Forecasting Global Weather with Graph Neural Networks

Ryan Keisler
My background

10 years in physics + cosmology
(building things, data analysis, lots of stats)

7 years in industry
(satellite imagery, weather, physical & statistical models)
Quick look at main result
Main result: a new data-driven system for forecasting global weather

See bit.ly/graph_weather for more.
Motivation
Motivation

- Intellectual
- New opportunities
- Wouldn’t it be cool if...
Motivation

- Intellectual

- New opportunities

- Wouldn’t it be cool if...

  for <$1, I could run a high-quality, global weather forecast with an object that is
  - easy to share (some code + 20 MB weights)
  - easy to modify (fine-tune for specific applications)
  - easy to inspect (autodiff)
  - easy to glue (python)
Warming up
Before tackling weather, let’s try a toy problem.
Before tackling weather, let’s try a toy problem.

Let’s learn chaotic dynamics on the sphere.
Learning chaotic dynamics on the sphere

Truth  ML forecast  Difference

Learning Kuramoto-Sivashinsky with jraph + haiku
Learning Global Weather
Design Philosophy

Traditional NWP works really well so let’s:

- model the variables that drive traditional NWP (z, t, q, u, v, w)
- model on a dense physical grid (whatever $ and GPU allow)
- pick an architecture that enables this (~MeshGraphNet)
- pick a dataset that enables this (ERA5)
In this work, I used a 2 TB subset of ERA5:

- Horizontal resolution: 1.0 degrees in lat/lon
- Vertical resolution: 13 pressure levels
- Time: every 3 hours, from 1979 through 2020
- Fields: 6 fields (z, q, t, u, v, w)

Data stored as a single zarr array.
Spectrum of weather training data

- **Observations**
  - (e.g. weather sat data)
  - "Pure" data, no physical modeling imposed

- **Reanalysis data**
  - (e.g. ERA5)
  - A blend of observational data & physical modeling

- **Forecast data**
  - (e.g. GFS or ECMWF IFS)
  - "Pure" physical model
Spectrum of weather training data

Observations
(e.g. weather sat data)
“Pure” data, no physical modeling imposed

Reanalysis data
(e.g. ERA5)
A blend of observational data & physical modeling

Forecast data
(e.g. GFS or ECMWF IFS)
“Pure” physical model

Best-case ML scenario
You outperform NWP!

Best-case ML scenario
?

Best-case ML scenario
You perfectly emulate (but don’t outperform) the NWP engine
Architecture

“Learning Mesh-Based Simulation with Graph Networks”
by Tobias Pfaff, Meire Fortunato, Alvaro Sanchez-Gonzalez, Peter W. Battaglia
arXiv:2010.03409
Input: just use current state, i.e. no history

Output: just predict next state
Input: just use current state, i.e. no history

Model: lat/lon grid

Output: just predict next state

lat/lon grid  icosahedral grid  lat/lon grid
Input: just use current state, i.e. no history

Output: just predict next state

Model
1. **Encode**
   - From physical variables on lat/lon grid to latents on icosahedron grid using message-passing GNN.

2. **Process**
   - Using 9 rounds of message-passing GNN on icosahedron grid.

3. **Decode**
   - From latents on icosahedron grid to physical variables on lat/lon grid using message-passing GNN.

4. **Add**
   - The state change to input state to determine new state.
GNNs are well suited to NWP

- Easy to handle the \textit{spherical geometry} of earth
  - Just nodes in 3d space

- Potential for \textit{multi-resolution} models
  - e.g. learn from GFS \textit{and} HRRR?

- Potential for \textit{adaptive meshing}
  - i.e. put the compute where it is needed
Counting bits

25 MB

Model weights
## Counting bits

<table>
<thead>
<tr>
<th>Size</th>
<th>Description</th>
<th>Size</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 PB</td>
<td>ERA5 full</td>
<td>2 TB</td>
<td>ERA5 used</td>
</tr>
<tr>
<td>&gt;&gt;</td>
<td></td>
<td>&gt;&gt;</td>
<td></td>
</tr>
<tr>
<td>20 GB</td>
<td>ERA5 used, compressed</td>
<td>25 MB</td>
<td>Model weights</td>
</tr>
<tr>
<td>&gt;&gt;</td>
<td></td>
<td>~</td>
<td></td>
</tr>
<tr>
<td>5 MB</td>
<td>GFS FV3+FMS source code</td>
<td>&gt;&gt;</td>
<td>GFS atmos_model.F90</td>
</tr>
<tr>
<td>&gt;&gt;</td>
<td></td>
<td>&gt;&gt;</td>
<td></td>
</tr>
<tr>
<td>20 kB</td>
<td></td>
<td>1 kB</td>
<td>Primitive Equations</td>
</tr>
</tbody>
</table>
Counting bits

ERA5 full  ERA5 used  ERA5 used, compressed  Model weights  GFS FV3+FMS source code  GFS atmos_model.F90  Primitive Equations

No overfitting

There is a lot more data available.

But do you need it?
Results
6-hour Differences
3-day Rollout

- ERA5, 0 hours
- Initial conditions
- ERA5, 24 hours
- ML forecast, 24 hours
- ERA5, 48 hours
- ML forecast, 48 hours
- ERA5, 72 hours
- ML forecast, 72 hours
Hurricane Sandy
Hurricane Sandy
Hurricane Sandy
Hurricane Sandy
Hurricane Sandy
Hurricane Sandy
Hurricane Sandy
1-year rollout
1-year rollout (final frame)

- Grid pattern
- Overly smooth
- Location-specific problems (topography, land-sea, etc)
Anticipating GFS
Improves upon previous data-driven approaches
Comparable to Operational NWP

...when high-res op models are evaluated at ~1-deg scale

...when using reanalysis initial conditions

...but still, it works surprisingly well!
Thank you

Please see bit.ly/graph_weather for more.