The role of machine learning in a seamless modeling approach from weather to climate time scales

Second NOAA Workshop on Leveraging AI in the Environmental Sciences

V. Balaji

NOAA/Geophysical Fluid Dynamics Laboratory (GFDL) and Princeton University

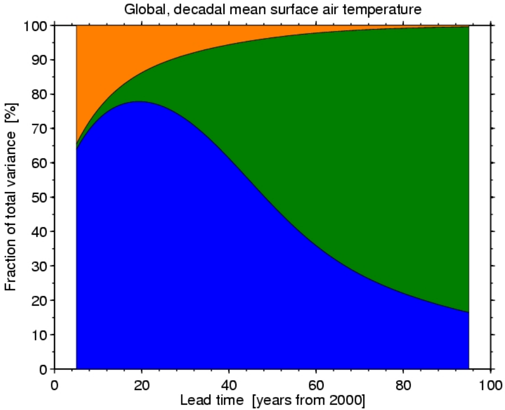
24 September 2020

Outline

1. Seamless modeling
   - What is seamlessness?
   - Seamless modeling systems
2. Machine learning for models
   - Learning parameterizations from high-resolution simulations
   - Parameter calibration
   - Training on models, training on observations
3. Ideas and challenges

Outline

Science requires going beyond observations

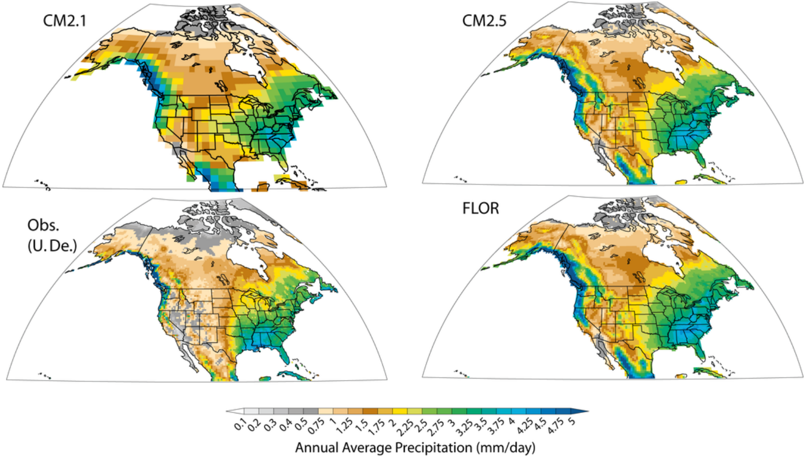


Sources of uncertainty in weather and climate simulation:

- *chaotic or aleatoric uncertainty* or internal variability
- *scenario uncertainty* dependent on policy and human actions.
- *structural or epistemic uncertainty*, imperfect understanding.

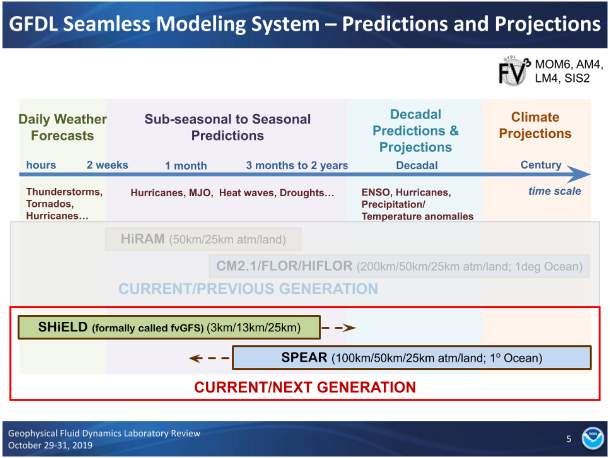
The premise of seamlessness is that the same model can be used for solving both initial value problems and boundary value problems, including counterfactual values.
From Hawkins and Sutton (2009).

Model configuration and calibration in a seamless modeling system



Equivalent predictability at lower cost in FLOR vs CM2.5.
Figure courtesy Gabe Vecchi, Princeton University.

Current generation GFDL models



Note: IPCC models GFDL-CM4 and ESM4 use higher ocean resolution (0.25°, 0.5°). Figure courtesy Tom Delworth, NOAA/GFDL.

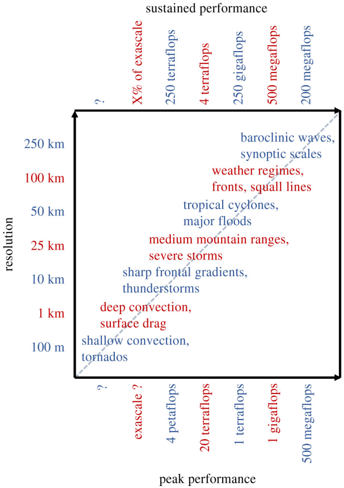
Outline

What does any of this have to do with ML?

- Models, even “seamless” ones, may be configured or calibrated differently for different problems (e.g forecast horizons).
- Each problem carries an implicit cost function by which a model configuration is declared suitable.
- Models do not converge cleanly with resolution: much unresolved physics is not yet “scale-aware”.
- Computation alone is not going to make the problem go away (see below...)
- Important new constraints on models from observations (new generation of satellites, Argo...)
- While data science is a misnomer (what is non-data science?) the convergence of computation and statistics that we call ML provides paths forward toward seamlessness: traceable hierarchies of scale.

What can we expect at an exaflop?

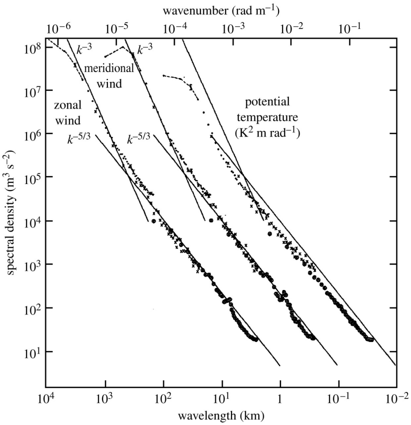
Will exascale be the rescue? Neumann et al (2019).



- ICON projects that a 1 km global model will run at 0.06 SYPD on "pre-exascale" technology: 17X improvement needed for 1 SYPD.
- This will be on 200,000 nodes (roughly 2xGaea).
- DECK: 1000 SY.
- A full suite of hindcasts for seasonal forecasting: 10,000 SY.

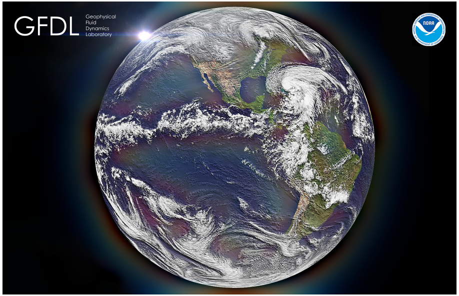
Hypothesis: vastly reduced uncertainty at 1 km.

No separation of "large" and "small" scales

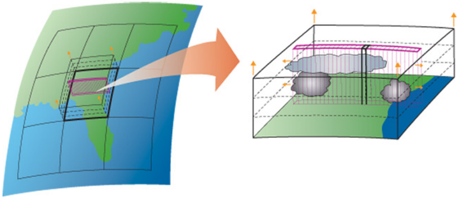


Nastrom and Gage (1985). More fidelity, more complexity over time in small scales ("physics"). The backscatter idea (Jansen and Held 2014) provides an energetically consistent framework for SGS.

Learn from short duration high resolution simulations



(Courtesy: S-J Lin, NOAA/GFDL).



(Courtesy: D. Randall, CSU; CMMAP).

- Global-scale CRMs (e.g 7 km simulation on the left) and even super-parameterization using embedded cloud models (right) remain prohibitively expensive.
- Can we learn the statistical aggregate of small scales? See Schneider et al 2017, Gentine et al (2018), O'Gorman and Dwyer (2018), Bolton and Zanna (2019), ...
- GFDL-Vulcan collaboration begun.

Model calibration

Model calibration or "tuning" consists of reducing overall model error (relative to some goal of modeling) by modifying parameters. In principle, minimizing some cost function:

$$C(p_1, p_2, ...) = \sum_{1}^{N} \omega_i \|\phi_i - \phi_i^{obs}\|$$

- Usually the p must be chosen within some observed or theoretical range $p_{min} \leq p \leq p_{max}$.
- "Fudge factors" (applying known wrong values) generally frowned upon (see Shackley et al 1999 on "flux adjustments".)
- The choice of ω_i is part of the lab's "culture". Cost also plays a role.
- The choice of ϕ_i^{obs} is also troublesome:
  - "Over-tuning": remember "reality" is but one ensemble member...
  - overlap between "tuning" metrics and "evaluation" metrics.

See for example, Hourdin et al (BAMS 2017)

Problems with parameter optimization

- Parametric uncertainty vs structural uncertainty.
- A two stage process: process-level constraints followed by global constraints.
- The choice of cost function.
- Metric weights and normalization.
- Do observations sample the space sufficiently?
- If models "higher" in the hierarchy are used for calibration, are they representative of all possible states? What the associated uncertainties?
- Internal feedbacks and compensating errors.

Danny Williamson (Exeter) has been arguing that we should look at it differently, as a problem of eliminating implausible regions of phase space rather than optimization.

Formulating the problem

$$\frac{\partial \mathbf{x}}{\partial t} = D(\mathbf{x}) + \sum_n \mathcal{P}_n(\mathbf{x}, \lambda_n)$$

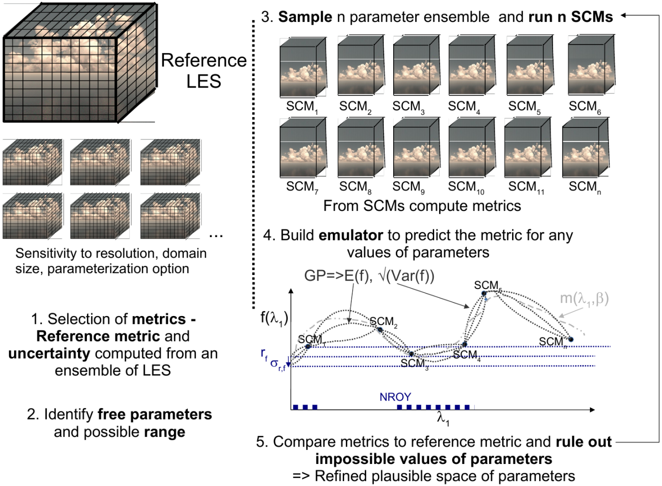
- Structure is given by $\mathcal{P}$, we are trying to calibrate values of a vector of parameters λ
- Multiple metrics we wish to satisfy. For each metric f we can define a distance given by:

$$I_f(\lambda) = \frac{\| r_f - E_f[\lambda] \|}{\sigma^2_{r,f} + \sigma^2_{d,f} + Var[f(\lambda)]}$$

- Euclidean distance over history normalized by error (observational, structural, chaotic)
- Sample λ space as exhaustively as practical for $I < T$, the NROY space. Iterate in *waves*. Can use different metrics in subsequent waves.

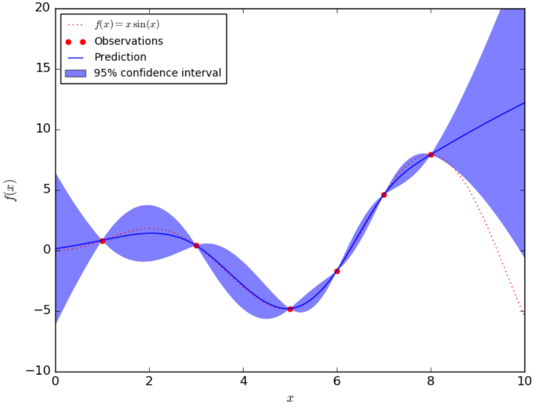
$$\mathrm{NROY}^n = \cap_k \mathrm{NROY}_{f_k}$$

Couvreux et al (2020)



- LES as ground truth, multiple variants to get "observational error".
- Emulate LES using SCMs encoding all the $\mathcal{P}$.
- Latin hypercube sampling of λ
- Fit Gaussian processes to SCMs to densely sample all values of λ

Gaussian processes

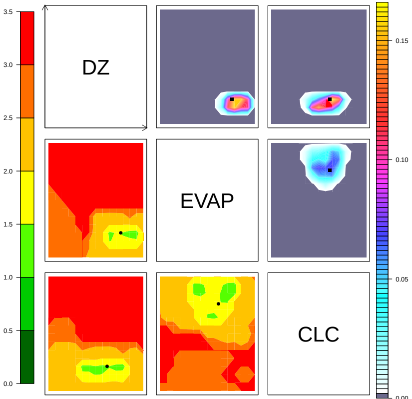


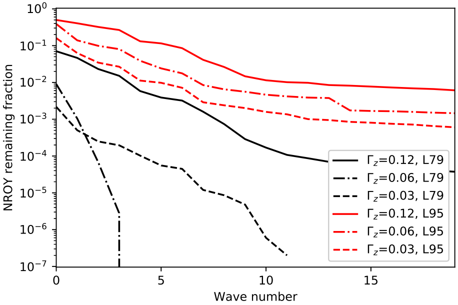
- Extremely standard emulator, widely available in python libraries
- Very poor at extrapolation, so training data must span phase space!

Couvreux et al 2020, some highlights

- Importance of a library of distinct physical regimes (e.g marine, continental) sampled by LES
- Results are sensitive to LES turbulence closure and numerics.
- Don't do sensitivity analysis on the full phase space (premise is that most of it is unphysical). But see discussion of order of imposition of metrics.
- Even individual $\mathcal{P}$ may have multiple tunable subsystems with compensating errors, e.g EDMF.
- Rule of thumb: need $10 \times \mathrm{rank}(\lambda)$ SCM runs.

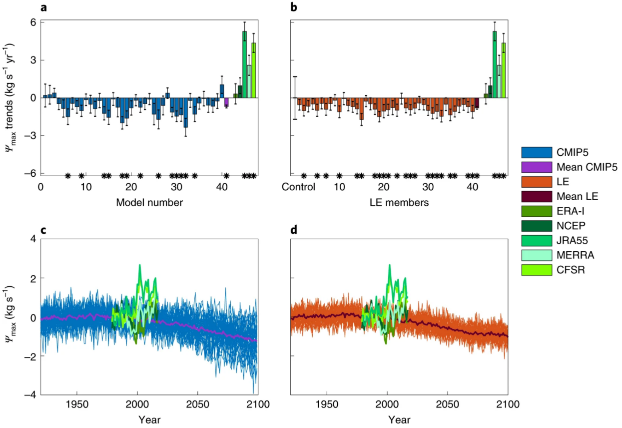
Hourdin et al (2020)





- Eliminate implausible parameter space comparing SCMs with LES.
- ... leaving irreducible ("structural") model error.

Training on models or observations?



From Chemke and Polvani (2019). Hadley cell strength is likely correct in models and not in “observations”!

Outline

Using ML in seamless modeling: ideas and challenges

- Big data, machine learning, AI: not a step change but massive computation applied to existing methods (regression, classification, assimilation)
- Seamless models may be a hierarchy of scale, resolution, cost: ML-inspired emulators help navigate the hierarchy.
- Models are calibrated in multiple stages: ML can play a role at process-level as well as global constraints.
- Training data may come from model hierarchy (e.g CRM, LES) or observations and reanalysis (even for directly predictive methods, e.g Ham et al, *Nature* 2019)).
- Ensure training data and ML methods are well-anchored in theory (see next talk by Maike Sonnewald).
- Fundamental questions still unanswered:
  - How much physics should be learnt?
  - Can we assume a structure for the physical formulation?
  - Process fidelity vs overall model error.
  - Will ML give us differentiable models?

Bibliography

- Couvreux et al 2020: Process-based climate model development harnessing machine learning: I. a calibration tool for parameterization improvement.
- Hourdin et al 2020: Process-based climate model development harnessing machine learning: II. model calibration from single column to global
- Hourdin et al 2017: The art and science of climate model tuning.
- Balaji 2020: Climbing down Charney's ladder: Machine Learning and the post-Dennard era of computational climate science
- Williamson et al 2013: History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble
- Williamson et al 2017: Tuning without over-tuning: parametric uncertainty quantification for the NEMO ocean model