Environmental Modelling & Software 54 (2014) 165–181

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

# A modelling methodology for the analysis of radon potential based on environmental geology and geographically weighted regression

Antonio Pasculli [a,*], Sergio Palermi [b], Annalina Sarra [c], Tommaso Piacentini [a], Enrico Miccadei [a]

[a] Department of Engineering and Geology, University G. d'Annunzio, Chieti-Pescara, V.le dei Vestini 31, 65013 Chieti, Italy
[b] Agency of Environmental Protection of Abruzzo (ARTA), V.le G. Marconi, 178, 65127 Pescara, Italy
[c] Department of Economics, University G. d'Annunzio, Pescara, V.le Pindaro, 42, 65127 Pescara, Italy

A R T I C L E I N F O

*Article history:*
Received 19 October 2013
Received in revised form
22 December 2013
Accepted 1 January 2014
Available online 28 January 2014

*Keywords:*
Geostatistic
Environmental geology
Geographically Weighted Regression
Local spatial effect
Indoor radon
Environmental radon potential modelling

A B S T R A C T

Many countries have promoted environmental studies and established national radon programmes in order to identify those geographical areas where high indoor exposure risk of people to this radioactive gas are more likely to be found (often referred to as 'radon-prone areas'). Traditionally, the evaluation of radon potential has been pursued by means of global inference techniques. Conversely, in this paper we present a novel modelling approach, based on well established environmental software, best suited to capture the spatial variability of local relationships between indoor radon measurements and some environmental geology-related factors. The proposed strategy consists of three stages. First, a multilevel model based standardisation of indoor radon data should be carried out in order to reduce the building related variability. Then, the global and local autocorrelation indexes have to be employed to highlight the role of the local effects. The last step implies the use of the *Geographically Weighted Regression* (GWR) to show the differences in associations between indoor radon and the geological factors across space. The method was tested using an available geo-referenced dataset including both radon indoor measurements and geological data related to the territory of an Italian region (Abruzzo). The results are encouraging, although there are several critical issues to be addressed.



## 1. Introduction

It is well known that radon ($^{222}Rn$, a gaseous, chemically inert and radioactive member of the decay chain starting from the isotope $^{238}U$ of uranium) is a major contributor to the ionizing radiation exposure of the general population (World Health Organization, 2009).

In recent years evidence of an association between radon concentration at home and lung cancer has been supported by many epidemiological studies (Kreienbrock et al., 2001; Darby et al., 2005; Krewski et al., 2005). This knowledge is generating a growing attention by national and international authorities aimed at assessing the exposure of people to this radioactive gas and identifying those geographical areas where high indoor radon concentrations are more likely to be found (often referred to as "radon-prone areas"). Over the last decades, national radon surveys were 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) and Italy (Bochicchio et al., 1996). These surveys, whose results are often displayed as "radon maps", represent a fundamental step of national radon programmes, in order to get the necessary awareness for adequate future actions aimed at reducing the risks from the exposure to radon of the general population. The use of radon measurements in existing dwellings and strategies restricted to small spatial scale (Clifford, 2008) are not the only methods to build indicators of radon risk, as geo-statistically and geologically-based large scale approaches were proposed as well. The essential feature of these studies is the assessment of a *geogenic radon potential*, which is a quantity directly related to the geological environment (Gruber et al., 2013; Tung et al., 2013). A properly defined radon potential 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. In principle, the sole geological information (e.g.

* Corresponding author. Tel.: +39 0871 3556159; fax: +39 0871 3555364.
*E-mail address:* a.pasculli@unich.it (A. Pasculli).

1364-8152/$ – see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.envsoft.2014.01.006

radiometric and geochemical data, concentration of radon in soil gas, permeability and moisture content of surface rock and soil), may be quite sufficient to infer the radon potential (Gundersen and Schumann, 1996; Orlando et al., 2000; Kemski et al., 2001; Ielsch et al., 2010). Alternative approaches, combining to varying degrees indoor radon measurements and geological data, were also explored (Smethurst et al., 2008; Appleton et al., 2008, 2011). In some cases (Friedmann, 2005; Bossew et al., 2008), radon data have been pre-processed in order to filter out, as far as possible, the building related variability (due to floor level, building materials, building type, presence of a basement etc.).

Usually, the assessment of radon potential has been pursued by means of global estimation techniques assuming spatial homogeneity of the relationships under study (Apte et al., 1999; Smith and Field, 2007; Bossew et al., 2008). Conversely, following a preliminary study of Nissi et al. (2012), we explore the consequences related to significant spatial variations in the relationship between properly pre-processed indoor radon data and soil/geological features. Indeed, we believe that the evaluation of factors influencing indoor radon (geological and geochemical parameters, such as soil and rocks radioactivity, permeability, porosity etc.), can be better performed by accounting for spatial effects: spatial autocorrelation and spatial heterogeneity (variation over space of the relationship between variables under study, no longer constant from one location to the next). The occurrence of spatial autocorrelation violates the assumption that the observed value of a variable at one locality is independent of the values of the variable at neighbouring sites. This would imply that many statistical tools for modelling the phenomenon under study may be inappropriate. For instance, ordinary least squares (OLS) regression analysis assumes that observations have been selected randomly whilst the occurrence of spatial autocorrelation reduces the number of independent observations. Accordingly, we chose an appropriate statistical method to account for spatial heterogeneity through local analysis, the GWR (Fotheringham et al., 2002), discussed in detail in Section 3.2.2., in order to quantify the dependence of the response quantity (Rn field) on factors which are known or assumed to be its physical controls. Under spatial heterogeneity we understand here that the observed relationship between variables under study is not longer constant from one location to the next. Heterogeneity can occur at least for two reasons (Fotheringham et al., 2002): there are intrinsic differences in the relationship between variables over space as well as the regression equation might include either incorrect functional forms of the relationships and/or leaves out relevant factors (model misspecification). In addition, the local variation in model parameters can be ascribed to miss-classified factors. Thus, even if the model itself is correct and also the factors are accurately included, the predictors could be wrongly categorized. For instance, in the radon researches, some geological units might have been erroneously classified or its borders inaccurate, owing to limited resolution of the geological map. It is important to notice that local modelling by GWR is particularly sensitive to the representativeness of sample data. More precisely, local deviations from representativeness, which inevitably occur when a sampling plan is designed on a wide geographic scale, can induce biases in local relationships, whereas global modelling shows a greater robustness, as long as the deviations are not themselves a global effect, due to, for example, an inadequate sampling plan. As for our case study, the Rn dataset derives from the union of several subsets of data, each coming from a different measurement survey. Anyway we believe that the standardisation of Rn data with respect to the housing characteristics may help to reduce possible distortive effects attributable to the merging of data generated by different sampling plans. In addition, it should be noted that we have no reason to think that the global distribution of sampling data is really unrepresentative with respect to the spatial distributions of any radon control quantity (environmental or anthropic).

Whatever the cause of the spatial heterogeneity, its inclusion in a modelling approach may enhance the accuracy of results of data exploration since the model being fitted locally is more tuned to local circumstances.

Further, the GWR offers the possibility to promptly generate maps integrating diverse spatial information and visualize how the radon potential factors' influence changes over the geographical space. In particular, the use of this technique in our research allows answering to the following questions with regard to indoor radon–geological factors relationships:

i. Does the GWR model describe the data better than a global model?
ii. Given the data and given the study area is there any statistical evidence for non-random, heterogeneous processes?
iii. To what extent is radon indoor affected by geological and geochemical features?
iv. What factors predict the spatial variance of the regression parameters?

Finally, the basic contribution of this study is to provide, by using the GWR procedure, an explorative analysis of the phenomenon under study, able to identify areas where the relationship between radon measurements and geological characteristic is stronger or weaker than in others. To our knowledge, few attempts have been made to employ this approach in environmental modelling and particularly in the radon mapping context (see Nissi et al., 2012). The credibility and limitations of our multi-stage environmental modelling process is assessed by employing composite performance criteria. Essentially, quantitative metrics ($R^2$, RMSE and AIC), discussed in the following sections, are adopted to detect significant divergences between calibration and testing models performance as well as to determine the relative ranking among analysed models. However, the overall modelling strategy evaluation is also strongly dependent on qualitative considerations, which, as pointed out by Bennett et al. (2013), become crucial in highly complex situations, such as our case study.

We present an application of our method to a dataset related to Abruzzo region, in Italy, including indoor radon measurements, radiometric soil data and geological data. Radon data were collected within several monitoring surveys conducted by the Agency of Environmental Protection (ARTA) of Abruzzo (Palermi and Pasculli, 2008; Nissi et al., 2012). In particular, we refer to the results of a research undertaken in 94 municipalities of L'Aquila district (AQ) which has involved 509 buildings. The radiometric soil data are provided by Bellotti et al. (2007), updated with a few unpublished data recently obtained by ARTA. The geological data (permeability, porosity, fracturation etc.) were provided by integrating not officially released data with geological maps of the Abruzzo and L'Aquila area (Ghisetti and Vezzani, 1998; APAT 2006a,b,c,d).

The paper is organised as follows. Section 2 describes the connection between indoor radon and geology. Section 3 covers the various statistical techniques employed for: (a) pre-processing data, (b) exploring the spatial autocorrelation patterns and (c) implementing the GWR, along with a description of the available software resources. In the Section 4 the case study is illustrated, introducing the geological setting of the area and the dataset (indoor radon, radiometric soil and geological data). Section 5 and 6 are devoted, respectively, to the presentation and the discussion of the main analysis results.