State Estimation and Data Assimilation for an Agent-Based Model using a Probabilistic Framework ⋆

Nick Malleson[1,3][0000−0002−6977−0615], Luke Archer[3], Jonathan A. Ward[2][0000−0003−3726−9217], Alison Heppenstall[1,3][0000−0002−0663−3437], and IMPROBABLE?[4]

[1] School of Geography, University of Leeds, LS2 9JT, UK
http://geog.leeds.ac.uk/
n.s.malleson@leeds.ac.uk
[2] School of Mathematics, University of Leeds, LS2 9JT, UK
http://maths.leeds.ac.uk
[3] Leeds Institute for Data Analytics (LIDA), University of Leeds, LS2 9JT, UK
http://lida.leeds.ac.uk
[4] Improbable, 30 Farringdon Road, London, EC1M 3HE, UK
http://www.improbable.io

Abstract. XXXXX NICK TO WRITE ABSTRACT

Keywords: Agent-based modelling · Probabilistic programming · Uncertainty · Data assimilation · State estimation · Bayesian inference

1   Introduction and Objectives

Calibration – the process of finding optimal values for a model's parameter – is well studied in the field of agent-based modelling. Two aligned topics that are much less well studied, however, are those of *state estimation* and *data assimilation*. State estimation refers to the practice of estimating the *true* state of a system. Although their complexity precludes the true state of human systems ever being known precisely, it can be estimated by combining a model of the system with current data. The practice of estimating a system's current state using a model and some data is often termed data assimilation[5]. Data assimilation has, and continues to be, extremely well studied in fields such as meteorology [4] where it has been credited as being part of the reason that weather forecasts have improved so substantially in recent years [1].

Applications of data assimilation in agent-based modelling are scarce, however; only a handful of attempts have been published [5–7]. The aim of this research is to contribute to this emerging field by performing data assimilation on a simple agent-based model of a hypothetical crowd. A series of 'identical twin' experiments are performed, whereby the agent-based model first generates 'pseudo-truth' data that reflect the 'true' state of the system (in the real world such data are never available) and then the model is then re-run in a data assimilation framework that attempts to replicate the truth data. Importantly, by varying the amount of information about the true system state that are provided to the data assimilation algorithm it is possible to estimate the amount of information that might be required were a real crowd of people to be simulated. Here it is assumed that some individuals in the hypothetical data are tracked, so their spatio-temporal locations are known, but in future work the framework will be extended to aggregate data as well (i.e. counts of people rather than individual traces).

Data assimilation is a Bayesian approach. The prior estimate of the state of the (pseudo) real system is the current estimate of the system state as provided by the (agent-based) model.The posterior is the new estimate that is obtained once the most recent observations from the (pseudo) real system have been assimilated. Probabilistic programming [2, 3] is a relatively new approach to writing computer programs whereby variables represent probability distributions rather than actual values. This approach is therefore ideally suited to Bayesian modelling. Hence this paper makes a further contribution: it explores the value of a probabilistic programming library, *keanu*, and the probabilistic approach to modelling in general, as a means of performing data assimilation on an agent-based model.

---

[5] It is worth noting that although data assimilation techniques can be used to adjust model parameters in response to new data, here the technique is applied solely to the task of state estimation.

2   Background

> How this work fits in to the wider data assimilation schema (is it 'nudging'? and how it compares to traditional data assimilation. Basically a very, very brief literature review.

> Outline what probabilistic programming is, and what Keanu is.

3   An Example Agent-Based Model: *StationSim*

> NM: Briefly outline station sim to show that it has some of the normal characteristics of an ABM (one paragraph).

4   Data Assimilation Framework

> Explain the basic framework here, e.g. number of iterations, number of windows, calculating the posterior for the state, etc. We apply the same framework to the simple model and station sim.

5   Results

5.1   Full Knowledge of the System

> Experiments when the probabilistic model has full knowledge of the system

5.2   Knowledge of Only Some Agents

> Experiments when we only give the probabilistic model access to partial information in the state vector (i.e. only a few agents)

6   Conclusions

XXXX Conclusions

Link to ABMUS Workshop Themes

This paper contributes to the challenge set out in the workshop theme in two ways. Firstly, by developing methods that support "trusted models that can be used by industry and governments to enhance decision-making, and that can incorporate real (and real-time) data sets in a meaningful way". Secondly, regarding the 'humans and devices' theme specifically, by beginning to estimate the volume of sensors that would be required in order to successfully model crowds in real time. Without a reliable means of incorporating real-time data into urban models, their use will continue to be limited to scenario evaluation based on historical data rather than providing the most likely estimates of the *current* state of urban systems as well as short-term forecasts.

References

1. Bauer, P., Thorpe, A., Brunet, G.: The quiet revolution of numerical weather prediction. Nature **525**(7567), 47–55 (2015). https://doi.org/10.1038/nature14956
2. Ghahramani, Z.: Bayesian non-parametrics and the probabilistic approach to modelling. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **371**(1984), 20110553–20110553 (Dec 2012). https://doi.org/10.1098/rsta.2011.0553
3. Ghahramani, Z.: Probabilistic machine learning and artificial intelligence. Nature **521**(7553), 452–459 (May 2015). https://doi.org/10.1038/nature14541
4. Kalnay, E.: Atmospheric Modeling, Data Assimilation and Predictability. Cambridge University Press (2003)
5. Malleson, N., Ward, J.A., Heppenstall, A., Adcock, M., Tang, D., Coello, J., Crols, T.: Understanding Input Data Requirements and Quantifying Uncertainty for Successfully Modelling 'Smart' Cities. In: 3rd International Workshop on Agent-Based Modelling of Urban Systems. Stockholm, Sweden (2018)
6. Wang, M., Hu, X.: Data assimilation in agent based simulation of smart environments using particle filters. Simulation Modelling Practice and Theory **56**, 36–54 (2015). https://doi.org/10.1016/j.simpat.2015.05.001
7. Ward, J.A., Evans, A.J., Malleson, N.S.: Dynamic calibration of agent-based models using data assimilation. Royal Society Open Science **3**(4) (2016). https://doi.org/10.1098/rsos.150703