Spatial distribution of Biomphalaria mollusks at São Francisco River Basin, Minas Gerais, Brazil, using geostatistical procedures

Ricardo J.P.S. Guimarães a,b, Corina C. Freitas c, Luciano V. Dutra c, Carlos A. Felgueiras c, Ana C.M. Moura d, Ronaldo S. Amaral e, Sandra C. Drummond f, Ronaldo G.C. Scholte a,b, Guilherme Oliveira a,b, Omar S. Carvalho a,∗

a Centro de Pesquisas René Rachou/FIOCRUZ-MG, Av. Augusto de Lima, 1715, Barro Preto, CEP 30190-002, Belo Horizonte-MG, Brazil
b Programa de Pós-Graduação da Santa Casa de Misericórdia de Belo Horizonte/MG, Av. Francisco Sales, 1111, Santa Efigênia, CEP 30150-221, Belo Horizonte-MG, Brazil
c Instituto Nacional de Pesquisas Espaciais/INPE, Av. dos Astronautas, 1758, Jd. Granja, CEP 12227-010, São José dos Campos-SP, Brazil
d Universidade Federal de Minas Gerais/UFGM, Av. Antônio Carlos, 6627, Pampulha, CEP 31270-901, Belo Horizonte-MG, Brazil
e Secretaria de Vigilância em Saúde/MS, Setor Hoteleiro Sul, Q-06, Conjunto A, Bl.C, sala 711, CEP 70 322-915, Brasília-DF, Brazil
f Instituto Nacional de Pesquisas Espaciais/INPE, Av. dos Astronautas, 1758, Jd. Granja, CEP 12227-010, São José dos Campos-SP, Brazil

Abstract

Geostatistics is used in this work to make inferences about the presence of the species of Biomphalaria (B. glabrata, B. tenagophila and/or B. straminea), intermediate hosts of Schistosoma mansoni, at the São Francisco River Basin in Minas Gerais, Brazil. One of these geostatistical procedures, known as indicator kriging, allows the classification of categorical data, in areas where the data are not available, using a punctual sample set. The result is a map of species and risk area definition. More than a single map of the categorical attribute, the procedure also permits the association of uncertainties of the stochastic model, which can be used to qualify the inferences. In order to validate the estimated data of the risk map, a fieldwork in five municipalities was carried out. The obtained results showed that indicator kriging is a rather robust tool since it presented a very good agreement with the field findings. The obtained risk map can be thought as an auxiliary tool to formulate proper public health strategies, and to guide other fieldwork, considering the places with higher occurrence probability of the most important snail species. Also, the risk map will enable better resource distribution and adequate policies for the mollusk control. This methodology will be applied to other river basins to generate a predictive map for Biomphalaria species distribution for the entire state of Minas Gerais.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Schistosomiasis mansoni is an endemic disease encountered in approximately 54 American and African countries (WHO, 1985; Chitsulo et al., 2000). It is caused by Schistosoma mansoni and the intermediate hosts are mollusks of Biomphalaria genus (Mollusca: Pulmonata, Planorbidae).

In Brazil, there are ten species and one subspecies of Biomphalaria genus: B. glabrata, B. tenagophila, B. straminea, B. peregrina, B. schrammi, B. kuhniana, B. intermedia, B. amazonica, B. oligoza, B. occidentalis and B. tenagophila guaibensis. Come in these, only B. glabrata, B. tenagophila and B. straminea have been been found naturally infected by S. mansoni. Other two species, B. amazonica and B. peregrinates, were infected experimentally with this parasite, being considered hosts in potential (Correa and Paraense, 1971; Paraense and Correa, 1973).

Among the three intermediate host species of S. mansoni present in Brazil (B. glabrata, B. tenagophila and B. straminea), B. glabrata is of great importance, due to its extensive geographic distribution, high infection indices and effectiveness in the schistosomiasis transmission. Moreover, its distribution is almost always associated with disease occurrence. B. tenagophila was found naturally infected by S. mansoni in state of Minas Gerais, and it is responsible for the focus maintenance in the city of Itajubá (Katz and Carvalho, 1983). B. straminea, although had not been found infected in state of Minas Gerais, was considered responsible for Paracatu’s focus (Carvalho et al., 1988).

The mollusks of the Biomphalaria genus live in a wide of habitats, particularly in shallow and slow running waters, as lakes, lagoons, wells, cisterns, swamps, brooks, irrigation ditches and of drainage, where the substratum can be the muddy or rocky bed and with
floating or rooted vegetation. As these mollusks they distribute for extensive geographic areas, and present suitable populations the different environmental conditions could tolerate great variations in the physical, chemical and biological characteristics of the environment where they live.

The substitution of one species of Biomphalaria with another one, in the nature, already is an observed fact. The substitution of B. glabrata with B. tenagophila was observed in Belo Horizonte, Minas Gerais, Brazil, and B. glabrata with B. straminea, as described by Paraense (1970), Michelson and Dubois (1979) and Barbosa (1973). Paraense (1970) relates that a colony of B. glabrata was introduced accidentally in 1917, in Manguinhos, in the city of Rio de Janeiro, Brazil, in a natural breeding site of B. tenagophila, excluding the autochthon species of that place. This author raises the hypothesis by spawning transportation for birds (Jarne and Delay, 1991), gives the dispersion, be by floods, plants and fishes commercialization or above all, of endemic zones (Barreto, 1967; Barbosa, 1986). This of the transmission existence, is related to populations migration, high prevalence of schistosomiasis. Moreover, the presence of para-
other risk factors, contributes to the presence of communities with endemic areas, the high concentration of the hosts, associated with
São Francisco River Basin with area approximate of 300,760 km²
2.1. Data set acquisition

The research area is situated in the state of Minas Gerais, in the São Francisco River Basin with area approximate of 300,780 km² in 239 municipalities. Fig. 1a shows the state of Minas Gerais, and its municipalities, and the spatial distribution of the mollusks of Biomphalaria genus at the São Francisco River Basin. About 50% of the municipalities inside the São Francisco River Basin were investigated and one out of eight classes of the categorical attribute was assigned to each of them. The classes were defined as: B. glabrata, B. tenagophila, B. straminea, B. glabrata + B. tenagophila, B. glabrata + B. straminea, B. tenagophila + B. straminea, B. glabrata + B. tenagophila + B. straminea and, without Biomphalaria.

Data on the distribution of Biomphalaria mollusks were provided by the Laboratory of Helminthiasis and Medical Malacology of the Rene Rachou Research Center (C pqRR/Fiocruz-MG). Mollusks were collected in breeding places from different municipalities in Minas Gerais at different periods, using scoops and tweezers, and then packed to be transported to the laboratory (Souza and Lima, 1990). Specific identification was performed according to the morphology of the shells, reproductive system and renal ridge of the mollusks (Deslandes, 1951; Paraense and Deslandes, 1955a,b, 1959; Paraense, 1975, 1981), or recently through low stringency polymerase chain reaction and restriction fragment length polymorphism (Vidigal et al., 2000). In the last 40 years, great part of identification was undertaken by Dr. Wladimir Lobato Paraense, Department of Malacology of the Fundação Oswaldo Cruz (Fiocruz-RJ). In the last three decades technicians and researchers from state of Minas Gerais, mainly from CpqRR/Fiocruz-MG, Fundação Nacional de Saúde (FNS), and Universidade Federal de Minas Gerais (UFMG), have done their utmost to collect and classify the planorbid mollusks from this state (Souza et al., 1981; Souza, 1986; Carvalho et al., 1989, 1994, 1997, 1998; Souza and Lima, 1990; Souza et al., 1998; FNS, 1995, 1998).

2.2. Indicator kriging

A spatial attribute inside a region A of the earth surface can be modeled, from a geostatistical point of view, as a random function (Isaaks and Srivastava, 1989; Deutsch and Journel, 1998). For each position \( u \in A \), the attribute value of a spatial data can be considered as a random variable (r.v.) \( Z(u) \), which can assume different values with an associated probability of occurrence. For the n sample positions \( u_\alpha \), \( \alpha = 1, 2, \ldots, n \), it is considered that the r.v. has the probability 1 of occurrence. The distribution function, for categorical data, of the r.v. \( Z(u) \), conditioned to \( n \) sample points can be defined by (Deutsch and Journel, 1998):

\[
F(u; z(n)) = \text{Prob}[Z(u) = z(n)]
\]

The \( F(u; z(n)) \) models the uncertainty about \( Z(u) \) values, on the unsampled positions \( u \) considering the \( n \) sample points. The univariate distribution function of a categorical r.v. might be approximated using indicator kriging. The process begins in the transformation of the r.v. \( Z(u) \) in an indicator r.v. \( I(u; z_k) \) defined by:

\[
I(u; z_k) = \begin{cases} 
1 & \text{for } Z(u) = z_k \\
0 & \text{otherwise} 
\end{cases}
\]

where \( z_k \) is a cutoff value (classes of interest) belonging to the attribute domain.

The expectation value of the indicator r.v. \( E[I(u; z_k)](n) \), yields an estimation \( F^*(u; z_k(n)) \) for the distribution function of \( Z(u) \), at the cutoff value \( z_k \) and conditioned to the \( n \) sample data:

\[
E[I(u; z_k)(n)] = 1 \cdot \text{Prob}(I(u; z_k) = 1(n)) + 0 \cdot \text{Prob}(I(u; z_k) = 0(n)) = 0(n) + 1 \cdot \text{Prob}(I(u; z_k) = 1(n)) = F^*(u; z_k(n))
\]

This estimation, made through ordinary kriging over the indicator values, yields a least-square estimate for the distribution...
function at the cutoff value \( z_k \) (Deutsch and Journel, 1998). A set of \( F^*(u; z_k(n)) \) estimates, obtained for different \( k \) cutoff values, can lead to an approximation of the full distribution function of \( Z(u) \).

The optimal estimation \( z^*(u) \), of a categorical attribute represented by \( K \) classes \( c_k, k = 1, \ldots, K \), can be defined as:

\[
z^*(u) = c_k \text{ iff } p_k(u) > p_i(u) \quad \forall i, k = 1, \ldots, K, \quad i \neq k
\]

where \( p_k(u) = F^*(u; z_k(n)) \), when \( z_k = c_k \), is the estimated probability of the class \( k \) at the location \( u \). This estimator is known as the mode estimator since it considers the highest probability, and the classifier that assigns classes at each position \( u \) using the mode estimator is known as mode classifier.

The species classification probability for each position, inside the region of interest, is obtained with the values for indication (0 or 1) of the neighboring known samples (nearer) of \( u \). This probability is estimated by using ordinary kriging, where the weighted value of each neighboring sample of \( u \) is calculated considering the variability model defined by the semivariograms.

Usually, inferences about local uncertainties for categorical attributes, \( \text{Unc}(u) \), are made using the complement of the mode
The probability, which is defined as:

$$\text{Unc}(u) = 1 - \frac{1}{K} \sum_{k=1}^{K} p_k(u)$$

(5)

The mode estimator, as defined in Eq. (4) has the advantage of classifying all locations $u$ inside the region of interest. An alternate methodology consists on considering the uncertainty as a restriction to the classification process. In this case, only locations with an uncertainty below a pre-defined threshold ($U_{\text{max}}$) are classified, that is:

$$z^*(u) = \begin{cases} c_k & \text{iff } (p_k(u) > p_i(u) \land \text{Unc}(u) < U_{\text{max}}) \forall i, k = 1, \ldots, K \land i \neq k \\ \phi & \text{otherwise} \end{cases}$$

(6)

where $z^*(u) = \phi$ means that the value was not estimated, or the location was not classified.

2.3. Model fitting

Most of the data have no information about their geolocation, which is a prerequisite for applying geostatistical procedures. To overcome this limitation, the mollusk attributes (class of species and localization) were distributed along the drainage network for the investigated municipalities, by automatically converting the polyline river representation to a point sequence representation.$^1$

The mollusk presence class (which is known for each municipality) is then assigned to each of these points which belong to same municipality. The employed operation results in an average distance between points of 3 km, in sinuous river sections, to an about 12 km, in non-sinuous sections (see Fig. 1b).

Variogram models were fitted for each class, through exploratory analysis, using the geostatistical procedures as implemented in the SPRING (Georeferenced Information Processing System) software (Camara et al., 1996). These procedures involved the creation of experimental variograms and fitting them to mathematical theoretical models. Fig. 2 shows some of the experimental variograms and theoretical models, for the classes without Biomphalaria, B. straminea, B. glabrata + B. straminea and, B. glabrata + B. tenagophila + B. straminea. The exponential model was the best model for all the classes, except for the class containing all the species, whose the chosen model was the spherical. The fittings were not automatic, but iterative, since the user, after a first fit, check its suitability to the theoretical model (Camargo, 1997). After model fittings, indicator kriging procedures were applied to obtain an approximation of the conditional distribution function of the random variables. Based on the estimated function, maps of mollusk spatial distributions along with the corresponding uncertainties for the entire basin were built.

3. Results and discussion

The indicator kriging procedure, based on the fitted semivariograms, were applied using the sample data, presented in Fig. 1b, to generate a regular grid of 250 m of resolution ($x, y$) over the São Francisco River Basin. The resulting map of the species distribution generated by applying the mode estimator (Eq. (4)) is presented in Fig. 1c.

Fig. 1d presents a map of the uncertainties associated with the classification, computed by using Eq. (5). The map of uncertainties shows that the higher uncertainties are concentrated along class transition areas. As a consequence, regions where several classes may occur, more transitions are found and, therefore, higher uncertainties. This can be observed in the South
of study area where there exists a higher classes' variability.

Fig. 1c, e, f illustrates the use of the classifier that considers restrictions for the uncertainty level. In Fig. 1c the classes associates to the species of Biomphalaria, determined for a maximum level of uncertainty of 0.75; in Fig. 1e for a maximum level of uncertainty of 0.50 and in Fig. 1f for a maximum level of uncertainty of 0.25. It is observed, in these maps, that as the threshold of uncertainty admitted in the classification is decreased, the classified regions diminish. In this manner, it is possible to establish several scenarios based on distinct levels of uncertainties, depending on the decision to be taking.

To validate the methodology, a sampling procedure was conducted in five municipalities where no information existed about the presence of the mollusks, located at the areas which were estimated as having the mollusks B. glabrata and/or B. straminea on the north of Minas Gerais State. The research of mollusks was accomplished dividing the municipality in quadrants (north, south, east and west). A team was sent to one of the quadrants in different places of sampling, with the goal to identify the Biomphalaria, once found, it was sent to the analysis of the species in the CqRR and initiated the research in another municipality. Table 1 shows the five municipalities chosen for this purpose: Buritizeiro, Miravânia, São João das Missões, Matias Cardoso, and Bonito de Minas. These municipalities are shown in Fig. 1a. Also, Table 1 shows the estimated and found species, as well as the value of the uncertainty.

Table 1 shows the estimated map presented in Fig. 1c. Table 1 shows that, where uncertainty is higher (between 25% and 36%) at least one of the predicted species have been detected and where the uncertainty is very low (less than 1% for the Bonito de Minas municipality), the very two predicted species, B. straminea and B. glabrata have been found.

4. Conclusions and future work

The methodology presented in this paper, and validated in fieldwork, allowed to determine and to delimit, respectively, the distribution of the species of Biomphalaria and the areas of risk (map of uncertainty of the species of Biomphalaria).

Kriging is an auxiliary tool useful to guide the fieldwork, indicating the places with higher probability of occurrence of the considered species, with particular attention to those species that are more important for disease transmission, as B. glabrata in the case of Minas Gerais. The results of this tool can be used to better allocate the always limited resources for distribution studies and development of strategies for mollusk control.

Some important issues, related to the nature and precision of the Biomphalaria species data, need to be considered when looking at the results: the Biomphalaria species data were obtained from historical records (most occurring before the broad usage of GPS equipment), and the information is given in a municipality level basis. Because of this, an assumption was made that the species found in the municipalities are uniformly distributed inside the municipality drainage network. The authors believe, however, that other type of distribution hypothesis would not greatly affect the results.

Indicator kriging showed to be a rather robust tool since its results presented a very good agreement with the field findings. This methodology will be applied to the other basins to generate a predictive map for Biomphalaria species distribution for the entire state of Minas Gerais.

Conditioned to appropriate funds existence, an extensive malacologic survey is recommended for better evaluation of the methodology and also GPS utilization in all future fieldworks.

Acknowledgment

The authors would like to acknowledge the support of CNPq (grants # 305546/2003-1; 380203/2004-9; 304274/2005-4), Fapemig (EDP 1775/03; EDT 61775/03; CRA 0070/04) and NIH-Fogarty (grant # 5D43TW007012).

References


