

Application of GWR for assessing spatial non-stationarity of residential radon concentration: evidence from Abruzzo data

Eugenia Nissi¹ - Annalina Sarra¹ - Sergio Palmeri²

¹Dipartimento di Metodi Quantitativi e Teoria Economica,
Università “G.d’Annunzio”

²ARTA (Agenzia Regionale per la Tutela dell’Ambiente)
nissi@dmqte.unich.it - asarra@dmqte.unich.it
s.palermi@artaabruzzo.it

Abstract: Radon is a noble gas coming from natural decay of uranium. It can migrate from the underlying soil into buildings, where sometimes very high concentration can be found, particularly in the basement or at ground floor. It contributes up to about the 50% of the radiation dose received by the population, constituting a real health hazard. In this study, we use the geographically weighted regression (GWR) technique, to detect spatial non-stationarity of relationship between indoor radon concentration and the radioactivity content of soil in the Province of L’Aquila, in the Abruzzo region (Central Italy). Radon measurements have been taken in a sample of 455 dwellings. Local estimates are obtained and discussed. The significance of the spatial variability in the local parameter estimates is examined by performing a Monte Carlo test.

Keywords: radon gas, uranium, spatial non-stationarity, geographically weighted regression, Monte Carlo testing.

1 Introduction

Radon (^{222}Rn) is a naturally-occurring radioactive noble gas. It derives from the decay of uranium which can be found in varying concentrations in both rocks and soils. Radon becomes a very human health concern when it enters in confined air spaces, as prolonged exposure to high levels of this substance is well recognised to be the second cause of lung cancer after tobacco. Many developed countries have performed systematic studies in order to assess the exposure of people to this radioactive gas and identify areas where there is a high probability of exceeding reference values and for which remedial actions are necessary (*radon prone areas*). Mainly, geostatistical methods and Bayesian hierarchical models are traditionally employed to map, model and predict radon levels. For the present research, a statistical technique, known as Geographically Weighted Regression (GWR), is used to investigate the spatial association between indoor radon measurements, expressed in Bq/m^3 , collected in a sample of 455 dwellings of Province dell’Aquila (Central Italy) and some naturally occurring radioactive elements (uranium, potassium, thorium) which characterize the underlying soil. This study was motivated by the Regional Agency for the Environment Protection of Abruzzo (ARTA), interested in characterizing the spatial distribution of radon levels in such region and in identifying *radon prone areas*. Since uranium in the

soil is the primary source of indoor radon, we stress that measuring their local relationships, via GWR, can provide useful insights in ascertaining regional zone with high radon concentration.

The remainder of this paper is organised as follows: Section 2 gives a short description of Geographically Weighted Regression technique. Section 3 presents data and results as well as their interpretation.

2 Methodology

This section offers a very brief introduction to a theoretical framework for GWR. Despite it being a relative new technique, GWR has become a standard tool for constructing spatial models in a variety of disciplines, recognising that the relationship between variables, measured at different geographic points, might not be constant over space. GWR extends the traditional global regression by allowing local rather than global parameters to be estimated. A typical model of GWR can be written as:

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) x_{ik} + \varepsilon_i \quad (1)$$

Where β_{ki} ($k=1,2,...,M$) are the regression coefficients for each location i and each variable k ; (u_i, v_i) denotes the coordinates of i -th point in space and ε_i are error terms. Equation 1 creates a continuous surface of estimated parameters values. The local parameters $\beta_k(u_i, v_i)$ are estimated by weighted least-squares estimator, given by :

$$\hat{\beta}(i) = (x^T W(i) x)^{-1} x^T W(i) y \quad (2)$$

In (2) the weight matrix is no longer constant but varies according to the location of point i . It is worth noting that localised parameters can be derived for any point in space regardless whether or not that point is one at which data are measured. In GWR an observation is weighted in accordance with its proximity to regression point i : data points near to location i (u_i, v_i) will be assigned higher weights in the model than data points farther away. Hence, an important step in calculating GWR estimates is the choice of weighting scheme: this requires the specification of a kernel shape and a bandwidth. Several weighting functions (“*kernels*”) can be considered and calibrated, although they tend to be Gaussian or “Gaussian like”, reflecting the type of dependency found in many spatial processes. Whichever weighting function is used, the results are very sensitive to the bandwidth. This is a measure of distance-decay in the weighting function and indicates the extent to which the resulting local calibration results are smoothed. One can use a constant bandwidth around every regression point.

However, it is desirable to have larger bandwidths where data are sparse and have smaller bandwidths where data are plentiful. This consideration justifies the employment of variable (or *adaptive*) weighting function. Consequently, 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 in order to

define the finest bandwidth value or the appropriate value of N . Among them, there are the Aikake Information Criterion (AIC) and the *cross-validation* procedure. The cross-validation criteria is a least square approach which can be formalised as follows:

$$CV = \sum_{i=1}^n [y_i - \hat{y}_{\neq i}(b)]^2 \quad (3)$$

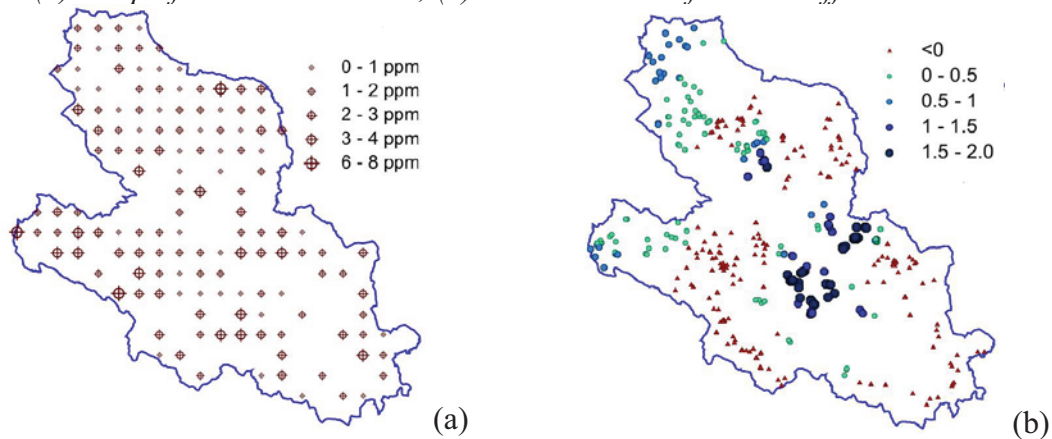
where $\hat{y}_{\neq i}(b)$ is the fitted value of y_i with the observation for point i omitted from the calibration process. Finally, to assess whether any of the local parameter estimates are significantly non-stationary, a Monte Carlo significance test can be performed. For a comprehensive overview about GWR methodology see Fotheringham *et al.* (2002).

3 Data and empirical results

The Regional Agency for the Environment Protection has carried out various indoor radon surveys in Abruzzo since early nineties. This has resulted in a database of radon levels in more than 1800 dwellings, mostly at ground level. A detailed description of these surveys, as well as the sampling strategies adopted, can be found in Palmeri and Pasculli (2008). In this paper, we refer to the most mountainous Province of Abruzzo (Province dell'Aquila), identified in previous studies as the highest indoor radon risk zone. The Province dell'Aquila is a region of about 5000 Km² of extension, moderately populated (60 inhab/ Km²). As stated above, the present analysis explains the zone of high radon concentration as a function of radioactivity of the underlying soil. At this purpose, the recent local modelling technique of GWR, illustrated in the previous section, has employed. In particular, the logarithm of radon concentration, collected in 455 homes, is included in the model as dependent variable whereas the following radiometric soil data: uranium (eU), potassium (K), thorium (eTh) are used as covariates. These latter radiometric explanatory variables were gathered by an experiment conducted by Bellotti *et.al* (2007). They carried out systematic measurements of γ -ray activity in a sample of soils in the area of Province dell'Aquila.

Several other studies have used uranium data as a predictor of indoor radon. Often, such kind of data are available at aggregate level averages. In contrast, in this study, we can rely on point-referenced measurements (see Fig.1-a). However, in our data set, point-referenced measurements of the radioactive elements included in the analysis, are not available at the same sites where radon measurements are taken. In order to overcome this limitation, predicted uranium, potassium and thorium data must be used. Accordingly, the interpolation through the block kriging method (wich provides estimates of the average value of the square blocks centered on the nodes of a 2 km spaced grid) is employed to obtain simulated radiometric areal data. Indoor Radon concentrations normally show an increasing trend along altitudinal gradient due to the decrease of temperatures and the change in other relevant factors. For this reason, we also add this variable (altitude) in our models as it acts as surrogate for mean annual temperature. Our analysis consider as spatial information the geographical coordinates of each dwellings in which the radon measurements are taken, whereas the Euclidean distance between them is used as distance between the corresponding data points.

Fig.1. (a): Map of eU measurements; (b): GWR estimates for eU coefficients



With regard to the Geographically Weighted Regression model, we use the adaptive weighted scheme (fixed number of nearest neighbours for each local model) and the common bisquare function of calculating the weights. The number of nearest neighbour data points to be included within the calibration of the local model is determined from the cross-validation criteria, as formalised in (3). After examining collinearity diagnostics and stepwise regression, only the uranium (eU) remained in the linear regression: all other predictor variables were not significant. By comparing the standard linear regression and the local one (GWR), we found that incorporating spatial information can improve model predictability. Casewise diagnostics, expressed in terms of local R^2 , show values ranging from 0.13 to 0.47 which can be deemed quite good in comparison with the global R^2 obtained by OLS method.

Local findings suggest that there are meaningful spatial variations in the relationships between radon measurements and uranium soil content ($pvalue=0.0000$ for Monte Carlo significance test for spatial variability of parameters). Information on the spatial variation of this relationship (see Fig.1-b) provides useful clues to understanding local factors playing an important role in affecting radon distribution, allowing to distinguish those related to its geochemical source (soil uranium) from those related to its transport through rocks and soil towards the surface (permeability, faults and fractures). GWR is similar to kernel regression except that weights are based on location in geographic rather than attribute space. Future research may want to compare results obtained by applying these two locally weighted regressions.

References

- Bellotti, E., Di Carlo, G., Di Sabatino, D., Ferrari, N., Laubenstein, M., Pandola, L., Tomei, C. (2007) γ -ray spectrometry of soil samples from the Provincia dell'Aquila (Central Italy) *Applied Radiation and Isotopes* 65, 858-865
- Fotheringham, A.S., Brunson, C., and Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester, UK: John Wiley & Sons
- Palermi S., Pasculli A., (2008) Radon mapping in Abruzzo, Italy, *Proceedings of 4th Canadian Conference on Geohazards Québec City Canada, May 20-24th*