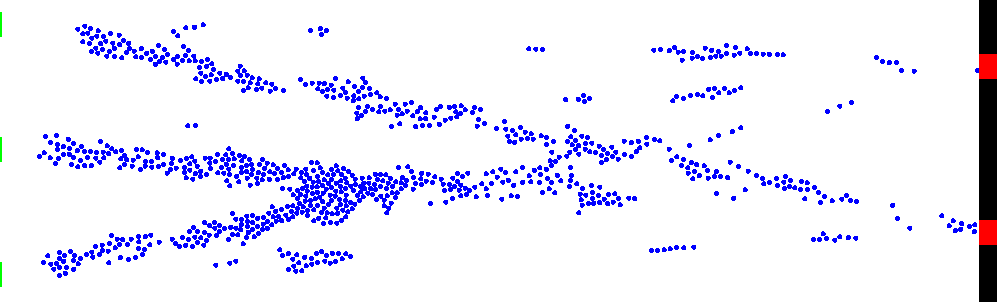
**Project Background**

Agent-based models (ABMs) are ideally suited to modelling the behaviour and evolution of social systems. However, input data are noisy and sparse, and human behaviour is extremely uncertain. Therefore one of the key challenges facing the discipline relates to the quantification of uncertainty within ABMs. This work presents initial steps towards the development of new methods that will us to better understand uncertainty in ABMs and, ultimately, to allow streams of data to be incorporated into models in *real time*.

**Data and Methods**

The broad aims of this research programme are to model cities using real-time data feeds. This project begins the necessary methodological groundwork by exploring how much real-world data are needed to successfully model a system of pedestrians. It focusses on a simple hypothetical system: that of a train station which has crowding as people try to exit (Fig 1). While simple in nature, this model was designed to include stochastic elements and so produced differing types of crowding on different runs.

To address the aims of this work in better understanding the uncertainty in ABMs, a probabilistic programming library that is currently being developed by Improbable was utilised to perform Bayesian inference.

**

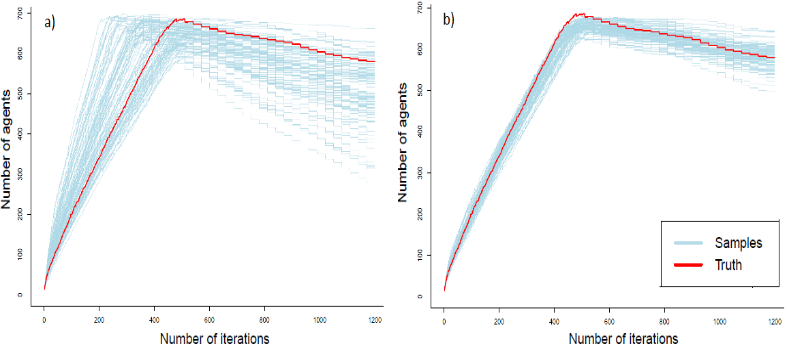
*Fig. 1: A snapshot of the simple, hypothetical model. Agents arrive from the green entrances on the left and move towards the red exits on the right. During the course of the runtime crowding occurs in different areas.*

Our model was used to generate ‘pseudo-truth’ data to represent a hypothetical reality. These data were then used by the probabilistic model (they were ‘*observed’*) to create a posterior distribution over the model parameters. This distribution was sampled using a Metropolis Hastings algorithm. These samples allow us to see whether the probabilistic model is able to constrain the ABM to the ‘true’ observations. In other words, can the probabilistic model find the parameters in the ABM that were used to generate the pseudo-truth data?

**Key Findings**

The model was successfully constrained, allowing us to reduce the uncertainty in the outputs from the ABM with respect to the truth data. In other words, the probabilistic model was able to estimate approximately what the inputs to the ABM were, and to produce output that was similar to the pseudo-truth.

This was demonstrated using the following model output: the number of people in the simulation at each step. When *observing* the pseudo-truth data, the model was shown to have been constrained (the samples are much closer to the ‘truth’) (Fig 2). When the pseudo-truth data were not observed, the samples varied widely (as would be expected).

****

*Fig. 2:* *Results of sampling the prior (a) and the posterior (b) (i.e. after observing the pseudo-truth data). When the ‘truth’ data are used to constrain the posterior distribution, the sampling routine is much better able to estimate the input model state, so the outcomes of the samples are much closer to the ‘truth’ data.*

**Value of the Research**

This work lays the groundwork for a better understanding about how real data can be used to reduce uncertainty in ABMs. This an important initial step for the application of these methods to smart city modelling. By utilising these methods we can produce more robust agent based models for urban systems that can be of greater use when making policy decisions. Working with Improbable has been mutually beneficial: the LIDA team have gained access to an invaluable probabilistic modelling library, and Improbable have been able to test the library on a real use case with an agent-based model.