Miguel, Amblard, Barceló & Madella (eds.) Advances in Computational Social Science and Social Simulation
Barcelona: Autònoma University of Barcelona, 2014, DDD repository <http://ddd.uab.cat/record/125597>

Sensitivity analysis for agent-based models: a low complexity test-case

Guus ten Broeke
Biometris
Wageningen University and Research
Wageningen, the Netherlands

George van Voorn
Biometris Wageningen U&R
Wageningen, the Netherlands
Email: george.vanvoorn@wur.nl

Arend Ligtenberg
Laboratory of Geo-information
Science and Remote Sensing
Wageningen, the Netherlands

***Abstract*—Methodologies for sensitivity analysis are considered to be of great importance for analyzing agent-based models (ABMs), even more because calibration and validation of ABMs often prove problematic. Different methodologies for sensitivity analysis may help to understand ABM dynamics and (thus) aid in the calibration and validation of ABMs. However, model complexity of ABMs is a significant hinderance for a detailed research on which (combination of) sensitivity analysis methods may provide the best option. We present here an agent-based model of low complexity to be used as test case for different methodologies of sensitivity analysis.**

I. Introduction

Calibration and validation are considered to be key challenges for agent-based modelling [1]. Sensitivity analysis is a statistical tool to analyze the effects of variations and uncertainty in input on the resulting output [2]. Sensitivity analysis can be helpful for model calibration and validation. For instance, it may reveal what level of data is required for certain inputs, and thus what type of experimental design is minimally required for model validation. A protocol for the numerical experimental design of sensitivity analysis for ABMs has recently been proposed [3].

One weak point of currently available methodologies for sensitivity analysis is that they are not particularly well-suited for models with strong nonlinearities and tipping points [4]. Yet, systems simulated by ABMs may display exactly these characteristics typical of Complex Adaptive Systems (CAS). For instance, an ABM simulating a fishery system in the Philippines shows a near linear increase in revenue under a decrease in the tightness of fishing quota, followed by a tipping point at which the ecological system - and hence the revenues - completely collapses [5]. This result was found using one-at-a-time sensitivity analysis.

It is conceivable that a combination of different methods for sensitivity analysis offers better insight on the sensitivity of models representing CAS. We therefore explore here a strategy of combining methodologies for sensitivity analysis to analyze an ABM. In practice, however, many ABMs have a rather high complexity, which means it may prove cumbersome to perform sensitivity analysis - up to the point of being practically infeasible. We therefore propose a low complexity test case ABM to ascertain the best strategy for doing sensitivity analysis. This approach allows us to start out simple, while letting us add further CAS properties to the model once the present version has been sufficiently analyzed.

TABLE I: Process overview of the model

sites: grow and diffuse resource
agents: observe
agents: harvest?
agents: move?
agents: pay maintenance
agents: die?
agents: breed?

TABLE II: Table containing the main variables and parameters

i	Agent index
n_{tot}	Total number of agents
$N_x \times N_y$	Field size
t	Time index
x, y	Spatial orientation indices
$E_i[t]$	Internal energy content of agent
$n_{x,y}[t]$	Number of agents at a site
$R_{x,y}[t]$	Resource density at a site
$\alpha_{h,i}$	Harvest coefficient of an agent
$\alpha_{m,i}$	Move coefficient of an agent
D	Diffusion coefficient
E_a	Energy expenditure on maintenance per time unit
E_{birth}	Minimal energy for procreation
E_h	Energy expenditure on harvest
E_m	Energy expenditure on move
H_{max}	Maximum harvest by agent at a site per time unit
k	Carrying capacity per site
n_0	Initial number of agents
r	Growth rate of resource
R_{unc}	Uncertainty in agent's observation of resource
γ	Mortality coefficient
η	Efficiency of converting resource to internal energy
κ	Procreation coefficient
ξ	Variation in offspring characteristics

II. Test case description

The test case ABM considers agents that harvest a resource in a spatially explicit environment. The ABM contains two types of entities: lattice sites and agents. The lattice sites form an $N_x \times N_y$ lattice that represents the spatial environment. Each site contains an abstracted resource density $R_{x,y,t}$.

Because agents have a basic energy expenditure, they need to harvest resource from sites to maintain their internal energy balance. Every time-step, agents decide to either harvest at their current location, move to a neighbouringing site (which costs energy), or stay (preserving energy with the risk of finding the resource taken by another agent). This decision depends on both the external environment and the internal state of the agent.

The ABM is implemented in NetLogo [6]. The main process overview for a single time-step is shown in Table I. All parameters and variables are given in Table II. A summarized description of the submodels in the process overview follows below.

1) Grow and diffuse resource: The resource grows logistically on each site and diffuses between sites following following Fick's second law [7],

$$\frac{dR(x,y,t)}{dt} = rR(x,y,t)\left(1 - \frac{R(x,y,t)}{k}\right) + D\nabla^2 R(x,y,t). \quad (1)$$

We use the exact solution for the logistic growth, avoiding numerical destabilization. The diffusion equation is solved using a forward Euler algorithm, which is stable for $D < \frac{1}{4}$.

2) Observe: Agents check their current site and the 4 Von Neumann neighbours for resource $R_{x,y}$, the number of agents nx, y, and the internal energy E_i of each agent. Agents do not have perfect information on the amount of resource or internal energy; there is a random difference between the observed amount and the true amount. The difference is distributed normally around zero with standard deviation R_{unc}.

3) Harvest?: Agents decide whether they will harvest this time-step. The agent is more likely to harvest if E_i is low (ie., the agent is 'hungy'), if $R_{x,y}$ is high, if $n_{x,y}$ is low (ie., there is no need to share) and if the resource on neighbouring sites is low. Agents also have an individual harvest coefficient α_h. Agents with a low value of α_h harvest more often. If an agent decides to harvest, an amount of resource up to a maximum harvest H_{max} is converted to internal energy with an efficiency η, .

4) Move?: If an agent does not decide to harvest, it may spend energy to move. Moving is likely in the direction of the highest expected harvest. That is, a move is made in the direction with the highest $R_{x,y}$ and the lowest number of agents. In addition, agents have an individual move coefficient α_m. Agents with a low value of α_m are more likely to move.

5) Pay maintenance: Agents pay a constant amount of energy for maintenance E_a each time-step.

6) Die?: Each agent has a probability to die each time-step. This probability is higher if the internal energy of the agent is low.

7) Breed?: Agents with a sufficient internal energy level have a probability of procreating by dividing their internal energy over two newly created 'daughter' agents. These inherit the characteristics of the 'parent' agent with minor deviations.

III. Results

A. Local sensitivity analysis

Local sensitivity analysis is a low-cost methodology that can expose tipping points and other strong nonlinearities. Local analysis starts from a nominal parameter set. From this set, each parameter, including the initial number of agents n_0, was varied individually over a large range. As output variables, we chose n_{tot}, and the average values of α_h and α_m, taken over all agents. Many runs showed periodic solutions (see Fig. 1). For simplicity, we averaged over the second half of each run and did not consider the periodicity. Each run lasted 10^3 time-steps. A number of runs of 10^5 time-steps showed this to be

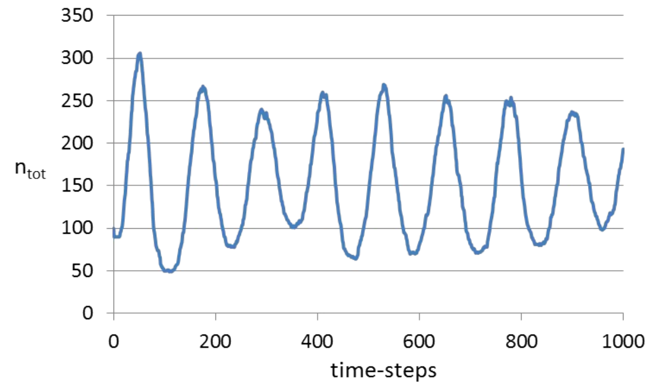

Fig. 1: Time-series of the total number of agents in the nominal parameter set.

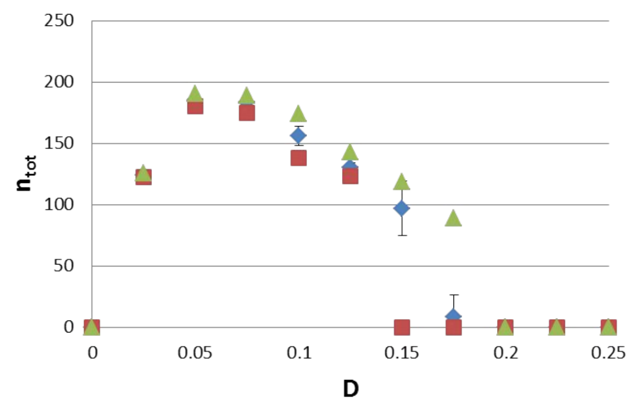

Fig. 2: The number of agents n_{tot} for different values of D. The reported values are the averages over the second half of each run, and over 10 replication runs. The mimimum and maximum among replicates are shown in red squares and green triangles respectively. The error bars show the 95% confidence interval.

sufficient to capture the long-term model behaviour. Each run was repeated 10 times to account for stochastic variations.

The results for the diffusivity D are shown in Fig. 2. If there is no diffusion ($D = 0$), the number of agents goes to zero. This is to be expected, since sites cannot gain resource after having been emtied completely. The agents thus empty the sites one by one, until the system collapses. As the diffusivity is increased, the number of agents increases, up to an optimal value of D where the number of agents is maximal. As D increases further, n_0 decreases and eventually goes to zero (depicted in Fig. 2 as red squares). Visual inspection of simulations reveals that this results from an increase in the amplitude of the oscillations in the number of agents. For $D = 0.15$ and $D = 0.175$ some of the runs went to zero as the mimimum of the oscillations reached $n_{tot} = 0$. For higher values of D, this happened in all of the runs.

Figs. 3 and 4 show the effect of the diffusivity on α_h and α_m. A low value of α_h or α_m means that agents harvest or

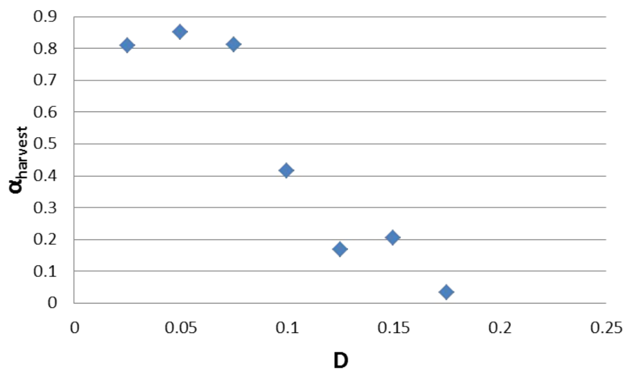

Fig. 3: OAT sensitivity analysis of D on α_h. The reported values are the averages over the second half of each run, and over 10 replication runs.

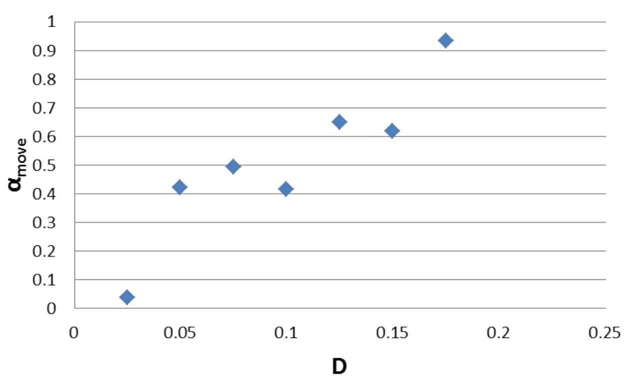

Fig. 4: OAT sensitivity analysis of D on α_m. The reported values are the averages over the second half of each run, and over 10 replication runs.

move more often respectively. For low values of D agents move more often, trying to find sites that are rich in resource. For high values of D the resource spreads out more evenly, so that agents tend to harvest at their present location more often. Further one-at-a-time analyses for all other parameters revealed that η, k, E_h, E_a, E_{max} and θ_{die} all have tipping points where n_0 goes to zero, i.e. this is a robust property of the ABM.

B. Global sensitivity analysis

After revealing nonlinearities and tipping points by using a local method, a global variance-based analysis was used to sample from the full parameter space and to obtain information about interactions between parameters. For a variance-based method one assumes probability distributions to account for uncertainty in the input parameters. The method then decomposes the variance of the output in terms that are attributed to these input uncertainties. We chose uniform input distributions over a large range, to explore the model behaviour over a large part of the parameter space. The analysis was based on an independent draw of 1000 parameter sets from these input distributions. Inspection of the histograms showed that the input sample sufficiently covered the uncertainty range of each parameter. Furthermore, there were no significant correlations between the input parameters in the sample. As was done in the local analysis, all outputs were averaged over the second half of the simulation. For each sample point, 10 replicates were generated to account for stochastic effects. Table III shows the mean of the variance among replicates, expressed as a percentage of the variance over all model runs. Stochastic effects did not cause a large amount of variation in n_{tot}, but α_h and α_m showed larger variations between replicates.

Following the method employed earlier by [9], we fitted a regression model on the averaged output of the 10 replicates. The resulting polynomial was then used to attribute output variance to parameter uncertainties. A linear regression of all the parameters and initial conditions explained 49.6 % of the output variance in total. The first order sensitivity of a parameter is the percentage of the output variance that is accounted for by that parameter, excluding interactions with other parameters. These are printed for all parameters in table IV. The total order sensitivity of a parameter is the percentage of the output variance accounted for by that parameter and its interaction with other parameters. For our test case, the differences between the first and total order sensitivities were found to be very small. This indicates that interactions between parameters did not account for a significant part of the variance.

To also take into account nonlinearities, we computed the sensitivities based on a regression model with third order splines for all parameters. The difference between the linear terms and terms with splines is small, indicating that nonlinearities up to that order also do not account for a large percentage of the variance. The best fit was obtained by taking into account all possible interactions up to second order and third order nonlinearities for all parameters. This fit explained 64.0 % of variance.

Fig. 2 suggests that nonlinearities account for some of the unexplained variance. The number of agents collapses rapidly as the diffusivity increases, which is not easily captured using a regression based method. The one-at-a-time sensitivity analysis shows similar tipping points for other parameters. In the results for the global sensitivity analysis, the number of agents went to zero in about half of the runs, again suggesting the presence of tipping points. As a test, regression based sensitvity analysis was also applied to the remainder of the runs, omitting all the runs that went to zero. This increased the amount of explained variance to 76.0 %. However, omitting these runs introduces correlations between the input parameters, which makes it impossible to fully separate the contributions of those parameters to the output.

TABLE III: Mean variance among replications, as percentage of variance among all runs

Output variable	% of variance
n_{tot}	0.4
α_h	6.0
α_m	13.5

TABLE IV: First order sensitivities, ranked from most to least influential on the model output n_{tot}. The total % of explained variance in n_{tot} is 49.6%

Parameter	Uncertainty range	% of variance n_{tot}
η	0 - 1	18.4
r	0 - 0.5	11.5
k	1 - 5	8.0
M	0 - 0.4	3.8
γ	0 - 5	3.0
$E_{harvest}$	0 - 0.5	2.2
E_{move}	0 - 2	1.4
E_{birth}	1.0 - 10	1.3
D	0 - 0.25	1.2
E_{max}	0.2 - 2	0.2
N_0	0 - 1000	0.0
R_0	0 - 1	0.0
R_{unc}	0 - 1	0.0
ξ	0 - 0.5	0.0

IV. Discussion

We have been able to explain 64.0 % of the output variance by regression methods, including third order nonlinearities and interaction effects. This leaves a significant part of the variance unexplained. Table III shows that stochastic effects do not explain this variance. Nonlinearities may be another source of unexplained variance. Nonlinearities up to third order were considered, but this does not account for a result like in Fig. 2, where the output variable abruptly goes to zero as the diffusivity is increased. Similar tipping points were found for other parameters.

An important aim of sensitivity analysis is to better understand model behaviour. Our test case shows that one-at-a-time sensitivity analysis helps to achieve this aim by uncovering tipping points and strong nonlinearities. We therefore recommend to use one-at-a-time sensitivity analysis in addition to global sensitivity analysis. While global sensitivity analysis does not allow for the straightforward identification of tipping points, it does provide summary measures for the sensitivities of (interactions between) parameters over a large region of the parameter space. For complex relations between inputs and outputs, such as in our test-case, regression-based sensitivity analysis does not yield accurate estimations for these sensitivities. “Model-free” methods of sensitivity analysis, like the Sobol method [2] [8], may aid in this case. In contrast with regression-based methods, model-free methods do not a priori assume any specific form for the relation between the input parameters and output variables.

A. Future work

The analysis so far has used outputs that were averaged over time. However, the sensitivities can vary strongly in time. To understand the dynamics of an ABM, it is relevant to compute the sensitivities as a function of time [10]. In our test case model, the existence of oscillations in the output variables (see Fig.1) makes the computation of time-dependent sensitivities particularly relevant.

Furthermore, the outputs were also averaged over space. Since the model features local interactions it is relevant to consider spatial correlations. Due to the difficulties in dealing with spatial correlations, this is not a common approach [11], but may be worthwhile for a simple test model.

We plan to perform a model-free method of global sensitivity analysis. This should yield accurate estimations for the global sensitivities, even in the presence of tipping points such as the one in Fig. 2. We also plan to examine such tipping points in more detail. For example, we will zoom in around the tipping point in the region of $D = [0.15, 0.2]$ using a smaller step-size in D and more replicates to unravel the exact dynamics around that point.

References

[1] A. Crooks, C. Castle, and M. Batty, “Key challenges in agent-based modelling for geo-spatial simulation,” Comp. Environ. Urban Syst., vol. 32, pp. 417–430, 2008.
[2] A. Saltelli, S. Tarantola, F. Campolongo, M. Ratto, M., Sensitivity Analysis in Practice. A Guide to Assessing Scientific Models, John Wiley & Sons, 2004.
[3] I. Lorscheid, B.-O. Heine, and M. Meyer, “Opening the ‘black box’ of simulations: increased transparency and effective communication through the sytematic design of experiments,” Comput. math. Organ. Theory, vol. 18, pp. 22-62, 2012.
[4] G. A. K. van Voorn, G. A. Ten Broeke, and A. Ligtenberg, “Concepts and methods for sensitivity analysis of agent-based models”, *Proc. ESSA 2013*, Warsaw, Poland, Sept. 16–20, 2013.
[5] G. A. Ten Broeke, S. Libre, and G. A. K. van Voorn, “Using an agent-based model to study the effects of enforced quota for sustainable tuna fishery’,’ in *Proc. ESSA2013*, Warsaw, Poland, Sept. 16–20, 2013.
[6] U. Wilensky. Netlogo. http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern University. Evanston, IL.
[7] A. Fick, “Ueber diffusion”, Annalen der Physik, vol. 170, pp. 59-86.
[8] J. Cariboni, D. Gatelli, R. Liska, and A. Saltelli, “The role of sensitivity analysis in ecological modelling,” Ecol. Mod., vol. 203, pp. 167–182, 2007.
[9] Burgers, Saskia LGE, et al. 2010. ”Sensitivity analysis of an agent-based model of cultures consequences for trade.” Progress in Artificial Economics. Springer Berlin Heidelberg, 253-264, 2010.
[10] A. Ligmann-Zielinska and L. Sun, ”Applying time-dependent variance-based global sensitivity analysis to represent the dynamics of an agent-based model of land use change”. Internat. J. Geograp. Inf. Sci., vol. 24, pp. 1829-1850, 2010.
[11] L. Lilburne, and S. Tarantola, ”Sensitivity analysis of spatial models.” Internat. J. Geograph. Inf. Sci., vol. 23, pp. 151-168, 2009.