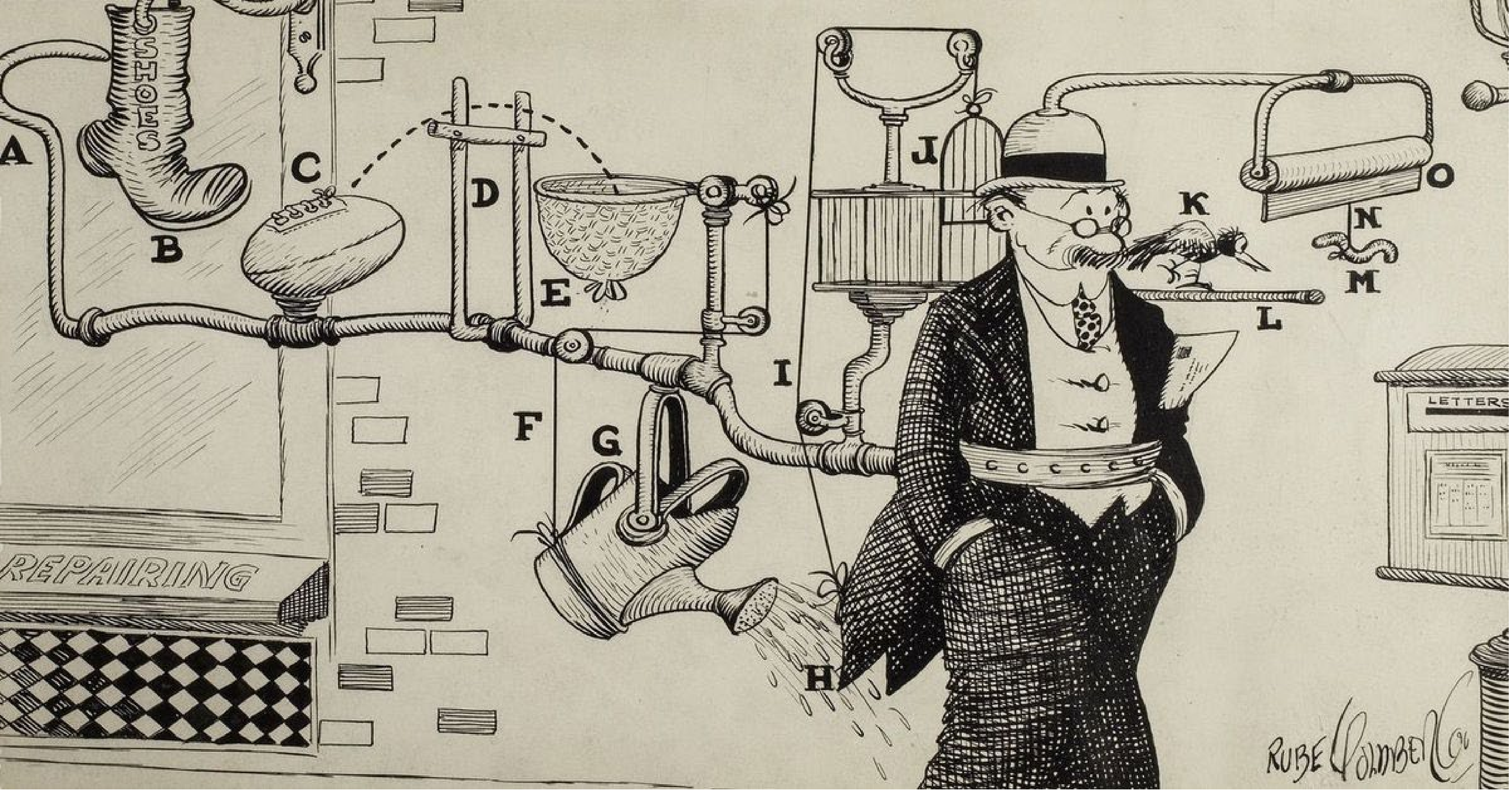
Earth & Environmental Systems Modeling

3. Methods in Model Integration, Hierarchical Modeling, Model Complexity

Ruby Leung (PNNL; ruby.leung@pnnl.gov) and Jennie Rice (PNNL; jennie.rice@pnnl.gov)
Topic Leads: Rob Jacob, Ian Kraucunas, Jonathan Lamontagne, Jian Lu, Brian O'Neill, Scott Painter, Claudia Tebaldi, Chris Vernon

U.S. DEPARTMENT OF ENERGY | Office of Science

2024 EESM PI Meeting
August 6–9, 2024


SHOES
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
REPAIRING
LETTERS
RUBE GOLDBERG

What are the current and unique strengths and foundational capabilities of DOE for this topic?

- Large range of types of model complexity, including sectoral, processes, resolution, coupling, computational frameworks
- Cross-cutting expertise: universities, labs, multiple domains
- System-level thinking (to bring different disciplines to a problem and develop a broader perspective)
- Connection to observational programs and facilities and staff to inform modeling
- Capabilities in leveraging mathematical and computational expertise and infrastructure (e.g., UC/UQ)
- MSD human-natural system integration

What are the gaps in research / infrastructure / coordination that prevent advances?

All of these are general gaps for complex modeling, not specific to DOE

- Terminology issues
  - Same term can mean different things in different domains, and vice versa
- Lack of a theoretical framework to determine ideal model coupling strategies
- Reusability of approaches, tools, frameworks, etc.
- Different modeling paradigms–lack of awareness and understanding
- Staffing and Training
  - Software engineers
  - Multisectoral expertise; interdisciplinary training
  - AI/ML with domain science knowledge
- Incentives and Resources
  - Need alignment with integrated modeling needs

What opportunities exist to overcome each of those gaps?

- CoP for EESM (build on MSD CoP)
  - Sharing best practices
  - Develop common solutions
  - Access to data and models (heard in other breakouts as well)
- Develop approaches to best leverage AI/ML
  - Emulation
  - Propagation of uncertainty and identifying sensitivities
  - Identifying optimal model complexity
  - Partner more broadly, including industry and other agencies
- It's important to continue with hi-res models and the necessary new processes even if we lack the necessary constraints because it provides the tools to improve our understanding of the system
- Additional stakeholder engagement to inform modeling choices and ensure societal relevance

Spin up of E3SM – slow process/many processes across time scales – deeper ocean
Time needed is a limiting factor – evenn 2000 years still drifting ocean heat content

key idea: rely on uncoupled ocean simulation (some questions about what the uncoupled configuration looks like in terms of atmospheric forcings)
Alternate coupled and uncoupled then start piControl, compare to standard spinup
For drifting quantities and climatology

Achieve 40-50%

In general well designed simulation approach can improve efficiency (maybe adaptive design can be invoked; maybe the idea of different time steps, like used in tracers spinup can be used)
Possibly Data Assimilation can help further

IM3 integrated modeling

Many models coupled – the crux is in the arrows

Challenges: spatio-temporal scales are different
Models not built to talk to each other; processes may be called the same but be spurious between models

Sociology of interdisciplinarity (language, institutional cultures)
Time resources for QA/QC
Open science requirements…

Uncertainty quantification/cascades; computational challenge
Prioritization of uncertainty should be according to the relevance of the outcome of the modeling

If we have a loss function (societally relevance) dependent on modeling outcomes we can prioritize sources
Not just about computational challenges but actually understanding of sources and processes still difficult, maybe ML/AI can help.

Hard/soft coupling should be guided by sensitivity analysis for each model component (systematic approach to it) in order to decide if hard coupling is warranted.

IM3 (cont'd) example of GCAM and power siting – harmonization/units/scale/information buried

Need for multisectoral translator
Need a culture of team work
Need homogenization of languages
Experts of QA/Qc
Multisectoral review

Leader responsibility to enforce standard, documentation; accountability; willingness to get into the weeds; work to translate the science question in practical steps; hands on management

Sponsor sustained funding is key and is hard to get (~4/5 years at least)

Flash Talk : River-Ocean two-way coupling (ICOM)

Compound flooding (= coastal extremes)simulations

E3SM v2 to simulate fine scale processes in coastal regions

Variable resolution regional refined meshes for the different model components (lnd/ocean/atm/river)

Extreme TC in the Mid_Atlantic region, improved river discharge/sea surface height simulations/representation of water discharge driven by tides

Could storyline approach be used similarly to what was done in Sherwood et al., start from the outcome and ask what should happen to generate the outcome of concern

How important is to be a multidisciplinary (multimodel) expert? Recognizing the importance of being a domain expert for career development, the more you can branch out the more beneficial for the project.

Slab ocean model

To simplify the ocean component (reaches equilibrium within 20 years)
Useful to address coupled model questions

SOM may be used for model tuning
SOM responding to wind to

Develop a more interactive E3SM-SOM

Hyperfacets – model coupling and stakeholder engagement

Engaging stakeholders may guide use of model hierarchies/linkages

Really focusing on linkages (Climate model output and impact models)
Perspective is general, not for a specific problem to solve

Many models for the same purpose (climate models/ ML vs. statistic)
Mismatch of scales
Uncertainty cascade
Communication of uncertainties

Climate Water Energy system linkages

Matching one model scale to the other

Difference in response from different climate models

Adding climate change changes costs/partition of renewables for decarbonization

Process representation using AI/ML

Using AI/ML for modelparameterization

ML to improve rain formulation results
Embedded NN in the microphysics
Solves the drizzling problem

Looking at portability of NN across models

Hierarchy of modeling but also in the sense of using observations at very fine scales for parametrizations at the grid scale

AI can simplify the model hierarchy for domain science– emulate pieces from different models/mix and match across domains and scales

Foundation models quick/fast
Need to be supplemented by other models for other processes.

Need to partner with industry for synergy

Need observations to train these models

Gaps/challenge: robustness uncertainty interpretabilty

AMOC Scidac project

Coupled phenomenon / emergent over hundred of years

Forced Ocean Sea Ice simulations: have limited feedbacks but shorter emergence time scales

Still difficult to run more than 20-30 ensembles – trying touse ML (but needs lots of training data)

Trade off about the dynamics prioritized.

What do we mean by hierarchy? Nested, coupled versus loosely coupled, fidelity/complexity - different speakers have different concepts.

All couplings are custom one-off. Are there general abstraction that we can make to help with coupling?

Structural issues:

Technical professional to help with software.

Training to think in interdisciplinary and system-focused ways.

Inconsistent use of terminology around coupling and hierarchy.

Definition and approach depends on the application.

Common examples of hierarchies - are there a few representative cases that might lead to a small number of design patterns?

Communities of practice useful - like the MSD community of practice - help define linkages.

Need ability to think beyond the next paper.

MSD community doesn't necessarily think explicitly about hierarchy (or use the terminology).

Emulators of closed models may be one transferable solution.

In terms of learning from other communities, may also consider communities adjacent to Earth System modeling (e.g. watershed)

Complexity

Hierarchical machine learning - pattern effect?

Food in GCAM: aggregation/grouping into small number of primary types with similar supply chains.

Energy in GCAM: importance of common language, trade-off complexity with science goal, eventually translate detailed models into statistical relationships?

What do social scientists bring to the hierarchical modeling? Systematizing set of challenges and solutions among different communities.

Shared bibliography possibly constructed by AI (like community of practice in MSD).

When do we need process-based versus statistical? How do we determine the appropriate level of complexity? Non-stationarity, ability to validate,

Too much complexity can lead to larger uncertainty because of difficulty to constrain.

5 or 10 year challenge: Emulators as framework for analyzing multimodal complexity.

Another solution: Theoretical frameworks useful. Standardizing and system

Improving resolution and adding complexity requires higher demands on data for validation and informing parameters. Process oriented (physical) models still needs to be fostered for the new models for the processes that are not properly constrained. Can still learn from high-resolution simulations even when they are not adequately constrained.

Access to tools