APPLIED CATEGORY THEORY FOR MODELING

JOHN C. BAEZ

Mathematicians have long used category theory to organize their thinking: this is a very general framework, first invented in 1945 and greatly extended since then, which organizes many different branches of mathematics into a coherent whole. Since the 1990s, category theory has become important in theoretical computer science. Starting around 2010, *applied category theory* has become a tool for developing models of real-world systems, from public health [1] to the specification and verification of tasks for artificial intelligence [2] and beyond.

Principles. Applied category theory is not an alternative to other modeling techniques, but rather a complement: a systematic framework that makes it easier for interdisciplinary teams to build, refine and adapt models. Too often, a scientific model is a monolithic piece of code whose inner workings are understood only by the experts who created it. Model assumptions are expressed as algorithms, making it hard to *inspect* the model, *question* its assumptions, and *change* the model by changing these assumptions. Monolithic models gradually becomes harder to use as more and more features are added. Further, combining several such models to form a larger one is difficult because they are based on assumptions that are nowhere clearly spelled out.

We need a framework for modeling that more closely resembles the common language of science, so that teams in different disciplines can more easily work collaboratively. A model should be, among other things, a meeting ground for communication. Thus, a model should be a mathematical structure, made of clearly organized parts, ideally presented in a visual way that requires little training to understand.

Applied category theory addresses all these issues, using principles such as these:

(1) **Compositionality:** Models of specific subsystems may be constructed individually and then coupled together to build larger models. This gives a flexible approach to modeling that supports ongoing interaction between different teams. To achieve this, each model is not merely a piece of code, but a crisply defined structure designed from the start to be easily inspected and easily combined with other models. The "interface" of each model must be clearly specified, to make it clear how this model can affect and be affected by the computations in other models. Technically, this is done by treating models as "morphisms" in a category, and the coupling of models as "composition" of these morphisms. Thus, this approach to model design is known as "compositionality".

(2) **Functorial Semantics:** There is a clear distinction between the model and a specific way of running the model and extracting information from it. For example, one can take essentially the same model and run it either deterministically or stochastically, apply analyses to determine the effect of changing parameters, etc. This is achieved by treating different ways of running a model as different "functors" from a category whose morphisms are models to various other categories.

(3) **Interoperability:** Models expressed in different frameworks can be coupled or interoperated. This is, in fact, a consequence of compositionality and functorial semantics.
(4) **Stratification:** There are systematic methods for refining models, which do not require labor-intensive and error-prone rewriting of large amounts of code. These methods are particularly well-developed for compartmental models. Such models have "stocks" representing populations, quantities of substances such as nutrients, etc., and mathematically described "flows" between these stocks. To refine a compartmental model one can"stratify" it: that is, subdivide stocks into smaller stocks that differ in various characteristics. Applied category theory uses the mathematics of "pullbacks" to do this.

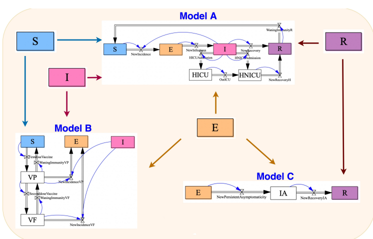
Instead of delving into the details of the mathematics that make these principles precise and actionable, it may be more helpful here to survey who is doing applied category theory, and some of the things they are doing.

Institutions and Projects. Applied category theorists can be found at many universities; they publish in journals such as *Applied Categorical Structures* and *Compositionality*, and they have an annual conference series Applied Category Theory as well as many smaller conferences.

The Topos Institute in Berkeley is a major center for research in applied category theory. Their software tools include Catlab, a software platform for scientific programming with category theory, and various packages built using Catlab, most notably:

(1) **AlgebraicDynamics**: a tool for building models of dynamical systems (that is, systems of ordinary differential equations). This allows various styles of composition, including:
   - **Undirected wiring diagrams**, where dynamical systems are composed by identifying variables of one with variables of another. For example, if several species of plankton affect the concentration of oxygen in the water, these can be separately modeled and then the oxygen concentration variables of all these models can be identified as a single variable in a larger model.
   - **Directed wiring diagrams**, where dynamical systems are composed by letting a quantity computed by one serve as a parameter in another. For example, one model can compute the rate at which organic detritus produced in the upper ocean falls to the bottom, and this (time-dependent) rate can serve as a parameter in a model of the benthic layer.

(2) **StockFlow**: a tool for creating compartmental models compositionally. This is being extensively used and developed by a team led by Nathaniel Osgood, who helped lead Canada's COVID modeling during the pandemic [1, 5].



Composing several compartmental models

(3) **AlgebraicRewriting**: Generalized term rewriting software, suited to agent-based modeling.
(4) **Decapodes**: a diagrammatic tool for representing, composing, and solving partial differential equations. This is being developed by a team led by James Fairbanks at the University of Florida [4].

More recently an affiliated organization called Topos Research UK has been formed, based in Oxford. This plays a major role in ARIA's £59 million project to create safeguarded AI [2]. Crucial to Topos UK's vision for this project is Double Operadic Systems Theory, a more powerful applied category framework for compositional modeling [6]. Ideas from this theory are being implemented in CatColab, a new software platform for collaborative scientific models, based on category-theoretic principles but intended to be used by the broader scientific community.

Many other applied category theorists in the UK are also involved in the ARIA project. Additionally, Dominic Orchard, co-director of the Institute of Computing for Climate Science at the University of Cambridge, is knowledgeable in applied category theory.

References

[1] J. C. Baez, X. Li, S. Libkind, N. D. Osgood and E. Redekopp, A categorical framework for modeling with stock and flow diagrams, in *Mathematics of Public Health: Mathematical Modelling from the Next Generation*, Springer, 2023, pp. 175–207.
[2] Advanced Research and Invention Agency, Safeguarded AI: TA1 scaffolding (theory).
[3] J. C. Baez, X. Li, S. Libkind, N. D. Osgood and E. Patterson, Compositional modeling with stock and flow diagrams, *Proceedings of the Fifth International Conference on Applied Category Theory*, EPTCS **380** (2022), 77–96.
[4] L. Morris, A. Baas, J. Arias, M. Gatlin, E. Patterson and J. P. Fairbanks, Decapodes: a diagrammatic tool for representing, composing, and computing spatialized partial differential equations, *Journal of Computational Science* **81** (2024).
[5] X. Li, E Patterson, P. L. Mabry and N. D. Osgood, Compositional system dynamics: the higher mathematics underlying system dynamics diagrams and practice.
[6] S. Libkind and D. J. Myers, Towards a double operadic theory of systems.