Interactive comment on "GRACE storage-streamflow hystereses reveal the dynamics of regional watersheds" *by* E. A. Sproles et al.

Anonymous Referee #1

Received and published: 5 December 2014

General comment:

The paper addresses:

1. The possibility to forecast runoff at certain times during summer from terrestrial water storage in spring measured by GRACE.

2. The behavior of TWS and groundwater storage during the seasonal cycle

This study uses GRACE observations of terrestrial water storage observations to expand upon a fundamental concept in watershed hydrology – that the temporal relationship between storage and runoff can be used to quantify complex watershed behavior

at broad scales, including groundwater recharge amounts and timing, baseflow recession characteristics, and long lead-time streamflow prediction. The methods is implemented over three catchments of minimum size (for the application of GRACE data) and similar climatic conditions

Interesting highlights in this study:

1. Clock-wise behavior of groundwater hysteresis in the presence of a counter clockwise behavior of the hysteresis for the GRACE signal.

2. Prediction of seasonal runoff using GRACE data

However, the focus of the presented results is not reflected in the title of the paper nor in the abstract. Both, title and abstract claim a much more general result with respect to the dynamic behavior of systems. This is not covered in the presented investigations. The claimed methodology cannot be applied to other regional scale studies. As the authors mentioned in the text, the paper is based to a large extend on the work by Riegger & Tourian 2014, who have investigated and modelled the runoff–storage(R-S) relationship for large scale catchments in different climatic regions in detail.

Here in this paper there is a lot of text claiming an explanation for system dynamics by hysteresis only in a qualitative way, mainly repeating the results of Riegger & Tourian yet not being supported by own investigations. There is also a lengthy text trying to explain the deviations from the expected results by anthropogenic management, pumping and local groundwater level measurements.

The investigations presented here are not performed in a sound scientific way and are insufficiently documented, because

1. For an assessment of the forecast potential a comparison of predicted and measured discharge is needed. This comparison should be supported by reporting appropriate metrics like RMSE, Nash-Sutcliffe and correlation (The claimed correlation coefficients seem to correspond to the fitting curves and not to the observed data).

2. No reasons for the selection of points in time are given, neither for the use of TWSA (March) nor for the predictions of Qseason and QAug. No alternatives months are investigated or discussed.

3. If there are anthropogenic impacts in the chosen catchments, why isn't other catchment or time periods used for which no anthropogenic effects occur?

4. Details in the calculation of GWSA are not given nor are the corresponding data sets (time series of mean GW-level, recharge, soil moisture, snow water equivalent, reservoir volumes, TWSA, discharge) displayed. Thus the calculation steps and conclusions with respect to the GWSA hysteresis cannot be retraced. An appropriate visualization of time series of different compartments is needed.

5. In the comparison of TWSA vrs point specific well data data (Fig.7) no explanation for the inconsistency of GWSA and mean groundwater level is given. Instead further detailed studies are proposed.

6. The authors should clearly describe the behavior of the groundwater system GWSA compared to TWSA and provide possible physical reason for the behavior. For a better understanding, one might think of showing the hysteresis of slow and fast discharge vrs precipitation and different storage compartments like soil moisture, snow water equivalent etc.

In addition:

7. Results and conclusions need to be related to the presented investigations

8. Conclusions have to be explained in detail.

9. Title and abstract have to reflect the actual investigations

Based on the aforementioned points, the paper can only be accepted after a major revision, in which the aforementioned points are taken into account.

Specific Comments:

12029 L14

The correlation values seem to correspond to data and fitting curves rather than to predicted and measured runoff.

12029 L15

This is very general sentence. In fact to apply the same methodology, one should characterize each basin individually. Indeed, this characterization sometimes becomes cumbersome due to the heterogeneous behavior of many large scale basins. In fact, the prerequisite of applying this method is not only availability of GRACE data. So please be precise in your statement.

12032 L8

There might be confusion in the citation of the respective publication of Reager et al. 2014: In the cited paper the relationship between regional water storage and specific streamflow is not addressed neither the corresponding hysteresis. In the cited paper GRACE TWSA is just used as additional forcing term in an auto-regression approach. Possibly it is from another publication, please cite the correct paper.

12032 L10

In the paper of Riegger and Tourian 2014 the hysteresis between runoff and storage is described in detail for different climatic zones. For boreal regions they show that runoff is linear to coupled liquid storage. They claim that beside the time lag for runoff the uncoupled solid components of storage are responsible for the hysteresis. However their calculation of runoff from GRACE mass is based on a homogeneous distribution of aggregated snow mass and is not directly applicable to mountainous areas as investigated here. In mountainous areas the snow mass distribution very much depends on local conditions like topography, elevation etc.

12035 L2

Please report the scale factors of studied basins.

12035 L2

How the leakage is quantified. The scale factor would only deal with signal attenuation. Leakage is much more complicated to be quantified by a simple scale factor. Please also describe the measurement error as well. Does the measurement error come from propagating the calibration error of spherical harmonic coefficients?

12036 L9-17

The simultaneous display of TWSA and GWSA in one figure for each catchment might help to highlight the different dynamical behavior.

12036 L22

Similar confusion in the citation of the respective publication of Reager et al. 2014 is seen as above (12032 L8).

12036 L22

In order to describe the systems independently from catchment area I propose to use runoff rather than discharge for a comparison

12037 L25 – 12038 L8

To support this paragraph I suggest to show time series of soil moisture or distribution of snow coverage in time.

12039 L14

Fig.3b shows the total runoff i.e. surface runoff and baseflow vrs. GWSA. Thus total runoff should be separated into its fast and slow components and each of them being displayed vrs. GWSA.

The Clock-wise behavior of groundwater hysteresis is one of interesting finding of this study. Therefore, the authors should provide a physical explanation for that. Explana-

tion of Fig 3b is not clear: Vertical branch Jun-Oct: How can runoff decrease with a nearly constant GW storage? Left branch Oct – Mar: how can runoff increase with a decreasing GW storage? Is it matter of surface runoff? For an understanding of the behaviour it is essential to display time series of SM, SWE, RES, TWSA and RGW and RSW. A display of precipitation and evapotranspiration would be very helpful.

12039 L19

Confusion about Fig. numbering: possibly Fig 7 is meant

12039 L21–23

To me taking TWSA_march for prediction looks like cherry picking. Why not March and not February for TWSA or June, July for discharge, for instance? Please provide reason for taking this month!

Also, confusion about Fig. numbering: possibly Fig 6 is meant It is not clear from the text how stream flow is predicted and how this is related to Fig6: Qseason in Fig6: is it the mean observed discharge or predicted discharge? If it is the later, how is it calculated?

Are the correlations displayed in Fig6 and Table 1 the correlations between the measured Qseason and the fitted curves (power functions with which parameters?) or between the measured Qseason and TWSA (the high correlation value rather represents the curve fit than Qseason vrs TWSA).

The fitting curves in Fig6 do not already represent predictions, yet are the basis for predictions using measured TWSA to determine forecast discharge (via the curve fits) (as Fig6 shows, that Qseason and QAug are not very much depending on TWSA for smaller values of TWSA, yet only for bigger values.The calculation scheme is essential for an assessment of the method)

For an evaluation of the predictive potential of the investigated methods the parameters of the fitting curves determined on a training period should be used to calculate predic-

tions of discharge in an independent prediction period. The predicted discharge values should then be displayed versus the measured for the prediction period in a scatter plot and correlations should be calculated for forecasts vrs measured.

As a predicted value should always be better than a simple use of mean monthly values from the training period of forecasts, on top of conventional Nash Sutcliffe (NS) coefficient, in which the values are assessed w.r.t long term mean, the NS coefficient w.r.t. the seasonal signal (using the monthly residuals of the training period) should also be presented in table1 (NS_cycle: in the denominator instead of $\bar{Q_o}$ you should use monthly mean).

12040 L1

Again why August and no other month? How a reader should follow the story here? What does a seasonal average/aggregation of discharge mean for possible applications?

12041 L5–29

Over boreal regions the R-S hysteresis is determined by (Riegger and Tourian 2014):
- Climatic impacts i.e. the relative importance of aggregated solid precipitation (represented on the lower branch)

- The runoff time constant determining the slope of the linear part (upper branch)

- The time lag between mass and runoff being responsible for (a smaller) part of the hysteresis

The different forms of the hysteresis thus can be explained by the corresponding hydraulic time constants, which is shortest for steep slopes and fractured systems. This explains that the upper branch is steeper for the Upper Columbia than for the Snake River. This should be considered and discussed in this section.

12042 L10–17

See comment above on prediction accuracy in table 1 There are 9 years of measurements available. This period could be split into a training and a prediction period for a better estimation of prediction accuracy

12042 L18

Probably Fig 6, and Fig 7 is meant

12042 L21-26

The fact that Q is insensitive to TWSA < 100mm is reflected by the curves in Fig 6. This means that prediction could only be made for TWSA > 100mm. If these catchments are managed, a reliable prediction from TWSA cannot be made!! So either other, unmanaged catchments have to be chosen or the authors should only consider the time periods with no management. The explanation of the Q – TWSA relationship for TWSA < 100mm by water resources management is not sufficient to explain quantitative effects.

12042 L21-26 –12043 L11

Lengthy text, clear quantitative consequences are missing

12043 L16-19

Explanation of GWSA-Q not understood (see also above)! During the winter period in snow covered areas discharge is released only from the groundwater system. The groundwater recharge from the surface in this case is zero, i.e. GWSA should decrease with Q. How is distinguished between surface runoff from snow melt and runoff from groundwater. Possibly the behavior of the model based calculation of soil water content helps.

12043 L25

GWSA is nearly constant from June to October. Why? This does not fit to the timing of pumping test! What is the purpose of the pumping tests mentioned in this paper?

12044 L3

Probably Fig 7 instead of 6 Meaning and conclusion of Fig 7

12044 L3–10

GWSA does not fit to the overall GW-levels from observations. So either the calculated GWSA is wrong or the selection of observation wells is not representative for the general GW level. GWSA correspond to the total volume of the groundwater system and not to groundwater levels! The GW-storage coefficient determines the relationship between volume and level and is not mentioned here as it is probably not known on catchment scale. There is no scientific consequence from this observations mentioned in the paper. If this part is presented here, there should be a more detailed description of GW-level measurements, storage coefficients from hydrogeology, selection of observation points and a detailed discussion of the results and consequences for the message of the paper.

12045 Conclusions

The conclusion should represent the conclusion of this work. In case the authors would bring in arguments from other published works the bridges between studies should be clear. The last paragraph of conclusion is too general and does not reflect the results of the study.

12045 L10–13

Please report the prediction results here including RMSE, NS, NS_cycle, correlation

12045 L23 Please provide a citation for the background research in the text, otherwise please rephrase the sentence as this statement contradicts with your earlier statement in 12040 L22.