UNIVERSITY OF CALIFORNIA,
IRVINE

Understanding Large-scale Plant Water Relations Using Remote Sensing and Models

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Earth System Science

by

Meng Zhao

Dissertation Committee:
Professor Isabella Velicogna, Chair
Professor James Randerson
Professor Steven Davis
Professor Amir AghaKouchak
Professor Eric Rignot

2021

DEDICATION

To my beloved parents

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Isabella Velicogna for giving me the opportunity to work in her lab. Her guidance and support make my time at ESS incredibly fulfilling and life changing experience.

I am grateful for Dr. Geruo A's collaboration and guidance on all materials presented in this dissertation. Through numerous discussions with him, I learn to think critically and creatively.

I also want to thank my dissertation committee members Dr. James Randerson, Dr. Eric Rignot, Dr. Amir AghaKouchak, and Dr. Steven Davis for all the instructive comments and suggestions along the way of my entire Ph.D. program.

I am also grateful to the ESS Department at UC Irvine, as well as the graduate student community.

Special thanks to Zhuoya He for her endless love and support.

The American Meteorological Society and Nature Springer (publishers) have permitted the reprinting of previously published work in Chapters 3 to 5 of this dissertation. Full citations are provided below:

Zhao, M., Velicogna, I., Kimball, J. S. (2017). Satellite observations of regional drought severity in the continental United States using GRACE-based terrestrial water storage changes. Journal of Climate, 30(16), 6297-6308.

Zhao, M., Velicogna, I., Kimball, J. S. (2017). A global gridded dataset of GRACE drought severity index for 200214: comparison with PDSI and SPEI and a case study of the Australia millennium drought. Journal of Hydrometeorology, 18(8), 2117-2129.

Zhao, M., Geruo, A., Zhang, J., Velicogna, I., Liang, C., Li, Z. (2021). Ecological restoration impact on total terrestrial water storage. Nature Sustainability, 4(1), 56-62.

VITA

Meng Zhao

EDUCATION

Doctor of Philosophy in Earth System Science **2021**
University of California Irvine *Irvine, CA*

Master of Science in Earth and Environmental Science **2013**
Lehigh University *Bethlehem, CA*

Bachelor of Engineering in Remote Sensing Science and Technology **2011**
Wuhan University *Wuhan, China*

RESEARCH EXPERIENCE

Graduate Research Assistant **2016–2021**
University of California Irvine *Irvine, California*

Graduate Research Assistant **2011–2013**
Lehigh University *Bethlehem, Pennsylvania*

TEACHING EXPERIENCE

Teaching Assistant **2013–2016**
University of California Irvine *Irvine, CA*

REFEREED FIRST-AUTHORED JOURNAL PUBLICATIONS

Ecological restoration impact on total terrestrial water storage **2021**
Nature Sustainability

A global gridded dataset of GRACE drought severity index for 200214: Comparison with PDSI and SPEI and a case study of the Australia Millennium drought **2017**
Journal of Hydrometeorology

Satellite observations of regional drought severity in the continental United States using GRACE-based terrestrial water storage changes **2017**
Journal of Climate

Recent ice cap snowmelt in Russian High Arctic and anti-correlation with late summer sea ice extent **2014**
Environmental Research Letters

REFEREED CO-AUTHORED JOURNAL PUBLICATIONS

Satellite detection of varying seasonal water supply restrictions on grassland productivity in the Missouri basin, USA **2020**
Remote Sensing of Environment

Multicomponent satellite assessment of drought severity in the Contiguous United States from 2002 to 2017 using AMSRE and AMSR2 **2019**
Water Resources Research

Development of a daytime cloud and haze detection algorithm for Himawari8 satellite measurements over central and eastern China **2017**
Journal of Geophysical Research-Atmosphere

ABSTRACT OF THE DISSERTATION

Understanding Large-scale Plant Water Relations Using Remote Sensing and Models

By

Meng Zhao

Doctor of Philosophy in Earth System Science

University of California, Irvine, 2021

Professor Isabella Velicogna, Chair

Plant photosynthesis plays a major role in ecosystem-climate feedback. Drought-induced declines in photosynthesis and associated forest die-off events have increased considerably in the past decades due to climate warming. Currently, there is substantial uncertainty in understanding ecosystem vulnerability to drought. This impedes an accurate projection of climate change impact on ecosystem's capability to store carbon and to provide services to humans. Many ecosystems such as forests and shrubs tend to have deep root systems. During droughts, they could access water beyond the groundwater table. Therefore, to understand the drought response in an ecosystem would require information of moisture supply in the deep soil. However, existing observations of underground water supply over large areas are limited to a shallow soil depth (<5 cm). In this dissertation, we establish a framework that uses terrestrial water storage (TWS) from the Gravity Recovery and Climate Experiment (GRACE) to complement other atmospheric and shallow surface drought indictors (e.g., precipitation, temperature, surface soil moisture) by providing information about both root zone soil moisture and groundwater changes. We use this framework to study the correspondence between vegetation status and drought intensication as it moves from early-stage precipitation shortage to shallow and deep soil layers. In addition to droughts, we evaluate how ecosystems respond to shifts in long-term mean moisture supply and demand conditions. Finally, we evaluate human ecological restoration impact on total terrestrial water

storage in northern China. This work demonstrates that GRACE TWS is a reliable proxy for total plant-available water supply and advances our understanding of large-scale plant water relations from space.

Chapter 1

Introduction

Vegetation connects terrestrial water, carbon flux, and energy cycles, and has major economic and ecological implications[e.g. *Field et al.*, 2012]. Climatic factors such as temperature, radiation, and water supply interact to impose complex and varying constraints on vegetation growth in different parts of the world[*Nemani et al.*, 2003; *Seddon et al.*, 2016]. Characterizing vegetation response to these climatic constraints are not only important for evaluating and mitigating climatic change impacts on ecosystem functions and services, but also to determine the feedback mechanisms that ecosystem response may generate on the climate itself[e.g. *Richardson et al.*, 2013].

During the past decades, observations reveal that global vegetation has changed significantly in response to the warming climate and changing precipitation patterns[e.g. *Myneni et al.*, 1997; *Allen et al.*, 2010; *Xu et al.*, 2013]. However, current land surface models have large uncertainties in capturing these changes. This is partly due to our limited knowledge of climatic constraints on ecosystem phenology and productivity[*Randerson et al.*, 2009; *Richardson et al.*, 2012]. Recent studies show that we might have underestimated the role of water in comparison to temperature in controlling vegetation phenology and productivity

as well as carbon turnover rate in land surface models[*Carvalhais et al.*; *Forkel et al.*, 2015; *Humphrey et al.*, 2018]. Thus, we need to improve our understanding of climatic constraints, especially water stress on vegetation growth on a global scale.

Observations of water availability for plant use are rare and come with large uncertainties. *In-situ* measurements of soil moisture are sparse and limited to a shallow soil depth where plant roots can penetrate deep in the soil and use groundwater and rock moisture[*Fan et al.*, 2017; *Rempe and Dietrich*, 2018]. Microwave remote sensing observations of soil moisture are also limited to a shallow soil depth (usually top few centimeters) and perform poorly in densely vegetated regions[e.g. *Njoku and Entekhabi*, 1996]. Due to the lack of direct observations, drought indices are often used as proxies of plant water availability to assess plant-water relations on a global scale [e.g. *Vicente-Serrano et al.*, 2013]. Commonly used drought indices such as the Palmer Drought Severity Index (PDSI) have various limitations, for instance model bias in approximating terrestrial water balance and uncertainties associated with reanalysis precipitation datasets[e.g. *Keyantash and Dracup*, 2002]. In addition, many of the drought indices include temperature in their calculations (e.g., PDSI), which makes it difficult to separate the effect of temperature anomaly from water stress on vegetation. These drought indices are generally sensitive to only a few hydrological components such as snow, surface water, soil moisture, or groundwater, and cannot provide a complete representation of water deficit during droughts. Observing all the relevant hydrological components is important for characterizing drought propagation, recovery, and their impacts on ecosystems, especially on deep-rooted vegetation[e.g. *Van Loon*, 2015a]

Since 2002 the Gravity Recovery and Climate Experiment (GRACE) mission and GRACE Follow-On mission have provided continuous regional estimates of monthly changes in terrestrial water storage (TWS)[*Tapley et al.*, 2004a]. These changes include variations in both soil moisture affecting vegetation growth and deeper groundwater influencing soil moisture recharge and drought recovery. GRACE TWS estimates have been used to investigate re-

gional impacts of ecological disturbances[*Maness et al.*, 2013; *Chen et al.*, 2013], and to study water constraints on global CO_2[*Humphrey et al.*, 2018], as well as on vegetation growth in Australia[*Yang et al.*, 2014], Eurasia[*A et al.*, 2015], and tropical forests[*Guan et al.*, 2015]. GRACE TWS differs from other Earth surface soil moisture observations (such as those from microwave remote sensing) in that GRACE measures changes in Earths gravity field that are caused by the redistribution of water mass both at and underneath the Earth surface[*Wahr et al.*, 1998a]. As long as the water mass change signal is strong enough over a large region, GRACE can detect it no matter at which soil depth it occurs.

This work uses monthly GRACE TWS as an indicator of plant-available water supply based on the following two justifications:1) Using GRACE TWS as a proxy for land water supply, we show that on a global scale, vegetation productivity increases with land water supply but levels off in regions with ample land water supply, such as in forest-dominated ecosystems (Chapter 2). This demonstrates that GRACE TWS captures the non-linear characteristic response of vegetation growth to water availability[*Huxman et al.*, 2004; *Ponce-Campos et al.*, 2013]. 2) We develop a new drought severity index solely based on GRACE data (GRACE-DSI). This index shows favorable agreement with traditional drought monitoring tools including the United States Drought Monitor (USDM), PDSI, standardized precipitation and evapotranspiration index (SPEI), as well as remote sensing surface soil moisture estimates, and *in-situ* groundwater observations (Chapters 3 and 4). This index also captures drought impacts on a multitude of vegetation characteristics, such as canopy greenness, productivity, and phenology (Chapters 3 and 4).

The overarching goal of this work is to use GRACE TWS measurements in conjunction with other observation-based hydrometeorological and ecological datasets to improve the process-based understanding of water stress impact on terrestrial ecosystems to better inform model developments. In chapter 2, we characterize how the spatial pattern of long-term mean vegetation productivity is shaped by the mean annual land water supply and atmospheric

moisture demand across the global climate gradient. We use these observations to evaluate the CMIP5 Earth system model (ESM) outputs. We find large inter-model discrepancies and all models fail to reproduce our observed relationship. We discuss the potential causes of these discrepancies and suggest priorities in the next generation ESM developments. In chapter 3, we develop a new drought severity index from GRACE (GRACE-DSI) and use it to study the seasonal drought impact on canopy greenness in the Continental United States. In chapter 4, we evaluate GRACE-DSI with multiple existing drought monitoring tools and characterize GRACE-DSI uncertainty on a global scale. We also combine GRACE-DSI with other satellite environmental datasets to improve the characterization of vegetation response to the 2000s Australia Millennium Drought at shallow surface and subsurface soil layers. In chapter 5, we separate the human ecological restoration impacts on TWS from natural climate variability for a restoration hotspot in northern China using GRACE and multiple other satellite observations, government reports, and eco-hydrological modeling.

Chapter 2

Interactive effect of land water supply and atmospheric moisture demand on natural vegetation productivity

2.1 Abstract

The water cycle strongly controls the spatial pattern of terrestrial ecosystem productivity. Previous studies mostly rely on sparse observations of precipitation to evaluate how water supply affects the mean condition of plant productivity neglecting the role of atmospheric moisture demand. Studies that separate the impacts of water supply versus atmospheric demand on productivity are limited to hourly to interannual timescales; therefore, they do not account for the longer-term constraint on productivity. Here, we use satellite observations of terrestrial water storage, vapor pressure deficit, and solar-induced chlorophyll fluorescence to identify the relative control of land water supply versus atmospheric demand on the spatial pattern of mean vegetation productivity during 2007-2015. On a global scale, we find that

productivity increases with rising water supply but saturates at high supply level. High atmospheric demand suppresses productivity in warmer regions but benefits productivity in colder regions through the relaxation of low temperature constraint. In warmer regions, observations show that water supply exerts a stronger control on the spatial variations of productivity and its importance increases with rising atmospheric demand. Atmospheric demand is a weaker control than water supply alone, however its importance increases with depleting water supply. We use these observations to evaluate the CMIP5 Earth system models. We find large inter-model discrepancies and all models fail to reproduce our observed relationship. Our study provides an effective means to evaluate water stress sensitivities simulated in the upcoming CMIP6 ESMs.

2.2 Main

With rising temperatures and shifting rainfall patterns, heatwaves and droughts are increasing in frequency and intensity and are exerting devastating impacts on the functioning of terrestrial ecosystems[*Reichstein et al.*, 2013]. For example, the 2003 European heatwave and drought reduced gross primary productivity (GPP) over Europe by 30% and released 0.5PgC yr^{-1} CO_2 to the atmosphere, which was equivalent to four years of net ecosystem carbon sequestration[*Ciais et al.*, 2005]. Nevertheless, heatwave and drought impacts on terrestrial ecosystems are difficult to predict because ecosystem water and carbon fluxes respond to atmospheric dryness (high atmospheric moisture demand) and soil dryness (low land water supply) in complex ways[*McDowell and Allen*, 2015; *Williams et al.*, 2013]. High atmospheric moisture demand induces stomatal closure to save water, which reduces CO2 diffusion in leaves and thus CO2 supply for photosynthesis[*Novick et al.*, 2016]. Low land water supply not only causes stomatal closure but also impairs the biochemical capacity for photosynthesis[*Stocker et al.*, 2018]. Atmospheric moisture demand and land water supply

also co-regulate the optimal stomatal behavior across biomes and influence the spatial distribution of vegetation types[*Lin et al.*, 2015]. An accurate understanding of how these two constraints affect vegetation is imperative for the mechanistic understanding of ecosystem vulnerability to drought as well as for the projection of climate change impact on terrestrial ecosystems[*Rogers et al.*, 2016].

At shorter timescales (i.e., from hourly to year-to-year), previous studies, on forest and grassland ecosystems in the United States, find that stomatal conductance and vegetation productivity are more sensitive to atmospheric moisture demand than to available land water supply[*Novick et al.*, 2016; *Konings et al.*, 2017; *Sulman et al.*, 2016]. Space for time substitution approach is commonly used to study the vegetation response at longer timescales (decadal to centennial and beyond) over which ecosystems adjust to climate change through shifts in population density and community composition[e.g. *Pickett*, 1989]. Studies based on site-level measurements focus on the impact of mean annual precipitation on vegetation productivity[*Huxman et al.*, 2004; *Ponce-Campos et al.*, 2013; *Sala et al.*, 2012]. They find that precipitation explains a large portion of the spatial variability of vegetation productivity[*Huxman et al.*, 2004; *Ponce-Campos et al.*, 2013; *Sala et al.*, 2012]. However, they do not examine the impact of atmospheric moisture demand nor the impact of mutual interaction between atmospheric moisture demand and available land water supply on the spatial pattern of vegetation productivity. Still these effects should be considered to characterize climate change impact on plant growth at decadal to centennial timescales.

Separating the impacts of these two constraints is difficult. Both *in-situ* and satellite-based studies suffer from the lack of appropriate observations of land water supply. *In-situ* measurements of soil moisture are sparse and limited to a shallow depth[*Novick et al.*, 2016; *Sulman et al.*, 2016; *Stocker et al.*, 2018] whereas plant roots can penetrate deep in the soil and use groundwater and rock moisture[*Fan et al.*, 2017; *Rempe and Dietrich*, 2018]. Satellite-based studies generally use precipitation as a proxy for available water supply[*Konings et al.*,

2017; *Giardina et al.*, 2018]. However, precipitation can be significantly different from available land water supply, especially in regions with significant drainage and soil evaporation[*Seneviratne et al.*, 2010]. In addition, the use of precipitation or shallow soil moisture in conjunction with atmospheric moisture demand often leads to co-linearity issue. Precipitation covaries with atmospheric humidity and thus correlates with atmospheric moisture demand[*Seneviratne et al.*, 2010; *Giardina et al.*, 2018]. Shallow soil moisture is correlated with atmospheric demand at timescales of weeks to years[*Novick et al.*, 2016]. Thus, *in-situ* observational studies are restricted to separating their impacts at hourly or daily timescales[*Novick et al.*, 2016; *Sulman et al.*, 2016; *Stocker et al.*, 2018].

Terrestrial water storage (TWS) changes observed from the Gravity Recovery and Climate Experiment (GRACE) satellites have been successfully used as a proxy for vegetation available water supply to investigate water constraints on land photosynthesis and CO2 assimilation[*Guan et al.*, 2015; *Humphrey et al.*, 2018]. GRACE observations track water storage variations in the entire soil column, and effectively capture changes in land water supply at all timescales[**?***Zhao et al.*, 2017a]. Hence these data provide unique water supply information that is decoupled from the transient atmospheric moisture conditions. Here we use TWS from GRACE JPL mascon solutions[*Watkins et al.*, 2015] as a proxy for plant available land water supply, vapor pressure deficit (VPD) from the Atmospheric Infrared Sounder (AIRS)[*Aumann and Pagano*, 1994] as an indicator of atmospheric moisture demand, and solar-induced fluorescence (SIF) from GOME-2[*Joiner et al.*, 2013] as a proxy for vegetation productivity. We use these data to characterize the spatial pattern of vegetation productivity averaged during 2007-2015 as a function of available land water supply and atmospheric moisture demand over the global natural vegetation domain. We use these observations to evaluate water stress sensitivities simulated in the Coupled Model Intercomparison Project Phase 5 (CMIP5) ESMs and we discuss the implications of these results for the projection of climate change impact on terrestrial ecosystem productivity.

2.3 Spatial dependence of vegetation productivity on land water supply and atmospheric moisture demand

We investigate how water supply and atmospheric demand shape the spatial distribution of mean vegetation productivity during 2007-2015. This 9-year period is the longest period for which all the satellite data sets are available at high quality. We use the annual maximum SIF ($SIFmax$) as a proxy for the annual maximum gross primary productivity ($GPPmax$). $GPPmax$ represents the seasonal maximum capacity of CO_2 uptake and reflects plant physiological constraint on annual productivity across biomes[*Xia et al.*, 2015; *Zhou et al.*, 2016]. SIF has been shown to covary with GPP at large spatial and temporal scales[*Guanter et al.*, 2014; *Sun et al.*, 2017]. Multi-year average $SIFmax$ is linearly correlated with multi-year average $GPPmax$ derived from the FLUXCOM global carbon flux data set[*Tramontana et al.*, 2016] with $R^2 = 0.80$ (Figure 2.1). We use the annual GRACE TWS change (ΔTWS) as a proxy for the available land water supply. ΔTWS quantifies the amount of water storage that is released from land annually. Thus, ΔTWS represents the maximum available land water supply for plant use[*Guan et al.*, 2015; *Gao et al.*, 2014]. We use annual maximum VPD ($VPDmax$) as a proxy of the highest atmospheric demand constraint on vegetation growth during the growing season[*Giardina et al.*, 2018]. Figure 2.2 provides an conceptual description of the analyzed metrics. GRACE data consist of a set of $3^\circ \times 3^\circ$ mascon blocks[*Watkins et al.*, 2015]. Although observed at a coarser resolution than the other datasets we use, the GRACE TWS estimates can effectively capture the water supply pattern across the global climate gradient[*Humphrey et al.*, 2018]. For consistency we average all datasets to the GRACE spatial footprint (see Methods). As we are interested in characterizing the spatial patterns of and linkages between the mean conditions during the analyzed period, we calculate the 9-year average of $SIFmax$ ($\overline{SIFmax}$),

VPDmax ($\overline{VPDmax}$), and annual water storage change ($\overline{\Delta TWS}$) (Figure 2.3).

For the natural vegetated mascons over the globle (see Methods), we find that $\overline{SIFmax}$ increases with $\overline{\Delta TWS}$, but this effect saturates for higher $\overline{\Delta TWS}$ values (Figure 2.4a). This is consistent with the fact that in wet regions water availability is not anymore a limitation for vegetation growth, and other factors such as nutrients and light might be constraining $\overline{SIFmax}$[*Huxman et al.*, 2004]. The relationship between $\overline{SIFmax}$ and $\overline{\Delta TWS}$ is consistent with the non-linear productivity-precipitation relationship obtained from site-level observations which show that productivity increases with mean annual precipitation but saturates at sites with higher mean annual precipitation[*Huxman et al.*, 2004; *Ponce-Campos et al.*, 2013].

On a global scale, we observe that $\overline{SIFmax}$ increases with $\overline{VPDmax}$ when $\overline{VPDmax}$ is lower than 1.67kPa while $\overline{SIFmax}$ decreases with $\overline{VPDmax}$ when $\overline{VPDmax}$ is higher than 1.67kPa (Figure 2.4b and see Methods). We find that this $\overline{VPDmax}$ threshold corresponds to a mean annual maximum temperature ($\overline{Tmax}$) of 24°. We define $\overline{Tmax}$=24°as the temperature threshold that divides the analyzed domain into two groups with monotonic SIF-VPD relationships (See Methods and Figure 2.4c). VPD suppresses growth in warmer mascons ($\overline{Tmax}$ >24°) but benefits productivity in colder mascons ($\overline{Tmax}$ <24°) (Figure 2.4b-c). To analyze the impact of the selection of the $\overline{Tmax}$ threshold, we vary the threshold from 22°to 26°and find that the selection of this threshold within this range does not impact our conclusions (see Methods). The colder mascons are generally at high latitudes or in high mountain regions (Figure 2.4c) where the ecosystem productivity is mostly temperature and radiation constrained[*Seddon et al.*, 2016; *Nemani et al.*, 2003]. The warmer mascons are located in arid, semi-arid, temperate and tropical regions where water supply affects plant growth[*Guan et al.*, 2015; *Seddon et al.*, 2016; *Nemani et al.*, 2003]. In the colder mascons, $\overline{SIFmax}$ and $\overline{\Delta TWS}$ are small with low spatial variability and show no clear correspondence (Figure 2.4d). In contrast, in the warmer regions, $\overline{SIFmax}$ and $\overline{\Delta TWS}$ have a much

larger variability (Figure 2.4e) and show a relationship similar to that obtained using all the natural vegetated mascons in the world (Figure 2.4a).

To investigate if $\overline{\Delta TWS}$ or $\overline{VPDmax}$ is the dominant driver that determines the spatial distribution of productivity, we regress $\overline{SIFmax}$ against $\overline{\Delta TWS}$ and $\overline{VPDmax}$ separately. We interpret the regression slopes as the water supply and atmospheric demand sensitivity, respectively. We focus our analysis in the warmer regions, because in the colder regions atmospheric demand is not a constraint on long-term plant growth (Figure 2.4b), and here we do not identify a clear relationship between water supply and plant productivity (Figure 2.4d). Before the regression analysis, we log-transform $\overline{\Delta TWS}$ to linearize the productivity versus water supply relationship (Figure 2.4e). To obtain unitless slopes, which allows direct comparison of the sensitivities, we calculate the z-scores of all variables by subtracting the mean and dividing by the standard deviation. The z-score normalization also converts both explanatory variables to a unit range of variance such that the variable with a larger magnitude of regression slope will be the dominant driver of SIF spatial variation. $\overline{\Delta TWS}$ has negligible collinearity with $\overline{VPDmax}$ (Figures 2.5a-b), hence these two proxies can be used to separate the impacts of water supply and atmospheric demand on photosynthesis. Note that while precipitation is frequently used as a proxy for available water supply together with VPD to separate the water supply and atmospheric demand control on vegetation growth[*Konings et al.*, 2017; *Giardina et al.*, 2018], we find that the long-term mean annual precipitation is significantly correlated with VPDmax in warm mascons (Figures 2.5c-d). This indicates that precipitation is not suitable to separate the relative control of land water supply from atmospheric demand on the spatial pattern of mean vegetation productivity.

The regression slope between SIFmax and TWS, i.e. the water supply sensitivity, is 0.58±0.07 (hereafter all errors indicate 95% confidence level), and the atmospheric demand sensitivity is -0.33±0.07. Hence land water supply is the dominant driver of the multi-year mean vegetation productivity across the climate gradient in the warm regions. Previous studies

find that atmospheric demand is the dominant driver of daily to inter-annual variations of vegetation growth and leaf level carbon/water fluxes both at the site scale and for the grassland biome in the mid latitudes[*Novick et al.*, 2016; *Sulman et al.*, 2016; *Konings et al.*, 2017]. Temporal studies reflect transient vegetation response to short-term perturbations in water supply and atmospheric demand[*Sala et al.*, 2012; *Lauenroth and Sala*, 1992]. Spatial studies reflect long-term response of the ecosystem functions and structures to the mean water supply and atmospheric demand conditions[*Sala et al.*, 2012; *Lauenroth and Sala*, 1992; *Pickett*, 1989]. For instance, although stomatal regulation occurs quickly upon exposure to high VPD, for some species, acclimation of stomata to long-term exposure to high VPD can counteract the short-term responses and sustain or even increase plant water and carbon fluxes[*Marchin et al.*, 2016; *Grossiord et al.*, 2017]. Short-timescale observational studies are likely neglecting these acclimation processes occurring at longer timescales. In addition they do not account for the deeper soil moisture impact on vegetation growth as deeper soil moisture changes occur at longer timescales than shallow soil moisture[*Zhao et al.*, 2017a; *Wang et al.*, 2007].

2.4 Interactive effect of water supply and atmospheric demand

Previous studies suggest that the daily variability of certain tree species growth is governed by an interaction between soil water deficit and excessive atmospheric demand, i.e. the sensitivity of growth to one of the stressor differs depending on the level of the other stressor[*Perez-Martin et al.*, 2009; *Sulman et al.*, 2016]. We investigate whether this interaction exists at longer time scales and across large spatial climate gradients. To do so, we first divide the warmer mascons into three $\overline{VPDmax}$ quantiles, denoted as low, medium, and high demand groups. Within each quantiles, we regress $\overline{SIFmax}$ against $\overline{\Delta TWS}$. We interpret

the resulting slopes as the vegetation sensitivity to water supply for different stress levels of atmospheric demand. We use the same approach to estimate the vegetation sensitivity to atmospheric demand for different water supply levels.

We find that $\overline{SIFmax}$ decreases as $\overline{\Delta TWS}$ decreases but the regression slope, i.e. the water supply sensitivity, changes with atmospheric demand (Figure 2.6a). Water supply sensitivity is significantly ($p<0.05$) larger in the high atmospheric demand group than in the low demand group (Figure 2.6b). We attribute this change in sensitivity to the different plant physiology and traits across climate gradients. For instance, we analyze a global plant rooting depth dataset upscaled from in-situ measurements[*Schenk et al.*, 2009], and we find that plants grown under higher atmospheric demand tend to grow deeper roots (Figure 2.7), which may explain their larger sensitivity to soil water supply. These results imply that the expected increase in VPD under a warming climate can make natural plants more dependent on water supply for example by changing plant functional and structural traits such as rooting depth and shifting species composition[*McDowell and Allen*, 2015; *Fan et al.*, 2017]. We find that $\overline{SIFmax}$ decreases as $\overline{VPDmax}$ increases (Figure 2.6c). The plant sensitivity to $\overline{VPDmax}$ is significantly ($p<0.05$) larger in the low water supply group than in the high supply group where the sensitivity is not significant ($p>0.05$) (Figure 2.6d). This indicates that higher water supply may alleviate and potentially negate the expected productivity decline associated with the increase of atmospheric moisture stress.

These results document the interactive effects of long-term mean water supply and atmospheric demand on the spatial variability of vegetation productivity. We characterize the vegetation sensitivities to the climate gradient across natural ecosystems, spanning a larger spatio-temporal range than in previous in-situ short-timescale studies[*Sulman et al.*, 2016] and manipulated experiments[*Perez-Martin et al.*, 2009]. Current projections of climate change impact on vegetation productivity suggest that VPD will play a dominant role in limiting future carbon fluxes[*Williams et al.*, 2013; *Novick et al.*, 2016; *Restaino et al.*, 2016;

McDowell and Allen, 2015; *McDowell et al.*, 2016]. However, our results indicate that a potential rise in future VPD will enhance vegetation sensitivity to available land water supply, emphasizing the increasing importance of land water supply constraint on carbon fluxes under a warming climate. In addition, we find that vegetation sensitivity to VPD variations depends on the available land water supply. Hence, given the large uncertainty in predicting future terrestrial water supply patterns[*Lemordant et al.*, 2018; *Berg et al.*, 2017], it is challenging to accurately project the VPD impact on productivity.

2.5 Assessment of Earth system models

We use our results to evaluate vegetation sensitivity to water stress simulated in the CMIP5 ESMs (Table 2.1). In the previous section we separately evaluate the vegetation sensitivities to water supply and atmospheric demand as a function of different ranges of atmospheric demand and water supply, respectively (Figure 2.6). We now simultaneously estimate these sensitivities using a linear regression model that includes a multiplicative interaction term: $y \sim \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$ (eq.1), where x1, x2 and y are available land water supply, atmospheric moisture demand and vegetation productivity respectively. The interactive term $x_1 x_2$ accounts for the water supply sensitivity dependence on atmospheric demand, and vice versa. We use this regression model to compare the sensitivities from the observational record and the CMIP5 ESMs. For the models, we use GPP in lieu of SIF. As CMIP5 ESMs do not have an explicit TWS field, we calculate TWS from model hydrological fluxes (See Methods). In the regression analysis we use z-scores for all variables. Figure 2.8a shows the water supply sensitivity ($\beta_1 + \beta_3 x_2$) as a function of atmospheric demand and Figure 2.8b the atmospheric demand sensitivity ($\beta_2 + \beta_3 x_1$) as a function of water supply.

All models fail to reproduce the observed sensitivity and we find large differences between the model sensitivities (Figure 2.8). These results imply major model uncertainties in character-

izing the plant sensitivity to land water supply and atmospheric demand. In addition, most of the models underestimate the sensitivity to water supply under high atmospheric moisture demand and overestimate the sensitivity to water supply under low atmospheric moisture demand (Figure 2.8a). They also underestimate the atmospheric demand sensitivity in low water supply conditions and overestimate the atmospheric demand sensitivity in high water supply conditions (Figure 2.8b). Five of the nine ESMs show much larger collinearity between land water supply and atmospheric demand than the observations (Supplementary Table 2). This suggests that these CMIP5 ESMs overestimate the spatial coupling of long-term mean conditions of land water supply and atmospheric demand. These results indicate that most CMIP5 ESMs do not correctly represent the interactions between photosynthesis response, land water supply and atmospheric demand.

2.6 Conclusion

This work expands on prior site-level studies by using state-of-the-art satellite observations to separate the impact of land water supply vs. atmospheric moisture demand on the spatial distribution of the natural vegetation productivity. Our spatial model includes changes in water supply, atmospheric demand, vegetation structure, density, community composition, and biogeochemistry along large climate gradients. We show that at the ecosystem scale, during 2007-2015, vegetation productivity is more sensitive to available land water supply than to atmospheric moisture demand in warmer regions (mean annual maximum temperature higher than 24°C). We find that mean available land water supply and mean atmospheric demand conditions interact to shape the spatial pattern of vegetation productivity. We show that most of the CMIP5 ESMs do not capture this interaction and underestimate the water supply sensitivity under high atmospheric demand conditions. Given the fact that CMIP5 ESMs project a universal increase in VPD by the year 2100 under the Representative Con-

centration Pathways 8.5 experiment[*Lemordant et al.*, 2018], our results imply that CMIP5 ESMs can underestimate the land water supply impact on future ecosystem productivity. We present a new methodology to evaluate Earth system models on a global scale that can be used to assess the upcoming new generation of CMIP6 ESMs that include more realistic physical processes to better simulate vegetation water stress sensitivities such as an explicit representation of plant hydraulic processes linking water supply and atmospheric demand variations[*Bonan et al.*, 2014; *Sperry and Love*, 2015; *Xu et al.*, 2016].

2.7 Method

2.7.1 Satellite Data Sets

We use satellite data for the period 2007-2015 which is the longest period for which all satellite data are available at high quality. We use monthly GRACE terrestrial water storage (TWS) anomalies from the Jet Propulsion Laboratory (JPL) RL05 version 2 Mascon solutions[*Watkins et al.*, 2015]. Each mascon solves for the monthly gravity field averaged over an equal-area 3° × 3° mass concentration block (mascon). Every mascon is uncorrelated with neighboring mascons. Missing month in the GRACE solutions are interpolated from previous and following months' values using a quadratic spline. We use the amplitude of the annual TWS change (ΔTWS=TWSmax-TWSmin, Figure 2.2) as a proxy for annual available water supply. The mean water supply condition, $\overline{\Delta TWS}$, is calculated as the average of the ΔTWS over the study period. By propagating the uncertainty of the GRACE TWS observations[*Watkins et al.*, 2015], we find an average relative error of about 5% for $\overline{\Delta TWS}$(Figure 2.9).

We use level-3 SIF retrievals on a 0.5° × 0.5° grid from the GOME-2 optical spectrometer on board the MetOP-A (version 26)[*Joiner et al.*, 2013]. We use SIFmax as a proxy for

vegetation productivity. Compared to the annual sum of SIF values, SIFmax retrieval has the largest signal-to-noise ratio and is less likely contaminated by clouds and aerosols. We use $1^\circ \times 1^\circ$ gridded level-3 monthly near-surface relative humidity (RH, unit: %) and near-surface air temperature (T, unit: Celsius) from the ascending (daytime) orbit of the AIRS platform[*Aumann and Pagano*, 1994] to compute annual maximum vapor pressure deficit (VPDmax, unit: kPa) and annual maximum air temperature (Tmax). We calculate monthly VPD using the Clausius-Clapeyron equation: $VPD = (1 - RH) \times 0.61078 \times e^{\frac{17.27 \times T}{T + 237.3}}$(eq. 2). We use multi-year mean VPDmax ($\overline{VPDmax}$) rather than the growing season mean VPD or annual mean VPD because $\overline{VPDmax}$ yields the largest explanatory power on $\overline{SIFmax}$. T, VPD and SIF data sets are spatially averaged to match the GRACE resolution, i.e. $3^\circ \times 3^\circ$ mascon. This spatial averaging procedure reduces the noise in AIRS and GOME-2 data by a factor of 3 and 6, respectively. Multi-year average in the calculation of $\overline{Tmax}$, $\overline{VPDmax}$ and $\overline{SIFmax}$ further minimizes the impact of satellite measurement error on our analysis.

We calculate the areal extent of each land cover type within every $3^\circ \times 3^\circ$ mascon using the MODIS International Geosphere-Biosphere Programme (IGBP) land cover classification product (MCD12Q1)[*Friedl et al.*, 2010]. Because anthropogenic interventions such as cultivar improvements and irrigation can have a significant impact on photosynthesis and water resources in agricultural regions[*Haddeland et al.*, 2014; *Lobell et al.*, 2014], we only consider 589 mascons (Figure 2.4c) with over 75% area extent covered by natural vegetation types (i.e. land cover types excluding water bodies, barren or sparsely vegetation grounds, croplands, urban and built-up lands, cropland/natural vegetation mosaics and permanent snow and ice).

2.7.2 $\overline{VPDmax}$ and $\overline{Tmax}$ threshold

We first evaluate the $\overline{VPDmax}$ threshold that divides the overall SIF-VPD relationship into two monotonic ones (Figure 2.4b). To do so, we calculate the regression of $\overline{SIFmax}$ against $\overline{VPDmax}$ using a 2-segment piecewise linear model, and we define the turning point of the piecewise model as the $\overline{VPDmax}$ threshold. We then evaluate the correspondence between $\overline{VPDmax}$ and $\overline{Tmax}$ across all the mascons, which approximates an exponential form (Figure 2.10); we define the $\overline{Tmax}$ threshold as the mean annual maximum temperature corresponding to the $\overline{VPDmax}$ threshold. We find a $\overline{VPDmax}$ threshold of 1.67 $\pm$ 0.06kPa (mean $\pm$ 1 standard error) from the piecewise linear regression, which corresponds to a $\overline{Tmax}$ threshold of 24° C (Figure 2.10). We also test a range of $\overline{Tmax}$ thresholds from 22° C to 26° C and find that the choice of the threshold within this range does not affect our results (Figure 2.11).

2.7.3 Ancillary Data Sets

We use global rooting depth estimates from the International Satellite Land Surface Climatology Project Initiative (ISLSCP) II Data Collection[*Schenk et al.*, 2009] to explore the potential association between rooting depth and vegetation sensitivity to water supply. The ISLSCP rooting depth estimates are up-scaled from a global database of over 500 point measurements of vertical root profiles using global gridded datasets of land cover, climate, and soil characteristics. We use the depth that contains 95% of all roots at 1° grid resolution. We aggregate the dataset to GRACE 3° $\times$ 3° mascon resolution. We split warmer mascons ($\overline{Tmax}$ >24°) into three quantiles, corresponding to low, medium, and high $\overline{VPDmax}$ groups. We calculate the average rooting depth within each group in Figure 2.7.

We use mean annual precipitation from the ERA-Interim[*Dee et al.*, 2011] to illustrate the strong co-linearity between mean annual precipitation and $\overline{VPDmax}$ (Figure 2.5). Mean an-

nual precipitation is averaged over the same period as $\overline{VPDmax}$ and aggregated to GRACE footprint prior to analysis. Because of the co-linearity, we can not use mean annual precipitation and $\overline{VPDmax}$ to disentangle the relative impacts of water supply and atmospheric demand on plant growth.

2.7.4 $\overline{SIFmax}$-$\overline{GPPmax}$ relationship

We use GPP estimates from the FLUXCOM global carbon flux data set *Tramontana et al.* [2016] to evaluate the relationship the spatial patterns of $\overline{SIFmax}$ and $\overline{GPPmax}$ on a global scale. The FLUXCOM GPP is estimated based on three machine learning algorithms. These algorithms are then forced with gridded input driver data to estimate carbon fluxes at a 0.5° grid and daily time step. We use the mean GPP from the three algorithms and calculate the $\overline{GPPmax}$ during 2007-2013, a common period with the GOME-2 SIF retrievals. We find that $\overline{SIFmax}$ shows a strong linearity with $\overline{GPPmax}$ across space (2.1)

2.7.5 CMIP5 model simulations

We use monthly ESM Representative Concentration Pathway 8.5 (RCP8.5) outputs from an individual ensemble member (r1i1p1) of the CMIP5 collection. We use GPP in lieu of GOME-2 SIF. Because CMIP5 models do not output an explicit TWS field, we calculate TWS by accumulating model fluxes of precipitation, evapotranspiration and totoal runoff[*Rodell et al.*, 2004]. We calculate VPD from model outputs of near-surface relative humidity and near surface air temperautre using eq.2. We only use models that output all analyzed variables (Table 2.1). For consistency we aggregate model outputs to GRACE spatial footprint before analysis. We average GPPmax, ΔTWS, and Tmax over the same period as the observational record. We select warm mascons (modelled $\overline{Tmax}$ >24°C) and use the same regression model (eq.1) to quantify the water stress sensitivitiy simulated in ESMs. The regression results

are summarized in Supplementary Table 2. We also calculate the modelled water stress sensitivity using the warm mascons as defined from observations (aqua colored mascons in Figure 2.4c) and we find this does not change our conclusions drawn from Figure 2.8.

Table 2.1: Summary of analyzed CMIP5 models

Modeling Center (or group)	**Institute ID**	**Model Name**
Beijing Climate Center, China Meteorological Administration	BCC	BCC-CSM1.1
College of Global Change and Earth System Science, Beijing Normal University	GCESS	BNU-ESM
Canadian Centre for Climate Modelling and Analysis	CCCMA	CanESM2
National Center for Atmospheric Research	NCAR	CCSM4
NOAA Geophysical Fluid Dynamics Laboratory	NOAA GFDL	GFDL-ESM2M
NASA Goddard Institute for Space Studies	NASA GISS	GISS-E2-H
Institute for Numerical Mathematics	INM	INM-CM4
Institut Pierre-Simon Laplace	IPSL	IPSL-CM5A-LR IPSL-CM5A-MR

Table 2.2: Eq.1 Regression results of observations and CMIP5 RCP8.5 model outputs. All errors represents the 95% confidence level.

	β_1	β_2	β_3	Regression Model R^2	Collinearity between land available water supply and atmospheric demand, R^2
Observed	0.60±0.07	-0.29±0.07	0.22±0.08	0.59	0.08
BCC-CSM1.1	0.27±0.09	-0.55±0.09	-0.07±0.10	0.40	0.001
BNU-ESM	0.75±0.06	-0.34±0.05	-0.11±0.05	0.80	0.13
CanESM2	0.56±0.09	-0.06±0.09	0.07±0.09	0.35	0.14
CCSM4	0.59±0.06	-0.35±0.06	-0.19±0.05	0.85	0.37
GFDL-ESM2M	0.82±0.07	-0.05±0.07	-0.06±0.07	0.65	0.001
GISS-E2-H	0.89±0.07	-0.03±0.07	-0.07±0.06	0.81	0.44
INM-CM4	0.40±0.08	-0.58±0.08	-0.04±0.08	0.76	0.34
IPSL-CM5A-LR	1.01±0.06	0.29±0.06	0.15±0.04	0.83	0.41
IPSL-CM5A-MR	1.03±0.06	0.30±0.06	0.14±0.04	0.82	0.42

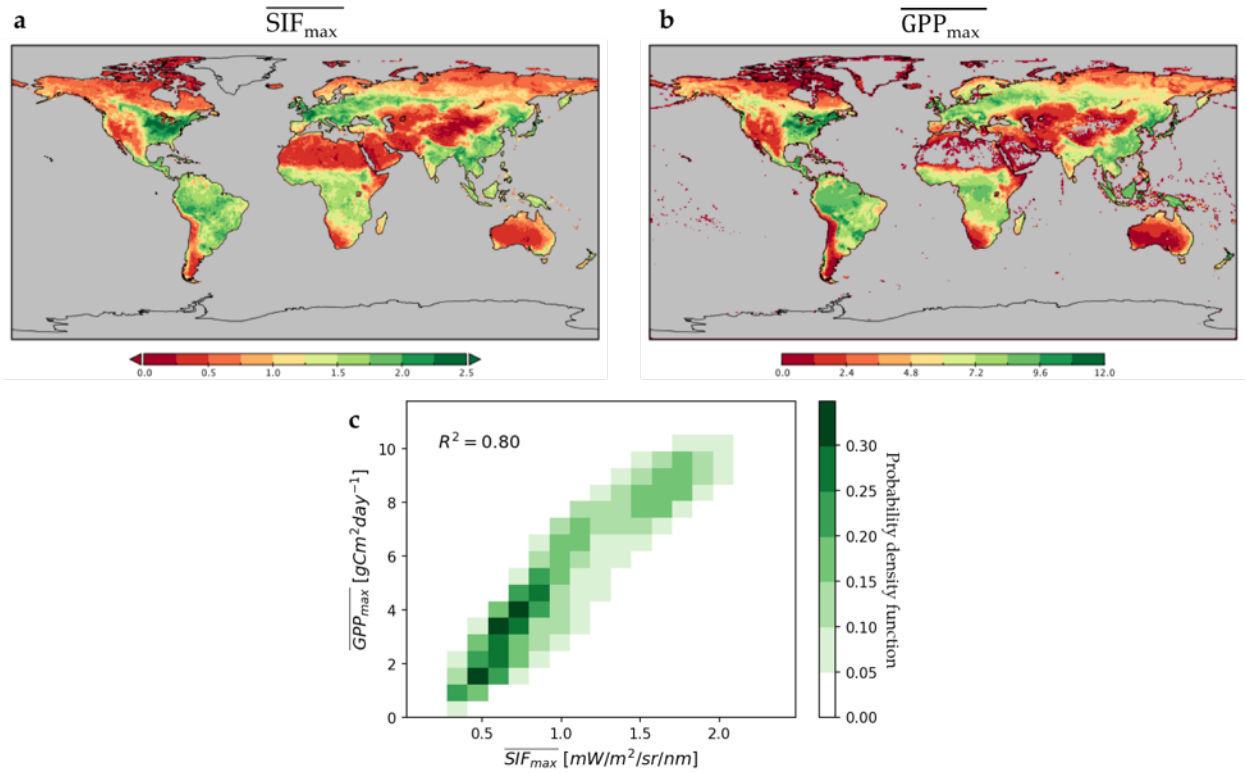

Figure 2.1: $\overline{SIFmax}$- $\overline{GPPmax}$ relationship. (a) is multi-year average annual maximum SIF from GOME-2. (b) is multi-year average annual maximum GPP from the FLUXCOM. (c) is joint probability density function between and for global 0.5 grid cells with over 75% area extent covered by natural vegetation.

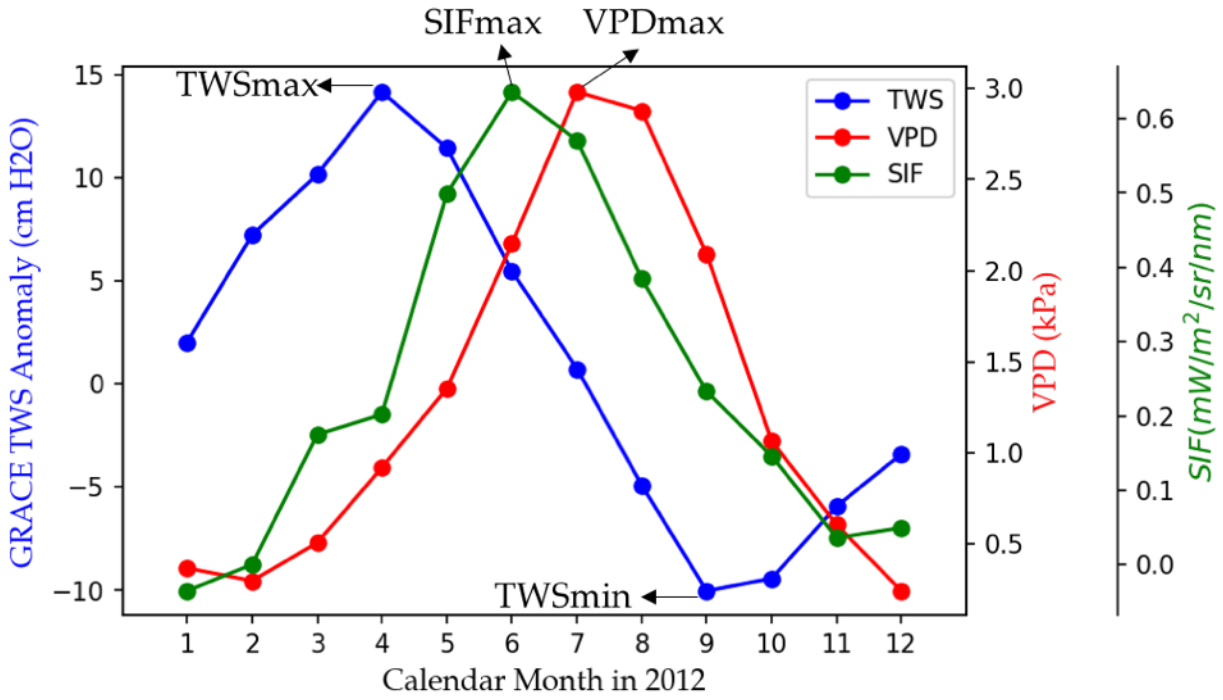

Figure 2.2: Time series example at a location in western US centered at (-115.28W, 42N) for 2012. GRACE does not measure absolute water storage. GRACE data are anomalies relative to the 2004.0-2009.999 time-mean baseline. Rather, we use ΔTWS, i.e., TWSmax-TWSmin, to represent the maximum land water supply for plants. Large ΔTWS likely suggest heavy consumption of water supply from storage such as from deeper soil or groundwater storages. Similar concept has been used by *Gao et al.* [2014]; *Guan et al.* [2015].

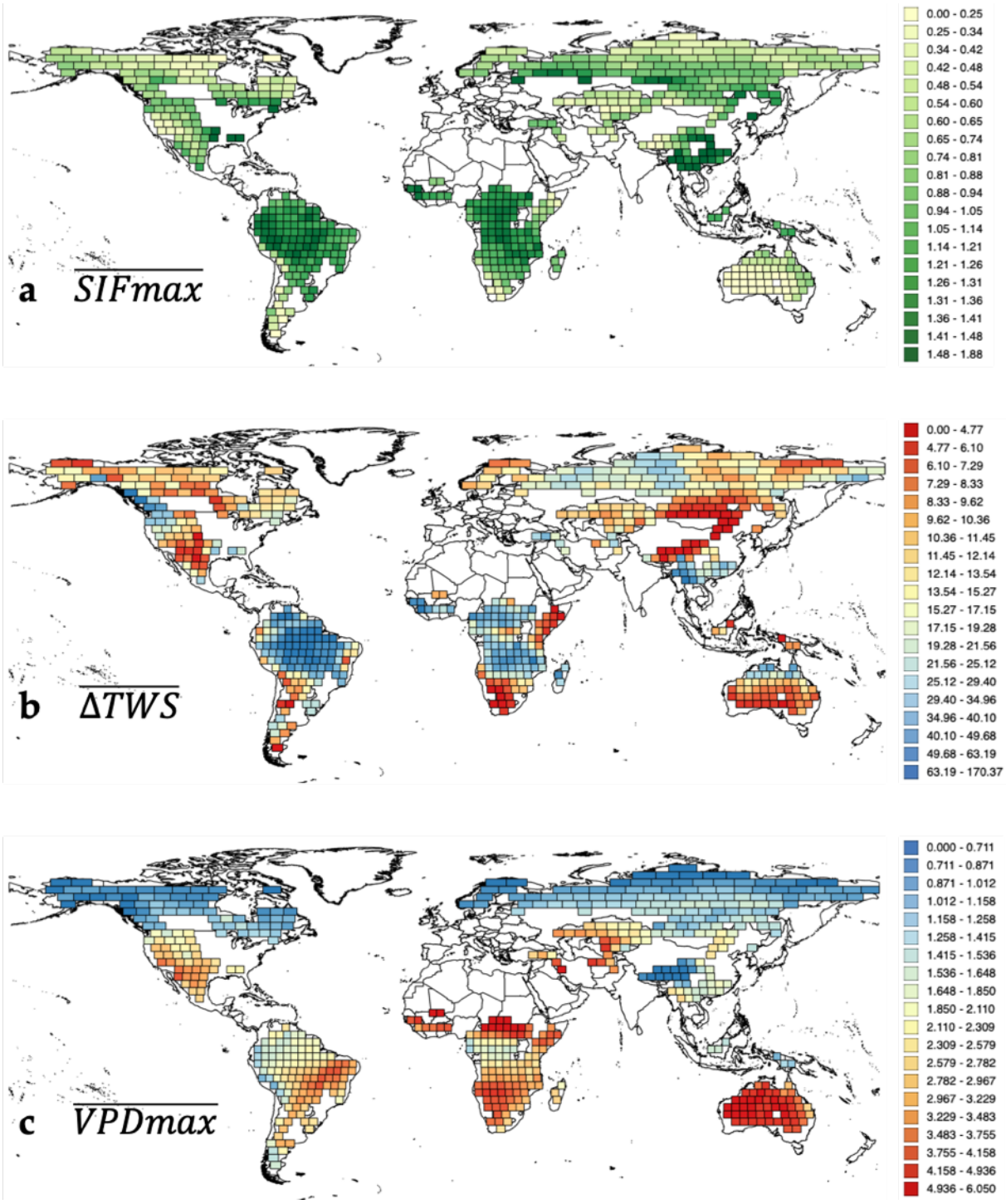

Figure 2.3: $\overline{SIFmax}$ mW/m^2/sr/nm (a), $\overline{\Delta TWS}$ cm water equivalent (b), and $\overline{VPDmax}$ kPa (c) during 2007-2015.

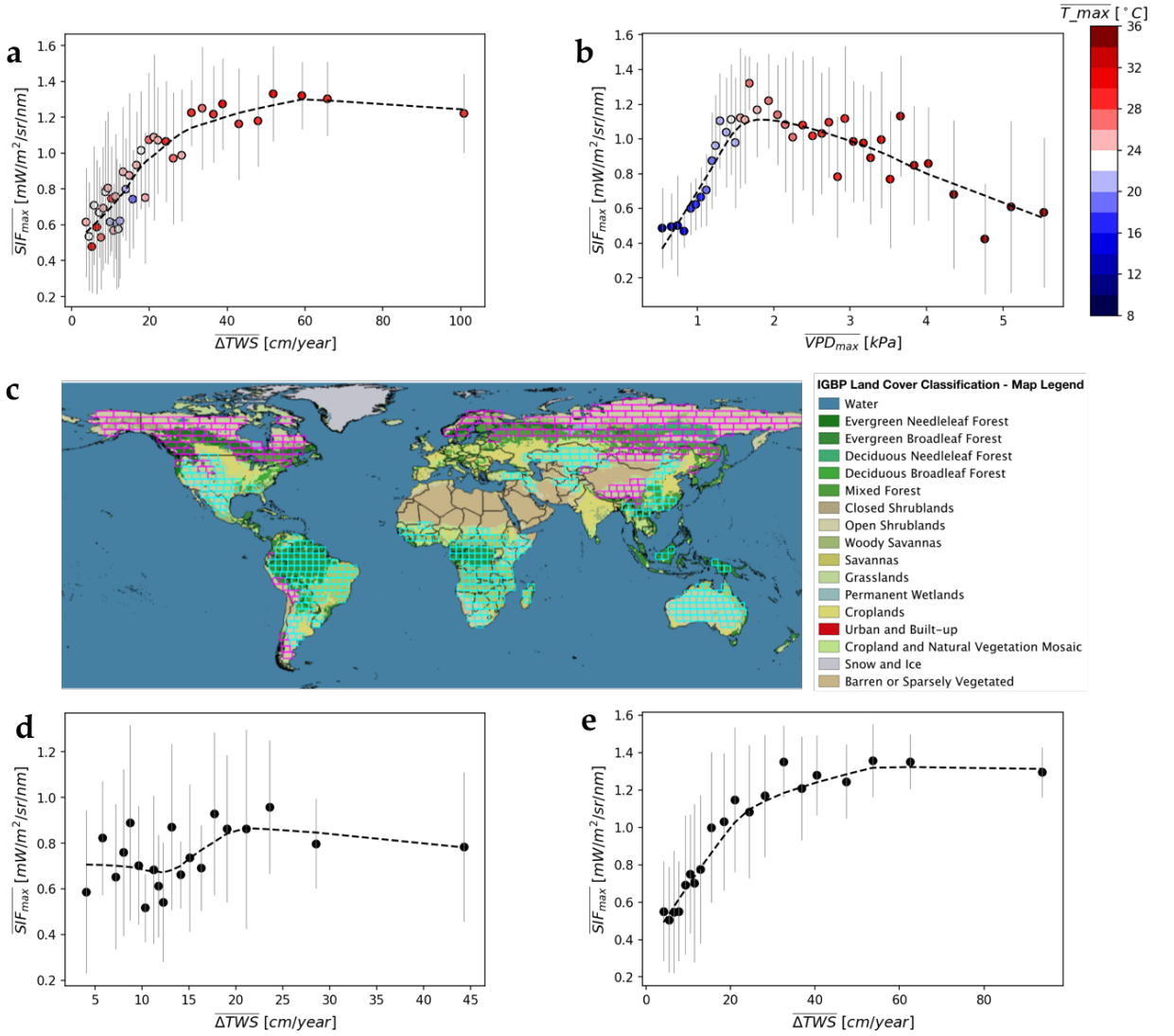

Figure 2.4: Global distribution of vegetation productivity in relation to land water supply and atmospheric moisture demand. a, scatterplot of $\overline{\Delta TWS}$ and $\overline{SIFmax}$. All natural vegetation dominant GRACE mascons are grouped into 40 equal-sized bins based on $\overline{\Delta TWS}$. Circle and error bar denote the mean and standard deviation of $\overline{SIFmax}$ within each bin, respectively. Color represents the mean $\overline{Tmax}$ within each bin. Dashed line is a LOWESS smoothing curve[*Cleveland*, 1979]. b, same as a but replacing $\overline{\Delta TWS}$ with $\overline{VPDmax}$. c, global distributions of cold mascons (magenta) and warm mascons (aqua) based on a $\overline{Tmax}$ threshold of 24°C. The base map is land cover map from the MODIS MCD12Q1 product. d-e, scatterplots of $\overline{\Delta TWS}$ and $\overline{SIFmax}$ for cold and warm mascons, respectively. Cold and warm mascons are grouped into 20 equal-sized bins.

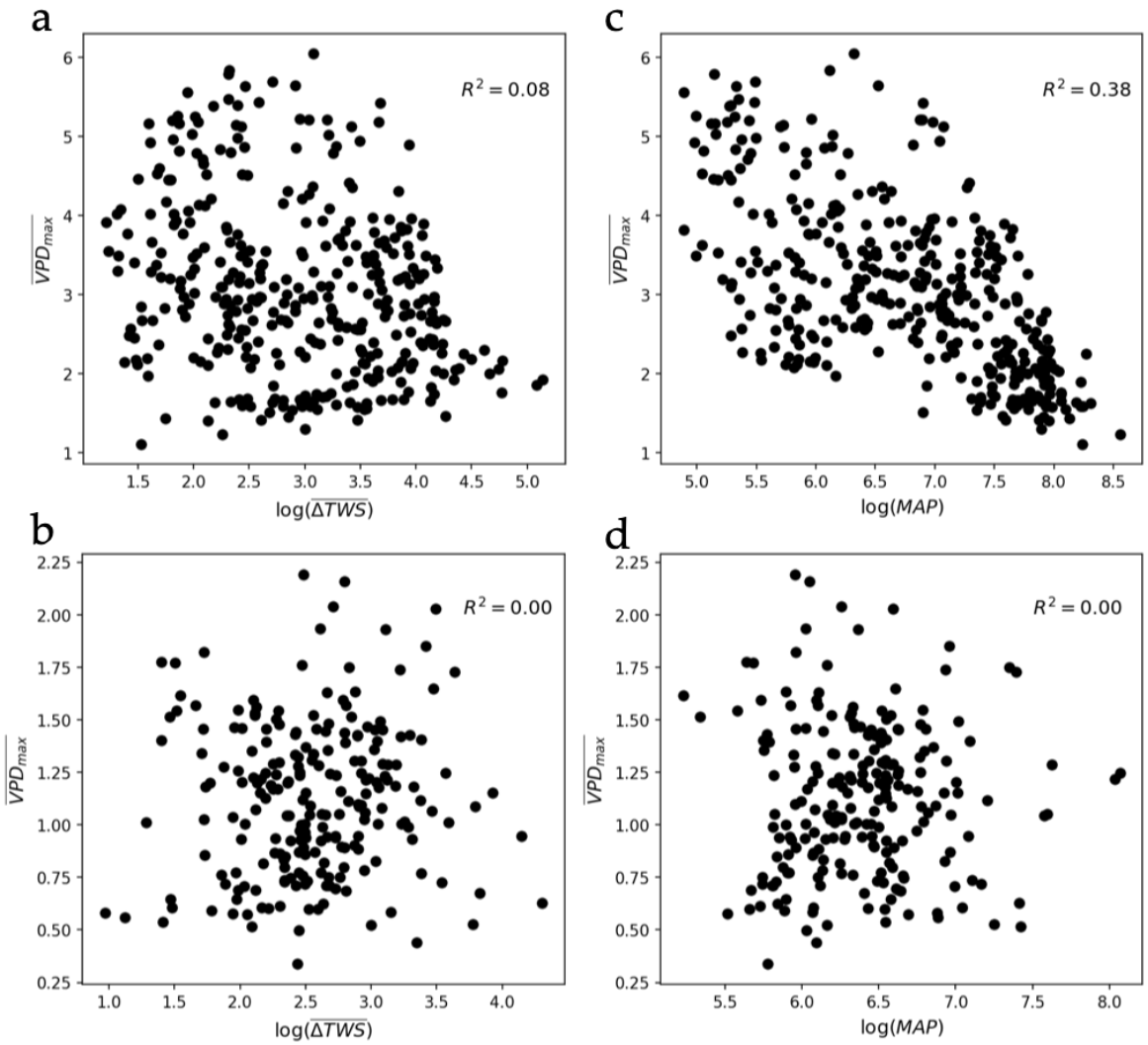

Figure 2.5: Co-linearity analysis. a and b are the scatterplots between log($\overline{\Delta TWS}$) and $\overline{SIFmax}$ for warm and cold mascons, respectively, based on a $\overline{Tmax}$ threshold of 24°C. c and d are the scatterplots between mean annual precipitation (MAP) and $\overline{SIFmax}$ for warm and cold mascons, respectively, based on a threshold of 24°C. MAP is derived from the ERA-Interim.

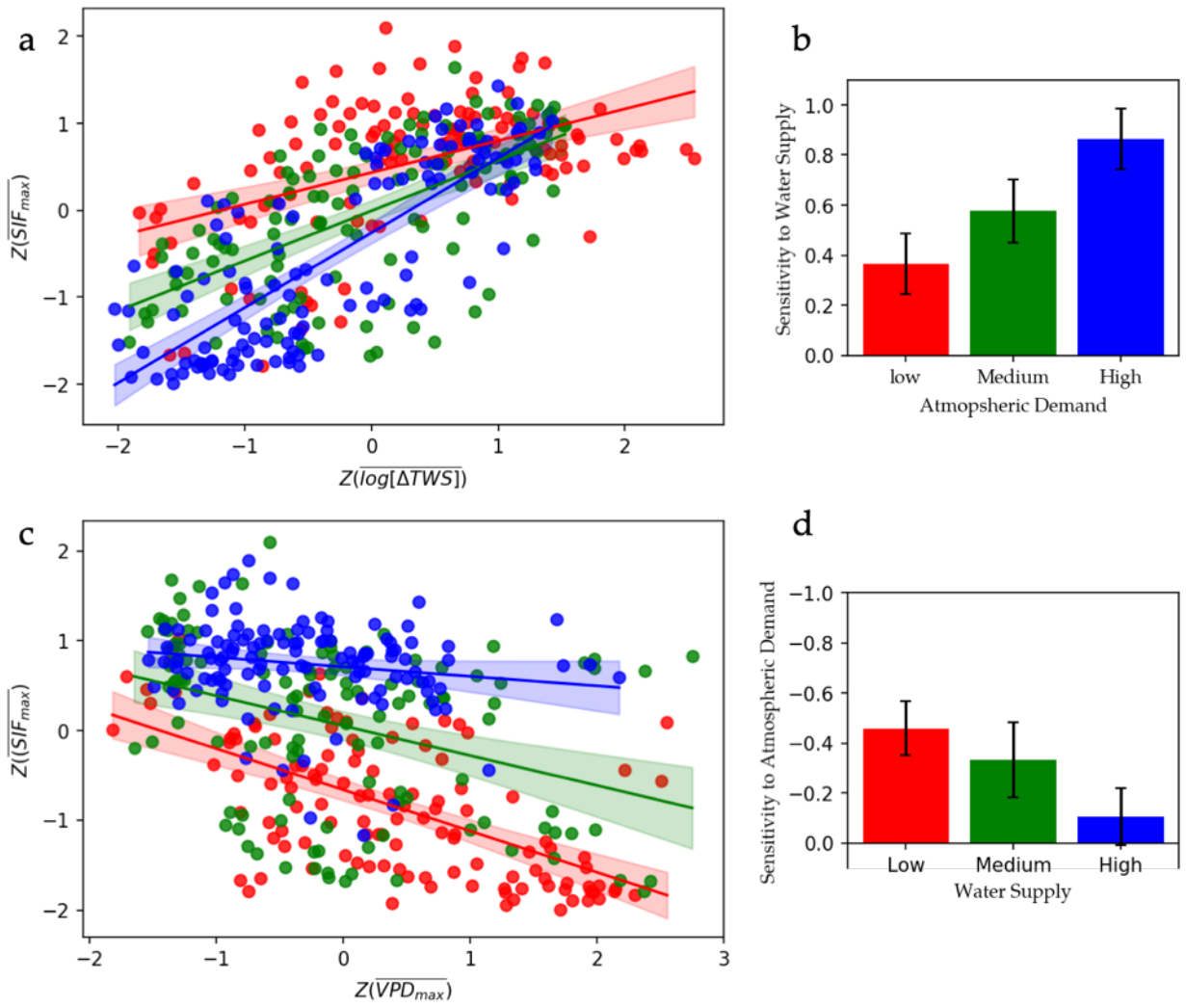

Figure 2.6: Interactive effect of water supply and atmospheric demand on vegetation productivity. a, scatterplot of z-scores of $\log(\overline{\Delta TWS})$ and $\overline{SIFmax}$ for three warm mascon groups characterized by $\overline{VPDmax}$ quantiles, denoted as low (red), medium (green) and high (blue) demand groups. b, barplot of the regression slope for each group in a, i.e., water supply sensitivity at different stress levels of atmospheric demand. c, scatterplot of z-scores of $\overline{VPDmax}$ and $\overline{SIFmax}$ for three warm mascon groups characterized by $\log(\overline{\Delta TWS})$ quantiles, denoted as low (red), medium (green) and high (blue) supply groups. d, barplot of the regression slope for each group in c, i.e., atmospheric demand sensitivity at different stress levels of land water supply. Solid lines in a and c represent linear regression models with 95% confidence intervals (shaded colors). Error bars in b and d represent the 95% confidence levels of the sensitivities.

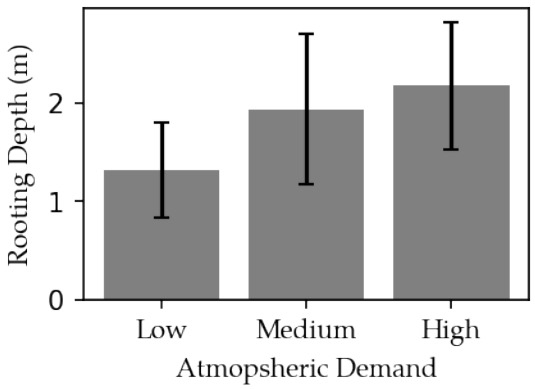

Figure 2.7: Rooting depth in relation to atmospheric moisture demand. We split global warm mascons into three quantiles, corresponding to low, medium, and high $\overline{VPDmax}$ groups. Each gray column represents the average depth that contain 95% of all roots and the corresponding error bar represents one standard deviation of the rooting depths within each group.

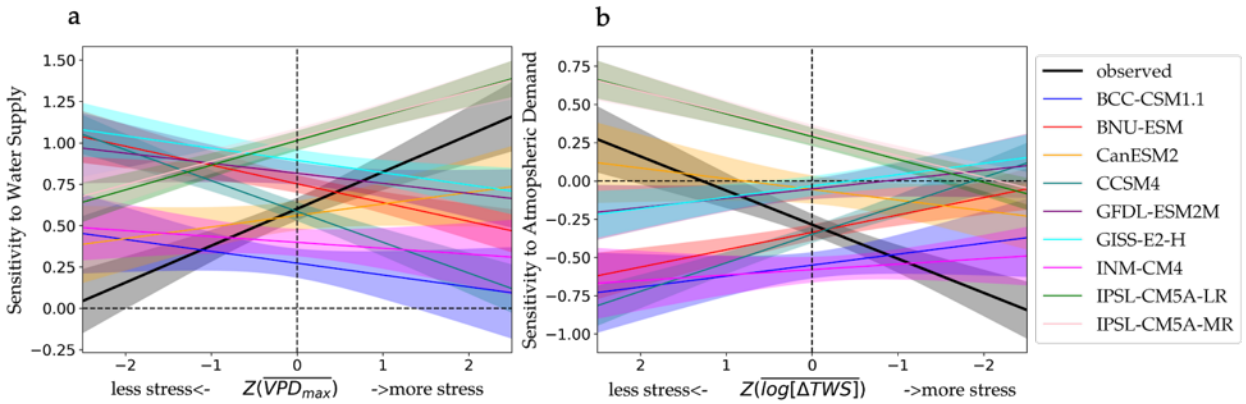

Figure 2.8: Comparison of vegetation sensitivity to water stress between observational and Earth system model results. a, sensitivity to water supply conditioned on various atmospheric demand levels, i.e., $\beta_1 + \beta_3 * Z(\overline{VPDmax})$ from eq.1. b, sensitivity to atmospheric demand conditioned on various water supply levels, i.e., $\beta_2 + \beta_3 * Z(log(\overline{\Delta TWS}))$ from eq.1. Shaded color represents the 95% confidence level of the estimated sensitivity.

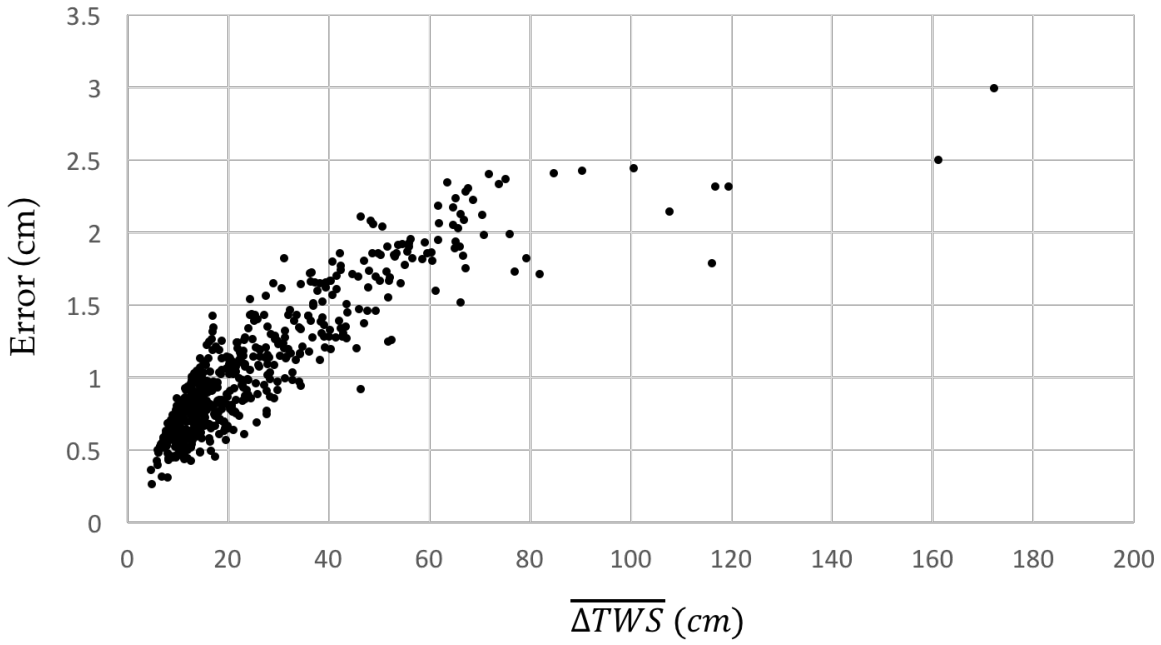

Figure 2.9: Error of $\overline{\Delta TWS}$. Each dot represents $\overline{\Delta TWS}$ and its error at each mascon. We estimate the error of $\overline{\Delta TWS}$ by Monte Carlo simulation. At each mascon, we generate an ensemble of 1000 normally distributed zero mean random noise time series (σ=GRACE error provided by JPL). We compute the $\overline{\Delta TWS}$ for each of those 1000 simulations and set the one standard deviation of resulting simulations as the error of $\overline{\Delta TWS}$.

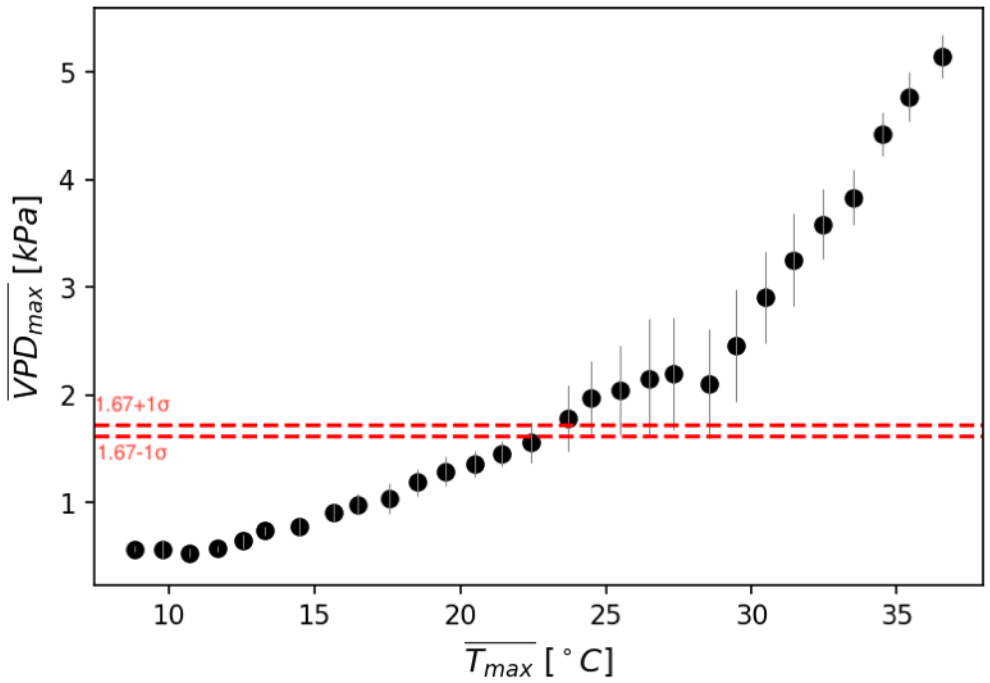

Figure 2.10: The relationship between $\overline{VPDmax}$ and $\overline{Tmax}$. All analyzed mascons are binned based on $\overline{Tmax}$. Circle and error bar denote the mean and standard deviation of $\overline{VPDmax}$ within each bin, respectively. The horizontal dashed red lines indicate the $\overline{VPDmax}$ threshold ($\pm$ 1 error) that divides the global SIF-VPD relationship (Figure 2.4a in the main text) into two monotonic ones.

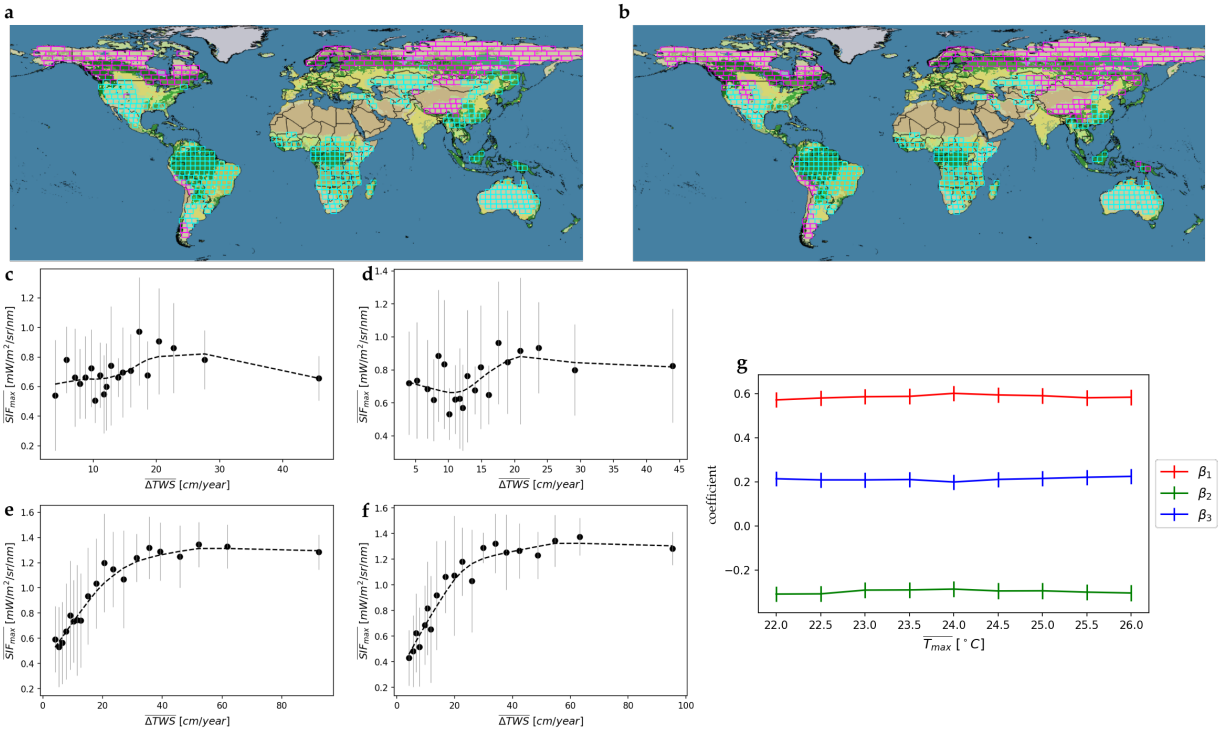

Figure 2.11: Temperature threshold sensitivity test. (a) is the distribution of cold mascons (magenta) and hot mascons (aqua) based on a $\overline{Tmax}$ threshold of 22°C. (b) is the same as (a) but for a $\overline{Tmax}$ threshold of 26°C. (c-d) is the relationship between $\overline{SIFmax}$ and $\overline{\Delta TWS}$ for cold mascons in (a-b), respectively. (e-f) is the relationship between $\overline{SIFmax}$ and $\overline{\Delta TWS}$ for hot mascons in (a-b), respectively. Dashed lines are LOWESS smoothing curves. (g) is the regression coefficient of eq. 1 in the main text as a function of threshold ranging from 22°C to 26°C.

Chapter 3

Satellite observations of regional drought severity in the continental United States using GRACE-based terrestrial water storage changes

As appears in:

Zhao, M., G. A, I. Velicogna, and J.S. Kimball, "Satellite Observations of Regional Drought Severity in the Continental United States Using GRACE-Based Terrestrial Water Storage Changes," *Journal of Climate*, 30, 6297-6308, 2017.

3.1 Abstract

Drought monitoring is important for characterizing the timing, extent, and severity of drought for effective mitigation and water management. Presented here is a novel satellite-based drought severity index (DSI) for regional monitoring derived using time-variable terrestrial water storage changes from the Gravity Recovery and Climate Experiment (GRACE). The GRACE-DSI enables drought feature comparison across regions and periods, it is unaffected by uncertainties associated with soil water balance models and meteorological forcing data, and it incorporates water storage changes from human impacts including groundwater withdrawals that modify land surface processes and impact water management. Here, the underlying algorithm is described, and the GRACE-DSI performance in the continental United States during 2002–14 is evaluated. It is found that the GRACE-DSI captures documented regional drought events and shows favorable spatial and temporal agreement with the monthly Palmer Drought Severity Index (PDSI) and the U.S. Drought Monitor (USDM). The GRACE-DSI also correlates well with a satellite-based normalized difference vegetation index (NDVI), indicating sensitivity to plant-available water supply changes affecting vegetation growth. Because the GRACE-DSI captures changes in total terrestrial water storage, it complements more traditional drought monitoring tools such as the PDSI by providing information about deeper water storage changes that affect soil moisture recharge and drought recovery. The GRACE-DSI shows potential for globally consistent and effective drought monitoring, particularly where sparse ground observations (especially precipitation) limit the use of traditional drought monitoring methods.

3.2 Introduction

Drought indices are convenient ways to characterize drought because they compress the complexity of the drought phenomenon into a single number. Commonly used indices are generally sensitive to only a few hydrological components and cannot provide a complete representation of the water deficit during drought. For instance, the Palmer Drought Severity Index (PDSI) is sensitive to atmospheric moisture demand and near-surface soil moisture content, and has been frequently used as a measure of meteorological and soil moisture drought[*Mishra and Singh*, 2010; *Trenberth et al.*, 2014]. Observing all relevant hydrological variables (i.e., snow, surface water, soil moisture, and groundwater) is important for fully characterizing drought propagation and recovery, and associated ecosystem impacts[*Van Loon*, 2015b]. Previous studies have used Gravity Recovery and Climate Experiment (GRACE)-derived terrestrial water storage (TWS) estimates to examine regional-scale droughts[*Yirdaw et al.*, 2008; *Chen et al.*, 2009; *Leblanc et al.*, 2009; *Long et al.*, 2013; *Castle et al.*, 2014; *Cao et al.*, 2015] and to quantify drought-induced water storage deficits[*Thomas et al.*, 2014]. The drought index for Texas in *McCandless* [2014] combines TWS, precipitation, and the satellite-based normalized vegetation difference index (NDVI) and it is sensitive to vegetation drought response over semiarid areas. These methods are not suitable for comparing drought features for different locations and time periods as they do not account for the spatiotemporal variability of local hydroclimate. For example, the same amount of water deficit may have a larger impact on arid and humid biomes than on semiarid and semihumid biomes[*Vicente-Serrano et al.*, 2013]. The same amount of water deficit may also induce more severe damage to vegetation during reproductive growth stages than during green-up and senescence[*Ji and Peters*, 2003]. *Houborg et al.* [2012] accounted for these differences by deriving region-specific cumulative distribution of dry and wet conditions from GRACE data assimilation system, but this approach may not be readily useable outside North America.

To overcome these limitations, we develop a new standardized drought severity index (DSI)

based solely on GRACE TWS estimates, herein referred to as GRACE-DSI. Here, we introduce the GRACE-DSI algorithm, evaluate how it captures the space and time evolution of documented regional drought events during 2002–14, compare it with the PDSI and the U.S. Drought Monitor (USDM), and demonstrate its synergistic use with traditional drought monitoring tools across the continental United States (CONUS). The CONUS domain was selected owing to a dense surface station network, reliable PDSI, and diverse climate and vegetation conditions. We also compare the GRACE-DSI against the NDVI from the Moderate Resolution Imaging Spectroradiometer (MODIS) used as a proxy for vegetation growth changes. Finally, we discuss the merits of combining the GRACE-DSI with other datasets for drought characterization and potential applications for monitoring water supply and ecosystem interactions in other areas.

3.3 Data and Methodology

3.3.1 Data

We use release-05 GRACE gravity solutions in the form of spherical harmonic coefficients truncated to degree 60, from the Center for Space Research at the University of Texas, for the period April 20022014. The GRACE-derived C_{20} coefficients are replaced with satellite laser ranging estimates[*Chen et al.*, 2013]. We include degree-1 coefficients calculated as in *Swenson et al.* [2008]. We correct the glacial isostatic adjustment signal using *A et al.* [2013]. To reduce correlated errors, we filter each monthly field following *Swenson and Wahr* [2006]. We smooth the Stokes coefficients using a 350-km radius Gaussian averaging function [*Wahr et al.*, 1998a] and calculate regular $1^\circ \times 1^\circ$ latitudemonthly TWS mass anomalies relative to the 2002–14 mean.

We use the monthly self-calibrated PDSI from *Dai* [2011] on a global 2.5° grid. The PDSI

uses a two-layer bucket model to assess soil water balance by accounting for water supply and demand[*Palmer*, 1965]. *Dai et al.* [2004] found good correlations of PDSI with soil moisture observations over the United States, the former Soviet Union, Mongolia, and China, and with streamflow over major global river basins.

We use remotely sensed monthly soil moisture (SM) data from the European Space Agency Climate Change Initiative[*Liu et al.*, 2011, 2012]. The SM record employs passive and active microwave satellite data with improved spatial and temporal coverage and resolution.

We use the cloud-free MODIS monthly Climate Model Grid 0.05° NDVI product (MOD13C2; *Huete et al.* [2002]) as a measure of vegetation activity. We only use data from the growing season (April–October) to avoid snow-related NDVI noise[*Ji and Peters*, 2003; *Karnieli et al.*, 2010; *Mu et al.*, 2013; *A et al.*, 2015].

We use the USDM weekly shapefiles distributed by the National Drought Mitigation Center (http://droughtmonitor.unl.edu). The USDM integrates information from many existing drought indicators, including the PDSI, along with local reports from state climatologists and observers across the country. The shapefiles are converted to monthly raster composites in a 1° grid. For computational purposes, the USDM drought classes are mapped to numerical values with no drought assigned a value of 0, D0 = 1 (abnormally dry), D1 = 2 (moderate drought), D2 = 3 (severe drought), D3 = 4 (extreme drought), and D4 = 5 (exceptional drought).

3.3.2 GRACE-DSI

For each 1° grid cell, we calculate the GRACE-DSI as the standardized anomalies of GRACE TWS for month j and year i, as GRACE-DSI$_{i,j}$=(TWS$_{i,j}$-$\langle TWS_j \rangle$)/σ_j, with i ranging from 2002 to 2014, where $\langle TWS_j \rangle$ and σ_j are the mean and standard deviation of TWS anomaly

for month j, respectively. The global GRACE-DSI follows a pseudo-standard normal distribution. We classify the GRACE-DSI into five drought categories (Table 3.1) by matching their ranking percentiles to thresholds used by the USDM (i.e., 30%, 20%, 10%, 5%, 2%)[*Svoboda et al.*, 2002]. For example, the cumulative relative frequency for GRACE-DSI less than -2.0 is 2%; therefore, we set -2.0 as the upper cutoff value for exceptional drought. The GRACE-DSI detects both drought and abnormally wet events (Table 3.1).

3.3.3 GRACE-DSI evaluation

Because of the truncation and smoothing applied to reduce short scale errors, each GRACE-DSI grid cell represents conditions averaged over a 350-km radius footprint. For consistency, all datasets are processed as the GRACE datathat is, converted in spherical harmonic, truncated to degree 60, filtered, smoothed using a 350-km Gaussian averaging function, and converted in the spatial domain on a $1^\circ \times 1^\circ$ lat/lon grid. This process preserves the time and spatial variability of the original signal but reduces its amplitude[*Velicogna and Wahr*, 2006]. To minimize the impact of these changes on the data comparisons, we standardize PDSI and SM data relative to the GRACE-DSI period (2002–14), herein referred to as PDSI-Z and SM-Z, and we employ the same drought classification scheme as for the GRACE-DSI (Supplementary Materials).

We compare the spatial patterns of GRACE-DSI, PDSI-Z, and USDM during the record-setting 2011–12 drought[*Hoerling et al.*, 2014]. We examine the temporal correspondence between the GRACE-DSI and PDSI-Z over the CONUS using correlation analysis. The USDM is a discrete metric and does not characterize wetter-than-average conditions, preventing a rigorous comparison with GRACE-DSI through correlation analysis. Instead, we map continuous GRACE-DSI to categorical series following the classification scheme in Table 1 and convert the positive side of GRACE-DSI to the no-drought category. At each

grid cell, we calculate the ratio of the number of months for which both GRACE-DSI and USDM have the same drought category, relative to the total number of months. We rescale GRACE-like processed USDM using the method described in *Landerer and Swenson* [2012] and round the rescaled value up to the nearest drought category.

We calculate the correlation coefficients between GRACE-DSI and NDVI for each growing season month separately, as the vegetation drought response varies considerably within different phenological stages and thus cannot be adequately represented by simple correlation or time series comparison without accounting for the seasonal effect[*Ji and Peters*, 2003; *Karnieli et al.*, 2010; *A et al.*, 2015; *Forkel et al.*, 2015].

In the GRACE-DSI algorithm, we normalize the TWS deficit by the regional hydroclimatological variability to account for the fact that a TWS deficit of a given magnitude may indicate different drought levels in an arid or humid region. To illustrate this, we show that the same USDM-classified drought severity level corresponds to different TWS deficits depending on the regional hydroclimatology. We focus on the 2011–12 drought and at each grid cell we define for each month the corresponding drought category using the USDM drought classification (D1-D4 drought categories). We then categorize the months into four groups each corresponding to one of the drought classifications. For each group, we identify the month when the maximum TWS deficit occurred, and we calculate the corresponding TWS standard deviations using the entire analyzed period, 2002–14. We investigate the spatial patterns of the maximum TWS deficit in relation to the monthly TWS standard deviation and the regional precipitation climatology.

3.4 Results

The GRACE-DSI captures documented regional drought events during the past decade in the CONUS (Supplementary Materials and Figure 4.8). We also find a remarkable agreement in spatial pattern between the GRACE-DSI, PDSI-Z, and USDM over different seasons and across a wide range of land covers, despite the fact that vegetation, snow, and other cold season land surface processes are not explicitly treated in the PDSI model[*Dai*, 2013; *van der Schrier et al.*, 2013]. In 2011, GRACE-DSI and PDSI-Z capture similar drought and wet spatial extent across the CONUS especially from February to August (Figure 5.1). The USDM does not provide wetter-than-average information and thus does not reflect the wetting pattern in 2011. When dry weather conditions expand northward after August[*NOAA*, 2012], the PDSI-Z and USDM detect dryness in the western Great Lakes, upper Mississippi Valley, parts of the northern plains, and the far West, whereas the GRACE-DSI indicates above- or near-normal underlying TWS in these regions. In 2012, a strong heat wave hit the CONUS[*Wang et al.*, 2014]. The PDSI-Z and USDM indicate severe drought conditions for all of 2012. In contrast, the GRACE-DSI shows a persistent water storage surplus throughout the year in the Northwest.

We compare GRACE-DSI and PDSI-Z monthly time series at six drought locations representing different land cover and climate zones (Figure 5.2). While we find overall favorable correspondence between GRACE-DSI and PDSI-Z across the CONUS, the two indices also provide complementary information regarding surface and total water supply conditions. In California (location 4), both indices capture the water deficit during 2007–09 and the exceptional drought starting in 2012 (Figure 5.2e). In late 2014, the PDSI returns to normal values after a short-term rainfall increase[*NOAA*, 2015], while the GRACE-DSI is still indicative of exceptional drought conditions, in agreement with the ongoing groundwater crisis and the extraordinary cumulative precipitation deficit[*Famiglietti*, 2014; *Savtchenko et al.*, 2015].

The monthly GRACE-DSI and PDSI-Z correlation is significant ($p < 0.01$) over the entire CONUS (Figure 5.3a). The correlation is strongest in the South and Southeast and has relatively smaller magnitude over the northwestern plains, consistent with the spatial and temporal comparisons in Figs. 1 and 2. To investigate the relatively lower correlation in the northwestern Plains, we compare both indices with the satellite SM record and water table measurements from three wells in the USGS groundwater climate response network (http://groundwaterwatch.usgs.gov) (Figure 5.3a). These locations are not impacted by pumping or injection, have 10 years of measurements, and the well records are good representatives of regional groundwater variability in the Missouri watershed[*Reager et al.*, 2015]. The GRACE-DSI and PDSI-Z show overall consistent temporal variations, but with a time lag and a shift in magnitude (Figures 5.3b). In 2004–08, following the 2000–04 drought[*Schwalm et al.*, 2012], the PDSI-Z indicates an overall wetting with episodic dryness consistent with the SM variations, whereas the GRACE-DSI indicates drier-than-normal conditions in alignment with observed negative anomalies in groundwater level during this time period despite the difference in spatial scale of GRACE-DSI and well measurements. This is in agreement with a previous study[*Anderson et al.*, 2013] of the northwestern Plains Snake River subarea, where long-term hydrological drought persisted during 2004–08 as shown by the USDM despite increasing surface SM and decreasing evapotranspiration deficit from temporary surface wetting.

Following the warming trend in early 2012[*Wang et al.*, 2014], the PDSI-Z decreases and reaches exceptional drought in the summer while SM declines later and remains above normal until spring (Figs. 3b). The GRACE-DSI decreases later than the PDSI-Z and by a smaller amount, which is consistent with the near-normal to positive groundwater levels throughout 2012.

We evaluate the agreement between GRACE-DSI, PDSI, and USDM in terms of drought severity classification. We calculate the percentage of the number of months for which the

USDM and GRACE-DSI (Figure 5.4a), USDM and PDSI-Z (Figure 5.4d), and USDM and PDSI (Figure 5.4g) display the same drought category. We repeat the same calculation with the USDM index biased by one (Figures 5.4b,e,h) and two drought categories (Figures 5.4c,f,i) to determine the percentage of the total number of months for which GRACE-DSI, PDSI-Z, and PDSI underestimate the USDM drought classification by one or two categories respectively (note that the original no-drought observations remain the same). We find that the GRACE-DSI, PDSI-Z, and original PDSI all tend to underestimate USDM-classified drought by one to two categories in the western United States (Figure 5.4).

The GRACE-DSI and NDVI correlation (Figure 5.5) is stronger and more widespread during June–September than at the beginning (April–May) or the end (October) of the growing season. At lower latitudes, the correlation becomes significant earlier and extends later in the year compared to higher-latitude areas.

Figure 4.6 shows the spatial distribution of the maximum TWS deficit for each USDM-classified drought category during 2011–12 and the corresponding monthly TWS standard deviation. We find that it takes a larger TWS deficit to reach the same USDM-classified drought category in the East, where the climate is humid and the TWS variability is large, than in the West where the climate is drier and the TWS variability is small. This spatial pattern agrees with the eastdecreasing gradient in long-term mean and variability of annual precipitation (Figure 4.9), illustrating that it is important to normalize TWS deficit with regional hydroclimatological variability in the GRACE-DSI algorithm for drought severity categorization and drought comparison across space and time.

3.5 Discussion

The GRACE-DSI is based on direct measurements of soil water balance that account for water supply and demand, which enables a globally consistent hydrological drought monitoring. The GRACE-DSI differs from *Thomas et al.* [2014] and other indices using GRACE alone because it accounts for the regional variability in TWS that directly affects the characterization of drought (Figure 4.6). For instance, *Thomas et al.* [2014] observe a peak TWS deficit of 66 km3 in southeastern United States in November 2007 and a peak TWS deficit of 68 km3 in Texas in January 2013, both of which equal about a 9-cm water-equivalent deficit. The January TWS variability is larger in Texas than November TWS variability in southeastern United States, which means that Texas naturally experiences a wider range of dryness and wetness in January than the southeastern United States in November. Therefore, the GRACE-DSI ranks the same 9-cm water-equivalent deficit one to two categories milder in Texas than in the southeastern United States (Figures 5.2b,c). This result agrees with the USDM classification of these two drought events (Figure 4.10).

The GRACE-DSI provides consistent and complementary information to the PDSI that strengthens the analysis of drought conditions. The GRACE-DSI is sensitive to terrestrial water storage changes, whereas the PDSI is responsive to surface SM and atmospheric moisture deficits*Dai* [2011]. The GRACE-DSI can lag the PDSI by 1 month in detecting drought onset and recovery (e.g., Figures 5.2g), which is consistent with the different response time of near-surface conditions and overall water storage*Van Loon* [2015b]. During the 2012 drought in the northwestern Plains, the PDSI decreased earlier than SM probably because of warming-induced atmospheric moisture stress as the PDSI is sensitive to air temperature[*Hu and Willson*, 2000]. In contrast, the GRACE-DSI decreases later than SM, indicating more rapid depletion of surface SM than deeper groundwater. An apparent drought recovery is also detected by both GRACE-DSI and PDSI in spring 2013, consistent with previous studies[*Hoerling et al.*, 2014; *Wang et al.*, 2014]. During the analyzed period,

GRACE-DSI and PDSI also detect different trends in the northwestern Plains associated with changes in overall and shallow-depth water balance, respectively (Figures 5.3b). We find that for location 1 (Figure 5.3b), for instance, their temporal correlation increases from 0.25 ($p < 0.01$) to 0.44 ($p < 0.001$) after prewhitening (Supplementary Materials and Figure 4.11). For the rest of the CONUS, correlation between prewhitened GRACE-DSI and PDSI time series is significant but generally of smaller magnitude compared to those shown in Figure 5.3, indicating that both drought indices capture consistent trends.

The GRACE-DSI and PDSI-Z both underestimate the severity of USDM-classified drought by one to two categories in the western United States (Figure 5.4). The mismatch pattern between the original PDSI (normalized using 1950–79 as the baseline period) and USDM is consistent with the patterns between PDSI-Z and USDM and between GRACE-DSI and USDM. This indicates that the short normalization baseline period (2002–14) in GRACE-DSI and PDSI-Z does not account for their differences with USDM. Instead, the composite nature of USDM and the inclusion of subjective information from local experts might account for its mismatch with GRACE-DSI and PDSI[*Anderson et al.*, 2011; *Hao and Singh*, 2015].

The favorable GRACE-DSI and NDVI correspondence indicates that GRACE-DSI is sensitive to water supply constraints influencing vegetation growth. The correlation is higher in the middle of the growing season than during green-up and senescence when solar radiation and temperature are important factors controlling vegetation growth[*Ji and Peters*, 2003; *Karnieli et al.*, 2010; *Forkel et al.*, 2015]. The GRACE-DSI also captures the characteristic latitudinal shift of the vegetationrelationship[*Karnieli et al.*, 2010; *Yi et al.*, 2010], whereby the NDVI correlation weakens from lower to higher latitudes due to increasing cold temperature and energy constraints on vegetation growth. Similar correlation patterns are also found between the PDSI-Z and NDVI (Supplementary Materials and Figure 3.12).

The 2002–14 period may be too short to characterize the full range of dryness and wetness for which a climatology of at least 30 years is desired. We use Levenes test[*Levene*, 1960]

to evaluate the difference in PDSI variability of the 13-yr (2002–14) period from a 33-yr (1982–2014) climatology. We find that in 40% of the CONUS, the standard deviations of PDSI are different between the two periods ($p < 0.1$); these regions have experienced a narrower range of wetness and dryness in 2002–14 than in 1982–2014. We calculate the root-mean-square error between PDSI-Zs referenced to those two periods. We find that, using the 13-yr climatology, we underestimate the drought by one category in 40% of the CONUS due to the fact that the 2002–14 period is drier than 1982–2014 (Figure 4.7a). Therefore, we infer that using the 13-yr period to calculate GRACE-DSI does not affect drought category characterization in the majority of the CONUS.

Because of the relatively short data record, using all available GRACE solutions to estimate the monthly TWS climatology mean and standard deviation in GRACE-DSI calculation is currently optimal. This limitation may change the characterization of past drought events when new GRACE solutions become available. To understand how the baseline affects the GRACE-DSI results, we calculate seven alternative GRACE-DSI records using the TWS mean and standard deviation calculated from baseline lengths ranging from 6-yr (2002–07) to 12-yr (2002–13) sequentially. The spatial patterns of these alternative GRACE-DSI records are very similar in the CONUS. Figures 4.7b and c show the area-weighted annual spatial correlation coefficients (R) and Nashefficiency coefficients (NS; [*Nash and Sutcliffe*, 1970]) between these alternative GRACE-DSIs and the 2002–14 baseline GRACE-DSI over the CONUS. When there are at least 9 years of GRACE record, the GRACE-DSI does not change significantly over longer baseline periods. In other words, the characterization of a drought event in 2010 using the 2002–10 baseline would not change significantly when the GRACE-DSI is updated four years later using the 2002–14 baseline. The GRACE-DSI can eventually be normalized to a fixed baseline when the GRACE record will expand to a climatological length.

3.6 Conclusions

We present a novel drought severity index (DSI) derived solely from GRACE satellite observations. The large footprint of GRACE-DSI makes it useful for regional- to global-scale hydrological drought assessment. We demonstrate that the GRACE-DSI complements traditional drought metrics such as the PDSI by providing complementary information about deeper water storage changes which affect soil moisture recharge and drought recovery. This is of potential use in drought propagation research, the knowledge of which is imperative to the prediction of hydrological drought[*Van Loon*, 2015b]. This is also of potential use in developing and improving composite and multi-indicator drought models[*AghaKouchak et al.*, 2015], such as the USDM, with which the GRACE-DSI has good spatial agreement in drought detection. In addition, the GRACE-DSI includes moisture variations from the plant root zone and the correspondence between GRACE-DSI and MODIS NDVI over the CONUS manifests spatial and seasonal characteristics of water supply constraints influencing vegetation growth. The GRACE-DSI can therefore be useful to study waterrelations. The GRACE-DSI also captures human impacts on drought and water resource management, which makes it advantageous to study drought in the Anthropocene[*Sivapalan et al.*, 2012; *Van Loon et al.*, 2016]. Combing GRACE-DSI with traditional drought metrics might enable a partition of drought impact into natural and anthropogenic components. The GRACE-DSI is also of use to the hydrologic modeling community because it provides an independent observation benchmark for evaluating model-based drought monitoring tools. Currently, GRACE rapid solutions are generated using the L1B data product within 24 h of data acquisition, but with limited data availability. For the GRACE Follow-On mission scheduled for launch between December 2017 and January 2018, this product will become a standard level 3 product, thereby providing near real-time information for operational drought monitoring.

3.7 Supplementary Materials

3.7.1 Standardization of PDSI

Original self-calibrated PDSI[*Dai*, 2011] is normalized with respect to the long-term climatology from 1950 to 1979, and hence cannot be directly compared to GRACE-DSI which is normalized to the period from 2002 to 2014. To account for this inconsistency, we standardize the PDSI relative to the same period used for GRACE-DSI, i.e. we define the standardized PDSI for year i and month j as PDSI-$Z_{i,j}$=($PDSI_{i,j}$-$\langle PDSI_j \rangle$)/σ_j, where $\langle TWS_j \rangle$ and σ_j are the monthly mean PDSI value and standard deviations calculated from the same years and months used for GRACE-DSI.

The global PDSI-Z calculated relative to 2002-2014 follows a pseudo normal distribution similar to the one for the global GRACE-DSI, therefore we can use the same drought classification scheme for both indices. This standardization process has been employed in other drought index studies, e.g. *Anderson et al.* [2011, 2013]; *Mu et al.* [2013]. The PDSI-Z can underestimate drought intensity by up to one category compared to the original PDSI because the 2002-2014 period is relatively dry compared to the CONUS long-term climatology[*Anderson et al.*, 2011; *Dai*, 2013].

3.7.2 Examination of past drought events using the GRACE-DSI

Figure 4.8 shows the monthly GRACE-DSI from 2002-2014. The index captures the 2002 drought over much of the CONUS eastern seaboard, northwest, central and northern Plains. From 2003 to early 2004, the index detects the continuing drought and its recovery across much of the west, the central and northern Plains and the central Great Lakes. In 2004, the GRACE-DSI captures the protracted hydrological drought in the western US*Schwalm et al.*

[2012] even though this multiyear hydrological drought is not well delineated by climate-based indices including the PDSI[*Anderson et al.*, 2011]. The index captures the onset of wetter conditions over much of the west by the end of 2004. The index also detects a winter drought in the Pacific Northwest and northern Rockies in early 2005 and the consequent severe drought conditions for most of the year. Other documented droughts detected by the index include southern Plains and lower Great Lakes (2005-2006), West and Southeast (2007), Southeast and northwest Great Lakes (2008) and Southwest (2009) areas[*Luo and Wood*, 2007]. The GRACE-DSI also captures the extended (2011-2014) Southern US drought, the severe California drought starting in 2012, and the 2012 central Great Plains drought[**?***Hoerling et al.*, 2014]. The GRACE-DSI also detects abnormally wet periods, such as the 2005-2006 California wetting[*Parrett and Hunrichs*, 2006], 2007 Southern Great Plains flooding[*Dong et al.*, 2011] and the 2010/2011 Mississippi River basin flooding[*Villarini et al.*, 2013].

3.7.3 The GRACE-DSI correlation with the PDSI-Z after pre-whitening

The GRACE-DSI and the PDSI-Z represent respectively the overall and shallower-depth water balance, and therefore can capture different trends and persistence (i.e. autocorrelation signature) due to different variations in overall and near-surface water components. To evaluate this, we pre-whiten both time series using an autoregressive integrated moving average (ARIMA) model with a first-degree difference and a first-order autoregressive process. This pre-whitening procedure retains the monthly and inter-annual variability but removing the trend and autocorrelation in both time series[*Chatfield*, 1975]. We show correlation result using the pre-whitened time series in Figure 4.11. We find significant correlation ($p<0.01$) over the entire CONUS. The strength of correlation enhances in the northwestern Plains compared to the un-pre-whitened result (Fig 3a), indicating that in that region

the GRACE-DSI and the PDSI-Z capture different trends and persistence in overall and near-surface water balance. For the rest part of the CONUS, the pre-whitened result shows smaller magnitude of correlation, implying that both indices capture consistent trends and persistence signatures.

3.7.4 Growing season correlation of PDSI with NDVI

Similar to GRACE-DSI, PDSI-Z captures the seasonal timing and latitudinal dependency of water constraints on NDVI. Generally, PDSI-Z has a slightly stronger correlation with NDVI than GRACE-DSI, because PDSI-Z is sensitive to surface moisture conditions that are more directly related to vegetation growth. GRACE-DSI complements PDSI by providing continuous estimates of bulk water changes. These changes include variations in both soil moisture affecting vegetation growth and deeper groundwater influencing soil moisture recharge and drought recovery. Therefore, combining GRACE-DSI with existing observations/estimates of plant-available water (such as PDSI) may improve the characterization of drought evolution as well as its impact on terrestrial ecosystems.

Table 3.1: Dynamic range and relative categories for wet (W) and dry (D) conditions of GRACE-DSI and PDSI-Z. The drought classification scheme is consistent with the USDM

Category	**Description**	**GRACE-DSI and PDSI-Z**
W4	Exceptionally wet	2.0 or greater
W3	Extremely wet	1.60 to 1.99
W2	Very wet	1.30 to 1.59
W1	Moderately wet	0.80 to 1.29
W0	Slightly wet	0.50 to 0.79
WD	Near normal	0.49 to -0.49
D0	Abnormally dry	-0.50 to -0.79
D1	Moderate drought	-0.80 to -1.29
D2	Severe drought	-1.30 to -1.59
D3	Extreme drought	-1.60 to -1.99
D4	Exceptional drought	-2.0 or less

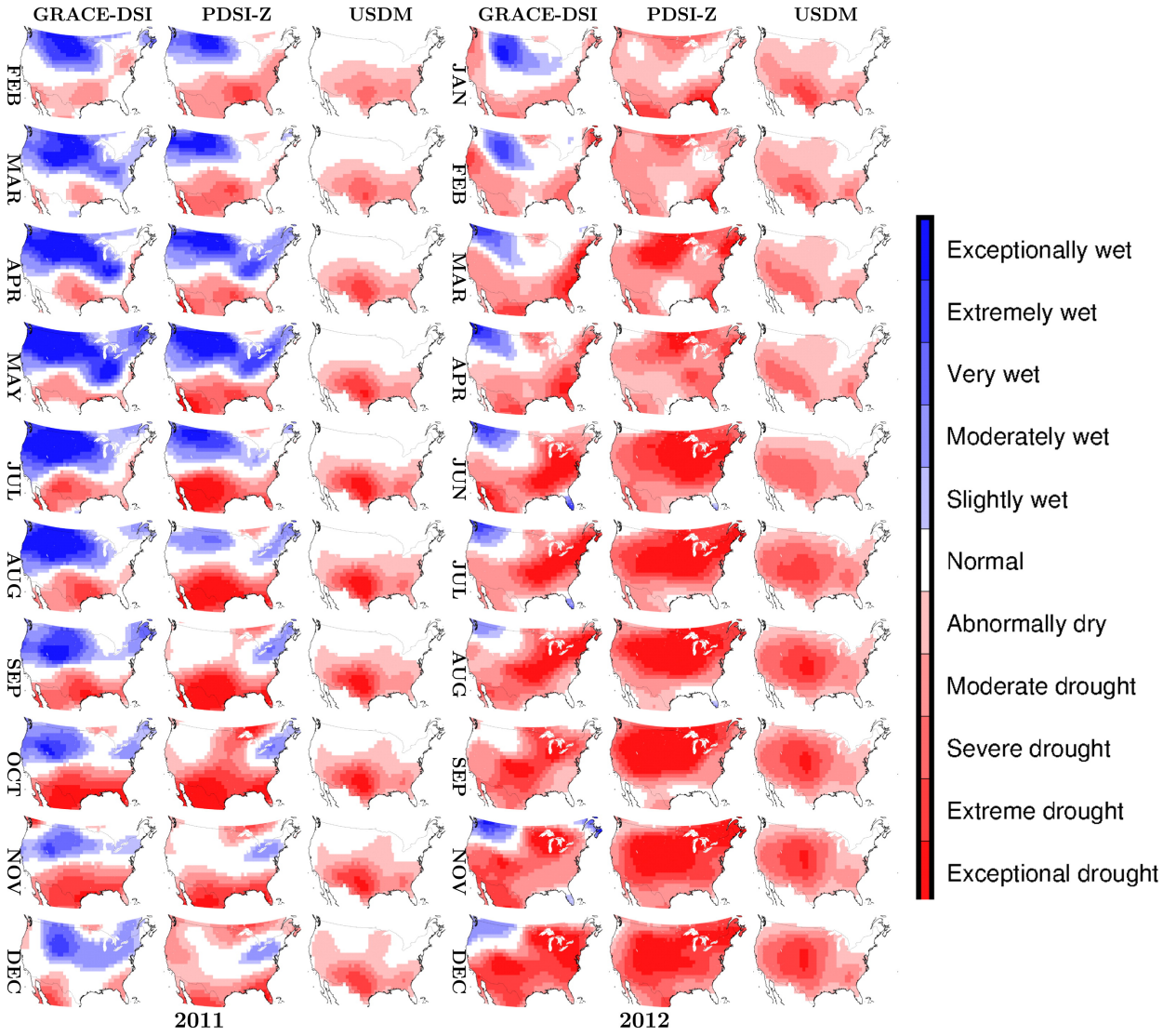

Figure 3.1: Spatial comparison of monthly GRACE-DSI, PDSI-Z, and USDM during 2011–12. Months with missing GRACE data are skipped. Color bar shows the drought/wetting categories defined in Table 3.1. PDSI-Z and GRACE-DSI use the same classification scheme. The drought classification scheme is consistent with the USDM. Note that USDM does not provide wetter-than-average information.

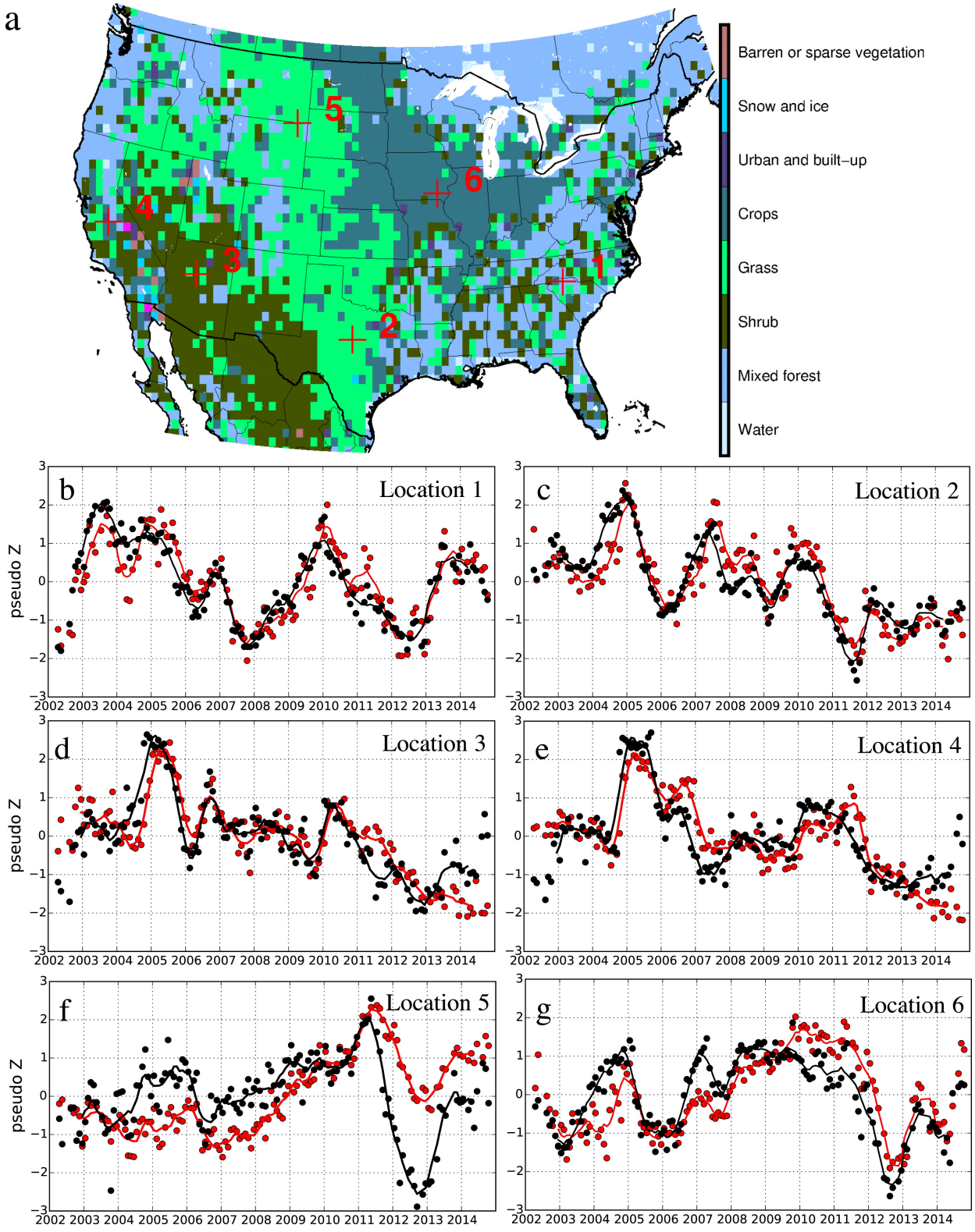

Figure 3.2: (a) Land cover map of the CONUS from MODIS land cover type (MCD12Q1)[*Friedl et al.*, 2010]. (b)–(g) Time series of GRACE-DSI (red) and PDSI-Z (black) at six locations shown in (a), respectively: (35°N, 82°W); (32°N, 99°W); (35°N, 112°W); (37°N, 120°W); (45°N, 105°W); and (41°N, 92°W). In (b)–(g), dots are index values and lines are smoothed values using a quadratic polynomial filter with a 13-month window.

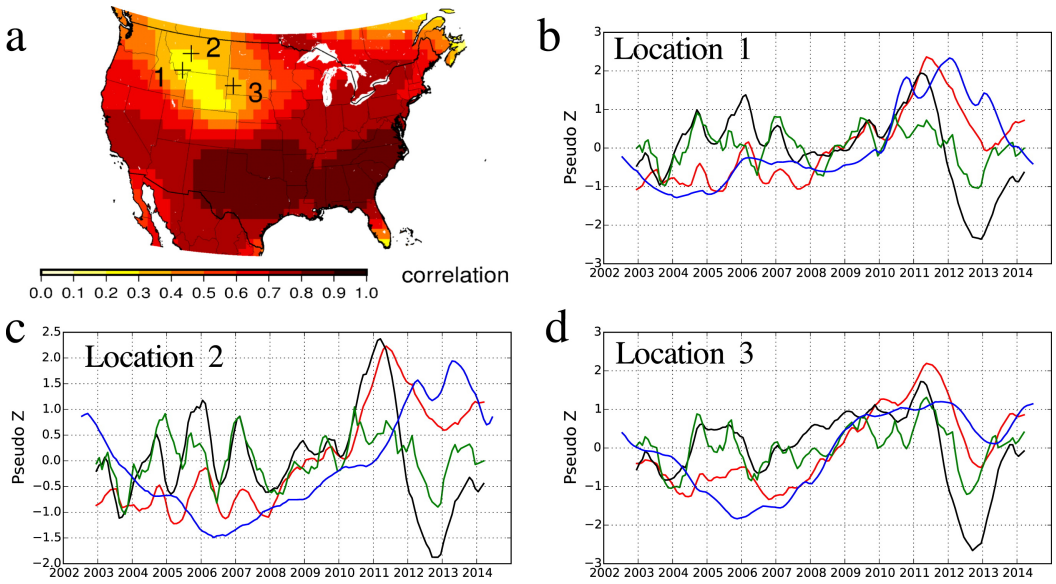

Figure 3.3: (a) Temporal cross-correlation coefficient between monthly GRACE-DSI and PDSI-Z during study period. (b)–(d) Times series of GRACE-DSI (red), PDSI-Z (black), satellite-retrieved SM-Z (green), and standardized groundwater depths (blue) for locations 1 to 3 in (a). Geographic coordinates for locations 1, 2, and 3 are (45.09°N, 112.64°W), (47.37°N, 111.19°W), and (44.30°N, 103.44°W), respectively. All lines are smoothed values using a quadratic polynomial filter with a 13-month window.

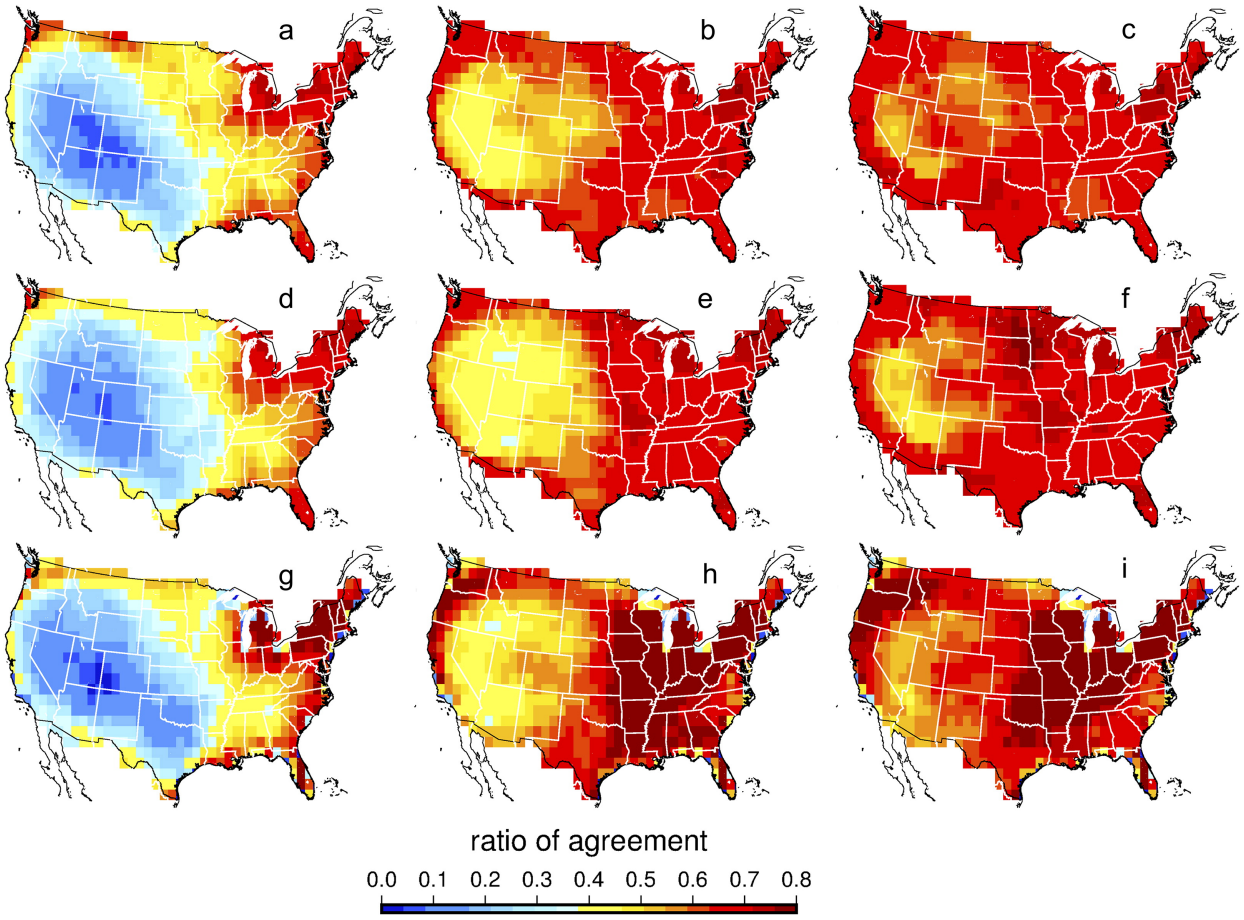

Figure 3.4: (a) The ratio of the number of months for which GRACE-DSI and USDM yield the same drought category to the total number of months considered. (b) As in (a), but with the USDM series biased by one category milder (i.e., USDM-1) at every grid cell. (c) As in (b), with the USDM series biased by two categories milder (i.e., USDM-2). Note that the no-drought category minus one or two categories is still considered no-drought. (d)-(f) As in (a)–(c), but replacing GRACE-DSI with PDSI-Z. (g)–(i) As in (a)–(c) but replacing GRACE-DSI with the original PDSI.

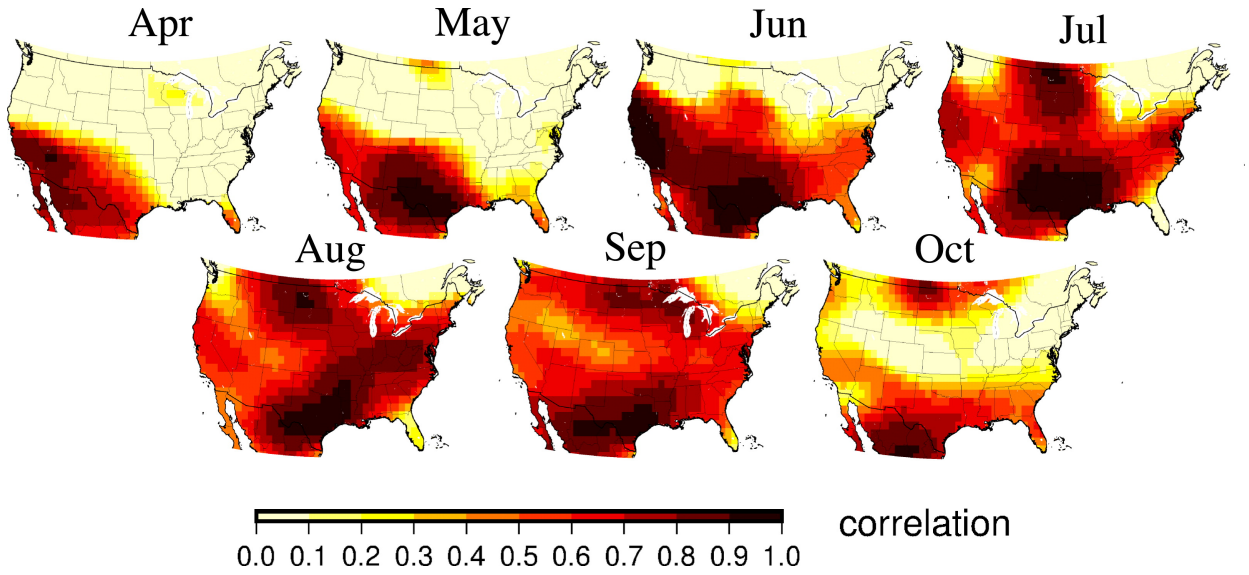

Figure 3.5: Monthly GRACE-DSI and MODIS NDVI correlation coefficients during the growing season (April). Correlation coefficient higher than 0.5 is significant at 90% confidence level.

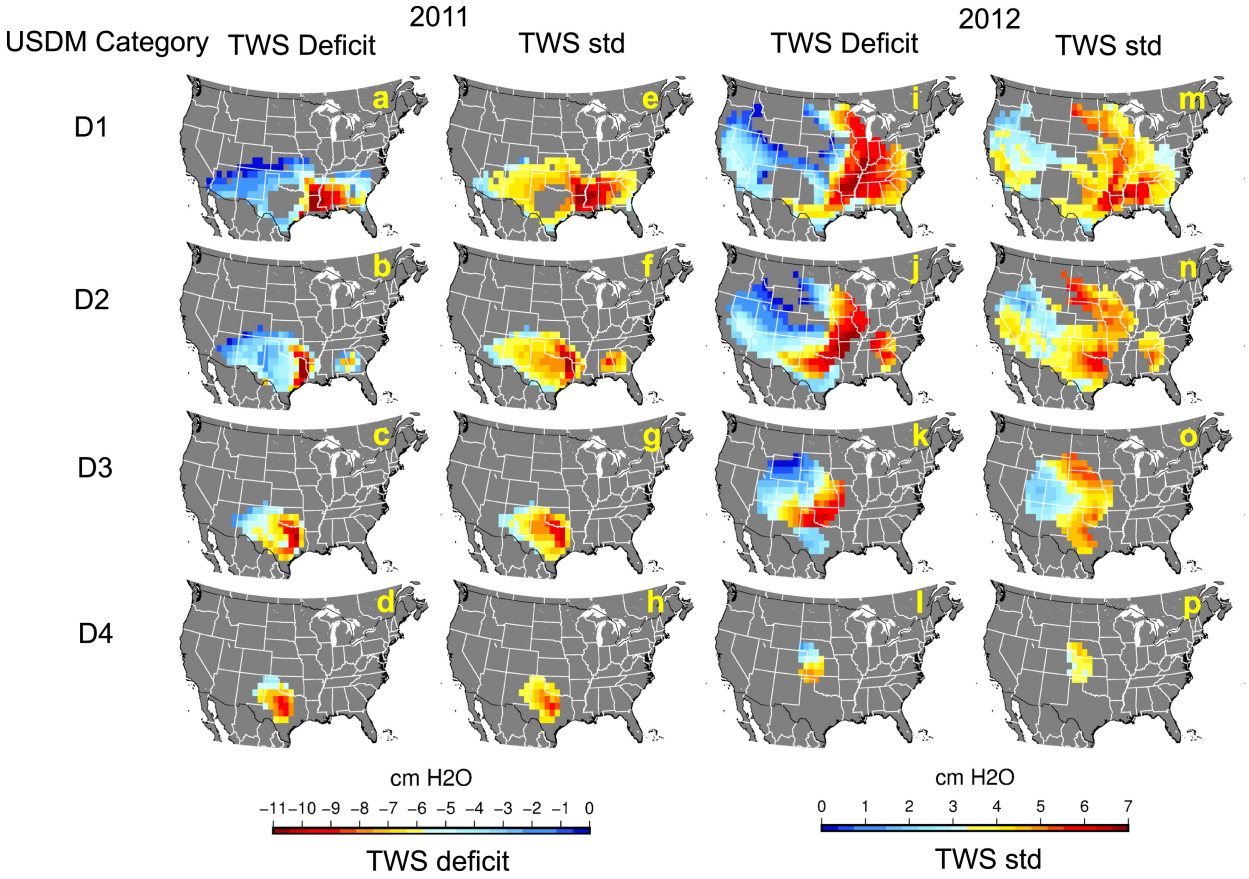

Figure 3.6: (a)–(d) The maximum TWS deficit observed in regions where D14 drought, respectively, has been reported by the USDM in 2011. (e)–(h) The TWS standard deviation for the calendar month when the maximum TWS deficit is observed in 2011, as seen in (a)–(d). (i)–(l) As in (a)–(d), but for 2012. (m)–(p) As in (e)–(h), but for 2012.

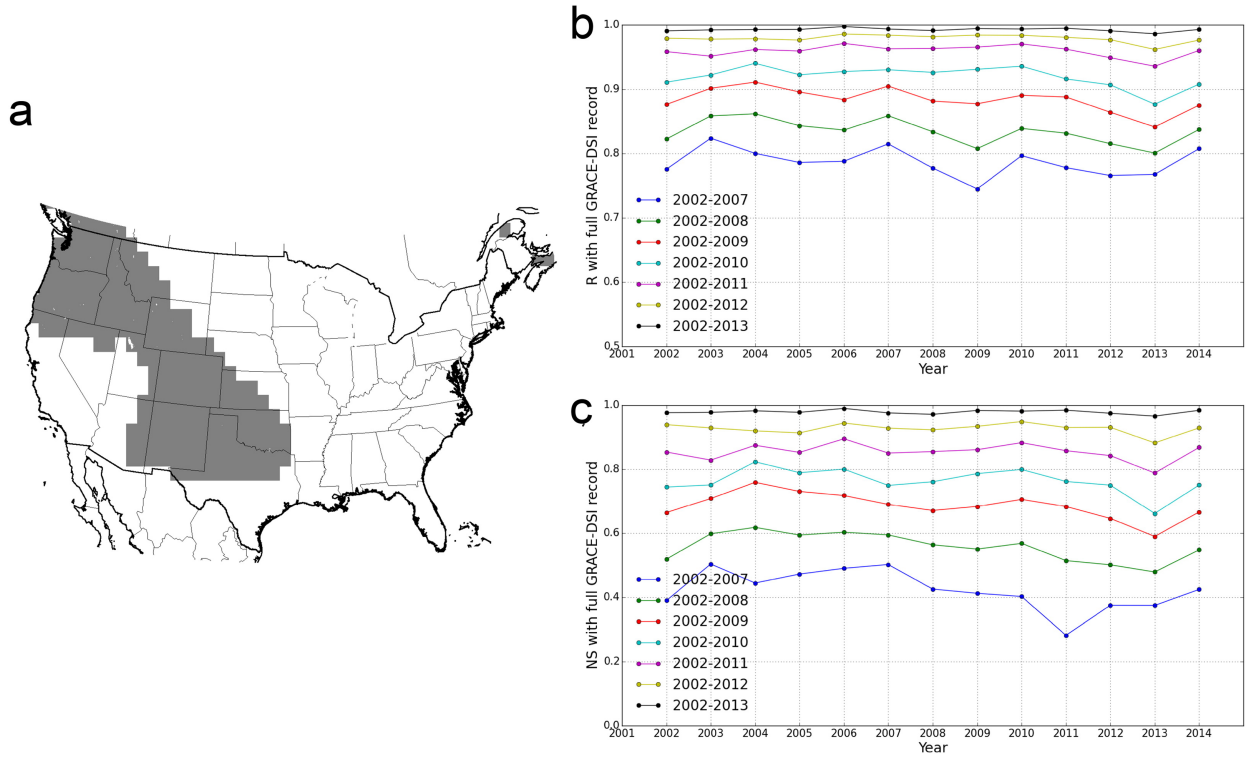

Figure 3.7: (a) In the gray regions, PDSI-Z underestimates drought by one category when using the 2002–14 climatology rather than the 1982–2014 climatology. (b) The area-weighted annual spatial correlation coefficients between the 2002–14 baseline GRACE-DSI and seven alternative GRACE-DSI calculated from varying baselines. (c) As in (b), but for the Nash-efficiency coefficients.

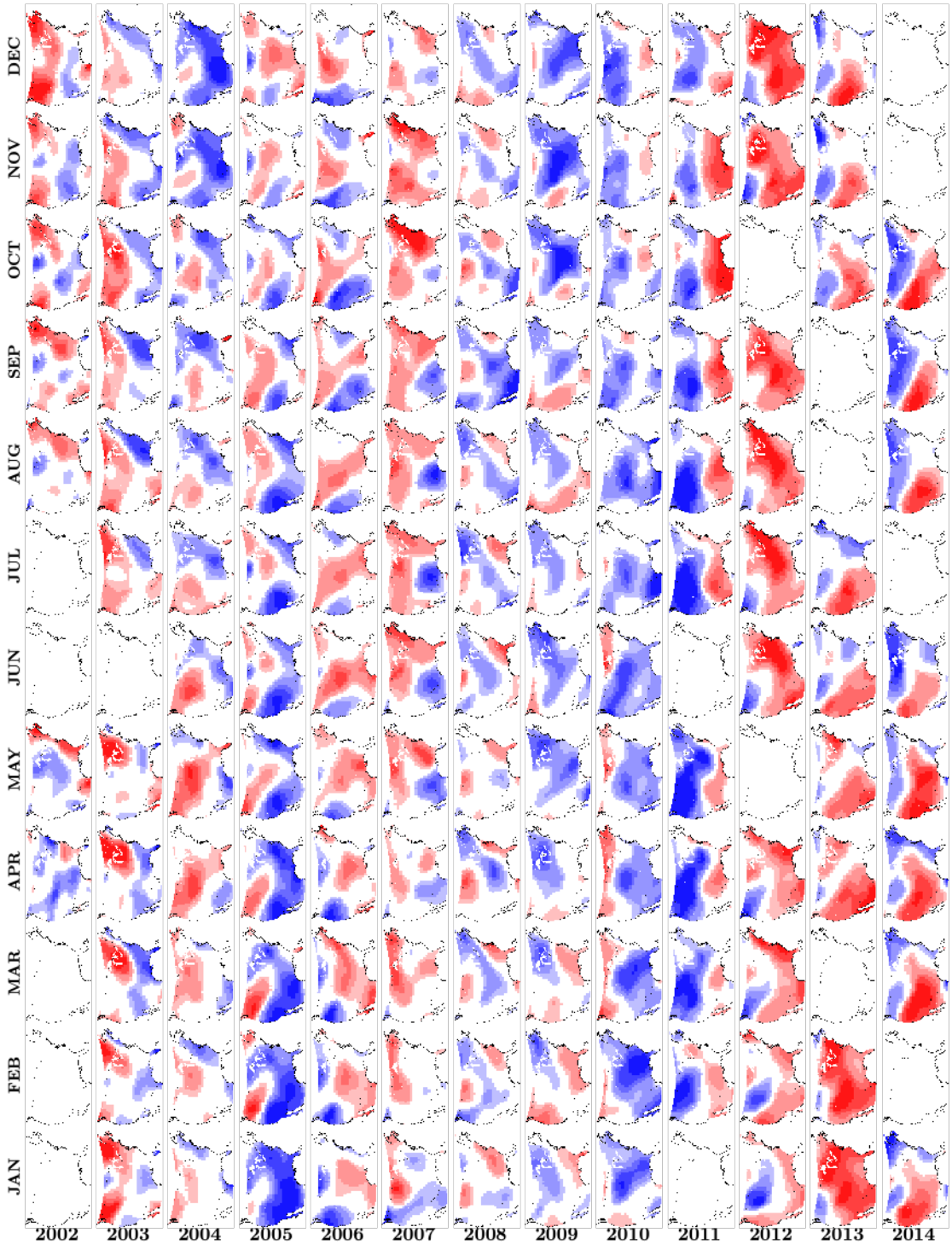

Figure 3.8: Monthly GRACE-DSI over the CONUS from 2002-2014. Months with missing GRACE data are left blank. Color schedule is the same as Figure 5.2 in the main text.

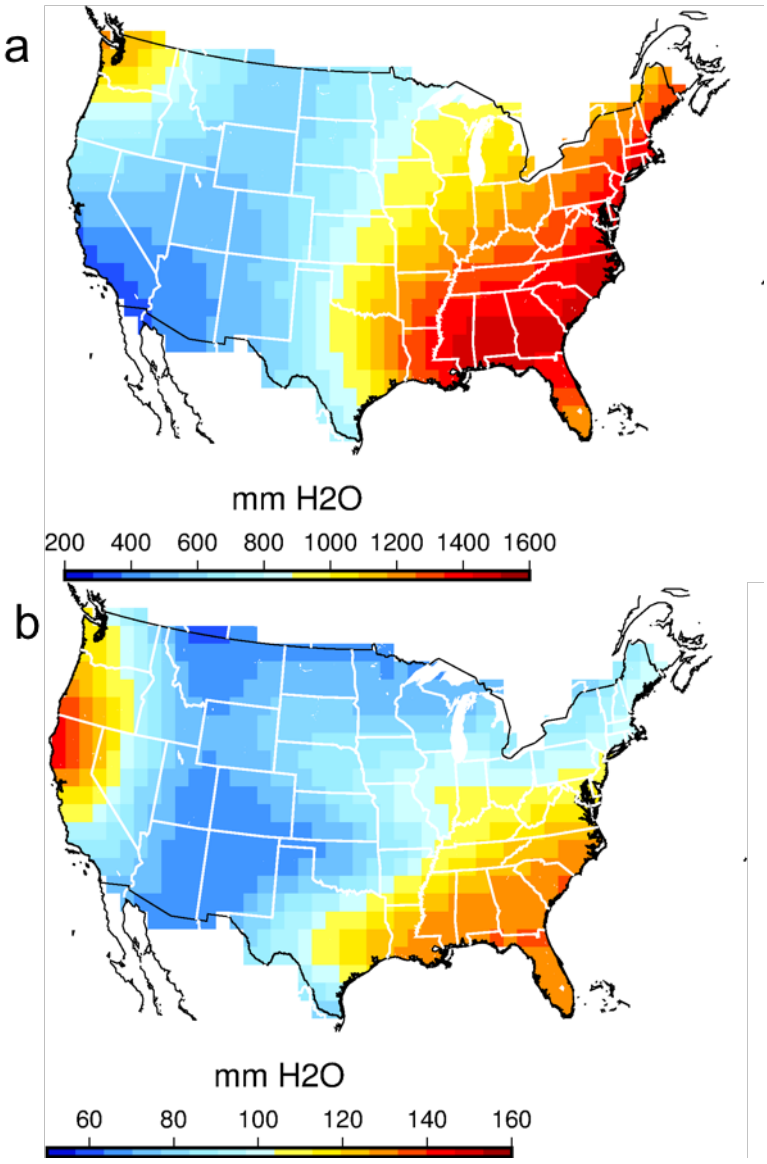

Figure 3.9: a) is the annual mean precipitation over CONUS from 1979-2009, derived from the Global Precipitation Climatology Project. b) is the standard deviation of annual mean precipitation.

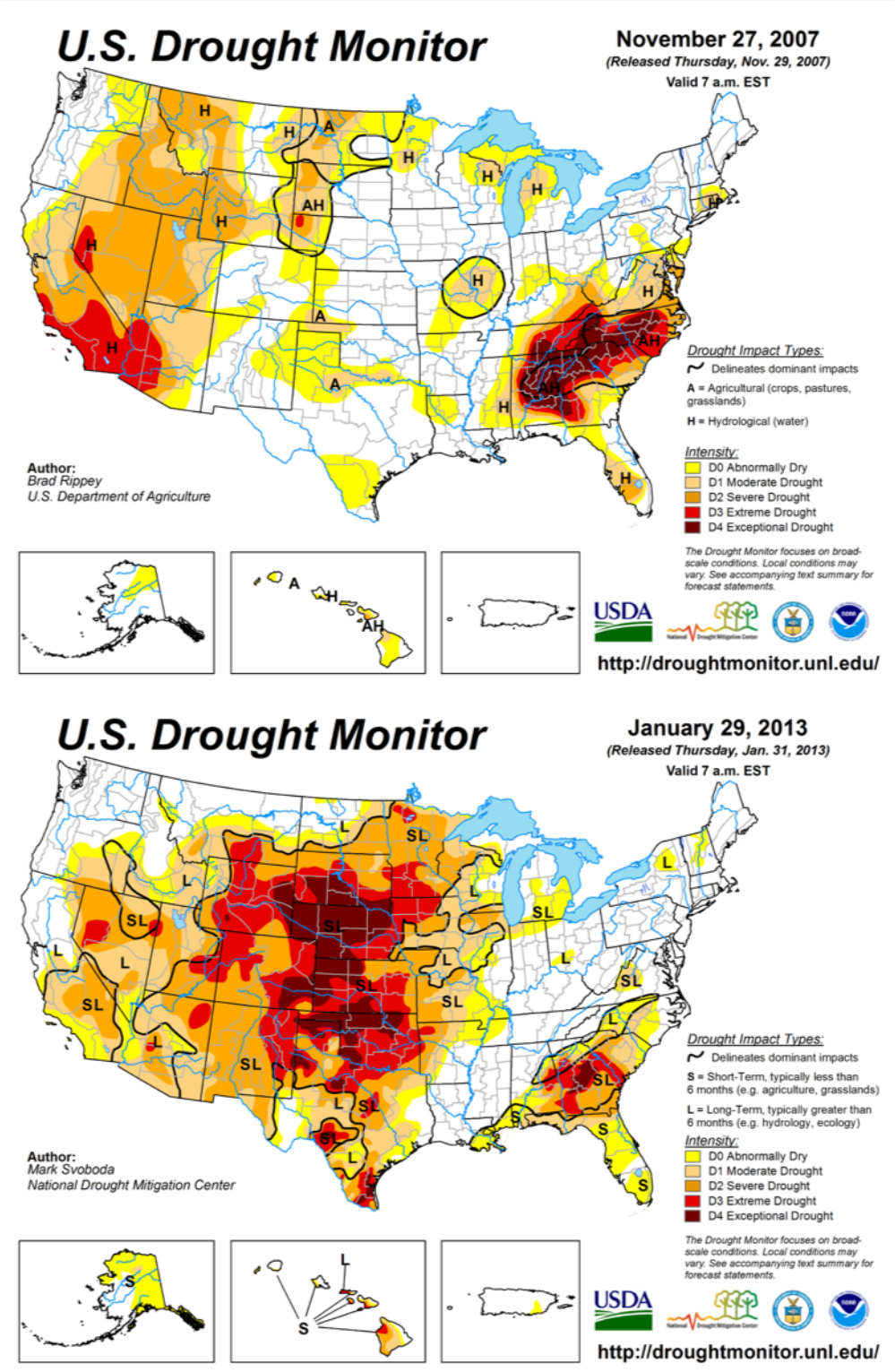

Figure 3.10: USDM drought monitoring maps for November 2007 and January 2013.

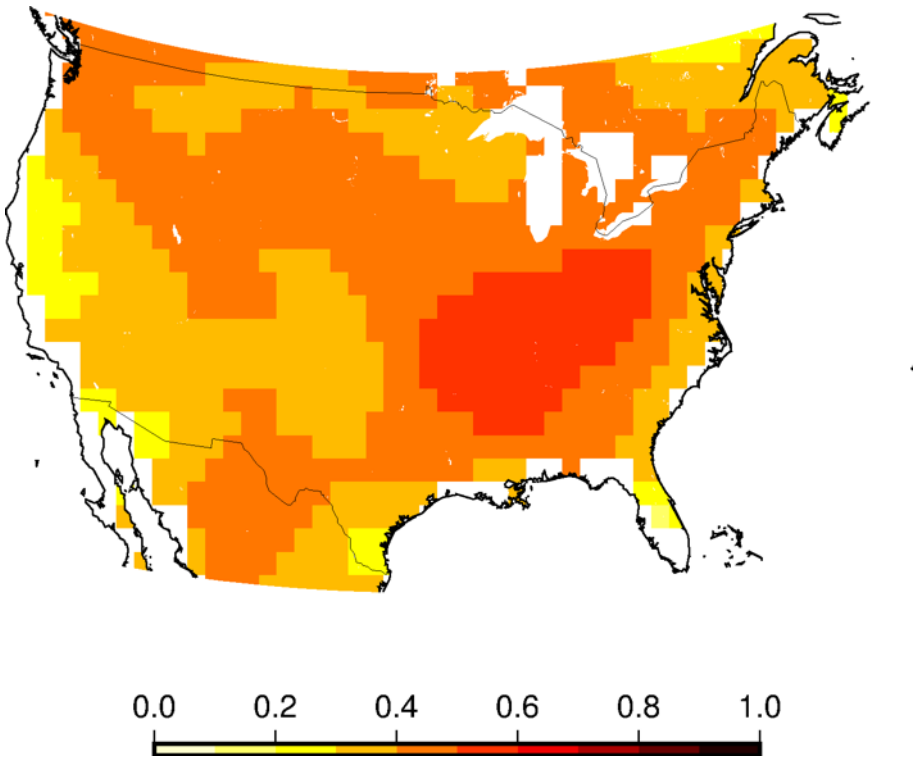

Figure 3.11: Correlation coefficient between prewhitened GRACE-DSI and PDSI-Z time series. The correlation is significant over the entire CONUS at 99% confidence level.

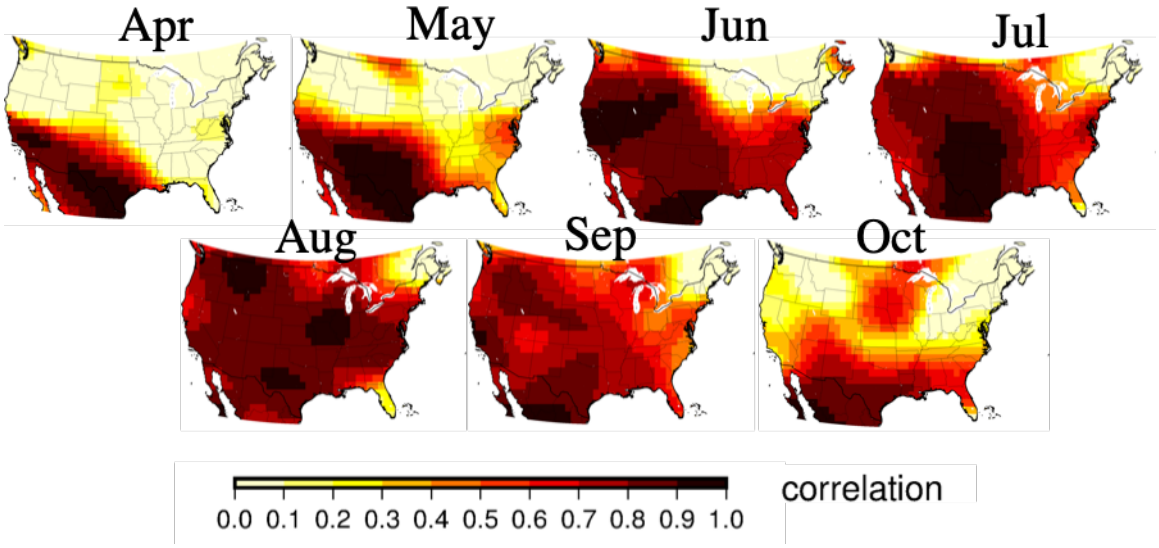

Figure 3.12: Monthly PDSI-Z and MODIS NDVI correlation coefficients during the growing season (April to October). Correlation coefficient higher than 0.5 is significant at 90% confidence level.

Chapter 4

A global gridded dataset of GRACE drought severity index for 2012-14: Comparison with PDSI and SPEI and a case study of the Australia Millennium drought

As appears in:

Zhao, M., G. A, I. Velicogna, and J.S. Kimball, "A Global Gridded Dataset of GRACE Drought Severity Index for 2002-14: Comparison with PDSI and SPEI and a Case Study of the Australia Millennium Drought," *Journal of Hydrometeorology*, 2017.

4.1 Abstract

A new monthly global drought severity index (DSI) dataset developed from satellite-observed time-variable terrestrial water storage changes from the Gravity Recovery and Climate Experiment (GRACE) is presented. The GRACE-DSI record spans from 2002 to 2014 and will be extended with the ongoing GRACE and scheduled GRACE Follow-On missions. The GRACE-DSI captures major global drought events during the past decade and shows overall favorable spatiotemporal agreement with other commonly used drought metrics, including the Palmer drought severity index (PDSI) and the standardized precipitation evapotranspiration index (SPEI). The assets of the GRACE-DSI are 1) that it is based solely on satellite gravimetric observations and thus provides globally consistent drought monitoring, particularly where sparse ground observations (especially precipitation) constrain the use of traditional model-based monitoring methods; 2) that it has a large footprint (350 km), so it is suitable for assessing regional- and global-scale drought; and 3) that it is sensitive to the overall terrestrial water storage component of the hydrologic cycle and therefore complements existing drought monitoring datasets by providing information about groundwater storage changes, which affect soil moisture recharge and drought recovery. In Australia, it is demonstrated that combining GRACE-DSI with other satellite environmental datasets improves the characterization of the 2000s Millennium Drought at shallow surface and subsurface soil layers. Contrasting vegetation greenness response to surface and underground water supply changes between western and eastern Australia is found, which might indicate that these regions have different relative plant rooting depths.

4.2 Introduction

Drought indices are convenient tools for evaluating drought and its social and ecological impacts, as well as for decision-making in drought prevention and mitigation. Many global gridded drought indices such as the Palmer drought severity index (PDSI) and the standardized precipitation index (SPI) rely on the accuracy of meteorological inputs and/or simple water balance models. They become unreliable where ground observations (especially precipitation and soil properties) are sparse. Therefore, satellite remote sensing of drought-related variables such as normalized difference vegetation index (NDVI), surface soil moisture (SM), and terrestrial water storage (TWS) have been proposed for complementing more traditional meteorological drought indices[*Anderson et al.*, 2011; *Mu et al.*, 2013; *Thomas et al.*, 2014; *AghaKouchak et al.*, 2015].

Drought indices are generally sensitive to a specific part of the hydrological cycle. The PDSI is sensitive to atmosphere moisture demand and near-surface soil moisture content and has been frequently used as a measure of meteorological and soil moisture drought[*Mishra and Singh*, 2010]. The SPI measures accumulated precipitation deficits over varying time scales. Shorter time scales are sensitive to surface SM variability, medium scales are sensitive to streamflow and reservoir levels, and longer time scales are sensitive to groundwater variations[*McKee et al.*, 1993]. However, this sensitivity pattern varies by region and land cover, which might complicate the choice of SPI time scale in characterizing drought onset, duration, and recovery[*Zeng et al.*, 2008; *Wang et al.*, 2015]. An index that reflects total water storage changes in the hydrological cycle is desired for fully characterizing drought evolution and ecosystem response to water supply variations[*Van Loon*, 2015b].

TWS estimates derived from the Gravity Recovery and Climate Experiment (GRACE) have been widely used to examine regional-scale droughts worldwide[*Yirdaw et al.*, 2008; *Leblanc et al.*, 2009; *Chen et al.*, 2010a; *Long et al.*, 2013]. *Li et al.* [2012] assimilated GRACE TWS

into a land surface model to monitor drought in Europe. *Thomas et al.* [2014] proposed a framework to quantify drought-induced water storage deficits. However, previous studies did not consider the spatial and temporal variability of local hydroclimatology, which is important for drought comparison across space and time[*Zhao et al.*, 2017b]. *Houborg et al.* [2012] account for this issue by deriving local cumulative distribution of dry and wet conditions from the GRACE data assimilation system; however, this system may not be readily available outside of North America. Recently, *Zhao et al.* [2017b] developed a new drought severity index (DSI) based solely on GRACE TWS estimates. Compared to previous GRACE TWS-based drought studies, the GRACE-DSI is calculated without model assimilation and considers spatial and temporal variability of local hydroclimatology. *Zhao et al.* [2017b] evaluate the performance of the GRACE-DSI in the continental United States, where robust drought characterization is available attributed to dense surface observation network and reliable PDSI estimates, whereby GRACE-DSI shows significant agreement with the PDSI and overall favorable spatiotemporal correspondence with satellite retrievals of NDVI and SM.

Here we present a global gridded GRACE-DSI record derived from the standard spherical harmonic approach extending from 2002 to 2014. We compare the GRACE-DSI record with traditional drought indices including the PDSI and the standardized precipitation evapotranspiration index (SPEI; [*Vicente-Serrano et al.*, 2010a]) over the global domain, excluding Antarctica and Greenland. We also exclude the barren or sparsely vegetated land cover based on the Moderate Resolution Imaging Spectroradiometer (MODIS) land-cover type data product (MCD12Q1; [*Friedl et al.*, 2010]) because these regions have low hydroclimatic variability and large DSI error[*Dai*, 2011]. We conduct a regional case study of the 2000s Millennium Drought in Australia[*van Dijk et al.*, 2013] to evaluate potential synergy between the GRACE-DSI and other satellite environmental data records, including SM and NDVI, for monitoring drought-related impacts on terrestrial ecosystems. We quantify the uncertainty of GRACE-DSI in characterizing drought. We evaluate the impact of using a 13-yr

reference period on climatology estimates rather than a longer record. We also discuss DSI results derived from the GRACE mascon solutions.

4.3 Data and Methodology

We use Release-05 (RL05) Center for Space Research (CSR) GRACE gravity solutions in the form of spherical harmonic coefficients truncated to degree 60 for the period from April 2002 to October 2014. We substitute the GRACE-derived C20 (degree 2 order 0 spherical harmonic coefficient) coefficients with monthly estimates from satellite laser ranging[*Cheng et al.*, 2013b]. We include degree-1 coefficients calculated as in *Swenson et al.* [2008] and correct the glacial isostatic adjustment signal following [*A et al.*, 2013]. To reduce correlated errors, we filter each monthly field following [*Swenson and Wahr*, 2006]. We convolve the filtered coefficients with a 350-km radius Gaussian averaging function[*Wahr et al.*, 1998b] and calculate the monthly TWS mass anomalies relative to the 2002-14 mean on a $1^\circ \times 1^\circ$ latitude-longitude grid.

We use the 2.5°global self-calibrated PDSI developed by *Dai* [2011]. The PDSI uses a two-layer model to assess soil water balance by accounting for both water supply and demand [*Palmer*, 1965]. The PDSI values correlate with top 1-m depth soil moisture observations over the United States, the former Soviet Union, Mongolia, and China, and with streamflow over major global river basins[*Dai et al.*, 2004; *Dai*, 2011].

We use the 0.5°global SPEI developed by *Vicente-Serrano et al.* [2010a]. The SPEI evaluates accumulated precipitation minus potential evapotranspiration (PET) over multiple time scales up to 48 months[*Vicente-Serrano et al.*, 2010b]. The SPEI extends the SPI by incorporating PET in determining drought and shows advantages over the SPI in capturing temperature impact on water demand[*Vicente-Serrano et al.*, 2014].

In the Australian case study, we use remotely sensed 0.25°SM data from the European Space Agency Climate Change Initiative[*Liu et al.*, 2011, 2012]. The SM record employs satellite passive and active microwave sensor data with improved spatial and temporal coverage and resolution. We also use the MODIS monthly Climate Model Grid 0.05°NDVI (MOD13C2; [*Huete et al.*, 2002]) as a proxy for drought-sensitive vegetation growth changes. Previous studies have shown that, at the site scale, changes in NDVI are sensitive to root-zone SM variations in dry regions[e.g. *Wang et al.*, 2007; *Schnur et al.*, 2010]. Recent studies confirmed this result at the regional scale in Australia[*Chen et al.*, 2014; *Yang et al.*, 2014; *De Keersmaecker et al.*, 2017].

For each grid cell, the GRACE-DSI is defined as the standardized anomalies of GRACE TWS as follows:

$$\mathrm{GRACE-DSI}_{i,j} = \frac{TWS_{i,j} - \overline{TWS_j}}{\sigma_j} \tag{4.1}$$

where i is year ranging from 2002 to 2014; j is month ranging from January to December; and and are the mean and standard deviation of TWS anomalies in month j, respectively. The GRACE-DSI is a dimensionless quantity that detects both drought and abnormally wet events: less than 2.0 is an exceptional drought, from 1.99 to 1.60 is an extreme drought, from 1.59 to 1.30 is a severe drought, from 1.29 to 0.80 is a moderate drought, from 0.79 to 0.50 is abnormally dry, from 0.49 to 0.49 is near normal, from 0.50 to 0.79 is slightly wet, from 0.80 to 1.29 is moderately wet, from 1.30 to 1.59 is very wet, from 1.60 to 1.99 is extremely wet, and higher than 2.0 is exceptionally wet.

Because of the truncation and filtering applied to reduce short-scale errors in GRACE, each GRACE-DSI grid cell represents conditions averaged over a 350-km radius footprint[*Velicogna and Wahr*, 2006, 2013]. For consistency, all datasets are processed the same as the GRACE data, that is, expanded in spherical harmonic, truncated to degree 60, filtered, spatially

smoothed, and projected onto a 1° × 1° grid[*Velicogna et al.*, 2012]. To bring the data comparison to the same reference period, we standardize the PDSI, SPEI, SM, and NDVI relative to the GRACE-DSI period (2002-14), herein referred to as PDSI-Z, SPEI-Z, SM-Z, and NDVI-Z using the following equation:

$$\mathrm{X}_{i,j} = \frac{X_{i,j} - \overline{X_j}}{\sigma_j} \tag{4.2}$$

where X represents PDSI, SPEI, SM, or NDVI and and are the monthly mean and standard deviation of X calculated from the same years and months used for the GRACE-DSI. This standardization process has also been employed in previous drought index studies[*Anderson et al.*, 2011, 2013; *Mu et al.*, 2013].

We calculate the GRACE-DSI uncertainty due to GRACE measurement error and leakage error. At every grid cell, we estimate the GRACE measurement error following *Wahr et al.* [2006] and the leakage error following *Landerer and Swenson* [2012]. A synthetic TWS field from the Community Land Model, version 4.5 (CLM4.5), is used as a realistic representation of the observed TWS signal. The leakage error is defined at each grid cell as the RMS difference between the original and the GRACE-like processed CLM4.5 TWS time series (i.e., converted into harmonics, truncated to degree 60, de-striped, spatially smoothed, and converted to 1° × 1° regular longitude-latitude grids). The total GRACE error is the summation of the measurement and leakage error in quadrature. We then use a Monte Carlo simulation to estimate the DSI error. At each grid cell, we generate an ensemble of 1000 normally distributed zero mean random noise time series (σ = total GRACE error). We compute the GRACE-DSI for each of those 1000 simulations and calculate the sample distances from their monthly sample means. The resulting sample distances follow a normal distribution, and we set the one standard deviation of the sample distances as the error of GRACE-DSI. Our GRACE-DSI uncertainty estimate is conservative as it assumes that all nonannual and nonsemiannual variations in GRACE data are due to measurement errors.

This assumption overestimates the GRACE-DSI error in regions where large nonseasonal variability is observed, for instance due to extreme drought and flooding[*Wahr et al.*, 2006; *Tiwari et al.*, 2009]. For the Australian case study, we adopt the same approach to provide a conservative error estimate for the NDVI-Z and SM-Z records.

4.4 Results

Figure 5.1 shows the global distribution of GRACE-DSI, PDSI-Z, and SPEI-Z at selected time scales for July 2010, a month in which major drought conditions occurred worldwide[*AghaKouchak et al.*, 2015]. The three indices show agreement in the intensity and spatial distribution of drought conditions in western Russia, Southeast Asia, and northern India. Differences are observed in regions including northern China and mainland Australia. The GRACE-DSI and 1-month SPEI-Z show severe drought in northern China, whereas the PDSI-Z and longer-time-scale SPEI-Z ($>$ 6 months) show only moderate drought or near-normal conditions. In mainland Australia, the GRACE-DSI shows a similar drought pattern as the 36-month SPEI-Z, but more severe and widespread drought than the PDSI-Z. In the Amazon, the GRACE-DSI and PDSI-Z capture similar drought intensity and extent. The SPEI-Z also detects drought in the Amazon, but with varying extent and intensity for different time scales. Wet events are also observed by these indices, but with different extents and magnitudes. For instance, all indices capture the flooding in China except the 36-month SPEI-Z. The GRACE-DSI, PDSI-Z, and shorter-than-15-months SPEI-Z observe the wetting in Pakistan[*Webster et al.*, 2011], whereas the long-time-scale SPEI-Z does not detect this wetting event. The GRACE-DSI characterizes the southern Africa wetting as a very wet to extremely wet event, which is consistent with other observation-based drought indices[*AghaKouchak et al.*, 2015]. This event, however, is missed by the PDSI-Z and short-time-scale SPEI-Z ($<$6 months).

Figure 5.2 shows time series of the GRACE-DSI, PDSI-Z, and SPEI-Z at selected time scales for four locations where major drought events have been reported. The GRACE-DSI shows generally close agreement with the PDSI-Z, with a 1-month delay behind PDSI-Z in response to drought evolution for the 2010 western Russian drought (Figure 5.2a; [*Yoshida et al.*, 2015]) and the 2005-06 vegetation drought in East Africa (Figure 5.2c; [*Rulinda et al.*, 2012]). In the Amazon, the GRACE-DSI is concurrent with the PDSI-Z and captures the extraordinary 2005 and 2010 droughts in this region as well as the exceptional 2009 flooding (Figure 5.2b; [*Zeng et al.*, 2008; *Chen et al.*, 2010b; *Lewis et al.*, 2011]).

The GRACE-DSI agrees well with the 6-12-month SPEI-Z (Figures 5.2i-p). The GRACE-DSI also captures drought events detected by the shorter-time-scale SPEI-Z. For instance, the GRACE-DSI captures the short-term 2005/06 winter drought in western Russia (Figures 5.2e,i,m). This event is well captured by the 1-12-month SPEI-Z but is missed by the 24-0month SPEI-Z. The GRACE-DSI captures the severe short-term 2010 drought and the 2011/12 winter drought in northern China[*Barriopedro et al.*, 2012], which are both missed by other indices except for the 1-month SPEI-Z (Figure 5.2h). In the Amazon, the GRACE-DSI agrees better with the SPEI-Z at time scales shorter than 6 months (Figures 5.2f,j).

Figure 5.3 shows the spatial distribution of monthly GRACE-DSI correlations with the PDSI-Z and SPEI-Z at selected time scales. The correlation between the GRACE-DSI and PDSI-Z is significant at the 95% confidence level over 91% of our study domain with an area-weighted mean correlation coefficient of 0.54 (Figure 5.3a). The 1-month SPEI-Z has a low correlation with the GRACE-DSI in most regions (Figure 5.3b). The correlation increases in magnitude and spatial extent at time scales of 3-18 months and weakens at time scales of 18-48 months (Figures 5.3c-i). The GRACE-DSI correlation with SPEI-Z is significant at the 95% confidence level over 90% of our study domain at time scales from 5 to 10 months with area-weighted average correlation coefficients above 0.43. Generally, the GRACE-DSI and SPEI-Z have the highest correlation at time scales of 6-18 months (Figures 5.3d-f).

Distinct rainfall and drought patterns were documented between western and eastern Australia during the Millennium Drought period[*Beard et al.*, 2011; *van Dijk et al.*, 2013]. Figures 5.4b and c show the time series of the GRACE-DSI, SM-Z, and NDVI-Z for western Australia (location 1) and eastern Australia (location 2). The GRACE-DSI and NDVI-Z have much smaller uncertainty than SM-Z. These satellite-based metrics show overall consistent temporal variations, but differences in trends and magnitude. From 2002 to late 2009, the GRACE-DSI shows statistically significant drying trends in both locations ($0.27 \pm 0.03\ yr^1$, $p < 0.001$ for western Australia and $0.07 \pm 0.02\ yr^1$, $p < 0.005$ for eastern Australia). A significant NDVI-Z decreasing trend ($0.08 \pm 0.04\ yr^1$, $p < 0.05$) is observed in western Australia but not in eastern Australia. The SM-Z shows no trend in either location. During the extreme 2010/11 La Nina-induced flooding[*Beard et al.*, 2011], the SM-Z increases earlier than GRACE-DSI at both locations. The NDVI-Z increases concurrently with GRACE-DSI in western Australia but responds simultaneously with SM-Z in eastern Australia. For the entire study period, the GRACE-DSI has consistently good agreement with NDVI-Z at both locations with a correlation coefficient R of 0.80 and root-mean-square error (RMSE) of 0.60. SM-Z has good agreement with NDVI-Z in eastern Australia ($R = 0.84$, $RMSE = 0.57$) but less agreement with NDVI-Z in western Australia ($R = 0.53$, $RMSE = 0.96$).

Figure 5.5 shows the GRACE-DSI uncertainty in drought category caused by GRACE measurement and leakage errors. We find that GRACE-DSI characterizes drought severity with an uncertainty less than one category (i.e., DSI error < 0.4, one drought category interval) in 47% of our study domain (all land regions excluding Antarctica, Greenland, and barren grounds), with one category uncertainty ($0.4 \leq$ DSI error < 0.8) in 51% and with two categories uncertainty (DSI error ≥ 0.8) in 2%.

4.5 Discussion

4.5.1 GRACE-DSI comparison with PDSI and SPEI

The three indices adopt different water balance concepts and address different hydrological processes. The GRACE-DSI is based on direct measurements of water balance that account for water supply from precipitation and water demand from actual evapotranspiration and runoff. The PDSI also accounts for the net water changes but through a simplified two-layer bucket model, which usually represents the water balance of shallow soil depth*Dai* [2011]. The GRACE-DSI captures total water supply conditions while PDSI-Z captures conditions relative to the shallow-depth water storage. For instance, in July 2010 over eastern Australia (Figure 5.1), the GRACE-DSI captures the groundwater deficit near the end of the Millennial Drought[*van Dijk et al.*, 2013; *Leblanc et al.*, 2009, 2012] while the PDSI shows surface water replenishment from the 2010 La Nina event[*Beard et al.*, 2011]. The difference in water balance concepts causes the two indices to capture different long-term trends in surface or deeper water storage; this is the case for northern China (Figure 5.2b) where the GRACE-DSI indicates a drying trend in overall water storage from 2002 to 2012 while the PDSI-Z suggests a surface wetting trend from 2006 to 2014. This agrees with the observation from *Qin et al.* [2015] in the same area that surface soil moisture and NDVI are increasing during our study period despite significantly decreasing annual precipitation trend. For this reason, we find a low temporal correlation in this region between GRACE-DSI and PDSI-Z for the entire time series ($R = 0.29$, $p < 0.05$) while the correlation between the detrended time series increases to 0.50 ($p < 0.05$).

In addition, the response time of shallow soil moisture to drought can be shorter than the overall water storage[e.g. *Van Loon*, 2015b]. This explains the 1-month lag delay between the GRACE-DSI and PDSI-Z for the 2010 western Russian drought and the 2005-06 East Africa drought (Figures 5.2a,c; [*Rulinda et al.*, 2012; *Yoshida et al.*, 2015]). Groundwater pumping

and agricultural irrigation can also delay the response of GRACE-DSI to natural drought evolution, which would affect its agreement with PDSI in heavily irrigated areas[*Dai*, 2011].

The SPEI-Z assesses the climatic water balance (precipitation minus potential evapotranspiration) over multiple time scales[*Vicente-Serrano et al.*, 2010a]. The GRACE-DSI represents moisture variations from all hydrological components. The response time of these components to climatic water balance generally lengthens as they go deeper into the ground[*McKee et al.*, 1993; *Vicente-Serrano et al.*, 2010a]. Therefore, the component that dominates changes in the hydrologic cycle will determine at which time scale SPEI-Z agrees with GRACE-DSI. For instance, the GRACE-DSI only has good agreement with 1-month SPEI-Z in detecting the northern China drought (Figure 5.1), indicating that shallow-depth water storage deficit dominates this drought event, in agreement with *Barriopedro et al.* [2012] describing this event as short-term drought caused by transient summer precipitation shortage. In Australia (Figure 5.1), where the GRACE-DSI is dominated by severe deep-water depletion near the end of the Millennium Drought[e.g. *Leblanc et al.*, 2012; *van Dijk et al.*, 2013], GRACE-DSI shows a better agreement in terms of spatial pattern with SPEI-Z for time scales longer than 30 months (Figure 5.1). In the Amazon we find a maximum correlation for 6-month SPEI-Z (Figures 5.2j, 5.3k), indicating that surface and shallow soil moisture dominates the changes in the hydrologic cycle, in agreement with earlier studies in the same area[*Han et al.*, 2009; *Kim et al.*, 2009; *Frappart et al.*, 2012, 2013]. These results demonstrate that the time scale at which the SPEI-Z achieves maximum temporal correlation with the GRACE-DSI (optimal time scale) is of potential use to constrain the relative depth of the water component that dominates the overall changes in the water cycle at a specific location. For instance, the optimal time scale in northern and western Australia is much smaller than in southeastern Australia (Figure 5.3k). This result suggests that a relatively shallower hydrologic component dominates the overall changes in the water cycle in northern and western Australia compared to southeastern Australia, where severe groundwater depletion dominates the changes in TWS during the Millennium Drought*Leblanc et al.* [2009, 2012]. We evaluate

this hypothesis using NDVI-Z as a proxy for root-zone soil moisture variations[*Chen et al.*, 2014; *Yang et al.*, 2014; *De Keersmaecker et al.*, 2017]. We calculate the optimal time scale between SPEI-Z and NDVI-Z in Figure 4.6a. In northern and western Australia, we find that the optimal SPEI-Z time scale for NDVI-Z is consistent with the optimal time scale for GRACE-DSI (Figure 4.6b), suggesting that root-zone soil moisture variability dominates the overall changes in the water cycle over these regions. This result therefore confirms that changes in the water cycle in northern and western Australia are dominated by a relatively shallower hydrologic component (root-zone soil moisture) than in southeastern Australia (groundwater).

The optimal time scale (Figure 5.3k) shows large spatial variability. This variability is consistent among different GRACE processing approaches (see Fig. S3b in the supplemental material). We also find that the spatial pattern of Figure 5.3k is insensitive to GRACE measurement errors. Therefore, the result of Figure **??** is robust. The large spatial variability highlights the geographical complexity of the translation of climatic water balance into hydrologic system[*Tallaksen et al.*, 2009; *Teuling et al.*, 2013; *Van Loon et al.*, 2012, 2014; *Van Loon*, 2015b]. This large spatial variability is also observed in previous studies comparing SPEI with PDSI at global scale[*Vicente-Serrano et al.*, 2010b] and comparing SPEI and SPI at local and regional scales with hydrological drought proxies using groundwater and streamflow observations[*Bloomfield and Marchant*, 2013; *López-Moreno et al.*, 2013; *Li and Rodell*, 2015; *Kumar et al.*, 2016]. We observe very long time scales (>40 months) over regions such as western Africa; southern Africa; southern South America; and parts of Australia, western Russia, and North America (Figure 5.3k). During the analyzed period, those regions have experienced persistent trends in TWS[*Long et al.*, 2017], mainly driven by long-term changes in climatic conditions such as the sustained wetting trend in western and southern Africa[*Maidment et al.*, 2015] and persistent drought conditions in southern South America and Australia[*Chen et al.*, 2010a; *van Dijk et al.*, 2013]. In those regions, the GRACE-DSI is largely influenced by long-term accumulation of climatic water balance,

therefore achieving maximum correlation with long-time-scale SPEI-Z.

4.5.2 Australian case study

From 2002 to late 2009, surface soil moisture (SM-Z) shows no trends but GRACE-DSI indicates significant drying trends in both western and eastern Australia. This suggests gradual depletion of deeper water storage, which is beyond the depth that microwave satellites can sense, such as root-zone moisture and groundwater. Our analyses for location 2 agree with the results from *Leblanc et al.* [2009, 2012] and *van Dijk et al.* [2013] in the Murray-Darling basin, where soil moisture droughts stabilize at low levels since 2002 while groundwater levels gradually decline until late 2009. *Leblanc et al.* [2012] also observe a small increase in total water storage from 2007 to 2008 and estimate that about two-thirds of this increase is used to replenish the shallow surface soil moisture reservoir. This is highly consistent with our results in eastern Australia (Figure 5.4c). The small increase in total water storage in 2008 greatly reduces the magnitude of the drying trend in eastern Australia from 2002 to late 2009.

The NDVI-Z has different degrees of agreement with SM-Z in western and eastern Australia, but has an overall good and consistent agreement with GRACE-DSI in both locations. This suggests that the GRACE-DSI is a more consistent indicator of plant water availability than surface soil moisture across Australia, in agreement with earlier studies[e.g. *Chen et al.*, 2014; *Yang et al.*, 2014; *Wu et al.*, 2015]. From 2002 to late 2009, although SM-Z has no trend, NDVI-Z shows a decreasing trend in western Australia that coincides with the drying trend in GRACE-DSI. This possibly reflects increasing water stress on plants as the drought propagates from the shallow surface to the deep underground. During and after the 2010/11 La Nina-induced extreme wetting, NDVI-Z closely follows GRACE-DSI in western Australia. These together suggest that deep-rooted and groundwater-dependent plants might dominate

vegetation greenness in western Australia. In contrast, NDVI-Z does not show significant decreasing trend from 2002 to late 2009 in eastern Australia, suggesting that vegetation might not be influenced by the gradual depletion of deeper water storage. Overall, NDVI-Z agrees better with SM-Z both in magnitude and temporal variation than with GRACE-DSI in eastern Australia, indicating that shallow-rooted plants might dominate vegetation greenness in this region. This contrasting NDVI response pattern is consistent with the regional distribution of drought-sensitive vegetation in eastern Australia and drought-resistant vegetation in western Australia where water-table depth is much deeper than in the east[*Khan et al.*, 2008; *Fan et al.*, 2013; *De Keersmaecker et al.*, 2015; *Seddon et al.*, 2016].

4.5.3 Short data record length

The initial GRACE-DSI period (2002-14) may not be long enough to sample the full range of wetness and dryness required for a climatological index, for which a climatology of at least 30 years is preferred. To evaluate the potential impact of short data record length on drought characterization, we calculate the RMSE between two PDSI-Z records normalized to 2002-14 and 1982-2014, respectively, using equation4.2. We find that using the 13-yr climatology, PDSI-Z underestimates drought by one category over 15% of the study domain (all land regions excluding Antarctica, Greenland, and barren ground) and by two categories over 1% of the study domain (Figure 4.7). These regions experienced a significantly drier mean climate during 2002-14 than during 1982-2014. We also find that using the 13-yr climatology, PDSI-Z overestimates drought by one category over 12% of the study domain where the 13-yr mean climate is much wetter and/or the hydroclimatic variability is much smaller than the 33-yr climatology. Therefore, when using GRACE-DSI in these regions, we suggest that the stakeholders correct for the category bias to mitigate the short baseline issue. When comparing the GRACE-DSI with other long-term drought metrics, the bias can also be effectively removed by referencing to the 2002-14 period using the normalization

method employed in this study.

4.5.4 GRACE-DSI calculated from JPL mascon solutions

We also calculate GRACE-DSI using the Jet Propulsion Laboratory (JPL) RL05 Mascon solutions[*Watkins et al.*, 2015]. Each mascon solves for the monthly gravity field averaged over an equal-area 3°spherical cap mass concentration block. The spatial representation of the GRACE-DSI from the mascon is therefore different from the spherical harmonics-derived GRACE-DSI for which each 1°grid cell represents conditions averaged over a 350-km radius footprint. We repeat the drought index intercomparison for the mascon-derived GRACE-DSI (see supplemental material for details). Overall, the mascon results are in agreement with the spherical harmonics results except in the Amazon rain forest, where for over a few months we observe north-south stripes (e.g., Figure 4.8) in regions characterized by larger errors in the mascon[*Watkins et al.*, 2015; *Wiese et al.*, 2016]. The stripes are not visible in the spherical harmonics-derived GRACE-DSI and in the other drought indices (Figures. 4.9,4.10,4.11).

4.6 Conclusions

We present a new monthly global DSI developed from satellite-observed time-variable terrestrial water storage changes from GRACE. The initial global GRACE-DSI record extends from 2002 to 2014 and will be updated on a regular basis when newly processed GRACE solutions become available. Although currently not in real time, the GRACE-DSI has potential to provide near-real-time drought monitoring capability with the launch of the GRACE Follow-On mission in 2017. Moreover, the GRACE-DSI has several unique advantages that make it valuable to study drought-related processes and socioecological impacts. The GRACE-DSI

is unaffected by uncertainties associated with traditional model-based indices (e.g., PDSI, SPI, and SPEI), such as dependency on sparse weather station observations or uncertainty from the use of simple water balance models or reanalysis meteorological data. Therefore, it provides an independent observation benchmark for evaluating model-based drought monitoring tools while providing drought information even where sparse ground observations may constrain other approaches.

The GRACE-DSI is sensitive to changes in the bulk terrestrial water storage component of the hydrologic cycle extending from a plants root-zone soil moisture to deeper groundwater. In contrast, PDSI represents the water balance of shallow soil depth, and SPEI-Z assesses the climatic water balance (precipitation minus potential evapotranspiration) over multiple time scales. We interpret the differences between GRACE-DSI and those traditional drought indicators in terms of which hydrological component dominates the drought signal. For instance, the low correlation between GRACE-DSI and PDSI-Z in northern China is due to the difference in trend between TWS and shallow water storage. The time scales of maximum correlation between GRACE-DSI and SPEI-Z indicate which component dominates the hydrologic cycle: maximum correlation occurs with short-time-period SPEI-Z when the hydrological cycle is dominated by changes at shallow depth; maximum correlation occurs with longer-period SPEI-Z when the hydrological cycle changes at depth. For eastern and southeastern Australia, for instance, maximum correlation between GRACE-DSI and SPEI-Z occurs on a longer time scale because the Millennium Drought is dominated by deep-water depletion. The synergistic use of GRACE-DSI with other existing environmental data and drought monitoring tools therefore has potential for improving the characterization of drought (e.g., propagation and recovery) and associated ecological impacts at regional and global scales. In this paper, we demonstrate that combing the GRACE-DSI with other satellite environmental records improves the characterization of the 2000s Australia Millennium Drought, as well as associated vegetation response to water supply changes at surface and subsurface soil layers. A caveat to the application of GRACE-DSI is that it should be

used with caution in glaciated areas where the GRACE-DSI might contain ice mass change signals[*Jacob et al.*, 2012].

4.7 Supplementary Materials

We repeat our drought index inter-comparison using DSI calculated from GRACE mascon solutions. We use NASA JPL version 2 mascon data with no Coastline Resolution Improvement (CRI) filter applied (RL05M_1.MSCNv02, [*Watkins et al.*, 2015]). Each mascon solves for monthly gravity field averaged over an equal-area 3-degree spherical cap mass concentration blocks. For consistency in spatial scale, we process other drought indices for each mascon by averaging all index values within the same mascon.

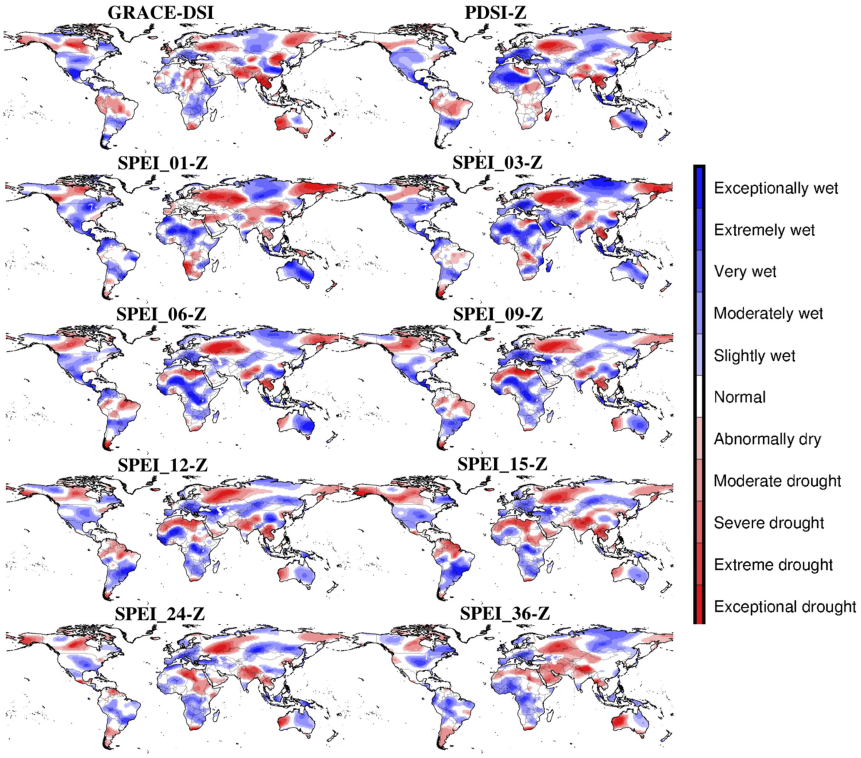

Figure 4.1: Global patterns of the GRACE-DSI, PDSI-Z, and SPEI-Z drought metrics at selected time scales (1,3,6,9,12,15,24,and 36 months) for July 2010.

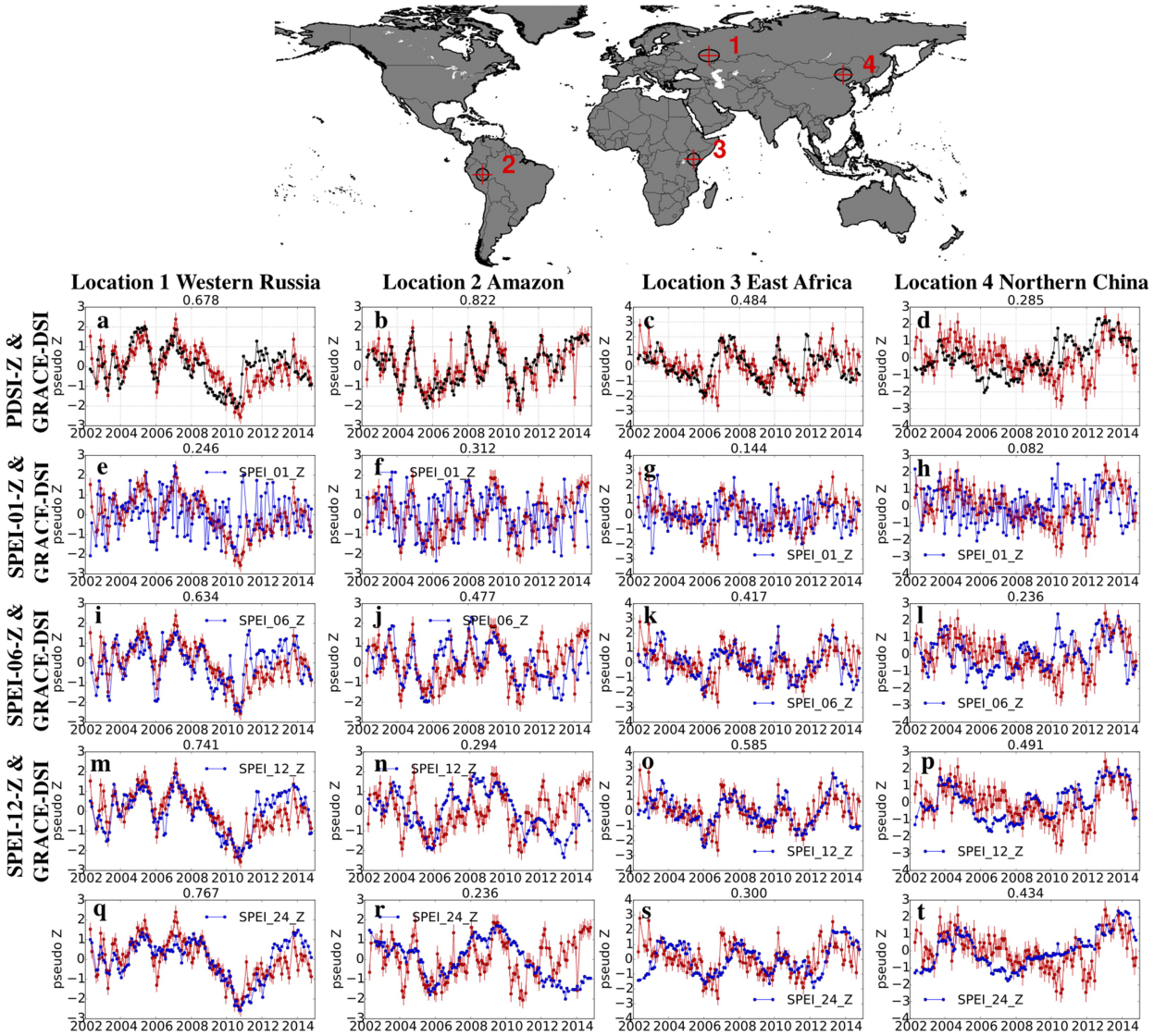

Figure 4.2: Time series of the GRACE-DSI (red), PDSI-Z (black), and SPEI-Z (blue) at selected time scales at four locations annotated with 350-km radius footprints in the map. The geographic coordinates are (54°N, 46°E), (8°S, 72°W), (0°, 38°E), and (44°N, 116°E) for locations 1-4, respectively. Note that the GRACE-DSI is the same for all plots in the same location. Error bar on GRACE-DSI represents the GRACE-DSI uncertainty due to GRACE measurement and leakage errors. Pearson correlation coefficient of each comparison is shown on top of each plot. Correlation coefficients larger than 0.17 are significant at the 95% confidence level.

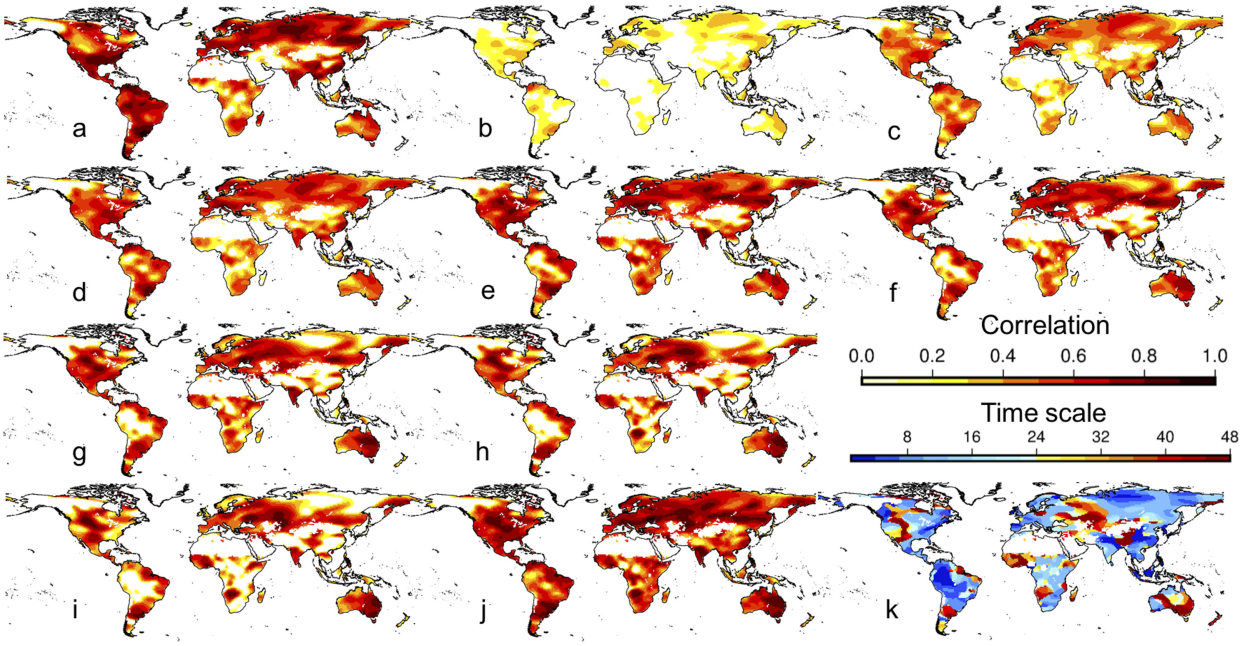

Figure 4.3: Correlation between monthly (a) GRACE-DSI and PDSI-Z and (b)-(i) GRACE-DSI and SPEI-Z at time scales of 1, 3, 6, 12, 18, 27, 36, and 48 months, respectively. (j) Max correlation between monthly GRACE-DSI and SPEI-Z at various time scales. (k) Time scale of SPEI-Z in which the max correlation in (j) is recorded. Insignificant correlation coefficients ($p > 0.05$) are masked out in (a)-(j).

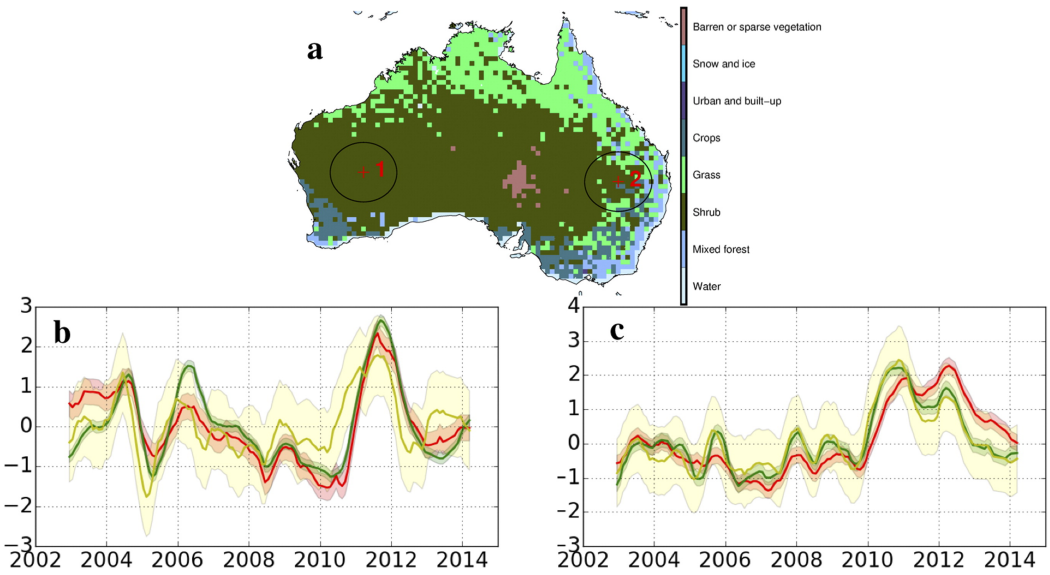

Figure 4.4: (a) Time series of GRACE-DSI (red), satellite-retrieved SM-Z (yellow), and NDVI-Z (green) for two locations in mainland Australia annotated with 350-km footprints in the land-cover map. (b) Location 1 (27°S, 121°E) in western Australia and (c) location 2 (28°S, 148°E) in eastern Australia. Time series are smoothed using a quadratic polynomial filter with a 13-month window[*Savitzky and Golay*, 1964]. Uncertainties of these satellite records are shaded in corresponding colors. The errors of SM-Z and NDVI-Z are estimated conservatively in a similar manner as the GRACE-DSI considering both measurement error and leakage error.

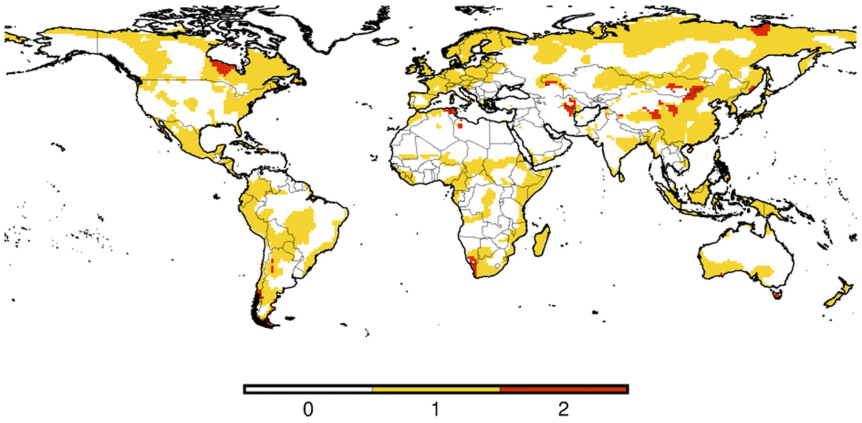

Figure 4.5: Global distribution of GRACE-DSI uncertainty in drought category excluding Antarctica, Greenland, and barren grounds.

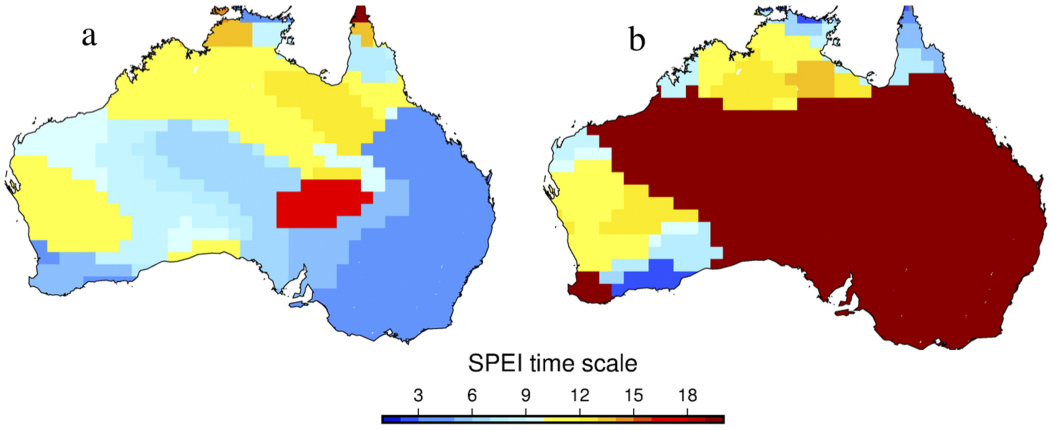

Figure 4.6: (a) Time scale by which NDVI-Z achieves max correlation coefficient with SPEI-Z. (b) The Australia subregion in Figure 5.3, that is, the time scale by which GRACE-DSI achieves max correlation coefficient with SPEI-Z. Note that a large area of (b) saturates at time scales over 20 months. Corresponding max correlation coefficients are significant at the 99% confidence level for both plots.

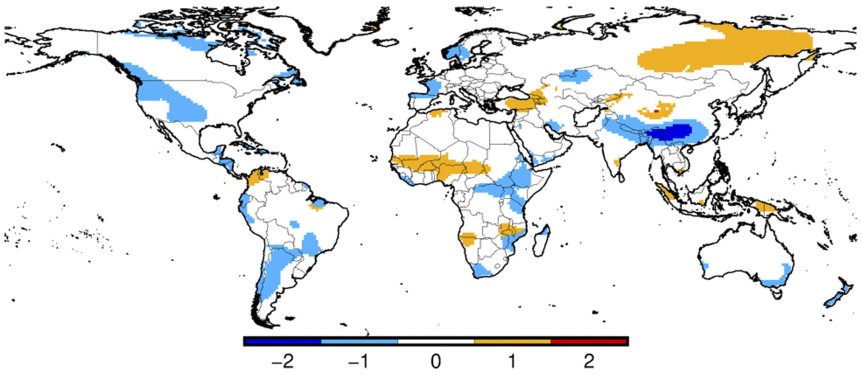

Figure 4.7: Drought category overestimation (positive value) and underestimation (negative value) using the 2002-14 reference period rather than the 1982-2014 reference period for the PDSI-Z drought index.

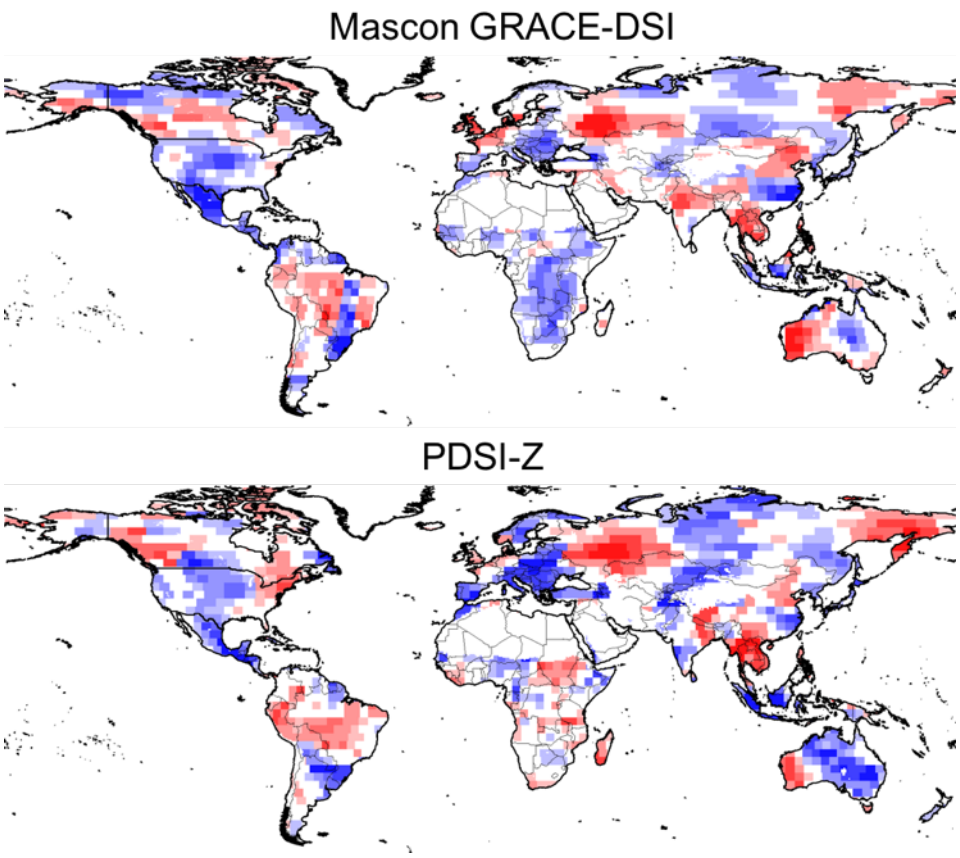

Figure 4.8: Global patterns of the mascon GRACE-DSI and PDSI-Z for July 2010. Color scheme is the same as Figure 5.1 in the main text.

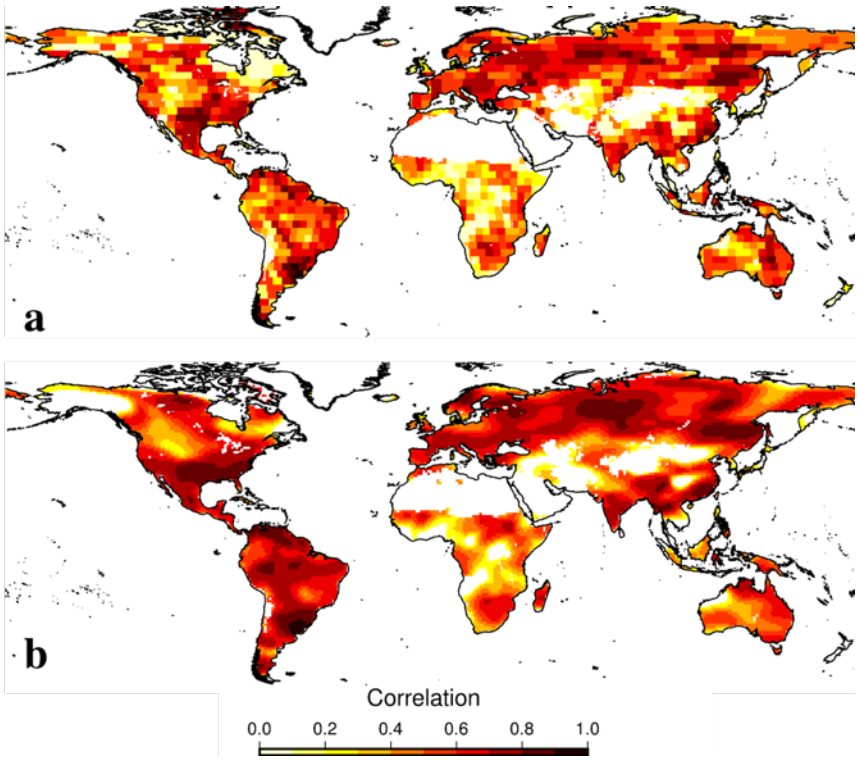

Figure 4.9: (a) is the correlation between mascon GRACE-DSI and PDSI-Z. (b) is the correlation between spherical harmonic GRACE-DSI and PDSI-Z (same as Figure 5.3a in the main paper).

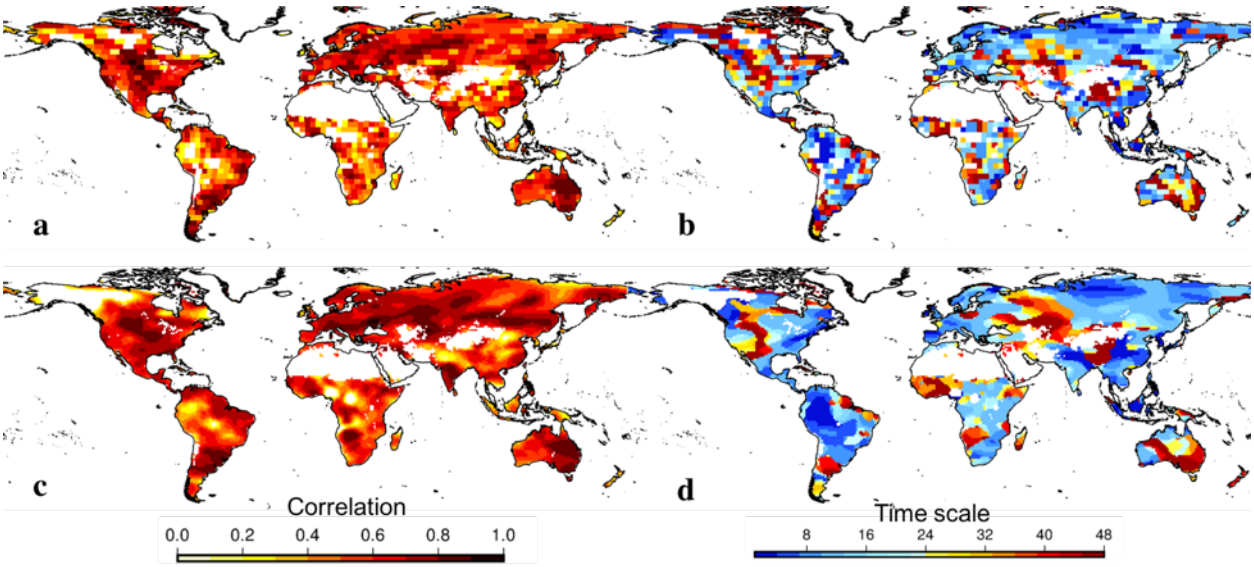

Figure 4.10: (a) is the maximum correlation between mascon GRACE-DSI and SPEI-Z at various time scales. (b) is the time scale of SPEI-Z in which the maximum correlation in (a) is recorded. (c) and (d) are similar to (a) and (b) respectively except for the spherical harmonic processing.

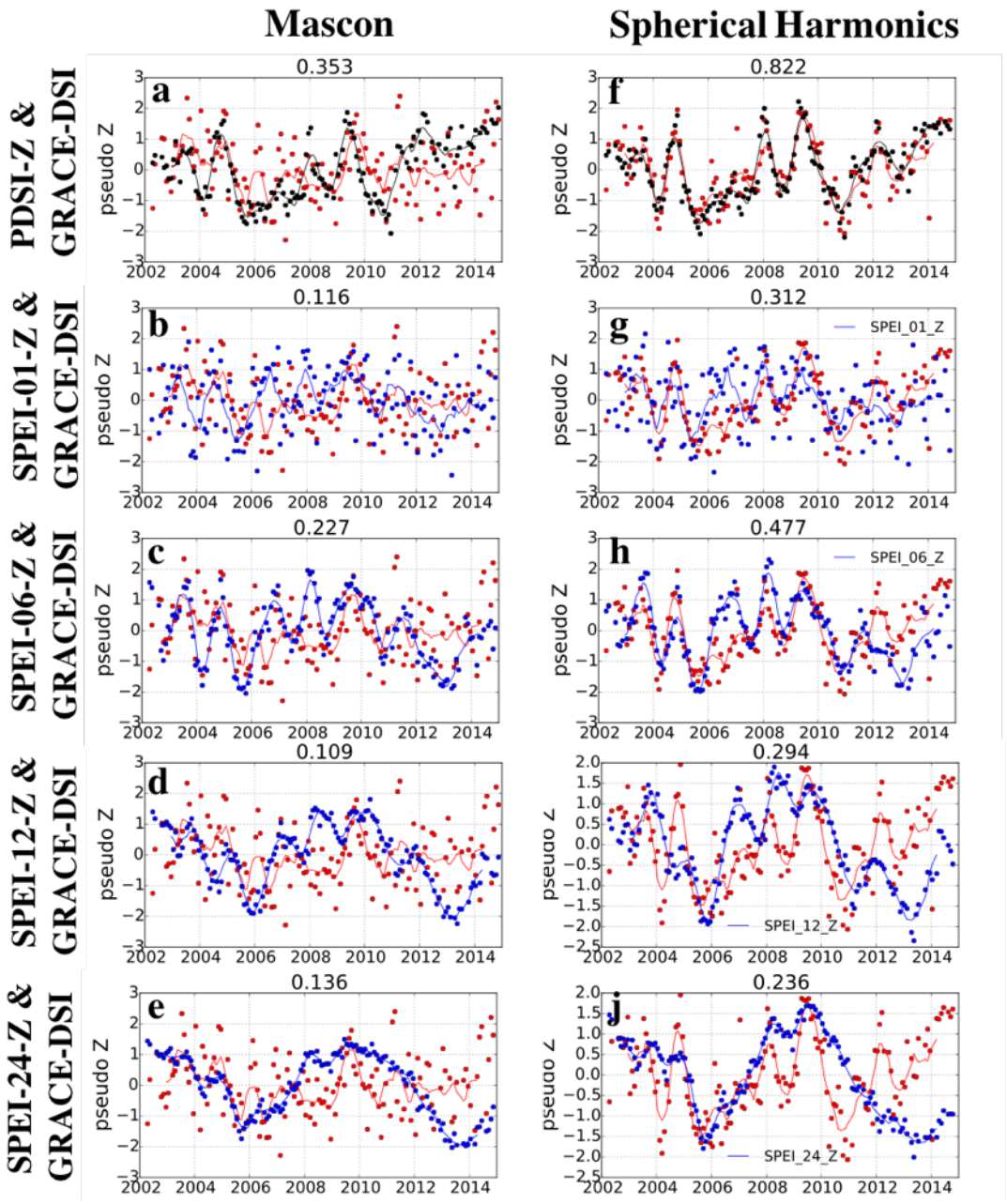

Figure 4.11: (a-e) are time series inter-comparisons of the mascon GRACE-DSI (red), PDSI-Z (black), and SPEI-Z (blue) at selected time scales for a mascon close to the location 2 in the Amazon annotated in the top panel of Figures 5.2. (f-j) are similar to (a-e) respectively except for the spherical harmonic processing. Note that the spatial representation of a mascon and a one-degree grid cell is different. Dots are original index values and lines are smoothed values using a quadratic polynomial filter with a 13-month window[*Savitzky and Golay*, 1964]. Pearson correlation coefficient is shown on top of each plot. Correlation coefficient larger than 0.17 is significant at 95% confidence level.

Chapter 5

Ecological restoration impact on total terrestrial water storage

As appears in:

Zhao, M., G. A, J. Zhang, I. Velicogna, C. Liang, Z. Li, "Ecological restoration impact on total terrestrial water storage," ***Nature Sustainability*** **4, 56-62, 2021.**

5.1 Abstract

Large-scale ecological restoration (ER) has been successful in curbing land degradation and improving ecosystem services. Previous studies have shown that ER changes individual water flux or storage, but its net impact on total water resources remains unknown. Here we quantify ER impact on total terrestrial water storage (TWS) in the Mu Us Sandyland of northern China, a hotspot of ER practices. By integrating multiple satellite observations and government reports, we construct a TWS record that covers both the pre–ER (1982–1998)

and the post-ER (2003–2016) periods. We observe a significant TWS depletion ($P<0.0001$) after ER, a substantial deviation from the pre–ER condition. This contrasts with a TWS increase simulated by an ecosystem model that excludes human interventions, indicating that ER is the primary cause for the observed water depletion. We estimate that ER has consumed TWS at an average rate of 16.6 ± 5.0 mm yr^{-1} in the analysed domain, equivalent to a volume of 21 km^3 freshwater loss during the post-ER period. This study provides a framework that directly informs the water cost of ER. Our findings show that ER can exert excessive pressure on regional water resources. Sustainable ER strategies require optimizing ecosystem water consumption to balance land restoration and water resource conservation.

5.2 Introduction

Since 1998, China has been implementing multiple ecological restoration (ER) programmes to combat desertification, air pollution and climate change[*Bryan et al.*, 2018; *Ouyang et al.*, 2016]. These policy-driven programmes, covering over 6.24 million km^2 of land, have contributed to a strong greening trend and desertification reversion in China[*Bryan et al.*, 2018; *Ouyang et al.*, 2016; *Chen et al.*, 2019; *Tong et al.*, 2018; *Lu et al.*, 2018]. Meanwhile, they have exerted a notable influence on regional water balance[*Ouyang et al.*, 2016; *Feng et al.*, 2016; *Jia et al.*, 2017; *Chen et al.*, 2015; *Tong et al.*, 2020]. Previous studies have primarily focused on evaluating how ER influences hydrological fluxes (for example, evapotranspiration or runoff)[*Feng et al.*, 2016; *Jackson et al.*, 2005] or an individual water storage component (for example, soil moisture)[*Jia et al.*, 2017; *Chen et al.*, 2015; *Tong et al.*, 2020]. Changes in these individual fluxes and storage components affect the land water budget in different directions, leaving their net impact on total water resources unknown (Figure 5.1). For example, recent studies have shown that revegetation increases evapotranspiration (ET) and decreases soil moisture[*Feng et al.*, 2016; *Jia et al.*, 2017; *Tong et al.*, 2020]; however,

the total amount of freshwater resources may remain stable if vegetation recovery increases precipitation (potential ecosystem feedback to climate)[*Li et al.*, 2018; *Green et al.*, 2017; *Branch and Wulfmeyer*, 2019], enhances soil infiltration through root development, and thus increases groundwater recharge and decreases runoff[*Ellison et al.*, 2017]. Quantifying the changes in total water resources is essential for monitoring and understanding the overall water constraint on ER practices[*McDonnell et al.*, 2018; *Rodell et al.*, 2018]. It also directly informs sustainable land and water management for balancing the needs of socioeconomic development and the protection of limited water supply to humans and ecosystems. However, due to limited observations, few studies have detected changes in total water resources before and after ER. Climate variability and other human activities such as groundwater pumping also complicate the attribution of changes in total water resources to ER[*Rodell et al.*, 2009; *Scanlon et al.*, 2018].

In this study, we quantify ER impact on total water resources using direct measurements from the Gravity Recovery and Climate Experiment (GRACE) satellites[*Tapley et al.*, 2004b, 2019] in conjunction with multiple other environmental observations. GRACE measures monthly changes in Earths gravity field with unprecedented precision[*Tapley et al.*, 2004b, 2019]. The primary process driving the measured gravitational variations at monthly timescales is the redistribution of water[*Tapley et al.*, 2004b]. This allows GRACE to monitor changes in total terrestrial water storage (TWS), where TWS = groundwater + soil moisture + surface waters + snow + ice, that were impossible to measure before its launch in 2002. Since GRACE directly measures gravity changes, it integrates the impacts of both natural climate and human activities on TWS[*Tapley et al.*, 2019]. Using GRACE measurements, other available environmental observations such as a long-term vegetation index and a water balance approach, we construct a continuous TWS record covering both the pre- and post-ER periods. To separate the impact of ER on TWS changes, we simulate climate- and atmosphere-driven TWS changes using a dynamic vegetation model that excludes anthropogenic influences. We then remove the simulated climate and atmospheric impact from the constructed TWS

record to estimate the contribution from ER.

We focus our analysis on the Mu Us Sandyland and its surrounding areas in northern China (Figure 5.2a,b and Methods) because they are among the most successful examples of revegetation and reversing desertification[*Tian et al.*, 2015; *Zhang and Huisingh*, 2018]. Climate change combined with human activities such as over-grazing contributed to the desertification of these semiarid regions in history[*Huang et al.*, 2009]. Since 1999, multiple policy-driven ER programmes, such as 'Grain for Green', 'Three-North Shelter Forest' and the afforestation subsidy policy, have reduced grazing pressure and promoted grassland restoration[*Lu et al.*, 2018] and desertification reversion in this region[*Xu et al.*, 2010]. Multi-satellite observations of the normalized difference vegetation index (NDVI) reveal that these regions have experienced a strong vegetation growth since the early 2000s (Figure 5.2c). Several studies suggest that ER programmes and favourable climate conditions jointly drove this trend[*Yan et al.*, 2015; *Li et al.*, 2016; *Xu et al.*, 2018]. Based on global annual land-cover maps from the European Space Agency Climate Change Initiative (ESA CCI)[*Poulter et al.*, 2015], bare and sparsely vegetated areas were rapidly restored to grassland during 1999–2002 (Figure 5.2d). We thus use 1999–2002 as a divider for the pre- and post-ER periods.

ER is the most prominent human practice in our study region since the late 1990s. To stabilize dune fields, local communities create 'checkerboards' of straw or sticks, within which they plant vegetation (such as *Artemisca ordosica, Caragana microphylla, Sabina vulgaris and Salix spp.*)[*Xu et al.*, 2015]. Aerial seeding is also used to increase vegetation cover[*Yan et al.*, 2015], and small-sized croplands such as liquorice are cultivated for economic return. These revegetation activities are expected to enhance the consumption of soil moisture and groundwater via the ET process[*Cheng et al.*, 2013a; *Yu et al.*, 2018]. Outside of ER, other human impacts on water resources are negligible. There is no water diversion or large reservoir construction in this region. Groundwater pumping for crop irrigation in the analysed domain is negligible due to the small area of cultivated croplands[*Feng et al.*, 2016] (Figure

5.2b and Supplementary Discussion). It is not affected by intense groundwater pumping in the North China Plain because the aquifer underneath our study region is disconnected from the aquifers in that plain[*Li et al.*, 2013] (Supplementary Discussion). The impact of coal mining on groundwater consumption is also shown to be small[*Xie et al.*, 2018]. These conditions make the Mu Us Sandyland and its surrounding areas an ideal region to investigate the impact of ER on TWS.

5.3 Observed TWS for the post-ER period (2003-2016)

The GRACE satellite mission was launched in March 2002 and therefore provides direct TWS observations for the post-ER period. We observe an average TWS depletion rate of 7.1 ± 0.6 mm yr^{-1} (hereafter, all uncertainties are reported as 1 σ errors) equivalent water thickness for our study region from 2003 to 2016 (Figure 5.3a and Methods).

The post-ER period is wetter than average (Figure 5.6); and the impact of drought on ecohydrology appears to cover a limited time span. For instance, 2005 recorded the lowest annual precipitation in the period (Figure 5.6). In 2007, NDVI and ET had returned to above their pre-drought levels (Figure 5.2c and 5.7), consistent with previous studies that have shown transitory drought impact on semiarid ecosystems[*Griffin-Nolan et al.*, 2018; *Poulter et al.*, 2014; *Ahlström et al.*, 2015]. For the GRACE TWS trend analysis, excluding 2005 leads to a trend of -7.0 mm yr^{-1}. Excluding 2005 and 2006 leads to a trend of -7.3 mm yr^{-1}. These estimates are consistent with what we find using the entire post-ER period (-7.1 mm yr^{-1}). These results indicate that climate variability and drought are not responsible for the observed decreasing trend in the post-ER TWS.

5.4 Estimated TWS for the pre-ER period (1982-1998)

We estimate the changes in TWS during the pre-ER period based on the balance between precipitation (P), runoff (R) and ET. We use P estimates from the Climate Research Unit (CRU) because they agree well with rain gauge measurements (Figure 5.6b). We calculate R from river gauge measurements. We calculate pre-ER ET by first calculating the post-ER ET using GRACE TWS and the water balance approach (Methods and Figure5.7). We build an empirical relationship between GRACE-derived ET and the NDVI, and use this relationship to extrapolate pre-ER ET (Methods and Figures 5.7 and 5.8). By accumulating P-ET-R on annual scales, we find a TWS trend of 6.8 $\pm$ 10.8 mm yr^{-1} during 1982–1998 (Figure 5.3 and Methods).

The estimates of ET in the pre-ER period rely on the ET-NDVI relationship derived from the post-ER period, which reflects water-use efficiency for the restored grassland. Compared with its natural counterpart, restored grassland is actively selected by ER practitioners and probably features a higher water-use efficiency[*Cho et al.*, 2019]. For this reason, the extrapolation might underestimate the actual ET and lead to an upper-bound estimate of the TWS trend in the pre-ER period. To complement this analysis, we simulate TWS variations using a dynamic vegetation model Lund-Potsdam-Jena General Ecosystem Simulator (LPJ-GUESS) and find an increasing TWS trend of 3.4 $\pm$ 4.8 mm yr^{-1} for the pre-ER period (Figure 5.4a and Methods). The model has a limited water storage capacity and tends to underestimate TWS trends under natural conditions[*Scanlon et al.*, 2018; *Swenson and Lawrence*, 2015](Methods). Therefore, the two methods (one using LPJ-GUESS, the other ET-NDVI extrapolation) provide lower- and upper-bound estimates of the pre-ER TWS trend.

5.5 Separation of ER impact

Our calculations show a significant depletion of TWS ($P<0.0001$) during the post-ER period (Figure 5.3a), a substantial deviation from the pre-ER condition (Figures 5.3b and 5.4a). This change is a result of combined climate and human influences. To isolate the human impact, we simulate the climate-driven variations in TWS using LPJ-GUESS. The model accounts for climate forcing, rising levels of CO_2 and nitrogen deposition, but it does not account for revegetation (Methods). The simulation produces a TWS trend of 9.5 ± 5.0 mm yr^{-1} in the post-ER period, which is driven by increasing precipitation (Figure 5.6) and an underestimated vegetation productivity in the absence of human interventions (Figure 5.4b). By removing the simulated TWS trend from the GRACE results, we estimate that ER has depleted TWS at an average rate of 16.6 ± 5.0 mm yr^{-1} during 2003-2016. This estimate is conservative since LPJ-GUESS probably underestimates the TWS trend due to insufficient storage capacity[*Scanlon et al.*, 2018; *Swenson and Lawrence*, 2015] and the model forcing might be subject to a dry bias since 2011 (Methods).

5.6 Restoration strategies and future TWS trend

Whether the current TWS decreasing trend will continue depends on the interactions between ER strategies and local hydrometeorological conditions, which we illustrate in Figure 5.5. The study domain experienced a gradual increase in precipitation (Figure 5.6), and model results show that natural conditions favoured an increase in TWS since the late 1990s (Figure 5.4a). Ecosystem water consumption associated with initial revegetation would be offset by precipitation, and TWS would increase or remain stable (regime I in Figure 5.5). As ER intensifies, water consumption would surpass the input from precipitation, and TWS would decline (regime II in Figure 5.5). This phase describes the current status of ER in

the study domain, where the sustained greening is at the expense of TWS. Depending on the socioeconomic cost and benefit of trading TWS for revegetation, the local stakeholders may opt to (1) reduce ER practices and regulate plant water uptake by monitoring plant water-use efficiency and thinning densely vegetated areas[*Chen et al.*, 2015]; (2) maintain the current level of ER practices; or (3) elevate revegetation efforts. Strategy (1) will reduce ecosystem water consumption and slow down or reverse the current TWS trend (that is, a shift towards regime I; Figure 5.5). On the other hand, strategies (2) and (3) are expected to exacerbate water consumption, and the current trend in TWS would probably continue. A sustained decrease in land water supply will elevate plant water stress, undermining the normal functioning of the existing plant community. The adoption of these two strategies will therefore require active monitoring of local water table depth and plant survival rates. A decrease in survival rates may indicate that vegetation density is approaching or exceeding the maximum capacity that the local environment can carry[*Feng et al.*, 2016] (regime III in Figure 5.5). At that point, further revegetation will impair both water resources and ER achievements. Regional climate models predict future warming and wetting in the study area[*Guo et al.*, 2017], which is expected to promote both TWS replenishment and plant growth. This will probably provide more room for ER to function under regime I and favour a strategy with reduced human revegetation but more natural regeneration.

5.7 Discussion

We compare our broader spatial and longer timescale results with ground-level observations. Our TWS trend calculations for the pre- and post-ER periods are in line with groundwater statistics in Yellow River Resource Bulletins. Yellow River Resource Bulletins report shallow aquifer water storage changes for two subregions occupying about 15,000 km^2 of our study region (Table 5.1). The cumulative change in total shallow aquifer storage of the two

subregions was 0.043 km^3 from 1980 to 2003, equivalent to an average increase rate of 0.12 mm yr^{-1}. In contrast, the cumulative change from 2003 to 2017 was 0.553 km^3, equivalent to an average rate of 2.45 mm yr^{-1}. This number is smaller than the TWS depletion rate observed by GRACE (7.1 mm yr^{-1}), indicating that soil moisture is probably depleting at a mean rate of 4.7 mm yr^{-1}. This soil moisture depletion rate is consistent with previous research which has shown that the top 4 m total soil moisture content declined at a mean rate of 5.1 mm yr^{-1} from 2004 to 2012 due to human plantations in Shenmu County of Mu Us Sandyland[*Jia et al.*, 2017]. In addition, we show that ER has depleted TWS at a rate of 16.6 mm yr^{-1}, equivalent to a total water depletion of 232 mm during 2003-2016. The decrease in TWS is accompanied by an increase of NDVI from about 0.16 to 0.20 in the study area (Figure 5.2c). The correspondence between TWS loss and NDVI gain is consistent with field-level results from [*Gong et al.*, 2017] (Table 5.2 and Supplementary Discussion).

Previous studies discussing anthropogenic influences on TWS have often considered unsustainable groundwater use for crop irrigation, water diversion and reservoir construction[*Rodell et al.*, 2009; *Feng et al.*, 2013]. For our study region, we conservatively estimate that ER has depleted TWS at a mean rate of 16.6 $\pm$ 5.0 mm yr^{-1} during 20032016. This rate is comparable to groundwater depletion in the nearby North China Plain (22 $\pm$ 3 mm yr^{-1} from 2003 to 2010) and Californias Central Valley (20.4 $\pm$ 3.9 mm yr^{-1} from 2003 to 2010), both of which are heavily influenced by groundwater pumping for crop irrigation[*Feng et al.*, 2013; *Famiglietti et al.*, 2011]. Therefore, we argue that large-scale policy-driven ecological restoration is an important anthropogenic contributor to TWS changes. China has been implementing over 16 policy-driven ER programmes covering two-thirds of its territory[*Bryan et al.*, 2018; *Ouyang et al.*, 2016]. For instance, the 'Three-North Shelter Forest' programme (planned until 2050), covering our study region and 13 provincial regions in northern China, is restoring degraded land via grassland conservation, forest plantations, mountain closure and sandy area revegetation[*Lu et al.*, 2018]. Recent studies have shown strong decreasing trends in surface soil moisture and TWS[*Rodell et al.*, 2018; *Wang et al.*,

2018; *Chen et al.*, 2016] despite increasing precipitation trends over these regions[*Peng and Zhou*, 2017]. Our results indicate that ER may be one of the important factors that drive these decreasing trends. Currently, hydrological models do not consider ER impact on water resources[*Scanlon et al.*, 2018] and therefore they may be underestimating future human stress on freshwater resources.

In summary, we propose an analytical framework to study ER impact on TWS by achieving three objectives: (1) calculate the change in TWS before and after the implementation of ER; (2) isolate climate change and other human influences on the calculated TWS change; and (3) inform the potential impacts of different restoration strategies on future TWS trend. Applying this framework with independent datasets (based on different satellite platforms, dynamic vegetation modelling and government reports), we conclude that ER programmes are the primary cause of the sustained TWS decrease in the Mu Us Sandyland and its surrounding areas. On a global scale, restoring deforested and degraded land is an increasingly popular tactic to enhance carbon sequestration and mitigate climate change[*Ellison et al.*, 2017; *Grassi et al.*, 2017; *Griscom et al.*, 2017; *Bastin et al.*, 2019]. The impact of ER on different hydrological fluxes and storage components differs by soil type, biome and climate zone; therefore, their net impact on total terrestrial water storage also varies by region. Future work is needed to understand how different environmental settings influence ERs impact on total freshwater availability. Our proposed framework is applicable to other restoration-intensive regions for this purpose, which is essential to ensure sustainable utilization of freshwater resources in restoration practices[*Rodell et al.*, 2018].

5.8 Methods

5.8.1 GRACE data and the trend estimate

We use monthly GRACE TWS anomalies from the Jet Propulsion Laboratory (JPL) RL06 version 1 mascon solutions from 2003 to 2016[*Watkins et al.*, 2015]. The mascon solutions are represented on a 0.5°longitude-latitude grid. Each mascon solves for the monthly gravity field averaged over an equal-area 3°× 3°mass concentration block (mascon). The Mu Us Sandyland extends over two mascons (Figure 5.9). To best represent the Mu Us Sandyland, we define a 3°× 3°region (∼ 86,000 km2) that best covers our study area while maintaining the native resolution of a single mascon (Figures 5.2a,b). We calculate the TWS time series of our defined region by averaging the two mascons. We linearly interpolate GRACE missing months from the nearest previous and following non-missing GRACE values[*Rodell et al.*, 2018]. We calculate GRACE TWS error by propagating the error estimates of JPL mascon solutions[*Wiese et al.*, 2016]. We also evaluate the potential signal leakage to our study domain from the groundwater depletion in the North China Plain and find the leakage effect negligible (Figure 5.10 and Supplementary Discussion).

We calculate the TWS trend by simultaneously fitting an annual and a semiannual signal, a linear trend, and a constant to the GRACE time series. From the best fit model, the regression error of the trend is 0.39 mm yr^{-1}. We use a Monte Carlo simulation to estimate the trend error caused by GRACE error. We generate an ensemble of 1,000 normally distributed pseudo GRACE time series with zero mean random noise (σ = GRACE error). We compute the linear trend for each of those 1,000 simulations. The resulting sample trends follow a normal distribution, and we set the 1 s.d. of the sample trends as the trend error caused by GRACE error, which is 0.45 mm yr^{-1}. The overall error of the calculated TWS trend is the summation of the s.d. and regression error in quadrature, namely 0.60 mm yr^{-1}.

5.8.2 NDVI and land cover change

NDVI, calculated as the ratio of the difference between near-infrared and red visible reflectance values to their sum, measures canopy greenness and represents the composite property of leaf area, fractional vegetation cover, canopy structure and total chlorophyll content[*Myneni et al.*, 1995]. Therefore, NDVI is a good proxy for photosynthesis and evapotranspiration, which are driven by light absorption at the top of the canopy, especially in arid and semiarid ecosystems[*Gong et al.*, 2017; *Glenn et al.*, 2007].

Two NDVI datasets are used to calculate the annual mean NDVI, namely the Moderate Resolution Imaging Spectroradiometer (MODIS) Collection 6 MOD13C2 and the latest Global Inventory Modeling and Mapping Studies 3rd generation (GIMMS-3g). MOD13C2 NDVI is available monthly from February 2000 to the present with a spatial resolution of 0.05°. We only use high-quality pixels with pixel reliability of 'good data' or a usefulness value of '0010' or better according to the user manual. GIMMS-3g NDVI is generated from the National Oceanic and Atmospheric Administration's Advanced Very High Resolution Radiometer data, available bi-monthly on a 0.083°grid from July 1981 to December 2015. These data have been corrected for various deleterious effects, such as calibration loss, orbital drift and volcanic eruptions[*Pinzon and Tucker*, 2014]. We only use pixels without apparent issues. We linearly interpolate missing pixel values from nearby time points for both MODIS and GIMMS-3g datasets. To obtain a consistent NDVI record over the entire study period, we calibrate the GIMMS-3g NDVI values to the MODIS reference using a simple linear regression (R^2= 0.97) between their annual values from 2003 to 2015[*Fan and Liu*, 2018]; we use this consistent record for our analysis.

We quantify the land cover change using global annual land-cover maps from the ESA CCI. CCI land-cover maps (v2.0.7) range from 1992 to 2015 with a spatial resolution of 300m. These maps are produced by combining global daily surface reflectance observations from

five satellite systems in a consistent manner[*Poulter et al.*, 2015].

During the 1999-2002 period, the NDVI shows large variability due to the 1999-2000 drought (Figures 5.2c and 5.6). Revegetation in this period is captured by the land conversion from bare and sparsely vegetated to grassland (Figure 5.2d). From the early 2000s to present, the land cover distribution in the study area remains stable while NDVI gradually increases. In general, NDVI reflects changes in plant structure and states; it can capture restoration efforts both on barren land[*Xu et al.*, 2015] and in pre-existing vegetated regions[*Yan et al.*, 2015]. NDVI is therefore a more comprehensive metric than land cover to quantify the impact of ER on biomass.

5.8.3 Precipitation

We obtain monthly precipitation measurements from six stations located within our study region (Figure 5.2b) for the period 1985-2016 from the National Meteorological Administration of China (`http://data.cam.cn`). For an extended spatial coverage, we evaluate the gridded monthly precipitation data from the Tropical Rainfall Measuring Mission (TRMM) 3B43 product (available after 1998 on a 0.25°grid)[*Huffman et al.*, 2007] and CRU TS v.4.03 (available after 1901 on a 0.5°grid)[*Harris et al.*, 2020]. We average each dataset over the grid cells that contain the six stations and compare the aggregated time series with the station average. We find that the TRMM data agree with the station average in the post-ER period, while the CRU data agree in the pre-ER period; but it is too dry since 2011 (Figure 5.6b). We therefore adopt the CRU and the TRMM data to represent the pre-ER and post-ER precipitation, respectively, in the water budget analysis. We quantify the uncertainties in the gridded precipitation datasets using their mean absolute error relative to the station measurements.

5.8.4 Runoff

Half of our study region is contained within the Ordos endorheic region, where water does not drain outside the region[*Wang et al.*, 2018] (Figure 5.11a). A quarter of our study region drains to the Yellow River, and another quarter drains to the Wuding River, a tributary of the Yellow River. We obtain annual runoff during 1982-2013 from the Baijiachuan hydrological station that gauges total runoff out of the Wuding River Basin[*Zhou et al.*, 2013] (Figure 5.11a). However, runoff data is unavailable for the quarter that directly drains to the Yellow River. We estimate its runoff to be 65% of Wuding River runoff because its long-term mean precipitation is 65% of that in the quarter region that drains to the Wuding River. We interpolate for 20142016 by using the 2013 runoff value. We assume a ±50% error on our final runoff estimates (Figure 5.11b). Because the annual runoff is negligible (less than 5% of annual total precipitation amount), even assuming a ±100% error on the final runoff estimates does not change our conclusions despite slightly increasing our error estimates.

5.8.5 Calculation of 2003-2016 evapotranspiration

We calculate ET for the post-restoration period (2003-2016) as the residual of the water budget equation:

$$ET = P - R - dTWS/dt \tag{5.1}$$

where P is precipitation, R is runoff, and dTWS/dt is the change in terrestrial water storage (TWS) for a given period[*Rodell et al.*, 2004]. The dTWS/dt rate on monthly timescales has a low signal-to-noise ratio due to the large month-to-month GRACE error. We therefore calculate ET on annual scales using equation 5.1. To reduce the influence of seasonal variability on dTWS/dt, for a given year we calculate dTWS/dt as the linear trend of a model

that simultaneously fits an annual, a semiannual and a linear trend to the 12 monthly points. The error of dTWS/dt is approximated by the summation of model trend statistic error and GRACE annual error in quadrature. We calculate annual P from TRMM because it has no bias compared to station measurements during 2003-2016 (Figure 5.6b). The error of ET is then approximated by the summation of dTWS/dt error, TRMM precipitation error and runoff error in quadrature (Figure 5.7).

5.8.6 Calculation of 1982-2002 ET

To estimate annual ET before 2003, we first build a regression equation that relates water balance ET and NDVI values during 20032016 (ET $\approx$ slope $\times$ NDVI + intercept; slope = 2886.6 $\pm$ 559.7, intercept = 191.4 $\pm$ 98.1, R^2 = 0.7) (Figure 5.8). We input annual NDVI from 1982 to 2002 into the regression equation to calculate annual ET as well as its 1σ prediction error. We propagate post-ER ET errors to pre-ER ET estimates using a Monte Carlo simulation. We generate an ensemble of 1,000 normally distributed post-ER ET time series with zero mean random noise (σ = our estimated post-ER ET errors). For each simulation, we build the regression equation and input pre-ER NDVI values to simulate pre-ER ET. We set the one s.d. of the 1,000 sample pre-ER ET values at each year as post-ER ET error impact. The final errors of pre-ER ET estimates are approximated by the summation of post-ER ET error impact and 1σ prediction error in quadrature (Figure 5.7).

Here the analysis employs the NDVI values calibrated by the MODIS data. If we use the original GIMMS-3g NDVI as an additional test, we estimate the regression slope, intercept and R^2 to be 2715.8$\pm$727.1, 159.3$\pm$125.5 and 0.6, respectively, for the 20032015 period, consistent with what we find using the MODIS-calibrated data.

5.8.7 The TWS trend before ER

We calculate annual TWS before restoration by accumulating P from CRU, R from river gauge measurements and our ET estimates during 1982-1998 using equation 5.1 (Figure 5.3b). We estimate the TWS trend by fitting a simple linear model to the annual TWS time series. We propagate errors of P, R and ET to TWS using a Monte Carlo simulation. We generate an ensemble of 1,000 normally distributed pseudo P, R and ET time series with zero mean random noise (σ = our estimated errors for P, R and ET, respectively). For each simulation, we accumulate P, R and ET to get a pseudo annual TWS time series and calculate its linear trend. We then calculate the s.d. of the 1,000 sample trends. Finally, we approximate the overall TWS trend error as the summation of the s.d. and trend regression error in quadrature. The large error bar of the estimated TWS trend results from the accumulation of errors when integrating P, ET and R to calculate TWS.

5.8.8 Dynamic vegetation modelling

We simulate the response of ecohydrology to climate and atmospheric drivers using the dynamic vegetation model LPJ-GUESS[*Gerten et al.*, 2004]. We select LPJ-GUESS because it has been used to effectively disentangle climate and atmosphere drivers from the ER impact on carbon sequestration[*Tong et al.*, 2018] and ET[*Feng et al.*, 2016] in China. LPJ-GUESS considers 11 plant functional types for natural vegetation and simulates the occurrence of each plant functional type within a grid cell based on the bioclimatic limits and vegetation dynamics including establishment, growth, succession, competition, mortality and disturbances[*Smith et al.*, 2014]. LPJ-GUESS accounts for the hydrological effects of dynamic changes in vegetation structure and functioning, including potential physiological effects of increasing CO_2[*Haxeltine and Prentice*, 1996]. Policy-driven ER is an emerging human land use and is not represented in the global models[*Chen et al.*, 2019; *Tong et al.*, 2018; *Prestele*

et al., 2017]. For instance, previous research has demonstrated that global models (including LPJ-GUESS) substantially downplay the impact of human restoration on recent greening trends in China despite historical land use data being used to drive these models[*Piao et al.*, 2018]. Therefore, we turn off the land use and land use change module, and only use the model to simulate the impact of climatic forcing on natural vegetation dynamics. The simulations are performed for 1920 to 2016 by applying gridded monthly climate data (temperature, precipitation, sunshine and rain days) from CRU TS 4.03, monthly nitrogen deposition[*Tian et al.*, 2018] and annual mean atmospheric CO_2 concentration[*Etheridge et al.*, 1996; *Keeling et al.*, 1995] as forcing. The model is run from 'bare ground' for 500 years to create an equilibrium state of vegetation. For this spin-up, we use climate forcing from 19201949 to capture realistic climatic variations; we use 1920 values for CO_2 concentration and nitrogen deposition. The simulated vegetation type is predominantly grassland, which is consistent with observations. Although LPJ-GUESS has a fixed soil layer depth of 1.5 m, we find that the water balance (equation 5.1) is closed at each grid cell by substituting total soil moisture for TWS. Therefore, we use total soil moisture to approximate TWS in our study region. We calculate the simulated TWS trend in the same way as for the observations. We assume that the error of the simulated TWS trend is $\pm 50\%$ of the simulated TWS trend itself. This is equivalent to assuming the width of the simulated TWS trend uncertainty interval is equal to the simulated TWS trend itself. The final simulated TWS trend error is approximated by the summation of the $\pm 50\%$ error and trend regression error in quadrature.

Aquifers in our study domain have good connections with soil[*Cheng et al.*, 2013a]. Because LPJ-GUESS does not simulate groundwater, it may lack sufficient storage capacity to accommodate the full range in TWS trend[*Scanlon et al.*, 2018; *Swenson and Lawrence*, 2015]. This might lead to an underestimate of the TWS trend under natural conditions[*Scanlon et al.*, 2018]. Removing the model simulated trend from the GRACE results will therefore provide a conservative estimate of the ER-induced TWS depletion rate. Moreover, we use a self-consistent multivariate climate dataset from the CRU to drive LPJ-GUESS. The dry

bias since 2011 in the CRU precipitation (Figure 5.6) may lead to an underestimate of the post-ER TWS trend under natural conditions.

5.9 Supplementary Materials

5.9.1 Human water use

Human water use is slight in our study region. According to the Yellow River Resource Bulletins (available in Chinese at `http://www.yrcc.gov.cn/other/hhgb/`), from 2003 to 2017, the average human water use from surface and subsurface water storage (including from deep aquifers) is 8.4 mm yr^{-1} for the Ordos endorheic basin (Figure 5.11a). Within this figure, water use for crop irrigation is 5.7 mm yr^{-1}. These numbers are less than 3% of the annual total ET flux in our study region, suggesting human water use is negligible, consistent with previous research[*Feng et al.*, 2016]. These numbers are also small compared to the nearby Northern China Plain[*Feng et al.*, 2013], where groundwater irrigation is estimated to be 60 mm yr^{-1}.

5.9.2 Potential GRACE signal leakage from the North China Plain (NCP)

For the JPL mascon solution (3°by 3°) at high latitudes, there is nominally no leakage between two adjacent mascons. This is based on the analysis of the posteriori covariance matrix (David N. Wiese, NASA JPL, personal communication). At lower latitudes, mascon leakage is generally negligible but not in all cases, and the posterior covariance matrix should be examined (David N. Wiese, NASA JPL, personal communication). As the posteriori covariance matrix is not publicly available, we evaluate the potential leakage from the NCP signal

into mascon #822 and #823 using a synthetic as described below. We uniformly distribute over the NCP region a groundwater mass loss trend of -18.8 mm yr^{-1} (Figure 5.10a), which was calculated based on in-situ groundwater measurements by ref1. To estimate the leakage effect of this signal in mascons #822 and #823, we converted this signal into spherical harmonics and truncated to degree 75 (corresponding to the spatial scale of a 3°JPL mascon), and converted the signal back in the spatial domain on a half-degree grid consistent with the mascon representation. We then averaged all the half-degree grid cells within each mascon to obtain the averaged leakage from the NCP region. Note that having equal-area 3°by 3°mascons as basis functions imply an inherent smoothing function on the data. This is equivalent to applying a de-striping and a 300 km smoothing to the spherical harmonic solutions4,5. We find a potential leakage on the trend over mascons #822 and #823 of -0.07 mm yr^{-1} and +0.67 mm yr^{-1}, respectively, from the NCP (Figure 5.10b). These leakage estimates are very small compared to the observed TWS trends of -3.1 mm yr^{-1} and -11.2 mm yr^{-1}, respectively (Figure 5.9). Therefore, the signal leakage from NCP into our analyzed JPL mascons is negligible. In addition, recent research have shown that JPL mascon solution reproduced NCP groundwater depletion rate calculated from in-situ observations with no leakage correction[*Gong et al.*, 2018]. This suggests that the JPL mascon solution is successful in minimizing the leakage effect in the broader regions of NCP.

5.9.3 TWS loss and ET and NDVI gain from field-level results

We show that ER has increased multi-year mean annual ET by 42.9 mm from the pre-ER period to the post-ER period (Figure 5.7) and has depleted TWS at a rate of 16.6 mm yr^{-1}, equivalent to a total of 232 mm water depletion during 2003-2016. The decrease in TWS is accompanied by an increase of NDVI from about 0.16 to 0.20 in our study area. These results are consistent with field knowledge from in-situ study at Yulin eddy covariance flux tower site located in the Mu Us Sandyland*Gong et al.* [2017]. Gong et al. monitored

variations of ET due to land use/cover change during a four-year period (from July 2011 to June 2015) at a grassland site. Within the tower footprint ($\sim$ 1km by 1km), land degraded from year 1 to year 2. However, the degraded land was gradually restored from year 3 to year 4 with growing season peak NDVI increasing from about 0.21 to 0.25. This NDVI increase is associated with 40.1 mm increase in growing season ET (Table 5.2). Meanwhile, land water storage has reduced by about 333.2-199.8 = 132.4 mm (ET minus Precipitation; Table 5.2). This contrasts with the situation in year 3 where ET minus Precipitation is -38.1 mm. This suggests that ER could have depleted 132.4+38.1 = 170.5 mm potential land water storage. Outside the growing season, Gong et al. find a similar depletion rate in soil moisture and groundwater storage as in the growing season (Figure 5 from Gong et al. 2017). If we linearly extrapolate the growing season depletion rate to the non-growing season, we get an annual land water storage depletion of 170.5/(5month/12month) = 409.2 mm. This number likely represents an upper bound on water storage depletion because vegetation activity (and ET) during the non-growing season is expected to be smaller than that during the growing season. Therefore, field knowledge from Gong et al. (2017) implies that annual peak NDVI change from about 0.21 to 0.25 can induce land water storage depletion in the range of 170.5 to 409.2 mm. Our larger spatial and temporal scale estimate is well within this range.

Table 5.1: Shallow aquifer water storage changes for two subregions of our study region. Data are adopted from the Yellow River Resource Bulletins (YRRB) prepared Yellow River Conservancy Commission of Ministry of Water Resources of the Peoples Republic of China (available in Chinese at `http://www.yrcc.gov.cn/other/hhgb/`)

	Ordos Plateau along Yellow River			Northern Shaanxi Sandy Beach		
year	area (km^2)	annual change (10^8 m^3)	cumulative change (10^8 m^3)	area (km^2)	annual change (10^8 m^3)	cumulative change (10^8 m^3)
1980 to 2003	2150		2.248	126322		-1.820
2003	2150	0.396		12632	0.435	
2004	2150	0.154		12905	-1.89	
2005	2150	-0.512		12905	-0.736	
2006	2150	0.03		12905	-0.27	
2007	2150	-0.17		12905	0.851	
2008	2150	0.069		12905	-2.529	
2009	2150	-0.277		12905	0.361	
2010	2150	-0.214		12905	0.778	
2011	2150	-0.314		12905	-2.602	
2012	2150	-0.082		12905	1.25	
2013	2150	-0.402		12905	-0.336	
2014	2150	-0.08		12905	0.1	
2015	2150	-0.37		12905	0.07	
2016	2150	0.12		12905	1.45	
2017	2150	-0.27		12905	-0.54	
2003 to 2017	2150		-1.922	12905		-3.608

Table 5.2: In-situ measurements taken from Tables 1 and 2 in [*Gong et al.*, 2017]. These values are growing season (May 1st to Sep 30th) totals at Yulin eddy covariance flux tower site

Measurement	Year 3	Year 4
Precipitation (mm)	330.2	199.8
ET (mm)	292.1	332.2
0-160cm soil moisture (m^3 $m^{-}3$)	0.075	0.064
Groundwater level	-3.0	-3.5

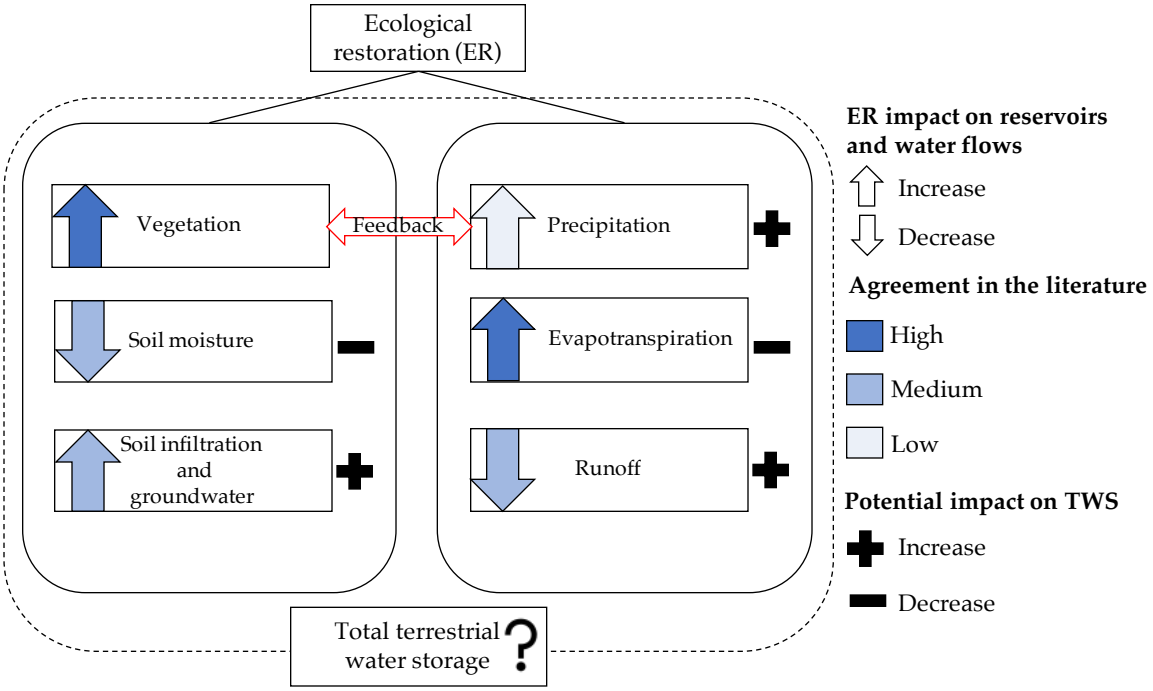

Figure 5.1: Agreement in the literature is analysed from [*Feng et al.*, 2016; *Jia et al.*, 2017; *Chen et al.*, 2015; *Tong et al.*, 2020; *Jackson et al.*, 2005; *Li et al.*, 2018; *Branch and Wulfmeyer*, 2019; *Ellison et al.*, 2017]. The question mark after total TWS suggests that the impact of ER on it is unknown.

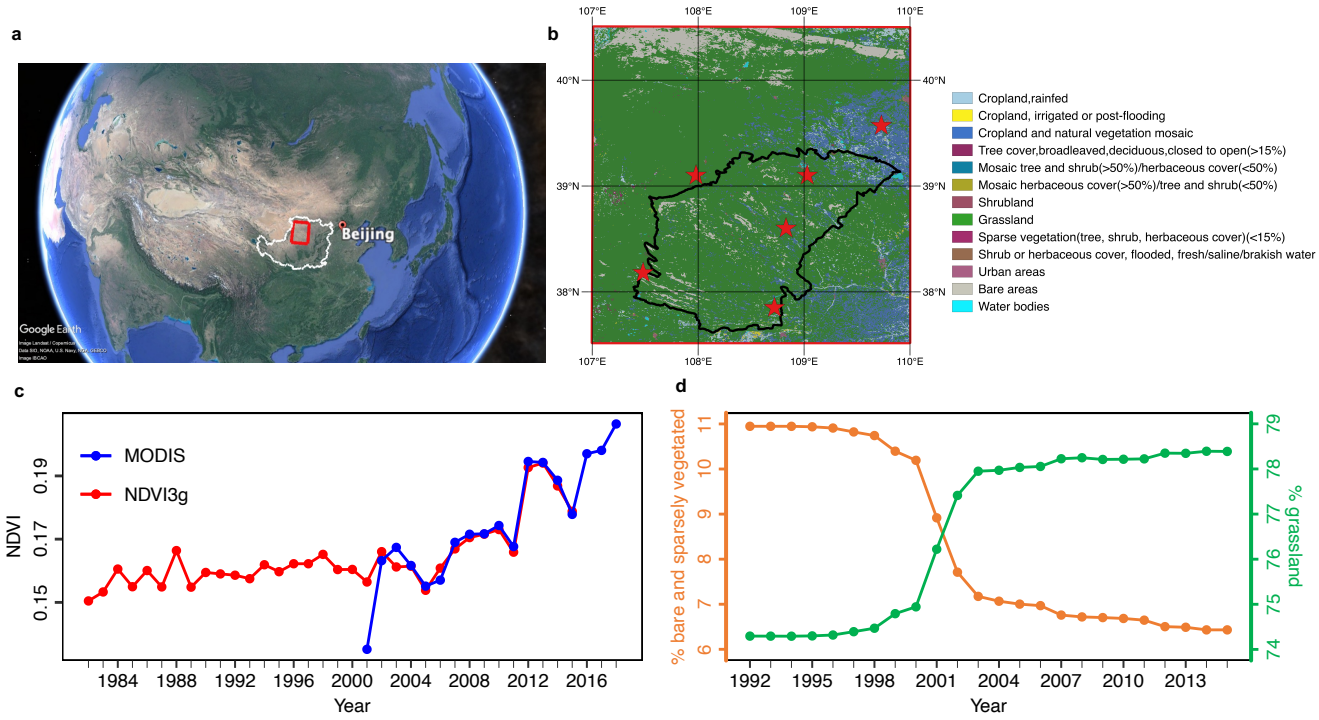

Figure 5.2: (a) Location of our study region (red polygon) within Chinas Loess Plateau (white polygon). (b) Land-cover types of our study region in 2015 from ESA CCI land-cover maps. Red stars represent rain gauge stations. The black polygon represents the Mu Us Sandyland. (c) Annual mean NDVI time series from MODIS and GIMMS-3g averaged over our study region. (d) Time series of major land-cover conversions within our study region from 1992 to 2015 based on ESA CCI land-cover maps. Other land-cover types in b do not change substantially from 1992 to 2015. Credit: Google EarthImage: Landsat/Copernicus; Data: SIO, NOAA, US Navy, NGA, GEBCO; Image: IBCAO (a); European Space Agency Climate Change Initiative (b).

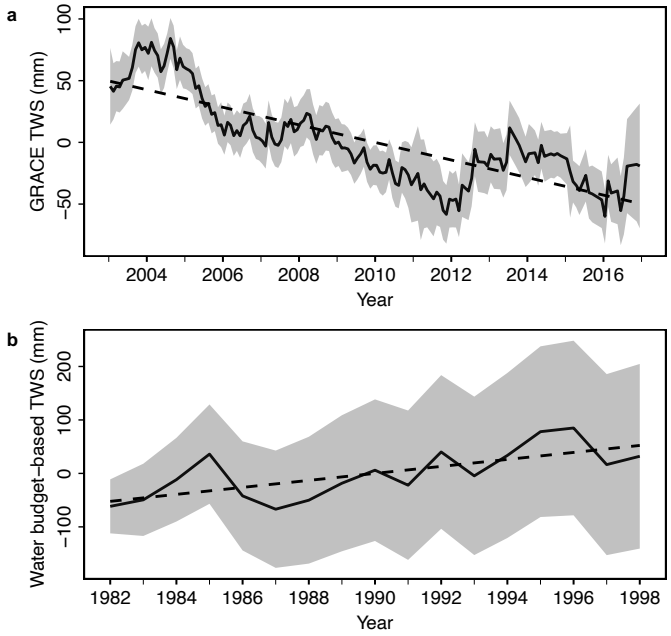

Figure 5.3: (a) GRACE TWS time series from JPL Mascon RL06 solutions from 2003 to 2016. The shaded grey area represents 1 error of TWS. The dashed line represents the best fit linear trend. (b) Estimated TWS time series from 1982 to 1998 calculated by accumulating P, R and our estimated ET using equation (1). The shaded grey area represents 1 error of our estimated TWS. The increasing error band results from the accumulation of errors in P, R and ET when we accumulate PETR to get TWS. The dashed line represents the best fit linear trend.

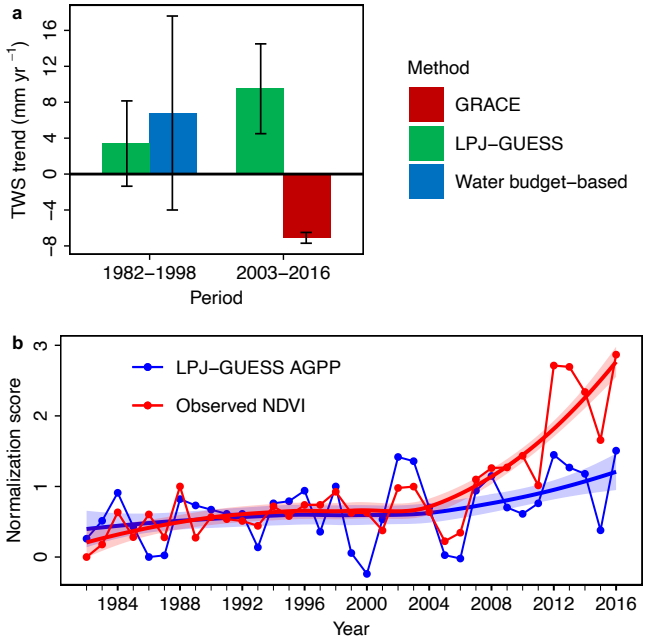

Figure 5.4: (a) GRACE TWS time series from JPL Mascon RL06 solutions from 2003 to 2016. The shaded grey area represents 1 error of TWS. The dashed line represents the best fit linear trend. (b) Estimated TWS time series from 1982 to 1998 calculated by accumulating P, R and our estimated ET using equation (1). The shaded grey area represents 1 error of our estimated TWS. The increasing error band results from the accumulation of errors in P, R and ET when we accumulate PETR to get TWS. The dashed line represents the best fit linear trend.

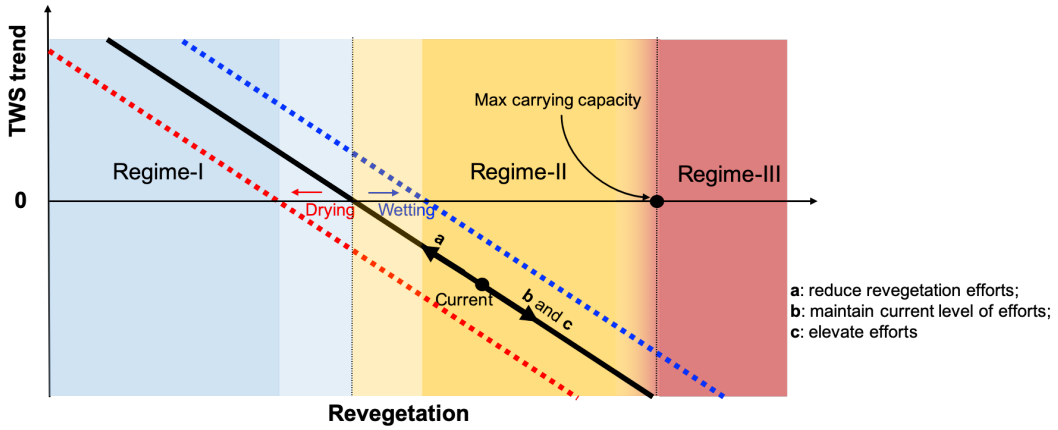

Figure 5.5: The black line illustrates the TWS trends at different revegetation regimes under the current climate in the Mu Us Sandyland. Regime I (blue shade) represents a low level of ER efforts when TWS increases or remains stable. Regime II (yellow shade) presents a higher level of revegetation efforts when plant water consumption exceeds water storage replenishment; TWS declines. Regime III (red shade) occurs when the level of revegetation exceeds the maximum carrying capacity of the local environment. The arrows on the black line illustrate how the TWS trend changes under (1) reduced ER efforts; (2) ER maintained at the current level; and (3) elevated efforts. The blue and red dotted lines illustrate how the TWS trend would shift under a wetting and a drying climate, respectively.

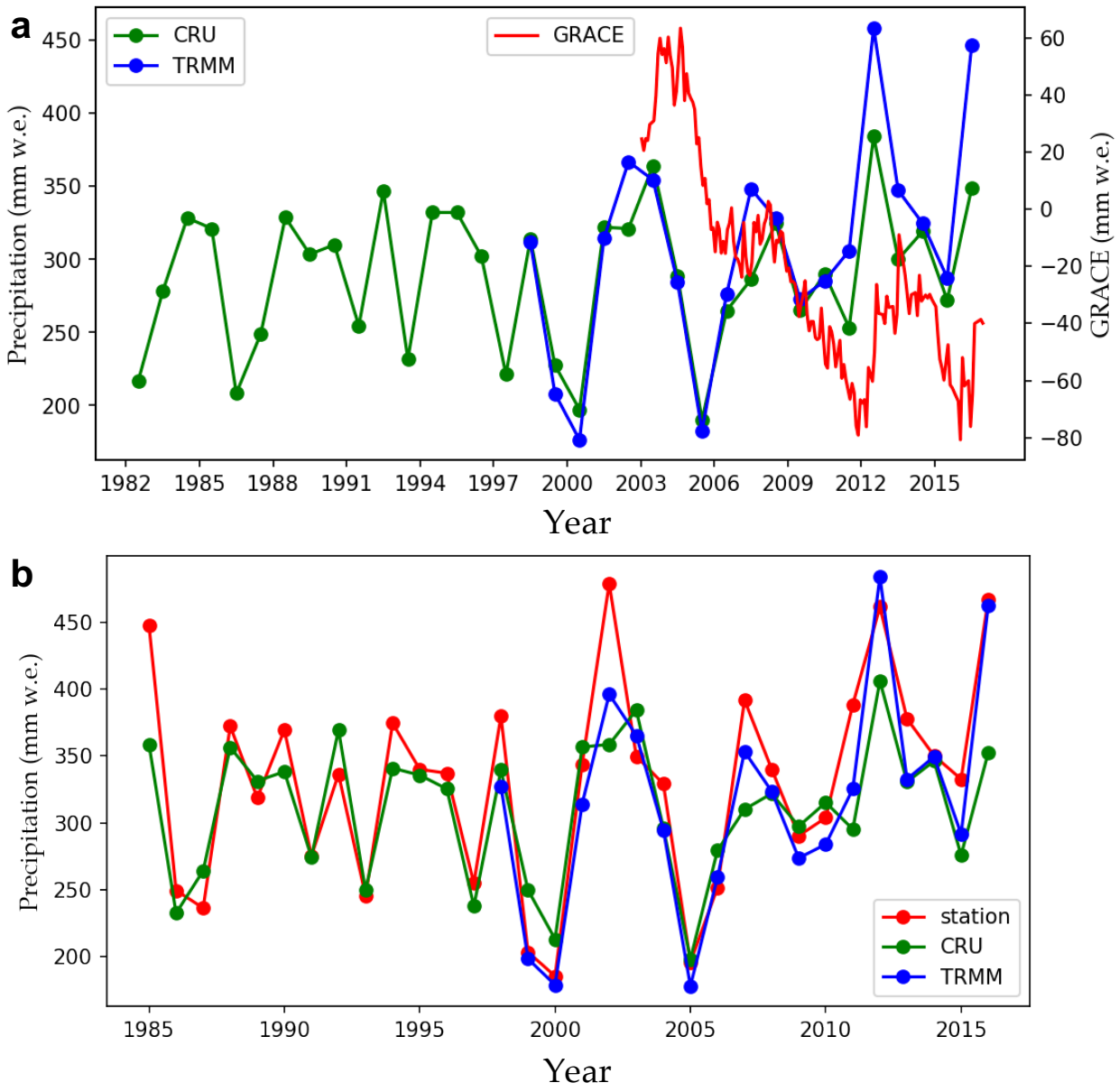

Figure 5.6: (a) GRACE TWS time series and domain-averaged annual total precipitation from the Climate Research Unit (CRU) and the Tropical Rainfall Measuring Mission (TRMM). (b) Gridded precipitation data (TRMM and CRU) with precipitation time series averaged at six weather stations (Figure 5.2b). The correlation coefficients of station vs. CRU and station vs. TRMM are 0.8 and 0.9, respectively.

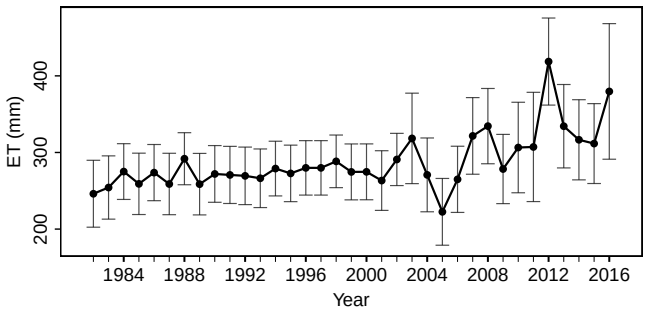

Figure 5.7: ET estimates from 1982–2016. Error bars represent 1-σ errors.

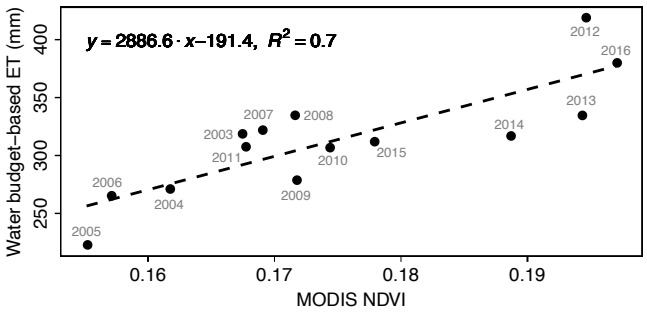

Figure 5.8: Scatterplot of interannual MODIS NDVI values and GRACE water budget-based ET estimates during 2003–2016. The dashed line represents the best fit regression line.

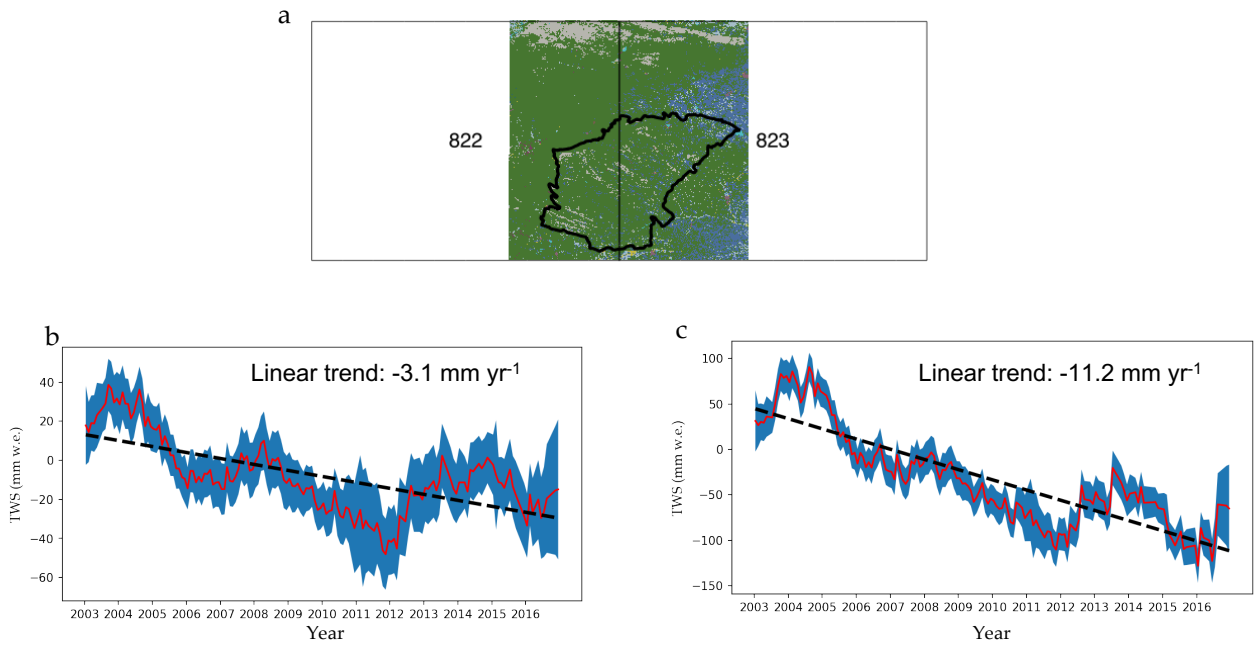

Figure 5.9: (a) The location of JPL mascons (that is #822 and #823) that include Mu Us Sandyland. The land-cover legend is the same as in Figure 5.2b. (b) and (c) are GRACE TWS time series for the two mascons, respectively. The shaded blue area represents 1σ error of TWS.

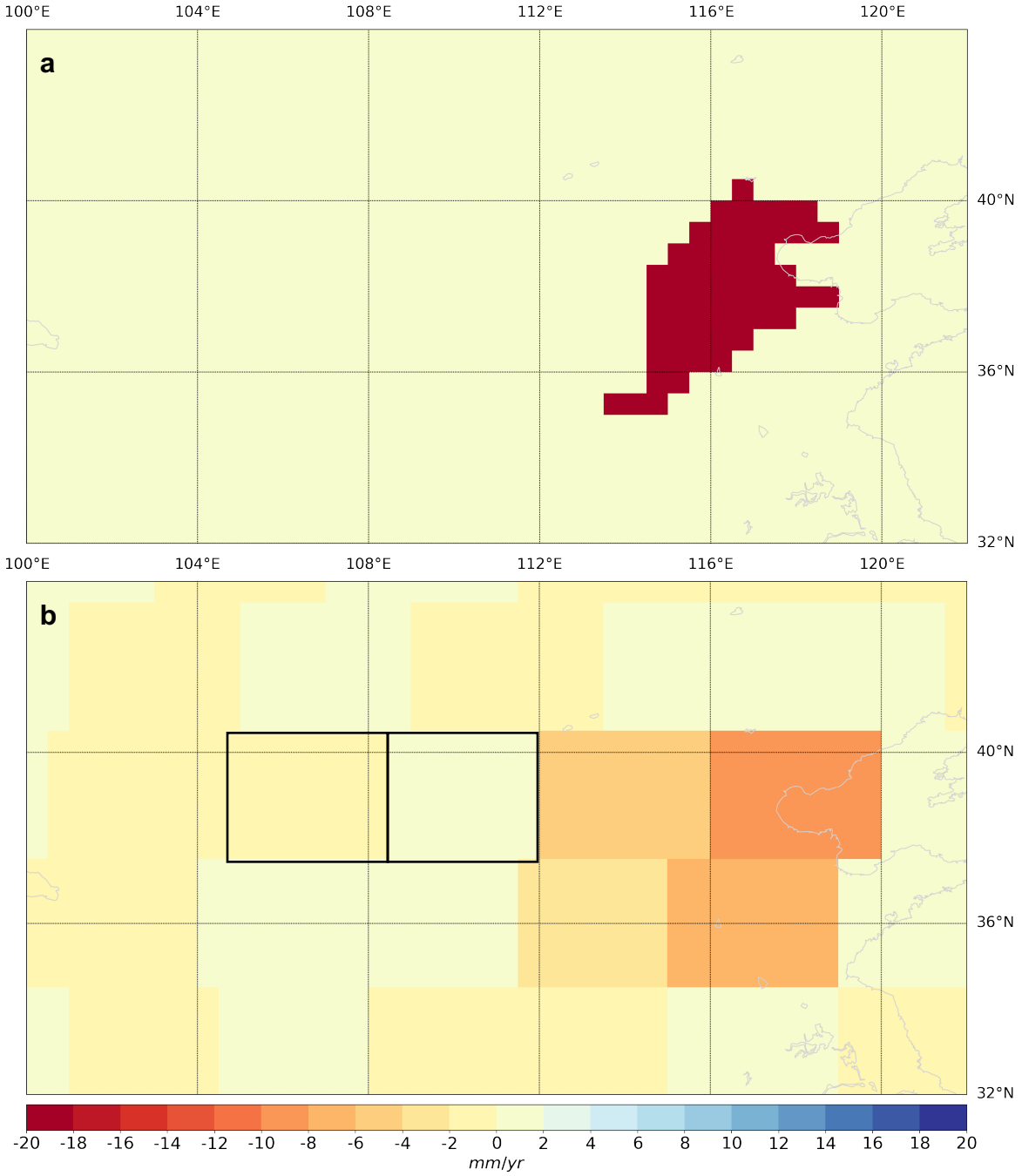

Figure 5.10: Potential signal leakage from the groundwater depletion in North China Plain (NCP) to our study domain. (a) NCP groundwater depletion rate derived from previous research. (b) Mascon representation of (a). The two black boxes in (b) represent JPL mascon #822 and #823.

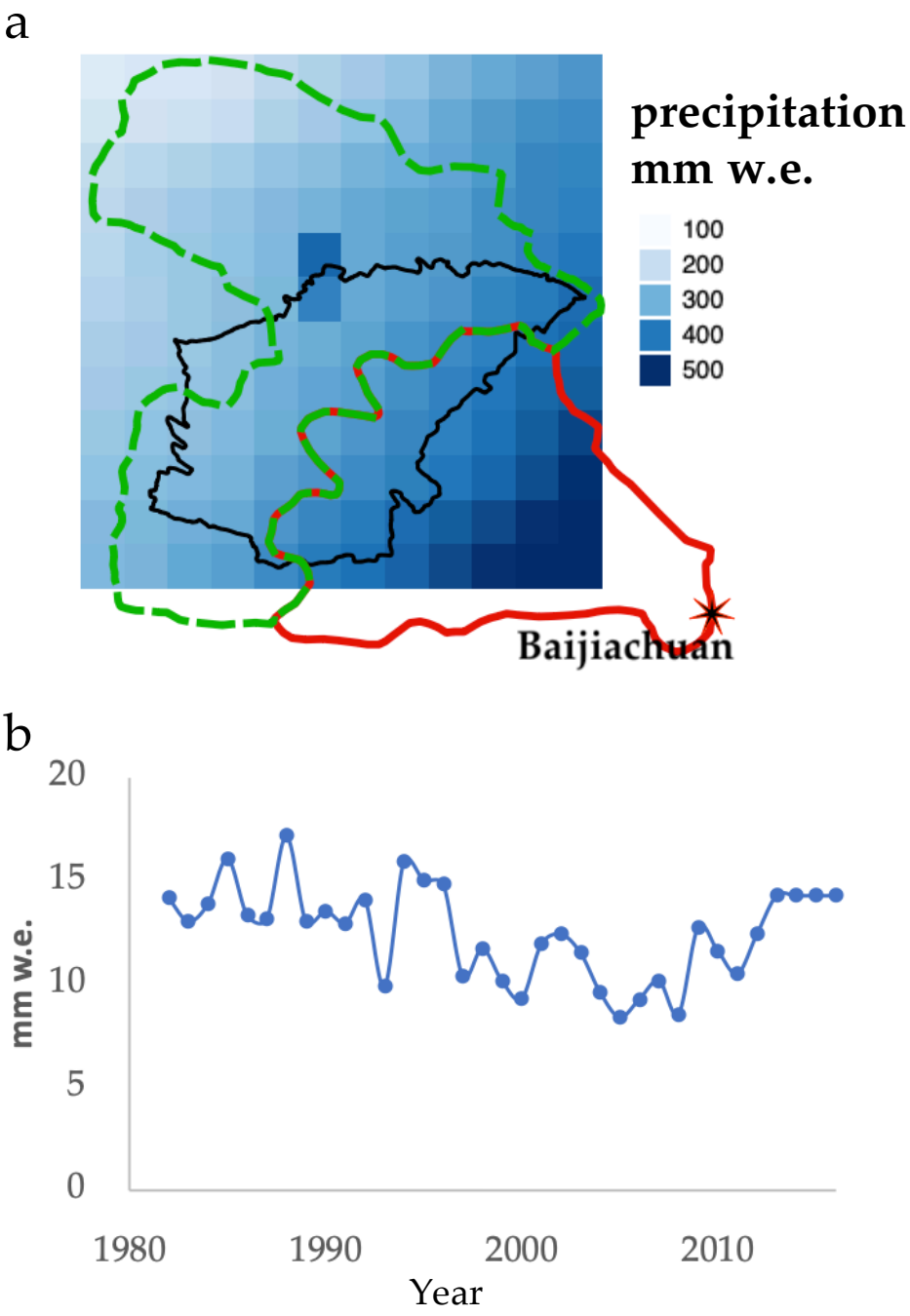

Figure 5.11: (a) Drainage basins in our study region. The green basin is the Ordos endorheic basin. The red basin is the Wuding River drainage basin. Baijiachuan is a hydrological gauge station that measures the total runoff out of the Wuding River basin. The background raster represents the long-term mean annual total precipitation from TRMM during 1998–2018. (b) Time series of estimated runoff for our entire study region.

Chapter 6

Conclusions

Terrestrial ecosystems connect land water, carbon flux, and energy cycles, and have major economic and ecological implications. Terrestrial ecosystems are maintained by complex biological, physical, and chemical processes such as photosynthesis and nutrient cycling which are profoundly affected by the climate. Characterizing these ecological processes are not only important for evaluating and mitigating climate change impacts on ecosystem functions and services, but also to determine the feedback mechanisms that ecosystem response may generate on the climate itself. This work uses satellite remote sensing observations in conjunction with state-of-the-art Earth system models (ESMs) to explore how terrestrial ecosystems respond and feed back to climate.

Atmospheric moisture demand and soil water supply co-regulate the opening of leaf stomata and constrain the land photosynthetic CO2 assimilation during periods of drought. High atmospheric moisture demand induces stomatal closure, which reduces CO2 diffusion in leaves and thus CO2 supply for photosynthesis. Low soil water supply not only causes stomatal closure but also impairs the biochemical capacity for photosynthesis. Accurate understanding of the effects of soil water supply and atmospheric moisture demand on vegetation productivity

is key for the mechanistic understanding of ecosystem vulnerability to drought. Separating the effects of these two constraints is difficult because 1) we lack appropriate observations of terrestrial water supply, and 2) soil water supply and atmospheric moisture demand are often co-linear in time. To overcome the first difficulty, this work calculates a novel soil water supply metric from GRACE TWS. To circumvent the second difficulty, this work analyzes the spatial functional response of global vegetation productivity to atmospheric moisture demand and GRACE-derived water supply metric. This work finds that vegetation productivity is more sensitive to soil water supply than to atmospheric moisture demand. This work also finds an important interaction between soil water supply and atmospheric moisture demand in shaping the spatial distribution of global natural vegetation productivity. This global scale observation-based analysis reflects slow-changing controls on plant-water relations such as vegetation structure, density, community composition, and biogeochemistry that are determined by long-term exposure to hydroclimate.

Due to the process complexity and the lack of observational constraints for model validation, the effects of atmospheric moisture demand versus soil water supply on photosynthesis are parameterized in distinct ways in ESMs. This contributes a large model divergence in projecting climate change impact on terrestrial ecosystem productivity. This work uses the global scale observation-based analysis described in the previous paragraph to evaluate vegetation water stress sensitivities simulated in ESMs from the Coupled Model Intercomparison Project Phase 5 (CMIP5). This work shows that most CMIP5 ESMs fail to reproduce the observation-based water stress sensitivities mainly because they underestimate the interaction between soil water supply and atmospheric moisture demand. The next generation of ESMs are including more realistic physical processes to better simulate vegetation water stress sensitivities such as including explicit representation of plant hydraulic processes linking water supply and atmospheric moisture demand variations. My work provides an effective means to benchmark these model developments.

Climate change such as more frequent heatwaves and drought is threatening our food security and ecosystem by reducing water availability to plants. Characterizing how plants will respond to this changing water availability is imperative for predicting the state of ecosystem under a warming climate. Such characterization is not well-known on a global scale partly because observations of underground water over large areas are limited to shallow soil depth (about 5 centimeters from the surface). These observations are not suitable for studying drought impact on deep-rooted plants such as tropical rainforests and shrubs which can draw water from tens of meters deep into the ground. To overcome this shortfall, I develop a new satellite-based drought severity index (DSI) using the time-variable total water storage estimates (TWS) from the Gravity Recovery and Climate Experiment mission (GRACE). I establish a framework for combing GRACE-DSI with other drought metrics to better characterize hydrological drought, including its propagation, duration, and recovery. I show that GRACE-DSI complements existing proxies of plant water availability by providing information about both root zone soil moisture and deeper groundwater. In combination with other satellite-derived surface soil moisture observations, GRACE-DSI can help elucidate the role of rooting depth in plant-water relations. This work is highlighted in the 2018 March issue of the Bulletin of American Meteorology Society. GRACE-DSI record is now available to the public at `https://www.ess.uci.edu/~velicogna/drought_data.php`.

Large-scale ecological restoration (ER) such as tree planting is an increasingly popular human practice to combat land degradation and climate change. Meanwhile, ER projects have exerted a notable influence on regional water balance. Previous studies show that ER increases evapotranspiration and decreases soil moisture. However, the total amount of freshwater resources may remain stable if vegetation recovery increases precipitation (potential ecosystem feedback to climate) or enhances soil infiltration through root development. To demystify the ER impact on water resources, we must know its effects on total terrestrial water storage. Using GRACE and multiple other satellite observations, government reports, and eco-hydrological modeling, I separated the ER impacts from natural climate variability

for an ER hotspot in northern China. I observed an ERinduced water consumption rate of 16.6 ±5.0 mm yr^{-}1 from 2003 to 2016, an alarming rate comparable to that caused by groundwater pumping for irrigation in Californias Central Valley. This study has important implications for future ER policy development.

Bibliography

A, G., J. Wahr, and S. Zhong, 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*, 557–572, 2013.

A, G., I. Velicogna, J. S. Kimball, and Y. Kim, Impact of changes in grace derived terrestrial water storage on vegetation growth in eurasia, *Environmental Research Letters*, *10*, 124,024, 2015.

AghaKouchak, A., A. Farahmand, F. Melton, J. Teixeira, M. Anderson, B. D. Wardlow, and C. Hain, Remote sensing of drought: Progress, challenges and opportunities, *Reviews of Geophysics*, *53*, 452–480, 2015.

Ahlström, A., et al., The dominant role of semi-arid ecosystems in the trend and variability of the land co2 sink, *Science*, *348*, 895–899, 2015.

Allen, C. D., et al., A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests, *Forest Ecology and Management*, *259*, 660–684, 2010.

Anderson, M. C., C. Hain, B. Wardlow, A. Pimstein, J. R. Mecikalski, and W. P. Kustas, Evaluation of drought indices based on thermal remote sensing of evapotranspiration over the continental united states, *Journal of Climate*, *24*, 2025–2044, 2011.

Anderson, M. C., C. Hain, J. Otkin, X. Zhan, K. Mo, M. Svoboda, B. Wardlow, and A. Pimstein, An intercomparison of drought indicators based on thermal remote sensing and nldas-2 simulations with us drought monitor classifications, *Journal of Hydrometeorology*, *14*, 1035–1056, 2013.

Aumann, H. H., and R. J. Pagano, Atmospheric infrared sounder on the earth observing system, *Optical Engineering*, *33*, 776–785, 1994.

Barriopedro, D., C. M. Gouveia, R. M. Trigo, and L. Wang, The 2009/10 drought in china: possible causes and impacts on vegetation, *Journal of Hydrometeorology*, *13*, 1251–1267, 2012.

Bastin, J.-F., Y. Finegold, C. Garcia, D. Mollicone, M. Rezende, D. Routh, C. M. Zohner, and T. W. Crowther, The global tree restoration potential, *Science*, *365*, 76–79, 2019.

Beard, G., E. Chandler, A. Watkins, and D. Jones, How does the 2010–11 la niña compare with past la niña events, *Bulletin of the Australian Meteorological and Oceanographic Society*, *24*, 17–20, 2011.

Berg, A., J. Sheffield, and P. C. Milly, Divergent surface and total soil moisture projections under global warming, *Geophysical Research Letters*, *44*, 236–244, 2017.

Bloomfield, J., and B. Marchant, Analysis of groundwater drought building on the standardised precipitation index approach, *Hydrology and Earth System Sciences*, *17*, 4769–4787, 2013.

Bonan, G., M. Williams, R. Fisher, and K. Oleson, Modeling stomatal conductance in the earth system: linking leaf water-use efficiency and water transport along the soil–plant–atmosphere continuum, *Geoscientific Model Development*, *7*, 2193–2222, 2014.

Branch, O., and V. Wulfmeyer, Deliberate enhancement of rainfall using desert plantations, *Proceedings of the National Academy of Sciences*, *116*, 18,841–18,847, 2019.

Bryan, B. A., et al., Chinas response to a national land-system sustainability emergency, *Nature*, *559*, 193–204, 2018.

Cao, Y., Z. Nan, and G. Cheng, Grace gravity satellite observations of terrestrial water storage changes for drought characterization in the arid land of northwestern china, *Remote Sensing*, *7*, 1021–1047, 2015.

Carvalhais, N., M. Forkel, M. Khomik, J. Bellarby, M. J. Nature, and 2014, Global covariation of carbon turnover times with climate in terrestrial ecosystems, *nature.com*.

Castle, S. L., B. F. Thomas, J. T. Reager, M. Rodell, S. C. Swenson, and J. S. Famiglietti, Groundwater depletion during drought threatens future water security of the colorado river basin, *Geophysical Research Letters*, *41*, 5904–5911, 2014.

Chatfield, C., *Bivariate Processes*, pp. 169–185, Springer US, Boston, MA, 1975.

Chen, C., et al., China and india lead in greening of the world through land-use management, *Nature sustainability*, *2*, 122–129, 2019.

Chen, J., C. Wilson, B. Tapley, Z. Yang, and G. Niu, 2005 drought event in the amazon river basin as measured by grace and estimated by climate models, *Journal of Geophysical Research: Solid Earth*, *114*, 2009.

Chen, J., C. Wilson, B. Tapley, L. Longuevergne, Z. Yang, and B. Scanlon, Recent la plata basin drought conditions observed by satellite gravimetry, *Journal of Geophysical Research: Atmospheres*, *115*, 2010a.

Chen, J. L., C. R. Wilson, and B. D. Tapley, The 2009 exceptional amazon flood and interannual terrestrial water storage change observed by grace, *Water Resources Research*, *46*, 2010b.

Chen, T., R. De Jeu, Y. Liu, G. Van der Werf, and A. Dolman, Using satellite based soil moisture to quantify the water driven variability in ndvi: A case study over mainland australia, *Remote Sensing of Environment*, *140*, 330–338, 2014.

Chen, X., Y. Su, J. Liao, J. Shang, T. Dong, C. Wang, W. Liu, G. Zhou, and L. Liu, Detecting significant decreasing trends of land surface soil moisture in eastern china during the past three decades (1979–2010), *Journal of Geophysical Research: Atmospheres*, *121*, 5177–5192, 2016.

Chen, Y., I. Velicogna, J. S. Famiglietti, and J. T. Randerson, Satellite observations of terrestrial water storage provide early warning information about drought and fire season severity in the amazon, *Journal of Geophysical Research: Biogeosciences*, *118*, 495–504, 2013.

Chen, Y., K. Wang, Y. Lin, W. Shi, Y. Song, and X. He, Balancing green and grain trade, *Nature Geoscience*, *8*, 739–741, 2015.

Cheng, D.-h., Y. Li, X. Chen, W.-k. Wang, G.-c. Hou, and C.-l. Wang, Estimation of groundwater evaportranspiration using diurnal water table fluctuations in the mu us desert, northern china, *Journal of Hydrology*, *490*, 106–113, 2013a.

Cheng, M., B. D. Tapley, and J. C. Ries, Deceleration in the earth's oblateness, *Journal of Geophysical Research: Solid Earth*, *118*, 740–747, 2013b.

Cho, S., B. Ser-Oddamba, N.-O. Batkhuu, and H. Seok Kim, Comparison of water use efficiency and biomass production in 10-year-old populus sibirica and ulmus pumila plantations in lun soum, mongolia, *Forest Science and Technology*, *15*, 147–158, 2019.

Ciais, P., et al., Europe-wide reduction in primary productivity caused by the heat and drought in 2003, *Nature*, *437*, 529–533, 2005.

Cleveland, W. S., Robust locally weighted regression and smoothing scatterplots, *Journal of the American statistical association*, *74*, 829–836, 1979.

Dai, A., Characteristics and trends in various forms of the palmer drought severity index during 1900–2008, *Journal of Geophysical Research: Atmospheres*, *116*, 2011.

Dai, A., Increasing drought under global warming in observations and models, *Nature Climate Change*, *3*, 52, 2013.

Dai, A., K. E. Trenberth, and T. Qian, A global dataset of palmer drought severity index for 1870–2002: Relationship with soil moisture and effects of surface warming, *Journal of Hydrometeorology*, *5*, 1117–1130, 2004.

De Keersmaecker, W., S. Lhermitte, L. Tits, O. Honnay, B. Somers, and P. Coppin, A model quantifying global vegetation resistance and resilience to short-term climate anomalies and their relationship with vegetation cover, *Global Ecology and Biogeography*, *24*, 539–548, 2015.

De Keersmaecker, W., S. Lhermitte, M. Hill, L. Tits, P. Coppin, and B. Somers, Assessment of regional vegetation response to climate anomalies: A case study for australia using gimms ndvi time series between 1982 and 2006, *Remote Sensing*, *9*, 34, 2017.

Dee, D. P., et al., The era-interim reanalysis: Configuration and performance of the data assimilation system, *Quarterly Journal of the royal meteorological society*, *137*, 553–597, 2011.

Dong, X., et al., Investigation of the 2006 drought and 2007 flood extremes at the southern great plains through an integrative analysis of observations, *Journal of Geophysical Research: Atmospheres*, *116*, 2011.

Ellison, D., et al., Trees, forests and water: Cool insights for a hot world, *Global Environmental Change*, *43*, 51–61, 2017.

Etheridge, D. M., L. Steele, R. L. Langenfelds, R. J. Francey, J.-M. Barnola, and V. Morgan, Natural and anthropogenic changes in atmospheric co2 over the last 1000 years from air in antarctic ice and firn, *Journal of Geophysical Research: Atmospheres*, *101*, 4115–4128, 1996.

Famiglietti, J. S., The global groundwater crisis, *Nature Climate Change*, *4*, 945, 2014.

Famiglietti, J. S., M. Lo, S. L. Ho, J. Bethune, K. Anderson, T. H. Syed, S. C. Swenson, C. R. de Linage, and M. Rodell, Satellites measure recent rates of groundwater depletion in california's central valley, *Geophysical Research Letters*, *38*, 2011.

Fan, X., and Y. Liu, Multisensor normalized difference vegetation index intercalibration: A comprehensive overview of the causes of and solutions for multisensor differences, *IEEE Geoscience and Remote Sensing Magazine*, *6*, 23–45, 2018.

Fan, Y., H. Li, and G. Miguez-Macho, Global patterns of groundwater table depth, *Science*, *339*, 940–943, 2013.

Fan, Y., G. Miguez-Macho, E. G. Jobbágy, R. B. Jackson, and C. Otero-Casal, Hydrologic regulation of plant rooting depth, *Proceedings of the National Academy of Sciences*, *114*, 201712,381–10,577, 2017.

Feng, W., M. Zhong, J.-M. Lemoine, R. Biancale, H.-T. Hsu, and J. Xia, Evaluation of groundwater depletion in north china using the gravity recovery and climate experiment (grace) data and ground-based measurements, *Water Resources Research*, *49*, 2110–2118, 2013.

Feng, X., et al., Revegetation in chinas loess plateau is approaching sustainable water resource limits, *Nature Climate Change*, *6*, 1019–1022, 2016.

Field, C. B., V. Barros, T. F. Stocker, and Q. Dahe, Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change, 2012.

Forkel, M., M. Migliavacca, K. Thonicke, M. Reichstein, S. Schaphoff, U. Weber, and N. Carvalhais, Codominant water control on global interannual variability and trends in land surface phenology and greenness, *Global change biology*, *21*, 3414–3435, 2015.

Frappart, F., F. Papa, J. S. da Silva, G. Ramillien, C. Prigent, F. Seyler, and S. Calmant, Surface freshwater storage and dynamics in the amazon basin during the 2005 exceptional drought, *Environmental Research Letters*, *7*, 044,010, 2012.

Frappart, F., G. Ramillien, and J. Ronchail, Changes in terrestrial water storage versus rainfall and discharges in the amazon basin, *International Journal of Climatology*, *33*, 3029–3046, 2013.

Friedl, M. A., D. Sulla-Menashe, B. Tan, A. Schneider, N. Ramankutty, A. Sibley, and X. Huang, Modis collection 5 global land cover: Algorithm refinements and characterization of new datasets, *Remote sensing of Environment*, *114*, 168–182, 2010.

Gao, H., M. Hrachowitz, S. Schymanski, F. Fenicia, N. Sriwongsitanon, and H. Savenije, Climate controls how ecosystems size the root zone storage capacity at catchment scale, *Geophysical Research Letters*, *41*, 7916–7923, 2014.

Gerten, D., S. Schaphoff, U. Haberlandt, W. Lucht, and S. Sitch, Terrestrial vegetation and water balancehydrological evaluation of a dynamic global vegetation model, *Journal of hydrology*, *286*, 249–270, 2004.

Giardina, F., A. G. Konings, D. Kennedy, S. H. Alemohammad, R. S. Oliveira, M. Uriarte, and P. Gentine, Tall amazonian forests are less sensitive to precipitation variability, *Nature Geoscience*, p. 1, 2018.

Glenn, E. P., A. R. Huete, P. L. Nagler, K. K. Hirschboeck, and P. Brown, Integrating remote sensing and ground methods to estimate evapotranspiration, *Critical Reviews in Plant Sciences*, *26*, 139–168, 2007.

Gong, H., et al., Long-term groundwater storage changes and land subsidence development in the north china plain (1971–2015), *Hydrogeology Journal*, *26*, 1417–1427, 2018.

Gong, T., H. Lei, D. Yang, Y. Jiao, and H. Yang, Monitoring the variations of evapotranspiration due to land use/cover change in a semiarid shrubland, *Hydrology and Earth System Sciences*, *21*, 863–877, 2017.

Grassi, G., J. House, F. Dentener, S. Federici, M. den Elzen, and J. Penman, The key role of forests in meeting climate targets requires science for credible mitigation, *Nature Climate Change*, *7*, 220–226, 2017.

Green, J. K., A. G. Konings, S. H. Alemohammad, J. Berry, D. Entekhabi, J. Kolassa, J.-E. Lee, and P. Gentine, Regionally strong feedbacks between the atmosphere and terrestrial biosphere, *Nature Geoscience*, *10*, 410–414, 2017.

Griffin-Nolan, R. J., C. J. Carroll, E. M. Denton, M. K. Johnston, S. L. Collins, M. D. Smith, and A. K. Knapp, Legacy effects of a regional drought on aboveground net primary production in six central us grasslands, *Plant ecology*, *219*, 505–515, 2018.

Griscom, B. W., et al., Natural climate solutions, *Proceedings of the National Academy of Sciences*, *114*, 11,645–11,650, 2017.

Grossiord, C., et al., Tree water dynamics in a drying and warming world, *Plant, cell & environment*, *40*, 1861–1873, 2017.

Guan, K., et al., Photosynthetic seasonality of global tropical forests constrained by hydroclimate, *Nature Geoscience*, *8*, 284, 2015.

Guanter, L., et al., Global and time-resolved monitoring of crop photosynthesis with chlorophyll fluorescence, *Proceedings of the National Academy of Sciences*, p. 201320008, 2014.

Guo, J., G. Huang, X. Wang, Y. Li, and Q. Lin, Investigating future precipitation changes over china through a high-resolution regional climate model ensemble, *Earth's Future*, *5*, 285–303, 2017.

Haddeland, I., et al., Global water resources affected by human interventions and climate change, *Proceedings of the National Academy of Sciences*, *111*, 3251–3256, 2014.

Han, S.-C., H. Kim, I.-Y. Yeo, P. Yeh, T. Oki, K.-W. Seo, D. Alsdorf, and S. B. Luthcke, Dynamics of surface water storage in the amazon inferred from measurements of intersatellite distance change, *Geophysical Research Letters*, *36*, L09,403, 2009.

Hao, Z., and V. P. Singh, Drought characterization from a multivariate perspective: A review, *Journal of Hydrology*, *527*, 668–678, 2015.

Harris, I., T. J. Osborn, P. Jones, and D. Lister, Version 4 of the cru ts monthly high-resolution gridded multivariate climate dataset, *Scientific data*, *7*, 1–18, 2020.

Haxeltine, A., and I. C. Prentice, Biome3: An equilibrium terrestrial biosphere model based on ecophysiological constraints, resource availability, and competition among plant functional types, *Global biogeochemical cycles*, *10*, 693–709, 1996.

Hoerling, M., J. Eischeid, A. Kumar, R. Leung, A. Mariotti, K. Mo, S. Schubert, and R. Seager, Causes and predictability of the 2012 great plains drought, *Bulletin of the American Meteorological Society*, *95*, 269–282, 2014.

Houborg, R., M. Rodell, B. Li, R. Reichle, and B. F. Zaitchik, Drought indicators based on model-assimilated gravity recovery and climate experiment (grace) terrestrial water storage observations, *Water Resources Research*, *48*, 2012.

Hu, Q., and G. D. Willson, Effects of temperature anomalies on the palmer drought severity index in the central united states, *International Journal of Climatology*, *20*, 1899–1911, 2000.

Huang, Y., N.-a. Wang, T. He, H. Chen, and L. Zhao, Historical desertification of the mu us desert, northern china: a multidisciplinary study, *Geomorphology*, *110*, 108–117, 2009.

Huete, A., K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira, Overview of the radiometric and biophysical performance of the modis vegetation indices, *Remote sensing of environment*, *83*, 195–213, 2002.

Huffman, G. J., D. T. Bolvin, E. J. Nelkin, D. B. Wolff, R. F. Adler, G. Gu, Y. Hong, K. P. Bowman, and E. F. Stocker, The trmm multisatellite precipitation analysis (tmpa): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, *Journal of hydrometeorology*, *8*, 38–55, 2007.

Humphrey, V., J. Zscheischler, P. Ciais, L. Gudmundsson, S. Sitch, and S. I. Seneviratne, Sensitivity of atmospheric CO 2 growth rate to observed changes in terrestrial water storage, *Nature*, *560*, 628–631, 2018.

Huxman, T. E., et al., Convergence across biomes to a common rain-use efficiency, *Nature*, *429*, 651, 2004.

Jackson, R. B., et al., Trading water for carbon with biological carbon sequestration, *science*, *310*, 1944–1947, 2005.

Jacob, T., J. Wahr, W. T. Pfeffer, and S. Swenson, Recent contributions of glaciers and ice caps to sea level rise, *Nature*, *482*, 514, 2012.

Ji, L., and A. J. Peters, Assessing vegetation response to drought in the northern great plains using vegetation and drought indices, *Remote Sensing of Environment*, *87*, 85–98, 2003.

Jia, X., Y. Zhu, Y. Luo, et al., Soil moisture decline due to afforestation across the loess plateau, china, *Journal of Hydrology*, *546*, 113–122, 2017.

Joiner, J., L. Guanter, R. Lindstrot, M. Voigt, A. Vasilkov, E. Middleton, K. Huemmrich, Y. Yoshida, and C. Frankenberg, Global monitoring of terrestrial chlorophyll fluorescence from moderate-spectral-resolution near-infrared satellite measurements: methodology, simulations, and application to gome-2, *Atmospheric Measurement Techniques*, *6*, 2803–2823, 2013.

Karnieli, A., N. Agam, R. T. Pinker, M. Anderson, M. L. Imhoff, G. G. Gutman, N. Panov, and A. Goldberg, Use of ndvi and land surface temperature for drought assessment: Merits and limitations, *Journal of climate*, *23*, 618–633, 2010.

Keeling, C. D., T. P. Whorf, M. Wahlen, and J. Van der Plichtt, Interannual extremes in the rate of rise of atmospheric carbon dioxide since 1980, *Nature*, *375*, 666–670, 1995.

Keyantash, J., and J. A. Dracup, The quantification of drought: An evaluation of drought indices, *Bulletin of the American Meteorological Society*, *83*, 1167–1180, 2002.

Khan, S., H. Gabriel, and T. Rana, Standard precipitation index to track drought and assess impact of rainfall on watertables in irrigation areas, *Irrigation and Drainage Systems*, *22*, 159–177, 2008.

Kim, H., P. J.-F. Yeh, T. Oki, and S. Kanae, Role of rivers in the seasonal variations of terrestrial water storage over global basins, *Geophysical Research Letters*, *36*, L17,402, 2009.

Konings, A., A. Williams, and P. Gentine, Sensitivity of grassland productivity to aridity controlled by stomatal and xylem regulation, *Nature Geoscience*, *10*, 284, 2017.

Kumar, R., J. L. Musuuza, A. F. V. Loon, A. J. Teuling, R. Barthel, J. Ten Broek, J. Mai, L. Samaniego, and S. Attinger, Multiscale evaluation of the standardized precipitation index as a groundwater drought indicator, *Hydrology and Earth System Sciences*, *20*, 1117–1131, 2016.

Landerer, F. W., and S. Swenson, Accuracy of scaled grace terrestrial water storage estimates, *Water resources research*, *48*, 2012.

Lauenroth, W., and O. E. Sala, Long-term forage production of north american shortgrass steppe, *Ecological Applications*, *2*, 397–403, 1992.

Leblanc, M., S. Tweed, A. Van Dijk, and B. Timbal, A review of historic and future hydrological changes in the murray-darling basin, *Global and planetary change*, *80*, 226–246, 2012.

Leblanc, M. J., P. Tregoning, G. Ramillien, S. O. Tweed, and A. Fakes, Basin-scale, integrated observations of the early 21st century multiyear drought in southeast australia, *Water resources research*, *45*, 2009.

Lemordant, L., P. Gentine, A. S. Swann, B. I. Cook, and J. Scheff, Critical impact of vegetation physiology on the continental hydrologic cycle in response to increasing co2, *Proceedings of the National Academy of Sciences*, p. 201720712, 2018.

Levene, H., Robust tests for equality of variances, *Contributions to probability and statistics. Essays in honor of Harold Hotelling*, pp. 279–292, 1960.

Lewis, S. L., P. M. Brando, O. L. Phillips, G. M. van der Heijden, and D. Nepstad, The 2010 amazon drought, *Science*, *331*, 554–554, 2011.

Li, B., and M. Rodell, Evaluation of a model-based groundwater drought indicator in the conterminous us, *Journal of Hydrology*, *526*, 78–88, 2015.

Li, B., M. Rodell, B. F. Zaitchik, R. H. Reichle, R. D. Koster, and T. M. van Dam, Assimilation of grace terrestrial water storage into a land surface model: Evaluation and potential value for drought monitoring in western and central europe, *Journal of Hydrology*, *446*, 103–115, 2012.

Li, Q., Y.-N. Wei, G. Liu, M. Jing, M. Zhang, W. Fei, and X. Li, Feasibility of the combination of co2 geological storage and saline water development in sedimentary basins of china, *Energy Procedia*, *37*, 4511–4517, 2013.

Li, S., W. Liang, B. Fu, Y. Lü, S. Fu, S. Wang, and H. Su, Vegetation changes in recent large-scale ecological restoration projects and subsequent impact on water resources in china's loess plateau, *Science of the Total Environment*, *569*, 1032–1039, 2016.

Li, Y., et al., Divergent hydrological response to large-scale afforestation and vegetation greening in china, *Science Advances*, *4*, eaar4182, 2018.

Lin, Y.-S., et al., Optimal stomatal behaviour around the world, *Nature Climate Change*, *5*, 459–464, 2015.

Liu, Y. Y., R. M. Parinussa, W. A. Dorigo, R. A. M. De Jeu, W. Wagner, A. I. J. M. van Dijk, M. F. McCabe, and J. P. Evans, Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals, *Hydrology and Earth System Sciences*, *15*, 425–436, 2011.

Liu, Y. Y., W. A. Dorigo, R. Parinussa, R. A. de Jeu, W. Wagner, M. F. McCabe, J. Evans, and A. Van Dijk, Trend-preserving blending of passive and active microwave soil moisture retrievals, *Remote Sensing of Environment*, *123*, 280–297, 2012.

Lobell, D. B., M. J. Roberts, W. Schlenker, N. Braun, B. B. Little, R. M. Rejesus, and G. L. Hammer, Greater sensitivity to drought accompanies maize yield increase in the us midwest, *Science*, *344*, 516–519, 2014.

Long, D., B. R. Scanlon, L. Longuevergne, A. Y. Sun, D. N. Fernando, and H. Save, Grace satellite monitoring of large depletion in water storage in response to the 2011 drought in texas, *Geophysical Research Letters*, *40*, 3395–3401, 2013.

Long, D., Y. Pan, J. Zhou, Y. Chen, X. Hou, Y. Hong, B. R. Scanlon, and L. Longuevergne, Global analysis of spatiotemporal variability in merged total water storage changes using multiple grace products and global hydrological models, *Remote sensing of environment*, *192*, 198–216, 2017.

López-Moreno, J. I., S. M. Vicente-Serrano, J. Zabalza, S. Beguería, J. Lorenzo-Lacruz, C. Azorin-Molina, and E. Morán-Tejeda, Hydrological response to climate variability at different time scales: A study in the ebro basin, *Journal of hydrology*, *477*, 175–188, 2013.

Lu, F., et al., Effects of national ecological restoration projects on carbon sequestration in china from 2001 to 2010, *Proceedings of the National Academy of Sciences*, *115*, 4039–4044, 2018.

Luo, L., and E. F. Wood, Monitoring and predicting the 2007 us drought, *Geophysical Research Letters*, *34*, 2007.

Maidment, R. I., R. P. Allan, and E. Black, Recent observed and simulated changes in precipitation over africa, *Geophysical Research Letters*, *42*, 8155–8164, 2015.

Maness, H., P. J. Kushner, and I. Fung, Summertime climate response to mountain pine beetle disturbance in british columbia, *Nature Geoscience*, *6*, 65, 2013.

Marchin, R. M., A. A. Broadhead, L. E. Bostic, R. R. Dunn, and W. A. Hoffmann, Stomatal acclimation to vapour pressure deficit doubles transpiration of small tree seedlings with warming, *Plant, cell & environment*, *39*, 2221–2234, 2016.

McCandless, S. E., Utilizing grace tws, ndvi, and precipitation for drought identification and classification in texas, Master's thesis, The University of Texas at Austin, 2014.

McDonnell, J., et al., Water sustainability and watershed storage, *Nature Sustainability*, *1*, 378–379, 2018.

McDowell, N. G., and C. D. Allen, Darcy's law predicts widespread forest mortality under climate warming, *Nature Climate Change*, *5*, 669, 2015.

McDowell, N. G., et al., Multi-scale predictions of massive conifer mortality due to chronic temperature rise, *Nature Climate Change*, *6*, 295, 2016.

McKee, T. B., N. J. Doesken, J. Kleist, et al., The relationship of drought frequency and duration to time scales, in *Proceedings of the 8th Conference on Applied Climatology*, vol. 17, pp. 179–183, American Meteorological Society Boston, MA, 1993.

Mishra, A. K., and V. P. Singh, A review of drought concepts, *Journal of hydrology*, *391*, 202–216, 2010.

Mu, Q., M. Zhao, J. S. Kimball, N. G. McDowell, and S. W. Running, A remotely sensed global terrestrial drought severity index, *Bulletin of the American Meteorological Society*, *94*, 83–98, 2013.

Myneni, R. B., F. G. Hall, P. J. Sellers, and A. L. Marshak, The interpretation of spectral vegetation indexes, *IEEE Transactions on Geoscience and Remote Sensing*, *33*, 481–486, 1995.

Myneni, R. B., C. Keeling, C. J. Tucker, G. Asrar, and R. R. Nemani, Increased plant growth in the northern high latitudes from 1981 to 1991, *Nature*, *386*, 698, 1997.

Nash, J. E., and J. V. Sutcliffe, River flow forecasting through conceptual models part ia discussion of principles, *Journal of hydrology*, *10*, 282–290, 1970.

Nemani, R. R., C. D. Keeling, H. Hashimoto, W. M. Jolly, S. C. Piper, C. J. Tucker, R. B. Myneni, and S. W. Running, Climate-driven increases in global terrestrial net primary production from 1982 to 1999, *science*, *300*, 1560–1563, 2003.

Njoku, E. G., and D. Entekhabi, Passive microwave remote sensing of soil moisture, *Journal of Hydrology*, *184*, 101 – 129, 1996, soil Moisture Theories and Observations.

NOAA, State of the climate: National climate report for annual 2011, *Tech. rep.*, National Centers for Environmental Information, 2012, retrieved on April 22, 2019 from https://www.ncdc.noaa.gov/sotc/national/201113.

NOAA, State of the climate: National climate report for annual 2014, *Tech. rep.*, National Centers for Environmental Information, 2015, retrieved on April 22, 2019 from https://www.ncdc.noaa.gov/sotc/national/201413.

Novick, K. A., et al., The increasing importance of atmospheric demand for ecosystem water and carbon fluxes, *Nature Climate Change*, *6*, 1023–1027, 2016.

Ouyang, Z., et al., Improvements in ecosystem services from investments in natural capital, *Science*, *352*, 1455–1459, 2016.

Palmer, W. C., Meteorological drought, *US weather bureau research paper*, *58*, 1965.

Parrett, C., and R. A. Hunrichs, Storms and flooding in california in december 2005 and january 2006-a preliminary assessment, *Tech. rep.*, 2006.

Peng, D., and T. Zhou, Why was the arid and semiarid northwest china getting wetter in the recent decades?, *Journal of Geophysical Research: Atmospheres*, *122*, 9060–9075, 2017.

Perez-Martin, A., J. Flexas, M. Ribas-Carbó, J. Bota, M. Tomàs, J. Infante, and A. Díaz-Espejo, Interactive effects of soil water deficit and air vapour pressure deficit on mesophyll conductance to co2 in vitis vinifera and olea europaea, *Journal of Experimental Botany*, *60*, 2391–2405, 2009.

Piao, S., et al., Lower land-use emissions responsible for increased net land carbon sink during the slow warming period, *Nature Geoscience*, *11*, 739–743, 2018.

Pickett, S. T., Space-for-time substitution as an alternative to long-term studies, in *Long-term studies in ecology*, pp. 110–135, Springer, 1989.

Pinzon, J. E., and C. J. Tucker, A non-stationary 1981–2012 avhrr ndvi3g time series, *Remote sensing*, *6*, 6929–6960, 2014.

Ponce-Campos, G. E., et al., Ecosystem resilience despite large-scale altered hydroclimatic conditions, *Nature*, *494*, 349, 2013.

Poulter, B., et al., Contribution of semi-arid ecosystems to interannual variability of the global carbon cycle, *Nature*, *509*, 600–603, 2014.

Poulter, B., et al., Plant functional type classification for earth system models: results from the european space agency's land cover climate change initiative, *Geoscientific Model Development*, *8*, 2315–2328, 2015.

Prestele, R., A. Arneth, A. Bondeau, N. d. Noblet-Ducoudré, T. A. Pugh, S. Sitch, E. Stehfest, and P. H. Verburg, Current challenges of implementing anthropogenic land-use and land-cover change in models contributing to climate change assessments, *Earth System Dynamics*, *8*, 369–386, 2017.

Qin, Y., D. Yang, H. Lei, K. Xu, and X. Xu, Comparative analysis of drought based on precipitation and soil moisture indices in haihe basin of north china during the period of 1960–2010, *Journal of Hydrology*, *526*, 55–67, 2015.

Randerson, J. T., et al., Systematic assessment of terrestrial biogeochemistry in coupled climate–carbon models, *Global Change Biology*, *15*, 2462–2484, 2009.

Reager, J., A. Thomas, E. Sproles, M. Rodell, H. Beaudoing, B. Li, and J. Famiglietti, Assimilation of grace terrestrial water storage observations into a land surface model for the assessment of regional flood potential, *Remote Sensing*, *7*, 14,663–14,679, 2015.

Reichstein, M., et al., Climate extremes and the carbon cycle, *Nature*, *500*, 287–295, 2013.

Rempe, D. M., and W. E. Dietrich, Direct observations of rock moisture, a hidden component of the hydrologic cycle, *Proceedings of the National Academy of Sciences*, *115*, 2664–2669, 2018.

Restaino, C. M., D. L. Peterson, and J. Littell, Increased water deficit decreases douglas fir growth throughout western us forests, *Proceedings of the National Academy of Sciences*, *113*, 9557–9562, 2016.

Richardson, A. D., T. F. Keenan, M. Migliavacca, Y. Ryu, O. Sonnentag, and M. Toomey, Climate change, phenology, and phenological control of vegetation feedbacks to the climate system, *Agricultural and Forest Meteorology*, *169*, 156–173, 2013.

Richardson, A. D., et al., Terrestrial biosphere models need better representation of vegetation phenology: results from the North American Carbon Program Site Synthesis, *Global Change Biology*, *18*, 566–584, 2012.

Rodell, M., J. Famiglietti, J. Chen, S. Seneviratne, P. Viterbo, S. Holl, and C. Wilson, Basin scale estimates of evapotranspiration using grace and other observations, *Geophysical Research Letters*, *31*, 2004.

Rodell, M., I. Velicogna, and J. S. Famiglietti, Satellite-based estimates of groundwater depletion in india, *Nature*, *460*, 999–1002, 2009.

Rodell, M., J. S. Famiglietti, D. N. Wiese, J. Reager, H. K. Beaudoing, F. W. Landerer, and M.-H. Lo, Emerging trends in global freshwater availability, *Nature*, *557*, 651–659, 2018.

Rogers, A., et al., A roadmap for improving the representation of photosynthesis in Earth system models, *New Phytologist*, *213*, 22–42, 2016.

Rulinda, C. M., A. Dilo, W. Bijker, and A. Stein, Characterising and quantifying vegetative drought in east africa using fuzzy modelling and ndvi data, *Journal of Arid Environments*, *78*, 169–178, 2012.

Sala, O. E., L. A. Gherardi, L. Reichmann, E. Jobbagy, and D. Peters, Legacies of precipitation fluctuations on primary production: theory and data synthesis, *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 3135–3144, 2012.

Savitzky, A., and M. J. Golay, Smoothing and differentiation of data by simplified least squares procedures., *Analytical chemistry*, *36*, 1627–1639, 1964.

Savtchenko, A. K., G. Huffman, and B. Vollmer, Assessment of precipitation anomalies in california using trmm and merra data, *Journal of Geophysical Research: Atmospheres*, *120*, 8206–8215, 2015.

Scanlon, B. R., et al., Global models underestimate large decadal declining and rising water storage trends relative to grace satellite data, *Proceedings of the National Academy of Sciences*, *115*, E1080–E1089, 2018.

Schenk, H., R. Jackson, F. HALL, G. COLLATZ, B. MEESON, S. LOS, E. BROWN DE COLSTOUN, and D. LANDIS, Islscp ii ecosystem rooting depths, *ORNL DAAC*, 2009.

Schnur, M. T., H. Xie, and X. Wang, Estimating root zone soil moisture at distant sites using modis ndvi and evi in a semi-arid region of southwestern usa, *Ecological Informatics*, *5*, 400–409, 2010.

Schwalm, C. R., et al., Reduction in carbon uptake during turn of the century drought in western north america, *Nature Geoscience*, *5*, 551, 2012.

Seddon, A. W., M. Macias-Fauria, P. R. Long, D. Benz, and K. J. Willis, Sensitivity of global terrestrial ecosystems to climate variability, *Nature*, *531*, 229, 2016.

Seneviratne, S. I., T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, I. Lehner, B. Orlowsky, and A. J. Teuling, Investigating soil moisture–climate interactions in a changing climate: A review, *Earth-Science Reviews*, *99*, 125–161, 2010.

Sivapalan, M., H. H. Savenije, and G. Blöschl, Socio-hydrology: A new science of people and water, *Hydrological Processes*, *26*, 1270–1276, 2012.

Smith, B., D. Wårlind, A. Arneth, T. Hickler, P. Leadley, J. Siltberg, and S. Zaehle, Implications of incorporating n cycling and n limitations on primary production in an individual-based dynamic vegetation model, *Biogeosciences*, *11*, 2027–2054, 2014.

Sperry, J. S., and D. M. Love, What plant hydraulics can tell us about responses to climate-change droughts, *New Phytologist*, *207*, 14–27, 2015.

Stocker, B. D., J. Zscheischler, T. F. Keenan, I. C. Prentice, J. Peñuelas, and S. I. Seneviratne, Quantifying soil moisture impacts on light use efficiency across biomes, *New Phytologist*, *218*, 1430–1449, 2018.

Sulman, B. N., D. T. Roman, K. Yi, L. Wang, R. P. Phillips, and K. A. Novick, High atmospheric demand for water can limit forest carbon uptake and transpiration as severely as dry soil, *Geophysical Research Letters*, *43*, 9686–9695, 2016.

Sun, Y., et al., Oco-2 advances photosynthesis observation from space via solar-induced chlorophyll fluorescence, *Science*, *358*, eaam5747, 2017.

Svoboda, M., et al., The drought monitor, *Bulletin of the American Meteorological Society*, *83*, 1181–1190, 2002.

Swenson, S., and D. Lawrence, A grace-based assessment of interannual groundwater dynamics in the c ommunity l and m odel, *Water Resources Research*, *51*, 8817–8833, 2015.

Swenson, S., and J. Wahr, Post-processing removal of correlated errors in grace data, *Geophysical Research Letters*, *33*, 2006.

Swenson, S., D. Chambers, and J. Wahr, Estimating geocenter variations from a combination of grace and ocean model output, *Journal of Geophysical Research: Solid Earth*, *113*, 2008.

Tallaksen, L. M., H. Hisdal, and H. A. Van Lanen, Space–time modelling of catchment scale drought characteristics, *Journal of Hydrology*, *375*, 363–372, 2009.

Tapley, B. D., S. Bettadpur, J. C. Ries, P. F. Thompson, and M. M. Watkins, Grace measurements of mass variability in the earth system, *Science*, *305*, 503–505, 2004a.

Tapley, B. D., S. Bettadpur, J. C. Ries, P. F. Thompson, and M. M. Watkins, Grace measurements of mass variability in the earth system, *Science*, *305*, 503–505, 2004b.

Tapley, B. D., et al., Contributions of grace to understanding climate change, *Nature climate change*, *9*, 358–369, 2019.

Teuling, A. J., et al., Evapotranspiration amplifies european summer drought, *Geophysical Research Letters*, *40*, 2071–2075, 2013.

Thomas, A. C., J. T. Reager, J. S. Famiglietti, and M. Rodell, A grace-based water storage deficit approach for hydrological drought characterization, *Geophysical Research Letters*, *41*, 1537–1545, 2014.

Tian, H., C. Cao, W. Chen, S. Bao, B. Yang, and R. B. Myneni, Response of vegetation activity dynamic to climatic change and ecological restoration programs in inner mongolia from 2000 to 2012, *Ecological Engineering*, *82*, 276–289, 2015.

Tian, H., et al., The global n2o model intercomparison project, *Bulletin of the American Meteorological Society*, *99*, 1231–1251, 2018.

Tiwari, V., J. Wahr, and S. Swenson, Dwindling groundwater resources in northern india, from satellite gravity observations, *Geophysical Research Letters*, *36*, 2009.

Tong, X., et al., Increased vegetation growth and carbon stock in china karst via ecological engineering, *Nature sustainability*, *1*, 44–50, 2018.

Tong, X., et al., Forest management in southern china generates short term extensive carbon sequestration, *Nature communications*, *11*, 1–10, 2020.

Tramontana, G., et al., Predicting carbon dioxide and energy fluxes across global fluxnet sites with regression algorithms, 2016.

Trenberth, K. E., A. Dai, G. Van Der Schrier, P. D. Jones, J. Barichivich, K. R. Briffa, and J. Sheffield, Global warming and changes in drought, *Nature Climate Change*, *4*, 17, 2014.

van der Schrier, G., J. Barichivich, K. Briffa, and P. Jones, A scpdsi-based global data set of dry and wet spells for 1901–2009, *Journal of Geophysical Research: Atmospheres*, *118*, 4025–4048, 2013.

van Dijk, A. I., H. E. Beck, R. S. Crosbie, R. A. de Jeu, Y. Y. Liu, G. M. Podger, B. Timbal, and N. R. Viney, The millennium drought in southeast australia (2001–2009): Natural and human causes and implications for water resources, ecosystems, economy, and society, *Water Resources Research*, *49*, 1040–1057, 2013.

Van Loon, A., M. Van Huijgevoort, and H. Van Lanen, Evaluation of drought propagation in an ensemble mean of large-scale hydrological models, *Hydrology and Earth System Sciences*, *16*, 4057–4078, 2012.

Van Loon, A., E. Tijdeman, N. Wanders, H. J. Van Lanen, A. Teuling, and R. Uijlenhoet, How climate seasonality modifies drought duration and deficit, *Journal of Geophysical Research: Atmospheres*, *119*, 4640–4656, 2014.

Van Loon, A. F., Hydrological drought explained, *Wiley Interdisciplinary Reviews: Water*, *2*, 359–392, 2015a.

Van Loon, A. F., Hydrological drought explained, *Wiley Interdisciplinary Reviews: Water*, *2*, 359–392, 2015b.

Van Loon, A. F., et al., Drought in the anthropocene, *Nature Geoscience*, *9*, 89, 2016.

Velicogna, I., and J. Wahr, Measurements of time-variable gravity show mass loss in antarctica, *science*, *311*, 1754–1756, 2006.

Velicogna, I., and J. Wahr, Time-variable gravity observations of ice sheet mass balance: Precision and limitations of the grace satellite data, *Geophysical Research Letters*, *40*, 3055–3063, 2013.

Velicogna, I., J. Tong, T. Zhang, and J. S. Kimball, Increasing subsurface water storage in discontinuous permafrost areas of the lena river basin, eurasia, detected from grace, *Geophysical research letters*, *39*, 2012.

Vicente-Serrano, S. M., S. Beguería, and J. I. López-Moreno, A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index, *Journal of climate*, *23*, 1696–1718, 2010a.

Vicente-Serrano, S. M., S. Beguería, J. I. López-Moreno, M. Angulo, and A. El Kenawy, A new global 0.5 gridded dataset (1901–2006) of a multiscalar drought index: comparison with current drought index datasets based on the palmer drought severity index, *Journal of Hydrometeorology*, *11*, 1033–1043, 2010b.

Vicente-Serrano, S. M., et al., Response of vegetation to drought time-scales across global land biomes, *Proceedings of the National Academy of Sciences*, *110*, 52–57, 2013.

Vicente-Serrano, S. M., et al., Evidence of increasing drought severity caused by temperature rise in southern europe, *Environmental Research Letters*, *9*, 044,001, 2014.

Villarini, G., J. A. Smith, and G. A. Vecchi, Changing frequency of heavy rainfall over the central united states, *Journal of Climate*, *26*, 351–357, 2013.

Wahr, J., M. Molenaar, and F. Bryan, Time variability of the earth's gravity field: Hydrological and oceanic effects and their possible detection using grace, *Journal of Geophysical Research: Solid Earth*, *103*, 30,205–30,229, 1998a.

Wahr, J., M. Molenaar, and F. Bryan, Time variability of the earth's gravity field: Hydrological and oceanic effects and their possible detection using grace, *Journal of Geophysical Research: Solid Earth*, *103*, 30,205–30,229, 1998b.

Wahr, J., S. Swenson, and I. Velicogna, Accuracy of grace mass estimates, *Geophysical Research Letters*, *33*, 2006.

Wang, H., S. Schubert, R. Koster, Y.-G. Ham, and M. Suarez, On the role of sst forcing in the 2011 and 2012 extreme us heat and drought: A study in contrasts, *Journal of Hydrometeorology*, *15*, 1255–1273, 2014.

Wang, H., J. C. Rogers, and D. K. Munroe, Commonly used drought indices as indicators of soil moisture in china, *Journal of Hydrometeorology*, *16*, 1397–1408, 2015.

Wang, J., et al., Recent global decline in endorheic basin water storages, *Nature geoscience*, *11*, 926–932, 2018.

Wang, X., H. Xie, H. Guan, and X. Zhou, Different responses of modis-derived ndvi to root-zone soil moisture in semi-arid and humid regions, *Journal of hydrology*, *340*, 12–24, 2007.

Watkins, M. M., D. N. Wiese, D.-N. Yuan, C. Boening, and F. W. Landerer, Improved methods for observing earth's time variable mass distribution with grace using spherical cap mascons, *Journal of Geophysical Research: Solid Earth*, *120*, 2648–2671, 2015.

Webster, P., V. E. Toma, and H.-M. Kim, Were the 2010 pakistan floods predictable?, *Geophysical research letters*, *38*, 2011.

Wiese, D. N., F. W. Landerer, and M. M. Watkins, Quantifying and reducing leakage errors in the jpl rl05m grace mascon solution, *Water Resources Research*, *52*, 7490–7502, 2016.

Williams, A. P., et al., Temperature as a potent driver of regional forest drought stress and tree mortality, *Nature Climate Change*, *3*, 292, 2013.

Wu, D., X. Zhao, S. Liang, T. Zhou, K. Huang, B. Tang, and W. Zhao, Time-lag effects of global vegetation responses to climate change, *Global change biology*, *21*, 3520–3531, 2015.

Xia, J., et al., Joint control of terrestrial gross primary productivity by plant phenology and physiology, *Proceedings of the National Academy of Sciences*, *112*, 2788–2793, 2015.

Xie, X., C. Xu, Y. Wen, and W. Li, Monitoring groundwater storage changes in the loess plateau using grace satellite gravity data, hydrological models and coal mining data, *Remote Sensing*, *10*, 605, 2018.

Xu, D., X. Kang, D. Zhuang, and J. Pan, Multi-scale quantitative assessment of the relative roles of climate change and human activities in desertification–a case study of the ordos plateau, china, *Journal of Arid Environments*, *74*, 498–507, 2010.

Xu, L., et al., Temperature and vegetation seasonality diminishment over northern lands, *Nature Climate Change*, *3*, 581–586, 2013.

Xu, X., D. Medvigy, J. S. Powers, J. M. Becknell, and K. Guan, Diversity in plant hydraulic traits explains seasonal and inter-annual variations of vegetation dynamics in seasonally dry tropical forests, *New Phytologist*, *212*, 80–95, 2016.

Xu, Z., J. A. Mason, and H. Lu, Vegetated dune morphodynamics during recent stabilization of the mu us dune field, north-central china, *Geomorphology*, *228*, 486–503, 2015.

Xu, Z., R. Hu, K. Wang, J. A. Mason, S.-Y. Wu, and H. Lu, Recent greening (1981–2013) in the mu us dune field, north-central china, and its potential causes, *Land degradation & development*, *29*, 1509–1520, 2018.

Yan, F., B. Wu, and Y. Wang, Estimating spatiotemporal patterns of aboveground biomass using landsat tm and modis images in the mu us sandy land, china, *Agricultural and forest meteorology*, *200*, 119–128, 2015.

Yang, Y., D. Long, H. Guan, B. R. Scanlon, C. T. Simmons, L. Jiang, and X. Xu, Grace satellite observed hydrological controls on interannual and seasonal variability in surface greenness over mainland australia, *Journal of Geophysical Research: Biogeosciences*, *119*, 2245–2260, 2014.

Yi, C., et al., Climate control of terrestrial carbon exchange across biomes and continents, *Environmental Research Letters*, *5*, 034,007, 2010.

Yirdaw, S. Z., K. R. Snelgrove, and C. O. Agboma, Grace satellite observations of terrestrial moisture changes for drought characterization in the canadian prairie, *Journal of Hydrology*, *356*, 84–92, 2008.

Yoshida, Y., et al., The 2010 russian drought impact on satellite measurements of solar-induced chlorophyll fluorescence: Insights from modeling and comparisons with parameters derived from satellite reflectances, *Remote Sensing of Environment*, *166*, 163–177, 2015.

Yu, X., Y. Huang, E. Li, X. Li, and W. Guo, Effects of rainfall and vegetation to soil water input and output processes in the mu us sandy land, northwest china, *Catena*, *161*, 96–103, 2018.

Zeng, N., J.-H. Yoon, J. A. Marengo, A. Subramaniam, C. A. Nobre, A. Mariotti, and J. D. Neelin, Causes and impacts of the 2005 amazon drought, *Environmental Research Letters*, *3*, 014,002, 2008.

Zhang, Z., and D. Huisingh, Combating desertification in china: Monitoring, control, management and revegetation, *Journal of Cleaner Production*, *182*, 765–775, 2018.

Zhao, M., I. Velicogna, and J. S. Kimball, A global gridded dataset of grace drought severity index for 2002–14: Comparison with pdsi and spei and a case study of the australia millennium drought, *Journal of Hydrometeorology*, *18*, 2117–2129, 2017a.

Zhao, M., I. Velicogna, and J. S. Kimball, Satellite observations of regional drought severity in the continental united states using grace-based terrestrial water storage changes, *Journal of Climate*, *30*, 6297–6308, 2017b.

Zhou, S., Y. Zhang, K. K. Caylor, Y. Luo, X. Xiao, P. Ciais, Y. Huang, and G. Wang, Explaining inter-annual variability of gross primary productivity from plant phenology and physiology, *Agricultural and forest meteorology*, *226*, 246–256, 2016.

Zhou, Y., C. Shi, J. Du, and X. Fan, Characteristics and causes of changes in annual runoff of the wuding river in 1956–2009, *Environmental earth sciences*, *69*, 225–234, 2013.