

Mapping of aquifer geological composition from space

Kyongsik Yun

kyun@jpl.nasa.gov

California Institute of Technology

Kyra H. Adams

California Institute of Technology

J. T. Reager

California Institute of Technology

Zhen Liu

California Institute of Technology

Michael Turmon

California Institute of Technology

Article

Keywords:

Posted Date: October 5th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11029247/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Land subsidence due to groundwater recharge and withdrawal has emerged as a widespread, societally impactful consequence of human development. However, the relationship between groundwater depletion and land subsidence is incredibly complex. It depends on a suite of environmental variables and non-linear processes at and beneath the land surface. One major constraint on land surface subsidence and rebound is the complex geologic matrix and depth-varying responses of distinct geologic features to groundwater drainage and/or filling. In many regions globally, sufficient soil geology information at depth is unavailable. In this work, for a study section of California's Central Valley aquifer, we show that time-series characteristics of surface deformation detected using interferometric SAR (Synthetic Aperture Radar) observations, when combined with deep neural network learning approaches, can successfully provide screening-level insight into integrated soil-column responses to predict a map of basic geologic composition and coarse grain percentages of an aquifer. As aquifer grain sizes dictate aquifer elasticity, elucidating grain-size composition based on surface signals is valuable for predicting future aquifer behavior and groundwater depletion for many of the world's critical aquifers.

Key Points

- In an experiment over the Central Valley, surface deformation signals from InSAR correlate spatially with soil geology
- Machine learning techniques applied to time-series signals from InSAR showed skill in predicting vertically-integrated soil geologic composition
- This approach may have value for subsiding aquifers globally that lack in-situ geological information

Main text

The relationship between groundwater depletion and surface subsidence is complex . Groundwater extraction increases effective stress within pore spaces in sedimentary systems, compacting the geologic unit and causing land subsidence on the surface . Subsidence then poses a two-fold threat to the aquifer system at hand: it decreases allowable groundwater storage space, and it damages civil infrastructure such as buildings, pipes, and roads . Many of the world's major metropolitan areas are witnessing groundwater-related surface subsidence, and it is estimated that groundwater extraction is involved in nearly 60% of the world's cases of human-induced subsidence . Recent satellite and in situ studies continue to highlight the intensifying rate of subsidence in aquifer systems under prolonged drought and increasing water demand, including in California's Central Valley .

However, subsidence trends from the surface are difficult to decipher, as the trends represent a vertically integrated signal from varying compaction behaviors across different geologic layers at depth . While sandy bodies can elastically "bounce back" after compaction due to ease of recharge, more fine-grained bodies often undergo irreversible, inelastic compaction . Fine-grained layers can also be slow to

equilibrate, driving long-term signals (e.g., years to decadal), while coarser layers cause higher frequency signals from their shorter response times (e.g., monthly to seasonal) . As such, geologic heterogeneity renders the relationship between groundwater depletion and land subsidence highly non-linear and difficult to track and predict . Probabilistic and hybrid machine learning models such as Gaussian processes and deep networks have recently been shown to improve predictive capabilities for complex aquifer behavior in Central Valley .

The recent advent of Synthetic Aperture Radar (SAR) systems has provided a great advantage in tracking and understanding surface subsidence . In particular, Interferometric SAR (InSAR) has shown great success in quantifying the subsidence in regions such as California, Beijing, and Mexico City by analyzing persistent point targets across interferometric phase pairs . However, interlinking subsidence behavior to contrasting groundwater signals or using surface trends to make predictions on groundwater requires a deeper contextual understanding of the geologic structure beneath . Recent syntheses of remote sensing capabilities highlight the importance of integrating multiple satellite-based measurements to improve the interpretation of subsurface aquifer responses . Within the Central Valley, contrasting signals from groundwater storage and land surface are often exhibited, especially on shorter timescales (monthly), and require geologic explanations to contextualize this behavior . However, geologic data are often labor-intensive and costly, requiring dense borehole networks or detailed modeling . The challenge of sparse observational networks further underscores the need for proxy-based or remote approaches to characterize geologic variability .

In this work, we leverage the wealth of existing geologic information compiled for the Central Valley aquifer system in California to exhibit the feasibility of using surface subsidence signals to inversely predict the geologic composition at depth. Using InSAR, regions were classified by their temporal surface movement dynamics and trained against a kriged geologic dataset using a neural network. We find a significant correlation between InSAR-based land deformation and geologic composition using the deep neural network regression model, even in cases with sparse sampling, suggesting that the model has the potential to be generalized to other regions for indirect estimation of geologic composition .

Predicting Geologic Composition from InSAR-Derived Surface Deformation Using Deep Neural Networks

Deep neural networks successfully displayed the predictive power for geologic composition from InSAR remote sensing data, with a mean correlation coefficient of $R = 0.88$ across all aquifer layers (Fig. 2). The decision tree method ($R = 0.65$) and random forest model ($R = 0.85$) were tested as a baseline. There were some variances in model performance across the ten model layers, with model layers 7–10 achieving the most success (Fig. 2C). This may be attributed to the fact that deeper confined aquifers (Corcoran clay and the units underneath) are often more fine-grained and, therefore, responsible for more secular, long-term subsidence signals that are more prominent and larger in amplitude than short-term elastic signals.

The model demonstrated robust performance even when trained on a reduced dataset, achieving a correlation coefficient of $R = 0.80$ with only 40% of the training data. While the overall accuracy decreased compared to the full dataset, this strong performance suggests that the model can generalize well to other regions for indirect estimation of geologic composition.

To further test the model's resilience and ability to handle reduced spatial correlation between data points, we implemented a *distant data sampling* strategy. In this approach, training data points were selectively sampled to ensure a minimum separation distance of 10 kilometers between any two points (Fig. 3). This method was designed to reduce the spatial autocorrelation in the training data, effectively challenging the model to rely more heavily on learned temporal and geologic patterns rather than spatial proximity.

During this run, increased errors were observed in predictions for Layer 4. This discrepancy may be attributed to the presence of fine-grained lenses within this layer, which exhibit high contrast in grain size compared to surrounding materials (Supplementary Fig. S2). Despite this localized drop in accuracy, the model maintained a strong overall correlation with ground-truth data across all layers, with an $R = 0.83$. This result underscores the model's robustness and ability to provide meaningful predictions even under less-than-ideal spatial sampling conditions.

The neural network was able to adequately predict the percentages of the coarse- or fine-grained binary bins by model cell, even going down into the deeper model layers. While the spatial dynamics were well-replicated, further analysis was conducted to see neural network performance on temporal characteristics. As groundwater and precipitation patterns display fluctuations over various timescales (monthly, seasonal, and interannual), further analysis was conducted to determine which month of data was the most critical to neural network performance. We conducted a 10-fold cross-validation performance test, systematically leaving out data from one selected month per iteration (Fig. 4). Results revealed that excluding data from October and December led to the most significant decline in model correlation. This suggests that InSAR data from these two months play a particularly important role in capturing temporal patterns essential for accurately estimating geologic composition. This may be related to the pulse of winter recharge the Central Valley receives, with most of the precipitation starting around late fall and lasting in spurts throughout the winter months. This is expected to drive significant, telling signals in the surface movement, with elastic regions uplifting, responding to winter recharge¹³. Therefore, without such elastic winter signals included as part of the training dataset, the model performance degrades.

Geologic Interpretability and Broader Applicability of InSAR-Driven Neural Network Models for Aquifer Characterization

Aquifers have high heterogeneity both laterally and vertically, which can cause spatial variability in model performance (Figs. 2 & 3). For the Central Valley, the thick, fine-grained Corcoran Clay deposit supported better prediction of geologic composition at corresponding depths, most likely owing to the clay driving

long-term, secular subsidence signals showing up across several InSAR acquisitions with amplitudes greater than seasonal surface movements. The approaches presented here, mainly targeted to test the rigor of the method, leveraged the wealth of CVHM kriged geotextural data available to this region and tested prediction capabilities for all 10 model layers. However, for applications to other aquifer regions without much borehole data or detailed geologic information, it may be a more conservative approach to train and run neural networks to predict bulk coarse grain percentage integrated across the entire aquifer depth, rather than producing predictions by depth.

The model was able to closely replicate the kriged dataset from CVHM. This highlights the capability of using InSAR surface deformation signals as input for the neural network to generate reliable estimations of relative geologic composition, particularly when simplified into coarse- and fine-grained bins.

The grain size difference between these two bins in the Central Valley spans a wide range (2–2,000 μm), aligning with typical sedimentary environments experiencing subsidence. While exact grain size classifications for these bins cannot be determined without direct sediment sampling and ground truthing, the model's outputs still provide valuable screening-level geologic insights. These preliminary results can serve as a useful starting point for identifying aquifer regions susceptible to compaction or for highlighting connected features of interest that warrant further investigation.

Recent work highlighted the severity of groundwater pumping-related subsidence across global aquifers²² using InSAR signals and machine learning methods. While the study was able to quantify the extent and rate of subsiding regions, management of groundwater pumping or Managed Aquifer Recharge (MAR) requires details on geologic heterogeneity, feature connectivity, and overall composition¹⁹, to better predict aquifer response to management strategies. It is likely that results from the LSTM network will require coupling with process-based groundwater numerical models for specific quantification of aquifer responses, but nonetheless, the approach and results presented here perform well and are robust enough to be applied to other regions. Many regions around the world that are under threat of subsidence are also mostly sedimentary in nature¹¹, casting a hopeful outlook on the scalability of InSAR-based LSTM neural networks in geologic composition predictions. Further, this work shows an optimistic application pathway for upcoming SAR missions such as NISAR, a collaborative SAR mission between NASA and the Indian Space Research Organization (ISRO). NISAR is expected to produce surface deformation measurements at resolutions of 5–10 meters, four to six times a month. Such data can be integrated into the methods presented here to yield high-resolution geologic composition estimates for subsidence aquifers worldwide.

We showed that InSAR land deformation signals can be used to remotely estimate geologic composition. While understanding geologic heterogeneity is critical to contextualizing the non-linear relationship between surface deformation signals and groundwater depletion, obtaining such in-situ data can be expensive and time-consuming. The approaches highlighted here, while very general and column-integrated, can serve as a first estimate of geologic structure based on surface deformation behavior across wet and dry periods. This may be helpful in some regions to address sub-surface sampling

limitations and showcase future application scenarios for future InSAR missions such as NISAR, where geologic composition can be quickly estimated at a high resolution using remote sensing. This work has implications for better understanding critical groundwater-bearing aquifer systems around the world, as well as applications for groundwater management and monitoring.

Declarations

- **Funding Declaration**

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004). Funding was provided by Jet Propulsion Laboratory's Research and Technology Development (R&TD) program.

Data Availability Statement

The Sentinel-1 InSAR-derived surface deformation data are available through the NASA Alaska Satellite Facility (<https://asf.alaska.edu>). The Central Valley Hydrologic Model (CVHM) geotexture data are publicly available from the U.S. Geological Survey (<https://pubs.usgs.gov/pp/1766/>). Groundwater well data are accessible via the USGS National Water Information System (<https://waterdata.usgs.gov>). Precipitation data are available from PRISM Climate Group (<https://prism.oregonstate.edu>). The neural network model code and processed datasets generated in this study are available from the corresponding author upon reasonable request.

Author Contribution

K.Y. conceived the study, developed the neural network model, led the analysis, and wrote the manuscript. K.H.A. contributed to scientific interpretation and manuscript revision. J.T.R. contributed to hydrological interpretation. Z.L. processed the InSAR time-series data. M.T. contributed to statistical validation. All authors reviewed and approved the final manuscript.

References

1. Liu, Z. et al. Monitoring Groundwater Change in California's Central Valley Using Sentinel-1 and GRACE Observations. *Geosciences* **9**, 436 (2019).
2. Farr, T. G., Jones, C. E. & Liu, Z. Progress Report: Subsidence in California, March 2015 – September 2016. 1–37 (2016).
3. Meinzer, O. E. Compressibility and elasticity of artesian aquifers. *Econ. Geol.* **23**, 263–291 (1928).
4. Terzaghi, K. K. & Terzaghi Principles of Soil Mechanics I—Phenomena of Cohesion of Clays, Engineering News-Record, Vol. 95, No. 19, pp. 742–746. - References - Scientific Research Publishing. (1925). <https://www.scirp.org/reference/referencespapers?referenceid=790260> (1925).

5. Faunt, C. C., Sneed, M., Traum, J. & Brandt, J. T. Water availability and land subsidence in the Central Valley, California, USA. *Hydrogeol. J.* **24**, 675–684 (2016).
6. Faunt, C. C. & Sneed, M. Water availability and subsidence in California's Central Valley. *San Francisco Estuary Watershed Sci.* **13**, 0–8 (2015).
7. Bagheri-Gavkosh, M. et al. Land subsidence: A global challenge. *Sci. Total Environ.* **778**, 146193 (2021).
8. Liu, P. W. et al. Groundwater depletion in California's Central Valley accelerates during megadrought. *Nat. Commun.* **13**, 7825 (2022).
9. Poland, J., Lofgren, B. E., Ireland, R. L. & Pugh, R. G. Land Subsidence in the Prepared in cooperation with the California Department of Water Resources. *Geological Surv. Prof. Paper* **437–H**, (1972).
10. Galloway, D. L., Jones, D. R. & Ingebritsen, S. E. *Land Subsidence in the United States* Vol. 1182 (US Geological Survey, 1999).
11. Mirzadeh, S. M. J. et al. Characterization of Irreversible Land Subsidence in the Yazd-Ardakan Plain, Iran From 2003 to 2020 InSAR Time Series. *Journal of Geophysical Research: Solid Earth* **126**, e2021JB022258 (2021).
12. Liu, Z. et al. Monitoring groundwater change in california's central valley using sentinel-1 and grace observations. *Geosciences (Switzerland)* **9**, (2019).
13. Vasco, D. W. et al. Using Sentinel-1 and GRACE satellite data to monitor the hydrological variations within the Tulare Basin, California. *Scientific Reports* **12**, (2022).
14. Pradhan, A. et al. Modeling Groundwater Levels in California's Central Valley by Hierarchical Gaussian Process and Neural Network Regression. *Journal of Geophysical Research: Machine Learning and Computation* **1**, e2024JH000322 (2024).
15. Farr, T. G. & Liu, Z. Monitoring subsidence associated with groundwater dynamics in the Central Valley of California using interferometric radar. *Remote Sens. Terr. Water Cycle.* **206**, 397 (2014).
16. Herrera-García, G. et al. Mapping the global threat of land subsidence. *Science* **371**, 34–36 (2021).
17. Adams, K. H. et al. Remote Sensing of Groundwater: Current Capabilities and Future Directions. *Water Resources Research* **58**, e2022WR032219 (2022).
18. Kim, K. H. et al. An Evaluation of Remotely Sensed and In Situ Data Sufficiency for SGMA-Scale Groundwater Studies in the Central Valley, California. *J. Am. Water Resour. Assoc.* **57**, 664–674 (2021).
19. Faunt, C. C. & Ed *Groundwater Availability of the Central Valley Aquifer, California: U.S. Geological Survey Professional Paper 1766. Program* (2009).
20. Hastings, D. *California's Central Valley: Producing America's Fruits and Vegetables.* (2014).
21. Murray, K. D. & Lohman, R. B. Short-lived pause in Central California subsidence after heavy winter precipitation of 2017. *Science Advances* **4**, (2018).
22. Hasan, M. F., Smith, R., Vajedian, S., Pommerenke, R. & Majumdar, S. Global land subsidence mapping reveals widespread loss of aquifer storage capacity. *Nat. Commun.* **14**, 6180 (2023).

23. Page, R. W. *Geology of the Fresh Ground-Water Basin of the Central Valley, California: With Texture Maps and Sections* (US Government Printing Office, 1986).
24. Williamson, A. K., Prudic, D. E. & Swain, L. A. *Ground-Water Flow in the Central Valley, California* (US Government Printing Office, 1989).
25. Phillips, S. P. & Belitz, K. Calibration of a Texture-Based Model of a Ground-Water Flow System, Western San Joaquin Valley, California'. (1991).
26. Burow, K. R. *Hydrogeologic Characterization of the Modesto Area, San Joaquin Valley, California* (US Department of the Interior, US Geological Survey, 2004).

Figures

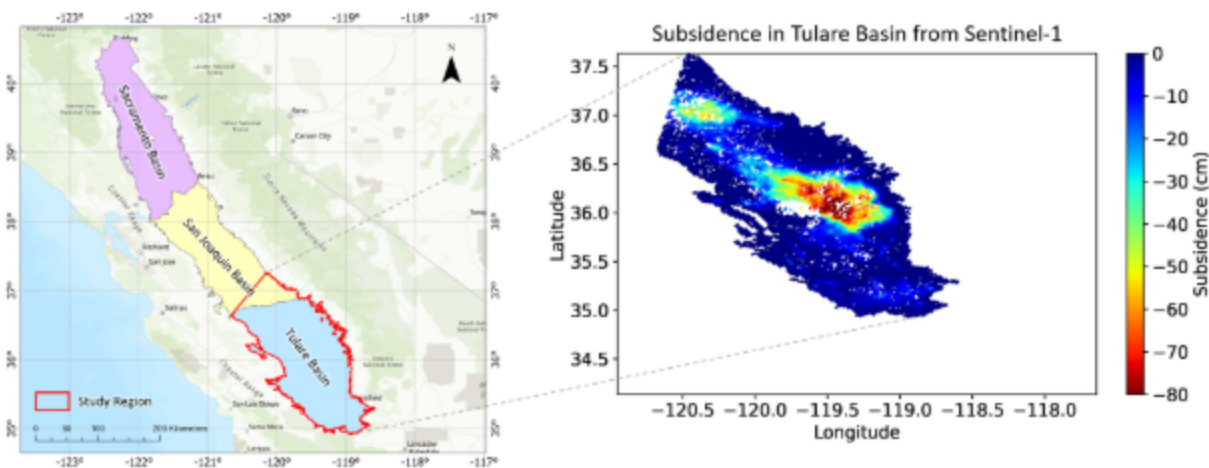


Figure 1

Map of the study site and study region. Central Valley has three major water basins: Sacramento, San Joaquin, and Tulare from North to South. The study region encompassed the Tulare Basin and southern San Joaquin, where subsidence hotspots have been identified using Sentinel-1 (right panel). Subsidence up to 80 cm is recorded from November 8, 2014 to January 22, 2019.

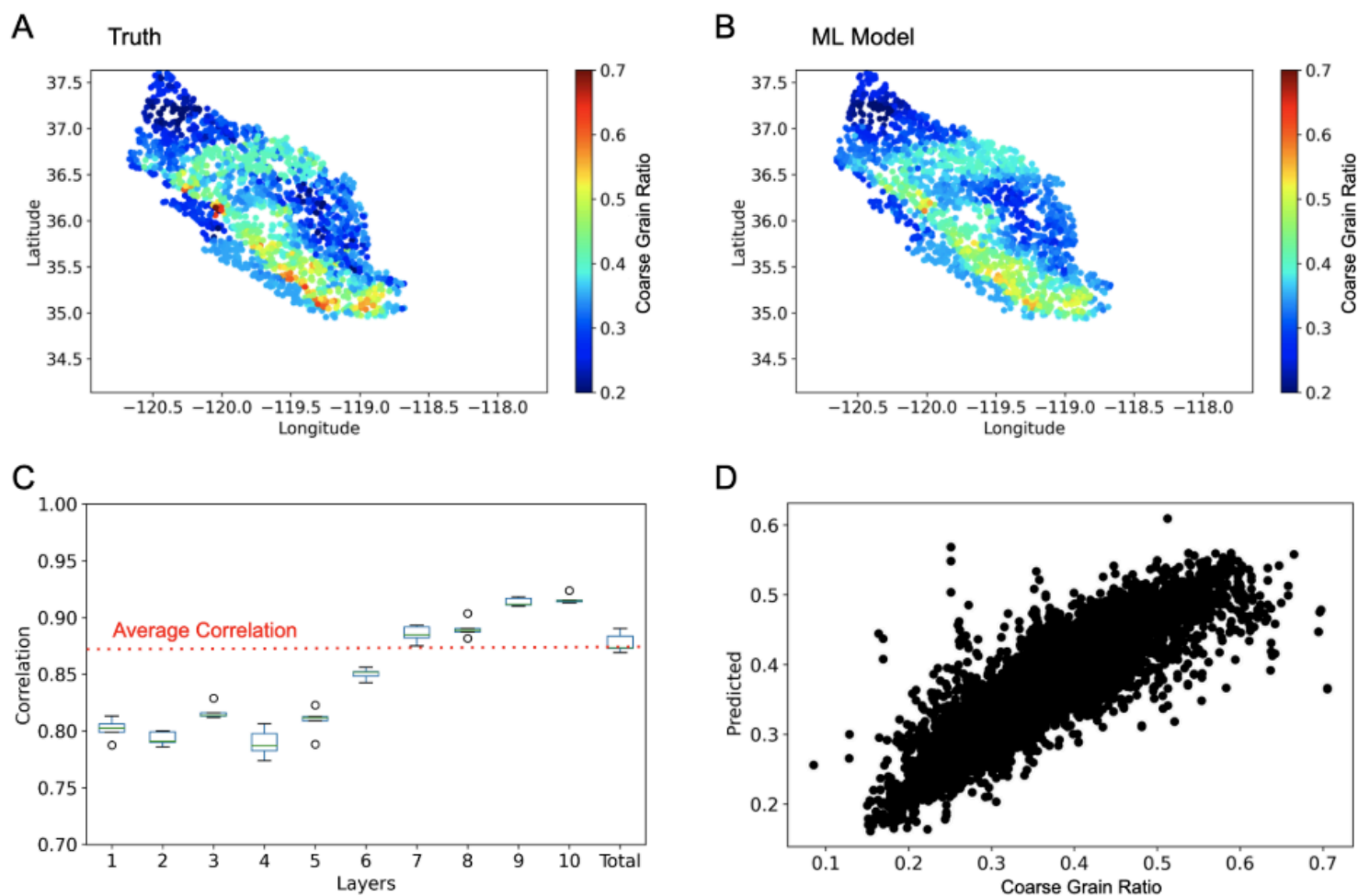


Figure 2

Geologic composition prediction using InSAR land deformation data. A: Ground-truth coarse grain percentage across all 10 layers of the geotexture model and B. estimated coarse grain percentage from the model. C. Correlation between model output and ground truth data across the different layers. D. Correlative scatter plot between the ground truth data and estimated geologic composition for the entire vertical section (no layer distinction).

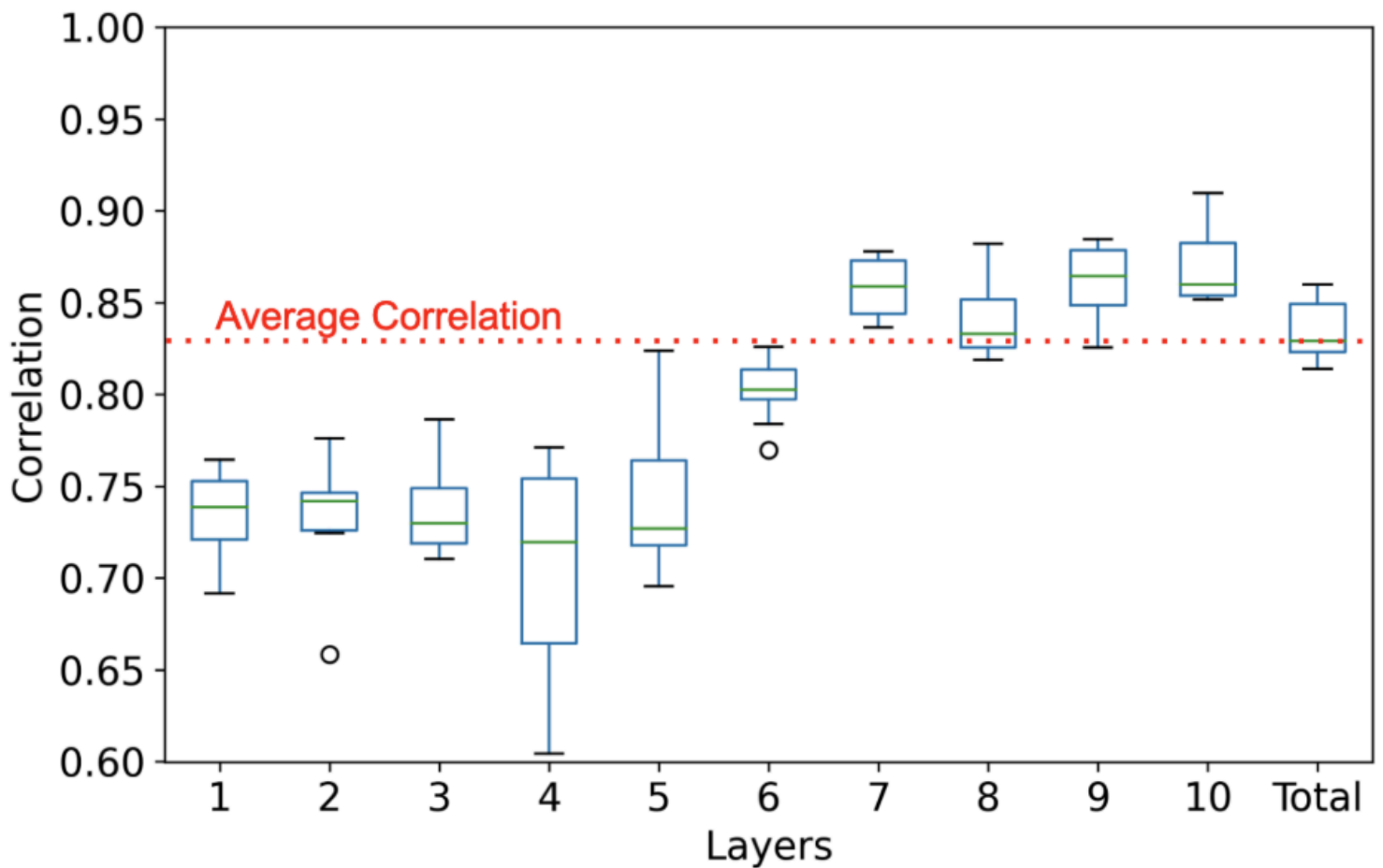


Figure 3

Geologic composition prediction with distant data sampling (minimum distance between samples was 10km) using InSAR land deformation data. Distant data sampling was performed to reduce the impact of spatial correlation of adjacent data points. Total prediction performance dropped from 0.88 to 0.83 but remained largely unchanged.

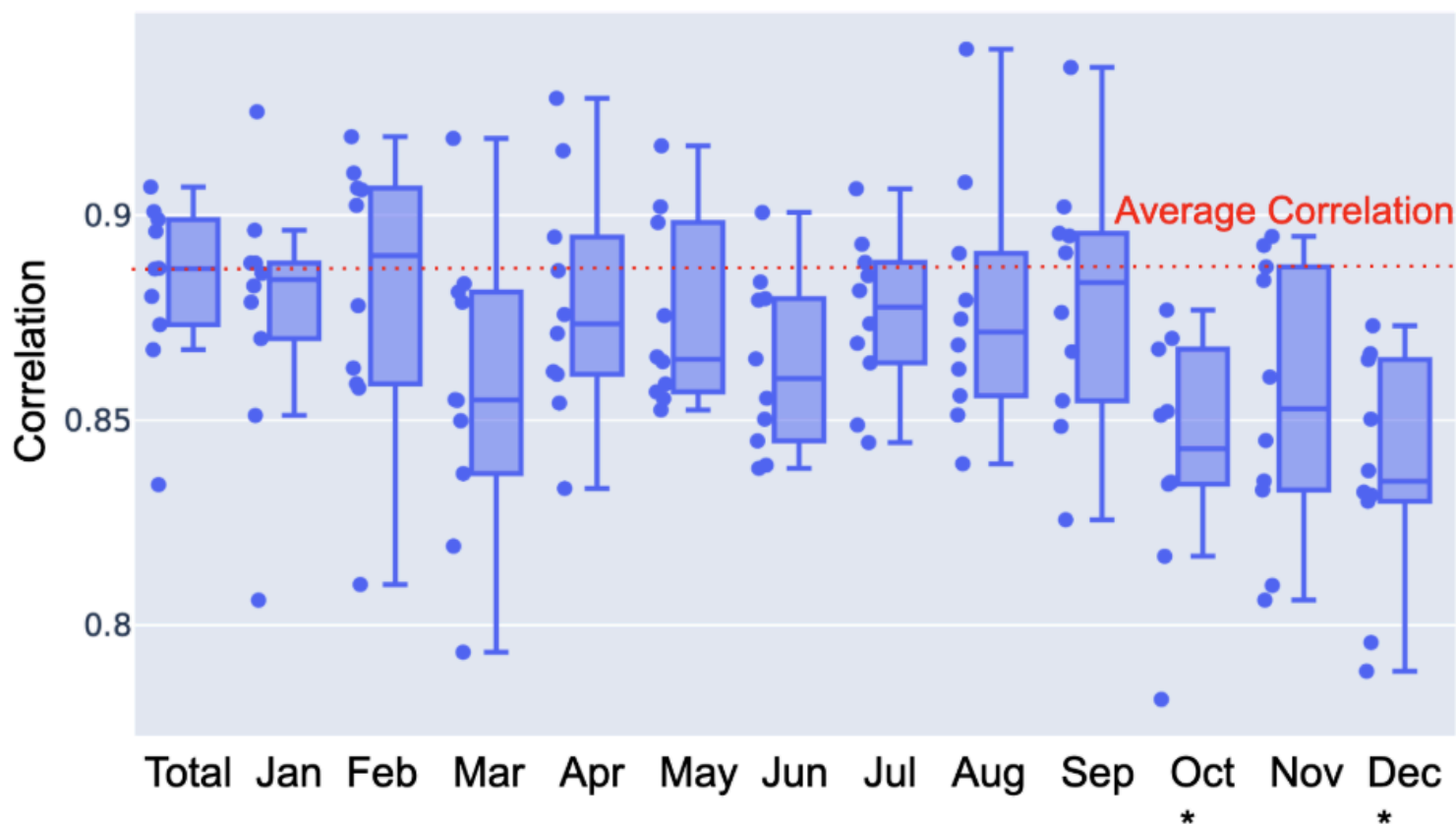


Figure 4

Performance results across 10-layers for the 10-fold cross-validation test where one month of data was intentionally excluded. Correlation degradation coefficients were computed and labeled according to the month of data that was excluded from the input. October ($p=0.00175$) and December ($p=0.000654$) showed significant degradation compared to the total data correlation. The statistical significance cut-off is at $p=0.004$ considering Bonferroni multiple comparisons ($0.05/12$ comparisons= 0.004).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [floatimage5.png](#)
- [floatimage6.png](#)
- [SupplementaryInformation.docx](#)