RainBench: Enabling Data-Driven Precipitation Forecasting on a Global Scale

Catherine Tong – Christian Schroeder de Witt
Valentina Zantedeschi – Daniele De Martini – Freddie Kalaitzis – Matthew Chantry
Duncan Watson-Parris – Piotr Biliński
Global Precipitation Forecasting

Motivation

Climate change: rising extreme precipitation events

Global Precipitation Forecasting

Motivation

Numerical models: heavy data and resource requirements

Recent Machine Learning models: regional nowcasting (<8 hours)

This work: introduce a multi-modal benchmark dataset to advance global precipitation forecasting in the medium-range (3-5 days)
Rainbench

- Global precipitation estimation product provided by NASA
  - Native resolution 0.1°

SimSat
2016-present

- Generated from ECMWF
- Emulates 3 spectral channels from the Meteosat-10 SEVIRI satellite
  - Native resolution 0.1°

IMERG
2000 - present

- Global precipitation estimation product provided by NASA
  - Native resolution 0.1°

ERA5
1979-present

- ERA5 Reanalysis Product
- Broad spectrum of physical and atmospheric variables at different heights (e.g. humidity, temperature)
  - Includes precipitation
  - Native resolution 0.25°
PyRain

Efficient data loading pipeline
Performance Analysis

Benchmark Tasks

3 input data settings: (a) SimSat only, (b) ERA only, (c) Simsat + ERA

Forecasting precipitation values from: ERA5, or, IMERG

Model: ConvLSTM conditioned on lead-time

<table>
<thead>
<tr>
<th></th>
<th>ERA5</th>
<th>IMERG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1-day</td>
<td>3-day</td>
</tr>
<tr>
<td>Persistence</td>
<td>0.6249</td>
<td>0.6460</td>
</tr>
<tr>
<td>Climatology</td>
<td>0.4798</td>
<td>0.4802</td>
</tr>
<tr>
<td>SimSat</td>
<td>0.4610</td>
<td>0.4678</td>
</tr>
<tr>
<td>ERA</td>
<td>0.4562</td>
<td>0.4655</td>
</tr>
<tr>
<td>SimSat + ERA</td>
<td><strong>0.4557</strong></td>
<td><strong>0.4655</strong></td>
</tr>
</tbody>
</table>

Performance Analysis

Class Imbalance

<table>
<thead>
<tr>
<th></th>
<th>Slight</th>
<th>Moderate</th>
<th>Heavy</th>
<th>Violent</th>
<th>Micro Avg.</th>
<th>Macro Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unbalanced</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERA</td>
<td>0.20</td>
<td>4.08</td>
<td>16.2</td>
<td>63.1</td>
<td>0.65</td>
<td>20.9</td>
</tr>
<tr>
<td>SimSat</td>
<td>0.20</td>
<td>4.38</td>
<td>16.8</td>
<td>54.1</td>
<td>0.65</td>
<td>18.9</td>
</tr>
<tr>
<td>SimSat + ERA</td>
<td>0.20</td>
<td>4.03</td>
<td>16.5</td>
<td>53.0</td>
<td><strong>0.65</strong></td>
<td><strong>18.4</strong></td>
</tr>
<tr>
<td>Balanced</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERA</td>
<td>1.05</td>
<td><strong>2.75</strong></td>
<td>12.4</td>
<td>58.0</td>
<td>1.40</td>
<td>18.6</td>
</tr>
<tr>
<td>SimSat</td>
<td>1.17</td>
<td>3.10</td>
<td>13.3</td>
<td>50.1</td>
<td>1.26</td>
<td>16.9</td>
</tr>
<tr>
<td>SimSat + ERA</td>
<td>1.30</td>
<td>3.15</td>
<td><strong>11.8</strong></td>
<td><strong>44.3</strong></td>
<td>1.38</td>
<td><strong>15.1</strong></td>
</tr>
</tbody>
</table>

Model: LightGBM
RainBench

Future Work

1. Limited extreme precipitation events
   class-balanced sampling
2. Modelling earth topology
   neural network architectures for spherical data
3. Using high-resolution data
   multi-fidelity approach
4. Making use of atmospheric state variables
   physics-informed ML approach
RainBench: Enabling Data-Driven Precipitation Forecasting on a Global Scale

Release expected by Dec 2020.

Thank you for listening.

Link to code:
https://github.com/FrontierDevelopmentLab/PyRain