Journal of Environmental Radioactivity 151 (2016) 264–274

Contents lists available at ScienceDirect

Journal of Environmental Radioactivity

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

Geostatistical simulations for radon indoor with a nested model including the housing factor

C. Cafaro [a, *], C. Giovani [b], M. Garavaglia [b]

[a] Department of Physics, University of Trieste, Italy
[b] ARPA-FVG, Natural Radioactivity Lab, Italy

ARTICLE INFO

Article history:
Received 13 March 2015
Received in revised form
3 October 2015
Accepted 4 October 2015
Available online 5 November 2015

Keywords:
Radon
Housing
Geostatistics
Co-kriging
Risk

ABSTRACT

The radon prone areas definition is matter of many researches in radioecology, since radon is considered a leading cause of lung tumours, therefore the authorities ask for support to develop an appropriate sanitary prevention strategy. In this paper, we use geostatistical tools to elaborate a definition accounting for some of the available information about the dwellings. Co-kriging is the proper interpolator used in geostatistics to refine the predictions by using external covariates. In advance, co-kriging is not guaranteed to improve significantly the results obtained by applying the common lognormal kriging. Here, instead, such multivariate approach leads to reduce the cross-validation residual variance to an extent which is deemed as satisfying. Furthermore, with the application of Monte Carlo simulations, the paradigm provides a more conservative radon prone areas definition than the one previously made by lognormal kriging.

1. Introduction

Radon is an α-emitter noble gas. Epidemiological studies have confirmed that the presence of radon in homes increases the risk of lung cancer in the general population. Indeed, radon has been shown to be the second most important cause of lung cancer after smoking in many countries (W. H. Organization, 2009). The carcinogenic risk is typically assumed as related to the radon density of activity concentration (UNSCEAR, 2000; Tirmarche et al., 2010), measured in Bq/m^3 and from now on simply called concentration (C in the graphs).

The main contribution to the radon-related dose is due to indoor concentrations. Therefore, considerable efforts have been made in many countries to create geographic maps of radon risk by using indoor measurements. The maps identify the areas which are expected to be affected by high concentrations, i.e. the so called *Radon Prone Areas* (RPA). The final aim is to suggest a cost-effective risk-reduction strategy to the authorities in charge of the preventive actions.

In this specific case, Friuli Venezia Giulia (FVG), the north-easternmost region of Italy, is under scrutiny. In Italy, RPAs are conventionally defined on municipal basis, for the sake of simplicity and for administrative purposes.

Amongst all the possible ways to infer an RPA map from real measurements, a very direct and reliable one is the so-called Miles' method (Miles, 1998). Alternatively, using geostatistics (Goovaerts, 1997; Chiles and Delfiner, 2009; Cressie and Cassie, 1993) for radon mapping is state of art today, because of its efficiency in mimicking the geological patterns (Dubois, 2005). It is widely known (Nazaroff, 1992) indeed that the soil is the primary source of radon, since radon is exhaled from the uppermost soil layers beneath the basement of a building. The exhalation rate is affected by the geological properties of the soil, hence geostatistics uses their spatial (auto-)correlation structure to infer predictions.

Geostatistics partially overtakes the need for geological maps, by replacing it with a quantitative and more feasible tool, i.e. the variogram. On the other hand, geological knowledge, gathered from expert insight and provided by a geographic information system (GIS), can be helpful, in principle, to refine a radon map. Indeed, the statistical relation with known geological parameters has been confirmed multiple times to hold (Borgoni et al., 2011; Appleton and Miles, 2010).

However, the mere statement that indoor radon concentration is related with geological covariates does not imply necessarily that such covariates can be used to improve a radon map. I.e., in Cafaro et al. (2014) it has been verified that a very common implementation of

* Corresponding author.
E-mail address: shinacronistic@gmail.com (C. Cafaro).

http://dx.doi.org/10.1016/j.jenvrad.2015.10.002
0265-931X/

geologic informations within a geostatistical framework, namely kriging with external drift, does not yield any improvement on this very same dataset. In other cases, instead, such technique (or slightly different ones) has yielded better results (Dubois et al., 2007; Bossew et al., 2008; Hauri et al., 2012; Cinelli et al., 2011).

Arguably, the failure may be typical of specific geographic regions, i.e where the geologic covariates affecting concentrations do not identify different uranium soil concentration, but, at most, different permeabilities. Such feature is very sensitive to even very small variations of the meteorological conditions. Furthermore, it rarely influences, even under homogeneity conditions, the entrance into an indoor environment with a process well summarized by a categorical classification, i.e. the one provided by a geological map.

Therefore, in this case it has been chosen to follow an alternative approach, involving the so-called *housing factor*, i.e. a “proxy” of the dwelling characteristics. Indeed, the spatial variability of radon concentration is influenced by geology as much as by its conditions of entrance inside the indoor environment, collectively referred to as housing. Removing such proxy is a common theoretical attempt and it leads to the definition of a (*geogenic*) *radon potential* (Gruber et al., 2013), sometimes described as the radon concentration we would expect if the house was a “standard” one (see, e.g. Friedmann et al., 1996).

The housing problem is relatively overlooked in the applied geostatistical literature about radon, beside very few exceptions (Smith and Field, 2007). It is usually deemed as a mere source of local shocks, therefore the principal cause of the prevailing nugget effects in the radon variograms. The very definition of a radon potential comes from such statement and it is clearly a strategy to “normalise away” the housing characteristics.

The local nature of housing follows by intuition and it is substantiated by evidence, though indirectly. E.g., it is pointed out by the nugget differences between schools-related variograms and the analogous dwellings-related variograms, computed by measurements collected over a comparable timespan and within the same region. The nugget effect is, indeed, typically lower for schools than for dwellings (Bossew et al., 2014).[1] A possible reason is that the building criteria of schools, as public facilities, follow very stringent rules (mainly due to safety issues). Thus, the schools are usually more “standard” structurally than private dwellings. As far as living habits are concerned, the very same rationale can be applied, since schools are working places, hence living habits can be considered quite homogeneous in a relatively small region like FVG.

The starting point of this paper is that the housing factor can be used to improve radon predictions, accordingly with the aforementioned premises. Unfortunately, the only way to define housing operatively is through a categorical classification, hence it suffers of the very same, if not worse, limits of the geogenic contribution. Nonetheless, the attempt can be worthwhile, since the variability of indoor radon can be affected by housing more than geology (Borgoni et al., 2014). In those cases, a small amount of information, provided by a very coarse housing classification, could yield a greater improvement than a slightly less coarse geologic information.

2. Dataset description

This dataset is the same one thoroughly described in Cafaro et al. (2014), i.e. a database of 2452 annual concentrations, with relative questionnaires, collected during the RPA-FVG survey in a timespan between 2005 and 2006. The measurements were performed through passive detectors (CR39).

Summary statistics are displayed in Fig. 1 and Table 1. As largely verified in the literature (Nero et al., 1986; Bossew, 2010), they confirm that radon follows an approximately lognormal distribution (Allen et al., 2001; Limpert et al., 2001; Mitzenmacher, 2004).

The database had been designed as *stratified* and *preferential*, i.e. higher sampling where higher concentrations are expected. The stratification was made through an intermediate geographical layer called *carta tecnica regionale numerica* (*CTRN*), which is an almost-squared partition of the region (Fig. 2, right). Since logistical issues occurred at the time of the actual positioning, the stratification, and its related preferentiality, has partially disappeared, resulting in a dataset affected mainly by natural clustering. However, a very weak preferentiality has still been verified, though it is not part of the present analysis.

Such composite clustering is the main reason to perform geostatistical interpolation on a regular grid. Indeed, the target areas, i.e. the municipalities (Fig. 2, left), are very different from the CTRN cells as well as very diverse in size, resulting in biased direct areal estimations. This problem is referred to in the literature as *change of support* (Gotway and Young, 2002).

3. Models and methods

3.1. Model hypotheses

Theoretically, the concentration is usually assumed to be decomposable in two leading contributions, represented by geological and housing features. Under lognormal hypothesis, a fair decomposition would then be

$$Z(x) = \log(C(x)) \sim G(x) + H(x), \tag{1}$$

where all the different factors relative to either contribution has been aggregated into a single function (Dubois et al., 2007). x is evidently the 2D vector of coordinates. G is usually referred to as *geogenic* component. It is usual practice to decompose H further (e.g. see the famous argument on lognormality of indoor radon in Gunby et al. (1993)) and separately consider each component.

Eq. (1) is hardly realistic, because it considers the influence of housing as a mere multiplicative factor (in linear scale). Furthermore, it does not contemplate the possibility of synergy between the components. In this paper, instead, the radon log-concentration is defined as

$$Z = \sum_{\alpha} \mathbf{1}^{\alpha} Z^{\alpha}, \tag{2}$$

where $\mathbf{1}^{\alpha}$ is the standard indicator function defined as

$$\mathbf{1}^{\alpha}(x_i) = \begin{cases} 1 & \text{if} \quad H(x_i) = \alpha \\ 0 & \text{if} \quad H(x_i) \neq \alpha \end{cases}. \tag{3}$$

Namely, the variables Z^{α} s are defined as *the logarithmic concentration relative to a house of class* α. The classes are obviously a simplification of a possibly continuous family. Their definition will be derived in the result section.

Such model, though seemingly an excess of abstraction, is instrumental for the consequent geostatistical application. It allows to deal with the geogenic part by geostatistics, without further refinement, while focusing the analysis on H.

3.2. Cross-variogram

Beside standard statistical tests, the present paper uses mainly multivariate geostatistics (Castrignanò et al., 2000; Clark et al.,

[1] This very same feature, i.e. a strikingly lower nugget, has been verified by the authors to hold for FVG schools as well. Schools are the workplaces most continuously monitored by ARPA (Regional Environmental Protection Agency) within the region.