 36 out of the 244 municipalities of the Bergamo province. The majority of them (30) are grey areas in the maps based on local kriging analysis, mostly located in the centre of the province, which become white in figure 11. This is due to the trend effect which tends to over-smooth the map being less sensitive to local behaviours.

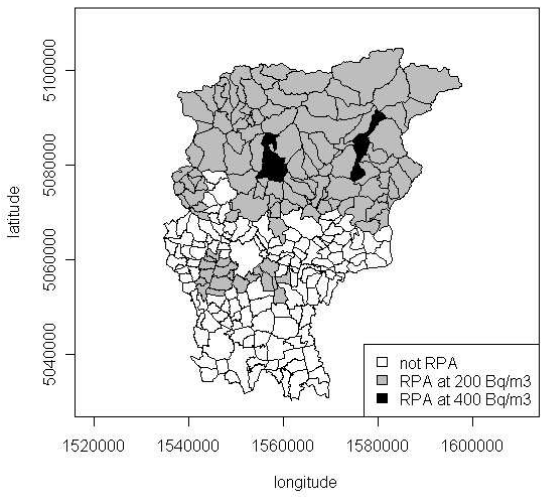

Fig. 11. Radon prone areas in the province of Bergamo using a simple kriging and GAM trend

5 Conclusions

In order to assess the exposure of people to indoor radon and to identify radon prone areas, monitoring surveys have been recently implemented in a number of western European countries. Such campaigns are designed to provide concentration values in each single monitored dwelling whereas it is often necessary to have measurements of the pollutant concentration which refer to sub-areas of the region under study. This requires a realignment of the spatial data from the level at which they are collected (points) to the level at which they are necessary (areas).
In this paper, we considered a methodology based on geostatistical simulations in a Multi-Gaussian context in order to solve this problem and to provide a method to identify radon prone areas.
We applied the proposed methodology to the province of Bergamo, one of the most exposed Italian provinces to high radon concentration in order to identify radon prone areas in terms of municipalities.
The procedure proposed in this paper seems to be quite effective in identifying radon prone areas. We found that municipalities located close to the Alps have a higher probability of being radon prone. This is consistent with findings from other studies, that is, sites located on mountains tend to have higher radon concentration mostly due to geological reasons. It is also found that area classification is quite insensitive to different kriging algorithms used for prediction inside the Monte Carlo simulations once the prediction procedure is able to grasp non-stationary behaviors in the mean.
Finally, two other points were observed. Firstly, the approach was implemented assuming that radon prone areas are identified in terms of municipalities mostly because such administrative units are able to promote remediation campaigns or to suggest further investigation in the highlighted areas. However, this can be adopted for other administrative units too. Secondly, although computationally intensive, the proposed approach has been implemented in a standard machine and may therefore be successfully applied to larger regions too and eventually adopted to define RPA on a national scale.

Acknowledgments Giorgio Somà was supported by a grant assigned to the project RAPRAR “Radon Prone Areas in Lombardia” funded by Sovvenzione Globale INGENIO on behalf of: Fondo Sociale Europeo, Ministero del Lavoro e della Previdenza Sociale and Regione Lombardia. We warmly thank Denise Kilmartin for editing the paper.

References

Apte MG, Price PN, Nero AV, Revzan KL (1999) Predicting New Hampshire indoor radon concentrations from geologic information and other covariates Environmental Geology 37, 181-194

Arbia G (1989) Statistical effect of data transformations: a proposed general framework, in the accuracy of spatial data bases, eds. M. Goodchild and S. Gopal- Taylor and Francis London, 249–259

Arbia G, Lafratta G (2002) Exploring nonlinear spatial dependence in the tails, Geographical Analysis, 37: 423-437

Bochicchio F, Campos Venuti G, Nuccetelli C, Piermattei S, Risica S, Tommasino L, Torri G. (1996) Results of the representative Italian National Survey on Radon Indoors – Health Physics 17(5): 741-748

Chilès D, Delfiner P (1999) Geostatistics Modeling Spatial Uncertainty, Wiley, New York

Cleveland, WS (1979) Robust locally weighted regression and smoothing scatterplots”, Journal of the American Statistical Association, 74: 829-836

Cramér CH (1930) On the Mathematical Theory of Risk, Forsakringsaktiebolaget Skandias Festskrift, Stockholm: Centraltryckeriet

Cressie N (1993), Statistics for spatial data, Wiley, New York

de Bartolo D, Alberici A, Gallini R, Maggioni T, Arrigoni S, Cazzaniga P, Cugini A, Olivieri F, Romanelli M, Gallinari G (2005) Piano di monitoraggio per l'individuazione della radon prone areas nella regione Lombardia. Atti del convegno AIRP, Catania 15-17 settembre 2005

Dubois G, Bossew P (2006) A European Atlas of Natural Radiations including harmonized radon maps of the European Union. What do we have, what do we know, quo vadimus? European Commission – DG Joint Research Center, Institute for Environment and Sustainability, Italy

Dubois G, Bossew P, Friedmann H (2007) A geostatistical autopsy of the Austrian indoor radon survey (1992–2002). Science of the Total Environment. In press

Emery X (2005) Variograms of Order ω: A Tool to Validate a Bivariate Distribution Model, Mathematical Geology, Vol. 37

Emery X (2006) Ordinary multigaussian kriging for mapping conditional probabilities of soil properties Geoderma, 132: 75-88

Emery X, Ortiz JM (2005) Histogram and variogram inference in the multigaussian model, Stochastic Environment Res. Risk Assess, Vol. 19: 48-58

Goovaerts P (1997) Geostatistics for Natural Resources Evaluation, Oxford University Press, New York

Gotway CA, Young LJ (2002) Combining incompatible spatial data, Journal of the American Statistical Association, 97, 458: 632-648

Hall P, DiCiccio TJ, Romano JP (1989) On smoothing and the bootstrap. Annals of Statistics 17: 692-704

Hastie TJ, Tibshirani RJ (1990) Generalized additive models, Chapman & Hall

Franco-Marina F, Villalba-Caloca J, Segovia N, Tavera L. (2003) Spatial indoor radon distribution in Mexico City. Science of Total Environment. 317 (1–3): 91–103

Price PN, Nero AV, Gelman A (1996) Bayesian prediction of mean indoor radon concentrations for Minnesota counties Health Physics. 71: 922-936

Regione Veneto, ARPAV (2000) Indagine regionale per l'individuazione delle aree ad alto potenziale di radon nel territorio veneto.

Sesana L, Polla G, Facchini U, de Capitani L (2005) Radon-prone areas in the Lombard plain. Journal of Environmental Radioactivity 82: 51-62

Smith BJ, Cowles MK (2007) Correlating point-referenced radon and areal uranium data arising from a common spatial process. JRSS Series C, 56: 313-326

Smith BJ, Field RW (2007) Effect of housing factor and suficial uranium on the spatial prediction of residential radon in Iowa. Environ-metrics, 18: 481-497

Verly G (1983) The Multi-Gaussian approach and its applications to the estimation of local reserves. Journal of the International Association for Mathematical Geology, 15: 259-286

Zhu H-C, Charlet JM, Poffijn A. (2001) Radon risk mapping in southern Belgium: an application of geostatistical and GIS techniques. Science of the Total Environment, 272(1–3): 203–10