9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

Large Scale Calibration of Agent-Based Models in Social Systems with Sensitive Data

Keywords: agent-based, modeling, calibration, large-scale, sensitive data

Extended Abstract

Social systems are complex and non-linear, with numerous interacting elements that can influence each other in unexpected ways, making them difficult to model and understand. However, computational methods, such as agent-based modeling (ABM), offer a way to study these systems. ABM involves simulating the behavior of individual agents and their interactions over time, allowing researchers to capture the complexity and non-linearity of social systems, test different scenarios and generate new insights. However, calibrating an ABM to accurately reflect real-world behavior is a major challenge, as social behavior is highly variable and context-dependent, and the data required for calibration is often incomplete, unreliable or include sensitive data. Furthermore, ABMs can be computationally expensive to run, requiring significant resources and time.

In this work we present our approach to calibrating large-scale agent-based models using supercomputing. Specifically, we focused on a real-world model that explores the effects of non-pharmaceutical interventions on the spread of COVID-19 in the Netherlands. To the best of our knowledge, there is no prior work that tackled the challenge of calibrating ABMs on a full population level on a 1:1 scale (one agent representing one citizen) because it is computationally expensive. However, with the use of a novel ABM platform called BioDynaMo [1], we were able to tackle this challenge and introduce fine-grained data on a population scale to calibrate the model. Furthermore, due to the 1:1 scale modeling, we were able to use register data from Statistics Netherlands to initialize the model. By using this data, we were able to calibrate the model to better reflect the actual behavior and interactions of individuals in the Netherlands, improving the accuracy of the simulated data. It is worth noting that the register data obtained from Statistics Netherlands is sensitive data that is not publicly accessible. Therefore, to ensure the security and privacy of the data, we needed to use a secure supercomputing environment [2] to work with the data. This environment provided us with the necessary security measures to handle the data appropriately while also allowing us to run the computationally intensive ABM simulations at a large scale. The use of such a secure environment is becoming increasingly important in social science research as more and more sensitive data is being used to study complex social systems.

Our contributions can be summarized as follows:

- Present the Multi-Simulation Manager that is able to run multiple agent-based models in parallel as part of an optimization algorithm using MPI and OpenMP, and is implemented as an extension to BioDynaMo.

- Implement an epidemiological model of the first COVID-19 wave in the Netherlands in BioDynaMo and use register data from CBS for enriching the model.

- Explore the parameter space of the COVID-19 model using a distributed version of the Particle Swarm Optimization method.

– Run the model at different resolutions and analyze the difference in simulated results. Report preliminary results on the accuracy and the runtime.

In order to allow for calibration of large agent-based models in BioDynaMo, we developed a new class that can be used for distributing multiple simulations as separate processes using MPI, the Multi-Simulation Manager. Any iterative algorithm that can run its iterations independently of each other can be used within this class, but for the scope of this study we focus on the Particle Swarm Optimization (PSO) [3].

In short, PSO is a population-based optimization algorithm used to find the global optimum solution in a search space. In PSO, a set of particles is initialized randomly within the search space, and each particle moves in the space to find the global optimum. The movement of each particle is guided by its own best known position and the global best position found so far by the entire swarm. The algorithm continues until the convergence criteria are met or a maximum number of iterations is reached. PSO is a widely used optimization algorithm due to its simplicity and efficiency.

Figure 1 shows an overview of the implementation. In the Multi-Simulation Manager, the master rank initiates a PSO algorithm with N particles and at most max_iter iterations. Each particle, within each iteration, represents a full simulation and is executed by one of the worker ranks in the worker pool and is repeated M times for statistical significance. The simulation is executed in multi-threaded fashion with OpenMP with a user-specified number of threads. Each worker rank computes the mean-squared error (MSE) between the simulated values and the observed (expected) values. After a full iteration, the master receives the MSE from each worker and updates the weights of the optimization algorithm. The algorithm stops when convergence is reached or max_iter is reached.

Our experimental setup includes the ODISSEI Secure Supercomputer (OSSC), which is a secure environment that sits on top of the Dutch national supercomputer (Snellius) that disallows any incoming or outgoing internet traffic to protect data that is accessible within the environment. The customized version of BioDynaMo and the model code are embedded in a Docker container, which is used to launch an MPI job to the underlying Slurm cluster that the OSSC is based on.

Preliminary results show the ability to efficiently scale out the Particle Swarm Optimization routine that is illustrated in Figure 1. We aim to produce the results to show how the resolution of the agent-based model affects the accuracy of the simulated data. We anticipate that the error bar of each simulated data point will be smaller as the resolution increases, which gives rise to producing an accurate digital twin of the Dutch population.

References

[1] L. Breitwieser, A. Hesam, J. De Montigny, *et al.*, "Biodynamo: A modular platform for high-performance agent-based simulation," *Bioinformatics*, vol. 38, no. 2, pp. 453–460, 2022.

[2] M. Scheerman, N. Zarrabi, M. Kruiten, *et al.*, *Secure platform for processing sensitive data on shared hpc systems*, 2021. DOI: `10.48550/ARXIV.2103.14679`. [Online]. Available: `https://arxiv.org/abs/2103.14679`.

[3] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95-international conference on neural networks*, IEEE, vol. 4, 1995, pp. 1942–1948.

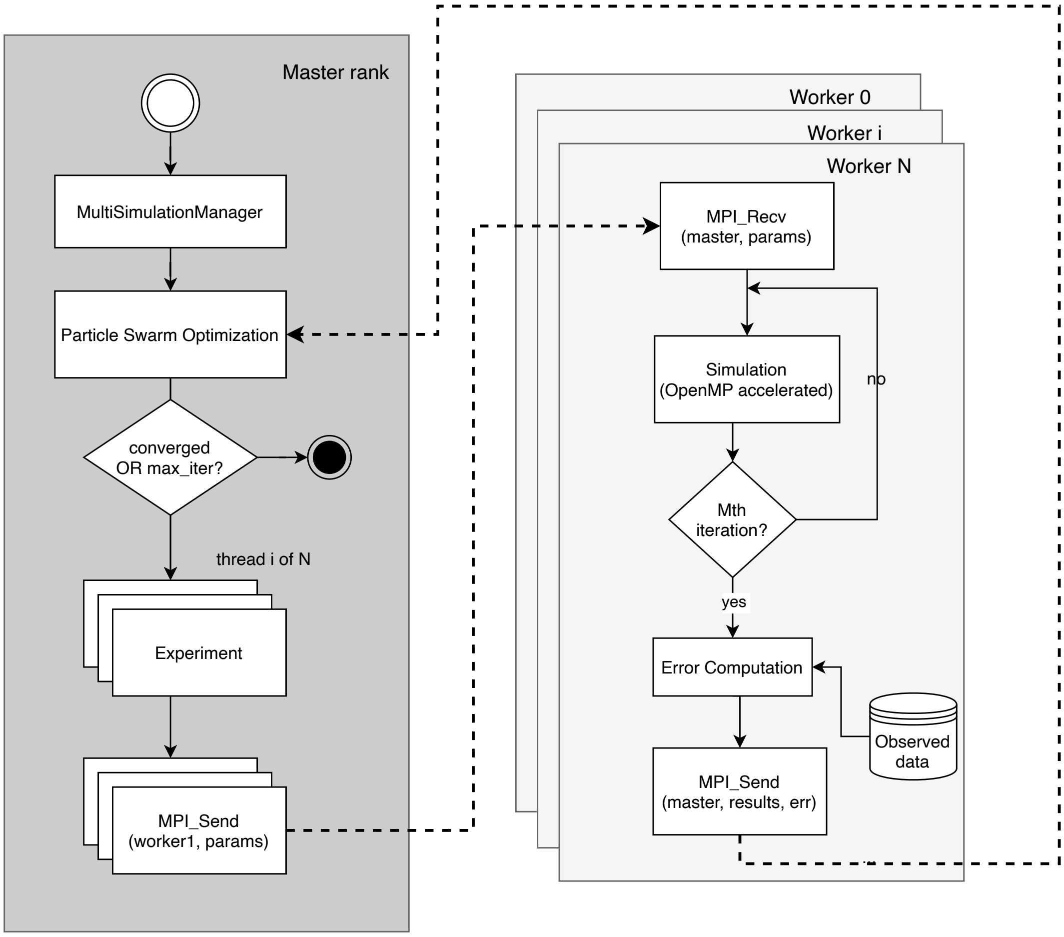

Figure 1: Flowchart of the Multi-Simulation Manager using the Particle Swarm Optimization algorithm for calibrating against observed data.