Mohamed Akl[1], M Akl[2,3], and B F Thomas[2]

[1]Affiliation not available
[2]School of Engineering, Newcastle University
[3]Faculty of Engineering, Tanta University

January 16, 2023

Evaluating the Use of "Goodness-of-Fit" Metrics in GRACE Validation: GRACE Accuracy for Monitoring Groundwater Dynamics.

M. Akl[1,2,] *, and B. F. Thomas[1]

[1] School of Engineering, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK.

[2] Faculty of Engineering, Tanta University, Tanta, Egypt.

* The researcher, Mohamed Akl, is funded by a full scholarship from the Ministry of Higher Education of the Arab Republic of Egypt.

Abstract:

The Gravity Recovery and Climate Experiment (GRACE) satellite has proven to be an excellent tool for monitoring changes in total water storage (TWS), which vertically integrate water storage changes from the land surface to the deepest aquifers. The objective of many GRACE studies is to isolate groundwater storage changes from changes in TWS using independent in-situ, remotely sensed, simulated, or assimilated data to remove other water budget components. Using auxiliary datasets to account for water budget components have revealed large biases and uncertainties, especially over high latitude regions, leading to accumulating errors in GRACE-GW estimates. Comparisons with in-situ groundwater observations permit assessments to evaluate how accurately we can isolate groundwater storage signals from TWSA. Goodness-of-fit (GOF) indices e.g., spearman correlation, mean square error (MSE), Nash-Sutcliffe Efficiency (NSE), and the Kling-Gupta Efficiency (KGE), are commonly applied hydrologic fit metrics that express similarity of time series. Such metrics are used here to compare GRACE-GW estimations and in-situ groundwater observations. The use of GOF indices is constrained by their substantial sampling uncertainty, and controversial interpretation, which may lead to wrong judgement on GRACE-GW estimations. Bias, nonlinearity, and non-normality introduce challenges in our use and interpretation of GOF applied to GRACE-GW time series. The goal of this work is to improve interpretation and use of GOF metrics to validate GRACE-GW estimates, highlighting the importance of assessing multiple GOF criteria beyond simply correlation often applied in GRACE studies. Our results document that poor performance of GOF metrics do not simply translate to inaccurate extraction of GRACE-GW time series but may be attributed to the GOF metric applied. We show that a rigorous assessment of GOF enhances our ability to interpret GRACE-GW change.

M. Akl[1,2], and B. F. Thomas[1,3]

[1] School of Engineering, Newcastle University, Newcastle upon Tyne, United Kingdom. (m.a.a.akl2@newcastle.ac.uk)
[2] Faculty of Engineering, Tanta University, Tanta, Egypt.
[3] Department of Earth Science, University College London, London, United Kingdom.

Evaluating the Use of "Goodness-of-Fit" Metrics in GRACE Validation: GRACE Accuracy for Monitoring Groundwater Dynamics. Abstract number:(H25T-1356)

Newcastle University

AGU FALL MEETING
Chicago, IL & Online Everywhere
12–16 December 2022

Introduction

The objective of many GRACE studies is to isolate groundwater storage changes from TWSA. Using auxiliary datasets to account for water budget components have revealed large biases and uncertainties, leading to accumulating errors in GRACE-GWA. Goodness-of-fit (GOF) indices e.g., spearman correlation, Nash-Sutcliffe Efficiency (NSE), and the Kling-Gupta Efficiency (KGE), are used here to compare GRACE-GWA and in-situ groundwater observations. The use of GOF indices is constrained by their substantial sampling uncertainty, and controversial interpretation, which may lead to unsound judgement on GRACE-GWA . In this work we explore the performance of GOF indices in validation of GRACE-GWA, highlighting the importance of assessing multiple GOF criteria.

Data sets

- Case study conducted on Red-Saskatchewan basin (Fig. 1; basin area = ~673000 km2) over 2003 - 2016.
- Independent datasets (Table 1) used to conduct unique combinations (Table 2) to isolate groundwater changes from GRACE solutions.
- Water-level observations for seven lakes (Fig. 1) were retrieved from The Canadian Water Office.

Water component	n.	Data sets
TWSA	5	CSR-M, JPL-M, CSR-SH, JPL-SH, GFZ-SH
SWEA	6	NOAH 1°, 0.25°, 0.1°, VIC 1°, CLSM 1°, WGHM 0.5°
SMA	8	NOAH 1°, 0.25°, 0.1°, VIC 1°, CLSM 1°, WGHM 0.5°, GlobSnow 0.25°, HGSWE 0.25°
SWA	1	In-situ observations

Table 1: Data sets used in the study.

SWEA \ SMA	NOAH 1°	NOAH 0.25°	NOAH 0.1°	VIC 1°	CLSM 1°	HGSWE 0.25°	WGHM 0.5°	GlobSnow 0.25°
NOAH 1°	1	7	13	19	25	31	37	43
VIC 1°	2	8	14	20	26	32	38	44
CLSM 1°	3	9	15	21	27	33	39	45
NOAH 0.25°	4	10	16	22	28	34	40	46
WGHM 0.50°	5	11	17	23	29	35	41	47
NOAH 0.10°	6	12	18	24	30	36	42	48

Table 2: SWEA and SMA for 48 unique combinations (red-dashed lines show an example for how to detect combination datasets).
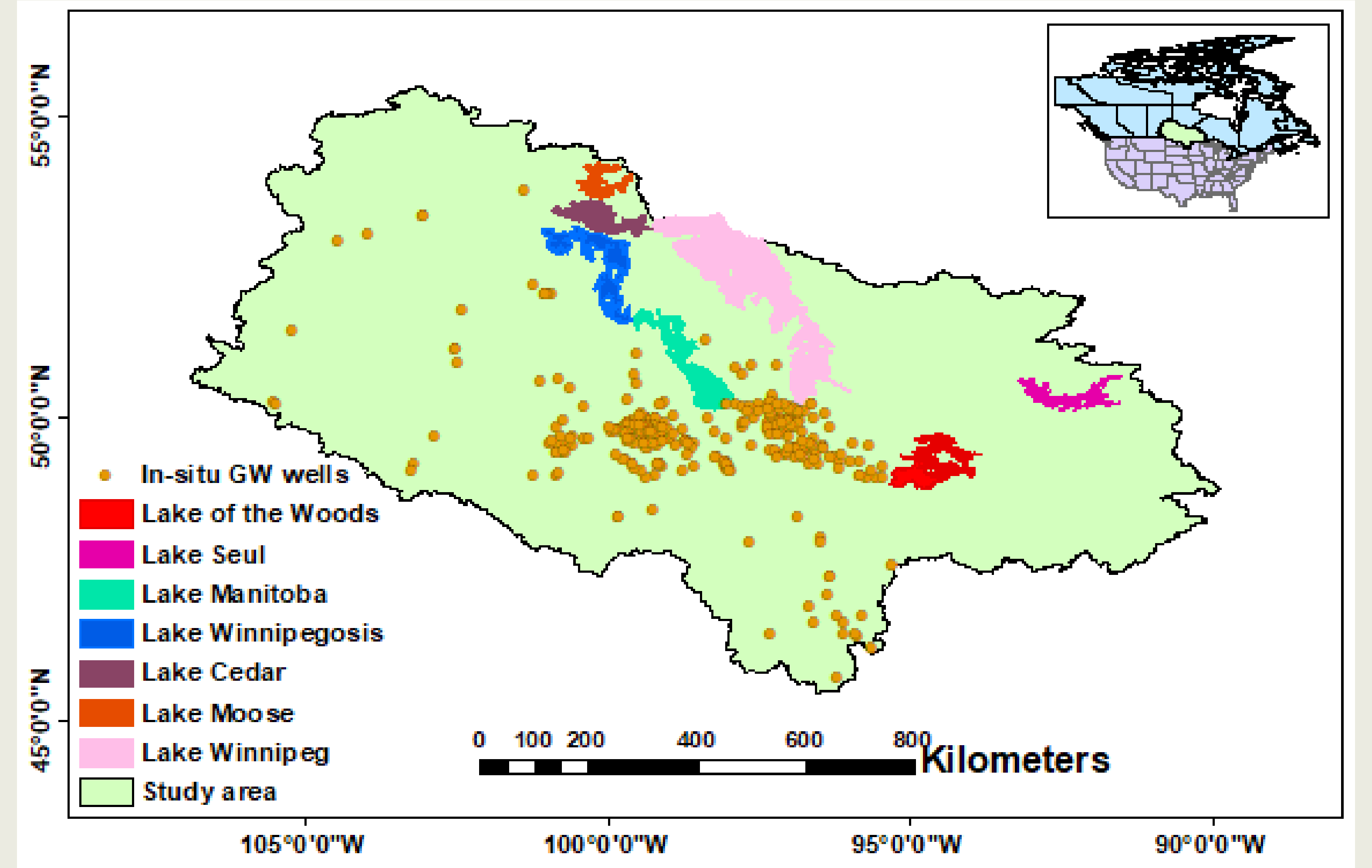

Figure 1: Case study basin: Red-Saskatchewan basin.

Conclusions

- Correlation results suggest that GRACE can accurately observe changes in groundwater storage across Red-Saskatchewan basin.
- Correlation metrics fail to capture dynamic groundwater variability. Spearman correlation requires a monotonic relationship, an assumption violated given the seasonal amplitudes of GRACE-GW. Using a Pearson correlation is ill-advised given serial correlation and non-normality.
- NSE is biased and influenced by skewness and periodicity. GLDAS combinations exhibit moderate skewness in GRACE-GW time series and periodicity since the data is seasonal in nature.
- KGE is based on the assumptions of data linearity and data normality. Non-normality in GRACE-GW time series in addition to skewness from GLDAS combinations violates the implicit assumptions underlying KGE.
- An assessment of multiple GOF indices in GRACE-GWA validation enhances interpretation of GRACE-GW changes. Multiple GOF indices can capture the different characteristics of GRACE-GW time series.

Results

Spearman correlation (Zar, J. H., 1972), which is a non-parametric alternative to Pearson's correlation, evaluates the monotonic relationship between GRACE-GWA and in-situ groundwater, based on the ranked values for each variable rather than the raw data (Fig. 2).
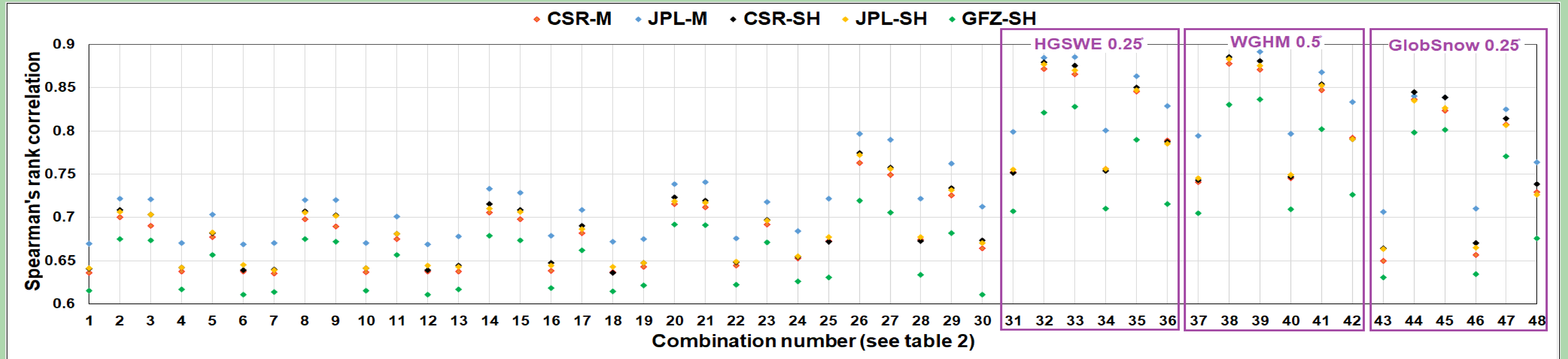

Figure 2: Spearman's rank correlation coefficients between GRACE-GWA estimations and in-situ GWA (P-value in correlation test <0.05 for all combinations).

NSE (Nash and Sutcliffe, 1970) assess the predictive accuracy of GRACE-GWA (Fig. 3). NSE = 1 indicates perfect correspondence between GRACE-GWA and in-situ GWA; NSE = 0 indicates that GRACE-GWA have the same explanatory power as the mean of the in-situ GWA; and NSE <0 indicates that GRACE-GWA is worse than the mean of the observations.
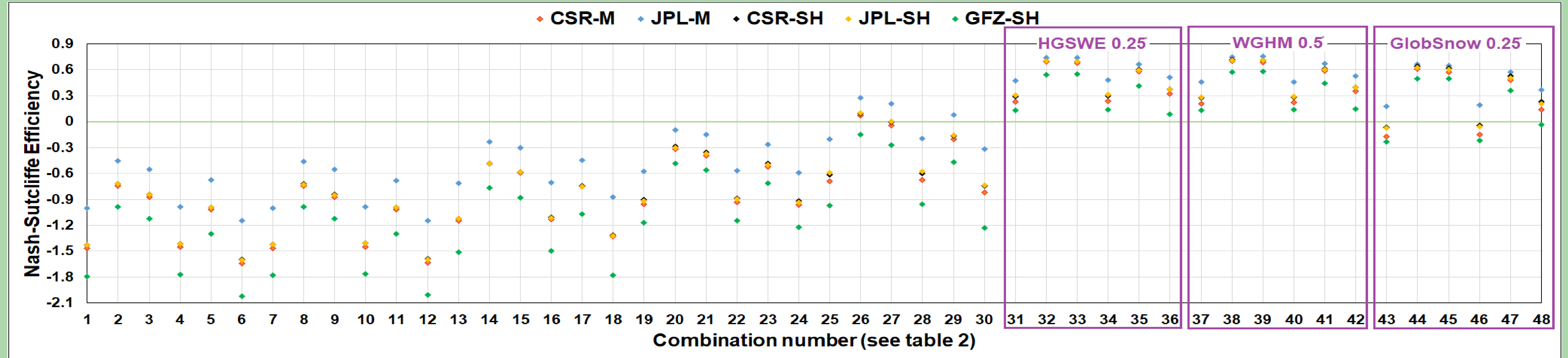

Figure 3: Nash-Sutcliffe Efficiency coefficients between GRACE-GWA estimations and in-situ GWA .

KGE (Gupta et al., 2009) is based on a decomposition of NSE into its constitutive components, including correlation, variability bias and mean bias (Fig. 4). KGE = 1 indicates perfect agreement between GRACE-GWA and in-situ GWA. Poor model performance benchmarks are typically viewed as negative KGE values.
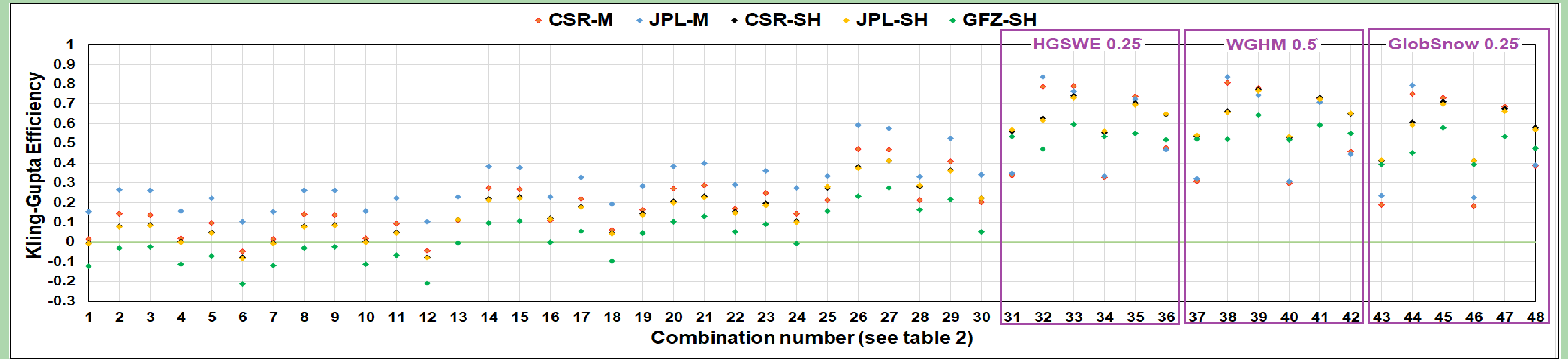

Figure 4: Kling-Gupta Efficiency coefficients between GRACE-GWA estimations and in-situ GWA .

Methodology

<u>Isolating Groundwater Storage Variations from GRACE</u>

- Water budget components (SWEA, SMA, SWA) were converted to anomalies by removing the time series mean for 2004 - 2009 to be consistent with GRACE processing.
- Groundwater storage anomalies (GWA) were extracted from TWSA by removing contributions from other water components, using a water budget equation given as:

$$GWA_t = TWSA_t - (SWEA_t + SMA_t + SWA_t) \qquad (1)$$

<u>In Situ Groundwater Data from Monitoring Wells</u>

- A total of 866 wells were extracted across Red-Saskatchewan basin.
- A well selection algorithm was applied with no more than one missing seasonal observation measurements through each calendar year, identifying 317 quality-controlled monitoring wells (Fig. 1).
- Data gaps were filled using geospatial and temporal gap filling approaches (Zeng and Levy, 1995).
- In-situ GW level anomalies (GWLAs) were converted to GWAs using an effective storage coefficient, with values ranging from 0.01 to 0.05, based on the relationship between GRACE GWA and GWLAs.
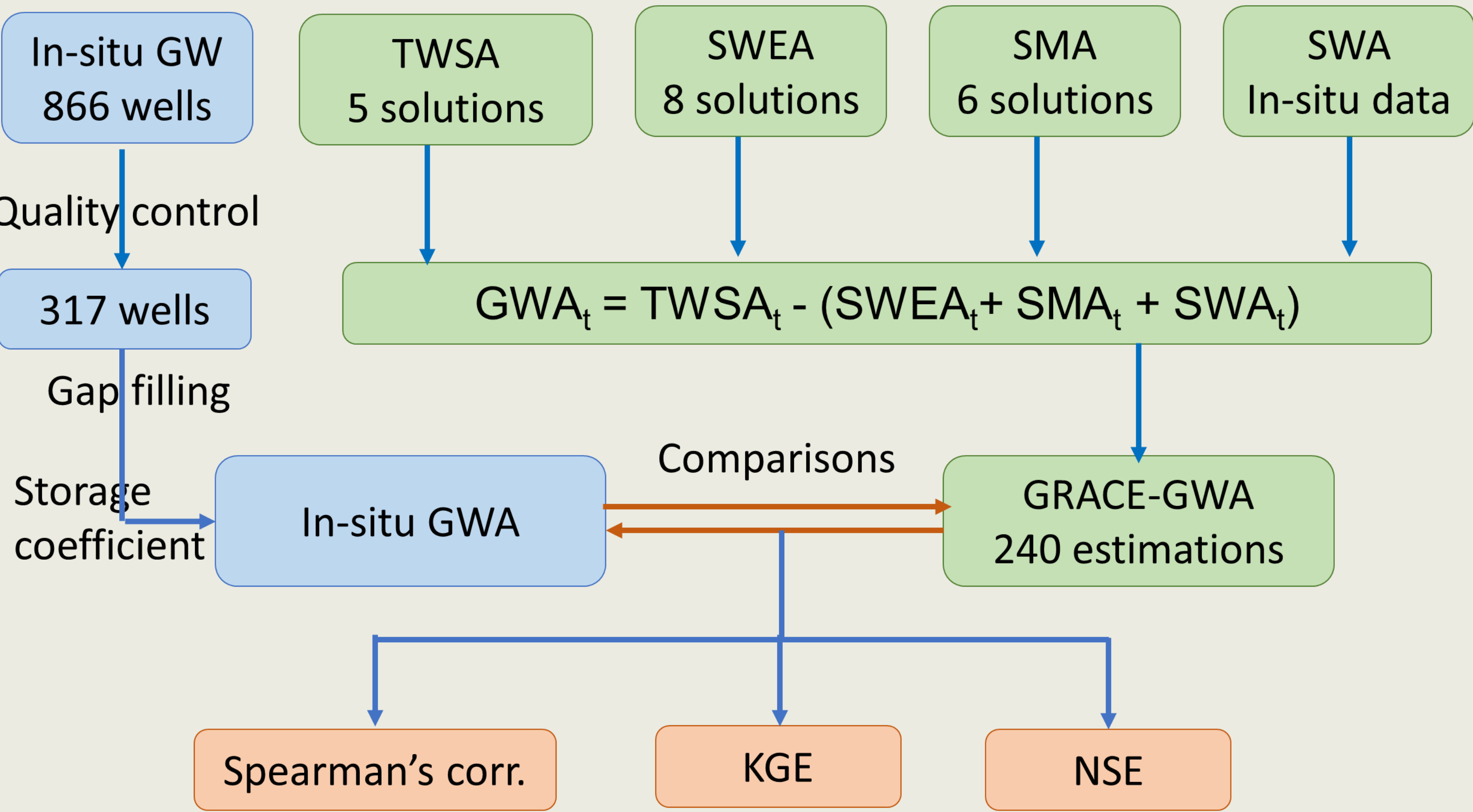

Figure 5: Flowchart describing the analysis conducted in this study.

References

- Gupta, H. V. et al. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of hydrology, 377(1-2)*, 80-91.
- 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.
- 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.
- Zar, J. H. (1972). Significance testing of the Spearman rank correlation coefficient. *Journal of the American Statistical Association, 67(339)*, 578-580.

Acknowledgement