Jensen, L., Eicker, A., Dobslaw, H., Stacke, T., Humphrey, V. (2019): Long-Term Wetting and Drying Trends in Land Water Storage Derived From GRACE and CMIP5 Models. - *Journal of Geophysical Research*, *124*, 17-18, pp. 9808—9823.

DOI: http://doi.org/10.1029/2018JD029989

JGR Atmospheres

RESEARCH ARTICLE
10.1029/2018JD029989

Long-Term Wetting and Drying Trends in Land Water Storage Derived From GRACE and CMIP5 Models

L. Jensen[1], **A. Eicker**[1], **H. Dobslaw**[2], **T. Stacke**[3], **and V. Humphrey**[4]

[1]Geodesy and Geoinformatics, HafenCity University Hamburg, Hamburg, Germany, [2]Helmholtz Centre Potsdam, German Research Centre for Geosciences (GFZ), Potsdam, Germany, [3]Max Planck Institute for Meteorology, Hamburg, Germany, [4]Institute for Atmospheric and Climate Science, ETH Zürich, Zürich, Switzerland

Abstract Coupled climate models participating in the CMIP5 (Coupled Model Intercomparison Project Phase 5) exhibit a large intermodel spread in the representation of long-term trends in soil moisture and snow in response to anthropogenic climate change. We evaluate long-term (January 1861 to December 2099) water storage trends from 21 CMIP5 models against observed trends in terrestrial water storage (TWS) obtained from 14 years (April 2002 to August 2016) of the GRACE (Gravity Recovery And Climate Experiment) satellite mission. This is complicated due to the incomplete representation of TWS in CMIP5 models and interannual climate variability masking long-term trends in observations. We thus evaluate first the spread in projected trends among CMIP5 models and identify regions of broad model consensus. Second, we assess the extent to which these projected trends are already present during the historical period (January 1861 to August 2016) and thus potentially detectable in observational records available today. Third, we quantify the degree to which 14-year tendencies can be expected to represent long-term trends, finding that regional long-term trends start to emerge from interannual variations after just 14 years while stable global trend patterns are detectable after 30 years. We classify regions of strong model consensus into areas where (1) climate-related TWS changes are supported by the direction of GRACE trends, (2) mismatch of trends hints at possible model deficits, (3) the short observation time span and/or anthropogenic influences prevent reliable conclusions about long-term wetting or drying. We thereby demonstrate the value of satellite observations of water storage to further constrain the response of the terrestrial water cycle to climate change.

Key Points:
- By comparing terrestrial water storage trends from CMIP5 models and GRACE satellite data, we identify hot spot regions of wetting and drying
- Model analysis reveals that regional long-term trends start to emerge from interannual variations after 14 years
- Large model spread in water storage trends demonstrates importance of GRACE to constrain response of the water cycle to climate change

Correspondence to:
L. Jensen,
laura.jensen@hcu-hamburg.de

Citation:
Jensen, L., Eicker, A., Dobslaw, H., Stacke, T., & Humphrey, V. (2019). Long-term wetting and drying trends in land water storage derived from GRACE and CMIP5 models. *Journal of Geophysical Research: Atmospheres*, *124*, 9808–9823. https://doi.org/10.1029/2018JD029989

Received 14 NOV 2018
Accepted 17 AUG 2019
Accepted article online 29 AUG 2019
Published online 3 SEP 2019

1. Introduction

The terrestrial branch of the global water cycle is an important component of the Earth's coupled climate system: Water available in the soil critically determines biomass production that effectively takes up carbon dioxide from the atmosphere and thus constitutes the land cover and consequently also the albedo of the Earth's surface. The availability of water at the surface influences the rate of evapotranspiration and thereby the amount of latent heat absorbed by the atmosphere locally and advected to distant regions along with the tropospheric winds; and water in the form of snow cover thermally isolates the soil from the air above it. The accurate representation of the terrestrial water dynamics and its various feedbacks to the atmospheric water, energy and carbon cycles is thus critically important for interactively coupled global numerical climate models that are used to infer information about the current state and the future evolution of the Earth's climate conditions (Trenberth, 2010).

Due to their direct effect on the availability of freshwater resources, investigating climate change impacts on the global water cycle is of great societal relevance. Changes in terrestrial water storage (TWS) might reflect long-term wetting or drying in various regions of the world, and the identification of such regions is of substantial importance for water resources management. However, coupled climate models used to predict future climatic conditions still exhibit a spread in the representation of long-term trends in soil moisture and other land water related variables (Guo & Dirmeyer, 2006; Figure 12.23 in Berg et al., 2017; Collins et al., 2013; Yuan & Quiring, 2017).

Comparing the output of numerical models with observations is crucial to demonstrate their reliability and to test predictive capacities, but measurements of water storage changes are difficult to obtain. A classical

approach for the determination of TWS at basin scale is the integration of the water balance equation (precipitation minus evapotranspiration minus runoff), see Rodell et al. (2004). However, this is challenging on a global scale, since streamflow measurements are sparse and evapotranspiration is generally difficult to measure (Wartenburger et al., 2018). Especially, trends in water storage cannot be recovered well by this method due to biases in the water flux observations (Hirschi & Seneviratne, 2017).

Complementary to conventional meteorologic observations of atmospheric water fluxes, the satellite mission Gravity Recovery And Climate Experiment (GRACE; Tapley et al., 2004) in operation from 2002 to 2017 allowed for the first time the observation of water storage changes with global coverage from space. By evaluating relative distance changes between two spacecraft at very low altitudes of 400–500 km, time variations in the Earth's gravity field are mapped that can be unambiguously related to changes in TWS. Due to the indirect observation concept, GRACE essentially senses water mass anomalies independently of their surface exposure and thus integrates all mass changes vertically from the surface down to the deepest aquifers. This unique capability of the gravimetric method makes GRACE highly complementary to alternative radiometric satellite techniques of soil moisture remote sensing that are only sensitive to changes in the top few centimeters of soil (Dorigo et al., 2015). GRACE mission data have been used in various hydrometeorological applications, for example, Famiglietti and Rodell (2013), and it is rated among the top five priorities of the future Earth observation capacity by the most recent National Aeronautics and Space Administration decadal survey (Committee on the Decadal Survey for Earth Science and Applications from Space et al., 2018). The successor mission GRACE-FO (Follow On), launched in May 2018, is expected to continue this important observational record over the next decades (Flechtner et al., 2016), which will facilitate the separation between interannual variability and long-term climatological trends in TWS. Because the limited time span of GRACE data makes the identification of climate-related signals still challenging, this study aims to investigate how GRACE TWS trends could (and should) be compared to model-derived trends.

TWS as observed with GRACE has already been used to validate both global hydrological models (Döll et al., 2014; Eicker et al., 2014; Güntner, 2008; Syed et al., 2008) and land surface models (Scanlon et al., 2018; Zhang et al., 2017) which are driven by a prescribed meteorological forcing. In this study, we focus on interactively coupled Earth System Models (ESMs) participating in CMIP5 (Coupled Model Intercomparison Project Phase 5, Taylor et al., 2011). Comparing GRACE trends with long-term coupled climate model projections is challenging in mainly two aspects: (i) In contrast to GRACE TWS (i.e., the full integrated water column, including all water reservoirs), TWS in the models is reflected typically only by means of snow storage and soil moisture. The representation of the latter critically depends on the depth of the soil column and the number of vertical layers considered. In particular, current ESMs do not explicitly simulate groundwater storage changes. As groundwater-surface interactions play an important role in the global hydrological cycle, this poses an additional source of uncertainty in long-term model projections of wetting and drying. (ii) Coupled runs in CMIP5 starting from preindustrial conditions and extending over the whole historical period until the present day are forced with temporally variable solar radiation, aerosols, CO_2 concentrations, and land use. Those experiments are thus expected to reproduce the climate variability in a statistical sense only. As a result, different realizations of the interannual and decadal climate variability are superimposed over the climatological trends so that a direct comparison with the 14-year GRACE TWS time series only has limited explanatory power. While a regional study for the Mississippi Basin (Freedman et al., 2014) showed reasonably good agreement for the annual amplitude of GRACE data and a subset of CMIP5 models, Fasullo et al. (2016) found the trends from historical CESM1-CAM5 runs compared to GRACE to be dominated by internal variability rather than by the forced response. Different drivers of TWS trends observed by GRACE were investigated by Rodell et al. (2018), who also made use of CMIP5 model precipitation projections to attribute wetting and drying tendencies in some regions to climate-driven precipitation changes. To our knowledge, an extensive global comparison of soil moisture and snow trends modeled over more than two centuries (in the following referred to as bicentennial) against GRACE observations has never been conducted with an ensemble of models such as CMIP5.

In response to these challenges, we focus in this study in particular on the correspondence of bicentennial trends in TWS as simulated by the majority of CMIP5 models and TWS tendencies as observed by GRACE and investigate regions of agreement and disagreement on wetting or drying trends in models and satellite observations.

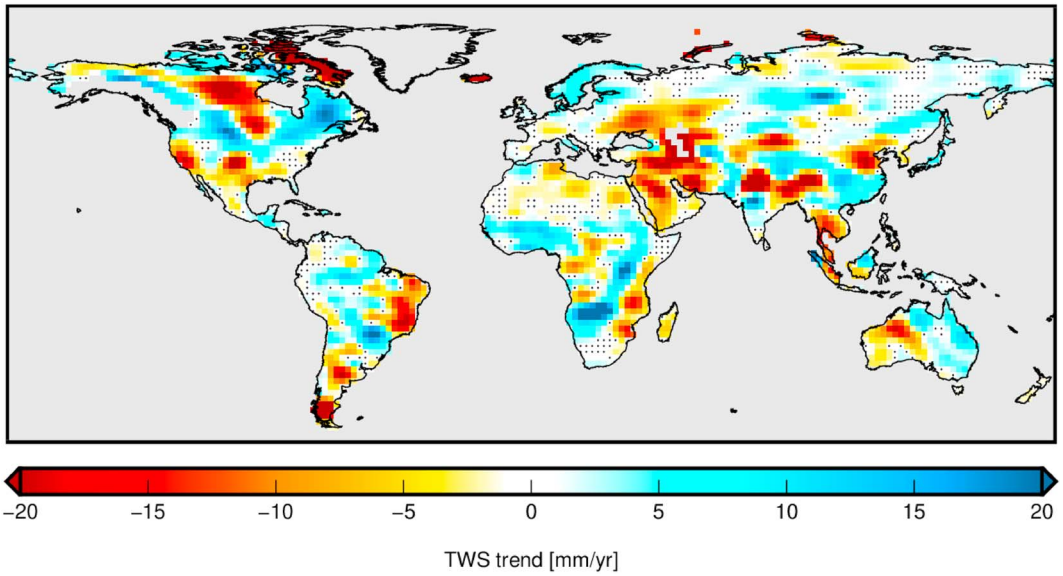

Figure 1. TWS trends from ITSG-Grace2018s (preliminary) for the time span April 2002 to August 2016 (without Greenland, Svalbard, Gulf Coast of Alaska, and Antarctica). Stippling indicates regions with nonsignificant trends ($\alpha = 0.05$).

This paper is structured as follows: First, we compute global maps of TWS trends from GRACE data (section 2) and CMIP5 models (section 3) together with an evaluation of the variability among different models and within historical and future time spans. We estimate the influence of the different time series lengths for GRACE and models by means of two model studies using model TWS tendencies from time periods ranging from 14 to more than 200 years (section 4). The TWS trend maps from GRACE and CMIP5 models are subsequently compared (section 5). Next, we investigate hot spot and noncompliance regions of wetting and drying trends regarding their uncertainty (section 6), which might be caused by model deficits or natural interannual variability and human impacts affecting the GRACE-derived trends. Section 7 summarizes the results and addresses future work.

2. TWS Trends From GRACE Data

To obtain a global grid of observed TWS trends we use the ITSG-Grace2018s trend Level 2 data (Mayer-Gürr et al., 2018), which was obtained from estimating a long-term mean gravity field model together with linear trend and annual cycle from all available GRACE Level 1B RL03 data. The ITSG-Grace2018s trend used here is a preliminary version containing Level 1B data of the time span April 2002 to August 2016 (~14 years). It will be updated once the complete time series of Level 1B RL03 data (April 2002 to June 2017) is available. However, for the trend only minor changes are expected by extending the time series by less than 1 year.

The spherical harmonic coefficients (Level 2) of the trend in gravitational potential are given up to degree $n_{\max} = 120$ and are postprocessed as follows: The effect of geocenter motion is taken into account by augmenting the GRACE data with the linear trends of degree 1 harmonic coefficients provided by Swenson et al. (2008). The zonal Δc_{20} trend coefficient is replaced using a result from Satellite Laser Ranging (Cheng et al., 2013). To reduce mass trends originating from glacial isostatic adjustment (GIA), we subtract a model from A et al. (2013) and to mitigate the effect of correlated noise a DDK4 filter (Kusche, 2007) is applied. We calculate the TWS trend on a 2° × 2° geographical grid (Figure 1) according to

$$tws(\lambda,\theta) = \frac{M}{4\pi R^2 \rho_w} \sum_{n=1}^{n_{\max}} \sum_{m=-n}^{n} \frac{(2n+1)}{(1+k'_n)} \Delta c_{nm} Y_{nm}(\lambda,\theta) \tag{1}$$

where λ and θ denote the spherical coordinates, M and R are the mass and the radius of the Earth, $\rho_w = 1{,}000$ kg/m^3 is the density of water, k'_n denote the Load Love Numbers (Lambeck, 1988), Δc_{nm} are the filtered spherical harmonic coefficients of the gravitational potential, and $Y_{nm}(\lambda,\theta)$ are the surface

spherical harmonic functions. Corresponding standard deviations of the TWS trends are obtained by variance propagation from realistic error assumptions provided with the Δc_{nm} coefficients of the ITSG-Grace2018s trend.

The significance of the trend can be tested with a parameter test. The estimated trend divided by its estimated standard deviation is compared to the critical value of the normal distribution for a certain significance level $1 - \alpha$ which we set to 95% in this study. Generally, the reliability of trends from GRACE is high. Among the solutions of different GRACE processing centers trends over the same time period are very similar (Scanlon et al., 2018), even if a different representation (mascons instead of spherical harmonics) is chosen. Thus, selecting another GRACE solution (e.g., from JPL or CSR) does not alter the findings of our study (not shown).

GRACE-derived trends might not originate purely from TWS changes everywhere, as residual tectonic effects from GIA (Caron et al., 2018), postseismic deformation after large earthquakes (Han et al., 2008, 2010), or residual atmospheric mass variability (Fagiolini et al., 2015) can overlay TWS trends. Furthermore, leakage of signal into neighboring grid cells due to filtering and residual noise that could not be removed during filtering might also distort TWS trends.

As the GRACE TWS trends are only calculated from 14 years of data, the results can be dominated by low-frequency climate variability related to El Niño–Southern Oscillation (Ni et al., 2018; Phillips et al., 2012), the solar cycle (Bhattacharyya & Narasimha, 2005), the quasi-biennial oscillation and other coupled climate modes (Gray et al., 2018), and episodic events as volcanic eruptions (Iles et al., 2013), which may either conceal the long-term trend or produce a spurious transient trend. Approaches to reduce these interannual variabilities in the GRACE record are currently being discussed (e.g., Eicker et al., 2016).

3. TWS Trends From CMIP5 Model Data

As CMIP5 models do not provide a standard output variable for total water storage, we use the sum of total soil moisture content (mrso) and surface snow amount (snw) as an approximation of it. In the remaining part of the paper we refer to this TWS approximation as model TWS (mTWS). The mTWS differs in several aspects from GRACE-derived TWS: Soil moisture layers in ESMs have a depth that can vary widely between just a few and up to tens of meters depending on the model and thus does not necessarily capture the full soil moisture content at every location. Furthermore, groundwater and surface water are not explicitly included in mTWS as these states are generally not represented in CMIP5 models. However, a certain fraction of these quantities might be implicitly included in total soil moisture as the transport to ocean and atmosphere is limited and the water balance is largely closed by most of the models (Liepert & Lo, 2013). Moreover, historical CMIP5 runs do not contain regional anthropogenic intervention other than land use changes in their setup (e.g., groundwater depletion or dam building is not represented), whereas GRACE observations include their consequences. The representation of mTWS differs from model to model due to different root depths, number of soil layers, and model physics (Huang et al., 2016). Snw also exhibits large intermodel differences in representation (Brutel-Vuilmet et al., 2013). We therefore note that mTWS of different models might not be fully compatible.

After adding monthly mrso and snw for each model, we concatenate the corresponding mTWS simulations of the historical runs (1850–2005) and the RCP8.5 scenarios (2006–2100) to calculate trends for time spans that go beyond the year 2006. For those models, where more than one run is available, we calculate the ensemble mean which we regard as the most robust realization for long-term mTWS trend estimates. Afterward, the mTWS values are remapped to a common $2° \times 2°$ geographical grid.

A bicentennial mTWS trend map (time span January 1861 to December 2099, i.e., earliest/latest common date of all models for historical/RCP8.5 experiments) is calculated from the time series of mTWS grids for each model. For each grid cell the linear trend is calculated by fitting a function

$$f(t) = a + b \cdot t + c \cdot \cos(\omega t) + d \cdot \sin(\omega t) + e \cdot \cos(2\omega t) + f \cdot \sin(2\omega t) \quad (2)$$

with parameters for bias (a), linear trend (b), annual and semiannual cycle (c, d, e, f) to the time series by means of least squares adjustment. The standard deviation of the trend is estimated from the postfit residuals. Note that we exclude the glaciated regions of Greenland, Svalbard, Gulf Coast of Alaska, and Antarctica, since not all models properly represent glacier mass balance dynamics dominating TWS in those regions.

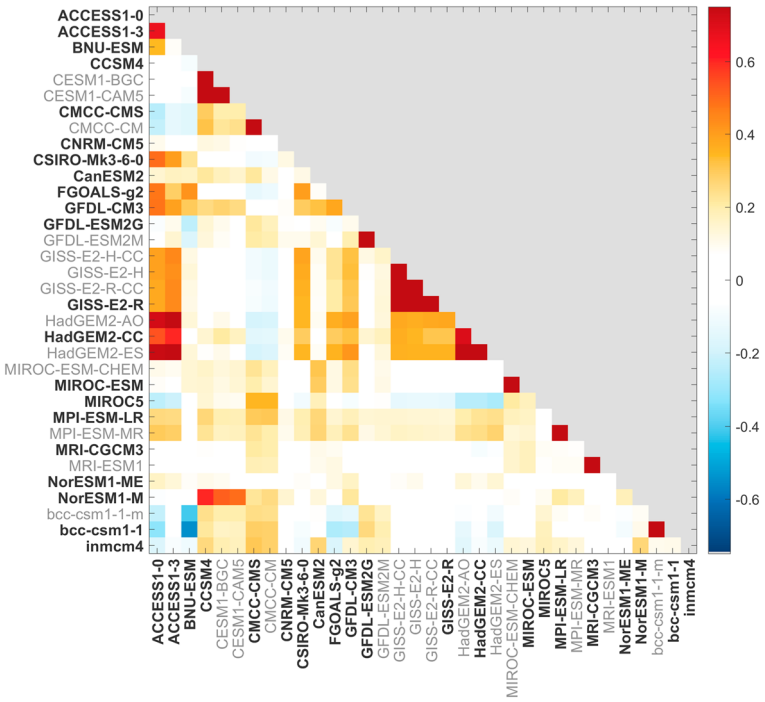

Figure 2. Correlations of the bicentennial mTWS trend maps for 34 CMIP5 models.

In total 34 CMIP5 models provide at least one run for mrso and snw. However, as some of these 34 models are either different versions of the same model or are runs with partly identical components (land surface and/or atmosphere model), it cannot be assumed that each model produces a completely independent estimate for the mrso and snw fields (Knutti et al., 2013). In order to obtain an unbiased multimodel average mTWS trend map and a reliable conclusion about model consensus, we identify the independent models by comparing the similarity of mTWS trend maps for all models. As a measure for the similarity of two maps we use the Pearson product-moment correlation coefficient r^2 calculated from the vectorized maps, giving every land pixel of the 2° × 2° grid equal weight. As trend outliers in single pixels can distort the correlation coefficient we apply a simple threshold to the mTWS trend maps, excluding absolute trend values above 23 mm/year. This is the 2σ boundary of the 14-year GRACE TWS trend (Figure 1), thus it is very unlikely that bicentennial trends above these threshold are realistic.

The correlations of the bicentennial mTWS trend maps (after applying the 23 mm/year threshold) are calculated for all 34 models and arranged in a matrix (Figure 2). Detailed information and references for the models listed in Figure 2 are given, for example, in Flato et al. (2013) and are not reiterated here. As expected, models that use common atmosphere or land surface components exhibit a very high correlation. In order to only consider models that are independent and to justify the application of equal weight to each model result, in the remaining part of the study we use only one instance from each group of models that are highly correlated ($r^2 > 75\%$). In Figure 2 the models that are excluded due to this threshold are denoted in gray font and the remaining 21 models are highlighted in bold font. The criteria for choosing a specific model among highly correlated models was based on its estimated age (most recent publication), degree of specialization (most general), or spatial resolution (closest to 2° × 2°). Generally, after excluding all but one from the highly correlated models, the correlation among trend maps from different models is very low (mean $r^2 = 10\%$, maximum $r^2 = 67\%$) and for some pairs of models it is even negative (minimum $r^2 = -55\%$). This analysis demonstrates the large inhomogeneity among CMIP5 models regarding trends in mTWS.

In order to further investigate model spread we define different time spans (Table 1) for which we calculate and discuss mTWS trend maps in the following. First, the 21 models that remain after excluding highly correlated models, are used to calculate a median trend map for the bicentennial time span January

Table 1
Notation for Different Time Spans That Are Investigated for mTWS Trends

Time span	Notation
Jan 1861 to Dec 2099	bicentennial trend
Jan 1861 to Aug 2016	historical trend
Sep 2016 to Dec 2099	RCP8.5 trend
Jan 1986 to Dec 2035	50a tendency
Jan 1996 to Dec 2025	30a tendency
Apr 2002 to Aug 2016	14a tendency

1861 to December 2099 (Figure 3a), that is, for each geographical grid cell the median of the trends of all models is determined. We use the median instead of the arithmetic unweighted mean because it is much less affected by outliers and thus can be assumed to be a more robust estimate for the trend. However, to identify nonsignificant trends in the median map (stippled regions in Figure 3a) we carry out error propagation for the arithmetic mean, because this is not straightforward for the median. Figure 3a is not affected by a model drift in mTWS, as trends from preindustrial control simulations (i.e., model runs only forced with natural, nonevolving atmospheric concentrations) of the same CMIP5 models were found to be an order of magnitude smaller and thus are negligible (not shown). According to the 21 models, the largest trends occur mainly in southern Europe and Turkey, in Central America and in the west of North America, in the north of South America and in the Himalaya region. The climatological trends derived here are in agreement with the results of a previous study (Berg et al., 2017) that focused on total soil moisture, even though significant differences are present in high latitudes since mTWS also includes snowpack.

To assess the reliability of the median mTWS trends, we compute the level of consensus of the 21 models, that is, the number of models with the same bicentennial trend direction for a given grid cell (Figure 3b). The higher the consensus, the higher the certainty that the agreement is not by chance, for example, if 15 or more of 21 models agree on the sign, the probability that this is just chance is only 4% or less (Dirmeyer et al., 2013). Hence, the higher the consensus in a grid cell, the more we can trust the direction of the trend in this grid cell according to the models. In many regions high consensus corresponds to large trends and vice versa. However, this is not valid everywhere, meaning that also the sign of small trends can be represented by a majority of models (e.g., India) and inversely, there might be model disagreement about the direction of large trends (e.g., Northern Russia). For the bicentennial mTWS trend, we find 39% of the global land area to exhibit a drying (30%) or wetting trend (9%) that is supported by at least 71% (15 of 21) of the models. These findings are not free of uncertainties as the consensus map (Figure 3b) might be affected by systematic deficits in CMIP5 models, such as in particular the lack of groundwater storage in aquifers at different depth and thus very different residence times (Pokhrel et al., 2014).

To investigate if mTWS trends as calculated for the bicentennial time span are in principle already detectable in observational records available today, we compute (in addition to the bicentennial time span) mTWS trends for a historical time span January 1861 to August 2016 (until the end of the GRACE time span; Figure 4a). For comparison, also the mTWS trends for the RCP8.5 time span September 2016 to December 2099 are displayed (Figure 4b). Overall, we find a similar pattern for the historical and the RCP8.5 trend (pattern correlation of 55%), though the historical trend has a much smaller magnitude (only about 20% of RCP8.5). Furthermore, the portion of land area where the median trends are not significant (stippled areas, 95% confidence level) is larger for the historical time span than for the RCP8.5 time span. However, in 73% of the land area the historical trend is already significant and in 68% it is in agreement with the RCP8.5 trend map. In high-consensus regions (agreement of bicentennial trend sign in $\geq$71% of the models, Figure 3b) to which we restrict the analysis in section 5 and 6, the area of agreement between significant historical and RCP8.5 trends is 92%. This indicates that in most regions the current trends are set to continue in the same direction and even increase in the future, thereby suggesting that the processes shaping the climate change footprint on TWS are already acting today. The consensus among the CMIP5 models is generally lower for the historical time span (Figure 4c) than for the RCP8.5 time span (Figure 4d), which is related to the fact that stronger trends generally imply higher consensus and RCP8.5 is the scenario with the strongest climate change signal. The patterns of the consensus maps are similar for all three time spans (bicentennial, historical, and RCP8.5), thus we infer that regions of large model agreement are largely independent from the selected time span (for centennial trends).

4. Influence of Observation Time Span

From Figures 3 and 4 it can be concluded that for centennial time spans a temporally stable pattern of drying and wetting trends exists in the models. However, we cannot expect to readily find these trend patterns in

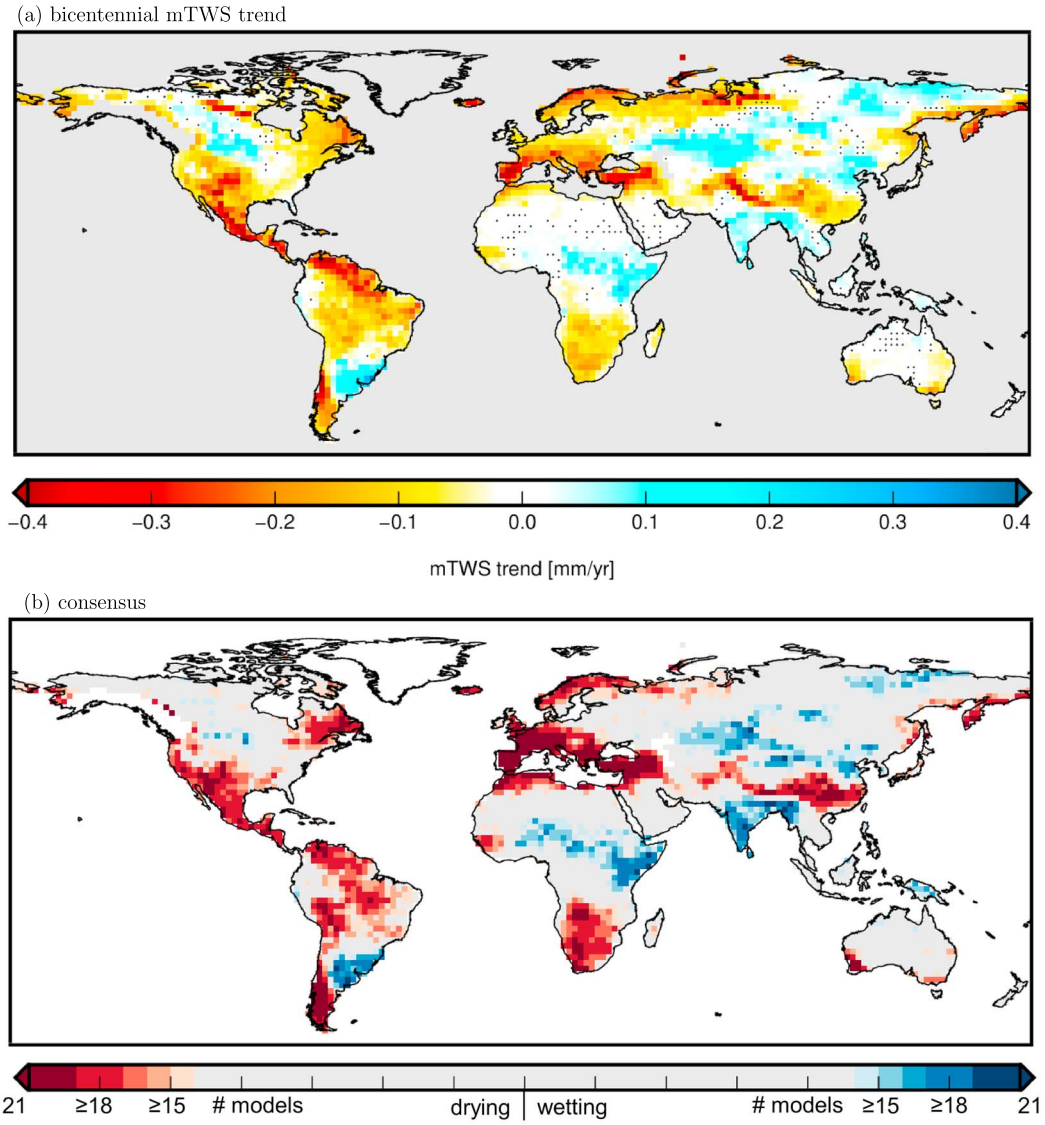

Figure 3. (a) Median of bicentennial mTWS trend maps from 21 CMIP5 models (without Greenland, Svalbard, Gulf Coast of Alaska, and Antarctica). Stippling indicates regions where the mean trend is not significantly different from zero ($\alpha = 0.05$). (b) Consensus map for bicentennial mTWS trends from 21 CMIP5 models. Red colors indicate that ≥#models agree on a negative (i.e., drying) trend, blue colors indicate that ≥#models agree on a positive (i.e., wetting) trend.

a short time period of only 14 years for which GRACE observations are available. For short time periods, interannual variations may be dominating the trend estimation in many regions of the world.

To estimate the influence of the observation time span on the expected agreement with the bicentennial trend, we perform two model studies using tendency maps for different time spans calculated from the CMIP5 models. In the first model study we investigate after which time span long-term climatic trends in mTWS might be clearly distinguished from interannual variations. In the second model study we estimate the degree to which even after long time spans the observed trends might still be in disagreement with the long-term climatic trend. In contrast to the other sections of the paper, where we rely on the ensemble means, for these model studies we only use one individual run (r1i1p1) per model in order to preserve interannual variability. This is important as natural variations would largely average out by calculating ensemble means. By using CMIP5 model output for simulating differently long observation time spans, we presume that individual model runs represent natural variability realistically in terms of relative magnitude, frequency, and duration.

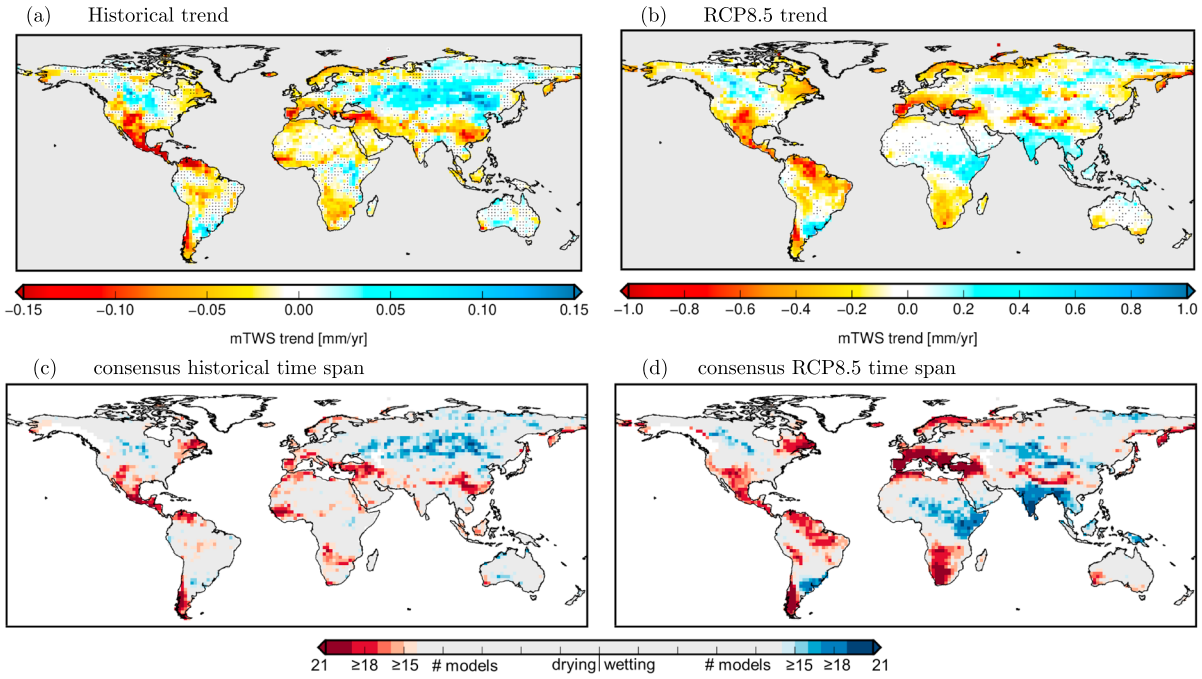

Figure 4. (top) Median of mTWS trend maps from 21 CMIP5 models for (a) historical (January 1861 to August 2016) and (b) RCP8.5 (September 2016 to December 2099) time span. Stippling indicates regions where the mean trend is not significantly different from zero ($\alpha = 0.05$). Please note the different color scales in (a) and (b). (bottom) Consensus maps for bicentennial mTWS trends from 21 CMIP5 models for (c) historical and (d) RCP8.5 time span.

For the first study we fit trends in a least squares sense for time spans of different lengths ranging from 14 to 100 years in steps of 5 years with the center year 2010. Examples for tendency maps obtained for the 50a, 30a, and 14a time periods are given in Figure 5. The median tendency maps for all 18 time spans are each correlated to the bicentennial median trend map (Figure 5a). For the 14a observation period the global correlation is only 23%, but with increasing time span it asymptotically approaches 100% (blue curve in Figure 5a). After the 30a time span the global correlation is 57% which is the same order of similarity that we find for the historical and RCP8.5 time spans (55%). Thus we conclude that around 30 years of TWS observations would be the minimum time to globally obtain a TWS trend comparable to long-term model results. However, even though the agreement between trend patterns might be low at 14a globally, this might not be the case locally, for instance, when only considering regions that exhibit strong model agreement. When calculating the correlation of the 14a tendency and the bicentennial trend only for grid cells with a model consensus of ≥71%, the correlation coefficient increases to 39% (red curve in Figure 5a), when additionally excluding nonsignificant grid cells, it increases to 52% (yellow curve in Figure 5a).

In this model study we evaluate the (global) spatial pattern correlation which only contains limited information about the agreement of trends for individual grid cells. This means that though this experiment brings out what to expect from the similarity of the spatial patterns, it does not provide the likelihood for a local mTWS tendency computed from a certain time span to actually match the bicentennial trend in that grid cell. As we are interested in regions where 14a GRACE TWS tendencies agree with bicentennial mTWS model trends and want to rate the results with respect to what to expect from this short time span, we perform a second model study: For each of the 21 CMIP5 models we cut 22 slices of 14a mTWS data with a distance of 5 years (centered around the year 1970) and estimate 22 14a tendencies. For each grid cell the fraction of tendencies that agree or disagree (in terms of sign) with the bicentennial mTWS trend from that particular model is calculated. Subsequently, the global mean of all fractions and all models is computed. This procedure is repeated for different time spans from 1 to 100 years in steps of 5 years (Figure 6). According to the models the probability that a 14a tendency is in agreement with the long-term trend is on average 53%, which is slightly better than random chance. Even after a century there is still a chance of 27% that an individual tendency does not match the bicentennial trend even though the global pattern correlation is already high with 87%. This indicates that there is natural variability in the models even over long time periods of 100 years and more, which Laepple and Huybers (2014) found to be caused by sea surface temperature

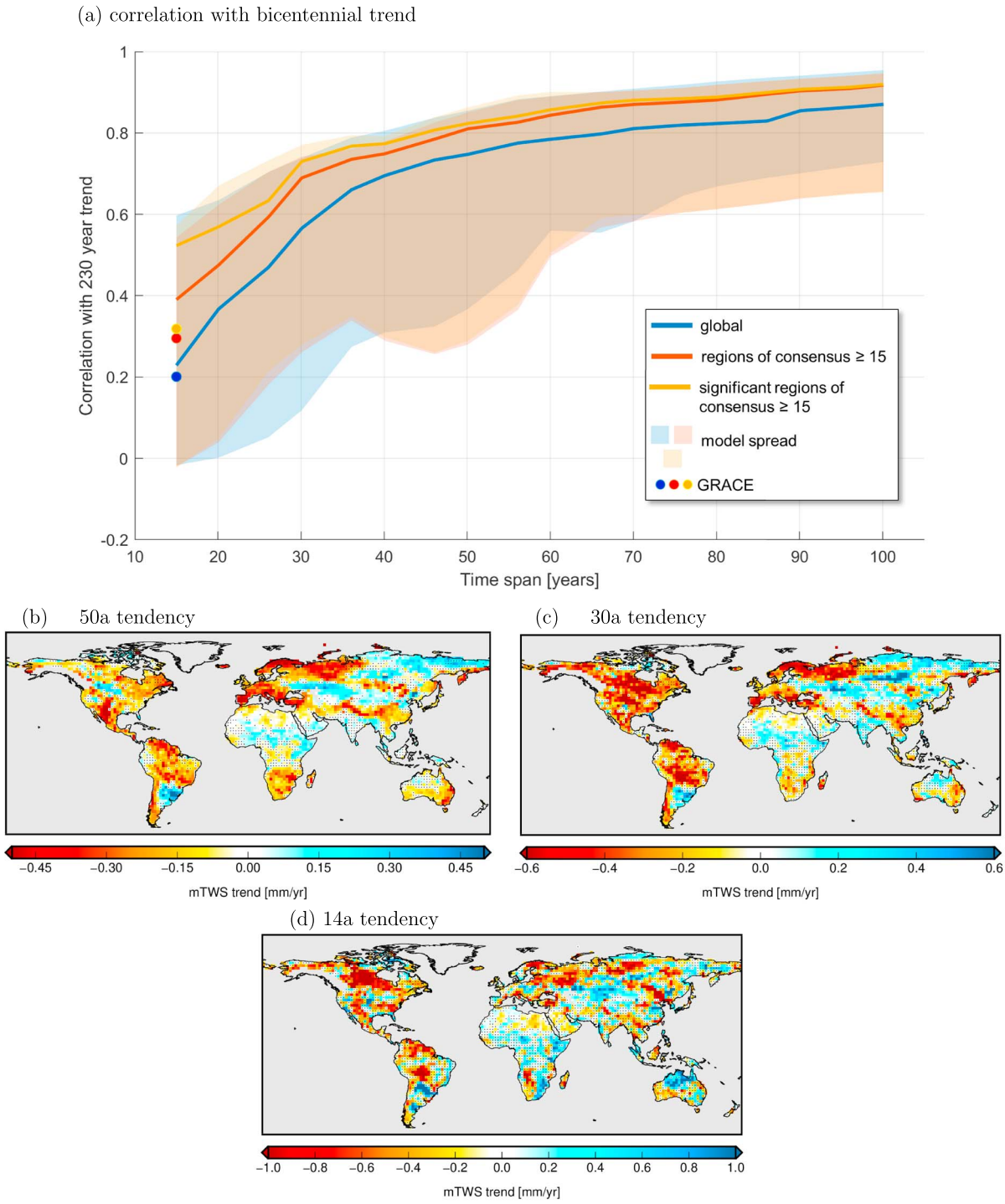

Figure 5. (a) Correlation of the mTWS tendency maps for different time spans (center year 2010) with the bicentennial mTWS trend map. (b–d) Median of mTWS tendency maps from 21 CMIP5 models for (b) 50a time span, (c) 30a time span, and (d) 14a time span. Each time span is centered around the year 2010. Note the different color scales due to larger variability for shorter time spans. Stippling indicates regions with nonsignificant trends ($\alpha = 0.05$).

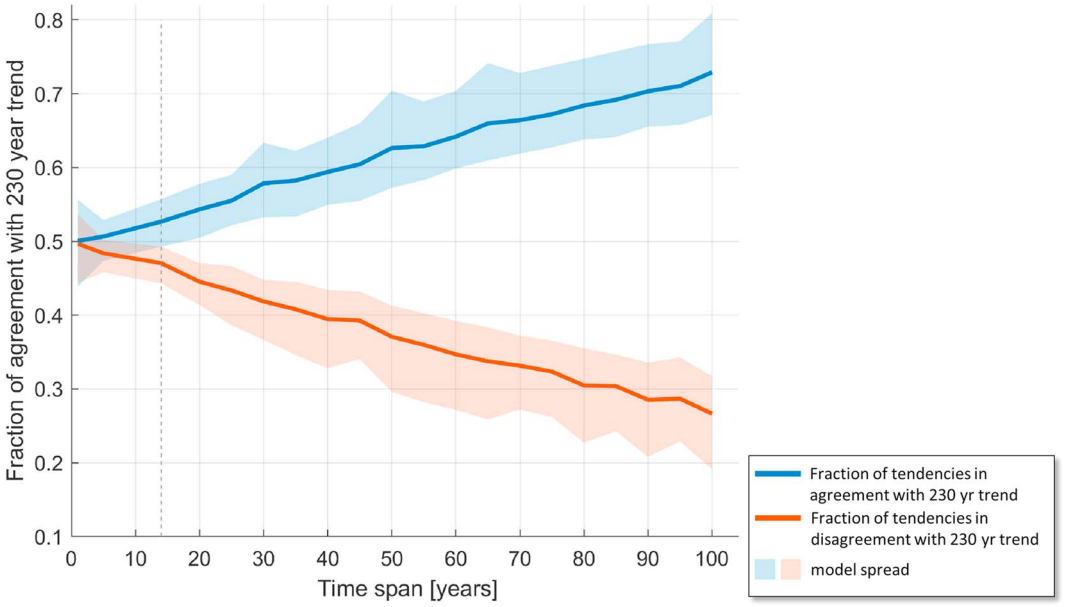

Figure 6. Proportion of grid cells from mTWS tendency maps for different time spans in agreement and disagreement with the direction of the bicentennial mTWS trend. Stippled line indicates 14 years.

variability on these time scales. However, when taking into account the model spread, we conclude that after about 14 years (stippled line in Figure 6) the two curves are just starting to be clearly distinguishable, that is, an agreement with the long-term trend can be significantly considered to being not just by chance any more. This is an indication that it is possible—at least locally—to distinguish long-term climate-driven TWS trends from interannual variations even in short observation time series and thus a comparison of 14a tendencies from GRACE to bicentennial model trends is reasonable already today.

5. Comparison of CMIP5 mTWS and GRACE TWS Trends

The two model studies in section 4 indicate that when comparing TWS tendencies from a 14a observation period to bicentennial mTWS trends we cannot expect too much agreement: a fairly low global pattern correlation and a probability of agreement only slightly above 50%. However, these model studies represent only the global mean and according to Figure 6 after 14 years we are starting to be able to clearly identify regions of long-term wetting and drying. We thus compare the TWS trends derived from GRACE observations for the 14a time period (Figure 1) to the bicentennial median mTWS trend (Figure 3a).

We note that the magnitude of the TWS trend from GRACE is substantially larger than the bicentennial median mTWS trend. One reason for this is that generally interannual TWS trends (as seen from 14 years of GRACE) exhibit a larger magnitude than bicentennial trends mostly unaffected by low-frequency climate variability. This can, for example, be seen from the trends of the median 14a tendency maps from the CMIP5 models (Figure 5d), which are already much larger than the median bicentennial trends. Individual models even exhibit a larger 14a tendency range than such maps where trend magnitudes are dampened due to the median calculation. In addition, TWS and mTWS do not represent the same physical entity everywhere as described in sections 2 and 3 and mTWS might have a lower amplitude as soil moisture depth is often limited. Recent results by Scanlon et al. (2018) show that also in off-line global hydrological models and land surface models TWS trends are generally of smaller magnitude compared to the observed GRACE trends. This indicates that even high-resolution uncoupled models driven with realistic atmospheric forcings still have difficulties to simulate realistic TWS trend magnitudes. We also note that agreement between hydrological models and GRACE could be improved by considering groundwater storage in the models (Pokhrel et al., 2013), which is entirely omitted in all CMIP5 models considered here.

As expected, the correlation between the global patterns of the 14a TWS tendency from GRACE and the bicentennial median mTWS trend is low (r^2 = 20%, blue dot in Figure 5a), but it largely meets the correlation expected from the model study where we compared the 14a median mTWS tendency to the bicentennial

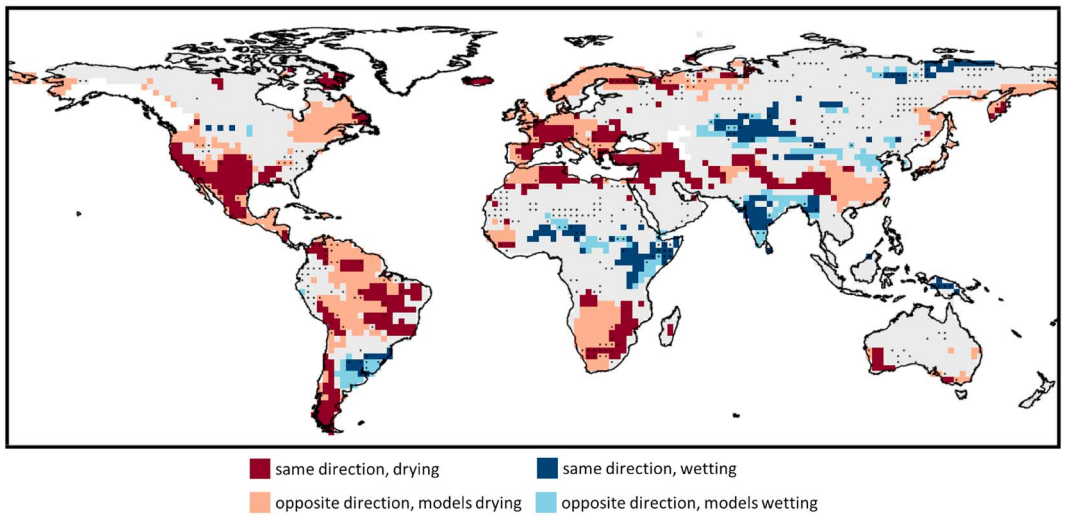

Figure 7. Regions where at least 71% of CMIP5 models show the same direction of bicentennial mTWS trend, distinguished into regions where the TWS trend from GRACE has the same or opposite direction. Stippling indicates regions with nonsignificant GRACE trends ($\alpha = 0.05$).

median mTWS trends ($r^2 = 23\%$, see section 4). When calculating the correlation only for grid cells with a model consensus of ≥71%, the correlation coefficient for GRACE increases to 30% (red dot in Figure 5a), which is slightly lower than expected from the model study ($r^2 = 39\%$) but largely within the model spread. After excluding nonsignificant GRACE trends the correlation further rises to 32%, which is, compared to the model study, in the middle of the model spread, but lower than the median ($r^2 = 52\%$). The reason for this is likely due to the different number of nonsignificant trend grid cells for GRACE and for the model median (compare Figure 1 and Figure 5d).

In Figure 7 areas are displayed in red where at least 71% (15 of 21) of the models show a negative bicentennial trend direction (Figure 3). A grid cell is marked in dark red if at the same location the trend from GRACE (Figure 1) exhibits a negative sign as well, or in light red if the trend from GRACE has the opposite (i.e., positive) sign. Analogously, positive high-consensus model trends are marked in blue; dark blue where GRACE has the same (positive) sign and light blue where GRACE has the opposite (negative) sign. Dark red or dark blue areas in Figure 7 indicate hot spot regions where trends in GRACE data may already be related to climate change signals because the majority of CMIP5 models supports this trend (at least regarding its sign) on the bicentennial time scale. According to Figure 6, hot spot regions of drying trends are mainly the region around the Mediterranean Sea, southwestern United States, the southern tip of South America, and the Himalaya region.

Few wetting trends are identified as hot spots, potentially in South America, Central Africa, Central and North Asia, and India. This is quite consistent with the bicentennial mTWS trend map (Figure 3a), which is dominated by negative trends (66% of trends are negative, 34% positive, global mean -0.04 mm/year, excluding Greenland, Alaska, and Antarctica). A reason for dominating drying trends in CMIP5 models might be that the models have a limited ability to capture anomalously high water storage. In most models that do not include groundwater, rivers, lakes, and wetlands, water that exceeds the soil moisture storage capacity is allocated to surface runoff and does not reenter the land surface scheme. In addition, an underestimation of persistence times of water in the soil might prevent the simulation of high accumulations of water. Thus, models tend to simulate drying trends more easily than wetting trends.

Besides the dark red and blue areas there are several regions of opposite signs between 14a GRACE TWS tendencies and bicentennial mTWS model trends (light areas in Figure 7). Major regions of disagreement are located in northeastern Canada, Fennoscandia, southeastern China, southern Africa, and central South America. The proportion of high-consensus areas agreeing with GRACE versus those disagreeing with GRACE is 49% to 51%. When not restricting to high-consensus regions but calculating the proportion globally it is 50% versus 50%. From the second model study in section 4 we obtained a higher percentage of agreement of about 53%, but this number is estimated by comparing each 14a model tendency map to the

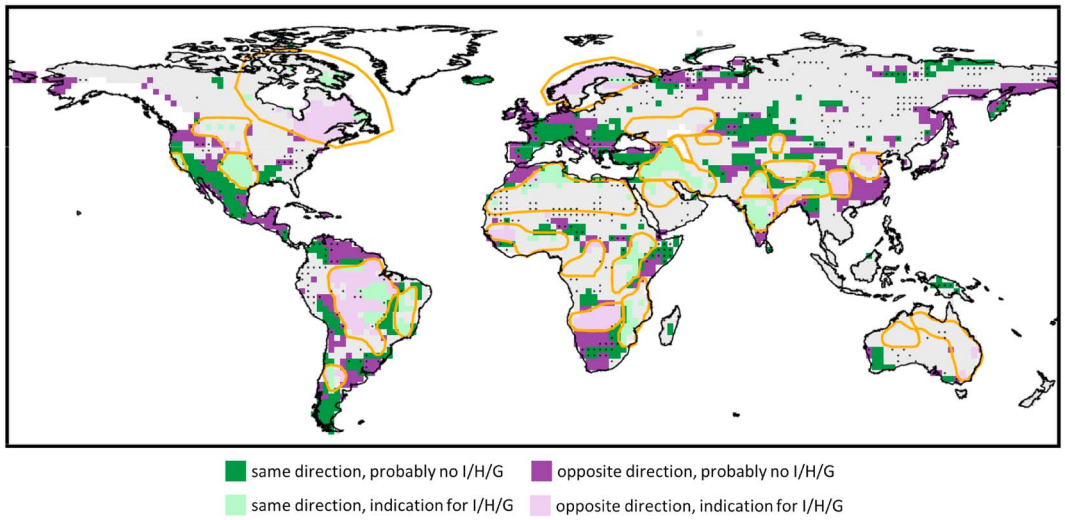

Figure 8. Regions where at least 71% of CMIP5 models show the same direction of bicentennial mTWS trend, distinguished into regions where the TWS trend from GRACE has the same (green color) or opposite (violet color) direction. Regions affected by interannual variability (I), human impact (H), or glacial isostatic adjustment (G) are outlined by orange polygons and shaded in light green/violet. Stippling indicates regions with nonsignificant GRACE trends ($\alpha = 0.05$).

bicentennial trend map from the same model. As we do not have a bicentennial reference for GRACE but compare to the model median instead, it is expected that the match is not at the same level. Furthermore, results from the model study might be too optimistic if modeled interannual variability is lower than in reality. In addition, the results from section 4 are not directly comparable to GRACE as human impact such as dam building and groundwater abstraction is not reflected in the model median which further affects the level of agreement. From the low level of agreement between modeled and observed TWS trends we conclude that in many regions the influence of interannual variations in the 14a time span is still dominant at this stage and thus the comparison to long-term modeled trends is affected by large uncertainties. However, the analysis of the existing hot spot and noncompliance regions in Figure 7 can give valuable information on potential climate-related wetting and drying as well as indications for possible model shortcomings and can be the focus of future investigations.

6. Uncertainty Analysis

Due to interannual variability in the short GRACE observation time span and human impacts that are not considered by the models, we cannot assume the same level of certainty for the results presented in section 5 in every region. In order to identify regions where climate-related wetting or drying trends may be overlaid by interannual variability (I) or human impact (H) we access a study by Rodell et al. (2018), who attribute the TWS trends from GRACE for April 2002 to March 2016 to their different dominating origins for 34 major basins. The regions where I or H is present in GRACE TWS trends according to Rodell et al. (2018) are outlined as orange polygons in Figure 8. Within these orange polygons the results from Figure 7 about agreement/disagreement have to be regarded as uncertain due to I/H effects. In addition, we marked regions affected by glacial isostatic adjustment (GIA; G) as regions possibly not comparable with mTWS from ESMs as residual GIA effects in GRACE might remain in the trend even after removing a state-of-the-art GIA model. There might be other regions affected by I/H/G that are wrongly not considered as uncertain because they are not accounted for by Rodell et al. (2018) who focused on the 34 study regions with the most prominent trend signals in GRACE. Furthermore, there might be additional regions where mTWS trends from CMIP5 models are systematically wrong—for example, due to missing groundwater and deeper soil layers, but these regions are also not explicitly identified and marked as uncertain here.

By means of the orange polygons in Figure 8 we can distinguish the hot spot and noncompliance regions (see Figure 7) into four categories:

1. Same direction of model majority and GRACE trend, and no indication for I/H/G (dark green). In these regions we assume the climate-related wetting or drying simulated by the models to be confirmed by observations.
2. Same direction, but indication for I/H/G (light green). Here we do not know at this time if a possible underlying climatic trend has the same or opposite direction as modeled trends, thus we mark these regions as uncertain. By extending the observation record it might turn out in the future that the current match in some of these regions is only chance due to I/H/G effects counteracting the long-term trend.
3. Opposite direction, and no indication for I/H/G (dark violet). In these regions we have no reason to assume that the direction of the 14a tendency from observations differs from the real long-term climate trend, thus there might be deficits in CMIP5 models leading to a mismatch. Of course, also other I/H/G effects that were not yet identified might be the reason for noncompliance.
4. Opposite direction, but indication for I/H/G (light violet). As in Category 2, we do not know the direction of a possible underlying climatic trend, thus consider these regions as uncertain. The current mismatch might be only chance due to I/H/G effects dominating the 14a GRACE trends and in the long term these regions might turn out to be agreement regions.

In Category 1, where observations already today confirm what a majority of models predict for the long-term, mainly three larger regions remain: drying conditions in southwestern United States/Mexico and around the Mediterranean Sea (central southern Europe and Turkey), and wetting conditions in Central Asia (indicated by a cluster of dark green grid cells). These regions we regard as hot spot regions of drying and wetting that can already today be attributed to climate change with the help of GRACE.

Category 2 indicates regions where potential climate signals are overlaid by I/H/G effects that might be the cause for the current agreement. However, in some regions it is quite likely that the long-term trend has the same direction as the I/H/G effects, for example, the light green spots in central southern United States and the Middle East (i.e., Syria, Iraq, and Iran), where relatively large connected regions of agreement are adjacent to Category 1 regions. According to Rodell et al. (2018) in both regions the large negative TWS trends seem to be due to a combination of drought and enhanced groundwater depletion. Additionally, in the Middle East dam building in Turkey further intensifies the lack of water (Voss et al., 2013). Also at parts of the Mediterranean coast in northern Africa models and GRACE see a common drying that might be connected to the long-term drying conditions in southern Europe, even though the currently observed trend is assumed to be dominated by groundwater abstraction (Döll et al., 2014). Another location where I/H/G effects might enhance a common long-term trend is the light green spot in India. Here, possibly climate-related precipitation increase is overlaid by human impact in form of a groundwater policy change that is reflected in overall wetting conditions (Bhanja et al., 2017).

In Category 3, a mismatch of model majority and GRACE is unlikely to be due to I/H/G, thus these regions indicate possible systematical errors in CMIP5 models. For example, CMIP5 models were found to underestimate summer precipitation over southeastern China (Chen & Frauenfeld, 2014), which might explain the negative soil moisture trend not supported by GRACE. Furthermore, systematic rainfall biases in CMIP5 models over southern Africa were identified by Munday and Washington (2018), another possible reason for noncompliance with GRACE. However, this needs to be further investigated, as according to Munday and Washington (2018) rainfall is largely overestimated whereas soil moisture is simulated to decrease. Large connected regions of noncompliance in western and northeastern Europe are likely affected by interannual variations. As the apparent positive TWS trends seen by GRACE are very small here (see Figure 1), hardly significant, and dominated by the annual cycle they were not accounted for by Rodell et al. (2018). The same holds for a dark violet region in northern South America.

Category 4 denotes regions affected by I/H/G that are not in agreement with long-term predicted trends. We consider these regions as uncertain as possible climate signals masked by I/H/G might actually agree with model trends. Large Category 4 areas are the GIA-affected regions Fennoscandia and northeastern Canada where the uncertainty from the applied GIA model might be dominating the GRACE trend. In the Amazon region in central South America a recovery from a drought in the early GRACE period is responsible for an overall wetting trend in GRACE. However, the mismatch could also result from model deficits due to missing groundwater. Considering groundwater buffering effects in CMIP5 models causes a shift in the evapotranspiration regime resulting in much less drying trends in the Amazon region (Pokhrel et al., 2014). In southern Africa the Category 4 region is associated with interannual variability, a progression from dry

to wet during the GRACE period. The light violet spot in southeastern China can be connected to the Three Gorges Dam reservoir filling that models do not capture. However, it is not clear to what extent also model deficits in southeastern China and southern Africa are responsible for this mismatch as these are adjacent to Category 3 regions.

7. Conclusions

We analyzed in this study long-term trends in TWS derived from a selection of 21 coupled climate models stored in the CMIP5 archive and compared them to satellite observed TWS trends from 14 years (April 2002 to August 2016) of GRACE data. We found a large disagreement among the bicentennial (January 1861 to December 2099) TWS trends (sum of mrso and snw) from different models: the mean correlation among individual models is only 10%, which reflects the still high uncertainties in TWS variability simulated by present-day global coupled climate models. While a significant part of intermodel variation might result from the different atmosphere components, differences in land surface parameterization, particularly the soil moisture and snow storage capacities, are likely to add to this. We nevertheless identified several regions of high model consensus regarding the direction of the trend on which we focused for the comparison with GRACE-derived TWS. Furthermore, we found a large agreement between modeled TWS trends for the historical (January 1861 to August 2016) and the RCP8.5 (September 2016 to December 2099) time span, indicating that long-term TWS trends are already emerging in present time and thus can be expected to be contained in observational records available today.

By computing model TWS tendencies for differently long time spans (from 14 to 100 years) and comparing them to the bicentennial trend map, we assessed the influence of interannual variations on the degree of agreement between long-term and short-term trends. For the global pattern correlation we concluded that a time span of 30 years or more would be sufficient to globally distinguish interannual climate variability in TWS from long-term climate trends. However, for even 14 years we obtained a global correlation of 23% with the bicentennial TWS trend, which regionally is substantially higher, for instance, when limiting to significant trends in regions of strong model consensus only (52%). When estimating the fraction of grid cells with the same direction of the trend for the bicentennial and different shorter time spans we found that after 14 years the proportion is 53% agreement versus 47% disagreement, and even after a century there is still a substantial probability of disagreement (27%). We identified 14 years as the minimum observation time span required to distinguish long-term trends from interannual variations, with a (global) probability that is better than just chance.

By comparing the bicentennial modeled median TWS trend map against the TWS trend map obtained from GRACE data over the period April 2002 to August 2016, we found a similar global correlation (20%) as we expected from the model study. Focusing on regions of strong model consensus, we identified hot spot regions where modeled TWS trends have the same direction as the GRACE TWS trend. Drying hot spots were mainly found in the region around the Mediterranean Sea, the west coast of North America, the southern tip of South America, and the Himalaya region. Wetting hot spots are sparse and only found for small areas in South America, Central Africa, Central and North Asia, and India. The largest regions of noncompliance between trends from models and GRACE are located in northeastern Canada, Fennoscandia, southeastern China, southern Africa, and central South America. In total, the proportion of regions in agreement versus regions in disagreement in high model consensus regions is 49% versus 51%, which indicates that the 14-year GRACE time series is in many regions still dominated by interannual variations.

We further investigated the existing hot spot and noncompliance regions regarding possible natural interannual variability or human impact in order to classify the results regionally. In this classification, 36% of the high-consensus area was marked as uncertain and thus remains undetermined at this time if and how much it is affected by climate change. In the other 64% of the high-consensus area the climate signal was assumed to dominate the TWS trends. Within these 64% the proportion of agreement versus disagreement areas is 49% versus 51%, which is the same proportion as in the entire high-consensus area without uncertainty assessment. From this we conclude, that in half of the area marked as certain either model deficits or unidentified interannual variability in observations cause disagreement. This is consistent with our findings from the model study that revealed a large influence of interannual variations on TWS trends even after long time spans.

Nevertheless, from the GRACE record available today, the classification leads to the identification of hot spot regions in southern United States/Mexico, central southern Europe, and central Asia, where GRACE confirms modeled long-term drying or wetting trends already today. In some regions, for example, in southeastern China, the Amazon region, and in southern Africa the results hint at possible model deficits, for example, due to missing groundwater modeling. Furthermore, we identified regions where mismatch of models and observations might be due to interannual variations or other effects (e.g., dam building, groundwater withdrawal, glacial isostatic adjustment) in GRACE that are not included in the CMIP5 models.

As the time series of GRACE data will soon be extended with new observations from the GRACE-FO mission launched in May 2018, observed TWS trends are expected to become even more representative for the long-term climate response. At the same time, climate model outputs from the upcoming CMIP6 will include new model versions with altered representations of soil moisture-related variables, which might either improve or degrade the model consensus depending on the impact of new parameterizations and the effect of a higher degree of model freedom and complexity. Comparisons between observed and modeled TWS trends as presented in this paper will thus remain of high relevance for future climate model assessments.

Acknowledgments
We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Figure 2 of this paper) for producing and making available their model output (e.g., at "https://esgf-data.dkrz.de/search/cmip5-dkrz/"). For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. We are grateful to Torsten Mayer-Gürr and his group at TU Graz, Austria, for providing the ITSG-Grace2018s data at this website (https://www.tugraz.at/institute/ifg/downloads/gravity-field-models/itsg-grace2018/). The degrees 1 and 2 harmonic coefficients used during GRACE processing were received from this site (http://grace.jpl.nasa.gov).

References

A, G., Wahr, J., & Zhong, S. (2013). Computations of the viscoelastic response of a 3-D compressible Earth to surface loading: An application to Glacial Isostatic Adjustment in Antarctica and Canada. *Geophysical Journal International*, *192*(2), 557–572. https://doi.org/10.1093/gji/ggs030

Berg, A., Sheffield, J., & Milly, P. C. D. (2017). Divergent surface and total soil moisture projections under global warming. *Geophysical Research Letters*, *44*, 236–244. https://doi.org/10.1002/2016GL071921

Bhanja, S. N., Mukherjee, A., Rodell, M., Wada, Y., Chattopadhyay, S., Velicogna, I., et al. (2017). Groundwater rejuvenation in parts of India influenced by water-policy change implementation. *Scientific Reports*, *7*(1), 7453. https://doi.org/10.1038/s41598-017-07058-2

Bhattacharyya, S., & Narasimha, R. (2005). Possible association between Indian monsoon rainfall and solar activity. *Geophysical Research Letters*, *32*, L05813. https://doi.org/10.1029/2004GL021044

Brutel-Vuilmet, C., Ménégoz, M., & Krinner, G. (2013). An analysis of present and future seasonal Northern Hemisphere land snow cover simulated by CMIP5 coupled climate models. *The Cryosphere*, *7*(1), 67–80. https://doi.org/10.5194/tc-7-67-2013

Caron, L., Ivins, E. R., Larour, E., Adhikari, S., Nilsson, J., & Blewitt, G. (2018). GIA model statistics for GRACE hydrology, cryosphere, and ocean science. *Geophysical Research Letters*, *45*, 2203–2212. https://doi.org/10.1002/2017GL076644

Chen, L., & Frauenfeld, O. W. (2014). A comprehensive evaluation of precipitation simulations over China based on CMIP5 multimodel ensemble projections. *Journal of Geophysical Research: Atmospheres*, *119*, 5767–5786. https://doi.org/10.1002/2013JD021190

Cheng, M., Tapley, B. D., & Ries, J. C. (2013). Deceleration in the Earth's oblateness. *Journal of Geophysical Research: Solid Earth*, *118*, 740–747. https://doi.org/10.1002/jgrb.50058

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., et al. (2013). Long-term climate change: Projections, commitments and irreversibility. In T. F. Stocker (Ed.), *Climate change 2013: The physical science basis. Contribution of working group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 1029–1136). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.024

Committee on the Decadal Survey for Earth Science and Applications from Space, Space Studies Board, Division on Engineering and Physical Sciences, & National Academies of Sciences, Engineering, and Medicine (2018). *Thriving on our changing planet: A decadal strategy for earth observation from space*. Washington, DC: National Academies Press. https://doi.org/10.17226/24938

Dirmeyer, P. A., Jin, Y., Singh, B., & Yan, X. (2013). Trends in land-atmosphere interactions from CMIP5 simulations. *Journal of Hydrometeorology*, *14*(3), 829–849. https://doi.org/10.1175/JHM-D-12-0107.1

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*, 5698–5720. https://doi.org/10.1002/2014WR015595

Dorigo, W. A., Gruber, A., De Jeu, R. A. M., Wagner, W., Stacke, T., Loew, A., et al. (2015). Evaluation of the ESA CCI soil moisture product using ground-based observations. *Remote Sensing of Environment*, *162*, 380–395. https://doi.org/10.1016/j.rse.2014.07.023

Eicker, A., Forootan, E., Springer, A., Longuevergne, L., & Kusche, J. (2016). Does GRACE see the terrestrial water cycle "intensifying"? *Journal of Geophysical Research: Atmospheres*, *121*, 733–745. https://doi.org/10.1002/2015JD023808

Eicker, A., Schumacher, M., Kusche, J., Döll, P., & Schmied, H. M. (2014). Calibration/data assimilation approach for integrating GRACE data into the WaterGAP Global Hydrology Model (WGHM) using an ensemble Kalman filter: First results. *Surveys in Geophysics*, *35*, 1285–1309. https://doi.org/10.1007/s10712-014-9309-8

Fagiolini, E., Flechtner, F., Horwath, M., & Dobslaw, H. (2015). Correction of inconsistencies in ECMWF's operational analysis data during de-aliasing of GRACE gravity models. *Geophysical Journal International*, *202*(3), 2150–2158. https://doi.org/10.1093/gji/ggv276

Famiglietti, J. S., & Rodell, M. (2013). Water in the balance. *Science*, *340*(6138), 1300–1301. https://doi.org/10.1126/science.1236460

Fasullo, J. T., Lawrence, D. M., & Swenson, S. C. (2016). Are GRACE-era terrestrial water trends driven by anthropogenic climate change? *Advances in Meteorology*, *2016*, 4830603. https://doi.org/10.1155/2016/4830603

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., et al. (2013). Evaluation of climate models, *Climate change 2013: The physical science basis. Contribution of working group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 741–882). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.020

Flechtner, F., Neumayer, K.-H., Dahle, C., Dobslaw, H., Fagiolini, E., Raimondo, J.-C., & Güntner, A. (2016). What can be expected from the GRACE-FO laser ranging interferometer for earth science applications? *Surveys in Geophysics*, *37*(2), 453–470. https://doi.org/10.1007/s10712-015-9338-y

Freedman, F. R., Pitts, K. L., & Bridger, A. F. C. (2014). Evaluation of CMIP climate model hydrological output for the Mississippi River basin using GRACE satellite observations. *Journal of Hydrology*, *519*, 3566–3577. https://doi.org/10.1016/j.jhydrol.2014.10.036

Gray, L. J., Anstey, J. A., Kawatani, Y., Lu, H., Osprey, S., & Schenzinger, V. (2018). Surface impacts of the Quasi Biennial Oscillation. *Atmospheric Chemistry and Physics*, *18*(11), 8227–8247. https://doi.org/10.5194/acp-18-8227-2018

Güntner, A. (2008). Improvement of global hydrological models using GRACE data. *Surveys in Geophysics*, *29*(4), 375–397. https://doi.org/10.1007/s10712-008-9038-y

Guo, Z., & Dirmeyer, P. A. (2006). Evaluation of the second global soil wetness project soil moisture simulations: 1. Intermodel comparison. *Journal of Geophysical Research*, *111*, D22S02. https://doi.org/10.1029/2006JD007233

Han, S.-C., Sauber, J., & Luthcke, S. (2010). Regional gravity decrease after the 2010 Maule (Chile) earthquake indicates large-scale mass redistribution. *Geophysical Research Letters*, *37*, L23307. https://doi.org/10.1029/2010GL045449

Han, S.-C., Sauber, J., Luthcke, S. B., Ji, C., & Pollitz, F. F. (2008). Implications of postseismic gravity change following the great 2004 Sumatra-Andaman earthquake from the regional harmonic analysis of GRACE intersatellite tracking data. *Journal of Geophysical Research*, *113*, B11413. https://doi.org/10.1029/2008JB005705

Hirschi, M., & Seneviratne, S. I. (2017). Basin-scale water-balance dataset (BSWB): An update. *Earth System Science Data*, *9*(1), 251–258. https://doi.org/10.5194/essd-9-251-2017

Huang, Y., Gerber, S., Huang, T., & Lichstein, J. W. (2016). Evaluating the drought response of CMIP5 models using global gross primary productivity, leaf area, precipitation, and soil moisture data. *Global Biogeochemical Cycles*, *30*, 1827–1846. https://doi.org/10.1002/2016GB005480

Iles, C. E., Hegerl, G. C., Schurer, A. P., & Zhang, X. (2013). The effect of volcanic eruptions on global precipitation. *Journal of Geophysical Research: Atmospheres*, *118*, 8770–8786. https://doi.org/10.1002/jgrd.50678

Knutti, R., Masson, D., & Gettelman, A. (2013). Climate model genealogy: Generation CMIP5 and how we got there. *Geophysical Research Letters*, *40*, 1194–1199. https://doi.org/10.1002/grl.50256

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

Laepple, T., & Huybers, P. (2014). Ocean surface temperature variability: Large model-data differences at decadal and longer periods. *Proceedings of the National Academy of Sciences*, *111*(47), 16,682–16,687. https://doi.org/10.1073/pnas.1412077111

Lambeck, K. (1988). *Geophysical geodesy: The slow deformations of the earth*. Oxford [Oxfordshire], New York: Clarendon Press: Oxford University Press. https://trove.nla.gov.au/version/21392411

Liepert, B. G., & Lo, F. (2013). CMIP5 update of 'Inter-model variability and biases of the global water cycle in CMIP3 coupled climate models'. *Environmental Research Letters*, *8*(2), 29401. https://doi.org/10.1088/1748-9326/8/2/029401

Mayer-Gürr, T., Behzadpour, S., Ellmer, M., Kvas, A., Klinger, B., & Zehentner, N. (2018). ITSG-Grace2018—Monthly, daily and static gravity field solutions from GRACE. GFZ Data Services, https://doi.org/10.5880/ICGEM.2018.003

Munday, C., & Washington, R. (2018). Systematic climate model rainfall biases over Southern Africa: Links to moisture circulation and topography. *Journal of Climate*, *31*(18), 7533–7548. https://doi.org/10.1175/JCLI-D-18-0008.1

Ni, S., Chen, J., Wilson, C. R., Li, J., Hu, X., & Fu, R. (2018). Global terrestrial water storage changes and connections to ENSO events. *Surveys in Geophysics*, *39*(1), 1–22. https://doi.org/10.1007/s10712-017-9421-7

Phillips, T., Nerem, R. S., Fox-Kemper, B., Famiglietti, J. S., & Rajagopalan, B. (2012). The influence of ENSO on global terrestrial water storage using GRACE. *Geophysical Research Letters*, *39*, L16705. https://doi.org/10.1029/2012GL052495

Pokhrel, Y. N., Fan, Y., & Miguez-Macho, G. (2014). Potential hydrologic changes in the Amazon by the end of the 21st century and the groundwater buffer. *Environmental Research Letters*, *9*(8), 84004. https://doi.org/10.1088/1748-9326/9/8/084004

Pokhrel, Y. N., Fan, Y., Miguez-Macho, G., Yeh, PatJ.-F., & Han, S.-C. (2013). The role of groundwater in the Amazon water cycle: 3. Influence on terrestrial water storage computations and comparison with GRACE. *Journal of Geophysical Research: Atmospheres*, *118*, 3233–3244. https://doi.org/10.1002/jgrd.50335

Rodell, M., Famiglietti, J. S., Chen, J., Seneviratne, S. I., Viterbo, P., Holl, S., & Wilson, C. R. (2004). Basin scale estimates of evapotranspiration using GRACE and other observations. *Geophysical Research Letters*, *31*, L20504. https://doi.org/10.1029/2004GL020873

Rodell, M., Famiglietti, J. S., Wiese, D. N., Reager, J. T., Beaudoing, H. K., Landerer, F. W., & Lo, M.-H. (2018). Emerging trends in global freshwater availability. *Nature*, *557*(7707), 651. https://doi.org/10.1038/s41586-018-0123-1

Scanlon, B. R., Zhang, Z., Save, H., Sun, A. Y., Müller Schmied, H., van Beek, L. P. H., et al. (2018). Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data. *Proceedings of the National Academy of Sciences*, *115*(6), E1080–E1089. https://doi.org/10.1073/pnas.1704665115

Swenson, S., Chambers, D., & Wahr, J. (2008). Estimating geocenter variations from a combination of GRACE and ocean model output. *Journal of Geophysical Research*, *113*, B08410. https://doi.org/10.1029/2007JB005338

Syed, T. H., Famiglietti, J. S., Rodell, M., Chen, J., & Wilson, C. R. (2008). Analysis of terrestrial water storage changes from GRACE and GLDAS. *Water Resources Research*, *44*, W02433. https://doi.org/10.1029/2006WR005779

Tapley, B. D., Bettadpur, S., Watkins, M., & Reigber, C. (2004). The gravity recovery and climate experiment: Mission overview and early results. *Geophysical Research Letters*, *31*, L09607. https://doi.org/10.1029/2004GL019920

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2011). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

Trenberth, K. E. (Ed.) (2010). *Climate system modeling*. Cambridge: Cambridge University Press. OCLC: 845631695.

Voss, K. A., Famiglietti, J. S., Lo, M., Linage, C. d, Rodell, M., & Swenson, S. C. (2013). Groundwater depletion in the Middle East from GRACE with implications for transboundary water management in the Tigris-Euphrates-Western Iran region. *Water Resources Research*, *49*, 904–914. https://doi.org/10.1002/wrcr.20078

Wartenburger, R., Seneviratne, S. I., Hirschi, M., Chang, J., Ciais, P., Deryng, D., et al. (2018). Evapotranspiration simulations in ISIMIP2a—Evaluation of spatio-temporal characteristics with a comprehensive ensemble of independent datasets. *Environmental Research Letters*, *13*, 75001. https://doi.org/10.1088/1748-9326/aac4bb

Yuan, S., & Quiring, S. M. (2017). Evaluation of soil moisture in CMIP5 simulations over the contiguous United States using in situ and satellite observations. *Hydrology and Earth System Sciences*, *21*(4), 2203–2218. https://doi.org/10.5194/hess-21-2203-2017

Zhang, L., Dobslaw, H., Stacke, T., Güntner, A., Dill, R., & Thomas, M. (2017). Validation of terrestrial water storage variations as simulated by different global numerical models with GRACE satellite observations. *Hydrology and Earth System Sciences*, *21*(2), 821–837. https://doi.org/10.5194/hess-21-821-2017