Supplement of Earth Syst. Sci. Data, 15, 5597–5615, 2023
https://doi.org/10.5194/essd-15-5597-2023-supplement

Supplement of

GTWS-MLrec: global terrestrial water storage reconstruction by machine learning from 1940 to present

Jiabo Yin et al.

Correspondence to: Jiabo Yin (jboyn@whu.edu.cn)

Text S1: Skill metrics computation

Eight skill metrics are used to evaluate model performance: Pearson's Correlation Coefficient (PCC), Nash-Sutcliffe efficiency coefficient (NSE); Kling-Gupta Efficiency coefficient (KGE), Coefficient of Determination (R^2), Root Mean square error (RMSE); normalized Root Mean Square Error (nRMSE), Mean Absolute Percentage Error (MAPE), and Percent bias (Pbias, unit: %).

PCC measures the linear correlation between modeled and observed TWS anomalies, and is expressed as:

$$\text{PCC} = \frac{COV(Q_m, Q_o)}{\sigma Q_m \sigma Q_o} \tag{1}$$

where Q_m and Q_o are the reconstructed and observed TWS anomalies respectively; COV is the covariance of Q_m and Q_o; σQ_m and σQ_o are the standard deviations of the modeled and observed TWS anomalies, respectively.

The NSE metric is widely used to determine overall model efficiency in hydrological fields, and is computed from model-simulated and observed TWS anomalies time series as follows:

$$\text{NSE} = 1 - \frac{\sum_{t=1}^{T}(Q_m^t - Q_o^t)^2}{\sum_{t=1}^{T}(Q_o^t - \overline{Q_o})^2} \tag{2}$$

where Q_m^t and Q_o^t are modeled and observed TWS anomalies at time *t*. $\overline{Q_o}$ is the mean observed TWS anomalies. NSE can range from $-\infty$ to 1, and the closer the NSE is to 1, the more reliable is the match between modeled and inferred TWS anomalies time series.

KGE measures the Euclidean distance between a point and the optimal point, and is calculated as:

$$\mathrm{KGE} = 1 - \sqrt{(PCC-1)^2 + (BR-1)^2 + (RV-1)^2} \tag{3}$$

where

$$\mathrm{BR} = \overline{Q_m}/\overline{Q_o} \tag{4}$$

and

$$\mathrm{RV} = (\sigma Q_m/\overline{Q_m})/(\sigma Q_o/\overline{Q_o}) \tag{5}$$

A KGE of 1 indicates perfect agreement between simulations and simulated TWS anomalies.

R^2 measures the proportion of variation in the dependent variable explained by the predictors included in the model, which is expressed as:

$$\mathrm{R}^2 = 1 - \left(\frac{\sum_{t=1}^{T}(Q_m^t - Q_0^t)^2}{\sum_{t=1}^{T}(Q_0^t - \overline{Q_o})^2}\right) \tag{6}$$

When the R^2 is approaching 1, it means that the model has better performance.

The RMSE is a frequently used measure of the differences between predictors and the observations:

$$\mathrm{RMSE} = \frac{\sqrt{\sum_{t=1}^{T}(Q_m^t - Q_0^t)^2}}{T} \tag{7}$$

RMSE can be used to compare different models. However, RMSE does not perform well if comparing models fits for different response variables or if the response variable is standardized, log-transformed, or otherwise modified. To overcome these issues, the NRMSE is also used:

$$nRMSE = \frac{RMSE}{\sigma Q_o} = \frac{\sqrt{\sum_{t=1}^{T}(Q_m^t - Q_0^t)^2}}{\sum_{t=1}^{T}(Q_0^t - \overline{Q_o})^2} \tag{8}$$

The MAPE measures the accuracy of model forecasts as a percentage. It can be calculated as the average of the absolute differences between predicted and actual values, divided by the actual values, for each time period.

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{Q_m^t - Q_0^t}{Q_0^t}\right| \quad (9)$$

Pbias measures the percentage bias of modeled TWS anomalies that are larger or smaller than the corresponding inferred natural TWS anomalies. A Pbias of 0 indicates perfect alignment. Pbias is computed as:

$$\text{Pbias} = \frac{\sum_{t=1}^{T}(Q_m^t - Q_0^t)}{\sum_{t=1}^{T} Q_0^t} \times 100\% \quad (10)$$

For the metrics of RMSE, nRMSE, MAPE and Pbias, a smaller metric value indicates better performance of the model simulations.

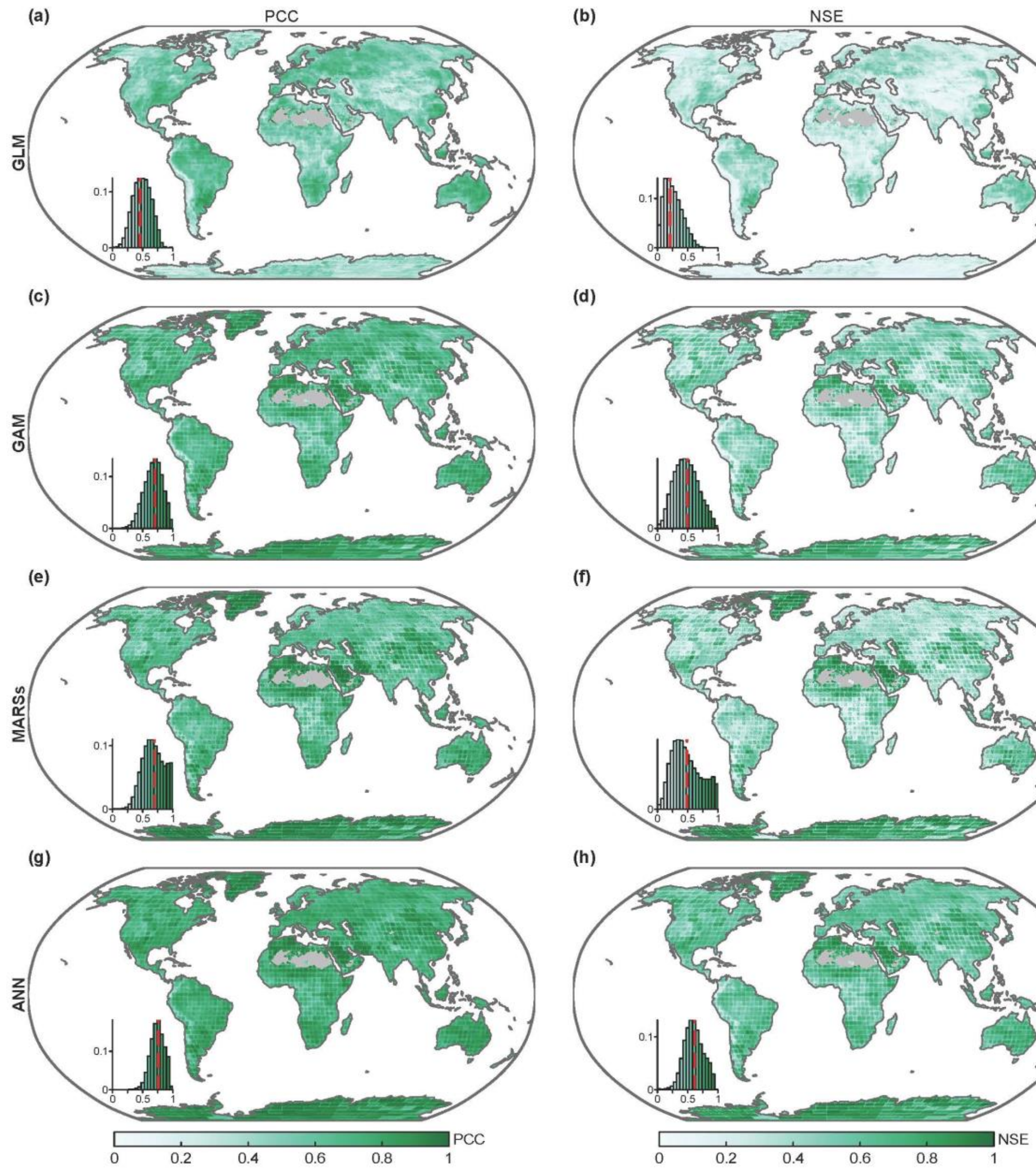

Figure S1. Performance of different machine learning models in simulating JPL TWS anomalies under scheme 8 during the test period. The left plots indicate the value of PCC, and the right plots show the value of NSE. Insets in each figure show the histogram of these metrics, with the dashed vertical line showing the median value. Data-sparse areas without reconstruction are marked in grey.

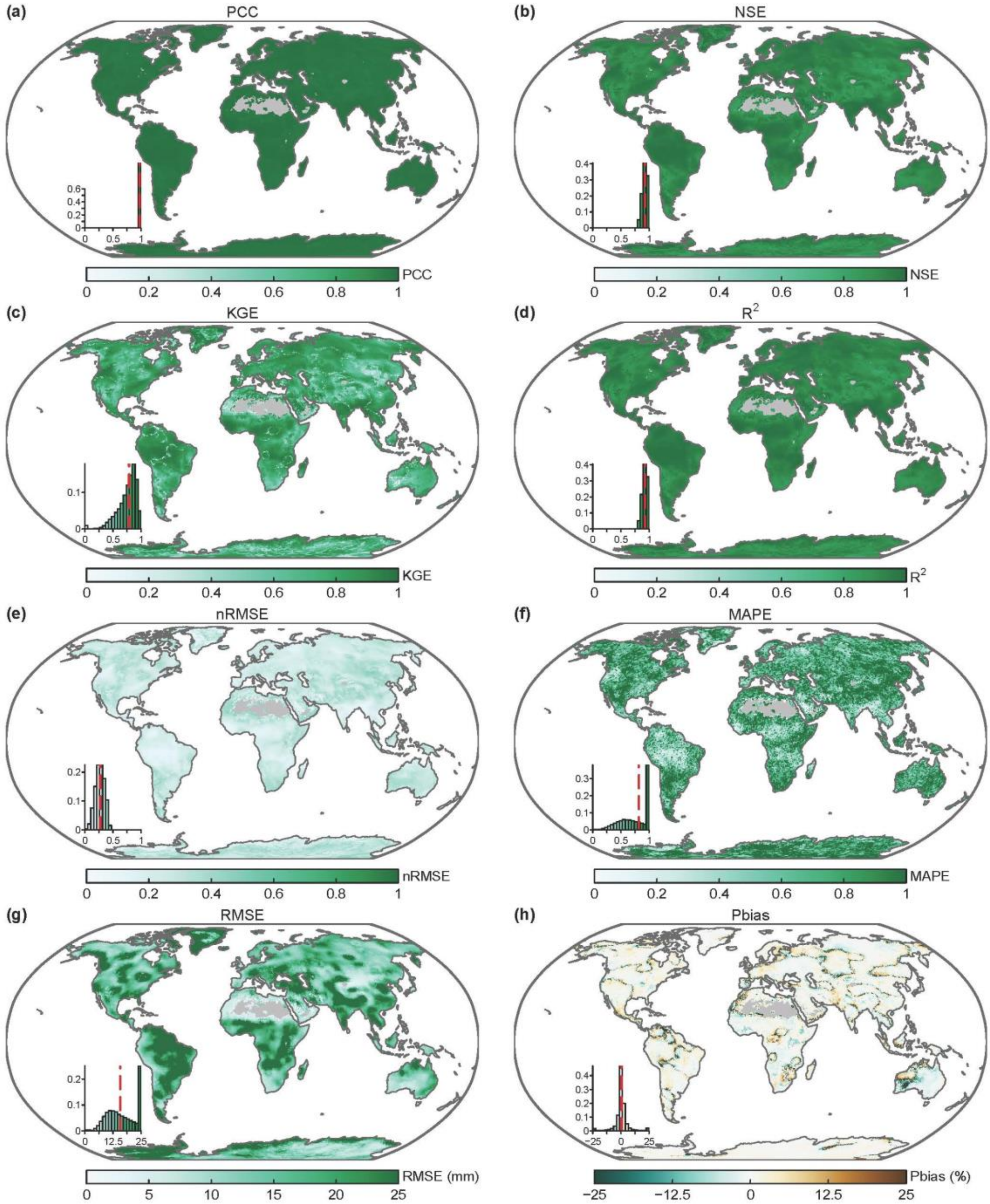

Figure S2. Comparison of our reconstructed CSR TWS anomalies against the GRACE/GRACE-FO observations. Insets in each figure show the histogram of these metrics, with the dashed vertical line showing the median value. Data-sparse areas without reconstruction are marked in grey.

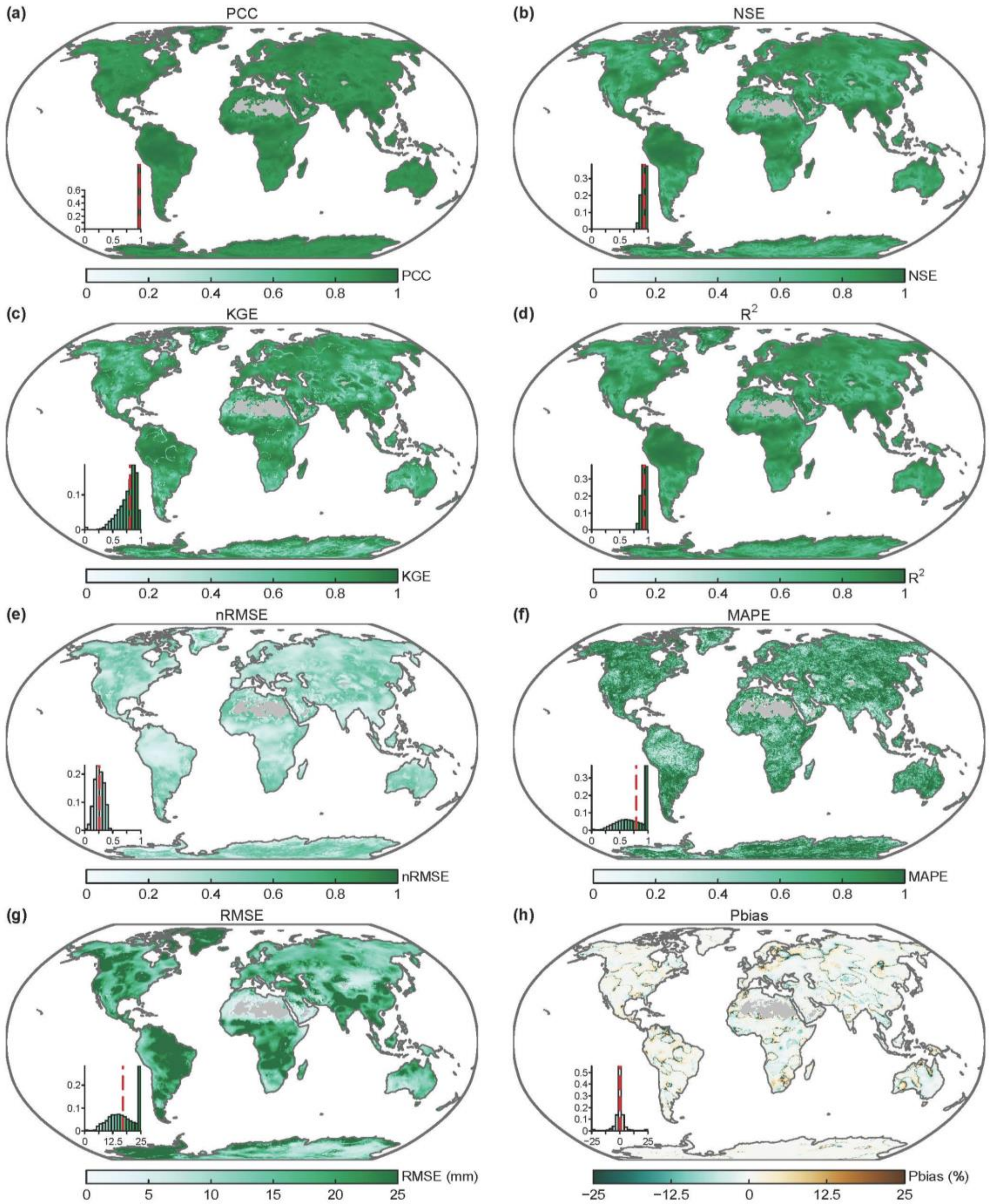

Figure S3. Comparison of our reconstructed GSFC TWS anomalies against the GRACE/GRACE-FO observations. Insets in each figure show the histogram of these metrics, with the dashed vertical line showing the median value. Data-sparse areas without reconstruction are marked in grey.

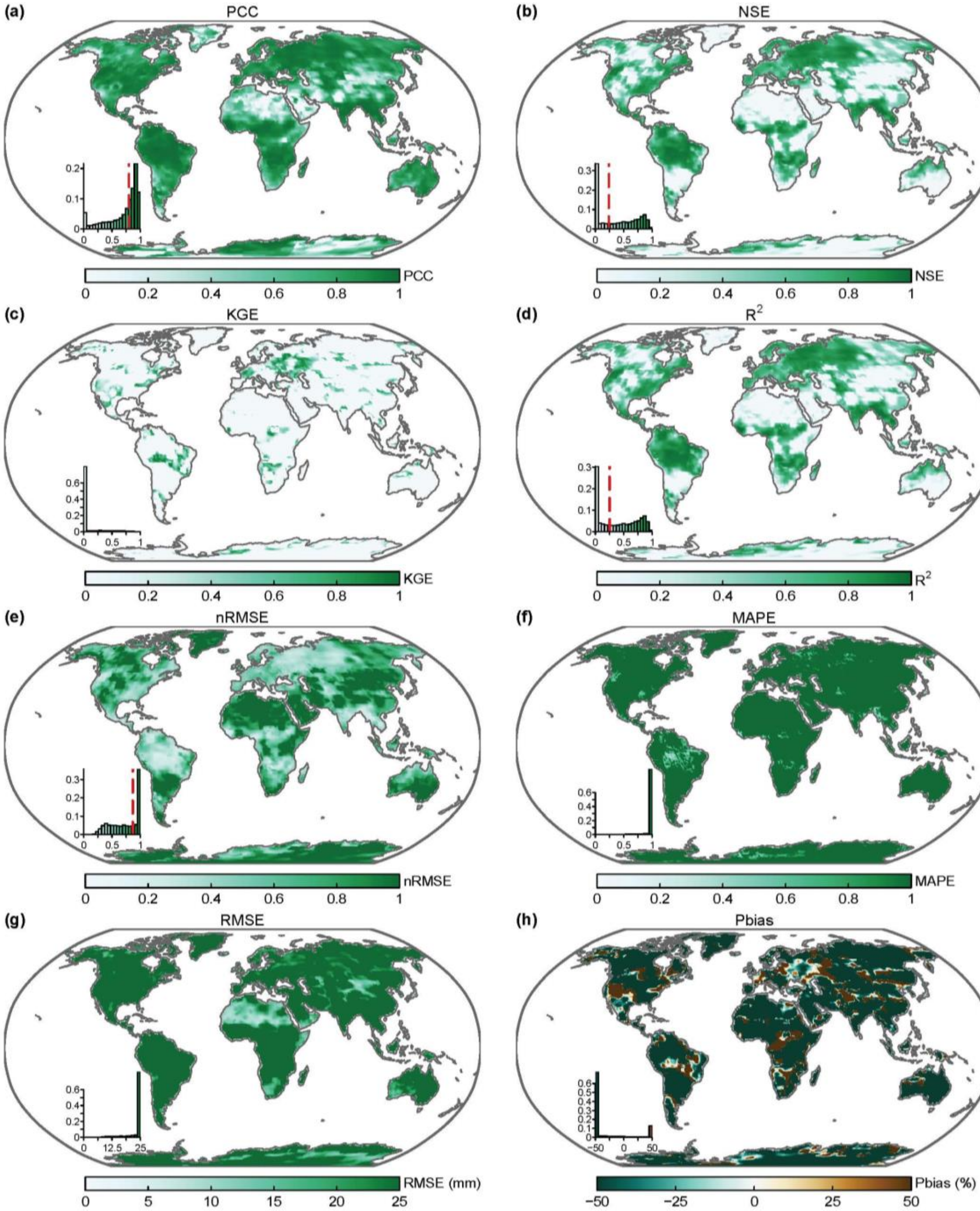

Figure S4. Comparison of the GRACE-REC dataset against the GRACE/GRACE-FO observations. Insets in each figure show the histogram of these metrics, with the dashed vertical line showing the median value. Data-sparse areas without reconstruction are marked in grey.

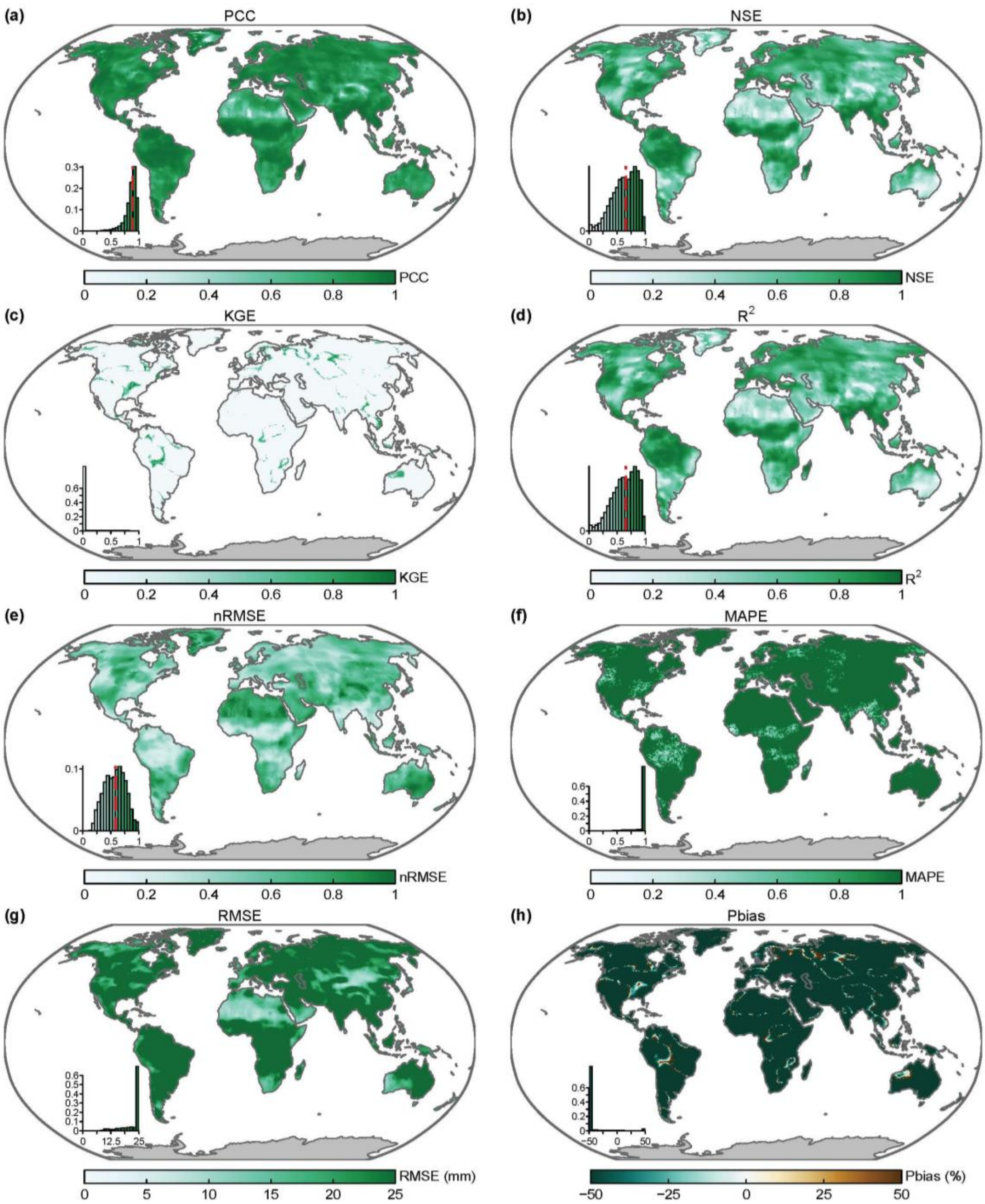

Figure S5. Comparison of the GRL reconstructed dataset against the GRACE/GRACE-FO observations. Insets in each figure show the histogram of these metrics, with the dashed vertical line showing the median value. Data-sparse areas without reconstruction are marked in grey.

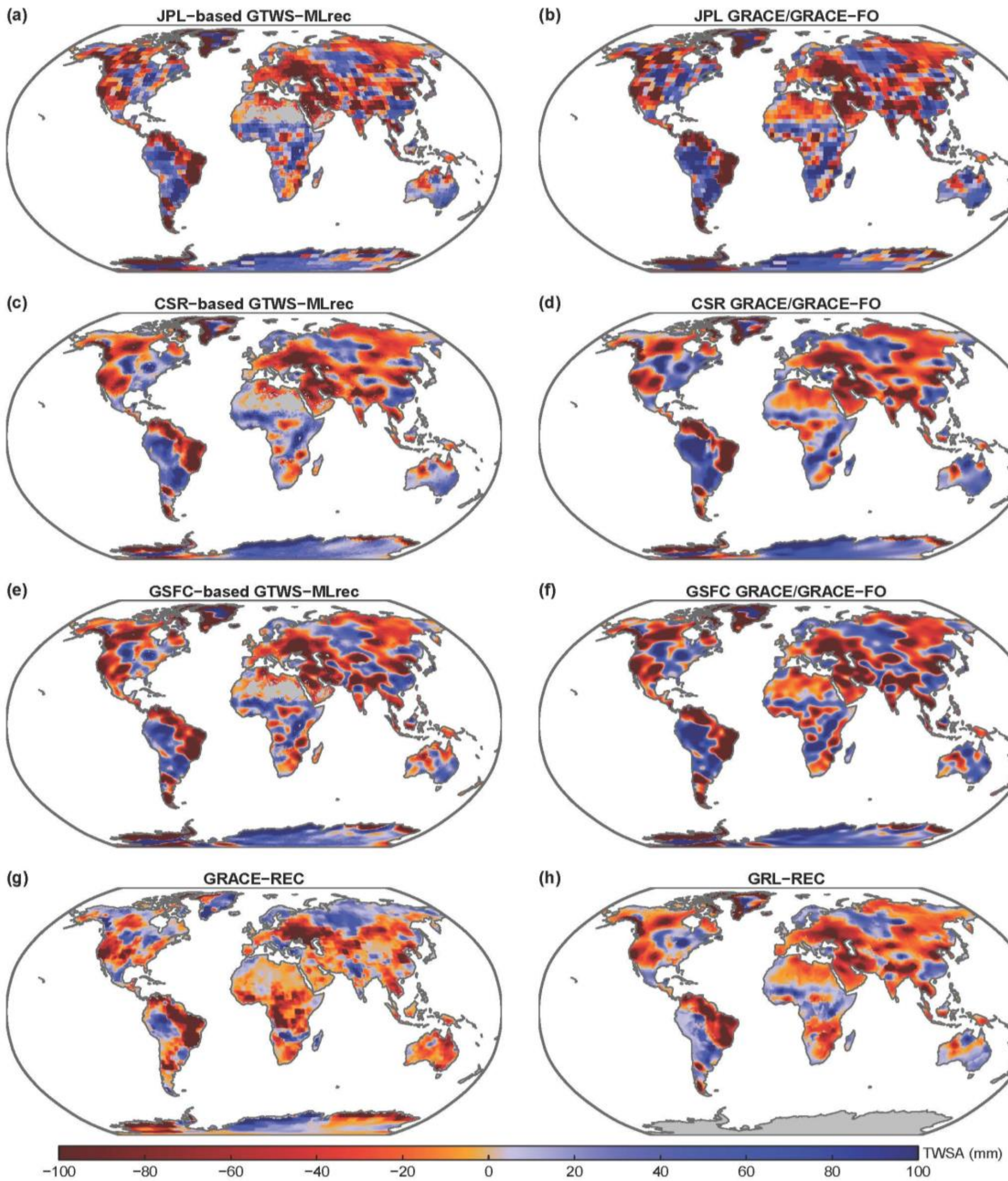

Figure S6. Global map of TWS anomalies in 2015 under GRACE/GRACE-FO (right column) and for the different reconstruction datasets (left column).

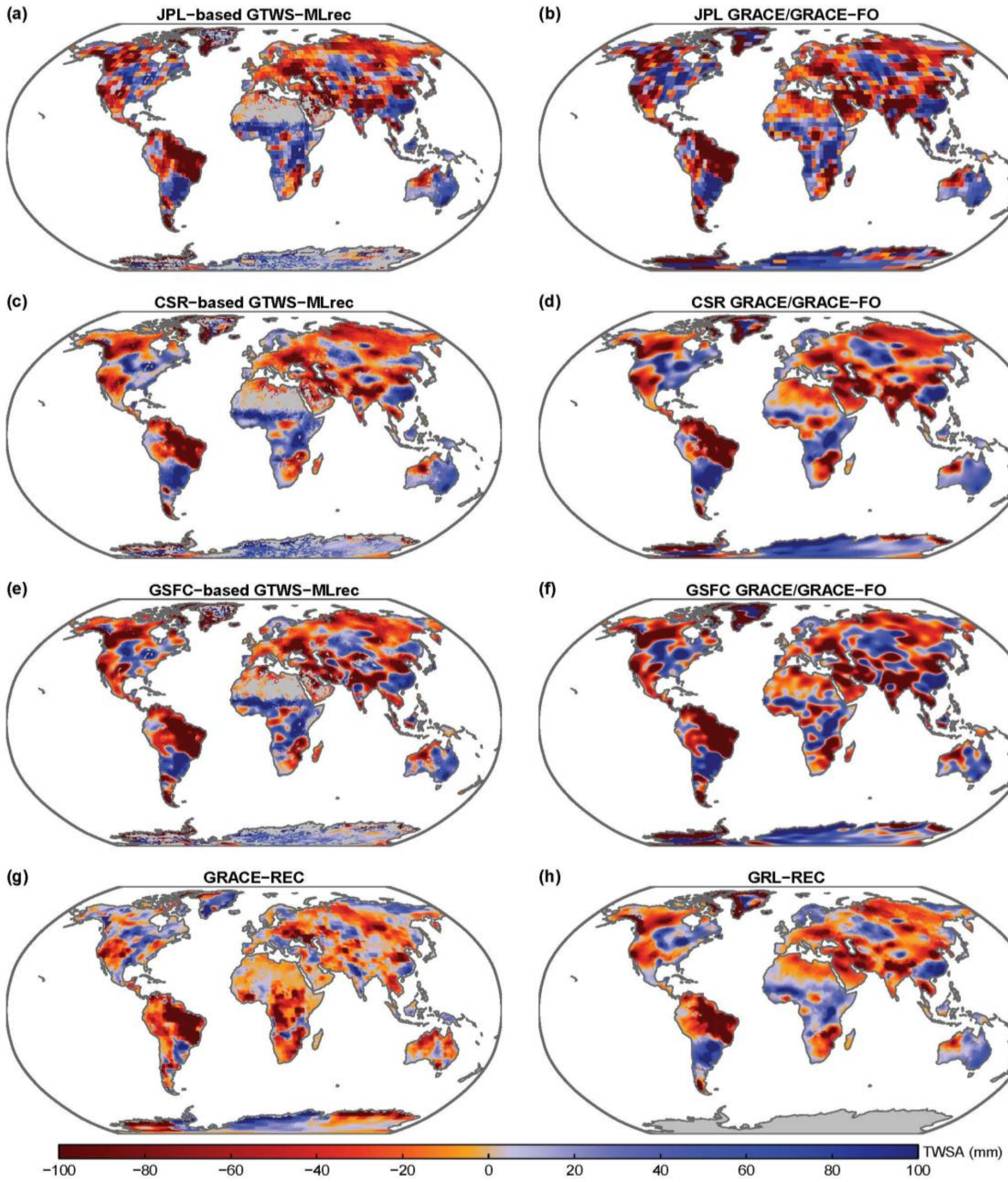

Figure S7. Global map of TWS anomalies in 2016 under GRACE/GRACE-FO and for the different reconstruction datasets.

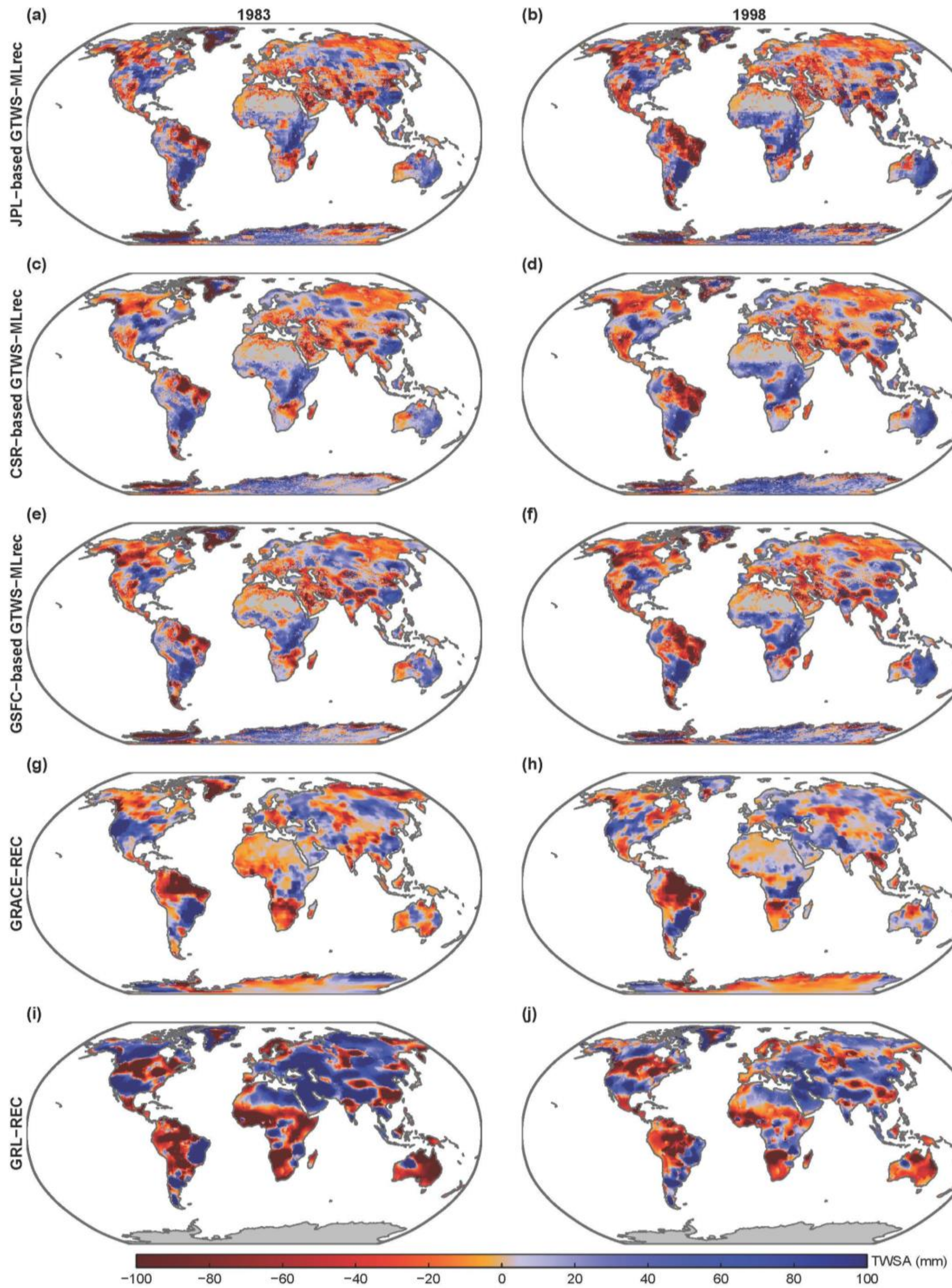

Figure S8. Global map of TWS anomalies in 1983 (left column) and 1998 (right column) for the different reconstruction datasets.