Uncertainties in GRACE-Groundwater Assessments and Their Implications for Groundwater Drought

Mohamed Akl Abdallah Akl

A Thesis submitted to Newcastle University in partial fulfilment of the requirements for the degree of Doctor of Philosophy (PhD) in the School of Engineering – Faculty of Science, Agriculture and Engineering

School of Engineering
Newcastle University
Newcastle upon Tyne
NE1 7RU

May 2025

Abstract

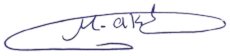
Groundwater is an essential freshwater resource that underpins ecological resilience, agricultural productivity, and long-term water security. Despite its importance, large-scale assessment remains constrained by the limited spatial and temporal coverage of conventional monitoring networks. The Gravity Recovery and Climate Experiment (GRACE) and its follow-on mission (GRACE-FO) have transformed groundwater monitoring by providing global estimates of terrestrial water storage anomalies (GRACE-TWSA), from which groundwater storage anomalies (GRACE-GWA) can be derived. However, isolating GRACE-GWA remains methodologically challenging due to uncertainties in auxiliary water storage components, such as soil moisture, snow, and surface water, and limited validation against in-situ groundwater observations. This thesis addresses these challenges by refining the derivation and evaluation of GRACE-GWA and examining the implications of associated uncertainties for groundwater drought characterisation. A systematic integration of diverse auxiliary datasets into the water balance framework reveals that discrepancies among water budget components introduce significant bias and variability in GRACE-GWA estimates. Analysis across three hydrologically complex basins demonstrates strong sensitivity to input selection, with in-situ correlation coefficients ranging from 0.32 to 0.89. To improve evaluation fidelity, a multi-objective comparison framework, combining Nash–Sutcliffe Efficiency (NSE) and Kling–Gupta Efficiency (KGE), is implemented, offering a more holistic assessment of time series agreement than traditional correlation-based methods. This approach reveals substantial inter-model divergence in GRACE-GWA trends and amplitudes, reinforcing the value of ensemble-based strategies. Building on these insights, the GRACE-Groundwater Drought Index (GGDI) is revisited using multi-model GRACE-GWA realisations to investigate how uncertainty influences drought indicators across 37 major aquifers. Results reveal that GRACE-GWA uncertainty compromise drought reliability and that aquifers with high memory, quantified via GGDI autocorrelation, experience fewer but longer and more severe droughts. Overall, this research delivers a scalable, uncertainty-aware framework to advance the accuracy and applicability of GRACE-based groundwater monitoring under intensifying climatic and anthropogenic pressures.

Declaration

I hereby confirm that this thesis is my original work and has not been submitted, in whole or in part, for consideration for any other degree or qualification at this or any other university.

Mohamed Akl

May 2025



Acknowledgements

First and foremost, I extend my deepest and most heartfelt thanks to my primary supervisor, **Dr Brian F. Thomas** (University College London). Your unwavering support, expert guidance, and steadfast encouragement have shaped every step of this journey. Your belief in my work, and in me, has not only made this thesis possible but has also profoundly influenced my growth as a researcher and scholar. I am truly privileged to have worked under your mentorship.

I am immensely grateful to my supervisor, **Prof. Peter Clarke** (Newcastle University), whose expertise, thoughtful feedback, and constant support have been invaluable throughout this journey. Your clarity of thought and dedication to academic excellence have greatly enriched this research, and I am sincerely thankful for the opportunity to learn from you.

Special thanks are due to **Prof. Jon Mills** (Newcastle University), played a pivotal role in making this PhD a reality by supporting my application to Newcastle University. I also wish to express my heartfelt appreciation to **Prof. Hafez Afify** of Tanta University, whose guidance during my master's studies helped me take the first steps toward this academic path.

I am especially thankful to **Dr Achraf Koulali** (Newcastle University), not only for our many enriching discussions, but also for his unwavering encouragement and the shared moments over coffee that brought balance to this journey. I also extend my appreciation to **Dr Ahmed Elsherif** (Tanta University) for his kind support and guidance during the PhD application process.

I am deeply grateful to **the Egyptian Ministry of Higher Education and Scientific Research** for funding my PhD research and to **the Egyptian Cultural and Educational Bureau in London** for their steadfast support throughout this endeavour.

On a personal level, I owe a profound debt of gratitude to my family. To **my parents**, your unwavering love and prayers have been my strength and guiding light. To my beloved wife, **Omnia**, your patience, resilience, and unfailing belief in me have been the cornerstone of this achievement. Your sacrifices, love, and quiet strength have carried me through even the most difficult moments, and this milestone would not have been possible without you.

Finally, to my dear children, **Yahia** and **Amr**, thank you for bringing joy, laughter, and perspective into my life. Your smiles have been a daily reminder of what truly matters and have given me the inspiration to persevere through every challenge.

Research Outputs

The research undertaken as part of this PhD has led to a number of scholarly contributions, including peer-reviewed journal articles, international conference presentations, and an invited talk.

Peer-Reviewed Journal Articles:

- **Akl, M.**, & Thomas, B. F. (2024). Challenges in applying water budget framework for estimating groundwater storage changes from GRACE observations. *Journal of Hydrology*, *639*, 131600. https://doi.org/10.1016/j.jhydrol.2024.131600
- **Akl M.**, Thomas B. F., & Clarke P. J. (under review). A Multi-objective Comparative Framework for Enhanced GRACE-Groundwater Comparative Analysis. Journal of Hydrology, HYDROL69104. Submitted: April 28, 2025.
- **Akl M.**, Thomas B. F., & Clarke P. J. (under review). Global groundwater drought assessment revisited: a holistic re-evaluation of the GRACE-Groundwater Drought Index across major aquifers. Water Resources Res., 2025WR040389. Submitted: March 3, 2025.

Conference Contributions:

- **Akl, M.**, Thomas, B., & Mills, J. (2022, May). GRACE-derived groundwater storage estimation: Lake/Reservoir storage controls across Canada. In *EGU General Assembly Conference Abstracts* (pp. EGU22-2495). https://doi.org/10.5194/egusphere-egu22-2495
- **Akl, M.**, & Thomas, B. (2023). Evaluating the Use of "Goodness-of-Fit" Metrics in GRACE Validation: GRACE Accuracy for Monitoring Groundwater Dynamics. *AGU Fall Meeting 2022, Chicago, IL*. https://doi.org/10.22541/essoar.167388164.42151735/v1
- **Akl, M.**, Thomas, B., & Clarke, P. (2023, May). Evaluation of GRACE-derived Groundwater Signal Accuracy using Developed Statistical Framework. In *EGU General Assembly Conference Abstracts* (pp. EGU-3357). https://doi.org/10.5194/egusphere-egu23-3357
- **Akl, M.**, Thomas, B., & Clarke, P. (2023). Re-evaluation of GRACE Groundwater Drought Index (GGDI): Precise Accounting of Surface Water Changes. *General Assembly of the International Union of Geodesy and Geophysics (IUGG) (Berlin 2023)*. https://doi.org/10.13140/RG.2.2.33648.64002

Invited Presentation:

- **Akl, M.**, Thomas, B., & Clarke, P. (2024, September 19). Challenges in Large-Scale Groundwater from GRACE. *The 50th Anniversary Celebration Event of The Hydrogeological Group, Burlington House, London*. https://www.hydrogroup.org.uk/50th-anniversary-celebration/

Contents

List of Figures

List of Tables

List of Abbreviations

C_i	Climatology
CLM	Community Land Model
CLSM	Catchment Land Surface Model
CMIP5	Coupled Model Intercomparison Project Phase 5
CSR	Centre for Space Research
FLDAS	Famine Early Warning System Network (FEWS NET) Land Data Assimilation System
GFZ	GeoForschungsZentrum
GGDI	GRACE Groundwater Drought Index
GIN	Groundwater Information Network
GLDAS	Global Land Data Assimilation System
GLOBathy	GLObal lakes Bathymetry datasets
GOF	Goodness-of-Fit
GRACE	Gravity Recovery and Climate Experiment
GRACE-FO	Gravity Recovery and Climate Experiment Follow-on
GRACE-GWA	GRACE Groundwater Storage Anomalies
GRACE-TWSA	GRACE Terrestrial Water Storage Anomalies
G-REALM	Global REservoirs and Lakes Monitor
GSD	Groundwater Storage Deviation
HGSWE	Historical Gridded Snow Water Equivalent and snow water fraction over Canada
HydroSHEDS	Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales
IDW	Inverse Distance Weighting
JPL	Jet Propulsion Laboratory
KGE	Kling-Gupta Efficiency
LDAS	Land Data Assimilation System
LOOCV	Leave-One-Out Cross-Validation
LSM	Land Surface Model
Mascon, M	Mass Concentration
MLR	Multi Linear Regression
MSE	Mean Square Error
NGWMN	National Ground-Water Monitoring Network

NSE	Nash-Sutcliffe Efficiency
OK	Ordinary Kriging
OK-Log	Ordinary Kriging-Log transform
PCR-GLOBWB	PCRaster GLOBal Water Balance model
RMSE	Root Mean Square Error
SH	Spherical Harmonics
SMA	Soil Moisture Anomalies
SMK	Seasonal Mann-Kendall
STK	Spatio-Temporal Kriging
SWA	Surface Water Anomalies
SWEA	Snow Water Equivalent Anomalies
TP	Thiessen Polygon
UK	Universal Kriging
VIC	Variable Infiltration Capacity
WGHM	WaterGAP Global Hydrological Model
WHYMAP	Worldwide Hydrogeological Mapping and Assessment Program

Chapter 1. Introduction

1.1 Research Context

Water security ranks among the most pressing global sustainability challenges of the 21st century (Gain et al., 2018; Vanham et al., 2018). Nearly 80% of the global population is confronted by substantial water scarcity risk, placing unprecedented pressure on freshwater resources already burdened by excessive extraction, contamination, and escalating climate change (Vörösmarty et al., 2010; Mekonnen et al., 2016; Mishra et al., 2021). As surface water supplies become increasingly uncertain and climate-vulnerable (Vliet et al., 2021; Jones et al., 2024), groundwater has emerged as critical hydrological buffer, a hidden but indispensable resource that sustains human livelihoods and ecological resilience, particularly during extended drought periods (Wada et al., 2010; Döll et al., 2014; Scanlon et al., 2023; Kuang et al., 2024). Globally, groundwater supplies drinking water for more than 2.5 billion people (Morris et al., 2003) and supports over 40% of irrigated agriculture (Alley et al., 2002; Siebert et al., 2010; Döll et al., 2012), cementing its vital role in global public health and food security. Its strategic importance is projected to increase dramatically amid mounting anthropogenic demands and intensifying climate-related stressors (Döll, 2009; Kundzewicz & Döll, 2009; United Nations, 2022; Loaiciga & Doh, 2024).

However, this growing reliance on groundwater occurs against a troubling backdrop of unsustainable exploitation. In many regions worldwide, groundwater extraction rates significantly exceed natural recharge capacities, pushing aquifers toward severe and potentially irreversible depletion (Wada et al., 2010; Gleeson et al., 2010; Famiglietti, 2014; Bierkens et al., 2019; Jasechko & Perrone, 2021; Loaiciga & Doh, 2024). The consequences of groundwater exploitation are far-reaching and multidimensional: declining groundwater levels disrupt ecosystem function (Sophocleous, 2002; Rohde et al., 2024), degrade water quality (Fogg & LaBolle, 2006), reduce streamflow and wetland extent (Döll et al., 2012; Li et al., 2021), and jeopardize the stability of deltaic regions through land subsidence and saltwater intrusion (Shirzaei et al., 2021; Werner et al., 2013). Furthermore, competition for dwindling groundwater supplies is increasingly recognised as a catalyst for social conflict and geopolitical tension (Llamas et al., 2005; Döring, 2020). In this context, robust groundwater monitoring is not merely a scientific interest but a strategic imperative, crucial for detecting emerging crises, guiding adaptive water governance, and ensuring the long-term resilience of groundwater-dependent communities and ecosystems (Famiglietti, 2014).

Despite underpinning vital ecological and societal functions, groundwater remains one of the least monitored and most poorly characterised components of the global hydrological cycle (Famiglietti, 2014; Condon et al., 2021). Traditional observation networks are often sparse, inconsistent, or entirely absent, particularly across large, remote, or transboundary aquifer systems (Rodell et al., 2007; Chen et al., 2016a; Jasechko & Perrone, 2021; United Nations, 2022; Jasechko et al., 2024), hampering efforts to track groundwater dynamics (Kuang et al., 2024). In response, satellite gravimetry has emerged as a transformative tool for monitoring global water storage change (Syed et al., 2008). The Gravity Recovery and Climate Experiment (GRACE; Tapley et al., 2004), operational from 2002 to 2017, and its successor, GRACE Follow-On (GRACE-FO; Landerer et al., 2020), launched in 2018, have enabled unprecedented monitoring of large-scale groundwater variations through the estimation of groundwater storage anomalies (GRACE-GWA; Rodell & Famiglietti, 1999). These twin-satellite missions detect temporal variations in Earth's gravity field by precisely measuring changes in the distance between the satellites along a ~220 km orbital baseline (Tapley et al., 2004). Each observed change in gravity corresponds to mass redistribution, with one gigaton of water roughly equivalent to 1 km^3 (Tapley et al., 2004). The resulting terrestrial water storage anomalies (GRACE-TWSA) reflect monthly changes in equivalent water thickness relative to a climatological baseline (2004–2009), after removal of non-hydrological signals such as tides, atmospheric pressure, oceanic circulation, and solid Earth deformation (Bettadpur, 2007; Syed et al., 2008). However, the intricate post-processing necessary for deriving GRACE-TWSA has prompted data centres to implement diverse methodologies, culminating in the release of multiple products that often display substantial inter-product discrepancies (Jing et al., 2019).

To derive meaningful gravity field estimates from GRACE, two principal processing methodologies have been developed: the spherical harmonics (SH) approach and the mass concentration (mascon) technique. The SH approach expresses Earth's gravity field using a global basis set of surface spherical harmonic functions (Wahr et al., 1998; Bettadpur, 2007). While computationally efficient, SH solutions are inherently global and lack the spatial localization necessary to distinguish between land and ocean signals. As a result, high-amplitude terrestrial signals often leak into adjacent low-signal ocean regions, compromising spatial accuracy. Additionally, unconstrained SH solutions are poorly equipped to resolve east–west gravity gradients, leading to the characteristic "striping" artifacts oriented in the north–south direction. To suppress these artifacts and reduce noise, SH-based solutions undergo extensive postprocessing, including truncation to lower spherical degrees (typically degree and order 60), application of destriping filters (Swenson & Wahr, 2006), and spatial smoothing—

often using a Gaussian filter with a 300 km radius. However, while these techniques improve signal clarity, they also distort or dampen the underlying geophysical signal, especially at regional or basin scales.

To compensate for the attenuation of hydrological signals introduced during GRACE-SH processing, scaling factors are commonly applied. These factors are empirically derived by subjecting land surface model (LSM)-simulated water storage outputs to the same filtering, truncation, and destriping applied to GRACE data, thereby enabling the calculation of correction coefficients that reconcile filtered and original values (Landerer & Swenson, 2012; Long et al., 2015). However, this approach is not without caveats (Landerer & Swenson, 2012). Most LSMs simplify complex hydrological processes and often exclude critical dynamics such as variable surface water storage, groundwater abstraction, and regulated reservoir operations (Zeng, 1999; Swenson & Milly, 2006; Han et al., 2010; Werth & Güntner, 2010). Moreover, the use of a single scaling factor implicitly assumes uniform spectral behaviour across all water storage components, an assumption that fails in reality, as storage components exhibit distinct spatial and temporal signatures (Landerer & Swenson, 2012). These simplifications and assumptions undermine the reliability of derived scaling factors, particularly in hydrologically heterogeneous or surface-water-dominated regions (Kull, 2006; Awange et al., 2007; Long et al., 2015), ultimately introducing biases that can propagate into GRACE-GWA estimates. As a more physically consistent alternative, some researchers have proposed applying GRACE-like processing steps directly to individual water budget components prior to their use in the water balance equation (Lenk, 2013; Fatolazadeh et al., 2016; Nanteza et al., 2016). This harmonized processing strategy enhances spatial and spectral consistency across components and reduces the need for externally derived correction factors (Nanteza et al., 2016). In a notable application across East Africa, Nanteza et al. (2016) demonstrated that implementing GRACE-like processing at the component level led to a ~75% improvement in the correlation between GRACE-GWA and in-situ groundwater observations, alongside a 14% reduction in root-mean-square error (RMSE), compared to conventional scaling approaches. These findings underscore the diagnostic value of harmonized processing, particularly in hydrologically complex regions where residual GRACE signals may be confounded by contributions from surface water or other non-groundwater sources (Nanteza et al., 2016).

In contrast to SH analysis, the mascon approach offers a more spatially localized and physically intuitive alternative, with growing adoption in GRACE data processing (Watkins et al., 2015; Save et al., 2016). Rather than relying on global basis functions, mascon solutions estimate mass change over predefined geographic blocks, or “mass concentrations”, allowing for

separation of land and ocean areas and inherently minimizing signal leakage. This approach also permits the application of regularization constraints and incorporation of *a priori* information during the inversion process, which suppresses correlated noise (e.g., striping) without the need for empirical filtering or externally derived scaling factors. As a result, mascon-based solutions tend to preserve more of the true geophysical signal, reduce amplitude attenuation, and offer improved spatial resolution, especially in high-latitude regions with denser satellite track coverage (Scanlon et al., 2106).

Three leading processing centres, the Centre for Space Research (CSR) at the University of Texas, GeoForschungsZentrum (GFZ) in Potsdam, and the Jet Propulsion Laboratory (JPL), routinely generate and distribute monthly GRACE-TWSA solutions (Wouters & Sasgen, 2024). While all three centres begin with the same fundamental satellite observations, their processing pipelines diverge in terms of parameter choices, basis functions, filtering methods, and regularization techniques, resulting in differences in the derived gravity field solutions, whether SH- or mascon-based (https://podaac.jpl.nasa.gov/gravity/grace-documentation; Sakumura et al., 2014). These differences are particularly evident in mascon-based solutions, where CSR employs regularization constraints based exclusively on GRACE data (Save et al., 2016), while JPL integrates additional constraints from external geophysical models (Watkins et al., 2015). Spatial discretization also varies: CSR-M uses a geodesic equal-area tiling scheme comprising approximately 41,000 hexagonal cells (~120 km wide, ~12,400 km² each), intentionally oversampling the gravity field at the Equator to enhance resolution at higher latitudes where satellite ground track density is greater (Save et al., 2016). Conversely, JPL-M adopts a 3˚ × 3˚ grid (~330 km at the Equator), resulting in a much coarser product of about 4,551 grid cells globally (Watkins et al., 2015). While this coarser approach aligns with the native equatorial resolution of GRACE, it reduces spatial granularity in regions where finer-scale hydrological features may be present (Scanlon et al., 2106).

Discrepancies among GRACE-TWSA solutions, driven by conceptual differences and methodological variations across processing centres, underscore the importance of rigorous validation (Jing et al., 2019). Yet, the integrated nature of GRACE-TWSA complicates validation against independent in-situ observations (Scanlon et al., 2016), as the observations capture vertically aggregated change across multiple storage compartments, including groundwater storage (GRACE-GWA), snow water equivalent anomalies (SWEA), soil moisture anomalies (SMA), and surface water anomalies (SWA). Although this integrative capacity enables powerful insights into large-scale hydrological variability, it simultaneously obscures the behaviour of individual storage components. A widely adopted approach for

isolating GRACE-GWA involves the application of a water budget equation (Equation 1.1; Figure 1.1), assuming knowledge of the other water budget components (i.e., SWEA, SMA, and SWA; Rodell & Famiglietti, 1999). However, the reliability of the water budget method hinges on the accuracy of both GRACE-TWSA and water budget components (Rodell & Famiglietti, 1999; Swenson & Lawrence, 2015), a challenging task due to methodological discrepancies across GRACE processing and the lack of direct spatio-temporal observations of water budget components.

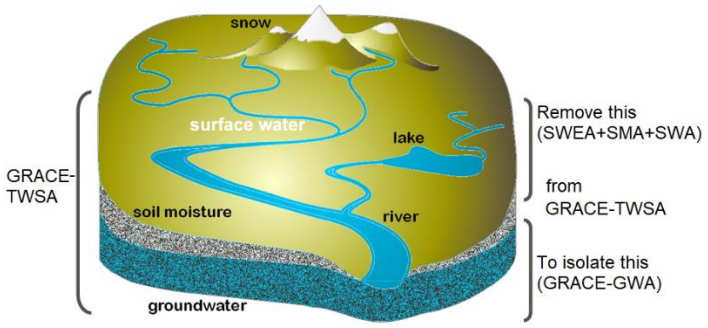


Figure 1.1 Schematic representation of terrestrial water storage anomalies (GRACE-TWSA), which integrates observations of groundwater (GRACE-GWA), snow (SWEA), soil moisture (SMA), and surface water (SWA) (rivers, lakes, reservoirs). To isolate GRACE-GWA, components of SWEA, SMA, and SWA need to be removed (Rodell and Famiglietti, 2002; source: Fu et al., 2024).

$$\text{GRACE-TWSA}_t = \text{GRACE-GWA}_t + \text{SMA}_t + \text{SWEA}_t + \text{SWA}_t \qquad (Eq.\,1.1)$$

Spatiotemporally continuous estimates of the water budget components are often generated through auxiliary datasets, including LSMs, hydrological models, reanalysis products, and remote sensing observations, particularly in regions where direct observations are sparse or unavailable (Rodell et al., 2007; Rodell et al., 2009; Xavier et al., 2010; Frappart et al., 2011; Shamsudduha et al., 2012; Feng et al., 2013; Chen et al., 2014; Richey et al., 2015b; Thomas & Famiglietti, 2019; Shamsudduha & Taylor, 2020; Li & Wang, 2022). While these datasets are indispensable for large-scale hydrological assessments, they vary considerably in terms of underlying algorithms, physical assumptions, and parameterisations used to simulate and partition terrestrial water fluxes and storage, introducing substantial variability across estimates of individual storage components (Beven, 2002; Elsner et al., 2014; Mizukami et al., 2014; Bierkens, 2015; Clark et al., 2015; Mendoza et al., 2015; Mizukami et al., 2017; Saxe et al., 2021). This inter-model variability, referred to here as model structural bias, reflects conceptual divergences in how water is represented within different modelling frameworks, broadening the term beyond its typical usage in prior studies (Clark et al., 2008; Ruddell et al., 2019; Konapala et al., 2020). Disagreement among models in simulating water fluxes and stores can be substantial, sometimes exceeding the magnitude of the

measurements themselves (Saxe et al., 2021). This variability complicates the partitioning of GRACE-TWSA, making the isolation of GRACE-GWA particularly sensitive to the choice of auxiliary datasets. As a result, differences in water budget inputs can propagate through residual calculations, introducing systematic biases into GRACE-GWA estimates and ultimately affecting the interpretation of groundwater dynamics (Chen et al., 2016a; Saxe et al., 2021).

To mitigate the influence of model selection on water budget components, many studies employ an ensemble approach, averaging outputs from multiple LSMs, commonly 1˚ resolution datasets available through the Global Land Data Assimilation System (GLDAS), to produce composite estimates of water storage components and corresponding GRACE-GWA uncertainty (Rodell et al., 2009; Shamsudduha et al., 2012; Voss et al., 2013; Feng et al., 2013; Castle et al., 2014; Nanteza et al., 2016; Huang et al., 2016). GRACE-GWA uncertainty is typically assessed through a structured framework, wherein uncertainties associated with each input (e.g., GRACE-TWSA, SMA, SWEA and SWA) are quantified and propagated through to the residual GRACE-GWA estimates, often using least-squares fitting or error propagation techniques (e.g., Rodell & Famiglietti, 2002; Famiglietti et al., 2011; Voss et al., 2013; Shamsudduha et al., 2020). However, ensemble averaging inherently assumes statistical independence among model outputs, an assumption that is often violated, as many LDAS models are driven by the same meteorological forcings (Rodell et al., 2004; Mitchell et al., 2004; McNally et al., 2017). This dependence can lead to underestimation of uncertainty and a false sense of robustness. Moreover, averaging may mask model-specific structural biases, limiting the ability to trace systematic deviations that propagate into groundwater storage estimates. While the sensitivity of GRACE-GWA estimates to variability in water budget components has been increasingly recognised (Ahmed et al., 2016; Scanlon et al., 2019; Müller Schmied et al., 2021; Hu et al., 2021; Ju et al., 2023), critical questions remain regarding the integration of model output and different sets of data in combination with GRACE-TWSA to systematically evaluate groundwater storage variability. A comprehensive framework, incorporating an ensemble of GRACE-TWSA solutions and water budget components offers an improved foundation for evaluating the heterogeneity of GRACE-GWA estimates. This approach yields important insights into the complex dynamics shaping groundwater storage estimates and emphasizes the influence of methodological choices on GRACE-based groundwater interpretations.

Validating GRACE-GWA against in-situ groundwater observations, where such observations are available, is essential for assessing the extent to which GRACE-based estimates capture genuine groundwater behaviour. This is particularly important given the uncertainties inherent

in applying the water balance framework (Equation 1.1; Akl & Thomas, 2024; see also Chapter 2 of this thesis). Chapter 2 in this thesis reveals that GRACE-GWA estimates are susceptible to significant bias and variability, arising from the selection of individual water budget components, underscoring the critical need for comprehensive evaluation framework. Although numerous studies have sought to validate GRACE-GWA time series (e.g., Yeh et al., 2006; Strassberg et al., 2007; Rodell et al., 2007; Shamsudduha et al., 2012; Scanlon et al., 2012; Castle et al., 2014; Nanteza et al., 2016; Feng et al., 2018; Rateb et al., 2020; Ouatiki et al., 2022; Amiri et al., 2023; Alghafli et al., 2023; Huang et al., 2023; Nenweli et al., 2024; Yang et al., 2024), the majority have predominantly relied on correlation-based metrics to evaluate the agreement between GRACE-GWA estimates and in-situ groundwater measurements. While correlation provides insight into the synchronicity of temporal patterns, it does not adequately reflect key features such as anomaly magnitude, seasonal amplitude, and phase alignment—attributes that are critical for capturing the true hydrological dynamics of groundwater systems (Legates & McCabe, 1999; Moore et al., 2006). The neglect of these crucial time series characteristics triggers bias in the comparison efficacy, provoking consequential challenges in truthful tracking of groundwater change (Alley & Konikow, 2015).

The limited adoption of advanced evaluation metrics in GRACE-GWA comparison has resulted in a methodological gap when compared to the hydrologic modelling community, where a rich body of literature supports the use of multi-objective goodness-of-fit (GOF) metrics (Nash & Sutcliffe, 1970; Gupta et al., 2009; Althoff et al., 2021; Mathevet et al., 2023). Metrics such as the Nash–Sutcliffe Efficiency (NSE; Nash & Sutcliffe, 1970) and Kling-Gupta Efficiency (KGE; Gupta et al., 2009) provide a more comprehensive and diagnostically rigorous basis for evaluating time series agreement. Unlike simple correlation-based measures, these metrics capture a broader array of hydrologic signal characteristics, including amplitude variability, seasonal dynamics, and phase shifts. Integrating multi-objective GOF metrics into GRACE-GWA comparison frameworks offers a more robust means of identifying errors introduced through GRACE-TWSA and water budget component selection. Chapter 3 in this thesis documents improved identification of GRACE-GWA realisations that accurately reflect true groundwater variability, ultimately strengthening the interpretability, reliability, and policy relevance of large-scale groundwater assessments.

GRACE-GWA forms the basis for a wide range of groundwater applications, yet uncertainties arising from GRACE-TWSA and water budget component selection risk propagating through to these subsequent analyses. A critical application is the GRACE-Groundwater Drought Index (GGDI; Thomas et al., 2017a), which provides a normalised framework for assessing the onset,

duration, and severity of large-scale groundwater drought. As highlighted by Thomas et al. (2017a), the robustness and interpretability of GGDI is inherently tied to the quality of the underlying GRACE-GWA, which is itself sensitive to differences in GRACE processing strategies and uncertainties in auxiliary water budget components (Saxe et al., 2021; Akl & Thomas, 2024; see also Chapter 2 of this thesis). Despite this sensitivity, a substantial body of subsequent GGDI studies has adopted a single-model configuration for representing non-groundwater components, most commonly the GLDAS NOAH model at 1˚ spatial resolution, without adequately addressing the structural uncertainty inherent in GRACE-GWA estimation (Wang et al., 2020; Paredes-Trejo et al., 2021; Satish Kumar et al., 2021; Ali et al., 2022; Liu et al., 2022; Wang et al., 2022; Huang et al., 2023; Aon et al., 2024; Nandi & Biswas, 2024; Neves, 2024; Nigatu et al., 2024; Nikraftar et al., 2024; Song et al., 2024; Zhang et al., 2024; Zheng et al., 2024; Liu et al., 2025). This singular approach implicitly assumes representativeness and reliability, yet it obscures model-dependent variability and narrows the epistemic space through which groundwater drought dynamics are interpreted.

Recent advances achieved through this PhD research in characterising the methodological and structural uncertainties inherent in the water balance approach to GRACE-GWA estimation have raised critical questions regarding the reliability of groundwater drought assessments based on these data. Specifically, while variability among GRACE-GWA realisations has been observed, the extent to which such uncertainties influence key drought attributes, such as frequency, duration, severity, and intensity, remains inadequately explored. Chapter 4 in this thesis revisits the GGDI framework through the integration of a multi-model GRACE-GWA, enabling a systematic assessment of the sensitivity of groundwater drought diagnostics to input uncertainty. By embedding structural variability into the GGDI framework, this approach enhances the robustness, transparency, and policy relevance of GRACE-informed drought assessments—particularly in regions where observational data are limited or absent.

1.2 Research Aim, Questions, and Objectives

1.2.1 Research aim

The overarching aim of this research is to strengthen the scientific basis and practical utility of GRACE-based groundwater assessments by addressing key sources of uncertainty and methodological variability. The aim is pursued through a structured, multi-phase investigation encompassing three interrelated analytical dimensions. First, the study constructs a comprehensive ensemble of GRACE-GWA realisations by combining multiple GRACE-TWSA datasets with diverse representations of water budget components, enabling a critical evaluation of how component selection influences GRACE-based groundwater estimates

(Chapter 2). Second, the research applies a multi-objective comparative framework utilizing commonly applied goodness of fit metrics (e.g., NSE and KGE), to augment the diagnostic precision of comparing multi-model GRACE-GWA realisations against basin-scale in-situ groundwater observations, with the aim of identifying estimates that faithfully reflect observed groundwater dynamics ***(Chapter 3)***. Third, the study integrates these multi-model GRACE-GWA realisations into the GGDI framework to investigate how model-driven variability shapes interpretations of groundwater drought dynamics and aquifer memory behaviour across diverse hydrological settings ***(Chapter 4)***.

In addition, the thesis offers a technical commentary that explores the challenges in extracting large-scale groundwater signals from GRACE, synthesizing the insights gained throughout the research into a coherent framework aimed at enhancing the transparency and reliability of GRACE-GWA monitoring practices ***(Chapter 5)***. Collectively, this research aims to advance the methodological rigour, interpretive clarity, and decision-making relevance of GRACE-based groundwater analyses, thereby contributing to more robust, data-informed groundwater resource monitoring—particularly in regions where conventional observations are sparse or absent.

1.2.2 Research questions

1. To what extent do GRACE-GWA estimates faithfully reflect observed changes in in-situ groundwater observations?
2. What inconsistencies in GRACE-GWA estimates can be attributed solely to the selection of GRACE-TWSA datasets and water budget components?
3. Are conventional correlation-based approaches sufficient for evaluating the agreement between multi-model GRACE-GWA estimates and basin-scale in-situ groundwater observations?
4. What added diagnostic value is achieved by applying a multi-objective comparative framework—employing NSE and KGE metrics—in evaluating GRACE-GWA comparison?
5. How does variability among GRACE-GWA realisations influence the interpretation of groundwater drought characteristics, including frequency, duration, severity, and intensity?
6. In what ways does inter-model variability affect the estimation of aquifer memory for perceived groundwater drought?

1.2.3 Research objectives

The main research objectives are to:

1. Examine how the selection of GRACE-TWSA solutions and water budget components within the water balance equation influences GRACE-GWA estimates—*addressing Questions 1 and 2 (Chapter 2).*
2. Assess whether incorporating SWA within the water budget equation enhances the agreement between GRACE-GWA and basin-scale in-situ groundwater anomalies— *further addressing Questions 1 and 2 (Chapter 2)*.
3. Utilize a multi-objective comparative framework, employing NSE and KGE metrics, to identify optimal GRACE-GWA realisations that accurately reflect observed groundwater dynamics—*addressing Questions 3 and 4 (Chapter 3).*
4. Evaluate variability in multi-model GRACE-GWA trends and amplitudes using the multi-objective comparative framework—*further addressing Questions 1, 2 and 4 (Chapter 3).*
5. Examine how variability in multi-model GRACE-GWA realisations within the GGDI framework impacts key groundwater drought indicators, influencing drought characterisation and assessment—*addressing Question 5 (Chapter 4).*
6. Quantify and interpret variability in aquifer memory, providing insights into its relationship with perceived groundwater drought—*addressing Question 6 (Chapter 4).*

1.3 Significance and Contribution

This research contributes to the advancement of large-scale groundwater monitoring by enhancing the methodological rigour, uncertainty awareness, and practical utility of GRACE-based groundwater assessments. GRACE has become an indispensable tool for evaluating groundwater storage dynamics globally (e.g., Chen et al., 2016a; Frappart & Ramillien, 2018), yet significant challenges persist in ensuring the reliability of its outputs. In particular, uncertainties stemming from variations in GRACE-TWSA processing methodologies and inconsistencies across water budget components datasets continue to hinder confidence in derived groundwater estimates. These uncertainties are often overlooked or masked through ensemble averaging, a practice that may inadvertently obscure meaningful hydrological signals and contribute to the misinterpretation of groundwater trends, seasonal dynamics, and drought events.

A cornerstone of this research is the application of a **multi-model framework** that systematically integrates a range of GRACE-TWSA solutions and water budget component datasets. By explicitly accounting for the variability across model realizations, the framework improves the transparency and reliability of GRACE-GWA estimates. In doing so, the applied

framework builds a compelling case for adopting ensemble-based, multi-model strategies as the methodological standard in GRACE-based groundwater assessments (Akl & Thomas, 2024).

Complementing this, the research employs a **multi-objective comparative framework** that integrates the NSE and KGE goodness-of-fit metrics to benchmark multi-model GRACE-GWA realisations against in-situ groundwater observations. This approach moves beyond traditional correlation-based validations by capturing key time series attributes—such as magnitude fidelity, seasonal amplitude, and timing—thereby identifying the most hydrologically representative GRACE-GWA realizations. In doing so, the framework significantly elevates the quality and credibility of GRACE-based comparative analyses (Akl M., Thomas B. F., & Clarke P. J. (under review). *Journal of Hydrology*, HYDROL69104. Submitted: April 28, 2025).

Extending its contribution further, the study embeds these **multi-model GRACE-GWA realisations into the GGDI framework** to critically evaluate how model-driven variability influences groundwater drought interpretation. By quantifying differences in drought indicators—such as frequency, severity, intensity, and duration—across multi-model realizations, the study reveals the sensitivity of groundwater drought assessments to input uncertainty. This insight underscores the importance of ensemble-informed drought analysis for more reliable groundwater resilience forecasting (Akl M., Thomas B. F., & Clarke P. J. (under review). *Water Resources Research*, 2025WR040389. Submitted: March 03, 2025).

To unify and operationalize the methodological advances developed across the thesis, Chapter 5 offers a **technical commentary** that critically interrogates the challenges of isolating large-scale groundwater signals from GRACE. Drawing on the collective insights of the preceding chapters, it formulates a transparent and scalable framework for GRACE-GWA monitoring—articulating best practices for model selection, SWA incorporation, and comparison evaluation. This synthesis not only reinforces the thesis's contribution to advancing methodological rigour but also provides a practical roadmap for improving the reproducibility, interpretability, and policy relevance of satellite-informed groundwater assessments. This technical commentary is intended for submission to *Groundwater*.

Together, these contributions emphasize the transformative potential of **multi-model, multi-objective frameworks** in addressing the limitations of current GRACE-based groundwater assessments. This study not only establishes a new methodological standard but also provides

a critical foundation for more informed, adaptive, and sustainable groundwater resource practices— essential for navigating the escalating complexity of global water challenges.

1.4 Thesis Structure

This thesis consists of this introductory chapter followed by five additional chapters, including three core research chapters (Chapters 2, 3, and 4), a technical commentary (Chapter 5), and a concluding chapter (Chapter 6). Each research chapter is designed as a standalone contribution, comprising an abstract, an introduction with a focused literature review, a detailed account of the datasets and methodological framework, followed by results presentation, discussion of key findings, and a concluding section that synthesizes the chapter's contributions. **Chapter 2** systematically allocates water budget components from diverse auxiliary sources to illustrate biases and distortions in GRACE-GWA estimates resulting from water storage component selection; this chapter has been published in *Journal of Hydrology*. **Chapter 3** employs a multi-objective comparative framework that integrates NSE and KGE metrics to assess multi-model GRACE-GWA against basin-scale groundwater observations, refining GRACE-GWA comparison; this chapter is currently under review in *Journal of Hydrology*. **Chapter 4** re-evaluates the GRACE-Groundwater Drought Index (GGDI) by incorporating multi-model GRACE-GWA estimates to assess the impact of inter-model variability on groundwater drought interpretation across 37 major aquifers; this chapter is currently under review in *Water Resources Research*. **Chapter 5** is framed as a technical commentary, addressing the challenges of isolating large-scale groundwater signals from GRACE, and builds on the advanced understanding developed throughout this PhD research; this chapter is intended for submission to *Groundwater* as a technical paper. Finally, **Chapter 6** presents the overall conclusions, synthesizing the key findings and contributions of the research while outlining future research directions.

Chapter 2. Challenges in Applying Water Budget Framework for Estimating Groundwater Storage Changes from GRACE Observations

This chapter is based on a manuscript published in the Journal of Hydrology (Akl & Thomas, 2024)

Chapter abstract: The application of a water budget framework to isolate Gravity Recovery and Climate Experiment (GRACE) groundwater storage anomalies (GRACE-GWA) from GRACE terrestrial water storage anomalies (GRACE-TWSA) is hindered by the lack of direct observations of water budget components. In GRACE groundwater studies, water budget components are frequently applied to isolate changes in a storage component from various auxiliary methods (e.g., land surface or hydrology models, reanalysis, or remote sensing) and are used as a supplement to, or a substitute for, in-situ measurements when direct observations are sparse or unavailable. The contribution of select auxiliary datasets to isolate GRACE-GWA from GRACE-TWSA is an enduring quandary, attributed to the various assumptions applied to resemble hydrologic processes in model formulations. This study systematically allocates water budget components from assorted auxiliary sources to demonstrate bias and distortion of GRACE-GWA resulting simply from water storage component data selection. Whereas previous GRACE-GWA studies used combinations of water budget component datasets with uncertainty derived from the variance of the combined water budget components, this study applies single estimates for each component to measure bias and variability. An initial comparative analysis focused on three basins with suitable in-situ groundwater observations and complex hydrology stores (e.g., large lakes and seasonal snow cover), with correlation coefficients ranging from 0.32 to 0.89. The analysis was extended to 12 additional basins that capture a range of hydrologic characteristics to highlight the inconsistency in GRACE-GWA estimates. The variability evident in GRACE-GWA estimates highlights that the selection of water budget components carries the risk of misleading outcomes when disaggregating GRACE-TWSA into GRACE-GWA. Water budget intercomparison provides an important assessment of the limitation in extraction of GRACE-GWA, highlighting the usefulness of comprehensive appraisal of water budget components for GRACE groundwater studies.

2.1 Introduction

Groundwater is a valuable natural resource for maintaining sustainable economic development and healthy ecosystems (Morris et al., 2003). Groundwater resources provide drinking water for more than 2.5 billion people (Morris et al., 2003) and account for approximately 43% of the world's irrigation water (Alley et al., 2002; Siebert et al., 2010). The global reliance on

groundwater to meet water demand has increased due to natural influences (e.g., droughts) (Scanlon et al., 2012; Castle et al., 2014; Thomas et al., 2017a) and anthropogenic interventions (e.g., increasing population, agricultural activities) (Gleick, 2000; Morris et al., 2003; Scanlon et al., 2007). Groundwater observations are limited in many parts of the world (Giordano, 2009; Sophocleous, 2010; Scanlon et al., 2012), and even when groundwater records do exist, interpretation can be difficult due to data availability, effective quality control, and sparse metadata (Rodell et al., 2007; Famiglietti et al., 2011). Increased global reliance combined with limited and inaccurate storage surveillance render groundwater resources in many parts of the world to be exploited (Famiglietti, 2014), a phenomenon observed at regional (e.g., Rodell et al., 2009; Scanlon et al., 2012; Thomas & Famiglietti, 2015) to global scales (e.g., Gleeson et al., 2012; Taylor et al., 2013; Döll et al., 2014; Richey et al., 2015a, b).

Remote sensing has produced robust datasets to augment groundwater surveillance, providing vital information to quantify changes in groundwater storage. The Gravity Recovery and Climate Experiment (GRACE) mission, launched in March 2002, stands alone in its ability to measure variations in the gravity field due to redistribution of mass in the Earth system (Tapley et al., 2004). Gravity variations observed by GRACE are interpreted as changes in terrestrial water storage anomalies (GRACE-TWSA) after removal of signals from tides, atmosphere, oceans, and solid earth (Tapley et al., 2004; Bettadpur, 2007; Syed et al., 2008). GRACE-TWSA represents monthly estimates of vertically integrated terrestrial water storage, which includes snow water equivalent anomalies (SWEA), soil moisture storage anomalies (SMA), surface water storage anomalies (SWA), and groundwater storage anomalies (GRACE-GWA) (Syed et al., 2008). Individual water budget components (SWEA, SMA, and SWA) are utilized to isolate GRACE-GWA from GRACE-TWSA across sufficiently large basins ($\geq$~100,000 km^2; Scanlon et al., 2016).

The water budget equation, also known as the mass balance equation, is consistently employed to isolate GRACE-GWA from GRACE-TWSA assuming that removal of signals attributed to other terrestrial water stores (i.e., SWEA, SMA, and SWA) would leave a signal representative of groundwater storage (e.g., Rodell & Famiglietti, 2002). In this approach, GRACE-GWA is isolated from GRACE-TWSA using the equation:

$$\text{GRACE-GWA}_t = \text{GRACE-TWSA}_t - (\text{SWEA}_t + \text{SMA}_t + \text{SWA}_t) \quad (2.1)$$

where the subscript ($_t$) accounts for time (Rodell & Famiglietti, 1999). Robust estimation of GRACE-GWA relies on accurate accounting of water budget components in Equation 2.1, a challenging task due to the lack of direct observations of water budget components over space

and time. To overcome this challenge, researchers have commonly turned to auxiliary datasets, such as outputs from land surface models (LSM), hydrological models, reanalysis products, or remotely sensed observations, to estimate water budget components (Rodell et al., 2007, 2009; Xavier et al., 2010; Frappart et al., 2011a; Shamsudduha et al., 2012; Feng et al., 2013; Chen et al., 2014; Richey et al., 2015b; Thomas & Famiglietti, 2019; Shamsudduha & Taylor, 2020; Li & Wang, 2022).

Different models employ different algorithms, assumptions, and parameterisations to partition water within the terrestrial environment which result in variable estimates of storage components (Saxe et al., 2021). Differential water partitioning within the modelled terrestrial environment is referred to here as model structural bias, terminology that deviates from previous works (Clark et al., 2008; Ruddell et al., 2019; Konapala et al., 2020). An approach applied to mitigate the influence of model selection in the estimation of water budget components is to average multiple model outputs, which further is used to generate GRACE-GWA error (Rodell et al., 2009; Shamsudduha et al., 2012; Voss et al., 2013; Feng et al., 2013; Castle et al., 2014; Nanteza et al., 2016; Huang et al., 2016). Accounting for variance of the sum in this fashion requires that the data are independent, an assumption that is violated given that many Land Data Assimilation System (LDAS) models assign identical forcing data (Rodell et al., 2004; Mitchell et al., 2004; McNally et al., 2017). Further, averaging water budget components masks the transfer of bias that may impede a robust assessment of groundwater storage change. The question of model structural bias in relation to GRACE applications is not new, where studies documented discrepancies between simulated water storage against GRACE-TWSA (Ahmed et al., 2016; Scanlon et al., 2019; Müller Schmied et al., 2021; Hu et al., 2021; Ju et al., 2023). However, open questions remain regarding the use of model output and different sets of data in combination with GRACE-TWSA to estimate variability in groundwater storage.

Previous work has called attention to accounting inefficiencies of the water budget framework, highlighting their impact on GRACE-GWA (Rodell et al., 2007, 2009; Moore and Fisher, 2012; Jin and Feng, 2013; Chen et al., 2016a; Feng et al., 2018; Bonsor et al., 2018; Yin et al., 2020; Saxe et al., 2021; Li & Wang, 2022; Li et al., 2023). Xiao et al. (2015) revealed noteworthy distinctions in groundwater storage changes when comparing combinations from GRACE-TWSA and LSMs included in the NASA Global Land Data Assimilation System (GLDAS) against in-situ groundwater measurements in the Mid-Atlantic region, with the most substantial differences observed in correlation (0.50) and standard deviation (3.00 cm). Śliwińska et al. (2019) conducted a study comparing groundwater storage changes across two Polish river

basins using combinations of GRACE data and 10 LSMs included in GLDAS and the Coupled Model Intercomparison Project Phase 5 (CMIP5), alongside in-situ groundwater data. They noted correlation coefficients varying from -0.67 to 0.74 and root mean square error (RMSE) values between 8.52 and 13.29 cm for the Odra basin, and between -0.64 and 0.80 and 5.37 to 10.89 cm for the Vistula basin. In a study across Tasmania, Australia, Yin et al. (2020) assessed the accuracy of groundwater storage anomaly estimations using GRACE-TWSA with GLDAS and the WaterGAP Global Hydrology Model (WGHM), comparing results with in-situ groundwater data. Notably, over study period 2003 to 2015, significant differences were observed in Pearson correlation, RMSE, and Nash–Sutcliffe efficiency (NSE) with maximum deviations of -58.8%, -32%, and -126%, respectively. The observed variability in GRACE-GWA underscores irregularity in groundwater assessments, emanating from inefficiencies in water budget components. However, these studies often employed a restricted number of auxiliary datasets, lacked comprehensive comparisons with in-situ groundwater data, or cantered their attention on regional-scale basins characterised by similar intricate hydrological dynamics (Xiao et al., 2015; Śliwińska et al., 2019; Yin et al., 2020). A comprehensive and systematic framework is necessary to assess GRACE-GWA variability due to selection of water budget components, and rather than apply an averaging approach to account for model structural bias, this study argue that multi-model combinations can bring about new insights about groundwater behaviours.

No study to date has approached the question of inefficiency by systematically collecting water budget components to assess the practicality of GRACE-GWA estimation. The goal of this study is to address this research gap by systematically assessing distortions in GRACE-GWA, with a specific emphasis on the consequences of water storage component data selection. The study framework incorporates a comprehensive comparative analysis, specifically focus on three hydrologically similar basins with significant snow and surface water storage components that possess in-situ groundwater data. In the comparative analysis, basin-scale in-situ groundwater observations are deemed “truth” to compare GRACE-GWA estimates extracted using assorted water budget components. This detailed comparative analysis integrates five GRACE-TWSA solutions, eight SWEA solutions, six SMA solutions, and three approaches for SWA, resulting in a total of 720 GRACE-GWA realisations across each study basin. This extensive dataset, sourced from existing literature, allows scrutiny of variability in GRACE-GWA estimates while capturing heterogeneity, offering insights into the intricate dynamics influenced by the selection of different water budget components. By utilizing a diverse set of GRACE-TWSA solutions and water budget components, this study aims to provide a nuanced

understanding of how these choices impact GRACE-GWA outcomes. This systematic analysis is further extended to encompass 12 additional basins, each distinguished by unique hydrologic features and varying levels of aridity. This multifaceted approach highlights inconsistencies in accounting for water budget components within Equation 2.1 to regulate estimation of GRACE-GWA (e.g., Rodell et al., 2007; Famiglietti et al., 2011; Castle et al., 2014).

The importance of this study arises from the ubiquitous application of GRACE to assess changes in groundwater storage (e.g., Chen et al., 2016a; Frappart & Ramillien, 2018). At question is the reliability in retrieving signals representative of groundwater changes from GRACE-TWSA using the water budget framework. Do the predictions of GRACE groundwater storage change capture "true" groundwater changes? What inconsistencies in GRACE-GWA occur simply due to selection of water budget components? Rather than offer a recipe for extraction of GRACE-GWA, this study highlights the importance of robust multi-model output and exhaustive accounting of water budget components to enhance assessment of GRACE-GWA variability that may be obscured by averaging model output. Further, the accounting of water budget components is restricted to readily available datasets and the analysis is applied to the GRACE (i.e., not GRACE-FO) hydrology datasets.

2.2 Data and Methods

2.2.1 Study basins

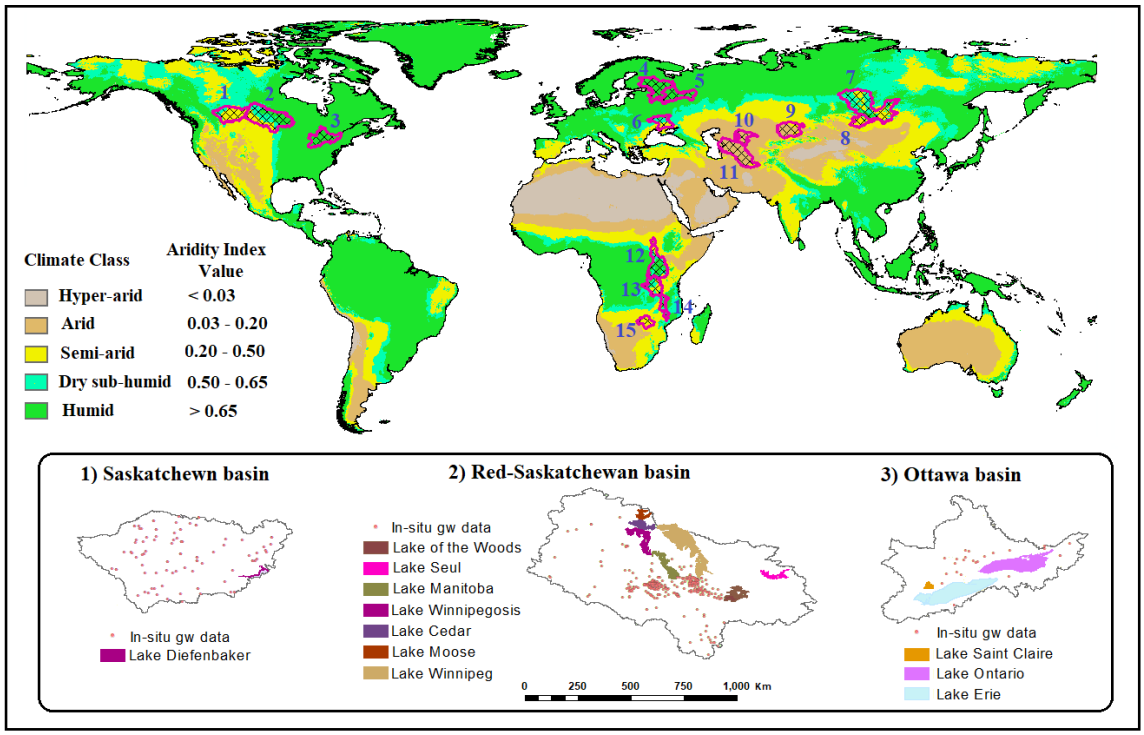


Figure 2.1 Map illustrating locations of the 15 study basins (Table 2.1) as adapted from the Hydrological data and maps based on Shuttle Elevation Derivatives at multiple Scales (HydroSHEDS) (Lehner et al., 2008). Colour shading depicts aridity index based on the Global Potential Evapotranspiration and Global Aridity Index (Zomer et al., 2022). The locations of

surface water bodies and groundwater observation wells are depicted only for selected comparative study basins, namely (1) Saskatchewan, (2) Red-Saskatchewan, and (3) Ottawa basins.

This comparative study required basins with adequate in-situ groundwater observations over the GRACE record, a suitable mix of water budget components (i.e., prominent lakes, seasonal snowpack), and be of large-enough size (i.e., ≥~100,000 km^2). This led us to select three basins: (1) Saskatchewan, (2) Red-Saskatchewan, and (3) Ottawa, located in Canada and the contiguous United States (Figure 2.1), which offer a wide range of constructive attributes. The study was further expanded to encompass an additional 12 basins, each characterised by diverse hydrologic features and aridity (Figure 2.1, Table 2.1). The incorporation of these additional basins significantly strengthens the capacity to assess disaggregating GRACE-TWSA based on the selection of water budget components within Equation 2.1.

Basin ID	Basin name	Continent	Basin area (km2)	Climate zone based on Aridity index	Water budget components
1	Saskatchewan	North America	329163	Semi-arid	SWEA, SMA, SWA, and GWA
2	Red-Saskatchewan	North America	673354	Dry-sub-humid	
3	Ottawa	North America	332328	Humid	
4	Ladoga/Onega	Europe	286995.7	Humid	
5	Rybinsk	Europe	235216.6	Humid	
6	Kremenchuk-Kakhovka	Europe	175696.4	Dry-sub-humid	
7	Baikal/Bratsk	Europe	754575.9	Dry-sub-humid	
8	Khovsgol	Asia	151160.9	Semi-arid	
9	Balkhash	Asia	286013.7	Arid	
10	Large Aral Sea	Asia	164010.8	Arid	
11	Kara-Bogaz-Gol/ Sarygamysh	Asia	570843.7	Arid	
12	Victoria	Africa	589284.1	Dry-sub-humid	SMA, SWA, and GWA
13	Tanganyika	Africa	346122.7	Dry-sub-humid	
14	Malawi	Africa	169843.1	Dry-sub-humid	
15	Kariba	Africa	165737.5	Semi-arid	

Table 2.1 Detailed information of 15 Study basins including basin name, continent, basin area, climate zone based on Aridity index, and water budget components in each basin.

2.2.2 GRACE terrestrial water storage anomalies (GRACE-TWSA)

Five GRACE-TWSA solutions were employed into the framework, including spherical harmonics (SH) (Landerer & Swenson, 2012) and mass concentration (mascon) solutions (M) (Watkins et al., 2015; Wiese et al., 2016). The utilization of multiple GRACE-TWSA solutions, each with its own distinct characteristics (Scanlon et al., 2016), was applied given that solutions can yield varying estimates of GRACE-TWSA. TWSA-SH were processed from the latest release RL-06 V04 with 1˚ × 1˚ global grids by three data processing centres: the Centre for Space Research (CSR-SH), Jet Propulsion Laboratory (JPL-SH), and GeoForschungsZentrum (GFZ-SH) (https://grace.jpl.nasa.gov/data/get-data/). TWSA-M were provided by CSR (CSR-M) from RL-06 V02 with 0.25˚ × 0.25˚ global grids (https://www2.csr.utexas.edu/grace/RL06_mascons.html) and by JPL (JPL-M) from RL-06 V02 with 0.5˚ × 0.5˚ global grids (https://grace.jpl.nasa.gov/data/get-data/). In this study, scale factors derived from the Community Land Model (CLM) were excluded from the processing of JPL-M. These scale factors serve the purpose of redistributing and effectively reducing mass to smaller sub-mascon spatial scales, with the primary aim of minimizing leakage errors (i.e., signal distortion) when precise averaging kernels are used to calculate mass flux within specific hydrological basin boundaries (Wiese et al., 2016). Unlike harmonic-based grids, these scale factors typically have much smaller magnitudes, often approaching unity (Wiese et al., 2016). The decision to not use provided scale factors is derived from the recognition of CLM limitations to accurately simulate snow processes, a dominant factor in the majority of basins examined in this study (Wiese et al., 2016). GRACE-TWSA solutions (CSR-M, JPL-M, CSR-SH, JPL-SH, GFZ-SH), which are anomalies relative to the baseline average (January 2004 to December 2009), were compared over the study period March 2003 to December 2016. Gridded GRACE-TWSA datasets were clipped by basin boundaries while retaining their original spatial resolution.

Uncertainty in TWSA-SH encompasses both measurement and leakage errors. Measurement errors are derived from GRACE-TWSA residuals, excluding the long-term trend and annual/interannual signals (Wahr et al., 2006). Leakage errors are assessed through the root mean square (RMS) difference between unfiltered and filtered monthly mean TWSA estimates from the GLDAS-NOAH, which are then scaled by the ratio of RMS variability between the filtered GRACE-TWSA and GLDAS-NOAH time series (Landerer & Swenson, 2012). Due to spatial correlation of grid cell errors, actual basin scale uncertainties were calculated following Landerer and Swenson (2012; calculation code is available at: https://grace.jpl.nasa.gov/data/get-data/monthly-mass-grids-land/). CSR-M uncertainty

includes the measurements errors, which were estimated from GRACE-TWSA residuals after removing long-term trends and interannual, annual, and semi-annual amplitudes (Wahr et al., 2006; Scanlon et al., 2016). The RMS of the residual approximates the measurement uncertainty in CSR-M. Uncertainty in JPL-M was derived by considering measurement errors provided by JPL (https://grace.jpl.nasa.gov/data/get-data/). Given grid resolution of CSR-M and JPL-M, leakage errors were assumed to be negligible through this study (Scanlon et al., 2016).

2.2.3 Auxiliary water budget datasets

Soil moisture:

Monthly soil moisture anomalies (SMA) were extracted from GLDAS (Rodell et al., 2004), the Famine Early Warning System Network (FEWS NET) Land Data Assimilation System (FLDAS) (McNally et al., 2017), the WaterGAP Global Hydrology Model (WGHM) (Müller Schmied et al., 2014), and ERA5-Land (Muñoz-Sabater et al., 2021). Four land surface models were used to calculate SMA, namely: NOAH at $1^{\circ} \times 1^{\circ}$, and $0.10^{\circ} \times 0.10^{\circ}$ (McNally et al., 2017; Koren et al., 1999), Variable Infiltration Capacity (VIC) at $1^{\circ} \times 1^{\circ}$ (Liang et al., 1996), and Catchment Land Surface Model (CLSM) at $1^{\circ} \times 1^{\circ}$ (Li et al., 2019a). Land surface models (i.e., NOAH, VIC and CLSM) simulate soil moisture through a constant-thickness multi-layer soil profile using different model parameterisation (Table 2.2; Rodell & Famiglietti, 2001; Thomas et al., 2017a). WGHM simulates monthly soil moisture within the effective root zone (Müller Schmied et al., 2021) at a resolution of $0.5^{\circ} \times 0.5^{\circ}$ (Döll et al., 2003; Müller Schmied et al., 2016). WGHM utilizes a one-layer soil water storage compartment, considering land-cover, soil-specific maximum storage capacity, and soil texture. ERA5-Land stands as a reanalysis dataset that offers a consistent view of changes in land-related variables, enriched by a higher resolution of $0.10^{\circ} \times 0.10^{\circ}$. Within ERA5-Land, volumetric soil water is represented across a constant thickness of four soil layers, with considerations for soil texture, soil depth, and the underlying groundwater level (Muñoz-Sabater et al., 2021).

In this study, SMA were constrained to the upper 0–1 m soil profile, as deeper layers are more likely to remain saturated (Thomas et al., 2017a; Fan et al., 2013). However, soil depth representation remains a critical and often underappreciated source of structural uncertainty in water balance modelling. Shallow soil schemes, frequently used in land surface and hydrological models, tend to exhibit artificially persistent soil moisture signals due to limited water-holding capacity and insufficient linkage with deeper subsurface processes (Houborg et al., 2012; Swenson & Lawrence, 2015). Additionally, most models apply fixed saturation thresholds and lack explicit simulation of groundwater–soil moisture interactions (Rodell & Famiglietti, 2001). These structural limitations can introduce bias in the estimation of soil

moisture anomalies (SMA), particularly when deriving groundwater storage anomalies (GRACE-GWA) using the water budget approach. In humid regions, where shallow water tables frequently intersect the root zone, modelled SMA may inadvertently include shallow groundwater contributions, resulting in underestimation of GRACE-GWA (Fan et al., 2013). Conversely, in arid and semi-arid regions with deep unsaturated zones, models may fail to fully capture actual storage changes in the subsurface, leading to overestimation of GRACE-GWA (Scanlon et al., 2009; Fan et al., 2013; Shamsudduha & Taylor, 2020).

Soil moisture products were converted to anomalies by removing the time series mean for January 2004-December 2009 to be consistent with GRACE-TWSA processing. Gridded soil moisture datasets were clipped by basin boundaries while retaining their original spatial resolution.

Model	Spatial Resolution	Soil Layering / Depth Representation	References
NOAH	1.0˚ and 0.10˚	Four layers: 0–10 cm, 10–40 cm, 40–100 cm, 100–200 cm	McNally et al., 2017; Koren et al., 1999
VIC	1.0˚	Three layers: 0–30 cm (surface), variable-depth middle, bottom layer	Liang et al., 1996
CLSM	1.0˚	Three conceptual reservoirs: Surface (0–2 cm), Root Zone (0–100 cm), Profile	Li et al., 2019a
WGHM	0.5˚	Single-layer: Effective root zone	Müller Schmied et al., 2014
ERA5-Land	0.10˚	Four layers: 0–7 cm, 7–28 cm, 28–100 cm, 100–289 cm	Muñoz-Sabater et al., 2021

Table 2.2 Summary of soil moisture representation in the land surface models, hydrological models, and reanalysis product used in this study.

Snow water:

This study encompassed a diverse set of 15 basins, each residing within a range of distinct climate conditions. Notably, the initial 11 basins, numerically designated as 1 to 11 (Figure 2.1), exhibit seasonal snowpack. The remaining basins, situated across Africa and numbered 12 to 15 in Figure 2.1, typically do not experience notable seasonal snowpack.

The derivation of SWEA involved the integration of multiple distinct datasets. These datasets encompass NOAH at $1^{\circ} \times 1^{\circ}$, and $0.10^{\circ} \times 0.10^{\circ}$ (McNally et al., 2017; Koren et al., 1999), VIC at $1^{\circ} \times 1^{\circ}$ (Liang et al., 1996), CLSM at $1^{\circ} \times 1^{\circ}$ (Li et al., 2019a), WGHM at $0.5^{\circ} \times 0.5^{\circ}$ (Döll et al., 2003; Müller Schmied et al., 2016), along with remote sensing/observation based datasets, such as GlobSnow at $0.25^{\circ} \times 0.25^{\circ}$ (Luojus et al., 2021). Additionally, reanalysis data from ERA5-Land at $0.10^{\circ} \times 0.10^{\circ}$ (Muñoz-Sabater et al., 2021) were incorporated into our analysis.

Furthermore, the Historical Gridded Snow Water Equivalent and snow water fraction over Canada (HGSWE) at 0.25° × 0.25° (Mudryk et al., 2015) were included. It is important to note that the inclusion of HGSWE datasets was spatially limited to the first three basins, specifically the Saskatchewan, Red-Saskatchewan, and Ottawa basins. Similar to SMA, the time series mean for January 2004-December 2009 was removed from the snow water products, to ensure consistency with the processing of GRACE-TWSA. Additionally, gridded snow water datasets were clipped by basin boundaries while preserving their original spatial resolution.

Surface water storage:

Surface water storage anomalies (SWA) represent temporal changes in surface water bodies, including lakes, reservoirs, rivers, and floodplains. These components play a central role in regulating basin-scale hydrological processes, acting as buffers that modulate both short- and long-term water availability. Despite their hydrological importance, observational data for surface water storage, particularly for rivers and floodplains, remain sparse, inconsistently distributed, and often lacking the spatial and temporal resolution needed for integration with GRACE-based studies (Alsdorf et al., 2007; Busker et al., 2019; Tortini et al., 2020).

Dynamic storage changes in river channels and floodplains were not explicitly incorporated due to the absence of globally consistent and reliable observational records. In the absence of direct measurements, surface water components are commonly inferred using river-routing modules embedded within large-scale hydrological models (Han et al., 2009; Kim et al., 2009) or estimated through land surface models (Thomas et al., 2017b; Shamsudduha & Taylor, 2020). However, these model-based estimates often lack adequate calibration and validation, especially in data-scarce regions, and are associated with substantial structural and parametric uncertainties (Gao et al., 2012). Accordingly, this study focused on lakes and reservoirs as the primary SWA component, using datasets derived from remote sensing and available in-situ records.

Lake water-level observations for the initial three basins, namely Saskatchewan, Red-Saskatchewan, and Ottawa basins (as shown in Table 2.3 and Figure 2.1), were obtained from the Canadian Water Office (https://wateroffice.ec.gc.ca/). Meanwhile, for the remaining study basins (Table 2.3), lake water-level observations were sourced from the Global Reservoirs and Lakes Monitor (G-REALM; https://ipad.fas.usda.gov/cropexplorer/global_reservoir/; Tortini et al., 2020). Lake water-level observations were combined with lake area from the global lakes bathymetry datasets (GLOBathy; Khazaei et al., 2022) to calculate monthly average lake water storage anomalies, a function of lake area and water height (Refer to Figures 2.2 and 2.3 for the

initial three basins and Figure A.1 in Appendix A for the remaining basins). The calculations considered the variations in lake-surface area due to fluctuations in lake water-level. Monthly average lake water storages were converted to anomalies by removing the time series mean for January 2004-December 2009 to align with the processing of GRACE-TWSA.

Basin	Lake	Reference Area (km^2)	Standard Deviation (km^3)
1- Saskatchewan (329163 km^2)	Lake Diefenbaker	430	0.61
2- Red-Saskatchewan (673354 km^2)	Lake Winnipeg	23923	5.55
	Lake Winnipegosis	5150	0.95
	Lake Manitoba	4610	0.36
	Lake of the Woods	4350	0.45
	Lake Seul	1450	0.67
	Lake Moose	1340	0.88
	Lake Cedar	1320	0.20
3- Ottawa (332328 km^2)	Lake Erie	25821	4.65
	Lake Ontario	19009	4.30
	Lake Saint Clair	1161	0.14

Table 2.3 Ranked reference lake area per basin and standard deviation in lake storage changes across (1) Saskatchewan, (2) Red-Saskatchewan, (3) and Ottawa basins (Lehner et al., 2008; Khazaei et al., 2022). Detailed information on the remaining study basins can be found in Table A.1 within Appendix A.

After estimation of lake storage anomalies, the resulting anomalies must be spatially distributed to estimate water equivalent depth anomalies. In this study, two simple approaches were examined: the basin average approach ($SWA_{av.}$) and the grid approach (SWA_{grid}). The basin average approach applies uniform mass-change distribution across the basin as a water equivalent thickness. For each basin, individual lake storage anomalies were summed and scaled by basin area. The grid approach projected lake water height anomalies to GRACE grid resolutions. This required subjecting lake water height anomalies to processing steps mirroring those applied to GRACE data. Notably, the grid approach was implemented differently for SH versus mascon solutions, reflecting the distinct processing steps employed in each approach. For SH solutions ($SWA_{grid\ 1^{\circ}}$), lake water height anomalies were projected to GRACE grids and underwent similar processing steps, including truncation at harmonic degree 60, destriping using a decorrelation filter following Swenson and Wahr (2006), and smoothing with a 300 km radius Gaussian filter. The grid approach for mascon solutions ($SWA_{grid\ 0.25^{\circ}}$ for CSR-M and $SWA_{grid\ 0.5^{\circ}}$ for JPL-M) assumed that potential leakage across GRACE grids is minimized during processing, and thus lake water height anomalies were directly removed from respective GRACE-TWSA grids prior to averaging to the basin scale.

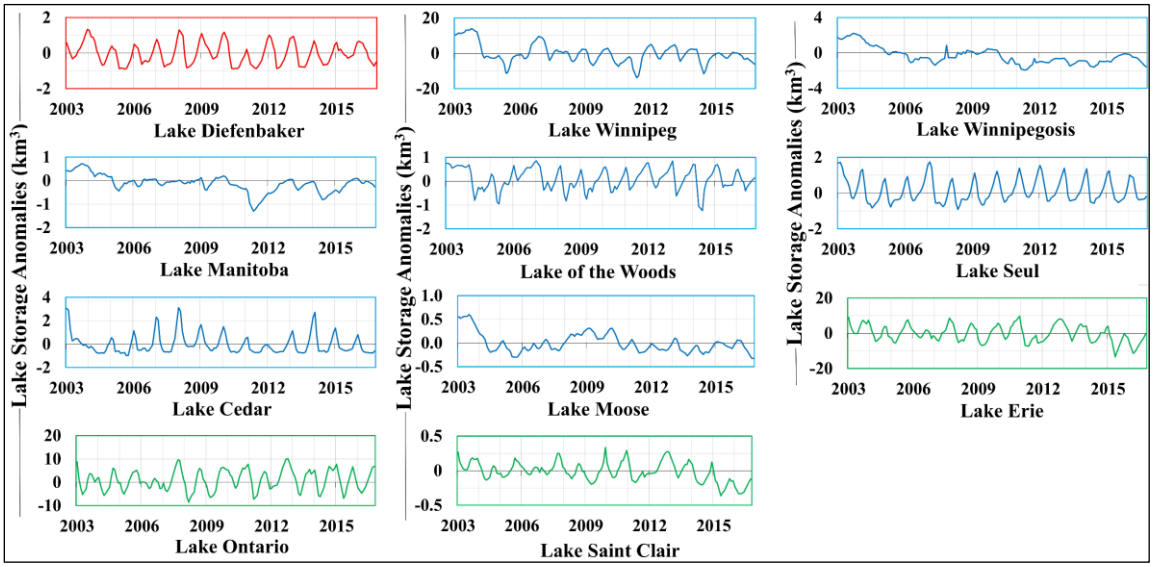


Figure 2.2 Lake storage anomalies from March 2003 till December 2016 across (1) Saskatchewan (red), (2) Red-Saskatchewan (blue), (3) and Ottawa (green) basins (https://wateroffice.ec.gc.ca/; Khazaei et al., 2022).

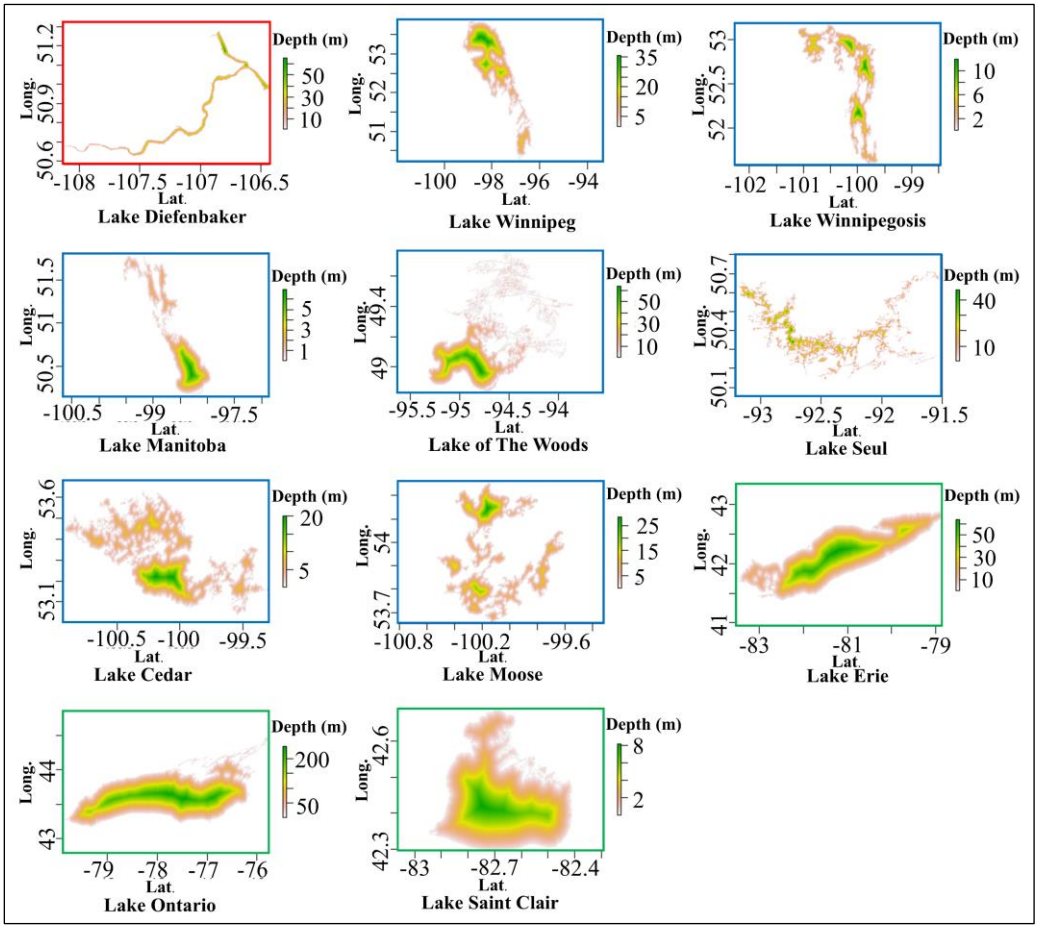


Figure 2.3 Bathymetry data in resolution of 1 arc-seconds and WGS84 projection system for individual study lakes across (1) Saskatchewan (red), (2) Red-Saskatchewan (blue), and (3) Ottawa (green) basins (Khazaei et al., 2022). Variability in bathymetry and lake area are notable in spatial and depth scales. Bathymetry data of lakes across the remaining study basins can be found in Figure A.1 within Appendix A.

2.2.4 Extraction of GRACE-GWA

Groundwater storage anomalies (GRACE-GWA) were extracted by removing water budget components from GRACE-TWSA using the water budget equation (Equation 2.1) (Rodell & Famiglietti, 1999). The application of the water budget equation varied across study basins to account for the unique water budget components present in each basin. For basins 1-11 (Figure 2.1), which include seasonal snowpack, the calculation of GRACE-GWA from GRACE-TWSA involved incorporating SWEA into Equation 2.1. For basins 12-15 (Figure 2.1), SWEA was assumed to be zero in Equation 2.1 due to no significant seasonal snowpack. The combinations of five GRACE-TWSA solutions (CSR-SH, JPL-SH, GFZ-SH, CSR-M, JPL-M), eight SWEA datasets ($SWEA_{NOAH\ 1°}$, $SWEA_{NOAH\ 0.1°}$, $SWEA_{VIC\ 1°}$, $SWEA_{CLSM\ 1°}$, $SWEA_{WGHM\ 0.5°}$, $SWEA_{GlobSnow\ 0.25°}$, $SWEA_{HGSWE\ 0.25°}$, $SWEA_{ERA5\text{-}Land\ 0.1°}$), six SMA datasets ($SMA_{NOAH\ 1°}$, $SMA_{NOAH\ 0.1°}$, $SMA_{VIC\ 1°}$, $SMA_{CLSM\ 1°}$, $SMA_{WGHM\ 0.5°}$, $SMA_{ERA5\text{-}Land\ 0.1°}$), and three SWA datasets (no-SWA (i.e., assuming SWA = 0), $SWA_{av.}$, SWA_{grid}) produced 720 realisations of GRACE-GWA across Saskatchewan, Red-Saskatchewan, and Ottawa basins (numbered 1 to 3 in Figure 2.1). For the other snow basins (numbered 4 to 11 in Figure 2.1), where the spatial coverage of $SWEA_{HGSWE\ 0.25}$ was limited, 610 realisations of GRACE-GWA were generated. Finally, for the basins without a snowpack signal (numbered 12 to 15 in Figure 2.1), the exclusion of the SWEA component from Equation 2.1 resulted in 90 realisations of GRACE-GWA for each basin. Whereas previous GRACE-GWA studies used combinations (i.e., averages) of water budget component datasets which allowed for estimation of uncertainty following Rodell and Famiglietti (2002), this study applied a single estimate for each component (i.e., SWEA, SWA, SMA) to account for water storage heterogeneity and thus uncertainty in GRACE-GWA realisations was not estimated.

2.2.5 Basin-scale in-situ groundwater observations

In-situ groundwater observations were accessible specifically for the Saskatchewan, Red-Saskatchewan, and Ottawa basins. These observations were provided by the Groundwater Information Network (GIN) (https://gin.gw-info.net/service/api_ngwds:gin2/en/gin.html) and the National Groundwater Network (NGWMN) (https://cida.usgs.gov/ngwmn/index.jsp). A total of 608, 866, and 524 observation wells data were extracted across Saskatchewan, Red-Saskatchewan, and Ottawa basins, respectively, many with irregular temporal observations. A well selection algorithm was applied to isolate suitable observation well data with no more than one missing seasonal measurements through each calendar year, identifying 95, 317, and 21 quality-controlled observation wells across Saskatchewan, Red-Saskatchewan, and Ottawa basins, respectively (Figure 2.1 and Figures A.2-A.13 and Table A.2-A.8 in Appendix A).

Interpolation to monthly in-situ groundwater storage changes applied geospatial and temporal gap filling approaches (Zeng & Levy, 1995) given as:

$$gw_{est} = \frac{\sum_{k=1}^{n} w_k gw_k}{\sum_{k=1}^{n} w_k} \tag{2.2}$$

$$w_k = \frac{2 - \left[\frac{(x_k - x_o)^2 + (y_k - y_o)^2}{D^2} + \frac{(t_k - t_o)^2}{T^2}\right]}{2 + \left[\frac{(x_k - x_o)^2 + (y_k - y_o)^2}{D^2} + \frac{(t_k - t_o)^2}{T^2}\right]} \tag{2.3}$$

The interpolator utilizes both temporal and spatial information to estimate missing values, taking advantage of conditions where spatial information is substituted for temporal information if it is missing, and vice versa (Zeng & Levy, 1995). Equations 2.2 and 2.3 were applied with a 1-month time step (T), and an arbitrary 100 km spatial range (D). The selected spatial range was set high to capture numerous water level observations, while recognizing that near well data exerts a larger influence on missing observations as the weight diminishes with the distance. The variables x_k, y_k, and t_k represent the location and time of the non-missing groundwater observation gw_k within the distance D and time range T. Similarly, x_o, y_o, and t_o represent the location and time of the missing groundwater observation gw_{est} within the same distance D and time range T. The weight assigned to each non-missing monthly value, w_k, is based on its temporal and spatial gap from the missing monthly groundwater observation, meaning that higher weights are given to close wells with concurrent groundwater observations. Previous studies successfully applied the interpolator to rectify irregular in-situ groundwater records for comparisons with GRACE-GWA (Nanteza et al., 2016; Thomas et al., 2017a).

Spatial bias, whereby observation wells are not uniformly distributed across a basin, is commonly encountered in hydrogeology studies (Rodell et al., 2007; Huang et al., 2016; Chen et al., 2016a; Li & Wang, 2022). To evade spatial bias concerns in development of a basin average time series, Rodell et al. (2007) constructed an average groundwater time series using Thiessen Polygon (TP) interpolation. Chen et al. (2016) upscaled in-situ observations to $1^{\circ}x1^{\circ}$ grids, later using grids to obtain a groundwater level time series for the study region. In this study, in-situ groundwater observations were spatially interpolated using TP interpolation (Brassel et al., 1979) to produce a basin-scale groundwater observation time series (Figure 2.4). The TP interpolation method is applied given the simplicity and rapid replication of the

approach. Bootstrapping interpolations were conducted to validate this spatial averaging approach, whereby different numbers of wells were removed from spatial averaging methods to assess variability in the resulting basin-scale groundwater observation time series. Results documented an unremarkable influence of missing observations on the basin-scale groundwater observation time series (Figures A.14 to A.16 in the Appendix A).

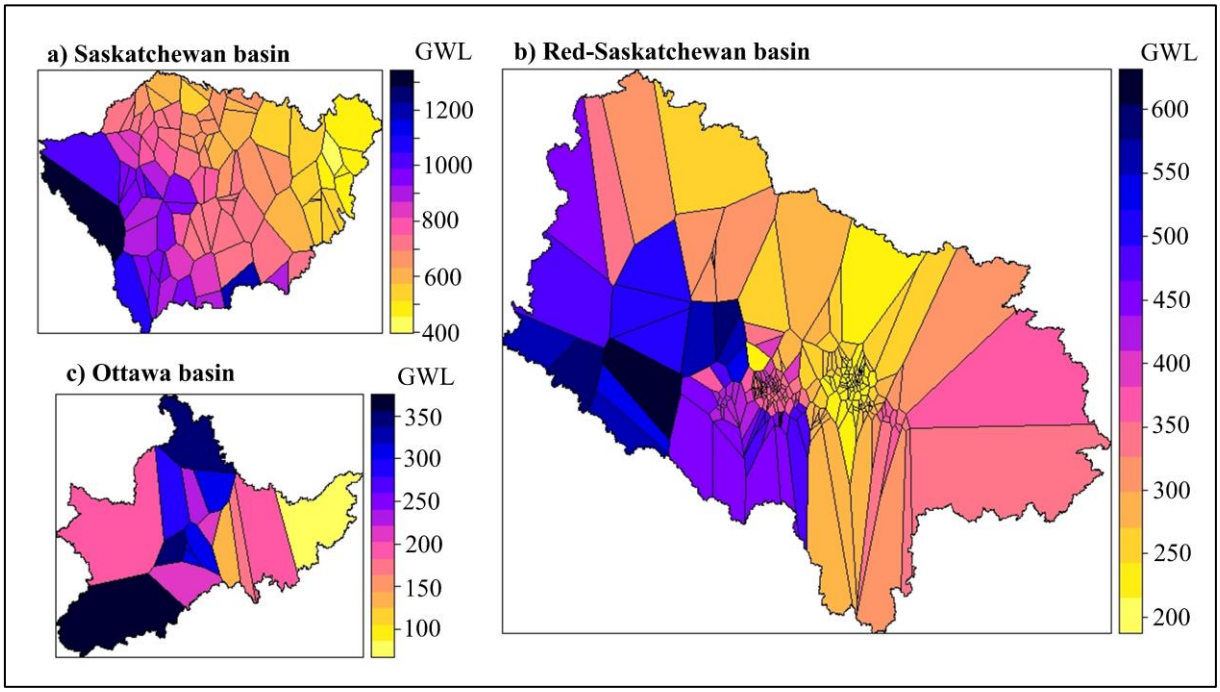


Figure 2.4 The application of the Thiessen Polygons interpolation method across the study basins, visualizing in-situ groundwater observations for March 2003.

Recognising the potential limitations of relying on a single interpolation method, Chapter 3 provides an extended evaluation of alternative spatial interpolation techniques. Specifically, TP, Inverse Distance Weighting (IDW), Universal Kriging (UK), Ordinary Kriging (OK), Log-transformed Ordinary Kriging (OK-Log), Multi-Linear Regression (MLR), and Spatio-Temporal Kriging (STK) were tested (Brassel & Reif, 1979; Oliver et al., 1990; Lu et al., 2008; Li & Heap, 2014; Gräler et al., 2016). The performance of these methods was assessed through leave-one-out cross-validation (LOOCV) following the procedures of Van Zoest et al. (2020) and Ferreira et al. (2023). The comparative results (Figures B.1–B.7 in Appendix B) demonstrated high consistency in the derived basin-scale time series, thereby validating the initial use of TP in this chapter and establishing a solid methodological foundation for subsequent analyses.

The basin-scale groundwater observation time series was transformed to an in-situ anomaly time series by removing the time series mean for January 2004-December 2009. Converting an in-situ anomaly time series to groundwater storage anomalies requires information on the

aquifer storativity (i.e., Sy for unconfined aquifers). Rodell et al. (2007) heralded the importance of applying representative storage coefficients despite the recognised challenges of estimating Sy (Sun et al., 2010). As no Sy data was readily available for study basins, the in-situ anomaly time series was normalised to facilitate meaningful comparisons with GRACE-GWA. For consistency, the normalised time series was referred as the normalised in-situ groundwater anomaly.

2.2.6 Metrics of comparison

Spearman's rank correlation coefficient (ρ) and an amplitude ratio were applied to conduct comparisons between various time series. Spearman's ρ was utilized to measure the strength of the monotonic relationship between variables and is applied here for consistency with previous GRACE studies (Sun, 2013; Nanteza et al., 2016; Ouatiki et al., 2021; Han et al., 2021; Zhang et al., 2021a; Amiri et al., 2023). The amplitude ratio was utilized to assess the degree of variability or difference between two time series (Müller Schmied et al., 2021). The annual amplitude of time series was calculated using *cosinor* R-package (Sachs, 2015; https://github.com/sachsmc/cosinor). The outcome of time series (Y) with time variable (t) and fixed period (D) is of interest to fit periodic model given as:

$$Y(t) = \alpha + A * \cos 2 * \pi * t/D - \beta_1 + \varepsilon \quad (2.4)$$

where α is the intercept, A is the amplitude of the time series, β_1 is the acrophase, and ε is an error term with mean 0. The amplitude ratio between two time series $A_{1,2}$ was calculated as

$$A_{1,2} = A_1/A_2 \quad (2.5)$$

An amplitude ratio of one implies identical amplitudes (e.g., $A_1 = A_2$), while any deviation from one signifies differences in the amplitudes.

2.3 Results

2.3.1 Comparisons of GRACE-terrestrial water storage anomalies (GRACE-TWSA)

The primary objective of this study is to assess variability in GRACE-GWA resulting from water budget component selection. However, it is crucial to recognise that the choice of GRACE-TWSA solutions can also impact the extraction of GRACE-GWA. Thus, the analysis commences with a comparison of various GRACE-TWSA solutions, focusing on three basins: Saskatchewan, Red-Saskatchewan, and Ottawa. Agreement in GRACE-TWSA amplitudes were visibly noted during the anomaly calculation period 2004-2009 (pink boxes in Figure 2.5) as compared to the remainder of the study period. A correlation matrix using Spearman's ρ was applied to depict the relationship between GRACE-TWSA solutions (Figure 2.6). High correlations were noted for Saskatchewan, Red-Saskatchewan, and Ottawa basins, with correlation magnitudes ranging from 0.93 to 0.98, 0.94 to 0.98, and 0.91 to 0.98, respectively,

documenting GRACE-TWSA solutions similitude across study basins. While accounting for GRACE-TWSA uncertainty relative to SH solutions and mascon solutions, GRACE-TWSA solutions were found to be within estimated uncertainty (Figure 2.5). The consistency observed among GRACE-TWSA solutions serves as an important validation of the approach involving the clipping of gridded GRACE-TWSA datasets based on basin boundaries while preserving their original spatial resolution.

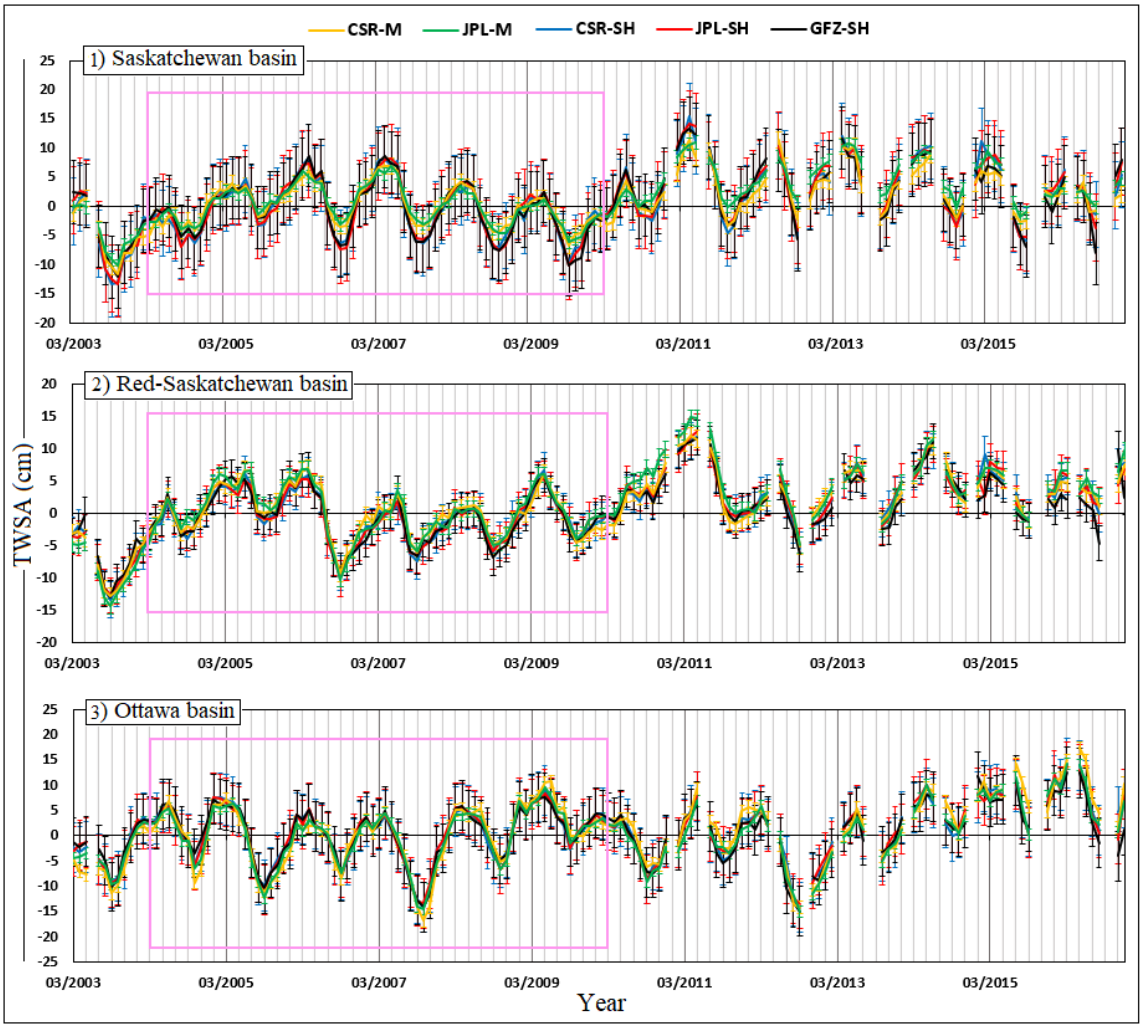


Figure 2.5 Monthly time series of GRACE-Terrestrial Water Storage Anomalies (GRACE-TWSA) derived from CSR-M, JPL-M, CSR-SH, JPL-SH, and GFZ-SH solutions across (1) Saskatchewan, (2) Red-Saskatchewan, and (3) Ottawa basins. Error bars depict computed uncertainty in GRACE-TWSA solutions and pink boxes represent the anomaly calculation time (2004-2009) of GRACE-TWSA solutions. A total of 17 data gaps were noted during the study period, represented by temporal spaces within the plotted time series.

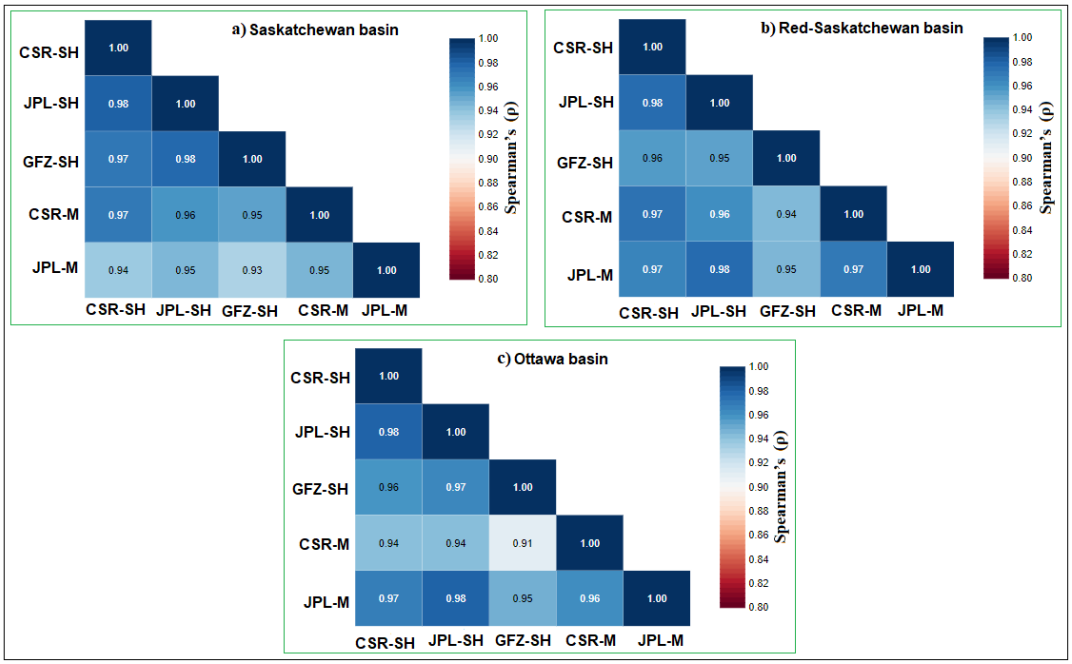


Figure 2.6 Spearman's rank correlation coefficients (ρ) between 5 GRACE-TWSA solutions (CSR-SH, JPL-SH, GFZ-SH, CSR-M, and JPL-M) over the period 2003-2016 across (a) Saskatchewan, (b) Red-Saskatchewan, and (c) Ottawa basins.

2.3.2 Quantifying variability in water budget components

SMA and SWEA products:

SWEA / SMA	NOAH 1°	NOAH 0.1°	VIC 1°	CLSM 1°	HGSWE 0.25°	WGHM 0.5°	GlobSnow 0.25°	ERA5-L 0.1°
NOAH 1°	1	7	13	19	25	31	37	43
NOAH 0.1°	2	8	14	20	26	32	38	44
VIC 1°	3	9	15	21	27	33	39	45
CLSM 1°	4	10	16	22	28	34	40	46
WGHM 0.5°	5	11	17	23	29	35	41	47
ERA5-L 0.1°	6	12	18	24	30	36	42	48

Table 2.4 Summary of 48 datasets combinations consisting of eight Snow Water Equivalent Anomalies (SWEA) and six Soil Moisture Anomalies (SMA), across (1) Saskatchewan, (2) Red-Saskatchewan, and (3) Ottawa basins.

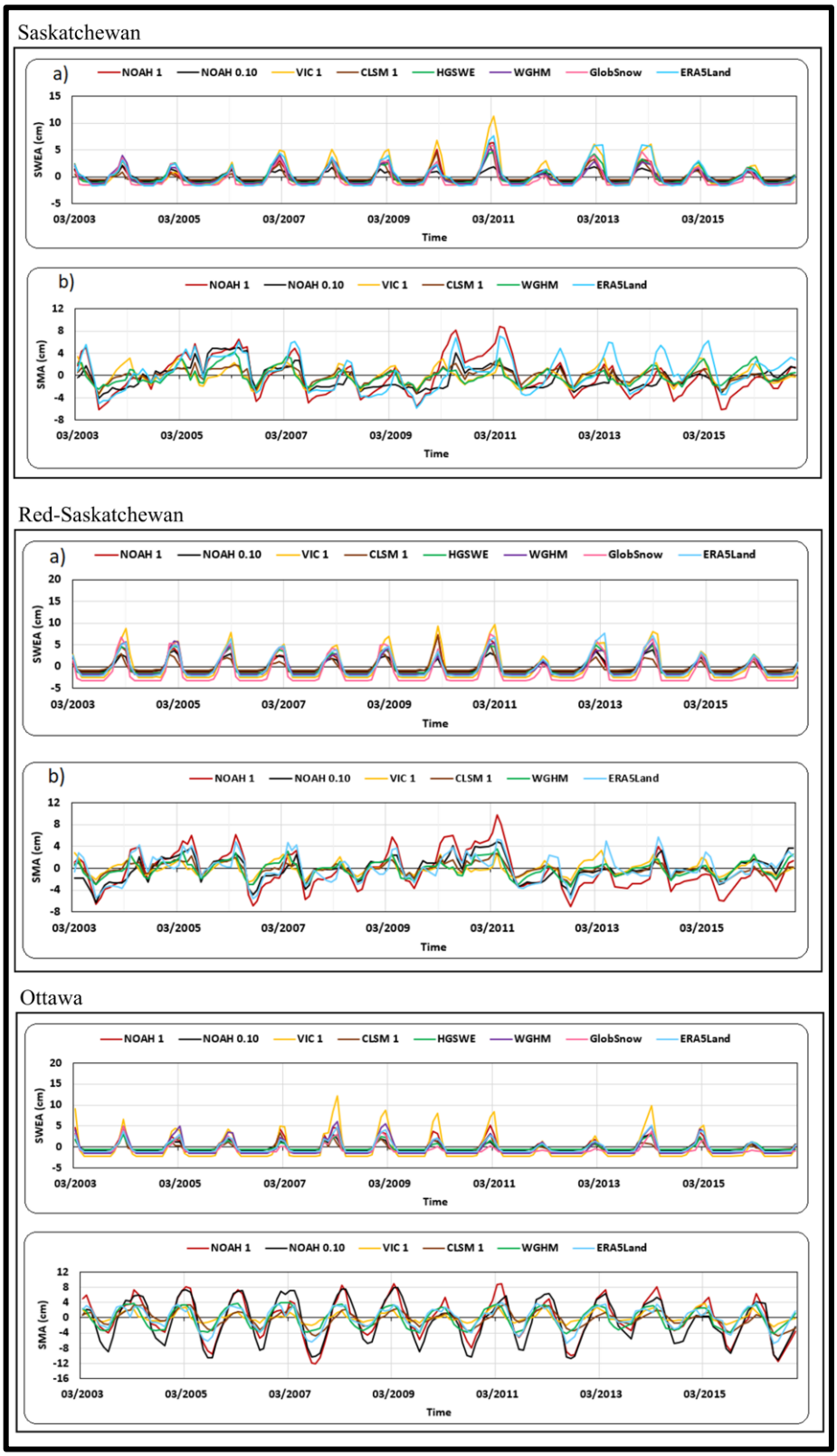


Figure 2.7 Monthly time series of (a) Snow Water Equivalent Anomalies (SWEA) derived from NOAH 1°, NOAH 0.10°, VIC 1°, CLSM 1°, HGSWE 0.25°, WGHM 0.5°, GlobSnow 0.25°, and ERA5-Land 0.10° and (b) Soil Moisture Anomalies (SMA) derived from NOAH 1°, NOAH 0.10°, VIC 1°, CLSM 1°, WGHM 0.5°, and ERA5-Land 0.10° over the period 2003-2016 across Saskatchewan, Red-Saskatchewan, and Ottawa basins.

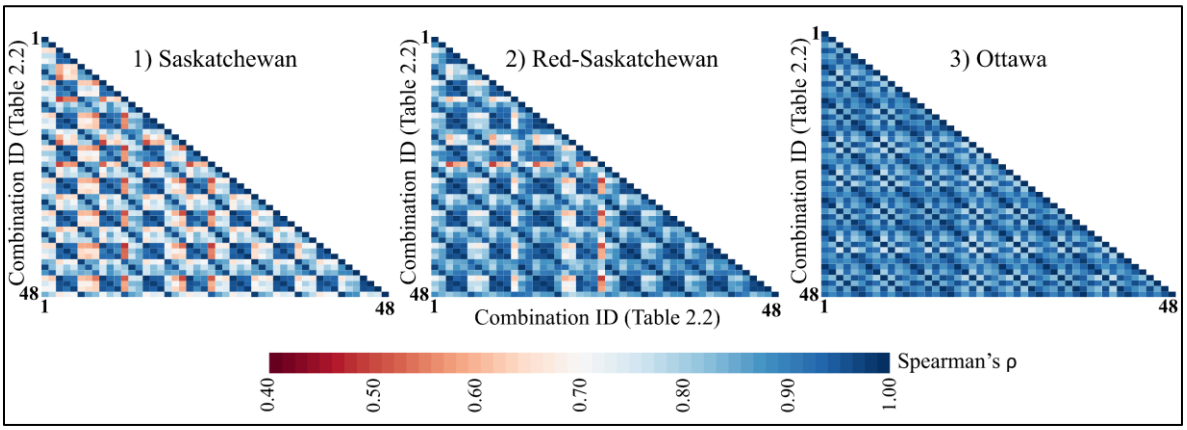


Figure 2.8 Spearman's ρ between 48 time series estimations of SWEA + SMA (Table 2.4) for (1) Saskatchewan, (2) Red-Saskatchewan, and (3) Ottawa basins.

SWEA and SMA water budget components are often applied as the primary water budget components in extraction of GRACE-GWA, with some studies estimating GRACE-GWA assuming effectively no signal contribution from SWA (Rodell et al., 2007, 2009; Chen et al., 2014; Hachborn et al., 2017; Meghwal et al., 2019). Combinations of SWEA+SMA datasets exhibit variable seasonality and amplitudes (Figure 2.7), factors that can influence the timing and amplitude of GRACE-GWA. Correlation matrix depicting Spearman's ρ between combinations of SWEA and SMA are shown in Figure 2.8. Correlation coefficients ranged from 0.48 to 1.00, 0.49 to 1.00, and 0.76 to 1.00 for Saskatchewan, Red-Saskatchewan, and Ottawa basins, respectively. Correlation amongst SWEA+SMA combinations was lower when VIC, CLSM, and WGHM were used to account for SMA as compared to combinations from NOAH and ERA5-Land. The lowest correlation among the combinations was found when comparing VIC combinations to those that applied ERA5-Land across Saskatchewan and Red-Saskatchewan basins, and to those that applied CLSM across Ottawa basin. Notably, it is important to recognise that these widespread Spearman's ρ magnitudes occur between datasets assumed to represent the same combination of water budget components (i.e., SWEA + SMA).

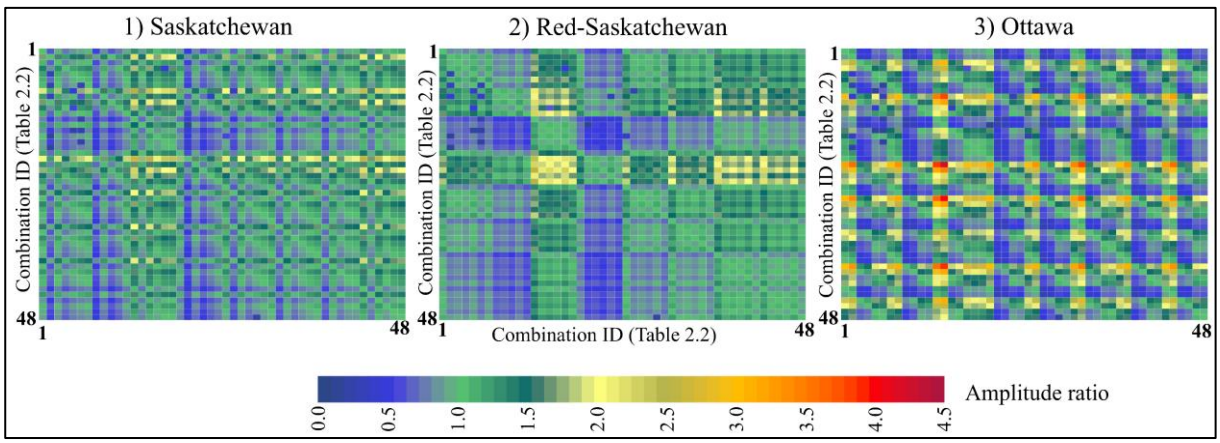


Figure 2.9 Amplitude ratio (A1,2) between 48 time series estimations of SWEA + SMA (Table 2.4) for (1) Saskatchewan, (2) Red-Saskatchewan, and (3) Ottawa basins.

Amplitude ratios similarly captured distinct differences in combined SMA and SWEA from the various datasets (Figure 2.9). Amplitude ratios ranged from 0.12 to 2.66, 0.08 to 2.56, and 0.08 to 4.26 for Saskatchewan, Red-Saskatchewan, and Ottawa basins, respectively (Figure 2.9). Differences in correlation and amplitude across unique combinations reflect limitations between reanalysis products, simulating snow processes (Singh et al., 2015) and model structure for soil moisture (Clark et al., 2015). Amplitude ratios were clearly influenced by select model budget accounting, where large ratios were attributed to VIC in Saskatchewan and Red-Saskatchewan and NOAH 0.10° for Ottawa.

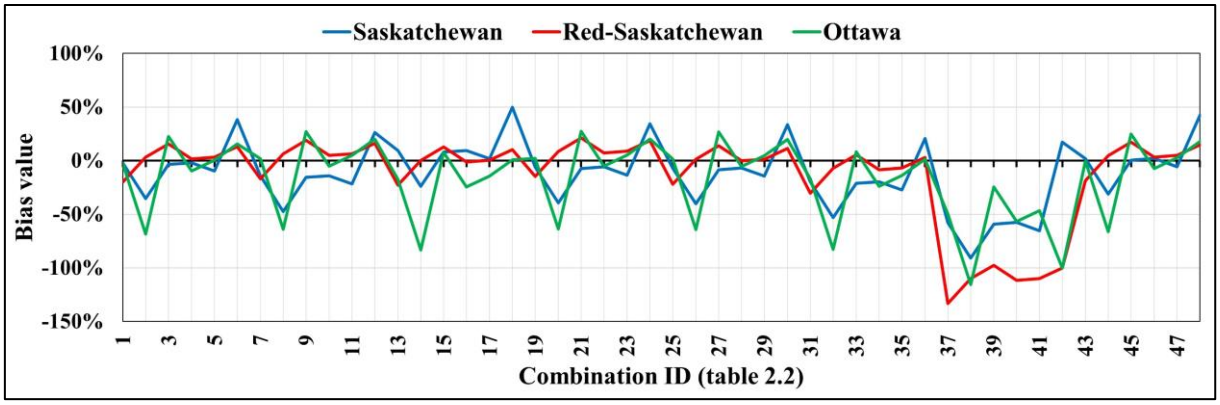


Figure 2.10 Bias of 48 combinations of SWEA + SMA (Table 2.4) relative to the average SWEA and SMA taken from 1-degree GLDAS model output (NOAH, VIC, and CLSM) across Saskatchewan (blue), Red-Saskatchewan (red), and Ottawa (green) basins.

Bias was applied to provide a measure of the differences between various combinations of SWEA and SMA. Bias of SWEA and SMA combinations were measured relative to average SWEA and SMA from 1-degree GLDAS output (NOAH, VIC, CLM) to provide a measure of comparison with previous GRACE studies (e.g., Rodell et al., 2007, 2009; Famiglietti et al., 2011; Shamsudduha et al., 2012; Voss et al., 2013; Feng et al., 2013; Castle et al., 2014; Richey et al., 2015b; Nanteza et al., 2016; Huang et al., 2016; Xu et al., 2021; Ali et al., 2022). Wide inconsistencies were noted (Figure 2.10) with bias exceeding -100% when applying the GlobSnow product to account for SWEA. The varying degrees of bias suggest significant inconsistency between GRACE-GWA time series that result due to model data selection.

SWA accounting:

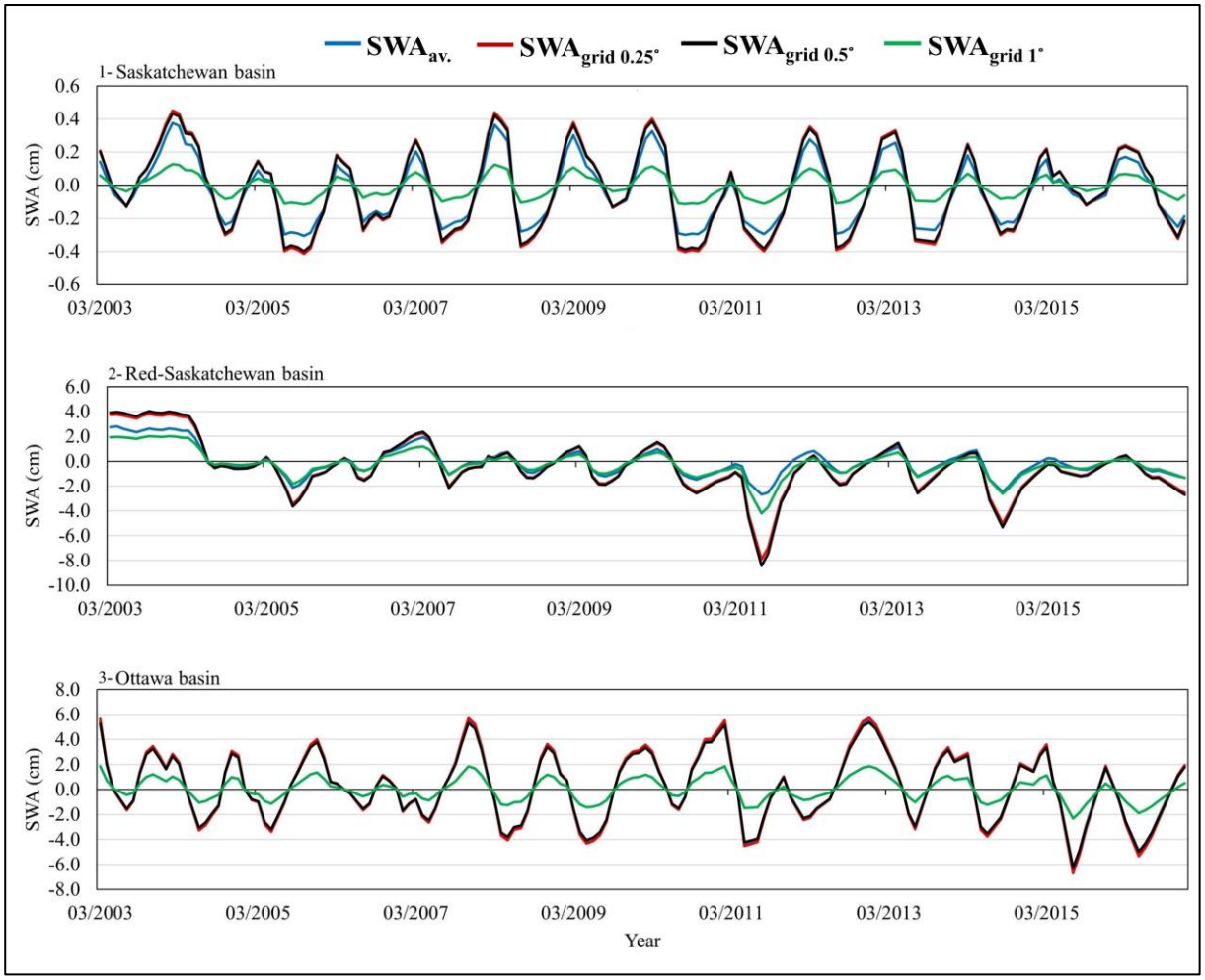


Figure 2.11 Comparisons of SWA derived from lake storage datasets using the basin average approach (SWAav.) and the grid approach (SWAgrid 1°, SWAgrid 0.5°, and SWAgrid 0.25°) across (1) Saskatchewan, (2) Red-Saskatchewan, and (3) Ottawa basins.

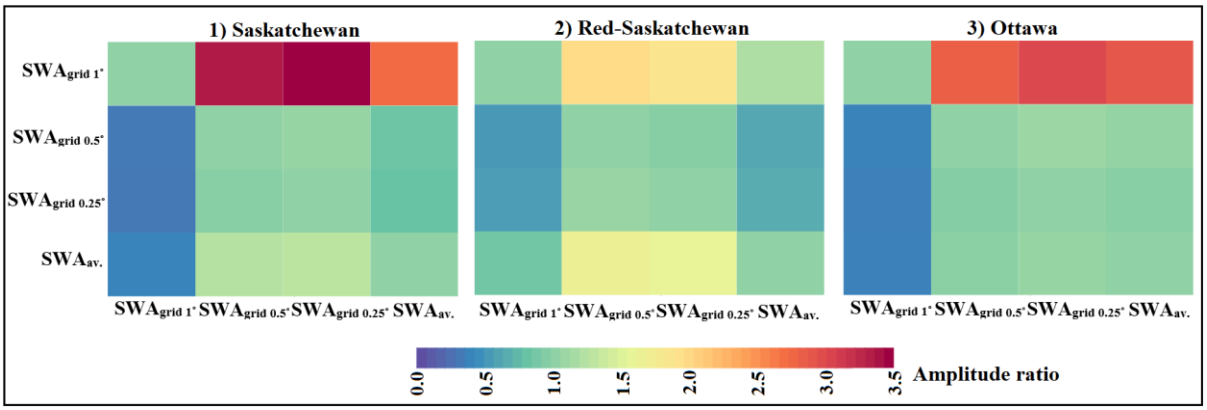


Figure 2.12 Amplitude ratio (A1,2) between SWA estimates from the grid approach (SWAgrid 1°, SWAgrid 0.5°, SWAgrid 0.25°) and the basin average approach (SWAav.) for (1) Saskatchewan, (2) Red-Saskatchewan, and (3) Ottawa basins (Amplitude ratio obtained by dividing

The current state of GRACE-GWA studies in basins with surface water stores reveals a lack of a well-defined framework for converting lake storage data to SWA within Equation 2.1. The

study explored the influence of assumptions on SWA using the basin average approach ($SWA_{av.}$) and the grid approach ($SWA_{grid\ 1°}$, $SWA_{grid\ 0.5°}$, $SWA_{grid\ 0.25°}$) (Figures 2.11 and 2.12). Notable differences in amplitude values were observed among applied approaches (Figure 2.12). Amplitude ratios ranged from 0.29 to 3.50, 0. 51 to 1.98, and 0. 33 to 3.01 for Saskatchewan, Red-Saskatchewan, and Ottawa basins, respectively (Figure 2.12). Notable differences in time series amplitude are attributed to repercussions of spherical harmonic solutions including mass leakage (Longuevergne et al., 2013) and truncation (Longuevergne et al., 2010) influenced by the location of lake storage mass change within the basin (Figure 2.1). For example, the sole lake in the Saskatchewan basin was located at the edge of the basin, while in the Red-Saskatchewan basin, lakes are concentrated toward the centre of the basin (Figure 2.1). The location of the lake storage mass change within the basin impacts lake mass leakage (Longuevergne et al., 2013), whereby significant mass was lost for the Saskatchewan basin as compared to the Red-Saskatchewan basin.

2.3.3 Comparison of GRACE-GWA estimates with normalised in-situ groundwater anomaly

A total of 720 realisations of GRACE-GWA were extracted for the three comparative basins, namely, (1) Saskatchewan, (2) Red-Saskatchewan, and (3) Ottawa basins, from combinations of water budget components (Figure 2.13). These estimates were compared against basin-scale in-situ groundwater level anomalies (normalized to z-scores rather than converted to storage due to a lack of reliable specific yield data). Correlation magnitudes ranged from 0.51 to 0.76 across Saskatchewan basin, 0.60 to 0.87 across Red-Saskatchewan basin and 0.33 to 0.77 across Ottawa basin (Figure 2.14). The wide range of correlation magnitudes highlights the influence of seasonality and temporal offsets known to effect dynamic change (i.e., timing) in GRACE-GWA (Akhtar et al., 2022; Li & Wang, 2022; Thomas et al., 2017a).

Accounting for SWA within the water budget equation resulted in incremental improvements in the correlation between GRACE-GWA and normalised in-situ groundwater anomalies (Figure 2.14). Correlation increases ranging from 0 to 0.02, 0 to 0.10, and 0 to 0.20 were observed across Saskatchewan, Red-Saskatchewan, and Ottawa basins, respectively, when accounting for SWA (Figure 2.14). However, the application of different SWA accounting approaches ($SWA_{av.}$, SWA_{grid}) demonstrated a differential impact on the correlation between GRACE-GWA estimates and basin-scale in-situ observations (Figure 2.14). For mascon solutions, SWA_{grid} demonstrated incremental improvements in the correlation compared to $SWA_{av.}$, while for SH solutions, $SWA_{av.}$ exhibited incremental improvements in the correlation compared to SWA_{grid} (Figure 2.14).

In general, correlation results revealed variable outcomes when comparing GRACE-GWA estimates with normalised in-situ groundwater anomalies. However, in all cases, correlations were statistically significant, preventing any rigorous statements about a change in correlation significance with and without SWA. Increases in correlation magnitudes of up to 0.20 are notable, especially given that the Spearman's ρ is a bounded metric (i.e., [-1,1]). Thus, although the results cannot be supported by statistical measures, a clear benefit is observed as measured by incremental improvements in correlation magnitude.

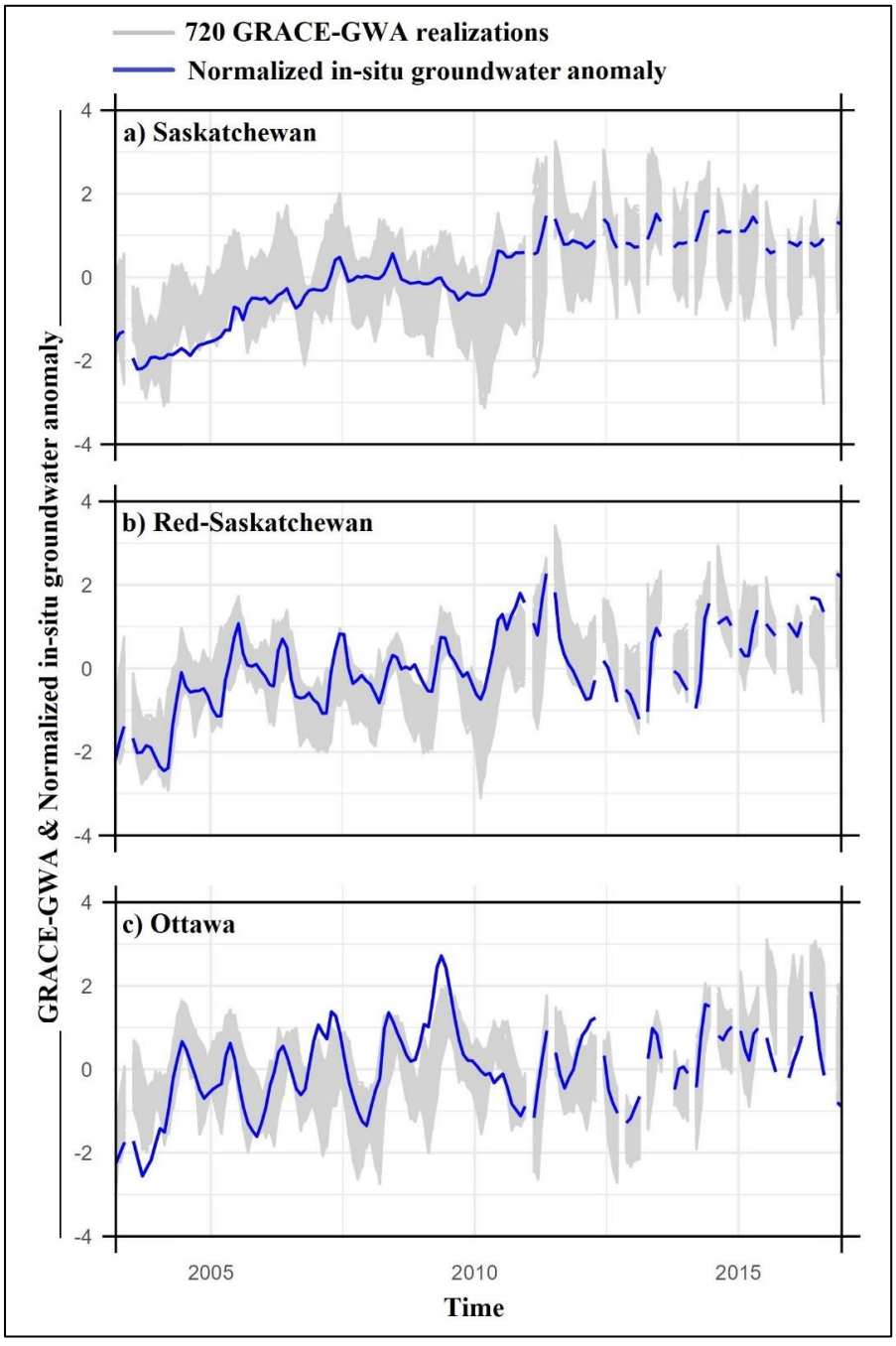


Figure 2.13 Comparisons of 720 realisations of GRACE-Groundwater Anomaly (GRACE-GWA) depicted in grey and normalised basin-scale in-situ groundwater (level) anomaly represented in blue, across (a) Saskatchewan, (b) Red-Saskatchewan, and (c) Ottawa basins. The study spans from March 2003 to December 2016. Due to the unavailability of specific yield

information across the study basins, a normalisation approach was employed to facilitate meaningful comparisons between the datasets.

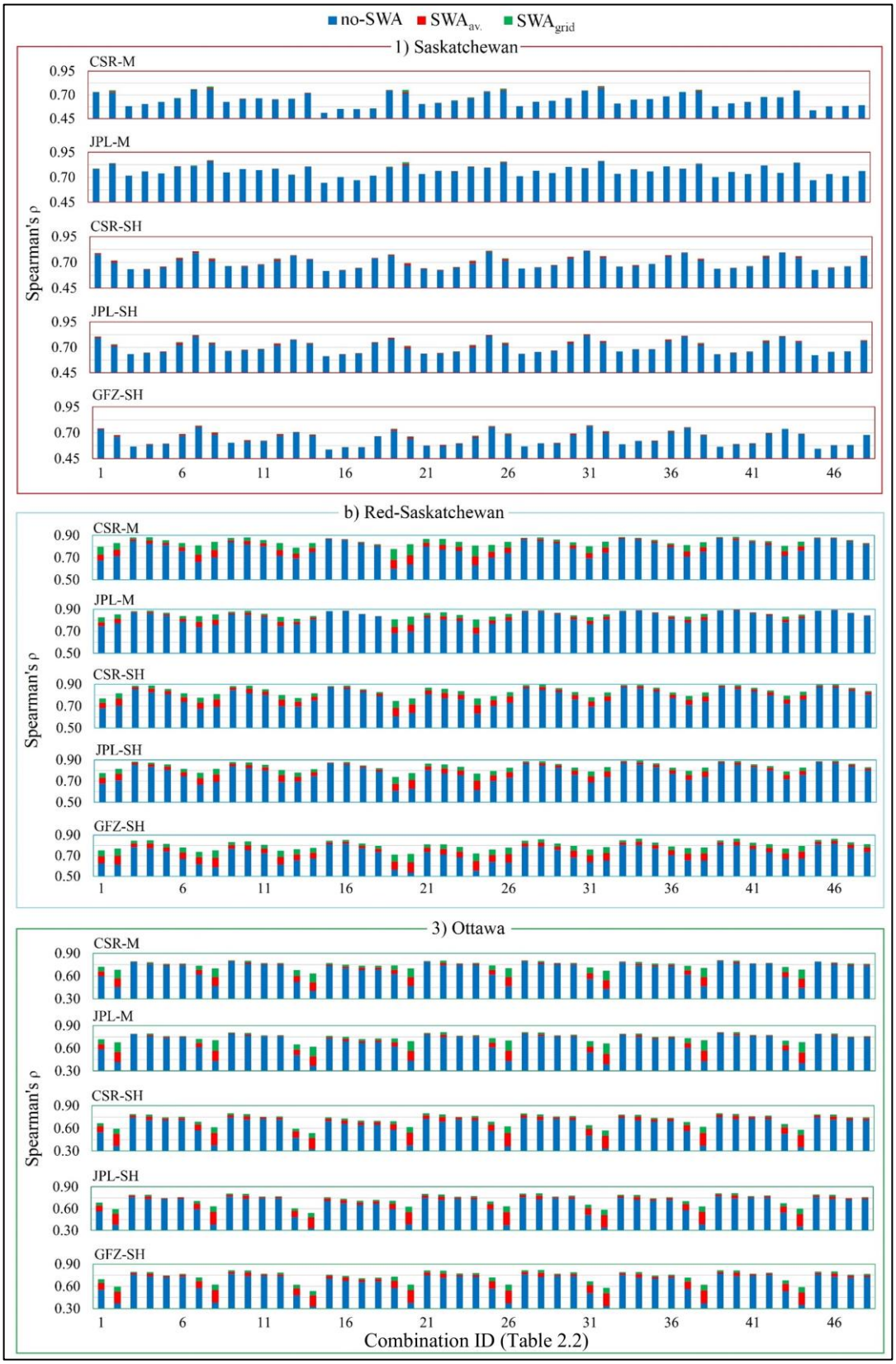


Figure 2.14 Comparisons of 48 unique estimations of GRACE-GWA (Table 2.4) and normalised in-situ groundwater anomaly before and after accounting of SWA using Spearman's rank correlation coefficient for (1) Saskatchewan, (2) Red-Saskatchewan, and (3) Ottawa basins. GRACE-TWSA were derived from CSR-M, JPL-M, CSR-SH, JPL-SH, and GFZ-SH

solutions. Blue bars denote correlation values between GRACE-GWA estimates and normalised in-situ groundwater anomaly before accounting of SWA (no-SWA). Red and green bars signify the enhancement in correlation attributed to accounting for surface water influence, employing basin average approach ($SWA_{av.}$) and grid approach (SWA_{grid}), respectively. All correlations were significant ($p < 0.05$).

2.3.4 Variability in GRACE-GWA estimates across the additional basins

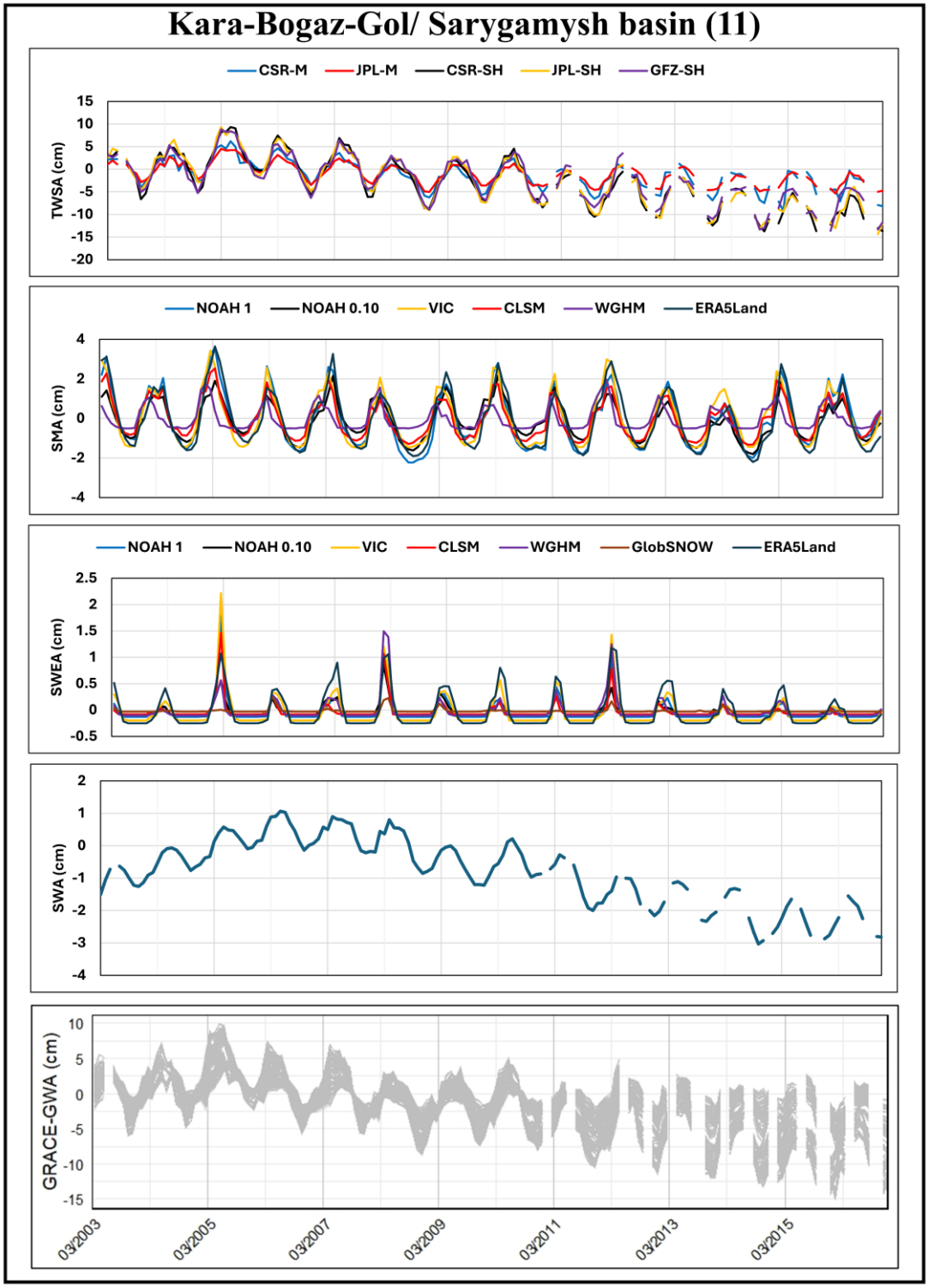


Figure 2.15 Monthly time series from 2003 to 2016 over the Kara-Bogaz-Gol/Sarygamysh basin (Basin 11 in Figure 2.1), illustrating: (i) GRACE-derived Terrestrial Water Storage Anomalies (TWSA) from CSR-M, JPL-M, CSR-SH, JPL-SH, and GFZ-SH solutions; (ii) Soil Moisture Anomalies (SMA) from NOAH (1˚ and 0.1˚), VIC (1˚), CLSM (1˚), WGHM (0.5˚),

and ERA5-Land (0.1˚); (iii) Snow Water Equivalent Anomalies (SWEA) from NOAH (1˚ and 0.1˚), VIC (1˚), CLSM (1˚), HGSWE (0.25˚), WGHM (0.5˚), GlobSnow (0.25˚), and ERA5-Land (0.1˚); (iv) Surface Water Anomalies (SWA); and (v) 630 GRACE-based Groundwater Storage Anomaly (GRACE-GWA) realisations derived from multi-model combinations of water budget components. For the complete set of 12 additional basins, refer to Figures A.17–A.28 in Appendix A.

To improve the generalisability and applicability of multi-model GRACE-GWA framework, the analysis was extended beyond the initial three Canadian–US basins to an additional twelve basins (Basins 4–15 in Figure 2.1), representing a wide range of hydroclimatic and geographic conditions. While the initial basins were selected due to their hydrological complexity, characterised by significant snow accumulation and substantial surface water influence, they also benefit from relatively high data availability, allowing for more robust benchmarking of methodological choices. Recognising that many regions of interest for GRACE-based groundwater monitoring are data-scarce and hydrologically distinct, this extension was not intended as a direct transfer of conclusions, but rather as a means to stress-test the framework across varied environmental settings. The additional basins encompass hyper-arid, arid and semi-arid regions, basins with minimal snow or surface water contributions, and those with differing levels of model uncertainty in soil moisture and snow water equivalent datasets.

The water budget framework facilitated the generation of 630 GRACE-GWA estimates for basins numbered 4 to 11 (Figure 2.15 and Figures A.17-A.24 in Appendix A). These combinations applied 5 GRACE-TWSA solutions, 7 SWEA datasets, 6 SMA datasets, and 3 SWA scenarios, as detailed in Table A.9 and Figures A.17 to A.24 in Appendix A. In parallel, 90 GRACE-GWA estimates were derived for basins numbered 12 to 15 (Table A.10 and Figures A.25 to A.28 in Appendix A), where the absence of snowpack led to the exclusion of the SWEA component from the water budget equation (Equation 2.1). Since in-situ groundwater data was unavailable for these basins, no comparisons were conducted between GRACE-GWA estimates and basin-scale in-situ groundwater observations. Instead, a cross-correlation matrix was employed between GRACE-GWA estimates, which allowed us to reveal variability existing among these estimates (Figures A.29 to A.40 in Appendix A).

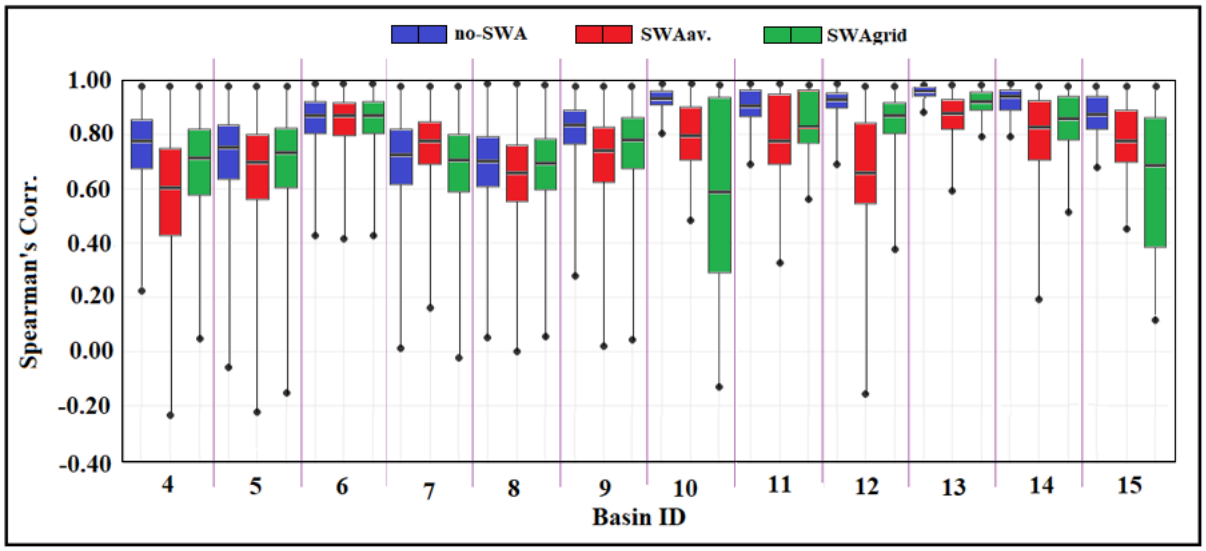


Figure 2.16 Box and whisker plots summarizing cross-correlation results for GRACE-GWA estimates across the additional basins (numbered from 4 to 15 as shown in Figure 2.1).

Box and whisker plots depicting results of the cross-correlation analysis reveal wide correlation magnitudes among GRACE-GWA estimates (Figure 2.16). In cases where SWA is ignored (i.e., no-SWA), the absence of SWEA signal in the African basins (numbered 12 to 15 in Figure 2.1), or the presence of low SWEA amplitudes in basins numbered 10 and 11, mitigated the effect of model structural bias. As a result, a decrease in the variability among GRACE-GWA estimates is noted when compared to the remaining basins (Figure 2.16).

Inclusion of SWA in the water budget equation expanded the range of correlation magnitudes among GRACE-GWA estimates. The application of different SWA accounting approaches ($SWA_{av.}$, SWA_{grid}) demonstrated varying impacts on the correlation magnitudes between GRACE-GWA estimates (Figure 2.16). The location of lake storage mass change within the basin played a crucial role in defining the distinctions between $SWA_{av.}$ and SWA_{grid} when linking lake storage to SWA in Equation 2.1 (as described in section 2.2.3). Noteworthy, the most substantial effect of incorporating SWA within the water budget equation was observed in basin 12 (Victoria basin), driven by significant volume variations in Lake Victoria. Conversely, the smallest influence of SWA on GRACE-GWA estimates was noticed in basin 6, primarily due to its minimal lake volume variations. The findings document that integration of SWA into the water budget equation may increase the range of variability among GRACE-GWA estimates (Figure 2.16).

2.4 Discussion

GRACE has been heralded as a key tool to characterize large-scale changes in groundwater storage (Famiglietti, 2014). In the clamour to use GRACE to assess groundwater storage changes, it is sometimes overlooked that extraction of GRACE-GWA from GRACE-TWSA is

troublesome despite validation studies (e.g., Yeh et al., 2006; Rodell et al., 2007; Scanlon et al., 2012; Feng et al., 2013; Castle et al., 2014). Early GRACE studies applied large-scale (1-deg) GLDAS output to account for water budget components to extract GRACE-GWA, employing an averaging approach to level out recognised inconsistencies in model representations of water budget components (e.g., Rodell et al., 2009; Shamsudduha et al., 2012; Voss et al., 2013; Feng et al., 2013; Castle et al., 2014; Nanteza et al., 2016; Huang et al., 2016). Although the averaging approach provides a method to estimate uncertainty by calculating the variance of water budget sums, this study followed an experimental design whereby individual model output was applied to account for a water budget component, thus allowing for numerous realisations of GRACE-GWA using Equation 2.1. Results document the influence of model water partitioning differences and model structural bias to isolate GRACE-GWA (Figures 2.8 and 2.9). Comparison of GRACE-GWA and normalised in-situ groundwater anomalies clearly document a range of correlation magnitudes, suggesting that the use of Equation 2.1 conflicts both the timing of groundwater dynamics and the seasonal variabilities in groundwater storage. Notably, combinations of water budget components were found to exhibit a pronounced inability to faithfully capture observed groundwater changes (Figure 2.14).

Various combinations of SWEA and SMA produced a range of bias as compared to 1-deg GLDAS LSM output (Figure 2.10). In the extreme case, a bias of -135% was documented, meaning that comparative differences due only to water budget data selection can equate to alarmingly high unreliability. Amplitude ratios exhibited similar patterns (Figure 2.9). Although not quantified here, amplitude disagreements in water budgets influence the magnitude of storage change assessments, particularly with respect to trend estimation (Scanlon et al., 2016). GRACE users adhere to the message that the lack of global data requires using output from hydrologic models and LSMs to account for water budget variables, and although users are aware of structural bias in how models partition water (Clark et al., 2008; Ruddell et al., 2019; Konapala et al., 2020), there is a general devotion to using a select few model outputs. Recent studies have used various model output to assess groundwater storage trends (Rateb et al., 2020; Scanlon et al., 2021) but have ignored the monthly storage variations that may capture short-term changes driven by climate and anthropogenic use (Thomas & Famiglietti, 2019; Thomas & Nanteza, 2023). Additionally, by considering the range of outcomes in extraction of GRACE-GWA from different combinations of water budget components, the analysis captures the heterogeneity in the outcomes, each of captures legitimate adherence to Equation 2.1. In this analysis, comparison between various GRACE-GWA time series have applied nonparametric correlation and thus informs of ranked correlation but fails to capture seasonal

differences and amplitudes. The application of the amplitude ratio further resolves the influence of water budget components, assessing the efficacy of applying Equation 2.1. Comparisons for the Ottawa basin (Figure 2.9(3)) highlight extreme amplitudes magnitudes, with seasonal variability between SWEA+SMA resulting in amplitude ratios of up to 4, meaning that combinations of SWEA+SMA result in amplitude differences of 4x that of other combinations.

A comprehensive analysis comparing 720 GRACE-GWA estimates spanning three basins, namely, Saskatchewan, Red Saskatchewan, and Ottawa, documented a considerable range of groundwater changes (Figures 2.13 and 2.14). Differences in seasonality and amplitudes in water budget components from assorted auxiliary datasets were found to affect dynamic change in GRACE-GWA, corroborating previous studies (Akhtar et al., 2022; Li & Wang, 2022; Thomas et al., 2017a). The wide-ranging variability underscored the disparities in the capability of GRACE-GWA estimates to accurately capture the true behaviour of groundwater resources, with certain estimates falling short in adequately depicting observed dynamic behaviour (Figure 2.14).

The lack of a clear framework to account for SWA in Equation 2.1 has resulted in a wide range of assumptions and methods to convert surface water (i.e., lakes and reservoir) storage to SWA (Shamsudduha et al., 2012; Sultan et al., 2014; Nanteza et al., 2016; Agutu et al., 2019; Ramjeawon et al., 2022). Assuming reduced importance of SWA in Equation 2.1 has led researchers to frequently omit or inaccurately account for this factor in the water budget equation. This is despite the acknowledged influence of SWA on GRACE-GWA estimates (Kim et al., 2009; Han et al., 2009; Shamsudduha et al., 2012; Zhou et al., 2016; Getirana et al., 2017). Theoretically, this results in estimating available water (AW; Castle et al., 2014; Thomas & Nanteza, 2023), i.e., a time series representing groundwater and surface water storage, instead of solely groundwater. The strict adherence to Equation 2.1 means that ignoring SWA leads to inappropriate groundwater interpretations. Global datasets providing lake bathymetry and surface water elevation changes provide new ways to account for SWA (Yigzaw et al., 2018; Li et al., 2019b; Khazaei et al., 2022; Busker et al., 2019; Tortini et al., 2020; Hou et al., 2022). Although explicit statements regarding statistical significance change while accounting for surface water storage change cannot be made, inclusion of surface water storage to account for SWA incrementally increased correlation for most cases considered. Results document that the ideal estimate of GRACE-GWA should thus account for lake and reservoir change, even if only incremental improvements in correlation were noted in this study. These incremental improvements demonstrate an enhanced capacity of GRACE-GWA in accurately capturing observed groundwater dynamics. The findings add to the numerous studies documenting the

importance of surface water accounting to isolate a groundwater signal (Scanlon et al., 2012; Shamsudduha et al., 2012; Longuevergne et al., 2013; Nanteza et al., 2016).

A systematic analysis of an additional 12 basins, each showcasing a diverse range of hydrological characteristics and aridity, underscored the inconsistency in GRACE-GWA retrievals. Cross-correlation results for GRACE-GWA revealed considerable variability among the realisations (Figure 2.16). Notably, the incorporation of SWA within the water balance equation (Equation 2.1) resulted in reduced cross-correlation values, signifying that time series integrating SWA increase precision in GRACE-GWA realizations. This extended basin analysis provided a valuable opportunity to examine the sensitivity of GRACE-GWA estimates to different combinations of input datasets under varying hydroclimatic contexts. Rather than assuming uniform applicability, the inclusion of these diverse basins allowed for a more rigorous evaluation of the framework's robustness and limitations, ultimately contributing to a more globally transferable and context-aware approach for large-scale groundwater monitoring using GRACE.

The application of this comparative study was restrained by limitations. As if often the case, a lack of storativity data limited estimation of observed groundwater storage anomalies for direct comparison to GRACE-GWA. Instead, the in-situ groundwater time series was normalised to capture seasonal changes. Further research to address model output uncertainty within a multi-model approach to extract GRACE-GWA is warranted, as this study focused on single model output contributions to water budget components. Spearman's ρ simply measures the strength of the monotonic relationship between two variables and thus fails to account for important GRACE-GWA signals including seasonal amplitudes. By applying an amplitude metric, the amplitude ratio (Equations 2.4 and 2.5), this study shows how combinations of water budget components can result in inaccurate or misleading groundwater storage loss, potentially resulting in misleading decisions about groundwater use and conservation. Lastly, limitations in surface water representation persist in several of the study basins, where lakes, reservoirs, and rivers remain poorly captured in existing datasets. This issue reflects a broader systemic challenge in global hydrological observation (Alsdorf et al., 2007; Busker et al., 2019; Tortini et al., 2020). As a result, GRACE-GWA estimates in these regions may retain residual signals linked to unresolved surface water variability (Castle et al., 2014; Thomas & Nanteza, 2023), complicating direct comparisons with basin-scale in-situ observations.

2.5 Conclusion

The successful achievement of groundwater sustainability is contingent upon the implementation of effective surveillance of groundwater resources (Morris et al., 2003; Giordano, 2009; Gleeson et al., 2020). Through effective surveillance, stakeholders can track changes in groundwater levels, identify potential threats to groundwater resources, and make informed decisions about the use of these resources. Model structural bias pertaining to how hydrologic models and land surface models vertically partition water regulated estimation of GRACE-GWA, while simple accounting methods to account for surface water were found to consistently increase correlations with respect to basin in-situ groundwater time series (Figure 2.14). The wide range of variability observed in GRACE-GWA underscores the uncertainty in the current state of knowledge within groundwater resources (Figures 2.13, 2.14, and 2.15). These findings underscore the paramount importance of relying on accurate multi-model output. This precision is not only a nicety but a necessity in elevating the reliability and robustness of GRACE-GWA estimates, thereby advancing the understanding of groundwater dynamics with unparalleled accuracy.

This study not only contributes substantially to a comprehensive understanding of the challenges inherent in isolating GRACE-GWA from GRACE-TWSA using the water balance equation (Equation 2.1) but also advocates for a comprehensive appraisal of water budget components in GRACE groundwater studies. Such a holistic approach promises to yield reliable and trustworthy estimates of GRACE-GWA, providing a solid foundation for future assessments. Additionally, the findings presented here serve as a catalyst for revisiting and advancing critical GRACE-GWA studies, particularly those focused on basins experiencing groundwater depletion. This re-examination aims a nuanced assessment of groundwater storage dynamics, crucial for effectively mitigating challenges associated with this vital resource.

Chapter 3. A Multi-objective Comparative Framework for Enhanced GRACE-Groundwater Comparative Analysis

This chapter is based on a manuscript submitted for publication and currently under peer review by Journal of Hydrology (JH).

Chapter abstract: Accurate monitoring of groundwater resources is essential for sustainable water management, especially under escalating pressures from climate variability and intensive human activities. Despite significant advancements provided by the Gravity Recovery and Climate Experiment (GRACE) satellites in monitoring terrestrial water storage anomalies (GRACE-TWSA), isolating representative groundwater signals (GRACE-GWA) remains challenging primarily due to uncertainties in estimating complementary water budget components. While multi-model approaches to deriving GRACE-GWA can account for these uncertainties, systematic frameworks to objectively compare and constrain multi-model realisations against observed groundwater data remain scarce. To address this gap, this study applies a multi-objective comparative framework employing Nash–Sutcliffe Efficiency (NSE) and Kling-Gupta Efficiency (KGE) metrics to compare multi-model GRACE-GWA realisations against in-situ basin-scale groundwater anomalies. Unlike conventional correlation-based approaches, this framework captures critical aspects of time series similarity, including seasonal amplitude fidelity and magnitude consistency, thus enabling clearer identification of optimal groundwater storage realizations. The findings reveal significant uncertainty between multi-model groundwater storage trend and seasonal amplitude, emphasizing critical limitations often overlooked in standard GRACE-GWA assessments. By systematically isolating the most hydrologically consistent realizations, this framework significantly enhances the reliability, interpretability, and applicability of GRACE-based groundwater estimates. This methodological framework supports more accurate groundwater monitoring, strengthens data-driven decision-making processes, and ultimately contributes toward ensuring the long-term sustainability and resilience of vital groundwater resources.

3.1 Introduction

Aquifer exploitation threatens the resilience of groundwater resources, prompting groundwater depletion (Famiglietti, 2014), land subsidence (Herrera-García et al., 2021), and irreversible environmental impacts (Kundzewicz et al., 2008). The inconspicuous nature of groundwater allows it to evade observation, particularly due to the lack of consistent in-situ monitoring (Rodell et al., 2007; Lall et al., 2020; Condon et al., 2021; Jasechko et al., 2024). Gravimetry missions including the Gravity Recovery and Climate Experiment (GRACE) have enriched

monitoring of Earth's water resources by measuring temporal gravity variations (Tapley et al., 2004), enabling the identification of integrated changes in terrestrial water storage (GRACE-TWSA; Syed et al., 2008). Extracting a time series attributable solely to groundwater storage (GRACE-GWA) from GRACE-TWSA remains a challenge, due in part to the lack of direct observations of other water budget components (Rodell & Famiglietti, 2002; Saxe et al., 2021; Akl & Thomas, 2024; see also Chapter 2 of this thesis), yet has proven to be a useful tool to assess large-scale groundwater change (Feng et al., 2013; Richey et al., 2015a; Nanteza et al., 2016; Thomas et al., 2017b). Remote sensing surveillance promotes the generation of groundwater system understanding as the resource responds to changing water cycle and human water use (Rodell et al., 2009; Strassberg et al., 2009; Richey et al., 2015a; Nanteza et al., 2016; Thomas & Famiglietti, 2019) but requires the notion that the GRACE-GWA signal truthfully tracks temporal variations in groundwater. This notion, however, requires us to ask, "What is truthful?".

Evaluating GRACE-GWA against in-situ groundwater observations provides a foundation for groundwater surveillance in regions lacking suitable direct observations. A comprehensive literature review examined GRACE-GWA validation, citing over 90 relevant studies (Yeh et al., 2006; Swenson et al., 2006; Strassberg et al., 2007; Rodell et al., 2007; Sun et al., 2010; Henry et al., 2011; Shamsudduha et al., 2012; Scanlon et al., 2012; Feng et al., 2013; Castle et al., 2014; Shen et al., 2015; Xiao et al., 2015; Nanteza et al., 2016; Huang et al., 2016; Long et al., 2016; Feng et al., 2018; Chen et al., 2019; Rateb et al., 2020; Skaskevych et al., 2020; Chen et al., 2020; Liu et al., 2021; Xu et al., 2021; Han et al., 2021; Ouatiki et al., 2022; Amiri et al., 2023; Alghafli et al., 2023; Huang et al., 2023b; Nenweli et al., 2024; Rzepecka et al., 2024; Yang et al., 2024; [complete list available in Table B.1 in Appendix B]). Notably, 59 studies (65%) relied exclusively on correlation metrics, including the Pearson correlation coefficient or Spearman's ranked coefficient (ρ), to assess goodness-of-fit (GOF) between extracted GRACE-GWA time series and representative in-situ groundwater change. Despite their simplicity and wide application in the hydrologic sciences, these correlation metrics overlook additive and proportional differences, and disregard complex patterns, temporal shifts, and amplitude variations (Legates & McCabe, 1999; Moore et al., 2006). Simple correlation metrics are sensitive to outliers which skew results towards extreme events (Legates & Davis, 1997; Legates & McCabe, 1999). The neglect of these crucial time series characteristics triggers bias in the comparison efficacy, provoking consequential challenges in truthful tracking of groundwater change (Alley & Konikow, 2015).

The insufficient evaluation of GRACE-GWA comparison beyond simple correlation metrics has lagged the hydrologic modelling community, where a rich literature with numerous alternative GOF metrics is applied (Nash & Sutcliffe, 1970; Gupta et al., 2009; Althoff et al., 2021; Mathevet et al., 2023). The pervasive application of GRACE-GWA to assess large-scale groundwater change (e.g., Jin & Feng, 2013; Richey et al., 2015a; Chen et al., 2016a; Thomas et al., 2017b; Frappart & Ramillien, 2018) warrants deeper study of groundwater time series estimation from GRACE. This study employs multiple-objective functions, namely the Nash–Sutcliffe Efficiency (NSE; Nash & Sutcliffe, 1970) and the Kling-Gupta Efficiency (KGE; Gupta et al., 2009), alongside correlation metrics, to evaluate the consistency between multi-model GRACE-GWA and in-situ groundwater observations. This study aims to elevate the capacity of GRACE-GWA comparisons by utilizing multi-objective functions that effectively capture critical time series similarities. The analysis avoids prescribing preferred outcomes for GRACE-GWA and instead focus on the broadened understanding of groundwater change that could be gained by performing multi-model/multi-objective function studies for water resources assessments.

Robust comparison assessment should exploit embedded time series information to quantify synchrony between multi-model GRACE-GWA and in-situ groundwater observations. Applying GOF metrics for GRACE-GWA corroboration enables the capture of crucial time series characteristics, including seasonal patterns, amplitudes, temporal shifts, and extreme values (Gupta et al., 2009; Althoff et al., 2021). A broader evaluation of GRACE-GWA, employing NSE and KGE metrics which capture elements of comparison beyond measures of agreement in simply increasing or decreasing, would theoretically improve analysis of the efficacy of large-scale groundwater measurement. However, these metrics are not without challenges (Gupta et al., 2009; Pool et al., 2018; Knoben et al., 2019). NSE, a normalised version of Mean Square Error (MSE) using the mean of observations as a baseline, tends to underestimate variability in time series comparisons (Gupta et al., 2009; Kling et al., 2012; Pool et al., 2018; Lamontagne et al., 2020). KGE decomposes NSE into constitutive components (correlation, variability, and mean bias), potentially introducing bias for skewed data and correlation between variability and mean bias within its framework (Kling et al., 2012; Pool et al., 2018; Lamontagne et al., 2020). Despite these challenges, the multi-objective nature of GOF metrics has the potential to capture essential aspects of correspondence that were obscured by simple correlation.

This study introduces a comparative framework to capture the multidimensional nature of groundwater time series behaviour. The adoption of NSE and KGE responds to the need for

performance metrics that account for not only co-variation but also the structural, temporal, and statistical properties of the signals being compared (Nash & Sutcliffe, 1970). NSE evaluates how well GRACE-GWA reproduces the variance observed in in-situ groundwater anomalies, thereby detecting mismatches in amplitude and overall variability. In contrast, KGE offers a more integrated perspective by disaggregating performance into three diagnostic components: correlation (timing and synchrony), bias (systematic offset), and variability (relative scaling) (Gupta et al., 2009). This structure enables the framework to identify and quantify complex discrepancies between GRACE-derived and observed time series, such as phase shifts, amplitude attenuation, or mean displacement, that are often masked by simple correlation metrics (Legates & Davis, 1997; Legates & McCabe, 1999). Given that GRACE-GWA signals are susceptible to temporal lags and smoothed hydrologic responses, especially in deeper or slower-responding aquifer systems, such diagnostic resolution is critical. By jointly applying NSE and KGE, the framework advances beyond traditional assessments of directional agreement to deliver a more nuanced and hydrologically meaningful evaluation of GRACE-GWA performance across diverse groundwater settings.

This study suggests that knowledge gained employing a multi-model, multi-objective framework improves the capacity to designate GRACE-GWA time series that faithfully reflect in-situ groundwater changes, thus aiding the development of water management schemes.

3.2 Data and Methods

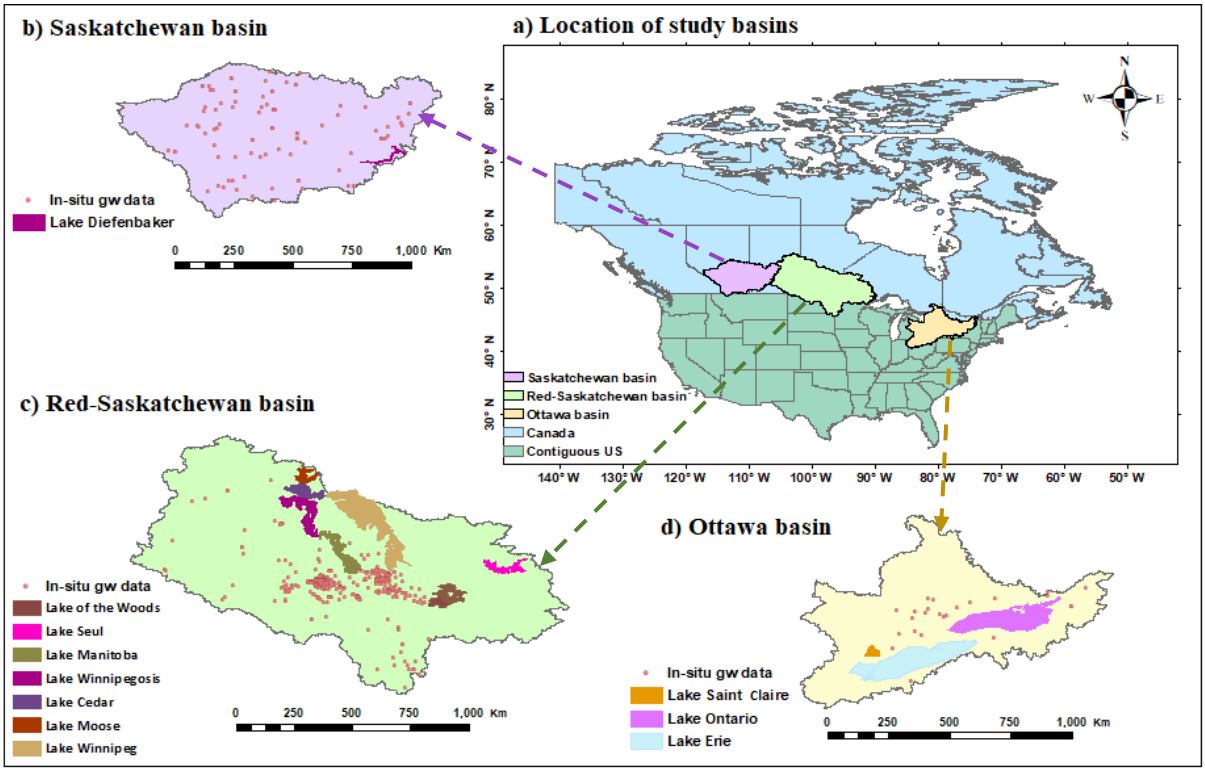


Figure 3.1 Map depicting (a) study basins with distribution patterns of in-situ groundwater (gw) wells and surface waters in (b) the Saskatchewan, (c) Red-Saskatchewan, and (d) Ottawa basins.

Three basins spanning the Canadian-US border were selected given the availability of in-situ groundwater observations and appreciable water budget components, including seasonal snowpack. In-situ groundwater observations were collected from the Groundwater Information Network (GIN) (https://gin.gw-info.net/service/api_ngwds:gin2/en/gin.html) and the National Groundwater Monitoring Network (NGWMN) (https://cida.usgs.gov/ngwmn/index.jsp). Pre-processing steps first applied a data quality algorithm to reject wells with more than one missing seasonal observation per calendar year. Selected wells, exhibiting different measurement frequencies, underwent gap filling using spatial and temporal methods (Zeng & Levy, 1995) to generate monthly datasets of observed groundwater head (Nanteza et al., 2016; Thomas et al., 2017a; Akl & Thomas, 2024; see also Chapter 2 of this thesis). A total of 95, 317, and 21 observation wells in the Saskatchewan, Red-Saskatchewan, and Ottawa basins, respectively (Figure 3.1), were used to generate basin-scale groundwater change for comparison to GRACE-GWA.

Various spatial interpolation methods were utilized to upscale observation well datasets to a broader basin scale. Spatial interpolation experiments applied Thiessen Polygons (TP), Inverse Distance Weighting (IDW), Universal Kriging (UK), Ordinary Kriging (OK), Ordinary Kriging-Log Transform (OK-Log), Multi Linear Regression (MLR), and Spatio-Temporal

Kriging (STK) (Brassel & Reif, 1979; Oliver et al., 1990; Lu et al., 2008; Li & Heap, 2014; Gräler et al., 2016). Accuracy assessment via leave-one-out cross-validation (LOOCV; Van Zoest et al., 2020; Ferreira et al., 2023) confirmed the robustness of groundwater observations against various spatial interpolation methods (Figures B.1–B.7 in Appendix B). TP was selected to generate basin-scale groundwater observation time series (Nanteza et al., 2016; Thomas et al., 2017a; Akl & Thomas, 2024; see also Chapter 2 of this thesis), which were subsequently transformed into in-situ anomaly time series by removing the mean for the period January 2004 to December 2009.

Basin-scale in-situ groundwater anomalies represent anomaly groundwater head change, not groundwater storage changes, due to the lack of storativity data. Previous GRACE-GWA studies have applied a range of storativity values estimated from geologic characteristics (Bhanja et al., 2016; Li et al., 2019c; Scanlon et al., 2012; Rodell et al., 2007; Chen et al., 2016a, 2019; Skaskevych et al., 2020; Yeh et al., 2006; Long et al., 2016; Strassberg et al., 2009; Thomas et al., 2017a), generally applying specific yield (Sy)=0.10. GRACE-GWA studies across Canada have applied generally larger estimates of Sy, ranging from 0.03-0.44, attributed to the variability in hydrogeologic parameters in glacial deposits (Huang et al., 2016; Hachborn et al., 2017; Bhanja et al., 2018; Wang et al., 2022b). The PCRater GLOBal Water Balance model (PCR-GLOBWB; Sutanudjaja et al., 2018) offers estimated hydrogeologic parameters at 5 arcmin spatial resolution, with basin average Sy of 0.04, 0.06, 0.07 for Saskatchewan, Red-Saskatchewan, and Ottawa, respectively. In this study, a range of Sy of 0.10, 0.15, and 0.30 was applied across study basins and PCR-GLOBWB-derived Sy was examined, with detailed results provided in Appendix B. These broader Sy values were selected to cover the spectrum of storativity conditions observed in prior investigations, accounting for the uncertainty introduced by variations in glacial geology. Additionally, a normalised anomaly time series was employed to bypass the lack of storativity (i.e., Sy for unconfined aquifer) estimates (Figures B.8 in Appendix B) (Castle et al., 2014; Akl & Thomas, 2024; see also Chapter 2 of this thesis).

3.2.1 Extracting GRACE-GWA from GRACE observations

Datasets	Source	Spatial resolution	Reference
GRACE-TWSA	CSR-SH	1.00˚	Landerer & Swenson, 2012
	JPL-SH		
	GFZ-SH		
	CSR-M	0.25˚	Save et al., 2016
	JPL-M	0.50˚	Watkins et al., 2015
SWEA	NOAH	1.00˚	Koren et al., 1999
		0.10˚	McNally et al., 2017
	VIC	1.00˚	Liang et al., 1996
	CLSM	1.00˚	Li et al., 2019a
	WGHM	0.50˚	Müller Schmied et al., 2021
	GlobSnow	0.25˚	Luojus et al., 2021
	HGSWE	0.25˚	Mudryk et al., 2015
	ERA5-Land	0.10˚	Muñoz-Sabater et al., 2021
SMA	NOAH	1.00˚	Koren et al., 1999
		0.10˚	McNally et al., 2017
	VIC	1.00˚	Liang et al., 1996
	CLSM	1.00˚	Li et al., 2019a
	WGHM	0.50˚	Müller Schmied et al., 2021
	ERA5-Land	0.10˚	Muñoz-Sabater et al., 2021
SWA	lake/reservoir water level	-	https://wateroffice.ec.gc.ca/
	lakes bathymetry datasets	-	Khazaei et al., 2022

Table 3.1 Summary of GRACE-TWSA and water budget components datasets utilized in this study.

This study utilized five GRACE-TWSA datasets (Table 3.1), including three spherical harmonic (SH) solutions—CSR-RL06 V04, JPL-RL06 V04, and GFZ-RL06 V04— (Landerer & Swenson, 2012) and two mass concentration (mascon) solutions: CSR-RL06 V02 (Wiese et al., 2016) and JPL-RL06 V02 (Watkins et al., 2015). The JPL-M solution was applied without scaling factors, recognizing limitations of the Community Land Model (CLM) in accurately simulating snow processes, a dominant factor in study basins (Wiese et al., 2016; Akl & Thomas, 2024; see also Chapter 2 of this thesis). The analysis covered March 2003 to December 2016, excluding 17 months of missing GRACE data.

GRACE-TWSA was decomposed into individual water storage components using a water balance equation (Equation 3.1), which represents terrestrial water storage as the sum of groundwater storage anomalies (GRACE-GWA), soil moisture anomalies (SMA), snow water equivalent anomalies (SWEA), and surface water anomalies (SWA):

$$\text{GRACE-TWSA}_t = \text{GRACE-GWA}_t + \text{SMA}_t + \text{SWEA}_t + \text{SWA}_t \quad (3.1)$$

where the subscript ($_t$) represents time (Rodell & Famiglietti, 2002).

Soil moisture anomalies (SMA) were derived from six datasets (Table 3.1), including three models from the NASA Global Land Data Assimilation System (GLDAS; Rodell et al., 2004): NOAH, Variable Infiltration Capacity (VIC) and Catchment Land Surface Model (CLSM), each at a resolution of 1˚×1˚. Additional sources included NOAH at 0.10˚×0.10˚ from the Famine Early Warning System Network (FEWS NET) Land Data Assimilation System (FLDAS; McNally et al., 2017), WaterGAP Global Hydrology Model (WGHM; Müller Schmied et al., 2021) at 0.50˚×0.50˚, and ERA5-Land at 0.10˚×0.10˚ (Muñoz-Sabater et al., 2021). While WGHM simulates monthly soil moisture in a single-layer representing the effective root zone (Müller Schmied et al., 2021), the other models (GLDAS, FLDAS, and ERA5-Land) simulate multi-layer soil moisture profiles. This study considered only the upper 1 m of the soil profile, as deeper layers frequently reach saturation conditions (Rodell & Famiglietti, 2001; Thomas et al., 2017a).

Snow water equivalent anomalies (SWEA) were obtained from eight datasets (Table 3.1). These included the six datasets utilized for SMA, supplemented by GlobSnow (Luojus et al., 2021), a remote sensing/observation-based dataset at 0.25˚×0.25˚ resolution, and the Historical Gridded Snow Water Equivalent and snow water fraction over Canada (HGSWE; Mudryk et al., 2015), which integrates remote sensing and land surface model outputs, also at 0.25˚×0.25˚ resolution.

Surface water storage anomalies (SWA) were estimated using lake water-level observations from the Canadian Water Office (https://wateroffice.ec.gc.ca) for 11 lakes across the Saskatchewan, Red-Saskatchewan, and Ottawa basins (Figure 3.1). Monthly average lake water storage estimates were calculated from the head-Area-Volume (*h*-*A*-*V*) relationships provided by the global lakes bathymetry datasets (GLOBathy; Khazaei et al., 2022).

To ensure consistency with GRACE processing methodologies (Lenk, 2013; Fatolazadeh et al., 2016; Nanteza et al., 2016), water budget components (i.e., soil moisture, snow and surface water) were processed in accordance with GRACE standards. For spherical harmonic (SH) solutions, water budget data were projected onto GRACE grids and underwent truncation at harmonic degree 60, destriping with a decorrelation filter (Swenson & Wahr, 2006) and smoothing using a 300 km radius Gaussian filter. Subsequently, anomalies were calculated by removing the mean of the time series from January 2004 to December 2009 and then subtracted from GRACE-TWSA grids. In contrast, for mascon solutions, potential leakage across GRACE grids was presumed to be effectively addressed during processing (Save et al., 2012; Watkins et al., 2015; Richey et al., 2015a; Bhanja et al., 2016; Scanlon et al., 2016; Wiese et al., 2016; Save et al., 2016; Thomas et al., 2017a, b; Rodell et al., 2018; Thomas & Famiglietti, 2019;

Neves et al., 2020; Wang et al., 2020b; Scanlon et al., 2021, 2022). Accordingly, water budget components for mascon solutions were converted to anomalies by subtracting mean values for January 2004 to December 2009 and subsequently subtracted from GRACE grids prior to aggregation at the basin scale. Both SH and mascon grids were spatially clipped to basin boundaries, preserving their original resolutions to ensure spatial integrity (Akl & Thomas, 2024; see also Chapter 2 of this thesis).

Integrating the five GRACE-TWSA solutions with eight SWEA, six SMA, and the SWA estimates (Table 3.1) resulted in a total of 240 GRACE-GWA realizations, computed using Equation 3.1 (Tables B.2 and B.3 in Appendix B).

3.2.2 GRACE-GWA time series analysis

This study analysed GRACE-GWA time series by applying trend analysis and seasonal amplitude estimation. Trend magnitude and significance were evaluated using the Seasonal Mann-Kendall (SMK) test and Sen's Slope estimator (Sen, 1968; Theil, 1950), both of which are nonparametric methods ideal for non-normal data (Helsel & Hirsch, 2002; Onyutha, 2016). The SMK test extends the traditional Mann-Kendall (MK) test (Mann, 1945; Kendall, 1955) by performing individual MK tests for each season, mitigating the impact of serial correlation and eliminating the need for pre-whitening techniques (Hirsch et al., 1982). The seasonal trend statistics are then aggregated into the overall S_k statistic:

$$S_k = \sum_{g=1}^{m} S_g \tag{3.2}$$

where m is the number of seasons, and S_g, the Kendall S statistic for each season, is calculated as:

$$S_g = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sgn(x_j - x_i) \tag{3.3}$$

where, x_j and x_i represent GRACE-GWA values at times j and i, respectively, while $sgn(x_j - x_i)$ is 1 if $x_j > x_i$, −1 if $x_j < x_i$ and 0 otherwise.

To quantify trend magnitude, the Sen's Slope estimator (b) was applied, calculated as the median of all pairwise slopes between data points:

$$b = median\left[\frac{Y_j - Y_i}{X_j - X_i}\right] \;\; \forall \; j > i \tag{3.4}$$

where Y represents GRACE-GWA values and X denotes the corresponding observation times.

The average amplitude of the GRACE-GWA time series was evaluated using an R package developed for estimating and predicting cosinor models for periodic data (Sachs, 2015). This approach fits a periodic model to the time series, considering both the time variable (t) and a fixed period (D). The periodic model (Y) is expressed as:

$$Y(t) = \alpha + A * \cos\left(2 * \pi * \frac{t}{D}\right) - \beta_1 + \varepsilon \quad (3.5)$$

Where α is the intercept; A is the time series amplitude; β_1 is the acrophase, indicating the timing of the peak within the cycle; and ε is the error term, assumed to have a mean of zero.

3.2.3 GRACE-GWA goodness-of-fit (GOF) metrics

This study focusses on the efficacy of alternative GOF metrics to reconcile GRACE-GWA and basin-scale in-situ groundwater anomalies. To conform to previous GRACE-GWA studies, Spearman's ρ (Zar, 1972) and Pearson (Cohen et al., 2009) correlation coefficients were first applied to quantify the strength and direction of association between time series. Spearman's ρ captures the monotonic relationship, while the Pearson correlation coefficient captures linear strength. In any case, the correlation metrics capture the similarity in the rise and fall of the time series, giving no information about irregular timing or amplitude disagreement (Legates and McCabe, 1999).

To overcome these limitations, a multi-objective evaluation framework was implemented, incorporating the NSE (Nash and Sutcliffe, 1970) and KGE (Gupta et al., 2009). These metrics provide a more holistic evaluation of agreement by accounting for both the magnitude and distributional characteristics of the compared datasets. NSE evaluates the proportion of observed variance that is captured by the derived GRACE-GWA and is defined as:

$$NSE = 1 - \frac{\sum_{t=1}^{T}(GW_{in-situ}^{t} - GW_{GRACE}^{t})^2}{\sum_{t=1}^{T}(GW_{in-situ}^{t} - \overline{GW_{in-situ}})^2}$$

Where, $GW_{in-situ}^{t}$ and GW_{GRACE}^{t} represent in-situ groundwater anomalies and GRACE-GWA at time t, and $\overline{GW_{in-situ}}$ is the mean of the in-situ groundwater anomalies. An NSE value of 1 indicates perfect agreement, while values below zero suggest that the in-situ groundwater anomalies mean outperforms the GRACE-GWA. NSE is sensitive to extreme values and may underperform with seasonal or anomaly-based data due to its reliance on the mean as a reference (Krause et al., 2005; Schaefli and Gupta, 2007; Gupta et al., 2009).

The KGE offers a more balanced evaluation, combining correlation, bias, and variability into a single metric:

$$KGE = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 - (\alpha-1)^2}$$

where, r is the Pearson correlation coefficient, $\beta = \mu_s/\mu_o$is the bias ratio, and $\alpha = \sigma_s/\sigma_o$is the variability ratio. Here, μ_s and μ_o are the means, and σ_s and σ_o are the standard deviations of the GRACE-GWA and in-situ groundwater anomalies, respectively. KGE enables a more interpretable decomposition of model performance, though it may still produce biased results when applied to normalized time series or datasets with small mean values, conditions often encountered in GRACE-GWA applications.

While both NSE and KGE aim to quantify model fidelity, their methodological emphases differ (Krause et al., 2005; Schaefli and Gupta, 2007; Gupta et al., 2009; Kling et al., 2012; Waseem et al., 2017; Pool et al., 2018; Knoben et al., 2019; Lamontagne et al., 2020; Althoff and Rodrigues, 2021). NSE provides a variance-based measure and is particularly sensitive to the overall spread and magnitude of deviations. As a result, it tends to underperform in datasets with strong seasonal components or low variability, characteristics typical of normalized groundwater anomaly records (Gupta et al., 2009). KGE, on the other hand, distributes performance assessment across multiple dimensions, yielding a more balanced and interpretable metric. Its inclusion of bias and variability terms makes it especially relevant for evaluating GRACE-GWA outputs, which may deviate systematically from in-situ observations due to scale mismatches or model structural uncertainty. However, KGE is not without limitations; it can produce misleading values when applied to time series with very small means or standard deviations (Pool et al., 2018; Knoben et al., 2019; Althoff and Rodrigues, 2021).

Incorporating both NSE and KGE within a multi-objective framework strengthens the robustness of the evaluation and captures a broader spectrum of performance characteristics than any single metric alone. This approach enhances interpretability and provides more comprehensive insight into the reliability of GRACE-GWA estimates across diverse hydroclimatic contexts.

3.3 Results and Discussions

3.3.1 Comparative analysis of correlation coefficients (Pearson, Spearman's ρ) and GOF (NSE, KGE) metrics in evaluating GRACE-GWA comparison

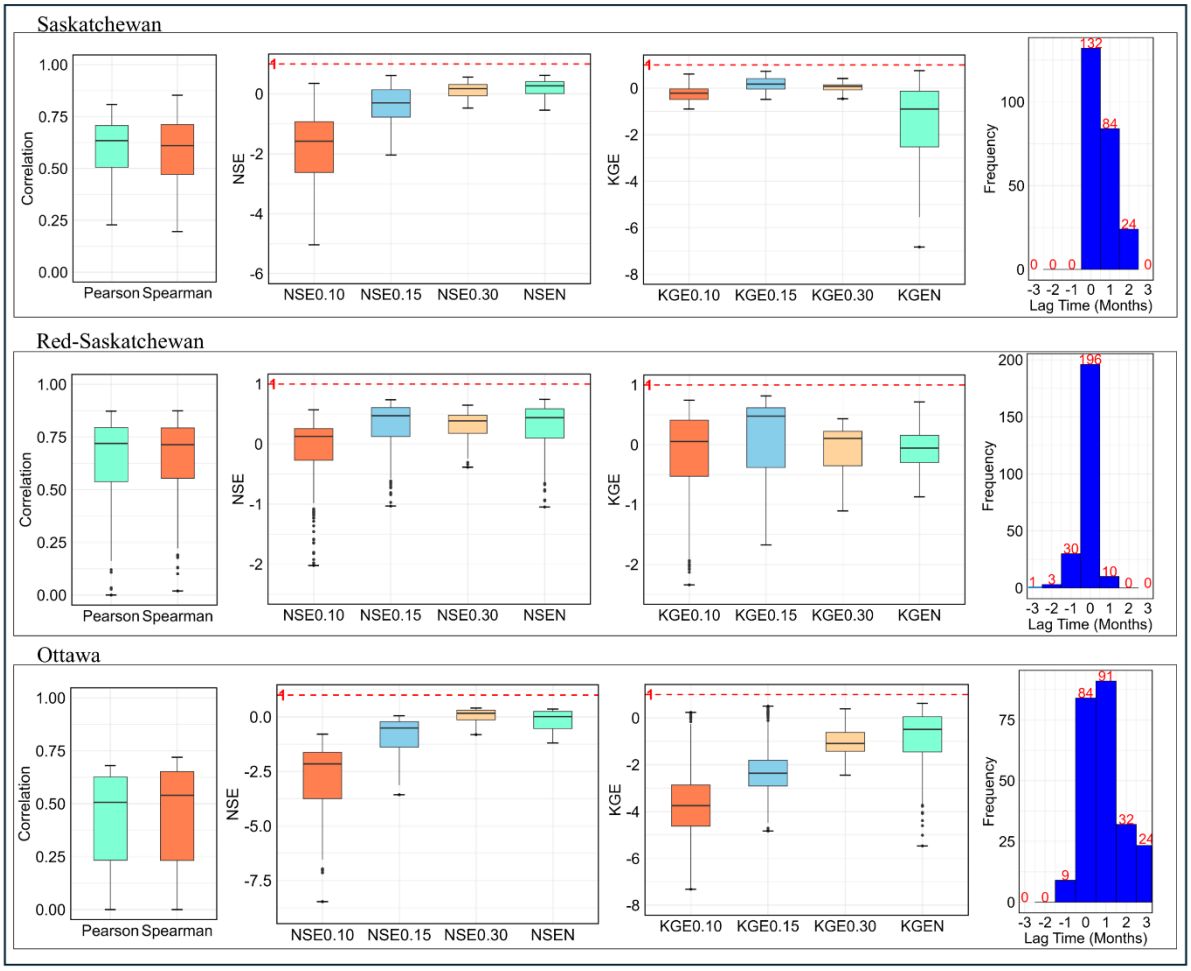


Figure 3.2 Box plots illustrating the distribution of correlation coefficients (Pearson and Spearman's ρ), GOF metrics (NSE and KGE), and lag time frequency (in Months) for 240 GRACE-GWA realizations, compared against basin-scale in-situ groundwater anomalies across the Saskatchewan, Red-Saskatchewan and Ottawa basins. Analysis includes normalised time series comparisons (denoted as NSEN and KGEN), as well as evaluations employing varying Sy values: 0.10, 0.15, and 0.30, represented as NSE0.10, NSE0.15, NSE0.30, and KGE0.10, KGE0.15, KGE0.30, respectively. Additional analyses based on PCR-GLOBWB model-derived Sy are provided in Figure B.9 in Appendix B.

Pearson cross-correlation revealed temporal discrepancies between GRACE-GWA and in-situ groundwater observations with lags of -3 to +3 months (Figure 3.2). Notably, the analysis identified 0, 5, and 33 instances of non-significant correlations (p-value > 0.05) for the Saskatchewan, Red-Saskatchewan, and Ottawa basins, respectively. Although both Pearson and Spearman's ρ correlation coefficients consistently exceeded 0.50 (Figure 3.2), regression analysis showed high correlation often coincided with deviations from the ideal gradient and intercept values (1 and 0, respectively). Conversely, comparisons with the highest NSE and

KGE metrics yielded gradients and intercepts much closer to the ideal, reflecting better alignment between GRACE-GWA and basin-scale in-situ groundwater anomalies.

NSE exhibited a systematic bias attributed to the in-situ time series mean (Figure 3.2) where the NSE magnitude is correlated to the applied Sy. NSE magnitude reflects performance in capturing temporal groundwater changes in comparison to the reference time series that would simulate groundwater changes superior to the mean in-situ groundwater observation. Thus, there is a bias toward GRACE-GWA time series that reflect dynamic amplitude changes. The bias is notable for normalised time series, where the reference model, $\overline{GW_{in-situ}}$, is equal to zero after normalisation.

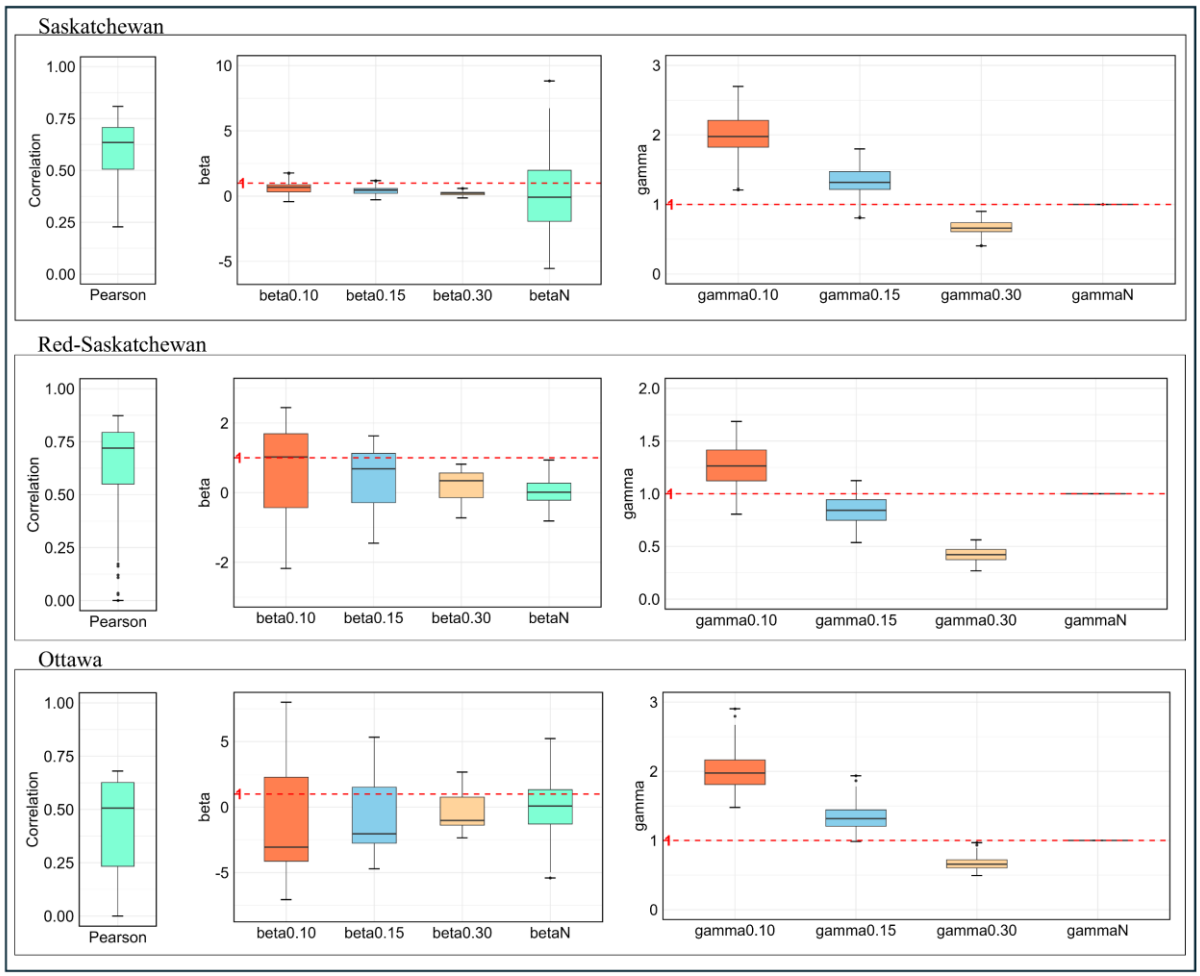


Figure 3.3 Box plots illustrating the distribution of KGE components including, correlation (Pearson), bias (beta), and variability (gamma) for 240 GRACE-GWA realizations, compared against basin-scale in-situ groundwater anomalies across the Saskatchewan, Red-Saskatchewan and Ottawa basins. Analysis includes normalised time series comparisons (denoted as betaN and gammaN), as well as evaluations employing varying Sy values: 0.10, 0.15, and 0.30, represented as beta0.10, beta0.15, beta0.30, and gamma0.10, gamma0.15, gamma0.30, respectively.

KGE demonstrated basin-specific patterns in relation to specific yield (Figure 3.2). Normalisation results in mean values near zero, which amplify mean bias in KGE (Figure 3.3). The influence of Sy on the observed time series is notable, due to the combination of variability and bias terms (Kling et al., 2012; Pool et al., 2018; Lamontagne et al., 2020). The effect was prominent in the Ottawa basin, where the KGE values, despite a few exceptions, were consistently below zero.

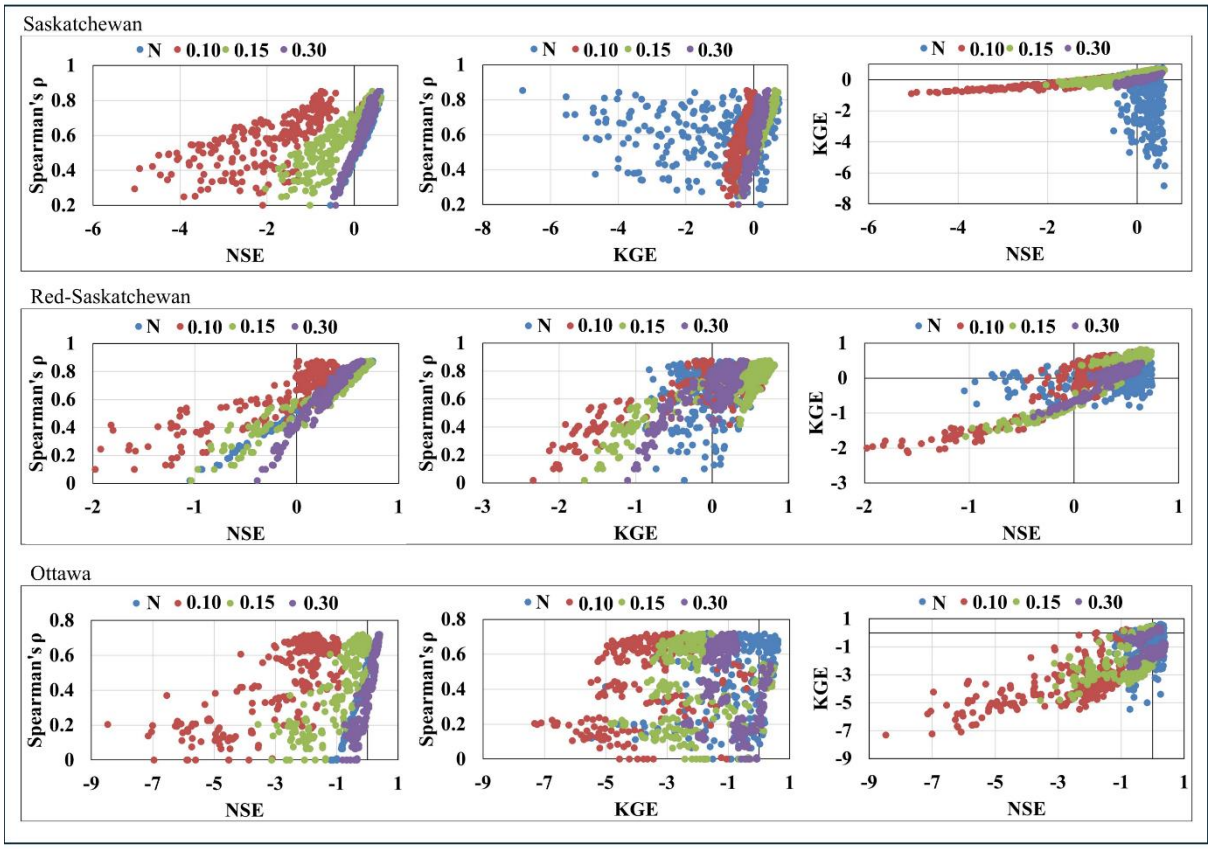


Figure 3.4 Scatter plots illustrating relationships among Spearman's ρ, NSE, and KGE for 240 GRACE-GWA comparisons against basin-scale in-situ groundwater anomalies across the Saskatchewan, Red-Saskatchewan, and Ottawa basins. Panels display Spearman's ρ versus NSE (left), Spearman's ρ versus KGE (middle), and NSE versus KGE (right). The analysis incorporates both normalised time series comparisons (denoted as N) and comparisons using Sy values of 0.10, 0.15, and 0.30, denoted as 0.10, 0.15, and 0.30, respectively. Additional comparisons utilizing PCR-GLOBWB model-derived Sy are provided in Figure B.10 in Appendix B.

Figure 3.4 reveals notable divergence between Spearman's ρ and the NSE/KGE metrics, highlighting inconsistencies in magnitude and variability that correlation alone cannot capture. The strength of association between ρ and NSE was sensitive to the Sy applied. The strongest relationship occurred with Sy = 0.30 (denoted as 0.30), where correlations reached 0.97 across all basins. In contrast, the weakest relationships were observed for PCR-GLOBWB model-derived Sy, 0.04 for Saskatchewan, 0.06 for Red–Saskatchewan, and 0.07 for Ottawa, with correlations of 0.35, 0.24, and 0.57, respectively (Figure B.10 in Appendix B). When

normalised time series were used (denoted as N), the association between Spearman's ρ and NSE consistently approached unity across study basins, suggesting strong agreement but potentially overlooking amplitude and bias.

The relationship between Spearman's ρ and KGE exhibited greater inter-basin variability (Figure 3.4). In Saskatchewan and Red-Saskatchewan, the strongest relationships were observed at Sy = 0.15 (denoted as 0.15) (ρ = 0.95 and 0.87), while the weakest corresponded to normalised comparisons (denoted as N), where ρ values dropped to –0.04 and 0.19. In Ottawa, the highest correlation occurred at Sy = 0.07 (denoted as 0.07) (ρ = 0.37), and the lowest at Sy = 0.30 (denoted as 0.30) (ρ = –0.24), further illustrating the sensitivity of KGE to both variability and mean bias.

Leveraging both NSE and KGE revealed unique strengths in evaluating GRACE-GWA comparison (Figure 3.4): NSE provided a clear measure of overall GRACE-GWA predictive accuracy, while KGE illuminated critical biases and variability. Saskatchewan exhibited the strongest NSE-KGE agreement (correlation = 0.94 for Sy = 0.10), whereas Ottawa showed the weakest alignment (correlation = -0.03 for Sy = 0.30), underscoring the value of multi-metric evaluation to capture diverse comparison errors.

These results underscore the limitations of correlation-based assessments in GRACE-GWA comparisons and advocate for a multi-objective evaluation approach. Time series normalisation, while improving monotonic agreement, often distorts absolute performance by inflating KGE bias and dampening NSE responsiveness to seasonal dynamics. Instead, applying representative Sy values enables a more physically meaningful assessment of GRACE-GWA performance. By leveraging the complementary strengths of NSE and KGE, this framework provides a more reliable and diagnostic evaluation of multi-model GRACE-based groundwater estimates.

3.3.2 Multi-objective evaluation of multi-model GRACE-GWA realisations employing NSE and KGE

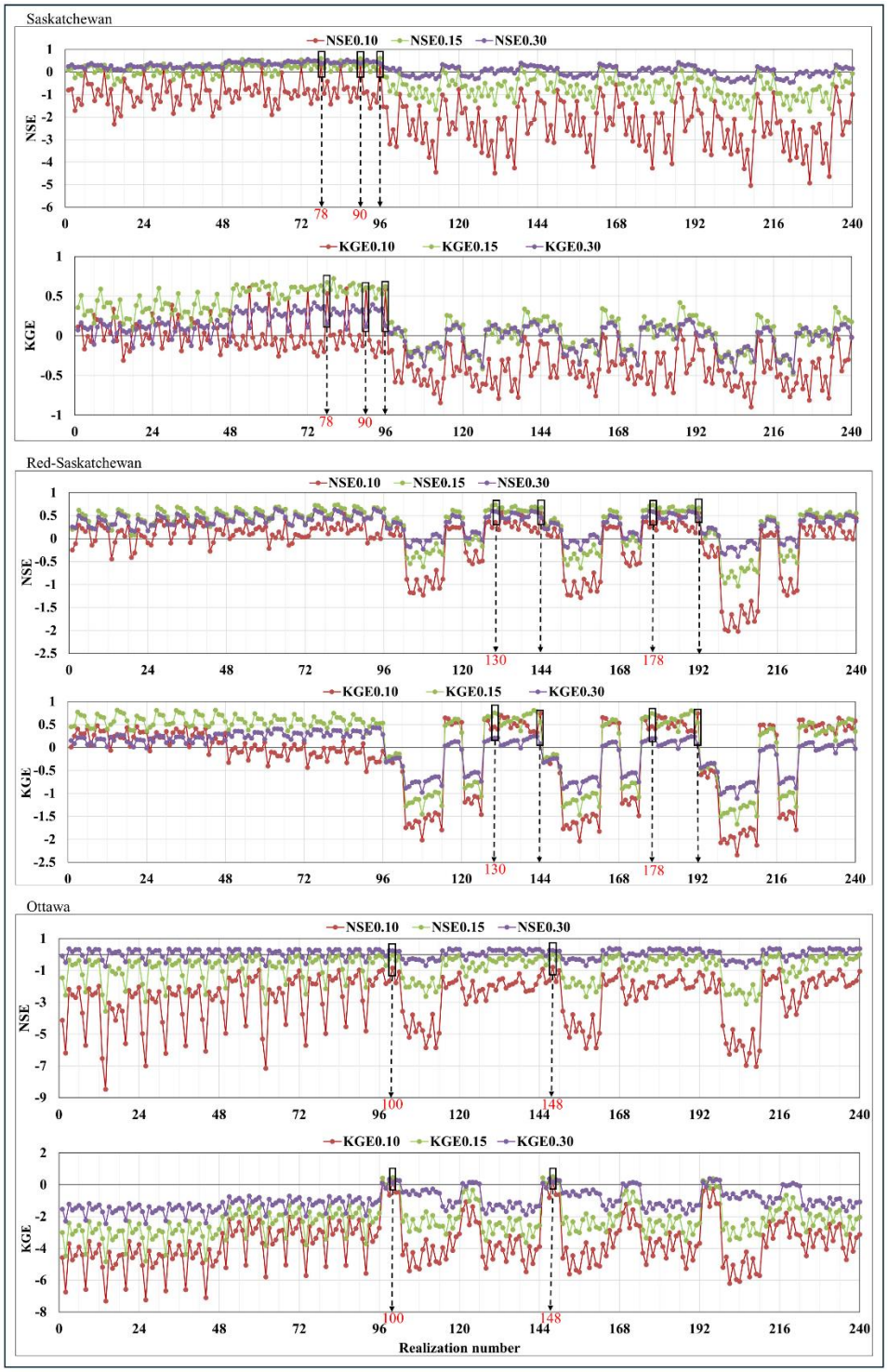


Figure 3.5 Performance of 240 GRACE-GWA realisations evaluated against basin-scale in-situ groundwater anomalies using NSE and KGE metrics across Saskatchewan, Red-Saskatchewan, and Ottawa basins. Comparisons incorporate three Sy values—0.10, 0.15, and 0.30— denoted as NSE0.10, NSE0.15, NSE0.30 and KGE0.10, KGE0.15, KGE0.30, respectively. Dashed black lines highlight the optimal realisations within each basin, reflecting the highest agreement with in-situ groundwater anomalies based on corresponding NSE and KGE metrics.

Systematic biases in NSE and KGE, driven by uncertainties in Sy, convolute time series agreement, highlighting the importance of a priori knowledge of the aquifer systems under study (Figure 3.5). To robustly identify optimal realizations, a structured selection process was employed across the different Sy values. Specifically, for each Sy value (0.10, 0.15, and 0.30), the top 10 GRACE-GWA realisations were identified based on their NSE and KGE rankings, resulting in six high-performance subsets. The final set of optimal realisations was defined as those appearing in all six subsets, ensuring consistently strong performance across both metrics and Sy assumptions. This strategy mitigates the influence of any single Sy value or performance metric and enhances confidence in the hydrological reliability of the selected time series realisations. For clarity of presentation, GRACE-GWA comparisons employing Sy from the PCR-GLOBWB model, exhibiting considerable bias, were excluded from Figure 3.5.

In the Saskatchewan basin, GRACE-GWA realisations96, 90, and 78 yielded the highest performance against in-situ groundwater anomalies, with NSE values of 0.35, 0.61, and 0.40 and corresponding KGE values of 0.61, 0.67, and 0.15 across Sy = 0.10, 0.15, and 0.30, respectively. These top-performing realisations were generated using the JPL-M GRACE-TWSA solution, GlobSnow, HGSWE, and WGHM for SWEA, and ERA5-Land for SMA. In the Red-Saskatchewan basin, realisations130, 178, 144, and 192 demonstrated the strongest agreement with in-situ groundwater anomalies. These were derived from CSR-SH and JPL-SH GRACE-TWSA solutions, paired with ERA5-Land and GlobSnow for SWEA, and ERA5-Land and CLSM for SMA, achieving NSE values of 0.56, 0.73, and 0.60 and KGE values of 0.74, 0.76, and 0.21 across Sy = 0.10, 0.15, and 0.30, respectively. In the Ottawa basin, realisations148 and 100 exhibited relatively higher performance despite the basin's high bias, with NSEs of –0.79, –0.03, and 0.25 and KGEs of 0.23, 0.50, and 0.12, across Sy = 0.10, 0.15, and 0.30, respectively. These were derived from CSR-SH and JPL-SH for GRACE-TWSA, NOAH 1˚ for SWEA and CLSM for SMA.

Collectively, these results highlight the sensitivity of GRACE-GWA accuracy to the selection of GRACE-TWSA solution and water budget components and demonstrate the value of a multi-objective framework for identifying the most reliable groundwater storage estimates across contrasting hydrogeologic contexts.

3.3.3 Evaluating trends and average amplitudes in multi-model GRACE-GWA

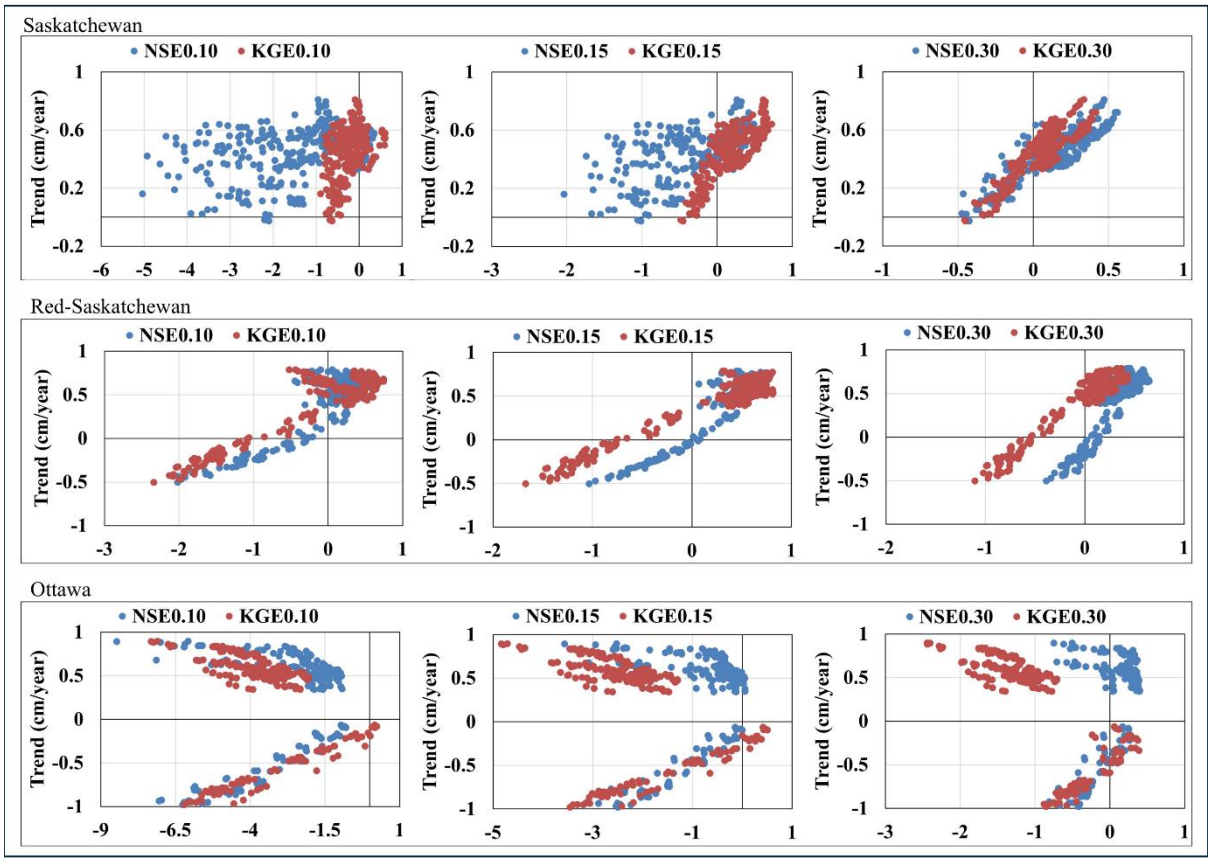


Figure 3.6 Scatter plots examining the relationships between GRACE-GWA trends and GOF metrics (NSE and KGE) across 240 GRACE-GWA realisations for the Saskatchewan, Red-Saskatchewan, and Ottawa basins. NSE and KGE were computed by evaluating GRACE-GWA realisations against basin-scale in-situ groundwater anomalies, using Sy values of 0.10, 0.15, and 0.30 (denoted as NSE0.10, NSE0.15, NSE0.30, KGE0.10, KGE0.15, and KGE0.30, respectively). Additional evaluations applying PCR-GLOBWB-derived Sy values and normalised GRACE-GWA time series are provided in Figure B.11 in Appendix B.

The study assessed multi-model GRACE-GWA trends and amplitudes applying NSE and KGE for GRACE-GWA realisations against in-situ groundwater anomalies, identifying time series characteristics that yield optimal realizations. Across the 240 realizations, substantial variability was observed, with trend magnitudes differing by up to ~12-fold and amplitudes by ~5-fold, highlighting complexities in groundwater assessment due to inefficiencies in water budget accounting (Figures 3.6 and 3.7). Despite these discrepancies, Pearson and Spearman's ρ correlations were generally >0.50, a correlation threshold typically applied in GRACE-groundwater validation to suggest consistency between remotely sensed and observed groundwater change. However, the variability in NSE and KGE values pointed to more complex inconsistencies across the GRACE-GWA realisations that were not captured by correlation alone.

The Seasonal Mann-Kendall (SMK) test identified 32, 20, and 6 non-significant GRACE-GWA trends across Saskatchewan, Red-Saskatchewan, and Ottawa, respectively. Notably, the strength of association between GRACE-GWA trends and NSE/KGE metrics varied systematically with the applied Sy. In Saskatchewan, low Sy (e.g., Sy = 0.04) yielded negligible correlation between trends and NSE, and a negative association with KGE (ρ = –0.34), suggesting that weak trends coincided with inflated GOF metrics. As Sy increased, these relationships strengthened, reaching peak correlations at Sy = 0.30 (ρ = 0.83 for both NSE and KGE), indicating improved alignment under reduced systematic bias. Similar behaviour was observed in the Red-Saskatchewan basin, where correlations progressed from weak at Sy = 0.06 (ρ = 0.25 for NSE, ρ = 0.41 for KGE) to strong at Sy = 0.30 (ρ = 0.83 for both metrics). In contrast, Ottawa, characterised by larger bias, displayed minimal correlation between trend and NSE (ρ = –0.04 to 0.22 across Sy values), while KGE exhibited a progressively negative relationship with trend magnitude, culminating in a strong inverse correlation at Sy = 0.30 (ρ = –0.83). Under normalisation, trends maintained a moderate-to-strong positive correlation with NSE across all basins, but relationships with KGE remained weak and inconsistent (Figure B.11 in Appendix B).

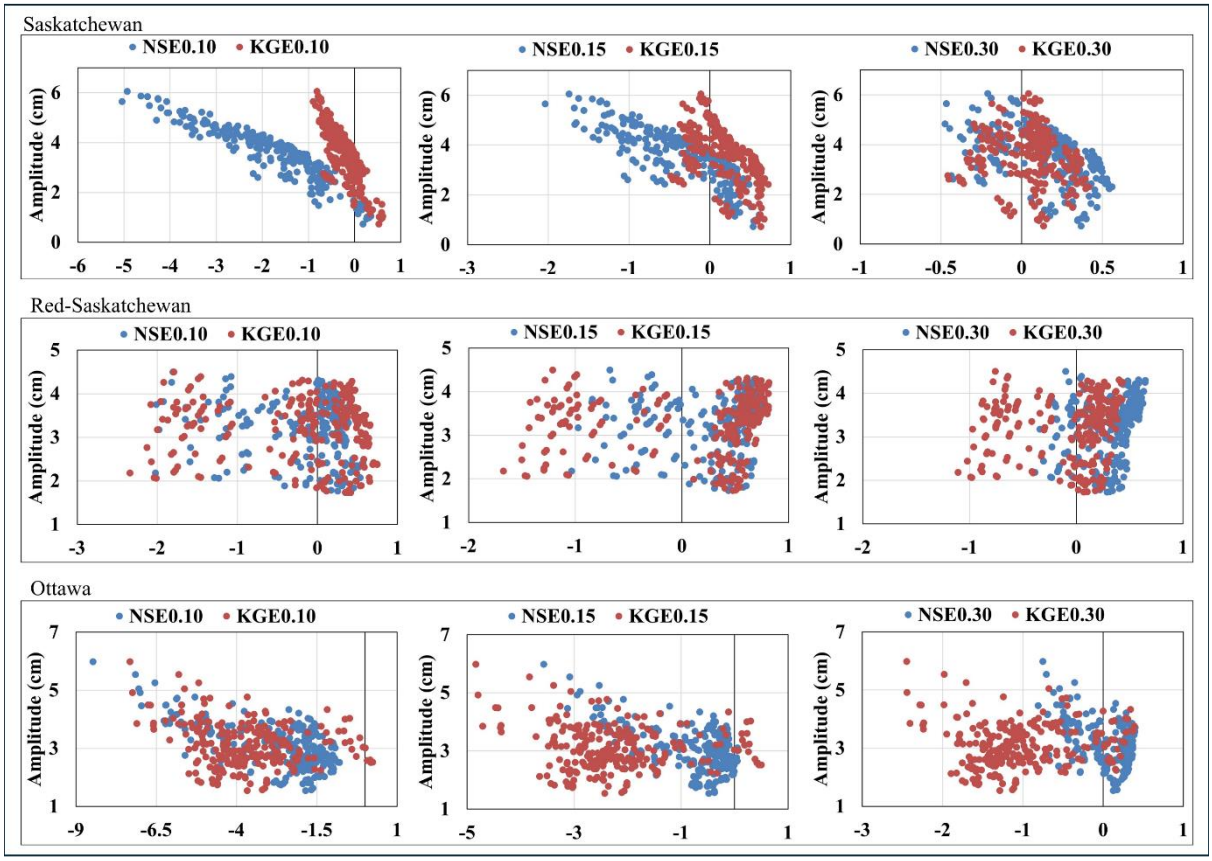


Figure 3.7 Scatter plots examining the relationships between GRACE-GWA amplitudes and GOF metrics (NSE and KGE) across 240 GRACE-GWA realisations for the Saskatchewan, Red-Saskatchewan, and Ottawa basins. NSE and KGE were computed by evaluating GRACE-GWA realisations against basin-scale in-situ groundwater anomalies, using Sy values of 0.10,

0.15, and 0.30 (denoted as NSE0.10, NSE0.15, NSE0.30, KGE0.10, KGE0.15, and KGE0.30, respectively). Additional evaluations applying PCR-GLOBWB-derived Sy values and normalised GRACE-GWA time series are provided in Figure B.12 in Appendix B.

Amplitude variability was similarly pronounced, with seasonal amplitude varying by 2.6- to 8-fold across realisations (Figure 3.7), underscoring substantial discrepancies in seasonal signal representation among water budget estimates. In Saskatchewan, amplitudes were strongly inversely related to GOF metrics, particularly under high bias conditions (Sy 0.04). At Sy = 0.04, correlations reached –0.90 for NSE and –0.70 for KGE, indicating that over-amplified seasonal signals corresponded to lower GRACE-GWA performance. These relationships remained negative, though weaker, even at Sy = 0.30. Red-Saskatchewan exhibited a similar transition: weak negative relationships at Sy = 0.06 (ρ = –0.57 for NSE; ρ = –0.17 for KGE) became positive at Sy = 0.30 (ρ = 0.40 for NSE; ρ = 0.23 for KGE). In Ottawa, amplitude-GOF relationships remained inconsistent, with amplitude–NSE correlation persisting as negative and amplitude–KGE turning only marginally positive at Sy = 0.30. Under normalised comparisons, amplitude-GOF relationships varied widely across basins, further indicating that normalisation may obscure signal fidelity (Figure B.12 in Appendix B).

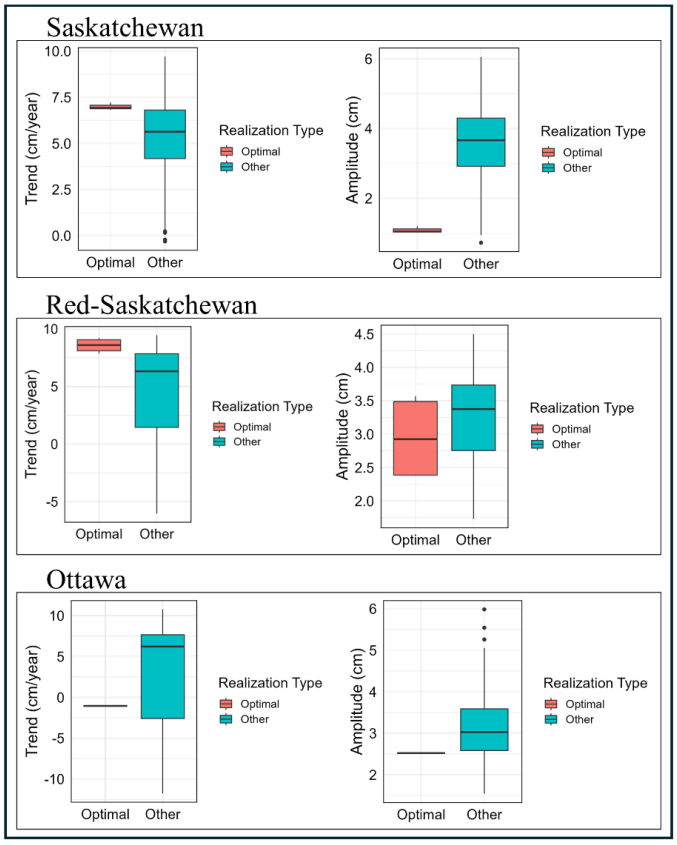


Figure 3.8 Box plots illustrating GRACE-GWA trend and amplitude estimates across 240 multi-model realisations for the Saskatchewan, Red–Saskatchewan, and Ottawa basins, contrasting "optimal" estimates (IDs 78, 90, 96 in Saskatchewan; 130, 144, 178, 192 in Red–Saskatchewan; and 100, 148 in Ottawa) against all other realizations.

Figure 3.8 highlights the spread of trends and amplitudes in Saskatchewan, Red–Saskatchewan, and Ottawa. In Saskatchewan, trends spanned –0.31 to 9.70 cm/year and amplitudes 0.73 to 6.00 cm, while three optimal realisations (e.g., 78, 90, 96) clustered at higher trends (6.82 to 7.21 cm/year) but maintained low amplitudes (1.02 to 1.20 cm). A similar outcome emerged in Red–Saskatchewan, where four optimal realisations (e.g., 130, 144, 178, 192) occupied the upper end of the trend range (7.86 to 9.21 cm/year) yet retained moderate amplitudes (2.38 to 3.57 cm). Both basins, exhibiting lower bias, suggest that optimal solutions maximize trend consistency while avoiding extremes in amplitude. By contrast, Ottawa's wide range of trends (–11.75 to 10.76 cm/year) and amplitudes (1.54 to 5.98 cm) yielded two optimal realisations(e.g., 100, 148) featuring modest negative trends (–0.89 and –1.12 cm/year) and matching amplitudes (2.52 cm), underscoring that in high-bias contexts, "optimal" does not necessarily align with the most extreme trend values.

Collectively, these findings demonstrate that uncertainties in specific yield propagate systematic bias into GRACE-GWA evaluations, influencing both trend and amplitude diagnostics. Although optimal realisations in lower-bias basins converged on higher trends paired with lower amplitudes, the high-bias conditions in Ottawa blurred the trends pattern. Nevertheless, across all basins, optimal realisations generally avoided the lowest trend values and the highest amplitude extremes. The results affirm that while GRACE-GWA realisations can broadly capture groundwater variability, identifying reliable outputs requires a multi-objective evaluation framework that accounts for both systematic bias and temporal dynamics.

3.4 Conclusions

Groundwater monitoring is fundamental to understand and manage hydrological systems under mounting anthropogenic and climatic pressures (Taylor et al., 2013; Famiglietti, 2014; Thomas & Famiglietti, 2019). Such understanding supports informed decision-making, ensures groundwater sustainability, and fosters resilience in water management systems. While GRACE has significantly improved the ability to track large-scale groundwater dynamics, isolating time series that reflect true groundwater changes remains challenging primarily due to persistent uncertainties in water budget components. These uncertainties introduce biases and obscure variability in GRACE-GWA, complicating validation and reducing confidence in GRACE-based groundwater assessments (Saxe et al., 2021; Akl & Thomas, 2024; see also Chapter 2 of this thesis). To address these challenges, this study applied a multi-objective comparative framework integrating NSE and KGE metrics to compare multi-model GRACE-GWA realisations against basin-scale in-situ groundwater anomalies. By moving beyond traditional correlation-based comparison, this approach systematically captures both the

magnitude and temporal variability of groundwater signals, enabling a more robust and hydrologically meaningful assessment of GRACE-GWA performance (Figure 3.5).

The detailed analysis of GRACE-GWA trends and amplitudes demonstrated substantial variability across multi-model realizations, highlighting critical uncertainties inherent in remote groundwater estimations (Figures 3.6 and 3.7). Specifically, trends varied by approximately a factor of 12, and amplitude estimates differed by up to five-fold among realizations. Such wide-ranging variability emphasizes the significant impact of selecting GRACE-TWSA and water budget components on groundwater estimates. Although correlation coefficients often exceeded 0.50, deeper inconsistencies emerged under NSE/KGE analyses, especially in high-bias basins like Ottawa (Figures 3.2, 3.6 and 3.7). These findings closely align with recent research by Akl and Thomas (2024), who emphasized the necessity of adopting multi-model approaches to quantify uncertainty and enhance the reliability of GRACE-based groundwater assessments. The results reinforce the imperative for employing complementary GOF metrics in a unified multi-objective comparative framework, enabling more accurate identification of optimal realization and increasing confidence in interpreting groundwater dynamics from GRACE observations.

Through a multi-objective comparative methodology, this research systematically identified optimal GRACE-GWA realisations tailored to each basin. Importantly, the results revealed limitations inherent in normalisation approaches commonly used for GRACE-GWA comparisons (e.g., Castle et al., 2014; Liu et al., 2021), including inflation of biases within KGE and reduced sensitivity of NSE to seasonal variations. Furthermore, uncertainties associated with specific yield (Sy) introduced significant systematic biases, critically influencing the accuracy of GRACE-GWA evaluations. Therefore, integrating representative, basin-specific Sy emerged as essential to ensuring meaningful and robust comparisons. Despite these complexities, the multi-objective framework successfully distinguished optimal realisations that most effectively represent groundwater variability in terms of both magnitude and seasonal dynamics (Figure 3.5). The results reveal a pronounced contrast in how optimal realisations arise under varying bias conditions. In Saskatchewan and Red–Saskatchewan—where bias is relatively low—optimal GRACE-GWA realisations cluster around higher trends coupled with lower amplitudes. Conversely, elevated bias in the Ottawa basin produces optimal realisations characterised by moderate trends and lower amplitudes, demonstrating that trends/amplitudes values are not intrinsically linked to improved reliability. Collectively, these results highlight the importance of adapting trend and amplitude criteria to each basin's distinct hydrogeological context, rather than relying on extremes alone to identify robust GRACE-GWA outcomes.

While the primary objective of this research was methodological advancement, rather than characterisation of groundwater dynamics in specific basins, several limitations were recognised. For example, surface water anomalies (SWA), derived from spatially limited in-situ observations, potentially overlapped with groundwater signals, complicating the isolation of pure groundwater changes within GRACE-GWA (Castle et al., 2014; Busker et al., 2019; Thomas & Nanteza, 2023). Incorporating recent remote sensing advancements (e.g., Gerdener et al., 2023; Hou et al., 2024) could significantly enhance the accuracy of SWA within the water budget framework, thereby improving GRACE-GWA reliability. Moreover, identified optimal realisations inherently reflect basin-specific hydrological conditions, choice of GRACE-TWSA solutions, and selected water budget datasets. Finally, the effectiveness of the presented multi-objective comparative framework fundamentally relies on the availability and quality of basin-scale in-situ groundwater anomaly observations. Producing spatially representative and reliable groundwater datasets remains challenging (Rodell et al., 2007; Sun et al., 2010), highlighting an urgent research priority for developing robust basin-scale in-situ groundwater reference datasets to validate GRACE-GWA estimates more effectively (Figure 3.5).

Overall, this study represents a significant methodological advancement by introducing a rigorous, multi-objective comparative framework for evaluating multi-model GRACE-GWA realizations. This approach systematically addresses biases, quantifies uncertainty from water budget accounting, and substantially surpasses conventional correlation-based methods that frequently obscure critical differences and mislead water resource assessments. The findings presented here advocate strongly for reconsidering traditional GRACE-GWA evaluation methods reliant solely on correlation metrics, cautioning against their limitations in capturing essential groundwater dynamics. Ultimately, this research sets a new standard for GRACE-based groundwater analysis, providing a robust methodological foundation that enhances the reliability, accuracy, and practical utility of remotely sensed groundwater assessments, thereby empowering researchers, stakeholders, and decision-makers to sustainably manage and protect critical groundwater resources.

Chapter 4. Global Groundwater Drought Assessment Revisited: A Holistic Re-evaluation of the GRACE-Groundwater Drought Index Across Major Aquifers

This chapter is based on a manuscript submitted for publication and currently under peer review by Water Resources Research (WRR).

Chapter abstract: The Gravity Recovery and Climate Experiment (GRACE) and GRACE Follow-On missions have enriched global groundwater monitoring, forming the basis for tools that detect groundwater drought, including the GRACE-Groundwater Drought Index (GGDI). The reliability of GGDI is fundamentally tied to the accurate isolation of a representative groundwater storage anomaly (GRACE-GWA) from GRACE observations, a challenge heightened by the scarcity of direct water budget measurements and the diverse methodologies applied in GRACE data processing. This study integrates multi-model GRACE-GWA estimates, derived from diverse GRACE solutions and water budget datasets, into the GGDI framework to assess how variability among these estimates influences groundwater drought interpretation across 37 study aquifers. Five key drought indicators are evaluated: number of drought events, maximum duration, average duration, severity, and maximum intensity. The analysis reveals substantial uncertainty in these indicators, driven by differences among GRACE-GWA realizations. Aquifer memory, quantified via autocorrelation in GGDI time series, also exhibits pronounced variability, most notably in the Nubian Basin where memory estimates range from 3 to 61 months amongst multi-model realizations. The findings underscore that even modest discrepancies in GRACE-GWA methodologies can translate into considerable uncertainties in both drought indicators and aquifer memory, thereby compromising the reliability of groundwater drought assessments. Furthermore, the results indicate that basins with higher aquifer memory typically experience fewer drought events yet endure more prolonged and intense drought conditions when they do occur.

4.1 Introduction

Drought, characterised as an extended period of water scarcity, disrupts ecosystems, economies, and societies by virtue of an anomalous reduction in precipitation (Dracup et al., 1980; Wilhite & Glantz, 1985; Mishra & Singh, 2010; Wilhite et al., 2014). Although drought commences with a precipitation deficit, incoming water shortages propagate through the hydrological system, leading to diminished groundwater storage (Mishra & Singh, 2010). Groundwater drought poses a complex challenge because groundwater storage deficits arise from multiple factors: reduced groundwater recharge (Goodarzi et al., 2016), excessive withdrawals (Mishra

& Singh, 2010; Bloomfield & Marchant, 2013; Bloomfield et al., 2015), and reduced discharge (Peters et al., 2001). Intensifying climate extremes coupled with escalating human interventions (Taylor et al., 2013; Famiglietti, 2014; Thomas & Famiglietti, 2019) have magnified the frequency and severity of groundwater drought (Panda et al., 2007; Tallaksen & VanLanen, 2023). Often seen as a natural buffer during hydrological (i.e., reduced streamflow) or agricultural (i.e., depleted soil moisture) droughts (Famiglietti, 2014), groundwater has reliably supported human consumption and agricultural production (Siebert et al., 2010; Hughes et al., 2012; Scanlon et al., 2012; Castle et al., 2014). However, unsustainable pumping practices undermine this resilience, compounding water scarcity and intensifying drought impacts (Famiglietti et al., 2011; Famiglietti & Rodell, 2013; Castle et al., 2014). Operational groundwater drought monitoring is hampered by the challenges of directly measuring groundwater storage—particularly in transboundary or remote aquifers where in-situ monitoring is sparse or absent (Giordano, 2009; Famiglietti et al., 2011; Lall et al., 2020; Condon et al., 2021; Jasechko et al., 2024).

Thomas et al. (2017a) introduced the Gravity Recovery and Climate Experiment (GRACE) groundwater drought index (GGDI), a normalised groundwater storage-based indicator to enumerate groundwater drought. By leveraging GRACE-derived groundwater storage anomalies (GRACE-GWA), the GGDI captures both surpluses and deficits, offering a comprehensive evaluation of groundwater drought conditions (Thomas et al., 2017a). Since its development, GGDI has been widely applied to analyse groundwater drought and characterize its spatial and temporal variability across diverse aquifer systems (Wang et al., 2020b; Paredes-Trejo et al., 2021; Satish Kumar et al., 2021; Ali et al., 2022; Liu et al., 2022a; Wang et al., 2022a; Huang et al., 2023a; Aon et al., 2024; Nandi & Biswas, 2024; Neves, 2024; Nigatu et al., 2024; Nikraftar et al., 2024; Song et al., 2024; Zhang et al., 2024; Zheng et al., 2024; Liu et al., 2025). Despite the growing application of GGDI, its interpretation hinges on the variability in GRACE-GWA time series (Thomas et al., 2017a). This variability is driven by differences in GRACE processing methods and uncertainties in other water budget components (e.g., soil moisture, snow, surface water) (Saxe et al., 2021; Akl & Thomas, 2023, 2024). Many GGDI applications since Thomas et al. (2017a) have overlooked this inherent variability, raising questions about the reliability of groundwater drought interpretations. For instance, Wang et al. (2020) identified a severe drought event in the North China Plain (NCP) from August 2013 to September 2014 (GGDI = –1.36) within a study period of 2003 to 2015, while Ali et al. (2022) cited an extreme event in the Indus Basin in August 2010 (GGDI = –1.81) within a study period of 2003 to 2016. Both studies relied on a single land surface model (LSM) to disaggregate

GRACE-TWSA into GRACE-GWA, thereby disregarding structural model biases (Clark et al., 2008; Ruddell et al., 2019; Konapala et al., 2020) and possibly conflating groundwater signals with unaccounted surface water components (Castle et al., 2014; Akl et al., 2022; Thomas & Nanteza, 2023; Akl & Thomas, 2024; see also Chapter 2 of this thesis). The evolving knowledge of the challenges in applying the water budget framework for estimating GRACE-GWA (Akl & Thomas, 2024; see also Chapter 2 of this thesis) translates to a compelling need to re-examine the GGDI framework. Central to this re-examination is elucidating how variability amongst multi-model GRACE-GWA estimates impacts GGDI interpretation, which can be achieved by quantifying and characterising variability in key groundwater drought indicators derived from the GGDI framework. This study seeks to (1) assess the influence of multi-model GRACE-GWA realisations on key groundwater drought indicators and (2) assess aquifer memory variability for perceived groundwater drought. The analysis centres on the world's largest aquifers, which collectively store the majority of global groundwater and are vital to water and food security (Margat, 2008; Gleeson et al., 2012; Margat & Van der Gun, 2013; Van der Gun, 2022).

This study employs an integrated water balance framework that synthesizes various data combinations to generate multi-model GRACE-GWA realisations across 37 aquifer systems. Specifically, five GRACE-TWSA solutions are combined with a comprehensive suite of auxiliary datasets, including six soil moisture datasets, six snow water datasets, and remote sensing data from 429 lakes and reservoirs, to capture both surface and subsurface water storage change. Leveraging this unprecedented data fusion, the GRACE-Groundwater Drought Index (GGDI) is re-evaluated for the period spanning April 2002 to December 2022, incorporating data from both GRACE and GRACE Follow-On (GRACE-FO) missions. This framework delineates how the variability among multi-model GRACE-GWA estimates, originating from differences in GRACE-TWSA methodologies and water budget accounting, impacts GGDI interpretation. By evaluating a comprehensive suite of drought indicators, including the number of drought events, maximum and average durations, severity, and maximum intensity, the results reveal significant uncertainty that undermines the reliability of groundwater drought characterization. Furthermore, the framework quantifies aquifer memory in groundwater storage deviations by leveraging the autocorrelation of GGDI. This analysis reveals marked uncertainty across many basins, impeding the ability to assess aquifer recovery and response dynamics. Finally, the analysis demonstrates that basins with higher aquifer memory, while generally experiencing fewer drought events, exhibit drought episodes that are more prolonged and intense when they occur. Ultimately, the findings highlight the critical importance of

integrating multi-model GRACE-GWA realisations within the GGDI framework to understand uncertainties in groundwater drought assessments that might otherwise be concealed by limited model outputs.

4.2 Data and Methods

This study focuses on 37 large aquifer systems as defined by the Worldwide Hydrogeological Mapping and Assessment Program (WHYMAP; Margat & Van der Gun, 2013) for the period spanning April 2002 to December 2022 (Figure 4.1 and Table 4.1). These aquifers are among the world's most productive, holding a substantial share of accessible global groundwater resources (Margat, 2007; Gleeson et al., 2012; Richey et al., 2015; Thomas et al., 2017b; Shamsudduha & Taylor, 2020; Van der Gun, 2022). In addition, each aquifer exceeds the minimum 100,000 km² spatial threshold necessary for reliable GRACE-based assessments (Scanlon et al., 2016). These aquifers span a wide range of climatic zones, from hyper-arid to humid conditions, offering diverse contexts for investigating hydrological processes (Figure 4.1).

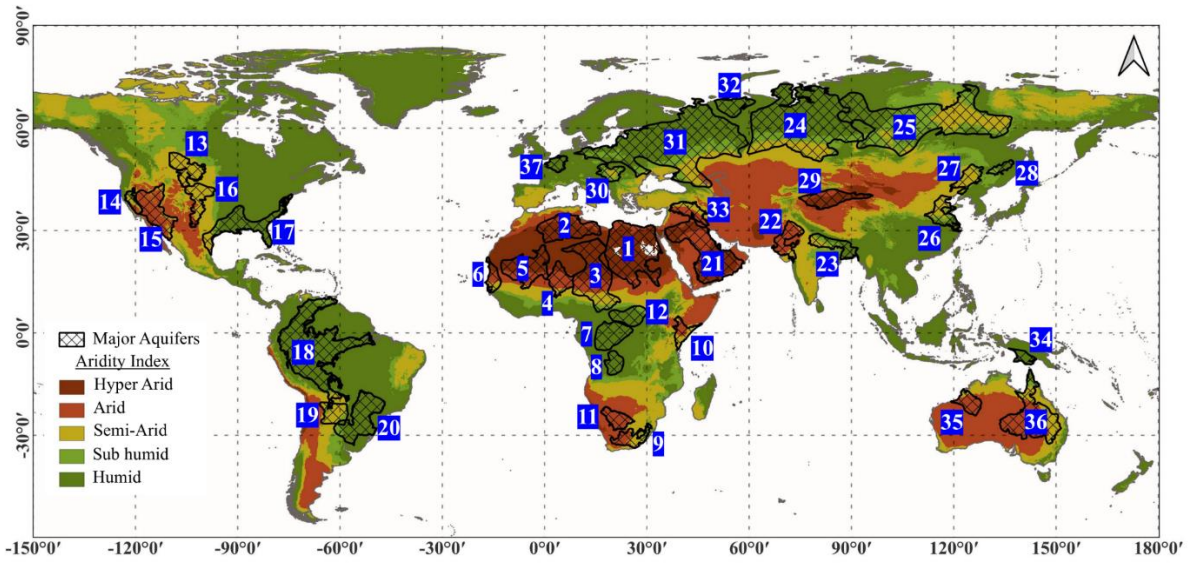


Figure 4.1 Map of the 37 study aquifers, sourced from the World-wide Hydrogeological Mapping and Assessment Programme (WHYMAP; Margat & Van der Gun, 2013). Numeric labels refer to aquifer names listed in Table 4.1. The colour scale represents the aridity index, sourced from the CGIAR-CSI Global Aridity Index and Potential Evapotranspiration Climate Database V3 (Global-AI_PET; Zomer & Trabucco, 2022).

n	Aquifer name	Continent	Area km^2	Climate zone
1	Nubian Aquifer System (NAS)	Africa	2455638	Hyper Arid
2	Northwest Sahara Aquifer System (NWSAS)		1042049	Hyper Arid
3	Lake Chad Basin		1990500	Arid
4	Iulluemeden-Irhazer Basin		510056.4	Arid
5	Taoudeni-Tanezrouft Basin		1097585	Arid
6	Senegalo-Mauretanian Basin		293085.5	Arid
7	Cuvette Basin		794661.8	Humid
8	Coango Basin		330251.4	Humid
9	Karoo Basin		379080.8	Semi-Arid
10	Ogaden-Juba Basin		330924.2	Semi-Arid
11	Stampriet-Kalahari Basin		361382.2	Arid
12	Karoo-Carbonate Basin		546942.2	Humid
13	Northern Great Plains Aquifer-Judith River	North America	450771.3	Semi-Arid
14	California Central Valley Aquifer System		156295.3	Semi-Arid
15	Basin and range basing-fill aquifers		863740.6	Arid
16	Ogallala Aquifer (High Plains)		471906.5	Semi-Arid
17	Atlantic and Gulf Coastal Plains Aquifer		1118877	Humid
18	Amazonas Basin	South America	3655269	Humid
19	Yrendá-Toba-Tarijeño Basin		484591.2	Semi-Arid
20	Guarani Basin (Paraná Basin)		1213632	Humid
21	Arabian Aquifer System	Asia	2021297	Hyper Arid
22	Indus River Basin		507928.6	Arid
23	Ganges-Brahmaputra Basin		495449.6	Humid
24	West Siberian Artesian Basin		2735273	Humid
25	Yakut Basin		2673045	Sub humid
26	North China Plains Aquifer System		461819.6	Sub humid
27	Song-Liao Plain		307573.9	Semi-Arid
28	Middle Heilongjiang-Amur River Basin		113233.5	Humid
29	Tarim Basin		576224.6	Hyper Arid
30	North Caucasus Basin	Europe	207496.8	Sub humid
31	Russian Platform Basin		3764406	Humid
32	Pechora Basin		194799.1	Humid
33	Taurus-Zagros Basin	Asia	360327.1	Semi-Arid
34	Merauke-Ketu Basin Aquifer	Australia	206043.6	Humid
35	Canning Basin		379204.2	Arid
36	Great Artesian Basin		1737982	Arid
37	Paris basin	Europe	149369.4	Humid

Table 4.1 Identification number (n), name, geographic location (continent), area (km2), and climatic classification of the world's 37 major aquifer systems, as determined by the mean aridity index.

4.2.1 GRACE-terrestrial water storage anomalies (GRACE-TWSA)

Terrestrial water storage anomalies (GRACE-TWSA) were obtained from five GRACE-based solutions covering the interval from April 2002 to December 2022, including both spherical harmonic (SH; Landerer & Swenson, 2012) and mass concentration (mascon; Watkins et al., 2015, Wiese et al., 2016) approaches. The SH solutions were provided on 1˚×1˚ global grids by three data processing centres: the Centre for Space Research (CSR-SH), Jet Propulsion Laboratory (JPL-SH), and GeoForschungsZentrum (GFZ-SH) (https://grace.jpl.nasa.gov/). Meanwhile, mascon solutions were supplied by CSR (CSR-M) on a 0.25˚×0.25˚ global grid (https://www2.csr.utexas.edu/grace/) and by JPL (JPL-M) on a 0.5˚×0.5˚ global grid (https://grace.jpl.nasa.gov/). Data gaps between the GRACE and GRACE-FO missions (July 2017–May 2018) were omitted, and shorter missing intervals (one to two months) were filled using cubic spline interpolation (Wei et al., 2021; Aon et al., 2024). This study avoided JPL scale factors derived from the Community Land Model (CLM; Lawrence et al., 2011) citing the CLM limitations in accurately representing snow processes (Wiese et al., 2016)— a critical consideration for many basins investigated.

4.2.2 Auxiliary datasets

Soil Moisture Anomalies (SMA) and Snow Water Equivalent Anomalies (SWEA)

Monthly SMA and SWEA were derived from six modelling and reanalysis systems, each offering varied spatial resolutions and unique physical representations. SMA and SWEA were obtained from LSMs provided by the NASA Global Land Data Assimilation System (GLDAS; Rodell et al., 2004), including NOAH at 1˚×1˚, Variable Infiltration Capacity (VIC) at 1˚×1˚, and Catchment Land Surface Model (CLSM) at 1˚×1˚. High-resolution SMA and SWEA data were also sourced from NOAH at 0.10˚×0.10˚ via the Famine Early Warning System Network (FEWS NET) Land Data Assimilation System (FLDAS; McNally et al., 2017). Additional datasets included the WaterGAP Global Hydrology Model v2.2e (WGHM; Müller Schmied et al., 2024) at 0.5˚ ×0.5˚ resolution and the ERA5-Land dataset at 0.10˚×0.10˚ resolution (Muñoz-Sabater et al., 2021).

The models used in this study capture soil moisture dynamics through distinct methodologies. The NOAH model (at 1˚ and 0.10˚ resolutions), simulates soil moisture through a multi-layer soil profile divided into four depths: 0-10cm, 10-40cm, 40-100cm, 100-200cm (Rodell et al., 2004). VIC represents soil moisture within three layers: a surface layer (0–30 cm), a variable-depth second layer, and a bottom layer. CLSM conceptualizes soil moisture within three reservoirs: the Surface layer (0-2 cm), Root Zone (0-100 cm), and the Profile reservoir (varies grid-by-grid). WGHM simulates monthly soil moisture within the effective root zone using a

single-layer soil water storage compartment that integrates land cover, soil-specific maximum storage capacity, and soil texture (Müller Schmied et al., 2024). ERA5-Land provides volumetric soil moisture estimates across four layers: 0–7 cm, 7–28 cm, 28–100 cm, and 100–289 cm (Muñoz-Sabater et al., 2021). In this study, soil moisture within the 0–1 m depth range was considered, as the deeper layers have the potential to become fully saturated (Thomas et al., 2017a; Akl & Thomas, 2024; see also Chapter 2 of this thesis).

Surface Water Storage Anomalies (SWA)

To address the spatial and temporal limitations of in-situ surface water observations (Alsdorf et al., 2007; Busker et al., 2019; Tortini et al., 2020), this study integrated satellite-based datasets from GloLakes (Hou et al., 2024) and the Copernicus Climate Change Service (https://cds.climate.copernicus.eu/). The Hydrological Data and Maps Based on Shuttle Elevation Derivatives at Multiple Scales (HydroSHEDS; Lehner et al., 2008) dataset was used to delineate lake and reservoir extents (Figure C.1 in Appendix C). A quality-control algorithm excluded lakes with more than one missing seasonal reading in any calendar year, producing a final dataset of 429 lakes (415 from GloLakes and 14 from Copernicus; Figure 4.2 and Table 4.2). The data were aggregated to produce consistent monthly intervals, with any remaining gaps filled via cubic spline interpolation.

While GloLakes offered lake-storage data, Copernicus supplied lake water-level observations that required integration with time-varying lake area. This integration was achieved using the global lake bathymetry dataset (GLOBathy; Khazaei et al., 2022), enabling a reliable representation of surface water storage changes (Akl & Thomas, 2024; see also Chapter 2 of this thesis).

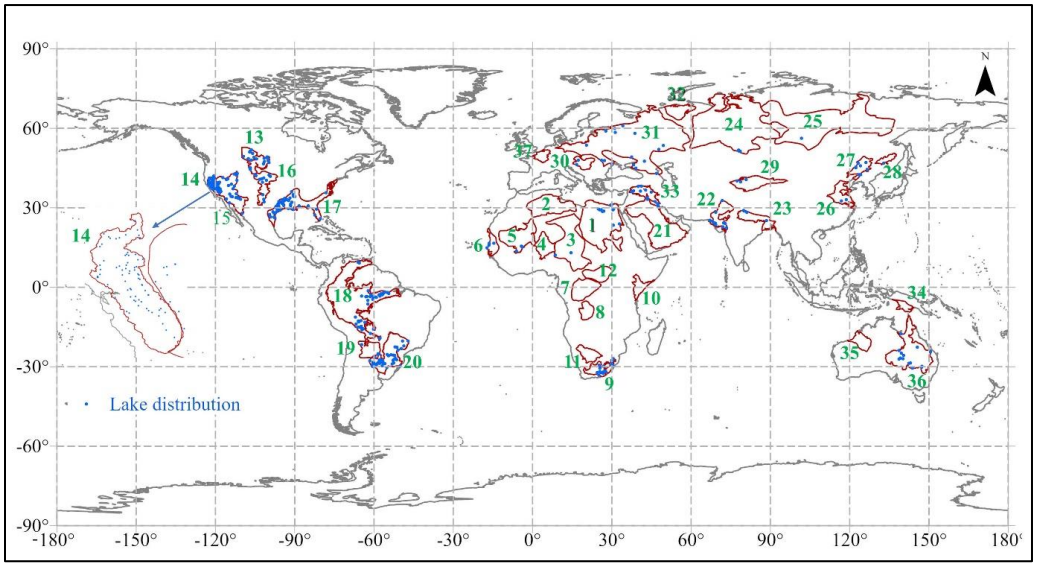


Figure 4.2 Mapping of surface water bodies, including lakes and reservoirs (depicted in blue), utilized in this study across the 37 largest aquifer systems.

Basin	N. Lakes	Basin	N. Lakes	Basin	N. Lakes	Basin	N. Lakes
1	14	11	-	21	3	31	17
2	-	12	-	22	11	32	-
3	2	13	19	23	4	33	10
4	-	14	72	24	3	34	-
5	3	15	36	25	1	35	-
6	5	16	12	26	2	36	15
7	-	17	60	27	9	37	-
8	-	18	45	28	1	**Total**	**429**
9	18	19	5	29	3		
10	-	20	57	30	2		

Table 4.2 Total number of lakes and reservoirs detected by the quality control algorithm across all study basins including 415 lakes from GloLakes and 14 lakes from Copernicus.

4.2.3 Disaggregating GRACE-TWSA into GRACE-GWA

GRACE-TWSA offers an integrated estimate of terrestrial water storage changes, encompassing soil moisture anomalies (SMA), snow water equivalent anomalies (SWEA), surface water anomalies (SWA), and groundwater anomalies (GRACE-GWA). Isolating GRACE-GWA from GRACE-TWSA relies on the water balance equation (Equation 4.1), which assumes that accurate accounting for water budget components (i.e., SMA, SWEA, SWA) allows the extraction of a representative GRACE-GWA signal (Rodell & Famiglietti, 1999).

$$GRACE\text{-}TWSA = SMA + SWEA + SWA + GRACE\text{-}GWA \quad (4.1)$$

To ensure consistency with GRACE processing methodologies (Nanteza et al., 2016), all water budget components (i.e., soil moisture, snow, and surface water) were processed in accordance with GRACE standards. For spherical harmonic (SH) solutions, with significant leakage potential (Swenson & Wahr, 2002; Klees et al., 2007; Landerer & Swenson, 2012; Lenk, 2013; Longuevergne et al., 2013; Awange et al., 2014; Fatolazadeh et al., 2016; Guo et al., 2016; Nanteza et al., 2016), water budget data were projected onto GRACE grids and underwent truncation at harmonic degree 60, destriping with a decorrelation filter (Swenson & Wahr, 2006) and smoothing using a 300 km radius Gaussian filter. Subsequently, these processed components were converted to anomalies by removing the mean of the time series over the period from January 2004 to December 2009, before being subtracted from GRACE-TWSA grids. In contrast, for mascon solutions, potential leakage across GRACE grids was presumed to be effectively addressed during processing (Save et al., 2012; Watkins et al., 2015; Richey et al., 2015a; Bhanja et al., 2016; Scanlon et al., 2016; Wiese et al., 2016; Save et al., 2016; Thomas et al., 2017a, b; Rodell et al., 2018; Thomas & Famiglietti, 2019; Neves et al., 2020;

Wang et al., 2020b; Scanlon et al., 2021, 2022). Water budget components were converted to anomalies by subtracting the mean over January 2004 to December 2009 and further subtracted from the respective GRACE grids prior to averaging at the basin scale. The original spatial resolutions of both SH and mascon grids were preserved, with data clipped to basin boundaries to maintain spatial fidelity and integrity (Akl & Thomas, 2024; see also Chapter 2 of this thesis).

This analysis utilized five GRACE-TWSA solutions, six SMA datasets, six SWEA datasets, and a SWA estimate within the water balance equation (Equation 4.1) to derive GRACE-GWA across the 37 study aquifers. These aquifers span a diverse range of climatic and hydrological conditions, reflecting substantial variability in the composition of water budget components incorporated into the water balance equation (Equation 4.1; Figure 4.1). Consequently, GRACE-GWA estimates for individual aquifers yielded up to 180 realizations, depending on the combination of water budget inputs (Table C.1 in Appendix C).

4.2.4 GRACE-groundwater drought index (GGDI)

The GRACE-Groundwater Drought Index (GGDI; Thomas et al., 2017a) is applied in this study to detect and characterize groundwater drought across the investigated aquifers. GGDI is a standardized anomaly index derived from GRACE-based groundwater storage anomalies (GRACE-GWA), representing deviations from long-term seasonal climatology. By capturing both deficits and surpluses in groundwater storage, GGDI reflects the combined influence of natural variability and anthropogenic pressures on aquifer systems.

This broader interpretability aligns with the multi-causal drought framework proposed by Van Loon et al. (2016), which emphasizes the interaction between climatic drivers and human activity. GGDI accommodates a full spectrum of groundwater conditions: negative values indicate storage deficits (drought), while positive values represent groundwater surpluses or 'wetness', signalling recharge or recovery. This makes GGDI particularly valuable for assessing not only drought onset and severity, but also groundwater resilience and post-drought recovery across hydroclimatic contexts.

To derive GGDI, the influence of seasonality was first removed by calculating a monthly climatology (C_i) for each calendar month (i=1,2,…....,12), given as:

$$C_i = \frac{\sum_1^{n_i} GRACE\text{-}GWA_i}{n_i} \qquad Eq.4.2$$

where n_i represents the number of years of data available for month i. This climatology defines the expected seasonal pattern of groundwater storage and serves as the reference baseline for anomaly detection. The climatology (C_i) is removed from GRACE-GWA to compute the

Groundwater Storage Deviation (GSD), which represents net seasonal variation in groundwater storage. The GSD was further standardized to derive the GGDI using the formula:

$$GGDI_t = \frac{\mathrm{GSD}_t - \overline{\mathrm{X}}_{GSD}}{\sigma_{GSD}} \qquad Eq.4.3$$

where t accounts for time and $\overline{\mathrm{X}}_{GSD}$ and σ_{GSD} are the mean and standard deviation of the GSD time series, respectively. Following Thomas et al. (2017a), a groundwater drought event is defined as a period beginning with at least three consecutive months of GGDI < 0, and it terminates after three consecutive months of GGDI ≥ 0.

4.2.5 Groundwater drought indicators and aquifer memory

GGDI indicators were applied to characterize groundwater drought: *the number of drought events*, representing the total count of distinct drought occurrences; *the maximum drought duration*, defined as the total number of months characterising the longest drought over the observational period; *the average drought duration*, capturing the mean length of all drought events occurring during the observational period; *the drought severity*, quantified as the cumulative sum of GGDI values during drought periods; and *drought maximum intensity*, representing the minimum GGDI value recorded across all drought events, captures the most extreme intensity observed among all identified drought periods. The GGDI indicators were compared against *aquifer memory*, reflecting the temporal persistence of groundwater anomalies.

4.2.6 Aquifer Memory

Aquifer memory refers to the temporal persistence of groundwater storage anomalies, capturing how long the hydrological impacts of a drought event continue to influence subsequent conditions. It reflects the intrinsic response time of an aquifer system and serves as a proxy for its resilience to climatic variability and anthropogenic stress (Opie et al., 2020; Neves, 2024). In this study, aquifer memory was quantified by examining the autocorrelation structure of the GGDI time series. Specifically, memory was defined as the number of months required for the autocorrelation coefficient to decline below 0.1, a threshold commonly used to indicate the point at which the influence of antecedent conditions becomes statistically insignificant (Bloomfield and Marchant, 2013; Dubois and Larocque, 2024).

This metric provides an informative characterisation of groundwater system persistence, enabling cross-aquifer comparison of recovery dynamics. Longer memory values indicate sustained drought impacts and delayed hydrological recovery, while shorter memory signals a more responsive system. To examine how memory correlated with drought behaviour,

estimated memory were systematically compared with key GGDI-derived drought indicators, including event frequency, maximum and average duration, severity, and maximum intensity.

4.3 Results

4.3.1 Propagation of uncertainty through GRACE-groundwater drought index (GGDI)

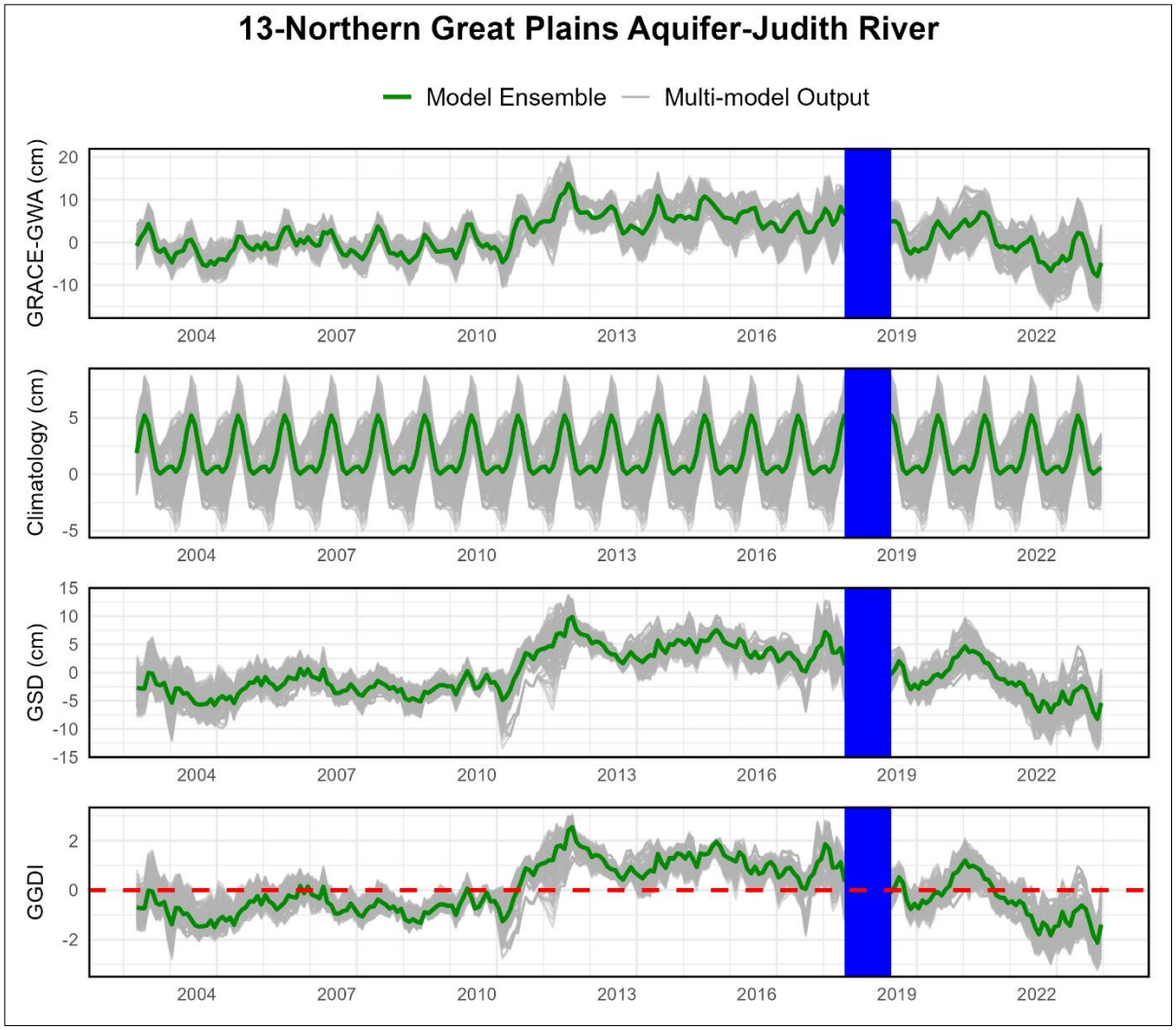


Figure 4.3 Time series of GRACE-GWA and associated climatology, groundwater storage defects (GSD), and GRACE-groundwater drought index (GGDI) across study basin number 13 (Northern Great Plains Aquifer-Judith River) from April 2002 to December 2022. The blue-shaded area marks the gap period between the GRACE and GRACE-FO missions, which was excluded from the analysis. For the time series of all 37 study basins, please refer to Figures C.2–C.38 in Appendix C.

Figure 4.3 depicts the variability in multi-model GRACE-GWA for the Northern Great Plains Aquifer–Judith River (Basin 13), emphasizing how discrepancies in GRACE processing methods and water budget accounting, propagate through each sequential stage of GGDI derivation. The figure is arranged in four panels—GRACE-GWA (top), Climatology (second), GSD (third), and GGDI (bottom)—with each panel displaying the multi-model ensemble (grey lines/shading) and its mean (green line). In the top panel, the GRACE-GWA time series is anchored to a January 2004–December 2009 baseline, rendering anomalies within this

reference window comparably subdued. Post-2009, however, the amplitude volatility increases, particularly between 2010–2012 when the aquifer recharged due to reported episodes of rapid snowmelt and intense precipitation (Dumanski et al., 2015; Coles et al., 2017). The divergence of ensemble members in both the magnitude and timing reflects the influence of model structural bias when disaggregating GRACE-TWSA into GRACE-GWA (Akl & Thomas, 2024; see also Chapter 2 of this thesis).

In the second panel (Climatology), the derived monthly mean captures the aquifer's seasonal cycle of storage increases and decreases, exposing significant discrepancies among estimates driven by divergent model assumptions. These discrepancies are particularly pronounced during seasonal peaks and troughs, where impediments in water partitioning led to pronounced variations in climatology amplitudes. In the third panel, the Groundwater Storage Deficit (GSD) quantifies deviation from the seasonal climatology, delineating intervals of groundwater surplus (positive GSD) or deficits (negative GSD). Notably, abrupt shifts—such as those noted in early 2010—coincide with notable expansions in the GSD ensemble spread, signalling heightened ambiguity in model water budget allocation.

Finally, in the bottom panel (GGDI), GSD values are standardized to allow comparisons of groundwater drought intensity. By converting absolute storage deficits to a z-score measure, GGDI is thought to reduce amplitude-driven uncertainty. Variability in GGDI, attributed to GRACE processing methodologies and water budget component selection, increases notably during abrupt hydrologic transitions, indicating heightened model water allocation discord under rapid change. For example, GGDI values range from 0.25 to -2.80 in January 2003, 0.15 to -2.75 in February 2010, 2.40 to -1.50 in March 2011, and 1.00 to -2.15 in June 2022. When the ensemble means hovers near zero (e.g., in 2010–2012 or 2019–2020), small deviations can shift apparent conditions between mild drought and near-normal, fundamentally altering drought classification.

4.3.2 Groundwater drought indicators across study basins

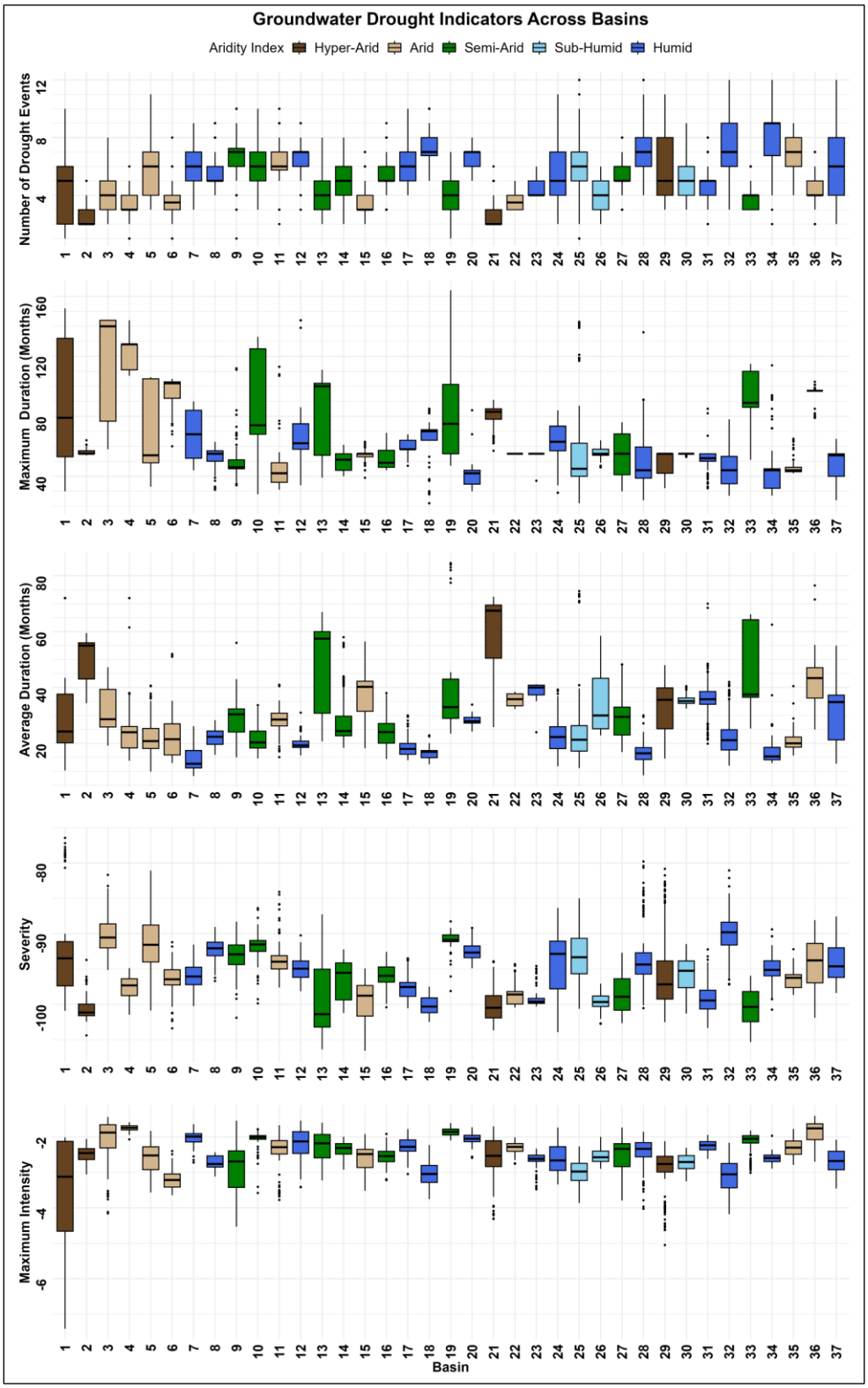


Figure 4.4 Groundwater drought indicators derived from GGDI time series for 37 study aquifers spanning April 2002 to December 2022. Five indicators are displayed across panels: the number of drought events, maximum duration, average duration, severity, and maximum intensity. Aquifers are ordered by basin number, with a colour scale indicating aridity class—hyper-arid, arid, semi-arid, sub-humid, and humid—to facilitate climate-based comparisons. The gap period between the GRACE and GRACE-FO missions was excluded from this analysis.

Basin	Climate zone	Range of variability (maximum – minimum)				
		Number of drought events	Maximum duration (m)	Average duration (m)	Severity	Max intensity
1	Hyper Arid	9	122	61.69	24.47	5.39
2	Hyper Arid	3	10	25.13	10.68	1.00
3	Arid	6	86	28.04	13.46	2.72
4	Arid	5	37	58.17	6.59	0.48
5	Arid	8	73	30.73	19.81	1.74
6	Arid	7	45	39.00	12.19	1.25
7	Humid	6	46	17.85	8.71	1.08
8	Humid	5	32	12.42	7.66	0.69
9	Semi-Arid	9	78	40.98	13.59	2.99
10	Semi-Arid	7	105	19.04	13.49	1.80
11	Arid	8	82	25.90	13.53	2.11
12	Humid	6	110	15.30	7.89	1.88
13	Semi-Arid	6	72	46.25	19.12	1.62
14	Semi-Arid	6	21	39.57	9.01	0.93
15	Arid	5	24	38.25	11.66	1.60
16	Semi-Arid	6	25	23.58	7.85	1.30
17	Humid	6	21	16.00	7.06	1.27
18	Humid	5	63	10.38	5.00	1.52
19	Semi-Arid	6	117	61.00	9.85	0.49
20	Humid	3	54	9.63	5.70	0.86
21	Hyper Arid	4	34	46.67	9.76	2.61
22	Arid	2	0	6.17	6.14	0.74
23	Humid	2	18	17.17	5.74	1.15
24	Humid	9	55	27.42	17.59	1.61
25	Sub humid	11	121	63.33	15.63	1.59
26	Sub humid	4	18	35.63	5.66	0.90
27	Semi-Arid	5	46	31.33	9.95	2.06
28	Humid	8	112	21.44	20.76	2.17
29	Hyper Arid	8	23	33.42	21.71	2.93
30	Sub humid	6	2	8.00	9.83	0.94
31	Humid	6	53	50.08	11.12	0.67
32	Humid	9	51	29.94	16.15	2.01
33	Semi-Arid	3	64	40.92	9.37	1.19
34	Humid	10	87	49.56	11.36	0.92
35	Arid	5	23	24.83	6.47	1.01
36	Arid	5	24	51.50	13.81	1.29
37	Humid	10	41	42.23	10.82	1.37

Table 4.3 Climate zones across study basins and the variability ranges of groundwater drought indicators, defined as the difference between the maximum and minimum values, including the number of drought events, maximum duration (months), average duration (months), severity, and maximum intensity. For comprehensive statistical details, please refer to Tables C2–C6 in Appendix C.

Figure 4.4 presents five key groundwater drought indicators for the 37 aquifer systems. The number of drought events exhibits substantial inter-basin variability, with ranges extending up to 11 events (Figure 4.4 and Table 4.3). For instance, Basin 25 (Yakut; Sub-Humid) displayed

the widest range—from 1 to 12 events, while Basin 22 (Indus; Arid) showed the narrowest, with counts between 3 and 5. In general, humid basins tend to register both a higher number of drought events and greater variability, whereas arid basins typically record fewer events with less variability. Nonetheless, notable outliers exist among arid basins: Basin 1 (Nubian) recorded a range of 1 to 10 events, Basin 5 (Taoudeni-Tanezrouft) ranged from 3 to 11, and Basin 29 (Tarim) also varied between 3 and 11 events.

Variability in maximum drought duration also differed markedly across basins, with ranges spanning from 0 to 122 months (Figure 4.4 and Table 4.3). For example, Basin 1 (Nubian; Hyper-Arid) exhibited the greatest variability, with maximum drought durations ranging from 30 to 152 months, whereas Basin 22 (Indus; Arid) demonstrated no variability, with a consistently recorded maximum drought duration of 55 months. Average drought duration followed a similar pattern, varying between 6 and 63.5 months. Notably, Basin 25 (Yakut; Sub-Humid) recorded the highest variability in average duration, ranging from 11 to 74.5 months, while Basin 22 (Indus; Arid) again exhibited the lowest variability, with average durations confined to a narrow band between 32.5 and 38.5 months. The overall findings suggest that arid basins tend to exhibit longer drought duration and greater variability in both maximum and average drought durations, in contrast to shorter drought duration and reduced variability observed in humid basins.

Drought severity revealed pronounced inter-basin differences (Figure 4.4 and Table 4.3). Basin 1 (Nubian; Hyper-Arid) exhibited the highest variability, with severity values ranging from -101 to -76.5 (a range of 24.5), whereas Basin 18 (Amazonas; Humid) showed the lowest variability, with values spanning from -102.5 to -97.5 (a range of 5.0). In general, arid basins tended to display high variability in drought severity—except for Basins 4 (Iulluemeden-Irhazer; Arid), 16 (Ogallala; Semi-Arid), 22 (Indus; Arid), and 35 (Canning; Arid), which exhibited narrower ranges of 6.5, 7.8, 6.1, and 6.5, respectively. Conversely, while humid basins typically exhibited lower variability, outliers were noted in Basins 24 (West Siberian), 28 (Middle Heilongjiang-Amur River), and 32 (Pechora), with ranges of 17.5, 20.8, and 16.1, respectively.

Maximum intensity revealed significant heterogeneity across the study basins (Figure 4.4 and Table 4.3). Notably, Basin 1 (Nubian; Hyper-Arid) exhibited the highest variability, with GGDI ranging from -2.0 to -7.5, a range of approximately 5.5 points. Following this apex of variability, other arid basins, specifically Basins 3 (Chad), 9 (Karoo), 21 (Arabian) and 29 (Tarim), displayed variability with ranges around 3 points. Meanwhile, Basins 4 (Iulluemeden-

Irhazer; Arid), 8 (Coango; Humid), 19 (Yrendá-Toba-Tarijeño; Semi-Arid), and 31 (Russian Platform; Humid) exhibited minimal variability, confined to roughly 0.50 points. The findings underscore that arid basins are distinguished by both elevated maximum intensity values and markedly greater variability, whereas humid basins consistently register lower maximum intensity values accompanied by reduced variability.

In humid basins, drought events occur more frequently, with a broad range of variability, yet these events tend to have relatively short and consistent durations and severity. In contrast, arid basins, while recording fewer events with a narrower variability range, often exhibit prolonged durations and more extreme intensity values, accompanied by high variability. Moreover, several outlier basins underscore the complex influence of multi-model GRACE-GWA variability on drought characterization.

4.3.3 Aquifer memory across study basins

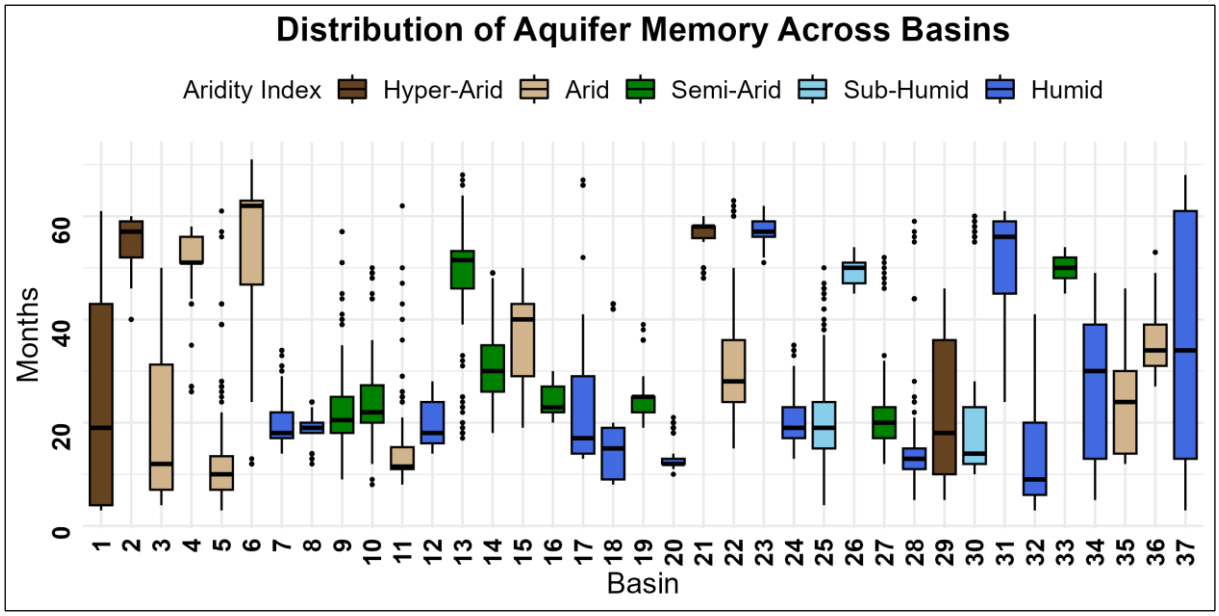


Figure 4.5 Box blots of aquifer memory (in months) across 37 study aquifers, ordered by basin number, with a colour scale indicating aridity class (hyper-arid, arid, semi-arid, sub-humid, and humid).

Basin	Climate zone	Aquifer memory statistics				
		Min	Q1	Median	Q3	Max
1	Hyper Arid	3	4	19	43	61
2	Hyper Arid	40	52	57	59	60
3	Arid	4	7	12	31.25	50
4	Arid	26	51	51	56	58
5	Arid	3	7	10	13.5	61
6	Arid	12	46.75	62	63	71
7	Humid	14	17	18	22	34
8	Humid	12	18	19	20	24
9	Semi-Arid	9	18	20.5	25	57
10	Semi-Arid	8	20	22	27.25	50
11	Arid	8	11	11.5	15.25	62
12	Humid	14	16	18	24	28
13	Semi-Arid	17	46	51.5	53.25	68
14	Semi-Arid	18	26	30	35	49
15	Arid	19	29	40	43	50
16	Semi-Arid	20	22	23	27	30
17	Humid	13	14	17	29	67
18	Humid	8	9	15	19	43
19	Semi-Arid	19	22	25	25	39
20	Humid	10	12	12	13	21
21	Hyper Arid	48	55.75	58	58	60
22	Arid	15	24	28	36	63
23	Humid	51	56	57	59	62
24	Humid	13	17	19	23	35
25	Sub humid	4	15	19	24	50
26	Sub humid	45	47	50	51	54
27	Semi-Arid	12	17	20	23	52
28	Humid	5	11	13	15	59
29	Hyper Arid	5	10	18	36	46
30	Sub humid	10	12	14	23	60
31	Humid	24	45	56	59	61
32	Humid	3	6	9	20	41
33	Semi-Arid	45	48	50	52	54
34	Humid	5	13	30	39	49
35	Arid	12	14	24	30	46
36	Arid	27	31	34	39	53
37	Humid	3	13	34	61	68

Table 4.4 Climate zones across study basins and the aquifer memory statistics, including minimum values (Min), lower quartile (Q1), median, upper quartile (Q3), and maximum values (Max).

Figure 4.5 displays aquifer memory estimates for the 37 study basins. These memory estimates are derived from a series of GGDI calculations based on multi-model GRACE-GWA estimates. Aquifer memory demonstrated striking inter-basin variability, ranging from 9 to 65 months (Figure 4.5 and Table 4.4). For example, Basin 37 (Paris; Humid) exhibited the broadest range—with estimates spanning from 3 to 68 months—whereas Basins 26 (North China Plains;

Sub-Humid) and 33 (Taurus-Zagros; Semi-Arid) showed much narrower ranges, varying from 45 to 54 months.

In general, hyper-arid and arid basins tended to display both higher aquifer memory values and greater variability. For instance, Basin 1 (Nubian) exhibited a variability range of 58 months (3 to 61 months), and Basin 6 (Senegalo-Mauretanian) recorded some of the highest values, ranging from 12 to 71 months. An intriguing outlier was Basin 21 (Arabian), which, despite having relatively high memory values (48 to 60 months), showed only a 12-month variability range—possibly reflecting the combined effects of minimal precipitation and intensive groundwater extraction for agricultural purposes (Siebert et al., 2010). In contrast, humid and sub-humid basins generally exhibited lower aquifer memory values with minimal variability. For example, Basin 12 (Karoo-Carbonate) recorded memory estimates ranging from 14 to 28 months—a narrow variability of just 14 months. Yet, outliers such as Basin 23 (Ganges-Brahmaputra) and Basin 26 (North China Plains), despite maintaining relatively narrow ranges, reported considerably higher memory values (51–62 and 45–54 months, respectively), likely reflecting the impact of dense populations and intense irrigation demand (Richey et al., 2015a). Notably, Basin 37 (Paris) emerges as an extreme case, demonstrating the most pronounced variability with a range spanning 65 months.

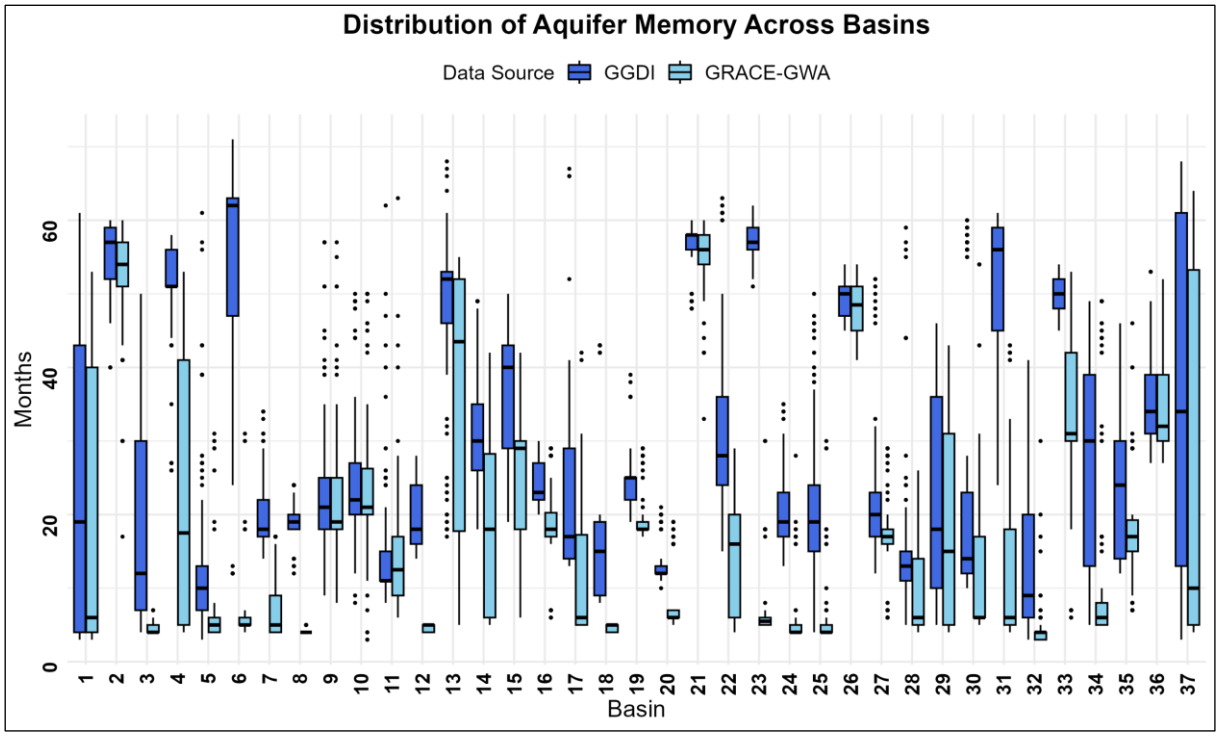


Figure 4.6 A comparison of aquifer memory (in months), calculated from GGDI and GRACE-GWA, across 37 study aquifers.

Removing the climatology signal within the GGDI framework exerts a profound and multifaceted influence on aquifer memory, significantly altering both its values and the range

of variability (Figure 4.6). For example, variability in Basin 18 (Amazonas; Humid) increased from 1 to 35 months, in Basin 3 (Chad; Arid) from 3 to 46 months, in Basin 12 (Karoo-Carbonate; Humid) from 1 to 14 months, and in Basin 8 (Coango; Humid) from 1 to 12 months. Conversely, the variability decreased markedly in Basin 33 (Taurus-Zagros; Semi-Arid), from 47 to 9 months; in Basin 23 (Ganges-Brahmaputra; Humid), from 25 to 11 months; in Basin 21 (Arabian; Hyper-Arid), from 27 to 12 months; in Basin 16 (Ogallala; Semi-Arid), from 23 to 10 months; and in Basin 2 (Northwest Sahara; Hyper-Arid), from 43 to 20 months. Notably, Basins 30 (North Caucasus; Sub-Humid) and 36 (Great Artesian; Arid) showed only minimal increases—from 49 to 50 and from 25 to 26 months, respectively. Furthermore, the range of aquifer memory values was profoundly affected in few basins; for example, aquifer memory range in Basin 23 expanded from 5–30 months before to 51–62 months after climatology removal. These results suggest that eliminating the climatology signal substantially amplifies the variability range of aquifer memory and significantly increases its values in humid basins, while exerting only minimal effects in arid regions (Figure C.39 in Appendix C).

4.3.4 Correlation analysis of aquifer memory and drought indicators.

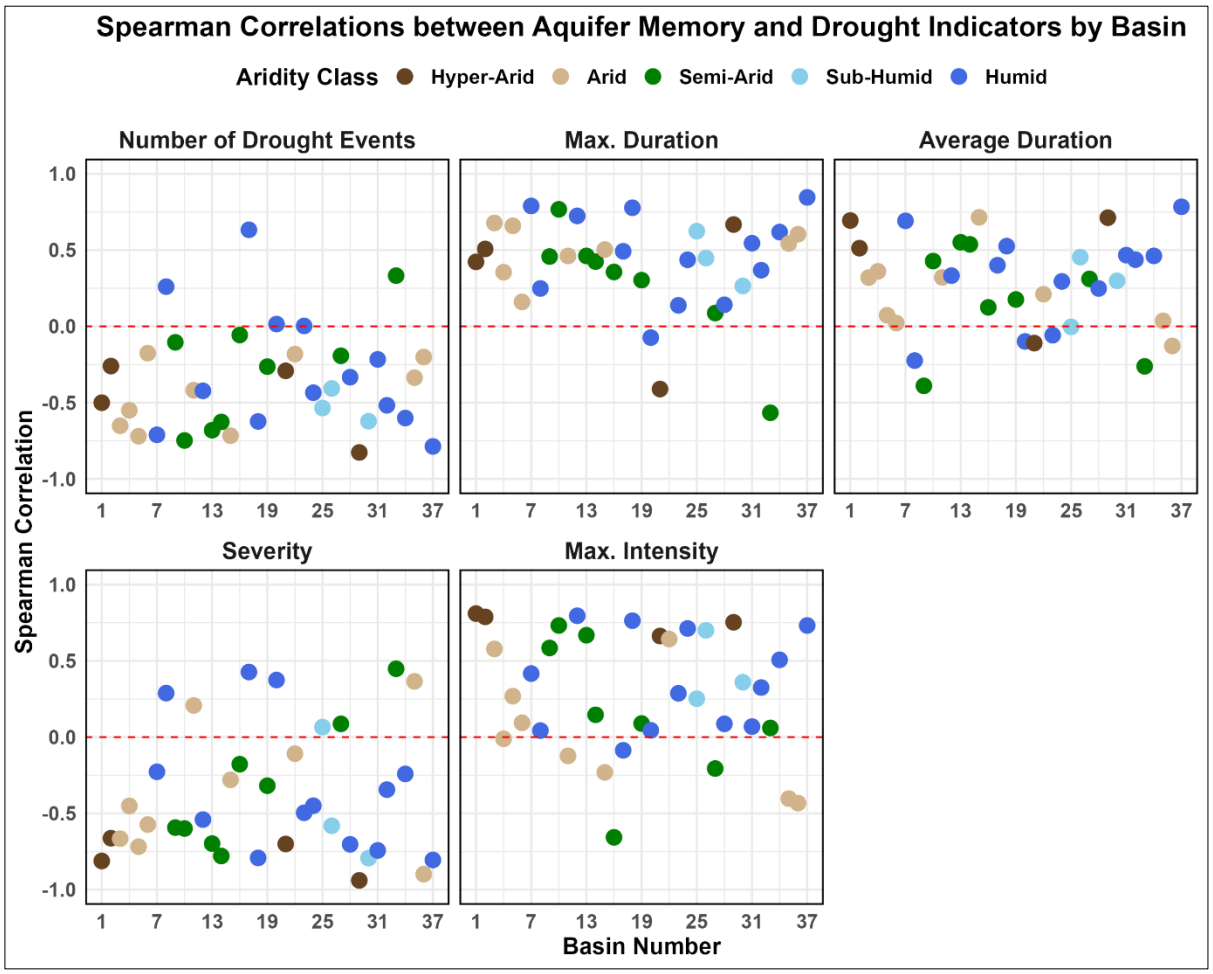


Figure 4.7 Spearman correlation coefficients depicting the relationship between aquifer memory and five drought indicators, including number of drought events, maximum duration,

average duration, severity, and maximum intensity, across 37 study basins, each ordered by basin number, with a colour scale indicating aridity class (hyper-arid, arid, semi-arid, sub-humid, and humid).

Spearman's correlation analysis was conducted to document the relation between aquifer memory and groundwater drought indicators across study basins (Figure 4.7). When comparing aquifer memory and drought event frequency, average Spearman correlation coefficients (ρ) were -0.77 for hyper-arid, -0.64 for arid, -0.56 for semi-arid, -0.64 for sub-humid, and -0.54 for humid basins. These negative correlations indicate that aquifer memory is generally linked to reduced drought event frequency, suggesting that aquifer memory may serve as an effective buffer against frequent drought occurrences. However, certain basins—such as Basin 8 (Coango; Humid), Basin 17 (Atlantic and Gulf Coastal Plains; Humid), and Basin 33 (Taurus–Zagros; Semi-Arid)—display positive correlations (0.26, 0.63, and 0.33, respectively).

In contrast, the relationship between aquifer memory and maximum drought duration generally shows a positive correspondence (Figure 4.7). Average correlations are ρ = 0.50 for hyper-arid, 0.49 for arid, 0.50 for semi-arid, 0.36 for sub-humid, and 0.30 for humid basins, indicating that while aquifer memory may hinder frequent droughts, it prolongs the maximum duration of drought events when they do occur. However, notable deviations emerge in specific cases—Basin 20 (Guarani; Humid), Basin 21 (Arabian; Hyper-Arid), and Basin 33 (Taurus–Zagros; Semi-Arid) exhibit negative correlations (-0.07, -0.41, and -0.57, respectively). A similar positive association is observed when comparing aquifer memory with mean drought duration, with average correlations of ρ = 0.77 for hyper-arid, 0.15 for arid, 0.50 for semi-arid, 0.21 for sub-humid, and 0.54 for humid basins. Yet again, significant outliers are evident: Basins 8 (Coango; Humid), 9 (Karoo; Semi-Arid), 20 (Guarani; Humid), 21 (Arabian; Hyper-Arid), 23 (Ganges–Brahmaputra; Humid), 33 (Taurus -Zagros), and 36 (Great Artesian) display negative correlations ranging from -0.05 to -0.41.

Drought severity generally exhibits a negative relationship with aquifer memory (Figure 4.7). Hyper-arid basins exhibit an average correlation of ρ = -0.80, arid basins ρ = -0.53, semi-arid basins ρ = -0.52, sub-humid basins ρ = -0.73, and humid basins ρ = -0.51. This pattern suggests that aquifer memory is capable of mitigating the severity of drought conditions. However, variability persists, as evidenced by basins such as Basin 8 (Coango; Humid), Basin 11 (Stampriet-Kalahari; Arid), Basin 17 (Atlantic and Gulf Coastal Plains; Humid), Basin 20 (Guarani; Humid), Basin 25 (Yakut; Sub-humid), Basin 27 (Song-Liao; Semi-Arid), Basin 33 (Taurus–Zagros; Semi-Arid), and Basin 35 (Canning; Arid), which exhibit correlations ranging from 0.06 to 0.45.

The analysis of maximum drought intensity reveals an overall positive association with aquifer memory (Figure 4.7). In hyper-arid basins, the average correlation is $\rho = 0.64$, while in arid, semi-arid, sub-humid, and humid basins, the average correlations are $\rho = 0.048$, $\rho = 0.27$, $\rho = 0.54$, and $\rho = 0.21$, respectively. This suggests that, in general, aquifer memory may contribute to higher peak drought intensities, possibly due to a delayed groundwater response that allows more severe water deficits to develop. Yet, this trend is not universal; significant outliers exist. For example, several arid basins—such as Basin 4 (Iulluemeden-Irhazer), Basin 11 (Stampriet-Kalahari), Basin 15 (Basin and Range Basin-fill), Basin 35 (Canning), and Basin 36 (Great Artesian)—along with Basin 16 (Ogallala; Semi-Arid), Basin 17 (Atlantic and Gulf Coastal Plains; Humid), and Basin 27 (Song-Liao; Semi-Arid), exhibit negative correlations ranging from -0.08 to -0.66, indicating that in these cases, aquifer memory may be associated with reduced maximum drought intensity.

4.4 Discussion

This study demonstrates that biases and variability in multi-model GRACE-GWA realizations, arising from differences among GRACE solutions and water budget assumptions, introduce uncertainty into GGDI interpretation (Figures 4.3, 4.4 and 4.5). These findings reinforce the concerns raised in the second chapter of this thesis, highlighting the necessity of a comprehensive evaluation of water budget components to enhance the accuracy and credibility of GRACE-groundwater based studies. Although normalisation within the GGDI framework mitigates some amplitude-related discrepancies, it cannot fully resolve model divergences, particularly under extreme hydrological events or near-threshold conditions where even minor shifts in ensemble spreads can alter drought classification (Figure 4.3). The broad uncertainty ranges observed in groundwater drought indicators (Figure 4.4) and aquifer memory (Figure 4.5) across study basins highlight critical deficiencies in our current knowledge of groundwater drought dynamics.

Previous GGDI studies (Wang et al., 2020b; Paredes-Trejo et al., 2021; Satish Kumar et al., 2021; Ali et al., 2022; Liu et al., 2022a; Wang et al., 2022a; Huang et al., 2023a; Zhang et al., 2024; Aon et al., 2024; Nandi & Biswas, 2024; Neves, 2024; Nigatu et al., 2024; Nikraftar et al., 2024; Song et al., 2024; Zheng et al., 2024; Liu et al., 2025) relied on a single-model approach to disaggregate GRACE-TWSA into GRACE-GWA. This strategy overlooks inherent model structure biases (Clark et al., 2008; Ruddell et al., 2019; Konapala et al., 2020) and underestimates the influence of surface water storage anomalies within the water balance equation (Kim et al., 2009; Han et al., 2009). In contrast, the multi-model analysis presented in this study reveals that discrepancies in amplitude and timing across different water budget

components datasets introduce substantial uncertainty into GGDI outcomes (Figure 4.4). Consequently, reliance on a single-model approach or even on an ensemble average for deriving GRACE-GWA embeds critical biases into groundwater drought assessments, thereby compromising their overall robustness and reliability. For example, in Basin 1 (Nubian), the ensemble average yields six drought events, a maximum duration of 70 months, an average duration of 14 months, a severity of -92.50, a maximum intensity of -2.90, and an aquifer memory of four months. However, when an ensemble spread is considered, these values vary dramatically ranging from 1 to 10 events, 30 to 152 months in maximum duration, 10 to 72 months in average duration, -76.50 to -101 in severity, -2.00 to -7.40 in maximum intensity, and 3 to 61 months in aquifer memory. Such pronounced divergences underscore the critical importance of employing multi-model GRACE-GWA realizations, which consider different GRACE solutions and various water budget component data, within the GGDI framework to understand the inherent uncertainty in groundwater drought assessments. Notably, the multi-model approach adopted here does not seek to select an optimal GRACE-GWA time series for GGDI analysis but takes the position that a multiple realization approach derives a comprehensive understanding of the groundwater system.

The considerable uncertainty in GGDI indicators observed across study basins highlights the challenge of accurately characterising groundwater drought conditions (Figure 4.4). The most pronounced variability includes a fluctuation of 11 drought events in Basin 25 (Yakut), a 63-month range in average drought duration within the same basin, and a significant 122-month range in maximum drought duration for Basin 1 (Nubian), in addition to severity fluctuations of 24.47 and a maximum intensity variation of 5.39. These pronounced discrepancies reveal fundamental limitations in groundwater drought assessment, emphasizing the challenges in achieving consistent and reliable drought classifications. Groundwater drought indicators are essential for informing resource allocation, aquifer recharge strategies, usage regulations, and long-term sustainability planning (e.g., Thomas et al., 2017b; Zhang et al., 2024; Liu et al., 2025). However, failure to account for GRACE-GWA variability increases the risk of misjudging groundwater deficits—potentially leading to either resource depletion due to underestimated drought severity or overly restrictive management actions that unnecessarily constrain water use.

The wide range of variability in aquifer memory estimates across study basins hampers the ability to anticipate groundwater response and implement efficient sustainable strategies (Figure 4.5). For example, Basin 1 (Nubian), Basin 3 (Lake Chad), Basin 6 (Senegalo-Mauretanian), and Basin 37 (Paris) exhibit aquifer memory ranges spanning from 3 to 61

months, 4 to 50 months, 12 to 71 months, and 3 to 68 months, respectively. Such variability makes it difficult to predict when an aquifer will recover from drought, complicating long-term planning. If aquifer memory is highly uncertain, it is unclear whether prolonged dry spells will quickly subside once wetter conditions return or persist for extended periods. This ambiguity can lead to either overconfidence or excessive caution in water-use restrictions.

The findings suggest a surprisingly consistent relationship between aquifer memory and various groundwater drought indicators across study basins (Figure 4.7) despite variability attributed to GRACE-GWA extraction. Generally, aquifer memory corresponds to fewer drought events and lower severity, while also correlating with extended drought durations and more intense peaks. However, certain basins depart from this pattern. For instance, Basin 33 (Taurus-Zagros; Semi-Arid) shows a contrasting trend, with a positive correlation between aquifer memory and drought frequency ($\rho = 0.33$), negative correlations with maximum ($\rho = -0.57$) and average ($\rho = -0.26$) drought duration, a positive correlation with drought severity ($\rho = 0.45$), and virtually no relationship with maximum drought intensity ($\rho = 0.06$). Likewise, Basin 21 (Arabian; Hyper-Arid) presents negative correlations between aquifer memory and both maximum ($\rho = -0.42$) and average ($\rho = -0.11$) drought duration. Basin 33 is predominantly populated rangeland, while Basin 21 comprises barren, remote rangeland interspersed with populated areas (Richey et al., 2015a). Both basins experience limited recharge and considerable groundwater pumping for agriculture, conditions that likely contribute to their atypical drought behaviour.

This study intentionally avoids relying on precipitation data to validate multi-model GGDI time series, as the objective is not to isolate a single optimal GRACE-GWA realization for drought characterization. Instead, it adopts a multiple-realization framework to promote a more nuanced understanding of groundwater system behaviour. Nevertheless, a supplementary comparison was undertaken between the ensemble of GGDI outputs and precipitation anomalies from the Global Precipitation Climatology Centre (GPCC) over the Central Valley aquifer system (Basin 14 in Figure 4.1). The comparison reveals considerable divergence among GGDI realizations, which complicates direct alignment with precipitation anomalies (Figure C.40, Appendix C). These findings underscore the value of ensemble-based analysis in capturing the structural uncertainty inherent in GRACE-informed groundwater drought assessments.

One limitation of this study arises from the spatial resolution mismatch between the employed water budget components and GRACE products, requiring interpolation to align their grid resolution. While necessary, this adjustment introduces an additional layer of uncertainty. Furthermore, to minimize uncertainties, the data gap between the GRACE and GRACE-FO

missions was excluded; however, this omission may inadvertently obscure critical information essential for a more comprehensive evaluation of groundwater drought indicators. Notably, the GRACE observation period was sufficient for aquifer memory calculations, ensuring that this gap did not affect the robustness of memory estimates. Although this study leveraged recent advancements in remote sensing to account for surface water across the study basins, data limitations persist in many regions where surface water bodies remain underrepresented, potentially influencing the calculations (Figure 4.2 and Figure C.1 in Appendix C). Furthermore, as with any normalised index, GGDI results are inherently dependent on the chosen study period. Any modification to this period could alter GGDI calculations, thereby impacting the estimated range of uncertainty.

4.5 Conclusions

Reliable groundwater drought assessment is paramount, given groundwater's critical role during periods of water scarcity (Hughes et al., 2012; Famiglietti, 2014; Castle et al., 2014). This study integrates multi-model GRACE-GWA, sourced from diverse GRACE-TWSA solutions and water budget data, within the GGDI framework across 37 study aquifers. By evaluating key five groundwater drought indicators, i.e. number of drought events, maximum duration, average duration, severity, and maximum intensity, the findings demonstrate that variability among multi-model GRACE-GWA introduces significant uncertainty to GGDI, thereby constraining the reliability of groundwater drought assessments (Figure 4.4). Furthermore, the exploration of aquifer memory, quantified via the autocorrelation in GGDI time series, reveals substantial uncertainty, limiting the ability to accurately capture aquifer response and recovery dynamics, complicating the development of long-term water sustainability strategies (Figure 4.5). Additionally, this study uncovered the relation between aquifer memory and groundwater drought indicators documenting that aquifer memory correlates with fewer drought events yet results in longer and more intense drought episodes when they occur (Figure 4.7). These findings underscore the inherent uncertainty in large-scale groundwater drought assessments and reinforce the necessity of employing a multi-model GRACE-GWA approach within the GGDI framework. By capturing variability, this approach enhances uncertainty awareness and improves the reliability of groundwater drought assessments.

Chapter 5. Toward Robust Groundwater Monitoring: Confronting the Complexities of GRACE-Derived Estimates

This chapter has been prepared in the format of a technical commentary intended for submission to the journal Groundwater.

5.1 Introduction

Over the past two decades, the Gravity Recovery and Climate Experiment (GRACE; Tapley et al., 2004) and its successor, GRACE Follow-On (GRACE-FO; Landerer et al., 2020), have revolutionized the monitoring of large-scale (≥~100,000 km^2) terrestrial water storage, offering transformative insights into the global hydrological cycle (Rodell & Reager, 2023). GRACE terrestrial water storage anomalies (GRACE-TWSA) represent integrated changes across all water storage compartments—including snowpack, soil water, surface water storage (i.e., rivers, reservoirs, lakes) and groundwater (GRACE-GWA) (Rodell & Famiglietti, 1999)—and are now routinely applied to estimate monthly changes in global groundwater. GRACE has played a pivotal role in documenting groundwater depletion (Strassberg et al., 2009; Voss et al., 2013; Castle et al., 2014; Chen et al., 2014; Huang et al., 2015; Richey et al., 2015a; Nanteza et al., 2016; Thomas & Famiglietti, 2019), evaluating groundwater sustainability (Thomas et al., 2017b; Ahmed et al., 2019; Wang et al., 2020a; Nikraftar et al., 2024) and assessing drought impacts on subsurface storage (Thomas et al., 2017a; Wang et al., 2020b; Liu et al., 2022a; Nandi & Biswas, 2024). Beyond advancing hydrological science, GRACE data have informed water policy decisions (Famiglietti et al., 2011) and deepened understanding of socio-ecological dynamics related to water scarcity and resilience (Wang et al., 2020a).

Nevertheless, critical challenges persist. GRACE's coarse spatial resolution (Alley & Konikow, 2015; Scanlon et al., 2016) constrains its utility for management-relevant applications (Alley & Konikow, 2015). Future missions may enhance resolution (Daras et al., 2024), which in cooperation with downscaling approaches (Girotto et al., 2016; Khaki et al., 2017; Li et al., 2019c; Fatolazadeh et al., 2022) may allow for small-scale management applications. Nonetheless, GRACE has provided valuable new information about global groundwater change (Castle et al., 2014; Richey et al., 2015a, b; Nanteza et al., 2016; Thomas et al., 2017a, b). Despite these advancements (Chen et al., 2016a; Frappart et al., 2018), accurately isolating groundwater dynamics from GRACE (i.e., deriving GRACE-GWA) remains challenging (Alley & Konikow, 2015; Akl & Thomas, 2023, 2024). A concurrent challenge is the insufficient comparison of GRACE-GWA against ground truth (e.g., Yeh et al., 2006; Rodell et al., 2007; Feng et al., 2013; Castle et al., 2014; Huang et al., 2016; Nanteza et al., 2016; Rateb

et al., 2020), an important factor given that GRACE is often used to document depletion/groundwater mining (Rodell et al., 2009; Feng et al., 2013; Voss et al., 2013; Castle et al., 2014; Chen et al., 2014; Huang et al., 2015; Richey et al., 2015a; Chen et al., 2016a; Nanteza et al., 2016; Frappart et al., 2018; Thomas & Famiglietti, 2019) in regions with sparse to non-existent in-situ groundwater observations. When combined with the uncertainty attributed to water budget component selection to isolate a representative GRACE-GWA signal from integrated GRACE-TWSA (Saxe et al., 2021; Akl & Thomas, 2023, 2024**)**, these challenges introduce significant uncertainties, diminishing the reliability of GRACE-based groundwater assessments.

Addressing these methodological challenges is critical given the ubiquitous, and potentially misguided, application of GRACE to assess stresses on groundwater resources (Taylor et al., 2013; Famiglietti, 2014; Haddeland et al., 2014; Richey et al., 2015a; Thomas et al., 2016, 2017a, b). This technical commentary explores the complexities inherent in isolating large-scale groundwater signal from GRACE. This technical commentary confronts these challenges to foster robust surveillance for one of the planet's most vital freshwater reserves.

5.2 The Selection of Water Budget Components

Extracting reliable groundwater anomalies (GRACE-GWA) from GRACE-TWSA hinges on the rigorous accounting of water budget components like snow water equivalent, soil moisture, and surface water storage within a water balance framework (Rodell & Famiglietti, 2002; Akl & Thomas, 2024; see also Chapter 2 of this thesis). In the absence of direct water budget measurements (Saxe et al., 2021; Akl & Thomas, 2024; see also Chapter 2 of this thesis), researchers typically rely on auxiliary datasets such as land surface models, hydrological models, reanalysis products, or remote sensing outputs. Auxiliary datasets to account for water budget components are susceptible to numerous distortions. For example, land surface models and hydrological models are prone to structural biases in water partitioning (Clark et al., 2008; Ruddell et al., 2019; Konapala et al., 2020). Reanalysis products vary depending on spatial resolutions, assimilation systems and data sources used, factors which can lead to consistent bias (Steri, 2004; Haddeland et al., 2011; Lindsay et al., 2014; Hassler & Lauer, 2021; Saxe et al., 2021). Early GRACE studies attempted to mitigate these inter-model discrepancies by averaging Global Land Data Assimilation System (GLDAS; Rodell et al., 2004) outputs at 1-degree resolution (e.g., Rodell et al., 2009; Shamsudduha et al., 2012; Castle et al., 2014; Richey et al., 2015a; Nanteza et al., 2016; Thomas et al., 2017b; Alshehri & Mohamed, 2023). However, this technique presupposes that model outputs are independent—even though many Land Data Assimilation System (LDAS) models rely on shared forcing datasets (Rodell et al.,

2004)—and can obscure biases embedded in individual model structures, ultimately undermining the reliability of the resulting groundwater estimates (Akl & Thomas, 2023, 2024).

Much of the GRACE groundwater literature has relied on a select few model frameworks for water budget accounting (Rodell et al., 2009; Shamsudduha et al., 2012; Voss et al., 2013; Castle et al., 2014; Richey et al., 2015a; Bhanja et al., 2016; Thomas et al., 2017b; Yin et al., 2017; Sarkar et al., 2020; Wang et al., 2020a; Alshehri & Mohamed, 2023; Shalby et al., 2023; Nenweli et al., 2024; Nikraftar et al., 2024; Zhao et al., 2024). Variation in model-derived amplitudes for snow, soil moisture and surface water can shift GRACE-GWA seasonality (Akl & Thomas, 2024; see also Chapter 2 of this thesis). These factors limit consequential analysis of GRACE-GWA, particularly for assessing short-term changes induced by climatic and anthropogenic factors (Thomas & Famiglietti, 2019; Thomas & Nanteza, 2023) or trends in groundwater storage (Shamsudduha & Taylor, 2020; Xu et al., 2021; Shao & Liu, 2023). More recent efforts (Rateb et al., 2020; Scanlon et al., 2021) that incorporate multiple models to improve groundwater storage estimates often overlook monthly storage variations that encapsulate short-term changes induced by climatic and anthropogenic factors (Thomas & Famiglietti, 2019; Thomas & Nanteza, 2023). Consequently, a comprehensive appraisal of each water budget component is indispensable for accurately characterising GRACE-GWA, thus ensuring more robust and reliable evaluations of groundwater dynamics (Akl & Thomas, 2023, 2024).

5.3 Accounting for Surface Water Storage Anomalies (SWA)

Surface water storage anomalies (SWA), encompassing temporally dynamic water stored in lakes, reservoirs, rivers, and floodplains, represent a fundamental yet often underrepresented component of the terrestrial water budget framework. Despite their importance, SWA are frequently deprioritized or excluded in GRACE-based groundwater assessments, leading to potential inaccuracies and incomplete characterizations of groundwater storage dynamics (Rodell et al., 2007, 2009; Strassberg et al., 2007; Voss et al., 2013; Chen et al., 2014; Richey et al., 2015a; Bhanja et al., 2016, 2020; Lezzaik & Milewski, 2017; Thomas et al., 2017b; Meghwal et al., 2019; Melati et al., 2019; Shamsudduha & Taylor, 2020; Mohamed et al., 2021; Xanke & Liesch, 2022; Nikraftar et al., 2024). This limited emphasis largely reflects the absence of a robust framework for incorporating SWA into water balance analyses, resulting in considerable methodological variability in estimating and integrating surface water contributions. Several commonly adopted approaches introduce significant sources of uncertainty. For instance, the use of runoff outputs from land surface models as proxies for SWA (e.g., Thomas et al., 2017b; Shamsudduha & Taylor, 2020) fails to capture the dynamics

of surface water bodies and overlooks anthropogenic regulation, such as reservoir operations (Haddeland et al., 2006; Hanasaki et al., 2006; Zhou et al., 2016). Likewise, assuming a constant lake area when converting water level variations to volumetric storage (e.g., Nanteza et al., 2016) neglects temporal fluctuations in lake extent and ignores bathymetric variability, introducing systematic biases into surface water storage estimates (Li et al., 2019b; Khazaei et al., 2022). Moreover, inconsistencies across methods used to upscale individual lake storage to basin-wide SWA exacerbate magnitude differences in SWA time series, propagating biases into GRACE-GWA estimates (Akl & Thomas, 2024; see also Chapter 2 of this thesis).

While most efforts have focused on lakes and reservoirs, the omission of river water storage remains a critical oversight, especially in river-dominated basins where seasonal discharge, floodplain inundation, and channel storage represent major components of the water balance. In such settings, excluding river storage can lead to an overestimation of groundwater anomalies, as GRACE-TWSA will reflect surface contributions that are not otherwise accounted for in the water budget (Shamsudduha et al., 2012). This concern is particularly acute in tropical and monsoonal systems, where river stage and floodplain extent can fluctuate considerably over short time scales.

Compounding these challenges is the scarcity of in situ observations detailing the dynamics of surface water storage, namely, water levels, bathymetry, and areal extent (Duan & Bastiaanssen, 2013; Busker et al., 2019). When SWA is inadequately represented, GRACE-based estimates tend to reflect changes in overall available water, encompassing both surface and subsurface contributions (Castle et al., 2014; Thomas & Nanteza, 2023), rather than isolating a groundwater-specific signal, compromising the reliability and interpretability of groundwater assessments (Akl & Thomas, 2024; see also Chapter 2 of this thesis).

Advances in remote sensing technologies have begun to address these limitations by enabling consistent global monitoring of surface water dynamics. Satellite altimetry, optical sensors, and radar observations now provide increasingly accurate measurements of water level fluctuations and surface extent (Alsdorf et al., 2007; Huang et al., 2018; Busker et al., 2019; Tortini et al., 2020; Cooley et al., 2021; Hou et al., 2022). In parallel, improvements in bathymetric datasets have enhanced the ability to estimate lake and reservoir volumes with greater accuracy (Yigzaw et al., 2018; Li et al., 2019b; Khazaei et al., 2022). A growing suite of global databases—including G-REALM (Tortini et al., 2020), GRanD (Lehner et al., 2011), HydroLAKES (Messager et al., 2016), GLCP (Meyer et al., 2024), and GLWD (Lehner & Doll, 2004; Lehner et al., 2024), now facilitate temporal tracking of surface water bodies, while repositories such

as ReGeom (Yigzaw et al., 2018), GLWS (Gerdener et al., 2023), and GloLakes (Hou et al., 2024) integrate water level and bathymetric information, supporting more accurate volumetric estimations. Crucially, emerging technologies such as SWOT (Surface Water and Ocean Topography) offer the potential to directly observe river widths, slopes, and water surface elevations, paving the way for more accurate river storage quantification at global scales.

The integration of such datasets into the water balance framework will enable more reliable separation of surface and subsurface storage dynamics, thereby improving the precision and policy relevance of GRACE-based groundwater assessments. Incorporating SWA, including lake, reservoir, and river storage, within a structured and consistent water balance framework is essential for reducing uncertainty and enhancing the hydrological fidelity of GRACE-GWA estimates.

5.4 GRACE-GWA Comparison

A fundamental pillar of GRACE's utility for large-scale groundwater assessments lies in rigorous validation against well-documented aquifer systems. Traditionally, validation efforts have relied predominantly on correlation-based metrics like the Pearson correlation coefficient or Spearman's rank correlation to gauge the alignment between GRACE-GWA and in-situ groundwater observations (e.g., Yeh et al., 2006; Rodell et al., 2007; Sun et al., 2010; Cao et al., 2012; Shamsudduha et al., 2012; Castle et al., 2014; Panda & Wahr, 2015; Huang et al., 2016; Yin et al., 2017; Rateb et al., 2020; Alghafli et al., 2023; Amiri et al., 2023; Nenweli et al., 2024; Rzepecka et al., 2024). While widely adopted and straightforward to apply, these metrics offer limited diagnostic depth, as they quantify statistical association but fail to reflect essential features of hydrological time series, such as amplitude discrepancies, seasonal phase shifts, and the influence of outliers (Legates & Davis, 1997; Legates & McCabe, 1999; Akl & Thomas, 2023). Consequently, exclusive reliance on correlation metrics risks masking meaningful variations in groundwater behaviour, undermining the robustness of GRACE-GWA evaluations and potentially leading to misrepresentations of groundwater storage dynamics (Alley & Konikow, 2015; Akl & Thomas, 2023).

In contrast, hydrological modelling communities have long recognised the importance of more comprehensive goodness-of-fit (GOF) metrics for time series evaluation (Nash & Sutcliffe, 1970; Gupta et al., 2009; Althoff et al., 2021; Mathevet et al., 2023). Metrics such as the Nash–Sutcliffe Efficiency (NSE; Nash & Sutcliffe, 1970) and Kling–Gupta Efficiency (KGE; Gupta et al., 2009) incorporate both additive and proportional error components, providing a more rigorous and hydrologically relevant evaluation of time series behaviour (Gupta et al., 2009;

Althoff et al., 2021). When embedded within a multi-objective comparative framework, these metrics capture key structural attributes of hydrological signals, such as timing, amplitude, and variability, that are often overlooked by correlation-based approaches (Akl & Thomas, 2023). Their application offers a richer diagnostic perspective on GRACE-GWA performance and facilitates more informed assessments of its ability to capture true groundwater storage dynamics. Although both NSE and KGE have been subject to critique, particularly regarding sensitivity to bias and skewness (Legates & McCabe, 1999; Krause et al., 2005; Kling et al., 2012; Pool et al., 2018; Knoben et al., 2019; Althoff & Rodrigues, 2021), their capacity to encapsulate multiple dimensions of time series similarity makes them valuable tools for identifying discrepancies introduced by GRACE-TWSA processing and water budget component selection (Akl & Thomas, 2023). As such, they provide a robust basis for isolating GRACE-GWA realisations that most faithfully represent true groundwater storage dynamics, marking a critical methodological shift that enhances the transparency, reliability, and hydrological relevance of GRACE-based groundwater assessments (Akl & Thomas, 2023).

5.5 Implausible GRACE-GWA estimates

A critical source of uncertainty in GRACE-GWA estimates arises from amplitude mismatches between GRACE-TWSA and the auxiliary water budget components, specifically SMA, SWEA, and SWA, that are subtracted to isolate the groundwater signal. In certain configurations, the combined amplitude of these components exceeds that of the GRACE-TWSA signal itself, resulting in physically implausible GRACE-GWA estimates, including large negative anomalies or exaggerated variability not supported by hydrological processes (Shamsudduha and Taylor, 2020; Arifin et al., 2025).

These inconsistencies typically reflect a combination of factors, including differences in temporal misalignment and underlying model assumptions. For example, surface water estimates may assume constant lake extent, or soil moisture retrievals may underrepresent saturation dynamics in deeper layers. When left unaddressed, these mismatches can obscure the actual behaviour of groundwater systems and compromise the interpretability of GRACE-based assessments.

However, such discrepancies also present an opportunity to inform quality control in multi-model GRACE-GWA accounting. In data-scarce regions lacking in-situ groundwater observations for validation, the physical plausibility of GRACE-GWA estimates, particularly their amplitude relative to the driving TWSA signal, can serve as a diagnostic constraint. Component combinations that consistently yield anomalous amplitudes or violate expected

hydrologic balances can be flagged and excluded from further analysis. This strategy allows for refinement of multi-model realizations, improving the robustness and credibility of GRACE-GWA outputs in regions where direct validation is not feasible.

5.6 Concluding Remarks

Reliable groundwater assessments are indispensable for characterising aquifer dynamics under the compounded pressures of climate change and growing anthropogenic water demands (Taylor et al., 2014; Thomas & Famiglietti, 2019). Despite the growing reliance on GRACE-GWA to support large-scale assessments, their reliability remains constrained by insufficient validation against in-situ observations and persistent uncertainties in the selection and representation of water budget components (Akl & Thomas, 2023, 2024). These methodological shortcomings can obscure key hydrological signals and undermine the interpretive value of GRACE-based analyses. Addressing these limitations is critical for improving the scientific integrity, operational robustness, and policy relevance of satellite-based groundwater monitoring frameworks.

This technical commentary introduces three interrelated methodological enhancements that collectively advance the diagnostic power of GRACE-GWA assessments. First, the implementation of multi-model disaggregation frameworks enables more comprehensive representation of water budget variability, capturing structural uncertainties and reducing bias in groundwater estimates. Second, the rigorous accounting for SWA—including lakes and reservoirs—improves the separation of surface and subsurface water signals, enhancing the fidelity and hydrological realism of GRACE-derived outputs. Third, the incorporation of multi-objective GOF metrics into comparative frameworks transcends the limitations of conventional correlation-based methods, offering a more nuanced assessment of GRACE-GWA comparison. Fourth, the explicit screening of physically implausible GRACE-GWA estimates, which arise from amplitude inconsistencies between GRACE-TWSA and auxiliary water budget components, introduces an additional layer of constraint that can help refine model selection in basins where in-situ validation data are sparse or unavailable.

Collectively, these advancements, multi-model disaggregation, rigorous SWA accounting, multi-objective evaluation, and plausibility screening, form a cohesive and resilient framework for GRACE-based groundwater assessment. By enhancing methodological transparency, quantifying uncertainty, and strengthening interpretability, this framework offers a critical step toward more informed, data-driven groundwater monitoring in the face of intensifying environmental and societal pressures.

Chapter 6. Conclusions and Future Work

This thesis is guided by a unifying objective: to enhance the reliability and diagnostic value of GRACE-based groundwater assessments by addressing uncertainties arising from the representation and integration of auxiliary water budget components. While GRACE and GRACE-FO have transformed large-scale groundwater monitoring, their reliance on ancillary data to resolve non-groundwater storage components introduces significant uncertainty. Despite widespread awareness of structural biases in model-based water partitioning, GRACE-based studies have often remained limited to a small set of models, without adequately accounting for inter-model variability and its implications for groundwater signal interpretation. Although recent studies have incorporated multiple model outputs to assess long-term groundwater trends, they frequently overlook monthly-scale dynamics that reflect short-term variability driven by climate fluctuations and human withdrawals. Compounding these challenges is the widespread reliance on correlation-based metrics to evaluate GRACE-GWA comparison, which, while convenient, offer only a partial view of time series behaviour—failing to account for differences in magnitude, seasonal amplitude, and temporal coherence. As a result, these practices can obscure meaningful hydrological signals, reduce comparison efficacy, and embed systematic bias into GRACE-GWA interpretations. Crucially, such biases may propagate into dependent applications—such as the characterisation of groundwater drought—ultimately undermining the scientific robustness and practical relevance of GRACE-derived groundwater insights. In response, this thesis makes three key contributions: (1) it systematically elucidates how the selection of auxiliary water storage components introduce bias and variability in GRACE-GWA estimates; (2) it applies a multi-objective comparison framework that integrates Nash–Sutcliffe Efficiency (NSE) and Kling–Gupta Efficiency (KGE), offering a more comprehensive and hydrologically meaningful evaluation of GRACE-GWA comparison; and (3) it demonstrates how these uncertainties influence the interpretation of groundwater drought through a re-assessment of the GRACE-Groundwater Drought Index (GGDI) across hydrologically diverse aquifer systems.

6.1 Conclusions

This thesis presents a substantive contribution to the advancement of GRACE-based groundwater assessments by systematically addressing the methodological uncertainties that compromise the accuracy, interpretability, and policy relevance of its outputs. Central to the findings is the recognition that the selection and combination of auxiliary water budget components introduce substantial variability and bias into GRACE-GWA *(Chapter 2)*. These uncertainties affect not only storage trends but also seasonal signal characteristics, thereby

influencing interpretations of groundwater dynamics across multiple scales *(Chapters 3)*. Importantly, the propagation of these uncertainties into derived applications—specifically the GRACE-Groundwater Drought Index (GGDI)—raises concerns about the robustness of groundwater drought diagnostics if model-driven variability is not explicitly accounted for *(Chapter 4).*

A key insight emerging from this work is the underappreciated role of surface water storage—particularly lakes and reservoirs—in shaping GRACE-GWA outputs. Although often omitted or simplified in previous studies, this component was shown to meaningfully enhance alignment with in-situ groundwater observations when properly incorporated *(Chapter 2).* This finding underscores the imperative for comprehensive, regionally tailored water budget accounting to strengthen the hydrological credibility of GRACE- based assessments *(Chapter 2).*

The thesis also challenges the continued reliance on traditional correlation-based validation approaches, which frequently mask meaningful discrepancies across multi-model realisations when compared against basin-scale in-situ data *(Chapter 3).* In response, a multi-objective comparison framework, integrating NSE and KGE metrics, was implemented. This framework captures critical facets of time series behaviour, including signal magnitude, variability, and seasonal coherence, offering a more nuanced and diagnostically rich evaluation of multi-model GRACE-GWA comparison *(Chapter 3).*

The research further reveals the critical sensitivity of GRACE-GWA comparison to uncertainty in specific yield (Sy), a pivotal, yet frequently under characterised, aquifer parameter *(Chapter 3).* Poorly constrained or generalized Sy values systematically bias the outcomes of multi-objective evaluation frameworks, confounding the alignment between GRACE-derived and in-situ groundwater time series. This underscores the necessity of robust, site-specific a priori knowledge of aquifer properties to ensure the integrity and hydrological validity of GRACE-GWA comparison *(Chapter 3).*

Taken together, the findings demonstrate the critical value of integrating multi-model disaggregation with multi-objective evaluation to improve the precision, robustness, and interpretive confidence of GRACE-based groundwater assessments. This dual-framework approach enables a more nuanced understanding of inter-model variability and systematically reduces structural bias introduced by auxiliary components, thereby enhancing the ability to isolate realisations that most accurately represent observed groundwater behaviour *(Chapter 3)*. Importantly, the benefits of this framework extend beyond data-rich environments: even in

regions where in-situ observations are sparse or absent, the multi-model disaggregation strategy alone provides a defensible means of quantifying GRACE-GWA uncertainty. As evidenced in *Chapter 4*, this capability is particularly valuable for advancing reliable groundwater monitoring and drought diagnostics across hydrologically diverse and data-limited basins.

Lastly, the research provides insight into the relationship between aquifer memory and groundwater drought behaviour. Results indicate that aquifers with higher memory exhibit fewer drought events; however, when droughts do occur, they are generally more persistent and intense *(Chapter 4)*.

In sum, this thesis establishes a rigorous and scalable methodological foundation for future GRACE-based groundwater research. By confronting key sources of uncertainty *(Chapter 5)*, it enhances the transparency, reliability, and operational utility of GRACE-GWA products—paving the way for more informed and resilient water resource management in an era of accelerating climate and anthropogenic stress.

6.2 Future Research Directions

- Advancing Multi-Model GRACE-GWA Frameworks through Model-Specific Uncertainty Integration

Future efforts should focus on incorporating model-specific uncertainty into the multi-model GRACE-GWA frameworks. In this study, GRACE-GWA estimates were derived by combining single deterministic outputs from individual models, capturing structural variability across combinations but omitting uncertainty internal to each component. As a result, the ensemble reflects inter-model spread while neglecting intra-model sensitivity. A more rigorous approach would involve generating internal ensembles by perturbing model parameters, initial conditions, and meteorological forcings. This would yield component-specific probability distributions that can be propagated through the water balance equation using analytical or numerical uncertainty propagation techniques. Such a dual-layered uncertainty framework—capturing both intra- and inter-model variability—would enable structured error decomposition, improved uncertainty attribution, and the derivation of statistically robust confidence intervals.

- Enhancing the Reliability of In-situ Benchmarks for Multi-objective Validation

The robustness of GRACE-GWA multi-objective evaluation frameworks hinges on the integrity of the in-situ groundwater observations used as references. Yet, in-situ measurements—by nature localized and point-based—are prone to spatial and temporal distortions stemming from instrumentation and human errors, site-specific heterogeneities,

discontinuities in data coverage, and external influences such as groundwater abstraction and surface–water interactions. While this study implemented structured well selection and robust interpolation to produce basin-scale time series, the broader sensitivity of GRACE-GWA validation outcomes to in-situ data quality remains insufficiently addressed. Future research should pursue formalized, data-driven strategies to identify high-fidelity observation wells that could accurately produce basin-scale variability while minimizing localized noise. AI-driven techniques, including unsupervised clustering, anomaly detection, and supervised classification informed by geostatistical and temporal features, offer a powerful means to refine well selection criteria. Embedding such methods into validation workflows would markedly improve the transparency, coherence, and diagnostic strength of satellite-based groundwater assessments.

- AI-Driven Knowledge Transfer for Constraining GRACE-GWA in Data-Limited Regions

A persistent challenge in GRACE-based groundwater assessment is the lack of in-situ observations in many regions, which limits the ability to evaluate and constrain multi-model GRACE-GWA realizations. To address this, future research should explore AI-driven knowledge transfer approaches that leverage information from data-rich basins to inform groundwater estimation in data-scarce environments. State-of-the-art machine learning models—such as spatiotemporal neural networks, graph-based architectures, or transformer models—can be trained on well-instrumented basins to learn complex relationships between GRACE-GWA ensemble characteristics and hydrological plausibility, guided by in-situ benchmarks. Embedding physics-informed constraints (e.g., mass conservation, hydrological closure, and regional climate coherence) ensures physical consistency while enhancing generalizability. Once trained, these models can be transferred to data-poor regions through domain adaptation techniques, enabling the selection or weighting of GRACE-GWA realisations that most plausibly represent actual groundwater dynamics. This AI-enhanced knowledge transfer paradigm offers a scalable, data-efficient, and physically coherent solution for improving the robustness and interpretability of GRACE-based groundwater assessments in observation-limited regions.

- Re-evaluating Foundational GRACE-GWA Studies with Multi-model, Multi-objective Frameworks

Given the methodological advances highlighted in this thesis—and the growing dependence on GRACE-derived insights for groundwater resource evaluation—there is a compelling need to revisit foundational GRACE-GWA studies. Many efforts relied on limited model ensembles and correlation-based validation approaches, often overlooking critical uncertainties in water budget component data selection. These limitations may have contributed to overly simplified,

and potentially misleading, interpretations of groundwater trends, seasonality, and drought severity in key regions. Future work should systematically apply expanded multi-model ensembles and rigorous multi-objective frameworks to reassess and refine these foundational conclusions. Doing so will not only increase the scientific rigor of GRACE-based analyses but also enhance their relevance, credibility, and applicability for long-term groundwater sustainability planning under intensifying climate and anthropogenic stress.

Appendix A. Supporting Information for Chapter 2

Contents of this appendix:

Introduction:

This supporting information includes 40 figures and 10 tables that complement the main analysis of chapter (2) by providing detailed insights into the in-situ groundwater observations and the water budget components across the three primary comparative basins. Additionally, it presents extended results for GRACE-TWSA, associated water budget terms, and cross-correlation analyses of GRACE-GWA across additional basins examined in the chapter.

Figures:

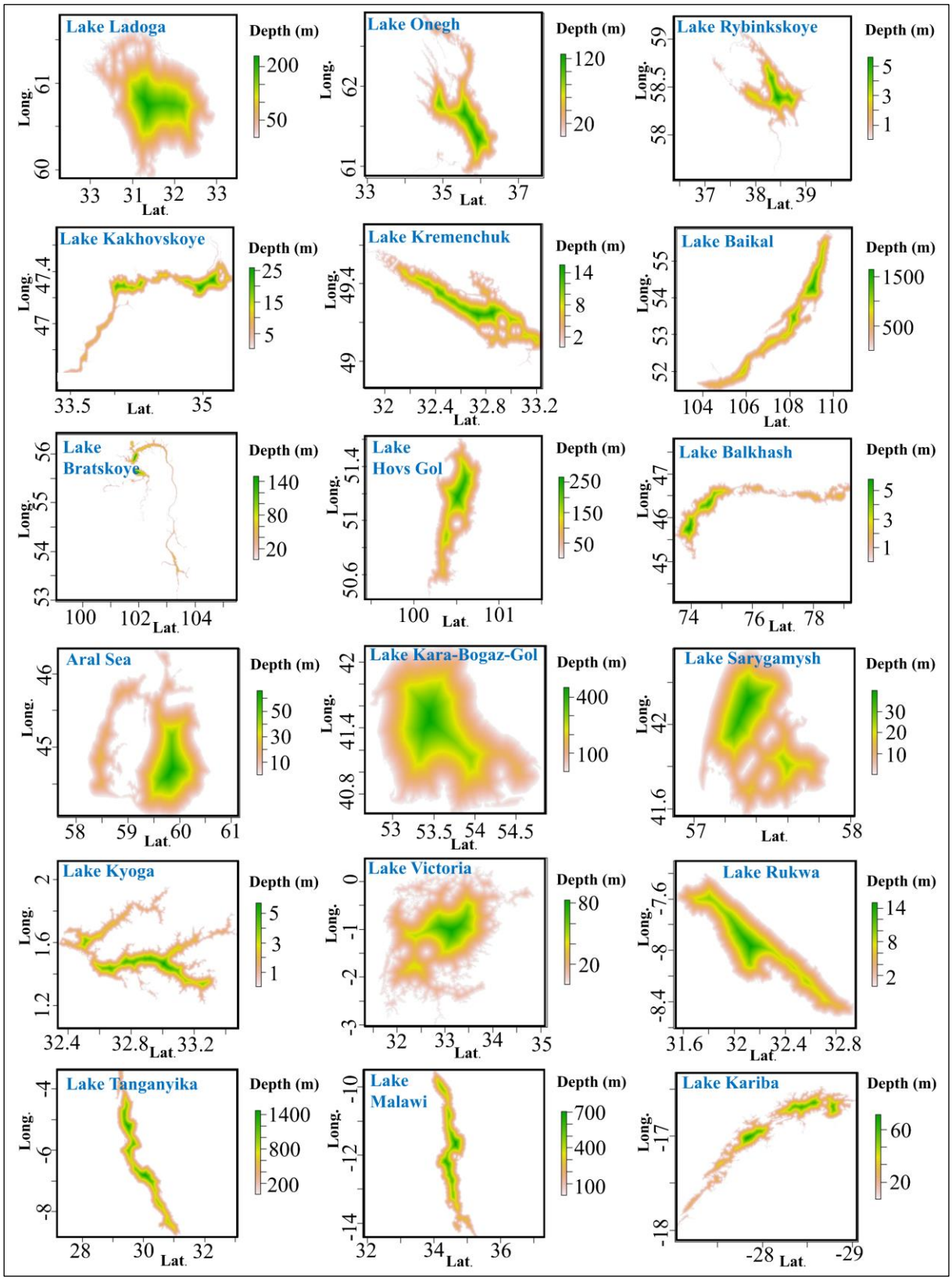


Figure A.1 Bathymetry data in resolution of 1 arc-seconds and WGS84 projection system for individual study lakes across basins (4:15) (Khazaei et al., 2022). Variability in bathymetry and lake area are notable in spatial and depth scales.

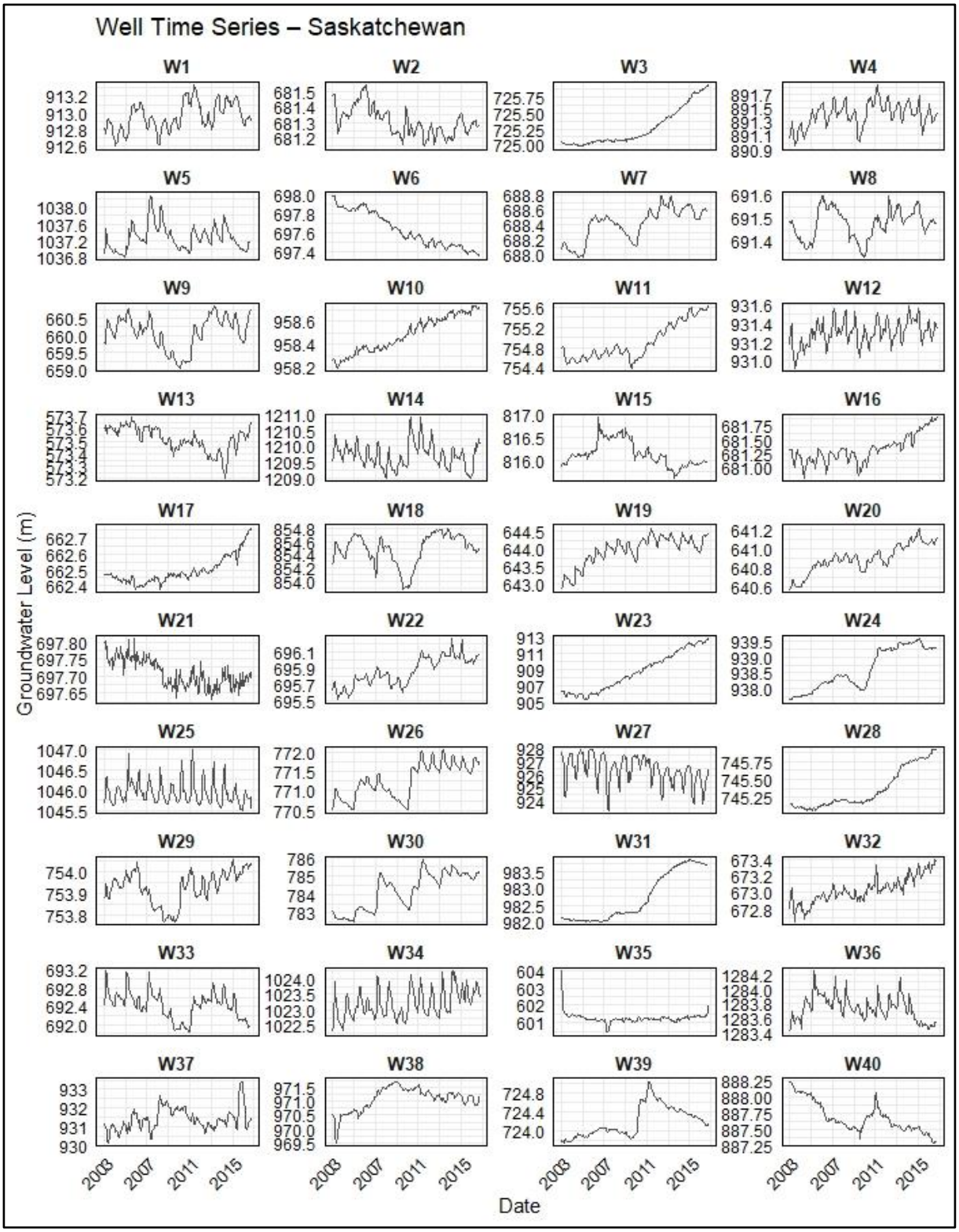


Figure A.2 Monthly groundwater level time series (in meters) for individual wells (W1–W40; see Table A.2) across the Saskatchewan basin over the study period from March 2003 to December 2016.

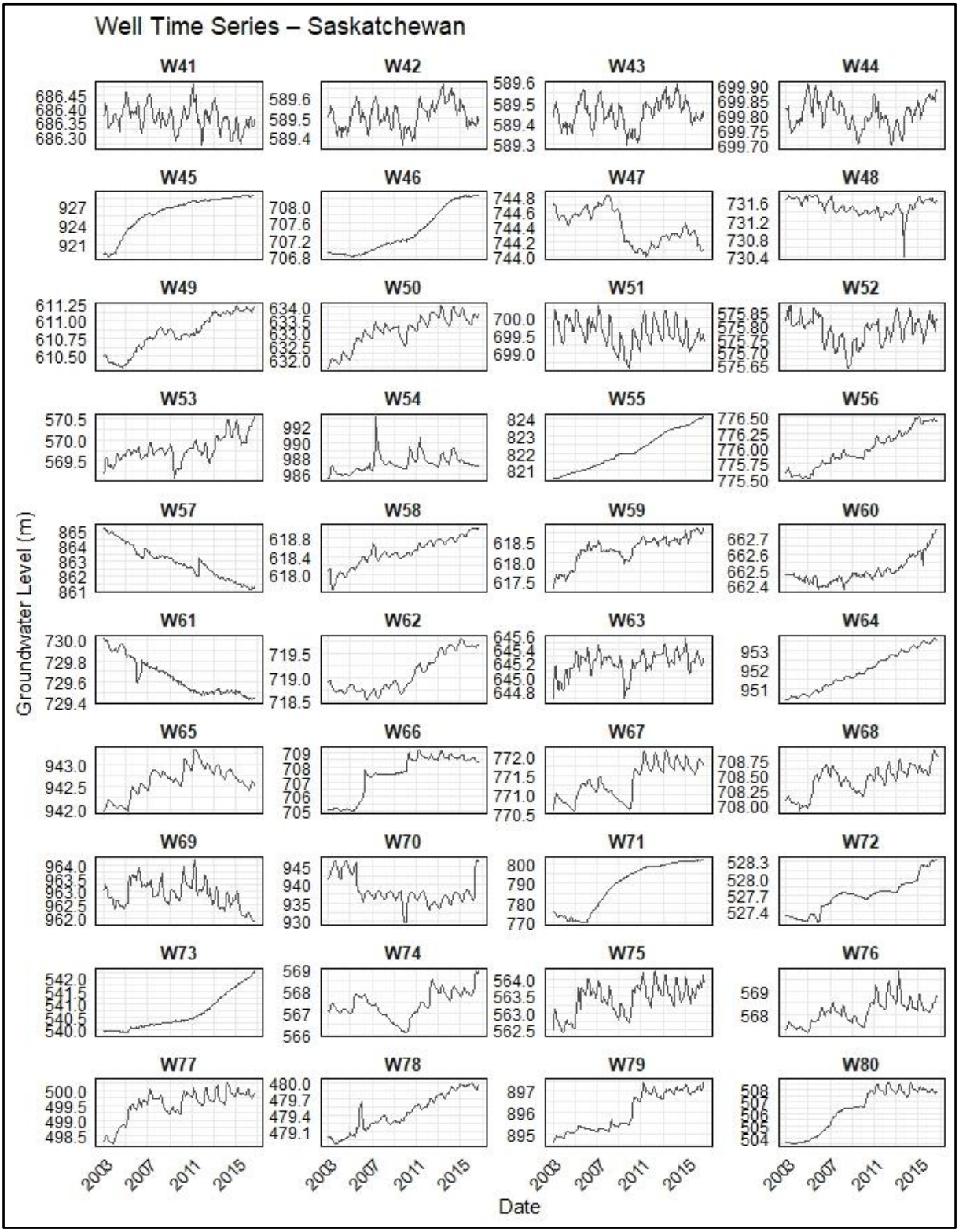


Figure A.3 Monthly groundwater level time series (in meters) for individual wells (W41–W80; see Table A.2) across the Saskatchewan basin over the study period from March 2003 to December 2016.

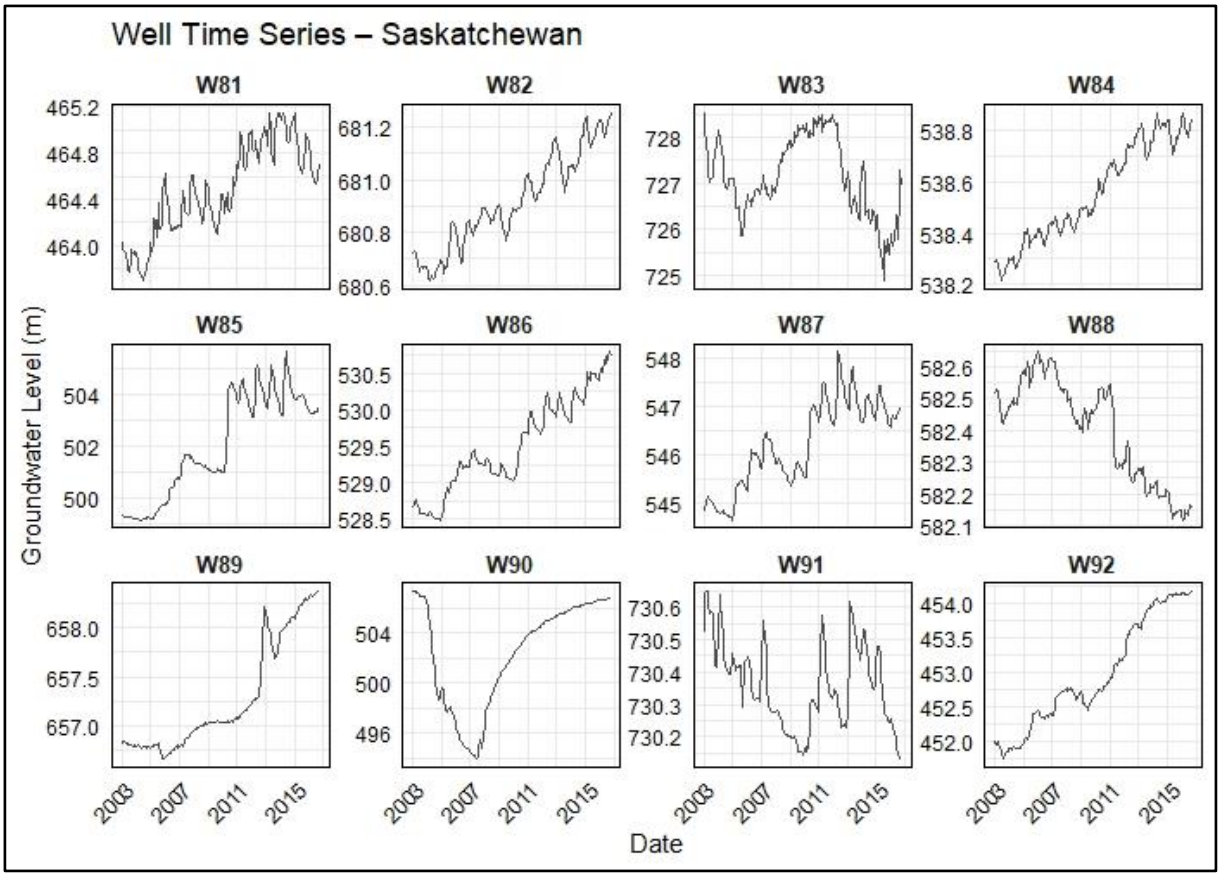


Figure A.4 Monthly groundwater level time series (in meters) for individual wells (W81–W92; see Table A.2) across the Saskatchewan basin over the study period from March 2003 to December 2016.

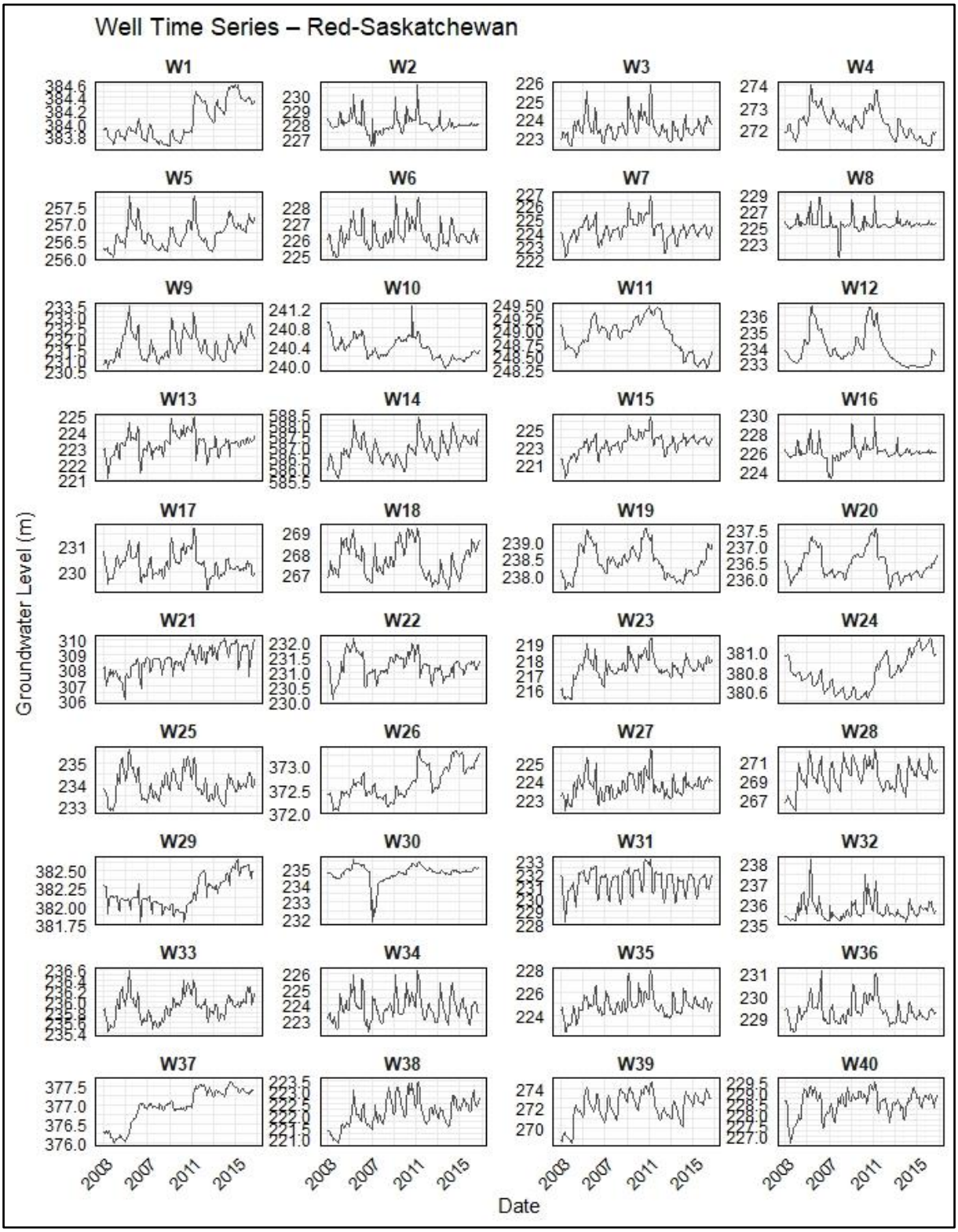


Figure A.5 Monthly groundwater level time series (in meters) for individual wells (W1–W40; see Table A.2) across the Red-Saskatchewan basin over the study period from March 2003 to December 2016.

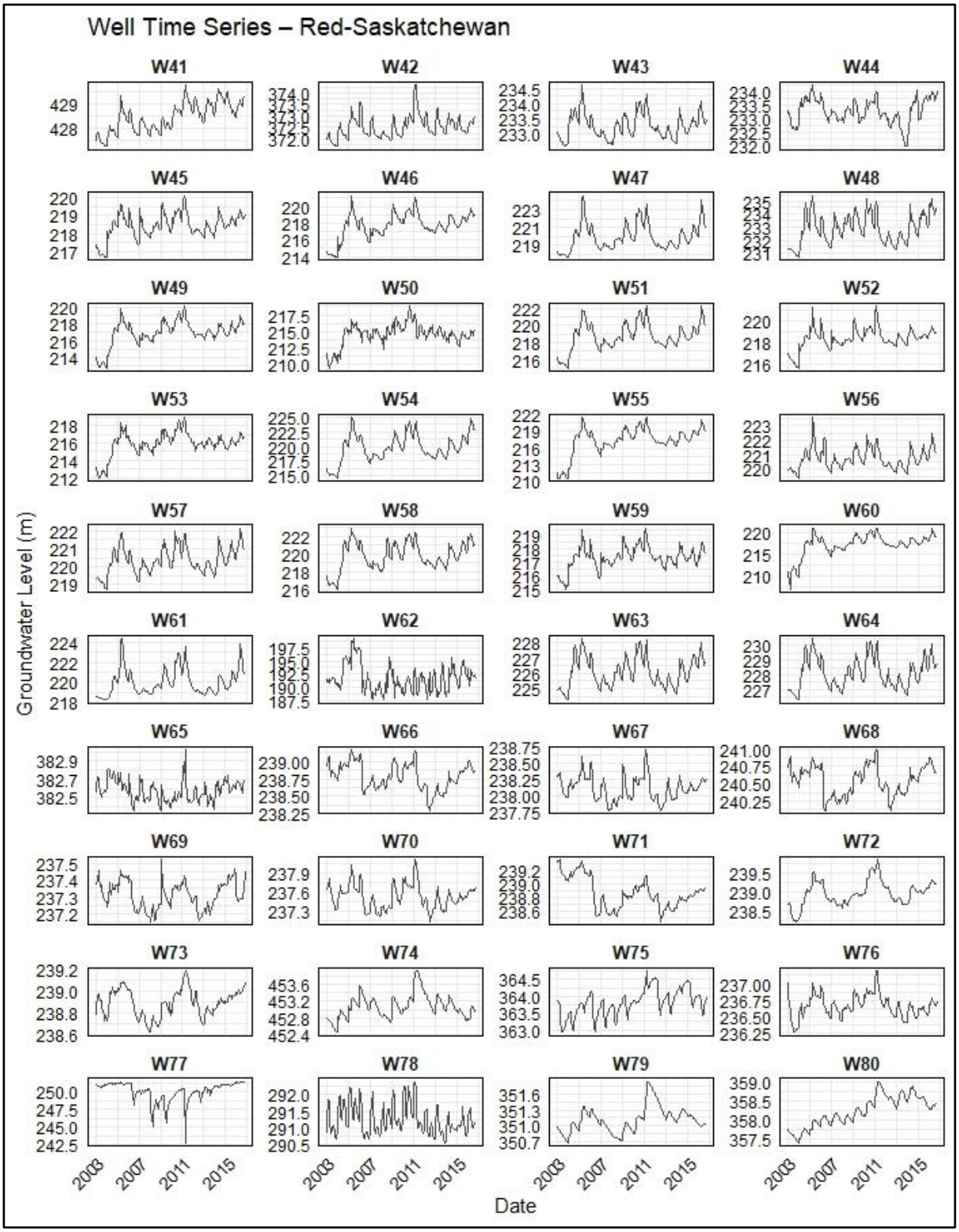


Figure A.6 Monthly groundwater level time series (in meters) for individual wells (W41–W80; see Table A.2) across the Red-Saskatchewan basin over the study period from March 2003 to December 2016.

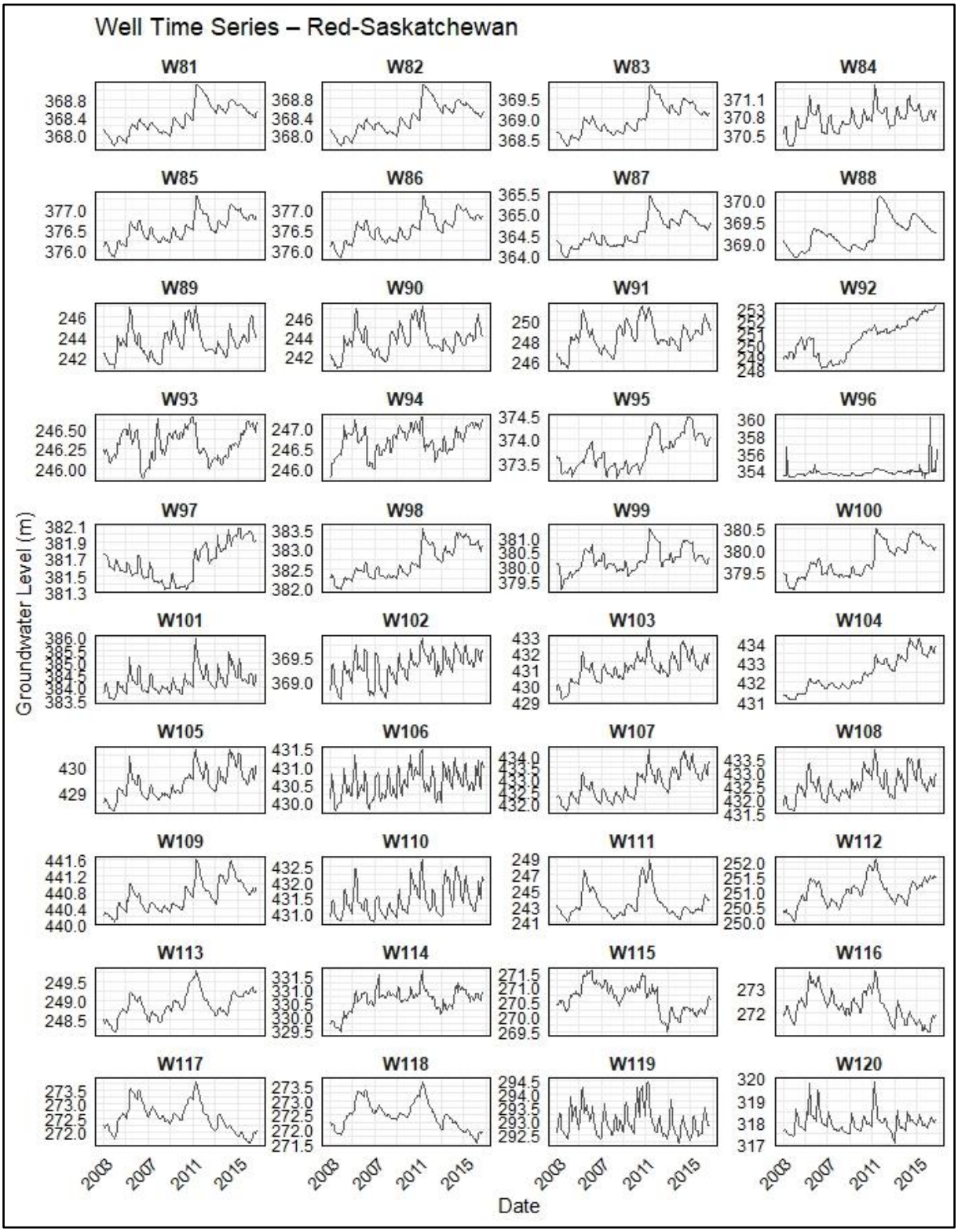


Figure A.7 Monthly groundwater level time series (in meters) for individual wells (W81–W120; see Table A.2) across the Red-Saskatchewan basin over the study period from March 2003 to December 2016.

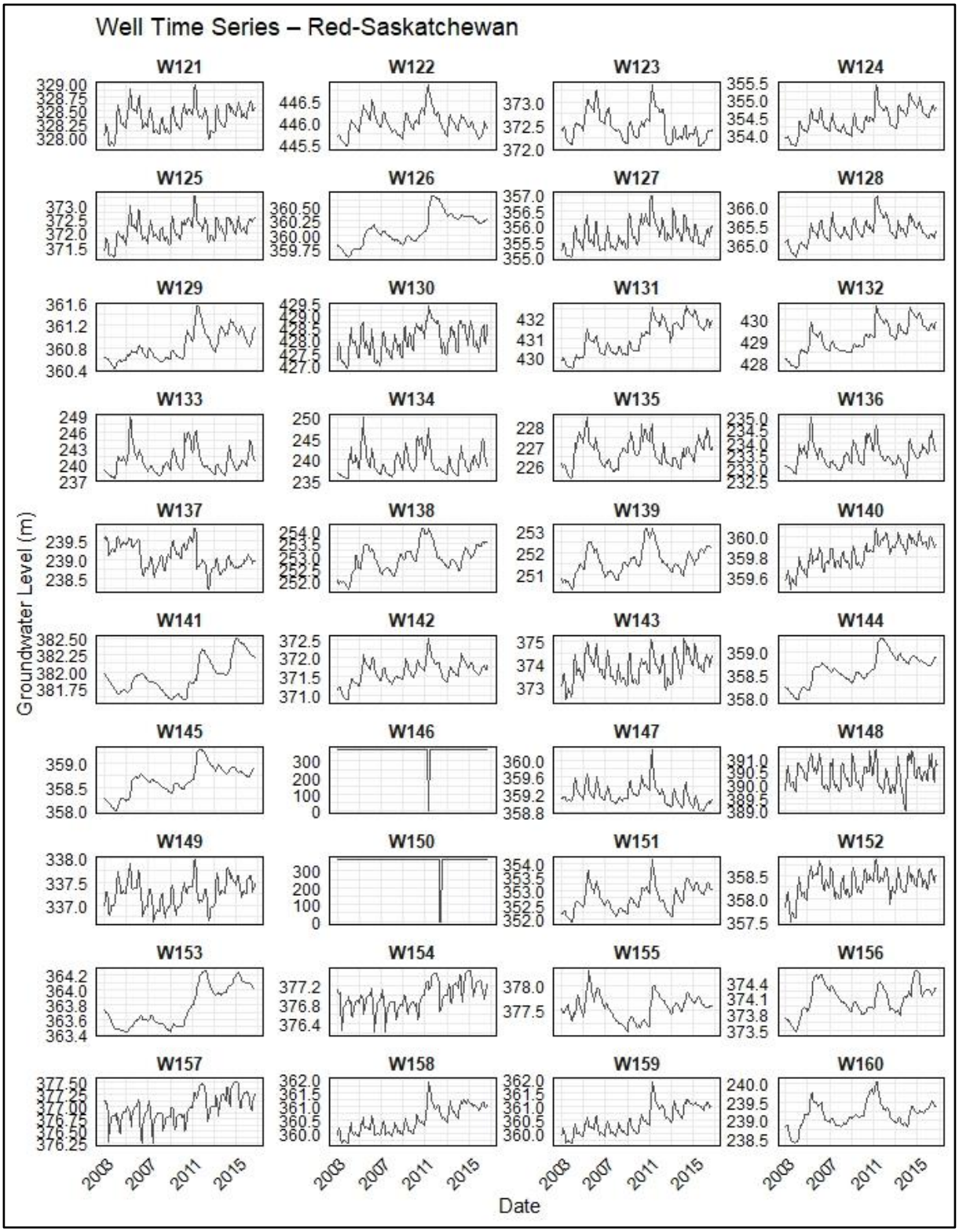


Figure A.8 Monthly groundwater level time series (in meters) for individual wells (W121–W160; see Table A.2) across the Red-Saskatchewan basin over the study period from March 2003 to December 2016.

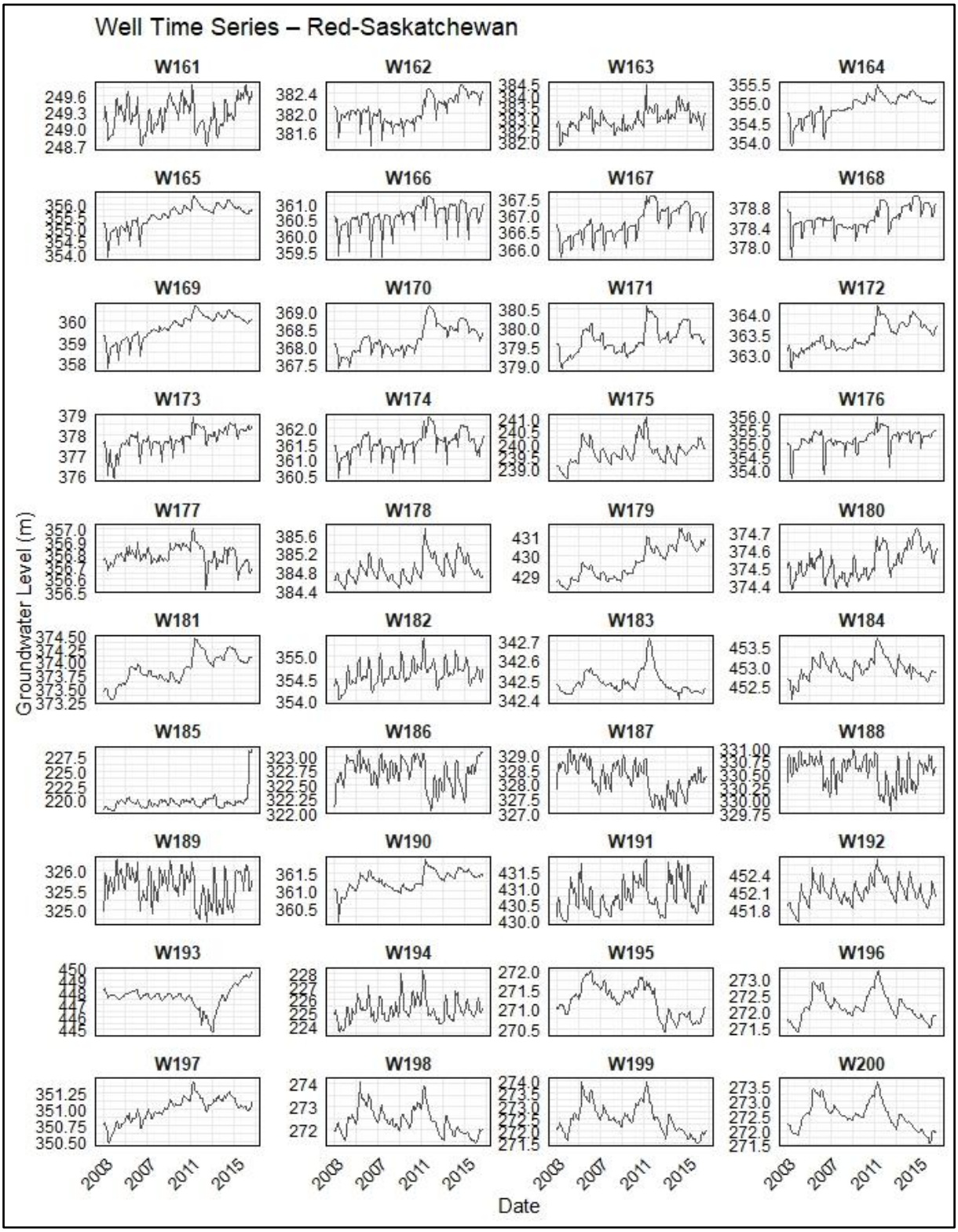


Figure A.9 Monthly groundwater level time series (in meters) for individual wells (W161–W200; see Table A.2) across the Red-Saskatchewan basin over the study period from March 2003 to December 2016.

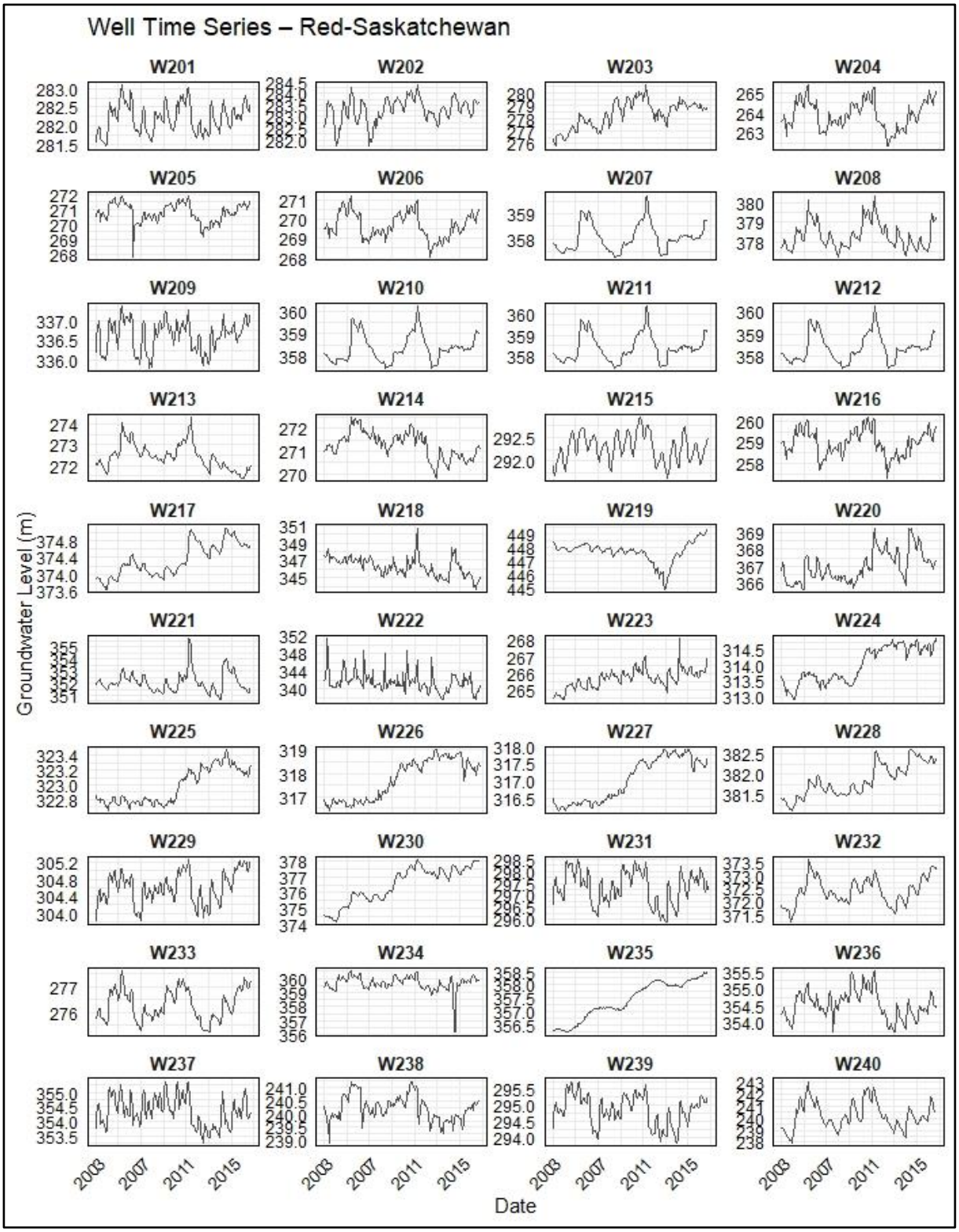


Figure A.10 Monthly groundwater level time series (in meters) for individual wells (W201–W240; see Table A.2) across the Red-Saskatchewan basin over the study period from March 2003 to December 2016.

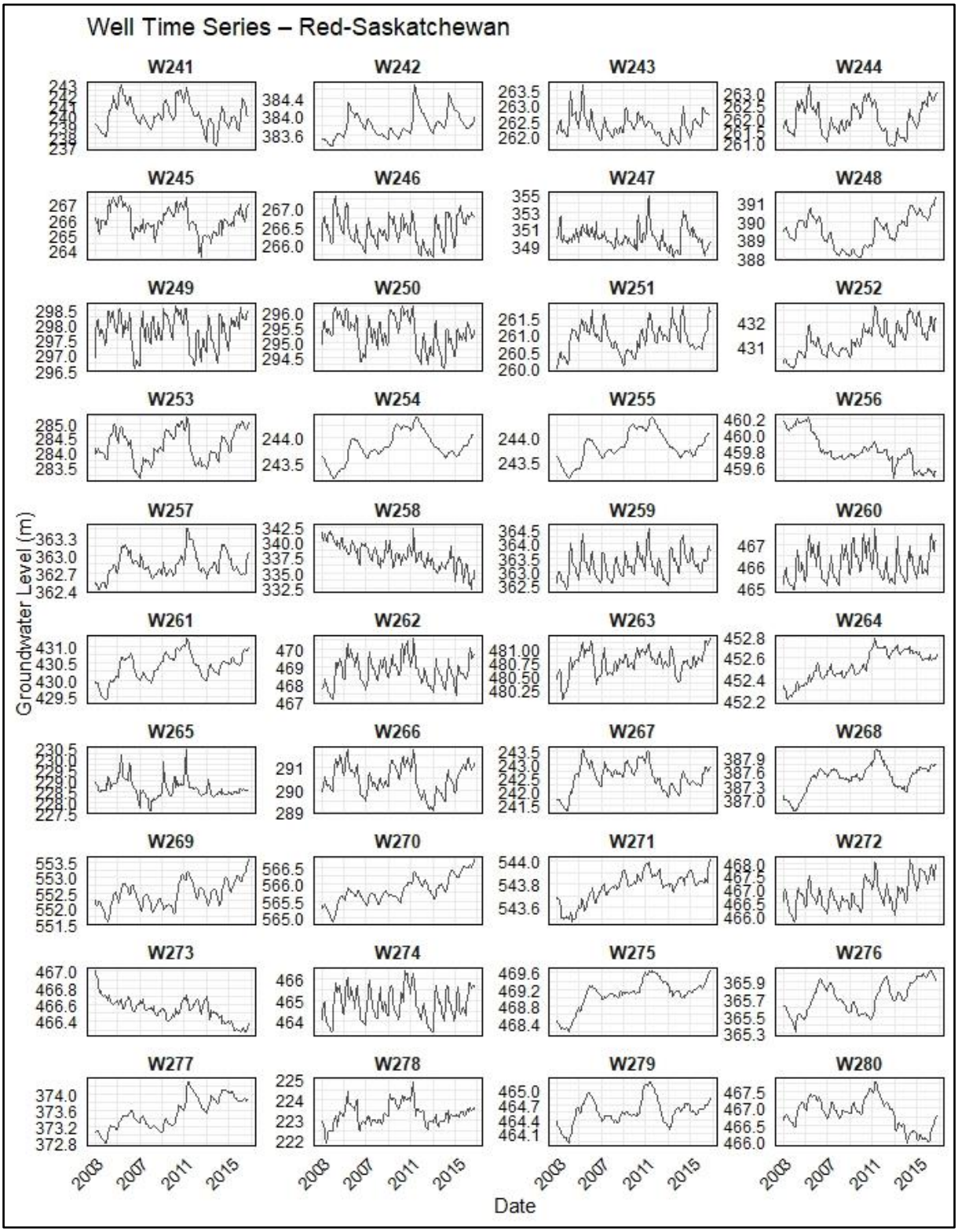


Figure A.11 Monthly groundwater level time series (in meters) for individual wells (W241–W280; see Table A.2) across the Red-Saskatchewan basin over the study period from March 2003 to December 2016.

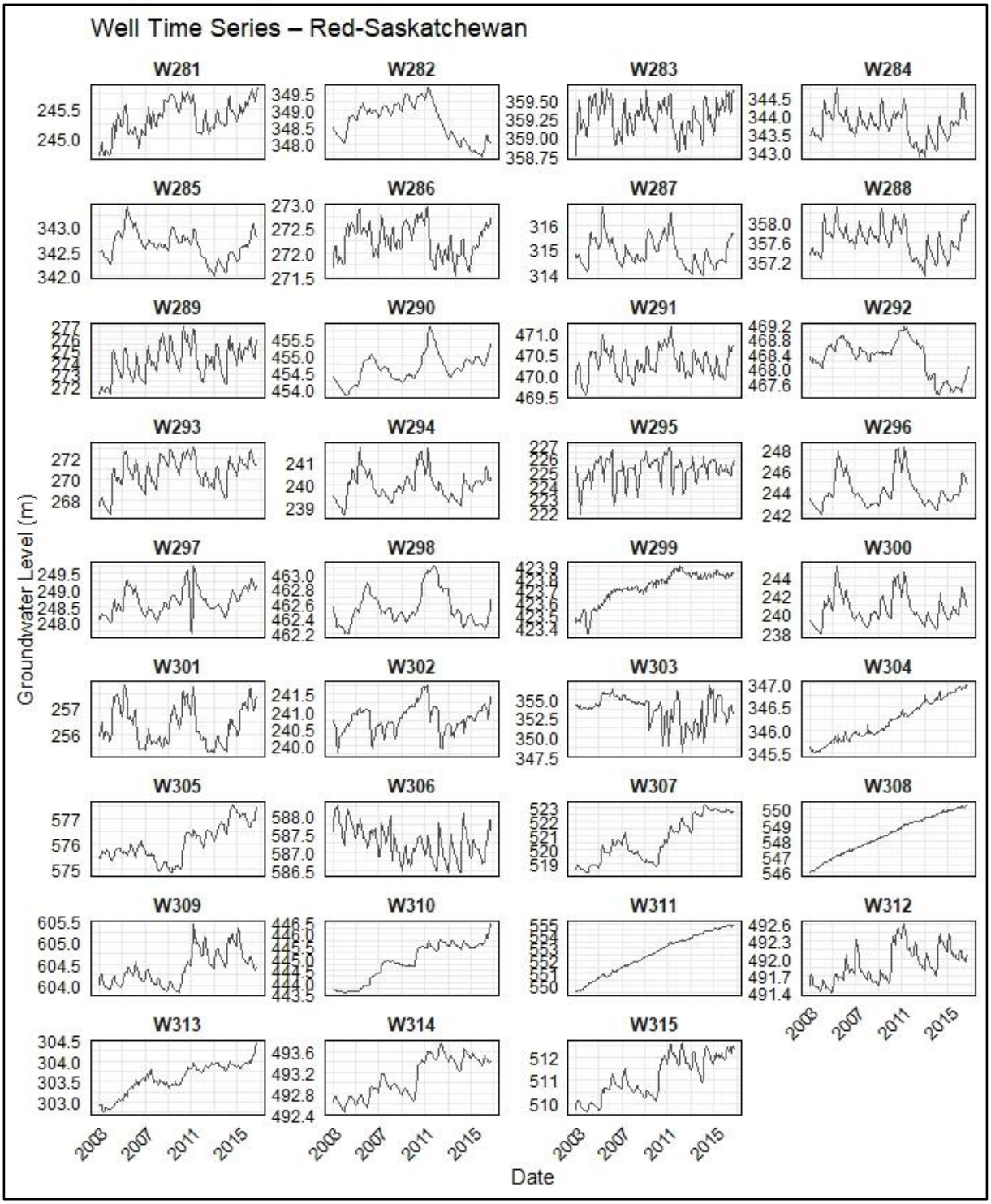


Figure A.12 Monthly groundwater level time series (in meters) for individual wells (W281–W315; see Table A.2) across the Red-Saskatchewan basin over the study period from March 2003 to December 2016.

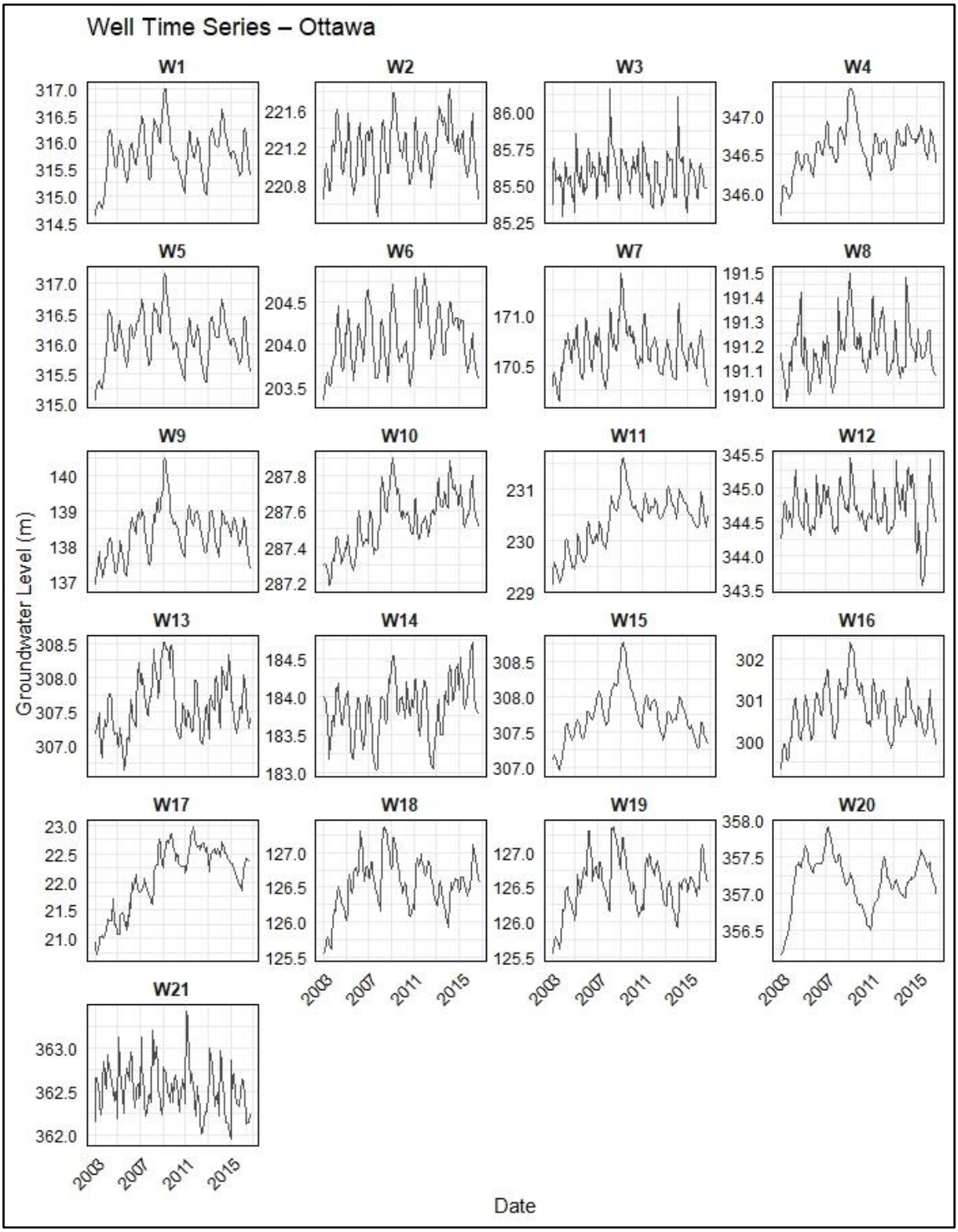


Figure A.13 Monthly groundwater level time series (in meters) for individual wells (W1–W21; see Table A.2) across the Ottawa basin over the study period from March 2003 to December 2016.

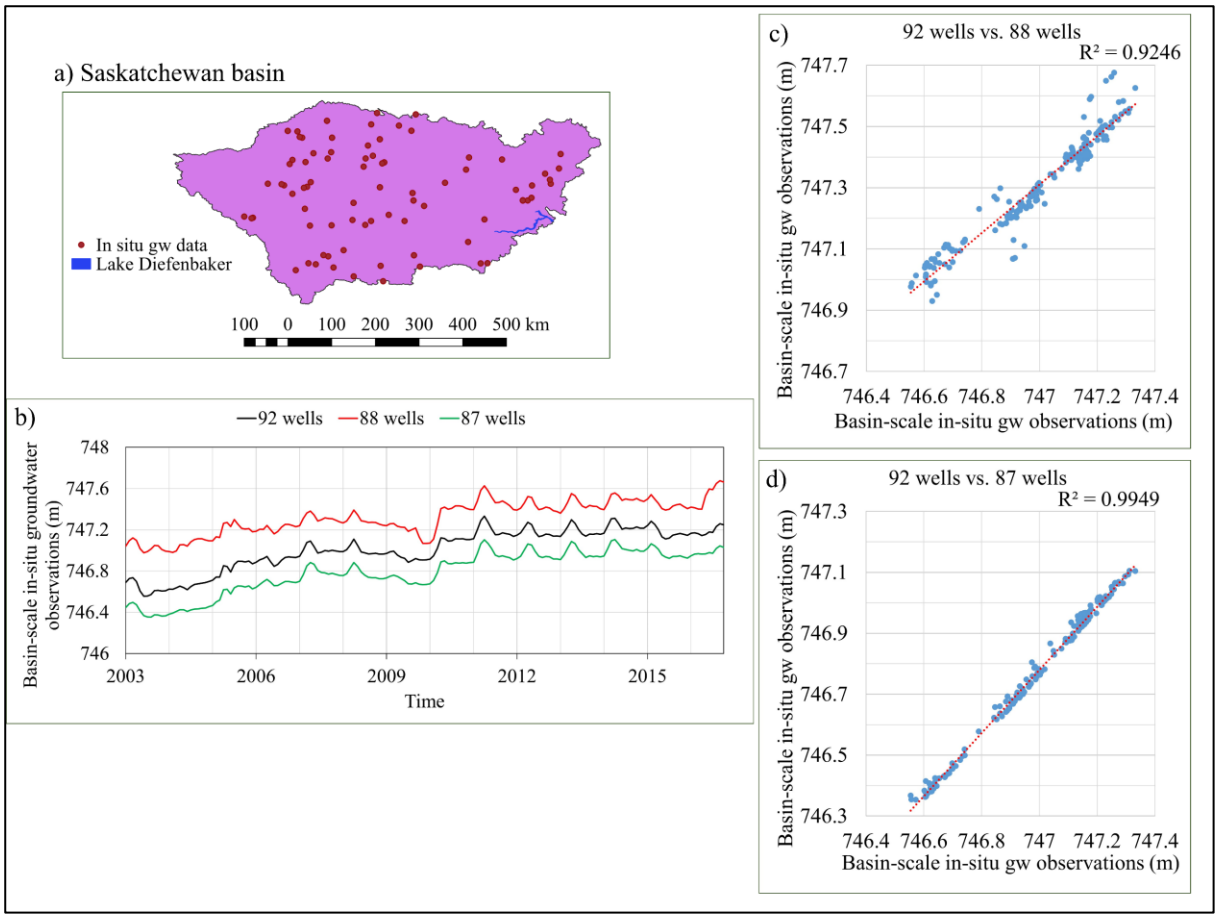


Figure A.14 Bootstrapping interpolation of in-situ groundwater observations, applying Thiessen Polygons (TP), to generate time series of basin-scale in-situ groundwater observations across Saskatchewan basin. The figure comprises four panels: (a) the spatial distribution of in-situ groundwater observations across the basin, and (b) the time series of the basin-scale in-situ groundwater observations using three different well counts: 92 wells (black), 88 wells (red), and 87 wells (green). Panel (c) shows the Spearman's rank correlation coefficient (ρ) and coefficient of determination (R^2) between the basin-scale in-situ groundwater observation time series using 92 wells and 88 wells, which are 0.97 and 0.92, respectively. Panel (d) displays the Spearman's rank correlation coefficient (ρ) and coefficient of determination (R^2) between the basin-scale in-situ groundwater observation time series using 92 wells and 87 wells, which are 0.99, both.

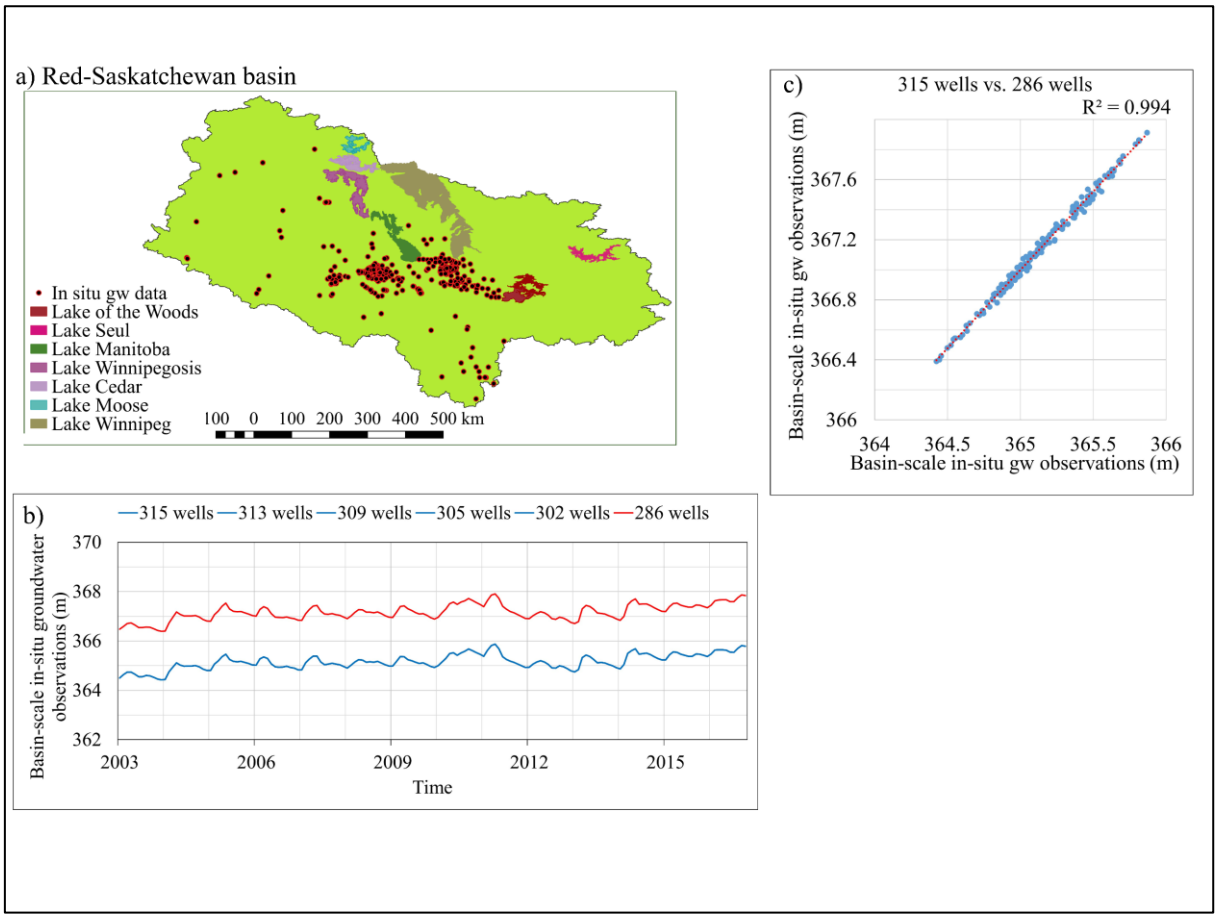


Figure A.15 Bootstrapping interpolation of in-situ groundwater observations, applying Thiessen Polygons (TP), to generate time series of basin-scale in-situ groundwater observations across Red-Saskatchewan basin. The figure comprises three panels: (a) the spatial distribution of in-situ groundwater observations across the basin, and (b) the time series of the basin-scale in-situ groundwater observation using six different well counts: 315 wells (blue), 313 wells (blue), 309 wells (blue), 305 wells (blue), 302 wells (blue), and 286 wells (red). Panel (c) shows the Spearman's rank correlation coefficient (ρ) and coefficient of determination (R^2) between the basin-scale in-situ groundwater observation time series using 315 wells and 286 wells, which are 0.99, both. The produced time series of basin-scale in-situ groundwater observations were identical for the first five well counts (315, 313, 309, 305, 302).

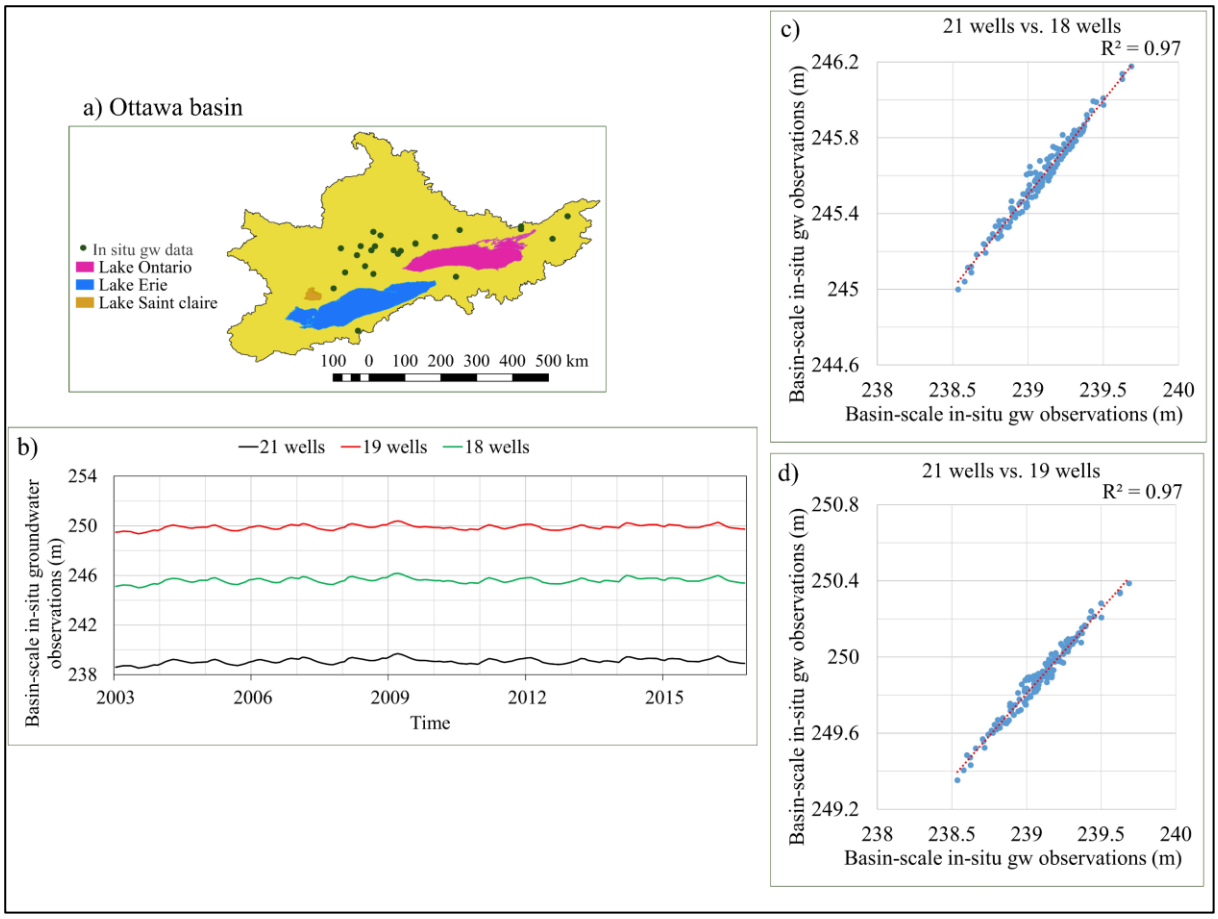


Figure A.16 Bootstrapping interpolations of in-situ groundwater observations, applying Thiessen Polygons (TP), to generate time series of basin-scale in-situ groundwater observations across Ottawa basin. The figure comprises four panels: (a) the spatial distribution of in-situ groundwater observations across the basin, and (b) the time series of the basin-scale in-situ groundwater observation using three different well counts: 21 wells (black), 19 wells (red), and 18 wells (green). Panel (c) shows the Spearman's rank correlation coefficient (ρ) and coefficient of determination (R^2) between the basin-scale in-situ groundwater observation time series using 21 wells and 19 wells, which are 0.98 and 0.97, respectively. Panel (d) displays the Spearman's rank correlation coefficient (ρ) and coefficient of determination (R^2) between the basin-scale in-situ groundwater observations time series using 21 wells and 18 wells, which are 0.98 and 0.97, respectively.

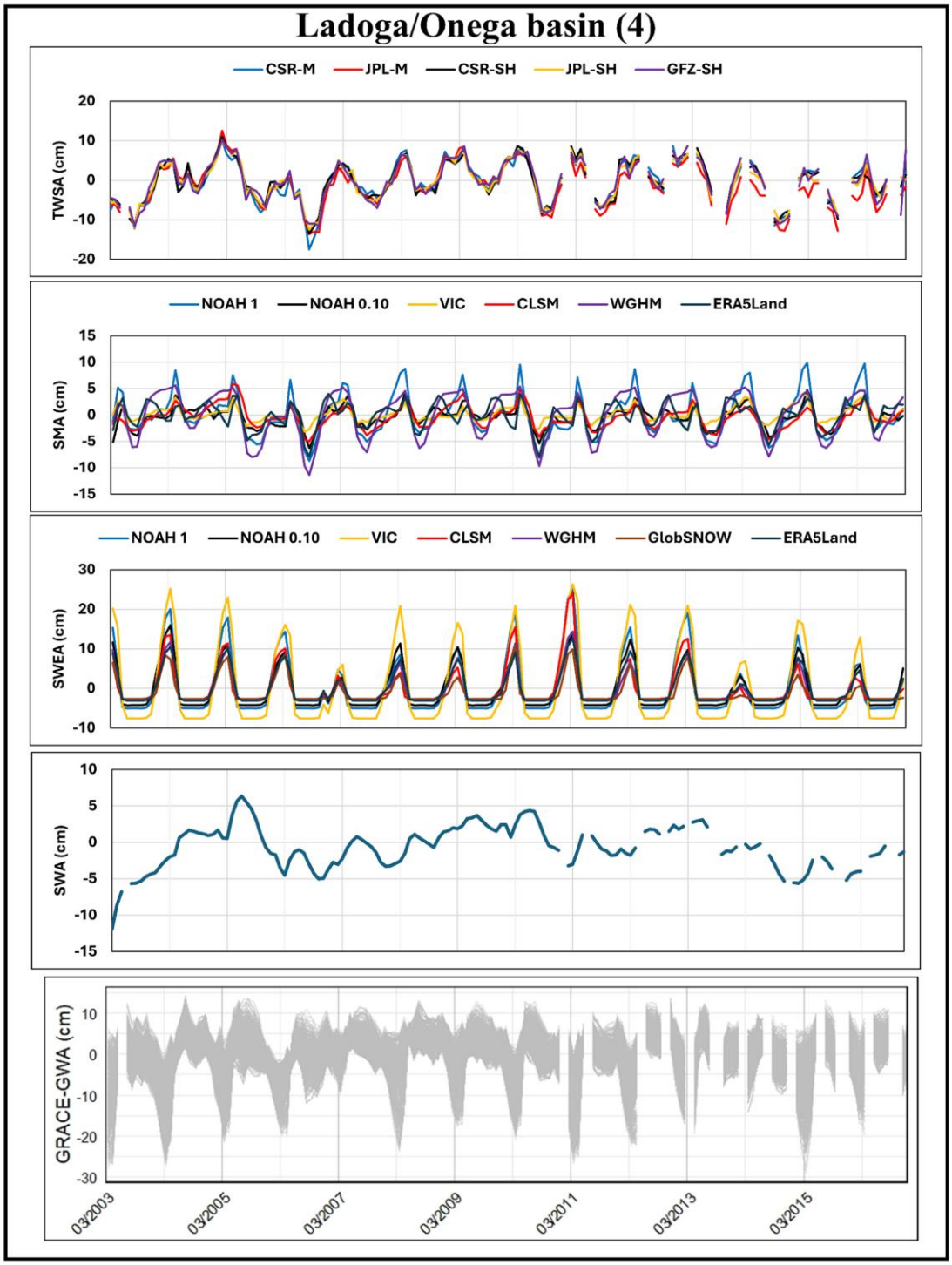


Figure A.17 Monthly time series from 2003 to 2016 over the Ladoga/Onega Basin (Basin 4 in Figure 2.1), illustrating: (i) GRACE-derived Terrestrial Water Storage Anomalies (TWSA) from CSR-M, JPL-M, CSR-SH, JPL-SH, and GFZ-SH solutions; (ii) Soil Moisture Anomalies (SMA) from NOAH (1˚ and 0.1˚), VIC (1˚), CLSM (1˚), WGHM (0.5˚), and ERA5-Land (0.1˚); (iii) Snow Water Equivalent Anomalies (SWEA) from NOAH (1˚ and 0.1˚), VIC (1˚), CLSM (1˚), HGSWE (0.25˚), WGHM (0.5˚), GlobSnow (0.25˚), and ERA5-Land (0.1˚); (iv) Surface Water Anomalies (SWA); and (v) 630 GRACE-based Groundwater Storage Anomaly (GRACE-GWA) realisations derived from multi-model combinations of water budget components.

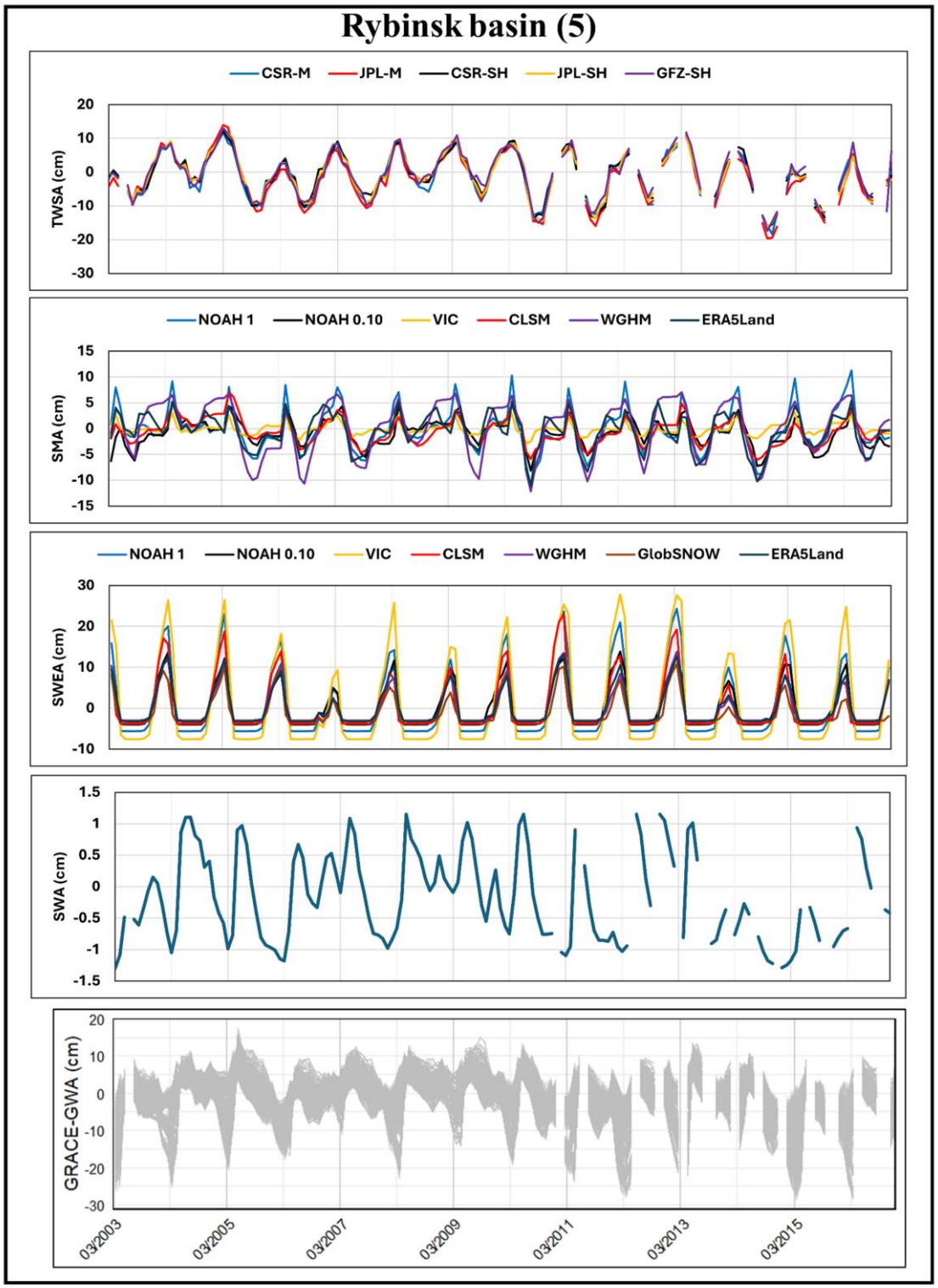


Figure A.18 Same as figure **A.17** for Rybinsk basin (Basin 5 in Figure 2.1).

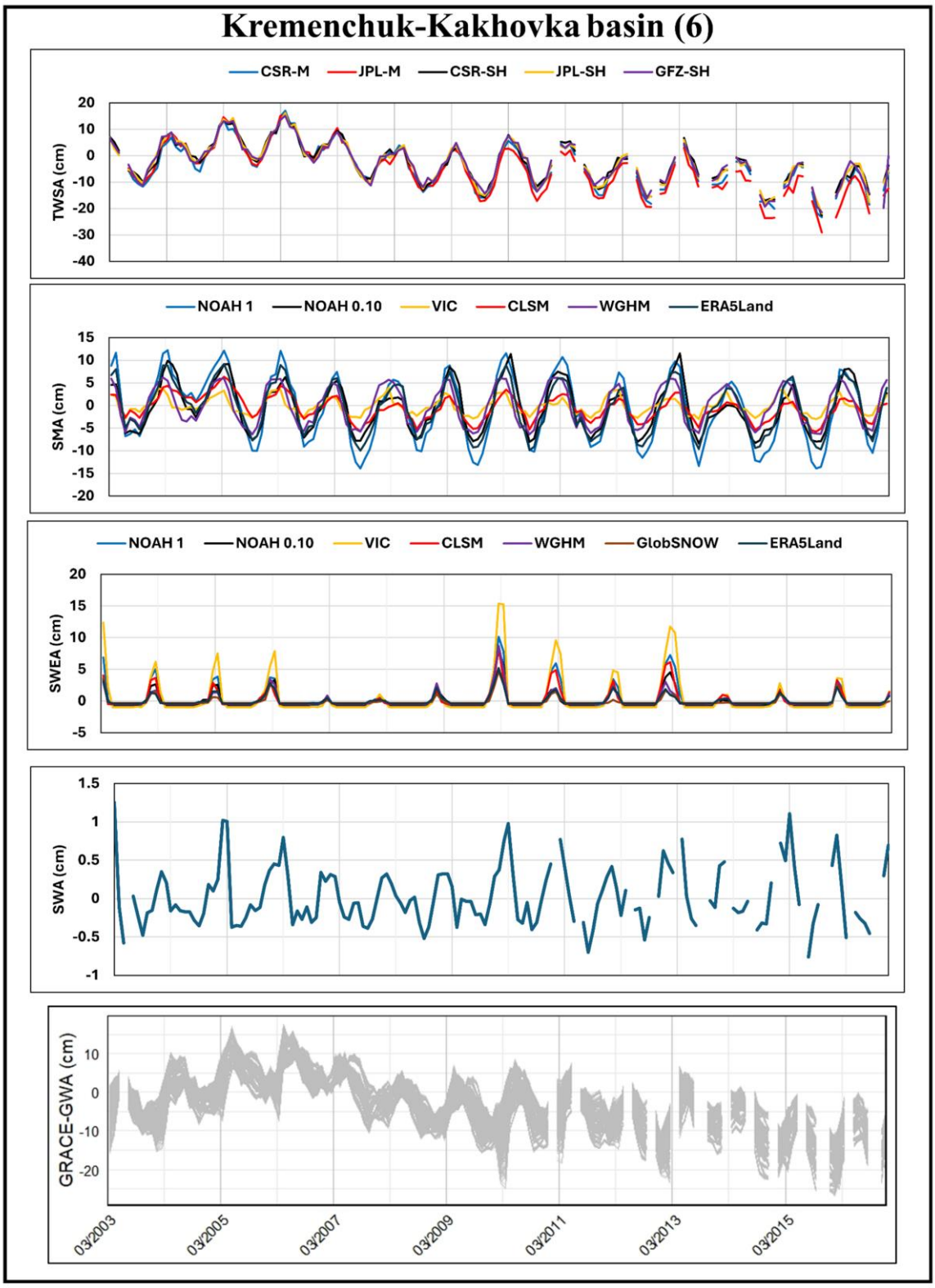


Figure A.19 Same as figure **A.17** for Kremenchuk-Kakhovka basin (Basin 6 in Figure 2.1).

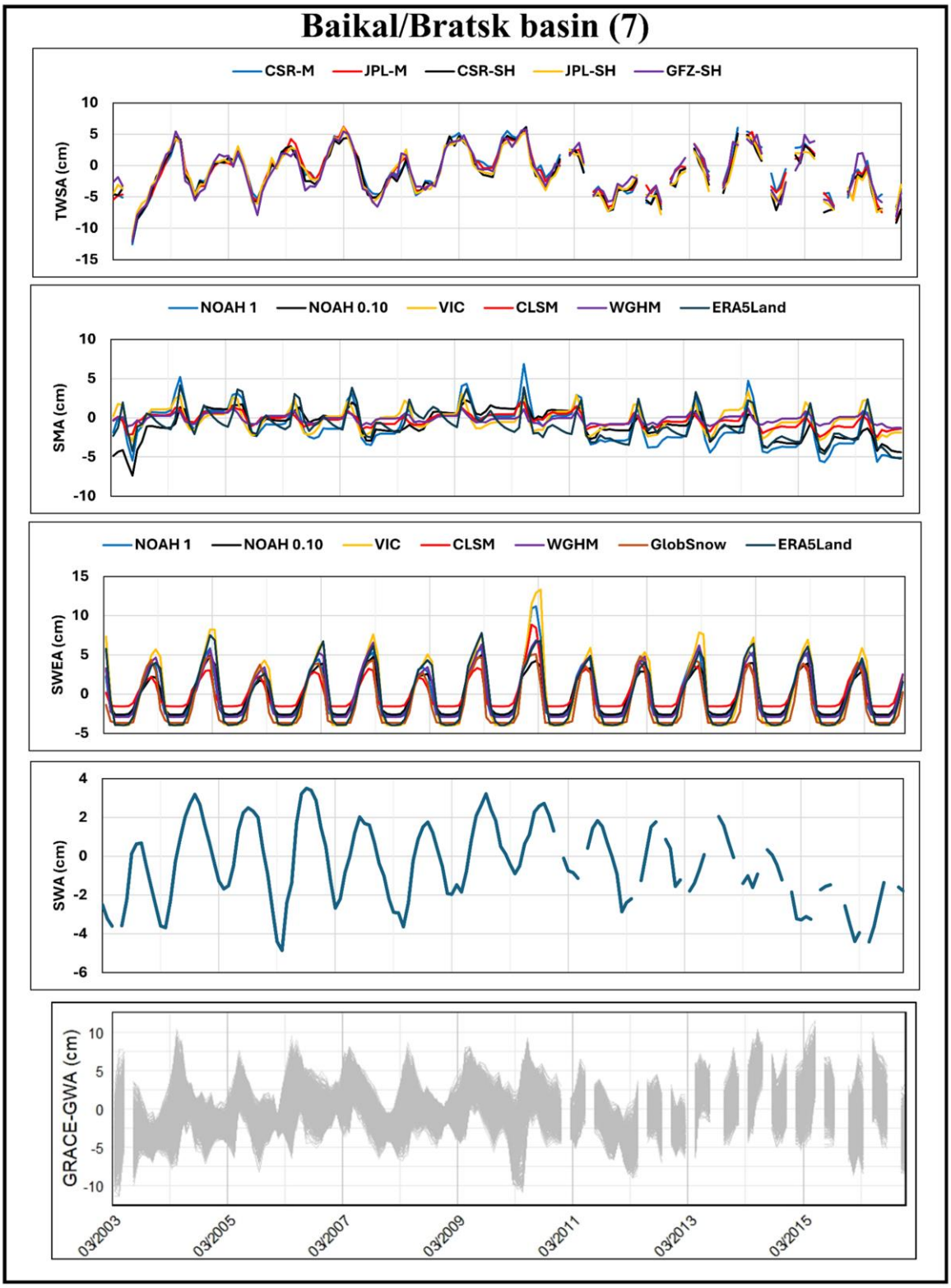


Figure A.20 Same as figure **A.17** for Bikal-Bratsk basin (Basin 7 in Figure 2.1).

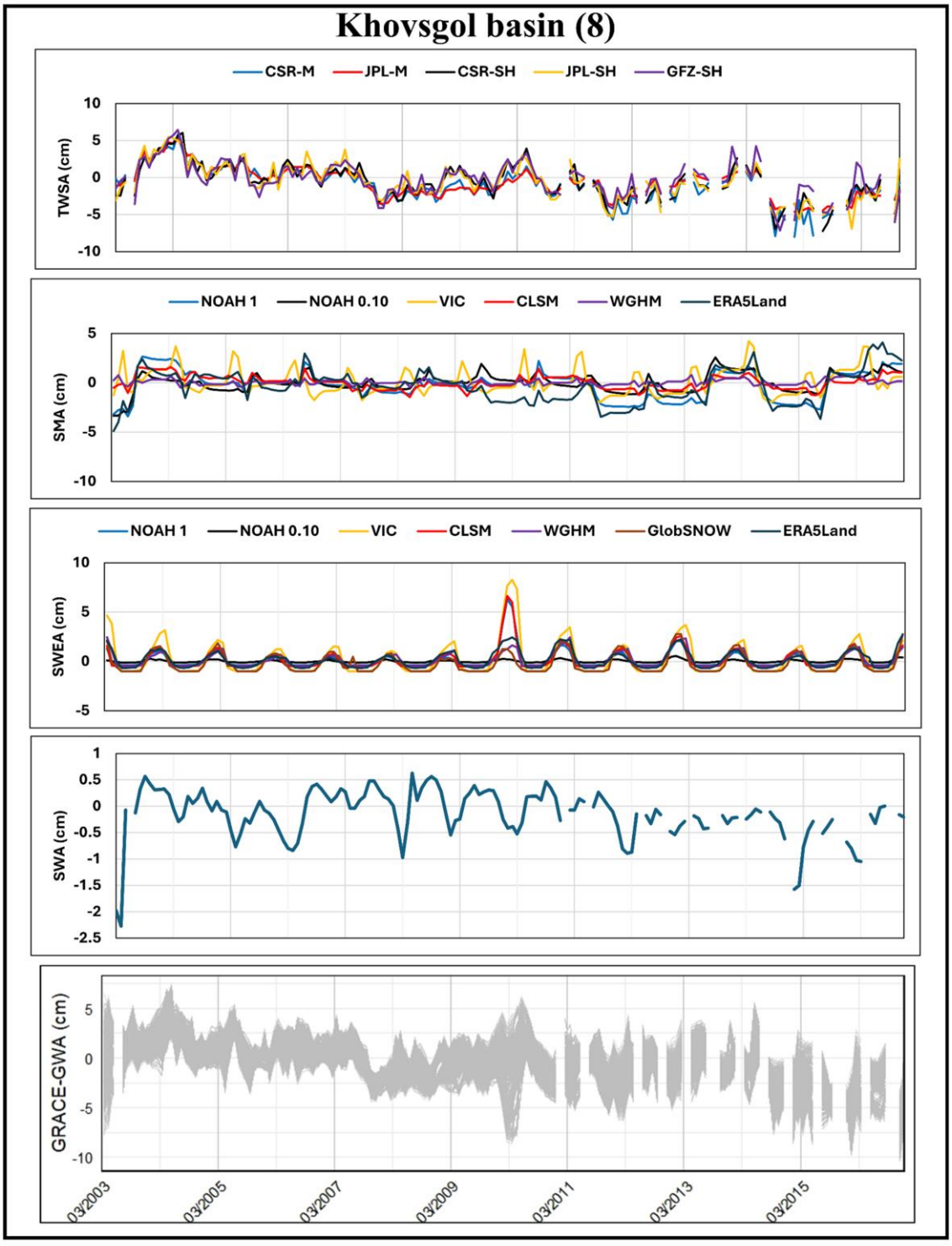


Figure A.21 Same as figure **A.17** for Khovsgol basin (Basin 8 in Figure 2.1).

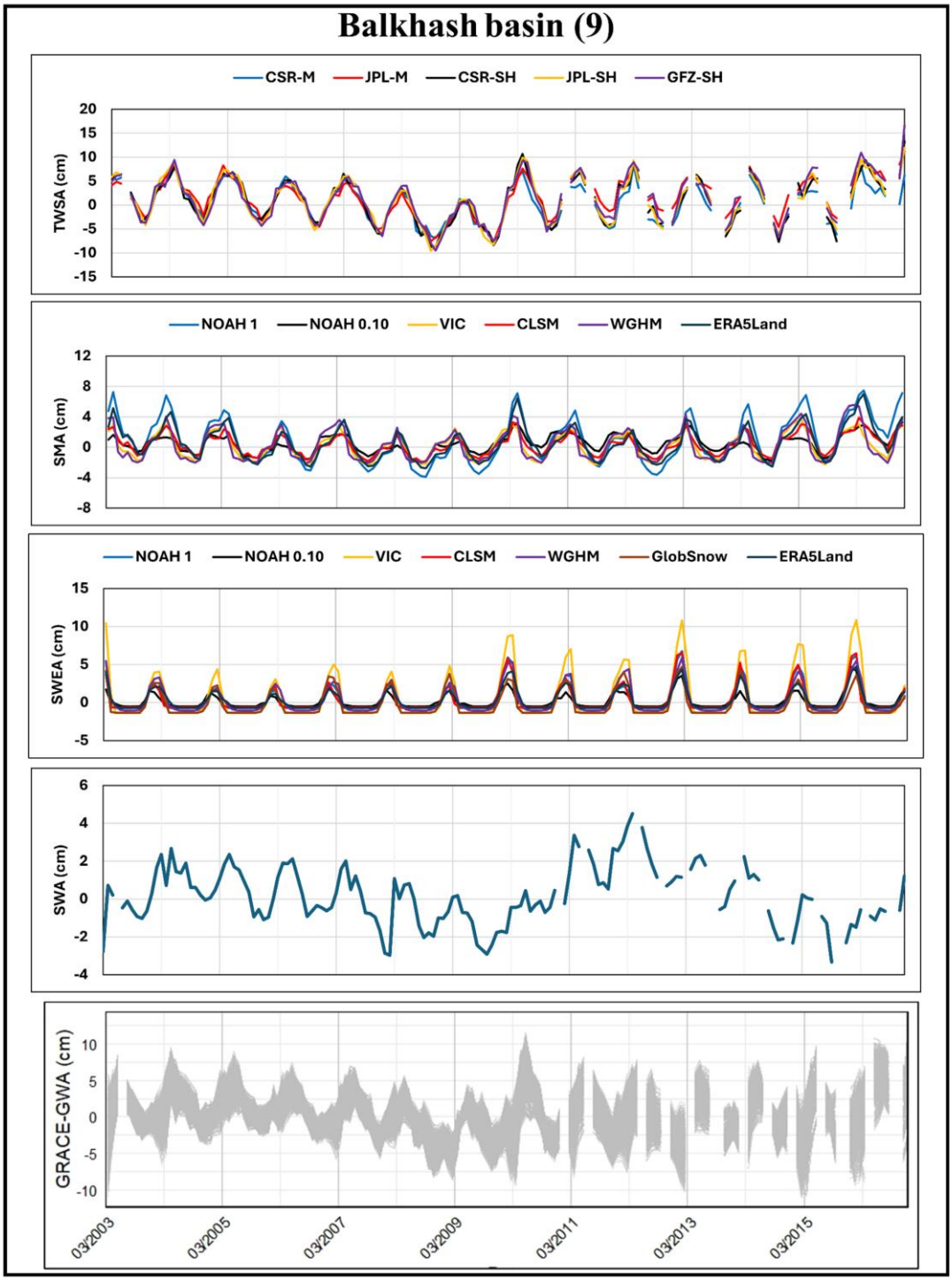


Figure A.22 Same as figure **A.17** for Balkhash basin (Basin 9 in Figure 2.1).

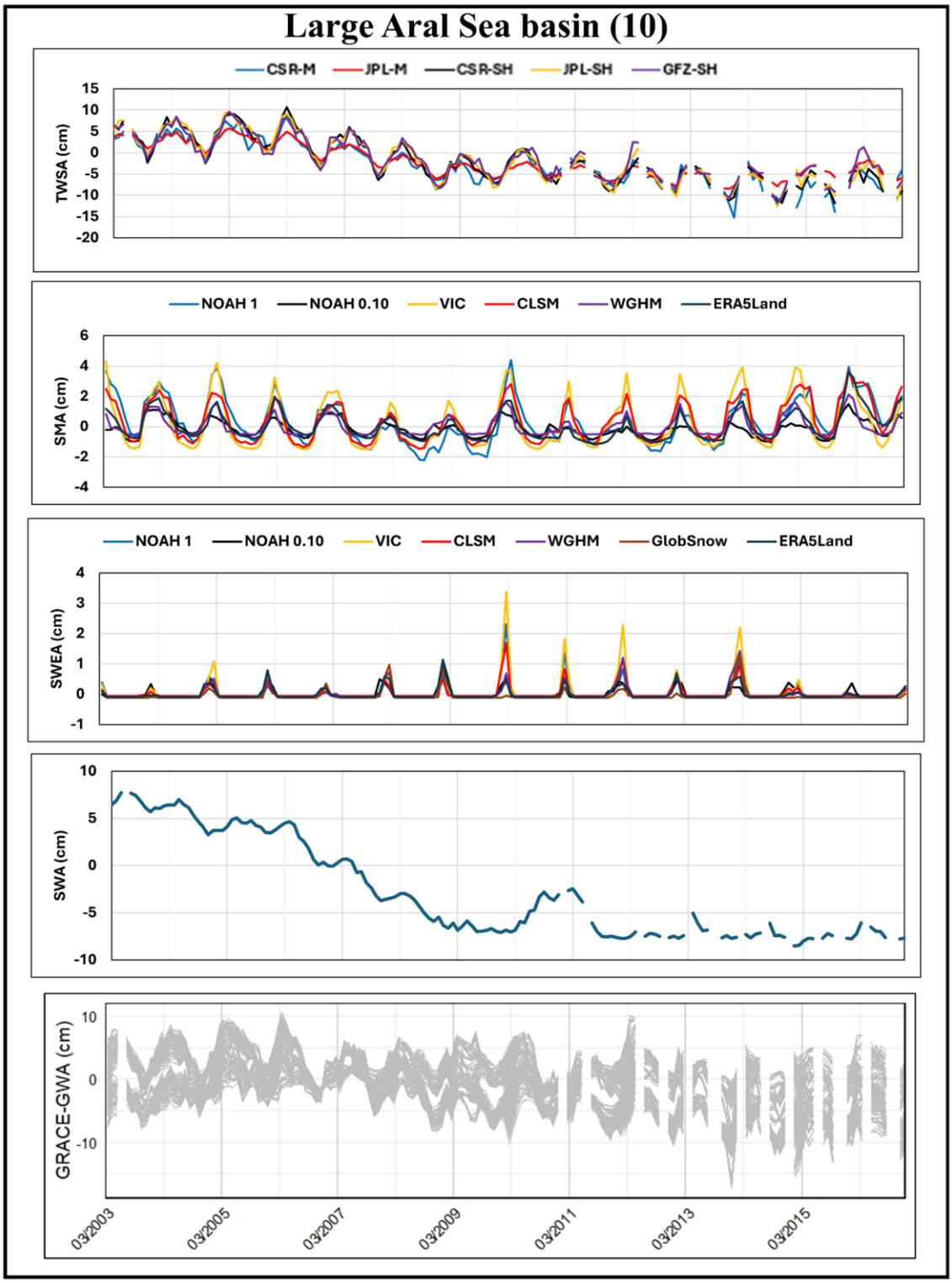


Figure A.23 Same as figure **A.17** for Large Aral Sea basin (Basin 10 in Figure 2.1).

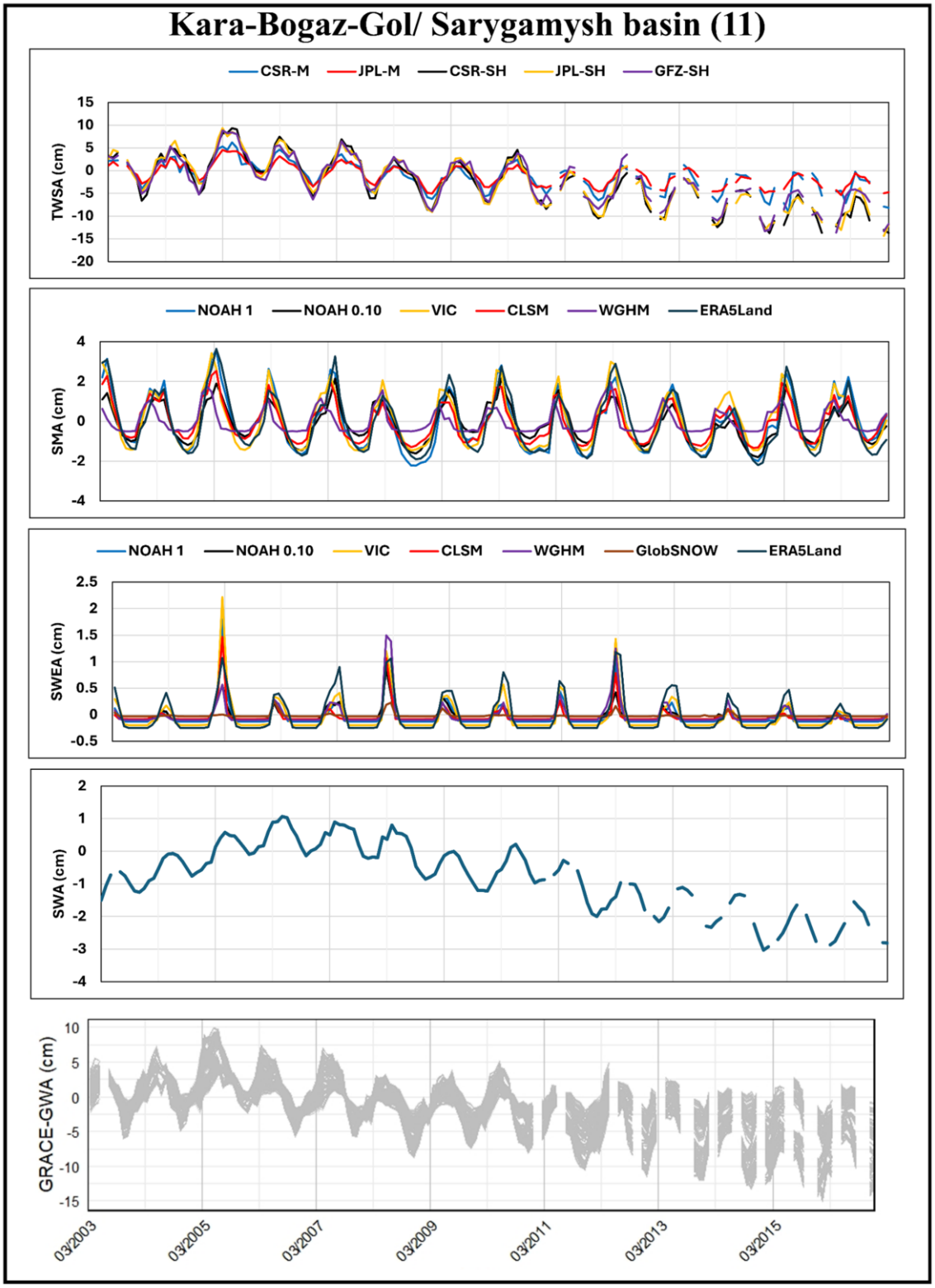


Figure A.24 Same as figure **A.17** for Kara-Bogaz-Gol/Sarygamysh basin (Basin 11 in Figure 2.1).

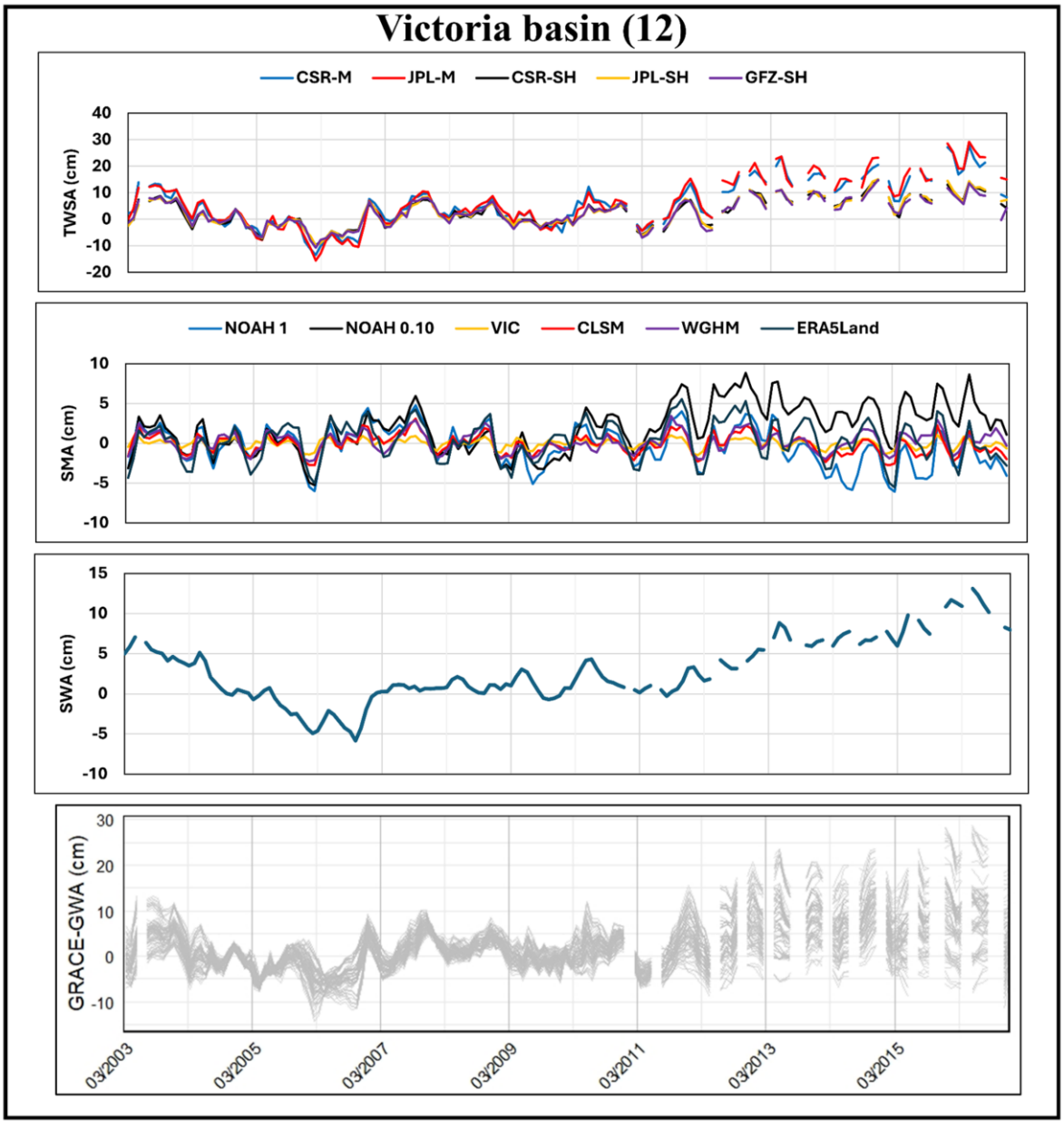


Figure A.25 Monthly time series from 2003 to 2016 over the Victoria basin (Basin 12 in Figure 2.1), illustrating: (i) GRACE-derived Terrestrial Water Storage Anomalies (TWSA) from CSR-M, JPL-M, CSR-SH, JPL-SH, and GFZ-SH solutions; (ii) Soil Moisture Anomalies (SMA) from NOAH (1˚ and 0.1˚), VIC (1˚), CLSM (1˚), WGHM (0.5˚), and ERA5-Land (0.1˚); (iii) Surface Water Anomalies (SWA); and (iv) 90 GRACE-based Groundwater Storage Anomaly (GRACE-GWA) realisations derived from multi-model combinations of water budget components.

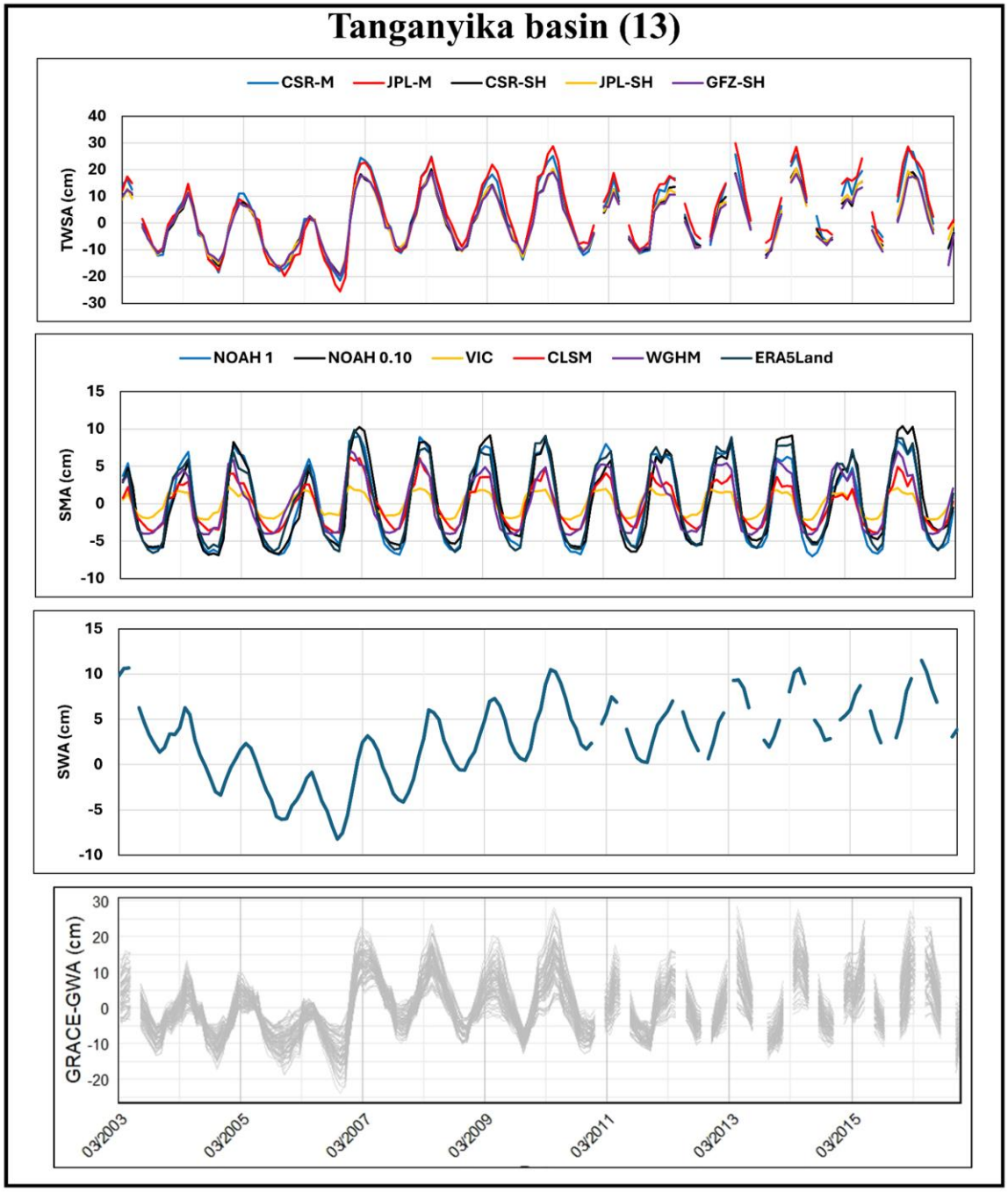


Figure A.26 Same as figure **A.25** for Tanganyika basin (Basin 13 in Figure 2.1).

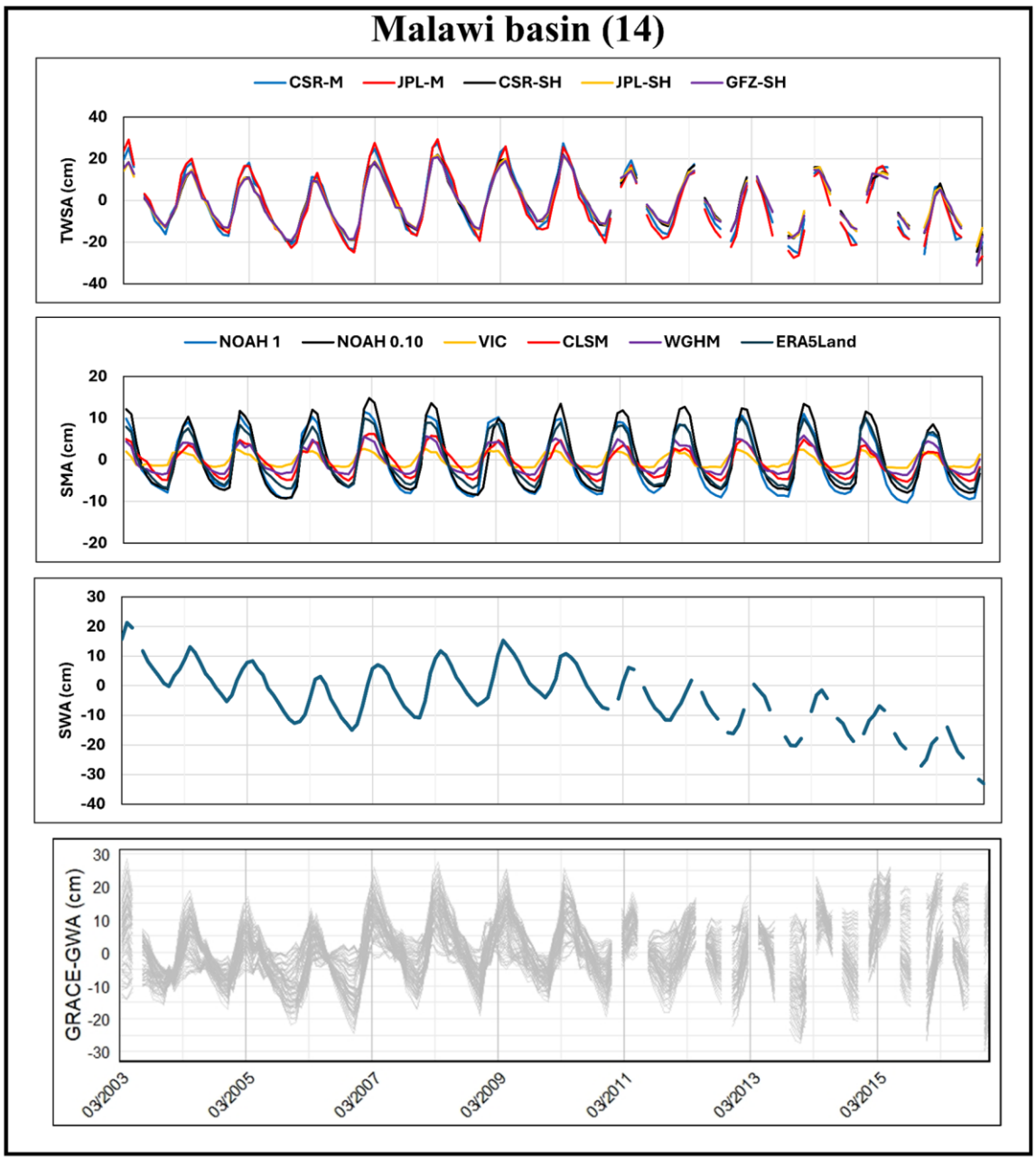


Figure A.27 Same as figure **A.25** for Malawi basin (Basin 14 in Figure 2.1).

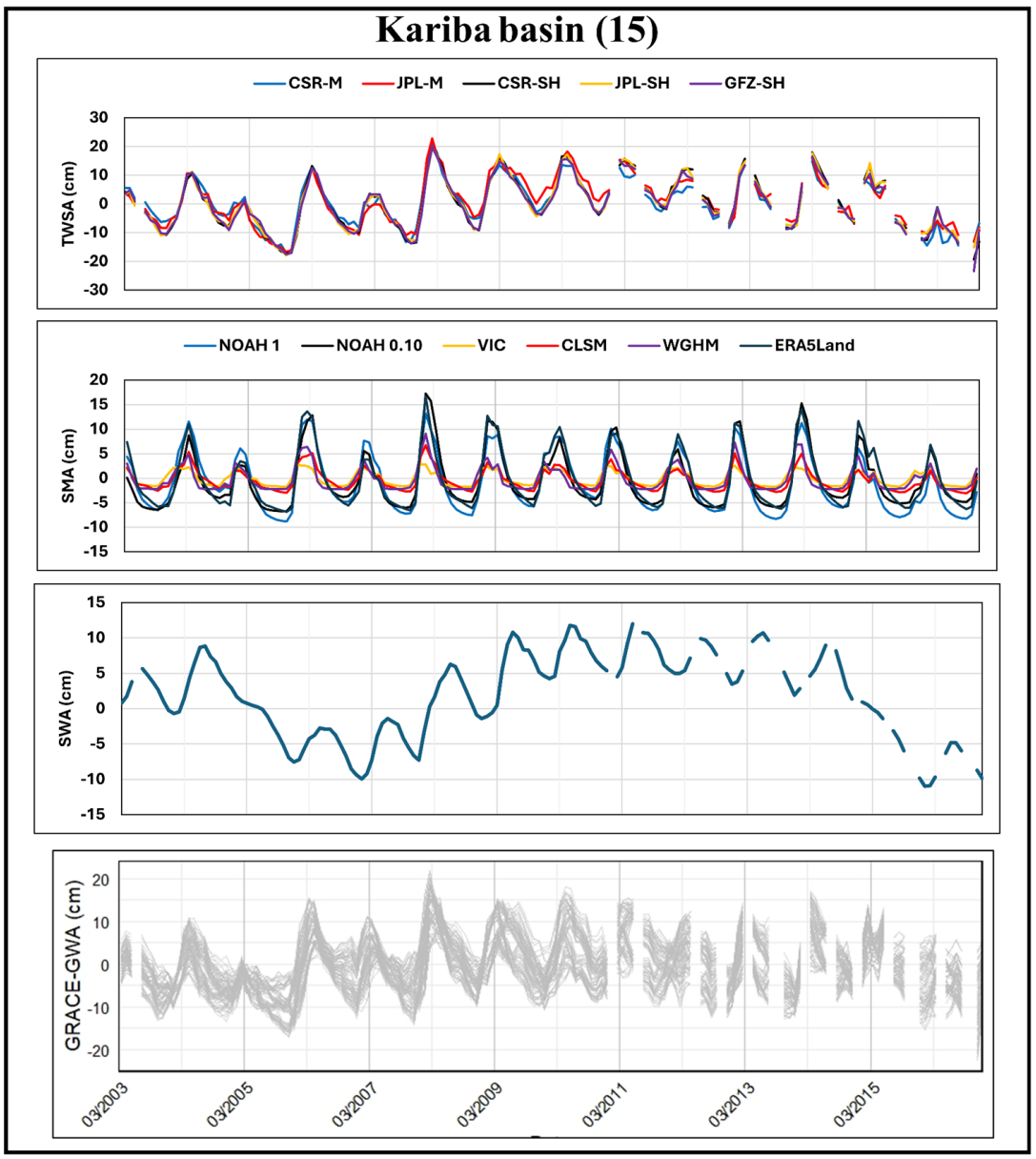


Figure A.28 Same as figure **A.25** for Kariba basin (Basin 15 in Figure 2.1).

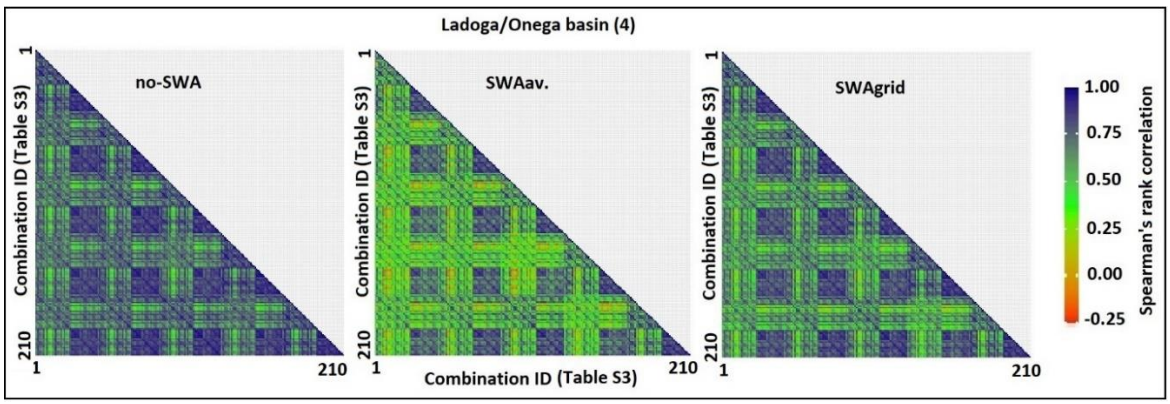


Figure A.29 Spearman's rank correlation coefficients (ρ) between 210 time series estimates of GRACE-GWA (Table S3; no-SWA, $SWA_{av.}$, SWA_{grid}) for Ladoga/Onega basin (Basin 4 in Figure 2.1).

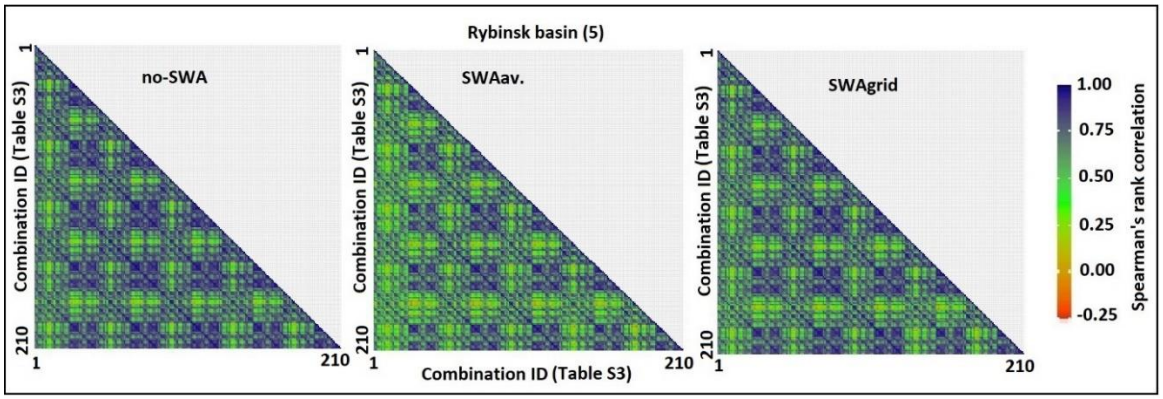


Figure A.30 Same as figure **A.29** for Rybinsk basin (Basin 5 in Figure 2.1).

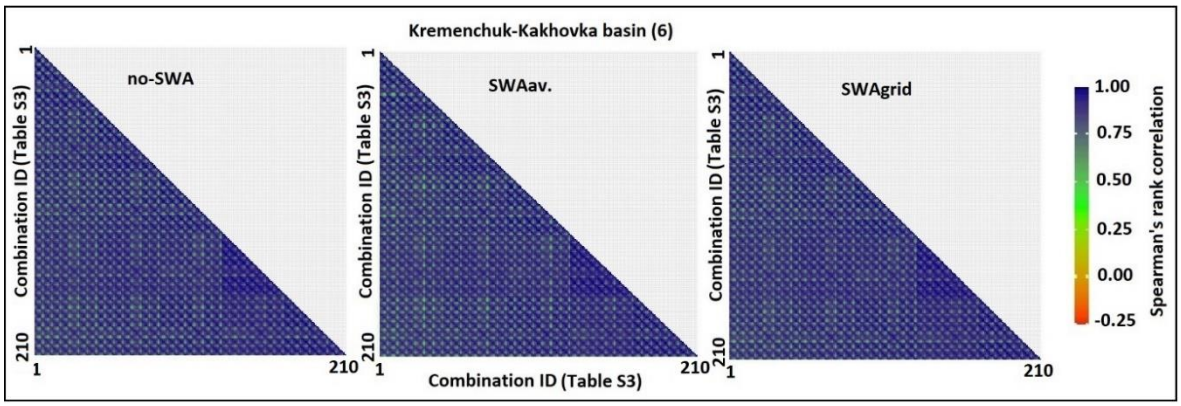


Figure A.31 Same as figure **A.29** for Kremenchuk-Kakhovka basin (Basin 6 in Figure 2.1).

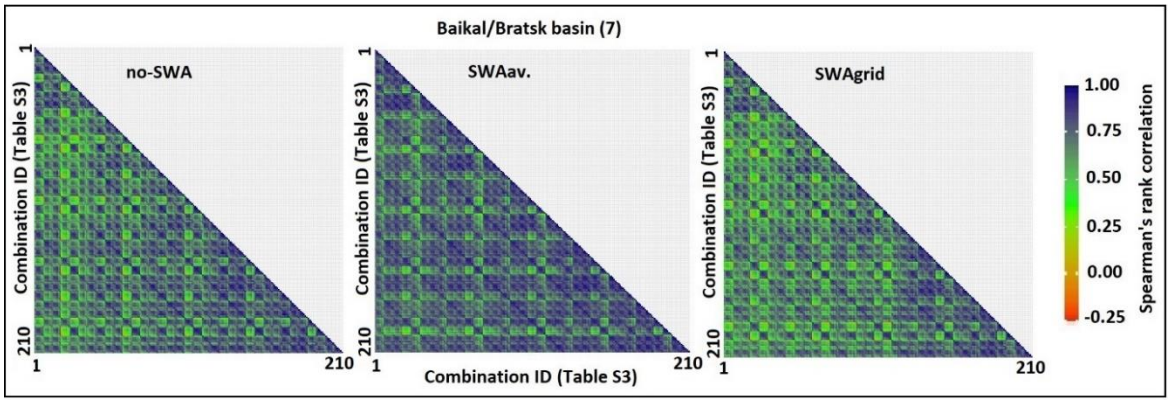


Figure A.32 Same as figure **A.29** for Baikal/Bratsk basin (Basin 7 in Figure 2.1).

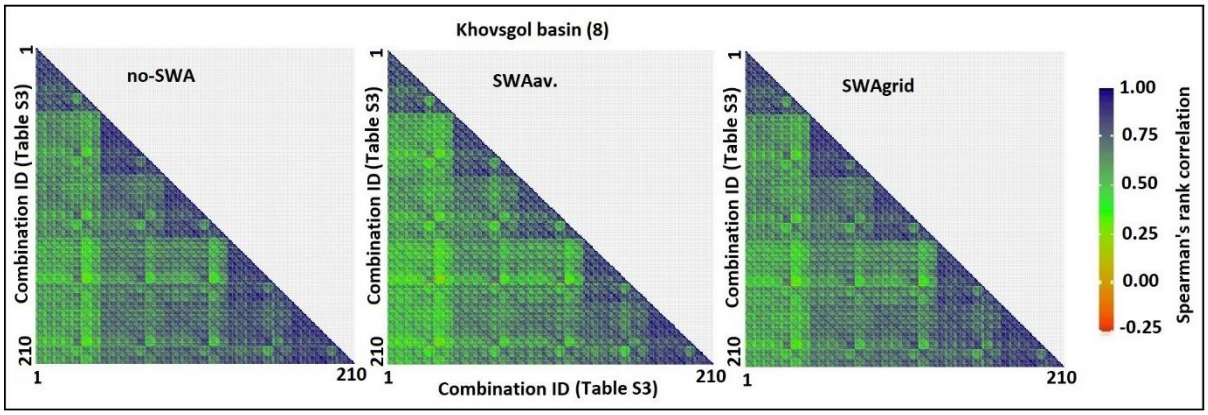


Figure A.33 Same as figure **A.29** for Khovsgol basin (Basin 8 in Figure 2.1).

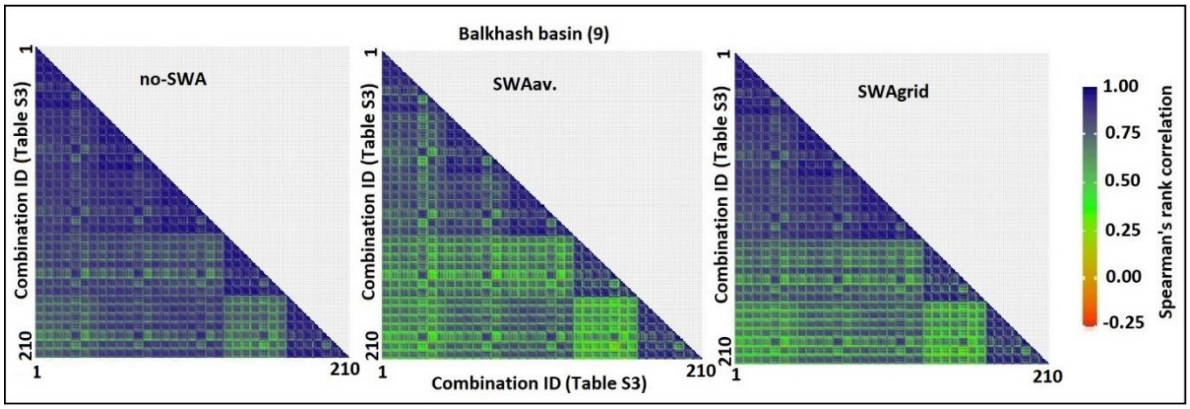


Figure A.34 Same as figure **A.29** for Balkhash basin (Basin 9 in Figure 2.1).

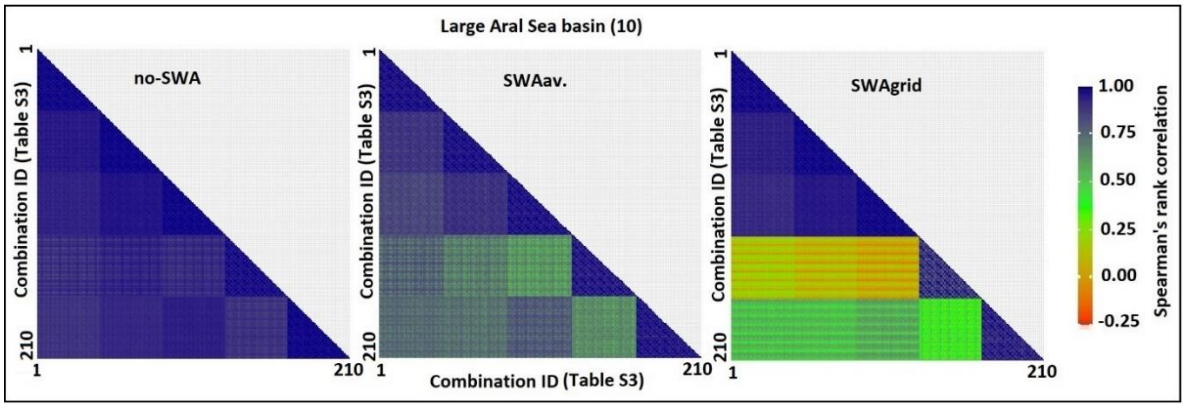


Figure A.35 Same as figure **A.29** for Large Aral Sea basin (Basin 10 in Figure 2.1).

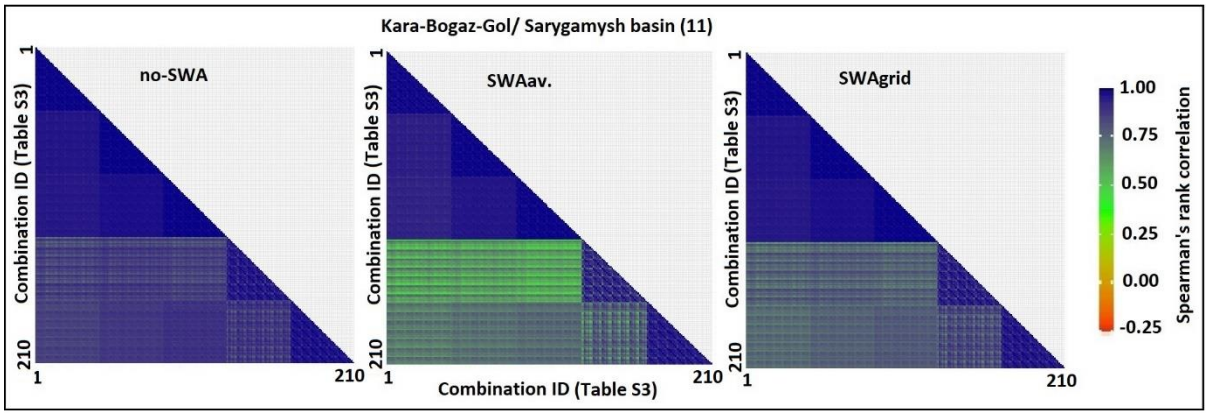


Figure A.36 Same as figure **A.29** for Kara-Bogaz-Gol/ Sarygamysh basin (Basin 11 in Figure 2.1).

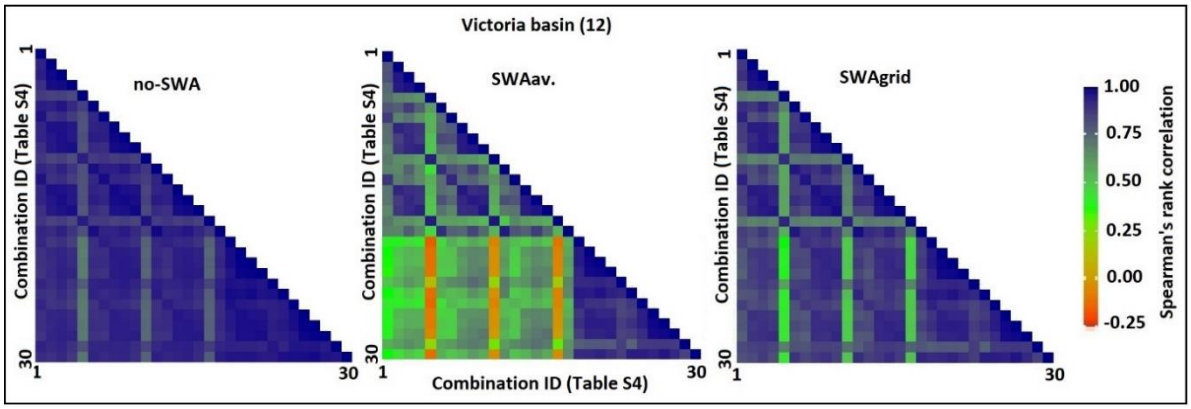


Figure A.37 Spearman's rank correlation coefficients (ρ) between 30 time series estimates of GRACE-GWA (Table S4; no-SWA, $SWA_{av.}$, SWA_{grid}) for Victoria basin (Basin 12 in Figure 2.1).

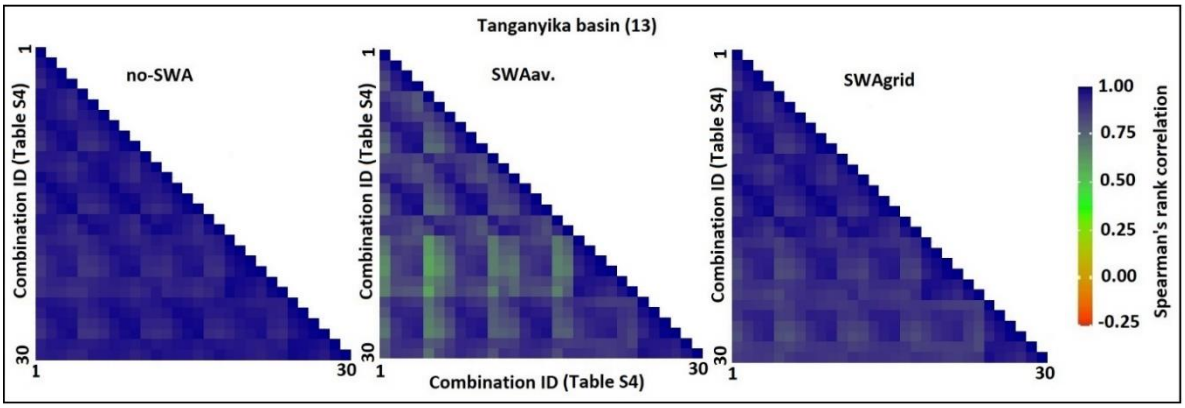


Figure A.38 Same as figure **A.37** for Tanganyika basin (Basin 13 in Figure 2.1).

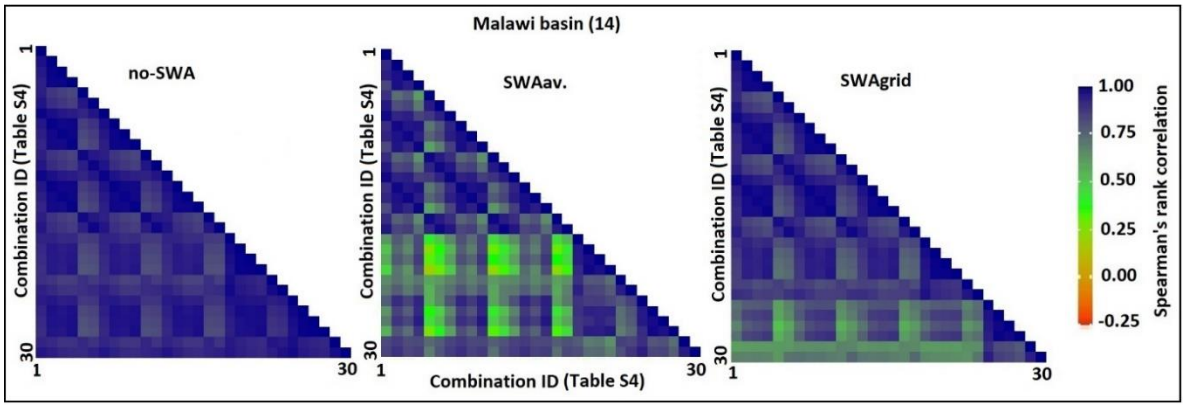


Figure A.39 Same as figure **A.37** for Malawi basin (Basin 14 in Figure 2.1).

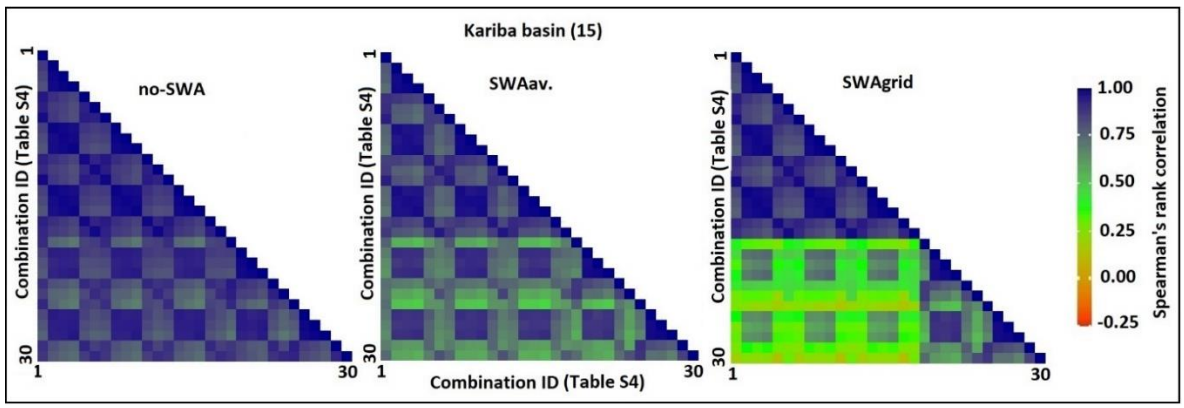


Figure A.40 Same as figure **A.37** for Kariba basin (Basin 15 in Figure 2.1).

Tables:

Basin ID	Basin name	Area (km2)	Lake storage	Area (km2)
4	Ladoga/Onega	286995.7	Ladoga	17444
			Onegh	9961
5	Rybinsk	235216.6	Rybinkskoye	4042.5
6	Kremenchuk-Kakhovka	175696.4	Kakhovskoye	2092.3
			Kremenchuk	1849.09
7	Baikal/Bratsk	754575.9	Baikal	31968
			Bratskoye	4811
8	Khovsgol	151160.9	Hovs_Gol	2768
9	Balkhash	286013.7	Balkhash	16718
10	Large Aral Sea	164010.8	Aral_Sea	23866
11	Kara-Bogaz-Gol/ Sarygamysh	570843.7	Kara-Bogaz-Gol	18666.8
			Sarygamysh	3772.03
12	Victoria	589284.1	Kyoga	2789
			Victoria	67166
13	Tanganyika	346122.7	Rukwa	5895
			Tanganyika	32827
14	Malawi	169843.1	Malawi	29544
15	Kariba	165737.5	Kariba	5277

Table A.1 Summary of additional study basins (4:15), featuring basin names, basin areas, and rankings of reference lake areas for each respective basin. (Lehner et al., 2008; Khazaei et al., 2022).

Well ID	x	y	Well ID	x	y
W1	-112.786	49.63514	W43	-112.975	53.87596
W2	-110.504	51.78894	W44	-112.855	52.96839
W3	-110.502	51.7885	W45	-113.467	50.84466
W4	-113.445	52.1027	W46	-111.7	50.97978
W5	-114.684	52.05721	W47	-112.829	53.38762
W6	-110.475	51.57186	W48	-112.829	53.3876
W7	-110.475	51.57194	W49	-110.409	54.06098
W8	-110.475	51.57199	W50	-110.408	54.06089
W9	-113.828	53.56897	W51	-113.762	53.41031
W10	-112.203	49.37844	W52	-110.534	53.58826
W11	-113.366	52.93811	W53	-111.697	53.77241
W12	-114.216	52.01171	W54	-113.642	51.93813
W13	-110.534	53.58811	W55	-112.495	50.13547
W14	-110.288	49.66097	W56	-112.238	51.00854
W15	-111.424	49.68289	W57	-114.053	52.62541
W16	-111.648	52.86495	W58	-111.322	52.67534
W17	-111.378	52.64519	W59	-111.319	52.67348
W18	-113.972	52.7452	W60	-111.378	52.64519
W19	-110.895	53.73692	W61	-114.109	53.58363
W20	-111.447	53.34187	W62	-114.109	53.58368
W21	-110.733	50.97921	W63	-111.447	53.34177
W22	-111.916	52.5503	W64	-112.939	49.95835
W23	-113.614	51.33197	W65	-112.843	50.85436
W24	-112.208	51.50361	W66	-111.869	50.86625
W25	-113.877	49.55814	W67	-111.445	51.95366
W26	-111.445	51.95367	W68	-110.169	51.41464
W27	-113.299	49.72788	W69	-113.077	49.99295
W28	-111.19	51.15663	W70	-114.288	52.04256
W29	-111.445	51.95367	W71	-113.596	52.68341
W30	-112.955	52.78546	W72	-108.869	52.81576
W31	-112.208	51.50355	W73	-107.174	51.57395
W32	-111.858	52.78743	W74	-107.315	51.5793
W33	-113.692	53.38843	W75	-107.314	51.57026
W34	-113.551	51.96683	W76	-107.315	51.57206
W35	-111.527	54.10055	W77	-106.223	52.91534
W36	-115.115	51.07367	W78	-106.223	52.91534
W37	-113.511	49.75703	W79	-108.533	49.75391
W38	-113.973	51.76559	W80	-106.508	52.05741
W39	-110.461	49.98837	W81	-106.27	52.49776
W40	-111.351	49.2376	W82	-109.564	52.08424
W41	-111.79	53.16117	W83	-108.336	49.75939
W42	-112.975	53.87591	W84	-107.917	52.75763

Table A.2 Geographic coordinates (latitude [x] and longitude [y]; WGS84) of groundwater monitoring wells (W1–W84) across the Saskatchewan basin.

Well ID	x	y			
W85	-106.532	52.17187			
W86	-107.065	51.64979			
W87	-107.514	51.88062			
W88	-108.424	51.02432			
W89	-108.955	52.46499			
W90	-107.049	52.0065			
W91	-108.896	50.37291			
W92	-106.671	52.34641			

Table A.3 Geographic coordinates (latitude [x] and longitude [y]; WGS84) of groundwater monitoring wells (W85–W92) across the Saskatchewan basin.

Well ID	x	y	Well ID	x	y
W1	49.985	-99.4634	W43	50.201	-97.183
W2	49.87	-96.9457	W44	50.167	-97.137
W3	50.009	-96.9569	W45	50.161	-96.869
W4	49.24	-97.9986	W46	50.127	-96.889
W5	50.012	-98.3316	W47	50.133	-96.945
W6	49.758	-97.1279	W48	50.12	-96.727
W7	49.852	-97.0036	W49	50.132	-96.875
W8	49.925	-96.964	W50	50.147	-96.879
W9	49.96	-96.9787	W51	50.122	-96.921
W10	49.961	-96.933	W52	50.123	-96.891
W11	50.018	-96.888	W53	50.142	-96.87
W12	49.883	-96.821	W54	50.144	-96.906
W13	49.971	-97.012	W55	50.127	-96.898
W14	50.669	-100.813	W56	50.132	-96.826
W15	49.893	-97.099	W57	50.133	-96.826
W16	49.884	-96.981	W58	50.165	-96.895
W17	49.709	-96.973	W59	50.137	-96.855
W18	49.693	-96.548	W60	50.127	-96.898
W19	49.974	-96.797	W61	50.133	-96.945
W20	49.783	-96.776	W62	50.147	-96.879
W21	52.215	-101.228	W63	50.196	-96.942
W22	50.01	-97.145	W64	50.195	-97.002
W23	50.151	-96.866	W65	50.019	-99.477
W24	50.041	-99.34	W66	49.429	-97.008
W25	49.985	-97.231	W67	49.443	-97.167
W26	49.798	-99.189	W68	49.441	-97.005
W27	49.987	-97.086	W69	49.473	-97.301
W28	50.89	-97.614	W70	49.495	-97.225
W29	50.024	-99.365	W71	49.384	-97.187
W30	49.839	-97.329	W72	49.887	-97.767
W31	49.868	-97.325	W73	49.577	-97.607
W32	49.957	-97.315	W74	49.252	-99.159
W33	49.885	-97.322	W75	49.798	-99.712
W34	49.883	-97.132	W76	49.96	-96.956
W35	49.807	-97.135	W77	49.68	-98.028
W36	49.768	-97.3	W78	49.366	-98.182
W37	49.848	-99.349	W79	49.651	-99.429
W38	50.905	-97.181	W80	49.709	-99.621
W39	50.875	-97.855	W81	49.771	-99.577
W40	49.917	-97.171	W82	49.77	-99.577
W41	49.689	-100.395	W83	49.813	-99.623
W42	49.547	-99.302	W84	49.835	-99.583

Table A.4 Geographic coordinates (latitude [x] and longitude [y]; WGS84) of groundwater monitoring wells (W1–W84) across the Red-Saskatchewan basin.

Well ID	x	y	Well ID	x	y
W85	49.874	-99.553	W127	49.62	-99.53
W86	49.875	-99.552	W128	49.65	-99.645
W87	49.775	-99.644	W129	49.769	-99.689
W88	49.843	-99.717	W130	49.665	-100.303
W89	50.378	-97.263	W131	49.727	-100.577
W90	50.385	-97.264	W132	49.459	-100.687
W91	50.398	-97.252	W133	50.257	-97.344
W92	49.535	-96.866	W134	50.3	-97.299
W93	49.536	-96.867	W135	50.255	-97.045
W94	49.546	-96.867	W136	50.241	-97.179
W95	49.813	-99.212	W137	49.532	-96.986
W96	49.685	-99.267	W138	50.226	-97.735
W97	49.96	-99.36	W139	50.222	-97.62
W98	49.931	-99.513	W140	49.9	-99.179
W99	49.932	-99.659	W141	50.108	-99.431
W100	49.91	-99.591	W142	49.812	-99.44
W101	49.975	-99.592	W143	49.798	-99.086
W102	49.879	-99.712	W144	49.634	-99.078
W103	49.579	-100.554	W145	49.635	-99.079
W104	49.575	-100.69	W146	49.604	-99.234
W105	49.459	-100.737	W147	49.739	-99.325
W106	49.607	-100.827	W148	50.064	-98.883
W107	49.724	-100.759	W149	49.931	-99.042
W108	49.435	-100.85	W150	49.828	-98.95
W109	49.606	-100.962	W151	49.827	-98.748
W110	49.657	-100.895	W152	49.739	-98.891
W111	50.256	-97.477	W153	49.974	-99.271
W112	50.241	-97.873	W154	49.916	-99.293
W113	50.24	-98.01	W155	49.842	-99.345
W114	50.742	-99.537	W156	49.853	-99.144
W115	49.148	-97.932	W157	49.916	-99.293
W116	49.236	-97.979	W158	50.034	-99.202
W117	49.266	-98.003	W159	50.035	-99.201
W118	49.283	-98.024	W160	49.931	-97.735
W119	49.371	-98.205	W161	49.517	-96.827
W120	49.69402	-98.9603	W162	49.92	-99.39
W121	49.69406	-98.9604	W163	49.95	-99.643
W122	49.236	-99.065	W164	49.709	-99.66
W123	49.576	-99.144	W165	49.706	-99.649
W124	49.742	-99.154	W166	49.75	-99.649
W125	49.53	-99.338	W167	49.802	-99.661
W126	49.589	-99.393	W168	49.927	-99.31

Table A.5 Geographic coordinates (latitude [x] and longitude [y]; WGS84) of groundwater monitoring wells (W85–W168) across the Red-Saskatchewan basin.

Well ID	x	y	Well ID	x	y
W169	49.724	-99.644	W211	49.638	-98.665
W170	49.831	-99.706	W212	49.631	-98.695
W171	49.926	-99.666	W213	49.248	-98.001
W172	49.778	-99.667	W214	49.139	-97.966
W173	49.891	-99.547	W215	50.263	-99.124
W174	49.795	-99.74	W216	49.541	-96.711
W175	49.935	-97.712	W217	49.813	-99.303
W176	49.595	-99.401	W218	49.829	-99.872
W177	49.901	-99.218	W219	49.118	-100.806
W178	50.048	-99.481	W220	49.876	-100.853
W179	49.665	-100.592	W221	49.841	-99.825
W180	49.887	-99.304	W222	49.837	-99.899
W181	49.842	-99.264	W223	51.081	-99.488
W182	49.671	-99.598	W224	52.075	-101.038
W183	49.606	-99.389	W225	52.085	-100.916
W184	49.249	-99.136	W226	52.09	-100.989
W185	50.328	-99.662	W227	52.087	-101.013
W186	49.569	-95.639	W228	49.903	-99.476
W187	49.015	-95.531	W229	49.267	-96.473
W188	49.017	-95.436	W230	49.384	-96.217
W189	49.023	-95.684	W231	49.429	-96.512
W190	49.754	-99.303	W232	49.29	-96.15
W191	49.458	-100.839	W233	49.385	-96.776
W192	49.244	-99.111	W234	49.4	-96.116
W193	49.073	-100.841	W235	49.373	-96.29
W194	49.775	-97.16	W236	49.089	-95.818
W195	49.178	-98.001	W237	49.088	-95.819
W196	49.266	-98.047	W238	49.96	-96.843
W197	49.697	-99.649	W239	49.574	-96.406
W198	49.237	-98.004	W240	50.069	-97.276
W199	49.252	-98.003	W241	50.098	-97.276
W200	49.281	-98.054	W242	49.813	-99.891
W201	49.798	-98.188	W243	49.385	-96.856
W202	50.27	-99.079	W244	49.43	-96.782
W203	50.287	-99.055	W245	49.556	-96.617
W204	49.503	-96.688	W246	49.556	-96.616
W205	49.518	-96.624	W247	49.846	-99.907
W206	49.517	-96.624	W248	49.798	-99.936
W207	49.635	-98.711	W249	49.524	-96.398
W208	49.562	-98.847	W250	49.524	-96.398
W209	49.621	-98.551	W251	53.802	-101.384
W210	49.631	-98.695	W252	49.503	-100.712

Table A.6 Geographic coordinates (latitude [x] and longitude [y]; WGS84) of groundwater monitoring wells (W169–W252) across the Red-Saskatchewan basin.

Well ID	x	y	Well ID	x	y
W253	49.267	-96.646	W295	49.954	-97.116
W254	50.239	-96.588	W296	50.147	-97.482
W255	50.238	-96.589	W297	50.138	-97.689
W256	49.073	-101.223	W298	49.355	-99.745
W257	49.831	-99.848	W299	49.429	-99.834
W258	49.823	-99.825	W300	50.159	-97.345
W259	49.83	-99.873	W301	49.604	-96.658
W260	49.533	-98.596	W302	49.634	-96.763
W261	49.591	-98.664	W303	49.858	-99.967
W262	49.406	-98.681	W304	53.05564	-103.951
W263	49.238	-98.545	W305	50.29575	-105.51
W264	49.252	-99.16	W306	50.26302	-105.49
W265	49.915	-96.934	W307	50.95145	-102.455
W266	49.348	-96.556	W308	49.26526	-103.185
W267	50.032	-96.562	W309	49.70556	-102.863
W268	50.61	-99.508	W310	52.95006	-104.448
W269	50.652	-101.09	W311	49.13863	-103.255
W270	50.506	-100.6	W312	51.45725	-105.199
W271	50.211	-100.394	W313	53.36656	-103.057
W272	50.004	-100.733	W314	51.82104	-102.41
W273	49.326	-99.698	W315	51.16701	-102.508
W274	49.059	-99.209			
W275	49.147	-98.74			
W276	49.872	-99.07			
W277	49.813	-99.267			
W278	49.977	-97.008			
W279	49.044	-99.334			
W280	49.252	-99.563			
W281	50.091	-96.358			
W282	49.156	-95.598			
W283	49.302	-95.968			
W284	49.29	-95.787			
W285	49.178	-96.16			
W286	49.866	-96.297			
W287	49.605	-95.915			
W288	49.237	-95.439			
W289	51.333	-98.352			
W290	49.147	-99.436			
W291	49.195	-99.537			
W292	49.252	-99.585			
W293	50.734	-97.763			
W294	50.02	-97.551			

Table A.7 Geographic coordinates (latitude [x] and longitude [y]; WGS84) of groundwater monitoring wells (W253–W315) across the Red-Saskatchewan basin.

Well ID	x	y	Well ID	x	y
W1	-80.2161	43.45139	W12	-79.4625	46.37908
W2	-79.5081	44.10917	W13	-79.5722	44.52528
W3	-75.6244	45.09778	W14	-81.7375	44.03778
W4	-80.6008	43.36444	W15	-80.3667	43.33694
W5	-80.2161	43.45139	W16	-80.3853	43.29194
W6	-80.5106	42.65861	W17	-72.5281	46.3839
W7	-78.4078	44.05778	W18	-71.9098	46.8547
W8	-78.2992	44.08417	W19	-71.6671	46.7719
W9	-78.9969	43.93472	W20	-81.3319	41.42811
W10	-80.9281	44.14806	W21	-72.1939	44.665
W11	-79.7419	44.435			

Table A.8 Geographic coordinates (latitude [x] and longitude [y]; WGS84) of groundwater monitoring wells (W1–W21) across the Ottawa basin.

n	GRACE-TWSA	SWEA	SMA	SWA		
				a	b	c
1	CSR-M	NOAH1	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
2	CSR-M	NOAH1	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
3	CSR-M	NOAH1	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
4	CSR-M	NOAH1	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
5	CSR-M	NOAH1	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
6	CSR-M	NOAH1	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
7	CSR-M	NOAH 0.1	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
8	CSR-M	NOAH 0.1	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
9	CSR-M	NOAH 0.1	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
10	CSR-M	NOAH 0.1	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
11	CSR-M	NOAH 0.1	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
12	CSR-M	NOAH 0.1	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
13	CSR-M	VIC	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
14	CSR-M	VIC	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
15	CSR-M	VIC	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
16	CSR-M	VIC	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
17	CSR-M	VIC	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
18	CSR-M	VIC	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
19	CSR-M	CLSM	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
20	CSR-M	CLSM	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
21	CSR-M	CLSM	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
22	CSR-M	CLSM	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
23	CSR-M	CLSM	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
24	CSR-M	CLSM	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
25	CSR-M	WGHM	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
26	CSR-M	WGHM	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
27	CSR-M	WGHM	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
28	CSR-M	WGHM	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
29	CSR-M	WGHM	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
30	CSR-M	WGHM	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
31	CSR-M	GlobSnow	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
32	CSR-M	GlobSnow	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
33	CSR-M	GlobSnow	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
34	CSR-M	GlobSnow	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
35	CSR-M	GlobSnow	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
36	CSR-M	GlobSnow	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
37	CSR-M	ERA5Land	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
38	CSR-M	ERA5Land	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
39	CSR-M	ERA5Land	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
40	CSR-M	ERA5Land	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
41	CSR-M	ERA5Land	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
42	CSR-M	ERA5Land	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}

43	JPL-M	NOAH1	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
44	JPL-M	NOAH1	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
45	JPL-M	NOAH1	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
46	JPL-M	NOAH1	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
47	JPL-M	NOAH1	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
48	JPL-M	NOAH1	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
49	JPL-M	NOAH 0.1	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
50	JPL-M	NOAH 0.1	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
51	JPL-M	NOAH 0.1	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
52	JPL-M	NOAH 0.1	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
53	JPL-M	NOAH 0.1	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
54	JPL-M	NOAH 0.1	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
55	JPL-M	VIC	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
56	JPL-M	VIC	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
57	JPL-M	VIC	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
58	JPL-M	VIC	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
59	JPL-M	VIC	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
60	JPL-M	VIC	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
61	JPL-M	CLSM	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
62	JPL-M	CLSM	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
63	JPL-M	CLSM	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
64	JPL-M	CLSM	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
65	JPL-M	CLSM	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
66	JPL-M	CLSM	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
67	JPL-M	WGHM	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
68	JPL-M	WGHM	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
69	JPL-M	WGHM	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
70	JPL-M	WGHM	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
71	JPL-M	WGHM	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
72	JPL-M	WGHM	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
73	JPL-M	GlobSnow	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
74	JPL-M	GlobSnow	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
75	JPL-M	GlobSnow	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
76	JPL-M	GlobSnow	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
77	JPL-M	GlobSnow	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
78	JPL-M	GlobSnow	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
79	JPL-M	ERA5Land	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
80	JPL-M	ERA5Land	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
81	JPL-M	ERA5Land	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
82	JPL-M	ERA5Land	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
83	JPL-M	ERA5Land	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
84	JPL-M	ERA5Land	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
85	CSR-SH	NOAH1	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
86	CSR-SH	NOAH1	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}

87	CSR-SH	NOAH1	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
88	CSR-SH	NOAH1	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
89	CSR-SH	NOAH1	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
90	CSR-SH	NOAH1	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
91	CSR-SH	NOAH 0.1	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
92	CSR-SH	NOAH 0.1	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
93	CSR-SH	NOAH 0.1	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
94	CSR-SH	NOAH 0.1	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
95	CSR-SH	NOAH 0.1	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
96	CSR-SH	NOAH 0.1	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
97	CSR-SH	VIC	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
98	CSR-SH	VIC	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
99	CSR-SH	VIC	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
100	CSR-SH	VIC	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
101	CSR-SH	VIC	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
102	CSR-SH	VIC	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
103	CSR-SH	CLSM	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
104	CSR-SH	CLSM	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
105	CSR-SH	CLSM	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
106	CSR-SH	CLSM	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
107	CSR-SH	CLSM	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
108	CSR-SH	CLSM	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
109	CSR-SH	WGHM	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
110	CSR-SH	WGHM	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
111	CSR-SH	WGHM	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
112	CSR-SH	WGHM	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
113	CSR-SH	WGHM	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
114	CSR-SH	WGHM	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
115	CSR-SH	GlobSnow	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
116	CSR-SH	GlobSnow	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
117	CSR-SH	GlobSnow	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
118	CSR-SH	GlobSnow	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
119	CSR-SH	GlobSnow	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
120	CSR-SH	GlobSnow	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
121	CSR-SH	ERA5Land	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
122	CSR-SH	ERA5Land	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
123	CSR-SH	ERA5Land	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
124	CSR-SH	ERA5Land	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
125	CSR-SH	ERA5Land	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
126	CSR-SH	ERA5Land	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
127	JPL-SH	NOAH1	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
128	JPL-SH	NOAH1	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
129	JPL-SH	NOAH1	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
130	JPL-SH	NOAH1	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}

131	JPL-SH	NOAH1	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
132	JPL-SH	NOAH1	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
133	JPL-SH	NOAH 0.1	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
134	JPL-SH	NOAH 0.1	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
135	JPL-SH	NOAH 0.1	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
136	JPL-SH	NOAH 0.1	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
137	JPL-SH	NOAH 0.1	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
138	JPL-SH	NOAH 0.1	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
139	JPL-SH	VIC	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
140	JPL-SH	VIC	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
141	JPL-SH	VIC	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
142	JPL-SH	VIC	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
143	JPL-SH	VIC	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
144	JPL-SH	VIC	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
145	JPL-SH	CLSM	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
146	JPL-SH	CLSM	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
147	JPL-SH	CLSM	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
148	JPL-SH	CLSM	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
149	JPL-SH	CLSM	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
150	JPL-SH	CLSM	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
151	JPL-SH	WGHM	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
152	JPL-SH	WGHM	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
153	JPL-SH	WGHM	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
154	JPL-SH	WGHM	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
155	JPL-SH	WGHM	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
156	JPL-SH	WGHM	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
157	JPL-SH	GlobSnow	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
158	JPL-SH	GlobSnow	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
159	JPL-SH	GlobSnow	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
160	JPL-SH	GlobSnow	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
161	JPL-SH	GlobSnow	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
162	JPL-SH	GlobSnow	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
163	JPL-SH	ERA5Land	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
164	JPL-SH	ERA5Land	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
165	JPL-SH	ERA5Land	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
166	JPL-SH	ERA5Land	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
167	JPL-SH	ERA5Land	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
168	JPL-SH	ERA5Land	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
169	GFZ-SH	NOAH1	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
170	GFZ-SH	NOAH1	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
171	GFZ-SH	NOAH1	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
172	GFZ-SH	NOAH1	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
173	GFZ-SH	NOAH1	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
174	GFZ-SH	NOAH1	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}

175	GFZ-SH	NOAH 0.1	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
176	GFZ-SH	NOAH 0.1	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
177	GFZ-SH	NOAH 0.1	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
178	GFZ-SH	NOAH 0.1	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
179	GFZ-SH	NOAH 0.1	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
180	GFZ-SH	NOAH 0.1	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
181	GFZ-SH	VIC	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
182	GFZ-SH	VIC	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
183	GFZ-SH	VIC	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
184	GFZ-SH	VIC	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
185	GFZ-SH	VIC	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
186	GFZ-SH	VIC	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
187	GFZ-SH	CLSM	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
188	GFZ-SH	CLSM	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
189	GFZ-SH	CLSM	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
190	GFZ-SH	CLSM	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
191	GFZ-SH	CLSM	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
192	GFZ-SH	CLSM	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
193	GFZ-SH	WGHM	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
194	GFZ-SH	WGHM	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
195	GFZ-SH	WGHM	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
196	GFZ-SH	WGHM	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
197	GFZ-SH	WGHM	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
198	GFZ-SH	WGHM	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
199	GFZ-SH	GlobSnow	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
200	GFZ-SH	GlobSnow	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
201	GFZ-SH	GlobSnow	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
202	GFZ-SH	GlobSnow	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
203	GFZ-SH	GlobSnow	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
204	GFZ-SH	GlobSnow	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
205	GFZ-SH	ERA5Land	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
206	GFZ-SH	ERA5Land	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
207	GFZ-SH	ERA5Land	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
208	GFZ-SH	ERA5Land	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
209	GFZ-SH	ERA5Land	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
210	GFZ-SH	ERA5Land	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}

Table A.9 Summary of datasets combinations for 630 GRACE-GWA estimates across basins numbered 4:11. The first 210 realizations exclude surface water storage anomalies (no-SWA), the second 210 apply a basin-average surface water storage estimate ($SWA_{av.}$), and the final 210 incorporate surface water storage using the grid approach SWA_{grid}.

n	GRACE-TWSA	SMA	SWA		
			a	b	c
1	CSR-M	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
2	CSR-M	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
3	CSR-M	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
4	CSR-M	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
5	CSR-M	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
6	CSR-M	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
7	JPL-M	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
8	JPL-M	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
9	JPL-M	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
10	JPL-M	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
11	JPL-M	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
12	JPL-M	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
13	CSR-SH	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
14	CSR-SH	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
15	CSR-SH	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
16	CSR-SH	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
17	CSR-SH	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
18	CSR-SH	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
19	JPL-SH	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
20	JPL-SH	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
21	JPL-SH	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
22	JPL-SH	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
23	JPL-SH	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
24	JPL-SH	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}
25	GFZ-SH	NOAH1	no-SWA	$SWA_{av.}$	SWA_{grid}
26	GFZ-SH	NOAH 0.1	no-SWA	$SWA_{av.}$	SWA_{grid}
27	GFZ-SH	VIC	no-SWA	$SWA_{av.}$	SWA_{grid}
28	GFZ-SH	CLSM	no-SWA	$SWA_{av.}$	SWA_{grid}
29	GFZ-SH	WGHM	no-SWA	$SWA_{av.}$	SWA_{grid}
30	GFZ-SH	ERA5Land	no-SWA	$SWA_{av.}$	SWA_{grid}

Table A.10 Summary of datasets combinations for 30 GRACE-GWA estimates across basins numbered 12:15.

Appendix B. Supporting Information for Chapter 3

Contents of this file:

Introduction

This supporting information includes 12 figures and 3 tables that complement chapter 3 by presenting the results of in-situ groundwater spatial interpolation method tests conducted in the study, as well as detailed outputs from the normalisation procedure and the specific yield (Sy) values derived from the PCR-GLOBWB model.

Figures:

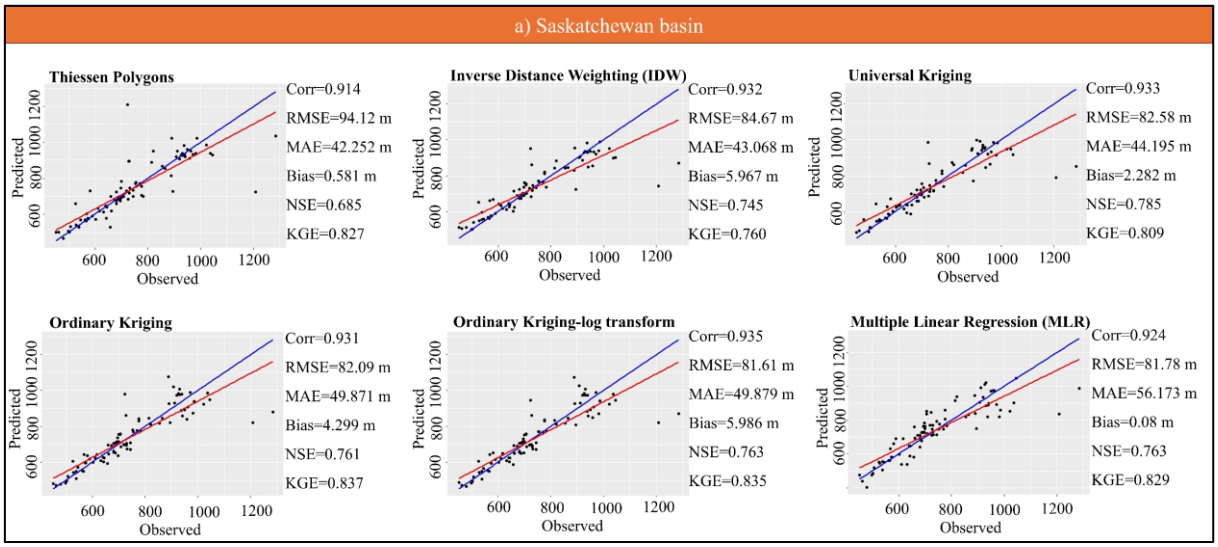


Figure B.1 Leave-One-Out Cross-Validation for different spatial interpolation methods across the Saskatchewan basin, focusing on the month of March 2003.

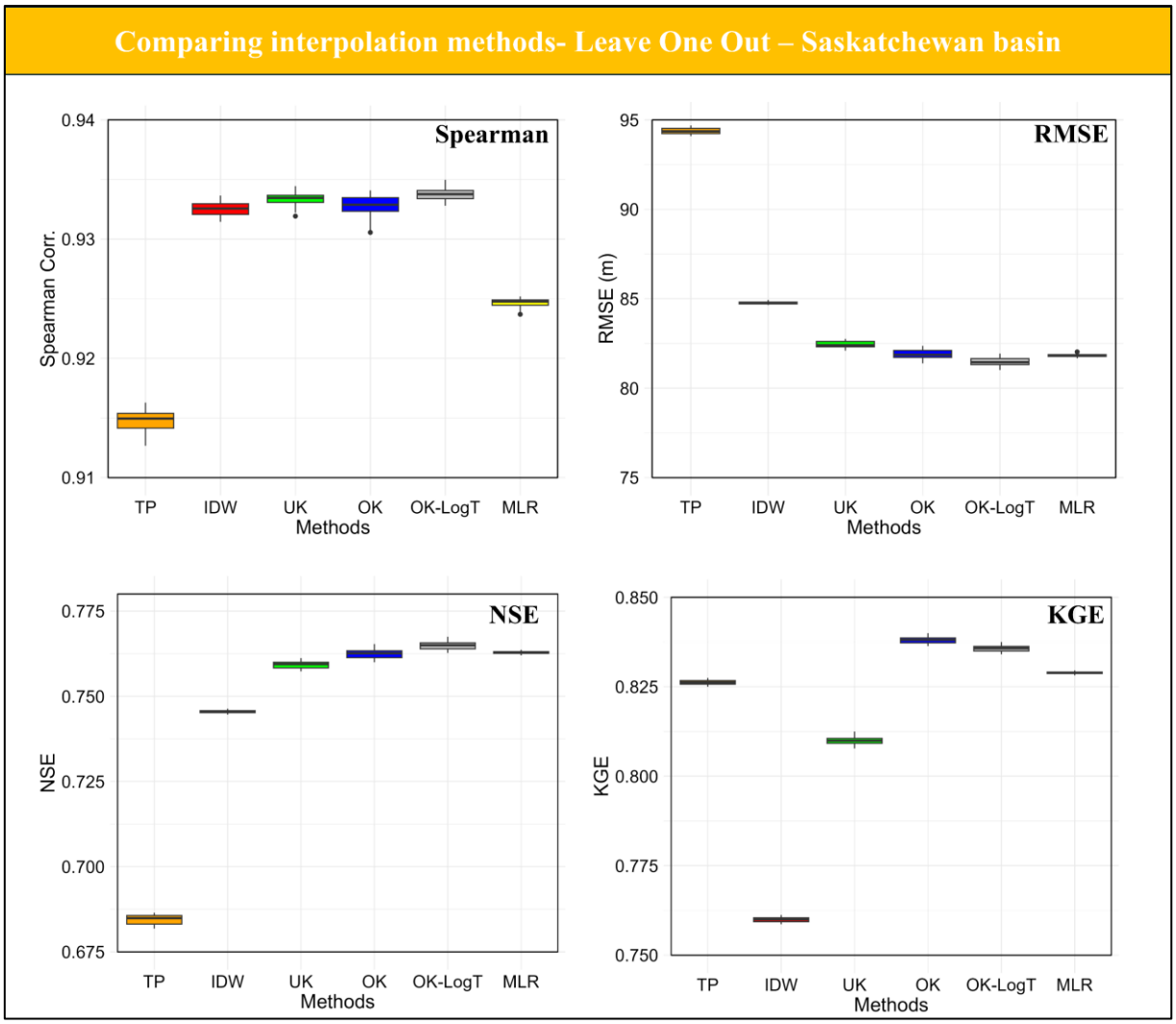


Figure B.2 Correlation and Goodness-Of-Fit metrics derived from Leave-One-Out Cross-Validation, evaluating the performance of multiple spatial interpolation methods across the Saskatchewan basin over the entire study period.

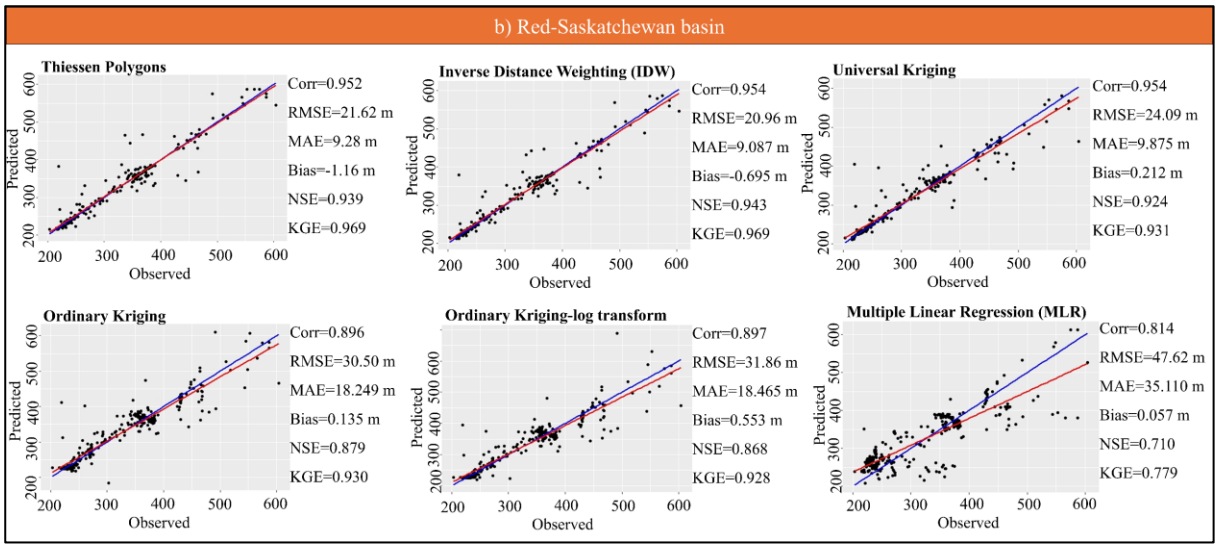


Figure B.3 Leave-One-Out Cross-Validation for different spatial interpolation methods across the Red-Saskatchewan basin, focusing on the month of March 2003.

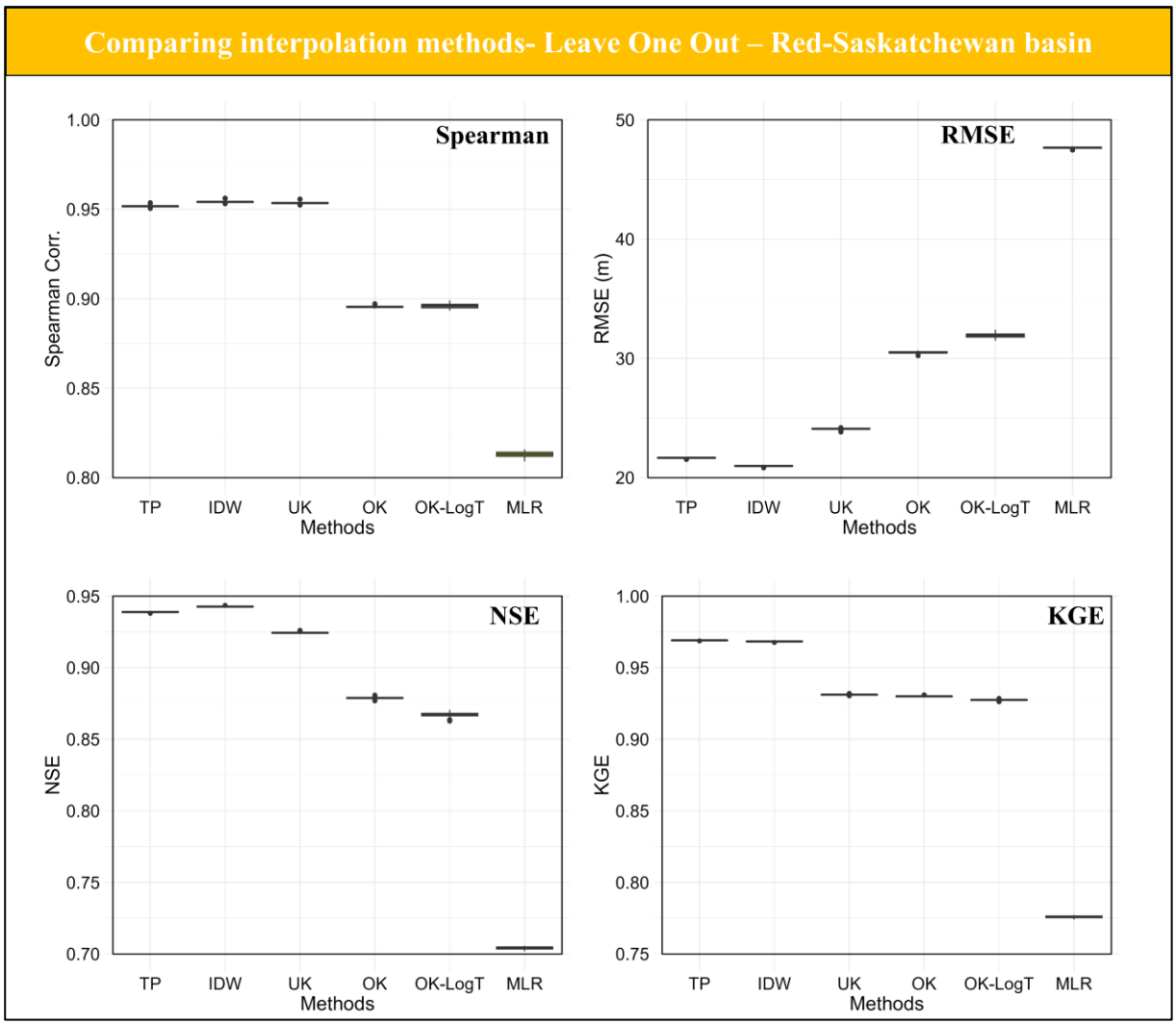


Figure B.4 Correlation and Goodness-Of-Fit metrics derived from Leave-One-Out Cross-Validation, evaluating the performance of multiple spatial interpolation methods across the Red-Saskatchewan basin over the entire study period.

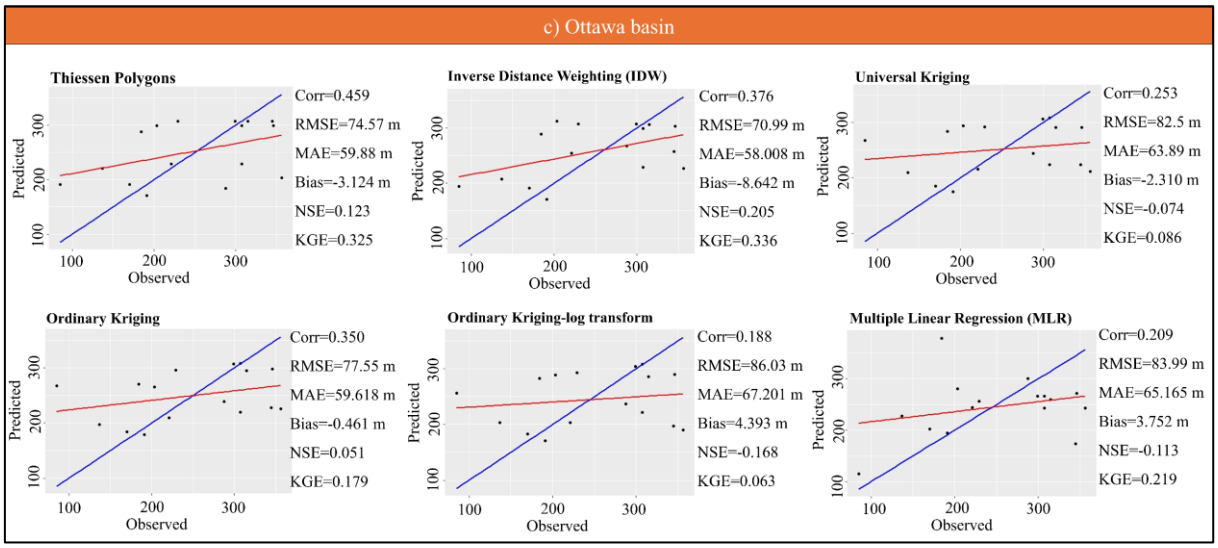


Figure B.5 Leave-One-Out Cross-Validation for different spatial interpolation methods across the Ottawa basin, focusing on the month of March 2003.

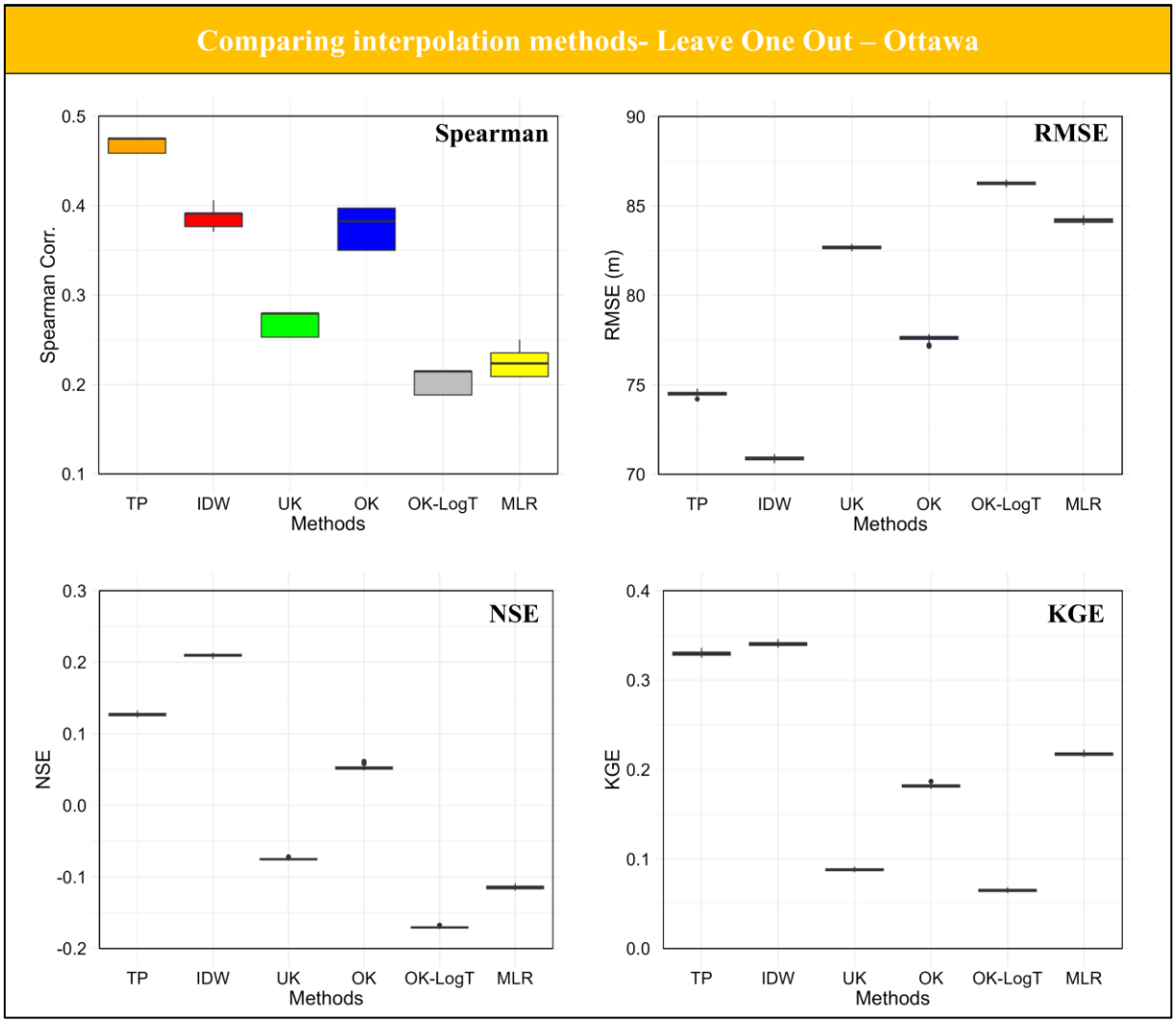


Figure B.6 Correlation and Goodness-Of-Fit metrics derived from Leave-One-Out Cross-Validation, evaluating the performance of multiple spatial interpolation methods across the Ottawa basin over the entire study period.

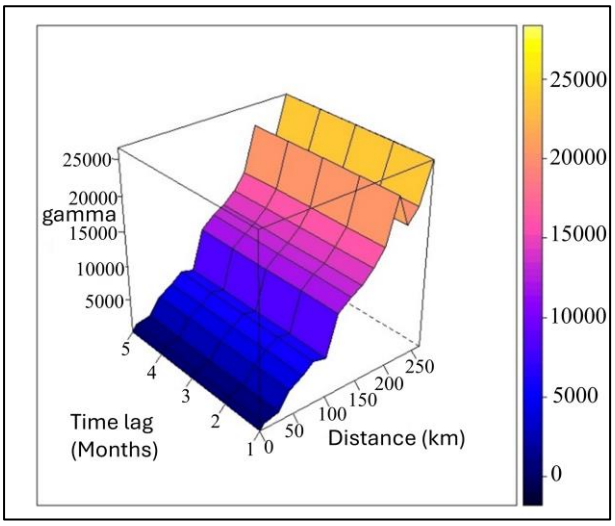


Figure B.7 Simple Spatio-temporal variogram, revealing the lack of significance in the temporal dimension across the Saskatchewan basin. A similar temporal behaviour was observed across the Red-Saskatchewan and Ottawa basins.

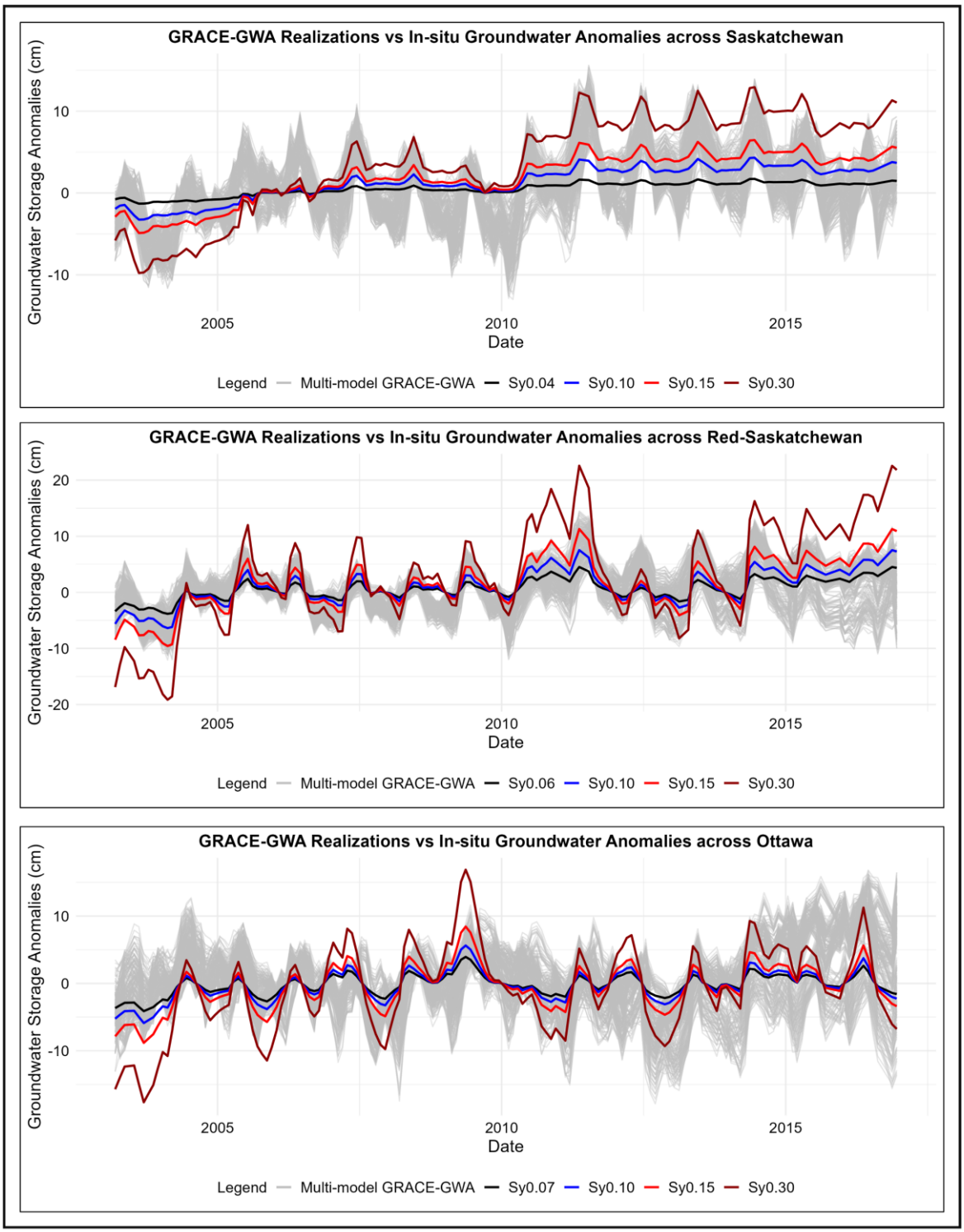


Figure B.8 Time series comparison of 240 GRACE-GWA realisations (grey lines) and in-situ groundwater anomalies derived using specific yield (Sy) values of 0.04/0.06/0.07, 0.10, 0.15, and 0.30 (black, blue, red, and dark red lines, respectively) for the Saskatchewan, Red-Saskatchewan, and Ottawa basins.

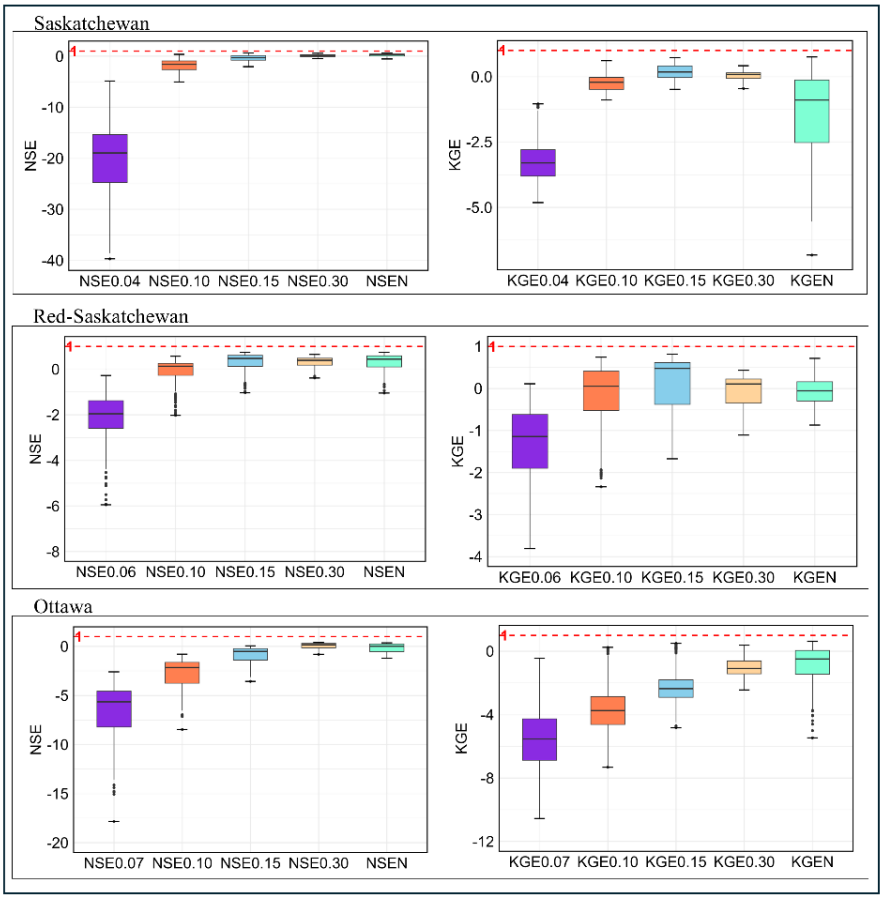


Figure B.9 Box plots illustrating the distribution of NSE and KGE metrics for 240 GRACE-GWA realizations, compared against basin-scale in-situ groundwater anomalies across the Saskatchewan, Red-Saskatchewan, and Ottawa basins. Analysis includes normalised time series comparisons (denoted as NSEN and KGEN), as well as evaluations employing varying Sy values: 0.10, 0.15, and 0.30, represented as NSE0.10, NSE0.15, NSE0.30, and KGE0.10, KGE0.15, KGE0.30, respectively. Additional assessments incorporate Sy values derived from the PCR-GLOBWB model: 0.04 for the Saskatchewan basin, 0.06 for the Red–Saskatchewan basin, and 0.07 for the Ottawa basin.

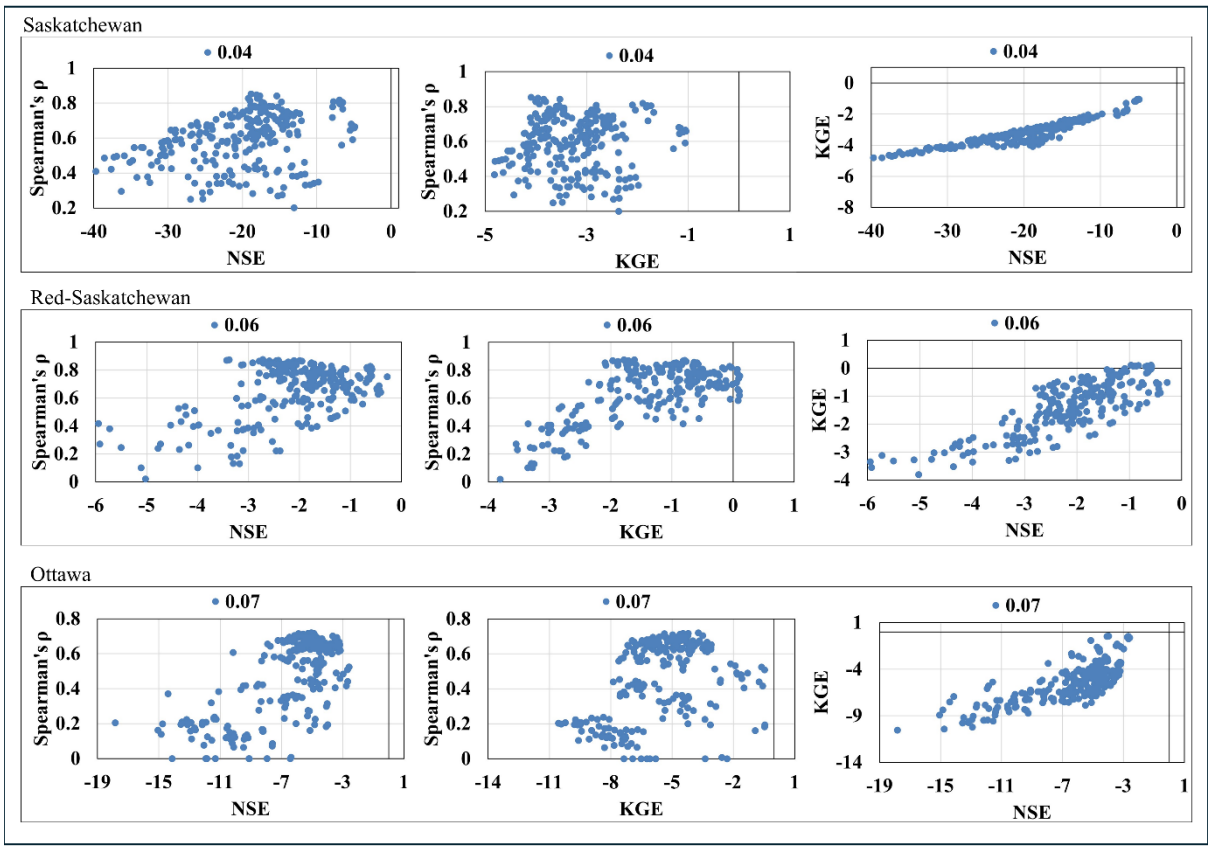


Figure B.10 Scatter plots illustrating relationships among Spearman's ρ, NSE, and KGE for 240 GRACE-GWA comparisons against basin-scale in-situ groundwater anomalies across the Saskatchewan, Red-Saskatchewan, and Ottawa basins. Panels display Spearman's ρ versus NSE (left), Spearman's ρ versus KGE (middle), and NSE versus KGE (right). The analysis incorporates Sy values derived from the PCR-GLOBWB model: 0.04 for the Saskatchewan basin, 0.06 for the Red–Saskatchewan basin, and 0.07 for the Ottawa basin.

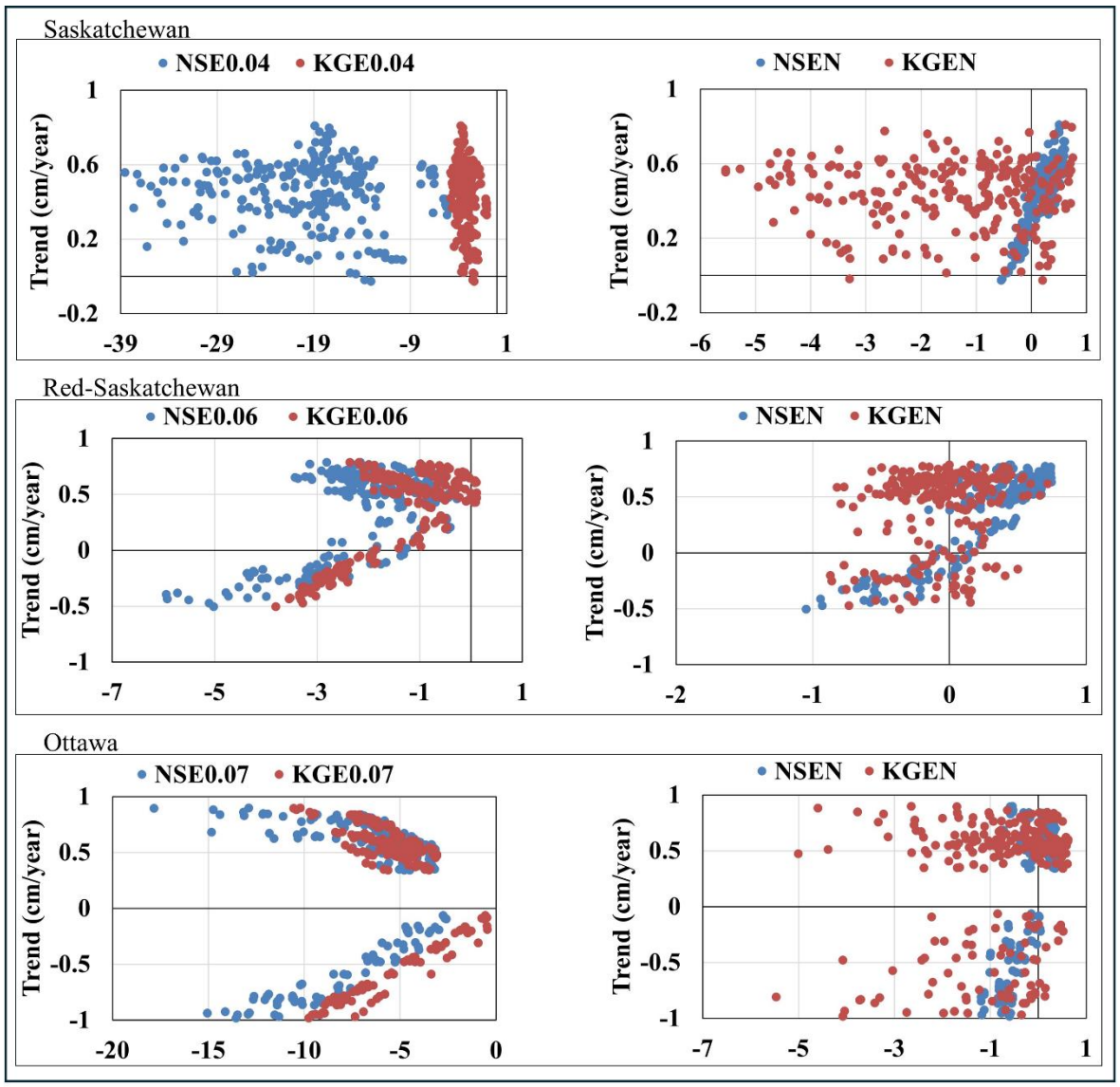


Figure B.11 Scatter plots examining the relationships between GRACE-GWA trends and GOF metrics (NSE and KGE) across 240 GRACE-GWA realisations for the Saskatchewan, Red-Saskatchewan, and Ottawa basins. NSE and KGE were computed by evaluating GRACE-GWA realisations against basin-scale in-situ groundwater anomalies, using normalisation (NSEN, KGEN) and Sy values derived from the PCR-GLOBWB model: 0.04 for the Saskatchewan basin (NSE0.04, KGE0.04), 0.06 for the Red–Saskatchewan basin (NSE0.06, KGE0.06), and 0.07 for the Ottawa basin (NSE0.07, KGE0.07).

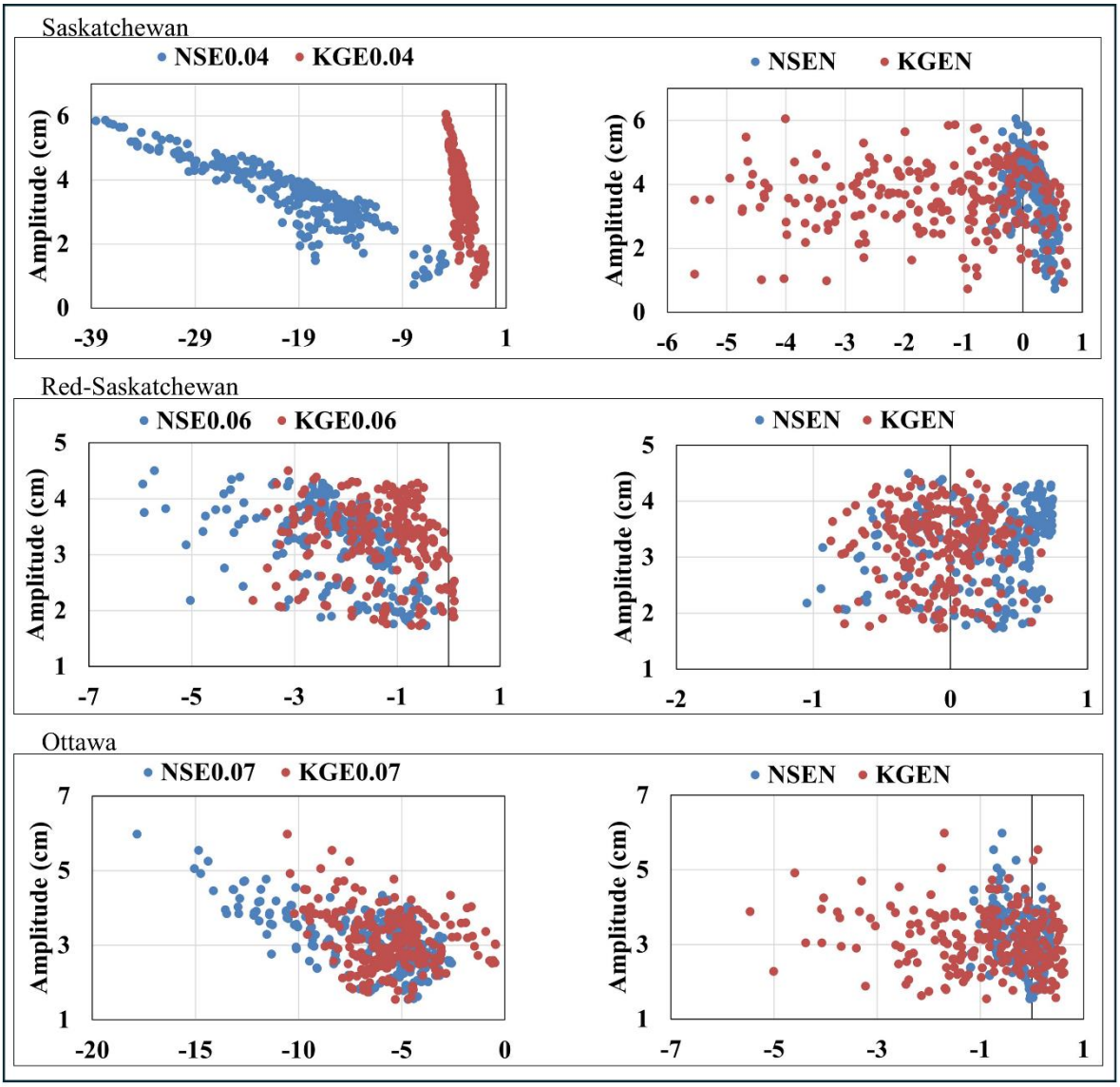


Figure B.12 Scatter plots examining the relationships between GRACE-GWA amplitudes and GOF metrics (NSE and KGE) across 240 GRACE-GWA realisations for the Saskatchewan, Red-Saskatchewan, and Ottawa basins. NSE and KGE were computed by evaluating GRACE-GWA realisations against basin-scale in-situ groundwater anomalies, using normalisation (NSEN, KGEN) and Sy values derived from the PCR-GLOBWB model: 0.04 for the Saskatchewan basin (NSE0.04, KGE0.04), 0.06 for the Red–Saskatchewan basin (NSE0.06, KGE0.06), and 0.07 for the Ottawa basin (NSE0.07, KGE0.07).

Tables:

Year	n	References
2006	2	Swenson et al., 2006; Yeh et al., 2006
2007	2	Rodell et al., 2007; Strassberg et al., 2007
2008	1	Swenson et al., 2008
2009	2	Leblanc et al., 2009; Strassberg et al., 2009
2010	2	Chen et al., 2010; Sun et al., 2010
2011	2	Frappart et al., 2011b; Henry et al., 2011
2012	6	Cao et al., 2012; Moore & Fisher, 2012; Scanlon et al., 2012; Shamsudduha et al., 2012; Tregoning et al., 2012; Van Dijk et al., 2012
2013	3	Chinnasamy et al., 2013; Feng et al., 2013; Jin & Feng, 2013
2014	3	Castle et al., 2014; Forootan et al., 2014; Joodaki et al., 2014
2015	3	Panda & Wahr, 2015; Shen et al., 2015; Xiao et al., 2015
2016	11	Bhanja et al., 2016; Chen et al., 2016b; Fatolazadeh et al., 2016; Huang et al., 2016; Iqbal et al., 2016; Liesch & Ohmer, 2016; Long et al., 2016; Nanteza et al., 2016; Rzepecka et al., 2016; Seo & Lee, 2016; Touré et al., 2016
2017	2	Hachborn et al., 2017; Yin et al., 2017
2018	6	Bhanja et al., 2018; Brookfield et al., 2018; Feng et al., 2018; Katpatal et al., 2018; Xie et al., 2018; Zhong et al., 2018
2019	6	Abou Zaki et al., 2019; Chen et al., 2019; Majid & Ardalan, 2019; Meghwal et al., 2019; Melati et al., 2019; Śliwińska et al., 2019
2020	14	Bhanja et al., 2020; Chanu et al., 2020; Chen et al., 2020; Neves et al., 2020; Rateb et al., 2020; Rzepecka & Birylo, 2020; Salam et al., 2020; Sarkar et al., 2020; Singh & Saravanan, 2020; Skaskevych et al., 2020; Su et al., 2020; Wang et al., 2020a; Zhang et al., 2020; Zheng et al., 2020
2021	6	Han et al., 2021; Kinouchi, 2021; Liu et al., 2021; Xu et al., 2021; Zhang et al., 2021a, b
2022	7	Akhtar et al., 2022; Akl et al., 2022; Ali et al., 2022; Fatolazadeh & Goïta, 2022; Liu et al., 2022b; Ouatiki et al., 2022; Ramjeawon et al., 2022
2023	9	Akl & Thomas, 2023; Alghafli et al., 2023; Amiri et al., 2023; Huang et al., 2023a, b; Nazari et al., 2023; Shao & Liu, 2023; Tariq et al., 2023; Wang et al., 2023
2024	7	Camilo et al., 2024; Jawadi et al., 2024; Nenweli et al., 2024; Qu et al., 2024; Rzepecka et al., 2024; Upadhyay et al., 2024; Yang et al., 2024

Table B.1 Complete list of 94 GRACE-GWA validation studies retrieved from Google Scholar and Scopus, accessed on July 20, 2024.

SMA / SWEA	NOAH 0.1˚	NOAH 1˚	VIC	CLSM	WGHM	ERA5-Land
NOAH 0.1˚	1	2	3	4	5	6
NOAH 1˚	7	8	9	10	11	12
VIC	13	14	15	16	17	18
CLSM	19	20	21	22	23	24
HGSWE	25	26	27	28	29	30
WGHM	31	32	33	34	35	36
GlobSnow	37	38	39	40	41	42
ERA5-Land	43	44	45	46	47	48

Table B.2 Listing of 48 combinations of Snow water Equivalent Anomalies (SWEA) and Soil Moisture Anomalies (SMA). These combinations were subsequently utilized in conjunction with GRACE-TWSA and Surface Water Anomalies (SWA) within the water budget equation to derive GRACE-GWA estimates.

GRACE-TWSA	SWEA+SMA	SWA	GRACE-GWA estimates
CSR-M	48 combs. (Table S2)	SWA	1:48
JPL-M			49:96
CSR-SH			97:144
JPL-SH			145:192
GFZ-SH			193:240

Table B.3 Numbering system for 240 realisations of GRACE-GWA, derived through the integration of 5 GRACE-TWSA solutions, 48 combinations of SWEA and SMA (as presented in Table S2), and SWA within water budget framework.

Appendix C. Supporting Information for Chapter 4

Contents of this file

- Figures C.1 – C.40

- Tables C.1 – C.6

Introduction

This supporting information comprises 40 figures and 6 tables that supplement chapter 4 by providing details of the water budget components across study basins, along with the results of the GRACE-groundwater drought index (GGDI).

Figures:

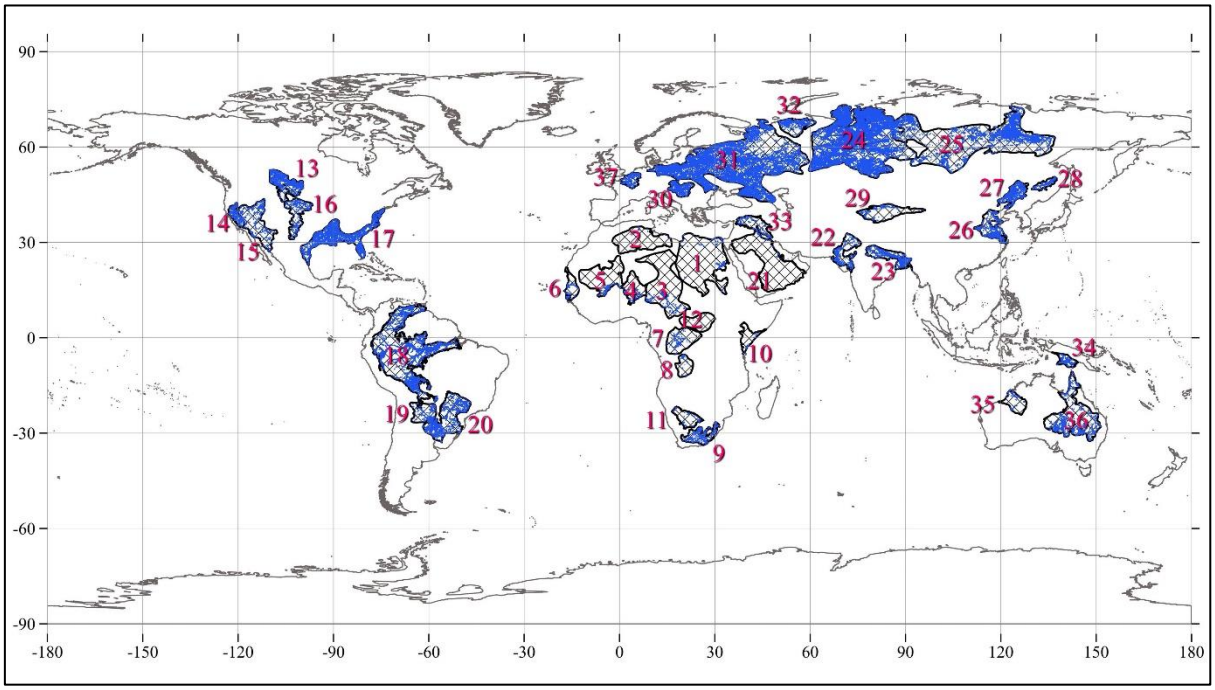


Figure C.1 Spatial distribution of surface water bodies, including lakes and reservoirs (depicted in blue), across the 37 largest aquifer systems. Data is sourced from the Hydrological Data and Maps Based on Shuttle Elevation Derivatives at Multiple Scales (HydroSHEDS; Lehner et al., 2008).

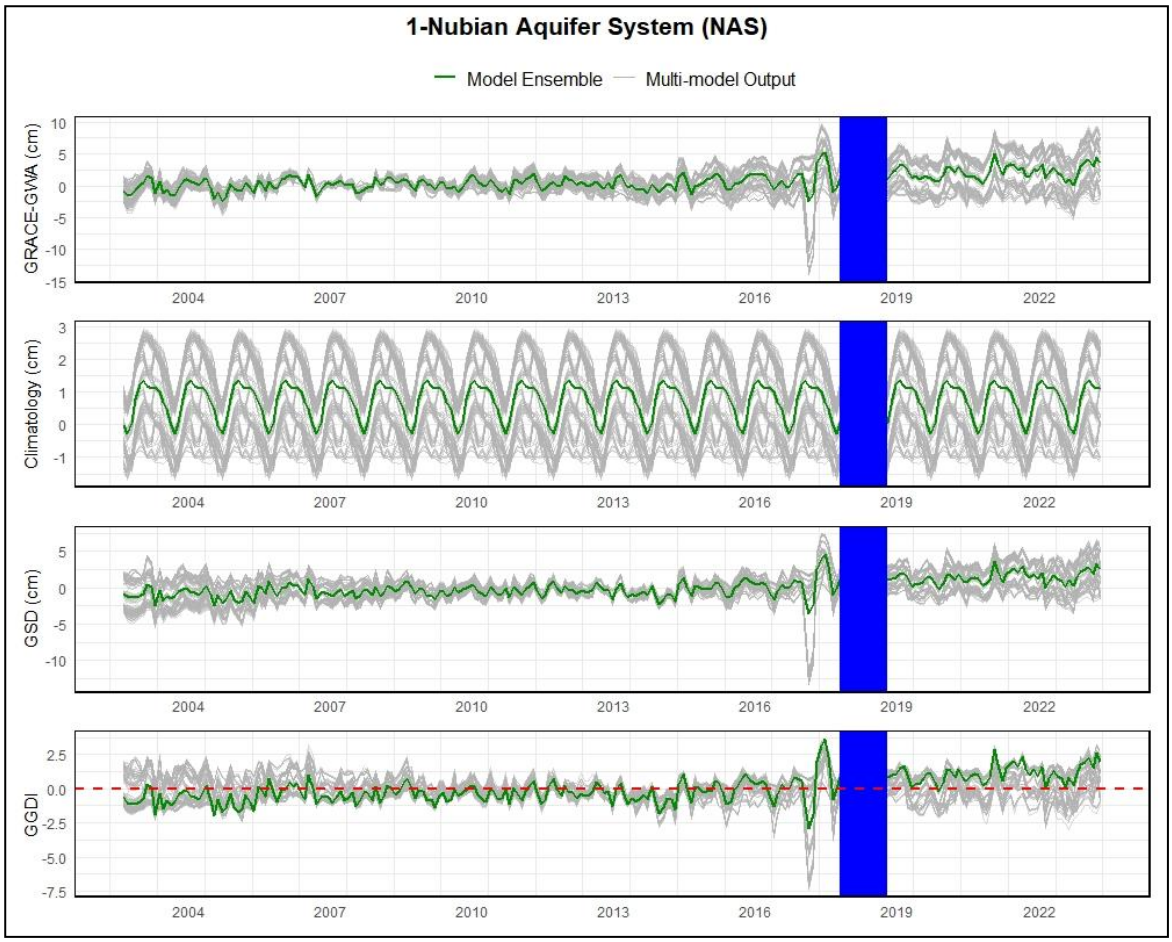


Figure C.2 Time series of GRACE-ground water storage anomalies (GRACE-GWA) and associated climatology, groundwater storage defects (GSD), and GRACE-groundwater drought index (GGDI) across Basin 1(Nubian Aquifer System [NAS]) from April 2002 to December 2022. The grey-shaded region marks the gap period between the GRACE and GRACE-FO missions, which was excluded from the analysis.

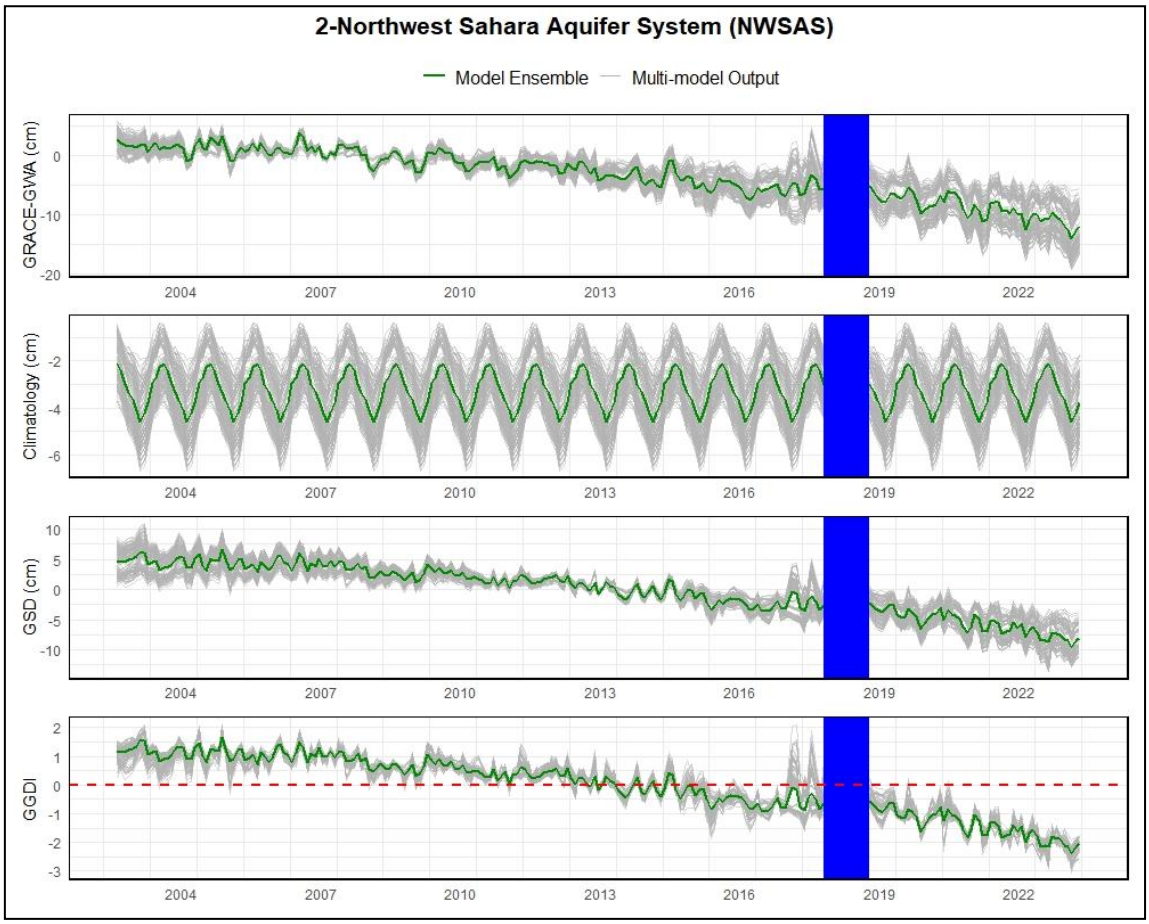


Figure C.3 Same as figure **C.3** for Basin 2 (Northwest Sahara Aquifer System [NWSAS]).

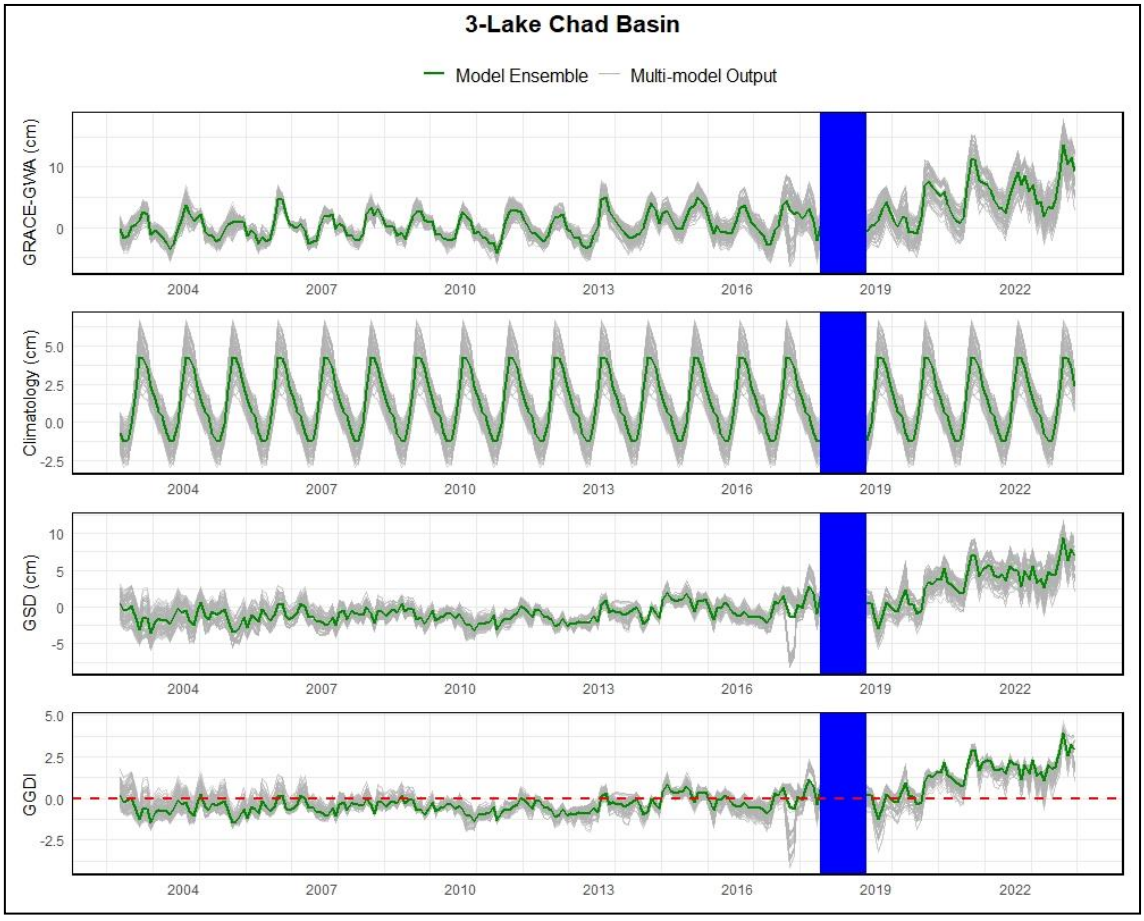


Figure C.4 Same as figure **C.3** for Basin 3 (Lake Chad Basin).

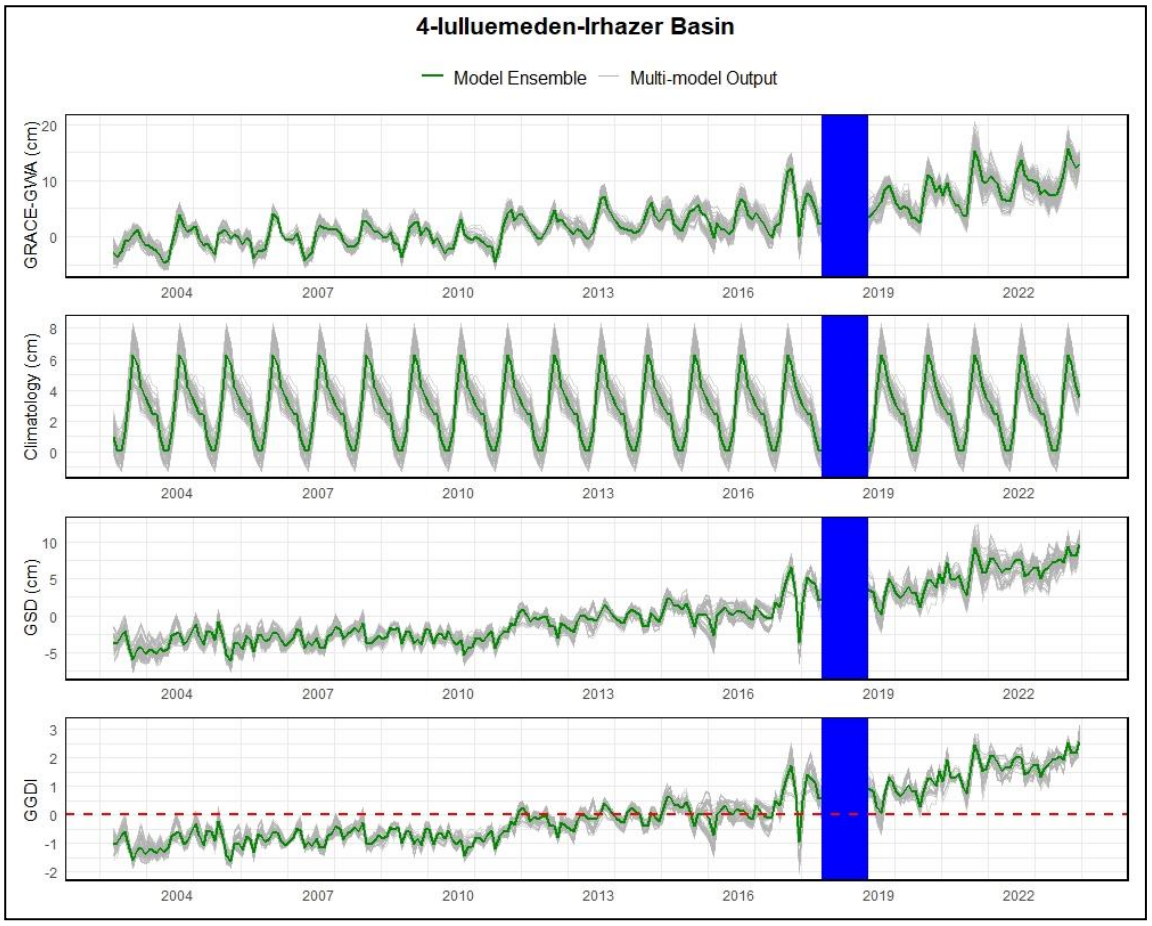


Figure C.5 Same as figure **C.3** for Basin 4 (Iulluemeden-Irhazer Basin).

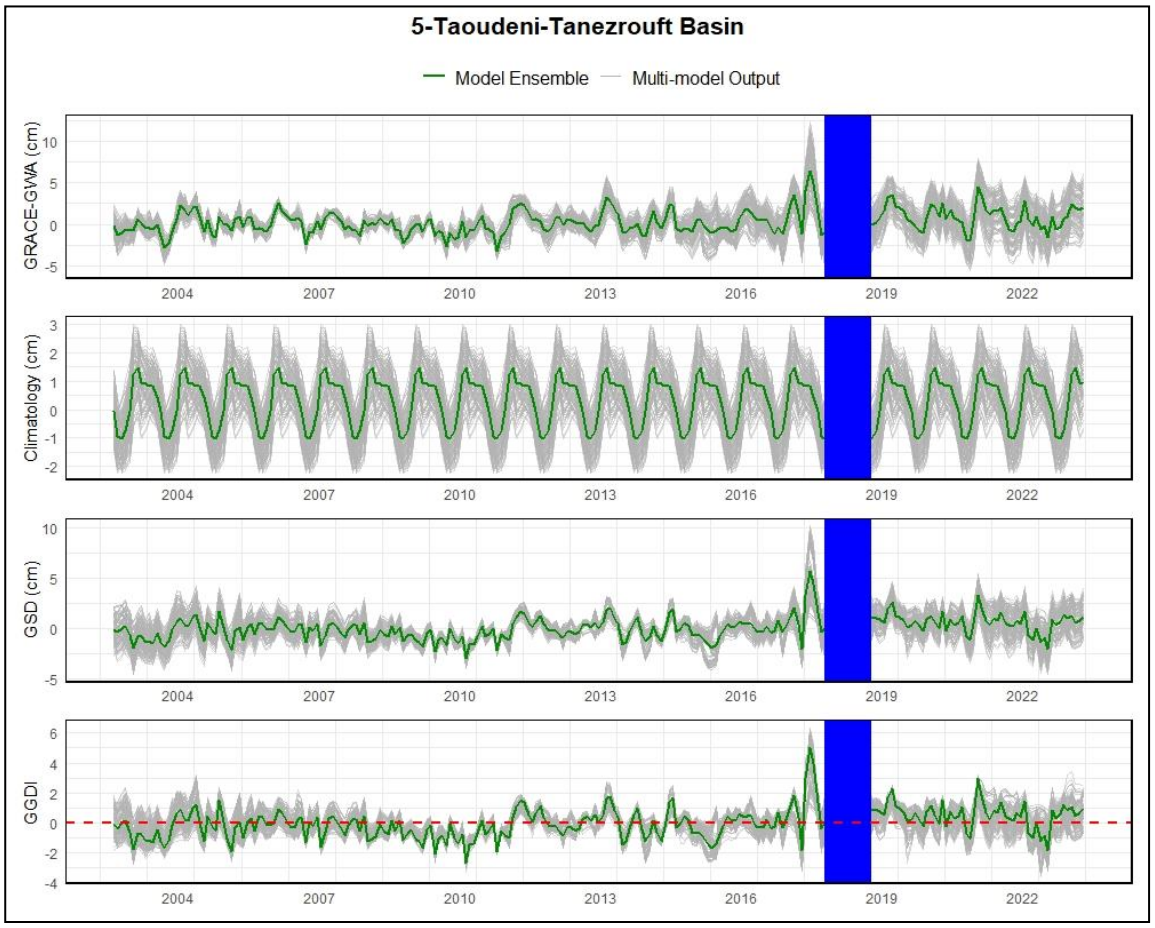


Figure C.6 Same as figure **C.3** for Basin 5 (Taoudeni-Tanezrouft Basin).

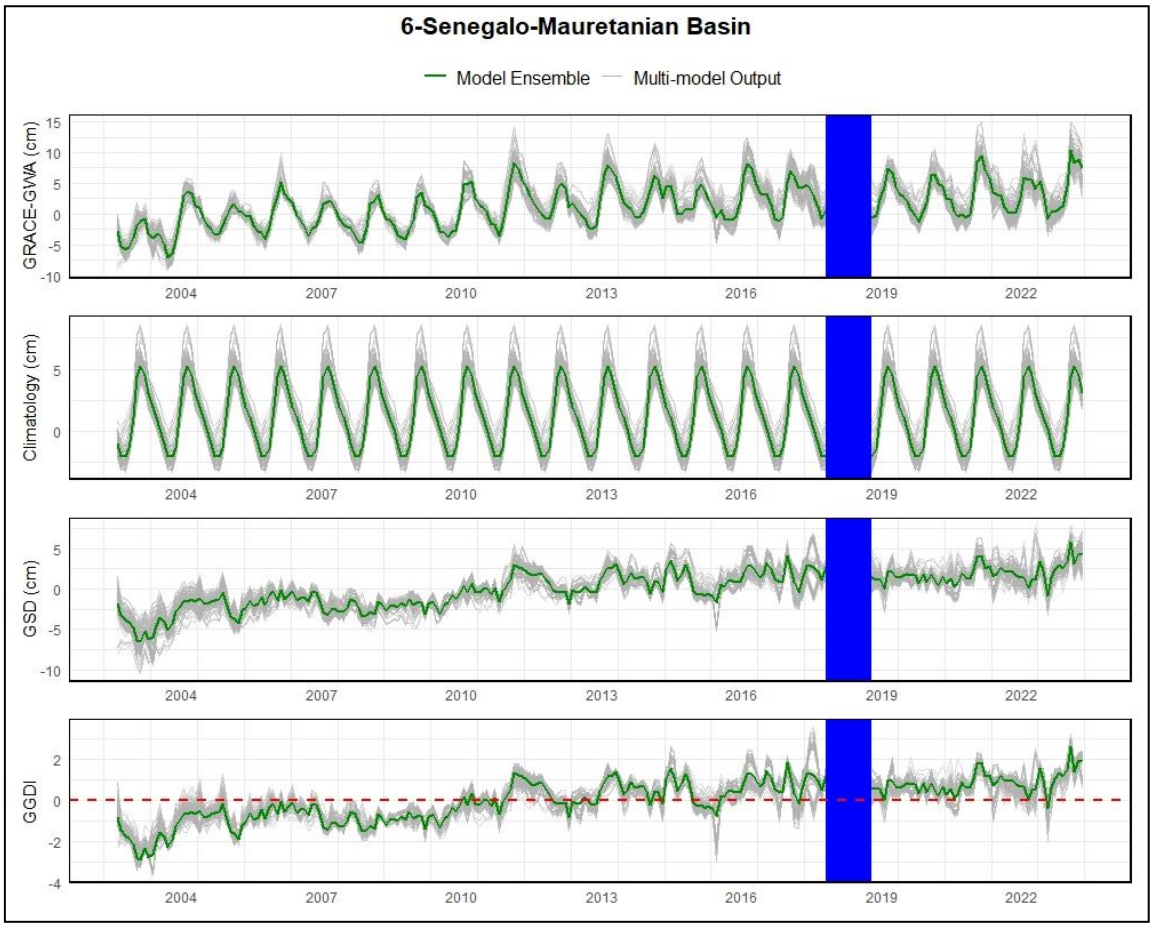


Figure C.8 Same as figure **C.3** for Basin 7 (Senegalo-Mauretanian Basin).

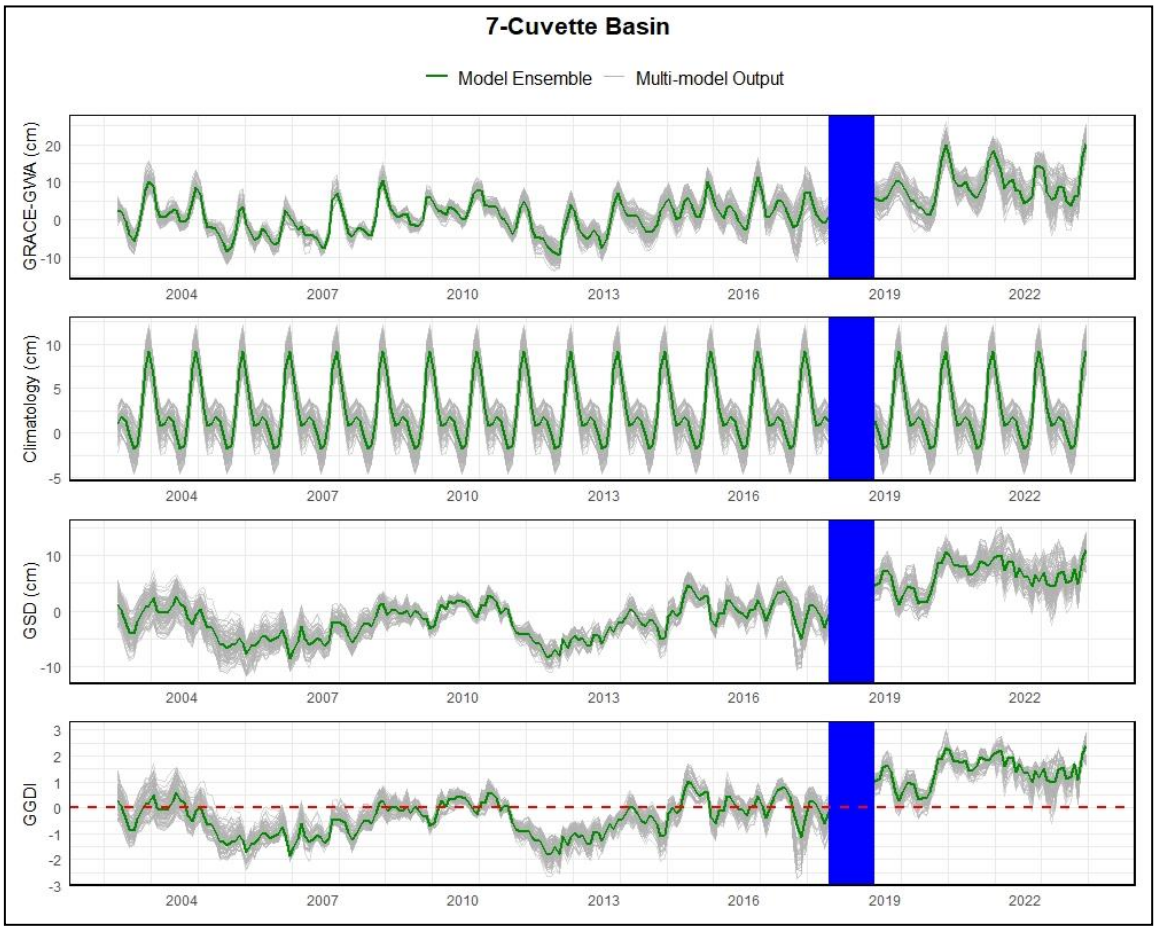


Figure C.8 Same as figure **C.3** for Basin 7 (Cuvette Basin).

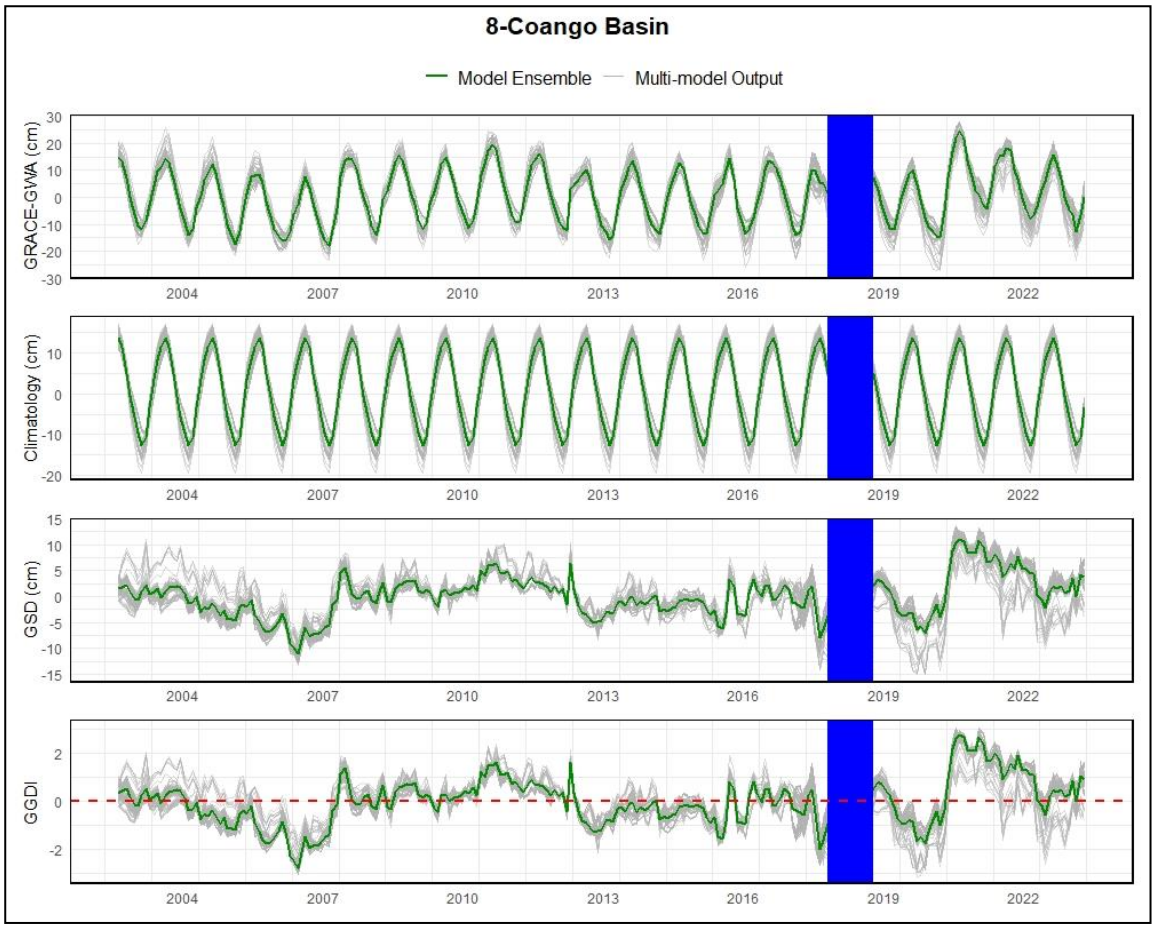


Figure C.9 Same as figure **C.3** for Basin 8 (Coango Basin).

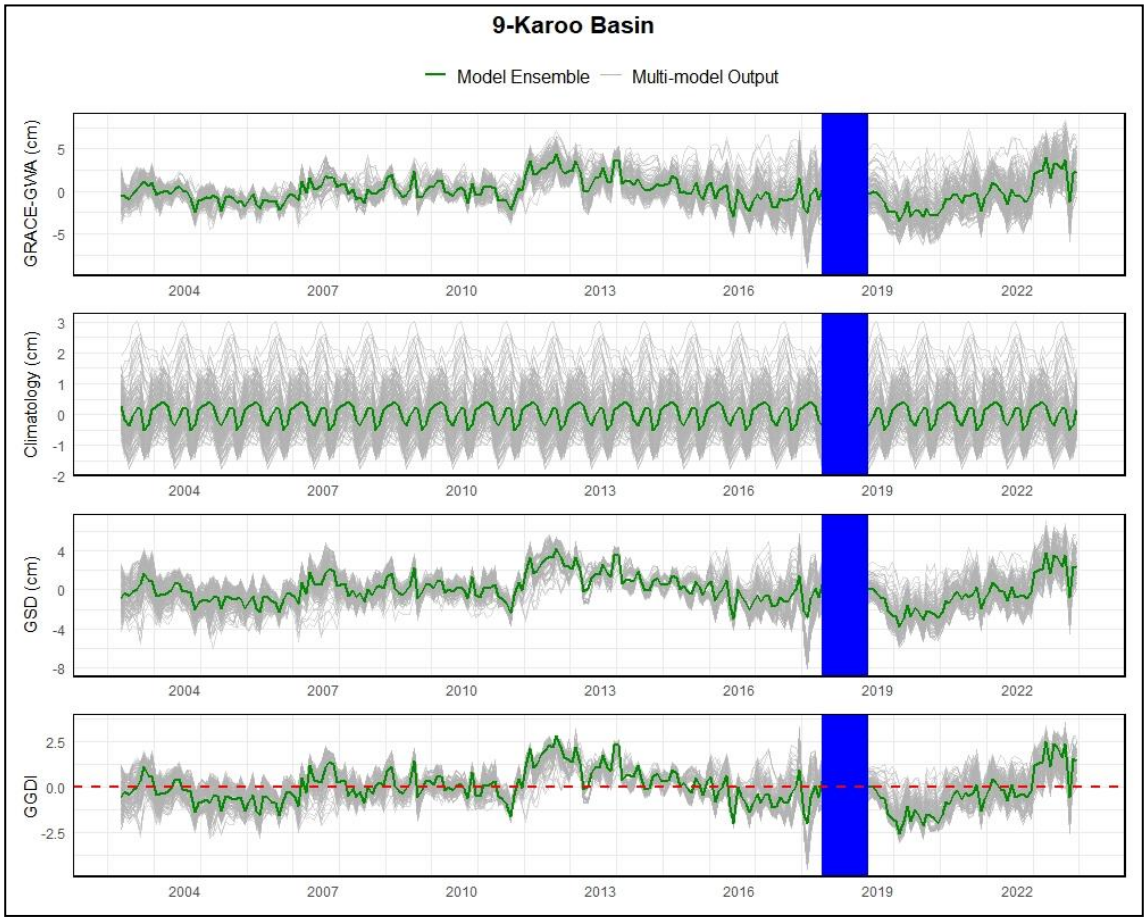


Figure C.10 Same as figure **C.3** for Basin 9 (Karoo Basin).

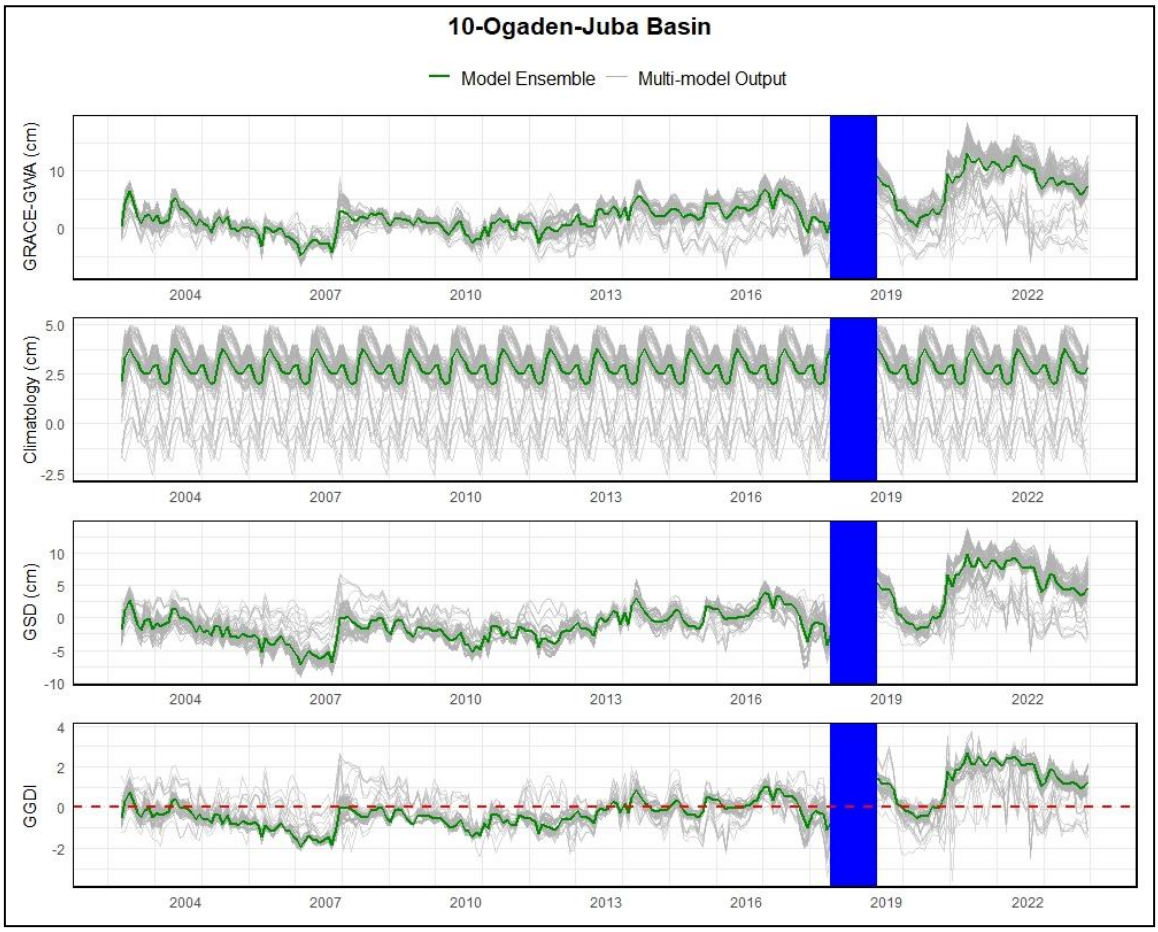


Figure C.11 Same as figure **C.3** for Basin 10 (Ogaden-Juba Basin).

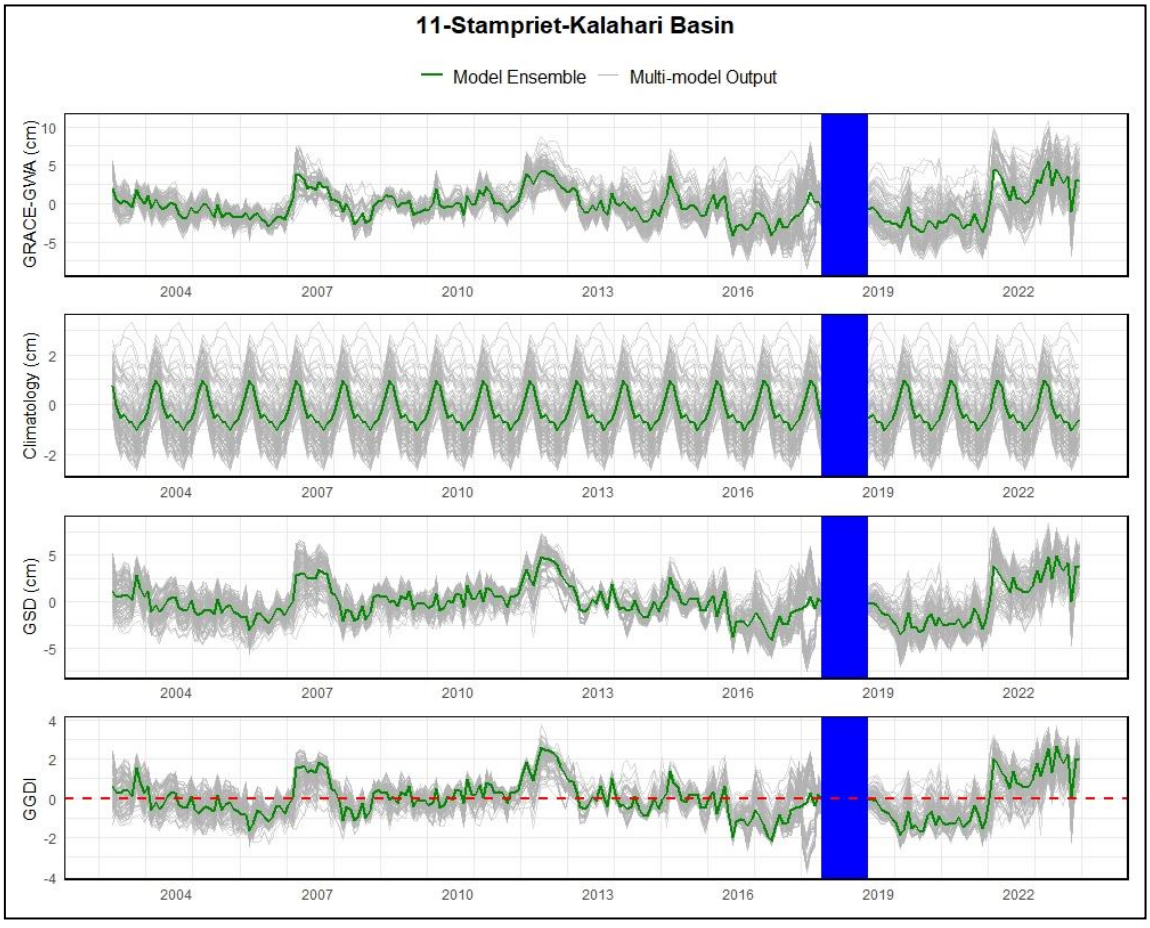


Figure C.12 Same as figure **C.3** for Basin 11 (Stampriet-Kalahari Basin).

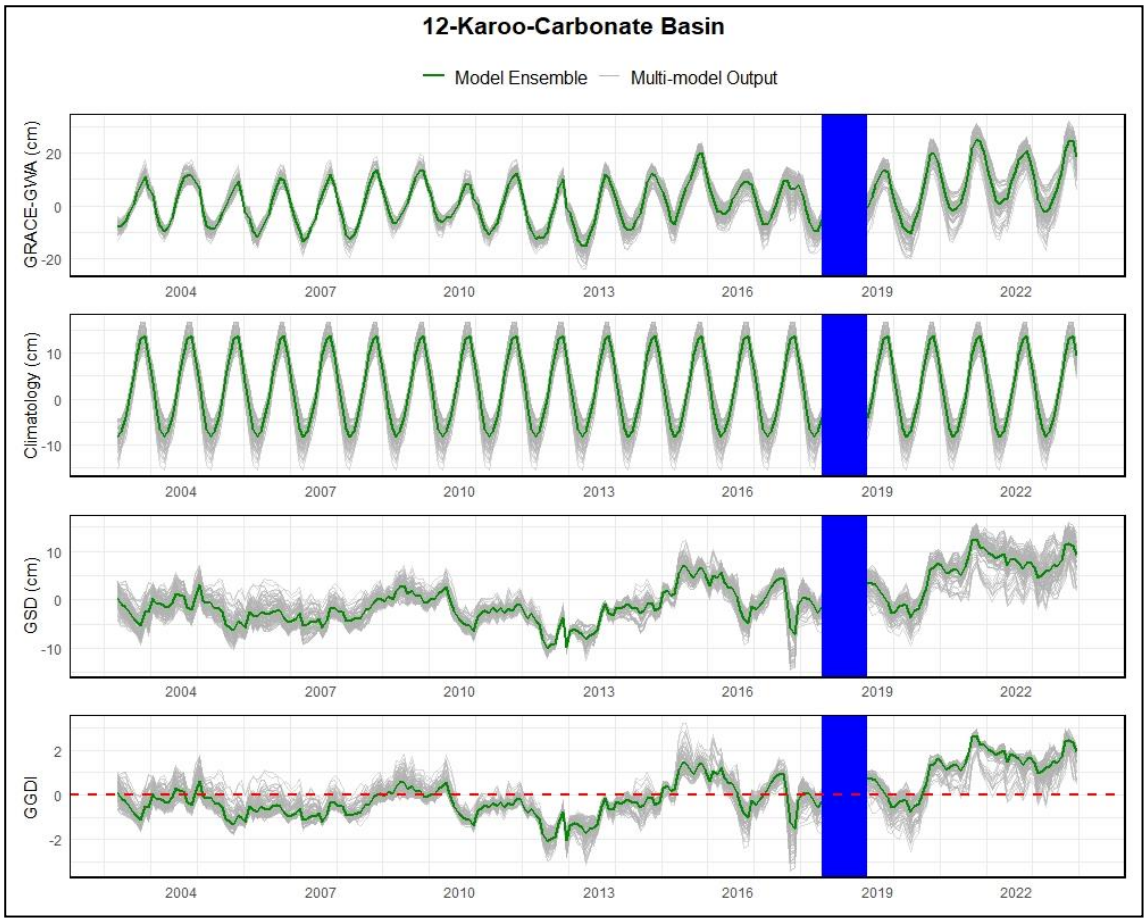


Figure C.13 Same as figure **C.3** for Basin 12 (Karoo-Carbonate Basin).

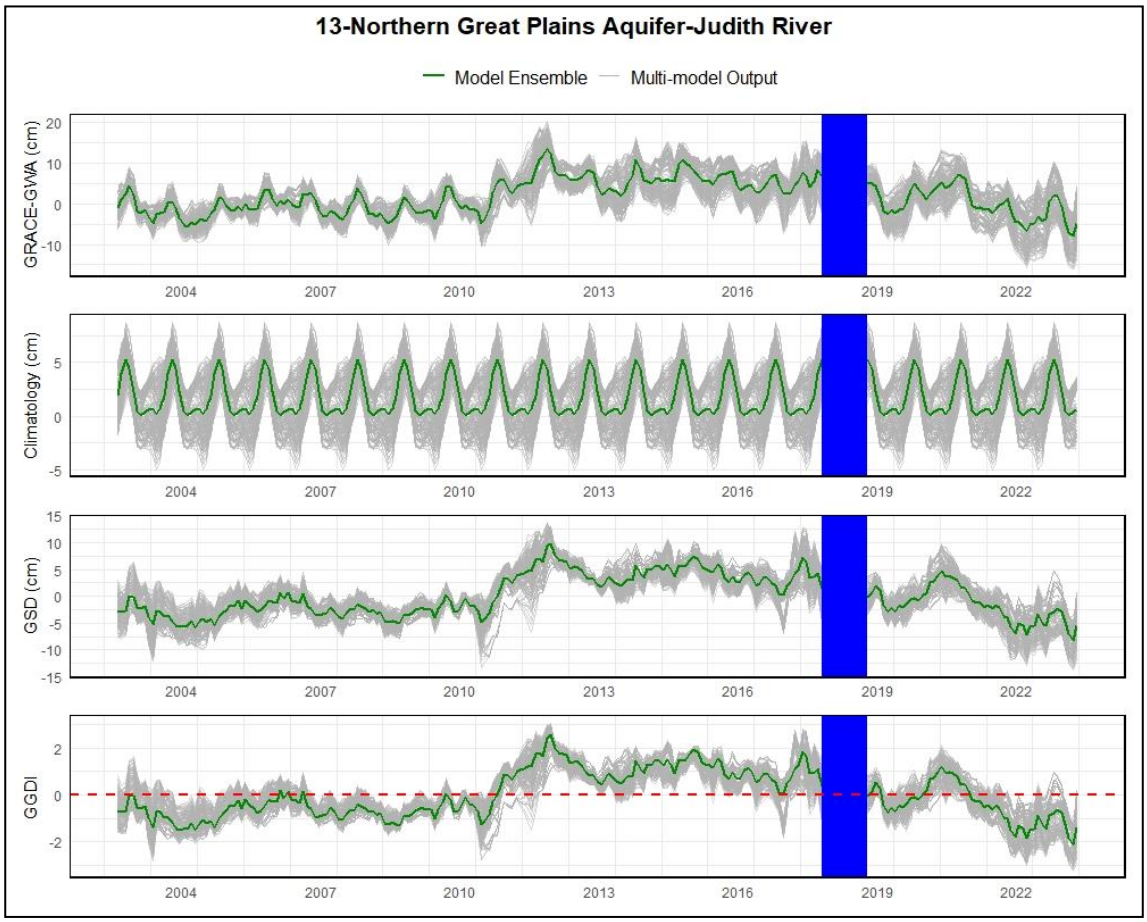


Figure C.14 Same as figure **C.3** for Basin 13 (Northern Great Plains Aquifer-Judith River).

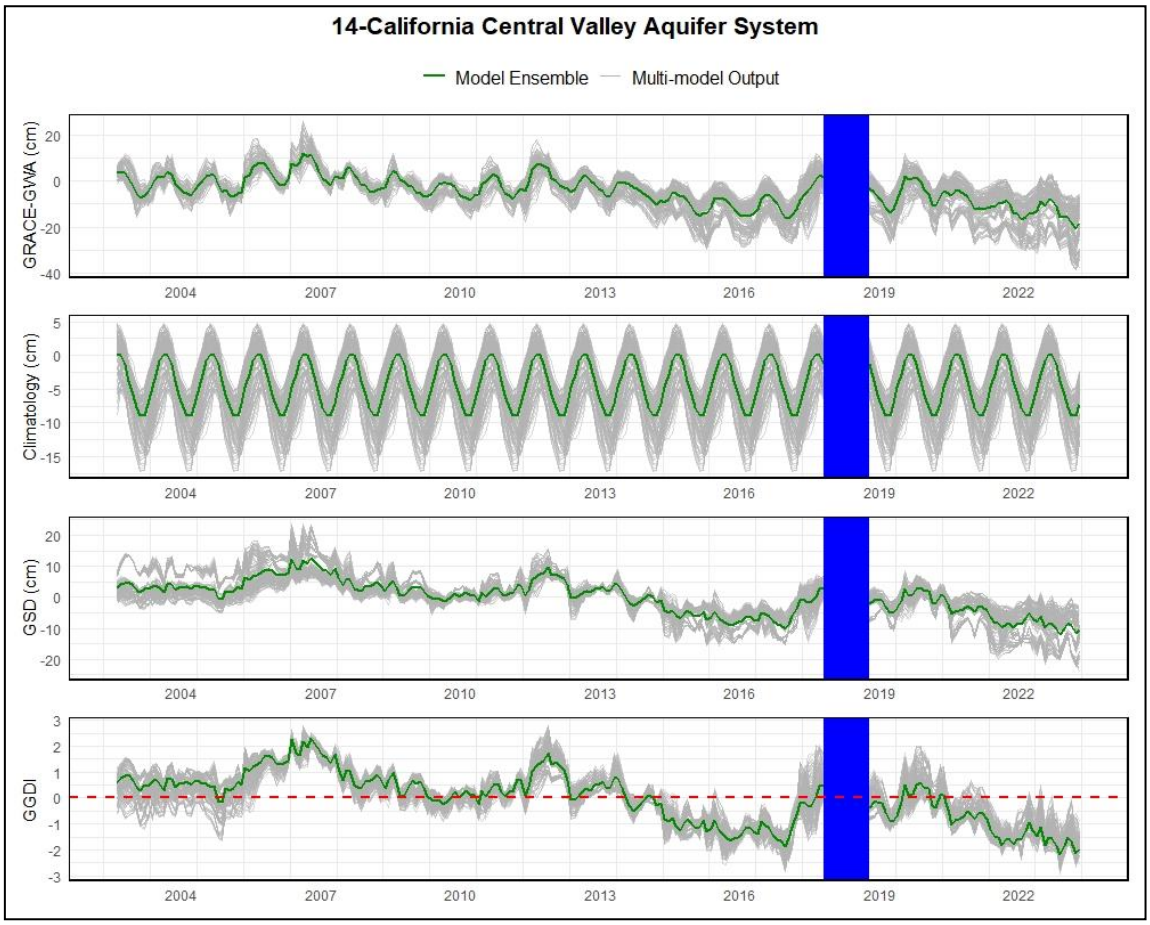


Figure C.15 Same as figure **C.3** for Basin 14 (California Central Valley Aquifer System).

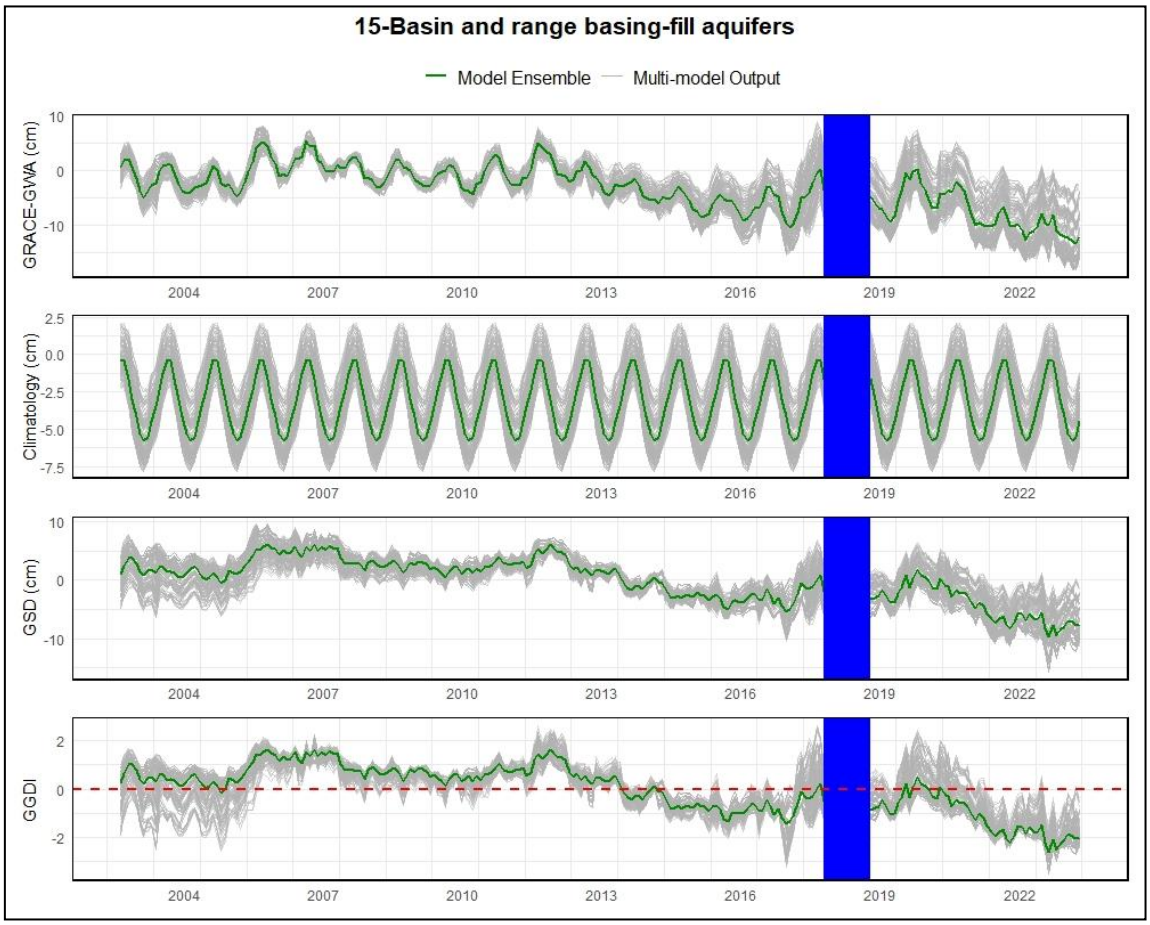


Figure C.16 Same as figure **C.3** for Basin 15 (Basin and range basing-fill aquifers).

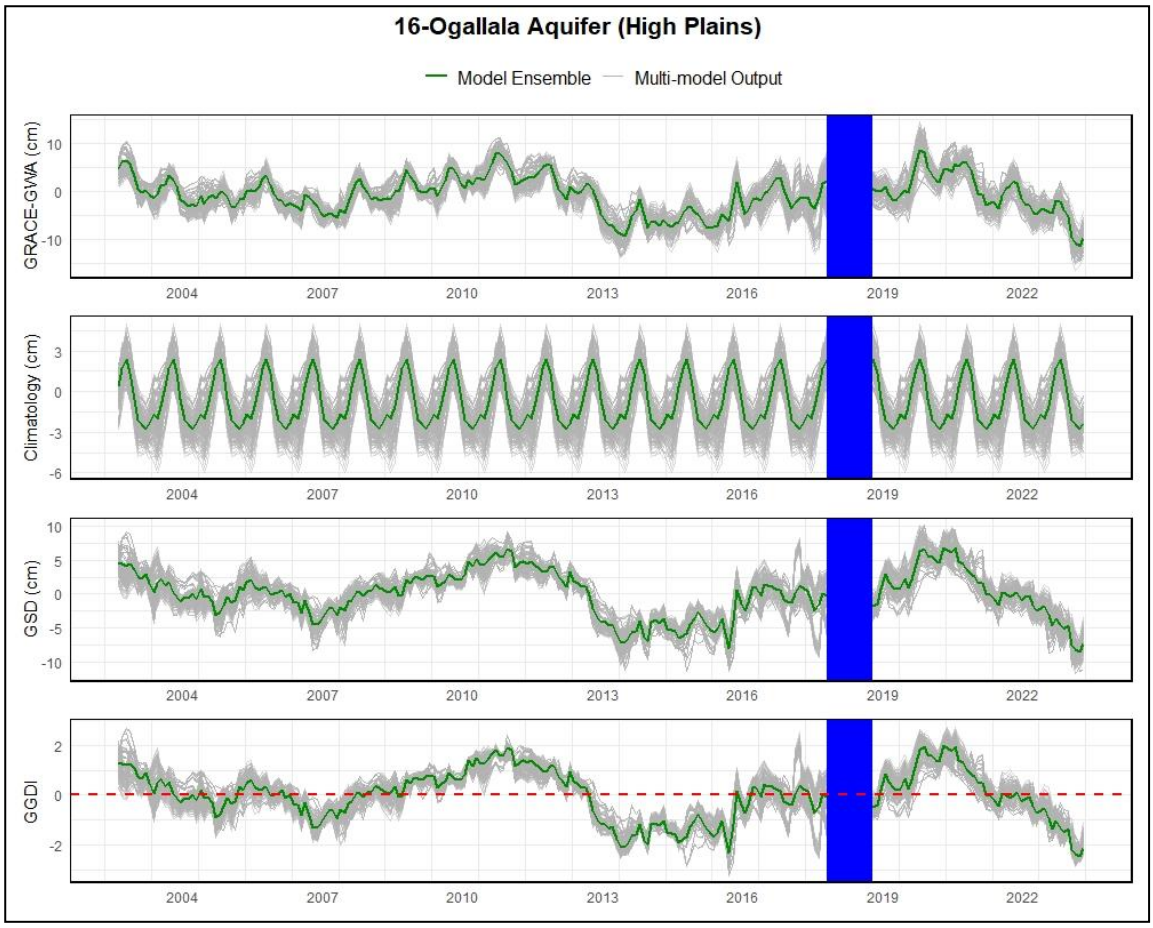


Figure C.17 Same as figure **C.3** for Basin 16 (Ogallala Aquifer [High Plains]).

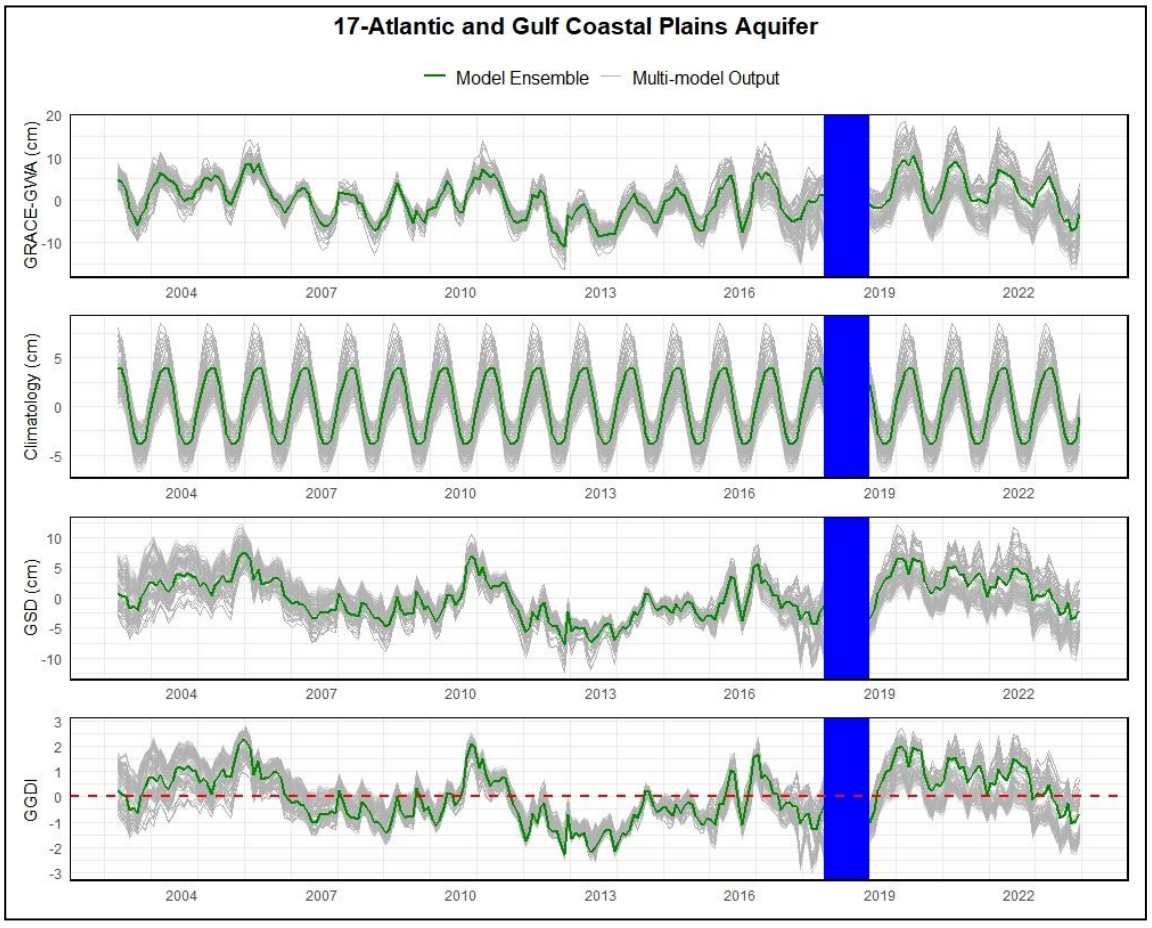


Figure C.18 Same as figure **C.3** for Basin 17 (Atlantic and Gulf Coastal Plains Aquifer).

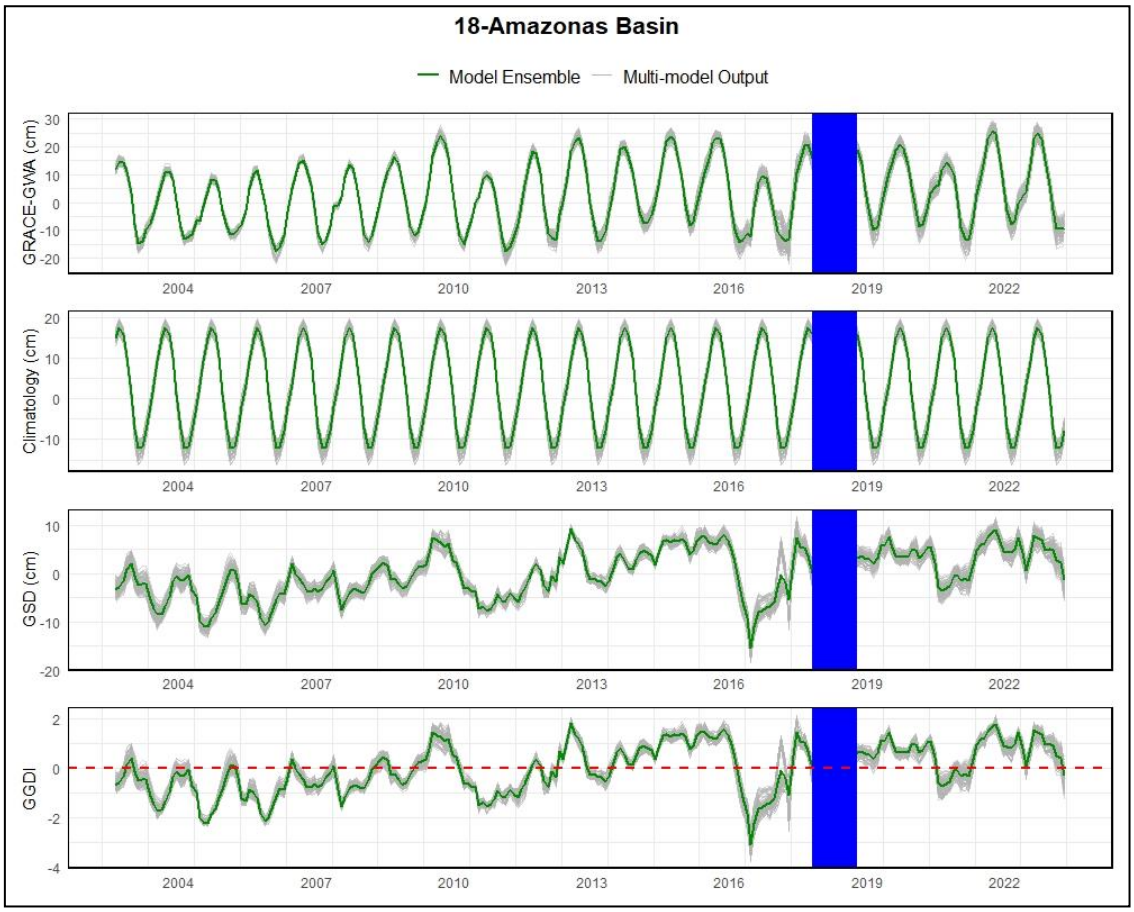


Figure C.19 Same as figure **C.3** for Basin 18 (Amazonas Basin).

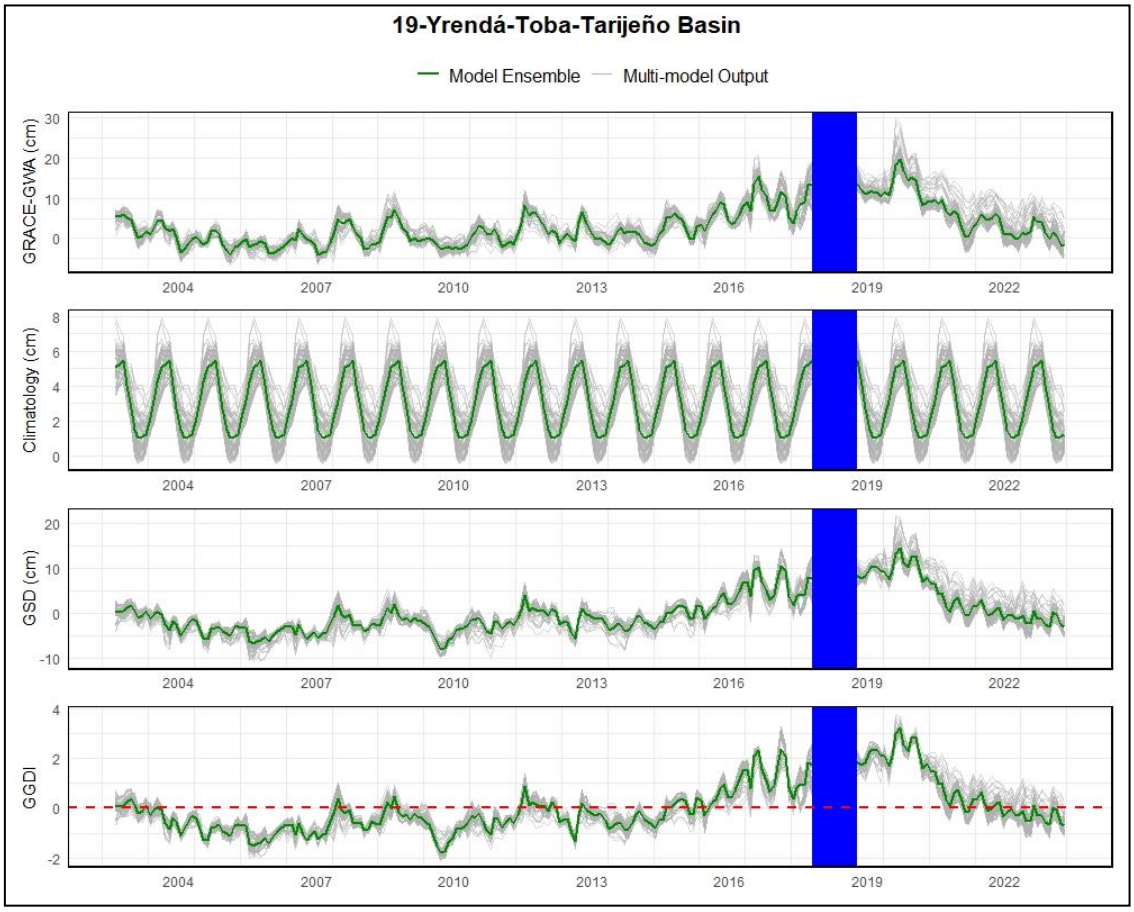


Figure C.20 Same as figure **C.3** for Basin 19 (Yrendá-Toba-Tarijeño Basin).

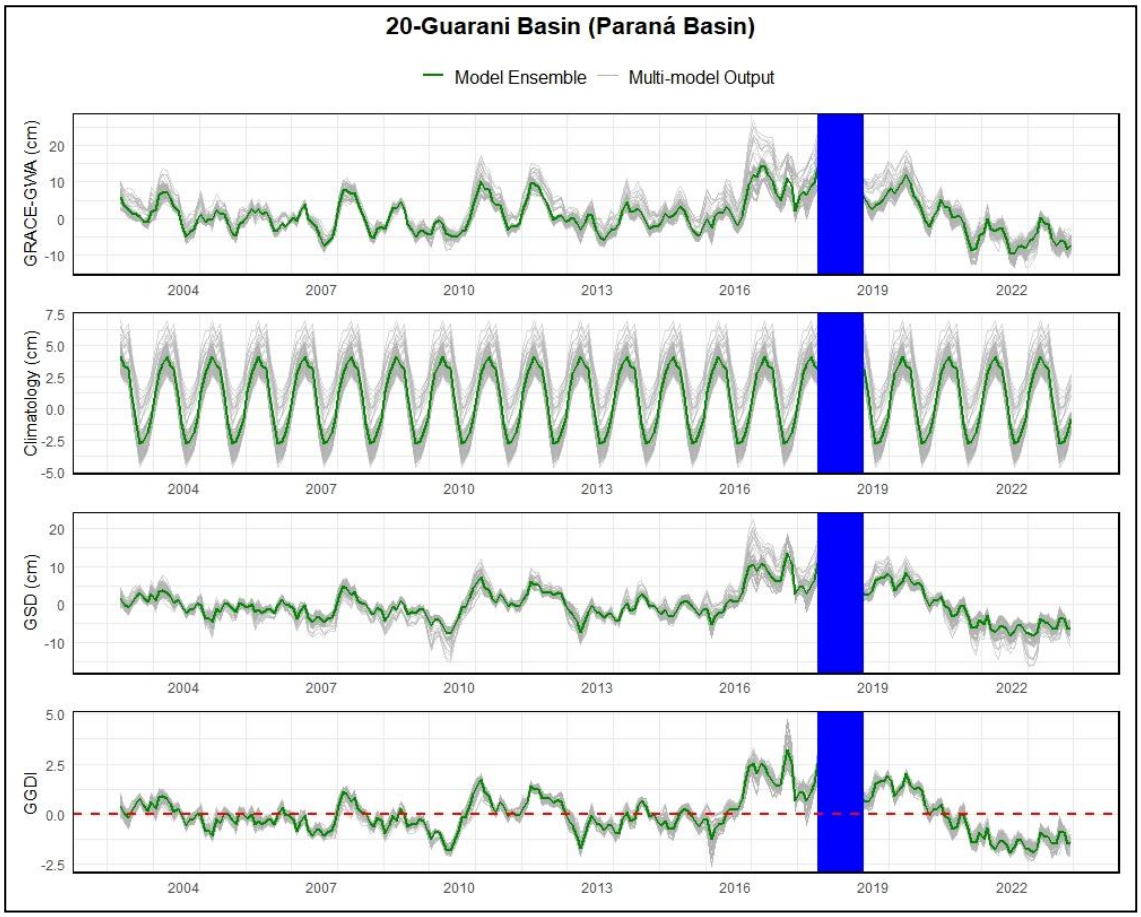


Figure C.21 Same as figure **C.3** for Basin 20 (Guarani Basin [Paraná Basin]).

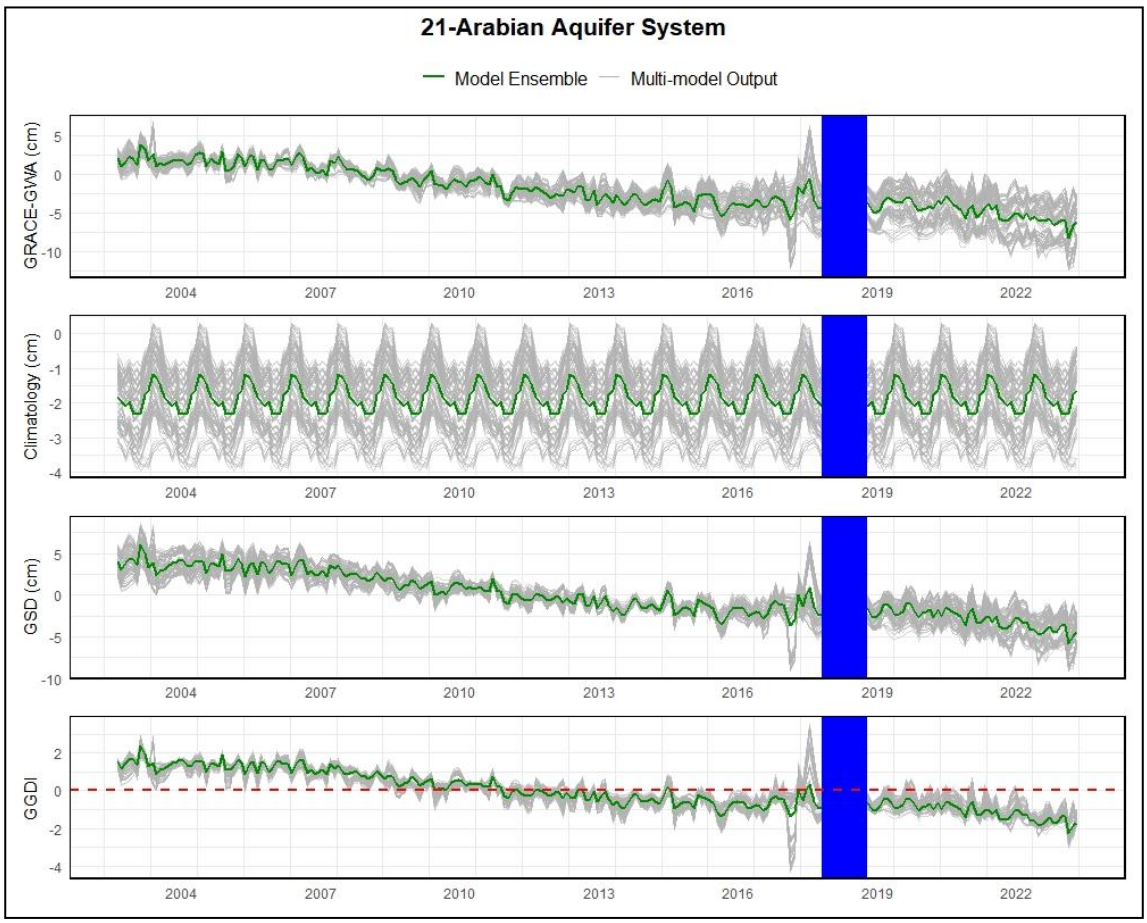


Figure C.22 Same as figure **C.3** for Basin 21 (Arabian Aquifer System).

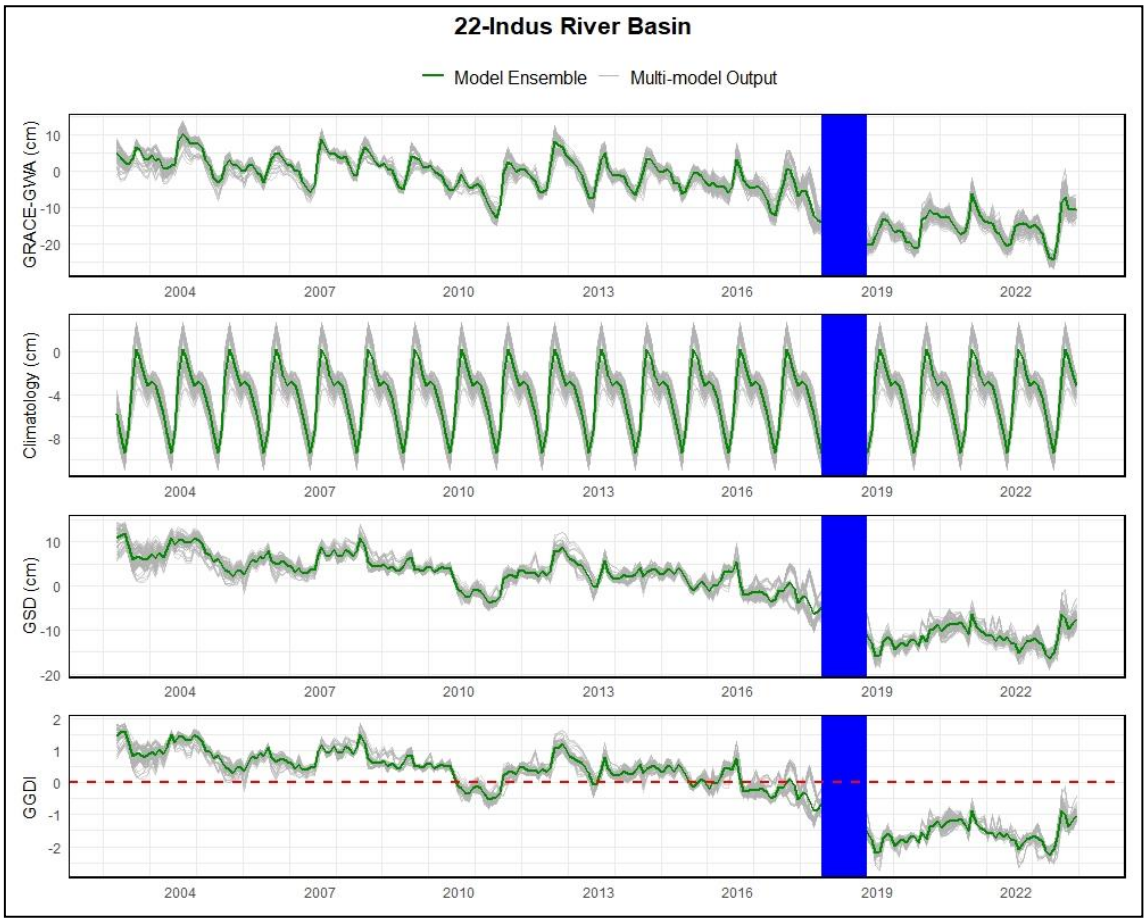


Figure C.23 Same as figure **C.3** for Basin 22 (Indus River Basin).

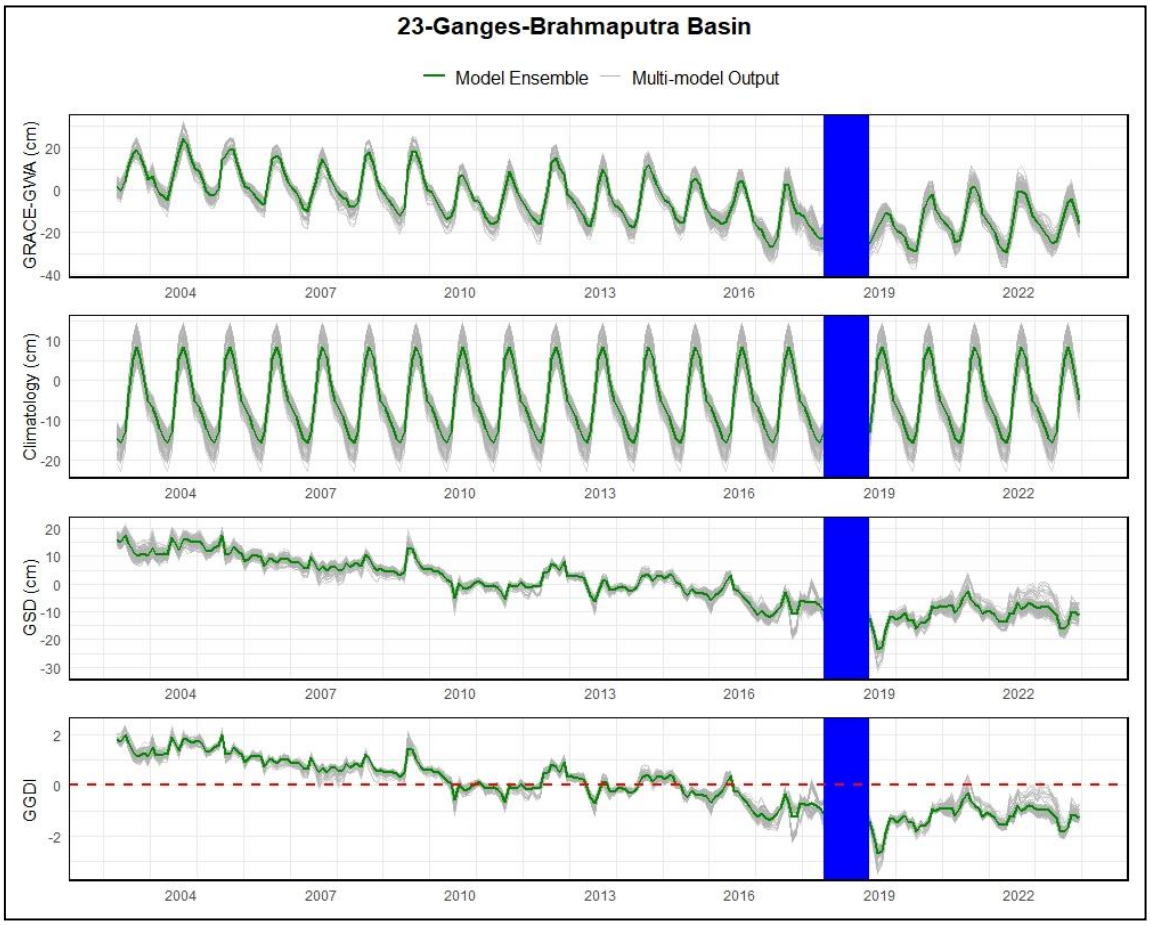


Figure C.24 Same as figure **C.3** for Basin 23 (Ganges-Brahmaputra Basin).

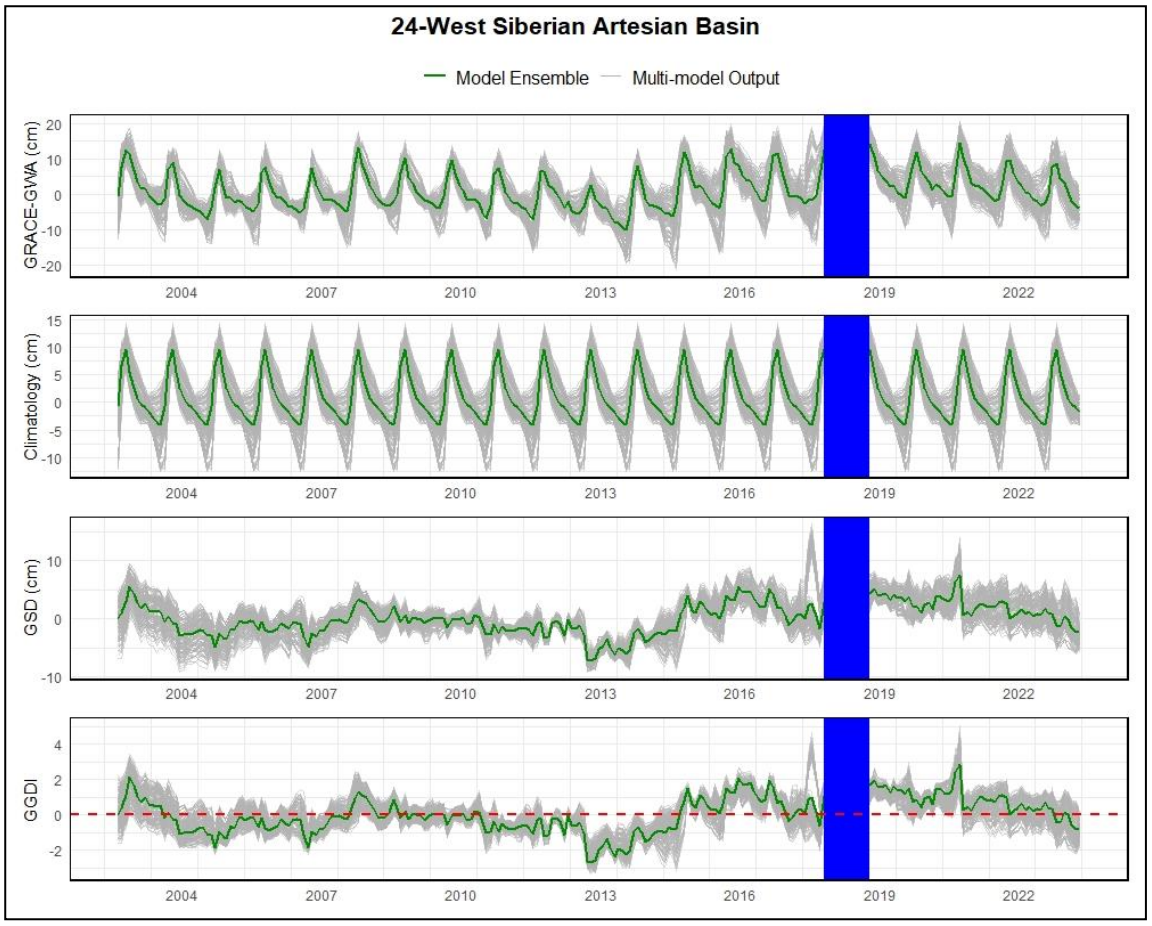


Figure C.25 Same as figure **C.3** for Basin 24 (West Siberian Artesian Basin).

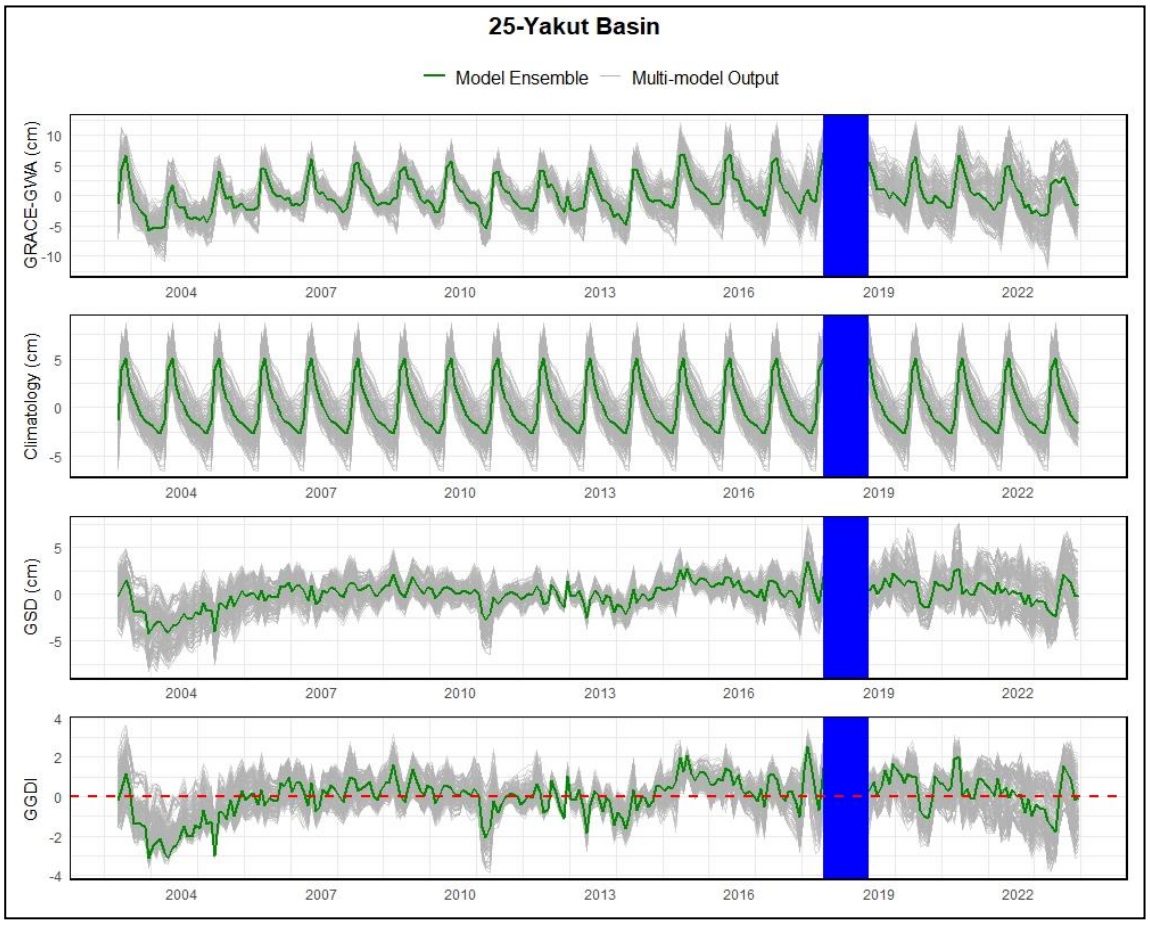


Figure C.26 Same as figure **C.3** for Basin 25 (Yakut Basin).

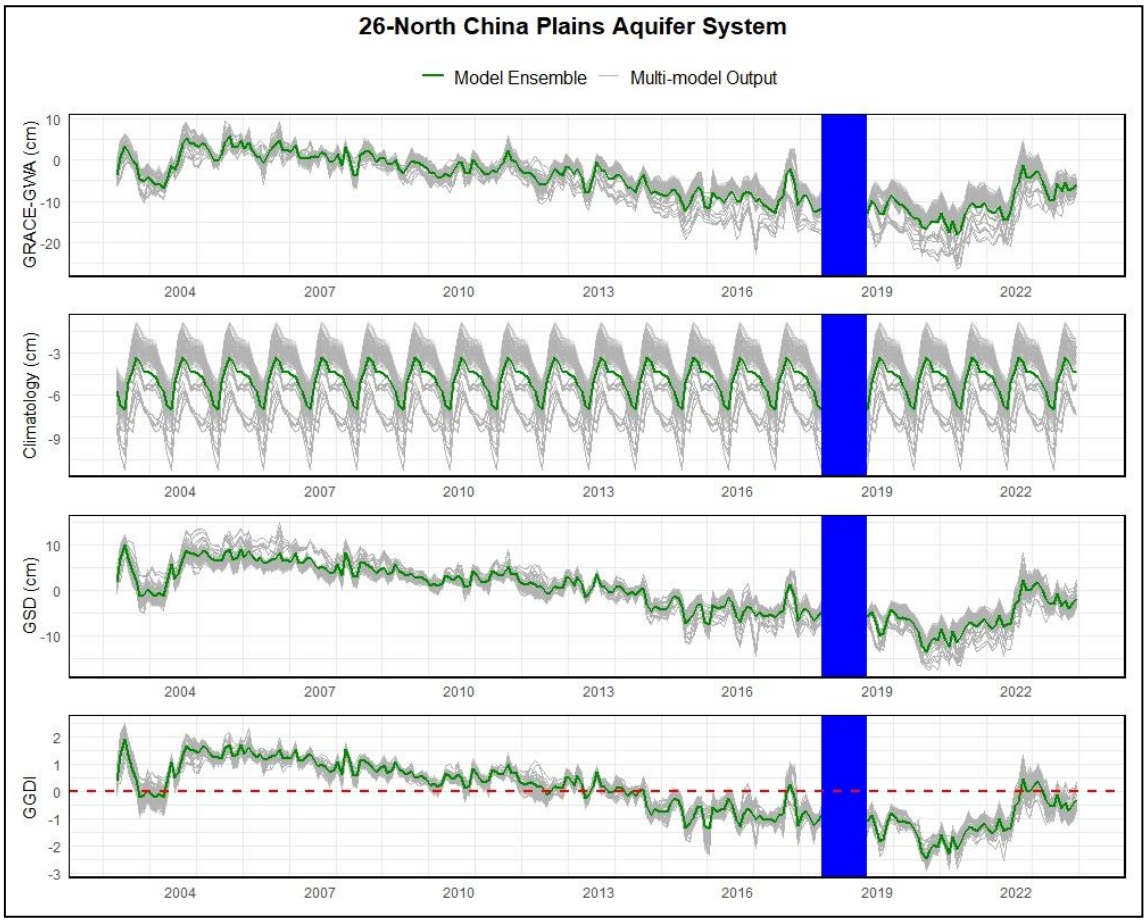


Figure C.27 Same as figure **C.3** for Basin 26 (North China Plains Aquifer System).

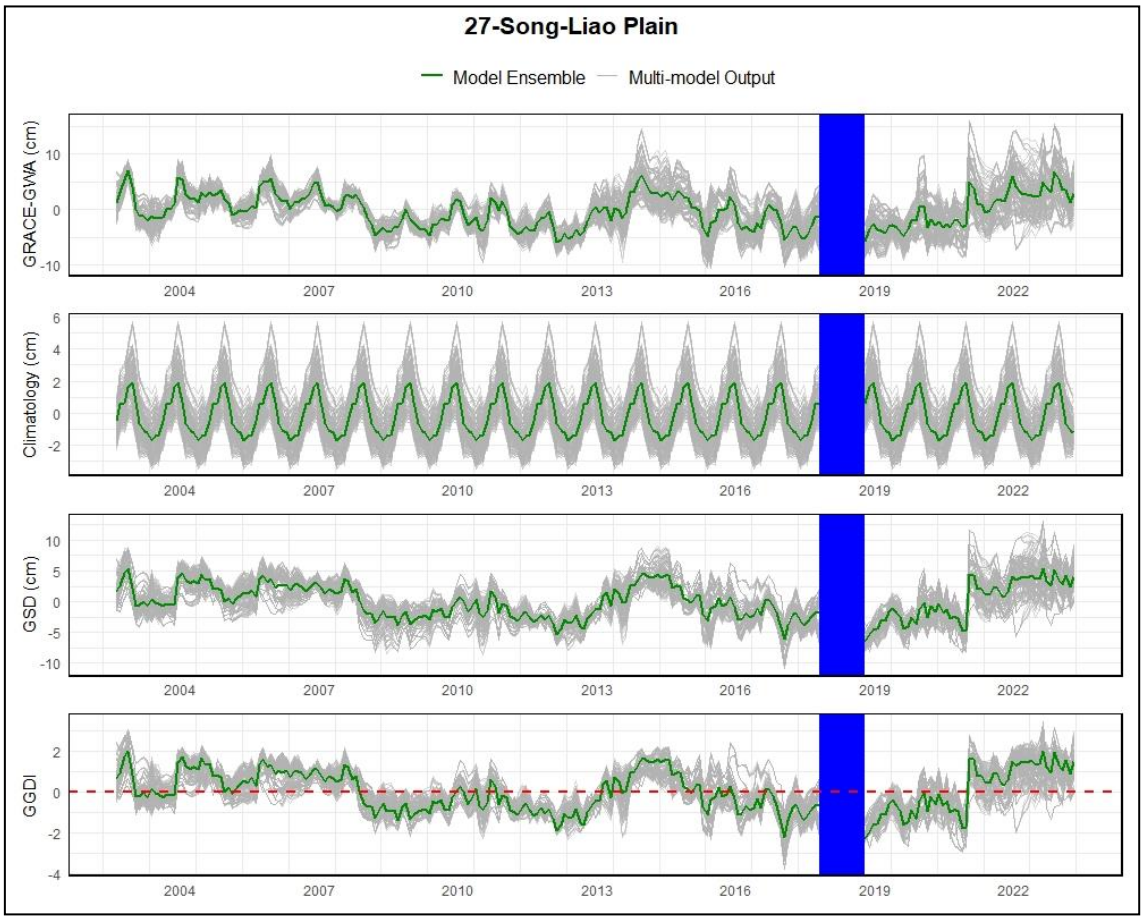


Figure C.28 Same as figure **C.3** for Basin 27 (Song-Liao Plain).

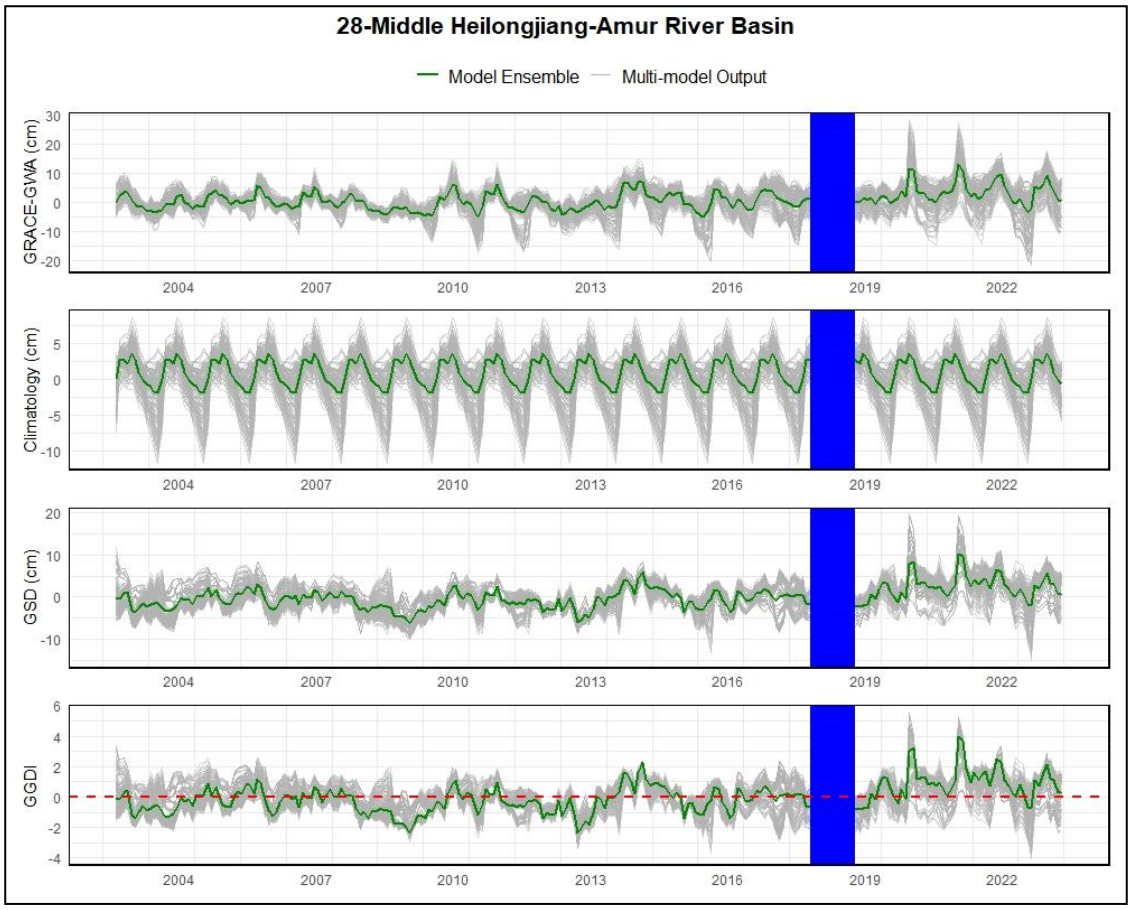


Figure C.29 Same as figure **C.3** for Basin 28 (Middle Heilongjiang-Amur River Basin).

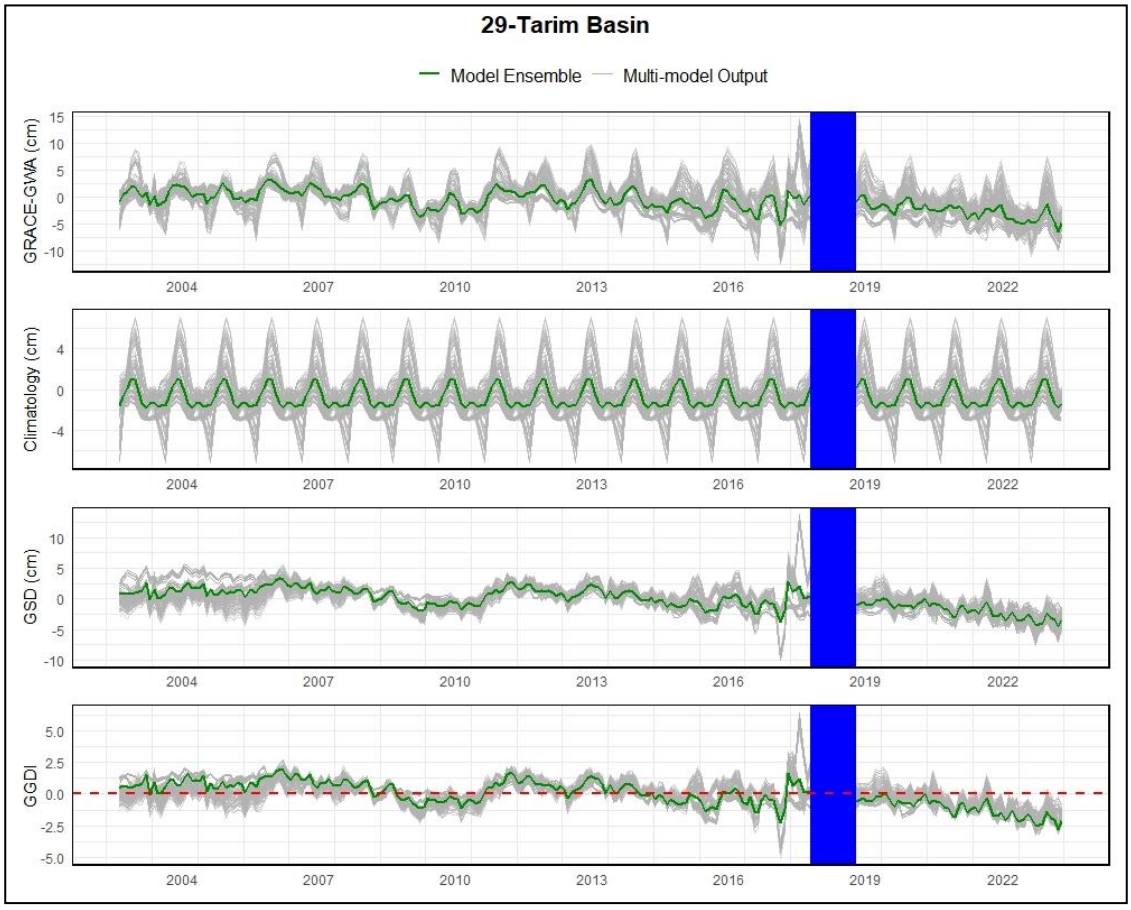


Figure C.30 Same as figure **C.3** for Basin 29 (Tarim Basin).

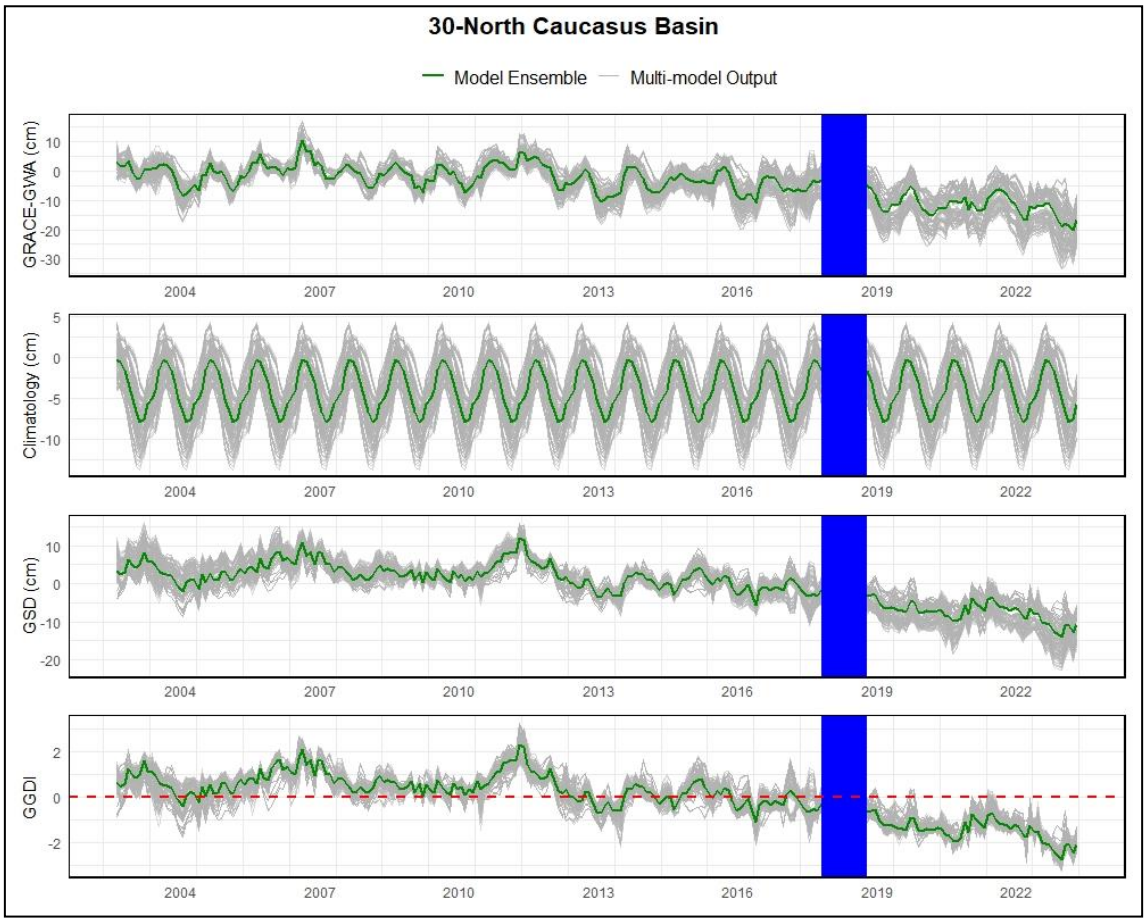


Figure C.31 Same as figure **C.3** for Basin 30 (North Caucasus Basin).

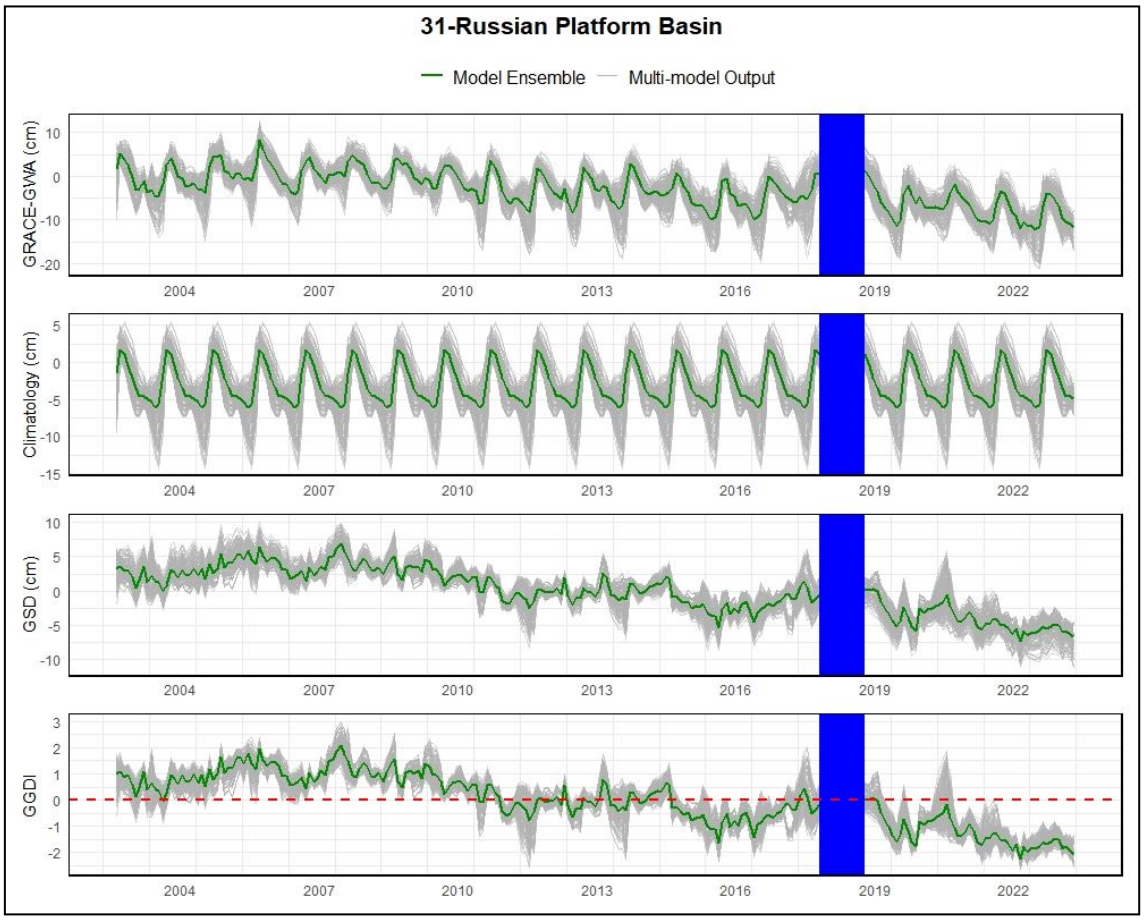


Figure C.32 Same as figure **C.3** for Basin 31 (Russian Platform Basin).

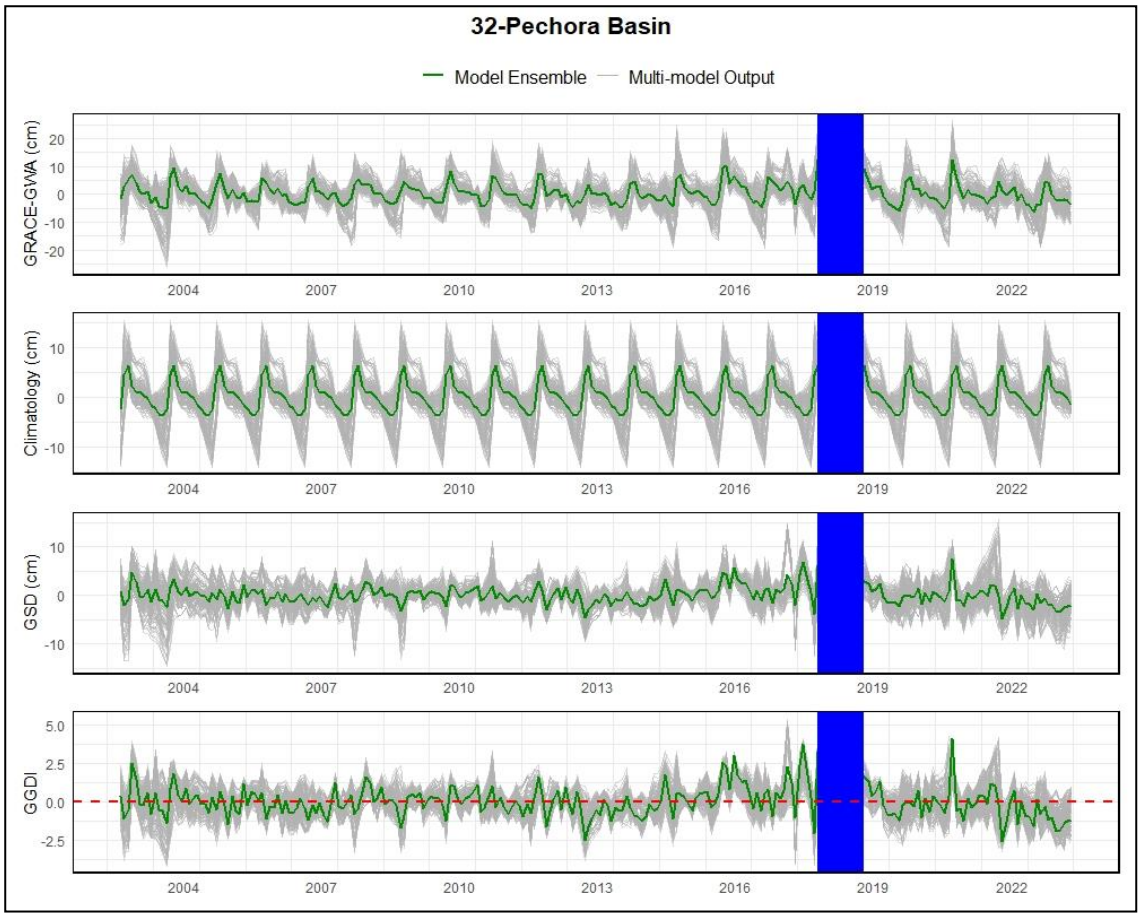


Figure C.33 Same as figure **C.3** for Basin 32 (Pechora Basin).

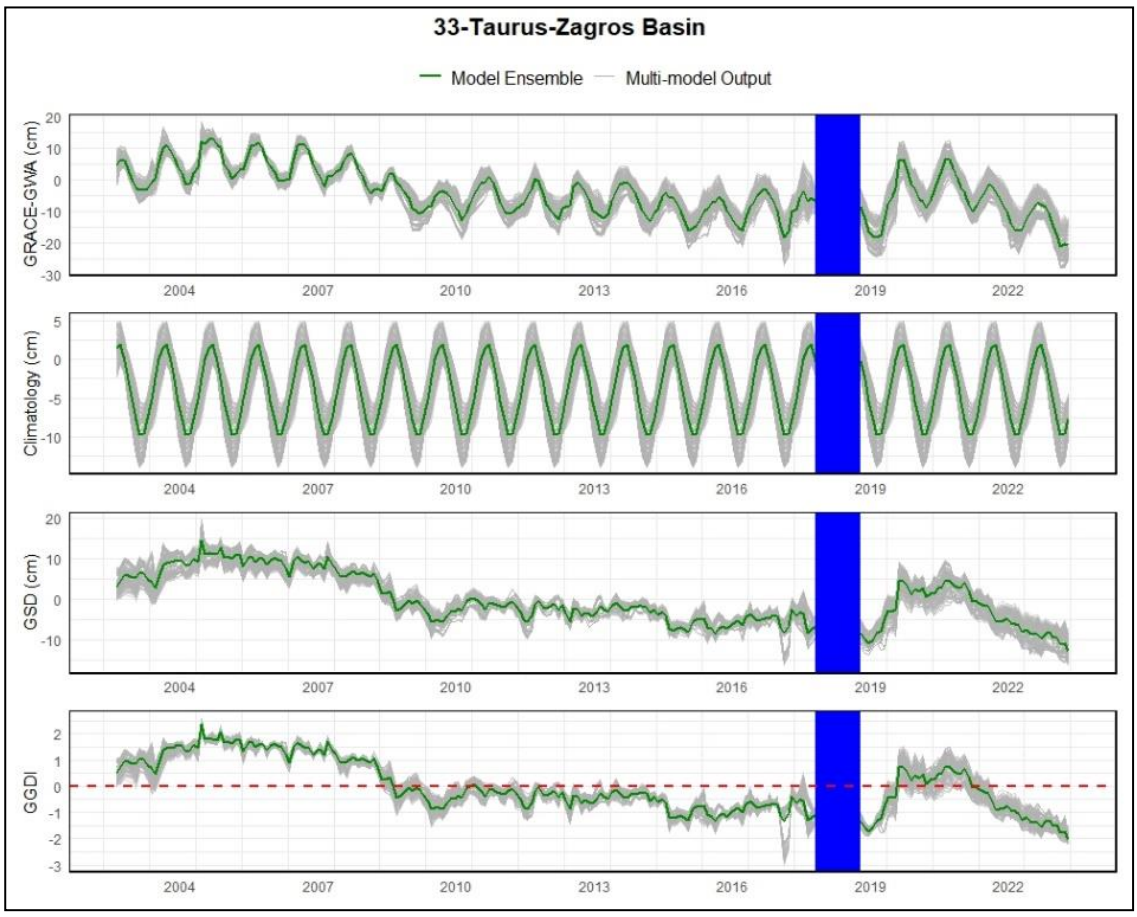


Figure C.34 Same as figure **C.3** for Basin 33 (Taurus-Zagros Basin).

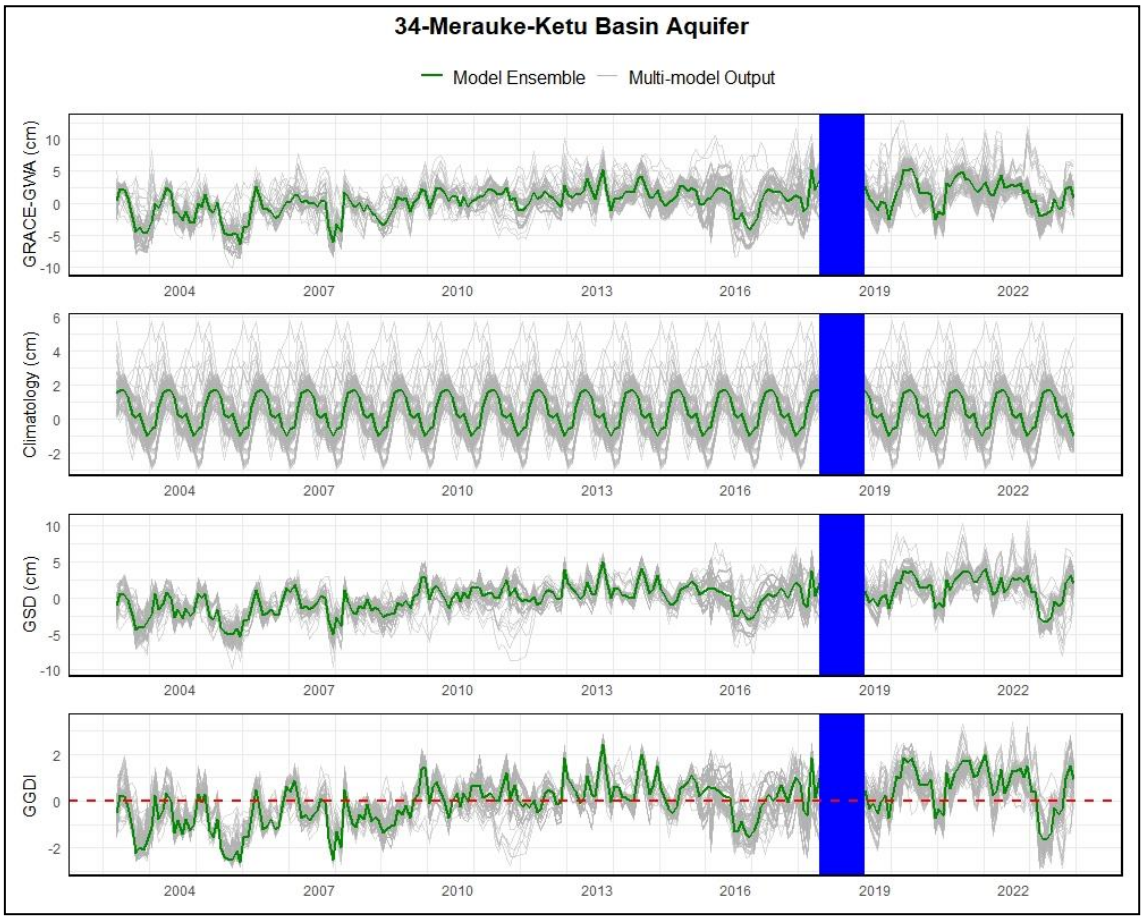


Figure C.35 Same as figure **C.3** for Basin 34 (Merauke-Ketu Basin Aquifer).

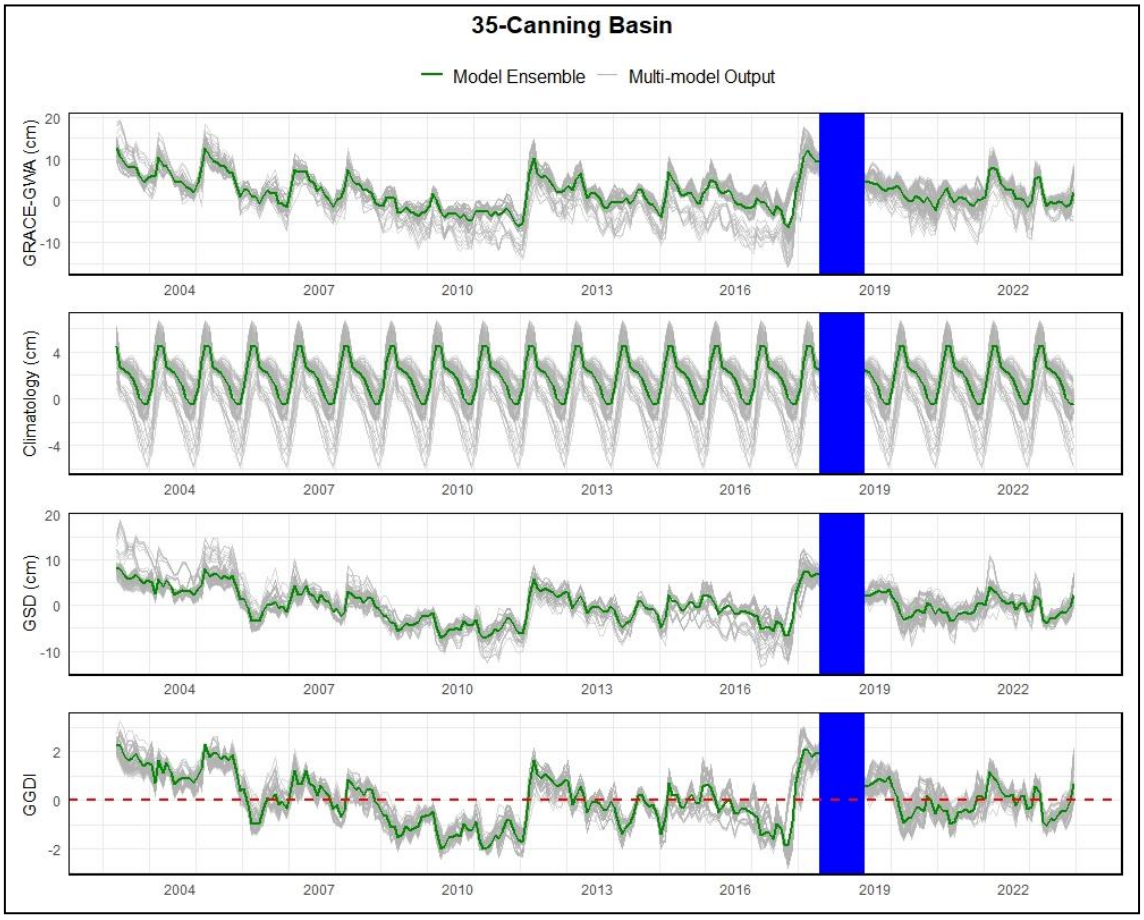


Figure C.36 Same as figure **C.3** for Basin 35 (Canning Basin).

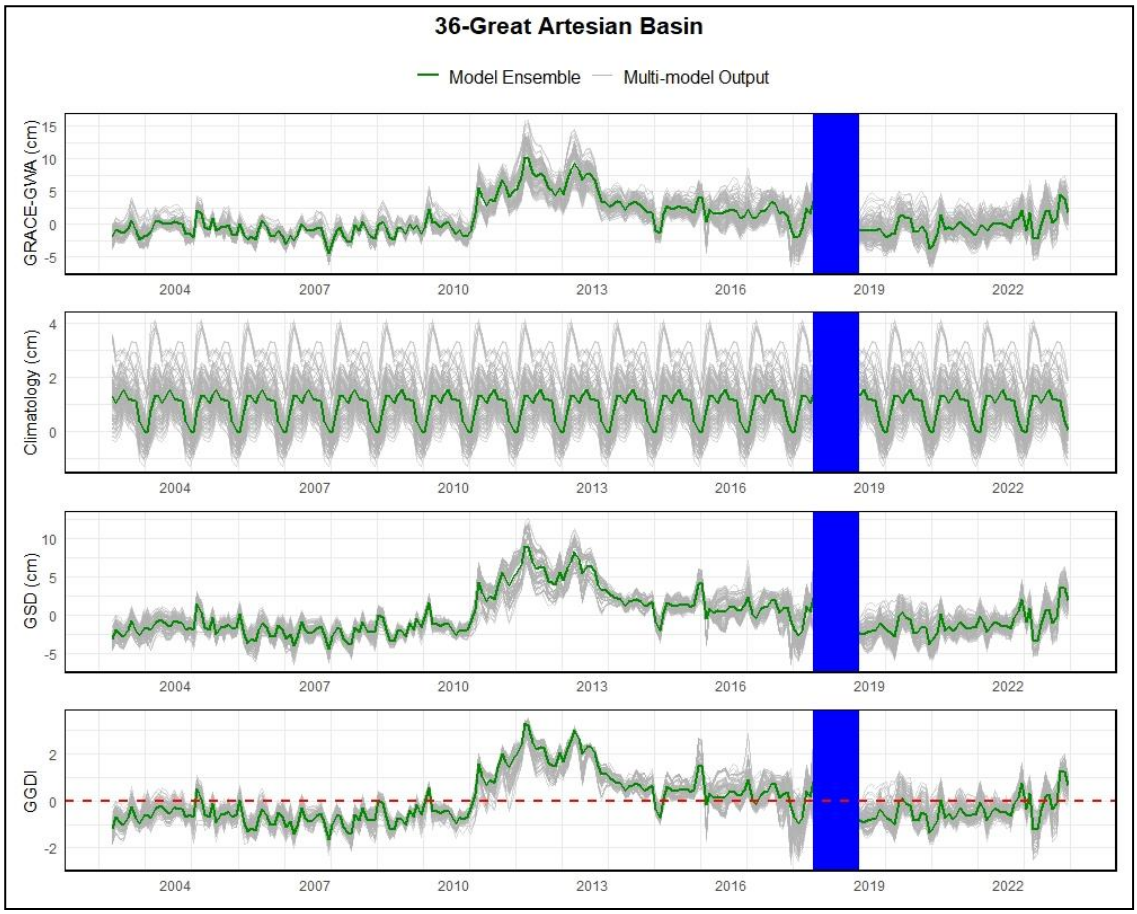


Figure C.37 Same as figure **C.3** for Basin 36 (Great Artesian Basin).

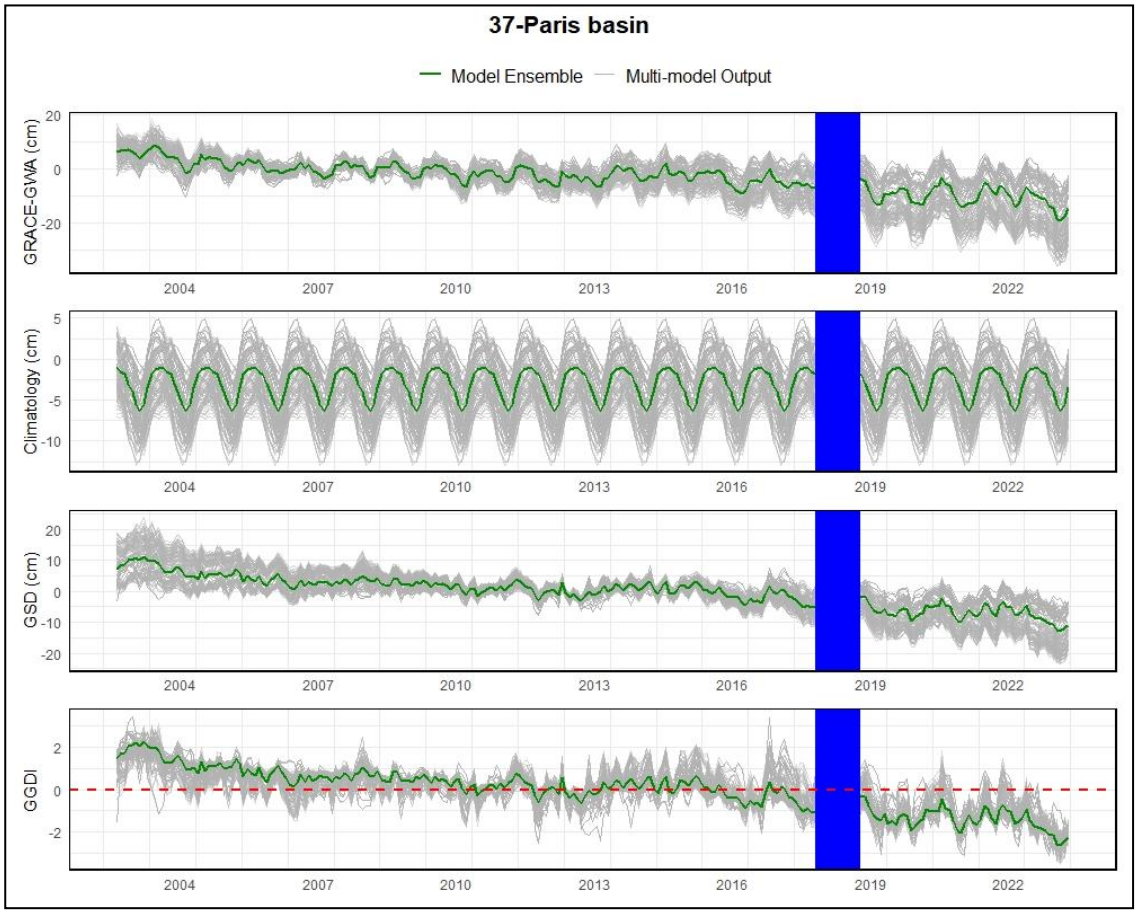


Figure C.38 Same as figure **C.3** for Basin 37 (Paris basin).

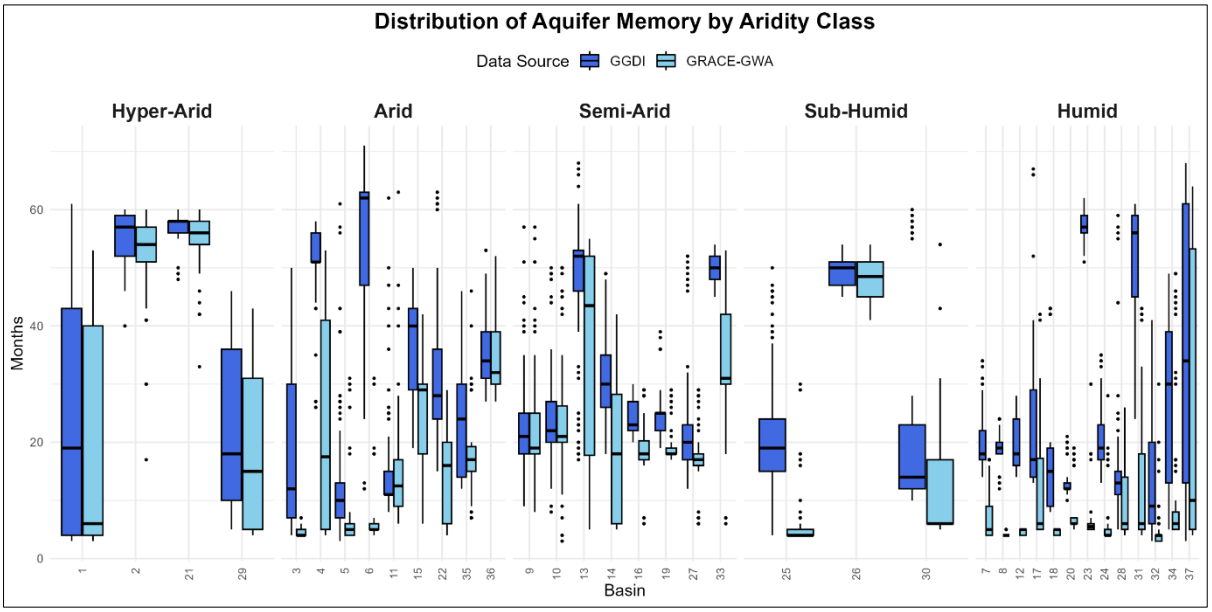


Figure C.39 A comparison of aquifer memory (in months), calculated from GGDI and GRACE-GWA, across 37 study aquifers, each classified by aridity level (hyper-arid, arid, semi-arid, sub-humid, and humid).

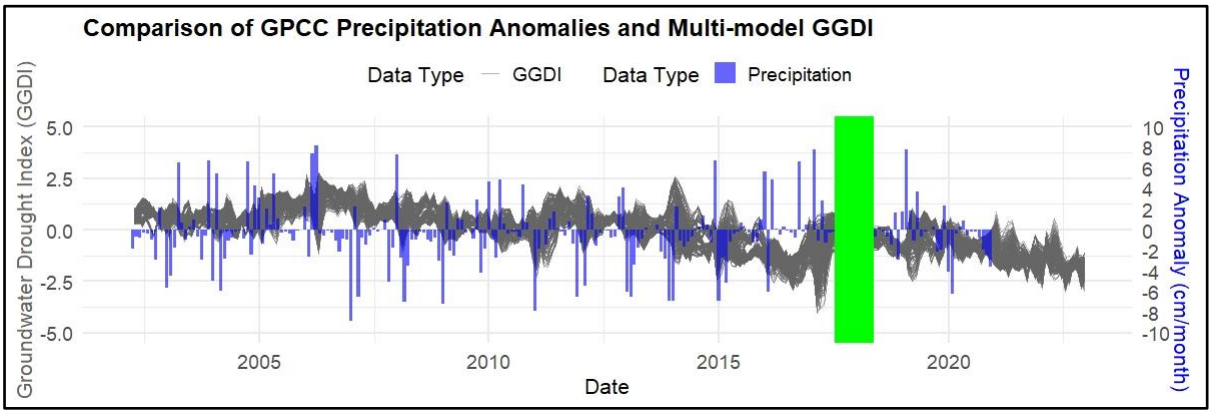


Figure C.40 Comparison of multi-model GRACE-Groundwater Drought Index (GGDI) estimates (grey lines; left axis) and precipitation anomalies from the Global Precipitation Climatology Centre (GPCC) (blue bars; right axis) across the Central Valley aquifer system (Basin 14 in Figure 1). Precipitation anomalies are expressed in cm/month relative to the long-term mean (1901–2020) and are shown from April 2002 to December 2020. GGDI time series extend from April 2002 to December 2022. The green-shaded region indicates the observational gap between the GRACE and GRACE-FO missions, during which GGDI data are unavailable. This figure illustrates the relationship between meteorological variability and groundwater drought across a critical agricultural basin. Multi-model GGDI estimates document significant variability complicating the comparison with precipitation datasets.

Tables:

Basin	SMA	SWEA	SWA	Basin	SMA	SWEA	SWA
1	√	x	√	20	√	x	√
2	√	x	x	21	√	x	√
3	√	x	√	22	√	x	√
4	√	x	x	23	√	x	√
5	√	x	√	24	√	√	√
6	√	x	√	25	√	√	√
7	√	x	x	26	√	√	√
8	√	x	x	27	√	√	√
9	√	x	√	28	√	√	√
10	√	x	x	29	√	√	√
11	√	x	x	30	√	√	√
12	√	x	x	31	√	√	√
13	√	√	√	32	√	√	x
14	√	√	√	33	√	√	√
15	√	√	√	34	√	x	x
16	√	√	√	35	√	x	x
17	√	√	√	36	√	x	√
18	√	x	√	37	√	√	x
19	√	x	√				

Table C.1 Overview of water budget components, including soil moisture anomalies (SMA), snow water equivalent anomalies (SWEA) and surface water anomalies (SWA) across major aquifers.

Basin	Number of drought events					
	Min	Q1	Median	Q3	Max	Ens
1	1	2	5	6	10	6
2	2	2	2	3	5	2
3	2	3	4	5	8	3
4	1	3	3	4	6	4
5	3	4	6	7	11	8
6	1	3	3.5	4	8	3
7	3	5	6	7	9	8
8	4	5	5	6	9	4
9	1	6	7	7.25	10	7
10	3	5	6	7	10	5
11	2	5.75	6	7	10	7
12	3	6	7	7	9	7
13	2	3	4	5	8	3
14	2	4	5	6	8	4
15	2	3	3	4	7	3
16	3	5	5	6	9	6
17	4	5	6	7	10	6
18	5	6.75	7	8	10	7
19	1	3	4	5	7	3
20	5	6	7	7	8	6
21	2	2	2	3	6	2
22	3	3	3.5	4	5	3
23	4	4	4	5	6	5
24	2	4	5	7	11	5
25	1	5	6	7	12	4
26	2	3	4	5	6	4
27	3	5	5	6	8	4
28	4	6	7	8	12	8
29	3	4	5	8	11	4
30	3	4	5	6	9	5
31	2	4	5	5	8	5
32	3	6	7	9	12	7
33	3	3	4	4	6	3
34	2	6.75	9	9	12	9
35	4	6	7	8	9	8
36	2	4	4	5	7	3
37	2	4	6	8	12	3

Table C.2 Summary statistics of number of drought events across study basins, including the minimum (Min), first quartile (Q1), median, third quartile (Q3), maximum (Max), and the model ensemble estimate (Ens).

Basin	Maximum duration (Months)					
	Min	Q1	Median	Q3	Max	Ens
1	30	53	79	132	152	70
2	54	55	55	57	64	56
3	58	76.75	140	144	144	142
4	107	111	128	128	144	107
5	33	49	54	105	106	53
6	60	92	102	103	105	103
7	44	52	68	84	90	68
8	31	50	55	57	63	56
9	34	45	46	51	112	44
10	28	68	74	125	133	131
11	31	36	42	49	113	34
12	34	58	62	75	144	61
13	39	54	100	102	111	101
14	40	44	51	55	61	54
15	39	53	55	55	63	54
16	44	46	49	57.25	69	50
17	47	58	58	64	68	58
18	22	64	70	71	85	71
19	47	55	75	101.25	164	101
20	30	34.75	42	44	84	45
21	57	78	83	85	91	83
22	55	55	55	55	55	55
23	37	55	55	55	55	55
24	29	57	63	73.5	84	59
25	22	40	45	62	143	55
26	46	54	55	58	64	57
27	30	41	55	68.25	76	73
28	24	38.75	44	59.5	136	33
29	32	42	55	55	55	55
30	53	55	55	55	55	55
31	32	50	52	55	85	52
32	27	35	44	53.25	78	50
33	51	86	89	110	115	110
34	27	32	44	45	114	45
35	42	43	44	46	65	43
36	79	97	97	97	103	97
37	24	40	54	55	65	55

Table C.3 Summary statistics of maximum duration (in months) across study basins, including the minimum (Min), first quartile (Q1), median, third quartile (Q3), maximum (Max), and the model ensemble estimate (Ens).

Basin	Average duration (Months)					
	Min	Q1	Median	Q3	Max	Ens
1	10.31	20.22	24.25	37.63	72.00	14.17
2	34.38	43.06	55.00	56.00	59.50	55.50
3	19.21	25.88	28.65	39.31	47.25	45.75
4	13.83	18.38	24.00	26.04	72.00	18.63
5	9.94	18.20	20.83	25.35	40.67	17.29
6	13.00	15.88	21.50	27.00	52.00	21.67
7	8.31	11.24	12.71	17.46	26.17	9.81
8	15.92	19.68	22.38	24.42	28.33	27.50
9	15.02	24.07	30.42	32.35	56.00	31.83
10	14.63	18.35	20.33	24.38	33.67	25.88
11	15.10	26.25	28.54	30.82	41.00	27.58
12	15.70	18.56	19.29	20.83	31.00	18.50
13	20.75	30.81	57.50	60.00	67.00	61.00
14	18.43	22.75	24.42	29.75	58.00	28.25
15	18.25	31.50	40.25	42.25	56.50	40.50
16	14.42	20.10	24.05	27.13	38.00	23.90
17	14.00	16.10	18.02	20.00	30.00	18.50
18	12.50	14.92	17.00	17.44	22.88	17.17
19	23.50	28.96	33.00	43.00	84.50	43.25
20	24.25	27.29	27.92	29.33	33.88	29.70
21	25.83	50.50	67.50	69.50	72.50	69.00
22	32.33	33.47	35.83	37.75	38.50	38.00
23	24.00	37.34	40.00	40.83	41.17	37.50
24	11.83	18.16	22.29	25.97	39.25	19.00
25	11.17	17.25	21.31	26.35	74.50	30.25
26	22.88	25.25	30.00	43.31	58.50	30.25
27	16.92	23.00	29.50	33.00	48.25	34.83
28	8.56	14.25	16.46	18.50	30.00	14.50
29	14.58	25.25	35.59	39.88	48.00	39.17
30	32.50	34.30	35.13	36.25	40.50	33.63
31	19.92	34.00	35.81	38.58	70.00	36.38
32	12.06	17.68	21.15	24.91	42.00	34.25
33	25.33	36.50	37.50	64.25	66.25	64.75
34	12.94	14.13	15.32	18.57	62.50	14.25
35	15.67	18.65	20.04	22.30	40.50	18.50
36	25.00	36.23	43.38	47.08	76.50	54.50
37	12.77	21.25	34.80	37.28	55.00	41.00

Table C.4 Summary statistics of average duration (in months) across study basins, including the minimum (Min), first quartile (Q1), median, third quartile (Q3), maximum (Max), and the model ensemble estimate (Ens).

Basin	Severity					
	Min	Q1	Median	Q3	Max	Ens
1	-100.91	-97.39	-93.49	-91.13	-76.44	-92.31
2	-104.39	-101.61	-101.16	-99.97	-93.72	-101.49
3	-95.14	-92.01	-90.54	-88.61	-81.68	-89.40
4	-101.47	-98.76	-97.32	-96.38	-94.87	-97.06
5	-100.87	-94.00	-91.57	-88.78	-81.06	-89.51
6	-103.40	-97.26	-96.46	-95.10	-91.21	-98.26
7	-100.25	-97.22	-96.06	-94.75	-91.54	-95.51
8	-96.70	-93.12	-92.06	-91.21	-89.04	-92.41
9	-101.90	-94.37	-92.92	-91.59	-88.31	-91.46
10	-99.87	-92.50	-91.52	-90.97	-86.38	-90.93
11	-97.61	-94.99	-93.96	-93.12	-84.09	-93.98
12	-98.13	-96.18	-94.95	-93.87	-90.24	-95.06
13	-106.37	-103.22	-101.40	-95.02	-87.25	-103.61
14	-101.23	-99.38	-95.54	-94.14	-92.22	-97.50
15	-106.57	-101.66	-98.78	-97.33	-94.91	-98.93
16	-100.42	-96.85	-95.97	-94.71	-92.57	-94.25
17	-100.55	-98.87	-97.55	-96.90	-93.50	-100.27
18	-102.50	-101.18	-100.32	-99.09	-97.50	-100.44
19	-98.12	-91.12	-90.88	-90.18	-88.27	-89.99
20	-94.88	-93.40	-92.67	-91.74	-89.19	-92.65
21	-103.67	-101.92	-100.49	-98.77	-93.91	-101.94
22	-100.45	-99.96	-98.60	-98.19	-94.31	-98.71
23	-100.31	-99.86	-99.63	-99.15	-94.57	-99.86
24	-103.93	-97.81	-92.85	-91.06	-86.34	-94.70
25	-100.62	-95.67	-93.34	-90.65	-85.00	-89.46
26	-102.74	-100.31	-99.67	-98.82	-97.08	-99.96
27	-102.65	-100.83	-98.93	-96.38	-92.71	-102.79
28	-100.54	-95.72	-94.36	-92.70	-79.78	-90.32
29	-102.53	-99.25	-97.17	-93.86	-80.82	-98.62
30	-101.30	-97.64	-95.25	-93.86	-91.46	-95.25
31	-103.34	-100.66	-99.46	-98.01	-92.22	-99.89
32	-97.20	-91.60	-89.81	-88.38	-81.05	-89.05
33	-105.33	-102.46	-100.37	-98.19	-95.96	-101.37
34	-100.75	-95.97	-95.13	-93.84	-89.39	-93.88
35	-98.67	-97.62	-96.25	-95.79	-92.20	-97.06
36	-101.89	-96.95	-93.78	-91.38	-88.08	-94.70
37	-98.38	-96.20	-94.62	-92.00	-87.56	-94.70

Table C.5 Summary statistics of severity across study basins, including the minimum (Min), first quartile (Q1), median, third quartile (Q3), maximum (Max), and the model ensemble estimate (Ens).

Basin	Maximum Intensity					
	Min	Q1	Median	Q3	Max	Ens
1	-7.41	-4.66	-3.12	-2.13	-2.01	-2.89
2	-3.06	-2.64	-2.46	-2.33	-2.06	-2.40
3	-4.16	-2.31	-1.88	-1.66	-1.44	-1.45
4	-2.07	-1.80	-1.74	-1.69	-1.59	-1.65
5	-3.57	-2.92	-2.52	-2.28	-1.83	-2.64
6	-3.65	-3.41	-3.22	-3.04	-2.40	-2.88
7	-2.72	-2.14	-1.99	-1.91	-1.64	-1.86
8	-3.11	-2.85	-2.76	-2.54	-2.43	-2.77
9	-4.53	-3.42	-2.69	-2.39	-1.54	-2.55
10	-3.58	-2.07	-2.01	-1.96	-1.78	-1.96
11	-3.78	-2.48	-2.29	-2.10	-1.67	-2.19
12	-3.42	-2.47	-2.13	-1.85	-1.54	-2.10
13	-3.23	-2.59	-2.18	-1.93	-1.60	-2.13
14	-2.92	-2.48	-2.31	-2.19	-2.00	-2.18
15	-3.52	-2.86	-2.48	-2.35	-1.92	-2.61
16	-3.22	-2.70	-2.54	-2.41	-1.92	-2.47
17	-3.05	-2.40	-2.28	-2.09	-1.78	-2.31
18	-3.75	-3.28	-3.04	-2.81	-2.23	-3.09
19	-2.10	-1.94	-1.86	-1.78	-1.61	-1.80
20	-2.59	-2.13	-2.05	-1.96	-1.73	-1.94
21	-4.31	-2.84	-2.53	-2.11	-1.70	-2.27
22	-2.76	-2.41	-2.28	-2.19	-2.01	-2.26
23	-3.48	-2.69	-2.62	-2.52	-2.33	-2.68
24	-3.34	-2.95	-2.66	-2.28	-1.74	-2.73
25	-3.86	-3.23	-2.98	-2.75	-2.27	-3.14
26	-2.90	-2.69	-2.57	-2.40	-2.00	-2.47
27	-3.79	-2.84	-2.34	-2.19	-1.74	-2.33
28	-4.03	-2.56	-2.34	-2.16	-1.85	-2.36
29	-5.05	-2.99	-2.76	-2.55	-2.12	-2.75
30	-3.25	-2.89	-2.71	-2.53	-2.31	-2.74
31	-2.62	-2.36	-2.24	-2.13	-1.94	-2.23
32	-4.18	-3.43	-3.05	-2.75	-2.17	-2.61
33	-3.01	-2.17	-2.05	-1.96	-1.82	-2.03
34	-2.89	-2.70	-2.60	-2.51	-1.97	-2.63
35	-2.79	-2.49	-2.30	-2.11	-1.77	-1.99
36	-2.70	-2.08	-1.76	-1.63	-1.41	-1.63
37	-3.45	-2.93	-2.68	-2.40	-2.08	-2.60

Table C.6 Summary statistics of maximum intensity across study basins, including the minimum (Min), first quartile (Q1), median, third quartile (Q3), maximum (Max), and the model ensemble estimate (Ens).

References

Agutu, N. O., Awange, J. L., Ndehedehe, C., Kirimi, F., & Kuhn, M. (2019). GRACE-derived groundwater changes over Greater Horn of Africa: temporal variability and the potential for irrigated agriculture. *Science of The Total Environment*, 693, 133467. https://doi.org/10.1016/j.scitotenv.2019.07.273

Ahmed, K., Shamsuddin, S., Demirel, M. C., Nadeem, N., & Najeebullah, K. (2019). The changing characteristics of groundwater sustainability in Pakistan from 2002 to 2016. *Hydrogeology Journal*, *27*(7), 2485-2496. https://doi.org/10.1007/s10040-019-02023-x

Ahmed, M., Sultan, M., Yan, E., & Wahr, J. (2016). Assessing and improving land surface model outputs over Africa using GRACE, field, and remote sensing data. *Surveys in Geophysics*, 37, 529-556. https://doi.org/10.1007/s10712-016-9360-8

Akhtar, F., Nawaz, R. A., Hafeez, M., Awan, U. K., Borgemeister, C., & Tischbein, B. (2022). Evaluation of GRACE derived groundwater storage changes in different agro-ecological zones of the Indus Basin. *Journal of Hydrology*, 605, 127369. https://doi.org/10.1016/j.jhydrol.2021.127369

Akl, M., & Thomas, B. F. (2023). Evaluating the Use of" Goodness-of-Fit" Metrics in GRACE Validation: GRACE Accuracy for Monitoring Groundwater Dynamics. Authorea Preprints. https://doi.org/10.22541/essoar.167388164.42151735/v1

Akl, M., & Thomas, B. F. (2024). Challenges in applying water budget framework for estimating groundwater storage changes from GRACE observations. *Journal of Hydrology*, 639, 131600. https://doi.org/10.1016/j.jhydrol.2024.131600

Akl, M., Thomas, B., & Mills, J. (2022, May). GRACE-derived groundwater storage estimation: Lake/Reservoir storage controls across Canada. In *EGU General Assembly Conference Abstracts* (pp. EGU22-2495). https://doi.org/10.5194/egusphere-egu22-2495

Alghafli, K., Shi, X., Sloan, W., Shamsudduha, M., Tang, Q., Sefelnasr, A., & Ebraheem, A. A. (2023). Groundwater recharge estimation using in-situ and GRACE observations in the eastern region of the United Arab Emirates. *Science of the Total Environment*, *867*, 161489. http://dx.doi.org/10.1016/j.scitotenv.2023.161489

Ali, S., Liu, D., Fu, Q., Cheema, M. J. M., Pal, S. C., Arshad, A., ... & Zhang, L. (2022). Constructing high-resolution groundwater drought at spatio-temporal scale using GRACE

satellite data based on machine learning in the Indus Basin. *Journal of Hydrology*, 612, 128295. https://doi.org/10.1016/j.jhydrol.2022.128295

Ali, S., Wang, Q., Liu, D., Fu, Q., Rahaman, M. M., Faiz, M. A., & Cheema, M. J. M. (2022). Estimation of spatio-temporal groundwater storage variations in the Lower Transboundary Indus Basin using GRACE satellite. *Journal of Hydrology*, 605, 127315. https://doi.org/10.1016/j.jhydrol.2021.127315

Alley, W. M., Healy, R. W., LaBaugh, J. W., & Reilly, T. E. (2002). Flow and storage in groundwater systems. *Science*, *296*(5575), 1985–1990. https://doi.org/10.1126/science.1067123

Alley, W. M., & Konikow, L. F. (2015). Bringing GRACE down to earth. *Groundwater*, *53*(6), 826-829. https://doi.org/10.1111/gwat.12379

Alsdorf, D. E., Rodríguez, E., & Lettenmaier, D. P. (2007). Measuring surface water from space. *Reviews of Geophysics*, *45*(2). https://doi.org/10.1029/2006RG000197

Alshehri, F., & Mohamed, A. (2023). Analysis of groundwater storage fluctuations using GRACE and remote sensing data in Wadi As-Sirhan, Northern Saudi Arabia. *Water*, *15*(2), 282. https://doi.org/10.3390/w15020282

Althoff, D., & Rodrigues, L. N. (2021). Goodness-of-fit criteria for hydrological models: Model calibration and performance assessment. *Journal of Hydrology*, 600, 126674. https://doi.org/10.1016/j.jhydrol.2021.126674

Amiri, V., Ali, S., & Sohrabi, N. (2023). Estimating the spatio-temporal assessment of GRACE/GRACE-FO derived groundwater storage depletion and validation with in-situ water quality data (Yazd province, central Iran). *Journal of Hydrology*, 620, 129416. https://doi.org/10.1016/j.jhydrol.2023.129416

Aon, S., Nandi, S., Sen, S., & Biswas, S. (2024). GRACE based groundwater drought evaluation of Ganga Basin and analysis of drought propagation using wavelet based quantitative approach. *Science of The Total Environment*, 951, 175666. https://doi.org/10.1016/j.scitotenv.2024.175666

Arifin, Taylor, R., Shamsudduha, M., & Ramdhan, A. M. (2025). Groundwater storage dynamics and climate variability in the Lower Kutai Basin of Indonesia: reconciling GRACE ΔGWS to piezometry. *EGUsphere*, *2025*, 1-34. https://doi.org/10.5194/egusphere-2025-2941

Awange, J. L., Forootan, E., Fleming, K., & Odhiambo, G. (2014). Dominant patterns of water storage changes in the Nile Basin during 2003–2013. *Remote Sensing of the Terrestrial Water Cycle*, 367-381. https://doi.org/10.1002/9781118872086.ch22

Awange, J. L., Sharifi, M. A., Ogonda, G., Wickert, J., Grafarend, E. W., & Omulo, M. A. (2008). The falling Lake Victoria water level: GRACE, TRIMM and CHAMP satellite analysis of the lake basin. *Water Resources Management*, *22*, 775-796. https://doi.org/10.1007/s11269-007-9191-y

Bettadpur, S. (2007). Level-2 gravity field product user handbook. The GRACE Project (Jet Propulsion Laboratory, Pasadena, CA, 2003).

Beven, K. (2002). Towards a coherent philosophy for modelling the environment. *Proceedings of the royal society of London. Series A: mathematical, physical and engineering sciences*, *458*(2026), 2465-2484. https://doi.org/10.1002/hyp.343

Bhanja, S. N., Mukherjee, A., & Rodell, M. (2020). Groundwater storage change detection from in situ and GRACE-based estimates in major river basins across India. *Hydrological Sciences Journal*, *65*(4), 650-659. https://doi.org/10.1080/02626667.2020.1716238

Bhanja, S. N., Mukherjee, A., Saha, D., Velicogna, I., & Famiglietti, J. S. (2016). Validation of GRACE based groundwater storage anomaly using in-situ groundwater level measurements in India. *Journal of Hydrology*, *543*, 729-738. https://doi.org/10.1016/j.jhydrol.2016.10.042

Bhanja, S. N., Zhang, X., & Wang, J. (2018). Estimating long-term groundwater storage and its controlling factors in Alberta, Canada. *Hydrology and Earth System Sciences*, *22*(12), 6241-6255. https://doi.org/10.5194/hess-22-6241-2018

Bierkens, M. F., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., ... & Sudicky, E. (2015). Hyper-resolution global hydrological modelling: what is next?. *Hydrological processes*, *29*(2). https://doi.org/10.1002/hyp.10391

Bierkens, M. F., & Wada, Y. (2019). Non-renewable groundwater use and groundwater depletion: a review. *Environmental Research Letters*, *14*(6), 063002. https://doi.org/10.1088/1748-9326/ab1a5f

Bloomfield, J. P., & Marchant, B. P. (2013). Analysis of groundwater drought building on the standardised precipitation index approach. *Hydrology and Earth System Sciences*, *17*(12), 4769-4787. https://doi.org/10.5194/hess-17-4769-2013

Bloomfield, J. P., Marchant, B. P., Bricker, S. H., & Morgan, R. B. (2015). Regional analysis of groundwater droughts using hydrograph classification. *Hydrology and Earth System Sciences*, *19*(10), 4327-4344. https://doi.org/10.5194/hess-19-4327-2015

Bonsor, H. C., Shamsudduha, M., Marchant, B. P., Macdonald, A. M., & Taylor, R. G. (2018). Seasonal and decadal groundwater changes in African sedimentary aquifers estimated using GRACE products and LSMs. *Remote Sensing*, *10*(6), 904. https://doi.org/10.3390/rs10060904

Brassel, K. E., & Reif, D. (1979). A procedure to generate Thiessen polygons. *Geographical Analysis*, *11*(3), 289-303. https://doi.org/10.1111/j.1538-4632.1979.tb00695.x

Busker, T., de Roo, A., Gelati, E., Schwatke, C., Adamovic, M., Bisselink, B., ... & Cottam, A. (2019). A global lake and reservoir volume analysis using a surface water dataset and satellite altimetry. *Hydrology and Earth System Sciences*, *23*(2), 669-690. https://doi.org/10.5194/hess-23-669-2019

Cao, Y., Nan, Z., & Hu, X. (2012, July). Estimating groundwater storage changes in the Heihe river basin using grace. *In 2012 IEEE International Geoscience and Remote Sensing Symposium* (pp. 798-801). IEEE.

Castle, S. L., Thomas, B. F., Reager, J. T., Rodell, M., Swenson, S. C., & Famiglietti, J. S. (2014). Groundwater depletion during drought threatens future water security of the Colorado River Basin. *Geophysical research letters*, *41*(16), 5904-5911. https://doi.org/10.1002/2014GL061055

Chen, H., Zhang, W., Nie, N., & Guo, Y. (2019). Long-term groundwater storage variations estimated in the Songhua River Basin by using GRACE products, land surface models, and in-situ observations. *Science of the Total Environment*, 649, 372-387. https://doi.org/10.1016/j.scitotenv.2018.08.352

Chen, J., Famigliett, J. S., Scanlon, B. R., & Rodell, M. (2016a). Groundwater storage changes: present status from GRACE observations. *In Remote Sensing and Water Resources* (pp. 207-227). Springer, Cham. https://doi.org/10.1007/s10712-015-9332-4

Chen, J., Li, J., Zhang, Z., & Ni, S. (2014). Long-term groundwater variations in Northwest India from satellite gravity measurements. *Global and Planetary Change*, *116*, 130-138. https://doi.org/10.1016/j.gloplacha.2014.02.007

Chen, J. L., Wilson, C. R., Tapley, B. D., Scanlon, B., & Güntner, A. (2016b). Long-term

groundwater storage change in Victoria, Australia from satellite gravity and in situ observations. *Global and Planetary change*, 139, 56-65. http://dx.doi.org/10.1016/j.gloplacha.2016.01.002

Chen, X., Jinbao, J., Tianjie, L., & Chong, Y. (2020). GRACE satellite monitoring and driving factors analysis of groundwater storage under high-intensity coal mining conditions: A case study of Ordos, northern Shaanxi and Shanxi, China. *Hydrogeology Journal*, *28*(2), 673-686. https://doi.org/10.1007/s10040-019-02101-0

Clark, M. P., Fan, Y., Lawrence, D. M., Adam, J. C., Bolster, D., Gochis, D. J., ... & Zeng, X. (2015). Improving the representation of hydrologic processes in Earth System Models. *Water Resources Research*, 51(8), 5929-5956. https://doi.org/10.1002/2015WR017096

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., ... & Hay, L. E. (2008). Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research*, *44*(12). https://

Cohen, I., Huang, Y., Chen, J., Benesty, J., Benesty, J., Chen, J., ... & Cohen, I. (2009). Pearson correlation coefficient. *Noise reduction in speech processing*, 1-4. https://doi.org/10.1007/978-3-642-00296-0

Coles, A. E., McConkey, B. G., & McDonnell, J. J. (2017). Climate change impacts on hillslope runoff on the northern Great Plains, 1962–2013. *Journal of Hydrology*, *550*, 538-548. https://doi.org/10.1016/j.jhydrol.2017.05.023

Condon, L. E., Kollet, S., Bierkens, M. F., Fogg, G. E., Maxwell, R. M., Hill, M. C., ... & Abesser, C. (2021). Global groundwater modeling and monitoring: Opportunities and challenges. *Water Resources Research*, *57*(12), e2020WR029500. https://doi.org/10.1029/2020WR029500

Cooley, S. W., Ryan, J. C., & Smith, L. C. (2021). Human alteration of global surface water storage variability. *Nature*, 591(7848), 78-81. https://doi.org/10.1038/s41586-021-03262-3

Daras, I., Massotti, L., Willemsen, P., March, G., Francois, M., & Carnicero Dominguez, B. (2024, April). Next Generation Gravity Mission (NGGM) status overview and scientific outlook. In EGU General Assembly Conference Abstracts (p. 11271). https://doi.org/10.5194/egusphere-egu24-11271

Döll, P. (2009). Vulnerability to the impact of climate change on renewable groundwater resources: a global-scale assessment. *Environmental Research Letters*, *4*(3), 035006. https://doi.org/10.1088/1748-9326/4/3/035006

Döll, P., Hoffmann-Dobrev, H., Portmann, F. T., Siebert, S., Eicker, A., Rodell, M., ... & Scanlon, B. R. (2012). Impact of water withdrawals from groundwater and surface water on continental water storage variations. *Journal of Geodynamics*, *59*, 143-156. https://doi.org/10.1016/j.jog.2011.05.001

Döll, P., Kaspar, F., & Lehner, B. (2003). A global hydrological model for deriving water availability indicators: Model tuning and validation. *Journal of Hydrology*, *270*(1–2), 105–134. https://doi.org/10.1016/S0022-1694(02)00283-4

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

Döring, S. (2020). Come rain, or come wells: How access to groundwater affects communal violence. *Political geography*, *76*, 102073. https://doi.org/10.1016/j.polgeo.2019.102073

Dracup, J. A., Lee, K. S., & Paulson Jr, E. G. (1980). On the statistical characteristics of drought events. *Water resources research*, *16*(2), 289-296. https://doi.org/10.1029/WR016i002p00289

Duan, Z., & Bastiaanssen, W. G. M. (2013). Estimating water volume variations in lakes and reservoirs from four operational satellite altimetry databases and satellite imagery data. *Remote Sensing of Environment*, 134, 403-416. https://doi.org/10.1016/j.rse.2013.03.010

Dubois, E., & Larocque, M. (2024). Contribution of standardized indexes to understand groundwater level fluctuations in response to meteorological conditions in cold and humid climates. *Journal of Hydrology*, *634*, 131105. https://doi.org/10.1016/j.jhydrol.2024.131105

Dumanski, S., Pomeroy, J. W., & Westbrook, C. J. (2015). Hydrological regime changes in a Canadian Prairie basin. *Hydrological Processes*, *29*(18), 3893-3904. https://doi.org/10.1002/hyp.10567

Elsner, M. M., Gangopadhyay, S., Pruitt, T., Brekke, L. D., Mizukami, N., & Clark, M. P. (2014). How does the choice of distributed meteorological data affect hydrologic model calibration and streamflow simulations?. *Journal of Hydrometeorology*, *15*(4), 1384-1403. https://doi.org/10.1175/JHM-D-13-083.1

Famiglietti, J. S. (2014). The global groundwater crisis. *Nature Climate Change*, *4*(11), 945-948. https://doi.org/10.1038/nclimate2425

Famiglietti, J. S., Lo, M., Ho, S. L., Bethune, J., Anderson, K. J., Syed, T. H., ... & Rodell, M. (2011). Satellites measure recent rates of groundwater depletion in California's Central Valley. *Geophysical Research Letters*, *38*(3). https://doi.org/10.1029/2010GL046442

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

Fan, Y., Li, H., & Miguez-Macho, G. (2013). Global patterns of groundwater table depth. *Science*, *339*(6122), 940-943. https://doi.org/10.1126/science.1229881

Fatolazadeh, F., Eshagh, M., & Goïta, K. (2022). New spectro-spatial downscaling approach for terrestrial and groundwater storage variations estimated by GRACE models. *Journal of Hydrology*, *615*, 128635. https://doi.org/10.1016/j.jhydrol.2022.128635

Fatolazadeh, F., Voosoghi, B., & Naeeni, M. R. (2016). Wavelet and Gaussian approaches for estimation of groundwater variations using GRACE data. *groundwater*, *54*(1), 74-81. https://doi.org/10.1111/gwat.12325

Feng, W., Shum, C., Zhong, M., & Pan, Y. (2018). Groundwater Storage Changes in China from Satellite Gravity: An Overview, *Remote Sens.*, *10*, 674. https://doi:10.3390/rs10050674

Feng, W., Zhong, M., Lemoine, J. M., Biancale, R., Hsu, H. T., & Xia, J. (2013). Evaluation of groundwater depletion in North China using the Gravity Recovery and Climate Experiment (GRACE) data and ground-based measurements. *Water Resources Research*, *49*(4), 2110-2118. https://doi.org/10.1002/wrcr.20192

Ferreira, M. M., Santos, J. A., da Silva, L. R., Abrahao, R., Gomes, F. D. S. V., & Braz, H. D. M. (2023). A new index to evaluate renewable energy potential: A case study on solar, wind and hybrid generation in Northeast Brazil. *Renewable Energy*, *217*, 119182. https://doi.org/10.1016/j.renene.2023.119182

Fogg, G. E., & LaBolle, E. M. (2006). Motivation of synthesis, with an example on groundwater quality sustainability. *Water Resources Research*, *42*(3). https://doi.org/10.1029/2005WR004372

Frappart, F., Papa, F., Güntner, A., Werth, S., Da Silva, J. S., Tomasella, J., ... & Bonnet, M. P. (2011a). Satellite-based estimates of groundwater storage variations in large drainage basins with extensive floodplains. *Remote Sensing of Environment*, *115*(6), 1588-1594. https://doi.org/10.1016/j.rse.2011.02.003

Frappart, F., & Ramillien, G. (2018). Monitoring groundwater storage changes using the Gravity Recovery and Climate Experiment (GRACE) satellite mission: A review. *Remote Sensing*, *10*(6), 829. https://doi.org/10.3390/rs10060829

Gain, A. K., Giupponi, C., & Wada, Y. (2016). Measuring global water security towards sustainable development goals. *Environmental Research Letters*, *11*(12), 124015. https://doi.org/10.1088/1748-9326/11/12/124015

Gerdener, H., Kusche, J., Schulze, K., Döll, P., & Klos, A. (2023). The global land water storage data set release 2 (GLWS2. 0) derived via assimilating GRACE and GRACE-FO data into a global hydrological model. *Journal of Geodesy*, *97*(7), 73. https://doi.org/10.1007/s00190-023-01763-9

Getirana, A., Kumar, S., Girotto, M., & Rodell, M. (2017). Rivers and floodplains as key components of global terrestrial water storage variability. *Geophysical Research Letters*, *44*(20), 10-359. https://doi.org/10.1002/2017GL074684

Giordano, M. (2009). Global groundwater? Issues and solutions. *Annual review of Environment and Resources*, *34*(1), 153-178. https://doi.org/10.1146/annurev.environ.030308.100251

Girotto, M., De Lannoy, G. J., Reichle, R. H., & Rodell, M. (2016). Assimilation of gridded terrestrial water storage observations from GRACE into a land surface model. *Water Resources Research*, *52*(5), 4164-4183. https://doi.org/10.1002/2015WR018417

Gleeson, T., Cuthbert, M., Ferguson, G., & Perrone, D. (2020). Global groundwater sustainability, resources, and systems in the Anthropocene. *Annual review of earth and planetary sciences*, 48, 431-463. https://doi.org/10.1146/annurev-earth-071719055251

Gleeson, T., VanderSteen, J., Sophocleous, M. A., Taniguchi, M., Alley, W. M., Allen, D. M., & Zhou, Y. (2010). Groundwater sustainability strategies. *Nature geoscience*, *3*(6), 378-379. https://doi.org/10.1038/ngeo881

Gleeson, T., Wada, Y., Bierkens, M. F., & Van Beek, L. P. (2012). Water balance of global aquifers revealed by groundwater footprint. *Nature*, *488*(7410), 197-200. https://doi.org/10.1038/nature11295

Gleick, P. H. (2000). A look at twenty-first century water resources development. *Water international*, *25*(1), 127-138. https://doi.org/10.1080/02508060008686804

Goodarzi, M., Abedi-Koupai, J., Heidarpour, M., & Safavi, H. R. (2016). Development of a new drought index for groundwater and its application in sustainable groundwater extraction. *Journal of Water Resources Planning and Management*, *142*(9), 04016032. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000673

Gräler, B., Pebesma, E. J., & Heuvelink, G. B. (2016). Spatio-temporal interpolation using gstat. *R J.*, *8*(1), 204. https://doi.org/10.32614/RJ-2016-014

Guo, J., Mu, D., Liu, X., Yan, H., Sun, Z., & Guo, B. (2016). Water storage changes over the Tibetan Plateau revealed by GRACE mission. *Acta Geophysica*, *64*, 463-476. https://doi.org/10.1515/acgeo-2016-0003

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of hydrology*, *377*(1-2), 80-91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Hachborn, E., Berg, A., Levison, J., & Ambadan, J. T. (2017). Sensitivity of GRACE-derived estimates of groundwater-level changes in southern Ontario, Canada. *Hydrogeology Journal*, *25*(8), 2391-2402. https://doi.org/10.1007/s10040-017-1612-2

Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voß, F., Arnell, N. W., ... & Yeh, P. (2011). Multimodel estimate of the global terrestrial water balance: setup and first results. *Journal of Hydrometeorology*, *12*(5), 869-884. http://www.jstor.org/stable/24912973

Haddeland, I., Heinke, J., Biemans, H., Eisner, S., Flörke, M., Hanasaki, N., ... & Wisser, D. (2014). Global water resources affected by human interventions and climate change. Proceedings of the National Academy of Sciences, 111(9), 3251-3256. https://doi.org/10.1073/pnas.1222475110

Haddeland, I., Skaugen, T., & Lettenmaier, D. P. (2006). Anthropogenic impacts on continental surface water fluxes. *Geophysical Research*

Letters, *33*(8). https://doi.org/10.1029/2006GL026047

Han, S. C., Kim, H., Yeo, I. Y., Yeh, P., Oki, T., Seo, K. W., ... & Luthcke, S. B. (2009). Dynamics of surface water storage in the Amazon inferred from measurements of inter-satellite distance change. *Geophysical Research Letters*, *36*(9). https://doi.org/10.1029/2009GL037910

Han, S. C., Yeo, I. Y., Alsdorf, D., Bates, P., Boy, J. P., Kim, H., ... & Rodell, M. (2010). Movement of Amazon surface water from time-variable satellite gravity measurements and implications for water cycle parameters in land surface models. *Geochemistry, Geophysics, Geosystems*, *11*(9). https://doi.org/10.1029/2010GC003214

Han, Z., Huang, S., Huang, Q., Leng, G., Liu, Y., Bai, Q., ... & Shi, W. (2021). GRACE-based high-resolution propagation threshold from meteorological to groundwater drought. *Agricultural and Forest Meteorology*, 307, 108476. https://doi.org/10.1016/j.agrformet.2021.108476

Hanasaki, N., Kanae, S., & Oki, T. (2006). A reservoir operation scheme for global river routing models. *Journal of Hydrology*, *327*(1-2), 22-41. https://doi.org/10.1016/j.jhydrol.2005.11.011

Hassler, B., & Lauer, A. (2021). Comparison of reanalysis and observational precipitation datasets including ERA5 and WFDE5. *Atmosphere*, *12*(11), 1462. https://doi.org/10.3390/atmos12111462

Helsel, D. R., & Hirsch, R. M. (2002). Statistical methods in water resources. *U.S. geological survey*. https://doi.org/10.3133/twri04A3

Henry, C. M., Allen, D. M., & Huang, J. (2011). Groundwater storage variability and annual recharge using well-hydrograph and GRACE satellite data. *Hydrogeology Journal*, *19*(4), 741. https://doi.org/10.1007/s10040-011-0724-3

Herrera-García, G., Ezquerro, P., Tomás, R., Béjar-Pizarro, M., López-Vinielles, J., Rossi, M., ... & Ye, S. (2021). Mapping the global threat of land subsidence. *Science*, *371*(6524), 34-36. https://doi.org/10.1126/science.abb854

Hirsch, R. M., Slack, J. R., & Smith, R. A. (1982). Techniques of trend analysis for monthly water quality data. *Water resources research*, *18*(1), 107-121. https://doi.org/10.1029/WR018i001p00107

Hou, J., Van Dijk, A. I., & Renzullo, L. J. (2022). Merging Landsat and airborne LiDAR

observations for continuous monitoring of floodplain water extent, depth and volume. *Journal of Hydrology*, *609*, 127684. https://doi.org/10.1016/j.jhydrol.2022.127684

Hou, J., Van Dijk, A. I., Renzullo, L. J., & Larraondo, P. R. (2024). GloLakes: water storage dynamics for 27 000 lakes globally from 1984 to present derived from satellite altimetry and optical imaging. *Earth System Science Data*, *16*(1), 201-218. https://doi.org/10.5194/essd-16-201-2024

Houborg, R., Rodell, M., Li, B., Reichle, R., & Zaitchik, B. F. (2012). Drought indicators based on model-assimilated Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage observations. *Water Resources Research*, *48*(7). https://doi.org/10.1029/2011WR011291, 2012

Hu, Z., Zhang, Z., Sang, Y. F., Qian, J., Feng, W., Chen, X., & Zhou, Q. (2021). Temporal and spatial variations in the terrestrial water storage across Central Asia based on multiple satellite datasets and global hydrological models. *Journal of Hydrology*, 596, 126013. https://doi.org/10.1016/j.jhydrol.2021.126013

Huang, C., Chen, Y., Zhang, S., & Wu, J. (2018). Detecting, extracting, and monitoring surface water from space using optical sensors: A review. Reviews of Geophysics, 56(2), 333-360. https://doi.org/10.1029/2018RG000598

Huang, J., Cao, L., Wang, L., Liu, L., Yu, B., & Han, L. (2023a). Identification and Spatiotemporal Migration Analysis of Groundwater Drought Events in the North China Plain. *Atmosphere*, *14*(6), 961. https://doi.org/10.3390/atmos14060961

Huang, J., Pavlic, G., Rivera, A., Palombi, D., & Smerdon, B. (2016). Mapping groundwater storage variations with GRACE: a case study in Alberta, Canada. *Hydrogeology Journal*, *24*(7), 1663-1680. https://doi.org/10.1007/s10040-016-1412-0

Huang, Z., Pan, Y., Gong, H., Yeh, P. J. F., Li, X., Zhou, D., & Zhao, W. (2015). Subregional-scale groundwater depletion detected by GRACE for both shallow and deep aquifers in North China Plain. *Geophysical Research Letters*, *42*(6), 1791-1799. https://doi.org/10.1002/2014GL062498

Huang, Z., Yeh, P. J. F., Jiao, J. J., Luo, X., Pan, Y., Long, Y., ... & Zheng, L. (2023b). A new approach for assessing groundwater recharge by combining GRACE and baseflow with case studies in karst areas of southwest China. *Water Resources Research*, *59*(2), e2022WR032091. https://doi.org/10.1029/2022WR032091

Hughes, J. D., Petrone, K. C., & Silberstein, R. P. (2012). Drought, groundwater storage and stream flow decline in southwestern Australia. *Geophysical Research Letters*, *39*(3). https://doi.org/10.1029/2011GL050797

Jasechko, S., & Perrone, D. (2021). Global groundwater wells at risk of running dry. *Science*, *372*(6540), 418-421. https://www.science.org/doi/10.1126/science.abc2755

Jasechko, S., Seybold, H., Perrone, D., Fan, Y., Shamsudduha, M., Taylor, R. G., ... & Kirchner, J. W. (2024). Rapid groundwater decline and some cases of recovery in aquifers globally. *Nature*, *625*(7996), 715-721. https://doi.org/10.1038/s41586-023-06879-8

Jin, S., & Feng, G. (2013). Large-scale variations of global groundwater from satellite gravimetry and hydrological models, 2002–2012. *Global and Planetary Change*, 106, 20-30. https://doi.org/10.1016/j.gloplacha.2013.02.008

Jing, W., Zhang, P., & Zhao, X. (2019). A comparison of different GRACE solutions in terrestrial water storage trend estimation over Tibetan Plateau. *Scientific reports*, *9*(1), 1765. https://doi.org/10.1038/s41598-018-38337-1

Jones, E. R., Bierkens, M. F., & van Vliet, M. T. (2024). Current and future global water scarcity intensifies when accounting for surface water quality. *Nature Climate Change*, *14*(6), 629-635. https://doi.org/10.1038/s41558-024-02007-0

Ju, J., Wu, C., Li, J., Yeh, P. J. F., & Hu, B. X. (2023). Global evaluation of model agreement and uncertainty in terrestrial water storage simulations from ISIMIP 2b framework. *Journal of Hydrology*, 129137. https://doi.org/10.1016/j.jhydrol.2023.129137

Kendall, M. G. (1955). Further contributions to the theory of paired comparisons. *Biometrics*, *11*(1), 43-62. https://www.jstor.org/stable/3001479

Khaki, M., Hoteit, I., Kuhn, M., Awange, J., Forootan, E., Van Dijk, A. I., ... & Pattiaratchi, C. (2017). Assessing sequential data assimilation techniques for integrating GRACE data into a hydrological model. *Advances in Water Resources*, *107*, 301-316. https://doi.org/10.1016/j.advwatres.2017.07.001

Khazaei, B., Read, L. K., Casali, M., Sampson, K. M., & Yates, D. N. (2022). GLOBathy, the global lakes bathymetry dataset. *Scientific data*, *9*(1), 1-10. https://doi.org/10.1038/s41597-022-01132-9

Kim, H., Yeh, P. J. F., Oki, T., & Kanae, S. (2009). Role of rivers in the seasonal variations of terrestrial water storage over global basins. *Geophysical Research Letters*, *36*(17).

https://doi.org/10.1029/2009GL039006,2009

Klees, R., Zapreeva, E. A., Winsemius, H. C., & Savenije, H. H. G. (2007). The bias in GRACE estimates of continental water storage variations. *Hydrology and Earth System Sciences*, *11*(4), 1227-1241. https://doi.org/10.5194/hess-11-1227-2007

Kling, H., Fuchs, M., & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of hydrology*, *424*, 264-277. https://doi.org/10.1016/j.jhydrol.2012.01.011

Knoben, W. J., Freer, J. E., & Woods, R. A. (2019). Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrology and Earth System Sciences*, *23*(10), 4323-4331. https://doi.org/10.5194/hess-23-4323-2019

Konapala, G., Kao, S. C., & Addor, N. (2020). Exploring hydrologic model process connectivity at the continental scale through an information theory approach. *Water Resources Research*, *56*(10), e2020WR027340. https://doi.org/10.1029/2020WR027340

Koren, V., Schaake, J., Mitchell, K., Duan, Q. Y., Chen, F., & Baker, J. M. (1999). A parameterization of snowpack and frozen ground intended for NCEP weather and climate models. Journal of Geophysical Research: *Atmospheres*, *104*(D16), 19569-19585. https://doi.org/10.1029/1999JD900232

Krause, P., Boyle, D. P., & Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advances in geosciences*, 5, 89-97. https://doi.org/10.5194/adgeo-5-89-2005

Kuang, X., Liu, J., Scanlon, B. R., Jiao, J. J., Jasechko, S., Lancia, M., ... & Zheng, C. (2024). The changing nature of groundwater in the global water cycle. *Science*, *383*(6686), eadf0630. https://doi.org/10.1126/science.adf0630

Kull, D. (2006). Connections between recent water level drops in Lake Victoria, dam operations and drought. *International Rivers Org*.

Kundzewicz, Z. W., & Doell, P. (2009). Will groundwater ease freshwater stress under climate change?. *Hydrological sciences journal*, *54*(4), 665-675. https://doi.org/10.1623/hysj.54.4.665

Kundzewicz, Z. W., Mata, L. J., Arnell, N. W., Döll, P., Jimenez, B., Miller, K., ... & Shiklomanov, I. (2008). The implications of projected climate change for freshwater resources and their management. *Hydrological sciences journal*, *53*(1), 3-10.

https://doi.org/10.1623/hysj.53.1.3

Lall, U., Josset, L., & Russo, T. (2020). A snapshot of the world's groundwater challenges. *Annual Review of Environment and Resources*, *45*(1), 171-194. https://doi.org/10.1146/annurev-environ-102017-025800

Lamontagne, J. R., Barber, C. A., & Vogel, R. M. (2020). Improved estimators of model performance efficiency for skewed hydrologic data. *Water Resources Research*, *56*(9), e2020WR027101. https://doi.org/ 10.1029/2020WR027101

Landerer, F. W., Flechtner, F. M., Save, H., Webb, F. H., Bandikova, T., Bertiger, W. I., ... & Yuan, D. N. (2020). Extending the global mass change data record: GRACE Follow-On instrument and science data performance. *Geophysical Research Letters*, *47*(12), e2020GL088306. https://doi.org/10.1029/2020GL088306

Landerer, F. W., & Swenson, S. C. (2012). Accuracy of scaled GRACE terrestrial water storage estimates. *Water resources research*, *48*(4). https://doi.org/10.1029/2011WR011453

Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J., ... & Slater, A. G. (2011). Parameterization improvements and functional and structural advances in version 4 of the Community Land Model. *Journal of Advances in Modeling Earth Systems*, *3*(1). https://n2t.org/ark:/85065/d7rf5vhd

Legates, D. R., & Davis, R. E. (1997). The continuing search for an anthropogenic climate change signal: Limitations of correlation-based approaches. *Geophysical Research Letters*, *24*(18), 2319-2322. https://doi.org/10.1029/97GL02207

Legates, D. R., & McCabe Jr, G. J. (1999). Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water resources research*, *35*(1), 233-241. https://doi.org/10.1029/1998WR900018

Lehner, B., Anand, M., Fluet-Chouinard, E., Tan, F., Aires, F., Allen, G. H., ... & Thieme, M. (2024). Mapping the world's inland surface waters: An update to the Global Lakes and Wetlands Database (GLWD v2). Earth System Science Data Discussions, 2024, 1-49. https://doi.org/10.5194/essd-2024-204

Lehner, B., & Döll, P. (2004). Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of hydrology*, 296(1-4), 1-22. https://doi.org/10.1016/j.jhydrol.2004.03.028

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., ... & Wisser,

D. (2011). High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Frontiers in Ecology and the Environment*, *9*(9), 494-502. https://doi.org/10.1890/100125

Lehner, B., Verdin, K., Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. *Eos, Transactions, American Geophysical Union*, *89*(10): 93–94. https://doi.org/10.1029/2008EO100001

Lenk, O. (2013). Satellite based estimates of terrestrial water storage variations in Turkey. *Journal of Geodynamics*, *67*, 106-110. https://doi.org/10.1016/j.jog.2012.04.010

Lezzaik, K., & Milewski, A. (2018). A quantitative assessment of groundwater resources in the Middle East and North Africa region. Hydrogeology Journal, 26(1), 251-266. https://doi.org/10.1007/s10040-017-1646-5

Li, B., Rodell, M., Kumar, S., Beaudoing, H. K., Getirana, A., Zaitchik, B. F., ... & Bettadpur, S. (2019c). Global GRACE data assimilation for groundwater and drought monitoring: Advances and challenges. *Water Resources Research*, *55*(9), 7564-7586. https://doi.org/10.1029/2018WR024618

Li, B., Rodell, M., Sheffield, J., Wood, E., & Sutanudjaja, E. (2019a). Long-term, non-anthropogenic groundwater storage changes simulated by three global-scale hydrological models. *Scientific reports*, *9*(1), 1-13. https://doi.org/10.1038/s41598-019-47219-z

Li, J., & Heap, A. D. (2014). Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling & Software*, *53*, 173-189. https://doi.org/10.1016/j.envsoft.2013.12.008

Li, J., & Wang, S. (2022). Seasonal variations and long-term trends of groundwater over the Canadian landmass. *Hydrogeology Journal*, *30*(2), 401-415. https://doi.org/10.1007/s10040-022-02460-1

Li, P., Karunanidhi, D., Subramani, T., & Srinivasamoorthy, K. (2021). Sources and consequences of groundwater contamination. *Archives of environmental contamination and toxicology*, *80*, 1-10. https://doi.org/10.1007/s00244-020-00805-z

Li, Q., Pan, Y., Zhang, C., & Gong, H. (2023). Quantifying Multi-Source Uncertainties in GRACE-Based Estimates of Groundwater Storage Changes in Mainland China. *Remote Sensing*, *15*(11), 2744. https://doi.org/ 10.3390/rs15112744

Li, Y., Gao, H., Jasinski, M. F., Zhang, S., & Stoll, J. D. (2019b). Deriving high-resolution

reservoir bathymetry from ICESat-2 prototype photon-counting lidar and landsat imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(10), 7883-7893. https://doi.org/10.1109/TGRS.2019.2917012

Liang, X., Lettenmaier, D. P., & Wood, E. F. (1996). One-dimensional statistical dynamic representation of subgrid spatial variability of precipitation in the two-layer variable infiltration capacity model. Journal of Geophysical Research: *Atmospheres*, *101*(D16), 21403-21422. https://doi.org/10.1029/96JD01448

Lindsay, R., Wensnahan, M., Schweiger, A., & Zhang, J. (2014). Evaluation of seven different atmospheric reanalysis products in the Arctic. *Journal of Climate*, *27*(7), 2588-2606. https://doi.org/10.1175/JCLI-D-13-00014.1

Liu, M., Pei, H., & Shen, Y. (2022a). Evaluating dynamics of GRACE groundwater and its drought potential in Taihang Mountain Region, China. *Journal of Hydrology*, 612, 128156. https://doi.org/10.1016/j.jhydrol.2022.128156

Liu, X., Hu, L., Sun, K., Yang, Z., Sun, J., & Yin, W. (2021). Improved understanding of groundwater storage changes under the influence of river basin governance in northwestern China using GRACE data. *Remote Sensing*, *13*(14), 2672. https://doi.org/10.3390/rs13142672

Liu, X., Wang, N., Wang, Y., Meng, N., Wang, Y., Qiao, B., ... & Yang, D. (2025). Research on Groundwater Drought and Sustainability in Badain Jaran Desert and Surrounding Areas Based on GRACE Satellite. *Land*, *14*(1), 173. https://doi.org/10.3390/land14010173

Llamas, M. R., & Martínez-Santos, P. (2005). Intensive groundwater use: silent revolution and potential source of social conflicts. *Journal of water resources planning and management*, *131*(5), 337-341. https://doi.org/10.1061/(ASCE)0733-9496(2005)131:5(337)

Loaiciga, H. A., & Doh, R. (2024). Groundwater for people and the environment: a globally threatened resource. *Groundwater*, *62*(3), 332-340. https://doi.org/10.1111/gwat.13376

Long, D., Chen, X., Scanlon, B. R., Wada, Y., Hong, Y., Singh, V. P., ... & Yang, W. (2016). Have GRACE satellites overestimated groundwater depletion in the Northwest India Aquifer?. *Scientific reports*, *6*(1), 24398. http://dx.doi.org/10.1038/srep24398

Long, D., Yang, Y., Wada, Y., Hong, Y., Liang, W., Chen, Y., ... & Chen, L. (2015). Deriving scaling factors using a global hydrological model to restore GRACE total water storage

changes for China's Yangtze River Basin. *Remote Sensing of Environment*, *168*, 177-193. https://doi.org/10.1016/j.rse.2015.07.003

Longuevergne, L., Scanlon, B. R., & Wilson, C. R. (2010). GRACE Hydrological estimates for small basins: Evaluating processing approaches on the High Plains Aquifer, USA. *Water Resources Research*, *46*(11). https://doi.org/10.1029/2009WR008564

Longuevergne, L., Wilson, C. R., Scanlon, B. R., & Crétaux, J. F. (2013). GRACE water storage estimates for the Middle East and other regions with significant reservoir and lake storage. *Hydrology and Earth System Sciences*, *17*(12), 4817-4830. https://doi.org/10.5194/hess-17-4817-2013

Lu, G. Y., & Wong, D. W. (2008). An adaptive inverse-distance weighting spatial interpolation technique. *Computers & geosciences*, *34*(9), 1044-1055. https://doi.org/10.1016/j.cageo.2007.07.010

Luojus, K., Pulliainen, J., Takala, M., Lemmetyinen, J., Mortimer, C., Derksen, C., ... & Venäläinen, P. (2021). GlobSnow v3. 0 Northern Hemisphere snow water equivalent dataset. *Scientific Data*, *8*(1), 1-16. https://doi.org/10.1038/s41597-021-00939-2

Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica: Journal of the econometric society*, 245-259. https://doi.org/10.2307/1907187

Margat, J. (2007). Which crises in the 21st century?. Making Peace with the Earth: What Future for the Human Species and the Planet.

Margat, J. (2008). Exploitation and use of groundwater in the world. Co-edition: UNESCO and BRGM, 52p .

Margat, J., & Van der Gun, J. (2013). Groundwater around the world: a geographic synopsis. Crc Press.

Mathevet, T., Le Moine, N., Andréassian, V., Gupta, H., & Oudin, L. (2023). Multi-objective assessment of hydrological model performances using Nash–Sutcliffe and Kling–Gupta efficiencies on a worldwide large sample of watersheds. *Comptes Rendus. Géoscience*, *355*(S1), 1-25. https://doi.org/10.5802/crgeos.189

McNally, A., Arsenault, K., Kumar, S., Shukla, S., Peterson, P., Wang, S., ... & Verdin, J. P. (2017). A land data assimilation system for sub-Saharan Africa food and water security applications. *Scientific data*, *4*(1), 1-19. https://doi.org/10.1038/sdata.2017.12

Meghwal, R., Shah, D., & Mishra, V. (2019). On the changes in groundwater storage variability in western India using GRACE and well observations. *Remote Sensing in Earth Systems Sciences*, 2, 260-272. https://doi.org/10.1007/s41976-019-00026-6

Mekonnen, M. M., & Hoekstra, A. Y. (2016). Four billion people facing severe water scarcity. *Science advances*, *2*(2), e1500323. https://doi.org/10.1126/sciadv.1500323

Melati, M. D., Fleischmann, A. S., Fan, F. M., Paiva, R. C., & Athayde, G. B. (2019). Estimates of groundwater depletion under extreme drought in the Brazilian semi-arid region using GRACE satellite data: application for a small-scale aquifer. *Hydrogeology Journal*, *27*(8), 2789-2802. https://doi.org/10.1007/s10040-019-02065-1

Mendoza, P. A., Clark, M. P., Mizukami, N., Newman, A. J., Barlage, M., Gutmann, E. D., ... & Arnold, J. R. (2015). Effects of hydrologic model choice and calibration on the portrayal of climate change impacts. *Journal of Hydrometeorology*, *16*(2), 762-780. https://doi.org/10.1175/JHM-D-14-0104.1

Messager, M.L., Lehner, B., Grill, G., Nedeva, I., Schmitt, O. (2016). Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nature Communications*, 7: 13603. https://doi.org/10.1038/ncomms13603

Meyer, M. F., Virdis, S. G., Yang, X., Brousil, M. R., McClure, R. P., Sharma, S., ... & Shi, H. (2024). The extended Global Lake area, Climate, and Population (GLCP) dataset: Extending the GLCP to include ice, snow, and radiation-related climate variables. https://doi.org/10.6073/pasta/e0bf4571ca6cbfb81c3ed7caefc85fc6

Mishra, A. K., & Singh, V. P. (2010). A review of drought concepts. *Journal of hydrology*, *391*(1-2), 202-216. https://doi.org/10.1016/j.jhydrol.2010.07.012

Mishra, B. K., Kumar, P., Saraswat, C., Chakraborty, S., & Gautam, A. (2021). Water security in a changing environment: Concept, challenges and solutions. *Water*, *13*(4), 490. https://doi.org/10.3390/w13040490

Mitchell, K. E., Lohmann, D., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., ... & Bailey, A. A. (2004). The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *Journal of Geophysical Research: Atmospheres*, *109*(D7). https://doi.org/10.1029/2003JD003823

Mizukami, N., Clark, M. P., Newman, A. J., Wood, A. W., Gutmann, E. D., Nijssen, B., ... & Samaniego, L. (2017). Towards seamless large-domain parameter estimation for hydrologic models. *Water Resources Research*, *53*(9), 8020-8040. https://doi.org/10.1002/2017WR020401

Mizukami, N., P. Clark, M., G. Slater, A., D. Brekke, L., M. Elsner, M., R. Arnold, J., & Gangopadhyay, S. (2014). Hydrologic implications of different large-scale meteorological model forcing datasets in mountainous regions. *Journal of Hydrometeorology*, *15*(1), 474-488. https://doi.org/10.1175/JHM-D-13-036.1

Mohamed, A., Ragaa Eldeen, E., & Abdelmalik, K. (2021). Gravity based assessment of spatio-temporal mass variations of the groundwater resources in the Eastern Desert, Egypt. *Arabian Journal of Geosciences*, 14, 1-15. https://doi.org/10.1007/s12517-021-06885-y

Moore, D. S., Notz, W. I., & Notz, W. (2006). Statistics: Concepts and controversies. *Macmillan*.

Moore, S., & Fisher, J. B. (2012). Challenges and opportunities in GRACE-based groundwater storage assessment and management: an example from Yemen. *Water resources management*, *26*, 1425-1453. https://doi.org/10.1007/s11269-011-9966-z

Morris, B. L., Lawrence, A. R., Chilton, P. J. C., Adams, B., Calow, R. C., & Klinck, B. A. (2003). Groundwater and its susceptibility to degradation: a global assessment of the problem and options for management. [Available at https://nora.nerc.ac.uk/id/eprint/19395/]

Mudryk, L. R., Derksen, C., Kushner, P. J., & Brown, R. (2015). Characterization of Northern Hemisphere snow water equivalent datasets, 1981–2010. *Journal of Climate*, *28*(20), 8037-8051. https://doi.org/10.1175/JCLI-D-15-0229.1

Müller Schmied, H., Adam, L., Eisner, S., Fink, G., Flörke, M., Kim, H., ... & Döll, P. (2016). Variations of global and continental water balance components as impacted by climate forcing uncertainty and human water use. *Hydrology and Earth System Sciences*, *20*(7), 2877-2898. https://doi.org/10.5194/hess-20-2877-2016

Müller Schmied, H., Cáceres, D., Eisner, S., Flörke, M., Herbert, C., Niemann, C., ... & Döll, P. (2021). The global water resources and use model WaterGAP v2. 2d: Model description and evaluation. *Geoscientific Model Development*, 14(2), 1037-1079. https://doi.org/10.5194/gmd-14-1037-2021

Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M., & Döll, P. (2014). Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration. *Hydrology and Earth System Sciences*, *18*(9), 3511-3538. https://doi.org/10.5194/hess-18-3511-2014

Müller Schmied, H., Trautmann, T., Ackermann, S., Cáceres, D., Flörke, M., Gerdener, H., ... & Döll, P. (2024). The global water resources and use model WaterGAP v2. 2e: description and evaluation of modifications and new features. *Geoscientific Model Development*, *17*(23), 8817-8852. https://doi.org/10.5194/gmd-17-8817-2024

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., ... & Thépaut, J. N. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth system science data*, *13*(9), 4349-4383. https://doi.org/10.5194/essd-13-4349-2021

Nandi, S., & Biswas, S. (2024). Spatiotemporal distribution of groundwater drought using GRACE-based satellite estimates: a case study of Lower Gangetic Basin, India. *Environmental Monitoring and Assessment*, *196*(2), 151. https://doi.org/10.1007/s10661-024-12309-7

Nanteza, J., De Linage, C. R., Thomas, B. F., & Famiglietti, J. S. (2016). Monitoring groundwater storage changes in complex basement aquifers: An evaluation of the GRACE satellites over East Africa. *Water Resources Research*, *52*(12), 9542-9564. https://doi.org/10.1002/2016WR018846

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of hydrology*, *10*(3), 282-290. https://doi.org/10.1016/0022-1694(70)90255-6

Nenweli, R., Watson, A., Brookfield, A., Münch, Z., & Chow, R. (2024). Is groundwater running out in the Western Cape, South Africa? Evaluating GRACE data to assess groundwater storage during droughts. *Journal of Hydrology: Regional Studies*, 52, 101699. https://doi.org/10.1016/j.ejrh.2024.101699

Neves, M. C. (2024). Integrating standardized indices and performance indicators for better drought assessment in semi-arid coastal aquifers. *Groundwater for Sustainable Development*, 27, 101341. https://doi.org/10.1016/j.gsd.2024.101341

Neves, M. C., Nunes, L. M., & Monteiro, J. P. (2020). Evaluation of GRACE data for water resource management in Iberia: a case study of groundwater storage monitoring in the

Algarve region. *Journal of Hydrology: Regional Studies*, *32*, 100734. https://doi.org/10.1016/j.ejrh.2020.100734

Nigatu, Z. M., You, W., & Melesse, A. M. (2024). Drought Dynamics in the Nile River Basin: Meteorological, Agricultural, and Groundwater Drought Propagation. *Remote Sensing*, *16*(5), 919. https://doi.org/10.3390/rs16050919

Nikraftar, Z., Parizi, E., Saber, M., Hosseini, S. M., Ataie-Ashtiani, B., & Simmons, C. T. (2024). Groundwater sustainability assessment in the Middle East using GRACE/GRACE-FO data. *Hydrogeology Journal*, *32*(1), 321-337. https://doi.org/10.1007/s10040-023-02717-3

Oliver, M. A., & Webster, R. (1990). Kriging: a method of interpolation for geographical information systems. *International Journal of Geographical Information System*, *4*(3), 313-332. https://doi.org/10.1080/02693799008941549

Onyutha, C. (2016). Identification of sub-trends from hydro-meteorological series. *Stochastic environmental research and risk assessment*, *30*(1), 189-205. https://doi.org/10.1007/s00477-015-1070-0

Opie, S., Taylor, R. G., Brierley, C. M., Shamsudduha, M., & Cuthbert, M. O. (2020). Climate–groundwater dynamics inferred from GRACE and the role of hydraulic memory. *Earth System Dynamics*, *11*(3), 775-791. https://doi.org/10.5194/esd-11-775-2020

Ouatiki, H., Boudhar, A., Leblanc, M., Fakir, Y., & Chehbouni, A. (2022). When climate variability partly compensates for groundwater depletion: An analysis of the GRACE signal in Morocco. *Journal of Hydrology:* Regional Studies, 42, 101177. https://doi.org/10.1016/j.ejrh.2022.101177

Panda, D. K., Mishra, A., Jena, S. K., James, B. K., & Kumar, A. (2007). The influence of drought and anthropogenic effects on groundwater levels in Orissa, India. *Journal of hydrology*, *343*(3-4), 140-153. https://doi.org/10.1016/j.jhydrol.2007.06.007

Panda, D. K., & Wahr, J. (2016). Spatiotemporal evolution of water storage changes in India from the updated GRACE-derived gravity records. *Water Resources Research*, *52*(1), 135-149. http://dx.doi.org/10.1002/2015WR017797

Paredes-Trejo, F., Barbosa, H. A., Giovannettone, J., Kumar, T. L., Thakur, M. K., Buriti, C. D. O., & Uzcátegui-Briceño, C. (2021). Drought assessment in the São Francisco River Basin using satellite-based and ground-based indices. *Remote Sensing*, *13*(19), 3921. https://doi.org/10.3390/rs13193921

Peters, E., Lanen, H. A. J., Bradford, R. B., de Abia, J. C., & Cortina, L. M. (2001). Droughts derived from groundwater heads and groundwater discharge. In Assesment of the Regional Impact of Droughts in Europe (pp. 35-39). Institute of Hydrology.

Pool, S., Vis, M., & Seibert, J. (2018). Evaluating model performance: towards a non-parametric variant of the Kling-Gupta efficiency. *Hydrological Sciences Journal, 63*(13-14), 1941-1953. https://doi.org/10.1080/02626667.2018.1552002

Ramjeawon, M., Demlie, M., & Toucher, M. (2022). Analyses of groundwater storage change using GRACE satellite data in the Usutu-Mhlatuze drainage region, north-eastern South Africa. *Journal of Hydrology: Regional Studies, 42*, 101118. https://doi.org/10.1016/j.ejrh.2022.101118

Rateb, A., Scanlon, B. R., Pool, D. R., Sun, A., Zhang, Z., Chen, J., ... & Zell, W. (2020). Comparison of groundwater storage changes from GRACE satellites with monitoring and modeling of major US aquifers. *Water Resources Research, 56*(12), e2020WR027556. https://doi.org/10.1029/2020WR027556

Richey, A. S., Thomas, B. F., Lo, M. H., Famiglietti, J. S., Swenson, S., & Rodell, M. (2015b). Uncertainty in global groundwater storage estimates in a Total Groundwater Stress framework. *Water resources research, 51*(7), 5198-5216. https://doi.org/10.1002/2015WR017351

Richey, A. S., Thomas, B. F., Lo, M. H., Reager, J. T., Famiglietti, J. S., Voss, K., ... & Rodell, M. (2015a). Quantifying renewable groundwater stress with GRACE. *Water resources research, 51*(7), 5217-5238. https://doi.org/10.1002/2015WR017349.

Rodell, M., Chen, J., Kato, H., Famiglietti, J. S., Nigro, J., & Wilson, C. R. (2007). Estimating groundwater storage changes in the Mississippi River basin (USA) using GRACE. *Hydrogeology Journal, 15*(1), 159-166. https://doi.org/10.1007/s10040-006-0103-7

Rodell, M., & Famiglietti, J. S. (1999). Detectability of variations in continental water storage from satellite observations of the time dependent gravity field. *Water Resources Research, 35*(9), 2705-2723. https://doi.org/10.1029/1999WR900141b

Rodell, M., & Famiglietti, J. S. (2001). An analysis of terrestrial water storage variations in Illinois with implications for the Gravity Recovery and Climate Experiment (GRACE). *Water Resources Research, 37*(5), 1327-1339. https://doi.org/10.1029/2000WR900306

Rodell, M., & Famiglietti, J. S. (2002). The potential for satellite-based monitoring of groundwater storage changes using GRACE: the High Plains aquifer, Central US. *Journal of Hydrology*, *263*(1-4), 245-256. https://doi.org/10.1016/S0022-1694(02)00060-4

Rodell, M., Famiglietti, J. S., Wiese, D. N., Reager, J. T., Beaudoing, H. K., Landerer, F. W., & Lo, M. H. (2018). Emerging trends in global freshwater availability. *Nature*, *557*(7707), 651-659. https://doi.org/10.1038/s41586-018-0123-1

Rodell, M., Houser, P. R., Jambor, U. E. A., Gottschalck, J., Mitchell, K., Meng, C. J., ... & Toll, D. (2004). The global land data assimilation system. *Bulletin of the American Meteorological society*, *85*(3), 381-394. https://doi.org/10.1175/BAMS-85-3-381

Rodell, M., & Reager, J. T. (2023). Water cycle science enabled by the GRACE and GRACE-FO satellite missions. *Nature Water*, *1*(1), 47-59. https://doi.org/10.1038/s44221-022-00005-0

Rodell, M., Velicogna, I., & Famiglietti, J. S. (2009). Satellite-based estimates of groundwater depletion in India. *Nature*, *460*(7258), 999-1002. https://doi.org/10.1038/nature08238

Rohde, M. M., Albano, C. M., Huggins, X., Klausmeyer, K. R., Morton, C., Sharman, A., ... & Stella, J. C. (2024). Groundwater-dependent ecosystem map exposes global dryland protection needs. *Nature*, *632*(8023), 101-107. https://doi.org/10.1038/s41586-024-07702-8

Ruddell, B. L., Drewry, D. T., & Nearing, G. S. (2019). Information theory for model diagnostics: Structural error is indicated by trade-off between functional and predictive performance. *Water Resources Research*, *55*(8), 6534-6554. https://doi.org/10.1029/2018WR023692

Rzepecka, Z., Birylo, M., Jarsjö, J., Cao, F., & Pietroń, J. (2024). Groundwater Storage Variations across Climate Zones from Southern Poland to Arctic Sweden: Comparing GRACE-GLDAS Models with Well Data. *Remote Sensing*, *16*(12), 2104. https://doi.org/10.3390/rs16122104

Sachs, M. (2015). Tools for Estimating and Predicting the Cosinor Model. [Available at https://github.com/sachsmc/cosinor]

Sakumura, C., Bettadpur, S., & Bruinsma, S. (2014). Ensemble prediction and intercomparison analysis of GRACE time-variable gravity field models. *Geophysical Research Letters*, *41*(5), 1389-1397. https://doi.org/10.1002/2013GL058632

Sarkar, T., Kannaujiya, S., Taloor, A. K., Ray, P. K. C., & Chauhan, P. (2020). Integrated study of GRACE data derived interannual groundwater storage variability over water stressed Indian regions. *Groundwater for sustainable development*, 10, 100376. https://doi.org/10.1016/j.gsd.2020.100376

Satish Kumar, K., AnandRaj, P., Sreelatha, K., Bisht, D. S., & Sridhar, V. (2021). Monthly and seasonal drought characterization using grace-based groundwater drought index and its link to teleconnections across south indian river basins. *Climate*, *9*(4), 56. https://doi.org/10.3390/cli9040056

Save, H., Bettadpur, S., & Tapley, B. D. (2012). Reducing errors in the GRACE gravity solutions using regularization. *Journal of Geodesy*, *86*, 695-711. https://doi.org/10.1007/s00190-012-0548-5

Save, H., Bettadpur, S., & Tapley, B. D. (2016). High-resolution CSR GRACE RL05 mascons. Journal of Geophysical Research: *Solid Earth*, *121*(10), 7547-7569. https://doi.org/10.1002/2016JB013007

Saxe, S., Farmer, W., Driscoll, J., & Hogue, T. S. (2021). Implications of model selection: a comparison of publicly available, conterminous US-extent hydrologic component estimates. *Hydrology and Earth System Sciences*, *25*(3), 1529-1568. https://doi.org/10.5194/hess-25-1529-2021

Scanlon, B. R., Fakhreddine, S., Rateb, A., de Graaf, I., Famiglietti, J., Gleeson, T., ... & Zheng, C. (2023). Global water resources and the role of groundwater in a resilient water future. *Nature Reviews Earth & Environment*, *4*(2), 87-101. https://doi.org/10.1038/s43017-022-00378-6

Scanlon, B. R., Jolly, I., Sophocleous, M., & Zhang, L. (2007). Global impacts of conversions from natural to agricultural ecosystems on water resources: Quantity versus quality. *Water resources research*, *43*(3). https://doi.org/10.1029/2006WR005486

Scanlon, B. R., Longuevergne, L., & Long, D. (2012). Ground referencing GRACE satellite estimates of groundwater storage changes in the California Central Valley, USA. *Water Resources Research*, *48*(4). https://doi.org/10.1029/2011WR011312

Scanlon, B. R., Rateb, A., Anyamba, A., Kebede, S., MacDonald, A. M., Shamsudduha, M., ... & Xie, H. (2022). Linkages between GRACE water storage, hydrologic extremes, and climate teleconnections in major African aquifers. *Environmental Research Letters*, *17*(1), 014046. https://doi.org/10.1088/1748-9326/ac3bfc

Scanlon, B. R., Rateb, A., Pool, D. R., Sanford, W., Save, H., Sun, A., ... & Fuchs, B. (2021). Effects of climate and irrigation on GRACE-based estimates of water storage changes in major US aquifers. Environmental Research Letters, 16(9), 094009. https://doi.org/10.1088/1748-9326/ac16ff

Scanlon, B. R., Stonestrom, D. A., Reedy, R. C., Leaney, F. W., Gates, J., & Cresswell, R. G. (2009). Inventories and mobilization of unsaturated zone sulfate, fluoride, and chloride related to land use change in semiarid regions, southwestern United States and Australia. *Water Resources Research*, *45*(7). https://doi.org/10.1029/2008WR006963

Scanlon, B. R., Zhang, Z., Rateb, A., Sun, A., Wiese, D., Save, H., ... & Reedy, R. C. (2019). Tracking seasonal fluctuations in land water storage using global models and GRACE satellites. *Geophysical Research Letters*, *46*(10), 5254-5264. https://doi.org/10.1029/2018GL081836

Scanlon, B. R., Zhang, Z., Save, H., Wiese, D. N., Landerer, F. W., Long, D., ... & Chen, J. (2016). Global evaluation of new GRACE mascon products for hydrologic applications. *Water Resources Research*, *52*(12), 9412-9429. https://doi.org/10.1002/2016WR019494.

Schaefli, B., & Gupta, H. V. (2007). Do Nash values have value?. *Hydrological processes*, *21*, 2075-2080. https://doi.org/10.1002/hyp.6825

Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *Journal of the American statistical association*, *63*(324), 1379-1389. http://www.jstor.com/stable/2285891

Shalby, A., Emara, S. R., Metwally, M. I., Armanuos, A. M., El-Agha, D. E., Negm, A. M., & Gado, T. A. (2023). Satellite-based estimates of groundwater storage depletion over Egypt. Environmental Monitoring and Assessment, 195(5), 594. https://doi.org/10.1007/s10661-023-11171-3

Shamsudduha, M., & Taylor, R. G. (2020). Groundwater storage dynamics in the world's large aquifer systems from GRACE: uncertainty and role of extreme precipitation. *Earth System Dynamics*, *11*(3), 755-774. https://doi.org/10.5194/esd-11-755-2020

Shamsudduha, M., Taylor, R., & Longuevergne, L. (2012). Monitoring groundwater storage changes in the highly seasonal humid tropics: Validation of GRACE measurements in the Bengal Basin. *Water Resources Research*, *48*, W02508. https://doi.org/10.1029/2011WR010993

Shao, C., & Liu, Y. (2023). Analysis of Groundwater Storage Changes and Influencing Factors in China Based on GRACE Data. *Atmosphere, 14*(2), 250. https://doi.org/10.3390/atmos14020250

Shen, H., Leblanc, M., Tweed, S., & Liu, W. (2015). Groundwater depletion in the Hai River Basin, China, from in situ and GRACE observations. *Hydrological Sciences Journal, 60*(4), 671-687. https://doi.org/10.1080/02626667.2014.916406

Shirzaei, M., Freymueller, J., Törnqvist, T. E., Galloway, D. L., Dura, T., & Minderhoud, P. S. (2021). Measuring, modelling and projecting coastal land subsidence. Nature Reviews Earth & Environment, 2(1), 40-58. https://doi.org/10.1038/s43017-020-00115-x

Siebert, S., Burke, J., Faures, J. M., Frenken, K., Hoogeveen, J., Döll, P., & Portmann, F. T. (2010). Groundwater use for irrigation-A global inventory. *Hydrology and Earth System Sciences, 14*(10), 1863–1880. https://doi.org/10.5194/hess-14-1863-2010

Singh, R. S., Reager, J. T., Miller, N. L., & Famiglietti, J. S. (2015). Toward hyper-resolution land-surface modeling: The effects of fine-scale topography and soil texture on CLM 4.0 simulations over the S southwestern US. *Water Resources Research, 51*(4), 2648-2667. https://doi.org/10.1016/j.quaint.2018.10.036

Skaskevych, A., Lee, J., Jung, H. C., Bolten, J., David, J. L., Policelli, F. S., ... & Ichoku, C. M. (2020). Application of GRACE to the estimation of groundwater storage change in a data-poor region: A case study of Ngadda catchment in the Lake Chad Basin. *Hydrological Processes, 34*(4), 941-955. https://doi.org/10.1007/978-981-13-5889-0_7

Śliwińska, J., Birylo, M., Rzepecka, Z., & Nastula, J. (2019). Analysis of groundwater and total water storage changes in Poland using GRACE observations, in-situ data, and various assimilation and climate models. *Remote Sensing, 11*(24), 2949. https://doi:10.3390/rs11242949

Sterl, A. (2004). On the (in) homogeneity of reanalysis products. Journal of Climate, 17(19), 3866-3873. https://doi.org/10.1175/1520-0442(2004)017<3866:OTIORP>2.0.CO;2

Strassberg, G., Scanlon, B. R., & Chambers, D. (2009). Evaluation of groundwater storage monitoring with the GRACE satellite: Case study of the High Plains aquifer, central United States. *Water Resources Research, 45*(5). https://doi.org/10.1029/2008WR006892

Strassberg, G., Scanlon, B. R., & Rodell, M. (2007). Comparison of seasonal terrestrial water storage variations from GRACE with groundwater-level measurements from the High

Plains Aquifer (USA). *Geophysical Research Letters*, *34*(14). https://doi.org/10.1029/2007GL030139

Song, X., Chen, H., Chen, T., Qin, Z., Chen, S., Yang, N., & Deng, S. (2024). GRACE-based groundwater drought in the Indochina Peninsula during 1979–2020: Changing properties and possible teleconnection mechanisms. *Science of The Total Environment*, 908, 168423. https://doi.org/10.1016/j.scitotenv.2023.168423

Sophocleous, M. (2002). Environmental implications of intensive groundwater use with special regard to streams and wetlands. *Intensive use of groundwater: challenges and opportunities. CRC Press, Boca Raton, Louisiana*, 93-112.

Sophocleous, M. (2010). groundwater management practices, challenges, and innovations in the High Plains aquifer, USA—lessons and recommended actions. *Hydrogeology Journal*, *18*(3), 559-575. https://doi.org/10.1007/s10040-009-0540-1

Sultan, M., Ahmed, M., Wahr, J., Yan, E., & Emil, M. K. (2014). Monitoring aquifer depletion from space: case studies from the Saharan and Arabian aquifers. *Remote sensing of the terrestrial water cycle*, 347-366. https://doi.org/10.1002/9781118872086.ch21

Sun, A. Y. (2013). Predicting groundwater level changes using GRACE data. *Water resources research*, *49*(9), 5900-5912. https://doi:10.1002/ wrcr.20421.

Sun, A. Y., Green, R., Rodell, M., & Swenson, S. (2010). Inferring aquifer storage parameters using satellite and in situ measurements: Estimation under uncertainty. *Geophysical Research Letters*, 37(10). https://doi.org/10.1029/2010GL043231

Sutanudjaja, E. H., Van Beek, R., Wanders, N., Wada, Y., Bosmans, J. H., Drost, N., ... & Bierkens, M. F. (2018). PCR-GLOBWB 2: a 5 arcmin global hydrological and water resources model. *Geoscientific Model Development*, *11*(6), 2429-2453. https://doi.org/10.5194/gmd-11-2429-2018

Swenson, S. C., & Lawrence, D. M. (2015). A GRACE-based assessment of interannual groundwater dynamics in the Community Land Model. *Water Resources Research*, *51*(11), 8817-8833. https://doi.org/10.1002/2015WR017582

Swenson, S. C., & Milly, P. C. D. (2006). Climate model biases in seasonality of continental water storage revealed by satellite gravimetry. *Water Resources Research*, *42*(3). https://doi.org/10.1029/2005WR004628

Swenson, S., & Wahr, J. (2002). Methods for inferring regional surface-mass anomalies from

Gravity Recovery and Climate Experiment (GRACE) measurements of time-variable gravity. *Journal of Geophysical Research: Solid Earth, 107*(B9), ETG-3. https://doi.org/10.1029/2001JB000576

Swenson, S., & Wahr, J. (2006). Post-processing removal of correlated errors in GRACE data. *Geophysical research letters, 33*(8). https://doi.org/10.1029/2005GL025285

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

Syed, T. H., Famiglietti, J. S., Rodell, M., Chen, J., & Wilson, C. R. (2008). Analysis of terrestrial water storage changes from GRACE and GLDAS. *Water Resources Research, 44*(2). https://doi.org/10.1029/2006WR005779

Tallaksen, L. M., & Van Lanen, H. A. (Eds.). (2023). Hydrological drought: processes and estimation methods for streamflow and groundwater.

Tapley, B. D., S. Bettadpur, J. C. Ries, P. F. Thompson, and M. M. Watkins (2004), GRACE measurements of mass variability in the Earth system, *Science, 305*(5683), 593–505. https://doi.org/10.1029/2006WR005779

Taylor, R. G., Scanlon, B., Döll, P., Rodell, M., Van Beek, R., Wada, Y., ... & Treidel, H. (2013). Ground water and climate change. *Nature climate change, 3*(4), 322-329. https://doi.org/10.1038/nclimate1744

Theil, H. (1950). A rank-invariant method of linear and polynomial regression analysis. *Indagationes mathematicae, 12*(85), 173.

Thomas, B. F., Behrangi, A., & Famiglietti, J. S. (2016). Precipitation intensity effects on groundwater recharge in the southwestern United States. Water, 8(3), 90. https://doi.org/10.3390/w8030090

Thomas, B. F., Caineta, J., & Nanteza, J. (2017b). Global assessment of groundwater sustainability based on storage anomalies. *Geophysical Research Letters, 44*(22), 11-445. https://doi.org/10.1002/2017GL076005

Thomas, B. F., & Famiglietti, J. S. (2015). Sustainable groundwater management in the arid Southwestern US: Coachella Valley, California. *Water resources management, 29*(12), 4411-4426. https://doi.org/10.1007/s11269-015-1067-y

Thomas, B. F., & Famiglietti, J. S. (2019). Identifying climate-induced groundwater depletion in GRACE observations. *Scientific reports*, *9*(1), 1-9. https://doi.org/10.1038/s41598-019-40155-y

Thomas, B. F., Famiglietti, J. S., Landerer, F. W., Wiese, D. N., Molotch, N. P., & Argus, D. F. (2017a). GRACE groundwater drought index: Evaluation of California Central Valley groundwater drought. *Remote Sensing of Environment*, *198*, 384-392. https://doi.org/10.1016/j.rse.2017.06.026

Thomas, B. F., & Nanteza, J. (2023). Global assessment of the sensitivity of water storage to hydroclimatic variations. *Science of The Total Environment*, 879, 162958. https://doi.org/10.1016/j.scitotenv.2023.162958

Tortini, R., Noujdina, N., Yeo, S., Ricko, M., Birkett, C. M., Khandelwal, A., ... & Lettenmaier, D. P. (2020). Satellite-based remote sensing data set of global surface water storage change from 1992 to 2018. *Earth System Science Data*, *12*(2), 1141-1151. https://doi.org/10.5194/essd-12-1141-2020

United Nations, The United Nations World Water Development Report 2022: Groundwater: Making the invisible visible. UNESCO, Paris. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000380721

Van der Gun, J. (2022). Large aquifer systems around the world. *The Groundwater Project, Guelph, ON*. https://doi.org/10.21083/978-1-77470-020-4

Van Loon, A. F., Stahl, K., Di Baldassarre, G., Clark, J., Rangecroft, S., Wanders, N., ... & Van Lanen, H. A. (2016). Drought in a human-modified world: reframing drought definitions, understanding, and analysis approaches. *Hydrology and Earth System Sciences*, *20*(9), 3631-3650. https://doi.org/10.5194/nhess-24-3173-2024

Van Vliet, M. T., Jones, E. R., Flörke, M., Franssen, W. H., Hanasaki, N., Wada, Y., & Yearsley, J. R. (2021). Global water scarcity including surface water quality and expansions of clean water technologies. *Environmental Research Letters*, *16*(2), 024020. https://doi.org/10.1088/1748-9326/abbfc3

Van Zoest, V., Osei, F. B., Hoek, G., & Stein, A. (2020). Spatio-temporal regression kriging for modelling urban NO2 concentrations. *International journal of geographical information science*, *34*(5), 851-865. https://doi.org/10.1080/13658816.2019.1667501

Vanham, D., Hoekstra, A. Y., Wada, Y., Bouraoui, F., de Roo, A., Mekonnen, M. M., ... & Bidoglio, G. (2018). Physical water scarcity metrics for monitoring progress towards SDG target 6.4: An evaluation of indicator 6.4. 2 “Level of water stress”. *Science of the total environment*, *613*, 218-232. https://doi.org/10.1016/j.scitotenv.2017.09.056

Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., ... & Davies, P. (2010). Global threats to human water security and river biodiversity. nature, 467(7315), 555-561. https://doi.org/10.1038/nature09440

Voss, K. A., Famiglietti, J. S., Lo, M., De Linage, C., Rodell, M., & Swenson, S. C. (2013). Groundwater depletion in the Middle East from GRACE with implications for transboundary water management in the Tigris-Euphrates-Western Iran region. *Water resources research*, *49*(2), 904-914. https://doi.org/10.1002/wrcr.20078

Wada, Y., Van Beek, L. P., Van Kempen, C. M., Reckman, J. W., Vasak, S., & Bierkens, M. F. (2010). Global depletion of groundwater resources. *Geophysical research letters*, *37*(20). https://doi.org/10.1029/2010GL044571

Wahr, J., Molenaar, M., & Bryan, F. (1998). 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*(B12), 30205-30229. https://doi.org/10.1029/98JB02844

Wahr, J., Swenson, S., & Velicogna, I. (2006). Accuracy of GRACE mass estimates. *Geophysical Research Letters*, *33*(6). https://doi.org/10.1029/2005GL025305

Wang, F., Lai, H., Li, Y., Feng, K., Zhang, Z., Tian, Q., ... & Yang, H. (2022a). Identifying the status of groundwater drought from a GRACE mascon model perspective across China during 2003–2018. *Agricultural Water Management*, 260, 107251. https://doi.org/10.1016/j.agwat.2021.107251

Wang, F., Wang, Z., Yang, H., Di, D., Zhao, Y., & Liang, Q. (2020b). Utilizing GRACE-based groundwater drought index for drought characterization and teleconnection factors analysis in the North China Plain. *Journal of Hydrology*, 585, 124849. https://doi.org/10.1016/j.jhydrol.2020.124849

Wang, H., Xiang, L., Steffen, H., Wu, P., Jiang, L., Shen, Q., ... & Hayashi, M. (2022b). GRACE-based estimates of groundwater variations over North America from 2002 to 2017. *Geodesy and Geodynamics*, *13*(1), 11-23. https://doi.org/10.1016/j.geog.2021.10.003

Wang, S., Liu, H., Yu, Y., Zhao, W., Yang, Q., & Liu, J. (2020a). Evaluation of groundwater sustainability in the arid Hexi Corridor of Northwestern China, using GRACE, GLDAS and measured groundwater data products. *Science of the Total Environment*, 705, 135829. https://doi.org/10.1016/j.scitotenv.2019.135829

Waseem, M., Mani, N., Andiego, G., & Usman, M. (2017). A review of criteria of fit for hydrological models. *International Research Journal of Engineering and Technology (IRJET)*, *4*(11), 1765-1772.

Watkins, M. M., Wiese, D. N., Yuan, D. N., Boening, C., & Landerer, F. W. (2015). Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap mascons. *Journal of Geophysical Research: Solid Earth*, *120*(4), 2648-2671. https://doi.org/10.1002/2014JB011547

Wei, L., Jiang, S., Ren, L., Tan, H., Ta, W., Liu, Y., ... & Duan, Z. (2021). Spatiotemporal changes of terrestrial water storage and possible causes in the closed Qaidam Basin, China using GRACE and GRACE Follow-On data. *Journal of Hydrology*, 598, 126274. https://doi.org/10.1016/j.jhydrol.2021.126274

Werner, A. D., Bakker, M., Post, V. E., Vandenbohede, A., Lu, C., Ataie-Ashtiani, B., ... & Barry, D. A. (2013). Seawater intrusion processes, investigation and management: Recent advances and future challenges. *Advances in water resources*, *51*, 3-26. https://doi.org/10.1016/j.advwatres.2012.03.004

Werth, S., & Güntner, A. (2010). Calibration analysis for water storage variability of the global hydrological model WGHM. *Hydrology and Earth System Sciences*, *14*(1), 59-78. https://doi.org/10.5194/hess-14-59-2010

Wiese, D. N., Landerer, F. W., & Watkins, M. M. (2016). Quantifying and reducing leakage errors in the JPL RL05M GRACE mascon solution. *Water Resources Research*, *52*(9), 7490-7502. https://doi.org/10.1002/2016WR019344

Wilhite, D. A., & Glantz, M. H. (1985). Understanding: the drought phenomenon: the role of definitions. *Water international*, *10*(3), 111-120. http://digitalcommons.unl.edu/droughtfacpub/20

Wilhite, D. A., Sivakumar, M. V., & Pulwarty, R. (2014). Managing drought risk in a changing climate: The role of national drought policy. *Weather and climate extremes*, *3*, 4-13. https://doi.org/10.1016/j.wace.2014.01.002

Wouters, B., & Sasgen, I. (2024). The Gravity Recovery and Climate Experiment (GRACE). In *Remote Sensing for Characterization of Geohazards and Natural Resources* (pp. 131-145). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-59306-2_6

Xanke, J., & Liesch, T. (2022). Quantification and possible causes of declining groundwater resources in the Euro-Mediterranean region from 2003 to 2020. Hydrogeology Journal, 30(2), 379-400. https://doi.org/10.1007/s10040-021-02448-3

Xavier, L., Becker, M., Cazenave, A., Longuevergne, L., Llovel, W., & Rotunno Filho, O. C. (2010). Interannual variability in water storage over 2003–2008 in the Amazon Basin from GRACE space gravimetry, in situ river level and precipitation data. *Remote Sensing of Environment*, *114*(8), 1629-1637. https://doi.org/10.1016/j.rse.2010.02.005

Xiao, R., He, X., Zhang, Y., Ferreira, V. G., & Chang, L. (2015). Monitoring groundwater variations from satellite gravimetry and hydrological models: A comparison with in-situ measurements in the mid-atlantic region of the United States. *Remote Sensing*, *7*(1), 686-703. https://doi.org/10.3390/rs70100686

Xu, Y., Gong, H., Chen, B., Zhang, Q., & Li, Z. (2021). Long-term and seasonal variation in groundwater storage in the North China Plain based on GRACE. *International Journal of Applied Earth Observation and Geoinformation*, 104, 102560. https://doi.org/10.1016/j.jag.2021.102560

Yang, J., Pan, Y., Zhang, C., Gong, H., Xu, L., Huang, Z., & Lu, S. (2024). Comparison of groundwater storage changes over losing and gaining aquifers of China using GRACE satellites, modeling and in-situ observations. *Science of The Total Environment*, 173514. https://doi.org/10.1016/j.scitotenv.2024.173514

Yeh, P. J. F., Swenson, S. C., Famiglietti, J. S., & Rodell, M. (2006). Remote sensing of groundwater storage changes in Illinois using the Gravity Recovery and Climate Experiment (GRACE). *Water Resources Research*, *42*(12). https://doi:10.1029/2006WR005374

Yigzaw, W., Li, H. Y., Demissie, Y., Hejazi, M. I., Leung, L. R., Voisin, N., & Payn, R. (2018). A new global storage-area-depth data set for Modeling reservoirs in land surface and earth system models. *Water Resources Research*, *54*(12), 10-372. https://doi.org/10.1029/2017WR022040

Yin, W., Hu, L., & Jiao, J. J. (2017). Evaluation of groundwater storage variations in northern China using GRACE data. *Geofluids*, *2017*(1), 8254824. https://doi.org/10.1155/2017/8254824

Yin, W., Li, T., Zheng, W., Hu, L., Han, S. C., Tangdamrongsub, N., ... & Huang, Z. (2020). Improving regional groundwater storage estimates from GRACE and global hydrological models over Tasmania, Australia. *Hydrogeology Journal*, *28*(5). https://doi.org/10.1007/s10040-020-02157-3

Zar, J. H. (1972). Significance testing of the Spearman rank correlation coefficient. *Journal of the American Statistical Association*, *67*(339), 578-580. https://doi.org/10.1080/01621459.1972.10481251

Zeng, L., & Levy, G. (1995). Space and time aliasing structure in monthly mean polar-orbiting satellite data. *Journal of Geophysical Research: Atmospheres*, *100*(D3), 5133-5142. https://doi.org/10.1080/01621459.1972.10481251

Zeng, N. (1999). Seasonal cycle and interannual variability in the Amazon hydrologic cycle. *Journal of Geophysical Research: Atmospheres*, *104*(D8), 9097-9106. https://doi.org/10.1029/1998JD200088

Zhang, H., Ding, J., Wang, Y., Zhou, D., & Zhu, Q. (2021a). Investigation about the correlation and propagation among meteorological, agricultural and groundwater droughts over humid and arid/semi-arid basins in China. *Journal of Hydrology*, 603, 127007. https://doi.org/10.1016/j.jhydrol.2021.127007

Zhang, Q., Sun, J., Dai, C., Zhang, G., & Wu, Y. (2024). Sustainable development of groundwater resources under the large-scale conversion of dry land into rice fields. *Agricultural Water Management*, 298, 108851. https://doi.org/10.1016/j.agwat.2024.108851

Zhao, J., Li, G., Zhu, Z., Hao, Y., Hao, H., Yao, J., ... & Yeh, T. C. J. (2024). Analysis of the spatiotemporal variation of groundwater storage in Ordos Basin based on GRACE gravity satellite data. *Journal of Hydrology*, 632, 130931. https://doi.org/10.1016/j.jhydrol.2024.130931

Zheng, W., Askari, K., Song, C., Shi, P., Ge, W., Shi, S., ... & Wang, F. (2024). Increasing vulnerability of vegetation to Meteorological and Groundwater drought: A Case study in Argentina. *Journal of Hydrology: Regional Studies*, 55, 101931. https://doi.org/10.1016/j.ejrh.2024.101931

Zhou, T., Nijssen, B., Gao, H., & Lettenmaier, D. P. (2016). The contribution of reservoirs to global land surface water storage variations. *Journal of Hydrometeorology*, *17*(1), 309-325. https://doi.org/10.1175/JHM-D-15-0002.1

Zomer, R.J., Xu, J. & Trabucco, A. Version 3 of the Global Aridity Index and Potential Evapotranspiration Database. *Sci Data 9*, 409 (2022). https://doi.org/10.1038/s41597-022-01493-1

References From Appendix B

Abou Zaki, N., Torabi Haghighi, A., M. Rossi, P., J. Tourian, M., & Kløve, B. (2019). Monitoring groundwater storage depletion using gravity recovery and climate experiment (GRACE) data in Bakhtaran Catchment, Iran. *Water, 11*(7), 1456. https://doi.org/10.3390/w11071456

Brookfield, A. E., Hill, M. C., Rodell, M., Loomis, B. D., Stotler, R. L., Porter, M. E., & Bohling, G. C. (2018). In situ and GRACE-based groundwater observations: Similarities, discrepancies, and evaluation in the High Plains aquifer in Kansas. *Water Resources Research*, *54*(10), 8034-8044. https://doi.org/10.1029/2018WR023836

Camilo, B. K., Andrade, P. L., Eleutério, J. C., & Rodrigues, A. F. (2024). The importance of local observation wells, reanalysis, and satellite data on gravity anomaly, climate, and land use to improve groundwater management in the Urucuia Aquifer System. *Journal of South American Earth Sciences*, 105018. https://doi.org/10.1016/j.jsames.2024.105018

Chanu, C. S., Munagapati, H., Tiwari, V. M., Kumar, A., & Elango, L. (2020). Use of GRACE time-series data for estimating groundwater storage at small scale. *Journal of Earth System Science*, 129, 1-19. https://doi.org/10.1007/s12040-020-01465-2

Chen, B., Xia, J., Wang, Q., Chen, C., Miller, R. D., & Liang, Q. (2010, April). Estimating Groundwater Storage Changes in the Western Kansas Using Grace Data. In 23rd EEGS Symposium on the Application of Geophysics to Engineering and Environmental Problems (pp. cp-175). European Association of Geoscientists & Engineers. https://doi.org/10.3997/2214-4609-pdb.175.SAGEEP031

Chinnasamy, P., Hubbart, J. A., & Agoramoorthy, G. (2013). Using remote sensing data to improve groundwater supply estimations in Gujarat, India. *Earth Interactions*, *17*(1), 1-17. https://doi.org/10.1175/2012EI000456.1

Fatolazadeh, F., & Goïta, K. (2022). Reconstructing groundwater storage variations from

GRACE observations using a new Gaussian-Han-Fan (GHF) smoothing approach. *Journal of Hydrology, 604*, 127234. https://doi.org/10.1016/j.jhydrol.2021.127234

Forootan, E., Rietbroek, R., Kusche, J., Sharifi, M. A., Awange, J. L., Schmidt, M., ... & Famiglietti, J. (2014). Separation of large scale water storage patterns over Iran using GRACE, altimetry and hydrological data. *Remote Sensing of Environment*, *140*, 580-595. http://dx.doi.org/10.1016/j.rse.2013.09.025

Frappart, F., Ramillien, G., Leblanc, M., Tweed, S. O., Bonnet, M. P., & Maisongrande, P. (2011b). An independent component analysis filtering approach for estimating continental hydrology in the GRACE gravity data. *Remote Sensing of Environment*, *115*(1), 187-204. https://doi.org/10.1016/j.rse.2011.02.003

Huang, Q., Wang, L., Jia, B., Lai, X., & Peng, Q. (2023b). Impact of Climate Change on the Spatio-Temporal Variation in Groundwater Storage in the Guangdong–Hong Kong–Macao Greater Bay Area. *Sustainability*, *15*(14), 10776. https://doi.org/10.3390/su151410776

Iqbal, N., Hossain, F., Lee, H., & Akhter, G. (2016). Satellite gravimetric estimation of groundwater storage variations over Indus Basin in Pakistan. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*(8), 3524-3534. http://dx.doi.org/10.1109/JSTARS.2016.2574378

Jawadi, H. A., Farahmand, A., Fensham, R., & Patel, N. (2024). Evaluating groundwater storage variations in Afghanistan using GRACE, GLDAS, and in-situ measurements. *Modeling Earth Systems and Environment*, 1-17. https://doi.org/10.1007/s40808-024-02084-2

Joodaki, G., Wahr, J., & Swenson, S. (2014). Estimating the human contribution to groundwater depletion in the Middle East, from GRACE data, land surface models, and well observations. *Water Resources Research*, *50*(3), 2679-2692. http://dx.doi.org/10.1002/2013WR014633

Katpatal, Y. B., Rishma, C., & Singh, C. K. (2018). Sensitivity of the Gravity Recovery and Climate Experiment (GRACE) to the complexity of aquifer systems for monitoring of groundwater. *Hydrogeology Journal*, *26*(3). https://doi.org/10.1007/s10040-017-1686-x

Kinouchi, T. (2021). Synergetic application of GRACE gravity data, global hydrological model, and in-situ observations to quantify water storage dynamics over Peninsular India during 2002–2017. *Journal of Hydrology*, 596, 126069. https://doi.org/10.1016/j.jhydrol.2021.126069

Leblanc, M. J., Tregoning, P., Ramillien, G., Tweed, S. O., & Fakes, A. (2009). Basin-scale, integrated observations of the early 21st century multiyear drought in southeast Australia. *Water resources research, 45*(4). https://doi.org/10.1029/2008WR007333

Liesch, T., & Ohmer, M. (2016). Comparison of GRACE data and groundwater levels for the assessment of groundwater depletion in Jordan. *Hydrogeology Journal, 24*(6), 1547. http://dx.doi.org/10.1007/s10040-016-1416-9

Liu, P. W., Famiglietti, J. S., Purdy, A. J., Adams, K. H., McEvoy, A. L., Reager, J. T., ... & Rodell, M. (2022b). Groundwater depletion in California's Central Valley accelerates during megadrought. *Nature Communications*, *13*(1), 7825. https://doi.org/10.1038/s41467-022-35582-x

Majid, R., & Ardalan, E. S. (2019). Performance of the Gravity Recovery and Climate Experiment (GRACE) method in monitoring groundwater-level changes in local-scale study regions within Iran. *Hydrogeology Journal*, *27*(7), 2497-2509. https://doi.org/10.1007/s10040-019-02007-x

Nazari, A., Zaryab, A., & Ahmadi, A. (2023). Estimation of groundwater storage change in the Helmand River Basin (Afghanistan) using GRACE satellite data. *Earth Science Informatics*, *16*(1), 579-589. https://doi.org/10.1007/s12145-022-00899-0

Qu, W., Zhang, P., Chen, P., Li, J., & Gao, Y. (2024). Spatiotemporal Variations and Sustainability Characteristics of Groundwater Storage in North China from 2002 to 2022 Revealed by GRACE/GRACE Follow-On and Multiple Hydrologic Data. *Remote Sensing*, *16*(7), 1176. https://doi.org/10.3390/rs16071176

Rzepecka, Z., & Birylo, M. (2020). Groundwater storage changes derived from GRACE and GLDAS on smaller river basins—A case study in Poland. *Geosciences*, *10*(4), 124. https://doi.org/10.3390/geosciences10040124

Rzepecka, Z., Birylo, M., & Nastula, J. (2016). Assessment of resultant groundwater calculated on the basis of GRACE and GLDAS models. *In 16th International Multidisciplinary Scientific GeoConference SGEM 2016 Conference Proceedings* (pp. 125-132). http://dx.doi.org/ 10.5593/SGEM2016/B22/S09.017

Salam, M., Cheema, M. J. M., Zhang, W., Hussain, S., Khan, A., Bilal, M., ... & Zaman, M. A. (2020). Groundwater storage change estimation using grace satellite data in Indus Basin. *Big data in water resources engineering* (BDWRE), 1, 13-18. http://doi.org/10.26480/bdwre.01.2020.10.15

Seo, J. Y., & Lee, S. I. (2016). Integration of GRACE, ground observation, and land-surface models for groundwater storage variations in South Korea. *International journal of remote sensing*, *37*(24), 5786-5801. https://doi.org/10.1080/01431161.2016.1249301

Singh, L., & Saravanan, S. (2020). Satellite-derived GRACE groundwater storage variation in complex aquifer system in India. *Sustainable Water Resources Management*, *6*(3), 43. https://doi.org/10.1007/s40899-020-00399-3

Su, Y., Guo, B., Zhou, Z., Zhong, Y., & Min, L. (2020). Spatio-temporal variations in groundwater revealed by GRACE and its driving factors in the Huang-Huai-Hai Plain, China. *Sensors*, *20*(3), 922. https://doi.org/10.3390/s20030922

Swenson, S., Famiglietti, J., Basara, J., & Wahr, J. (2008). Estimating profile soil moisture and groundwater variations using GRACE and Oklahoma Mesonet soil moisture data. *Water Resources Research*, *44*(1). https://doi.org/10.1029/2007WR006057

Tariq, A., Ali, S., Basit, I., Jamil, A., Farmonov, N., Khorrami, B., ... & Hatamleh, W. A. (2023). Terrestrial and groundwater storage characteristics and their quantification in the Chitral (Pakistan) and Kabul (Afghanistan) river basins using GRACE/GRACE-FO satellite data. *Groundwater for Sustainable Development*, *23*, 100990. https://doi.org/10.1016/j.gsd.2023.100990

Touré, M. Y., Goïta, K., Magagi, R., & Touré, A. M. (2016, July). Comparison of in situ and grace estimated groundwater in the Canadian prairies. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 7639-7642). IEEE. https://doi.org/10.1109/IGARSS.2016.7730992

Tregoning, P., McClusky, S., Van Dijk, A. I. J. M., Crosbie, R. S., & Peña-Arancibia, J. L. (2012). Assessment of GRACE satellites for groundwater estimation in Australia. *National Water Commission*, Canberra, 82.

Upadhyay, S., Shrestha, S., Loc, H. H., Mohanasundaram, S., Dhungana, S., Lim, S., & Tangdamrongsub, N. (2024). Satellite-based estimates of declining groundwater storage in the transboundary Cambodia-Mekong River Delta Aquifer of the Lower Mekong region, Southeast Asia. *Hydrogeology Journal*, *32*(2), 601-619. https://doi.org/10.1007/s10040-023-02746-y

Van Dijk, A. I., Crosbie, R. S., Peña-Arancibia, J. L., Tregoning, P., & McClusky, S. (2012, July). Analysis of uncertainties in the inference of groundwater dynamics from gravity recovery and climate experiment observations over Australia. *In 2012 IEEE International*

Geoscience and Remote Sensing Symposium (pp. 1318-1320). IEEE. https://doi.org/10.1109/IGARSS.2012.6351295

Wang, S., Cui, G., Li, X., Liu, Y., Li, X., Tong, S., & Zhang, M. (2023). GRACE Satellite-Based Analysis of Spatiotemporal Evolution and Driving Factors of Groundwater Storage in the Black Soil Region of Northeast China. *Remote Sensing*, *15*(3), 704. https://doi.org/10.3390/rs15030704

Xie, X., Xu, C., Wen, Y., & Li, W. (2018). Monitoring groundwater storage changes in the Loess Plateau using GRACE satellite gravity data, hydrological models and coal mining data. *Remote Sensing*, *10*(4), 605. https://doi.org/10.3390/rs10040605

Zhang, C., Duan, Q., Yeh, P. J. F., Pan, Y., Gong, H., Moradkhani, H., ... & Guo, X. (2021b). Sub-regional groundwater storage recovery in North China Plain after the South-to-North water diversion project. *Journal of Hydrology*, *597*, 126156. https://doi.org/10.1016/j.jhydrol.2021.126156

Zhang, J., Liu, K., & Wang, M. (2020). Seasonal and interannual variations in China's groundwater based on GRACE data and multisource hydrological models. *Remote Sensing*, *12*(5), 845. doi: https://doi.org/10.3390/rs12050845

Zheng, L., Pan, Y., Gong, H., Huang, Z., & Zhang, C. (2020). Comparing groundwater storage changes in two main grain producing areas in china: Implications for sustainable agricultural water resources management. *Remote Sensing*, *12*(13), 2151. http://doi.org/10.3390/rs12132151

Zhong, Y., Zhong, M., Feng, W., Zhang, Z., Shen, Y., & Wu, D. (2018). Groundwater depletion in the West Liaohe River Basin, China and its implications revealed by GRACE and in situ measurements. *Remote Sensing*, *10*(4), 493. https://doi.org/10.3390/rs10040493