The Modeling Process and the Modeling Stack Study

Tomasz D. Loboda

January 25, 2020

1 The Modeling Process

Agriculture and food systems is the application for which we intend to develop the AI-assisted modeling methods we discuss here. However, we see it as one of many applications of scientific modeling and whenever possible we intent to make these methods generalize (to, e.g., biological sciences). To this end, by analyzing the process modelers employ, this study will decompose it into *elementary modeling operations*. When put together into sequences and cycles, these operations will comprise specific *modeling tasks* such as de novo modeling, initial model calibration and evaluation, and recurring model update and improvement (e.g., recalibration, restructuring in light of new scientific evidence, etc.). This approach informs and interacts with the Maia synthetic assistant.

Once elementary modeling operations are organize into a structure (likely a multigraph), that structure will serve several important purposes. First, it will be a detailed map which we will use to chart direction of further AI-assisted modeling technology. Second, that map will also obviate the most efficient paths of education for people who seek to become adept modelers. Third, by being externalized from any one person's mind, it will be subject of discussion, scrutiny, and critique not only from us, but importantly, from the scientific modeling community. This is one way in which we intend to achieve the aforementioned generalizability from agriculture and food systems. Four, the map will make it easier to assign implementation responsibilities. For example, programmers may not fully grasp the scale and direction of the endeavor (i.e., the big picture) without seeing all the pieces and appreciating how they interact. Five, by making it clear how the elementary operations interact with one another, the map itself will be an application programming interface (API) making it even easier to maintain the entirety of the machinery and, when needed, substitute existing implementations with better alternatives (or providing multiple options all at the same time).

2 The Modeling Stack

One of the elementary operations is composing a new model (or recomposing an extant one). A model is composed from a set of modeling building blocks. These building blocks are organized into a graph of its own (i.e., it is distinct from the elementary modeling operations graph we have described earlier). That graph specifies operations like generalization and specialization commonly used in object oriented programming languages. To give an easy example, the abstract primitive `Distribution` specializes into, among others, the `TruncatedNormal`. We at the University of Pittsburgh have already begun developing those building blocks or *modeling primitives*. Those primitives are part of the *Modeling Stack*, an idea that formalizes how complicated models can be composed from smaller parts. The stack is currently being implemented and is already partially functional. Our initial modeling and simulation framework of choice the Modeling Stack has been PRAMs [2] which are based on ABMs and extend them to offer a number of benefits.

The Modeling Stack consists of three layers. On the highest level are the modeling primitives. Those primitives are brought together in various configurations on the second level where they form models. Those models, so far represented symbolically, are subsequently represented as executable code on the final, lowest level of the stack.

In agreement with generalizability, the some modeling primitives are domain agnostic and represent, e.g., some of the fundamental processes observed in nature (such as the `CountingProcess`). As we have mentioned, hierarchical dependencies exist between some primitives; this allows for specialized and domain-specific primitives. For example, `TimeInvariantMarkovChain`, itself very general, becomes the basis for the `SIRSModel`, one of the fundamental epidemiological models. Of course, this is just an example and the scope of the primitives graph is much larger. The point is that domain-specific modeling is built upon general primitives bringing unification to the current practice of scientific modeling which is otherwise unnecessarily fragmented into many fields of inquiry. This unification paves the way to polymathic expertise and education in modeling.

3 Team Coherence

As we have mentioned above, the technology around the Modeling Stack is already partially operational. We are currently able to turn models composed from modeling primitives into executable code that implements PRAM simulations and subsequently run that code to obtain results. To use another simple example, by composing a model from two `SIRSModel` primitives, each with different parameters, we are able to investigate the interaction between two concurrent epidemics. By injecting stochasticity to the simulations, we can, for instance, gain insight into the possible number of infected people region-by-region and consequently anticipate the optimal hospital locations in a large municipal area. In this study, we intend to use the same principle to enable interaction of multiple crop models (e.g., multiple instances of the DSSAT model, each for a specific regionally-congruent type of crop). However, our plans go beyond that.

First, being work-in-progress, we currently have no user interface that would let modelers compose and recompose models in a natural fashion. While creating such an interface is one of our priorities in another on-going project, this study will give MAIA the ability to automatically compose and recompose models via interaction with the modeler.

Bugs' Vignette. A state of Washington agriculture policy-maker Bugs wants to investigate the extent of damage an extremely hot summer can have on a specific cultivar of orange. Bugs asks MAIA to create such a model. MAIA recognizes that such a problem has been addressed before and suggests to import (one of the elementary modeling operations) one of the previously used models. Unfortunately, none of those models match the spacial distribution of orange orchards in the area. Bugs asks MAIA to suggest an initial distribution of farms based on information from published sources (including tabular data extracted by MODX). Unhappy with the results, Bugs requests MAIA to instead use the spacial distribution implied by the latest aerial images of the area. Finally content, Bugs instructs MAIA to run an ensemble of five thousand simulations with stochasticity in parameter values based on empirical distributions. He also makes sure that the distribution for temperature is shifted into the extreme heat region. Bugs manually selects the three most commonly observed farmer behavior mechanics implying he's interested in simulating farmers as agents. The final tally of simulations MAIA will execute after Bugs calls it a day ends up being 15,000. Thanks to automatic scheduling of those simulation done by OCCAM, Bugs will be able to review the results in the morning to make a policy recommendation or plan further analyses.

References

[1] Childers, B., Mosse, D., & Jones, A. (2019) OCCAM: Open Curation for Computer Architecture Modeling. *Computer software.* `https://occam.cs.pitt.edu`

[2] Cohen, P.R. & Loboda T.D. (2019) Probabilistic Relational Agent-Based Models. International Conference on Social Computing, Behavioral-Cultural Modeling & Prediction and Behavior Representation in Modeling and Simulation.