Towards a Multi-Model Approach to Agent-Based Social Simulation for Policymaking

Emil Johansson[1,2][0000−0002−2399−8817], Fabian Lorig[1,2][0000−0002−8209−0921], and Paul Davidsson[1,2][0000−0003−0998−6585]

[1] Department of Computer Science and Media Technology, Malmö University, Malmö, Sweden
[2] Sustainable Digitalisation Research Centre, Malmö University, Malmö, Sweden
{emil.johansson,fabian.lorig,paul.davidsson}@mau.se

Abstract. This paper explores a multi-model approach to agent-based social simulation; that is, using several independent model to simulate the same system and then aggregating results and insights from each of them. An example study is presented, where five different models are used to simulate disease spread in a town. This is followed by a discussion of some key questions when designing a multi-model study, and the feasibility of such a study. Though challenges exist, the multi-model approach has the potential to become a powerful tool against structural uncertainty in social simulation models.

Keywords: Multi-Model · Policymaking · Structural Uncertainty.

1 Introduction

The idea of a multi-model approach (as we define the term here) is simple: instead of performing a simulation study using a singular model, you use several different models and aggregate the results and insights you receive from each of them. The approach, though not new (see, e.g., [3,4,12,16]), is rarely seen in Agent-Based Social Simulation (ABSS) studies [8]. This could be considered surprising. One of the largest challenges for ABSS models to be appealing in real-world policymaking is the high structural uncertainty of models, often stemming from challenges surrounding modeling human behavior [1], and the quantification of this uncertainty [7]. The main purpose of multi-model studies, that is, exploring and to some degree counteracting structural uncertainty, should thus be appealing to the ABSS community.

One argument for multi-model studies is a wisdom-of-the-crowd type of reasoning: that several independent models pointing toward the same conclusions make these more reliable than had only one model been used to arrive at them. But this is not the only strength of an ABSS multi-model study. Were model outputs to differ between models, this could open for the exploration of how different modeling assumptions lead to different outcomes, leading to a better understanding about the simulated system itself or what models are more appropriate to study it. If differences between model results are large, this could at the very least serve as evidence to policymakers (and perhaps the modelers themselves) that the real-world system is currently not sufficiently well understood to draw the desired conclusions from simulations thereof.

The development of a multi-model approach for ABSS-models faces two main challenges. First, it needs to be *feasible*. A multi-model study will not be possible if project time and required resources scale linearly with the number of models used, especially if the study's purpose is aimed for real-world policy advice where timelines are often short. The other challenge is *justifiability.* Including a large number of models would often come at the cost of model complexity; it is necessary to understand when the advantages of having multiple simulation models of the same system would justify this loss of complexity and what is given up with it.

The aim of this paper is to explore the application of a multi-model approach to ABSS development for policymaking. This is done in two ways: through an example study where several different models of disease spread are used and their conclusions aggregated, and a discussion of key considerations when designing a multi-model ABSS study. Through this, steps are taken towards understanding how a multi-model approach can be designed to be a feasible option to a typical single-model study. Overall, the paper should not be considered an unconditional endorsement of multi-model ABSS studies, but a call for the further exploration of their full potential.

The remainder of the paper is structured as follows. Section 2 provides an example of a multi-model study, using models of epidemic spread. Section 3 discusses some key questions to be addressed when a multi-model study is designed, while section 4 briefly tackles the matter of how a multi-model approach can be made feasible. Section 5 concludes the paper.

2 The Case: A Multi-Model Study of Epidemic Spread

In this section, we provide an example of a multi-model study, using five simple models of epidemic spread. This is done in order to illustrate the multi-model approach, a potential process for how it could be executed, and what payoffs it could provide. Sections 3 and 4 will further discuss considerations and challenges regarding the process.

The example study is performed in the following steps (see the respective section below for more details):

1. The simulation study's aim and questions to be studied are defined.
2. A set of different models are implemented, verified and validated, each of them individually attempting to address the study's aim.
3. Independent experiments are run with the models, and an answer for each model - question pair is given.
4. General conclusions are drawn for the questions where most or all of the models agree (*model consensus*).

5. For questions where the individual models' conclusions differ, explore whether these differences are statistically significant or can be explained by parametrical uncertainty in the models.
6. When appropriate, aggregate the numerical outputs from the individual models to acquire new estimates, uncertainty measures or scenarios.
7. Next steps for the study are determined. This could include developing additional models to aid where results are non-conclusive, or identifying new questions of future studies based on the results.

2.1 The Scenario: Aim and Model Purpose

The scenario used for this example is the spread of an airborne disease in a town. Though no cases have yet been discovered in the town, policymakers are considering implementing population-wide testing and telling anyone who has tested positive to isolate themselves. The aim of the simulation study is to investigate the impact of this intervention on the spread of the disease. Specifically, the following policy-relevant questions are considered:

1. Would population-wide testing reduce the total number of infected individuals?
2. Would population-wide testing be able to mitigate the epidemic entirely?
3. If population-wide testing is implemented, how many individuals would at most be infected at the same time?

While this example scenario is fictional, plenty of ABSS models have been used for similar purposes to this in the past [13,9].

2.2 The models

Five models are created and implemented in NetLogo [15]. All of these are intended to be able to fulfill the aim of the study and provide answers to the questions above, but vary in what assumptions they make and what parts of the simulated system are included. First, the following two models were developed:

– **Model 1:** A Susceptible-Exposed-Infected-Recovered (SEIR) compartmental model on a network, extended from the Virus on a Network Model from the NetLogo library [11].
– **Model 3:** Agents are assigned a household and a workplace and visit both during each timestep. Uses an Susceptible-Infected-Recovered (SIR) compartmental model for disease transmission and progression.

These models were then extended in different ways to create new models:

– **Model 2:** Based on model 1, but splits the agent population into four different age groups with different infection probabilities.
– **Model 4:** Based on model 2, but includes the arrival and departure of agents from other towns, modeled on a network.

– **Model 5:** Based on model 1, but replaces the random risk of being infected by non-contacts with a model of infections within a grocery store.

This approach of creating models by extending each other represents a trade-off. While not having to implement every model from scratch is more time-efficient, it also reduces the diversity of assumptions represented in the model set. For this reason, we start with two models instead of one.

More in-depth explanations for the models and their source code are available at: https://github.com/emjolund/ABSS-combination-models (models 1 through 5 are here referred to as Network, WorkHome, AgeGroups, Integration and Modules respectively). We assume here that the models have all been verified and validated, and have individually been deemed appropriate for the aim of the study had they been the only model that was used.

2.3 Simulation Results

Next, experiments are run with each model. Every model simulates the spread of the disease with and without population-wide testing, and the models are run 1000 times each for each case. The model parameters used in these experiments can be found in Appendix A.

Table 1 shows the total number of infected individuals for the two cases for each model, as well as the largest number of infected individuals at any point in time. For the models where there exist an Exposed state for agents in addition to an Infected state, the reported maximum number is the sum of the number of agents with these two states. Each value is the average of 1000 runs, with standard deviation reported in parenthesis. Figures 1 and 2 show the number of infections over time for a single example run of each of the models, without and with population-wide testing respectively.

Based on these results, answers to the three questions of the study are determined for each model separately, as if they had been the only model used in the study. Table 2 list these conclusions.

Table 1. The total and maximum simultaneous numbers of infection in the five different models, with or without population-wide testing. Values averaged over 1000 runs, standard deviation in parentheses.

	Model 1	Model 2	Model 3	Model 4	Model 5
Total Inf. No Testing	643 (±57)	660 (±92)	792 (±3)	820 (±55)	709 (±15)
Total Inf. Testing	40 (±27)	6 (±2)	476 (±109)	613 (±52)	166 (±111)
Max Inf. No Testing	186 (±35)	92 (±21)	571 (±19)	530 (±29)	321 (±27)
Max Inf. Testing	16 (±7)	6 (±2)	154 (±47)	260 (±55)	36 (±19)

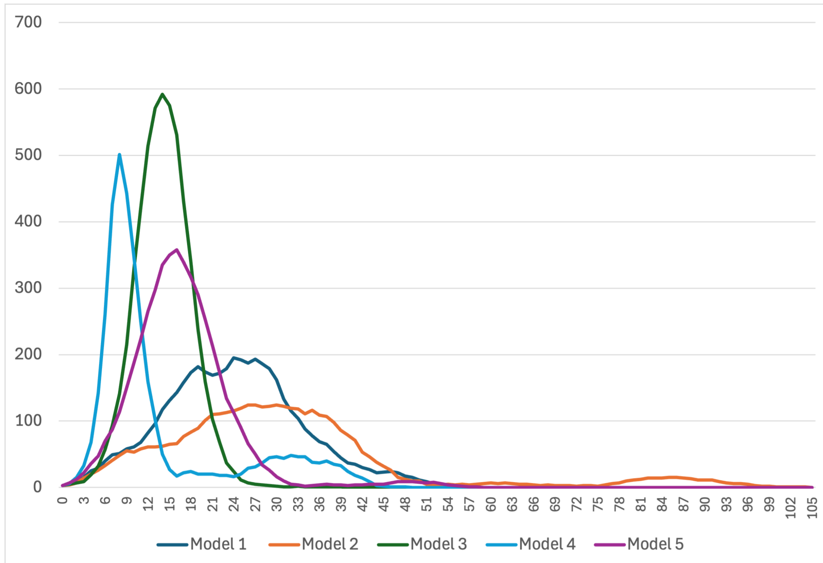

Fig. 1. Examples of a single run of each of the models in the scenario **without** population-wide testing.

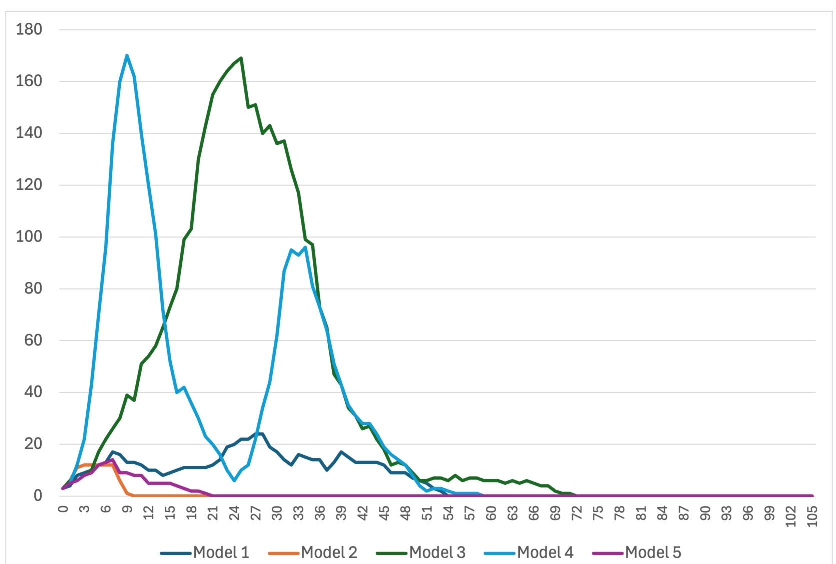

Fig. 2. Examples of a single run of each of the models in the scenario **with** population-wide testing.

Table 2. Conclusions drawn from each of the five models separately, with regards to the questions stated in the study's aim.

Question	Model 1	Model 2	Model 3	Model 4	Model 5
Testing reduces infections?	Yes	Yes	Yes	Yes	Yes
Testing prevents outbreak?	Yes	Yes	No	No	No
How many max infections with testing?	10-20	Almost none	100-200	200-300	20-60

2.4 Model Consensus

Looking at table 2, we see that all five models indicate that population-wide testing would decrease the total number of infections. This *consensus* between models thus could be considered to increase the confidence with which the conclusion can be drawn, compared to had only one model been used for this purpose.

That said, a consensus between models is no guarantee that the conclusion is correct. Since the number of models are still relatively small, with some of them sharing code and all of them sharing some basic assumptions of the system, it is still completely possible that all of them point at the wrong conclusions about the real-world system. Even with a much larger set of models, it is hard to avoid that shared "incorrect" assumptions, whether stemming from an incorrect understanding of the simulated system or from computational necessity, could potentially steer the consensus in the wrong direction. Thus, while it is apparent that a model consensus regarding a conclusion can increase the confidence with which that conclusion can be drawn, it is important that this confidence does not turn into overconfidence.

2.5 Diverging Conclusions

Question 2 of the case asks whether preemptive population-wide testing could prevent an outbreak of the disease entirely. Looking again at table 2, we see that the model results disagree here. Models 1 and 2 indicate that only a small percentage of the population would be infected in the scenario with population-wide testing, whereas models 3, 4 and 5 still see a large part of the population infected.

Understanding the reasons behind these differences can be tackled in many different ways. Here, we start by exploring whether the different results can be explained by parametrical uncertainty. While there exist rigorous statistical approaches to such a comparison (such as performing sensitivity analyses of all models and comparing their distributions), for the sake of this example we simply perform a manual recalibration. We focus on the infection probability, as it is likely the most difficult parameter to estimate and is used somewhat differently in the models. Changing the value of the infection probability in models 3, 4 and 5 so that number of total infections match those in models 1 and 2 in the no-testing case, we then observe whether the number of total infections in the population-wide-testing case match as well.

Table 3. The total number of infections in the five different models, with or without population-wide testing, after models 3, 4 and 5 have ben re-calibrated. Values averaged over 1000 runs, standard deviation in parentheses.

	Model 1	Model 2	Model 3	Model 4	Model 5
Total Inf. No Testing	643 (±57)	660 (±92)	670 (±50)	613 (±49)	617 (±30)
Total Inf. Testing	40 (±27)	6 (±2)	19 (±16)	301 (±46)	36 (±29)

Table 3 show the results from the recalibration of models 3, 4 and 5. After lowering the probability of disease spread between contacts, models 3 and 5 now give results similar to those of models 1 and 2. For model 4, however, no reasonable parametrization was found in which both scenarios matched the other models. As model 4 is the only model that includes agents entering and leaving the city, we could from this form the hypothesis that inter-city mobility reduces the effectiveness of the simulated testing policy (we call this a hypothesis instead of a conclusion of the study since there exist several other potential explanations to the differences in the models).

2.6 Numeric Estimations

The third question of the case asks for numeric estimations of the maximum number of infections at any given time in case population-wide testing is implemented. Generally, this type of prediction is not commonly performed using agent-based models and is sometimes advised against, given the structural uncertainty of the simulated system [5]. While a multi-model approach does not fully mitigate this issue, it could potentially put us in a better position to address this type of questions:

- Were the estimates of the different models to be very similar, it could greatly increase the confidence of the estimations. Such predictions should still be used cautiously, especially when working together with policymakers, but in policy situations where they are necessary and no better alternative sources of estimates exist they could still be of use.
- When model outputs are similar but not so similar that a singular estimate can be formed, they could instead form a range of or different cases for the variable in question. Again, caution needs to be taken that result presented in this way become misleading, since the real number may very well go outside of the "best" or "worst" case scenario that the study arrives at.
- When the model outputs are vastly different, this could be used as evidence to policymakers (and perhaps the modelers themselves) that the simulated

system is not sufficiently well understood for a simulation system to provide the requested estimates. As model uncertainty can be very difficult to communicate to policymakers without a statistical background, making this point when presenting only a singular, seemingly trustworthy, estimate can be more difficult.

In our case, the estimates of the maximum number of simultaneous infections (see table 2) fall into the third category. In this case, an uncertainty analysis of the different models might have been sufficient to conclude that the models are ill-suited to make these estimations, as indicated by the effects of the recalibration in the previous section. This might however not always be the case, since an uncertainty analysis cannot account for the structural uncertainty of the system; thus the multi-model approach could provide a better picture.

2.7 Conclusions

From the example study above, we would conclude the following:

- Population-wide testing would likely be able to reduce the number of infections of the studied disease.
- More precise estimates on the number of infections and whether the epidemic can be prevented entirely cannot be concluded from the study, as this is highly dependent on the infectiousness of the disease which is not currently sufficiently well-defined.
- There are signs that the impact of the population-wide testing intervention is affected by the level of mobility between the simulated town and outside areas, though this would have to be further examined.

3 Designing a Multi-Model Approach

While one step-by-step method for conducting a agent-based multi-model study was suggested above, there does not yet exist sufficiently many examples of these studies to develop a definite set of best-practices for how they should be performed. Instead, this section discusses a number of key questions that whould have to be addressed whenever a multi-model study is performed.

3.1 Homogeneous or Heterogeneous Models?

One could consider two different fashions in which the models in a multi-model study could be varied. The first is to keep models as similar to each other as possible, only differing with respect to a few assumptions regarding especially uncertain aspects of the model. Such models could then be used to systematically study the effects of key assumptions. Since these models would then share a large part of their code and interface, the added effort of developing one additional model could be relatively small. However, this approach only explores a small part of the uncertainty space, so while it let's us better understand the effects of some specific assumptions, it may greatly underestimate the structural uncertainty of the system as a whole.

On the other hand, one could develop models with the intention of them being as diverse as possible (while still staying within the space of "reasonable" models). Models could then be created and implemented by different members of the modeling team, in different programming languages and even using other paradigms than agent-based modeling. Such an approach would render models more difficult to compare to each other and would require significantly more work hours, but the models would then to a greater extent represent the full space of potential models.

Rather than fully committing to one of these options, both could serve a purpose in a multi-model approach. One could for instance start with a diverse set of models, then zoom in and slightly vary these once differences between model outputs have been identified. Conversely, a study could first study smaller variations of one model to understand a certain aspect of a system, then use more diverse models as a means of validating findings.

3.2 New or Reused Models?

The reuse of existing simulation models has been lifted as one of the current largest opportunities and challenges in simulation research [14]. Using existing implemented models in place of creating all models from scratch could in theory greatly reduce the amount of work needed for a multi-model study. This would also anchor the study in existing advancements, giving it a much more stable foundation.

However, model reuse in itself can also be time-consuming and costly [2], especially for models that were not documented with reusability in mind. Importantly, while included models can look very different on both a conceptual and implementation level, the intended purpose and use of the models need to match, something that is difficult to achieve for a model created by a different team and might not even be obvious if it is the case. Including models without understanding their original intended purpose or limitations can be directly harmful for the study [6]. Of course, reusing models instead of implementing new ones also assumes that there exist appropriate models available in the first place, which will not always be the case.

3.3 Are Models Weighted Equally?

In the example above, the outputs of all five models were considered equal contributions to the final conclusions. However, the study could also put more weight on models within the set that are considered more credible, for instance based on the validation process of each model or if any of the models are already well-established in their field. How to perform this weighting is not straight-forward; for example, how many somewhat credible models arriving at one conclusion does it take to challenge a model already accepted by the community that arrives at the opposite conclusion?

3.4 How Many Models Are Enough?

When exploring parametrical uncertainty in a single model, one can design experiments that run the model with millions of parameter combination, to fully map out the uncertainty space. For exploring structural uncertainty, no such automated methods currently exist. A handful of models implemented by hand can't come anywhere close to accurately estimate the structural uncertainty of a simulated system. For this reason, multi-model studies have sometimes been referred to as "a collection of best guesses" [10] rather than a uncertainty estimation method.

Because of the non-negligible cost of every new model added to the study, the question of when a sufficient number of models has been reached becomes relevant. Ultimately, there is no magic number for this, since it depends on the simulated system, the diversity of the included models, and what is meant by "sufficient", among many other things. As previously noted, modeler's always need to be careful about what conclusions are drawn from simulation studies and with what confidence these conclusions are communicated to policymakers; there is no feasible number of included models that would mitigate this need for caution.

4 Can a Multi-Model Approach Be Feasible?

While there exist several benefits with using a multi-model approach for a simulation study, it will not always be possible or desirable to do so. If the available time and resources of the study are too constrained, developing or re-applying more than one model might either not be possible at all or could require a decrease in quality and detail of included models. However, if the process is designed cleverly, a multi-model study could require less additional effort than it might appear at first glance:

- Using one or a few "base" models that are then modified or extended is cheaper than creating each included model from scratch. If parts of this modification of models to create new ones could be automatized (similar to parameter sweeps when exploring parametrical uncertainty) while still ensuring that all these models are reasonable and consistent with the current purpose, the feasible number of models to be included would increase even further.
- While each model would typically have to be individually validated, the process would likely look very similar for the models. Many of the most time-consuming activities during model validation (data acquisition, getting into contact with domain experts, etc) would not have to be repeated for each model.

– As discussed above, if the libraries of publicly available ABSS models continue to grow and efforts toward higher level of model reusability are taken by the community, the inclusion of previously implemented models could become an appealing option.

There is much room for future research to explore how to design ABSS multi-model studies that are both feasible and rewarding. This includes topics such as automating the model modification process, how to best aggregate results, how validation and documentation of models should be performed and whether aggregated numerical predictions can be sufficiently accurate to make ABSS models useful for this purpose. If these efforts prove successful, the multi-model approach can become an answer to some of the current largest challenges in the agent-based community.

5 Conclusion

This paper has explored a multi-model approach to agent-based social simulation. An example study using five models of epidemic spread was presented, and key considerations for designing a multi-model study were discussed. Though more research is needed regarding their design, the potential of multi-model studies to tackle the issue of structural uncertainty in simulated systems make them a promising candidate to solve several of the largest challenges that the ABSS community faces.

Acknowledgments. This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program Humanities and Society (WASP-HS) funded by the Marianne and Marcus Wallenberg Foundation.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. An, L., Grimm, V., Sullivan, A., Turner Ii, B., Malleson, N., Heppenstall, A., Vincenot, C., Robinson, D., Ye, X., Liu, J., et al.: Challenges, tasks, and opportunities in modeling agent-based complex systems. Ecological Modelling **457**, 109685 (2021)
2. Balci, O., Ball, G.L., Morse, K.L., Page, E., Petty, M.D., Tolk, A., Veautour, S.N.: Model reuse, composition, and adaptation. In: Research Challenges in Modeling and Simulation for Engineering Complex Systems, pp. 87–115. Springer (2017)
3. Diomede, T., Davolio, S., Marsigli, C., Miglietta, M.M., Moscatello, A., Papetti, P., Paccagnella, T., Buzzi, A., Malguzzi, P.: Discharge prediction based on multi-model precipitation forecasts. Meteorology and atmospheric physics **101**(3), 245–265 (2008)
4. Duan, Q., Ajami, N.K., Gao, X., Sorooshian, S.: Multi-model ensemble hydrologic prediction using bayesian model averaging. Advances in water Resources **30**(5), 1371–1386 (2007)

5. Edmonds, B.: Different modelling purposes. In: Simulating social complexity: A handbook, pp. 39–58. Springer (2017)
6. Edmonds, B., ní Aodha, L.: Using agent-based modelling to inform policy–what could possibly go wrong? In: International Workshop on Multi-Agent Systems and Agent-Based Simulation. pp. 1–16. Springer (2018)
7. Heppenstall, A., Crooks, A., Malleson, N., Manley, E., Ge, J., Batty, M.: Future developments in geographical agent-based models: Challenges and opportunities. Geographical Analysis **53**(1), 76–91 (2021)
8. Johansson, E., Lorig, F., Davidsson, P.: Combination of agent-based social simulation models: Approaches and challenges. In: 2025 Annual Modeling and Simulation Conference (ANNSIM). pp. 1–13. IEEE (2025)
9. Lorig, F., Johansson, E., Davidsson, P.: Agent-based social simulation of the covid-19 pandemic: A systematic review. JASSS: Journal of Artificial Societies and Social Simulation **24**(3) (2021)
10. Parker, W.S.: Ensemble modeling, uncertainty and robust predictions. Wiley interdisciplinary reviews: Climate change **4**(3), 213–223 (2013)
11. Stonedahl, F., Wilensky, U.: NetLogo Virus on a Network model. http://ccl.northwestern.edu/netlogo/models/VirusonaNetwork (2008)
12. Tebaldi, C., Knutti, R.: The use of the multi-model ensemble in probabilistic climate projections. Philosophical transactions of the royal society A: mathematical, physical and engineering sciences **365**(1857), 2053–2075 (2007)
13. Tracy, M., Cerdá, M., Keyes, K.M.: Agent-based modeling in public health: current applications and future directions. Annual review of public health **39**(1), 77–94 (2018)
14. Uhrmacher, A.M., Frazier, P., Hähnle, R., Klügl, F., Lorig, F., Ludäscher, B., Nenzi, L., Ruiz-Martin, C., Rumpe, B., Szabo, C., et al.: Context, composition, automation, and communication: The c2ac roadmap for modeling and simulation. ACM Transactions on Modeling and Computer Simulation **34**(4), 1–51 (2024)
15. Wilensky, U.: NetLogo. http://ccl.northwestern.edu/netlogo (1999)
16. Xiao, H., Cinnella, P.: Quantification of model uncertainty in rans simulations: A review. Progress in Aerospace Sciences **108**, 1–31 (2019)

Supplementary Material

Appendix A: Model Parameter Values

Table 4. Model parameter values used for the experiments performed in the paper.

Parameter	Value	Parameter	Value
number-of-nodes / number-of-people	800	child-factor	0.3
average-node-degree	6	YA-factor	1.5
initial-outbreak-size	3	adult-factor	1.0
virus-check-frequency	2	elderly-factor	0.7
virus-spread-infected-contacts	3.0%	store-opening-hour	8h
virus-spread-exposed-contacts	10.0%	store-closing-hour	24h
virus-spread-exposed-random	5.0%	virus-spread-store-risk	0.2
virus-spread-children	5.0%	number-of-nodes-country	50
virus-spread-YA	15.0%	average-node-degree-country	6
virus-spread-adults	10.0%	number-of-agents-country	50000
virus-spread-elderly	7.0%	initial-outbreak-size-country	50
infection-rate-home	10.0%	travelers-per-city	20
infection-rate-work	5%	virus-spread-chance-country	50%

Appendix B: Model Code

The code for the five models used in the study can be found at: https://github.com/emjolund/ABSS-combination-models.