Secure Water Future
Water Hack Challenge
April 21-23, 2023

Three challenge tracks:
Liquid-Flood
Solid-Snow
Gas-Evapotranspiration

Each team competes in one track.

Sponsored by:

Made possible by:

STRUCTURE AND RULES

Regardless of which problem statement you or your team expressed interest in during the registration phase of this Water Hack Challenge, once you have read through the three problem statements detailed below, please choose one “track” in which to participate. The three tracks are **liquid-flood**, **solid-snow**, and **gas-ET**. Winners will be chosen irrespective of the track (i.e., there is not one winner per track).

Before deciding on a track, we recommend the following: 1) read all three problem statements; 2) watch any recommended introduction videos; and 3) discuss with your team which challenge you find the most compelling. Once ready, declare your team’s hack track using the associated role-select in Discord and start hacking!

You and your team are welcome to deviate from the described challenges as long as you utilize one or more of the provided data sets. In other words, pursuing a hack approach different from the recommended challenge in the hack problem statement is allowed as long as the original data are used. You may also bring in any other datasets that help you solve interesting problems or make great visualizations and applications.

If you have questions or need general guidance, please contact a mentor during posted mentor office hours. We are here to help you find creative solutions!

Your hacks will help the State of California better manage its water by better balancing supply and demand through data-driven decision-making. Innovative solutions to water measurement and management are key to a sustainable future in California.

Your participation in this Water Hack Challenge has the potential to transform current practices and technologies!

PARTICIPATION AND JUDGING

1. Participants must create a repository in the SWF Water Hack Challenge GitHub organization and commit their code for review before judging begins. Code may run on Windows, Mac, or Linux, but compiled languages must include a binary (either committed or attached as a "release" upload to the GitHub repository) and all languages should have steps needed to run the code documented in the README or a location linked from the README.
2. Projects using one of the algorithmic challenges should include measures of algorithm performance in their READMEs – including performance criteria appropriate for the chosen algorithm type, as well as a confusion matrix for classification tasks, and access to both training and validation data.
3. Presentations and judging will be from 12:30 pm to 3:00 pm on Sunday, April 23. Judging will commence in a "snowball" format wherein each judge will rank initial groups using the scoring rubric, which is included below. The top six teams selected by each judge will go on for a second round of judging. Those competitors will present to a judge who has not previously seen their work. Judges will then select their top two teams. Those competitors will present to the final judge who has not previously seen their work. Results will be based on the cumulative total from all three judges.
4. Delivery of awards may be subject to verification of code and algorithm performance after the end of the hackathon. Awards will require additional processing time commensurate with team composition and preferences.

Scoring Rubric for SWF Water Hack Challenge

Team Name: ____

Circle Track: Liquid-Flood, Solid-Snow, Gas-Evapotranspiration

Judge's Name: ____

Total Score: ________________

Category	Guiding Questions	Score (1-5 Points)
1. Understanding of the Problem	How well has the problem been defined?	
2. Innovation	Does the product introduce a new technical, analytical, or visual approach or perspective?	
3. Technology	Has skill been demonstrated? Was there illustrated use of unique coding implementations?	
4. Presentation and Teamwork	How well was the team able to communicate their ideas and demonstrate their team strengths and individual contributions?	
5. Feasibility	How easily could decision-makers adopt the algorithm or visualization?	

Index of Key Terms and Concepts

Automated Snow Sensors
These devices are designed to automatically detect and measure snow depth, snow-water equivalent (water content of the snowpack), and other meteorological parameters. They are used to monitor snow conditions in real-time and are especially useful for managing water resources in areas where snowmelt is an important water source.

Evapotranspiration (ET)
This is the process by which water is transferred from the land to the atmosphere through evaporation from the soil and transpiration from plants. It is an important factor in the water cycle and can affect agricultural productivity, water availability, and the climate.

Meteorological Variables
These are measurements of various weather conditions, such as temperature, humidity, pressure, wind speed and direction, and precipitation. These variables are used to study and predict weather patterns and to inform various industries, such as agriculture, transportation, and energy.

Fallow
This refers to a period when land is left unplanted and is allowed to rest to improve soil quality and conserve moisture. It is a common practice in agriculture and can help to prevent soil erosion and reduce water usage.

Precipitation
This refers to any form of water that falls from the sky, such as rain, snow, sleet, or hail. Precipitation is an important component of the water cycle and is critical for sustaining life on Earth.

Runoff
This is the movement of water over the land's surface and into streams, rivers, and other bodies of water. It is an important part of the water cycle and can impact water quality and availability.

Snow Observations
These are measurements of snow-related parameters, such as snow depth, snow density, and snow water equivalent. They are used to monitor snow conditions and to help manage water resources in areas where snowmelt is an important source of water.

Snow Water Equivalent
This is the water content of the snow measured in terms of its liquid water equivalent. It is used to determine the potential amount of water available once the snow melts.

Stage Height

This is the height of water in a river or stream above a specified reference point. It is a common measurement used to monitor water levels and to inform water management decisions.

Streamflow

This is the volume of water that is flowing in a river or stream at a given point in time. It is an important component of the water cycle and is used to inform water management decisions.

Water Balance

This concept describes the relationship between the inputs and outputs of water in a particular system, such as a watershed or ecosystem. It is important for understanding water availability and for managing water resources.

Water Management

This refers to the process of managing water resources, including the allocation, conservation, and distribution of water. It is an important practice for ensuring sustainable water use and addressing water-related challenges, such as drought and water scarcity.

Water Year

This is a period used to measure water availability and precipitation, typically from October 1st to September 30th of the following year. It is an important tool for water management and planning.

LIQUID PROBLEM STATEMENT

Real-World Context

Bear Creek winds through the City of Merced, where UC Merced and Merced College are located. In early January 2023, flooding displaced upwards of 1,600 residents, many of whom live along the waterway. The flooding garnered national attention and prompted questions about how forecasting could assist with future potential disasters.

Floods are not always harmful, but they can be dangerous when they put people and property at risk. Scientists suggest using floodplains in a strategic manner to reduce the danger of floods while also providing additional benefits like groundwater recharge, ecosystem habitat creation, and carbon storage. See the linked videos if you would like to know more.

Data Sets

1. Daily maximum streamflow and daily maximum stage height for Bear Creek @ McKee, data from 1997-current.
2. Observed daily precipitation totals over the upstream contributing watershed to Bear Creek @ McKee during 1979-2023.
3. Time series of daily precipitation from 20 different climate models run during the 2030-2099 period.

Potential Challenges

1) Algorithms

Will flood risk in Merced change in the coming decades with climate change? This challenge asks you to develop an approach for assessing flood risk along Bear Creek in Merced for the next 30 years (2031-2060) using future climate scenarios. Your task is to develop a model that predicts streamflow, as measured by the water height on Bear Creek, using weather datasets.

To accomplish this task, you will need a model that estimates runoff – the water that flows over the ground during a rainstorm – from precipitation data. Models of precipitation-runoff take into account both immediate and delayed effects. This is because saturated soils can cause water to flow quickly on the ground. Runoff is affected by the amount of daily rain, as well as the overall amount of precipitation and water balance in the preceding days. Once you create a precipitation-runoff model, you can use precipitation from modeled future climate conditions in your model to evaluate risk. However, because the future trajectory of precipitation is uncertain due to climate change, we recommend using multiple future scenarios to understand the risks. We specifically want to know: ***what is the risk of a flood comparable to the 2023 event for the 2031-2060 period and how has this changed from the contemporary period (i.e., recent years)?***

2) Visualization

Model results, such as those from climate models, come with some uncertainty based on the model's assumptions. One way to understand the uncertainty is by visualizing results from many models or model runs. Visualizing uncertainty is an art, and can take many potential forms, including charts, infographics, animations, videos, and more. For this challenge, produce a visualization that helps the viewer understand the trajectory of change and how certain or uncertain those model predictions are.

Further Reading and Information

Public Policy Institute of California, short video on flood management in California.

Whipple, A.A., Viers, J.H. and Dahlke, H.E., 2017. Flood regime typology for floodplain ecosystem management as applied to the unregulated Cosumnes River of California, United States. *Ecohydrology*, *10*(5), p.e1817. https://onlinelibrary.wiley.com/doi/full/10.1002/eco.1817

Viers, J.H. and N.R. Santos, 2014. Hydrolapse Videography: A Coupled Hydroinformatic Stack For Improved Visual Assessment Of River Dynamics. Proceedings of the 11th International Conference on Hydroinformatics. *CUNY Academic Works.* https://academicworks.cuny.edu/cc_conf_hic/51

This interview with UC Merced Professor Josh Viers discusses some of these recent approaches to flood management, as well as this in-depth video on hydroinformatics.

SOLID PROBLEM STATEMENT

Real-World Context

The [2023 water year will be in the record books for the most snow]in California's data recorded history, but how we use snow data is a big challenge for scientists and water managers. Snow observations are important for water management in the western US because snow is a major water source. Snow falls on the ground during the winter and melts in the spring and summer, providing water for drinking, agriculture, recreation, and the environment. Accurate measurements of snow depth and water content are used by water managers, policymakers, and other stakeholders to make decisions about how to allocate water, prepare for droughts, and prevent floods.

Automated snow sensors are an important part of water monitoring systems. In the past, snow observations were taken manually once or twice a month, which was time-consuming and labor-intensive. Automated sensors can provide sub-daily measurements that can improve monitoring and help with short-term decision-making about reservoir operations. However, automated sensors also have potential problems, such as the possibility of faulty data. It is important to be aware of these problems to avoid making decisions based on bad data.

Video Introduction

https://ucmerced.box.com/s/wqt53uxbocwa056fckxs4jzkyh0v1mqg

Data Sets

The California Department of Water Resources has about 100 automated snow observing stations in mountains in the state, primarily in the Sierra Nevada. The sensors measure snow water content, precipitation, and temperature. This challenge focuses on the Tuolumne Meadows DWR station in Yosemite at an elevation of 8600 feet. We provide a csv file of daily cumulative precipitation since October 1st, snow water equivalent, and mean daily air temperature from Oct 1999-March 2023.

Potential Challenge

There are numerous quantitative approaches for quality-controlling datasets to screen for data errors to ensure that bad data does not inform water management decisions. In this challenge, you will develop algorithms for flagging questionable data in daily snow observations. One of the challenges is in devising screening methods that correctly identify poor quality or suspect data without incorrectly identifying legitimate data (e.g., flagging good data as erroneous). Examples of potential approaches for doing this involve: (a) a machine learning approach using data from valid data to provide a check on subsequent data; (b) developing stringent physical and statistical measures on daily changes in snow water equivalent. Reservoir operators reported that the snow water equivalent data was good for the first 12 years (Oct 1999 - Sep 2011), after which the data showed a tendency to show errors. Your task is to devise an approach for quality controlling the snow water equivalent data from Oct 2011-March 2023. For this exercise, you can assume that the temperature and cumulative precipitation data are correct and focus on algorithms for 'cleaning' the snow data. We specifically want to know: can you develop an approach for identifying suspect snow water equivalent data and apply it to the Oct 2011-March 2023 observations?

Further Reading and Information

Serreze, M.C., Clark, M.P., Armstrong, R.L., McGinnis, D.A. and Pulwarty, R.S., 1999. Characteristics of the western United States snowpack from snowpack telemetry (SNOTEL) data. Water Resources Research, 35(7), pp.2145-2160.

Public Policy Institute of California, short video on the role of Snowpack in feeding our Headwaters.

Yarnell, Sarah. California Water Virtual Tour: Sierra Nevada Rivers.

GAS PROBLEM STATEMENT

Real-World Context

Water managers need to track the origin and destination of water to determine how much water is available for use. This may appear to be a straightforward task, but water usage takes various forms, and some cannot be measured easily or directly. For instance, water used in agriculture can flow into rivers or seep into groundwater. It can also evaporate or transpire through plants into the atmosphere.

The combined evaporation of water from the surface of the Earth and transpiration through plants form an important process called evapotranspiration (ET). Knowing how much water is consumed by ET is essential for determining the available water. Water that goes into the atmosphere through ET is known as "consumptive use" because it cannot be reused or recycled by the same individuals and ecosystems that were using it.

ET data can be readily estimated across wide areas via satellite observations, allowing it to fill in important variables in water balance equations that estimate sources and destinations of water in a given area. Satellite observations, when coupled with field-based sensors, allow for widespread though coarse estimations of "actual" ET as opposed to "potential" ET, which is the theoretical maximum ET based on weather and atmospheric conditions. More generally, ET is a combination of air vs plant pressure as well as temperature, relative humidity, and wind speed.

Data Sets

This challenge involves timeseries weather and evapotranspiration information. These datasets will be provided in two forms – as timeseries entries in a SQLite database and also as spatial data (rasters). You may work on the challenge using only the timeseries data, or using both - the spatial data is there to provide more flexibility to teams with spatial data experience.

All data for the challenges below are included in this folder. See the README in the folder for more information on how to use the data, suggested tools, and more.

Potential Challenges

1) Algorithms 1

You will be given a partial year monthly time series of ET values (e.g April, May, and June) as well as past weather data, and projections of remaining seasonal weather data - with these data, your model should predict the remaining ET time series for the rest of the water year. To facilitate this, you will be given full-year timeseries data for two previous years in order to train potential models. This data has all been extracted for you so that you have the mean value for each variable/agricultural field/month combination in tabular formats. See the README in the data folder for more details.

Successful algorithms will produce reasonable estimates for ET that minimize total error across all agricultural fields (that is, values for some agricultural fields may be low, some may be high, but they will approximately cancel each other), and very high-performing algorithms will minimize the error in individual fields.

2) Algorithms 2

A common need is estimating what crops are grown in a field, and ET data can help us use satellite data to get landscape-scale cropping data. Given ET time-series data for agricultural fields and weather/climate information for the fields, develop a model that predicts the crop types into five categories: grasses and pasture, rice, perennial crops (trees and vines), annual crops, and not cropped (fallow).

As before, you will have two seasons of training data for 2018 and 2019 where you have both the timeseries data, as well as the crop classification, and one season with at least partial timeseries ET data, but no crop classification.

3) Data Visualization and Applications

Water managers frequently want to understand total water use in a district or region. Build an interactive tool that lets water managers quickly see total usage by the provided regional polygons, along with any other metrics of interest, based on the spatial field data and ET data provided for the Algorithms 2 challenge.

Further Reading and Information

https://openetdata.org/methodologies/

Melton, F.S., Huntington, J., Grimm, R., Herring, J., Hall, M., Rollison, D., Erickson, T., Allen, R., Anderson, M., Fisher, J.B. and Kilic, A., 2022. OpenET: Filling a critical data gap in water management for the western United States. *JAWRA Journal of the American Water Resources Association*, *58*(6), pp.971-994. https://onlinelibrary.wiley.com/doi/full/10.1111/1752-1688.12956

Kalua, M., Rallings, A.M., Booth, L., Medellín-Azuara, J., Carpin, S. and Viers, J.H., 2020. sUAS Remote Sensing of Vineyard Evapotranspiration Quantifies Spatiotemporal Uncertainty in Satellite-Borne ET Estimates. *Remote Sensing*, *12*(19), p.3251. https://www.mdpi.com/2072-4292/12/19/3251

Public Policy Institute of California, short video on drought impacts.