### Project Name

California Snowpack

### Project Team Members (list first and last name of each member)

Malachy Moran, Kayla Wopschall, Derrick Hee, Tilman Bayer

### Problem Statement

Current predictions of snow water equivalent (SWE) have either poor temporal or spatial control. As a result, it is difficult to predict the amount of snowfall in the Sierra Nevadas, which in turn makes it difficult to predict the amount of snow melt that makes it into catchment areas that feed critical agricultural practices. As climate change continues, having accurate predictions of water availability for agriculture is critical for mitigation and resource conservation and management.

### Target Customer / user

There are several customers of this project and more users. The main customers are two professors in the Environmental Science Department at UC Berkeley, Manuela Girotto and Paolo D'odorico, who currently work on SWE predictions through mathematical modeling.

Secondarily, there are many government agencies and agricultural entities that are the end customers of this modeling, as they will be utilizing the predictions to inform policy and practice. They will be the main users of these results.

### How big would be the impact?

If done well, the impact of this project could be quite transformative for the state of California in its efforts to manage and adapt to fluctuating snowfall and snow melt resources. This has far reaching impacts to the greater health of the California ecosystem, as well as real impact to the livelihoods of many California residents and business owners. Outside of California, there is significant impact to the agriculture supply chain as it relates to crops currently only available from California.

### Minimal Viable Product (MVP)

The MVP for this project would be a model that successfully outputs a snow water equivalent (SWE) by 1km squares that predicts the relative direction and scale of the hand measured SWE.

### What is the competitive landscape (existing solutions or approaches)?

There are currently a variety of mathematical models to predict snow water equivalent (SWE), and several attempts to use those predictions to then inform the amount of water available in the Central Valley of CA. We will be leveraging insights gained from building the mathematical models to incorporate important variables in our CNN, as well and pulling in existing staged datasets.

In addition, since this is a potentially large gain outside of the Sierra Nevadas and CA, there is a large competition with over 700 teams to tackle this type of modeling for all of the Western US. At the conclusion of the competition, there will be a wide variety of solutions available for comparison.

### Mission

Our mission is to predict waterflow into California’s Central Valley Basin based on a two stage process: In stage one, we will try to model the Snow Water Equivalent (SWE) in the Sierra Nevadas and stage two we will try to use the outputs of stage one to predict water outflow.

### Data sets

Data sets utilized in this project can all be found here:

<https://www.drivendata.org/competitions/86/competition-reclamation-snow-water-dev/page/417/>

### Technical Approach

What data science algorithms are you intending to use, develop and build for the project?

* We intend to build and train a CNN/Transformer to predict SWE based on tabular and satellite imagery data
* We will implement an LSTM algorithm, utilizing previously trained CNN features as inputs alongside additional weather data

### What is the primary method of communication for you team (ie. Slack, email, etc)?

What is the maximum delay for responding to Slack/email from team members? Under what conditions does it apply? (e.g. weekends, etc)

* Slack
  + Ideally ~24 hr response time
* Google Drive
  + Shared documents and collaboration of presentations, papers, etc.
  + Shared data for early stage model development
  + Meeting notes
* Trello
  + Task management
* Google CoLab
  + Shared notebooks and code for early stage model development.
* GitHub
  + Shared repo (public or private) for production notebooks and code
  + Will be final repository for project docs, code, etc

### Hours per week of project work

Group members will spend an average of 15hrs/wk on the project; Due to the nature of the project for some individuals this time will be front loaded and others back loaded in the semester

### Weekly meeting schedule

Consult your individual schedules, and find days and times that will work for all of you. List the days and times of meetings, and what platform you will use.

* Zoom Meeting: ~1-2x per week depending on need. Need will be communicated via Slack
  + Thursday 6pm PST
  + Sat/Sun TDB

### Other collective responsibilities and conduct

Shared responsibilities include:

* Peer review of each others work
* Collaborative input on direction and improvements
* Communication on status and needs for individual responsibilities
* Participation in final presentation, communication, and dissemination of final project

### Individual members' roles and responsibilities

Individual responsibilities listed below. Please note that this is early in the project and these are currently fluid and likely to change. Everybody should feel free to take on work, it’s mostly about having point persons for each area who feel responsible to ensure the necessary work in that area gets done.

Malachy:

* Data wrangling - initial organization of SWE data and incorporation of satellite imagery and ground station data
* Lead on initial contest details and early modeling
* Paper Writing
* Lead on ML architecture research
* Containerization and model deployment

Tilman:

* Data wrangling - assistance with identifying additional data sources and staging data in usable format
* Lead on PM and organizational task tracking
* Literature review
* Paper writing
* Modeling

Derrick:

* Data wrangling - assistance with identifying additional data sources and staging data in usable format
* Incorporation of geologic/geographic flags relevant to station clustering for initial modeling
* Lead on main CNN model training/testing

Kayla:

* Data wrangling - Identification of variables used in previous research and tying to available data.
* At conclusion of stage 1, testing of resolution impacts for the Sierra Nevadas (1k vs. 30m squares)
* Stage 1 evaluation and accuracy
* Lead on paper and communication of results
* Subject matter expert
* Literature review

### Conflict resolution process

As a group, we’ve agreed to make every attempt to discuss conflict in person (e.g. zoom) in a respectful and productive environment. If needed, we will include a third party mediation (if the conflict is between two members this mediator would likely be another team member).

We will provide space for all members involved to voice concerns, opinions, and work together to find a collaborative solution. If we are unable to resolve a conflict, we will reach out to the course instructors to set up a time to discuss the nature of the issue and suggestions for how to proceed.

**I participated in formulating these standards, roles and procedures, and understand that I am obligated to abide by these terms and conditions.**

Name Malachy Moran

Name Kayla Wopschall

Name Derrick Hee

Name Tilman Bayer