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

Hesam, A.S.; Pijpers, Frank; Rademakers, Fons; Al-Ars, Z.

Publication date
2023

Document Version
Final published version

Citation (APA)
Hesam, A. S., Pijpers, F., Rademakers, F., & Al-Ars, Z. (2023). *Large Scale Calibration of Agent-Based Models in Social Systems with Sensitive Data*. Poster session presented at 9th International Conference on Computational Social Science, Copenhagen, Denmark.

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

A.S. Hesam[1,2], F.P. Pijpers[2,3], F. Rademakers[4], Z. Al-Ars[1]

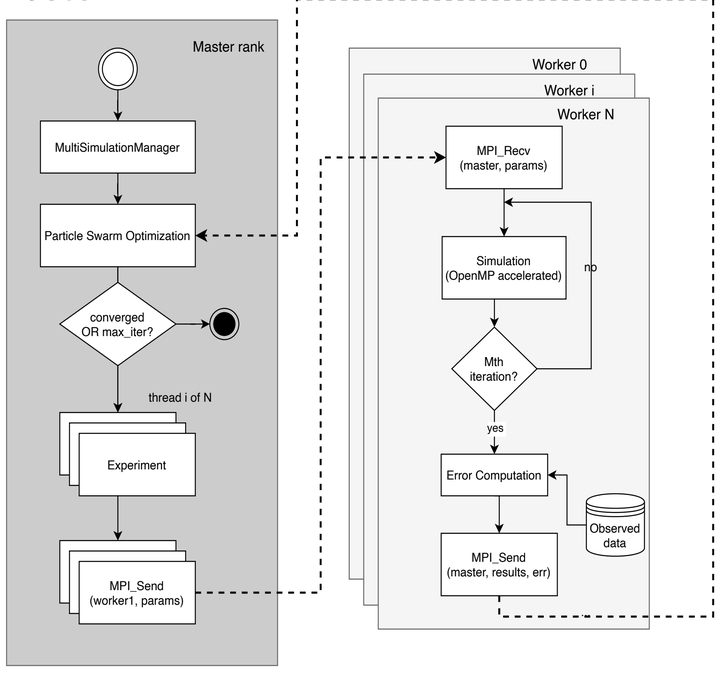
1 Computer Engineering Lab, Delft University of Technology, Delft, Netherlands
2 Statistics Netherlands, The Hague, Netherlands
3 Korteweg-deVries Institute for Mathematics, University of Amsterdam, Amsterdam, Netherlands
4 CERN, Geneva, Switzerland

Introduction

- Agent-based modeling (ABM) has emerged as a powerful tool for exploring the behavior of complex social systems over time
- ABMs are inherently challenging to calibrate due to the lack of knowledge regarding the parameter space
- The computational time for large-scale and complex ABMs becomes exorbitant, making the utilization of calibration methods, which often require numerous runs, unfeasible
- In numerous social science studies, ABM parameters are unknown and require calibration to be established
- We extended a state-of-the-art agent-based simulation platform, BioDynaMo [1], to support calibration on HPC clusters
- We created a workflow to support reproducible research on ABM on HPC systems that require sensitive data
- We implemented a proof-of-concept COVID-19 model in BioDynaMo to showcase the system
- We outline future directions of this work

Distributed calibration of agent-based models

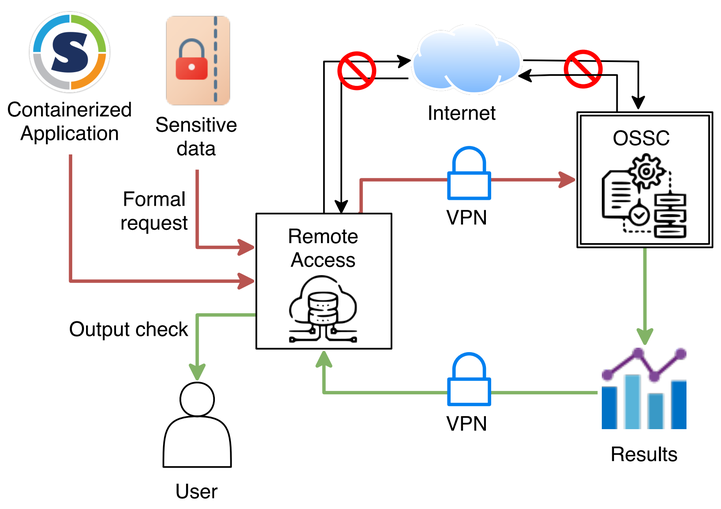
- Iterative optimization algorithms can be run in parallel (1 iteration = 1 simulation)
- A new module, MultiSimulationManager, was developed in BioDynaMo to facilitate distributed calibration
- We implemented a Master-Worker pattern using MPI
- The Master rank performs the optimization loop and dispatches parameter sets
- Workers execute simulations with the provided parameter set, compute error with observed data, and send back error to Master



- Each simulation is executed in multi-threaded fashion using OpenMP
- Optimization algorithms implemented: Particle Swarm, Parameter Sweep
- ROOT [2] is used for parameter distribution (serialization)
- A web-based dashboard enables graphical exploration of the searched parameter space

Secure Supercomputing Environment

- We made use of the ODISSEI Secure Supercomputer (OSSC) [3]
- The OSSC is a virtual (Slurm) cluster that runs on the Dutch national supercomputer (Snellius). It is developed and hosted by SURF
- There is no ingoing or outgoing internet traffic to guarantee security
- A remote access (RA) environment within Statistics Netherlands acts as the stepping stone portal
- Sensitive data, once formally requested, are uploaded to the RA, which can then be securily uploaded to the OSSC via a dedicated VPN
- Users request heir (containerized) application to be uploaded to the RA

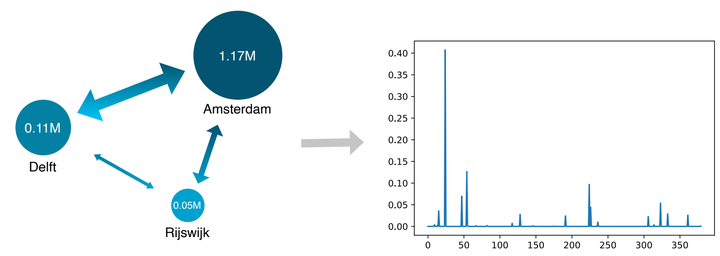


- One (thin) node of the OSSC has 128 cores, 256GB of RAM
- Sensitive data can be mounted inside the container to enable analysis in a secure and reproducible way
- Results can be sent back to the RA for post-simulation analysis
- Final results can be sent to the users after an output check

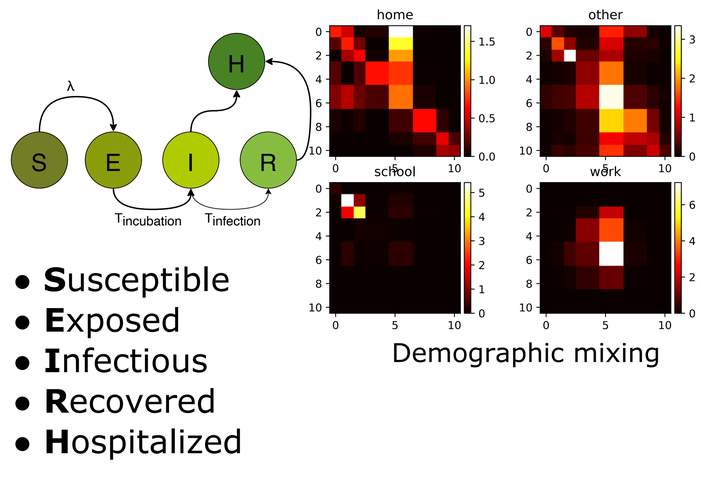
COVID-19 model: first wave in the Netherlands

- We implemented an existing COVID-19 model [4] in BioDynaMo
- There are 4 free parameters representing the effects of non-pharmaceutical interventions on the force of infection (one for each phase of interventions)
- Calibration of these 4 parameters is done through matching simulated hospital admissions with observed data (period of doubling of admissions)
- Using BioDynaMo we were able to increase the resolution from 1 agent : 100 persons to 1 agent: 1 person
- Gravity model used instead of tracing mobile data for mobility behavior
- Register data of Statistics Netherlands was used to initialize agents with micro-level data for high demographic fidelity

Mobility behavior



Infection behavior



- **S**usceptible
- **E**xposed
- **I**nfectious
- **R**ecovered
- **H**ospitalized

Conclusions & future work

- We developed a scalable system to enable calibration and parameter exploration of large scale agent based models that require sensitive data
- We are working to **quantifying the effects of high resolution** pandamic models
- We want to explore the difference between the use of **microdata vs synthetic (random) data** for initialization
- Perform **performance benchmarks** to quantify the distributed calibration system

References

[1] Breitwieser, Lukas, et al. "BioDynaMo: a modular platform for high-performance agent-based simulation." Bioinformatics 38.2 (2022): 453-460.
[2] Brun, Rene, and Fons Rademakers. "ROOT—An object oriented data analysis framework." Nuclear instruments and methods in physics research section A: accelerators, spectrometers, detectors and associated equipment 389.1-2 (1997): 81-86.
[3] Scheerman, Michel, et al. "Secure platform for processing sensitive data on shared HPC systems." arXiv preprint arXiv:2103.14679 (2021).
[4] Dekker, Mark M., et al. "Reducing societal impacts of SARS-CoV-2 interventions through subnational implementation." Elife 12 (2023): e80819.

Contact details: Ahmad Hesam, a.s.hesam@tudelft.nl