GRACE storage-runoff hystereses reveal the dynamics of regional watersheds

E.A. Sproles[1,2,3]; S.G. Leibowitz[4]; J.T. Reager[5]; P.J. Wigington Jr[4,6]; J.S. Famiglietti[5]; S.D. Patil[7]

*corresponding author: eric.sproles@gmail.com

[1] Oak Ridge Institute for Science and Technology, c/o National Health and Environmental Effects Research Laboratory, US Environmental Protection Agency, 200 SW 35th Street, Corvallis, OR, 97333, USA

[2] Current address: Centro de Estudios en Zonas Áridas, Universidad de La Serena, Raul Bitran 1305, La Serena, Chile

[3] College of College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, OR, 97331-5503, USA

[4] National Health and Environmental Effects Research Laboratory, US Environmental Protection Agency, 200 SW 35th Street, Corvallis, OR, 97333, USA

[5] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, 91109, USA

[6] Retired

[7] School of Environment, Natural Resources and Geography, Bangor University, Bangor, LL57 2UW, Wales

Abstract

We characterize how regional watersheds function as simple, dynamic systems through a series of hysteresis loops using measurements from NASA's Gravity Recovery and Climate Experiment (GRACE) satellites. These loops illustrate the temporal relationship between runoff and terrestrial water storage in three regional-scale watersheds (>150,000 km^2) of the Columbia River Basin, USA and Canada. The shape and size of the hysteresis loops are controlled by the climate, topography, and geology of the watershed. The direction of the hystereses for the GRACE signals move in opposite directions from the isolated groundwater hystereses. The subsurface water (soil moisture and groundwater) hystereses more closely resemble the storage-runoff relationship of a soil matrix. While the physical processes underlying these hystereses are inherently complex, the vertical integration of terrestrial water in the GRACE signal encapsulates the processes that govern the non-linear function of regional-scale watersheds. We use this process-based understanding to test how GRACE data can be applied prognostically to predict seasonal runoff (mean Nash-Sutcliffe Efficiency of 0.91) and monthly runoff during the low flow/high demand month of August (mean Nash-Sutcliffe Efficiency of 0.77) in all three watersheds. The global nature of GRACE data allows this same methodology to be applied in other regional-scale studies, and could be particularly useful in regions with minimal data and in trans-boundary watersheds.

1. Introduction

At the most fundamental level, watershed processes can be described as the collection, storage, and release of water (Black, 1996; McDonnell et al., 2007). The runoff from these processes is governed by threshold mediated relationships across scales that result in storage–runoff hystereses (Spence, 2010). These threshold relationships between storage and runoff (*S*–*R*) are not uniform across a watershed, functioning as a series of discontinuous processes in soils and hillslopes that provide an integrated *S*–*R* relationship at the watershed scale (Spence, 2010). Kirchner (2009) described the *S*–*R* relationship to be non-linear and stated that watersheds typically function as dynamic systems governed by their unique climate and geology. These conceptual models of hydrologic behaviors help provide a process-based understanding of watersheds as dynamic environmental systems (Aspinall, 2010), and identify connections that advance hydrologic science and hydrologic prediction (Wagener et al., 2007).

At the local scale, *in situ* instrumentation can quantify the non-linear relationship between streamflow and water stored in a watershed as snow, soil moisture, groundwater and reservoirs (Appleby, 1970; Brutsaert, 2008; Kirchner, 2009; Sayama et al., 2011). These four primary storage components, along with climate, topography, and geology govern the fluxes of water through a catchment, and play an important role in the hysteretic nature of storage and runoff dynamics (McGlynn and McDonnell, 2003; McNamara et al., 2011). Knowledge of these processes is fundamental to developing an understanding of a watershed's hydrologic behavior. However, observations over larger regions can be technically challenging and costly, and *in situ* measurements from small basins do not necessarily represent the complexity inherent to watersheds at more broad scales (Spence, 2010). This scaling problem limits our understanding and predict regional hydrologic processes, which is often the practical scale of watershed management (Blöschl, 2001; Western et al., 2002; Skøien et al., 2003; Peel and Blöschl, 2011; Thompson et al., 2011).

In the absence of broad-scale observations, past hydrological studies have typically relied on *in situ* measurements as a proxy for regional scale hydrological processes. For example, in higher latitude or mountainous regions measurements of snow water storage have provided a simple metric that has been used in water resource planning for decades (Cayan, 1996; United States Army Corps of Engineers, 2001), and are often correlated to streamflow gauged downstream (Dozier, 2011). While informative, this approach can often provide hydrological forecasts that are misleading, because point-based measurements do not fully represent the broad-scale variability of rugged mountain terrain (Dozier, 2011; Nolin, 2012; Webster et al., 2014; Ayala et al., 2014). Similarly, measurements of soil moisture in the upper 2000 mm of the soil rely on point-based data that are often distributed at the regional scale, but do not effectively represent the true variability of soil moisture found at the regional scale (Western et al., 2002; Brocca et al., 2010). A complete understanding of groundwater stores and fluxes (deeper than 2000 mm) at regional scales also remains elusive, despite its increasing importance in water resources management (Wagener et al., 2007; Gleeson et al., 2012; Famiglietti and Rodell, 2013; Barthel, 2014). In addition to contributing to runoff, groundwater serves as an important water resource for consumptive use (Gleeson et al., 2012).

While local-scale methods have been applied with moderate success in the past, current trends in climate and in consumptive water demand suggest that long-term changes in hydrological fluxes will have a major impact at the regional scale (Milly et al., 2008). As a result, the supply and demand of water is also expected to shift, especially at the regional scale (Wagener et al., 2010; Gleick, 2014a).

Hydrologic models can help address the questions of scale and bridge the gap between local scale observations and regional-scale processes by estimating the primary components of water storage (snow, soil moisture, reservoir, and groundwater) across a larger spatial grid. Regional-scale modeling approaches are integrated into water resource management operations for navigation, human consumptive use, irrigation, and hydropower (Payne et al., 2004; Rodell et al., 2004). Models can also be applied diagnostically to test scientific hypotheses and provide a better understanding of the physical processes that govern real world systems, such as the connections between snowmelt, streamflow, and groundwater (Beven, 2007, 2010; Moradkhani and Sorooshian, 2008; Kirchner, 2009; Clark et al., 2011; Capell et al., 2012). Despite their utility, developing and validating a model can be both time consuming and reliant on multiple data inputs, which even in the most well-instrumented basins provides sparse geographic coverage (Bales et al., 2006; Zang et al., 2012). The lack of an integrated measurement of water storage and streamflow has limited regional-scale hydrologic insights to model-based studies (Koster et al., 2010; Mahanama et al., 2011).

Since 2002, broad-scale measurements of changes in the amount of water stored across and through the earth have been available from NASA's Gravity Recovery and Climate Experiment (GRACE) satellites (Tapley et al., 2004). GRACE measures monthly changes in the Earth's gravitational field that are proportional to regional changes in total water storage (Wahr et al., 2006). GRACE satellites provide a monthly record of terrestrial water storage anomalies (*TWSA*), which represent the changes in the vertical sum of water at the Earth's surface stored in snow, surface, soil and groundwater. Water losses to runoff and evapotranspiration are implicit in the GRACE storage signal, which greatly simplifies calculations of changes in terrestrial water storage.

GRACE data, coupled with modeled and measured variations of water stored in snow, surface reservoirs and soils, have successfully been decomposed to quantify regional groundwater changes (Rodell et al., 2009; Famiglietti et al., 2011; Voss et al., 2013; Castle et al., 2014) and have contributed to improving water balance calculations (Zaitchik et al., 2008; Li et al., 2012). More recent efforts have quantified the relationship between regional water storage and specific streamflow events (Reager and Famiglietti, 2009; Reager et al., 2014). Riegger and Tourian (2014) coupled GRACE data using data-driven and model-based approaches to better understand the relationship between storage and runoff across climatic zones globally. Their study found that coupled liquid storage is linear to runoff, and that in climatic regions with snow and ice the relationship between storage and runoff is more hysteretic. These novel analyses, which are more diagnostic in nature, have provided new insights into regional watershed hydrology using GRACE measurements as a core data input. These studies have not explored how topography and geology can also help describe the *S*–*R* relationship of regional watersheds. Nor did these studies examine the ability of GRACE measurements to predict seasonal runoff.

In this paper, we use terrestrial water storage data from GRACE to better understand the hydrology of regional watersheds and the relationship between storage and runoff. The temporal relationships between coincident *TWSA* and discharge observations at three scales in the Columbia River Basin (CRB) of western North America are investigated using climate, topography, and geology as a framing principle to describe the shape of the storage-streamflow hysteresis. We associate regional and temporal differences in the hystereses with varying watershed dynamics. Finally, we compare the prognostic abilities of GRACE observations with individual modeled estimates of snow and soil moisture to predict seasonal streamflow at regional scales.

2. Study Area

Our study area is the Columbia River Basin (CRB; 41-53ºN and 110-122ºW; Fig. 1). This basin has dry summers and wet winters. Up to 70% of annual precipitation falls between November and March, 50–60% of which occurs as snow (Serreze et al., 1999; Nolin et al., 2012). The spring months (April to June) are also wet, but warmer. Precipitation during the spring combines with snowmelt to swell rivers and potentially exacerbate flooding. Snowmelt also serves as a critical component of the hydrologic cycle, recharging aquifers and filling streams later in the year. These contributions bridge the temporal disconnect between wet winters and dry summers when demand is at its peak as farmers, fish, hydropower and municipal users vie for over-allocated water resources (United States Army Corps of Engineers, 2001; Oregon Water Supply and Conservation Initiative, 2008). However, concerns with winter surplus and summer scarcity are not uniform across the CRB, since climate and geology vary greatly (Nolin, 2012). Two of the study watersheds, the Upper Columbia (155,000 km^2) and the Snake River basin (182,000 km^2), represent distinctly different climatic, topographic, and geologic provinces of the CRB (described and illustrated in Fig. 1). The Upper Columbia is wet and is characterized by steep topography of fractured rock and poor groundwater storage. In contrast, the arid Snake River basin is bowl-shaped with mountains on three sides. The interior of the Snake River basin is a broad plain with well-developed soils underlain by a highly transmissive aquifer (Whitehead, 1992, Fig. 1). The Columbia River at The Dalles (614,000 km^2) encompasses the Upper Columbia and the Snake River sub-basins, and its climate and geology are an integration of the two (Fig. 1). A distinct climatic feature of the Columbia River at The Dalles is the western slope of the Cascade Mountains, where over 3000 mm of mean annual precipitation at higher elevations sustains a considerable seasonal snowpack. The scale of this study was constrained to watersheds larger than 150,000 km^2, the optimal minimum geographic limit of GRACE data (Yeh et al., 2006; Landerer and Swenson, 2012).

3. Methods and Data

We used 108 months of GRACE and streamflow data over nine water years (WY; Oct – Sep; 2004–2012). This data comprises positive, neutral, and negative phases of the El Niño-Southern Oscillation and negative and positive phases of the Pacific Decadal Oscillation (Feng et al., 2014; Iizumi et al., 2014). As a result, the data provides years of above- and below-average precipitation, snowpack, and streamflow for the region. The three watersheds were delineated upstream from United States Geological Survey (USGS) stream gages at 1° resolution, which is the resolution of GRACE data. In the CRB, these grid cells represent a dimension of approximately 80 km by 120 km. The Upper Columbia consists of the area upstream of the Columbia River at the International Boundary gage (USGS 12399500), just downstream of the confluence of the Columbia and Pend-Oreille Rivers. The Pend-Oreille is a major watershed in the upper portions of the CRB. The Snake River gage at Weiser (USGS 13269000) provides gauged streamflow data above Hell's Canyon Reservoir, the largest impoundment in the Snake River basin. The USGS gage at The Dalles (USGS 14105700) provides the most downstream streamflow data for the CRB.

Monthly mean runoff (*R*; mm) was calculated for each of the three gages using the USGS streamflow data.

Measurements of *TWSA* were obtained from the GRACE RL-05 (Swenson and Wahr, 2006; Landerer and Swenson, 2012) data set from NASA's Tellus website (http://grace.jpl.nasa.gov). The errors present in the gridded GRACE data exist primarily as a result of truncation (i.e., a low number of harmonics) in the spherical harmonic solution, and smoothing and systematic noise removal (called "de-striping") that is applied after GRACE level-2 processing to remove spatially correlated noise (called "stripes") (Swenson and Wahr, 2006). This smoothing tends to smear adjacent signals together (within the radius of the filtering function), resulting in smaller signals being lost, and larger signals having a coarser footprint and a loss of spatial information.

To restore the GRACE signal lost during processing, the data were scaled using 1° Land-Grid Scale Factors produced by putting a 1° land surface model through identical processing (truncation and filtering) as the GRACE solutions, then measuring the decrease in the signal amplitude at each 1° grid. These procedures are described on the Tellus website and detailed in Landerer and Swenson (2012). Monthly 1° GRACE estimates of *TWSA*, and the associated 1° leakage and measurement errors, were spatially averaged over each of the three study watersheds following the procedures described in the Tellus website.

GRACE represents monthly storage anomalies relative to an arbitrary record-length mean value, analogous to the amount of water above or below the long-term mean storage of a bucket, and should balance with the equation:

$$\Delta Storage = TWSA = \Delta GW + \Delta SM + \Delta SWE + \Delta RES \qquad (1)$$

where all components are at monthly time steps; *GW* represents groundwater, *SM* represents soil moisture (from 0–2000 mm depth), *SWE* represents snow water equivalent (the equivalent depth of water held in snowpack), and *RES* represents reservoir storage. The Δ used here represents the anomaly from the study-period mean, rather than a monthly change. To isolate monthly groundwater storage anomalies ($\Delta GW = GWSA$) in the above equation, ΔSM, ΔSWE and ΔRES estimates were subtracted from the monthly *TWSA* data using methods described in Famiglietti et al. (2011). Similarly, the combined signal of water storage anomalies of subsurface moisture ($TWSA_{\mathrm{sub}}$), *SM* and GW, was isolated by subtracting *SWE* and *RES* from *TWSA* values.

Monthly *SM* values over the study basins were obtained from the mean of the North American and Global Land Data Assimilation Systems (NLDAS at 1/8º resolution (Cosgrove et al., 2003) and GLDAS at 1/4º resolution (Rodell et al., 2004), respectively), and were spatially averaged over the three study watersheds. Monthly 1-km resolution *SWE* values were obtained from the mean of NLDAS and Snow Data Assimilation System (SNODAS; National Operational Hydrologic Remote Sensing Center, 2004) and were spatially averaged over the three watersheds. SNODAS data were used in place of the GLDAS data product, which considerably underestimated *SWE* in mountainous areas when compared to point-based measurements. Changes in monthly reservoir storage were calculated for the five largest reservoirs in the CRB (see Table A1). Other smaller reservoirs in the CRB were excluded when it was determined that fluctuations in their levels were below the detection limits of GRACE.

Like all measurements, estimates of *TWSA* from GRACE contain error. For all of the study basins, the range of error is well below the *TWSA* signal strength, approximately an order of magnitude below the annual amplitude (200 – 300 mm) of the *TWSA* signal in the CRB. The basin-averaged *TWSA* errors (time invariant) for the three study basins are 37 mm (Upper Columbia), 22 mm (Snake), and 25 mm (The Dalles).

The model data from LDAS and SNODAS simulations are driven by *in situ* measurements, and represents the best available data for broad scales. We address any structural error from an individual model by using an ensemble of outputs. Calculation of the error in individual terms followed standard methodologies (Famiglietti et al., 2011), where error in *SM* is the mean monthly standard deviation, and standard errors for *SWE* and *RES* are 15% of mean absolute changes. *GWSA* and $TWSA_{sub}$ anomaly errors are calculated as the sum of basin-averaged errors (added as variance) in the individual terms in each respective calculation (eq. 1), including the error in *TWSA* (Swenson et al., 2006). The basin-averaged error variance for *GWSA* (time invariant) in the three study basins are 45 mm (Upper Columbia), 26 mm (Snake), and 33 mm (The Dalles). For $TWSA_{sub}$ these values are 37 mm (Upper Columbia), 22 mm (Snake), and 25 mm (The Dalles). The individual error components (*SM*, *SWE*, *RES* respectively) for each basin are Upper Columbia (24 mm, 6 mm, 0.01 mm), Snake (14 mm, 3 mm, 0.01 mm), and The Dalles (21 mm, 4 mm, 0.01mm). Note that these error estimates are distributed across an entire regional watershed and do not represent the error at individual monitoring sites. A time series of these values and basin-averaged errors is provided in Fig. 2.

Based on an approach similar to Reager et al. (2014) and Riegger and Tourian (2014), we plotted the temporal relationship between *TWSA* and *R* to examine hysteresis relationships in all three of the study watersheds for each individual water year and for the monthly mean across all water years. Expanding from the integrated terrestrial component of water storage, we also plotted the relationships of $TWSA_{sub}$ and *GWSA* with *R*. We examined the branches of these hysteresis plots to better understand how the size, shape, and direction of the hystereses varied across years in each of the three regional watersheds.

In order to verify groundwater hysteresis, we compared the GRACE-derived *GWSA* to groundwater depths from well measurements at 33 sites throughout the study region (Fig. 1 and Table A2). These data were normalized by their standard deviation, and the mean of the 33 wells was calculated. The standard deviation of the GRACE-derived *GWSA* for The Dalles was normalized to provide a direct comparison of *GWSA* and *in situ* measurements.

We further hypothesized that because peak *SWE* accumulation occurs between February and April, that *TWSA* for these months could be used to predict *R* for an individual month and the cumulative seasonal runoff (R_{season}) that occurs after peak *SWE* accumulation. To test this prognostic hypothesis we used a two-parameter power function (The MathWorks, 2013):

$$R_{predicted} = a(TWSA_{month})^b + c \qquad (2)$$

where $R_{predicted}$ is runoff for the predicted time interval; $TWSA_{month}$ represents terrestrial water storage for an individual month, and a, b, and c are fittted parameters from the power function.

We tested this relationship for *TWSA* in February, March and April to predict R_{season} (April – September) and for the individual months of July (R_{July}), August (R_{Aug}), and September (R_{Sep}); these represent the lower-flow months when demand is near its peak.

Additionally, we tested and compared the modeled-values of *SWE* from NLDAS and SNODAS and *SM* from NLDAS and GLDAS, and the model-derived values of $TWSA_{sub}$ to predict R_{season} and for the individual months using the same power-function analysis.

Because our data set was constrained to nine water years, we used a double-pass approach to fit and test the empirical relationship between *S*–*R*. This approach allowed us double our data inputs for calculating standard hydrologic evaluation metrics such as Root Mean Square Error (RMSE), standard Nash-Sutcliffe Efficiency (NSE) and Coefficient of Determination (R^2); (Legates and McCabe, 1999; Nash and Sutcliffe, 1970). The nine years were divided into two sets (Set 1, even years 2004–2012; Set 2, odd years 2005–2011). The first pass calculated the power function of *S*–*R* to Set 1, and the parameters were then tested against Set 2. The roles of the datasets were then reversed and the datasets were again tested against each other. The empirical model results from Sets 1 and 2 were then combined into a single set of solutions for the model fit and tested against measured values to calculate RMSE, NSE, and R^2. In order to maximize the limited data inputs, once we tested the two independent sets for model performance, we combined the input data values into a single set for all nine years to calculate a single power function curve. The observed data were tested against the simulated data from the complete, but limited data record. The final model curve was fit to these data, and the evaluation metrics include all of the years for each respective dataset.

4. Results

4.1. Storage-runoff hysteresis

The filling and emptying of the study basins at the regional-scale over the course of an individual WY results in a hysteretic relationship between storage and runoff (Fig. 3a). The hysteresis loops begin at the onset of the wet season in October, with *TWSA* increasing (Figs. 3a, 4a-c) as precipitation is stored as snow and soil moisture. An increase in storage that is not offset by an increase in discharge indicates a predominance of snow inputs and the freezing of soil water. The lower branch of the hysteresis plot (storage increase unmatched by runoff) can be used to estimate cumulative snow water equivalent and soil moisture in the basin. This is the water that later contributes to streamflow and groundwater recharge in the spring.

The hysteresis shifts direction from Feb-Apr (inflection 1, Fig. 3a), and represents each watershed's transition from storage to release. This response is evident (Figs. 3a, 4a-c, and 5), as each hysteresis loop contains a vertical branch of the curve during which storage is relatively constant, but streamflow increases rapidly. This timing also represents the groundwater recharge branch of the loop. As snow melts and the ground thaws, runoff is generated, recharge into soils occurs, and basins tend to be at peak storage during this branch. Storage losses and additional precipitation inputs during this period are re-organized internally. A second shift (inflection 2, Fig. 3a) occurs from Apr-June when peak *TWSA* begins to decrease, representing spring snowmelt and a switch from precipitation that falls primarily as snow to rain; these combine to contribute to peak *R*.

Once peak *R* values are reached, the loop shifts direction a third time (inflection 3, Fig. 3a), receding on both axes as contributions from snowmelt diminish while presumably groundwater sustains streams and provides a source for irrigated agriculture. During this period, the relationship between *TWSA* and discharge is more linear, corresponding to baseflow-driven runoff processes in which each monthly change in storage causes a proportional change in the generation of streamflow.

The hysteresis plots of *TWSA–R* for an individual water year demonstrate that the timing and quantity of precipitation governs the size of a hysteresis loop for an individual WY (Figs. 3a, 4a-c, 5). For instance wet years (e.g., 2008) have bigger loops, while dry years (e.g., 2005) are more compressed along both axes. However, the general shape of the loops is distinct for each basin. Plotting multiple WYs provides a family of curves for each basin that helps describe how climate, topography, and geology govern the timing and magnitude of the relationship between *TWSA* and *R* (Figs. 3a, 5).

4.2. Subsurface water ($TWSA_{sub}$) – runoff hysteresis

The $TWSA_{sub}$ hysteresis curve contracts horizontally when the snow signal is removed from *TWSA* values for both the Upper Columbia and The Dalles (Figs. 3b, 4d-f), which collapses the loops and takes a form similar to a plot-scale hysteresis of soil. In the initial stages of the WY, the direction of the curve changes directions 2-3 times along the *TWSA* axis. Similar hysteresis fluctuations have also been documented at the more local scale as the soil profile moves towards saturation (Penna et al., 2011). Peak $TWSA_{sub}$ occurs in June, which corresponds to the spring melt of mountain snowpack and the end of the wet season (Figs. 4d-f). In contrast to the near linear relationship between $TWSA_{sub}$ and *R* in the Upper Columbia and The Dalles, the Snake River retains a more complex relationship. In this watershed the hysteresis curve still retains a loop, but the timing of maximum $TWSA_{sub}$ is also earlier, reaching its peak during March and April (Fig. 4e). It is also noteworthy that in the Snake River the $TWSA_{sub}$–*R* hysteresis loop temporally progresses in the opposite direction, but stays in phase with precipitation inputs.

4.3. Groundwater-runoff hysteresis

The hysteresis loops describing the temporal relationship between *GWSA* and *R* are equally informative, with one distinct difference—they temporally progress in opposite directions of the hysteresis loops of *TWSA* and *R* (Fig. 3). For all three watersheds, *GWSA* decreases from Oct–Feb/Mar (Fig. 4h-j), which is out of phase with the onset of the wet season. *GWSA* does not shift towards positive gains until early spring and the initial stages of melt before reaching its maximum in June. From June-Sept, *GWSA* decreases minimally across all years during the runoff recession limb, indicating groundwater contributions to streamflow. This decreasing *GWSA* signal does not stand out in Fig 2, as during WY 2011 the *GSWA* increased from 12.8 mm in June to 71.2 mm in September due a large snowpack that melted several months later than normal. This considerable anomaly muted the overall *GWSA* recession from June-Sept that is found in all other water years (supplementary data and the interactive visualizations described subsequently).

The 33 point-specific well data located across the CRB show considerable individual variability throughout a water year, and the mean of the normalized standard deviations of well levels was close to zero for all months. The temporal variability for the well data provides no discernable correlation with the derived *GWSA* signal (Fig. A1).

4.4. Individual basin hysteresis plots of *TWSA*, $TWSA_{sub}$, *GWSA* and *R*

Of the three study basins, the Upper Columbia is the most hydrologically active, showing the largest annual range for *TWSA*, $TWSA_{sub}$, *GWSA*, and *R* (Fig. 6). The *TWSA*–*R* hysteresis loops are more open (Fig. 4), corresponding to the fluxes of water moving through watershed. When *SWE* is removed and subsurface water is highlighted, the $TWSA_{sub}$–*R* hysteresis loops collapse horizontally and more closely resemble the hystereses associated with soil (Figs. 4d). However the inter-annual range ($WY_{max} – WY_{min}$) for $TWSA_{sub}$ in the Upper Columbia is considerably greater than the other two basins (median range = 234 mm; Fig. 6). As the hysteresis reverses directions for *GWSA*–*R*, the loops shift to a more open shape (Figs. 4d), but the inter-annual range remains similar.

In contrast to the rapid response of the Upper Columbia, the Snake River receives ~60% less annual precipitation, but has an annual *TWSA* range that is only 22% less (median annual range = 192 mm; *R*=7 mm; Figs. 4, 5, and 6). However, the *TWSA* hysteresis loops for the Snake River are collapsed vertically (Fig. 4b). In the more arid Snake River, removing the snow signal does not collapse the $TWSA_{sub}$–*R* hysteresis loops ($TWSA_{sub}$ = 89 mm). Similarly, the *GWSA* loops suggest that subsurface moisture plays a more prominent role in the Snake River.

The climate, topography, and geology of the Columbia River at The Dalles are an integration of the Upper Columbia and Snake River, seen in the shape of the hysteresis loops (Figs. 4, 5, 6; median annual range *TWSA*=195 mm; *R*=27 mm). The period from Feb–June more closely resembles the Snake River basin, with gradual increases in *TWSA* and sharp increases in *R*. The slope of the recession from June-Sept has the same general shape for The Dalles as the Upper Columbia (Figs. 4a, 4c), presumably from snowmelt-generated runoff.

Interactive visualizations that compliment all of the hysteresis figures are online at: https://public.tableau.com/profile/sprolese - !/vizhome/GRACE_hystereses/WSADash.

4.5. Streamflow forecasting

We next present how *TWSA* was applied prognostically to predict streamflow. Using the double-pass calibration and validation approach, $TWSA_{\mathrm{Mar}}$ provided the best overall predictive capabilities for R_{season} with a mean NSE ($\overline{\mathrm{NSE}}$) and mean R^2 ($\overline{\mathrm{R}}^2$) of 0.75 and 0.91, respectfully (Fig. 7a, Table 1), for all three basins. The Dalles had the highest NSE and R^2, and lowest RMSE values (0.98, 0.98, 6 mm; Table 1). The results in the Upper Columbia were also robust (0.82, 0.86, 33 mm; Table 1), while the Snake River performed with less skill (0.46, 0.59, and 14 mm, Table 1). Applying $TWSA_{\mathrm{April}}$ also provided similar results, but with a lower degree of skill in predicting *R* ($\overline{\mathrm{NSE}}$ = 0.57, $\overline{\mathrm{R}}^2$= 0.69). $TWSA_{\mathrm{Apr}}$ provided improved predicted capabilities in the Upper Columbia (0.87, 0.88, and 28 mm, Table 1), but inferior results in the other two watersheds. $TWSA_{\mathrm{Feb}}$ had a low degree of skill in predicting *R* in all three watersheds (Table A3).

$TWSA_{\mathrm{Mar}}$ and $TWSA_{\mathrm{April}}$ also served as a good predictor of monthly runoff in July and August for the Upper Columbia and to a lesser degree in The Dalles (Tables 1 and A3). In the Snake River, *TWSA* did not serve as a good predictor for *R* in an individual month.

Snowpack and soil moisture play a considerable role in the hydrology of the CRB and are commonly used to help predict water demand and availability later in the year (Koster et al., 2010). We compared the capabilities of the modeled snow (*SWE*) and soil moisture (*SM*) products to predict *R* to the skill of measured GRACE *TWSA* data (Table 1). In the Upper Columbia and The Dalles, $TWSA_{\mathrm{Mar}}$ predicts seasonal and monthly runoff

(July and August) with considerably more skill than *SWE* or *SM* (Figure 7, Table 1). In the Snake River, SM_{Mar} has a higher degree of skill than $TWSA_{\text{Mar}}$ in predicting R_{season} and R_{Aug}. SWE_{Mar} provided inferior results in all three watersheds, but with some predictive skill in the Upper Columbia and The Dalles (NSE of 0.24 and 0.46 respectively, Table 1). In all three watersheds, $TWSA_{\text{sub}}$ provided extremely poor predictions (Tables 1 and A3).

When the results of the empirical model using two independent sets of data proved robust for some of the storage metrics, the observed data were tested against the simulated data from the complete, but limited data record. The performance of the empirical model improved using the complete data set (Tables 2 and A4), with the same general results. $TWSA_{\text{Mar}}$ provided the best model fit for seasonal runoff in the Upper Columbia (NSE = 0.93, RMSE = 19.8 mm) and The Dalles (NSE = 0.98, RMSE = 5.7 mm). In the Snake River, predictive capabilities improved more dramatically (NSE = 0.83, RMSE = 7.4 mm), but soil moisture still served as a better predictor of seasonal streamflow (NSE = 0.93, RMSE = 5.2 mm). Similarly, $TWSA_{\text{Mar}}$ provided the best model fit for runoff in August, one of the drier months when demand is at its peak (Tables 2 and A4).

5. Discussion

5.1. Storage-runoff hysteresis

Decades of data collection and monitoring at individual gage sites indicate that watersheds collect, store and release water. Using one integrated measurement from the GRACE satellites, our results show these same processes at the regional scale in the hysteresis loops of storage (*TWSA*) and runoff (*R*). While hystereic processes have previously been identified in local-scale measurements (McDonnell, 2003; McGlynn and

McDonnell, 2003), only recently has streamflow-storage hysteresis been identified at the regional scale (Riegger and Tourian, 2014).

Our work builds on Riegger and Tourian's (2014) results, and employs GRACE data to describe how regional watersheds function as integrated, non-linear systems governed by climate, topography, and geology. Climate controls the size of the hysteresis loops by providing a first-order control on hydrologic inputs and the storage of solid water, which in turn governs the ranges of *TWSA* and *R*. However, runoff response to precipitation and snowmelt does not act independently from topography and geology (Jefferson et al., 2008; Tague et al., 2008), which controls how liquid water is stored and routed through a watershed, even at the regional scale. The climatic, topographic, and geological characteristics of each watershed provide an explanation of the *S*–*R* relationship that govern the shape and size of its respective hysteresis curve. GRACE offers a single, integrated measurement of changes in water storage through and across a watershed that can be applied to predict regional streamflow using an empirical model. Where these predictive capabilities succeed and fail help better describe the climatic, topographic, and geological characteristics in each watershed.

For example, in the Upper Columbia, steep topography and wet climate fills subsurface storage quickly before reaching a threshold in April or May. This transition represents the watershed's transition from winter storage to spring runoff. After this watershed-scale threshold is reached, the steep topography moves snowmelt and rain quickly through the terrestrial system and into the river channel until cresting in June (Figs. 4, 5, and 6), followed by declines in *TWSA* and *R* from June-September. These large fluxes of water create a more open hysteresis loop, expanding non-linearly on both the horizontal and vertical axes.

The Upper Columbia also has the broadest range of annual $TWSA_{sub}$ and *GWSA* during the study period (Figs. 5 and 6), despite having limited aquifer capacity. Conceptually, this demonstrates that the upper limit of storage is greater than in the Snake River or The Dalles, but that it also loses the most water. Its minimums at the end of the WY are also the lowest (median $TWSA_{Sep} = -98$mm; Figs. 5 and 6). This range across *TWSA*, $TWSA_{sub}$, and *GWSA* supports the conceptual model that the watershed fills during the wet season, and is then drained more quickly due to steep topography and limited water storage. The predictive capability of *TWSA* also strongly suggests that the components and temporal relationships of storage across this watershed are interconnected, and that incorporating April snowpack improves the model results.

In contrast, the arid Snake River basin provides a very different family of hysteresis curves (Figs. 4, 5) that identify groundwater and soil moisture as primary components of watershed function. The curves are compressed vertically (R) as compared to the Upper Columbia, and are more constrained horizontally (Fig. 6). The onset of spring melt runoff in February does not deplete *TWSA* for the Snake River. Instead, *TWSA* continues to increase until May, when peak runoff occurs. As *TWSA* decreases to the end of the water year in September, the median $TWSA_{Sep}$ measurement (−78 mm) is 20 mm greater than in the Upper Columbia. This indicates that the lower drainage threshold of the Snake River watershed is relatively greater than the Upper Columbia, potentially explained by a less severe topography and higher aquifer capacity.

The $TWSA_{sub}$ hysteresis curves in the Snake River retain a similar shape to the *TWSA* signal. While they reverse direction they do stay temporally connected to the onset of the wet season in October, indicating that subsurface moisture is a central control on the filling of the watershed through May. The capabilities of *SM* to empirically predict *R* better than *TWSA* further highlight the importance of subsurface water in this watershed. The intra-annual range of *GWSA* in the Snake River is also more limited than in the more hydrologically responsive Upper Columbia. This more limited range of data supports the conceptual model of a watershed that retains comparatively more winter precipitation in soils and aquifers throughout the spring season, and that sustains flow later in the year and until the onset of melt the following winter.

The greater Columbia River Basin upstream from The Dalles integrates the climatic, topographic, and geologic characteristics of the Snake River and Upper Columbia as well as other areas within the CRB. The western slope of the Cascades (Fig. 1), which is outside of the Upper Columbia, accumulates up to several meters of *SWE* each winter. Due east of the Cascades, a broad basalt plain that provides aquifer storage helps dampen the snowmelt pulse in the spring. The hysteresis loops for The Dalles reflect these combined characteristics.

Storage at The Dalles increases along the horizontal axis (*TWSA*) until peak storage is reached in March or April (Figs 3, 4, and 5). This *TWSA* threshold of approximately 100 mm responds with an increase in *R* that continues through June. In July, the hysteresis begins to recede along both axes closing out the loop. The *GWSA* has the most limited range, potentially explained by the extensive basalt aquifer moderating the relationship between storage and runoff. In The Dalles, $TWSA_{Sep}$ has a median value of −88mm (Fig. 6), between the lower drainage thresholds of the Upper Columbia and Snake River watersheds; indicating an integration of the contributing climate, topography, and geology.

5.2. Distinguishing the difference between $TWSA_{sub}$ and *GWSA*

Conceptually $TWSA_{sub}$ represents changes in the amount of water stored as soil moisture and groundwater, where as *GWSA* represents water changes greater than 2000mm below the soil surface. The goals of evaluating these metrics were to see if monthly changes in soil moisture were linked to changes in groundwater storage, and the role of snowpack in the *S*–*R* relationship.

The $TWSA_{sub}$ hysteresis curves in the Upper Columbia and The Dalles collapse into a more linear relationship that is more commonly associated with the *S*–*R* relationship of a soil matrix (Fig. 3 and 4). These stand in contrast to the Snake River where the $TWSA_{sub}$–*R* hystereses retain a loop, indicating a more complex relationship between storage and runoff. The hydrological processes that create these differences warrant investigation, but lie outside the scope of this study.

Although the annual fluctuations of *SM* are similar in all three basins (Fig. 2), its impact in the Upper Columbia is more pronounced. This watershed has poor groundwater storage, and relies on soils to provide seasonal storage of rain and snowmelt. In the Upper Columbia once the *SM* signal is removed, the intra-annual range of *GWSA* shifts to considerably lower values (Fig. 6). Shifts of similar magnitudes in *GWSA* are not found in the Snake River and The Dalles. These watersheds have excellent groundwater storage and are less reliant on soils to provide seasonal storage of rain and snowmelt. Fluctuations in reservoir levels are minimal with regards to the water fluxes in this region, and have minimal impact on calculations of *GWSA*.

The *GWSA–R* hystereses are represented by loops that show an out-of-phase relationship between precipitation and groundwater recharge from the start of the wet season in October until February or March. The $TWSA_{sub}$ and *GWSA* hysteresis plots demonstrate that in the Upper Columbia and The Dalles changes in monthly soil moisture are not always temporally aligned with *GWSA*. This can be explained by the physical reality that soil moisture and groundwater are not always interconnected, and that there is not a fixed depth (i.e., 2000 mm) that separates the two components of water storage.

GRACE-derived calculations of *GWSA* also provide insights into the hydrological processes governing groundwater recharge and depletion, as evidenced in the *GWSA* hysteresis loops. The *GWSA–R* curves show an out-of-phase relationship between precipitation and groundwater recharge from the start of the wet season in October until February or March.

This response in all of the *GWSA–R* hystereses suggests that even at the watershed scale groundwater recharge requires soils and geologic materials to fill to a certain moisture threshold and for the onset of snowmelt (Figs. 3a, 4a-c, 5, and web-based interactive visualizations). This also suggests that snowmelt inputs to groundwater are considerable. In the CRB this is critical as current climate trends are projected to reduce snowpack accumulation and exacerbate melt in the region (Wu et al., 2012; Rupp et al., 2013; Sproles et al., 2013).

Additionally, our analysis identifies summer as the time of peak groundwater storage in all three regional watersheds. This finding is of value for groundwater management and policy decisions, as peak groundwater levels in June correspond to the timing of groundwater pump tests that are used to develop groundwater withdrawal regulations (Jarvis, 2011, 2014). Our data suggest that groundwater pump tests should not be limited to an individual month, and should also include periods of reduced storage particularly during the winter months. The inclusion of multiple pump tests throughout the year could be particularly relevant as the population and water demand is projected to increase in the region.

The point-specific well data are not conclusive and show considerable variability with no consistent pattern regarding the timing of recharge and peak groundwater levels. This is presumably a function of how site characteristics (i.e., usage, depth, location, elevation) are extremely variable across a region. Rather than excluding these results or selecting individual wells that match GRACE data, we discuss the results from all 33 wells to help demonstrate the high variability that exists from well to well, and that measurements of groundwater changes at a fixed location does not represent watershed-scale characteristics (Jarvis, 2011, 2014). The disconnect between sites also highlights the concept brought forward by Spence (2010), that storage is not uniform across a watershed, and functions as a series of discontinuous processes at the watershed scale.

5.3. Applying the *S*–*R* relationship as a predictive tool

We applied these climatic, topographic, and geologic insights to develop and test the hypothesis that spring *TWSA* could predict *R* later in the year, based on two observations: First, the shapes of the hysteresis curves for each basin are similar (Figs. 4a-c, 5), but vary by magnitude of annual *TWSA*. Second, peak *TWSA* occurs before the peak runoff. We show that the integrated GRACE signal is a good baseline measurement to empirically predict seasonal streamflow across a range of water years with regards to precipitation and streamflow. In essence, our data suggest that the water stored across and through the Columbia River Basin in March describes the water available for the remainder of the water year.

In the CRB and in the northwestern United States, peak snowpack occurs in March or April, and is commonly used as a metric for predicting spring runoff. Despite the importance of snowpack to the hydrologic cycle of the region, measurements of $TWSA_{\mathrm{Mar}}$ from GRACE provide a better prediction of R_{season}, R_{July}, and R_{Aug} than model-derived estimates of snowpack. GRACE $TWSA_{\mathrm{Mar}}$ also provided a better prediction for runoff than soil moisture, except for the Snake River watershed. There SM_{Mar} provided a better indicator of runoff for the rest of the year. $TWSA_{\mathrm{Feb}}$ provided inferior predictive capacity, as the annual maximum *TWSA* values have not been reached.

These results are promising with regards to using GRACE as a predictive tool for water resources in both wet and dry years. Our limited data record represents a wide-range of conditions with regards to climate and streamflow, which is captured in our empirical models and is shown in the box plots to the right of Figs. 7a - b. These same results also indicate that *R* is insensitive to $TWSA_{\mathrm{Mar}}$ values below 100 mm in the Upper Columbia and at The Dalles and below 85 mm in the Snake River watershed. This limitation is not the empirical model, which works well, but that the basins have a limited response (i.e., R_{season} in the Snake River only changes about 60 mm, from around 20 to 85 mm). Given these responses the model works, and provides a lower threshold that describes with some certainty the amount of runoff that will be available for operations for the remainder of the year.

We recognize that all three of these regional watersheds are managed through a series of dams and reservoirs that create an altered runoff signal. Water resources managers use point-specific and model-based estimates of water storage in the region to optimize their operations for the water year. Additionally, in the fertile plains of the Snake River and lower CRB, broad-scale agriculture relies on both ground- and surface water for irrigation. Water withdrawals would be implicit in the *TWSA* signal and reduce *R*. However, a more detailed analysis of withdrawals lies outside the scope of this study.

Regardless of the length of record or anthropogenic influence, climate, topography, and geology still provide the first-order controls on water storage that are found in the hysteresis loops. GRACE encapsulates these hydrologic processes through measurements of *TWSA*. The hysteresis loops expand and contract accordingly during wet and dry years, as the intra-annual relationship between *TWSA* and Q represents the fluxes of water into and out of the watershed. Despite intra-annual differences, a family of hysteresis curves can describe each of the sub-regional watersheds. The predicative capability using *TWSA*, the vertical sum of water, as compared to snowpack and soil moisture further highlights the integrated nature of water storage in regional hydrology. These predictive capabilities highlight the potential of GRACE to improve upon seasonal forecast predictions and regional hydrological models.

5.4. GRACE as an analysis tool for regional watersheds

Where previous approaches to modeling watershed behavior have focused on separate storage compartments, new approaches should include the magnitude and direction of hysteresis (Spence, 2010). This integrated approach would provide new ways forward to classify watersheds not only by runoff, but also on the first-order controls that govern the non-linear hydrological processes.

Even though GRACE is somewhat of a blunt instrument with regards to temporal (monthly) and spatial (1°) resolution, this emerging technology provides a new dimension to regional watershed analysis by providing an integrated measurement of water stored across and through the Earth. These measurements continue to prove their value in retrospective analysis of regional hydrology (Rodell et al., 2009; Castle et al., 2014). However, the hysteresis loops presented by Riegger and Tourian (2014) and further developed in this paper demonstrate the ability of GRACE data to help develop a process-based understanding of how regional watersheds function as simple, dynamic systems. As the temporal record of GRACE continues to increase, its value as both a diagnostic and predictive tool will continue to grow. In the mean time, these data have value in augmenting existing management strategies.

Perhaps one of the most important facets of GRACE data is that it does not distinguish political boundaries. It is not linked to a specific *in situ* monitoring agency with limited data access and has the capacity to bridge sparse and inconsistent on-the-ground hydrologic monitoring networks that exist in many regions of the world. Previous GRACE-based analysis has shown its value in highlighting negative trends in terrestrial water storage in trans-boundary watersheds (Voss et al., 2013; Castle et al., 2014), and resulting regional conflict exacerbated by water shortages (Gleick, 2014b). GRACE provides an objective measurement of a region's water resources that can provide valuable insights into potential shortages or surpluses of water resources, and simple empirical predictions of seasonal and monthly runoff that are easily deployable in places with limited data.

6. Conclusions

We have shown how GRACE-based measurements of *TWSA* distill the complexity of regional hydrology into a simple, dynamic system. *TWSA* and derived estimates of *GWSA* reveal hysteretic behavior for regional watersheds, which is more commonly associated with hydrologic measurements at local scales. While the magnitude of the hysteresis curves vary across years, they retain the same general shape that is unique to each watershed. We demonstrated the utility of these hysteresis curves by showing how the complete *TWSA* record during March and April can be used to empirically predict *R* for the remainder for the water year ($TWSA_{\text{Mar}}$, mean NSE = 0.91) and during the drier summer months ($TWSA_{\text{Mar}}$, mean NSE for July = 0.76, August = 0.72; Tables 1 and 2).

Because GRACE *TWSA* can augment prediction, managers could start to interpret each year's hysteresis curve for the upcoming spring and summer, providing greater clarity and validation for model-based forecasts presently used by water resource managers. Our results demonstrate a way forward, expanding GRACE from a diagnostic tool, into a conceptual model and predictive resource.

Although this study focused on the CRB, which has a rich data record, GRACE data are available at a global scale and could be readily applied in areas with a paucity of data to understand how watersheds function and to improve streamflow forecasting capabilities. GRACE does not discern political boundaries and provides an integrated approach to understanding international watersheds (Voss et al., 2013). This resource could serve as a valuable tool for managers in forecasting surplus and scarcity, and in developing strategies that include changes in supply and demand due to human consumptive needs and current climate trends (Wagener et al., 2010; Gleick, 2014a).

Author Contributions

E.A.S., S.G.L., and P.J.W. developed the hysteresis concept based upon background research by J.R. and J.S.F. The data analysis was led by E.A.S., but represents a combined effort from all of the authors. J.R. provided expertise in the GRACE data product, groundwater, and error analysis. E.A.S. prepared the manuscript with contributions from all co-authors.

Acknowledgments

The authors would like to thank Matthew Rodell and Felix Landerer for their expertise in understanding GRACE data during the initial stages of the research. GRACE terrestrial data were processed by Sean Swenson, supported by the NASA MEaSUREs Program, and are available at http://grace.jpl.nasa.gov. The GLDAS and NLDAS data used in this study were acquired as part of the mission of NASA's Earth Science Division and archived and distributed by the Goddard Earth Sciences Data and Information Services Center. We would like to thank the reviewers of the manuscript that have helped improve its overall quality. Additionally, Tim Kerr provided objective feedback and comments on the research findings. The information in this document has been funded entirely by the US Environmental Protection Agency, in part by an appointment to the Internship/Research Participation Program at the Office of Research and Development, U.S. Environmental Protection Agency, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and EPA. This manuscript has been subjected to Agency review and has been approved for publication.

Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

References

Appleby, F. V.: Recession and the baseflow problem, Water Resour. Res., 6(5), 1398–1403, doi:10.1029/WR006i005p01398, 1970.

Aspinall, R.: A Century of physical geography research in the Annals, Ann. Assoc. Am. Geogr., 100(5), 1049–1059, 2010.

Ayala, A., McPhee, J. and Vargas, X.: Altitudinal gradients, midwinter melt, and wind effects on snow accumulation in semiarid midlatitude Andes under La Niña conditions, Water Resour. Res., 2014.

Bales, R. C., Molotch, N. P., Painter, T. H., Dettinger, M. D., Rice, R. and Dozier, J.: Mountain hydrology of the western United States, Water Resour. Res., 42(8), W08432, doi:10.1029/2005wr004387, 2006.

Barthel, R.: A call for more fundamental science in regional hydrogeology, Hydrogeol. J., 22(3), 507–510, 2014.

Beven, K.: Towards integrated environmental models of everywhere: uncertainty, data and modelling as a learning process, Hydrol. Earth Syst. Sci., 11(1), 460–467, 2007.

Beven, K. J.: Preferential flows and travel time distributions: defining adequate hypothesis tests for hydrological process models, Hydrol. Process., 24(12), 1537–1547, 2010.

Black, P. E.: Watershed hydrology, CRC Press, Boca Raton, FL., 1996.

Blöschl, G.: Scaling in hydrology, Hydrol. Process., 15(4), 709–711, 2001.

Brocca, L., Melone, F., Moramarco, T. and Morbidelli, R.: Spatial-temporal variability of soil moisture and its estimation across scales, Water Resour. Res., 46(2), 2010.

Brutsaert, W.: Long-term groundwater storage trends estimated from streamflow records: Climatic perspective, Water Resour. Res., 44(2), W02409, doi:10.1029/2007WR006518, 2008.

Capell, R., Tetzlaff, D., Hartley, A. J. and Soulsby, C.: Linking metrics of hydrological function and transit times to landscape controls in a heterogeneous mesoscale catchment, Hydrol. Process., 26(3), 405–420, 2012.

Castle, S. L., Thomas, B. F., Reager, J. T., Rodell, M., Swenson, S. C. and Famiglietti, J. S.: Groundwater Depletion During Drought Threatens Future Water Security of the Colorado River Basin, Geophys. Res. Lett., doi:10.1002/2014GL061055, 2014.

Cayan, D. R.: Interannual Climate Variability and Snowpack in the Western United States, J. Clim., 9(5), 928–948, doi:10.1175/1520-0442(1996)009<0928:ICVASI>2.0.CO;2, 1996.

Clark, M. P., Kavetski, D. and Fenicia, F.: Pursuing the method of multiple working hypotheses for hydrological modeling, Water Resour. Res., 47(9), 2011.

Cosgrove, B. A., Lohmann, D., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Marshall, C., Sheffield, J. and Duan, Q.: Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project, J. Geophys. Res., 108(D22), 8842, 2003.

Dozier, J.: Mountain hydrology, snow color, and the fourth paradigm, Eos, Trans. Am. Geophys. Union, 92(43), 373, doi:10.1029/2011EO430001, 2011.

Famiglietti, J. S., Lo, M., Ho, S. L., Bethune, J., Anderson, K. J., Syed, T. H., Swenson, S. C., de Linage, C. R. and Rodell, M.: Satellites measure recent rates of groundwater depletion in California's Central Valley, Geophys. Res. Lett., 38(3), L03403, doi:10.1029/2010gl046442, 2011.

Famiglietti, J. S. and Rodell, M.: Water in the Balance, Science (80-.)., 340(6138), 1300–1301, 2013.

Feng, J., Wang, L. and Chen, W.: How does the East Asian summer monsoon behave in the decaying phase of El Niño during different PDO phases?, J. Clim., 27(7), 2682–2698, 2014.

Gleeson, T., Wada, Y., Bierkens, M. F. P. and van Beek, L. P. H.: Water balance of global aquifers revealed by groundwater footprint, Nature, 488(7410), 197–200, 2012.

Gleick, P.: Water, Drought, Climate Change, and Conflict in Syria, Weather. Clim. Soc., 6, 331–340, 2014a.

Gleick, P. H.: The World's Water Volume 8: The Biennial Report on Freshwater Resources, Island Press., 2014b.

Iizumi, T., Luo, J.-J., Challinor, A. J., Sakurai, G., Yokozawa, M., Sakuma, H., Brown, M. E. and Yamagata, T.: Impacts of El Niño Southern Oscillation on the global yields of major crops, Nat. Commun., 5, 2014.

Jarvis, W. T.: Unitization: a lesson in collective action from the oil industry for aquifer governance, Water Int., 36(5), 619–630, 2011.

Jarvis, W. T.: Contesting Hidden Waters: Conflict Resolution for Groundwater and Aquifers, Routledge., 2014.

Jefferson, A., Nolin, A., Lewis, S. and Tague, C.: Hydrogeologic controls on streamflow sensitivity to climate variation, Hydrol. Process., 22(22), 4371–4385, doi:10.1002/hyp.7041, 2008.

Kirchner, J. W.: Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward, Water Resour. Res., 45(2), W02429, doi:10.1029/2008WR006912, 2009.

Koster, R. D., Mahanama, S. P. P., Livneh, B., Lettenmaier, D. P. and Reichle, R. H.: Skill in streamflow forecasts derived from large-scale estimates of soil moisture and snow, Nat. Geosci, 3(9), 613–616, doi:http://www.nature.com/ngeo/journal/v3/n9/suppinfo/ngeo944_S1.html, 2010.

Landerer, F. W. and Swenson, S. C.: Accuracy of scaled GRACE terrestrial water storage estimates, Water Resour. Res., 48(4), W04531, doi:10.1029/2011wr011453, 2012.

Legates, D. R. and McCabe, G. J.: Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation, Water Resour. Res., 35(1), 233–241, 1999.

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

Mahanama, S., Livneh, B., Koster, R., Lettenmaier, D. and Reichle, R.: Soil Moisture, Snow, and Seasonal Streamflow Forecasts in the United States, J. Hydrometeorol., 13(1), 189–203, doi:10.1175/JHM-D-11-046.1, 2011.

McDonnell, J. J.: Where does water go when it rains? Moving beyond the variable source area concept of rainfall-runoff response, Hydrol. Process., 17(9), 1869–1875, doi:10.1002/hyp.5132, 2003.

McDonnell, J. J., Sivapalan, M., Vache, K., Dunn, S., Grant, G., Haggerty, R., Hinz, C., Hooper, R., Kirchner, J., Roderick, M. L., Selker, J. and Weiler, M.: Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology, Water Resour. Res., 43(7), doi:10.1029/2006wr005467, 2007.

McGlynn, B. L. and McDonnell, J. J.: Quantifying the relative contributions of riparian and hillslope zones to catchment runoff, Water Resour. Res., 39(11), doi:10.1029/2003WR002091, 2003.

McNamara, J. P., Tetzlaff, D., Bishop, K., Soulsby, C., Seyfried, M., Peters, N. E., Aulenbach, B. T. and Hooper, R.: Storage as a Metric of Catchment Comparison, Hydrol. Process., 25(21), 3364–3371, doi:10.1002/hyp.8113, 2011.

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P. and Stouffer, R. J.: Stationarity Is Dead: Whither Water Management?, Science (80-.)., 319(5863), 573–574, doi:10.1126/science.1151915, 2008.

Moradkhani, H. and Sorooshian, S.: General Review of Rainfall-Runoff Modeling: Model Calibration, Data Assimilation, and Uncertainty Analysis, in Hydrological Modelling and the Water Cycle, pp. 1–24., 2008.

Nash, J. E. and Sutcliffe, J. V: River flow forecasting through conceptual models part I ,Äî A discussion of principles, J. Hydrol., 10(3), 282–290, 1970.

National Operational Hydrologic Remote Sensing Center: Snow Data Assimilation System (SNODAS) Data Products at NSIDC, [2003-2012], National Snow and Ice Data Center, Boulder, Colorado USA., 2004.

Nolin, A. W.: Perspectives on Climate Change, Mountain Hydrology, and Water Resources in the Oregon Cascades, USA, Mt. Res. Dev., 32(S1), S35–S46, doi:10.1659/mrd-journal-d-11-00038.s1, 2012.

Nolin, A. W., Sproles, E. A. and Brown, A.: Climate change impacts on snow and water resources in the Columbia, Willamette, and McKenzie River Basins, USA: A nested watershed study. , in Transboundary River Governance in the Face of Uncertainty: The Columbia River Treaty, edited by B. Cosens, Oregon State University Press, Corvallis, OR., 2012.

Oregon Water Supply and Conservation Initiative: Southern Willamette valley municipal water providers final report., 2008.

Payne, J. T., Wood, A. W., Hamlet, A. F., Palmer, R. N. and Lettenmaier, D. P.: Mitigating the Effects of Climate Change on the Water Resources of the Columbia River Basin, Clim. Change, 62(1), 233–256, 2004.

Peel, M. C. and Blöschl, G.: Hydrological modelling in a changing world, Prog. Phys. Geogr., 35(2), 249–261, 2011.

Penna, D., Tromp-van Meerveld, H. J., Gobbi, A., Borga, M. and Dalla Fontana, G.: The influence of soil moisture on threshold runoff generation processes in an alpine headwater catchment, Hydrol. Earth Syst. Sci., 15(3), 689–702, doi:10.5194/hess-15-689-2011, 2011.

Reager, J. T. and Famiglietti, J. S.: Global terrestrial water storage capacity and flood potential using GRACE, Geophys. Res. Lett., 36(23), L23402, doi:10.1029/2009gl040826, 2009.

Reager, J. T., Thomas, B. F. and Famiglietti, J. S.: River basin flood potential inferred using GRACE gravity observations at several months lead time, Nat. Geosci., 7, 588–592, doi:10.1038/ngeo2203, 2014.

Riegger, J. and Tourian, M. J.: Characterization of runoff-storage relationships by satellite gravimetry and remote sensing, Water Resour. Res., 50(4), 3444–3466, 2014.

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C. J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D. and Toll, D.: The Global Land Data Assimilation System, Bull. Am. Meteorol. Soc., 85(3), 381–394, doi:10.1175/bams-85-3-381, 2004.

Rodell, M., Velicogna, I. and Famiglietti, J. S.: Satellite-based estimates of groundwater depletion in India, Nature, 460(7258), 999–1002, 2009.

Rupp, D. E., Abatzoglou, J. T., Hegewisch, K. C. and Mote, P. W.: Evaluation of CMIP5 20th century climate simulations for the Pacific Northwest USA, J. Geophys. Res. Atmos., 2013.

Sayama, T., McDonnell, J. J., Dhakal, A. and Sullivan, K.: How much water can a watershed store?, Hydrol. Process., 25(25), 3899–3908, doi:10.1002/hyp.8288, 2011.

Serreze, M. C., Clark, M. P., Armstrong, R. L., McGinnis, D. A. and Pulwarty, R. S.: Characteristics of the western United States snowpack from snowpack telemetry (SNOTEL) data, Water Resour. Res., 35(7), 2145–2160, doi:10.1029/1999wr900090, 1999.

Skøien, J. O., Blöschl, G. and Western, A. W.: Characteristic space scales and timescales in hydrology, Water Resour. Res., 39(10), 2003.

Spence, C.: A paradigm shift in hydrology: Storage thresholds across scales influence catchment runoff generation, Geogr. Compass, 4(7), 819–833, 2010.

Sproles, E. A., Nolin, A. W., Rittger, K. and Painter, T. H.: Climate change impacts on maritime mountain snowpack in the Oregon Cascades, Hydrol. Earth Syst. Sci., 17(7), 2581–2597, doi:10.5194/hess-17-2581-2013, 2013.

Swenson, S. and Wahr, J.: Post-processing removal of correlated errors in GRACE data, Geophys. Res. Lett., 33(8), L08402, doi:10.1029/2005gl025285, 2006.

Swenson, S., Yeh, P. J. F., Wahr, J. and Famiglietti, J.: A comparison of terrestrial water storage variations from GRACE with in situ measurements from Illinois, Geophys. Res. Lett., 33(16), L16401, doi:10.1029/2006gl026962, 2006.

Tague, C., Grant, G., Farrell, M., Choate, J. and Jefferson, A.: Deep groundwater mediates streamflow response to climate warming in the Oregon Cascades, Clim. Change, 86(1), 189–210, 2008.

Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F. and Watkins, M. M.: GRACE measurements of mass variability in the Earth system, Science (80-.)., 305(5683), 503–505, 2004.

The MathWorks: MATLAB and Statistics Toolbox Release 2013a, 2013.

Thompson, S. E., Harman, C. J., Schumer, R., Wilson, J. S., Basu, N. B., Brooks, P. D., Donner, S. D., Hassan, M. A., Packman, A. I. and Rao, P. S. C.: Patterns, puzzles and people: implementing hydrologic synthesis, Hydrol. Process., 25(20), 3256–3266, 2011.

United States Army Corps of Engineers: Comprehensive water supply study - An examination of current water supply issues, edited by T. M. Hillyer., 2001.

Voss, K. A., Famiglietti, J. S., Lo, M., Linage, C., Rodell, M. and Swenson, S. C.: Groundwater depletion in the Middle East from GRACE with implications for transboundary water management in the Tigris-Euphrates-Western Iran region, Water Resour. Res., 2013.

Wagener, T., Sivapalan, M., Troch, P. A., McGlynn, B. L., Harman, C. J., Gupta, H. V, Kumar, P., Rao, P. S. C., Basu, N. B. and Wilson, J. S.: The future of hydrology: An evolving science for a changing world, Water Resour. Res., 46(5), 2010.

Wagener, T., Sivapalan, M., Troch, P. and Woods, R.: Catchment classification and hydrologic similarity, Geogr. Compass, 1(4), 901–931, doi:10.1111/j.1749-8198.2007.00039.x, 2007.

Wahr, J., Swenson, S. and Velicogna, I.: Accuracy of GRACE mass estimates, Geophys. Res. Lett., 33(6), L06401, doi:10.1029/2005gl025305, 2006.

Webster, C. S., Kingston, D. G. and Kerr, T.: Inter-annual variation in the topographic controls on catchment-scale snow distribution in a maritime alpine catchment, New Zealand, Hydrol. Process., 2014.

Western, A. W., Grayson, R. B. and Blöschl, G.: Scaling of soil moisture: A hydrologic perspective, Annu. Rev. Earth Planet. Sci., 30(1), 149–180, 2002.

Whitehead, R. L.: Geohydrologic framework of the Snake River Plain regional aquifer system, Idaho and eastern Oregon, US Government Printing Office., 1992.

Wu, H., Kimball, J. S., Elsner, M. M., Mantua, N., Adler, R. F. and Stanford, J.: Projected climate change impacts on the hydrology and temperature of Pacific Northwest rivers, Water Resour. Res., 48(11), W11530, doi:10.1029/2012wr012082, 2012.

Yeh, P. J. F., Swenson, S. C., Famiglietti, J. S. and Rodell, M.: Remote sensing of groundwater storage changes in Illinois using the Gravity Recovery and Climate Experiment (GRACE), Water Resour. Res., 42(12), W12203, doi:10.1029/2006wr005374, 2006.

Zaitchik, B. F., Rodell, M. and Reichle, R. H.: Assimilation of GRACE terrestrial water storage data into a land surface model: Results for the Mississippi River Basin, J. Hydrometeorol., 9(3), 535–548, doi:10.1175/2007JHM951.1, 2008.

Zang, C. F., Liu, J., Velde, M. and Kraxner, F.: Assessment of spatial and temporal patterns of green and blue water flows under natural conditions in inland river basins in Northwest China, Hydrol. Earth Syst. Sci., 16(8), 2859–2870, 2012.

Table 1: Comparison of performance metrics using the dual-pass approach to apply GRACE TWSA data, model derived snow (SWE), and soil moisture (SM) products in predicting seasonal (R_{season}) and August (R_{Aug}) runoff by watershed. Average values for the three basins are also provided. RMSE values are in mm. Complete results can be found in Appendix table A3.

	R_{season}				R_{Aug}			
	$TWSA_{Mar}$	**$TWSA_{Apr}$**	**SWE_{Mar}**	**SM_{Mar}**	**$TWSA_{Mar}$**	**$TWSA_{Apr}$**	**SWE_{Mar}**	**SM_{Mar}**
Upper Columbia								
NSE	0.82	0.87	0.46	< 0	0.71	0.70	< 0	< 0
RMSE	33.06	27.62	56.10	> 1000	5.71	5.38	13.08	143.17
R^2	0.86	0.88	0.58	0.00	0.71	0.71	0.28	0.05
Snake River								
NSE	0.46	0.29	< 0	0.85	< 0	< 0	< 0	< 0
RMSE	14.03	15.71	21.53	7.38	13.59	0.76	0.78	0.72
R^2	0.59	0.47	0.08	0.86	0.15	0.08	0.27	0.29
The Dalles								
NSE	0.98	0.54	0.24	< 0	0.80	0.29	< 0	< 0
RMSE	6.01	26.50	26.48	122.88	1.86	3.31	18.91	22.10
R^2	0.98	0.71	0.39	0.00	0.82	0.71	0.03	0.02
Average								
NSE	0.75	0.57	0.35	0.85	0.76	0.50	< 0	< 0
RMSE	17.70	23.28	34.70	65.13	7.05	3.15	10.92	55.33
R^2	0.81	0.69	0.35	0.29	0.56	0.50	0.19	0.12

Table 2: Comparison of performance metrics from applying all nine water years of GRACE TWSA data, model derived snow (SWE), and soil moisture (SM) products in predicting seasonal (R_{season}) and August (R_{Aug}) runoff by watershed. Average values for the three basins are also provided. RMSE values are in mm. R^2 values are the same as NSE for this linear regression. Complete results can be found in Appendix table A4.

	R_{season}				R_{Aug}			
	$TWSA_{Mar}$	**$TWSA_{Apr}$**	**SWE_{Mar}**	**SM_{Mar}**	**$TWSA_{Mar}$**	**$TWSA_{Apr}$**	**SWE_{Mar}**	**SM_{Mar}**
Upper Columbia								
NSE	0.93	0.92	0.82	0.03	0.76	0.73	0.56	0.09
RMSE	22.18	23.18	36.19	82.90	6.60	6.90	8.92	12.79
Snake River								
NSE	0.83	0.75	0.34	0.93	0.68	0.52	0.62	0.76
RMSE	8.76	10.55	17.23	5.80	0.43	0.52	0.47	0.37
The Dalles								
NSE	0.98	0.91	0.67	0.00	0.88	0.91	0.46	0.02
RMSE	6.22	13.00	24.60	42.67	1.55	1.30	3.30	4.40
Average								
NSE	0.91	0.86	0.61	0.32	0.77	0.72	0.55	0.29
RMSE	12.39	15.58	26.01	43.79	2.86	2.91	4.23	5.85

Table 3: Parameters from the power function curves in each of the three watersheds using TWSA to predict streamflow. Figure 7 provides these results visually.

	Upper Columbia		**Snake River**		**The Dalles**	
	$TWSA_{Mar}$ R_{season}	**$TWSA_{Mar}$ R_{Aug}**	**$TWSA_{Mar}$ R_{season}**	**$TWSA_{Mar}$ R_{Aug}**	**$TWSA_{Mar}$ R_{season}**	**$TWSA_{Mar}$ R_{Aug}**
a	2.12E-10	4.83E-06	5.69E-05	2.26E-04	7.40E-10	3.61E-15
b	4.99	3.41	2.88	1.89	5.25	7.28
c	41.06	273.99	23.97	3.30	124.21	15.54

Table A1: The reservoirs used in the GRACE analysis.

Reservoir Name	Operating Agency	Normal Operating Capacity (m^3)
Grand Coulee	US Department of Interior	1.16×10^{10}
Libby	US Army Corps of Engineers	7.17×10^{9}
Hungry Horse	US Department of Interior	4.28×10^{9}
Dworsha	US Army Corps of Engineers	4.26×10^{9}
American Falls	US Department of Interior	2.10×10^{9}

Table A2: The groundwater wells used in the analysis that compares GRACE-derived groundwater with location-specifc wells. USGS is the United States Geological Survey and IDWR is the Idaho Department of Water Resources.

Well Number	Operating Agency
434400121275801	USGS
442242121405501	USGS
452855119064701	USGS
453239119031501	USGS
453845121191401	USGS
453937121215801	USGS
453944121211301	USGS
454013121225901	USGS
454027121212501	USGS
454040121222901	USGS
454047121203701	USGS
454100119164801	USGS
454416119212801	USGS
455418118333001	USGS
461518114090802	USGS
463750114033001	USGS
465520114074001	USGS
470049113035401	USGS
470946114013201	USGS
473442118162201	USGS
474011117072901	USGS
474251114385201	USGS
475439116503401	USGS
480519114091001	USGS
480621115244901	USGS
02S20E-01ACC2	IDWR
07S06E-29BBA1	IDWR
08S06E-03BDC1	IDWR
07S06E-34BCA1	IDWR
09S14E-03BAA1	IDWR
08S14E-16CBB1	IDWR
05S31E-27ABA1	IDWR
07N38E-23DBA1	IDWR

Table A3: Comparison of performance metrics using the dual-pass approach to apply GRACE TWSA, model derived snow (SWE), soil moisture (SM), and subsurface ($TWSA_{sub}$) data in predicting seasonal (R_{season}) and August (R_{Aug}) runoff by watershed. RMSE values are in mm.

Upper Columbia

		TWSA			SM			SWE			$TWSA_{sub}$		
		Feb	**Mar**	**Apr**	**Feb**	**Mar**	**Apr**	**Feb**	**Mar**	**Apr**	**Feb**	**Mar**	**Apr**
R_{season}	**NSE**	< 0	0.82	0.87	< 0	< 0	< 0	< 0	0.46	< 0	< 0	< 0	< 0
	RMSE	84	33	28	>1000	>1000	134	110	56	309	>1000	>1000	354
	R^2	0.43	0.86	0.88	0.01	0.00	0.07	0.23	0.58	0.27	0.15	0.02	0.02
R_{July}	**NSE**	< 0	0.90	0.84	< 0	< 0	< 0	< 0	< 0	< 0	< 0	< 0	< 0
	RMSE	32	7	8	>1000	71	56	28	25	108	>1000	>1000	123
	R^2	0.19	0.93	0.92	0.01	0.00	0.00	0.32	0.45	0.24	0.05	0.01	0.01
R_{Aug}	**NSE**	< 0	0.71	0.70	< 0	< 0	< 0	< 0	< 0	< 0	< 0	< 0	< 0
	RMSE	228	6	5	>1000	143	32	12	13	51	>1000	>1000	30
	R^2	0.19	0.71	0.71	0.07	0.05	0.30	0.25	0.28	0.12	0.18	0.11	0.01
R_{Sept}	**NSE**	< 0	< 0	< 0	< 0	< 0	< 0	< 0	< 0	< 0	< 0	< 0	< 0
	RMSE	2	21	104	4	28	10	>1000	3	50	20	587	6
	R^2	0.12	0.06	0.12	0.09	0.24	0.20	0.04	0.07	0.04	0.04	0.02	0.03

Snake River

		TWSA			SM			SWE			$TWSA_{sub}$		
		Feb	**Mar**	**Apr**	**Feb**	**Mar**	**Apr**	**Feb**	**Mar**	**Apr**	**Feb**	**Mar**	**Apr**
R_{season}	**NSE**	< 0	0.46	0.29	0.58	0.85	< 0	< 0	< 0	0.09	< 0	< 0	< 0
	RMSE	258	14	16	12	7	52	5	22	8	>1000	108	474
	R^2	0.21	0.59	0.47	0.64	0.86	0.29	0.00	0.08	0.13	0.04	0.11	0.01
R_{July}	**NSE**	< 0	< 0	< 0	< 0	< 0	< 0	< 0	< 0	< 0	< 0	< 0	< 0
	RMSE	23	3	2	2	2	40	1	2	1	99	>1000	35
	R^2	0.00	0.05	0.01	0.01	0.09	0.11	0.15	0.00	0.04	0.00	0.06	0.02
R_{Aug}	**NSE**	< 0	< 0	-0.70	< 0	< 0	< 0	< 0	< 0	0.65	< 0	< 0	< 0
	RMSE	11	13.59	0.76	1	1	2	0	1	1	>1000	>1000	474
	R^2	0.05	0.15	0.08	0.06	0.29	0.10	0.00	0.27	0.67	0.04	0.11	0.01
R_{Sept}	**NSE**	< 0	< 0	-0.94	< 0	< 0	< 0	< 0	< 0	0.03	< 0	< 0	< 0
	RMSE	16	1	1	1	1	1	0	1	0	140	8	435
	R^2	0.01	0.04	0.03	0.07	0.15	0.11	0.03	0.00	0.11	0.00	0.00	0.01

The Dalles

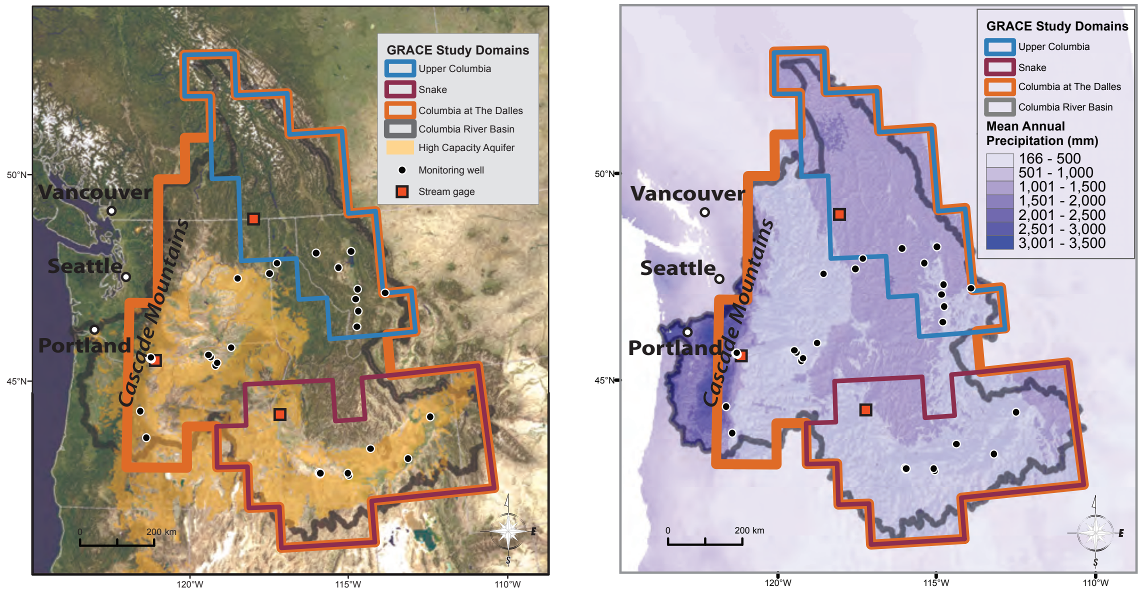
		TWSA			SM			SWE			$TWSA_{sub}$		
		Feb	**Mar**	**Apr**	**Feb**	**Mar**	**Apr**	**Feb**	**Mar**	**Apr**	**Feb**	**Mar**	**Apr**
R_{season}	**NSE**	< 0	0.98	0.54	< 0	< 0	< 0	< 0	0.24	0.14	< 0	< 0	< 0
	RMSE	84	61	27	267	122	363	>1000	26	26	13	5231	737
	R^2	0.20	0.98	0.71	0.01	0.00	0.02	0.13	0.39	0.29	0.02	0.00	0.00
R_{July}	**NSE**	< 0	0.86	< 0	< 0	< 0	< 0	< 0	0.28	< 0	< 0	< 0	< 0
	RMSE	19	3	10	>1000	16	80	>1000	4	6	4	4	311
	R^2	0.05	0.86	0.64	0.00	0.00	0.02	0.03	0.30	0.10	0.00	0.00	0.00
R_{Aug}	**NSE**	< 0	0.80	0.29	< 0	< 0	< 0	< 0	< 0	0.05	< 0	< 0	< 0
	RMSE	9	2	3	>1000	22	16	>1000	19	2	2	1	3
	R^2	0.04	0.82	0.71	0.04	0.02	0.00	0.02	0.03	0.12	0.00	0.00	0.12
R_{Sept}	**NSE**	< 0	0.41	< 0	< 0	< 0	< 0	< 0	< 0	< 0	< 0	0.05	< 0
	RMSE	5	1	3	756	3	7	1	$5x10^9$	$7x10^{10}$	6	1	2
	R^2	0.00	0.42	0.28	0.03	0.01	0.03	0.06	0.02	0.02	0.22	0.06	0.14

Table A4: Comparison of performance metrics from applying all nine water years of GRACE TWSA, model derived snow (SWE), soil moisture (SM), and subsurface ($TWSA_{sub}$) data in predicting seasonal (R_{season}) and August (R_{Aug}) runoff by watershed. RMSE values are in mm. R^2 values are the same as NSE for this linear regression.

Upper Columbia		**TWSA**			**SM**			**SWE**			**$TWSA_{sub}$**		
		Feb	**Mar**	**Apr**	**Feb**	**Mar**	**Apr**	**Feb**	**Mar**	**Apr**	**Feb**	**Mar**	**Apr**
R_{season}	**NSE**	0.84	0.93	0.92	0.01	0.03	0.33	0.63	0.82	0.62	0.15	0.22	0.22
	RMSE	28.62	19.81	20.72	8.38	14.30	36.80	37.78	30.27	37.85	28.22	32.50	32.50
R_{July}	**NSE**	0.75	0.95	0.96	0.01	0.00	0.18	0.53	0.79	0.60	0.05	0.22	0.22
	RMSE	10.38	5.00	4.74	2.16	1.34	9.10	11.95	9.80	11.73	5.38	9.86	9.86
R_{Aug}	**NSE**	0.62	0.76	0.73	0.07	0.09	0.44	0.37	0.56	0.34	0.18	0.11	0.23
	RMSE	6.02	5.31	5.48	3.12	3.50	6.15	6.00	6.16	5.87	4.80	3.95	5.22
R_{Sept}	**NSE**	0.20	0.07	0.13	0.31	0.28	0.40	0.04	0.04	0.10	0.39	0.15	0.51
	RMSE	1.60	1.05	1.32	1.85	1.80	1.96	0.80	0.80	1.22	1.95	1.42	2.00

Snake River		**TWSA**			**SM**			**SWE**			**TWSA.sub**		
		Feb	**Mar**	**Apr**	**Feb**	**Mar**	**Apr**	**Feb**	**Mar**	**Apr**	**Feb**	**Mar**	**Apr**
R_{season}	**NSE**	0.39	0.83	0.75	0.84	0.93	0.91	0.09	0.34	0.60	0.35	0.39	0.42
	RMSE	9.59	7.39	8.48	7.15	5.16	5.64	5.60	9.37	9.65	9.39	9.63	9.71
R_{July}	**NSE**	0.07	0.43	0.43	0.41	0.63	0.51	0.09	0.21	0.70	0.05	0.19	0.23
	RMSE	0.41	0.80	0.80	0.79	0.78	0.81	0.46	0.66	0.74	0.34	0.63	0.68
R_{Aug}	**NSE**	0.35	0.68	0.52	0.56	0.76	0.61	0.24	0.62	0.91	0.13	0.09	0.12
	RMSE	0.34	0.33	0.35	0.35	0.30	0.34	0.30	0.34	0.21	0.24	0.20	0.22
R_{Sept}	**NSE**	0.18	0.53	0.58	0.60	0.88	0.66	0.08	0.30	0.91	0.16	0.18	0.18
	RMSE	0.34	0.44	0.44	0.43	0.29	0.42	0.25	0.41	0.25	0.32	0.34	0.34

The Dalles		**TWSA**			**SM**			**SWE**			**TWSA.sub**		
		Feb	**Mar**	**Apr**	**Feb**	**Mar**	**Apr**	**Feb**	**Mar**	**Apr**	**Feb**	**Mar**	**Apr**
R_{season}	**NSE**	0.48	0.98	0.91	0.00	0.01	0.22	0.21	0.67	0.65	0.19	0.23	0.27
	RMSE	19.82	5.70	11.43	2.10	3.59	16.53	16.06	18.65	18.95	15.43	16.74	17.61
R_{July}	**NSE**	0.27	0.89	0.89	0.04	0.03	0.09	0.07	0.52	0.51	0.20	0.38	0.40
	RMSE	4.05	2.90	2.87	1.73	1.52	2.64	2.27	4.55	4.55	3.66	4.43	4.47
R_{Aug}	**NSE**	0.29	0.88	0.91	0.04	0.02	0.24	0.05	0.45	0.42	0.34	0.44	0.49
	RMSE	1.89	1.34	1.22	0.77	0.65	1.78	0.88	2.07	2.05	1.96	2.06	2.08
R_{Sept}	**NSE**	0.20	0.57	0.53	0.03	0.03	0.13	0.02	0.29	0.34	0.37	0.15	0.35
	RMSE	0.75	0.94	0.94	0.34	0.31	0.63	0.28	0.86	0.90	0.92	0.67	0.90



Watershed	Physical Characteristics	Climate
Upper Columbia (155,000 km^2)	- Steep topography - Low aquifer storage - Responsive runoff	Wet maritime conditions with 900 mm of annual precipitation, and substantial snowpack accumulation
Snake River (182,000 km^2)	- Low relief topography - High aquifer storage - Muted runoff	Dry continental climate with 350 mm of annual precipitation, falling primarily at upper elevations as snow
The Dalles (614,000 km^2)	- Includes the Upper Columbia and Snake River - Topography is a blend of the Upper Columbia and Snake River - High storage capacity - Large system that floods	Mix of dry continental climate plains and wet uplands. Mean annual precipitation is 625 mm, with considerable snowpack accumulation in the mountains

Fig. 1: Context map and descriptions of the three study watersheds and the locations of the groundwater wells used in the study.

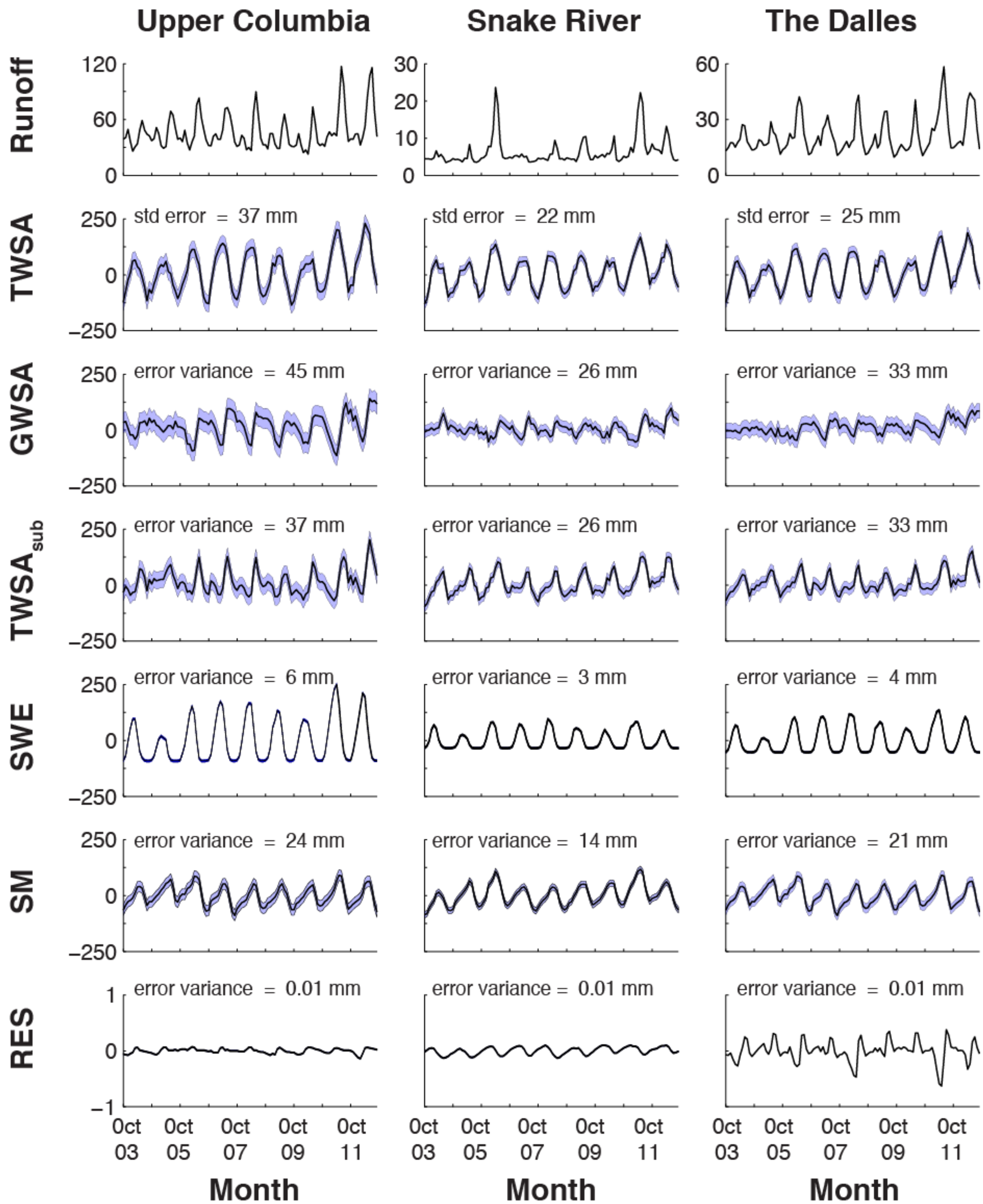


Fig. 2 Monthly storage anomalies for Runoff, *TWSA*, and the subcomponents of terrestrial water for the three watersheds. Standard errors and error variance for hydrological component are noted in each sub-figure, and represented by the blue shading. All units on the vertical axes are on the same scale and in mm. Note the different vertical scales for Runoff and *RES*. Horizontal axes are in months and on the same scale.

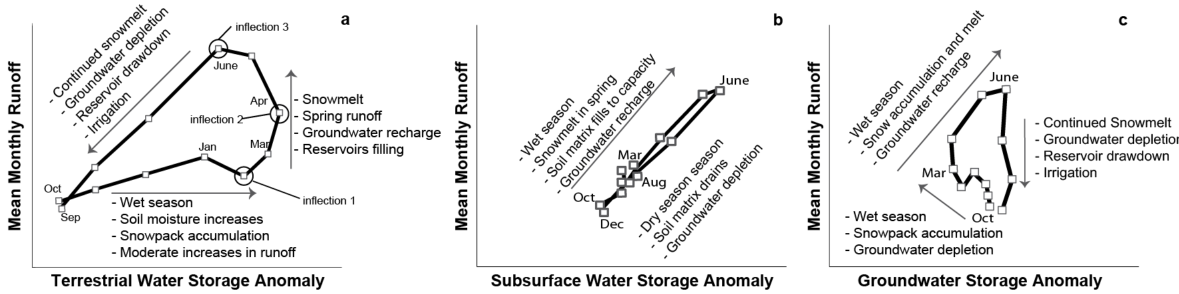


Fig. 3a-c: Annotated hysteresis curves of terrestrial water storage anomalies (a), the subsurface water storage anomalies ($TWSA_{sub}$; b), and groundwater storage anomalies (c) based upon the nine-year mean for the Columbia River at The Dalles. These curves describe the fluxes of water moving through the watershed.

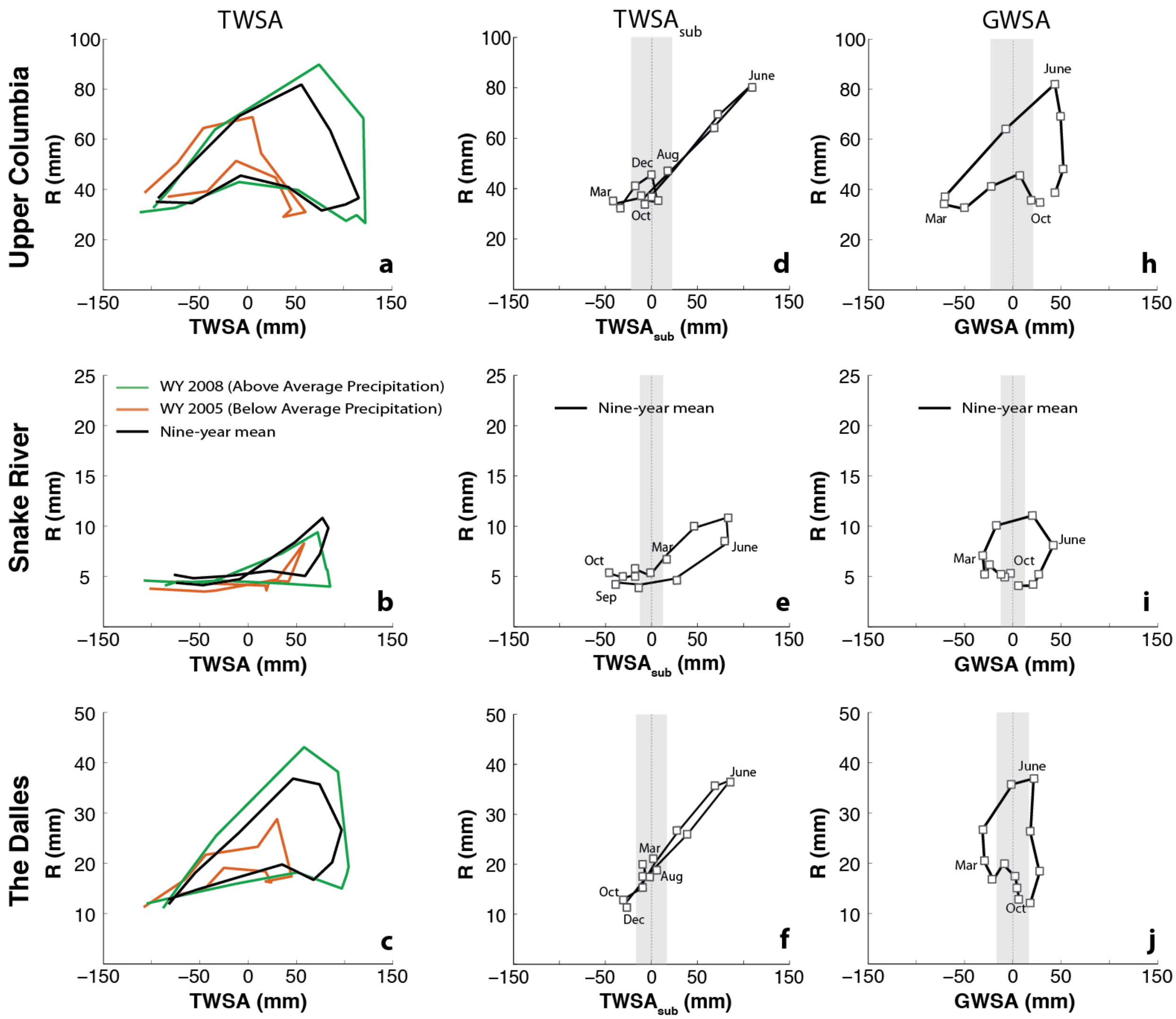


Fig. 4a-f: Individual hysteresis curves for the three study watersheds for *TWSA* (a-c), $TWSA_{sub}$ (d-f), and GWSA (h-j). $TWSA_{sub}$ in the Upper Columbia and The Dalles collapses to represent a shape more commonly associated with the hysteresis of a soil matrix. The Snake River retains a similar looping shape. The grey areas in the $TWSA_{sub}$ and *GWSA* plots provide a visual reference of the *TWSA* error variance for each watershed. The low topography and high storage capacity of the Snake aquifer provides a consistent groundwater signal, as compared to the limited aquifer of the Upper Columbia, which fills and drains quickly. Note the different scales on the y-axes.

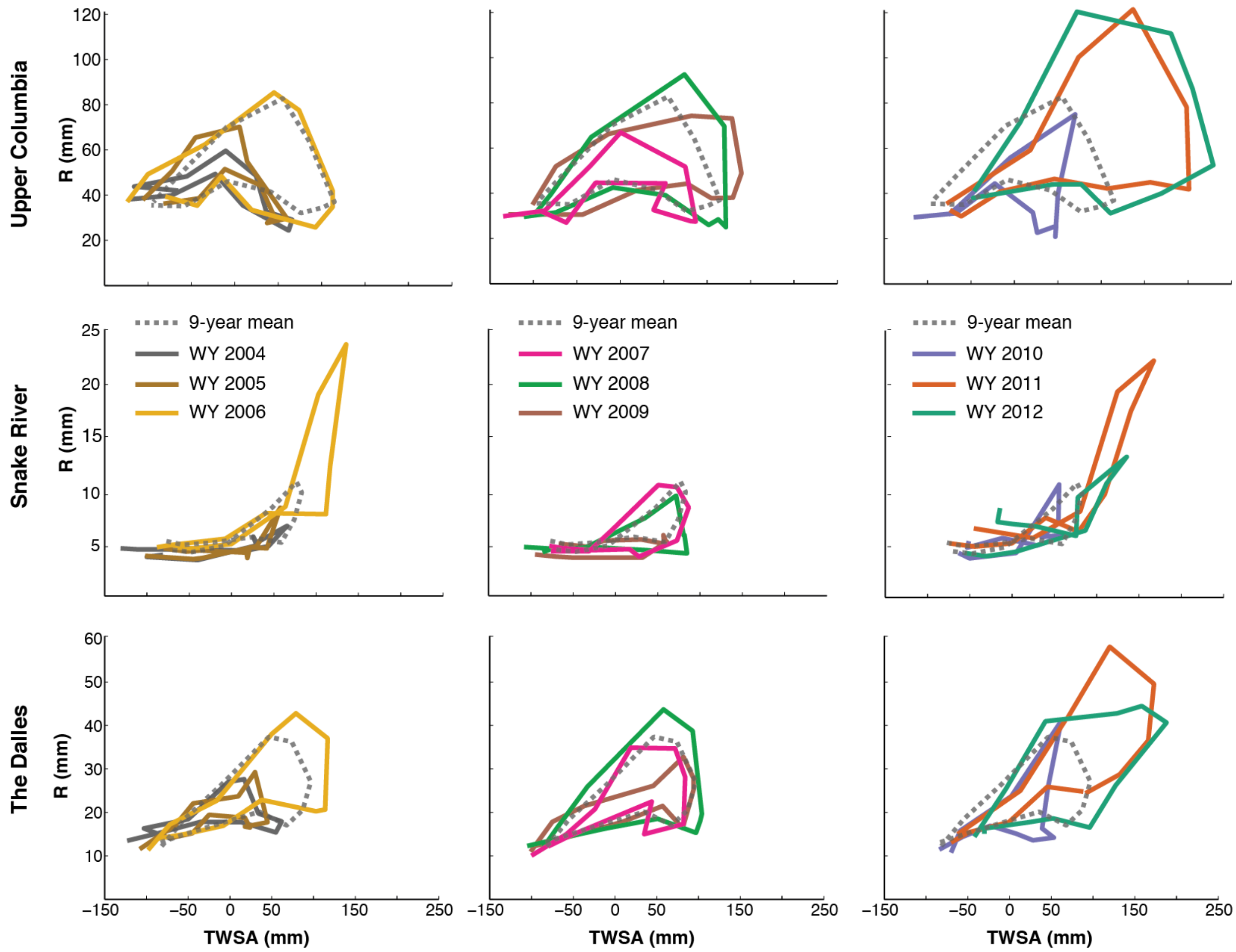


Fig. 5: Plots of the hysteresis curves for *TWSA* in each of the three study watersheds across all nine water years. For visual clarity, each plot contains three water years and the nine-year mean. Note the different scales on the y-axes for each basin.

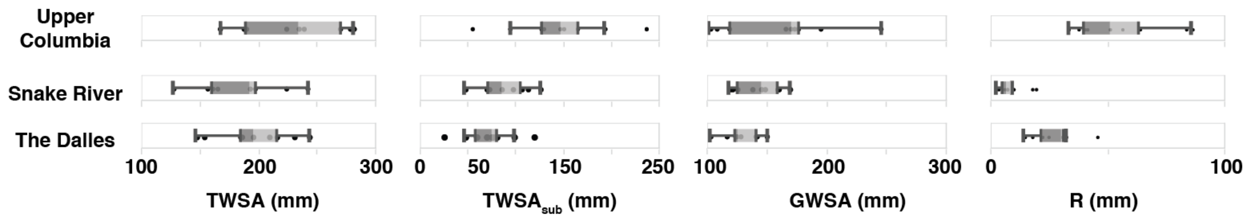


Fig. 6: The intra-annual range of *TWSA*, $TWSA_{sub,}$ *GWSA*, and *R* for the nine water years of the study period.

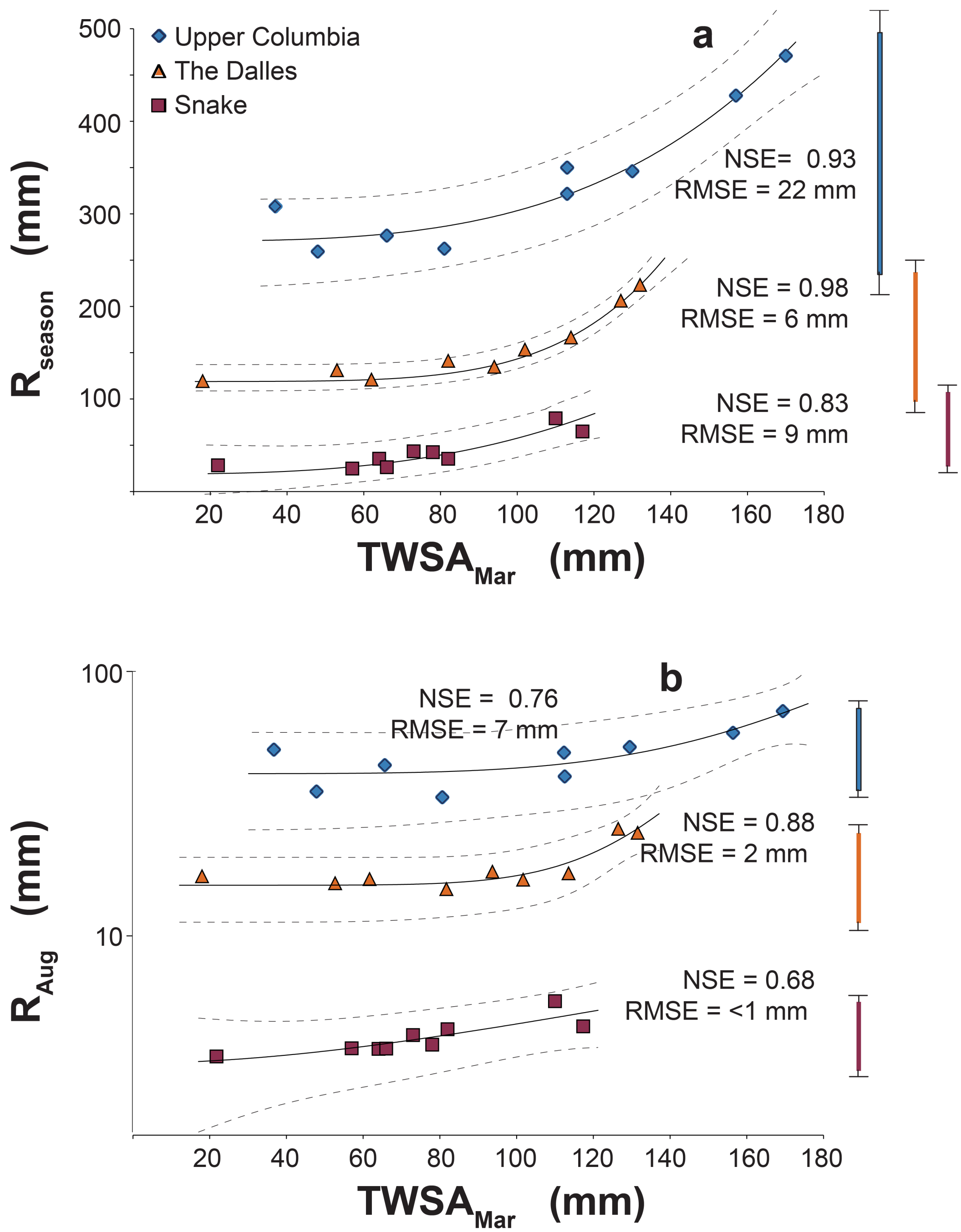


Fig. 7a-b: Measurements of terrestrial water storage anomalies in March ($TWSA_{Mar}$) effectively predict the cumulative runoff for April – September (R_{season}; a), and help describe how these three regional watersheds function as simple non-linear systems. $TWSA_{Mar}$ also predicts mean runoff for August (R_{Aug}; b), one of the driest months of the year when demand for water is at its peak. The hashed lines represent the 95% confidence intervals. The box plots to the right of each plot represent the range of R for the respective watershed from WY's 1969 – 2012. Note the semi-log y-axis on (b). For complete results and parameters from the empirical model please refer to Tables 1, 2, 3, A3, and A4.