Streamflow Prediction in Ungauged Basins Located Within Data-scarce Regions

2019

Mohammadhossein Alipour
University of Central Florida

STREAMFLOW PREDICTION IN UNGAUGED BASINS LOCATED WITHIN DATA-SCARCE REGIONS

by

MOHAMMADHOSSEIN ALIPOUR
B.Sc. Azad University, 2006
M.A.Sc. University of British Columbia, 2012

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Civil, Environmental and Construction Engineering
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Spring Term
2019

Major Professor: Kelly Kibler

ABSTRACT

Preservation and or restoration of riverine ecosystem requires quantification of alterations inflicted by water resources development projects. Long records of streamflow data are the first piece of information required in order to enable this analysis. Ungauged catchments located within data-scarce regions lack long records of streamflow data. In this dissertation, a multi-objective framework named Streamflow Prediction under Extreme Data-scarcity (SPED) is proposed for streamflow prediction in ungauged catchments located within large-scale regions of minimal hydrometeorologic observation. Multi-objective nature of SPED allows for balancing runoff efficiency with selection of parameter values that resemble catchment physical characteristics. Uncertain and low-resolution information are incorporated in SPED as soft data along with sparse observations. SPED application in two catchments in southwestern China indicates high runoff efficiency for predictions and good estimation of soil moisture capacity in the catchments. SPED is then slightly modified and tested more comprehensively by application to six catchments with diverse hydroclimatic conditions. SPED performance proves satisfactory where traditional flow prediction approaches fail. SPED also proves comparable or even better than data-intensive approaches. Utility of SPED versus a simpler catchment similarity model for the study of flow regime alteration is pursued next by streamflow prediction in 32 rivers in southwestern China. The results indicate that diversion adversely alters the flow regime of the rivers while direction and pattern of change remain the same regardless of the flow prediction method of choice. However, the results based on SPED consistently indicate more substantial alterations to the flow regime of the rivers after diversion. Finally, the value added by a limited number of streamflow observations to improvement of predictions in an ungauged catchment located within a data-scarce region is

studied. The large number of test scenarios indicate that there may be very few near-universal schemes to improve flow predictions in such catchments.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude toward my committee chair and advisor, Dr. Kelly Kibler, for her invaluable guidance throughout my PhD program as well as my deep appreciation for the encouragement and advice she provided me with whenever I faced a challenging and seemingly unresolvable problem in my research. She was always available to answer my questions and discuss my concerns, and her guidance and insightful advice was the key to the progress of my research and academics. I would also like to thank Dr. Dingbao Wang for providing insightful advice as well as my other PhD committee members Dr. Talea Mayo and Dr. Christopher Emrich, for their support and guidance.

Special thanks to my colleagues at Kibler Ecohydraulics Lab who taught me new things both research-wise and life-wise. Meetings, discussions and informal chats with them through the four academic years I spent at UCF substantially contributed to the progress of my research and made life sweeter for me. This unforgettable experience will be with me for the rest of my life.

I would like to express my gratitude toward my lovely parents Behnaz and Ahmad, for their support and advice throughout my life. I would also like to thank my lovely wife Samaneh, who has been my best friend in life and greatest support in the face of challenges.

Finally, I would like to acknowledge that several chapters of this dissertation have previously been published in or are under review by Journal of Hydrology, Hydrological Sciences Journal, and River Research and Applications. These journals did not require written permission to use the published material in my dissertation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW

1.1 Problem Statement and Literature Review

Similar to well-gauged catchments, ecosystem of ungauged catchments located within data-scarce regions of the world may undergo substantial stress levels and alterations as a result of deployment of water resources development projects. Analysis and quantification of these alterations is a necessity in order to identify appropriate schemes to preserve and or restore the riverine ecosystem. Since riverine ecosystem is highly dependent on natural flow regime of the river (Bunn and Arthington, 2002), a thorough impact analysis will not be achieved without a complete assessment of flow regime alterations (Matos et al., 2010; Wang et al., 2016a). Thus, a sufficiently long record of natural streamflow in a river is the first piece of information required for analyzing the impacts of development projects. Such long-term records of streamflow data are also advantageous, for instance, in the planning phase of development projects as well as multiple other tasks such as allocation of water among competing users. Today many rivers in the world still remain ungauged and evidence suggests that globally the number of hydrologic monitoring stations is decreasing rather than growing in recent decades (Fekete and Vörösmarty, 2007; Fekete et al., 2012). In the absence of observed data, alternative approaches are required to simulate and estimate streamflow in ungauged basins. While comprehensive research has taken place in the field of flow prediction (e.g., Seibert and McDonnell, 2002, 2013; Song et al., 2016; Heřmanovský et al., 2017), prediction in smaller catchments within vast areas of minimum hydrometeorologic observation has received little attention from researchers. This, however, does not resonate with the large amount of human interventions/developments in these smaller catchments (Kibler and Alipour, 2017). Besides lack of observed streamflow data, other hydroclimatological observations, such as precipitation and temperature, might be scarce in such regions as well so low-resolution regional/global databases or remotely-sensed data, which are associated with high uncertainties, may be the only source of available data in some of these catchments.

Once streamflow prediction has been enabled, analysis of flow regime alterations can be carried out within ungauged catchments. In particular, diversion hydropower is one type of development that has largely taken place in small catchments of the world (Kibler and Tullos, 2013; Fantin-Cruz et al., 2015; McManamay et al., 2016). Many of these projects are implemented in steep catchments located in remote mountainous areas which are not equipped with flow measurement devices (Narula, 2012; Li et al., 2013; Tuna, 2013). While flow regime alterations due to river impoundment/regulation have largely been studied by researchers around the world (e.g., Timpe and Kaplan, 2017; Batalla et al. 2004; Ngor et al., 2018), evaluation and quantification of the impacts of diversion hydropower projects on natural flow regime of the smaller rivers has only recently received attention from the research community (Anderson et al., 2015; Fantin-Cruz, 2015; Gibeau et al., 2016; Kibler and Alipour, 2017).

1.1.1 Prediction in Ungauged Basins (PUB)

PUB research has recently been an area of significant attention from hydrologists around the world through initiatives of the International Association of Hydrological Sciences (PUB Decade 2003-2012 and the ongoing Panta Rhei Decade 2013-2022). Such initiatives were considered partly as a response to the desire that process understanding and model structural diagnostics replace parameter fitting as the focus of hydrologic research (Hrachowitz et al., 2013). Through completion of the first PUB Decade in 2012, progress was made with respect to multiple aspects of flow prediction and a synthesis framework was proposed as a platform for prediction in ungauged basins. The framework recognizes that catchments are complex and diverse systems, wherein distinct hydrologic response signatures manifest as a results of varied streamflow generation processes (see Blöschl et al. 2013). The framework emphasizes the role of comparative hydrology as an effective tool for learning about catchment functionality through analysis of similarities, for instance with respect to physical characteristics or runoff signatures. Such hydrologic similarities may be employed for predicting streamflow in ungauged catchments (Wagener et al. 2013). While the framework forms a strong foundation upon which practical applications may develop, prescriptive methods to bridge research advancements and practice in genuinely ungauged catchments may make such advances more accessible to water managers in poorly-gauged regions. Without prescriptive methods, routine application of hydrologic modeling to prediction within severely data-scarce regions faces several barriers (Hughes 2016, Koutsouris et al. 2017, Tegegne et al. 2017). Water resources managers in these regions may struggle to overcome severe challenges, including: over-calibration of models (i.e., when the identified optimum parameter set over the calibration period is not the optimum set over a different period; Andréassian et al., 2012; Bardossy et al., 2016; Gelfan et al., 2015); equifinality of potential models (i.e. when many different parameter sets perform equally well at reproducing an output signal; Beven, 1993, 2006; Savenije, 2001); great uncertainty in available information for constraining parameter values (Merz et al., 2009; 2011); limited availability of gauged reference catchments and thus potential for hydrologic dissimilarity to ungauged catchments (Peñas et al. 2014); and lack of observed streamflow data for validation of predictions in ungauged catchments (van Emmerik et al. 2015).

A prescriptive method should be based on simplicity and avoid excessive parameterization or calibration (Ajmal et al., 2015; Skaugen et al., 2015). Over-parameterization and or over-calibration of hydrologic models may lead to misrepresentation of processes contributing to streamflow generation (Kirchner, 2006; Archibald et al., 2014). As an example, Gharari et al. (2013) suggested a practical approach to reduce over-calibration by studying calibration performance over different sub-periods so that time-consistent parameters could be identified (Gharari et al.,2013; Brigode et al., 2015). A prescriptive method should also at least partially address the equifinality problem. Equifinality is often a problem when the selected hydrological model has large parameter spaces and high parameter interdependence so that many parameter sets may be identified as behavioral (Arsenault and Brissette, 2014; Poissant et al, 2017). Utilizing simple hydrological models with a few parameters can help reduce equifinality (Arsenault and Brissette, 2014). Several other approaches have been proposed for reduction of equifinilaity as well (Lüdtke et al., 2014; Ford et al., 2017; Kelleher et al., 2017). Accounting for uncertainties in the available data and information for constraining parameter values is another important quality for a prescriptive hydrological model. Many approaches have been devised/employed to address modeling uncertainties (e.g., Renard et al., 2010; Spence et al., 2013; Zhang et al., 2016). Melching (1995) reviewed the uncertainty sources in hydrological modeling and the methods for addressing these uncertainties. Robustness to some level of catchment dissimilarity is another characteristic desired for a prescriptive hydrological model. Wagener et al. (2007) introduced the need for a catchment classification framework as a starting point to identify hydrologically similar reference catchments to ungauged target catchments. Such a framework would require the global hydrologic community to create a database of gauged catchments classified according to the framework and its associated metrics. Meanwhile, multiple researchers have proposed measures for quantifying dissimilarity between catchments (e.g., McIntyre et al., 2004; 2005; Fry et al., 2013). Hydrologic modeling within data-scarce regions with a limited selection of gauged reference catchments (may consist of only one gauged catchment) is particularly prone to catchment dissimilarity. Finally, alternative methods for validation of predictions in ungauged catchments is another important feature desired for a prescriptive hydrological model. Validation in genuinely ungauged catchments is a challenging task that has received only little attention from the research community (van Emmerik et al., 2015). van Emmerik et al. (2015) argued that creativity and in-depth knowledge of the local hydrology should be utilized to find validation data other than observations from gauging stations. Inclusion of soft data in hydrological modeling can facilitate introduction of other criteria besides runoff efficiency in calibration and validation phases (Seibert and McDonnell, 2002; Parajka et al., 2007a; Rinderer and Seibert, 2012; Seibert and McDonnell, 2013; Arnold et al., 2015). By incorporation of multiple criteria in model validation, dependency on observed streamflow data for validation is lowered.

1.1.1.1 Regionalization

A large number of streamflow prediction studies and methods are founded upon regionalization concepts. Regionalization is enabled through using a regional hydrologic network for development of relationships between a hydrologic model parameters and catchment characteristics and or runoff signatures (Zhang et al. 2008, Yadav et al. 2007). Such regionalization of model parameters has been implemented widely and proven efficient for streamflow prediction over vast regions (e.g., McIntyre et al., 2005; Parajka et al., 2005; Eng et al, 2007; Parajka et al., 2007a; Post, 2009;

Samaniego et al., 2010; Samuel et al., 2011; Wang et al., 2012; Kult et al., 2014; Ibrahim et al., 2015; Heřmanovský et al., 2017; Swain and Patra, 2017; He et al. (2011); Razavi and Coulibaly (2012) provide a review of such techniques). The results of regionalization are often acceptable, but performance may vary over scales and location within the region of interest (e.g., Kult et al., 2014). Several studies have evaluated and compared the performance of various regionalization techniques. For instance, Vandewiele and Elias (1995) used two different regionalization techniques, including kriging and using parameter values from only a limited number of neighboring basins, to estimate monthly water balance in 75 basins in Belgium. The kriging method performed well in 72% of the catchments analyzed while regionalization with a limited number of neighboring basins presented a good performance only in 44% of the catchments. Oudin et al. (2008) compared three classical regionalization techniques, including regression, spatial proximity, and physical similarity, to estimate daily streamflow over 913 catchments in France using two lumped rainfall-runoff models. The results indicated that spatial proximity had the best performance followed by physical similarity and regression. Merz and Blöschl (2004) had a similar observation where analysis of 308 catchments in Austria illustrated that the best regionalization techniques were regionalization by kriging as well as using the average parameter values from immediate upstream and downstream neighbors. Parajka et al. (2005) analyzed multiple regionalization techniques (based on local/global averaging, spatial proximity, regression, and similarity) in 320 Austrian catchments, and found that kriging based on spatial correlation performed among the two best methods. Arsenault and Brissette (2014) also compared the three common regionalization techniques, including multiple linear regression, spatial proximity and physical similarity, for calibrating a conceptual model (HSAMI; Fortin, 2000) to estimate streamflow in 268 basins in Quebec. The results were similar to the previous studies, and it was found that four to seven donor catchments were required to achieve optimal performance in regional calibration of the conceptual model used.

The fact that spatial proximity methods were found among the best techniques in all these studies hints at how important the density of a network of gauged catchments might be to the success of regionalization techniques. The results, thus, provide insight into some of the intrinsic deficiencies and limitations of regionalization techniques such as high dependence on the density of gauged catchments (Oudin et al. 2008, Parajka et al. 2015; Lebecherel et al., 2016). Robustness of regionalization techniques based on spatial proximity to the density of the network of gauged catchments was further analyzed and compared using two different methods by Lebecherel et al. (2016). The robustness assessment methods included the hydrometrical random reduction (HRand) and the hydrometrical desert method (HDes). HRand is based on random thinning of a hydrometrical network while HDes is based on progressive exclusion of the closest donor catchments. Application of these two methods over a dataset of 609 small to medium-size catchments in France indicated that HDes is a more conservative approach as it results in the fastest decrease in model efficiency. Thus, HDes is recommended over HRand for providing a more realistic (though more pessimistic as well) view on robustness of spatial proximity regionalization to the density of gauging stations.

Despite the promise and wide application of regionalization techniques, it was clarified that the density of available gauging stations in the region surrounding a particular catchment of interest (target catchment) highly influences the reliability of streamflow predictions (Oudin et al. 2008, Parajka et al. 2015; Lebecherel et al., 2016). Data sparsity can also lead to introduction of higher levels of catchment dissimilarity since it makes it harder to find similar reference catchment(s) to target catchment(s). Thus, applicability of regionalization methods may be limited in areas where densities of climatic/hydrologic monitoring are (far) below the recommendations of World Meteorological Organization (WMO, 2008). WMO has recommended minimum densities for climatic/hydrologic monitoring given physiographic units (Table 1). This minimum network is intended to avoid serious deficiencies in water resources management on a scale corresponding to the overall level of economic development and environmental needs of a country. After this minimum network density has been established, general hydrological characteristics such as rainfall and runoff can be estimated through regionalization. Although these recommendations can be understood as minimum recommended densities for basic management functions (e.g. long-term water resources planning), even this minimal level of coverage has not been achieved in many areas of the world (Kundzewicz, 1997; Mishra and Coulibaly, 2009). For instance, 65% of mountainous basins do not meet the WMO recommendations (Perks et al. 1996). While quality of a regionalization model can lower the required density, evidence suggests that globally the number of hydrologic monitoring stations is decreasing rather than growing in recent decades (Fekete and Vörösmarty, 2007; Fekete et al., 2012). Therefore, alternative methods are required for streamflow prediction in ungauged catchments within vast areas of minimal data. An alternative modeling approach in such cases must rely on the existing data, including regional/global data, and be robust to the extensive uncertainty inherent in sparse and low-resolution data. It should also be robust to some level of catchment dissimilarity.

Table 1. Recommended minimum densities of stations for developing regional hydrological relationships (area in km^2 per station; WMO, 2008)

Physiographic unit	Precipitation		Evaporation	Streamflow	Sediments	Water quality
	Non-recording	Recording				
Coastal	900	9,000	50,000	2,750	18,300	55,000
Mountains	250	2,500	50,000	1,000	6,700	20,000
Interior plains	575	5,750	5,000	1,875	12,500	37,500
Hilly/undulating	575	5,750	50,000	1,875	12,500	47,500
Small islands	25	250	50,000	300	2,000	6,000
Urban areas	-	10-20	-	-	-	-
Polar/arid	10,000	100,000	100,000	20,000	200,000	200,000

1.1.1.2 Multi-Objective Hydrological Modeling

Multi-objective calibration of hydrologic models has been established as an approach to lowering modeling uncertainties and developing a more realistic portrayal of hydrologic mechanisms (Madsen, 2000; Ajami et al., 2004; Fenicia et al., 2007; Parajka et al., 2007b, Khu et al., 2008; Shafii and Smedt, 2009; Zhou et al., 2014; Wang and Brubaker, 2015; Efstratiadis and Koutsoyiannis (2010) provide a review of such techniques). Whether modeling is performed using a conceptual model (e.g., Le and Nguyen, 2018), a physically distributed model (e.g., Shrestha and Rode, 2008), or even a regionalization technique (e.g., Zhang et al., 2008), there is opportunity to incorporate multiple criteria in the objective function of the calibration problem. Some of the early research in this area was conducted by Yapo (1996) and Yapo et al. (1998). Yapo (1996) developed a new multi-objective optimization algorithm and used it to calibrate the Soil Moisture Accounting model of the National Weather Service River Forecasting System in Leaf River Watershed near Collins, Mississippi. This was carried out twice, once by taking daily root mean square and heteroscedastic maximum likelihood estimator as the two objectives of the problem and once by matching the rising and falling limbs of the hydrograph as the two objectives of the problem. This study was important since it illustrated the possibility of inclusion of more than a single objective in efficient model calibration.

Incorporation of fuzzy theory and 'soft' data (i.e., data associated with high uncertainties) in multi-objective calibration of hydrological models has been pursued by several researchers in order to account for problem uncertainties and utilize qualitative knowledge in model calibration. For instance, Yu and Yang (2000) set the mean absolute percent error (MPE) of 11 different flow stages as their objectives and prioritized the objectives by their fuzzy membership functions. The authors applied their method to Gao-Ping Creek in Taiwan using the Hydrologiska Byråns Vattenbalansavdelning (HBV) hydrological model (Bergström, 1976). The fuzzy multi-criteria objective function proved better than a single RMSE (Root Mean Squared Error) or MPE. Yang et al. (2004) created a fuzzy multi-criteria objective function consisting of three hydrograph characteristics including time to peak flow, peak flow, and total runoff volume. They applied their method to estimation of streamflow in Pa-Chang Creek in Taiwan. The results indicated that their multi-criteria objective function, which used only partial information from the hydrograph, was comparable to a single-criterion objective function (weighted root-mean square error) which used the entire hydrograph for calibration. In another interesting study, Seibert and McDonnell (2002) combined soft data, in the form of qualitative information from experimentalists, with traditional goodness of fit criteria to form a multi-criteria objective function for the Maimai catchment in New Zealand. Calibration results indicated that inclusion of soft data slightly lowered the goodness of fit of the estimations to observations. However, estimated parameter values provided a better overall performance than the traditionally estimated parameter values. The authors argued that reduction of parameter uncertainty through introduction of soft data and providing a more realistic representation of catchment phenomena should be favorable to methods based only on goodness of fit to observations; arguing that less accurate answers for the right reasons should be preferable to right answers for the wrong reasons. Among more recent works that take advantage of fuzzy theory for hydrologic model calibration, Kamali and Mousavi (2014) performed fuzzy multi-objective calibration of HEC-HMS hydrologic model (USACE, 2008) to estimate flood in the Tamar basin in Iran. Four objectives, including RMSE, flood volume, time to peak, and peak flow, were considered. Two different multi-criteria objective functions, each consisting of three objectives, and a single-criterion objective function were evaluated. The fuzzy multi-objective calibrated model proved better than the single-objective model in estimating flood magnitude.

As an example of a hydrological model that has been extensively evaluated by researchers for multi-objective calibration, the process-based and semi-distributed Soil and Water Assessment Tool (SWAT) can be named. Confesor and Whittaker (2007) used a multi-objective optimization algorithm for automatic calibration of SWAT, and generated daily streamflow data for the Calapooia River watershed in Oregon. The objectives taken into account included RMSE of peak flows as well as RMSE of low flows. The modeling results indicated Nash-Sutcliff efficiencies (NSEs) of 0.86 and 0.81 respectively for calibration and validation periods. Muleta and Nicklow (2005) performed a two-stage calibration of SWAT to optimize streamflow and sediment concentration estimations in Big Creek watershed. Parameter screening (for example, through estimating the value of some parameters using field data) was performed first to reduce the number of calibrable parameters of SWAT. Parameter estimation was then performed using a genetic algorithm. While the calibration results indicated considerable improvement over default simulations and previous works, the verification results were poor for both streamflow and sediment concentration. Bekele and Nicklow (2007) also performed estimation of streamflow and sediment concentration in Big Creek watershed using SWAT. Two different calibration scenarios were considered, once by fitting different portions of the streamflow time series using relevant objective functions and once by calibrating the model to two gauging stations in the watershed. The results indicated better performance by the second scenario. However, validation results were poor most probably due to the short period of data used for calibration. Rajib et al. (2016) calibrated SWAT in two watersheds in Indiana, The Upper Wabash and Cedar Creek, by incorporating spatially distributed data on soil moisture. Incorporation of these remotely sensed data improved simulation of surface soil moisture. Streamflow estimates as well as root zone soil moisture simulations made substantial improvement when root zone soil moisture estimates from limited field sensor data were incorporated as well.

A promising direction for improved representation of catchment function within models might be the incorporation of *a priori* parameter estimates/distributions in the multi-objective calibration of hydrological models (Parajka et al. 2007b). However, extensive observed data has normally been an important requirement for multi-objective techniques based on *a priori* parameter estimates. Thus, methods for streamflow prediction in ungauged catchments within regions of minimal available data are lacking. Merz et al. (2009, 2011), for instance, constrained model parameters to pre-defined Beta distribution functions to form their multi-objective calibration technique. The proposed technique was applied to a large number of gauging stations in Australia. Formation of the *a priori* parameter distributions was carried out based on observed catchment-specific data including previous local modeling experiments. This makes similar application of the proposed method very difficult, if not impossible, in remote catchments of a data-scarce region where available data and previous modeling experiments are so rare that there is no basis for *a priori* estimation of parameter distributions. Thus, estimating *a priori* parameter distributions, which may significantly improve modeling, is an additional challenge to streamflow prediction in data-scarce regions.

1.1.1.3 Gauging the Ungauged Catchment

Supporting the streamflow predictions in an ungauged catchment through taking a few runoff measurements in the ungauged catchment itself is a topic that has been studied by researchers recently and proven promising in improving accuracy of predictions (Rojas-Serna et al., 2006; Perrin et al., 2008; Juston et al., 2009; Seibert et al., 2015; Westerberg et al., 2013; Drogue and Plasse, 2014). Perrin et al. (2007) studied 12 catchments in the United States for calibration using only a limited number of streamflow measurements. They found that 350 daily measurements randomly selected from a longer dataset including both wet and dry conditions were sufficient to provide robust estimations of parameter values. Seibert and Beven (2009) studied potential improvements to streamflow prediction within 11 catchments in a region of Sweden through constraining the HBV hydrological model using different subsets of observed streamflow in each catchment. For this purpose, randomly selected daily observations of 1, 2, 4, 8, 16, 32, 64, 128, and 256 days were tested in each year within a 10-year period. The results indicated that ensemble predictions by utilizing the weighted mean of simulations from acceptable parameter sets outperformed predictions generated through using only the best parameter set. It was also found that good results could be achieved using only little observed runoff data, however this varied significantly between the catchments and depending on the days chosen for measurements. Seibert and McDonnell (2013) further incorporated limited runoff measurements as well as soft data in the calibration process of a simple conceptual hydrological model in Maimai watershed in New Zealand. The authors found that constraining the model using 10 observations during high flow was on par with doing so using three months of continuously measured streamflow. Finally, Viviroli and Seibert (2015) studied the combination of parameter regionalization with limited runoff measurements for streamflow prediction in 49 catchments in Switzerland. The results indicated different behavior for catchments dominated with either or both snow melt and ice melt versus catchments dominated by rainfall. Modeling in the former showed significant improvements with only a couple of measurements during spring or summer while modeling in the latter showed only moderate improvements without any season being particularly suitable for taking measurements.

While these studies have introduced a new direction in order to improve streamflow predictions in ungauged catchments, some of the proposed methods do not account for a number of inherent limitations faced within data-scarce regions. These limitations include presence of only one partially similar reference catchment to the ungauged catchment(s) of interest as well as little knowledge about hydrograph behavior to choose event-based days for measurements. Moreover, in a real-world scenario it is likely that collection of measurements in the ungauged catchment of interest is only possible during a short period prior to implementation of modeling and generating the required streamflow data. These concerns justify a study that evaluates the potential of limited measurements in ungauged catchments for improvement of predictions in the context of data-scarce regions with their inherent limitations.

1.1.2 Flow Regime Alteration

Study of flow regime alteration and design of environmental flows, especially in remote small steep catchments of the world developed with projects such as diversion hydropower, requires robust flow prediction techniques. Thus, it provides a platform to further test the performance and applicability of the streamflow prediction approach proposed in this study. Flow regime of a river is important since it shapes the composition, structure, and functionality of aquatic, wetland, and riparian ecosystems (Poff and Ward, 1990; Richter et al., 1996; Bunn and Arthington, 2002; Scott et al., 2005). Unsurprisingly, many researchers have focused on the study of flow regime alterations in order to analyze and/or quantify the impacts of human interventions on riverine ecosystems (e.g. Doyle et al., 2005; Merritt and Poff., 2010; Poff and Zimmerman (2010) provide a review of such techniques). Among these studies, pioneering work by Poff et al. (1997) and Richter et al. (1996) led to frameworks for systematically quantifying flow regime alterations (e.g. Indicators of Hydrologic Alteration (IHA) (The Nature Conservancy website, 2017; Richter et. al, 1997; Mathews & Richter, 2007)). The IHA model is based on analysis of alterations of vital flow characteristics as a result of some change in a river such as human interventions. These vital flow characteristics are the five main characteristics of flow regime including magnitude, frequency, timing, duration and rate of change of flow (Poff et al. 1997). Accordingly, 32 ecologically relevant descriptors of flow are considered, each associated with one or more of the five vital flow characteristics. These indices are calculated for each year within pre- and post-impact periods, and their measures of central tendency and dispersion are determined to identify percent deviations from pre- to post-impact period. This provides quantitative indication of how much the flow regime of a river and its vital characteristics have changed as a result of a development project such as diversion hydropower.

One of the best-known human interventions in river flow regime is through dam construction and operation. The role of dams in altering river flow regime has been studied widely by researchers (Batalla et al. 2004; Chen et al., 2010; Taylor et al. 2014; Mwedzi et al. 2016; Sojka et al. 2016; Wang et al. 2016a; Timpe and Kaplan 2017; Ngor et al., 2018). Zhang et al. (2018) used the IHA "eco-flow" regime metrics to identify long term variations of inflow and outflow series in the Chaishitan Reservoir in China. The authors identified the impact patterns due to reservoir regulation and hypothesized that the patterns are variable over time. Rheinheimer and Viers (2015) utilized only four flow metrics to test flow alterations due to dam regulation under future climatic conditions in the Sierra Nevada. They found that dam regulation altered the flow regimes much more significantly than climate warming. Pringle et al. (2000) identified imperilment of migratory fish, imperilment of small-bodied riverine taxa, reduction and imperilment of taxa dependent on flooding or freshwater inflows to estuarine habitats, and increase in exotic and lentic-adapted species as the prominent examples of ecological impacts of flow alterations, mainly as a result of dams, in temperate and tropical regions of North and South America and the Caribbean. Poff et al. (2007) quantitatively showed that dams had homogenized the flow regimes of intermediate-sized (third- to seventh-order) rivers in 16 historically distinctive hydrologic regions in the United States.

Petts and Gurnell (2005) analyzed the consequences of flow alteration by dams through studying sediment transport and channel morphology, and reviewed the advancements made in this field.

1.1.2.1 Diversion Hydropower

Diversion hydropower (McManamay et al., 2016; Kibler and Alipour, 2017) is a type of hydropower generation project that requires diversion of water from a river into a hydropower generation station. Small to moderate, steep river catchments located in remote mountainous areas have been the primary location for diversion hydropower projects (Narula, 2012; Li et al., 2013; Tuna, 2013). Diversion dams or weirs are often built high in such catchments and create impoundments to withdraw water. Water then flows through low-gradient pipes or canals to a forebay. From forebay, water drops through pressurized, high-gradient penstocks to a power generation station. Water that exits the tailrace returns either to the river of its origin or to a different river (Fig. 1). Therefore, flow alterations are likely to happen in the reaches between the dam and tailrace (Kibler and Alipour, 2017). While flow regime alterations as a result of larger dams has been well studied and recognized, evaluation and quantification of the impacts of diversion hydropower on natural flow regime of smaller rivers has received only little attention from the research community (Kibler and Tullos, 2013; Kibler and Alipour, 2017). This is probably associated with the assumption that diversion hydropower projects have little to no impact on riverine flow regime and ecosystem (Gibeau et al., 2016). To prove this assumption potentially wrong, quantitative comparisons between large storage systems without diversion versus cumulative impacts of diversion systems have shown that there is potential for greater environmental impact from diversion systems (Gleick, 1992; Kibler and Tullos, 2013). Recent studies have also shed light on the fact that diversion hydropower projects can substantially alter the natural flow regime of the reaches that are located downstream of impoundment/diversion and endanger the riverine ecosystem (Kibler and Tullos, 2013; Anderson et al., 2015; Fantin-Cruz et al., 2015; Wang et al., 2016b).

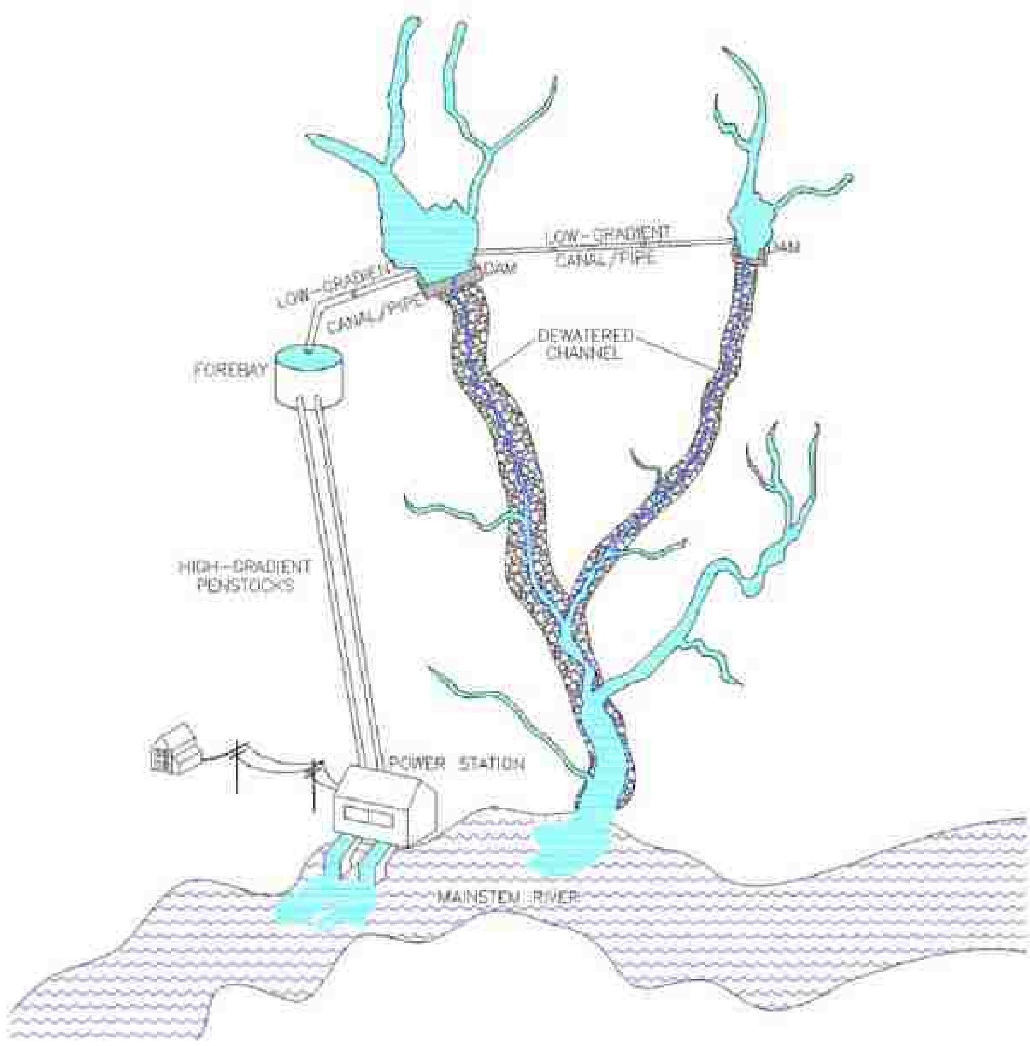

Figure 1. A schematic of diversion hydropower (Kibler and Alipour, 2017)

Diversion hydropower projects may often also be called run-of-river (ROR) hydropower. Anderson et al. (2015) reviewed the published literature on ROR to highlight the physical and ecological alterations caused by these projects. In terms of flow alteration, impacts mainly affected riverine habitat and connectivity. Reduction of lotic habitat, changes in habitat availability and water chemistry, and reduction in habitat complexity were among the reported impacts on riverine habitat while disruption of longitudinal connectivity and hindering upstream and downstream migration of fish species were among the reported impacts on connectivity. Another study by Fantin-Cruz et al. (2015) used the IHA model as well as flow duration curves to analyze flow regime alterations caused by a diversion hydropower facility on the Correntes River in Brazil. The authors found that besides changes in the seasonal regime, which was revealed by the flow duration curves, seven IHA metrics had significantly changed. Kibler and Alipour (2017) analyzed 32 developed rivers by such diversion hydropower projects in southwestern China. The results indicated that in terms of magnitude, low to moderate flows were highly altered across all rivers (e.g. mean annual flows decreased by a mean of 76 ± 12%). High flow duration decreased by a mean of 1.27 days across the rivers while low flow duration increased by a mean 8-fold degree from 3 days before alteration to 27 days after that. Flow constancy highly increased after diversion (a mean increase of 184 ± 49%) so that a static minimal flow replaced temporally dynamic low and moderate flows. The results also indicated sharper rates of change for flow after alteration as well as lower frequency of high flows.

Reliable flow prediction techniques in small steep ungauged catchments, suitable for diversion hydropower, can highly facilitate conduction of similar studies focused on the tradeoff between development and ecosystem function in other affected catchments. This at the same time can provide a great way to present the potential utilities of a flow prediction method and any improvements it might offer. For instance, Kibler and Alipour (2017) applied a catchment similarity model (Falkenmark & Chapman, 1989) to their 32 study rivers for streamflow prediction. Catchment similarity is based on the assumption that hydrologic routing processes are fully similar between a gauged reference catchment and ungauged target catchment(s) of interest. Clearly, streamflows predicted under this assumption may be associated with high levels of uncertainty in particular when you perform the analysis in a remote region where the choice of gauged reference catchments is limited to only one partially similar catchment. We noted before that many diversion hydropower projects are implemented in small, steep catchments located in remote mountainous areas, which are not equipped with flow measurement devices (Narula, 2012; Li et al., 2013; Tuna, 2013). This makes availability of similar gauged reference catchments very limited within these regions and consequently flow prediction and analysis of flow regime alterations becomes highly difficult. In addition, other hydrometeorological data such as precipitation and temperature are often scarcely measured within these remote regions as well. Therefore, a streamflow prediction approach suited to such severely data-scarce regions is required to enable the study of flow regime alterations due to diversion hydropower and may highly improve the accuracy of such studies.

1.2 References

Ajami, N.K., Gupta, H., Wagener, T., Sorooshian, S., 2004. Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system. Journal of Hydrology, 298(1), pp.112-135.

Ajmal, M., Waseem, M., Wi, S., Kim, T.W., 2015. Evolution of a parsimonious rainfall–runoff model using soil moisture proxies. Journal of Hydrology, 530, 623-633.

Anderson, D., Moggridge, H., Warren, P., Shucksmith, J., 2015. The impacts of 'run-of-river' hydropower on the physical and ecological condition of rivers. Water and Environment Journal, 29(2), pp.268-276.

Andréassian, V., Le Moine, N., Perrin, C., Ramos, M.H., Oudin, L., Mathevet, T., Lerat, J., Berthet, L., 2012. All that glitters is not gold: the case of calibrating hydrological models. Hydrological Processes, 26 (14), 2206-2210.

Archibald, J.A., Buchanan, B.P., Fuka, D.R., Georgakakos, C.B., Lyon, S.W., Walter, M.T., 2014. A simple, regionally parameterized model for predicting nonpoint source areas in the northeastern US. Journal of Hydrology: Regional Studies, 1, pp.74-91.

Arnold, J.G., Youssef, M.A., Yen, H., White, M.J., Sheshukov, A.Y., Sadeghi, A.M., Moriasi, D.N., Steiner, J.L., Amatya, D.M., Skaggs, R.W., Haney, E.B., 2015. Hydrological processes and model representation: Impact of soft data on calibration. Transactions of the ASABE, 58(6), pp.1637-1660.

Arsenault, R., Brissette, F.P., 2014. Continuous streamflow prediction in ungauged basins: The effects of equifinality and parameter set selection on uncertainty in regionalization approaches. Water Resources Research, 50 (7), 6135-6153.

Bárdossy, A., Huang, Y., Wagener, T., 2016. Simultaneous calibration of hydrological models in geographical space. Hydrology and Earth System Sciences, 20 (7), 2913-2928.

Batalla, R.J., Gomez, C.M., Kondolf, G.M., 2004. River impoundment and changes in flow regime, Ebro River basin, northeastern Spain. Journal of Hydrology, 290, pp.117-136.

Bekele, E.G., Nicklow, J.W., 2007. Multi-objective automatic calibration of SWAT using NSGA-II. Journal of Hydrology, 341(3), pp.165-176.

Bergström, S., 1976. Development and application of a conceptual runoff model for Scandinavian catchments. Report RHO 7, Swedish Meteorological and Hydrological Institute, Norrköping, Sweden.

Beven, K., 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. Advances in water resources, 16 (1), 41-51.

Beven, K., 2006. A manifesto for the equifinality thesis. Journal of hydrology, 320 (1), 18-36.

Blöschl, G., Sivapalan, M., Savenije, H., Wagener, T., Viglione, A. eds., 2013. Runoff prediction in ungauged basins: synthesis across processes, places and scales. Cambridge University Press.

Brigode, P., Paquet, E., Bernardara, P., Gailhard, J., Garavaglia, F., Ribstein, P., Bourgin, F., Perrin, C., Andréassian, V., 2015. Dependence of model-based extreme flood estimation on the calibration period: case study of the Kamp River (Austria). Hydrological sciences journal, 60(7-8), pp.1424-1437.

Bunn, S.E., Arthington, A.H., 2002. Basic principles and ecological consequences of altered flow regimes for aquatic biodiversity. Environmental Management, 30(4), 492-507.

Chen, Y.D., Yang, T., Xu, C.Y., Zhang, Q., Chen, X., Hao, Z.C., 2010. Hydrologic alteration along the Middle and Upper East River (Dongjiang) basin, South China: a visually enhanced mining on the results of RVA method. Stochastic Environmental Research and Risk Assessment, 24(1), pp.9-18.

Confesor, R.B., Whittaker, G.W., 2007. Automatic Calibration of Hydrologic Models With Multi-Objective Evolutionary Algorithm and Pareto Optimization. JAWRA Journal of the American Water Resources Association, 43(4), 981-989.

Doyle, M.W., Stanley, E.H., Strayer, D.L., Jacobson, R.B., Schmidt, J.C., 2005. Effective discharge analysis of ecological processes in streams. Water Resources Research, 41(11).

Drogue, G.P., Plasse, J., 2014. How can a few streamflow measurements help to predict daily hydrographs at almost ungauged sites? Hydrological Sciences Journal, 59(12), pp.2126-2142.

Efstratiadis, A., Koutsoyiannis, D., 2010. One decade of multi-objective calibration approaches in hydrological modelling: a review. Hydrological Sciences Journal, 55(1), pp.58-78.

Eng, K., Stedinger, J.R., Gruber, A. M., 2007. Regionalization of streamflow characteristics for the Gulf-Atlantic rolling plains using leverage-guided region-of-influence regression. EWRI World Water & Environmental Resources Congress, American Society of Civil Engineers.

Falkenmark, M., Chapman, T., (eds.), 1989. Comparative hydrology: an ecological approach to land and water resources. UNESCO, Paris, France.

Fantin-Cruz, I., Pedrollo, O., Girard, P., Zeilhofer, P., Hamilton, S. K., 2015. Effects of a diversion hydropower facility on the hydrological regime of the Correntes River, a tributary to the Pantanal floodplain, Brazil. Journal of Hydrology, 531, 810–820.

Fekete, B.M., Vörösmarty, C.J., 2007. The current status of global river discharge monitoring and potential new technologies complementing traditional discharge measurements. In

Predictions in Ungauged Basins: PUB kick-off (Proceedings of the PUB kick-off meeting held in Brasilia, 20–22 November 2002). IAHS Publication (Vol. 349).

Fekete, B.M., Looser, U., Pietroniro, A., Robarts, R.D., 2012. Rationale for monitoring discharge on the ground. Journal of Hydrometeorology, 13 (6), 1977-1986.

Fenicia, F., Savenije, H.H., Matgen, P., Pfister, L., 2007. A comparison of alternative multiobjective calibration strategies for hydrological modeling. Water Resources Research, 43(3).

Ford, W.I., Fox, J.F., Pollock, E., 2017. Reducing equifinality using isotopes in a process-based stream nitrogen model highlights the flux of algal nitrogen from agricultural streams. Water Resources Research, 53(8), pp.6539-6561.

Fortin, V., 2000. Le modèle météo-apport HSAMI: historique, théorie et application. Institut de recherche d'Hydro-Québec, Varennes, 68.

Fry, L.M., Hunter, T.S., Phanikumar, M.S., Fortin, V., Gronewold, A.D., 2013. Identifying streamgage networks for maximizing the effectiveness of regional water balance modeling. Water Resources Research, 49(5), pp.2689-2700.

Gelfan, A., Motovilov, Y., Krylenko, I., Moreido, V., Zakharova, E., 2015. Testing the robustness of the physically-based ECOMAG model with respect to changing conditions. Hydrological Sciences Journal, 60 (7-8), pp.1266-1285.

Gharari, S., Hrachowitz, M., Fenicia, F., Savenije, H.H.G., 2013. An approach to identify time consistent model parameters: sub-period calibration. Hydrology and Earth System Sciences, 17(1), pp.149-161.

Gibeau, P., Connors, B.M., Palen, W.J., 2016. Run-of-River hydropower and salmonids: potential effects and perspective on future research. Canadian Journal of Fisheries and Aquatic Sciences, 74(7), pp.1135-1149.

Gleick, P.H., 1992. Environmental consequences of hydroelectric development: the role of facility size and type. Energy, 17(8), pp.735-747.

He, Y., Bárdossy, A., Zehe, E., 2011. A review of regionalisation for continuous streamflow simulation. Hydrology and Earth System Sciences, 15(11), p.3539.

Heřmanovský, M., Havlíček, V., Hanel, M., Pech, P., 2017. Regionalization of runoff models derived by genetic programming. Journal of Hydrology, 547, 544-556.

Hrachowitz, M., Savenije, H.H.G., Blöschl, G., McDonnell, J.J., Sivapalan, M., Pomeroy, J.W., Arheimer, B., Blume, T., Clark, M.P., Ehret, U., Fenicia, F., Freer, J.E., Gelfan, A., Gupta, H.V., Hughes, D.A., Hut, R.W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P.A., Uhlenbrook, S., Wagener, T., Winsemius, H.C., Woods, R.A., Zehe, E., Cudennec, C., 2013. A decade of Predictions in Ungauged Basins (PUB)—a review. Hydrological sciences journal, 58 (6), 1198-1255.

Hughes, D.A., 2016. Hydrological modelling, process understanding and uncertainty in a southern African context: lessons from the northern hemisphere. Hydrological Processes, 30(14), pp.2419-2431.

Ibrahim, B., Wisser, D., Barry, B., Fowe, T., Aduna, A., 2015. Hydrological predictions for small ungauged watersheds in the Sudanian zone of the Volta basin in West Africa. Journal of Hydrology: Regional Studies, 4, 386-397.

Juston, J., Seibert, J., Johansson, P.O., 2009. Temporal sampling strategies and uncertainty in calibrating a conceptual hydrological model for a small boreal catchment. Hydrological Processes: An International Journal, 23(21), pp.3093-3109.

Kelleher, C., McGlynn, B., Wagener, T., 2017. Characterizing and reducing equifinality by constraining a distributed catchment model with regional signatures, local observations, and process understanding. Hydrology and Earth System Sciences, 21(7), p.3325.

Khu, S.T., Madsen, H., Di Pierro, F., 2008. Incorporating multiple observations for distributed hydrologic model calibration: An approach using a multi-objective evolutionary algorithm and clustering. Advances in Water Resources, 31(10), pp.1387-1398.

Kibler, K.M., Tullos, D.D., 2013. Cumulative biophysical impact of small and large hydropower development in Nu River, China. Water Resources Research 49 (6), 3104–3118.

Kibler, K.M., Alipour, M.H., 2017. Flow alteration signatures of diversion hydropower: An analysis of 32 rivers in southwestern China. Ecohydrology, 10 (5).

Kamali, B., Mousavi, S.J., 2014. Automatic Calibration of HEC-HMS Model Using Multi-Objective Fuzzy Optimal Models. Civil Engineering Infrastructures Journal, 47(1), pp.1-12.

Kirchner, J.W., 2006. Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. Water Resources Research, 42 (3).

Kult, J.M., Fry, L.M., Gronewold, A.D., Choi, W., 2014. Regionalization of hydrologic response in the Great Lakes basin: Considerations of temporal scales of analysis. Journal of Hydrology, 519, 2224-2237.

Kundzewicz, Z. W., 1997. Water resources for sustainable development. Hydrological Sciences Journal, 42 (4), pp. 467-480.

Le, V.T., Nguyen, L.T., 2018. Multi-objective calibration of conceptual rainfall-runoff model in Thac Mo Catchment, Vietnam. International Symposium on Lowland Technology, Hanoi, Vietnam.

Lebecherel, L., Andréassian, V., Perrin, C., 2016. On evaluating the robustness of spatial-proximity-based regionalization methods. Journal of Hydrology, 539, pp.196-203.

Li, B.S., Zhou, P.J., Wang, X.Y., Zhu, L., 2013. Opportunities and eco-environmental influence of cascade hydropower development and water diversion projects in Hanjiang river basin. Journal of the Geological Society of India, 82(6), pp.692-700.

Lüdtke, S., Apel, H., Nied, M., Carl, P., Merz, B., 2014. Reducing equifinality of hydrological models by integrating Functional Streamflow Disaggregation. In EGU General Assembly Conference Abstracts, Vol. 16.

Madsen, H., 2000. Automatic calibration of a conceptual rainfall–runoff model using multiple objectives. Journal of hydrology, 235(3), pp.276-288.

Mathews, R., Richter, B.D., 2007. Application of the indicators of hydrologic alteration software in environmental flow setting. Journal of the American Water Resources Association 43(6), 1400-1413.

Matos, J.P., Cohen, T., Boillat, J.L., Schleiss, A.J., Portela, M.M., 2010. Analysis of flow regime changes due to operation of large reservoirs on the Zambezi River. In Environmental Hydraulics-Proceedings of the 6th International Symposium on Environmental Hydraulics, Vol. 6, pp. 337-342.

McIntyre, N., Lee, H., Wheater, H., Young, A., 2004. Tools and approaches for evaluating uncertainty in streamflow predictions in ungauged UK catchments. International Congress on Environmental Modelling and Software, Osnabrück, Germany.

McIntyre, N., Lee, H., Wheater, H., Young, A., Wagener, T., 2005. Ensemble predictions of runoff in ungauged catchments. Water Resources Research, 41(12).

McManamay, R.A., Oigbokie, C.O., Kao, S.C., Bevelhimer, M.S., 2016. Classification of US hydropower dams by their modes of operation. River Research and Applications, 32(7), pp.1450-1468.

Melching C., 1995. Reliability estimation. In: Singh VP, editor. Computer models of watershed hydrology. Colorado, USA: Water Resource Publications, p. 69–118.

Merritt, D.M., Poff, N., 2010. Shifting dominance of riparian Populus and Tamarix along gradients of flow alteration in western North American rivers. Ecological Applications, 20(1), pp.135-152.

Merz, R., Blöschl, G., 2004. Regionalisation of catchment model parameters. Journal of Hydrology, 287(1), pp.95-123.

Merz, R., Parajka, J., Blöschl, G., 2009. Scale effects in conceptual hydrological modeling. Water Resources Research, 45 (9).

Merz, R., Parajka, J., Blöschl, G., 2011. Time stability of catchment model parameters: Implications for climate impact analyses. Water Resources Research, 47 (2).

Mishra, A.K., Coulibaly, P., 2009. Developments in hydrometric network design: A review. Reviews of Geophysics, 47 (2).

Muleta, M.K., Nicklow, J.W., 2005. Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. Journal of Hydrology, 306(1), pp.127-145.

Mwedzi, T., Katiyo, L., Mugabe, F.T., Bere, T., Bangira, C., Mangadze, T., Kupika, O.L., 2016. A spatial assessment of stream-flow characteristics and hydrologic alterations, post dam construction in the Manyame catchment, Zimbabwe. Water SA, 42(2), pp.194-202.

Narula, K.K., Gosain, A.K., Nischal, N., 2012. Modeling streamflows for estimating hydrological PLF of a small hydropower scheme in an ungauged mountainous watershed in Western Ghats in India using SWAT. SWAT Conference.

Ngor, P.B., Legendre, P., Oberdorff, T., Lek, S., 2018. Flow alterations by dams shaped fish assemblage dynamics in the complex Mekong-3S river system. Ecological Indicators, 88, pp.103-114.

Oudin, L., Andréassian, V., Perrin, C., Michel, C., Le Moine, N., 2008. Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments. Water Resources Research, 44 (3).

Parajka, J., Merz, R., and Blöschl, G., 2005. A comparison of regionalisation methods for catchment model parameters. Hydrology and Earth System Sciences Discussions, 9(3), pp.157-171.

Parajka, J., Blöschl, G., Merz, R., 2007a. Regional calibration of catchment models: Potential for ungauged catchments. Water Resources Research, 43 (6).

Parajka, J., Merz, R., Blöschl, G., 2007b. Uncertainty and multiple objective calibration in regional water balance modelling: case study in 320 Austrian catchments. Hydrological Processes, 21 (4), 435-446.

Parajka, J., Merz, R., Skøien, J. O., Viglione, A., 2015. The role of station density for predicting daily runoff by top-kriging interpolation in Austria. Journal of Hydrology and Hydromechanics, 63 (3), 228-234.

Peñas Silva, F.J., Barquín Ortiz, J., Snelder, T.H., Booker, D.J., Álvarez Díaz, C., 2014. The influence of methodological procedures on hydrological classification performance. Hydrology and Earth System Sciences, 18, 3393-3409.

Perks, A., Winkler, T., and Stewart, B., 1996. The adequacy of hydrological networks: A global assessment, Rep. WMO-740, World Meteorological Organization, Geneva, Switzerland.

Perrin, C., Oudin, L., Andreassian, V., Rojas-Serna, C., Michel, C., Mathevet, T., 2007. Impact of limited streamflow data on the efficiency and the parameters of rainfall—runoff models. Hydrological sciences journal, 52(1), pp.131-151.

Perrin, C., Andréassian, V., Rojas Serna, C., Mathevet, T., Le Moine, N., 2008. Discrete parameterization of hydrological models: Evaluating the use of parameter sets libraries over 900 catchments. Water Resources Research, 44(8).

Petts, G.E., Gurnell, A.M., 2005. Dams and geomorphology: research progress and future directions. Geomorphology, 71(1), pp.27-47.

Poff, N.L., Ward, J.V., 1990. Physical habitat template of lotic systems: recovery in the context of historical pattern of spatiotemporal heterogeneity. Environmental management, 14(5), p.629.

Poff, N.L., Allan, J.D., Bain, M.B., Karr, J.R., Prestegaard, K.L., Richter, B.D., Sparks, R.E., Stromberg, J.C., 1997. The natural flow regime. BioScience, 47(11), pp.769-784.

Poff, N.L., Olden, J.D., Merritt, D.M., Pepin, D.M., 2007. Homogenization of regional river dynamics by dams and global biodiversity implications. Proceedings of the National Academy of Sciences, 104(14), pp.5732-5737.

Poff, N.L., Zimmerman, J.K., 2010. Ecological responses to altered flow regimes: a literature review to inform the science and management of environmental flows. Freshwater Biology, 55(1), pp.194-205.

Poissant, D., Arsenault, R., Brissette, F., 2017. Impact of parameter set dimensionality and calibration procedures on streamflow prediction at ungauged catchments. Journal of Hydrology: Regional Studies, 12, pp.220-237.

Post, D.A., 2009. Regionalizing rainfall–runoff model parameters to predict the daily streamflow of ungauged catchments in the dry tropics. Hydrology Research, 40 (5), pp.433-444.

Pringle, C.M., Freeman, M.C., Freeman, B.J., 2000. Regional Effects of Hydrologic Alterations on Riverine Macrobiota in the New World: Tropical-Temperate Comparisons. BioScience, 50(9), 807–823.

Rajib, M.A., Merwade, V., Yu, Z., 2016. Multi-objective calibration of a hydrologic model using spatially distributed remotely sensed/in-situ soil moisture. Journal of Hydrology, 536, 192-207.

Razavi, T., Coulibaly, P., 2012. Streamflow prediction in ungauged basins: review of regionalization methods. Journal of hydrologic engineering, 18(8), pp.958-975.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., Franks, S.W., 2010. Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. Water Resources Research, 46(5).

Rheinheimer, D.E., Viers, J.H., 2015. Combined effects of reservoir operations and climate warming on the flow regime of hydropower bypass reaches of California's Sierra Nevada. River research and applications, 31(3), pp.269-279.

Richter, B.D., Baumgartner, J.V., Powell, J., Braun, D.P., 1996. A method for assessing hydrologic alteration within ecosystems. Conservation biology, 10(4), pp.1163-1174.

Richter, B.D., Baumgartner, J.V., Wigington, R., Braun, D.P., 1997. How much water does a river need? Freshwater biology, 37(1), 231-249.

Rinderer, M., Seibert, J., 2012. Soil information in hydrologic models: Hard data, soft data, and the dialog between experimentalists and modelers. In: Lin, H., Hydropedology: Synergistic Integration of Soil Science and Hydrology. Waltham: Elsevier, 515-536.

Rojas-Serna, C., Michel, C., Perrin, C., Andreassian, V., Hall, A., Chahinian, N., Schaake, J., 2006. Ungauged catchments: how to make the most of a few streamflow measurements? IAHS publication, 307, p.230.

Samaniego, L., Kumar, R., Attinger, S., 2010. Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. Water Resources Research, 46(5).

Samuel, J., Coulibaly, P., Metcalfe, R.A., 2011. Estimation of continuous streamflow in Ontario ungauged basins: comparison of regionalization methods. Journal of Hydrologic Engineering, 16(5), 447-459.

Savenije, H.H., 2001. Equifinality, a blessing in disguise? Hydrological processes, 15 (14), 2835-2838.

Scott, M.L., Brasher, A.M., Reynolds, E.W., Caires, A., Miller, M.E., Thomas, L.P., Garman, S., 2005. The structure and functioning of riparian and aquatic ecosystems of the Colorado Plateau–conceptual models to inform monitoring. US Geological Survey.

Seibert, J., McDonnell, J.J., 2002. On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration. Water Resources Research, 38(11).

Seibert, J. and Beven, K.J., 2009. Gauging the ungauged basin: how many discharge measurements are needed? Hydrology and Earth System Sciences, 13(6), pp.883-892.

Seibert, J., McDonnell, J.J., 2013. Gauging the ungauged basin: relative value of soft and hard data. Journal of hydrologic engineering, 20(1), p.A4014004.

Seibert, J., Vis, M., Pool, S., 2013. Gauging the Ungauged Catchment: the Value of Stream Level Observations. In AGU Fall Meeting Abstracts.

Shafii, M., Smedt, F.D., 2009. Multi-objective calibration of a distributed hydrological model (WetSpa) using a genetic algorithm. Hydrology and Earth System Sciences, 13(11), pp.2137-2149.

Shrestha, R.R., Rode, M., 2008. Multi-objective calibration and fuzzy preference selection of a distributed hydrological model. Environmental Modelling & Software, 23(12), pp.1384-1395.

Skaugen, T., Peerebom, I.O., Nilsson, A., 2015. Use of a parsimonious rainfall–run-off model for predicting hydrological response in ungauged basins. Hydrological Processes, 29 (8), 1999-2013.

Sojka, M., Jaskuła, J., Wicher-Dysarz, J., Dysarz, T., 2016. Assessment of dam construction impact on hydrological regime changes in lowland river–A case of study: the Stare Miasto reservoir located on the Powa River. Journal of Water and Land Development, 30(1), pp.119-125.

Song, J., Xia, J., Zhang, L., Wang, Z.H., Wan, H., She, D., 2016. Streamflow prediction in ungauged basins by regressive regionalization: a case study in Huai River Basin, China. Hydrology Research, 47(5), pp.1053-1068.

Spence, C., Burn, D.H., Davison, B., Hutchinson, D., Ouarda, T.B., St-Hilaire, A., Weber, F., Whitfield, P.H., 2013. A Canadian viewpoint on data, information and uncertainty in the context of prediction in ungauged basins. Hydrology Research, 44(3), pp.419-429.

Swain, J.B., Patra, K.C., 2017. Streamflow estimation in ungauged catchments using regionalization techniques. Journal of Hydrology, 554, pp.420-433.

Taylor, J.M., Seilheimer, T.S., Fisher, W.L., 2014. Downstream fish assemblage response to river impoundment varies with degree of hydrologic alteration. Hydrobiologia, 728(1), pp.23-39.

Tegegne, G., Park, D.K., Kim, Y.O., 2017. Comparison of hydrological models for the assessment of water resources in a data-scarce region, the Upper Blue Nile River Basin. Journal of Hydrology: Regional Studies, 14, pp.49-66.

The Nature Conservancy website, 2017. Available online at: www.conservationgateway.org

Timpe, K., Kaplan, D., 2017. The changing hydrology of a dammed Amazon. Science advances, 3(11), p.e1700611.

Tuna, M.C., 2013. Feasibility assessment of hydroelectric power plant in ungauged river basin: a case study. Arabian Journal for Science and Engineering, 38(6), pp.1359-1367.

USACE, 2008. Hydrologic modeling system HEC-HMS applications guide. US. Army Corps of Engineers, Hydrologic Engineering Center (HEC), Washington, DC.

Vandewiele, G.L., Elias, A., 1995. Monthly water balance of ungauged catchments obtained by geographical regionalization. Journal of Hydrology, 170(1-4), pp.277-291.

van Emmerik, T., Mulder, G., Eilander, D., Piet, M., Savenije, H., 2015. Predicting the ungauged basin: Model validation and realism assessment. Frontiers in Earth Science, 3 (62).

Viviroli, D., Seibert, J., 2015. Can a regionalized model parameterisation be improved with a limited number of runoff measurements? Journal of Hydrology, 529, pp.49-61.

Wagener, T., Sivapalan, M., Troch, P., Woods, R., 2007. Catchment classification and hydrologic similarity. Geography compass, 1(4), pp.901-931.

Wagener, T., Blöschl, G., Goodrich, D.C., Gupta, H., Sivapalan, M., Tachikawa, Y., Troch, P.A., Weiler, M., 2013. A synthesis framework for runoff prediction in ungauged basins. chapt, 2, pp.11-28.

Wang, G.Q., Zhang, J.Y., Jin, J.L., Pagano, T.C., Calow, R., Bao, Z.X., Liu, C.S., Liu, Y.L., Yan, X.L., 2012. Assessing water resources in China using PRECIS projections and a VIC model. Hydrology and Earth System Sciences, 16(1), pp.231-240.

Wang, Y., Brubaker, K., 2015. Multi-objective model auto-calibration and reduced parameterization: Exploiting gradient-based optimization tool for a hydrologic model. Environmental Modelling & Software, 70, pp.1-15.

Wang, Y., Rhoads, B.L., Wang, D., 2016a. Assessment of the flow regime alterations in the middle reach of the Yangtze River associated with dam construction: potential ecological implications. Hydrological Processes, 30(21), pp.3949-3966.

Wang, H., Chen, Y., Liu, Z., Zhu, D., 2016b. Effects of the “Run-of-River” Hydro Scheme on Macroinvertebrate Communities and Habitat Conditions in a Mountain River of Northeastern China. Water, 8(1), p.31.

Westerberg, I., Seibert, J., Beven, K., 2013, April. Gauging the ungauged catchment: a flow-duration curve approach. In EGU General Assembly Conference Abstracts (Vol. 15).

World Meteorological Organization, 2008. Guide to Hydrological Practices Volume I, Hydrology – From Measurement to Hydrological Information. WMO-No. 168, Sixth edition.

Yadav, M., Wagener, T., Gupta, H., 2007. Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. Advances in Water Resources, 30 (8), pp. 1756-1774.

Yang, T.C., Yu, P.S., Kuo, C.M., Wang, Y.C., 2004. Application of fuzzy multiobjective function on storm-event rainfall-runoff model calibration. Journal of Hydrologic Engineering, 9(5), pp.440-445.

Yapo, P.O., 1996. A multiobjective global optimization algorithm with application to calibration of hydrologic models. PhD dissertation, University of Arizona, Tucson.

Yapo, P.O., Gupta, H.V., Sorooshian, S., 1998. Multi-objective global optimization for hydrologic models. Journal of hydrology, 204(1-4), pp.83-97.

Yu, P.S., Yang, T.C., 2000. Fuzzy multi-objective function for rainfall-runoff model calibration. Journal of hydrology, 238(1), pp.1-14.

Zhang, Z., Wagener, T., Reed, P., Bhushan, R., 2008. Reducing uncertainty in predictions in ungauged basins by combining hydrologic indices regionalization and multiobjective optimization. Water Resources Research, 44 (12).

Zhang, J.L., Li, Y.P., Huang, G.H., Wang, C.X., Cheng, G.H., 2016. Evaluation of Uncertainties in Input Data and Parameters of a Hydrological Model Using a Bayesian Framework: A Case Study of a Snowmelt–Precipitation-Driven Watershed. Journal of Hydrometeorology, 17(8), pp.2333-2350.

Zhang, Y., Zhai, X., Zhao, T., 2018. Annual shifts of flow regime alteration: new insights from the Chaishitan Reservoir in China. Scientific reports, 8(1), p.1414.

Zhou, J., Ouyang, S., Wang, X., Ye, L., Wang, H., 2014. Multi-objective parameter calibration and multi-attribute decision-making: an application to conceptual hydrological model calibration. Water resources management, 28(3), pp.767-783.

CHAPTER 2:
A FRAMEWORK FOR STREAMFLOW PREDICTION IN THE WORLD'S MOST SEVERELY DATA-LIMITED REGIONS: TEST OF APPLICABILITY AND PERFORMANCE IN A POORLY-GAUGED REGION OF CHINA

2.1 Preface

This chapter describes development of Streamflow Prediction in Extreme Data-scarcity (SPED) framework and test of its applicability in two catchments located in southwestern China. The content of this chapter has been published in Journal of Hydrology[1].

2.2 Abstract

A framework methodology is proposed for streamflow prediction in poorly-gauged rivers located within large-scale regions of sparse hydrometeorologic observation. A multi-criteria model evaluation is developed to select models that balance runoff efficiency with selection of accurate parameter values. Sparse observed data are supplemented by uncertain or low-resolution information, incorporated as 'soft' data, to estimate parameter values *a priori*. Model performance is tested in two catchments within a data-poor region of southwestern China, and results are compared to models selected using alternative calibration methods. While all models perform consistently with respect to runoff efficiency (NSE range of 0.67 - 0.78), models selected using the proposed multi-objective method may incorporate more representative parameter values than those selected by traditional calibration. Notably, parameter values estimated by the proposed

[1] Alipour, M.H., Kibler, K.M., 2018. A framework for streamflow prediction in the world's most severely data-limited regions: Test of applicability and performance in a poorly-gauged region of China. Journal of Hydrology, 557, pp.41-54.

method resonate with direct estimates of catchment subsurface storage capacity (parameter residuals of 20 and 61 mm for maximum soil moisture capacity (C_{max}), and 0.91 and 0.48 for soil moisture distribution shape factor (B); where a parameter residual is equal to the centroid of a soft parameter value minus the calibrated parameter value). A model more traditionally calibrated to observed data only (single-objective model) estimates a much lower soil moisture capacity (residuals of C_{max} = 475 and 518 mm and B = 1.24 and 0.7). A constrained single-objective model also underestimates maximum soil moisture capacity relative to *a priori* estimates (residuals of C_{max} = 246 and 289 mm). The proposed method may allow managers to more confidently transfer calibrated models to ungauged catchments for streamflow predictions, even in the world's most data-limited regions.

2.3 Introduction

Prediction of streamflow in ungauged catchments is surrounded by uncertainty, and is thus a challenging, yet vital task for water managers. Substantial advancements to the science of flow prediction have been made, for instance through initiatives of the International Association of Hydrological Sciences Prediction in Ungauged Basins (PUB) Decade 2003-2012. The emergent framework synthesized from the first PUB Decade acknowledges catchment complexity and diversity, as well as hydrological response signatures, and highlights the importance of comparative hydrology for prediction in ungauged basins (see Blöschl et al., 2013). While this forms a strong foundation upon which practical applications may develop, prescriptive methods to bridge research advancements and practice may make such advances more accessible to water managers. Specific approaches to avoid and address pervasive flow prediction pitfalls such as over-calibration of models (Andréassian et al., 2012; Kirchner, 2006) and equifinality (Beven, 1993, 2006) are needed. Without such methods, ad hoc and less comprehensive methods may be adopted in practice (Efstratiadis et al., 2014; van Emmerik et al., 2015), especially in data-limited regions, where simple water balance methods may outperform rainfall-runoff models (Chiew, 2010).

Many flow prediction approaches utilize regional hydrologic networks to estimate streamflow in ungauged rivers. Such regionalization of model parameters has been applied widely and proven to be an efficient method for prediction over large spatial areas (e.g. Ibrahim et al., 2015; Parajka et al., 2007a; Wang et al., 2012). For instance, Arsenault and Brissette (2014) compared three common regionalization techniques for predicting streamflow in 268 basins in Quebec, finding that approaches based on physical similarity performed the best. Kult et al. (2014) then modeled the relationship between physical characteristics and hydrologic response in 163 watersheds in the Great Lakes basin through regionalization based on multiple linear regression and regression tree analysis. The authors achieved satisfactory model performance (median NSE of 0.53) for their 62 validation watersheds without use of a rainfall–runoff model. Despite the promise of regionalization techniques, density of available gauging stations in the region surrounding a particular catchment of interest (target catchment) influences reliability (Oudin et al., 2008; Parajka et al., 2015). Thus, methods based on regionalization may be limited in areas where densities of climatic/hydrologic monitoring are (far) below the recommendations of World Meteorological Organization (WMO, 2008). Such data sparsity is in fact directly associated with introduction of higher levels of catchment dissimilarity because it limits available choices for similar reference catchment(s) to target catchment(s). An alternative modeling approach in such cases must rely on the existing data, including regional/global data, and must be robust to the extensive uncertainty inherent in sparse and low-resolution data as well as some level of catchment dissimilarity.

Multi-objective analysis is a well-established approach to address and lower prediction uncertainties and develop a more realistic modeling of hydrologic mechanisms (see Efstratiadis and Koutsoyiannis (2010) for a review of such techniques). Multi-objective evaluation has been incorporated into flow prediction approaches such as regionalization (e.g. Zhang et al., 2008). Yapo et al. (1998) developed an algorithm to solve global optimization problems for multi-objective models based on watershed output fluxes. Later, Confesor and Whittaker (2007) used multi-objective modeling for automatic calibration of the Soil and Water Assessment Tool (SWAT) to generate daily streamflows in the Calapooia River in Oregon. SWAT was also calibrated in two watersheds in Indiana, by Rajib et al. (2016), using multiple objectives based on streamflow observations as well as remotely-sensed surface soil moisture data and root-zone soil moisture estimates from limited field sensor data. Merz et al. (2009, 2011) introduced a multi-objective calibration technique, based on constraining model parameters to pre-defined Beta distribution functions. Despite all advances in this field, extensive observed data has normally been an important requirement of such techniques; methods for applying such techniques to predict streamflow within regions of minimal data are lacking. For example, the approach proposed by Merz et al. (2009, 2011) may not be applicable or reliable where observed data and previous local modeling experiments provide insufficient information to estimate *a priori* parameter distribution function.

To address the aforementioned challenges, we propose a prescriptive framework methodology based on multi-objective modeling to predict streamflows in data-limited regions of the world. Accordingly, our research objectives are to 1) develop a flexible methodology for prediction in ungauged catchments within data-scarce regions, 2) undertake a realistic test of performance by generating streamflow data in two genuine poorly-gauged catchments embedded within a large sparsely-gauged region of southwestern China, and 3) test performance of our proposed method as compared to models selected through traditional single-objective and constrained single-objective calibration.

The proposed method targets the most data limited regions of the world, but can in principle be applied to any region. However, it may not offer advantage over existing well-established methods in data-abundant regions. The novelty of this work thus lies in three main factors:

1) The proposed multi-objective framework creates a platform to apply advanced conceptual hydrological models to ungauged catchments within severely data-limited regions, where existing methods requiring robust data are often inapplicable.

2) The proposed method is designed to be highly flexible and potentially applicable to a wide range of catchments and regions. We do not propose a new hydrologic model; rather our proposed method is designed to wrap around a manager's model of choice, allowing practitioners to select the best model for their location and objective. The approach is also designed to be robust to some level of catchment dissimilarity, providing flexibility for use in regions where choice of gauged catchments is limited and dissimilarity is inevitable.

3) Built upon the proposed framework, we develop and test models in a remote area of southwestern China. The experimental design is based on globally-available subsurface data, regional climatic data, and a rainfall-runoff model. While flow prediction methodologies are often developed in well-gauged regions, assuming they are ungauged, we develop and test our methods within a truly data-limited region of the world where gauging stations are rare and sparse. Thus, we demonstrate that the proposed methods may be suitable for practitioners in similar areas.

2.4 Methodology

2.4.1 A Framework for Flow Prediction in Severely Data-limited Regions

The first part of our proposed framework, preliminary model testing (Fig. 2a), consists of traditional single-objective (i.e. maximize runoff efficiency) model calibration and validation in a gauged reference catchment. Model simplicity and avoidance of over-parameterization should be prioritized in model selection (Kirchner, 2006). Sparse observations of forcing data may be supplemented with global or regional databases, though bias correction of meteorological data may be necessary (e.g. Wi et al., 2015). After a hydrologic model has been selected according to the practitioner's preference, the model is calibrated and validated in the reference catchment. Such preliminary testing verifies the suitability of the underlying perceptual/conceptual model for runoff prediction in the region.

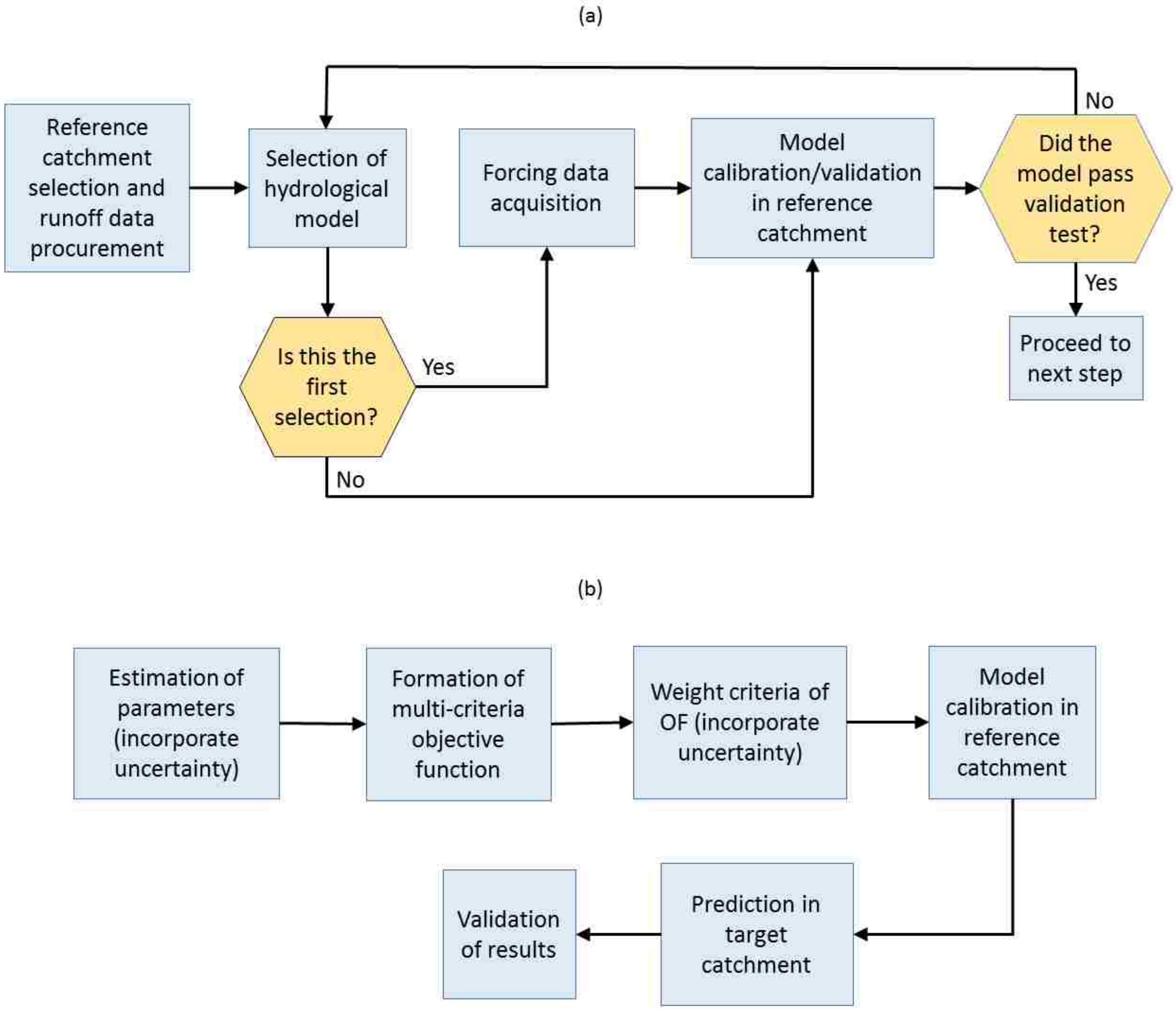

Figure 2. A proposed framework for prediction in data-limited regions: a) preliminary model testing; b) multi-criteria calibration and prediction

The second part of the framework (Fig. 2b) consists of *a priori* parameter estimation and multi-objective model calibration in the reference catchment, followed by transfer of the final model to the target catchment(s) for prediction. Values of influential parameters are separately estimated in both reference and target catchments from available physical data. There is opportunity for creativity as to how parameters may be estimated in the absence of hard data (e.g. Seibert & McDonnell, 2002). In lieu of detailed site-scale observations, global/regional databases may be used (as in this study) to constrain the range of acceptable parameter values. Uncertainty in estimations by soft data (i.e. qualitative knowledge that cannot be expressed by exact numbers or quantitative knowledge associated with high uncertainties) may be acknowledged, for instance by representing parameter values as fuzzy numbers. A multi-criteria objective function (Eq. 4) is then parsed, which aims to simultaneously maximize runoff efficiency and hone parameter values within the ranges estimated with soft and fuzzy data. The model is re-calibrated to this multi-objective function, using fuzzy arithmetic, and the resulting model is transferred to the target catchment(s) for prediction. The final step is to validate predicted streamflows in the target catchment(s).

The remainder of this article describes the application of our proposed framework to two catchments within a severely data-limited region of southwestern China. In addition to demonstrating the applicability of the proposed framework in a truly poorly-gauged region, we undertake a comparison of model performance, comparing our proposed method versus a single-objective model as well as a constrained single-objective model.

2.4.2 Model Testing in Upper Salween River and Upper Mekong River Basins

2.4.2.1 The Region of Study

The international Salween and Mekong Rivers (known as Nu and Lancang Rivers in China) originate from the eastern highlands of the Tibetan Plateau and flow from north to south through Yunnan Province (China) before entering Myanmar and Laos, respectively (Fig .3). Near the

Yunnan-Tibet border, the relief between the river valley and ridges varies by as much as 2500 m, while the terrain is less steep further south. There exist many tributaries to both rivers which, in Yunnan Province, drain small catchments in steep valleys. As opposed to the mainstem rivers, snow and glacial melt do not constitute a significant runoff source to tributaries (Chinese Academy of Sciences, 1990; Yunnan Bureau of Hydrology and Water Resources, 2005; Mekong River Commission, 2005). The climate is monsoonal, though climate varies considerably with local topography in Yunnan Province (Mekong River Commission, 2005). Rainfall is characterized by two seasonal pulses (first between February-May, and second between June-October) and high river flows correspond (Kibler and Tullos, 2013). Streamflow and precipitation measurements are sparse in this mountainous region, particularly in tributaries. Large-scale subsurface data for the region are limited to global databases. Land use/cover is similar in the Upper Salween River and Upper Mekong River basins: more than 90% of the Salween and 83% of the Mekong basin in this region is covered by forests and other types of vegetation (DeFries and Hansen, 2010). Croplands cover about 7% of the land in the Salween and about 15% of that in the Mekong basin. Less than 0.1% of land is urbanized in either catchment.

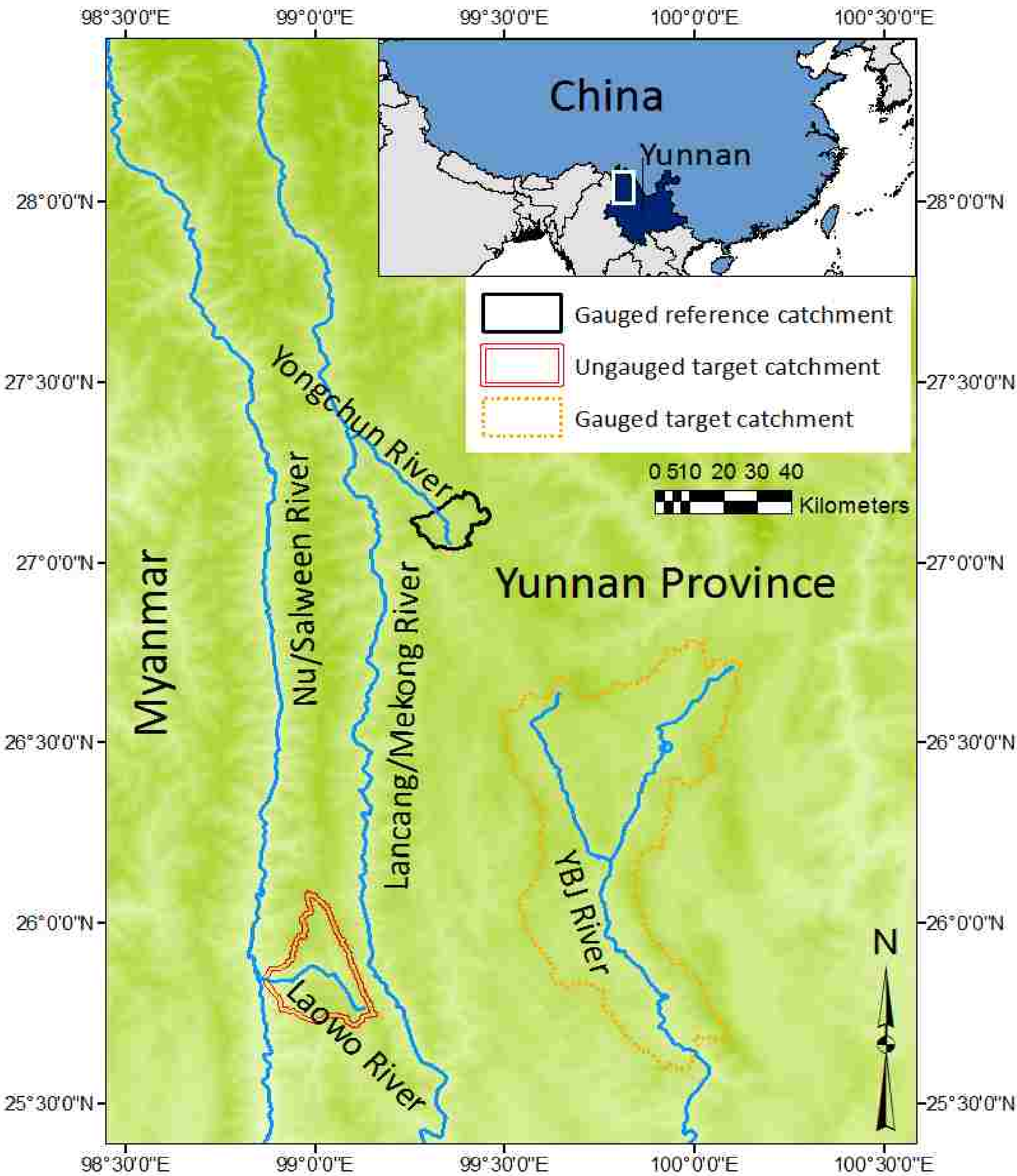

Figure 3. Study region and target (Laowo River and YBJ River) and reference (Yongchun River) catchments

2.4.2.2 *Target and Reference Catchments*

We consider two target catchments for streamflow prediction. Yang Bi Jiang (YBJ) River (approximately 4300 km^2), is a gauged catchment located in the Upper Mekong basin, with a 20-year observed streamflow record. We also pilot the method in a catchment with a much shorter record of observed streamflow data, to realistically demonstrate challenges faced by practitioners predicting flows in poorly gauged regions. Laowo River (approximately 575 km^2) in the Upper Salween River Basin (Fig. 3), is a fifth-order river that drains the western slopes of the ridge separating the Salween and Mekong River basins. Only one year of observed unregulated streamflow is available in Laowo River.

The reference catchment, Yongchun River (197 km^2), is a gauged tributary to the Upper Mekong River. The Tangshang daily stream gauge has been operational on the Yongchun River since 1960. Mean basin elevations in the Laowo, YBJ, and Yongchun catchments are similar (mean values of 2475, 2612, and 2935 m, respectively (Danielson and Gesch, 2011)) and river flows in all catchments are dominated by rainfall-runoff processes. Land cover is similar in all three catchments, with about 90% of each basin covered by forests and other types of vegetation while croplands cover about 10% of the land (DeFries and Hansen, 2010).

2.4.2.3 *Catchment Similarity*

Choosing a reference catchment that is hydrologically similar to target catchment(s) will generally yield more accurate runoff predictions (Singh et al., 2014). However, the choice of a reference catchment may be limited in sparsely-gauged regions. Here we test target catchments that are only partially similar to the reference catchment. While the target and reference catchments are similar with respect to topography, climate, and land cover, the contributing area of the reference catchment is much smaller than either target catchment. A Q-Q plot (Wilk and Gnanadesikan, 1968) of nine daily streamflow quantiles (Fig. 4) indicates a linear relationship between distribution functions of streamflow in the reference and target catchments, providing some superficial evidence of their hydrologic similarity. However, the rivers are potentially far from perfect hydrologic analogs due to their substantial size differences.

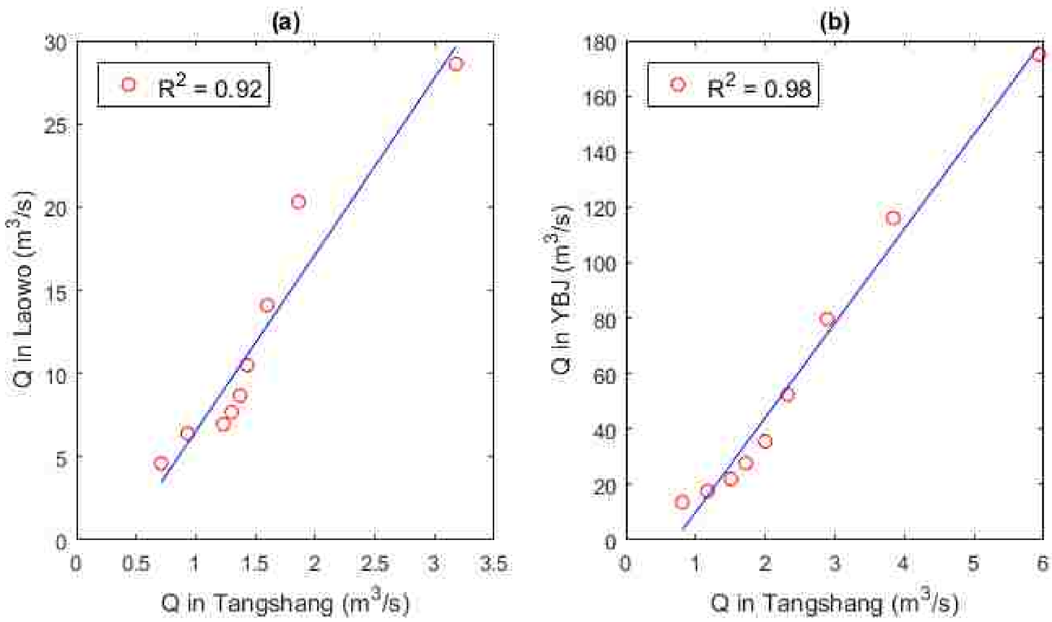

Figure 4. Q-Q plot of daily streamflow in a) target (Laowo) and reference (Yongchun) rivers, 1987; b) target (YBJ) and reference (Yongchun) rivers, 1962-1973

2.4.2.4 Model Selection and Initial Testing

We chose a simple conceptual model, HyMOD (Moore 1985, 1999), for flow prediction. HyMOD is a lumped, deterministic model that predicts river flows based on simulated probability distributions of soil moisture. To avoid over-parameterization, we chose the simplest form of

HyMOD, the 5-parameter model. In this model, a catchment consists of infinite points, each defined by a soil moisture capacity. Soil moisture capacities vary within the catchment as a result of variability in soil texture and depth (Wang et al., 2009). A cumulative distribution function (CDF) describes catchment soil moisture variability (Eq. 1):

$$F(c) = 1 - \left[1 - \frac{c}{C_{max}}\right]^{B}, 0 \leq c \leq C_{max} \quad (1)$$

Where c is soil moisture capacity, C_{max} is the maximum soil moisture capacity within the catchment, and B is a shape factor that is dependent on the degree of spatial variability in soil moisture capacities. Besides C_{max} and B, the other parameters of HyMOD include R_q, which is inverse of residence time in quick reservoirs, R_s, which is inverse of residence time in a slow reservoir, and α, which is a fraction coefficient for distribution of water between slow and quick reservoirs. For the sake of brevity, we refer our readers to Wang et al. (2009) and Moore (1985, 1999) for a comprehensive description of HyMOD. We coded HyMOD into MATLAB (Mathworks, 2012), in a configuration that can process 10,000 different combinations of model parameters (for 13 years of daily data) within 15 minutes on a typical processor (Intel Core i7-5500U @ 2.40GHz).

2.4.2.4.1 Estimating precipitation and PET

To estimate precipitation and potential evapotranspiration (PET), we extracted daily cumulative precipitation (mm) and daily mean temperature (°C), respectively from the APHRODITE precipitation dataset and AphroTemp dataset for Monsoon Asia (APHRODITE website, 2015). Daily precipitation and temperature were extracted from 0.25° resolution grids and mean values were computed over the reference and target catchments. We used Thornthwaite's approach (1948) to estimate PET from temperature data. We estimated daylight length using the model of Forsythe et al. (1995), which uses latitude and day of the year for its estimations.

We chose APHRODITE because it covers our period of analysis (before 1987) and has been shown to perform comparably to or better than other precipitation data products (e.g., Chen et al., 2017; Khandu et al, 2016; Zhao et al, 2015). However, the APHRODITE dataset is associated with underestimation of precipitation in high altitudes due poor representation of orographic effect (Kishore et al., 2015; Wi et al., 2015). To ensure accurate precipitation input, we compared APHRODITE precipitation with daily precipitation estimated from observations within the region's sparse ground-based precipitation network (1960-1987). As the region's limited locations of long-term observed rainfall data did not justify advanced methods, we applied the Inverse Distance Weighting (IDW) method (Chen & Liu, 2012; Keblouti et al., 2012). Comparisons of APHRODITE and IDW hyetographs with observed streamflow hydrographs indicate that timing of precipitation was better represented by APHRODITE (Fig. 5). However, we expect that long-term precipitation estimations given by observed data will provide more accurate estimations of annual cumulative precipitation.

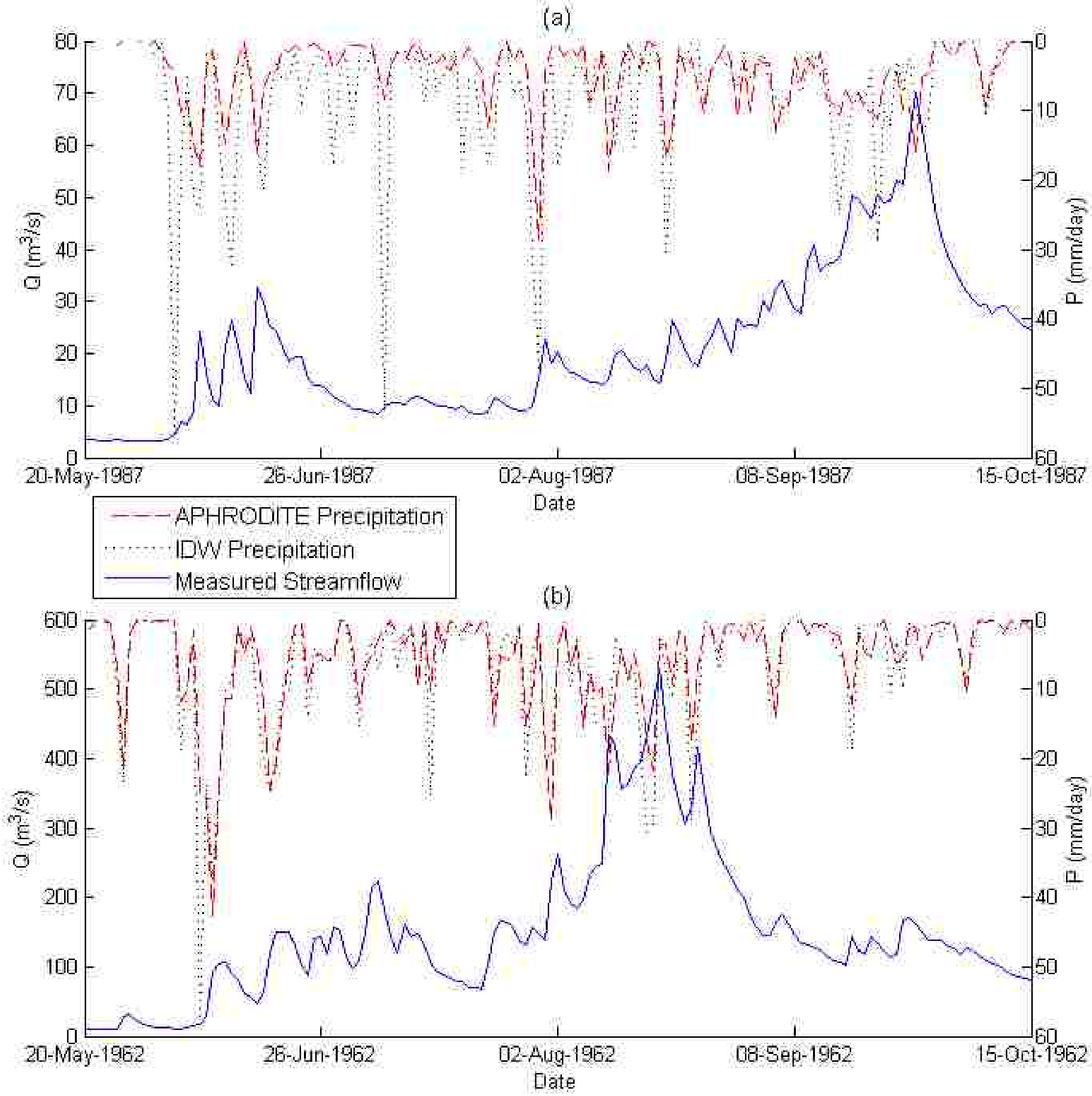

Figure 5. APHRODITE hyetograph versus observed discharge hydrograph for a) May 20 - Oct 15, 1987, in Laowo, b) May 20 – Oct 15, 1962 in YBJ

We followed the empirical quantile mapping approach (Bennett et al., 2014; Lafon et al., 2013) to bias correct magnitudes of catchment-averaged daily precipitation from APHRODITE using IDW cumulative distribution functions. To avoid overfitting of APHRODITE to IDW data, we mapped two discrete quantiles ($Q_0 - Q_{50}$ and $Q_{50} - Q_{100}$). In the Yongchun catchment, APHRODITE and IDW CDFs correspond well (Fig. 6a), thus transfer functions of unity were applied across all quantiles. In the Laowo catchment, CDFs illustrate that APHRODITE underestimated long-term precipitation as compared to IDW (Fig. 6b). Therefore, transfer functions of 1.26 and 1.4 were applied respectively to daily precipitation values below 3.96 mm and above 3.96 mm. Similarly in YBJ, transfer functions of 0.98 and 1.12 were applied respectively to precipitation values below 4.23 mm and above 4.23 mm.

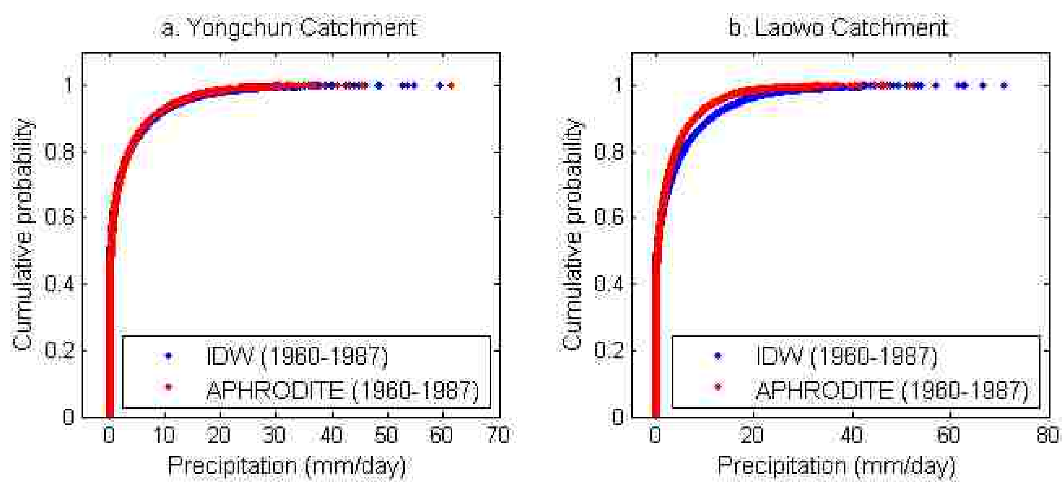

Figure 6. a) Precipitation CDFs in Yongchun River catchment, and b) Laowo River catchment

2.4.2.4.2 Single-objective calibration

We calibrated HyMOD in Yongchun catchment using observed streamflow data from 1961-1973, and validated the model from 1975-1980 (Chinese Ministry of Hydrology, 1970, 1971, 1974, 1977, 1982a, 1982b). We analyzed the impact on parameter estimates of swapping the calibration and validation periods. Observed discharge data were prepared for prediction periods in Laowo (1987) and YBJ (1962-1973, 1975-76, 1978-1980, 1984-85, and 1987) (Henck et al., 2010). We investigated a wide feasible range of HyMOD parameters (Table 2).

Table 2. Upper and lower bounds of parameters in HyMOD

Parameter	Lower bound (X_{min})	Upper bound (X_{max})
B	0.01	4
C_{max} (mm)	5	1500
R_q (day^{-1})	0.01	0.99
R_s (day^{-1})	0.0001	0.01
α	0.01	0.99

We calibrated the model by running 9375 different parameter combinations through a branch-and-bound method with the single objective to maximize Nash-Sutcliffe runoff efficiency (NSE). Through our branch-and-bound method, the model first discretized the feasible range given for each parameter into 4 equal sections with 5 boundary values. All combinations of boundary values ($5^5 = 3125$) were then tested to maximize runoff efficiency. A region of one third of the original parameter space was further investigated around each selected parameter, equally distributed on both sides of the selected value if possible (i.e. if the selected value was not too close to the lower or upper limits of the feasible range). The newly selected range for each parameter was discretized similarly to the previous step and 3125 different parameter combinations were again tested to optimize the objective function. This process was performed one more time, choosing only one third of the constrained parameter space (without crossing the borders to the previous range). The final parameter values selected through single-objective calibration were used to validate model performance in Yongchun River.

2.4.2.5 Multi-Objective Model Evaluation

2.4.2.5.1 Estimation of model parameters associated with soil moisture capacity using soft data

Rather than relying on rote calibration to discharge observations in the reference catchment, we seek to improve model performance in ungauged target catchments by increasing confidence in parameter estimates. However, only little and highly uncertain information is available to characterize soil moisture capacity in the region of study. We used International Soil Reference and Information Centre (ISRIC) databases to estimate the total available water capacity (TAWC) (WISE30sec dataset, Batjes, 2015) and depth to bedrock (SoilGrids1km dataset, Hengl et al., 2014). Both datasets are global in coverage. Combining TAWC (cm) and depth to bedrock (cm), we estimated the soil moisture capacity (cm) in each catchment at a spatial resolution of approximately 1 km^2. We then created empirical CDFs of soil moisture capacity at the catchment scale. We fitted the HyMOD soil moisture capacity CDF (Eq. 1) to the empirical CDFs to estimate values of B and C_{max} directly from soft data (Fig. 7).

Use of the global datasets potentially introduces substantial uncertainty, which we addressed by considering ranges within extreme potential end members and defining parameters as trapezoidal fuzzy numbers. Depth to bedrock in the global dataset is reported only to a maximum value of 2.4 m. However, it is possible that soil depth in parts of the study areas may exceed 2.4 m. In total, about 75 km^2 (38%) of Yongchun, 1427 km^2 (33%) of YBJ and 98 km^2 (21%) of Laowo are characterized by the maximum soil depth of 2.4 m. To characterize the additional uncertainty in soil moisture capacity of these particularly data-limited areas, we estimate two extreme potential end members, of maximum soil depth of 2.4 m and 4.8 m. Through this approach, we provide a reasonable estimation of the lower and upper bounds of the soil moisture capacity CDFs and HyMOD parameters values (Fig. 7). A more comprehensive explanation of our analysis approach has been provided as supplementary material published online with this manuscript.

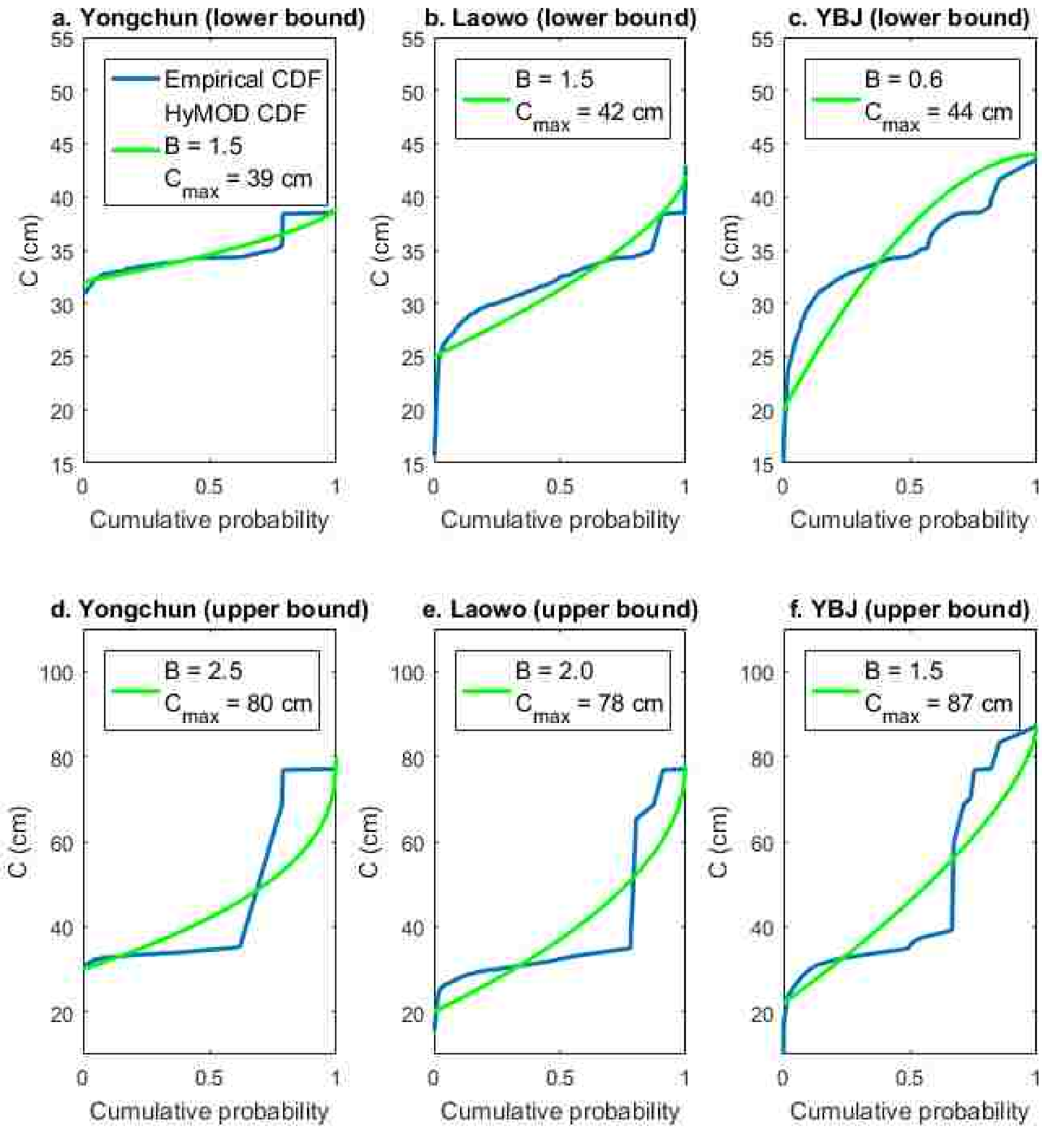

Figure 7. Soil moisture capacity (c) and parameter values estimated from soft data

Using the values of B and C_{max} fitted to empirical CDFs of soft data, we defined trapezoidal fuzzy numbers, $\tilde{a} = (x_l, x_{c1}, x_{c2}, x_r)$, for these parameters in each catchment (Table 3), as in Fig. 8.

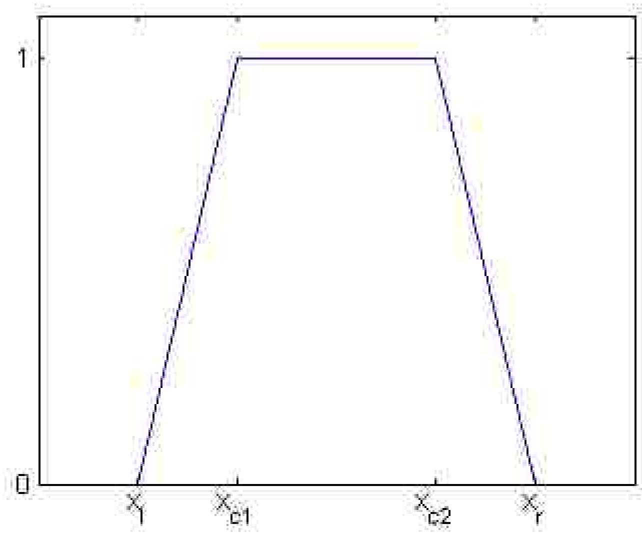

Figure 8. Construction of a trapezoidal fuzzy number

Table 3. Trapezoidal fuzzy *a priori* estimates of B and C_{max} in target and reference catchments.

Catchment	Parameter	x_l	x_{c1}	x_{c2}	x_r
Yongchun	B	1.2	1.5	2	3
	C_{max} (mm)	350	500	700	1000
Laowo	B	1	1.7	2	2.5
	C_{max} (mm)	400	550	650	950
YBJ	B	0.3	0.6	1.5	2.5
	C_{max} (mm)	400	500	850	1000

2.4.2.5.2 Creation of a multi-criteria objective function

To encompass additional information given by soft data, we developed a multi-criteria objective function (OF) for model calibration. The criteria include: minimize the ratio between model error and variation in observed data (i.e. maximize NSE, criterion 1), and minimize the difference between B and C_{max} values estimated through optimization and soft data (criteria 2 and 3, respectively). To standardize the three criteria into one coherent function, we present each as a value from 0 to 1. After determining the feasible range of each parameter, $[X_{min}\ \ X_{max}]$, (Table 2), criteria 2 and 3 were normalized as in Eq. 2:

$$\frac{\tilde{a}-X}{\tilde{a}-X_m}$$

$$\begin{gathered} If\ |defuzz\{\tilde{a}-X_{\max}\}| \geq |defuzz\{\tilde{a}-X_{\min}\}| \\ then\ X_m = X_{max} \\ else\ X_m = X_{min} \end{gathered} \tag{2}$$

Where $\tilde{a}$ is the estimated trapezoidal fuzzy value of the parameter of interest, X is the estimated value for the same parameter by model calibration, X_m is either X_{min} or X_{max}, and defuzz denotes the defuzzification process through the centroid method (Sugeno, 1985).

We allowed the strength of each individual criterion to vary with respect to the others by permitting unique weighting of the criteria. Weights were first assigned in the form of crisp (non-fuzzy) numbers, considering relative importance of the criteria, level of certainty in *a priori* parameter estimations, and degree of resonance between estimations in the reference and target catchments. Weights were then converted to fuzzy numbers, to address subjectivity associated with choice of weight. Therefore, the weights for criteria 2 and 3 were modified into the form of triangular fuzzy numbers, $\widetilde{W} = (x_l, x_c, x_r)$, (Table 4) as in Eq. 3.

$$\widetilde{W}(x) = \begin{cases} \frac{x-x_l}{(x_c-x_l)} & x_l \leq x < x_c, \\ 1 & x = x_c, \\ \frac{x_r-x}{(x_r-x_c)} & x_c < x \leq x_r, \\ 0 & otherwise. \end{cases} \quad (3)$$

Given the uncertainty associated with parameter estimation by soft data, we chose to weight criterion 1 (maximize NSE) considerably higher than criteria 2 and 3 (Table 4). Empirical analysis suggested a closer resonance between B estimations in Yongchun and Laowo than those in Yongchun and YBJ. In assigning weights to criterion 2 (fit B to value estimated by empirical analysis of soft data), we accordingly assigned a higher initial weight to criterion 2 in Laowo than in YBJ (Table 4). We tested sensitivity of parameter values to criteria weighting, detecting differences in some of parameter values returned by varying weights (from 0.77 to 1.99, for B in YBJ for instance).

Table 4. Weighting of multi-objective function terms

Catchment	Criteria	x_l	x_c	x_r
Laowo	1 (NSE)	0.8	0.8	0.8
	2 (B)	0.05	0.1	0.12
	3 (C_{max})	0.07	0.1	0.15
YBJ	1 (NSE)	0.84	0.84	0.84
	2 (B)	0.03	0.05	0.07
	3 (C_{max})	0.08	0.11	0.16

Finally, the general form of the multi-criteria objective function (OF) is as follows:

$$OF = \sum_{i=1}^{n} \left| defuzz\left\{\widetilde{W}_i * \frac{\tilde{a}_i - X_i}{\tilde{a}_i - X_{m_i}}\right\}\right| + |defuzz\{W_{n+1} * (1 - NSE)\}|$$

$$NSE = 1 - \frac{\sum_{t=1}^{T}(Q_o^t - Q_m^t)^2}{\sum_{t=1}^{T}(Q_o^t - \overline{Q_o})^2} \quad (4)$$

Where $\widetilde{W}_i$ is the fuzzy weight assigned to criterion i, W_{n+1} is the weight assigned to criterion $n+1$, Q_o^t is observed streamflow in time step t, Q_m^t is modeled streamflow in time step t, and $\overline{Q_o}$ is the mean value of all observed streamflows. The value of the OF is almost always between 0 and 1, where OF = 0 indicates perfect match between parameter estimates from calibration and soft data analysis, as well as a perfect match between observed and predicted streamflows.

2.4.2.5.3 Multi-objective calibration and validation

We again tested 9375 parameter combinations at the Tangshang gauge in Yongchun River during the calibration period, and optimized the model through the same branch-and-bound method to minimize our multi-objective function (Eq. 4). As the goal at this time is flow prediction in the target catchments, it was not necessary to validate the multi-objective model in the reference catchment. Validation was instead performed in the target catchments, Laowo and YBJ. The calibrated models for Laowo and YBJ were respectively validated using streamflow observations in 1987, including an initialization period in 1986, and 1962-1987 (with some years missing), including an initialization period in 1961.

2.4.2.6 Evaluation of Model Performance: Comparing Single- to Multi-objective Model Selection

We compared the performance of the multi-objective calibrated model in each catchment against a single-objective calibrated model and a constrained single-objective calibrated model (Tables 5 and 6). For the single-objective model, we applied parameters estimated by single-objective calibration to observed streamflow in the reference catchment. For the constrained single-objective model, parameters B and C_{max} were constrained to values between the lower and upper bounds estimated by empirical analysis of soft data (Table 3) and the model was calibrated to observed reference catchment streamflows. For the multi-objective model, we applied parameter estimates derived by the multi-objective calibration methods developed in Section 2.4.2.5. In addition to comparing models by runoff efficiency (NSE), we investigated parameter residuals (Eq. 5) and the combined OF (objective function) value (Eq. 4) of each model.

Parameter residual = centroid of parameter value estimated with soft data – calibrated parameter value (5)

2.5 Results and Discussion

While all tested models performed well with respect to runoff efficiency (Table 5), models selected using the proposed multi-objective method may incorporate more reasonable parameter values. When models selected by single-, constrained single- and multi-objective calibration in the reference catchment (Yongchun) are transferred to Laowo for prediction, runoff efficiencies are similar, respectively 0.78, 0.77, and 0.74 (Fig. 9 and Table 5). However, residuals of parameters B and C_{max} are relatively large for the single-objective model (1.24 and 475 mm, respectively) and OF value (0.33) is greater than for the constrained single-objective or multi-objective models, which may indicate the selection of less reasonable parameter values.

When the calibrated models are transferred to YBJ for prediction, the multi-objective model substantially outperforms both the single-objective and constrained single-objective models with respect to parameter residuals (0.48 versus 0.65 and 0.7 for B, and 61 mm versus 289 and 518 mm for C_{max}), which is reflected by a lower overall OF value (0.27 versus 0.32 and 0.41). Comparison with 20 years of observed data in YBJ also indicates that runoff efficiency of the single-objective model (NSE = 0.67) is lower than the multi-objective (NSE = 0.72) and the constrained single-objective (NSE = 0.73) models. Thus, benefits of incorporating catchment-specific *a priori* parameter estimates based on soft data can be seen, particularly in testing against longer observed time series.

Table 5. Performance of models with respect to NSE, parameter residuals, and OF value

Target catchment	Metric	Single-objective model	Constrained single-objective model	Multi-objective model
Laowo	OF value	0.33	0.28	0.27
	NSE	0.78	0.77	0.74
	B residual	1.24	0.79	0.91
	C_{max} residual (mm)	475	246	20
YBJ	OF value	0.41	0.32	0.27
	NSE	0.67	0.73	0.72
	B residual	0.7	0.65	0.48
	C_{max} residual (mm)	518	289	61

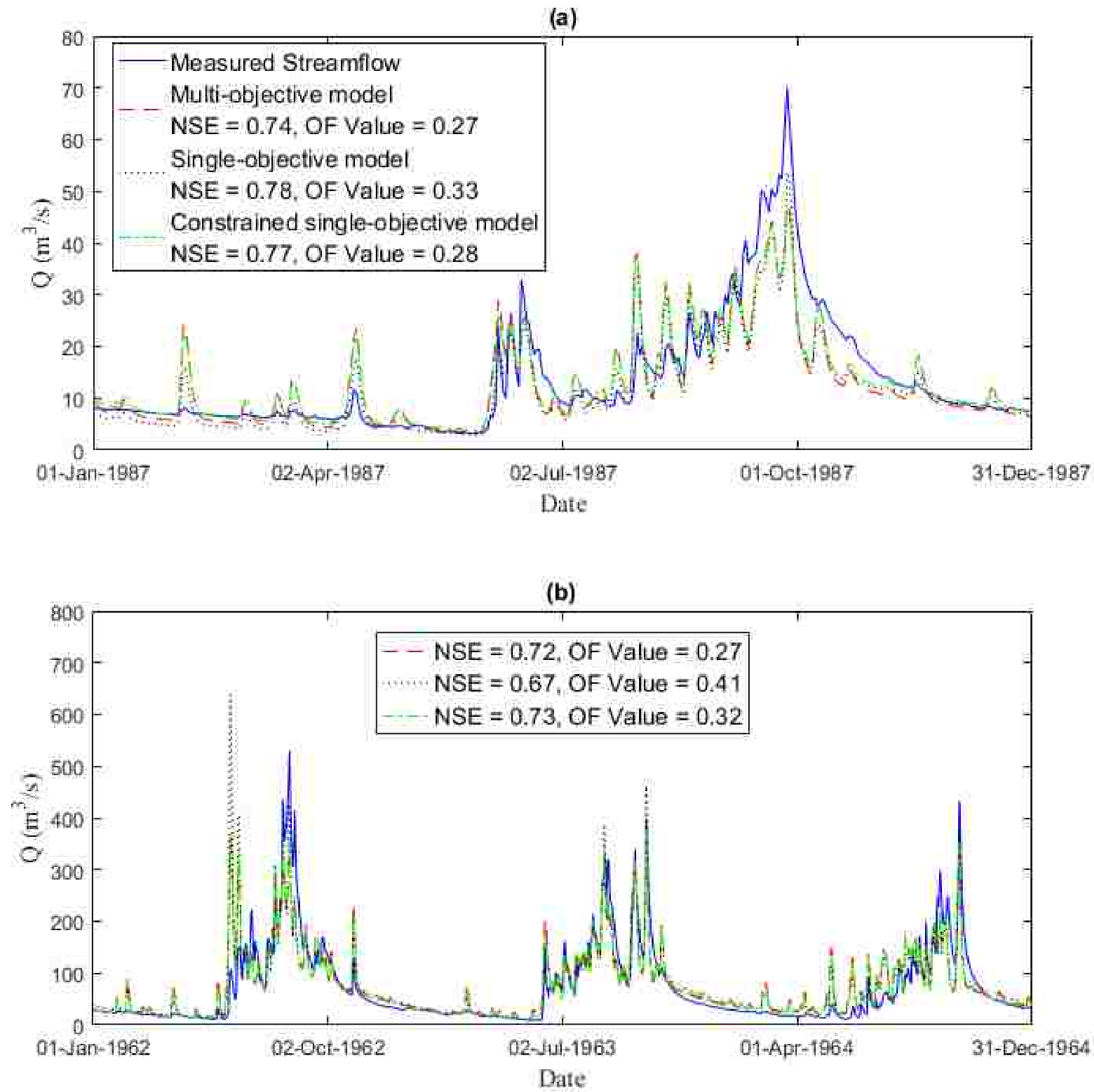

Figure 9. Flow prediction in a) Laowo and b) YBJ with parameter values selected by multi-objective, single-objective, and constrained single-objective calibration. The NSE and OF values for YBJ are calculated over a 20-year prediction period; data from 1962-1964 are shown.

2.5.1 Parameter Estimation in Target Catchments Laowo and YBJ

While values of R_q, R_s and α are estimated to be similar across all three models (Table 6), values of C_{max} and B vary across models according to the calibration method. As per design, the multi-objective model consistently characterized C_{max} as similar to the centroid of *a priori* estimates (residuals in Laowo (YBJ) of 20 (61) mm). By comparison, the single-objective and constrained single-objective models estimate values of C_{max} that are much lower than *a priori* estimates (residuals in Laowo (YBJ) of 475 (518) mm and 246 (289) mm, respectively). Estimates of B values also varied among models. Relative to centroids of *a priori* estimates, both the multi-objective and constrained single-objective models select similar values of B in both Laowo (B residual of 0.79 versus 0.91) and YBJ (B residual of 0.65 versus 0.48). Both potentially estimate B more accurately than the single-objective model (B residual of 1.24 and 0.70 in Laowo and YBJ, respectively). Overall, the incorporation of *a priori* catchment-specific parameter estimates or ranges led to very different representations of soil moisture capacity across both catchments as compared to models selected by single-objective calibration (Fig. 10), without compromising runoff efficiency. These different representations of soil moisture capacity may translate into considerably different hydrological behaviors. For example, the greater soil storage capacities indicated by the *a priori* estimates and reflected in the multi-objective model translate to a less flashy hydrologic response. Predictions may be characterized by more stable base flows, fewer zero flow days, and higher constancy of flow in comparison to the representations modeled by the constrained single-objective model and especially the single-objective model. While many possibilities exist for matching runoff prediction to an existing data record (Beven, 1993, 2006), the multi-objective calibration method may be more adept at parametrizing models to provide the "right answers for the right reasons". Thus, managers may feel more confident to utilize such models to predict flows in fully ungauged areas. However, it should be noted that more exhaustive evaluation of the methods presented here are required to ensure that similar results would be replicated in other data-limited regions of the world.

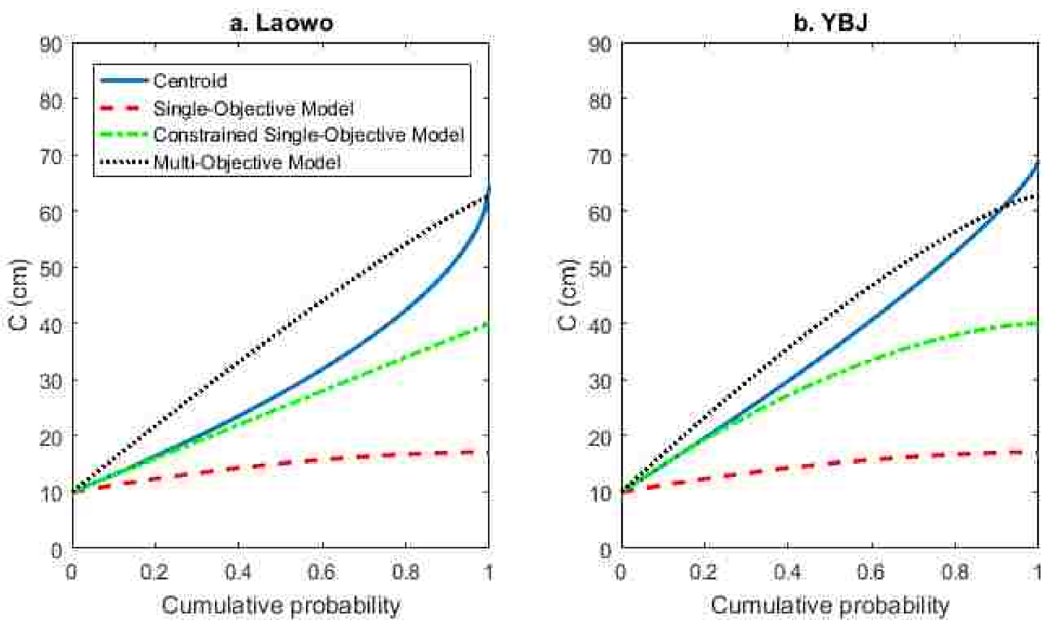

Figure 10. Soil moisture capacity modeling in a) Laowo catchment, B) YBJ catchment, by assuming a minimum capacity of 10 cm

Table 6. Parameter estimates for single-, constrained single-, and multi-objective models.

Parameters and metrics	Single-objective	To transfer to Laowo		To transfer to YBJ	
		Constrained single-objective	Multi-objective	Constrained single-objective	Multi-objective
B	0.55	1	0.88	0.6	0.77
$C_{max}\ (mm)$	171	400	626	400	628
$R_q\ (day^{-1})$	0.66	0.64	0.64	0.63	0.63
$R_s\ (day^{-1})$	0.01	0.01	0.01	0.01	0.01
α	0.36	0.39	0.445	0.47	0.45

2.5.2 Calibration Using a Multi-objective Function

Optimizing to a multi-criteria OF value during calibration proves a robust method to incorporate important characteristics of models, in this case including runoff efficiency and parameters controlling soil moisture capacity. In both catchments, the constrained single- and multi-objective models perform comparably with respect to runoff efficiency and representation of soil moisture capacity, though in both cases the multi-objective model matches *a priori* estimates of maximum soil moisture capacity more closely. By contrast, the single-objective model consistently represents soil moisture capacity as much lower than *a priori* parameter estimates. In YBJ, the single-objective model also underperforms with respect to NSE. This suite of information is collated through the OF value, which indicates that the multi-objective model is in both catchments the superior overall model. Simultaneously optimizing to multiple criteria in a weighted OF value may be a solution that helps managers differentiate between models with similar performance in runoff efficiency, yet which portray vastly different hydrological processes (equifinality), as discussed above.

Our results underscore prior findings that multi-objective calibration methods including objectives to match parameter estimates with 'ground-truthed' values provide better hydrologic representation than models parameterized using rote calibration to observed runoff (Parajka et al., 2007b; Rajib et al., 2016; Seibert and McDonnell, 2002). The performance of the constrained single-objective model, which takes advantage of parameter ranges constrained by *a priori* estimates, is comparable and computationally less demanding than the multi-objective model. However, the multi-objective model may be preferable for several reasons:

1) As in this study, the multi-objective model may be more adept to represent physical/hydrological characteristics of a catchment.

2) The possibility of defining new criteria, instead of using soft data in the form of hard constraints, provides the opportunity of weighting the criteria and differentiating between them based on their relative importance or confidence in information.

3) The lower and upper bounds of fuzzy numbers defined for a parameter may sometimes span across the entire feasible range of that parameter. Accordingly, constraining a parameter within its feasible range would not add any information to the single-objective model. The multi-objective model, on the other hand, could still use such soft data in calibration by putting more emphasis on the centroid of the estimates.

4) The multi-objective approach described herein allows the optimization process to search the entire feasible range of parameters while also incorporating the *a priori* estimates. On the other hand, hard constraints limit the search range based on these uncertain estimates.

2.5.3 Application of Multi-objective Calibration for Flow Prediction in Ungauged Basins

For practitioners wishing to simulate river flows in ungauged basins, the proposed multi-objective modeling framework allows for the application of advanced conceptual hydrological models within severely data-limited regions, where existing methods requiring robust data are often inapplicable. As opposed to a new hydrologic model, the proposed method is designed to work with a practitioner's preferred model, allowing managers to select the best hydrologic model for their location and objective. In many regions, the proposed method may be an improvement over existing methods, especially where data-intensive regionalization techniques are infeasible; however, this remains subject to further evaluation. In comparison to previous flow predictions in Laowo River generated through a catchment similarity modeling approach (Kibler and Alipour, 2017), the multi-objective model demonstrates substantial improvement with respect to runoff efficiency. The considerable improvement in predictive skill attributes to multiple factors, including improved rainfall bias correction, but is primarily due to the more realistic representation of the relationship between reference and target catchments. The proposed modeling approach is thus a proficient tool to bridge hydrologic non-similarity between reference and target catchments in the region of study, allowing for flexible use in the region where gauged catchments are few and dissimilarity is unavoidable.

Robustness to some level of catchment dissimilarity is indebted to initial testing for suitability in the region of study before application and the influence of *a priori* parameter estimates from the target catchment in calibration. For example, suitability of HyMOD for the region of study was confirmed in an initial phase of traditional calibration to the single objective of maximizing runoff efficiency (NSE = 0.74), and validation (NSE = 0.75) in the reference catchment (Fig. 11). Swapping the calibration and validation periods did not change parameter estimates considerably. *A priori* estimates derived from highly uncertain data may be the best available information in regions of sparse data. In such cases, practitioners may distribute unequal weights across OF criteria to acknowledge uncertainty, as demonstrated herein. Analysis of the sensitivity of parameter estimates to weighting (result provided as supplementary material to this manuscript) indicates that choice of weight is influential to estimates of B; C_{max} and α are less sensitive to weight, and R_q and R_s are not influenced by weighting. Overall, the weighting process should be performed based on confidence in *a priori* parameter estimates, degree of resonance between estimations in the reference and target catchments, expert opinions, and a decision maker's level of risk tolerance to go from traditional performance metrics (e.g. maximize NSE) to more physical/hydrological-based metrics (e.g. minimize parameter estimate residuals).

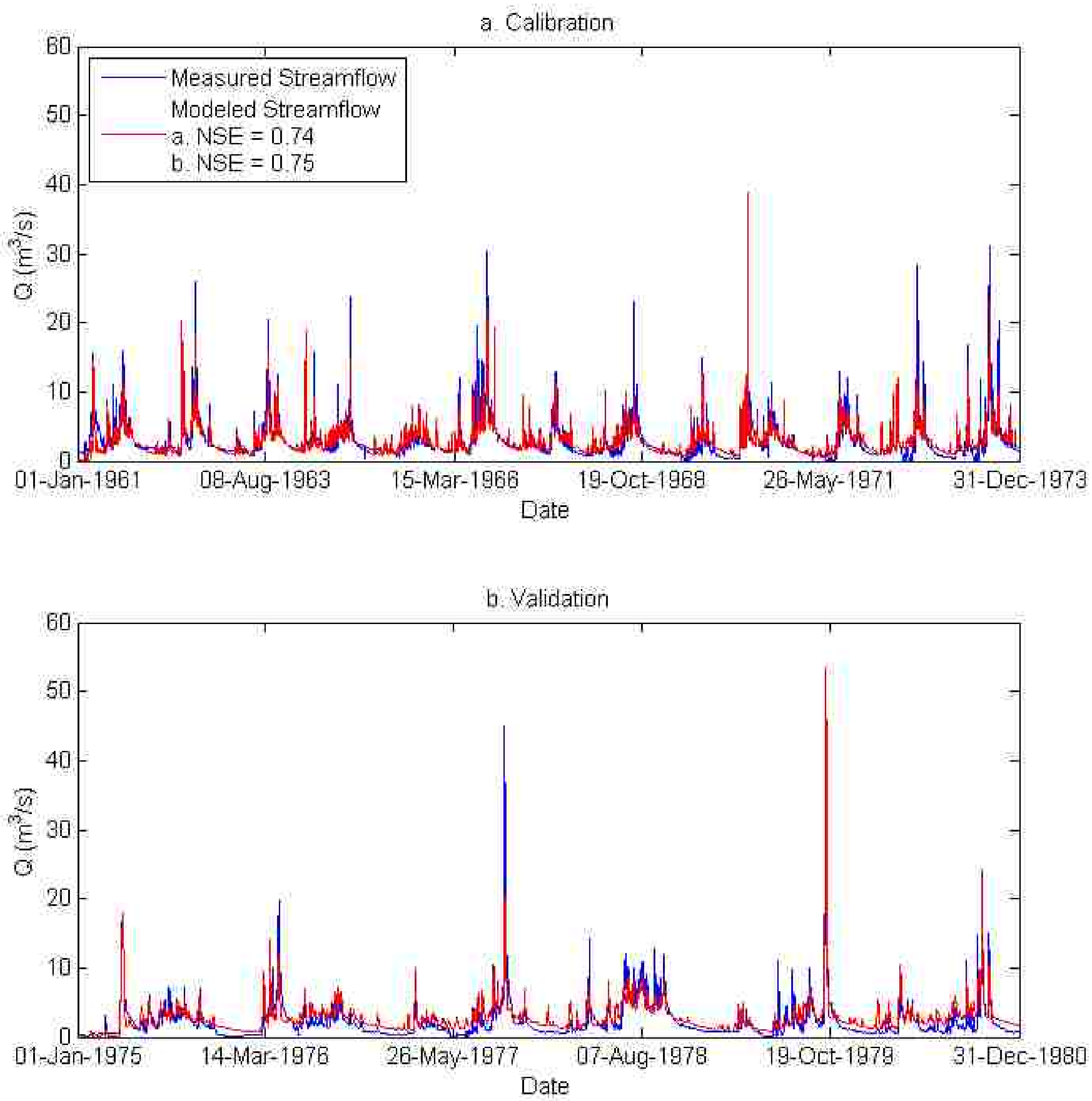

Figure 11. a) Calibration and b) validation at the Tangshang gauge in Yongchun River

The proposed method involves calibration to flow data from a reference catchment; however, optimization to the OF value simultaneously encourages parameter values to remain similar to *a priori* values estimated in a different (target) catchment. A loss in runoff efficiency during calibration, relative to models calibrated to the single objective of echoing reference catchment flows, is therefore sometimes to be expected. For example, calibration runoff efficiencies (obtained during calibration in the reference catchment) of the multi-objective (Laowo (YBJ) NSE = 0.67 (0.67), Fig. 12) and constrained single-objective models (Laowo (YBJ) NSE = 0.69 (0.71)) were lower than the NSE value obtained with single-objective calibration (NSE = 0.74). It is important to note, however, that lower runoff efficiencies at the calibration stage may not always translate to lower prediction efficiency when models are transferred to target catchments. For instance, the multi-objective and constrained single-objective models returned greater runoff efficiencies for prediction in YBJ as compared to the single-objective calibration, despite the lower efficiency at the calibration stage.

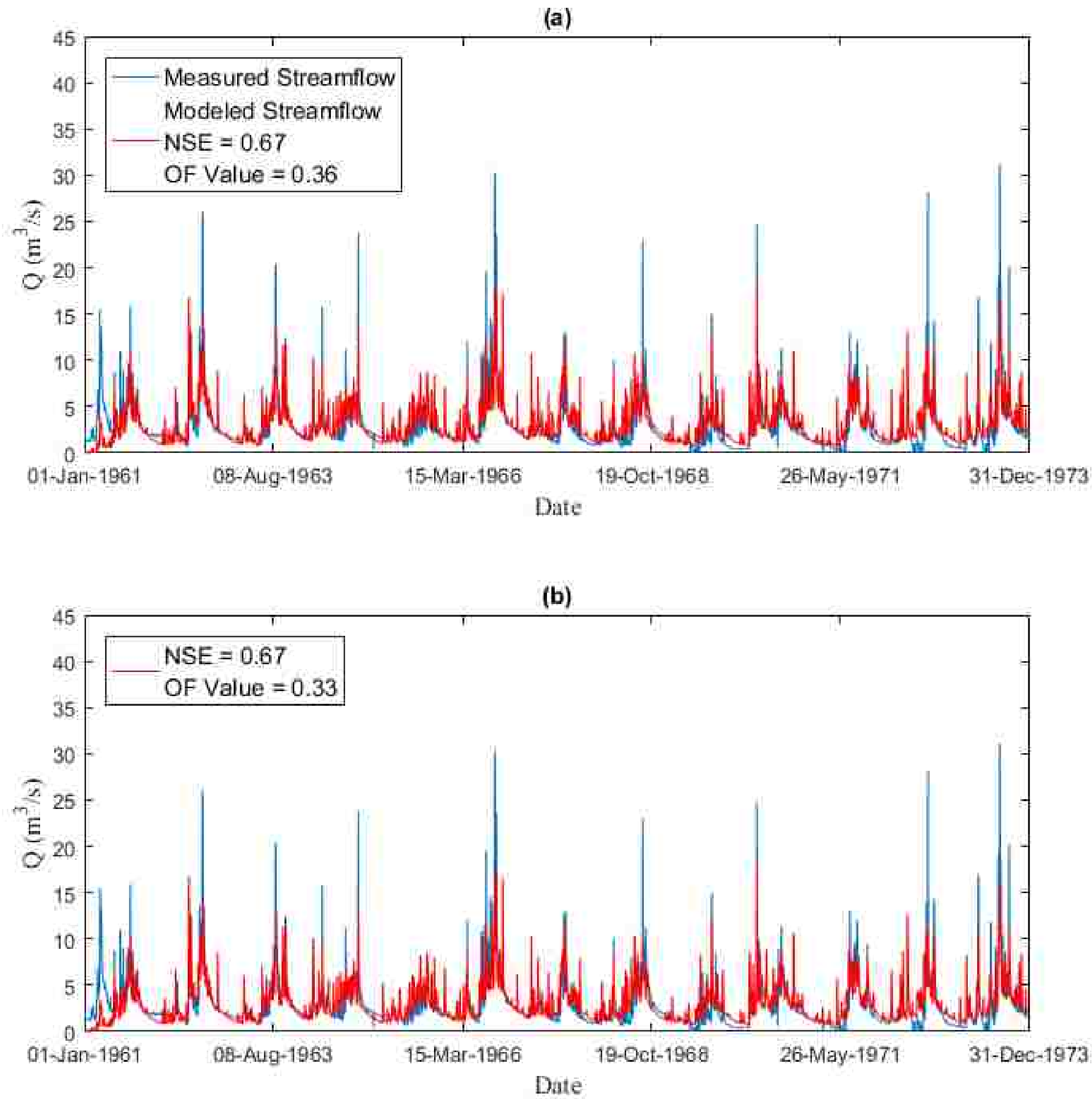

Figure 12. Calibration at the Tangshang gauge in Yongchun River using soft data and a multi-criteria objective function to transfer to a) Laowo, b) YBJ

2.6 Conclusion

A multi-objective framework for flow prediction in the most data-limited regions of the world was proposed and successfully tested in a remote, data-limited region of southwestern China. Despite data limitations, streamflow predictions generated by the proposed multi-objective model demonstrated reasonable runoff efficiencies in two target catchments (NSE = 0.72 and 0.74). Performance of the proposed multi-objective method was similar to that of single-objective and constrained single-objective models. However, the multi-objective model also selected values of influential parameters that more closely resonate with *a priori* estimates derived from soft data. Parameter residuals relative to *a priori* estimates of maximum soil moisture capacity were lowest for the multi-objective model (20 – 61 mm), and much greater for the constrained single-objective (246 - 289 mm) and single-objective (475 - 518 mm) models.

Managers predicting flows in regions of sparse data have been in some ways left behind in the wake of recent scientific advances in hydrologic modeling. Enhanced predictive tools that address the unique challenges faced in severely data-limited regions are needed. The proposed framework and approach to include *a priori* parameter estimates based on globally-available data in model calibration offers a preliminary step towards greater process understanding in regions of severe data limitations. For instance, when models are blindly calibrated to observed data in a reference catchment, managers may struggle to differentiate between competing models with similar performance but different representations of hydrological processes. The proposed calibration to a multi-objective function may allow practitioners to more confidently transfer calibrated models to predict flows in fully ungauged catchments. Future applications in more ungauged catchments and in other data-limited regions of the world can better clarify the merits of the proposed framework.

2.7 Acknowledgements

We would like to thank Dr. Dingbao Wang for his advice on hydrological model selection and student Fernanda Campos De Paulo for her contribution to data processing. We are also thankful to the China Hydrology Data Project (Henck et al., 2010) for digitizing observed data and making them available for research.

2.8 References

Andréassian, V., Le Moine, N., Perrin, C., Ramos, M.H., Oudin, L., Mathevet, T., Lerat, J., Berthet, L., 2012. All that glitters is not gold: the case of calibrating hydrological models. Hydrological Processes, 26 (14), 2206-2210.

APHRODITE website, 2015. Available at http://www.chikyu.ac.jp/precip/

Arsenault, R., Brissette, F.P., 2014. Continuous streamflow prediction in ungauged basins: The effects of equifinality and parameter set selection on uncertainty in regionalization approaches. Water Resources Research, 50 (7), 6135-6153.

Batjes, N.H., 2015. World soil property estimates for broad-scale modelling (WISE30sec, Ver. 1.0). Report 2015/01, ISRIC—World Soil Information, Wageningen (available via: http://www.isric.org/data/data-download)

Bennett, J.C., Grose, M.R., Corney, S.P., White, C.J., Holz, G.K., Katzfey, J.J., Post. D.A., Bindoff, N. L., 2014. Performance of an empirical bias-correction of a high-resolution climate dataset. International Journal of Climatology, 34 (7), 2189-2204.

Beven, K., 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. Advances in water resources, 16 (1), 41-51.

Beven, K., 2006. A manifesto for the equifinality thesis. Journal of hydrology, 320 (1), 18-36.

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

Chen, C.J., Senarath, S.U., Dima-West, I.M., Marcella, M.P., 2017. Evaluation and restructuring of gridded precipitation data over the Greater Mekong Subregion. International Journal of Climatology, 37, 180-196.

Chen, F.W., Liu, C.W., 2012. Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan. Paddy and Water Environment, 10 (3), 209-222.

Chiew, F.H., 2010. Lumped Conceptual Rainfall-Runoff Models and Simple Water Balance Methods: Overview and Applications in Ungauged and Data Limited Regions. Geography Compass, 4 (3), 206-225.

Chinese Academy of Sciences, 1990). Qinghai Tibet Atlas [in Chinese]. Science Publishing House, Beijing, China.

Chinese Ministry of Hydrology, 1970. Hydrologic Yearbook: Yunnan-Tibet International Rivers District, District 9 Region 2 1965, [in Chinese]. Ministry of Hydrology, Beijing, China.

Chinese Ministry of Hydrology, 1971. Hydrologic Yearbook: Yunnan-Tibet International Rivers District, District 9 Region 2 1966, [in Chinese]. Ministry of Hydrology, Beijing, China.

Chinese Ministry of Hydrology, 1974. Hydrologic Yearbook: Yunnan-Tibet International Rivers District, District 9 Region 2 1972, [in Chinese]. Ministry of Hydrology, Beijing, China.

Chinese Ministry of Hydrology, 1977. Hydrologic Yearbook: Yunnan-Tibet International Rivers District, District 9 Region 2 1978, [in Chinese]. Ministry of Hydrology, Beijing, China.

Chinese Ministry of Hydrology, 1982a. Hydrologic Yearbook: Yunnan-Tibet International Rivers District, District 9 Region 2 1979, [in Chinese]. Ministry of Hydrology, Beijing, China.

Chinese Ministry of Hydrology, 1982b. Hydrologic Yearbook: Yunnan-Tibet International Rivers District, District 9 Region 2 1979, [in Chinese]. Ministry of Hydrology, Beijing, China.

Confesor, R.B., Whittaker, G.W., 2007. Automatic Calibration of Hydrologic Models With Multi-Objective Evolutionary Algorithm and Pareto Optimization. JAWRA Journal of the American Water Resources Association, 43 (4), 981-989.

Danielson, J.J., Gesch, D. B., 2011. Global multi-resolution terrain elevation data 2010 (GMTED2010) (No. 2011-1073). US Geological Survey.

DeFries, R.S., Hansen, M., 2010. ISLSCP II University of Maryland global land cover classifications, 1992-1993. In Hall, F.G., Collatz, G., Meeson, B., Los, S., Brown de Colstoun, E., Landis, D. (Eds.). ISLSCP initiative II collection. Data set. Available online [http://daac.ornl.gov/] from Oak Ridge national laboratory distributed active archive center, Oak Ridge, Tennessee, USA.

Efstratiadis, A., Koussis, A.D., Koutsoyiannis, D., Mamassis, N., 2014. Flood design recipes vs. reality: can predictions for ungauged basins be trusted?. Natural Hazards and Earth System Sciences, 14 (6), 1417-1428.

Efstratiadis, A., Koutsoyiannis, D., 2010. One decade of multi-objective calibration approaches in hydrological modelling: a review. Hydrological Sciences Journal, 55 (1), 58-78.

Forsythe, W.C., Rykiel, E.J., Stahl, R.S., Wu, H.I., Schoolfield, R.M., 1995. A model comparison for daylength as a function of latitude and day of year. Ecological Modelling, 80 (1), 87-95.

Henck, A.H., Montgomery, D.R., Huntington, K.W., and Liang, C., 2010. Monsoon control of effective discharge, Yunnan and Tibet. Geology, 38, 975-978.

Hengl, T., de Jesus, J.M., MacMillan, R.A., Batjes, N.H., Heuvelink, G.B., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Gonzalez, M.R. (2014). SoilGrids1km—global soil information based on automated mapping. PLOS One, 9 (8), e105992.

Ibrahim, B., Wisser, D., Barry, B., Fowe, T., Aduna, A., 2015. Hydrological predictions for small ungauged watersheds in the Sudanian zone of the Volta basin in West Africa. Journal of Hydrology: Regional Studies, 4, 386-397.

Keblouti, M., Ouerdachi, L., Boutaghane, H., 2012. Spatial interpolation of annual precipitation in Annaba-Algeria-comparison and evaluation of methods. Energy Procedia, 18, 468-475.

Khandu, Awange, J.L., Forootan, E., 2016. An evaluation of high-resolution gridded precipitation products over Bhutan (1998–2012). International Journal of Climatology, 36, 1067-1087.

Kibler, K.M., Alipour, M., 2017. Flow alteration signatures of diversion hydropower: an analysis of 32 rivers in southwestern China. Ecohydrology, doi: 10.1002/eco.1846.

Kibler, K.M., Tullos, D.D. (2013). Cumulative biophysical impact of small and large hydropower development in Nu River, China. Water Resources Research, 49 (6), 3104-3118.

Kirchner, J.W., 2006. Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. Water Resources Research, 42 (3).

Kishore, P., Jyothi, S., Basha, G., Rao, S.V.B., Rajeevan, M., Velicogna, I., Sutterley, T.C., 2015. Precipitation climatology over India: validation with observations and reanalysis datasets and spatial trends. Climate Dynamics, 46 (1-2), 541-556.

Kult, J.M., Fry, L.M., Gronewold, A.D., Choi, W., 2014. Regionalization of hydrologic response in the Great Lakes basin: Considerations of temporal scales of analysis. Journal of Hydrology, 519, 2224-2237.

Lafon, T., Dadson, S., Buys, G., Prudhomme, C., 2013. Bias correction of daily precipitation simulated by a regional climate model: a comparison of methods. International Journal of Climatology, 33 (6), 1367-1381.

Mathworks, 2012. MATLAB software.

Mekong River Commission, 2005. Overview of the Hydrology of the Mekong Basin. Mekong River Commission, Vientiane 82.

Merz, R., Parajka, J., Blöschl, G., 2009. Scale effects in conceptual hydrological modeling. Water Resources Research, 45 (9).

Merz, R., Parajka, J., Blöschl, G., 2011. Time stability of catchment model parameters: Implications for climate impact analyses. Water Resources Research, 47 (2).

Moore, R.J., 1985. The probability-distributed principle and runoff production at point and basin scales. Hydrological Sciences Journal, 30 (2), 273-297.

Moore, R.J., 1999. Real-time flood forecasting systems: Perspectives and prospects. In Floods and landslides: Integrated risk assessment (pp. 147-189). Springer Berlin Heidelberg.

Oudin, L., Andréassian, V., Perrin, C., Michel, C., Le Moine, N., 2008. Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments. Water Resources Research, 44 (3).

Parajka, J., Blöschl, G., Merz, R., 2007a. Regional calibration of catchment models: Potential for ungauged catchments. Water Resources Research, 43 (6).

Parajka, J., Merz, R., Blöschl, G., 2007b. Uncertainty and multiple objective calibration in regional water balance modelling: case study in 320 Austrian catchments. Hydrological Processes, 21 (4), 435-446.

Parajka, J., Merz, R., Skøien, J. O., Viglione, A., 2015. The role of station density for predicting daily runoff by top-kriging interpolation in Austria. Journal of Hydrology and Hydromechanics, 63 (3), 228-234.

Rajib, M.A., Merwade, V., Yu, Z., 2016. Multi-objective calibration of a hydrologic model using spatially distributed remotely sensed/in-situ soil moisture. Journal of Hydrology, 536, 192-207.

Seibert, J., McDonnell, J.J., 2002. On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration. Water Resources Research, 38 (11).

Singh, R., Archfield, S.A., Wagener, T., 2014. Identifying dominant controls on hydrologic parameter transfer from gauged to ungauged catchments–A comparative hydrology approach. Journal of Hydrology, 517, 985-996.

Sugeno, M., 1985. An introductory survey of fuzzy control. Information sciences, 36 (1), 59-83.

Thornthwaite, C.W., 1948. An approach toward a rational classification of climate. Geographical review, 38 (1), 55-94.

van Emmerik, T., Mulder, G., Eilander, D., Piet, M., Savenije, H., 2015. Predicting the ungauged basin: Model validation and realism assessment. Frontiers in Earth Science, 3 (62).

Wang, D., Chen, Y., Cai, X., 2009. State and parameter estimation of hydrologic models using the constrained ensemble Kalman filter. Water resources research, 45 (11).

Wang, G.Q., Zhang, J.Y., Jin, J.L., Pagano, T.C., Calow, R., Bao, Z.X., Liu, C.S., Liu, Y.L., Yan, X. L., 2012. Assessing water resources in China using PRECIS projections and a VIC model. Hydrology and Earth System Sciences, 16 (1), 231-240.

Wi, S., Yang, Y.C.E., Steinschneider, S., Khalil, A., Brown, C.M., 2015. Calibration approaches for distributed hydrologic models in poorly gaged basins: implication for streamflow projections under climate change. Hydrology and Earth System Sciences, 19 (2), 857-876.

Wilk, M.B., Gnanadesikan, R., 1968. Probability plotting methods for the analysis for the analysis of data. Biometrika, 55 (1), 1-17.

World Meteorological Organization, 2008. Guide to Hydrological Practices Volume I, Hydrology – From Measurement to Hydrological Information. WMO-No. 168, Sixth edition.

Yapo, P.O., Gupta, H.V., Sorooshian, S. 1998. Multi-objective global optimization for hydrologic models. Journal of hydrology, 204 (1-4), 83-97.

Yunnan Bureau of Hydrology and Water Resources, 2005. Gutan River Ecological Hydropower Station Project Water Resources Appraisal, Nujiang Prefecture, Lushui [in Chinese]. Yunnan Bureau of Hydrology and Water Resources, Kunming, China.

Zhang, Z., Wagener, T., Reed, P., Bhushan, R., 2008. Reducing uncertainty in predictions in ungauged basins by combining hydrologic indices regionalization and multiobjective optimization. Water Resources Research, 44 (12).

Zhao, C., Yao, S., Zhang, S., Han, H., Zhao, Q., Yi, S., 2015. Validation of the accuracy of different precipitation datasets over Tianshan mountainous area. Advances in Meteorology.

CHAPTER 3:
STREAMFLOW PREDICTION UNDER EXTREME DATA-SCARCITY: A STEP TOWARD HYDROLOGIC PROCESS UNDERSTANDING WITHIN SEVERELY DATA-LIMITED REGIONS

3.1 Preface

This chapter provides a concise description of the modifications made to the SPED framework and focuses on comprehensive test of performance of the framework against data-intensive approaches in catchments located around the world. The content of this chapter has been submitted to Hydrological Sciences Journal[2] and is currently under review.

3.2 Abstract

Streamflow prediction in ungauged basins is necessary to support water resources management decisions. Herein we refine and evaluate the Streamflow Prediction under Extreme Data-scarcity (SPED) model, a framework designed for streamflow prediction within regions of sparse hydrometeorologic observation. With the SPED framework, inclusion of soft data directs optimization to balance runoff efficiency with selection of hydrologically-representative parameters. Here SPED is tested in catchments around the world, including four well-gauged catchments, by mimicking data-scarcity and comparing against data-intensive approaches. By differentiating equifinal models, SPED succeeds where traditional approaches are likely to fail: partially-dissimilar reference/target catchments. For instance, in a pair of reference/target catchments with different base flow regimes, SPED outperforms a model calibrated only to

[2] Alipour, M.H., Kibler, K.M., 2019 (under review). Streamflow Prediction under Extreme Data-scarcity: A Step Toward Hydrologic Process Understanding within Severely Data-limited Regions. Hydrological Sciences Journal.

maximize efficiency (NSE of 0.54 versus 0.08). SPED performs consistently (NSE range of 0.54-0.74) across the diverse climatological and physiographic settings tested and proves comparable to state-of-the-science methods that use robust data networks.

3.3 Introduction

As predictive abilities in catchment science become more advanced, methods are needed to translate new capabilities and technologies into wide application by water resources managers in diverse geographies. Ungauged catchments located in regions characterized by sparse hydrologic networks deserve particular attention from the research community, to ensure that the most rigorous and promising approaches to streamflow prediction may be applied broadly. For instance, 65% of mountainous basins do not meet the World Meteorological Organization (WMO) minimum recommended density of discharge gauging stations necessary for basic water resources management (Perks et al. 1996) and there is evidence that the number of long term stream gauges is in decline worldwide (Lanfear et al. 1999, Vörösmarty et al. 2001, Shiklomanov et al. 2002, Hannah et al. 2011). Lack of climatic and hydrologic data, such as long-term daily streamflow data, may lead to deficiencies in water resources management capabilities (see WMO 2008), and also presents particular challenges to application of advanced modeling techniques for streamflow prediction, such as regionalization and multi-criteria model calibration.

Prediction of streamflow in ungauged rivers has received significant attention from researchers around the world due to the vast applications in water resources planning and management (e.g., Gibbs et al. 2012, Seibert and McDonnell 2013, Arsenault and Brissette 2016, Kibler and Alipour 2017). Long-term, daily streamflow data are advantageous, for instance, in planning development projects, preservation or restoration of aquatic ecosystem function, and allocation of water. Substantial progress has been made in the science of streamflow prediction, for instance through the International Association of Hydrological Sciences Prediction in Ungauged Basins (PUB) Decade 2003-2012. The synthesized runoff prediction framework from this decade of research identifies catchments as complex and diverse systems, wherein varied streamflow generation processes manifest distinct hydrologic response signatures (see Blöschl et al. 2013). The role of comparative hydrology is emphasized as a tool to learn about catchment functionality based on similarities, for instance in physical characteristics or runoff signatures, such that hydrologic similarities may be utilized to predict streamflow in ungauged catchments (Wagener et al. 2013). Such initiatives are in part a response to the desire that research focus shift towards process understanding and model structural diagnostics (Hrachowitz et al. 2013). Based on the principles of comparative hydrology and catchment similarity, numerous flow prediction techniques have been developed and tested around the world (e.g., Post and Jakeman 1999, McIntyre et al. 2004, Post 2009, Parada and Liang 2010, Parajka et al. 2015). However, barriers to hydrologic modeling within severely data-scarce regions persist (Hughes 2016, Koutsouris et al. 2017, Tegegne et al. 2017), and managers struggle with problems such as equifinality of potential models (Beven 1993, 2006), uncertainty in information available to constrain parameter values (Alipour and Kibler 2018), limited selection of gauged reference catchments and thus potential for hydrologic dissimilarity to ungauged catchments (Peñas et al. 2014), and lack of streamflow data for validating predictions (van Emmerik et al. 2015).

Many state-of-the-science streamflow prediction techniques are founded upon concepts of regionalization, wherein the regional hydrologic network is utilized to develop relationships between catchment characteristics and/or runoff signatures and hydrologic model parameters (Zhang et al. 2008, Yadav et al. 2007). For example, McIntyre et al. (2005), Parajka et al. (2007a), Post (2009), Heřmanovský et al. (2017), and Swain and Patra (2017) developed regionalization techniques to estimate streamflows over large spatial areas. In each of these studies, methodologies were developed within regions equipped with robust hydrologic monitoring networks, which produced satisfactory streamflow predictions in the majority of catchments tested. Moreover, studies evaluating performance of regionalization techniques conclude that spatial proximity to locations of long-term monitoring is a primary predictor of model accuracy (e.g., Vandewiele and Elias 1995, Merz and Blöschl 2004, Parajka et al. 2005, Oudin et al. 2008). Thus, reliability of hydrologic prediction based on regionalization is closely associated with resolution of the available regional observation network (Oudin et al. 2008, Parajka et al. 2015).

Additionally, the value of multi-criteria calibration and evaluation of models has been highlighted as an avenue for improving process representation in hydrologic modeling (Yapo 1996, Yapo et al. 1998, Vrugt et al. 2003). Yu and Yang (2000), Seibert and McDonnell (2002), Yang et al. (2004), and Kamali and Mousavi (2014), for example have incorporated fuzzy theory or 'soft' data (i.e., data associated with high uncertainties such as regional/global data or qualitative knowledge) into multi-objective model calibration. Incorporating *a priori* (i.e., information acquired or estimated directly from physical/hydrological data and information in a catchment without the need for model calibration) predictions of parameter distributions in multi-objective calibration of hydrological models may be an especially promising direction for improving representation of catchment function within models (Parajka et al. 2007b; Merz et al. 2009, 2011). However, making such *a priori* estimations would be an additional challenge to streamflow prediction in data-scarce regions. Alipour and Kibler (2018) proposed the Streamflow Prediction under Extreme Data-scarcity (SPED) framework which incorporates *a priori* estimates of parameter values in the multi-criteria calibration of a hydrological model of choice. Preliminary testing suggested that even highly uncertain soft data available in data-scarce regions can support such *a priori* parameter estimates. While runoff prediction efficiency of SPED was similar to that of traditional single-objective methods, parameter values selected by SPED aligned more closely with *a priori* estimates based on soft data. This alignment suggests that the SPED process may allow managers to isolate parameter values that better represent hydrological processes in poorly gauged basins.

To further address the need for incorporating scientific advancements into practical techniques for poorly-gauged regions, in this study we conduct a comprehensive evaluation of the SPED framework to more fully explore its merits and limitations. The objectives of the study are thus to: 1) test the streamflow prediction skill of the SPED framework in diverse regions with different hydro-climatological conditions, 2) analyze whether multi-criteria SPED offers improvement in runoff efficiency over models calibrated only to maximize runoff efficiency, and 3) compare the quality of SPED simulations generated under severe data-scarcity to those achieved in prior studies utilizing robust data networks (e.g., availability of gauging stations for hydrometeorological variables, availability of multiple gauged catchments in the region, availability of data required for physically-distributed hydrologic modeling) and state of the art prediction methods. The current study makes important contributions to the study of streamflow prediction in ungauged basins by diversifying both the number and geography/hydrology of validation test cases for the SPED framework. By comparing SPED performance with prior models in well-gauged catchments, the limitations of SPED can be well understood. Moreover, the merits of SPED at addressing wicked problems in hydrologic modeling, such as equifinality, catchment dissimilarity and data uncertainty, are comprehensively explored.

3.4 Materials and Methods

3.4.1 Synopsis of the SPED Framework

The SPED framework proposed by Alipour and Kibler (2018) is not itself a model, but a systematic procedure within which any number of hydrologic models may be embedded. Before application of SPED, the choice of hydrologic model for general suitability in the region of interest should be preliminarily tested by traditional calibration (with the single objective to maximize runoff efficiency) and validation in a local, gauged reference catchment. A reference catchment is a gauged catchment located in relative proximity to the target catchment(s), which has some degree of hydrologic similarity to target catchment(s). The target catchment is an ungauged catchment in which streamflows are to be predicted.

After a model is confirmed to be generally suitable for the region, available physical and hydrologic data are used to create *a priori* estimates of influential model parameters in the reference catchment (Fig. 13a). In the absence of sufficient observed or ground-truthed data, 'soft' data, including low-resolution or highly uncertain data, such as remotely sensed data, may be utilized to inform *a priori* estimates. Uncertainties are incorporated by representing *a priori* estimates into the model as fuzzy numbers (Fig. 13a). A multi-criteria objective function (i.e., Eq. 9) is parsed to balance both agreement with *a priori* parameter estimates in the reference catchment (criteria 1 to n) and conformity to observed streamflows (criterion n+1) (Fig. 13b). Similar to the procedure performed in the reference catchment, *a priori* values of influential model parameters are estimated in the target catchment (Fig. 13c) in order to weight the criteria. Criteria are weighted based on their relative importance, level of certainty in *a priori* parameter estimates, and degree of resonance between *a priori* estimates in the reference and target catchments (Fig. 13c). Subjectivity in weighting may be unavoidable and uncertainties in criteria weights are acknowledged through the use of fuzzy weights (Fig. 13c). The model is calibrated to optimize the multi-criteria objective function in the reference catchment using *a priori* parameter estimates in the reference catchment and the weights assigned to the criteria (Fig. 13d). Through this multi-objective process, calibration aims at maximizing runoff efficiency while at the same time providing a true representation of physical and hydrological characteristics of the catchment. Finally, the selected model is transferred to the target catchment(s) for prediction and validation (Fig. 13e). Detailed examples of SPED implementation are provided in Section 3.4.3.

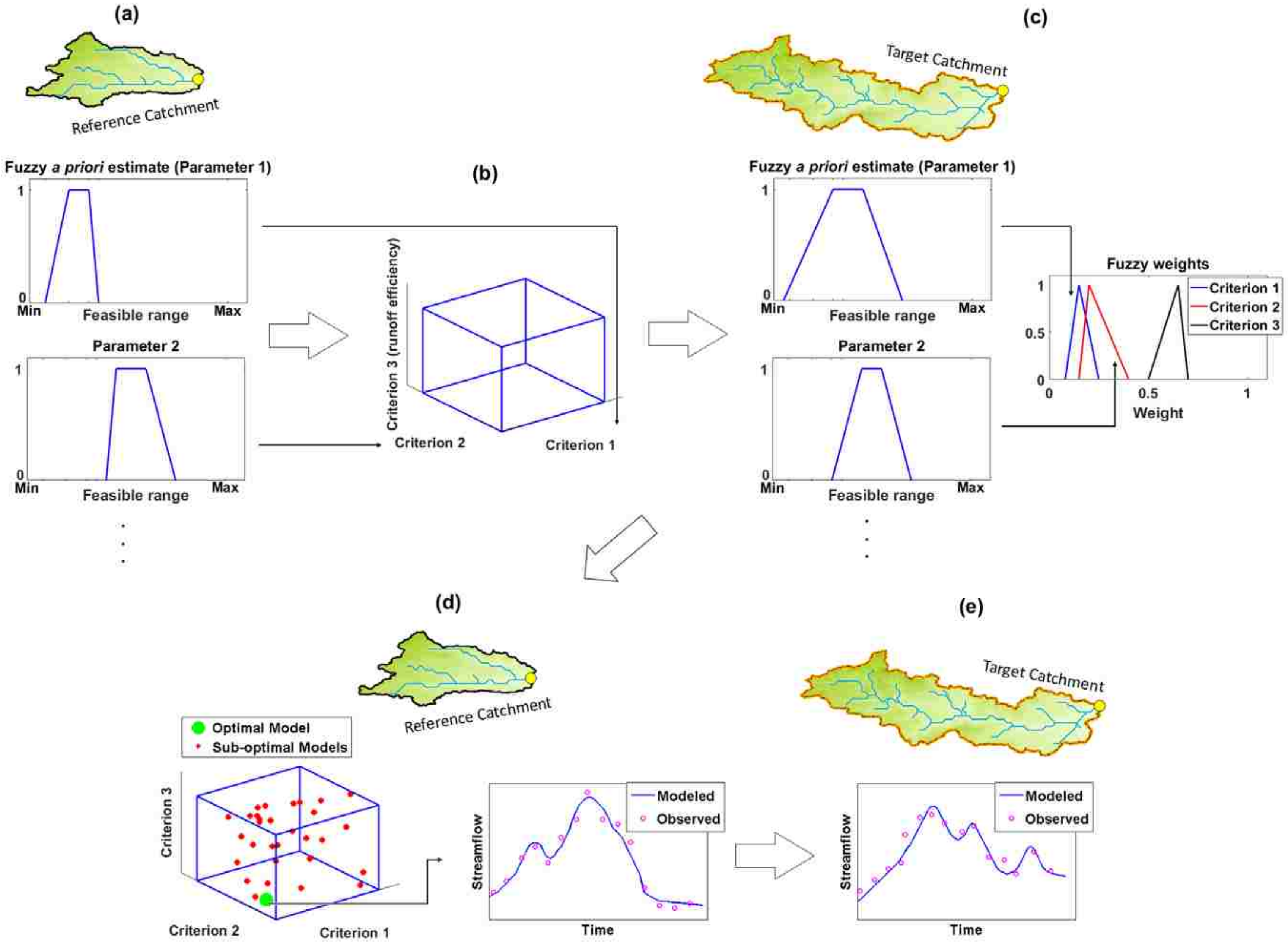

Figure 13. Application of the SPED procedure; a) parameter values are estimated *a priori* in the reference catchment; b) multi-criteria objective function is formed in the reference catchment; c) parameter values are estimated *a priori* in the target catchment and criteria are weighted; d) from multiple potential models (red points) the model with lowest OF value (Eq. 9) is selected as optimal (green point); e) the optimal model selected from the reference catchment is transferred to the target catchment(s) for prediction.

3.4.2 Study Catchments

We applied the SPED framework to predict streamflows in six target catchments from four countries (United Kingdom, United States, Australia, and China) located on four different continents and with diverse hydro-climatic regimes. The first target catchment, River Coquet (gauge at Morwick) (Table 7, Fig. 14a), with a size of 578 km^2 is located in the UK. Annual mean rainfall in the catchment is 850 mm (1961-1990). There is association between precipitation in the River Coquet basin and low pressure over Britain, and precipitation is characterized by a cyclonic weather type. In River Coquet basin, easterly air streams may cause an onshore flow off the North Sea and consequently generate river streamflow (Lavers 2011). Land cover in the catchment is dominated by grasslands while woodland, arable/horticultural, and mountain/heath/bog are the other major land cover types in the catchment. We chose a nearby gauged catchment, River Wansbeck (gauge at Mitford) (Table 7, Fig. 14a), with a size of 282 km^2 as the reference catchment. The River Wansbeck catchment has a mean annual rainfall of 794 mm (1961-1990) and a similar land cover to River Coquet (UK National River Flow Archive 2017). Flows in River Coquet (Table 7, Fig. 14a), have previously been modeled at Morwick using regionalization techniques in a study by McIntyre et al. (2004).

The North Fork Cache Creek watershed in California (Table 7, Fig. 14b) with a size of 510 km^2 is the second target catchment. Winter cyclonic storms create most of the precipitation in the catchment. Surface runoff corresponds with rainfall events which normally begin in November and occur frequently through mid-April. Very dry conditions accompanied by very low streamflow prevail for the rest of the year. Mean annual precipitation of the study area for the period of analysis is 880 mm (Parada and Liang 2010). Woodlands followed by wooded grasslands/shrublands are the major land cover types in the catchment (DeFries and Hansen 2010). We chose one gauged catchment, Eel River (below Scott Dam) (Table 7, Fig. 14b) with a size of 751 km^2 as the reference catchment. Woodlands are the single major land cover in the catchment (DeFries and Hansen 2010). North Fork Cache Creek has previously been modeled by Parada and Liang (2010).

Two locations within Broken River were selected as interchangeable reference and target catchments in the dry-tropical rangeland environment of Burdekin catchment (129,660 km^2) in Queensland in Australia. At Old Racecourse (Table 7, Fig. 14c) catchment area is 68 km^2 and further downstream at Urannah (Table 7, Fig. 14c) catchment area is 1,033 km^2. Burdekin catchment is characterized by a semi-arid climate where coldest month temperatures average above 0°C. Such climatic regime is not suitable for agriculture and is home to low population densities. The catchment routinely experiences large variabilities in its annual, mean monthly, as well as daily streamflows. Mean annual rainfall in the Burdekin catchment is 650 mm (Australian Bureau of Meteorology 2017). Wooded grasslands/shrublands dominate the land cover of catchment areas contributing to Urannah while woodlands are the major land cover at Old Racecourse (DeFries and Hansen 2010). Burdekin catchment was previously modeled through the regionalization study by Post (2009), including both Broken River at Old Racecourse and Broken River at Urannah, and prediction results were generated for both catchments. Because it was possible to compare prior modeled predictions in both catchments to results from SPED, both are modeled herein as interchangeable reference/target catchments. This was not possible in the other study locations.

Finally, two target catchments, including Yang Bijiang (YBJ) River catchment (Table 7, Fig. 14d) in the Upper Mekong River basin and Laowo River catchment (Table 7, Fig. 14d) in the Upper

Salween River basin in China were selected to analyze SPED performance in a truly poorly-gauged region. Yongchun River (Table 7, Fig. 14d), a gauged tributary to the Upper Mekong River, was selected as the reference catchment for both Chinese rivers. The regional climate is monsoonal, although precipitation and temperature vary considerably with local topography, which is mountainous and highly varied. In the rivers, high flows correspond with rainfall which is characterized by two seasonal pulses (between February-May, and between June-October). Annual rainfall in the Upper Mekong basin can range from 600 mm in the Tibetan Plateau to 1,700 mm in the mountains of Yunnan (Mekong River Commission 2018). In the Upper Salween basin, the annual precipitation ranges from 400 mm to 2000 mm and averages at 900 mm (Zhou et al. 2017). Land cover in the Upper Salween River and Upper Mekong Rivers basins is similar where forests and other types of vegetation dominate the catchments, followed by a substantially smaller portion of croplands (DeFries and Hansen 2010).

Table 7. Summary of reference and target catchments

Country	Catchment (location)	Catchment type	Catchment size (km^2)	Streamflow data source	Precipitation data source	Temperature data source
United Kingdom	River Wansbeck (at Mitford)	Reference	282	UK National River Flow Archive (2017)	version 16.0 of E-OBS gridded dataset (Haylock et al., 2008); resolution 0.25°	
	River Coquet (at Morwick)	Target	578			
United States	Eel River (below Scott Dam)	Reference	751	USGS (2017)	NOAA/OAR/ESRL PSD (2017); rainfall resolution 0.5°; temperature resolution 0.25° (Fan and Van den Dool, 2008)	
	North Fork Cache Creek (near Lower Lake)	Target	510			
Australia	Broken River (at Old Racecourse)	Reference/ Target	68	Australian Bureau of Meteorology (2017)	Australian Bureau of Meteorology; resolution 0.05°	NOAA/OAR /ESRL PSD (2017); resolution 0.25° (Fan and Van den Dool, 2008)
	Broken River (at Urannah)	Target/Reference	1,033			
China	Yongchun River (at Tangshang gauge)	Reference	197	Chinese Ministry of Hydrology (1970; 1971; 1974; 1977; 1982)	APHRODITE; Yatagai et al. (2012); Yasutomi et al. (2011); resolution 0.25°; precipitation data were bias corrected using inverse distance weighting to locations of observed precipitation (see Alipour and Kibler 2018)	
	Laowo River (above Laowo Dam)	Target	575	Henck et al. (2010)		
	YBJ River (near Yiqianpu)	Target	4,300	Henck et al. (2010)		

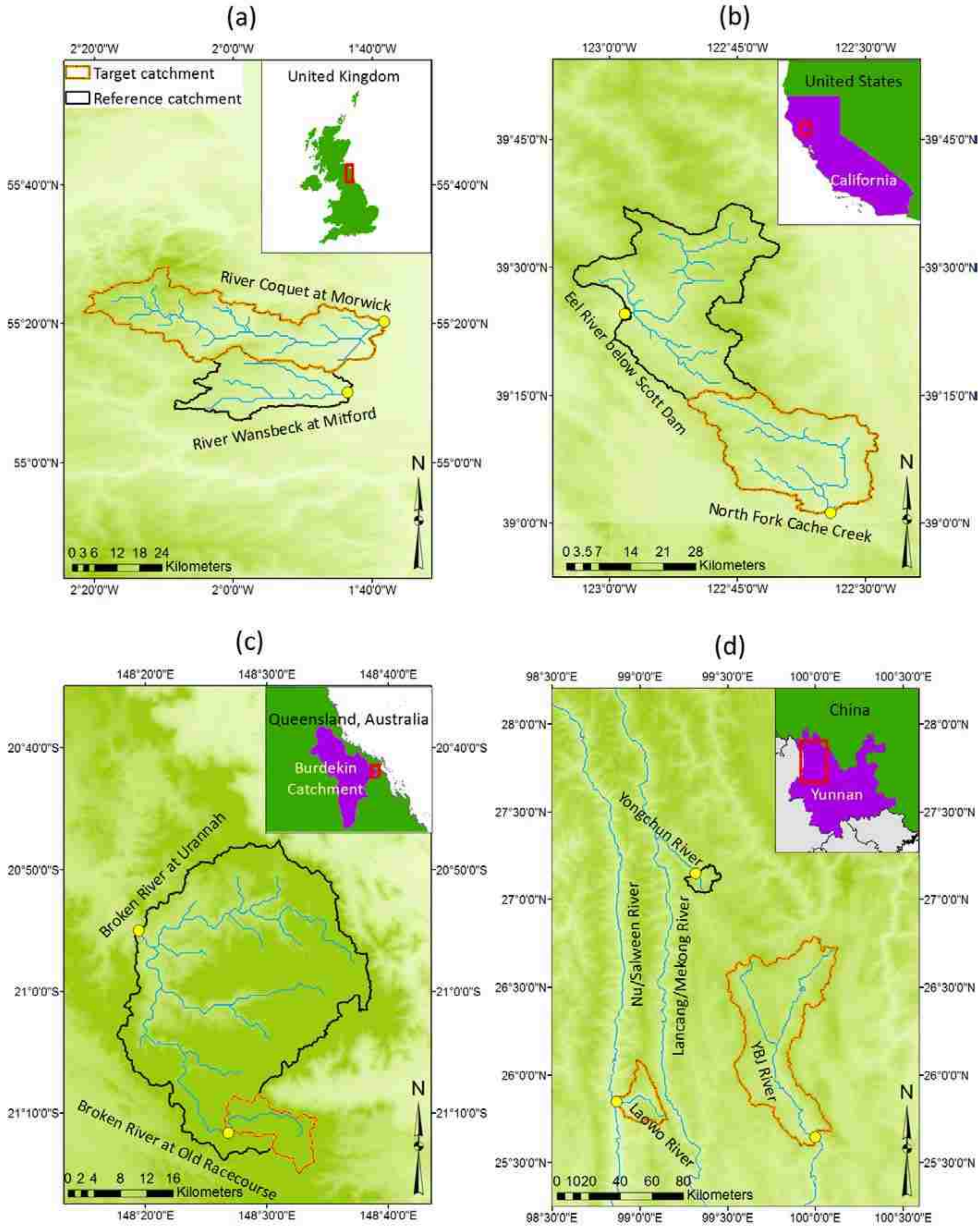

Figure 14. Study regions and reference/target catchments; a) United Kingdom, b) United States, c) Australia, d) China.

3.4.3 Application of SPED in Test Catchments

We perform preliminary validation within reference catchments to confirm suitability of the hydrologic model of choice (HyMOD, Moore 1985, 1999; Wang et al. 2009) in each region of study. Following Alipour and Kibler (2018), preliminary testing consists of traditional single-objective model calibration and validation against data observed within gauged reference catchments. Thus, calibration is performed to maximize runoff efficiency:

$$NSE = 1 - \frac{\sum_{t=1}^{T}\left(Q_o^t - Q_m^t\right)^2}{\sum_{t=1}^{T}\left(Q_o^t - \overline{Q_o}\right)^2} \qquad (6)$$

Where NSE is Nash-Sutcliffe runoff efficiency, Q_o^t is observed streamflow in time step *t*, Q_m^t is modeled streamflow in time step *t*, and $\overline{Q_o}$ is the mean value of observed streamflows in time *T*.

The hydrological model of choice, HyMOD, is a lumped conceptual model that predicts flow based on simulated probability distributions of soil moisture across a catchment (Wang et al., 2009). The distribution of soil moisture capacity across the catchment is represented using a cumulative distribution function (CDF):

$$F(c) = 1 - \left[1 - \frac{c}{C_{max}}\right]^B, 0 \leq c \leq C_{max} \qquad (7)$$

Where c is soil moisture capacity, C_{max} is the maximum soil moisture capacity of the catchment, and B is a shape factor that defines the degree of spatial variability in soil moisture capacity across the catchment. Other parameters of HyMOD (besides C_{max} and B) include R_q (inverse of residence time in quick reservoirs), R_s (inverse of residence time in a slow reservoir), and α (a fraction coefficient for distribution of water between slow and quick reservoirs). Interested readers are referred to Wang et al. (2009) and Moore (1985, 1999) for a more comprehensive description of HyMOD. We use a modified formulation of the 5-parameter HyMOD model, including addition of a minimum soil moisture capacity parameter, C_{min}, which may be greater than zero, to the HyMOD CDF of soil moisture capacity (c) (see Jayawardena and Zhou (2000), Post (2009), and Wang (2018), for similar modifications to soil moisture capacity modeling). The new soil moisture capacity CDF is formulated as:

$$F(c) = 1 - \left[1 - \frac{c - C_{min}}{C_{max} - C_{min}}\right]^{B}, C_{min} \leq c \leq C_{max}, 0 \leq C_{min} \leq C_{max} \quad (8)$$

Parameters B, C_{max}, and C_{min} define the capacity of soil for storing water. For instance, a high value for C_{min} indicates that during dry periods, even a large rainfall event may not result in initiation of saturation excess overland flow and consequently increased river flows. A high value for C_{max} indicates more stable base flows, fewer days with zero flow, and higher constancy of flow. For B, a parameter describing the shape of the soil moisture CDF, a value between zero and one indicates a soil moisture capacity CDF that is convex, while values greater than one indicate a concave CDF.

The new formulation of HyMOD in this study was coded into MATLAB (Mathworks 2016) in a configuration that can process 62,500 different combinations of model parameters within 10 minutes on a typical processor (Intel Core i7-5500U at 2.40 GHz). In each reference catchment, we first calibrated HyMOD to maximize NSE by running 62,500 different parameter combinations through the branch-and-bound method described by Alipour and Kibler (2018) and validated with data from a different time period.

After confirming suitability of the model in each region, the SPED process was applied in each target-reference catchment pair. To mimic circumstances within severely data-scarce regions, we utilize regional-scale gridded databases to estimate precipitation and temperature in each study location (Table 7). Thornthwaite's approach (1948) is used to estimate potential evapotranspiration (PET) from temperature data. Daylight length is estimated using the model of Forsythe et al. (1995), which uses latitude and day of the year for its estimations. Global data on total available water capacity (WISE30sec dataset, Batjes 2015) and depth to bedrock (SoilGrids250m dataset, Hengl et al. 2017) were combined to estimate spatial variability in soil moisture capacity (cm) in each reference/target catchment at a resolution of about 0.06 km^2. From this spatial dataset, an empirical CDF of soil moisture capacity was constructed for each catchment (Fig. 15), to which we fitted the HyMOD soil moisture capacity CDF (Eq. 8). *A priori* parameter values of B, C_{max}, and C_{min} were thus estimated for each catchment using only globally-available data, as the best HyMOD CDF fit to the catchment's own empirical CDF (Fig. 15). Given the substantial uncertainties associated with global-scale coverage of soil properties, we consider these data as 'soft' data and estimate *a priori* parameter values in the form of trapezoidal fuzzy numbers. Soil moisture capacity distributions estimated by such highly uncertain data may contain errors. In some regions for example, the empirical CDF may exhibit a threshold behavior, such as near-vertical increase in the value of c as the CDF approaches the upper end of the distribution, C_{max}. Such behavior is consistent with likely presence of anomalous data contained in the highly uncertain soil databases. Application of quality assurance procedures to identify and disregard spurious data are recommended. We recommend that when the ratio of the top 5% of the distribution to the full distributional range exceeds a value of around 0.4, the 95th percentile value of the distribution should be selected as C_{max}.

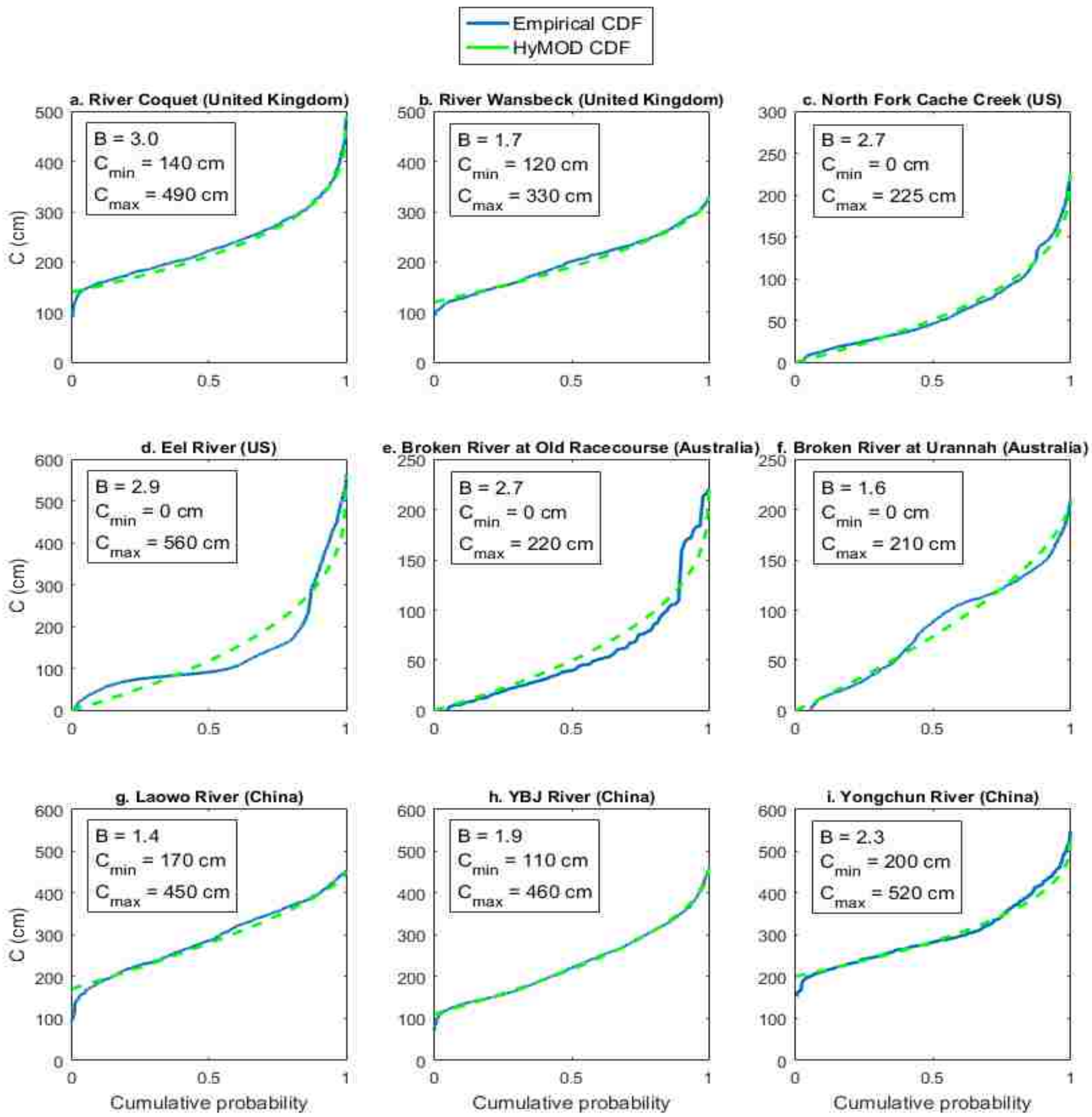

Figure 15. Soil moisture capacity (c) and parameter values estimated *a priori* from soft data (B, C_{max}, and C_{min} values are for fitted HyMOD CDFs to the empirical CDFs)

A priori values for parameters B, C_{max}, and C_{min} were incorporated into a four-part multi-criteria objective function (Eq. 9, similar to the three-part objective function developed by Alipour and Kibler (2018)), used to calibrate the model in each reference catchment. The multi-criteria objective function simultaneously maximizes runoff efficiency (minimizes 1-NSE) while minimizing the difference between *a priori* and calibrated estimates for each parameter. Since at this time we were unable to provide *a priori* estimates for the other HyMOD parameters (R_q, R_s and α), they were incorporated into the objective function only by their contribution to maximizing runoff efficiency. The four criteria (maximized NSE, calibrated values of B, C_{max}, and C_{min} that are similar to *a priori* estimates) were weighted based upon their relative importance, level of certainty in *a priori* parameter estimates, and degree of similarity between *a priori* estimates in the reference and target catchments (Fig. 13c). For instance, if the *a priori* estimate for parameter B (or C_{max}, or C_{min}) in a reference catchment was similar (dissimilar) to that in the corresponding target catchment, a higher (lower) weight was assigned to the criterion associated with that parameter. By comparing of empirical CDFs in paired reference-target catchments, degree of catchment similarity with respect to influential parameters can be discerned. If reference and target catchments were quite similar (dissimilar) with respect to a given parameter, we were able to signify (de-signify) the importance of calibrating this parameter close to its *a priori* value in the reference catchment through weighting. Since subjectivity in weighting may be unavoidable and the *a priori* estimates were associated with high uncertainties, we used triangular fuzzy weights to partially account for uncertainty in criteria weighting.

The feasible ranges of C_{max} (typically between 5-8000 mm), C_{min} (typically between 0-1500 mm), and B (typically between 0-6) were tailored to each study area based on ranges suggested by *a* *priori* parameter estimates in each region. To normalize objectives associated with *a priori* estimates of parameter values, we divided the difference between the *a priori* estimate and calibrated value of a parameter with half the feasible range of that parameter (Eq. 9). If the resulting value was greater than 1, we assigned a value of 1 to that objective (without application of its weight):

$$OF = \sum_{i=1}^{n} \left| defuzz\left\{ \widetilde{W}_i * \frac{\tilde{a}_i - X_i}{\frac{X_{max_i} - X_{min_i}}{2}} \right\} \right| + |defuzz\{W_{n+1} * (1 - NSE)\}| \qquad (9)$$

$$If\ defuzz\left\{ \frac{\tilde{a}_i - X_i}{\frac{X_{max_i} - X_{min_i}}{2}} \right\} > 1 \quad then \quad \widetilde{W}_i * \frac{\tilde{a}_i - X_i}{\frac{X_{max_i} - X_{min_i}}{2}} = 1$$

Where $\tilde{a}_i$ is the estimated trapezoidal fuzzy value of parameter i, X_i is the estimated value for parameter i by model calibration, X_{max_i} and X_{min_i} are the upper and lower feasible limits for parameter i, $\widetilde{W}_i$ is the fuzzy weight assigned to criterion i, W_{n+1} is the weight assigned to criterion n + 1, and defuzz denotes the defuzzification process through the centroid method (Sugeno 1985). Through this process, calibrated values that diverge sharply from *a priori* estimates are excluded and the optimization process is directed towards values that more closely match *a priori* parameter estimates, while balancing high runoff efficiency.

The multi-criteria objective function (Eq. 9) was optimized in each reference catchment by analyzing 62,500 different parameter combinations, using the same branch-and-bound method as applied in single-objective calibration. The value of the optimization objective function (OF value) is normally between 0 and 1 (OF can be greater than 1), where zero indicates streamflow predictions which perfectly mirror observed values as well as calibrated parameter values which perfectly mirror *a priori* estimates. The parameter values estimated through multi-objective calibration in each reference catchment are transferred to target catchment(s) for streamflow prediction.

3.4.4 Assessment of SPED Predictions

We evaluate predictive skill of SPED in the six target catchments, as compared to performance of models selected by traditional single-objective calibration. We perform single-objective calibration, to maximize runoff efficiency, in reference catchments and transfer models for prediction in all target catchments (Fig. 16a and b). SPED performance is evaluated relative to that of the single-objective model through comparisons of model prediction accuracy and model parameter residuals (Eq. 10) for influential parameters (B, C_{max} and C_{min}). To evaluate accuracy of model predictions, we compare NSE for daily streamflow and quantitative flow metrics indicating flow magnitude (3-day minimum and maximum, total runoff volume), and frequency and duration of low flows (average number of days below Q_{75} of observed flow) and high flows (average number of days above Q_{25} of observed flow) (Richter et al. 1997).

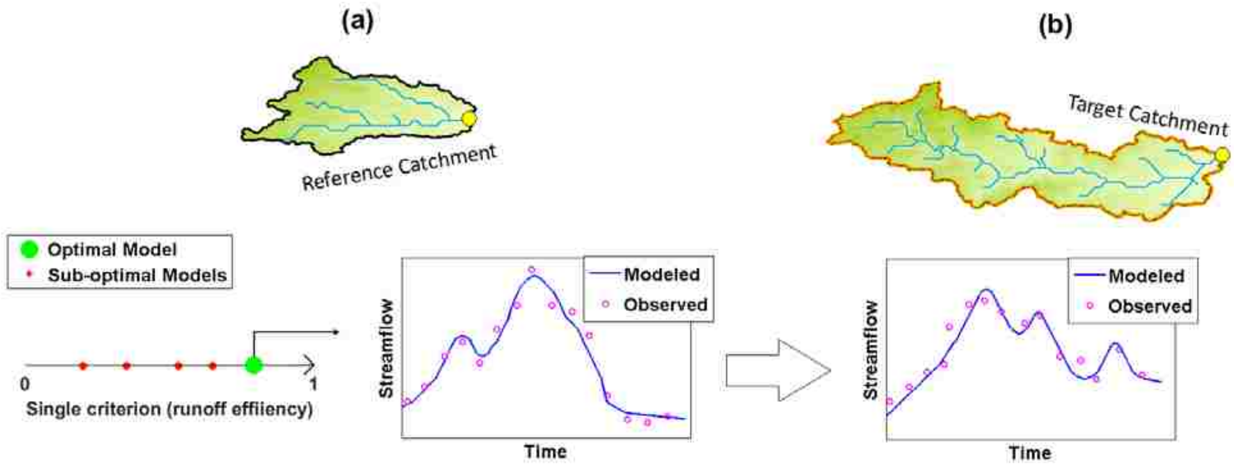

Figure 16. Application of single-objective calibration to runoff efficiency; a) from multiple potential models (red points) the model with greatest NSE is selected as optimal (green point); b) the optimal model selected from the reference catchment is transferred to the target catchment(s) for prediction.

Parameter residuals indicate how well the empirical CDF for soil moisture capacity in a catchment is represented by different modeling approaches. We note here that empirical CDFs are created using data associated with high uncertainties, and thus are perhaps themselves imperfect representations of true catchment condition. However, in areas of the world lacking more complete data, such soft data may still improve understanding of true catchment condition. Thus, inclusion of soft data in modeling has been recommended over rote calibration to only maximize runoff efficiency (e.g., Seibert and McDonnell (2002)). Parameter residuals for comparison between models are calculated as follows:

$$P_{res} = P_{a\ priori} - P_{calibrated} \tag{10}$$

Where P_{res} is the parameter residual, $P_{a\ priori}$ is the centroid of *a priori* parameter values estimated with soft data, and $P_{calibrated}$ is the calibrated parameter value.

Finally, we compare the runoff efficiencies of SPED simulations in the United Kingdom, United States, and Australia, which are generated under simulated severe data-scarcity, to those achieved in prior studies utilizing robust data networks and state-of-the-art prediction methods. The transformation of data-rich regions into synthetically data-scarce regions was accomplished by assuming availability of only one gauged reference catchment and by using lower-quality (regional/global scale) data to estimate precipitation and temperature (Table 7). By contrast, prior studies had utilized the full suite of information available in these well-gauged regions (e.g. multiple gauged reference catchments and high-quality observed climatic data). In each catchment, we modeled the same time periods presented in prior studies.

3.5 Results

3.5.1 Prediction Performance of SPED

With respect to runoff efficiency, streamflow predictions generated by the SPED framework in six geographically diverse target catchments are acceptable (NSE range of 0.54-0.74, Table 8). Since the single-objective model is calibrated to only maximize runoff efficiency in the reference catchments, calibration performance of the single-objective model with respect to NSE is expectedly always greater than that of SPED (NSE range of 0.61-0.82 versus 0.48-0.79, Table 9). However, prediction performance of the two models are comparable and even sometimes substantially higher by SPED (Table 8). Performance of SPED and the single-objective model is comparable in River Coquet for the period of October 1989 – September 1994 (both NSE of 0.63,

Fig. 17a, Table 8). Comparison of flows simulated by the single-objective model and SPED against observed data (Fig. 17a, Table 10) indicates that the single-objective model better represents low flow magnitudes (29% error for 3-day minimum) than SPED (-52%). While both models do not estimate extreme high flows very well, these events are slightly better modeled by the single-objective model (Fig. 17a). Moderate and high flows, on other hand, are more accurately predicted by SPED (-14% error for total runoff volume) than the single-objective model (-22%). While both models perform comparably in estimating high flow duration, low flow duration is overestimated by the single-objective model (229% error versus -14% for SPED).

In North Fork Cache Creek (Fig. 17b), during the validation period, January 1950 – August 1955, SPED substantially outperforms the single-objective model (NSE of 0.54 versus 0.08, total runoff volume residual of 3.30e+08 m^3 versus -7.68e+08 m^3, Table 8 and 10). Notably, the single-objective model predicts only 26% of the observed runoff volume. While both models struggle to estimate magnitudes of extreme high flows (Fig. 17b), moderate to high flows are better predicted by SPED (-55% error versus -87%, for 3-day maximum). Three-day minimum flows are better represented by the single-objective approach (Table 10), however the single-objective model overestimates duration of low flows (78% error) while SPED overestimates low flows such that duration below Q_{75} is predicted to be zero days. The case is however opposite for high flow duration, where SPED overestimates high flow duration (75% error) and the single-objective model estimates a duration lower than observed (-44% error).

For the 10-year period of September 1975 – August 1985, SPED and the single-objective model perform comparably in Broken River both at Old Racecourse (NSE of 0.74 versus 0.79, Fig. 17c, Table 8) and at Urannah (NSE of 0.71 versus 0.78, Fig. 17d, Table 8). At Old Racecourse, most flow metrics and the flow duration curves (FDCs) are modeled comparably by SPED and the single-objective approach (Table 10 and Fig. 17c). At Urannah, comparison of flows simulated by the single-objective model and SPED against observed data (Fig. 17d, Table 10) indicate a better performance by SPED in modeling high flows and total runoff volume (3% error for 3-day maximum, and 54% error for total runoff volume) than the single-objective model (27% and 72%). The single-objective model however performs considerably better in modeling extreme high flows (Fig. 17d) as well as low flows (695% error for 3-day minimum) than SPED (2670%).

In Laowo river during January 1987 – December 1987 (NSE of 0.71 versus 0.76, Fig. 17e, Table 8) and in YBJ river for the 20-year period of test (1962–1973, 1975–76, 1978–1980, 1984–85, and 1987) (NSE of 0.72 versus 0.76, Fig. 17f, Table 8), there is little difference in performance of SPED and single-objective models. In Laowo River, where only one year of observed data is available, FDCs indicate a better performance by SPED with respect to the majority of flow sizes (bottom 90%) while the very largest flows are better modeled by the single-objective model (Fig. 17e). In YBJ, the single-objective model better represents low flow magnitudes (9% error for 3-day minimum flow) and durations (37% error for low flow duration) than SPED (96% and 77%, respectively). Performance of the models is similar with respect to 3-day maximum flow (-21% error for SPED versus -10% for the single-objective model), total runoff volume (12% error for SPED versus 1% for the single-objective model) and high flow duration (-46% error for SPED versus -50% for the single-objective model). Similar to Laowo, the very largest flows (top 10%) in YBJ are better modeled by the single-objective model (Fig. 17f).

Table 8. Prediction performance of SPED and the single-objective model with respect to runoff efficiency and parameter residuals

Target catchment	Metric	Single-objective model	SPED
River Coquet at Morwick (United Kingdom)	NSE	0.63	0.63
	OF value	0.52	0.39
	B residual	-2.83	0.33
	C_{max} residual (mm)	3206	2465
	C_{min} residual (mm)	347	-14
North Fork Cache Creek (United States)	NSE	0.08	0.54
	OF value	0.77	0.37
	B residual	-1.47	-0.2
	C_{max} residual (mm)	1084	788
	C_{min} residual (mm)	-1267	-17
Broken River at Old Racecourse (Australia)	NSE	0.79	0.74
	OF value	1.18	0.22
	B residual	1.37	1.98
	C_{max} residual (mm)	557	779
	C_{min} residual (mm)	-1072	-3
Broken River at Urannah (Australia)	NSE	0.78	0.71
	OF value	1.17	0.22
	B residual	-0.31	0.90
	C_{max} residual (mm)	583	435
	C_{min} residual (mm)	-1086	-3
Laowo River (China)	NSE	0.76	0.71
	OF value	1.32	0.27
	B residual	-3.65	-0.66
	C_{max} residual (mm)	2227	302
	C_{min} residual (mm)	420	-154
YBJ River (China)	NSE	0.76	0.72
	OF value	1.28	0.28
	B residual	-3.39	-0.40
	C_{max} residual (mm)	2070	441
	C_{min} residual (mm)	-123	-696

Table 9. Calibration performance of SPED and the single-objective model with respect to runoff efficiency

Country	Reference catchment	Single-objective model (NSE)	SPED (NSE)
United Kingdom	River Wansbeck (at Mitford)	0.61	0.60
United States	Eel River (below Scott Dam)	0.62	0.48
Australia	Broken River at Old Racecourse	0.82	0.74
	Broken River at Urannah	0.82	0.79
China	Yongchun River (to transfer to Lawo River)	0.79	0.69
	Yongchun River (to transfer to YBJ River)	0.79	0.70

Table 10. Observed flow metric values versus modeled values by the single-objective approach and SPED in target catchments (For SOM* and SPED the first value under each river is the absolute value and the second value is percent change with respect to the observed value; *Single-Objective Model)

Metric	Model	River Coquet		North Fork Cache Creek		Broken River at Old Racecourse		Broken River at Urannah		YBJ River	
3-day minimum (m^3/s)	Observed	0.92		0.02		0.02		0.23		9.10	
	SOM	1.19	29	0.04	100	0.50	2400	2.45	965	9.91	9
	SPED	0.44	-52	1.41	6950	0.50	2400	6.37	2670	17.87	96
3-day maximum (m^3/s)	Observed	68.90		100.10		26.89		238.20		387.80	
	SOM	61.49	-11	13.02	-87	31.40	17	302.70	27	347.70	-10
	SPED	62.31	-10	45.10	-55	19.86	-26	246.00	3	307.90	-21
Total runoff volume (m^3)	Observed	1.14e+09		1.04e+09		4.64e+08		4.06e+09		4.44e+10	
	SOM	8.89e+08	-22	2.72e+08	-74	4.71e+08	2	7.00e+09	72	4.49e+10	1
	SPED	9.77e+08	-14	1.37e+09	32	5.67e+08	22	6.25e+09	54	4.98e+10	12
Low flow duration (days)	Observed	11.53		63.30		13.08		12.29		25.13	
	SOM	37.92	229	112.6	78	0	-100	17.3	41	34.49	37
	SPED	9.87	-14	0	-100	0	-100	0	-100	5.88	-77
High flow duration (days)	Observed	8.98		19.53		15.54		10.42		22.77	
	SOM	6.94	-23	10.87	-44	6.09	-61	80.23	670	11.46	-50
	SPED	6.33	-30	34.12	75	6.46	-58	122.95	1080	12.39	-46

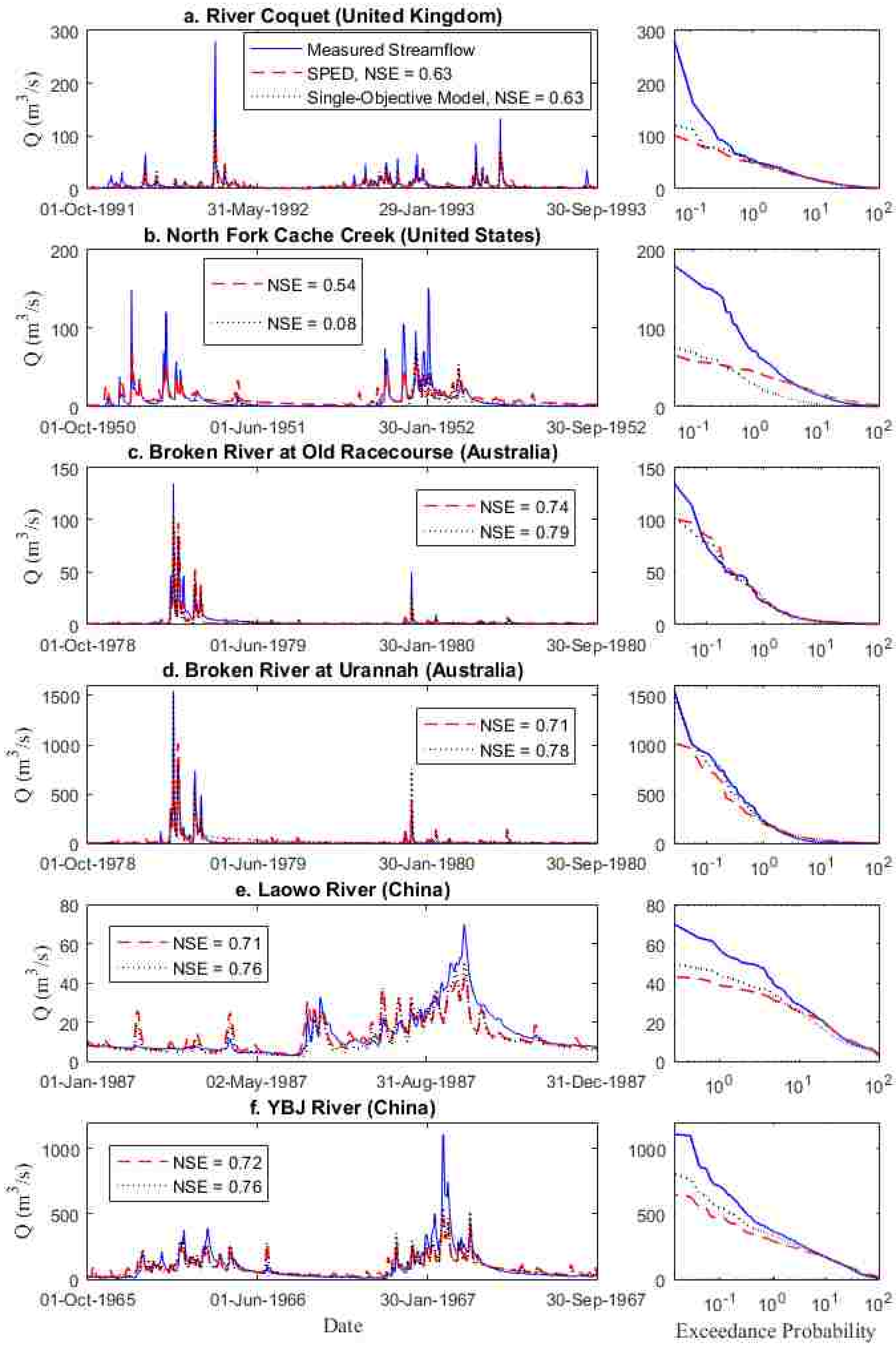

Figure 17. Flow Duration Curves (FDCs) and example validation hydrographs for portions of each prediction period in a) River Coquet at Morwick, b) North Fork Cache Creek, c) Broken River at Old Racecourse, d) Broken River at Urannah, e) Laowo River, and f) YBJ River. Runoff efficiencies displayed are relative to the entire prediction periods.

3.5.2 Parameter Estimation Performance

As SPED is designed to select parameter sets which are similar to *a priori* estimates while the single-objective model only aims at maximizing runoff efficiency, soil moisture capacity CDFs modeled by SPED exhibit closer resonation with empirical CDFs than the CDFs modeled by the single-objective approach in all six target catchments (Fig. 18). However, in some of the catchments, in particular River Coquet (Fig. 18a), the CDF modeled by SPED still displays significant difference from the empirical CDF. All three parameters of B, C_{max} and C_{min} are modeled with lower residuals by SPED than the single-objective approach in River Coquet (residuals of 0.33 versus -2.83 for B, 2465 versus 3206 mm for C_{max}, and -14 versus 347 mm for C_{min}, Table 8), North Fork Cache Creek (residuals of -0.2 versus -1.47, 788 versus 1084 mm, and -17 versus -1267 mm respectively, Table 8), and Laowo River (residuals of -0.66 versus -3.65, 302 versus 2227 mm, and -154 versus 420 mm respectively, Table 8). In Broken River at Urannah, SPED performs better in modeling C_{max} (residual of 435 mm versus 583 mm) and C_{min} (residual of -3 mm versus -1086 mm) but B (residual of 0.90 versus -0.31) is better modeled by the single objective approach. Similarly, in YBJ two parameters exhibit lower residuals when modeled by SPED (residuals of -0.4 versus -3.39 for B and 441 versus 2070 mm for C_{max}, while residual of -696 versus -123 mm for C_{min}, Table 8). In Broken River at Old Racecourse, despite that the single-objective modeling yields lower residuals than SPED for the two parameters of B (residuals of 1.98 versus 1.37, Table 8) and C_{max} (residuals of 779 versus 557 mm, Table 8), SPED does a much better job in modeling C_{min} (residual of -3 versus -1072 mm) so that its CDF is closer to the empirical CDF than the single-objective model. Overall, OF values summarizing performance in terms of both runoff efficiency and soil moisture capacity modeling are lower (with zero as the ideal value) for SPED than the single-objective modeling in all six target catchment (Table 8).

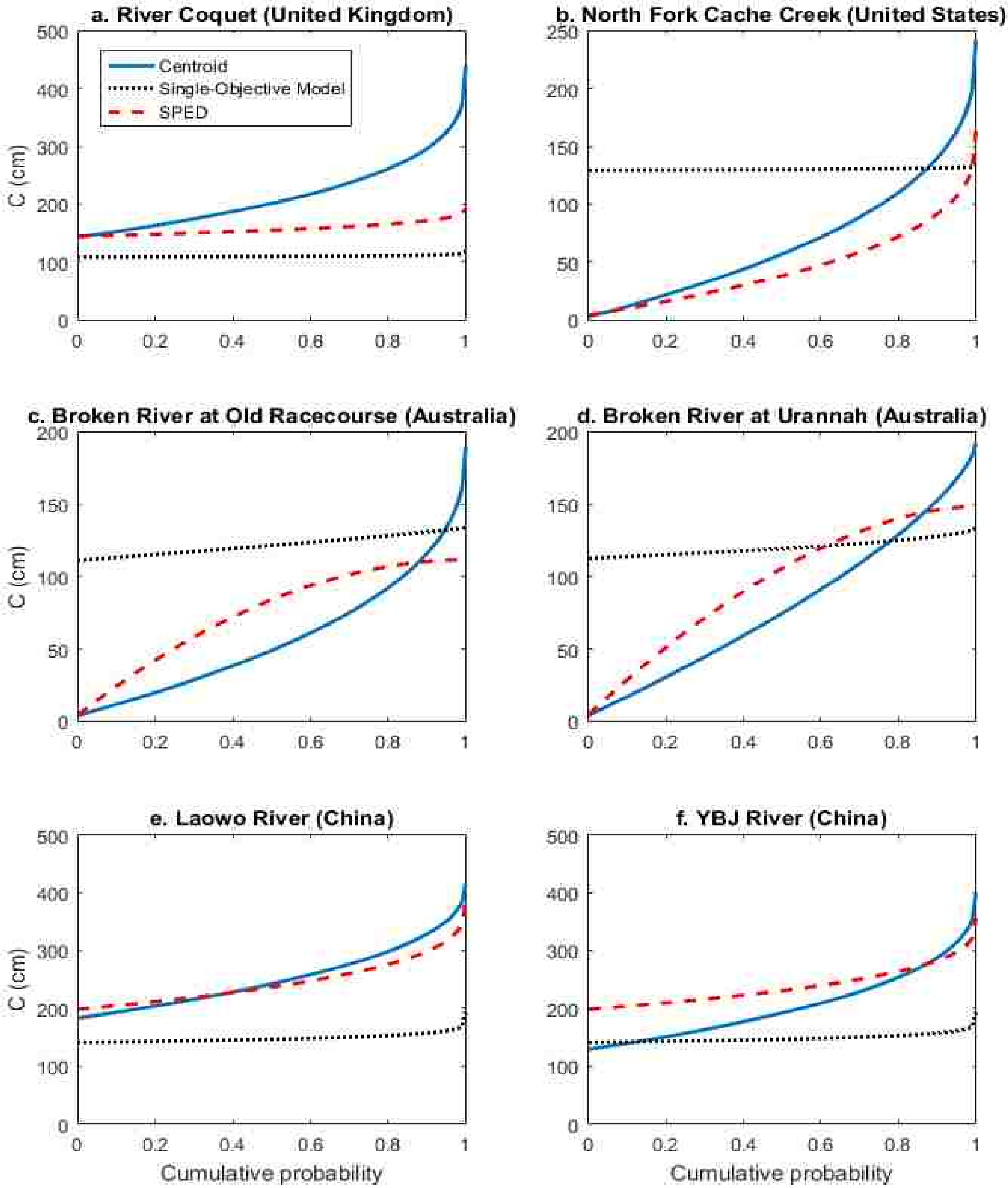

Figure 18. Soil moisture capacity CDFs, as modeled by the single-objective model and SPED, compared to those derived through analysis of soft data in target catchments

3.5.3 SPED Performance as Compared to Predictions Using Robust Data Networks

The SPED framework performs reasonably close to, and sometimes even exceeds performance of prediction approaches based on more robust data networks in all study locations. In River Coquet, SPED performs similarly to the best regionalization case tested by McIntyre et al. (2004) with respect to runoff efficiency, with the same NSE of 0.63 (Table 11). For the case of North Fork Cache Creek in California, the proposed approach by Parada and Liang (2010) performs marginally better than SPED with respect to overall runoff efficiency (NSE of 0.66 versus 0.54, Table 11). However, SPED performs significantly better than the regionalization approach of Post (2009) in predicting streamflow for the 10-year period of September 1975 - August 1985 for both Broken River at Old Racecourse (NSE of 0.74 versus 0.43, Table 11) and Broken River at Urannah (NSE of 0.71 versus 0.22, Table 11).

Table 11. Runoff efficiency of streamflow predictions generated by SPED using a synthetically sparse data network and by previous studies using the full data network

Target catchment	SPED	Robust data network
River Coquet at Morwick (United Kingdom)	0.63	0.63
North Fork Cache Creek (United States)	0.54	0.66*
Broken River at Old Racecourse (Australia)	0.74	0.43*
Broken River at Urannah (Australia)	0.71	0.22*

* Calculated from the data provided in the study.

3.6 Discussion

Streamflow predictions generated using the SPED framework and (artificially or truly) sparse observed data networks are satisfactory in terms of runoff efficiency across a diverse geography of six target catchments (NSE range of 0.54-0.74, Table 8). SPED exhibits a high degree of flexibility and performs consistently and acceptably within diverse hydro-climatological regions. Furthermore, quality of streamflows predicted by SPED are consistent, whether applied in a truly poorly-gauged region of southwestern China (Laowo and YBJ catchments) or within well-gauged regions with a synthetically sparse data set (Table 8, Fig. 17). Regional-scale gridded precipitation and temperature data used in some of the synthetically data-poor regions were of higher resolution and quality (e.g. precipitation data in Australia has 0.05° resolution) than may be available in many regions worldwide. However, data of the quality and resolution applied to model the truly poorly-gauged catchments in China may now be found at a global scale. That streamflows are predicted with reasonable accuracy (NSE of 0.71 - 0.72) in the truly poorly-gauged basins, using only low-quality available data, indicates that SPED may be applicable even in the world's most data-poor regions.

3.6.1 SPED May Alleviate Problems of Equifinality as Compared to Single-objective Calibration

With respect to runoff efficiency, SPED performed similarly to models calibrated to the single objective of maximizing NSE (Table 8). The slightly greater efficiency achieved by single-objective models reflects that the model calibration process was unconstrained, with a single objective of attaining the greater runoff efficiency. By comparison, the multi-criteria SPED calibration process balanced runoff efficiency with selection of a representative parameter set. The trade-offs in runoff efficiency were, for the most part negligible. But, in comparison with the single-objective approach, values of influential model parameters selected by SPED were almost always more similar to parameter values estimated *a priori* by analysis of catchment data (Table 8). We note that there are few instances where residuals of individual parameters are lower in the single-objective model (e.g. YBJ, Broken River at Urannah and Old Racecourse). However, even in these cases, soil moisture capacity CDFs modeled by SPED more closely resemble empirical CDFs compared to those modeled by single-objective models (Fig. 18). This behavior is not surprising, given that the multi-criteria objective function engine of SPED is designed to do just this; select parameter sets which are similar to *a priori* estimates. This design allows SPED to rank equifinal models – models with similar NSE but different parameter sets, and to assign preference to models that represent a best approximation of catchment conditions from available data. For instance, the relatively low parameter residuals achieved for SPED parameter sets (Table 8) indicate that SPED has substantially improved representation of subsurface processes in study catchments (Fig. 18). However, the corresponding NSE values are similar to those attained by models selected solely based on ability to replicate observed streamflows. This indicates that little compromise to runoff efficiency is needed to achieve the observed improvement to process representation.

That runoff efficiencies of SPED and single-objective models are similar, even as parameters estimated by the two models often indicate different hydrologic behaviors, reflects the classic and persistent problem of equifinality in calibration of hydrological models. Uncertainty related to equifinality is particularly troublesome for prediction in ungauged basins, given lack of validation capacity in truly ungauged catchments. In such catchments, the transfer of parameters calibrated in donor catchments is often done “in the blind”, with limited capacity for validation. Increased assurance that selected parameters represent true hydrologic behavior is a benefit to a practitioner who must make decisions based on their confidence in the model. For instance, this problem is exemplified in North Fork Cache Creek in California. It is interesting to note that runoff efficiency achieved during calibration of the single-objective model in the donor catchment (in this case, Eel River below Scott Dam, NSE of 0.62, Table 9) exceeds that achieved through SPED (NSE of 0.48, Table 9). However, when the single-objective model is transferred for prediction in North Fork Cache Creek, the selected model utterly fails to estimate observed streamflows (NSE of 0.08). On the other hand, SPED predictions in North Fork Cache Creek are more robust, estimating streamflows with acceptable accuracy (NSE of 0.54). This example indicates that, by representing soil moisture distributions based on empirical CDFs, SPED is more robust and consistent in performance than the single-objective model, even when efficiency of predictions is the only factor that matters to a decision maker.

3.6.2 SPED Performance Is Comparable to Prediction Using Robust Data Networks

In the United Kingdom, McIntyre et al. (2004) tested whether Bayesian model averaging is preferable to regression for regionalization of hydrological models. The performance of SPED in River Coquet is similar to the best regionalization case presented by McIntyre et al. (2004), both with NSE of 0.63 (Table 11). McIntyre et al. (2004) achieved this performance level by selecting the optimum parameter sets given by the two most similar catchments (in terms of catchment descriptors such as catchment area, standardized annual average rainfall, and base flow index), which were chosen from a network of 30 gauged catchments across UK. Additionally, high-resolution observed daily rainfall data were applied to force the models. SPED, on the other hand, was calibrated for prediction in River Coquet using only one gauged reference catchment and lower-quality regional precipitation data for predictions.

In the United States, Parada and Liang (2010) applied the physically distributed VIC-3L model to predict streamflows in North Fork Cache Creek in California. While runoff efficiency of SPED predictions in North Fork Cache Creek are acceptable (NSE of 0.54), Parada and Liang's approach predicts streamflow with slightly greater efficiency (NSE of 0.66). Parada and Liang (2010) ran the default VIC-3L model using daily gridded meteorological data (precipitation, wind speed, and minimum and maximum daily temperature), as well as soil and vegetation parameters at a resolution of 0.125°. The authors also used observed streamflow data from two gauged reference catchments. The SPED predictions reported, on the other hand, were generated using a lumped conceptual hydrologic model (HyMOD), which is much less data-intensive than VIC-3L model, precipitation, temperature and subsurface information at much lower resolution, and streamflow data from only one reference catchment.

In Australia, Post (2009) regionalized the IHACRES rainfall-runoff model for prediction in several locations within the Burdekin catchment. In Broken River, regionalization proved a more efficient prediction technique as compared to local calibration/simulation at the Old Racecourse gauge, while the opposite was true for the Urannah gauge. We report substantially higher runoff efficiencies based on application of SPED (Table 11, NSE of 0.74 at Old Racecourse and 0.71 at Urannah) in comparison with Post's approach (Table 11, NSE of 0.43 at Old Racecourse and 0.22 at Urannah). Post (2009) reports results of a simplified regionalization technique, requiring only daily rainfall, mean wet season rainfall, stream length, and percent cropping/percent forest.

However, such data are required for a large number of reference catchments to develop regionalization relationships. Quality of available precipitation and land cover data may thus heavily influence the performance of the approach in data-poor regions. Using regional rainfall and temperature data, streamflow data from one reference catchment, and globally available subsurface data with high uncertainties, the SPED framework was able to generate more accurate predictions in both catchments.

In summary, SPED performs comparably or exceeds runoff efficiency performance in three of the four catchments for which previous studies predict streamflows using the full, more robust data networks. In the remaining catchment (North Fork Cache Creek in California), runoff efficiencies are reasonably close (Table 11), suggesting little degradation in model performance despite the lower quantity and quality of data used. While the prior studies utilized robust data networks, which enabled more sophisticated modeling, including regionalization, SPED performed comparably or better while using data from fewer reference catchments, and data associated with greater potential uncertainties (e.g., low-resolution coverages for subsurface characterization, precipitation and temperature). However, some models used in prior studies, such as the regionalization technique by Post (2009), are considerably simpler than the modeling approach used in this paper and can be applied to a large number of catchments efficiently. Thus, indeed, part of the noted improvement in some of the study catchments can be attributed to the modeling approach applied herein (choosing one nearby gauged reference catchment), as well as the hydrologic model used (HyMOD) rather than the SPED framework itself. For instance, runoff efficiencies achieved for the single-objective model in some catchments are similar to or exceed those reported in previous modeling studies (e.g., in the Australian catchments and River Coquet, Tables 8 and 11). While maintaining this advantage, the SPED framework adds further performance consistency and realistic parameter estimation to the modeling process. Consequently, SPED performs reasonably close to the Parada and Liang (2010) modeling approach in North Fork Cache Creek, where the single-objective model fails to predict streamflow acceptably. Such comparable or improved performance of streamflow prediction using SPED is particularly meaningful given the volume of data-poor places in the world where water managers struggle to apply data-rich technologies.

3.6.3 Scientific Contribution of SPED to Hydrologic Process Understanding in Poorly-Gauged Regions

The ability to systematically differentiate equifinal parameter sets to select models that accurately represent estimated catchment conditions is perhaps the most important contribution of SPED to hydrologic process understanding within poorly-gauged regions. While understanding catchment condition through use of soft data is associated with uncertainty, broadly available soft datasets can be incorporated into modeling by water managers around the world to ground traditional hydrologic model calibration within realistic parameter sets. The SPED approach is a mechanism for using such data while acknowledging the inherent uncertainties. Alipour and Kibler (2018) presented preliminary analyses suggesting that the incorporation of preliminary process understanding through inclusion of soft global-scale data led to more accurate model parameterization, without sacrificing runoff efficiency in two catchments of southwestern China. In this study, it is demonstrated that the SPED framework is sufficiently flexible to perform consistently in different climatological conditions and physiographic settings (NSE range of 0.54-0.74, Fig. 17). Comparable performance with previous flow prediction studies that require robust data also indicates the merits of the SPED framework. Since the number of study catchments is limited, we however need to be careful about making general conclusions about the associations between modeling results and climatic and physiographic characteristics of the study regions.

A closer look at the modeled hydrographs and FDCs (Fig. 17) sheds some light into why differentiating equifinal models can be so important in modeling. In five of the six target catchments, the single-objective model more accurately models extreme high flows, while moderate to low flows are consistently modeled either comparably or better by SPED (Fig. 17). This performance difference can be explained by examining differences in soil moisture capacity CDFs generated by the two models (Fig. 18). The difference between C_{max} and C_{min} in CDFs modeled by the single-objective model is small (relative to SPED CDFs) in all six target catchments (Fig. 18). Under this condition, when a catchment reaches C_{min}, from there it can quickly attain C_{max}. Reaching C_{max} means that the entire catchment is saturated and additional rainfall becomes surface runoff, quickly contributing to streamflow. The single-objective model thus often predicts somewhat binary behavior, whereby the catchment is either "on" or "off", producing flashy streamflows. By contrast, the greater difference between C_{min} and C_{max} modeled by SPED indicates a more balanced catchment behavior, where different parts of the catchments become gradually saturated, approximating expansion of variable source areas. Thus, SPED CDFs prove to be more efficient in modeling moderate flow magnitudes (Fig. 17).

The influence of model parameterization to process representation can further be illustrated by North Fork Cache Creek (Fig. 17b and 18b) where SPED remains skillful (NSE of 0.54) while traditional model calibration fails (NSE of 0.08). The single-objective model here estimates a value for C_{min} close to C_{max}, while the empirical (and thus SPED) CDFs indicate a C_{min} value close to zero (Fig. 18b). A high value for C_{min} indicates that small events often cannot attain sufficient soil moisture in any part of the catchment to generate streamflow. Streamflows will be substantial only after long/intense rainfall events. This clearly does not resemble the hydrograph behavior in the catchment (Fig. 17b). Consequently, the hydrograph modeled by the single-objective approach fails to represent low and medium flows (except near zero flows) and only captures some of the highest flows in the catchment (Fig. 17b). Please note that a sufficiently long spin-up period was assigned for modeling in all catchments and about two-thirds of the single-objective predicted streamflow data between October 1950 and December 1951 presented in Fig. 5b are non-zero values. Thus, poor performance of the single-objective model cannot be associated with lack of a spin-up period for filling up the soil reservoir. The SPED parameterization, on the other hand, is able to correctly represent the impact of smaller rainfall events and the fluctuations they cause in the streamflow (Fig. 17b) by estimating more representative soil moisture capacity in the catchment (Fig. 18b). That extreme high flows are better modeled by the single-objective model indicates that processes controlling generation of peak flows are perhaps not sufficiently represented by the 6-parameter HyMOD model. While this model deficiency is obscured in the single-objective model by strict calibration to observed flows, this does not necessarily indicate that the single-objective model is able to better model these processes.

Further analysis of results in North Fork Cache Creek clarifies why the two models deviate substantially in estimating soil moisture capacity CDF and exemplifies another aspect of contribution to hydrologic process understanding that is made possible through application of the SPED framework. A key challenge of flow prediction in ungauged basins is reliance on data from hydrologically similar gauged catchments, which is scarce in regions with poor observed flow networks. Poor prediction skill results when reference and target catchments are dissimilar. Results from North Fork Cache Creek indicate that the SPED framework is robust to some level of catchment dissimilarity. Analysis of streamflow hydrographs in North Fork Cache Creek (target catchment) and Eel River (reference catchment) indicate divergent hydrologic behaviors (Fig. 19), though the two catchments are adjacent (Fig. 14b). Eel River has a stable base flow regime (base flow index of 0.16) and flow almost never drops below 0.1 mm/day. On the other hand, baseflow in North Fork Cache Creek is very low (base flow index of 0.003) and total river flow is frequently below 0.1 mm/day. Because traditional single-objective calibration is tuned to estimate parameters solely based on best fit between calibrated and observed streamflows in the reference catchment (Eel River), the resultant single-objective model aims at fitting a curve that has a stable base flow regime and simultaneously accounts for seasonal high flow events. For instance, a soil moisture capacity CDF with high threshold for initiation of saturation excess (high C_{min}, Fig. 18b) is selected. This allows for greater soil storage, which translates into stable base flows. High flows occur in the case of large events that fill the high storage capacity of soil. While this CDF well serves the task of maximizing calibration runoff efficiency (NSE of 0.62), global soil data indicate that it may not correspond with reality very well. There could be several other parameterizations that produce comparable runoff efficiencies for the calibration (equifinality), but managers have no way of differentiating them. Due to the only partial similarity between Eel River and North Fork Cache Creek, just a few of these parameterizations should work well in North Fork Cache Creek as well. By incorporating soft data in calibration, the SPED procedure selects parameters that better resonate with global soil data, even though this comes with a lower calibration efficiency (NSE of 0.48). When the two models are transferred to North Fork Cache Creek for prediction, the single-objective model fails to predict streamflow (NSE of 0.08), while SPED still takes advantage of the partial similarity between the two catchments and keeps a satisfactory prediction skill (NSE of 0.54). This example illustrates that through incorporation of soft data, rather than relying only on rote calibration, SPED also contributes to hydrologic process understanding within poorly-gauged regions by eliminating parameter sets that could only be identified as poorly-performing if the reference and target catchments were significantly dissimilar with respect to some catchment processes.

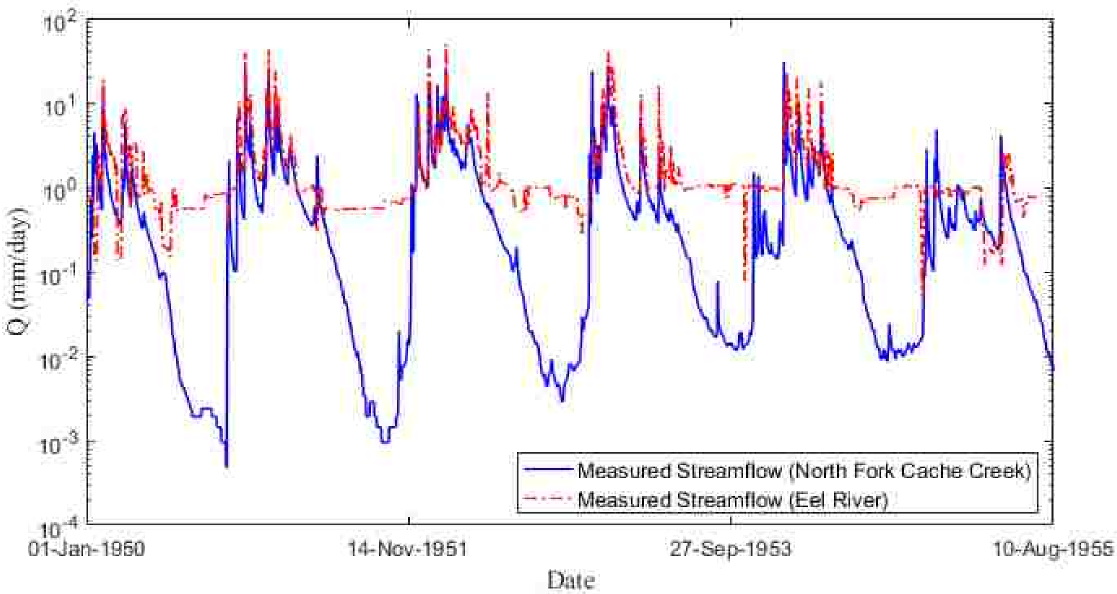

Figure 19. Observed hydrographs in North Fork Cache Creek and Eel River (United States)

3.6.4 SPED Limitations and Directions for Future Research

Questions may arise about the high performance level of the single-objective approach in the target catchments, in particular in the Australian catchments. As in North Fork Cache Creek, the soil moisture CDFs selected by the single-objective model do not match well the catchment conditions

(Fig. 18c and d). The single-objective model reproduces streamflow accurately despite the poor representation of soil moisture capacity because the target and reference Australian catchments are hydrologically similar (Fig. 17c and d). A model that estimates streamflow with high efficiency in the reference catchment is likely to also generate high efficiency predictions in a hydrologically similar target catchment, even if the model does not represent true catchment processes. This should be particularly true when calibration and prediction periods correspond. Similarly, though the SPED procedure is able to partially represent true catchment processes to distinguish equifinal models, the high weight applied to matching NSE in calibration still leads to mismatch in calibrated and estimated parameter values from soft data. Until the uncertainties of global soil data are highly reduced, there may be no other way than giving the highest weight to NSE and relying heavily on hydrograph matching for calibration.

The SPED framework has been tested using only one hydrologic model (HyMOD). Future research and performance testing of SPED using other hydrological models could further clarify limitations/merits of the framework. Application of SPED requires specific attention to each target catchment, for example to assemble *a priori* catchment information. In comparison with regionalization techniques, such as those described by Post (2009) and McIntyre et al. (2004), for application within large geographical areas consisting of several ungauged catchments, catchment-by-catchment application of SPED seems tedious. Another caveat to SPED performance, is exemplified by the described anomalies in the soil moisture capacity empirical CDF and difficulty in matching modeled and empirical data. Despite substantial improvements in estimation of model parameters, there still are cases where parts of the modeled CDF significantly deviate from the empirical CDF (e.g., River Coquet, Fig. 18a), reflecting the challenge of using highly uncertain data. A bimodal calibration and prediction approach and or a better conceptualization of soil moisture capacity and hydrologic behavior (i.e., a better hydrologic model) can lead to further improvements in this regard. However, models selected using a procedure such as SPED will reflect the quality of available data. The SPED procedure is capable of overcoming equifinality challenges only to the extent that data uncertainties allow representation of true catchment condition. As ability to accurately describe catchment and atmospheric conditions improves, for instance through advances in remote sensing, applications for modeling with frameworks such as SPED also improve accordingly.

3.7 Conclusions

Herein the SPED framework was tested in diverse hydro-climatic regions. Accuracy of multi-criteria SPED predictions were tested against single-objective models and also compared to results of previous modeling studies in four synthetically poorly-gauged catchments. SPED performance was also compared to single-objective models in two catchments located in a truly poorly-gauged region of southwestern China, demonstrating its potential for wide applicability in data poor regions. Preliminary process understanding by more representative modeling of catchment soil moisture capacity and associated processes helps SPED better decipher equifinal models. This enhances SPED performance where traditional flow prediction models are likely to fail: handling partial dissimilarity between reference and target catchments. In North Fork Cache Creek in California where partial dissimilarity with Eel River (reference catchment) in precipitation pattern and physiographic setting leads to different base flow regimes, SPED outperforms the single-objective model (NSE of 0.54 versus 0.08). In other study regions where reference and target catchments are similar in climatological conditions and physiographic settings, SPED and single-objective models perform comparably in predicting streamflow (NSE range of 0.63-0.74 versus 0.63-0.79). SPED performance is robust and consistent across the diverse climatological conditions and physiographic settings of test (NSE range of 0.54-0.74 in all six target catchments). Additionally, model parameters selected by SPED may offer superior representation of soil moisture capacity in all study catchments, as compared to empirical distributions. This is reflected in the multi-criteria OF value range of the models selected by SPED (0.22-0.39, Table 8) compared to the single-objective models (0.52-1.32, Table 8). Finally, SPED prediction skill within synthetically poorly-gauged regions with minimum hydrometeorologic observation is comparable to or exceeds that achieved by previous state-of-the-science methods applied within the same (well-gauged) regions when the entire data with highest available quality are used (NSE range of 0.54-0.74 versus 0.22-0.66). Thus, SPED represents an important contribution to the science of flow prediction in regions of sparse hydrologic observation, by addressing flow prediction pitfalls such as equifinality, catchment dissimilarity, and difficulty utilizing uncertain data.

3.8 Acknowledgments

We acknowledge data from the UK National River Flow Archive. We acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (http://ensembles-eu.metoffice.com) and the data providers in the ECA&D project (http://www.ecad.eu). We acknowledge the use of Copernicus data and information funded by the European Union - EU-DEM layers for creation of Fig. 14a. We acknowledge the use of HydroSHEDS data available online at http://hydrosheds.cr.usgs.gov. We acknowledge CPC US Unified Precipitation data and GHCN Gridded V2 data provided by the

NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at https://www.esrl.noaa.gov/psd/. We acknowledge the use of data from ArcGIS website (HydroSHEDS (RIV)) (2017), Geoscience Australia (2010), Lehner et al. (2006; 2008), Morris and Flavin (1990; 1994); US Census Bureau website (2018) and USGS website (Watershed Boundary Dataset) (2017) in creating Fig. 14. We are thankful to the China Hydrology Data Project (Henck et al., 2010) for digitizing observed data and making them available for research. We would like to thank Dr. Dingbao Wang for his advice related to improvement of HyMOD. Finally, we would like to thank Dr. David Post, the other two anonymous reviewers as well as the Associate Editor, Dr. Alberto Viglione, for providing insightful and invaluable comments which substantially improved the quality of this paper.

3.9 References

Alipour, M.H., and Kibler, K.M., 2018. A framework for streamflow prediction in the world's most severely data-limited regions: Test of applicability and performance in a poorly-gauged region of China. *Journal of Hydrology*, 557.

ArcGIS website, 2017. HydroSHEDS (RIV) - Europe river network (stream lines) at 30s resolution (Data basin dataset), 2010. Available from: https://databasin.org/datasets/17c21f42e65545589e000c339734f857 [Accessed 2 November 2017].

Arsenault, R., and Brissette, F., 2016. Multi-model averaging for continuous streamflow prediction in ungauged basins. *Hydrological Sciences Journal*, 61 (13), pp.2443-2454.

Australian Bureau of Meteorology website, 2017. Precipitation data owner: Queensland - Department of Natural Resources and Mines (downloaded from TERN AusCover website (2017)). Available from: http://data.auscover.org.au/xwiki/bin/view/Product+pages/Product+User+Page+Melbourne+2 [Accessed 26 November 2017].

Batjes, N.H., 2015. World soil property estimates for broad-scale modelling (WISE30sec, Ver. 1.0). Report 2015/01, ISRIC—World Soil Information, Wageningen. Available from: http://www.isric.org/data/data-download [Accessed 10 November 2016].

Beven, K., 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources*, 16 (1), 41–51.

Beven, K., 2006. A manifesto for the equifinality thesis. *Journal of Hydrology*, 320 (1), 18–36.

Blöschl, G., *et al.*, (Eds.), 2013. Runoff prediction in ungauged basins: synthesis across processes, places and scales. Cambridge University Press, Cambridge.

Chinese Ministry of Hydrology, 1970. Hydrologic Yearbook: Yunnan-Tibet International Rivers District, District 9 Region 2 1965, [in Chinese]. Ministry of Hydrology, Beijing, China.

Chinese Ministry of Hydrology, 1971. Hydrologic Yearbook: Yunnan-Tibet International Rivers District, District 9 Region 2 1966, [in Chinese]. Ministry of Hydrology, Beijing, China.

Chinese Ministry of Hydrology, 1974. Hydrologic Yearbook: Yunnan-Tibet International Rivers District, District 9 Region 2 1972, [in Chinese]. Ministry of Hydrology, Beijing, China.

Chinese Ministry of Hydrology, 1977. Hydrologic Yearbook: Yunnan-Tibet International Rivers District, District 9 Region 2 1978, [in Chinese]. Ministry of Hydrology, Beijing, China.

Chinese Ministry of Hydrology, 1982. Hydrologic Yearbook: Yunnan-Tibet International Rivers District, District 9 Region 2 1979, [in Chinese]. Ministry of Hydrology, Beijing, China.

Fan, Y., and van den Dool, H., 2008. A global monthly land surface air temperature analysis for 1948-present, *Journal of Geophysical Research: Atmospheres*, 113(D1).

Forsythe, W.C., *et al*., 1995. A model comparison for daylength as a function of latitude and day of year. *Ecological Modelling*, 80 (1), 87-95.

Geoscience Australia, 2010. Geoscience Australia, 3 second SRTM Digital Elevation Model (DEM) v01. Bioregional Assessment Source Dataset. Available from: http://data.bioregionalassessments.gov.au/dataset/12e0731d-96dd-49cc-aa21-ebfd65a3f67a [Accessed 27 November 2017].

Gibbs, M.S., Maier, H.R., and Dandy, G.C., 2012. A generic framework for regression regionalization in ungauged catchments. *Environmental Modelling & Software*, 27, pp.1-14.

Hannah, D.M., *et al*., 2011. Large-scale river flow archives: importance, current status and future needs. *Hydrological Processes*, 25(7), pp.1191-1200.

Haylock, M.R., *et al*., 2008. A European daily high-resolution gridded dataset of surface temperature and precipitation. *Journal of Geophysical Research D: Atmospheres*, 113, D20119, doi:10.1029/2008JD10201.

Hrachowitz, M., *et al*., 2013. A decade of Predictions in Ungauged Basins (PUB)—a review. *Hydrological Sciences Journal*, 58 (6), 1198-1255.

Henck, A.H., *et al*., 2010. Monsoon control of effective discharge, Yunnan and Tibet. *Geology*, 38, 975-978.

Hengl, T., *et al*., 2017. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS One*, 12 (2), p.e0169748.

Heřmanovský, M., *et al*., 2017. Regionalization of runoff models derived by genetic programming. *Journal of Hydrology*, 547, pp.544-556.

Hughes, D.A., 2016. Hydrological modelling, process understanding and uncertainty in a southern African context: lessons from the northern hemisphere. *Hydrological Processes,* 30(14), pp.2419-2431.

Jayawardena, A.W., and Zhou, M.C., 2000. A modified spatial soil moisture storage capacity distribution curve for the Xinanjiang model. *Journal of Hydrology*, 227(1-4), pp.93-113.

Kamali, B., and Mousavi, S.J., 2014. Automatic calibration of HEC-HMS model using multi-objective fuzzy optimal models. *Civil Engineering Infrastructures Journal*, 47(1), pp.1-12.

Kibler, K.M., and Alipour, M.H., 2017. Flow alteration signatures of diversion hydropower: An analysis of 32 rivers in southwestern China. *Ecohydrology*, 10 (5).

Koutsouris, A.J., Seibert, J., and Lyon, S.W., 2017. Utilization of global precipitation datasets in data limited regions: A case study of Kilombero Valley, Tanzania. *Atmosphere*, 8(12).

Lanfear, K.J., and Hirsch, R.M., 1999. USGS study reveals a decline in long-record streamgages. *Eos, Transactions American Geophysical Union*, 80(50), pp.605-607.

Lavers, D.A., 2011. Seasonal hydrological prediction in Great Britain–an assessment. Doctoral dissertation, University of Birmingham.

Lehner, B., Verdin, K., and Jarvis, A., 2006. HydroSHEDS technical documentation. World Wildlife Fund US, Washington, DC. Available from http://hydrosheds.cr.usgs.gov [Accessed 9 February 2018].

Lehner, B., Verdin, K., and Jarvis, A., 2008. New global hydrography derived from spaceborne elevation data. *Eos, Transactions, AGU*, 89 (10): 93-94.

Mathworks, 2016. MATLAB software.

McIntyre, N., *et al*., 2004. Tools and approaches for evaluating uncertainty in streamflow predictions in ungauged UK catchments. International Congress on Environmental Modelling and Software, Osnabrück, Germany.

McIntyre, N., *et al*., 2005. Ensemble predictions of runoff in ungauged catchments. *Water Resources Research*, 41 (12).

Mekong River Commission website, 2018. Available from: http://www.mrcmekong.org/ [Accessed 14 October 2016].

Merz, R., and Blöschl, G., 2004. Regionalisation of catchment model parameters. *Journal of Hydrology*, 287(1), pp.95-123.

Merz, R., Parajka, J., and Blöschl, G., 2009. Scale effects in conceptual hydrological modeling. *Water Resources Research*, 45(9).

Merz, R., Parajka, J., and Blöschl, G., 2011. Time stability of catchment model parameters: Implications for climate impact analyses. *Water Resources Research*, 47(2).

Moore, R.J., 1985. The probability-distributed principle and runoff production at point and basin scales. *Hydrological Sciences Journal*, 30 (2), 273-297.

Moore, R.J., 1999. Real-time flood forecasting systems: Perspectives and prospects. In Floods and landslides: Integrated risk assessment (pp. 147-189). Springer Berlin Heidelberg.

Morris, D.G., and Flavin, R.W., 1990. A digital terrain model for hydrology. Proc 4th International Symposium on Spatial Data Handling. Vol 1 Jul 23-27 Zurich, pp 250-262.

Morris, D.G., and Flavin, R.W. 1994. Sub-set of UK 50 m by 50 m hydrological digital terrain model grids. NERC, Institute of Hydrology, Wallingford.

NOAA/OAR/ESRL PSD website, 2017. Boulder, Colorado, USA. Available from: https://www.esrl.noaa.gov/psd/ [Accessed 27 December 2017].

Oudin, L., *et al*., 2008. Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments. *Water Resources Research*, 44 (3).

Parada, L.M., and Liang, X., 2010. A novel approach to infer streamflow signals for ungauged basins. *Advances in Water Resources*, 33 (4), pp.372-386.

Parajka, J., Merz, R., and Blöschl, G., 2005. A comparison of regionalisation methods for catchment model parameters. *Hydrology and Earth System Sciences Discussions*, 9(3), pp.157-171.

Parajka, J., Blöschl, G., and Merz, R., 2007a. Regional calibration of catchment models: Potential for ungauged catchments. *Water Resources Research*, 43(6).

Parajka, J., Merz, R., and Blöschl, G., 2007b. Uncertainty and multiple objective calibration in regional water balance modelling: case study in 320 Austrian catchments. *Hydrological Processes*, 21(4), pp.435-446.

Parajka, J., *et al*., 2015. The role of station density for predicting daily runoff by top-kriging interpolation in Austria. *Journal of Hydrology and Hydromechanics*, 63 (3), pp.228-234.

Peñas, F.J., *et al*., 2014. The influence of methodological procedures on hydrological classification performance. *Hydrology and Earth System Sciences*, 18, 3393-3409

Perks, A., Winkler, T., and Stewart, B., 1996. The adequacy of hydrological networks: A global assessment, Rep. WMO-740, World Meteorological Organization, Geneva, Switzerland.

Post, D.A., and Jakeman, A.J., 1999. Predicting the daily streamflow of ungauged catchments in SE Australia by regionalising the parameters of a lumped conceptual rainfall-runoff model. *Ecological Modelling*, 123 (2-3), pp.91-104.

Post, D.A., 2009. Regionalizing rainfall–runoff model parameters to predict the daily streamflow of ungauged catchments in the dry tropics. *Hydrology Research*, 40 (5), pp.433-444.

Quan, Z., *et al*., 2015. Evaluation of the HYMOD model for rainfall–runoff simulation using the GLUE method. Proceedings of the International Association of Hydrological Sciences, 368, pp.180-185.

Richter, B.D., *et al.*, 1997. How much water does a river need? *Freshwater Biology*, 37, pp. 231-249.

Seibert, J., and McDonnell, J.J., 2002. On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration. *Water Resources Research*, 38(11).

Seibert, J., and McDonnell, J.J., 2013. Gauging the ungauged basin: relative value of soft and hard data. *Journal of Hydrologic Engineering*, 20 (1).

Shiklomanov, A.I., Lammers, R.B., and Vörösmarty, C.J., 2002. Widespread decline in hydrological monitoring threatens pan-Arctic research. *Eos, Transactions American Geophysical Union*, 83(2), pp.13-17.

Sugeno, M., 1985. An introductory survey of fuzzy control. *Information Sciences*, 36 (1), 59-83.

Sun, W., *et al*., 2015. Estimating daily time series of streamflow using hydrological model calibrated based on satellite observations of river water surface width: Toward real world applications. *Environmental Research*, 139, pp.36-45.

Swain, J.B., and Patra, K.C., 2017. Streamflow estimation in ungauged catchments using regionalization techniques. *Journal of Hydrology*, 554, pp.420-433.

Tegegne, G., Park, D.K., and Kim, Y.O., 2017. Comparison of hydrological models for the assessment of water resources in a data-scarce region, the Upper Blue Nile River Basin. *Journal of Hydrology: Regional Studies*, 14, pp.49-66.

Thornthwaite, C.W., 1948. An approach toward a rational classification of climate. *Geographical Review*, 38 (1), 55-94.

UK National River Flow Archive website, 2017. Available from: https://nrfa.ceh.ac.uk [Accessed 3 November 2017].

US Census Bureau website, 2018. Cartographic boundary shapefiles - states. Available from: https://www.census.gov/geo/maps-data/data/cbf/cbf_state.html [Accessed 9 February 2018].

USGS website, 2017. USGS water data for the nation. Available from: https://www.usgs.gov [Accessed 27 December 2017].

USGS website, 2017. Watershed Boundary Dataset (WBD). Watershed boundary dataset for California. Available from: http://datagateway.nrcs.usda.gov [Accessed 27 December 2017].

van Emmerik, T., *et al*., 2015. Predicting the ungauged basin: Model validation and realism assessment. *Frontiers in Earth Science*, 3 (62).

Vandewiele, G.L., and Elias, A., 1995. Monthly water balance of ungauged catchments obtained by geographical regionalization. *Journal of Hydrology*, 170(1-4), pp.277-291.

Vörösmarty, C., *et al*., 2001. Global water data: A newly endangered species. *Eos, Transactions American Geophysical Union*, 82(5), pp.54-58.

Vrugt, J.A., *et al*., 2003. Effective and efficient algorithm for multiobjective optimization of hydrologic models. *Water Resources Research*, 39(8).

Wagener, T., *et al*., 2013. A synthesis framework for runoff prediction in ungauged basins. chapt, 2, pp.11-28.

Wang, D., Chen, Y., and Cai, X., 2009. State and parameter estimation of hydrologic models using the constrained ensemble Kalman filter. *Water Resources Research*, 45 (11).

Wang, D., 2018. A new probability density function for spatial distribution of soil water storage capacity leads to SCS curve number method. *Hydrology and Earth System Sciences*, 22(12), pp.6567-6578.

Wi, S., *et al*., 2015. Calibration approaches for distributed hydrologic models in poorly gaged basins: implication for streamflow projections under climate change. *Hydrology and Earth System Sciences*, 19(2), pp.857-876.

World Meteorological Organization, 2008. Guide to hydrological practices volume I, hydrology – from measurement to hydrological information. WMO-No. 168, Sixth edition.

Yadav, M., Wagener, T., and Gupta, H., 2007. Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. *Advances in Water Resources*, 30 (8), pp. 1756-1774.

Yang, T.C., *et al.*, 2004. Application of fuzzy multiobjective function on storm-event rainfall-runoff model calibration. *Journal of Hydrologic Engineering*, 9(5), pp.440-445.

Yapo, P.O., 1996. A multiobjective global optimization algorithm with application to calibration of hydrologic models. PhD dissertation, University of Arizona, Tucson.

Yapo, P.O., Gupta, H.V., and Sorooshian, S., 1998. Multi-objective global optimization for hydrologic models. *Journal of Hydrology*, 204 (1-4), pp.83-97.

Yasutomi, N., Hamada, A., and Yatagai, A., 2011. Development of a long-term daily gridded temperature dataset and its application to rain/snow discrimination of daily precipitation. *Global Environmental Research*, 15(2), pp.165-172.

Yatagai, A., *et al*., 2012. APHRODITE: Constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges. *Bulletin of the American Meteorological Society*, 93(9), pp.1401-1415.

Yu, P.S., and Yang, T.C., 2000. Fuzzy multi-objective function for rainfall-runoff model calibration. *Journal of Hydrology*, 238(1), pp.1-14.

Zhang, Z., *et al*., 2008. Reducing uncertainty in predictions in ungauged basins by combining hydrologic indices regionalization and multiobjective optimization. *Water Resources Research*, 44 (12).

Zhou, X., *et al*., 2017. Remapping annual precipitation in mountainous areas based on vegetation patterns: a case study in the Nu River basin. *Hydrology and Earth System Sciences*, 21(2).

CHAPTER 4:
FLOW ALTERATION BY DIVERSION HYDROPOWER IN TRIBUTARIES TO THE SALWEEN RIVER: A COMPARATIVE ANALYSIS OF TWO STREAMFLOW PREDICTION METHODOLOGIES

4.1 Preface

This chapter describes application of the SPED framework on a regional scale to predict streamflow in 32 ungauged catchments in southwestern China developed with diversion hydropower projects. The generated data are used to study flow regime alterations due to diversion hydropower and the results are compared to those based on streamflow data simulated by a more simplistic catchment similarity approach. The content of this chapter has been submitted to River Research and Applications[3] and is currently under review.

4.2 Abstract

A multi-model approach was applied to reconstruct long-term flow records in 32 ungauged rivers developed with small diversion hydropower stations. Hydrologic alteration was assessed for flow records simulated by a catchment similarity model and the multi-criteria Streamflow Prediction under Extreme Data-scarcity (SPED) framework. Both flow prediction techniques indicated that flow signatures were altered substantially by diversion hydropower. Mean annual flows decreased by a mean of 76-86% across the 32 rivers and flow became more predictable in most rivers (47-94% mean increase in predictability). Frequency and duration of high flows decreased and duration of low flow events increased substantially. Slopes of rising hydrograph limbs and recession limbs

[3] Alipour, M.H., Kibler, K.M., 2019 (under review). Flow alteration by diversion hydropower in tributaries to the Salween River: a comparative analysis of two streamflow prediction methodologies. River Research and Applications.

increased respectively by a mean of 123-161% and 254-720%. While direction of detected flow alteration was similar regardless of model choice, severity of alteration was consistently greater based on analysis of flows simulated by the multi-objective SPED model. Model validation based on limited observed data suggests that the SPED flow predictions are substantially more accurate than those generated by the catchment similarity model (NSE of 0.74 and 0.22, respectively). Overall, the agreement of the multi-model analysis indicates that the signal of flow alteration by diversion hydropower in the study rivers supersedes uncertainty associated with flow prediction. While both models may be appropriate for applications such as change detection analysis, prescriptive management actions, such as establishing flow targets for environmental flow regimes, should be based on flow records generated by models adept at simulating rainfall-runoff processes targeted to individual basins, such as SPED.

4.3 Introduction

Composition, structure and functionality of aquatic ecosystems are shaped by flow regime (Richter et al., 1996; Bunn and Arthington, 2002). Thus, human alteration to river flow regimes, for instance through water withdrawal or flow regulation, is often associated with substantial ecosystem-level impact (Merritt et al., 2010). While flow regime alterations due to river regulation have been extensively reported (e.g., Timpe and Kaplan, 2017; Zhang et al., 2018), alteration by diversion hydropower has received comparatively little analysis. It is often assumed that diversion projects (also commonly known as Run of River projects) have little to no impact on the natural flow regime and ecosystem of the rivers (Gibeau et al., 2016). However, comparisons between large storage systems without diversion and diversion systems have indicated potential for greater environmental impact from diversion systems (Gleick, 1992; Kibler and Tullos, 2013; Kibler 2017). Recent studies of flow alteration have clarified that diversion hydropower projects can substantially alter the flow regime of the depleted downstream reaches (Anderson et al., 2014; Fantin-Cruz, 2015; Wang et al., 2016; Kibler and Alipour, 2017). For instance, Kibler and Alipour (2017) analyzed flow regime alterations across 32 small rivers developed with diversion hydropower projects in southwestern China. The study concluded that diversion of water for the small hydropower projects altered all aspects of flow regime, and that low to moderate flow variability was all but eliminated, replaced by a minimum residual flow.

Quantitative analysis of flow regime change by diversion hydropower is often limited by lack of data. Diversion hydropower is often developed in steep, ungauged catchments (Li et al., 2013; Tuna, 2013). In areas of the world with sparse hydrometeorological networks, such as remote, mountainous areas, quantitative flow predictions and analyses can be challenging. For example, to assess hydrologic alteration in the 32 largely ungauged rivers, Kibler and Alipour (2017) utilized a catchment similarity model (Falkenmark & Chapman, 1989) to create historical baseline and regulated flows. This approach is based on the presumption of hydrologic similarity between streamflow generation processes in a gauged reference catchment and ungauged target catchment(s). Accuracy of simulated streamflows may be compromised in data-poor study regions where few reference catchments are available. Indeed, validation of catchment similarity model performance as applied by Kibler and Alipour (2017) indicated that the model was largely unable to reproduce accurate moderate to high flows. To enable quantitative assessment of flow regime alterations within such data-scarce regions, streamflow prediction techniques suited to severely data-scarce regions are required.

Much progress has been made in the science of streamflow prediction (Blöschl et al., 2013; Hrachowitz et al., 2013; Seibert and McDonnell, 2013). However, application of many cutting-edge techniques within severely data-scarce regions remains limited due to barriers such as equifinality (Beven, 1993), hydrologic dissimilarity (Peñas et al., 2014), difficulty utilizing highly uncertain regional/global low-resolution data (Alipour and Kibler, in revision), and difficulty validating predictions in the absence of streamflow observations (van Emmerik et al., 2015). In response to this need, Alipour and Kibler (2018) proposed a framework for Streamflow Prediction under Extreme Data-scarcity (SPED) which proved adept to predict streamflow with high efficiency in a remote area of southwestern China. Testing in diverse hydro-climatic catchments on four continents illustrated that SPED could improve streamflow prediction efficiency in comparison with traditional methods (Alipour and Kibler, in revision).

The objective of this study is to assess hydrologic alteration in a suite of ungauged rivers developed for diversion hydropower production, and to heighten certainty of results by applying a multi-model approach to streamflow prediction. This is achieved by comparing results derived through analysis of streamflow data generated by SPED to those deriving from application of the catchment similarity model (as reported by Kibler and Alipour, 2017). Herein we address two questions: 1) How does diversion hydropower alter flow regimes of the 32 study rivers, and 2) Do conclusions regarding flow regime alteration vary depending on streamflow prediction method? For the first time the SPED framework is utilized and evaluated for large-scale flow prediction over a truly poorly gauged region. Furthermore, the relative importance of flow prediction accuracy to assess direction and severity of hydrologic change is investigated. Finally, potential merits and time/effort tradeoffs of sophisticated versus more simplistic flow prediction approaches are analyzed from the perspective of a water resources manager.

4.4 Methodology

4.4.1 The Region of Study and the Rivers

The catchments of interest are located in southwestern China, within the Salween and Mekong River basins (Table 12, Fig. 20). The international Salween (locally known as Nu River) and Mekong Rivers (locally Lancang River) originate in the eastern highlands of the Tibetan Plateau and flow southward through Yunnan Province before entering Myanmar and Laos, respectively. Within Yunnan Province, tributaries to the mainstem rivers flow through steep valleys draining small, mountainous catchments. Regional climate is monsoonal, and seasonal rainfall pulses result in corresponding high flows in tributary rivers (Institute of Water Resources, 2006). While mainstem rivers contain snow and glacial melt, these are not significant sources of runoff to tributaries in Yunnan (Chinese Academy of Sciences, 1990; Yunnan Bureau of Hydrology and Water Resources, 2005; Mekong River Commission, 2005). Catchments are dominated by forested land cover, with limited urbanization (less than 0.1%) and some agriculture (7-15%) (DeFries and Hansen, 2010).

Diversion hydropower projects have been implemented on many steep tributaries to the Salween River within Yunnan Province (Kibler and Tullos, 2013). Each hydropower project incorporates at least one dam, but several dams may divert multiple tributaries to one power generation station. Small impoundments behind dams retain water on the order of hours (Kibler and Tullos, 2013). Diverted water to power stations may be returned to the same river downstream of the impoundment, or more often, is discharged to a different river (e.g., mainstem of the Salween River). Among the tributaries of the Salween River developed with diversion hydropower, we analyze flow in 32 diverted rivers (Fig. 20) which contribute to 23 hydropower projects. The catchments (Table 12) range from very small in size (13.9 km^2) to larger catchments (457 km^2). Observed streamflow data are only available for one year prior to dam development (1987), and only in the largest catchment of study (Laowo River, Table 12, Fig. 20). The Yongchun River (197 km^2 catchment area), a tributary to the Mekong River, is adopted as a reference catchment for estimating streamflow in the 32 ungauged catchments. Daily flows have been monitored in this gauged river since 1960.

Table 12. Rivers and hydropower dams studied. Mean annual flow metrics and diversion index are based on data simulated using the SPED framework.

Number	River name	*Project installed capacity (MW)	*Dam height (m)	*Basin area at dam (km^2)	*Designed diversion (m^3/s)	Mean annual river flow (m^3/s)			Diversion index
						without diversion	2.05	Percent difference	
1	Pula River	24.8	19.3	70.1	3.21	1.80	1.05	-91	2.05
2	Qiqiluo River	20.0	18.7	257.9	4.95	5.81	2.95	-67	1.05
3	Dimaluo River	56.0	24.6	162.4	7.53	3.28	2.37	-90	2.95
4	Galabo River	14.0	17.8	127.5	4.90	2.67	5.53	-86	2.37
5	Mujiajia River	18.9	6.0	262.1	3.26	0.74	5.43	-78	5.53
6	Mujiajia tributary 2	NA	5.0	13.9	1.24	0.29	5.16	-79	5.43
7	Mujiajia tributary 3	NA	6.0	20.3	1.81	0.44	9.32	-80	5.16
8	Mujiajia River (US)	10.0	10.0	32.3	5.38	0.72	6.04	-94	9.32
9	Mukeji River	31.5	10.5	57.0	6.04	1.25	4.84	-95	6.04
10	Lishiluo River	6.4	14.5	41.8	2.36	0.62	3.39	-84	4.84
11	Yamu River	49.0	8.7	78.3	4.28	1.66	3.22	-92	3.39
12	Yamu tributary	NA	8.7	66.3	3.63	1.44	6.41	-92	3.22
13	Alu River	12.6	5.5	24.6	2.73	0.53	3.70	-81	6.41
14	Zhali River	2.6	4.0	40.9	2.50	0.87	3.70	-93	3.70
15	Ganbu River	3.8	4.0	32.1	1.85	0.68	3.81	-88	3.70
16	Guquan River	22.0	11.0	34.9	2.34	0.79	3.64	-90	3.81
17	Wuke River	NA	10.0	28.7	1.93	0.67	3.81	-90	3.64
18	Zema River	15.0	4.0	56.6	3.61	1.22	3.99	-93	3.81
19	Zema tributary	NA	3.0	14.2	0.94	0.31	4.87	-94	3.99
20	Pushi River	10.0	5.0	43.0	3.70	0.98	3.28	-94	4.87
21	Zilijia River	6.4	7.0	31.5	1.76	0.69	4.31	-93	3.28
22	Zileng River	24.0	7.0	41.7	3.07	0.93	8.43	-89	4.31
23	Zileng tributary 2	NA	8.0	10.9	1.49	0.24	2.27	-83	8.43
24	Zileng tributary 3	NA	5.5	20.3	0.80	0.46	4.29	-87	2.27
25	Labuluo River	26.0	10.3	84.2	5.79	1.73	3.77	-75	4.29
26	Toulu River	NA	10.0	29.4	2.03	0.70	11.10	-79	3.77
27	Nalai River	24.0	9.1	21.6	4.41	0.52	4.21	-94	11.10
28	Duduluo River	48.0	15.5	84.6	7.26	2.23	1.79	-89	4.21
29	Jidu River	16.0	4.6	71.9	2.42	1.79	1.83	-79	1.79
30	Jidu tributary	NA	4.6	69.7	2.34	1.69	1.88	-79	1.83
31	Gutan River	7.5	4.0	99.0	3.92	2.83	1.83	-80	1.88
32	Laowo River	25.0	17.0	457.2	17.18	12.62	NA	-81	1.83
Ref	Yongchun River	NA	NA	197.0	NA	2.59	NA	NA	NA

***Kibler and Tullos (2013)

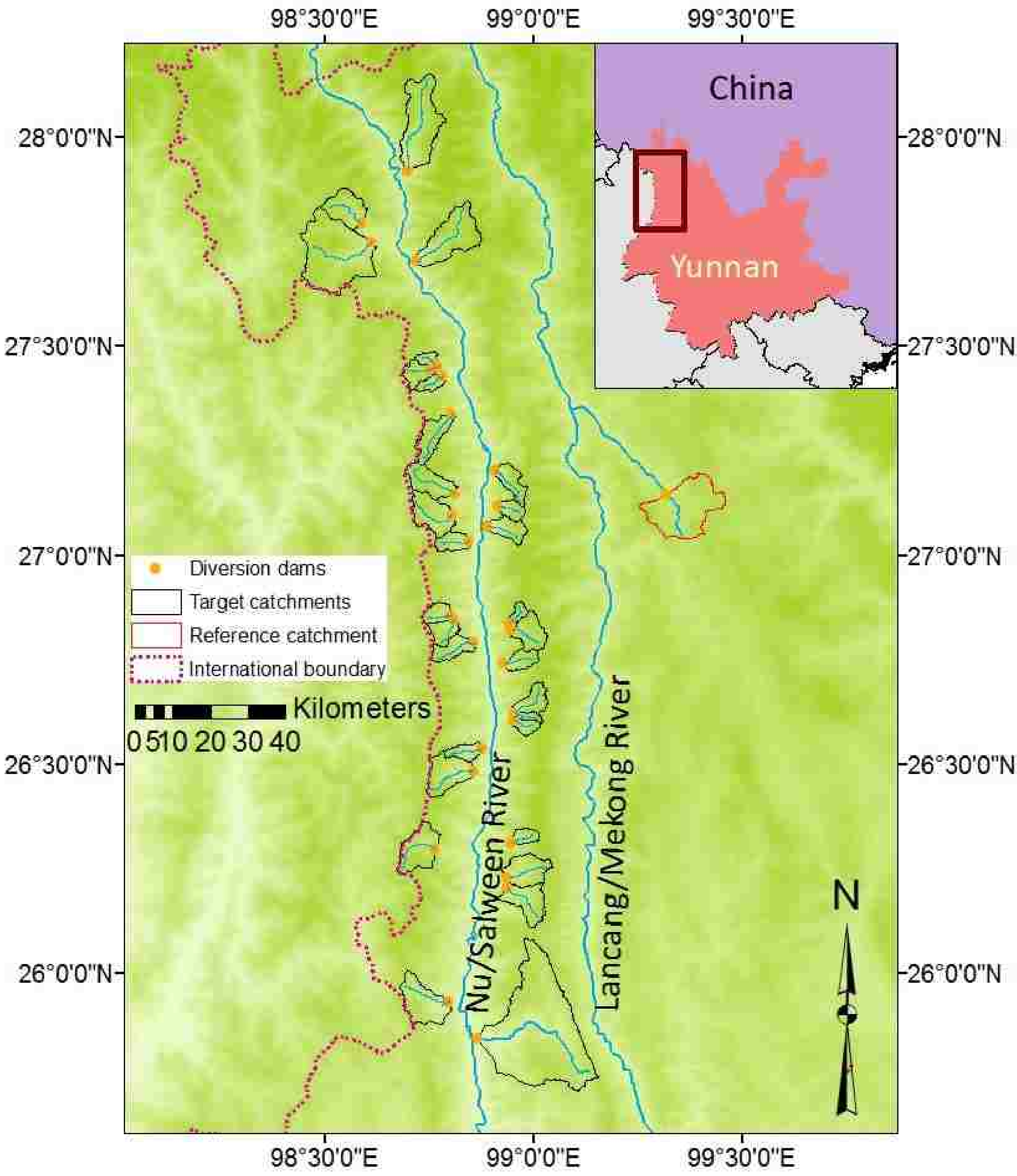

Figure 20. The region and the rivers of study

4.4.2 Streamflow Prediction

We simulated 27-year historical timeseries of daily mean unaltered flows in the 32 rivers using two different streamflow prediction methods. We then modeled effects of hydropower diversion by removing reported absolute withdrawal rates from the two simulated unaltered flow timeseries. We quantified hydrologic changes caused by diversion hydropower by comparing the regulated and unregulated timeseries. The two flow prediction methodologies applied make use of data from a gauged reference catchment to predict flows in nearby ungauged target catchments. The same observed flow data, from the Yongchun River (Fig. 20) were applied to both models. The first flow prediction method, a catchment similarity model, presumes that hydrologic routing processes are similar in target and reference catchments (Kibler and Alipour, 2017). Hence, under the catchment similarity approach, flows observed in the reference site are scaled by comparative 3-day mean precipitation and catchment size to approximate target catchment flows. A detailed description of the catchment similarity streamflow prediction approach can be found in Kibler and Alipour (2017).

The second flow prediction method, the SPED framework (Alipour and Kibler, 2018; Alipour and Kibler, in revision) balances multiple objectives in the calibration process. Through a multi-criteria objective function, soft data (i.e., qualitative knowledge or highly uncertain quantitative knowledge) are combined with hard data to simultaneously maximize runoff efficiency while accurately representing catchment characteristics. In calibration, SPED incorporates *a priori* parameter estimates derived directly from available data. In data-poor regions, these *a priori* values can be estimated from soft data and represented by fuzzy numbers to account for uncertainties.

The 5-parameter HyMOD hydrologic model (Moore 1985) was chosen as the runoff engine within the SPED framework. The model was calibrated with dual objectives: to minimize residuals between *a priori* estimates of influential model parameters and their calibrated value, and to maximize Nash-Sutcliff Efficiency (NSE). Influential model parameters included C_{max}, maximum soil moisture capacity within a catchment, and B, a shape factor describing the catchment's spatial variability in soil moisture capacity. Both parameters influence the catchment's capacity to store water and thus control the partitioning of rainfall between overland and subsurface flowpaths. For instance, high C_{max} indicates that a catchment can store a comparatively large fraction of precipitation, which translates into few zero flow days, stable base flows and high flow constancy. Parameter B controls the curvature of a cumulative distribution function (CDF) of soil moisture capacity across the catchment. Curvature of the CDF describes variable sources areas that become saturated during a rainfall event and how quickly saturation occurs.

A priori values of C_{max} and B were estimated from global databases (soil type WISE30sec, Batjes, 2015 and soil depth to bedrock SoilGrids250m, Hengl et al., 2017), which are the only source of large-scale subsurface data in the region. The two datasets were combined to create empirical CDFs of soil moisture capacity in each catchment. The HyMOD CDF for soil moisture capacity was fitted to each empirical CDF to estimate *a priori* values of B and C_{max} in each catchment. Multi-objective calibration and validation was performed in the reference catchment (Yongchun River). Finally, calibrated models were transferred to target catchments for flow prediction. Interested readers may refer to Alipour and Kibler (2018) and Alipour and Kibler (in revision) for a more detailed description of the SPED modeling approach using HyMOD.

4.4.2.1 Precipitation and PET

Precipitation was estimated from the APHRODITE spatial precipitation product (Yatagai et al., 2012). As APHRODITE is associated with underestimation of rainfall in high altitudes due to poor representation of orographic effect (Wi et al., 2015), data were bias-corrected through the empirical quantile mapping approach (Lafon et al., 2013) using long-term regional precipitation observations. Details of precipitation bias correction approach are available in Alipour and Kibler (2018). Monthly potential evapotranspiration (PET) estimates were generated for the 32 target catchments and the reference catchment using Thornthwaite's approach (1948). PET was estimated from a regional-scale spatial temperature data product, AphroTemp (APHRODITE website, 2015; Yasutomi et al., 2011) and daylight length was estimated as in Forsythe et al. (1995).

4.4.3 Analysis of Hydrologic Alteration

The two unregulated streamflow data series simulated in each catchment were used to estimate the regulated (modified) streamflows for the same climatic/hydrologic conditions after perturbation by diversion hydropower. Following Kibler and Alipour (2017), the modified flows were computed as:

$$Q_{mod,i,j} = Q_{i,j} - Q_{div,j} \quad (11)$$

Where $Q_{mod,i,j}$ denotes the mean daily streamflow downstream of the dam on day *i* in river *j*; $Q_{i,j}$ is the unregulated mean daily flow downstream of the dam on day *i* in river *j*; and $Q_{div,j}$ is the static diverted flow for hydropower generation from river j. Diverted flow consists of hydropower design flow as well as additional water withdrawn to compensate for leakage losses in the system. We make a conservative assumption that minimum residual flows equal to 5% of the unregulated river's mean annual flow are maintained below each diversion dam, though this is not reported at all dams (Kibler and Tullos, 2013). Accordingly, the following decision rules are used to estimate the modified flows from unregulated streamflow data in each river (Kibler and Alipour, 2017):

$$\begin{gathered} If\ Q_{i,j} < Q_{min,j},\ \ then\ \ Q_{mod\_final,i,j} = Q_{i,j} \\ If\ Q_{mod,i,j} < Q_{min,j},\ \ then\ \ Q_{mod\_final,i,j} = Q_{min,j} \\ If\ Q_{mod,i,j} > Q_{min,j},\ \ then\ \ Q_{mod\_final,i,j} = Q_{mod,i,j} \end{gathered} \quad (12)$$

Where $Q_{mod\_final,i,j}$ denotes the final modified flow on day *i* in river *j*; and $Q_{min,j}$ is the minimum residual flow considered for river *j*.

Flow regime metrics from unregulated and modified flow records were compared (Richter et. al, 1997; Olden and Poff, 2003; Mathews & Richter, 2007) to identify flow alteration patterns as a result of diversion hydropower across the 32 rivers of study. The Nature Conservancy's Indicators of Hydrologic Alteration (The Nature Conservancy website, 2017) software package was used to compute flow regime descriptors associated with the five main characteristics of flow: magnitude, timing (predictability), frequency, duration, and rate of change of flow. Flow alterations were assessed two times: once comparing unregulated and regulated flows generated using the SPED framework, and once comparing unregulated/regulated flow data generated using the catchment similarity model. Flow alterations detected using both models were compared to assess the impact of model choice.

4.5 Results

Streamflows predicted by the SPED framework are similar to observed values across almost all flow magnitudes (NSE = 0.74, Percent Bias = -7%, Fig. 21). By comparison, streamflows simulated by the catchment similarity model often underestimate observed flows [NSE = 0.22, Percent Bias = -56%, Fig. 21].

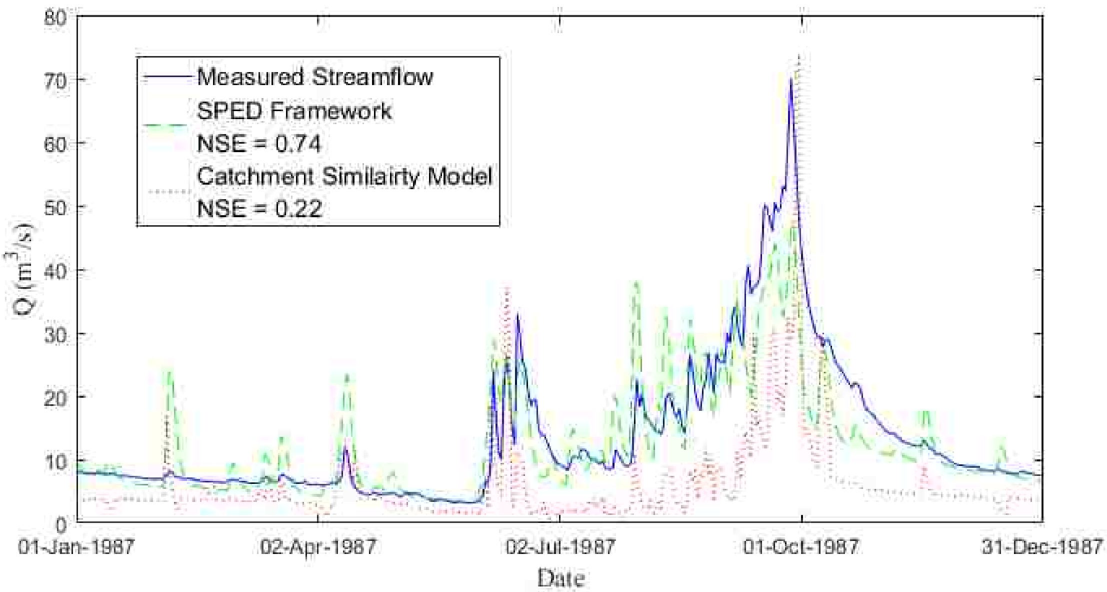

Figure 21. Modeled streamflow by the SPED framework and catchment similarity modeling approach in comparison with observed streamflow in Laowo River

Regardless of which model is used, analysis of hydrologic alteration indicates that hydrograph behavior in all rivers is greatly affected by diversion (Fig. 22). This overall effect transcends the differences between models, and translates into substantial alterations with respect to flow descriptors. Detailed comparative analysis allows for differentiation of model effects as well. Results of hydrologic alteration analysis are reported as percent difference in regulated versus unregulated flows computed across the 32 rivers. Results based on the SPED framework are reported first, followed by those based on the catchment similarity model in brackets. Annual minimum flows decreased with a mean of -68 ± 23% [-12 ± 16%] across the 32 rivers and annual maximum flows decreased with a mean of -69 ± 25% [-26 ± 13%]. Moderate events were also affected by diversion; mean annual flows decreased with a mean of -86 ± 7% [-76 ± 12%] (Fig. 23a), 7-day minimum flows decreased with a mean of -71 ± 21% [-41 ± 23%] (Fig. 23b), and 7-day maximum flows decreased with a mean of -78 ± 20% [-52 ± 18%] (Fig. 23c).

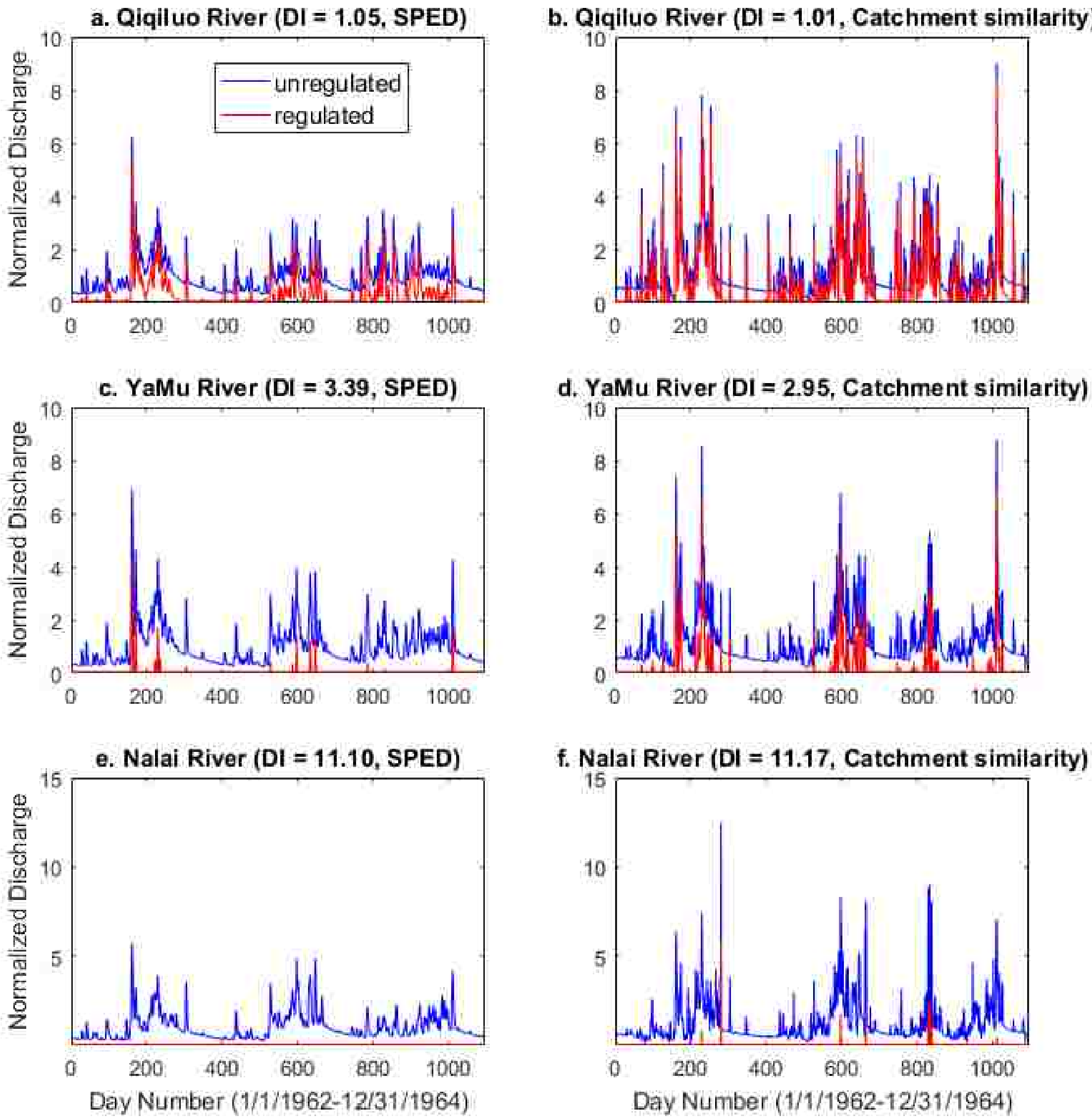

Figure 22. Normalized discharge before and after diversion across rivers with a range of diversion indices for SPED (left) and catchment similarity modeling (right)

In terms of flow frequency, while regulated flows surpassed the pre-diversion 25th exceedance probability flow (Q_{25}) a mean of 13.66 [19] times per year before diversion, this was surpassed only 1.13 [7] times per year after diversion (-91 ± 14% decrease; [-64 ± 16% decrease], Fig. 23d). Duration of high flow events (number of days per event spent above Q_{25}) also decreased from a mean of 4 [3] days to a mean of 2 [2] days (-52 ± 22% decrease; [-40 ± 19% decrease], Fig. 23e). Low flow events (below the pre-diversion 75th exceedance probability flow (Q_{75})) occurred with similar or lower frequency after regulation, with the two models predicting slightly different changes. Low flows occurred a mean of 3.72 [11] times per year after diversion versus 8.95 [11] times per year before diversion (-59 ± 39% decrease; [-2 ± 44% decrease]). However, both models detect that duration of low flow events increased substantially from a mean of 11 [3] days to a mean of 175 [27] days (Fig. 24a).

With respect to rate of change, slopes of rising hydrograph limbs exhibit a mean increase of 161 ± 213% [123 ± 71%] across the rivers after diversion (Fig. 24b) while slopes of recession limbs increase even more substantially (720 ± 316% mean increase; [254 ± 137% mean increase], Fig. 24c). Some aspects of flow timing and predictability were substantially altered by diversion as well. While timing of annual maximum flow was altered in only a few rivers, timing of annual minimum flow was altered across all rivers ($p < 0.001$, -15 ± 19% decrease, [-23 ± 18% decrease]). Moreover, flow became highly predictable in most rivers (47 ± 16% mean increase in predictability of flow [94 ± 22% mean increase]) as a result of substantial increase in flow constancy (92 ± 29% mean increase; [184 ± 49% mean increase], Fig. 24d). On the other hand, flow contingency/periodicity reduced in all rivers (-80 ± 22% mean decrease) [-50 ± 19% mean decrease] primarily due to long periods of flow sustained at minimum residual flow. The increase in flow constancy can very well be described by the diversion index (DI, Eq. 13) proposed by Kibler and Alipour (2017):

$$DI_j = \frac{Q_{div,j}}{Q_{50,j}} \quad (13)$$

Where $Q_{div,j}$ is the design diversion flow in river j, and $Q_{50,j}$ is the pre-diversion median annual flow in river j. Diversion indices (Table 12) range from 1.05 [1.01] (Qiqiluo River) to 11.10 [11.17] (Nalai River) across the rivers with a mean of 4 ± 2 [4 ± 2]. As diversion index increases, the impact of diversion on the river's natural flow regime increases as well. This normally translates into longer periods of flow sustained at minimum residual flow as well as other substantial hydrological changes in the rivers. Rivers with a high diversion index flow at the minimum residual flow for much of the time. While flows in all rivers exceeded the minimum residual flow more than 99 [96] percent of the time before diversion, flows in almost all rivers exceeded minimum residual flow less than 24 [22] percent of the time after diversion (Fig. 24e). The exceptions are Qiqiluo and Dimaluo Rivers, which are both characterized by relatively low diversion indices.

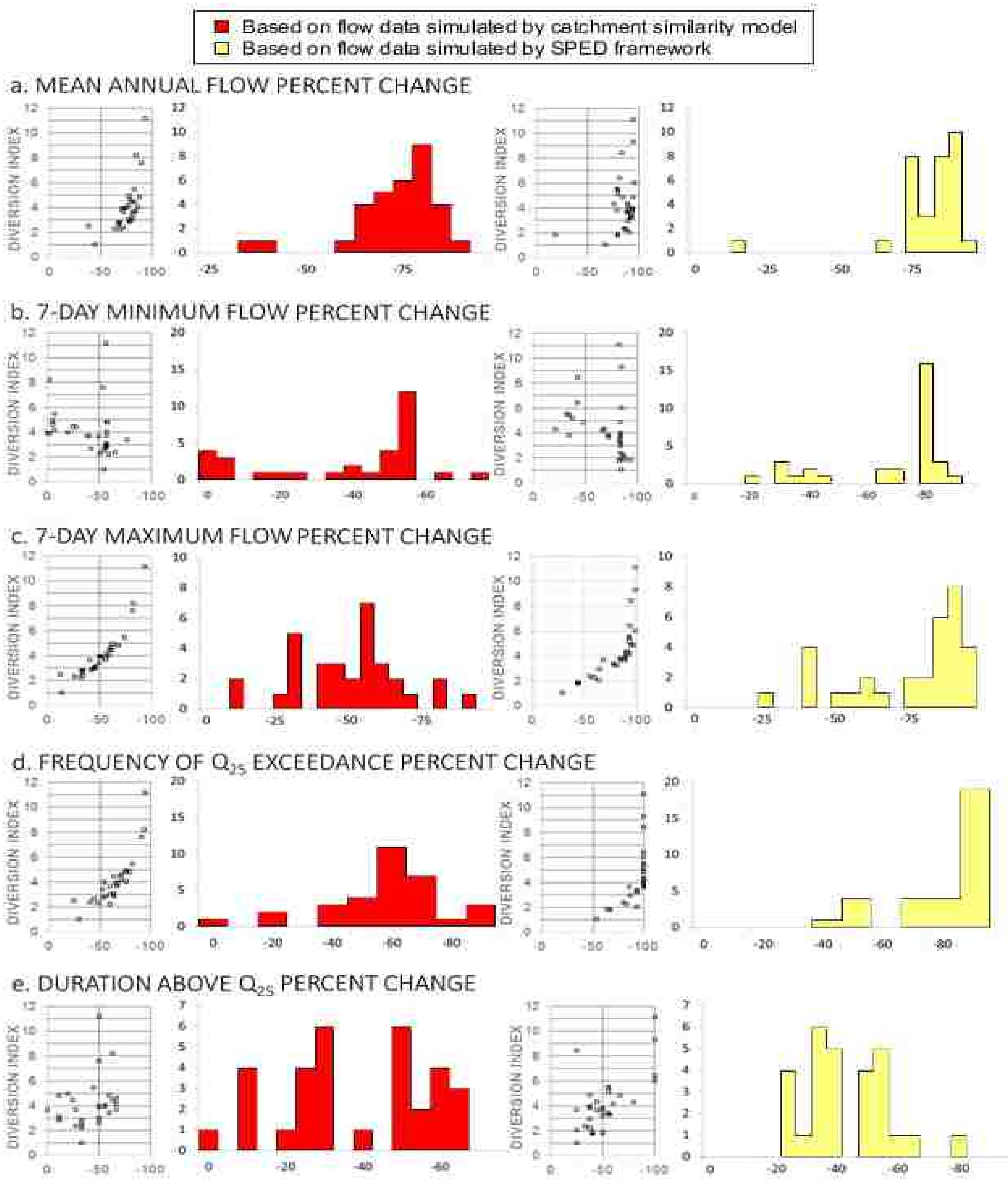

Figure 23. Histograms of percent change to flow metrics across the rivers and variation of response with diversion index

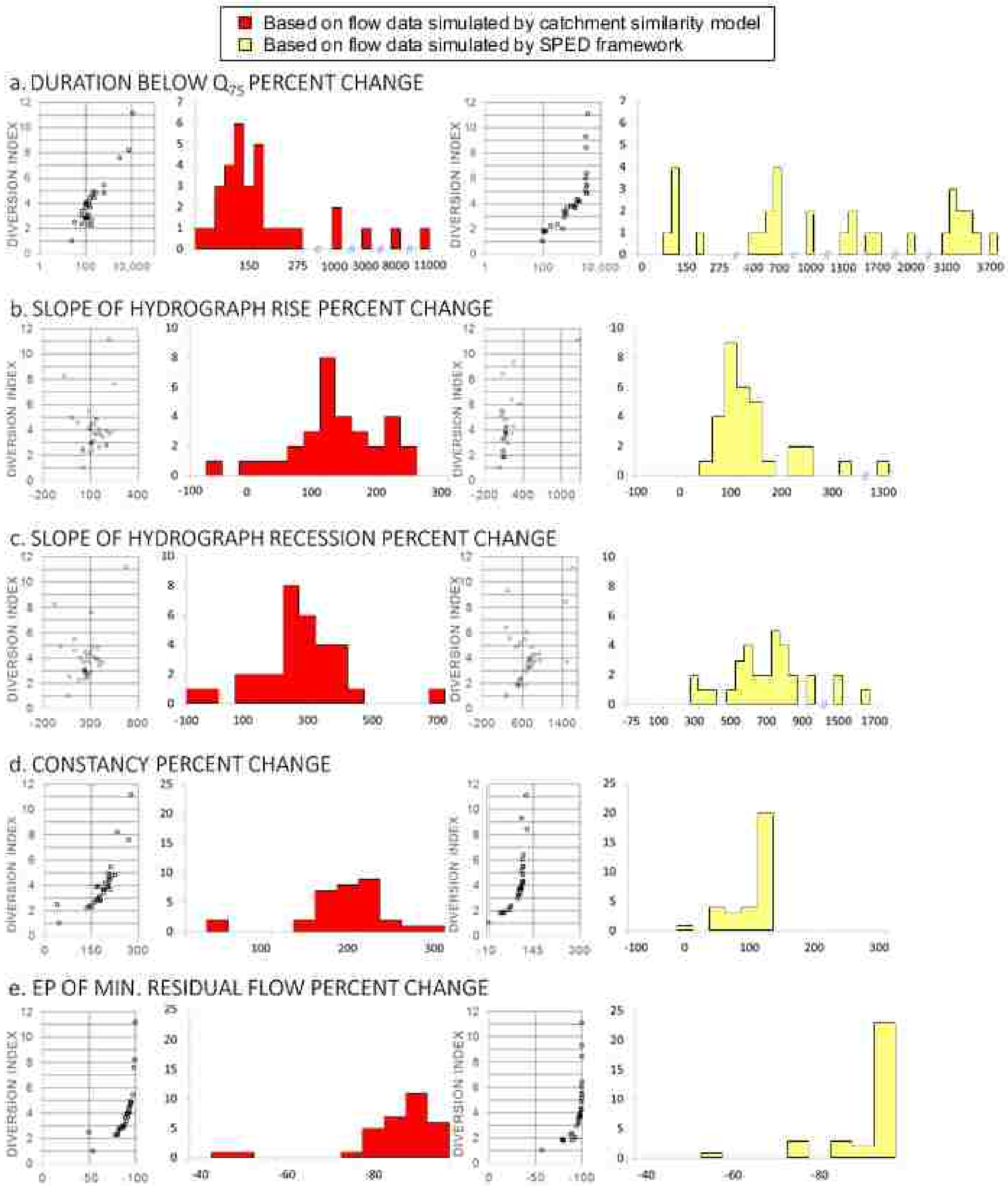

Figure 24. Histograms of percent change to flow metrics across the rivers and variation of response with diversion index

4.6 Discussion and Conclusions

4.6.1 The Flow Alteration Signal Supersedes Model Variability

The overall message remains the same regardless of the flow prediction method of choice: diversion hydropower alters flow signatures substantially (Fig. 23 and 24). Results from both the SPED and catchment similarity models indicate substantial changes to the magnitude and variability of flows, such that regulated hydrographs are often maintained at a static minimum residual flow for long periods of time (Fig. 22). Variability is suppressed, especially in low to moderate flow magnitudes (Fig. 23 and 24). The diversion index correlates well with severity of alteration, such that rivers with the highest diversion indices exhibit longer periods of flow at minimum residual value and consequently more substantial alterations to flow metrics. For instance the striking level of change in the duration of low flow events (from a mean of 11 [3] days to a mean of 175 [27] days, Fig. 24a) is mainly attributed to rivers with the highest diversion indices while rivers with low diversion indices showed only little difference with this regard. Thus, although diversion hydropower usually requires construction of only a small dam or weir, the impacts particularly to reaches immediately downstream of the dam may be consequential to the aquatic ecosystem or downstream water users.

4.6.2 Data Simulated by SPED Indicate More Severe Hydrologic Alteration

While direction and pattern of hydrologic alterations detected are consistent between analyses using the SPED framework and the catchment similarity approach (Fig. 23 and 24), the severity of effects vary systematically across the two prediction models. Alterations are consistently estimated to be more severe when flows simulated by the SPED framework are analyzed (Fig. 23

and 5). Particular attention to the metric of flow constancy may reveal the reason for this divergence. Constancy of flow describes uniformity of flow events through time. High constancy implies that flow variability is low through time and consequently flow is more predictable. Lower constancy indicates substantial variability of flow through time and thus less predictable flow. Both flow prediction approaches detect substantial alteration to constancy in most rivers as a result of diversion (Fig. 24d), which suggests that flow has become more uniform and less variable due to diversion. Mean relative alteration (percent change) in flow constancy based on the SPED framework (92 ± 29% mean increase) is lower than that detected with the catchment similarity model [184 ± 49% mean increase]. However, across the 32 rivers, the SPED framework also predicts that unregulated flow regimes have greater constancy as compared to the catchment similarity model (0.46 versus [0.26] before diversion). This difference is associated with how the two approaches predict flow. Both approaches use the same reference catchment (Yongchun) for extrapolating information to the ungauged target catchments. However, the SPED framework incorporates *a priori* parameter estimates from both reference and target catchments in calibration of the runoff model. This results in models better tuned to the behavior of individual catchments (Alipour and Kibler, 2018, Alipour and Kibler, in revision). For instance, *a priori* parameter estimates from study catchments indicate high capacities for storage of water within the soils of all target catchments. The SPED framework uses this *a priori* knowledge and calibrates parameters to reflect the high water storage capacity of soil in the catchments. The high storage capacity estimated by SPED translates into less variable flow due to higher and more stable base flows, even during dry periods, as well as moderated high flows even during wet periods. Consequently, the constancy of unregulated flows modeled by SPED are high. Indeed, a constancy value of 0.51

for observed flow in the reference catchment, which is close to the average of that modeled by SPED in the target catchments (0.46), substantiates the *a priori* information on high water storage capacity of soil in the reference catchment. One might think that the catchment similarity modeling approach, which is based on the assumption of full similarity between hydrologic routing processes in the reference and target catchments, should also be able to model the high water storage capacity of soil in the target catchments and estimate constancy close to that of the reference catchment. However, similarity in water storage capacity of soil in the reference and target catchments may not necessarily translate into similarity in terms of hydrologic routing processes. In fact, it could even be a source of dissimilarity if the catchments are of significant size difference. Thus, the catchment similarity model partially fails to address the high water storage capacity of soil in the target catchments and produces a mean constancy of 0.26 across the target catchments. The small quantity of observed flow data available to validate models also suggest that the SPED framework is indeed more adept at accurately predicting magnitudes of baseflows (Fig. 21).

4.6.3 Management Purpose of Flow Analysis Can Guide the Choice of Flow Prediction Approach

Significant alterations to the natural flow regime of the rivers were detected similarly using flows predicted through both a catchment similarity modeling approach and the SPED framework (Fig. 22, 23, and 24). However, the level of impact sometimes differed between the two approaches (Fig. 23 and 24). Thus, the choice of which approach a manager should use for flow prediction depends on the management purpose of the analysis. For instance, the SPED framework is likely to produce more accurate flow predictions, but requires more time and effort to implement as compared to the catchment similarity approach (Fig. 21). An approach such as SPED should be chosen when analysis is performed for a sensitive task, such as the design of instream flows for preservation of a riverine ecosystem. If the objective of flow alteration analysis is primarily to detect direction and patterns of change, the catchment similarity modeling approach may be appropriate.

A limitation to the present study, which is common to streamflow prediction in many poorly-gauged places of the world, is a lack of observed data to validate models. In this study, models can be directly validated based on observed data in only one catchment for one year. Managers can partially address such challenges through indirect validation. For example, in another nearby gauged catchment with 20 years of observed daily streamflows in the region (Yang Bi Jiang River), SPED predicted streamflow with high efficiency (NSE of 0.72) (Alipour and Kibler, 2018). While such indirect performance evaluation can increase confidence of model predictions, alternative techniques to validate model performance in truly ungauged catchments should become a fruitful research area. Innovative validation techniques, for instance using remotely-sensed data, citizen science, or crowd-sourced data, must be explored to evaluate streamflow prediction techniques for truly ungauged areas.

4.7 Acknowledgements

We would like to thank the China Hydrology Data Project (Henck et al., 2010) for digitizing observed data and making them available for research.

4.8 References

Alipour, M.H., Kibler, K.M., 2018. A framework for streamflow prediction in the world's most severely data-limited regions: Test of applicability and performance in a poorly-gauged region of China. Journal of Hydrology, 557.

Alipour, M.H., Kibler, K.M., in revision. Streamflow Prediction under Extreme Data-scarcity (SPED): A Step Toward Hydrologic Process Understanding within Severely Data-limited Regions. Hydrological Sciences Journal.

Anderson, D., Moggridge, H., Warren, P., Shucksmith, J., 2015. The impacts of 'run-of-river' hydropower on the physical and ecological condition of rivers. Water and Environment Journal, 29(2), pp.268-276.

APHRODITE website, 2015. Available at http://www.chikyu.ac.jp

Batjes, N.H., 2015. World soil property estimates for broad-scale modelling (WISE30sec, Ver. 1.0). Report 2015/01, ISRIC—World Soil Information, Wageningen (available via: http://www.isric.org/data/data-download).

Beven, K., 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. Advances in Water Resources, 16 (1), 41–51.

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

Bunn, S.E., Arthington, A.H., 2002. Basic principles and ecological consequences of altered flow regimes for aquatic biodiversity. Environmental Management, 30(4), 492-507.

Chinese Academy of Sciences, 1990. Qinghai Tibet Atlas [in Chinese]. Science Publishing House, Beijing, China.

DeFries, R.S., Hansen, M., 2010. ISLSCP II University of Maryland global land cover classifications, 1992-1993. In Hall, F.G., Collatz, G., Meeson, B., Los, S., Brown de Colstoun, E., Landis, D. (Eds.). ISLSCP initiative II collection. Data set. Available online [http://daac.ornl.gov/] from Oak Ridge national laboratory distributed active archive center, Oak Ridge, Tennessee, USA.

Falkenmark, M., Chapman, T., (eds.), 1989. Comparative hydrology: an ecological approach to land and water resources. UNESCO, Paris, France.

Fantin-Cruz, I., Pedrollo, O., Girard, P., Zeilhofer, P., Hamilton, S. K., 2015. Effects of a diversion hydropower facility on the hydrological regime of the Correntes River, a tributary to the Pantanal floodplain, Brazil. Journal of Hydrology, 531, 810–820.

Forsythe, W.C., Rykiel, E.J., Stahl, R.S., Wu, H.I., Schoolfield, R.M., 1995. A model comparison for daylength as a function of latitude and day of year. Ecological Modelling, 80 (1), 87-95.

Gibeau, P., Connors, B.M., Palen, W.J., 2016. Run-of-River hydropower and salmonids: potential effects and perspective on future research. Canadian Journal of Fisheries and Aquatic Sciences, 74(7), pp.1135-1149.

Gleick, P.H., 1992. Environmental consequences of hydroelectric development: the role of facility size and type. Energy, 17(8), pp.735-747.

Henck, A.H., Montgomery, D.R., Huntington, K.W., and Liang, C., 2010. Monsoon control of effective discharge, Yunnan and Tibet. Geology, 38, 975-978.

Hengl, T., de Jesus, J.M., Heuvelink, G.B., Gonzalez, M.R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., & Guevara, M.A. (2017). SoilGrids250m: Global gridded soil information based on machine learning. PLoS one, 12 (2), p.e0169748.

Hrachowitz, M., Savenije, H.H.G., Blöschl, G., McDonnell, J.J., Sivapalan, M., Pomeroy, J.W., Arheimer, B., Blume, T., Clark, M.P., Ehret, U., Fenicia, F., Freer, J.E., Gelfan, A., Gupta, H.V., Hughes, D.A., Hut, R.W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P.A., Uhlenbrook, S., Wagener, T., Winsemius, H.C., Woods, R.A., Zehe, E., Cudennec, C., 2013. A decade of Predictions in Ungauged Basins (PUB)—a review. Hydrological sciences journal, 58 (6), 1198-1255.

Institute of Water Resources (IWR), 2006. Alu River Power Station Feasibility Study, Yunnan Province, Fugong County [in Chinese]. Vol. I, Hydropower Survey and Design, Chenzhou, Hunan.

Kibler, K. M., 2017. The Hydraulic Size Of A Dam: A Metric Indicating Potential Hydrologic Impact For Storage And Diversion. International Journal of Hydrology, 1(4): 00020. DOI: 10.15406/ijh.2017.01.00020.

Kibler, K.M., Alipour, M.H., 2017. Flow alteration signatures of diversion hydropower: An analysis of 32 rivers in southwestern China. Ecohydrology, 10 (5).

Kibler, K.M., Tullos, D.D., 2013. Cumulative biophysical impact of small and large hydropower development in Nu River, China. Water Resour. Res. 49 (6), 3104–3118.

Lafon, T., Dadson, S., Buys, G., Prudhomme, C., 2013. Bias correction of daily precipitation simulated by a regional climate model: a comparison of methods. International Journal of Climatology, 33 (6), 1367-1381.

Li, B.S., Zhou, P.J., Wang, X.Y., Zhu, L., 2013. Opportunities and eco-environmental influence of cascade hydropower development and water diversion projects in Hanjiang river basin. Journal of the Geological Society of India, 82(6), pp.692-700.

Mathews, R., Richter, B.D., 2007. Application of the indicators of hydrologic alteration software in environmental flow setting. Journal of the American Water Resources Association 43(6), 1400-1413.

Mekong River Commission, 2005. Overview of the Hydrology of the Mekong Basin. Mekong River Commission, Vientiane, p. 82.

Merritt, D.M., Poff, N., 2010. Shifting dominance of riparian Populus and Tamarix along gradients of flow alteration in western North American rivers. Ecological Applications, 20(1), pp.135-152.

Moore, R.J., 1985. The probability-distributed principle and runoff production at point and basin scales. Hydrological Sciences Journal, 30 (2), 273-297.

Olden, J. D., Poff, N. L., 2003. Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. River Research and Applications, 19(2), 101-121.

Peñas, F.J., Barquín, J., Snelder, T.H., Booker, D.J., Álvarez, C., 2014. The influence of methodological procedures on hydrological classification performance. Hydrology and Earth System Sciences, 18, 3393-3409.

Poff, N.L., Zimmerman, J.K., 2010. Ecological responses to altered flow regimes: a literature review to inform the science and management of environmental flows. Freshwater Biology, 55(1), pp.194-205.

Richter, B.D., Baumgartner, J.V., Powell, J., Braun, D.P., 1996. A method for assessing hydrologic alteration within ecosystems. Conservation biology, 10(4), pp.1163-1174.

Richter, B.D., Baumgartner, J.V., Wigington, R., Braun, D.P., 1997. How much water does a river need? Freshwater biology, 37(1), 231-249.

Seibert, J., McDonnell, J.J., 2013. Gauging the ungauged basin: relative value of soft and hard data. Journal of hydrologic engineering, 20 (1).

The Nature Conservancy website, 2017. Available online at: www.conservationgateway.org

Thornthwaite, C.W., 1948. An approach toward a rational classification of climate. Geographical review, 38 (1), 55-94.

Timpe, K., Kaplan, D., 2017. The changing hydrology of a dammed Amazon. Science advances, 3(11), p.e1700611.

Tuna, M.C., 2013. Feasibility assessment of hydroelectric power plant in ungauged river basin: a case study. Arabian Journal for Science and Engineering, 38(6), pp.1359-1367.

van Emmerik, T., Mulder, G., Eilander, D., Piet, M., Savenije, H., 2015. Predicting the ungauged basin: Model validation and realism assessment. Frontiers in Earth Science, 3 (62).

Wang, H., Chen, Y., Liu, Z., Zhu, D., 2016. Effects of the “Run-of-River” Hydro Scheme on Macroinvertebrate Communities and Habitat Conditions in a Mountain River of Northeastern China. Water, 8(1), p.31.

Wi, S., Yang, Y.C.E., Steinschneider, S., Khalil, A., Brown, C.M., 2015. Calibration approaches for distributed hydrologic models in poorly gaged basins: implication for streamflow projections under climate change. Hydrology and Earth System Sciences, 19 (2), 857-876.

Yasutomi, N., Hamada, A., Yatagai, A., 2011. Development of a long-term daily gridded temperature dataset and its application to rain/snow discrimination of daily precipitation. Global Environmental Research, 15(2), pp.165-172.

Yatagai, A., Kamiguchi, K., Arakawa, O., Hamada, A., Yasutomi, N., Kitoh, A., 2012. APHRODITE: Constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges. Bulletin of the American Meteorological Society, 93(9), pp.1401-1415.

Yunnan Bureau of Hydrology and Water Resources, 2005. Gutan River Ecological Hydropower Station Project Water Resources Appraisal, Nujiang Prefecture, Lushui [in Chinese]. Yunnan Bureau of Hydrology and Water Resources, Kunming, China.

Zhang, Y., Zhai, X., Zhao, T., 2018. Annual shifts of flow regime alteration: new insights from the Chaishitan Reservoir in China. Scientific reports, 8(1), p.1414.

CHAPTER 5:
VALUE ADDED BY GAUGING THE UNGAUGED CATCHMEN WITHIN A DATA-SCARCE REGION

5.1 Preface

This chapter describes the valued added by limited streamflow observations in an ungauged catchment to improvement of accuracy of long term flow predictions in the catchment.

5.2 Abstract

Streamflow prediction in ungauged basins has become more important than ever given the increasing developments within small catchments located in remote areas of the world. The value of few observations collected through limited field campaigns to improvement of predictions has received attention recently through several research studies. We study some specific situations a manager of an ungauged catchment located within a data-scarce region may face in reality, which have not been explored previously by other researchers. Our experimental design is thus for a situation where decisions on a development project within the ungauged catchment have to be made quickly. We study whether it is beneficial to go ahead with a limited field campaign within a timeframe of maximum one year to collect observed streamflow data in the ungauged catchment. We take a traditional streamflow prediction approach in such regions based on calibration in a partially similar gauged catchment to transfer parameter values to the ungauged catchment as reference. We then test 528 different scenarios for combination of this traditional approach with the collected observed data in the ungauged catchment to see if improvements are made possible and to what extent. This combination is realized through defining a two-criterion objective function considering runoff efficiency of calibration in the reference catchment as well as runoff

efficiency of calibration in the ungauged catchment using the collected data. The scenarios are based on 11 different schemes for selection of the days for data collection, 6 different combinations for the weights of the two criteria, and 8 different numbers of data collection days. We apply our approach to ten catchments located on four different continents including six catchments in the United States. We find that there are two scenarios which almost universally lead to improvements in runoff efficiency of predictions over the traditional approach.

5.3 Introduction

Streamflow data are a significant element required for water resources planning and management projects. Degree of success in activities such as river restoration, reservoir operation, water allocation and many other water development projects is highly associated with the availability and quality/accuracy of long-term streamflow data. However, World Meteorological Organization has made it clear that even the minimum recommended density of streamflow gauges required for basic water resources management has not been met in many areas of the world (Perks et al., 1996). Ironically, the number of gauging stations worldwide is even in decline (Lanfear et al. 1999, Vörösmarty et al. 2001, Shiklomanov et al. 2002, Hannah et al. 2011). Streamflow modeling and prediction may thus be the only way to meet the requirements for long-term streamflow data. In acknowledgment of this increasing demand and in response to the desire to shift research focus toward process understanding and model structural diagnostics (Hrachowitz et al., 2013), the International Association of Hydrological Sciences initiated the Prediction in Ungauged Basins (PUB) decade (2003-2012). Significant advancements were made during this decade of research with respect to multiple aspects of the science of flow prediction. In this context, contribution and the value added by a limited number of streamflow measurements in the field to improvement of long term streamflow predictions has recently been considered and studied by researchers (e.g., Drogue and Plasse, 2014; Perrin et al., 2007; Viviroli and Seibert, 2015).

Rojas-Serna et al. (2006) calibrated the four-parameter GR4J model (Edijatno et al., 1999; Perrin et al., 2003) using only *a priori* knowledge acquired from the large number of prior calibration experiments as well as a combination of both *a priori* knowledge and a few streamflow measurements in the ungauged catchments of the study. Substantial improvements in runoff efficiency of predictions were observed by incorporating up to 50 daily streamflow measurements in the calibration process. In a relevant study, Perrin et al. (2008) tested calibration of two lumped daily rainfall-runoff models using only a library of prior parameter estimates from a large number of catchments. The method was found sensible and more robust in comparison to classical calibration schemes when the available data for calibration were less than 2 years long. This would make the approach applicable in poorly gauged catchments where there is limited availability of streamflow measurements. Combination of few streamflow measurements with groundwater well level observations (Juston et al., 2009), glacial mass balance observations (Konz and Seibert, 2010) and soft data (on maximum and minimum groundwater levels, frequencies of groundwater levels as well as the contribution of new event precipitation water to event runoff) (Seibert and McDonnell, 2013), and the consequential improvements in model calibration/prediction have also been subject to research recently. Seibert and Beven (2009) calibrated the HBV mode (Bergström, 1992; Lindström et al., 1997) in 11 catchments located in central Sweden north of Uppsala. Uniform distributions were used to sample 10,000 random parameter sets. Limited number of daily runoff measurements from a varying one-year period (of 1, 2, 4, 8, 16, 32, 64, 128 and 256) were then added to evaluate the performance of the random parameter sets and rank them. Top 100 parameter sets were evaluated for runoff efficiency over a 10-year period in each catchment. The authors found that, on average, a few daily measurements could importantly improve model performance. However, there were significant variations in model performance based on the catchment being modeled as well as the days chosen for runoff measurement. Perrin et al. (2007) studied 12 catchments located in the United States with continuous high-quality streamflow data of 39 years. They chose different number of observed streamflow data randomly distributed over these streamflow records for calibration of two hydrological models. While drier catchments proved more difficult, in general models reached stability in parameter estimates when the number of observed streamflow data were greater than 350.

In this study we aim at answering some very specific questions, which have not been explored in previous studies, regarding the potential advantages of making few streamflow measurements in an ungauged catchment located within a poorly-gauged region. We design these questions in a realistic scenario where a manager has to make a decision on whether or not to go ahead with a limited field campaign for collection of streamflow data. Our experimental design is thus developed for a case where we have an ungauged target catchment which is going to be developed with a water resources project such as diversion hydropower. Time is essential here as development plans need to be finalized as quickly as possible, and there is only limited time for evaluation studies such as a data collection field campaign. We assume there is a maximum of one year of time available for completion of data collection and making streamflow predictions. To mimic circumstances of data-scarce regions, we also limit the availability of a similar gauged reference catchment to only one catchment with partial similarity to the target catchment of interest. So the first question to answer is whether or not it is beneficial to have the data collection field campaign. To elaborate, we want to answer whether conducting the field campaign and combining the collected data from the ungauged target catchment with the available streamflow data in the reference catchment is more favorable and advantageous than the traditional streamflow prediction in the ungauged target catchment using only data from the gauged reference catchment. Moreover, in a real case we would not know for sure what days of the year ahead are going to experience high flows, low flows or even moderate flows so that we can schedule field measurements on these specific days. We would only have the historical data from the gauged reference catchment to give us hints on this. We would not know how many measurements are required to ensure improvement in modeling over the traditional approaches either. Thus, another question we would like to explore is that whether there is a (almost) universal approach that can be followed to ensure with high probability that improvements are going to be made over traditional approaches, and if so, what are the required resources associated with that. This question is particularly important since in truly ungauged catchments we do not have streamflow data to validate the performance of a model. Thus, in order to go ahead with a field data collection campaign, a manager needs to know that improvements are almost guaranteed in comparison to what can be achieved right away without allocating resources to this task. We aim at answering these questions by testing a large number of scenarios for field data collection, including number and timing of measurements, and combination of these data with streamflow data available in a gauged reference catchment.

5.4 Methodology

5.4.1 Multi-criteria Objective Function for Calibration

We base our analysis on traditional single-objective calibration to maximize runoff efficiency in a gauged reference catchment and transferring the calibrated parameter values to an ungauged target catchment for prediction. We however test if predictions could be improved by adding a limited number of daily streamflow measurements in the ungauged target catchment to this process. Since we work in poorly-gauged regions or mimic circumstances of such regions in well-gauged regions, we use only one gauged reference catchment in each region which is at least partially similar to the ungauged target catchment for calibration of the hydrological model of choice. Thus, the objective function of the calibration process consists of two criteria: maximizing runoff efficiency in the gauged reference catchment and maximizing runoff efficiency based on the limited available observations in the ungauged target catchment. Calibrated parameter values are then used for long-term predictions (5-10 years) in the ungauged target catchment.

5.4.1.1 Runoff Efficiency in the Gauged Reference Catchment

Nash-Sutcliff Efficiency (NSE) is used in the reference catchment in each region as the metric to measure the runoff efficiency of the parameter sets tested during the calibration process:

$$NSER = 1 - \frac{\sum_{t=1}^{T}\left(Q_o^t - Q_m^t\right)^2}{\sum_{t=1}^{T}\left(Q_o^t - \overline{Q_o}\right)^2} \qquad (14)$$

Where NSER is Nash-Sutcliffe runoff efficiency in the reference catchment, Q_o^t is observed streamflow in time step *t*, Q_m^t is modeled streamflow in time step *t*, and $\overline{Q_o}$ is the mean value of observed streamflows in time *T*.

5.4.1.2 Runoff Efficiency in the Ungauged Target Catchment

NSE is also used for the limited available streamflow observations in the target catchment in each region as the metric to measure the runoff efficiency of the parameter sets tested during the calibration process:

$$NSET = 1 - \frac{\sum_{n=1}^{N}(Q_o^n - Q_m^n)^2}{\sum_{n=1}^{N}(Q_o^n - \overline{Q_o})^2} \qquad (15)$$

Where NSET is Nash-Sutcliffe runoff efficiency in the target catchment, Q_o^n is observed streamflow for the nth measurement day in the target catchment, Q_m^n is modeled streamflow for the nth measurement day in the target catchment, and $\overline{Q_o}$ is the mean value of observed streamflows for all *N* available measurement days in the target catchment.

5.4.1.3 Weighting the Criteria

In order to explore the entire domain of potential solutions to our research questions, we discretize the weight range for each criterion (0-1) and test all their possible combinations (Table 13).

Table 13. Weight combinations for the criteria

Combination	NSER weight (%)	NSET weight (%)
1	0	100
2	10	90
3	30	70
4	50	50
5	70	30
6	90	10

The final form of the multi-criteria objective function used for model calibration is thus as follows:

$$OF = W_1 * NSER + W_2 * NSET \quad (16)$$

Where OF is the multi-criteria objective function to be maximized, W_1 is the weight assigned to maximization of NSE in the reference catchment, and W_2 is the weight assigned to maximization of NSE in the target catchment. NSER, NSET and OF all range between negative infinity and one with one being the ideal value where observed and modeled streamflows match through the entire analysis period.

5.4.2 Scenarios

As described in the Introduction section, in this study we focus on a realistic case that a manager in a poorly-gauged region might face. In this case, the manager has to make a decision whether or not to go ahead with a limited field campaign for collecting streamflow data in an ungauged catchment which is going to be developed by a water resources project. We limit the available time for the field campaign before beginning of planning/development to one year. Thus, in the regions we study, we assume that limited daily streamflow measurements are available in the ungauged target catchments only within the last year of calibration period (assuming they have been collected through the field campaign). Other important factors to consider are the number and timing of these measurements. In our search for a potentially widely applicable solution, we aim at investigating almost the entire range of the number of collection days. Thus, we test a wide range of possible scenarios having 2, 4, 8, 16, 32, 64, 128 or 256 collection days.

With respect to the timing of measurements, we again consider a realistic case where at the beginning of the field campaign the best source of information about the flow regime (timing of low flows, high flows, moderate flows, etc.) of the ungauged target catchment within the next year is the historical streamflow data in the gauged reference catchment. Thus, since the reference and target catchments must be at least partially similar, we rely on the available streamflow data in the reference catchment (up to the beginning of the field campaign) to estimate the flow regime in the target catchment within the following year. Using this information we select the data collection days. Here again, since we search for a widely applicable solution, we investigate a wide range of scenarios for selection of data collection days. To do so, first, historical daily streamflow data in the gauged reference catchment are averaged for each day of the year over the entire years of available data. Next, days of the year are ranked based on their associated average flow value (ranking 1). This ranking gives us an estimation of what days of the year are more likely to experience highest flows (days with highest ranks), lowest flows (days with lowest ranks), median flows (days with median ranks), etc. Moreover, average value of daily streamflow over all days and all years is calculated and the absolute value of its difference with daily averages is calculated. Days are then ranked based on this difference (ranking 2). This ranking gives us an estimation of what days of the year are more likely to experience flows that are closest to the flow average over the entire available data (days with lowest ranks). Finally, 11 different scenarios are defined based on these two rankings to select the filed data collection days in the ungauged target catchment, including:

1. Half of the collection days are selected from the days with the highest rankings (ranking 1) and half of them are selected from the days with the lowest rankings (ranking 1);

2. All collection days are selected from the days with the lowest rankings (ranking 1);

3. All collection days are selected from the days with the highest rankings (ranking 1);

4. All collection days are selected from the days with the median rankings (ranking 1; for example day 182 and 183 if we only have two collection days, or days 181, 182, 183, 184 if we have four collection days);

5. Half of the collection days are selected from the days with the highest rankings (ranking 1) and half of them are selected from the days with the median rankings (ranking 1);

6. Half of the collection days are selected from the days with the lowest rankings (ranking 1) and half of them are selected from the days with the median rankings (ranking 1);

7. Half of the collection days are selected from the days with the highest rankings (ranking 1) and half of them are selected from the days with the lowest rankings (ranking 2).

8. Half of the collection days are selected from the days with the lowest rankings (ranking 1) and half of them are selected from the days with the lowest rankings (ranking 2).

9. All collection days are selected from the days with the lowest rankings (ranking 2);

10. One third of the collection days are selected from the days with the lowest rankings (ranking 1), one third of the collection days are selected from the days with the highest rankings (ranking 1) and one third of the collection days are selected from the days with the median rankings (ranking 1);

11. One third of the collection days are selected from the days with the lowest rankings (ranking 1), one third of the collection days are selected from the days with the highest rankings (ranking 1) and one third of the collection days are selected from the days with the lowest rankings (ranking 2).

Given the 8 tested scenarios for the number of collection days, 11 tested scenarios for the timing of collection days, and 6 tested combinations of weights for the criteria in the OF, a total of 528 scenarios are tested for each reference/target catchment pair in search of a widely applicable solution.

5.4.3 Hydrological Model of Choice

HyMOD hydrological model (Moore 1985, 1999) was selected for evaluating the performance of the scenarios in this study. This is a lumped conceptual model which uses simulated probability distribution of soil moisture across a catchment for streamflow modeling (Wang et al., 2009). This cumulative distribution function is formulated as follows:

$$F(c) = 1 - \left[1 - \frac{c}{C_{max}}\right]^{B}, 0 \leq c \leq C_{max} \qquad (17)$$

Where F is the cumulative probability, c is soil moisture capacity, C_{max} is the maximum soil moisture capacity across the catchment, and B is a shape factor associated with the degree of spatial variability in soil moisture capacity across the catchment. R_q (inverse of residence time in quick reservoirs), R_s (inverse of residence time in a slow reservoir), and α (a fraction coefficient for distribution of water between slow and quick reservoirs) are the other parameters of HyMOD. For interested readers, Wang et al. (2009) and Moore (1985, 1999) provide a more detailed description of HyMOD. In this study, a modified formulation of the 5-parameter HyMOD is used. This includes addition of a minimum soil moisture capacity parameter, C_{min}, which may be greater than zero (Alipour and Kibler, in revision). Thus, the new soil moisture capacity CDF is formulated as:

$$F(c) = 1 - \left[1 - \frac{c - C_{min}}{C_{max} - C_{min}}\right]^{B}, C_{min} \leq c \leq C_{max}, 0 \leq C_{min} \leq C_{max} \qquad (18)$$

HyMOD was calibrate by running 62,5000 different parameter combinations in each reference/target catchment pair using the branch-and-bound method described by Alipour and Kibler (2018).

5.4.4 Catchments of Study

Five of the target catchments where we apply and test our approach come from the study by Alipour and Kibler (in revision). These include River Coquet in the United Kingdom, Broken River at Urannah and at Old Racecourse in Australia, North Fork Cache Creek in the US, and YBJ River in China. YBJ River is located within a truly poorly gauged region while the other regions were transformed into synthetically data-scarce regions by assuming availability of only one gauged reference catchment in each region and by using lower-quality (regional/global scale) data to estimate precipitation and temperature (Table 13). More detailed information on the catchments is available through the study by Alipour and Kibler (in revision).

To expand the test of applicability and reliability of our approach, we added five other catchments located in the United States. The catchments were mainly located in areas with negligible snow contribution so that the 6-parameter HyMOD model used in our study would be applicable (this version of HyMOD does not account for snowmelt). The catchments include Pea River near Ariton (Alabama), Murder Creek below Eatonton (Georgia), Bayou Grand Cane near Stanley (Louisiana), Horse Creek near Arcadia (Florida) and Myakka River at Myakka City (Florida). These regions were transformed into synthetically data-scarce regions by assuming availability of only one gauged reference catchment in each region.

5.5 Results

There is considerable variability with respect to flow prediction performance from catchment to catchment and also depending on the scenario (weights of criteria, number of daily observations, and timing of daily observations) being tested (Fig. 25, 26 and 27).

5.5.1 *Number of Runoff Observations*

Generally, an increase in the number of observation days up to 128 is associated with an increasing trend in the flow prediction performance in the ungauged target catchments (Fig. 25). However, this is associated with significant variability for individual scenarios and from catchment to catchment. From 128 to 256 observation days, the prediction performance varies more significantly from catchment to catchment, and can be ascending or descending (Fig. 25).

5.5.2 *Timing of Data Collection*

Timing of the data collection days in the ungauged target catchments can substantially influence the prediction performance in these catchments (Fig. 26). There is not a single scenario for timing of data collection that clearly outperforms the other scenarios. Variability is high from catchment to catchment and from scenario to scenario (Fig. 26).

5.5.3 *Weights of Criteria*

Assigning more weight to the NSE of modeling in the gauged reference catchments (NSER) over the NSE of modeling in the ungauged target catchments (NSET) leads to an increasing trend in the streamflow prediction performance in the target catchments (Fig. 27). However, this trend is less visible and sometimes even reverses between weight combinations of 5 (NSER weight = 70% and NSET weight = 30%) and 6 (NSER weight = 90% and NSET weight = 10%).

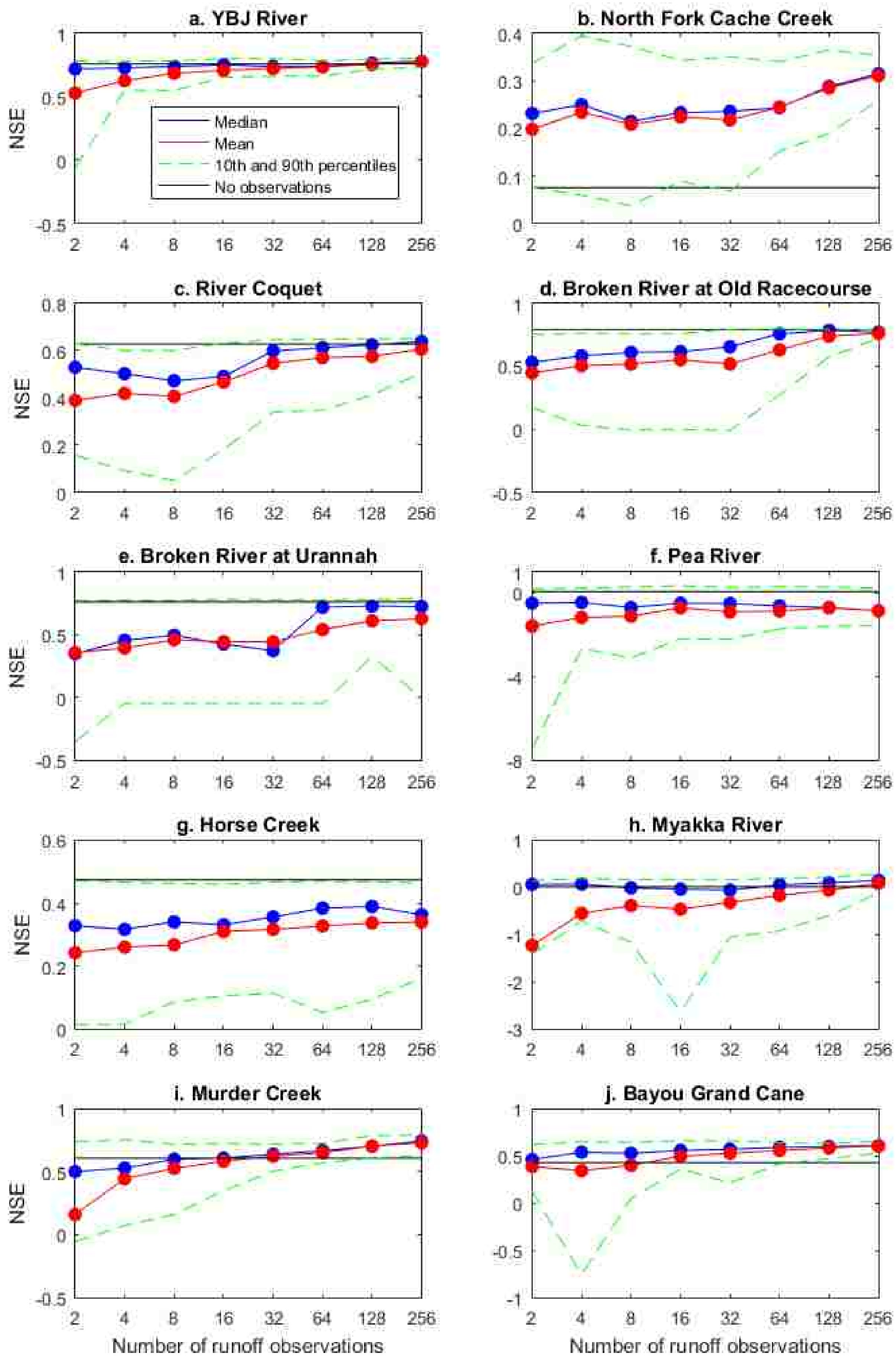

Figure 25. Median, mean, 10th and 90th percentile of prediction NSEs for corresponding number of runoff observations in the target catchments, and prediction NSEs for no observations in the target catchments

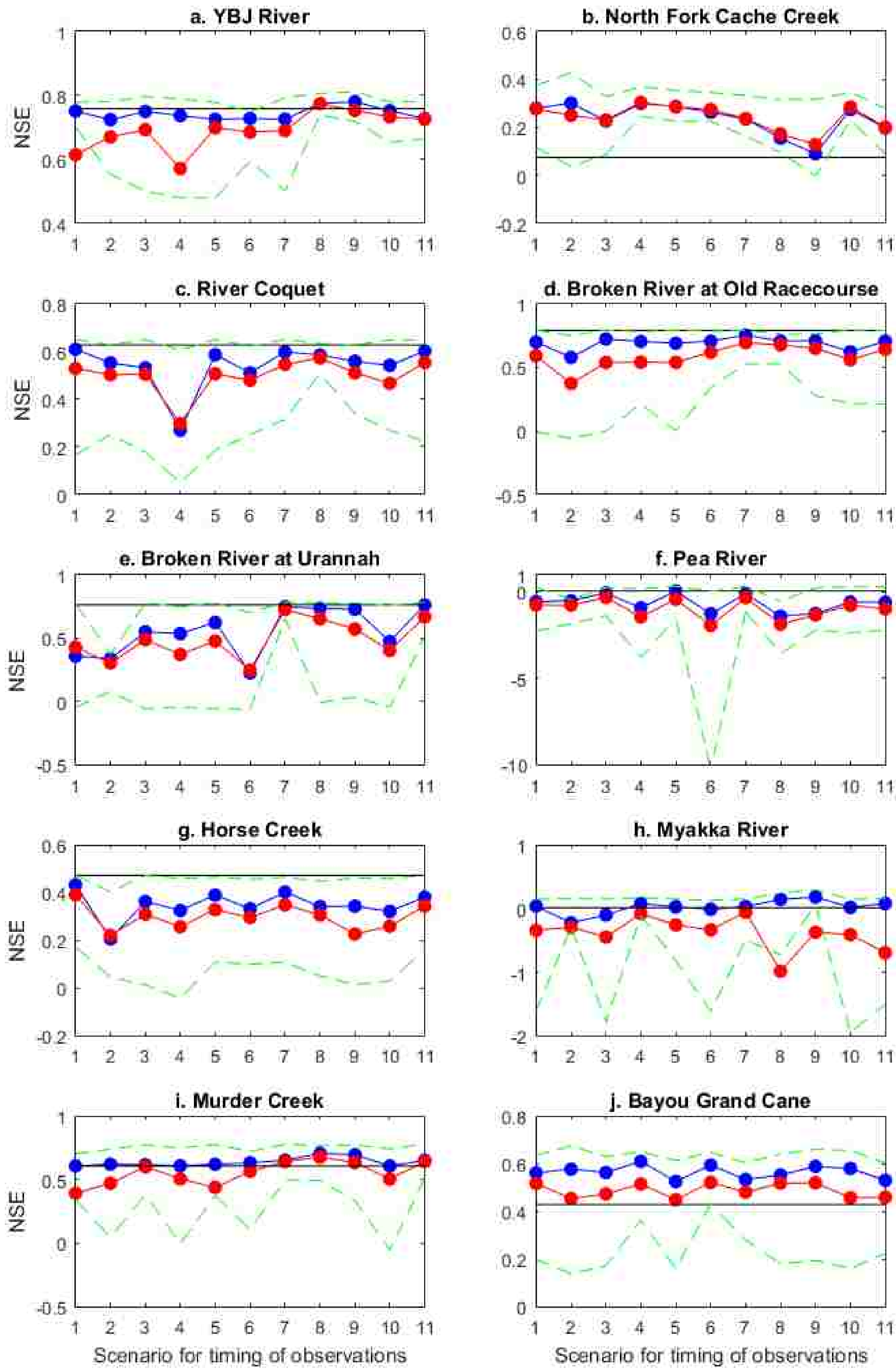

Figure 26. Median, mean, 10^{th} and 90^{th} percentile of prediction NSEs for corresponding field data collection scenarios in the target catchments, and prediction NSEs for no observations in the target catchments

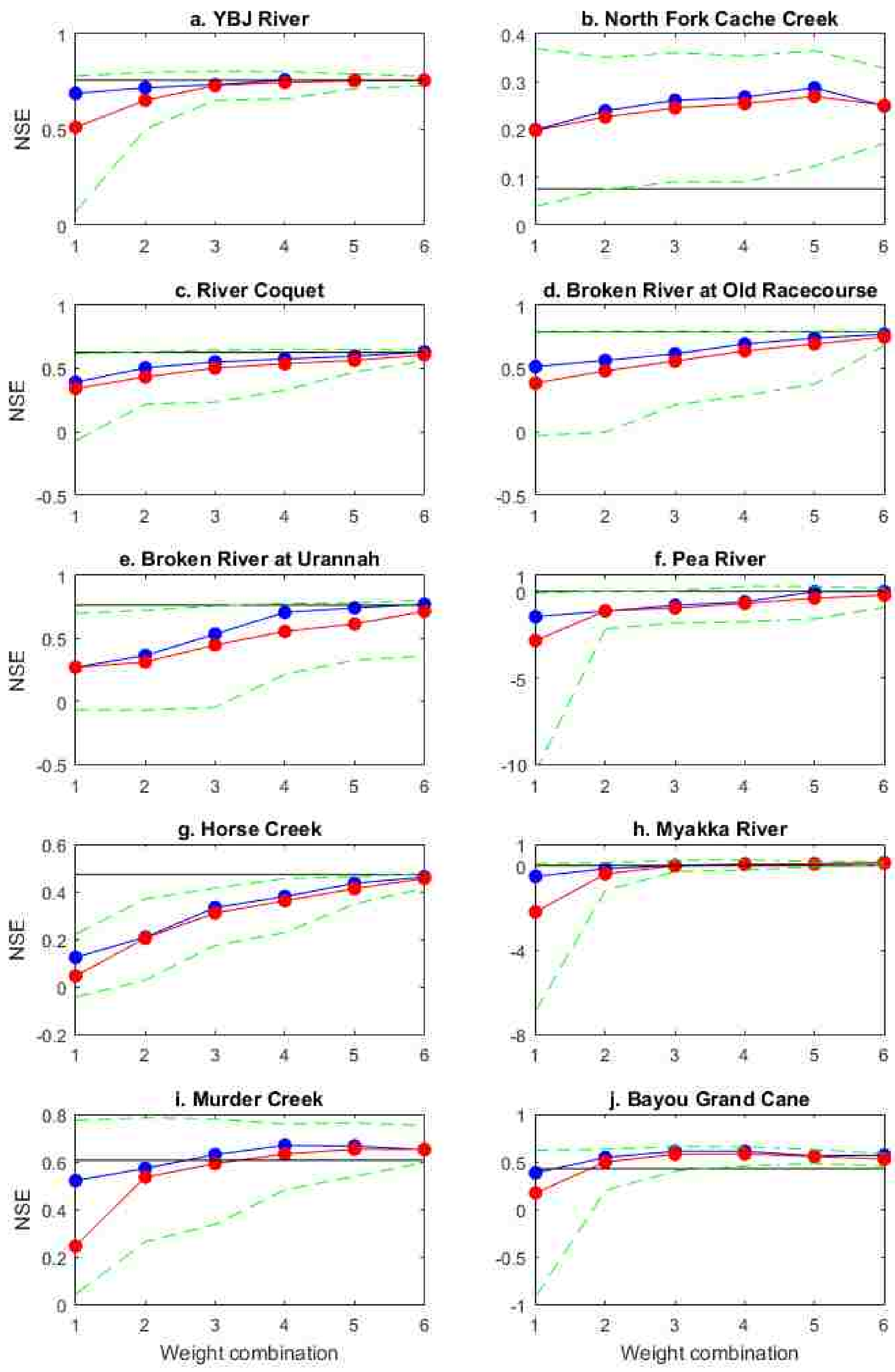

Figure 27. Median, mean, 10th and 90th percentile of prediction NSEs for corresponding weight combinations in the target catchments, and prediction NSEs for no observations in the target catchments

5.6 Discussion

Streamflow prediction performance of different scenarios, based on weights of criteria, timing of data collection days, and number of days for data collection, indicate that almost in all of the target catchments improvements are possible through incorporation of limited observed data in these catchments with available observed data in the gauged reference catchments (Fig. 25, 26 and 27). However, in some catchments there are only very few scenarios that result in improvements, while in others a large number of scenarios may lead to substantial improvements in streamflow prediction performance (Fig. 25, 26 and 27). Among the three variables defining the scenarios in each catchment, weights of criteria display the clearest trend with respect to NSE of predictions so that almost in all catchments assigning a weight of 0.7 or higher to NSER is preferable (Fig. 27). However, it should be noted that the observations (Fig. 27) are for median, mean and confidence intervals. Thus, there might still be individual scenarios with an NSER weight of less than 0.7 that perform better than scenarios with a higher weight for NSER. Moreover, the fact that the increasing trend of NSE sometimes stops or even reverses between weight combinations of 5 (NSER weight = 70% and NSET weight = 30%) and 6 (NSER weight = 90% and NSET weight = 10%) proves that there actually are cases where limited observations in the ungauged target catchments can supplement the data in the gauged reference catchments and improve the streamflow prediction performance in the target catchments.

There is an increasing trend of NSE with respect to increase in the number of observation points (data collection days) as well (Fig. 25). Similar to weights of criteria, however, sometimes this trend stops or reverses between 128 and 256 observation points. While we take it into consideration that there may be individual scenarios with fewer observation points than 128 that work better than scenarios with higher observation points, it is safe to say that the best results are probably achieved with at least 128 observation points in the ungauged target catchments. The choice of timing for data collection days also indicates very important impacts on the flow prediction performance in the target catchments (Fig. 26). However, there is not a clear trend based on this factor alone and there is no single timing scenario that indicates a better performance than the other timing scenarios when we consider mean, median and confidence intervals of predictions (Fig. 26). This indicates that other factors such as the catchment of study, number of observation days and weighting of criteria are also influential on the streamflow prediction performance in the target catchments.

5.6.1 Search for a Universal Solution for Improvement of Streamflow Prediction Efficiency

We analyzed the NSE performance of all 528 scenarios in each single target catchment of the study in search of a single scenario that led to improvements in flow prediction performance in all ten target catchments. Our search indicated that there was no single scenario that was advantageous over the traditional streamflow prediction method in all ten target catchments. However, we did find two near universal scenarios that led to improvements in nine of the target catchments and performed close to the traditional approach in the remaining target catchment. The two scenarios were both for 128 data collection days and an NSER weight of 0.7 (NSET weight of 0.3). The only difference between the two scenarios is with respect to the data collection timing. One of the scenarios is associated with timing scenario 10 and the other with timing scenario 11. The catchment where improvements were not made over the traditional approach for these two scenarios is Horse Creek in Florida. The results (Fig. 25, 26 and 27) indicate that there are only very few scenarios that lead to partial improvements in this catchment. While the two scenarios do not lead to improvements in this catchment, their prediction performance (NSEs of 0.42 and 0.45) is close to the prediction performance of the traditional flow prediction method (NSE of 0.47). Overall, the results indicate that it is safe and advantageous to choose one of these two scenarios for limited field data collection in an ungauged target catchment to supplement calibration process in a similar gauged reference catchment and improve the flow prediction performance in the target catchment. The two scenarios respectively lead to an average NSE improvement of 0.06 and 0.07 in the 10 target catchments.

5.7 Conclusions

Value added by a limited field data collection campaign to improve the accuracy of streamflow prediction in an ungauged catchment located within a data-scarce region was studied. In this sense, a realistic case was studied where there is up to one year of time available for completion of the field campaign. Calibration only in a gauged reference catchment was taken as reference traditional approach, and two-criterion calibration to maximize runoff efficiency in the gauged reference catchment and maximize runoff efficiency in the target catchment based on the limited collected data was tested to explore potential improvements. A total of 528 scenarios were tested by combination of 11 scenarios for timing of the data collection days, 8 scenarios for the number of data collection days and 6 scenarios for the weights of criteria. Ten pairs of reference/target catchment were studied through this approach including six target catchments in the United States. The results indicated high variability from catchment to catchment and from scenario to scenario. However, two overall trends were discernible where better performances were normally achieved as the number of data collection days increased from 2 to 128, and the weight assigned to the criterion of maximizing NSE in the reference catchment increase from 0% to 70%. The search for a universal scenario that resulted in improvements over the traditional approach in all ten target catchments indicated that none of the scenarios were universal in this sense. However, there existed two near-universal scenarios that results in improvements in nine of the target catchments and performed closely to the traditional approach in the remaining catchment.

5.8 References

Alipour, M.H., and Kibler, K.M., 2018. A framework for streamflow prediction in the world's most severely data-limited regions: Test of applicability and performance in a poorly-gauged region of China. *Journal of Hydrology*, 557.

Alipour, M.H., and Kibler, K.M., in revision. Streamflow Prediction under Extreme Data-scarcity (SPED): A Step Toward Hydrologic Process Understanding within Severely Data-limited Regions. *Hydrological Sciences Journal*.

Bergström, S., 1992. The HBV Model: Its Structure and Applications. Swedish Meteorological and Hydrological Institute, Norrköping.

Drogue, G.P., and Plasse, J., 2014. How can a few streamflow measurements help to predict daily hydrographs at almost ungauged sites? *Hydrological Sciences Journal*, 59(12), pp.2126-2142.

Edijatno, Nascimento, N.O., *et al.*, 1999. GR3J: a daily watershed model with three free parameters. *Hydrological Sciences Journal*, 44(2), pp.263-277.

Hannah, D.M., *et al.*, 2011. Large-scale river flow archives: importance, current status and future needs. *Hydrological Processes*, 25(7), pp.1191-1200.

Hrachowitz, M., *et al*., 2013. A decade of Predictions in Ungauged Basins (PUB)—a review. *Hydrological Sciences Journal*, 58 (6), 1198-1255.

Juston, J., Seibert, J., and Johansson, P.O., 2009. Temporal sampling strategies and uncertainty in calibrating a conceptual hydrological model for a small boreal catchment. *Hydrological Processes*, 23(21), pp.3093-3109.

Konz, M., and Seibert, J., 2010. On the value of glacier mass balances for hydrological model calibration. *Journal of hydrology*, 385(1-4), pp.238-246.

Lanfear, K.J., and Hirsch, R.M., 1999. USGS Study reveals a decline in long-record streamgages. *Eos, Transactions American Geophysical Union*, 80(50), pp.605-607.

Lindström, G., *et al.*, 1997. Development and test of the distributed HBV-96 hydrological model. *Journal of Hydrology*, 201(1-4), pp.272-288.

Moore, R.J., 1985. The probability-distributed principle and runoff production at point and basin scales. *Hydrological Sciences Journal*, 30 (2), 273-297.

Moore, R.J., 1999. Real-time flood forecasting systems: Perspectives and prospects. In Floods and landslides: Integrated risk assessment (pp. 147-189). Springer Berlin Heidelberg.

Perrin, C., Michel, C. and Andréassian, V., 2003. Improvement of a parsimonious model for streamflow simulation. *Journal of hydrology*, 279(1-4), pp.275-289.

Perrin, C., *et al.*, 2007. Impact of limited streamflow data on the efficiency and the parameters of rainfall—runoff models. *Hydrological sciences journal*, 52(1), pp.131-151.

Perrin, C., *et al.*, 2008. Discrete parameterization of hydrological models: Evaluating the use of parameter sets libraries over 900 catchments. *Water Resources Research*, 44(8).

Rojas-Serna, C., *et al*., 2006. Ungauged catchments: how to make the most of a few streamflow measurements? *IAHS publication*, 307, p.230.

Seibert, J., and Beven, K.J., 2009. Gauging the ungauged basin: how many discharge measurements are needed? *Hydrology and Earth System Sciences*, 13(6), pp.883-892.

Seibert, J., and McDonnell, J.J., 2013. Gauging the ungauged basin: relative value of soft and hard data. *Journal of Hydrologic Engineering*, 20(1), p.A4014004.

Shiklomanov, A.I., Lammers, R.B., and Vörösmarty, C.J., 2002. Widespread decline in hydrological monitoring threatens pan-Arctic research. *Eos*, *Transactions American Geophysical Union*, 83(2), pp.13-17.

Viviroli, D., and Seibert, J., 2015. Can a regionalized model parameterisation be improved with a limited number of runoff measurements? *Journal of Hydrology*, 529, pp.49-61.

Vörösmarty, C., *et al*., 2001. Global water data: A newly endangered species. *Eos, Transactions American Geophysical Union*, 82(5), pp.54-58.

Wang, D., Chen, Y., and Cai, X., 2009. State and parameter estimation of hydrologic models using the constrained ensemble Kalman filter. *Water Resources Research*, 45 (11).

CHAPTER 6: CONCLUSIONS

The multi-objective SPED framework was proposed for streamflow prediction in ungauged catchments located within data-scarce regions of the world. SPED relied on *a priori* parameter estimates derived from highly uncertain and low resolution regional and global data to from its multi-criteria objective function. These data were treated as soft data, and fuzzy theory was utilized to partially account for their associated uncertainties. The SPED objective function aimed at maximizing runoff efficiency while simultaneously minimizing the difference between *a priori* parameter estimates and calibrated values for the same parameters. Tradeoffs were made by assigning triangular fuzzy weights to the criteria. SPED proved capable of predicting streamflow with high efficiency in two catchments located in a truly poorly gauged region of southwestern China (NSE = 0.72 and 0.74). In comparison, a single-objective and a constrained single-objective model performed comparably to SPED in terms of runoff efficiency. SPED, however, estimated the value of influential model parameters more closely to *a priori* estimates. The ability of SPED to perform comparably to models that have been designed solely to maximize runoff efficiency and do not significantly consider true representation of underlying phenomena contributing to runoff generation, was important in the sense that SPED provided the opportunity to pay substantial attention to process understanding without sacrificing the ability of the model to predict streamflow with high efficiency. This is particularly important within data-scarce regions where any attempt toward process understanding is highly hindered by lack of sufficiently high quality data.

SPED was further tested by application to four well-gauged catchments located on three different continents with diverse hydro-climatic conditions. The catchments were selected from previous flow prediction studies so that comparison was enabled between the SPED performance and that of sophisticated flow prediction approaches. Since minor modifications were made to the SPED procedure, the two previously studied catchments in China were subject to test again. As a reference, a traditional single-objective model was applied to all catchments as well. The well-gauged regions/catchments were transformed into synthetically poorly-gauged regions by using only regional and global data for hydro-climatic variables and soil properties, and by using data from only one gauged reference catchment in each region. Previous flow prediction studies, on the other hand, used the available robust data networks in each region and more than one gauged reference catchment. SPED ability in identifying and differentiating between equifinal models assisted it with handling partial dissimilarity between reference and target catchments. Thus, in North Fork Cache Creek in California, where the catchment was dissimilar to its reference catchment in terms of baseflow regime, SPED performed well (NSE = 0.54) while the traditional single-objective model completely failed (NSE = 0.08). SPED also proved robust and consistent in performance across the different hydro-climatic and physiographic settings of test (NSE range of 0.54-0.74). In comparison with flow prediction studies based on robust data networks, SPED performance was comparable or even exceeded that achieved previously (NSE range of 0.54-0.74 for SPED versus 0.22-0.66). SPED, thus, makes an important contribution to the science of flow prediction within data-scarce regions by addressing flow prediction pitfalls such as equifinality, catchment dissimilarity and difficulty of utilizing highly uncertain data.

SPED applicability on a regional scale for an application such as analysis of flow regime alterations due to diversion hydropower was tested against a simpler catchment similarity approach. Both SPED and catchment similarity model were applied to 32 small catchments developed with diversion hydropower projects in southwestern China. The results indicated that magnitude and variability of flow were highly altered, and regulated hydrographs were maintained at a static minimum residual flow for long periods of time. For instance, mean annual flows decreased by a mean of 76-86% across the 32 rivers and flow became more predictable in most rivers (47-94% mean increase in predictability). Frequency and duration of high flows decreased and duration of low flow events increased substantially. Slopes of rising hydrograph limbs and recession limbs increased respectively by a mean of 123-161% and 254-720%. The choice of a flow prediction method between SPED and catchment similarity did not alter this conclusion. However, analysis results based on SPED predicted flow data constantly indicated more severe effects on the natural flow regime of the rivers due to diversion hydropower. Purpose and application of the results of the analysis would therefore justify the choice of a flow prediction method: more simplistic catchment similarity model for applications such as detecting the direction and pattern of change, and more sophisticated SPED framework for applications such as more sensitive tasks such as design of instream flows.

Finally, the value added by limited streamflow observations collected through a field campaign to improvement of the accuracy of long term flow predictions in an ungauged catchment was studied in the context of a real-world case scenario. To this end, the case scenario was defined for a manger of an ungauged catchment located within a data-scarce region who needs to make a decision on whether or not to go ahead with a limited field data collection campaign in the ungauged catchment. Since the catchment is subject to a development project, all analysis needs to be completed shortly. The manager can rely on the traditional approach of calibration solely based on data in a gauged reference catchment or calibration based on a combination of data from the gauged reference catchment and limited data collected in the ungauged target catchment. We defined a two-criterion objective function for the latter and studied its value over the former. To simulate real-world circumstances, we assumed that there was up to one year of time available to collect data in the field before a decision had to be made. We defined 528 different scenarios based on a combination of 11 scenarios for timing of data collection days, 8 scenarios for the number of data collection days, and 6 scenarios for the weights of the two criteria. Ten catchments located on four different continents were subject to this test, including six catchments in the US. We found that there were two near universal scenarios that almost always (except in one catchment) resulted in improvements in flow prediction accuracy over the traditional approach.

Overall, our proposed methods and findings enable important improvements to flow prediction accuracy and process understanding in ungauged catchments located within data-scarce regions of the world.