Long-term, non-anthropogenic groundwater storage changes simulated by three global-scale hydrological models

*Bailing Li [1,2], Matthew Rodell[2], Justin Sheffield[3,4], Eric Wood[3] and Edwin Sutanudjaja[5]

[1] ESSIC University of Maryland

[2] NASA Goddard Space Flight Center

[3] Princeton University

[4]University of Southampton

[5]Utrecht University

Abstract

This study examined long-term, natural (i.e., excluding anthropogenic impacts) variability of groundwater storage worldwide. Groundwater storage changes were estimated by forcing three global-scale hydrological models with three 50+ year meteorological datasets. Evaluation using in situ groundwater observations from the U.S. and terrestrial water storage derived from the Gravity Recovery and Climate Experiment (GRACE) satellites showed that these models reasonably represented inter-annual variability of water storage, as indicated by correlations greater than 0.5 in most regions. Empirical orthogonal function analysis revealed influences of the El Niño Southern Oscillation (ENSO) on global groundwater storage. Simulated groundwater storage, including its global average, exhibited trends generally consistent with that of precipitation. Global total (natural) groundwater storage decreased over the past 5-7 decades with modeled rates ranging from 0.01 to 2.18 mm year^{-1}. This large range can be attributed in part to groundwater's low frequency (inter-decadal) variability, which complicates identification of real long-term trends even within a 50+ year time series. Results indicate that non-anthropogenic variability in groundwater storage is substantial, making knowledge of it fundamental to quantifying direct human impacts on groundwater storage.

1. Introduction

Groundwater is crucial for meeting agricultural, industrial and municipal water needs, especially in arid, semi-arid and drought impacted regions where other types of fresh water are scarce[1]. Groundwater also sustains streams and rivers during dry periods through its contributions to baseflow[2]. Nevertheless, groundwater variations and their relationship to climate change are poorly understood at the global-scale due to the unavailability of observational data[3,4]. A few studies based on in situ observations have revealed the influence of large-scale climate signals such as the El Niño Southern Oscillation (ENSO) and Pacific Decadal Oscillation (PDO) on aquifers in the U.S.[5,6,7,8] and Canadian Prairies[9]. Outside of North America similar studies are lacking because groundwater measurements typically are inaccessible because their distribution is restricted[10]. Further complicating matters, in situ observations are often unsuitable for climate studies because of short and discontinuous records, because they are made in confined aquifers where groundwater heads are controlled by factors other than climate or are affected by groundwater pumping[11]. The lack of knowledge about groundwater response to climate change has been noted in the 5th assessment report (AR5) of Intergovermental Panel on Climate Change[12] (IPCC).

It has been suggested that groundwater responses to future climate scenarios, at the global scale, will vary spatially, mainly depending on recharge rates, with arid and semi-arid areas more vulnerable than wetter areas such as the northern high latitudes, where recharge may increase[4]. Regional-scale studies support these predictions. A study found that projected changes in precipitation and temperature did not have significant impacts on groundwater recharge in the Grand Forks aquifer of British Columbia, Canada[13]. Another study showed that

47 groundwater recharge in the drier and warmer part of the High Plains Aquifer exhibited higher
48 sensitivity to rainfall changes than in the cool and wetter part of the aquifer and that the rate of
49 changes in groundwater recharge amplified that of changes in precipitation[14]. Similarly, it was
50 predicted that climate change would reduce recharge in the southern and mountainous parts of
51 the western U.S., while recharge would be largely unchanged in the northwestern U.S.[15] These
52 studies suggest that land evapotranspiration (ET), in addition to precipitation, may be an
53 important factor controlling groundwater responses to climate change. Nevertheless, they are
54 mainly focused on precipitation and temperature as the drivers of groundwater change,
55 without considering other factors such as wind and air pressure for which changes are difficult
56 to generalize due to internal variability and regional variation[16]. As a slower component of the
57 hydrological cycle, groundwater may reflect small but persistent changes in the climate
58 system[8]. To better predict the future, we need to improve our understanding on how
59 groundwater has responded to changes that have occurred in the climate system over the past
60 several decades. Such studies are noticeably lacking and limited to small spatial[17] or short
61 temporal scales[18,19].

62 In this study, we investigate variations in global groundwater storage simulated by three global-
63 scale hydrological models, the Catchment land surface model[22] (CLSM), the WaterGAP[21] and
64 PCRaster Global Water Balance[23] (PCR-GLOBWB) water resource models. CLSM is a physics
65 based model that was developed for coupled land and atmospheric modeling including
66 seasonal weather forecasting at NASA Goddard Space Flight Center with an emphasis on land-
67 atmosphere fluxes, which are constrained by both energy and water balances. WaterGAP and
68 PCR-GLOBWB are two leading models developed for assessing the spatial distribution and

69 temporal variability of water resources around the world. They rely on empirical relations to
70 compute all water budget components including surface water and human water use which are
71 neglected by most land surface models including CLSM.

72 Forced with multiple decades of meteorological data that reflect changes in the climate system,
73 these models are able to generate spatially and temporally continuous groundwater estimates
74 suitable for studying climate change impacts on groundwater at regional to global
75 scales[24,21,25,26]. Multi-model analysis also helps to reduce uncertainties associated with model
76 physics and meteorological forcing fields, as trends in terrestrial water storage simulated by
77 different models have been shown to vary considerably[27]. Modeled groundwater storage was
78 evaluated using long-records of in situ data in the U.S. and modeled terrestrial water storage
79 (TWS) was evaluated using TWS derived from Gravity Recovery and Climate Experiment
80 (GRACE) satellites.

81 Because CLSM does not simulate human impacts such as groundwater abstraction, this study
82 focuses on the temporal variability of natural (excluding anthropogenic impacts) groundwater
83 storage changes associated with atmospheric effects (precipitation and ET). However,
84 groundwater output from WaterGAP and PCR-GLOBWB with human water use is also analyzed
85 in order to assess the uncertainty associated with evaluation using in situ and satellite data
86 which may be affected by groundwater abstractions and to provide insight on the relative
87 contributions of natural and anthropogenic groundwater variability to the total groundwater
88 variability as groundwater depletion caused by abstraction has gained wide attention[20,21,1]

89 **2. Data and method**

2.1 Models

CLSM, which was developed at NASA Goddard Space Flight Center, simulates water storage changes in the surface layer (0 to 2 cm below the surface), the root zone (0 to 100 cm), and the full soil profile, and in three snow layers[22]. Water transfers among these stores are governed by empirical parameters derived based on topography, soil properties and average moisture transfer rates within different profiles. Profile depths are determined by CLSM parameter, bedrock depth, which varies spatially from 3 to 6 m and has direct impacts on the dynamics of simulated TWS and groundwater storage, with deeper bedrock generally producing slower groundwater responses to atmospheric drivers (precipitation and ET) [28]. The bedrock depth was increased worldwide by 2 m in this study to better simulate the full dynamic range of terrestrial water storage, particularly during severe drought, following previous studies [28,29,30]. No other parameters were adjusted or calibrated for this study. CLSM does not explicitly model water table variations but groundwater storage can be computed by subtracting root zone water storage from total soil profile storage.

CLSM accounts for sub-grid scale heterogeneity in soil wetness, including the fractions of the land surface where the soil is at or below the wilting point and transpiration has ceased, where the soil is unsaturated and above the wilting point and where it is saturated, to better simulate ET and runoff. Modeled state and flux updates occur at a 20 min time step to accommodate their highly non-linear relationships. CLSM simulated ET has been shown comparable with other land surface models in terms of capturing mean annual ET and its spatial patterns[31,32]. CLSM was run at 0.25° spatial resolution for this study.

111 The WaterGAP model (version 2), which was developed at the Centre for Environmental
112 Systems Research of the University of Kassel, Germany, simulates water storage changes in the
113 soil (with a single layer representing effective root depths depending on landcover type),
114 groundwater, snow and surface water[21]. Water exchanges among these states and with the
115 atmosphere are represented by simple empirical relations that facilitate model calibration and
116 tuning. This version of WaterGAP was calibrated using mean annual streamflow data from
117 more than 1,300 river basins covering more than 50% of the global land area. Outside of these
118 basins, model parameters were estimated through regionalization. Benefiting from the simple
119 (more linearized) model physics, the model was run on larger time steps (daily, compared to 20
120 min of CLSM) and coarser spatial resolution, 0.5°.

121 WaterGAP includes options to simulate anthropogenic effects such as reservoir operation and
122 withdrawals from surface waters and groundwater. Water demands from all sectors (irrigation,
123 livestock, industry and households) and net withdrawals are calculated based on the water use
124 efficiency of each country or county, population development, and industrial and energy
125 output. These are then input to the underlying hydrological model, along with atmospheric
126 forcing data (precipitation, temperature and potential ET). Return flows are calculated as the
127 difference between water inputs and net withdrawals. Anthropogenic groundwater storage
128 estimates used in this study are based on an assumption of 70% irrigation efficiency, which is
129 the ratio of consumptive water use to abstractions[21].

130 PCR-GLOBWB has a model structure similar to that of WaterGAP. It also employs relatively
131 simple empirical relationships to partition precipitation into surface runoff and infiltration and
132 to simulate water transfers among different water stores, including surface water[23]. The model

was not calibrated but its parameterizations were evaluated using GRDC runoff data. Sub-grid heterogeneity on land cover is also considered for calculating fluxes. PCR-GLOBWB can simulate anthropogenic effects but takes a different approach to estimating consumptive irrigation water use. Irrigation water demand is simulated by wetting soils to their field capacity in irrigated cells (or fractions of cells) during growing seasons. The estimated water demand is then subtracted from surface waters or groundwater. As a result, its simulated irrigation water demands depend heavily on model physics such as those for ET. Similar to WaterGAP, groundwater recharge in PCR-GLOBWB can come from the upper soil (either natural recharge or irrigation return flow), or from river-beds. Output used in this study is based on model simulation at 5 arc minute spatial resolution and a daily time step.

2.2 Meteorological forcing data

The CLSM simulation was driven by the Princeton forcing dataset which comprises time series of precipitation, short and long wave radiation, air temperature, relative humidity, surface pressure and wind speed. The dataset, which spans 1948 to 2014, was constructed by bias correcting the precipitation, solar radiation, and air temperature fields from the National Center for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) re-analysis using in situ and satellite observations[33]. The data are provided on a 1° global grid (excluding the Antarctica) with a 3-hourly timestep. The Princeton dataset has been used for a wide variety of applications including examining long-term trends of drought[34], runoff[35], and changes in the water cycle[36]. CLSM was first spun up from uniform initial conditions over the period, 1948 to 2014. The first 10 years of that simulation were discarded and the temporal (1959-2014) mean of each variable at each grid pixel on December 31 was used to initialize the

155 experimental simulation on January 1, 1948. Use of multi-year average initial conditions is
156 optimal for reducing the impacts of climatological anomalies on an experimental model
157 simulation[37].

158 WaterGAP simulations from 1960 to 2009 were driven by gridded time series of monthly
159 precipitation and potential ET. Monthly precipitation obtained from Global Precipitation
160 Climate Center[38] (GPCC), Version 6, were downscaled by evenly distributing the monthly values
161 to wet days of the month determined from Climate Research Unit (CRU) data[39], version TS 3.10.
162 CRU temperature and cloud cover data were used to derive monthly potential ET based on the
163 Priestley and Taylor equation[40,41] which requires fewer atmospheric variable inputs than the
164 Penman-Monteith method used by most land surface models including CLSM. Monthly time
165 series of potential ET were further downscaled to daily values prior to model simulation.

166 The forcing data used for PCR-GLOBWB simulations during 1958-2015 were based on monthly
167 CRU (TS 3.21.2 for 1958-2010 and TS.3.24 for 2011-2015) precipitation, temperature and
168 potential ET, which were spatially and temporally downscaled using ERA40 and ERA-Interim
169 data[23]. CRU data are provided at 5 arc minute resolution but the temperature data were
170 further downscaled for better simulation of snow dynamics. Monthly CRU precipitation was
171 proportionally distributed to ERA rainy days for a given threshold (0.1 mm day^{-1}) with daily
172 values re-scaled to match monthly CRU precipitation. In areas with too few CRU stations, CRU
173 precipitation was replaced by ERA precipitation, thus this precipitation dataset is referred to as
174 CRU-ERA in this study. Monthly reference ET was estimated outside of the model simulation
175 using CRU meteorological data and the Penman-Monteith method and was downscaled to daily
176 values using a daily temperature-based ET product. PCR-GLOBWB was first repeatedly

177 executed for 150 years using the mean atmospheric forcing conditions under the non-
178 anthropogenic scenario from 1968-2000 to allow simulated groundwater to reach a steady
179 state. The resulting conditions were then use to initialize the model under non-anthropogenic
180 and anthropogenic simulation scenarios.

181 **2.3 In situ and GRACE satellite data**

182 In situ groundwater observations were obtained for the four Mississippi sub-basins and four
183 northeastern U.S. regions (see Supplementary Fig. S1) from the USGS web site. These 181 wells
184 were selected because they are located in unconfined or semi-confined aquifers, are not
185 directly affected by groundwater withdrawals or injections and have been continuously
186 monitored for 10-30 years. These data have been used in previous studies for validating GRACE
187 TWS[11,42,43,44] and investigating scale-dependency of groundwater storage[45]. Depth-to-water
188 measurements were converted to groundwater storage anomalies using individual estimates of
189 specific yield for each well following published procedures[28,11].

190 The GRACE satellites mapped Earth's gravity field on a monthly basis from 2002 to 2017. The
191 resulting time variable gravity observations were used to derive TWS anomalies which reflect
192 changes in soil moisture, groundwater, snowpack, and surface water storage (lakes and rivers).
193 GRACE also detected water storage changes associated with anthropogenic effects such as
194 groundwater abstractions which, along with surface water, are not simulated by CLSM. The
195 GRACE data used in this study were provided approximately monthly during the mission period
196 on a 0.5° global grid. They were developed using a regularization method that eliminates the
197 need for post-processing and thus better preserves TWS amplitudes and trends[46]. Despite the

198 grid resolution, the effective resolution of this GRACE product is still around 150,000 km^2 due to
199 the limitations of the method used[47,48].

200 For comparison with in situ data, modeled groundwater storage at the grid cell nearest each
201 well was selected. Because this study focuses on large-scale interannual variability, monthly
202 precipitation, ET, groundwater storage and TWS were averaged to 1° spatial resolution and
203 monthly anomalies were calculated by removing the long-term mean of each calendar month.
204 Anomalies were further standardized (thus unitless) so that they had zero mean and standard
205 deviation of one. Annual anomalies are averages of the standardized monthly anomalies in
206 each year. Unless otherwise noted, anomalies discussed in following sections refer to
207 standardized anomalies.

208 **2.4 Methods**

209 Regular Empirical Orthogonal Function (EOF) analysis[49] was applied to annual anomalies
210 (weighted by area) of Princeton forcing data, CLSM ET estimates, GPCC and CRU-ERA
211 precipitation, and simulated groundwater storage to identify patterns of variability and their
212 relationship with large-scale climate signals. The non-parametric Mann-Kendall method[50] was
213 used to calculate linear trend and quantify changes over the simulation periods.

214 **3. Results**

215 **3.1 Evaluation of model estimates**

216 Figure. 1 shows that the correlation coefficient between groundwater storage anomalies
217 simulated by the three models and observations generally exceeds 0.5.The exception is the

combined Red River and Lower Mississippi (Red-LM) region, where the mean water table depth (around 17 m) is much deeper than other regions[28]; as a result, groundwater response to atmospheric effects is considerably lagged and may not be well represented by the simplified groundwater schemes in these models. PCR-GLOBWB groundwater lags and shows considerably lower correlation with in situ data than the others in New Jersey, Massachusetts, Pennsylvania and Ohio-Tennessee.

Groundwater is a significant source of fresh water in some of these regions and thus it is possible that certain wells were directly affected by groundwater withdrawals, despite our effort to exclude such wells. Supplementary Figs.S2 and S3 show that simulating withdrawals had more persistent impacts on groundwater in Long Island where groundwater is heavily abstracted for municipal water supplies[51] and Red-LM where groundwater is abstracted to support irrigated agriculture[21,23], but it did not always improve the correlation with in situ observations in these two and other regions. These results suggest that the chosen observation wells outside of Long Island and Red-LM are generally unaffected by pumping and that simulation of withdrawals does not guarantee improved correlation even where withdrawals are significant. Simulating groundwater abstractions are challenged by lack of information on irrigation water demand, efficiency, timing and duration of irrigation[21].

Across the models, the correlation between WaterGAP and CLMS natural groundwater output exceeds 0.5 in most land areas except the Arabian Peninsula, the Amazon, high-mountain Asia and high latitudes of North America (see Supplementary Fig.S4). In drier climates where the latency between atmospheric phenomena and groundwater responses is large, discrepancies in model physics and precipitation data may enhance the differences in simulated groundwater.

240 In regions where snowmelt is a major source of groundwater recharge, differences in simulated
241 snow pack among models [26] translate to lower correlation among simulated groundwater time
242 series. CLSM does not represent surface water and its associated recharge, resulting in lower
243 correlation with models that do (WaterGAP and PCR-GLOBWB) in humid tropical and boreal
244 river basins in which surface water is a significant part of TWS[52]. CLSM groundwater correlation
245 with PCR-GLOBWB resemble that with WaterGAP (see Supplementary Fig. S5) including those
246 areas of low correlations. WaterGAP and PCR-GLOBWB do not simulate natural groundwater
247 storage changes in the Sahara Desert, hence the zero correlation between their groundwater
248 and that of CLSM in that area.

249 To help overcome the lack of spatial coverage of well data, modeled TWS was compared with
250 GRACE TWS. Fig. 2 shows that modeled monthly non-anthropogenic TWS anomalies generally
251 correlate well with the observations, with correlation coefficients exceeding 0.5 in large areas
252 of North America, South America, Northern Eurasia, Southern Asia and Australia (for CLSM and
253 PCR-GLOBWB only). In northern Africa and northwestern China, the correlations are low,
254 possibly due to the scarcity of in situ observations for bias correcting and calibrating forcing
255 data. Correlation is also reduced in regions with significant groundwater depletion such as
256 Northern India[10], which was detected by GRACE but not represented in these simulated non-
257 anthropogenic TWS. Similarly, rapid glacier and ice cap melt in the Gulf of Alaska[53] and the
258 Canadian archipelago[54] are also not simulated, causing poor correlations in those areas.
259 Negative correlation in the Congo basin may be attributed to lack of surface water in CLSM.
260 Despite their representation of surface water, WaterGAP and PCR-GLOBWB do not always show
261 better correlation with GRACE TWS than CLSM in large river basins such as the Amazon and the

Nile (Fig. 2), underscoring the challenge of simulating TWS and its components. As noted previously, despite the 0.5° resolution of the GRACE TWS product we used, the effective resolution of GRACE is still around 150,000 km[2,47,48] and thus small scale TWS variability not resolved by GRACE may contribute to diminished correlation with the models. Indeed, CLSM TWS exhibits much higher correlation (>0.6) with GRACE TWS when averaged over large basins[26].

Simulating human water use in WaterGAP and PCR-GLOBWB led to improved correlation with GRACE TWS (see Supplementary Figs.S6 and S7) in most regions where excessive groundwater withdrawals have led to groundwater depletion such as Northern India, the Middle East and the North China Plain. In the Great Plains, recharge from ephemeral playas may lead to overestimation of irrigation water requirement in these models[21] and hence degraded correlation with GRACE TWS. Overall, simulating withdrawals and water management changed the correlation of simulated TWS and with GRACE TWS by more than 0.02 in about 10% of the 1° grid cells over the global land area (excluding Antarctic and the Greenland,). This highlights the importance of evaluating non-anthropogenic TWS using GRACE data as in Fig.2.

3.2 Modes of variability

Empirical Orthogonal Function (EOF) analysis has been widely used to find important patterns of variability in climate datasets[55,56]. By construction, EOF analysis yields modes of variability, consisting of a spatial pattern and a temporal component (or principal component), with their explained variance over the total variance in decreasing order[57]. Natural variability of groundwater is mainly governed by precipitation and ET[45,58]. Thus, we first examine modes of

284 variability of global precipitation and ET. Fig. 3 shows that the principal component (PC) of the
285 leading mode of variability in Princeton precipitation, PC1, is significantly (significance levels are
286 provided at figure captions) correlated with the Southern Oscillation Index (SOI). PC1 contains
287 a 'trough'-shaped low-frequency variability, i.e., decreasing from 1948 to mid-1980 and
288 increasing afterwards which is also observed in SOI. The spatial pattern of EOF 1 shows the
289 low-frequency variability is mainly associated with changes in the Sahel, which experienced a
290 multi-decadal drought starting from 1960[59]. EOF 2 identified the low latitude regions where
291 ENSO is known to have stronger impacts on precipitation, namely, Australia, Southern Africa
292 and Northern South America[56]. Indeed, PC2 of Princeton precipitation exhibits higher
293 correlation (0.64) with SOI. These areas, including the Sahel, were also identified in a previous
294 study[60] which showed stronger correlation (0.82) between PC1 of global precipitation and
295 ENSO, likely due to the longer period (1900-1982) of that study.

296 The leading modes of variability in global GPCC precipitation (used as input to WaterGAP)
297 exhibit spatial and temporal patterns (Fig.4) similar to those of Princeton precipitation,
298 including the low frequency variability in PC1. In contrast, EOF 1 of CRU-ERA precipitation (used
299 as input to PCR-GLOBWB) is an apparent trend (Fig.5), mainly associated with changes in the
300 Sahara Desert, Alaska, the high-mountains of Asia and the southern Arabian Peninsula. Its EOF
301 2 and 3 exhibit spatial patterns similar to EOF 1 and EOF 2 of Princeton and GPCC precipitation.
302 EOF analysis of CRU precipitation produced similar results (not shown) as those of Princeton
303 and GPCC precipitation, suggesting that the trend in EOF1 of CRU-ERA precipitation was caused
304 by replacing CRU precipitation with ERA estimates at locations with few CRU gauges.

ET can have a significant impact on shallow groundwater and a warm climate would tend to favor increased ET. Supplementary Figs.S8 and S9 show that the leading modes of variability in Princeton longwave radiation, air temperature, relative humidity, air pressure and wind are apparent trends, presumably reflecting climate change[61,62]. In contrast, EOF 1 of global CLSM ET is similar to that of Princeton precipitation (see Supplementary Fig.S10). Both display the trough-shaped variability in PC1, suggesting strong control of precipitation on ET. EOF 2 of global CLSM ET is an apparent trend associated with increasing ET in northern Eurasia, the northeastern U.S. and Central Australia, and decreasing ET in northern South America and the Sahel. This trend only accounts for 7.3% of the total variance while most of the leading modes in the atmospheric forcing fields account for more than 20% of the total variance (see Supplementary Figs.S8 and S9). PCR-GLOBWB and WaterGAP ET output was not analyzed because the methods and high-resolution data used by those models to downscale potential and reference ET may also impact modes of variability in actual ET estimates, making it difficult to interpret the results.

The leading modes of variability in CLSM and WaterGAP groundwater show spatial and temporal patterns similar to those of their corresponding precipitation inputs (Figs. 3 and 4). The PCs of CLSM groundwater contain less high-frequency variability than those of Princeton precipitation, reflecting the low-pass filtering effect of groundwater recharge[63]. PC1 of CLSM groundwater exhibits the largest change in 1960s and 1970s, when PC2 of ET and PC1 of precipitation also exhibit their largest changes. These are mainly attributed to significant decrease in precipitation in the Sahel and increases in ET in northern Europe, the northeastern US and central Australia during that period. PCs of groundwater from both CLSM and

WaterGAP show lower correlation with SOI than that between precipitation and SOI, reflecting the influence of other processes and fluxes on groundwater such as ET.

On the other hand, the leading modes of variability in PCR-GLOBWB groundwater bear less resemblance to those of CRU-ERA precipitation. Its EOF 1 is still a trend, but its spatial pattern differs from that of EOF 1 of CRU-ERA precipitation, suggesting other factors may have had significant impacts on simulated groundwater. Similarly, its EOF 2 does not resemble that of CRU-ERA precipitation while its EOF 3 shows spatial and temporal patterns similar to EOF 3 of CRU-ERA precipitation.

3.3 Long-term trends

Figure 6 shows the Mann-Kendal trend of annual standardized precipitation and non-anthropogenic groundwater storage anomalies from the three models. As in the EOF analysis, Princeton and GPCC precipitation datasets exhibit similar trends including strong decreases in the Sahel and sub-tropical Africa and increases in the northeastern US and Western Europe, changes also reported in IPCC/AR5[16]. In contrast, CRU-ERA precipitation exhibits different trends, especially in the Sahara Desert, Alaska and the Amazon, most of which were identified in its EOF 1 (Fig.5).

The CLSM and WaterGAP groundwater trends resemble their corresponding precipitation trends in most land areas, particularly in much of Africa, eastern Asia and Northern Russia, where groundwater decreased significantly due to below normal precipitation. In the Hudson Bay and the Great Lake regions, the two models produced different trends in groundwater.

348 This may be attributed to the fact that CLSM does not simulate surface water and its related
349 recharge. Decreasing CLSM groundwater in much of northern Europe appears to result from
350 increasing ET (see Supplementary Fig.S11), and from a combination of ET increases and
351 precipitation decreases in east-central Asia. Northern Eurasia experienced decreased surface
352 air pressure and, in some areas, increased wind speed (see Supplementary Fig. S12), both of
353 which favor increased ET and thus decreased groundwater storage. Changes in surface air
354 pressure in the high latitudes were noted in IPCC AR5[16]. Several re-analysis products also show
355 increased storminess in northern, western, and central Europe towards the end of the 20th
356 century, more so in the North Sea and Baltic Sea regions[64]. More frequent cyclones (associated
357 with a low pressure center) in the lower arctic Canada in 1953- 2002 were also reported[65].
358 Since WaterGAP estimates potential ET using the Priestly-Taylor equation, which does not
359 require wind and air pressure data, its estimated ET likely would not reflect these changes.

360 As in other analyses of this study, PCR-GLOBWB groundwater trends exhibit spatial patterns
361 different from those of CLSM and WaterGAP (Fig.6). PCR-GLOBWB groundwater increased
362 significantly in high altitude regions of northern South America and Iran, despite no associated
363 increases in CRU-ERA precipitation. The model uses lapse rates derived from a high-resolution
364 temperature climatology data to downscale CRU temperature data in order to better simulate
365 snow dynamics. It is possible this downscaling scheme led to earlier and increased snowmelt,
366 and hence increased groundwater recharge in those areas.
367

368 Figure 7 shows that globally, CLSM and WaterGAP groundwater storage (ignoring
369 anthropogenic effects) decreased during their respective simulation periods, consistent with

370 the trends in their corresponding global precipitation datasets. Both time series also exhibit
371 low frequency variability, namely a change in the mid-1980s from generally positive to
372 generally negative anomalies, resulting in negative trends. Apparently, ET did not contribute to
373 the negative trend in CLSM groundwater because global averaged ET was relatively stable over the
374 simulation period with an insignificant decreasing trend (see Supplementary Fig. S13). Changes in
375 other atmospheric forcing fields such as increased humidity and air pressure and decreased wind
376 speed (see Supplementary Fig. S14) may suppress the effect of rising temperature on ET. The
377 global standardized CLSM groundwater time series has larger amplitude variation than that of
378 WaterGAP (Fig.7), owing to larger amplitude variations in global Princeton precipitation than in
379 global GPCC precipitation, and hence a larger (and significant) decreasing trend. In contrast,
380 Global PCR-GLOBWB groundwater exhibits an increasing trend, consistent with the trend in
381 global CRU-ERA precipitation (Fig.7).

382 Groundwater responds to surface wet and dry conditions asymmetrically[58], .e.g., an intense
383 rainfall event in which the water runs off rapidly may do less to replenish an aquifer than a
384 slow, soaking rain. Moreover, groundwater levels reflect the cumulative effects of atmospheric
385 conditions, so that one rainy month or year may not be enough to return groundwater levels to
386 normal after a drought, for example. Differences in model physics such as those controlling
387 partitioning of precipitation into surface runoff and infiltration may further amplify these
388 asymmetrical responses and discrepancies among the models. This is why the three global
389 groundwater storage time series exhibit weaker cross-correlations than the global precipitation
390 time series (Table 1). In the case of this study, global PCR-GLOBWB groundwater was
391 consistently the least well correlated among groundwater output of the three models (Table 1).

392 The standardized anomalies discussed above enable examination of modes of variability and
393 comparison of trends across different regions. Fig. 7 also shows the trends of global, non-
394 standardized annual groundwater storage anomalies (in mm $year^{-1}$) from three models which
395 range from -0.01 to -2.18 mm $year^{-1}$. Note that the trend in global non-standardized PCR-
396 GLOBWB groundwater is negative, in contrast to the positive trend in the standardized time
397 series. This is because groundwater storage anomalies in dry climate may exhibit significant
398 trend but their contributions to the global total non-standardized groundwater storage
399 anomalies (mm) are diminished due to their small amplitude variations (mm).

400 For comparison, global total groundwater by WaterGAP and PCR-GLOBWB, including
401 anthropogenic processes, decreased significantly (Figs.S15 and S16). Global standardized PCR-
402 GLOBWB groundwater exhibits much smaller decreasing (0.001 vs 0.002 from WaterGAP)
403 trend, owing to the increasing trend in global PCR_GLOBWB non-anthropogenic groundwater.
404 EOF 1 in the global anthropogenic groundwater is a trend (see Supplementary Figs.S17 and S18)
405 mainly associated with decreases in known regions of intensive groundwater abstractions such
406 as the Arabian Peninsula, Northern India, the North China Plain and the Great Plains[21]. Its
407 remaining EOFs reflect the influence of climate variability, also identified in the corresponding
408 non-anthropogenic groundwater output (Figs.4 and 5). Once again, WaterGAP and PCR-
409 GLOBWB produced drastically different rates of decline in global non-standardized
410 anthropogenic groundwater, 46.15 and 0.35 mm $year^{-1}$, respectively (see Supplementary Figs.
411 S15 and S16).

412

4. Discussions and conclusions

This study investigates the long-term variability of global non-anthropogenic groundwater storage and its environmental controls. Due to the scarcity of in situ groundwater observations, we analyzed simulated groundwater storage from the Catchment land surface model (CLSM), WaterGAP and PCR-GLOBWB water resource models. The outputs from these three models may be considered independent given that the model parameterizations are quite distinct and each model was driven by a different atmospheric forcing dataset. Modeled groundwater storage and terrestrial water storage (TWS) were evaluated using in situ data in the eight U.S. regions and GRACE derived TWS at the globe scale. The three output datasets compared reasonably well with the observations based on correlations greater than 0.5 in most regions, demonstrating skill in reproducing interannual variability of water storage. Output from the models was somewhat consistent, especially between CLSM and WaterGAP, owing to the more consistent precipitation data used by those two models.

Neglecting anthropogenic impacts, global total groundwater storage decreased during the past 50-70 years, at a rate of 0.25 mm $year^{-1}$ based on the ensemble average. It is possible that the long-term decline is a symptom of climate change, but considering that low-frequency variability in the global groundwater storage time series and in regional groundwater of deeper aquifers[45], the 50 to 67 years of simulation periods may not be long enough to identify a secular trend with confidence. This low frequency variability together with the sensitivity of trends to anomaly amplitudes that vary among models may have also contributed to the discrepancies

among trends estimated by the three models. Therefore, better characterizing that low frequency variability would be valuable for improving the models and interpreting the apparent groundwater storage trends.

The EOF analysis revealed that ENSO influenced the variability of global groundwater through its control of precipitation patterns. Because ENSO also exhibits low frequency variability, it is plausible that multidecadal scale variability in global total groundwater can be partly attributed to ENSO, though more research would be needed to describe those relationships.

Our analysis shows that simulated regional and global groundwater anomalies are sensitive to trends in precipitation. Therefore, the linear trend in global groundwater reported here must be interpreted in the context of the reliability and temporal consistency of the forcing datasets, which currently remain incompletely established. For the Princeton dataset, the effectiveness of bias-correction depends on regional availability of in situ data, and may rely more on the fidelity of the NCEP/NCAR re-analysis (used as a baseline) when and where in situ data are scarce. Meteorological data tend to be scarce in the deep tropics, at high latitudes and during earlier years of the 1948-2014 period, especially for fields such as pressure and wind speed[66]. The GPCC and CRU precipitation data are based on gauge observations and their spatial and temporal variability and consistency may be adversely affected by changes in the numbers and locations of gauges. On the other hand, simulated ET did not exhibit a significant trend and thus was not a factor in the simulated global groundwater depletion trend.

Our analysis shows that representing groundwater abstraction for supporting irrigated agriculture increased the magnitude of the decreasing trends in global groundwater storage

455 simulated by WaterGAP and PCR-GLOBWB. In addition, groundwater depletions became the
456 leading mode of variability in global groundwater, surpassing ENSO. However, the resulting
457 trends in global groundwater storage, 46.15 to 0.35 mm year^{-1} for WaterGAP and PCR-
458 GLOBWB, respectively, are directly impacted by non-anthropogenic trends and the low
459 frequency variability in global WaterGAP and PCR-GLOBWB non-anthropogenic groundwater
460 (Fig.7) are embedded in their corresponding global anthropogenic groundwater time series
461 (Figs.S15 and S16). This emphasizes the importance of evaluating natural groundwater
462 variability when assessing human impacts on groundwater resources. This is particularly true
463 for models like PCR-GLOBWB which rely on the underlying hydrological model to estimate
464 consumptive irrigation water use.

465 Finally, we caution that our evaluation using in situ groundwater data was limited to regions of
466 the U.S. that do not represent the climates of Africa, east-central Asia, and Northern Eurasia,
467 where some of the significant groundwater storage changes were simulated. Uncertainty in
468 simulated groundwater time series may be higher in drier climates where aquifers tend to be
469 deeper and therefore have longer-scale variability and disconnection from the atmosphere
470 which are difficult to model. Evaluation using GRACE data is limited to less than 15 years when
471 GRACE was active and provides no insight on a model's ability to separate groundwater from
472 other TWS changes.

473

474

475

Acknowledgements:

This study was funded by NASA's GRACE and GRACE Follow On Science Team and Terrestrial Hydrology Program. CLSM model output are available through the Goddard Earth Sciences Data and Information Services Center, https://disc.gsfc.nasa.gov/datasets?keywords=Hydrology. We thank Somayeh Shadkam and Hannes Müller Schmied for providing the GPCC precipitation data used in this version of WaterGAP.

Contributions

B. L. and M.R. conceived the study and wrote the paper. J. S. and E. W. provided an updated version of the Princeton forcing data and reviewed the paper. E. S. provided the output of PCR-GLOBWB and CRU-ERA precipitation data and reviewed the paper.

Competing Interests

The authors declare that they have no competing interests.

Figures

Fig. 1. Standardized monthly groundwater storage anomalies from in situ observations and CLSM, WaterGAP and PCR-GLOBWB under the non-anthropogenic scenario for the four northeastern U.S. regions and the four sub-basins of the Mississippi river. "r" represents correlation between in situ observations and modeled groundwater with subscribes c, w and p representing CLSM, WaterGAP and PCR-GLOBWB, respectively.

Fig. 2. Correlation between monthly TWS anomalies from GRACE and those from CLMS, WaterGAP and PCR-GLOBWB under the non-anthropogenic scenario. Staples indicate correlation greater than 0.5.

Fig.3 Spatial and temporal patterns of the first two EOFs of annual standardized Princeton precipitation and CLSM groundwater storage anomalies. Numbers in parentheses represent percentages of variance explained over total variance. Red lines represent annual SOI. "r" represents the correlations between normalized PCs and SOI which are significant at the 0.01 level.

Fig.4 same as Fig. 3 but for annual standardized GPCC precipitation and WaterGAP groundwater storage anomalies.

Fig.5 Same as Fig. 3 but for the first three EOFs of annual standardized CRU-ERA precipitation and PCR-GLOBWB groundwater storage anomalies. Correlations are significant at the 0.01 level except that between PC2 of groundwater and SOI.

Fig. 6. Mann-Kendal trend of annual standardized Princeton, GPCC, CRU-ERA precipitation (top row), and CLSM, WaterGAP and PCR-GLOBWB non-anthropogenic groundwater storage anomalies (bottom row). Stipples represent significant trend at the 0.05 level.

513 Fig. 7. Global-averaged non-anthropogenic groundwater storage anomalies from the three
514 models and their ensemble average (left column) and their corresponding global precipitation
515 (right column). Slopes are Mann-Kendal trend (numbers in black indicate significant trend at the
516 0.05 level). Numbers in parentheses are trend of non-standardized global total groundwater
517 storage anomalies.

518

519

520

521

522 Table 1. Cross-correlation between global total groundwater storage of the three models (Fig.7)
523 and between global total precipitation of the three datasets.

	CLSM & WaterGAP	CLSM & PCR-GLOBWB	WaterGAP & PCR-GLOBWB
global groundwater	0.67	0.48	0.47
	Princeton & GPCC	Princeton and CRU-ERA	GPCC and CRU-ERA
global precipitation	0.91	0.90	0.86

524

References:

1. Famiglietti, J.S. The global groundwater crisis. *Nature Climate Change*, **4**(11), 945-948 (2014), doi:10.1038/nclimate2425.
2. Berghuijs, W. R. & Kirchner, J. W. The relationship between contrasting ages of groundwater and streamflow, *Geophys. Res. Lett.*, **44**, 8925–8935(2017), doi:10.1002/2017GL074962.
3. Famiglietti, J. S. & Rodell, M. Water in the Balance. *Science*, **340 (6138),** 1300-1301(2013), 10.1126/science.1236460.
4. Taylor et al. Ground water and climate change, *Nature Climate Change* **3**, pages 322–329 (2013), doi:10.1038/nclimate1744.
5. Gurdak et al. Climate variability controls on unsaturated water and chemical movement, High Plains aquifer, USA. *Vadose Zone J.* **6(3),** 533–547 (2007), https://doi.org/10.2136/vzj2006.0087.
6. Anderson, W. P., Jr., & Emanuel, R. E. Effect of interannual and interdecadal climate oscillations on groundwater in North Carolina. *Geophys. Res. Lett*. **35**, L23402(2008), doi:10.1029/2008GL036054.
7. Barco, J., Hogue T. S., Girotto, M., Kendall, D. R. & Putti, M. Climate signal propagation in southern California aquifers. *Water Resour. Res*. **46**, W00F05(2010), doi:10.1029/2009WR008376.
8. Velasco, E. M., Gurdak, J. J., Dickinson, J. E., Ferré, T. P. A. & Corona, C. R. Interannual to multidecadal climate forcings on groundwater resources of the U.S. West Coast. *J. Hydrol. Reg. Stud.* **11,** 250-265(2015), doi:10.1016/j.ejrh.2015.11.018.

9. Perez-Valdivia, C., Sauchyn, D. & Vanstone, J. Groundwater levels and teleconnection patterns in the Canadian Prairies. *Water Resour. Res*. **48,** W07516, oi:10.1029/2011WR010930 (2012).
10. Rodell, M., Velicogna, I. & Famiglietti, J.S. Satellite-based estimates of groundwater depletion in India. *Nature* **460,** 999–1002(2009), http://dx.doi.org/10.1038/nature08238.
11. Rodell et al. Estimating ground water storage changes in the Mississippi River basin (USA) using GRACE. *Hydrogeology J.* **15,** 159-166(2007), doi:10.1007/s10040-006-0103-7.
12. Jiménez Cisneros et al. Freshwater resources. In: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Field, C.B., V.R. Barros, D.J. Dokken, K.J. Mach, M.D. Mastrandrea, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L. White (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 229-269, 2014.
13. Allen D. M., Mackie, D. C. & Wei, M. Groundwater and climate change: a sensitivity analysis for the Grand Forks aquifer, southern British Columbia, Canada. *Hydrogeol. J.* 12, 270–290(2004), DOI 10.1007/s10040-003-0261-9.
14. Crosbie et al. Potential climate change effects on groundwater recharge in the High Plains Aquifer, USA. *Water Resour. Res.*, **49(7),** 3936-3951(2013), http://dx.doi.org/10.1002/wrcr.20292.
15. Meixner et al. Implications of projected climate change for groundwater recharge in the western United States. *J. Hydrol.* **534,** 124-138(2016), doi:10.1016/j.jhydrol.2015.12.027.

16. Hartmann et al. Observations: Atmosphere and Surface. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (eds. Stocker et al.). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

17. Small, E. E. Climatic controls on diffuse groundwater recharge in semiarid environments of the southwestern United States. *Wat. Resour. Res*. **41(4),** W04012 (2005), https://doi.org/10.1029/2004WR003193.

18. Van Dijk, A.I., Renzullo, L.J., Wada, Y. & Tregoning, P. A global water cycle reanalysis (2003-2012) merging satellite gravimetry and altimetry observations with a hydrological multi-model ensemble. *Hydrol. Earth Syst. Sci*. **18**, 2955-2973(2014), https://doi.org/10.5194/hess-18-2955-2014.

19. Richey et al. Uncertainty in global groundwater storage estimates in a total groundwater stress framework. *Water Resour. Res*., **51,** 5198–5216(2015), doi:10.1002/2015WR017351.

20. Wada et al. Multimodel projections and uncertainties of irrigation water demand under climate change. *Geophys. Res. Lett.*, **40,** 4626–4632(2013), doi:10.1002/grl.50686.

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

22. Koster, R.D., Suarez, M.J., Ducharne, A., Stieglitz, M., Kumar, P. A catchment based approach to modeling land surface processes in a general circulation model, 1, Model structure. *J. Geophys. Res*. **105**, 24809–24822 (2000).

591 23. Sutanudjaja et al. PCR-GLOBWB 2: a 5 arcmin global hydrological and water resources
592 model, Geosci. Model Dev., **11**, 2429–2453, https://doi.org/10.5194/gmd-11-2429-2018.
593 24. Niu, G.-Y., Yang, Z.-L., Dickinson, R. E., Gulden, L. E. & Su H. Development of a simple
594 groundwater model for use in climate models and evaluation with Gravity Recovery and
595 Climate Experiment data, *J. Geophys. Res.*, **112,** D07103(2007),
596 doi:10.1029/2006JD007522.
597 25. Niraula et al. How might recharge change under projected climate change in western US?
598 *Geophys. Res. Lett*., **44,** 10, 407–10,418(2017), doi:10.1002/2017GL075421.
599 26. Xia et al. Comparison and assessment of three advanced land surface models in simulating
600 terrestrial water storage components over the United States. *J. Hydrometeorol.* **18 (3),** 625-
601 649 (2017), 10.1175/jhm-d-16-0112.1.
602 27. Scanlon et al. Global models underestimate large decadal declining and rising water storage
603 trends relative to GRACE satellite data. *PNAS*, 2018 115(6),
604 https://doi.org/10.1073/pnas.1704665115.
605 28. Li, B. & Rodell, M. Evaluation of a model-based groundwater drought indicator in the
606 conterminous U.S. *J. Hydrol*., **526,** 78-88(2014),
607 http://dx.doi.org/10.1016/j.jhydrol.2014.09.027 (2014).
608 29. Houborg, R., Rodell, M., Li, B., Reichle, R., & Zaitchik, B.F. Drought indicators based on
609 model-assimilated Gravity Recovery and Climate Experiment (GRACE) terrestrial water
610 storage observations. *Water Resour. Res*. **48,** W07525(2012)
611 http://dx.doi.org/10.1029/2011WR011291.

30. Li et al. Assimilation of GRACE terrestrial water storage into a land surface model: evaluation and potential value for drought monitoring in western and central Europe. *J. Hydrol*. 446–447, 103–115 (2012), http://dx.doi.org/10.1016/j.jhydrol.2012.04.035.

31. Mueller et al. Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations. *Geophys. Res. Lett*. **38,** L06402(2011), doi:10.1029/2010GL046230.

32. Jiménez et al. Global intercomparison of 12 land surface heat flux estimates. *J. Geophys. Res.* **116**, D02102(2011), doi:10.1029/2010JD014545.

33. Sheffield, J., Goteti, G. & Wood, E. F. Development of a 50-yr high-resolution global dataset of meteorological forcings for land surface modeling. *J. Clim.* **19,** 3088–3111(2006).

34. Sheffield, J. & Wood, E. F. Global Trends and Variability in Soil Moisture and Drought Characteristics, 1950–2000, from Observation-Driven Simulations of the Terrestrial Hydrologic Cycle. *J. Clim*. **21,** 432–458 (2008), https://doi.org/10.1175/2007JCLI1822.1.

35. Clark, E. A., Sheffield, J., van Vliet, M. T. H., Nijssen, B. & Lettenmaier, D. P. Continental Runoff into the Oceans (1950–2008). *J. Hydrometeor.* **16**, 1502–1520 (2015), https://doi.org/10.1175/JHM-D-14-0183.1.

36. Rodell et al. The Observed State of the Water Cycle in the Early Twenty-First Century. J. Climate, 28, 8289–8318(2015), doi: http://dx.doi.org/10.1175/JCLI-D-14-00555.1.

37. Rodell, M., Houser, P.R., Berg, A.A & Famiglietti, J.S. Evaluation of 10 methods for initializing a land surface model, *J. Hydrometeoro*., **6,** 146-155 (2005), https://doi.org/10.1175/JHM414.1.

38. Schneider et al. GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle, Theor. Appl. Climatol., 115, 15–40, 2013, doi:10.1007/s00704-013-0860-x.

635 39. Harris, I., Jones, P., Osborn, T. and Lister, D. Updated high-resolution grids of monthly
636 climatic observations—The CRU TS3.10 dataset, Int. J. Climatol., 34, 623–642, 2014,
637 doi:10.1002/joc.3711.
638 40. Priestley, C. & Taylor, R. On the assessment of surface heat flux and evaporation using
639 large scale parameters, Mon. Weather Rev., 100 (1972), pp. 81-92.
640 41. Doll, P., Kaspar, F. & Lehner, B. A global hydrological model for deriving water
641 availability indicators: Model tuning and validation, J. Hydrol., 270, 105–134, 2003,
642 doi:10.1016/S0022-1694(02)00283-4.
643 42. Rodell, M. & Famiglietti. J. S. An analysis of terrestrial water storage variations in Illinois
644 with implications for the Gravity Recovery and Climate Experiment (GRACE). *Water*
645 *Resour. Res.* **37 (5)**, 1327-1339(2001), 10.1029/2000WR900306.
646 43. Zaitchik, B.F., Rodell, M., Reichle, R.H. Assimilation of GRACE terrestrial water storage
647 data into a land surface model: results for the Mississippi river basin. *J. Hydrometeorol*. **9,**
648 535–548 (2008), http://dx.doi.org/10.1175/2007JHM951.1.
649 44. Kumar et al. Assimilation of gridded GRACE terrestrial water storage estimates in the North
650 American Land Data Assimilation System. *J. Hydrometeor.* **17 (7),** 1951-1972(2016),
651 10.1175/jhm-d-15-0157.1.
652 45. Li., B., Rodell, M. & Famiglietti, J. S. Groundwater variability across temporal and spatial
653 scales in the central and northeastern U.S. *J. Hydrol.*, **525,** 769-780(2015),
654 http://dx.doi.org/10.1016/j.jhydrol.2015.04.033.
655 46. Save, H., Bettadpur, S. & Tapley, B.D. Reducing errors in the GRACE gravity solutions
656 using regularization. *J. of Geodesy* **86(9),** 695-711(2012), 10.1007/s00190-012-0548-5.

47. Rowlands et al. Resolving mass flux at high spatial and temporal resolution using GRACE intersatellite measurements. Geophys. Res. Lett., 32, L04310, 2005, doi:10.1029/2004GL021908.

48. Swenson, S., Yeh, P. J. F., Wahr, J., & Famiglietti, J. A comparison of terrestrial water storage variations from GRACE with in situ measurements from Illinois. Geophysical Research Letters, 33(16), 2006, https://doi.org/10.1029/2006GL026962.

49. Quadrelli, R., Bretherton, C. S. & Wallace, J. M. On Sampling Errors in Empirical Orthogonal Functions. *J. Clim.* **18**, 3704-3710(2005), https://doi.org/10.1175/JCLI3500.1.

50. Mishra V, Cherkauer, K. A. & Shukla, S. Assessment of drought due to historic climate variability and projected future climate change in the midwestern United States. *J Hydrometeorol.* **11,** 46–68(2010), https://doi.org/10.1175/2009JHM1156.1.

51. Busciolano, R. Water-table and potentiometric-surface altitudes of the upper glacial, Magothy, and Lloyd aquifers on Long Island, New York, in March-April 2000, with a summary of hydrogeologic conditions. 2002. U.S. Geological Survey Water-Resources Investigations Report 2001-4165, 17 p.

52. Getirana, A., Kumar, S., Girotto, M., & Rodell, M. Rivers and floodplains as key components of global terrestrial water storage variability. Geophysical Research Letters, 44, 2017, DOI: 10.1002/2017GL074684.

53. Luthcke et al. Antarctica, Greenland and Gulf of Alaska land ice evolution from an iterated GRACE global mascon solution. *J. Glaciol.* **59**, 613-631(2013), doi:10.3189/2013JoG12J14.

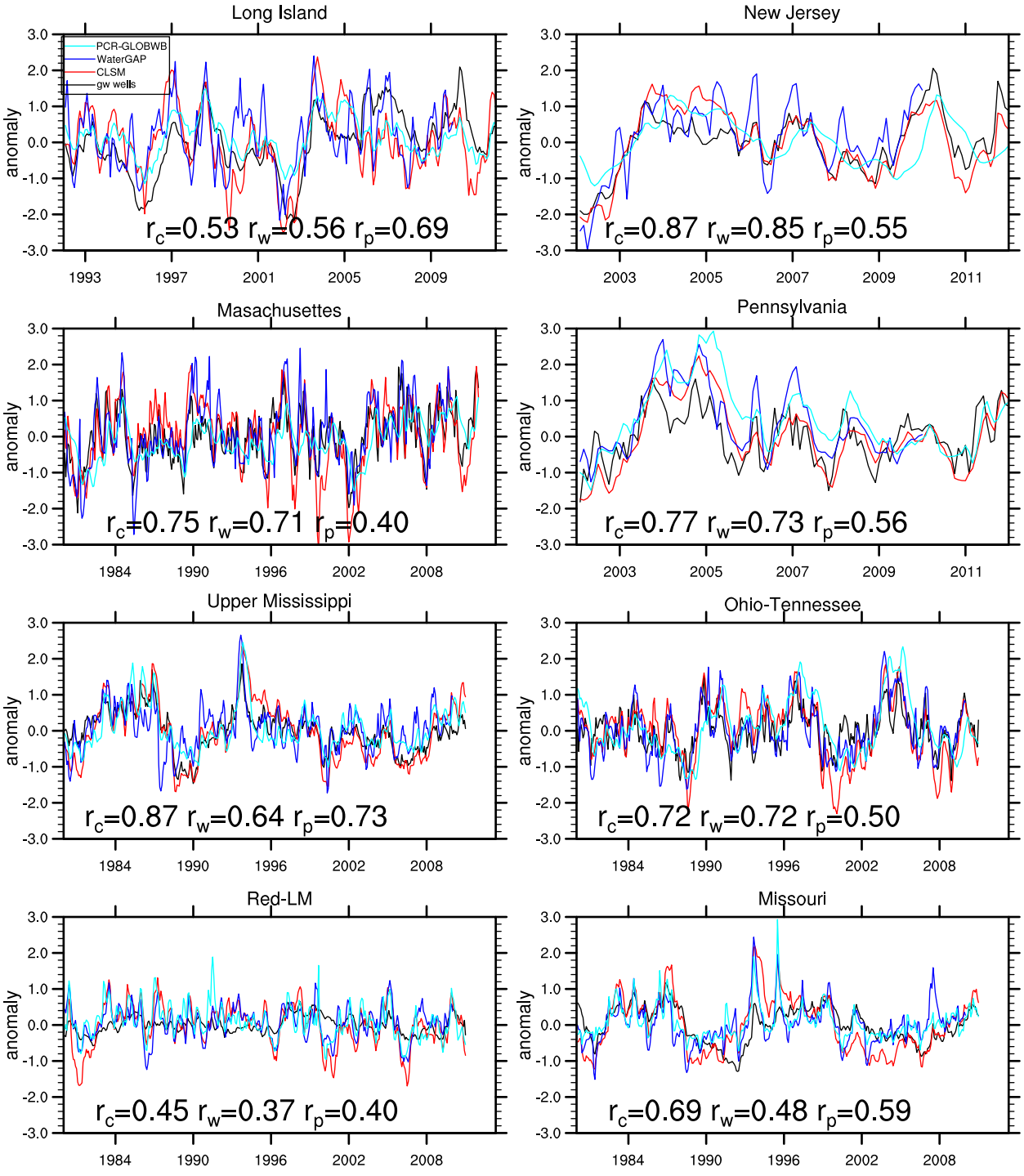
54. Jacob, T., Wahr, J., Pfeffer, W. T., & Swenson, S. Recent contributions of glaciers and ice caps to sea level rise. *Nature* **482**, 514-518 (2012), doi:10.1038/nature10847.

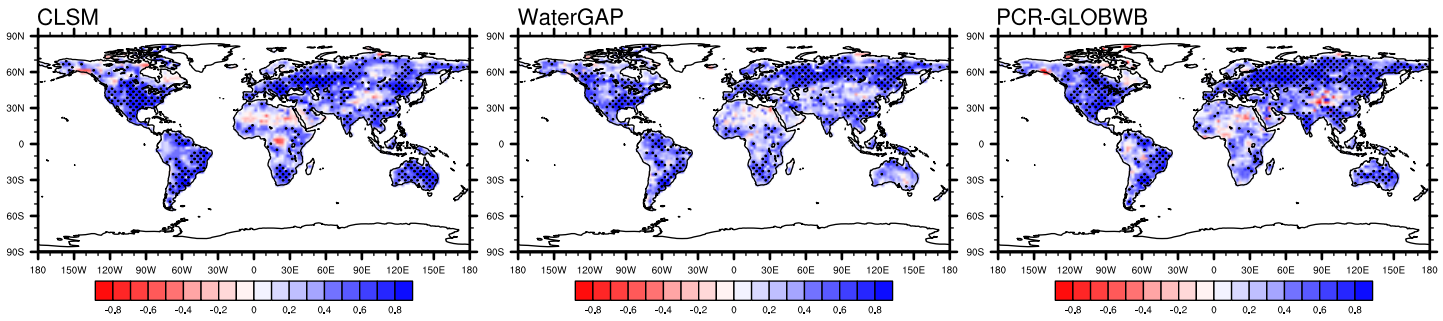
680 55. Thompson, D. W. J., & Wallace, J. M. Annular modes in the extratropical circulation. Part I:
681 Month-to-month variability. *J. Clim.* **13,** 1000-1016(2000), https://doi.org/10.1175/1520-
682 0442(2000)013<1000:AMITEC>2.0.CO;2.
683 56. Dai, A. G., & Wigley, T. M. L. Global patterns of ENSO induced precipitation. *Geophys.*
684 *Res. Lett.* **27,** 1283–1286 (2000), https://doi.org/10.1029/1999GL011140.
685 57. Jolliffe, I. T., Principal Component Analysis, 2nd Edition, Springer, New York, 2002.
686 58. Eltahir, E.A.B. & Yeh, P.J.-F. On the asymmetric response of aquifer water level to floods
687 and droughts in Illinois. *Water Resour. Res*. **35** (4), 1199–1217(1999),
688 https://doi.org/10.1029/1998WR900071.
689 59. Held, I.M., Delworth, T. L., Lu, J., Findell, K. L. & Knutson, T. R. Simulation of Sahel
690 drought in the 20th and 21st centuries. *Proc. Nat. Acad. Sci*., **102(50),** 17891-17896.
691 doi:10.1073/pnas.0509057102 (2005).
692 60. Dai, A., Fung, I. Y. & Del Genio, A. D. Surface Observed Global Land Precipitation
693 Variations during 1900–88. *J. Clim.* **10,** 2943-2962(1997), https://doi.org/10.1175/1520-
694 0442(1997)010<2943:SOGLPV>2.0.CO;2.
695 61. Dai, A. Recent climatology, variability, and trends in global surface humidity. *J. Clim.* **19**,
696 3589–3606(2006), https://doi.org/10.1175/JCLI3816.1.
697 *62*. McVicar et al. Global review and synthesis of trends in observed terrestrial near-surface
698 wind speed: Implications for evaporation, *J. Hydrol*. **416–417**, 182–205,
699 doi:10.1016/j.jhydrol.2011.10.024 (2012).
700 63. Wu, W., Geller, M. A. & Dickinson, R. E. The response of soil moisture to long-term
701 variability of precipitation. *J. Hydrometeorol*. **3,** 604-613 (2002),
702 https://doi.org/10.1175/1525-7541(2002)003<0604:TROSMT>2.0.CO;2.

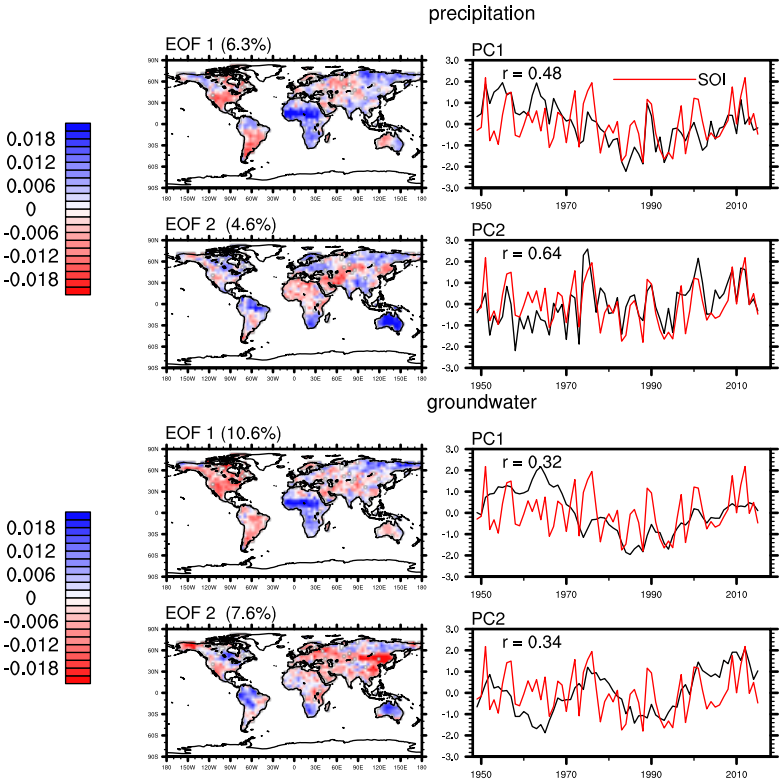
64. Donat et al. Reanalysis suggests long-term upward trends in European storminess since 1871. *Geophys. Res. Lett.* **38**, L14703(2011), doi:10.1029/2011GL047995.

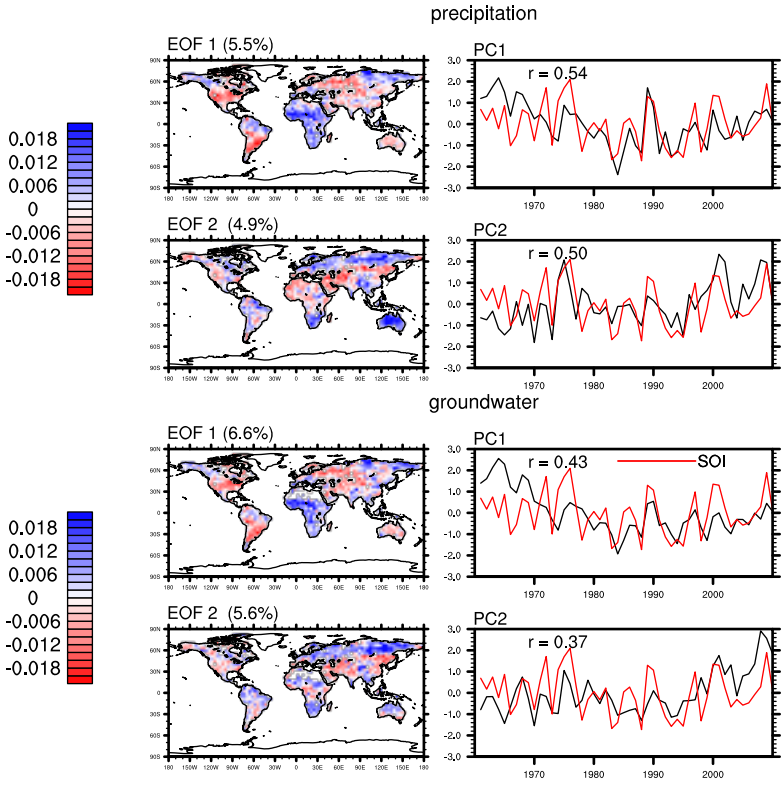
65. Wang, X., Wan, H. & Swail, V. Observed changes in cyclone activity in Canada and their relationships to major circulation regimes. *J. Clim.* **19,** 896–915 (2006), https://doi.org/10.1175/JCLI3664.1.

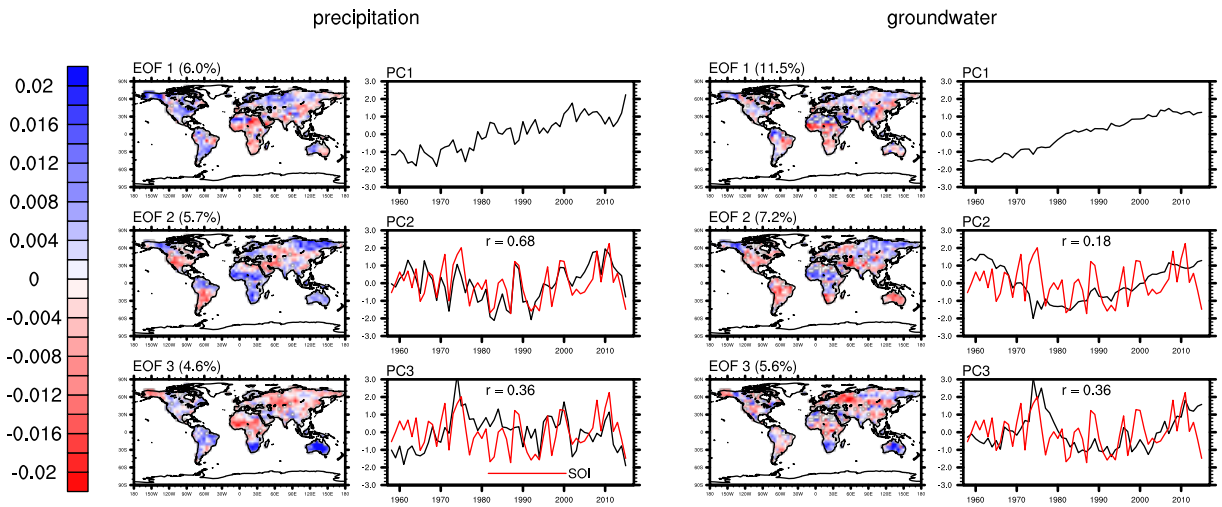
66. Sheffield, J., Wood, E. F. & Roderick, M. L. Little change in global drought over the past 60 years, *Nature* **491**, 435-438(2012), doi:10.1038/nature11575.

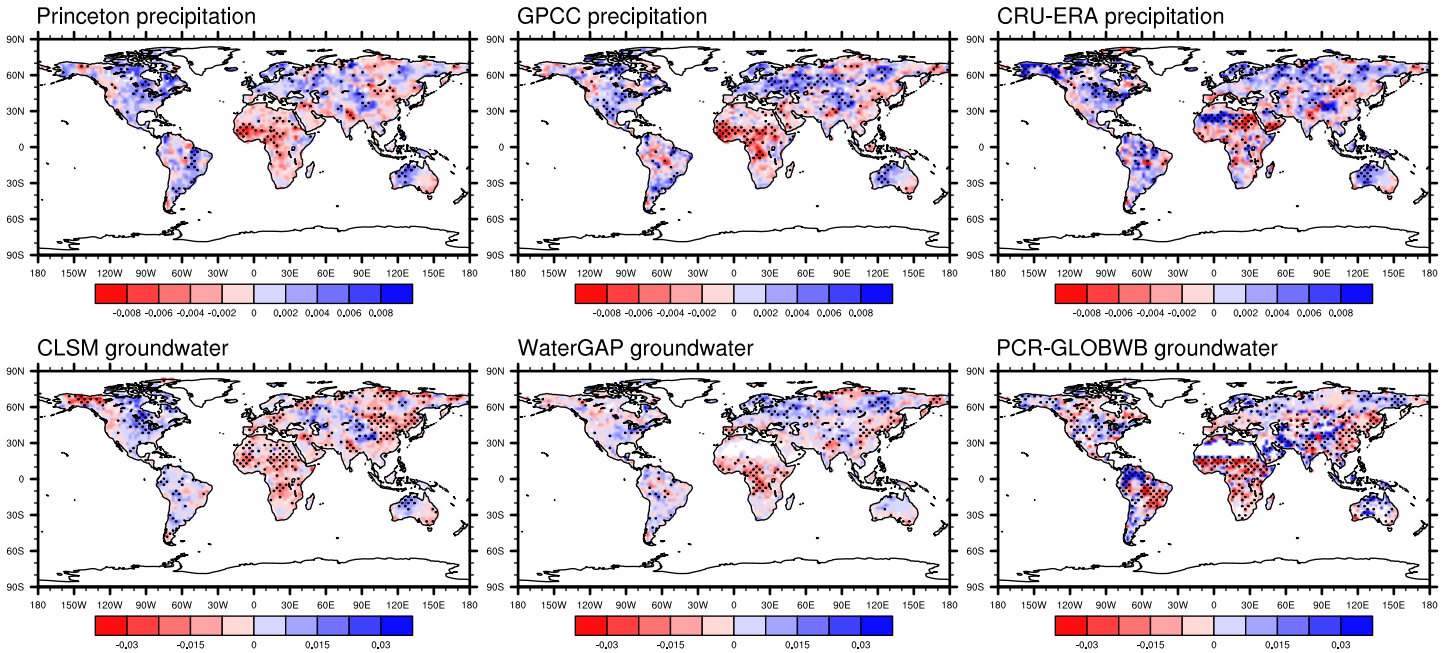

Long Island
PCR-GLOBWB
WaterGAP
CLSM
gw wells
anomaly
r_c=0.53 r_w=0.56 r_p=0.69
New Jersey
r_c=0.87 r_w=0.85 r_p=0.55
Masachusettes
r_c=0.75 r_w=0.71 r_p=0.40
Pennsylvania
r_c=0.77 r_w=0.73 r_p=0.56
Upper Mississippi
r_c=0.87 r_w=0.64 r_p=0.73
Ohio-Tennessee
r_c=0.72 r_w=0.72 r_p=0.50
Red-LM
r_c=0.45 r_w=0.37 r_p=0.40
Missouri
r_c=0.69 r_w=0.48 r_p=0.59

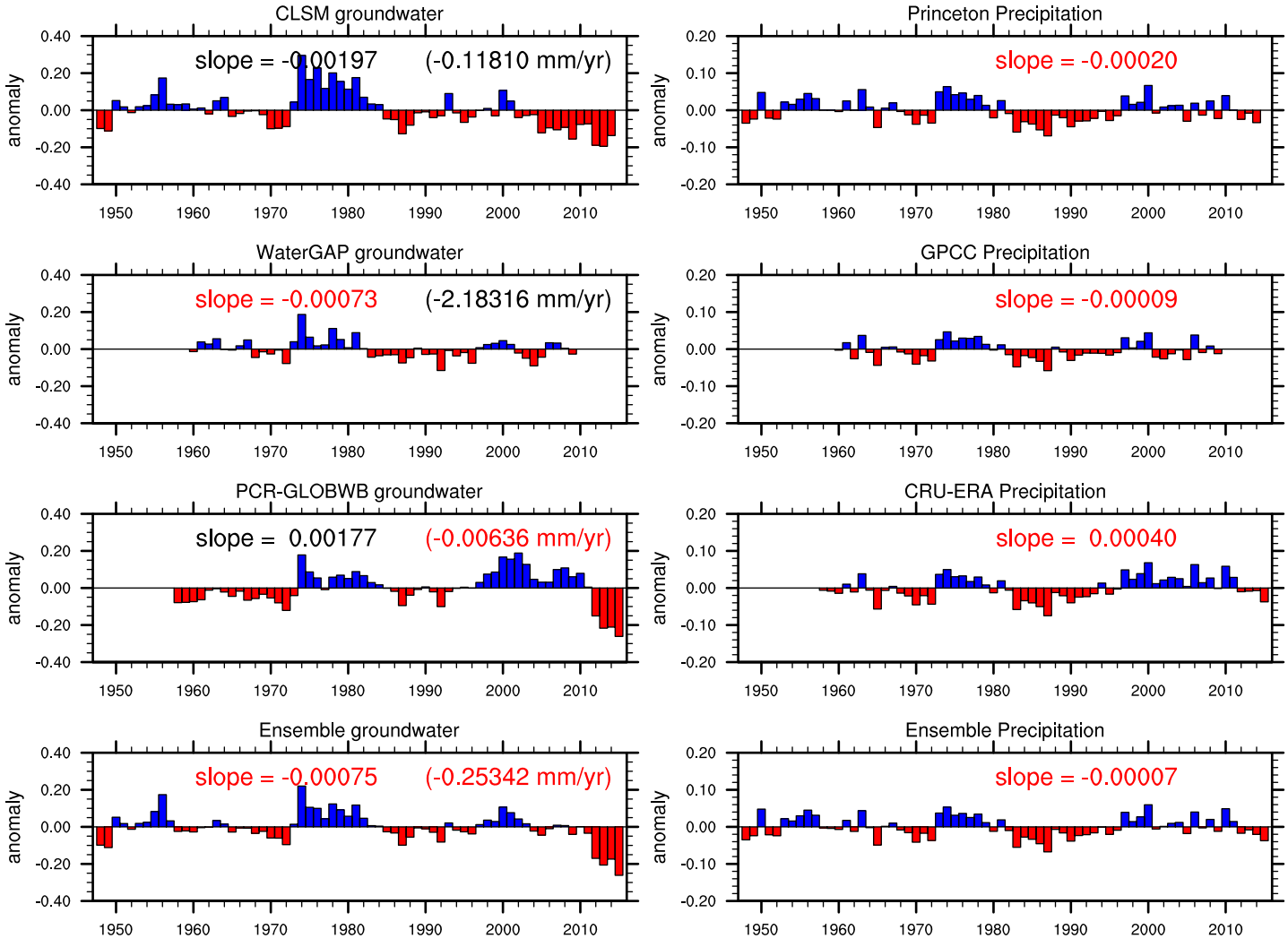

CLSM
WaterGAP
PCR-GLOBWB
90N
60N
30N
0
30S
60S
90S
180 150W 120W 90W 60W 30W 0 30E 60E 90E 120E 150E 180
-0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8


precipitation
EOF 1 (6.3%)
PC1
r = 0.48
SOI
EOF 2 (4.6%)
PC2
r = 0.64
0.018
0.012
0.006
0
-0.006
-0.012
-0.018
groundwater
EOF 1 (10.6%)
PC1
r = 0.32
EOF 2 (7.6%)
PC2
r = 0.34
1950
1970
1990
2010


precipitation
EOF 1 (5.5%)
PC1
r = 0.54
EOF 2 (4.9%)
PC2
r = 0.50
0.018
0.012
0.006
0
-0.006
-0.012
-0.018
groundwater
EOF 1 (6.6%)
PC1
r = 0.43
SOI
EOF 2 (5.6%)
PC2
r = 0.37
1970
1980
1990
2000


precipitation
groundwater
0.02
0.016
0.012
0.008
0.004
0
-0.004
-0.008
-0.012
-0.016
-0.02
EOF 1 (6.0%)
PC1
EOF 2 (5.7%)
PC2
r = 0.68
EOF 3 (4.6%)
PC3
r = 0.36
SOI
EOF 1 (11.5%)
PC1
EOF 2 (7.2%)
PC2
r = 0.18
EOF 3 (5.6%)
PC3
r = 0.36


Princeton precipitation
GPCC precipitation
CRU-ERA precipitation
CLSM groundwater
WaterGAP groundwater
PCR-GLOBWB groundwater
90N
60N
30N
0
30S
60S
90S
180 150W 120W 90W 60W 30W 0 30E 60E 90E 120E 150E 180
-0.008 -0.006 -0.004 -0.002 0 0.002 0.004 0.006 0.008
-0.03 -0.015 0 0.015 0.03


CLSM groundwater
slope = -0.00197
(-0.11810 mm/yr)
Princeton Precipitation
slope = -0.00020
WaterGAP groundwater
slope = -0.00073
(-2.18316 mm/yr)
GPCC Precipitation
slope = -0.00009
PCR-GLOBWB groundwater
slope = 0.00177
(-0.00636 mm/yr)
CRU-ERA Precipitation
slope = 0.00040
Ensemble groundwater
slope = -0.00075
(-0.25342 mm/yr)
Ensemble Precipitation
slope = -0.00007
anomaly
1950
1960
1970
1980
1990
2000
2010