

VERIFICATION THROUGH CALIBRATION: AN APPROACH AND A CASE STUDY OF A MODEL OF CONFLICT IN SYRIA

Maciej M. Łatek, Seyed M. Mussavi Rizi and Armando Geller

Scensei GmbH
Zürich, SWITZERLAND

ABSTRACT

In this paper we introduce a workflow for multiagent modeling that relies on piecemeal calibration to verify the model, and discuss how modelers can organize this workflow to accelerate model building, improve the quality and technical soundness of the final model and be able to attribute dynamics of model outputs to causal mechanisms represented in the model. To this end, we apply the proposed workflow step by step to the development process of a multiagent model of the civil war in Syria, and visualize model validity and dynamics across individual development sprints.

1 INTRODUCTION

The predominant narrative of building multiagent models depicts modeling as a linear process in which modelers develop *the* model, verify that the model code does what it is supposed to do, calibrate the model parameters to minimize divergence between model outputs and a training dataset, and validate model outputs with another dataset (Edmonds and Moss 2005, Railsback and Grimm 2012). In reality, building multiagent models is a hectic, iterative process in which modelers tinker with model components, groping for economical representations and resolutions of the biophysics of the model world, and formulations of agent behaviors that are faithful to empirical data. Their efforts evolve a series of models into *a* model developed under budget constraints and time pressure. They then examine the model code for bugs, calibrate model parameters and empirically validate the calibrated model. In this paper we propose an alternative modeling workflow that uses incremental calibration of evolving models to (a) formalize the search process over model components; (b) verify the model code, and (c) identify the model components that *persistently* drive changes in model outputs. We demonstrate the usefulness of the proposed workflow by applying it to the outputs of a multiagent model of the current civil war in Syria.

The rest of the paper is organized as follows: We outline the general setup of the proposed workflow and discuss its mathematical properties in Section 2; describe the model of civil war that we will use as a case study and show an implementation of the proposed workflow on the model in Section 3, and discuss the results of applying the workflow to the model in terms of model calibration and empirical validation in Section 4.

2 SETUP

2.1 Model development process

Imagine a multiagent modeling project. Suppose some time into the project t_1 , modelers come to believe that they have built a model M_1 with K_1 parameters $s_1 = [s_1 \ s_2 \ \dots \ s_{K_1}]$ that can produce some of the outputs r_1 they wish to explain or match with corresponding real values of R_1 . For example, six months into a project on building a multiagent model of housing markets, the modelers have built a model that generates rents and house prices, but not the spatial distribution of property values in a city. Assuming they

have verified the model code, at this point the modelers can either add more components to M_1 or calibrate it. Intuition dictates the former; we advocate the latter, as follows:

1. Use some score function to minimize the divergence between R_1 and r_1 . Let S_1^* denote the $N_1 \times K_1$ matrix of solutions to the minimization problem, where N_1 is the number of solutions. For example, if M_1 has $K_1 = 4$ parameters, the minimization procedure may produce $N_1 = 3$ solutions, all of which minimize the divergence between R_1 and r_1 . Each row $s_1^*, s_2^*, \dots, s_{K_1}^*$ in S_1^* denotes one solution to the minimization problem.
2. Perturb the parameter values in each row of S_1^* . For simplicity, we split some range around the parameter values into equal intervals. Continuing with the example, if the first solution is $s_1^* = [2 \ 4 \ 7 \ 1]$, we will construct intervals $[0, 4], [0, 8], [0, 14]$ and $[0, 2]$ around 2, 4, 7 and 1 respectively, and split each interval into say 10 values. We then run M_1 for all possible combinations of parameters a number of times and record the mean value of M_1 outputs.

We can now inspect the response surfaces of outputs for some local neighborhoods of the minima. These response surfaces may deviate unexpectedly from the minima either in magnitude or direction. For example, in the housing model, *ceteris paribus*, a parameter is thought to have a positive impact on house prices. The calibration sweep in the second step above may reveal that at some range the model is not responsive to changes in the value of the parameter or that house prices decrease with increasing parameter value at some value of the parameter. These effects may be due to bugs in the code or threshold effects in the parameter space. Assuming the modelers reverify the code, they have now gained some knowledge into the behavior of the parameter space for M_1 .

At some time t_2 , the modelers will come up with another model M_2 . In general, the transition from M_1 to M_2 is discontinuous. M_2 may not in general share any components with M_1 . However, for simplicity, let's assume some form of continuity in the model space where any subsequent model can inherit business logic, output measures and *calibrated parameters* from preceding models. For example, in M_1 households that are deciding to buy a house may use parameters a and b in their decision making and come up with an initial price for a specific house. In M_2 households may use the calibrated values of a and b to form an initial price, but change the business logic of decision making or keep the same parameters and business logic, but change the output measure from initial price to the final price. They may also add another parameter c to the business logic.

Once the modelers repeat the same procedure outlined above with M_2, R_2, r_2 , and S_2^* ; they face three questions:

- How does the fit between r_1 and R_1 change if S_2^* values of the common parameters between S_1^* and S_2^* are plugged into M_1 ? The answer to this question determines whether the transition from M_1 to M_2 contributes to the empirical validity of the model.
- How different are the values of common parameters between S_1^* and S_2^* ? The answer to this question sheds some light on the distribution of optimum values in the parameter space.
- How different are the threshold behaviors of S_1^* and S_2^* ? The answer to this question sheds some light on the shape of the parameter space.

In the next section we will describe an implementation of the procedure outlined above to a multiagent model of civil war in Syria.

2.2 A brief description of the model

The model of civil war in Syria integrates data layers such as logistic and infrastructure networks, synthesized population with ethnic and religious attributes, information on militias and political alliances that are supported with weapons, money and training by external actors such as the U.S., UK, France, Russia,

Turkey, Saudi Arabia, Qatar, Iran and the like. Neighboring countries can also decide to close their borders with Syria.

Households decide whether to flee violence or stay put and what routes to pick if they decide to flee. They also form grievances and perceive insecurity as a function of violent events around them. Military age males can get recruited by militias. Political alliances distribute resources to and set strategic goals for militias they represent. Militias distribute resources to their fighter units and set tactical goals. Fighter units wage battles, form grievances, loot and can commit mass atrocities. In this article, we will show the results of the workflow in Section 2 applied to household behaviors: Grievance formation and violence news process; fleeing or staying; choosing destinations, and planning routes.

3 HUMANITARIAN OUTPUTS OF A MULTIAGENT MODEL OF A CIVIL WAR

The model underwent six major development phases summarized in Table 3. This section describes the final logic of the simulation.

3.1 Behavior 1: Deciding to flee or stay

Each week a household evaluates if it should stay or flee. This decision is based on grievances and perceptions of violence. Table 1 gives the exact algorithm.

Table 1: Behavior 1: Deciding to flee or stay put by households.

Actor	A household
Decision making	<ol style="list-style-type: none"> 1. Read violence and total grievance toward self dominant identity. 2. If the sum of violence and total grievance exceeds the flee threshold δ, but is less than the stuck threshold ε, activate the route planning behavior. 3. Otherwise, maintain the household status as is and wait for a week.
Triggers and links	A household activates route planning if it decides to flee.
Parameters	δ : Household-specific real number in $(0, 1)$. Default average value: 0.05. ε : A real number in $(0, 1)$ greater than δ . Average default value: 0.2.

3.2 Behavior 2: Grievance formation and violence news process

Each armed engagement, revenge killing, looting or ethnic cleansing event generates a news item. The events can be either directly observed by households or spread on the logistics network. The number of nodes the news of an event travels depends on the its magnitude, that is, the number of victims. Households sharing a common location and dominant identity process their own observations and the news they receive every week. See Table 2 for details.

Table 2: Behavior 2: Grievance calculations and violence news process.

Actor	Households sharing a dominant identity in neighborhood l
Decision making	<ol style="list-style-type: none"> 1. Processing violence events. For each fatality, increase perception of local violence $l.v$ by $l.v/l.p$ where $l.p$ is the neighborhood population. 2. Spreading perceptions of violence. For each neighbor n of l within the radius of $r = c^\gamma$ $n.v_{t+1} = (1 - \alpha) \times n.v_t + \alpha \times l.v_t$ <p>where c denotes the number of fatalities in a given kinetic event, γ the news magnitude exponent and α the violence spread coefficient.</p> 3. Updating grievance. For each fatality increase own grievance toward the perpetrator, if the victim belongs to own dominant identity, and increase victim's dominant identity grievance toward own dominant identity, if the perpetrator belongs to own dominant identity, by $1/l.p$. 4. Dissipating perceptions of violence. Set $l.v_{t+1} = (1 - \beta) \times l.v_t$ <p>where β is a discount rate.</p> 5. Applying the above formula to dissipating grievances, making sure that grievances and perceptions of violence fall into $(0, 1)$.
Triggers and links	Grievance and perceptions of violence influence a household's decisions to stay or flee, choose routes and join militias, and a fighter unit's decision to engage in revenge killings and ethnic cleansing.
Parameters	<p>β: Rate at which memories of grievance and perceptions of violence dissipate. Default mean value: 0.025, corresponding to a memory length of roughly 35 weeks.</p> <p>α: Rate at which perceptions of violence diffuse among neighborhoods. Default value: 0.25 per week. Given the mean number of neighbors on the logistics network, diffusion is less intense in rural areas than in dense urban neighborhoods.</p> <p>γ: Determines how far news of violence travels through the logistics network as a function of the number of fatalities. Default value: 0.5. The whole population will learn about a battle in a central location with more than 250 fatalities in a week.</p>

Table 3: Model variants created during the development process. Changes in model logic, new parameters and new output are with respect to the preceding variant of the model.

Model variant	Change in logic	New parameters	New outputs
1a Migration only due to local violence to refugee camps		$\alpha, \beta, \delta, \varepsilon, \tau, \nu, \pi, \sigma, \mu$	People deciding to move, their destinations and routes
1b Household socioeconomic status and cities in other countries	Dominant identity determines the cost of sustaining a household in neighboring country cities. Behavior 3 has changed. Households consider more potential destinations and account for the monetary cost of staying at all destinations.	θ	Refugees outside camps
1c Internal destinations	Population migrations before 2011 added to population synthesis. Behavior 3 has changed: Households consider one more destination in every decision cycle.		IDPs
2 Grievance routing	Grievance calculation and attribution of events added to Behavior 1. β affects grievance dissemination.	ρ	
3 News process	Event magnitude determines how many times event news disseminates. Behavior 1 has changed to add news propagation.	γ	
4 Pre-computed routeset evaluation	Pre-computed routesets from internal locations to cities and refugee camps outside Syria. Parallelized routing.		

3.3 Behaviors 3 and 4: Destination choice, route planning and moving

Tables 4 and 5 provide outlines of destination choice, route planning and moving by distressed households.

4 RESULTS

4.1 Model outputs and validation data

We exposed the model households to historical data streams on violence, killing off household members as indicated by a casualty dataset we compiled from open sources such as Syria Tracker and Syrian Revolution Martyr Database. The dataset contains information on 69854 civilian deaths in 23584 events during January 1, 2011 – May 30, 2013. Each event is geolocated and timed, and has information about the total number of adults and children killed.

We distinguish between refugees who end up in refugee camps and those who stay with relatives or rent a dwelling outside Syria. For IDPs, we report the number of times violence forces a member of a household to move, and whether people who intended to reach locations outside Syria “get stuck” and turn into IDPs instead. To prepare validation data we harvested information from the UN Office for the Coordination of Humanitarian Affairs, UN High Commissioner for Refugees and Humanitarian Information Unit of the United States Department of State, and collected data on past, present and planned refugee camps, and refugee flows. Reconciling, imputing and averaging data from these sources, we estimated displacement patterns of the Syrian population since the outbreak of the crisis. These data sources produce data of varying temporal and spatial resolution and use different labels for refugees, for example, registered, in camp and total, and IDPs, for example “people in need of assistance”.

To match real-life refugee figures for every quarter q of 2012 and the first quarter of 2013, that is, $Q = \{q_{2012}^1, q_{2012}^2, \dots, q_{2013}^1\}$, with model outputs, we calibrated model parameters that control household behaviors, news processing and grievance formation by minimizing the mean absolute relative deviation (MARD) of simulated refugees counts sX and real-life refugee counts ${}^\ell X$ for each external refugee destination $C = \{\text{Jordan, Iraq, Lebanon, Turkey}\}$, and the numbers of IDPs between simulation sY and in real life ${}^\ell Y$:

$$\text{MARD} \left({}^sX_{q,c}, {}^\ell X_{q,c}, {}^sY_q, {}^\ell Y_q \right) = \frac{\sum_{q \in Q} \left(\sum_{c \in C} \frac{|{}^sX_{q,c} - {}^\ell X_{q,c}|}{{}^sX_{q,c} + {}^\ell X_{q,c}} \right) + \frac{|{}^sY_q - {}^\ell Y_q|}{{}^sY_q + {}^\ell Y_q}}{\|C\| \times \|Q\|}.$$

4.2 Regression tests

After incorporating a new data layer into the model or implementing a new household behavior, we ran parameter sweeps that tie the changes in fatalities, refugees and IDPs as response variables to changes in the combinations of behavioral and technical parameters. The model underwent 18 such regression tests, with each test exploring 56–121 parameter sets over 3 or more independent simulation runs.

Figure 1 presents some response surfaces across six version of the simulation and provides the following insights:

- γ has no meaning in the model until version **3**. This gives rise to vertical contours.
- Up until version **2**, the mean value of MARD changes dramatically whenever a new behavior or data layer is added to the model. Transitions to versions **3** and **4** see somewhat smaller improvements.
- Surprisingly, constraining route optimization in version **4** significantly improves MARD.
- Grievances do not matter for empirical validity unless there is also a global news process.
- There are potentially two optimal values of γ in version **4**.

Table 4: Behavior 3: Destination choice and route planning.

Actor	A household
Decision making	<ol style="list-style-type: none"> 1. Create a list of possible destinations: <ol style="list-style-type: none"> (a) Add all possible internal destinations where relatives live. (b) Add under capacity refugee camps to which borders are open. (c) Add external destinations other than refugee camps for which self wealth score is greater than $m\theta$ where m is the monetary cost of staying in the destination and θ is the staying cost multiplier. (d) If the list of possible destinations is empty, force-add all refugee camps as destinations. This may result in overcrowded refugee camps. 2. Find a destination that minimizes the weighted sum of risk-adjusted costs of traveling and staying in the destination: <ol style="list-style-type: none"> (a) Risk-adjusted cost of traveling is the sum over route segments of four components: <ol style="list-style-type: none"> i. Random bias with household routing bias ν as weight. ii. Travel cost, as the distance to the destination times household routing length multiplier μ. iii. Route risk, derived from sum of mean perceptions of violence at the start and end points of each road segment weighted by household routing risk multiplier π. iv. Route grievance, derived from the sum of mean grievances held by the populations at the start and end points of each road segment toward the traveling household weighted by household routing grievance multiplier ρ. (b) Risk-adjusted cost of staying at the destination weighted by destination conditions multiplier σ. Risk-adjusted cost of staying at refugee camps is 0. The cost of staying at a location inside Syria is equal to the perception of violence at the location. 3. Memorize the planned route.
Triggers and links	Household moving is activated.
Parameters	θ : Monetary costs of staying at a destination other than refugee camps; varies by location and dominant identity. Default value: 5.
	σ : Weights security conditions at the destination. By default, external destinations are secure. Default value: 5.
	ν : Stochastic factor as a component of a household's traveling cost multiplied by a uniform random variable in the unit interval. Default value: 2.
	μ : Penalty for distance traveled. Default value: 0.02 per km.
	π : Penalty for taking riskier routes. Default value: 2.
	ρ : Penalty for traveling through areas in which the population is hostile to the household dominant identity. Default value: 2.

Table 5: Behavior 4: Deciding to flee or stay put.

Actor	A household
Decision making	<ol style="list-style-type: none"> 1. If perceptions of local violence is greater than stuck threshold ε, stop traveling and cancel the current route. 2. Otherwise, move to the next location on the current route at speed τ.
Triggers and links	<ol style="list-style-type: none"> 1. When a household arrives at an external destination, its behaviors are deactivated. 2. If a household arrives to or is stopped at an internal destination, its status is reset to <i>OK</i>, and move or stay behavior (Behavior 2) is activated.
Parameters	τ : Base household speed.

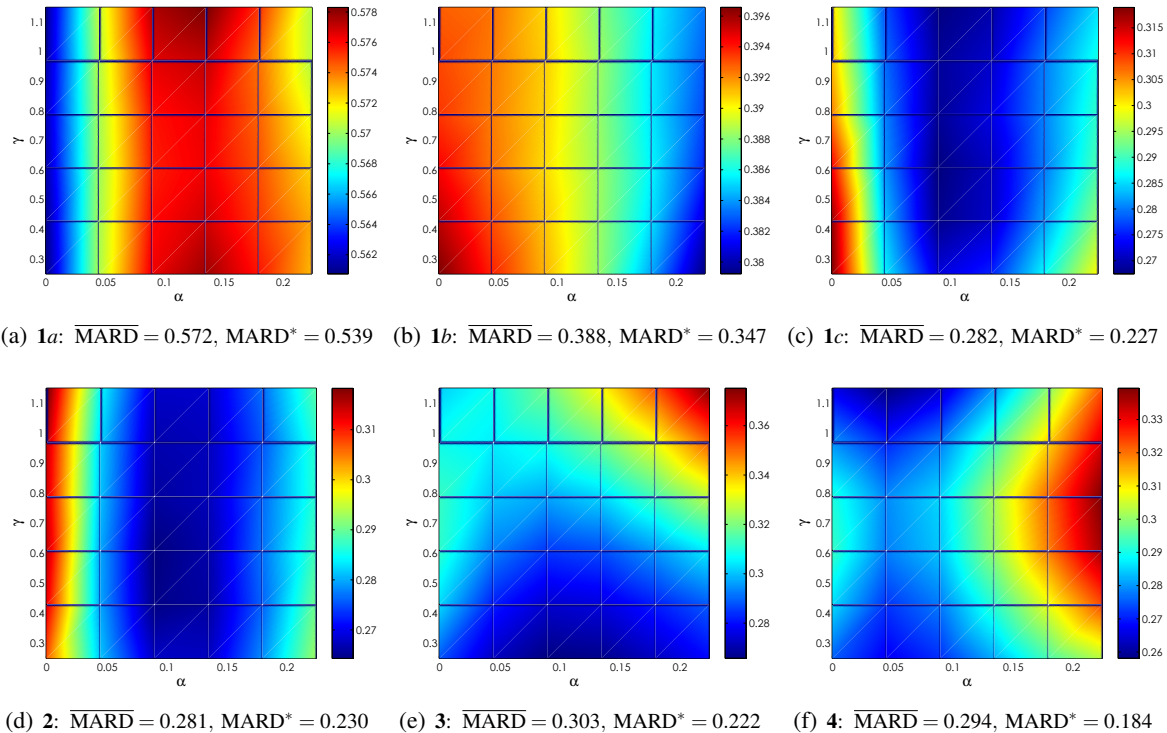


Figure 1: Selected regression test across model versions with MARD as the response variable. This regression test sweeps the product space of parameters α and γ . Color scales on panels are independent from one another. $\overline{\text{MARD}}$ indicates the mean and MARD^* the minimum of MARD.

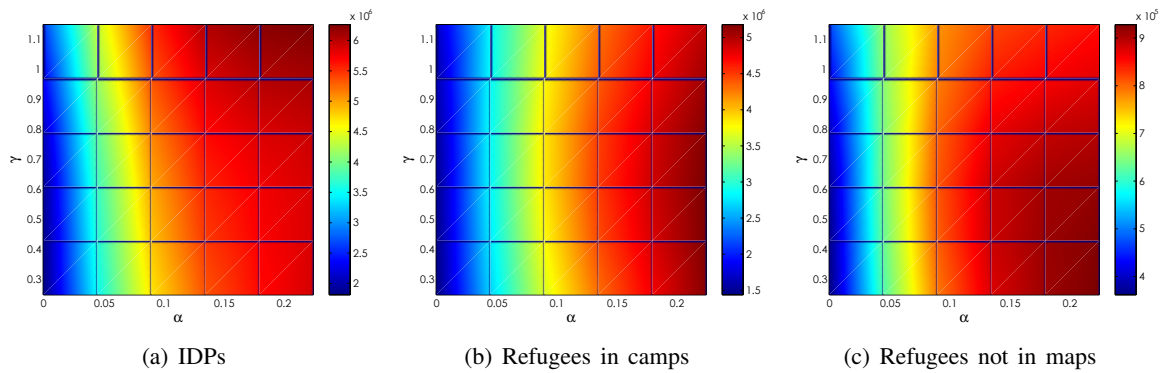


Figure 2: Humanitarian outputs for version 4, sweeping the product space of parameters α and γ .

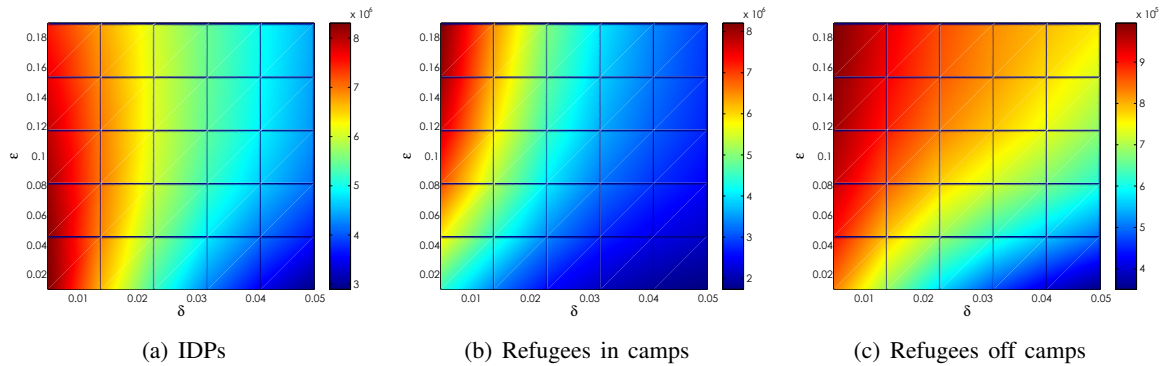


Figure 3: Humanitarian outputs for version 4, sweeping the product space of ϵ and δ .

Since other parameter values are kept at default settings, the best MARD on Figure 1 is not the best overall MARD, achieved when all parameters are calibrated simultaneously. We investigate this issue in the next section.

Figures 2 and 3 show refugees in camps, refugees off camps and IDPs as outputs of two parameter sweeps of version 4 of the model.

4.3 Calibration and validity

For the components of the models where real-life data on response variables are available, we use an automated calibration process to search for parameter sets that minimize MARD. This section describes how 12 parameters controlling the dynamics of news processes, grievance formation and household behaviors were set automatically by matching refugee flows generated by the model to historical refugee flows, when households were exposed to historical violence.

We used “Evolution strategies”, an optimization technique based on ideas of adaptation and evolution as an optimization algorithm (Beyer and Schwefel 2002). The population was composed of 100 individual solutions, each a real vector of 8 dimensions for version 1a to 12 dimensions for versions 3 and above. The fitness of a solution was derived by the average MARD across at least 3 independent simulation runs. No improvement over the last 5 generations set off termination.

Dashboard 6 presents selected diagnostics for the calibration process. Since the dimensionality of the solution space increased by every subsequent version of the model, the optimization algorithm required more generations before reaching the termination condition. Similarly, whenever we added new behaviors to the

agents, the computational complexity of agents and their interactions increased, which in turn increased the wall time necessary to perform a 128-week long simulation and evaluate the associated MARD. However, in version 4, model re-engineering and limiting households' escape routes to a library of routes instead of allowing them to search the logistics network globally decreased the computational time 10 fold. Even though the calibration of version 4 takes more evaluations than other versions of the model, this makes experimentation with this version significantly easier than other versions.

Table 6: Selected optimization statistics for the calibration process. Legend: GTC: Number of generations to convergence, MR: Mean runtime (wall clock) of a single instance of the simulation, MARD*: Global minimum of MARD, and μ^* through γ^* : Values of parameters for MARD*. * indicates not applicable.

Version	GTC	MR	MARD*	Optimal value of parameters									
				μ^*	β^*	π^*	ρ^*	θ^*	α^*	δ^*	σ^*	ϵ^*	γ^*
1a	58	1.09	0.399	4.1	0.004	0.21	*	*	0.589	0.022	2.8	0.005	*
1b	65	1.10	0.302	0.2	0.021	5.71	*	1.2	0.152	0.022	4.6	0.032	*
1c	66	2.05	0.215	0.2	0.000	8.48	*	5.5	0.048	0.028	4.3	0.178	*
2	72	1.95	0.229	0.4	0.027	2.14	4.5	5.7	0.784	0.012	5.6	0.011	*
3	106	1.81	0.210	4.1	0.001	6.97	1.5	5.4	0.062	0.031	5.5	0.289	1.748
4	127	0.22	0.126	0.2	0.005	9.51	9.8	4.5	0.071	0.032	9.9	0.022	1.775

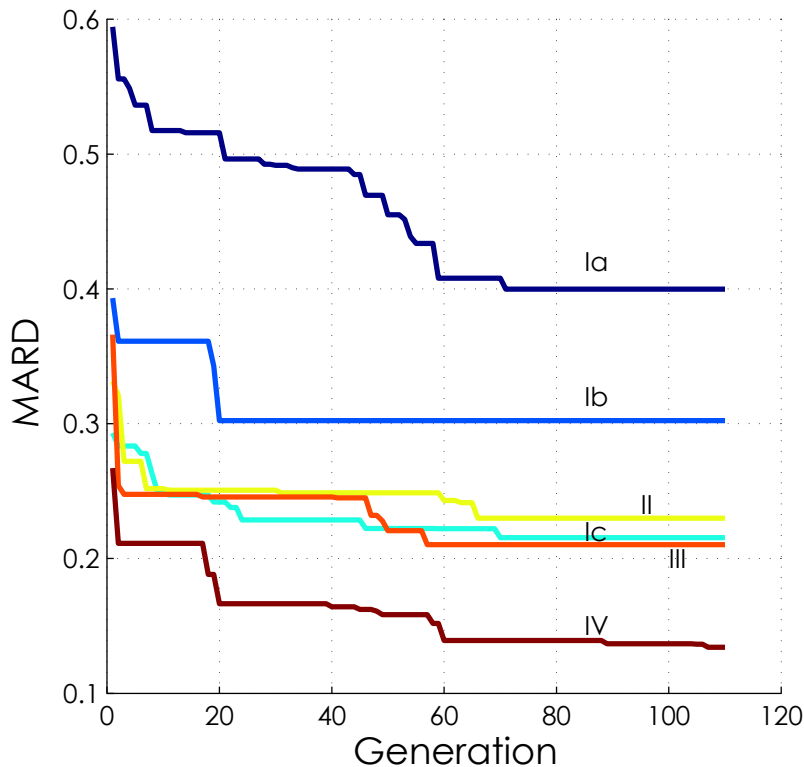


Figure 4: Dynamics of the optimization process: MARD of the best solution as a function of optimization generation.

5 CONCLUSIONS

In this paper we introduced a workflow for large-scale parameters sweeps and calibration experiments that can guide and verify the process of developing multiagent simulation models. Using a multiagent model of civil war in Syria as a case, we presented the steps in the verification and calibration workflow and visualized model validity and dynamics across individual development sprints.

We used the process outlined in this paper to build a tool housed at <http://www.syriafutures.ch> to understand the dynamics of the Syrian civil war. The tool does predict selected characteristics of the conflict, but it is not merely or primarily a prediction machine. Instead, it is designed to help observers explore what conditions may give rise to what kind of outcomes and why. The robustness of the tool relies on the systematic use of parameter sweeps to test and calibrate the multiagent model that powers the tool.

REFERENCES

- Beyer, H. G., and H. P. Schwefel. 2002. "Evolution Strategies: A Comprehensive Introduction". *Natural Computing* 1 (1): 3–52.
- Edmonds, B., and S. Moss. 2005. "From KISS to KIDS – An 'Anti-simplistic' Modelling Approach". In *Multi-Agent and Multi-Agent Based Simulation 2004*, edited by P. Davidsson, 130–144. Berlin: Springer.
- Humanitarian Information Unit, the United States Department of State 2013. "Syria Refugee Camps Dataset". <https://hiu.state.gov/data/data.aspx>. Last accessed 1 June 2013.
- Railsback, S. F., and V. Grimm. 2012. *Agent-Based and Individual-Based Modeling: A Practical Introduction*. Princeton University Press.
- Syria Tracker. <https://syriatracker.crowdmap.com>. Last accessed 1 June 2013.
- Syrian Revolution Martyr Database. <http://www.syrianshuhada.com>. Last accessed 1 June 2013.
- The United Nations High Commissioner for Refugees. "Syria Regional Refugee Response". <http://data.unhcr.org/syrianrefugees/regional.php>. Last accessed 1 July 2013.
- The United Nations Office for the Coordination of Humanitarian Affairs. "Humanitarian Response Syria". <http://syria.humanitarianresponse.info>. Last accessed 1 June 2013.
- Understanding the Syrian Civil War 2013. <http://www.syriafutures.ch>. Last accessed 1 June 2013.

AUTHOR BIOGRAPHIES

MACIEJ M. ŁATEK is a partner at Scensei and a Research Assistant Professor at George Mason University from where he also has a Ph.D. in computational social science. Dr. Latek is an operations researcher with an M.A. in quantitative methods and information systems from the Warsaw School of Economics. His dissertation focused on using simulation to help decision makers who face intelligent adversaries or competitors to gain insight into the most likely and effective strategies of their adversaries and to determine which of their own possible strategies should be used to respond. He has co-developed a multiagent model for counterterrorism applications and a country-scale model of the drug industry in Afghanistan. Dr. Latek has recently worked issues of border security and ways to influence low intensity conflicts and intra-state criminal organizations. His email address is maciej@scensei.ch.

SEYED MOHAMMAD MUSSAVI RIZI is a partner at Scensei and a Research Assistant Professor at George Mason University. He received a Ph.D. in computational social science from George Mason University, an M.A. in economics from Tufts University and a MALD in international relations from the Fletcher School. As a political economist and statistician he has expertise in data analytics and has worked extensively on algorithms for social networks, geospatial and remote sensing data fusion into simulations. Dr. Mussavi Rizi has co-developed a simulation of the Afghan rural economy, another to measure the effectiveness of development projects in low intensity conflicts and finally a simulation to plan for humanitarian crises of the current conflict in Syria. His email address is seyed@scensei.ch.

ARMANDO GELLER is a partner at Scensei. He earned a D.Phil. in political science from the University of Zurich; conducted postdoctoral studies at Manchester Metropolitan University and was a Research Assistant Professor at George Mason University where he is currently affiliated with the School for Conflict Analysis and Resolution. With over 10 years of fieldwork experience in Central Asia, Dr. Geller specializes in information elicitation in conflict areas and evidence-driven model design. His fieldwork included interviews on power relations among Afghan urban elites; and implementing role playing games with rural elites to determine development needs and estimating effective portfolios of development projects. His email address is armando@scensei.ch.