Water Cycle-Driven Infectious Diseases as Multiscale, Reliable, Continuously Updating Water Cycle Sensors

Authors

Amy Powell[1,*], Erin C.S. Acquesta[2,*], Warren L. Davis[3,*], Jeffrey J. Nichol[3,*,&], Irina Tezaur[4,+], Kara Peterson[5,*], Susan Rempe[6,*], Jose Gabriel Huerta[7,*]

[1] System Design and Architecture, [2] Computational Decision Science, [3] Scalable Analysis and Visualization, [4] Quantitative Modeling and Analysis, [5] Computational Mathematics, [6]Chemical, Biological, Radiological and Nuclear Defense and Energy Technologies, [7]Statistical Sciences; [*] Sandia National Laboratories, Albuquerque, NM USA; [+] Sandia National Laboratories Livermore, CA, USA; [&] University of New Mexico, Department of Computer Science, Albuquerque, NM USA

Focal Area(s)

We propose a novel synthesis of observational and simulated data (climatological and biological) to enhance understanding of the real-world interplay between climate (here, the water cycle) and the epidemiology of water cycle - driven infectious disease. Aligning with Focal Area 2, *predictive modeling using AI techniques,* we will develop systems of hierarchical models to discover latent features of the water cycle. Our approach will leverage state-of-the-art in Artificial Intelligence (AI) to measure the degree to which climate change-driven shifts in water cycle can be predicted by supplementing sparse and irregular climate data with water cycle – driven infectious disease resources.

Science Challenge

Rapid climate change introduces ahistorical and cryptic factors of influence that existing climate models are unable to capture and forecast. Lack of knowledge about emerging climate states, together with limited, sparse, and irregular data, highlight the need to augment E3SM with biosphere information[1]. We will mine the relatioship between climate change-driven perturbations in the water cycle together with the epidemiology of water cycle-driven infectious diseases, such as cholera (*Vibrio cholerae*), dengue fever (Dengue virus) and malaria (*Plasmodium falciparum*). The well-understood chain of causality provides a framework for discovering latent water cycle trends driving increasing disease outbreaks. Our framework promises disruptive insight into climate model uncertainty using calibration indicated by inferred latent water cycle variables.

Our scientific goals are to: 1) discover latent features of the regional, global and seasonal and multiannual water cycle through the lens of water cycle-driven transmission of infectious disease, 2) determine water cycle quantities of interest (QoI) that provide predictive power for infectious disease outbreaks, and 3) measure climate model-form epistemic uncertainty.

Our proposed efforts are aligned with the "Integrated Water Cycle," "Drivers and Responses in the Earth System," and "Data-Model Integration" scientific challenges highlighted in the Earth and

DOI: https://doi.org/10.2172/1769797
DOE BER, AI4ESP 2021 #AI4ESP1106 (https://www.ai4esp.org)

Environmental Systems Science Division Strategic Plan. The proposed R&D also supports goals summarized in the "Biological Systems Science Division Strategic Plan[2]": "Develop the enabling computational, visualization, and characterization capabilities to integrate genomic data with functional information on biological processes". Using E3SM climate simulation datasets to transform our understanding of water cycle-driven infectious disease will provide a framework for formal design of experiments (*e.g.*, genomic surveillance) to verify causal links between climate and epidemiology, to move towards state-of-the-art coupled analyses of correlated complex systems.

Rationale

Water cycle-driven disease outbreaks (and associated case counts) are effectively multiscale "sensors," potentially useful for understanding interannual variation, multiannual phenomena (*e.g.*, El Niño Southern Oscillation), departures from climatic norms and emerging extremes. These infectious disease "sensors" are never switched off, are globally distributed, and are always being "updated," both with new information about water cycle status and climate, and also with new "sensor prototypes" constantly being added by evolution.

Climate-change-driven perturbation of the water cycle across scales is ubiquitous, and there are modeling/simulation frameworks in place to understand and predict the accelerating physical changes[3]. In contrast, how emerging climate dynamics will impact global infectious disease burden is far from obvious, but at least two decade's worth of serious scientific consideration points to intensifying climate change ushering in an era of pandemics[4].

To begin addressing these expertise and infrastructure gaps, we will (1) discover new water cycle features by augmenting data models with epidemiological information, (2) enrich epidemiological models with E3SM water cycle output for different climate scenarios to expose and inventory latent water cycle variables. New expertise and tools (that interact with established resources, such as E3SM Diagnostics[1]), will provide a pioneering first framework for predicting shifts and extremes in the water cycle, and discovery of novel water cycle attributes and behaviors. This framework can also be used to develop quantitative, risk-qualified predictions of outbreak potential using newly constrained, calibrated and

[1] https://github.com/E3SM-Project/e3sm_diags

elaborated climate models**.** Our goals unify BER's EESSD (climate) and BSSD (biology) missions.

Narrative

Cholera, dengue fever, malaria and Zika are but a few examples of disease with strong sensitivity to the water cycle [5-7]. Higher annual mean temperatures, sea level rise, more frequent and intense storms, and altered precipitation and flooding regimes of the past 75 years have been accompanied by marked increase in water – associated disease outbreaks. Our own recent work in climate, together with expertise in epidemiological modeling, pathogenesis, machine learning (ML) and data analytics, bring together a unique capability for discovering latent water cycle features useful for calibrating, constraining and elaborating the next generation of climate models. These newly updated climate models will in turn be used to forecast emerging and re-emerging, climate-mediated infectious disease[8,9].

We will develop climate-informed epidemiological models to simulate water cycle impacts on rates of disease transmission, using time-dependent parameterization of force-of-infection function in compartmental models. We will employ deep neural networks (DNNs) and recurrent neural networks (RNNs) to: (1) calibrate and inform mechanistic epidemiological models using climate data (observational and simulated), and (2) detect/predict extreme water cycle events (*e.g.*, precipitation, flooding, cyclones, changes in major ocean currents), and correlate them with disease outbreaks. We will prototype integration of epidemiological models with E3SM, making the DOE flagship climate code a critical tool for quantitative estimation of infectious disease occurence. In parallel, we will explore using ML to advance multi-scale and multi-modal modeling. By supporting sensitivity analysis and parameter optimization, ML is useful for inferring the underlying dynamics of a physical system. These inferences will be central to enabling integration of climate and disease models, an idea that remains largely unexplored[6]. Towards quantifying uncertainties in water cycle-driven disease transmission, we will perform independent studies to learn how uncertainty propagates in these models, enabling rigorous prediction of both unexpected water cycle behavior and outbreak potential. We will use Bayesian Inference together with disease case counts to discover latent water cycle variables that only become evident in interaction with the biosphere.

To frame workshop goals, we develop exemplars representing key transmission classes of water cycle-driven infectious disease. We will use relevant disease and environmental data, including information from Next Generation Ecosystem Experiments (NGEE), Environmental Molecular Sciences Laboratory (EMSL), the Joint Genome Institute (JGI) and the National Institutes of Health (NIH), among others, together with E3SM climate simulations to develop integrated, sophisticated predictive views of the water cycle and the diseases it drives.

Suggested Partners/Experts (Optional)

- Aaron Bernstein and Renee Salas, Climate Change & Health Initiative (https://globalhealth.harvard.edu/domains/cc-health/team/)
- Professor Melanie Moses, University of New Mexico (https://moseslab.cs.unm.edu/)
- Professor Helen Wearing, University of New Mexico (https://biology.unm.edu/core-faculty/wearing.shtml)
- Cosmin Safta, Sandia National Laboratories
- Joint Genome Institute (Igor Grigoriev, Susannah Tringe, et al.)
- Professor Micaela Martinez, Columbia University (https://www.publichealth.columbia.edu/people/our-faculty/mem2352; mem2352@cumc.columbia.edu)

References (Optional)

1 Stevens, R. *et al.* AI for Science. Report No. ANL-20/17; Other: 158802 United States 10.2172/1604756 Other: 158802 ANL English, Medium: ED (; Argonne National Lab. (ANL), Argonne, IL (United States), 2020).

2 Anderson, T. *et al.* Biological Systems Science Division Strategic Plan. (2015).

3 Golaz, J.-C. *et al.* The DOE E3SM Coupled Model Version 1: Overview and Evaluation at Standard Resolution. *Journal of Advances in Modeling Earth Systems* **11**, 2089-2129, doi:https://doi.org/10.1029/2018MS001603 (2019).

4 Morens, D. M. & Fauci, A. S. Emerging Pandemic Diseases: How We Got to COVID-19. *Cell* **182**, 1077-1092, doi:https://doi.org/10.1016/j.cell.2020.08.021 (2020).

5 Council, N. R. *Under the Weather: Climate, Ecosystems, and Infectious Disease*. (The National Academies Press, 2001).

6 Martinez, M. E. The calendar of epidemics: Seasonal cycles of infectious diseases. *PLOS Pathogens* **14**, e1007327, doi:10.1371/journal.ppat.1007327 (2018).

7 Yang, K. *et al.* Global Distribution of Outbreaks of Water-Associated Infectious Diseases. *PLOS Neglected Tropical Diseases* **6**, e1483, doi:10.1371/journal.pntd.0001483 (2012).

8 Powell A, T. I., Davis W,Peterson K,Rempe S,Smallwood C,Roesler E. Predicting Future Disease Burden in a Rapidly Changing Climate. (Sandia National Laboratories, 2020).

9 Peterson K, P. A., Tezaur I, Roesler E, Nichol J, Peterson M, Davis W, Jakeman J, Stracuzzi D, Bull D. Arctic Tipping Points Triggering Global Change LDRD Final Report. (Sandia National Laboratories, 2020).