Elsevier Editorial System(tm) for Journal of Environmental Radioactivity
Manuscript Draft

Manuscript Number:

Title: Geographically weighted regression and geostatistical techniques to construct the Geogenic Radon Potential map of the Lazio region: a methodological proposal for the European Atlas of Natural Radiation

Article Type: SI: Geogenic Radon

Keywords: soilgas, radon, GWR, geostatistics, radon potential map

Corresponding Author: Dr. Giancarlo Ciotoli, PhD Earth Sciences

Corresponding Author's Institution: National Research Council

First Author: Giancarlo Ciotoli, PhD Earth Sciences

Order of Authors: Giancarlo Ciotoli, PhD Earth Sciences; Paola Tuccimei; Michele Soligo; Antonio Pasculli; Stanley Eugene Beaubien; Mario Voltaggio

Abstract: In many countries, radon programmes are carried out to identify radon prone areas, where people may be exposed to high indoor radon values. Some attempts have been conducted to detailed map these areas based on the relationships between indoor radon values and some geological and environmental factors (i.e., lithology, permeability, etc.). These data are used to optimise the radon hazard maps and to assess the potential radon risk in building zones at the scale of the regions and/or municipalities. In this work, Geographical Weighted Regression and geostatistics are used to construct the geogenic radon potential (GRP) of the Latium Region, assuming that the radon risk only depends on the geological and environmental characteristics of the study area. A wide geodatabase has been organised including about 8000 samples of soil gas and indoor radon, as well as other geological (i.e., rock permeability, faults, topography) and geochemical (i.e., radium and uranium content of rocks) proxy variables strictly correlated with the radon production in the shallow environment. All these data have been elaborated within a GIS by using geospatial analysis and geostatistics to produce base thematic maps in a 1000x1000 m grid format. Global Ordinary Least Squared regression and local Geographical Weighted Regression have been applied and compared assuming that the relationships between radon activities and the environmental variables are not spatially stationary, but vary locally according to the GRP. The spatial regression model has been elaborated considering soil gas radon concentrations as the response variable and the proxy variables as predictors by using training dataset. Then a validation procedure was used to predict soil gas radon values at test dataset. The predicted values were then elaborated by kriging algorithm to obtain the GRP map of the Lazio region. The map highlights areas characterised by high GRP mainly linked to radionuclide content of rocks in correspondence of the volcanic areas (central-northern sector of Lazio region), and high GRP mainly linked to faulted and fractured carbonate rocks (central-southern and eastern sectors of the Lazio region). This typical local variability of autocorrelated phenomena can be taken into account only by using local methods for spatial data

analysis. The constructed GRP map can be a useful tool to implement radon policies at both national and local level, providing the priority to a better knowledge of the territory for land use and planning purposes.

Geographically weighted regression and geostatistical techniques to construct the Geogenic Radon Potential map of the Lazio region: a methodological proposal for the European Atlas of Natural Radiation

Ciotoli G.[1,2]✉, Voltaggio M.[1],Tuccimei P.[3], Soligo M.[3], Pasculli A. [4], Beaubien S.E. [5],

1 Istituto di Geologia Ambientale e Geoingegneria, Consiglio Nazionale delle Ricerche – IGAG-CNR, Area della Ricerca di Roma1 Rome, Italy - giancarlo.ciotoli@igag.cnr.it
2 Istituto Nazionale di Geofisica e Vulcanologia, Sezione Roma 2, Rome, Italy
3 Università degli Studi "Roma Tre", Dipartimento di Scienze - Rome, Italy
4 Dipartimento di Ingegneria e Geologia, Università di Chieti, Italy
5 Dipartimento di Scienze della Terra, Sapienza Università di Roma, Rome, Italy

Highlights

- soil gas radon sampling
- Identification of homogeneous geological units of the Lazio region
- Global and local regression models by using soil gas radon and geological and geochemical proxy variables
- Geogenic Radon Potential map of the Lazio region obtained by using local Geographycally Weighted Regression and geostatistical analysis.

Abstract

In many countries, radon programmes are carried out to identify radon prone areas, where people may be exposed to high indoor radon values. Some attempts have been conducted to detailed map these areas based on the relationships between indoor radon values and some geological and environmental factors (i.e., lithology, permeability, etc.). These data are used to optimise the radon hazard maps and to assess the potential radon risk in building zones at the scale of the regions and/or municipalities. In this work, Geographical Weighted Regression and geostatistics are used to construct the geogenic radon potential (GRP) of the Latium Region, assuming that the radon risk only depends on the geological and environmental characteristics of the study area. A wide geodatabase has been organised including about 8000 samples of soil gas and indoor radon, as well as other geological (i.e., rock permeability, faults, topography) and geochemical (i.e., radium and uranium content of rocks) proxy variables strictly correlated with the radon production in the shallow environment. All these data have been elaborated within a GIS by using geospatial analysis and geostatistics to produce base thematic maps in a 1000x1000 m grid format. Global Ordinary Least Squared regression and local Geographical Weighted Regression have been applied and compared assuming that the relationships between radon activities and the environmental variables are not spatially stationary, but vary locally according to the GRP. The spatial regression model has been elaborated considering soil gas radon concentrations as the response variable and the proxy variables as predictors by using training dataset. Then a validation procedure was used to predict soil gas radon values at test dataset. The predicted values were then elaborated by kriging algorithm to obtain the GRP map of the Lazio region. The map highlights areas characterised by high GRP mainly linked to radionuclide content of rocks in correspondence of the volcanic areas (central-northern sector of Lazio region), and high GRP mainly linked to faulted and fractured carbonate rocks (central-southern and eastern sectors of the Lazio region). This typical local variability of autocorrelated phenomena can be taken into account only by using local methods for spatial data analysis. The constructed GRP map can be a useful tool to implement radon policies at both national and local level, providing the priority to a better knowledge of the territory for land use and planning purposes.

Key words: *soilgas, radon, GWR, geostatistics, radon potential map*

1. Introduction

Indoor Air Quality (IAQ) in public, and residential buildings has become a highly important environmental issue, especially in large, densely populated urban areas. Furthermore, the introduction of new building criteria, such as the use of new techniques to improve thermal insulation and therefore energy savings, compound this problem because they tend to reduce air exchange. On average people spend about 80-90% of their time in confined spaces (i.e. homes, workplaces, schools, etc.) and this percentage rises for children, elderly, patients, etc. The monitoring of the healthiness of such environments is fundamental to reduce the exposure of the population to pollutants.

Natural radioactivity is the main source of human exposure to ionizing radiation. The inhalation of radon (^{222}Rn) and its progeny contributes 50% of the annual dose from ionising radiation. Whereas radon concentrations are extremely low in outdoor air, concentrations can become dangerously high indoors due to its accumulation in closed spaces. Sources for indoor radon include seepage from the surrounding soil and rock geology (so called "geogenic" radon), from the building materials used, or degassed from tap water having a groundwater origin. Accumulation, instead, is a function of ventilation within the building.

Radon is a gaseous trace element, chemically inert and ubiquitous in soil and groundwater. Radon is produced via the decay chain of primordial radionuclides ^{238}U, ^{232}Th and ^{235}U. The most abundant isotope is ^{222}Rn (from the decay chain of ^{238}U), has a half-life of 3.82 days, and decays itself to stable lead ^{206}Pb through an intermediate decay chain. Radon gas is colourless, tasteless, odourless and is not detected by the human senses even at high concentrations. Being a noble gas, radon is not very reactive, and is generally eliminated from the body. However, the real health hazard are its daughters (i.e., Pb, Po, Bi) which are also radioactive; a portion of the radon breathed-in decays to these daughters, which bind to dust particles and irradiate lung and bronchial tissues as they decay.

Radon was classified as a human carcinogen in 1988 by the IARC (International Agency for Research on Cancer). More recently, the health effects linked to indoor radon exposure have been considered in the EC Directive 2013/59/EURATOM of 5/12/2013. It stated that recent epidemiological findings from residential studies demonstrate a statistically significant increase of lung cancer risk from prolonged exposure to indoor radon at levels of the order of 100 Bq m^{-3}. It is estimated that about 9 to 15% of the approximately 14,000 annual cases of lung cancer in Europe can be attributed to radon and its progeny (Darby et al., 2005; Krewski et al., 2005; Charles, 2001; Kreienbrock et al., 2001; IARC, 1988). For this reason, indoor radon in public and residential buildings constitutes one of the main environmental problem in urban areas (UNSCEAR, 2000; European Commission, 1990, 2013).

In general, it is accepted that areal variation of radon levels in houses primarily depends on the geological features of the investigated areas, because the bedrock and soil type constitute the main

Rn sources, and because soil permeability controls Rn transport toward surface (Bossew, 2015, 2014, 2013; Ciotoli et al., 2007; Shi et al. 2006; Friedmann, 2005; Kemski et al. 2005, 2001; Killip 2005; Miles and Appleton 2005; Apte et al. 1999; Gates et al., 1992).

Over the last few decades, various national indoor radon surveys have been performed in several European countries. These surveys, whose results are collected within the European Atlas of Natural Radiation by the European Joint Research Centre (Tollefsen et al., 2014), often display their results as contoured "radon maps", and are considered as a preliminary action directly related to risk assessment. However, considering the lack of spatial correlation between houses having different structural characteristics and owner habits related to ventilation, this approach could be misleading. In some studies, the building-related variability (e.g., floor level, building materials, building type, presence of a basement, etc.) was recorded and filtered out to obtain, as far as possible, "true" radon indoor values. However, it is difficult to justify the interpolation of such data to predict the indoor radon levels of yet un-measured houses, or to cover unpopulated areas and draw conclusions about how to build new houses.

Another approach consists in the assessment of the Geogenic Radon Potential (GRP) of a region, which is a quantity directly related to the local geology. A properly defined GRP based on a spatially continuous parameter might provide a reasonable guide for identifying radon-prone areas, particularly when the number and/or the quality of available indoor radon data is inadequate. The geological information by itself (e.g. lithological types, U and Ra content, soil gas radon and permeability), may be sufficient to infer the radon potential. However, to date there is no generally accepted method of radon risk mapping.

The modelling approach proposed in this work uses different appropriate geospatial techniques, such as Geographical Weighted Regression (GWR) and geostatistics (kriging) to account for spatial autocorrelation and to produce a map of the Geogenic Radon Potential (GRP) of the Latium region. Geological data and soil gas data are provided by the Soil Protection and Remediation Department of Regione Lazio and by the Fluid Chemistry Laboratory of the Earth Science Department, Rome University Sapienza, respectively. This wide database was elaborated in the GIS environment by using ArcGIS 10.2 (Copyright © 1999-2013 Esri Inc.). All produced maps are constructed according to a grid format with 1000 x 1000 m unit cell resulted by using vector to raster transformation, reclassification and interpolation of primary geological, geomorphological and geochemical data.

2. Radon in the shallow environment

The distribution of radon in soil gas and, consequently the indoor activities, is strictly related to the geological characteristics of the studied territory (Kemski et al., 2009; Barnet et al., 2008; Ciotoli et al., 2007). Three main factors are known which predispose houses to elevated indoor radon levels. First, the regional geochemical and geological characteristics of the soil / rock will establish the in situ conditions. For example, uranium (^{238}U, ^{235}Th) and radium (^{226}Ra) content will control the amount of radon generated. Uranium and radium occur in all rocks at concentrations from 0.5-5 mg/kg, depending on the rock type. Igneous and metamorphic rocks (granites, acid lavas, tuffs, etc.) typically have very high uranium/radium contents and sedimentary rocks generally have lower contents (but high in some types like organic rich rocks, phosphates, reworked igneous or magmatic clastic rocks, etc.) (Drolet et al., 2013). Second, environmental conditions will control the rate of movement of soil radon toward the surface and into buildings. The escape of radon atoms at the grain scale is controlled by porosity, soil moisture and grain-size, whereas the migration toward the shallow environment is controlled by the large scale geological features including rock thickness, permeability, fractures and karst (Castelluccio et al., 2012; Nazaroff, 1992; Etiope et al., 2002; Nazaroff et al., 1988; Tanner, 1980). Also meteorological parameters like wind, barometric pressure, relative humidity and rainfall can affect radon exhalation from soil to the atmosphere (Galli et al., 2015; Szabo et al., 2013; Vasilyev et al., 2013; Zafrir et al., 2012; Baykut et al., 2010; Crockett, et al., 2010; Fujiyoshi et al., 2006; Al-Shereideh et al., 2006; Winkler et al., 2001). Both these phenomena affect the GRP in terms of source and migration mechanisms. The third factor is the building characteristics that will influence radon entry into buildings (e.g., discontinuities or fractures in the foundation that can provide gas entry pathways) while particular building materials can also be a source of radon production inside the building itself. Therefore, geology, quantified by a categorical classification system and or according to proxy variables (e.g., U/Ra content, permeability, etc.), can provide predictors of the GRP which make the problem of the estimate across the geological boundaries reduced and more realistic (e.g. Tondeur et al., 2014; Gruber et al., 2013; Appleton and Miles, 2010; Cinelli et al., 2010; Kemski et al. 2009; Bossew et al., 2008).

3. The mapping radon problem

Over the last decades, a number of national radon projects have been carried out in several countries, e.g. in the U.S. (White et al., 1992), U.K. (Green et al., 2002), Ireland (Fennell et al., 2002), Finland (Weltner et al., 2002), Germany (Kemski et al., 1996), Austria (Friedmann, 2005), Czeck Republic (Neznal et al., 2004) and Italy (Bochicchio et al., 1996).

Considerable work is being invested into methods of estimating GRP from observed geological data and/or indoor Rn measurements (Bossew, 2014 and references therein) according to two main techniques of data processing. The first approach combines geological data and indoor radon measurements, sometimes including building characteristics and meteorological parameters that can affect radon entry into buildings (Pasculli et al., 2014; Bossew P., 2013; Gruber et al., 2013; Tung et al., 2013; Cinelli et al., 2010; Smethurst et al., 2008; Bossew et al., 2008; Appleton et al., 2008, 2011; Neznal et al., 2004; Kemski et al., 2001; Friedmann, 2005). The second approach consists in the direct interpolation of indoor radon values to identify RPAs. This last approach could be a difficult and non-robust procedure to accomplish at large scale, and could have a meaning only within the inhabited areas and in standard conditions (Miles, 1998a, b). However, because of its multifactorial dependence, indoor radon usually shows strong variability at least on short geographic scale (i.e., non-stationary spatial behaviour). Furthermore, the spatial distribution of indoor radon samples is often linked to the clustered distribution of houses within the inhabited zones, therefore a declustering procedure for estimating unbiased means and other statistics is required.

An alternative approach provides the construction of GRP maps considering only proxy information (i.e., soil permeability, faults, U and Ra content, emanation coefficient, etc.), calibrated through “what earth delivers”, i.e., soil gas radon measurements (Tondeur et al., 2014; Gruber et al., 2013; Ielsch et al., 2010; Kemski et al., 2001; Gundersen and Schumann, 1996). In these maps, RPAs can be recognized where the GRP coincides with indoor radon values above the reference level.

Usually, the construction of GRP maps involved global estimation techniques assuming spatial homogeneity of the relationships among radon and the other geological information. However, significant spatial variations characterize the relationships between pre-processed soil gas and indoor radon data, and soil/geochemical geological features. Therefore, the evaluation of factors influencing soil gas and indoor radon (i.e., geological and geochemical parameters) can be better performed by accounting for spatial autocorrelation. This means that the spatial variation of the relationship between radon in the environment and related variables is not constant but is dependent of the values of the variables at neighbouring sites.

The assessment of GRP of an area has been obtained by combining geochemical and geological parameters by using classical regression techniques (i.e., ordinary least squares) that imply the independence of the observations. However, the spatial analysis of the environmental variables that govern geogenic radon need to take into account their spatial autocorrelation (i.e., the observed value of a variable at one location is dependent of the values of the variable at neighbouring sites). This implies that multivariate classical statistical methods may be inappropriate for the modelling of this phenomenon. Therefore, more robust techniques of geospatial analysis taking into account the spatial variability of the direct and proxy variables should be considered.

4. Material and methods

The GRP mapping is a multivariate problem that can be addressed through the construction of a conceptual model based on the selection of the variables that most influence the presence of radon in the shallow environment. In this paper, a conceptual model is proposed based on geological, geochemical, structural and geomorphological data collected from the literature, as well as available soil gas field sampled data. These data are more suitable to construct GRP maps because they are characterised by: (i) higher spatial autocorrelation; (ii) lower variability; (iii) and not depend by anthropogenic factors with respect to the indoor radon data mainly affected by the building parameters.

The spatial relationships between geological data, in a broad sense, and the soil gas radon concentrations were then modelled by using global (Ordinary Least Squares, OLS) and spatial (Geographically Weighted Regression, GWR) multivariate regressions. In particular, the regression model includes a response variable (i.e., the radon concentration in soil gas) and some explanatory variables (i.e., the radium content of the rocks, the rock permeability, the presence of faults and fractures and the Digital Terrain Model, DTM). The final spatial model was used to estimate the response variable at unknown locations (Fig.1).

The GWR (i.e., local regression models) conducted in this study as a complementary approach to radon global spatial regression modelling, were calibrated using a computer software program, GWR 4.0 (https://geodacenter.asu.edu/gwr, Nakaya et al., 2009; Fotheringham et al., 2002). As the GWR outputs are location specific, they were integrated with ESRI ArcGIS software for computation, exploratory spatial data analysis, mapping and visualization. This software was chosen because it presents numerous extensions for spatial statistical and geostatistical modelling (Krivoruchko, 2011a, 2011b). Generally, these techniques were used to map spatial pattern, test relationships, check for redundancy among the explanatory variables and geo- visualization. The model's workflow is shown in figure 2.

4.1 Spatial Autocorrelation Analysis

Among the Exploratory Spatial Data Analysis (ESDA) techniques, the concept of spatial autocorrelation, i.e. the correlation of a single variable between pairs of neighbouring observations, constitutes one of the main topics for the analysis of geographical point data. The distribution of any natural phenomenon (i.e., radon potential) or its associated values (e.g., Rn in soil gas, Ra in rocks, etc.) within a space will produce a pattern. The geographic patterns range from completely clustered at one extreme to completely dispersed at the other. Patterns that fall between these extremes are assumed to be random. Knowing whether there is a pattern is useful for gaining a better understanding of a geographic phenomenon, monitoring conditions on the ground, comparing patterns or tracking changes (Mitchell, 2005).

The first "law" of geography, which states "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970), is a crucial idea in geography and particularly in spatial data analysis. In statistical terms, this law is related to the concept of spatial autocorrelation. In other words, when high values in a place tend to be associated with high values at nearby locations, or low values with low values for the neighbours, positive spatial autocorrelation or spatial clustering is said to occur. In contrast, when high values at a location are surrounded by nearby low values, or vice versa, negative spatial autocorrelation is present in the form of spatial outliers. In the analysis of spatial autocorrelation the reference point distribution is spatial randomness, e.g. the lack of any pattern structure. Global and local spatial autocorrelation indexes can be calculated to evaluate the existence of a pattern and, therefore, clusters in the spatial arrangement of a given variable.

In this work Moran's Index (I) and Getis-ord (G) indexes were used to preliminary test the global spatial autocorrelation of the studied variables (Moran, 1950; Getis et al., 1992). These indexes were used to estimate the strength of the correlation between observations as a function of the distance separating them. They share many similarities with Pearson's correlation coefficient: its numerator is a covariance, while its denominator is the sample variance. In addition, like a correlation coefficient, their values range from +1 meaning strong positive spatial autocorrelation/high values clustering, to 0 meaning a random pattern to -1 indicating strong negative spatial autocorrelation/low values clustering.

The Moran's I statistics for spatial autocorrelation of a variable is given as:

$$I = \frac{n}{S_0} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}, \quad (1)$$

where $\bar{x}$ is the mean of the x variable, w_{ij} are the elements of the weight matrix, and S_0 is the sum of the elements of the weight matrix: $S_0 = \sum_i \sum_j w_{ij}$.

Getis and Ord (1992) have recently proposed a different approach to measuring spatial association based on the definition of a neighbourhood for each location given by those observations that fall within a critical distance *d*. Getis-ord (G) measures the degree of clustering for either high values or low values. The general G statistics is given as:

$$G = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} x_i x_j}{\sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j} \qquad (2)$$

where x_i and x_j are attributes value for feature i and j, and w_{ij} are the elements of the weight matrix between feature i and feature j, n is the number of feature in the dataset. The G statistic takes values ranging between 0 and 1, where values close to 1 indicate clustering of high values, while values close to 0 indicate clustering of low values.

The major limitations of global Moran's *I* and Getis-ord G indexes are rooted in the fact that the former distinguish the clustering of high and low values, but does not capture the presence of negative spatial correlation; the latter is able to detect both positive and negative spatial correlations, but clustering of high or low values are not distinguished.

Global measures of spatial autocorrelation emphasize the average spatial dependence over the study region, hence they will only be useful if spatial dependence is relatively uniform over the study region. If the underlying spatial process is not stationary, global measures may not be representative, particularly if the size of the study region is relatively large. Local measures of spatial association aim at identifying patterns of spatial dependence within the local study regions. This has induced statisticians to develop local indices of spatial association (LISA) in order to examine the local level of spatial autocorrelation and to identify areas where values of the variable are both extreme and geographically homogeneous. This approach is most useful for the identification of so-called hot spots regions where the considered phenomenon is extremely pronounced across localities, as well of spatial outliers. The index to examine local autocorrelation is the Luc Anselin's LISA (Local Indicator of Spatial Association), which can be seen as the local equivalent of Moran's I.

Local Moran's I was proposed by Anselin (1995) and it is defined as follows:

$$I_i = \frac{x_i - \overline{X}}{S_0} \sum_{j=1, j \neq i}^{n} w_{i,j} (x_j - \overline{X}) \qquad (3)$$

For each location, local value of Moran's *I* allows for the computation of its similarity with its neighbours and to test its significance. A positive value for *I* indicates that a feature has neighbouring features with similarly high or low attribute values; this feature is part of a cluster. A negative value for *I* indicates that a feature has neighbouring features with dissimilar values: this feature is an outlier. In either instance, the p-value for the feature must be small enough for the cluster or outlier to be considered statistically significant.

Results of local value of Moran's *I* are expressed in term of statistically significance, p, at the 0.05 level, or in terms of z score. Very high or very low (negative) z-scores are associated with very small p-values (found in the tails of the normal distribution). According to the z-score values at the significance level of 0.05, the following scenarios may emerge:

- z-score > 1.96 indicates locations with high values with similar neighbours (*high-high, H-H)*, also known as "hot spots", as well as locations with low values with similar neighbours (*low-low, L-L)*, also known as "cold spots"
- z-score <-1.96 indicates locations with high values with low-value neighbours (*high-low, H-L)* and vice-versa low values with high-value neighbours (*low-high, L-H)*, indicating potential "spatial outliers"
- z-score >-1.96 and <1.96 indicates locations with no significant local autocorrelation.

In the same way of local Moran *I*, the Getis-Ord (Gi* statistics) measures the degree of clustering for either high values or low values. In particular, a high z-score and small p-value for a feature indicates a spatial clustering of high values. A low negative z-score and small p-value indicates a spatial clustering of low values. The higher (or lower) the z-score, the more intense the clustering. A z-score near zero indicates no apparent spatial clustering.

This index works by looking at each feature within the context of neighbouring features. A feature with a high value is interesting but may not be a statistically significant hot spot. To be a statistically significant hot spot, a feature will have a high value and be surrounded by other features with high values as well. The Getis-Ord local statistics is defined by the following equation:

$$G^*_i = \frac{\sum_{j=1}^{n} w_{i,j} x_j - \overline{X} \sum_{j=1}^{n} w_{i,j}}{S\sqrt{\dfrac{\left[n \sum_{j=1}^{n} w^2_{i,j} - \left(\sum_{j=1}^{n} w_{i,j}\right)^2 \right]}{n-1}}} \qquad (4)$$

where x_j is the attribute value for feature *j*, $w_{i,j}$ is the spatial weight between feature *i* and *j*, *n* is the equal to the total number of features and:

$$\overline{X} = \frac{\sum_{j=1}^{n} w_{i,j}}{n} \qquad (5) \qquad\qquad S_i = \sqrt{\frac{\sum_{j=1}^{n} x^2_j}{n} - (\overline{X})^2} \qquad (6)$$

The Gi* statistic results obtained for each feature in the dataset is also a z-score. For statistically significant positive z-scores (>1.96), the larger the z-score is, the more intense the clustering of high values (hot spot). For statistically significant negative z-scores (<-1.96), the smaller the z-score is, the more intense the clustering of low values (cold spot). In this paper both indexes are calculated by using the "Spatial Statistics Tool" of ArcGIS 10.2 (copyright©1999-2014 ESRI Inc.).

4.2 The Exploratory Regression

The procedure to find the optimal global model by using OLS regression can be difficult especially when there are lots of potential explanatory variables that may contribute to the modelled, dependent variable. The Exploratory Regression can help to try all possible combinations of explanatory variables to see which models pass all of the necessary OLS diagnostics (see next section). By evaluating all possible combinations of the candidate explanatory variables, the chances of finding the best model are greatly increased. ER is similar to stepwise regression (found in many statistical software packages). Rather than only looking for models with high Adjusted R^2 values, it looks for models that meet all of the requirements and assumptions of the OLS regression method. ER runs OLS on every possible combination of the candidate explanatory variables for models which include at least the minimum number of explanatory variables and not more than the maximum number of explanatory variables. The model that pass all the specified requirements and assumptions of the OLS will be candidate for the OLS method. To be a passable model, a set of parameters should be evaluated: the adjusted R^2 of 0.50 or higher, significance of the β coefficients (p- values that are less than 0.05), a Variance Inflation Factor (VIF) of less than 7.5, a Jarque-Bera statistics (p-value greater than 0.10), and a spatial autocorrelation test (p-value greater than 0.10). A brief description of each of these tests follows.

Adjusted R^2. The coefficient of determination (R^2) provides a summary of how much variation in a dependent variable's values is explained by a set of explanatory variables (Weisburd & Piquero, 2008). The *adjusted* R^2 is a recalibration of the R^2 value which generally artificially increases as more independent variables are added to a model (Theil, 1961). Thus, in a multivariate model, the *adjusted R^2* always lower than the 'raw' R^2. Like R^2, an *adjusted R^2* ranges between 0 and 1 (i.e., *adj. R^2* = 0.90 indicates that 90% of the variability in the dependent variable is explained by changes in the set of explanatory variables being modelled. In this study, a model that failed to explain at least 50% (after taking the *adjusted R^2* penalty) of the variability in soil gas radon concentrations resulted in its elimination from further consideration.

P-Value. Statistical inferences are typically made in the context of the null hypothesis. In the case of OLS regression modelling, the null hypothesis states that there is no linear relationship between a set of explanatory variables and a dependent variable. For OLS modelling, coefficients are produced which describe the y-intercept and the linear relationship between each independent/dependent variable. If a coefficient value is too large to be due simply to random chance, the null hypothesis should be rejected. The p-value provides the basis to take this decision because it quantifies the probability of obtaining a particular coefficient value when there is no relationship between the explanatory and the dependent variables (Kleinbaum, et al, 1998).

Variance Inflation Factor (VIF). This value represents a description of multicollinearity in a model. For models with two or more explanatory variables there may be correlations between them, which can result in highly unstable correlation coefficients (Kleinbaum, et al., 1998). The *VIF* measures multicollinearity by determining the extent of the increase caused by the correlations between explanatory variables (Kleinbaum, et al., 1998). Thus, the larger the *VIF* value, the more increase is present, and the model becomes more unstable. As a general heuristic, a *VIF* of 10.0 or higher is regarded as problematic. For this study, the *VIF* threshold was set more conservatively at 7.5.

Jarque-Bera Statistics. After relationships in a dataset have been modelled, predicted values can be computed using the observed independent variables; the differences between the predicted values and the observed values are the residuals of the regression model. If the model is robust, the residuals should be independent and normally distributed. Non-normally distribution of the residuals indicates a lack of organization and structure of the model error. Biased residuals indicate model misspecification, which in turn renders the results untrustworthy (Kleinbaum, 1998; Jarque and Bera, 1987) proposed a procedure to test a model's residuals for skewness and kurtosis (i.e., for normality). The null hypothesis for this procedure states that residuals are normally distributed. If the *Jarque-Bera* score is too high to be due to random chance the null hypothesis should be rejected. For this study, the *p-value* threshold for the Jarque-*Bera* test was set conservatively at 0.10, so that a models' residuals had an increased chance of being considered biased, which then increased the likelihood that a model would be excluded from consideration.

Spatially Autocorrelated Residuals. As mentioned above, a basic assumption of regression modelling is that there is no systematic structure to model residuals. While the *Jarque-Bera* test determines whether or not residuals are biased, if a model built on geographic data produces spatially biased predictions, this also violates the assumption of residual normality (Cliff & Ord, 1972). But the *Jarque-Bera* test cannot capture if residuals are spatially clustered (i.e., autocorrelated). Spatial autocorrelation techniques (detailed in section 4.1) can be applied to model over/under-predictions in order to ascertain the geographic pattern of residuals. If, for example, there are significant clusters of model residuals this would provide evidence of model misspecification. In short, autocorrelation of the residuals can increase the probability of finding significant coefficients that are not really significant; and/or it can mean that a key variable is missing from the model (Dormann et al., 2007). For this study, the *p-value* threshold for the spatial autocorrelation (i.e., *Moran's I*) test was set conservatively at 0.10, so that a models' residuals had an increased chance of being considered spatially autocorrelated, which then increased the likelihood that a model would be excluded from consideration. In summary, an appropriate set of 1990 predictor variables that did not violate regression assumptions needed to be established in order to move on the next phase and be tested as an OLS model.

In particular, a properly specified OLS model should provide:

- explanatory variables with the regression coefficients statistically significant and explaining justifiable relationships between each explanatory variable and the dependent variable;
- non redundant explanatory variables with small Variance Inflaction Factor (VIF) (in general less than 7.5);
- normally distributed residuals indicating your model is free from bias (the Jarque-Bera p-value should not be statistically significant);
- randomly distributed over and under predictions indicating model residuals are normally distributed (the spatial autocorrelation test should not provide significant statistics).

4.3 Ordinary Least Squares (OLS)

Spatial data often do not fit traditional, non-spatial regression requirements because they are: i) spatially autocorrelated (features near each other are more similar than further away); ii) non-stationary (features behave differently based on their location/regional variation). The general purpose of linear regression analysis is to find a (linear) relationship between a dependent variable and a set of explanatory variables, in the form:

$$y= \beta_0+\beta_1x_1+ \beta_2x_2+\ldots.. \beta_nx_n+\varepsilon \qquad (7)$$

where y is the dependent variable to predict, x_i is the explanatory variables and β_i are the coefficients computed by the regression tool, representing the strength and type of relationship between x and y, and ε are the residuals, i.e., the unexplained portion of the dependent explanatory variables. Large residuals indicate a poor fitting of the regression model.

Once a passable model had been established the OLS regression algorithm in ArcGIS will be applied. OLS is the best known of all regression techniques. It provides a global model of the variable or process you are trying to understand or predict, and creates a single regression equation to represent that process. The OLS regression estimates β coefficients minimizing the sum of squared prediction errors, hence, least squares.

The OLS tool in ArcGIS provided the *Akaike's Information Criterion* (*AIC)* as an additional information about model performance, and it also produced a map layer of model residuals, which allowed for the visualization of the global model's over/under-predictions. Furthermore, Qui and Wu (2011) pointed out that prior to conducting a GWR analysis, it is necessary to first confirm that predictor variables are statistically valid and significant through OLS regression (and the accompanying tests for violations of regression assumptions).

4.4 Geographical Weighted Regression (GWR)

Geographical Weighted Regression (GWR) is a local spatial statistical technique used to analyse spatial non-stationarity, i.e., the measurement of relationships among variables may differ at different locations. Unlike conventional regression, which produces a single regression equation to summarize global relationships among the explanatory and dependent variables, GWR generates spatial data that express the spatial variation in the relationships among variables. Maps generated from these data play a key role in exploring and interpreting spatial non-stationarity. Instead of calibrating a single regression equation, GWR provides separate regression equations for each observation of the dataset, consisting of a dependent (response) variable y and a set of k independent (explanatory) variables x_k, k=1…m, and of n observations for which their positions are available in a suitable coordinate system. Each equation is calibrated using a different weighting of the observations contained in the dataset. The equation for a typical GWR model is (Fotheringham et al., 2001; 1998):

$$y_i(u) = \beta_{0i}(u,v) + \beta_{1i}(u,v)x_{1i} + \beta_{2i}(u,v)x_{2i} + \ldots + \beta_{mi}(u,v)x_{mi} \quad (8)$$

The notation β_{0i} (u,v) indicates that the parameter β describes a relationship around location i (u,v) and it is specific of each location. A prediction of the dependent variable may be made if measurements for the independent variables are also available at the location i (u,v).

The calibration of the GWR model requires a decision regarding size of the subset of n observations to be included in the neighbour of the predicted values. This is referred to as the bandwidth (or "kernel" (Brundson et al., 1998) size for estimating the local regression parameters. For GWR it is ordinarily (but not necessarily) assumed that Tobler's first law applies to a given dataset. Thus, the default weighting scheme is that the soil gas values near to point i have more influence in the estimated regression values than values located far away from that same point (Fotheringham et al., 2001).

An adaptive or fixed kernel size can be selected. Using a fixed kernel ensures that area is preserved, so even though the number of local observations in the kernel area will change, the area represented by each local equation will remain constant (Brundson, et al., 1998). Alternatively, an adaptive kernel will ensure when the area of the kernel changes, the number of observations within each kernel area will remain the same. In case of highly irregular distributed observations, the most appropriate selection is the adaptive kernel (Fotheringham, et al., 2001). In this study we adopt the Gaussian fixed kernel type that weights continuously and gradually decreases from the centre of the kernel but never reaches zero. Gaussian kernel is suitable for fixed kernels since it can avert or mitigate the risk of there being no data within a kernel. The kernel shape is defined by the following equation which take into account only the nth nearest neighbours:

$$w_{ij} = \exp(-d^2_{ij} / \theta^2) \qquad (9)$$

where

i is the regression point index; j is the locational index;

w_{ij} is the weight value of observation at location j for estimating the coefficient at location i.

d_{ij} is the Euclidean distance between i and j;

θ is a fixed bandwidth size defined by a distance metric measure.

The calibration of the model involves also the choice of n, the number of data point to be included in the estimation of local parameters. Different methods are traditionally used to define the finest bandwidth value or the appropriate value of n. The GWR algorithm in GWR4 software provides different methods of doing this: the Akaike Information Criteria (Hurvich et al., 1998, Akaike, 1974) and the Cross- Validation score (CV) procedure (Cleveland, 1979). The bandwidth was optimized using an algorithm that seeks to minimize the corrected AICc score, given as:

$$AICc = 2n\log_e(\hat{\sigma}) + n\log_e(2\pi) + \left[\frac{n + tr(S)}{n - 2 - tr(S)}\right] \qquad (10)$$

In the Eq. (10), n is the sample size, $\hat{\sigma}$ is the estimated standard deviation of error term, and tr(S) is the trace of the hat matrix S of GWR, which is defined as:

$$\hat{y} = Sy \qquad (11)$$

where y is the vector of the dependent variable and is ŷ the vector of the GWR estimated value. Lower values of AICc indicate better model performance.

Following Pasculli et al. (2014) and Slagle (2010), and based on the work of Burnham and Anderson (2002), the *AIC* of the regression and the kernel size should be evaluated. AIC provides a description of the goodness-of-fit for a statistical model by comparing its complexity to its residual sum of squares. Models with lower AIC values are more performant (Fotheringham et al., 2003). AIC provides a way for comparing a global OLS model to a local GWR one Ultimately, the GWR algorithm produced a series of four maps depicting a continuous surface of regression coefficients for each predictor variable. GWR to construct a radon map was recently applied to elaborate a RPA model of Abruzzo region incorporating indoor radon and geological data (Pasculli et al., 2014).

In order to have an indication if there was a significant improvement in model performance of GWR over the ordinary global model, an ANOVA test was formulated (Brunsdon et al., 1998). It is an approximate likelihood ratio test based on F-test defined as follows:

$$F = \frac{RSS_0 / d_0}{RSS_1 / d_1} \quad (12)$$

Where RSS_1 and RSS_0 are the residual sum of squares of GWR model and global regression model (OLS) while d_1 and d_0 are the degrees of freedom for GWR and global models, respectively.

4.5 Study area

Lazio region is located in central Italy and has an area of 17,207 km^2 with 5.7 million of inhabitants. Lazio extends in central Italy along the Tyrrhenian Sea and is surrounded by the regions of Tuscany, Umbria, Marche to the north, Abruzzo, Molise to the east and Campania to the south. The region is divided into five administrative provinces**:** Rome, Frosinone, Latina, Rieti and Viterbo. The far most important city in Lazio is the Italian capital Rome, which is, in terms of area and population, the largest province of Lazio (Fig. 3).

Geologically, the study area is part of a passive Neogene-to-present continental margin along the eastern side of the Tyrrhenian Sea back arc basin. The Tyrrhenian basin has developed, since the Miocene, as a consequence of back arc extension associated with the west-dipping and “easterly” retreating Apennine subduction zone (Carminati et al., 2102, and references therein) (Fig. 3). Along the Latium Tyrrhenian margin, the extensional tectonics controlled the development of basins that trend primarily NNW-SSE/NW-SE and subordinately NE-SW. During the Plio-Pleistocene, clastic sediments (Milli, 1997) and volcanoclastic deposits belonging to the volcanic complexes of the Roman Magmatic Province (Peccerillo, 2005; Karner et al., 2001; De Rita et al., 1993) filled these basins. The Lazio region is characterized by a considerable lithological variability represented by different units of sedimentary and volcanic origin. Lithologies of the region of Lazio are shown in the

Geological Map of Lazio Region (scale 1:250,000) in vector format (Cosentino et al., 2012); moving from west to east the study area presents: Plio-Pleistocene marine sediments of the coastal floodplains (several hundreds of meters thick), volcanic deposits (lavas, tuffs, pyroclastics) of the Roman Comagmatic Province (these deposits constitute the main lithology of the region), and to the east the mountain ridges of the Apennine Chain, characterized by thick carbonatic sequences affected by karst, and incised by deep river valleys.

5. Results

5.1 Selection of proxy variables

The proposed modelling approach consists of two fundamental working phases (Fig. 2): the selection of appropriate geological and geochemical data and the preparation and pre-processing of the primary GIS layers. The objective of these preliminary phases was to set up a homogeneous database and calculate derived GIS layers to be used in the GWR analysis for spatial autocorrelation analysis, as well as modelling and calculation of the final GRP map of the Lazio region.

Data used in the present study include: (1) soil gas radon measurements (kBq/m^3); (2) the natural content of radiogenic elements (Bq/kg) (i.e., U and Ra) of the outcropping rocks; (3) the permeability (m^2) of the outcropping rocks; (4) the presence of faults and fracture as proxy variable of secondary permeability; (5) the Digital Terrain Model (DTM) (m) as proxy variable of meteorological parameters (i.e., rainfall and temperature).

The primary GIS layers include the following base maps:

- Soil gas radon concentrations (kBq/m^3) collected in the Lazio region (Fig.4);
- Geological Map of Lazio Region (scale 1:25,000) in vector format (Cosentino et al., 2012) (see supplementary material);
- Hydrogeological map of Lazio Region (scale 1:25,000) in vector format (Capelli et al., 2012 (Fig. 5);
- Map of the main faults obtained by Regione Lazio (Fig. 6);
- Digital Terrain Model in a grid format at 20x20m cell resolution (Fig 7);

Soil gas radon concentration (7625 samples) collected in the entire region was used as direct information of the GRP of an area (i.e., dependent variable in the GWR analysis). Soil gas data (published and unpublished data) are provided by the Fluid Geochemistry Laboratory of the Earth Science Department of Rome University Sapienza (see Ciotoli et al., 2003; Annunziatellis et al., 2003, 2008, 2010; Beaubien et al., 2003; Bigi et al., 2014) and by ARPA Lazio (2006).

The Geological Map of the Lazio Region was reclassified into homogeneous geological units (HGUs) according to the following geological formations: carbonate platform (limestone, marls), pelagic-slope basin (limestone, chert, marls), continental deposits (clay, sand and conglomerate), marine deposits (clay and sand), volcanics (lava, tuff, pyroclastics), flysch (Fig. 8). As volcanic rocks are derived by different volcanic systems of different age and rock types, the volcanic domain was further subdivided into four different homogeneous units from south to north, respectively: the Alban Hills volcano, the Sabatini volcano, the Vico volcano and the Vulsini volcano.

Considering that radon is produced by radioactive decay of radiogenic elements (mainly U, Th and Ra) contained in the substratum beneath the buildings, available radiometric data were used as proxy variable of radon production in the rocks. Then, the radium content in term of "equivalent" uranium (eU), and of the average radium content (Bq/kg), obtained from the literature (Castelluccio, 2010; Tuccimei et al., 2006; Voltaggio et al., 2006; Locardi, 1967) was assigned to each new HGUs of the Geological Map of the Lazio Region (Fig. 9). Similarly, permeability values (m^2) obtained by Spitz and Moreno (1996) were also assigned to each HGU to obtain a map of the rock permeability (Fig 10). Furthermore, as faults and fractures may constitute main pathways of radon movement in the subsoil (Baubron et al., 2002; Fu et al., 2005; Ciotoli et al., 2007; Walia et al., 2009; Bigi et al., 2014) the network of the main faults and fractures of the region has been used as proxy of secondary permeability, this is also consistent with the selection of the rock permeability as explanatory variable (Fig. 11). Finally, as climate can strongly affect radon exhalation at the soil/atmosphere boundary, the digital terrain model of the Lazio region was used as a proxy of the meteorological parameters (i.e., temperature, barometric pressure and rainfall) (Fig. 7).

Radium content, permeability, faults and DTM were then used as explanatory variables, whereas radon concentration in soil gas was used as response variable within all the applied regression models (OLS and GWR) to calculate the final GRP map. The considered geological and geochemical parameters are assumed homogeneous at the working scale, though they could show a high variability at more local scale.

5.2 Data pre-processing

Data pre-processing provided 1000x1000m raster grids obtained from the primary GIS layers (i.e., radium content and rock permeability) by using "polygon to raster" transformation tool in ArcGIS (Copyright © 1995-2014 Esri) (Figg. 9 and 10). The problem caused by the presence of multiple polygons within the unit grid cell was overcome by assigning to the grid cell the value of the largest polygon inside. The layer of faults was gridded by using Kernel Density algorithm to obtain a

1000x1000m fault density map (m/km^2) (Fig. 11). Kernel Density calculates the density of point/line features around each output raster cell by fitting a smoothly curved surface over each point/line. The surface value is highest at the location of the point/line and diminishes with increasing distance, reaching zero at the search radius distance from the point/line. The kernel function is based on the quadratic kernel function described in Silverman (1986). The DTM at 20x20m grid was re-sampling at 1000x1000m grid resolution (Fig. 7).

This pre-elaboration phase provided four 1000 x 1000 m grid maps representing the predictor variables, respectively radium content, permeability, fault density, and DTM. Then a regular point layer (12911 points) corresponding to the centroid of all the grid cells was generated. Table 1 reports summary statistics of the raw radon values, as well as the other considered parameters.

The distribution of the 7625 soil gas samples was regularised in order to match with the 1000x1000 grid of the other layers by using the "point to raster" tool. As more than one soil gas sample may occur within the unit grid, the geometric mean (GM) of radon values measured at these locations was assigned to the grid cell. Then the application of the "Extract multivalues to point" tool assigned to the regular 1000x1000 point layer (the centroids of all the 12911 cells) the corresponding grid values of each created raster map. The final attribute table includes 12911 records with complete information of all the predictors and 2529 records with the radon data (the response variable). GWR model was calculated by using only the records (centroids) that have measured/calculated values of all the variables (2529 samples). This dataset was used as training dataset to construct the GWR model that then will be validated to the total dataset (7625 samples). After the validation and the selection of appropriate model, this was applied to all the extracted centroids of the 1km x 1 km grid to obtain the final radon potential map of the Lazio region. All these techniques are available in the ArcGIS "Geostatistical Analyst" extension and in the "Spatial Statistics" tool.

5.3 Preliminary analysis of the soil gas radon

Table 2 reports specific statistics for radon data collected in the Lazio region. The similarity between the Geometric Mean (GM) and the median suggest a log-normal distribution for this variable. In general, the quite large variability observed in soil gas radon values can be caused to local variations of the geological, geochemical and geophysical characteristic of the soil. Some authors proposed a standardisation, in order to emphasize the role of geology filtering out the variability due to the shallower environmental factors (i.e., porosity, humidity, etc.) (Miles, 1998; Friedmann, 2005). In the case of radon data, the log-transformation could be a suitable compromise to reduce its variability (Fig. 12).

Furthermore, radon data were intersected with the HGUs layer in order to calculate statistics within each of the HGUs (Tab. 3). The table highlights that highest mean values occur in correspondence of the main volcanic areas of the Lazio region, though the number of samples are different for each of the HGU. These high radon values can be linked to the high radionuclide content of these rocks. Carbonate rocks and flysch also show high mean values probably caused by the presence of faults and fractures, as well as by the enhanced groundwater circulation. The box-plot of figure 13 well highlights this particular behaviour of radon within the HGUs.

5.4 Spatial autocorrelation

Spatial autocorrelation analysis was carried out in order to assess if the studied variables exhibit a random pattern or if they show a significant spatial structure. Initially global indexes were applied to determine the general pattern; following the application of local indicators allowed the identification the presence of significant clusters of high or low values as well as some interesting spatial outliers. Global and local indexes of autocorrelation were applied to the variable selected in the section 4.5 to detect their spatial autocorrelation, spatial pattern and the presence of cluster and outliers in the study area.

Global Moran's I and Getis-ord (G) indexes were obtained by using the *Spatial autocorrelation* and *High/Low Clustering* commands of the *Spatial Statistics Tools* in ArcGIS 10.2. Results for both indexes expressed as z-score and p values are reported in table 4. Morans' I result highlights that all variables are positively spatially autocorrelated (significant p-values). Then further analysis at global scale should indicate if one would expect that explanatory variables and soil gas radon values appear clustered or positively associated throughout the Lazio region. In other words, the global Getis-ord (G) statistics is used to test if the sites with relatively high values of the analysed variables (respectively low) are localized close to other sites with high values of those variables (respectively low), or if they show a purely random pattern. Getis-ord (G) results indicate that at global scale high values of the studied variables seem to be clustered.

As reported in the section 4.1, these results of global spatial autocorrelation can be refined by using Local Indexes of Spatial Autocorrelation (LISA). Further details of the clustered patterns of the studied variables can be highlighted by examining the extent of local spatial autocorrelation by using the Cluster and Outliers Analysis (Anselin Local Moran) and Hot Spot Analysis (Local Getis-ord G) tools in the ArcGIS environment.

As LISA indexes are scale-dependent, they need an appropriate selection of optimal distance threshold for the working spatial scale. The "Spatial Statistics Tool" of the ArcGIS software provides the "Incremental Spatial Autocorrelation" tool to conceptualize the spatial relationships among data. The "Incremental Spatial Autocorrelation" tool runs the Global Moran's I test for a series of increasing distances, measuring the intensity of spatial clustering for each distance. The intensity of clustering is determined by the z-score returned (e.g., p-value). Typically, as the distance increases (and consequently the z-score), intensification of clustering is present in the data. At some particular distance, however, the z-score generally peak. The peak reflects the distance where the spatial processes causing clustering is most pronounced. The "Incremental Spatial Autocorrelation" tool was applied to soil gas radon data in order to conceptualize the spatial relationships for the following "Hot Spot Analysis" by using LISA indexes. In this case, the following input data were used: beginning distance = 1000 m, distance increment = 5000 m and number of increments = 10. Results indicate a significant max peak distance value of 16000 m (z-score=206.87 and p=0.000) (Fig 14). This distance represents the distance where the spatial processes promoting clustering is most pronounced.

Figure 15 shows the combined results of Anselin Local Moran and Local Getis-ord G statistics in terms of hot spots (areas where locations with high radon values are surrounded by high values: HH), cold spots (low values surrounded by low values: LL) and local outliers (HL). Major hot spots are located in the volcanic areas of Cimini Mts. (Viterbo Province, northern Lazio), and of the Alban Hills (south of Rome); further hot spot zone is located along the NW-SE Lepini Mts. carbonate ridge (central-southern Lazio). Major cold spots occur in the Tolfa Mts. and Comino valley area. Finally, potential spatial outliers are mapped, i.e. locations with high values with low-value neighbours (HL), they occur between the Bracciano lake and the Tolfa Mts.

5.5 Exploratory regression results

Preliminary Exploratory Regression will try all possible combinations of explanatory variables to see which model passes all of the necessary OLS diagnostics. This analysis greatly increases the chances of finding the best model. Summary statistics on candidate models must respect: high goodness of fit, model significance, collinearity, normal distributed residuals, spatially uncorrelated residuals. Table 5 reports the overall statistics of the ER model. The model did not pass all the required cutoff, though each independent variable was significantly related to the soil gas radon except for the DTM (Tab. 6). Significant Jarque-Bera statistic indicates a biased model with not normally distributed residuals. The null hypothesis for this test is that the residuals are normally distributed. Furthermore, the spatial autocorrelation test (Global Moran's I) highlights that residuals are not spatially random, but significant clustering of high and /or low residuals (model under- and overpredictions).

5.6 OLS regression model

The OLS regression is also the proper starting point for all spatial regression analyses. It provides a global model of the variable or process you are trying to understand or predict; it creates a single regression equation to represent that process. The previous model was tested with the Ordinary Least Square regression in order to evaluate the effects of the geological environment on soil gas radon measurements and test for the possibility that the effect of the predictor variables on the dependent variable varies continuously over space. Results confirm that the model explains approximately 15% of the variation in the explanatory variables. The model significance is assessed by the Joint F-Statistic and the Joint Wald Statistic. The Joint F-Statistic is trustworthy only when the Koenker (BP) statistic (see table 7) is not statistically significant. In this case the the Koenker (BP) statistic is significant, therefore the the Joint Wald Statistic highlights a statistically significant model.

Furthermore, the Koenker (BP) (Koenker's studentized Bruesch-Pagan statistic, BP) statistic determines whether the explanatory variables in the model have a consistent relationship to the dependent variable both in geographic space and in data space. When the model is consistent in geographic space, the spatial processes represented by the explanatory variables behave the same everywhere in the study area (the processes are stationary). When the model is consistent in data space, the variation in the relationship between predicted values and each explanatory variable does not change with changes in explanatory variable magnitudes (there is no heteroscedasticity in the model). In this case the significance of the Koenker (BP) statistic indicates heteroscedasticity and/or non-stationarity of the model; this model results, therefore, a good candidate for Geographically Weighted Regression analysis. The Jarque-Bera statistics reported in table 7 indicates that residuals are non-normally distributed; a statistically significant Jarque-Bera test can also occur when there is strong heteroscedasticity.

Table 8 reports the coefficient diagnostic table that captures important elements of the OLS regression. The table highlights that coefficients of the explanatory variables are consistent (low Standard Error), but confirms that DTM is not significant in the OLS model (Robust Probability not significant). Therefore, the DTM is not used as explanatory variables in the GWR regression model.

The resulting OLS regression equation is:

$$y\ (soil\ gas\ Rn) = 5.12 + 0.0003\ (Fault) - 0.0012\ (DTM) + 0.041\ (Ra) + 0.218\ (Perm) \qquad (13)$$

Radium content and rock permeability highlight the highest coefficients.

The histogram of the regression residual highlights that residuals are non-normally distributed and the Global Moran's I indicates that are spatially autocorrelated (Fig. 16).

5.7 Spatial regression model

A spatial regression model by using GWR taking into account the relationships between the radon concentration in soil gas (i.e., dependent variable) and the radium content, the rock permeability, the fault density (i.e., explanatory variables) was then calculated. The GWR model was constructed by using GWR4 software was applied to the training dataset (20558 points), for a faster performance, and then applied to the test data set. All data are standardised to reduce variability and avoid the problem caused by different measure unit. Coefficients of the explanatory variables are estimated using nearby feature values by using the kernel density algorithm. The Fixed Gaussian kernel type was used to solve each local regression analysis at fixed distance (i.e., bandwidth parameter). The optimal bandwidth (4000m) has been calculated automatically by the GWR4 software.

Table 9 and table 10 report the main statistics of the overall GWR model parameters as well as for the local parameters. The model provides a total Adjusted R-Squared (Adj.R^2) parameter of 0.935. Local R^2 values range between 0.0 and 1.0, and indicate how well the local regression model fits observed *y* values. The map of the Local R^2 values indicates where GWR predicts well, and where it predicts poorly; it may provide clues about important variables that may be missing from the regression model (Fig. 17).

Very high amount of variance was explained by the selected GWR model (R2 Adjusted= 0.935). Also the model performance (AICc = 143.74) results better than that calculated by the OLS model (AICc = 19512). Sigma value is the estimated standard deviation for the residuals. Smaller values of this statistic are preferable. The histogram of the standardized residuals of the regression shows a Gaussian distribution confirming the good performance of the model (Fig. 18a), and the Morans'I autocorrelation test highlights that residuals are not spatially autocorrelated (Fig. 18b). The ANOVA test allowed the comparison between the significance of the Global Residuals (OLS model) and the Local residuals (GWR model) indicating the improvement provided by the GWR model (Tab. 11).

The mean values of the coefficients calculated by the GWR model confirmed that soil gas radon is mainly positively related to the radium content, the rock permeability and the fracture density across the study region (Tab. 10). It is well known that after radium (i.e., radionuclides) concentrations in soils and rocks, the "interconnection" of pore space (primary permeability) in soil, as well as mainly the rock fracturing, are probably the most significant factors influencing the concentration of radon in rocks, soils, as well as in buildings. The GWR model is described by the following equation:

$$SoilRn_i = \beta_0(u,v) + \beta_1(u,v)Ra + \beta_2(u,v)Perm + \beta_3(u,v)Fault + \varepsilon_i \qquad (14)$$

The calculated GWR model was then used to predicted values at the 1 x 1 km square grid by using Simple Kriging (SK) (Webster and Oliver, 2007; Bohling, 2005; Stein, 1999; Isaaks and Srivastava, 1997) to obtain the final Radon Potential map of the Lazio Region. Kriging interpolation requires the analysis and modelling of the experimental variogram. The variogram map highlights an anisotropic behaviour of the GWR predictions along N330 direction (Fig. 19a). Therefore, the anisotropic experimental variogram was then modelled along this direction by using an exponential model with a nugget value of 0.07, a sill of 1, as well as a maximum range of anisotropy of 50000 m and a minimum range of anisotropy of 25000 m (γ = 0.07nugget + 1.0 Exp (50000, 25000, 337) (Fig. 19b). The model fitting was provided by cross-validation graph that indicates a Root-Mean-Square Standardised of 0.790 and a Gaussian standardised error distribution (Fig. 20). Figure 21 shows the final GRP map.

6 Discussion

Soil gas radon is a complex multivariate phenomenon, affected by different environmental factors, including geochemical and mechanical characteristics of rocks and soil, as well as meteorological parameters. This work is one of the first attempt to study the correlations between some of these factors and a high number of soil gas data in order to define the "geogenic radon potential" of an area (Dubois et al., 2010).

In this paper a modelling approach accounting for spatial effects by means of the Geographically Weighted Regression, a local spatial regression technique, was used to assess the effects of the most affecting factors on soil gas radon, by using more than 7000 radon data from soil gas field surveys and coupled to some geological and geochemical parameters collected provided by literature review. The GWR was tested for the entire Lazio region characterised by heterogeneous morphology, fractured and faulted areas with high permeable rocks, as well as by four volcanic complexes of different age with a high radionuclide content.

The proposed procedure involves the construction and the comparison of global and local multivariate regression models, as well as autocorrelation indexes, as suitable tools to highlight the presence of local effects and the differences among the variables associations across space. The models were constructed starting from a conceptual model which include the following variables: soil gas radon (i.e., the response variable), and radium content and permeability of outcropping rocks, fault and fractures distribution and the Digital Terrain Model as a proxy variable of meteorological factors (i.e., explanatory variables).

In this specific case study, global and local indicator of spatial association (i.e., Morans'I and Getis-ord statistics), indicated clear evidences of spatial correlation in the distribution of the considered geological geochemical variables, as well as soil gas radon data throughout the study area. In addition to the identification of clustered distribution, the local Moran test reveals the presence of many locations that do not follow the global process of spatial dependence (i.e., spatial outliers). In fact, the preliminary OLS global regression model confirmed these results highlighting spatial autocorrelation and non-random distribution of the residuals. Furthermore, OLS indicated that morphology (DTM) was not a significant variable in the model, then DTM was excluded by further elaborations.

The GWR was carried out to study the spatial distributions and the strengths of the local relationships between soil gas radon and the considered geological and geochemical factors. Results suggests that the spatial variations in the radium content and permeability of rocks, as well as the presence of a network of faults and fractures significantly affect the radon concentrations in soil gas.

Local results identify areas where the model predicts well ($R^2 > 0.5$) and where it predicts poorly ($R^2 < 0.5$). However, GWR models has an advantage over the global regression models, which often mask the geographic heterogeneity and the complex associations that might exist between variables over space.

Fotheringham (1997) suggested that in global models non-complete dataset with missing information may cause spatial heterogeneity. Since complete datasets are difficult to obtain in the case of multivariate phenomena, the inclusion of spatial information into local modelling techniques can significantly improve model predictability. However, the local GWR coefficients should be interpreted with caution especially in the case of local multicollinearity of the explanatory variables; this increases the variances of the estimated regression coefficients and can invalidate conclusions about the relationships based on the estimated coefficients (Pasculli et al., 2014). Many works (Páez et al., 2011; Griffith, 2008; Wheeler, 2007) highlighted that the lack of multi-collinearity in global regression models is not a guarantee for high performant GWR models.

Some critical issues to the obtained results can be resumed as follows. First of all, the Lazio territory shows a high degree of morphological and geological complexity which could be properly represented only by very accurate and detailed dataset. This condition may explain why the DTM did not result as significant in the global regression model. Maybe morphology should be considered in more detail including local characteristics such as the presence of dolines, caves, sinkhole, as well as other forms (i.e., ridges and valleys). Furthermore, instead of considering the DTM as proxy variable of meteorological parameters, it would be better to include in the model direct variables such as, rainfall, air temperature, soil moisture, etc.

In this study, the values and the spatial distribution of the soil gas radon data can be mainly related to the geochemical characteristics of the recognised HGUs, which in the central-northern part of the region highlight middle-to high concentrations of radionuclide content (i.e., U, Ra, Th) due to the presence of volcanism. Whereas in the southern and eastern sectors the GRP could be affected by the high fracturing (i.e., secondary permeability) despite the overall low radionuclide content of the outcropping carbonate rocks. However, the distribution of U and Ra in rocks could also be high variable within the same HGU due to the presence of volcanics in sedimentary deposits, as well as for alteration processes that can enrich or deplete the radionuclides content. At this regard, radionuclide content in soil and soil permeability could be important factors to be included in the conceptual model for the definition of GRP.

7 Conclusions

The use of GWR as a local spatial regression technique with respect to global regression models suggests that the local spatial variations of the bedrock radium content, the rock permeability, as well as the karst and fractured areas significantly affect soil gas radon concentrations. Therefore, the GWR technique highlights a high performance in the mapping of the GRP at the scale of the geological/geographical scenario of the Latium Region. The presented work can be considered as the first step toward the development of a conceptual model that should include as much as robust and correlated explanatory variables to predict the GRP of an area. Obviously, the proposed model is only one of the possible models that can be applied to define the GRP of a region. This model was obtained by using literature data trying to respect the principle of parsimony of included variables. This does not mean that more complex models may not be the best performing.

Specific conclusions of this work are following:

- the mapping of the GRP is a multivariate phenomenon; therefore, the application of multivariable local regression techniques seems to be more appropriate than the global regression models;
- GWR model highlights a higher mapping performance than the global regression models which often mask the spatial autocorrelation, and the complex associations among spatial variables;

- the proposed procedure was applied to a set of «a priori» selected variables that could affect the radon emission in the shallow environment, however this does not mean that the proposed model cannot be implemented by including/excluding other variables;
- to achieve this goal, significant improvements in the modelling could derive from more detailed analysis of the phenomenon including more accurate dataset of explanatory variables such as the total gamma radiation, the soil characteristics (i.e., permeability and radionuclide content), as well as the use of real climate data instead of the altitude. The further, and ambitious step, is to include also the indoor radon as response variable to obtain the map of the "radon prone areas".
- the proposed method is quite fast to carry out and, starting from soil gas data, can be applied preliminarily by using a geological map and literature data.

An accurate GRP map can be a useful tool to implement radon policies at both national and local level, providing the priority to better know the territory for which they have to take decisions. GRP maps can help in allocating resources to plan more efficiently denser surveys of both soil gas and indoor radon, as well as remediation and monitoring of affected houses and targeting regulation in priority areas. For example, it can provide some hints to determine the proper sample size of radon surveys, as more buildings need to be monitored in those areas characterised by high GRP.

Acknowledgment

The authors would like to thank all the researchers of the Fluid Geochemistry Laboratory of the University of Rome Sapienza that contributed to collect the large database used in this work. Furthermore, we are grateful to Antonio Gerardi and Giacomo Catalano of the Direzione Regionale Infrastrutture, Ambiente e Politiche Abitative – Area Difesa del Suolo e Bonifiche of the Regione Lazio for their support and precious suggestions over the course of the research.

References

Al-Shereideh, S.A., Bataina, B.A., Ershaidat, N.M., 2006. Seasonal variations and depth dependence of soil radon concentration levels in different geological formations in DeirAbu-Said district, Irbid e Jordan. Radiat. Meas. 41, 703-707.

Akaike H., 1974. A new look at the statistical model identification. IEEE Trans Automat Contr AC 19:716–723

Annunziatellis A., Ciotoli G., Guarino P.M., Nisio S., 2010. Nuovi dati sui sinkholes del bacino delle Acque Albule (Tivoli, Roma). Atti del 2° Workshop Internazionale “I sinkholes. Gli sprofondamenti catastrofici nell’ambiente naturale ed in quello antropizzato”, Roma 3-4 dicembre 2009, Auditorium ISPRA, Via Curtatone 7, 00185 Roma.

Annunziatellis, A., Ciotoli, G., Lombardi, S., Nolasco, F., 2003. Short- and long-term gas hazard: the release of toxic gases in the Alban Hills volcanic area (central Italy). J. Geochem. Explor. 77, 93–108.

Annunziatellis, A., Beaubien, S.E., Bigi, S., Ciotoli, G., Coltella, M., Lombardi, S., 2008. Gas migration along fault systems and through the vadose zone in the Latera caldera (central Italy): implications for CO2 geological storage. Int. J. Greenhouse Gas Control 2, 353–372. http://dx.doi.org/10.1016/j.ijggc.2008.02.003.

Anselin, L., 1995. The Local Indicators of Spatial Association “LISA". Geographical Analysis. 27, 93-115.

Appleton, J. D., Miles, J. C., Green, B. M., and Larmour, R. , 2008. Pilot study of the application of Tellus airborne radiometric and soil geochemical data for radon mapping. J. Environ. Radioact. 99, 1687–1697.

Appleton, J. D., Doyle, E., Fenton, D., and Organo, C., 2011. Radon potential mapping of the Tralee-Castleisland and Cavan areas (Ireland) based on airborne gamma-ray spectrometry and geology. J. Radiol. Prot. 31, 221–235.

Apte, M.G., Price P.N., Nero A.V., Revzan K.L.,1999. Predicting New Hampshire indoor radon concentrations from geologic information and other covariates. Environ. Geol. 37(3), 181-194, doi: 10.1007/s002540050376.

ARPA Lazio, 2006. Individuazione delle aree a rischio radon. Final Report, DOCUP Obiettivo 2 Lazio 2000-2006 Asse I “Valorizzazione ambientale” Misura I.4 “Azioni di controllo, monitoraggio e informazione ambientale”.

Barnet, I., Pacherová, P., Neznal, M., Neznal, M., 2008. Radon in Geological Environment – Czech Experience. Czech Geological Survey Special Paper 19. Prague: Czech Geological Survey.

Baubron, J. C., A. Rigo, and J. P. Toutain, 2002. Soil gas profiles as a tool to characterise active tectonic areas: The Jaut Pass example (Pyrenees, France). Earth Planet. Sci. Lett. 196, 69– 81.

Baykut, S., Akgül, T., Inan, S., Seyis, C., 2010. Observation and removal of daily quasi- periodic components in soil radon data. Radiat. Meas. 45, 872-879.

Bigi S., Beaubien S.E., Ciotoli G., D’Ambrogi C., Doglioni C., Ferrante V., Lombardi S., Milli S., Orlando L., Ruggiero L., Tartarello M.C., Sacco P., 2014. Mantle derived CO2 migration along active faults within an extensional basin margin (Fiumicino, Rome, Italy). Tectonophysics. 637, 137-149

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 Phys. 71(5), 743–750.

Bohling, G., 2005. Introduction to Geostatistics and Variogram Analysis. Kansas Geological Survey. Available at: http://people.ku.edu/~gbohling/cpe940/Variograms.pdf.

Bossew P., 2015. Mapping the Geogenic Radon Potential and Estimation of Radon Prone Areas in Germany. Radiation Emergency Medicine. 4(2), 13-20

Bossew P., 2014. Determination of radon prone areas by optimized binary classification. Journal of Environmental Radioactivity. 129, 121-132

Bossew, P., 2013. A radon risk map of Germany based on the geogenic radon potential. In: Pardo-Igúzquiza, E., et al. (Eds.), Mathematics of Planet Earth, Lecture Notes in Earth System Sciences. Springer, pp. 527-531. Pres., IAMG (Intl. Ass. Math. Geosciences) 2013, Madrid 2e6 Sept 2013. http://dx.doi.org/10.1007/9783-642-32408-6_115.

Bossew P., Dubois G., Tollefsen T., 2008. Investigations on indoor Radon in Austria, part 2: Geological classes as categorical external drift for spatial modelling of the Radon potential. J. Environ Radioact. 99(1), 81–97.

Brundson, C., Fotheringham, A.S., & Charlton, M.E., 1998. Geographically weighted regression-modeling spatial non-stationarity. Journal of Royal Statistical Society (Series D). 47(3), 431-443.

Burnham K.P., Anderson D.R., 2002. Model selection and multimodel inference: a practical information-theoretic approach, 2nd edn. Springer, New York

Capelli G., Mastrorillo L., Mazza R., Petitta M., Baldoni T., Banzato F., Cascone D., Di Salvo C., La Vigna F., Taviani S., Teoli P., 2012. Carta Idrogeologica della Regione Lazio, scala 1:100000 (4 fogli). Regione Lazio, S.EL.CA. Firenze. A cura della Regione Lazio, Area Difesa del Suolo.

Carminati E., Lustrino M., Doglioni C., 2012. Geodynamic evolution of the central and western Mediterranean: Tectonics vs. igneous petrology constraints. Tectonophysics, 579, 173–192, doi: 10.1016/j.tecto.2012.01.026

Castelluccio M., Giannella G., Lucchetti C., Moroni M., Tuccimei P., 2012. La classificazione della pericolosità radon nella pianificazione territoriale finalizzata alla gestione del rischio. Classification of radon hazard in urban planning focused to risk management. Italian Journal of Engineering Geology and Environment. 2, 5-16, DOI: 10.4408/IJEGE.2012-02. O-01

Castelluccio M., 2010. Soil radon concentration survey in Caffarella Valley test site (Rome) Ph.D. Thesis in Geodynamics at the “Roma Tre” University.

Charles M., 2001. UNSCEAR Report 2000: sources and effects of ionizing radiation. J. Radiol. Prot. 21(1), 83-85.

Ciotoli, G., Lombardi, S., Annunziatellis, A., 2007. Geostatistical analysis of soil gas data in a high seismic intermontane basin: Fucino Plain, central Italy. J. Geophys. Res. 112, B05407. http://dx.doi.org/10.1029/2005JB004044.

Beaubien S.E., Ciotoli G., Lombardi S., 2003. Carbon dioxide and radon gas hazard in the Alban Hills area (central Italy). Journal of Volcanology and Geothermal Res. 123 (1-2), 63-80

Cinelli, G., Tondeur, F., Dehandschutter, B., 2010. Development of an indoor radon risk map of the Walloon region of Belgium, integrating geological information. Environmental Earth Sciences 62 (4), 809-819. http://dx.doi.org/10.1007/s12665-010-0568-5.

Cleveland, W., 1979. Robust locally weighted regression and smoothing scatterplots. J. Am. Stat. Assoc. 74, 829-836.

Cliff, A., Ord, J., 1972. Testing for spatial autocorrelation among regression residuals. Geographic Analysis. 4, 267–284.

Cosentino D., Pasquali V., 2012. Carta Geologica Informatizzata della Regione Lazio. Università degli Studi Roma Tre Dipartimento di Scienze Geologiche, Regione Lazio Agenzia Regionale Parchi Area Difesa del Suolo. Coordinamento editoriale G. Catalano, C. Fattori, D. Mancinella, F. Meloni.

Crockett R.G.M., Perrier F., Richon, P., 2010. Spectral-decomposition techniques for the identification of periodic and anomalous phenomena in radon time-series. Nat. Hazard. Earth Sys. 10, 559-564.

Darby S., Hill D., Auvinen A., Barros-Dios J.M., Baysson H., Bochicchio F., 2005. Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case-control studies. Br. Med. J. 330 (7485), 223. http://dx.doi.org/10.1136/bmj.38308.477650.63.

De Rita D., Funiciello R., Corda L., Sposato A., Rossi U., 1993. Volcanic unit. In: Di Filippo, M. (Ed.), Sabatini Volcanic Complex, Consiglio Nazionale Delle Ricerche. Progetto Finalizzato “Geodinamica” Monografie Finali 11, 33–79.

Dormann C., McPherson J., Araujo M., Bivand R., Bollinger J., Carl G., Davies R., Hirzel A., Jetz W., Kissling D., Kuhn I., Ohlemuller R., Peres-Neto P., Reineking B., Schroder B., Schurr F., Wilson R., 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography. 30 (5), 609-628.

Drolet J.P., Martel R., Poulin P., Dessau J.C., Lavoie D., Parent M., Levesquem B., 2013. An approach to define potential radon emission level maps using indoor radon concentration measurements and radiogeochemical data positive pro- portion relationships. J. Environ. Radioact. 124, 57-67.

Dubois, G., Bossew, P., Tollefsen, T., De Cort, M., 2010. First steps towards a European atlas of natural radiation: status of the European indoor radon map. J. Environ. Radioactiv. 101, 786-798.

EC (European Commission), 1990. Protection of the public against indoor exposure to radon. Official Journal of the European Communities 1990. L80/26.

EC (European Commission), 2013. Council Directive 2013/59/Euratom of 5 December 2013 laying down Basic Safety Standards for Protection against the Dangers Arising from Exposure to Ionising Radiation. Official Journal L13 of 17/01/2014 European Commission, Bruxelles.

Etiope G., Martinelli G., 2002. Migration of carrier and trace gases in the geosphere: an overview. Phys. Earth Planet. Inter. 129, 185-204.

Fennell S.G., Mackin G.M., Madden J.S., McGarry A.T., Duffy J.T., O'Colmain M., Colgan P.A., Pollard, D., 2002. Radon in Dwellings. The Irish National Radon Survey. RPII- 02/1, Dublin: Radiological Protection Institute of Ireland.

Fotheringham S., Brunsdon C., Charlton M., 2003. Geographically weighted regression: the analysis of spatially varying relationships. John Wiley & Sons, eds.

Fotheringham A. S., Brunsdon C., Charlton M. E., 2002. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships, Wiley, Chichester.

Fotheringham A., Charlton M., Brunsdon C., 2001. Spatial variations in school performance: a local analysis using geographically weighted regression. Geographical and Environmental Modeling. 5(1), 43-66.

Fotheringham A.S., Brunsdon C., Charlton M.E., 1998. Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. Environ. Plann. A 30 (11), 1905-1927.

Friedmann, H., 2005. Final results of the Austrian radon project. Health Physics. 89 (4), 339-348.

Fujiyoshi R., Sakamoto K., Imanishi T., Sumiyoshi T., Sawamura S., Vaupotic J., Kobal I., 2006. Meteorological parameters contributing to variability in 222Rn activity concentrations in soil gas at a site in Sapporo, Japan. Sci. Total Environ. 370, 224-234.

Fu C. C., Yang T. F., Walia V., Cheng C.H., 2005. Reconnaissance of soil gas composition over the buried fault and fracture zone in southern Taiwan. Geochem. J. 39, 427–439.

Gates A.E., Gundersen L.C.S., 1992. Geologic Controls on Radon; Geological Society of America: Washington, DC, USA, 1992; Special Paper 271.

Getis A., Ord J. K.., 1992. The Analysis of Spatial Association by Use of Distance Statistics. Geographical Analysis 24 (3).

Green B.M.R., Miles J.C.H., Bradley E.J., Rees D.M., 2002. Radon Atlas of England and Wales. NRPB-W26, Didcot, UK: National Radiological Protection Board.

Griffith D.A., 2008. Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR). Environ. Plan. A 40, 2751-2769.

Gruber, V., Bossew, P., De Cort, M., Tollefsen, T. 2013. The European map of the geogenic radon potential. J. Radiol. Prot. 33, 51-60.

Gundersen L.C., Schumann R.R., 1996. Mapping the radon potential of the United States: examples from the Appalachians. Environ International. 22, 829-837

Hurvich C.M., Simonoff J.S., Tsai C.L., 1998. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. J. R. Stat. Soc. B 60, 271-293

IARC, 1988. Man-made Mineral Fibres and Radon; WHO: Geneva, Switzerland, 1988; Available on line: http://monographs.iarc.fr/ENG/Monographs /vol43/volume43.pdf.

Ielsch G., Cushing M.E., Combes Ph., Cuney M., 2010. Mapping of the geogenic radon potential in France to improve radon risk management: methodology and first applications to region Bourgogne. J. Environ. Radioact. 101, 813-820.

Isaaks, E. H., Srivastava R.M., 1989. An Introduction to Applied Geostatistics, 561 pp., Oxford Univ. Press, New York.

Karner D.B., Marra F., Renne P.R., 2001. The history of the Monti Sabatini and Alban Hills volcanoes: groundwork for assessing volcanic–tectonic hazards for Rome. J. Volcanol. Geotherm. Res. 107, 185–219.

Kemsk J., Klingel R., Siehl A., Valdivia-Manchego M. R., 2009. From radon hazard to risk prediction - based on geological maps, soil gas and indoor radon measurements in Germany. Environ. Geol. 56, 1269-1279.

Kemski J., Klingel R., Siehl A., Stegemann R., 2005. Radon transfer from ground to houses and prediction of indoor radon in Germany based on geo- logical information. **In:** J.P. McLaughlin, E.S. Simopoulos, and F. Steinhäusler (eds.), The Natural Radiation Environment *VII*, Ser. Radioactivity in the Environment **7**, Seventh International Symposium on the Natural Radiation Environment (NRE-VII), Rhodes, Greece, 20-24 May 2002, Elsevier, 820-832, DOI: 10.1016/ S1569-4860(04)07103-7.

Kemski J., Siehl A., Stegemann R., Valdivia-Manchego M., 2001. Mapping the geogenic radon potential in Germany. Sci. Total Environ. 272, 217-230.

Kemski J., Klingel R., Siehl A., 1996. Classification and mapping of radon affected areas in Germany. Environment International, Supplement 22. 789-798.

Killip I.R., 2005. Radon hazard and risk in Sussex, England and the factors affecting radon dwellings in chalk terrain. Radiat. Prot Dosimetry. 113(1), 99-107

Kleinbaum D.G., Kupper L.L., Muller K.E., Nizam A. 1998. Applied Regression Analysis and Other Multivariate Methods (3rd ed). Belmont: Brooks/Cole.

Kreienbrock L., Kreuzer M., Gerken M., Dingerkus G., Wellmann J., Keller G., Wichmann H.E., 2001. Case control study on lung-cancer and residential radon in Western Germany. Am. J. Epidemiol. 153, 42-52.

Krewski D., Lubin J.H., Zielinski J.M., Alavanja M., Catalan V.S., Field R.W., Klotz J.B., Letourneau E.G., Lynch C.F., Lyon J.I., Sandler D.P., Schoenberg J.B., Steck D.J., Stolwijk J.A., Weinberg C., Wilcox H.B., 2005. Residential radon and risk of lung cancer: a combined analysis of 7 North American case-control studies. Epidemiology 16, 137-145.

Krivoruchko K., Gribov A., Krause E., 2011a. Multivariate Areal Interpolation for Continuous and Count Data. Procedia Environmental Sciences. 3, 14-19.

Krivoruchko, K., 2011b. Spatial Statistical Data Analysis for GIS Users. Redlands, CA: Esri Press, 928 pp.

Jarque C., Bera A., 1987. A test for normality of observations and regression residuals. International Statistical Review. 55(2), 163-172.

Locardi E., 1967. Uranium and thorium in the volcanic processes. Bulletin Volcanologique. 31 (1), 235-260.

Miles J. C. H., Appleton J. D., 2005. Mapping variation in radon potential both between and within geological units. Journal of Radiological Protection. 25, 257-276.

Miles J., 1998a. Mapping radon-prone areas by lognormal modelling of the house radon data. Health Phys. 74(3), 370–378.

Miles J., 1998b. Development of maps of radon-prone areas using radon measurements in houses. J. Hazard. Mater. 61, 53–58.

Milli S. 1997. Depositional setting and high-frequency sequence stratigraphy of the Middle-Upper Pleistocene to Holocene deposits of the Roman Basin. Geologica Romana. 33, 99- 136.

Mitchell A., 2005. The ESRI Guide to GIS Analysis, Spatial Measurements and Statistics. ESRI; Redlands, CA, USA.

Moran P.A.P. (1950). Notes on continuous stochastic phenomena. Biometrika. 37, 17-23.

Nakaya T., Fotheringham S., Charlton M., Brunsdon C., 2009. Semiparametric geographically weighted generalised linear modelling in GWR4.0. Proceedings of Geocomputation 2009 (online), 1-5.

Nazaroff W.W. 1992. Radon transport from soil to air. Rev. Geophys. 30 (2), 137-160

Nazaroff W.W., Moed B.A., Sextro R.G., 1988. Soil as a source of indoor radon: Generation, migration, and entry, in radon and its decay products in indoor air, edited by Nazaroff W.W., and Nero A.V., Jr., 57-112, John Wiley,New York.

Neznal M., Matolin M., Barnet I., Miksova J., 2004. The New Method for Assessing the Radon Risk of Building Sites. Czech Geol. Survey Special Papers, 16. Czech Geol. Survey, Prague, p. 47. http://www.radon-vos.cz/pdf/metodika.pdf (accessed 29.03.13).

Páez A., Farber S., Wheeler D., 2011. A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships. Environ. Plan. A 43, 2992-3010.

Pasculli A., Palermi S., Sarra A., Piacentini T., Miccadei E., 2014. A modelling methodology for the analysis of radon potential based on environmental geology and geographically weighted regression. Environmental Modelling & Software 54, 165-181.

Peccerillo A., 2005. Plio-Quaternary Volcanism in Italy, Petrology, Geochemistry, Geodynamics. Springer-Verlag Berlin Heidelberg, Berlin, Heidelberg.

Piersanti A., Cannelli V., Galli G., 2015. Long term continuous radon monitoring in a seismically active area. ANNALS OF GEOPHYSICS, 58, 4, 2015, S0437; doi:10.4401/ag-6735

Qui X., Wu S., 2013. Global and local regression analysis if factors of American College Test (ACT) scores for public high schools in Missouri. Annals of the Association of American Geographers. 101(1), 63-83.

Shi X., Hoftiezer D.J., Duell E.J., Onega T.L., 2006. Spatial association between residential radon concentration and bedrock types in New Hampshire. Environ. Geol. 51(1), 65-71, DOI: 10.1007/s00254-006-0304-3.

Silverman B. W., 1986. Density Estimation for Statistics and Data Analysis. New York: Chapman and Hall.

Slagle M. 2010. A comparison of spatial statistical methods in a school finance policy context. Journal of Education Finance 35(3), 199-216.

Smethurst M. A., Strand T., Sundal A. V., Rudjord A. L., 2008. Large-scale radon hazard evaluation in the Oslofjord region of Norway utilizing indoor radon concentrations, airborne gamma ray spectrometry and geological mapping. Sci. Total Environ. 15, 379–393.

Spiz K., Moreno J. 1996. A Practical Guide to Groundwater and Solute Transport Modeling. John Wiley & Sons, Inc., New York, N Y.

Stein, M. L. (1999). Interpolation of Spatial Data: Some Theory for Kriging. Springer Series in Statistics. New York: Springer

Szabó K.S., Jordan G., Horváth A., Szabó C., 2013. Dynamics of soil gas radon concentration in a highly permeable soil based on a long-term high temporal resolution observation series. Journal of Environmental Radioactivity. 124, 74-83.

Tanner A.B., 1980. Radon migration in the ground: A supplementary review, in proceedings of Natural Radiation Environment III, (eds) T. F. Gesell and w. M. Lowder, pp. 5-56 U. S. Dep. of Comm. Rep. CONF. 780422, National technical Information Service, Springerfield, Va (1980).

Theil H., 1961. Economic Forecasts and Policy, 2nd Edition, North-Holland, Amsterdam.

Tobler W.R., 1970. A computer movie simulating urban growth in the Detroit region. Econ. Geogr. 46, 234–240. doi: 10.2307/143141.

Tollefsen T., et al. (2014) From the European indoor radon map towards an atlas of natural radiation. Radiat. Prot. Dosim. 162(1–2): 129–134.

Tondeur F., Cinelli G., Dehandschutter B., 2014. Homogenity of geological units with respect to the radon risk in the Walloon region of Belgium. Journal of Environmental Radioactivity. 136, 140-151.

Tuccimei P., Moroni M., Norcia D., 2006. Simultaneous determination of 222Rn and 220Rn exhalation rates from building materials used in central Italy with accumulation chambers and a continuous solid state alpha detector: influence of particle size, humidity and precursors concentration. Applied Radiation and Isotopes. 64, 254-263.

Tung S., Leung J.K.C., Jiao J.J., Wiegand J., Wartenberg, W., 2013. Assessment of soil radon potential in Hong Kong, China, using a 10-point evaluation system. Environ. Earth Sci. 68, 679-689.

United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR), 2000. Sources and Effects of Ionizing Radiation, Report to the General Assembly, New York, United Nations, pp. 1-10.

Vasilyev A.V., Zhukovsky M.V., 2013. Determination of mechanisms and parameters which affect radon entry into a room. Journal of Environmental Radioactivity. 124, 185-190.

Voltaggio A, Masi U, Spadoni M, Zampetti G., 2006. A methodology for assessing the maximum expected radon flux from soils in Northern Latium (Central Italy). Environ. Geochem. Health. 28, 541–551. doi:10.1007/s10653-006-9051-3.

Walia V., Yang T. F., Hong W. L., S. J. Lin, Fu C. C., Wen K. J., Chen C. H., 2009. Geochemical variation of soil–gas composition for fault trace and earthquake precursory studies along the Hsincheng fault in NW Taiwan. Appl. Radiat. Isot. 67, 1855–1863.

Weltner A., Mäkeläinen I., Arvela H., 2002. Radon mapping strategy in Finland. Excerpta medica. International Congress Series 1225. High Levels of Natural Radiation and Radon Areas: Radiation Dose and Health Effects. Elsevier 2002, 63-69.

Webster, R. and M. A. Oliver (2007). Geostatistics for Environmental Scientists (2nd ed.). Statistics in Practice. Chichester: John Wiley & Sons, Ltd.

Wheeler D.C., 2007. Diagnostic tools and a remedial method for collinearity in geographically weighted regression. Environ. Plan. A 39, 2464-2481.

White S.B., Bergsten J.W., Alexander B.V., Rodman N.F., Phillips J.L.. 1992. Indoor 222Rn concentrations in a probability sample of 43,000 houses across 30 states. Health Phys. 62, 41–50.

Winkler R., Ruckerbauer F., Bunzl K., 2001. Radon concentration in soil gas: a comparison of the variability resulting from different methods, spatial het- erogeneity and seasonal fluctuations. Sci. Total Environ. 272, 273-282.

Zafrir H., Barbosa S.M., Malik U., 2012. Differentiation between the effect of temperature and pressure on radon within the subsurface geological media. Radiat. Meas. 49, 39-56. http://dx.doi.org/10.1016/j.radmeas.2012.11.019.

Figure Captions

Figure 1. Conceptual model of soil gas radon dependence from geological and geochemical factors. The global and local regression models include a response variable (i.e., radon in soil gas) and some explanatory variables (i.e., radium content of the outcropping rocks, rock permeability, presence of faults and fractures and the Digital Terrain Model, DTM).

Figure 2. Work flow of data processing and analysis

Figure 3. Geological framework of the study area. (modified from Mele et al., 2006)

Figure 4. Map of the soil gas sample distribution. Soil gas data (published and unpublished data) are provided by ARPA Lazio (2006), and by the Fluid Geochemistry Laboratory of the Earth Science Department of Rome University Sapienza (see Ciotoli et al., 2003; Annunziatellis et al., 2003, 2008, 2010; Beaubien et al., 2003; Bigi et al., 2014). Base Map from ESRI, De Lome, USGS, NPS.

Figure 5. Map of the Hydrogeological Complexes of the Lazio Region (scale 1:25,000) in vector format (Capelli et al., 2012). The color scale indicates the aquifer potential. Base Map from ESRI, De Lome, USGS, NPS.

Figure 6. Map of the main faults of the Lazio Region. Base Map from ESRI, De Lome, USGS, NPS.

Figure 7. Digital Terrain Model (DTM) of the Lazio region at 20 x 20 m resolution, then resampled at the working grid resolution of 1km x 1km. Base Map from ESRI, De Lome, USGS, NPS.

Figure 8. Maps of the Homogeneous Geological Units (HGUs) derived from the Geological Map of Lazio Region (scale 1:25,000) in vector format (Cosentino et al., 2012) (see supplementary material). Base Map from ESRI, De Lome, USGS, NPS.

Figure 9. Map of the radium content of outcropping rocks. The radium values have been obtained by the literature (Castelluccio, 2010; Tuccimei et al., 2006; Voltaggio et al., 2006; Locardi, 1967) and assigned to the HGUs considering.

Figure 10. Map of the rock permeability. The permeability values have been obtained by Spitz and Moreno (1996) and assigned to the HGUs considering the map of the hydrogeolgical complexes.

Figure 11. Map of the fault and fracture density of the Lazio region. The map was constructed by using Kernel Density algorithm to obtain a 1000x1000m fault density map (m/km^2).

Figure 12. Histograms of raw radon data (a); histogram of log-transformed data (b)

Figure 13. Box-plot of radon data within each HGU. The rounded square includes the radon concentration measured in the volcanic areas.

Figure 14. Incremental Spatial Autocorrelation graph for the soil gas radon data. The graph was calculated in order to conceptualize the spatial relationships for the following “Hot Spot Analysis” by using LISA indexes.

Figure 15. Anselin Local Moran and Local Getis-ord G statistics of soil gas radon values in terms of hot spots (areas where locations with high radon values are surrounded by high values: HH), cold spots (low values surrounded by low values: LL) and local outliers (HL).

Figure 16. Analysis of the OLS residuals. The histogram of the regression residual highlights that residuals are non-normally distributed (a); and the Global Moran’s I indicates that are spatially autocorrelated (b).

Figure 17. The map of the Local R^2 values indicates where GWR predicts well, and where it predicts poorly; it may provide clues about important variables that may be missing from the regression model.

Figure 18. The histogram of the standardized residuals of the GWR a Gaussian distribution confirming the good performance of the model (a), and the Morans’I autocorrelation test highlights that residuals are not spatially autocorrelated (b).

Figure 19. Variogram map (a) highlights the presence of anisotropy in the GWR predicted values. The calculated experimental variogram (b)was then modelled along this direction by using an exponential model with a nugget value of 0.07, a sill of 1, as well as a maximum range of anisotropy of 50000 m and a minimum range of anisotropy of 25000 m

Figure 20. Cross-validation graph (a) provides the fitting of the exponential model with a Root-Mean-Square Standardised of 0.790. The graphs of the standardised residuals (b) indicates a Gaussian distribution of the errors of the prediction.

Figure 21. Final GRP map obtained by constructing a local regression model by GWR.

Figure1

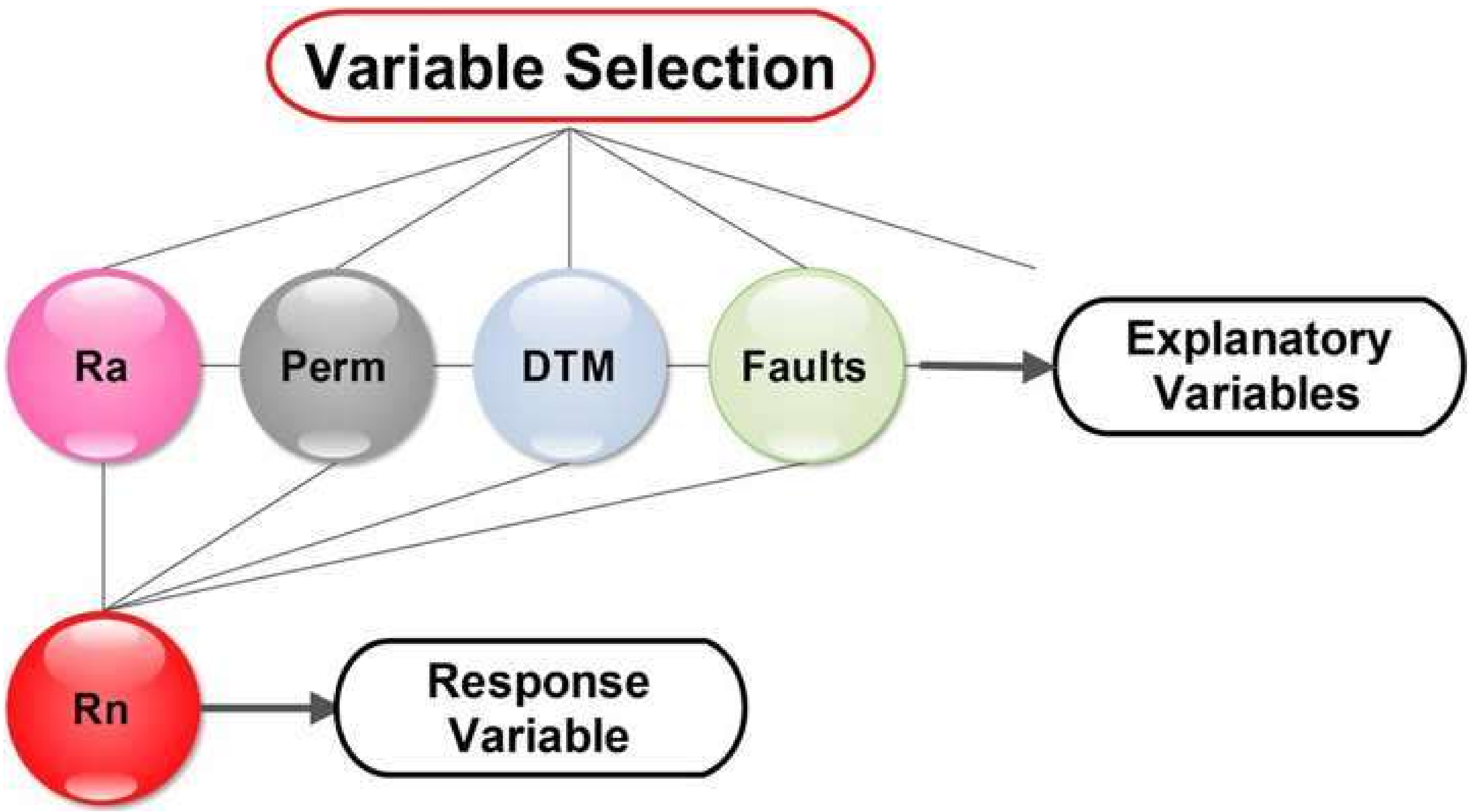

Variable Selection
Ra
Perm
DTM
Faults
Explanatory Variables
Rn
Response Variable

Figure2

Click here to download high resolution image

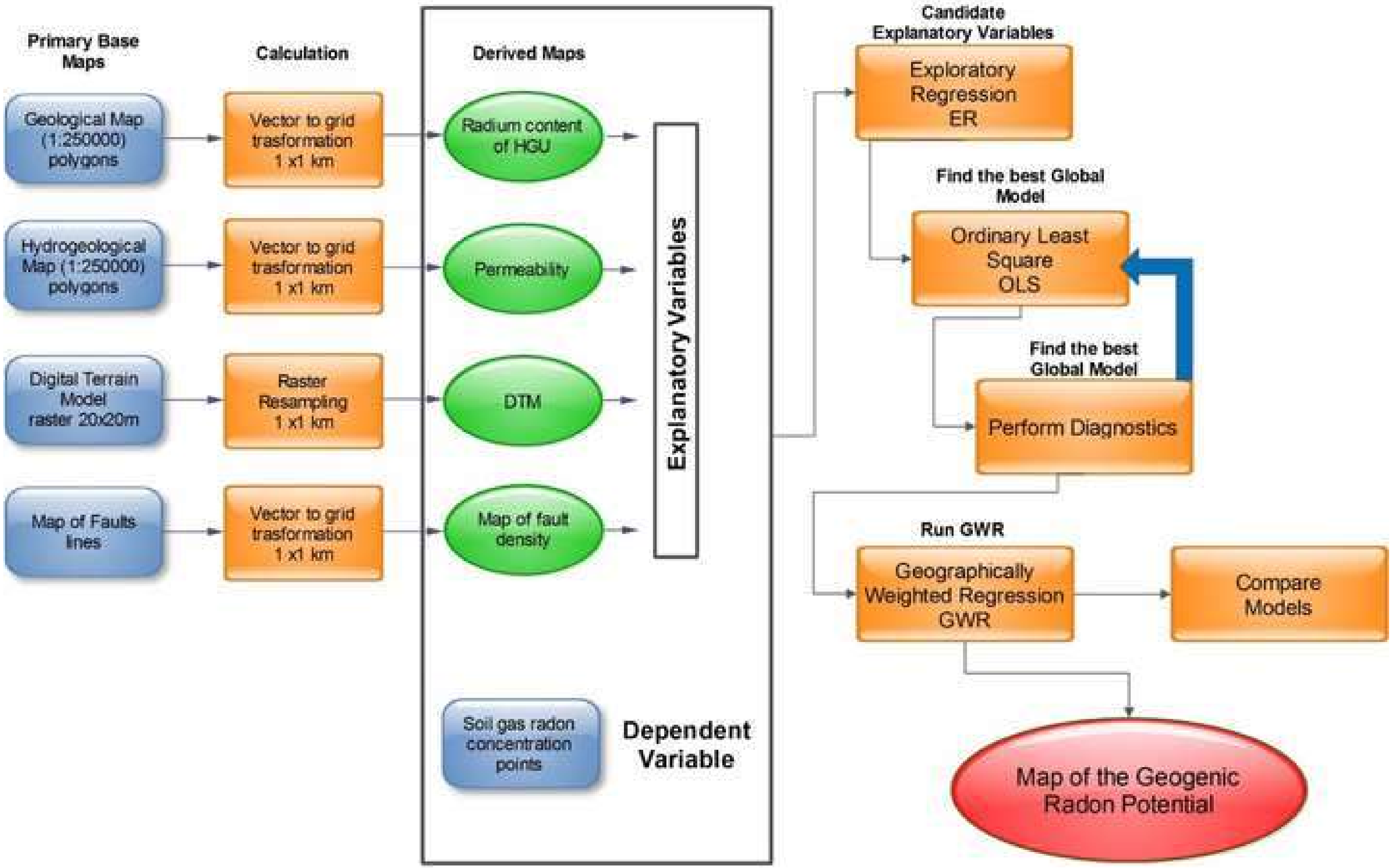
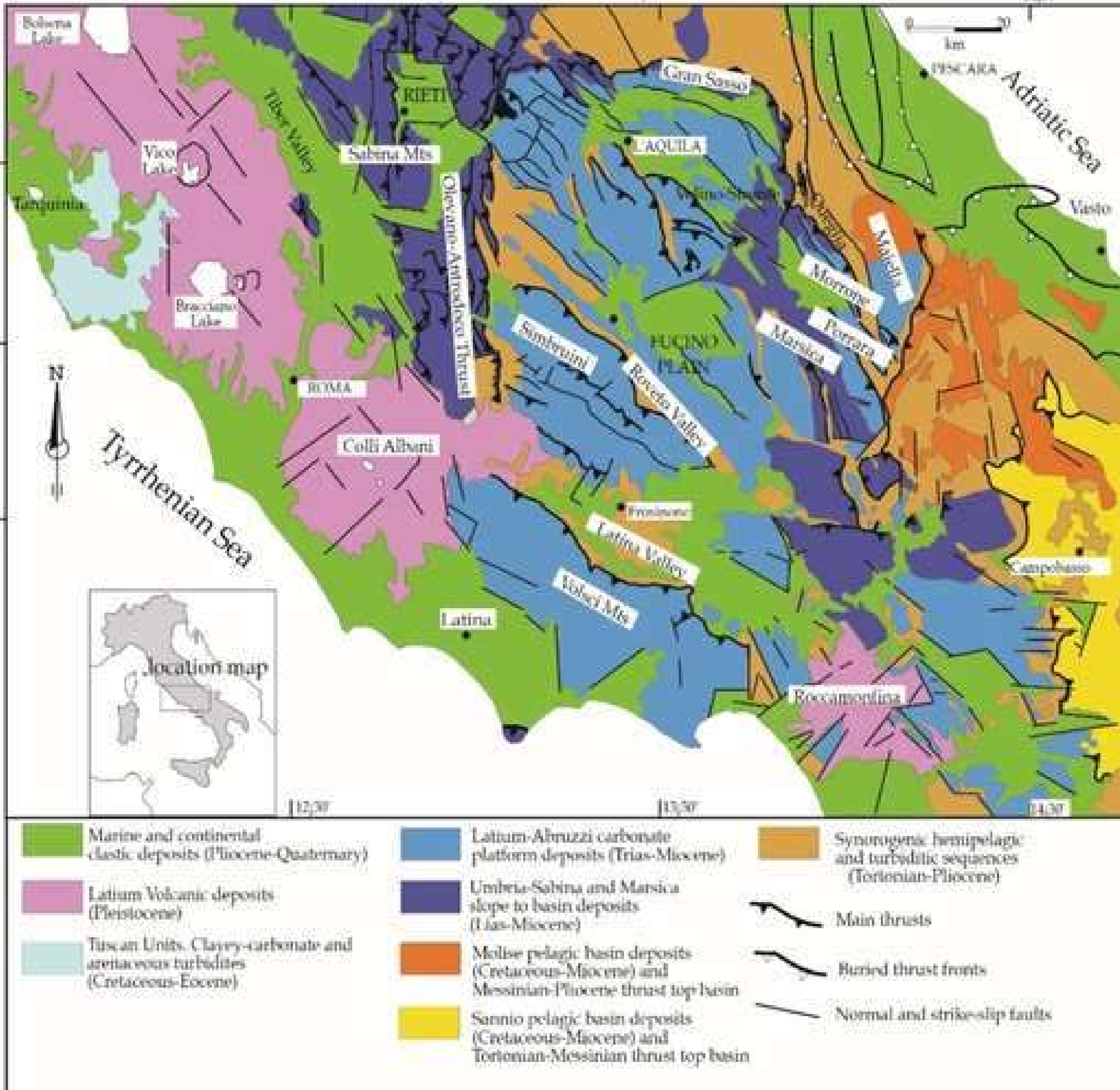


Figure3

Click here to download high resolution image



from Mele et alii (2006)

Figure4

Click here to download high resolution image

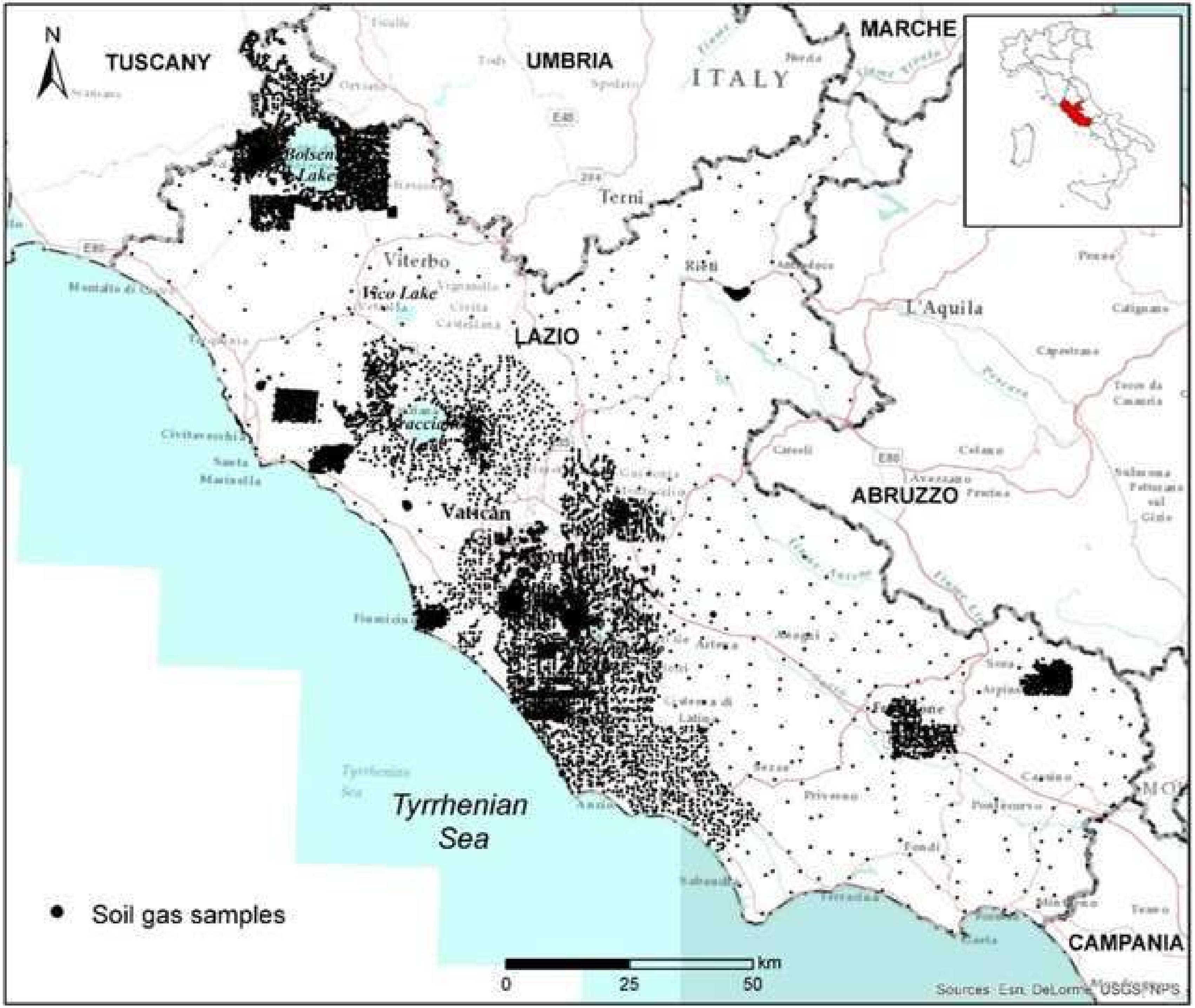


Figure5

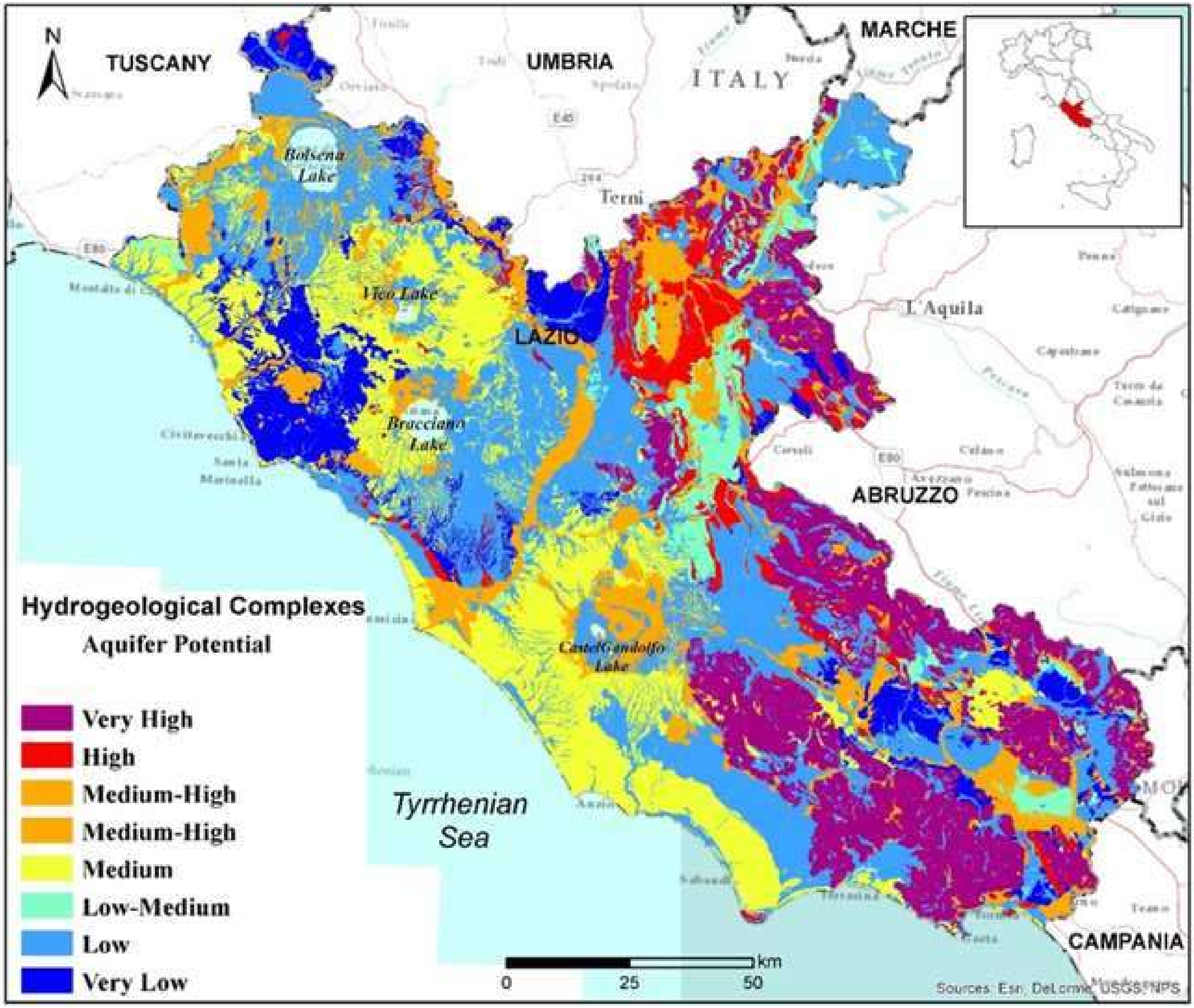

N
TUSCANY
UMBRIA
ITALY
MARCHE
Terni
L'Aquila
LAZIO
ABRUZZO
CAMPANIA
Bolsena Lake
Vico Lake
Bracciano Lake
CastelGandolfo Lake
Tyrrhenian Sea
Hydrogeological Complexes
Aquifer Potential
Very High
High
Medium-High
Medium-High
Medium
Low-Medium
Low
Very Low
0
25
50
km
Sources: Esri, DeLorme, USGS, NPS

Figure6

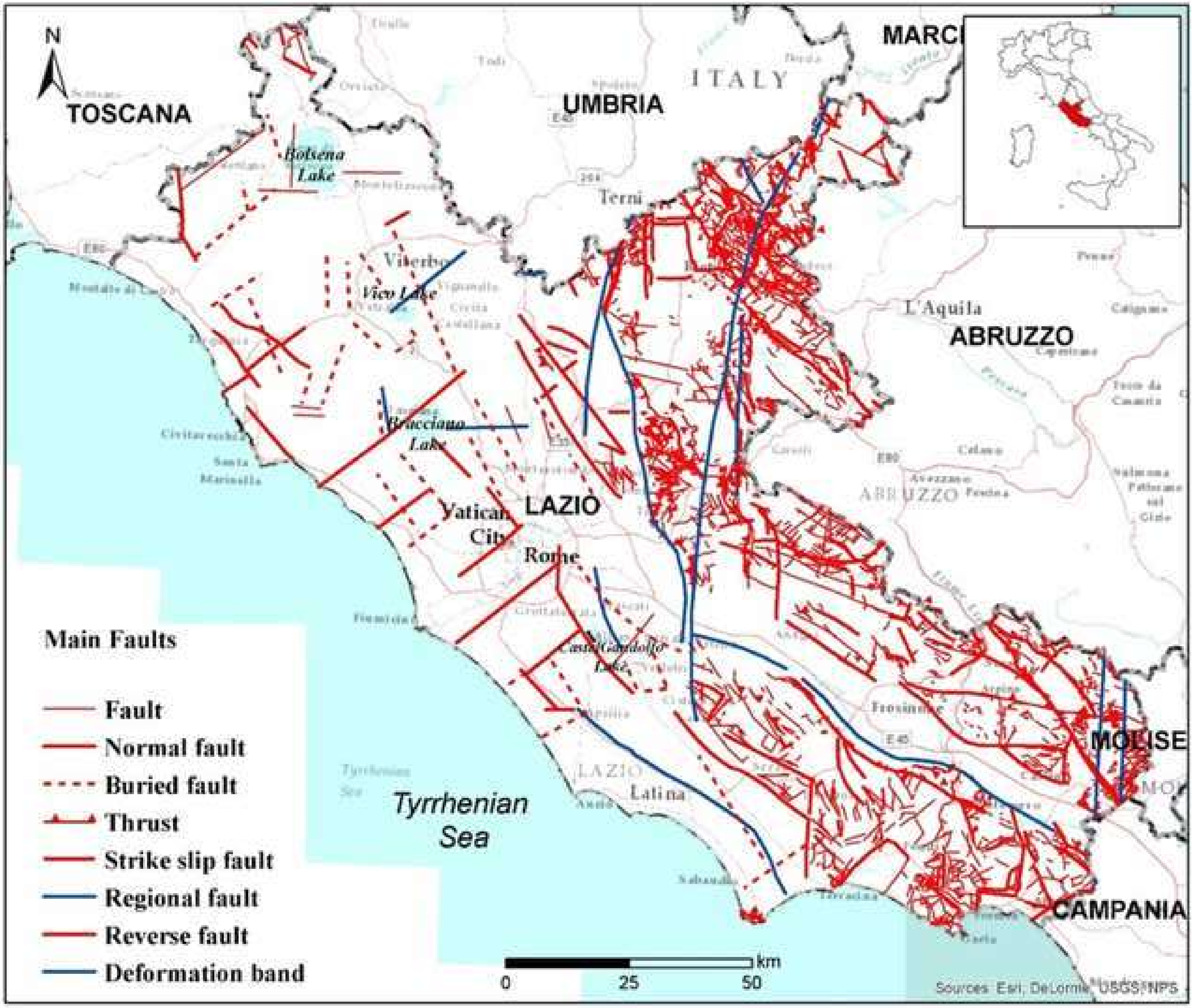
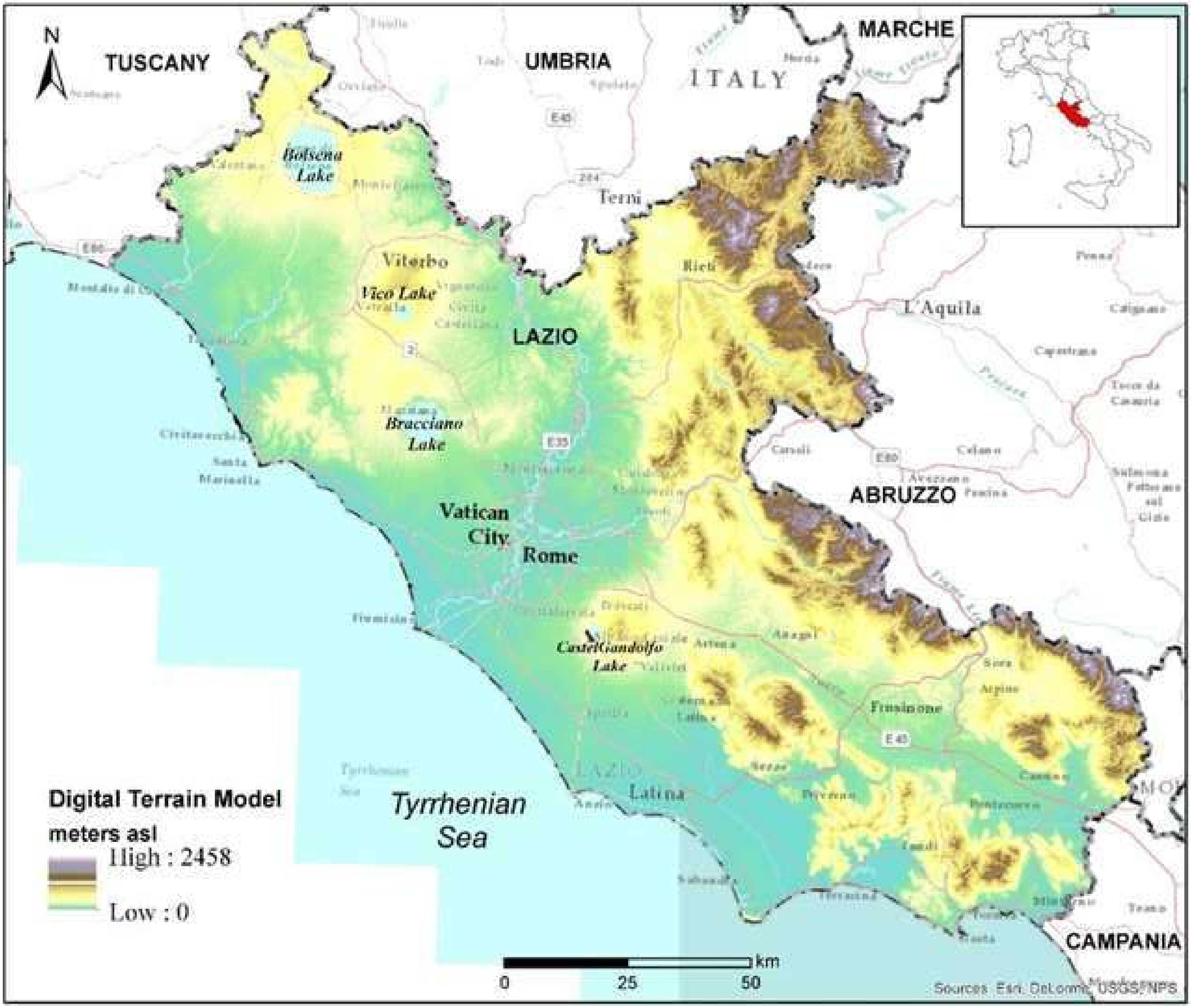

N
TOSCANA
UMBRIA
ITALY
MARC
ABRUZZO
L'Aquila
Terni
Viterbo
Bolsena Lake
Vico Lake
Bracciano Lake
Vatican City
LAZIO
Rome
Castel Gandolfo Lake
Latina
Frosinone
MOLISE
CAMPANIA
Tyrrhenian Sea
Main Faults
Fault
Normal fault
Buried fault
Thrust
Strike slip fault
Regional fault
Reverse fault
Deformation band
0
25
50
km
Sources: Esri, DeLorme, USGS, NPS

Figure7



N
TUSCANY
UMBRIA
ITALY
MARCHE
Bolsena Lake
Terni
Viterbo
Vico Lake
LAZIO
Rieti
L'Aquila
ABRUZZO
Bracciano Lake
Vatican City
Rome
CastelGandolfo Lake
Frosinone
Latina
Tyrrhenian Sea
CAMPANIA
Digital Terrain Model
meters asl
High : 2458
Low : 0
0
25
50
km
E40
E35
E80
Sources: Esri, DeLorme, USGS, NPS

Figure8


N
TOSCANA
UMBRIA
ITALY
MARC
ABRUZZO
L'Aquila
Terni
Bolsena Lake
Vico Lake
Bracciano Lake
Vatican City
LAZIO
Rome
CastelGandolfo Lake
Latina
Frosinone
MOLISE
CAMPANIA
Tyrrhenian Sea
Homogeneous Geological Units
HGU
Alluvial
Continental Deposits
Flysch
Marine Deposits
Pelagic Slope Carbonate
Platform Carbonate
Albani Volcanics
Sabatini Volcanics
Tolfa Volcanics
Vico Volcanics
Volsini Volcanics
0
25
50
km
Sources: Esri, DeLorme, USGS, NPS

Figure9

Click here to download high resolution image

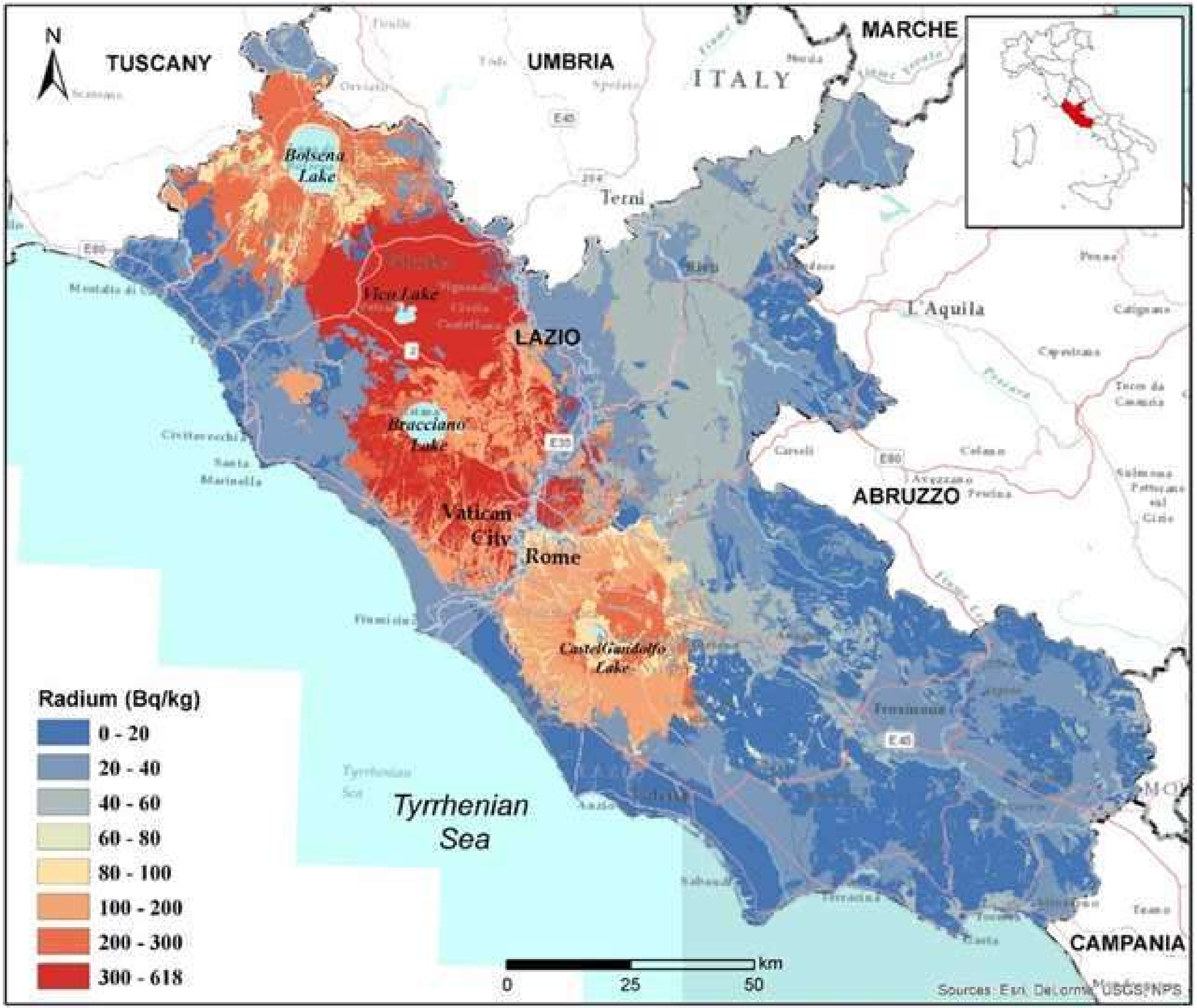


Figure10

Click here to download high resolution image

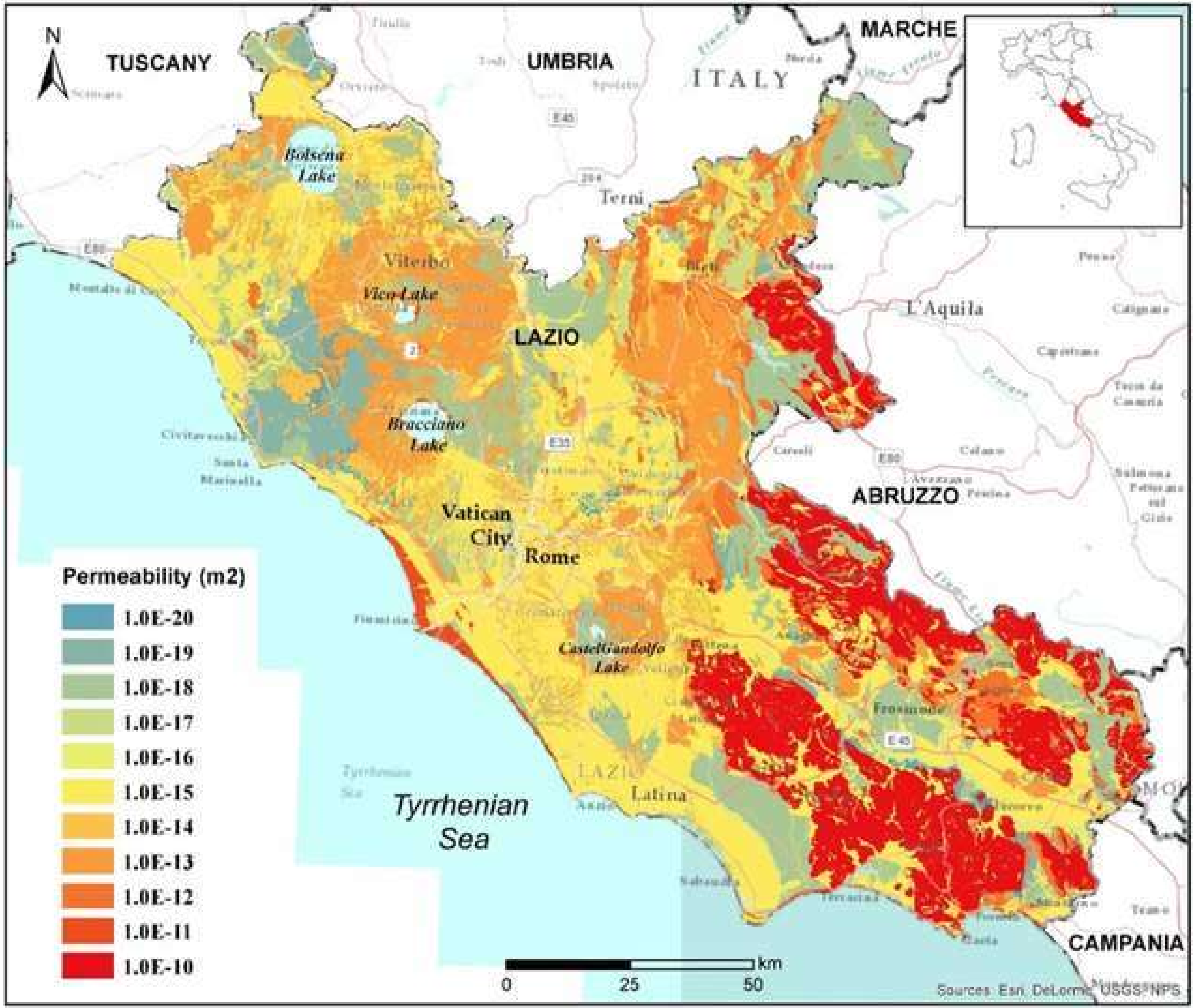


Figure11

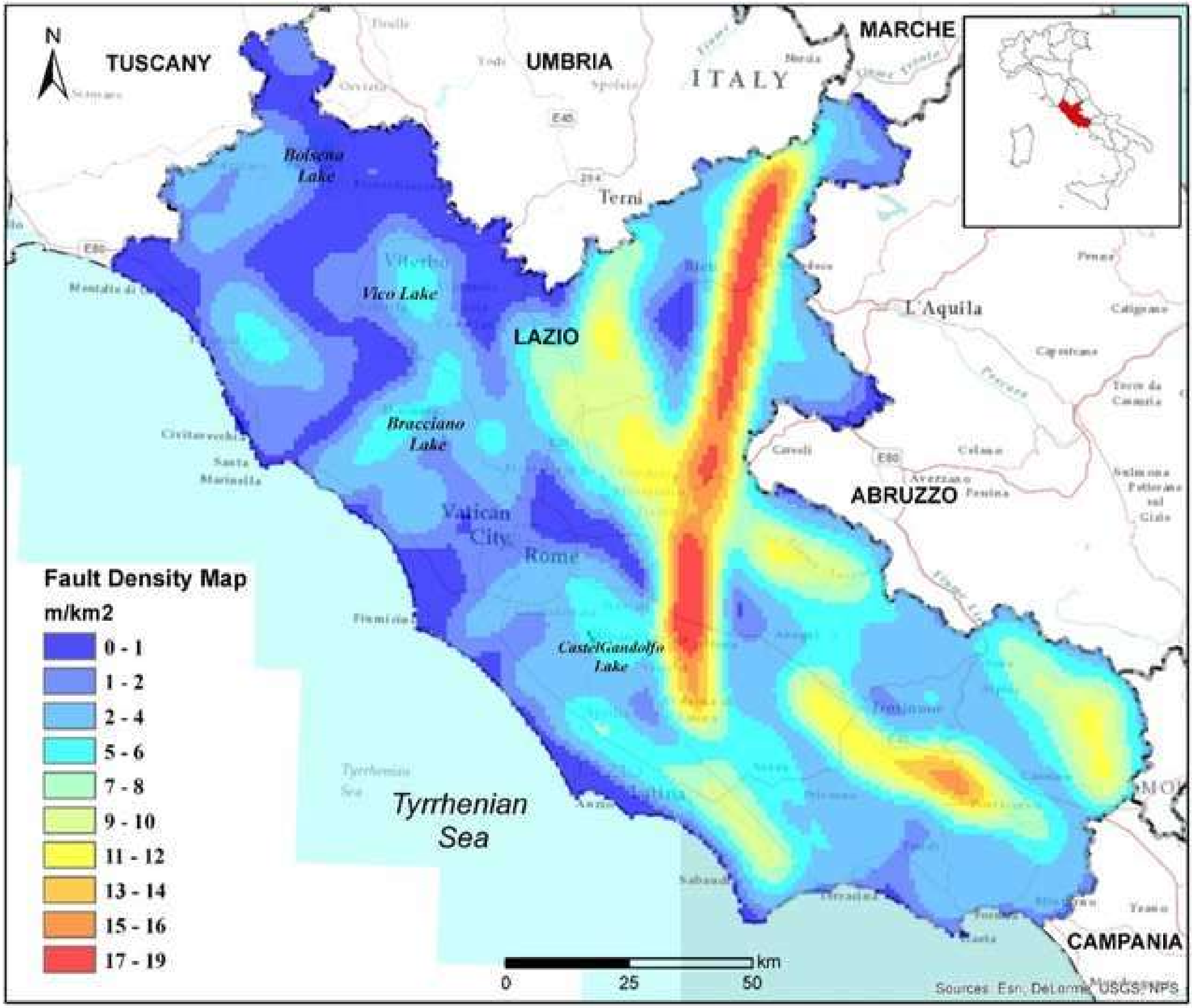

N
TUSCANY
UMBRIA
ITALY
MARCHE
LAZIO
ABRUZZO
CAMPANIA
Terni
L'Aquila
Bolsena Lake
Vico Lake
Bracciano Lake
Vatican City
Rome
CastelGandolfo Lake
Tyrrhenian Sea
Fault Density Map
m/km2
0 - 1
1 - 2
2 - 4
5 - 6
7 - 8
9 - 10
11 - 12
13 - 14
15 - 16
17 - 19
0
25
50
km
Sources: Esri, DeLorme, USGS, NPS

Figure12

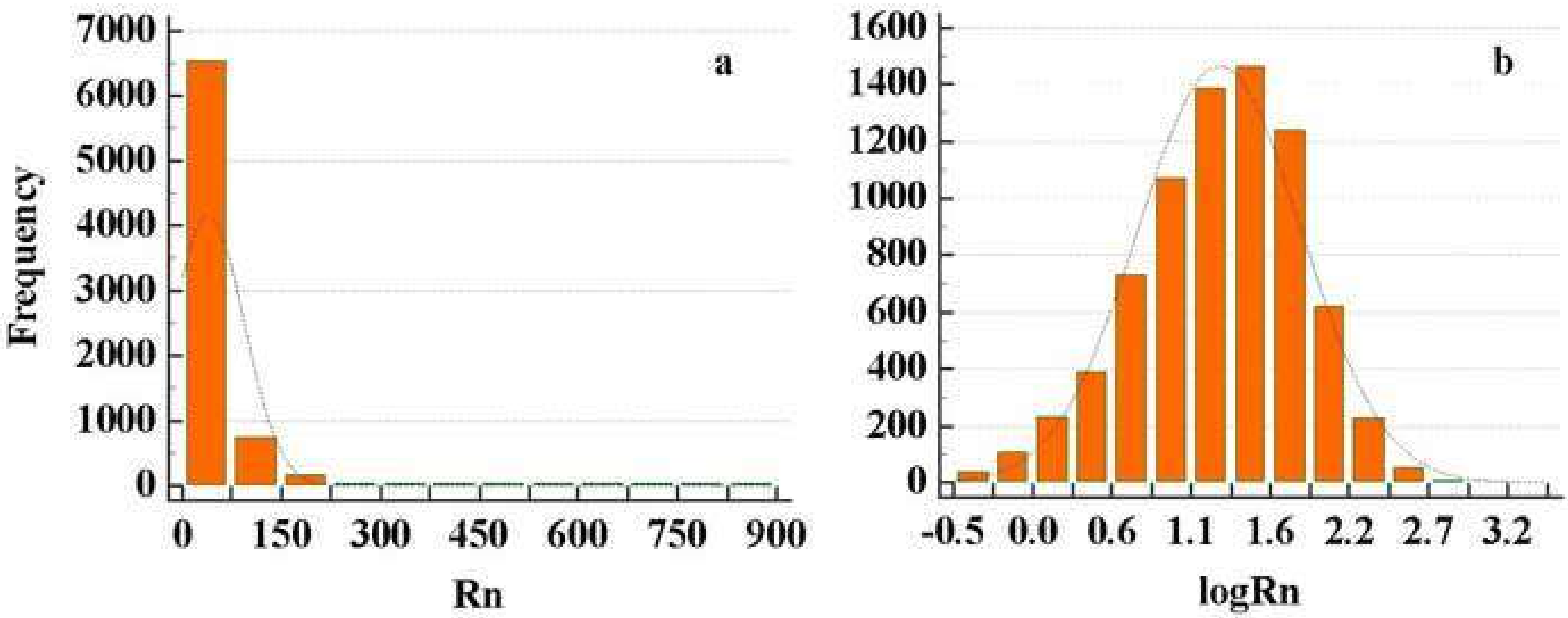

Frequency
7000
6000
5000
4000
3000
2000
1000
0
a
0
150
300
450
600
750
900
Rn
1600
1400
1200
1000
800
600
400
200
0
b
-0.5
0.0
0.6
1.1
1.6
2.2
2.7
3.2
logRn

Figure13

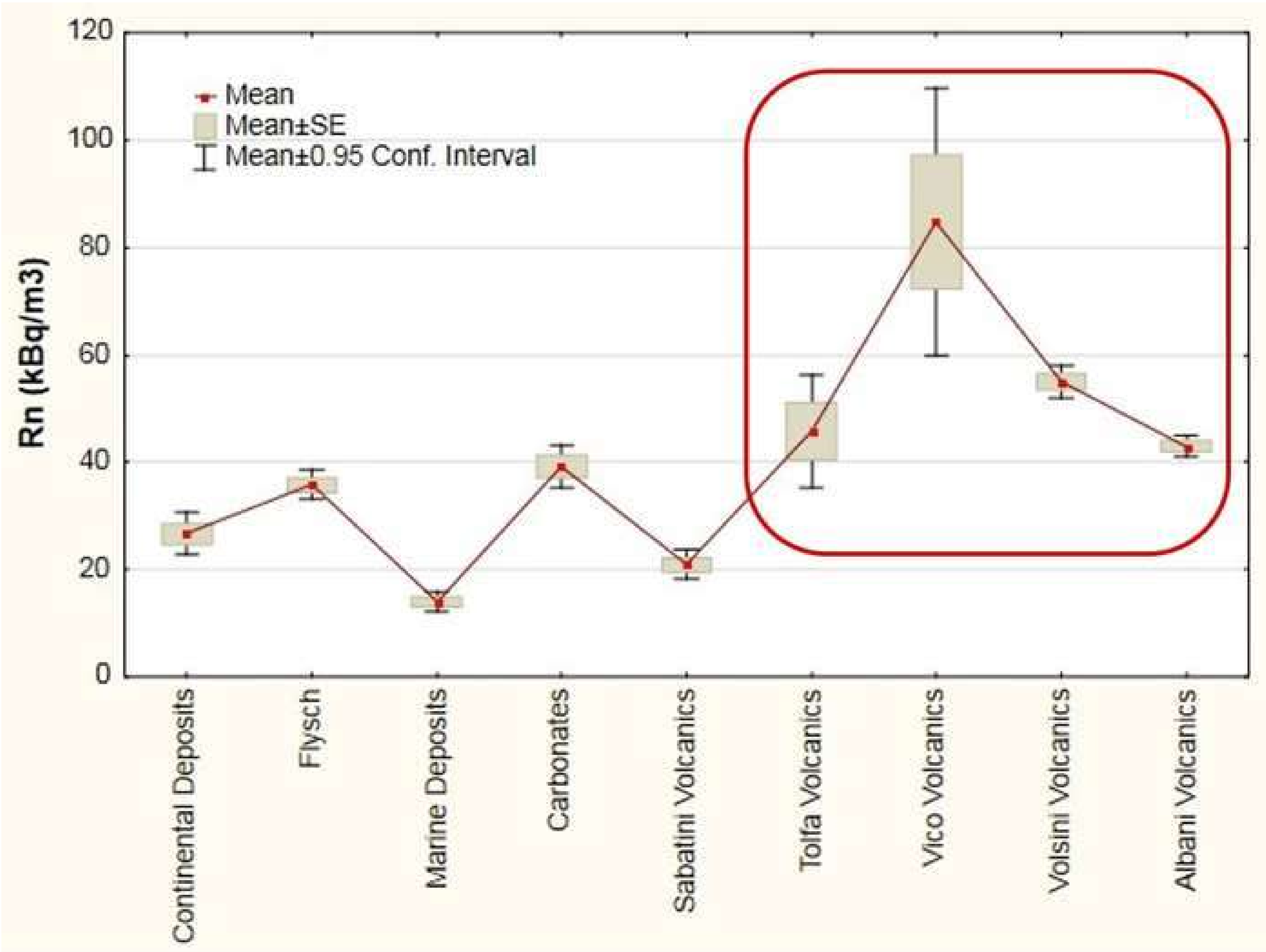

Mean
Mean±SE
Mean±0.95 Conf. Interval
Rn (kBq/m3)
0
20
40
60
80
100
120
Continental Deposits
Flysch
Marine Deposits
Carbonates
Sabatini Volcanics
Tolfa Volcanics
Vico Volcanics
Volsini Volcanics
Albani Volcanics

Figure14

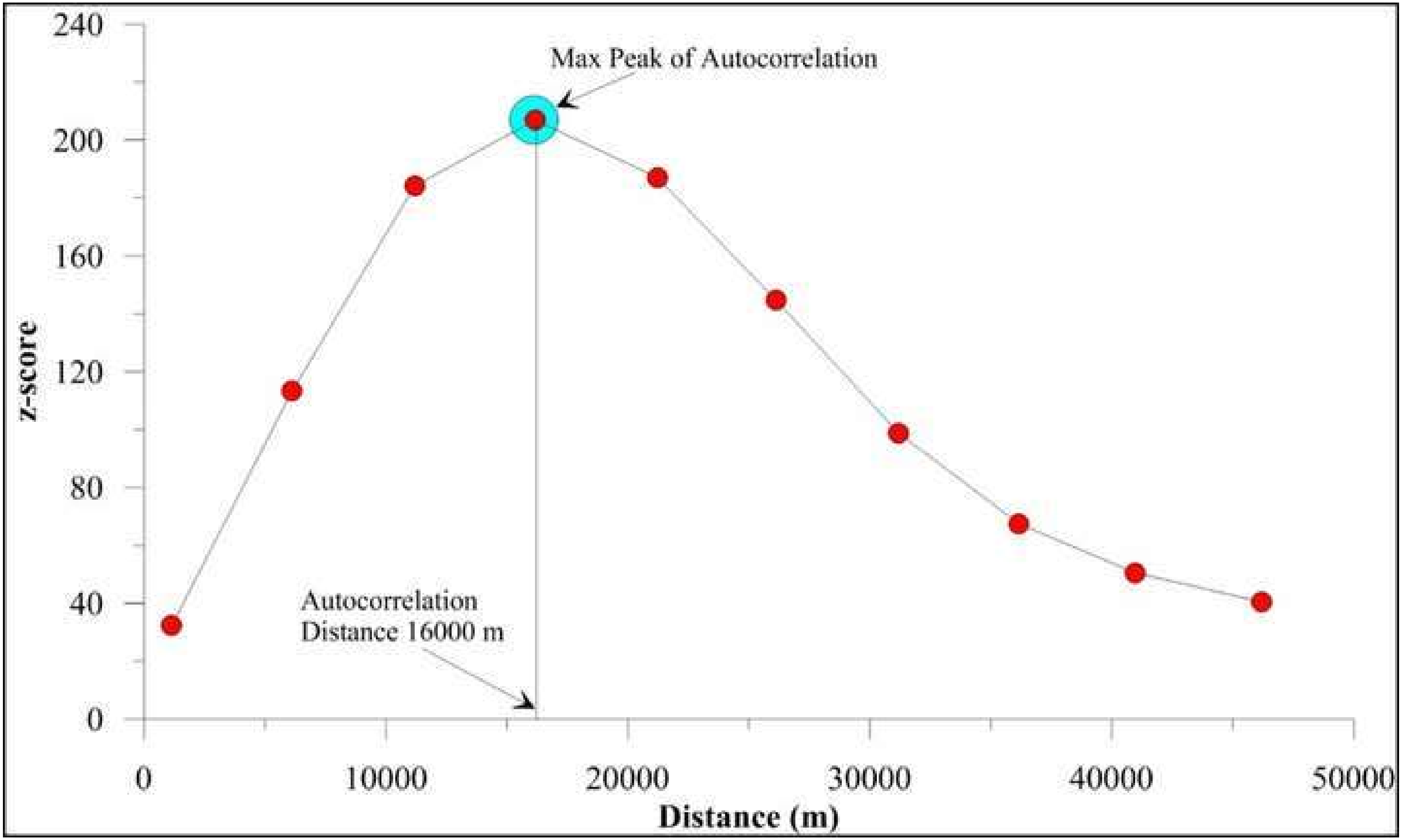

Max Peak of Autocorrelation
Autocorrelation
Distance 16000 m
z-score
240
200
160
120
80
40
0
0
10000
20000
30000
40000
50000
Distance (m)

Figure15

Click here to download high resolution image

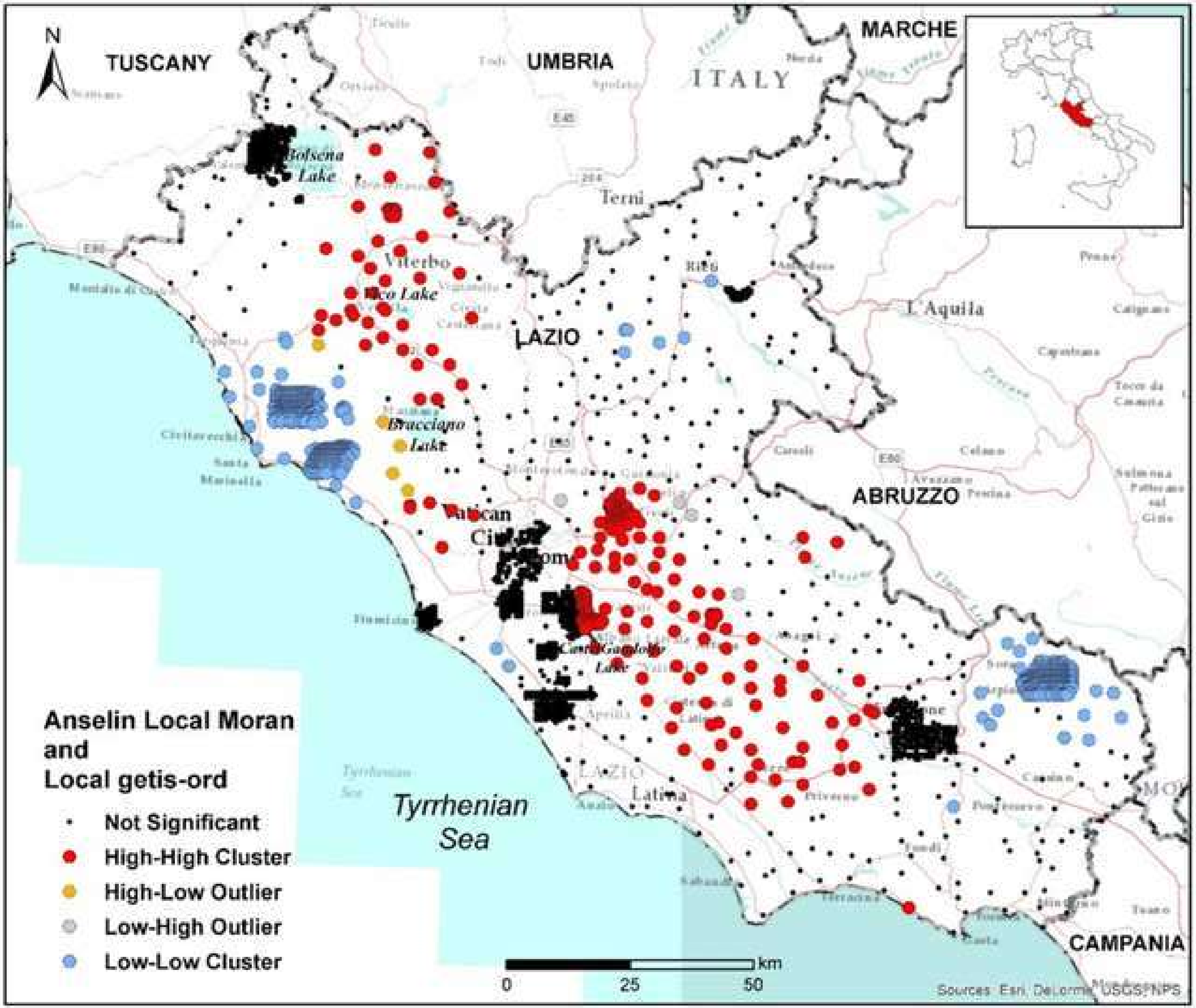
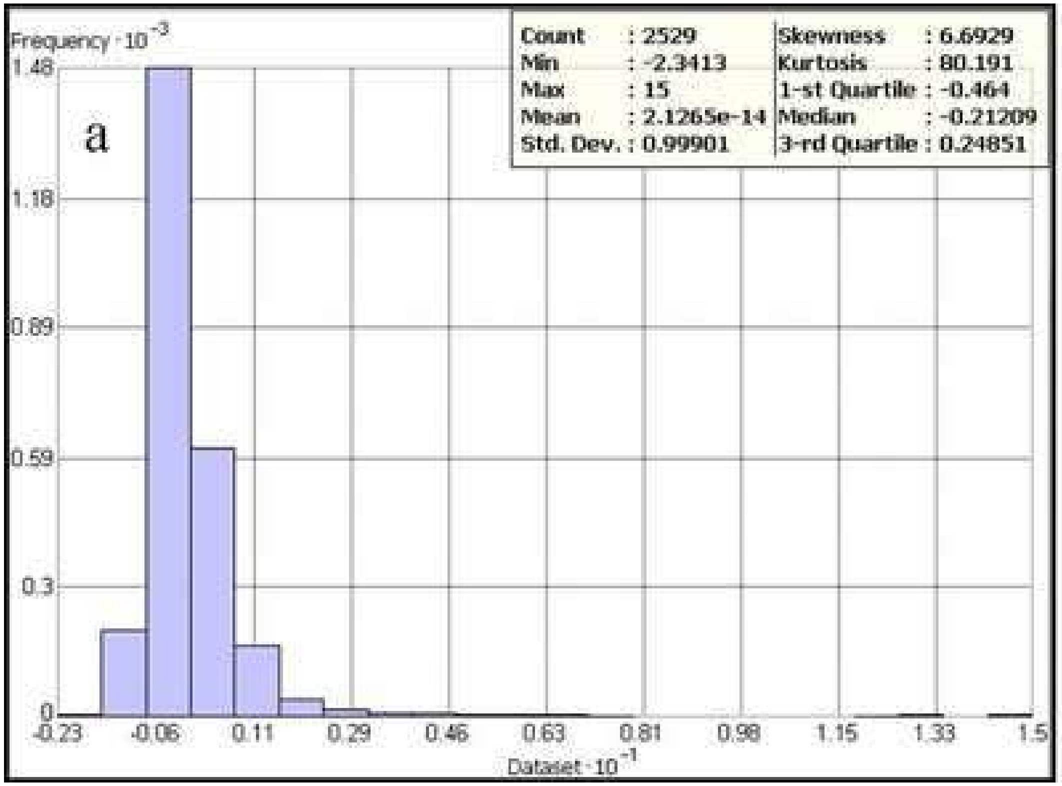


Figure16

Click here to download high resolution image



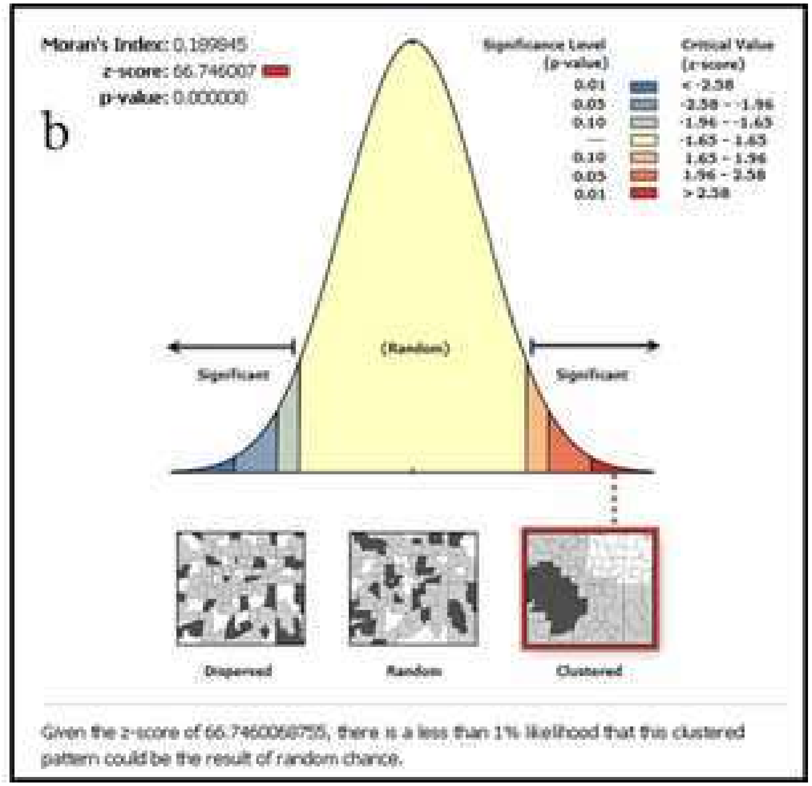


Figure17

Click here to download high resolution image

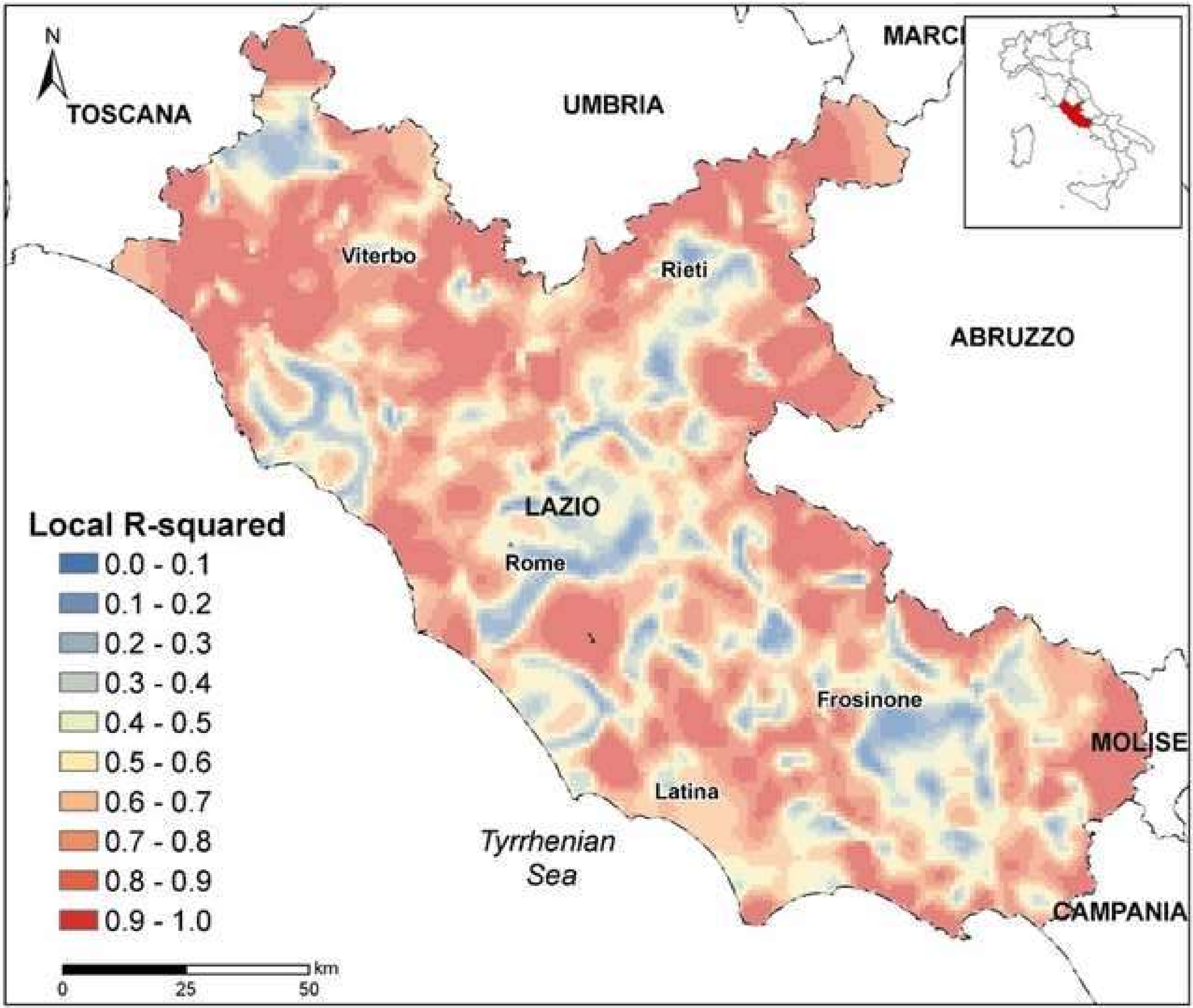


Figure18

Click here to download high resolution image

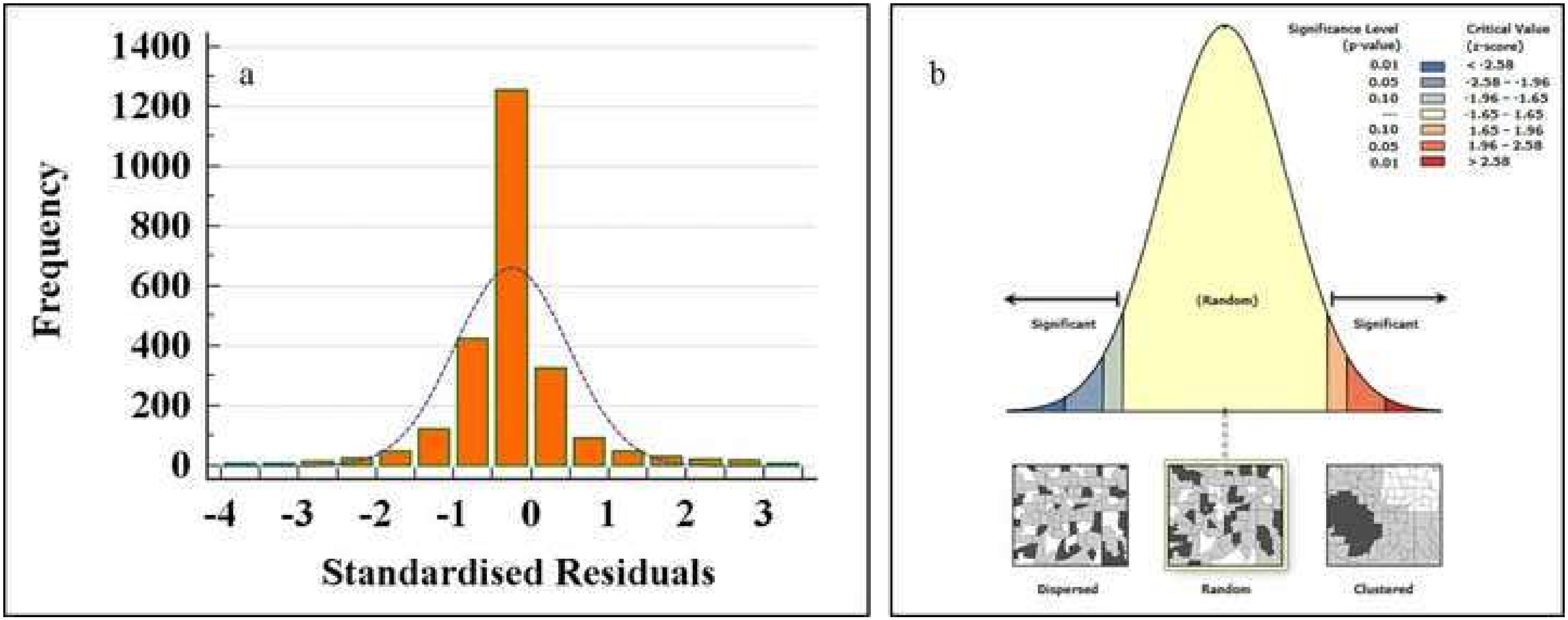


Figure19

Click here to download high resolution image

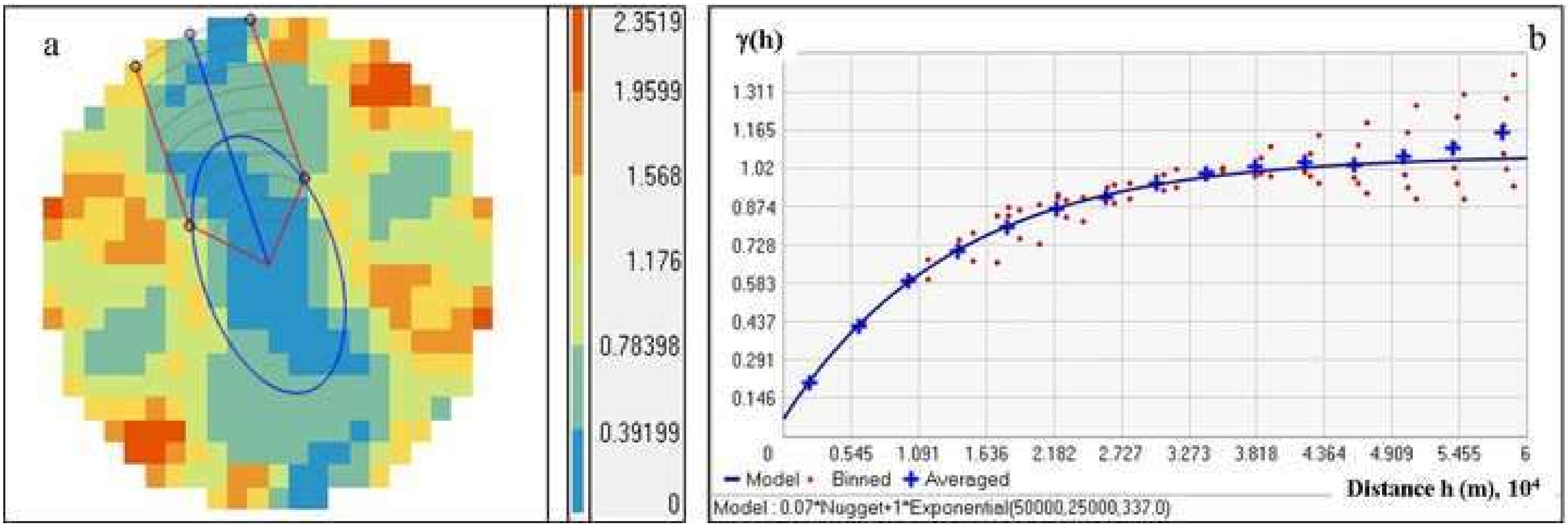


Figure20

Click here to download high resolution image

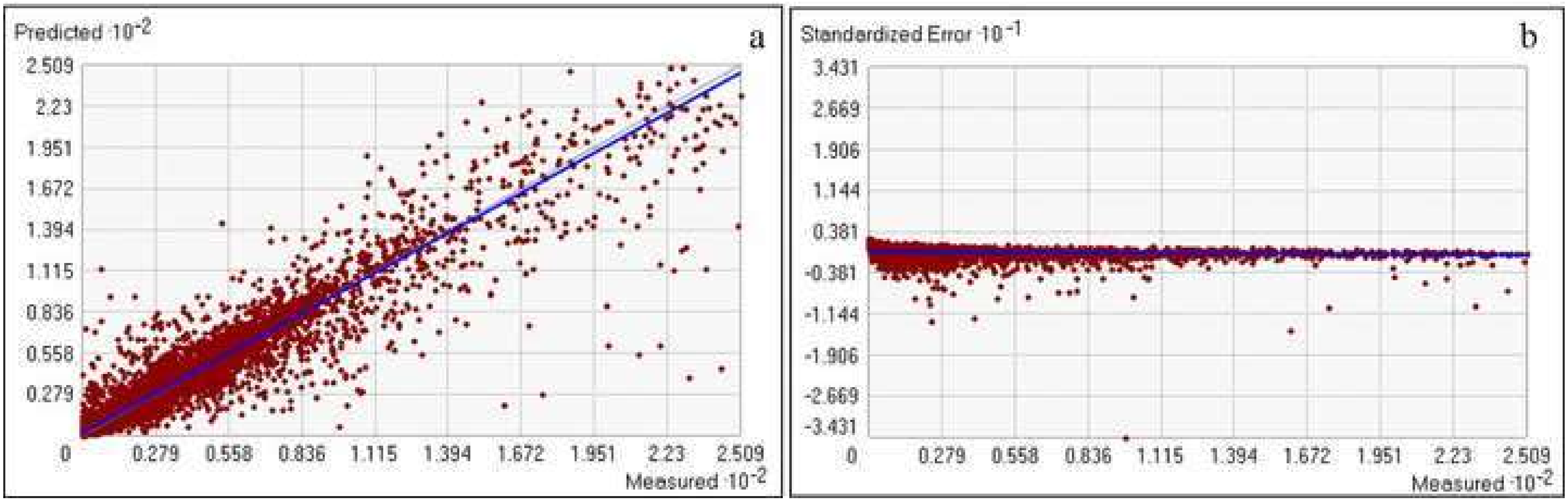


Figure21

Click here to download high resolution image

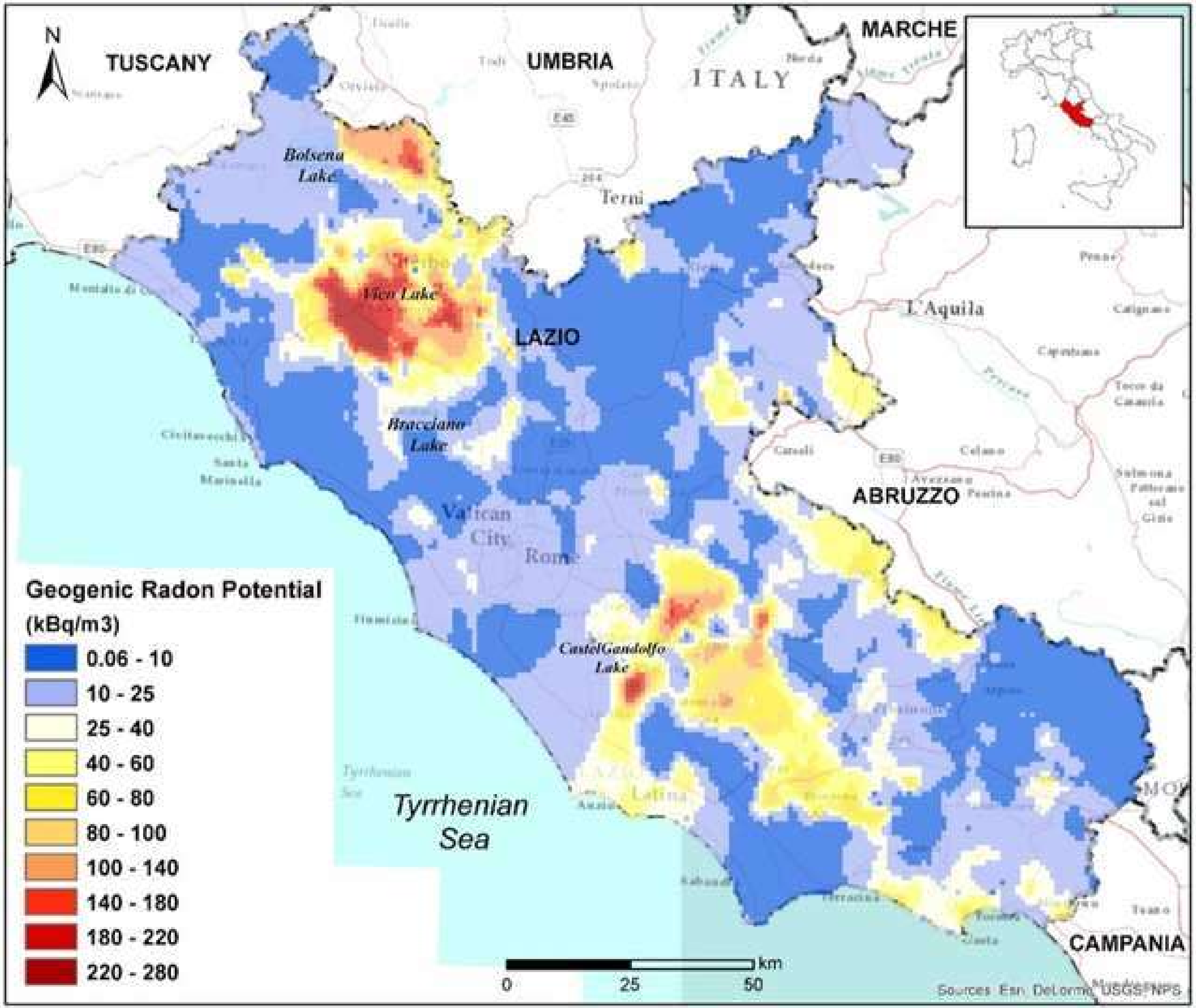


Table 1. Main statistic of the raw radon values, as well as the other parameters to be considered under this study. The statistics of the other variables was calculated considering the 12911 point of the 1km x 1km grid.

Variable	N	Min	Max	Mean	St.dev.
^{222}Rn (kBq m^{-3})	7625	0.1	828	38.65	54.40
^{226}Ra (Bq/kg)	12877	6.5	617	130	137
Permeability (m^2)	12911	10^{-20}	10^{-10}	10^{-15}	-
Fault density (m/km^{-2})	12911	10-5	19160	4151	3643
DTM (m)	12911	0	2385	397	394

Table 2. Detailed statistics of soil gas radon data collected in the Lazio region. AM , arithmetic mean; GM, geometric mean; Std.Dev, standard deviation

N	AM	SE	GM	Median	Min	Max	Stdev
7610	33.83 (37.60-40.06)	0.62	19.51 (18.96- 20.08)	21.46 (20-77-22.57)	0.37	828	54.48

Table 3. Main statistics of radon data within the HGUs. AM , arithmetic mean; GM, geometric mean; Std.Dev, standard deviation

HGU	N	AM	GM	Median	Min	Max	Std.Dev.
Continental deposits	935	26.71	11.78	11.84	0.37	828.00	60.65
Flysch	1546	35.83	17.96	18.50	0.44	592.37	54.64
Marine Deposits	377	14.00	8.21	8.00	0.37	134.64	17.00
Carbonate	1159	39.24	15.65	17.76	0.37	797.00	68.67
Sabatini Volcanics	305	20.95	12.70	14.06	0.37	240.32	24.13
Tolfa Volcanics	53	45.76	32.30	34.04	4.81	200.54	37.88
Vico Volcanics	74	**84.81**	37.93	41.50	2.00	391.00	107.52
Volsini Volcanics	1163	54.93	34.05	38.85	0.37	480.26	53.37
Alban Hill Volcanics	1967	42.98	26.63	31.08	0.37	444.00	42.13

Table 4. Morans'I and Getis-ord tests for spatial autocorrelation of the selected variables.

Variable	Morans' I	z-score	p-value	Getis-ord G	z-score	p-value
Soil_Rn	0.56	204.00	0.000	0.00	13.20	0.000
Fault	0.54	188.80	0.000	0.00	10.57	0.000
Radium	0.35	219.90	0.000	0.15	20.95	0.000
DEM	0.63	218.80	0.000	0.11	2.75	0.005
Permeability	0.04	25.10	0.000	0.00	3.31	0.001

Table 5. Exploratory regression results.

Model Parameter	Cutoff	Trials	# Passed	% Passed	Result
Adjusted R2	> 0.5	15	0	0	0.15
VIF	< 7.5	15	15	15	1.16
Jarqe-Bera p-value	> 0.1	15	0	0	0.00
Spatial Autocorrelation p-value (Global Morans'I)	> 0.1	15	0	0	0.00
AICc					19658

Table 6. Variable significance in the model (** = 0.05, *** = 0.01)

Variable	% significant	VIF	Violation
Radium	100.00***	1.15	0
Perm	100.00**	1.08	0
Fault	78.00***	1.16	0
DTM	0.00	1.08	0

Table 7. Main parameters of the Ordinary Least Squared regression model. DoF=Degree of Freedom; * indicates significance

Parameter	Value	Prob	p-value
N. Obs.			
Multiple R2	0.154		
Adjusted R2	0.152		
AICc	19512		
Joint-F-statistic	92.03	Prob (>F), (5, 2523) DoF	0.000*
Joint Wald statistic	168.15	Prob (> χ^2), (5) DoF	0.000*
Koenker (BP) Statistic	120.69	Prob (> χ^2), (5) DoF	0.000*
Jarque-Bera Statistic	646757.82	Prob (>F), (2) DoF	0.000*

Table 8. Coefficient diagnostic table of the OLS global regression model (* = statistical significant).

Variable	Coeff	StdErr	Robust t	Robust Prob	VIF
Intercept	5.1268	1.4974	3.5883	0.0003*	--------
Fault	0.0003	0.0000	3.2495	0.0011*	1.16
DTM	**-0.0012**	**0.0012**	**-0.7051**	**0.4807**	**1.08**
Radium	0.0411	0.0026	6.0615	0.0000*	1.15
Perm	0.2187	0.1038	2.0085	0.0446*	1.08

Table 9. Main parameters of the GWR local model.

Parameter	Value
Bandwidth	4000
Sigma	1.63
AICc	1231.48
R2	0.950
Adj. R2	0.935

Table 10. Main statistics of the coefficients of the explanatory variables for the GWR model

Variable	Mean	Robust STD	Min	Max	Median	IQR
Intercept	13.49	7.71	-16.17	607.88	11.41	10.40
Fault	1.32	4.67	-97.18	503.97	0.35	6.30
Radium	1.61	1.12	-67.71	83.58	0.22	1.51
Permeability	-0.39	0.72	-42.87	121.28	-0.13	0.97

Table 11. ANOVA test for the comparison of the OLS and GWR regression models.

ANOVA	SS	DF	MS	F
Global Residuals	350288.448	2525.000		
GWR Improvement	326677.861	395.661	825.651	
GWR Residuals	23610.587	2129.339	11.088	74.462

Supplementary Material

Click here to download Supplementary Material: SupplementaryMaterial_Geological Map of Lazio.jpg

Supplementary Material

Click here to download Supplementary Material: SupplementaryMaterial_Geological Map of Lazio_Legend.jpg