4-2016

HYDROLOGIC MODELING AT UNGAUGED LOCATIONS IN SUPPORT OF THE DEVELOPMENT OF A VULNERABILITY RANKING PROTOCOL SYSTEM FOR ROAD-STREAM CROSSING INFRASTRUCTURE

Gordon Clark

HYDROLOGIC MODELING AT UNGAUGED LOCATIONS IN SUPPORT OF THE DEVELOPMENT OF A VULNERABILITY RANKING PROTOCOL SYSTEM FOR ROAD-STREAM CROSSING INFRASTRUCTURE

A Masters Project Presented

by

Gordon E. Clark

Submitted to the Department of Civil and Environmental Engineering

University of Massachusetts- Amherst

In partial fulfillment of the requirements for the degree of

Master of Science in Civil Engineering

April 2016

Hydrologic Modeling at ungauged locations in support of the development of a vulnerability ranking protocol system for road-stream crossing infrastructure

A Masters Project Presented

by

Gordon E. Clark

Approved as to style and content by:

Richard N. Palmer

Richard N Palmer, Chairperson

David Ahlfeld

David Ahlfeld, Member

Paula RS Rees

Paula Rees, Member

Scott Jackson

Scott Jackson, Member

Sanjay Arwade
Civil and Environmental Engineering Department

Table of Contents

ACKNOWLEDGEMENTS

My thesis work was funded by the Massachusetts Department of Transportation and I would like to thank them for their support in this research critical to sustaining healthy aquatic systems and safe road-stream crossings for future generations. I am proud to be part of an effort which I feel so passionately about.

I would also like to express my gratitude for my advisor, Dr. Richard Palmer, who has been incredibly supportive, kind, knowledgeable, and patient throughout my research process. These are the kinds of qualities that are most important in an adviser and are also tough to find. I am very lucky to have been given the opportunity to work with him. I would also like to thank all of my committee members for their thoughtful questions and suggestions with this research and with my writing. In addition, I would like to acknowledge Kuk-Hyun Keith Ahn for his help in guiding me through the regionalization and statistical components of this research. I am very grateful.

We exist because of others before us. As a result, I am extremely grateful for the support that I had from family and friends, close and far, throughout this journey and I will never succeed in any of my endeavors without them. In particular, I would like to thank my Mother and Father for their consistent guidance and my partner Liz for her love and patience.

Chapter One

1.1 Abstract

Recent tropical storms that have resulted in flood events with large economic impacts in the Northeastern US have catalyzed efforts to understand the complex interactions between human and natural systems. Specifically, the resilience of our transportation infrastructure to climate and the impact of our transportation systems on the aquatic environment are of significant interest on both the local and regional scales to state and federal agencies. It is important that new, innovative approaches be developed that consider both the robustness of our infrastructure today and its ability to cope with forecasted extremes due to climate change.

Because few streams are gauged to record their flows, road-stream crossings are almost always designed without adequate knowledge of what floods flows that will occur in the future. Hydrologic predictions at these ungauged locations are difficult and few techniques exist that can accurately estimate extreme flows without long-term precipitation and streamflows records. Robust analysis using either statistical or physically based models within a multi-model framework is a useful approach in quantifying the degree of uncertainty and variability across models at these ungauged, road-stream crossing locations.

This paper reviews background information associated with the development of a hydrologic vulnerability protocol system for road-stream crossings currently being piloted in the Deerfield River basin, focused primarily on predicting flows in ungauged basins and estimating flood flows for current and future climate scenarios. Finally, a framework for a decision support tool is discussed within the context of providing hydrologic flood predictions at road-stream crossings within an online interactive map interface.

1.2 Problem Description

Recent research has explored the ecological impacts of road-stream crossing infrastructure (MA DOT, 2010; Bates *et al.*, 2003; Jackson, 2003; Jackson *et al.*, 2011; MA DER, 2012; Januchowski-Hartley *et al.*, 2014; Januchowski-Hartley *et al.*, 2013; Zarriello and Barbaro, 2014; Pépino *et al,*. 2012; Andersson *et al.*, 2000; Nagrodski *et al.*, 2012). Specifically, understanding the spatial extent and magnitude of barriers to fish passage poses significant challenges to engineers, biologists, and decision-makers. Parallel to this ecological motivation also exists a regional, national, and global effort to better understand the vulnerability of our transportation network and road-stream crossing infrastructure in the face of climate uncertainty (CT DOT, 2013; VTRANS, 2012; U.S. FHWA, 2012; Furniss et al, 1998; Chang et al, 2010; Kalantari et al, 2014; Rosenberg et al, 2010). This includes both estimating the appropriate hydrologic data at road-stream crossings and the hydraulic modeling necessary for accurately representing the system.

The most challenging aspect of estimating hydrologic and hydraulic vulnerability at road-stream crossings is the systemic lack of data. More directly, there are very few (if any) long-term continuous records at these locations. Streamflow data are fundamentally important because they provide the best estimate of the true natural system. Currently, the problem of insufficient data are mostly due to the cost of building, operating, and maintaining streamflow gauges that are able to provide reliable and accurate measurements. It is not practical (or necessary) to have stream gauges at every location on a river. However, the importance of collecting streamflow data at systematic locations in a region cannot be overstated. These data allow an interpretation of the natural hydrologic response of a watershed to perturbances, such as intense precipitation events, as well as provide an ability to determine how other hydrologically similar systems might respond to these perturbances.

The estimation of hydrologic extremes (e.g. floods) at locations with no streamflow gauge data has great uncertainty. These uncertainties are categorized into aleatoric and epistemic uncertainties (Bevin, 2013). Aleatoric uncertainties are uncertainties that are derivative of the irreducible complexity of natural systems.

These uncertainties account for the apparent randomness of nature that are assumed or expected to be irreducible. Epistemic uncertainties are those that result in uncertainties associated with knowledge. These are the uncertainties that could be better understood under a classic reductionism approach and include for example: hydrologic model structural error or even a simple lack of data (precipitation, streamflow, climate, etc.). Of these epistemic uncertainties, hydrologic model choice can play a significant role in the uncertainty associated with estimating the high-flow events necessary for the engineering and design of road-stream crossing infrastructure.

This research provides estimations for the 2-, 5-, 10-, 25-, 50-, and 100-year recurrence intervals at road-stream crossings (ungauged) in the Northeastern US using statistical and process-based hydrologic modeling approaches for both current and future climate scenarios. Impacts of hydrologic model choice within a road-stream crossing vulnerability assessment framework are identified. These efforts support a multi-disciplinary, systems-based model for assessing the vulnerability of road-stream crossings across the Northeastern US region that include hydrologic/hydraulic vulnerability as well as ecologic, geomorphic, structural, and transportation network disruption vulnerability metrics. Taken together, these different disciplines provide a more holistic perspective on the risk and vulnerability of our road-stream crossings, meeting the needs for both human-use and ecological integrity. The multiple-model framework on the hydrologic side provides insight into the level of uncertainty associated with the many models that are available for making these predictions at ungauged locations.

Decision support tools have become increasingly common in the region for several objectives including prioritization of stream-barrier removal (Neeson *et al.*, 2015; McGarigal *et al.*, 2011), assessing ecological impacts of road-stream crossings (TU, 2015; Martin and Apse, 2011), predicting road culvert passability for migratory fishes (Januchowski-Hartley *et al.*, 2014), and assessing aquatic connectivity through network analysis methods (McKay et al, 2013). Significant interest exists globally to develop decision support tools towards these ends (Lawrence et al, 2014 ; O'Hanley, 2011; Gauthier et al, n.d; Kemp et al, 2010). This research integrates hydrologic flood forecasts at ungauged locations, including at road-stream crossings, within an interactive web-based decision support tool to provide a vehicle towards actionable science for both infrastructure design and ecological connectivity.

1.3 Background

1.3.1 Implications of Interacting Linear Networks

Humans profoundly transform river landscapes by altering watersheds, climate, and channels, which in turn modify the hydrologic, biotic, and sediment fluxes through river systems (James and Marcus, 2006; Blanton and Marcus. 2009). Transportation and fluvial systems both function as linear systems that are strongly dependent on continuity. A discontinuity for either system poses significant challenges in their ability to convey people and materials, in the case of transportation, and environmental services, in the case of rivers. The ubiquity of these linear networks across the landscape results in frequent interception and interaction between these systems. Ultimately, water flow is a "master variable" (Power et al, 1995) that is the primary influence on the fundamental nature of streams and rivers (Poff *et al.* 1997; Hart and Finelli 1999) and modification of flow regime from human activity intuitively suggests the consequential alteration of the structure and function of fluvial ecosystems (Petts, 1984; Chisolm, 1994; Yeager 1994; Ligon *et al.*, 1995; Ward and Stanford, 1995; Stanford *et al.*, 1996; Poff *et al.*, 1997; Bednarek, 2008). There are few studies that provide insight to the ecological effect of hydrologic alteration due to discontinuities at the catchment scale in the region, especially with respect to studies that provide information on the long-term ecological responses to conservation and restoration efforts like barrier mitigation.

The biological impacts from fragmentation of streams and rivers are complex and interconnected. For example, the inability for riverine (or native, non-migratory) fish species to traverse the length of the full extent of the river subsequently affects the ability for native freshwater mussels to complete the reproductive life cycle. Freshwater mussel larvae, or glochidia, typically need to attach to an aquatic vertebrate (usually to fish gills) for a period of a few days to a few months where they then release and develop into juvenile mussels in a new location (Nadeau, 2008). Loss of fish species in areas upstream of impassable barriers will eliminate already threatened mussel assemblages and likewise, restoring these species by barrier mitigation efforts will likely help restore these populations (Smith, 1985). This is just one example of an biological and ecological implication of fragmented fluvial networks in the region. It is estimated that three-quarters of the 297 native mussel species in North America are imperiled and 35 species alone have thought to have gone extinct in the last century (Bogan, 1996).

The success of aquatic passage at these road stream crossings is ultimately a function of the geometry of the road-stream crossing infrastructure (e.g. culvert or bridge), which may cause physical or velocity barriers for migrating species. For example, a culvert generally constricts flow area at a road-stream crossing and as a result, velocities are often increased through the structure (Bodhaine, 2015). This velocity is often further impacted by the culvert substrate which is generally of lower roughness than the natural streambed. The grade of the culvert can also seriously impact road-stream crossings as a raised slope further increases velocity. Some culvert outlets are raised above the downstream water surface to an elevation that prevents aquatic organisms from passing. Either the outlet pool depth is not sufficiently deep for species to generate an approach velocity or the elevation differential between the culvert outlet and the outlet pool is too great for a fish to overcome (i.e. perched).

Effective culvert design for both adequate drainage and aquatic organism passage requires an understanding of the local hydrology, insight into the maximum and minimum design flows, and anticipated flows during critical time periods for aquatic passage. The ubiquity of road-stream crossings, coupled with a systemic lack of streamflow data, makes the accurate estimation of local hydrology across a broad region challenging. Since most of these road-stream crossings exist in absence of stream gauge data, the design of culverts and bridges on these stream must be informed by "best estimates" of flow at these ungauged locations.

1.3.2 Road-Stream Crossing Vulnerability and Barrier Prioritization

Much interest has been paid recently to understanding the ecological impacts of road-stream crossing infrastructure in the northeast region. From a broad perspective, habitat loss and fragmentation are leading drivers of declining biodiversity and ecosystem services across the world (Sala *et al.*, 2000; Tilman *et al.*, 2001; Fahrig, 2004) and implementation of both landscape corridors and barrier mitigation, including stream restoration, ecologically informed culvert design, and dam removal, are widely used as effective strategies for reducing fragmentation (Bednarek, 2001; Damschen *et al.*, 2006; Neeson *et al.*, 2015). The concept of "connectivity" may serve to represent and encapsulate the ultimate goal in the assessment road-stream crossing vulnerability, barrier prioritization, and aquatic connectivity research. Connectivity is defined here in its most broad sense including the spheres of ecological, hydrological, geomorphological sciences and the feedback between these disciplines and the local, regional, and global human populations. Unlike the field of ecology, little consensus exists for a standard definition of connectivity with respect to hydrological and geomorphological systems. Connectivity is defined here as having three characteristics (Bracken and Croke, 2007): (1) landscape connectivity; (2) hydrological connectivity; and(3) sedimentological connectivity. These three types of connectivity, in addition to an established notion of ecological connectivity, serve to represent ideals for the fluvial and transportation networks for road-stream crossing vulnerability and barrier prioritization research.

The Federal Highway Administration (FHWA) created guidelines in the past decade to achieve greater interagency cooperative conservation with respect to infrastructure planning, design, review and construction that emphasizes approaches that are more sensitive to wildlife and their ecosystems (FHWA, 2006). Motivations for creating these guidelines include the principles stated in the National Environmental Policy Act (NEPA) and Executive Order (EO13352) on Facilitation of Cooperative Conservation as well as an Executive Order (EO13274) for Environmental Stewardship and Transportation Infrastructure Project Reviews. These guidelines set a national precedent in the U.S. within the transportation sectors to promote principles of conservation and connectivity in the context of infrastructure design.

At the state level, the Massachusetts Department of Transportation (MassDOT) has prepared its own guidelines to underscore the importance of designing new and replacement bridges and culverts to accommodate fish and other wildlife passage at road-stream crossings (MassDOT, 2010). In Massachusetts, state and federal regulations of stream crossings apply requirements based on the Massachusetts River and Stream Crossing Standards developed by the 'Massachusetts River and Stream Continuity Partnership' in 2006 (revised again in 2011). This partnership includes the University of Massachusetts Amherst, The Nature Conservancy, the Massachusetts Division of Ecological Restoration (DER) Riverways Program, and American Rivers. The stream crossing design standards developed through this partnership establish the paradigm to which all road-stream crossing infrastructure should be built to accommodate ecological, hydrologic, and geomorphic connectivity. The diverse technical committee that supported the development of these standards included the US Fish and Wildlife Service, USGS BRD, UE EPA, US Army Corps of Engineers, MA Division of Fisheries and Wildlife, Connecticut River Watershed Council, Connecticut DEP, and hydraulic engineering consultants. The standards adopted a "stream simulation" design approach in which, at a fundamental and conceptual level, flows through a road-stream crossing should be no less constrictive than the natural river channel. This implies that no impairment to movement for aquatic organisms should occur and thereby maintaining connectivity in a full sense of the word.

At a regional level, the North Atlantic Aquatic Connectivity Collaborative (NAACC) provides a participatory network of practitioners across thirteen states that serves to provide multiple levels of support for road-stream crossing field assessments including: unified protocols for road-stream crossing assessments; online field assessment training; database repository for field assessment data; a watershed level crossing assessment prioritization tool; and general support for conducting road-stream crossing assessments throughout the region (NAACC, 2014). Field assessments of road-stream crossings are currently an important step in the ecologic, hydraulic and geomorphological assessments. However there has been less research on estimating the passibility at road-stream crossings (Januchowski-Hartley *et al.*, 2014), as well as measuring the ecological integrity of road-stream crossings through remote sensing and GIS analysis alone (McGarigal *et al.*, 2011). Although research has addressed road-stream crossing vulnerability and barrier prioritization in the region (e.g. MA DOT, 2010; Bates *et al.*, 2003; Jackson, 2003; Jackson *et al.*, 2011; McGarigal *et al.*, 2011; MA DER, 2012; Januchowski-Hartley *et al.*, 2014; Januchowski-Hartley *et al.*, 2013; Zarriello and Barbaro, 2014; Pépino *et al.* 2012; Andersson *et al.*, 2000; Nagrodski *et al.*, 2012), understanding the spatial extent and magnitude of barriers to fish passage still poses significant challenges to engineers, biologists, and decision-makers.

1.3.3 Predicting Flows at Ungauged Locations

Road-stream crossings are numerous and as a result are almost always at locations that do not have a streamflow gauge. Therefore, these locations are considered "ungauged" from a hydrological perspective. Ungauged locations have inadequate records (in terms of both data quantity and quality) of hydrological observations to enable an accurate estimate of hydrological variables of interest (e.g. high flows) (Silvapalan *et al.*, 2003). Predictions in ungauged basins (PUB) have been a research topic for decades. More recently, the International Association of Hydrologic Sciences (IAHS) dedicated the decade from 2003 to 2013 to furthering the research of predictions in ungauged basins. Apart from a fundamental lack of data in making these predictions, one of the main challenges is understanding the uncertainty inherent to these predictions, whether it be from climatic inputs, land-cover, soils, vegetation, or even from the model structure used to inform the predictions/forecasts. This challenge is further exacerbated by a historic fragmentation of approaches and methods in making predictions in ungauged basins across the world. This has led to a "cacophony of noises" as opposed to a "harmonious melody" within the scientific hydrology community (Blöschl *et al.*, 2013). The authors believe that the gluttony of hydrologic models used in the field across the world is one symptom of this fragmentation phenomenon in the field of hydrology.

There are several methods that are used when predicting flows in ungauged basins. These methods can be broadly categorized into two different categories: statistical and process-based (Figure 1). Statistical methods use available runoff time series data from neighboring catchments (donor catchments) to estimate runoff hydrograph at ungauged locations based on one or more similarity measure. An advantage of statistically based runoff simulation methods is their simplicity of input, as they do not require variables like precipitation, evapotranspiration, or other climactic variables. In addition, these methods, once developed, are typically less data intensive than process-based methods to utilize. However, the process-based methods are often the preferred methods for hydrologists as they allow for more flexibility in the modeling process such as the ability to simulate different land-use or climate change scenarios, as well as provide a way to interpret the hydrologic landscape of a single catchment: an opportunity that many stochastic methods do not allow. A more detailed description of these two categories will be discussed in the following section.

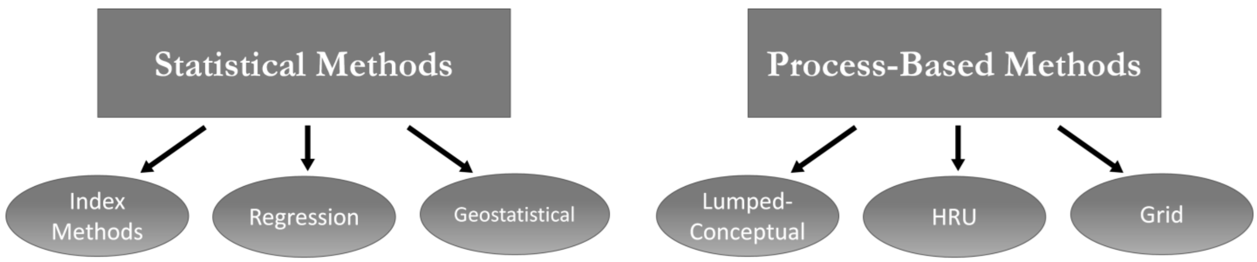


Figure 1: Different approaches for making hydrologic predictions in ungauged basins.

Implementation of any hydrologic model at an ungauged location can be particularly challenging. As previously mentioned, the largest challenge in predicting flows at ungauged locations is that there are no data to verify or calibrate a hydrologic model. Understanding this challenge and the uncertainties implied by this fact is essential when developing protocols for assessing the hydrologic/hydraulic vulnerability at a road-stream crossing. Due to the lack of streamflow gauge data at these road-stream crossing locations, stream discharge cannot be known with certainty. Instead hydrologic modeling tools applied intelligently can help inform what the stream discharge might be during a given period, but until long term continuous streamflow data are collected at these locations, it is impossible to know for certain.

1.4 Hydrologic Modeling at Ungauged Basins

1.4.1 Stochastic

Stochastic approaches to estimating streamflow include regression equations, index methods, and geostatistical methods. The use of regression equations to directly transfer the full hydrograph to an ungauged location is rather unusual (Blöschl *et al.*, 2013). More often, regression equations are used to estimate the flows for specific return intervals and not the full hydrograph. Geostatistical methods (Gandin, 1963; Martheron, 1963) to estimate the runoff hydrograph are relatively uncommon. Geostatistical methods exploit the spatial correlation of the variable of interest and provide an estimate of that variable as the weighted average of the measurements in the neighborhood (Blöschl, 2013). Skoien and Blöschl (2007) proposed 'spatio-temporal top-kriging' to estimate runoff time-series at all locations of a river network. This method avoids precipitation data errors and also avoids the parameter identifying issues associated with traditional process-based models (Blöschl , 2013).

The most commonly applied statistical approach to estimating flows are index methods. These methods are strongly reliant on an assumption of similarity between the ungauged (recipient) catchment and a gaged (donor) catchment. The simplest and most common index methods assume that the time series of runoff, once normalized by the mean flow, is identical between the donor catchment and the ungauged catchment. The drainage-area ratio method (Stedinger *et al.*, 1993) is the most widely used index method. The drainage-area (DA) ratio method assumes that the runoff at the donor and recipient ungauged catchments only differ because the sizes of the drainage areas at the respective catchments are different and that for a given time the runoff per unit area at the donor and recipient catchments are equal (Stedinger *et al.*, 1993).

1.4.2 Process-Based

Process-based methods of predicting runoff in ungauged basins is the alternative approach to statistically based models. These models, also known as physically-based or deterministic models, describe the spatial variability of hydrological processes through mathematically determining conservations of mass and momentum of water across a landscape. Process-based hydrologic models are an important evolutionary step in representing hydrological processes and spatially distributed data. Their complexity and application have increased since the first computer-based rainfall-runoff model (RRM) was developed in the 1960s (Crawford and Linsley, 1966). The need for better representation of physical processes in space and time is evident, especially considering the explosion of accessibility of digital products (e.g. elevation, soil, and vegetation) along with new technologies for measuring temporal and spatial variability in precipitation (Yu, 2003).

Process-based hydrologic models can be grouped into three categories that are representative of their respective discretization of space, or essentially how the models "interpret" the landscape (Figure 2). This characterization is model specific and involves transferring the continuous landscape into discrete counterparts that are used in ensemble to represent the spatial extent being modeled as a whole. Depending on the application, different RRM have different advantages.

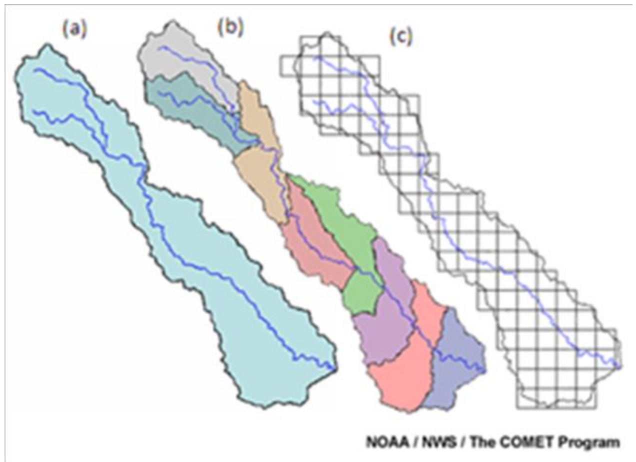


Figure 2: Representations of three types of physical model discretization: (a) lumped conceptual; (b) distributed hydrological response units (HRUs); (c) distributed grid-based. For the purposes of this report, we considered lumped conceptual and HRU-based discretization to be grouped as one under "physical HRU-based." [Borrowed from © the COMET Program]

Lumped conceptual models are generally applied at a single point or a region for the simulation of hydrological processes. These models are typically less complex than the distributed models. The parameterization of a lumped conceptual model is relatively simple. For example, input data (e.g. land use, soil type) used in these models are typically averaged or weighted across the extent of the landscape, effectively reducing landscape heterogeneity. The result of this discretization method is often described as an advantage in computational efficiency. One general disadvantage of lumped-conceptual watershed RRMs is that, in reducing the sub-basin scale variability, the model parameters that were designed to represent physical properties may become obscured during regression based regionalization.

Discretization using a grid-based method distributes the variability of landscape features uniformly based on the resolution (cell size) appropriate for the site being modeled and hence, these types of models are typically called 'distributed models.' Hydrologic response units (HRUs) lay between lumped-conceptual and grid-based discretization where smaller subbasins, or sometimes referred to as reaches, are appropriated across a catchment and are commonly referred to as having a 'semi-distributed' model structure. Landscape variability is distributed across these reaches (HRUs) which can vary in area, depending on the spatial extent of the site being modeled.

In applying a typical process-based hydrological model application (Figure 3), the first steps are to identify the watershed of interest as well as the streamflow gauges that have adequate and reliable historical records. It is almost always necessary to choose streamflow gauges that are not impacted by reservoir operations and are representative of natural flows in the system. The next step is to discretize the landscape, which will vary by model (Figure 2), and compute the necessary data needed for the model to represent the natural hydrologic system (e.g. river channel slope; land-use or soil types across the landscape; elevation data, etc.). Next, the input drivers to the model are identified, collected, and assimilated into the proper format. These data usually include at a minimum temperature and precipitation, however more complex models require more data such as potential evapotranspiration, cloud cover, and radiative energy (for example). Finally, the model is calibrated to the stream gauge at the basin outlet by adjusting model parameters. The model must be validated to ensure that it is appropriate and calibrated properly. It is important that there be clear calibration and validation periods and that they do not overlap. This step compares and statistically tests the fit of the calibrated model parameters and ensures that the physical processes represented adequately. If the model fails in its validation, it is then necessary to repeat the process to ensure input data are well-represented and/or assumptions are correct.

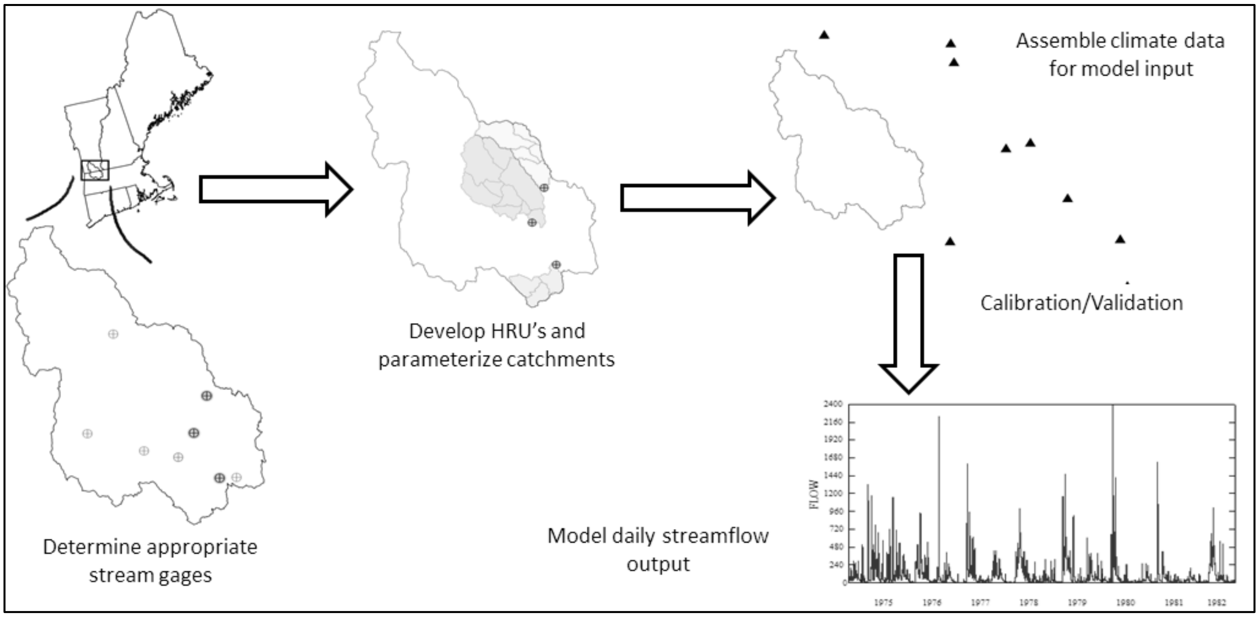


Figure 3: Conceptual flow diagram of a typical catchment based hydrological (rainfall-runoff) model application.

The extensive data input requirements of many of these process-based RRMs represent a significant challenge associated with use of these tools. Data collection, model development, parameter sensitivity analysis, and calibration/validation processes require significant resources of time and effort In addition, professional experience is required to properly implement these models. However, there are distinct advantages of process-based hydrologic modeling. One of the most important assets of these models is their ability to respond to perturbations introduced by the modeler when assessing how these impacts might impact the hydrology of the modeled system. For example, future climate change impacts can be interpreted directly using a well-calibrated RRM by modifying the meteorological drivers such as precipitation and temperature. Future drought conditions or flood conditions could be interpreted from scenarios with process-based models whereas statistical models are generally unable to easily incorporate such inputs. In addition, RRMs provide a more site-specific understanding of hydrologic response at the catchment level in comparison to statistical methods.

The primary challenge of process-based, rainfall-runoff modeling methods at an ungauged location is the lack of local runoff data that could be used for model selection and calibration (Blöschl, 2013). This is also called the regionalization problem in hydrological modeling. Regionalization can be defined as the process of transferring hydrological information (e.g. process-based model parameters) from one catchment to another (Blöschl and Sivapalan, 1995). Regionalization without runoff data can be a very difficult task and may be approached in several different ways (see Figure 4).

Catchment characteristics such as soil type, land-use type, stream hydraulic geometry, or topography (to only name a few) can be used to provide an estimate for what the model parameters where runoff data are unavailable. This "a-priori" approach is typically done without calibration of neighboring catchments and requires model specific knowledge as well empirical type relationships regarding how parameters are related to these catchment physical properties.

An alternative parameter estimation approach is the transfer of calibrated model parameters from a gauged catchment (or multiple gauged catchments) to the ungauged catchment. This is a commonly used technique and can be applied in several different ways. Spatial proximity, similarity, and model averaging methods are the most simple and straightforward methods that assume that the calibrated model parameters from hydrologically similar or adjacent catchments are also valid at an ungauged basin. Regression between calibrated model parameters and catchment characteristics is an alternative approach of parameter transfer.

Finally, parameters may also be constrained by runoff characteristics and/or dynamic proxy data. This approach involves the use of dynamic data in the ungauged catchment such as soil moisture or regionalized runoff to reduce uncertainty in the model parameters. For example, a short runoff record in an ungauged catchment, if it is available, may be leveraged to provide information regarding rainfall-runoff model parameters (Seibert and Beven, 2009). These approaches are not mutually exclusive and can be used in tandum to provide the most suitable estimate of model parameters at an ungauged site.

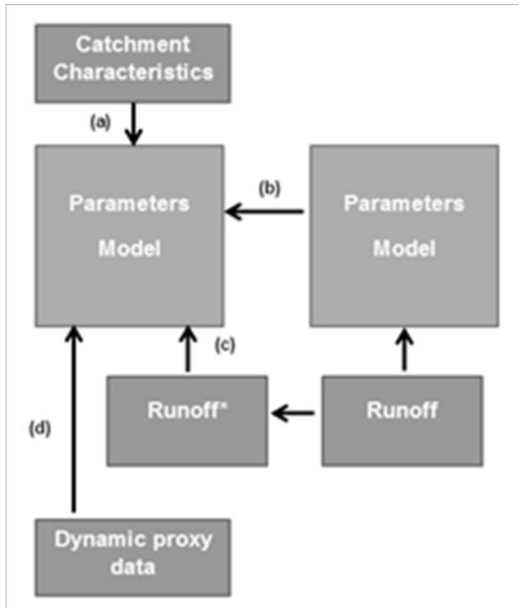


Figure 4: Conceptual schematic representation for estimating parameters in process-based models for ungauged basins. (a) A-priori estimation of model parameters from catchment characteristics; (b) transfer of calibrated model parameters from gaged catchments; (c) constraining model parameters by regionalized runoff characteristics; (d) constraining model parameters by dynamic proxy data. [Figure 10.20 and caption from Blöschl *et al.*, 2013. Runoff Prediction in Ungauged Basins, pg. 247]

1.4.3 Evaluating Model Performance

There are many ways to evaluate hydrologic model performance and there is much research on the subject (see Krause *et al.*, 2005). A brief accounting of some of the commonly used goodness-of-fit (GOF) performance criteria are listed in Table 1 and are described below.

The coefficient of determination (R^2) is used to describe and measure the amount of variance explained by the model. This value ranges from 0 to 1, with unity representing a model of perfect fit. The coefficient of determination is equal to the ratio of explained variation to the total variation and can be represented as the square of the correlation coefficient r. The R^2 value is one of the most common goodness-of-fit measures for hydrologic models.

The Nash-Sutcliffe Efficiency (NSE) value was proposed by Nash and Sutcliffe in 1970 as a modification to the mean-square-error (MSE) goodness-of-fit value. This is one of the most popular transformations of the MSE (Singh, 2014). The NSE can be interpreted as a classic skill score where skill is the comparative ability of a model with regard to a baseline model. The NSE value ranges from negative infinity to 1. If the NSE value is 0, then the model is no better than using the observed mean as predictor. If the MSE is zero, then NSE is unity indicating the model is a perfect fit.

The NSE value has been criticized for its inability to infer a sampling distribution (McCuen *et al.*, 2006) as well as the inadequacy of the metric to fully capture a model's performance (Jain and Sudheer, 2008). In addition, it ignores the degrees of freedom in the data, does not apply an exact probability function, is prone to subjective interpretations, has no lower bound, and is sensitive to outliers (personal communication with Richard McCuen, 2015). However, the NSE can be applied to a wide range of model types and is commonly used in the literature..

The KGE (Gupta *et al.,* 2009) is a criterion that decomposes the NSE (and MSE) value. It has been used in hydrologic modeling as an objective function that serves to mitigate some (but not all) of the shortcomings of the NSE value. The range of this value is between negative infinity to 1. The closer the model is to one, the more accurate the model is.

Several other measures are also used to help quantify overall model goodness-of-fit. Volumetric efficiency is proposed to circumvent some of the problems associated to the NSE value. It ranges from 0 to 1 and represents the fraction of water delivered at the proper time (Criss and Winston, 2008). The index of agreement is a standardized measure of the degree of model prediction error developed by Willmott (1981) and ranges between 0 and 1 with a value of 1 indicating a perfect match and 0 indicating no agreement at all. Its benefits include the ability to detect additive and proportional differences in the observed and simulated means and variances; however, similar to other measures, it has been demonstrated to be overly sensitive to extreme values due to the squared differences term (Legates and McCabe, 1999). The percent bias measure measures the average trend of simulated values to be greater or smaller than observed values with an optimal value of 0. Negative values indicate model underestimation. The result is often reported in a percentage.

Table 1: Commonly used goodness-of-fit criteria used in hydrological modeling.

	Name	Abrv.	Equation	Range
(1)	Root Mean Square Error	RMSE	$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - x_i)^2}$	0 to inf
(2)	Normalized Root Mean Square Error	NRMSE	$NRMSE = 100\ \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - x_i)^2}}{S_x}$	0 to inf
(3)	Percent Bias	PBIAS	$PBIAS = 100\frac{\sum_{i=1}^{N}(y_i - x_i)}{\sum_{i=1}^{N} x_i}$	0 to inf
(4)	Coefficient of Determination	R2	$\mathrm{R}^2 = \left[\frac{\frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}(\mathrm{x_i} - \bar{\mathrm{x}})(\mathrm{y_i} - \bar{\mathrm{y}})}{\mathrm{S_x S_y}}\right]^2$	0 to 1
(5)	Nash-Sutcliffe Efficiency Value	NSE	$\mathrm{NSE} = 1 - \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}}(\mathrm{x_i} - \mathrm{y_i})^2}{\sum_{\mathrm{i}=1}^{\mathrm{N}}(\mathrm{x_i} - \bar{\mathrm{x}})^2} = 1 - \frac{\mathrm{MSE}}{\mathrm{S_x}^2}$	-inf to 1
(6)	Kling Gupta Efficiency Value	KGE	$\mathrm{KGE} = 1 - \mathrm{ED}$ $\mathrm{ED} = \sqrt{(\mathrm{r}-1)^2 + (\mathrm{a}-1)^2 + (\mathrm{b}-1)^2}$ $\mathrm{a} = \frac{\mathrm{S_y}}{\mathrm{S_x}};\ \mathrm{b} = \frac{\bar{\mathrm{x}}}{\bar{\mathrm{y}}}$	-inf to 1
(7)	Volumetric Efficiency	VE	$\mathrm{VE} = 1 - \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}}\lvert\mathrm{y_i} - \mathrm{x_i}\rvert}{\sum_{\mathrm{i}=1}^{\mathrm{N}}\mathrm{x_i}}$	0 to 1
(8)	Index of Agreement	d	$d = 1 - \frac{\sum_{i=1}^{N}(x_i - y_i)^2}{\sum_{i=1}^{N}(\lvert y_i - \bar{x}\rvert + \lvert x_i - \bar{x}\rvert)^2}$	0 to 1

Notes: x_i is a set of observations; y_i is a set of predictions; $\bar{x}$ is the arithmetic mean of observed data, $\bar{y}$ is the arithmetic mean of the predicated data; Sx and Sy represent the standard deviations for the observed and predicted data, respectively; MSE represents the mean-square-error; ED represents the Euclidean distance from the ideal point in the scaled space; r is the correlation coefficient; a is a measure of relative variability of the predicted and observed values; and b is the bias defined as the ratio of the mean and predicted flows to the mean of the observed flows.

1.5 Estimating Flood Flows

1.5.1 Regression Equations

Developing empirical relationships between hydrologic variables of interest and catchment characteristics are used quite frequently to provide estimates in data sparse situations. These regression equations are typically developed and tested using many streamflow gauges over a region of interest. The greater the number of gauges used to develop relationships between catchment characteristics and hydrologic variables, the better. This approach harnesses the historical streamflow data over long periods of record to draw conclusions about a catchment's response. These analyses are then used to extrapolate what an ungauged catchment response is without any streamflow record at a site of interest.

For floods, the regression approach assumes that there is a relationship between a flood peak runoff of a given return period and catchment/climate characteristics (Thomas and Benson, 1970). In the U.S., peak flow regression equations have been developed on a state-by-state basis. In Massachusetts, the analyses reported in the USGS Water-Supply Technical Paper 2214 written by Wandle (1983) serve to provide the regional peak flow regression equations (RPFE) across the different regions in the Commonwealth. While applying these regression equations is straightforward, it should be noted that these flood estimations have not been updated by the state since they were published (to-date) and may be skewed by the stationary assumption (Milly *et al.*, 2007). The regression equations reported by Wandle (1983) are over 30 years old.

1.5.2 Index Scaling Approach

Index methods apply the principle of hydrologic similarity, including the assumption of temporal similarity. More precisely, the timing of flows in an adjacent or donor catchment is similar to the timing response of flows in an ungauged catchment. The most powerful assumption for this method is that a time-series of runoff, once normalized by the mean flow, is identical between the donor catchment and the ungauged catchment (Blöschl , 2013). For example, the drainage-area ratio method (Stedinger *et al.*, 1993) assumes that the runoff between donor and ungauged catchment only differ because of their differing drainage-areas. The assumption with this very simple and commonly applied model is that the runoff per unit area between donor and ungauged catchments are equal.

1.5.2 Using the Daily Runoff Hydrograph

When estimating flows from a continuous record of daily discharge, the first step is to determine the maximum average daily discharge for each year across the period of record in what is called an annual maxima series (AMS). The AMS is then used to fit an appropriate distribution for these flows to extrapolate flood flows from the continuous daily record.

There are two families of continuous probability distributions that are most commonly suggested as the initial choices for flood flow estimation: the Generalized Extreme Value (GEV) distribution and the log Pearson type 3 (LP3) distribution. Although the LP3 distribution was recommended by the U.S. Water Resource Council, recent research suggests that the GEV distribution is often preferred (Vogel, 1993), especially for the northeastern US region (Villarini and Smith, 2010; Vogel and Wilson, 1996). The GEV has a cumulative distribution function,

$$F(x;\ \mu;\ \sigma;\ \xi) = \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}$$

where μ, σ, and ξ are the location, scale, and shape parameters, respectively.

The parameters of the GEV distribution are often estimated using the method of L-Moments which are analogues of traditional moments. They were developed to provide estimators that were less sensitive to outliers and were therefore considered more robust (Hosking, 1990). L-Moments are used to calculate quantiles that are similar to standard deviation, skewness, and kurtosis. Although probability weighted moments can be used to estimate the distribution parameters, an advantage of the L-Moment approach is that they are easier to interpret, can calculate more accurate parameters for smaller sample sizes, and are nearly unbiased (Kochanek, 2010; Rowinski, 2001, Millington *et al.*, 2011). For a more detailed discussion of the application of L-moments for flood frequency analysis, see Kuczera and Franks (2011).

A component of flood frequency analysis is the probability plot. These plots present the annual exceedance probability (AEP) (defined as the inverse of the return storm year), and the discharge. These plots provide the opportunity to visually evaluate the adequacy of the fitted distribution as an empirical probability distribution. The AEP for each observed peak discharge on record is often referred to as the plotting position. The Cunnane plotting position, whose general form (Blom, 1958) can be represented by,

$$P_{(i)} = \frac{i - \alpha}{n + 1 - 2\alpha}$$

where i is the rank of the high flow in the annual maxima series (AMS), n is the number of years in the AMS, and α is a constant whose value is 0.4, is used to estimate unbiased quantities that are used to plot with the fitted distribution (Cunnane, 1978; Kuczera and Franks, 2006; Stedinger *et al.*, 1993). A more complete discussion of plotting positions can be found in Stedinger *et al.* (1993). In addition, statistical bootstrapping can provide the confidence intervals for the flood flow estimations providing more information regarding the uncertainty of the empirical distribution.

It is important to consider the advantages and disadvantages of flood frequency analysis. For one, flood peak flows are the results of complex interactions of many different components associated with a rainfall event, antecedent conditions, and rainfall-runoff transformation. Because peak flood records represent the integrated response of a storm event with the catchment in which the precipitation falls, they are able to provide a direct measure of flood exceedance probabilities (Franks and Kuczera, 2006). As a result, this approach is less susceptible to bias that can affect alternative methods such as design rainfall approaches (Kuczera *et al.*, 2003). This comes with the disadvantages of: the true probability distribution family for floods is unknown; short records may produce estimates with significant uncertainty; an inability to account for the physical processes that develop from land-use or climate changes in a catchment; and the issue of streamflow gauges often not being able to reliably capture the high flow events (Franks and Kuczera, 2006).

1.6 Pilot Study in the Deerfield River Basin

In May of 2014, the University of Massachusetts (UMass) Amherst, proposed to the Massachusetts Department of Transportation to develop risk-based and data driven protocols for assessing the present and future extreme flood vulnerability of road-stream crossing infrastructure in the Massachusetts's portion of the Deerfield River basin. This project was to incorporate multiple dimensions of vulnerability for the road and stream network including present and future flood hydrologic conditions, geomorphic stability, ecological system accommodation, structural flood resilience, and transportation/emergency response service disruption impact (Figure 5). The goals of the MassDOT Deerfield River Pilot (herein referred to as DRP) were to develop an innovative, systems-based approach to improve the assessment, prioritization, planning, protection, and maintenance of roads and road-stream crossings that are: (1) proactive with respect to upgrading structures to account for climate change; (2) complimentary of existing MassDOT project development and bridge design processes; and (3) provides a decision-support tool (DST) that can be used during project planning and development phases.

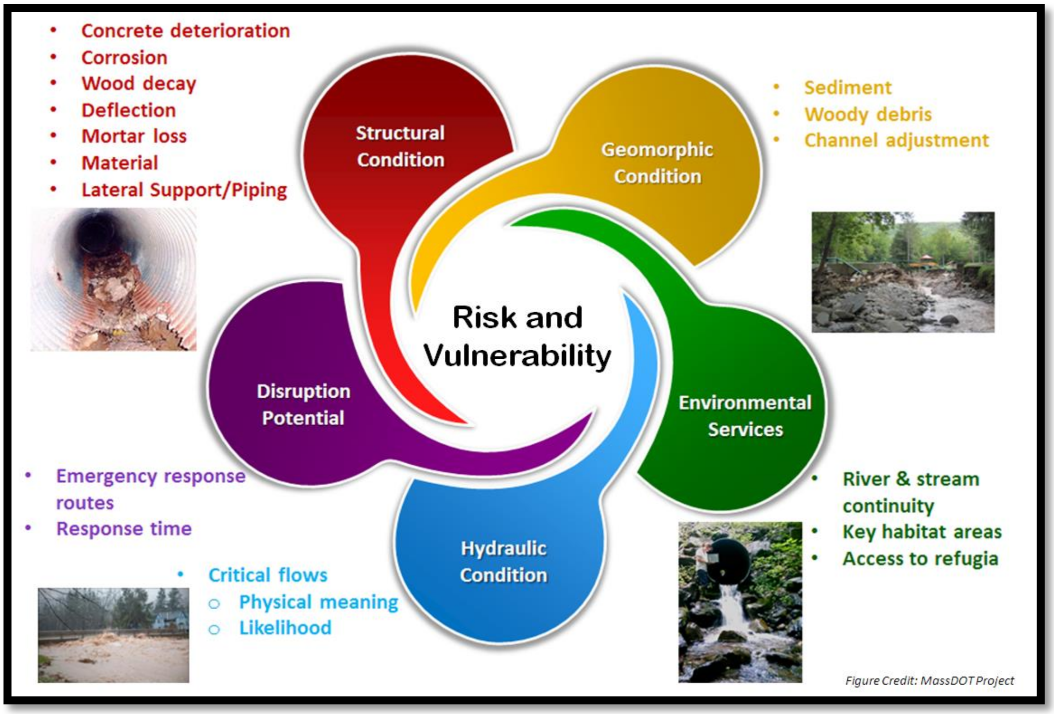


Figure 5: Holistic approach to vulnerability in the MassDOT Deerfield River Pilot (DRP) Project.

The Deerfield River basin (Figure 6) straddles north-western Massachusetts and southern Vermont with a drainage-area of approximately 1722 km^2. It is a major sub-basin of the Connecticut River. The largest tributary in the Deerfield basin is the North River with a total drainage-area of approximately 240 km^2. There is extensive hydroelectric-power generation (ten major dams) in the basin and the flows on this river are considered to be heavily altered by these activities (Friesz, 1996). The basin is mostly undeveloped with only about 5.3% of the total area classified as developed and about 82% as forest according to the 2011 National Land Cover Dataset (NLCD). There are approximately 1.48 km of stream length for every square kilometer of land in the basin calculated using the USGS National Hydrography Dataset (NHD+) dataset. There are three USGS streamgages operated in the basin that are not impacted by the reservoir and dam operations. The catchments of these gages represent about 23% of the total drainage-area of the Deerfield River basin (Table 2). Elevations in the Deerfield Basin range from about 35 meters above sea level in the Connecticut Valley Lowlands to about 1,202 meters in the ridges of the Berkshire Hills with a mean altitude of about 475 meters. Average annual precipitation in the basin is 107-112 cm in the low altitudes to 127-188 cm in the higher altitudes (PRISM, 2004; Knox and Nordenson, 1955). Snowmelt in spring and evapotranspiration in summer and fall cause annual cyclical trends in mean monthly runoff, even though mean monthly precipitation is evenly distributed throughout the year (Gay *et al.*, 1974).

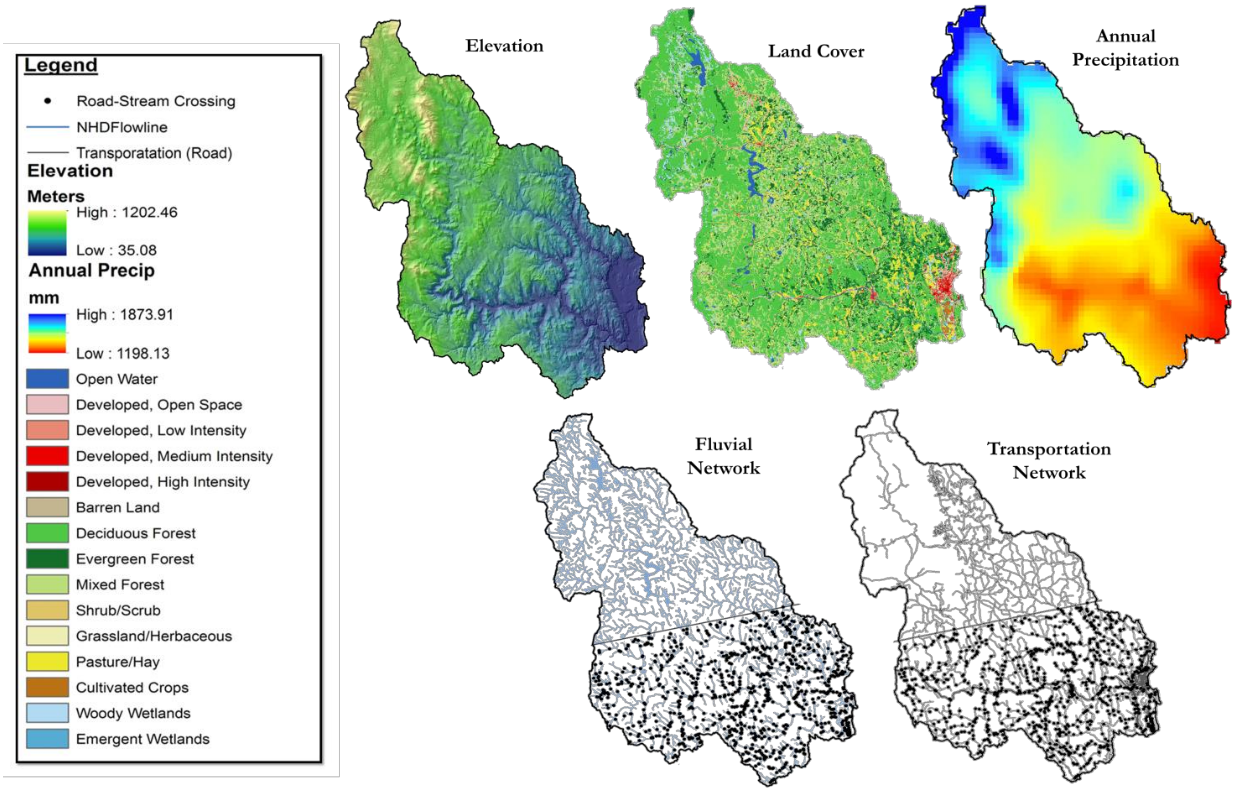


Figure 6: Deerfield River basin.

Table 2: Selected catchment characteristics in the Deerfield River basin and the three unimpaired USGS streamflow gages.

Catchment Property	Deerfield Basin	01170100 Green River (11)	01169000 North River (12)	01169900 South River (14)
Drainage Area (km2)	1718.23	107.66	231.24	62.78
Mean Annual Precipitation (mm)[a]	1374.88	1384.04	1378.52	1289.08
Mean Temperature (deg C)[a]	6.30	6.61	6.61	7.28
Max Temperature (deg C)[a]	12.10	12.44	12.36	13.15
Mean Elevation (m)[b]	475.11	413.51	430.79	343.22
Mean Slope (deg)[b]	9.0	9.8	8.6	8.8
North Facing (%)[b]	8.8	7.9	9.3	12.3
East Facing (%)[b]	17.3	16.9	17.6	17.9
Developed (%)[c]	5.3	3.0	4.4	6.8
Forest (%)[c]	82.0	90.3	84.0	78.6
Agriculture (%)[c]	5.9	3.8	7.8	10.0
Hydrological Group B (%)[d]	23.3	20.8	22.1	16.3
Hydrological Group C (%)[d]	1.9	0.7	0.7	0.8
Hydrological Group D (%)[d]	5.4	1.3	10.1	9.5
Stream Density (km/km2)[e]	1.48	1.67	1.40	1.31

Notes: [a] PRISM (2011); [b] National Elevation Dataset; [c] NLCD (2011); [d] NRCS SSURGO Dataset; [e] NHD High Resolution Dataset

Because there are a variety of hydrologic models that vary in complexity and structure, determining the effect of model choice on the vulnerability ranking of road-stream crossings is an important question for the DRP. Through an assessment of the differences between various hydrologic models at predicting flood flows, the uncertainty of these estimates can be better understood. Comparing the models at locations where historical data are available, it is possible to evaluate which models are more suitable for this particular region. These are the benefits of the multiple-model framework as applied to the vulnerability ranking analysis.

For the DRP, both process-based models and statistical models are employed to estimate a range of flood flows at the road-stream crossings in the basin (Figure 7). These models are also compared at locations where historical streamflow data exists to determine model performance. The data from the models used in this approach are incorporated into the hydrologic/hydraulic vulnerability ranking component of the vulnerability and risk assessment framework (Figure 8).

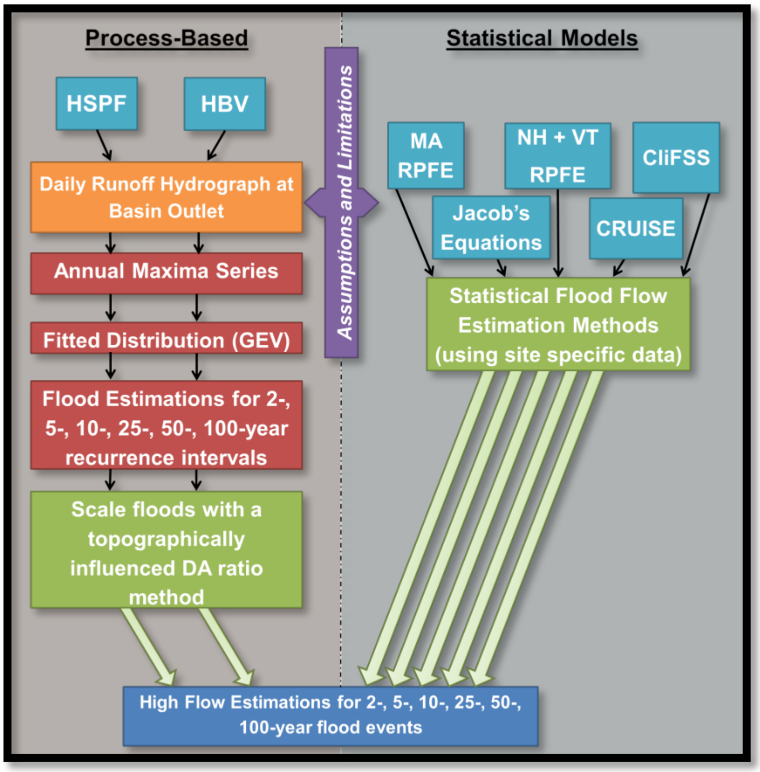


Figure 7: Multiple model framework for DRP.

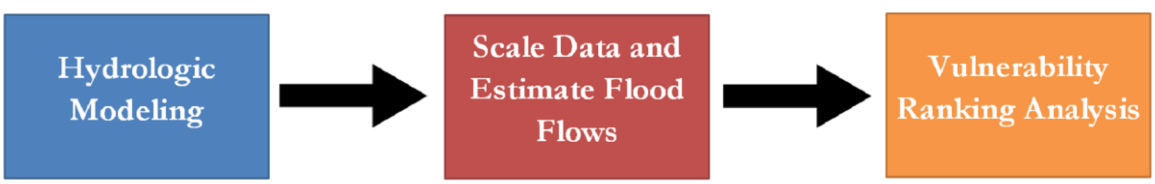


Figure 8: Conceptual flow diagram for estimating flood flows and performing vulnerability ranking analysis.

1.6.1 Vulnerability Analysis

To calculate the hydrologic vulnerability, it is necessary to assess the hydraulics of road-stream infrastructure on an individual location basis to determine hydraulic capacity. There are many different types of crossing infrastructure resulting in a wide variety of hydraulic responses. The site specific nature of these crossings requires a set of simplifying assumptions that allow for the calculation of hydraulic capacity. The hydraulic capacity is defined as the upper threshold of discharge that the structure is able to pass before it reaches a critical failure state. The critical failure state differs depending on the type of crossing infrastructure. For most crossings, hydraulic capacity can be defined as the total amount of discharge that can pass with an allowable headwater to culvert diameter ratio of 1.2. Literature suggests that it is at this point that the headwater elevation becomes greater than the culvert and submerges the inlet (Bodhaine, 1968; Normann *et al.*, 1985).

Once a critical discharge capacity has been established, the discharges at the design exceedance probabilities, mainly the 10- and 50-year storms, at the crossing location can inform a high, medium, or low probability of failure. A crossing falls into the high category if the critical discharge capacity ($Q_{critical}$) is greater than the discharge for the 10-year storm. A crossing will fall into a medium risk category if it lies between the 10- and 50-year storm discharge. If $Q_{critical}$ is above the 50-year design storm, the crossing is binned in the low-risk category (Figure 9).

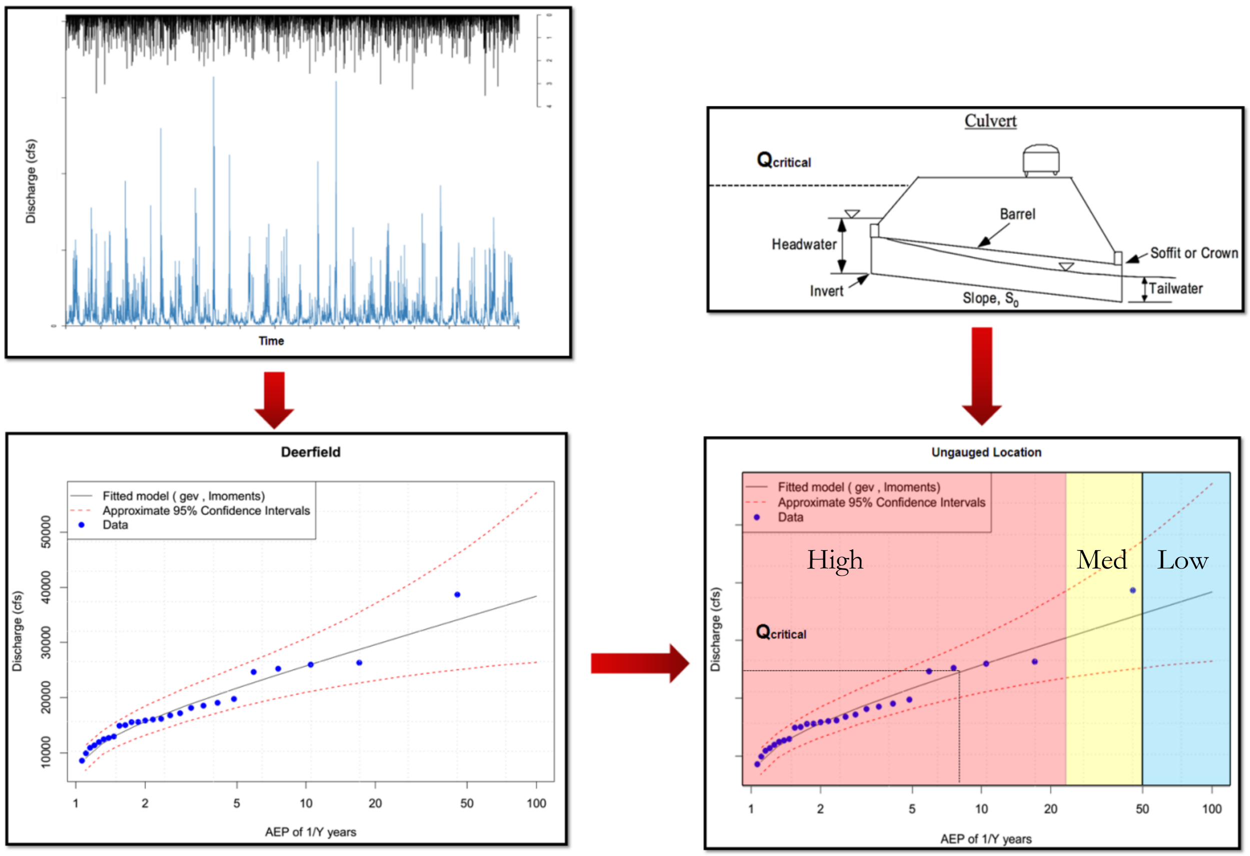


Figure 9: Conceptual schematic for road-stream crossing vulnerability analysis.

1.6.2 Future Climate Analysis

Climate change is expected to impact the range of extreme hydrologic events, however, the precise impact of these changes is difficult to estimate. Annual air temperature in the northeastern U.S. is projected to increase by an average of 5.3 degrees Celsius (˚C) by the end of the 21st century relative to 1961-1990 conditions (Hayhoe *et al.*, 2007). While future climate models (e.g. AOGCM simulations) are able to detect confident trends in average temperature increases, there is much more variability in the precipitation signal for the future making trend predictions much less robust (Hayhoe *et al.*, 2007). While at a seasonal time-step, future precipitation from future model scenarios suggest an increase in winter precipitation and little impacts on summer precipitation in the later part of the 21st century, even less is known about the change in frequency of higher intensity precipitation events in this region. There is much uncertainty in the modeling of future flood flows since the range of climate model simulations, which provide the driving input to the rainfall-runoff models, do not provide a clear consensus on higher intensity storm events. However, in the northeastern region both increasing trends in annual maximum instantaneous peak discharge and increasing trends in flood frequency have been noted (Collins, 2009; Armstrong *et al.*, 2012; Hodgkins, 2010). In addition, it has been suggested that New England hydroclimatic flood trends are congruent with the observed (increased) precipitation trends (Armstrong *et al.*, 2014).

Figure 10 provides a visualization of standard methods for estimating flood flows from the 75 years of record for the North River streamgauge in the Deerfield River basin. The two solid lines represent the flows predicted from the fitted GEV distribution or the first half of the historical record (1940 to 1975) to the second half of the historical record (1976 to 2015) and demonstrate the impact of choosing a time-frame for which low-frequency, high flow events are estimated. In this figure, the 90% confidence intervals for these two curves are presented to represent the uncertainty of both these estimations. The uncertainty increases as the storm return period increases. Based on this information, it appears that the magnitude and frequency of high-flow events in this region is increasing based on this historical streamflow record. A 20% increase in the 100 year storm between the two time periods is suggested.

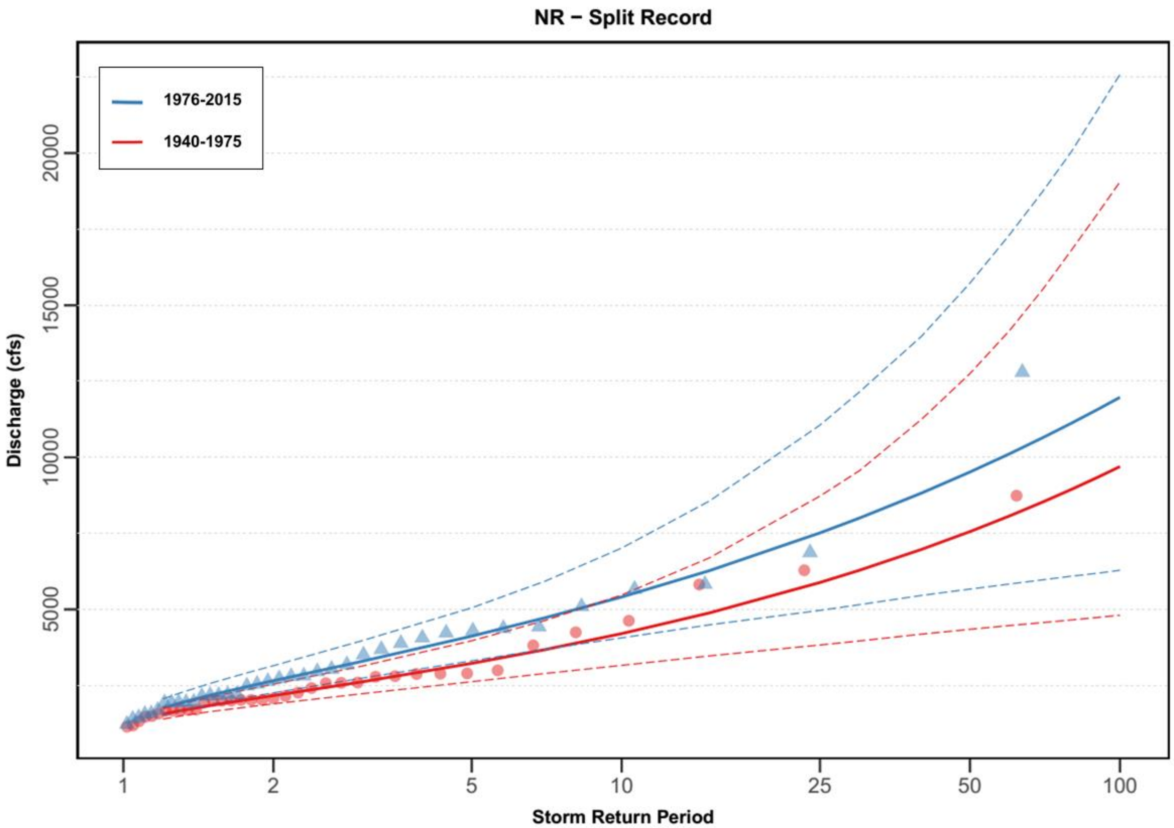


Figure 10: Split record at the North River USGS stream gauges (ID: 01169000). The solid lines represent the fitted GEV distribution for each half of the 75 years of record and the hashed-lines represent the 90% confidence interval for the GEV distribution model fit.

1.6.3 Decision Support Tool

As part of the MassDOT DRP, a decision support tool (DST) will be leveraged to provide flood flow estimations at ungauged locations throughout the Deerfield River watershed. This information will be made publicly available through the Spatial Hydro-Ecological Decision System (SHEDS) framework, which is a web-based interactive mapping tool developed by the USGS (Figure 11).

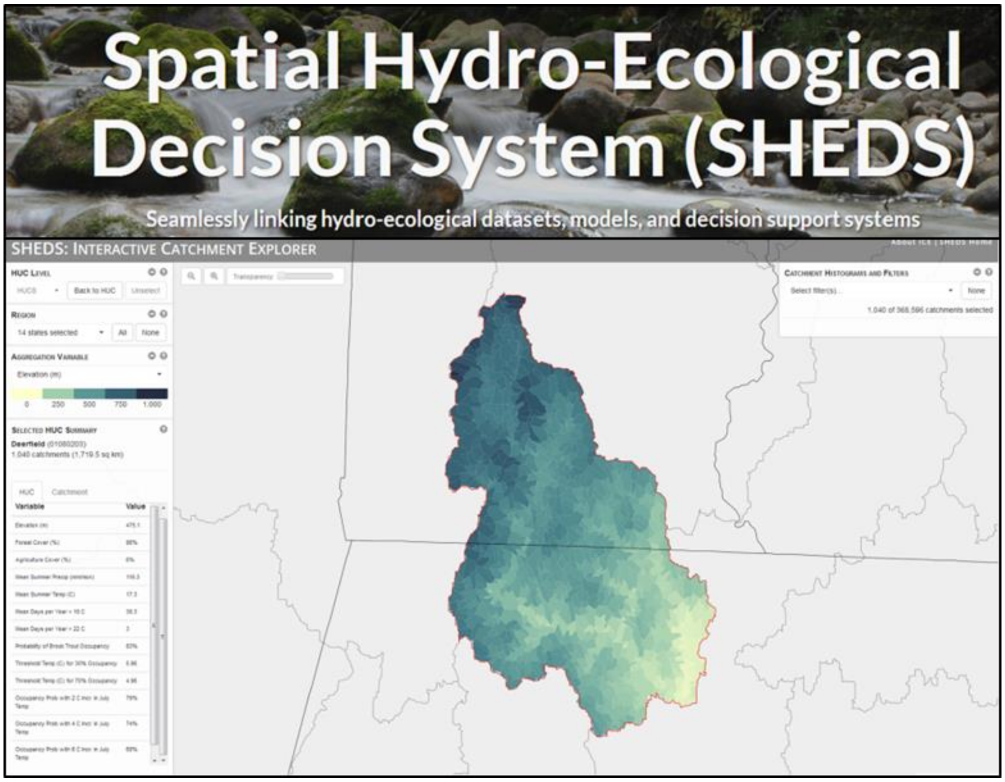


Figure 11: The USGS Spatial Hydro-Ecological Decision System (SHEDS) web-based interactive map.

1.7 References

Andersson, E., Nilsson, C., & Johansson, M. E. (2000). Effects of river fragmentation on plant dispersal and riparian flora. *Regulated Rivers Research & Management*, *16*(1), 83-89.

Armstrong, W. H., Collins, M. J., & Snyder, N. P. (2012). Increased Frequency of Low‐Magnitude Floods in New England1. JAWRA Journal of the American Water Resources Association, 48(2), 306-320.

Armstrong, W. H., Collins, M. J., & Snyder, N. P. (2014). Hydroclimatic flood trends in the northeastern United States and linkages with large-scale atmospheric circulation patterns. Hydrological Sciences Journal, 59(9), 1636-1655.

Bates, K., Bernard, B., Heiner, B. A., Klavas, J. P., & Powers, P. D. (2003). Design of road culverts for fish passage.

Bednarek, Angela T. (2001). Undamming Rivers: A review of the Ecological Impacts of Dam Removal. Environmental Management 27:6 pp 803-814.

Beven, K. (2013). So how much of your error is epistemic? Lessons from Japan and Italy. Hydrological Processes, 27(11), 1677-1680.

Blom, G. (1958). Statistical Estimates and Transformed Beta-Variables. Wiley. New York, NY, USA.

Blöschl, G., & Sivapalan, M. (1995). Scale issues in hydrological modeling: a review. Hydrological processes, 9(3 ‐ 4), 251-290.

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

Bodhaine, G. L. (1968). Measurement of peak discharge at culverts by indirect methods. US Government Printing Office.

Bogan, A. E. (1996). Decline and decimation: the extirpation of the unionid bivalves in North America. Journal of Shellfish Research, 15, 484.

Chang, H., Lafrenz, M., Jung, I. W., Figliozzi, M., Platman, D., & Pederson, C. (2010). Potential impacts of climate change on flood-induced travel disruptions: a case study of Portland, Oregon, USA. *Annals of the Association of American Geographers*, *100*(4), 938-952.

Chisholm, I., and L. Aadland. (1994). Environmental impacts of river regulation. Minnesota Department of Natural Re- sources, St. Paul, Minnesota, 31 pp.

Collins, M. J. (2009). Evidence for changing flood risk in New England since the late 20th century1.

Crawford, N. H., & Linsley, R. K. (1966). Digital Simulation in Hydrology'Stanford Watershed Model 4.

Cunnane, C. (1978). Unbiased plotting positions—a review. *Journal of Hydrology*, *37*(3), 205-222.

Damschen, E. I., Haddad, N. M., Orrock, J. L., Tewksbury, J. J., & Levey, D. J. (2006). Corridors increase plant species richness at large scales. Science, 313(5791), 1284-1286.

Fahrig, L. (2003). Effects of habitat fragmentation on biodiversity. Annual review of ecology, evolution, and systematics, 487-515.

Furniss, M. J., Ledwith, T. S., Love, M. A., McFadin, B. C., & Flanagan, S. A. (1998). *Response of roadstream crossings to large flood events in Washington, Oregon and northern California* (No. 9877 1806-SDTDC). Publication.

Friesz, P. J. (1996). Geohydrology of stratified drift and streamflow in the Deerfield River basin, northwestern Massachusetts. US Department of the Interior, US Geological Survey.

Gay, F. B., Toler, L. G., & Hansen, B. P. (1974). Hydrology and water resources of the Deerfield River basin, Massachusetts (No. 506).

Gauthier[1], M. E., Leroux, D., & Assani, A. Vulnerability of culvert to flooding.

O'Hanley, J. R. (2011). Open rivers: barrier removal planning and the restoration of free-flowing rivers. *Journal of Environmental Management*, *92*(12), 3112-3120.

Hayhoe, K., Wake, C. P., Huntington, T. G., Luo, L., Schwartz, M. D., Sheffield, J., ... & Troy, T. J. (2007). Past and future changes in climate and hydrological indicators in the US Northeast. Climate Dynamics, 28(4), 381-407.

Hodgkins, G. A. (2010). Historical changes in annual peak flows in Maine and implications for flood-frequency analyses. U. S. Geological Survey.

Hosking, JRM, (1990), L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. Journal of Royal Statistical Society, Series B, 52(2):105-124.

Jackson, S. D. (2003). Design and construction of aquatic organism passage at road-stream crossings: ecological considerations in the design of river and stream crossings.

Jackson, S. D., Bowden, A., Lambert, B., & Singler, A. (2011). Massachusetts River and Stream Crossing Standards.

Januchowski-Hartley, S. R., Diebel, M., Doran, P. J., & McIntyre, P. B. (2014). Predicting road culvert passability for migratory fishes. *Diversity and Distributions*, *20*(12), 1414-1424.

Januchowski-Hartley, S. R., McIntyre, P. B., Diebel, M., Doran, P. J., Infante, D. M., Joseph, C., & Allan, J. D. (2013). Restoring aquatic ecosystem connectivity requires expanding inventories of both dams and road crossings. *Frontiers in Ecology and the Environment*, *11*(4), 211-217.

Kalantari, Z., Briel, A., Lyon, S. W., Olofsson, B., & Folkeson, L. (2014). On the utilization of hydrological modelling for road drainage design under climate and land use change. *Science of the Total Environment*, *475*, 97-103.

Kemp, P. S., & O'hanley, J. R. (2010). Procedures for evaluating and prioritising the removal of fish passage barriers: a synthesis. *Fisheries Management and Ecology*, *17*(4), 297-322.

Knox, C. E., & Nordenson, T. J. (1955). Average annual runoff and precipitation in the New England-New York area (No. 7).

Kochanek, K., Markiewicz, I., Strupczewski, W,G. (2010). "On Feasibility of LMoments method for distributions with cumulative distribution function, and its inverse inexpressible in the explicit form". International Workshop: Advances in Statistical Hydrology. Taormina, Italy.

Krause, P., Boyle, D. P., & Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*, *5*, 89-97.

Kuczera, G, Lambert, M, Heneker, T, Jennings, S,Frost,A and Coombes,P. (2003), Joint Probability and Design Storms at the Crossroads, Proc2003. Hydrology and Water Resources Symposium, IEAust, Wollongong.

Kuczera, G., & Frank, S. (2006). Australian Rainfall and Runoff, Book IV, Estimation of Peak Discharge–Draft. *Engineers Australia*.

Lawrence, J. E., Cover, M. R., May, C. L., & Resh, V. H. (2014). Replacement of culvert styles has minimal impact on benthic macroinvertebrates in forested, mountainous streams of Northern California. *Limnologica-Ecology and Management of Inland Waters*, *47*, 7-20.

Ligon, F. K., W. E. Dietrich, and W. J. Thush, (1995). Downstream Ecological Effects of Dams. BioScience 45(3):183-192.

Martin, E. H., & Apse, C. D. (2011). Northeast aquatic connectivity: An assessment of dams on northeastern rivers. *The Nature Conservancy, Eastern Freshwater Program*.

McKay, S. K., Schramski, J. R., Conyngham, J. N., & Fischenich, J. C. (2013). Assessing upstream fish passage connectivity with network analysis. *Ecological Applications*, *23*(6), 1396-1409.

McGarigal, K., Compton, B. W., Jackson, S. D., Plunkett, E., Rolih, K., Portante, T., ... & Compton, B. (2012). Conservation Assessment and Prioritization System (CAPS) Statewide Massachusetts Assessment: November 2011. Landscape Ecology Program, Department of Environmental Conservation, University of Massachusetts, Amherst, MA.

Millington, N., Das, S., & Simonovic, S. P. (2011). The comparison of GEV, log-Pearson type 3 and Gumbel distributions in the Upper Thames River watershed under global climate models.

Milly, P. C. D., Julio, B., Malin, F., Robert, M., Zbigniew, W., Dennis, P., & Ronald, J. (2007). Stationarity is dead. *Ground Water News & Views*, *4*(1), 6-8.

Nadeau, E.J. (2008). Freshwater Mussels and the Connecticut River Watershed. Connecticut River Watershed Council, Greenfield, MA.

Nagrodski, A., Raby, G. D., Hasler, C. T., Taylor, M. K., & Cooke, S. J. (2012). Fish stranding in freshwater systems: Sources, consequences, and mitigation. *Journal of environmental management*, *103*, 133-141.

Neeson, T. M., Ferris, M. C., Diebel, M. W., Doran, P. J., O'Hanley, J. R., & McIntyre, P. B. (2015). Enhancing ecosystem restoration efficiency through spatial and temporal coordination. Proceedings of the National Academy of Sciences, 112(19), 6236-6241.

Normann, J. M., Houghtalen, R. J., & Johnston, W. J. (1985). Hydraulic design of highway culverts (No. FHWA-IP-85-15).

North Atlantic Aquatic Connectivity Collaborative NAACC. (2014). "About the NACC." Retrieved from < https://www.streamcontinuity.org/about_naacc/index.htm> on March 16, 2016.

Pépino, M., Rodríguez, M. A., & Magnan, P. (2012). Fish dispersal in fragmented landscapes: a modeling framework for quantifying the permeability of structural barriers. *Ecological Applications*, *22*(5), 1435-1445.

Petts, G. E. (1984). Impounded rivers: Perspectives for ecolog- ical management. John Wiley & Sons. Chichester, England, 322 pp.

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

PRISM Climate Group. (2004). Oregon State University, http://prism.oregonstate.edu, created 4 Feb 2004.

Rosenberg, E. A., Keys, P. W., Booth, D. B., Hartley, D., Burkey, J., Steinemann, A. C., & Lettenmaier, D. P. (2010). Precipitation extremes and the impacts of climate change on stormwater infrastructure in Washington State. *Climatic Change*, *102*(1-2), 319-349.

Sala, O. E., Chapin, F. S., Armesto, J. J., Berlow, E., Bloomfield, J., Dirzo, R., ... & Leemans, R. (2000). Global biodiversity scenarios for the year 2100. science, 287(5459), 1770-1774.

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

Smith, D. G. (1985). Recent range expansion of the freshwater mussel Anodonta implicata and its relationship to clupeid fish restoration in the Connecticut River system. Freshwater Invertebrate Biology, 105-108.

Stanford, J. A., J. V. Ward, W. J. Liss, C. A. Frissell, R. N. Williams, J. A. Lichatowich, and C. C. Coutant. (1996). A general protocol for restoration of regulated rivers. Regulated Rivers: Research and Management 12:391–413.

Stedinger, J. R. (1993). Frequency analysis of extreme events. *Handbook of hydrology*, *18*.

Stedinger, JR, Vogel, RM and Foufoula-Georgiou,E,(1993) .Frequency Analysis of Extreme Events in Handbook of Hydrology, Maidment, DR(ed.), McGraw-Hill, NewYork, NY,USA.

Thomas, D. M., & Benson, M. A. (1970). *Generalization of streamflow characteristics from drainage-basin characteristics*. Washington, DC, USA: US Government Printing Office.

Tilman, D., Fargione, J., Wolff, B., D'Antonio, C., Dobson, A., Howarth, R., ... & Swackhamer, D. (2001). Forecasting agriculturally driven global environmental change. Science, 292(5515), 281-284.

Villarini, G., Smith, J.A. (2010). Flood peak distributions for the eastern United States. Water Resour. Res. 46.

Vogel, R.M *et al.* (1993). "Flood-Flow Frequency Model Selection In Southwestern United States". Journal of Water Resources Planning and Management.

Vogel, R.M., Wilson, I. (1996). Probability distribution of annual maximum, mean, and minimum streamflows in the united states. J. Hydrol. Eng. 1, 69–76.

Wandle, S. W. (1983). *Estimating peak discharges of small, rural streams in Massachusetts*. US Department of the Interior, Geological Survey.

Ward, J. V. and J. A. Stanford, (1995). The Serial DisContinuity Concept —Extending the Model to floodplain Rivers. Regulated Rivers Research and Management 10(2-4):159-168.

Yeager, B. L. (1994). Impacts of reservoirs on the aquatic environment of regulated rivers. Tennessee Valley Authority, Water Resources, Aquatic Biology Department, Norris, Tennessee. TVA/WR/AB-93/1.

Zarriello, P. J., & Barbaro, J. R. (2014). Hydraulic assessment of existing and alternative stream crossings providing fish and wildlife passage at seven sites in Massachusetts (No. 2014-5146). US Geological Survey.

CHAPTER TWO

2.1 Abstract

Catchment-scale hydrologic predictions for current and future climate are of interest to river restoration and conservation efforts in the northeastern U.S. and well-calibrated rainfall-runoff models are useful towards this end. However, most of the catchments in this region are ungauged or poorly gauged posing a significant challenge to hydrologic modelers. This research uses a multiple model framework with regression-based regionalization at ungauged locations in the Deerfield River Basin, a major tributary to the Connecticut River Watershed. Two process-based rainfall-runoff models that differ in complexity are compared and evaluated in the region for accuracy of simulating the daily runoff hydrograph. Catchment characteristics are calculated across the study area and are correlated to the parameters of the rainfall-runoff models with a higher degree of accuracy compared to simpler, commonly applied approaches. This study provides a framework for hydrological modeling at ungauged locations in the region that may be more suitable for addressing hydrology at a small-catchment scale as well as provides a viable framework for addressing impacts such as climate change and flood flows at a local level. Furthermore, this study provides a superior method for regionalization of rainfall-runoff model parameters for ungauged basins in the northeastern U.S. region.

2.2 Introduction

The importance of accurate predictions of hydrologic processes are critical to the support of the sustainable management of our water resources and increasingly, society looks to science for these predictions, driving the field of hydrologic sciences to continuously improve its capacity and reliability (Blöschl, 2013). However, the majority of rivers and stream reaches and tributaries in the world are ungauged or poorly gauged (Blöschl, 2013; Sivapalan *et al.*, 2003; Young, 2006; Mishra and Coulibaly, 2009; Razavi and Coilibaly, 2012). These locations constitute what are called "ungauged basins" defined by Sivapalan *et al.* (2003) as locations that have inadequate records of hydrological observations, in terms of both data quantity and quality, to enable the reliable and accurate computation of hydrological variables of interest. Streamflow estimations serve as an important surrogate towards a more holistic interpretation of the hydrological and ecological regimes at a catchment scale and towards this end, there exist a variety of tools that can generate streamflow predictions over a range of time and space scales. However, these methods are almost always heavily dependent and driven by data. The effects of this fundamental lack of data drives a need to better understand and compare the variety of different methods and approaches for estimating hydrologic response at locations with little to no streamflow data.

There are two primary approaches to estimating streamflows at ungauged sites: process-based methods and stochastically-based methods. Stochastically-based methods include index methods for scaling flows from another gauged location, regression estimations of hydrologic variables such as flood flows, or geo-statistical methods that take into account spatial characteristics using, for example kriging techniques(Skøien *et al.*, 2006). Process-based methods include the use of rainfall-runoff models (RRMs) to represent the surface runoff at a location based on mathematical modeling of the physical processes behind the hydrological components across a landscape. The deterministic approach inherent to RRMs provide unique advantages to hydrologists and water resource engineers as they provide an ability to interpret the hydrologic landscape more fully and also provide the opportunity to impose perturbations in the model input data (e.g. land-cover, meteorological data, etc.) and measure hydrologic responses: a unique ability of process-based modeling. The ability to collect data and measure the goodness-of-fit (GOF) of simulated hydrologic variables is also an important advantage of a process-based modeling approach. However, RRMs are also heavily data driven and model parameters are estimated through calibration against observed historical streamflow data at a gauged location until the catchment and model behavior show sufficient agreement.

Regionalization refers to the process of transferring hydrological information, for instance the parameters of a RRM, from one catchment to another (Blöschl and Sivapalan, 1995) and is a well-recognized solution to provide time series of streamflow for ungauged basins (Young, 2006; Samuel *et al.*, 2011; Razavi and Coulibaly, 2012). The process of regionalization may be satisfactory if the catchments are similar (hydrologically, topographically, climatically, ecologically) but error-prone if not (Blöschl and Sivapalan, 1995; Razavi and Coilibaly, 2012). In addition, regionalization is widely regarded as a challenging task by hydrologists (Sivapalan *et al.*, 2003, Oudin *et al.* 2008; Stoll and Weiler, 2010; Samuel *et al.*, 2011; Razavi and Coilibaly, 2013) despite continued efforts in this area.

In recent decades, research has focused on the problem of regionalization for ungauged basins of which much effort has stemmed from the Predictions in Ungauged Basins (PUB) research decade from 2003-2012 initiated by the International Association of Hydrological Science (IAHS) (Hrachowitz *et al.*, 2013; Sivapalan *et al.*, 2003; Blöschl, 2013). Despite the significant research efforts over the past decade, the need to improve estimates and understand the effectiveness of different approaches remains (Hrachowitz *et al.*, 2013; Steinschneider *et al.*, 2014). Currently, there is no universal method of regionalization for a given region or catchment and the common approach is testing and applying various regionalization methods to attempt to shed light on the most appropriate for a specific location (Samuel *et al.*, 2011).

Several different regionalization approaches exist to estimate model parameters at ungauged locations. Three common regionalization approaches include the spatial proximity method, model averaging, and regression techniques (Blöschl, 2013). The most commonly employed and most straightforward methods are the spatial proximity and averaging methods in which one or more similar gauged catchments in the region are identified and the assumption is made that the parameter set derived from these donor (or analogue) catchments is also valid for the ungauged location (Blöschl, 2013). In the spatial proximity approach, hydrological model parameters are estimated based on the assumption that hydrologic and climactic similarity are a function of distance from a specified location and a model parameter set at an ungauged location is assumed most similar to the closest gauged catchment (Merz and Blöschl , 2004; Vandewiele and Elias, 1995; Parajka *et al.*, 2007; Bardossy, 2007; McIntyre *et al.*, 2005; Oudin *et al.*, 2008; Parajka *et al.*, 2005). The model averaging approach assumes that hydrologic and climactic similarity can be derivative of multiple donor catchments where catchments are selected based on proximity, catchment characteristics, or both (Goswami *et al.*, 2007; Kim and Kaluarachchi, 2008; Seibert and Beven, 2009; Blöschl, 2013).

A regression approach is often used in which calibrated RRM parameters are related to catchment physical and climate characteristics. This approach assumes that the model parameters are closely related to catchment attributes, since the model parameters are designed to be representative of the functional behavior of catchment physical and climate driven processes (Merz and Blöschl , 2004). While this method is popular in regionalization studies (Abdulla and Lettenmaier, 1997; Seibert, 1999; Merz and Blöschl , 2004; Hudecha and Bardossy, 2004; Wagener and Wheater, 2006; Oudin *et al.*, 2008; Kling and Gupta, 2009), often low correlations between model parameters and catchment attributes are discovered and this method has been strongly criticized (Bardossy, 2007; McIntyre *et al.*, 2005; Oudin *et al.*, 2008; Parajka *et al.*, 2007; Zhang *et al.*, 2009). However, regression-based regionalization of model parameters is inherently useful in elucidating the underlying influential characteristics of hydrologic response and has been successfully applied in the literature (see Razavi and Coulibaly, 2013). Although it has been suggested that parameter sets from neighboring gauged catchments may be more useful to estimate parameter sets on ungauged catchments than establishing relationships between catchment descriptors and model parameters (Merz and Blöschl, 2004; McIntyre *et al.*, 2005; Kay *et al.*, 2007; Oudin *et al.*, 2008), the authors believe that this conclusion may be affected by model structure and complexity.

Because of the numerous available hydrologic models, it is difficult to ascertain which RRM is best suited for a particular region or application or the degree of complexity that is appropriate for a specific application (Bevin, 2011). However, given a set of models for a catchment that are considered appropriate for a given objective (e.g. estimating the daily runoff hydrograph or predicting the 50-year flood), hypotheses testing can be used to gain insight towards model suitability (Clark *et al.*, 2011a; Beven, 2011; Beven *et al.*, 2012). An overabundance of RRMs may be symptomatic of an insufficient scientific understanding of environmental dynamics at the catchment scale (Clark *et al.*, 2011a; Clark *et al.*, 2011b). This is not surprising given the difficulties in measuring and representing the heterogeneity inherent to natural systems (Koren *et al.*, 2003; Duan *et al.*, 2006; Grayson *et al.*, 1992; Beven, 1989; McDonnell *et al.,* 2007; Beven, 2002; Kirchner, 2006).

Differences in the simulation of hydrologic processes and model structure can directly affect the accuracy of model results (Johnson *et al.*, 2003). In addition, there have been few studies that have compared the results of differing watershed models applied to the same catchment. Summaries of this literature are given by Perrin *et al.* (2001) and Refsgaard and Knudsen (1996). A recent study focused on the assessment of different models for predicting the daily runoff in a single catchment have discovered that models do indeed differ in their accuracy in predictions, however some models may be more well-suited than others depending on the hydrologic information of interest to the user such as peak flows (Linhart *et al.*, 2013). Gan *et al.* (1997) reports that significant differences among simulation results from differing models applied in a common catchment were primarily due to the differences in the models' runoff-generating mechanisms. Model complexity has also been investigated with respect to performance (Orth *et al.*, 2015), however the results show that the definition of accuracy is dependent on the hydrologic variable being assessed. There have been few, if any, studies in the literature that have compared regression-based regionalization approaches across different models of varying complexity for a particular location.

This paper assesses the responses of two different RRMs of varying complexity with a regression-based regionalization approach. A more parsimonious lumped-conceptual model (HBV) and a slightly more complex semi-distributed model (HSPF) are commonly used for catchment scale rainfall-runoff modeling and are compared in this study. The paper quantifies the accuracy of regression regionalization in forecasting the daily runoff hydrograph in ungauged basins. Despite the small-scale heterogeneity and process complexity, the hydrologic response at the catchment scale is often characterized by surprising simplicity (Sivapalan 2003a) which can often be represented quite well by a lumped-conceptual model (Sivapalan, 2005). Consequently, distributed and semi-distributed RRMs are considered to provide a more realistic representation of the spatial heterogeneity of hydrological processes because of their more complex model structure. A comparison of model performance as it relates to predictions in ungauged basins will be useful for directly comparing two different model structure types to help to identify appropriate regionalization approaches for the study region, provide an implicit accounting of emergent hydrological processes at the catchment scale, and create a framework for comparing the accuracy different RRMs within a small northeastern US catchment.

Finally, because the Northeastern US hydrologic regime is heavily affected by dam operations at a larger catchment scale, modeling the surface runoff hydrograph accurately becomes incredibly challenging when applying RRMs. However, this type of approach is often implemented at these scales in the northeastern US region and studies have mostly focused mainly on larger catchments (e.g. Marshall and Randhir, 2008; Parr and Wang, 2015a; Parr and Wang, 2015b; Parr *et al.*, 2014). The application of these models at this larger catchment scale may not be appropriate to draw conclusions about the hydrological responses at the smaller catchment scale, especially when using the daily runoff hydrograph generated by the RRMs to extrapolate other hydrologic variables of interest (e.g. flood flows). In this respect, this paper provides a framework for hydrological modeling at ungauged locations in the northeastern US that may be more suitable for addressing hydrology at a small-catchment scale as well as suggest a more suitable method for addressing impacts such as climate change and flood flows at a local level.

2.3 Study Area and Data

2.3.1 Study Area

The Deerfield River basin straddles the border between north-western Massachusetts and southern Vermont with a drainage-area of approximately 1722 km^2. It is a major sub-basin of the Connecticut River. The largest tributary in the Deerfield basin is the North River with a total drainage-area of approximately 240 km^2 (Figure 12). There is extensive hydroelectric-power generation (ten major dams) in the basin and the flows on this river are considered to be heavily altered by these activities (Friesz, 1996).

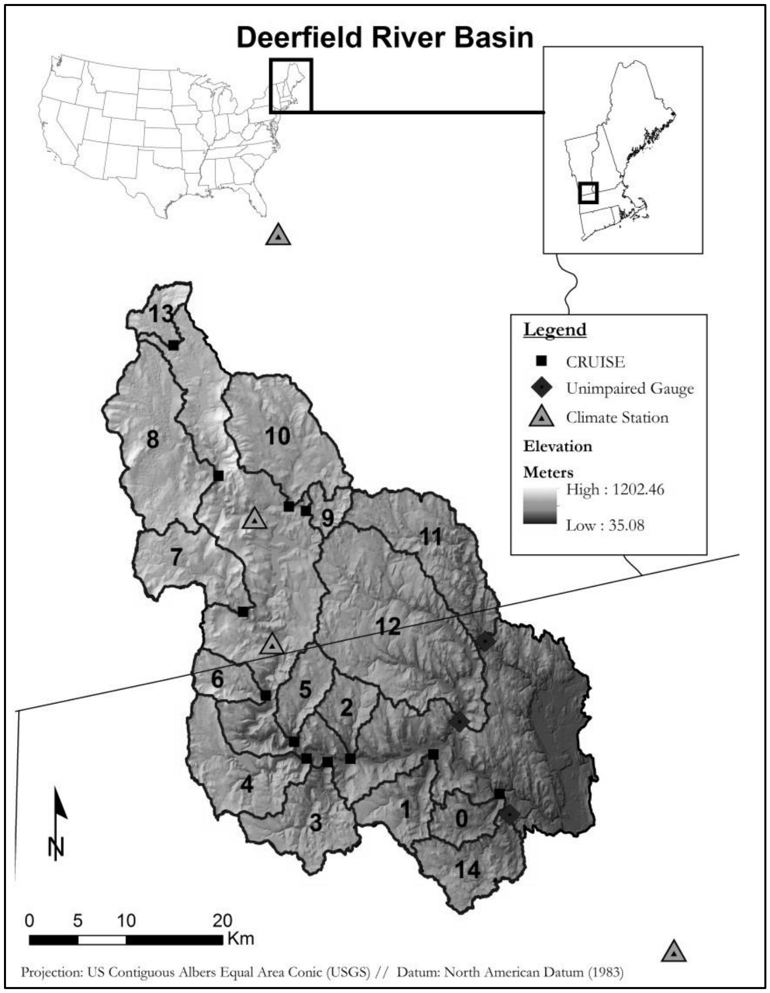


Figure 12: Overview of Deerfield River basin, selected watershed locations for both direct and indirect calibration/validation of RRMs, and climate station data locations.

The Berkshire Hills physiographic province contributes most of the drainage area of the Deerfield River Basin. It consists of narrow river valleys boarded by steep hillslopes. The southeastern part of the basin is part of the Connecticut Valley Lowlands physiographic province where the topography is flatter than the Berkshire Hills (Fenneman, 1938; Friesz, 1996). Elevations in the Deerfield Basin range from about 35 meters above sea level in the Connecticut Valley Lowlands to about 1,202 meters in the ridges of the Berkshire Hills with a mean altitude of about 475 meters. Average annual precipitation in the basin is 107-112 cm in the low altitudes to 127-188 cm in the higher altitudes (PRISM Climate Group, 2004; Knox and Nordenson, 1955). Snowmelt in spring and evapotranspiration in summer and fall cause annual cyclical trends in mean monthly runoff, even though mean monthly precipitation is evenly distributed throughout the year (Gay *et al.*, 1974).

2.3.2 Data

2.3.2.1 Climate Data

Historical climate data used for modeling included precipitation, temperature, and potential evapotranspiration. Hourly precipitation was used exclusively for the HSPF model because of its model structure. A total of four stations were selected based on the time-period of record and quality of data, two of which are located in the Deerfield River Basin and two which are adjacent to the basin (Figure 12). These observations were selected based on their proximity to the basin as well as the continuity and temporal overlap of the historical records.

Climate data was distributed across the subbasins in the HSPF model by the spatial proximity method. Climate data were averaged across the stations for the HBV model. Daily potential evapotranspiration (PET) was estimated using the Hamon (1963) method for the HBV model which computes the PET based on daytime length and the saturated vapor density calculated using the mean daily air temperature and a coefficient of 0.0065 (Hamon, 1963). The Hamon PET model was applied because of its simplicity of data inputs and accuracy (Lu *et al.*, 2005; Federer *et al.*, 1996; Vorosmarty *et al.*, 1998; McCable *et al.*, 2015).

2.3.2.2 Surrogate and Observed Streamflow in Deerfield Basin

There are seven streamflow gages as part of the USGS National Water Information System (NWIS) network, however, only three of these gages are considered unimpaired (Falcone, 2010), as there are several major dams in the basin (Table 3). These three unimpaired gages represent approximately 23% of the total drainage area of the Deerfield River basin. Forest cover is the dominant land-use type in each of these catchments and there is relatively little impervious or developed landscape. The North River is the Deerfield's largest gaged unimpaired tributary with a historical observation record of about 73 years. The Green River and South River gauges (1170100 and 01169900) have periods of record of about 49 and 48 years, respectively.

Table 3: Catchment characteristics of the three unimpaired streamflow gauges in the Deerfield River basin. These three catchments represent about 23% of the total drainage-area of the Deerfield River basin.

Catchment Property:	**01170100 Green River (11)**	**01169000 North River (12)**	**01169900 South River (14)**
Drainage Area (km2)	107.66	231.24	62.78
Mean Annual Precipitation (mm)[a]	1384.04	1378.52	1289.08
Mean Temperature (deg C)[a]	6.61	6.61	7.28
Max Temperature (deg C)[a]	12.44	12.36	13.15
Mean Elevation (m)[b]	413.51	430.79	343.22
Mean Slope (deg)[b]	9.8	8.6	8.8
North Facing (%)[b]	7.9	9.3	12.3
East Facing (%)[b]	16.9	17.6	17.9
Developed (%)[c]	3.0	4.4	6.8
Forest (%)[c]	90.3	84.0	78.6
Agriculture (%)[c]	3.8	7.8	10.0
Hydrological Group B (%)[d]	20.8	22.1	16.3
Hydrological Group C (%)[d]	0.7	0.7	0.8
Hydrological Group D (%)[d]	1.3	10.1	9.5
Stream Density (km/km2)[e]	1.67	1.40	1.31

Notes: [a] PRISM (2011); [b] USGS NED (2002); [c] NLCD (2011); [d] NRCS SSURGO Dataset; [e] USGS NHD High Resolution Dataset

The Connecticut River UnImpacted Streamflow Estimation (CRUISE) tool is used to estimate the streamflow at 12 additional major subbasins within the Deerfield River Basin to use as surrogate data for an indirect RRM calibration procedure. Vleeschouwer and Pauwels (2013) suggest that in the case of spatial gauging divergence, that is when no observed discharge records are available at the outlet of the ungauged catchment, the calibration can be carried out successfully based on a rescaled discharge time series of a “very similar” donor catchment. Because the Deerfield River Basin falls within the larger Connecticut River Basin, the CRUISE tool was applied. The CRUISE tool uses a geostatistical approach to select the donor catchment, calculates the cross-correlation coefficients of runoff with unimpacted streamflow gages in the Connecticut River Basin, and then interpolates these correlation coefficients in space using kriging (Blöschl *et al.*, 2013). For CRUISE, basin characteristics are computed using the online USGS Streamstats tool and these characteristics are then used in this procedure to identify the most suitable catchment. This method has been shown to give better runoff estimates than when choosing the nearest streamgage as the donor (Blöschl *et al.*, 2013; Archfield *et al.*, 2013) By increasing the number of subbasins modeled in the Deerfield basin using an indirect RRM calibration procedure in addition to the unimpaired NWIS gages, the authors believe that a more accurate representation of the physical hydrological processes and landscape heterogeneity could be obtained across the entire extent of the Deerfield basin (Figure 12).

2.3.2.3 Catchment Characteristics

Catchment characteristics were calculated from publically available raster datasets. The USGS National Elevation Dataset (NED) was downloaded and clipped to the catchment area where it was used to delineate the subbasins used in this study and derive the elevation, slope, and aspect characteristics. The USGS National Hydrography Dataset (NHD) was used to determine the total length of stream and the stream density of the subbasins. The National Land Cover Database 2011 (NLCD, 2011) was used to determine the different types of land cover in the subbasins, which was reclassified to represent agricultural, forest, and developed land categories (Homer *et al.*, 2011). The National Resources Conservation Service (NRCS) SSURGO database was used to estimate the different hydrological soil groups across the subbasins. Finally, the PRISM raster datasets were used to provide estimates of average annual climate characteristics (PRISM Climate Group, 2004). This particular dataset was chosen because it has been used extensively in evaluating annual normals for precipitation and temperature in addition to being homogeneously applied throughout the region as a single uniform dataset. Catchment characteristics were all chosen based on their hydrological value, but also based on their accessibility and ease of computation.

2.4. Methodology

Two process-based RRMs, Hydrologic Simulation Program Fortran (HSPF) and the Hydrologiska Byråns Vattenbalansavdelning (HBV) model, are applied to fifteen subbasins in the Deerfield River basin. Because of the lack of unimpaired stream gauge data within this basin, both direct and indirect calibration to historical streamflow data were applied. A split sample calibration and validation routine is applied using the shuffled complex evolutionary (SCE-UA) genetic algorithm for calibration with the Kling-Gupta Efficiency (KGE) criterion as the objective function. Finally, calibrated model parameters are related to catchment characteristics in the basin to inform regression based regionalization and the accuracy of this method is tested through comparison to two other commonly applied methods as well as through open and closed-form validation. The details of each step in the methodology are provided in the following subsections.

2.4.1 Hydrologic models

2.4.1.1 Hydrologic Simulation Program Fortran (HSPF)

The Hydrologic Simulation Program Fortran (HSPF) model was selected for use in this study. HSPF has been applied across the northeastern US with much success (Taner *et al.*, 2011; Srinivasan *et al.*, 1998; Johnson *et al.*, 2003; Filoso *et al.*, 2004) on the mid-Atlantic region (Seong *et al.*, 2015; Gutiérrez-Magness and McCuen, 2005; Kim *et al.*, 2007; Gao *et al.*, 2014; Doherty and Johnston, 2003) as well as other places throughout the world (Baloch et al, 2011; Bergman and Donnangelo, 2000; Iskra and Droste, 2007; Saleh and Du, 2004). This model was selected based on its widespread use in the region as well as being one of the most mature RRMs in the field of hydrology.

The HSPF model is a process-based, semi-distributed, continuous simulation watershed model for quantifying runoff and addressing water quality impairments associated with combined point and non-point sources (Bicknell *et al.*, 1996; Johnson *et al.*, 2003). The model was derived from the Stanford Watershed Model (SWM) developed by Norman Crawford and Ray Linsley, which was developed in the early 1960's and is credited as being the first computer based watershed model. The SWM was transformed into the Hydrologic Simulation Program Fortran (HSPF) in 1974 by the newly formed U.S. Environmental Protection Agency (EPA). The HSPF model is currently maintained by the EPA and exists as a core watershed model tool in EPA's software application BASINS (Better Assessment Science Integrating Point and Non-point Sources) in the current version 4.0 (2013).

BASINS was used to develop the HSPF model files as well as assemble the climate data needed for the model. The WinHSPF v3.0 interface within BASINS was used to automatically estimate the F-Tables for all the reaches in the model for the channel routing sub-routine. Reaches within each subbasin were defined using automatic watershed delineation methods in BASINS using a minimum drainage-area threshold of 2.5 km^2. A degree-day snow simulation was also applied for the HSPF model that uses a simple approach for estimating snow in the watersheds using minimum and maximum daily temperature data. The HSPF model was executed using WinHSPFLt called through the Windows 7 command line interface using RStudio and R (ver 3.2.1). Post-processing of the model output was performed using R coupled with the Python (ver 3.4.3) 'wdmtoolbox' package (ver 0.9.0) that allowed the extraction of the model output data from the HSPF binary WDM files.

In the HSPF model water mass and energy balances are simulated though the use of hydraulic response units (HRUs). The model is typically used at a spatial resolution that ranges in extremes from 10 to 100 km^2. HRUs provide a distributed calculation of surface runoff, interflow, and groundwater flow to streams by processes that determine the fate of water through losses and storage. Flows from the HRUs are typically directed to streams and routed by the kinematic-wave method to simulate streamflow. HSPF can simulate any period from a few minutes to hundreds of years using a time step ranging from sub-hourly to daily. Usually the model is executed for a time span ranging from 5 to 20 years or more using an hourly time step (Duda *et al.*, 2012). Input data includes both topographical controls and meteorological drivers. Meteorological drivers can include various climate data such as hourly precipitation, estimates of potential evapotranspiration, air temperature. Topographical controls include vegetation, digital elevation model (DEM), hydrography, and a land-use type layers.

Major elements of the HSPF model are reproduced from Crawford and Linsey (1966) in Figure 13. The calculations represented in this conceptual model diagram can be carried out by any number of reaches (HRUs) from any number of meteorological input stations. Upper and lower zone storages control overland flow, infiltration, interflow, and inflow to the groundwater while these two zone storages also combine together with groundwater storage to represent soil moisture profiles and groundwater conditions (Crawford and Linsey, 1966). Surface runoff is simulated as essentially an infiltration-excess process and the output from each HRU represent the average response of the HRU to precipitation and are routed to a stream channel (Johnson *et al.*, 2003). Flow is routed downstream from reach to reach by a kinematic wave method.

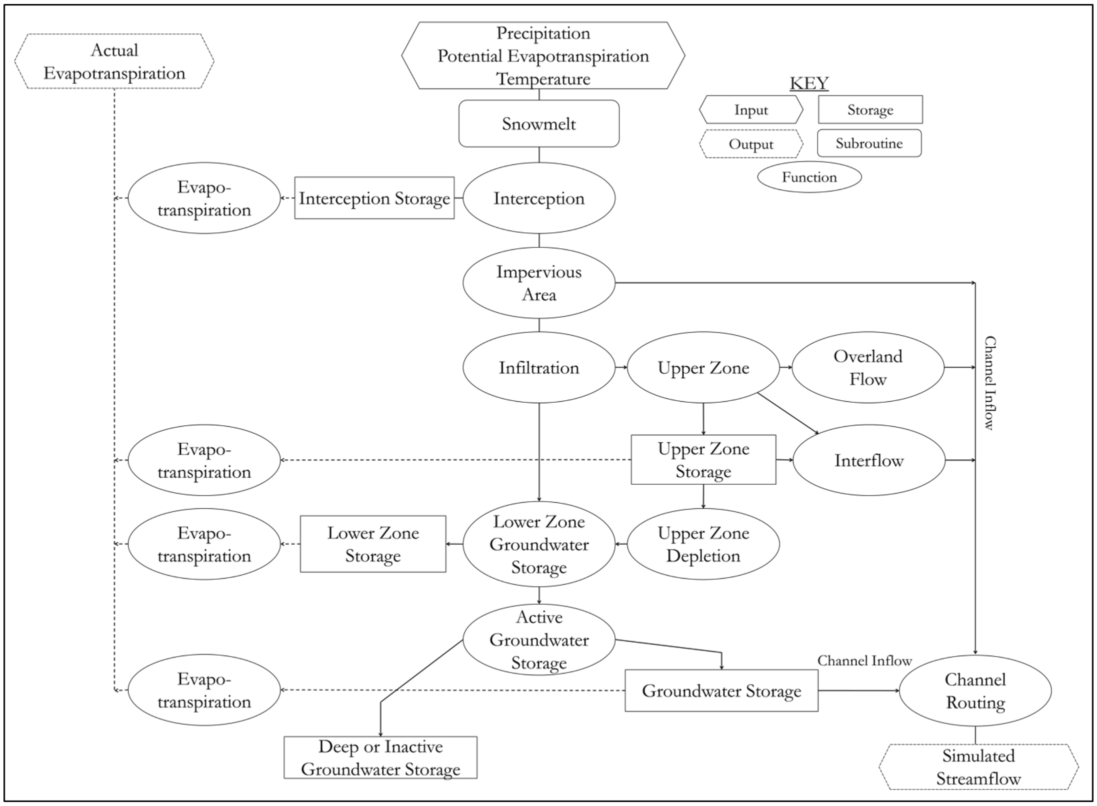


Figure 13: Conceptual flow diagram of the HSPF model based on a similar figure for the SWM published by Linsey and Crawford (1966).

2.4.1.2 Hydrologiska Byråns Vattenbalansavdelning (HBV Model)

The Hydrologiska Byråns Vattenbalansavdelning (HBV) model is a process-based, continuous streamflow simulation watershed model that has been characterized as a semi-distributed conceptual model (Lindstrom *et al.*, 2005; Parajka *et al.*, 2007), however it is often applied in a lumped-conceptual structure (eg. Merz and Blöschl , 2004; Yu and Yang, 2000; Singh and Woolhiser, 2002; Berstrom 1976). The HBV model was named after the Water Balance Department of the Hydrological Bureau Sweden (Berstrom and Lindstrom, 2015) and designed by the Swedish Meteorlogical and Hydrological Institute (Berstrom, 1992). Motivation for developing this model arose from a need for hydrological research purposes in Sweden followed by hydropower system forecasting. Although the SWM had been tested in Sweden, the HBV model provided a less complex approach that was congruent to the data availability in the observation network at the time (Wetterhall, 2014).

This model has been applied in more than 50 countries around the world and has been used extensively in Finland, Norway and Sweden, especially in the development of nationwide hydrological mappings (Bergström, 2006). In addition, new guidelines for flood prediction established in 1990 in Scandinavia included the HBV model as part of the design procedure (Flödeskommittén, 1990; Bergström *et al.*, 1992; Norstedt *et al.*, 1992; Bergström, 2006). Even though the model was originally developed for use in Scandinavia catchments, it has been effectively applied in tropical and subtropical areas as well (Bhatia *et al.*, 1984; Haggstrom *et al.*, 1990; Zhang and Lindstrom, 1997) and while the application of the HBV model has yet to be used extensively in the northeastern US, it holds significant promise as an applicable model to the region based on its original conception and application across cold, mountainous European climates.

Input to the TUW model includes daily precipitation, air temperature, and potential evapotransipiration estimations. There are three main routines for this model including a snow accumulation and melt routine, soil moisture accounting routine, and a response and channel routing routine (Figure 14). The snow routine consists of a simple degree day and threshold approach. The soil moisture accounting routine computes an index of the wetness of the entire basin and integrates interception and soil moisture storage (Bergstrom, 1992). The runoff response transforms the excess water from the soil moisture routine into river flow. This routine consists of two tanks (reservoirs) that represent different time dependent contributions to the river flow. Finally, a triangular distribution is used to attenuate the flood pulse at the basin outlet. For a further discussion on the history of the HBV model, including its application with respect to ungauged basins, see Bergstrom (2006).

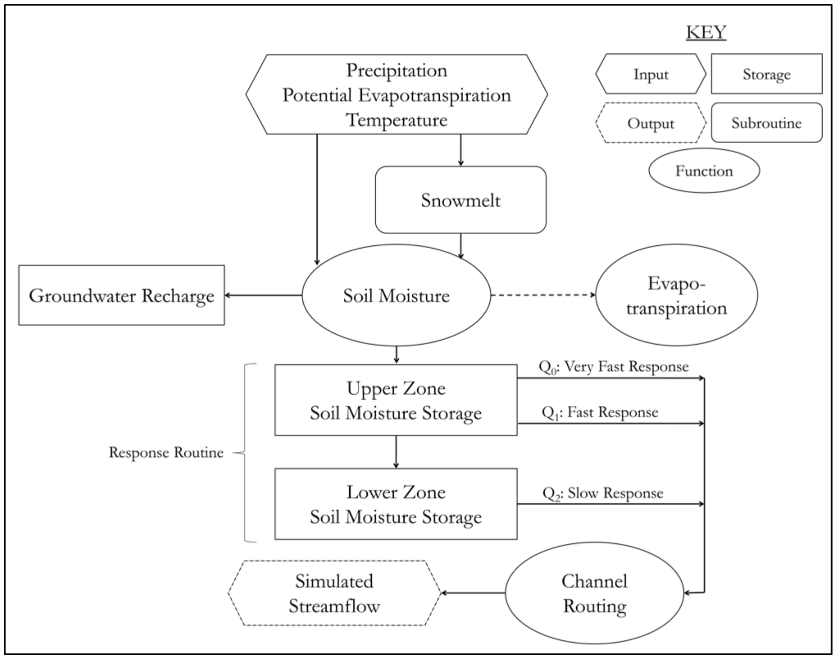


Figure 14: Conceptual flow diagram for the HBV model.

2.4.2 Quantifying Sensitive Parameters

The two RRMs used in this study have a large number of parameters which need to be calibrated to daily streamflow data (HSPF: +100; HBV: 15). Utilizing all the model parameters is almost impossible and it is necessary to identify the most sensitive parameters for the calibration process. Reducing the model parameters is important in reducing correlation and interdependence between parameters during the calibration process (Jackman and Hornberger, 1993; Zhang and Lindstrom, 1997). For this study, HSPF calibration parameters were selected based on peer-reviewed literature (Seong *et al.*, 2015; Kim *et al.*, 2007; Iskra and Droste, 2007; Gao *et al.*, 2014; Doherty *et al.*, 2003; Bicknell, 2000; Duda *et al.*, 2012, US EPA, 1999) and the personal modeling experience of the authors.

The HBV model uses fifteen parameters to simulate runoff response in a watershed. The parameter set was reduced to the eight most sensitive parameters informed through the literature (Zhang and

Lindstrom, 1997; Merz and Blöschl , 2004; Parajka *et al.*, 2007; Harlin and Kung, 1992) as well as from a Hornberger-Spear-Young generalized sensitivity analysis (HSY-GSA) method approach similar to Harlin and Kung (1992) (Hornberger and Spear, 1981; Young, 1983; Beck, 1987). A Monte Carlo approach is used to randomly generate parameter combinations from a uniform distribution based on the parameter constraints defined by Parajka and Viglione (2012) for 50,000 simulations. Model output is categorized as either behavioral or non-behavioral based on a threshold KGE value of 0.3. Behavioral simulations have a KGE value greater than 0.3 while non-behavioral simulations have a KGE value of less than 0.3 defined approximately by the average KGE value across all the Monte-Carlo simulation runs.

2.4.3 Calibration

To perform regionalization, calibrated model parameters are required for all gauged locations in the region. Model calibration was performed in R with the 'hydromad' library (ver. 0.9) that contained code for the shuffled complex evolution method developed at the University of Arizona (SCE-UA) which is very effective and efficient for global optimization for calibration of hydrological models (Wu and Zhu, 2006; Duan *et al.*, 1994; Seong *et al.*, 2015). The calibration period took place from January 1, 1980 to December 31, 1990 with a one-year warm-up period. Model validation was performed over January 1, 1991 to December 31, 1995.

Significant research exists concerning the differences between objective functions and their implication with respect to RRM calibration (Gutiérrez-Magness and McCuen, 2005; Gao *et al.*, 2014; Gupta *et al.*, 1998; Madsen, 2003). The objective function for the calibration period was to minimize the (–) KGE criterion at a daily-timestep. The KGE was defined by Gupta *et al.* (2009) and is a criterion that is essentially a decomposition of the NSE (and MSE) value. It can be expressed by the following equation:

$$\mathrm{KGE} = 1 - \mathrm{ED} \tag{1}$$

$$\mathrm{ED} = \sqrt{(\mathrm{r}-1)^2 + (\mathrm{a}-1)^2 + (\mathrm{b}-1)^2} \tag{2}$$

$$\mathrm{a} = \frac{\mathrm{S_y}}{\mathrm{S_x}};\ \mathrm{b} = \frac{\bar{\mathrm{x}}}{\bar{\mathrm{y}}} \tag{3}$$

where r is the correlation coefficient, $\bar{x}$ is the arithmetic mean of observed daily streamflow, $\bar{y}$ is the arithmetic mean of the modeled streamflow data; S_x and S_y represent the standard deviations for the observed and predicted data, respectively. The a term is a measure of relative variability of the predicted and observed values and b is the bias defined as the ratio of the mean and predicted flows to the mean of the observed flows. ED represents the Euclidean distance from the ideal point in the scaled space.

The KGE has been used in hydrologic modeling as an objective function that serves to mitigate some of the shortcomings of the NSE value. In particular, this metric has advantages over the NSE because it removes interactions between error components and reduces negative variability bias in simulation results (Steinschneider *et al.*, 2014). This criterion is composed of three independent components including mean bias, variability bias, and the correlation between simulated and observed flows. This value ranges from minus infinity to 1. Model accuracy is maximized as the KGE approaches unity. After calibration, model performance was evaluated by the authors' using a suite of performance criteria to achieve a more holistic interpretation of the model's performance (Table 4).

Table 4: Model goodness-of-fit was evaluated in this study over these criteria.

	Name	Abrv.	Equation	Range
(1)	Kling Gupta Efficiency Value	KGE	$\mathrm{KGE} = 1 - \mathrm{ED}$ $\mathrm{ED} = \sqrt{(\mathrm{r}-1)^2 + (\mathrm{a}-1)^2 + (\mathrm{b}-1)^2}$ $\mathrm{a} = \frac{\mathrm{S_y}}{\mathrm{S_x}}\,;\ \mathrm{b} = \frac{\bar{\mathrm{x}}}{\bar{\mathrm{y}}}$	-inf to 1
(2)	Coefficient of Determination	R2	$\mathrm{R}^2 = \left[\frac{\frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}(\mathrm{x_i}-\bar{\mathrm{x}})(\mathrm{y_i}-\bar{\mathrm{y}})}{\mathrm{S_x S_y}}\right]^2$	0 to 1
(3)	Nash-Sutcliffe Efficiency Value	NSE	$\mathrm{NSE} = 1 - \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}}(\mathrm{x_i}-\mathrm{y_i})^2}{\sum_{\mathrm{i}=1}^{\mathrm{N}}(\mathrm{x_i}-\bar{\mathrm{x}})^2} = 1 - \frac{\mathrm{MSE}}{\mathrm{S_x}^2}$	-inf to 1
(4)	Normalized Root Mean Square Error	NRMSE	$NRMSE = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - x_i)^2}}{S_x}$	0 to inf
(5)	Percent Bias	PBIAS	$PBIAS = 100\,\frac{\sum_{i=1}^{N}(y_i - x_i)}{\sum_{i=1}^{N} x_i}$	0 to inf
(6)	Volumetric Efficiency	VE	$\mathrm{VE} = 1 - \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}}\lvert\mathrm{y_i}-\mathrm{x_i}\rvert}{\sum_{\mathrm{i}=1}^{\mathrm{N}}\mathrm{x_i}}$	0 to 1

Notes: x_i is a set of observations; y_i is a set of predictions; $\bar{x}$ is the arithmetic mean of observed data, $\bar{y}$ is the arithmetic mean of the predicated data; Sx and Sy represent the standard deviations for the observed and predicted data, respectively; MSE represents the mean-square-error; ED represents the Euclidean distance from the ideal point in the scaled space; r is the correlation coefficient; a is a measure of relative variability of the predicted and observed values; and b is the bias defined as the ratio of the mean and predicted flows to the mean of the observed flows.

2.4.4 Regionalization

Attempts to define functional relationships can be assessed through correlating model parameters with catchment characteristics (e.g. topography and climate). This type of regionalization can be expressed by the following simple expression:

$$\hat{\theta}_L = F(\theta_R|\Phi) + \varepsilon_R \quad (4)$$

where $\hat{\theta}_L$is the estimated model parameter at the ungauged site, $F(\,)$ is a functional relation for the parameters, Φ is the set of catchment characteristics, θ_R is a set if regional model parameters, and ε_R is an error term. An ordinary-least-squares (OLS) linear regression approach defines the functional relationship between highly correlated catchment characteristics (Φ) with the RRM model parameters (θ_R) with the underlying assumption that the model parameters are independent. A Shapiro-Wilks test is used to determine if the set of calibrated model parameters and catchment are normally distributed and standard transformations are used on values from this test that were <0.05. If the distribution of a catchment characteristic was not correctable using a standard transformation, they were removed from the subsequent analysis.

Pearson's r value is calculated between the regional model parameters (θ_R) and the catchment characteristics across the subbasins (Φ) and significant relationships between these two independent data are determined. A threshold of 0.514 was used to identify significant relationships (p value<0.05). If a RRM parameter (dependent variable) had more than one significant relationship to a catchment characteristic (independent) variable, a principle component analysis (PCA) was applied to the significant independent variables and the first component was used in the OLS regression for each parameter. Otherwise, if the dependent variable had only one significant relationship or none that were above the threshold, the most significant independent variable was selected for the OLS regression. The use of the PCA in the regression development reduced both the dimensionality of the independent variables and eliminated the effects of colinearity between the catchment characteristics.

To compare the usefulness of our regression regionalization approach, two other methods are evaluated, namely spatial proximity and naïve approaches. Spatial proximity uses the parameter set from the closest donor catchment. The Euclidean distance is calculated between the ungauged catchment and the gauged catchments in the region and the catchment with the minimum distance is selected to be the donor catchment in this approach. The naïve mean is also compared in which the mean of the model parameters across the gauged sites are used at an ungauged site.

A "jack-knife" or "leave-one-out" cross validation (LOOCV) approach was used after the regression development. This LOOCV method was applied in a closed-form approach, in which the accuracy of hydrologic model parameter estimations were evaluated without running the model, as well as an open-form approach, in which the model parameters estimated were then used to simulate the streamflow and standard goodness-of-fit measurements were calculated. For the close-form analysis, the ordinary residuals and the leverages are used instead of fitting fifteen separate least-squares models and omitting each observation once. The hat matrix is calculated for each of the eight model parameter regressions, which describes the influence of each response value on each fitted value. The diagonal of the hat matrix is then used to calculate the deleted-residuals for each regression. These deleted-residuals are then used to create a plot with the estimated hydrologic model parameter value that has been left out through calculation of the deleted-residual with the actual calibrated values. In addition, the residuals from the OLS regression equations are calculated for each parameter in the hydrologic models and are mapped to the subbasins to identify any potential spatial clustering.

2.5 Results

For the HSPF model a total of eight parameters were adjusted to achieve an acceptable fit between the observed streamflow data (Table 5). These parameters were constrained using values suggested by Bicknell (2000) and initial parameter values were established using the HSPF Parameter Database (HSPFParm) (US EPA, 1999). Similar to Seong *et al.* (2015) the 'infilt' parameter was changed by a multiplier which retains difference between infiltration values across the different land use types. Unlike other studies (e.g. Seong *et al.*, 2015, Kim *et al.*, 2007), the nominal upper zone soil moisture storage term 'uzsn' was not allowed to vary monthly in order to reduce effects of colinearity between model parameters, increase numerical stability, and decrease non-uniqueness (equifinality) of calibrated parameter sets.

The analysis of the TUWmodel (HBV) model using the HSY-GSA method indicates that there is a range of sensitivities of parameters in the model(). The cumulative distribution of each parameter in the behavioral and non-behavioral sets are compared using the non-parametric Kolmogorov-Smirnov d statistic, which is used as an index of relative difference with higher d-values representing parameters that are more sensitive. A visualization of the behavioral and non-behavioral cumulative distribution curves provide a graphical representation of these results, with more sensitive parameters showing the most divergence between these two curves and the less sensitive parameters showing little to no change between these two groups (Figure 15). The Kolmogorov-Smirnov (K-S) statistic provides a quantitative accounting of this sensitivity analysis approach and shows agreement with the behavioral cumulative distribution curves. Both analyses suggest that the cperc and lsuz parameters are the most sensitive parameters for the Deerfield River Basin, followed by two flow recession parameters k1 and k0 and then several snow melt parameters including ‘SCF’ (snow correction factor), ‘DDF’ (degree day factor), ‘Tm’ (threshold melt temperature), followed by another soil moisture parameter ‘fc’ (field capacity, i.e. max soil moisture storage). These results correlate closely to other sensitivity analyses performed in the literature using this model (Harlin and Kung, 1992; Abebe *et al.*, 2010) providing support to this sensitivity analysis approach for this model. The HSY-GSA analysis and the K-S statistic was only applied to the TUWmodel (HBV) model to identify sensitive parameters while the literature was used to support the most sensitive parameters for the HSPF model.

Table 5: Selected sensitive parameters for each hydrological model. HBV parameters are the results of the most sensitive parameters from the HSY-GSA analysis. The lower and upper bounds for these parameters are represented as well as the final calibrated range between all of the subbasins.

	parameters	description	units	lower	range	upper
HSPF	**agwrc**	groundwater recession rate	(1/day)	0.85	(0.8503 - 0.9403)	0.999
	deepfr	fraction of infiltrating water lost to deep aquifers with the remaining fraction assigned to active groundwater storage	-	0	(0.00011-0.38796)	0.5
	infilt	index to mean soil infiltration rate	(in/hr)	0.001	(0.06745-0.49976)	0.5
	intfw	coefficient for the amount of water which enters ground from surface detention storage and becomes interflow	-	1	(1.8 - 9.998)	10
	irc	interflow recession coefficient	(1/day)	0.001	0.031	0.85
	kmelt	constant degree-day factor for the temp index snowmelt method	(in/d.F)	0	(0.03416 - 0.13815	none
	lzsn	lower zone nominal moisture storage	(in)	2	(2-10)	15
	uzsn	nominal upper zone soil moisture storage	(in)	0.05	(0.01-1.4)	2
HBV	**scf**	snow correction factor	-	0.9	(0.9-1.5)	1.5
	ddf	degree day factor	(mm/degC/day)	0	(1.11-2.78)	5
	tm	threshold temperature above which melt starts	(deg C)	-2	(-2-2)	2
	fc	field capacity, i.e. max soil moisture storage	(mm)	0	(5.6-600)	600
	k0	storage coefficient for very fast response	(day)	0	(0.555-2)	2
	k1	storage coefficient for fast response	(day)	2	(2.02-29.24)	30
	lsuz	threshold storage state, i.e. the very fast response start if exceeded	(mm)	1	(15.46-71.70)	100
	cperc	constant percolation rate	(mm/day)	0	(0-1.355)	8

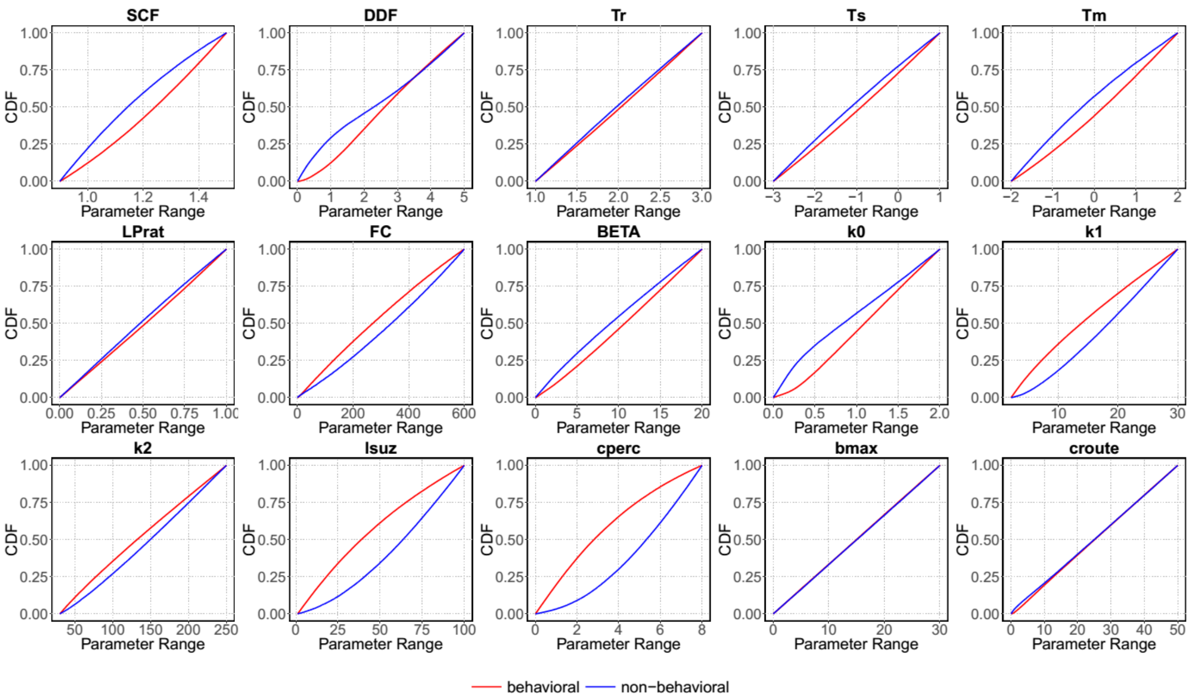


Figure 15: Results of the HSY-GSA for the 15 parameters of the TUWmodel.

Model performance on a daily time step was calculated with a split sample calibration and validation approach for both the HBV model and the HSPF model across the fifteen subbasins in the Deerfield River Basin using the period from 1981 to 1990 for calibration and from 1991 to 1995 for model validation (Table 6). The calibration performance over the ten-year period differed between the two models. HSPF tended to outperform the HBV across the subbasins and generally had higher KGE, R2, and NSE values. The model bias (PBIAS) was slightly lower across the subbasins for the HSPF model, although there was a slightly greater range as well for this model compared to the HBV model. The results from the model performance over the validation period showed slightly lower performance, as expected. However, the values are similar to the calibrated values indicating the models are not over-parameterized (parsimonious) and appropriate.

Table 6: Goodness-of-fit summary for HSPF and HBV calibration/validation.

	Calibration		Validation	
	HBV	**HSPF**	**HBV**	**HSPF**
KGE	0.67 *(0.58-0.73)*	0.78 *(0.71-0.86)*	0.66 *(0.41-0.74)*	0.73 *(0.67-0.86)*
R2	0.46 *(0.34-0.54)*	0.63 *(0.52-0.74)*	0.49 *(0.27-0.59)*	0.66 *(0.58-0.77)*
NSE	0.35 *(0.15-0.46)*	0.59 *(0.45-0.72)*	0.30 *(-0.21-0.46)*	0.53 *(0.37-0.74)*
NRMSE	0.80 *(0.73-0.92)*	0.64 *(0.52-0.74)*	0.83 *(0.73-1.10)*	0.68 *(0.51-0.80)*
PBIAS	-4.34 *(-12.1-4.1)*	-3.08 *(-16.9-8.4)*	1.04 *(-7.7-7.9)*	3.29 *(-23.9-11.7)*

Note: Values in parenthesis represent the minimum and maximum range of values across the subbasins. Calibration was performed from Jan 1, 1980 to Dec 31, 1990 (one-year ramp up period). Validation was performed from Jan 1, 1991 to Dec 31, 1995.

A Shapiro-Wilks test was performed on the RRM parameters as well as the catchment characteristics to assess the normality of these assumed to be independent variables. Standard log and square-root transformations were applied to variables that had a p-value of < 0.05. If the transformation increased the p-value from the Shapiro-Wilks test, it was used in the remainder of the analysis. However, there were several variables in which the transformations either reduced the normality of the variable or were ineffective for other reasons (e.g. domain included negative values). RRM parameters that could not be corrected by a standard transformation r for which the transformation reduced the normality (as estimated by the Shapiro-Wilks test) were noted (Table 7).

Table 7: Shapiro-Wilks test for normality for RRM parameters and catchment characteristics.

	Indep. Variable	p-value
HSPF Parameters	kmelt	*0.241*
	infilt	*0.585*
	lzsn_log	***0.005***
	agwrc	*0.233*
	deepfr_log	***0.016***
	intfw	*0.062*
	uzsn_log	*0.108*
	irc	*0.123*
HBV Parameters	SCF	*0.014*
	DDF	*0.870*
	Tm	*0.033*
	FC	*0.179*
	k0	*0.220*
	k1_log	*0.442*
	lsuz_log	*0.657*
	cperc_sqrt	*0.036*
Spatial	per_north	*0.267*
	per_east	*0.193*
	per_developed	*0.730*
	per_forest	*0.936*
	per_agr	*0.336*
	elev_MEAN	*0.693*
	slope_MEAN	*0.582*
	slope_std	*0.544*
	per_HGB	*0.190*
	per_HGC	*0.284*
	per_HGD_log	*0.748*
	TI_mn	*0.854*
	TI_max	*0.505*
	TI_min	*0.591*
	DA_log	*0.914*
	strm_len_log	*0.314*
Climate	ppt_log	*0.144*
	tmax	*0.732*
	tmean	*0.071*

Note: Values highlighted in bold indicate the paramters that could not be corrected using a log transformation. These parameters tended to bump into their upper/lower limits during calibration.

The correlation coefficients between the calibrated model parameters and the model catchment characteristics are represented in two matrixes, one for each model (Table 8 and Table 9). Prior to assessing these relationships, standard transformations were applied to the ‘cperc’, ‘k1’, and ‘lsuz’ parameters for the HBV model and the ‘deepfr’, ‘lzsn’, and ‘uzsn’ parameters for the HSPF model, to obtain a more normal distribution. The first principle component of the most significant catchment characteristics was selected using a PCA approach for use in regression for each RRM parameter. If there was only one significant catchment characteristic (or no statistically significant catchment characteristics), the catchment characteristic with the highest significance was used for regression.

Table 8: HSPF correlation coefficient values across catchment characteristics

	Pearson's r							
Catchment Characteristics	agwrc (1)	deepfr_log (2)	infilt (3)	intfw (4)	irc (5)	kmelt (6)	lzsn_log (7)	uzsn_log (8)
DA_km2_log	*-0.091*	*-0.100*	*0.069*	*-0.039*	*-0.325*	*-0.224*	*-0.062*	*0.078*
elev_MEAN	***-0.531***	*-0.478*	*-0.263*	***0.732***	***0.664***	*-0.007*	***-0.589***	***-0.684***
per_agr	***0.658***	*0.468*	*0.173*	***-0.656***	***-0.603***	*-0.146*	*0.366*	*0.397*
per_developed	*0.292*	*0.337*	*-0.097*	*-0.294*	***-0.703***	*-0.211*	*0.269*	*0.144*
per_east	***0.631***	*0.139*	*0.334*	*-0.352*	*0.017*	*0.078*	*0.356*	*0.219*
per_forest	*-0.287*	*0.010*	*-0.130*	*0.323*	***0.573***	***0.524***	*-0.049*	*-0.048*
per_HGB	*-0.181*	*0.269*	*-0.314*	*-0.121*	*0.028*	*0.356*	*-0.206*	*-0.106*
per_HGC	*-0.187*	*-0.051*	*0.132*	*-0.192*	*-0.340*	*-0.233*	*-0.289*	*-0.201*
per_HGD_log	***0.790***	*0.430*	*-0.110*	*-0.305*	*-0.133*	*0.277*	*0.484*	*0.279*
per_north	*-0.086*	*-0.093*	***0.408***	*-0.329*	*-0.279*	*0.161*	*-0.346*	*0.320*
ppt_log	***-0.571***	*-0.422*	*-0.235*	***0.724***	***0.690***	*-0.138*	***-0.619***	***-0.758***
slope_MEAN	*0.267*	*0.454*	*-0.004*	*-0.457*	*-0.494*	*0.365*	*0.431*	***0.628***
slope_std	*0.153*	*0.465*	*-0.063*	***-0.522***	*-0.414*	*0.431*	*0.392*	***0.643***
stream_length_km_log	*-0.251*	*-0.239*	*0.057*	*0.129*	*-0.225*	*-0.361*	*-0.204*	*-0.075*
TI_max	*-0.081*	*-0.182*	*-0.080*	*-0.096*	*-0.329*	*-0.227*	*-0.294*	*-0.059*
TI_min	*-0.387*	*-0.440*	*0.006*	*0.468*	*0.281*	*-0.354*	*-0.477*	*-0.513*
TI_mn	*-0.121*	*-0.350*	*-0.184*	*0.478*	***0.580***	*-0.032*	*-0.343*	***-0.566***
tmax	***0.536***	*0.473*	*0.239*	***-0.757***	***-0.647***	*0.088*	***0.572***	***0.725***
tmean	***0.538***	***0.518***	*0.154*	***-0.771***	***-0.621***	*0.293*	***0.573***	***0.810***

Notes: A threshold of r > 0.514 between calibrated HSPF model parameters and catchment characteristics were selected for regression. If this threshold was not satisfied, the relationship with the greatest r value was selected for regression. These results are highlighted in bold for each HSPF model parameter. Log transformations are indicated for both model parameters and catchment characteristics.

Table 9: HBV correlation coefficient values across catchment characteristics

Catchment Characteristics	Pearson's r							
	cperc_sqrt (1)	ddf (2)	fc (3)	k0 (4)	k1_log (5)	lsuz_log (6)	scf (7)	tm (8)
DA_km2_log	*-0.195*	*0.262*	*0.203*	*0.371*	*0.148*	*-0.034*	*0.210*	*0.300*
elev_MEAN	*-0.220*	*0.233*	*-0.423*	*-0.285*	*0.221*	*0.001*	*0.468*	*0.484*
per_agr	*0.069*	*-0.157*	*0.136*	*0.096*	*0.017*	*0.022*	*-0.283*	*-0.376*
per_developed	*-0.177*	*-0.136*	*0.030*	*-0.001*	*0.180*	*-0.093*	*0.093*	*-0.017*
per_east	***0.691***	*-0.076*	*0.084*	***0.528***	***-0.485***	*-0.343*	*-0.390*	*-0.362*
per_forest	*0.159*	*0.181*	*0.327*	*-0.017*	*-0.362*	*-0.014*	*-0.201*	*-0.109*
per_HGB	*-0.451*	*0.248*	*0.108*	*-0.057*	*0.206*	*-0.031*	*0.043*	*0.007*
per_HGC	*-0.136*	*-0.335*	*0.175*	*-0.043*	*0.048*	*-0.149*	*0.229*	*0.186*
per_HGD_log	*0.495*	***-0.391***	*0.251*	*-0.281*	*-0.291*	*-0.354*	*-0.336*	*-0.488*
per_north	*-0.267*	*0.022*	*-0.106*	*-0.145*	*0.320*	*0.235*	*0.139*	*-0.153*
ppt_log	*-0.326*	*0.176*	*-0.342*	*-0.248*	*0.258*	*0.208*	*0.445*	*0.550*
slope_MEAN	*0.088*	*0.071*	*0.503*	*0.193*	*-0.299*	*-0.005*	*-0.384*	*-0.441*
slope_std	*0.085*	*0.271*	***0.574***	*0.300*	*-0.287*	*-0.287*	*-0.480*	*-0.455*
stream_length_km_log	*-0.282*	*0.273*	*0.089*	*0.310*	*0.229*	*0.091*	*0.328*	*0.458*
TI_max	*-0.279*	*0.110*	*0.007*	*-0.056*	*0.319*	*-0.212*	*0.341*	*0.294*
TI_min	*-0.177*	*-0.108*	***-0.627***	*-0.311*	*0.136*	***0.414***	*0.332*	*0.278*
TI_mn	*0.043*	*0.255*	*-0.405*	*-0.079*	*0.108*	*-0.307*	*0.274*	*0.287*
tmax	*0.184*	*-0.206*	*0.391*	*0.244*	*-0.200*	*0.027*	*-0.476*	***-0.530***
tmean	*0.210*	*-0.122*	*0.473*	*0.176*	*-0.278*	*-0.073*	***-0.545***	***-0.648***

Notes: A threshold of r > 0.514 between calibrated HBV model parameters and catchment characteristics were selected for regression. If this threshold was not satisfied, the relationship with the greatest r value was selected for regression. These results are highlighted in bold for each HBV model parameter. Transformations are indicated for both model parameters and catchment characteristics.

In general, the HSPF had a greater number of significant catchment characteristics as well as higher correlation values with these significant parameters. The PCA approach was only needed for two of the HBV model parameters compared to five of the eight HSPF parameters. The percentage of variance explained for the RRM parameters in which the PCA approach was calculated was at least 70%. The percent of the catchment facing east and the minimum topographic index value in the catchment were most often related to HBV model parameters while the mean elevation and both annual average precipitation and temperature were most often correlated with HSPF parameters.

Residuals from the regressions for each parameter are mapped to their spatial location within the basin (Figure 16). While the sample size in this particular study is relatively small, there is no obvious clustering pattern occurring across the parameters. The only parameters that might demonstrate some level of patterning are the 'kmelt' HSPF parameter and the 'fc' HBV parameter. The 'kmelt' parameter seems to have more positive residuals in the lower part of the basin. However, when compared to the degree-day factor from the HBV model ('ddf'), there does not seem to be any similar patterning occurring. Also, the 'fc' parameter seems to display a negative tendency on the western side of the basin. However, there are a couple subbasins that are exceptions to this trend as well. It is difficult to infer from this sample size alone whether these possible clustering tendencies are significant, however we do gain some information as to the uncertainty of the regression regionalization from this analysis from both the positive negative bias and range of these residuals.

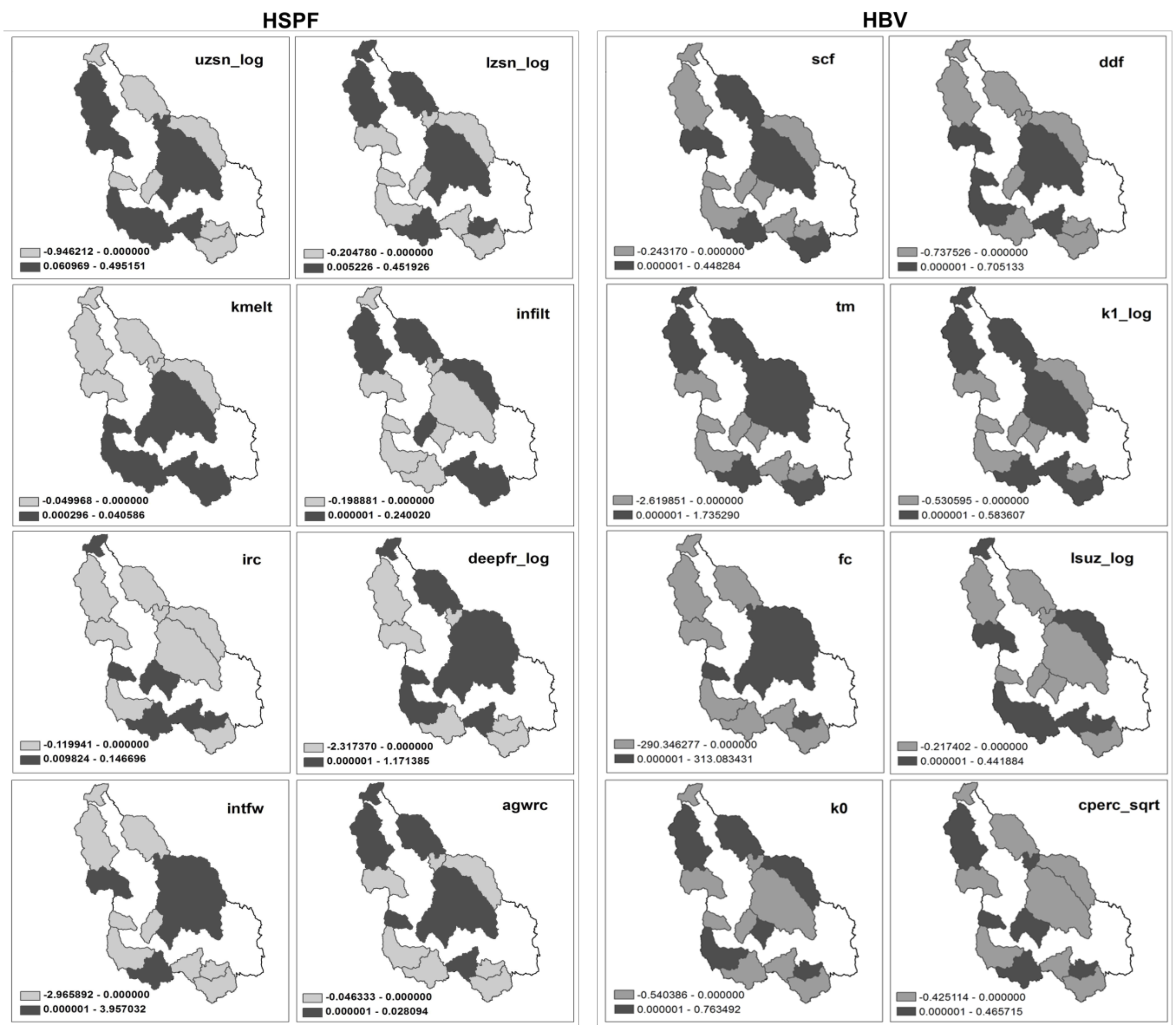


Figure 16: Spatial mapping of model residuals from the regionalization regressions by subbbasin.

The closed-form validation of the model parameters from the regression, proximity, and naïve mean regionalization methods are compared between models to evaluate the predicted RRM parameters in the calibrated parameter set (Figure 17). These values represent the average NRMSE between all the predicted and calibrated model parameters for each model. The regression method had the lowest NRMSE compared to the proximity method and the naïve mean method for both HBV and HSPF. There was greater variability in the NRMSE across the parameter predictions using the proximity method with the regression regionalization approach showing the least variance.

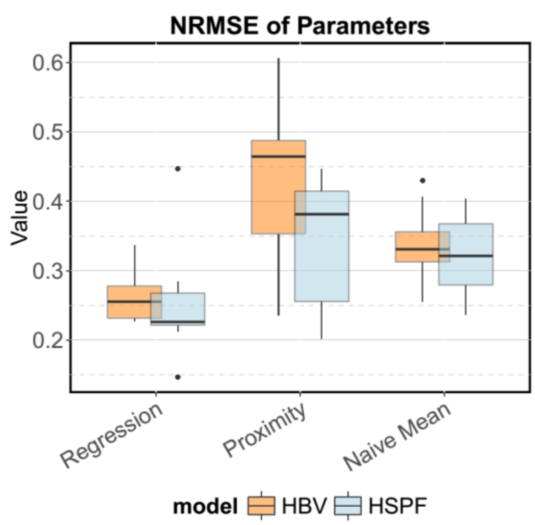


Figure 17: Closed-form validation results in predicting model parameters outside of running the RRMs. The regression method shows comparatively much less variance overall and lower NRMSE.

The open-form validation included running the RRMs with the estimated parameter set for the different regionalization methods. Multiple GOF criteria were used to gain a broader view of model performance using the regionalization methods (Figure 18). The calibrated model parameters are represented as the "Best" method and graphically indicate the spread of the performance across the fifteen subbasins for each of the RRMs. Again the HBV model generally seemed to perform less well across all the criteria compared to the HSPF model.

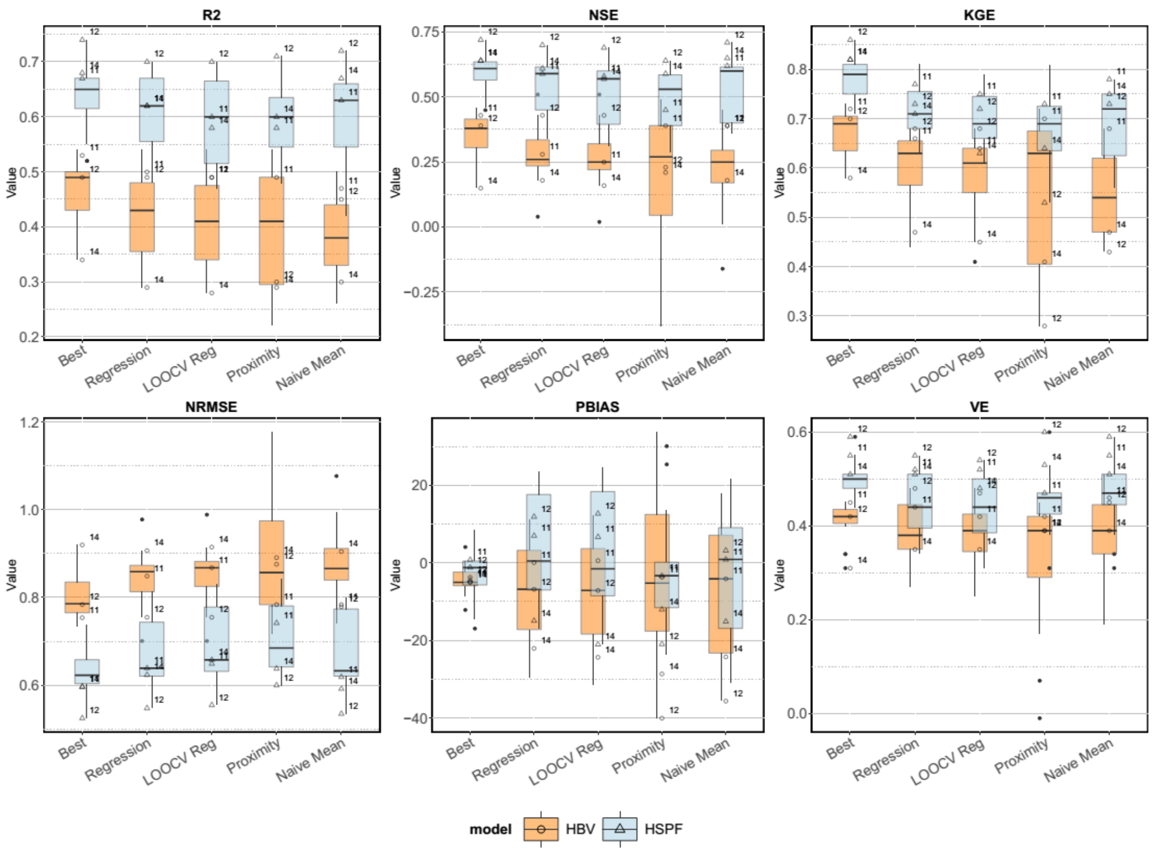


Figure 18: Open-form validation goodness-of-fit results across all subbasins for the HSPF and HBV models.

The regression method produces the best model results across most of the GOF performance criteria. The naïve mean method performs consistently better for the HSPF model while the proximity method often performs better than the naïve mean for the HBV model. The volumetric efficiency (VE) seemed to be increased when using the proximity and naïve mean method compared to the regression method.

Normalizing the GOF results from the regionalization method by the calibrated parameter set ('Best') yields a ratio wherein a value of unity suggests that the regionalization method performed the same as the calibrated model (Figure 19). In most cases, this ratio was less than one, as expected. However, in a few scenarios, the ratio was greater than one indicating that the estimated model parameters performed better than the calibrated model parameters. For the KGE GOF measure, this is a surprising result because the models were calibrated to this criterion. However it can interpreted from this analysis that there are distinct differences between these three GOF metrics and some of the estimated values may lead to improvements in one GOF measure while simultaneously decreasing another. While this is considered mostly an artifact of the differences between the GOF metrics, it appears that the HSPF model generally had less variation between the calibrated model simulations and the regionalization simulations. The regression method seemed to overall perform best for both models compared to the other approaches. Again, there is a wider distribution for the proximity method for both models across all the GOF measures.

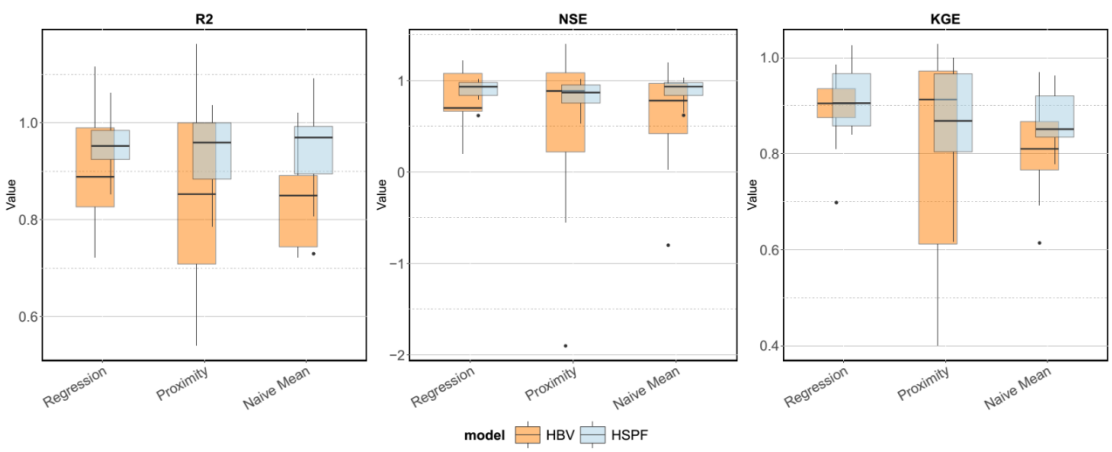


Figure 19: Ratios of goodness-of-fit (GOF) criteria between the calibrated “Best” paramter set and the different regionalization methods across all subbasins for the HSPF and HBV models.

2.6 Discussion

The HSY-GSA provided a qualitative and quantitative attempt to estimate the parameter uncertainty in the HBV model, similar to Harlin and Kung (1992). Literature has suggested shortcomings of this type of sensitivity analysis approach is the potential subjectivity in the determination of the behavioral and non-behavioral threshold in which the cumulative distributions of the model parameters are compared. The authors used the KGE criterion to measure sensitivity and attempted to minimize this subjectivity by setting a threshold that was the average of the KGE values across all of the Monte-Carlo simulations. This value was then adjusted 20% above and below this value and the KS statistics were compared across these three runs: 20% below 0.30, 0.30, and 20% above 0.30.

In ranking the model parameters based on the KS statistic for each of these three thresholds, the top 8 sensitive parameters did not change. Therefore, it was assumed that the most sensitive parameters to maximizing the KGE function were significantly accurate. Although this study did not address the differences in the parameter sensitivity from variations in the threshold value explicitly, unlike other RRM studies that have used this technique, it was at least considered in our analysis. Furthermore, the sensitive parameters for the HBV model found in this study agree with the ones found in other studies (Harlin and Kung, 1992; Abebe *et al*., 2010).

Several of the model parameters, both for HBV and HSPF, approached their upper or lower limits during the calibration process. For HSPF, the 'lzsn' and 'agwrc' parameters reached their lower limit and the 'infilt' and 'intfw' parameters approached the upper limits in several of the subbasins. For the HBV model, the parameters more often approached their respective upper and lower limits than the HSPF model. This suggests that the HBV model parameters are poorly represented in the model structure or that the homogeneous, lumped discretization of the subbasins contributed to the increased instability of the HBV model parameter space. The non-normal distribution of several of the RRM parameters may also be attributed to the tendency to approach upper and lower limits during the calibration process. While standard statistical transformations were applied to the model parameters prior to applying the regression regionalization approach, some transformations failed to help correct the normality of a few parameters (Table 7).

Overall, the HSPF model tended to perform better than the HBV model in the Deerfield over the calibration period with median KGE values of about 0.78 and 0.68, respectively. The HSPF model also had higher median R2 and NSE values over the calibration period of about 0.65 and 0.61, respectively with the HBV model having significantly lower median values of 0.58 and 0.33. The NRMSE error is lower for the HSPF model and the percent-bias (PBIAS) measure was closer to 0 on average for the HSPF model compared to HBV. In addition, volumetric efficiency (VE) was greater across the calibration and regionalization methods for the HSPF model compared to the

HBV model. The semi-distributed HSPF model seemed to perform better compared to the lumped-conceptual structure overall, suggesting this discretization difference might serve to better represent landscape heterogeneity as well as distribute the climate data more accurately.

In addition, there was less variance and fewer outliers in the regionalization GOF results for the HSPF model (Figure 18) compared to the HBV model. However, the regression method seemed to perform almost equally well for both RRMs with the KGE criterion (also used as the objective function for calibration), suggesting that the regression method was superior to the proximity and naïve mean methods overall across both RRMs. These results suggest that our method is quite useful and can be as reliable or more reliable compared to simpler methods.

Estimating the RRM parameter values using the proximity method resulted in both the highest average NRMSE as well as the greatest variance between the calibrated and predicted parameter sets (Figure 17). The proximity method consequently had the most variance in RRM simulations GOF measures and overall performed more poorly than the other two methods. For the HBV model, the naïve-mean method was equal or superior to the regression method for several of the subbasins (Figure 18). However, it also performed less well for the other subbasins suggesting that the wide variability in the mean NRMS (Figure 17) resulted in decreased performance in the RRM simulations (Figure 18). A mean NRMSE of about 0.26 for the regression regionalization method for both RRMs indicates that the regression method is again the more accurate means for estimating the RRM parameters and implies that this regionalization technique may be the preferred method for this particular region even through the uncertainty in the parameter estimations (Figure 16).

A comparison of Pearson's r (correlation coefficient) values between the RRMs indicate a distinct difference between models. The HSPF model had both more significant correlations to catchment characteristics as well as more significant correlations compared to the HBV model. While the mechanism for this is not clear, it might be indicative of the reduced resolution of the HBV model since this model is being used in a "lumped" approach. Because of the inherent heterogeneity of the subbasins, the model parameters may no longer be truly representative of the basin's physical processes. The lumping of land-use types, soil types, topography, and climate may serve to increase the noise between the lumped-model parameter set and the catchment characteristics. This concept has been noted previously and this study may serve to support those particular findings. In comparison to a more complex, semi-distributed model like HSPF, a reduction in the measure correlation between the model parameters and the catchment characteristics between these two RRMs is noted. However, these correlations are also most likely strongly influenced by calibration GOF, in which HSPF performed better than the HBV model, which generally had poorer performance. Climate stations were assigned to HSPF reaches (HRUs) independently based on proximity and infiltration was allowed to vary across the different land use types. In this way HSPF is considered semi-distributed in this study. However, as previously mentioned, it is difficult to identify why the correlations between the catchment characteristics were significantly less for the lumped-conceptual model (HBV) as compared to the semi-distributed (HSPF) model and most likely this result is a factor of both calibration fit and model structure.

In comparison to a more complex, semi-distributed model like HSPF, we do notice a reduction in the measure correlation between the model parameters and the catchment characteristics between these two RRMs. However, these correlations are also most likely strongly influenced by calibration GOF, in which HSPF performed better than the HBV model. Climate stations were assigned to HSPF reaches (HRUs) independently based on proximity and infiltration was allowed to vary across the different land use types. In this way HSPF is considered semi-distributed in this study. However, as previously mentioned, it is difficult to identify why the correlations between the catchment characteristics were significantly less for the lumped-conceptual model (HBV) as compared to the semi-distributed (HSPF) model and most likely this result is a factor of both calibration fit and model structure.

Dingman (1981) suggests that the primary influence on the hydrology of mountainous areas of New England is elevation; strong correlations observed in the HSPF model to the catchment characteristics support this notion. Dingman notes that "the effect of elevation is so dominant in the region that it can be used as the single independent variable in predicting many streamflow parameters" (Dingman, 1981). While elevation was not the only strongly correlated catchment characteristic, most of them (besides percent hydrologic soil group D and percent agricultural land use) were functions of the topography. Temperature variables as well as mean subbasin elevation were most often correlated with the HSPF parameters and both are complimentary of Dingman's hypothesis.

Limitations of lumped models include all the typical uncertainties associated with quality and availability of the input forcing data. One of the largest limitations of lumped-conceptual RRMs such as the HBV model is the inability to represent the true variability of the landscape and climate characteristics across the spatial domain. However, because lumped models tend to be parametrically parsimonious, they are also easier to calibrate and may even reduce issues with parameter identifiability and equifinality inherent to more complex distributed models (Beven, 2001; Kling and Gupta, 2009). In addition, because of the reduced representation of the landscape heterogeneity inherent to these models, attempts to derive significant relationships between catchment characteristics and lumped model parameters can become challenging (Kling and Gupta, 2009). This is of significance when these models are applied at ungauged locations, especially if regression based regionalization efforts are used towards this end.

In the case of HBV, there was less indication that elevation plays a major role in parameter estimation. This may be a product of two differences compared to the HSPF model: (1) calibrated parameters did not reproduce the daily hydrograph as well and (2) the model structure may have deconstructed the true process-based mechanisms of the HBV variables because of the lumped nature of the model discretization. The three most sensitive parameters ('cperc', 'lsuz', and 'k1') were all most correlated with the percent east-facing catchment attribute. Slope aspect can have a large role to play in forested catchments, affecting solar radiation, precipitation, wind speed, soil and air temperatures, snow accumulation, snowmelt, evapotranspiration, and vegetation type and growth (Chang, 2006). For example, in the northern hemisphere, forest transpiration is generally greater in northern than in southern slopes because of denser vegetative cover and deeper soils (Bethlamy, 1973). Also, west-facing forests tend to use more water than those on east-facing slopes (Chang, 2006). This attribute may also be indicative of the orientation of the subbasin, which can be important from a hydrological perspective. For example, because weather systems in mid-latitudes move from west to east, west-facing slopes and basins may receive more rain (Ward and Trimble., 2003). Ward and Trimble (2003) postulate that theoretically, an east-facing basin should have a more peaked hydrograph than a west-facing one, but there have been few investigations to conclusively support this idea. In the case of the HBV model, the correlation of these sensitive parameters, which include the constant percolation rate, recession rate parameter, and soil moisture parameter, with this potentially influential catchment attribute may support these concepts. However, percent east-facing was not correlated nearly as often to HSPF model parameters, serving to weaken Ward and Trimbles hypothesis.

2.7 Conclusions

Streamflow quantity and timing are essential components for the ecological integrity of river systems (Poff *et al.*, 1997) and are also vital for practical applications such as design of infrastructure, flood predictions, water supply and allocation, and climate impact analysis (Blöschl, 2013). Considering it is often difficult to measure these hydrologic processes directly, for example at ungauged locations such as road-stream crossings, rainfall runoff models (RRMs) are often used. When streamflow observations are not available to calibrate these models, the hydrologic regime can only be inferred from available physical and climatic characteristics of the catchments or by identifying hydrologically similar gauged catchments (Singh *et al.*, 2014). This study attempts to evaluate the effectiveness of different RRMs as well as assess a the accuracy of regression-based regionalization.

Two process-based RRMs that vary in complexity and structure are applied to fifteen subbasins in the Deerfield River Basin in the Northeastern US using indirect and direct streamgauge calibration. The Connecticut River UnImpacted Streamflow Estimation (CRUISE) tool (Archfield, 2013) is used to provide the streamflow data for twelve subbasins for the indirect calibration approach. Three USGS NWIS streamflow gauges within the Deerfield River Basin that are considered unimpaired are used to directly calibrate the RRMs. Calibration takes place over a ten year period (with one-year for warm-up) from January 1, 1980 to December 31, 1990 and validation of the models takes place over four years from January 1, 1991 to December 31, 1995. Goodness-of-fit (GOF) performance measures indicate an overall slight decrease in model performance between the calibration and validation model periods as expected, however the changes were not significant to indicate issues associated with over-parameterization.

PCA and OLS are used to develop a regression-based regionalization approach relating the sensitive parameters of RRMs with physical and climatic catchment characteristics. These regressions are used to predict the RRM parameters for the fifteen subbasins and are compared to the accuracy of using the two simpler regionalization methods, closest proximity and naïve-mean, in order to determine the effectiveness of this regionalization approach.

The comparison of the regionalization approaches suggest that the more complex regression approach used in this study was able to more accurately estimate the RRM parameters. In general, the NRMSE error of the regression method was noticeably lower than the proximity and the naïve-mean methods. When the models were run with the predicted RRM parameters from the different regionalization methods, the regression method seemed to provide the best results compared to the calibrated parameters. The proximity method showed the highest variation in model performance while the naïve-mean seemed to generally be lower on average than the other two methods.

Overall, the HSPF model tended to perform better as compared to the HBV model based on an evaluation of multiple GOF criteria. This may be attributed to several factors, including differences in model structure as well as a loss of accuracy associated with lumping climatic drivers to the HBV model. The semi-distributed nature of the HSPF model may lead to more accurate representations of the physical processes that drive surface runoff in the Deerfield River Basin. This hypothesis is supported in the calibration results for this model. However, because of the uncertainties associated with the modeling process for both HSPF and HBV, in addition to the lack of available data in the Deerfield Basin, it is difficult to more than surmise that either model is representing the physical processes at a basin scale accurately.

Wise stewardship of water and the environment requires a variety of predictive tools that can generate predictions of hydrological responses over a range of space-time scales and climates and is necessary for the sustainable management of river basins, integrating economic, social, and environmental perspectives (Sivapalan *et al.*, 2003). Because so many of our rivers across the globe are ungauged, it is necessary to evaluate the accuracy of different models and methods for making predictions of runoff at these ungauged locations. This study provides a framework for assessing the accuracy of estimating the parameters of process-based RRMs in the region and also provides an assessment of RRM performance as applied within a small forested northeastern U.S. catchment.

2.8 References

Abdulla, F. A., & Lettenmaier, D. P. (1997). Development of regional parameter estimation equations for a macroscale hydrologic model. Journal of Hydrology, 197(1-4), 230-257.

Abebe, N. A., Ogden, F. L., & Pradhan, N. R. (2010). Sensitivity and uncertainty analysis of the conceptual HBV rainfall–runoff model: Implications for parameter estimation. Journal of Hydrology, 389(3), 301-310.

Archfield, S. A., Steeves, P. A., Guthrie, J. D., & Ries III, K. G. (2013). Towards a publicly available, map-based regional software tool to estimate unregulated daily streamflow at ungauged rivers. Geoscientific Model Development, 6(1), 101-115.

Baloch, M. A., Ames, D. P., & Tanik, A. (2011). Application of BASINS/HSPF in the Koycegiz–Dalyan watershed in Turkey: a developing country case study in watershed modeling. River Basin Management VI, 146, 187.

Bárdossy, A. (2007). Calibration of hydrological model parameters for ungauged catchments. Hydrology and Earth System Sciences Discussions, 11(2), 703-710.

Beck, M. B. (1987). Water quality modeling: a review of the analysis of uncertainty. Water Resources Research, 23(8), 1393-1442.

Bergman, M. J., & Donnangelo, L. J. (2000). SIMULATION OF FRESHWATER DISCHARGES FROM UNGAGED AREAS TO THE SEBASTIAN RIVER, FLORIDA1.

Bergström, S. (2006). Experience from applications of the HBV hydrological model from the perspective of prediction in ungauged basins. IAHS publication, 307, 97.

Bergström, S. (1976). Development and application of a conceptual runoff model for Scandinavian catchments.

Bergström, S., & Lindström, G. (2015). Interpretation of runoff processes in hydrological modelling—experience from the HBV approach. Hydrological Processes, 29(16), 3535-3545.

Bergström, S., Harlin, J. & Lindström, G. (1992) Spillway design floods in Sweden. I. New guidelines. Hydrol. Sci. J.37(5), 10/1992, 505–519

Bethlahmy, N. (1973). Water yield, annual peaks and exposure in mountainous terrain. *Journal of Hydrology*, *20*(2), 155-169.

Beven, K. (1989). Changing ideas in hydrology—the case of physically-based models. Journal of hydrology, 105(1-2), 157-172.

Beven, K. (2001). How far can we go in distributed hydrological modelling?. Hydrology and Earth System Sciences Discussions, 5(1), 1-12.

Beven, K. (2002). Towards a coherent philosophy for modelling the environment. In Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences (Vol. 458, No. 2026, pp. 2465-2484). The Royal Society.

Beven, K. J. (2011). Rainfall-runoff modelling: the primer. John Wiley & Sons.

Beven, K., Smith, P., Westerberg, I., & Freer, J. (2012). Comment on "Pursuing the method of multiple working hypotheses for hydrological modeling" by P. Clark *et al.* Water Resources Research, 48(11).

Bhatia, P. K., Bergström, S., & Persson, M. (1984). Application of the distributed HBV-6 model to the upper Narmada basin in India. Swedish Meteorological and Hydrological Institute.

BICKNELL, B. R. (2000). Basins Technical Note 6: Estimating Hydrology and Hydraulic Parameters for HSPF. US: Environmental Protection Agency.

Bicknell, B. R., Imhoff, J. C., Kittle Jr, J. L., Donigian Jr, A. S., & Johanson, R. C. (1996). Hydrological simulation program-FORTRAN. user's manual for release 11. US EPA.

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

Blöschl, G., & Sivapalan, M. (1995). Scale issues in hydrological modeling: a review. Hydrological processes, 9(3-4), 251-290.

Chang, M. (2006). *Forest hydrology: an introduction to water and forests*. CRC press.

Clark, M. P., Kavetski, D., & Fenicia, F. (2011b). Pursuing the method of multiple working hypotheses for hydrological modeling. Water Resources Research, 47(9).

Clark, M. P., McMillan, H. K., Collins, D. B., Kavetski, D., & Woods, R. A. (2011a). Hydrological field data from a modeller's perspective: Part 2: process-based evaluation of model hypotheses. Hydrological processes, 25(4), 523-543.

Crawford, N. H., & Linsley, R. K. (1966). Digital Simulation in Hydrology'Stanford Watershed Model 4.

Dingman, S. L. (1981). Elevation: a major influence on the hydrology of New Hampshire and Vermont, USA/L'altitude exerce une influence importante sur l'hydrologie du New Hampshire et du Vermont, Etats-Unis. *Hydrological Sciences Journal*, *26*(4), 399-413.

Doherty, J., & Johnston, J. M. (2003). Methodologies for calibration and predictive analysis of a watershed model.

Duan, Q., Schaake, J., Andreassian, V., Franks, S., Goteti, G., Gupta, H. V., ... & Hogue, T. (2006). Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. Journal of Hydrology, 320(1), 3-17.

Duan, Q., Sorooshian, S., & Gupta, V. K. (1994). Optimal use of the SCE-UA global optimization method for calibrating watershed models. Journal of hydrology, 158(3), 265-284.

Duda, P. B., Hummel, P. R., Donigian Jr, A. S., & Imhoff, J. C. (2012). BASINS/HSPF: Model use, calibration, and validation. Transactions of the ASABE, 55(4), 1523-1547.

Falcone, J. A., Carlisle, D. M., Wolock, D. M., & Meador, M. R. (2010). GAGES: A stream gage database for evaluating natural and altered flow conditions in the conterminous United States: Ecological Archives E091 -045. Ecology, 91(2), 621-621.

Federer, C. A., Vörösmarty, C., & Fekete, B. (1996). Intercomparison of methods for calculating potential evaporation in regional and global water balance models. Water Resources Research, 32(7), 2315-2321.

Fenneman, N. M. (1938). Physiography of eastern United States: New York, McGraw-Hill Book Co.

Filoso, S., Vallino, J., Hopkinson, C., Rastetter, E., & Claessens, L. (2004). MODELING NITROGEN TRANSPORT IN THE IPSWICH RIVER BASIN, MASSACHUSETTS, USING A HYDROLOGICAL SIMULATION PROGRAM IN FORTRAN (HSPF) 1.

Flödeskommittén (1990) Slutrapport från Flödeskommittén. (Final report from the Swedish committee on spillway design). Swedish State Power Board, Swedish Power Association and SMHI, Norrköping (in Swedish).

Friesz, P. J. (1996). Geohydrology of stratified drift and streamflow in the Deerfield River basin, northwestern Massachusetts. US Department of the Interior, US Geological Survey.

Gan, T. Y., Dlamini, E. M., & Biftu, G. F. (1997). Effects of model complexity and structure, data quality, and objective functions on hydrologic modeling. *Journal of Hydrology*, *192*(1), 81-103.

Gao, W., Guo, H. C., & Liu, Y. (2014). Impact of Calibration Objective on Hydrological Model Performance in Ungauged Watersheds. Journal of Hydrologic Engineering, 20(8), 04014086.

Gay, F. B., Toler, L. G., & Hansen, B. P. (1974). Hydrology and water resources of the Deerfield River basin, Massachusetts (No. 506).

Goswami, M., O'connor, K. M., & Bhattarai, K. P. (2007). Development of regionalisation procedures using a multi-model approach for flow simulation in an ungauged catchment. Journal of Hydrology, 333(2), 517-531.

Grayson, R. B., Moore, I. D., & McMahon, T. A. (1992). Physically based hydrologic modeling: 1. A terrain-based model for investigative purposes. Water resources research, 28(10), 2639-2658.

Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1998). Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research*, *34*(4), 751-763.

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology, 377(1), 80-91.

Gutiérrez-Magness, A. L., & McCuen, R. H. (2005). Effect of flow proportions on HSPF model calibration accuracy. Journal of Hydrologic Engineering, 10(5), 343-352.

Häggström, M., Lindström, G., Cobos, C., Martinez, J. R., Merlos, L., Alonzo, R. D., ... & Alfaro, R. I. (1990). Application of the HBV model for flood forecasting in six Central American rivers (p. 73). Norrköping: SMHI.

Hamon, W. R. (1963). Computation of direct runoff amounts from storm rainfall. publisher not identified.

Harlin, J., & Kung, C. S. (1992). Parameter uncertainty and simulation of design floods in Sweden. Journal of hydrology, 137(1), 209-230.

Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J.D., and Megown, K., 2015, Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information. Photogrammetric Engineering and Remote Sensing, v. 81, no. 5, p.

Hornberger, G. M., & Spear, R. C. (1981). Approach to the preliminary analysis of environmental systems. J. Environ. Manage.;(United States), 12(1).

Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., ... & Fenicia, F. (2013). A decade of Predictions in Ungauged Basins (PUB)—a review. Hydrological sciences journal, 58(6), 1198-1255.

Hundecha, Y., & Bárdossy, A. (2004). Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model. Journal of Hydrology, 292(1), 281-295.

Iskra, I., & Droste, R. (2007). Application of non-linear automatic optimization techniques for calibration of HSPF. Water environment research, 647-659.

Jakeman, A. J., & Hornberger, G. M. (1993). How much complexity is warranted in a rainfall-runoff model?. Water resources research, 29(8), 2637-2649.

Johnson, M. S., Coon, W. F., Mehta, V. K., Steenhuis, T. S., Brooks, E. S., & Boll, J. (2003). Application of two hydrologic models with different runoff mechanisms to a hillslope dominated watershed in the northeastern US: a comparison of HSPF and SMR. Journal of Hydrology, 284(1), 57-76.

Kay, A. L., Jones, D. A., Crooks, S. M., Kjeldsen, T. R., & Fung, C. F. (2007). An investigation of site-similarity approaches to generalisation of a rainfall? runoff model. Hydrology and Earth System Sciences Discussions, 11(1), 500-515.

Kim, S. M., Benham, B. L., Brannan, K. M., Zeckoski, R. W., & Doherty, J. (2007). Comparison of hydrologic calibration of HSPF using automatic and manual methods. Water resources research, 43(1).

Kim, U., & Kaluarachchi, J. J. (2008). Application of parameter estimation and regionalization methodologies to ungauged basins of the Upper Blue Nile River Basin, Ethiopia. *Journal of Hydrology*, *362*(1), 39-56.

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).

Kling, H., & Gupta, H. (2009). On the development of regionalization relationships for lumped watershed models: The impact of ignoring sub-basin scale variability. Journal of Hydrology, 373(3), 337-351.

Knox, C. E., & Nordenson, T. J. (1955). Average annual runoff and precipitation in the New England-New York area (No. 7).

Koren, V., Smith, M., & Duan, Q. (2003). Use of a priori parameter estimates in the derivation of spatially consistent parameter sets of rainfall-runoff models. Calibration of Watershed Models, 239-254.

Lindström, G., Rosberg, J. & Arheimer, B. (2005) Parameter precision in the HBV-NP model and impacts on nitrogen scenario simulations in the Rönneå River, southern Sweden. Ambio 34(7), 533–537

Linhart, S. M., Nania, J. F., Christiansen, D. E., Hutchinson, K. J., Sanders Jr, C. L., & Archfield, S. A. (2013). *Comparison between two statistically based methods, and two physically based models developed to compute daily mean streamflow at ungaged locations in the Cedar River Basin, Iowa* (No. 2013-5111). US Geological Survey.

Lu, J., Sun, G., McNulty, S. G., & Amatya, D. M. (2005). A comparison of six potential evapotranspiration methods for regional use in the southeastern United States1.

Madsen, H. (2003). Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. Advances in water resources, 26(2), 205-216.

Marshall, E., & Randhir, T. (2008). Effect of climate change on watershed system: a regional analysis. *Climatic Change*, *89*(3-4), 263-280.

McCabe, G. J., Hay, L. E., Bock, A., Markstrom, S. L., & Atkinson, R. D. (2015). Inter-annual and spatial variability of Hamon potential evapotranspiration model coefficients. Journal of Hydrology, 521, 389-394.

McDonnell, J. J., Sivapalan, M., Vaché, K., Dunn, S., Grant, G., Haggerty, R., ... & Selker, J. (2007). Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology. Water Resources Research, 43(7).

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

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

Mishra, A. K., and Coulibaly, P. (2009).“Development in hydrometric networks design: A review.”Rev. Geophys., 47(2), RG2001

Norstedt, U., Brandesten, C.-O., Bergström, S., Harlin, J. & Lindströrn, G. (1992) Re-evaluation of hydrological dam safety in Sweden. International Water Power and Dam Construction, June 1992.

Orth, R., Staudinger, M., Seneviratne, S. I., Seibert, J., & Zappa, M. (2015). Does model performance improve with complexity? A case study with three hydrological models. *Journal of Hydrology*, *523*, 147-159.

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., & Viglione, A. (2012). TUWmodel: Lumped hydrological model developed at the Vienna University of Technology for education purposes, R package version 0.1-2.

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

Parajka, J., Merz, R., & Blöschl, G. (2005). A comparison of regionalisation methods for catchment model parameters. Hydrology and earth system sciences discussions, 9(3), 157-171.

Parr, D., & Wang, G. (2014). Hydrological changes in the US Northeast using the Connecticut River Basin as a case study: Part 1. Modeling and analysis of the past. *Global and Planetary Change*, *122*, 208-222.

Parr, D., Wang, G., & Ahmed, K. F. (2015). Hydrological changes in the US Northeast using the Connecticut River Basin as a case study: Part 2. Projections of the future. *Global and Planetary Change*, *133*, 167-175.

Parr, D., Wang, G., & Bjerklie, D. (2015). Integrating Remote Sensing Data on Evapotranspiration and Leaf Area Index with Hydrological Modeling: Impacts on Model Performance and Future Predictions. *Journal of Hydrometeorology*, *16*(5), 2086-2100.

Perrin, C., Michel, C., & Andréassian, V. (2001). Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. *Journal of Hydrology*, *242*(3), 275-301.

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

PRISM Climate Group. (2004). Oregon State University, http://prism.oregonstate.edu, created 4 Feb 2004.

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

Refsgaard, J. C., & Knudsen, J. (1996). Operational validation and intercomparison of different types of hydrological models. *Water Resources Research*, *32*(7), 2189-2202.

Saleh, A., & Du, B. (2004). Evaluation of SWAT and HSPF within BASINS program for the upper North Bosque River watershed in central Texas. Transactions of the ASAE, 47(4), 1039.

Samuel, J., Coulibaly, P., and Metcalfe, R. A. (2011).“Estimation of continuous streamflow in Ontario ungauged basins: Comparison of regionalization methods.”J. Hydrol. Eng., 16(5), 447–459.

Seibert, J. (1999). Regionalisation of parameters for a conceptual rainfall-runoff model. Agricultural and forest meteorology, 98, 279-293.

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

Seong, C., Her, Y., & Benham, B. L. (2015). Automatic calibration tool for Hydrologic Simulation Program-FORTRAN using a shuffled complex evolution algorithm. Water, 7(2), 503-527.

Singh, V. P., & Woolhiser, D. A. (2002). Mathematical modeling of watershed hydrology. Journal of hydrologic engineering, 7(4), 270-292.

Sivapalan, M. (2005). Pattern, process and function: elements of a unified theory of hydrology at the catchment scale. Encyclopedia of hydrological sciences.

Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., ... & Oki, T. (2003). IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences. Hydrological sciences journal, 48(6), 857-880.

Sivapalan, M. (2003a). Process complexity at hillslope scale, process simplicity at the watershed scale: is there a connection?. *Hydrological Processes*, *17*(5), 1037-1041.

Skøien, J. O., Merz, R., & Blöschl, G. (2006). Top-kriging-geostatistics on stream networks. *Hydrology and Earth System Sciences Discussions*, *10*(2), 277-287.

Srinivasan, M. S., Hamlett, J. M., Day, R. L., Sams, J. I., & Petersen, G. W. (1998). HYDROLOGIC MODELING OF TWO GLACIATED WATERSHEDS IN NORTHEAST PENNSYLVANIA1.

Steinschneider, S., Yang, Y. C. E., & Brown, C. (2015). Combining regression and spatial proximity for catchment model regionalization: a comparative study. Hydrological Sciences Journal, 60(6), 1026-1043.

Stoll, S., & Weiler, M. (2010). Explicit simulations of stream networks to guide hydrological modelling in ungauged basins. Hydrology and Earth System Sciences, 14(8), 1435-1448.

Taner, M. Ü., Carleton, J. N., & Wellman, M. (2011). Integrated model projections of climate change impacts on a North American lake. Ecological Modelling, 222(18), 3380-3393.

US EPA, 1999. HSPFParm: An Interactive Database of HSPF Model Parameters, Version 1.0. EPA-823-R-99-004. U.S. Environmental Protection Agency, Office of Water,Washington, DC. Available from the BASINS web site,http://www.epa.gov/ost/basins/support.htm.

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

Viglione, Alberto and Parajka, Juraj (2014). TUWmodel: Lumped Hydrological Model for Education Purposes. R package version 0.1-4.

Vleeschouwer, N. D., & Pauwels, V. R. (2013). Assessment of the indirect calibration of a rainfall-runoff model for ungauged catchments in Flanders. Hydrology and Earth System Sciences, 17(5), 2001-2016.

Vörösmarty, C. J., Federer, C. A., & Schloss, A. L. (1998). Potential evaporation functions compared on US watersheds: Possible implications for global-scale water balance and terrestrial ecosystem modeling. Journal of Hydrology, 207(3), 147-169.

Wagener, T., & Wheater, H. S. (2006). Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty. Journal of Hydrology, 320(1), 132-154.

Ward, A. D., & Trimble, S. W. (2003). *Environmental hydrology*. CRC Press.

Wetterhall, Fredrik. (2014). "HBV – The most famous hydrological model of all? An interview with its father: Sten Bergström." Retrieved from the HEPEX webpage on April 7, 2016 from <http://hepex.irstea.fr/the-hbv-model-40-years-and-counting/>.

Wu, J., & Zhu, X. (2006). Using the shuffled complex evolution global optimization method to solve groundwater management models. In Frontiers of WWW Research and Development-APWeb 2006 (pp. 986-995). Springer Berlin Heidelberg.

Young, A. R. (2006)."Streamflow simulation within UK ungauged catchments using a daily rainfall-runoff model." J. Hydrol., 320(1–2),155–172

Young, P. (1983). The validity and credibility of models for badly defined systems. In Uncertainty and forecasting of water quality (pp. 69-98). Springer Berlin Heidelberg.

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

Zhang, X., & Lindström, G. (1997). Development of an automatic calibration scheme for the HBV hydrological model. Hydrological Processes, 11(12), 1671-1682.

Zhang, Y., & Chiew, F. H. (2009). Relative merits of different methods for runoff predictions in ungauged catchments. Water Resources Research, 45(7).

Appendix

Not included in paper, but included to supplement writing.

3.1 Tables

Location	Station Name	Latitude	Longitude	Record Start	Record End
MA190120	Amherst	42.3833	-72.5333	5/1/1948	3/31/2006
VT430277	Ball Mountain Lake	43.1167	-72.8000	1/1/1970	12/31/2005
VT436761	Readsboro 1 SE	42.7500	-72.9333	2/1/1951	3/31/1998
VT437152	Searsburg Station	42.8667	-72.9167	1/1/1970	12/31/2005

Mean NRMSE of RRM Parameters		
Method	HBV	HSPF
Regression	*0.26*	*0.25*
Proximity	*0.43*	*0.34*
Naïve Mean	*0.34*	*0.32*

3.1.1 HSPF Modeling

	parameters	description	units	lower	range	upper
constant parameters	**snowcf**	multi factor for poor gage catch efficiency	-	1	1.3	2
	covind	max SWE depth at which entire land segment is covered in snow	(in)	0.1	6	10
	tbase	refernce temp for the temp index method	(F)	0	32	60
	rdcsn	density of new snow relative to water when temp < F	-	0.05	0.2	0.3
	tsnow	wet bulb air temp at which snow forms	(deg F)	30	32	40
	snowevp	factor to adjust sublimination from snowpack	-	0	0.1	0.5
	ccfact	factor to adjust rate of heat transfer from atm. to snowpack	-	0.5	1	8
	mwater	max liquid water holding capacity in snowpack	(in/in)	0.005	0.03	0.2
	mgmelt	maximim rate of snowmelt by ground heat	(in/day)	0	0.01	0.1
	kvary	groundwater recession flow parameter to describe non-linear groundwater recession rate	(1/in)	0	0	5
	petmax	temp below which ET will be reduced to 50% of that in the input time series	(deg F)	32	40	48
	petmin	temp at and below which ET will be zero	(deg F)	35	35	40
	infexp	exponent that determines deviation from mominal lower zone storage affects infiltration rate	-	1	2	3
	infild	ratio of max and mean soil infiltration capacities	-	1	2	3
	basetp	fraction of ET from riparian vegitation as active groundwater enters streambed	-	0	0.02	0.2
	agwetp	fraction of remaining PET that can be met from active groundwater storage	-	0	0	0.2
	cepsc	amount of rainfall retained by vegetation and never reaches land surface	(in)	0.01	0	0.4
	nsur	Manning's n for overland flow plane	-	0.05	0.35	0.5
	lzetp	index to lower zone ET	-	0.1	0.7	0.9

locationID	HSPF Calibration - Jan 01, 1980 to Dec 31 1990								HSPF Validation - Jan 1, 1991 to Dec 31, 1995							
	RMSE	**NRMSE**	**PBIAS**	**R2**	**NSE**	**KGE**	**VE**	**d**	**RMSE**	**NRMSE**	**PBIAS**	**R2**	**NSE**	**KGE**	**VE**	**d**
0	*22.78*	*62.3*	*8.4*	*0.65*	*0.61*	*0.79*	*0.53*	*0.89*	*19.91*	*68.4*	*9.6*	*0.65*	*0.53*	*0.74*	*0.52*	*0.89*
1	*49.61*	*72.1*	*0.1*	*0.55*	*0.48*	*0.74*	*0.45*	*0.86*	*40.59*	*78.5*	*4.5*	*0.59*	*0.38*	*0.67*	*0.43*	*0.86*
2	*28.92*	*61*	*3.4*	*0.67*	*0.63*	*0.81*	*0.5*	*0.9*	*24.29*	*69.5*	*9.2*	*0.68*	*0.52*	*0.7*	*0.48*	*0.89*
3	*66.64*	*61.1*	*-5.2*	*0.67*	*0.63*	*0.81*	*0.49*	*0.9*	*57.02*	*71.2*	*6.4*	*0.65*	*0.49*	*0.71*	*0.46*	*0.89*
4	*82.36*	*64.5*	*-0.5*	*0.63*	*0.58*	*0.8*	*0.48*	*0.89*	*70.07*	*73.7*	*2*	*0.62*	*0.46*	*0.71*	*0.43*	*0.87*
5	*33.49*	*61.1*	*0.8*	*0.66*	*0.63*	*0.81*	*0.51*	*0.9*	*30.46*	*74.9*	*6.2*	*0.65*	*0.44*	*0.67*	*0.46*	*0.88*
6	*29.85*	*64.5*	*-0.1*	*0.63*	*0.58*	*0.79*	*0.49*	*0.89*	*28.44*	*79.7*	*3.9*	*0.58*	*0.37*	*0.67*	*0.39*	*0.86*
7	*74.89*	*63.9*	*-6.2*	*0.62*	*0.59*	*0.78*	*0.51*	*0.88*	*55.91*	*61.3*	*4.2*	*0.72*	*0.62*	*0.78*	*0.48*	*0.91*
8	*142.46*	*67.1*	*-14.5*	*0.61*	*0.55*	*0.74*	*0.48*	*0.88*	*105.5*	*58.4*	*-6.8*	*0.72*	*0.66*	*0.81*	*0.49*	*0.92*
9	*19.28*	*73.9*	*-4.3*	*0.52*	*0.45*	*0.72*	*0.44*	*0.84*	*13.32*	*65.7*	*8.7*	*0.69*	*0.57*	*0.74*	*0.45*	*0.9*
10	*126.32*	*73.4*	*-6.4*	*0.52*	*0.46*	*0.71*	*0.46*	*0.84*	*88.74*	*66.4*	*6.1*	*0.67*	*0.56*	*0.75*	*0.48*	*0.9*
11	*83.34*	*59.6*	*0.8*	*0.67*	*0.64*	*0.82*	*0.55*	*0.9*	*70.77*	*64.8*	*11.7*	*0.69*	*0.58*	*0.74*	*0.51*	*0.9*
12	*193.59*	*52.5*	*-1.2*	*0.74*	*0.72*	*0.86*	*0.59*	*0.92*	*142.68*	*51.4*	*2*	*0.77*	*0.74*	*0.86*	*0.59*	*0.93*
13	*26.04*	*58.4*	*-16.9*	*0.7*	*0.66*	*0.76*	*0.5*	*0.91*	*25.66*	*67.2*	*-23.9*	*0.62*	*0.55*	*0.68*	*0.43*	*0.88*
14	*55.48*	*59.7*	*-4.4*	*0.68*	*0.64*	*0.82*	*0.51*	*0.9*	*48.84*	*69*	*5.5*	*0.65*	*0.52*	*0.74*	*0.48*	*0.89*

Notes: The calibration period includes a one-year warm up period. For validation, the model is run through the calibration period and then the goodness-of-fit measures are assessed from Jan 1, 1991 to Dec 31, 1995. All sub-basins were used in the regionalization regression process.

1029

Shapiro-Wilks Test for Normality		
	Indepent. Variable	p-value
HSPF Parameters	kmelt	*0.241*
	infilt	*0.585*
	lzsn_log	***0.005***
	agwrc	*0.233*
	deepfr_log	***0.016***
	intfw	*0.062*
	uzsn_log	*0.108*
	irc	*0.123*
Spatial	per_north	*0.267*
	per_east	*0.193*
	per_developed	*0.730*
	per_forest	*0.936*
	per_agr	*0.336*
	elev_MEAN	*0.693*
	slope_MEAN	*0.582*
	slope_std	*0.544*
	per_HGB	*0.190*
	per_HGC	*0.284*
	per_HGD_log	*0.748*
	TI_mn	*0.854*
	TI_max	*0.505*
	TI_min	*0.591*
	DA_log	*0.914*
	strm_len_log	*0.314*
Climate	ppt_log	*0.144*
	tmax	*0.732*
	tmean	*0.071*

Note: Values highlighted in bold indicate the paramters that could not be corrected using a log transformation. These parameters tended to bump into their upper/lower limits during calibration.

Results for the HSPF regionalization regressions								
Parameter	agwrc (1)	intfw (2)	irc (3)	lzsn_log (4)	uzsn_log (5)	deepfr_log (6)	infilt (7)	kmelt (8)
agwrc PCA comp.	-0.009*** (0.003)							
intfw PCA comp.		1.062*** (0.233)						
irc PCA comp.			0.035*** (0.009)					
lzsn.log PCA comp.				-0.060** (0.022)				
uzsn.log PCA comp.					-0.217*** (0.049)			
tmean						0.635** (0.291)		
per.north							1.780 (1.104)	
per.forest								0.348** (0.157)
Constant	0.892*** (0.006)	6.549*** (0.489)	0.267*** (0.020)	0.464*** (0.042)	-0.588*** (0.111)	-5.557*** (1.808)	0.102 (0.101)	-0.230 (0.135)
N	0.47	0.616	0.556	0.36	0.603	0.268	0.167	0.275
R2	0.43	0.586	0.522	0.311	0.572	0.212	0.103	0.219
Adjusted R2	0.022	1.895	0.077	0.162	0.43	1.146	0.122	0.025
Residual Std. Error (df = 13)	11.549***	20.832***	16.284***	7.328**	19.723***	4.759**	2.601	4.924**
F Statistic (df = 1; 13)	5.107**	15.011***	10.236***	6.908**	11.394***	4.759**	2.601	4.924**
Notes:	Parenthesis represent standard error of estimate. The first PCA component is used for the agwrc, intfw, irc, lzsn, and uzsn regressions. ***Significant at the 1 percent level; **Significant at the 5 percent level; *Significant at the 10 percent level.							

LOOCV - Comparison to Calibration Period Goodness-of-fit Measures					
locationID	**NSE ratio**	**KGE ratio**	**R2 ratio**	**RMSE ratio**	**VE ratio**
0	*0.984*	*1.000*	*1.046*	*1.003*	*1.019*
1	*0.750*	*0.824*	*0.945*	*1.085*	*0.711*
2	*0.952*	*0.852*	*1.030*	*1.042*	*0.860*
3	*0.984*	*0.951*	*0.970*	*1.002*	*1.000*
4	*0.845*	*0.850*	*0.937*	*1.111*	*0.771*
5	*0.825*	*0.815*	*0.985*	*1.135*	*0.765*
6	*0.638*	*0.835*	*0.794*	*1.229*	*0.633*
7	*1.017*	*0.949*	*0.984*	*0.986*	*1.000*
8	*0.564*	*0.892*	*0.836*	*1.204*	*0.792*
9	*0.911*	*0.931*	*0.904*	*1.005*	*1.000*
10	*0.826*	*0.972*	*0.923*	*1.040*	*0.935*
11	*0.906*	*0.915*	*0.896*	*1.073*	*0.945*
12	*0.958*	*0.837*	*0.946*	*1.055*	*0.915*
13	*0.970*	*1.000*	*0.986*	*1.003*	*0.980*
14	*0.891*	*0.768*	*0.853*	*1.077*	*0.941*

Notes: Values reported in this table are a ratio between the goodness-of-fit values estimated from the parameter regressions to the values obtained over the calibration period.

LOOCV - Deleted Residuals		
parameter	**model variance**	**deleted residual variance**
agwrc	*4.51E-04*	*5.67E-04*
deepfr_log	*1.22E+00*	*1.60E+00*
infilt	*1.37E-02*	*1.70E-02*
intfw	*3.33E+00*	*4.19E+00*
irc	*5.55E-03*	*8.42E-03*
kmelt	*5.86E-04*	*8.89E-04*
lzsn_log	*2.43E-02*	*3.12E-02*
uzsn_log	*1.71E-01*	*2.22E-01*

3.1.2 HBV Modeling

HBV Sensitivity Analysis: HSY-GSA	
Parameter	**d-value**
cperc	0.356
lsuz	0.272
k1	0.183
k0	0.182
SCF	0.171
DDF	0.169
Tm	0.134
FC	0.115
BETA	0.093
k2	0.086
Ts	0.063
LPrat	0.035
Tr	0.029
croute	0.028
bmax	0.005

	parameters	description	units	lower	range	upper
constant parameters	**tr**	threshold temperature above which precipitation is rain	(deg C)	1	1.96	3
	ts	threshold temperature below which precipitation is snow	(deg C)	-3	-1.75	1
	lprat	parameter related to the limit for potential evaporation	-	0	0.4819	1
	beta	the non linear parameter for runoff production	-	0	9.38	20
	k2	storage coefficient for slow response	(day)	30	122.1	250
	bmax	maximum base at low flows	(day)	0	12.48	30
	croute	free scaling parameter	(day^2/mm)	0	20.83	50

locationID	HBV Calibration - Jan 01, 1980 to Dec 31 1990								HBV Validation - Jan 1, 1991 to Dec 31, 1995							
	RMSE	NRMSE	PBIAS	R2	NSE	KGE	VE	d	RMSE	NRMSE	PBIAS	R2	NSE	KGE	VE	d
0	*2.32*	*79.6*	*-5.8*	*0.48*	*0.37*	*0.69*	*0.45*	*0.82*	*1.8*	*76.7*	*-6.1*	*0.53*	*0.41*	*0.71*	*0.49*	*0.84*
1	*3.15*	*89.5*	*4.1*	*0.36*	*0.2*	*0.6*	*0.34*	*0.75*	*2.47*	*91.3*	*5.4*	*0.37*	*0.17*	*0.6*	*0.38*	*0.76*
2	*2.9*	*77.6*	*0.3*	*0.5*	*0.4*	*0.71*	*0.41*	*0.83*	*2.28*	*82.8*	*2.8*	*0.51*	*0.31*	*0.67*	*0.44*	*0.83*
3	*3.21*	*85.7*	*-5.2*	*0.4*	*0.26*	*0.63*	*0.34*	*0.78*	*2.98*	*108.3*	*7.9*	*0.33*	*-0.17*	*0.48*	*0.26*	*0.72*
4	*2.96*	*77.7*	*-0.3*	*0.5*	*0.4*	*0.7*	*0.43*	*0.83*	*2.24*	*78.8*	*1.7*	*0.53*	*0.38*	*0.7*	*0.48*	*0.84*
5	*2.83*	*74.9*	*-1.9*	*0.52*	*0.44*	*0.72*	*0.42*	*0.84*	*2.16*	*77*	*0.7*	*0.54*	*0.41*	*0.72*	*0.44*	*0.85*
6	*2.97*	*73.4*	*-5.4*	*0.54*	*0.46*	*0.73*	*0.44*	*0.85*	*2.32*	*74.1*	*-3.2*	*0.57*	*0.45*	*0.73*	*0.47*	*0.86*
7	*3.2*	*78.6*	*-6*	*0.49*	*0.38*	*0.69*	*0.44*	*0.83*	*2.5*	*78.7*	*1.4*	*0.54*	*0.38*	*0.7*	*0.43*	*0.85*
8	*3.19*	*82.6*	*-8.6*	*0.43*	*0.32*	*0.65*	*0.42*	*0.8*	*2.67*	*78.9*	*-0.6*	*0.47*	*0.38*	*0.68*	*0.44*	*0.82*
9	*3.07*	*82.4*	*-7.7*	*0.43*	*0.32*	*0.64*	*0.41*	*0.79*	*2.33*	*77.8*	*4.6*	*0.53*	*0.39*	*0.71*	*0.42*	*0.84*
10	*3.16*	*84.4*	*-12.1*	*0.43*	*0.29*	*0.63*	*0.41*	*0.79*	*2.37*	*78.7*	*-1*	*0.54*	*0.38*	*0.7*	*0.44*	*0.85*
11	*2.36*	*75.4*	*-4.9*	*0.53*	*0.43*	*0.72*	*0.45*	*0.84*	*1.82*	*73.4*	*2.1*	*0.59*	*0.46*	*0.74*	*0.45*	*0.87*
12	*3.06*	*78.4*	*-3.7*	*0.49*	*0.39*	*0.7*	*0.42*	*0.83*	*2.27*	*77.2*	*2.2*	*0.58*	*0.4*	*0.7*	*0.43*	*0.86*
13	*3.29*	*73.6*	*-2.9*	*0.49*	*0.46*	*0.68*	*0.4*	*0.82*	*3.09*	*78.5*	*-7.7*	*0.42*	*0.38*	*0.61*	*0.4*	*0.78*
14	*3.26*	*92*	*-5*	*0.34*	*0.15*	*0.58*	*0.31*	*0.74*	*3.03*	*109.9*	*5.4*	*0.27*	*-0.21*	*0.47*	*0.25*	*0.69*

Notes: The calibration period includes a one-year warm up period. For validation, the model is run through the calibration period and then the goodness-of-fit measures are assessed from Jan 1, 1991 to Dec 31, 1995. All sub-basins were used in the regionalization regression process.

Shapiro-Wilks Test for Normality		
	Indepent. Variable	p-value
HBV Parameters	SCF	*0.014*
	DDF	*0.870*
	Tm	*0.033*
	FC	*0.179*
	k0	*0.220*
	k1_log	*0.442*
	lsuz_log	*0.657*
	cperc_sqrt	*0.036*
Spatial	per_north	*0.267*
	per_east	*0.193*
	per_developed	*0.730*
	per_forest	*0.936*
	per_agr	*0.336*
	elev_MEAN	*0.693*
	slope_MEAN	*0.582*
	slope_std	*0.544*
	per_HGB	*0.190*
	per_HGC	*0.284*
	per_HGD_log	*0.748*
	TI_mn	*0.854*
	TI_max	*0.505*
	TI_min	*0.591*
	DA_log	*0.914*
	strm_len_log	*0.314*
Climate	ppt_log	*0.144*
	tmax	*0.732*
	tmean	*0.071*

Note: Values highlighted in bold indicate the paramters that could not be corrected using a log transformation. These parameters tended to bump into their upper/lower limits during calibration.

Results for the HBV regionalization regressions

	cperc_sqrt (1)	ddf (2)	fc (3)	k0 (4)	k1_log (5)	lsuz_log (6)	scf (7)	tm (8)
fc PCA comp.	8.828*** (2.560)			7.833** (3.496)	-16.514			
tm PCA comp.		-0.258 (0.168)						
per.east			91.119*** (29.632)					
per.HGD.log						17.133 (10.455)		
TI.min							-0.129** (0.055)	
tmean								0.667** (0.250)
Constant	-1.118** (0.448)	1.580*** (0.285)	228.044*** (37.467)	0.031 (0.612)	1.879*** (0.503)	1.299*** (0.100)	1.991*** (0.342)	0.235 (0.339)
N	15	15	15	15	15	15	15	15
R^2	0.478	0.153	0.421	0.279	0.235	0.171	0.297	0.354
Adjusted R^2	0.438	0.088	0.377	0.223	0.177	0.107	0.242	0.304
Residual Std. Error (df = 13)	0.306	0.448	145.111	0.418	0.343	0.162	0.217	1.312
F Statistic (df = 1; 13)	11.897***	2.35	9.456***	5.020**	4.001*	2.685	5.480**	7.112**
Notes:	*Parenthesis represent standard error of estimate. The first PCA component is used for the agwrc, intfw, irc, lzsn, and uzsn regressions. ***Significant at the 1 percent level; **Significant at the 5 percent level; *Significant at the 10 percent level.*							

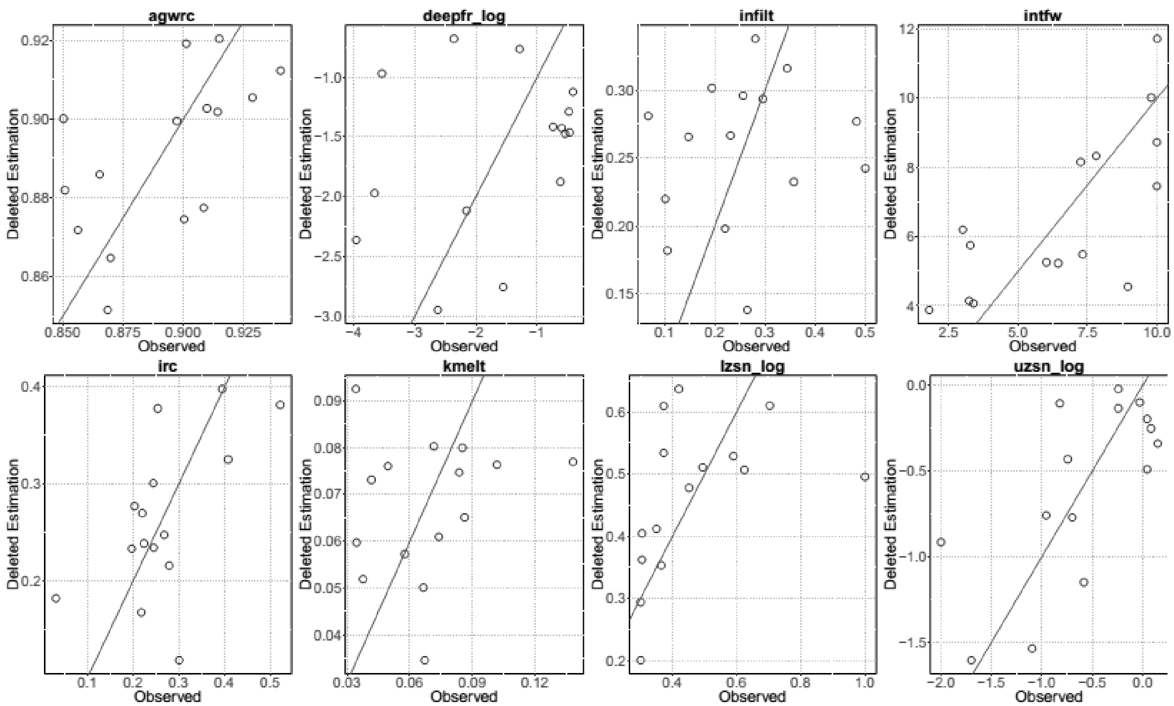
LOOCV - Comparison to Calibration Period Goodness-of-fit Measures					
locationID	**NSE ratio**	**KGE ratio**	**R2 ratio**	**RMSE ratio**	**VE ratio**
0	*0.595*	*0.913*	*0.979*	*1.108*	*0.867*
1	*0.100*	*0.883*	*0.861*	*1.105*	*0.735*
2	*0.650*	*0.873*	*0.820*	*1.110*	*0.829*
3	*0.885*	*0.651*	*0.775*	*1.025*	*1.059*
4	*0.625*	*0.914*	*0.840*	*1.115*	*0.837*
5	*0.523*	*0.847*	*0.750*	*1.173*	*0.786*
6	*0.457*	*0.836*	*0.704*	*1.215*	*0.727*
7	*1.053*	*0.986*	*0.980*	*0.988*	*1.023*
8	*1.188*	*0.908*	*1.093*	*0.956*	*1.024*
9	*0.688*	*0.844*	*0.744*	*1.068*	*1.000*
10	*0.862*	*0.889*	*0.837*	*1.025*	*1.000*
11	*0.581*	*0.889*	*0.925*	*1.153*	*0.933*
12	*1.103*	*0.971*	*1.000*	*0.961*	*1.119*
13	*1.109*	*0.956*	*1.102*	*0.951*	*1.200*
14	*1.067*	*0.776*	*0.824*	*0.994*	*1.129*

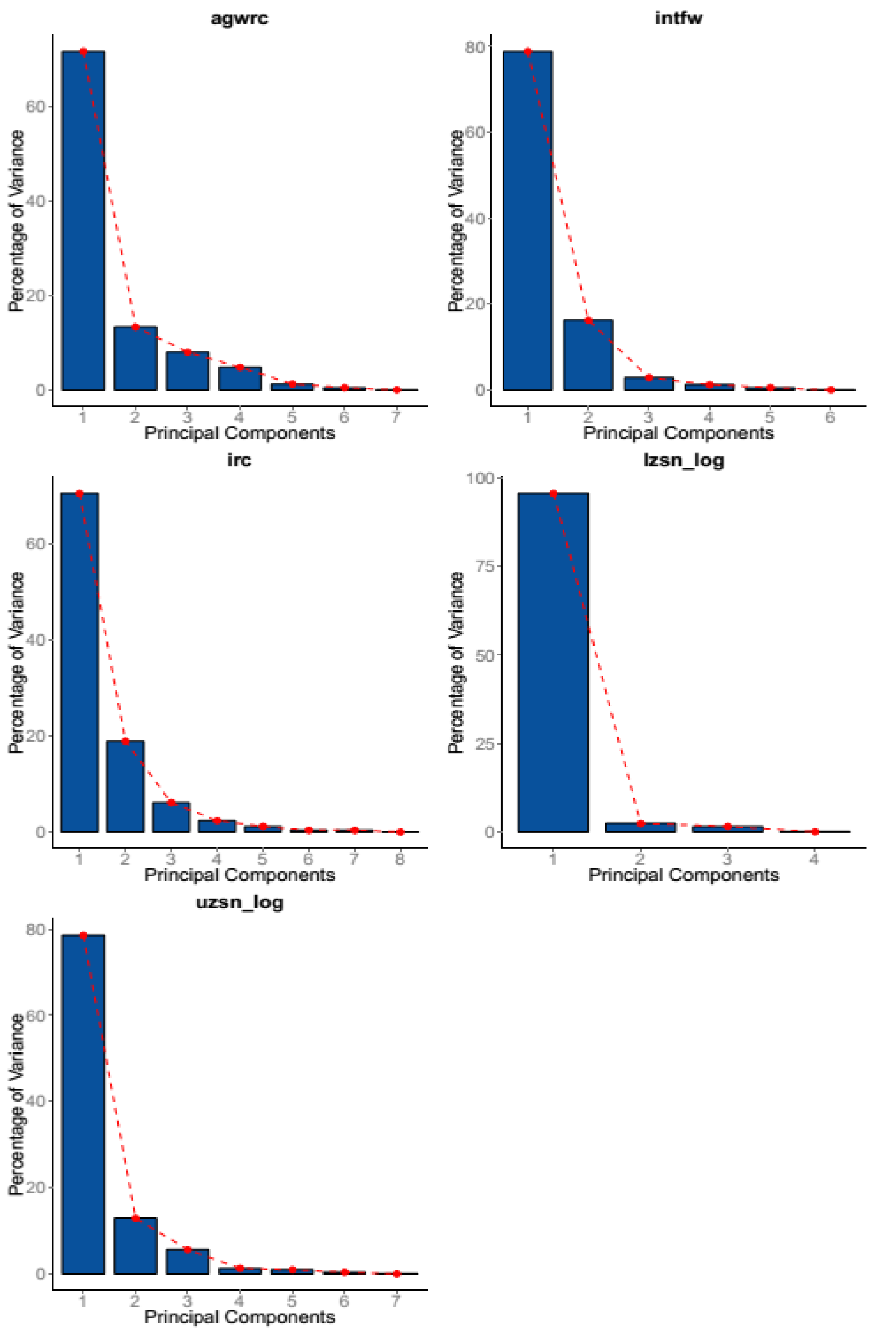
Notes: Values reported in this table are a ratio between the goodness-of-fit values estimated from the parameter regressions to the values obtained over the calibration period.

LOOCV- Deleted Residuals (HBV)		
parameter	**model variance**	**deleted residual variance**
cperc_sqrt	8.68E-02	1.35E-01
ddf	1.87E-01	2.58E-01
fc	1.96E+04	2.56E+04
k0	1.62E-01	2.13E-01
k1_log	1.09E-01	1.36E-01
lsuz_log	2.44E-02	2.89E-02
scf	4.37E-02	5.68E-02
tm	1.60E+00	1.95E+00

3.2 Figures

3.2.1 HSPF Modeling




agwrc
intfw
irc
lzsn_log
uzsn_log
Percentage of Variance
Principal Components

3.2.2 HBV Modeling