Some Thoughts on Using Machine Learning for Cumulus Parameterization

Zhiming Kuang
Scope

• Construct cumulus parameterizations from cloud resolving model simulations through machine learning.

• This is not a substitute for global cloud resolving models

• But it is still an interesting and useful problem given the success of super-parameterization.
Potential difficulties

- Intrinsic noise
- Memory
- Coupling with large-scale
- High dimensionality
Noise

- There are intrinsic noises in convective tendencies from CRM simulations.

Figure from Jones and Randall (2011)

For a 256km x 256km domain (dx=dy=2km) subject to constant forcing
Noise

• Care must be taken to ensure correct direction of causation.
\[ y_n = -\lambda x_n + \sigma_n \]

\[ x_{n+1} = x_n + y_n \Delta t \]

\[ \langle x, y \rangle = \frac{-1}{\langle x, x \rangle} \]

\[ \lambda = 1 \]

\[ \Delta t = 1 \]

\( x \) is the state variable

\( y \) is the convective tendency

\( n \) indicates the \( n \)th time step

\( x \) is uncorrelated with the noise
Do a two-point average (sum)

\[ X_n \equiv x_{2n-1} + x_{2n} \]
\[ Y_n \equiv y_{2n-1} + y_{2n} \]

\[
\frac{\langle X, Y \rangle}{\langle X, X \rangle} = -\frac{1}{2}
\]

\(X\) is no longer uncorrelated with the noise.
Noise

• Care must be taken to ensure correct direction of causation.

• For example, using hourly averages as in Krasnopolsy et al. (2013) will confuse the direction of causation, as noise can now influence the state vector. This influence will be interpreted incorrectly as the dependence of convective tendencies on the state vector.

• The use of instantaneous values can also be complicated by the presence of memory (in the noises).
Coupling with large-scale dynamics

• In a typical machine learning problem, minimizing the prediction error for the process in question is the end goal.

• Here, convection is a sub-component. Errors in convective tendencies along different directions can be amplified or damped by coupling.

• This makes the problem distinct from a typical machine learning problem.
Convectively coupled waves as an example

Space-time spectra (averaged over 15N-15S) (with background red noise removed)  
After Wheeler and Kiladis, J. Atmos. Sci., 1999
A linear model of convectively coupled waves

\[
\frac{d}{dt}\begin{pmatrix} \vec{x} \\ \vec{w} \end{pmatrix} = \begin{pmatrix} M & A \\ k^2C & D \end{pmatrix}\begin{pmatrix} \vec{x} \\ \vec{w} \end{pmatrix} + \vec{f}
\]

x: thermodynamic profiles (temperature and humidity)
w: vertical velocity profiles
f: forcing
k: horizontal wavenumber
M: thermodynamic linear response function for convection
A: effect of vertical advection on temperature and humidity
C: effect of temperature and humidity on vertical velocity
D: momentum damping
Reduce convection to 4 degrees of freedom

These are not set \textit{a priori}, and come from the order reduction procedure.
Learn this small linear system

- Stable linear system driven by random noise
- No noise in convective tendencies
- 4 degrees of freedom for convection (6 overall)
- A random forest approach (10 trees, 5 leaves)
- 1.2e5 training samples

- **Test set $R^2$:**
  - Lower troposphere temperature: 0.97
  - Upper troposphere temperature: 0.97
  - Free troposphere temperature: 0.99
  - Boundary layer moist static energy: 0.96
Spectrum from coupled run with the true matrix
True matrix replaced by random forest
Time series
Convective heating of the free troposphere

- blue: changing free troposphere temperature
- red: changing boundary layer MSE
- circles: random forest lines: true model
Binary decision tree

Figure from Hastie, Tibshirani and Friedman
Free troposphere temperature

Boundary layer MSE
Free troposphere temperature

Boundary layer MSE
Free troposphere temperature

Boundary layer MSE
True matrix replaced by random forest

The spurious peak corresponds to the dry wave for this wavelength
Higher dimensionality

• Use the full linear response function for convection (46 degrees of freedom instead of 4)
• 2.4e6 training samples
A more realistic example

- A cyclic CRM (512km x 512km) coupled with 2D linear gravity wave (horizontal wavelength of 5000km).
- Long period (~4 days) makes the memory issue less acute.
Reasonable test error

• 70000 training samples

Random forest
(10 trees, 5 leaves)

Neural network with
40 hidden nodes
Big errors when coupled

Random forest
(10 trees, 5 leaves)

Neural network with
40 hidden nodes

Period 1.3 days, phase speed
45m/s
The coupled system can behave differently even when test set errors are small for two reasons:

1. Coupling can selectively amplify or damp errors, which makes learning based on minimizing those errors less effective.

2. Erroneous behaviors can arise as the system travels to the boundaries of the training set.

The error function needs to take coupling into consideration.
“Active” learning?

• Unlike in typical machine learning problem, we can make new CRM runs as needed.

• We can build linear response functions for different mean states and stitch them together using e.g. a decision tree, and also build new ones on demand
  • This should be more efficient than a brute-force database (look-up table) of the state vectors and the tendencies.
Concluding remarks

• There are a number of issues to work through in order to learn a cumulus parameterization from CRM simulations:
  • Intrinsic noise
  • Memory (not explored here)
  • **Coupling with large-scale**
  • High dimensionality
• Some “active” learning is likely needed.
Model order reduction from control theory

Full linear response function

Reduced model
4 T&q modes with structural constraints
2 w modes
\[ X_n \equiv x_{2n-1} + x_{2n} \]
\[ Y_n \equiv y_{2n-1} + y_{2n} \]

\[ X_{n+1} = (2 - \lambda \Delta t) x_{2n-1} + \sigma_{2n-1} \Delta t \]

\[ Y_{n+1} = -\lambda X_{n+1} + \sigma_{2n-1} + \sigma_{2n} \]

\[ \frac{\langle X, Y \rangle}{\langle X, X \rangle} = -\frac{1}{2} \]

\[ X \text{ is no longer uncorrelated with the noise} \]