Response to reviewer 1

The interactive modules for simulating water abstraction etc. with the PCR-GLOBWB model are described in greater detail, but these did not be used in this study. Thus the model description should focus on the considered process.

R1: We agree with reviewer and the model description will be modified to be more concise in the revised manuscript.

The model parameterisation with respect to the soil hydraulic properties needs to be better described.

R2: More description of the soil properties will be added to the revised manuscript (please also see R40).

I suggest adding some plots showing the special distribution of simulated TWS for the different DA scenarios.

R3: An illustration of the spatial distribution of both DA scenarios will be shown in Fig. 11. The following discussion will also be added in the revised manuscript **lines 644 – 653**:

"It is also worth discussing the impact of GRACE DA on the spatial pattern of the water storage estimates. To demonstrate this, the update term (ΔA in Eq. (7)) of October 2002 from EnKF 1D and 3D cases is shown in Fig. 11. Only TWSV, SMSV, and GWSV are presented, since other components (snow, surface water, and interception) are small. As discussed above, EnKF 3D shows smaller update in all components. Due to a greater amplitude of GRACE-derived TWSV over northern and southern parts of the region (see Fig. 4), the update is mostly seen there. Almost all update is limited to the soil moisture layer. Higher precipitation is generally observed over the southern part, which leads to higher groundwater recharge (and GWSV) over that region. As such, a GWSV update is clearly seen over the southern part of the region."

In the DA scheme only TWS is considered. It is no clear, how "added" or "subtracted" water was distributed by DA to the different model storages (e.g. SM, GW, snow).

R4: As only TWSV is available from GRACE, TWSV is only used in the discussion in Sect. 6.1. However, the discussion of GRACE DA impact on individual stores are given Sect. 6.2 of the revised manuscript.

Compared to the model results the variations in GRACE determined TWS are much more pronounced. Possible reasons should be discussed in greater (e.g. influence of the pattern restauration procedure).

R5: This discussion will be added to the revised manuscript **lines 509 – 515**:

"It is seen that GRACE-derived TWSV has a greater annual amplitude compared to the model estimated TWSV. This can likely be attributed to the poor quality of the model parameter calibration and the accuracy of the meteorological input data over the data-sparse regions. In the absence of observations, model parameters are difficult to determine and only the best available knowledge (or guess) is generally used, leading to inaccurate model state estimates. Updating the water storage estimates using GRACE DA showed a clear improvement in this case."

It is unclear if at all or how groundwater abstraction was considered in the modelling. If this was considered, why was the groundwater abstraction not considered in the DA (e.g. by updating the groundwater abstraction parameter)?

R6: In this study, the state vector only contains the water storage. Groundwater abstraction is one of PCR-GLOBWB's model parameters, and it is not included in the state vector. Therefore, the groundwater abstraction is not updated or separately estimated in this study, but it is treated. Please also see R45 for more detail.

Title: The term "semi-arid" is not correct (see below)

R7: The "semi-arid" term is used based on Zhu et al. (2015). Please see also R10.

At times TWS variations are simply termed "TWS". This is somewhat confusing. The terms "TWS variations" or short "TWSV" should be always used.

R8: TWS variation will be changed to TWSV in the revised manuscript. Similarly, soil moisture storage variation and groundwater storage variation will be abbreviated as SMSV and GWSV.

L45: The groundwater well data should integrate of smaller areas than the catchment area of the streamflow data. Therefore, I am not convinced that this is a problem of spatial resolution.

R9: Reviewer statement is correct over a sufficiently large river basin. As GRACE spatial resolution is ~250 km or larger, the TWSV signal of the smaller basin can be easily interfered by the neighbouring basin. This is known as a leakage effect and such an effect is seen over the Hexi Corridor. Therefore, the limited spatial resolution of GRACE plays a very important role on the state estimates there.

L57-59: According to the Köppen climate classification this region belongs to "cold desert climate" (BWk).

R10: The "semi-arid" term is used based on Zhu et al. (2015); however, we also realize that much of the region has a cold desert climate, and this can be found in the submitted manuscript (line 152):

"Located next to the Gobi Desert, most parts of the region have a cold desert climate, …"

For clarity, we will include the references of both climate classifications (Zhu et al. (2015) and Peel et al. (2007)) in section 2 of the revised manuscript.

L67-68: This depends largely on the measured variable. For instance, streamwater discharge data provides integrated information for large catchment areas.

R11: We agree with reviewer. In the revised manuscript **lines 68 – 70**, this sentence will be written as follows:

"While streamflow gauges provide integrated information for large catchment areas, point observations of hydrometeorological variables and even groundwater levels can be very local in scope."

L81: In addition, hydrological models typically suffer from inadequate process representations (model structure errors).

R12: The suggested statement will be added to the introduction section of the revised manuscript **lines 85 – 86**.

L98: "jump" of what?

R13: "jump" will be extended to "jump of the water storage estimates" in the revised manuscript **line 101**.

L115: What is the size of the area?

R14: The size of the individual basin varies between 41,600 and 157,000 km^2. This can be found in lines 149 - 151 of the submitted manuscript:

"Shiyang River Basin (41,600 km2), the Heihe River Basin (143,000 km2), the Shule River Basin (157,000 km2), and a Desert Region (152,445 km2)"

L115-118: How do you know (e.g. the watershed area of the Rhine river is much smaller than the Hexi Corridor area)? Can you provide the SNR values for these different areas?

R15: The size of the individual basin of the Hexi Corridor is smaller than the mentioned basins, Mississippi (3,202,230 km2), Rhine (185,000 km2), Mackenzie (1,743,058 km2). The SNR values of the Hexi Corridor is approximately 2.5, compared to Mississippi (SNR ≈ 11), Rhine (SNR ≈ 17), Mackenzie (SNR ≈ 20).

L122: What is the difference between "surface water" and "inundated water"?

R16: The "surface water" in PCR-GLOBWB consists of river/channels, as well as lake and reservoir storages, while the term "inundated water" is conceptualized for the inundated water above the paddy field during the growing season. The terms are clearly described in PCR-GLOBWB literature (see e.g. Wada et al., 2014).

L128-129: In which way are the results validated against remote sensing?

R17: The validation is qualitatively analysed in terms of the correlation coefficient, Nash-Sutcliff coefficient, Root-Mean-Square different (RMSD). The statement will be added to the revised manuscript **lines 131 – 133**.

L147: The term "basin" is not appropriate.

R18: The term "basin" will be changed to "region" in the revised manuscript **line 145**.

L181: "distributed hydrological model"

R19: The term "global hydrological model" will be changed to "global distributed hydrological model" in the revised manuscript **line 179**.

L184-185: Also indicate the temporal resolution of the model.

R20: The statement "… and temporal resolution of 1 day" will be added to the revised manuscript **lines 183 – 184**.

L185-193: It is unclear, how or if at all these interactive modules for simulating water abstraction etc. have been used in this study. Clearly it was not the focus of this study. Thus I suggest removing this section incl. Appendix A.

R21: The section, including Appendix A, will be removed from the manuscript.

L197: Delete "an"

R22: "an" will be removed from the manuscript.

L208: Change "states" into "water storages"

R23: The term "states" will be changed to "water storage components" in the revised manuscript **line 198**.

L219: This is rather a conceptual model.

R24: Reviewer is correct, like many numerical models, it is conceptual in nature.

L230: Explain "complete to the degree and order 60"

R25: The Earth gravity field is generally presented using a set of spherical harmonic coefficients (SHC) to a certain degree and order. The GRACE CSR product is provided the gravity model up to SHC degree and order 60. Therefore, we compute the TWS variation using the SHC complete to the maximum degree and order 60 in this study.

L259: Does this increment correspond to the monthly change in TWS?

R26: The increment is not necessarily (or linearly) corresponding to the filtered TWS change. The increment rather reflects the missing signal that was caused by the filter applied. In other words, the spatial pattern of the restored TWS change (after signal restoration process applied) is not necessarily similar to the filtered TWS change (see Fig. 4a compared to Fig. 4f).

L261: Is this the general uncertainty of GRACE?

R27: Based on the previous GRACE literature (Wahr et al., 2006; Klees et al., 2008; Dahle et al., 2014), GRACE uncertainty averaged-globally is approximately 2 cm.

L263-264: By looking at Fig. 4 this procedure seems to have mainly intensified the already existing pattern. To which extent are the temporal variations in TWS estimates influenced by this procedure?

R28: The signal is generally damped after the filter is used, results in <4 cm of the TWSV amplitude (please see Fig. 4a). The signal restoration process is used to restore the mitigated signal that was caused by the filter applied. The process restores the signal back for each iteration and the TWSV amplitude becomes ~7 cm after 6 iterations (see Fig. 4f). The spatial pattern between Fig. 4a and Fig. 4f is also different (see the contour lines). As the signal restoration process acts differently (e.g., number of iteration) for different month, the temporal variations in TWSV estimates are also influenced by this procedure. Extensive discussion of the signal restoration process can be found in the given reference (e.g., Tangdamrongsub et al., 2016).

L287-289: It is well-know that global precipitation products show considerable uncertainties, which is also indicated by the low NS values. Since in-situ data is available, I suggest to correct the TRMM data product using the approach suggested by Condom et al. (2011).

R29: We agree with reviewer that correcting TRMM using the method proposed by Condom et al. (2011) is a good idea. However, since the in situ data over the Hexi Corridor is very sparse and does not cover all model grid cells, further analysis is needed to investigate the impact of the method on the spatial distribution. Particularly, the impact on higher frequency (daily) of the precipitation data used in this study (compared to monthly of Condom et al. (2001)). Also, there might be a chance of introducing artefacts into the TRMM data in the grid cells if no in situ data is available. Therefore, we do not apply any correction to TRMM data, and use the standard error the product provided to represent the data uncertainty.

L298: Actual or potential ET?

R30: "evapotranspiration" will be changed to "potential evapotranspiration" in the revised manuscript **line 287**.

L327-329: Actually, more appropriate data is available from other gauging stations in the Hexi Corridor for this study (see e.g. Zhang et al., 2015, 2016).

R31: We thank for reviewer's information. However, we only had an access to limited ground observations by the time this study is conducted. More ground observations will be considered in future work.

L307-322: Because of this conversion method any comparison of groundwater storage changes from in-situ and GRACE observations will not be independent. This needs to be discussed in some detail. In addition, in the procedure described in Tangdamrongsub et al. (2015) two parameter were used instead of one. Please comment on this difference.

R32: Due to the fact that the estimated scale factor values are in line with the specific yield from the field observations (please also see R33), the bias of the estimated parameter from our approach can be considered small over the Shiyang River Basin. However, we understand reviewer's concern, and therefore one additional paragraph will be added to the conclusion section **lines 809 – 815** as follows:

"The conversion approach between the groundwater head measurement and groundwater storage is proven feasible over the Shiyang River Basin. The approach delivers comparable ranges of scale factor estimates to the specific yield estimated from the field observation. However, it is noted here that the results of the conducted validation might be over-optimistic, since the well data processed with the adopted conversion procedure are not fully independent of assimilated GRACE data. The specific yield from the field observation must be used when available."

Additionally, the difference between 1 and 2 parameters are only the bias (first parameter, "a" parameter in Tangdamrongsub et al. (2015)) becomes very small (~1e-14) when the TWS variation and head variation are considered. Therefore, Eq. (1,2) of Tangdamrongsub et al. (2015) and Eq. (2,3) in the submitted manuscript provide the same result. However, for consistency, we restore the bias term in the revised manuscript as

$$\Delta \mathrm{GWS}_{(\mathrm{GRACE}-\Delta \mathrm{SM})} + e = b + f \cdot \Delta h \qquad (2)$$

$$\Delta \mathrm{GWS}_{\text{in situ}} = \hat{b} + \hat{f} \cdot \Delta h \qquad (3)$$

L317-318: Please provide a figure with the data and the regression.

R33: The figure of the regression analysis is shown below (Fig. R1). To reduce the redundancy, we do not include Fig. R1 in the manuscript, but instead we will include a discussion of the parameter estimation in the revised manuscript **lines 658 – 667** as follows:

"Yang et al. (2001) showed that the specific yield values obtained from the field measurements over the Shiyang River Basin was between 0.01 and 0.3. Although, the measurement was not conducted at the well stations used in this study, the values obtained can be used as a guidance of the specific yield of the Shiyang River Basin. In this study, the head measurements were converted to storage unit with the approach described in Sect. 4.3.1. The bias term in Eq. (3) was found to be very close to zero, as the variation (mean removed) was used in the regression analysis. The estimated scale factor was 0.23, 0.04, 0.24, 0.25, and 0.32 at W1 – W5, respectively, which was in line with the values obtained from the field measurement."

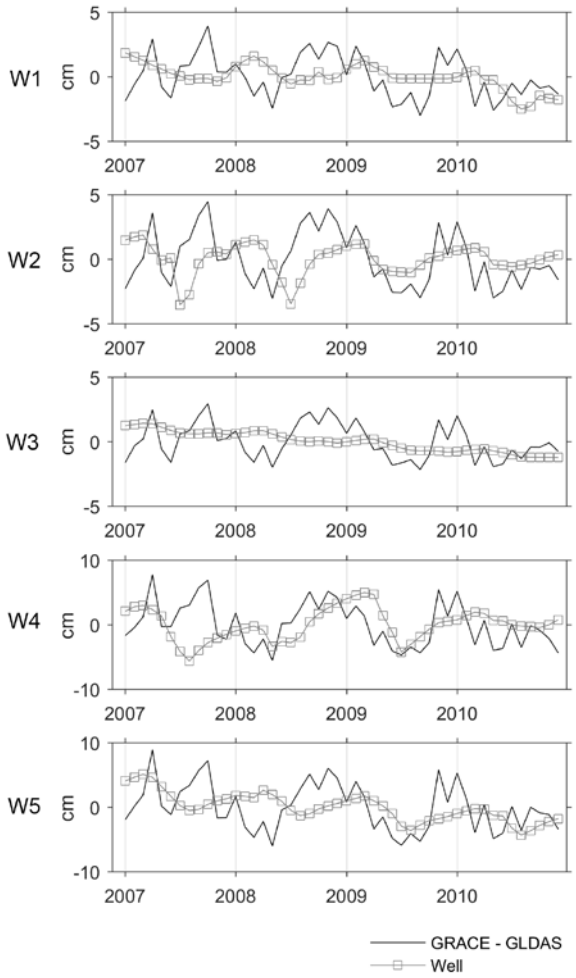


Figure R1. Regression analysis between GRACE-GLDAS and adjusted well measurements in 5 different locations.

L320: Why are you using an averaged f value to calculate the groundwater storage for each well? I would have thought that the variations in parameter f should represent local variations in storage parameters of the aquifers. Please explain the reasoning behind this procedure.

R34: The parameter is individually estimated and used for each well location. No average parameter is used. For clarity, we will extend the statement in **lines 313 – 314** as follows:

"… and ΔGWS(GRACE-ΔSM) at each individual location, a bias (b), a scale factor (f) …"

L451: Please explain how you selected these parameters (e.g. did you use a sensitivity test?).

R35: We selected these parameters based on several previous PCR-GLOBWB studies (e.g. Sutanudjaja et al., 2011, 2014), showing that these selected parameters are indeed the sensitive ones to model simulation results. For clarity, the reference will be added to the revised manuscript.

L526-527: Change into Figure 10

R36: "Fig. 9" will be changed to "Figure 10" in the revised manuscript.

L545: "on" instead of "of"

R37: "of" will be changed to "on" in the revised manuscript.

L549-550: Please provide information on the origin of these parameter values.

R38: The origin of the parameter values is given in Sutanudjaja et al. (2011, 2014), and the reference will be given in the revised manuscript.

Further information related to the origin of parameter values were provided in Appendix A of the submitted manuscript. However, they are removed based on reviewer suggestion (see R21).

The model parameters of PCR-GLOBWB are derived from several globally available datasets that are listed as follows. The Global Land Cover Characteristics Data Base Version 2.0 (GLCC 2.0, http://edc2.usgs.gov/glcc/globe int.php) and and FAO soil maps (1995) were used in order to parameterize the land cover and upper sub-surface properties. For mapping aquifers and estimating the groundwater recession coefficient, the GLobal HYdrogeology MaPS (GLHYMPS) global maps of permeability and porosity (Gleeson et al., 2014), as well as available global digital elevation models (e.g. HydroSHEDS, Lehner et al., 2008) were used. For further explanation about the PCR-GLOBWB model parameterization, the reader is referred to the technical reports (e.g. van Beek and Bierkens, 2009; van Beek, 2008); and other relevant publications (e.g. Sutanudjaja et al., 2011, 2014).

L543: How do you know that the groundwater store of the Desert Region is small.

R39: We realized that the statement is misleading and we change our statement in **lines 552 – 553** as " …the small amplitude of the groundwater variation of this region …". Small GWSV over the Desert Region is presented in Fig. 10k.

L553-554: Please explain in greater detail, why higher values of K_sat and lower values of J have led to a smaller amount of water addition.

R40: We realize that the interpretation the amount of water storage in terms of K_sat and J might be misleading as they do not have a linear relationship. Instead, soil water storage capacity (SC, see Table 1) and forcing data have greater impact on the water storage estimate. Note that greater SC value leads to greater amount of water stored in soil layer, and consequently lesser water percolate to the groundwater store. Therefore, we remove the statement related to K_sat and J, and change the analysis to:

"The impact of GRACE DA on different stores was influenced by both the model parameters and the forcing data. The 4 basins have similar soil water storage capacities (see Table 3), which indicates that the basins can store similar amounts of soil water and generate similar amounts of groundwater recharge under the same rainfall conditions. However, the 4 basins received different amounts of rainfall, which resulted in different SMSV and GWSV estimates. For example, the Shiyang River Basin received the greatest amount of rainfall (~ twice of Heihe River Basin), which led to the greatest amount of the SMSV estimate (~1 cm annual amplitude). Such a large amount was also sufficient to percolate into the groundwater layer, resulting in GWSV of ~0.7 cm (see Fig. 10i and Table 2). In contrast, the Desert Region received approximately 3 times less rainfall, which led to a somewhat smaller amount of SMSV (~0.7 cm annual amplitude) and a much smaller amount of GWSV, ~0.2 cm (see Fig. 10g, k).

The above paragraph will be added to the revised manuscript **lines 557 – 571**.

L599-600: I wonder whether the better agreement with the GRACE DA results is due to (or a least partly due to) the scaling procedure of the piezometer data. Please add a discussion on this.

R41: Due to the fact that the estimated scale factor values are in line with the specific yield from the field observations (please also see R33), the bias of the estimated parameter from our approach can be considered small over the Shiyang River Basin. However, we understand reviewer's concern, and therefore add one additional paragraph into the conclusion of the revised manuscript **lines 809 – 815** as follows:

"The conversion approach between the groundwater head measurement and groundwater storage is proven feasible over the Shiyang River Basin. The scale factor estimates produced with this approach are consistent with the specific yield estimated from the field observations. However, it is noted here that the results of the conducted validation might be over-optimistic, since the well data processed with the adopted conversion procedure are not fully independent of the assimilated GRACE data. The specific yield from the field observation must be used when available."

L642: Clearly, predictions for G2 were improved to a lesser degree.

R42: We agree with reviewer. For clarity, the statement will be changed to "A lesser improvement was observed at G2", **lines 713** in the revised manuscript.

L647-648: These are very low amounts of precipitation, indicating very local precipitation events. It would be interesting to see the spatial distribution of these rainfall events and the resulting modelled soil moisture distribution.

R43: The maps of rainfall and SM storage estimates of the discussed events (September 2007, 2008) are shown below (Fig. R2). However, this is beyond the scope of this study, and therefore Fig. R2 is not presented in the manuscript.

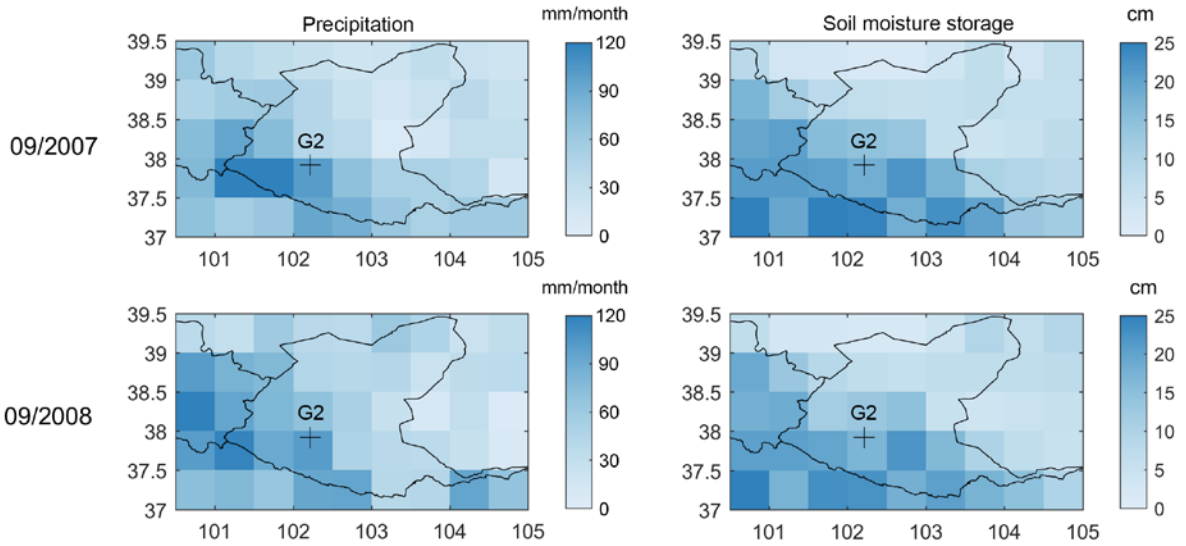


Figure R2. Monthly total precipitation (left) and SM storage estimates of September 2007 and 2008. Stream gauge location G2 is also shown.

L676-678: Why should the SM storage of the Desert Region decrease although precipitation shows an increasing trend? Please discuss.

R44: The discussion will be added to the revised manuscript **lines 749 – 755** as follows:

"In the Desert Region, in contrast to other basins, the minor decreasing TWS trend of -0.1 cm/yr was dominated by loss of SM storage. This was likely caused by inaccurate model parameter calibration over the Desert Region (i.e., too large SC value). Separation of the TWS into groundwater and soil moisture store was likely incorrect. As such, the annual

signal in GWS is much less than in SM there. Therefore, GRACE update was mostly attributed to the SM component, so that a groundwater-pumping signature (Jiao et al., 2015) was seen in the SM instead of the GWS layer."

Further discussion of this (and related) issue is also included in the conclusion, **lines 792 - 793**:

"It should be emphasized that GRACE does not fix a technical problem of the hydrological model, but it rather provides information which is not available otherwise."

L712-714: Until now, there was no indication that groundwater abstraction was considered in the modelling. Please add a description. Why was the groundwater abstraction not considered in the DA?

R45: In this study, the state vector only contains the water storage. As the groundwater abstraction is a parameter of PCR-GLOBWB, it is not included in the state vector. Therefore, the groundwater abstraction is not separately estimated in this study. However, the information of groundwater abstraction is contained in GRACE observation. Once GRACE DA is applied, such information is propagated into the state vector, particularly the groundwater layer. This is clearly seen in the negative trend of updated groundwater estimates. This discussion will be included in the conclusion of the revised manuscript **lines 792 – 797** as follows.

"It should be emphasized that GRACE does not fix a technical problem of the hydrological model, but rather it provides information, which is not available otherwise. Note that, in principle, the model may predict any long-term behaviour of water storage, but that information should be brought in "by hand" (e.g., via the groundwater abstraction parameter). As soon as that information is not available, reliable long-term predictions on the basis of hydrological modelling alone are conceptually impossible."

L734-735: See comment above. Would it be possible to update the groundwater abstraction parameter?

R46: Yes, it is possible to update the model parameter together with the state vector. We will consider reviewer's suggestion in the future work.

L744: Please provide quantitative information on groundwater abstraction.

R47: As the groundwater abstraction is not estimated by our GRACE DA approach, we do not quantify the amount of groundwater abstraction in this study. The groundwater abstraction can be quantified when the parameter is estimated together with the state vector.

References

Dahle, C., Flechtner, F., Gruber, C., König, D., König, R., Michalak, G., and Neumayer, K.-H.: GFZ RL05: An Improved Time-Series of Monthly GRACE Gravity Field Solutions, In Flechtner, F., Sneeuw, N., Schuh, W.-D. (Eds.), Observation of the System Earth from Space - CHAMP, GRACE, GOCE and future missions, (GEOTECHNOLOGIEN Science Report; 20; Advanced Technologies in Earth Sciences), Berlin, Springer, 29-39, http://doi.org/10.1007/978-3-642-32135-1_4, 2014.

Klees, R., Liu, X., Wittwe, T., Gunter, B. C., Revtova, E. A., Tenzer, R., Ditmar, P., Winsemius, H. C., and Savenije, H. H. G.: A Comparison of Global and Regional GRACE

Models for Land Hydrology, Surv. Geophys., 29, 335-359, doi:10.1007/s10712-008-9049-8, 2008.

Peel, M. C., Finlayson, B. L., and McMahon, T. A.: Updated would map of the Köppen-Geiger climate classification, Hydrol. Earth Syst. Sci., 11, 1633–1644, 2007.

Sutanudjaja, E. H., van Beek, L. P. H., de Jong, S. M., van Geer, F. C., and Bierkens, M. F. P.: Large-scale groundwater modeling using global datasets: a test case for the {Rhine-Meuse} basin, Hydrol. Earth Syst. Sci., 15(9), 2913–2935, doi:10.5194/hess-15-2913-2011, 2011.

Sutanudjaja, E. H., van Beek, L. P. H., de Jong, S. M., van Geer, F. C., and Bierkens, M. F. P.: Calibrating a large-extent high-resolution coupled groundwater-land surface model using soil moisture and discharge data. Water Resour. Res., 50, 687–705. doi:10.1002/2013WR013807, 2014.

Tangdamrongsub, N., Ditmar, P. G., Steele-Dunne, S. C., Gunter, B. C., and Sutanudjaja, E. H.: Exploring irregular flood events over Tonlé Sap basin in Cambodia using GRACE and MODIS satellite observations combined with altimetry observation and hydrological models, Remote Sens. Environ., 181, 162 – 173, http://dx.doi.org/10.1016/j.rse.2016.03.030, 2016.

Wada, Y., Wisser, D., and Bierkens, M. F. P.: Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources. Earth System Dynamics, 5, 15–40. doi:10.5194/esd-5-15-2014, 2014.

Wahr, J., Swenson, S., and Velicogna, I.: Accuracy of GRACE mass estimates, Geophys. Res. Lett., 33, L06401, doi:10.1029/2005GL025305, 2006.

Zhu, J. F., Winter, C. L., and Wang Z. J.: Nonlinear effects of locally heterogeneous hydraulic conductivity fields on regional stream–aquifer exchanges, Hydrol. Earth Syst. Sci., 19, 4531–4545, 2015.

Response to reviewer 2

Different GRACE gravity field models are available, CSR (this study, p6, l227ff), GFZ, JPL, CNES/GRGS (Sakumura et al 2014). Why was CSR selected and how are the differences between the different GRACE processing models for the study region. I understand that the focus of the article is on the added valued of the DA, however it would be interesting to see whether GRACE is actually providing added value based on the variability in GRACE processing models.

R1: Comparing to GFZ, JPL, and CNES/GRGS, the CSR product is the only product that provides the error variance covariance matrix of the spherical harmonic coefficients Therefore, it is selected in this study. Note here that the variance covariance matrix is the only information that reflects the true GRACE error. As this information is not available from GFZ, JPL, and CNES/GRGS, they are not considered in this study. We agree with reviewer that it would be interesting to see whether GRACE is consistently improving the water storage estimates based on different products used. The comparison can be conducted as soon as the error information from other data centre is released.

Groundwater head data can be quite complex depending on the well depth and the aquifer being pumped. So far the authors only use head data without information about the aquifer systems. Different aquifer systems also result in individual specific yields. This needs to be addressed. based on a quick literature search hydrogeologic studies (e.g. Ma et al. 2005) are available for the region. Please, do provide information on whether the wells access the same aquifer. Further, groundwater heads were converted to units of storage using a scale factor (p.8, l301ff) as specific yield data were not available. Ma et al. 2005 (and probably more papers as well) provide aquifer properties for the Shiyang basin. Given that the wells are in the same aquifer system, please, show how your units of storage compare to literature values for the region.

R2: We thank reviewer 2 for this valuable information. Unfortunately, the data we used does not come with the aquifer information, so we cannot guarantee whether the well accesses the same aquifer as in Ma et al. (2005). As such, the specific yield is computed based on the best hydrological knowledge (model) and observation. The estimated values are between 0.04 and 0.3, which is in line with the specific yield values Yang et al. (2001) determined from the pumping tests, 0.01 – 0.3. Therefore, our estimate value can be considered sufficiently accurate for the head conversion. For clarity, we add the additional statement to the revised manuscript **lines 658 – 667**:

"Yang et al. (2001) showed that the specific yield values obtained from the field measurements over the Shiyang River Basin was between 0.01 and 0.3. Although the measurements were not collected at the well stations used in this study, the values obtained can be used as a guidance of the specific yield of the Shiyang River Basin. In this study, the head measurements were converted to storage unit with the approach described in Sect. 4.3.1. The bias term in Eq. (3) was found to be very close to zero, as the variation (mean removed) was used in the regression analysis. The estimated scale factor was 0.23, 0.04, 0.24, 0.25, and 0.32 at W1 – W5, respectively, which is in line with the values obtained from the field measurement."

Regarding the precipitation errors the RMS of TRMM was used (p12, l440). As the authors also compared TRMM to station data, was that error included as well?

R3: As the error of other precipitation products are not available, no error is included in the analysis of Sect. 4.2 to avoid the inconsistency of the comparison.

The abstract is a bit too extensive, please, shorten.

R4: The abstract will be shortened in the revised manuscript.

p2, l57-59. Provide reference

R5: References (Gong et al., 2004; Zhu et al., 2015; Cui and Shao, 2005) will be given in the revised manuscript **line 59 and 61**.

Fig. 1. Include all symbols in the figure caption (crosses). Since color is used, the river networks could also be added (1b).

R6: The symbol will be added to Fig. 1 caption of the revised manuscript **lines 1130 – 1131** as " …the locations of considered groundwater wells (x) and river stream gauges (+)." The river network will also be added to Fig. 1b.

p6, l208/209. Please, explain 'the sum of different states'. What are e.g. '4 interception' states?

R7: TWS variation is computed from the sum of 27 different water storage components (layers), which are 8 soil moisture layers, 2 groundwater layers, 4 interception layers, 8 snow layers, 4 inundated top water layers, and 1 surface water layer. For clarity, we revise the statement in the revised manuscript **lines 197 – 199** to:

"… the total water storage (TWS) is computed as the sum of 27 different water storage components: 8 soil moisture layers, 2 groundwater layers, 4 interception layers, 8 snow layers, 4 inundated top water layers, and 1 surface water layer.".

p9, l331ff. What exactly was done with the NDVI values? Was the growing season length determined as the period above and below 0.2? If it was only used for visualization in Fig. 14, the section can be shortened to a couple of lines.

R8: NDVI and GWS variation were analysed together to determine if the growing season was being extended beyond the limited rainy period through groundwater extraction for irrigation. The reviewer is correct in that the growing season length is determined as the period above ~0.2. In the revised manuscript, we remove a few statements in Sect. 4.4.3 to make the section more concise.

Fig. 14a. Is the GW head relative to amsl? What is the depth to the surface?

R9: Yes, the measurement is relative to the mean sea level. For clarity, we will add an additional statement to the revise manuscript **line 295**:

"… form of piezometric heads (relative to the mean sea level), …"

The depth from to the surface is not available from the data provider, and therefore we cannot provide the value here.

References

Cui, Y. and Shao, J.: The Role of Ground Water in Arid/Semiarid Ecosystems, Northwest China, Groundwater, 43 (4), 471–477, doi:10.1111/j.1745-6584.2005.0063.x, 2005.

Gong, D. Y., Shi, P. J., and Wang, J. A.: Daily precipitation changes in the semi-arid region over northern China. J. Arid. Environ., 59 (4), 771–784, doi:10.1016/j.jaridenv.2004.02.006, 2004.

Zhu, J. F., Winter, C. L., and Wang Z. J.: Nonlinear effects of locally heterogeneous hydraulic conductivity fields on regional stream–aquifer exchanges, Hydrol. Earth Syst. Sci., 19, 4531–4545, 2015.

Response to reviewer 3

l. 368 and matrix D in Eq. (7): Burgers et al. (1998) showed that it is necessary to consider the observations as random variable, i.e. that not only an ensemble of predicted model states but also an ensemble of observations has to be considered when calculating the update of each model ensemble member. Perturbations for the observations can be drawn from the error covariance matrix R. Otherwise, the error statistics of the updated model ensemble are underestimated (i.e. not correctly treated). In a correct implementation, matrix D does not contain N identical columns as described in l. 368. This should be fixed or at least discussed by the authors.

R1: We implemented the EnKF as outlined by Evensen (2003). In our formulation, D contains the perturbed observations, i.e. each column is a replicate of the observation but perturbed with ~N(0,R). This was not articulated well in the previous version of the manuscript. The text will be corrected to make this clearer in the revised manuscript **lines 364 – 367** as follows:

"the GRACE observation vector is stored in the matrix D_{mxN}, in which each column is a replicate of the observation but perturbed with random noise ~N(0,R). The analysis equation can be expressed as (Evensen, 2003): "

In addition, it is not possible to draw random errors from the full error covariance matrix of GRACE TWS changes on a 0.5x0.5 degree grid, since the matrix has a rank deficiency. This is a critical issue and should be addressed by the authors as well.

R2: In our study, the error variance-covariance matrix associated with the post-processed GRACE data was used. We did not use the original error matrix since it did not represent the filtered GRACE signal used in our study. In our covariance computation (described in Sect. 5.2.2), the localization function with correlation length similar to the Gaussian smoothing used was applied. Although the main objective of the covariance localization is to reduce the spurious correlation at long distance caused by the limited realization number, the localization also affects the correlation at short distance, and a strong correlation at a short distance becomes slightly weaker. As a result, the error variance-covariance matrix derived based on our method has a full rank. Applying localization also improved the condition number of the covariance matrix, e.g., from $\sim 10^{14}$ to $\sim 10^{2}$ found in our study. Similar to Eicker et al. (2014), the matrix rank and condition number were determined using Matlab functions rank and cond, respectively. We thank reviewer for the advice. The clarification regarding rank deficiency will be included in the revised manuscript **lines 433 – 436**.

l. 507-508: The standard deviations of the EnKF results are however underestimated, since the observation vector was not treated as a random variable in Eq. (7). Therefore, the error statistics of the updated model states are not correct. This should be fixed or at least discussed.

R3: Please see R1

l. 588-589: This might change after correctly estimating the updated model ensemble spread by generating perturbations for the observations (revising Eq. (7)).

R4: Please see R1

Eq. (8): Since both error covariance matrices (from the model and the observations) have a rank-defect due to (1) the fact that usually the number of model states is much larger than the number of model ensemble members and (2) GRACE cannot actually resolve TWS changes on a 0.5x0.5 degree grid, the inverse in Eq. (8) does not exist. This should be pointed out and a reference to sections 5.2.1 and 5.2.2 might be provided that describe how the authors deal with this issue.

R5: Please see R2

l. 251: GRACE observations are highly correlated on such a fine spatial resolution (similar to the above comment). Did the authors investigate this? Was this the reason to use a maximum correlation length for the observation error covariance matrix?

R6: Reviewer is correct. In our covariance computation (described in Sect. 5.2.2), the localization function with correlation length similar to the Gaussian smoothing used was applied. The localization helps to improve the matrix stability and we investigated this by checking the rank and condition number of the matrix as explained in R2.

l. 414-415: If I understand it correctly, the error correlation length is set to 250 km and TWS changes outside of this radius are assumed to not be correlated to the center grid cell. Is this reasonable? It would be helpful to investigate the correlations of points with longer distances to verify this choice. Does the "local" error covariance matrix have a full rank?

R7: As the observation error variance-covariance matrix is derived based on the application of 250 km filter radius, the correlation error at distance beyond 250 km (correlation length) does not have a crucial impact on the result. In the submitted manuscript, we demonstrate the error characteristic in Fig. 8b. From the figure, the correlation reduces significantly beyond the correlation length. Additionally, the error variance-covariance matrix derived based on our method has a full rank.

Fig. 7: In the main text (l. 414-415), it is explained that a correlation length of 250 km is used (approx. four to five 0.5x0.5 degree (~50kmx50km at the equator) grid cells in each direction from the center grid cell). In Fig. 7, it is shown that only the neighboring grid cells are considered. Please clarify.

R8: Reviewer is correct. We realized that the figure caption was not explained clearly. To clarify this, we add an additional description in the figure caption **lines 1158 – 1160** as follows:

"The graphic demonstrates the case of 1 pixel (0.5 degree) correlation distance. The boundary stretches farther for larger correlation distance."

l. 419: Since the neighboring 0.5x0.5 degree grid cells are highly correlated, it is not reasonable - based on the GRACE error characteristics - to apply the EnKF without spatial error correlations on such a fine scale. A statement would be helpful to the reader.

R9: We thank reviewer for the recommendation, the statement will be added to the revised manuscript lines 399 – 400 as follows:

"Spatial correlations of model errors and observation errors were also taken into account in view of the fact that the latter are highly correlated at neighbouring 0.5^{o}x0.5^{o} grid cells."

l. 726-727: But: The authors do not use the full error covariance matrix as directly calculated from the observations. Instead a maximum correlation length of 250 km is

assumed, and thus a part of the information within the full error covariance matrix is neglected. Therefore, the statement might be misleading.

R10: We thank for reviewer comment. To clarify this, we will modify the statement in the revised manuscript **lines 830 – 832** as follows:

"…this is a reasonable price to pay as deriving the error variance-covariance matrix from the full (and only full) error covariance matrix noticeably improves the results of GRACE data assimilation."

l. 90-91: That seems to be incorrect. Zaitchik et al. (2008) used an ensemble Kalman smoother (EnKS) approach to partition the monthly update increment (based on comparing monthly means of modeled and observed TWS changes) equally to each day of the month. GRACE TWS changes are only assimilated once per month and not every 10 days.

R11: We thank for reviewer comment. The statement will be corrected in the revised manuscript **lines 197 – 199** as follows:

"…using a monthly observation value and distributing the update as daily increments (Zaitchik et al., 2008; Forman et al., 2012; Girotto et al. 2016)."

l. 95: This work adapts the method as proposed in Zaitchik et al. (2008) to a snow-dominated basin.

R12: Please see R11.

l. 98: Please also consider the disadvantage of computational costs: The method has some computational drawback since the model has to be evaluated twice over the same month.

R13: We thank for reviewer suggestion. The additional sentence will be added in the revised manuscript **lines 102 – 103** as follows:

"The only price to pay is the additional computational cost of running the model twice for the same month."

l. 106: In Forman et al. (2013), the authors did not use correlated errors for the data assimilation. They investigated for which spatial resolution errors of GRACE TWS changes might be considered as uncorrelated. According to these investigations, they assumed white noise for (sub-)basin averaged TWS changes from GRACE.

R14: We agree with reviewer. Forman et al. (2013) will be removed in this context to avoid the confusion.

l. 89-95: In this work, the authors performed an analysis of introducing the update increments completely at the beginning of a month, the end of a month or equally distributed over all days of a month. This is worth to be mentioned along with the other citations.

R15: We thank for reviewer suggestion. Girotto et al. (2016) will be cited in the revised manuscript.

l. 39-40 and l. 106-108: A first analysis of assessing the effect of considering or neglecting spatial error correlations of GRACE TWS changes was performed in Schumacher et al. (2016) in form of a synthetic experiment, for which one of the authors of this HESSD manuscript was the editor and should therefore be very familiar with the work. It seems that the paper is methodologically the closest related to the analysis presented here and,

therefore, should be cited and discussed. Findings should be compared to the findings in the published paper.

R16: At the time this study was conducted, Schumacher et al. (2016) was not published, therefore we conducted the analysis independently based on our method (proposed in this HESSD paper). However, we thank reviewer for the recommendation, and Schumacher et al. (2016) will be cited in the revised manuscript.

l. 577: This was also seen and discussed in Schumacher et al. (2016). The authors should compare their results with the findings in this paper, since the objective of both papers is to understand the effect of considering spatial error correlations of GRACE TWS changes on hydrological data assimilation results.

R17: We thank for reviewer suggestion. The suggestion will be considered in the revised manuscript.

l. 715-718: The authors should add something like "in agreement with the recommendation in Schumacher et al. (2016)."

R18: We thank for reviewer suggestion. The statement will be given in the revised manuscript **lines 823 – 824**.

l. 719-724: The findings in the HESSD manuscript allow for a clearer conclusion on improvements when error correlations of GRACE TWS changes are taken into account. What might be the reason for this? Differences in the study set up? Localization of model / observation error covariance matrices?

R19: The improvement is mainly due to a better representation of GRACE information in the EnKF. Ignoring error correlations in the DA led to an over-fit of the results to the observations, which led to less accurate state estimates. These explanation will be presented in the revised manuscript **lines 824 – 828** as follows:

"We explain this finding by the fact that GRACE errors at the neighbouring 0.5^{o}x0.5^{o} grid cells are highly correlated. As such, the simultaneous consideration of GRACE data at multiple neighbouring cells does not reduce data noise, as it would be the case if noise were white. In other words, the white-noise assumption may severely overestimate the information content of GRACE data."

l. 729: A reference to Schumacher et al. (2016) would strengthen this statement, since the HESSD manuscript is not the only study that concludes a benefit / more realistic GRACE data assimilation approach if implementing GRACE error correlations.

R20: We thank for reviewer for the suggestion, Schumacher et al. (2016) will be cited in the relevant context.

l. 752-753: Schumacher et al. (2016) should be added to the list of references.

R21: Schumacher et al. (2016) will be added to the list of references.

l. 755: Alternative methods have been investigated in Schumacher et al. (2016), namely a square root analysis scheme (SQRA) and the singular evolutive interpolated Kalman filter (SEIK). Especially the application of the SEIK filter showed promising results. A citation would support the authors expectation that alternative methods, e.g. the particle filter, would improve the data assimilation performance.

R22: We thank the review for the suggestion. We will consider this in the revision.

l. 583: "truth", i.e. to the independent measurements of individual water compartments. These measurements are also subject to uncertainties and not "true" values.

R23: To avoid the confusion, the statement will be changed to “Validating against the in situ groundwater and streamflow data will quantitatively reveal the performance of each approach”. This is given in lines 630 – 631 of the revised manuscript.

l. 756: "true" -> better "full" (true is difficult since often unknown / poorly known)

R24: “true” will be changed to “realistic”.

Improving estimates of water resources in a semi-arid region by assimilating GRACE data into the PCR-GLOBWB hydrological model

N. Tangdamrongsub[1,2], S. C. Steele-Dunne[3], B. C. Gunter[1,4], P. G. Ditmar[1], E. H. Sutanudjaja[5], Y. Sun[1], T. Xia[6], and Z. Wang[6,7]

[1] Department of Geoscience and Remote Sensing, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands
[2] School of Engineering, Faculty of Engineering and Built Environment, The University of Newcastle, Callaghan, New South Wales, Australia
[3] Department of Water Resources, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands
[4] School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, The United States of America
[5] Department of Physical Geography, Faculty of Geosciences, Utrecht University, Utrecht, The Netherlands
[6] Department of Hydraulic Engineering, Tsinghua University, Beijing 100084, China
[7] State Key Lab of Hydroscience and Engineering, Tsinghua University, Beijing 100084, China

Correspondence to: N. Tangdamrongsub (Natthachet.tangdamrongsub@newcastle.edu.au)

Abstract

An accurate estimation of water resources dynamics is crucial for proper management of both agriculture and the local ecology, particularly in semi-arid regions. Imperfections in model physics, uncertainties in model land parameters and meteorological data, as well as the human impact on land changes often limit the accuracy of hydrological models in estimating water storages. To mitigate this problem, this study investigated the assimilation of Terrestrial Water Storage Variation (TWSV) estimates derived from the Gravity Recovery And Climate Experiment (GRACE) data using an Ensemble Kalman Filter (EnKF) approach. The region considered was the Hexi Corridor in Northern China. The hydrological model used for the analysis was PCR-GLOBWB, driven by satellite-based forcing data from April 2002 to December 2010. The impact of the GRACE Data Assimilation (DA) scheme was evaluated in terms of the TWSV, as well as the variation of individual hydrological storage estimates. The capability of GRACE DA to adjust the storage level was apparent not only for the entire TWSV but also for the groundwater component. In this study, spatially-correlated errors in GRACE data were taken into account, utilizing the full error variance-covariance matrices provided as a part of the GRACE data product. The benefits of this approach were demonstrated by comparing the EnKF results obtained with and without taking into account error correlations. The results were validated against in situ groundwater data from 5 well sites. On average, the experiments showed that GRACE DA improved the accuracy of groundwater storage estimates by as much as 25%. The inclusion of error correlations provided an equal or greater improvement in the estimates. In contrast, a validation against in situ streamflow data from two river gauges showed no significant benefits of GRACE DA.

This is likely due to the limited spatial and temporal resolution of GRACE observations. Finally, results of the GRACE DA study were used to assess the status of water resources over the Hexi Corridor over the considered 9-year time interval. Areally-averaged values revealed that TWS, soil moisture, and groundwater storages over the region decreased with an average rate of approximately 0.2, 0.1, and 0.1 cm/yr in terms of equivalent water heights, respectively. A particularly rapid decline in TWS (approximately –0.4 cm/yr) was seen over the Shiyang River Basin located in the southeaster part of Hexi Corridor. The reduction mostly occurred in the groundwater layer. An investigation of the relationship between water resources and agricultural activities suggested that groundwater consumption required to maintain crop yield in the growing season for this specific basin was likely the cause of the groundwater depletion.

1. Introduction

The focus of this study is the Hexi Corridor. It is a semi-arid region located between the Gansu province of China and Mongolia (Fig. 1). A semi-arid region can be broadly classified as an area on the boundary of a larger desert, receiving just enough annual precipitation (300 mm or less) to sustain a limited amount of agriculture (Gong et al., 2004; Zhu et al., 2015). Inefficient use of the limited amount of surface water can often lead to overuse of groundwater resources and salinization of the soil (Cui and Shao, 2005). This can result in desertification, which not only reduces the amount of production but also may have long-term effects on the local ecology. All of this holds true for the Hexi Corridor (Wang et al., 2003).

Improving the water resources management of semi-arid regions requires accurate knowledge of the hydrological processes involved. For small areas, this can be partially obtained through a network of in-situ measurement systems, such as meteorological stations, river gauges, groundwater wells, evaporation trays, etc. (Dahlgren & Possling, 2007; Huo et al., 2007; Kang et al., 2004; Ma et al., 2005; Du et al., 2014). While streamflow gauges provide integrated information for large catchment areas, point observations of hydrometeorological variables and even groundwater levels can be very local in scope. A sensor at a point several kilometres away may record significantly different values. For large scales (> 10,000 km^2), such techniques are unlikely capable of delivering accurate results.

Two options for estimating the large-scale Terrestrial Water Storage Variation (TWSV) of a particular region are using observations from the Gravity Recovery And Climate Experiment satellite mission (GRACE, Tapley et al., 2004) or utilizing a regional or global hydrological model. A number of prior studies have reported on the potential of GRACE in the estimation of snow water equivalent (Niu et al., 2007), groundwater (Döll et al., 2014), and evapotranspiration (Long et al., 2014) in terms of temporal and spatial variability. However, GRACE only provides the total column of the water storage at a monthly time scale and large spatial scales (> 300 km). It is not possible to identify the contribution of separate hydrological components to the TWSV from GRACE data alone. On the other hand, a hydrological model can be used to estimate the individual storage components at very high spatial and temporal scales. The major drawback of the model approach is mainly the

significant uncertainties influenced by the quality of the model parameter calibration and the accuracy of the meteorological input data. In addition, hydrological models may suffer from inadequate process representations (model structure errors).

Data Assimilation (DA) can be employed to combine the strengths of GRACE and hydrological models while mitigating their respective weaknesses. A number of studies have shown that GRACE DA can be used to improve the estimation of groundwater and streamflow (Zaitchik et al., 2008; Tangdamrongsub et al., 2015), snow water equivalent (Forman et al., 2012; Su et al., 2012), and as well as for evaluation of drought events (Houborg et al., 2012; Li et al., 2012). Different temporal and spatial resolution of GRACE observations and hydrological models require proper design of the DA scheme. Several DA schemes have been developed to distribute GRACE observations into the model, which include using 5-day interpolated observations and updating the model every 5 days (Tangdamrongsub et al., 2015); using a monthly observation value and applying the model update only at the end of the month (Eicker et al., 2014); using a monthly value and distributing the update as a daily increments (Zaitchik et al., 2008; Forman et al., 2012; Girotto et al. 2016). Although all DA schemes are acceptable, the scheme proposed by Forman et al. (2012) is advantageous because it does not require an interpolation of the observations and can reduce the spurious jump of the water storage estimates caused by applying the update at the end of the month only. The only price to pay is the additional computational cost of running the model twice for the same month. A scheme similar to (Forman et al., 2012) is used in this study. Spatial disaggregation is also needed to reconcile the difference in horizontal resolution between the observations and the model. Recent studies by Eicker et al. (2014) and Schumacher et al. (2016) suggested including the GRACE variance-covariance error information in the spatial disaggregation step. Both studies proposed using 500-km GRACE spatial resolution to mitigate the ill-posedeness of the error covariance matrices in the spatial domain. In line with Eicker et al. (2014) and Schumacher et al. (2016), the assimilation scheme in this study accounts for spatially correlated errors by using full error variance-covariance matrices of GRACE data. This study will show that considering the GRACE error correlations leads to an improvement of the state estimates. Particularly, the Signal-to-Noise Ratio (SNR) of the TWSV is much lower than in the river basins considered in the previous studies, e.g., Mississippi (Zaitchik et al., 2008), Rhine (Tangdamrongsub et al., 2015), and Mackenzie (Forman et al., 2012).

Approximately 9 years of GRACE data – between April 2002 and December 2010 – are considered in this study. GRACE observations are assimilated into the PCRaster Global Water Balance (PCR-GLOBWB; Van Beek et al., 2011; Sutanudjaja et al., 2014; Wada et al., 2014) hydrological model over the Hexi Corridor. TWS is computed from PCR-GLOBWB as the sum of all the hydrological components (soil moisture, groundwater, surface water, inundated water, interception, and snow). The previous studies showed very good agreement of PCR-GLOBWB based estimates with GRACE observations in several river basins (Wada et al., 2014; Tangdamrongsub et al., 2016). However, the performance of PCR-GLOBWB has not yet been evaluated over the Hexi Corridor. In addition, to date the model has not been incorporated into any GRACE DA scheme, making this study the first attempt to do so.

Investigating the added value of GRACE DA in the Hexi Corridor is the main objective of this study.

First of all, the impact of GRACE DA and the effect of taking correlations in GRACE errors into account are assessed. Both the total terrestrial water storage and the individual hydrological storage compartments are considered.

Next, the results of the GRACE DA are validated with independent in-situ data. The agreement is analysed in terms of the correlation coefficient, Nash-Sutcliff coefficient, and Root-Mean-Square difference (RMSD). The groundwater storage variation (GWSV) and streamflow estimates after GRACE DA are validated with the well and river stream gauge measurements, respectively.

Finally, results from this GRACE DA study are used to assess the status of water resources over the Hexi Corridor. The connections between the water storage (including groundwater consumption) and agriculture in the area are also presented and discussed. At that stage, we use precipitation data from the Tropical Rainfall Measuring Mission (TRMM; Huffman et al., 2007) and the Moderate Resolution Imaging Spectroradiometer (MODIS) derived Normalized Difference Vegetation Index (NDVI; Huete et al., 2002).

2. Study region

The Hexi Corridor is a long and narrow area between the Qilian Mountain range and southern Mongolia (Fig. 1a). The region's elevation ranges from 5,200 m in the southern upstream area (Qilian Mountains) to 900 m in the northern downstream zone (Inner Mongolia) (Fig. 1b). The region is comprised of four typical inland arid and semi-arid regions (Zhu et al., 2015): the Shiyang River Basin (41,600 km^2), the Heihe River Basin (143,000 km^2), the Shule River Basin (157,000 km^2), and a Desert Region (152,445 km^2) (Geng and Wardlaw, 2013; Zhu et al., 2015). Located next to the Gobi Desert, most parts of the region have a cold desert climate (Peel et al., 2007), where precipitation is relatively low to sustain vegetation or crops. Approximately 60 to 80 % of the annual rainfall is concentrated during the timeframe from June to September. The inland rivers mainly originate from the Qilian Mountains and disappear after entering the midstream/downstream plains and oases. As such, the southern part of the region is more favourable for agriculture.

The four basins have distinct characteristics. First, the smallest river basin, Shiyang, has 8 main river streams, including the Xida and Xiying Rivers (Fig. 1c). The annual rainfall and the mean temperature are approximately 250 mm and 5 °C (Fig. 2a, b), respectively. The Shiyang River Basin is considered the wettest basin compared to the others, with relatively high mean total renewable annual water resources of approximately 1.66 billion m^3 (Zheng et al., 2013). However, a highly developed economy and population growth in the past decade have resulted in a severe water resources overexploitation problem (Zheng et al., 2013). The Heihe River Basin has a semi-arid climate and the mean daily temperature of ~6 °C (Fig. 2d). The average annual rainfall is ~150 mm (Fig. 2c) with high heterogeneity both in temporal

and spatial distribution. The mean total annual available water resources are estimated at 3.7 billion m^3 (Hu, 2015). Similar to the Shiyang River Basin, increased water exploitation, increasing population, and changing climate have aggravated the damage to the downstream ecology. The Shule River Basin has an arid climate, the mean temperature there is around 4 °C (Fig. 2f), and the average annual rainfall is only approximately 98 mm (Fig. 2e). Compared to the Shiyang River Basin, the Shule River Basin is approximately four times as large in terms of surface area, but has similar mean total annual water resources, ~1.6 billion m^3 (Hu, 2015). The district irrigation areas are mainly located in the middle of the Shule River Basin. Agricultural water consumption accounts for more than 80% of the total water use. Finally, the Desert Region has an extreme continental desert climate with an average temperature of 8 °C, and the annual rainfall of ~130 mm. Extensive groundwater abstraction was also observed over the region (Jiao et al., 2015).

3. Hydrology model

The global distributed hydrological model PCR-GLOBWB (van Beek et al., 2011; Sutanudjaja et al., 2016) simulates spatial and temporal continuous fields of fluxes and storages in various water storage components (soil moisture, groundwater, surface water, inundated water, interception, and snow). The model version used here (Sutanudjaja et al., 2016) has a spatial resolution of 30 arc minutes (approximately 50 km at the equator), and a temporal resolution of 1 day. Figure 3 illustrates the structure of PCR-GLOBWB model. The model includes 2 soil layers (SM_{upp}, SM_{low}), an underlying hydrologically active and replenishable groundwater (GWS_{active}) layer, a non-renewable groundwater (GWS_{fossil}) layer, as well as interception, surface water, and snow stores. The non-renewable groundwater is available for abstraction to satisfy water demands once the overlying hydrologically active groundwater storage is depleted. For soil, snow, inundated top water, and interception stores, an individual grid cell is divided into sub-grids associated with different types of topography, vegetation phenology, and soil properties, as well as land cover types. Specifically, there are 4 types of land covers defined: short natural vegetation, tall natural vegetation, irrigated non-paddy field, and irrigated paddy field. Soil components include the upper layer (SM_{upp}, 0 – 30 cm) and the lower layer (SM_{low}, 30 – 150 cm). The snow component includes snow water equivalent (SWE), as well as snow free water (SFW) representing the storage of melted snow. The water stored in the stream channels and lakes is also included in the TWS estimate. Based on the structure of PCR-GLOBWB, the total water storage (TWS) is computed as the sum of 27 different water storage components: 8 soil moisture layers, 2 groundwater layers, 4 interception layers, 8 snow layers, 4 inundated top water layers, and 1 surface water layer.

For each grid cell and for each daily time step, the model determines the water balance in two vertically stacked soil layers and the groundwater store. The model also computes the vertical water exchanges between the soil layers and between the inundated top water layer and the atmosphere, i.e. rainfall and snowmelt, percolation and capillary rise, as well as evaporation and transpiration fluxes. The active groundwater store underlies the soil, is fed by net groundwater recharge, discharges to baseflow as a linear reservoir, and is exempt from the

direct influence of evaporation and transpiration fluxes. However, capillary rise from the active groundwater store can occur depending on the simulated groundwater storage, the soil moisture deficit, and the unsaturated hydraulic conductivity. Fluxes are simulated according to the different land cover types. The model includes a physically-based scheme for infiltration and runoff, resulting in the direct runoff, interflow, as well as groundwater baseflow and recharge. River discharge is calculated by accumulating and routing the specific runoff along the drainage network. For further details, including model parameterization, the reader is referred to the technical reports and other relevant publications (van Beek and Bierkens, 2009; van Beek, 2008; Sutanudjaja et al., 2011, 2014).

4. Data and data processing

4.1 GRACE data

The GRACE gravity product release 5 (RL05), generated by the University of Texas at Austin's Center for Space Research (CSR, Bettadpur, 2012), was used as input. The product consists of monthly sets of spherical harmonic coefficients (SHC) complete to degree and order 60. On this basis, TWSVs were obtained for the study period between April 2002 and December 2010. The GRACE data were further processed in this study as follows:

- SHCs of degree 1 provided by Swenson et al. (2008) were restored, and all 5 coefficients of degree 2 were replaced by the values estimated from satellite laser ranging (Cheng and Tapley, 2004).
- SHC variations were computed by removing the long-term mean (computed between April 2002 and December 2010) from each monthly solution.
- A destriping filter (Swenson and Wahr, 2006) was applied to the SHC variations. The filter used a 5th degree polynomial (Savitsky-Golay) over a 5-point window to remove the correlations; orders below 8 remained unchanged.
- An additional 250-km radius Gaussian smoothing (Jekeli, 1981) was applied to SHC variations to suppress high-frequency noise, and the TWS variations ($\Delta\sigma$ [m]) were then computed using (Wahr et al 1998)

$$\Delta\sigma(\theta,\phi) = \sum_{l=1}^{60}\sum_{m=-l}^{l} W_l \overbrace{\frac{a_e(2l+1)}{3(1+k_l)}\frac{\rho_e}{\rho_w}}^{S_l} \Delta\bar{C}_{lm}\hat{Y}_{lm}(\theta,\phi), \quad (1)$$

where θ, ϕ are co-latitude and longitude in spherical coordinates, $\Delta\bar{C}_{lm}$ is the SHC variations of degree l and order m, $\hat{Y}_{lm}$ is the normalized surface spherical harmonic, W_l is the Gaussian smoothing function, S_l is a scaling factor used to convert dimensionless coefficients to TWS in terms of Equivalent Water Heights (EWH), a_e is the semi-major axis of the reference ellipsoid, k_l is the load love number of degree l, ρ_e and ρ_w are the average density of the Earth and water, respectively. In this study, the TWS variations were computed at every 0.5°x0.5° grid cell. This cell size was selected through trial and error as a balance between performance and resolution.

In general, filters suppress not only noise but also the genuine TWSV signal, and are a well-known source of signal leakage. To address this, a signal restoration method (Chen et al., 2014; Tangdamrongsub et al., 2016) was employed. The method iteratively determined the possible signal reduction caused by the Gaussian filter applied and added it back to the filtered signals. The errors of the procedure grew with the number of iterations, requiring a proper selection of the convergence criterion. In this study, the criterion was chosen empirically: the signal restoration process was iteratively repeated until the increment in every grid cell inside the Hexi Corridor became smaller than 0.5 cm. This value is 2-3 times smaller than the GRACE uncertainty (Wahr et al., 2006; Klees et al., 2008; Dahle et al., 2014). Figure 4 demonstrates the signal restoration for October 2002. The convergence criterion was met after approximately 6 iterations. The signal over the mountain range and Inner Mongolia became apparent after the signal restoration was applied (see Fig. 4f).

4.2 Forcing data

The forcing data required by PCR-GLOBWB are precipitation, air temperature, and potential evapotranspiration. Tangdamrongsub et al. (2015) showed that the use of high-quality precipitation data might lead to better estimates of hydrological fluxes (e.g., TWSV and streamflow). In principle, local precipitation and surface temperature measurements could be obtained from the China Daily Ground Climate Dataset provided by the China Meteorological Data Sharing Service System (http://cdc.cma.gov.cn/home.do). A total of 23 weather stations were found over the Hexi Corridor (see Fig. 1b). However, the measurements were spatially sparse and did not cover the entire region. Therefore, the global precipitation data were used to achieve a better spatial coverage. Four global precipitation products were considered for inclusion:

- The European Centre for Medium-range Weather Forecasts (ERA-Interim, spatial resolution: 0.75^{o}x0.75^{o}; Dee et al., 2011)
- The Tropical Rainfall Measuring Mission (TRMM 3B42, spatial resolution: 0.25^{o} x0.25^{o}; Huffman et al., 2007; Kummerow et al., 1998)
- The Climate Research Unit dataset (CRU, spatial resolution: 0.5^{o} x0.5^{o}; Mitchell and Jones, 2005; van Beek, 2008)
- The Princeton's Global Meteorological Forcing Dataset (Princeton, spatial resolution: 0.5^{o} x0.5^{o}; Sheffield et al., 2005)

To select the best product, the global precipitation values were first interpolated to the weather station locations and then the correlation coefficient, Nash-Sutcliffe (NS) coefficient, and RMSD between the interpolated and observed ground data were calculated. The mean values of the statistical estimates are shown in Fig. 5a. Overall, TRMM provided the best data quality, with the highest correlation (~0.85) and NS coefficients (~0.46), and an RMSD approximately 2–3 mm lower than other products. The high spatial resolution of TRMM is probably the reason for its better performance. Therefore, this product was chosen as the precipitation input. The low NS coefficient in all 4 cases suggests that the coarse spatial resolution of the global precipitation datasets prevents them from capturing all the local precipitation events.

A similar procedure was used to compare the air temperature data from ERA-Interim, CRU, and Princeton. The statistical estimates are shown in Fig. 5b. Although the results from all products were very similar, CRU provided the highest data quality in terms of correlation and RMSD values, and therefore, it was used as the temperature input. As far as potential evapotranspiration is concerned, few data are available for this region, so the data from (van Beek, 2008) were used.

4.3 Validation data

4.3.1 Groundwater

Monthly groundwater well measurements at 5 locations (Fig. 1c) were obtained from the ground network maintained by the Shiyang River Basin Management Bureau, and Institute of Water Resources and Hydropower of Gansu Province. The in situ data were provided in the form of piezometric heads (relative to the mean sea level), which needed to be converted to units of storage. For such a task, several parameters, e.g., storage coefficient and specific yield are required, but they are not available over the Hexi Corridor. To solve that problem, a scale factor computed using the information from GRACE and soil moisture (SM) from the Global Land Data Assimilation System (GLDAS, Rodell et al., 2004) was used for the conversion using the approach outlined by Tangdamrongsub et al. (2015). As discussed in Tangdamrongsub et al. (2015), it is ideally preferred to use the in-situ soil moisture data to represent the SM term, but they are not available at the well locations. The soil moisture estimated from remote sensing was also not appropriate due to the limitation of the penetration depth. The use of SM from PCR-GLOBWB is avoided to reduce the bias when compared the adjusted well measurements to the final DA result. Therefore, the GLDAS derived SM was used.

The adjustment procedure was as follows. First, GLDAS-based soil moisture storage variations (SMSV) were removed from GRACE-derived TWSV. Four variants of GLDAS model (NOAH, CLM, MOSAIC, and VIC; see Rodell et al., 2004) were considered and the average SMSV value was calculated. Taking into account that SMSV and groundwater storage variations (GWSV) are the major contributions to TWS variations, this resulted in GWSV ($GWSV_{(GRACE\text{-}SMSV)}$). Then, by conducting a regression analysis between the monthly time-series of piezometric head variation (Δh) and $\Delta GWS_{(GRACE\text{-}SMSV)}$ at each individual location, a bias (b) and a scale factor (f) were estimated using the following linear relationship:

$$\Delta GWS_{(GRACE-SMSV)} + e = b + f \cdot \Delta h, \quad (2)$$

where e indicates the observation error. Finally, the estimated bias ($\hat{b}$) and scale factor ($\hat{f}$) were used to convert the in situ head measurements into groundwater storage variation ($GWSV_{in\ situ}$) as:

$$GWSV_{in\ situ} = \hat{b} + \hat{f} \cdot \Delta h. \quad (3)$$

4.3.2 Streamflow

Monthly river gauge data were obtained from the same data centre as the groundwater measurements. Due to the coarse spatial resolution of PCR-GLOBWB, it models only the main river streams. Therefore, the gauge measurements of small river streams, as well as the gauge measurements that contained many data gaps (i.e., more than 24 months), were excluded. As a result, the measurements from only 2 gauges – at Xida and Xiying Rivers (see Fig. 1c) – were used in this study.

4.4.3 Normalized Difference Vegetation Index (NDVI)

NDVI (Carlson and Ripley, 1997) is an indicator of vegetation health or "greenness". In this study, NDVI and GWS were analysed to determine if the growing season was being extended beyond the limited rainy period through groundwater extraction for irrigation. NDVI was computed from the MODIS 8-day, 500-m spatial resolution surface reflectance product (Vermote et al., 2011) based on data from Aqua satellite (MYD09A1 product). Based on the location of the in situ groundwater measurements, the MODIS tiles h25v05 and h26v05 were selected. First, the data were quality controlled to exclude pixels with cloud cover. The 8-day NDVI was then computed as (Huete et al., 2002)

$$NDVI = \frac{\rho_{NIR}-\rho_R}{\rho_{NIR}+\rho_R}, \quad (4)$$

where ρ_{NIR} and ρ_R are the observed surface reflectances in the near-infrared and red portions of the electromagnetic spectrum, respectively. The monthly-averaged NDVI was then computed based on the derived 8-day NDVI values.

5. Methodology and implementation

5.1 Ensemble Kalman Filter (EnKF)

The Ensemble Kalman Filter (EnKF; Evensen, 2003) is used to assimilate GRACE derived TWSV into the PCR-GLOBWB model. The EnKF works in two steps, a forecast step and analysis (update) step. The forecast step involves propagating the states forward in time using the model (PCR-GLOBWB). Identical to how the EnKF is implemented by Forman et al. (2012), the state vector ($\boldsymbol{\psi}$ in this study is an nm x 1 vector, where n = 27 is the number of TWS-related states from PCR-GLOBWB (see Sect. 3), and m is the number of model grid cells. The model estimates are related to the GRACE observations by

$$\boldsymbol{d} = \mathbf{H}\boldsymbol{\psi} + \boldsymbol{\epsilon};\ \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}), \quad (5)$$

where $\boldsymbol{d}$ is an m x 1 vector containing the GRACE observations for the month of interest, and $\mathbf{H}$ is a measurement operator which relates the PCR-GLOBWB state $\boldsymbol{\psi}$ to the observation vector $\boldsymbol{d}$. Notice that the number of observations is equal to the number of grid cells because the GRACE-based estimates are obtained for all the grid cells of the PCR-GLOBWB model (see Sect. 4.1). The uncertainties in the observations are given in the random error $\boldsymbol{\epsilon}$, which is

assumed to have zero mean and covariance matrix $\mathbf{R}_{m\times m}$. As the sum of all state elements at a given cell is equal to TWSV, the $\mathbf{H}$ matrix is defined as:

$$\mathbf{H} = \left[\begin{pmatrix} (1\,1\,1\,\dots\,1)_{1\times n} & 0 & \cdots & 0 \\ 0 & (1\,1\,1\,\dots\,1)_{1\times n} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (1\,1\,1\,\dots\,1)_{1\times n} \end{pmatrix}\right]_{m\times nm}. \quad (6)$$

Let the ensemble of the states be stored in a matrix $\mathbf{A}_{nm\times N} = (\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \boldsymbol{\psi}_3, \dots, \boldsymbol{\psi}_N)$, where N is the number of ensemble members. Then, the ensemble perturbation matrix is defined as $\mathbf{A}' = \mathbf{A} - \overline{\mathbf{A}}$, where the matrix $\overline{\mathbf{A}}$ is of the same size as $\mathbf{A}$ and filled in with the mean values computed from all ensemble members. Similarly, the GRACE observation vector is stored in the matrix $\mathbf{D}_{m\times N} = (\boldsymbol{d}_1, \boldsymbol{d}_2, \boldsymbol{d}_3, \dots, \boldsymbol{d}_N)$, in which each column is a replicate of the observation but perturbed with random noise $\sim\mathcal{N}(\mathbf{0}, \mathbf{R})$. The analysis equation can be expressed as (Evensen, 2003)

$$\mathbf{A}^a = \mathbf{A} + \Delta\mathbf{A} = \mathbf{A} + \mathbf{K}(\mathbf{D} - \mathbf{H}\mathbf{A}) \quad (7)$$

with

$$\mathbf{K} = \mathbf{P}_e\mathbf{H}^T(\mathbf{H}\mathbf{P}_e\mathbf{H}^T + \mathbf{R})^{-1}, \quad (8)$$

where $\mathbf{A}^{\boldsymbol{a}}_{nm\times N}$ is the updated model state, $\Delta\mathbf{A}_{nm\times N}$ is the update from Kalman filter, and $\mathbf{K}_{nm\times m}$ is the Kalman gain matrix. The model error covariance matrix $(\mathbf{P}_e)_{nm\times nm}$ is computed as

$$\mathbf{P}_e = \mathbf{A}'(\mathbf{A}')^T/(N-1). \quad (9)$$

The matrix $\mathbf{R}$ is the error variance-covariance matrix of GRACE data in the spatial domain, its computation is discussed in Sect. 5.2.2.

In the initialization phase, which was needed to obtain the initial states, the model was spun up between 1 January 2000 and 31 December 2000 as a hot start. This time interval was sufficient to reach the dynamic equilibrium. The initial state $\boldsymbol{\psi}$ for 31 December 2000 obtained this way was perturbed to yield $N = 100$ ensemble members $\boldsymbol{\psi}_i, i = 1, 2, 3, \dots, N$. The N ensemble runs between 1 January 2001 and 31 March 2002 were then conducted independently based on the perturbed initial states. This resulted in an ensemble spread of the estimated states. The model was then propagated in time between 1 April 2002 and 31 December 2010 without assimilating any observation. This case is referred to hereafter as the Ensemble Open Loop (EnOL). For the EnKF, the model was also propagated beginning from 1 April 2002, but the observations (when available) were assimilated.

The processing diagram is shown in Fig. 6, and follows the methodology introduced by Forman et al. (2012). The state is first propagated in time from the first to the last day of the month without applying DA, and the monthly averaged states are calculated from the daily values. When the GRACE observation for that month is available, the DA routine is activated (otherwise, the model continues propagating to the next month without applying DA). The

DA routine computes the monthly averaged update $\Delta\mathbf{A}$ of the TWS-related states, cf. Eq. (7). The daily increment (DINC) of the update is then computed by dividing the monthly averaged update by the total numbers of days in that month ($\text{numday}_{\text{month}}$). The model propagation is then restarted (second run), using the last day of the previous month (month-1, $\text{numday}_{\text{month-1}}$) as the initial state. In this second run, the DINC is added to the current states every day up to the last day of the month. The DA scheme is repeated for each month up to the end of the study period.

Spatial correlations of model errors and observation errors were also taken into account in view of the fact that the latter are highly correlated at neighbouring 0.5°x0.5° grid cells. De Lannoy et al. (2009) proposed a so-called 3D-Fm (3-dimentional fine scale with multiple observation) approach, which is called EnKF 3D in this paper. The approach only considers the spatial correlations between the neighbouring grid cells. In this study, the neighbouring grid cells were assumed to be the ones inside the Gaussian smoothing radius applied, i.e., 250 km. This reduced the computational cost, as only a small subset of cells pairs was considered instead of all cells pairs. That approach was applied not only to observation errors, but also to model errors in TWSV and TWS-related components in this study. The EnKF 3D scheme is illustrated in Fig. 7. For a particular grid cell (centre grid cell), all TWS-related components of the neighbouring grid cells and the centre grid cell are used to form the state ($\mathbf{A}^{\mathbf{s}}_{np\times N}$) and observation ($\mathbf{D}^{\mathbf{s}}_{p\times N}$) matrices, where p is the number of the considered grid cells. The matrix notation with superscript **s** (e.g., $\mathbf{A}^{\mathbf{s}}$) is only used to emphasize the cell-dependent version, and it can be substituted into the original matrix notation (e.g., **A**) in Eqs. (5–9). It is emphasized here that EnKF 3D involves only p grid cells instead of all m grid cells. As such, the measurement operator, model error covariance matrix, and observation error covariance matrix become $\mathbf{H}^{\mathbf{s}}_{p\times np}$, $(\mathbf{P}^{\mathbf{s}}_{e})_{np\times np}$, and $\mathbf{R}^{\mathbf{s}}_{p\times p}$, respectively. The EnKF was then applied and the states of the centre grid cell (only) were updated. The procedure was repeated through all grid cells. To investigate the impact of including spatial correlations of errors, the EnKF 1D was also considered. The EnKF 1D scheme is similar to EnKF 3D, but the spatial correlations are omitted (i.e., the off-diagonal elements of the covariance matrices $\mathbf{P}^{\mathbf{s}}_{e}$ and $\mathbf{R}^{\mathbf{s}}$ are set to zero).

Furthermore, sampling errors caused by finite ensemble size might lead to spurious correlations in the estimated model error covariance matrices (Hamill et al., 2001). To reduce such an effect, a distance-dependent localization function is applied to $\mathbf{P}^{\mathbf{s}}_{e}$ (pair-wise). In this study, the Gaussian function ($c(\alpha)$) (Jekeli, 1981) was used for that:

$$c(\alpha_{j_1,j_2}) = \frac{e^{-b\left[1-\cos\left(\alpha_{j_1,j_2}/a_e\right)\right]}}{1-e^{-2b}} \tag{10}$$

$$\text{with } b = \frac{\ln(2)}{1-\cos(L/a_e)}, \tag{11}$$

where α_{j_1,j_2} is the distance on the Earth's surface between two grid cells (j_1 and j_2), and L is the correlation distance. The variogram analysis was used to derive the TWSV correlation distance (L) of PCR-GLOBWB, assuming that it is similar to the correlation distance of model errors. It was found to be approximately equal to 110 km over the Hexi Corridor. For

GRACE observations, to ensure that the spurious error correlations at distances greater than the Gaussian smoothing distance, 250 km, are insignificant, the localization applied to $\mathbf{R}^{s}$ was based on L = 250 km. The localization also makes the correlations at short distances slightly weaker. As a result, the condition number of the error covariance matrix is increased. In this study, for instance, the condition number increased from ~10^{14} to ~10^{2}. Thus, this matrix had a full rank after localization (see Sect. 5.2.2 for a further discussion).

5.2 Errors of PCR-GLOBWB model and errors in GRACE observations

5.2.1 Model errors

The two primary sources of considered errors in the PCR-GLOBWB model are the meteorological forcing data and the model parameters. For forcing data, the precipitation uncertainties were quantified as the RMS error provided by the TRMM product (Huffman, 1997). The uncertainties of temperature and potential evapotranspiration were not provided as parts of the corresponding products, and therefore errors of 2^{o}C, and 30% of the nominal potential evapotranspiration value were assumed, respectively. The error levels were chosen through trial-and-error, mainly to allow the ensemble to grow between updates. The precipitation and potential evapotranspiration were perturbed with additive lognormal noise while the temperature was perturbed with additive Gaussian noise. The forcing data uncertainties were assumed to be spatially correlated, which was accounted for using an exponential decay function. Based on a variogram analysis, the correlation distances of precipitation, temperature and potential evapotranspiration were found to be approximately 150 km, 450 km, and 450 km, respectively.

As far as model parameters are concerned, a total of 15 TWS-related parameters (see Table 1, Sutanudjaja et al., 2011, 2014) were perturbed using additive Gaussian noise without spatial correlations. The standard deviation of the perturbations of the parameters was set to 20% of the range of the nominal values.

5.2.2 GRACE observation errors

Spatial correlations of GRACE observation errors were also taken into account in the DA scheme. The uncertainties in the GRACE-derived TWSV over the Hexi Corridor were computed using the monthly calibrated error variance-covariance matrix of the SHCs ($\mathbf{\Sigma}$) provided by the CSR. Recalling the replacement of the low degree SHCs (see Sect. 4.1), the error (co-)variances of SHCs degree 2 were not provided by Cheng and Tapley (2004), and therefore the values obtained from the CSR were used. As for SHCs of degree 1, the error (co-) variances were not available from (Swenson et al., 2008) either and were set to zero. Note that $\mathbf{\Sigma}$ only reflects the error of the original GRACE data, i.e. before the GRACE processing described in Sect. 4.1 was applied. To obtain the error variance-covariance matrix associated with the post-processed GRACE data, an ensemble of SHC noise realizations $\mathbf{Q}^{c}$ was first generated based on $\mathbf{\Sigma}$ as follows:

$$\mathbf{Q}^c = (\boldsymbol{\Sigma})^{\frac{1}{2}}\,\mathbf{Q}^w, \tag{12}$$

where $\mathbf{Q}^w = (q_1^w, q_2^w, q_3^w, \dots, q_N^w)$ contains a set of white noise realizations and has the dimension of $s \times N$, where s = 1891 is the number of SHCs, and N = 100 is the number of realizations. The matrix $\mathbf{Q}^c = (q_1^c, q_2^c, q_3^c, \dots, q_N^c)$ has the same dimension as $\mathbf{Q}^w$ and contains an ensemble of correlated noise realizations in SHCs. Then, each noise realization (i.e., column of $\mathbf{Q}^c$) was post-processed in the same way as the GRACE data (Sect. 4.1), which resulted in $\widehat{\mathbf{Q}}^c = (\hat{q}_1^c, \hat{q}_2^c, \hat{q}_3^c, \dots, \hat{q}_N^c)$. The post-processing included applying the destriping and Gaussian smoothing filters, as well as the signal restoration using the same number of iterations as was used in the GRACE data post-processing. The error variance-covariance matrix $\widehat{\boldsymbol{\Sigma}}$ associated with the SHCs after post-processing was then computed as

$$\widehat{\boldsymbol{\Sigma}} = \left[\widehat{\mathbf{Q}}^c\left(\widehat{\mathbf{Q}}^c\right)^T\right]/(N-1). \tag{13}$$

Recalling Eq. (1), the TWSV over the Hexi Corridor can be computed as

$$\Delta\boldsymbol{\sigma} = \mathbf{YSx}, \tag{14}$$

where $\Delta\boldsymbol{\sigma}$ is the vector composed of the computed TWSV at grid cells, $\mathbf{Y}$ is the matrix of spherical harmonic synthesis (cf. Eq. (1)), $\mathbf{S}$ is the matrix containing the scaling factors S_l, and $\mathbf{x}$ is the vector composed of the dimensionless SHC variations after GRACE data post-processing described in Sect. 4.1. Then, the error covariance matrix $\mathbf{R}$ of the GRACE-based TWSV over the Hexi Corridor was computed with the error propagation law as

$$\mathbf{R} = \mathbf{YS}\,\widehat{\boldsymbol{\Sigma}}\,(\mathbf{YS})^T. \tag{15}$$

Some statistics of GRACE TWSV errors over the Hexi Corridor are shown in Fig. 8. The error standard deviation in Oct. 2002 varied with location (Fig. 8a), whereas the error correlation showed a distance-decay pattern in all directions (Fig. 8b). The areally-averaged standard deviations over 4 basins stayed in most of the months at a similar level of approximately 1 cm (Fig. 8c). The large uncertainty in September 2004 was likely caused by the near-repeat orbit of GRACE satellites during that month.

6. Results and discussion

The structure of this section is as follows. First, the impact of assimilation using EnKF 3D on the TWSV is considered in Sect. 6.1. Then, the impact of the EnKF 3D on the estimates of the individual stores is investigated in Sect. 6.2. The results of the EnKF 1D and EnKF 3D schemes are compared in Sect. 6.3 in terms of TWSV and the individual stores. Furthermore, the obtained results are validated against independent data in Sect. 6.4. Finally, in Sect. 6.5-6.6, the assimilation results are used together with ancillary remote sensing data to study water resources in the Hexi Corridor.

6.1 Impact of GRACE DA

To demonstrate the impact of DA, Fig. 9 shows the daily TWSV estimates over the Shiyang River Basin between 1 April 2002 and 31 December 2003. Several features associated with the EnKF can be observed. Firstly, when a GRACE observation is available, the EnKF moves the estimated TWSV towards it. As a result, the estimated TWSV lies between the EnOL estimate and the GRACE observation most of the time. It is seen that GRACE-derived TWSV has a greater annual amplitude compared to the model estimated TWSV. This can likely be attributed to the poor quality of the model parameter calibration and the accuracy of the meteorological input data over the data-sparse regions. In the absence of observations, model parameters are difficult to determine and only the best available knowledge (or guess) is generally used, leading to inaccurate model state estimates. Updating the water storage estimates using GRACE DA showed a clear improvement in this case. Secondly, the standard deviation across the EnKF ensemble of TWSV values is smaller than that of the EnOL and smaller than the GRACE observation error. Thirdly, at the first month (April 2002) the TWSV estimates of the EnOL and EnKF were similar at the forecast step (as the initial states were the same, see point (a) in Fig. 9), but became different when the daily increment was applied to the EnKF. Finally, discontinuities in the time-series before the update were observed at the end of the month e.g., in November and December 2002 (point (b) and (c)), and February 2003 (point (d)). Applying the daily increment (see Sect. 5.3) served as a smoother, and these stepwise changes were reduced.

Similar features were also seen in the EnKF 1D TWSV estimates (not shown).

6.2 Impact of GRACE DA on individual stores

The monthly-averaged values of the TWSV and individual stores in each of the 4 basins are presented in Fig. 10. Overall, TWSV estimates over the Hexi Corridor mostly reflect SMSV and GWSV components, while snow water storage variation (SNSV) and surface water storage variation (SFWV) are minor contributors, constituting less than 5% in most basins. Clear seasonal variations in TWSV were seen in all basins for GRACE, EnOL and GRACE DA (both EnKF 1D and EnKF 3D) time-series (Fig. 10 a,b,c,d). As observed in Fig. 10, the GRACE DA estimated TWSVs are generally between the GRACE observations and the EnOL estimates. As a result of assimilating GRACE data, both the EnKF 1D and EnKF 3D added water to all basins between 2002 and 2005 and reduced it from the basins between 2006 and 2010. This is also seen in the time-series of SMSV (Fig. 12 e,f,g,h) and GWSV (Fig. 12 I,j,k,l). Additionally, the annual amplitudes and phases of GRACE DA estimated TWSV were also found mostly in between the values computed from the GRACE observations and the EnOL results (see Table 2). In particular, the GRACE-DA estimated TWSV's phase was always closer to the GRACE observation. The phase shifts of approximately 1 month were seen in both GRACE DA estimated TWS and GRACE observations compared to the EnOL results. Similar phase differences of approximately 1 month were also observed in SMSV and GWSV components.

Differences in the long-term trends were also detected between the TWSV estimates from the model alone (EnOL) and the GRACE DA. The GRACE DA results showed decreasing TWSV trends similarly to the GRACE data, while the EnOL showed increasing trends (Fig. 10 a,b,c,d, see also Table 7). This change in TWSV trend was clearly a result of assimilating GRACE observations. The negative trends were also observed after DA in the GWSV component in most basins (Fig. 10 i,j,l). This indicates the potential of GRACE DA in adjusting GWSV. In this way, one can reveal continued groundwater consumption to support local agricultural activities (Li et al., 2013). Unlike over other basins, the negative trend of GWSV estimates was not clearly present over the Desert Region (Fig.10k). This could be due to the small amplitude of the groundwater variation of this region (see also below), and most of the update took place in the SM component. As a result, a relatively large negative trend was seen in SMSV rather than GWSV after GRACE DA (see also Table 7). Further discussions on the trends are given in Sect. 6.4.

The impact of GRACE DA on different stores was influenced by both the model parameters and the forcing data. The 4 basins have similar soil water storage capacities (see Table 3), which indicates that the basins can store similar amounts of soil water and generate similar amounts of groundwater recharge under the same rainfall conditions. However, the 4 basins received different amounts of rainfall, which resulted in different SMSV and GWSV estimates. For example, the Shiyang River Basin received the greatest amount of rainfall (~ twice of Heihe River Basin), which led to the greatest amount of the SMSV estimate (~1 cm annual amplitude). Such a large amount was also sufficient to percolate into the groundwater layer, resulting in GWSV of ~0.7 cm (see Fig. 10i and Table 2). In contrast, the Desert Region received approximately 3 times less rainfall, which led to a somewhat smaller amount of SMSV (~0.7 cm annual amplitude) and a much smaller amount of GWSV, ~0.2 cm (see Fig. 10g, k). As the uncertainty of the water storage variation is associated with the signal amplitude, the greater (smaller) water storage variation leads to greater (smaller) uncertainty, resulting in greater (smaller) update from GRACE DA. As such, a greater update (in particular, in GWSV) is seen over the Shiyang River Basin, as compared to other basins.

Snow estimates (SWE plus SFW) were very small (less than 0.2 cm) over the Hexi Corridor and therefore were only slightly updated by GRACE DA. Note that the large variability in the amount of snow seen as the sharp peaks (e.g., in January 2008) was caused by the precipitation and temperature variability. In January 2008, the precipitation records were 159 % higher than the January average value while the temperature was 2 – 3ºC lower. Such a condition resulted in a large amount of snow. Finally, GRACE DA influences the surface water, but the amplitude is still lower than that of the GRACE uncertainties. Validation of the surface water estimates in terms of river streamflow is given in Sect. 6.4.2.

6.3 Impact of taking spatial correlations of errors into account

The impact of accounting for the error correlations was clearly seen in the TWSV estimates (Fig. 10 a,b,c,d). When the error correlations were ignored (EnKF 1D), the TWSV estimate received a larger update from GRACE, particularly between 2002 and 2005. Hence, the

estimate was drawn significantly closer to the observation. The presence of error correlations effectively reduces the amount of information in the GRACE data, since spatial averaging of such data mitigates noise to a much less extent than averaging of data with uncorrelated errors. Therefore, the impact of GRACE data in the EnKF 3D case is reduced. As such, the EnKF 3D estimated TWSV was always between the EnOL and EnKF 1D results. Validating against the in situ groundwater and streamflow data will quantitatively reveal the performance of each approach (Sect. 6.4).

Taking error correlations into account also has a clear impact on the SMSV and GWSV components. For SMSV, similarly to TWSV, the EnKF 1D yielded a larger update between 2002 and 2005 compared to the EnKF 3D (Fig. 10 e,f,g,h). The difference between EnKF 1D and 3D results became smaller after 2005. This can be attributed to the fact that the ensemble spread in the soil moisture component becomes smaller after several years of updates. After 2005, the ensemble spread of SMS was lower than the GRACE uncertainty, and therefore taking the error correlations into account did not have a significant impact on the SMS estimates. For GWS, the impact of taking error correlations into account was even clearer, especially in terms of the long-term trend (Fig. 10, i,j,k,l). With the exception of the Desert Region, the EnKF 1D showed a steeper decreasing trend in all basins. For snow and surface water, the impact of considering error correlations was not significant due to the fact that the stores are small, as compared to SMS and GWS.

It is also worth discussing the impact of GRACE DA on the spatial pattern of the water storage estimates. To demonstrate this, the update term (ΔA in Eq. (7)) of October 2002 from EnKF 1D and 3D cases is shown in Fig. 11. Only TWSV, SMSV, and GWSV are presented, since other components (snow, surface water, and interception) are small. As discussed above, EnKF 3D shows smaller update in all components. Due to a greater amplitude of GRACE-derived TWSV over northern and southern parts of the region (see Fig. 4), the update is mostly seen there. Almost all update is limited to the soil moisture layer. Higher precipitation is generally observed over the southern part, which leads to higher groundwater recharge (and GWSV) over that region. As such, a GWSV update is clearly seen over the southern part of the region.

6.4 Validation against independent data

6.4.1 Validation of groundwater estimates against well data

The GWSs estimated from GRACE DA were validated against the well measurements at 5 locations shown in Fig. 1c. Yang et al. (2001) showed that the specific yield values obtained from the field measurements over the Shiyang River Basin was between 0.01 and 0.3. Although the measurements were not collected at the well stations used in this study, the values obtained can be used as a guidance of the specific yield of the Shiyang River Basin. In this study, the head measurements were converted to storage unit with the approach described in Sect. 4.3.1. The bias term in Eq. (3) was found to be very close to zero, as the variation (mean removed) was used in the regression analysis. The estimated scale factor was 0.23,

0.04, 0.24, 0.25, and 0.32 at W1 – W5, respectively, which is in line with the values obtained from the field measurement.

The GWSV estimate at each well location is shown in Fig. 12. Compared to the EnOL results, GRACE DA results were visually closer to the well measurements at all 5 locations. The EnKF 1D and EnKF 3D showed a noticeable difference at each location. The updated GWSV estimates were evaluated in terms of the correlation coefficient, RMSD, and long-term trend (Table 4, 5). Overall, the EnOL resulted in relatively poor correlation coefficients at most stations (except station W1), with the average value of only 0.06. Clear improvements were seen after GRACE DA was applied. The average correlation coefficient increased to approximately 0.6 – 0.7. Although the EnKF 1D introduced a greater update than the EnKF 3D, it only showed higher correlation coefficients at stations W1 and W3. Applying the EnKF 3D led to correlation coefficients greater than 0.45 in all stations, and on average it improved the correlation coefficient by approximately 0.1 over EnKF 1D. In terms of RMSD, applying GRACE DA reduced the difference by approximately 15 – 25% compared to the EnOL. Compared to EnKF 1D, the EnKF 3D significantly improved the RMSD in most stations. The EnKF 1D only performed better than EnKF 3D at station W1, where it reduced the RMSD by approximately 16 % compared to 8% reduction by the EnKF 3D. The noticeably low GWSV observed by the well data at station W2 in the summers of 2007 and 2008 (Fig. 12b) was probably caused by significant groundwater abstraction. These local features could not be reproduced by the model and GRACE observations due to a limited spatial resolution. As a result, neither of the EnKF algorithms could improve the GWSV estimates at the W2 location during those periods.

The long-term trend estimated between 2007 and 2010 was also used to evaluate the impact of GRACE DA and the effect of taking the error correlations into account (Table 5). The EnOL trend estimates were considered poor as they showed the largest RMSD respected to the in situ data. In fact, they were the least consistent with the in situ estimates at each individual station. Similar to the results in terms of correlation coefficient and RMSD (see Table 4), the EnKF 3D led to the largest improvement in the trend estimates (RMSD=0.54 compared to 0.93 after EnKF 1D). However, while the EnKF 3D showed closer long-term trends to the in situ measurements at stations W2, W4, W5, the EnKF 1D produced better estimates at station W1 and W3.

Thus, both EnKF 1D and 3D led to the improvement of the GWSV estimates in terms of all metrics. In terms of the average results and at the majority of well locations, the EnKF 3D provided more improvement than the EnKF 1D.

6.4.2 Validation of streamflow estimates against river gauge data

The streamflow estimates were validated against the river gauge measurements at locations G1 and G2 (Fig 1c). Results are shown in Figure 13 and Table 6. Only modest improvements in the streamflow estimates were observed in terms of the correlation coefficient, NS coefficient, and RMSD. This behaviour is similar to what was observed previously for the

Rhine River Basin, when a different hydrology model and input data were used (Tangdamrongsub et al., 2015). Figure 13 shows that taking error correlations into account had little impact, i.e. similar streamflow estimates were seen for EnKF 1D and 3D results. At location G1 (Fig. 13a), GRACE DA added more water to the stream channel between 2002 and 2006 and reduced it between 2008 and 2010. This behaviour is consistent with the TWSV estimates discussed in Sect. 6.2. GRACE DA increased the correlation coefficient from 0.82 to 0.84, increased the NS coefficient from 0.65 to 0.69, and reduced the RMSD by approximately 5 % (Table 6). A lesser improvement was observed at G2.

Comparing to the gauge measurements, both the EnOL and GRACE DA overestimated the streamflow in September 2007 and September 2008 at G2. The sudden surge in the estimated streamflow resulted from heavy rainfall recorded by precipitation data while the soil was, according to the model, already saturated (Fig. 14). For example, in September 2007, the second highest amount of SM storage in the record (~19.5 cm) was obtained when the third largest amount of rainfall (~90 mm/month) was observed. Similarly, in September 2008, large SM storage (~20 cm) and the heaviest rainfall (~100 mm/day) forced PCR-GLOBWB to generate a large amount of streamflow. In both cases, the modelled streamflow significantly exceeded the actual one observed at G2. Inaccurate precipitation data and model calibration likely led to these discrepancies. GRACE DA was unable to reduce these spurious peaks due to the limited spatial (~250 km) and temporal (1 month) resolution of GRACE data.

6.5 Declining water storages in the Hexi Corridor

The water resources situation over the Hexi Corridor was assessed using long-term trends estimated from the 9-year EnKF 3D results. This DA variant is primarily discussed here as it provided better agreement with well observations than the EnKF 1D (see Sect. 6.4.1). For completeness, however, the values estimated from GRACE, EnOL, EnKF 1D, and precipitation are also provided. The trends in the TWSV, SMSV and GWSV for the 4 basins, as well as the areally-averaged values for the entire Hexi Corridor, are given in Table 7. The average EnKF 3D trends are all negative: approximately –0.2, –0.1, and –0.1 cm/yr for TWSV, SMSV, and GWSV, respectively. This reduction in the water storages is observed despite the increased amount of rainfall, which shows a positive trend of about 0.4 (mm/month)/yr. The water storage reductions can likely be attributed to the extraction of groundwater to meet irrigation demands. In Sect. 6.6, it will be shown that groundwater extractions are essential for that purpose in the Hexi Corridor.

Focusing on individual river basins provides additional insight into the water storage issue, as the influence of the large desert area is removed. The water storage losses in the individual basins outside the desert are even more pronounced, particularly in the Shiyang River Basin. This basin had the greatest TWS loss (approximately 0.4 cm/yr), which was entirely caused by the reduction of GWS. This can be explained by groundwater abstraction to meet the irrigation demand in the region. The Heihe and Shule River Basins also experienced a TWS loss of ~0.2 cm/yr, which came from a reduction of both soil moisture and groundwater storages. Again, the negative GWS trend was likely caused by significant pumping of

groundwater to maintain crop production. This is consistent with the extreme water stress over the Heihe River basin between 2001 and 2010, which was documented in Table 11.7 of the study by Chen et al. (2014). In the Desert Region, in contrast to other basins, the minor decreasing TWS trend of -0.1 cm/yr was dominated by loss of SM storage. This was likely caused by inaccurate model parameter calibration over the Desert Region (i.e., too large SC value). Separation of the TWS into groundwater and soil moisture store was likely incorrect. As such, the annual signal in GWS is much less than in SM there. Therefore, GRACE update was mostly attributed to the SM component, so that a groundwater-pumping signature (Jiao et al., 2015) was seen in the SM instead of the GWS layer.

6.6 Connection to agriculture activity

Figure 15 shows the monthly averaged groundwater head measurements at wells W1 to W5 in the Shiyang River Basin (Fig. 1c). Monthly averaged precipitation and NDVI values are shown as well. Since extracted water can be used to support agriculture not only at the well location but also in the nearby area, precipitation and NDVI are reported as the average values within a circular area of the 10-km radius. These data will be used to ascertain if groundwater extractions to support agriculture might be the source of the negative GWS trends observed in Fig. 12 and Table 6. From Fig. 14, it is noticed that the growing period is approximately between May and October, where the amount of rainfall is higher than 15 mm/month and the NDVI is typically greater than 0.2. By observing well measurements, precipitation, and NDVI together, some groundwater extraction signatures can be explained by the extension of the growing period over the dry season. For example, at station W1, the groundwater in 2010 was lower than the average, showing a gradual decrease in summer (Fig. 15a). One may attribute this to the shortage of rainfall in July and August 2010, which was lower than the average (Fig. 15b). However, the NDVI value was higher than the average during summer 2010 (Fig. 15c), which implies that water from other sources than precipitation was probably used to maintain the growing period. This additional water was likely extracted from the ground, and such an activity led to a decreased groundwater table during summer 2010. A similar explanation can be applied to station W2, where low groundwater head, low rainfall, and high NDVI were observed in summer 2007 and summer 2008 (Fig. 15 d,e,f). At station W3, the behaviour is similar to station W1: the extension of the growing period was observed in summer 2010, where the GWS and precipitation were lower than the average, while NDVI was significantly higher (Fig. 15 g,h,i). Groundwater pumping signatures were not present at stations W4 and W5.

7. Conclusions

This study was focused on the estimation of water resources dynamics in the Hexi Corridor by assimilating GRACE-derived TWSV into the PCR-GLOBWB hydrological model. Validating against well data showed that DA led to noticeable improvement in the state estimates in terms of correlation, RMSD, and long-term trend. Furthermore, GRACE DA estimates

revealed the reduction of water storages between 2002 and 2010. The Shiyang River Basin – the southeaster part of the Hexi Corridor area – suffered the most from the water loss, which was likely caused by the overuse of the groundwater for irrigation. Due to inaccurate groundwater abstraction information, PCR-GLOBWB alone could not properly capture the downward trend of water storages. This highlights the value of the GRACE DA in this situation. It should be emphasized that GRACE does not fix a technical problem of the hydrological model, but rather it provides information, which is not available otherwise. Note that, in principle, the model may predict any long-term behaviour of water storage, but that information should be brought in "by hand" (e.g., via the groundwater abstraction parameter). As soon as that information is not available, reliable long-term predictions on the basis of hydrological modelling alone are conceptually impossible. GRACE DA acts as a provider of a missing puzzle piece here. Of course, the performance of GRACE DA needs to be further investigated in other geographical locations and with different hydrological models to confirm its benefits.

A substantial decrease in the water storage in the Hexi Corridor between 2002 and 2010, particularly over the Shiyang River Basin, took place in spite of the increased precipitation. The amount of water from rainfall was likely insufficient to support irrigation water requirements. Irrigation water demands increased significantly to maintain the crop production and, as a result, the region was under extreme water stress. Water consumption from all available sources was essential for bridging the deficit, including a sizeable amount of groundwater extraction. This study illustrates how ground observations and remote sensing data may reveal the connection between groundwater pumping and agricultural activity.

The conversion approach between the groundwater head measurement and groundwater storage is proven feasible over the Shiyang River Basin. The scale factor estimates produced with this approach are consistent with the specific yield estimated from the field observations. However, it is noted here that the results of the conducted validation might be over-optimistic, since the well data processed with the adopted conversion procedure are not fully independent of the assimilated GRACE data. The specific yield from the field observation must be used when available.

Furthermore, we demonstrate how the error covariance matrix **R** of GRACE-derived TWSV can be obtained from the error covariance matrix of GRACE SHCs (which is currently provided together with the SHCs themselves). This study shows that it is necessary to use the **R** matrix in order to properly take into account the error correlations in the DA scheme. To come to that conclusion, we considered 2 variants of the error variance-covariance matrix in the data assimilation: excluding and including error correlations. Validating against well data showed that ignoring error correlations in DA tended to over-fit results to the observations, and in many cases led to less accurate state estimates. This finding is in agreement with the recommendation in Schumacher et al. (2016). We explain this finding by the fact that GRACE errors at the neighbouring 0.5^{o}x0.5^{o} grid cells are highly correlated. As such, the simultaneous consideration of GRACE data at multiple neighbouring cells does not reduce data noise, as it would be the case if noise were white. In other words, the white-noise assumption may severely overestimate the information content of GRACE data. We recognize

that the derivation of GRACE-derived TWSV error variance-covariance matrices is very computationally demanding. Still, we believe that this is a reasonable price to pay as deriving the error variance-covariance matrix from the full (and only full) error covariance matrix noticeably improves the results of GRACE data assimilation.

To further improve the DA performance, an extended or an alternative DA framework can be considered. One of the points of attention is only a minor improvement in streamflow estimates, which is caused by an insufficient temporal and spatial resolution of GRACE data. A promising way to go is to improve the runoff scheme at a conceptual level, e.g., by extending GRACE DA with a simultaneous parameter calibration. To that end, the state vector should be extended to also include selected model parameters (Eicker et al., 2014; Wanders et al., 2014). This allows for the adjustment of the storage size and might lead to a more accurate estimate of model states, including streamflow (Wanders et al., 2014). Alternative ensemble-based DA approaches, such as particle filters (Weerts and El Serafy, 2006), can also be considered. Particle filters estimate a sample from the realistic posteriori distribution, which is not necessarily Gaussian, like in the EnKF. The approach has been shown very effective for the parameter calibration (Dong et al., 2015).

Finally, the usage of improved gravity solutions to be available after the launch of the GRACE Follow-on mission (Flechtner et al., 2014) will probably further increase the accuracy of the GRACE DA estimates.

Acknowledgement

This research was funded by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (Netherlands Organisation for Scientific Research, NWO; project number 842.00.006) and Ministry of Science and Technology of China (MoST, Project Number 2010DFA21750) under the Samenwerking China - Joint Scientific Thematic Research Programme (JSTP). The research was also sponsored by the NWO Exacte Wetenschappen, EW (NWO Physical Sciences Division) for the use of supercomputer facilities, with financial support from NWO. The research was also co-funded by National Natural Science Foundation of China (NSFC, project number 51279076). Authors would like to thank H. Bogena, M. Schumacher, and one anonymous reviewer for the insightful suggestions, which led to the improvement of the manuscript.

8. References

Bettadpur, S.: Insights into the Earth System mass variability from CSR-RL05 GRACE gravity fields, EGU Meeting, abstract #EGU2012-6409, Vienna, Austria, 2012.

Carlson, T. N. and Ripley, D. A.: On the relation between NDVI, fractional vegetation cover, and leaf area index, Remote Sens. Environ., 62(3), 241-252, doi:10.1016/S0034-4257(97)00104-1, 1997.

867 Chen, J., Li, J., Zhang, Z., and Ni, S.: Long-term groundwater variations in Northwest India
868 from satellite gravity measurements. Global Planet. Change, 116, 130–138.
869 doi:10.1016/j.gloplacha.2014.02.007, 2014.

870 Chen, Y.: Water resources research in northwest China, Springer Netherlands,
871 doi:10.1007/978-94-017-8017-9, 2014.

872 Cheng, M. and Tapley, B.: Variations in the Earth's oblateness during the past 28 years, J.
873 Geophys. Res., 109(B09402), doi:10.1029/2004JB003028, 2004.

874 Cui, Y. and Shao, J.: The Role of Ground Water in Arid/Semiarid Ecosystems, Northwest
875 China, Groundwater, 43 (4), 471–477, doi:10.1111/j.1745-6584.2005.0063.x, 2005.

876 Dahle, C., Flechtner, F., Gruber, C., König, D., König, R., Michalak, G., and Neumayer, K.-
877 H.: GFZ RL05: An Improved Time-Series of Monthly GRACE Gravity Field Solutions, In
878 Flechtner, F., Sneeuw, N., Schuh, W.-D. (Eds.), Observation of the System Earth from Space
879 - CHAMP, GRACE, GOCE and future missions, (GEOTECHNOLOGIEN Science Report;
880 20; Advanced Technologies in Earth Sciences), Berlin, Springer, 29-39,
881 http://doi.org/10.1007/978-3-642-32135-1_4, 2014.

882 Dahlgren, S. and Possling, B.: Soil Water Modelling In Arid/Semiarid Regions of Northern
883 China Using Land Information System (LIS): A Minor Field Study in Shiyang River Basin,
884 Dept. of Water Resources Engineering, Lund University, ISRN LUTVDG/TVVR-07/5019,
885 ISSN-1101-9824, 2007.

886 De Graaf, I. E. M., van Beek, L. P. H., Wada, Y., and Bierkens, M. F. P.: Dynamic attribution
887 of global water demand to surface water and groundwater resources: Effects of abstractions
888 and return flows on river discharges, Adv. Water Resour., 64, 21–33,
889 doi:10.1016/j.advwatres.2013.12.002, 2014.

890 De Lannoy, G. J. M., Reichle, R. H., Houser, P. R., Arsenault, K. R., Verhoest, N. E. C., and
891 Pauwels, R. N.: Satellite-Scale Snow Water Equivalent Assimilation into a High-Resolution
892 Land Surface Model, J. Hydrometeor., 11, 352–369,
893 doi:http://dx.doi.org/10.1175/2009JHM1192.1, 2009.

894 Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U.,
895 Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L.,
896 Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haiberger, L.,
897 Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M.,
898 McNally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P.,
899 Tavolato, C., Thépaut, J. N., and Vitart, F. (2011). The ERA-Interim reanalysis: configuration
900 and performance of the data assimilation system. Quarterly Journal of the Royal
901 Meteorological Society, 137, 553–597, doi:10.1002/qj.828.

902 Dong, J., Steele-Dunne, S. C., Judge, J., and van de Giesen, N.: A particle batch smoother for
903 soil moisture estimation using soil temperature observations, Adv. Water Resour., 83, 111 –
904 122, http://dx.doi.org/10.1016/j.advwatres.2015.05.017, 2015.

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

Du, T., Kang, S., Zhang, Z., and Zhang, J.: China's food security is threatened by the unsustainable use of water resources in North and Northwest China, Food and Energy Security, 3(1), 7–18, doi: 10.1002/fes3.40, 2014.

Eicker, A., Schumacher, M., Kusche, J., Döll, P., and Müller Schmied, H.: Calibration data assimilation approach for integrating GRACE data into the WaterGAP Global Hydrology Model (WGHM) using an Ensemble Kalman Filter: First Results, Surv. Geophys., 35(6), 1285-1309, doi:10.1007/s10712-014-9309-8, 2014.

Evensen, G.: The ensemble Kalman filter: Theoretical formulation and practical implementation, Ocean Dyn., 53(4), 343–367, doi:10.1007/S10236-003-0036-9, 2003.

Flechtner, F., Morton, P., Watkins, M., and Webb, F.: Status of the GRACE follow-on mission, in IAG symposium gravity, geoid, and height systems, 141, Venice, Italy, Springer, 117–121, 2014.

Forman, B. A., Reichle, R. H., and Rodell, M.: Assimilation of terrestrial water storage from GRACE in a snow-domained basin, Water Resour. Res., 48, W01507, doi:10.1029/2011WR011239, 2012.

Geng, G. T. and Wardlaw, R.: Application of Multi-Criterion Decision Making Analysis to Integrated Water Resources Management, Water Resour. Manage., 27, 3191–3207, 2013.

Girotto, M., De Lannoy, G. J. M., Reichle, R. H., and Rodell, M.: Assimilation of gridded terrestrial water storage observations from GRACE into a land surface model, Water Resour. Res., 52(5), 4164–4183, doi:10.1002/2015WR018417, 2016.

Gleick, P. H., Cooley, H., Cohen, M. J., Morikawa, M., Morrison, J., and Palaniappan, M.: The Worlds Water 2008–2009, The Biennial Report on Freshwater Resources, Island Press, Washington, DC, USA, 2009.

Gong, D. Y., Shi, P. J., and Wang, J. A.: Daily precipitation changes in the semi-arid region over northern China. J. Arid. Environ., 59 (4), 771–784, doi:10.1016/j.jaridenv.2004.02.006, 2004.

Hamill, T. M., Whitaker, J. S., and Snyder, C.: Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter, Mon. Weather Rev., 129, 2776–2790, 2001.

Houborg, R., Rodell, M., Li, B., Reichle, R., and Zaitchik, B. F.: Drought indicators based on model-assimilated Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage observations, Water Resour. Res., 48, W07525, doi:10.1029/2011WR011291, 2012.

941 Hu, Z. D.: Study on Study on Evolution Pattern of Water Resources, Ecology and Oasis
942 Migration in the Hexi Corridor, Doctoral Dissertation, Tsinghua University, Beijing, China,
943 2015.

944 Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., and Ferreira, L. G.: Overview of
945 the radiometric and biophysical performance of the MODIS vegetation indices, Remote, 83,
946 195–213, doi:10.1016/S0034-4257(02)00096-2, 2002.

947 Huffman, G.J.: Estimates of Root-Mean-Square Random Error for Finite Samples of
948 Estimated Precipitation, J. Appl. Meteor., 1191–1201, 1997.

949 Huffman, G. J., Adler, R. F., Bolvin, D. T., Gu, G., Nelkin, E. J., Bowman, K. P., Hong, Y.,
950 Stocker, E. F., and Wolf, D. B.: The TRMM multisatellite precipitation analysis (TMPA):
951 Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, J.
952 Hydrometeor., 8, 38–55, doi:10.1175/JHM560.1, 2007.

953 Huo, Z., Feng, S., Kang, s., Dai, x., Li, w., and Chen, S.: The Response of Water-Land
954 Environment to Human Activities in Arid Minqin Oasis, Northwest China, Arid Land Res.
955 Manag., 21(1), 21–36, 2007.

956 Jekeli, C.: Alternative methods to smooth the Earth's gravity field, Rep., 327, Dept. of Geod.
957 Sci. and Surv., Ohio State Univ., Columbus, 1981.

958 Jiao, J. J., Zhang, X., and Wang, X.: Satellite-based estimates of groundwater depletion in the
959 Badain Jaran Desert, China, Nature Sci. Rep., 5, 8960, doi:10.1038/srep08960, 2015.

960 Kang, S., Su, X., Tong, L., Shi, P., Yang, X., Abe, Y., Du, T., Shen, Q., and Zhang, J.: The
961 impacts of human activities on the water-land environment of the Shiyang River basin, an arid
962 region in northwest China, Hydrolog. Sci. J., 49(3), 2014.

963 Klees, R., Liu, X., Wittwe, T., Gunter, B. C., Revtova, E. A., Tenzer, R., Ditmar, P.,
964 Winsemius, H. C., and Savenije, H. H. G.: A Comparison of Global and Regional GRACE
965 Models for Land Hydrology, Surv. Geophys., 29, 335-359, doi:10.1007/s10712-008-9049-8,
966 2008.

967 Klein Goldewijk, K. and van Drecht, G.: Integrated modelling of global environmental
968 change: An overview of IMAGE 2.4, chap. HYDE 3: Current and historical population and
969 land cover, MNP – Netherlands Environmental Assessment Agency, Bilthoven, the
970 Netherlands, 93–112, 2006.

971 Kummerow, C., Barnes, W., Kozu, T., Shiue, J., and Simpson, J.: The Tropical Rainfall
972 Measuring Mission (TRMM) sensor package. J. Atmos. Ocean. Tech., 15, 809–817, 1998.

973 Li, B., Rodell, M., Zaitchik, B. F., Reichle, R. H., Koster, R. D., and van Dam, T. M.:
974 Assimilation of GRACE terrestrial water storage into a land surface model: Evaluation and
975 potential value for drought monitoring in western and central Europe, J. Hydrol., 446–447,
976 2012.

977 Li, F., Zhu, G., and Guo, C.: Shiyang River ecosystem problems and countermeasures, Agr.
978 Sci., 4, 72–78, doi: 10.4236/as.2013.42012, 2013.

979 Long, D., Longuevergne, L., and Scanlon, B. R.: Uncertainty in evapotranspiration from land
980 surface modeling, remote sensing, and GRACE satellites, Water Resour. Res., 50, 1131–1151,
981 doi:10.1002/2013WR014581, 2014.

982 Ma, J.Z., Wang, X.S., and Edmunds, W.M.: The characteristics of ground-water resources and
983 their changes under the impacts of human activity in the arid Northwest China—a case study
984 of the Shiyang River Basin, J. Arid Environ., 61, 277–295, 2005.

985 Mitchell, T. D. and Jones, P. D.: An improved method of constructing a database of monthly
986 climate observations and associated high-resolution grids. Int. J. Climatol., 25(6), 693–712,
987 2005.

988 Niu, G.-Y., Seo, K.-W., Yang, Z.-L., Wilson, C., Su, H., Chen, J., and Rodell, M.: Retrieving
989 snow mass from GRACE terrestrial water storage change with a land surface model,
990 Geophys. Res. Lett., 34, L15704, doi:10.1029/2007GL030413, 2007.

991 Peel, M. C., Finlayson, B. L., and McMahon, T. A.: Updated would map of the Köppen-
992 Geiger climate classification, Hydrol. Earth Syst. Sci., 11, 1633–1644, 2007.

993 Reichle, R. H. and Koster, R. D.: Bias reduction in short records of satellite soil moisture,
994 Geophys. Res. Lett., 31, L19501, doi:10.1029/2004GL020938, 2004.

995 Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C. J., Arsenault,
996 K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D.,
997 and Toll, D.: The global land data assimilation system, Bull. Amer. Meteor. Soc., 85(3), 381–
998 394, 2004.

999 Schumacher, M., Kusche, J., and Döll, P.: A Systematic Impact Assessment of GRACE Error
1000 Correlation on Data Assimilation in Hydrological Models, J. Geod., 90(6), 537–559.
1001 doi:10.1007/s00190-016-0892-y, 2016.

1002 Sheffield, J., Goteti, G., and Wood, E. F.: Development of a 50-yr high-resolution global
1003 dataset of meteorological forcings for land surface modeling, J. Climate, 19 (13), 3088–3111,
1004 2005.

1005 Shiklomanov, I. A.: Assessment of water resources and water availability in the world,
1006 Comprehensive assessment of the freshwater resources of the world, World Meteorological
1007 Organization and the Stockholm Environment Institute, Stockholm, Sweden, 1997.

1008 Steinfeld, H., Gerber, P., Wassenaar, T., Castel, V., Rosales, M., and de Haan, C.: Livestocks
1009 long shadow: Environmental issues and options, FAO, Rome, Italy, ISBN 978-92-5-105571-
1010 7, 2006.

1011 Su, H., Yang, Z. L., Dickinson, R. E., Wilson, C. R., and Niu, G. Y.: Multisensor snow data
1012 assimilation at the continental scale: The value of Gravity Recovery and Climate Experiment

terrestrial water storage information, J. Geophys. Res., 115, D10104, doi:10.1029/2009JD013035, 2010.

Sutanudjaja, E. H., de Jong, S. M., van Geer, F. C., and Bierkens, M. F. P.: Using ERS spaceborne microwave soil moisture observations to predict groundwater head in space and time, Remote Sens. Environ., 138, 172–188, 2013.

Sutanudjaja, E. H., van Beek, L. P. H., de Jong, S. M., van Geer, F. C., and Bierkens, M. F. P.: Large-scale groundwater modeling using global datasets: a test case for the {Rhine-Meuse} basin, Hydrol. Earth Syst. Sci., 15(9), 2913–2935, doi:10.5194/hess-15-2913-2011, 2011.

Sutanudjaja, E. H., van Beek, L. P. H., de Jong, S. M., van Geer, F. C., and Bierkens, M. F. P.: Calibrating a large-extent high-resolution coupled groundwater-land surface model using soil moisture and discharge data. Water Resour. Res., 50, 687–705. doi:10.1002/2013WR013807, 2014.

Sutanudjaja, E. H., van Beek, L. P. H., Drost, N., de Graaf, I. E. M., de Jong, K., Peßenteiner, S., Straatsma, M. W., Wada, Y., Wanders, N., Wisser, D. and Bierkens, M. F. P.: PCR-GLOBWB 2.0: a 5 arc-minute global hydrological and water resources model, Geosci. Model Dev. Diss., 2016 (in prep).

Swenson, S. and Wahr, J.: Post-processing removal of correlated errors in GRACE data, Geophys. Res. Lett., 33(L08402), doi:10.1029/2005GL025285, 2006.

Swenson, S., Chambers, D., and Wahr, J.: Estimating geocenter variations from a combination of GRACE and ocean model output, J. Geophys. Res., 113(B08410), doi:10.1029/2007JB005338, 2008.

Tangdamrongsub, N., Steele-Dunne, S. C., Gunter, B. C., Ditmar, P. G., and Weerts, A. H.: Data assimilation of GRACE terrestrial water storage estimates into a regional hydrological model of the Rhine River basin, Hydrol. Earth Syst. Sci., 19, 2079–2100, doi:10.5194/hess-19-2079-2015, 2015.

Tangdamrongsub, N., Ditmar, P. G., Steele-Dunne, S. C., Gunter, B. C., and Sutanudjaja, E. H.: Exploring irregular flood events over Tonlé Sap basin in Cambodia using GRACE and MODIS satellite observations combined with altimetry observation and hydrological models, Remote Sens. Environ., 181, 162 – 173, http://dx.doi.org/10.1016/j.rse.2016.03.030, 2016.

Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F., and Watkins, M.: GRACE Measurements of Mass Variability in the Earth System, Science, 305 (5683), 503–505, 2004.

van Beek, L. P. H.: Forcing PCR-GLOBWB with CRU data, Technical Report, Department of Physical Geography, Utrecht University, Utrecht, The Netherlands, http://vanbeek.geo.uu.nl/suppinfo/vanbeek2008.pdf, 2008.

Van Beek, L. P. H. and Bierkens, M. F. P.: The Global Hydrological Model PCR-GLOBWB: Conceptualization, Parameterization and Verification, Technical Report, Department of

1050 Physical Geography, Utrecht University, Utrecht, The Netherlands,
1051 http://vanbeek.geo.uu.nl/suppinfo/vanbeekbierkens2009.pdf, 2009.

1052 van Beek, L. P. H., Wada, Y., and Bierkens, M. F. P. (2011). Global monthly water stress: 1.
1053 Water balance and water availability. Water Resour. Res., 47, W07517.
1054 doi:10.1029/2010WR009791.

1055 Vermote, E. F., Kotchenova, S. Y., and Ray, J. P.: MODIS surface reflectance user's guide,
1056 verion 1.3, http://modis-sr.ltdri.org/guide/MOD09_UserGuide_v1_3.pdf, 2011.

1057 Vörösmarty, C. J., Leveque, C., and Revenga, C.: Millennium Ecosystem Assessment
1058 Volume 1: Conditions and Trends, chap. 7: Freshwater ecosystems, Island Press, Washington
1059 DC, USA, 165–207, 2005.

1060 Wada, Y., Wisser, D., and Bierkens, M. F. P.: Global modeling of withdrawal, allocation and
1061 consumptive use of surface water and groundwater resources. Earth System Dynamics, 5, 15–
1062 40. doi:10.5194/esd-5-15-2014, 2014.

1063 Wahr, J., Molenaar, M., and Bryan, F.: Time variability of the Earth's gravity field:
1064 Hydrological and oceanic effects and their possible detection using GRACE, J. Geophys.
1065 Res., 103(B12), 30205–30229, 1998.

1066 Wahr, J., Swenson, S., and Velicogna, I.: Accuracy of GRACE mass estimates, Geophys. Res.
1067 Lett., 33, L06401, doi:10.1029/2005GL025305, 2006.

1068 Wanders, N., Bierkens, M. F. P., Jong, S. M., Roo, A., and Karssenberg, D.: The benefits of
1069 using remotely sensed soil moisture in parameter identification of large-scale hydrological
1070 models, Water Resour. Res., 50, 6874–6891, doi:10.1002/2013WR014639, 2014.

1071 Weerts, A. H. and El Serafy G. Y. H.: Particle filtering and ensemble Kalman filtering for
1072 state updating with hydrological conceptual rainfall-runoff models, Water Resour. Res., 42,
1073 W09403, doi:10.1029/2005WR004093, 2006.

1074 World Resources Institute (WRI): World Resources: A Guide to the Global Environment
1075 1998–99, World Resources Institute, Washington DC, USA, 1998.

1076 Yang, Y. S., Kalin, R. M., Zhang, Y., Lin, X., and Zou, L.: Multi-objective optimization for
1077 sustainable groundwater resource management in a semiarid catchment, Hydrolog. Sci. J., 46
1078 (1), 55 – 72, doi:10.1080/02626660109492800, 2001.

1079 Zaitchik, B. F., Rodell, M., and Reichle, E. H.: Assimilation of GRACE terrestrial water
1080 storage data into a land surface model: Results for the Mississippi basin, Amer. Meteor. Soc.,
1081 J. Hydrometeor., 9, 535–548, doi:10.1175/2007JHM951.1, 2008.

1082 Zheng, H., Wang, Z. J., Hu, S. Y., Malano, H., and ASCE, A. M.: Seasonal Water Allocation:
1083 Dealing with Hydrologic Variability in the Context of a Water Rights System, J. of Water
1084 Resour. Plann. Manage., 139, 76–85, 2013.

Zhu, J. F., Winter, C. L., and Wang Z. J.: Nonlinear effects of locally heterogeneous hydraulic conductivity fields on regional stream–aquifer exchanges, Hydrol. Earth Syst. Sci., 19, 4531–4545, 2015.

Table 1. PCR-GLOBWB model parameters related to the TWS estimate. Parameters are functions of spatial coordinates, except DDF which is a constant.

Parameter	Description	unit
$K_{sat,up}$	Saturated hydraulic conductivity of the upper soil storage	m/day
$K_{sat,low}$	Saturated hydraulic conductivity of the lower soil storage	m/day
SC_{up}	Storage capacity of the upper soil	m
SC_{low}	Storage capacity of the lower soil	m
f_g^{min}, f_f^{min}, f_p^{min}, f_{np}^{min}	Minimum soil depth fraction of grassland (g), forest (f), paddy irrigation (p), non-paddy irrigation (np)	-
f_g^{max}, f_f^{max}, f_p^{max}, f_{np}^{max}	Maximum soil depth fraction of grassland (g), forest (f), paddy irrigation (p), non-paddy irrigation (np)	-
J	Groundwater recession coefficient	1/day
DDF	Degree-day factor in the snow pack	^{o}Cm/day
KC^{min}	Minimum crop coefficient	-

Table 2. TWSV, SMSV and GWSV estimated annual amplitude (A, cm) and phase (P, month) in 4 different basins computed between April 2002 and December 2010. Areally averaged values for the entire Hexi Corridor are also given.

			Shiyang	Heihe	Desert	Shule	Areally-average
TWSV	GRACE	**A**	2.05 ± 0.31	1.49 ± 0.21	1.79 ± 0.23	1.21 ± 0.27	1.43 ± 0.18
		P	6.97 ± 0.29	6.80 ± 0.27	6.49 ± 0.24	8.61 ± 0.42	7.05 ± 0.24
	EnOL	**A**	1.35 ± 0.16	0.90 ± 0.07	0.66 ± 0.07	0.37 ± 0.06	0.70 ± 0.06
		P	6.35 ± 0.23	5.61 ± 0.14	5.80 ± 0.19	5.40 ± 0.31	5.74 ± 0.16
	EnKF 1D	**A**	1.61 ± 0.16	0.87 ± 0.10	1.05 ± 0.11	0.40 ± 0.11	0.80 ± 0.09
		P	6.96 ± 0.19	6.80 ± 0.22	6.47 ± 0.19	8.35 ± 0.51	6.92 ± 0.23
	EnKF 3D	**A**	1.49 ± 0.13	0.80 ± 0.08	0.72 ± 0.07	0.26 ± 0.09	0.72 ± 0.07
		P	6.42 ± 0.17	6.12 ± 0.19	6.40 ± 0.20	8.48 ± 1.02	6.44 ± 0.22
SMSV	EnOL	**A**	1.03 ± 0.11	0.70 ± 0.06	0.62 ± 0.07	0.31 ± 0.05	0.59 ± 0.06
		P	5.77 ± 0.20	5.60 ± 0.16	5.82 ± 0.21	5.03 ± 0.32	5.62 ± 0.18
	EnKF 1D	**A**	0.88 ± 0.09	0.75 ± 0.09	0.99 ± 0.11	0.36 ± 0.10	0.67 ± 0.08
		P	6.55 ± 0.21	7.01 ± 0.22	7.08 ± 0.21	8.47 ± 0.54	7.26 ± 0.24
	EnKF 3D	**A**	1.30 ± 0.10	0.66 ± 0.07	0.71 ± 0.08	0.12 ± 0.08	0.55 ± 0.07
		P	5.59 ± 0.15	6.25 ± 0.20	6.44 ± 0.20	8.19 ± 0.37	6.32 ± 0.22
GWSV	EnOL	**A**	0.50 ± 0.08	0.19 ± 0.03	0.02 ± 0.004	0.09 ± 0.01	0.12 ± 0.01
		P	7.84 ± 0.29	7.13 ± 0.26	5.43 ± 0.34	6.91 ± 0.29	7.22 ± 0.21
	EnKF 1D	**A**	0.65 ± 0.05	0.12 ± 0.03	0.01 ± 0.01	0.05 ± 0.01	0.10 ± 0.01
		P	8.69 ± 0.16	7.82 ± 0.40	7.91 ± 1.90	8.49 ± 0.29	8.32 ± 0.25
	EnKF 3D	**A**	0.70 ± 0.06	0.11 ± 0.02	0.02 ± 0.01	0.05 ± 0.01	0.10 ± 0.01
		P	8.52 ± 0.16	7.50 ± 0.31	7.76 ± 1.00	8.66 ± 1.33	8.26 ± 0.23

Table 3. Averaged values and standard deviations of precipitation and model parameters for 4 different basins.

	Shiyang	Heihe	Desert	Shule
Precipitation (mm/month)	21 ± 12	13 ± 12	11 ± 2	8 ± 6
$SC_{,up}$ (m)	0.08 ± 0.02	0.09 ± 0.02	0.09 ± 0.01	0.08 ± 0.01
$SC_{,low}$ (m)	0.33 ± 0.08	0.37 ± 0.07	0.35 ± 0.04	0.33 ± 0.08

Table 4. Statistical values of the GWSV computed from the in situ well measurement and GRACE DA estimates between January 2007 and December 2010. The average values are computed by averaging the estimated statistical values from all well locations.

		W1	W2	W3	W4	W5	Average value
Correlation coefficient [-]	EnOL	0.74	0.17	-0.04	-0.05	-0.53	0.06
	EnKF 1D	0.84	0.32	0.90	0.45	0.64	0.63
	EnKF 3D	0.82	0.49	0.85	0.51	0.83	0.70
RMS difference [cm]	EnOL	0.69	1.67	0.77	3.34	3.81	2.06
	EnKF 1D	0.58	1.63	0.40	2.56	2.58	1.55
	EnKF 3D	0.63	1.43	0.38	2.24	1.27	1.19

Table 5. Long-term trends and standard deviations of the in situ data and the DA estimates. The RMS difference (RMSD) between the in situ data and the DA trend estimates are also provided.

	W1	**W2**	**W3**	**W4**	**W5**	**RMSD**
In situ	-0.49 ± 0.03	0.01 ± 0.06	-0.60 ± 0.004	0.56 ± 0.12	-1.40 ± 0.03	0
EnOL	-0.57 ± 0.01	-0.64 ± 0.002	-0.01 ± 0.01	-1.69 ± 0.01	1.29 ± 0.02	1.62
EnKF 1D	-0.52 ± 0.02	-0.58 ± 0.04	-0.74 ± 0.02	-1.33 ± 0.08	-1.99 ± 0.13	0.93
EnKF 3D	-0.83 ± 0.02	-0.51 ± 0.03	-0.38 ± 0.01	-0.44 ± 0.08	-1.18 ± 0.06	0.54

Table 6. Statistical values of the streamflow computed from the river stream gauge measurement and GRACE DA estimates between April 2002 and December 2010. The average values are calculated by averaging the estimated statistical values from both gauge locations.

		G1	G2	Average value
Correlation coefficient [-]	EnOL	0.82	0.76	0.79
	EnKF 1D	0.84	0.77	0.81
	EnKF 3D	0.84	0.78	0.81
NS coefficient [-]	EnOL	0.65	0.56	0.61
	EnKF 1D	0.69	0.57	0.63
	EnKF 3D	0.69	0.57	0.63
RMS difference [cm]	EnOL	5.49	3.09	4.29
	EnKF 1D	5.18	3.08	4.14
	EnKF 3D	5.23	3.04	4.14

Table 7. TWSV, SMSV, GWSV, and precipitation estimated long-term trends in 4 different basins computed between April 2002 and December 2010. Areally averaged values for the entire Hexi Corridor are also given.

		Shiyang	Heihe	Desert	Shule	Areally-average
TWSV (cm/yr)	GRACE	-0.73 ± 0.04	-0.64 ± 0.03	-0.72 ± 0.03	-0.34 ± 0.04	-0.59 ± 0.03
	EnOL	0.30 ± 0.15	0.24 ± 0.09	0.20 ± 0.04	0.18 ± 0.06	0.22 ± 0.07
	EnKF 1D	-0.72 ± 0.08	-0.41 ± 0.04	-0.33 ± 0.05	-0.34 ± 0.04	-0.39 ± 0.07
	EnKF 3D	-0.36 ± 0.02	-0.21 ± 0.02	-0.11 ± 0.03	-0.25 ± 0.03	-0.20 ± 0.03
SMSV (cm/yr)	EnOL	0.38 ± 0.05	0.21 ± 0.02	0.17 ± 0.03	0.14 ± 0.02	0.19 ± 0.02
	EnKF 1D	-0.11 ± 0.03	-0.20 ± 0.01	-0.29 ± 0.04	-0.22 ± 0.04	-0.23 ± 0.03
	EnKF 3D	0.10 ± 0.03	-0.12 ± 0.01	-0.12 ± 0.02	-0.14 ± 0.01	-0.11 ± 0.004
GWSV (cm/yr)	EnOL	-0.08 ± 0.12	0.03 ± 0.07	0.02 ± 0.007	0.04 ± 0.02	0.02 ± 0.04
	EnKF 1D	-0.61 ± 0.01	-0.16 ± 0.004	-0.01 ± 0.005	-0.12 ± 0.02	-0.16 ± 0.02
	EnKF 3D	-0.39 ± 0.01	-0.09 ± 0.003	0.01 ± 0.004	-0.11 ± 0.001	-0.11 ± 0.002
Precipitation ((cm/month)/yr)		0.04 ± 0.01	0.04 ± 0.01	0.05 ± 0.01	0.02 ± 0.01	0.04 ± 0.01

1125

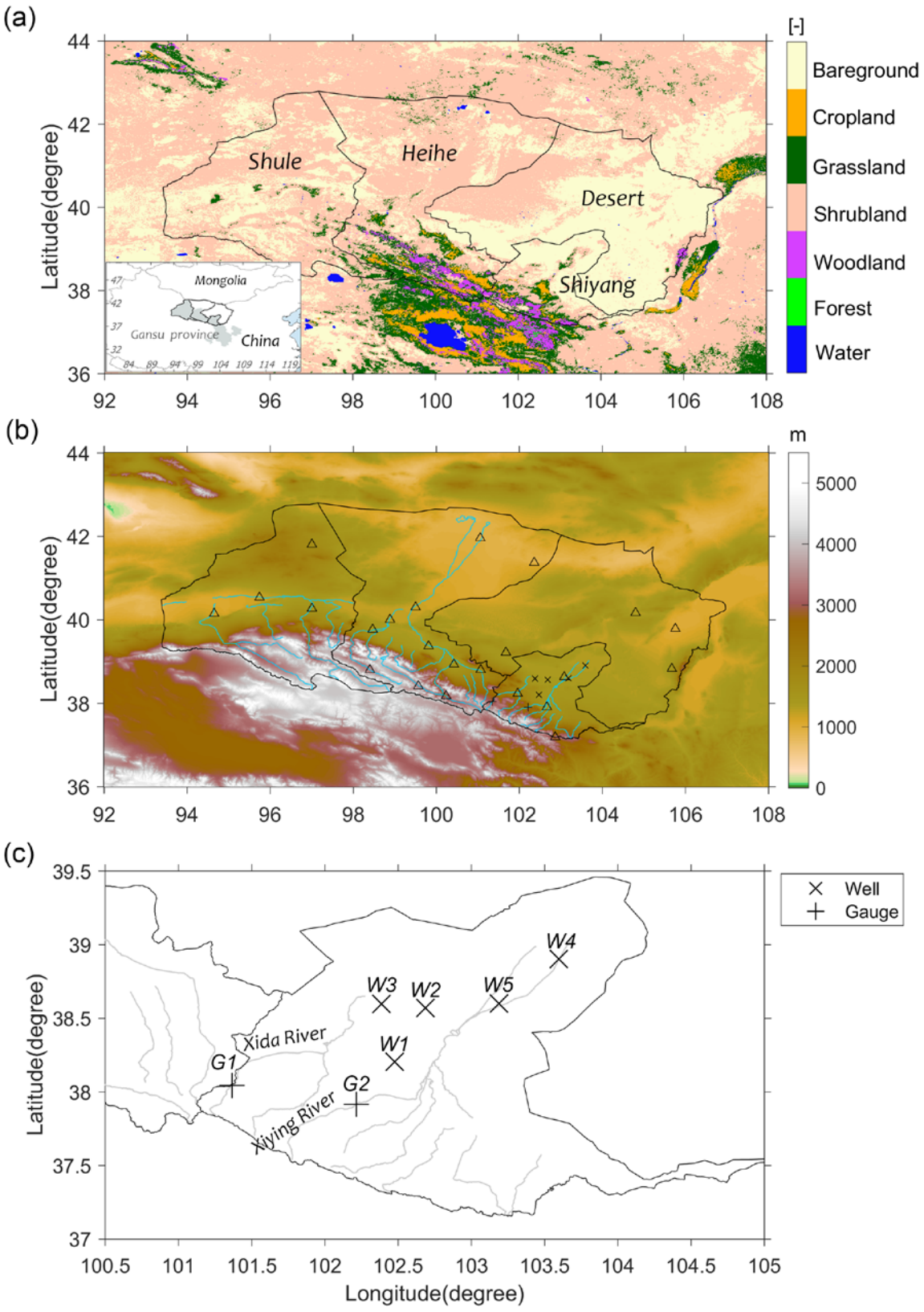


Figure 1. Geography of the Hexi Corridor. (a) Land cover and division into individual regions (Shiyang River Basin, Heihe River Basin, Shule River Basin, and a Desert), (b) Topography and locations of the local meteorological stations (triangles), (c) Zoom-in on the Shiyang River Basin, showing the locations of considered groundwater wells (x) and river stream gauges (+).

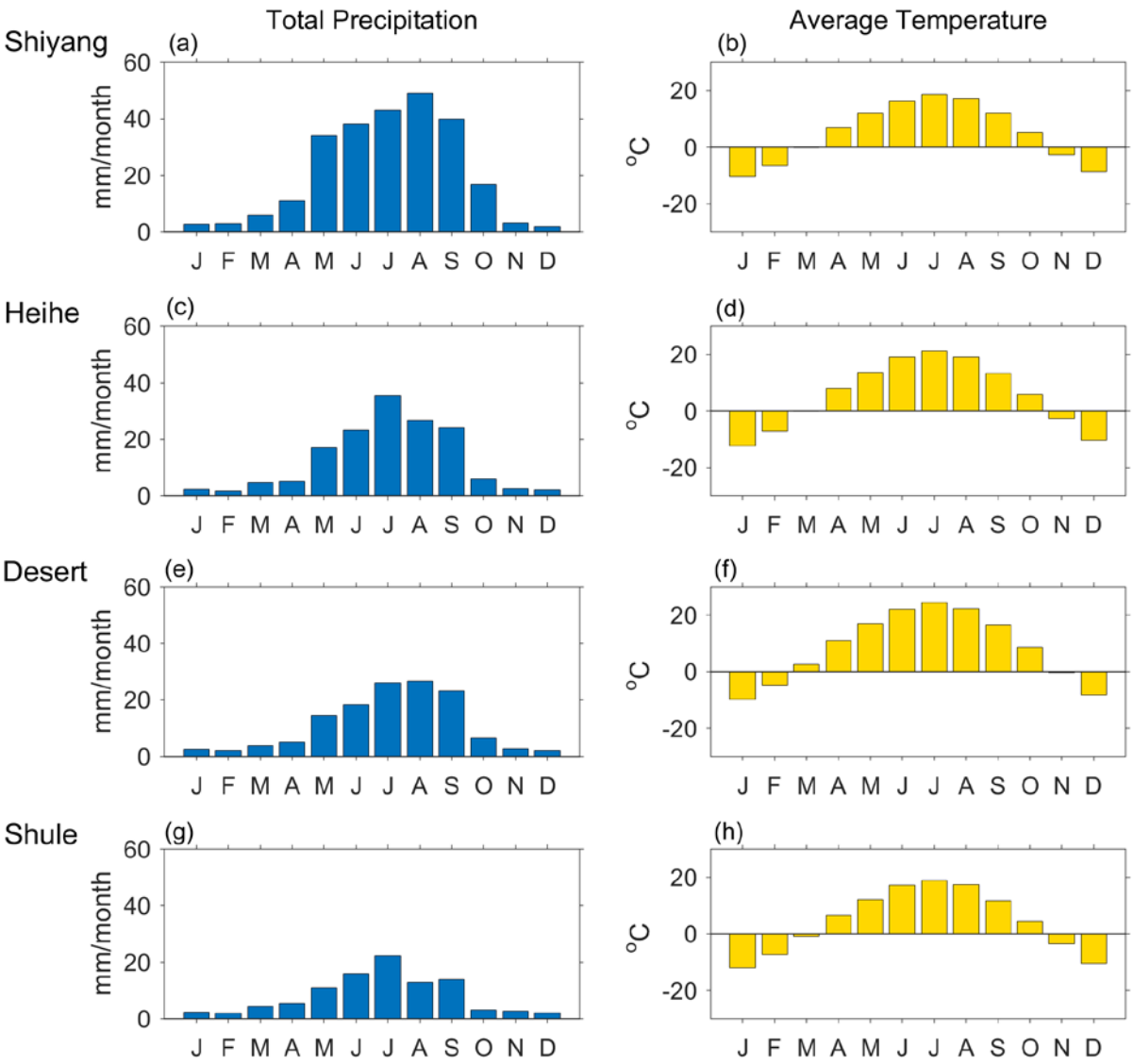
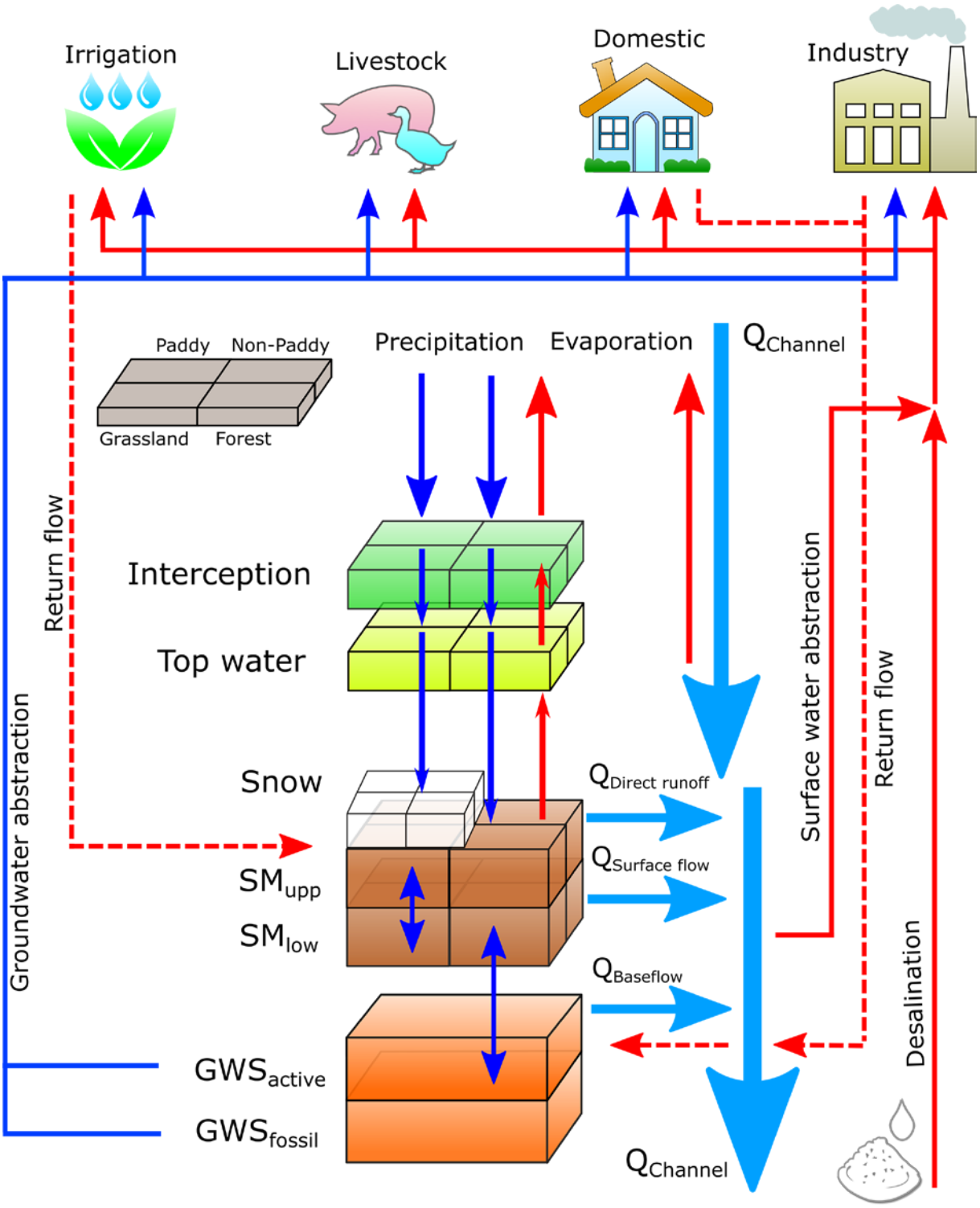
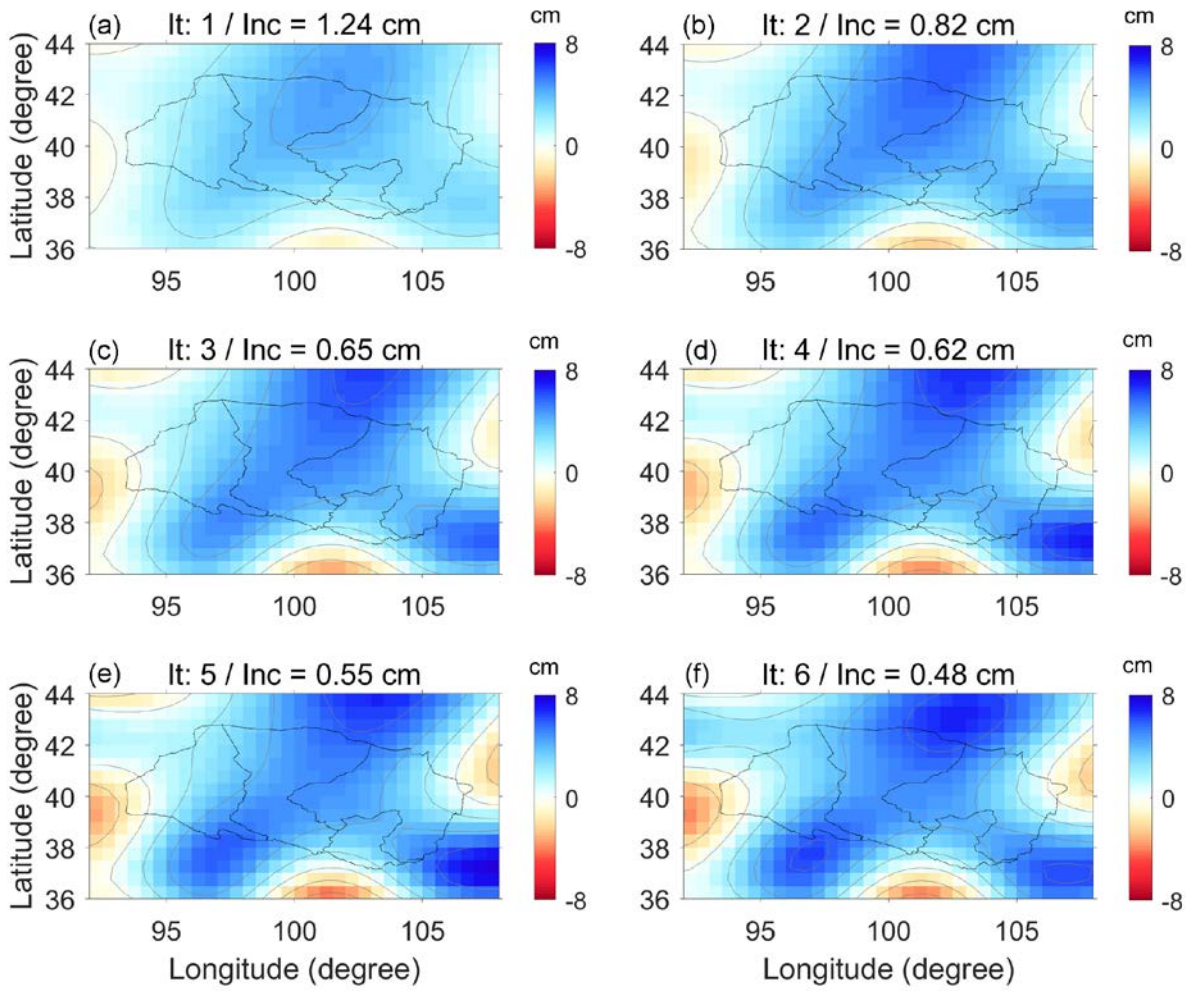


Figure 2. Monthly total precipitation and averaged temperature over 4 regions of the Hexi Corridor.



1136

1137 **Figure 3**. The structure of PCR-GLOBWB hydrological model.



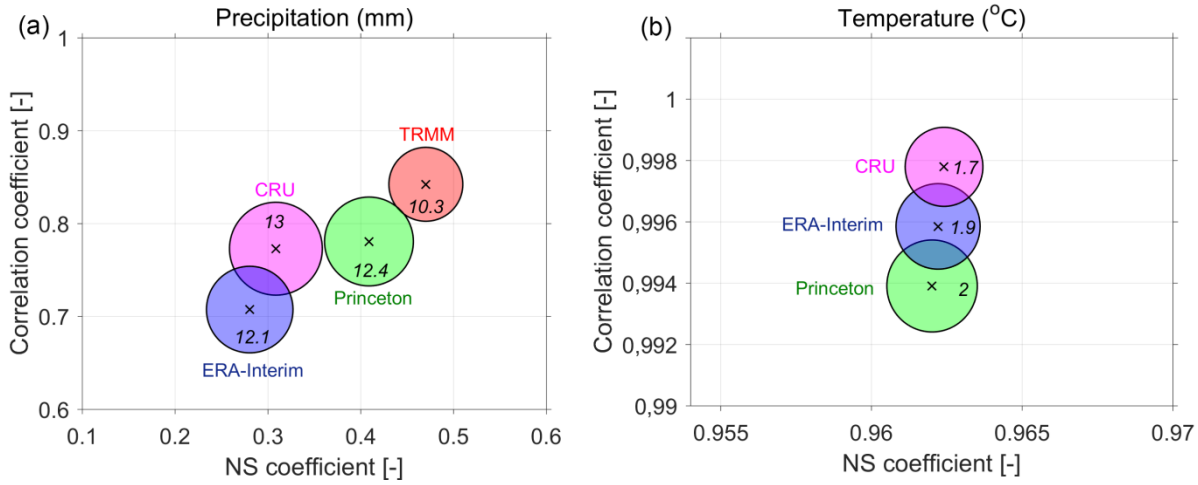
1138

1139 **Figure 4**. GRACE-derived TWS variation of October 2002. The signal restoration was
1140 applied to restore the signal mitigated by the applied Gaussian filter. After each iteration (It),
1141 the increment in each cell was computed. The procedure was stopped after six iterations,
1142 when the maximum increment (Inc) was lower than 0.5 cm (f).

1143

1144

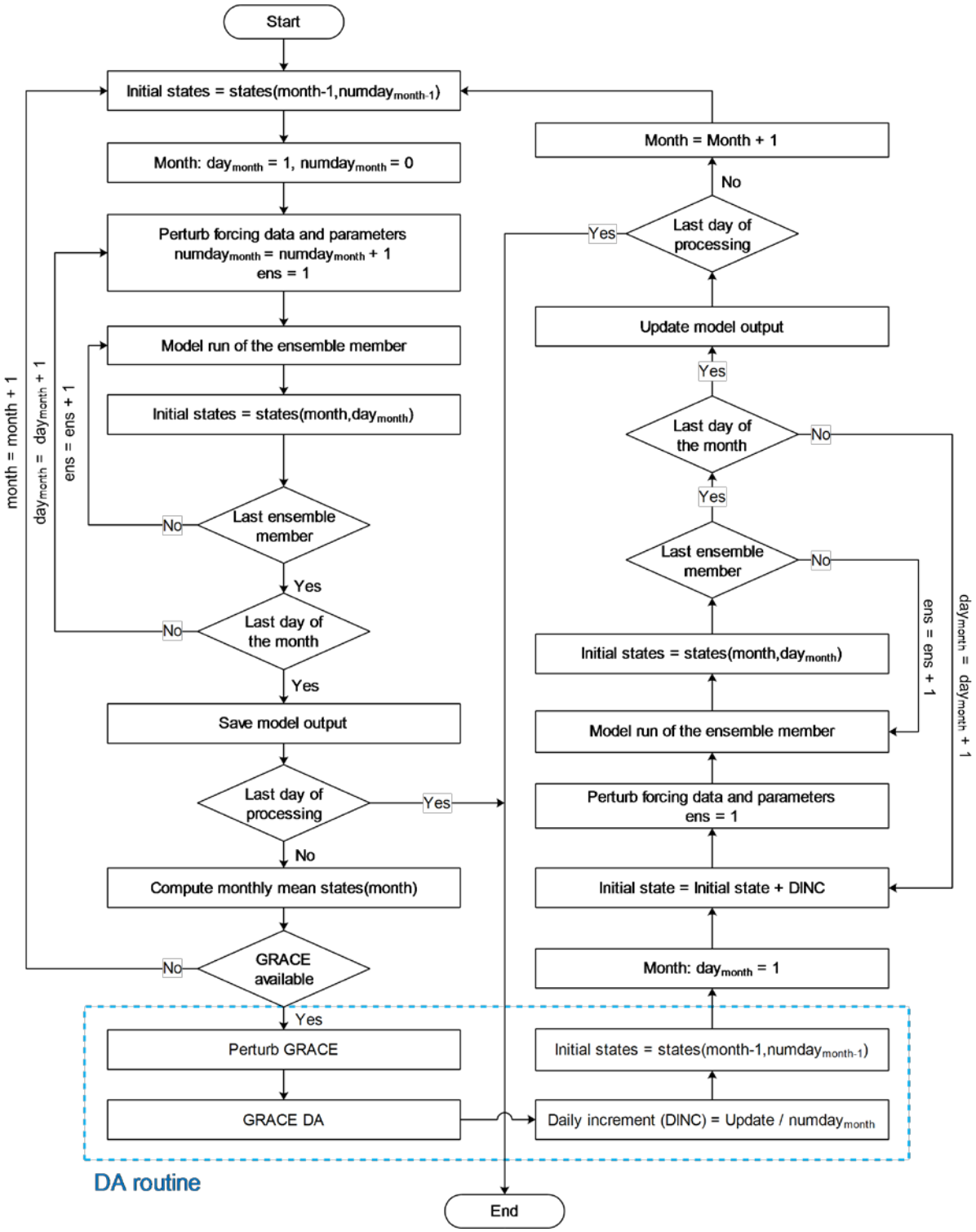
1145



1146

1147 **Figure 5**. The correlation coefficient, NS coefficient, and RMS difference computed between
1148 the local and different global forcing data. The RMS difference is shown as the radius of the
1149 circle (also explicitly provided as the number).

1150



1151

1152 **Figure 6**. DA diagram representing the disaggregation of monthly averaged TWS from
1153 GRACE into the daily PCR-GLOBWB state estimates.

1154

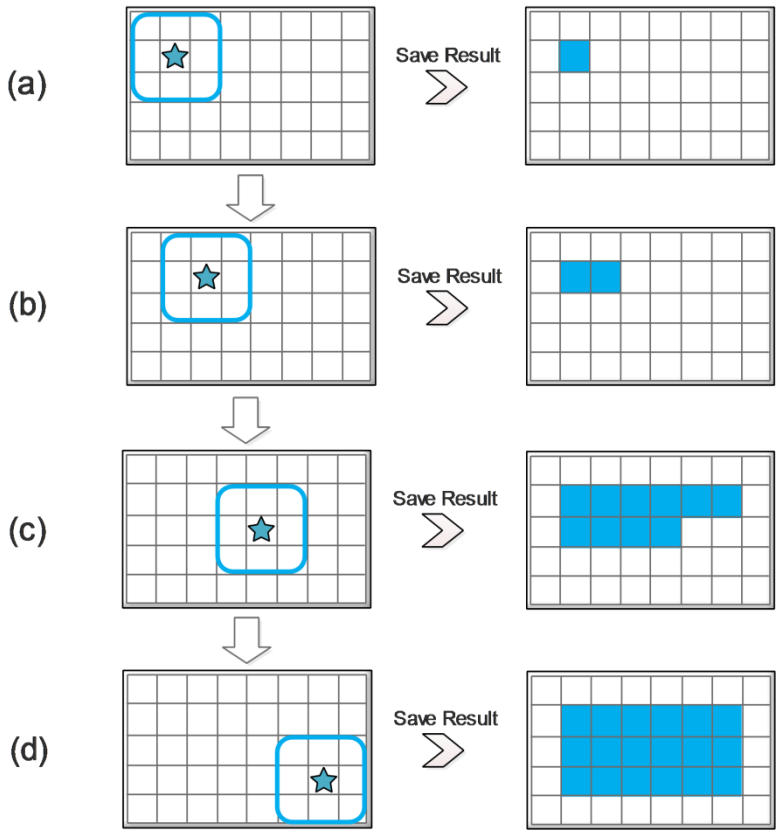


Figure 7. Demonstration of EnKF 3D scheme, accounting for the spatially-correlated errors. For a centre grid cell, the state and observation matrices contain all TWS-related components of the neighbouring grid cells and the centre grid cell (left). The graphic demonstrates the case of one pixel (0.5 degree) correlation distance. The boundary stretches farther for larger correlation distance. The covariance matrices $\mathbf{P_e}$ and $\mathbf{R}$ are computed based on the data from these grid cells. Then, the EnKF is applied and the states of the centre grid cell are updated (right). The procedure is repeated through all grid cells.

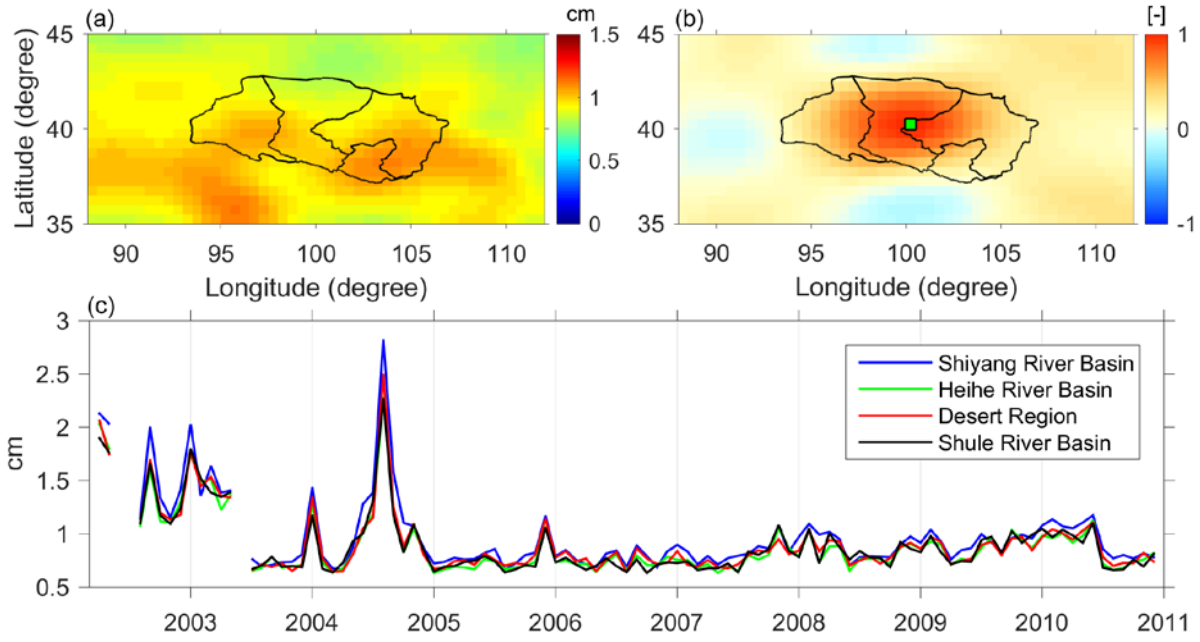


Figure 8. Some statistics of errors in GRACE-derived TWS variation over the Hexi Corridor. The standard deviation (a) and the correlation coefficient with respect to the green point (b) for a sample month, October 2002, are shown in the top. The time-series of averaged standard deviation computed over four different basins are shown in the bottom plot (c).

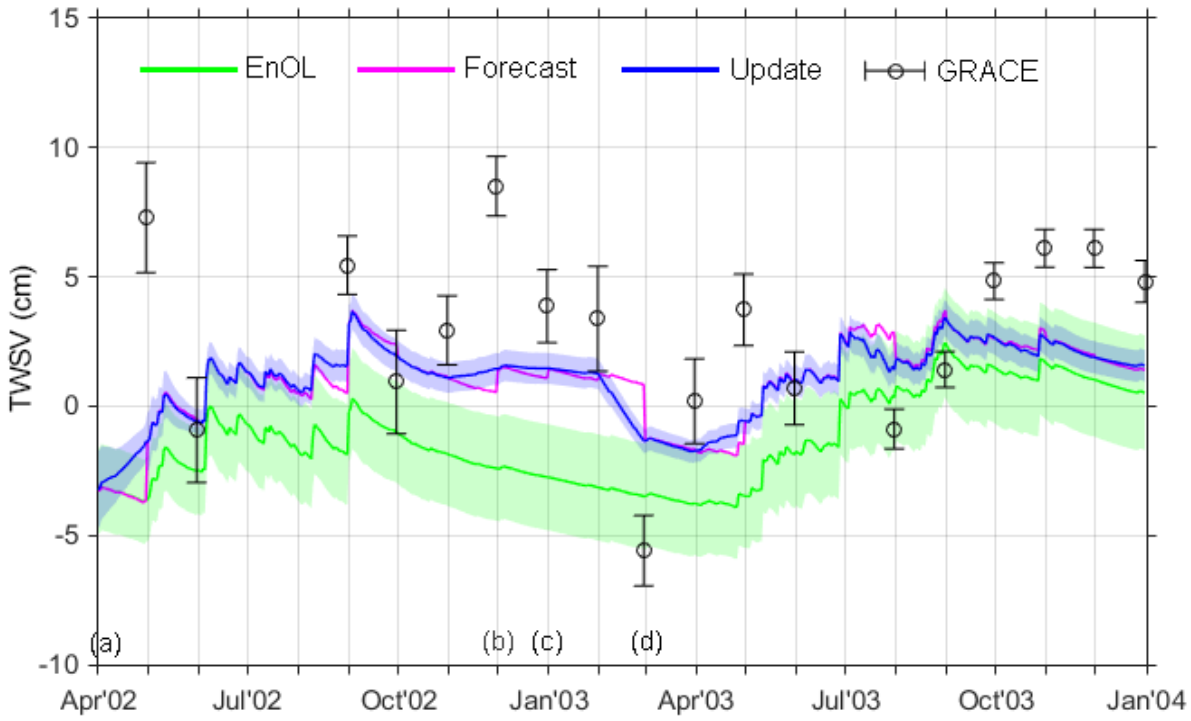


Figure 9. Daily TWS variations estimated between 1 April 2002 and 31 December 2003, averaged over Shiyang River Basin. The mean value of the ensemble is given as the solid line, and the standard deviation is shown as the shaded envelope. The TWS estimates from model only (EnOL), GRACE DA forecast (EnKF before the update), GRACE DA update (EnKF after update), and GRACE observations are shown. The x-axis labels represent the first day of the month. Some features of the DA scheme regarding the identical TWS estimate seen at the beginning of the update (point a) and the observed spurious jumps (point b,c,d) are also shown.

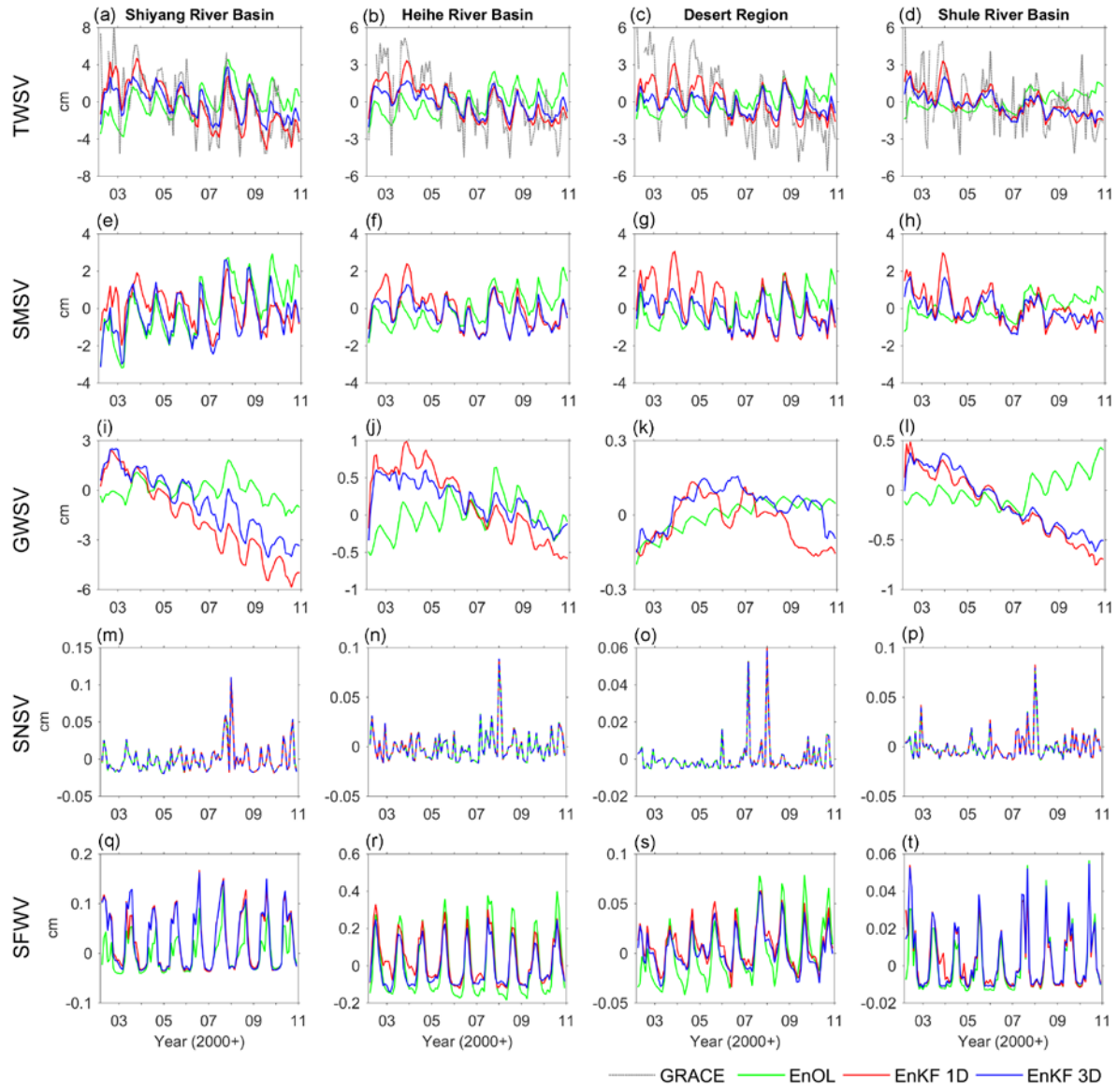


Figure 10. Monthly TWSV, SMSV, GWSV, snow water storage variation (SNSV), and surface water storage variation (SFWV) estimated between April 2002 and December 2010 from the EnOL, EnKF 1D, EnKF 3D, and GRACE observations over 4 basins.

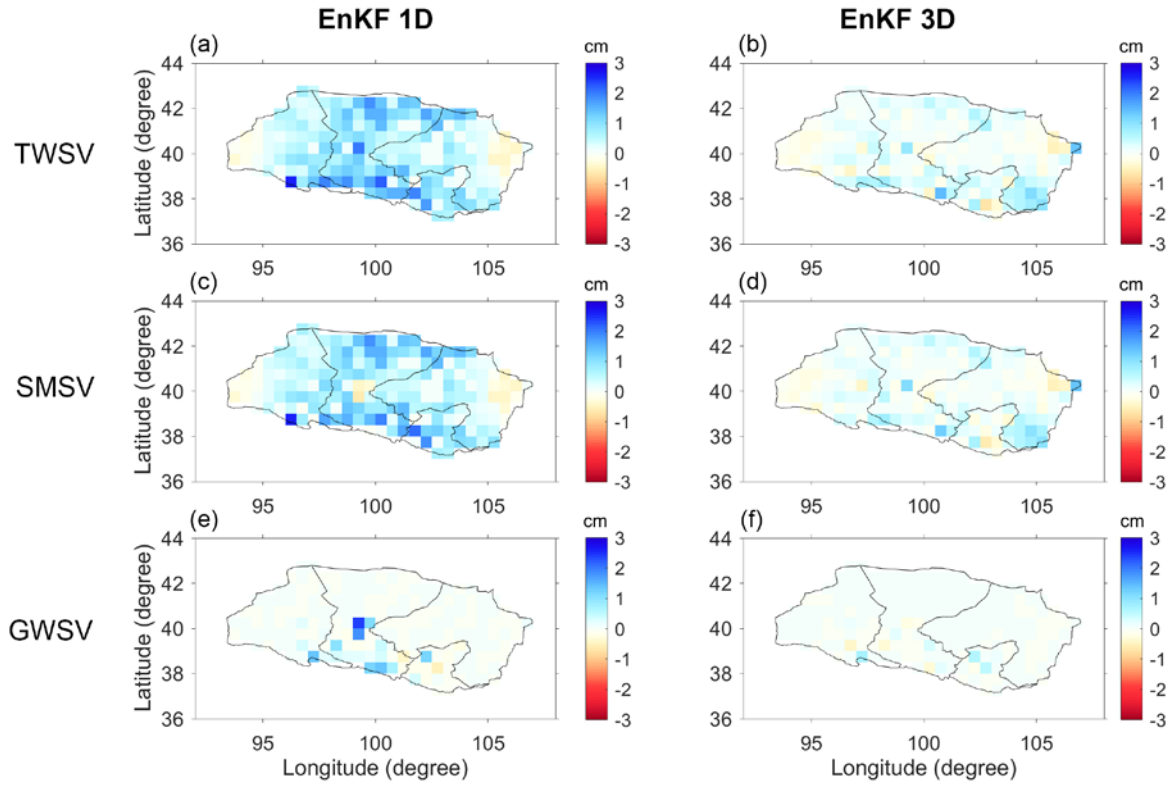
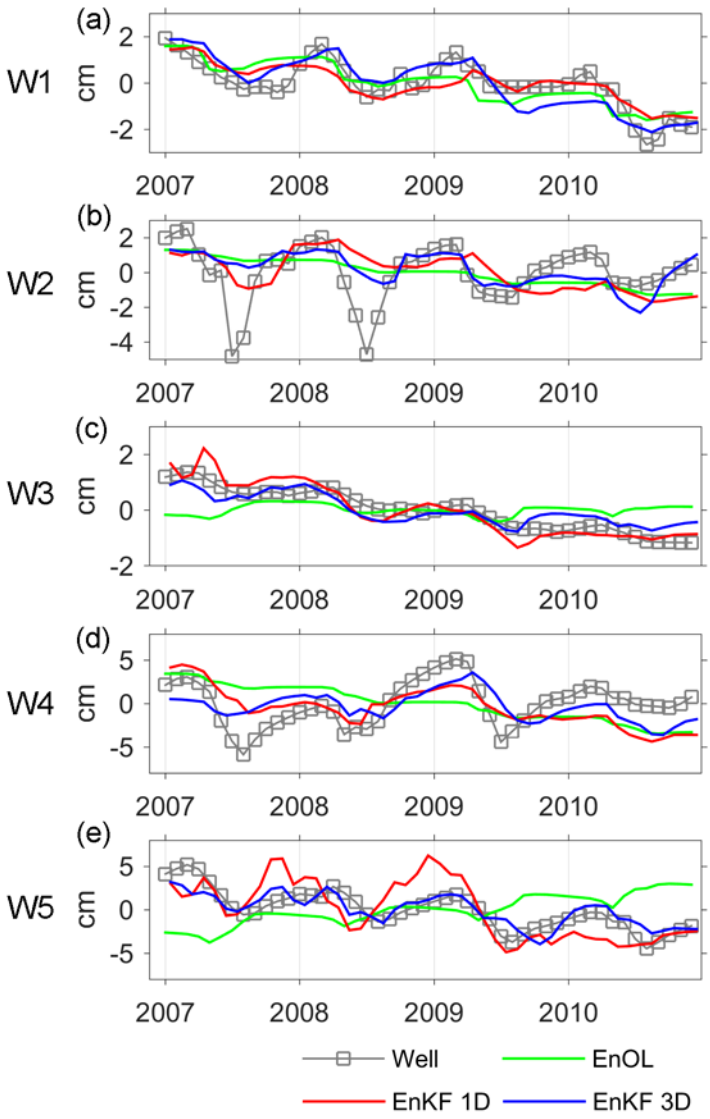


Figure 11. TWSV, SMSV, and GWSV updates of October 2002 without the correlation error applied (EnKF 1D) and with the correlation error applied (EnKF 3D).



1201

1202 **Figure 12**. Monthly GWS variation estimates from the in situ well measurements, as well as
1203 EnOL, EnKF 1D, and EnKF 3D results, between January 2007 and December 2010 at 5
1204 groundwater well locations. The chosen period is based on the availability of the well data.

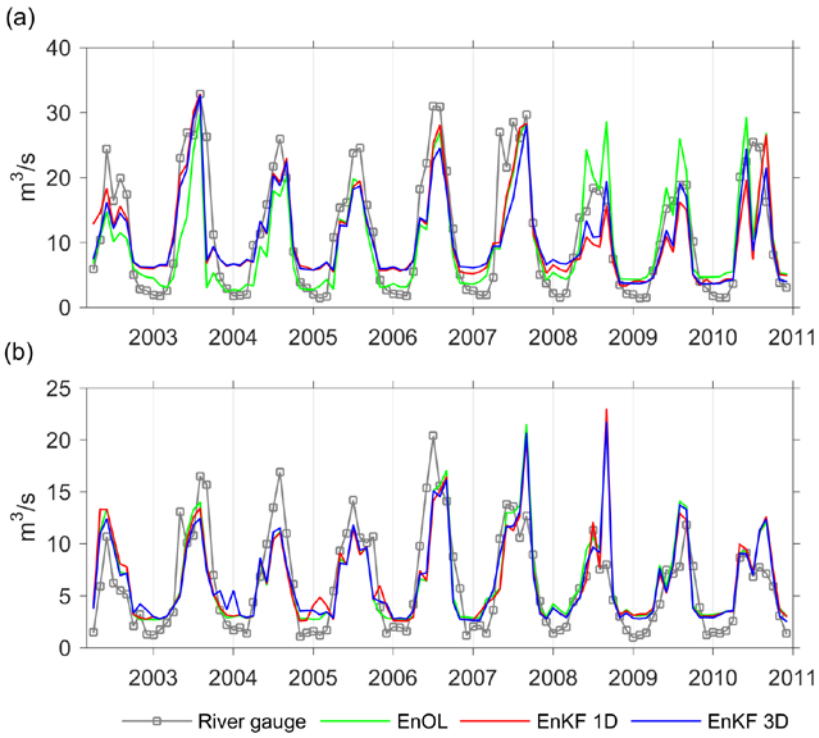


Figure 13. Monthly streamflow estimates from the in situ river gauge measurements, as well as EnOL, EnKF 1D, and EnKF 3D results, between April 2002 and December 2010 at 2 river gauge locations, G1 (a) and G2 (b).

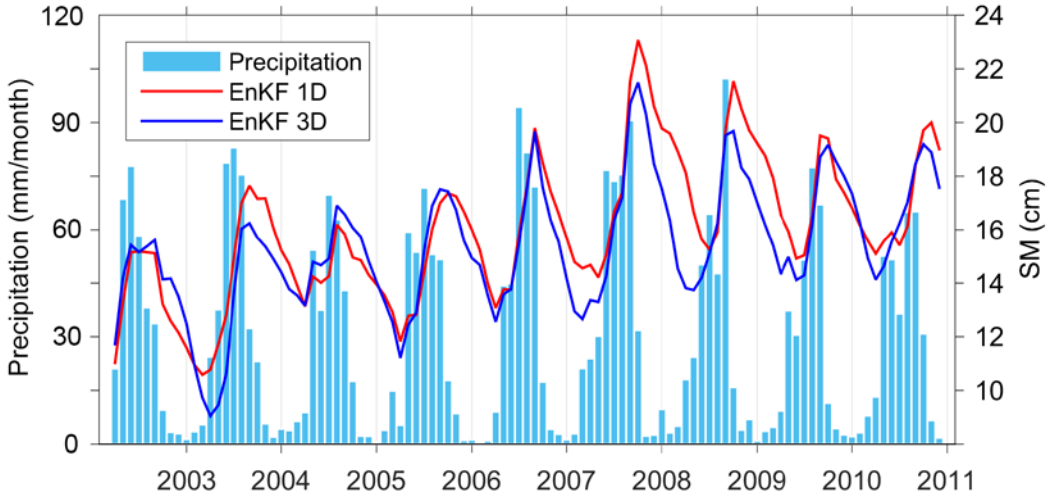


Figure 14. Monthly total precipitation (mm/month) and SM storage estimates (cm) from EnKF 1D and EnKF 3D results at river gauge G2 location.

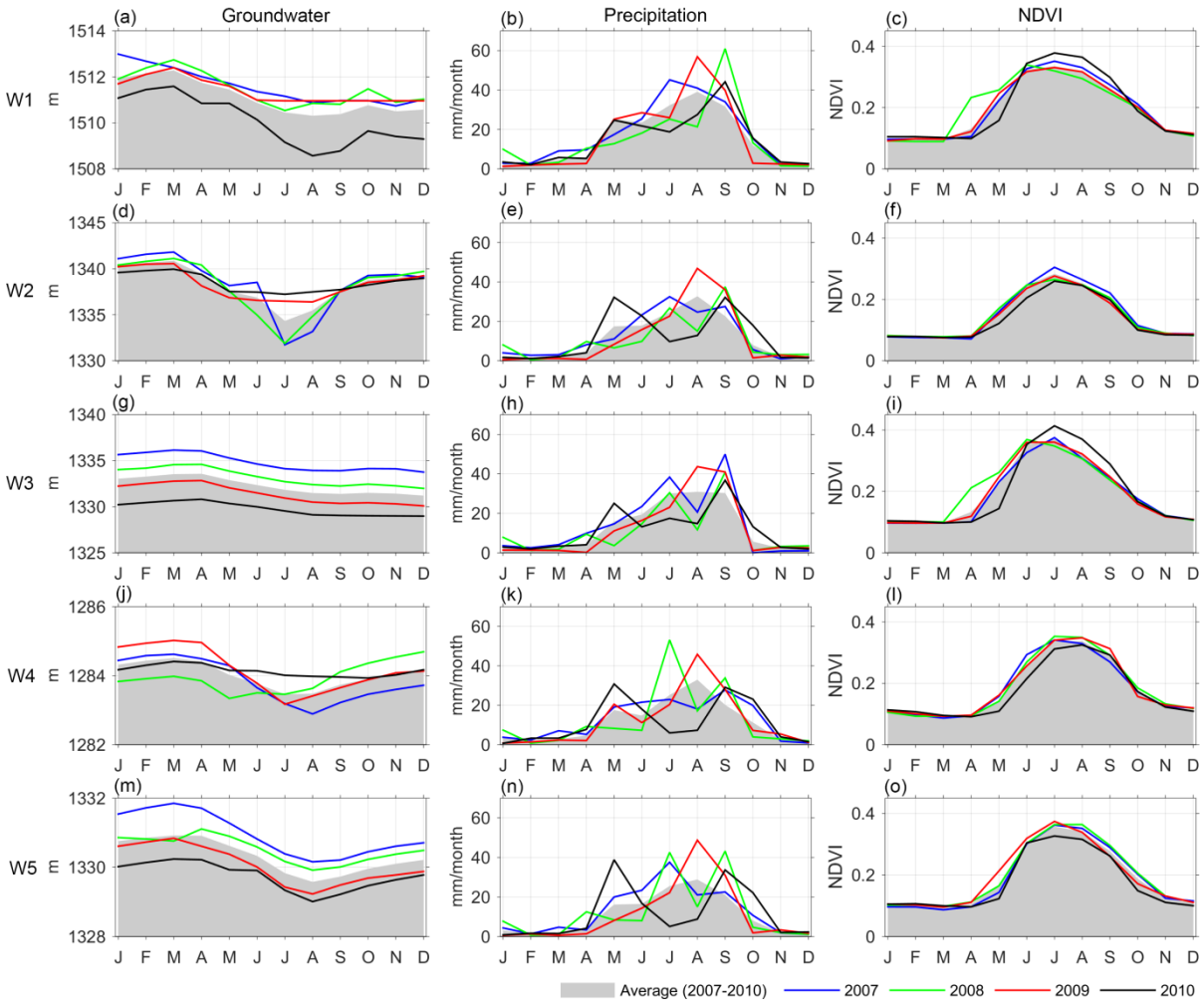


Figure 15. The monthly averaged groundwater head measurement (left), total precipitation (middle) and NDVI (right) for five groundwater well locations. Precipitation and NDVI data are reported as the average values within the circular areas of the 10-km radius. The long-term average values between January 2007 and December 2010 are shown in the grey shed, and the values in 2007, 2008, 2009, and 2010 are shown as blue, green, red, and black lines, respectively. The period is chosen based on the availability of the well data.