Science-integrated Artificial-intelligence for Flooding and precipitation Extremes (SAFE)

AUTHORS / AFFILIATIONS (Areas: Climate, Hydrology, AI, HPC, Models, Extremes, Risks)
1. Auroop R. Ganguly (Lead): Northeastern University, Boston, MA (Climate/Water, AI)
2. Sue Ellen Haupt: National Center for Atmospheric Research, Boulder, CO (Weather, AI)
3. Forrest M. Hoffman: Oak Ridge National Laboratory, Oak Ridge, TN (ESMs, AI)
4. Vipin Kumar: University of Minnesota, Twin Cities, MN (AI, HPC, Climate/Water)
5. Upmanu Lall: Columbia University, New York, NY (Hydrology, Climate, AI)
6. Claire Monteleoni: University of Colorado, Boulder, CO (ML, AI, Climate Informatics)
7. Jitendra Kumar: Oak Ridge National Laboratory, Oak Ridge, TN (Hydrology, Models)
8. Nagendra Singh: Oak Ridge National Laboratory, Oak Ridge, TN (GIS, Geoscience)
9. Julia Hopkins: Northeastern University, Boston, MA (Nature-Inspired Design)
10. Anuj Karpatne: Virginia Tech, Blacksburg, VA (Physics-Guided AI, Water)
11. Shafiqul Islam: Tufts University, Medford, MA (Hydrology, Water Diplomacy)
12. Samrat Chatterjee: Pacific Northwest National Laboratory, Richland, WA (Risks, OR)

FOCAL AREA(s)
• **Primary:** 3 (“Insights gleaned from complex data … physics- or knowledge-guided AI”)
• **Secondary:** 2 (“Predictive modeling through the use of AI … hierarchy of models”)

This white paper focuses on methods that blend AI and science (e.g., physics, biogeochemistry) by (a) guiding AI cost functions, trajectories and representations with science knowledge and/or with context-specific data-driven insights in a Bayesian-inspired framework; (b) framing AI models in the context of physics-informed dynamic, causal networks; (c) merging AI-enhanced science models with science-guided explainable AI; (d) focusing on statistics/processes related to extremes and translations to risks in a changing world; & (e) identifying model parameterizations or components that must be improved to minimize risks and add maximum value to stakeholders.

SCIENCE CHALLENGE
A grand challenge [1-10] in hydrologic science is to understand why signals of climate change and variability, which are often visible in precipitation extremes at aggregate scales, are not consistently observed in the case of extreme flooding. However, a solution to this challenge may prove elusive unless the water cycle is viewed in an integrative manner. Thus, for riverine flooding, while Hortonian (infiltration excess) runoff may have stronger correlation with precipitation extremes and hence perhaps to warming trends or climate oscillators, Dunne (saturation excess) runoff may have a more complex relationships with time series of precipitation and with evaporation and transpiration, but rain-on-snow and snowmelt events may depend on land-surface and atmospheric temperatures. Atmospheric rivers [11] and tropical cyclones [12] lead to precipitation or flooding and are impacted by climate. Flooding assessments need to consider long-term baselines [13], evolving risk factors [14], coupled natural-human systems [15-16], and novel adaptation such as nature-inspired design [17-18].

RATIONALE
The urgency of stakeholder needs pertaining to precipitation and flooding extremes requires transformative advances in long-standing challenges within earth systems sciences. Challenges for which emerging AI solutions have started to make a difference include (a) cloud physics and subgrid processes [19-24]; (b) spatiotemporal patterns and dependencies [25-29; 44]; (c) climate oscillations and teleconnections with regional hydrology [30-35]; (d) pattern search across multiple ensembles [35-36]; and (e) statistical downsampling [37-40; 45]. The AI solutions cited have ranged from machine learning (including Deep Learning), network-based approaches, and

DOI: https://doi.org/10.2172/1769776
DOE BER, AI4ESP 2021 #AI4ESP1047 (https://www.ai4esp.org)
spatiotemporal data mining with machine learning. The successful AI solutions [41-42] have mostly evolved along two categories: first, where process knowledge guides and informs AI methods to various degrees for analyzing observed and model-simulated data; and second, where AI gets embedded in computational models as data-driven equivalents of ‘parameterizations’. However, to be able to make a breakthrough in hydrologic (precipitation and flooding) extremes, three major barriers (among others) need to be overcome:

1. Hydrologic and hydroclimatic processes operate in highly heterogeneous environments, the properties of which are imprecisely known, causing scale-dependence of parameterizations; but as better/more data emerges, fine-tuning AI to address these challenges may be feasible.

2. Physics-based computer models of precipitation and flooding extremes rely too heavily on calibration parameters that do not generalize well under extrapolation; the ability of AI to address this challenge by considering a wider set of data and relations needs to be explored.

3. Predictability in hydrology or hydroclimatic is limited by a combination of long-standing gaps in process knowledge, human management, and internal variability; this limitation is further compounded by changes in risk elements such as vulnerability and exposure. Methodological and diagnostic understanding of spatiotemporal extremes generation processes, along with better appreciation of stakeholder risks, may suggest avenues for targeted improvements.

Our proposed “Focal Area(s)” described on the previous page, especially the three principled approaches indicated therein, are designed to directly target the three barriers discussed above.

**NARRATIVE**

A suite of science/knowledge integrated artificial intelligence (AI) methods may be needed to advance predictive understanding of precipitation and flooding extremes over 2 to 30-year projection horizons at stakeholder-relevant spatiotemporal resolutions. These methods may advance climate science and earth system modeling, develop credible projections, and lead to fundamental insights, and enable risk-informed decisions with a focus on precipitation and flooding extremes in the near term. Our overarching hypothesis is that intelligent blends of AI and physics can improve predictive understanding of crucial – yet long-standing – challenges.

Our team members are among the pioneers [62-80] in developing integrated science–AI models for earth systems: we will build upon our foundational work to inform workshops and plans for DOE that will inspire the development of a suite of solutions in physics-guided AI (PGAI), AI-enhanced physics (AIEP), including generative models, causal and interpretable AI, as well as uncertainty or risk assessments. Here we define ‘physics’ broadly to include biogeochemistry and process knowledge. We group the proposed advances under an umbrella phrase: Structure-Agnostic Process Models (SAPM). Rather than making a priori assumptions about the structural forms of the (parameterized) physics-based or AI-based models, SAPM will let the (ensemble of) structures (whether AIEP or PGAI) to evolve in a data-driven and physically (or scientifically) consistent manner. SAPM can therefore be thought of as a more general class of solutions that encapsulates both AIEP and PGAI, as well as their hybrid blends. One challenge in earth systems models (both physics and AI) is generalizability under changing conditions. Spatiotemporal analogues [37; 46-49] for model development and evaluation can be explored. Interpretability is crucial, a balance with accuracy may be possible by exploring manifolds of physically plausible hypotheses spaces. A combination of implicit and physics-guided explicit regularization will further advance generalizability. Evaluation will entail embedding process models within Earth System Models (ESMs) of differing complexity, as well as via spatiotemporal analogues (which will be extracted by mining historical space-time data that probabilistically resemble plausible future climate) to infer generalizability under (‘non-stationary’) change [43].
Uncertainty assessments (UA), including intrinsic variability and knowledge gaps [50-52], may rely on large-ensemble runs, what-ifs, and AI-based emulations of processes embedded in the latest-generation of ESMs along with adaptations of recent developments in information theory. We will discuss model-based ensembles (such as ESM large-ensembles) and observation-based ensembles to discuss a falsifiable theory for initial-condition and multi-parameter variability, in a way that balances skills, convergence and the ability to reproduce observed data associations and tail behavior. We will focus on how to bridge the gap between researchers in nonlinear physics and information theory who tend to work with relatively simplified models and computational climate modelers working with the latest ESMs [53-60]. Lessons learned with simplified models, from Lorenz and Held-Suarez to reduced-order ESM, will be adapted to more complex ESMs. AI-inspired risk quantification and prediction methods based on fundamental principles of operational research, data-driven science, and machine learning may characterize and quantify hazards, vulnerability, and exposure [13-18], with observations, model-simulations, and what-if scenarios. Diagnostic and prognostic insights on extremes generation processes in space-time can be combined with assessments of critical stakeholder needs (e.g., for adaptation or mitigation), which in turn may identify targeted model improvement goals while also informing decisions. A hypothesis-driven evaluation strategy needs to rely on experiments to examine causal interpretability and generalizability of the hybrid physics-AI methods under changing climate conditions. We will develop a causal-AI system for methods in parameter estimation, prediction, and uncertainty quantification. We propose a simulation-observation experimental testbed, where structure-agnostic hybrid physics-AI methods will be compared with observations (including those based on spatiotemporal analogues and observation-based ensembles) and examined by embedding within models of disparate complexity, e.g., from reduced order or constrained models to the latest generation of ESMs for multi-parameter and initial condition ensembles.

Organizing Principles for Workshops and a Five to Ten Year Vision
The workshops will characterize the state-of-the-art, develop requirements and specifications for novel methods (e.g., compare with [61]), and generate 5- to 10-year guidelines for future ESMs. We propose workshop organizing principles along 5 mutually supporting areas (a) understanding processes relevant for ESMs (Years 1-4), (b) SAPM methods and their evaluation (Years 2-5), (c) iterative feedback among: a–b (Years 3-6), (d) research into integration/evaluation of human system models (Year 1-6), (e) incorporation of human dynamics into future ESMs (Years 6-10). The timelines will be adjusted as physics-integrated AI and human system models mature.

Figure 1: The integrative water cycle, with core processes relevant for precipitation and flooding extremes, are shown on the left. The SAPM approach, including science-integrated AI and associated evaluation testbeds, risks and uncertainty, and ESM improvements, are on the right. Workshop organizing principles (a)–(c) relate to this figure. Research into human systems (e.g., lifelines, policy) and their integration into ESMs, (d)–(e), are not shown.
SUGGESTED PARTNERS / EXPERTS

Our 12-member core team is listed under Authors / Affiliations. In addition, we plan to select our invitees from the following other experts and specialists (with many of whom our team members have a long history of collaboration):

1. JCJH Aerts, Vrije Universiteit Amsterdam, Netherlands
2. Arindam Banerjee, University of Minnesota, Twin Cities, MN
3. Ana Barros, University of Illinois, Urbana-Champaign, IL
4. Ed Beighley, Northeastern University, Boston, MA
5. Udit Bhtatia, Indian Institute of Technology Gandhinagar, India
6. Rafael Bras, Georgia Institute of Technology, Atlanta, GA
7. Snigdhansu Chatterjee, University of Minnesota, Twin Cities, MN
8. Nitesh Chawla, University of Notre Dame, South Bend, IN
9. Clara Deser, National Center for Atmospheric Research, Boulder, CO
10. Jennifer Dy, Northeastern University, Boston, MA
11. Imme Ebert-Uphoff, Colorado State University & NOAA/CIRA, Fort Collins, CO
12. Stephen Flynn, Northeastern University, Boston, MA
13. David John Gagne, National Center for Atmospheric Research, Boulder, CO
15. Subimal Ghosh, Indian Institute of Technology Bombay, India
16. Dave Gochis, National Center for Atmospheric Research, Boulder, CO
17. Marta Gonzalez, University of California, Berkeley, CA
18. Tarik Gouhier, Northeastern University, Boston, MA
19. Carling Hay, risQ Corp., Cambridge, MA
20. Christiane Jablonowski, University of Michigan, Ann Arbor, MI
21. George Em Karniadakis, Brown University, Providence, RI
22. Evan Kodra, risQ Corp., Cambridge, MA
24. Amy McGovern, the University of Oklahoma, Norman, OK
25. Binayak Mohanty, Texas A&M University, College Station, TX
26. Ramakrishna Nemani, NASA Ames, Moffett Field, CA
27. Paul O’Gorman, Massachusetts Institute of Technology, Cambridge, MA
28. Nikunj Oza, NASA Ames, Moffett Field, CA
29. Ismael Pagan-Trinidad, University of Puerto Mayaguez, PR
30. Swapna Panickal, Indian Institute of Tropical Meteorology, Pune, India
31. Sai Ravela, Massachusetts Institute of Technology, Cambridge, MA
32. David Rolnick, McGill University, Montreal, Quebec, Canada
33. Jakob Runge, TU Berlin, Berlin, Germany
34. Tapio Schneider, Caltech, Pasadena, CA
35. Meytar Sorek-Hamer, NASA Ames, Moffett Field, CA
36. Aneesh Subramanian, University of Colorado Boulder, CO
37. Thomas J. Vandal, NASA Ames, Moffett Field, CA
38. Don Wuebbles, University of Illinois, Urbana-Champaign, IL
39. David Yates, National Center for Atmospheric Research, Boulder, CO
40. Moira Zellner, Northeastern University, Boston, MA
REFERENCES  

/Publications by the members of this team in blue/


Correlated Climate Data. In 2018 IEEE International Conference on Big Knowledge (ICBK) (pp. 146-153). IEEE.


Auroop Ganguly was the invited lead author of the AI section

Summary of the above was published in BAMS (Moss et al., 2019, 100(5), 897-907)

More details available from SCAN website: climateassessment.org