Hybridization of Physics-Based Modeling with Machine Learning in Numerical Weather/Climate Modeling Systems*

V. Krasnopolsky
NOAA/NWS/NCEP/EMC

Acknowledgments: A. Belochitski, M. Fox-Rabinovitz, H. Tolman, Y. Fan, J-H. Alves, R. Campos

* By no means this overview should be considered comprehensive.
Abstract

Numerical Weather and Climate Modeling Systems (NWMS) and related fields have been using Machine Learning (ML) for 25+ years.

Wide spectrum of physically based + ML hybrid approaches has been developed in this field.

Our current plans for using ML are build on the solid basis of our community previous experience with ML in Weather and Climate Modeling and related fields.

ML is a toolbox of versatile nonlinear statistical tools.

ML can solve or alleviate many problems but not any problem;
ML has a very broad but limited domain of application.
Outline

• I. Machin Learning
• II. Physically Based Modeling
• III. Hybrid Approach
• VI. Several examples
Spectrum of Hybridization

- ML Forecast Model Learned from Observations
- GFS with ML Dynamics
- GFS with ML Parameterization
- GFS with ML post-processing
- ML Emulation of GFS
- GFS with ML Physics
- GFS with multiple ML parts
- GFS
What is ML?

- ML is a subset of artificial intelligence (AI)
- ML algorithms build mathematical/statistical models based on training data - ML is Learning from Data Approach
- ML toolbox includes among other tools:
  - Artificial Neural Networks (ANN or NN)
  - Support Vector Machines (SVM)
  - Decision Trees
  - Bayesian networks
  - Genetic algorithms
  - DNN
  - CNN
  - RNN
  - …
Mapping

• **Mapping:** A *rule of correspondence established between two vectors* that associates each vector \(X\) of a vector space \(\mathbb{R}^n\) with a vector \(Y\) of another vector space \(\mathbb{R}^m\)

\[
Y = F(X) \\
X = \{x_1, x_2, \ldots, x_n\} \in \mathbb{R}^n \\
Y = \{y_1, y_2, \ldots, y_m\} \in \mathbb{R}^m
\]

\[
\begin{bmatrix}
y_1 = f_1(x_1, x_2, \ldots, x_n) \\
y_2 = f_2(x_1, x_2, \ldots, x_n) \\
\vdots \\
y_m = f_m(x_1, x_2, \ldots, x_n)
\end{bmatrix}
\]

**ML tools:** NNs, Support Vector Machines, Decision Trees, etc. are generic tools to approximate complex, nonlinear, multidimensional mappings.
**Continuous Input to Output Mapping**

**Multilayer Perceptron: Feed Forward, Fully Connected**

\[ Y = F_{NN}(X) \]

where:

1. **Input Layer**: \( X = [x_1, x_2, x_3, \ldots, x_n] \)
2. **Hidden Layer**: \( t_1, t_2, \ldots, t_k \)
3. **Output Layer**: \( y_1, y_2, \ldots, y_m \)

**Linear Part**

\[ a_j \cdot X + b_j = s_j \]

**Nonlinear Part**

\[ \phi(s_j) = t_j \]

\[ t_j = \phi(b_{j0} + \sum_{i=1}^{n} b_{ji} \cdot x_i) = \]

\[ = \tanh(b_{j0} + \sum_{i=1}^{n} b_{ji} \cdot x_i) \]

\[ y_q = a_{q0} + \sum_{j=1}^{k} a_{qj} \cdot t_j = a_{q0} + \sum_{j=1}^{k} a_{qj} \cdot \phi(b_{j0} + \sum_{i=1}^{n} b_{ji} \cdot x_i) = \]

\[ = a_{q0} + \sum_{j=1}^{k} a_{qj} \cdot \tanh(b_{j0} + \sum_{i=1}^{n} b_{ji} \cdot x_i); \quad q = 1, 2, \ldots, m \]
ML Models

• **Approach:**
  – Train ML model, using training set
  – Validate ML model, using independent validation set

• **Advantages:**
  – Approach backed by math (shallow NN approximates any mapping)
  – Relatively simple approach and model
  – Requires only data set for development
  – Fast model (after ML is trained)

• **Limitations:**
  – Too simple to model complex multiscale systems
  – Completely depends on data
  – Data are usually sparse in space, time, and scales
  – Generalization (extrapolation as well as interpolation) may be unstable

Aug 27, 2020  ML for Weather and Climate
General Circulation Model

The set of conservation laws (mass, energy, momentum, water vapor, ozone, etc.)

• **First Principles/Prediction 3-D Equations on the Sphere:**

\[
\frac{\partial \psi}{\partial t} = D(\psi, x) + P(\psi, x)
\]

- \( \psi \) - a 3-D prognostic/dependent variable, e.g., temperature
- \( X \) - a 3-D independent variable: \( x, y, z \) & \( t \)
- \( D \) - dynamics (spectral or gridpoint) – resolve physics
- \( P \) - physics or parameterization of subgrid physical processes

• **Continuity Equation**
• **Thermodynamic Equation**
• **Momentum Equations**
General Circulation Model

*Physics – \( P \), represented by "parameterized" or simplified 1-D (vertical) schemes*

- Major components of \( P = \{M, R, S, T\} \):
  - \( M \) – precipitation (moisture) processes
  - \( R \) - radiation (clouds + long & short wave processes)
  - \( S \) – surface model (land, ocean, ice – air interaction)
  - \( T \) – turbulent mixing (planetary boundary layer parameterization, vertical diffusion, and gravity wave drag)

- Each component of \( P \) is a 1-D parameterization of a very complicated set of multi-scale theoretical and empirical physical process models simplified for computational reasons

- Even after simplification, \( P \) is still *time consuming*

- The model physics components (or entire \( P \)) are very appropriate candidates for ML modeling
Resolution Challenge

Doubling of resolution requires 8X more processors

affordable power limit

8X more processors

[ECMWF, Bauer et al. 2015]

ML response to the challenge: Speed up model calculations
Model Physics Challenge

- With increased resolution, scales of subgrid processes become smaller and smaller
- Subgrid processes have to be parameterized
- Physics of these processes is usually more complex
- The parametrizations are complex and slow

**ML response to the challenge:** Speed up calculations via developing fast ML emulations of existing parameterizations and developing fast new ML parameterizations
Spectrum of Hybridization

- ML Forecast Model Learned from Observations
- GFS with ML Dynamics
- GFS with ML Parameterization
- GFS with ML post-processing
- ML Emulation of GFS
- GFS with ML Physics
- GFS with multiple ML parts
- GFS
Different Hybridization to Improve Numerical Weather/Climate Modeling Systems

- Better ML Retrieval Algorithms
- Fast ML Forward Models
- ML Observation Operators
- Fast ML Radiation
- Fast & Better Microphysics
- New ML Parameterizations
- Fast ML Physics
- ML Bias Corrections
- Nonlinear MOS
- Nonlinear ML ensemble averaging
Several Examples of Hybrid Approach Application in NWMS
Ingesting Satellite Data in DAS

- **Satellite Retrievals:**
  \[ G = f(S), \]
  - \( S \) – vector of satellite measurements;
  - \( G \) – vector of geophysical parameters;
  - \( f \) – transfer function or retrieval algorithm

- **Direct Assimilation of Satellite Data:**
  \[ S = F(G), \]
  - \( F \) – forward model

- Both \( F \) & \( f \) are mappings and NN can be used
  - Fast and accurate NN retrieval algorithms \( f_{NN} \)
  - Fast NN forward models \( F_{NN} \) for direct assimilation
SSM/I Wind Speed Satellite Retrievals

Wind speed fields retrieved from the SSM/I measurements for a mid-latitude storm. Two passes (one ascending and one descending) are shown in each panel. Each panel shows the wind speeds retrieved by (left to right) GSW (linear regression) and NN algorithms. The GSW algorithm does not produce reliable retrievals in the areas with high level of moisture (white areas). NN algorithm produces reliable and accurate high winds under the high level of moisture. 1 knot ≈ 0.514 m/s

Regression algorithm: Confuses high levels of moisture with high wind speeds

NN algorithm: Correctly retrieves the mostly energetic part of wind speed field

**NN Emulations of Parameterizations: The Magic of NN Performance (LWR)**

**Input/Output Dependency:** \( Y = F(X) \)

**Numerical Scheme for Solving Equations**

**Mathematical Representation of Physical Processes**

\[
\begin{align*}
F^+ (p) &= B(p) \cdot \varepsilon (p, p) + \int \alpha (p, p') \cdot dB(p') \\
F^- (p) &= R(p) - \int \alpha (p, p') \cdot dB(p') \\
B(p) &= \sigma \cdot T^4 (p) \quad \text{— the Stefan–Boltzmann relation} \\
\alpha (p, p') &= \frac{\int \frac{dB(p')}{dT(p')} \cdot (1 - \tau(p, p')) \cdot dv}{dB(p)/dT(p)} \\
\varepsilon (p, p') &= \frac{\int B(p') \cdot (1 - \tau(p, p')) \cdot dv}{B(p)} \\
\tau(p, p) &= \text{the Planck function}
\end{align*}
\]

**NN Emulation of Input/Output Dependency:**

\[
y_q = a_{q_0} + \sum_{j=1}^{k} a_{q_j} \cdot \tanh(b_{j_0} + \sum_{i=1}^{n} b_{j_i} x_i), \quad q = 1, \ldots, m
\]

**Input/Output Dependency:** \( \{X_i, Y_i\}_{i=1}^{N} \)

Aug 27, 2020 ML for Weather and Climate
Accurate and fast neural network (NN) emulations of long- and short-wave radiation parameterizations in NCEP GFS/CFS

- Neural Networks perform radiative transfer calculations *much faster* than the RRTMG LWR and SWR parameterizations they emulate:

<table>
<thead>
<tr>
<th></th>
<th>RRTMG LWR</th>
<th>RRTMG SWR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average Speed Up by NN, times</strong></td>
<td>16</td>
<td>60</td>
</tr>
<tr>
<td><strong>Cloudy Column Speed Up by NN, times</strong></td>
<td>20</td>
<td>88</td>
</tr>
</tbody>
</table>

- As a result of the speed up, GFS with NN radiation calculated with the same frequency as the rest of the model physics, or 12 times per model hour, takes up as much time as GFS with RRTMG radiation calculated only once per model hour.

- Neural network emulations are *unbiased* and affect model evolution only as much as round off errors (see next slide).

Individual LWR Heating Rates Profiles

Profile complexity

Blue improves upon red

Black – Original Parameterization
Red – NN with 100 neurons
Blue – NN with 150 neurons

PRMSE = 0.18 & 0.10 K/day
PRMSE = 0.11 & 0.06 K/day
PRMSE = 0.05 & 0.04 K/day

Aug 27, 2020
ML for Weather and Climate
CTL run with RRTMG LW and SW radiations

NN run with NN LW and SW radiations

Differences between two control runs with different versions of FORTRAN compiler

NCEP CFS PRATE – 17-year parallel

JJA runs

NN – CTL run differences
NN Full Suite of Atmospheric Physics

Atmospheric Physics Suite (GFS v16, C96L64):

- LW Radiation,
- SW Radiation,
- Planetary BL,
- Orographic and convective gravity wave drag,
- Deep convection,
- Shallow convection,
- Microphysics,
- \( \text{CO}_2(t) \), trace gases,
- Aerosols (tropo- and stratospheric),
- \( \text{O}_3 \) and \( \text{H}_2\text{O} \) photochemistry

**Training:** Data simulated by 24 10-day GFS v.16 forecasts, uniformly covering entire 2018 (radiation was calculated at each physics time step!).

**NN**
- 522 Inputs
- 304 Outputs
- 250 Hidden Neurons in one hidden layer
- 3 times faster

Validation: 24 parallel runs (10-day GFS v.16 forecasts each), uniformly covering entire 2018

In all 24 runs no signs of instability were observed!
Calculating Ensemble Mean

• Conservative ensemble (standard):
  \[ EM = \frac{1}{N} \sum_{i=1}^{N} p_i, \quad p_i \text{ is an ensemble member} \]

• If past data are available, a nonlinear ensemble mean can be introduced:
  \[ NEM = f(P) \approx NN(P) \]
  \[ P = \{ p_1, p_2, \ldots, p_N \} \]

• NN is trained on past data
NN wind-wave model ensemble (buoy data)

- Black: ensemble members
- Red: conservative ensemble mean (EM)
- Cyan: control run
- Green: NN ensemble (NEM)

NCEP Global Wave Ensemble System
21 ensemble members

Normalized bias (NBias) for GWES ensemble mean (EM, top), and for NN ensemble mean (bottom) on an independent test set. The columns represent U10 (left) and Hs (right). Red indicates overestimation of the model compared to altimeter observations while blue indicates underestimation. Great part of large-scale biases in the mid- to high-latitudes has been eliminated by the NN ensemble mean simulation.

Neural Network Improves CFS Week 3-4 2 Meter Air Temperature Forecasts

Y. Fan, et al., 2019: Using Artificial Neural Networks to Improve CFS Week 3-4 Precipitation and 2 Meter Air Temperature Forecasts, submitted
Hybrid Approach (HA)

• Approach:
  – Train ML component, using training set
  – Test ML component, using independent test set
  – Validate ML component in the hybrid model, running parallel runs

• Advantages:
  – HA uses simulated or mixed data (less noise, less sparse)
  – HA speeds up model calculations => higher resolution, more ensemble members
  – Can learn not well understood physics from data => better physics

• Limitations:
  – Accuracy of HA depends on accuracy of the ML component
  – Accuracy of the ML component depends on representativeness of data
  – If ML component is not sufficiently accurate, generalization (extrapolation as well as interpolation) may be unstable
I. ML for Model Initialization

• Developed NN Applications (examples)
  
  – *Satellite Retrievals*
    
    • Fast ML retrieval algorithms based on inversion of fast ML emulations of RT models
      
    
    • ML empirical (based on data) retrieval algorithms
      
  
  – *Direct Assimilation*
    
    • ML fast forward models
      
  
  – *Assimilation of surface observations and chemical and biological observations*
    
    • ML empirical biological model for ocean color
      
    
    • ML algorithm to fill gaps in ocean color fields
      
II. ML for Numerical Model

ML Applications developed & under development

- Fast and accurate ML emulations of model physics
  - Fast NN nonlinear wave-wave interaction for WaveWatch model
  - Fast NN long and short wave radiation for NCEP CFS, GFS, and FV3GFS models
  - Fast NN emulation of super-parameterization (CRM in MMF)
  - Fast NN PBL
  - New ML parameterizations
    - NN convection parameterization for GCM learned by NN from CRM simulated data
    - ML emulation of simplified GCM
III. ML for Post-processing

ML Applications Developed

- **Nonlinear ensembles**
  - Nonlinear multi-model NN ensemble for predicting precipitation rates over ConUS
  - Nonlinear NN averaging of wave models ensemble
  - Nonlinear NN ensemble for hurricanes: improving track and intensity

- **Nonlinear bias corrections**
  - Nonlinear NN bias corrections
  - Nonlinear NN approach to improve CFS week 3 an 4 forecast
Summary

- HA is a synergetic approach that provides new advanced capabilities in NWMSs
- There is no free lunch, ML part of HA has limitations:
  - ML, as any statistical modeling, requires data for training; it is Learning from Data approach
  - ML, as any nonlinear statistical modeling, requires more data, than linear models/regressions
  - As any numerical models, ML applications should be periodically updated; however, ML can be updated on-line
  - Interpretation of ML models, as any nonlinear statistical models, is not obvious
More to consider

• Shallow NNs is a mathematical solution of ML Problem (Vapnik, 2019)
• From theoretical point of view, DNNs can not guarantee solution of ML Problem and should be considered as a “heuristic” approach (Vapnik, 2019)
• DNNs require significantly more data for training
• DNNs may become excessively nonlinear, which may lead to unstable extrapolation and even interpolation (instability when integrating in the model).
• Parsimony principle is still valid!
Questions?
Some Additional References - 1:


Aug 27, 2020 ML for Weather and Climate


DAS: Propagating Information Vertically Using NNs, Assimilating Chemical and Bio data

Example of ML (NN)-based Ensemble: Nonlinear Multi-model Ensemble Mean 24 hour precipitation forecast over ConUS

Ensemble members:
NCEP (global and regional), UKMO, ECMWF, JMA, Canada (global and regional), German.

 Verification Data
ML-based Ensemble. Closer to CPC with maintained sharpness and minimal false alarm.

Reduced maximum and diffused sharpness and fronts. A lot of false alarms. Due to slightly shifted maps from ensemble members.

doi:10.1155/2012/649450
Why we need ML: Data challenge

ML response to the challenge: Speed up data processing by orders of magnitude; improve extraction of information from the data; enhance assimilation of data in DASs
ML Correction for Hurricane Track and Intensity

- NN was trained on 1,200 storms in the Atlantic and Pacific basins (2015 – 2017)
- NN was applied to storms not included in the training set (2018)

Composite NN consisting of CNN and two Fully connected NNs

Track, Max Wind, and Min SLP error for HWRF forecast (red),
AI corrected forecast for track only model (black), AI corrected forecast
for the composite model (blue), and AI corrected for the composite and
updated weights model (green) with respect to best track for bias (left)
and RMSE (right).

Combining Artificial Intelligence and Physics-Based Modeling techniques to Improve
Hurricane Track and Intensity Forecasting. In preparation.
## Approximation Statistics and Speedup

<table>
<thead>
<tr>
<th>Statistics for Differences in Kelvin/day</th>
<th>NCEP CFS/GFS (L = 64)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RRTMG LWR</td>
<td>RRTMG SWR</td>
</tr>
<tr>
<td>Bias</td>
<td>2.·10^{-3}</td>
<td>5.·10^{-3}</td>
</tr>
<tr>
<td>RMSD</td>
<td>0.49</td>
<td>0.2</td>
</tr>
</tbody>
</table>

### Speedup factor, $\eta$

<table>
<thead>
<tr>
<th>Times</th>
<th>Averaged speedup factor: 16</th>
<th>Averaged speedup factor: 60</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Speedup factor in cloudy conditions: 20</td>
<td>Speedup factor in cloudy conditions: 88</td>
</tr>
</tbody>
</table>

**Note:** Work in progress to extend radiation emulation for FV3GFS

Note: GFS with NN LWR and NN SWR calculated every time step takes as much time as GFS with RRTMG radiation calculated one time per hour

NOAA AI Strategy

• **Goal 1**: Establish an efficient organizational structure and processes to advance AI across NOAA.
• **Goal 2**: Advance AI research and innovation in support of NOAA’s mission.
• **Goal 3**: Accelerate the transition of AI research to operational capabilities.
• **Goal 4**: Strengthen and expand AI partnerships.
• **Goal 5**: Promote AI proficiency in the workforce.