Math Geosci (2013) 45:437–452
DOI 10.1007/s11004-013-9443-8

Stochastic Simulation Model for the Spatial Characterization of Lung Cancer Mortality Risk and Study of Environmental Factors

Ana Rita Oliveira · Cristina Branquinho · Maria Pereira · Amílcar Soares

Received: 5 March 2012 / Accepted: 13 January 2013 / Published online: 27 March 2013

Abstract This paper presents a study in which the lung cancer risk in males was characterized based on a simulation model of mortality rates. Block sequential simulation of mortality rates, measured in counties of different sizes, was implemented and applied to a normal grid of continental Portugal with high spatial resolution. The uncertainty in the mortality rate measurements, directly related to differences in the population size of each county, was integrated in a block direct sequential simulation through Poisson kriging of local means and variances. Three age groups were examined: 50–59, 60–69, and 70–79 years. After the continuous geographic patterns of lung cancer risk were obtained, factors potentially associated with the main areas of risk were analyzed for southern Portugal. Thus, a defined class of land use and dry weather events, related to airborne particulate matter, were found to be associated with high-risk areas, resulting in high local spatial correlation patterns in all three age groups.

Keywords Block sequential simulation · Cancer risk · Dry land · Drought

1 Introduction

The recognition of geographic patterns in diseases such as lung cancer has stimulated the use of mapping techniques to analyze their spatial distribution. As a result, spatial models of health data have come to play an important role both in studies on the

A.R. Oliveira (✉) · M. Pereira · A. Soares
CERENA, Instituto Superior Técnico, Universidade Técnica de Lisboa, Av. Rovisco Pais, 1049-001, Lisboa, Portugal
e-mail: rita.oliveira@ist.utl.pt

C. Branquinho
Centro de Biologia Ambiental, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, 1749-016, Lisboa, Portugal

incidence of various pathologies and in those of mortality with respect to environmental, demographic, and other covariates. Geostatistics, as a practical tool for the spatial characterization of health data, has been used in a wide variety of environmental applications (Soares et al. 1997; Gomez-Hernandez et al. 1999; others) and more recently in health geography (Goovaerts 2005). In addition, there have been a number of developments related to the geostatistical methodologies relevant to this field (Goovaerts 2005). One of the more recent applications of geostatistics to health data is the mapping of cancer data, especially in terms of the risk of developing cancer. Cancer data usually report the number of occurrences or the mortality rate recorded in a parish, county, district, or other geographic division. However, there are a number of difficulties, which in turn pose several challenges in the application of the usual geostatistical framework to these kinds of data. For example, there is the interest in computing the risk based on another variable, that is, mortality. Also, the existence of extreme relative risks in a small population leads to less reliable data (Goovaerts 2004). Finally, the data are usually collected in aggregate form (e.g., mortality rates are often aggregated over large geographic supports, such as counties or districts) and recorded over irregular spatial supports (Goovaerts 2006b).

Geostatistical methods to map cancer risk were first developed by Oliver et al. (1998) to estimate the risk of childhood cancer. In their work, risk was estimated with binomial cokriging, based on the assumption that cancer counts (conditioned to a fixed-risk function) followed a binomial distribution in which the parameters are the risk of developing cancer and the number of people at risk (Oliver et al. 1998; Goovaerts 2010). In 2005, Goovaerts adapted Poisson kriging to analyze cancer data (Goovaerts 2005). That study expanded on the seminal work of Monestiez et al. (2004, 2006), in which the authors used Poisson kriging to map the spatial distribution of wild animal species based on infrequent and sparse sightings and on spatially heterogeneous observation efforts (Monestiez et al. 2004, 2006). In the geostatistical approaches referred to above, the geographic entities of mortality rates (or death counts) are referenced by their centroids. However, this does not account for the irregular shape of the geographic entities of cancer data, and it assumes that the entities have the same geographic size and shape. Therefore, in subsequent work, Goovaerts (2006b) incorporated the size and shape of the geographic entities in Poisson kriging through methodologies, referred to as area-to-point and area-to-area kriging that were originally developed to predict punctual or areal values from areal data (Kyriakidis 2004). While all these methodologies show the most probable dispersion map of risk, they do not succeed in characterizing the spatial uncertainty of such maps and the probability of occurrence of extreme situations in given areas. In health studies, spatial uncertainty and extreme characteristics are important considerations if the goal is to understand the associations among the factors potentially causing abnormally high mortality. However, the use of traditional geostatistical methods to characterize spatial uncertainty, that is, stochastic simulations, has in this setting two drawbacks. Firstly, as there are no risk data (i.e., the available data are the mortality rates) the probability distribution law of the risk needed to condition the simulations must be inferred from another variable. Secondly, since the data are available at different spatial supports, any simulation approach cannot involve non-linear transformations (such as Gaussian) of the main variable to estimate the parameters (mean, variance) of the local distributions. In recent work, Goovaerts (2006a, 2009), after estimating the local

means and variances of the risk (assumed to be the mean and variance of Poisson kriging), assumed that for all nodes of a regular grid there is a Gaussian distribution with those two parameters. The P-field simulation was then used to produce spatial images of the risk, and consequently, to access the uncertainty (Goovaerts 2006a). The main disadvantages of this method are the assumption of local Gaussian distributions of risk and the well-known inconvenience associated with the use of P-field simulations (Pyrcz and Deutsch 2001).

In this work, the use of the block sequential simulation (BSSIM) algorithm (Liu and Journel 2009) to simulate lung cancer risk in continental Portugal and to obtain the associated spatial uncertainty is examined. This approach is based on direct sequential simulation (DSS) (Soares 2001), which does not require a non-linear transformation of the main variable; hence, different support data can be jointly used in the same model. As in Poisson kriging, the noise that results from the population size is integrated through the addition of an error variance term in the block kriging system. In addition, the global distribution of risk is estimated using the data on mortality rates. The final part of this study was aimed at identifying the associations of potential environmental risk factors that may account for the geographic patterns of extreme risks of lung cancer. We therefore chose the following environmental factors, related either directly or indirectly to airborne particle levels, to describe their association with lung cancer risk: dry climate conditions (RL10), defined as the number of days per year in which the amount of precipitation is lower than 10 mm, and land-cover classes, which can originate or mitigate the dispersion of airborne particles.

2 Methodology

2.1 Health Data

The data set used in the present work consisted of the mortality rates related to malignant tumor of the trachea, bronchi, and lung of male residents in continental Portugal. The data set, recorded at the county level, corresponded to the period between 1980 and 2008 and was obtained from Instituto Nacional de Estatística (INE). The methodology proposed in this work and described below was applied only to individuals between the ages of 50 and 79 years, specifically, on the following age groups: 50–59, 60–69, and 70–79 years, for which cancer is considered a rare event.

2.2 Environmental Factors Data

The two potential environmental risk factors considered were land cover class and the drought index RL10; the latter is defined as the number of days per year in which precipitation amounts are lower than 10 mm. These two factors were chosen based on their likely relation to airborne particulate matter, with potential sources especially being areas without vegetation or trees and susceptible to drought. RL10 was assessed in southern Portugal during the period 1961–2000. Data on the spatial dispersion of drought during that period (Durão et al. 2010) were used. Specifically, RL10 values were estimated based on a regular grid of 5×5 km^2 for 10-year periods, with reference to the data of 105 monitoring stations and the observed daily

precipitation recordings obtained from the National System of Water Resources Information (http://snirh.inag.pt) database. Then the RL10 40-year average was computed to study the association between the drought index and lung cancer risk. Information on the land-cover class potentially favoring the dispersion of airborne particles that is dry land was obtained from the CORINE Land Cover 1990 map of Portugal (http://www.igeo.pt).

2.3 Geostatistical Analysis of Cancer Data: Block Sequential Simulation in the Characterization of Risk Cancer Uncertainty

Consider $c(\mathbf{u}_\alpha)$, the number of mortality cases recorded in a certain entity (e.g., parish, county), referenced by its centroid $\mathbf{u}_\alpha$, and the size of the population at risk, $n(\mathbf{u}_\alpha)$. The mortality rate $z(\mathbf{u}_\alpha)$ can be written as

$$z(\mathbf{u}_\alpha) = \frac{c(\mathbf{u}_\alpha)}{n(\mathbf{u}_\alpha)}. \tag{1}$$

In the Poisson kriging model of cancer risk (Goovaerts 2005, 2009), the disease count $c(\mathbf{u}_\alpha)$ at each location $\mathbf{u}_\alpha$ is interpreted as a realization of a random variable $C(\mathbf{u}_\alpha)$ that follows a Poisson distribution with the parameter of the expected number of counts. This parameter is the product of the population size, $n(\mathbf{u}_\alpha)$, and the local risk, $R(\mathbf{u}_\alpha)$. The expectation of risk at any location is equal to the expectation of the mortality rate

$$E\big[Z(\mathbf{u}_\alpha)\big] = E\big[R(\mathbf{u}_\alpha)\big] = m. \tag{2}$$

The variance in the mortality rate is equal to the variance of the risk plus a term related to the dimension of the population

$$\mathrm{Var}\big[Z(\mathbf{u}_\alpha)\big] = \mathrm{Var}\big[R(\mathbf{u}_\alpha)\big] + E\big[R(\mathbf{u}_\alpha)/n(\mathbf{u}_\alpha)\big] = \sigma_R^2 + m/n(\mathbf{u}_\alpha). \tag{3}$$

The purpose of this work, which was based on this model, was to determine the uncertainty of cancer risk through a stochastic simulation methodology capable of integrating the different spatial supports comprising the data, that is, the geographic dimension of each county, and the uncertainty affiliated with the mortality rate, through the size of the population.

2.3.1 Block Sequential Simulation

The prerequisite of the majority of simulation algorithms (e.g., sequential simulations) is an estimate of the local parameters of the distribution of the main variable $Z(\mathbf{x})$. As block data are a linear average of their constituent point data, integration of the different support data in the geostatistical estimation of local means and variances implies the use of original variables such as in DSS (Soares 2001). Non-linear transformation of the original $Z(\mathbf{x})$, as in the Gaussian transformation of a sequential Gaussian simulation, cannot be applied. Block sequential simulation (BSSIM), developed by Liu and Journel (2009), is an extension of DSS that does not require the transformation of $Z(\mathbf{x})$. Furthermore, both the local mean and the variance of the global distribution law are estimated with block kriging. The algorithm was developed in order to integrate data with different volume supports, such as fine-scale support data (point data) and coarse-scale support data (block data) (Liu and Journel 2009).

Kriging with Block Data: Block simple kriging is applied to estimate the local mean and variance of $Z(\mathbf{x})$. This is a kriging technique that accounts simultaneously for point and block data (Liu and Journel 2009), with block data, $B(\boldsymbol{v}_\alpha)$, defined as the spatial linear average of point values $P(\mathbf{x}')$ within the block volume $\boldsymbol{v}_\alpha$

$$B(\boldsymbol{v}_\alpha) = \frac{1}{|\boldsymbol{v}_\alpha|} \int_{\boldsymbol{v}_\alpha} L_\alpha\big(P(\mathbf{x}')\big) d\mathbf{x}' \quad \forall \alpha, \tag{4}$$

where L_α is a known linear averaging function. The simple kriging estimator $Z^*_{SK}(\mathbf{x}_0)$ is conditioned to both point and block data and is written as

$$Z^*_{SK}(\mathbf{x}_0) - \boldsymbol{m}_0 = \Lambda^t \cdot \boldsymbol{D} = \sum_{\alpha=1}^{n(\mathbf{x}_0)} \lambda_\alpha(\mathbf{x}_0) \cdot \big[D(\mathbf{x}_\alpha) - m_0\big], \tag{5}$$

where $\Lambda^t = [\lambda_P \ \lambda_B]$ is the kriging weights for point data (P) and block data (B), $\boldsymbol{D}^t = [\boldsymbol{P} \ \boldsymbol{B}]$ is the data value vector, $D(\mathbf{x}_\alpha)$ is a specific datum at location $\mathbf{x}_\alpha$, $n(\mathbf{x}_0)$ denotes the number of data, and m_0 denotes the stationary mean.

Kriging weights Λ are the solution of the system of linear equations of the kriging system

$$\boldsymbol{K} \cdot \Lambda = \boldsymbol{k}, \tag{6}$$

where $\mathbf{K}$ is the data-to-data covariance matrix (point–point, average covariance point–block, and average covariance block–block), defined as

$$\boldsymbol{K} = \begin{bmatrix} \boldsymbol{C}_{PP} & \overline{\boldsymbol{C}}_{PB} \\ \overline{\boldsymbol{C}}^t_{PB} & \overline{\boldsymbol{C}}_{BB} \end{bmatrix}, \tag{7}$$

and $\mathbf{k}$ is the data-to-unknown covariance matrix (point–point and average covariance point–block), defined as

$$\boldsymbol{k} = \begin{bmatrix} \boldsymbol{C}_{P P_0} \\ \overline{\boldsymbol{C}}_{B P_0} \end{bmatrix}. \tag{8}$$

$\boldsymbol{C}$ is the point covariance submatrix, $\overline{\boldsymbol{C}}$ the covariance submatrix involving a block support, and $\mathbf{x}_0$ the point to be estimated. The simple kriging variance is

$$\sigma^2_{SK}(\mathbf{x}_0) = \mathrm{Var}\big\{Z(\mathbf{x}_0) - Z^*_{SK}(\mathbf{x}_0)\big\} = C(0) - \Lambda^t \cdot \boldsymbol{k}, \tag{9}$$

where $C(0) = \mathrm{Var}\{Z(\mathbf{x}_0)\}$ is the stationary variance.

Since there is noise/uncertainty attached to mortality rates, resulting from the size of the population (Goovaerts 2004), the population size can be used to quantify the uncertainty, through an error variance term, $m^*/n(\mathbf{u}_\alpha)$, for zero distance covariances [see Eq. (3)], which can be introduced into Eq. (7), thus transforming it into a Poisson kriging system (Goovaerts 2005). Note that in our study cancer mortality is considered to be a rare event in all age groups, even the oldest one. Otherwise a binomial kriging approach for the error variance (Goovaerts 2010) could be adopted.

Block Sequential Simulation Workflow: The workflow of the BSSIM algorithm can be described in the following steps:

1. Define the simulation visiting path of each node x of the grid (fully random or block-first).
2. For each node x, search the conditioning data (closest original point data, previously simulated values, and block data).
3. Compute or retrieve the local covariance values: block-to-block, block-to-point, point-to-block, and point-to-point; build and solve the block kriging system and obtain the kriging estimate and variance.
4. Select a value from the global probability distribution function and add the simulated value to the data set.
5. Repeat steps (i) to (iv) until all grid nodes are simulated.
6. Repeat steps (i) to (v) until all realizations are generated.

The estimate of the global probability distribution function of the risk (needed for step (iv)) is presented in Sect. 2.3.

2.3.2 Semivariogram of Cancer Risk

The semivariogram describes the main spatial continuity patterns of the risk. In this work, we applied an adaptation of the semivariogram proposed by Monestiez et al. (2004, 2006), which uses weights, $w(\mathbf{h})$, to account for the population size. The experimental semivariogram was then calculated as follows (Goovaerts 2005)

$$\hat{\gamma}_R(\mathbf{h}) = \frac{1}{2\sum_{\alpha=1}^{N(\mathbf{h})} w(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} \left\{ w(\mathbf{h}) \left[z(\boldsymbol{v}_\alpha) - z(\boldsymbol{v}_\alpha + \mathbf{h}) \right]^2 - m^* \right\}, \tag{10}$$

where m^* is the population-weighted mean of the mortality rates, $\mathbf{h}$ is the distance vector between paired points, and $w(\mathbf{h})$ is defined as

$$w(\mathbf{h}) = \frac{n(\boldsymbol{v}_\alpha) n(\boldsymbol{v}_\alpha + \mathbf{h})}{n(\boldsymbol{v}_\alpha) + n(\boldsymbol{v}_\alpha + \mathbf{h})}. \tag{11}$$

The vector $\mathbf{h}$ of the experimental semivariogram was calculated taking into account the spatial support of each entity, $\boldsymbol{v}_\alpha$, in which the distance between any two block data is given by the average distance

$$\mathrm{Dist}(\boldsymbol{v}_\alpha, \boldsymbol{v}_\beta) = \frac{1}{N_\alpha N_\beta} \sum_{i=1}^{N_\alpha} \sum_{j=1}^{N_\beta} \| \mathbf{u}_i - \mathbf{u}_j \|, \tag{12}$$

where N_α and N_β are the number of points $\mathbf{u}_i$ and $\mathbf{u}_j$ used to discretize the two blocks, $\boldsymbol{v}_\alpha$ and $\boldsymbol{v}_\beta$, respectively. As there are no point data, the variogram's point–point must be inferred from block data. Several algorithms were suggested by Goovaerts (2006b, 2008) and Kyriakidis (2004). In these cases, the block size (dimensions of the counties) is small compared to the range of the variograms; hence, the

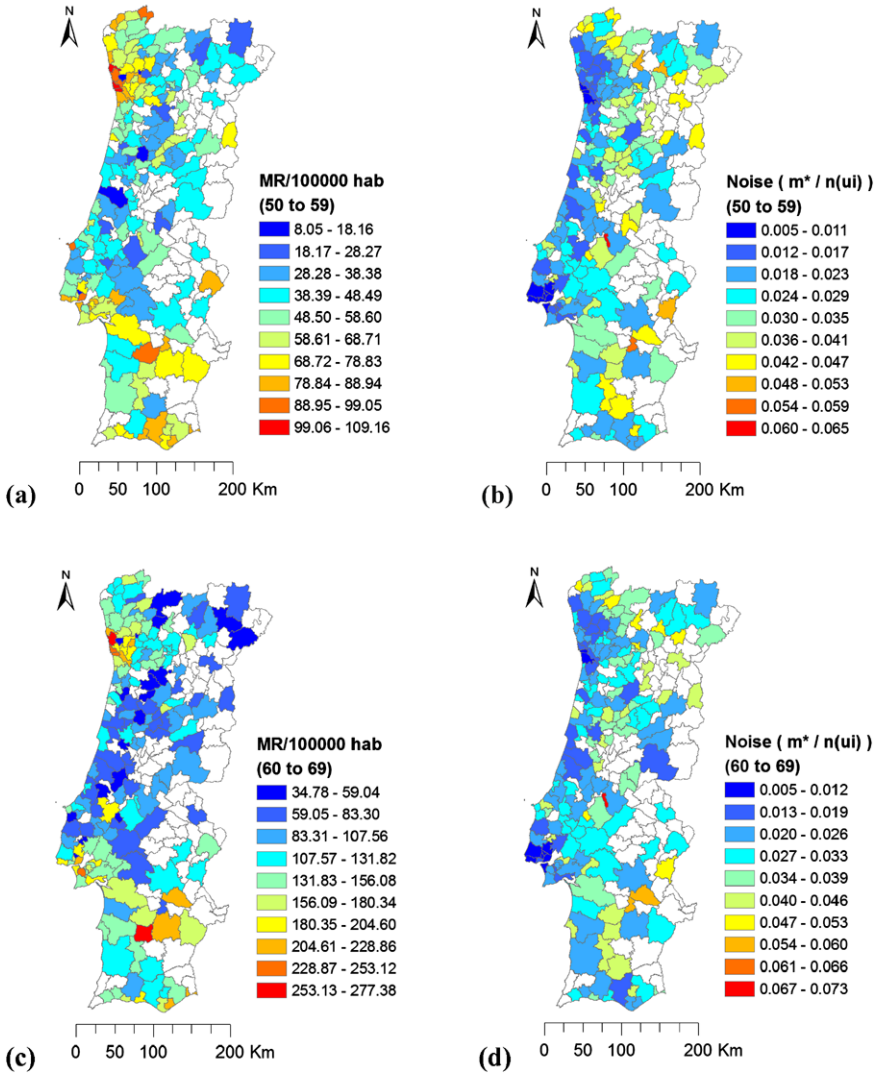

Fig. 1 Original data: mortality rates (*left maps*) and population size (*right maps*) for each age group. Maps (**a**), (**c**), and (**e**) are the mortality rates of individuals between 50–59, 60–69, and 70–79 years of age, respectively. Maps (**b**), (**d**), and (**f**) are the error variance term used ($m^*/n(\mathbf{u}_i)$) for the same three age groups, respectively

classical approach (Journel and Huigbreghts 1978) can be used, in which the block–block variogram is equal to the point–point variogram minus a constant, which is the average variogram inside the block

$$\gamma_v(h) = \gamma(h) - \overline{\gamma}(v, v). \tag{13}$$

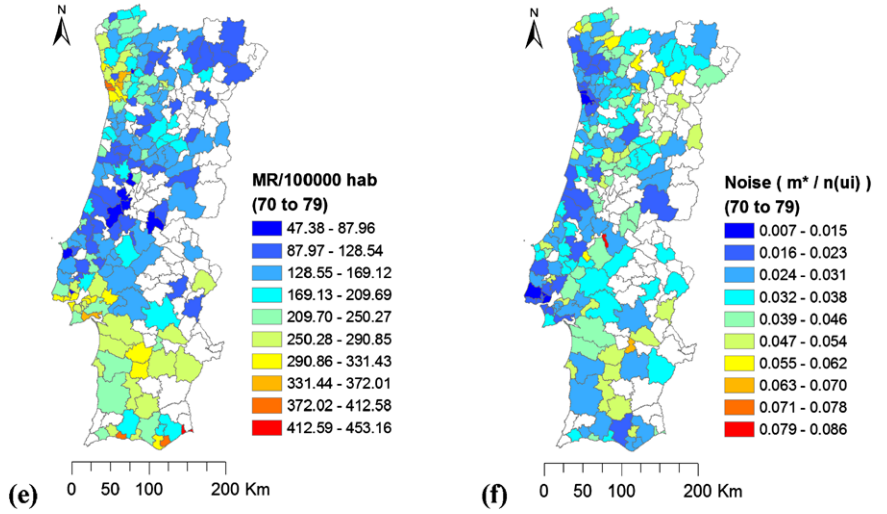

Fig. 1 (*Continued*)

2.3.3 Global Probability Distribution Function of Cancer Risk

To simulate the values of risk, an estimate of its global probability distribution function (pdf) is needed in our study. This value was estimated using the data on mortality rates. The expectation that the risk will be lower than any threshold t is equal to the equivalent expectation of the mortality rates

$$E\big[I_r(x,t)\big] = E\big[I_z(x,t)\big] \tag{14}$$

$$I_r(x,t) = \begin{cases} 1 & \text{if } r(x) < t \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

$$I_z(x,t) = \begin{cases} 1 & \text{if } z(x) < t \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

The global pdf $\text{prob}(z(x) < t) = E[I_z(x,t)]$ can be estimated by the global block kriging of indicator $I_z(x,t)$ for different thresholds t. Block indicator kriging accounts for the spatial dispersion of sample data and for the error in the dimension of the population. In this study, the following approximation was performed; it accounted only for the error in the dimension of the population, a globally weighted average in which the weights are determined by the population number $n(i)$ of each of the N counties

$$\text{prob}\big[z(x) < t\big]^* \approx \frac{1}{\sum_{i=1}^{Nt} n(i)} \sum_{i=1}^{N_t} n(i) I_z(i,t). \tag{17}$$

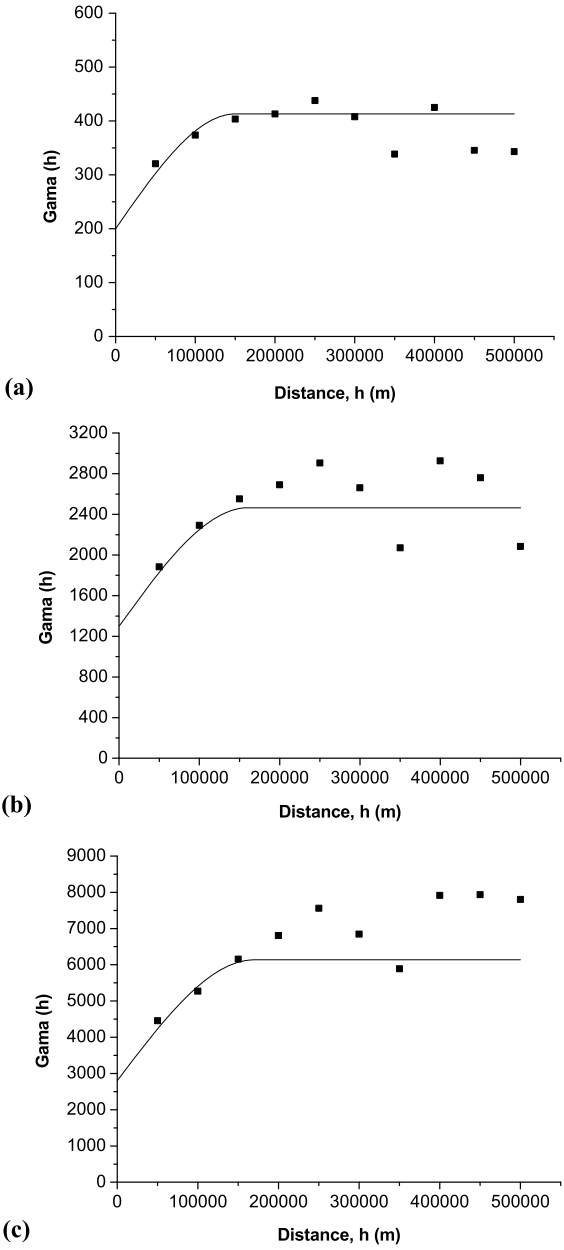

Fig. 2 Omnidirectional semivariograms for each age group, with the spherical model fitted. For the age group (**a**) 50–59 years, the model parameters are: $C = 213$, $a = 150{,}000$ m, and nugget effect = 200; (**b**) 60–69 years: $C = 1165$, $a = 160\,000$ m, and nugget effect = 1300; (**c**) 70–79 years: $C = 3338$, $a = 170\,000$ m, and nugget effect = 2800

3 Results and Discussion

The original data, mortality rates, and the noise related to the population size are shown in Fig. 1(a) to 1(f). Portugal has 278 counties. The 101 counties with less than three recorded cases of lung cancer were dismissed for confidentiality reasons. The experimental semivariograms (Eq. (10)) of this study are shown in Fig. 2, with

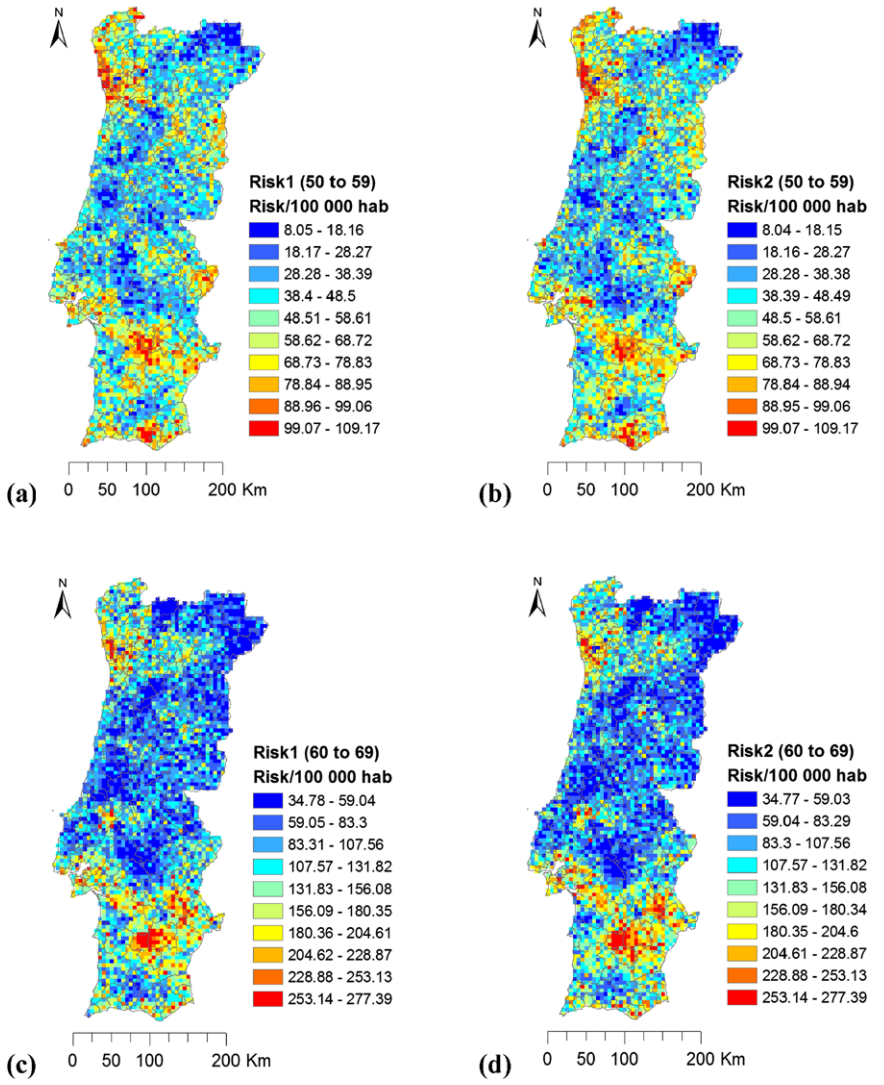

Fig. 3 Two examples of the realizations of lung cancer risk. Maps (**a**) and (**b**) pertain to the age group 50–59 years; maps (**c**) and (**d**) to the age group 60–69 years; and maps (**e**) and (**f**) to the age group 70–79 years

the spherical model fitted. The variogram ranges are much larger than the average dimensions of the counties (31,961 ha). Hence, the term $\overline{\gamma}(v, v)$ is considered to be negligible in order to make the approximation $\gamma_v(h) \cong \gamma(h)$ acceptable. The BSSIM algorithm was used to compute 50 realizations of lung cancer risk for each of the three age groups (50–59, 60–69, and 70–79 years). Two examples of the realizations of lung cancer risk for the three groups are shown in Fig. 3(a) to 3(f). The E-type estimator of cancer risk determined for all the simulations is presented in Fig. 4(a)

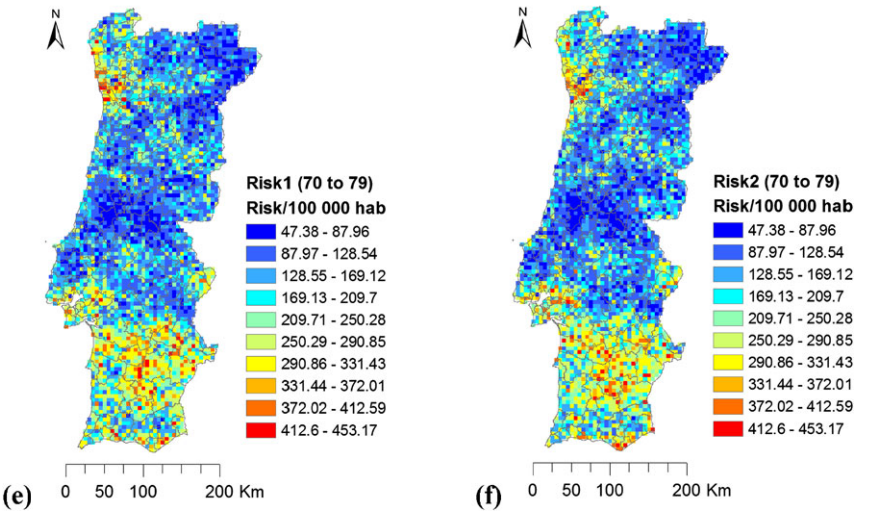

Fig. 3 (*Continued*)

to 4(c). It is clear that in southern Portugal there is an identical pattern of high risk for all age groups. The ability to evaluate the spatial uncertainty is the main advantage of using stochastic simulations in the spatial characterization of lung cancer risk. For illustrative purposes, the probability that any point will exceed the mean was computed for the age group 60–69 years, in which the average mortality rate is 120.17 persons per 100,000 habitants per year (Fig. 5). Several remarks can be made concerning the average risk maps and the spatial uncertainty of the three age groups:

(i) For all age groups, two separate areas with high risks of lung cancer were identified. The first of these corresponds to the two most populated urban areas, Lisbon and Oporto, and their corresponding counties, and the second to a large area in southern Portugal (Alentejo region), with a relatively low population density.
(ii) The Alentejo shows a very large and continuous pattern of high risk for all age groups.

The spatial continuity and the extent of the high-risk geographic patterns in southern Portugal for all age groups are so remarkable as to suggest, in principle, that some spatial phenomena may be associated with that spatial pattern.

As is well known, tobacco smoke is the most important risk factor in lung cancer. In Portugal, while there are no records for tobacco use at the county level, data are available for the country's five regions (Fig. 6). The presented map is based on data obtained in a national survey carried out at the regional level by the Instituto Nacional de Saúde Dr. Ricardo Jorge, which aims to study tobacco use in the Portuguese population (Machado et al. 2009). However, there is no reason to assume that tobacco use is higher in one area than another and it follows the spatial pattern shown in Fig. 6. Thus, in the high-risk areas of lung cancer (large urban areas) tobacco use is low to medium. Nonetheless, the association between environmental factors and lung cancer risk has thus far been studied assuming uniform tobacco use across the

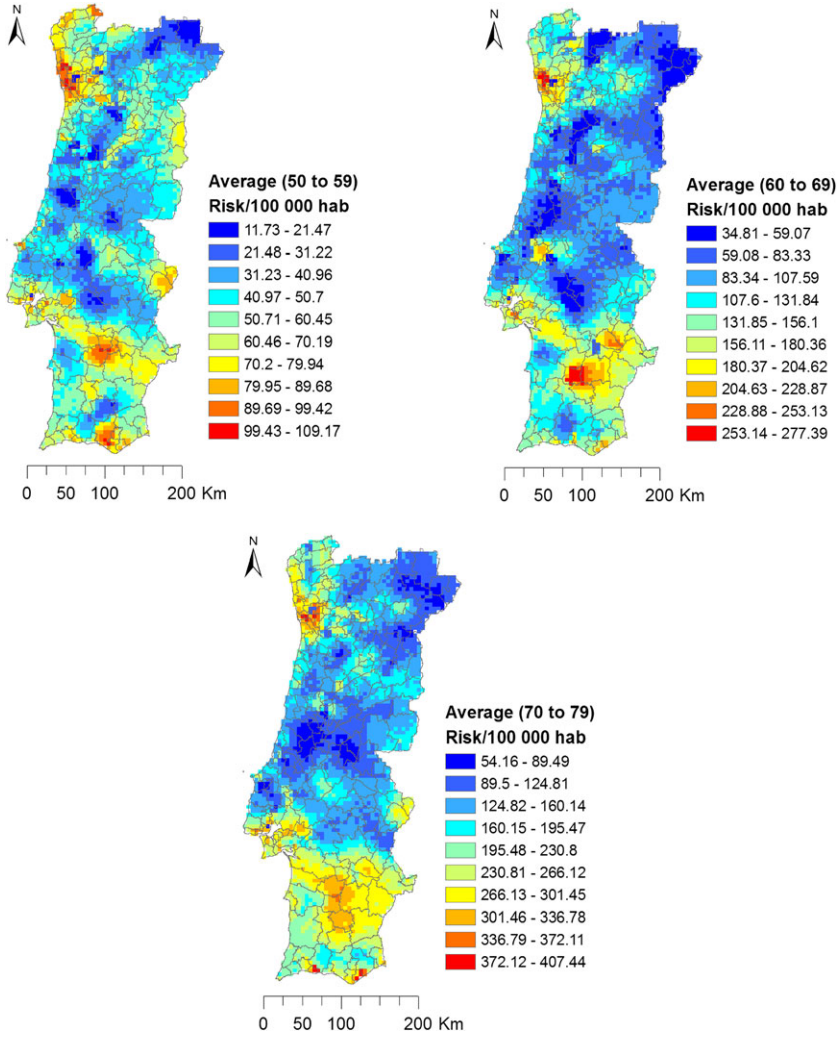

Fig. 4 E-type maps for the age groups 50–59, 60–69, and 70–79 years

southern part of the country. Since the Alentejo is the largest area with an extremely high risk of cancer, it was the only one further examined for an association between lung cancer and environmental factors. Accordingly, two environmental factors potentially related to the levels of airborne particulate matter, that is, the drought index RL10 and presence of dry land, were selected.

After the CORINE Land Cover 1990 Portugal map was scaled to a 5×5 km^2 grid, a map showing the percentage of dry land was constructed. Next, moving averages within a 20-km radius were calculated. The other environmental factor considered, also with a 5×5 km^2 resolution grid, was the 40-year average of the RL10 index. Figure 7 presents the environmental factors analyzed for southern Portugal. To determine whether lung cancer was associated with the chosen environmental factors, local

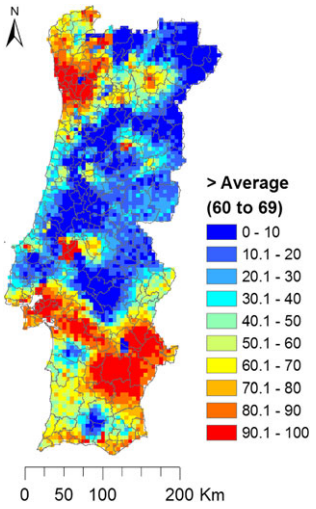

Fig. 5 Probability that the mortality rate of individuals in the age group 60–69 (for the 50 realizations) will exceed the average (**120.17** persons per 100,000 habitants per year)

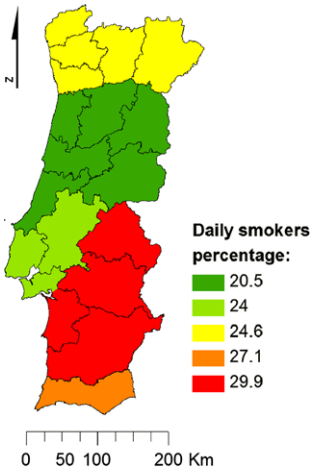

Fig. 6 Percentage of daily smokers in continental Portugal, recorded by region, for the period 2005/2006 (Machado et al., 2009)

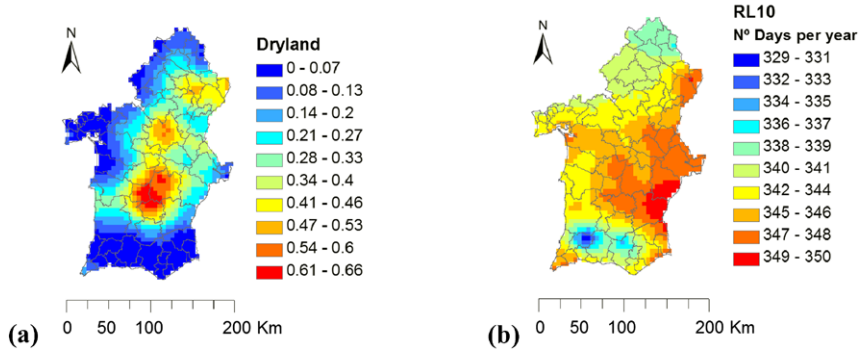

Fig. 7 Analyzed environmental factors potentially related to lung cancer: (**a**) dry land and (**b**) RL10

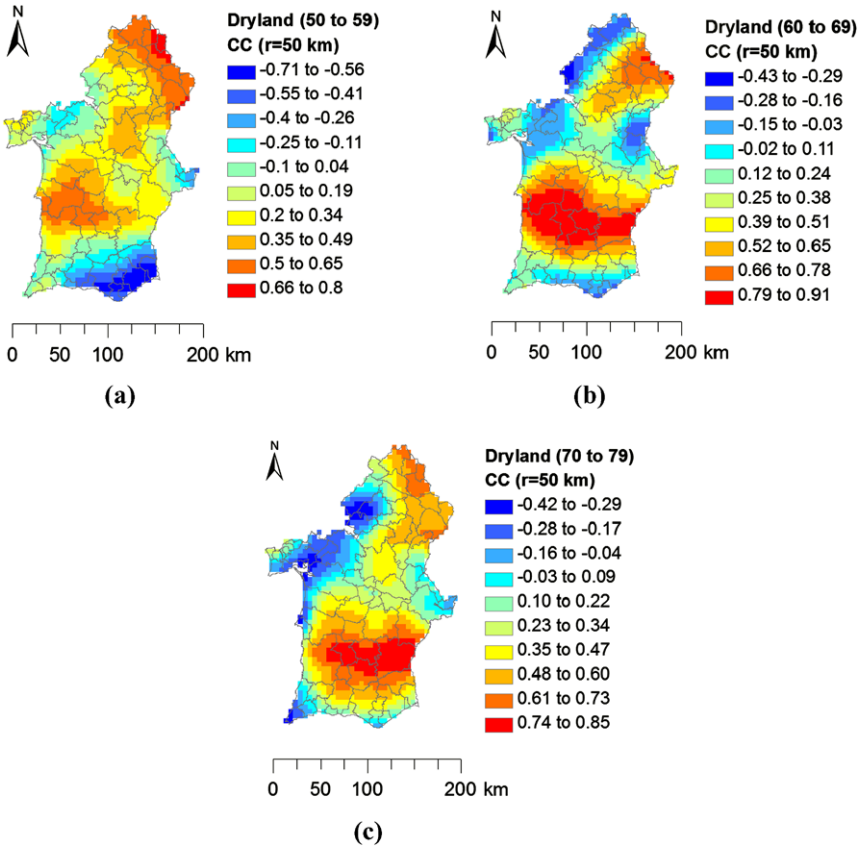

Fig. 8 Local correlation maps between E-type cancer risk map and dry land within a 50-km radius. Maps (**a**), (**b**), and (**c**) show the results for the age groups 50–59, 60–69, and 70–79 years, respectively

correlation maps (Pearson correlation coefficient) on a moving window of 50 km were computed (Figs. 8 and 9). The association between lung cancer and dry land (Fig. 8) correlated well with cancer risk in certain areas. For the RL10 index, the results (Fig. 9) also suggested good local correlations with lung cancer, especially for individuals between the ages of 60 and 69. The correlation maps illustrate the areas in which good local correlations for all three age groups were determined. The obtained results indicate that the chosen environmental factors can provide explanatory association factors for lung cancer mortality and/or enhance the role of tobacco smoke effects in southern Portugal.

4 Conclusion

In this work, block sequential simulation was used to assess the risk of cancer (of the lung, trachea, and bronchi). Additionally, the association between lung cancer risk and two environmental factors was evaluated. The results demonstrate the advantages

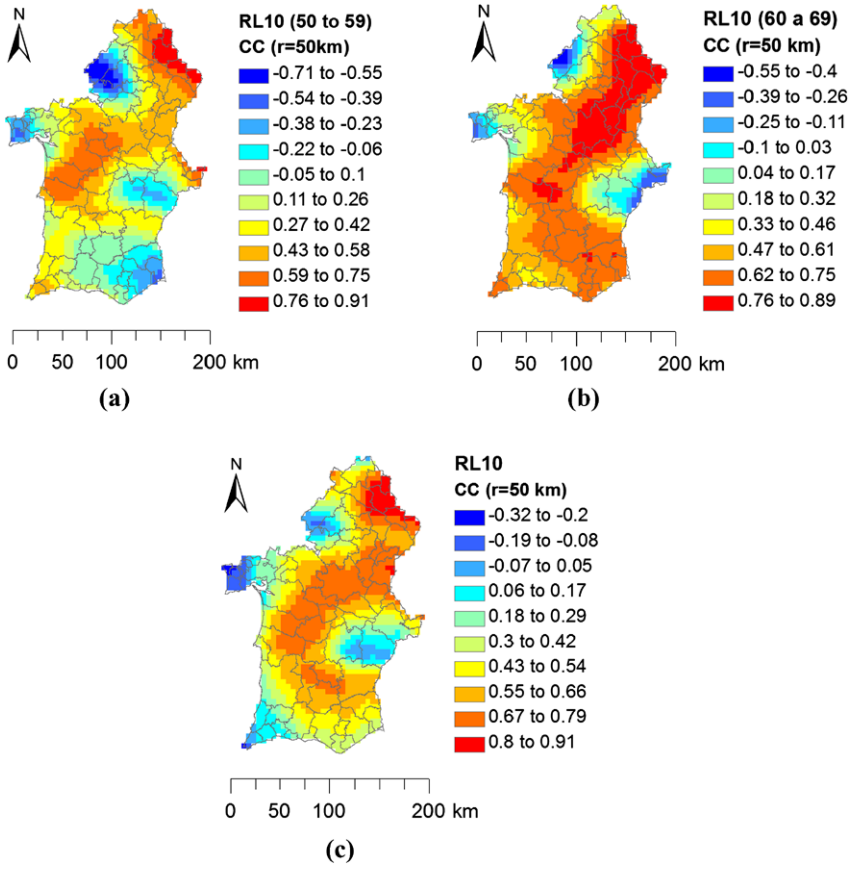

Fig. 9 Local correlation maps between E-type cancer risk map and the RL10 index within a 50-km radius. Maps (**a**), (**b**), and (**c**) show the results for the age groups 50–59, 60–69, and 70–79 years, respectively

of using block sequential simulation to map cancer risk. First, the algorithm accounts for both the spatial location and the size of the entities (here, counties), thus overcoming the problem of data measured over different and irregular supports. Moreover, discretization of the blocks is possible, which facilitates the study of causative factors with generally different scale supports, such as air pollution or other environmental factors. BSSIM also allows for the assessment of uncertainty, through the study of extreme risk values, which is very important for public health management policies. The maps obtained using the simulation approach showed that the algorithm reproduces the spatial patterns of the original data, taking into account their spatial variability (computed with the variogram model) and the variability of the population size. One of the advantages of geostatistical algorithms is that they can be used to create maps with the same covariate resolution and thereby assess the causative relationships among cancer risk factors. As all the maps were of the same resolution, local correlation coefficients could be calculated as well. The results revealed areas in

which the risk of cancer correlated well with the factors related to airborne particulate matter (RL10 and dry land) for the age groups 50–59, 60–69, and 70–79 years.

References

Durão RM, Pereira MJ, Costa AC, Delgado J, del Barrio G, Soares A (2010) Spatial-temporal dynamics of precipitation extremes in southern Portugal: a geostatistical assessment study. Int J Climatol 30:1526–1537

Gomez-Hernandez J, Soares A, Froidevaux R (1999) GeoENVII—Geostatistics for environmental applications. Proceedings of the second European conference on geostatistics for environmental applications, Valencia, Spain, November 18–20, 1998. Kluwer Academic, Dordrecht, 562p

Goovaerts P (2004) Simulation-based assessment of a geostatistical approach for estimation and mapping of the risk of cancer. In: Leuangthong O, Deutsch CV (eds) Geostatistics Banff. Kluwer Academic, Dordrecht, pp 787–796

Goovaerts P (2005) Geostatistical analysis of disease data: estimation of cancer mortality risk from empirical frequencies using Poisson kriging. Int J Health Geogr 4:31

Goovaerts P (2006a) Geostatistical analysis of disease data: visualization and propagation of spatial uncertainty in cancer mortality risk using Poisson kriging and p-field simulation. Int J Health Geogr 5:7

Goovaerts P (2006b) Geostatistical analysis of disease data: accounting for spatial support and population density in the isopleth mapping of cancer mortality risk using area-to-point Poisson kriging. Int J Health Geogr 5:52

Goovaerts P (2008) Kriging and semivariogram deconvolution in the presence of irregular geographical units. Math Geosci 40:101–128

Goovaerts P (2009) Medical geography: a promising field of application for geostatistics. Math Geosci 41:243–264

Goovaerts P (2010) Combining areal and point data in geostatistical interpolation: applications to soil science and medical geography. Math Geosci 42:535–554

Journel AG, Huigbreghts CJ (1978) Mining geostatistics. Academic Press, London

Kyriakidis P (2004) A geostatistical framework for area-to-point spatial interpolation. Geogr Anal 36(3):259–289

Liu Y, Journel A (2009) A package for geostatistical integration of coarse and fine scale data. Comput Geosci 35:527–547

Machado A, Nicolau R, Dias CM (2009) Consumo de tabaco na população Portuguesa: análise dos dados do. Inquérito Nacional de Saúde 2005/2006, Instituto Nacional de Saúde Doutor Ricardo Jorge, Departamento de Epidemiologia

Monestiez P, Dubroca L, Bonnin E, Durbec J, Guinet C (2004) Comparison of model based geostatistical methods in ecology: application to fin whale spatial distribution in northwestern Mediterranean sea. In: Geostatistics Banff. Kluwer Academic, Dordrecht, pp 777–786

Monestiez P, Dubroca L, Bonnin E, Durbec J, Guinet C (2006) Geostatistical modeling of spatial distribution of Balaenoptera physalus in the Northwestern Mediterranean Sea from sparse count data and heterogenous observation efforts. Ecol Model 193:615–628

Oliver MA, Webster R, Lajaunie C, Muir KR, Parkes SE, Cameron AH, Stevens MCG, Mann JR (1998) Binomial cokriging for estimating and mapping the risk of childhood cancer. IMA J Math Appl Med Biol 15:279–297

Pyrcz MJ, Deutsch CV (2001) Two artifacts of probability field simulation. Math Geol 33(7):775–799

Soares A (2001) Direct sequential simulation and cosimulation. Math Geol 33(8):911–926

Soares A, Gomez-Hernandez J, Froidevaux R (1997) GeoENVI—Geostatistics for environmental applications. Proceedings of the first European conference on geostatistics for environmental applications, Lisbon, Portugal, 18–19 November 1996. Kluwer Academic, Dordrecht