Towards Modeling and Predicting Water Table Levels in California

Haselden, EJ., Salinas, Edward A. and Thomas, Kieffer

Tuesday, April 12, 2022



Water is Precious

California is getting a very dry start to spring, with snowpack far below average

April 1, 2022 - 3:57 PM ET



https://www.npr.org/2022/04/01/1090342138/california-drought-dry-spring



Problem Definition

- Groundwater as a Mass-Balance Problem
- Changes in Groundwater Storage:
 - o In Sierra Nevada Mountains

- At Lower Elevations
 - $\Delta S = P + Q_{in} ET Q_{out} U$

S - Groundwater Storage

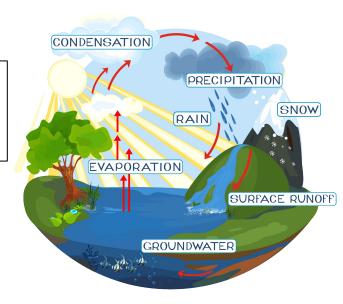
P - Precipitation

ET - Evapotranspiration

Q - Runoff

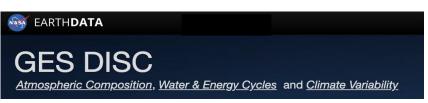
U - Usage (Agriculture etc.)

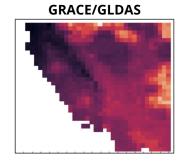
 Predict groundwater measurements in CA for the dry season (Apr-Sep) from historical trends and snowpack/precipitation in snow/rain season (Oct-Mar)

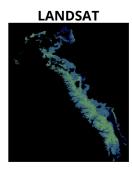




Data



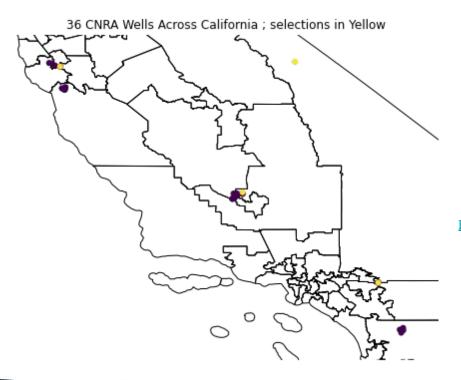




- Assimilates satellite imagery, GPS, and large volume of direct observational data to model hydrologic variables at granular scale
 - o Gravity Recovery and Climate Experiment Terrestrial Water Storage (GRACE-TWS)
 - <u>Catchment deficit (mm)</u>: ("water table deficit") average depth of water needed to saturate catchment; conceptually
 - Root-zone moisture content (rzmc): water contained in the root-zone soil layer (0-200cm)
 - <u>Surface excess (srfexc)</u>: amount of water in the surface soil layer (o-5cm) in excess of equilibrium
 - Global Land Data Assimilation System (GLDAS)
 - o Snow-Reanalysis Dataset from LANDSAT Satellite



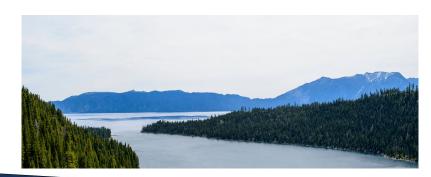
CNRA Data and Well Measurements Modeled





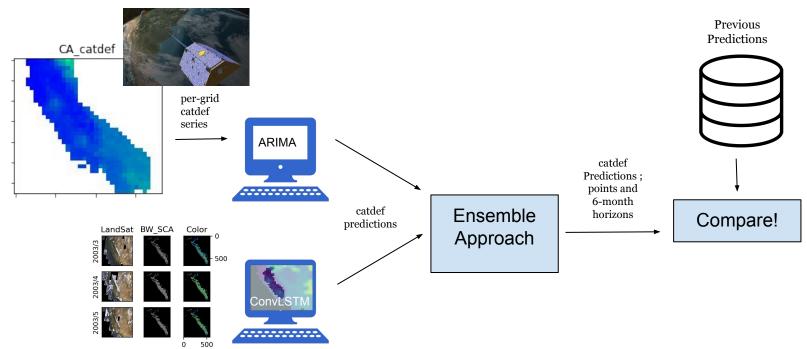
- Over 40,000 wells and their water table measurements
 - As old as 1888 or as new as 2022

https://data.cnra.ca.gov/dataset/periodic-groundwater-level-measurements





Modeling & Analysis Overview





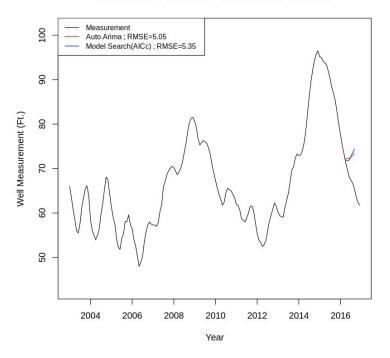
Modeling, Prediction, and Data Concerns

- Dr. Manuela Girotto advised that 6-month forecast outlooks of April September for a given year would be ideal
 - These months and the 6-month term based on water demands from agriculture as well as the summer season when extensive precipitation is not typically expected
- She also expressed concern that the GRACE data is not sufficiently immediate or itself too much the result of algorithms
 - o For such reasons she suggested analysis of the CNRA data

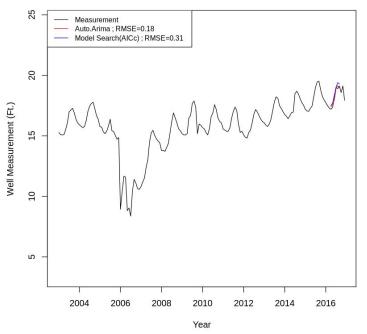


Analysis Results arima models of CNRA Data

March 2016 Forecasts: 373177N1219435W005



March 2016 Forecasts: 373922N1183430W001

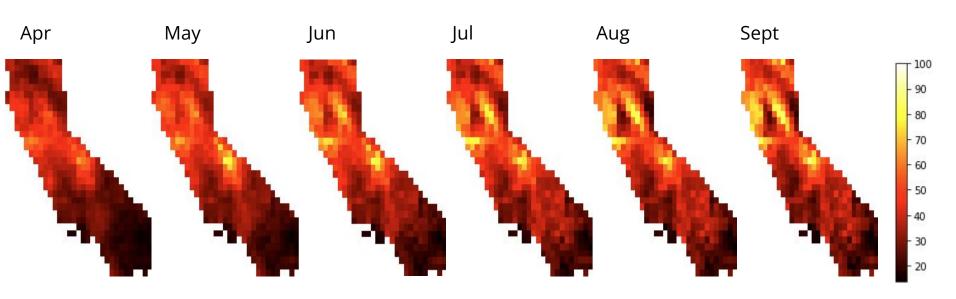


ARIMA models predict future values based on past values and relationships between past values.

ARIMA is an acronym for "<u>A</u>uto <u>Regressive</u> <u>Integrated</u> <u>M</u>oving <u>A</u>verage"



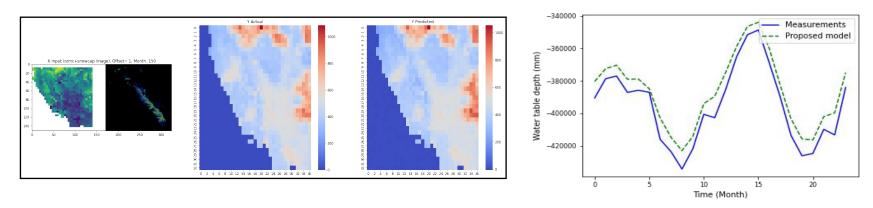
Analysis Results ARIMA Models of Grid-Location Water Tables



Average RMSE at each location. Lower is better on the scale; darker is better on the map.



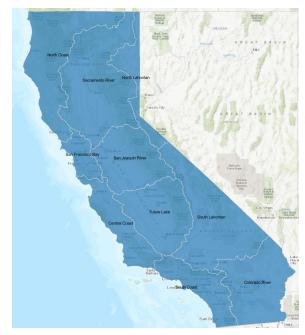
Inherited ConvLSTM Progress



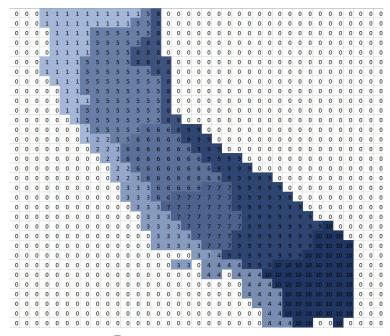
Rzmc + Snowpack ConvLSTM predicted catdef with average 88.3% accuracy



Regional ConvLSTM



Source: https://gis.data.ca.gov/datasets/2a572a181e094020bdaeb5203162de15_0/explore

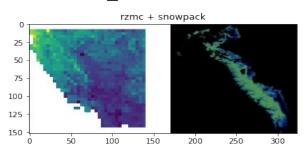


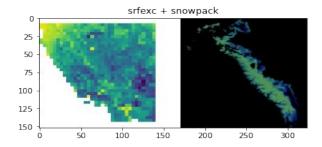




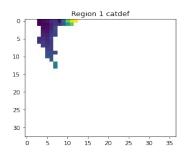
Regional ConvLSTM

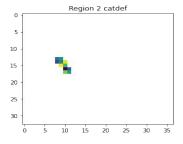
X_train





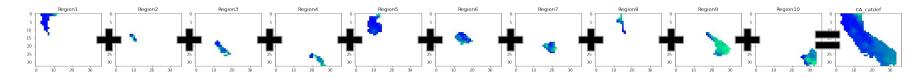
Y_train

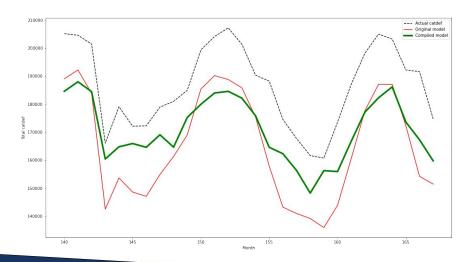






Regional ConvLSTM





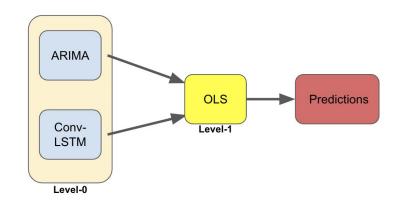
Reconstructed composite model predicted catdef with average 91.7% accuracy (increase of 3.4%)

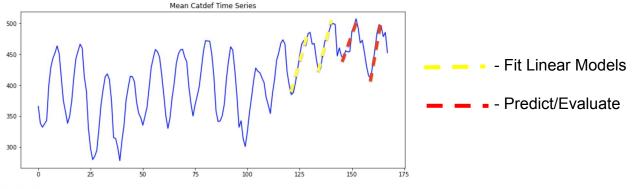


Ensemble

Linear Regression Models with the ARIMA and ConvLSTM Predictions as Input (by Month)

6 Models Fit on Apr-Sep 2013-2014 Data and Predictions Ensembled Predictions for Apr-Sep 2015-2016







Ensemble Results

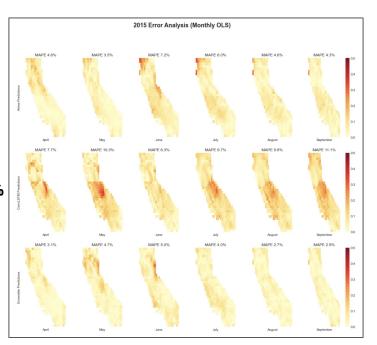
Comparison of Mean Absolute Percent Error for Arima/ConvLSTM/Ensemble 2015-2016

MAPE

Arima: 5.03%

ConvLSTM: 9.15%

Ensemble: 3.73%

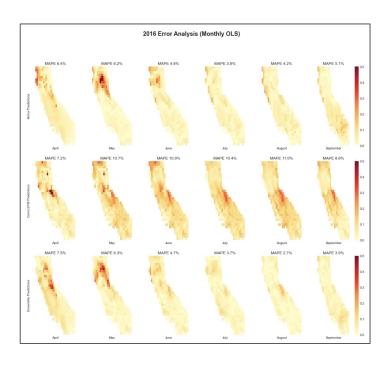


MAPE

5.08%

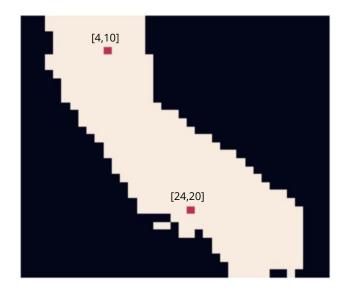
9.79%

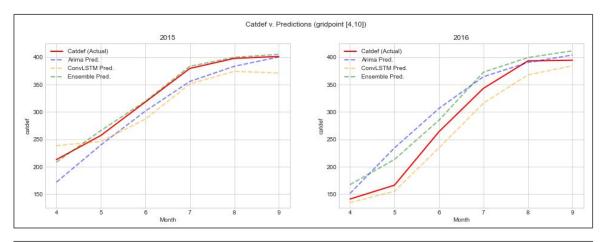
4.80%

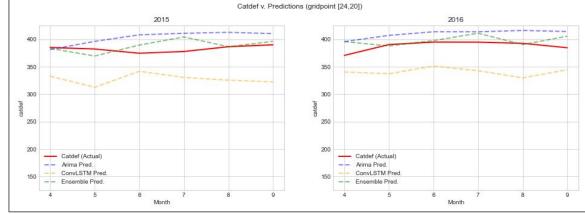




Ensemble Results

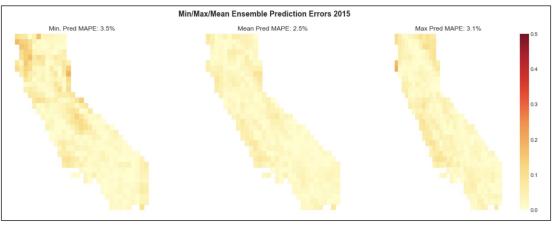


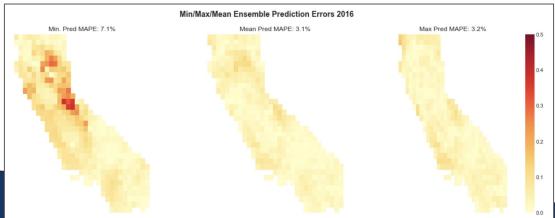






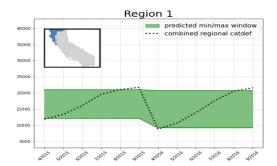
Ensemble Results

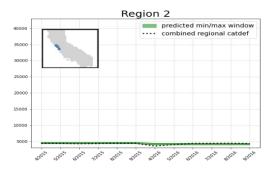


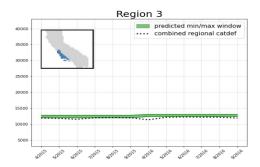


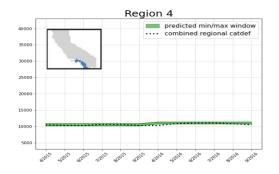


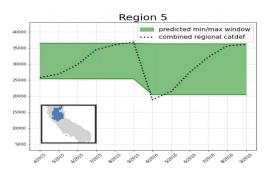
Prediction Envelopes

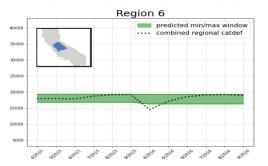














Conclusions

- Average Mean Absolute Errors of 3.73% and 4.80% in 2015 and 2016
 - Previous study (full-area ConvLSTM) had error of 11.7%
 - Reconstituted regional ConvLSTMs had error of 9.3%
- We Provide *min*, *mean*, and *max* Over the Seasons for a Given Year
 - These ranges likely more useful than point estimate for setting policy
 - Usage restrictions/conservation initiatives
 - Well-metering
 - Residential/commercial/industrial zoning
 - Agricultural allotment/crop rotation



Ideas for Next Steps

- Revisit stakeholder expectations/applications now that we can offer accuracy
 >90%
- Consider applying "explainability" procedures to applied deep learnings methods
 - e.g. "integrated gradients" to explain which snowpack area best predicts water tables at various locations
- Explore additional data sources for modeling and prediction



Thank You!

