GEM: a domain-specific modelling language for disease outbreak models

Lead supervisor: Chris Jewell <c.jewell@lancaster.ac.uk>, MARS, School of Mathematical Sciences, Lancaster University.

The GEM project is aiming to construct a domain-specific modelling language (DSML) for analysing infectious disease models for outbreaks such as Covid19, influenza, and malaria. The goal is to design a language that enables an epidemiologist to describe a disease system, with the computer then automatically build the model, generating the statistical and machine learning algorithms for training (fitting) and prediction from that model. In doing so, we aim to revolutionise the speed at which epidemiologists can respond to the next disease outbreak, providing quantitative information for policy makers quickly and reliably.

The GEM project currently lies in two halves -- *gemlang*: a high-level model description language, and *gemlib*: a low-level GPU-enabled Bayesian inference library written using Python/JAX. Whereas the former is capable of describing a model, the latter is able to implement a wide class of models. This project is about bringing the two halves together by developing an intermediate representation (IR) that can be machine-read to turn the *gemlang* model description into executable *gemlib* code.

Our approach to the GEM IR is to encode the mode as a directed-acyclic graph (DAG). Task 1 will develop a class hierarchy for representing different model components within a tree-like structure, with Task 2 developing a suite of tree-walking algorithms capable of manipulating, editing, and simplifying the tree. If time, Task 3 will be to develop an algorithm that takes the simplified DAG and generates code based on *gemlib* to implement the outbreak model on HPC hardware. The project will work in Python throughout, making use of a wide range of tools from the standard library, as well as advanced machine-learning libraries such as JAX.