Supplementary Materials for

High resolution US water table depth estimates reveal quantity of accessible groundwater

Yueling Ma, Laura E. Condon, Julian Koch, Andrew Bennett, Amy Defnet, Danielle Tijerina-Kreuzer, Peter Melchior, Reed M. Maxwell

Corresponding author: Laura E. Condon, lecondon@arizona.edu; Reed M. Maxwell, reedmaxwell@princeton.edu

The PDF file includes:

Materials and Methods

S1.1. Study Data

Long-term mean water table depth (WTD) observations over the contiguous United States (CONUS) were derived from historical WTD measurements over 1914-2023 at 502,839 groundwater monitoring wells archived by the United States Geological Survey (USGS), long-term mean WTD measurements over 1927-2009 at 452,139 wells from Fan et al.[4], and annual median WTD measurements over 1895-2022 at 190,613 wells in California and Texas provided by Jasechko et al.[2]. While we have WTD observations at over one million wells over the CONUS, WTD data are sparse at the city or individual farm level, where local decisions are often made. Over 60% of USGS groundwater monitoring wells only have one observation, which is presumed to represent the long-term mean for that well. This assumption might introduce additional uncertainty. Moreover, many wells are affected by anthropogenic activities such as groundwater pumping, resulting in long-term drying trends in observed WTD time series. All WTD observations have been validated using the same criteria as Ma et al.[30]*)*, which include limiting WTD to the range of 0 to 300 m. They were mapped to the CONUS domain at a resolution of 1 arc-second (~30 m) with nx = 209,904 and ny = 102,561 for constructing a random forest (RF) model. In cases where more than one well was mapped to a grid cell, we utilized the average of long-term mean WTD observations for all wells located in the same grid cell as the value for the grid cell. We filtered out the grid cells along streams with positive WTD. In total, we have WTD observations available at 726,828 grid cells over the CONUS (**Fig. S1**), ~0.003% of all the grid cells.

The inputs to the RF model are annual mean precipitation and temperature, precipitation minus evapotranspiration (PME), natural log of hydraulic conductivity (lnK), elevation, topographic slope, horizontal and vertical distances to streams, as well as mean sand and clay contents. Continuous spatial data covering the entire CONUS are accessible for all input variables, but at various resolutions. For input data with a spatial resolution coarser than 1 arc-second, we re-gridded the data to 1 arc-second × 1 arc-second via the nearest-neighbor interpolation method before feeding them into the RF model. The initial resolution of the annual mean precipitation and temperature, PME, and lnK data is 1 km. The annual mean precipitation and temperature data were computed from the precipitation and temperature datasets provided by the Center for Western Weather and Water Extremes (CW3E) for the years 1980-2020, respectively[49]. The PME data were extracted from the potential recharge data utilized in the ParFlow CONUS 2.0 platform[50]. The lnK data were derived from a continental-scale subsurface K dataset documented in Tijerina-Kreuzer et al.[34] and Swilley et al.[51]. The 1 arc-second resolution elevation data were obtained from the seamless USGS 3DEP digital elevation model (DEM) dataset[52]. Using the GDAL slope algorithm in QGIS 3.32[53], we calculated the topographic slope data at the same resolution from the elevation data. The stream networks for the CONUS were determined based on the 30-m resolution lateral position (LP) datasets for nine hydrologic orders generated by Belitz et al.[54], where LP=0 indicates that the grid cell is located along a stream. We also identified perennial streams using stream perenniality data produced by Brinkerhoff et al.[55] at the USGS Hydrologic Unit Code Level 4 (HUC4) scale. While the Belitz stream networks may contain ephemeral streams, the RF model significantly outperforms the model only considering perennial streams in identifying permanent surface water bodies. Thus, we apply the Belitz stream networks in this work. The horizontal distance to stream represents the closest linear distance to stream, calculated using the stream network data over the CONUS and the GDAL Proximity (raster distance) algorithm in QGIS 3.32. The vertical distance to

stream is the vertical distance to the interpolated stream network, computed as the difference between the actual elevation and the interpolated stream network base level elevation. The mean sand and clay contents were derived by averaging the 250-m resolution sand and clay contents from SoilGrids 2.0[56] over six layers ranging from 0 to 200 cm, respectively.

We also investigate the distribution of groundwater under agricultural regions (**Fig. 4**), which are closely related to food production and have been heavily impacted by anthropogenic activities. The land cover data were from the 30-m resolution NLCD 2021 Land Cover data[57] released by the National Land Cover Database (NLCD). Here we re-gridded the land cover data into the 1 arc-second grid cells, with a focus on crops.

S1.2. Random Forest Model

A RF model is trained to produce a WTD estimate map over CONUS at a spatial resolution of 1 arc-second. **Fig. S2** illustrates the schematic diagram of the RF model and the optimal settings of its hyperparameters (i.e., adjustable parameters that govern model behavior and complexity) that are determined by the cross-validation performance of the model during training (also referred to Ma et al.[30]). RF is an ensemble learning algorithm that operates by constructing multiple decision trees during the training period. Each decision tree provides independent estimates by utilizing a bootstrap sample of the original training dataset, generated through the "bagging" (or bootstrap aggregation) technique[58,59]. As such, the developed RF model internally produces a set of WTD estimates at the individual grid cell level with a size equal to the number of the decision trees involved (300 in this study). To eliminate the impact of anomalous tree outputs, the RF model returns the median prediction of the individual trees as the final WTD estimate for a grid cell, diverging from the typical practice of using the mean. Since the distributions of tree outputs at most grid cells are not Gaussian, we employ the 25th and 75th percentiles of tree outputs for each grid cell to quantify the uncertainty associated with the RF model. This approach, known as quantile regression forests, was introduced by Meinshausen[60] and has been successfully applied in many studies[31,61].

Although advanced machine learning (ML) methods, e.g., convolutional neural networks and Long Short-Term Memory networks, have shown success in groundwater hydrology[62-65], they depend on extensive spatially or temporally continuous data. Therefore, these methods are not suitable for this study, which deals with sparse point observations. The performance of physics-informed ML is hindered by the lack of pumping information in large-scale physically-based groundwater simulations and additional uncertainties associated with physical constraints. A known challenge for physics-informed neural networks (PINNs) is to balance the physics-informed and data-driven components in the loss function during training[66]. As such, recent applications of PINNs are mostly found in ideal cases. Instead, RF has a relatively simple structure with a small number of hyperparameters to tune, thus requiring less training data to maintain good model performance. Furthermore, RF usually does not require heavy hyperparameter tuning and data normalization, which simplifies the training process. Because of these advantages, RF has been successfully applied to produce a 1-km resolution WTD estimate map over most of the CONUS from limited WTD observations[30].

Here we utilize the data at 726,828 grid cells with WTD observations and randomly split them into 80% and 20% for training and testing the RF model, respectively. The test set is not exposed to the RF model during training, so it can be used to show the model's generalizability beyond the training data. Additionally, we introduce dummy data for 20,000 grid cells along the Belitz streams in the training dataset, where WTD = 0 is assumed. The number of grid cells with

dummy data is selected from a range of 1000 to 2M based on the optimal performance compared to true WTD data during training. The data augmentation aids the RF model in identifying saturated conditions along water bodies, including rivers and lakes[31]. The RF model effectively captures the nonlinear complex relationships between the input variables and WTD from gridded observations (as shown in **Fig. S3**), with a Pearson correlation coefficient (r) of 0.79 for the test set (0.96 when evaluating all data), a root mean square error (RMSE) of 14.94 m for the test set (6.72 m when evaluating all data), and a Nash-Sutcliffe efficiency (NSE) of 0.62 for the test set (0.93 when evaluating all data).

We assume that the input-output relationship learned from grid cells with WTD observations is applicable across the CONUS, and a map of WTD estimates for the CONUS (as displayed in **Fig. 2**) is justified using the trained RF model. We also assess the WTD map against unchanging permanent water surfaces from the 30-m Global Surface Water 1984-2021 dataset[67]. This dataset is based on remotely sensed data and have never been seen by the RF model. Our WTD product achieves a true positive rate of 78% (i.e., the rate of correctly identifying surface water bodies) and a true negative rate of 97% (i.e., the rate of correctly identifying groundwater). Zell and Sanford[68] performed a similar evaluation and compared simulated WTD with the USGS National Hydrography Dataset (NHD) perennial 1st order streams. Here we use a different surface water dataset for assessment, since the Belitz stream data were derived from the NHD streams and have been considered in the RF model.

We study the sensitivity of the RF model to input variables (**Fig. S4**) using permutation importance. The permutation importance is expressed by the relative increase in test RMSE when randomly shuffling the values of a single input variable. The five most important variables are elevation (99.55% increase in RMSE), annual mean temperature (94.63%), annual mean precipitation (52.72%), PME (20.42%), and vertical distance to stream (13.47%). The permutation importance calculated on the training set (not shown here) provides a similar order. The RF model is not mass conservative yet learns the variations in groundwater due to lateral flow likely using elevation as a surrogate for flow. The importance of elevation and climatology-related inputs indicates that lateral flow and atmosphere-land-surface-groundwater interactions dominate the spatial changes in groundwater over the CONUS. In contrast, inputs related to local hydrogeological processes (e.g., lnK and distances to streams) appear to be less significant, possibly because those processes are overshadowed by the distinct climate and topography gradients at the CONUS scale.

This study is performed on the Della cluster at Princeton University. The RF model is developed using scikit-learn[69], an open-source ML python library, with minor modifications to generate the median and percentiles of tree outputs. It takes ~ 0.5 hours to train the RF model on a 28-core 2.4 GHz Intel Broadwell processor with 128 GB of memory. We then divide the CONUS into 20 subdomains and run the trained RF model for each subdomain in parallel on a large-memory node with 6 TB of memory, requiring ~ 15 hours to complete. Hence, the RF model can provide fast and reliable estimates at a speed and scale not possible with traditional physically-based hydrological models.

We acknowledge that there is still much room for improvement in the WTD estimates for the CONUS, and the potential sources of bias are summarized as follows.

- Quality and quantity of the training data:

  The RF model is purely data-driven, and thus its behavior significantly depends on the quality and quantity of the data used for the model construction. Here we re-gridded data

originally at various spatial resolutions (from 30 m to 1 km) to 1 arc-second and mapped point WTD observations to 1 arc-second grid cells. The preprocessing processes may introduce uncertainty to the resulting WTD estimates. As mentioned in Section S1.1, many wells only have one observation, so it may be questionable how well the observation represents the long-term mean for that well. Moreover, relative to the large CONUS domain, our WTD observations are sparse, only covering ~0.003% of all the grid cells. They are unevenly distributed within the CONUS, as presented in **Fig. S1**. Ma et al.[30] found no obvious relationship between the density of WTD observations and the RF model performance in a large HUC2 basin. However, the limited WTD observations may not be representative for the distribution of groundwater resource at the city or individual farm level, thus constraining the model performance at very small scales.

- Uncertainty associated with the RF model:

As an ensemble learning method, the RF model provides information about the full distribution of the estimated WTD generated by the involved decision trees. **Fig. S5** and **S6** present the 25th and 75th percentiles of tree outputs at individual grid cells over the CONUS, and we utilize the difference between the 75th and 25th percentiles of tree outputs (i.e., the interquartile range, IQR) to represent the uncertainty associated with the RF model (as shown in **Fig. 4B**). Uncertain observations in combination with uncertain inputs make the splitting criteria more ambiguous resulting in varying tree results. As a result, the uncertainty quantified here is not limited to the uncertainty of the trees within the RF model. Instead, it also includes uncertainty in the observational data and uncertainty in inputs. The 25th-percentile map shows WTD ranging from 0 to 282.70 m, with a median of 6.10 m and a mean of 9.71 m, while the 75th-percentile map presents WTD ranging from 0 to 299.59 m, with a median of 24.84 m and a mean of 34.45 m. In addition to the median of the tree outputs, we estimate groundwater storage over the CONUS (see Section S1.3) using the 25th and 75th percentiles of tree outputs to quantify the impact of the uncertainty related to the RF model on the estimated groundwater storage.

- Anthropogenic influences and assumptions regarding the modeled scenario:

Here we focus on aquifers with WTD ≤ 300 m that have strong connections to the physics of the water system. Long-term groundwater depletion due to overexploitation (i.e., large and unsustainable groundwater pumping) has altered groundwater levels and changed base flow contributions from major aquifers[70,71]. Due to the lack of national pumping records, we do not explicitly incorporate anthropogenic impacts into the RF model. Nevertheless, the model somehow appears to learn or infer pumping from the WTD observations that will partly be affected by pumping, resulting in good performance.

Fig.S7 illustrates the testing performance of the trained RF model in the High Plains and California Central Valley aquifers, both known as hot spots of groundwater depletion[71]. Compared to pre-development groundwater conditions, water tables in many parts of the two aquifers have declined by > 30 m. In extreme cases, groundwater depletion in the southern Central Valley has exceeded 100 m[72]. The RF model has good testing performance in both aquifers with large groundwater depletion. As displayed in Fig. **S7C** and **S7F**, there are small differences between the estimated and observed WTD, on average < 2 m. The trained RF model obtains a test r of 0.9, a test RMSE of 12.27 m, and a test NSE of 0.81 in

the Central Valley and a test r of 0.82, a test RMSE of 12.88 m, and a test NSE of 0.67 in the High Plains.

We increase the estimated WTD in High Plains and California Central Valley aquifers by 4 m, 10 m, and 30 m to mimic heavy groundwater pumping. **Fig. S8** shows the comparison between the estimated and observed WTD in both aquifers. As the estimated WTD increases, the histograms of the estimated WTD show a greater mismatch with the one of the observed WTD. We recognize that this is an oversimplified experiment for pumping, since groundwater depletion varies spatially. However, the growing mismatch between the estimated and observed WTD after manually adding pumping provides further evidence that that pumping impacts have been already considered in our original WTD estimates.

- Time invariance:

This study aims to estimate spatial variations in WTD over the CONUS. The target variable is long-term mean WTD, which is static and time-invariant at each grid cell. Thus, the estimated WTD may not be truly representative of near real-time WTD, especially under changing climate and intensive human interventions. **Fig. S9** compares the estimated WTD to the most recent WTD observations (outside of the training set) from 1877 USGS groundwater monitoring wells that have at least one record from June 1 to August 31, 2024. Notably, this period is one of the hottest summers for the CONUS on record, and significant groundwater withdrawals happened during this period. It is encouraging that the data show good agreement, with a r of 0.85, a RMSE of 18.96 m, and a NSE of 0.72, and thereby aligning well with the test performance.

Section S2.1 presents the RF model testing performance for 18 HUC2 basins, including uncertainty maps showing the 25th and 75th percentiles of tree outputs at individual grid cells.

S1.3. Groundwater Storage Estimation

As expressed in Equation S1, groundwater storage across the CONUS is estimated based on the 1 arc-second resolution WTD estimates from the RF model and the 1-km resolution 10-layer porosity data for the same hydrostratigraphic categories employed in the ParFlow CONUS 2.0 modeling platform[50]. **Table S1** lists the porosity value for each hydrostratigraphic category. Similar to the ParFlow CONUS 2.0 platform, we assume that the subsurface is partitioned into ten vertical layers of varying thickness (200, 100, 50, 25, 10, 5, 1, 0.6, 0.3, and 0.1 m from bottom to top) over the CONUS, with a total subsurface depth (TSD) of 392 m[34]. Groundwater storage is given as

$$S_{GW\_CONUS} = \sum_{n\in\{0,N\}} \sum_{l=0}^{9} h_{l,n} \phi_{l,n} A, \quad \text{(S1)}$$

where N is the total number of grid cells over the CONUS, ϕ is the porosity for the layer l at the grid cell n, and A is the area of a grid cell. Depending on whether the layer l is fully saturated, the effective depth $h_{l,n}$ is calculated as follows.

$$h_{l,n} = \begin{cases} d_l & if\ TSD - WTD_n \geq \sum_{i=0}^{l} d_i\ (fully\ saturated) \\ (TSD - WTD_n) - \sum_{i=0}^{l-1} d_i & if\ TSD - WTD_n < \sum_{i=0}^{l} d_i\ (partially\ saturated) \end{cases}. \quad \text{(S2)}$$

Here d_l is the thickness of the layer l, and $\sum_{i=0}^{l-1} d_i$ is the accumulated thickness from the bottom (the layer 0) to the layer $l-1$.

We calculate groundwater storage for the CONUS using WTD estimates at 1 arc-second, 100 m, 1 km, 10 km, and 100 km resolutions (as displayed in **Fig. 3B**). The WTD estimates for the latter four resolutions are upscaled from the original 1 arc-second data using the mean function. During calculation, the porosity data at 1 km resolution are mapped to the corresponding resolution. To understand the influence of the uncertainty associated with the RF model on the groundwater storage estimation, we also estimate upper and lower groundwater storage limits for different resolutions (as illustrated in **Fig. 3B**) using the 25th and 75th percentiles of tree outputs from the model, respectively. Note that the upscaled WTD maps have smaller areas compared to the original 1 arc-second WTD map (as presented in **Table S2**). Nevertheless, the ratio of groundwater storage to area (**Table S2**) increases with the spatial resolution improves.

In addition to the potential biases in the WTD estimates discussed in Section S1.2 and the different spatial resolutions, the uncertainty in groundwater storage may result from the estimated porosity values and our simplified assumptions about the subsurface configuration. For example, we assume that there are ten vertically homogenous geologic layers in the subsurface, but in the real system, the geologic settings are more vertically heterogeneous. Moreover, we do not consider deep groundwater (below 392 m) in the calculation, most of which is brackish groundwater[37]. Ferguson et al.[35] recently found that deep groundwater has a minor contribution to the global hydrological cycle.

S1.4. Empirical Variogram

Here we utilize empirical variograms to quantify spatial dependence (also called spatial autocorrelation) in the estimated WTD at different resolutions, as shown in **Fig. S10**. As mentioned in Section S1.3, the WTD estimates with the resolutions of 100 m, 1 km, 10 km, and 100 km are upscaled from the 1 arc-second data using the mean function. The empirical variograms are estimated using the semivariogram function (Equation S3) defined by Matheron[73] as half the averaged squared difference in the estimated WTD between grid cells separated by a lag (i.e., Euclidean distance) L[74].

$$\gamma(L) = \frac{1}{2|N(L)|}\sum_{N(L)}(WTD_i - WTD_j)^2, \quad \text{(S3)}$$

where (WTD_i, WTD_j) is a pair of estimated WTD data separated by the lag $i-j=L$. $N(L)$ is the set of such data pairs, and $|N(L)|$ is the number of data pairs in $N(L)$.

A larger $\gamma(L)$ suggests a larger spatial difference in the estimated WTD for the lag L. For computational efficiency, we randomly sample 100,000 grid cells from the gridded WTD estimates for semivariogram model fitting, yielding ~ 5 billion data pairs. We also group the data pairs into 12 bins based on L (unit: m), which are [10,50], [50,10^2], [10^2,5*10^2], [5*10^2,10^3], [10^3,5*10^3], [5*10^3,10^4], [10^4,5*10^4], [5*10^4,10^5], [10^5,5*10^5], [5*10^5,10^6], [10^6,5*10^6], [5*10^6,10^7]. The resulting empirical variogram is the averaged estimate within each bin. The estimation of the empirical variograms is implemented using the open-source python package GSTools[75].

As illustrated in **Fig. S10**, utilizing finer resolution products allows us to obtain more information about the spatial changes in WTD at small scales, which is useful for farm scale decision making.

Supplementary Text

S2.1. Model Evaluation in 18 HUC2s

Fig. S11 to **S28** present the developed RF model's testing performance in 18 HUC2 basins compared with WTD observations, including maps showing the 25th and 75th percentiles of tree outputs from the RF model to represent the uncertainty associated with the RF model. In general, the RF model achieves reliable testing performance at the HUC2 level with r values ranging from 0.51 to 0.88 (mean: 0.68), RMSE values ranging from 6.25 m to 28.55 m (mean: 14.42 m), and NSE values ranging from 0.24 to 0.77 (mean: 0.46).

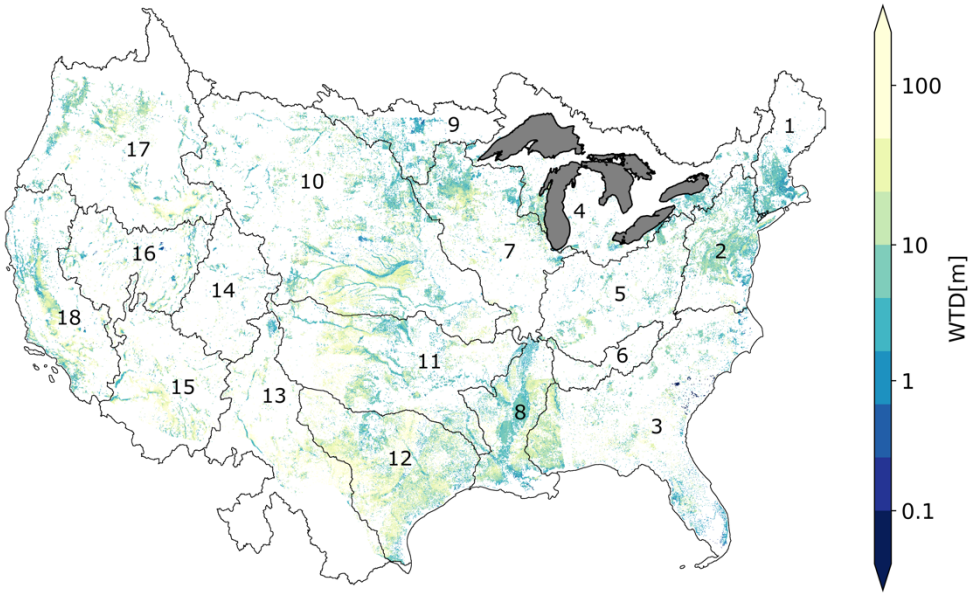
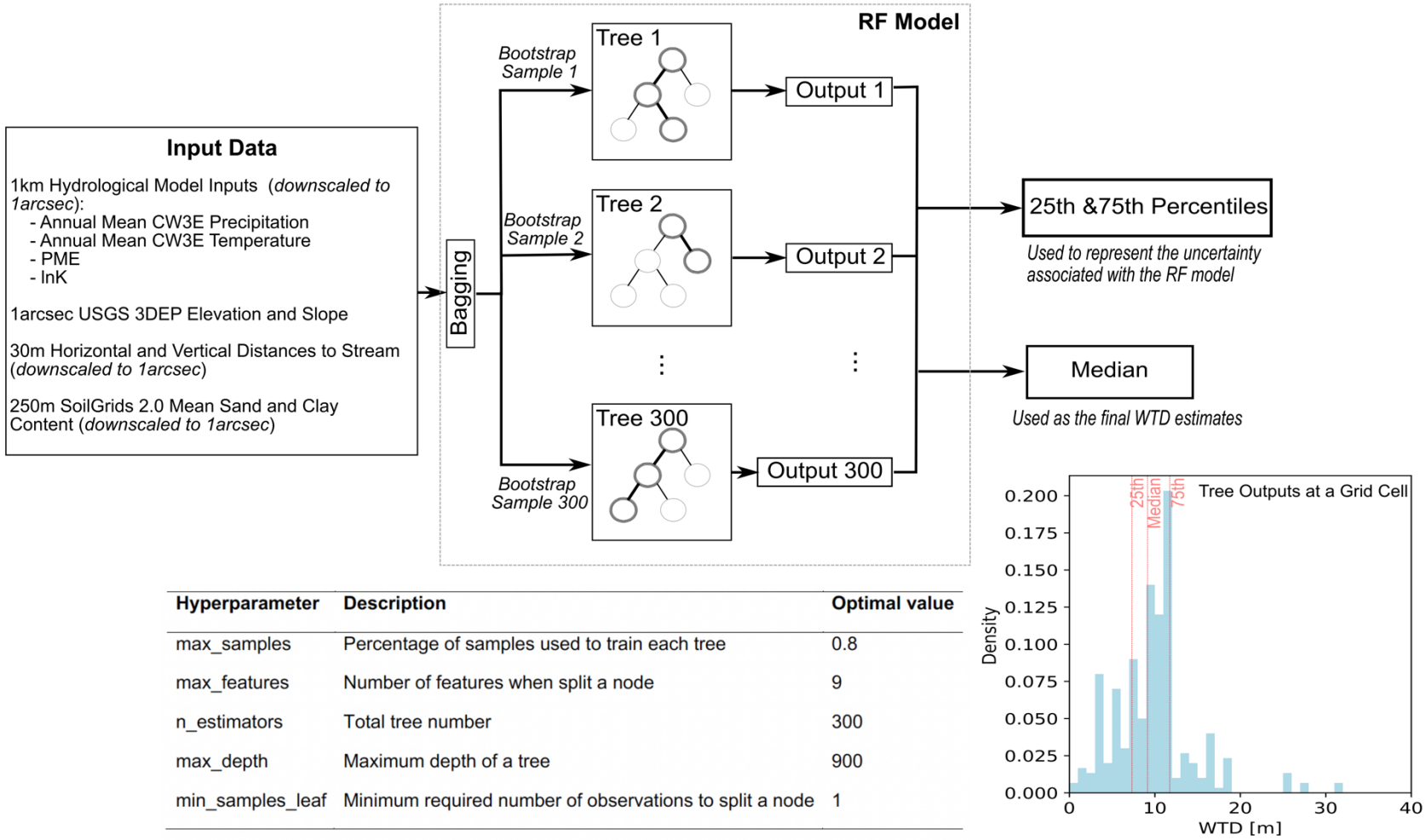


Fig. S1. Gridded long-term mean water table depth (WTD) map at 1 arc-second (~30 m) resolution for the contiguous United States (CONUS). Areas bounded by black polygons in the CONUS show 18 HUC2 basins, which are 1- New England Region, 2- Mid Atlantic Region, 3- South Atlantic-Gulf Region, 4-Great Lakes Region, 5- Ohio Region, 6-Tennessee Region, 7-Upper Mississippi Region, 8- Lower Mississippi Region, 9- Souris-Red-Rainy Region, 10-Missouri Region, 11- Arkansas-White-Red Region, 12- Texas-Gulf Region, 13- Rio Grande Region, 14-Upper Colorado Region, 15- Lower Colorado Region, 16- Great Basin Region, 17-Pacific Northwest Region, and 18- California Region.



Hyperparameter	Description	Optimal value
max_samples	Percentage of samples used to train each tree	0.8
max_features	Number of features when split a node	9
n_estimators	Total tree number	300
max_depth	Maximum depth of a tree	900
min_samples_leaf	Minimum required number of observations to split a node	1

Fig. S2. Schematic diagram of the proposed random forest (RF) model along with its optimal hyperparameter settings.

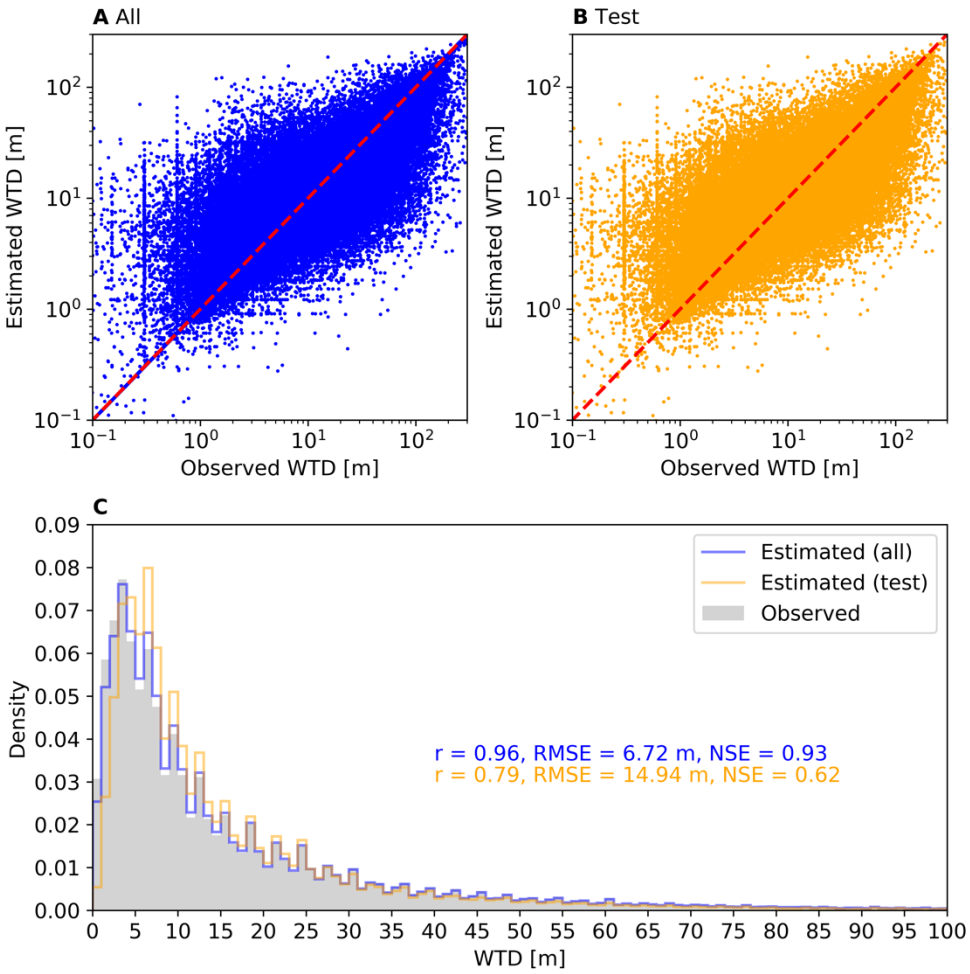


Fig. S3. Random forest model performance over the Contiguous United States (CONUS). Including A, Scatter plot of all data, estimated vs. observed water table depth (WTD), B, scatter plot of 20% test data, estimated vs. observed WTD, and C, histograms of estimated and observed WTD for the CONUS, where the estimated WTD are produced by the RF model at all grid cells with WTD observations (in total 726,828 grid cells) and the grid cells used for testing (in total 145,372 grid cells). The Pearson correlation coefficient (r), root mean square error (RMSE), and Nash-Sutcliffe efficiency (NSE) between estimated and observed WTD are presented here. Note the log scale used for the axes in A and B.

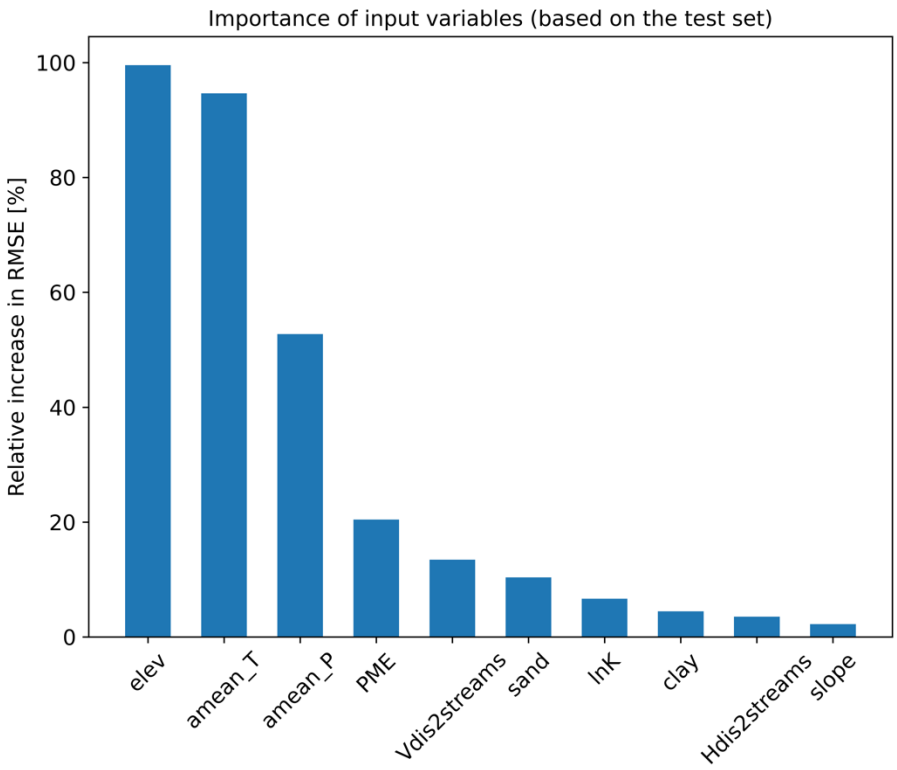


Fig. S4. Input variable importance of the random forest model. The permutation importance is calculated based on the relative increase in the test root mean square error (RMSE).

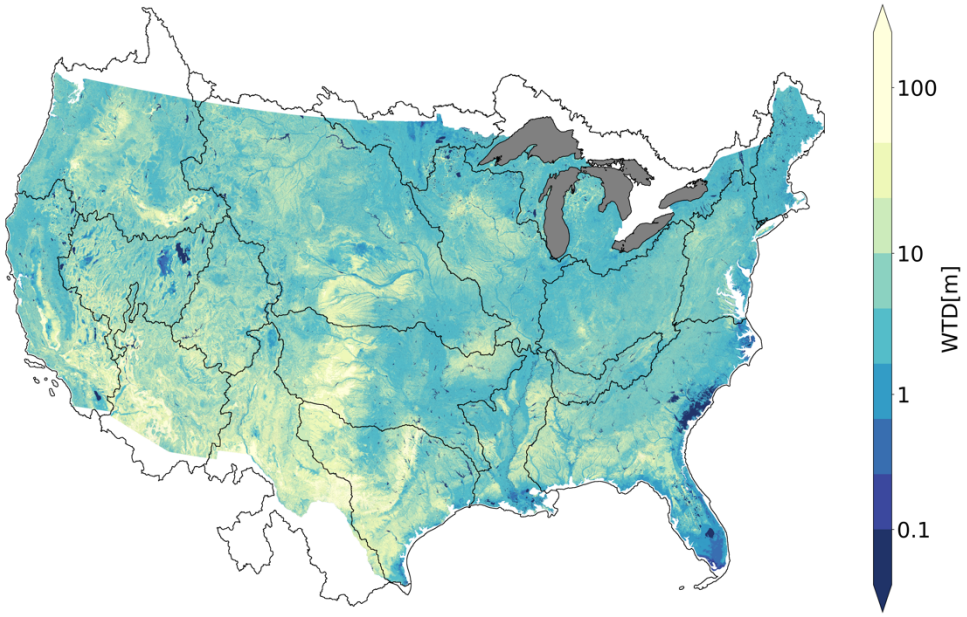


Fig. S5. A map showing the 25th percentile of tree outputs from the random forest (RF) model at the individual grid cell level over the contiguous United States (CONUS). Areas bounded by black polygons in the CONUS show 18 HUC2 basins.

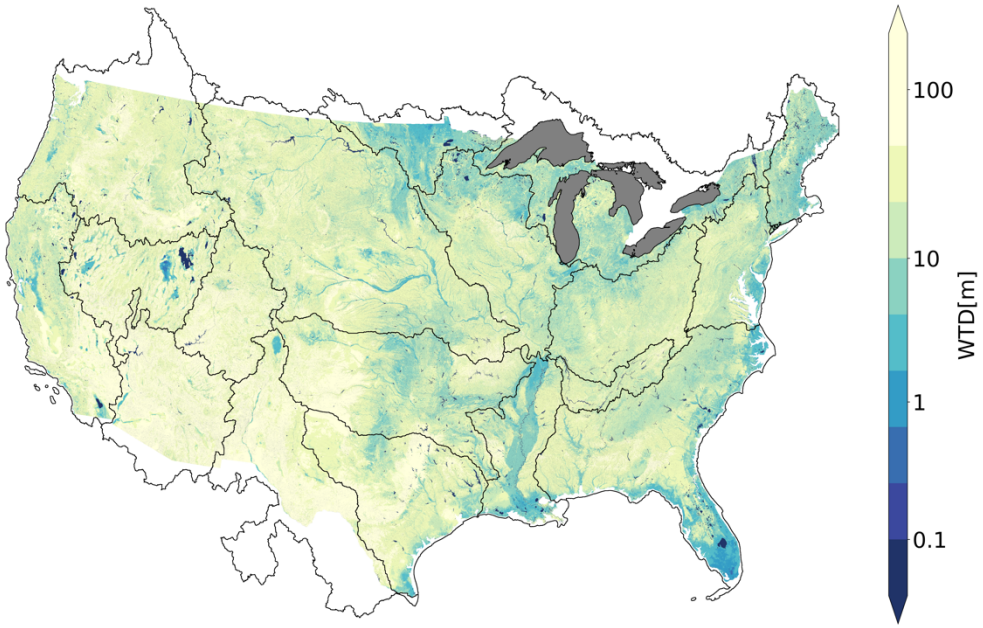


Fig. S6. A map showing the 75th percentile of tree outputs from the random forest (RF) model at the individual grid cell level over the contiguous United States (CONUS). Areas bounded by black polygons in the CONUS show 18 HUC2 basins.

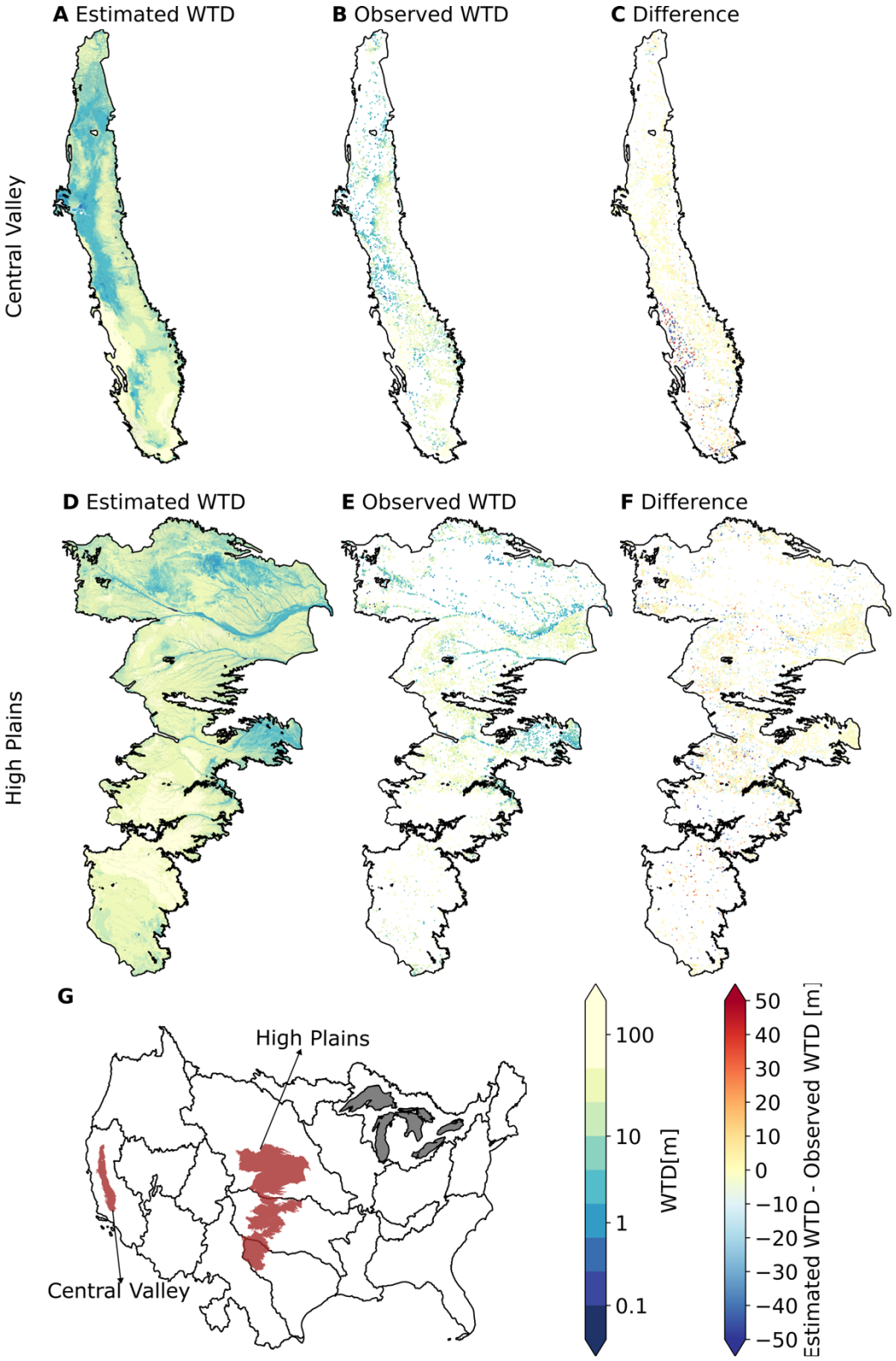


Fig. S7. Random forest model testing performance in the California Central Valley and High Plains aquifers, which are hot spots of groundwater depletion. Including A and D, maps of estimated water table depth (WTD), B and E, maps of observed WTD used for testing, C

and F, maps of difference between estimated and observed WTD during testing, and G, a map indicating the locations of the two aquifers in the Contiguous United States.

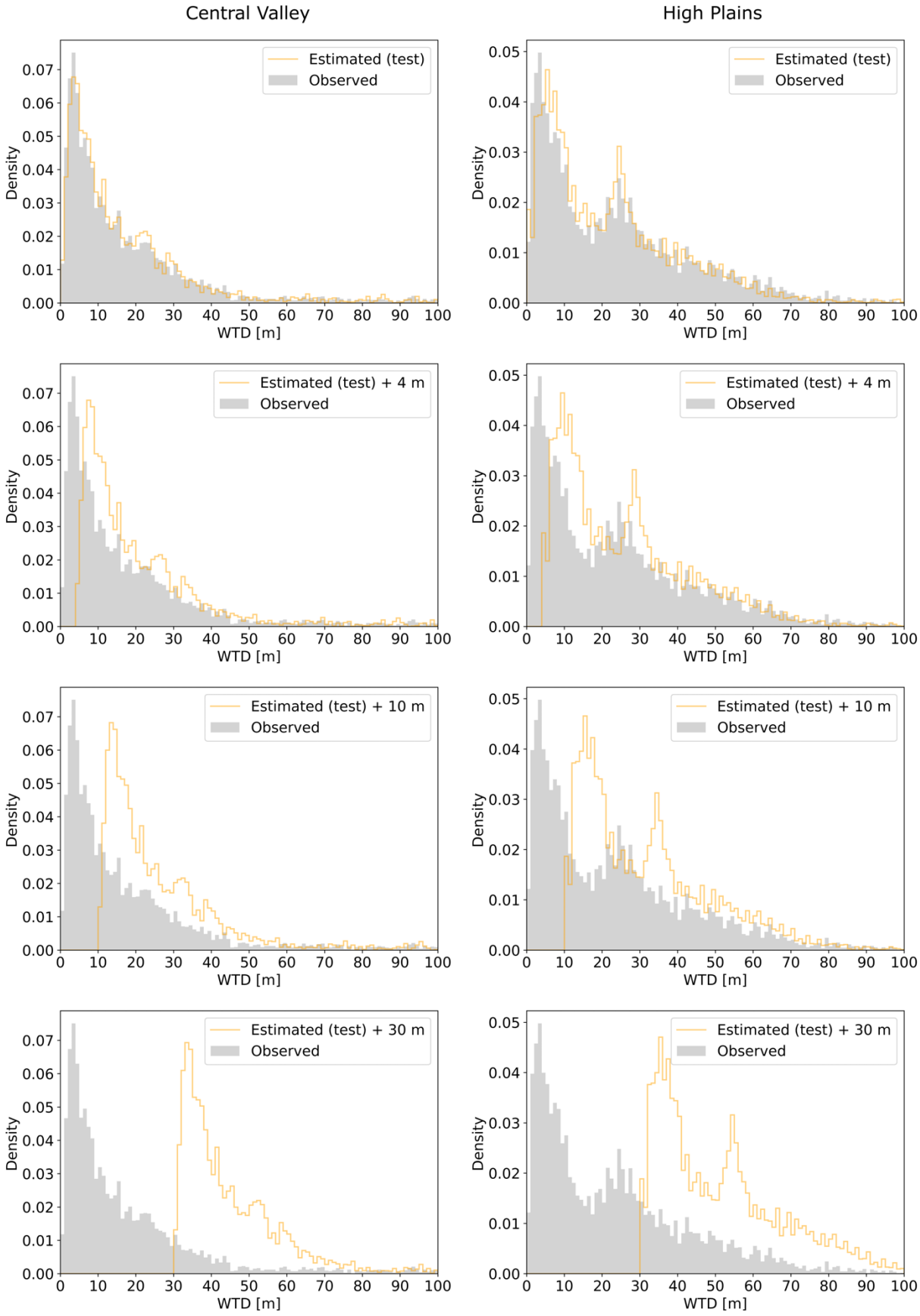

Central Valley
High Plains
Estimated (test)
Estimated (test) + 4 m
Estimated (test) + 10 m
Estimated (test) + 30 m
Observed
Density
WTD [m]

Fig. S8. Histograms showing the comparison between the estimated and observed water table depth (WTD) in the California Central Valley and High Plains aquifers after manually adding pumping (4 m, 10 m, and 30 m).

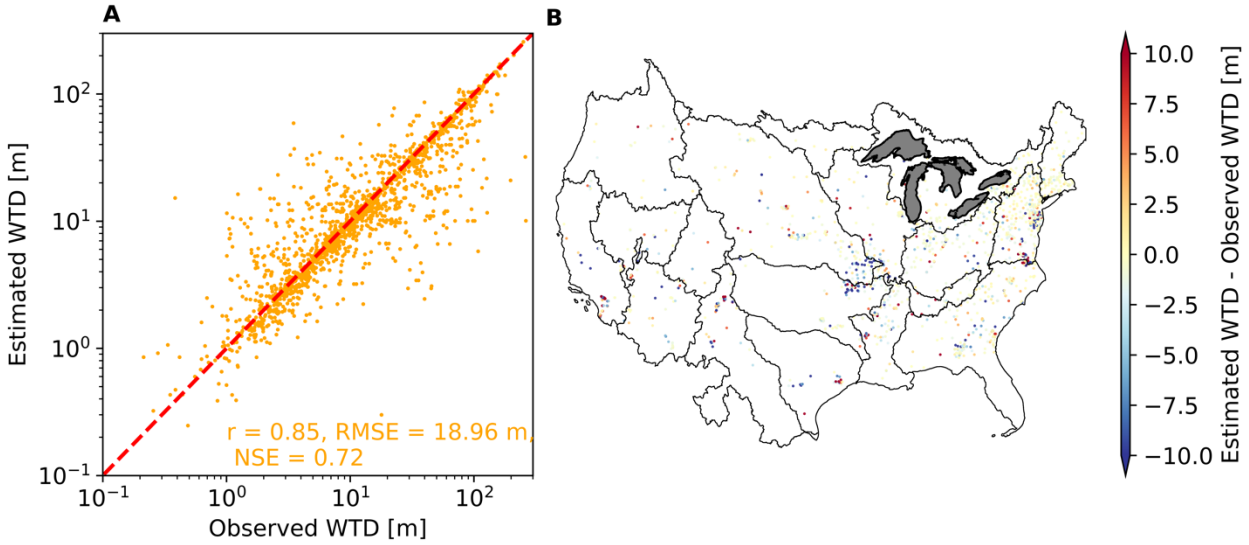


Fig. S9. Random forest model performance against the most recent water table depth (WTD) observations. Including A, Scatter plot of the estimated WTD vs. the most recent observed WTD from 1877 USGS groundwater monitoring wells that have at least one record from June 1 to August 31, 2024. B, Map showing the difference between the estimated WTD and the most recent observed WTD.

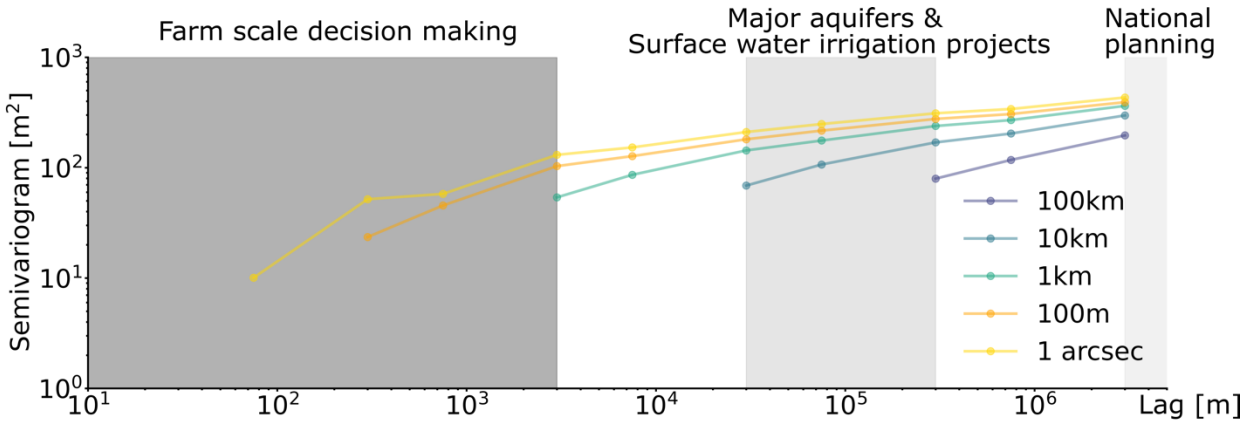


Fig. S10. Empirical variograms for the water table depth (WTD) estimates at 1 arc-second, 100 m, 1 km, 10 km, and 100 km resolutions. The WTD estimates for the latter four resolutions are upscaled from the original 1 arc-second data using the mean function.

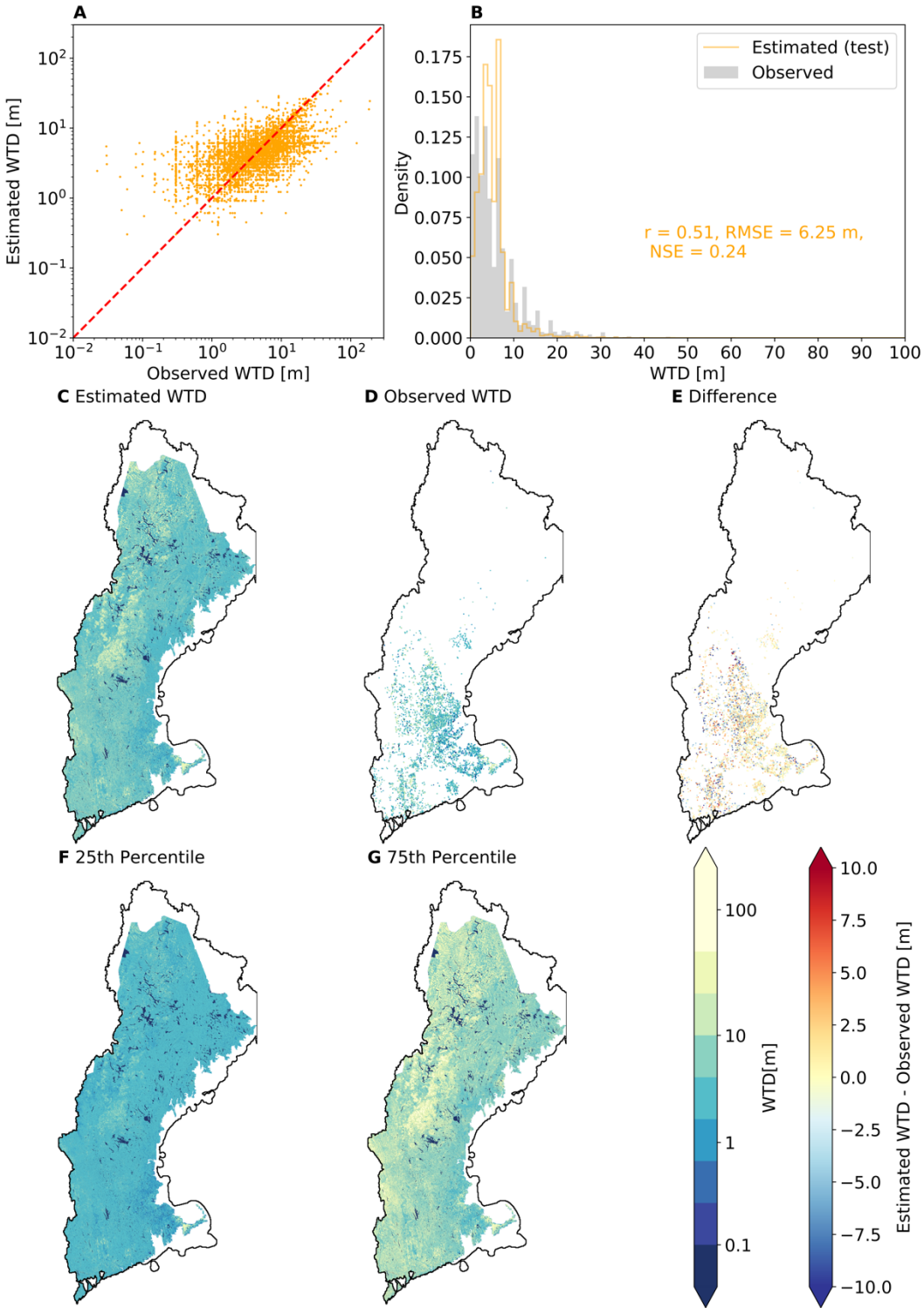


Fig. S11. Random forest (RF) model testing performance in the New England Region. Including A, a scatter plot of estimated vs. observed water table depth (WTD) during testing, B, a histogram of estimated and observed WTD during testing, C, a map of estimated WTD, D, a map of observed WTD used for testing, E, a map of the difference between estimated and observed WTD during testing, F, a map of the 25th percentile of tree outputs from the RF model

at the individual grid cell level, and G, a map of the 75th percentile of tree outputs from the RF model at the individual grid cell level.

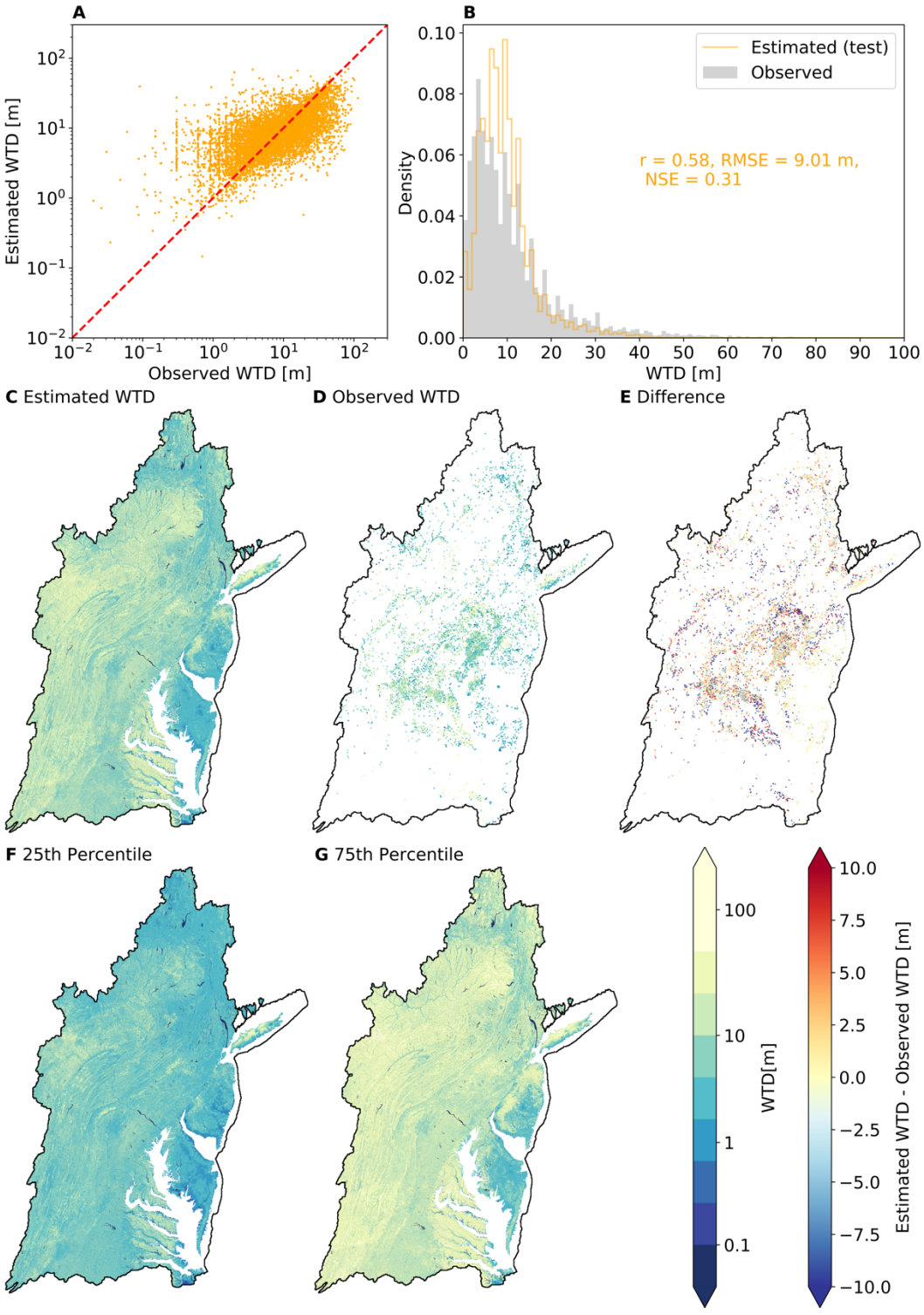


Fig. S12. Random forest model testing performance in the Mid Atlantic Region.

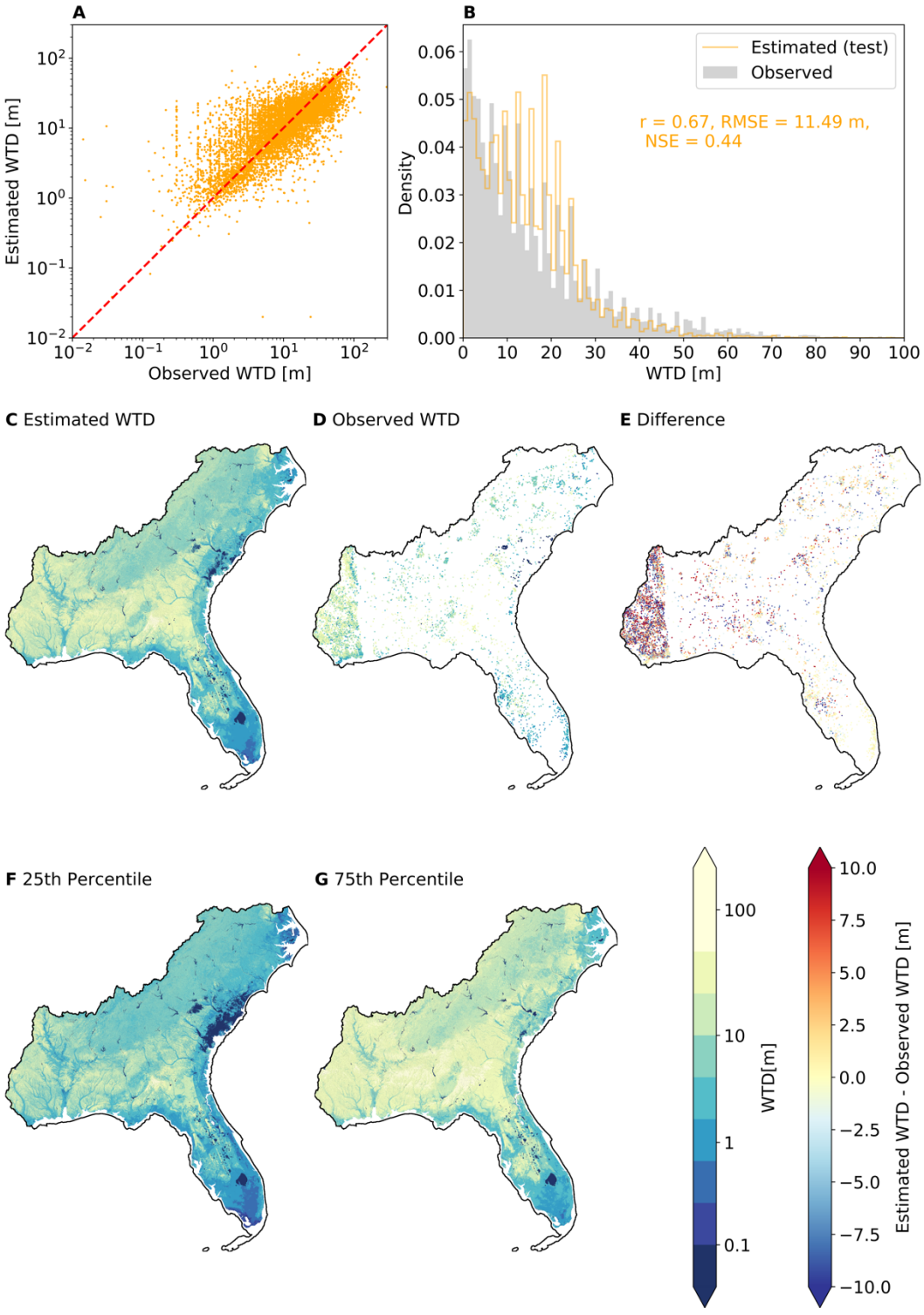


Fig. S13. Random forest model testing performance in the South Atlantic-Gulf Region.

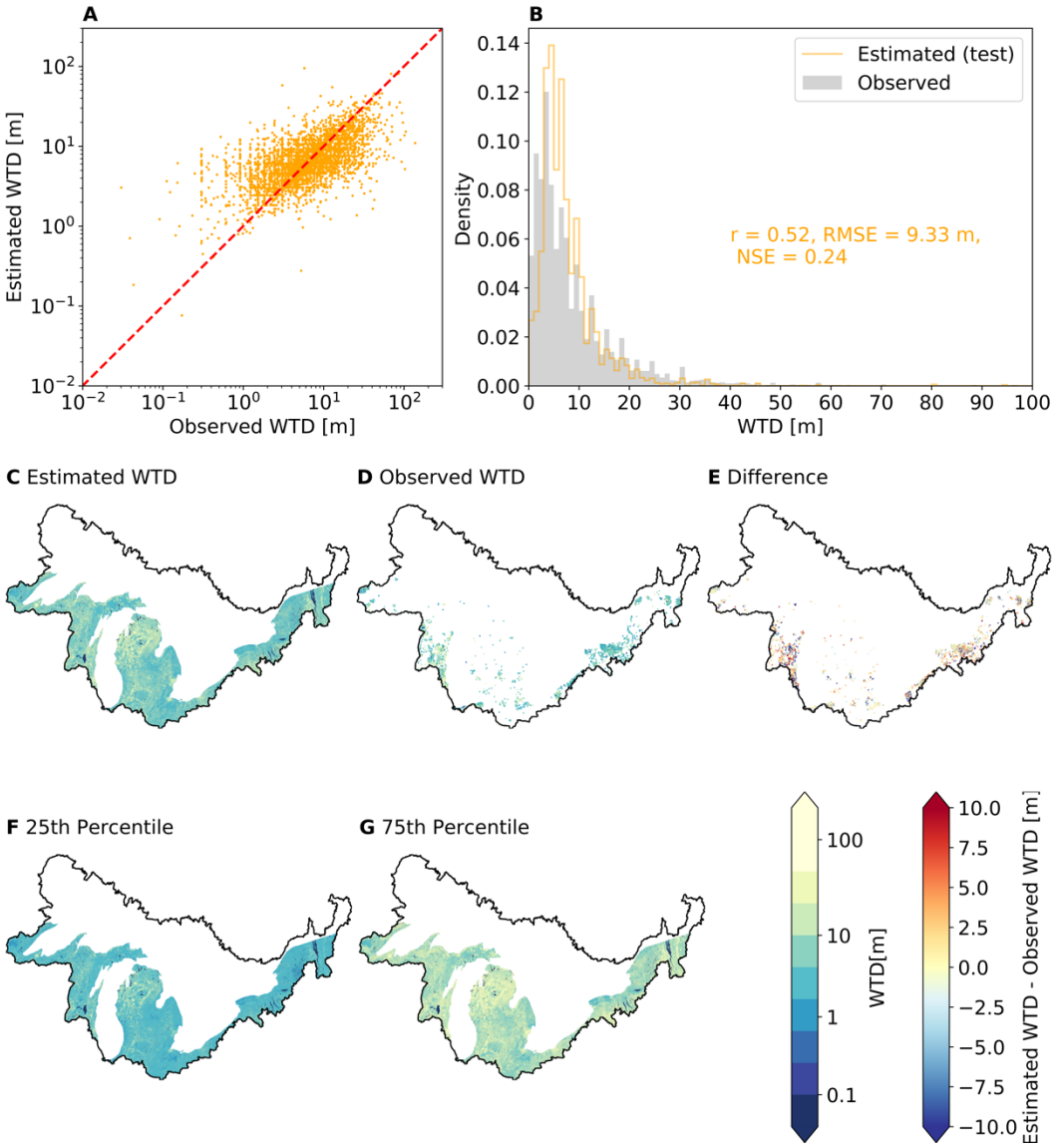


Fig. S14. Random forest model testing performance in the Great Lakes Region.

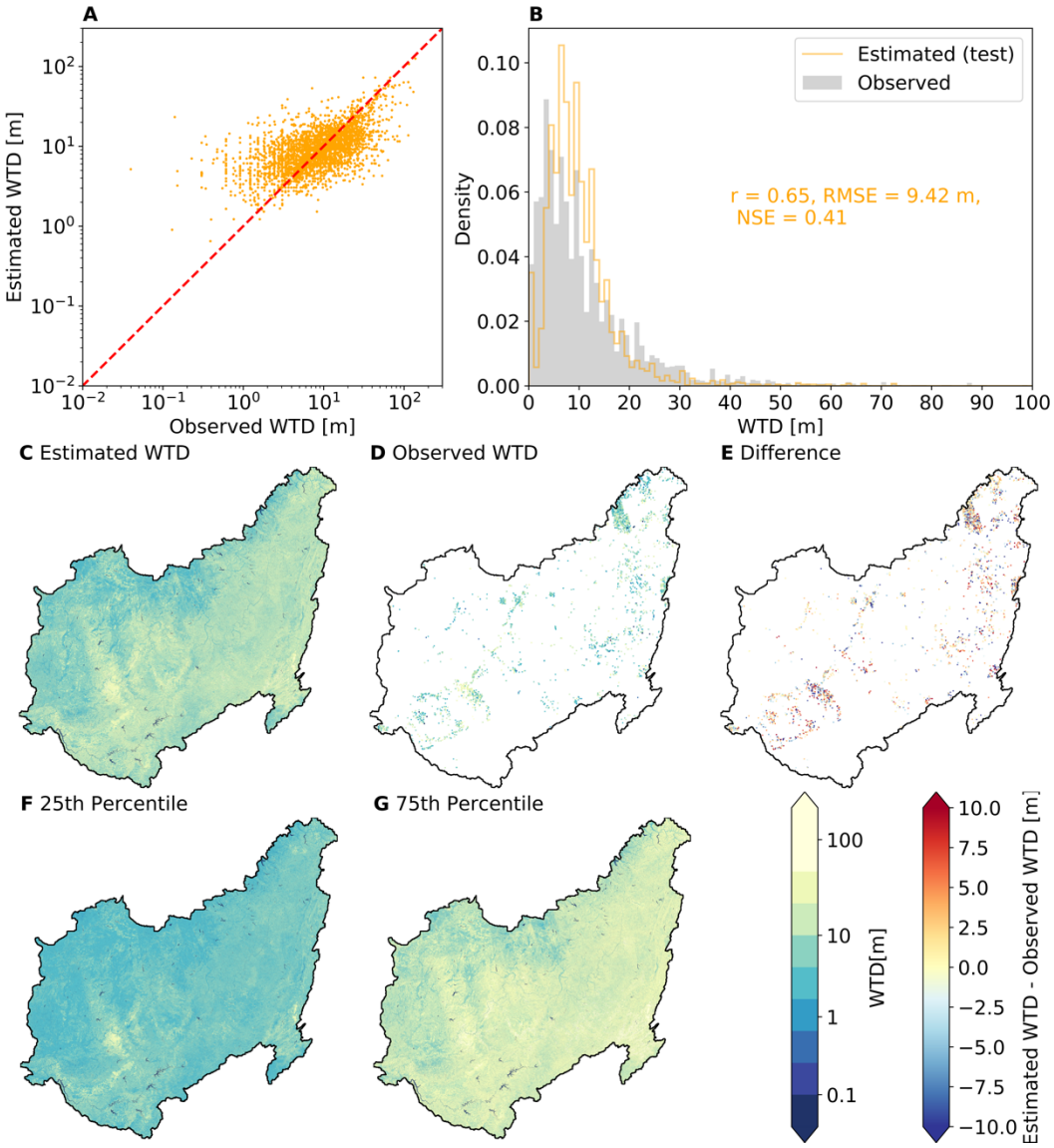


Fig. S15. Random forest model testing performance in the Ohio Region.

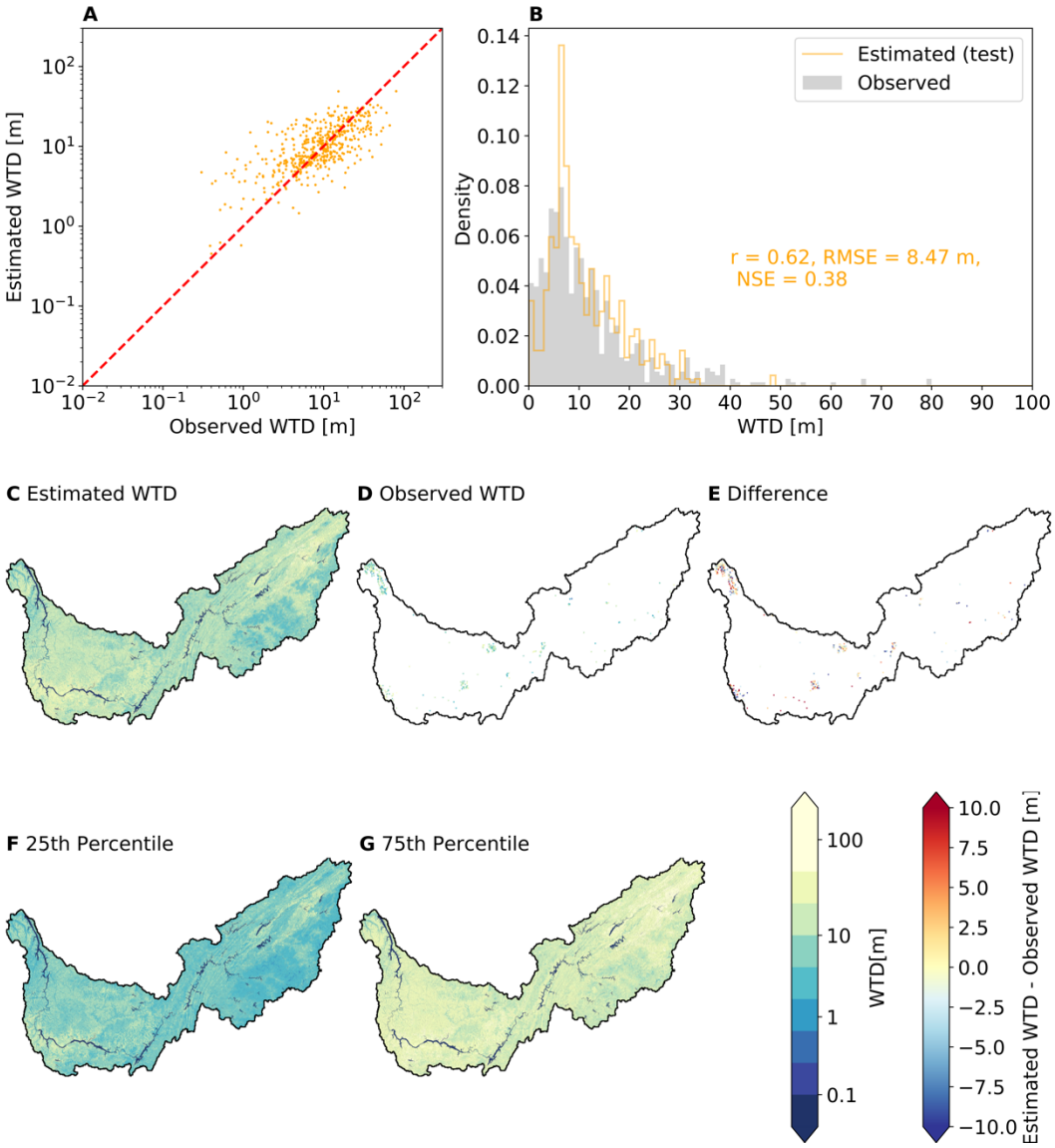


Fig. S16. Random forest model testing performance in the Tennessee Region.

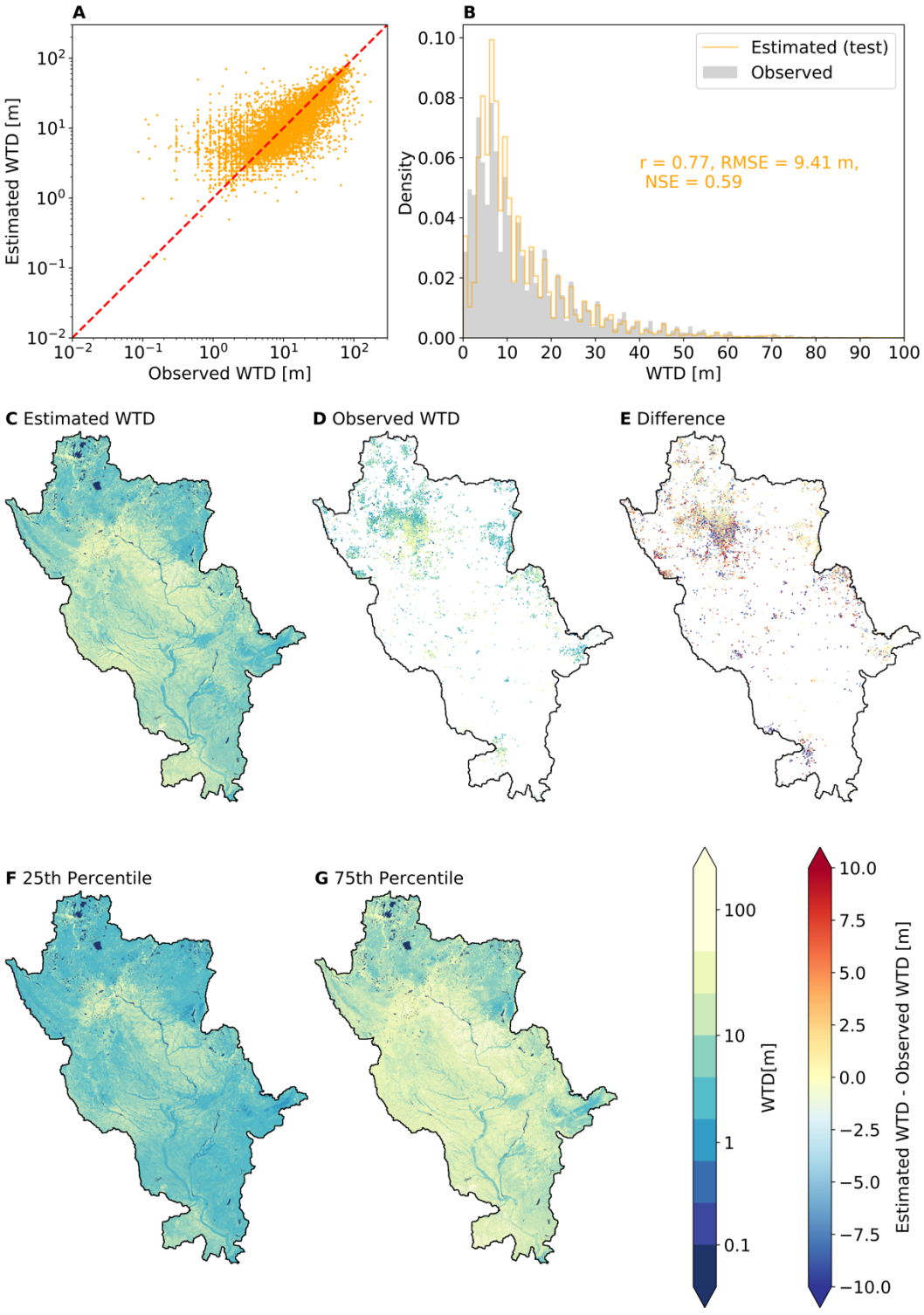


Fig. S17. Random forest model testing performance in the Upper Mississippi Region.

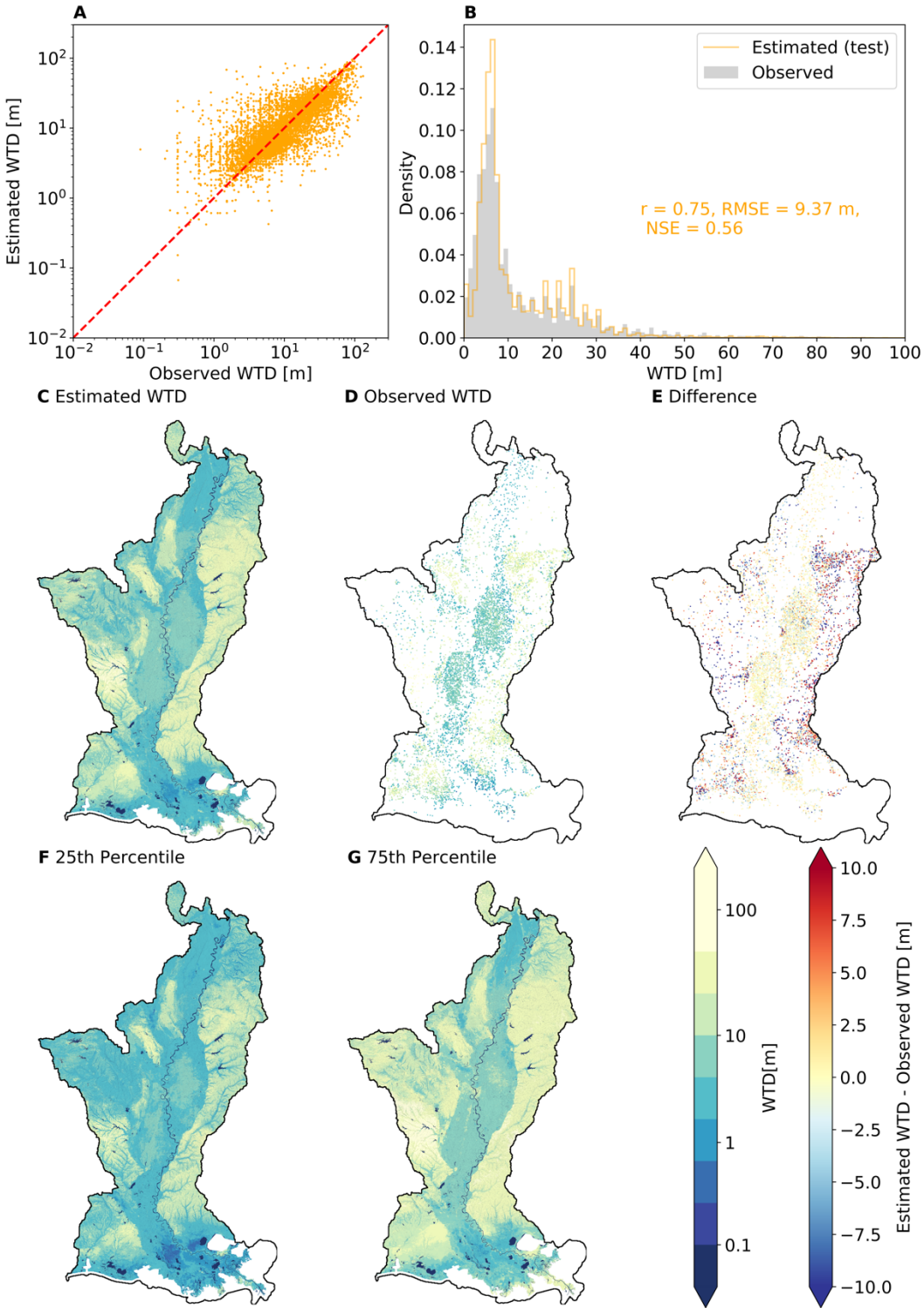


Fig. S18. Random forest model testing performance in the Lower Mississippi Region.

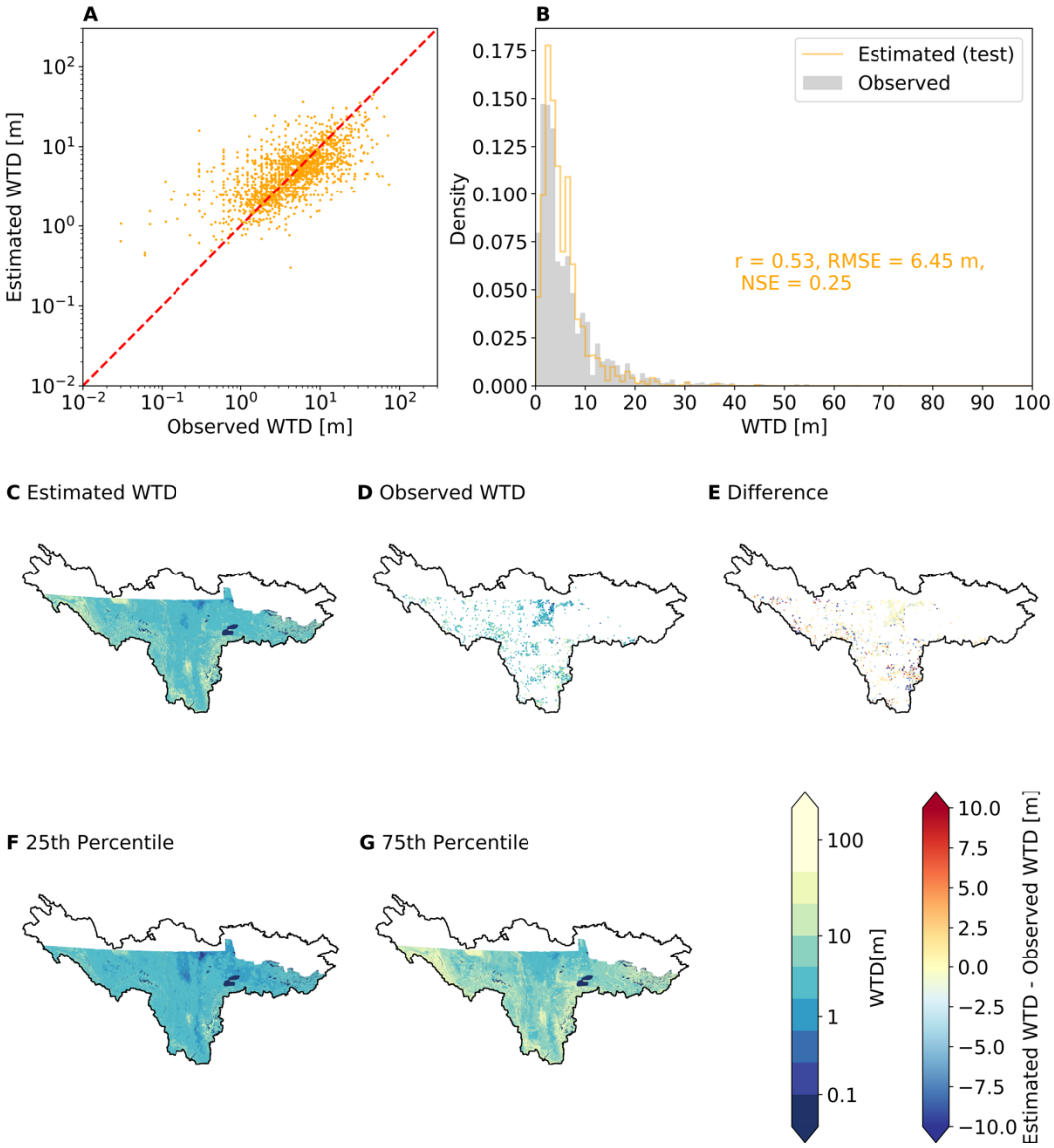


Fig. S19. Random forest model testing performance in the Souris-Red-Rainy Region.

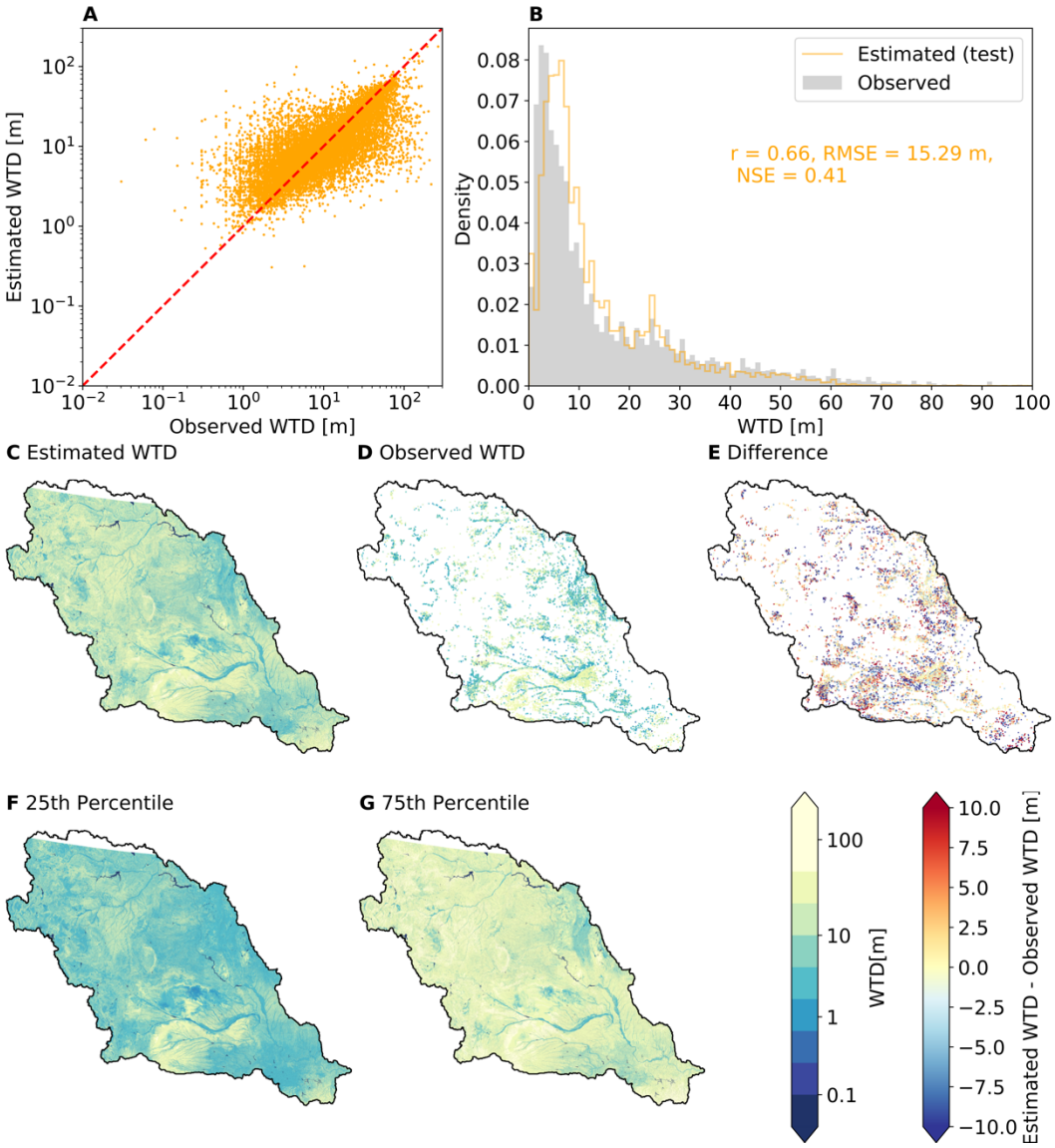


Fig. S20. Random forest model testing performance in the Missouri Region.

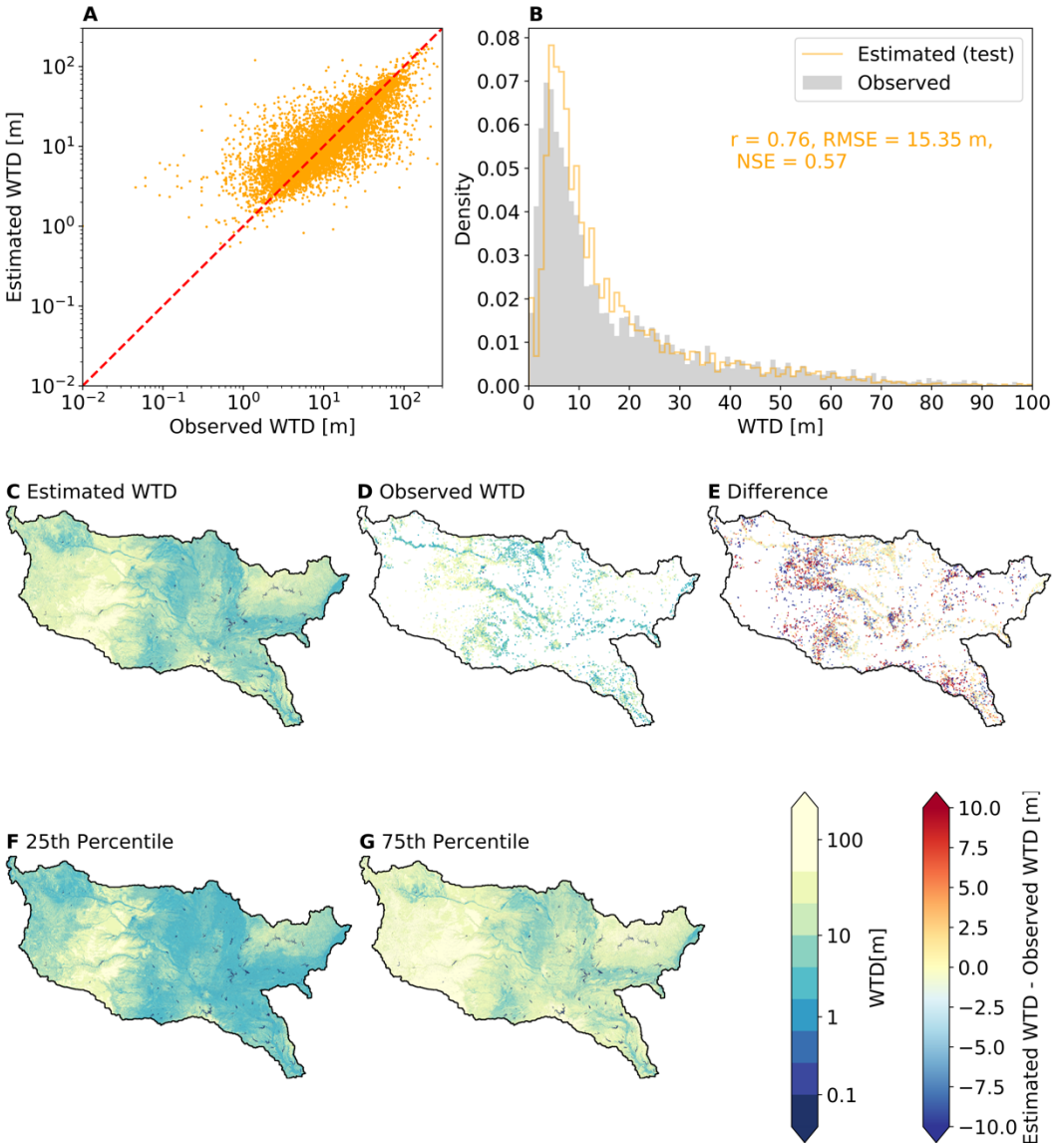


Fig. S21. Random forest model testing performance in the Arkansas-White-Red Region.

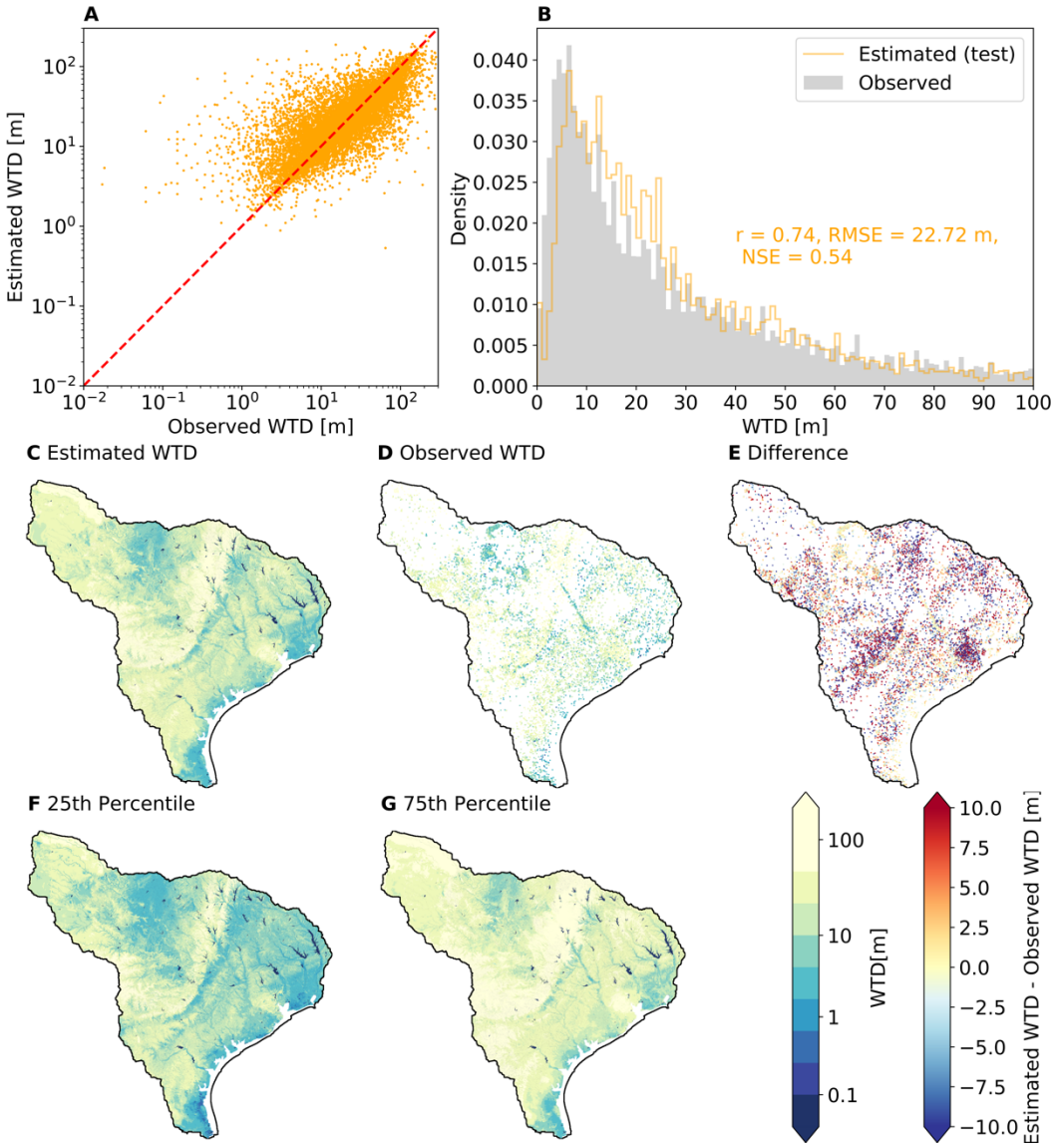


Fig. S22. Random forest model testing performance in the Texas-Gulf Region.

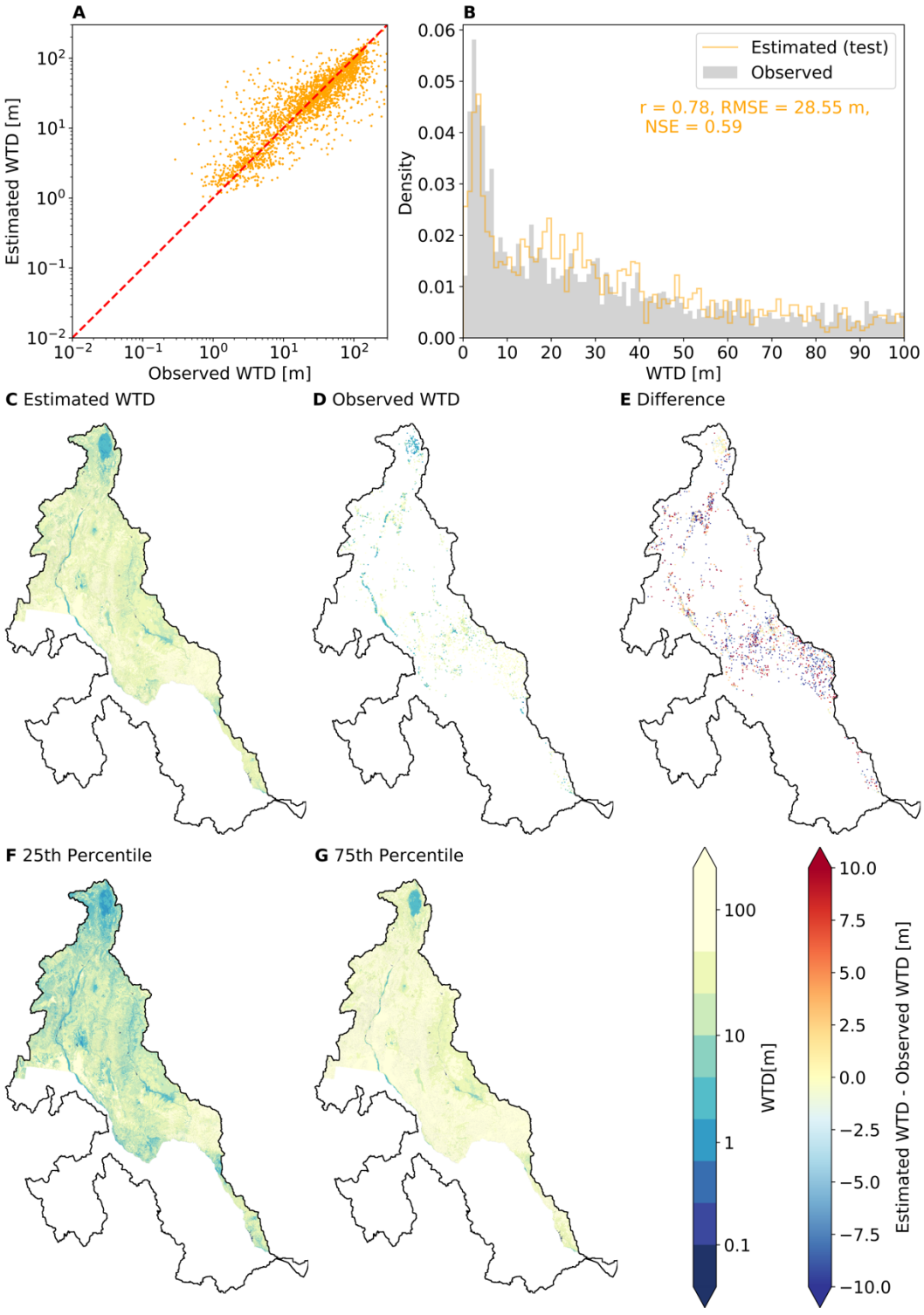


Fig. S23. Random forest model testing performance in the Rio Grande Region.

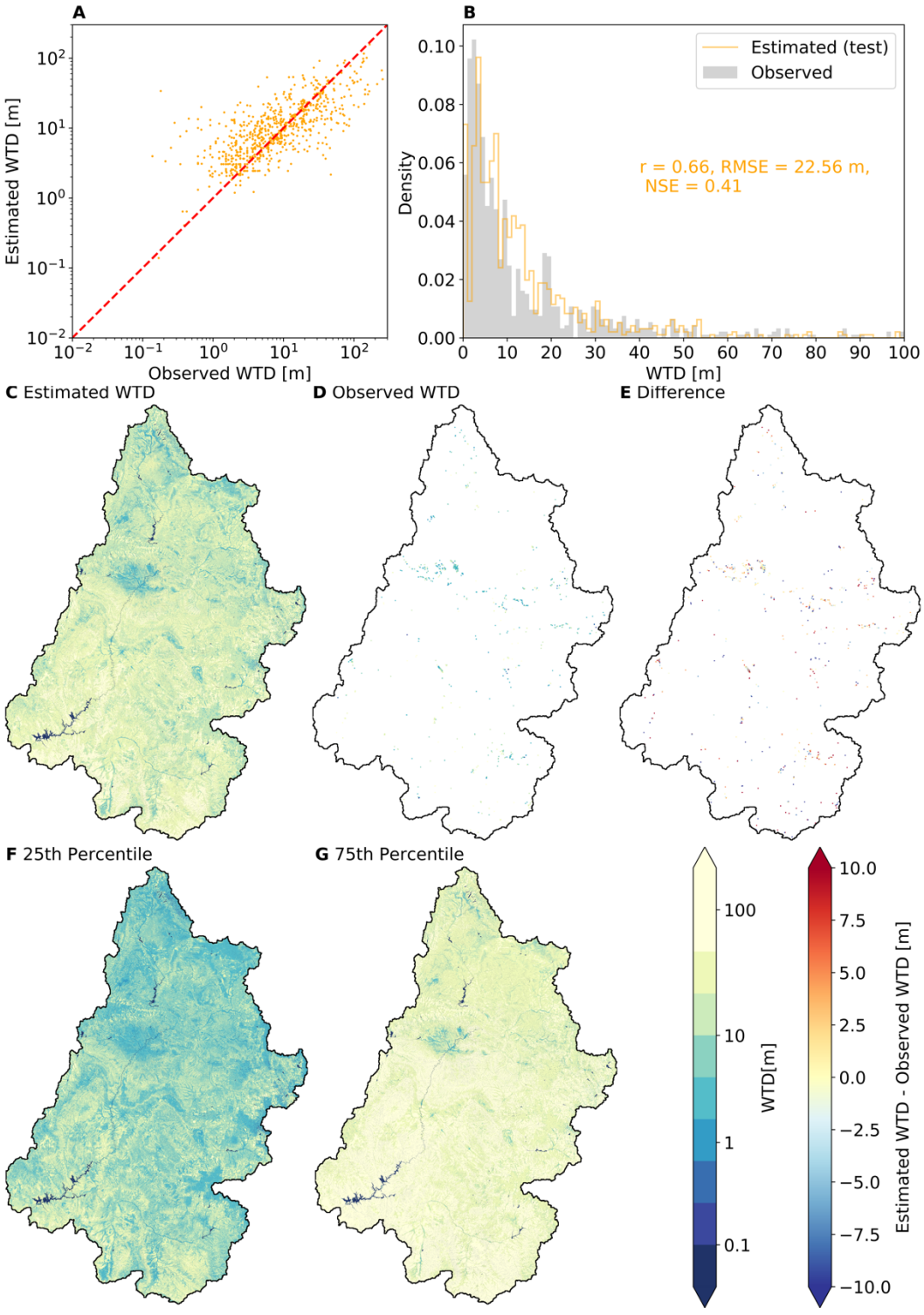


Fig. S24. Random forest model testing performance in the Upper Colorado Region.

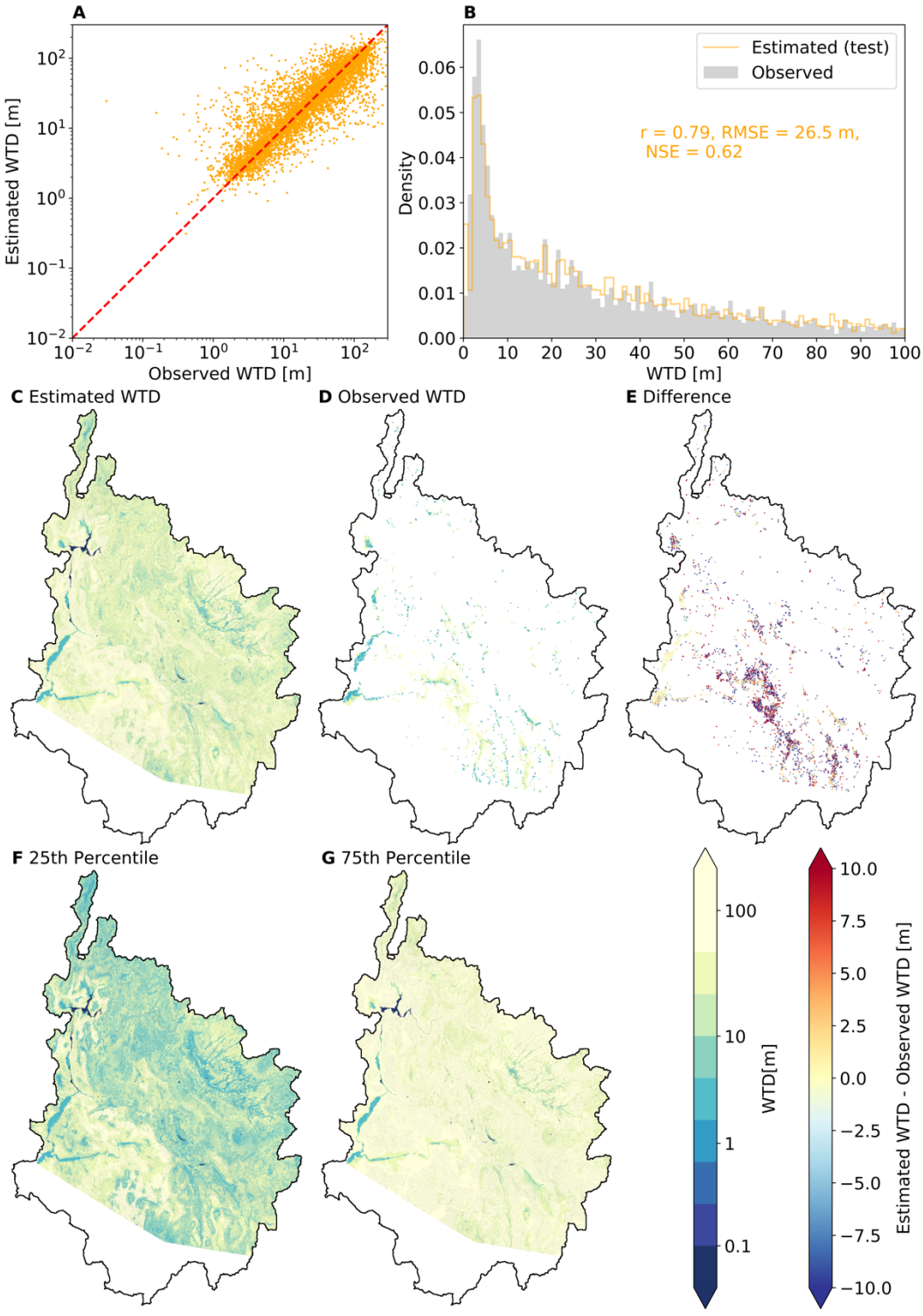


Fig. S25. Random forest model testing performance in the Lower Colorado Region.

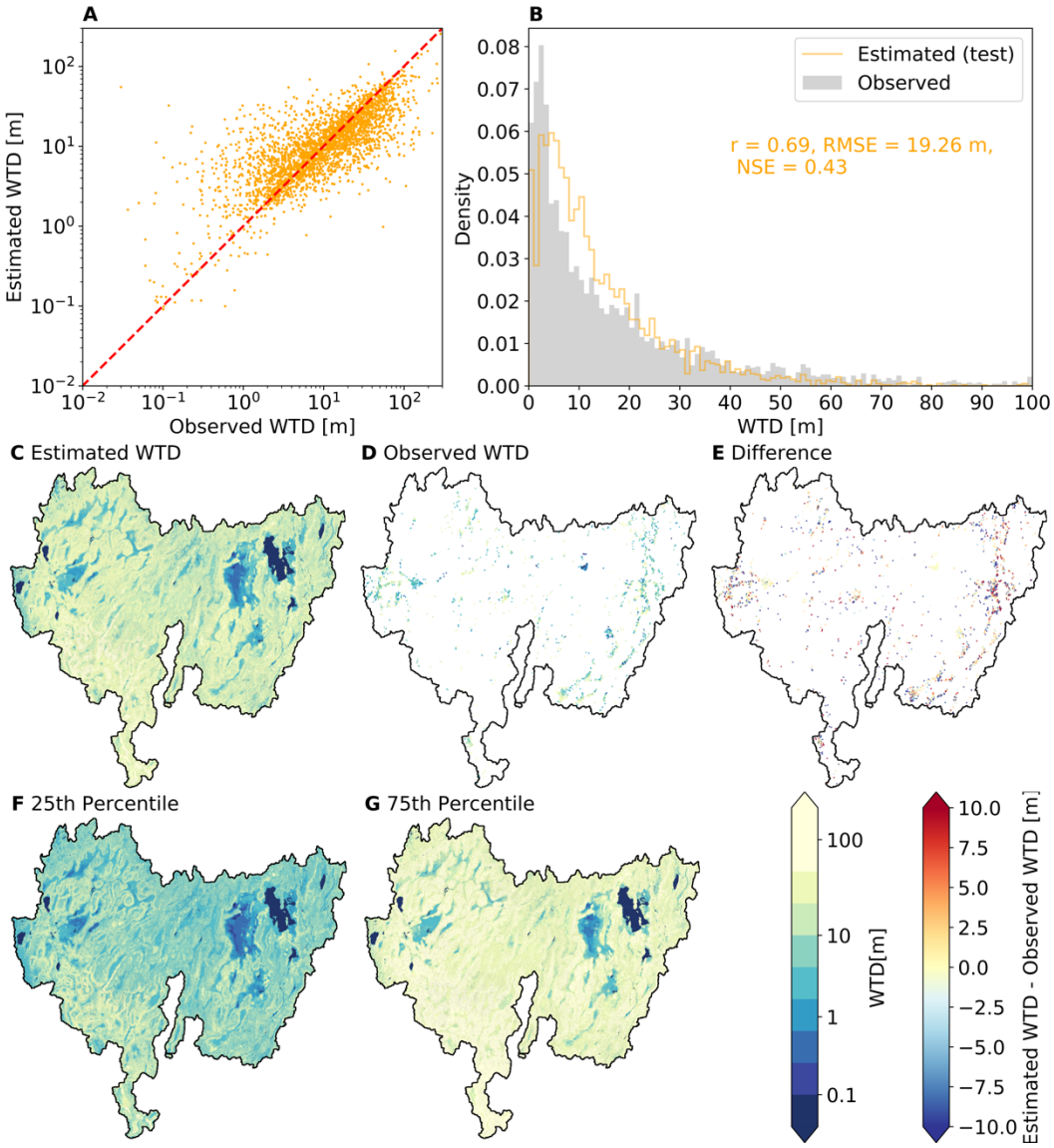


Fig. S26. Random forest model testing performance in the Great Basin Region.

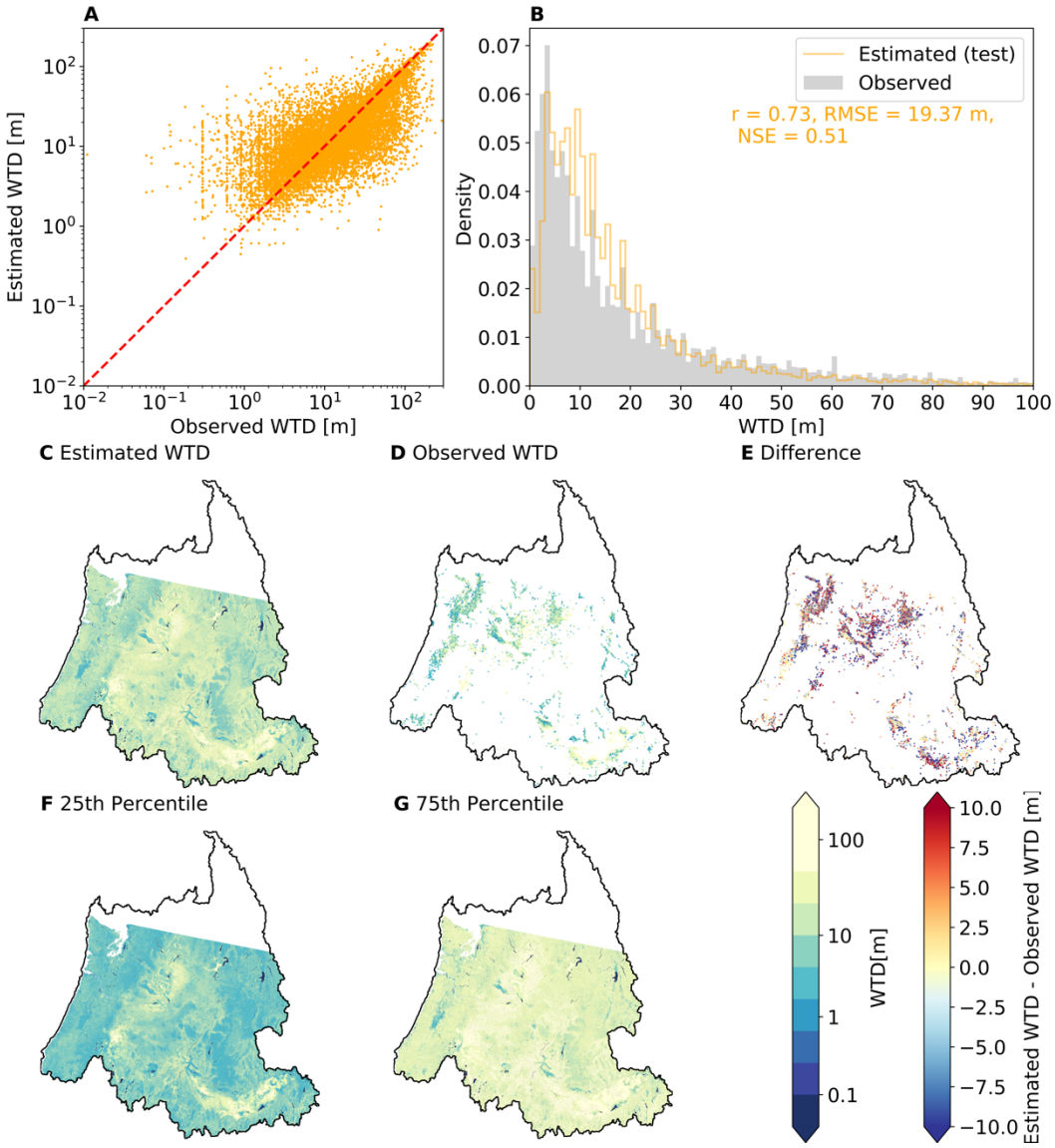


Fig. S27. Random forest model testing performance in the Pacific Northwest Region.

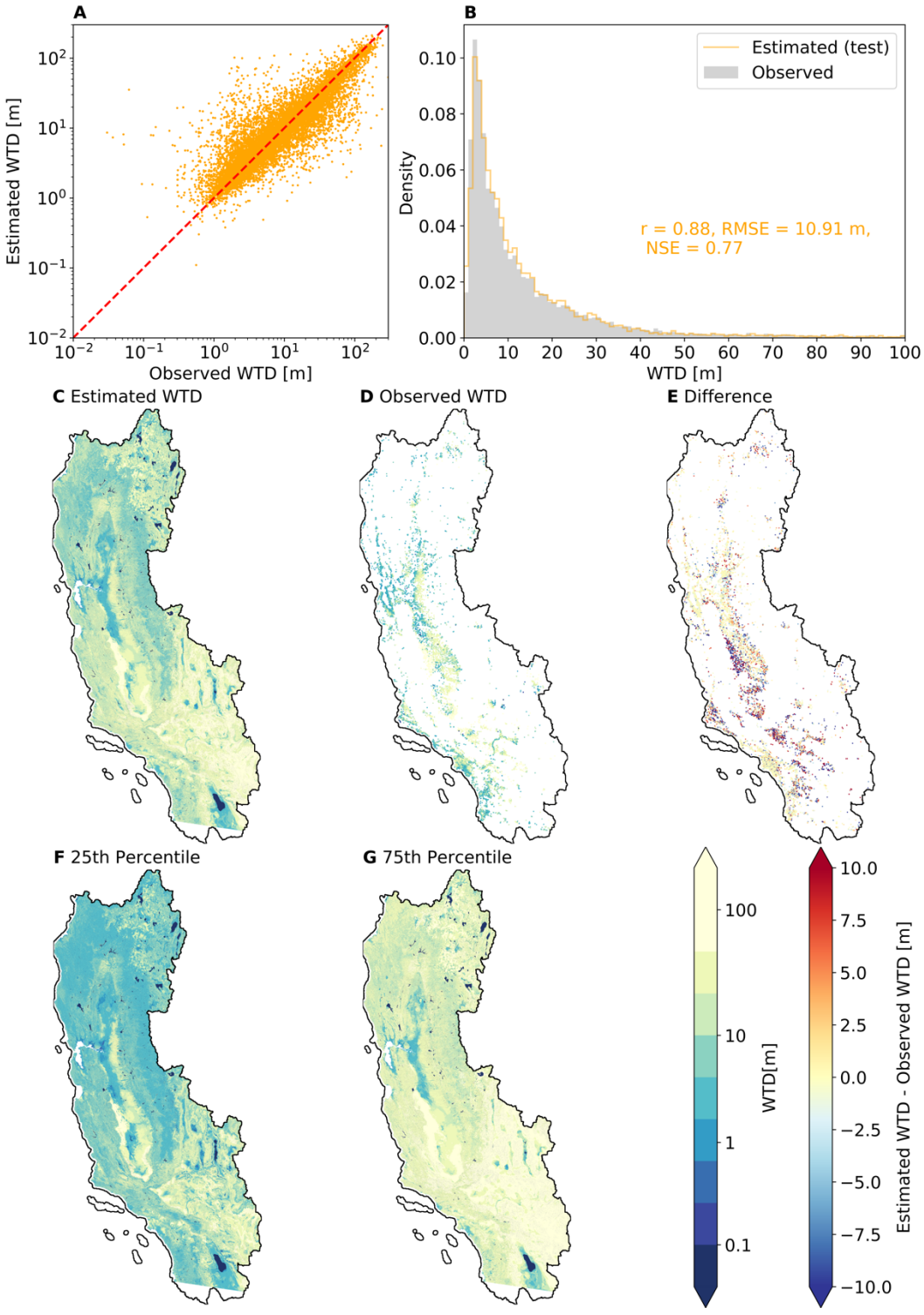


Fig. S28. Random forest model testing performance in the California Region.

Table S1. Porosity for each hydrostratigraphic category.

Hydrostatigraphy*	Porosity [-]	References
Bedrock 1	0.05	Worthington[76]
Bedrock 2	0.1	Worthington[76]
f.g. sil. sedimentary	0.12	Gleeson et al.[77]
sil. sedimentary	0.3	Worthington[76]; Worthington et al.[78]
crystalline	0.01	Gleeson et al.[77]
f.g. unconsolidated	0.15	Worthington[76]; Gleeson et al.[77]
unconsolidated	0.22	Gleeson et al.[77]
c.g. sil. sedimentary	0.27	Gleeson et al.[77]
carbonate	0.06	Worthington[76]; Gleeson et al.[77]
c.g. unconsolidated	0.3	Gleeson et al.[77]

**sil. sedimentary is siliciclastic sedimentary; c.g. and f.g. are coarse grained and fine grained, respectively.*

Table S2. Areas with water table depth (WTD) data and ratios of estimated groundwater storage to area for different spatial resolutions.

		Spatial resolution				
		100 km	**10 km**	**1 km**	**100 m**	**1 arc-second**
Areas with WTD data [10^6 km^2]		6.13	7.17	7.33	7.35	7.35
Ratio of estimated groundwater storage to area [m]	**Lower**	38.99	39.36	39.53	39.67	39.71
	Median	41.16	41.41	41.54	41.67	41.71
	Upper	42.68	42.85	42.95	43.07	43.1

References

49. CW3E. Center for Western Weather and Water Extremes. https://cw3e.ucsd.edu/ (2024).
50. Yang, C., Tijerina-Kreuzer, D. T., Tran, H. V., Condon, L. E. & Maxwell, R. M. A high-resolution, 3D groundwater-surface water simulation of the contiguous US: Advances in the integrated ParFlow CONUS 2.0 modeling platform. J Hydrol (Amst) **626**, 130294 (2023).
51. Swilley, J. S. et al. Continental Scale Hydrostratigraphy: Comparing Geologically Informed Data Products to Analytical Solutions. Groundwater (2023) doi:10.1111/gwat.13354.
52. USGS. The National Map—New Data Delivery Homepage, Advanced Viewer, Lidar Visualization. (2019) doi:10.3133/fs20193032.
53. QGIS.org. QGIS Geographic Information System. Open Source Geospatial Foundation Project. (2023).
54. Belitz, K., Moore, R. B., Arnold, T. L., Sharpe, J. B. & Starn, J. J. Multiorder Hydrologic Position in the Conterminous United States: A Set of Metrics in Support of Groundwater Mapping at Regional and National Scales. Water Resour Res **55**, 11188–11207 (2019).
55. Brinkerhoff, C. B., Gleason, C. J., Kotchen, M. J., Kysar, D. A. & Raymond, P. A. Ephemeral stream water contributions to United States drainage networks. Science (1979) **384**, 1476–1482 (2024).
56. Poggio, L. et al. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. SOIL **7**, 217–240 (2021).
57. Dewitz, J. National Land Cover Database (NLCD) 2021 Products. (2023) doi:10.5066/P9JZ7AO3.
58. Müller, A. C. & Guido, S. Ensembles of Decision Trees. in Introduction to machine learning with Python: A GUIDE FOR DATA SCIENTISTS 85–94 (O'Reilly Media, Inc., Sebastopol, 2017).
59. Breiman, L. Random forests. Mach Learn **45**, 5–32 (2001).
60. Meinshausen, N. Quantile Regression Forests. Journal of Machine Learning Resear **7**, 983–999 (2006).
61. Koch, J. et al. High Resolution Water Table Modeling of the Shallow Groundwater Using a Knowledge-Guided Gradient Boosting Decision Tree Model. Frontiers in Water **3**, (2021).
62. Wunsch, A., Liesch, T. & Broda, S. Deep learning shows declining groundwater levels in Germany until 2100 due to climate change. Nat Commun **13**, 1221 (2022).
63. Vu, M. T., Jardani, A., Massei, N. & Fournier, M. Reconstruction of missing groundwater level data by using Long Short-Term Memory (LSTM) deep neural network. J Hydrol (Amst) **597**, 125776 (2021).
64. Tran, H. et al. Development of a Deep Learning Emulator for a Distributed Groundwater–Surface Water Model: ParFlow-ML. Water (Basel) **13**, 3393 (2021).
65. Ma, Y., Montzka, C., Naz, B. S. & Kollet, S. Advancing AI-based pan-European groundwater monitoring. Environmental Research Letters **17**, 114037 (2022).
66. Raissi, M., Perdikaris, P. & Karniadakis, G. E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J Comput Phys 378, 686–707 (2019).
67. Pekel, J.-F., Cottam, A., Gorelick, N. & Belward, A. S. High-resolution mapping of global surface water and its long-term changes. Nature **540**, 418–422 (2016).

68. Zell, W. O. & Sanford, W. E. Calibrated Simulation of the Long-Term Average Surficial Groundwater System and Derived Spatial Distributions of its Characteristics for the Contiguous United States. Water Resour Res **56**, (2020).
69. Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011).
70. Maxwell, R. M. & Condon, L. E. Connections between groundwater flow and transpiration partitioning. Science (1979) **353**, 377–380 (2016).
71. Konikow, L. F. Long-Term Groundwater Depletion in the United States. Groundwater **53**, 2–9 (2015).
72. Scanlon, B. R. et al. Groundwater depletion and sustainability of irrigation in the US High Plains and Central Valley. Proceedings of the National Academy of Sciences **109**, 9320–9325 (2012).
73. Matheron, G. Principles of geostatistics. Economic Geology **58**, 1246–1266 (1963).
74. Ford, D. The Empirical Variogram. http://faculty.washington.edu/edford/Variogram.pdf.
75. Müller, S. & Schüler, L. GeoStat-Framework/GSTools: v1.5.0 'Nifty Neon' (v1.5.0). Preprint at https://doi.org/https://doi.org/10.5281/zenodo.8044720 (2023).
76. Worthington, S. R. H. Estimating Effective Porosity in Bedrock Aquifers. Groundwater **60**, 169–179 (2022).
77. Gleeson, T., Moosdorf, N., Hartmann, J. & van Beek, L. P. H. A glimpse beneath earth's surface: GLobal HYdrogeology MaPS (GLHYMPS) of permeability and porosity. Geophys Res Lett 41, 3891–3898 (2014).78. Worthington, S. R. H., Foley, A. E. & Soley, R. W. N. Transient characteristics of effective porosity and specific yield in bedrock aquifers. J Hydrol (Amst) **578**, 124129 (2019).