NUKLEONIKA 2020;65(2):99–104
doi: 10.2478/nuka-2020-0015

ORIGINAL PAPER

A new geostatistical tool for the analysis of the geographical variability of the indoor radon activity

Filomena Loffredo,
Antonio Scala,
Guido Maria Adinolfi,
Federica Savino,
Maria Quarto

Abstract. The population is continuously exposed to a background level of ionizing radiation due to the natural radioactivity and, in particular, with radon (^{222}Rn). Radon gas has been classified as the second leading cause of lung cancer after tobacco smoke [1]. In the confined environment, radon concentration can reach harmful level and vary accordingly to many factors. Since the primary source of radon in dwellings is the subsurface, the risk assessment and reduction cannot disregard the identification of the local geology and the environmental predisposing factors. In this article, we propose a new methodology, based on the computation of the Gini coefficients at different spatial scales, to estimate the spatial correlation and the geographical variability of radon concentrations. This variability can be interpreted as a signature of the different subsurface geological conditions. The Gini coefficient computation is a statistical tool widely used to determine the degree of inhomogeneity of different kinds of distributions. We generated several simulated radon distributions, and the proposed tool has been validated by comparing the variograms based on the semi-variance computation with those ones based on the Gini coefficient. The Gini coefficient variogram is shown to be a good estimator of the inhomogeneity degree of radon concentration. Indeed, it allows to better constrain the critical distance below which the radon geological source can be considered as uniform at least for the investigated length scales of variability; it also better discriminates the fluctuations due to the environmental predisposing factors from those ones due to the random spatially uncorrelated noise.

Keywords: Geographical variability • Geostatistics • Gini coefficient • Lorenz curve • Radon concentration • Variogram

F. Loffredo✉, M. Quarto
Advanced Biomedical Science Department
University of Naples, Federico II
Corso Umberto I, 40-80138 Naples, Italy
and National Institute of Nuclear Physics (INFN)
Strada Comunale Cinthia, 80126, Naples, Italy
E-mail: filomena.loffredo@unina.it

F. Savino
Advanced Biomedical Science Department
University of Naples, Federico II
Corso Umberto I, 40-80138 Naples, Italy

A. Scala, G. M. Adinolfi
Department of Physics, "E. Pancini"
University of Naples, Federico II
Corso Umberto I, 40-80138 Naples, Italy

Received: 2 December 2019
Accepted: 16 January 2020

Introduction

Several epidemiological studies have confirmed that the prolonged exposure to radon increases the risk of developing lung cancer for humans. In this regard, in recent years, the EU has consolidated the directive on basic safety standards (2013/59/Euratom), whose purpose is to reduce the number of cases of radon-induced lung cancer.

Radon is a natural radioactive noble gas produced from the decay of radium (^{226}Ra, itself a decay product of ^{238}U), and it is widely distributed in the geological environment in many kinds of rocks and soils. Radon and its progenies in air constitute the major natural exposure source to ionizing radiation for humans [2] because it can accumulate in enclosed space. In general, the outdoor radon concentrations are very low, while in confined environments, due to low ventilation, the concentration harmfully increases leading to a considerable health risk [3, 4]. Moreover, in confined environments, other factors can contribute to increase the radon concentration, such as the emanation from building materials, the floor, and the degassing from tap water having a groundwater origin [5–7].

To apply efficient locally based risk reduction actions, dense maps of indoor radon concentration are needed. Indeed, these maps are conceived to identify the areas of higher health risk, more prone to high indoor radon concentration [8]. However, these maps are often based on sparse measurements, e.g. [9], and hence require the use of interpolation techniques [6, 8].

To this aim, some statistical tools, generally referred to as geostatistical methods, are being currently used for the definition of these dense maps and to infer statistical correlation of the radon concentration with the geological conditions [8] and the dwellings and lifestyle features [6] of the sample areas. The geostatistical interpolations, like the kriging, are based on the use of regionalized variables whose spatial correlation is quantitatively estimated through the definition of a semi-variogram [10]. To compute a semi-variogram describing the spatial correlation in a reliable way is a challenging issue due to the fluctuations of the semi-variances as a function of distance and to the difficulties of finding a trend that fits both the correlated and uncorrelated part of the semi-variogram [8, 11]. With a view to define soon a radon risk map for the Campania, a region of southern Italy characterized by a complex tectonostratigraphic evolution that led to the formation of a great variety of geological environments [12], we here present an innovative method for variogram computation. This latter is based on the calculation of the Gini coefficient, which is a measure of the inequality of a distribution, widely used for the socioeconomic applications [13]. We demonstrated the efficiency of this approach applying it to several simulated radon concentrations and systematically comparing it with the classically used variograms.

Methods

The semi-variance and the variogram

One of the most challenging issues for radon risk assessment is to provide a realistic map of expected concentration exploiting sparse experimental measures. To this aim, several procedures have been proposed, generally referred to as geostatistic tools, like the kriging. These techniques are defined as distance-weighted interpolations of the experimental concentrations deriving from the spatial correlation analysis.

The spatial correlation among the measures is generally estimated through the analysis of the semi-variance variability with the inter-distance among all the measures. This variability, generally referred to as semi-variogram, is defined as [14]:

$$(1) \quad \gamma\left(h = \left\|x_1\, x_j\right\|\right) = \frac{1}{2nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\left\{z(x_i) - z(x_j)\,(x_i + h_j)\right\}^2$$

where n is the number of measure; m is the number of measure at locations having a distance h from the measure i; x_i is a geographic location ($i = 1,2,3,...,n$), and z is the concentration value.

It is worth to note that for practical applications, the distances h, also indicated as lags, are subdivided into equally spaced bins. In this case, the quantity $\gamma(h_j)$ is considered as computed in the middle of the bin $[h_j - ½l, h_j + ½l]$, with l being the size of the bins.

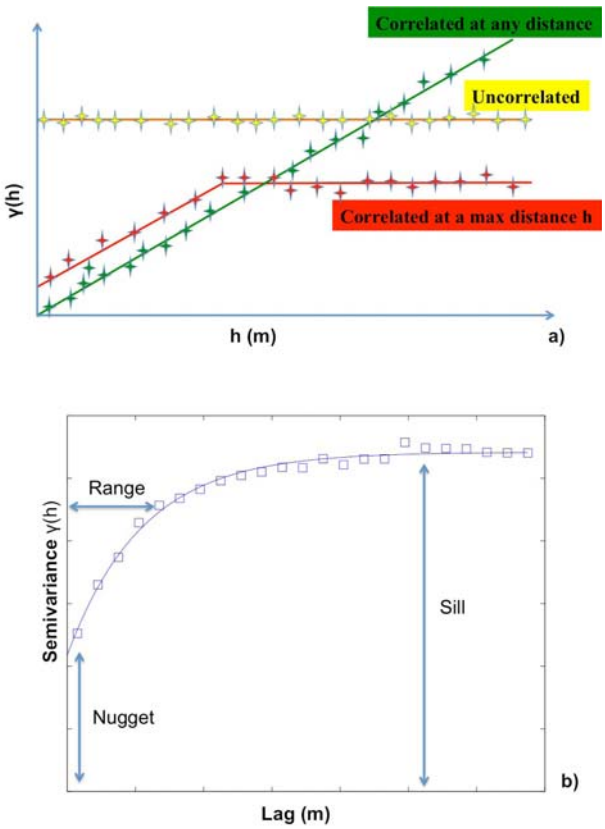

Fig. 1. (a) Different variogram trends depending on spatial correlation of the experimental measures and (b) an example of experimental semi-variance fitted by an exponential function. For sake of clarity, the parameters of the fit are reported within the panel.

The behaviour of γ as a function of h might evidence three different trends for the semi-variogram as shown in Fig. 1a. If the measurements are uncorrelated (in yellow), the values will oscillate around a constant value. Instead, if they are correlated in the whole range of investigation, $\gamma(h)$ will be expected to monotonically increase (green diamonds and line). Finally, if the measurements are correlated up to a certain distance and uncorrelated beyond, the trend will assume the behaviour described by the red line monotonically increasing before reaching a plateau level.

To use the $\gamma(h)$ discrete estimates for interpolation applications, like the kriging [15], these values must be inverted using a fit with a function accounting for the two phases of the semi-variogram.

A widely used function to fit such a trend is the following exponential relationship:

$$(2) \qquad \gamma(h) = S\left\{1 - e^{-\frac{|h|}{R}}\right\} + N$$

The model in Eq. (2) is characterized by three parameters: the range R is the distance at which the semi-variogram gets about 63% of the plateau level identified as the sill parameter S. In the other words, R can be considered as an estimator of the maximum distance for which two points are considered

spatially correlated. The nugget N is an estimate of the residual, spatially uncorrelated noise (Fig. 1b). Although crucial to limit the uncertainty on the estimation of expected concentration, it was shown that identify the correlation trend and hence the parameters R and S from an indoor concentration measurement campaign is not a trivial process, in particular for sparse and noisy sampling. In the next Section, we present a new tool for the variogram computation based on the spatial variability of the Gini coefficient [15–17].

Gini coefficient and Lorenz curve

The Lorenz curve and the Gini coefficient are two statistical tools generally used for socioeconomic applications [13]. The Lorenz curve is a visual representation of the degree of inhomogeneity of a distribution. From this curve, a quantitative estimation of the inhomogeneity can be inferred through the computation of the Gini coefficient. In the case of homogeneous distribution, the Lorenz curve takes the analytical form $y = x$, and the Gini coefficient $G = 0$ while for all other cases the Lorenz curve is convex to the y-axis, never rises above the line of equality [17] and leads to a G ranging from 0 to 1 (Fig. 2).

Figure 2 shows the trend of the Lorenz curve in the (p,q) reference system with p_i and q_i representing the ordered relative cumulative frequencies and the cumulative relative concentration, respectively.

They can be expressed as $p_i = i/n$; and $q_i = A_i/X$; the index i ranges from 1 to n with n being the number of experimental measurements $\{x_j\}$. The A quantities are the cumulative experimental values, that is $A_i = \Sigma^i_{j=1} x_j$, while $X = \Sigma^n_{i=1} A_i$. It is worth to note that to build the Lorenz curve, as shown in Fig. 2, the measurements $\{x_j\}$ must be sorted in an ascending order [17].

From a geometrical point of view, the Gini coefficient represents the ratio of the area between the

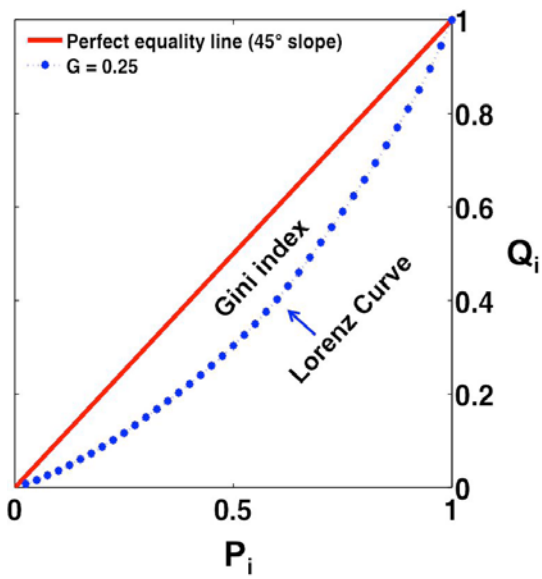

Fig. 2. Lorenz curve to evaluate the inhomogeneity degree of the radon concentrations.

equality line distribution and the observed Lorenz curve to the area under the uniform distribution: lower values of G indicate a pretty uniform distribution, while higher values stem from inhomogeneous distributions. It can be expressed as [17]:

$$G = \frac{\sum_i^{N-1}(p_i - q_i)}{\sum_i^{N-1} p_i} \tag{3}$$

The semi-variogram with Gini coefficient

By analogy with the classical semi-variogram described in Eq. (1), a different semi-variogram based on the computation of the Gini coefficient can be defined.

To this aim, we can define the semi-variogram as

$$\gamma_G(h) = \frac{1}{n_G}\sum_{i=1}^{n} G_i(h) \tag{4}$$

By analogy with Eq. (1), n is the number of experimental measure while h is a distance and as said, for discretized problems, it is assumed to be the center of equally spaced bins. G_i is the Gini coefficient computed as shown in Eq. (3), considering the concentration at the measure points i along with all the other measures having a distance d from the position i such that $h - (l/2) < d \le h + (l/2)$, with l being the bin size. If the measure at the position i has less than three points in the considered bin h, $G_i(h)$, is assumed to be 0. Therefore, n_G is the number of realizations for which $G_i(h) > 0$.

Simulated radon concentration dataset

The efficiency of the presented tool as compared to the classically used variogram has been tested by means of several simulated distributions of radon concentration defined over a simplified squared (100 × 100 km^2) surface. Over this surface, 1000 measure points are randomly extracted (Fig. 3a, black dots) from a uniform distribution.

Recently, it has been shown that a geographical variability of the geometrical means (GM(x,y)) of radon concentration can be interpreted in terms of different soil lithological properties [8]. Following this finding, the sample area has been divided into four squares; each of them is thought as characterized by different types of geological descriptive classes, hence by different expected GM(x,y) [8] (Fig. 3a). Several studies show a clear relationship between geological factors and indoor radon concentration; however other factors, to a lesser extent, such as soil structure, groundwater conditions, ventilation and condition of the building, and so on, definitely influence the measurements [16]. For our purposes, all these factors can be considered as a source of spatially uncorrelated noise.

Therefore, for each concentration dataset, on each measure point, the simulated concentration is randomly extracted from a log-normal distribution having μ = GM(x,y) and a fixed standard deviation of the logarithm of concentration σ ranging from

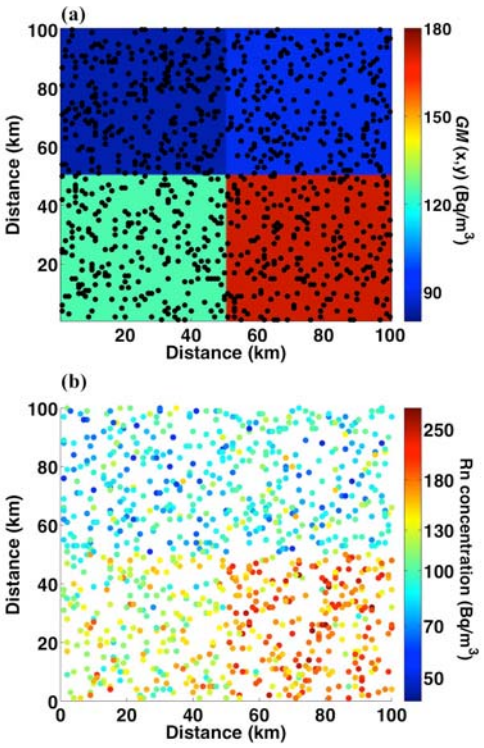

Fig. 3. (a) Measure points (black dots) on a sample surface (100×100 km^2) characterized by different types of geological descriptive classes. (b) One of the stochastically simulated radon concentration distribution on the sample surface (case with $\sigma = 0.2$).

0.2 to 0.5. These values are compatible with the uncertainties computed from previous concentration surveys in the Campania region [4, 7]. In Fig. 3b, there is an example of the distribution.

Results

In Fig. 4, the results are presented as a systematic comparison between the classical and Gini variograms. The variograms have been computed considering the whole simulated dataset shown in Fig. 3 and the figure shows the results for the case $\sigma = 0.2$ (Fig. 4a,b) and $\sigma = 0.5$ (Fig. 4c,d).

In Fig. 4, the blue and red squares represent the $\gamma(h)$ values calculated by means of Eqs. (1) and (4), respectively. By fitting both trends with Eq. (2) imposing $N = 0$ and $\sigma = 0.2$ (Fig. 4a), the Gini variogram, rapidly reaches the plateau level ($S = 0.17$) within the investigated distance range, leading to a small estimated range value ($R = 9.21$ km). At the sampled distances, for the classical variogram, the trend is still far from the plateau level not allowing to clearly identifying the distance at which the measures can be considered uncorrelated. The described results are even more evident when we apply the same analysis to noisier simulated data, imposing $\sigma = 0.5$ (Fig. 4c). Although the random noise affects the correlation distance leading to smaller R values, for the Gini case, the plateau level is still recognizable, and the trend has a good agreement with

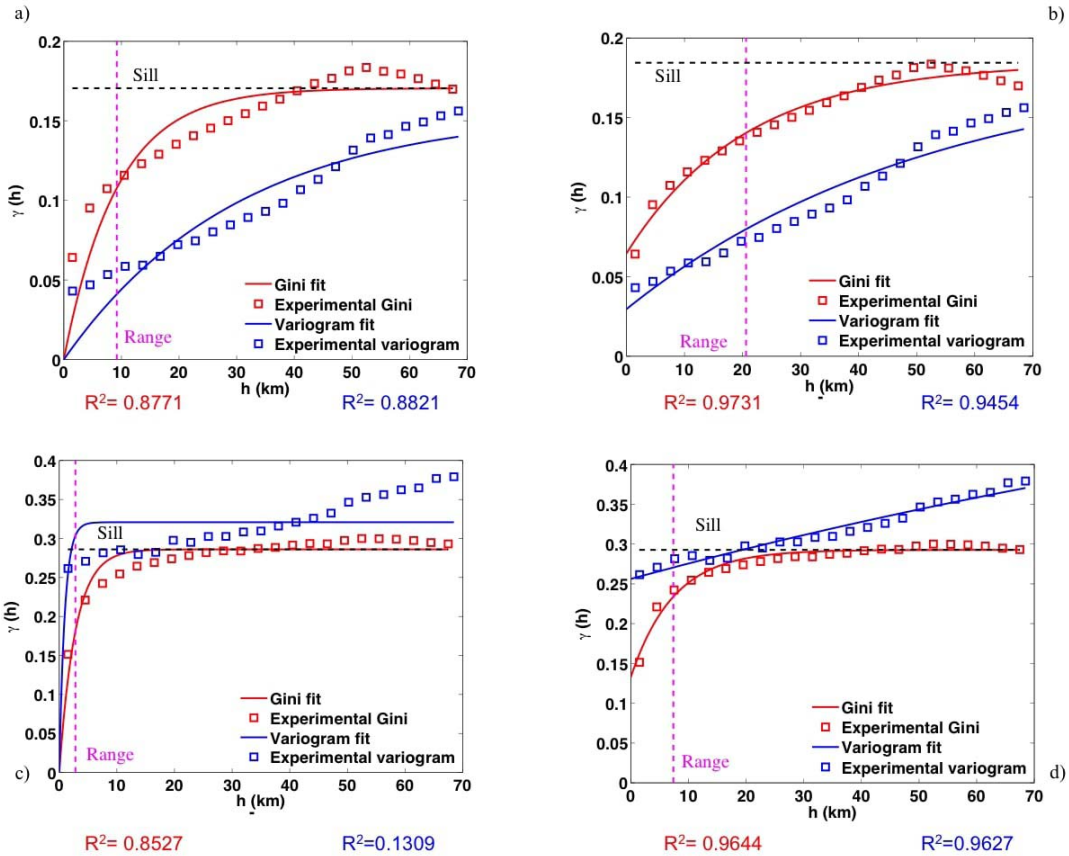

Fig. 4. Comparison between the classical variogram (blue squares and solid line) and the Gini variogram (red squares and solid line) with $N = 0$ (panels a and c) and $N \neq 0$ (panels b and d) for $\sigma = 0.2$ (panels a and b) and $\sigma = 0.5$ (panels c and d). Within all panels, the dashed magenta and black lines depict the range distances and the sill levels respectively.

exponential fit ($R^2 = 0.85$). Conversely, the classical variogram monotonically increases with distance and hence the exponential function fails to fit the experimental points ($R^2 = 0.13$). In this latter case, the low correlation distance ($R = 0.96$ km) is only due to the imposed constraint on the nugget ($N = 0$).

When we let the nugget N as a free parameter of the fit (Fig. 4b,d), we noticed an improvement of the fit agreement with $R^2 > 0.94$ for all cases, and also for this analysis, the Gini method better describes the spatial correlation. In the case $\sigma = 0.2$ (Fig. 4b), the trend asymptotically tends to the sill within the investigated distance range and $R = 20.61$ km is consistent with the simulated problem. The same observation holds for larger sigma ($\sigma = 0.5$, Fig. 4d); in this case, the classical variogram behaves as all the measurements were correlated at any investigated distance (compare with the sketch in Fig. 1a) hence leading to unrealistic estimates of range and sill.

Conclusions

In this article, we have presented a new method for the computation of the variogram, generally used for geostatistical applications such as the prediction of radon exhalation starting from sparse measurements. The development of this class of variograms is based on the calculation of the Gini coefficient, a tool widely used for several applications to quantify the degree of the inhomogeneity of a distribution. A systematic comparison of this method with respect to the classical one has been carried out on a synthetic dataset of radon concentration. We have verified that by using the Gini method, the transition between correlated and uncorrelated distances is better constrained and the parameters describing this spatial correlation are more consistent with the initial model. Moreover, the analyses of datasets generated with different noise levels have shown that the Gini method is more efficient in discriminating the variations due to the environmental conditions from those ones due to random noise at least for the length scales taken into account in the synthetic datasets. These features can be interpreted as intrinsic of the method since the $\gamma(h)$ parameters are obtained through an inhomogeneity quantification involving measure clusters in a given distance range rather than on the average concentration differences between pairs of measure points. We, therefore, conclude that this method can be applied to perform distance-weighted interpolation, for example, kriging, and hence to possibly identify the geological areas that act as homogeneous sources of radon exhalation. Future studies are needed to deepen these findings, in particular, to assess whether and for which length scales this method might better constrain the spatial variability of the radon concentration. These analyses will be performed through the application of the Gini-based kriging interpolation to both synthetic and real radon concentration datasets.

Acknowledment. The work was performed between the Advanced Biomedical Science Department and Department of Physics, "E. Pancini", University of Naples, Federico II, Naples, Italy

ORCID

G. M. Adinolfi http://orcid.org/0000-0002-3215-6566
F. Loffredo http://orcid.org/0000-0003-3105-1030
M. Quatro http://orcid.org/0000-0003-0119-2393
F. Savino http://orcid.org/0000-0001-5733-4325
A. Scala http://orcid.org/0000-0003-4584-463X

References

1. International Agency for Research on Cancer. (1988). *Manmade mineral fibres and radon.* (IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Vol. 43). Lyon, France: IARC.
2. United Nations Scientific Committee on the Effects of Atomic Radiation. (2000). *Sources and effects of ionizing radiation. United Nations Scientific Committee on the Effects of Atomic Radiation UNSCEAR 2000 Report to the General Assembly, with Scientific Annexes. Vol. 1: Sources*. New York: United Nations.
3. Kavasi, N., Somlai, J., Szeiler, G., Szabo, B., Schafer, I., & Kovacs, T. (2010). Estimation of effective doses to cavers based on radon measurements carried out in seven caves of the Bakony Mountains in Hungary. *Radiat. Meas.*, *45*, 1068–1071. https://doi.org/10.1016/j.radmeas.2010.07.017.
4. Quarto, M., Pugliese, M., Loffredo, F., Zambella, C., & Roca, V. (2014). Radon measurements and effective dose from radon inhalation estimation in the neapolitan catacombs. *Radiat. Prot. Dosim.*, *158*, 442–446. https://doi.org/10.1093/rpd/nct255.
5. Kendall, G. M. (2004). Controls on radioactivity in water supplies in England and Wales, with especial reference to radon. *J. Radiol. Prot.*, *24*, 409–412. DOI: 10.1088/0952-4746/24/4/005.
6. Demoury, C., Ielsch, G., Hemon, D., Laurent, O., Laurier, D., Clavel, J., & Guillevic, J. (2013). A statistical evaluation of the influence of housing characteristics and geogenic radon potential on indoor radon concentrations in France. *J. Environ. Radioact.*, *126*, 216–225. https://doi.org/10.1016/j.jenvrad.2013.08.006.
7. Quarto, M., Pugliese, M., Loffredo, F., & Roca, V. (2016). Indoor radon concentration and gamma dose rate in dwellings of the Province of Naples, South Italy, and estimation of the effective dose to the inhabitants. *Radioprotection*, *51*(1), 31–36. DOI: 10.1051/radiopro/2015021.
8. Bossew, P., Zunić, Z. S., Stojanovska, Z., Tollefsen, T., Carpentieri, C., Veselinovic, N., Komatina, S., Vaupotic, J., Simovic, R. D., Antignani, S., & Bochicchio, F. (2014). Geographical distribution of the annual mean radon concentrations in primary schools of Southern Serbia – application of geostatistical methods. *J. Environ. Radioact.*, *127*, 141–148. https://doi.org/10.1016/j.jenvrad.2013.09.015.

9. Menzler, S., Piller, G., Gruson, M., Rosario, A. S., Wichmann, H. E., & Kreienbrock, L. (2008). Population attributable fraction for lung cancer due to residential radon in Switzerland and Germany. *Health Phys*., *95*(2), 179–189. DOI: 10.1097/01.HP.0000309769.55126.03.
10. McBratney, A. B., Webster, R., & Burgess, T. M. (1981). The design of optimal sampling schemes for local estimation and mapping of regionalized variables-I: Theory and method. *Comput. Geosci*., *7*(4), 331–334. https://doi.org/10.1016/0098-3004(81)90077-7.
11. Zhu, H. C., Charlet, J. M., & Poffijn, A. (2001). Radon risk mapping in southern Belgium: an application of geostatistical and GIS techniques. *Sci. Total Environ*., *272*(1/3), 203–210. https://doi.org/10.1016/S0048-9697(01)00693-3.
12. Vitale, S., & Ciarcia, S. (2013). Tectono-stratigraphic and kinematic evolution of the southern Apennines/Calabria–Peloritani Terrane system (Italy). *Tectonophysics*, *583*, 164–182. https://doi.org/10.1016/j.tecto.2012.11.004.
13. Pandey, M. D., & Nathwani, J. S. (1996). Measurement of socio-economic inequality using the life-quality index. *Soc. Indic. Res*., *39*, 187–202.
14. Chiles, J. P., & Delfiner, P. (1999). *Geostatistics: Modeling spatial uncertainty*. New York: Wiley.
15. Lark, R. M. (2000). Estimating variograms of soil properties by the method-of-moments and maximum likelihood. *Eur. J. Soil Sci*., *51*(4), 717–728. https://doi.org/10.1046/j.1365-2389.2000.00345.x.
16. Borgoni, R., Quatto, P., Somà, G., & de Bartolo, D. (2010). A geostatistical approach to define guidelines for radon prone area identification. *Stat. Methods Appl.*, *19*, 255–276. DOI: 10.1007/s10260-009-0128-x.
17. Gini, C. (1912). *Memorie di metodologia statistica. Vol. 1. Variabilita concentrazione*. Rome: Libreria Eredi Virgilio Veschi.