A Data-Driven Approach to Deformation Forecasting: Machine Learning on InSAR Data

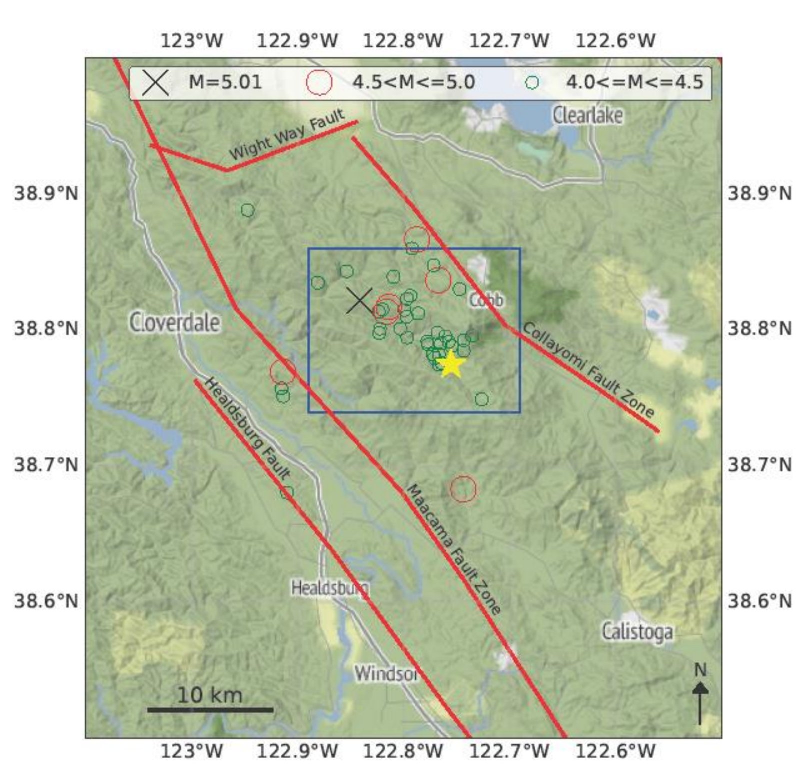
Joe Yazbeck[1] and John B. Rundle[1,2]

[1]University of California, Davis, [2]Santa Fe Institute

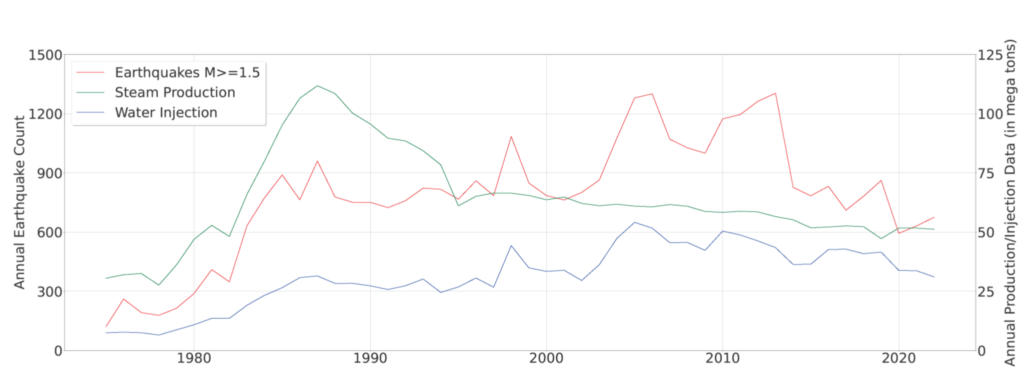
Abstract

Anthropogenic activities such as fluid injection, fluid extraction, mining, and hydraulic fracturing can all cause induced seismicity which can in turn result in land subsidence. This latter phenomenon is devastating to local infrastructure as well as underlying aquifers. It is for this reason that monitoring and predicting land deformation is of utmost importance. We relied on Interferometric Synthetic Aperture Radar (InSAR) images captured by Sentinel-1 to monitor deformation in the line-of-sight of the satellite. The Geysers geothermal field, where injection plays a direct role in induced seismicity, was used as the area of study and a deformation time series was built using LiCSBAS [1]. Two machine learning models (model **A** and model **B**) that included Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) layers were built to predict future deformation maps. The only difference between the models was the incorporation of geothermal injection and production data in model **B**. While both models outperformed a baseline linear model, it was model **B** that performed the best based on a mean squared error metric.

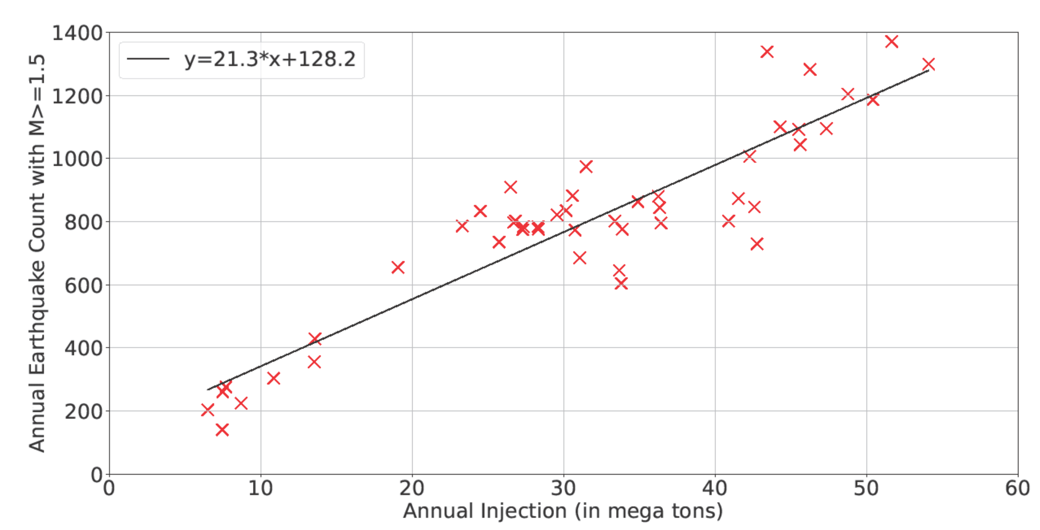
Case Study: The Geysers geothermal field



The Geysers geothermal field uses water injection to stimulate steam production which powers turbines that generate electricity. It is the largest geothermal power plant in the world producing around 800 MW of electricity which powers roughly 800,000 homes in the nearby counties. While a major earthquake (M>7.0) is unlikely to occur at the field since it does not directly lie on a major fault line, nearby fault zones do exist. Most notably, the Maacama fault zone is certainly capable of producing a major earthquake, and it is of concern to scientists and researchers what role the fluid injection at The Geysers could play in the triggering of such an earthquake [2].

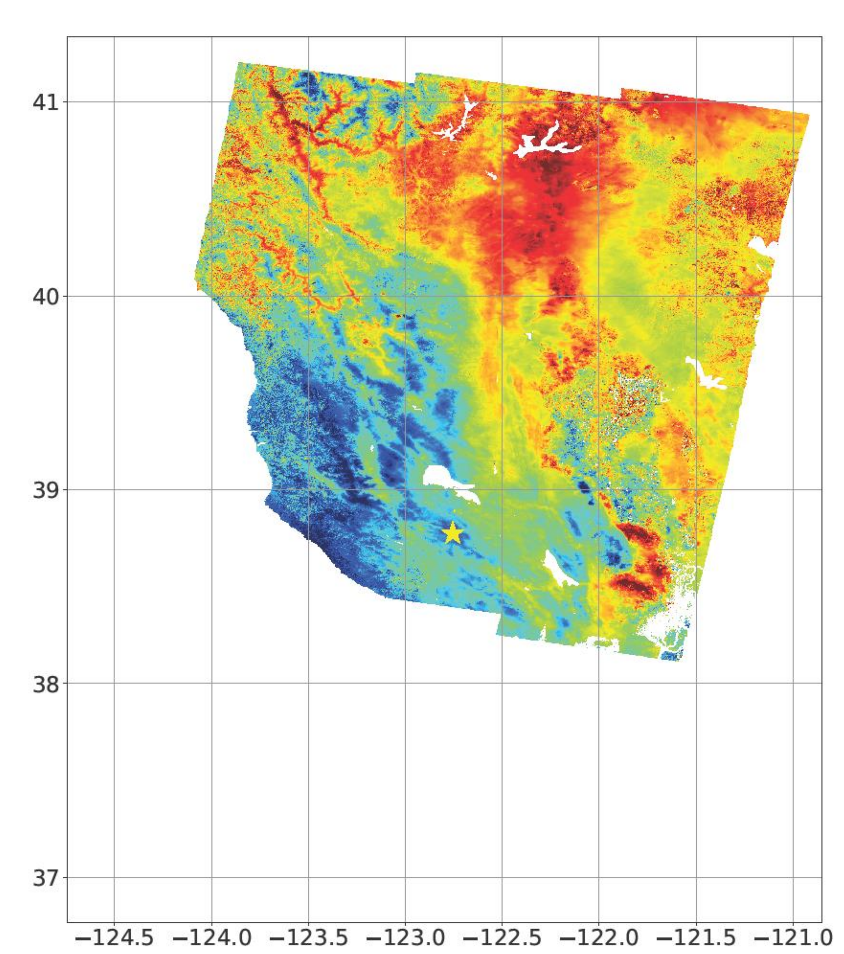


Seismicity at The Geysers



Using the Pearson and Spearman correlation coefficients, we find that the injection amount is highly correlated to the induced seismicity especially for earthquakes whose magnitudes range between 1.5 and 2.0. This indicates that the induced seismicity at The Geysers is relatively predictable given a certain injection amount. What this also means is that injection amounts can be used as a proxy for resulting deformation as the compounding earthquakes are the leading reason for the land subsidence occurring at The Geysers.

InSAR Study



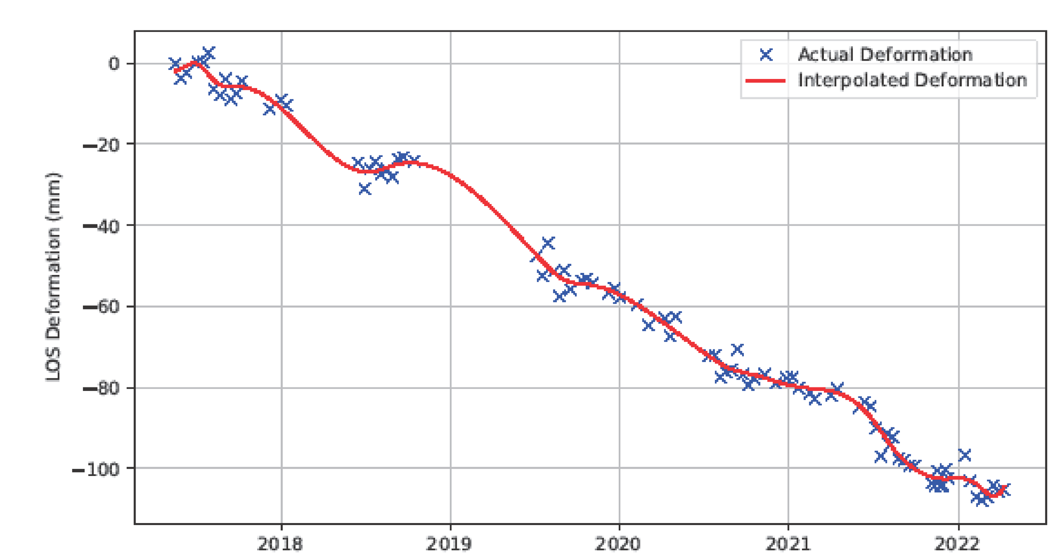
We build an InSAR velocity map from 2017 to 2022 using LiCSBAS. Cold colors indicate a velocity away from the satellite while warmer colors indicate a velocity towards the satellite. The yellow star indicates the location of The Geysers geothermal field, and the displayed velocities have a unit of mm/year. We can see that the field is experiencing subsidence at an approximate rate of 20 mm/year.

Data Preprocessing

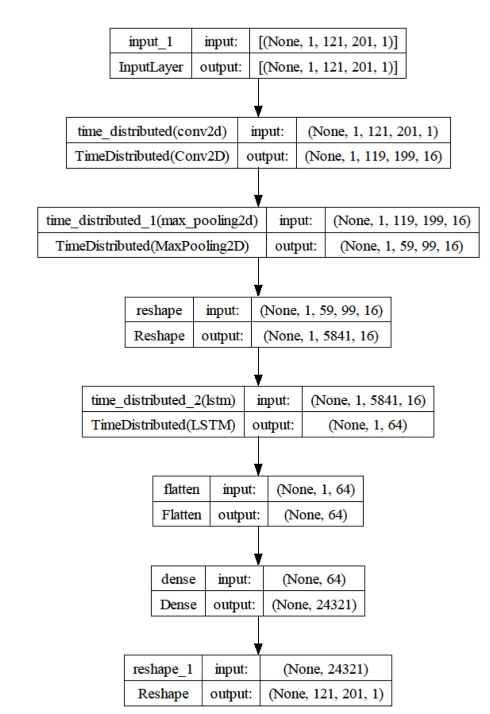
The resulting InSAR time series had 2 main issues:

1. It was not equally spaced
2. It had a couple large gaps

In order to fix these issues, we use a smoothing spline interpolation to which we enforce a limit on the amount of times it can oscillate up and down to avoid ringing issues.



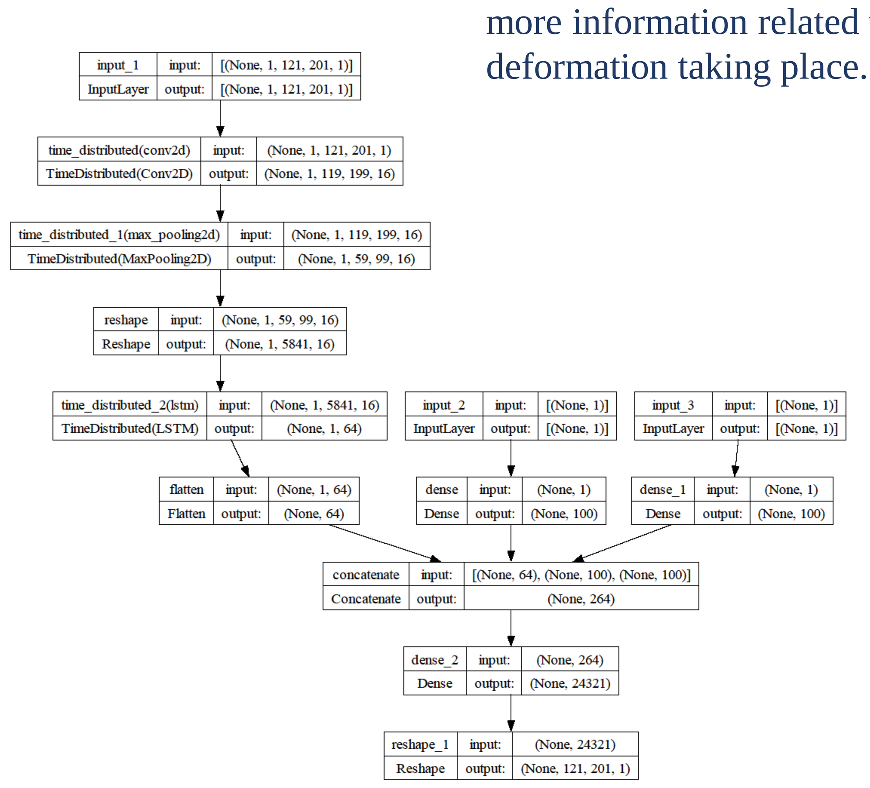
Model Architectures



Model **A**

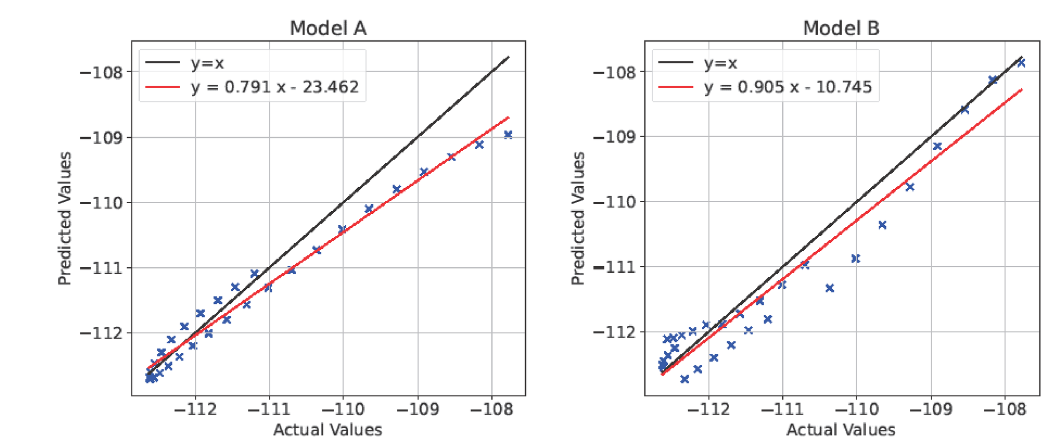
We build 2 different models: **A** and **B**. The models share similar elements such as LSTM and CNN layers. These layers are crucial to the models' learning as the LSTM layer learns the temporal patterns in the data, while the CNN layer learns the spatial patterns [3,4].

The key difference between the models is the incorporation of geothermal injection and production amounts as additional inputs in model **B**. This should give the model more information related to the deformation taking place.



Model **B**

Results and Conclusion



After running the models several times, we find that model **B** outperforms model **A** by a margin of 0.4 mean squared error. In order to highlight and visualize the difference in performance of the models, we plot the actual vs predicted plot of the averaged deformation images for a single run of the models. One can see that, in the case of model **B**, the fitted line is closer to the perfect y=x line.

This shows that incorporating additional data such as geothermal injection and production data into machine learning models that predict deformation can be beneficial and shows promise for use in hazard mitigation models that specifically deal with deformation as a result of induced seismicity.

Paper QR Code

References

[1] Morishita, Y., Lazecky, M., Wright, T.J., Weiss, J.R., Elliot, J.R. and Hooper, A., 2020. LiCSBAS: An open-source InSAR time series analysis package integrated with the LiCSAR automated Sentinel-1 InSAR processor. *Remote Sensing*, *12*(3), p.424.

[2] Luginbuhl, M., Rundle, J.B., Hawkins, A. and Turcotte, D.L., 2018. Nowcasting earthquakes: a comparison of induced earthquakes in Oklahoma and at the Geysers, California. *Pure and Applied Geophysics*, *175*, pp. 49-65.

[3] Yazbeck J. and Rundle, J.B., 2023. Predicting Short-Term Deformation in the Central Valley Using Machine Learning. *Remote Sensing*, *15*(2), p. 449.

[4] Cai, J., Gu, S. and Zhang, L., 2018. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Transactions on Image Processing*, 27(4), pp. 2049-2062.

Contact Information

Email: jyazbeck@ucdavis.edu