

Innovative modeling methodology for mapping of radon potential based on local relationships between indoor radon measurements and environmental geology factors

S. De Novellis¹, A. Pasculli¹ & S. Palermi²

¹*Department of Engineering and Geology, University G. d'Annunzio, Chieti-Pescara, Italy*

²*Agency of Environmental Protection of Abruzzo (ARTA), Pescara, Italy*

Abstract

The aim of this study is to investigate the relationship between environmental variables and indoor radon concentration. Such relationships are important for mapping the geogenic radon potential, considered as the quantity associated with the risk of radon that only depends on the geological and geophysical conditions of the area considered. In fact, a proper definition of the radon potential could provide a guide for the identification of radon-prone areas, in particular when the number and/or quality of indoor radon data are inadequate. Generally, the assessment of radon potential is obtained via a global inference technique, assuming that the relationship studied is spatially stationary. Our approach, instead, analyzes the consequences associated with significant changes of the spatial relationship between indoor radon measures, adequately corrected, and geological features. This allows us to estimate the indoor radon concentration by considering local environmental properties, in agreement with the geogenic radon potential of the underlying soil. Therefore the approach followed allows us to map more precisely any area of high radon concentration. The spatial statistic analysis has been performed by Geographically Weighted Regression, GWR. Thanks to the indices of spatial autocorrelation analysis and GWR, we have shown the presence of localized effects and differences between variables in space. The final map of



the geogenic radon potential obtained in this work provides guidance for the identification of radon-prone areas of the territory considered and is more accurate than that obtained in previous studies.

Keywords: radon potential, temperature, Geographically Weighted Regression.

1 Introduction

Radon is a natural radioactive gas amongst the greatest causes of population exposure due to natural ionizing radiation [1]. The radon is product from the decay chain of primordial radionuclides ^{238}U , ^{232}Th and ^{235}U . The most abundant isotope is ^{222}Rn with a half-life of 3.82 days (from the chain of the ^{238}U), which decays itself in stable lead ^{206}Pb through an intermediate decay chain. The short-life isotopes products, such as ^{218}Po , ^{214}Pb , ^{214}Bi and ^{214}Po , are of great importance for human exposure because they are easily inhaled and decay by emitting α (a few MeV) within the human body.

The ability of radon to penetrate and accumulate in indoor environments depends on the geological characteristics of the area considered, therefore by the concentration of the radium in soil and the transport properties of the gas itself. The transport is particularly affected by the permeability of the soil, ie the ability of a fluid to pass through it.

The indoor radon concentration is also determined by climatic conditions. In the presence of a thermal gradient between the internal and external environment, for the stack effect, the flow of air sucked from the soil increases, and then the amount of radon emanating. The outdoor temperature instead determines the lifestyle of the population living environments, increasing or decreasing the airflow and thereby influencing the natural processes of gas removal. Many countries, including Italy, carry out campaigns of indoor radon concentration measurements in order to identify the geographical areas in which is a greater probability of exceeding the maximum reference values (radon prone areas) chosen so as to minimize the risk for the population. In this study we use 757 measurements collected from 1991 to 2011 in buildings (schools, homes, offices, etc.) in the province of L'Aquila (5035 km², and about 200000 inhabitants), by the Regional Agency for the Protection of Environment (ARTA). The data are gathered with various sampling methods in the area of interest using dosimeters and nuclear track CR-39 [2–4].

The purpose of this paper is to analyze the heterogeneity of the relationship that describes the concentration of indoor radon as a function of environmental and geological features of the area studied.

For the modeling analysis we use Geographically Weighted Regression, GWR. The local analysis allows to study the spatial variability of relationship. Moreover it highlights areas where there are mis-specification errors, namely the regression equation includes and/or excludes relevant factors.



2 Selected assumptions

2.1 Log-normal distribution

The multitude of factor that influence the concentration of indoor radon can be considered independent of each other and multiplicative, so as to justify the usual log-normal distribution for measurements of radon concentrations in homes, whether the whole of a country or a small area is considered [5, 6] (see Fig. 2). Possible reasons for this deviation are due to the uncertainty of the measurement, the background subtraction which leads to an excess of low values and problems related to representativity sampling [6, 7]. Since extremely low values cause problems for the lower tail distribution, we decided to correct (with factor D) the value of concentration measured radon, the lowest values are shifted, and we left higher values virtually unchanged [7] (Fig. 1):

$$Rn_{in}^{new} = \frac{Rn_{in} - Rn_{out}}{2} + \sqrt{\left(\frac{Rn_{in} - Rn_{out}}{4}\right)^2 + D^2} \quad (1)$$

The best correction factor is obtained using the D'Agostino's normality test [8], namely, we used Rn_{out} and D values that maximize the p-value (null hypothesis is the normal distribution).

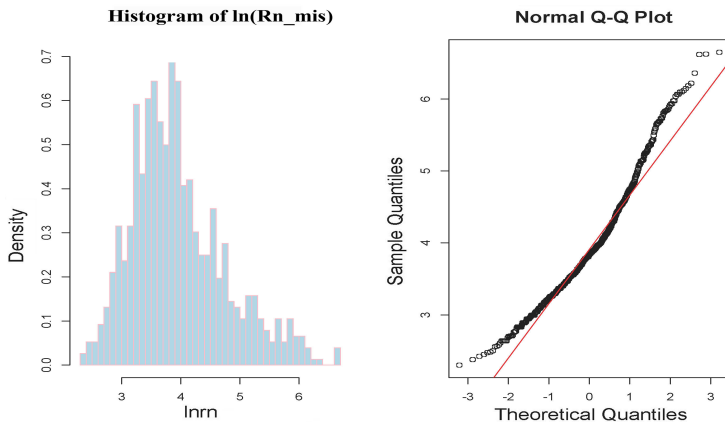


Figure 1: *left*: Histogram of the logarithm of indoor radon data, Rn_{mis} , without correction for background; *right*: graph quantile-quantile for the normal distribution $N(0, 1)$ and $\ln(Rn_{mis})$.

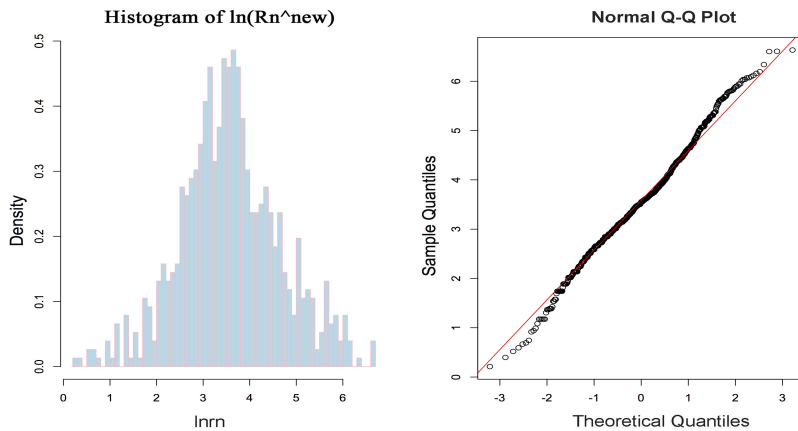


Figure 2: *left*: Histogram of the logarithm of indoor radon with correction for background, Rn_{in}^{new} , from (eq. 1), with $R_{out} = 12Bqm^{-3}$ and $D = 2Bqm^{-3}$. *right*: graph quantile-quantile for the normal distribution $N(0, 1)$ and $\ln(Rn_{in}^{new})$.

2.2 Basic standardization of measurement: normalization to a virtual ground level

In order to highlight the potential emission of the soil (the major source of indoor radon), the measurements have been normalized to a virtual ground level condition, employing multiplicative factors estimated from the data itself (in accord with [2]). The conversion coefficients are evaluated as

$$k_{i \rightarrow 0ct} = \frac{GM^{(0CT)}}{GM^{(i)}}$$

where $GM^{(0CT)}$ and $GM^{(i)}$ are the geometric means of all the measured values related respectively to the ground floor in contact with the soil (0CT) and the i^{th} floor.

Hence the normalized radon concentrations are

$$rn_{std}_{0ct}^{new} = k_{i \rightarrow 0ct} * (Rn_{in}^{new})_i$$

2.3 Explanatory variables

The radon standardized variable is used to define the direct relationship between indoor radon concentration and geological, or generally environmental, characteristics of the area of the building. The explanatory variables of indoor gas can be divided into three groups: radiometric content of the soil, geological/geomorphological features of soil below and around the building and climatic variables.

For each of these categories we examine a factor generally predominant (see [9]): the equivalent uranium content, the permeability of the soil and the average annual temperature for the climate variability.

The radiometric data of equivalent uranium (eU) are determined by the signal γ emitted by ^{214}Bi in the ground as short-lived ^{222}Rn decay product. The measures come from studies conducted by Bellotti *et al.* [10] of 200 soil samples collected uniformly in the Province of L'Aquila in order to estimate the activity of the rays γ in soil and water. The permeability is directly influenced by fracturing, the porosity and the grain size of the rocks and its components, so the average value is obtained by combining all this information, as a kind of derived variable [9].

The annual temperature average has been calculated using data collected from 43 thermometric stations of the Hydrographic Service of the Abruzzo Region and interpolated by kriging with external drift (altitude) [11]. These results will be discussed in incoming submitted article.

3 Material and methods

Geographically Weighted Regression (GWR) [12, 13] is a useful technique for modeling local spatial relationships. A typical model of GWR can be written as:

$$y_i = \beta_0(u_i, v_i) + \sum_k^M \beta_k(u_i, v_i) x_{ik} + \epsilon_i \quad (2)$$

where y_i is the dependent variable of point i with coordinates (u_i, v_i) ; $\beta_k(u_i, v_i)$ with $k = 1, \dots, M$, are regression coefficients for each point i and for each variable k ; x_{ik} is the value of k -th covariant predictor in i ; $\beta_0(u_i, v_i)$ is the intercept variable for each point i ; ϵ_i is the error term in the i that follows a normal distribution with zero mean and homogeneous variance.

In matrix form, regression coefficients are:

$$\hat{\beta}(i) = (X^T W(i) X)^{-1} X^T W(i) y \quad (3)$$

where $W(i) = \text{diag}[w_1(i), \dots, w_n(i)]$ is a diagonal matrix that varies for each calibration point i and applies weights to each of the n observations. A choice for the weight function is a Gaussian curve:

$$w_{ij} = \exp \left[-\frac{d_{ij}^2}{2b^2} \right] \quad (4)$$

where d_{ij} is the distance between the observed point i and N neighboring points ($j = 1, \dots, N$) and b is the bandwidth. Consequently, model calibration involves the selection of the number of neighboring points to be included in the estimation of local parameters (N), or the distance decay (b) of the weight function (bandwidth). Selecting the weight matrix, the value of b which produces the minimum of *Akaike Information Criterion* corrected (AIC_c) [14] is the optimal bandwidth. To perform the analysis, we used the software GWR4 by Nakaya [15],

available at website <http://gwr.nuim.ie>. Through software Arcgis we mapped the regressive coefficients estimated for each point i , the t-statistics and the R-squared for each local model.

4 Results GWR

The spatial variables generally exhibit autocorrelation values, therefore neighbors tend to assume similar values and among their dependent. This implies that the usual techniques of global regression (OLS) are inappropriate to model the phenomenon. Analytically, we can measure the spatial dependence between the values of a variable with the global Moran’s I statistic [16], listed in Table 1:

Table 1: Moran’s I statistic and p-values (with the null hypothesis of random distribution) for the indoor radon concentrations (log-transformed) and the three regressive variable (eU, permeability and temperature). These are calculated by the function *moran.test()* of package *spdep* in **R** [17].

	I di Moran	p.value
$\ln(\text{rn_std}_{0ct}^{new})$	0.036	0.007
$\ln \text{perm}$	0.247	0.000
$\ln T$	0.432	0.000
$\ln \text{eU}$	0.361	0.000

In agreement with what is expected, the Moran’s I for parameters is positive, highlighting the presence of spatial autocorrelation. The available measures don’t come from homes homogeneously distributed over the territory and spatial data density are proportional to the density of the houses themselves. To reduce this occurrence, it is preferable to implement a simple strategy of declustering, consisting of aggregating measures based on membership in square blocks of side equal to 2 km.

4.1 Discussion

Applying the GWR in our data set, we can analyze the spatial variability of the relationship between the concentration of indoor radon, standardized at virtual ground floor, with three main predictors: equivalent uranium concentration, soil permeability and average outdoor temperature.



$$\ln(rn_std_{oct}^{new}) = \beta_0(u_i, v_i) + \beta_{eU}(u_i, v_i) \ln(eU)_i + \beta_{perm}(u_i, v_i) \ln(permeability)_i + \beta_{temp}(u_i, v_i) \ln(temperature)_i + \epsilon_i \quad (5)$$

The GWR models show an improvement of the fit of the data with respect to multivariate linear regression (OLS), with regression coefficient constant, evidenced by the reduction of diagnostic test AIC , AIC_c , RSS (sum of squared residuals), R^2 and correct version R_{adj}^2 (Table 2).

model	R^2	R_{adj}^2	AIC	AIC_c	RSS
OLS	0.011	0.005	2000.747	2000.833	686.261
GWR	0.103	0.052	829.775	831.011	259.311

Table 2: Summary of characteristics for fitting regression models.

GWR allows us to explore the spatial non-stationarity of the radon concentration, intimately connected with the local characteristics of the area.

In the Fig. 3 the maps show significant spatial variability of regression coefficient and the associated t-values. The negative coefficients to equivalent uranium (β_{eU}) in the eastern part of the territory highlight local issues of our model, however these are the t-values are not significant ($-1.96 < t < 1.96$). In a specular manner, in smaller areas (south-east and eastern tip) the coefficients of temperature are positive, while the local statistic t is significant only in the central part of the territory, which correspond to high values of temperature and the correlation is positive. The regression coefficient of permeability is, however, always positive in line with the physical analysis, the t-values follow a trend similar to that for β_{eU} .

Finally, we report in Fig. 4 the spatial variation of the local coefficient of determination R^2 , which is a characteristic of the goodness of fit. For the radon studies, low R^2 values are typical (see for example Pasculli *et al.* [9]) due to a multifactorial dependency difficult to calculate. However, the results in our studies are satisfactory for the area to the south-east of the province and in the western part of the region ($R^2 = 0.18$). Note that the values of R^2 improve at both of the areas where the coefficients t-values have no abnormalities and, except for the temperature, can be considered significant ($|t| > 1.64$, $p\text{-value} < 0.10$). Our analysis also highlights the need to deepen and improve the survey in the eastern part of the area, given the difficulty in the description of the phenomenon ($R^2 < 0.07$). The introduction of the analysis of variability due to characteristics of the building will lead to a improvement of description.

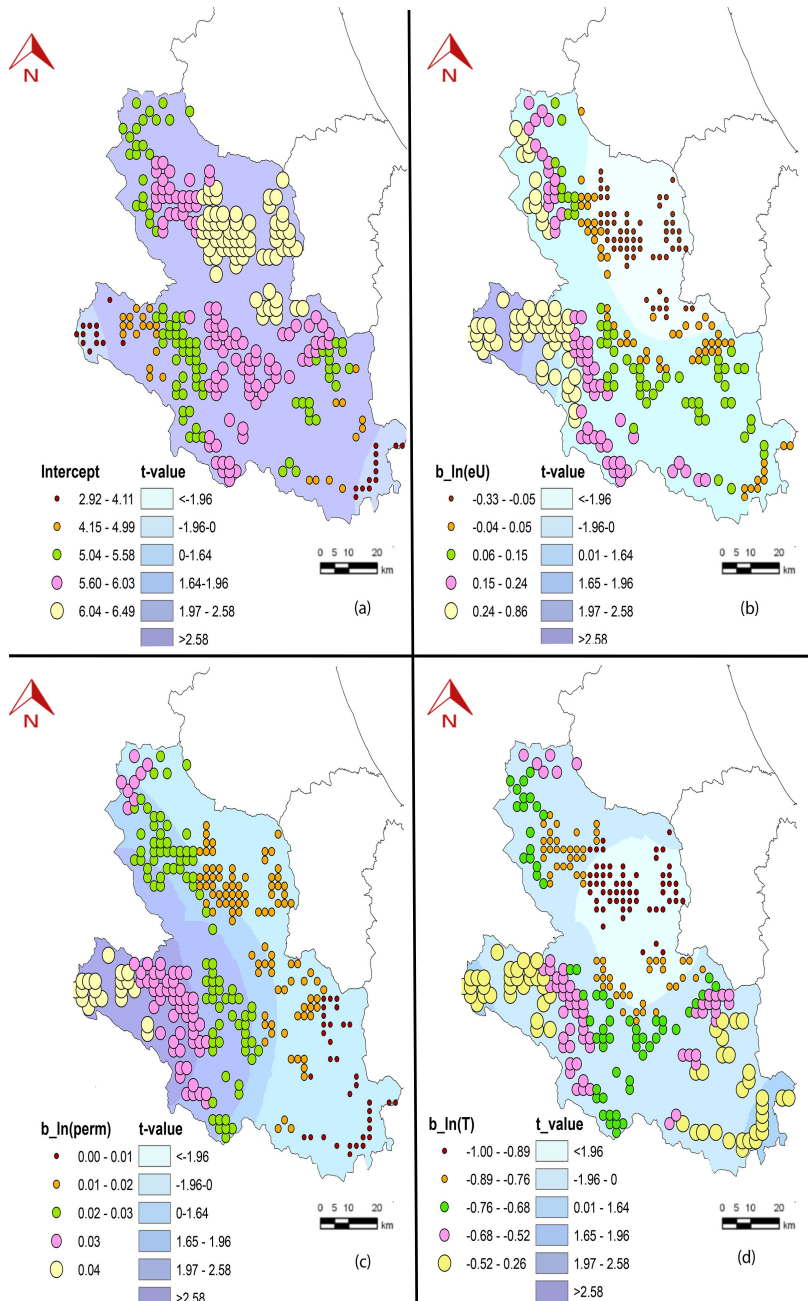


Figure 3: Maps of local regressor coefficients (coloured spots): a) intercept (β_0), b) β_{eU} , c) β_{perm} , d) β_{temp} ; maps of t-statistics for a) β_0 , b) β_{eU} , c) β_{perm} , d) β_{temp} .

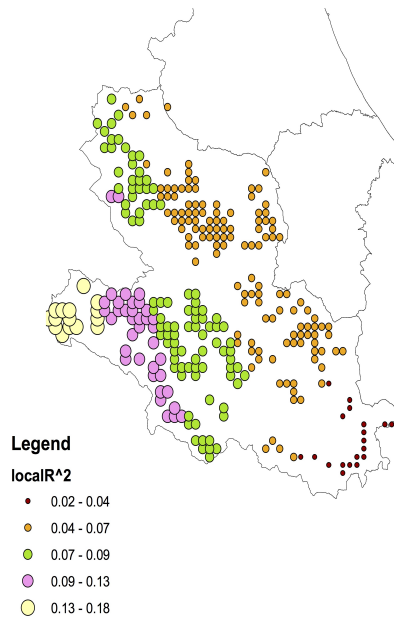


Figure 4: Distribution of local R^2 .

5 Conclusions

This work aimed to define a map of the geogenic radon potential (PRG), which is directly related to the risk of radon [18, 19]. We studied the relationship between the concentration of indoor radon (dependent variable) and the geophysical/environmental features of region (independent variables), using a local regression in order to highlight the importance of local condition to the determination of the gas concentration. The measures of the radon have been normalized to a virtual ground floor and analyzed with the GWR.

This enabled us to generate maps for the regression coefficients overcoming the uniformity of a global regression. Moreover, this approach allows to highlight areas where our model did not apply correctly, due to mis-specification errors and/or the quality of the available dataset. In future work we will analyze the component of variability due to the building features by the multilevel analysis, in order to better specify the amount PRG. The aim is to map the risk of radon even in areas where there are no measurements of radon and to provide information to the competent authorities for the establishment of an action plan for radon (Art. 103, Directive 2013/59 Euratom [20]).

References

- [1] UNSCEAR, *Sources and Effects of Ionizing Radiation.*, volume 1. United Nations Publications, 2000.
- [2] Palermi, S. & Pasculli, A., Radon mapping in Abruzzo, Italy. *4th Canadian Conference on Geohazards Québec City Canada, May*, 2008.
- [3] Nissi, E., Sarra, A. & Palermi, S., Radon level in dwellings and uranium content in soil in the Abruzzo region: a preliminary investigation by GWR. *Advanced Statistical Methods for the Analysis of Large Data-Sets*, Springer, pp. 265–275, 2012.
- [4] Sarra, A., Nissi, E. & Palermi, S., Residential radon concentration in the Abruzzo region (Italy): a different perspective for identifying radon prone areas. *Environmental and Ecological Statistics*, **10(2)**, pp. 219–247, 2012.
- [5] Nero, A., Schwehr, M., Nazaroff, W. & Revzan, K., Distribution of airborne radon-222 concentrations in U.S. homes. *Science*, **234(4779)**, pp. 992–997, 1986.
- [6] Miles, J., Mapping the proportion of the housing stock exceeding a radon reference level. *Radiation Protection Dosimetry*, **56(1–4)**, pp. 207–210, 1994.
- [7] Price, P. *et al.*, Predictions and maps of county mean indoor radon concentrations in the mid- atlantic states. *Health Physics*, **72(6)**, p. 893, 1997.
- [8] D’Agostino, R. & Pearson, E.S., Tests for departure from normality. Empirical results for the distributions of b_2 and $\sqrt{b_1}$. *Biometrika*, **60(3)**, pp. 613–622, 1973.
- [9] Pasculli, A., Palermi, S., Sarra, A., Piacentini, T. & Miccadei, E., A modelling methodology for the analysis of radon potential based on environmental geology and geographically weighted regression. *Environmental Modelling & Software*, **54**, pp. 165–181, 2014.
- [10] Bellotti, E., Di Carlo, G., Di Sabatino, D., Ferrari, N., Laubenstein, M., Pandola, L. & Tomei, C., γ -ray spectrometry of soil samples from the Provincia dell’Aquila (Central Italy). *ApplRad Isot*, **65**, pp. 858–865, 2007.
- [11] Hudson, G. & Wackernagel, H., Mapping temperature using kriging with external drift: Theory and an example from Scotland. *International Journal Climatology*, **14**, pp. 77–91, 1994.
- [12] Brunsdon, C., Fotheringham, A.S. & Charlton, M.E., Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical analysis*, **28(4)**, pp. 281–298, 1996.
- [13] Fotheringham, A., Brunsdon, C. & Charlton, M., *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley & Sons, Chichester,UK, 2002.
- [14] Hurvich, C., Simonoff, J. & Tsai, C., Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *J R Stat Soc B*, **60**, pp. 271–293, 1998.
- [15] Nakaya, T., *GWR4 User Manual*. GWR 4 Development Team, 2012.



- [16] Anselin, L., Local Indicators of Spatial Association—LISA. *Geographical Analysis*, **27**, pp. 93—115, 1995.
- [17] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.
- [18] Gruber, V., Bossew, P., De Cort, M. & Tollefsen, T., The European map of the geogenic radon potential. *J Radiol Prot*, **33**, pp. 51–60, 2013.
- [19] Dubois, G., Bossew, P., Tollefsen, T. & De Cort, M., First steps towards a European atlas of natural radiation: status of the European indoor radon map. *Journal of environmental radioactivity*, **101(10)**, pp. 786–798, 2010.
- [20] European Parliament and Council, *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, and repealing Directives 89/618/Euratom, 90/641/Euratom, 96/29/Euratom, 97/43/Euratom and 2003/122/Euratom*, 2014.

