A multi-discipline approach to seasonal drought prediction in California's Central Valley

Key Terms: groundwater, snowpack, ambient noise seismology, machine learning

Motivation: The California Central Valley (CCV) produces a quarter of US food on less than 1% of US farmland[1]. As the CCV has little summer rainfall (June, July, and August – JJA), it relies on meltwaters from Sierra Nevada Mountain Range (SNMR) snowpack alongside groundwater aquifers to maintain its prolific agricultural output. Recent CCV droughts have had billion dollar impacts on agricultural production and are expected to increase in frequency by 100% by 2100[2]. The lack of JJA rainfall in CCV implies that most of the water for summer irrigation is already stored the previous winter and spring in mountain snowpack and aquifers. The US Drought Monitor (USDM) bases its long-range drought forecasts on snowpack and weather patterns, assessing severity using a drought classification scheme, measured from D0 (abnormally dry) to D4 (exceptional drought)[3]. While a drought classification score lists potential impacts, it falls short of predicting whether water resources will be available locally, leading those reliant on water for irrigation or domestic use to guess if their wells will run dry. In CCV, an extensive record of the timing of well failures for CCV's ~100,000 wells has already been validated as a powerful metric for drought evaluation[4,6]. The dataset is available from the California Department of Water Resources.

Mountain snowpack, often measured as snow-water-equivalent (SWE), is measured at SNOTEL and Snow Course sites to provide daily snowpack measurements across the SNMR has been shown to be highly predictive of summer water supply[4,5]. As SNMR snowpack melts, the region of the CCV which the water ends up in is determined by the hydrologic unit (HUC) in which it fell as snow.

On the other hand, CCV groundwater monitoring is typically done by measuring the levels of specific aquifers using hydraulic head, which is prohibitively expensive and results in measurements that are spatially and temporally sparse, and thus unable to provide the agriculture industry with detailed enough measurements to estimate the extent of water resources available to them for the upcoming summer growing season[6]. Other methods for estimating groundwater changes include satellite mapping using tools that map surface deformation such as InSAR which fall short of measuring changes in underground structure[7]. A novel approach to groundwater volume prediction uses changes in seismic velocity (dv/v)[8,9]. The quantity dv/v can be calculated using cross correlation of the ambient seismic field and has been found to respond linearly to changes in groundwater[9].

Proposal: *I propose a novel approach for well failure prediction in the CCV by combining late-spring estimates of groundwater aquifer levels using seismic arrays with observational estimates of SWE across the SNMR.* This project, called SWEDV, will generate a comprehensive map of seismic velocity across CCV alongside estimates of spring SWE storage. I hypothesize that (1) seismic velocity changes can be used to predict well failure rate (WFR) in CCV, (2) lower SWE values across SNMR are positively correlated with WFR and that (3) SWE combined with seismic velocity changes in a machine learning (ML) model predict regional WFR better than the USDM 3-month forecast. I will focus on the time window 2000-present.

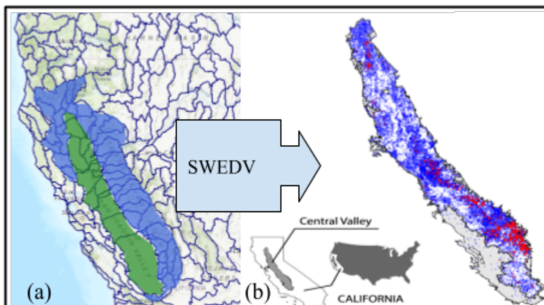

Figure 1: The SWEDV Model. (a) HUC8 regions: green shading denotes the CCV, blue denotes regions containing SNOTEL and Snow Course measuring sites [4]. Seismometers used will mostly drawn from CCV but may extend into the blue region. (b) Well locations (blue) and example failures (red) in CCV, see [6].

Aim 1: Calculate empirical map of seismic velocity changes and validate against well failure rates. I will use a portion of the 3000+ IRIS network stations located in and around CCV to compute spring (April) and summer (July) ambient noise cross-correlations (ANCCs) which can be used to estimate changes in seismic velocity between station pairs relative to a multi-year average[8,9]. To translate pairwise dv/v to a spatial grid of velocity structure changes we will distribute the dv/v value to grid cells within 2 or 3 kilometers of the direct path between station receivers and then report the total grid cell change as the average of all dv/v station pairs which overlap the cell, following the reconstruction technique of [8] and [10]. As I expect groundwater to respond linearly to dv/v changes, I will begin with linear and polynomial regression, using CCV HUC8 regions as fixed effects to capture regional differences, to assess the predictive power of dv/v changes on WFR. With accurate well failure predictions at high resolution ahead of the summer season, farms and households can adjust crop species and water usage.

Aim 2: Calculate SNMR SWE by HUC8 and assess correlation with well failure rates. I will estimate annual April 1st SWE at HUC8 or finer resolution using observations-based data from the over 250 SNOTEL and Snow Course (SC) measurement sites across the portion of the SNMR that feeds CCV[4]. To aggregate observations to estimates of HUC8 snowpack, I will either use a sample mean or employ a small-area estimation model to incorporate geographic information like elevation (Fig. 1a). For Aim 2, the response will be average July 1st average well failure rate by CCV HUC8. Using Multivariate Multiple Regression (MVMR) we can assess how predictive SNOTEL/SC measurements are of WFRs.

Aim 3: Combine seismic velocity changes with HUC8 SWE changes to predict well failure rates and compare to US Drought Monitor Predictions. Once I've validated that spring seismic velocity and SWE models correlate with summer WFRs, I will use the two datasets from Aim 1 and 2 to build a predictive model (SWEDV) to improve WFR estimation. Model inputs will be averaged dv/v change by CCV HUC8 region and estimated SWE by SNMR HUC8 region as defined in Aim 1 and 2, respectively. Our response variables will average July 1st summer well failure rate by HUC8, 3 months after our calculated prediction variables. Machine learning (ML) is particularly well suited to this task because we expect complex, and potentially highly non-linear interactions between the predictors and WFRs. I anticipate using expressive models for SWEDV, such as random forest regressors or neural networks, and expect to spend time tuning models to balance expressivity and parsimony.

While I expect to use SWEDV to predict WFR on a scale finer than HUC8 – extensions include estimating failure rates at the town or even well-level, validating its predictive power against the current standard, the USDM's 3-month prediction, will confirm the model's accuracy. To do this I will build a MVMR model to predict WFRs by HUC8 region from the USDM April prediction of July drought conditions (a 3-month prediction) by assigning each CCV HUC8 region a drought classification score by the classification which has the largest area in the region, no drought or D0-D4. By using a train-test split for our dataset, we can compare SWEDV to the current tool using metrics like classification accuracy and MSE.

Intellectual Merit: This project will improve WFR estimation for CCV by using machine learning across a terabyte-scale seismic dataset and observational SWE measurements. To my knowledge combining groundwater estimates from seismic data alongside climatological data to estimate drought intensity has never been done before. Furthermore, the scale at which we hope to use seismic data to estimate groundwater reserves is also novel and will produce a spatiotemporal map of dv/v estimates to infer groundwater at higher resolution than conventional hydraulic head methods. The project will exemplify how combining interdisciplinary datasets can improve prediction capabilities. The NSF Graduate Research Fellowship will provide the resources necessary to run and store the computationally intensive workflows.

Broader Impacts: As a proponent of open-science, I plan to make my workflow, seismic velocity datasets, and predictions for upcoming well failure rates publicly available, through online platforms (e.g., GitHub, Zenodo, AWS storage), conferences (e.g., AGU, AMS), and to policymakers and farmers who can use them to manage water resources, select drought resistant crops to plant in dry years, and support seasonal workforces. Focusing on groundwater *and* snowpack will improve prediction of summer drought severity in CCV as when both resources run dry, crop failures result in the loss of steady agricultural jobs, a degraded landscape, and increased food prices. While the impacts of drought are felt by all, disadvantaged communities, some of whom hold jobs in CCV agriculture, and who are most sensitive to food price increases, are likely to suffer most. Thus, while this research will help advance computational seismology and drought research, I hope to also use my funding as an NSF Fellow to ensure these predictions are shared with and accessible to the people who can apply it to address the water needs of the CCV's most vulnerable.

References: **[1]** California Water Science Center, USGS. **[2]** Swain, D. L. *et al. Nat. Clim. Change 8 427-433* (2018) **[3]** U.S. Drought Monitor, UNL. **[4]** National Water and Climate Center, NRCS, USDA. **[5]** Modi, P. A. *et al. J. Hydro* (2022) **[6]** R A Pauloo *et al. Environ. Res. Lett.* **15** 044010 (2020) **[7]** Smith, R. G. and Majumdar, S. *WRR 57* (2020) **[8]** Clements, T and Denolle, M. A. *GRL 45, 6459-6465* (2018) **[9]** Mao, S. *et al. Nat. Comm. 13, 4643* (2022) **[10]** Brenguier *et al. Nature Geosci* **1**, 126–130 (2008)