Machine learning models to improve parameterizations of climate models

Pierre Gentine – Columbia University
Mike Pritchard, Tom Beucler, Stephan Rasp, Wenli Zhao

Clivar August 2019
1. Global carbon uptake

Large intermodel spread, large interannual variability
Define concentrations $[CO_2] = f(\text{emission flux})$
2. Climate sensitivity

Still substantial spread in model climate sensitivity

global $T = f(\text{greenhouse gases})$:
Limits our climate mitigation and management capacity and increases cost

Mostly due to **clouds**

![Diagram showing relationship between ECS, allowable CO₂ concentration, and year of crossing 2°C threshold.](image)

**ECS** = Equilibrium climate sensitivity (T response do CO₂ doubling)

Schneider, T., et al. (2017). *Nature Climate Change*
3. Regional climate sensitivity

Cloud impact is not just global but also regional
(also circulation feedback)

Aquaplanet
+4K
(no SST feedback!)

Regional climate prediction is too uncertain
**1. Using ML for climate: (deep) clouds**

**Parameterization:** represent (physically or statistically) a physical process that cannot be resolved (e.g. clouds)

Typically physically based

\[
\frac{\partial \bar{X}}{\partial t} \bigg|_{\text{clouds}} = f(\bar{X}) \quad \text{with} \quad \bar{X} \quad \text{coarse-scale average of} \quad X
\]

However: it has failed for ~40 years (Randall et al. 2003)

This largely explains intermodal spread in climate prediction
1. Using ML for climate: (deep) clouds

How can we solve this issue?
Take advantage of **cloud-resolving simulations** (~1km, **alleviate most biases** but very expensive).

Not "physical" but **Data-driven approach**
(informed by cloud-resolving simulations)

**Cost function:** misfit to coarse-grained high-res. model

- Temperature $\bar{T}(z)$
- Specific humidity $\bar{q}(z)$
- Surface sensible heat flux $H$
- Surface evaporation $E$
- Surface pressure $P_s$
- Precipitation

**Deep Neural Net** or Convolutional NN

Global “CRM”

Coarse graining
1. Using ML for climate: (deep) clouds

Coarse-grained
Cloud-resolving Model
(superparameterization)

Machine learning
Coarse-resolution model

10 times cheaper than original coarse model, 1000 less expensive than high-res model

**Question:** generalization to unforeseen conditions? Climate change

- Gentine P., Pritchard M., Rasp S., Reinaudi G., GRL, 2018
- Rasp, Pritchard and Gentine, PNAS 2018
- Brenowitz and Bretherton, GRL 2018
1. Using ML for climate: (deep) clouds

Interactive model

Now we have good boundary condition to study hydrology 😊

10 times cheaper than original coarse model, 1000 less expensive than high-res model

Question: generalization to unforeseen conditions? Climate change: Poor!

Gentine P., Pritchard M., Rasp S., Reinaudi G., GRL, 2018
Rasp, Pritchard and Gentine, PNAS 2018

Regular CAM parameterization
Issues

1. **Physical Constraints**
   - Energy conservation
   - Mass conservation
   - Only *approximate with ML*

Gentine P., Pritchard M., Rasp S., Reinaudi G., GRL, 2018
Rasp, Pritchard and Gentine, PNAS 2018
2. **Extrapolation**

ML has mostly been about interpolations using lots of data, poor extrapolation

![Graph showing equatorial contraction and poleward upward movements](image-url)
2. **Extrapolation**
ML has mostly been about interpolations using lots of data, poor extrapolation.
Constraining physics within ML

1. Convection

Energy and mass conservations

Impose them within NN as function of inputs \( x \) and outputs \( y \):

\[ \begin{cases} c \mathbf{u} = 0 \\ \theta = 0 \end{cases} \]

Exact conservations

2 equations, reduce NN degrees of freedom to \( n-2 \) degrees of freedom

Beucler, Pritchard, Rasp, Gentine, PRL, submitted

Hybrid approaches
Hybrid approaches

Constraining physics within ML

1. Convection

Energy and mass conservations

Unconstrained

Constrained

Current climate

Future climate

Constrained physics + improved generalization 😊
Hybrid approaches

Constraining physics within ML

2. Land surface latent heat flux (LE)

Objective: predict LE from environmental variables

- Pure ML (feedforward NN) performs well
- But does not conserve surface energy budget

$$R_n - G \neq H + LE$$
Hybrid approaches

Constraining physics within ML

2. Land surface latent heat flux (LE)

Objective: predict LE + conserve energy + respect diffusion

\[ LE = \rho \frac{e_s - e_a}{r_s + r_a} \]

○ Hybrid ML performs as well as pure ML

○ Conserves surface energy budget 😊

\[ R_n - G = H + LE \]
Hybrid approaches

Constraining physics within ML

2. Land surface latent heat flux (LE)

Out-of-sample generalization/extremes

Test 1 and 99 percentiles

Hybrid systematically outperforms pure ML for extremes 😊
Conclusions

Machine learning is an appealing approach for subgrid parameterizations

Two working examples
1. Deep clouds
2. Land surface processes (evapotranspiration)

Issues:
1. Conservations, physical invariances, physical laws
2. Generalization

Solution:
Hybrid physical+ML approaches appear as powerful tools to tackle this
THANK YOU

Questions?