Characterization of Extremes and Compound Impacts: Applications of Machine Learning and Interpretable Neural Networks

Yan Feng[1], Romit Maulik[1], Jiali Wang[1], Prasanna Balaprakash[1], Whitney Huang[2], Vishwas Rao[1], Pengfei Xue[1], William Pringle[1], Julie Bessac[1], & Ryan Sullivan[1]
[1]Argonne National Laboratory; [2]Clemson University

Focal Area

This white paper responds to Focal area III by exploring data fusion, learning and explainable AI methods in characterizing hydrological extremes and interconnections. It also addresses Focal area II by using probabilistic AI and ensemble ML for predicting extremes and compound extremes.

Science Challenge

A key question associated with the integrated water (or hydrological) cycle grand challenge in the Earth and Environmental Systems Sciences Division (EESSD) strategic plan, is how the frequency and intensity of hydrological events will change. Prediction of the tail behavior (extremes) of the hydrological cycle is especially challenging, because of their stochasticity and low probability. These extreme events and their compound impacts have significant societal and economic consequences. It is anticipated for the next-generation Earth System models (ESMs), that model predictability of the water cycle will improve with increased resolution (e.g., regionally refined E3SM), advanced software and computational architectures, and improved model physics based on the data from ARM measurements and high-fidelity models. However, the challenges for predictability of low-probability high-impact extreme events will unlikely be alleviated with conventional modeling and data-driven approaches, as ESMs are calibrated largely for capturing the high-frequency mean climate states.

Recent AI and ML applications have shown great potential in quantifying well-defined climate extremes (e.g., supervised learning of tropical cyclones/atmospheric rivers by ClimateNet[1]) but few efforts are dedicated to compound events, extreme drivers and uncertainty estimation. We envision the opportunity to develop and apply ML and interpretable AI methods extended on the existing efforts, specifically, for: (1) *identification of compound extremes, (2) diagnosing drivers of extremes, (3) bias correction in extreme predictions and (4) probabilistic modeling of extremes*.

Rationale

In the following, we outline the limitations of the conventional approaches for:

(1) *Identification of compound extremes*. Identifying compound (including concurrent and sequential) extremes is far more challenging than identifying single events. Moving from univariate to multivariate frequency analysis typically introduces additional assumptions about the dependence structure to overcome the curse of dimensionality[2]. Quantification of the magnitude of compound extremes is also not straightforward, because a given return level is no longer a single value, but a set of vector values. Furthermore, compound extremes are often triggered by spatial-temporally non-local influences, e.g., the 2010 European mega heat wave in which sensible heat advected from the upwind regions led to abrupt increases in the surface air temperature that were further amplified by the local land–atmosphere feedbacks via soil desiccation[3]. Such spatial-temporal displacement makes it difficult

DOI: https://doi.org/10.2172/1769686
DOE BER, AI4ESP 2021 #AI4ESP1044 (https://www.ai4esp.org)

to characterize the interrelationships between multiple events. Utilizing supervised learning methods along with fusion of observations and model data could be potentially useful.

(2) *Diagnosing drivers of extremes*. Traditionally, drivers are explored based on empirical knowledge of event precursors. Those approaches could be limited by the lack of physical understanding and result in ambiguous conclusions. Another example of complexity is the compound flooding at the aquatic-terrestrial interface (TAI). It could result from a single extreme event (e.g., heavy rainfall); or multiple events (e.g., a storm surge combined with inland heavy rainfall), in which each individual event need not necessarily be classified as "extreme". It is non-trivial to distinguish the influences of local drivers from teleconnections. Development of surrogate models combined with explainable AI provides a means for tackling these issues.

(3) *Bias correction in extreme predictions*. Ensemble modeling is commonly used for prediction of regional extremes. Representing the total uncertainty in an ensemble is a key to the success of ensemble predictions. But the model uncertainty is often underestimated or misrepresented especially in the case of extreme events, because models are typically calibrated in fair weather or mean climate conditions and may have insufficient model resolution or physics to extract necessary variance in the fluid flow that will manifest itself at the tail of the distribution in the form of an extreme event. These limitations could benefit from ensemble learning, a class of supervised learning approaches that combines a number of ML methods for bias correction.

(4) *Probabilistic modeling of extremes*. Identification of hydrological extremes have largely relied on empirical metrics (e.g., block times; percentiles) or indices (e.g., drought indices). There are often large discrepancies between different detection algorithms for the same type of extreme events. Not surprisingly, different algorithms would lead to different estimates of occurrence, intensities, and spatial extents of the extreme events associated. Additionally, uncertainty associated with the data sources needs to be assessed in translating the ESM predictions to credible characterization of extremes for stakeholders. Thus, probabilistic projections with explainable AI are tractable to deal with these issues.

Narrative

While the progress with the ML and AI models in improving the predictability of extremes is promising, there are still open challenges. In the following, we describe the unique opportunities with supervised/unsupervised learning, explainable AI, and probabilistic AI for:

(1) *Identification of compound extremes*. A major obstacle for extreme detection is limited datasets with high quality, especially for compound extremes that are confined to the physical model capability. To overcome this, a multi-fidelity data fusion framework may be used to construct the training datasets, combining model outputs from high-resolution ESMs, hybrid or fully data-driven emulators along with observations (e.g., ARM and satellite observations). This data fusion may be performed by building linear[4] or nonlinear multi-fidelity models[5] which utilize Gaussian processes to build bias relationships between different datasets (models or observations) and propagate uncertainties through to the fused data outputs.

One opportunity for detection of extremes (or anomalies), single or compound, is to calibrate time-series models[6,7] based on the spectral content of fused flow-field information. For instance, for compound heat wave and drought, one could assume a specific distribution for the time-series data streams, e.g., surface air temperature and soil moisture, and calibrate with regular simulation data, which more often than not fail to capture extremes. With this supervised probabilistic calibration (any signal from real-time observations can be flagged as "anomalous" if it presents itself to be beyond a-priori specified confidence intervals), one may determine precursors to anomalous events before they are observed as flow-field features through conventional analyses. For compound extremes, dimension reduction techniques (i.e., deep autoencoders) may be used to identify a low-dimensional manifold first on which such extremes are detectable in time with the anomaly detection methods[8]. Alternatively, another data-driven approach is to decompose the problem of estimating multivariate distribution into i) estimation of univariate tail distribution and ii) estimation of conditional high quantiles given another variable being extremes. It allows for flexible modeling for the conditional dependence in the joint tail.

(2) *Drivers of extremes.* While traditional approaches are based on empirical knowledge, AI is helpful to identify any "missing" drivers not considered. First, we could build ML models for specific extreme events. In addition to the supervised learning used by ClimateNet[1], which requires a carefully curated datasets, the self-supervised learning approach may be adapted. A pretext task is defined such that labels can be automatically calculated directly from the raw data instances. During the learning process, the neural network teaches itself good representations based on the pretext task. This is useful in particular for events that are difficult to label. For example, to find the drivers that lead to a concurrent drought and heat event, we need form the pretext task by designing a ML architecture that include many input variables and the compound extremes as labels. The input variables may include the drivers based on the domain knowledge - although we may not know how they influence the extremes (e.g., linear or non-linear effects), as well as those we could not determine to be important or not.

If the ML models perform well on extremes, we can further examine the feature/variable (e.g., processes, regions, or seasons) importance through explainable AI methods such as backwards optimization and layer-wise relevance propagation (LRP)[9] for a better understanding of the drivers. Another approach is to perform sensitivity analysis[10,11] on the variables combined with observations. Through the analysis, one can identify processes and/or spatio-temporal regions that contribute the most to errors in the predicted extreme events or where additional observations are needed for improvement.

(3) *Bias correction in extreme predictions*. A promising approach for bias correction is using ML model ensembles. This also requires us first to develop computationally efficient emulators using ML surrogate models for extreme events. Specifically, we can build a large number of diverse ML models (e.g., using automated ML methods on exascale machines) from the same simulation data, where each model approximates input and output in a different way. The ensemble allows us to model the unknown relationships between the input factors and extremes using different functions (hypothesis) and overcome the limited functional physics space that the physical models can capture. The constructed ensemble ML models due to their sheer number can potentially reduce the bias in general. With the observations, we can further calibrate the models. Given a few extreme observations, we can leverage attention and model selection techniques to automatically calibrate the model ensembles to improve the prediction.

(4) *Probabilistic modeling of extremes*. Probabilistic modeling is very appealing for detecting or forecasting extremes, as it provides uncertainty quantification. To simplify the learning problem, data may need be formulated or labeled through preprocessing/augmentation like ClimateNet[1]. Such probabilistic segmentation and bounding box approaches could be further investigated against classical regression or classification techniques. The probabilistic formulation for the ML framework may range from full-blown Bayesian regression to computationally effective (if limited) variational inference[8,12], mixture density networks[13], and ensemble methods such as stochastic weight averaging[14] and deep ensembles[15]. Furthermore, a coupling with explainable AI, such as SHAP[16] or LRP[9], can shed light on the factors that engender these extreme events by performing causal analysis once a specific deep learning model has been constructed and calibrated.

Suggested Partners/Experts (Optional)

References (Optional)

1 Prabhat *et al.* ClimateNet: an expert-labeled open dataset and deep learning architecture for enabling high-precision analyses of extreme weather. *Geosci. Model Dev.* **14**, 107-124, doi:10.5194/gmd-14-107-2021 (2021).

2 Serinaldi, F. Can we tell more than we can know? The limits of bivariate drought analyses in the United States. *Stochastic Environmental Research and Risk Assessment* **30**, 1691-1704, doi:10.1007/s00477-015-1124-3 (2016).

3 Schumacher, D. L. *et al.* Amplification of mega-heatwaves through heat torrents fuelled by upwind drought. *Nat Geosci* **12**, 712-717, doi:10.1038/s41561-019-0431-6 (2019).

4 Kennedy, M. C. & O'Hagan, A. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63**, 425-464, doi:https://doi.org/10.1111/1467-9868.00294 (2001).

5 Perdikaris, P., Raissi, M., Damianou, A., Lawrence, N. D. & Karniadakis, G. E. Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **473**, 20160751, doi:10.1098/rspa.2016.0751 (2017).

6 Hundman, K., V. Constantinou, C. Laporte, Colwell, I. & Soderstrom, T. Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding. *CoRR* **abs/1802.04431** (2018).

7 Merkes, S. N. *et al.* in *AIAA Scitech 2021 Forum AIAA SciTech Forum* (American Institute of Aeronautics and Astronautics, 2021).

8 Maulik, R., Botsas, T., Ramachandra, N., Mason, L. R. & Pan, I. Latent-space time evolution of non-intrusive reduced-order models using Gaussian process emulation. *Physica D: Nonlinear Phenomena* **416**, 132797, doi:https://doi.org/10.1016/j.physd.2020.132797 (2021).

9 Toms, B. A., Barnes, E. A. & Ebert-Uphoff, I. Physically Interpretable Neural Networks for the Geosciences: Applications to Earth System Variability. *Journal of Advances in Modeling Earth Systems* **12**, e2019MS002002, doi:https://doi.org/10.1029/2019MS002002 (2020).

10 Daescu, D. N. & Todling, R. Adjoint sensitivity of the model forecast to data assimilation system error covariance parameters. *Quarterly Journal of the Royal Meteorological Society* **136**, 2000-2012, doi:https://doi.org/10.1002/qj.693 (2010).

11 Sandu, A., Cioaca, A. & Rao, V. Dynamic Sensor Network Configuration in InfoSymbiotic Systems Using Model Singular Vectors. *Procedia Computer Science* **18**, 1909-1918, doi:https://doi.org/10.1016/j.procs.2013.05.360 (2013).

12 Hoffman, M. D., Blei, D. M., Wang, C. & Paisley, J. Stochastic variational inference. *J. Mach. Learn. Res.* **14**, 1303–1347 (2013).

13 Bishop, C. M. *Mixture Density Networks*. (1994).

14 P. Izmailov & D. Podoprikhin, T. G., D. Vetrov, and A. G. Wilson. Averaging weights leads to wider optima and better generalization. *arXiv preprint* **arXiv:1803.05407** (2018).

15 Lakshminarayanan, B., Pritzel, A. & Blundell, C. in *Advances in neural information processing systems* 6402-6413 (2017).

16 Lundberg, S. M. & Lee, S.-I. in *Advances in Neural Information Processing Systems 30 (NIPS 2017)*. 4765-4774 (NeurIPS Proceedings).