systems.

2013 | volume 1
issue 3 | 83–99

ISSN 2305-6991
BCSSS

A two years simulation using a real data cellular automaton: *A predictive case study with the schistosomiasis expansion process along the coastline of Brazil*

Jones Albuquerque
Departamento de Estatística e Informática – Universidade Federal Rural de Pernambuco. Recife, PE, Brasil, joa@deinfo.ufrpe.br, +55.81.3320.6491

Silvana Bocanegra
Departamento de Estatística e Informática – Universidade Federal Rural de Pernambuco. Recife, PE, Brasil, silvana@deinfo.ufrpe.br, +55.81.3320.6491

Jordi Ferrer-Savall
Department de Física i Enginyeria Nuclear, Escola Superior d´Agricultura de Barcelona, Universitat Politècnica de Catalunya, Castelldefels, Spain, jordi.ferrer-savall@upc.edu, +34.93.552.11.28

Daniel López-Codina
Department de Física i Enginyeria Nuclear, Escola Superior d´Agricultura de Barcelona, Universitat Politècnica de Catalunya, Castelldefels, Spain, daniel.lopez-codina@upc.edu, +34.93.552.11.28

Marco Antônio de Souza
Departamento de Ciências da Saúde, CEUNES - Universidade Federal do Espírito Santo, São Mateus, ES, Brasil, marcosouza@ceunes.ufes.br, +55.27.3312.1544

Reinaldo Souza-Santos
Departamento de Endemias Samuel Pessoa, Escola Nacional de Saúde Pública, Fundação Oswaldo Cruz - FIOCRUZ. Rio de Janeiro, RJ, Brasil, rssantos@ensp.fiocruz.br, +55.21.2598.2525

Constança Barbosa
Centro de Pesquisas Aggeu Magalhães, Fundação Oswaldo Cruz - FIOCRUZ, Cidade Universitária, Recife, PE, Brasil, cbarbosa@cpqam.fiocruz.br, +55.81.2101.2572

Abstract: This work presents a Cellular Automata model to characterize the social and environmental factors, which contribute for the analysis of the expansion process of Schistosoma mansoni infection in Pernambuco - Brazil. The model has been experimented with a set of two years real data from a study area at North Coast of Pernambuco – Brazil. The main constraint equations, the modelling process and the results obtained until now with the simulating scenarios generated are presented here. The results identify, as in field works, endemic areas and human risk infection areas. Furthermore, predictive scenarios for a look ahead with a perspective into fifteen years are also presented.

Keywords: Cellular Automata; Schistosomiasis; Epidemiology; Computational Modeling

Acknowledgement: CNPq, CAPES, FACEP

1 Motivation

Schistosomiasis is a neglected primarily tropical parasitic disease caused by the larvae of flat worms of the genus *Schistosoma*. The helminth invades the human host through the skin, it reproduces in the digestive/excretory system and its larvae are released in human faeces. Then, they propagate via a freshwater snail found in tropical countries in Africa, the Caribbean, South America and Southeast Asia (Figure 1). In Pernambuco (Brasil), researchers from CPqAM/FIOCRUZ (Centro de Pesquisas Aggeu Magalhães/Fundação Oswaldo Cruz) have studied the correlation between rampant urbanization and spreading of schistosomiasis (Author et al, 2010), as well as its increasing adaptation to salt water along the coastline (Author et al., 2001).

The CPqAM/FIOCRUZ is currently implementing a project in Brasil entitled "Eco-epidemiology of Esquistossomose in the Coastline of Pernambuco" which focuses on the acute cases of schistosomiasis in coastal areas frequented by tourists and local holiday makers. Projects like this, aim to map out and characterize the sources and foci of schistosomiasis vectors, and to identify new sites of active transmission, correlating the biological determinants of the illness with the environmental context of their occurrence.

Despite the efforts of the National Health Foundation (FNS/PE), situations of chronic prevalence persist in rural regions, where up to 80% of humans are infected with the parasite, and cases of acute infection continually outbreak along the coastline of Pernambuco state, where the disease has been recently introduced. There is a lack of epidemiological data referencing the city of Recife and neighbouring areas, which would permit the identification and monitoring of these urban foci (Author et al., 2010; Author et al., 2001).

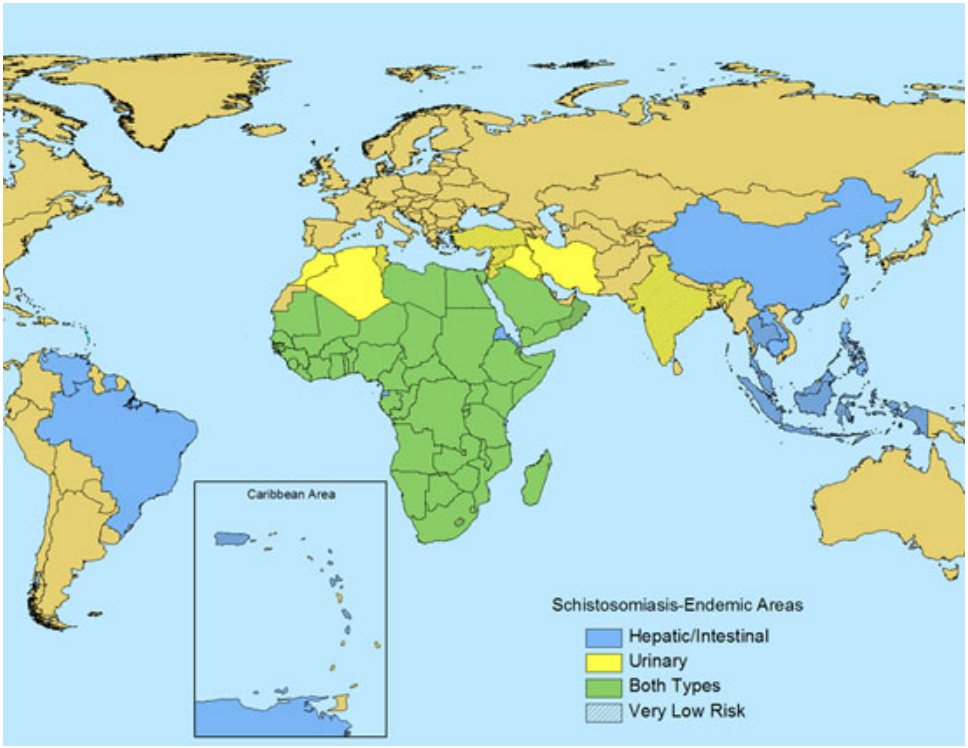

Figure 1: Schistosomiasis in the world and the case study location at Pernambuco-Brazil (CDC, 2011)

In this context, there is a need for epidemiological models for data gathering, planning of strategic control actions and for the prevention of the disease (Liu et al., 2010). This comprehensive approach should, not only allow the optimal use of public resources in data

gathering and monitoring of disease outbreaks, but it should also provide additional insight to the scientific community studying the epidemiology of the parasite and become a tool for practical purposes.

2 Material and Methods

A first step for developing a real-data model is to define its structure and to set the parameters that influence disease spread. This can be partially carried out through literature research (Habtemariam et al., 1990; Mangal et al., 2008) but, it also requires a specific study of the area to collect data in loco because, as it occurs with many Neglected Tropical Disease (NTD) the information available at web sites is not reliable or out of date (Author et al, 2010).

Data collection comprised years 2006-2007. Along with data gathering, the factors to be incorporated into the model are singled out in a two-steps procedure. The first is to examine all data and to identify incomplete or redundant information and how to infer it from other data sets. The second is to select the factors that must be incorporated into the model through performing virtual experiments in parallel to field trials in a sample location.

2.1 Area of study

The area under study is the village of Carne de Vaca, District of Pontas de Pedra, in the north coast of Pernambuco, Brazil. It is located in the municipality of Goiana and limited to the north by the Paraiba state, to the south by municipalities Itaquitinga, Igarassu, Itapissuma and Itamaracá, to the west by Condado and Itambé and to the east by the Atlantic Ocean (EMA, 2005). The main economic activities arise from subsistence farming and fishing of seafood and crab for local consumption and trade.

The village comprises around 1600 people in 1041 households distributed in 70 blocks and covering approximately 4 km^2. Most houses are located in the central part of Carne de Vaca, they have piped water distributed on alternate days, they require using tanks and wells for fresh-water storage and meagre sewer systems. The central area is cut by intermittent stream flows locally known as *maceió* (permanent lagoons formed by the sea tides and by rainfall). These streams spread over the southern sector of the village, they are piped only in a few central spots and they collect local drains without treatment.

Parasitological data were collected in a two-phase process (Author et al., 2010). First, an exhaustive parasitological study was carried out through a monthly stool analysis on the population, following the Kato-Katz method and examining two samples per patient between November 2006 and February 2007. Tests were performed on 1100 individuals, on average. In a second stage, on March 2007, the head of each household was formally interviewed using a semi-structured questionnaire on domestic habits concerning human water contact. This sampling comprised 202 households 44 blocks (62%). Since every block houses people under similar sanitary and socio-economic conditions, risk factors in daily routines were canvassed. Summer resorts, offices and dwellings with occasionally unavailable main holders (comprising 20 of the 70 blocks, on average) were not sampled.

Data regarding mollusc population ecology was gathered through monthly capture trials that tackled nine presumed hot spots along the *maceiós*, the coastline and the sewage system (Figure 2). These samples where used to estimate the mollusc population at each site, its health/fitness and the percentage of mollusc population infected with helminths. Meteorological data consisted of monthly pluviometry at a municipal scale.

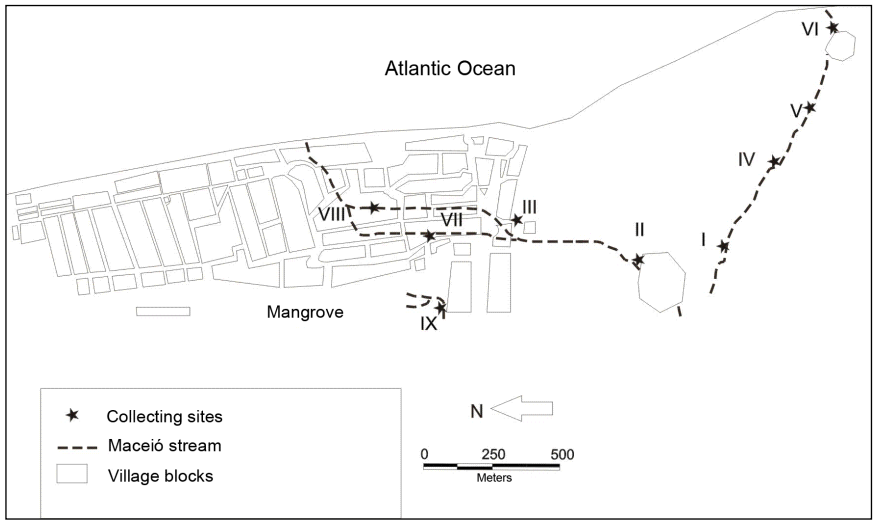

Figure 2: Sites for mollusc gathering along Carne de Vaca, District of Ponta de Pedra, Goiana, PE, 2007 (Author et al., 2010)

2.2 Summary of collected data

Parasitology

Carne de Vaca is a fairly endemic region for schistosomiasis, with a human prevalence of 17.3 cases per 100 inhabitants per year (Table 1).

	Male			Female			Total		
Age group	**Pop[1]**	**Posit[2]**	**Prev[3]**	**Pop**	**Posit**	**Prev**	**Pop**	**Posit**	**Prev**
up to 9	99	7	7.1	100	3	3.0	199	10	5.0
10 to 19	109	26	23.9	99	24	24.2	208	50	24.0
20 to 29	76	31	40.8	90	21	23.3	166	52	31.3
30 to 39	88	18	20.5	103	23	22.3	191	41	21.5
>= 40*	141	14	9.9	168	18	10.7	310	32	10.3
unreported	16	3	18.8	10	2	20.0	26	5	19.2
Total	**529**	**99**	**18.71**	**570**	**91**	**15.96**	**1100**	**190**	**17.3**

* No information on sex for one individual. 1 population. 2 Number of positives. 3 Prevalence per 100 inhabitants.

Table 1: Average number of humans infected with schohistosomiasis classified according to age group and sex detected in Carne de Vaca during years 2006 and 2007 (Author et al., 2010). Prevalence: percentage of infected people (total indicates the mean value).

Parasitological exams on 1100 residents (48.1% male) showed a male/female ratio of infection of 1.08:1, and no statistically significant difference between sexes ($p > 0.05$). The most affected age group was 20-29 years (16.31%) among men, and 10-19 years (12.63%) among women, the rate of infection for the group 30 - 39 years was very close to that of the most affected age group. No statistically significant differences were found between sexes per age group. The parasite load of 1 - 99 eggs per gram of stool was the most frequent among individuals of both sexes and for all age groups. Few individuals had burdens above

500 eggs per gram of faeces and a twelve year-old child presented a top load of 1992 eggs per gram of faeces.

Human population and habits

The survey on household conditions and sanitation routines rendered the following picture: 63.37% of the dwellings house 4 or more residents, 32.67% of them receive daily piped water supply, 2.97% hold weekly supply and 48.52% has intermediate regularity, usually on alternate days. Most of the drains (93.06%) led to a community pit, while only 0.49% of bathroom sinks are linked to the maceiós and 6.43% run elsewhere. A percentage of 80.7% of the households adopted alternative or complementary strategies for water use, such as community tanks, indoor reservoirs or home-made community wells. Most houses (59.9%) had no access to a well.

Cloth washing exclusively at home was reported in 68.81% of households, 12.8% was carried out in common washing places with home-made faucets, 2.5% in the streams nearby and 1% in the community main water reservoir. Many households combined cloth-washing at home or at the main reservoir with cloth washing at the maceió Riacho Doce, a spot suspect of being a focus of the disease.

Drinking water was mainly obtained from the main reservoir during both the dry and rainy seasons (at an estimated amount of 62.37% and 57.43%, respectively). Bathing and cooking with piped water was carried out by 45.54% and 44.06% of population, respectively. The ratio of houses with/without piped water was 5.31:1. The later, gathered water daily from diverse sources. No significant correlation was found between the holding water supply and risk of schistosomiasis.

Population ecology of mollusks

The sampling methodology established a fixed number of test sweeps at each collecting site. The number of captured seashells was a measure of the local population, the percentage of these snails that were captured alive, a measure of their health, and the fraction of them that turned out to be infested with helminths in the subsequent lab analysis, a measure of parasite prevalence.

Monthly data for each site were analyzed to define the model of mollusk population dynamics, which is specifically defined for each site and which also strongly depends on meteorological input data. Table 2 outlines the data regarding the mollusk population at each site averaged over a year period.

Collecting Sites	Alive	Dead	Positive to *S. mansoni*	% de infection
I	0	0		
II	1707	129	4	0,23
III	297	198	0	0
IV	0	0		
V	0	0		
VI	0	0		
VII	2355	322	37	1,57
VIII	76	125	3	3,95
IX	0	0		
Total	**4435**	**774**	**44**	**0,99**

Table 2: Number of monthly captures at nine fixed sites averaged over a year period (Author et al., 2010).

2.3 Modelling approach

A Cellular Automaton (Neumann & Burks, 1966) is used here to represent the evolution of schistosomiasis in a spatially explicit setting and at discrete time steps, as cellular automata (*CA*) are usually employed in computational epidemiology (Boccara et al., 1993; Rousseau et al., 1997; Fu et al., 2002).

Formally, a CA is a collection of cells on a grid of specified shape, whose states evolve through a number of discrete time steps according to a set of rules based on the states of neighboring cells. The number of states a cell can adopt is typically an integer (*k*), and the state of the cell *i* at a time *t* can be expressed with the following notation: $\sigma_i^t \in \{1 : k\}$

The evolution of the state of each cell is defined by a set of rules usually expressed as function (*f*) of the present own state and of the states of its neighbouring cells:

$$\sigma_i^{t+1} = f\left(\sigma_{i-L}^t, \ldots, \sigma_i^t, \ldots, \sigma_{i+L}^t\right),$$

where *L* is an index representing the size of the neighbourhood. In general, the evolutionary rule *f* is common to all cells and applied simultaneously. The most usual definition of the set of neighbourhoods within a CA is the Moore Neighbourhood, which comprises the nearest cells in the grid. Figure 3 presents the Moore Neighbourhood of a 2D square grid, the one that is employed in the current model:

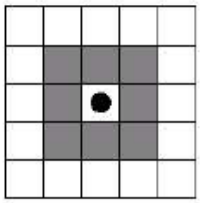

Figure 3: Moore Neighbourhood (in grey) of the cell marked with a dot in a 2D square grid (Neumann & Burks, 1966)

CA allow producing a great variety of unexpected behaviors with very simple rules and structures (Wolfram, 1984). In particular, this work presents a 2D square grid CA laid on a map representing the area under study, and its evolutionary function represents the spreading of schistosomiasis from site to site, following a set of rules that account for the epidemiological variables of each site, which are defined through the collected data. With it,

the evolution of the disease in various scenarios can be explored by running simulations with different initial conditions.

It must be stressed that the process of building, testing and using a model requires the continual close collaboration between modellers and field epidemiologists, in order to profit from heuristic knowledge and to avoid any unrealistic performance of the CA. Finally, the complexity of any CA can be increased at will, however complex models are more difficult to analyze and to interpret, and the computational time required per simulation in such models may become restrictive. Therefore, the maxim "*the simpler, the better*" fully applies when building a CA that aims to produce fruitful results.

3 Description of the model

The purpose of the model is to detect hot spots for the helminth propagation and to assist field epidemiologist in the design of cost-effective strategies to control its spreading.

The grid representing the area of study (Figure 4) is described by a 10 x 10 matrix. Each cell covers an area of around *4 ha* = 400 m² and it is occupied by 20 households (~100 people), at most. This high resolution for cells (20mX20m) is set to cope with the ecological data regarding the mollusc population. Therefore, the simulation space spreads over 4 km^2 and it has closed boundary conditions: no external fluxes of humans, molluscs or helminths are allowed.

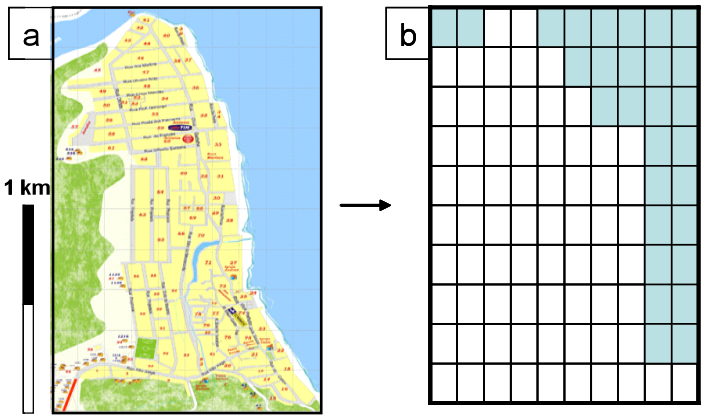

Figure 4: Scheme of the area of study: a) city map of Carne de Vaca, b) the corresponding simulation space, a 10x10 square grid.

The CA takes account of the data collected by the CPqAM/FIOCRUZ (*i.e.*: demographic, epidemiological and ecological). The model uses 15 quantities that describe the characteristics of each cell. Their values are extracted from the collected data and subsequently tuned to fit the behaviour observed on field during verification trials. Hence, they can be regarded as parameters of the model, except that they are not set to a fixed value, but defined through a statistical distribution among a range of values, which are listed in Table 3. In fact, these parameters should be preferably regarded as random variables that vary for every run of the simulator.

Parameter	Ranges (avg)	How were obtained?
Susceptible human population	0-23	social inquires (Author et al, 2010)
Infected human population	0-23	croposcological inquires (Author et al, 2010)
Recovered population of humans	0-23	social inquires (Author et al, 2010)
Rate of mobility of humans	0-26%	social inquires (Author et al, 2010)
Rate of mobility of molluscs	0-2%	malacological research (Author et al, 2010)
Population of healthy molluscs	0-1302	malacological research (Author et al, 2010)
Population of infected molluscs	0-11	malacological research (Author et al, 2010)
Area susceptible to flooding	0-45%	LAMEPE - Meteorological Laboratory of Pernambuco (lamepe, 2008) and environmental inquires (Author et al, 2010)
Connection to other cells	0-100%	LAMEPE - Meteorological Laboratory of Pernambuco (lamepe, 2008) and environmental inquires (Author et al, 2010)
Rate of human infection	0-100%	croposcological inquires and social inquires (Author et al, 2010)
Rate of human re-infection	0-100%	croposcological inquires and social inquires (Author et al, 2010)
Recovery rate	0-100%	croposcological inquires and social inquires (Author et al, 2010)
Mollusc infection rate	0-100%	malacological research (Author et al, 2010)
Rate of sanitation	0-93%	social and environmental inquires (Author et al, 2010)
Rainfall of the area	39-389mm	LAMEPE - Meteorological Laboratory of Pernambuco (Lamepe, 2008)

Table 3: Parameters and variables used in the CA, range of values adopted in the model and source for the range definition. Each parameter is defined at every site. When no units appear in the range columns, the variable is a whole number that represents the number of individuals that may occupy the spatial cell.

The model outcome presents 6 distinctly-coloured states per cell, which stand for the observed/predicted rate of human infection.

The temporal resolution of the model is one day, and the average duration of a simulation represents one year in the real world. The evolutionary function is set as an algorithm that accounts for human demography and routines, population dynamics of the molluscs, environmental conditions and spreading of the disease. The algorithm used to implement the temporal evolution establishes a set of arithmetic relations between the values contained in the 15-dimension matrix that characterizes the present state of the simulation space and updates the values for next time step in parallel. The algorithm, which considers a Moore neighbourhood and follows an explicit Euler method for integration, is outlined in Figure 5.

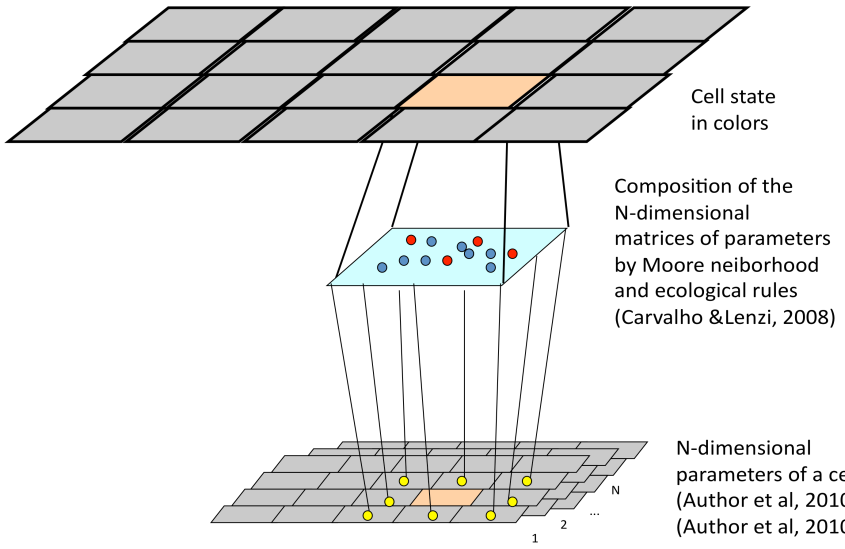

Figure 5: Scheme of the modelling process: at each time step, the future state of the system (top layer) is obtained from the operations carried out on a 15-D matrix that contains the characteristic parameters

and variables of each cell. These operations involve only the values of the variables of the nearest neighbouring cells. A single run of the simulator is a set of N iterations of this algorithm.

This process is iterated *N* times in order to generate a simulation run. Nevertheless, a complete simulation consists of *M* simulation runs, where *M* is a large number, so that a statistical analysis on the simulation landscape can be carried out. It should be kept in mind that this CA is essentially stochastic, in that the values of its parameters are set at random for each run, so its results can only be drawn through statistical inference. The distribution functions used to generate the values of the parameters under the range of observation are either uniform distribution functions or normal distributions. This choice is adopted o that the parameters are set mimicking the distribution of experimental data.

The set of rules used to define the evolution function are based on well-established population models of schistosomiasis epidemiology and risk analysis (Carvalho and Lenzi, 2008). They are explained with more detail in the following sections. A complete description of the simulation algorithm can be downloaded at Author Project Site. This site also offers a user guide on how to transform field data into matrix parameters.

3.1 Mollusk population dynamics

Mollusk population dynamics is modeled at each site separately using a growth model for the number of individuals (*N*) that considers the intrinsic growth rate (*r*) and the maximum sustainable yield or carrying capacity (*C*) defined at each site (Verhulst, 1838):

$$\frac{dN}{dt} = rN(1 - \frac{N}{C}) \qquad (1)$$

The analytical solution of the differential equation above is a logistic function. This means that the population increases exponentially when the population is much smaller than the carrying capacity and that it remains stable when the population matches the available resources. In the meantime, the number of individuals at any time *t* (*N(t)*) can be obtained in terms of the initial size of the population (N_0) and the local characteristics defined at each site *r* and *C:*

$$N(t) = \frac{C}{1 + \frac{C - N_0}{N_0} e^{-rt}} \qquad (2)$$

In practice, the model calculates the local increase of population using equation 1 and calculating *N(t+1)* out from *N(t)*. The values for *r* and *C* are set at each site and each time step, using monthly meteorological inputs and considering the ecological quality of the habitat (see Section 3.3).

There are two important factors that are not taken into account by this model. First, the model is continuous and deterministic, which entails several nonrealistic limitations. For instance, according to the model, preexisting populations in a site never become extinct while sites were samplings were negative have an unalterable null risk of snail appearance. And second, the model does not account for the potential migration of mollusk population from site to site along the *maceios*. Nevertheless, these considerations were not relevant in the present scenario, as the model correctly reproduced mollusk population dynamics.

3.2 Human infection dynamics

At each spatial cell, the evolution of the infection is modeled through a very simple compartmental model (SIR –SI) that couples human and mollusk populations. This model

splits the human population into three compartments: S (for susceptible), I (for infectious) and R (for recovered and not susceptible to infection) and the snail population into two compartments: M_S (for susceptible mollusk) and M_I (for infectious mollusk). It allows the computation of the infection dynamics at each site by using a set of coupled ordinary differential equations:

$$\frac{dS}{dt} = -p \cdot S \cdot M_I + \alpha R$$
$$\frac{dI}{dt} = p_H \cdot S \cdot M_I - \chi I \quad (3a)$$
$$\frac{dR}{dt} = \chi I - \alpha R$$

$$\frac{dM_S}{dt} = -p_M \cdot I \cdot M_S + rM_S \quad (3b)$$
$$\frac{dM_I}{dt} = p_M \cdot I \cdot M_S - rM_I$$

The key parameters in these equations are the transition rates between compartments. The transition between S and I depends on the contact between healthy humans and infected snails and which occurs at a rate p_H, The transition between I and R occurs at the rate of recovery of infected humans χ. Recovered humans may become susceptible to infection again at a rate α. For the mollusc population, the model is much simpler. Molluscs simpli get infected at a rate p_M in the presence of infectious humans.

This model has strong limitations as it describes the population at each site with continuous variables although the number of individuals is very small. In this sense, it has similar intrinsic problems to those of the model described in Section 3.1. It also does not consider a delay between being infected and becoming infectious. Withal, once its parameters have been calibrated to real phenomena, it renders valid results.

3.3 Socioeconomic and environmental factors

The diversity in human habits and routines is included in the CA as a local modulation for the transition rates of the infection model, and also as a set of bias to the model of human motion, which is implemented, in essence, a random-walk exploration of the simulation space and the subsequent return to the household.

Environmental and malacological data are include into the model by defining the quality index of each site, a number between 0 and 35 that increases with the impact of human actions in the quality of water (Callisto et al., 2002). This index (Figure 6) is used to set the values of r and C in the population growth model of mollusks.

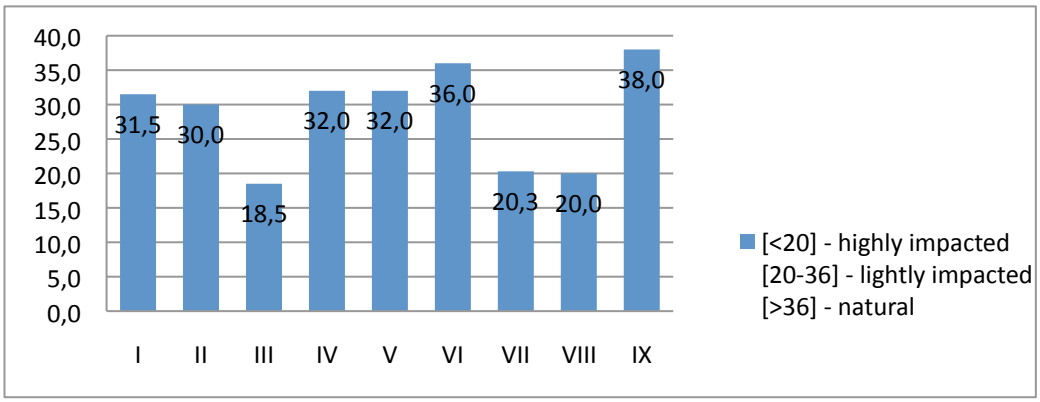

Figure 6: environmental quality of the nine collection sites in Carne de Vaca, according to the criteria of Callisto et al (Author et al., 2010).

Monthly rainfall in the region (Lamepe, 2008) is introduced as a global parameter governing mollusk population dynamics.

4 The Simulation Algorithm

A simulation of the CA comprises many simulation runs that take place simultaneously. Each simulation run consists in repeating iteratively a set of processes for each cell every time step. An overview of the processes that the algorithm executes every time step is listed below:

4.1 Main loop

The main loop implements a double-act iteration of the CA: during the first phase the human population spreads simulating daily routines. During the second one, the model calls the events occurring at that time step (Section 4.2) and it may also account for mollusk spreading, although this process is not actually called in most runs.

1. Choose a cell in the world;
2. For each human in the cell perform a random walk weighted by the “probability of movement" defined at each site.
3. Repeat these steps for every cell in the world. Then update data.
4. Choose a cell in the world;
5. Call the “Events” process;
6. Return the individual to his original cell after the infection phase;
7. Choose a cell in the world;
8. For the mollusk population in that cell, perform a diffusion process weighted by the “rate of movement" defined at each site;

   Repeat these steps for every cell in the world. Then update data.

4.2 Main events

The events occurring at every cell comprise the population growth of the mollusks and the infection process of humans and mollusks. They are carried out in the following order:

1. Increase the population of mollusks using the growth model described in Section 3.1;
2. Compute the transition between population compartments of humans using the set of equations (3b) defined in Section 3.2;
3. Compute the transition between population compartments of humans using the set of equations (3a) defined in Section 3.2;

   Update local data of the spatial cell.

Every time data is updated it may be stored and/or output to external files for further statistical analysis. The original source code of these simulations can be downloaded at Author Project Site.

5 Results and Discussions

5.1 Simulation performance

The simulations were performed using Mathematica 7.0 (Mathematica, 2011) with a processor Intel i5 3GHz, 4MB Cache, 8GB RAM. The scenarios generated were statistically

analyzed and compared with real situations. Field epidemiologists from CPqAM/FIOCRUZ evaluated the statistical data as well as the sets of simulation runs to discriminate realistic and unreal scenarios in order to refine and calibrate the model.

Figure 7 depicts the computational costs of a complete simulation when the CA sweeps values for an increasing number of parameters (while the values of the other parameters are maintained with a fixed value) for a fixed size of the simulated world (10x10) cells and duration of simulation runs (365 time steps).

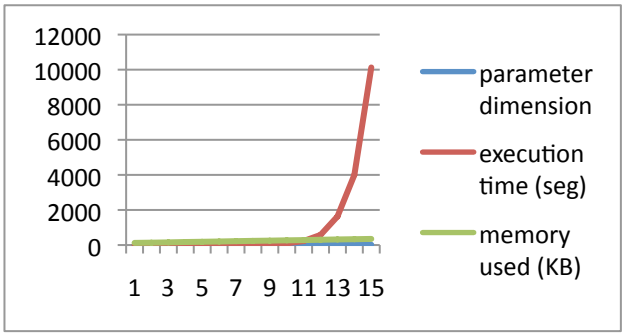

Figure 7: Computational costs of a complete simulation when assuming a fixed world size (10x10 cells) and extent (365 time steps) and an increasing number of parameters being swept for rejection sampling (from 1 to 15).

Figure 8 demonstrates the expected behavior of the simulation of any CA with increasing complexity (Wolfram, 1984): although the increase in memory used to store data and to program instructions for retrieval increases linearly, the execution time increases exponentially. These results show that 10x10-grid 1year-span simulations that include 15 parameters are feasible in a conventional computer or even a laptop (so they can be used to model spreading of schistosomiasis in a small village, as it has been done here). They also suggest that parallel computation would be required for larger world matrices or to deal with a greater number of parameters, because these choices would exponentially boost computation time. But that such strategy is not required when increasing the extent of simulations, because they cause a linear increase in execution time (Quinn, 1994).

5.2 Simulated risk scenarios

Simulation outcomes are risk scenarios that change every time step. Figure 8 presents several screenshots of the scenarios for human infection risks during a 1-year simulation. These scenarios can be compared to the number of residents of the corresponding blocks in Carne de Vaca, although strictly speaking, the correspondence is not one-to-one. It can be observed that:

- the risk of infection varies with time reproducing the patterns of infection spreading observed in reality
- by the end of a year, the model predicts that around 25% of the blocks (18 of 70) are at risk of being infected ($I \geq 20\%$) and that for at least one block, the infection is ensured.
- The blocks near the mollusk collection points III, VII and VIII, suggesting that they are the most active infection foci. These spots turned out to be the most environmentally affected areas, suggesting a correlation between contamination and risk.

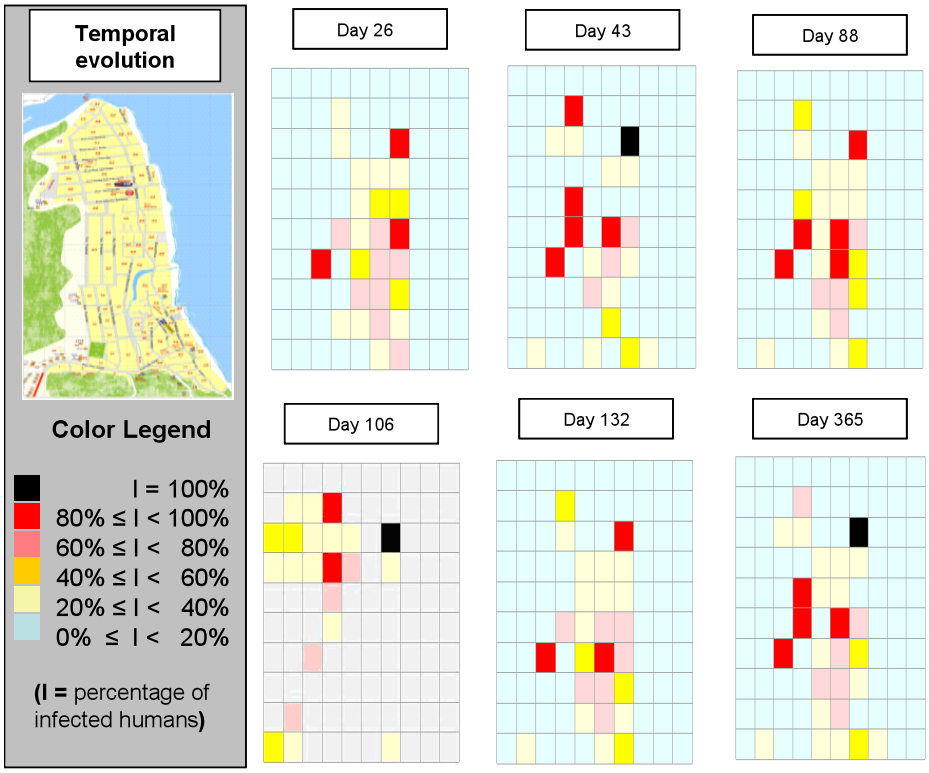

Figure 8: Screenshots of the human risk scenarios generated by a simulation with 15-dimension parameter matrix.

Noteworthy the simulation result are in complete agreement with the results presented in (Author et al., 2010), which point that:

> *"according to the risk indicator, in the scattering diagram of Moran represented in the Box Map (Figure 2), indicated 18 areas of highest risk for the schistosomiasis, all located in the central sector of the village. Areas with lower risk and areas of intermediate risk for occurrence of the disease were located in the north and central portions with some irregularity in the distribution." (Author et al., 2010, Cadernos de Saúde Pública (ENSP. Impresso), v. 26,1013-1023.)*

5.3 Predictive capacity of the CA

Once calibrated, the model can be used to generate long-term predictive scenarios assuming that the parameters are maintained within the simulation ranges, this meaning that no interventions are adopted in the real world, and that no changes in their routines are undertaken by the local community. Figure 9 shows the 15-year prediction scenarios for Carne de Vaca. The results show a rampant increase of the disease prevalence, which is an observed phenomenon in many of the villages of the Atlantic coastline of Brazil.

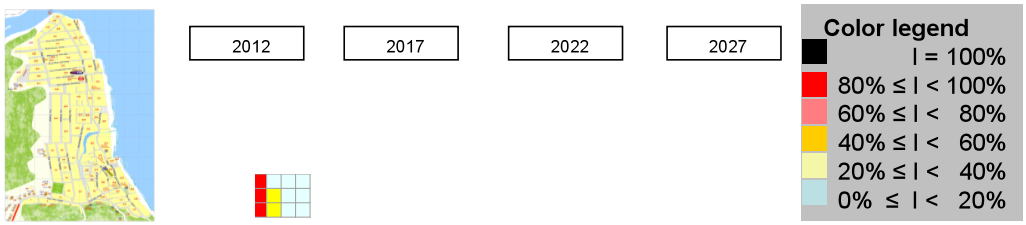

Figure 09: Predictive scenarios generated with the parameter calibration of the year 2007 that show endemic schistosomiasis. I stands for the average percentage of infected humans per spatial cell predicted by the model

These pessimistic scenarios point at the establishment of endemic schistosomiasis in Carne de Vaca if no actions are undertaken.

6 Conclusions

We have presented a Cellular Automata that models the expansion of *Schistosoma mansoni* in a mixed community of humans and freshwater mollusks. We have built and described a simulator that can be used as a tool to assist in the monitoring, diagnosis and preventive analysis of an affected or endemic area, and we have calibrated it to a specific location: Carne de Vaca, District of Ponta de Pedra, Goiana, PE, Brazil. We have made the simulator available at Author Project Site for trial-testing by field epidemiologists elsewhere.

In order to be applied to other specific locations, the model should be recalibrated: the range and distribution functions of the parameter swept at every simulation should be redefined and the simulator should be tuned and polished to prevent it from unrealistic scenarios. We believe that this process of calibration might take several weeks at most, once the epidemiological, malacological and meteorological data covering at least one year has been gathered.

Once the model has been calibrated using one year data, it has been proved to correctly reproduce the data of the following years -by the moment, only data regarding 2007 and 2008 has been used to check the model, epidemiological surveys to be performed by the CPqAM/FIOCRUZ during 2012 and 2013 are meant to be the acid test of the simulator-. Anyways, long-term predictions of the model are in agreement with the behavior observed in other regions of the Atlantic coastline of Brazil, which makes us optimistic on the model performance

The future validation process on the model will entail a comprehensive and detailed revision of its procedures and results. Further development of the model and of the simulator should address the modeling of specific interventions carried out to control the disease, and subsequently, the input of eventual modifications on the simulation parameters on-the-flow, in order to represent their potential effect in a given scenario. Finally, the working process undertaken during the development of the model should be formally described to the academic community and to field epidemiologists.

References

Author. (2010). Current epidemiological status of schistosomiasis in the state of Pernambuco, Brazil. Memórias do Instituto Oswaldo Cruz (Impresso), v. 105, 549-554.

Author (2001). Epidemia de esquistossomose aguda na praia de Porto de Galinhas, Pernambuco, Brasil. Cadernos de Saúde Pública - FIOCRUZ, 17(3), 725-728.

Boccara, N.; Cheong, K. & Oram, M. (1994). A probabilistic automata network epidemic model with births and deaths exhibiting cyclic behavior. Journal of Physics A: Mathematical and General. 27:1585-1597.

Callisto, M.; Ferreira, W.R.; Moreno, P.; Goulart, M. & Petrucio, M. (2002). Aplicação de um protocolo de avaliação rápida da diversidade de hábitats em atividades de ensino e pesquisa (MG-RJ). Acta Limnologica Brasiliensia 14(1):91-98.

Carvalho, O. & Lenzi, H. (Ed.) (2008). Schistosoma mansoni e Esquistossomose: Uma Visão Multidisciplinar. 1st ed. Rio de Janeiro: Editora FIOCRUZ.

CDC. (2011). Schistosomiasis Map. Retrieved December 17, 2011, from http://www.cdc.gov/parasites/schistosomiasis/

EMA - Ministério de Minas e Energia, (2005). Projeto Cadastro de Fontes de Abastecimento de Águas Subterrâneas – diagnóstico do Município de Goiana. Ministério de Minas e Energia: Secretaria de Planejamento e Desenvolvimento Energético, Secretaria de Geologia, Mineração e Transformação Mineral, Programa Luz Para Todos, Programa de Desenvolvimento Energético dos Estados e Municípios – PRODEEM, Serviço Geológico do Brasil – CPRM, Diretoria de Hidrologia e Gestão Territorial, Recife, 2 – 5 p.

Fu, S.C. (2002). Modeling Epidemic Spread using Cellular Automata. Master Thesis. Department of Computer Science and Software Engineering - The University of West-ern Australia, Australia, 70 pp.

Habtemariam R, 1990 Research in Computational Epidemiology. In: Mathematical and computer modelling [0895-7177] 14, 653 -658.

Lamepe. (2008). Laboratório de Meteorologia de Pernambuco (Lamepe). Retrieved December 17, 2011, from http://www.itep.br/index.php/laboratorios-lamepe/meteorologia-lamepe/links-lamepe

Liu, Y.; Chu, K.; Xu, X.; Wu, H. & Wan. C. (2010). A Cellular Automaton Model for the Transmission Dynamics of Schistosomiasis. 4th International Conference on Bioinformatics and Biomedical Engineering (iCBBE), 1-4, China: Chengdu.

Mangal, T.D., Paterson S. and Fenton, A (2008). Predicting the Impact of Long-Term Temperature Changes on the Epidemiology and Control of Schistosomiasis: A Mechanistic Model . PlosOne 3(1) e1438.

Malthus, T.R. (1826). An Essay on the Principle of Population. 6th ed. London: John Murray.

Mathematica. (2011). Wolfram Mathematica. Retrieved December 17, 2011, from http://www.wolfram.com/mathematica/.

Neumann, JV. & Burks, A.W. (1966). The Theory of Self-Reproducing Automata. University of Illinois Press, Urbana.

Author. (2010). Spatial patter, water use and risk level assciated with the transmission of schistosomiasis on the north coast of Pernambuco, Brazil. Cadernos de Saúde Pública (ENSP. Impresso), v. 26,1013-1023.

Quinn, M. J. (1994). Parallel computing: theory and practice. 2nd ed. New York, USA: McGraw-Hill, Inc.

Rousseau, G.; Giorgini, B. & Livi, R . (1997). Dynamical phases in a cellular automa-ton model for epidemic propagation. Physica D, 103:554-563.

Author (2010). Aspectos ecológicos e levantamento malacológico para identificação de áreas de risco para a transmissão da esquistossomose no litoral norte de Pernambuco, Brasil. Iheringia. Série Zoologia (Impresso), v. 100, 1-6.

Verhulst, P.F. (1838). Correspondence Mathmatique et Physique. Reprinted in Ecology (1965), E. J. Kormondy (Ed.). Englewood Cliffs, NJ: Prentice-Hall.

Wolfram, S. (1984). Computation theory of cellular automata. Communications in Mathematical Physics, 96:15-57.

About the Authors

Jones Albuquerque

received the Ph.D. degree in Computer Science from the Federal University of Minas Gerais, Brazil, in 2002. He is a Professor at the Department of Statistics and Informatics of the Federal Rural University of Pernambuco, Brazil, where he coordinates the Epischisto.org Group. He is currently a Post-Doctoral researcher with the Discrete Modelling and Simulation of Biological Systems (MOSIMBIO) Group at the Parc Mediterrani de la Tecnologia, from the Universitat Politécnica de Catalunya, Barcelona, Spain.

Silvana Bocanegra

received the Ph.D. degree in Computer Science from the Universidade Federal de Minas Gerais (2005). She is a Epichisto.org member and she is a Professor at the Department of Statistics and Informatics of the Federal Rural University of Pernambuco, Brazil.

Jordi Ferrer-Savall

is currently a Post-Doctoral researcher with the Discrete Modelling and Simulation of Biological Systems (MOSIMBIO) Group at the Parc Mediterrani de la Tecnologia, from the Universitat Politécnica de Catalunya, Barcelona, Spain.

Daniel López-Codina

coordinates the Discrete Modelling and Simulation of Biological Systems (MOSIMBIO) Group at the Parc Mediterrani de la Tecnologia, from the Universitat Politécnica de Catalunya, Barcelona, Spain.

Marco Antonio de Souza

received PhD in Parasitology from the Universidade Federal de Minas Gerais (2006). He served as a visiting researcher of Centro de Pesquisas Aggeu Magalhães - Fundação Oswaldo Cruz (FIOCRUZ) in Recife, Pernambuco(2006-2008). He is currently Associate Professor of Universidade Federal do Espirito Santo and he is a Epichisto.org member.

Reinaldo Souza-Santos

received PhD in Public Health from the Escola Nacional de Saúde Pública, Brasil(2000). He is a Titular Research at Fundação Oswaldo Cruz , Brasil and he is a Epichisto.org member.

Constança Barbosa

received Ph.D. degree in Public Health from the Escola Nacional de Saúde Pública, Brasil(1996). She is a Titular Research at Fundação Oswaldo Cruz , Brasil and she also coordinates the Epishisto.org Group.