Machine Learning for Detection and Prediction of Precipitation on Climate Timescales

Maria J. Molina, NCAR, Boulder, Colorado

Gerald A. Meehl, Katie Dagon, Jadwiga H. Richter, Sasha Glanville, Judith Berner, Aixue Hu, John Truesdale, James C. Biard (Climate AI), Kenneth E. Kunkel (NC State), and others
“The frequency and intensity of heavy precipitation events have increased since the 1950s over most land area for which observational data are sufficient for trend analysis (high confidence)...”

Image from: https://www.nersc.gov/research-and-development/data-analytics/climatenet/
Image from: https://www.nersc.gov/research-and-development/data-analytics/climatenet/

MCS labels from Feng et al. (2020).
Input fields from ERA5 (Hersbach et al. 2020).
Feature detection ML-based models trained with one dataset, and applied to another dataset, can work!

Prediction of precipitation on longer (than weather) timescales is challenging…


"Ensemble realignment"?
“Ensemble realignment”?  

Other ongoing projects: S2S bias correction with ML-based ensembles.

Weather and Climate Extreme Events in a Changing Climate

confidence about changes in sub-daily extreme precipitation due to the limited number of studies and available data.

11.4.3 Model Evaluation

Evaluating climate model competence in simulating heavy precipitation extremes is challenging due to a number of factors, including the lack of reliable observations and the spatial scale mismatch between simulated and observed data (Avila et al., 2015; Alexander et al., 2019). Simulated precipitation represents areal means, but station-based observations are conducted at point locations and are often sparse. The areal-reduction factor, the ratio between pointwise station estimates of extreme precipitation and extremes of the areal mean, can be as large as 130% at CMIP6 resolutions (about 100 km) (Gervais et al., 2014). Hence, the order in which gridded station based extreme values are constructed (i.e., if the extreme values are extracted at the station first and then gridded, or if the daily station values are gridded and then the extreme values are extracted) represents different spatial scales of extreme precipitation and needs to be taken into account in model evaluation (Wehner et al. 2020). This aspect has been considered in some studies. Reanalysis products are used in place of station observations for their spatial completeness as well as spatial-scale comparability (Sillmann et al., 2013a; Kim et al., 2020; Li et al., 2021). However, reanalyses share similar parametrizations to the models themselves, reducing the objectivity of the comparison.

Weather and Climate Extreme Events in a Changing Climate

confidence about changes in sub-daily extreme precipitation due to the limited number of studies and available data.

11.4.3 Model Evaluation

Evaluating climate model competence in simulating heavy precipitation extremes is challenging due to a number of factors, including the lack of reliable observations and the spatial scale mismatch between simulated and observed data (Avila et al., 2015; Alexander et al., 2019). Simulated precipitation represents areal means, but station-based observations are conducted at point locations and are often sparse. The areal-reduction factor, the ratio between pointwise station estimates of extreme precipitation and extremes of the areal mean, can be as large as 130% at CMIP6 resolutions (about 100 km) (Gervais et al., 2014). Hence, the order in which gridded station based extreme values are constructed (i.e., if the extreme values are extracted at the station first and then gridded, or if the daily station values are gridded and then the extreme values are extracted) represents different spatial scales of extreme precipitation and needs to be taken into account when model evaluation is performed (Wehner et al. 2020). This aspect has been considered in some studies. Reanalysis products are used in place of station observations for their spatial completeness as well as spatial-scale comparability (Sillmann et al., 2013a; Kim et al., 2020; Li et al., 2021). However, reanalyses share similar parameterizations to the models themselves, reducing the objectivity of the comparison.

“…lack of reliable observations and the spatial scale mismatch between simulated and observed data…”

“…reanalyses share similar parameterizations to the models themselves…”


Figures from:
Historically marginalized communities experience disproportionate vulnerability to extreme weather events and climate change (Thomas et al. 2019).
Pathways to mitigating biases in environmental science data for marginalized communities

**PI:** Amy Quarkume, Howard University

**NCAR Collaborators:** Maria J. Molina, Forrest Lacey, Curtis Walker

- Societal biases exist within Earth system science data.
- If not accounted for, AI models trained using these datasets can also learn systemic biases.
- Proliferation of AI for high-stakes tasks requires us to think carefully about bias leakage.
- Focus: Air quality (Louisiana), weather disasters (Texas/Florida), and transportation intersection.
Climate change: a threat to human wellbeing and health of the planet. Taking action now can secure our future _________
Prompt 1: Major challenge (*in situ* precipitation observations) – can we reimagine how we obtain observations?

Prompt 2: Feature detection and weather regime/patterns *could* be applied to assess features and their projected changes for CMIP7.

Prompt 3: Methods can (at times) be robust across data products.

Prompt 4: Extends to other problems (e.g., lack of features, prediction of other quantities).