

Mapping Groundwater Storage Anomalies at 1-km Resolution Across Arizona Within the Lower Colorado River Basin Using Downscaled GRACE Data

Md Rayhan^{1,2}, Karem Abdelmohsen¹, H.A. Chandanpurkar¹, Kaushik Gopalan³, James S. Famiglietti^{1,2}

¹School of Sustainability, Arizona State University, USA

²School of Civil, Environmental, and Sustainable Engineering, Arizona State University, USA

³Center for Interdisciplinary AI, FLAME University, India

Abstract

Arizona is at the center of one of the most serious groundwater crises in the United States, where decades of over-pumping and dwindling Colorado River flows have left aquifers under increasing pressure with inadequate information to manage them successfully. GRACE and GRACE-FO satellite missions provide monthly estimates of changes in terrestrial water storage, but the coarse spatial resolution of the satellite missions of about 25 km blurs the sub-basin variability that water managers must observe. This research fills this gap by creating a 1-km groundwater storage anomaly (GWSA) dataset of Arizona between April 2002 and December 2023. Eight hydroclimatic predictors, the combination of which [precipitation, evapotranspiration, runoff, soil moisture and NDVI] best represent the time-integrated nature of GRACE observations, were used to train a Random Forest model. This method resulted in model correlation of $r = 0.84$ in the final setup, with 92.4% of the original GRACE signal variance being retained. The resulting 1-km GWSA maps reveal localized depletion rates reaching -10.4 mm/year in specific sub-basins across Arizona's Active Management Areas and unmanaged groundwater basins, contrasting sharply with the uniform -3 to -7 mm/year picture presented by standard GRACE data. Comparison with the 811 active index wells of the Arizona Department of Water Resources confirms that the downscaled product is properly capturing the sign, magnitude and spatial pattern of the groundwater change across the domain. This framework offers a practical and scalable groundwater monitoring tool as the declining river supplies and rising pumping demand make high-resolution groundwater data more important than ever.

1 Introduction

Groundwater is now the backbone of water security across the American Southwest (Famiglietti, 2014). As Colorado River supplies continue to shrink, this dependence is deepening (Abdelmohsen

et al., 2025). The 1922 Colorado River Compact was drafted during an unusually wet period and it locked in allocations that the river has rarely been able to meet (Tadych et al., 2024; USBR, 2026). Streamflow has since fallen by roughly 20% over the past century, leaving less surface water for the communities and farms that depend on it (Abdelmohsen et al., 2025; Tadych et al., 2024). The Central Arizona Project was built to ease this pressure by routing Colorado River water into Arizona but declining river flows have eroded its reliability (Central Arizona Project, 2025; Ventura, S. J., 2023). The gap left behind is being filled by groundwater. Arizona is now pumping at rates that cannot be sustained (Milman et al., 2021; Scanlon et al., 2025).

These impacts are already evident. From 2002 to 2023, the Lower Colorado River Basin has lost $36.0 \pm 6.2 \text{ km}^3$ of total water storage, 71% attributable to groundwater depletion (Abdelmohsen et al., 2025; Scanlon et al., 2025). During the 2002–2014 drought, groundwater declined at $-5.6 \pm 0.4 \text{ km}^3$ per year (Castle et al., 2014). At a smaller scale, groundwater levels in agricultural regions such as Willcox and Douglas have dropped by more than 10–11 m (ADWR, 2024; Tadych et al., 2024). Arizona is also at risk as a junior water user of the Colorado River (O'Neill et al., 2016). However, the primary challenge is not just water scarcity but, more importantly, the lack of high-resolution data to manage groundwater across diverse regions.

Satellites such as GRACE and GRACE-FO have transformed large-scale groundwater monitoring by estimating monthly terrestrial water storage anomalies (TWSA) from Earth's gravity field (Landerer & Swenson, 2012; Tapley et al., 2004; Wahr et al., 2004). These have been extensively used to monitor droughts, floods and storage trends (Famiglietti et al., 2011; Houborg et al., 2012; Reager et al., 2014; Rodell et al., 2009). But their coarse spatial resolution ($\sim 300 \text{ km}$) renders them unsuitable for local water resource management (Famiglietti et al., 2011; Scanlon et al., 2016; Vishwakarma et al., 2018). Local-scale irrigation, groundwater recharge from CAP water and sub-basin variability may be obscured (Long et al., 2015; Scanlon et al., 2012; Zaitchik et al., 2008). Downscaling methods combine GRACE with higher-resolution hydroclimatic variables (Arshad et al., 2025; Kalu et al., 2024; Miro & Famiglietti, 2018; Sahour et al., 2020). Machine learning methods, such as Random Forest, XGBoost and neural networks have achieved resolutions of 1 km in some areas (Arshad et al., 2024, 2025; Fatolazadeh et al., 2022; Hamou-Ali et al., 2025; Zhong et al., 2021). However, most studies employ instantaneous variables which do not capture the lagged and integrated nature of groundwater systems (Xiong et al., 2025). Validation is also typically indirect, using limited in-situ data (Arshad et al., 2025; Kalu et al., 2024; Kim et al.,

2024; Vishwakarma et al., 2021; Wang et al., 2024). Directly validated, high-resolution GRACE downscaling has not yet been applied in Arizona.

This research aims to fill these knowledge gaps by producing a 1 km groundwater storage anomaly (GWSA) dataset for Arizona State. The method incorporates cumulative precipitation, evapotranspiration and runoff to capture groundwater system dynamics and mass balance, which is a novel approach at 1km scale downscaling. It is validated against well observations from the Arizona Department of Water Resources (ADWR), enabling spatial analysis across Arizona's Lower Colorado River sub-basin. At this resolution, hotspots of depletion and recovery or effects of managed aquifer recharge can be identified, providing more relevant information for Arizona groundwater management and policy.

2 Materials and Methods

2.1 Area of Interest

This study focuses on the state of Arizona, illustrated in **Figure 1**, located in the southwestern United States, covering a geographic extent of approximately 32°N to 37°N latitude and 109°W to 114°W longitude (Abdelmohsen et al., 2025). Topographically, Arizona is a diverse land with desert basins, plateaus and mountainous terrain, with large river systems like the Colorado River defining significant hydrological zones (Brandi et al., 2023; Lukas et al.).



Figure 1: Area of Interest (source: <https://www.istockphoto.com/photo/arizona-us-state-topographic-map-3d-render-color-county-border-gm2234985497-650362906>)

One of the main characteristics of the study area is the existence of active management areas (AMAs), in which the consumption of groundwater is strictly controlled because of the high demand and water stress in the long term (McGreal et al., 2021; Tillman & Leake, 2010). The key AMAs that will be discussed within the framework of the current research are Phoenix, Pinal, Tucson, Santa Cruz, Prescott, and Douglas. These areas are the most exploited groundwater basins in the state and play a vital role in urban supply, agriculture and sustainability of water in the region (McGreal et al., 2021).

2.2 Data Collection

Climatological variables, satellite data and in-situ observation well data have been collected for analysis, as shown in **Table 1**.

Table 1. Data Properties

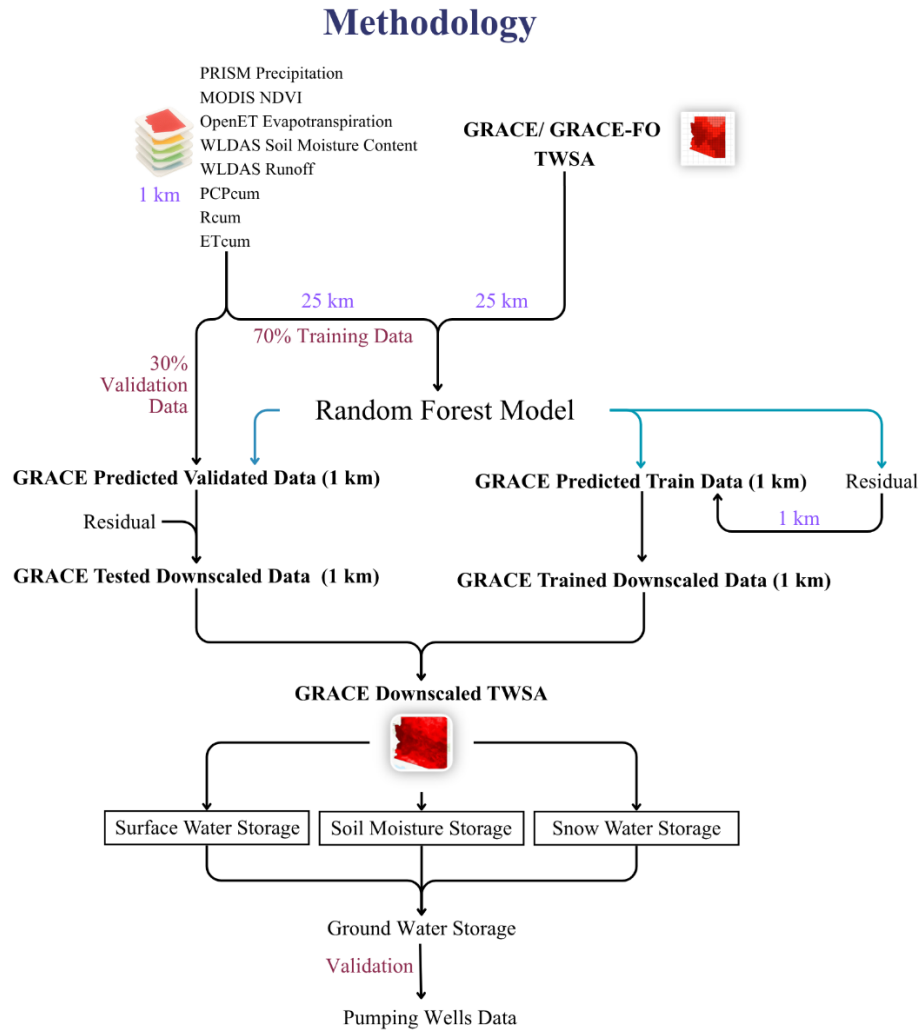
Data Property	Variable	Source	Spatial Resolution	Temporal Resolution	Begin Data	End Date
Gridded Hydro Climatic and Land Surface Variables	Precipitation	Prism	800m × 800m	Daily	April 2002	December 2023
	Evapotranspiration	OpenET	30m × 30m	Daily		
	Runoff	WLDAS	1000m × 1000m	Daily		
	NDVI	MODIS	1000m × 1000m	Monthly		
	Soil Moisture	WLDAS	1000m × 1000m	Daily		
Satellite-Based Terrestrial Water Storage Data	Terrestrial Water Storage Anomalies (TWSA)	CSR Mascon	25KM × 25KM 0.25° × 0.25°	Monthly		
In-Situ Observation	Wells Data	Arizona Department of Water Resources				

2.3 Methodology

The main goal of this study is to estimate groundwater storage anomalies (GWSA) at a spatial resolution of 1 km by developing a reliable downscaling model for GRACE terrestrial water storage anomalies (TWSA). To achieve this, a Random Forest regression model was used with hydroclimatic input variables. We initially worked with five variables: precipitation, evapotranspiration, runoff, NDVI and soil moisture. Among these, precipitation,

evapotranspiration and runoff are flux variables, meaning they represent short-term water movement. After doing some trials, we concluded that using only these instantaneous fluxes was not enough to explain long-term storage changes. To better capture the gravitational mass changes, we added cumulative precipitation, cumulative evapotranspiration and cumulative runoff as additional inputs, resulting in eight predictor variables in total. Before building the model, we examined the time series of the main variables from April 2002 to December 2023 to understand their seasonal behavior, timing of peaks and lagged responses. These checks helped confirm that the chosen variables were physically meaningful. After that, all input variables at 1 km resolution were first aggregated to the 25 km GRACE scale to match the resolution of the GRACE (target variable). Then the Random Forest model was tuned using the aggregated data and GRACE TWSA to figure out the best set of model hyperparameters. The dataset was then divided into 70% for training and 30% for testing using a spatially stratified split, as shown in **Figure 2(b)**. Finally, the RF model was trained and the trained model was then tested by predicting GRACE TWSA for the remaining 30% of the data and comparing it with the original GRACE observations. After confirming good performance at the GRACE scale, the model was applied to the full 1 km input data to produce downscaled TWSA. To ensure consistency with GRACE, the residual between observed and predicted GRACE TWSA at 25 km was calculated using Inverse Distance Weighting (IDW), interpolated to 1 km and added back to the downscaled results. Finally, groundwater storage anomalies were estimated from the downscaled TWSA using a water balance approach and were compared with in-situ groundwater level observations for validation.

(a)



(b)

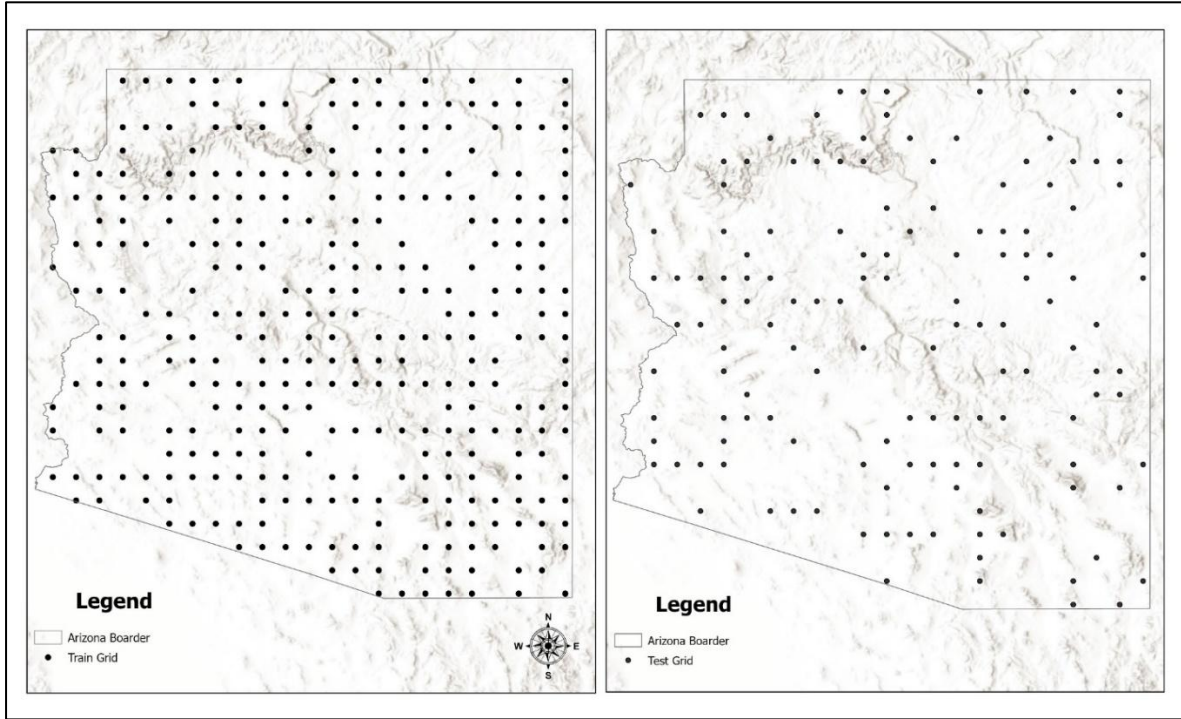


Figure 2. Overview of the downscaling framework and spatial sampling strategy. (a) Schematic of the Random Forest–based downscaling approach used to generate 1 km terrestrial water storage anomalies (TWSA) from GRACE/GRACE-FO observations (25 km). (b) Spatial distribution of grid points within the study area showing their division into training (70%) (Left) and testing (30%) (Right) subsets used for model development and validation.

3 Results and Analysis

3.1 Random Forest Model

We used a spatially stratified cross-validation structure to investigate the predictive performance of the Random Forest (RF) model in order to maximize the predictive power and avoid information leakage among the neighboring grid cells. The spatial cross-validation was used to evaluate each combination and the model performance was evaluated based on the average RMSE of all folds. The detailed description of hyperparameter tuning is given in Supporting Information **Text S1**. The final hyper-tuning parameters were 365 trees, an mtry of 6 and a minimum node size of 10. Developing a robust and generalizable framework to downscale the GRACE-derived groundwater storage anomalies (GWSA) involved an extensive iterative process. It was done through multiple iterations where combinations of input variables and alternative model configurations were

systematically tested. Based on this iterative evaluation and the finding that longer-term accumulation of water signals results in greater correlation, the final modeling configuration with 8 physically meaningful predictor variables was chosen. Three cumulative flux variables were added to the five instantaneous flux variables: Pcum (Cumulative Precipitation), ETcum (Cumulative Evapotranspiration) and Rcum (Cumulative Runoff).

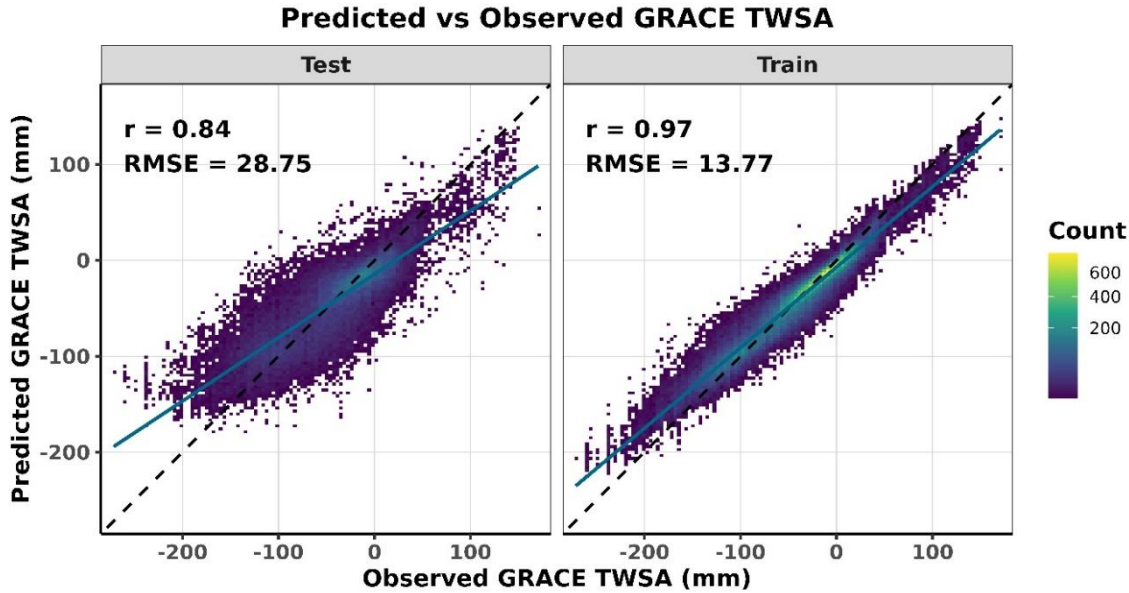


Figure 3. Model training and validation performance for GRACE-based downscaling. Density scatter plots comparing observed and predicted GRACE total water storage anomalies (TWSA) for testing (left) and the training (right) datasets. Color intensity represents data density, and the solid line indicates the fitted regression, with the dashed line denoting the 1:1 reference.

The initial evaluation of model performance compared predicted and observed total water storage anomalies (TWSA) using density scatter plots for both training and testing datasets, as illustrated in **Figure 3**. The model shows strong predictive performance during training, where the correlation coefficient ($r = 0.97$) is high and the RMSE of 13.77 mm is low. The density scatter plot correspondingly gives a linear fit along the 1:1 identity line with a high-density center of approximately 600 counts. This concentration implies that most of the predictions are similar to those observed and that the model effectively captures the prevailing spatio-temporal variability during learning. The model has a good predictive capability with a correlation of $r = 0.84$ when assessed in independent test data. Under an independent validation, the RMSE rises to 28.75 mm, indicating a wider scope of hydrological conditions that were not experienced during training. Although the test-phase density scatter plot has a less concentrated distribution, with the highest

density centers of approximately 400 counts, it maintains a linear relationship. While there is a slight tendency to underestimate at higher observed values, the overall agreement is significant, indicating that the model can be generalized well beyond the calibration space, as demonstrated in **Figure 3**.

3.2 Comparison with Wells Data

To assess the performance of the 1-km groundwater storage anomaly (GWSA) model, a ground-based dataset was compiled from active index (AI) wells in Arizona from the Arizona Groundwater Site Inventory (AGSI). A total of 1566 active index wells exist in the statewide monitoring network, of which 811 are located within the designated AMAs. These AI wells correspond to Arizona's Active Management Areas (AMAs) and other groundwater basins where groundwater monitoring networks are maintained, making them appropriate for groundwater monitoring and analysis (Arizona Department Of Water Resources, 2021).

To further assess the spatial fidelity of the downscaled GWSA product, a correlation between individual AMA index wells and their nearest GWSA grid cell was computed at the original 25 km GRACE resolution and at the downscaled 1 km resolution, as shown in **Figure 4**.

At 25 km resolution, correlations are spatially mixed with a prevalence of negative values in the central and southern parts of the domain, especially in the Phoenix and Tucson AMAs. At 1 km resolution, the pattern is more differentiated and positive correlations are visibly increased in the northern and western portions of the Phoenix AMA and clusters of improvement are also evident in the Prescott and Santa Cruz AMAs. However, a subset of wells in the southern and southeastern parts of the domain exhibit degraded or neutral correlation at 1 km.

The patterns of correlation between the AMA subregions have shown a significant diagnostic finding. The 1-km downscaled GWSA is effective in capturing the natural hydroclimatic aspect of groundwater storage change in basins which receive CAP water. The added variability of Managed Aquifer Recharge (MAR) and surface water imports - processes that do not respond to climate signals (Mohajer et al. 2026)- reflects a layer of anthropogenic influence overlaid onto this natural signal. The fact that the model can work well in most of the domain and also be able to detect the areas where this complexity occurs is in itself one of the key strengths. It shows that the 1-km product has the potential to spatially differentiate between climate-driven and infrastructure-influenced groundwater dynamics, which is a useful tool in CAP-related water management and planning.

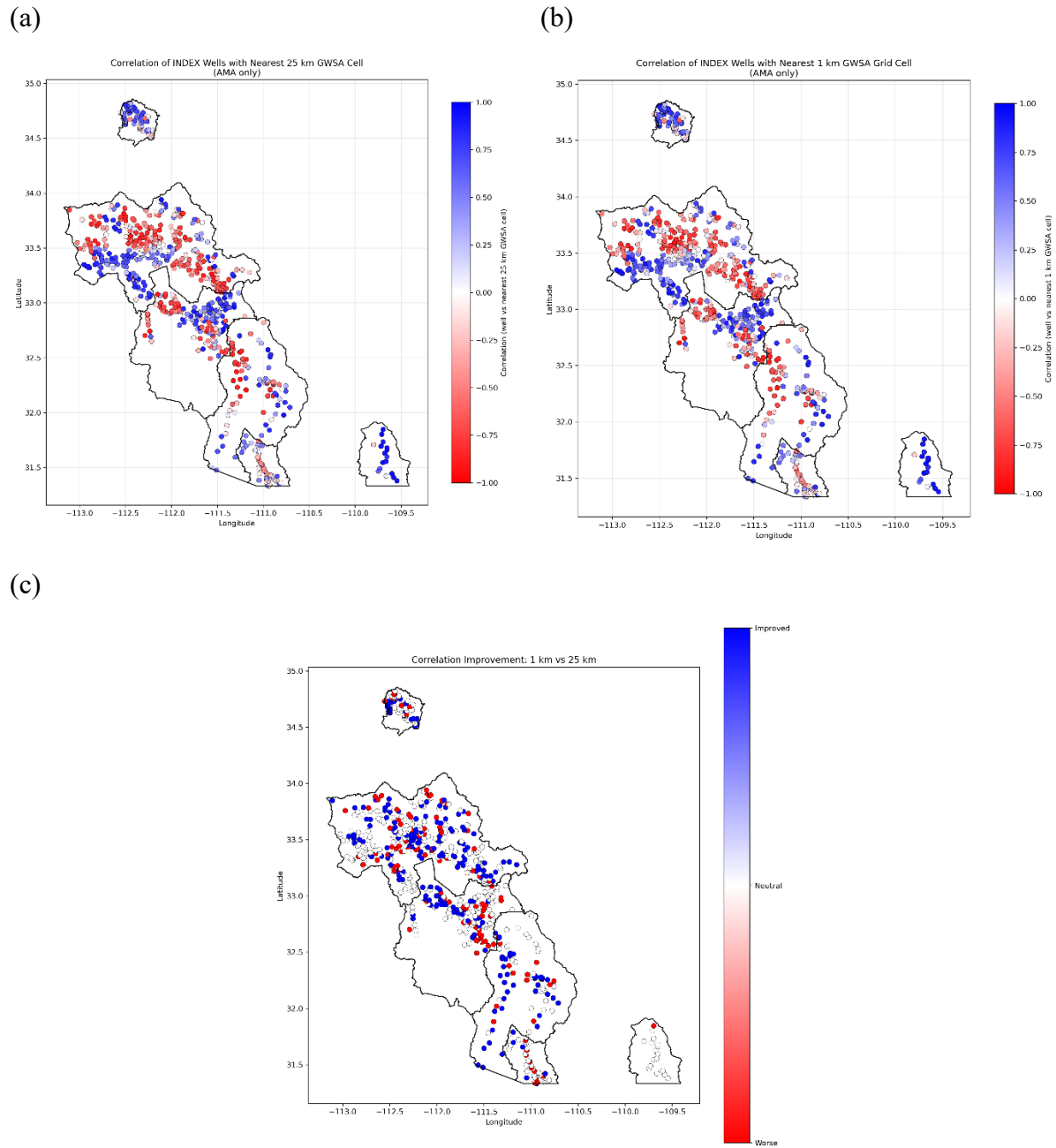


Figure 4. (a) Correlation of INDEX Wells with Nearest 25 km GWSA Cell (AMA only) (b) Correlation of INDEX Wells with Nearest 1 km GWSA Grid Cell (AMA only) (c) Correlation Improvement: 1 km vs 25 km

3.3 Importance of the regional observed GWSA at 1km

Figure 5 compares the groundwater storage anomaly (GWSA), derived by subtracting soil moisture storage and snow water storage from the downscaled TWSA, across three spatial scales: (i) original GRACE trends at 25-km resolution, (ii) downscaled trends at 1-km resolution, and (iii)

well-based point trends from April 2002 to 2023. The comparison highlights scale-dependent signal attenuation, spatial smoothing and depletion-recovery decoupling. The GRACE trend map of the 25 km (left panel) is characterized by large adjacent pixels with similar trend values, as shown in **Figure 5** (left). The majority of the pixels show negative GWSA patterns, with most in the range of about -3 to -7 mm yr⁻¹, represented by orange to dark brown colors. Each pixel represents an area of approximately 625 km², resulting in substantial spatial averaging. At this scale, trend gradients are small and spatial contrasts between adjacent hydrologic units are minimal. Entire AMAs and sub-basins can fall within the same trend class, making it difficult to understand localized groundwater responses to recharge projects or localized pumping. No discrete areas of depletion hotspots or recovery areas are identifiable; rather, a net loss across the region is visible in the regional signal.

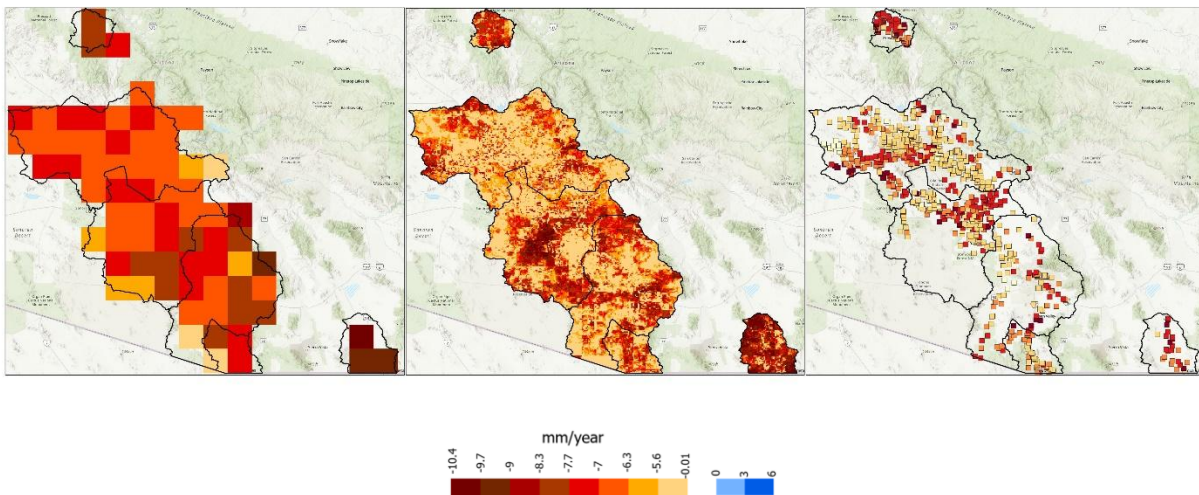


Figure 5. Scale-dependent spatial patterns of groundwater storage anomaly (GWSA) trends across Arizona. Comparison of (Left) original GRACE-derived GWSA trends at 25-km resolution, (Middle) downscaled GWSA trends at 1-km resolution and (Right) well-based groundwater storage trends within Active Management Areas (AMAs). Color scale represents annual trend magnitude (mm yr⁻¹).

The 1-km GWSA trend map (central panel) reveals considerable spatial heterogeneity that cannot be observed at the coarse resolution, as illustrated in **Figure 5** (middle). The trend magnitudes span a wider range, from near-neutral (range: -0.01 to -1 mm yr⁻¹, light yellow) to extreme depletion (-10.4 mm yr⁻¹, deep maroon to dark red). The depletion hotspots become spatially coherent clusters rather than isolated pixels. Notably, in areas that appear uniformly declining at

the 25-km scale, the 1-km map reveals sharp internal contrasts, with neighboring pixels varying by over 6–8 mm yr⁻¹ in trend strength.

The right panel superimposes well-derived GWSA trends on the spatial domain, as shown in **Figure 5** (right). The majority of wells show negative trends, consistent with the prevailing red and maroon areas in the 1-km map. The spatial correspondence between declining wells and high-intensity depletion areas confirms that the downscaled product captures the sign and relative magnitude of local groundwater loss.

This spatial analysis shows that coarse GRACE trends systematically underestimate localized extremes and suppress spatial variability, while 1-km downscaling recovers trend magnitude, contrast and heterogeneity that agree with observed well behavior, as demonstrated in **Figure 5**. The downscaled GWSA product is thus a more diagnostically informative representation of groundwater processes, particularly in detecting hotspot depletion and local recovery areas that cannot be identified at the native satellite resolution.

3.3.1 GW Basin Scale

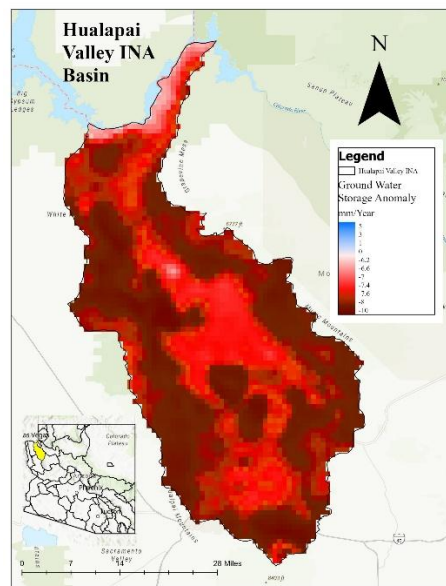
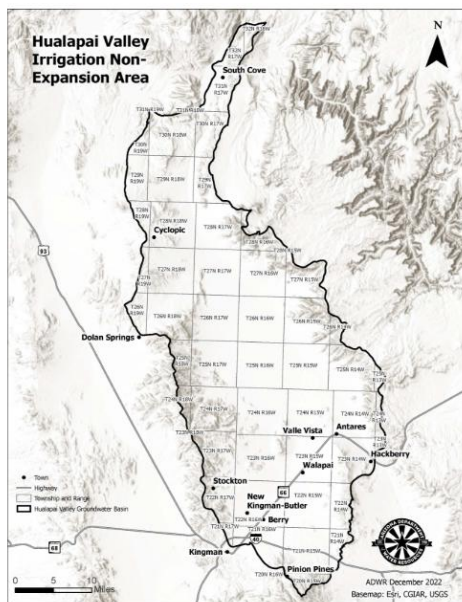
One of the greatest benefits of downscaling GRACE satellite data from 25 km to 1 km is that we can now see changes in groundwater at a much finer level, almost like zooming into a map from a regional view down to a neighborhood view. The Hualapai Valley and the Sacramento Valley, two groundwater basins in northwestern Arizona, demonstrate precisely why this matters (**Figure 6**).

The 1-km map of the Hualapai Valley Basin reveals concerning trends regarding groundwater levels (**Figure 6a**). The entire basin is experiencing a loss of groundwater, with most regions showing a decline of about –7.4 to –10 mm/year. The finer resolution is so useful, though, because it does not simply tell us that the basin is losing water. It shows us where that loss is happening the most. The steepest declines are being recorded in the central and southern parts of the basin, particularly around Valle Vista, Walapai and the area just north of Kingman. If we only had the original coarse GRACE data, all of this would look like one big average number spread across a large pixel. The 1-km product breaks that down and reveals the specific zones under the most pressure. This is especially relevant given that Hualapai Valley is already designated as an Irrigation Non-Expansion Area, meaning water authorities have recognized this basin as a place where groundwater use needs to be limited.

The same pattern holds for the Sacramento Valley Basin (**Figure 6b**). The land uses in this basin include agricultural fields, industrial facilities and communities such as Golden Valley and Yucca,

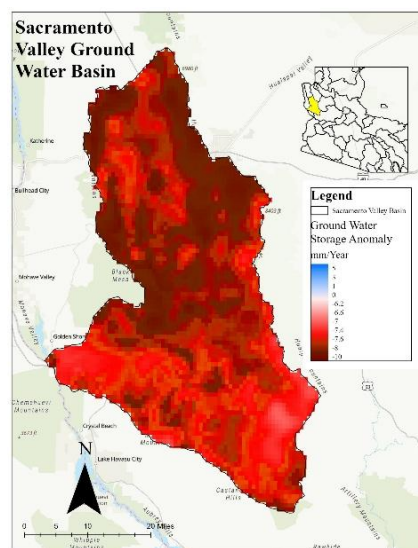
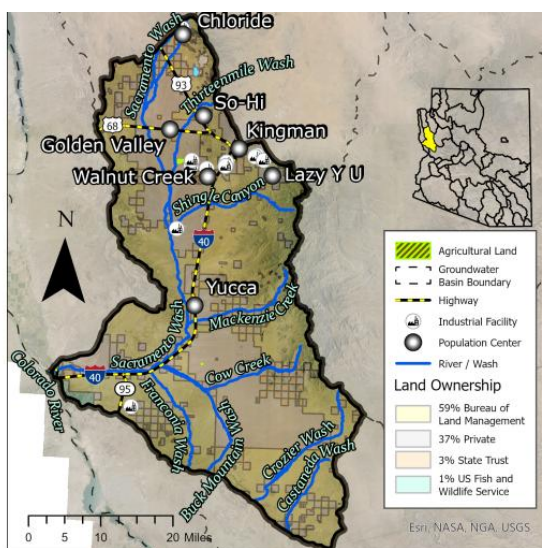
spread across an area where most of the land is federally managed. The groundwater depletion trends here also fall within the range of -7.6 to -10 mm per year across much of the basin. Looking at the 1-km map, two areas stand out as particularly stressed. The first is in the northern part of the basin, covering Golden Valley and west of Kingman, where urban and industrial water demand is likely driving higher extraction rates. The second runs along the Sacramento Wash corridor near Yucca, where agricultural activity appears to be contributing to continued storage losses.

(a)



Source: <https://www.azwater.gov/ama/ina/hualapai-ina>

(b)



Source:

https://www.azwater.gov/sites/default/files/2025-11/Sacramento%20Valley_0.pdf

Figure 6. Groundwater storage anomaly (GWSA) trends at 1-km resolution for the Hualapai Valley and Sacramento Valley basins, northwestern Arizona (2002–2023).

What both basins make clear is that the 1-km downscaled data is genuinely useful for decision-making, not only for researchers. It gives water managers a way to identify which specific parts of a basin need attention, rather than treating an entire basin as a single problem.

4 Discussion

The study generated a 1-km groundwater storage anomaly (GWSA) dataset for Arizona through downscaling of GRACE satellite data by a Random Forest model trained on cumulative hydroclimatic predictors. The most significant observation made on a water management scale is what the 1-km resolution reveals. At the original 25-km GRACE scale, the whole Lower Colorado River Basin would be seen as a patchwork of moderate negative storage trends distributed across large homogeneous pixels. In the same region, at 1 km, the same region tells a much more detailed story that the depletion hotspots as steep as ~ -10.4 mm/year emerge in certain sub-basins, separated by only a few kilometers by areas of near-neutral or recovering storage. In basins like Hualapai Valley and Sacramento Valley, both within the Lower Colorado River system, the 1-km data pinpoints concentrated groundwater loss tied directly to agricultural extraction and urban demand. This is precisely the type of information required to determine where groundwater restrictions, monitoring or recharge investments should be located first.

Connection to the Central Arizona Project (CAP) is the one that is directly connected to these results. CAP was built to reduce Arizona's dependence on groundwater by routing Colorado River water across the state, but as river flows have declined by roughly 20% over the past century, that reliability has eroded. The groundwater systems that CAP was built to defend are once more under strain - and that strain is not distributed evenly. The 1-km GWSA trends strongly distinguish between basins with rapid rates of depletion and basins with relatively stable conditions of storage. In the Phoenix and Pinal AMAs where CAP deliveries and managed aquifer recharge (MAR) programs have long been in operation, the 1-km product is found to capture the spatial signature of the CAP delivery and the managed aquifer recharge (MAR) program in a manner that simply

cannot be seen at 25 km. Areas with history of sustained recharge show significantly different trend patterns than areas where extraction is the only management activity. The correlation patterns with AMA subregions further indicate that in basins receiving CAP water, the natural hydroclimatic component of change in groundwater storage is successfully captured by the 1-km downscaled GWSA. The additional complexity of MAR and surface water imports, which are processes that do not depend on climate signals, is well documented by Mohajer et al. (2026) and is a layer of variability that sits over the natural signal that the model is resolving. The fact that the model can be used to determine exactly where managed recharge is creating a more complex signal, in itself is a valuable diagnostic outcome, demonstrating that the 1-km product can spatially differentiate between naturally driven and infrastructure-influenced groundwater behavior - a capability with direct value to CAP water management planning.

In addition to the AMAs, the 1-km data set is directly applicable to the wider Lower Colorado River Basin. Research based on conventional GRACE data has reported an overall loss of $36.0 \pm 6.2 \text{ km}^3$ of water storage across the basin since 2002 of which 71% is due to groundwater. However, these basin-wide figures do not provide much local action guidance. The downscaled version of the product developed here is already starting to fill that gap - it can be applied across the entire Arizona portion of the Lower Colorado River Basin and it will identify which sub-basins are contributing most to cumulative loss, which areas are showing the first signs of stabilization, and where recharge efforts are already producing measurable results on the ground. This type of sub-basin diagnosis has never been possible in the past with satellite data alone and represents a significant step in how the Lower Colorado River groundwater problem can be understood and communicated to decision-makers.

These results have a significant additional physical plausibility, attributable to the well-based validation of these findings across the Arizona AMA network. The majority of GRACE downscaling research throughout the globe is based on indirect validation, including comparisons of water balances or regional means. In this case, the downscaled GWSA was directly compared to in-situ water level measurements in the active index well network of ADWR. The 1-km product is able to capture the correct sign and overall magnitude of groundwater change across the AMAs, and well correlations improve at 1 km as compared to 25km across much of the domain - confirming that higher resolution does bring the satellite signal closer to what is happening in the aquifer. Such direct validation is particularly important, taking into account the fact that the

network of well monitoring in Arizona has not been evenly distributed. There are only a few long-term observation wells in many basins, especially those not located within the AMAs. In these, the 1-km satellite based GWSA could be the only spatially continuous groundwater data available, which is a crucial tool in the identification of stressed basins before the situation becomes irreversible.

To conclude, the 1-km downscaled GWSA dataset shifts Arizona groundwater monitoring significantly nearer to the spatial resolution practical water management demands. It demonstrates depletion and recovery trends not visible in the standard GRACE data, relates directly to the issues of CAP and the Colorado River system and is a framework that can be expanded across the entire Lower Colorado River Basin.

5 Conclusion

This paper shows that GRACE satellite data can effectively be downscaled to 1 km resolution to support feasible groundwater management on a statewide scale in Arizona and also on the larger Lower Colorado River Basin. The most significant advancement in the methodology is the introduction of cumulative mass predictors such as accumulated precipitation, evapotranspiration and runoff which enabled the Random Forest model to take into account the long-term memory of groundwater systems and not just short-term water movement. This design choice improved the model's correlation to $r = 0.84$ with the original GRACE signal variance, albeit with a minor systematic bias. The practical usefulness of this resolution enhancement is evident. Where standard GRACE data shows a broad, uniform band of groundwater loss across Arizona, the 1-km product reveals depletion hotspots reaching -10.4 mm/year in specific sub-basins, separated from areas of comparatively stable or recovering storage nearby. The spatial fingerprint of managed aquifer recharge programs is visible at this scale in a manner that just cannot be perceived at 25 km. The 1-km trends identify exactly where in each basin groundwater stress is most acute, providing water managers with the spatial information they need to take action.

Comparison with ADWR active index wells validates that the downscaled product is capturing the correct direction, magnitude and spatial pattern of groundwater change across the AMAs, with well correlations improving at 1 km compared to the native GRACE resolution across much of the domain.

As Colorado River supplies keep dwindling and pressure on Arizona aquifers keeps increasing, such tools are not an option anymore, but a necessity. The framework offers a framework upon which it is possible to build basin by basin across the entire Lower Colorado River system to support the data-driven groundwater decisions that the future of Arizona water depends on.

References

- Abdelmohsen, K., Famiglietti, J. S., Ao, Y. Z., Mohajer, B., & Chandanpurkar, H. A. (2025). Declining Freshwater Availability in the Colorado River Basin Threatens Sustainability of Its Critical Groundwater Supplies. *Geophysical Research Letters*, 52(10), e2025GL115593. <https://doi.org/10.1029/2025GL115593>
- ADWR. (2024). *ADWR holds informational meeting in Willcox on deteriorating groundwater conditions in the region.*
- Arizona Department Of Water Resources. (2021).
- Arshad, A., Mirchi, A., Taghvaeian, S., & AghaKouchak, A. (2024). Downscaled-GRACE Data Reveal Anthropogenic and Climate-Induced Water Storage Decline Across the Indus Basin. *Water Resources Research*, 60(7), e2023WR035882. <https://doi.org/10.1029/2023WR035882>
- Arshad, A., Shafeeque, M., Tran, T. N. D., Mirchi, A., Xiang, Z., He, C., AghaKouchak, A., Besnier, J., & Rahman, M. M. (2025). Multi-model ensemble machine learning-based downscaling and projection of GRACE data reveals groundwater decline in Saudi Arabia throughout the 21st century. *Journal of Hydrology: Regional Studies*, 60, 102552. <https://doi.org/10.1016/j.ejrh.2025.102552>
- Brandi, A., Balling, R. C., Iñiguez, P., & Georgescu, M. (2023). Climatological trends of mean and extreme daily precipitation in Arizona (USA). *Journal of Arid Environments*, 219, 105083. <https://doi.org/10.1016/j.jaridenv.2023.105083>
- Castle, S. L., Thomas, B. F., Reager, J. T., Rodell, M., Swenson, S. C., & Famiglietti, J. S. (2014). Groundwater depletion during drought threatens future water security of the Colorado River Basin. *Geophysical Research Letters*, 41(16), 5904–5911. <https://doi.org/10.1002/2014GL061055>
- Central Arizona Project. (2025). *History of CAP*. <https://www.cap-az.com/about/history-of-cap/>
- Famiglietti, J. S. (2014). The global groundwater crisis. *Nature Climate Change*, 4(11), 945–948. <https://doi.org/10.1038/nclimate2425>
- Famiglietti, J. S., Lo, M., Ho, S. L., Bethune, J., Anderson, K. J., Syed, T. H., Swenson, S. C., De Linage, C. R., & Rodell, M. (2011). Satellites measure recent rates of groundwater depletion in California's Central Valley. *Geophysical Research Letters*, 38(3), 2010GL046442. <https://doi.org/10.1029/2010GL046442>
- Fatolazadeh, F., Eshagh, M., & Goïta, K. (2022). New spectro-spatial downscaling approach for terrestrial and groundwater storage variations estimated by GRACE models. *Journal of Hydrology*, 615, 128635. <https://doi.org/10.1016/j.jhydrol.2022.128635>

- Hamou-Ali, Y., Karmouda, N., Mohsine, I., Bouramtane, T., Kacimi, I., Tweed, S., Tahiri, M., Kassou, N., El Bilali, A., Chafki, O., Ezzaouini, M. A., Laraichi, S., Zerouali, A., & Leblanc, M. (2025). Downscaling GRACE total water storage data using random forest: A three-round validation approach under drought conditions. *Frontiers in Water*, 7, 1545821. <https://doi.org/10.3389/frwa.2025.1545821>
- Houborg, R., Rodell, M., Li, B., Reichle, R., & Zaitchik, B. F. (2012). Drought indicators based on model-assimilated Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage observations. *Water Resources Research*, 48(7), 2011WR011291. <https://doi.org/10.1029/2011WR011291>
- Kalu, I., Ndehedehe, C. E., Ferreira, V. G., Janardhanan, S., Currell, M., & Kennard, M. J. (2024). Statistical downscaling of GRACE terrestrial water storage changes based on the Australian Water Outlook model. *Scientific Reports*, 14(1), 10113. <https://doi.org/10.1038/s41598-024-60366-2>
- Kim, J., Seo, K., Kim, B., Ryu, D., Chen, J., & Wilson, C. (2024). High-Resolution Terrestrial Water Storage Estimates From GRACE and Land Surface Models. *Water Resources Research*, 60(2), e2023WR035483. <https://doi.org/10.1029/2023WR035483>
- Landerer, F. W., & Swenson, S. C. (2012). Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resources Research*, 48(4), 2011WR011453. <https://doi.org/10.1029/2011WR011453>
- Long, D., Yang, Y., Wada, Y., Hong, Y., Liang, W., Chen, Y., Yong, B., Hou, A., Wei, J., & Chen, L. (2015). Deriving scaling factors using a global hydrological model to restore GRACE total water storage changes for China's Yangtze River Basin. *Remote Sensing of Environment*, 168, 177–193. <https://doi.org/10.1016/j.rse.2015.07.003>
- Lukas, J., Payton, L., McAfee, S., Wood, A., Woodhouse, C., Harding, B., Woelders, L., Smith, R., Gutmann, E., Lehner, F., Barsugli, J., Wolter, K., Rangwala, I., Duncan, B., Deems, J., Jerla, C., & Prairie, J. (n.d.). *Colorado River Basin Climate and Hydrology: State of the Science*.
- McGreal, B., Eden, S., & Polle, J. (2021). *ARIZONA GROUNDWATER MANAGEMENT - PAST, PRESENT AND FUTURE*.
- Milman, A., Bonnell, C., Maguire, R., Sorensen, K., & Blomquist, W. (2021). Groundwater Recharge for Water Security. *Case Studies in the Environment*, 5(1), 1113999. <https://doi.org/10.1525/cse.2020.1113999>
- Miro, M., & Famiglietti, J. (2018). Downscaling GRACE Remote Sensing Datasets to High-Resolution Groundwater Storage Change Maps of California's Central Valley. *Remote Sensing*, 10(1), 143. <https://doi.org/10.3390/rs10010143>
- Mohajer, B., Famiglietti, J. S., Chandanpurkar, H. A., & Hung, F. (2026). *Key natural influences on groundwater storage changes in central and southern Arizona*.
- O'Neill, B., Poupeau, F., Coeurdray, M., & Cortinas, J. (2016). *Laws of the river: Conflict and cooperation on the Colorado River*.

- Reager, J. T., Thomas, B. F., & Famiglietti, J. S. (2014). River basin flood potential inferred using GRACE gravity observations at several months lead time. *Nature Geoscience*, 7(8), 588–592. <https://doi.org/10.1038/ngeo2203>
- Rodell, M., Velicogna, I., & Famiglietti, J. S. (2009). Satellite-based estimates of groundwater depletion in India. *Nature*, 460(7258), 999–1002. <https://doi.org/10.1038/nature08238>
- Sahour, H., Sultan, M., Vazifedan, M., Abdelmohsen, K., Karki, S., Yellich, J., Gebremichael, E., Alshehri, F., & Elbayoumi, T. (2020). Statistical Applications to Downscale GRACE-Derived Terrestrial Water Storage Data and to Fill Temporal Gaps. *Remote Sensing*, 12(3), 533. <https://doi.org/10.3390/rs12030533>
- Scanlon, B. R., Faunt, C. C., Longuevergne, L., Reedy, R. C., Alley, W. M., McGuire, V. L., & McMahon, P. B. (2012). Groundwater depletion and sustainability of irrigation in the US High Plains and Central Valley. *Proceedings of the National Academy of Sciences*, 109(24), 9320–9325. <https://doi.org/10.1073/pnas.1200311109>
- Scanlon, B. R., Pool, D. R., Rateb, A., Conway, B., Sorensen, K., Udall, B., & Reedy, R. C. (2025). Multidecadal drought impacts on the Lower Colorado Basin with implications for future management. *Communications Earth & Environment*, 6(1), 214. <https://doi.org/10.1038/s43247-025-02149-9>
- Scanlon, B. R., Zhang, Z., Save, H., Wiese, D. N., Landarer, F. W., Long, D., Longuevergne, L., & Chen, J. (2016). Global evaluation of new GRACE mascon products for hydrologic applications. *Water Resources Research*, 52(12), 9412–9429. <https://doi.org/10.1002/2016WR019494>
- Tadych, D. E., Ford, M., Colby, B. G., & Condon, L. E. (2024). Historical patterns of well drilling and groundwater depth in Arizona considering groundwater regulation and surface water access. *JAWRA Journal of the American Water Resources Association*, 60(6), 1193–1208. <https://doi.org/10.1111/1752-1688.13234>
- Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F., & Watkins, M. M. (2004). GRACE Measurements of Mass Variability in the Earth System. *Science*, 305(5683), 503–505. <https://doi.org/10.1126/science.1099192>
- Tillman, F. D., & Leake, S. A. (2010). Trends in groundwater levels in wells in the active management areas of Arizona, USA. *Hydrogeology Journal*, 18(6), 1515–1524. <https://doi.org/10.1007/s10040-010-0603-3>
- USBR. (2026). *Colorado River Basin Post-2026 Operations*. <https://www.usbr.gov/ColoradoRiverBasin/post2026/index.html#:~:text=The%20Post%2D2026%20process%20includes%20the%20following%20documents,2026%20and%20ends%20on%20March%20%2C%202026.>
- Ventura, S. J. (2023). *Going Against the Current: Navigating Climate and Environmental Disparities in the Colorado River Basin*. Kent State University.
- Vishwakarma, B. D., Devaraju, B., & Sneeuw, N. (2018). What Is the Spatial Resolution of grace Satellite Products for Hydrology? *Remote Sensing*, 10(6), 852. <https://doi.org/10.3390/rs10060852>

- Vishwakarma, B. D., Zhang, J., & Sneeuw, N. (2021). Downscaling GRACE total water storage change using partial least squares regression. *Scientific Data*, 8(1), 95. <https://doi.org/10.1038/s41597-021-00862-6>
- Wahr, J., Swenson, S., Zlotnicki, V., & Velicogna, I. (2004). Time-variable gravity from GRACE: First results. *Geophysical Research Letters*, 31(11), 2004GL019779. <https://doi.org/10.1029/2004GL019779>
- Wang, Y., Li, C., Cui, Y., Cui, Y., Xu, Y., Hora, T., Zaveri, E., Rodella, A.-S., Bai, L., & Long, D. (2024). Spatial downscaling of GRACE-derived groundwater storage changes across diverse climates and human interventions with Random Forests. *Journal of Hydrology*, 640, 131708. <https://doi.org/10.1016/j.jhydrol.2024.131708>
- Xiong, Y., Feng, W., Bai, H., Chen, W., Jiang, Z., & Zhong, M. (2025). High-Resolution Terrestrial Water Storage Anomalies and Components in China From GRACE/GFO via Joint Inversion Downscaling. *Water Resources Research*, 61(7), e2024WR038996. <https://doi.org/10.1029/2024WR038996>
- Zaitchik, B. F., Rodell, M., & Reichle, R. H. (2008). Assimilation of GRACE Terrestrial Water Storage Data into a Land Surface Model: Results for the Mississippi River Basin. *Journal of Hydrometeorology*, 9(3), 535–548. <https://doi.org/10.1175/2007JHM951.1>
- Zhong, D., Wang, S., & Li, J. (2021). A Self-Calibration Variance-Component Model for Spatial Downscaling of GRACE Observations Using Land Surface Model Outputs. *Water Resources Research*, 57(1), e2020WR028944. <https://doi.org/10.1029/2020WR028944>

Supporting Information

Text S1: Random Forest Hyperparameter Tuning

To ensure that the Random Forest (RF) model provides reliable and generalizable predictions, a systematic hyperparameter tuning procedure was carried out using a spatial cross-validation framework. Instead of randomly splitting the data, the data were split according to spatial grid identifiers to prevent information leakage across adjacent pixels. There were 10 folds in total with each fold having independent spatial units and thus it was able to evaluate the model performance in space more realistically. Three important RF hyperparameters were optimised, namely the number of variables to be considered at each split (`mtry`), the size of the minimum node (`min.node.size`), and the number of trees in the ensemble (`num.trees`). These parameters regulate the complexity of models, stability and nonlinear relationship capturing. An exploration of parameter space was done efficiently using a randomized search approach. The following ranges were used in testing around 80 different combinations: `mtry`: 4 to 8, `min.node.size`: 10 to 50, `num.trees`: 200 to 550

The spatial cross-validation was used to evaluate each combination and the model performance was evaluated based on the average of the Root Mean Square Error (RMSE) of all folds. Further, standard deviation of RMSE was calculated to determine the stability of the model.

Interpretation of Hyperparameter Tuning Results

A series of diagnostic plots were used to understand how each parameter influenced model performance.

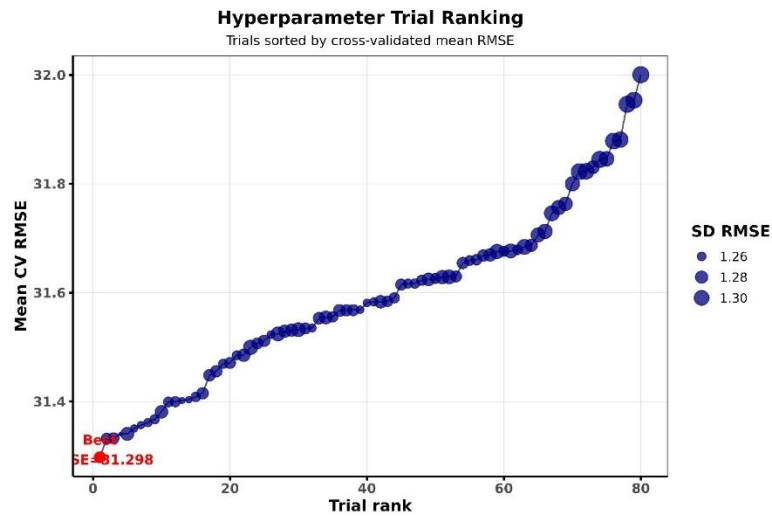


Figure S1. All tested parameter combinations sorted by their cross-validated RMSE. The results indicate a smooth improvement in performance toward the best models, with the optimal configuration achieving the lowest RMSE (~31.30 mm).

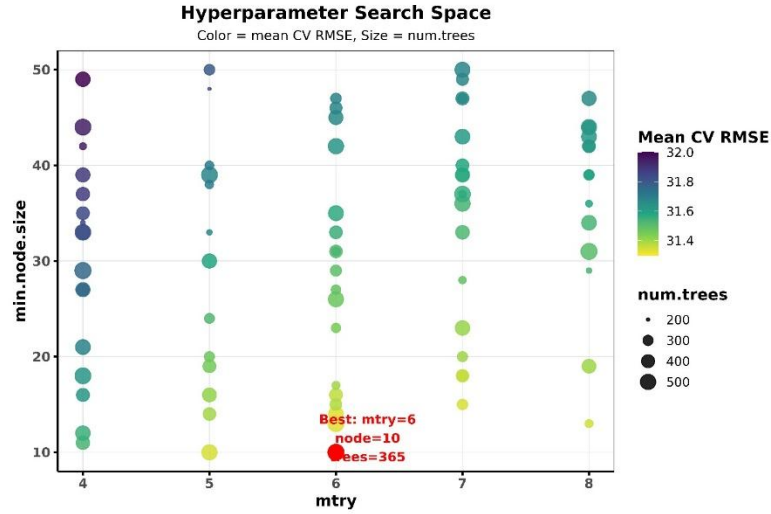


Figure S2. This plot illustrates the distribution of tested combinations in the parameter space. Color represents RMSE and point size represents the number of trees. Lower RMSE values are concentrated around moderate values of mtry and smaller node sizes, indicating that balanced model complexity leads to better performance.

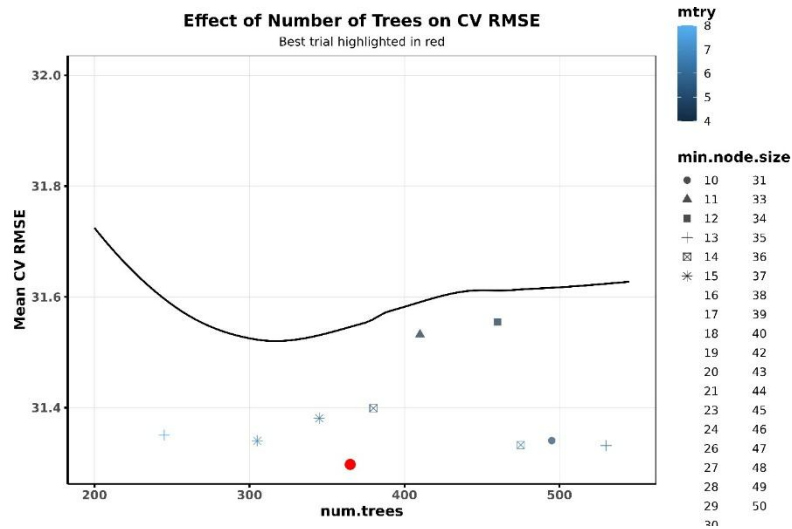


Figure S3. The relationship between the number of trees and model error shows a shallow U-shaped pattern. RMSE decreases as the number of trees increases up to around 300–400 trees, after which improvements plateau. This indicates that increasing the number of trees beyond this range does not significantly improve performance but increases computational cost.

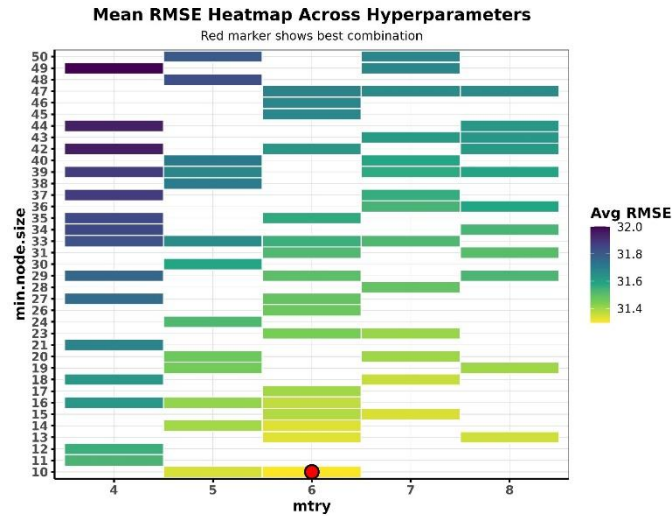


Figure S4: The heatmap highlights the combined effect of mtry and min.node.size. The lowest RMSE values are observed when mtry is around 6 and min.node.size is small (around 10–15). This suggests that allowing the model to consider a moderate number of variables while maintaining smaller terminal nodes improves predictive accuracy.

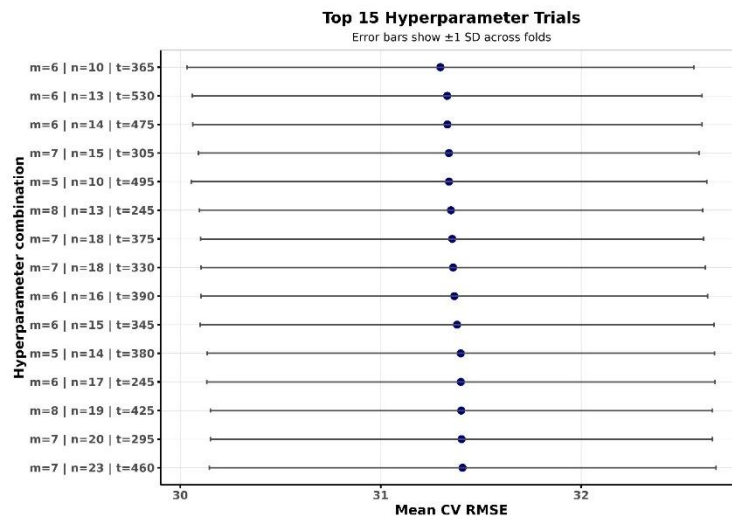


Figure S5. This figure shows the best-performing parameter combinations along with their variability across folds. The small error bars indicate that the top models are stable and consistent across different spatial subsets of the data.