This paper “**Moving beyond post-hoc XAI: Lessons learned from dynamical climate modeling**” (or a variation of it) is vitally important to understanding how AI/ML techniques will be adopted by the climate science community. Since climate science differs from other scientific disciplines in the impossibility of performing controlled laboratory experiments at scale, any claims need to stand the test of time or through extensive cross-validation against historical results. That’s the primary reason that new ideas are treated with suspicion in climate science (or in any of the other earth sciences) -- without experimental validation, all new models or hypotheses appear equally in doubt. Doesn’t matter if they are AI-generated or by humans, climate scientists will file it away with all the other candidates, likely to be ignored as they can’t easily be validated, e.g. as a new semiconductor device model can be validated by a lab experiment.

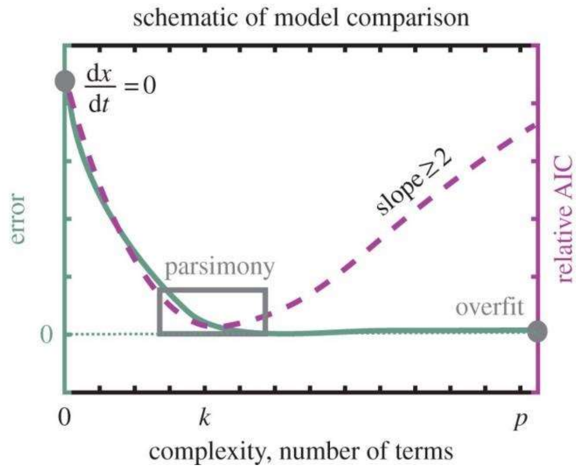
That is a daunting challenge but if nothing else, ML provides an extensive selection of cross-validation approaches that climate science can borrow from. That needs to be stated up front. As ML can easily generate matches to virtually any kind of data due to the magic of non-linear neural networks, cross-validation is necessary to weed out the many that over-fit the observations. The AI literature is full of cross-validation citations, as that is the lifeblood metric of the discipline of machine learning. The majority of neural network model fits would fail on non-training data without the benefit of rigorous cross-validation testing.

Yet, applying cross-validation alone is not enough. Climate science is further complicated by the fact that there is no consensus physics model that explains a climate behavior as erratic as the El Nino Southern Oscillation (ENSO). A cross-validation of an ML experiment matching ENSO observations would also need to explore possibly novel physical mechanisms, especially with respect to the fluid dynamics, that the GCM simulations may not be considering. That is part of the original promise of AI – that of discovering *emergent behavior* not previously considered. So that needs to be stated as well, as it could turn out that dynamical climate modeling could learn from XAI, as it’s not outside the realm of possibility that an AI experiment could find something that a GCM formulation missed.

In the context of XAI, it’s also important to acknowledge an oft-overlooked AI approach that links pure mathematical physics modeling to that of a physics-unaware neural network – that of *symbolic regression* coupled with a genetic algorithm to explore the solution space. Whereas a neural network will generate a tangled web of nonlinear interactions that are difficult to reverse engineer to a possible physical mechanism, a symbolic regression application will apply algebraic/calculus formulations with a selection of inputs to optimize a fit and perform cross-validation. Since the formulation is presented symbolically, it takes far less effort for a human to discern and sort through possible plausible physical mechanisms leading to the discovered equation.

There are certainly drawbacks to applying symbolic regression in comparison to a neural net, but there are cases that could be cited for offering a promising approach. One in particular is the independent discovery made by a symbolic regression tool called **Eureqa** in the plausible explanation of the mechanism behind the quasi-biennial oscillation (QBO) of equatorial stratospheric winds. Eureqa was able to discover a nonlinear interaction coupled to differential equation that reasonably fit to the data over the span of years that QBO data as collected, 1952 to the present. The human effort necessary was to supply a physical mechanism for the equation and numerical parameters. This was reported as an exact match to a non-linear lunar tidal interaction with the annual cycle, as described in *Mathematical Geoenergy*, P. Pukite, D. Coyne, D. Challou (Wiley/AGU, 2019). Alas, this AI-adjacent model has not gained any traction in the climate science community as it faces the same challenges of acceptance as any other model outside of the consensus.

So concerning the paper, in the discussion section, qualities such as robustness, faithfulness, *etc* are suggested. In practice, the critical factors include the 3 P's of plausibility, predictivity, and parsimony. Plausibility covers the allowable physics. Predictivity covers how well the model matches the observations, by minimizing the error. Parsimony covers the simplicity/complexity of the model in terms of the number of degrees of freedom (DOF) or terms expressed. Any standard model optimization technique features a *Pareto front* characterization curve that tracks predictivity (error minimized) versus parsimony (complexity minimized) as a metric. All symbolic reasoning tools feature this as a optimization metric.



A more rudimentary example of a Pareto front optimized symbolic reasoning solution is also described in Mathematical Geoenergy. This uses several factors to model the global temperature with a simple arithmetic superposition. The Pareto front is shown in the lower-right below.

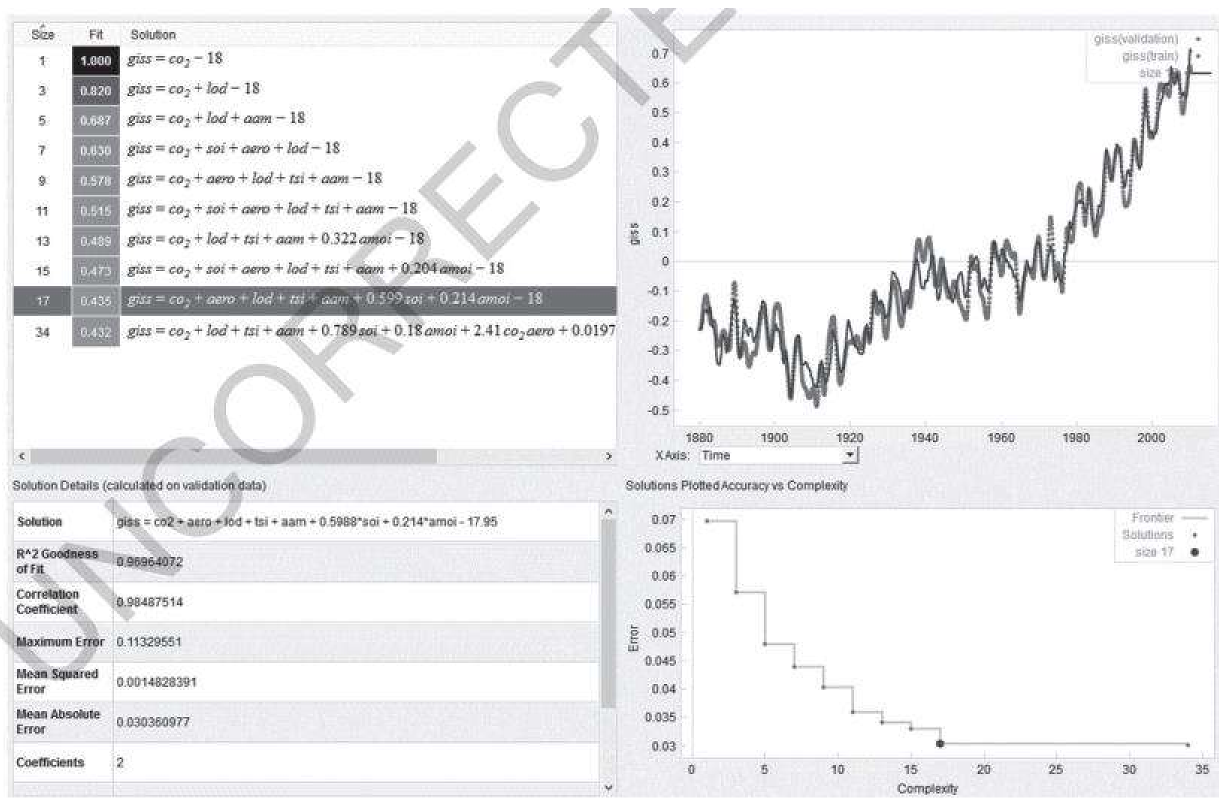


Figure 17.4 ☼ Symbolic reasoning solution to global temperature series showing a composite of main factors along a Pareto front of complexity and accuracy.

Suggestion then is to incorporate symbolic regression in addition to the neural network approach. Each of these approaches can be more suitable for different types of problems. Symbolic regression with genetic algorithms offers a more interpretable model which could be preferable in scientific applications where understanding the underlying phenomena is crucial – closer to what XAI implies. Neural networks might be more suitable for problems involving high-dimensional data, which also describes the requirements of a complex climate system. The jury is still out what will eventually work, perhaps a combination of the two, but should state the possible choices and lessons learned.