Supplemental Material

Supplemental Material: Predictive skill assessment for land water storage in CMIP5 decadal hindcasts by a global reconstruction of GRACE satellite data

L. Jensen[1], A. Eicker[1], T. Stacke[2], and H. Dobslaw[3]

[1]Geodesy and Geoinformatics, HafenCity University, Hamburg, Germany
[2]Helmholtz-Zentrum Geesthacht, Centre for Materials and Coastal Research, Geesthacht, Germany
[3]Helmholtz Centre Potsdam, German Research Centre for Geosciences (GFZ), Potsdam, Germany

S1: Comparison of GRACE and GRACE-REC

Figure 1 shows the yearly global mean anomaly time series of GRACE-REC (Humphrey & Gudmundsson, 2019), as displayed in figure 1 of the main text, together with the original GRACE observations (Luthcke et al., 2013) used for creation of the GRACE-REC data set for the time span 1970 – 2014 (2003 – 2014 for observations). The GRACE observations lie within the error bounds of the reconstruction and the correlation of the two time series is 0.92. Thus, we consider GRACE-REC as a reasonably reliable proxy within the realm of our study.

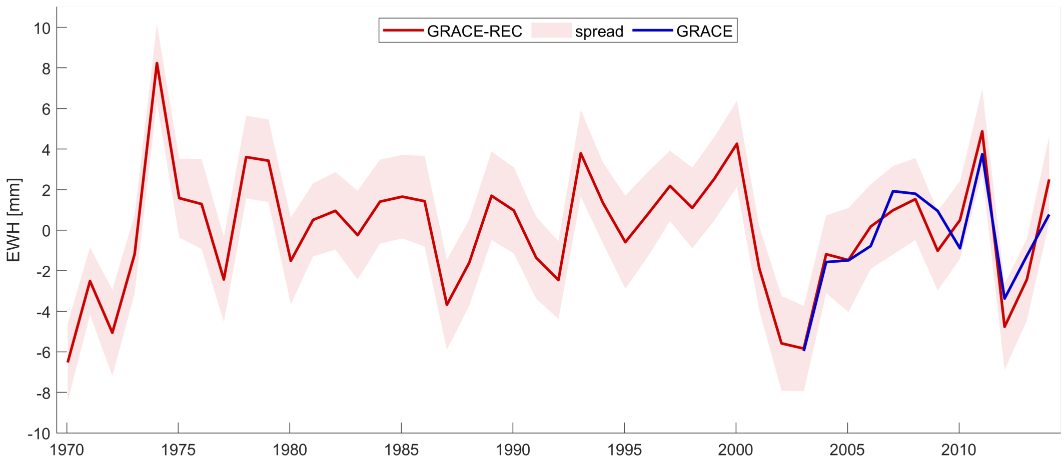

Figure 1. Yearly global mean anomaly time series of GRACE-REC (red) and original GRACE observations (blue) for the time span 1970 – 2014 (2003 – 2014 for observations).

Corresponding author: Laura Jensen, `laura.jensen@hcu-hamburg.de`

S2: Calculation of uncertainties

Along with the GRACE-REC time series $o(t)$, Humphrey and Gudmundsson (2019) provide an ensemble of randomly generated ensemble members $y_i(t)$ to incorporate the spatial and temporal error structure of the gridded TWS reconstruction into uncertainty estimates of spatial averages (e.g. global continental mean). The standard deviation of the GRACE-REC time series is computed as the unbiased sample standard deviation of these perturbed ensemble members $y_i(t)$ according to

$$\sigma_o(t) = \sqrt{\frac{1}{M-1}\sum_{i=1}^{M}(y_i(t) - \bar{y}(t))^2}. \tag{1}$$

Here, M is the number of ensemble members (here 100), $y_i(t)$ is the TWS anomaly of the GRACE-REC member i at time step t, and $\bar{y}(t)$ is the arithmetic mean of all ensemble members.

The Init and Hist multi-model mean (MMM) time series (see figure 1 of the main text) are each calculated as the weighted mean of all ensemble members, to avoid that a single model contributes to the MMM with a higher weight due to a larger number of ensemble members:

$$m(t) = \frac{\sum_{i=1}^{N} w_i x_i(t)}{\sum_{i=1}^{N} w_i} = \frac{1}{V}\sum_{i=1}^{N} w_i x_i(t) \tag{2}$$

The total number of ensemble members is denoted by N (e.g. 39 for Init), $x_i(t)$ is the mTWS value of ensemble member i at time step t, and w_i is the weight assigned to the ensemble member i. The weights $w_i = 1/K$ are calculated as the reciprocal value of the number K of ensemble members per model, with K varying between 3 and 10. E.g., if a model has 3 members, each of them gets a weight of 1/3. As a result, all weights w_i sum up to the number of models $V = \sum_{i=1}^{N} w_i$ (e.g. 5).

When calculating the uncertainties of the MMM time series also the internal model uncertainties have to be taken into account, i.e., the weighted standard deviation has to be applied. Analogously to the weighted mean, the (biased) weighted standard deviation is given by

$$\sigma_x(t) = \sqrt{\frac{1}{V}\sum_{i=1}^{N} w_i(x_i(t) - m(t))^2}. \tag{3}$$

However, to obtain the *unbiased* weighted standard deviation the factor $\frac{1}{V}$ in equation 3 has to be adjusted. This is similar to the bias correction in the case of an unweighted standard deviation, where instead of $\frac{1}{M}$ the factor $\frac{1}{M-1}$ is applied (see equation 1). Bias correction for weighted standard deviation is not straight forward (Gatz & Smith, 1995), but it can be shown that

$$\sigma_x(t) = \sqrt{\frac{V}{V^2 - \sum_{i=1}^{N} {w_i}^2}\sum_{i=1}^{N} w_i(x_i(t) - m(t))^2} \tag{4}$$

is a good estimate for the *unbiased* weighted standard deviation, with the adjusted factor according to Kish (1965). Equation 4 describes the sample standard deviation for $x_i(t)$, i.e. the ensemble spread around the MMM, but not the standard deviation $\sigma_m(t)$ of the MMM $m(t)$ itself. Formally, $\sigma_m(t)$ can be derived via variance propagation of equation 2 utilizing the full variance-covariance matrix of the ensemble members $x_i(t)$. However, to come up with this covariance matrix, the error correlations between all members (of all models) have to be determined, which is not trivial. There is an ongoing discussion about the dependence of models and derived accuracies of multi-model averages in the climate modeling community (Knutti et al., 2010; Pennell & Reichler, 2011; Abramowitz & Bishop, 2015). In this study, we make the practical assumption of full error correlations between the members belonging to a particular model, and no error correlations between the members of different models. We admit that this is a simplification of the real error structure, however, the outlined uncertainty assessment can easily be adjusted if more realistic correlation estimates become available. With the current assumptions the standard deviation of the MMM becomes

$$\sigma_m(t) = \frac{1}{\sqrt{V}}\sigma_x(t). \tag{5}$$

The standard deviations of the correlations and RMSDs between the MMM and GRACE-REC provided as 1-sigma boundaries in figures 2, 3, 4, and 6 of the main text, are obtained via variance propagation from the standard deviations $\sigma_m(t)$ of the Init/Hist time series and $\sigma_o(t)$ of the GRACE-REC data set. Given that the bias of the time series is removed and thus the temporal mean is zero, the correlation of the MMM time series $m(t)$ and the GRACE-REC observational record $o(t)$ is

$$\rho = \frac{s_{mo}}{s_m \cdot s_o}, \tag{6}$$

where

$$s_{mo} = \textstyle\sum_{t=1}^{T} m(t)o(t), \quad s_m = \sqrt{\textstyle\sum_{t=1}^{T} m(t)^2}, \quad \text{and} \quad s_o = \sqrt{\textstyle\sum_{t=1}^{T} o(t)^2}, \tag{7}$$

and T denoting the length of the yearly time series (41 years). The standard deviation σ_ρ of the correlation is calculated by inserting equations 7 into equation 6 and performing variance propagation with the covariance matrix of $m(t)$ and $o(t)$ obtained from $\sigma_m(t)$ and $\sigma_o(t)$. As our time series has a low temporal resolution of yearly averages, we assume the error correlations between subsequent time steps to be negligible and set the corresponding elements in the covariance matrix to zero. If any reasons for assuming different correlations should arise, these can easily be adopted within this framework.

The same approach as described for the correlations is applied for the derivation of the standard deviations of the RMSD. The RMSD is calculated with

$$RMSD = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(m(t) - o(t))^2}. \tag{8}$$

and by propagating the uncertainties of $m(t)$ and $o(t)$ we obtain its standard deviation σ_{RMSD}.

S3: Identification of regions with incompatibilities between TWS from models and observations

We are aware of three geophysical signals contained in GRACE-derived TWS (and thus to some extent in the GRACE-REC data set) that are not explicitly represented in the ESMs:

- surface water variability (S)
- anthropogenic groundwater abstraction and irrigation (G)
- mass displacement due to large earthquakes (E)

We note that also natural groundwater variability is a part of GRACE TWS which is probably not properly represented in the climate models, but we would like to emphasize it is difficult (1) to separate the groundwater signal from the remaining compartments in the observations, and (2) to assess the degree to which groundwater is implicitly contained in the variable *total soil moisture content* of an ESM. We therefore focus in the following exclusively on the S/G/E effects. In regions where the influence of S/G/E is particularly large, TWS and mTWS are probably not entirely compatible which leads to a degradation of the results. We thus aim to identify such regions in order to exclude them from the further analysis.
In the following we briefly describe the data sets used and processing applied for the estimation of the magnitude of S/G/E effects and the approach for deriving the mask of regions to be excluded:

Surface water variability
Within the realm of the Research Unit GlobalCDA (funded by the German Research Foundation) a data set was produced which describes the monthly (2003/01 – 2016/12) mean influence of surface water storage change in lakes and reservoirs (in total 283 lakes/reservoirs obtained from the DAHITI data base) on the GRACE TWS signal (Deggim et al., manuscript in preparation). For this data set the surface water extent (from remote sensing) was combined with surface water level time series (from satellite altimetry) and converted to the spatial resolution of the GRACE data by applying appropriate spatial filtering (DDK3 filter, Kusche, 2007). We project the global 0.5° maps of this data set to 2° resolution and calculate annual anomalies. Afterward, we compute the root mean square (RMS) over 2003 – 2016 (figure 2a). We interpret this as the local influence of the annual surface water variability on the GRACE observations.

Anthropogenic groundwater abstraction
To estimate the magnitude of groundwater abstraction we make use of data from the hydrological model WaterGAP 2.2a (Döll et al., 2014). Net abstraction in WaterGAP 2.2a is defined as groundwater withdrawals minus return flow from irrigation with both surface water and groundwater. The global grids are converted from rates (in m^3/month) to monthly cumulated storage changes (EWH in mm), averaged per year, and remapped to 2° spatial resolution. Subsequently, the RMS over 1996 – 2009 (14 years) is calculated from the annual anomalies. To estimate the influence of these net abstraction changes on the GRACE observations we apply a GRACE-like spatial filtering (DDK3 filter, Kusche, 2007) to the resulting map of net abstraction RMS (figure 2b). Compared to the magnitude of surface water variability the RMS of the net abstraction is substantially smaller. This is due to the fact that anthropogenic groundwater abstraction mainly occurs as a linear mass trend whereas year-to-year variations are minor. Thus, although the regions generally affected by groundwater abstraction are of considerable size (Taylor et al., 2013), for the results of this study that focuses on annual anomalies excluding linear trends, groundwater abstraction only has a minor influence.

Mass displacement due to large earthquakes
Large earthquakes involve mass displacements detectable with GRACE that in first approximation cause a step function in the GRACE-derived time series of mass variations (with the discontinuity at the time of the earthquake). This step function cannot be removed by subtracting bias and linear trend as is done for the creation of the GRACE-REC data set. Thus, these mass variations distort the TWS estimates of GRACE-REC in earthquake regions. In-

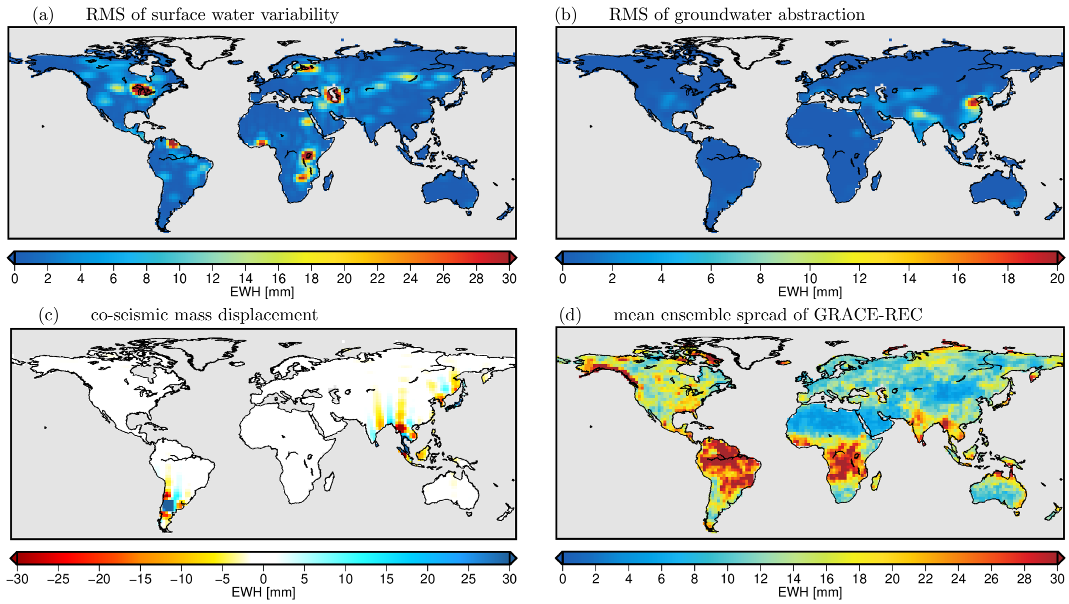

Figure 2. (a) RMS of annual surface water anomalies from 2003 – 2016 (b) RMS of annual net groundwater abstraction anomalies from 1996 – 2009 (c) co-seismic mass displacement due to Sumatra (2004), Chile (2010) and Tohoku (2011) earthquakes (d) mean standard deviation of 100 ensemble members of GRACE-REC annual anomalies between 2001 – 2014. The different time spans are due to data availability, but for comparability they all last 14 years (approx. GRACE time span).

formation on the spatial extent and magnitude of co-seismic mass variations contained in GRACE observations is, e.g., provided by Mayer-Gürr et al. (2018). They co-estimate the mass variations of the three largest earthquakes during the GRACE period (the Sumatra-Andaman 2004, the Chile 2010 and the Tohoku 2011 earthquake) together with the gravity field model ITSG-Grace2018s. The data are provided as spherical harmonic coefficients of the gravitational potential, and we convert them into EWH, apply a DDK3 filter (Kusche, 2007), and evaluate them on a 2° spatial grid (figure 2c). The assessment is restricted to the three largest earthquakes because GRACE is only sensitive to earthquakes with a magnitude of about > 8.5 (Pail et al., 2015; Han et al., 2013) and only these three earthquakes (with magnitudes 9.1, 8.8, and 9.0) are clearly above this threshold.

Derivation of a mask with regions to exclude

The magnitude of S/G/E effects (figure 2a-c) varies spatially and only substantially influences the GRACE-derived TWS in distinct regions. In order to identify these regions we use as a threshold for large S/G/E effects the noise floor of the GRACE-REC data set: For each grid cell the standard deviation of the 100 ensemble members of GRACE-REC is calculated for each year of the annual anomaly time series. We then average the standard deviations over the time span 2001 – 2014 (14 years) to obtain a mean observation spread (Figure 2d). The time span is slightly different to the GRACE time span (2003 – 2016) because the GRACE-REC data set ends in 2014. All grid cells where the RMS of the surface water variability, the groundwater abstraction, or the absolute influence of earthquakes (figure 2a-c) exceeds the spread of GRACE-REC (figure 2d) are excluded from the further analysis. These regions (figure 3) make up about 6.9% of the Earth's land surface (Greenland and Antarctica not considered).

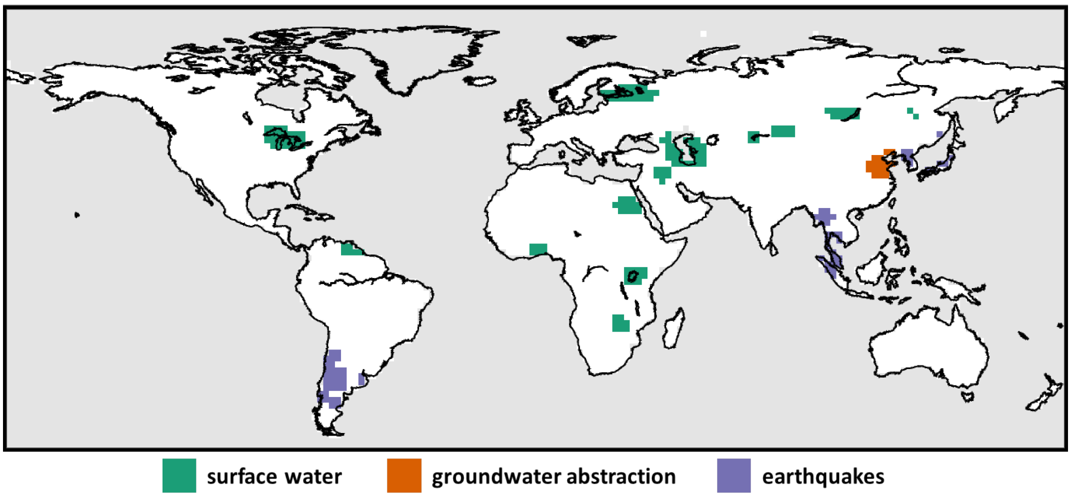

Figure 3. Mask with regions where the influence of surface water variability, anthropogenic groundwater abstraction, or the absolute mass displacement caused by earthquakes is larger than the spread of the GRACE-REC data set.

References

Abramowitz, G., & Bishop, C. H. (2015). Climate Model Dependence and the Ensemble Dependence Transformation of CMIP Projections. *Journal of Climate*, *28*(6), 2332–2348. (Publisher: American Meteorological Society) doi: 10.1175/JCLI-D-14-00364.1

Döll, P., Schmied, H. M., Schuh, C., Portmann, F. T., & Eicker, A. (2014). Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites. *Water Resources Research*, *50*(7), 5698–5720. doi: 10.1002/2014WR015595

Gatz, D. F., & Smith, L. (1995). The standard error of a weighted mean concentration—I. Bootstrapping vs other methods. *Atmospheric Environment*, *29*(11), 1185–1193. doi: 10.1016/1352-2310(94)00210-C

Han, S.-C., Riva, R., Sauber, J., & Okal, E. (2013). Source parameter inversion for recent great earthquakes from a decade-long observation of global gravity fields. *Journal of Geophysical Research: Solid Earth*, *118*(3), 1240–1267. (_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/jgrb.50116) doi: 10.1002/jgrb.50116

Humphrey, V., & Gudmundsson, L. (2019). GRACE-REC: a reconstruction of climate-driven water storage changes over the last century. *Earth System Science Data Discussions*, 1–41. doi: https://doi.org/10.5194/essd-2019-25

Kish, L. (1965). *Survey sampling.* New York: Chichester : Wiley.

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., & Meehl, G. A. (2010). Challenges in Combining Projections from Multiple Climate Models. *Journal of Climate*, *23*(10), 2739–2758. (Publisher: American Meteorological Society) doi: 10.1175/2009JCLI3361.1

Kusche, J. (2007). Approximate decorrelation and non-isotropic smoothing of time-variable GRACE-type gravity field models. *Journal of Geodesy*, *81*(11), 733–749. doi: 10.1007/s00190-007-0143-3

Luthcke, S. B., Sabaka, T. J., Loomis, B. D., Arendt, A. A., McCarthy, J. J., & Camp, J. (2013). Antarctica, Greenland and Gulf of Alaska land-ice evolution from an iterated GRACE global mascon solution. *Journal of Glaciology*, *59*(216), 613–631. doi: 10.3189/2013JoG12J147

Mayer-Gürr, T., Behzadpur, S., Ellmer, M., Kvas, A., Klinger, B., Strasser, S., & Zehentner, N. (2018). *ITSG-Grace2018 - Monthly, Daily and Static Gravity Field Solutions from GRACE.* GFZ Data Services. doi: 10.5880/ICGEM.2018.003

Pail, R., Bingham, R., Braitenberg, C., Dobslaw, H., Eicker, A., Güntner, A., ... IUGG Expert Panel (2015, November). Science and User Needs for Observing Global Mass Transport to Understand Global Change and to Benefit Society. *Surveys in Geophysics*, *36*(6), 743–772. doi: 10.1007/s10712-015-9348-9

Pennell, C., & Reichler, T. (2011). On the Effective Number of Climate Models. *Journal of Climate*, *24*(9), 2358–2367. (Publisher: American Meteorological Society) doi: 10.1175/2010JCLI3814.1

Taylor, R. G., Scanlon, B., Döll, P., Rodell, M., van Beek, R., Wada, Y., ... Treidel, H. (2013). Ground water and climate change. *Nature Climate Change*, *3*(4), 322–329. doi: 10.1038/nclimate1744