Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-449
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 17 September 2018

Hydrology and Earth System Sciences Discussions

Catchment-scale groundwater recharge and vegetation water use efficiency

Peter A. Troch, Ravindra Dwivedi, Tao Liu, Antonio Alves Meira Neto, Tirthankar Roy, Rodrigo Valdés-Pineda and Matej Durcik

Department of Hydrology and Atmospheric Sciences, The University of Arizona

Tucson, AZ 85721, USA

Saúl Arciniega-Esparza and José Agustín Breña-Naranjo

Instituto de Ingeniería, Universidad Nacional Autónoma de México, Mexico

Key Points:

- An index of vegetation water use efficiency at the catchment scale (Horton index) is a reliable predictor of long-term average recharge
- The Horton Index can be estimated using climate and catchment properties, such as its aridity index, elevation and slope
- Average recharge rates at the catchment scale can be estimated without the need for streamflow or groundwater observations

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-449
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 17 September 2018

Hydrology and Earth System Sciences
Open Access
Discussions

Abstract

Precipitation undergoes a two-step partitioning when it falls on the land surface. At the land surface and in the shallow subsurface, rainfall or snowmelt can either runoff as infiltration/saturation excess or quick subsurface flow. The rest will be stored temporarily in the root zone. From the root zone, water can leave the catchment as evapotranspiration or percolate further and recharge deep storage. It was recently shown that an index of vegetation water use efficiency, the Horton index (HI), could predict deep storage dynamics. Here we test this finding using 247 MOPEX catchments across the conterminous US. Our results show that the observed HI is indeed a reliable predictor of deep storage dynamics. We also find that the HI can reliably predict the long-term average recharge rate. Our results compare favorably with estimates of average recharge rates from the US Geological Survey. Previous research has shown that HI can be estimated based on aridity index, mean slope and mean elevation of a catchment (Voepel et al., 2011). We recalibrated Voepel's model and used it to predict the HI for our catchments. We then used these predicted values of the HI to estimate average recharge rates for our catchments, and compared them with those estimated from observed HI. We find that the accuracies of our predictions based on observed and predicted HI are similar. This provides a novel estimation method of catchment-scale long-term average recharge rates based on simple catchment characteristics, such as climate and topography, and free of discharge measurements.

Keywords: Horton Index, Total Storage, Deep Storage, Average Recharge, Water Use Efficiency

1. Introduction

Soils, vegetation, and landforms at catchment scales have coevolved with climate, geology and tectonics (*Troch et al*., 2015) and these internal and external catchment properties define the short- and long-term water balance components. Although the role that climate (*Budyko*, 1974), climate seasonality (*Milly*, 1994; *Gentine et al*., 2012), vegetation (*Zhang et al*., 2001; *Williams et al*., 2012; *Zhang et al*., 2016; *Donohue et al*., 2012; *Donohue et al.*, 2007), soil characteristics (*Porporato et al*., 2004; *Crosbie et al.*, 2010) and landscape features (*Shao et al*., 2012; *Scanlon et al.*, 2006) exert on the long-term water balance components have been thoroughly elucidated, the first-order controls of the inter-annual and inter-catchment variability of water balance remain less understood.

Beyond Budyko's steady-state framework, several studies have investigated the role of climate, vegetation and other catchment properties in hydrological partitioning (*Farmer et al*., 2003; *Troch et al*. 2009; *Voepel et al*., 2011; *Arciniega-Esparza et al*., 2016). For instance, *Farmer et al.* [2003] followed a top-down modeling approach to study the differences in water balance of a range of semi-arid and temperate catchments in Australia as a result of climate and landscape interactions. Their results showed that process sensitivity changes with time scale, where drier catchments are more sensitive to small time-scale perturbations. *Troch et al.* [2009] studied the role of vegetation on hydrological partitioning through catchment-scale water balance. They proposed the Horton Index (HI) as the ratio of catchment vaporization and wetting (*Horton*, 1933). Applying the HI (see Equation 1) across different ecosystems and spatial scales, they found that the water use efficiency of vegetation increases in water-limited conditions. They further showed that the adaptation of vegetation to climate change is similar across different ecosystem types, this is, ecosystems tend to use water more efficiently as water availability declines. *Voepel et al.* [2011] studied catchments from different ecoregions in the USA and found that, in addition to the Aridity Index (ratio of annual potential evaporation to precipitation), the HI is also dependent on mean catchment slope and elevation, which means that the HI is related to (and hence can be predicted from) the catchment characteristics controlling water retention in the catchment. Unlike *Troch et al.* [2009] work, where catchment-scale vegetation water use efficiency was solely derived using discharge measurements, *Voepel et al.* [2011] successfully managed to test statistical regression predictors with the potential to be further applied in ungauged catchments. More recently, *Arciniega-Esparza et al.* (2016) investigated the dynamics between maximum total and deep catchment storage and their relationship to vegetation water use efficiency as represented by the HI in 33 semi-arid (water-limited) catchments in Mexico. They found that this simple index of vegetation water use efficiency is a reliable predictor of deep storage dynamics, where catchments with highly water-use efficient ecosystems tend to generate lower amounts of baseflow that sustain riparian vegetation. Furthermore, the HI could also explain how vegetation water use affects flow persistence (perennial or ephemeral), as quantified by flow duration curves. These findings suggested that catchment deep storage is both a cause and consequence of vegetation dynamics and plant water use efficiency. Note that in our definition and subsequent use of the HI term, we considered vegetation water use as synonymous to catchment vaporization (V). The latter is usually justified, since the terrestrial water loss is dominated by transpiration water loss [see i.e. *Jasechko et al.*, 2013; *Maxwell and Condon*, 2016], even in semi-arid environments (Huxman et al., 2004).

The Horton Index is defined as:

$$HI = \frac{V}{W} = \frac{P - Q_T}{P - Q_d} \qquad (1)$$

where, V is vaporization (or ET, an estimate of vegetation water use), W is the catchment wetting (or precipitation retained by the catchment), P is the precipitation, Q_T is the total streamflow, and Q_d is direct or quick runoff. If we hypothesize that HI is a reliable predictor of deep storage dynamics, or specifically of long-term average recharge rates (R), Equation (1) can be used to state three general assumptions about the relationship between HI and R. (1) When the vaporization term is zero (HI = 0), it means that the vegetation is not using water; therefore we assume that any source of precipitation would maximize the recharge rates (R_{max}), and would replenish deep storage. (2) If the vaporization equals the wetting term (HI=1), we assume that all of the water retained in the catchment is used by the vegetation; therefore the recharge rates would be zero, and deep storage would not be able to sustain streamflow during dry periods. (3) We also assume that the vegetation cannot consume more water than the one retained in the catchment (V≤W), therefore HI cannot have values larger than one. Under these three assumptions we can expect that smaller values of HI (closer to zero) are related to larger long-term average recharge rates. On the other hand, larger values of HI (closer to one) are related to lower long-term average recharge rates. Figure 1 illustrates the different components of the annual water balance that define the Horton index, and the three assumptions stated above.

The ability of HI to predict deep storage dynamics in other climatic regions and varied geological settings has not been tested yet. In this study, we extend the analysis of *Arciniega-Esparza et al*. (2016) to 247 MOPEX catchments located across the conterminous US. Additionally, since the quantification of recharge rates at the catchment scale is a challenging task and no reliable direct measurements exist, we investigate whether the HI can be used to predict average baseflow conditions and long-term average recharge rates at catchment scales. Assuming that the effective recharge ultimately discharges to a stream and that baseflow consists entirely of groundwater discharge, baseflow can provide a good approximation to recharge (*Healy*, 2010).

Since it was shown that the HI can be predicted based on the aridity index, mean slope and mean elevation of a catchment (*Voepel et al*., 2011), we further test whether predicted HI values can reliably estimate average recharge rates for our catchments. This is of utmost relevance for catchment hydrology as it goes a step further in providing a connection between groundwater recharge, aquifer discharge to streams during dry periods, streamflow regime type, and vegetation water-use efficiency. In order to accomplish this, we employ bootstrapping to quantify the

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-449
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 17 September 2018

Hydrology and Earth System Sciences
Open Access
Discussions

sensitivity, robustness and reliability of HI-based deep storage predictions. Finally, we compared our estimates of average annual groundwater recharge rates to a recharge map for the conterminous US (*Wolock*, 2003).

2. Data and Methods

2.1 Study catchments and hydrological data

For this study, 247 catchments were selected from the MOPEX (Model Parameterization Estimation Experiment) database available at: http://www.nws.noaa.gov/ohd/mopex/mo_datasets.htm (*Schaake et al*., 2006). These catchments span a wide range of climate and geomorphological settings (Figure 2), and were selected because there were no missing records of daily precipitation (P), discharge (Q), and potential evapotranspiration (PET) for our period of analysis (1980-2002). We avoided missing data in order to exclude misinterpretation of results due to data gap filling. We used the North American Regional Reanalysis (NARR; *Mesinger et al*., 2006) database to estimate actual evapotranspiration for our selected catchments. The original dataset is provided in 3-hourly time steps with 32 km spatial resolution grid. We resampled the spatially distributed actual evapotranspiration (AET) data to 6 km spatial scale and averaged in space across the catchment areas and aggregated in time to obtain daily time series. The nearest neighbor method was used to resample NARR data. This resampled NARR dataset has been previously used by other studies (NARR; *Mesinger et al*., 2006; *Durcik et al*., 2009). Please note that in general, resampling is not necessary to compute spatially averaged AET values for the MOPEX basins; however, the resampling improved the spatial representation of the grid cells, especially along the boundaries of the catchments. We finally matched all the datasets and selected 23 hydrologic (water) years (1980-2002) with complete records of P, Q, PET, and AET data. In this study, a water year is defined as the period between October 1st and September 30th. Our catchment dataset further includes catchment's landscape properties such as drainage area, mean slope, mean elevation, and mean aspect calculated using the 3 arc-second (~90m) Shuttle Radar Topography Mission (SRTM) data (*Farr et al.*, 2007).

2.2 Estimation of Total and Deep Storage Dynamics

For the purpose of our analysis, we define catchment total storage within a given hydrologic year as derived from the time integration of the daily water balance relative to some arbitrary initial value (zero total storage). Further, we define catchment deep storage as the storage related to baseflow magnitudes, assuming linear reservoir dynamics with a given reservoir constant. Table 1 summarizes the nomenclature regarding different values and statistics of total and deep storage used throughout the rest of this paper.

We used the same method as *Arciniega-Esparza et al.* (2016) to estimate total and deep storage dynamics and statistics, and we summarize these methods here for consistency. Total storage change was estimated using the daily water balance:

$$\frac{dS_T}{dt} = P - ET - Q \quad (2)$$

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-449
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 17 September 2018

where S_T is total storage relative to some arbitrary value, P is daily precipitation, E is daily evapotranspiration and Q is daily discharge. Equation (2) was integrated every hydrologic year, starting at October 1st assuming zero total storage. We corrected the annual ET time series using the water balance method to ensure that the total change of storage for any hydrologic year is zero. The correction was performed proportional to the daily values of the initial ET data. We also tested whether uncorrected ET values led to different result. We refer to the Discussion section for more details. From the time series of S_T we then selected the maximum value to represent that year's annual maximum total storage (Figure 3). We denote this annual maximum total storage for a given year and a given catchment as $S_T^{\max}$ (Table 1). For each catchment we obtained 23 annual maximum values of total storage, and the average maximum total storage, $\bar{S}_T^{\max}$, was computed. This statistic was also used by *Sayama et al*. (2011) to compare average maximum total storage to catchment properties, such as average slope.

To estimate deep storage statistics, we performed streamflow separation and baseflow recession analysis (*Tallaksen*, 1995; *Wittenberg and Sivapalan*, 1999; *Sayama et al*., 2011; *Arciniega-Esparza et al*., 2016). Streamflow was partitioned into quick flow (Q_d) and baseflow (Q_b) components using a recursive low-pass filter (*Lyne and Hollick*, 1979). The one-parameter low-pass filter was passed three times over the time series, two times forward and one time backward, to smoothen the baseflow hydrograph (*Voepel et al*., 2011). We selected a parameter value of 0.925 for all catchments, which is similar to the approach used in previous studies [*Voepel* *et al.*, 2011; *Arciniega-Esparza et al.* 2016;]. This method has proven to be an effective tool to investigate the characteristics of storage feeding streams (*Brutsaert*, 2008; *Rupp and Woods*, 2008; *Sayama et al.*, 2011). Once the annual baseflow hydrograph was obtained, we performed baseflow recession analysis, assuming that the deep storage dynamics can be represented by means of a linear reservoir:

$$\frac{dS_D}{dt} = -Q_b = -\frac{S_D}{K} \qquad (3)$$

where S_D is deep storage, Q_b is baseflow, and K is the linear reservoir constant. Equation (3) can be rewritten as:

$$-\frac{dQ_b}{dt} = \frac{1}{K}Q_b \qquad (4)$$

Let $-dQ_b/dt=Y$ and $Q_b=X$. Equation (4) in terms of new variables X and Y is:

$$Y = (1/K)X \qquad (5)$$

Transforming Equation (5) in terms of square error (e_i) for i=1,2,...N, where N= total number of days when $-dQ_b/dt$ or Y is positive, leads to the following equation:

$$e_i^2 = (Y_i - \frac{1}{K}X_i)^2 \qquad (6)$$

An optimization of Equation (6) with respect to the variable 1/K leads to the following Equation:

$$\frac{d\sum_{i=1}^{N} e_i^2}{d(\frac{1}{K})} = 0 = -2\sum_{i=1}^{N} X_i(Y_i - \frac{1}{K}X_i) \qquad (7)$$

Thus, the value of the reservoir constant K for which Equation (6) represents a local minimum is:

$$K = \frac{\sum_{i=1}^{N} X_i^2}{\sum_{i=1}^{N} X_i Y_i} = \frac{\sum_{i=1}^{N} Q_{b,i}^2}{\sum_{i=1}^{N} Q_{b,i}(\frac{-dQ_b}{dt})_i} \qquad (8)$$

Equation (8) was used for estimating K for each hydrologic year and for each catchment. Annual maximum deep storage ($S_D^{\max}$) was then computed using the maximum baseflow value for each year (see Figure 3 for an illustration of our method), multiplied by the K value from the same year. The 23 annual maximum deep storage values obtained for each catchment were averaged to obtain the average maximum deep storage, $\bar{S}_D^{\max}$, of the catchments.

2.3 Predictive relationships of storage dynamics and groundwater recharge

Since total and deep storage are estimated using independent methods, the strength of their relationship was analyzed to determine common patterns of inter-annual and inter-catchment variability of storage dynamics. Several catchment geomorphological properties (i.e. mean slope, drainage area, mean elevation, and mean aspect, among others) were explored to predict storage dynamics. We also explored the relationships with catchment vaporization (V), catchment wetting (W), and the Horton Index, respectively. We further examined whether the HI can predict certain statistics of the flow duration curve (FDC). We selected the 50$^{\text{th}}$ percentile flow as a surrogate for average baseflow conditions. We tested whether the 50$^{\text{th}}$ percentile represents accurately the observed average recharge rates for the catchments. We also tested whether other streamflow percentiles would better represent average baseflow conditions (see Discussion section for more details). The advantage of using the FDC is that one does not have to perform hydrograph separation to estimate average baseflow conditions and the average recharge rates at catchment scale. As our MOPEX watershed dataset is composed of undisturbed hydrological systems (*Schaake et al.*, 2006) and our analysis is based on 23 hydrologic years of climate data, aquifer storage tends to remain constant over the long-term so that the steady-state hypothesis is valid (see Discussion in *Donohue et al.*, 2007). In such cases, the difference between average baseflow conditions and actual average recharge may be within the range of measurement uncertainty for baseflow (*Healy*, 2010). A summary of our research structure is presented in Figure 4.

2.4 Independent estimation of average groundwater recharge rates

It was previously shown that the HI can be reliably predicted based on the catchment's aridity index, mean slope and mean elevation (*Voepel et al.*, 2011). We tested whether predicted HI values can accurately estimate average recharge rates, and how these predictions compare to those based on observed HI values. For this purpose, we recalibrated Voepel's model, as that study used a different subset of MOPEX catchments.

2.5 Quantifying sensitivity, robustness and reliability of predictive relationships

We performed an uncertainty analysis using the Bootstrapping method to answer the following questions: How much of the variance in the relationship between $\bar{S}_D^{\max}$ and HI can be explained by a linear fit between the two variables as a function of the number of catchments used in the analysis (sensitivity)? How does the slope and intercept of the linear fit between $\bar{S}_D^{\max}$ and HI vary as a function of sample size (robustness)? How much does the estimation error vary with sample size (reliability)? For answering the first two questions, we randomly selected 1000 samples with replacement excluding x% of 247 catchments (called cutoff % here) and performed linear regression analysis to compare explained variance and the values of the slope and the intercept of the linear regression line for different cutoff levels. For performing the latter reliability analysis, we again randomly selected 1000 samples with replacements excluding x% of 247 catchments and performed a best-fit analysis for predicting the $\bar{S}_D^{\max}$ vs. HI relationship. We then used the predicted relationship to estimate $\bar{S}_D^{\max}$ for the excluded x% catchments. Subsequently, we computed the Mean Squared Error (MSE) at each of the cutoff % using the estimated and known $\bar{S}_D^{\max}$ values for those excluded x% catchments.

3. Results

3.1 Linear correlation between annual maximum total and deep storage

The inter-annual linear correlation between the time series of annual maximum total and annual maximum deep storage revealed that about 192 out of the 247 catchments show a positive correlation (Figure 5), and 55 catchments have negative correlations. Of the 192 catchments with positive correlation between total and deep storage, 96 catchments show a statistically significant correlation ($p < 0.05$), and only one catchment of the 55 catchments with a negative correlation was statistically significant. We removed the catchment that had a significantly negative correlation between total and deep storage to avoid including a catchment that possibly is affected by anthropogenic changes (e.g. pumping).

3.2 Predictors of average maximum total and deep storage

Inter-catchment variability: Similar to previous findings regarding relationships between catchment properties and storage dynamics (*Sayama et al*., 2011; *Voepel et al*., 2011; *Arciniega-Esparza et al*., 2016), we too found that the mean catchment slope is a strong control on average maximum total and average maximum deep storage (R=0.73 and R=0.69, respectively). We also found that the inter-catchment variability of total and deep storages is significantly correlated with mean catchment wetting, and this relationship seems to be stronger than with catchment slope (Figure 6, A1 and B1). On the other hand, vaporization does not seem to correlate significantly with any of the storages (Figure 6, A2 and B2). Interestingly and similar *to Arciniega-Esparza et al.* (2016), deep storage dynamics are strongly and significantly correlated with the mean HI, while the correlation between HI and total storage is weaker than between wetting and total storage. It shouldn't be surprising that the HI is related to the average maximum deep storage, as the latter is related to the average baseflow. Since the empirical HI is derived from baseflow separation, there is obviously a strong relationship with average baseflow. The linear pattern shown in Figure 6 (B3) reveals the nature of the relationship between empirical HI and average maximum deep storage, and indicates that HI can be an candidate to predict deep storage dynamics at regional scales. The fact that HI expresses catchment vegetation water use efficiency indicates the important role of terrestrial vegetation in controlling deep groundwater percolation which sustains baseflow conditions across a wide range of climates and geological settings (*Troch et al., 2009*). It remains to be investigated whether predictions of the empirical HI based on independent climate and catchment characteristics can be used to estimate storage dynamics at regional scales (see Section 3.3.2).

Inter-annual variability: Figure 7 shows the correlation coefficients between storage dynamics and water balance components (wetting, W, vaporization, V, and their ratio, HI=V/W) based on annual values. Considering the coefficients of determination (R^2, not shown), the inter-annual variability of storage dynamics compared to wetting, vaporization, and HI revealed that catchment wetting could significantly explain inter-annual variability of total storage for 68% of the catchments (Figure 7, A1). Catchment wetting inter-annual variability can also explain inter-annual variability of deep storage of about 46% of the catchments (Figure 7, B1). Catchment vaporization explained inter-annual variability of the total and deep storages for 51% and 22% of the catchments, respectively. The inter-annual variability of the HI could explain inter-annual variability of deep storage for 95% of the catchments (Figure 7, B3), but total storage inter-annual variability was only explained for 27% of the catchments (Figure 7, A3). Again, it is not surprising that the annual empirical HI is related to annual deep storage dynamics, and thus to annual average baseflow. Figure 7-B3 simply reveals the nature of this relationship.

3.3 The Horton Index, streamflow persistence and average recharge rates

3.3.1 Observed HI based estimates

We further investigated whether the observed HI is a good predictor for catchment-scale groundwater recharge. We found that there is a clear pattern between the catchment HI and the relative position of the flow duration curve (Figure 8). Low values of HI correspond to sustained higher flow, i.e., perennial streams, whereas high values of HI are related to much lower flows, i.e., ephemeral streams. We found that for HI=0.86, catchments switch between perennial to ephemeral flow regimes. The value of HI=0.86 was determined from the flow duration curves subject to the condition that the lowest HI for which streamflow first approaches a zero value. We call this the "Critical Horton Index". Arciniega-Esparza et al. (2016) found the same critical HI value for the semi-arid catchments in Mexico.

The relationship between mean baseflow derived from our baseflow separation method (Q_b) and the 50th percentile (Q_{50}) of the catchment streamflow is strongly significant (R^2=0.96; p=0.00) (Figure 9A). However, the slope of the relationship is different from 1 (0.7), therefore predictions of average baseflow based on Q_{50} could be overestimated. Q_{50} is only an efficient surrogate of average baseflow conditions for the catchments under consideration if the slope of the regression line is one. In the Discussion section, we address this limitation of the method. Nevertheless, we found that the linear correlation between Q_{50} and HI is strongly significant ($p<0.05$), and explains 72% of the variance (Figure 9B). These results suggest that under long-term steady state conditions, recharge rates can be predicted using an index of water use efficiency.

Our estimates of average recharge rates (using the results from Figure 9) compared against the average recharge rates map from the USGS [*Wolock*, 2003], revealed a strong similarity between their spatial variability (Figure 10). Moreover, the inter-catchment variability revealed a significant positive correlation while comparing the absolute values of average recharge rates from both sources (Figure 11). This finding is encouraging as the recharge rates in both sources are calculated differently. The USGS groundwater recharge estimates are derived from the baseflow index - the ratio of base flow to total flow - map for the conterminous US [*Wolock*, 2003], while ours is based on the Horton index.

3.3.2 Predicted HI based groundwater recharge estimations

The estimation of deep storage dynamics and hence average recharge rates based on the HI is only useful if we are able to estimate the HI independently. Without such independent estimates the method becomes circular: derive the HI from baseflow separation to predict baseflow characteristics. *Voepel et al*. (2011) used a multiple linear regression model to predict HI based on aridity index (AI), mean slope (β) and mean elevation (Z). Since *Voepel et al*. (2011) used a different subset of MOPEX catchments, we recalibrated his multiple regression model for our 246 catchments and found the best fit (R^2=0.72) using:

$$HI=0.78+0.11\ \ln(AI)-0.03\ \beta-0.04Z \qquad (9)$$

Note that β is fractional slope and ln(AI), β and Z are normalized variables using mean and standard deviations of the observed values for the 246 MOPEX catchments. For normalization of ln(AI), β and Z, the following values were used: (i) mean and standard deviation of ln(AI) is -0.06 and 0.45, resp.; (ii) mean and standard deviation of β (fractional) is 0.06 and 0.05, resp.; (iii) mean and standard

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-449
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 17 September 2018

deviation of Z (m) is 631.73 m amsl and 613.68 m amsl, resp. Figure 12 shows the comparison between observed and predicted HI based on Equation 9. The best linear fit corresponded to a coefficient of determination R^2 of 0.78.

As shown in Figure 13, estimates of catchment-scale groundwater recharge based on predicted HI values are very similar to those based on observed HI values. We thus have now a method that can estimate regional recharge rates based on easily obtainable catchment characteristics, such as climate and topography.

3.4 Reliability of estimation methods

In order to test the reliability of the methods developed, we performed a bootstrapping analysis to examine the effect of sample size on model performance. First, we investigated how the explained variance of our average recharge rate model changes with sample size (Figure A1). As we systematically decrease the number of catchments in our sample, we observe that the range of coefficients of determination increases symmetrically around a very constant mean. Even with 95% of the catchments removed from model fitting, the model can, on average, explain the same amount of variance as the model based on all 246 catchments. With 50% of the catchments removed, we see that the range of explained variance is between 0.7 and 0.9, a relatively narrow range.

Next, we looked at the robustness of the linear regression model in terms of slope and intercept value (Figure A2 and A3). Again, we observe that on average the slope and intercept values of the linear regression models are very stable, and that the uncertainty about the mean values grows symmetrically about the mean with decreasing sample size.

Finally, the reliability of the linear regression model was tested (Figure A4). After removing x% of catchments to fit a linear regression model and using the same model to predict the average recharge rate for the remaining x% catchments, we see that the MSE slightly increases. Moreover, removing 5% of all catchments results in mean square errors (MSE) having a wide range, from very low MSE to very high MSE. This is due to the fact that it is very likely that in 1000 runs, 5% 'good' catchments are selected (catchments that fall on the regression line estimated by the remaining 95%) but it is equally likely that 5% 'bad' catchments (catchments that don't fall on the regression line) are also selected. As the number of catchments left out of the analysis are increased, the probability of selecting all 'good' or all 'bad' catchments decreases, and this is reflected in the reduced range of MSE values. Only when leaving out 95% of the catchments, there is an increase in the range of MSE values, because the remaining 5% catchments are likely to result in unreliable regression models.

4. Discussion

The results presented in this paper confirm that average maximum total storage and average maximum deep storage are positively correlated at a similar level (78%) compared to a previous study (73%) performed across an arid region (Arciniega-

Esparza et al., 2016). Other similarities with *Arciniega-Esparza et al*. (2016) study were found when examining the relationship between maximum catchment storage and its mean catchment slope. In fact, the role of slope in total and deep storage was more prominent for the MOPEX dataset (R=0.73 and R=0.69, respectively) than in Mexico's drylands (R=0.59 and R=0.52), but very similar (R=0.74 only for maximum storage) to a group of small watersheds in Northern California (*Sayama et al*., 2011). This suggests that in arid regions, catchment slope become less relevant to predict its maximum storage capacity when compared to wetter regions.

Nevertheless, the variables that correlate strongest with deep storage dynamics were catchment wetting (R=0.70) and the Horton Index (R=-0.88), respectively. Surprisingly, the magnitude of the correlation between these variables and different storages across the region of study is very similar and only ~10% of the catchments seem to be out from these observed patterns.

In addition to the observed connection between deep storage and the HI, the HI turns out to be a simple but reliable classification tool for streamflow persistence. A threshold value (named the Critical Horton Index) that separates streams with perennial regimes from intermittent and/or ephemeral flows was found similar as in *Arciniega-Esparza et al*. (2016) study. This threshold determines the link between vegetation water-use efficiency and the time that a minimum amount of groundwater outflow (Q>0) to a stream is likely to occur. It suggests that catchments must reach high vegetation water-use efficiency levels (HI>0.86) in order to deplete the active catchment storage (storage that continuously interacts with the stream network) during some time of the hydrologic year.

In addition to the assumptions mentioned in the introduction of this paper, there are other important assumptions underlying the proposed method of using the Horton index estimated from climate and landscape characteristics that can limit its applicability. First, in order to interpret the Horton index as a vegetation water use efficiency, as in Troch et al. (2009), we have to assume that transpiration is the main component in catchment vaporization or evapotranspiration. This is generally the case in vegetated landscapes, but will obviously not the case in poorly vegetated arid catchments. Therefore, the proposed method cannot be used to estimate recharge rates in desert landscapes (except, perhaps, in the Sonoran Desert of the SW USA, which has significant green cover throughout the rainy seasons).

Second, the low pass filter of *Lyne and Hollick* (1979) is a reasonable method to separate quick runoff from baseflow. Without additional information about the chemistry of streamflow and the different sources of streamflow (groundwater, soil water and rainfall) there is no way to test the reliability of this baseflow separation method. In this study, we have opted to use a single parameter that defines the cut-off frequency of the low pass filter across all catchments, but if better information about end-members' chemistry is available this parameter could possibly be optimized for specific catchments.

Third, we assume that baseflow is equal to recharge, and that aquifer storage is constant over the long term. This will only be true in catchments with unregulated streamflow, without upstream extraction of surface and/or groundwater, without lateral groundwater flow into and out of the catchment topographic divide, and without significant evapotranspiration from the groundwater table. The MOPEX catchments were selected to avoid such effects on streamflow, but since the compilation of the database it is possible that some catchments have undergone some type of anthropogenic impacts.

Finally, we assumed that long-term average baseflow conditions can be estimated from the 50^{th} percentile of the flow duration curve (Q_{50}). This choice was more an intuitive guess than an informed decision. We tested this assumption for our catchments and found that Q_{65} in fact is a much better estimate of average baseflow conditions. For the selected catchments, there exists a one-to-one relationship between Q_{65} and average baseflow, and both variables are highly correlated (R^2 = 0.90). Before we can replace Q_{50} with Q_{65} more research is required to check how universal this finding is across different climates and geologies.

There can be a perceived possibility that some of the statistical relationships shown in this study are the result of spurious correlation (i.e. The HI is estimated from baseflow separation and is then used to estimate average baseflow conditions). There are several arguments that go against this statement, but the strongest that we can think of is the fact that we use predicted HI from climate and landscape properties to estimate average baseflow and long-term recharge. The fact that the HI is correlated with climate and landscape properties indicate that we can interpret the HI as an index of catchment co-evolution of soils, vegetation and landforms with climate and geology. It is therefore possible that the proposed method would only work in catchments where vegetation is in equilibrium with the local climate. Non-native species colonizing catchments could possibly off-set such equilibrium, and therefore caution is warranted to apply the method when that is the case.

5. Conclusions

In this paper, three different recharge estimation methods were compared: (i) a conventional method (*Wolock et al.*, 2003) based on physical hydrology (annual runoff and baseflow index), (ii) a methodology based on a catchment ecohydrological index derived from observed discharge records (*Troch et al.*, 2009), and (iii) a methodology based on a predicted ecohydrological index using climate and topography data (*Voepel et al.*, 2011). Differences between groundwater recharge estimates from the observed and predicted HI are minimal, suggesting that both methods are good proxies for long-term average groundwater recharge. In this work, we showed that the former can be widely used in gauged catchments while the latter has the potential to be applied in ungauged catchments.

This synthesis study tested a parsimonious modeling framework based on the Horton Index, a catchment index that combines in a simple way the effects of storage capacity of soils (*Horton*, 1933), topography (*Voepel et al.*, 2011; *Thompson et al.*, 2012), and ecosystems (*Troch et al.*, 2009; *Huxman et al.*, 2004) on hydrological partitioning. It effectively expresses how much water, available to the catchment's ecosystem, is being used by the plants. This water use efficiency metric increases with climate aridity while its value becomes less in temperate and humid regions. From our results we observe that when the HI crosses a critical value (0.86), catchments' main streams become intermittent or ephemeral. The flow duration curves of catchments rank quantitatively with HI (high to low flows corresponding to low and high HI, respectively), suggesting that the statistical distribution of streamflow regimes, including median flows (a proxy for recharge), is primarily controlled by vegetation water use.

This work (based on a large hydrological sample) provides further evidence that: (a) maximum total and deep catchment storage are mainly positively correlated and (b) vegetation water-use efficiency (the Horton Index) is a robust predictor for low-flows and groundwater recharge in catchments with different types of climate, soils, geology and vegetation cover. Future work could explore the suitability of the Horton Index for inferring low-flows and recharge at shorter timescales, analyze long-term changes in climate, water-use efficiency and catchment storage, and investigate the behavior of HI at smaller spatial scales (e.g. zero-order basins and experimental plots/hillslopes), among others.

6. Acknowledgements

This work was greatly improved from comments made by three anonymous reviewers. The authors acknowledge funding through a CAZMEX (UA-CONACYT) research grant. Antonio Meira would like to acknowledge financial support from the Brazilian Mobility Program, through the CAPES Foundation, Brazil. Rodrigo Valdés-Pineda would like to acknowledge funding from the National Commission of Science and Technology of Chile (CONICYT). The data used to produce this paper is available upon request with Matej Durcik.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-449
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 17 September 2018

7. References

Arciniega-Esparza, S., Breña-Naranjo, J. A., and Troch, P. A., On the connection between terrestrial and riparian vegetation: the role of storage partitioning in water-limited catchments, Hydrological Processes, 10.1002/hyp.11071, 2016.

Brutsaert, W., Long-term groundwater storage trends estimated from streamflow records: Climatic perspective. Water Resources Research, 44(2), 2008..

Budyko, M. I., Climate and Life, Int. Geophys. Ser., vol. 18, 508 pp., Academic, N. Y., 1974

Crosbie, R. S., Jolly, I. D., Leaney, F. W., and Petheram C., Can the dataset of field based recharge estimates in Australia be used to predict recharge in data-poor areas?, *Hydrology and Earth System Sciences*, *14*(10), 2023-2038, 2010

Donohue, R. J., Roderick, M. L., & McVicar, T. R., On the importance of including vegetation dynamics in Budyko's hydrological model. Hydrology and Earth System Sciences, 11, 983–995, 2007

Farmer, D., Sivapalan, M., & Jothityangkoon, C., Climate, soil, and vegetation controls upon the variability of water balance in temperate and semiarid landscapes: Downward approach to water balance analysis. Water Resources Research, 39(2), 2003.

• Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D., The shuttle radar topography mission, Rev. Geophys., 45, RG2004, doi:10.1029/2005RG000183, 2007.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-449
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 17 September 2018

Hydrology and Earth System Sciences
Discussions

Gentine, P., D'Odorico, P., Lintner, B. R., Sivandran, G., and Salvucci, G., Interdependence of climate, soil, and vegetation as constrained by the Budyko curve, Geophys. Res. Lett., 39, L19404, doi:10.1029/2012GL053492, 2012

Healy, R. W., Estimating Groundwater Recharge, Cambridge Univ. Press, Cambridge, U. K, 2010

Horton, R. E., The role of infiltration in the hydrologic cycle. Transactions of the American Geophysical Union 14: 446– 460, 1933.

Jasechko, S., Sharp, Z. D., Gibson, J. J., Birks, S. J., Yi, Y., and Fawcett, P. J., Terrestrial water fluxes dominated by transpiration, *Nature*, *496*(7445), 347-350, 2013.

Lyne, V. and Hollick, M., Stochastic time-variable rainfall runoff modeling, in Hydrology and Water Resources Symposium, Natl. Comm. on Hydrol. and Water Resour. of the Inst. of Eng., Perth, Western Aust., Australia., 1979.

Maxwell, R. M., and Condon, L. E., Connections between groundwater flow and Transpiration partitioning, *Science*, *353*(6297), 377-380, 2016.

Milly, P., Climate, soil-water storage, and the average annual water-balance, Water Resour. Res., 30(7), 2143–2156, doi:10.1029/94WR00586, 1994.

Porporato, A., Daly, E., and Rodríguez-Iturbe, I., Soil water balance and ecosystem response to climate change, Am. Nat., 164(5), 625–632, doi:10.1086/424970, 2004.

Rupp, D. E., and Woods, R. A., Increased flexibility in base flow modelling using a power law transmissivity profile. Hydrological processes, 22(14), 2667-2671, 2008.

Sayama, T., McDonnell, J. J., Dhakal, A., and Sullivan, K., How much water can a

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-449
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 17 September 2018

watershed store?. Hydrological Processes, 25(25), 3899-3908, 2011.

Scanlon, B. R., Keese, K. E., Flint, A. L., Flint, L. E., Gaye, C. B., Edmunds, W. M., and Simmers, I., Global synthesis of groundwater recharge in semiarid and arid regions, *Hydrological Processes*, *20*(15), 3335-3370, 2006.

Schaake, J., Cong, S., and Duan, Q., U.S. MOPEX DATA SET, NOAA National Weather Service, Silver Spring, MD, 23 p., 2006.

Shao, Q., Traylen, A., and Zhang, L., Nonparametric method for estimating the effects of climatic and catchment characteristics on mean annual evapotranspiration, Water Resour. Res., 48, W03517, doi:10.1029/2010WR009610, 2012.

Tallaksen, L. M., A review of baseflow recession analysis. Journal of hydrology, 165(1-4), 349-370, 1995.

Troch, P. A., Martinez, G. F., Pauwels, V. R. N., Durcik, M., Sivapalan, M., Harman, C., Brooks, P. D., Gupta, H. and Huxman, T., Climate and vegetation water use efficiency at catchment scales, Hydrol. Process., 23(16), 2409–2414, doi:10.1002/hyp.7358, 2009.

Troch, P. A., Lahmers, T., Meira, A., Mukherjee, A., Pedersen, J. W., Roy, T., and Valdes-Pineda, R., Catchment coevolution: A useful framework for improving predictions of hydrological change?, Water Resour. Res., 51, doi:10.1002/2015WR017032, 2015.

Voepel, H., Ruddell, B., Schumer, R., Troch, P. A., Brooks, P. D., Neal, A., ... & Sivapalan, M., Quantifying the role of climate and landscape characteristics on hydrologic partitioning and vegetation response. Water Resources Research, 47(10), 2011.

Christopher A. W., Reichstein M., Buchmann N., Baldocchi D., Beer C., Schwalm C., Wohlfahrt G., Hasler N., Bernhofer C., Foken T., Papale D., Schymanski S., Schaefer K., Climate and vegetation controls on the surface water balance: Synthesis of evapotranspiration measured across a global network of flux towers, Water Resour. Res., 48, W06523, doi:10.1029/2011WR011586, 2012.

Wittenberg, H., and Sivapalan, M., Watershed groundwater balance estimation using streamflow recession analysis and baseflow separation. Journal of hydrology, 219(1), 20-33, 1999.

Wolock, D. M., Estimated mean annual natural ground-water recharge in the conterminous United States, edited, U.S. Geological Survey, VA, 2003.

Zhang, L., Dawes, W. R., and Walker, G. R., Response of mean annual evapotranspiration to vegetation changes at catchment scale. Water resources research, 37(3), 701-708, 2001.

Zhang, S., Yang, H., Yang, D., & Jayawardena, A. W., Quantifying the effect of vegetation change on the regional water balance within the Budyko framework. Geophysical Research Letters, 43(3), 1140-1148, 2016.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-449
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 17 September 2018

Hydrology and Earth System Sciences
Open Access
Discussions

Appendix A

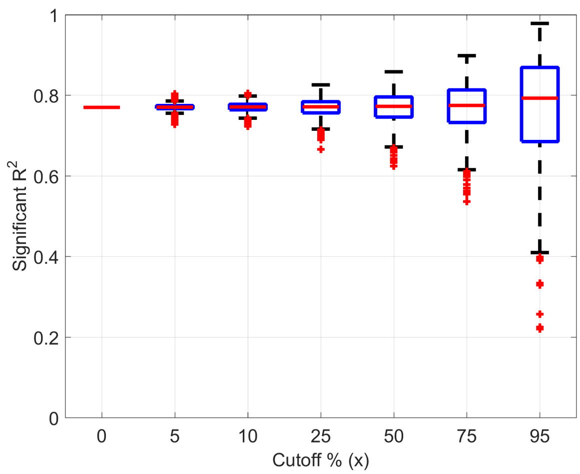


Figure A1: Box-Whisker plots of explained variance (R^2: coefficients of determination) with increasing cutoff level

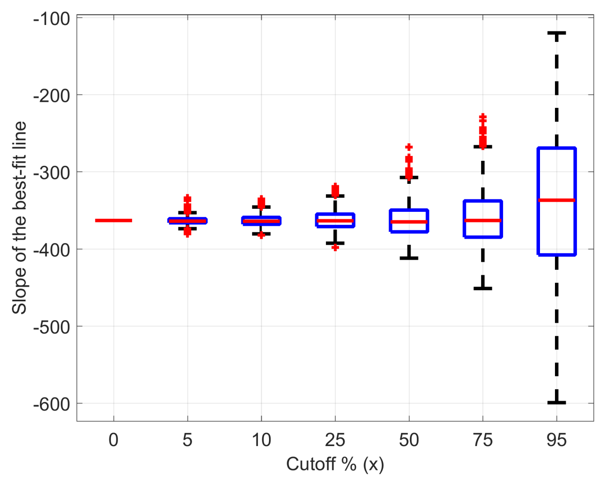


Figure A2: Box-Whisker plots of slope of the linear regression model with increasing cutoff level

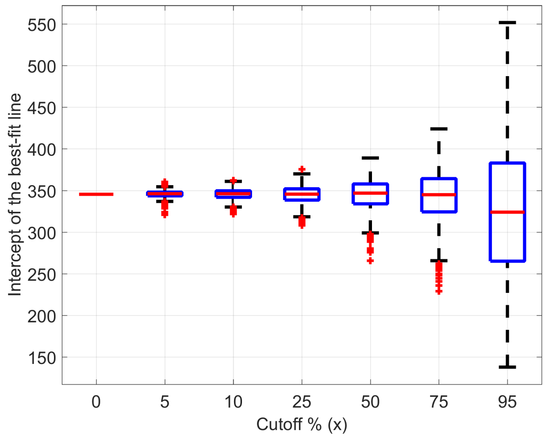


Figure A3: Box-Whisker plots of intercept of the linear regression model with increasing cutoff level

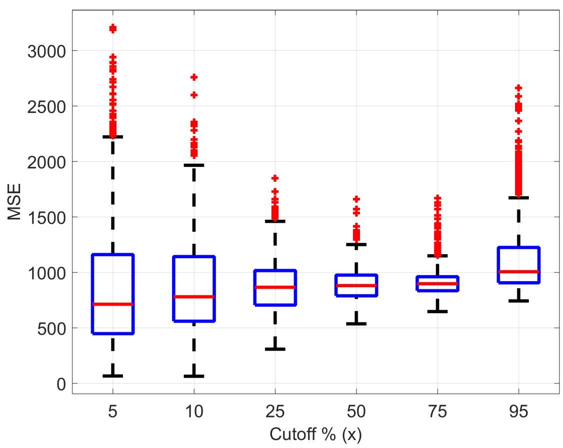


Figure A4: Box-Whisker plots of the mean squared error (MSE) with increasing cutoff level

Figures

$$HI = \frac{V}{W} = \frac{P - Q_T}{P - Q_d} = \frac{P-(Q_d+Q_b)}{P-(Q_T-Q_b)} = \frac{\text{P that does not become streamflow}}{\text{P retained by the catchment}}$$

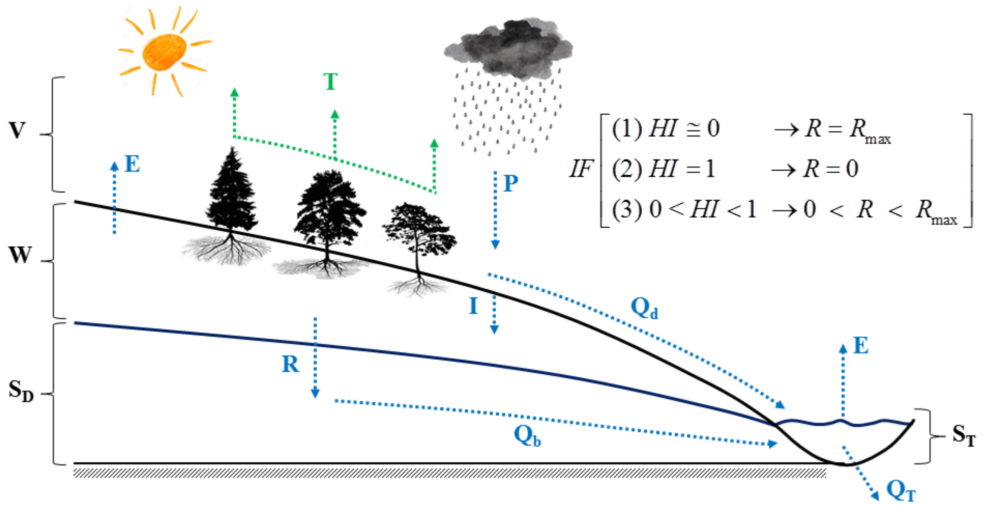


Figure 1: A conceptual-level representation of hydrological processes and annual water balance components that define the Horton index (HI). P is annual precipitation, E is annual evaporation, T is annual transpiration, I is annual infiltration, R is annual recharge, Q_d is annual quick or direct runoff, Q_b is annual baseflow, and Q_T is total annual streamflow. V is vaporization as the sum of E and T, W is wetting or annual infiltration, S_D is deep storage and S_T is total storage as derived from the catchment water balance. The three general assumptions underlying the relationship between HI and R are also displayed.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-449
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 17 September 2018

Hydrology and Earth System Sciences Discussions

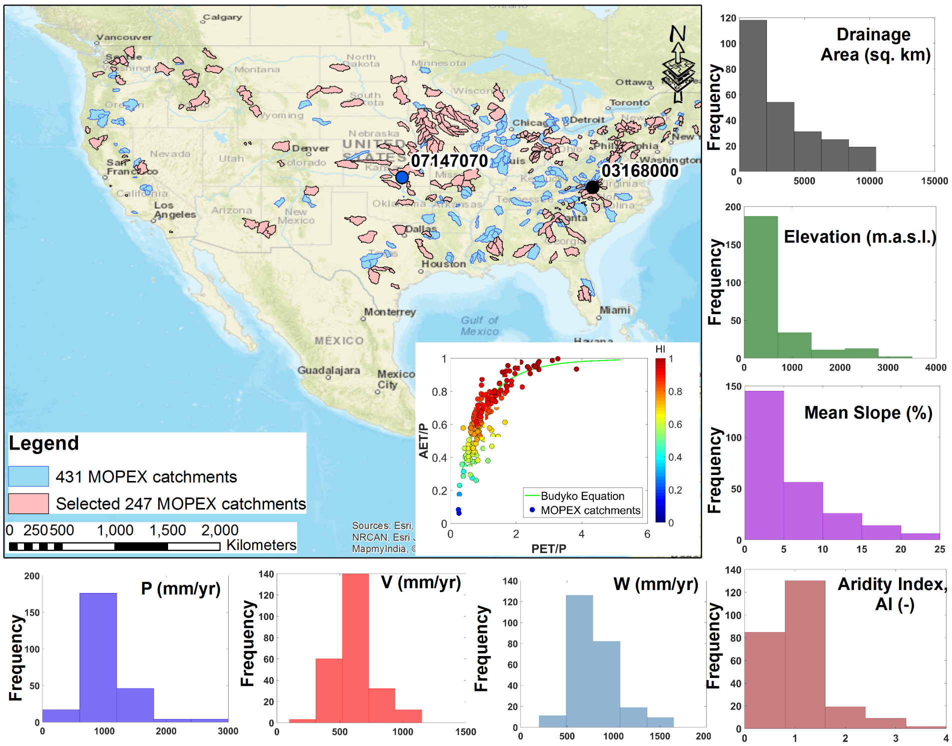


Figure 2: Study area and selected 247 catchments out of a total of 431 MOPEX catchments. The Budyko curve (color-coded using the catchments' Horton index) and the histograms of some catchment climate and landscape properties (i.e. precipitation (P), vaporization (V), wetting (W), aridity index, drainage area, slope, and elevation) are presented as insets. The two catchments for which the USGS stream gaging codes are provided in this plot correspond to the two catchments used in Figure 3.

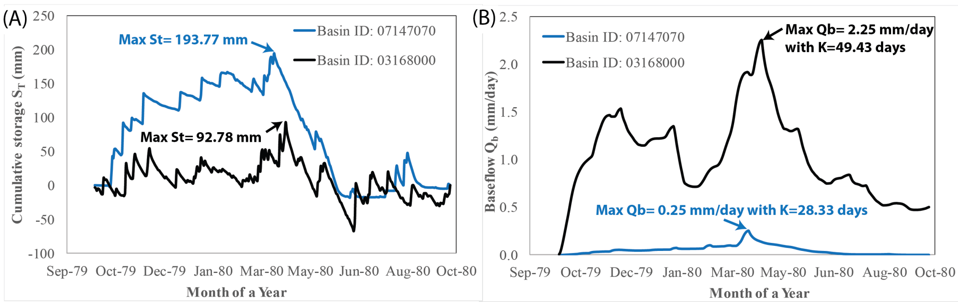


Figure 3: Examples of daily time series of total (S_T) and deep storage (S_D) for the catchments highlighted in Figure 2. (A) Time series of S_T; (B) Time series of S_D. The locations of the two catchments shown in this figure are given in Figure 2.

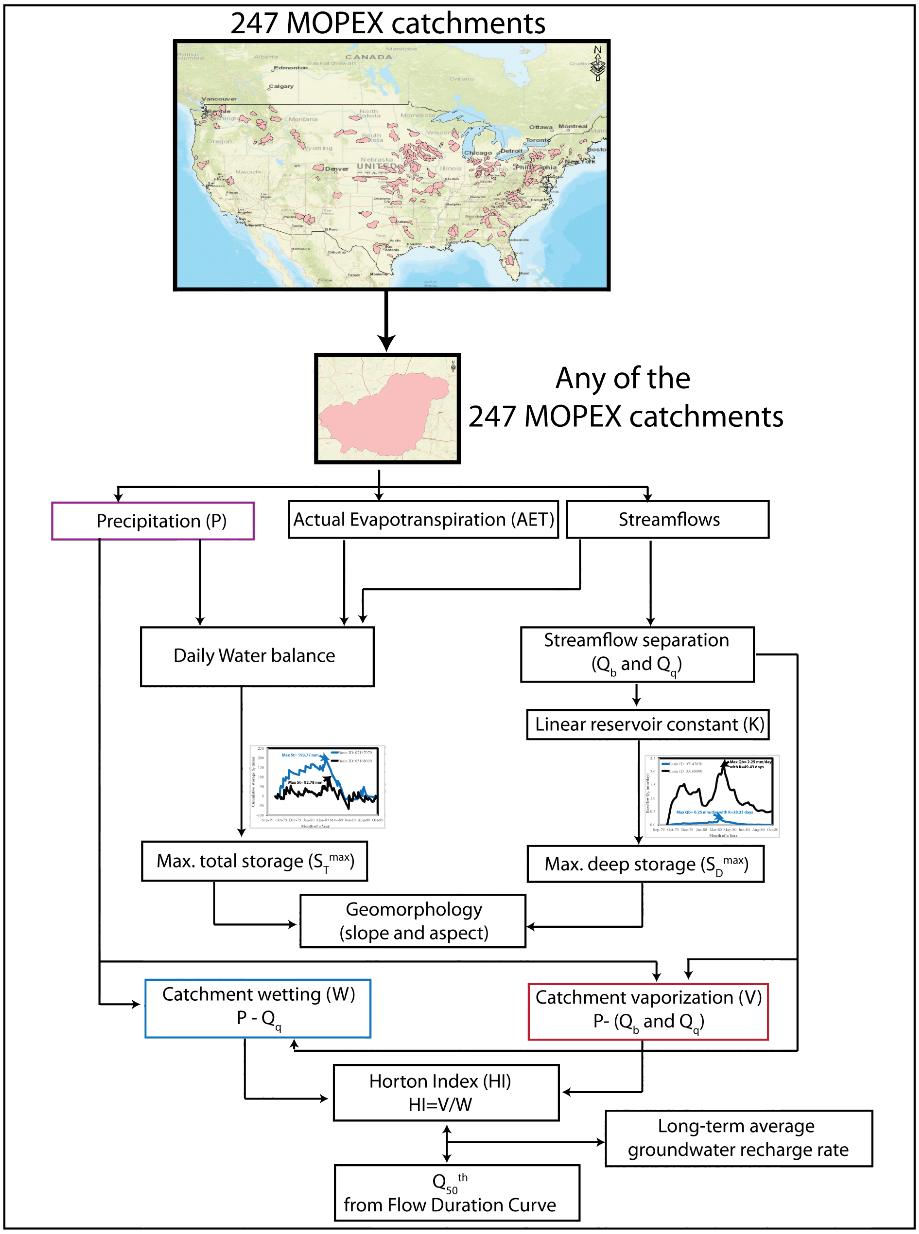


Figure 4: Research structure used to determine storage dynamics and potential predictors of baseflow conditions and groundwater recharge.

Hydrology and Earth System Sciences
Discussions

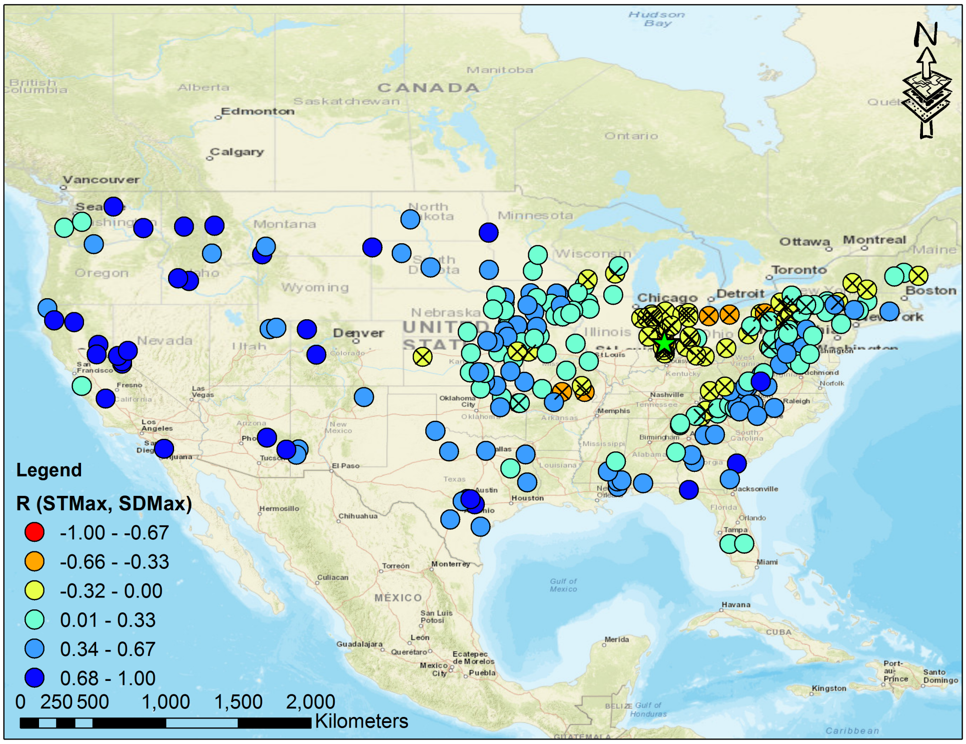


Figure 5: Correlation coefficients between annual maximum total (ST_{Max}) and annual maximum deep storage (SD_{Max}) for 247 MOPEX catchments in the conterminous US. Crossed markers show the catchments with a negative relationship (55 catchments) between the two types of storage, but the relationship is only significant for one catchment (shown by the green star).

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-449
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 17 September 2018

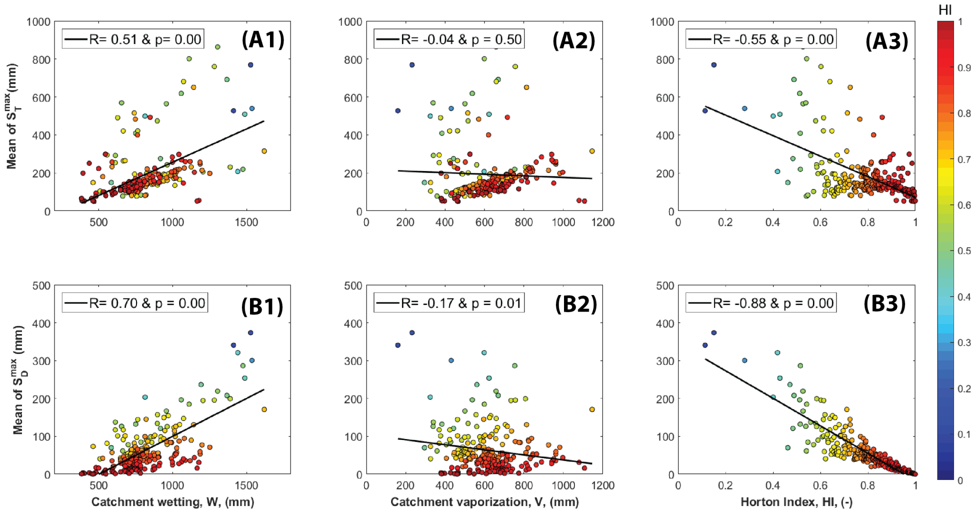
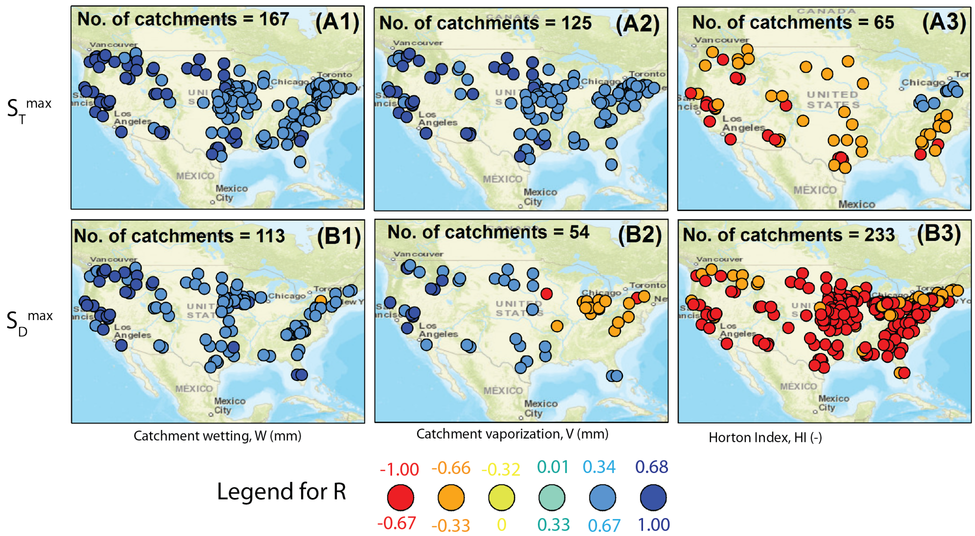


Figure 6: Linear relationships between average catchment wetting, vaporization, and the Horton index versus (A) average maximum total storage and (B) average maximum deep storage for the selected 246 MOPEX catchments.

697 

698 **Figure 7**: Maps of statistically significant linear correlation coefficients between
699 annual wetting, vaporization and the Horton index, versus (A) annual maximum
700 total and (B) annual maximum deep storage for the selected 246 MOPEX
701 catchments.

702

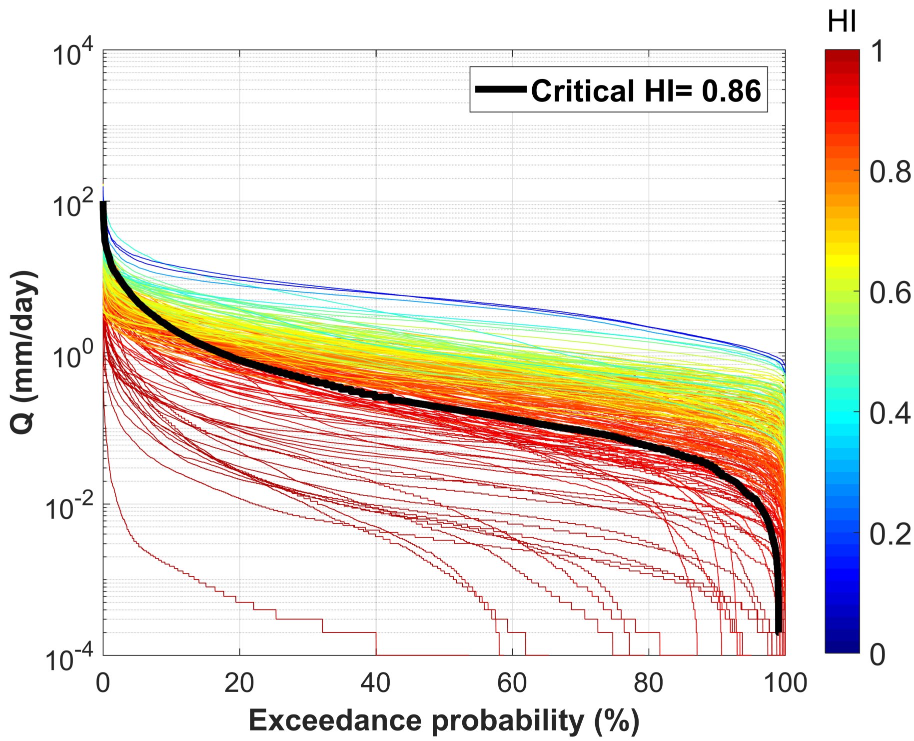


Figure 8: Flow duration curves for 247 MOPEX catchments under study. The color-scheme is related to the catchment average Horton Index. The black line separates perennial streams from ephemeral streams at the critical HI=0.86.

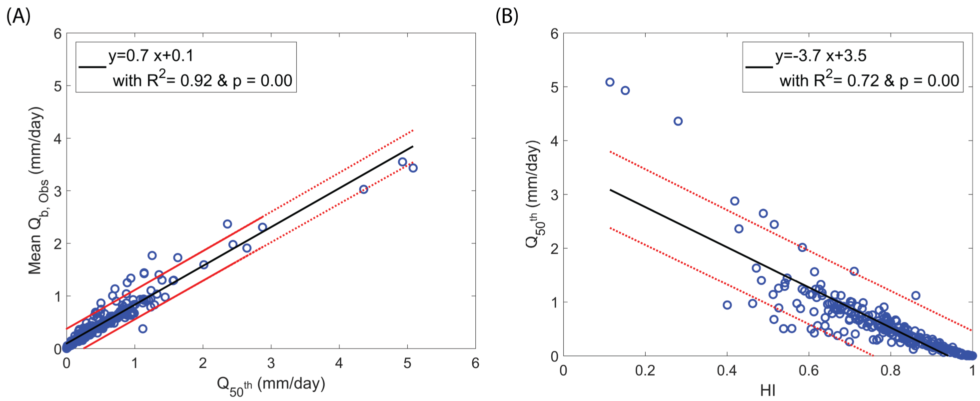


Figure 9: (A) Linear relationship between the Q_{50} flow of the FDC (a proxy of average recharge rates at the catchment scale) and the mean baseflow estimated from baseflow hydrographs. (B) Relationship between the Horton index and the Q_{50} flow of the FDC for all catchments under study.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-449
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 17 September 2018

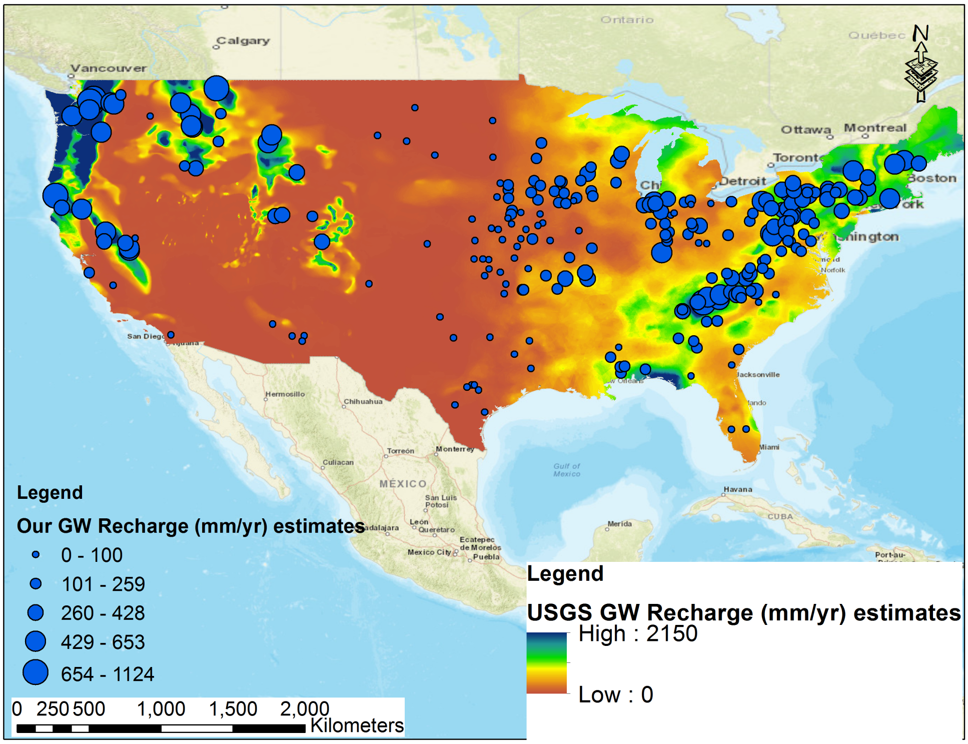
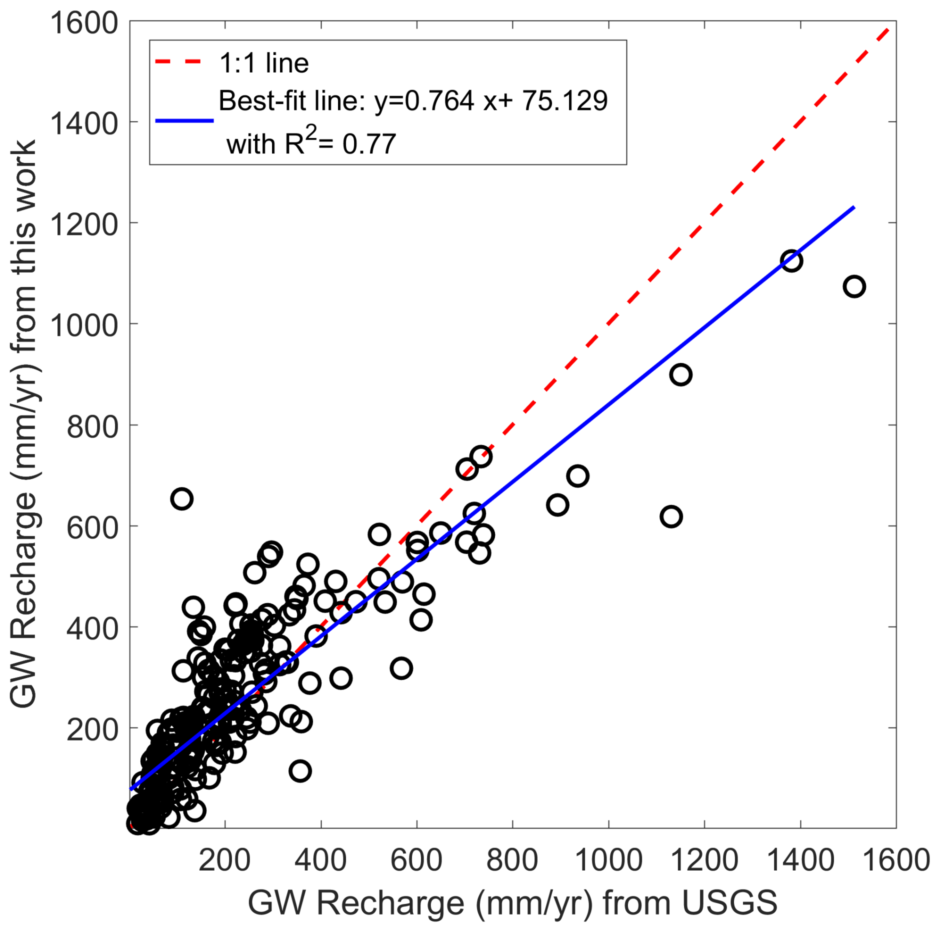


Figure 10: Horton Index-based estimates of average recharge rates (blue-markers) overlain onto a USGS map [*Wolock*, 2003] of average recharge rates for the conterminous US.



721

722 **Figure 11**: Inter-catchment comparison of HI based average recharge rates and

723 USGS-derived recharge rates (Wolock, 2003). The blue line is the regression line

724 through the data points and the red line is the 1:1 line.

725

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-449
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 17 September 2018

Hydrology and Earth System Sciences
Open Access
Discussions

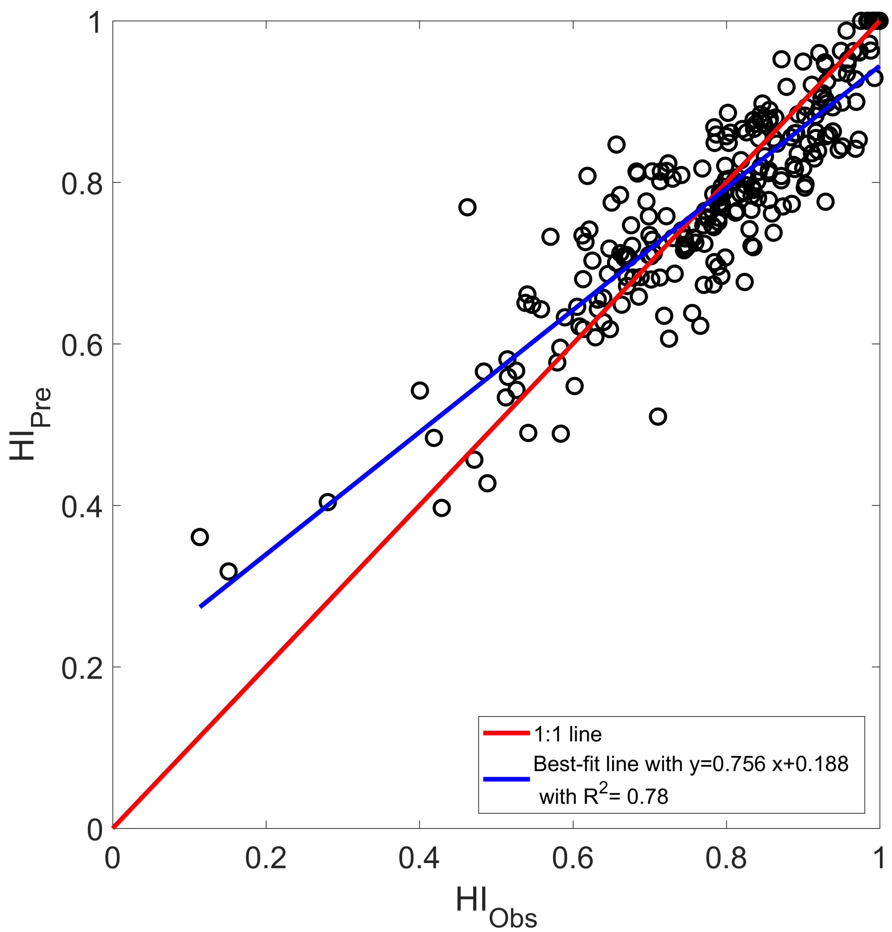


Figure 12: Comparison between predicted and observed HI using Equation (9). The blue line is the best linear fit through the data points and the red line is the 1:1 line.

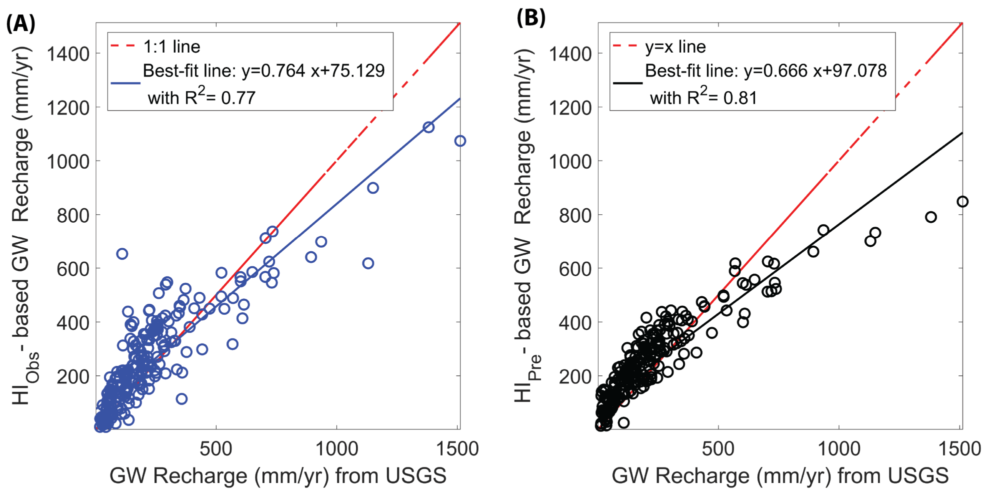


Figure 13: (A) Scatterplot between observed HI based groundwater recharge (GWR) vs. USGS groundwater recharge with regression line (blue) and 1:1 line (red); (B) scatterplot between predicted HI based GWR and USGS GWR with regression line (black) and 1:1 line (red).

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-449
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 17 September 2018

Hydrology and Earth System Sciences
Discussions

Tables

Table 1: Notation for Total and Deep Storage used in this Study

Description	**Total storage**	**Deep storage**
Storage	S_T	S_D
Storage at a given day t, $t = 1, 2, \ldots, 365$	$S_T(t)$	$S_D(t)$
Maximum storage for a given year	$S_T^{\max}$	$S_D^{\max}$
Average maximum storage; N is number of years	$\bar{S}_T^{\max} = \frac{1}{N}\sum_{i=1}^{i=N} S_T^{\max}(i)$	$\bar{S}_D^{\max} = \frac{1}{N}\sum_{i=1}^{i=N} S_D^{\max}(i)$