AI4ESP 2021 SAIL White Paper

Title: Reliable modeling and prediction of precipitation & radiation for mountainous hydrology

Authors:
Daniel Feldman[1], V. Chandrasekar[2], P. James Dennedy-Frank[1], Dipankar Dwivedi[1], Mayur Mudigonda[1], Karthik Mukkavilli[3], Utkarsh Mital[1], Erica Siirila-Woodburn[1]

[1] LBNL; [2] Colorado State University; [3] University of California-Irvine

Focal Area

This White Paper focuses on data-driven atmospheric process model emulation and atmospheric process surrogate model development. It proposes leveraging recent AI advances in these approaches to fill in unavoidable observational gaps and enable high-fidelity modeling/predictability of the atmosphere and land-surface interactions in mountainous watersheds. This approach will support studies and predictability of water cycle extremes.

Science Challenge

As described in the 2019 ARM Mobile Facility Workshop Report, "Diagnosing model biases has been difficult [in mountain regions] due to the sparse networks of observations available to constrain models." Some mountainous regions are heavily-studied, yielding large amounts of process understanding, but the necessary and sufficient conditions for demonstrable process model skill that enable the transfer of predictive understanding of atmospheric forcings from heavily-studied watersheds to under-studied ones have not been established.

Rationale

Mountainous watersheds are under-observed and will be for the foreseeable future due to resource and logistical constraints, leaving an understanding gap in how they respond to and alter water cycle extremes. Models are needed to fill this gap, but models across a wide range of complexity, from Earth System Models (ESMs) to process models, are in need of improvement: for example, recent work by *Chen et al.*, [2014], *Wu et al.*, [2017], and *Rhoades et al.,* [2016, 2018a,b,c], all of these models exhibit limited predictability in the date of peak snowpack timing and spring snowmelt rate across the Western United States. However, there is a large chasm between identifying these common problems and fixing the models so that, at the very least, they produce unbiased estimates of mountainous hydrology.

Advances in mountainous hydrological science are needed to understand and improve the prediction of these systems from synoptic to decadal time-scales and with spatial scales ranging from meters to the entire watershed (potentially thousands of km) [*Bales et al.*, 2006]. Such advances require coordinated modeling and observational data collection [*Viviroli et al.,* 2011]. Spatially- and temporally-resolved precipitation and radiation fields are extremely heterogeneous in complex high-altitude terrain (CHAT), and networks of point observations under-resolve the processes that contribute to this heterogeneity. Significant resources are being devoted by the DOE to resolve these processes in targeted watersheds, such as the East River Watershed in the

DOI: https://doi.org/10.2172/1769771
DOE BER, AI4ESP 2021 #AI4ESP1043 (https://www.ai4esp.org)

Upper Colorado River Basin. There, a wide range of new scientific opportunities is arising with the 2021-2023 Surface Atmosphere Integrated Field Laboratory (SAIL) Campaign (https://sail.lbl.gov). At the same time, the scientific findings from SAIL, in conjunction with the Watershed Function SFA [*Hubbard et al.*, 2020], will be of limited scientific benefit unless knowledge from these and other campaigns can be demonstrably transferred to models that will supplement the limited observational network outside of the East River for mountainous hydrology studies in understudied regions. The long-term vision of this White Paper is to use AI model emulation and surrogate models to upscale field campaigns that focus on atmospheric processes, optimize data collection and accelerate transferable scientific understanding in CHAT.

Narrative:

For mountainous hydrology, the traditional observation and modeling work-flow is broken. That is, the approach whereby researchers collect observational data in the mountains, confront models with those data, identify model skill and reveal model deficiencies, and make improvements to those models accordingly, is not viable. The primary reason for this is that even the most heavily-studied mountain watersheds are so under-observed that traditional atmospheric process models, such as convection-resolving Weather Research and Forecasting (WRF), are more reliable for forcing hydrological models than observational datasets like PRISM [*Lundquist et al.*, 2019]. Consequently, the model improvement pathway is ill-posed: the ways in which new observations can improve atmospheric process models is often not immediately clear.

For this White Paper, we focus on methods that address this broken work-flow in CHAT for precipitation and radiation because (1) these fields are so central to scientific understanding of mountainous hydrology and (2) because the physical processes responsible for heterogeneity in these fields are qualitatively known and are therefore amenable to AI modeling [*Reichstein et al.*, 2019]. The micro- and macro-physical processes, which exert immense effects control orographic precipitation, are enumerated (e.g., *Stoelinga et al*., [2013]); as are radiative processes (e.g., *Hoch et al*., [2011], and both can be resolved observationally.

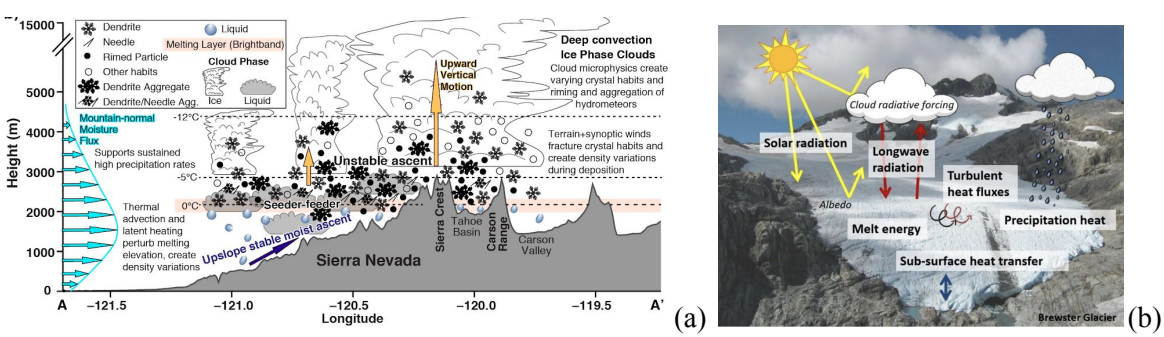

Figure 1: (a) Profile of macro- and micro-physical processes across the Sierra Nevada that influence orographic precipitation (from Hatchett et al., In Prep), (b) Cartoon of radiative processes in CHAT (NIWA).

AI methods have a number of advantages for augmenting and, in some cases acting as a surrogate for, traditional process models (TPMs) such as WRF. First, the work-flow to set-up

models such as WRF to support scientific understanding at a given location is an engineering exercise that involves numerous, unconstrained choices on physical parameterization schemes, boundary and initial conditions, domain and resolution. All of these impact modeled precipitation and radiation. Second, WRF simulations are computationally-expensive, and experiments are heavily constrained by CPU hour limitations. Finally, the process to fix WRF schemes with observations is painstaking and laborious.

While AI methods can be no better than (1) the information used to train them and (2) the insights into process modeling that researchers build into their structure, they can readily absorb a wide range of distinct observational data-streams with varying degrees of sensitivity to the physical processes that impact precipitation and radiation in CHAT. To that end, ***AI methods can identify where multiple, diverse observational datastreams exhibit enhanced sensitivity to the physical processes of interest***. They therefore represent an attractive alternative to advancing mountainous hydrology with process models, because AI approaches, such as conditional Generative Adversarial Networks (cGANS), can optimally ingest atmospheric process information from dozens of unique datasets, which field campaigns like SAIL provide.

AI-based model emulation and surrogate models are needed in CHAT, especially where they enhance scientific value of data collection during field campaigns. Nascent efforts have already indicated the scientific value of the computational efficiency gained from AI-based model emulation. cGAN-model emulation runs ~100x faster than WRF, and produces computationally-inexpensive gradients that enable the rank-ordering of the drivers of hydrological model uncertainty in CHAT [*Manepalli et al.*, 2019]. This capability can therefore quantify how new observations, such as from an IOP, reduce model uncertainty. ***AI-emulation may help field campaigns dynamically respond to their findings to date***: the superior efficiency of emulators allows for exploration, in mid-campaign, of the scientific utility of additional observations. Emulators enable exploration in far greater detail than TPMs.

AI-based surrogate atmospheric process modeling can also be of great utility for assessing the value added by additional observations, and provides an alternative pathway that avoids structural uncertainties in TPMs. Surrogate models can readily incorporate rich datasets, such as from SAIL [*Chen et al*, 2020; *Liu et al*, 2020; *Mital et al*, 2020], and can quickly discern obvious and non-obvious relationships that impact precipitation and radiation without having to first sleuth out WRF model errors. Surrogate models, such as deep neural networks, have already been shown to discern these relationships [*Weber et al*, 2020]. While there are real risks of surrogate model over-training and challenges with model error diagnostics, surrogate models can quickly transfer learning from SAIL to other under-studied watersheds [*Chandrasekar,* 2020]. With strict tests of such models, they may be superior to TPMs for scientific understanding and observational gap filling in CHAT.

Finally, AI-based model emulation and surrogate models in CHAT can and should adhere to FAIR principles in order to ensure reproducibility and demonstrate their utility to AI skeptics. They can do so by using freely available data and open-source, accessible code repositories.

Suggested Partners/Experts:

The development of AI methods to advance mountainous hydrology, in conjunction with the SAIL Campaign, to demonstrate transferability of process information from the East River Watershed to elsewhere would benefit from the following partners and experts:

- Close collaboration with each Instrument Mentor for each data stream from the SAIL Campaign to ensure high quality data is used for AI training and testing.
- Partnerships with atmospheric process experts in observational science and atmospheric process model testing and development to test AI methods for process representations.
- Partnerships with surface and sub-surface hydrologists to guide the levels of precision and accuracy needed in precipitation and radiation fields.
- Partnerships with AI researchers that have experience with geophysical sciences and can bridge the numerics between AI and geophysical modeling, to develop optimal and transferable emulation and surrogate models.

References:

Bales, R. C., N. P. Molotch, T. H. Painter, M. D. Dettinger, R. Rice, and J. Dozier (2006), Mountain hydrology of the western United States, *Water Resour. Res.*, **42**, W08432, doi: https://doi.org/10.1029/2005WR004387.

Chandrasekar, V., (2020), AI in Weather Radars, 2020 IEEE Radar Conference (RadarConf20), Florence, Italy, 1-3, doi: https://doi.org/10.1109/RadarConf2043947.2020.9266442.

Chen, F., et al. (2014), Modeling seasonal snowpack evolution in the complex terrain and forested Colorado Headwaters region: A model intercomparison study, *J. Geophys. Res. Atmos.*, **119**, 13,795– 13,819, doi: https://doi.org/10.1002/2014JD022167.

Chen, H., V. Chandrasekar, R. Cifelli and P. Xie, (2020), A Machine Learning System for Precipitation Estimation Using Satellite and Ground Radar Network Observations, *IEEE Transactions on Geoscience and Remote Sensing*, **58**(2), 982-994, doi: https://doi.org/10.1109/TGRS.2019.2942280.

Hoch, S. W., Whiteman, C. D., & Mayer, B. (2011). A Systematic Study of Longwave Radiative Heating and Cooling within Valleys and Basins Using a Three-Dimensional Radiative Transfer Model, *Journal of Applied Meteorology and Climatology*, **50**(12), 2473-2489, doi: https://doi.org/10.1175/JAMC-D-11-083.1.

Hubbard, SS, Varadharajan, C, Wu, Y, Wainwright, H, Dwivedi, D. (2020) Emerging technologies and radical collaboration to advance predictive understanding of watershed hydrobiogeochemistry. *Hydrological Processes*. **34**: 3175– 3182, doi: https://doi.org/10.1002/hyp.13807.

Liu, H., Chandrasekar, V., & Xu, G. (2001). An Adaptive Neural Network Scheme for Radar Rainfall Estimation from WSR-88D Observations, *Journal of Applied Meteorology*, **40**(11), 2038-2050, doi: https://doi.org/10.1175/1520-0450(2001)040<2038:AANNSF>2.0.CO;2.

Lundquist, J., et al., (2019) Our Skill in Modeling Mountain Rain and Snow is Bypassing the Skill of Our Observational Networks, *Bull. Amer. Meteo. Soc.*, 2473-2490, doi: https://doi.org/10.1175/BAMS-D-19-0001.1.

Manepalli, A., A. Albert, A. Rhoades, D. Feldman (2019) Emulating Numeric Hydroclimate Models with Physics-Informed cGANs, NeurIPS 2019, reference: https://www.climatechange.ai/papers/neurips2019/39/paper.pdf

Mital, U., et al., (2020) Sequential Imputation of Missing Spatio-Temporal Precipitation Data Using Random Forests, *Frontiers in Water*, **2**(20), doi: https://doi.org/10.3389/frwa.2020.00020

Reichstein, M., et al. (2019) Deep learning and process understanding for data-driven Earth system science, *Nature*, **566**, 195–204, doi: https://doi.org/10.1038/s41586-019-0912-1.

Rhoades, A. M., P. A. Ullrich, C. M. Zarzycki (2018a). Projecting 21st century snowpack trends in western USA mountains using variable-resolution CESM, *Clim. Dyn.*, **50**, 261-268, doi: https://doi.org/10.1007/s00382-017-3606-0.

Rhoades, A. M., Ullrich, P. A., Zarzycki, C. M., Johansen, H., Margulis, S. A., Morrison, H., et al. (2018b). Sensitivity of mountain hydroclimate simulations in variable-resolution CESM to microphysics and horizontal resolution. Journal of Advances in Modeling Earth Systems, 10, 1357– 1380. doi: https://doi.org/10.1029/2018MS001326.

Rhoades, A. M., Jones, A. D., & Ullrich, P. A. (2018c). Assessing mountains as natural reservoirs with a multimetric framework. *Earth's Future*, **6**, 1221– 1241. doi:https://doi.org/10.1002/2017EF000789.

Stoelinga, M. T., R. E. Stewart, G. Thompson, and J. M. Theriault (2013), Microphysical processes within winter orographic cloud and precipitation systems, in *Mountain Weather Research and Forecasting*, edited by F. Chow et al., pp. 345–408, Springer Atmospheric Sciences, doi: https://doi.org/10.1007/978-94-007-4098-3_7.

U.S. DOE. 2019. Atmospheric Radiation Measurement (ARM) User Facility ARM Mobile Facility Workshop Report, DOE/SC-0197 Office of Biological and Environmental Research, U.S. Department of Energy Office of Science, Germantown, Maryland, USA. https://science.osti.gov/-/media/ber/pdf/community-resources/2019/ARM_Mobile_Facility_Workshop_Report.pdf

Viviroli, D., et al. (2011), Climate change and mountain water resources: overview and recommendations for research, management and policy, *Hydrol. Earth Syst. Sci.*, **15**, 471-504, doi: https://doi.org/10.5194/hess-15-471-2011.

Weber, T., Corotan, A., Hutchinson, B., Kravitz, B., and Link, R. (2020) Technical note: Deep learning for creating surrogate models of precipitation in Earth system models, *Atmos. Chem. Phys.*, **20**, 2303–2317, doi: https://doi.org/10.5194/acp-20-2303-2020.

Wu, C., X. Liu, Z. Lin, A.M. Rhoades, P.A. Ullrich, C.M. Zarzycki, Z. Lu & S.R. Rahimi-Esfarjani (2017). Exploring a variable-resolution approach for simulating regional climate in the Rocky Mountain region using the VR-CESM, *J. Geophys. Res. Atmos.*, 122, 10,939–10,965. doi: https://doi.org/10.1002/2017JD027008.