Comparing spatial and temporal transferability of hydrological model parameters

Patil, S.D.; Stieglitz, M.

Journal of Hydrology

DOI:
10.1016/j.jhydrol.2015.04.003

Published: 09/04/2015

Peer reviewed version

Comparing spatial and temporal transferability of hydrological model parameters

Sopan D. Patil[1], Marc Stieglitz[2]

[1] School of Environment, Natural Resources and Geography,

Bangor University,

Deiniol Road, Bangor, LL57 2UW, United Kingdom

[2] School of Civil and Environmental Engineering,

Georgia Institute of Technology,

790 Atlantic Drive, Atlanta, GA 30332, United States of America

Submission to: Journal of Hydrology

Correspondence to: Sopan D. Patil (email: s.d.patil@bangor.ac.uk, Tel: +44 1248 388294)

Highlights:

1) We compare three different schemes for transfer of hydrological model parameters
2) Temporal transfer scheme outperforms spatial and spatiotemporal transfer schemes
3) Differences between spatial and spatiotemporal transfer schemes are negligible
4) Temporal gap in calibration and validation periods reduces difference among schemes

Abstract

Operational use of hydrological models requires the transfer of calibrated parameters either in time (for streamflow forecasting) or space (for prediction at ungauged catchments) or both. Although the effects of spatial and temporal parameter transfer on catchment streamflow predictions have been well studied individually, a direct comparison of these approaches is much less documented. Here, we compare three different schemes of parameter transfer, viz., temporal, spatial, and spatiotemporal, using a spatially lumped hydrological model called EXP-HYDRO at 294 catchments across the continental United States. Results show that the temporal parameter transfer scheme performs best, with lowest decline in prediction performance (median decline of 4.2%) as measured using the Kling-Gupta efficiency metric. More interestingly, negligible difference in prediction performance is observed between the spatial and spatiotemporal parameter transfer schemes (median decline of 12.4% and 13.9% respectively). We further demonstrate that the superiority of temporal parameter transfer scheme is preserved even when: (1) spatial distance between donor and receiver catchments is reduced, or (2) temporal lag between calibration and validation periods is increased. Nonetheless, increase in the temporal lag between calibration and validation periods reduces the overall performance gap between the three parameter transfer schemes. Results suggest that spatiotemporal transfer of hydrological model parameters has the potential to be a viable option for climate change related hydrological studies, as envisioned in the "trading space for time" framework. However, further research is still needed to explore the relationship between spatial and temporal aspects of catchment hydrological variability.

Keywords: Hydrological model; parameter transfer; catchment; streamflow prediction

1 Introduction

All hydrological models contain parameters whose values must be calibrated by comparing the observed and simulated streamflow values from the past record [*Refsgaard*, 1997; *Beven*, 2001]. Calibrated parameters represent the unique combination of climatic and physiographic factors that influence the hydrological behaviour of a catchment [*Merz and Blöschl*, 2004; *Wagener and Wheater*, 2006]. However, operational use of hydrological models is always outside of the calibration period and/or catchment, which is where the parameters face their true test [*Klemeš*, 1986; *Refsgaard and Knudsen*, 1996; *Coron et al.*, 2012]. Parameter transfer away from this calibration domain can be in time (for streamflow forecasting) or space (for prediction at ungauged catchments) or both.

Temporal transfer of calibrated parameters is perhaps the most common and straightforward procedure used in catchment hydrological modelling. The first step involves choosing a specific historical time period for which all the input and output data required for running the model are available for the catchment. These data are used to calibrate the model parameters by finding the best match between the simulated and observed streamflow values. This procedure is followed by the application of the calibrated model at some other time period in the same catchment. *Klemeš* [1986] recommends that testing of hydrological models outside the calibration period is critical to establish their credibility as useful forecasting tools. An implicit assumption here is that the calibrated model parameters are temporally stable, i.e., they are suitable for application beyond the calibration period. However, numerous recent studies have shown that hydrological model parameters are not always temporally stable [*Merz et al.*, 2011; *Brigode et al.*, 2013; *Westra et al.*, 2014], and their values depend on the duration as well as the specific physioclimatic conditions of the calibration period [*Xia et al.*, 2004; *Juston et al.*, 2009; *Vaze et al.*, 2010; *Razavi and Tolson*, 2013]. *Wagener et al.* [2003] used dynamic identifiability analysis (DYNIA) to estimate the

parameters of a spatially lumped hydrological model and found that parameter values varied significantly when calibrated to different parts of the hydrograph. *Merz et al.* [2011] calibrated the parameters of a semi-distributed version of HBV model for six consecutive 5 year periods between 1976 and 2006 at 273 Austrian catchments, and found that (1) optimal parameter values were variable across the six calibration periods, and (2) the assumption of time invariant parameters had a significant impact on model simulations outside the calibration period. Similar findings were reported by *Coron et al.* [2012] in their study on temporal parameter transfer using three rainfall-runoff models at 216 catchments in southeast Australia. *Razavi and Tolson* [2013] compared three different calibration approaches for the SWAT2000 model at a catchment in the state of New York, USA and concluded that "…model calibration solely to a short data period may lead to a range of performances from poor to very well depending on the representativeness of the short data period which is typically not known a priori".

Spatial transfer of calibrated parameters is another widely used procedure in catchment hydrological modelling and is primarily required for streamflow prediction at ungauged basins (PUB) [*Sivapalan et al.*, 2003]. A considerable amount of research has been conducted over the years in the development and comparison of approaches to transfer hydrological model parameters from gauged to ungauged catchments [*Post and Jakeman*, 1999; *Kokkonen et al.*, 2003; *McIntyre et al.*, 2005; *Young*, 2006; *Oudin et al.*, 2008; *Zhang and Chiew*, 2009; *Patil and Stieglitz*, 2014]. *Blöschl et al.* [2013] and *Hrachowitz et al.* [2013] provide a comprehensive summary and synthesis of the progress made in PUB research during the International Association of Hydrological Sciences' (IAHS) PUB decade initiative (2003-2012) [*Sivapalan et al.*, 2003]. Donor gauged catchments, from which model parameters can be transferred to the receiver ungauged catchments, are typically identified using an approach that is either based on spatial proximity or physical similarity to the

ungauged catchments. *Oudin et al.* [2008] compared the spatial proximity and physical similarity approaches at 913 catchments in France and found that the spatial proximity approach outperformed the physical similarity approach. *Zhang and Chiew* [2009] tested multiple parameter transfer approaches at 210 catchments in southeast Australia and found that an integrated similarity approach that combined spatial proximity and physical similarity slightly outperformed the spatial proximity approach. *Patil and Stieglitz* [2014] compared two different methods of spatial parameter transfer at 323 catchments in the United States and found that simulation performance at ungauged catchments is more sensitive to the types of parameters that are transferred than to the method used for transferring them. However, regardless of the chosen approach, spatial parameter transfer tends to cause deterioration in simulation performance (compared to calibration) due to the differences in physiographic properties and meteorological inputs between the donor and receiver catchments.

Although hydrological model simulation following temporal and/or spatial parameter transfer is expected to cause deterioration in catchment streamflow prediction, not many studies have focused on a direct comparison of these two approaches. A few PUB focused studies that have made such a comparison show results that range from a large performance difference between temporal and spatial parameter transfer (in favour of temporal) [*Merz and Blöschl*, 2004; *Parajka et al.*, 2005] to minor performance difference between them [*Oudin et al.*, 2008]. In our view, further exploration is therefore needed on how the spatial and temporal parameter transfer approaches compare against each other, especially in the context of increasing appeal and popularity of the "trading space for time" approaches that are proposed for assessing the hydrological implications of anthropogenic climate change [*Wagener et al.*, 2010; *Peel and Blöschl*, 2011; *Singh et al.*, 2011; *Ehret et al.*, 2014; *Refsgaard et al.*, 2014]. The trading space for time framework assumes that the spatial variability in catchment hydrological properties (including model parameters) can be used as

99 a proxy for the climate change induced temporal variability in those properties [*Merz et al.*,
100 2011]. Studies such as *Singh et al.* [2011, 2014] have already demonstrated that the spatial
101 parameter regionalization techniques developed for PUB can also be applied to make
102 temporal modifications in model parameters for streamflow predictions under change (PUC)
103 [*Montanari et al.*, 2013]. Therefore, we argue that a systematic comparison of the spatial and
104 temporal parameter transfer approaches is likely to provide further insights into the
105 connections between the PUB and PUC paradigms, and could even help refine the trading
106 space for time methods.
107 In this paper, we compare three schemes of model parameter transfer, viz., temporal,
108 spatial, and spatiotemporal, using a hydrological model called EXP-HYDRO [*Patil and*
109 *Stieglitz*, 2014; *Patil et al.*, 2014a, 2014b] at 294 catchments across the continental United
110 States. The temporal parameter transfer scheme is implemented using a split-sample test
111 procedure where the available data is divided into two periods, one for calibration and the
112 other for validation. For the spatial parameter transfer scheme, we use the nearest neighbour
113 catchment as a donor of calibrated parameters. Comparison of different spatial parameter
114 transfer techniques is beyond the scope of this study (and has already been done by *Patil and*
115 *Stieglitz* [2014]). In the spatiotemporal parameter transfer scheme, calibrated model
116 parameters are transferred simultaneously in the spatial (to the nearest neighbour catchment)
117 and temporal (to a different time period) domain.
118

119 ## 2 Data and Methods

120 ### *2.1 Hydrological Model*

121 We use the spatially lumped version of EXP-HYDRO model [*Patil and Stieglitz*,
122 2014; *Patil et al.*, 2014a, 2014b] to simulate daily streamflow (Figure 1). This model solves
123 the following two coupled ordinary differential equations simultaneously at each time step:

$$\frac{dS_{Snow}}{dt} = P_{Snow} - Q_{Melt} \quad (1a)$$

$$\frac{dS}{dt} = P_{Rain} + Q_{Melt} - ET - Q_{Bucket} - Q_{Spill} \quad (1b)$$

where S and S_{Snow} are, respectively, the amounts of stored water (mm) in the catchment and snow accumulation buckets. P_{Snow} and P_{Rain} are the snowfall and rainfall amounts (mm/day). ET is the actual evapotranspiration (mm/day) from the catchment bucket. Q_{Melt} is the snowmelt (mm/day) from the snow accumulation bucket, Q_{Bucket} is the subsurface runoff (mm/day) generated from the catchment bucket, and Q_{Spill} is the capacity-excess surface runoff (mm/day) that is generated when the catchment bucket is filled to capacity.

The incoming daily precipitation P is classified as snowfall or rainfall based on the following conditions:

If $T_a < T_{\min}$,

$$\begin{aligned} P_{Snow} &= P \\ P_{Rain} &= 0 \end{aligned} \quad (2a)$$

Else,

$$\begin{aligned} P_{Snow} &= 0 \\ P_{Rain} &= P \end{aligned} \quad (2b)$$

where T_a is the actual air temperature on a given day, $T_{\min}$ is the air temperature below which the precipitation occurs as snowfall and falls directly into the snow accumulation bucket.

Snowmelt Q_{Melt} is modelled using a simple thermal degree day model as follows:

If $T_a > T_{\max}$,

$$Q_{Melt} = \min\left\{S_{Snow}, D_f \cdot (T_a - T_{\max})\right\} \quad (3a)$$

Else,

$$Q_{Melt} = 0 \quad (3b)$$

where $T_{\max}$ is the air temperature above which the snow in snow accumulation bucket starts melting, and D_f is the thermal degree day factor that controls the rate of snowmelt. Evapotranspiration ET from the catchment bucket is calculated as follows:

$$ET = PET \cdot \left(\frac{S}{S_{\max}} \right) \quad (4)$$

where PET is the potential evapotranspiration (mm/day), and is calculated from the daily air temperature using Hamon's formula [*Hamon*, 1963]. $S_{\max}$ is the total storage capacity (mm) of the catchment bucket. The surface and subsurface runoff generated from the catchment bucket are calculated as follows:

If $S \le S_{\max}$,

$$\begin{aligned} Q_{Bucket} &= Q_{\max} \cdot \exp\left(-f \cdot \left(S_{\max} - S\right)\right) \\ Q_{Spill} &= 0 \end{aligned} \quad (5a)$$

If $S > S_{\max}$,

$$\begin{aligned} Q_{Bucket} &= Q_{\max} \\ Q_{Spill} &= S - S_{\max} \end{aligned} \quad (5b)$$

where $Q_{\max}$ is the maximum subsurface runoff produced (mm/day) when the catchment bucket reaches its capacity, and f is the parameter controlling the storage-dependent exponential decline in subsurface runoff (1/mm). Daily streamflow at the catchment outlet is the sum of Q_{Bucket} and Q_{Spill}.

We have now made the entire source code of the spatially lumped version of the EXP-HYDRO model (described above) freely available to the research community. This source code is written in Python ® programming language and can be downloaded from the following web link: http://sopanpatil.weebly.com/exp-hydro.html.

2.2 Model Calibration

There are six calibration parameters in the EXP-HYDRO model: f, $S_{\max}$, $Q_{\max}$, D_f, $T_{\max}$, and $T_{\min}$. For each catchment, we calibrate these six model parameters using the Particle Swarm Optimisation (PSO) algorithm [*Kennedy and Eberhart*, 1995]. PSO is a stochastic population-based search algorithm that has been used in numerous hydrological studies for model parameter calibration [*Gill et al.*, 2006; *Goswami and O'Connor*, 2007; *Liu*, 2009; *Zhang et al.*, 2009]. PSO is initialised with a group of random particles (parameter sets in our case), and this 'swarm' of particles searches for an optimal solution within the parameter domain by iteratively updating the velocity and position of each particle. We initialise the PSO algorithm with 10 randomly generated EXP-HYDRO parameter sets (sampled from a uniform distribution) and allow for a maximum of 50 swarm iterations to find the optimal solution. The upper and lower bound values of all six parameters are same as those in *Patil and Stieglitz* [2014]. We use Kling-Gupta efficiency (KGE) [*Gupta et al.*, 2009] as the objective function that is to be maximised during calibration:

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \quad (6)$$

where r is the Pearson's linear correlation coefficient between the observed and simulated streamflow, α is the ratio of standard deviations of observed and simulated streamflow, and β is the ratio of mean values of observed and simulated streamflow. The value of KGE varies from $-\infty$ to 1, with KGE = 1 being a perfect fit between the observed and simulated values.

As shown by *Gupta et al.* [2009], KGE consists of three main components, correlation (g_1), variability (g_2), and bias (g_3), whose relative contribution to the KGE value varies from 0 to 1 and is calculated as follows:

188 $$g_i = \frac{G_i}{\sum_{j=1}^{3} G_j} \quad (7a)$$

189 and,

190 $$\begin{aligned} G_1 &= (r-1)^2 \\ G_2 &= (\alpha-1)^2 \\ G_3 &= (\beta-1)^2 \end{aligned} \quad (7b)$$

191 ***2.3 Study Catchments***

192 We begin the catchment selection process by implementing the EXP-HYDRO model
193 at 756 catchments in the continental United States (Figure 2a). These are the same
194 catchments that have been used in our two previous studies [*Patil and Stieglitz*, 2012, 2014];
195 they belong to the U. S. Geological Survey's Hydro-Climate Data Network (HCDN) [*Slack et*
196 *al.*, 1993] and have a continuous daily streamflow record from water year (WY) 1970 to 1988
197 (i.e., 1st October, 1969 to 30th September, 1988). Daily precipitation and air temperature
198 data for each catchment are obtained from the gridded meteorological dataset developed by
199 *Maurer et al.* [2002]. This data has a spatial resolution of 0.125 degree and covers the entire
200 continental United States.

201 We split the timeline from WY 1970 to 1988 into the following three periods: WY
202 1970 is the spin-up period that is not used for parameter calibration, WY 1971 to 1978 is
203 calibration period 1, and WY 1979 to 1988 is calibration period 2 (see Figure 2b). For each
204 catchment, we calibrate the EXP-HYDRO model parameters separately for calibration
205 periods 1 and 2 using the methods described in Section 2.2. Only those catchments where the
206 simulated streamflow provides KGE > 0.6 for both calibration periods are retained as our
207 study catchments. This is done to ensure that only those catchments for which the EXP-
208 HYDRO model structure seems suitable for providing good hydrological simulations are
209 used for any further analyses. The above condition reduces the number of acceptable study

catchments to 294 (Figure 2a). The geographic distribution of the acceptable catchments is similar to that obtained by *Patil and Stieglitz* [2014]. The drainage area of the study catchments ranges from 24 km^2 to 4790 km^2, with median drainage area of 620 km^2. The majority of catchments are located in the eastern part of the continental United States, to the east of Mississippi River. In the western United States, the study catchments are primarily located along the Rocky, Cascade and Pacific Coastal mountain ranges. Mean annual precipitation among the study catchments ranges from 340 mm to 2556 mm (median = 1160 mm).

2.4 Parameter Transfer Schemes

For the 294 study catchments, we test the following three schemes of model parameter transfer:

(1) *Temporal transfer:* For the same catchment, model parameters from calibration period 1 are transferred to calibration period 2, and vice versa.

(2) *Spatial transfer:* Model parameters of a catchment are obtained from its nearest neighbour catchment over the same time period. This is done separately for calibration periods 1 and 2.

(3) *Spatiotemporal transfer:* Model parameters of a catchment are obtained from its nearest neighbour catchment across different time periods (i.e., from calibration period 1 to 2, and vice versa).

3 Results

We first compare the simulation performance of EXP-HYDRO model that is obtained across the two calibration periods. Figure 3 shows a 1:1 comparison of the KGE values obtained at the 294 catchments during calibration periods 1 and 2. The relationship between calibrated KGE values of the two periods is somewhat weak (Pearson's r = 0.53), with data

235 points scattered along both sides of the 1:1 line. Figure 4 shows the comparison of g_1, g_2,
236 and g_3 (the three components of KGE) during calibration periods 1 and 2. Similar to KGE, a
237 weak relationship exists between the values obtained at these two different periods (r = 0.55
238 for g_1, r = 0.4 for g_2, r = 0.65 for g_3). Figures 5a-f show a 1:1 comparison of the values of
239 all six EXP-HYDRO parameters during calibration periods 1 and 2. The relationship among
240 parameter values across the two calibration periods is strongest for f (r = 0.81, Figure 5a),
241 followed by $S_{\max}$ (r = 0.59, Figure 5b), D_f (r = 0.36, Figure 5d), $Q_{\max}$ (r = 0.26, Figure 5c),
242 $T_{\min}$ (r = 0.26, Figure 5e), and $T_{\max}$ (r = 0.17, Figure 5f).

243 We next compare the performance of EXP-HYDRO model across the three parameter
244 transfer schemes. Figure 6 shows a box-plot comparison of the KGE values among the
245 following four modelling scenarios: calibration, temporal transfer, spatial transfer, and
246 spatiotemporal transfer. Note that the KGE values shown in Figure 6 are the average values
247 of two sub-scenarios that are present in each scenario. For example, KGE of each catchment
248 in the calibration scenario is an average of its KGE values from calibration period 1 and
249 calibration period 2. As seen in Figure 6, the overall model performance is highest for the
250 calibration scenario (median KGE = 0.72), followed by temporal transfer (median KGE =
251 0.69; decline of 4.2% (compared to calibration)), spatial transfer (median KGE = 0.63;
252 decline of 12.4%), and spatiotemporal transfer (median KGE = 0.62; decline of 13.9%)
253 scenarios.

254 Figure 7 shows the box-plot comparison of the above four scenarios with respect to
255 the three KGE components, g_1, g_2, and g_3. For correlation component g_1, calibration
256 scenario has the highest overall contribution value (median g_1 = 0.85), and is followed by
257 temporal (median g_1 = 0.68), spatial (median g_1 = 0.54), and spatiotemporal (median g_1 =

258 0.54) transfer scenarios. For variability component g_2, calibration scenario has the lowest
259 contribution value (median g_2 = 0.06), and is followed by the temporal (median g_2 = 0.17),
260 spatial (median g_2 = 0.23), and spatiotemporal (median g_2 = 0.24) transfer schemes. The
261 bias component g_3 has a similar trend as g_2, but is less prominent. The calibration scenario
262 has the lowest contribution value (median g_3 = 0.05), and is followed by the temporal
263 (median g_3 = 0.11), spatial (median g_3 = 0.12), and spatiotemporal (median g_3 = 0.13)
264 transfer schemes.

265 Results from Figures 6 and 7 demonstrate that the overall model performance of the
266 temporal parameter transfer scheme is superior to that of the spatial and spatiotemporal
267 parameter transfer schemes. However, it is not clear from these results whether our
268 experimental setup provides any undue advantage to the temporal parameter transfer scheme
269 over the other two schemes. Below, we briefly mention two such potential scenarios:
270 *Scenario 1*: It is likely that for some of our study catchments, the distance between them and
271 their nearest neighbour catchment is too high. Such a scenario puts the spatial and
272 spatiotemporal transfer schemes at a clear disadvantage compared to the temporal transfer
273 scheme.
274 *Scenario 2*: There is no temporal lag between calibration periods 1 and 2, i.e., calibration
275 period 2 immediately follows calibration period 1 (Figure 2b). For catchments where the
276 meteorological input patterns have not changed much across the two periods, the temporal
277 parameter transfer scheme is much more likely to outperform the other two schemes.
278 To mitigate the impacts from above two potential scenarios on our results, we repeat the
279 parameter transfer experiment for the two following special conditions:
280 *Special Condition 1*: Eliminate all study catchments that have a nearest neighbour catchment
281 more than 50 km away. The 50 km distance limit is slightly less than the median nearest

neighbour distance (53.1 km) among all the 294 study catchments. This reduces the number of catchments from 294 to 138.

Special Condition 2: Shorten the calibration period 1 to span from WY 1971 to 1975 (instead of WY 1971 to 1978) and calibration period 2 to span from WY 1984 to 1988 (instead of WY 1979 to 1988). This creates a temporal lag of 8 years between the two calibration periods. Nonetheless, unlike Special Condition 1, all 294 study catchments are used for simulations. Note that these new calibration periods span 5 years each, which is the minimum time span recommended by some studies to adequately capture the temporal hydrological variability of a catchment [*Merz et al.*, 2009].

Figures 8a and 8b show the box-plot comparison of KGE values from the four modelling scenarios for Special Conditions 1 and 2 respectively. In both cases, the results are similar to those observed in Figure 6. For Special Condition 1, the highest model performance is obtained for the calibration scenario (median KGE = 0.74), followed by the temporal (median KGE = 0.71; decline of 4%), spatial (median KGE = 0.67; decline of 9.5%), and spatiotemporal (median KGE = 0.66; decline of 10.8%) parameter transfer schemes. For Special Condition 2, the median KGE values are 0.72 for calibration, 0.66 for temporal transfer scheme (decline of 8.3%), 0.62 for spatial transfer scheme (decline of 13.9%), and 0.62 for spatiotemporal transfer scheme (decline of 13.9%).

4 Discussion

Comparison of the optimal KGE values (Figure 3) and the individual KGE components (Figure 4) between calibration periods 1 and 2 demonstrates that the performance of a hydrological model can, at least in some cases, vary considerably in the same catchment for different calibration periods. Similar findings have been reported by *Vaze et al.* [2010] and *Razavi and Tolson* [2013]. The data points shown in Figures 3 and 4

307 are scattered along both sides of the 1:1 line. This suggests that for our study catchments no
308 systematic bias exists in terms of one calibration period providing better calibration
309 performance than the other. For the majority of study catchments (213 out of 294), the
310 difference between optimal KGE values for the two calibration periods is less than 10%
311 (median = 6%). This is consistent with *Merz et al.* [2011], who found that the calibrated
312 Nash-Sutcliffe efficiency values [*Nash and Sutcliffe,* 1970] of the HBV model, averaged
313 across their 273 Austrian catchments, showed small variability across six different calibration
314 periods. However, they found this variability to increase when only drier study catchments
315 were considered in their calculations. Although our data set does have a few catchments that
316 show large performance difference between the calibration periods (largest difference is
317 35%), we did not detect any specific geographic or climatic pattern among those catchments.
318 Results from Figure 5 show that the temporal variability of parameter values is
319 different for each parameter, with values of the f parameter showing the highest correlation
320 (and therefore lowest variability) between the two calibration periods, followed by $S_{\max}$, D_f,
321 $Q_{\max}$, $T_{\min}$, and $T_{\max}$. Interestingly, this trend is similar to the parameter sensitivity trend
322 shown in our previous study [*Patil and Stieglitz,* 2014]. *Patil and Stieglitz* [2014] performed
323 a sensitivity analysis of all 6 EXP-HYDRO parameters and found that f, $S_{\max}$, and D_f
324 were the most sensitive parameters (i.e., sensitive to the objective function) with better
325 defined optimal values and posterior distributions. On the other hand, $Q_{\max}$, $T_{\min}$, and $T_{\max}$
326 were characterised as the insensitive parameters with virtually no difference between their
327 prior (uniform) and posterior distributions. Combined, these results suggest that the high
328 sensitivity model parameters are also more likely to have low variability across different
329 calibration periods, thereby making them more representative of the catchment's intrinsic
330 physiographic conditions rather than the specific input conditions during the calibration

period. For the EXP-HYDRO model, the two most temporally stable parameters (f and $S_{\max}$) represent: (1) the rate of storage decline within the soil bucket in response to subsurface runoff, and (2) the total soil bucket capacity. Both these parameters are, at least in theory, linked to the intrinsic soil and topographic properties of a catchment that are unlikely to undergo drastic temporal change. Nonetheless, we suspect that the proportion of sensitive and temporally stable parameters is likely to be different for different types of hydrological models. For instance, *Merz and Blöschl* [2004] calibrated the lumped version of HBV model (containing 11 parameters) at 308 catchments in Austria for two different calibration periods. They found that the correlation coefficient (R^2) between the parameter values of the two calibration periods ranged from 0.09 to 0.64, with only 5 of the 11 parameters having $R^2 >$ 0.5. *Oudin et al.* [2008] compared the parameters of two spatially lumped hydrological models, GR4J (4 parameters) and TOPMO (6 parameters), for 913 French catchments across two calibration periods, and showed that the correlation across calibration periods was higher for GR4J parameters compared to TOPMO parameters. However, it must be noted we still do not have a complete understanding of how these model parameter values will change in response to land cover changes within a catchment [*Eckhardt et al.*, 2003; *Croke et al.*, 2004; *Wang and Kalin*, 2011].

Results from the parameter transfer experiment (Figures 6 and 7) demonstrate the overall superior performance of the temporal parameter transfer scheme over the spatial and spatiotemporal parameter transfer schemes. Analysis of the three KGE components (Figure 7) shows that the correlation component (g_1) is the most dominant contributor to KGE value, which is consistent with *Gupta et al.* [2009]. However, this component also undergoes the most decline when moving away from calibration to the parameter transfer scenarios. The decline in g_1 is compensated by a proportional increase in the contribution from the other two

components (variability g_2 and bias g_3) for the parameter transfer scenarios compared to calibration. *Gupta et al.* [2009] note that the relative contributions of the bias and variability components tend to increase for non-optimal parameter sets, as is observed in all our parameter transfer scenarios. When the three parameter transfer schemes are individually compared across each of the 294 catchments, the temporal transfer scheme shows best performance at 204 catchments (and worst at 50 catchments), the spatial transfer scheme is best at 65 catchments (and worst at 106 catchments), whereas the spatiotemporal transfer scheme is best at 27 catchments (and worst at 147 catchments). Figure 9 shows the map of catchment locations where either the spatial or spatiotemporal parameter transfer scheme is the best performing scheme. No specific geographic pattern is noticeable from this figure in terms of the catchments that prefer the spatial and spatiotemporal transfer schemes over the temporal transfer scheme. Table 1 shows the comparison of these two catchment groups (Group 1: spatial or spatiotemporal transfer scheme performing best; Group 2: temporal transfer scheme performing best) with respect to three commonly used hydro-climatic metrics, viz., mean annual rainfall (P), annual runoff ratio (Q/P) and climate aridity index (PET/P). Although the median values of these metrics suggest that Group 1 catchments are slightly drier (lower P and higher PET/P) and flashier (higher Q/P), there does not seem to be much difference among the two catchment groups. Nonetheless, Figure 9 clearly illustrates that even in regions with low catchment density and larger distances among neighbouring catchments (e.g., in the western US), the temporal parameter transfer schemes does not always outperform the spatial and spatiotemporal schemes.

Parameter transfer experiments under the two special conditions (see Section 3) show that the temporal parameter transfer scheme still preserves its advantage over the spatial and spatiotemporal schemes. For Special Condition 1, i.e., when only those catchments with nearest neighbour < 50 km away are retained, the spatial and spatiotemporal schemes exhibit

performance improvement as the KGE difference between them and calibration is about 3% lower than the base scenario. This is an expected result because lower spatial distances between catchments will most likely reduce the spatial variability of hydrological behaviour [*Oudin et al.*, 2008]. On the other hand, for Special Condition 2 (when the temporal gap between calibration periods is increased to 8 years), the KGE difference between calibration and the temporal scheme is about 4% higher than the base scenario, and is virtually unchanged between calibration and the spatial and spatiotemporal schemes. Thus, an increase in the temporal distance between calibration and validation periods reduces the model performance gap between the temporal scheme and the spatial and spatiotemporal schemes. This has important implications for the "trading space for time" framework [*Peel and Blöschl*, 2011; *Refsgaard et al.*, 2014] because a sufficiently large time lag between the calibration and validation periods (as is common in climate change scenarios) has the potential to make spatiotemporal parameter transfer a more viable option than temporal parameter transfer. However, even though the introduction of an 8 year temporal gap between the calibration and validation periods shows reduced performance gap among the parameter transfer schemes, it is not entirely clear how these schemes will compare for much larger (> 40-50 years) temporal gaps.

5 Conclusion

In this paper, we compared three different schemes for the transfer of hydrological model parameters, viz., temporal, spatial, and spatiotemporal, using a spatially lumped hydrological model called EXP-HYDRO at 294 catchments across the continental United States. In our view, such a comparison is highly relevant especially within the context of increasing appeal and popularity of the "trading space for time" framework proposed for assessing the hydrological implications of anthropogenic climate change. Results showed

405 that the temporal parameter transfer scheme performs best, with lowest decline in prediction
406 performance compared to calibration (median decline of 4.2%); whereas negligible difference
407 in prediction performance was observed between the spatial and spatiotemporal parameter
408 transfer schemes (median decline of 12.4% and 13.9% respectively). These results suggest
409 that the stability of hydrological model parameters tends to be higher in the temporal domain
410 than in the spatial domain, and are consistent with previous studies conducted in different
411 parts of the world [*Parajka et al.*, 2005; *Zhang and Chiew*, 2009]. We also demonstrated that
412 the relative superiority of temporal parameter transfer scheme is preserved even when: (1) the
413 spatial distance between donor and receiver catchments is reduced, or (2) the temporal lag
414 between calibration and validation periods is increased. Nonetheless, we found that an
415 increase in the temporal lag between calibration and validation periods reduces the model
416 performance gap between the temporal scheme and the spatial and spatiotemporal schemes.
417 This finding, combined with the negligible difference observed between the spatial and
418 spatiotemporal schemes, suggest that spatiotemporal transfer of hydrological model
419 parameters has the potential to be a viable option for climate change related studies, as
420 envisioned in the trading space for time framework. However, further research is still needed
421 to better understand the relationship between the spatial and temporal aspects of catchment
422 hydrological variability with increasing time lag between the calibration and validation
423 periods.
424

425 **Acknowledgements**

426 We are thankful to Mark Rayment and an anonymous reviewer for providing valuable
427 comments and suggestions that have greatly improved this paper. This research was
428 supported in part by the US National Science Foundation (NSF) grants 0922100 and

1027870. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

References

Beven, K. J. (2001), *Rainfall-runoff modelling: the primer*, Wiley, Chichester.

Blöschl, G., M. Sivapalan, and T. Wagener (2013), *Runoff prediction in ungauged basins: synthesis across processes, places and scales*, Cambridge University Press.

Brigode, P., L. Oudin, and C. Perrin (2013), Hydrological model parameter instability: A source of additional uncertainty in estimating the hydrological impacts of climate change?, *J. Hydrol.*, *476*, 410–425, doi:10.1016/j.jhydrol.2012.11.012.

Coron, L., V. Andréassian, C. Perrin, J. Lerat, J. Vaze, M. Bourqui, and F. Hendrickx (2012), Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments, *Water Resour. Res.*, *48*(5), W05552, doi:10.1029/2011WR011721.

Croke, B. F. W., W. S. Merritt, and A. J. Jakeman (2004), A dynamic model for predicting hydrologic response to land cover changes in gauged and ungauged catchments, *J. Hydrol.*, *291*(1–2), 115–131, doi:http://dx.doi.org/10.1016/j.jhydrol.2003.12.012.

Eckhardt, K., L. Breuer, and H. G. Frede (2003), Parameter uncertainty and the significance of simulated land use change effects, *J. Hydrol.*, *273*(1-4), 164–176, doi:10.1016/S0022-1694(02)00395-5.

Ehret, U. et al. (2014), Advancing catchment hydrology to deal with predictions under change, *Hydrol. Earth Syst. Sci.*, *18*(2), 649–671, doi:10.5194/hess-18-649-2014.

Gill, M. K., Y. H. Kaheil, A. Khalil, M. McKee, and L. Bastidas (2006), Multiobjective particle swarm optimization for parameter estimation in hydrology, *Water Resour. Res.*, *42*(7), W07417, doi:10.1029/2005WR004528.

Goswami, M., and K. M. O'Connor (2007), Comparative assessment of six automatic optimization techniques for calibration of a conceptual rainfall—runoff model, *Hydrol. Sci. J.*, *52*(3), 432–449, doi:10.1623/hysj.52.3.432.

Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez (2009), Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *J. Hydrol.*, *377*(1-2), 80–91, doi:10.1016/j.jhydrol.2009.08.003.

Hamon, W. R. (1963), Computation of direct runoff amounts from storm rainfall, *Int. Assoc. Sci. Hydrol. Publ*, *63*, 52–62.

Hrachowitz, M. et al. (2013), A decade of Predictions in Ungauged Basins (PUB)—a review, *Hydrol. Sci. J.*, *58*(6), 1198–1255, doi:10.1080/02626667.2013.803183.

Juston, J., J. Seibert, and P.-O. Johansson (2009), Temporal sampling strategies and uncertainty in calibrating a conceptual hydrological model for a small boreal catchment, *Hydrol. Process.*, *23*(21), 3093–3109, doi:10.1002/hyp.7421.

Kennedy, J., and R. Eberhart (1995), Particle swarm optimization, in *IEEE International Conference on Neural Networks Proceedings, Vols 1-6*, pp. 1942–1948.

Klemeš, V. (1986), Operational testing of hydrological simulation models, *Hydrol. Sci. J.*, *31*(1), 13–24, doi:10.1080/02626668609491024.

Kokkonen, T. S., A. J. Jakeman, P. C. Young, and H. J. Koivusalo (2003), Predicting daily flows in ungauged catchments: model regionalization from catchment descriptors at the Coweeta Hydrologic Laboratory, North Carolina, *Hydrol. Process.*, *17*, 2219–2238, doi:10.1002/hyp.1329.

Liu, Y. (2009), Automatic calibration of a rainfall–runoff model using a fast and elitist multi-objective particle swarm algorithm, *Expert Syst. Appl.*, *36*(5), 9533–9538, doi:10.1016/j.eswa.2008.10.086.

Maurer, E. P., A. W. Wood, J. C. Adam, D. P. Lettenmaier, and B. Nijssen (2002), A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States, *J. Clim.*, *15*, 3237–3251, doi:10.1175/1520-0442(2002)015<3237:althbd>2.0.co;2.

McIntyre, N., H. Lee, H. Wheater, A. Young, and T. Wagener (2005), Ensemble predictions of runoff in ungauged catchments, *Water Resour. Res.*, *41*, W12434, doi:10.1029/2005wr004289.

Merz, R., and G. Blöschl (2004), Regionalisation of catchment model parameters, *J. Hydrol.*, *287*(1-4), 95–123, doi:10.1016/j.jhydrol.2003.09.028.

Merz, R., J. Parajka, and G. Blöschl (2009), Scale effects in conceptual hydrological modeling, *Water Resour. Res.*, *45*(9), W09405, doi:10.1029/2009WR007872.

Merz, R., J. Parajka, and G. Blöschl (2011), Time stability of catchment model parameters: Implications for climate impact analyses, *Water Resour. Res.*, *47*(2), W02531, doi:10.1029/2010WR009505.

Montanari, A. et al. (2013), "Panta Rhei—Everything Flows": Change in hydrology and society—The IAHS Scientific Decade 2013–2022, *Hydrol. Sci. J.*, *58*(6), 1256–1275, doi:10.1080/02626667.2013.809088.

Nash, J. E., and J. V Sutcliffe (1970), River flow forecasting through conceptual models part I — A discussion of principles, *J. Hydrol.*, *10*, 282–290, doi:10.1016/0022-1694(70)90255-6.

498 Oudin, L., V. Andreassian, C. Perrin, C. Michel, and N. Le Moine (2008), Spatial proximity,
499 physical similarity, regression and ungaged catchments: A comparison of regionalization
500 approaches based on 913 French catchments, *Water Resour. Res.*, *44*(3), W03413–
501 W03413, doi:10.1029/2007WR006240.

502 Parajka, J., R. Merz, and G. Blöschl (2005), A comparison of regionalisation methods for
503 catchment model parameters, *Hydrol. Earth Syst. Sci.*, *9*(3), 157–171.

504 Patil, S., and M. Stieglitz (2012), Controls on hydrologic similarity: role of nearby gauged
505 catchments for prediction at an ungauged catchment, *Hydrol. Earth Syst. Sci.*, *16*(2),
506 551–562, doi:10.5194/hess-16-551-2012.

507 Patil, S., and M. Stieglitz (2014), Modelling daily streamflow at ungauged catchments: what
508 information is necessary?, *Hydrol. Process.*, *28*(3), 1159–1169, doi:10.1002/hyp.9660.

509 Patil, S. D., P. J. Wigington, S. G. Leibowitz, E. A. Sproles, and R. L. Comeleo (2014a),
510 How does spatial variability of climate affect catchment streamflow predictions?, *J.*
511 *Hydrol.*, *517*, 135–145, doi:10.1016/j.jhydrol.2014.05.017.

512 Patil, S. D., P. J. Wigington, S. G. Leibowitz, and R. L. Comeleo (2014b), Use of Hydrologic
513 Landscape Classification to Diagnose Streamflow Predictability in Oregon, *JAWRA J.*
514 *Am. Water Resour. Assoc.*, *50*(3), 762–776, doi:10.1111/jawr.12143.

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

517 Post, D. A., and A. J. Jakeman (1999), Predicting the daily streamflow of ungauged
518 catchments in S.E. Australia by regionalising the parameters of a lumped conceptual
519 rainfall-runoff model, *Ecol. Modell.*, *123*(2-3), 91–104, doi:10.1016/S0304-
520 3800(99)00125-8.

521 Razavi, S., and B. A. Tolson (2013), An efficient framework for hydrologic model calibration
522 on long data periods, *Water Resour. Res.*, *49*(12), 8418–8431,
523 doi:10.1002/2012WR013442.

524 Refsgaard, J. C. (1997), Parameterisation, calibration and validation of distributed
525 hydrological models, *J. Hydrol.*, *198*(1-4), 69–97, doi:10.1016/S0022-1694(96)03329-
526 X.

527 Refsgaard, J. C., and J. Knudsen (1996), Operational Validation and Intercomparison of
528 Different Types of Hydrological Models, *Water Resour. Res.*, *32*, 2189–2202,
529 doi:10.1029/96wr00896.

530 Refsgaard, J. C. et al. (2014), A framework for testing the ability of models to project climate
531 change and its impacts, *Clim. Change*, *122*(1-2), 271–282, doi:10.1007/s10584-013-
532 0990-2.

533 Singh, R., T. Wagener, K. van Werkhoven, M. E. Mann, and R. Crane (2011), A trading-
534 space-for-time approach to probabilistic continuous streamflow predictions in a

changing climate – accounting for changing watershed behavior, *Hydrol. Earth Syst. Sci.*, *15*(11), 3591–3603, doi:10.5194/hess-15-3591-2011.

Singh, R., K. van Werkhoven, and T. Wagener (2014), Hydrological impacts of climate change in gauged and ungauged watersheds of the Olifants basin: a trading-space-for-time approach, *Hydrol. Sci. J.*, *59*(1), 29–55, doi:10.1080/02626667.2013.819431.

Sivapalan, M. et al. (2003), IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences, *Hydrol. Sci. J.*, *48*, 857–880, doi:10.1623/hysj.48.6.857.51421.

Slack, J. R., A. Lumb, and J. M. Landwehr (1993), *Hydro-Climatic Data Network (HCDN) Streamflow Data Set, 1874-1988:*, U.S. Geological Survey, Reston, VA.

Vaze, J., D. A. Post, F. H. S. Chiew, J.-M. Perraud, N. R. Viney, and J. Teng (2010), Climate non-stationarity – Validity of calibrated rainfall–runoff models for use in climate change studies, *J. Hydrol.*, *394*(3-4), 447–457, doi:10.1016/j.jhydrol.2010.09.018.

Wagener, T., and H. S. Wheater (2006), Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty, *J. Hydrol.*, *320*, 132–154, doi:10.1016/j.jhydrol.2005.07.015.

Wagener, T., N. McIntyre, M. J. Lees, H. S. Wheater, and H. V Gupta (2003), Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis, *Hydrol. Process.*, *17*(2), 455–476, doi:10.1002/hyp.1135.

Wagener, T., M. Sivapalan, P. A. Troch, B. L. McGlynn, C. J. Harman, H. V Gupta, P. Kumar, P. S. C. Rao, N. B. Basu, and J. S. Wilson (2010), The future of hydrology: An evolving science for a changing world, *Water Resour. Res.*, *46*(5), W05301, doi:10.1029/2009WR008906.

Wang, R., and L. Kalin (2011), Modelling effects of land use/cover changes under limited data, *Ecohydrology*, *4*(2), 265–276, doi:10.1002/eco.174.

Westra, S., M. Thyer, M. Leonard, D. Kavetski, and M. Lambert (2014), A strategy for diagnosing and interpreting hydrological model nonstationarity, *Water Resour. Res.*, *50*(6), 5090–5113, doi:10.1002/2013WR014719.

Xia, Y., Z.-L. Yang, C. Jackson, P. L. Stoffa, and M. K. Sen (2004), Impacts of data length on optimal parameter and uncertainty estimation of a land surface model, *J. Geophys. Res. Atmos.*, *109*(D7), D07101, doi:10.1029/2003JD004419.

Young, A. R. (2006), Stream flow simulation within UK ungauged catchments using a daily rainfall-runoff model, *J. Hydrol.*, *320*, 155–172, doi:10.1016/j.jhydrol.2005.07.017.

Zhang, X., R. Srinivasan, K. Zhao, and M. Van Liew (2009), Evaluation of global optimization algorithms for parameter calibration of a computationally intensive hydrologic model, *Hydrol. Process.*, *23*(3), 430–441, doi:10.1002/hyp.7152.

571 Zhang, Y., and F. H. S. Chiew (2009), Relative merits of different methods for runoff
572 predictions in ungauged catchments, *Water Resour. Res.*, *45*, W07412–W07412,
573 doi:10.1029/2008WR007504.

574

575

576 Table 1: Comparison of the two catchment groups (from Figure 9) in terms of mean annual
577 rainfall, annual runoff ratio and climate aridity index.
578

Catchment group	**Median value of mean annual rainfall (P)**	**Median value of annual runoff ratio (Q/P)**	**Median value of climate aridity index (PET/P)**
Group 1: Spatial or spatiotemporal transfer performs best	1110 mm	0.43	0.73
Group 2: Temporal transfer performs best	1185 mm	0.41	0.69

579

580

Figures

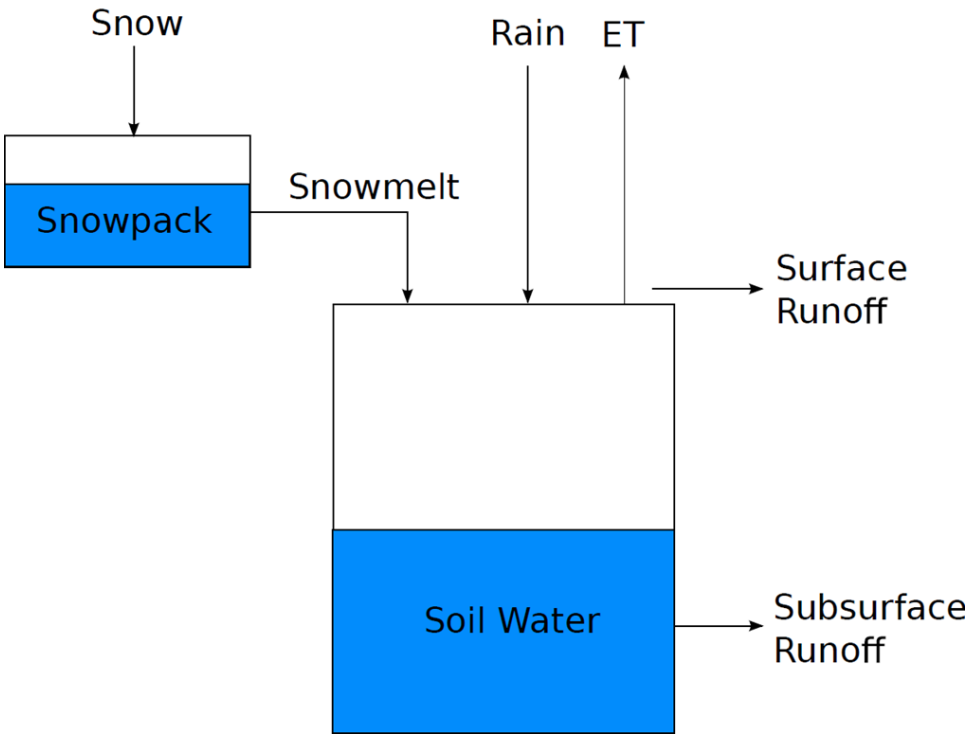

Figure 1: Overview of the EXP-HYDRO model components and fluxes (from *Patil et al.* [2014a]).

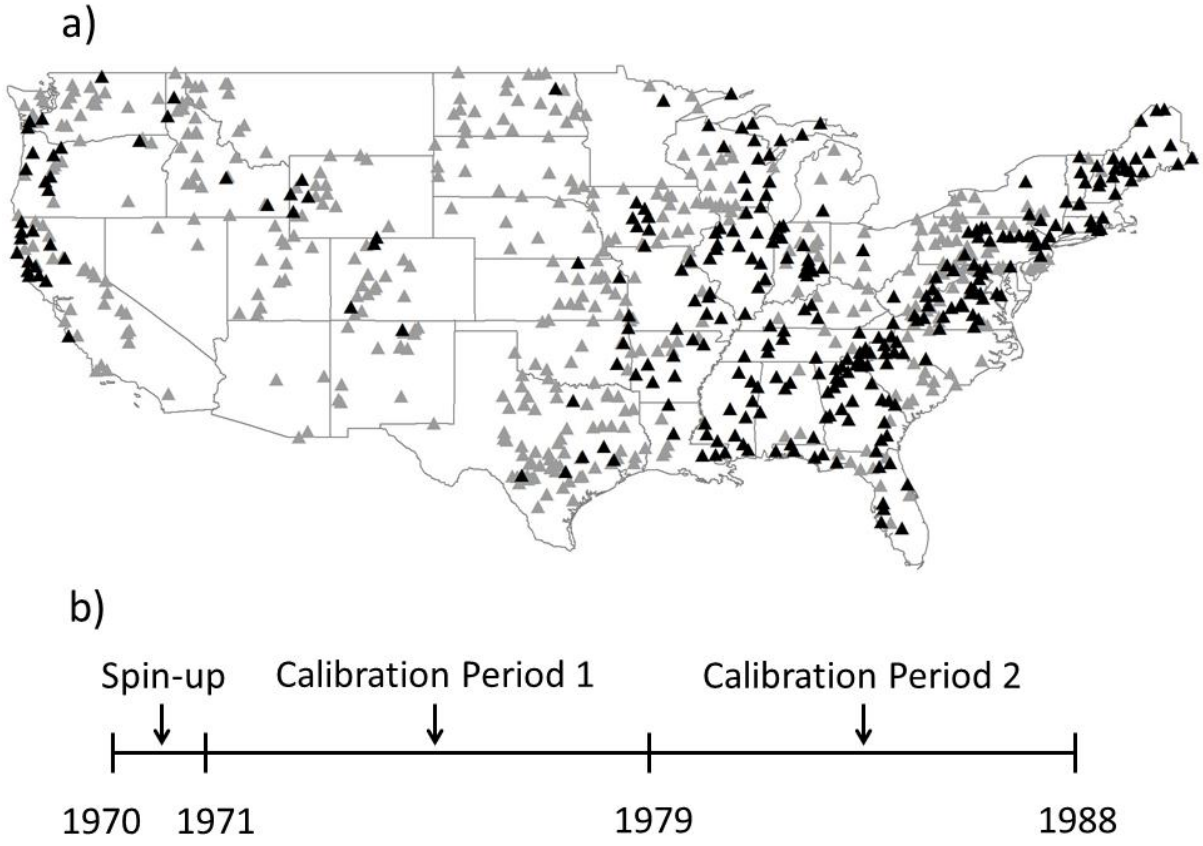

Figure 2: a) Location of all 756 catchments within the continental US where the EXP-HYDRO model is implemented; the 294 study catchments that are retained for parameter transfer experiments are shown in black triangles, and b) Schematic representation of the timeline of the calibration periods.

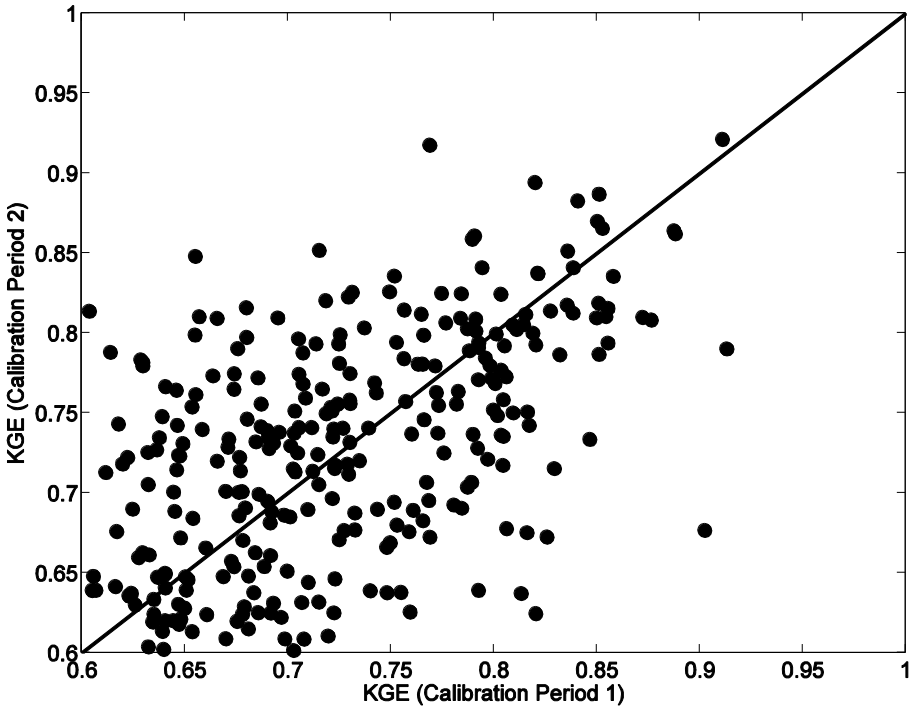

Figure 3: A 1:1 comparison of the KGE values for Calibration Periods 1 and 2.

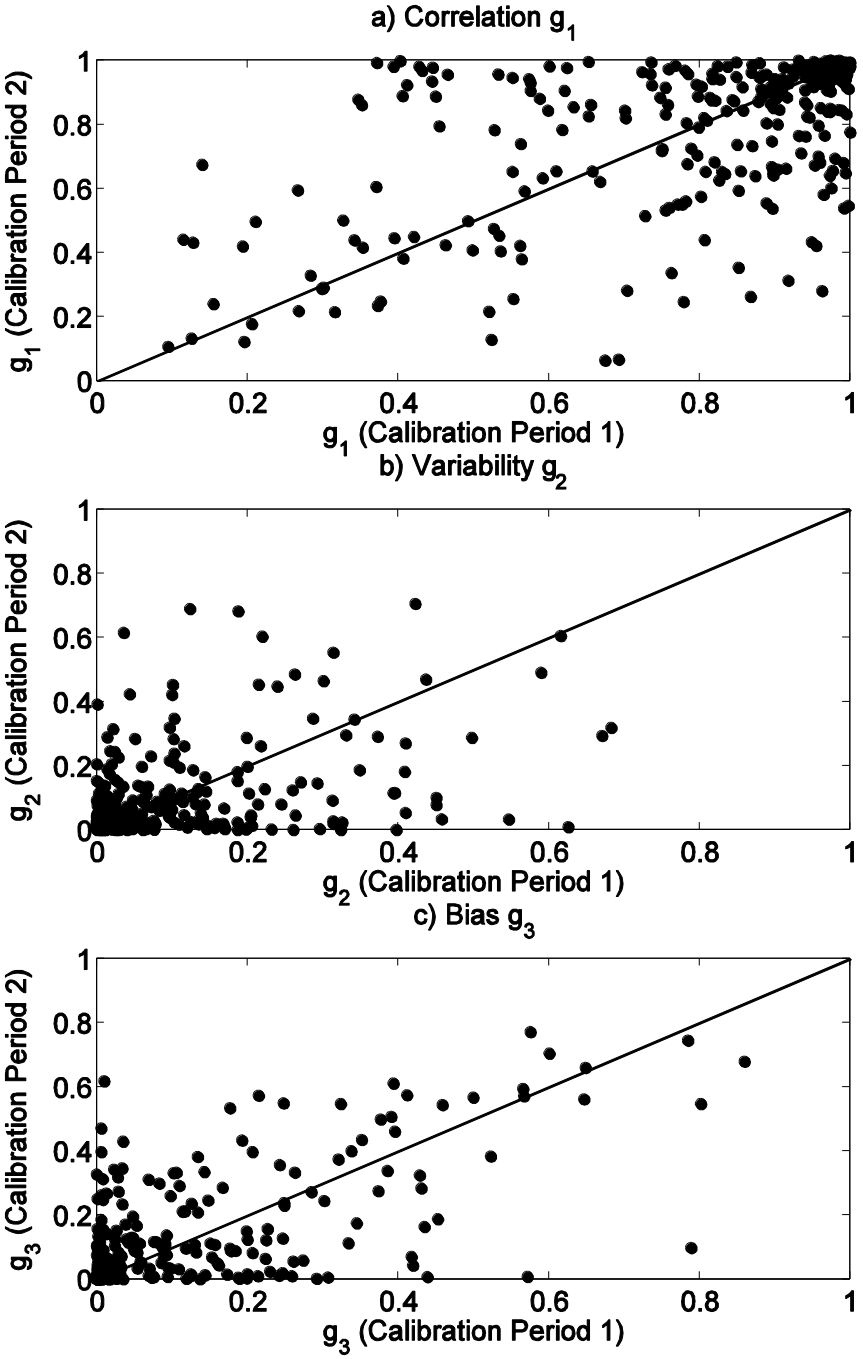

Figure 4: A 1:1 comparison of the three KGE components for Calibration Periods 1 and 2: a) Correlation g_1, b) Variability g_2, and c) Bias g_3.

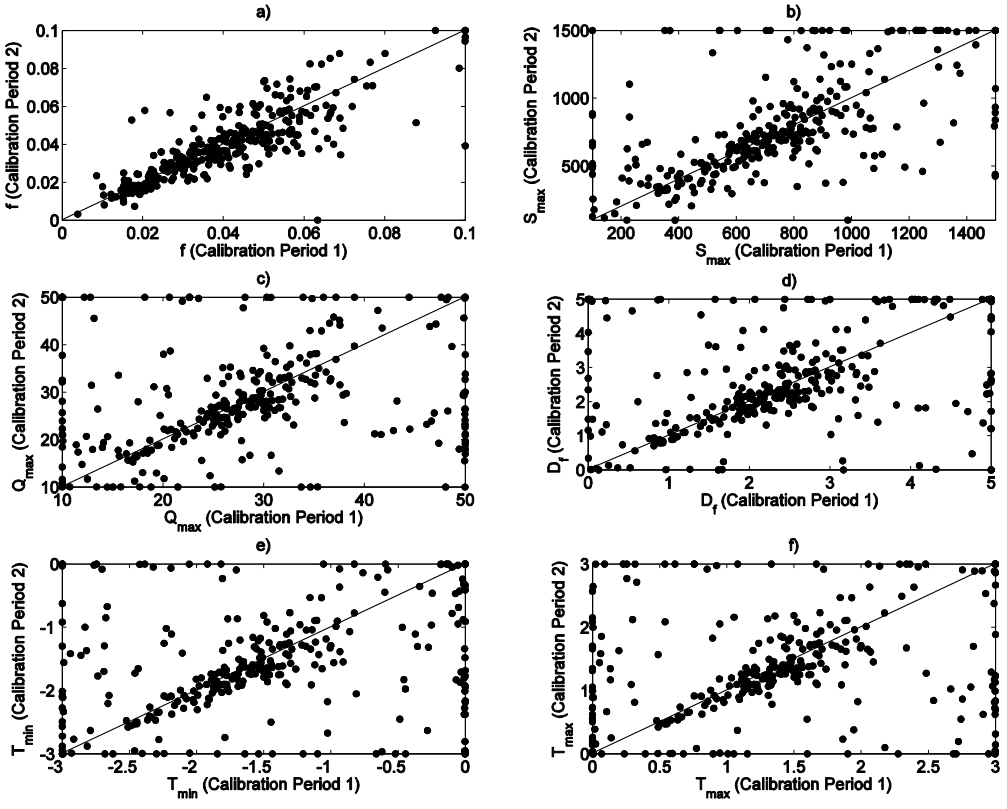

Figure 5: A 1:1 comparison of all 6 EXP-HYDRO parameter values for Calibration Periods 1 and 2.

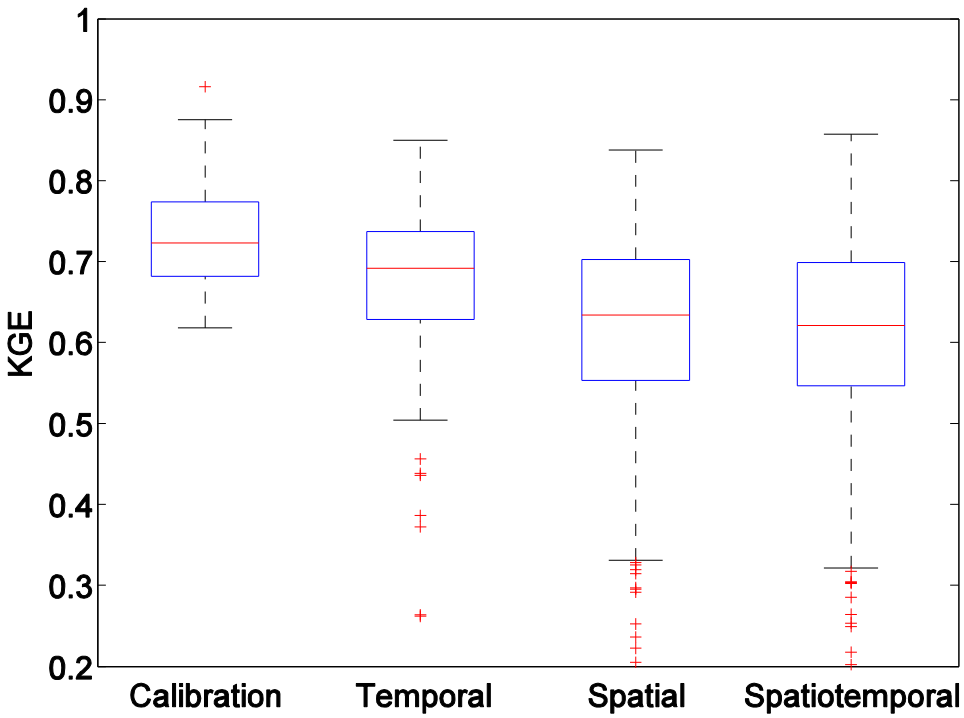

Figure 6: A box-plot comparison of the KGE values for calibration and the three parameter transfer scenarios.

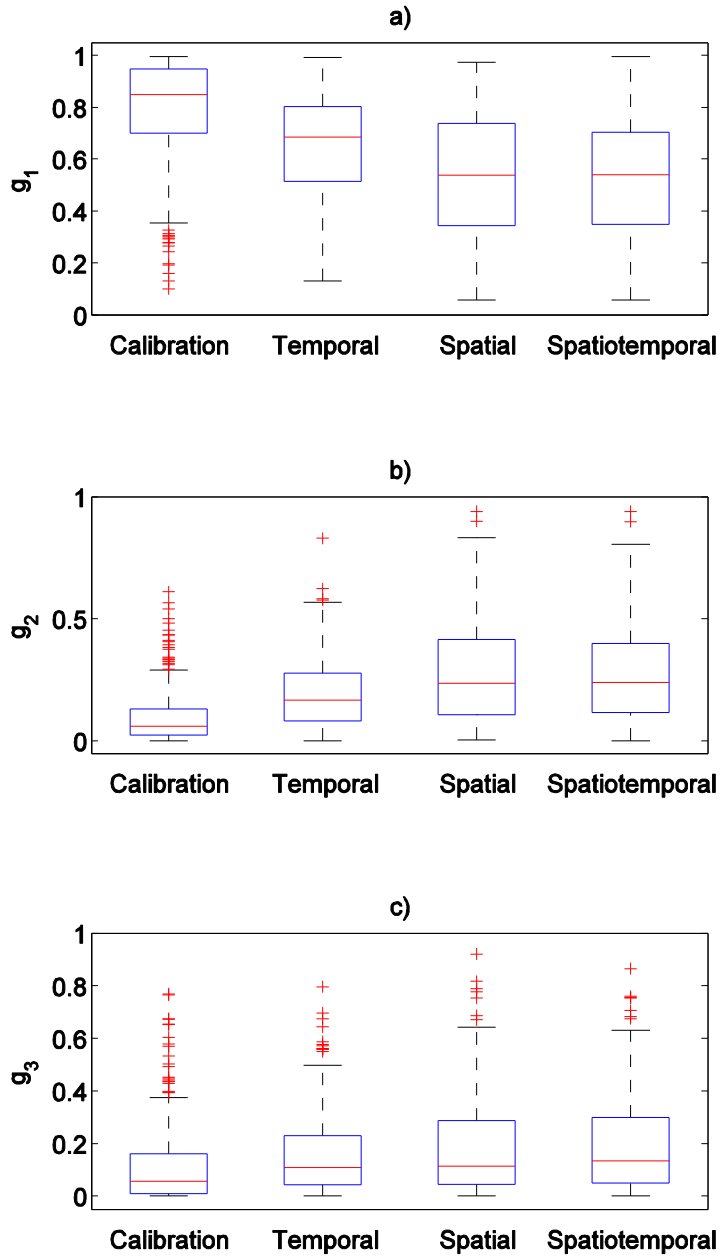

Figure 7: A box-plot comparison of the three KGE components for calibration and the three parameter transfer scenarios.

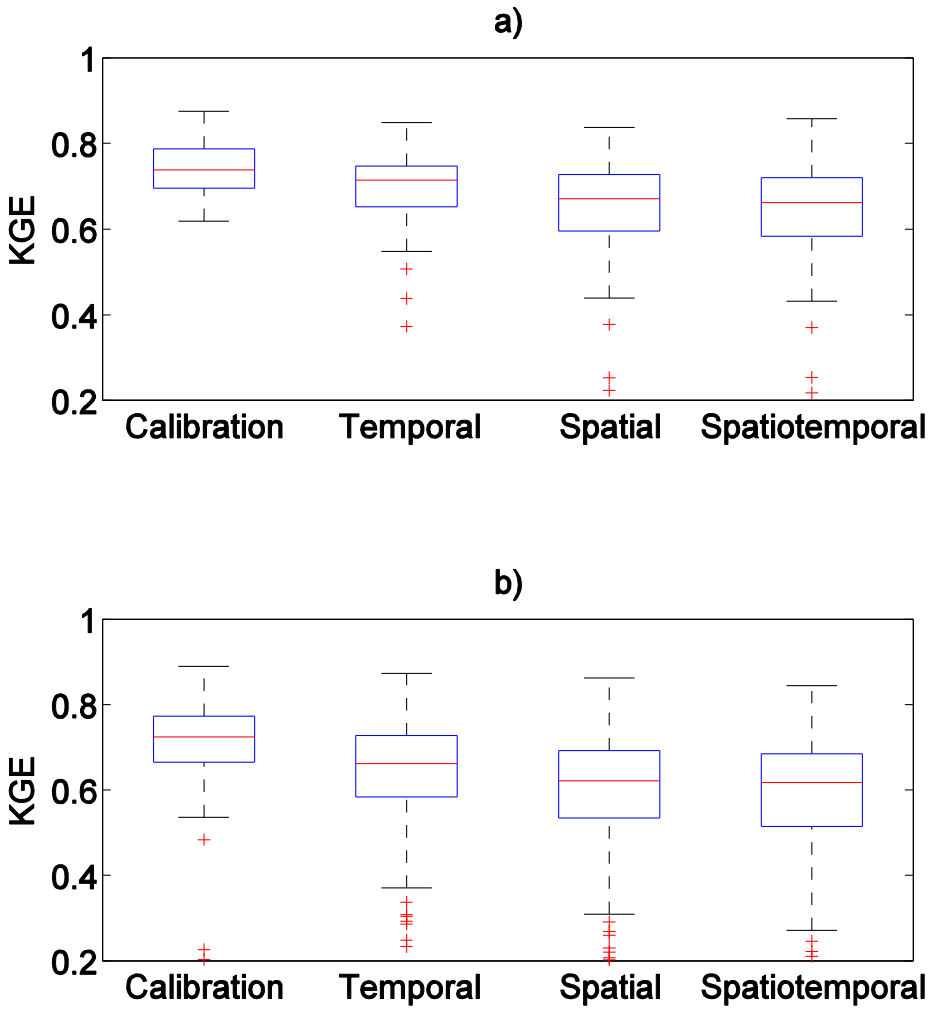

Figure 8: A box-plot comparison of the KGE values for calibration and the three parameter transfer scenarios, shown separately for a) Special Condition 1, and b) Special Condition 2.

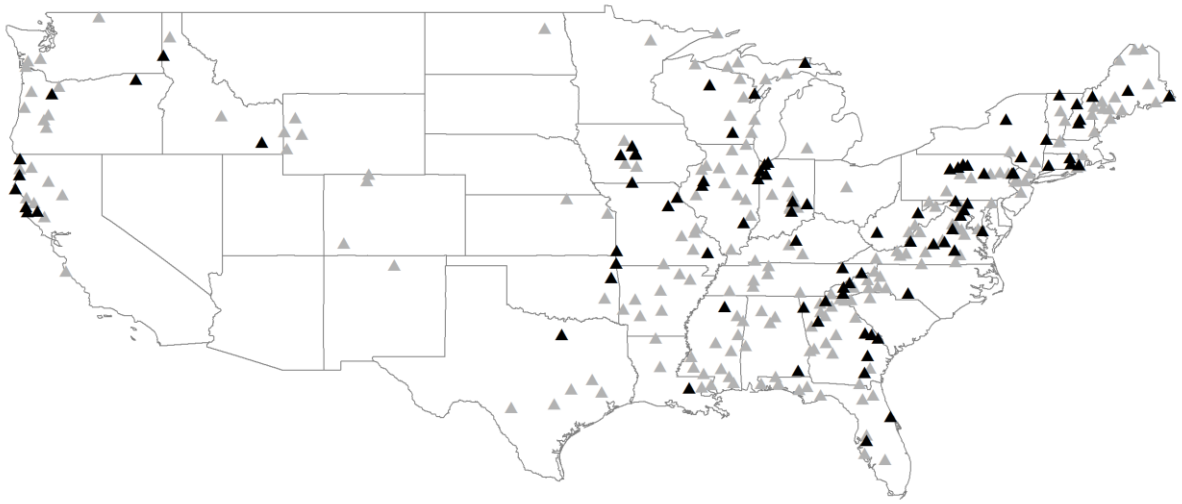

Figure 9: Location of the catchments where either the spatial or spatiotemporal parameter transfer scheme performs best (Black triangles). Catchments where the temporal parameter transfer scheme performs best are shown as grey triangles.