Understanding Climate Change's Impact on the Forest Ecosystem: Developing a Reproducible Machine Learning Approach to Ecohydrologic Model Outputs

PROPOSERS

Alex Clippinger | Master of Environmental Data Science Candidate, 2022 | Bren School of Environmental Science & Management | alexclippinger@bren.ucsb.edu | 860-326-6795

Peter Menzies | Master of Environmental Data Science Candidate, 2022 | Bren School of Environmental Science & Management | petermenzies@bren.ucsb.edu | 919-623-9987

CLIENT & FACULTY SUPPORT

Naomi Tague, PhD | Professor - Hydrology | Bren School of Environmental Science & Management | ctague@bren.ucsb.edu | 805-893-8579

OBJECTIVES

The objective of this project is to extract meaningful insights into the effect of climate change on forest ecosystem health using model outputs from the Tague Team Lab's Regional Hydro-Ecologic Simulation System (RHESSys). RHESSys is a state of the art watershed scale model of interactions among climate, hydrology, biogeochemical cycling, and forest growth. Extracting meaningful patterns from the large volume of output this model generates is challenging and often relies on the user having deep experience with earth system modelling. In this project, we will use machine learning and other tools to simplify the post-processing of model output. We will then visualize findings to make them more accessible for forest managers and to educate the public on how forest health is changing with climate. Specific objectives include:

1. Develop a machine learning algorithm to predict impacts on forest ecosystem health as a result of different climate warming scenarios. RHESSys model output for a forested watershed in the California Sierra will be used.
2. Create a reproducible workflow that can be leveraged for applications of RHESSys to other watersheds and climate scenarios.
3. Construct an interactive web application that decision-makers and the public can use to visualize and communicate relationships between climate change and watershed health derived from model output.

SIGNIFICANCE

Ecohydrologic models are core tools when investigating how climate can influence water, carbon, and energy cycles in natural and urbanizing landscapes. RHESSys and other similar models are becoming increasingly sophisticated due to advances in software engineering, increased data availability, and a growing body of knowledge within the field of ecohydrology. Due to the complexity of these models, interpreting the high volume of data they output is a key challenge—both for scientists and the public. Machine learning algorithms have the capability of quantifying patterns in data that are not revealed by statistical analysis. For example, machine learning regression techniques, such as random forest, provide significantly better predictions of streamflow compared with physical models (Guswa et al. 2020). A 2021 study conducted by Burke et al. demonstrated the viability of using random forests to analyze RHESSys output. Despite this, there remains limited guidance on suitable machine learning methods in ecohydrology, requiring collaboration between computer scientists and earth scientists (Guswa et al. 2020). As a result, the vast majority of studies using RHESSys to analyze watershed processes are only assessed using statistical summaries and time-series data visualizations. Utilizing the predictive power of machine learning algorithms, these projections can be delivered with greater precision and confidence.

This project will make RHESSys output more usable for a wider range of researchers by creating an accessible and reproducible machine learning workflow. The potential results of this project include an increase in the number of users analyzing watersheds globally and more robust data on these systems. In addition, detailed documentation will help bridge the gap between scientific research and public-facing communication. This supports the broader goals of spurring involvement in local natural resource protection and creating a deeper connection between society and the natural systems that we depend on.

BACKGROUND

RHESSys is a GIS-based model designed to simulate ecological and hydrological process interactions at varying spatial scales. The model simulates the effects of different climate trajectories on multiple variables (such as transpiration, soil and vegetation carbon sequestration, net primary productivity, and streamflow) over space and time. It is a powerful tool to better understand ecohydrologic systems and to predict how particular systems may respond to stressors like climate change (Tague and Band 2004). RHESSys has continually evolved since its creation in 1993—recent developments include integration of a fire model and new advances to represent forest carbon cycling responses to drought. With open source code on GitHub, RHESSys has been used by research teams worldwide in analyzing watershed carbon cycling, vegetation health, and hydrology to inform more sustainable management practices (Chen et al. 2020).

The model output this project will analyze was generated using input parameters that reflect the Sagehen Creek Basin. Spanning roughly 3600 hectares, Sagehen Creek Basin is located on the east slope of the northern Sierra Nevada section of the Sierra-Cascade Mountains. Much of the Basin is experimental forest managed by the UC Natural Reserve System and is home to the Sagehen Creek Field Station. Sagehen Creek is a headwater tributary flowing East into Stampede Reservoir on the Truckee River (Mast & Clow 2000).

California's upper watersheds in the Sierra Nevada are a critical source of water supply for the state's 39 million residents (Null et al. 2010). Assessing the responses of these areas to climate change is therefore required in watershed management and planning. Additionally, the state's forests sequester roughly 33 million metric tons of carbon each year (Gray et al. 2018). Sagehen, which consists primarily of land within Tahoe National Forest along with several hundred hectares of land owned by a private lumber company, is representative of much of the forest and watersheds of California. Thus, gaining a better understanding of climate impacts on this particular site has much broader management implications.

EQUITY

The Northeast Sierra Nevada region, where the project site is located, has provided natural resources for indigenous and rural communities in the area for centuries (Vale 2013). As a headwater tributary to large and small municipal drinking water supplies, protection of this ecosystem is important to the public health of a diverse population. A primary goal of this project is to provide all stakeholders with equal access to insights that inform the decision-making process surrounding watershed and forest management. Through interactive visualizations, we hope to address the need for scientific communication that bridges the information gap between experts and the public's understanding of climate change (Chalecki 2000). Additionally, as a project focused on reproducibility, our methods can be applied to other forest ecosystems near vulnerable communities.

DATA

This project will utilize existing RHESSys output associated with the Sagehen Creek Field Station in Truckee, California. The model output is in CSV format and consists of 30-meter gridded cells containing carbon and water information. The project will build on a RHESSys implementation for Sagehen that is part of a National Science Foundation Critical Zone Observatory project. The model implementation includes multiple drought scenarios, and includes output that examines model uncertainty and parameter sensitivity. The following is a link to the dataset with accompanying metadata (Graup 2021): http://www.hydroshare.org/resource/5dfb547718ae4125b122ea890f60047b

POSSIBLE APPROACHES

The general approach will analyze RHESSys model output by creating a machine learning algorithm to evaluate how forest water use and carbon cycling may respond to climate change. This includes identifying key sensitivities (e.g. which climate drivers are most important? At what time scale?), impacts of climate uncertainty, and significant spatial patterns. These predictions will then be visualized. More specifically, the approach will include the following tasks:

1. Develop a machine learning method, likely a regression model such as random forest. Forest carbon uptake, or net primary productivity (NPP), will be used as an indicator of forest health. The machine learning method will be used to determine how different climate variables, forest parameters, and spatial location (e.g. slope, elevation, soil depth) influence NPP and it's variability. If time permits, the climate sensitivity of other ecosystem service outputs, such as streamflow, will be analyzed.
2. Create a reproducible, comprehensively documented workflow to make our machine learning method transferable to other RHESSys applications. This project has the potential to be a framework for future projects using RHESSys, both from the Tague Lab and other RHESSys users.
3. Create an interactive and accessible web application to visualize what our analysis reveals about climate sensitivity of forest health. Machine learning will identify the most salient climate indicators and how these influence forest health—users will then have the ability to manipulate these variables to explore different outcomes and drivers of forest sensitivity to climate change.

DELIVERABLES

1. **Machine learning code** that is well-documented and reproducible for simulation of this specific site. The code will be open source and publicly available on GitHub.
2. **Machine learning workflow** that is well-annotated and supported with educational material to extend utility of code to similar projects using RHESSys data. This will take the form of an RMarkdown file that will be open source and publicly available on GitHub.

3. **Interactive web application**, likely in the form of an RShiny app, designed to communicate the importance of ecosystem health and the potential impact of climate change. This can initially be hosted on Dr. Naomi Tague's website with the ultimate goal of hosting on NSF's HydroLearn toolkit for further visibility.

BUDGET

It is not expected that the proposed project will require funding.

REFERENCES

Burke, William D., Christina Tague, Maureen C. Kennedy, and Max A. Moritz. 2021. "Understanding How Fuel Treatments Interact With Climate and Biophysical Setting to Affect Fire, Water, and Forest Health: A Process-Based Modeling Approach." Frontiers in Forests and Global Change 3: 143. https://doi.org/10.3389/ffgc.2020.591162.

Chalecki, E. L. 2000. "Same Planet, Different Worlds : The Climate Change Information Gap," June. https://www.osti.gov/etdeweb/biblio/20191734.

Chen, Benxin, Zhifeng Liu, Chunyang He, Hui Peng, Pei Xia, and Yu Nie. 2020. "The Regional Hydro-Ecological Simulation System for 30 Years: A Systematic Review." Water 12 (10): 2878. https://doi.org/10.3390/w12102878

Graup, L. 2021. RHESSys model output for Sagehen Creek hillslope, HydroShare, http://www.hydroshare.org/resource/5dfb547718ae4125b122ea890f60047b

Gray, A., Christensen, G., & Kuegler, O. 2018. "Working with Cal Fire to monitor carbon sequestration in California." Retrieved from https://www.fs.fed.us/research/highlights/highlights_display.php?in_high_id=1428

Guswa, Andrew J., Doerthe Tetzlaff, John S. Selker, Darryl E. Carlyle-Moses, Elizabeth W. Boyer, Michael Bruen, Carles Cayuela, et al. 2020. "Advancing Ecohydrology in the 21st Century: A Convergence of Opportunities." Ecohydrology 13 (4): e2208. https://doi.org/10.1002/eco.2208.

Mast, M.A., and Clow, David W. 2000. "Environmental characteristics and water-quality of Hydrologic Benchmark Network stations in the Western United States." U.S. Geological Survey Circular 1173-D, 115 p., https://pubs.usgs.gov/circ/circ1173/circ1173d/pdf-d/chapter04d.pdf.

Null, Sarah E., Joshua H. Viers, and Jeffrey F. Mount. 2010. "Hydrologic Response and Watershed Sensitivity to Climate Warming in California's Sierra Nevada." PLOS ONE 5 (4): e9932. https://doi.org/10.1371/journal.pone.0009932.

"RHESSys Homepage." n.d. Accessed October 6, 2021. http://fiesta.bren.ucsb.edu/~rhessys/.

Tague, C. L., and L. E. Band. 2004. “RHESSys: Regional Hydro-Ecologic Simulation System—An Object-Oriented Approach to Spatially Distributed Modeling of Carbon, Water, and Nutrient Cycling.” Earth Interactions 8 (19): 1–42. https://doi.org/10.1175/1087-3562(2004)8<1:RRHSSO>2.0.CO;2.

Tague, Christina, and James Frew. 2021. “Visualization and Ecohydrologic Models: Opening the Box.” Hydrological Processes 35 (1): e13991. https://doi.org/10.1002/hyp.13991.

Vale, Thomas. 2013. Fire, Native Peoples, and the Natural Landscape. Island Press.

UNIVERSITY OF CALIFORNIA, SANTA BARBARA

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO SANTA BARBARA • SANTA CRUZ

DONALD BREN SCHOOL OF ENVIRONMENTAL SCIENCE & MANAGEMENT
CHRISTINA TAGUE, PROFESSOR, ROOM 4516 BREN HALL

SANTA BARBARA, CA 93106-5131
http://www.bren.ucsb.edu/

October 15, 2021

I am writing to confirm my support for the MEDS capstone proposal entitled: "**Understanding Climate Change's Impact on the Forest Ecosystem: Developing a Reproducible Machine Learning Approach to Ecohydrologic Model Outputs"**. The proposal will undertake a machine learning analysis of outputs from RHESSys (Regional hydrologic ecosystem simulation system) applied to the Sagehen Creek Watershed, a forested watershed in the Central California Sierra. The goal of the analysis is to reduce the complexity of model output – that includes spatial and time series data on forest health and water resource variable under different climate scenarios. The project will generate workflows and also develop education/communication related visualizations of relationships between key forest health and climate variables that emerge from the machine learning analysis.

I confirm that the TagueTeamLab will provide all of the necessary RHESSys simulation output. This will include model estimates of forest net primary productivity, carbon sequestration, evaporation, transpiration for spatial patches distributed across the watershed and watershed scale streamflow. Model estimates will be provided at a daily time step for historic climate (1960-2010) and at least two projected future climate scenarios. Relevant climate input variables (air temperature, precipitation, humidity) will also be provided along with estimates of output uncertainty due to uncertainty in input parameters. A preliminary dataset is included in the proposal. Additional simulations may be run if required and other variables (such as leaf area, rooting depth or soil moisture) can be provided if needed.

I will serve as the primary point of contact. Louis Graup, a PhD student will also be available for questions about RHESSys simulation results for Sagehen Creek.

C Tague

.
Dr. Christina (Naomi) Tague

PHONE: (805) 893-8579 •ctague@bren.ucsb.edu • FAX: (805) 893-7612