Reliability Engineering and System Safety 118 (2013) 8–17

Contents lists available at SciVerse ScienceDirect

Reliability Engineering and System Safety

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

Global sensitivity/uncertainty analysis for agent-based models

Maria Fonoberova [a,*], Vladimir A. Fonoberov [a], Igor Mezić [b]

[a] Aimdyn, Inc., 1919 State St., Ste. 207, Santa Barbara, CA 93101, USA
[b] Center for Control, Dynamical Systems and Computation, University of California - Santa Barbara, Santa Barbara, CA 93106, USA

ARTICLE INFO

Article history:
Received 18 October 2010
Received in revised form
3 April 2013
Accepted 10 April 2013
Available online 23 April 2013

Keywords:
Global sensitivity analysis
Uncertainty analysis
Agent-based model
Variance-based global sensitivity
Derivative-based global sensitivity
Model reduction

ABSTRACT

Agent-based models simulate simultaneous actions and interactions of multiple agents, in an attempt to re-create and predict the appearance of complex phenomena. We propose to use global sensitivity analysis as a tool for analyzing and evaluating agent-based models. A general approach for applying the global sensitivity analysis to agent-based models is presented and tested on the example of a socio-cultural agent-based model we developed earlier [45]. We identify the most significant parameters in the model and uncover their contributions to the outputs of interest. Methodology of model reduction for agent-based models is discussed and demonstrated for the aforementioned model.

1. Introduction

An agent-based simulation is a computational technique in which behavior of individual agents or group of agents is encoded by simple rules, and the outcomes are observed at the scale of the system. Agent-based modeling is a widely used technique in different areas such as computer science [1–3], economics [4–7], biology [8], ecology [9], social phenomena [10–16]. The agent-based model is basically a big Markov chain, which however is too big for the standard analysis. The agent-based modelling is gaining popularity in the context of risk analysis and reliability [17].

Sensitivity analysis for agent-based models provides understanding of the influence of the different input parameters and their variations on the model outcomes. The objective of the sensitivity analysis is to identify the most significant parameters in the model and to quantify how the parameter uncertainty influences the outcomes. To perform the sensitivity analysis, the considered model is evaluated a specified number of times with different values of the input parameters. Based on the results, a reduced model with a smaller set of parameters can be produced. Sensitivity analysis is important for understanding relationship between input parameters and outputs, testing the robustness of the output, and identifying errors in the model. Sensitivity analysis strategies are well presented in [18]. The review on calibration, validation, and sensitivity analysis and survey of sampling-based methods for uncertainty and sensitivity analysis can be found in [19,20]. Some sensitivity analysis techniques for agent-based models have been discussed in [21–25].

In Ref. [21] an agent-based model of the spread of the communicable disease measles is considered. The authors demonstrated that the dynamic spatial interactions within the population lead to high numbers of exposed individuals who perform stationary activities in areas after they have finished commuting. The univariate technique, when the model outcome is analyzed with respect to the variation of one parameter at a time with the other parameters of the system being constant, was used for the sensitivity analysis. To analyze the impact of the parameters, the model outputs (daily numbers of susceptible, exposed, infected, recovered individuals) were visually compared. The results indicated that the model is sensitive to the rate of infection parameter, based on the population density, and the time spent in daily activities.

In Ref. [22] the agent-based model called Agricultural Policy Simulator, which shows the agricultural structural development on the regional level, is studied. The authors selected five parameters and one output (average economic land rent per hectare in the region) for analysis. The least squares method is used to fit the data to the linear response surface function. The graphical analysis (mean, scatter and block plots) is used to determine the most important parameter which is the interest rate level followed by the technological change and the managerial ability. The same parameters are identified by applying the linear regression model in which the output is regressed on factor level settings and two-factor interactions.

* Corresponding author. Tel.: +1 805 687 6999; fax: +1 805 687 7999.
E-mail address: mfonoberova@aimdyn.com (M. Fonoberova).

0951-8320/$ - see front matter
http://dx.doi.org/10.1016/j.ress.2013.04.004

In Ref. [23] an agent-based model describing granuloma formation during *M. tuberculosis* infection is presented. The authors pick 12 of the 27 parameters of the model and study their effect on the development and spread of infection. Latin hypercube sampling method with 1000 samples and uniform distribution for all parameters was used. The sensitivity analysis was done by using partial rank correlation coefficients on each parameter set. In such a way sensitivity of each considered parameter was calculated at every specified time moment. The analysis showed that the intracellular growth rate of the bacteria is strongly and positively correlated with the number of extracellular bacteria at early and late times of the infection and negatively correlated at intermediate times.

In Ref. [24] an agent-based model of Leishmania major infection is considered. It is based on the work from [23]. The authors select five parameters out of 25 parameters for the sensitivity analysis, which is performed by using Gaussian processes to approximate the computer code. The functional analysis of variance (ANOVA), where the total functional variance of the Gaussian process is decomposed into variance due to the main and interaction effects of the parameters, was used to find important parameters at every specified time moment. It was shown that the growth rate is the most important parameter for any time point.

In Ref. [25] an agent-based model of mosquitofish population dynamics is considered. The model includes 30 parameters driving the model expectancy and 11 parameters driving the model variability. In order to estimate local sensitivity coefficients by linear regressions, all parameters were varied simultaneously, at random, around their prior values. Also, ANOVA was carried out with complete factorial design for each output (two levels per factor). It was shown that the most important parameter is the effect of fish population biomass.

As seen from the above brief literature review, sensitivity analysis techniques for agent-based models consist mainly of visual analysis, one-parameter-at-a-time local techniques (e.g. partial derivative), and techniques assuming linear or monotonic relationships between model parameters and model outputs (e.g. least squares linear fit, linear regression, Gaussian process, partial rank correlation coefficient). None of the methods listed in this paragraph are capable of providing sensitivity indices for non-monotonic input-output dependencies typically observed in agent-based models. Moreover, local sensitivity analysis involves computation of the derivative of the model response with respect to the input parameters and does not take into account interactions between parameters. Other reasons against using local sensitivity techniques can be found in Ref. [26].

On the other hand, global sensitivity analysis methods evaluate the effect of a parameter while all other parameters are varied as well and thus account for interactions between parameters. Global sensitivity techniques can be applied to arbitrary nonlinear functions. In order to avoid the "curse of dimensionality" of the factorial analysis, global sensitivity indices are computed by sampling the space of uncertain parameters. Computational complexity of global variance sensitivity indices scales linearly with the number of samples and the number of parameters [31]. Agent-based models typically have many parameters and each model evaluation can take minutes. Thus direct evaluation of global sensitivity indices can be prohibitive.

In this paper we present a global sensitivity approach based on a meta-model (surrogate model or response surface) corresponding to a given agent-based model. Unlike techniques based on a linear fit of model outputs with respect to model parameters, our meta-model is an accurate support-vector regression based analytical model representation, which preserves interactions between model parameters. We discuss global variance-based and derivative-based sensitivity indices. We demonstrate agent-based model sensitivity analysis on an example of the civil violence/criminal activity agent-based model. We also perform model reduction process for the given high dimensional agent-based model.

2. Model analysis

Global sensitivity analysis [27] is a comprehensive approach to the model analysis. Both variance-based and derivative-based global sensitivity can be calculated. The input factors responsible for model variability are identified and their contribution to variability of model outputs is quantified.

2.1. Variance-based global sensitivity

Let $f(x_1,\dots,x_n)$ be a square integrable function defined in the domain R^n. The inputs are treated as random variables and their probability density functions represent the associated uncertainty. The impact of the multiple input variables on the output can be independent as well as cooperative, and the analysis of variance (ANOVA) expresses the model output $f(\mathrm{x})$ as a finite hierarchical cooperative function expansion in terms of its input variables. In order to express the input–output relationship of complex models with a large number of input variables, the mapping between the input variables $x_1,\dots,x_n$ and the output variables $f(\mathrm{x})=f(x_1,\dots,x_n)$ in the domain R^n can be written in the following form [27]:

$$f(\mathrm{x})=f_0+\sum_i f_i(x_i)+\sum_{i<j} f_{i,j}(x_i,x_j)+\cdots+f_{1,2,\dots,n}(x_1,x_2,\dots,x_n),$$

where f_0 is the constant mean effect (zeroth order), function $f_i(x_i)$ is a first order term describing the effect of variable x_i acting independently upon the output $f(\mathrm{x})$, function $f_{i,j}(x_i,x_j)$ is a second order term describing the cooperative effects of variables x_i and x_j upon the output $f(\mathrm{x})$. The higher order terms reflect the cooperative effects of increasing numbers of input variables acting together to influence the output $f(\mathrm{x})$. The last term $f_{1,2,\dots,n}(x_1, x_2,\dots,x_n)$ contains any residual nth order cooperative contribution of all input variables. All terms in ANOVA decomposition are orthogonal to each other.

The total variance D is computed as follows:

$$D=\int (f(\mathrm{x})-f_0)^2\rho(\mathrm{x})\,d\mathrm{x}, \tag{1}$$

where $\rho(\mathrm{x})$ is the probability density of distribution of input variables.

Partial variances are defined as follows:

$$D_{i_1,\dots,i_s}=\int f^2_{i_1,\dots,i_s}(x_{i_1},\dots,x_{i_s})\rho(\mathrm{x})\,d\mathrm{x}.$$

The total partial variances D_i^{tot} for each parameter x_i, $i=\overline{1,n}$, can be obtained as

$$D_i^{tot}=\sum_{\langle i\rangle} D_{i_1,\dots,i_s};\quad 1\le s\le n,$$

where $\langle i\rangle$ means summation over all $D_{i_1,\dots,i_s}$ that contain index i.

After the total partial variances are determined, the total sensitivity indices can be calculated as follows:

$$S_i^{tot}=\frac{D_i^{tot}}{D};\quad 0\le S_i^{tot}\le 1. \tag{2}$$

By definition, the total partial variance D_i^{tot} for each parameter x_i is

$$D_i^{tot}=D-\mathrm{Var}(\mathbb{E}(f|\mathrm{x}_{-i}))\equiv\mathbb{E}(\mathrm{Var}(f|\mathrm{x}_{-i})), \tag{3}$$

where $f|\mathrm{x}_{-i}$ is a function of x_i with all other parameters fixed.

For independent parameters, Eq. (3) can be rewritten in the following way [28,30]:

$$D_i^{tot} = \frac{1}{2}\int\int [f(x) - f(x_1, \ldots, x_{i-1}, x'_i, x_{i+1}, \ldots, x_n)]^2 \rho(x'_i)\, dx'_i \rho(x)\, dx. \tag{4}$$

It is easy to see that direct application of Eqs. (1)–(4) requires $(n+1)N$ model evaluations to compute the total sensitivity indices for all input parameters, where n is the number of parameters and N is the number of samples needed to compute each integral.

2.2. Derivative-based global sensitivity

Derivative sensitivity indices N_i^{tot} in norm L_2 for each parameter x_i, $i = \overline{1, n}$, can be calculated as follows:

$$N_i^{tot} = \frac{\alpha_i \sigma_i^2}{D}\int \left[\frac{\partial f(x)}{\partial x_i}\right]^2 \rho(x)\, dx, \tag{5}$$

where D is given by Eq. (1), variance $\sigma_i^2 = \frac{1}{2}\int (x_i - x'_i)^2 \rho(x_i)\, dx_i \rho(x'_i)\, dx'_i$, and α_i is a constant for each distribution $\rho(x_i)$. For example, $\alpha_i = 1$ for normal distribution, $\alpha_i = 12/\pi^2$ for uniform distribution, $\alpha_i = 4$ for exponential distribution. Derivative sensitivity indices N_i^{tot} are upper bounds of the corresponding variance-based sensitivity indices D_i^{tot}: $D_i^{tot} \le N_i^{tot}$ (see Ref. [28]).

L_1-norm derivative sensitivity indices can be calculated as follows:

$$L_i^{tot} = \sqrt{\frac{\alpha_i \sigma_i^2}{D}}\int \left|\frac{\partial f(x)}{\partial x_i}\right| \rho(x)\, dx. \tag{6}$$

As in the case of variance-based sensitivity indices, one has to perform $(n+1)N$ model evaluations when directly applying Eq. (5) or (6) to compute the derivative-based sensitivity indices.

2.3. Meta-model-based estimation of global sensitivity indices

Advanced agent-based models have tens to thousands parameters (large n). Thus, the number of required model evaluations $(n+1)N$ is usually prohibitively large for global sensitivity analysis. We have developed software GoSUM (Global Optimization, Sensitivity and Uncertainty in Models) [47] that makes the number of model evaluations independent on n. This makes GoSUM a perfect tool for analysis of agent-based models with large running times and large number of parameters.

The appropriate number of samples is created, for which an agent-based model is executed, then it is learnt (a meta-model is created) by using support vector regression (SVR) with Gaussian kernels. The global sensitivity indices are still calculated for each variable, but the required samples are extracted from the meta-model, which is many orders of magnitude faster than the original agent-based model. In the case of independent parameters, techniques described in Ref. [29] are used to estimate variance sensitivity indices. In the case of correlated parameters, we use Eq. (3) directly. Global derivative sensitivity indices are computed using Eqs. (5) and (6).

Meta-models are nowadays often used to model computationally intensive problems from various areas such as aerospace, electronics, automotive industry, chemistry, finance. The basic approach is to construct the simplified model that is computationally efficient and can accurately predict the characteristics of a product. There are a lot of metamodeling techniques: the response surface methodology [32,33], Kriging model [34], neural networks [35], radial basis functions [36], multivariate adaptive regression splines [37], inductive learning [38]. In the case of highly nonlinear models the regression version of support vector machines is applied [39–41]. The support vector regression based meta-models are used in various applications such as automobile, ship and aircraft design, crashworthiness design and many others [42–44].

3. Global sensitivity analysis of a socio-cultural agent-based model

In this section we demonstrate the global sensitivity analysis techniques on the example of the civil violence/criminal activity agent-based model developed by the present authors in Ref. [45]. The coarse-grained version of the model can be found in Ref. [46]. There are two kinds of agents in the model: citizens and law enforcement officers (LEO's). Citizens are members of the population and LEO's are the forces of the authority. All events transpire on a lattice with periodic boundaries. All citizens move once per day, and LEO's move several times per day.

3.1. Description of the agent-based model and main results

Each citizen is assigned hardship H drawn from the uniform distribution $U(0,1)$. The hardship is heterogeneous across citizens and is fixed for each citizen. The perceived legitimacy L of the law enforcement authority is equal across citizens and can be between 0 and 1. The hardship and the legitimacy are used to define citizen's grievance $E = H(1-L)$. Some citizens are more willing to pursue criminal activity than others, this is encoded by the risk aversion K. Each individual's risk aversion is drawn from $U(0,1)$ and is fixed for each citizen. The citizen's vision ν is a circle of radius ν that comprises lattice positions that the citizen is able to inspect. The vision is equal across citizens. The citizen's perceived net risk N is defined as follows: $N = KP$, where P is a function of the ratio of LEO's to active citizens. If for a law-abiding citizen the difference $E-N$ exceeds T, where T is some threshold, then the citizen becomes criminally active. If, for an active citizen, the difference $E-N$ exceeds T, then the citizen stays active. Otherwise, he/she becomes law-abiding. In summary, the citizen's rule for being active or quiescent is the following: If $E-N > T$ be active; otherwise, be quiescent.

The citizen's perceived risk function P is defined as

$$P(C/A) = 1 - \exp(-k'(C/A)) \sum_{i=0}^{15} \frac{(k'(C/A))^i}{i!}, \tag{7}$$

where A is the number of active citizens (including self) within citizen's vision, C is the number of LEO's within citizen's vision, constant $k' = 62.6716$ is found from the condition that $P(1/4) = 0.5$ (see Fig. 1).

The perceived risk is in fact zero up to a threshold value, after which it increases monotonically, thus giving it a sigmoidal shape. The sigmoidal shape encodes a level of irrational behavior by citizens, where the real risk of being incarcerated is being diminished by the

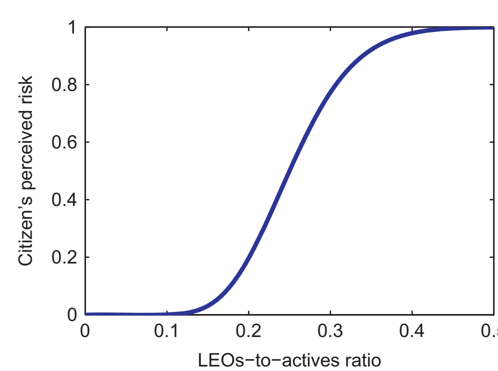

Fig. 1. Citizen's perceived risk function.

proportion of others in the same situation. The law enforcement operates through the rule that a LEO agent arrests the nearest active citizen. This rule leads to dynamics in which local crime hotspots attract police action, thus reflecting the modern problem-oriented policing law enforcement strategies. For more information on roots of the model in criminology theory see [45].

In general, citizens are less likely to engage in violence as the local ratio between number of LEO's and active citizens increases due to a fear of being identified and incarcerated. The citizen state (active or quiescent) can be regarded as a function of threshold T. Fig. 2 shows the structure of population depending on threshold. In particular, in the case when $T>1-L$, all citizens are always quiescent regardless of the lattice situation; we call these citizens "never active" and denote their fraction in the population as G. In the case when $T<-1$, all citizens are "always active", and we denote their fraction in the population as R. When $-1<T<0$, $G=0$ and all citizens are either "always active" or "conditionally active" (active or quiescent depending on the lattice situation). The case when $0<T<1-L$, is the most realistic one with all three groups of population present. In practice the threshold T and legitimacy L for the model run can be found using statistical data on fractions of R and G in the population

$$T=\frac{2R}{1-G}; \quad (8)$$

$$L=1-\frac{2R}{G(1-G)}. \quad (9)$$

LEO's seek out and arrest active citizens. The LEO's vision w is a circle of radius w that comprises lattice positions that the LEO is able to inspect. It is equal across LEO's. The LEO's rule is the following: Inspect all sites within w and arrest the nearest active citizen. Jail terms for arrested actives are assigned randomly from $U(0,J_{max})$, where J_{max} is the maximum jail term. Citizens and LEO's move on the lattice by using the following movement rule: Pick a random neighboring location on the lattice (from Moore neighborhood: 8 adjacent cells), if that location is unoccupied – move there, if the location is occupied – stay put.

In such a way, citizens can be active in criminal and/or violent activity, stay quiescent, or be jailed (see Fig. 3).

The procedure of a run is as follows. A citizen or a LEO is selected at random. The probability of selecting a LEO is higher according to the number of moves a LEO can make per day. If the selected person is a non-jailed citizen, then he/she moves according to the movement rule; if the citizen is in jail, then the days at jail are calculated and if the number of days at jail is equal to the assigned jail term the released citizen is put on a random unoccupied site on the lattice. After that the state of the citizen is calculated depending on the current lattice situation. If the selected person is a LEO, then he/she inspects all sites within the vision, arrests the nearest active citizen (if any) and jumps to the last location of the arrested citizen. The jailed citizens are

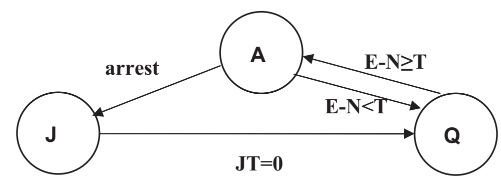

Fig. 3. The active citizen (A) can retain its active state, go to jail (J) if the LEO arrests him/her, or become quiescent (Q) if the difference between the citizen's grievance E and the citizen's perceived net risk N is smaller than the specified threshold. The quiescent citizen (Q) can retain its quiescent state or become active (A) if the difference between the citizen's grievance E and the citizen's perceived net risk N is bigger or equal to the specified threshold. The jailed citizen (J) becomes quiescent (Q) when its jail term (JT) is over.

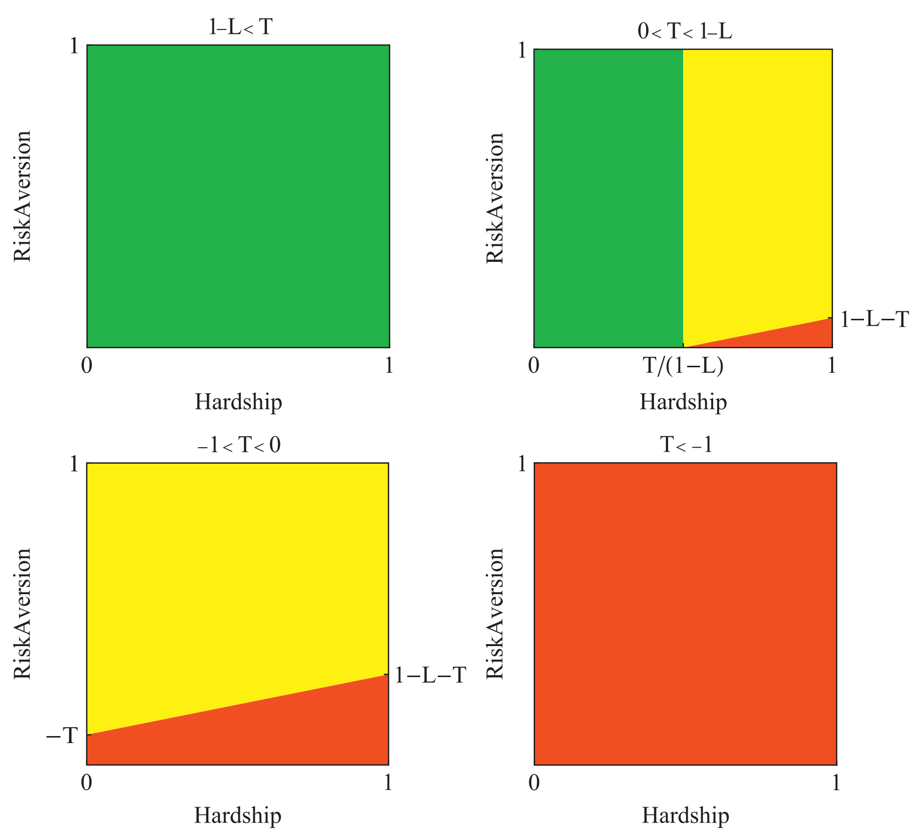

Fig. 2. Citizen state as a function of threshold. Green – fraction of citizens who are never active, red – fraction of citizens who are always active, yellow – fraction of citizens who use arrest probability to decide their state. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

placed outside the lattice. Then the LEO moves under the movement rule. The model iterates this procedure until the simulation time is reached.

In [45] it was shown that the proportion of law enforcement officers required to maintain a steady low level of criminal activity increases with the size of the population of the city. The nature of violence changes from global outbursts of criminal/violent activity in small cities to spatio-temporally distributed decentralized outbursts of activity in large cities, indicating that in order to maintain peace, bigger cities need larger ratio of law enforcement officers than smaller cities. It was also deduced that the number of criminal/violent events per 1,000 inhabitants of a city shows non-monotonic behavior with size of the population. The existence of tipping points for communities of all sizes in the model was observed: reducing the number of law enforcement officers below a critical level can rapidly increase the incidence of criminal/violent activity. These trends are in complete agreement with the FBI data [45].

3.2. Uncertainty analysis and global sensitivity analysis

In the following we analyze the global sensitivity of the agent-based model described in this section. Eight input parameters and nine outputs of the model were considered. Lattice size, citizen vision, LEO vision, LEO speed (number of times a LEO moves per day), maximum jail term, LEO density, always actives ratio, and never actives ratio are input parameters in the model (see Table 1). Number of actives not in jail per 1000 citizens, number of violent outbursts per year, peak number of active citizens per 1000 citizens, rate of violence per 1000 citizens, number of times an always active citizen is arrested, number of times a conditionally active citizen is arrested, probability of not being arrested for an always active citizen, probability of not being arrested for a conditionally active citizen, and probability of not being arrested for a never active citizen are considered to be outputs of the model (see Table 2). The probability of not being arrested for a citizen is defined by the value of the citizen's perceived risk function P given by Eq. (7).

Since the model is stochastic (agents can move randomly), there is intrinsic uncertainty in the model outputs even when all model parameters are fixed. First, we performed uncertainty analysis in the case of fixed model parameters. The agent-based model was run 5040 times for the duration of 5000 model days with 40,000 agents on the lattice. The parameters of the model are as follows: lattice size$=$240; citizen vision$=$14; LEO vision$=$14; LEO speed$=$4; maximum jail term$=$120; LEO density (LEOs per cell)$=$0.01; always actives ratio$=$0.025; never actives ratio$=$0.5. Average number of times a citizen is arrested and the probability of not being arrested for a citizen are presented in Fig. 4. As it can be clearly seen the sections in this figure repeat those from Fig. 2 in the case $0<T<1-L$. The plot of the number of times a citizen is arrested is self-explanatory. Never active citizens are not arrested and always active citizens are arrested often. The more active is the citizen, the higher probability he/she has of being arrested. The plot of the probability of not being arrested for a citizen shows that the probability of not being arrested for an always active citizen is high, what can be explained by the fact that an always active citizen is almost always in jail.

Table 1
Model input parameters

Input parameter number	Input parameter name
1	Lattice size
2	Citizen vision
3	LEO vision
4	LEO speed
5	Maximum jail term
6	LEO density
7	Always actives ratio
8	Never actives ratio

Table 2
Model outputs.

Output number	Output name
1	Number of actives not in jail per 1000 citizens
2	Number of violent outbursts per year
3	Peak number of active citizens per 1000 citizens
4	Rate of violence per 1000 citizens
5	Number of times an always active citizen is arrested
6	Number of times a conditionally active citizen is arrested
7	Probability of not being arrested for an always active citizen
8	Probability of not being arrested for a conditionally active citizen
9	Probability of not being arrested for a never active citizen

In the following 10% uncertainty in the model parameters was introduced. The 5040 samples within the space of parameters were generated by using the GoSUM software [47]. GoSUM software is a tool for producing sampling points, evaluating global uncertainty in the model outputs and global contributions of a large number of uncertain parameters to the model outputs. The model parameters were distributed uniformly in the following ranges: lattice size: [216; 264]; citizen vision: [12; 16]; LEO vision: [12; 16]; LEO speed: [2; 6]; maximum jail term: [108; 132]; LEO density: [0.009; 0.011]; always actives ratio: [0.0225; 0.0275]; never actives ratio: [0.45; 0.55]. For each sample the model was run for the duration of 5000 model days with 40,000 agents on the lattice. In this case the number of uncertain parameters is only increased by 8 comparing to the initial model, where the number of uncertain parameters is of order of $2\cdot\texttt{agents}\cdot(\texttt{model days}+1)$. For each agent for every model day we approximately need two random numbers, one of which is used for the random order of selecting an agent and the other one is used for the random positioning of the agent in the Moore neighborhood. Average number of times a citizen is arrested and the probability of not being arrested for a citizen in this case are presented in Fig. 5.

As seen from Figs. 4 and 5, when the model parameters are uncertain, the number of times a “conditionally active” citizen is arrested is lower than in the case with no uncertainty. The number of times an “always active” citizen and a “never active” citizen are arrested are the same in both cases. The probability of not being arrested for an “always active” citizen is as high as in the case with no uncertainty, and the probability of not being arrested for a “conditionally active” citizen and a “never active” citizen is lower than in the case with no uncertainty.

In Fig. 6 we compare the distribution of model outputs with (10%) and without uncertainty in model parameters. The histograms of all 9 outputs of the model in the case of 10% uncertainty are plotted in green and the fitting curves for histograms in the case with no uncertainty are plotted in red. As it can be seen from the figure, the red line is located around the mean value of the output in the model with 10% uncertainty in parameters. Because of some combinations of parameters which produce no outbursts of activity during the whole time of the simulation or the constant outburst of activity, there are many zero output values. Relatively small uncertainty in the input parameters produces large output uncertainty. The values of mean and standard deviation (uncertainty) in units of mean are indicated on the top of the histogram for each output. The only output with small uncertainty is the number of times an always active citizen is arrested.

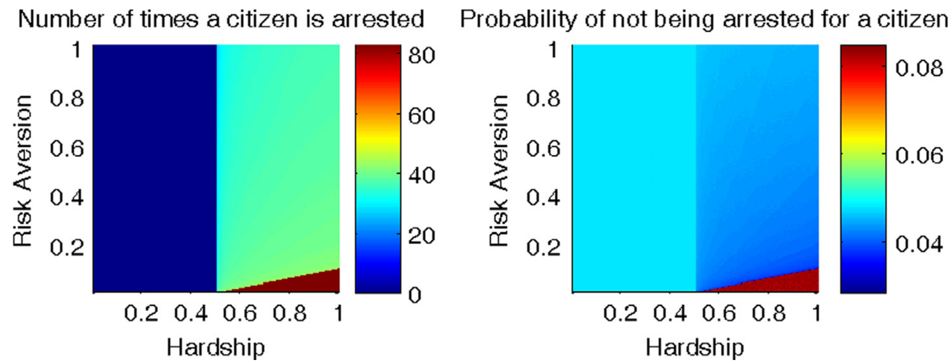

Fig. 4. Average number of times a citizen is arrested (left panel) and the probability of not being arrested for a citizen (right panel) in the model with fixed parameters. Each pixel in these plots corresponds to one citizen.

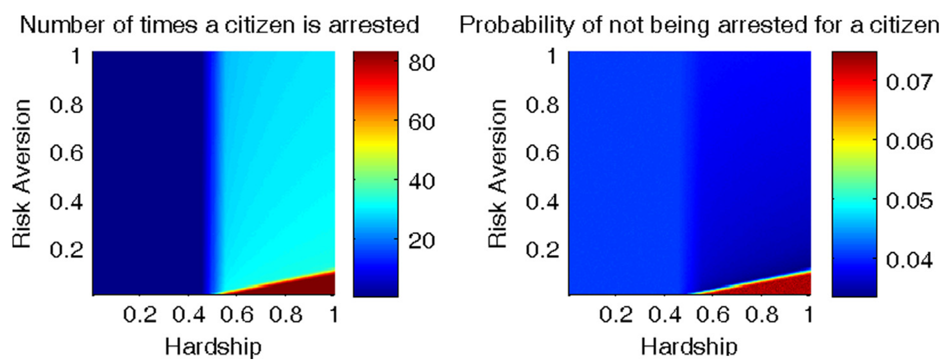

Fig. 5. Average number of times a citizen is arrested (left panel) and the probability of not being arrested for a citizen (right panel) in the model with 10% uncertainty in parameters. Each pixel in these plots corresponds to one citizen.

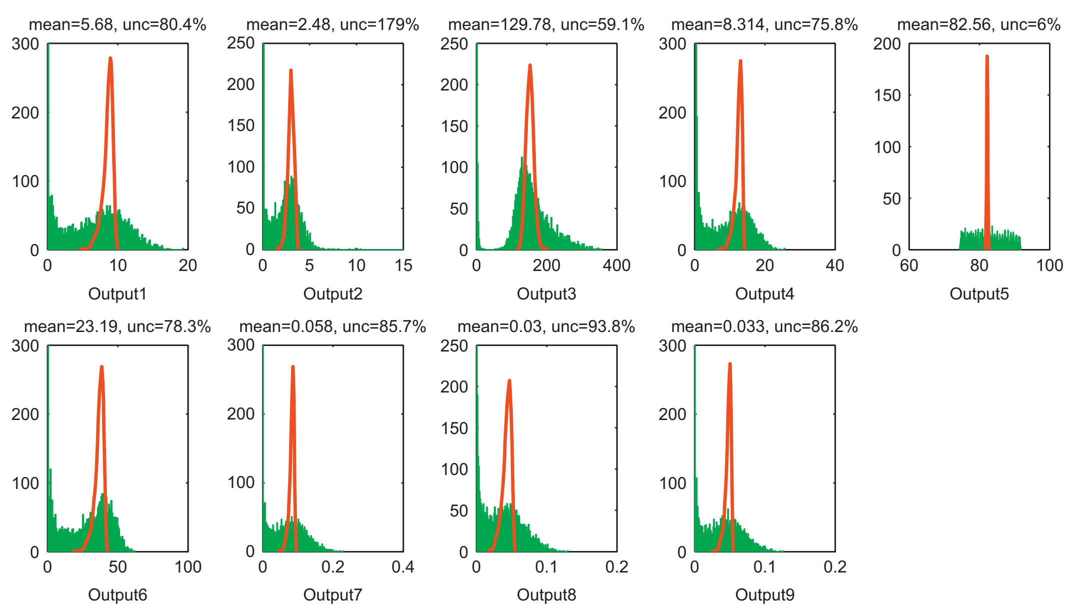

Fig. 6. Distribution of model outputs. No uncertainty in parameters (red), 10% uncertainty in parameters (green). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

The bi-modality can be seen from the figure as well. This demonstrates the fact that a small change in parameters can cause significant changes in the model outputs. The scenario with periodic outbursts can be changed to the ones with no outbursts of activity or the constant outburst of activity.

By using the GoSUM software [47], the global sensitivity analysis of all model outputs for all parameters was performed. Global derivative sensitivity in L_2 norm is presented in Fig. 7, global derivative sensitivity in L_1 norm is presented in Fig. 8, and global variance sensitivity is presented in Fig. 9.

Both global variance and derivative sensitivities identify LEO vision (input parameter 3) as the most important parameter for all outputs except the output of number of times an always active citizen is arrested. The biggest influence of the LEO vision is on the number of violent outbursts per year (output 2) and the peak number of active citizens per 1000 citizens (output 3) as well as

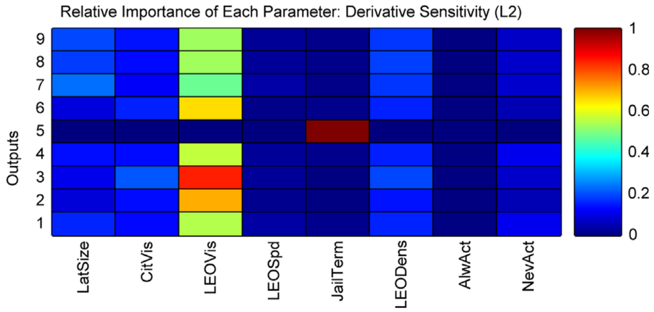

Fig. 7. Heat map of global derivative sensitivity in norm L_2.

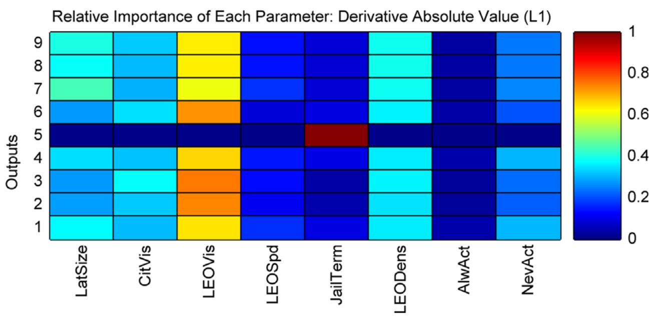

Fig. 8. Heat map of global derivative sensitivity in norm L_1.

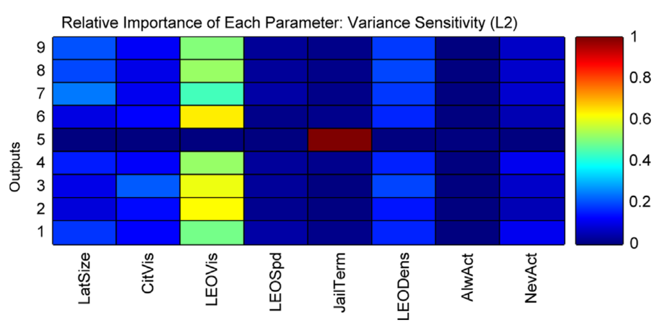

Fig. 9. Heat map of global variance sensitivity.

the number of times a conditionally active citizen is arrested (output 6). LEO density (input parameter 6), lattice size (input parameter 1) and citizen vision (input parameter 2) are also important parameters as can be clearly seen from the global derivative sensitivity in norm L_1 heat map in Fig. 8. Typically, L_2 norm global derivative sensitivity indices and global variance sensitivity indices are similar. On the other hand, L_1 norm global derivative sensitivity indices allow us to identify all parameters responsible for function variability, even those parameters with small contribution to the total variance.

3.3. *Rate of violence and model reduction*

In this subsection we perform model reduction and deduce important relationships between model parameters for our example agent-based model. Model reduction is the process of decreasing the number of model parameters by taking off the least important ones. In such a way the decision making process for a high-dimensional model can become low-dimensional. Note that parameters' importance depends on their range of variation. A parameter can be very important when varied in a wide range. However, if the range of variation is known to be small, the parameter may not be important compared to other parameters in the model.

In the following we introduce the notion of the rate of violence as an average number of citizens active at the end of a day or jailed during the day per 1000 citizens. We study the dependence of rate of violence on model parameters. We consider that rate of violence to be high if there are more than 2 active citizens per day per 1000 citizens and we consider that rate of violence to be low if there are less or equal than 2 active citizens per day per 1,000 citizens. We deduce relationships between the parameters of the agent-based model which are required in order to have low rate of violence.

First, by using the results of the simulation with 10% uncertainty in parameters, we investigate the dependence of rate of violence on all pairs of parameters. We detect that the strong differentiation between the high rate of violence and the low rate of violence is observed depending on the values of LEO vision and citizen vision. In Fig. 10 the red area shows high rate of violence (outbursts of activity) and the blue area represents low rate of violence (stable situation). To account for this differentiation, we introduce the notion of relative LEO vision: $\texttt{Relative LEO vision} = 3{*}\texttt{LEO vision} - 2{*}\texttt{Citizen vision}$, which is parallel to the green line in Fig. 10. In such a way, we do the coordinate transformation from citizen vision and LEO vision to citizen vision and relative LEO vision. The fixed value of the relative LEO vision at about 15.35 separates the high rate of violence and the low rate of violence.

When the relative LEO vision is fixed, LEO density is the most important parameter that discriminates between stable and unstable situations on the lattice [45]. We showed that there is a critical value of LEO density, or a tipping point, below which the situation becomes unstable [45]. In the following we are going to find a mathematical expression for the critical LEO density. In order to estimate the critical LEO density as a function of other model parameters, we consider a very wide range for all model parameters to see how the model depends on the input parameters not only around nominal value but in the wide range. We also introduce the relative LEO vision parameter in the model. In such a way we consider the following parameter ranges: lattice size: [210; 350]; citizen vision: [5; 20]; relative LEO vision: [5; 20]; LEO speed: [2; 8]; maximum jail term: [50; 300]; LEO density: [0.005; 0.02]; always actives ratio: [0.01; 0.04]; never actives ratio: [0.3; 0.6]. The LEO vision is calculated as follows: $\texttt{LEO vision} = (\texttt{Relative LEO vision} + 2{*}\texttt{Citizen vision})/3$. The number of realizations of this model is 10,000 with the duration of each realization of 5000 model days and 40,000 agents on the lattice. As above, we consider the dependence of rate of violence on all

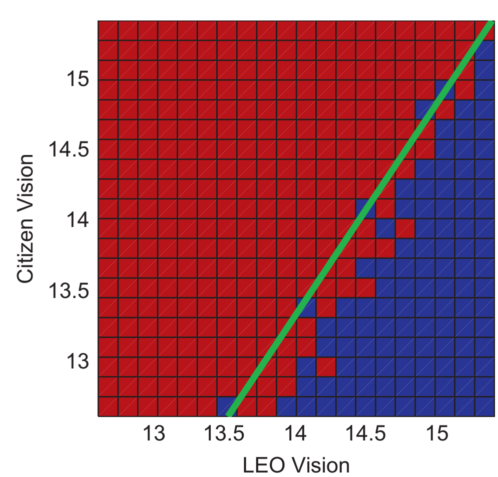

Fig. 10. Differentiation between the high rate of violence (red) and the low rate of violence (blue) depending on LEO vision and citizen vision. Green line corresponds to relative LEO vision = 15.35. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

pairs of parameters. The strong differentiation between the high rate of violence and the low rate of violence was observed depending on the values of citizen vision and LEO density (see Fig. 11, where the red area shows high rate of violence and the blue area represents low rate of violence). In such a way, the following relationship for the optimal LEO density per area to have a stable situation was deduced: $\texttt{LEO Density} = 0.005 + 0.75/(\texttt{Citizen Vision}^2)$, which is presented by green curve in Fig. 11. In terms of

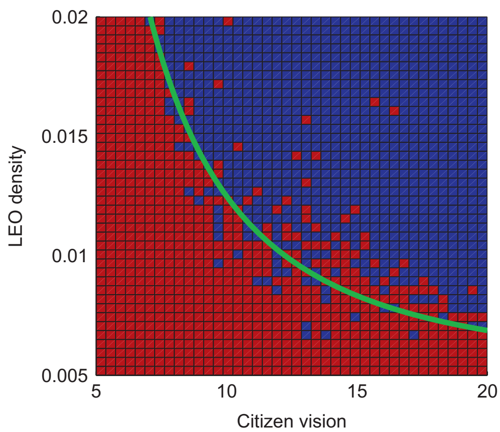

Fig. 11. Differentiation between the high rate of violence (red) and the low rate of violence (blue) depending on citizen vision and LEO density. Green curve corresponds to $\texttt{LEO Density} = 0.005 + 0.75/(\texttt{Citizen Vision}^2)$. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

the number of visible neighbors N_v, the number of LEOs per citizen needed to maintain the same rate of violence can be approximately calculated as follows: $\texttt{LEOs per citizen} = 0.01 + 0.75 * \pi / N_v$. Thus, the more informed are the citizens, the less LEOs is needed to maintain peace.

In real life, many of the model parameters are tightly controlled and could be considered less important than city-dependent parameters like citizen density, number of citizens, and number of LEOs per citizen. In the following we are going to find the critical number of LEOs as a function of population size and density. We fix LEO vision, citizen vision, LEO speed, maximum jail term, always actives ratio, never actives ratio and vary citizen density, number of citizens, and number of LEOs per citizen. We consider the following ranges of parameters: citizens density: [0.2; 0.8]; number of citizens: [1000; 400000] in log scale; LEOs per citizen: [0.015; 0.025]. Other parameters are fixed as follows: LEO vision = 12.5; citizen vision = 12.5; LEO speed = 5; maximum jail term = 175; always actives ratio = 0.025; never actives ratio = 0.45. Lattice size is calculated as: $LD = \sqrt{\texttt{Citizens/Citizens density}}$. $\texttt{LEO density} = \texttt{LEOs per citizen} * \texttt{Citizens density}$. The number of realizations of this model is 10,000 with the duration of each realization of 1000 model days. Comparing the high rate of violence and the low rate of violence, we observe almost no dependence on citizen density and logarithmic dependence on population size (see Fig. 12). Thus, in order to keep stability in a city, the number of LEOs per citizen should increase with logarithm of population.

Another quantity of interest in our agent-based model is the waiting time between outbursts of activity. By the waiting time between outbursts of activity we define the time (in model days) between the peak of one outburst and the peak of the next

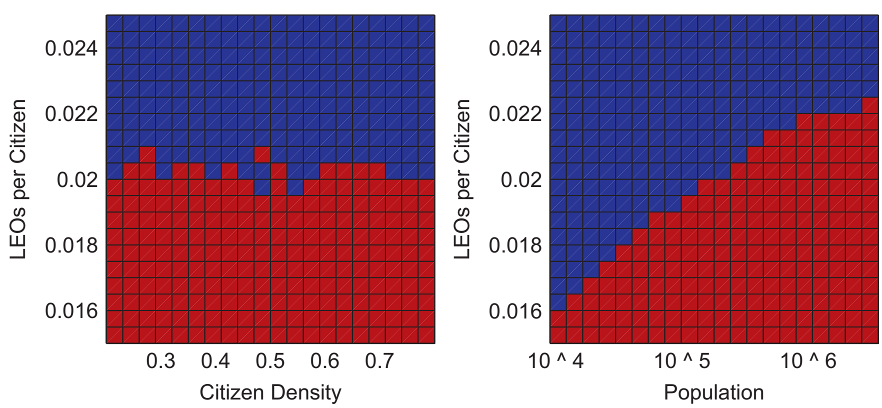

Fig. 12. Differentiation between the high rate of violence (red) and the low rate of violence (blue) depending on LEOs per citizen and citizens density as well as depending on LEOs per citizen and number of citizens. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

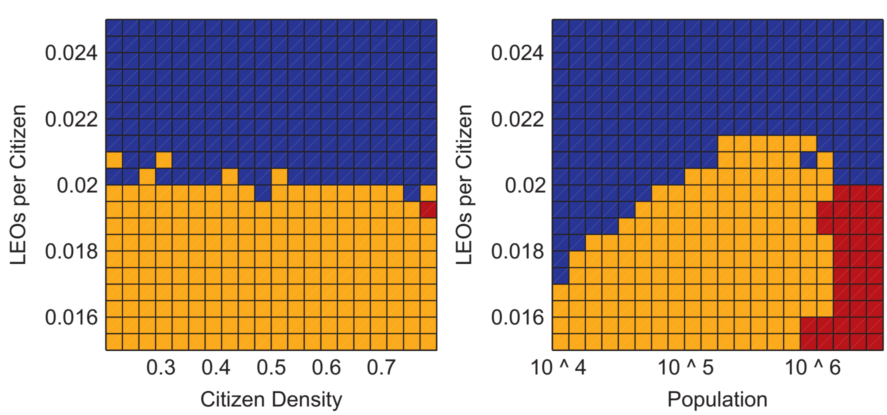

Fig. 13. Differentiation between no outbursts of activity (blue), waiting time between outbursts more or equal 87.5 (orange), waiting time between outbursts less than 87.5 (red). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

outburst. We differentiate between a scenario with no outbursts of activity, a scenario with waiting time between outbursts more or equal 87.5 model days and a scenario with waiting time between outbursts less than 87.5 model days (see Fig. 13). Similarly to the case of the rate of violence, almost no dependence on citizen density is found for the waiting time. The dependence of the waiting time between outbursts on the number of LEOs per citizen and on the population size is also similar to those for the rate of violence. In Fig. 13 the blue area shows no outbursts, the orange area corresponds to periodic outbursts and the red area shows immediate outbursts. As seen, the correct proportion of LEOs is very important for cities with over a million citizens. In such cities, incorrect number of LEOs per citizen would result in an immediate outburst of activity.

4. Conclusions

Global sensitivity analysis was found to be an important tool for analyzing and evaluating agent-based models. We proposed a meta-model-based technique to evaluate various global sensitivity indices for all parameters in the model. Our technique requires only N expensive agent-based model evaluations, where N is almost independent of the number of parameters in the problem and is chosen to provide the requested accuracy of the fit. Support vector regression based meta models preserve all interactions between model parameters and are applicable for non-monotonic and nonlinear model outputs.

We identified the most significant and non-significant parameters in the example socio-cultural agent-based model. By using both global variance and global derivative sensitivities, we found that LEO vision is the most important parameter for all outputs of the model (except the output of number of times an always active citizen is arrested). We observed that the biggest influence of the LEO vision is on the number of violent outbursts per year and the peak number of active citizens per 1000 citizens as well as the number of times a conditionally active citizen is arrested. These findings are in good agreement with the reality [45].

We also provided examples of the model reduction process, which allowed us to investigate the dependence of high and low rate of violence and outbursts waiting time on the model parameters. We found that less LEOs are needed when all citizens are well-informed (have more neighbors). It was also shown that in order to keep stability in a city, the number of LEOs per citizen should increase approximately as logarithm of population. Finally, the correct number of LEOs was found to be critical to keep the violence rate low for very large cities.

Acknowledgments

This work was partially supported by AFOSR Grants FA9550-08-1-0217 and FA9550-12-1-0230.

References

[1] Bonabeau E, Dorigo M, Theraulaz G. Swarm intelligence: from natural to artificial systems. New York: Oxford University Press; 1999.
[2] Weiss G, editor. Multiagent systems: a modern approach to distributed artificial intelligence, MIT Press, Cambridge, MA; 2000.
[3] Wooldridge M. An introduction to multiagent systems. Chichester, UK: Wiley; 2002.
[4] LeBaron B. Empirical regularities from interacting long- and short-memory investors in an agent-based stock market. IEEE Transactions on Evolutionary Computation 2001;5(5):442–55.
[5] De Wilde P. Fuzzy utility and equilibria. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 2004;34(4):1774–85.
[6] Sim KM, Wong E. Toward market-driven agents for electronic auction. IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans 2001;31(6):474–84.
[7] Huang C-C. Using intelligent agents to manage fuzzy business processes. IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans 2001;31(6):508–23.
[8] Troisi A, Wongt V, Ratnert M. An agent-based approach for modeling molecular self-organization. Proceedings of the National Academy of Sciences of the United States of America 2005;102(2):255–60.
[9] Grimm V, Revilla E, Berger U, Jeltsch F, Mooij W, Railsback S, et al. Pattern-oriented modeling of agent-based complex systems: lessons from ecology. Science 2005;310:987–91.
[10] Axelrod R. The complexity of cooperation. Princeton: Princeton University Press; 1997.
[11] Epstein JM, Axtell RL. Growing artificial societies: social science from the bottom up. Cambridge, MA: MIT Press; 1996.
[12] Epstein JM. Modeling civil violence; an agent-based computational approach. Proceedings of the National Academy of Sciences of the United States of America 2002;99:7243–50.
[13] Epstein JM. Generative social science: studies in agent-based computational modeling. Princeton, NJ: Princeton University Press; 2006.
[14] Bonabeau E. Agent-based modeling methods and techniques for simulating human systems. Proceedings of the National Academy of Sciences of the United States of America 2002;99(Suppl. 3):7280–7.
[15] Helbing D, Farkas I, Vicsek T. Simulating dynamical features of escape panic. Nature 2000;407:487–90.
[16] Read DW. A multitrajectory competition model of emergent complexity in human social organization. Proceedings of the National Academy of Sciences of the United States of America 2002;99(Suppl. 3):7251–6.
[17] Kaegi M, Mock R, Kroger W. Analyzing maintenance strategies by agent-based simulations: a feasibility study. Reliability Engineering and System Safety 2009;94:1416–21.
[18] Saltelli A, Ratto M, Tarantola S, Campolongo F. Sensitivity analysis practices: strategies for model-based inference. Reliability Engineering and System Safety 2006;91(10–11):1109–25.
[19] Trucano TG, Swiler LP, Igusa T, Oberkampf WL, Pilch M. Calibration, validation, and sensitivity analysis: what's what. Reliability Engineering and System Safety 2006;91(10–11):1331–57.
[20] Helton JC, Johnson JD, Sallaberry CJ, Storlie CB. Survey of sampling-based methods for uncertainty and sensitivity analysis. Reliability Engineering and System Safety 2006;91(10–11):1175–209.
[21] Perez L, Dragicevic S. An agent-based approach for modeling dynamics of contagious disease spread. International Journal of Health Geographics 2009;8:50.
[22] Happe K. Agent-based modelling and sensitivity analysis by experimental design and metamodelling: an application to modelling regional structural change. In: XIth international congress of the European association of agricultural economists, the future of rural Europe in the global agri-food system, Copenhagen, Denmark; August 24–27, 2005.
[23] Segovia-Juarez JL, Ganguli S, Kirschner D. Identifying control mechanisms of granuloma formation during *M. tuberculosis* infection using an agent-based model. Journal of Theoretical Biology 2004;231:357–76.
[24] Dancik GM, Jones DE, Dorman KS. Parameter estimation and sensitivity analysis in an agent-based model of Leishmania major infection. Journal of Theoretical Biology 2010;262(3):398–412.
[25] Beaudouin R, Monoda G, Ginot V. Selecting parameters for calibration via sensitivity analysis: an individual-based model of mosquitofish population dynamics. Ecological Modelling 2008;218(1–2):29–48.
[26] Saltelli A, Annoni P. How to avoid perfunctory sensitivity analysis. Environmental Modelling & Software 2010;25:1508–17.
[27] Sobol I. Sensitivity estimates for nonlinear mathematical models. Matematicheskoe Modelirovanie 1990;2(1):112–8 [in Russian]; English translation: Mathematical and Computer Modelling Experiment 1993;1(4):407–14.
[28] Sobol IM, Kucherenko S. Derivative based global sensitivity measures and their link with global sensitivity indices. Mathematics and Computers in Simulation 2009;79:3009–17.
[29] Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, Tarantola S. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Computer Physics Communications 2010;181:259–70.
[30] Sobol IM. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Maths and Computers in Simulation 2001;55(1–3):271–80.
[31] Saltelli A. Making best use of model evaluations to compute sensitivity indices. Computer Physics Communications 2002;145:280–97.
[32] Myers R, Montgomery D. Response surface methodology. 2nd ed. New York: Wiley-Interscience; 2002.
[33] Simpson TW, Peplinski JD, Koch PN, Allen JK. Metamodels for computer-based engineering design: survey and recommendations. Engineering Computations 2001;17(2):129–50.
[34] Sakata S, Ashida F, Zako M. Structural optimization using Kriging approximation. Computer Methods in Applied Mechanics and Engineering 2003;192(7–8):923–39.
[35] Cheng B, Titterington DM. Neural networks: a review from a statistical perspective. Statistical Science 1994;9(1):2–54.
[36] Hussain MF, Barton RR, Joshi SB. Metamodeling: radial basis functions, versus polynomials. European Journal of Operational Research 2002;138(1):142–54.

[37] Friedman JH. Multivariate adaptive regressive splines. The Annals of Statistics 1991;19(1):1–67.
[38] Langley P, Simon HA. Applications of machine learning and rule induction. Communications of the ACM 1995;38(11):55–64.
[39] Vapnik VN. The nature of statistical learning theory. Springer; 2000.
[40] Vapnik V, Golowich S, Smola A. Support vector method for function approximation, regression estimation, and signal processing. Advances in Neural Information Processing Systems 1997;9:281–7.
[41] Scholkopf B, Burges CJC, Smola AJ. Advances in kernel methods: support vector learning. Cambridge, Massachusetts: MIT Press; 1999.
[42] Pan F, Zhu P, Zhang Y. Metamodel-based lightweight design of B-pillar with TWB structure via support vector regression. Computers and Structures 2010;88(1–2):36–44.
[43] Backlund P, Shahan D, Seepersad C. Metamodeling techniques for multi-dimensional ship design problems. Ship Science and Technology, North America 2010;4(7):43–54.
[44] Wang H, Li E, Li GY. Probability-based least square support vector regression metamodeling technique for crashworthiness optimization problems. Computational Mechanics 2011;47(3):251–63.
[45] Fonoberova M, Fonoberov VA, Mezic I, Mezic J, Brantingham J. Nonlinear dynamics of crime and violence in urban settings. Journal of Artificial Societies and Social Simulation 2012;15(1) (url: ⟨http://jasss.soc.surrey.ac.uk/15/1/2.html⟩).
[46] Zou Y, Fonoberov VA, Fonoberova M, Mezic I, Kevrekidis IG. Model reduction for agent-based social simulation: Coarse-graining a civil violence model. Physical Review E 85, 066106 (13 pages); 2012.
[47] Aimdyn GoSUM Software, Global optimization, sensitivity and uncertainty in models (GoSUM) [online]. Santa Barbara, CA: Aimdyn Inc. Available from: ⟨http://aimdyn.com⟩; 2012.