Estimating *a-priori* Kinematic Wave Model Parameters Based on Regionalization for Flash Flood Forecasting in the Conterminous United States

Humberto Vergara[a,b,c,d], Pierre Kirstetter[c,b], Jonathan Gourley[b], Zac Flamig[a,b,c], Yang Hong[c,d], Ami Arthur[a,b], Randall Kolar[d]

Affiliation(s):

a – Cooperative Institute for Mesoscale Meteorological Studies (CIMMS), The University of Oklahoma, Norman, OK, USA

b – NOAA/National Severe Storms Laboratory (NSSL), Norman, OK, USA

c – Advanced Radar Research Center (ARRC), The University of Oklahoma, Norman, OK, USA

d – School of Civil Engineering and Environmental Science, The University of Oklahoma, Norman, OK, USA

Corresponding author:

Humberto Vergara
120 David L. Boren Blvd. Ste. 4600
National Weather Center
Norman OK 73072-7307
humber@ou.edu

Abstract

This study presents a methodology for the estimation of *a-priori* parameters of the widely used kinematic wave approximation to the unsteady, 1-D Saint-Venant equations for hydrologic flow routing. The approach is based on a multi-dimensional statistical modeling of the macro scale spatial variability of rating curve parameters using a set of geophysical factors including geomorphology, hydro-climatology and land cover/land use over the Conterminous united States. The main goal of this study was to enable prediction at ungauged locations through regionalization of model parameters. The results highlight the importance of regional and local geophysical factors in uniquely defining characteristics of each stream reach conforming to physical theory of fluvial hydraulics. The application of the estimates is demonstrated through a hydrologic modeling evaluation of a deterministic forecasting system performed on 1,672 gauged basins and 47,563 events extracted from a 10-year simulation. Considering the mean concentration time of the basins of the study and the target application on flash flood forecasting, the skill of the flow routing simulations is significantly high for peakflow and timing of peakflow estimation, and shows consistency as indicated by the large sample verification. The resulting *a-priori* estimates can be used in any hydrologic model that employs the kinematic wave model for flow routing. Furthermore, probabilistic estimates of kinematic wave parameters are enabled based on uncertainty information that is generated during the multi-dimensional statistical modeling. More importantly, the methodology presented in this study enables the estimation of the kinematic wave model parameters anywhere over the globe, thus allowing flood modeling in ungauged basins at regional to global scales.

Keywords: Kinematic wave routing, ungauged prediction, regionalization, fluvial hydraulics, multi-dimensional analysis, large sample hydrology.

1. Introduction

Providing useful estimates of the response of a hydrologic system (i.e. a catchment or watershed) at all locations (i.e. gauged and ungauged) is arguably *The Challenge* in rainfall-runoff modeling. This was the main subject of the past decade-long focus of the International Association of Hydrological Sciences (IAHS) through its Prediction at Ungauged Basins (PUB) initiative (Sivapalan et al. 2003), which, although promoted scientific productivity, was largely unsuccessful in achieving its main goal (Hrachowitz et al. 2013). The underlying challenge of PUB can be phrased as *how do we generate equally skillful model estimates at all locations regardless of whether there are measurements of the model output or not?* A key aspect involved in this challenge is the regionalization problem in hydrologic modeling, which is primarily concerned with the estimation of parameters at ungauged locations (Beven 2011). The parameters' main role is to enable the versatility of the model in simulating a diverse set of hydrologic processes and responses, thus facilitating the application of the model at all locations.

The estimation of hydrologic model parameters has been the concentration of many studies for the past two decades or so, the majority featuring model calibration techniques (e.g., Sorooshian et al. 1993; Boyle et al. 2000; Duan 2003; Gupta et al. 2003; Vrugt et al. 2006; Vrugt et al. 2008). However, model calibration is a technique primarily developed for lumped hydrologic models. This is because the spatially aggregated conceptualization of processes and parameterization in lumped models makes it difficult to employ an approach based on characterizations of the spatial variability of the basin physical structure (e.g., topography or soil texture properties such as hydraulic conductivity). Process-based distributed hydrologic models, on the other hand, are specifically designed to take advantage of the ever-increasing availability of geospatial datasets from geographical information systems and remote-sensing platforms to resolve the dominant spatial patterns of the hydrologic system. Consequently, distributed hydrologic models can be configured using *a-priori* methods for parameter estimation, which are naturally consistent with the PUB challenge and the regionalization problem.

While work on *a-priori* estimates for water balance model parameters based on soil properties have been reported to the literature (e.g. Koren et al. 2000; Yao et al. 2012), few efforts have been devoted to derive spatially-distributed flow routing parameter estimates without conditioning from calibration (e.g. Naden et al. 1999). The primary objective of routing models is to describe the space-time evolution of water flow throughout a watershed, catchment or stream network. Moreover, flow routing is essential in the description of flood wave timing, which not only establishes when a flooding event occurs, but also the magnitude and duration of the flood. Flood wave timing is critical in forecasting approaches that rely on threshold-based methodologies for detection (e.g. Reed et al. 2007). Some studies like the ones of Montgomery and Gran (2001) and Finnegan et al. (2005) have analyzed controlling factors of the downstream variability of channel characteristics related to routing parameters. Koren et al. (2004) discuss a methodology in which rating curve data at the basin outlet can be propagated upstream to populate all grids within the watershed with estimates of the flow routing parameters. However, and to the knowledge of the authors, no study has reported a methodology to estimate flow routing parameters at continental scales.

In this work, the spatial variability of parameter estimates of a physics-based distributed routing model was studied at the continental scale to devise an estimation approach based on regionalization. The choice of a physics-based model (i.e. models formulated from physical laws) is centered on the fact that model parameters are either based on or correspond to actual measurements of the physical system (Boyle et al. 2000), which facilitates the process of *a-priori* estimation. Moreover, the approach used herein to study the spatial characteristics of parameter estimates explores associations with several geophysical properties of the land surface. Using a model whose conceptualization of the physical system significantly departs from reality would prove difficult (if not impossible) to find aforesaid associations. The study was developed in the context of the Flooded Locations and Simulated Hydrographs (FLASH) project, whose main objective is *"to improve the accuracy, timing, and specificity of flash flood warnings in the US"* (NSSL 2014).Consequently, the overall goal of this study is find *a-priori* estimates of kinematic wave routing parameters in order to enable regional forecasting of floods and flash floods at a continental scale with a distributed hydrologic modeling system.

2. Physics-based distributed flow routing model

In general, there are two types of flow routing models: lumped routing models and distributed routing models, sometimes referred to as *hydrologic routing* and *hydraulic routing* respectively (Chow et al. 1988; Bedient et al. 2008). Lumped routing models usually employ empirical or conceptual ideas to describe the true mechanisms of water flow process in a hydrologic system. Distributed routing models, on the other hand, consider both space and time. Furthermore, and because water flow is a continuous variable, these models solve partial differential equations related to the physical laws governing the water movement mechanisms in a hydrologic system. Depending on the assumptions and approximations applicable to a particular hydrologic system, different distributed routing models can result.

The model selected herein was the kinematic wave approximation to the one-dimensional unsteady open channel flow equations developed by Barré de Saint-Venant in the 1800s (Beven 2011). The full implementation of the Saint-Venant equations represents the closest description of the 1-D water movement in a watershed. However, the use of alternative models by simplification of the governing equations is motivated by simpler and computationally less expensive methods for distributed flow routing. Additionally, these simpler models can capture the dominant physical processes depending on specific flow conditions. Kinematic wave model is arguably the most widely used distributed flow routing method in hydrologic modeling, given its simplicity as compared to the diffusion or dynamic wave models. A general criterion to support the use of the kinematic wave approximation is based on the slope: in watersheds with predominantly steep slopes, the flow conditions are such that the kinematic wave concept reasonably approximates the unsteady flow phenomena (Ponce 1986). Moreover, Ponce (1991) claimed that for most overland flow situations, kinematic wave approximation requirements are satisfied. Kazezyilmaz-Alhan and Medina (2007) define a minimum slope of 0.002 as a general guidance value required for kinematic wave applicability. Figure 1 presents a map of the applicability of the kinematic wave approximation over the Conterminous United States (CONUS) based on the aforementioned criterion. It can be observed that the kinematic wave approximation applies for the majority of CONUS.

Several well-known models or modeling frameworks implement kinematic wave for the flow routing component. A list of some past studies and modeling systems employing kinematic wave are presented in Table 1. In the majorities of these studies, the parameters of the routing model are derived from assumptions on the channel geometry (e.g. Feldman 1995; Feldman 2000; Liu and Todini 2002). In other cases, the estimation of the kinematic wave parameters relies on model calibration (e.g. Beldring et al. 2003). In this study, a methodology that does not employ assumptions of channel geometry nor relies on model calibration for the estimation of kinematic wave parameters is presented.

2.1. Derivation of the Kinematic Wave approximation

The one-dimensional nature of Saint-Venant equations relate to the fact that spatial variations of velocity can be neglected both horizontally and vertically across the channel when the interest is in the main direction of water flow (i.e. along the channel). Similarly, the water surface elevation is assumed to be constant horizontally at any section of the channel. In hydrologic applications at the watershed, catchment or stream network scales (e.g. hundreds of meters to a few kilometers), the aforementioned approximations are acceptable. The Saint-Venant equations are derived from the Eulerian view of motion, where physical laws are applied to the continuum of a fluid as it passes through a control volume. The concept is applied through the Reynolds transport theorem, which relates the time rate of change of a mass-dependent property of the fluid to the external factors causing this change (Chow et al. 1988). Applying the theorem to conservation of mass and momentum, Newton's second law of motion, and neglecting lateral inflow, wind shear and eddy losses, the Saint-Venant equation for continuity is given as:

$$\frac{\partial Q}{\partial x} + \frac{\partial A}{\partial t} = 0 \qquad (1)$$

where Q is the flow, A is the channel cross-section area, x is a horizontal distance and t is time. Likewise, the equation for momentum is given as:

$$\frac{1}{A}\frac{\partial Q}{\partial t} + \frac{1}{A}\frac{\partial}{\partial x}\left(\frac{Q^2}{A}\right) + g\frac{\partial y}{\partial x} - gS_o + gS_f = 0 \qquad (2)$$

where g is the acceleration due to gravity, S_o is the slope of the bottom of the channel, and S_f is the friction slope. Equation (2) above can be broken down into the different physical processes governing flow momentum represented in each equation term (from left to right): the local acceleration, the convective acceleration, the pressure force, the gravity force and the friction force.

Equations (1) and (2) above represent the governing equations for one-dimensional, unsteady, open channel flow. Simplifications in the Saint-Venant equations result in different distributed routing models. When equation (1) and (2) are applied in full (i.e., no simplifications), the method is called a dynamic wave model. When the acceleration (i.e., inertial) terms are neglected in (2), the method is called a diffusion wave model. Finally, if the acceleration and force (i.e. pressure) terms are ignored in (2), the method is called a *kinematic* *wave model*. Depending on the attributes of the channel and flow magnitude, the application of diffusion or kinematic wave approximations might be limited. For example, in the case of overbank flow during flood events where the geometry and material of the floodplain significantly differ from the main channel section and the vertical variability of velocities are not negligible (Moussa and Bocquillon 2000). However, it is a common practice to select one approximation depending on other criteria such as computational efficiency and availability of the necessary information to estimate model parameters. The latter two criteria were the main considerations for the choice of the kinematic wave model in the implementation featured in this study.

The kinematic wave simplifications yield $S_o = S_f$, which means that the flow is assumed uniform and, thus, a function of depth or channel's cross-section area alone. Consequently, the form of the kinematic wave equation for momentum becomes:

$$Q = \alpha A^{\beta} \tag{3}$$

where α and β are the *kinematic wave model parameters*. Substitution of (3) in (1) yields an expression for solving for Q as the only dependent variable (Chow et al. 1988):

$$\frac{\partial Q}{\partial x} + \alpha\beta Q^{\beta-1}\frac{\partial Q}{\partial t} = q \tag{4}$$

where q is the lateral inflow to the channel.

2.2. Methods for the estimation of the kinematic wave parameters

The standard method to estimate the kinematic wave parameters is based on an assumed channel cross-section shape and the application of Manning's equation, which accounts for the slope and the roughness of the channel (Bedient et al. 2008). Commonly used shapes to model natural streams' channel cross-section are rectangular, trapezoidal and parabolic (Dingman 2009). Each of these has explicit functions for the estimation of α and β derived from Manning's equation. A caveat of this method is precisely the need for explicit specification of channel cross-section shape. Because of the mathematical manipulation of Manning's equation, it is difficult (if not impossible) to use the actual cross-section shapes of natural streams, which are rather irregular. Moreover, the assumption of regular shapes, on the other hand, consequently leads to the assumption of prismatic channels (i.e. assuming the entire channel has a constant shape). There have also been attempts to employ geomorphological characteristics of basins and empirical relationships with channel geometry (e.g. Vélez et al. 2009; Reggiani et al. 2014). However, these empirical relationships are based on limited samples of river reaches and are usually followed by model calibration.

An alternative method is based on statistical analysis of rating curve data. Field measurements at stream gauges provide a mean to estimate the parameters α and β directly. Based on the form of the momentum equation shown in (3), a power function relating streamflow and channel cross-sectional area can be fitted to data measured in the field (Fig. 2). The field data needs to encompass a wide range of flows to have a representative sample able to describe the relationship. Usually, the majority of the data come from flows of low to average magnitudes (although it can also include some significantly high flows), because of the low frequency of high flows and difficulties in measuring in the field under flooding conditions (Beven 2011). Also, certain locations display a rather high irregularity in channel geometry, which leads to multiple relationships between streamflow and cross-section area. Panel b) of Fig. 2 shows an example of this kind of behavior, where at approximately 600 m^3/s, the relationship changes abruptly indicating a significant change in geometry. These changes occur when the material of the channel transitions from fine to coarse sediments such as in the case of water flowing out of the riverbanks to the floodplain (Ryan and Porth 2007). Although a multi-segment fit to field measurements data is possible and can more accurately parameterize rating curves (Reitan and Petersen-Øverleir 2009), a single segment fit approach was chosen for simplicity and avoid the necessity of modifying the implementation of the kinematic wave model to employ variable parameters. Nevertheless, this approach offers a way to directly estimate kinematic wave parameters, which implicitly accounts for channel cross-section shape, roughness, and slope.

This method has been described for the configuration of the HL-RMS distributed model in Koren et al. (2004) and in unpublished work by the Office of Hydrologic Development (OHD). They present a methodology to propagate the estimates of the rating curve parameters obtained at gauged locations to upstream locations (i.e. ungauged) using several empirically derived geomorphological functions based on drainage area solely. While their results show reasonable skill, their methodology is aimed at estimating routing parameters at the local scale. Additionally, some aspects in their methodology, such as the use of drainage area alone to define the variability of the parameter estimates, and the upstream propagation approach are simplistic and subject to unverified assumptions. Intuitively, flow conditions in non-regulated streams (i.e. no regulation or diversion structures) are defined by both local and upstream regional factors and, thus, a downstream approach is preferred.

3. Methodology of the *a-priori* estimation

The approach to kinematic wave parameters estimation presented herein is based on the rating curve method described in Section 2.2. The main aspect of the strategy was the investigation of explanatory geophysical factors of the spatial variability of rating curve parameters at a macro scale, with the aim of estimating kinematic wave parameters. This data intensive exercise represents a case of what has been called the "*fourth paradigm of science*" (Hey 2012) and the concept of "*large sample hydrology*" (Gupta et al. 2014). The ultimate goal of this study was to enable river flow routing simulation with a distributed hydrologic model for flash flood forecasting over CONUS without calibration (i.e. without model parameter fitting to a streamflow time-series).

3.1. Geospatial datasets over CONUS

3.1.1. Field measurements of streamflow and channel cross-section area

Using the record of stream gauge stations in the database described in Gourley et al. (2013), field measurement data from the U.S. Geological Survey (USGS) archive were obtained. A series of filtering steps were taken in order to robustly generate an appropriate sample for the statistical analysis of the spatial variability of rating curve parameters. First, the selection of stations was limited to those within CONUS, which amounts to approximately 9,000 gauges. Secondly, a filter was applied to the record in order to study natural streams only (i.e. no regulation or diversion of any degree). The identification of regulated stations was done by examination of the annual peak flow historical record of each USGS gauge station, where flags indicating the level of impact by regulation or diversion are specified. Lastly, an automatic processing script was employed to fit the streamflow and channel cross-sectional area data to a power-law function following Equation (3) for each of the selected USGS stations (see example in Fig. 2). An evaluation of the goodness-of-fit yielded a final sample size of 4,943 stream gauges employed in the analysis of this work.

3.1.2. Watershed characteristics as explanatory variables

Streamflow results from the natural integration in space and time of the different hydrologic processes occurring in a watershed (or basin), the main physical unit subject to measurements and modeling in hydrology (Bedient et al. 2008). For effects of analysis and the hydrologic model implementation, the pixel of a rectangular grid is defined herein as the elementary unit representing a stream reach and the immediately adjacent overland area (i.e. hillslope). The particular characteristics of each stream reach, assumed to be uniform within the pixel, are uniquely determined by the flow contributed by its drainage basin, its current and past geology, topography, pedology and climate, and are part of a spatial continuum that includes the entire watershed (Dingman 2009). Therefore, several of these geophysical characteristics of watersheds were explored as potential explanatory factors of the variability of rating curve parameters.

All geospatial datasets employed in this study were rendered on a rectangular grid with a 1-km pixel resolution. The grid was specifically chosen to match the Digital Elevation Model (DEM) grid on which the flash flood forecasting system is configured over the CONUS. Using DEM data, it is possible to derive geomorphological parameters of any given watershed or catchment. DEM is virtually available everywhere over the globe at high resolution (e.g. 30 meters), which enables the ability to generate geomorphological information at all gauged and ungauged locations. The DEM data used herein were based on the USGS' National Elevation Dataset (NED; Gesch et al. 2009). Geomorphological variables considered herein were selected based on the studies by Schumm (1956) and, in particular, Costa (1987) who analyzed relationships between characteristics of watersheds and flash floods over the CONUS. The variables include drainage basin area, elongation ratio, relief ratio, slope index, slope at the outlet, and river length.

The hydro-climatology of basins was considered by examining mean annual precipitation and average temperature. The data correspond to the 30-year datasets prepared by the PRISM Climate Group (PRISM Climate Group 2012) covering the period 1981 - 2010. Soil datasets from the STATSGO database (Soil Survey Staff 1994; Miller and White 1998) were examined herein. Variables explored from this dataset include soil class, mean rock volume percent, mean depth-to-rock, and erodability factor (K factor). Lastly, land cover and land use data from the National Land Cover Dataset (NLCD 2006; Fry et al. 2011) were utilized to explore the impact of the runoff (i.e. USDA NRCS) curve number.

3.2. Multidimensional analysis of kinematic wave parameters' variability over CONUS

In this work, the spatial variability of the kinematic wave parameters was analyzed through conditional distribution functions. The sets of α and β distributions were studied using the Generalized Additive Models for Location, Scale, and Shape (GAMLSS; Stasinopoulos and Rigby 2007) technique. The GAMLSS model aims to simulate the parameters of a distribution of the response variable (i.e., α or β) according to the values assumed by some explanatory variables (i.e., the geophysical characteristics of basins). GAMLSS was chosen over other multidimensional analysis methods (e.g., principal component analysis or a canonical correlation analysis) because modeling the complete conditional distributions enables diagnostic capabilities on the resulting estimates. More importantly, this method explicitly acknowledges the inherent uncertainty of the estimates, which can be employed for probabilistic applications.

Both parameters α and β were analyzed separately following the same approach. To simplify the description of the methodology, the GAMLSS modeling procedure on α alone is explained as follows. Two main assumptions were made: 1) the response variable α is a random variable following a known parametric distribution with density f conditional on the location parameter μ and the scale parameter σ, and 2) the observed α values are mutually independent given the parameter vectors μ and σ. Each distribution parameter was modeled as a function of the explanatory variable using monotonic (linear/nonlinear or smooth) link functions. More details are provided by Rigby and Stasinopoulos (2001; 2005), Akantziliotou et al. (2002) and Stasinopoulos and Rigby (2007), particularly on the model fitting and selection. It involves identifying a suitable distribution of α, the explanatory variables and the link functions. The estimation method is based on the maximum likelihood principle and the model selection is carried out by checking the significance of the fitting improvement in terms of information criteria such as the Akaike Information Criterion (AIC), the Schwarz Bayesian Criterion (SBC) and the generalized AIC (GAIC; Stasinopoulos and Rigby 2007). Forward, backward, and step-wise procedures were applied to select the meaningful explanatory variables, supervised by diagnostic plots to check the fitting performance as discussed in Stasinopoulos and Rigby (2007).

A wide variety of distributional forms are available within GAMLSS. A number of conditional two-parameter density functions (lognormal, normal, reverse gumbel, logistic, gamma, etc.) were tested to fit the data. The goodness-of-fit on the whole dataset was checked with the AIC for each of the semi-parametric density fits. The logistic distribution was found to be the most appropriate:

$$f_y\left(y \mid \mu,\sigma\right)=\frac{1}{\sigma}e^{-\left(\frac{y-\mu}{\sigma}\right)}\left\{1+e^{-\left(\frac{y-\mu}{\sigma}\right)}\right\}^{-2} \tag{5}$$

The function above was used to model the conditional α distributions, where the location μ is linked to the expected α value, and the scale σ is representative of prediction uncertainty. After selecting the distribution family, the structure of the model was refined through an iterative procedure by trying several combinations of explanatory variables. The trends for each parameter are fitted using penalized splines, which are more flexible than polynomials or fractional polynomials for modeling complex nonlinear relationships. Lastly, the goodness-of-fit was checked by computing the residuals, first four moments, their Filliben correlation coefficient, and quantile-quantile plots (Stasinopoulos and Rigby 2007).

3.3. Hydrologic Validation strategy

3.3.1. Hydrologic modeling using a-priori estimates of the kinematic wave parameters

Additional to the statistical verification explained above, a strategy based on a hydrologic evaluation was employed herein. The methodology evaluates the estimates of the kinematic wave parameters through an assessment of a deterministic hydrologic model implementation over CONUS. A probabilistic application of the *a-priori* estimates is possible given the uncertainty information that is part of their multi-dimensional modeling (Section 3.2). However, implementing the kinematic wave parameter estimates in their probabilistic form is not a trivial task because most hydrologic models are formulated in a deterministic way. Even an ensemble-based method poses challenges in terms of the multivariate nature of uncertainty. This is particularly difficult in this case because the α and β parameters were modeled independently and, thus, no information about their covariance is available. The level of difficulty added by a probabilistic implementation warrants a dedicated study in future work. Because of the focus of this study, a deterministic implementation of the kinematic wave parameter estimates for the hydrologic modeling evaluation is preferred.

The hydrologic model employed in this study was an implementation of the Coupled Routing Excess and Storage (CREST) distributed hydrologic model (Wang et al. 2011) that is used in a modeling framework for flood and flash flood prediction entitled the Ensemble Framework For Flash Flood Forecasting (EF5;Flamig et al. 2010). EF5 is a flexible modeling framework that enables the combination of different physical representations for hydrologic simulation. The configuration used herein consisted of the water balance component of CREST coupled to the kinematic wave model for surface flow routing. Subsurface flow routing was modeled using a distributed version of the linear reservoir (Nash 1957), a lumped routing model commonly used in hydrology (Moore 1985; Chow et al. 1988; Vrugt et al. 2003). The water balance model is based on the variable infiltration curve (Zhao et al. 1980; 1995) for the computation of excess rainfall, which is partitioned into its surface and subsurface components through a conceptual mechanism based on hydraulic conductivity (Wang et al. 2011). The surface excess rainfall component is routed as overland flow with an implementation of the kinematic wave model for a wide shallow (sheet) flow as:

$$\frac{\partial q}{\partial x}+\alpha_0\frac{3}{5}q^{3/5-1}\frac{\partial q}{\partial t}=i-f \tag{6}$$

where q is the overland flow in $m^3/s.m^2$ and the lateral inflow term of equation (4), $i-f$ is the surface excess rainfall from the water balance in m/s, and α_0 is an overland conveyance parameter defined as a function of Manning's roughness coefficient and overland slope alone.

The hydrologic model was configured with *a-priori* estimates for all of its parameters. This includes seven parameters for the water balance and the excess rainfall routing (subsurface and surface), and the kinematic wave parameters α and β for river routing subject of this study (see Table 2). Climatological mean monthly potential evapotranspiration data (Koren et al. 1998) were used as part of the hydrologic model inputs. High resolution (1-km/5-min) quantitative precipitation estimation data from the Multi-Radar/Multi-Sensor system (MRMS; Zhang et al. 2011; Zhang et al. 2015) were utilized to force the hydrologic model. A period of 10 years (2002 – 2011) was used to generate simulations of streamflow at a 5-min time step.

3.3.2. Event-based Skill Assessment

An event-based approach to skill evaluation was followed herein. Individual streamflow events were selected with an algorithm that utilizes a threshold value and a hydrograph separation procedure. An event was defined as that exceeding the 90th percentile flow value of the historical record at each gaged location. The evaluation employed stream gauge stations with no regulation and drainage area less than 1,000 km^2, which is a representative scale for the majority of drainages over CONUS (> 95%; Fig. 3). Additionally, locations with poor radar coverage and significant snow in the annual precipitation were filtered out. Radar coverage was quantified using the Hybrid Scan Reflectivity Height (HSRH; km), which is part of the MRMS suite of products. The percentage of pixels within a basin with an HSRH below 2 km was computed, and a subjectively chosen threshold of 80% was used to select basins with adequate coverage. Mean percentage of snow contribution to total annual precipitation was obtained from the Geospatial Attributes of Gages for Evaluating Streamflow (GAGE) dataset (Falcone et al. 2010), and a threshold of 30% was used to filter out snowmelt-dominated basins.

The aforementioned screening procedure resulted in an evaluation sample consisting of 47,563 events from 1,672 basins. This filtering was performed in order to reduce the impact of uncertainty from sources unrelated to the estimation of kinematic wave model parameters. Naturally, not all sources of uncertainty can be effectively neglected or accounted for. However, the quantitative approach to skill evaluation employed herein is able to target specific signatures of the modeling of flood wave routing. Two metrics to assess the skill of the simulations were used in these experiments: Peak Time Error (in units of hours) and Relative Peak Error (in units of %). Vergara et al. (2013) demonstrated the use of these two metrics to disentangle the impact of rainfall and flow routing uncertainty. The Peak Time Error was computed using serial date numbers, which represent the fractional number of hours from a reference date and time (e.g. 01-Jan-2000 00h):

$$Peak\_Time\_Error(hours) = Dt_{sim}^{peak} - Dt_{obs}^{peak} \quad (6)$$

where Dt_{obs}^{peak} is the serial date number of the observed peak flow in hours and Dt_{sim}^{peak} is the serial date number of the simulated mean peak flow in hours. A negative value of the Peak Time Error indicates peak flow is simulated early, while a positive value indicates peak flow is simulated late. To further the interpretation of the peak timing skill, mean concentration time of each of the selected basins in computed according to the method described by Mockus (1961) and used as reference value for the magnitude of the Peak Time Error. Lastly, the Relative Peak Error is computed according to the following:

$$Peak\_Error(\%) = \left(\frac{Q_{sim}^{peak} - Q_{obs}^{peak}}{Q_{obs}^{peak}} \right) \times 100\% \quad (7)$$

where Q_{obs}^{peak} is the event's observed peak flow in m^3/s and Q_{sim}^{peak} is the event's simulated peak flow in m^3/s. A negative value of the Relative Peak Error indicates underestimation of the event's peak flow, while a positive value indicates overestimation of the event's peak flow.

4. Discussion of modeling results

4.1. Estimation of parameters α and β

4.1.1. Association of α and β with watershed geophysical characteristics

Figure 4 presents the values of the rating curve parameters from all selected USGS stations over CONUS. An initial visual assessment of the spatial variability of both parameters reveals distinct patterns associated with the hydro-climatology and topography across the CONUS. Specifically, α variability appears correlated with the mean annual precipitation and β shows a strong association with relief ratio (Fig. 5). The β parameter also presents features corresponding to some clusters observed in the mean rock volume percent. This is consistent with findings of Finnegan et al. (2005) in relation to the scaling of channel geometrical characteristics depending on the material in which the channel is developed.

Scatterplots illustrating the aforementioned associations are presented in Fig. 6. The scaling effect of drainage area on the α parameter is arguably not surprising given its well-known relationship with channel width used in fluvial hydraulics (Montgomery and Gran 2001; Dingman 2009). An interesting feature, however, is the conditioning of this scaling by the hydro-climatology of the basins. Likewise, the indirect relationship between the β parameter and the relief ratio of the basins shows dependency on the mean rock volume percent. The aforementioned conditioning is a consequence of the interactions between different geophysical factors, which are evidenced by the clustering of points shown with the color scales. Further analysis of associations between the rating curve parameters and geophysical characteristics was performed through 2-D and 3-D methods such as density-colored scatterplots. However, it was not possible to observe additional significant relationships because the conditioning of the associations, which are a consequence of the interactions of several geophysical factors considered, needs to be assessed through high-dimensional analytical methodologies, such as GAMLSS.

4.1.2. *Multi-dimensional modeling with GAMLSS*

The GAMLSS model was constructed following the methodology explained in Section 3.2. The geophysical variables retained by GAMLSS and their corresponding statistical significance values are presented in Table 3. Additionally, diagnostic scores of the *goodness-of-fit* are included in Table 4. The model identified several of the important factors that were discussed in the simpler 2-D analysis discussed in Section 4.1.1. Drainage area, relief ratio, rock volume and the hydro-climatic variables are highlighted by their significance levels. This can be interpreted as a sign of robustness of the GAMLSS model.

An evaluation of the resulting model is shown in Fig. 7. Panels a) and b) in the figure present scatter density plots for α and β. Overall, GAMLSS displays skill to predict the expected values of α and β as indicated by the high densities of data points close to the 1-to-1 line and by their correlation coefficient values of 0.73 and 0.63, respectively. However, significant inaccuracies can be observed on the upper end of the rating curve α and the lower end of the rating curve β. An investigation of the rating curves associated with these estimates revealed a flow rate-dependent hysteresis at the corresponding gauged locations. The methodology followed herein for the fitting of rating curves does not account for this behavior and, thus, the estimates of the power-law regression parameters will have significant uncertainty. Moreover, the conditions that need an elaborate description of the hydraulics in an open channel (e.g. dynamic wave model) are out of the scope of the flow routing modeling subject of this work. A summary of the conditional distributions of the predicted values of α and β are shown in quantile plots in panels c) and d) of Fig. 7. It can be observed that in both cases the estimates display heteroscedasticity with respect to the reference values. These plots also show the significant variability on the upper end of the range of α values and the lower end of β values. This information is useful for applications that consider uncertainty such as in probabilistic forecasting frameworks.

The model fit with GAMLSS was employed to produce 1-km grids of the kinematic wave parameters over the CONUS. Each of the geophysical variables used in the analysis was available over the entire computational grid for which the hydrologic model was configured as explained in Section 3.1.2. Some of the ranges of the explanatory variables for the prediction dataset are larger than those for the training dataset (Table 5). The methodology, however, allows for a supervised extrapolation that was implemented herein. Figure 8 presents samples of the *a-priori* estimates of the kinematic wave parameters α and β and their corresponding grids of standard deviation. The main spatial patterns observed on the grids clearly correspond to climatology of precipitation and relief. A closer examination on the α grid also shows the influence of catchment size as indicated by high values at large streams. This is consistent with the analysis on geophysical characteristics discussed in Section 4.1.1 above. Additionally, it can be observed that the estimates have low standard deviations, which indicates that the GAMLSS model has good precision. Some regions display noticeably higher standard deviation such as in Nebraska, northwestern Kansas, Iowa, Illinois, the Mississippi valley, Florida, and southern California. Locations with significantly higher deviations are generally scattered although some clusters can be observed for the β estimates over Florida, the Mississippi valley and on the coast of North Carolina. Visual inspection of the maps of the different geophysical variables points to flat areas where the kinematic wave model may not apply and sandy soils as possible factors for this variability in the estimates. A rigorous and elaborate analysis of this particular aspect of the estimation should be performed in future works to understand these specific factors of uncertainty.

4.2. Hydrologic modeling evaluation

4.2.1. Discussion on event-based evaluation and flow routing signatures

Streamflow at any given location (e.g. an outlet) results from the convolution of flood wave routing of upstream reaches. Therefore, the analysis herein on streamflow simulation is representative of the integrated impact of the estimates of the kinematic wave parameters. A sample of the simulation of streamflow events demonstrating model skill and different signatures of the simulated flood wave routing is presented in Fig. 9. The events were selected from a historic group of floods occurring in September of 2009 in the southeast of the United States, where eleven fatalities resulted from flash floods and floods and a total of $270M USD of damage occurred (NWS 2010). In general, the hydrologic model with its *a-priori* configuration (i.e. no calibration) shows good skill in reproducing the hydrologic response to rainfall in each of the cases. The variability in the magnitude and timing of the peaks is due to uncertainty from several sources including those in radar rainfall estimates and the hydrologic model itself.

General signatures of flow routing modeling in streamflow hydrographs can be described with the cases shown in Fig. 9. Early and high (overestimated) peaks indicate that, overall, the flood wave is routed too fast (panels c and d), displaying a tendency for “flashy” responses. Late and low (underestimated) peaks indicate that the flood wave is routed too slowly (panel b) and shows attenuated responses. Both types of model behavior have an impact on the detection and prediction of floods in systems that rely on flooding thresholds: too fast flow routing will tend to over predict the occurrence of floods (i.e. increased false alarm rates), while too slow flow routing will tend to under predict the occurrence of floods (i.e. increased miss rates). In addition to the aforementioned cases, there are events that display strong signatures of the interaction between uncertainty in the runoff generation component (i.e. the water balance) and in the flow routing. In panel e) of Fig. 9, there is overestimation of the magnitude with a late peak, which indicates overestimation of excess rainfall in combination with a slow flow routing. On the other hand, panel f) shows a case where the peak is underestimated but occurs early, which indicates underestimation of the excess rainfall and fast flow routing. Lastly, the “ideal” case is presented in panel a) with a near perfect flood wave timing, although minor overestimation of the total volume can be observed.

4.2.2. Event-based evaluation over CONUS

Taking into consideration the aspects discussed in Section 4.2.1, an evaluation of the 47,563 events from the selected 1,672 basins was performed. Histograms of peak time error and relative peak error are shown in Fig. 10. The peak timing obtained from the *a-priori* estimation of routing parameters is remarkably skillful. The peaks tend to be early only 15 to 25 minutes on average. Figure 11 shows the contrast of peak time error to mean concentration time. It can be observed that the conditional median of peak time errors in much smaller than the concentration time, which further illustrates the significant skill of the kinematic wave parameter estimates. Moreover, the standard deviation is about 3.7 hours, which represents a skill arguably acceptable for flash flood forecasting.

The peak magnitude, on the other hand, tends to be underestimated. Furthermore, its frequency displays significant variability indicating that high underestimation can occur. Peak magnitude errors are more likely to be related to water balance uncertainty, in which quantitative precipitation estimates from radar can play a significant role. However, routing could also explain some of the magnitude errors of peak flow as discussed in Section 4.2.1.

5. Summary and conclusions

In this work, a methodology was devised to generate *a-priori* estimates for the parameters of the widely used kinematic wave approximation to the unsteady, 1-D Saint-Venant equations for hydrologic flow routing. The approach is based on an analysis of the conditional distribution of rating curve parameters over the Conterminous United States given a set of geophysical basin characteristics including geomorphology, hydro-climatology, pedology and land cover/land use. The main goal of this study was to enable prediction at ungauged locations through regionalization of model parameters. Key remarks of this work can be summarized as follows:

- The results of this work demonstrate the value of *a-priori* parameter estimation in a successful configuration of a hydrologic modeling system. The expected skill of the flow routing simulations, considering the mean concentration time of the basins and that no calibration was performed, is significantly high for peakflow and timing of peakflow estimation. More importantly, the skill shows consistency as indicated by the large sample verification. Attaining such level of skill and consistency is crucial in extending forecasting capabilities to ungauged locations.
- The resulting grids of *a-priori* estimates can be used in any hydrologic model that employs the kinematic wave model for flow routing. Moreover, the methodology presented in this study enables the estimation of the kinematic wave model parameters anywhere over the globe, thus allowing flood modeling in ungauged basins at regional to global scales.
- Even though the demonstration of the estimates in a hydrologic modeling exercise was deterministic, the multi-dimensional analysis on the kinematic wave parameters considers uncertainty during the estimation of *a-priori* values. This uncertainty information of the estimates can be utilized for probabilistic applications.
- The approach to parameter estimation featured herein combines the power of large sample hydrology, statistical multi-dimensional analysis and physical theory to investigate regional and local controls of the spatial variability of channel characteristics that can be parameterized using the rating curve. The results highlight the importance of regional and local geophysical factors in uniquely defining characteristics of each stream reach conforming to physical theory of fluvial hydraulics.

- An important aspect of this approach is its consistency with the scale of flood and flash flood modeling (commensurability). Furthermore, it addresses challenges in standard methodologies that rely on information whose availability might not be adequate for regional to global modeling, and whose scale is not explicitly resolved at the scale of the application.

Overall, this contribution illustrates the advantages of investigating relationships of model parameters with geophysical variables whose availability in the form of geospatial datasets is increasing. The particular exercise on the kinematic wave parameters leaves room for further development in terms of accuracy and adaptability to different basin physical structures. The latter is specifically needed to extend this work to modeling applications at the global scale. Future research will tackle some of the simplifications of the implementation of the kinematic wave used herein, such as the flow-independent nature of the parameter estimates.

Acknowledgements: Some of the computing for this project was performed at the OU Supercomputing Center for Education & Research (OSCER) at the University of Oklahoma (OU). This work was supported by the Disaster Relief Appropriations Act of 2013 (P.L. 113-2), which funded NOAA research grant NA14OAR4830100 and partial support from HyDROS Lab of National Weather Center, Norman, OK.

References

Akantziliotou, C., R. Rigby, and D. Stasinopoulos, 2002: The R implementation of generalized additive models for location, scale and shape. *Proc.* Proceedings of the Statistical modelling in Society: 17th International Workshop on Statistical Modelling, Chania,

Bedient, P. B., W. C. Huber, and B. E. Vieux, 2008: *Hydrology and floodplain analysis* Prentice-Hall,

Beldring, S., K. Engeland, L. A. Roald, N. R. Sælthun, and A. Voksø, 2003: Estimation of parameters in a distributed precipitation-runoff model for Norway. *Hydrology and Earth System Sciences Discussions*, **7**, 304-316

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

Boyle, D. P., H. V. Gupta, and S. Sorooshian, 2000: Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. *Water Resour. Res.*, **36**, 3663-3674 10.1029/2000wr900207

Chow, V. T., D. R. Maidment, and L. W. Mays, 1988: *Applied hydrology*. McGraw-Hill, Inc.,

Cosby, B., G. Hornberger, R. Clapp, and T. Ginn, 1984: A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. *Water Resources Research*, **20**, 682-690

Costa, J. E., 1987: Hydraulics and basin morphometry of the largest flash floods in the conterminous United States. *Journal of Hydrology*, **93**, 313-338

Dingman, S. L., 2009: *Fluvial hydraulics*. oxford university press New York,

Duan, Q. 2003. Global Optimization for Watershed Model Calibration. In *Calibration of Watershed Models*, edited by Q. Duan, S. Sorooshian, H. V. Gupta, A. N. Rousseau and R. Turcotte. Washington, DC: American Geophysical Union.

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

Feldman, A., 1995: HEC-1 flood hydrograph package. *Computer Models of Watershed Hydrology*, 119-150

Feldman, A., 2000: Hydrologic Modeling System HEC-HMS Technical Reference Manual. U.S. Army Corps of Engineers,

Finnegan, N. J., G. Roe, D. R. Montgomery, and B. Hallet, 2005: Controls on the channel width of rivers: Implications for modeling fluvial incision of bedrock. *Geology*, **33**, 229-232

Fischer, G., F. Nachtergaele, S. Prieler, H. Van Velthuizen, L. Verelst, and D. Wiberg, 2008: Global agro-ecological zones assessment for agriculture (GAEZ 2008). *IIASA, Laxenburg, Austria and FAO, Rome, Italy*,

Ensemble Framework For Flash Flood Forecasting (EF5) version v0.5 (Hydrologic Modeling Framework), Flamig, Z., H. Vergara, and J. J. Gourley. The University of Oklahoma, Norman, Oklahoma, USA, http://ef5.ou.edu.

Fry, J. A., G. Xian, S. Jin, J. A. Dewitz, C. G. Homer, Y. LIMIN, C. A. Barnes, N. D. Herold, and J. D. Wickham, 2011: Completion of the 2006 national land cover database for the conterminous United States. *Photogrammetric Engineering and Remote Sensing*, **77**, 858-864

Gesch, D., G. Evans, J. Mauck, J. Hutchinson, and W. J. Carswell Jr, 2009: The national map: Elevation. *US geological survey fact sheet*, **3053**,

Gourley, J. J., Y. Hong, Z. L. Flamig, A. Arthur, R. Clark, M. Calianno, I. Ruin, T. Ortel, M. E. Wieczorek, and P.-E. Kirstetter, 2013: A unified flash flood database across the United States. *Bulletin of the American Meteorological Society*, **94**, 799-805

Gupta, H. V., C. Perrin, G. Bloschl, A. Montanari, R. Kumar, M. Clark, and V. Andréassian, 2014: Large-sample hydrology: a need to balance depth with breadth. *Hydrology and Earth System Sciences*, **18**, p. 463-p. 477

Gupta, H. V., S. Sorooshian, T. S. Hogue, and D. P. Boyle. 2003. Advances in Automatic Calibration of Watershed Models. In *Calibration of Watershed Models*, edited by Q. Duan, S. Sorooshian, H. V. Gupta, A. N. Rousseau and R. Turcotte. Washington, DC: American Geophysical Union.

Hey, T. 2012. The Fourth Paradigm: Data-Intensive Scientific Discovery. In *E-Science and Information Management*: Springer.

Hrachowitz, M., H. H. G. Savenije, G. Blöschl, J. J. McDonnell, M. Sivapalan, J. W. Pomeroy, B. Arheimer, T. Blume, M. P. Clark, U. Ehret, F. Fenicia, J. E. Freer, A. Gelfan, H. V. Gupta, D. A. Hughes, R. W. Hut, A. Montanari, S. Pande, D. Tetzlaff, P. A. Troch, S. Uhlenbrook, T. Wagener, H. C. Winsemius, R. A. Woods, E. Zehe, and C. Cudennec, 2013: A decade of Predictions in Ungauged Basins (PUB)—a review. *Hydrological Sciences Journal*, **58**, 1198-1255 10.1080/02626667.2013.803183

Huber, W., and V. Singh, 1995: EPA Storm Water Management Model-SWMM. *Computer models of watershed hydrology.*, 783-808

Kazezyilmaz-Alhan, C., and M. Medina, 2007: Kinematic and Diffusion Waves: Analytical and Numerical Solutions to Overland and Channel Flow. *Journal of Hydraulic Engineering*, **133**, 217-228 doi:10.1061/(ASCE)0733-9429(2007)133:2(217)

Koren, V., S. Reed, M. Smith, Z. Zhang, and D. J. Seo, 2004: Hydrology laboratory research modeling system (HL-RMS) of the US national weather service. *Journal of Hydrology*, **291**, 297-318

Koren, V., J. Schaake, Q. Duan, M. Smith, and S. Cong, 1998: PET Upgrades to NWSRFS, Project Plan. In *Unpublished Report*,

Koren, V., M. Smith, D. Wang, and Z. Zhang, 2000: Use of soil property data in the derivation of conceptual rainfall-runoff model parameters. *Proc.* 15th Conference on Hydrology, AMS, 2, Long Beach, CA, 103–106

Liu, Z., and E. Todini, 2002: Towards a comprehensive physically-based rainfall-runoff model. *Hydrology and Earth System Sciences Discussions*, **6**, 859-881

Miller, D., and R. A. White, 1998: A Conterminous United States Multi-Layer Soil Characteristics Data Set for Regional Climate and Hydrology Modeling. Earth Interactions,

Mockus, V., 1961: Watershed lag. U.S. Dept. of Agriculture, Soil Conservation Service,

Montgomery, D. R., and K. B. Gran, 2001: Downstream variations in the width of bedrock channels. *Water Resources Research*, **37**, 1841-1846

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

Moussa, R., and C. Bocquillon, 2000: Approximation zones of the Saint-Venant equations f flood routing with overbank flow. *Hydrology and Earth System Sciences Discussions*, **4**, 251-260

Naden, P., P. Broadhurst, N. Tauveron, and A. Walker, 1999: River routing at the continental scale: use of globally-available data and an a priori method of parameter estimation. *Hydrology and Earth System Sciences Discussions*, **3**, 109-123

Nash, J., 1957: The form of the instantaneous unit hydrograph. *IAHS Publ*, **45**, 114-121

NSSL, 2016: The Flooded Locations And Simulated Hydrographs (FLASH) project, <http://blog.nssl.noaa.gov/flash/>, Cited

NWS, 2010: Southeast United States Floods, September 18-23, 2009.

Pokhrel, P., H. Gupta, and T. Wagener, 2008: A spatial regularization approach to parameter estimation for a distributed watershed model. *Water Resources Research*, **44**, W12419

Ponce, V., 1986: Diffusion Wave Modeling of Catchment Dynamics. *Journal of Hydraulic Engineering*, **112**, 716-727 doi:10.1061/(ASCE)0733-9429(1986)112:8(716)

Ponce, V. M., 1991: Kinematic wave controversy. *Journal of Hydraulic Engineering*, **117**, 511-525

PRISM Climate Group, July 2012: 30-yr Normals, 1981 - 2010, Oregon State University, http://prism.oregonstate.edu, Accessed September, 2013

Reed, S., J. Schaake, and Z. Zhang, 2007: A distributed hydrologic model and threshold frequency-based method for flash flood forecasting at ungauged locations. *Journal of Hydrology*, **337**, 402-420 10.1016/j.jhydrol.2007.02.015

Reggiani, P., E. Todini, and D. Meißner, 2014: Analytical solution of a kinematic wave approximation for channel routing. *Hydrology Research*, **45**, 43-57

Reitan, T., and A. Petersen-Øverleir, 2009: Bayesian methods for estimating multi-segment discharge rating curves. *Stochastic Environmental Research and Risk Assessment*, **23**, 627-642

Rigby, R., and D. Stasinopoulos, 2001: The GAMLSS project: a flexible approach to statistical modelling. *Proc.* New trends in statistical modelling: Proceedings of the 16th international workshop on statistical modelling, 337-345

Rigby, R. A., and D. M. Stasinopoulos, 2005: Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **54**, 507-554

Ryan, S. E., and L. S. Porth, 2007: A tutorial on the piecewise regression ap- proach applied to bedload transport data. In *Gen. Tech. Rep. RMRS-GTR-189*, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station., 41 p.

Schumm, S. A., 1956: Evolution of drainage systems and slopes in badlands at Perth Amboy, New Jersey. *Geological Society of America Bulletin*, **67**, 597-646

Sivapalan, M., K. Takeuchi, S. Franks, V. Gupta, H. Karambiri, V. Lakshmi, X. Liang, J. McDonnell, E. Mendiondo, and P. O'connell, 2003: IAHS Decade on Predictions in Ungauged Basins (PUB), 2003,Äì2012: Shaping an exciting future for the hydrological sciences. *Hydrological Sciences Journal*, **48**, 857-880

Soil Survey Staff, 1994: State Soil Geographic Database (STATSGO) data users guide. In *USDA Natural Resources Conservation Service Misc. Publ. 1492*,

Sorooshian, S., Q. Duan, and V. K. Gupta, 1993: Calibration of rainfall-runoff models: Application of global optimization to the Sacramento Soil Moisture Accounting Model. *Water Resour. Res.*, **29**, 1185-1194 10.1029/92wr02617

Stasinopoulos, D. M., and R. A. Rigby, 2007: Generalized additive models for location scale and shape (GAMLSS) in R. *Journal of Statistical Software*, **23**, 1-46

Vélez, J., M. Puricelli, F. López Unzu, and F. Francés, 2009: Parameter extrapolation to ungauged basins with a hydrological distributed model in a regional framework. *Hydrology and Earth System Sciences*, **13**, 229-246

Vergara, H., P. Kirstetter, Y. Hong, J. Gourley, and X. Wang, 2013: Impact of Uncertainty Characterization of Satellite Rainfall Inputs and Model Parameters on Hydrological Data Assimilation with the Ensemble Kalman Filter for Flood Prediction. *Proc.* AGU Fall Meeting Abstracts, 1, 1306

Vrugt, J. A., C. J. F. Braak, H. V. Gupta, and B. A. Robinson, 2008: Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? *Stochastic Environmental Research and Risk Assessment*, **23**, 1011-1026 10.1007/s00477-008-0274-y

Vrugt, J. A., H. V. Gupta, W. Bouten, and S. Sorooshian. 2003. A Shuffled Complex Evolution Metropolis Algorithm for Estimating Posterior Distribution of Watershed Model Parameters. In *Calibration of Watershed Models*, edited by Q. Duan, S. Sorooshian, H. V. Gupta, A. N. Rousseau and R. Turcotte. Washington, DC: American Geophysical Union.

Vrugt, J. A., H. V. Gupta, S. C. Dekker, S. Sorooshian, T. Wagener, and W. Bouten, 2006: Application of stochastic parameter optimization to the Sacramento Soil Moisture Accounting model. *Journal of Hydrology*, **325**, 288-307 10.1016/j.jhydrol.2005.10.041

Wang, J., Y. Hong, L. Li, J. J. Gourley, S. I. Khan, K. K. Yilmaz, R. F. Adler, F. S. Policelli, S. Habib, and D. Irwn, 2011: The coupled routing and excess storage (CREST) distributed hydrological model. *Hydrological Sciences Journal*, **56**, 84-98

Wei, Y., S.-K. Santhana-Vannan, and R. B. Cook, 2009: Discover, visualize, and deliver geospatial data through OGC standards-based WebGIS system. *Proc.* geoinformatics, 2009 17th international conference on, 1-6

Wigmosta, M. S., L. W. Vail, and D. P. Lettenmaier, 1994: A distributed hydrology-vegetation model for complex terrain. *Water Resour. Res.*, **30**, 1665-1679 10.1029/94wr00436

Woolhiser, D. A., R. Smith, and D. C. Goodrich, 1990: *KINEROS: a kinematic runoff and erosion model: documentation and user manual*. US Department of Agriculture, Agricultural Research Service,

Yao, C., Z. Li, Z. Yu, and K. Zhang, 2012: A priori parameter estimates for a distributed, grid-based Xinanjiang model using geographically based information. *Journal of Hydrology*, **468**, 47-62

Zhang, J., K. Howard, C. Langston, B. Kaney, Y. Qi, L. Tang, H. Grams, Y. Wang, S. Cocks, and S. Martinaitis, 2015: Multi-Radar Multi-Sensor (MRMS) Quantitative Precipitation Estimation: Initial Operating Capabilities. *Bulletin of the American Meteorological Society*,

Zhang, J., K. Howard, C. Langston, S. Vasiloff, B. Kaney, A. Arthur, S. Van Cooten, K. Kelleher, D. Kitzmiller, F. Ding, D.-J. Seo, E. Wells, and C. Dempsey, 2011: National Mosaic and Multi-Sensor QPE (NMQ) System: Description, Results, and Future Plans. *Bulletin of the American Meteorological Society*, **92**, 1321-1338 10.1175/2011bams-d-11-00047.1

Zhao, R., X. Liu, and V. Singh, 1995: The Xinanjiang model. *Computer models of watershed hydrology.*, 215-232

Zhao, R., Y. Zhang, L. Fang, X. Liu, and Q. Zhang, 1980: The Xinganjiang Model. *Proc.* Hydrological Forecasting Proceedings Oxford Symposium, IASH 129, 351-356

Figure 1: Applicability of the kinematic wave approximation over the Conterminous United States based on slope. The slope grid is based on a 1-km Digital Elevation Model (DEM) grid.

Figure 2: Power fit to rating curve data for streamflow (x-axis) and cross-section area (y-axis) measured in the field for USGS stations: a) 01118010 (~531 km^2) and b) 02083500 (~5654 km^2). The dots correspond to the field measurements and the dashed line to the power law regression fit.

Figure 3: Cumulative distribution of drainage areas over CONUS, computed from the 1-km drainage area grid.

Figure 4: Spatial distribution of rating curve parameters for the catchments of the selected USGS stream gauges over the CONUS: a) α in log scale; and b) β.

Figure 5: Sample of geospatial datasets used in the analysis of spatial variability of rating curve parameters: a) Relief ratio (log scale); b) K factor (Erodability); c) Mean annual precipitation (log scale; mm/year); d) Mean temperature (Celsius); e) Mean rock volume percent (log scale; %); and f) Runoff Curve Number.

Figure 6: A sample of the obtained results from the analysis of associations of kinematic wave model parameters to geophysical variables.

Figure 7: Evaluation of the *goodness-of-fit* of the GAMLSS model estimates of kinematic wave model parameters α and β : a) Scatter density plots of the reference rating curve parameter values and estimates produced with GAMLSS for α; b) Same as a) but for β; c) Conditional percentile plot of α estimates given reference rating curve parameter values; and d) Same as c) but for β. The 1-to-1 line and values of linear correlation coefficient are included for each fit in panels a) and b).

Figure 8: Samples of a) α *a-priori* estimates and b) β *a-priori* estimates, c) standard deviation of α *a-priori* estimates and d) standard deviation of β *a-priori* estimates. Standard deviation colormaps are stretched to 2% and 98% percentiles.

Figure 9: Sample hydrographs showing different simulated flow routing skill signatures. The hydrographs correspond to events occurred during September of 2009 on the Southeast of the United States: a) near perfect routing (Mississippi), b) late and low peak (Arkansas), c) early and high peak (Tennessee), d) early and high peak (Tennessee), e) late and high peak (Georgia) and f) early and low peak (near Atlanta, Georgia).

Figure 10: Histograms of the a) Peak Time Error (hours) and b) Relative Peak Error (%) for the approximately 47,563 events. Measures of location and scale are included for each case.

Figure 11: Quantile plot of conditional distributions of peak time errors with respect to the mean concentration time of the basins. The diagonal dashed lines show the direct (upper line) and indirect (lower line) 1-to-1 relationships. The gradients of gray area depict different distribution bounds: 1 – 99th percentiles, 5 – 95th percentiles, 10 – 90th percentiles and 25 – 75th percentiles. The black solid line represents the median of the conditional distributions.

Table 1: List of past study cases of kinematic wave application.

Model Name	**Institution**	**Reference**
Hydrology Laboratory's Distributed Hydrologic Model (HL-DHM)	Office of Hydrologic Development, National Weather Service	Koren et al. 2004
Hydrologiska Byråns Vattenbalansavdelning (HBV) model	Norwegian Water Resources and Energy Directorate, Norway	Beldring et al. 2003
TOPographic Kinematic APproximation and Integratio (TOPKAPI)	University of Bologna, Italy	Liu and Todini 2002
Hydrologic Engineering Center's Hydrologic Modeling System HEC-HMS	U.S. Army Corp of Engineers	Feldman 2000
Storm Water Management Model (SWMM)	Environmental Protection Agency	Huber and Singh 1995
Hydrologic Engineering Center's Flood Hydrograph Package (HEC-1)	U.S. Army Corp of Engineers	Feldman 1995
Distributed Hydrology-Vegetation Model (DHVM)	University of Washington	Wigmosta et al. 1994
KINEmatic Runoff and EROSion (KINEROS)	U.S. Department of Agriculture	Woolhiser et al. 1990

Table 2: CREST model parameters and *a-priori* estimates

Parameter (Units)	Description	Range	Source
PWM (mm)	Soil water capacity	0 - 690	STATSGO dataset (Miller and White 1998)
PIM (%)	Percent of impervious surface area	0 – 100	URB_2000 - built-up land (residential and infrastructure)" From Harmonic World Soil Database (HWSD; Fischer et al. 2008)
PB (-)	Infiltration curve exponent	0 – 11.55	STATSGO dataset (Miller and White 1998) and look-up table in Cosby et al. (1984)
PFC (mm/h)	Hydraulic conductivity	0 – 50.8	STATSGO dataset (Miller and White 1998)
UNDER (m/h)	Speed of subsurface flow	0 – 0.051	A scaled value of Hydraulic Conductivity (PFC parameter above)
LEAKI (-)	Interflow linear reservoir leakage factor	0 - 1	STATSGO dataset (Miller and White 1998) and empirical relationship based on CN number (Pokhrel et al. 2008)
COEM (-)	Inverse of Manning's coefficient for overland routing	8.3 – 66.67	UMD vegetation category from 2007 MODIS (Wei et al. 2009)
PKE (-)	Linear adjustment factor on Potential Evapotranspiration	0 - 1	Set to 1.0
TH (# of grid cells)	A threshold number of grid cells above which a pixel is defined as a stream	-	Set to 5.0

Table 3: Statistical significance of explanatory variables in GAMLSS model. Not retained or not considered variable are marked with '-'. Significance is expressed as a probability of rejection.

Variable (Units)	α	β
Basin Area (km^2)	0	-
Elongation Ratio (-)	0.001	-
Relief Ratio (-)	0	0
Slope Index (-)	0.001	-
Slope to Outlet (-)	0.001	0.001
Annual Precipitation (mm/yr)	0	0
Mean Temperature (Celsius)	0	0
K Factor (Erodability)	0	0
Depth-to-Rock (cm)	0.001	-
Rock Volume (%)	0	0
Soil Texture (*b* parameter)	0.05	-
Curve Number (-)	0.001	0
River Length (m)	-	0

Table 4: Score values of goodness-of-fit for GAMLSS models for α and β.

Summary of the Quantile Residuals	α	β	**Ideal - Gaussian**
Mean	0.03	-0.01	0.00
Variance	1.00	1.00	1.00
Skewness	0.38	0.03	0.00
Kurtosis	3.36	3.41	3.00
Filliben Correlation	0.99	1.00	1.00

Table 5: Explanatory variables retained by GAMLSS. The minimum, mean and maximum values of each variable are included for the training and prediction datasets.

Variable (Units)	**Training Dataset**			**Prediction Dataset**		
	Min	**Mean**	**Max**	**Min**	**Mean**	**Max**
Basin Area (km^2)	1	2,421	2,926,080	0.71	804	3,138,200
Elongation Ratio (-)	0.262	0.819	2.718	0.197	1.104	7.899
Relief Ratio (-)	8×10^{-6}	0.022	0.421	0	0.020	1.099
Slope Index (-)	2×10^{-5}	0.012	0.375	0	0.032	1.417
Slope to Outlet (-)	2×10^{-4}	0.023	0.208	0.000	0.037	3.005
Annual Precipitation (mm/yr)	121	1,053	4,463	2.8×10^{-3}	792	5,675
Mean Temperature (Celsius)	0.0	11.0	22.9	-5.5	11.0	25.5
K Factor (-)	0.000	0.256	0.640	0.000	0.259	0.640
Depth-to-Rock (cm)	9	130	176	9	125	191
Rock Volume (%)	0	12	100	0	14	100
Soil Texture (*b* parameter)	2.79	5.29	11.55	2.79	5.49	11.5
Curve Number (-)	8	70	92	0	70	100
River Length (m)	10,071	68,879	5,282,430	638	10,506	5,440,000

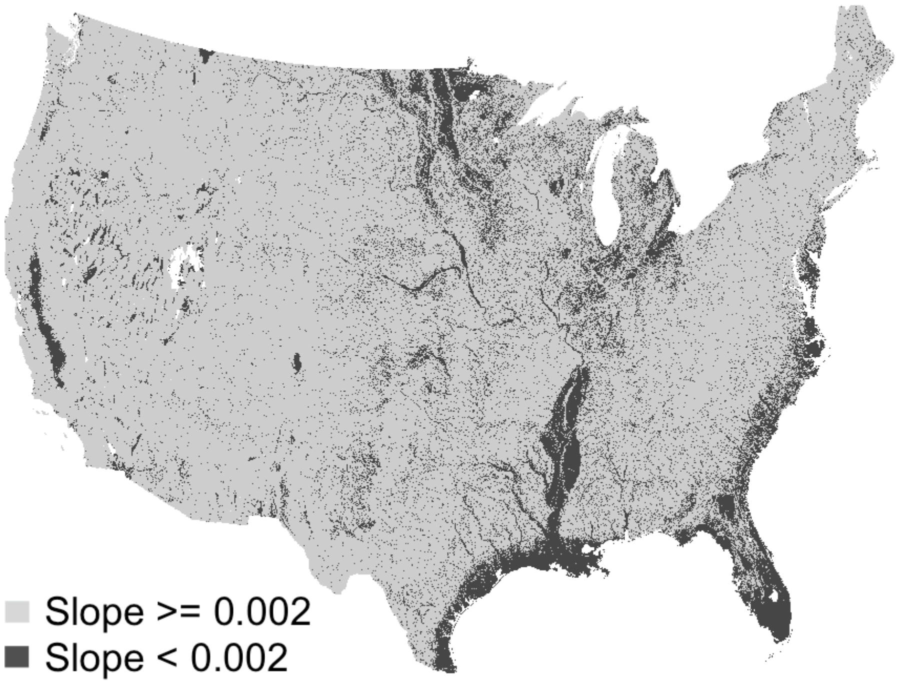

Figure 1: Applicability of the kinematic wave approximation over the Conterminous United States based on slope. The slope grid is based on a 1-km Digital Elevation Model (DEM) grid.

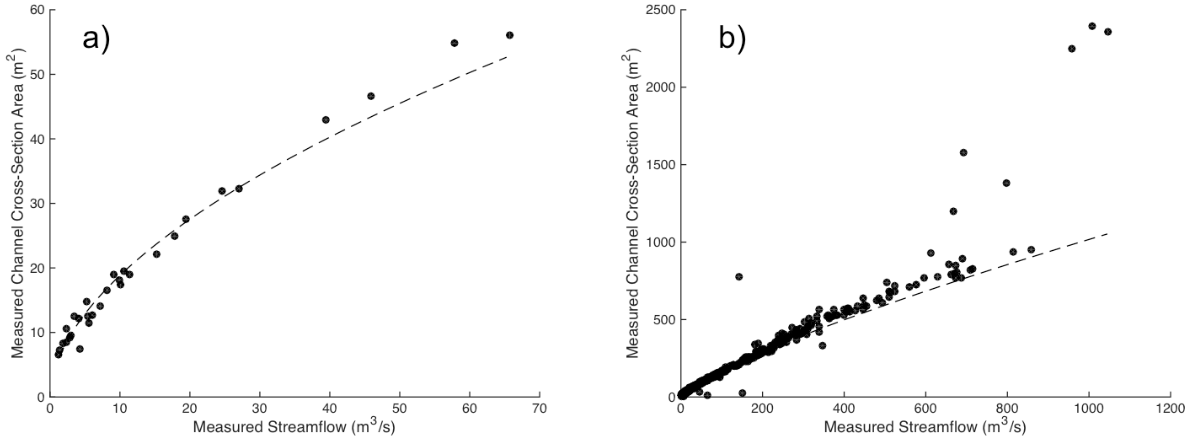

Figure 2: Power fit to rating curve data for streamflow (x-axis) and cross-section area (y-axis) measured in the field for USGS stations: a) 01118010 (~531 km^2) and b) 02083500 (~5654 km^2). The dots correspond to the field measurements and the dashed line to the power law regression fit.

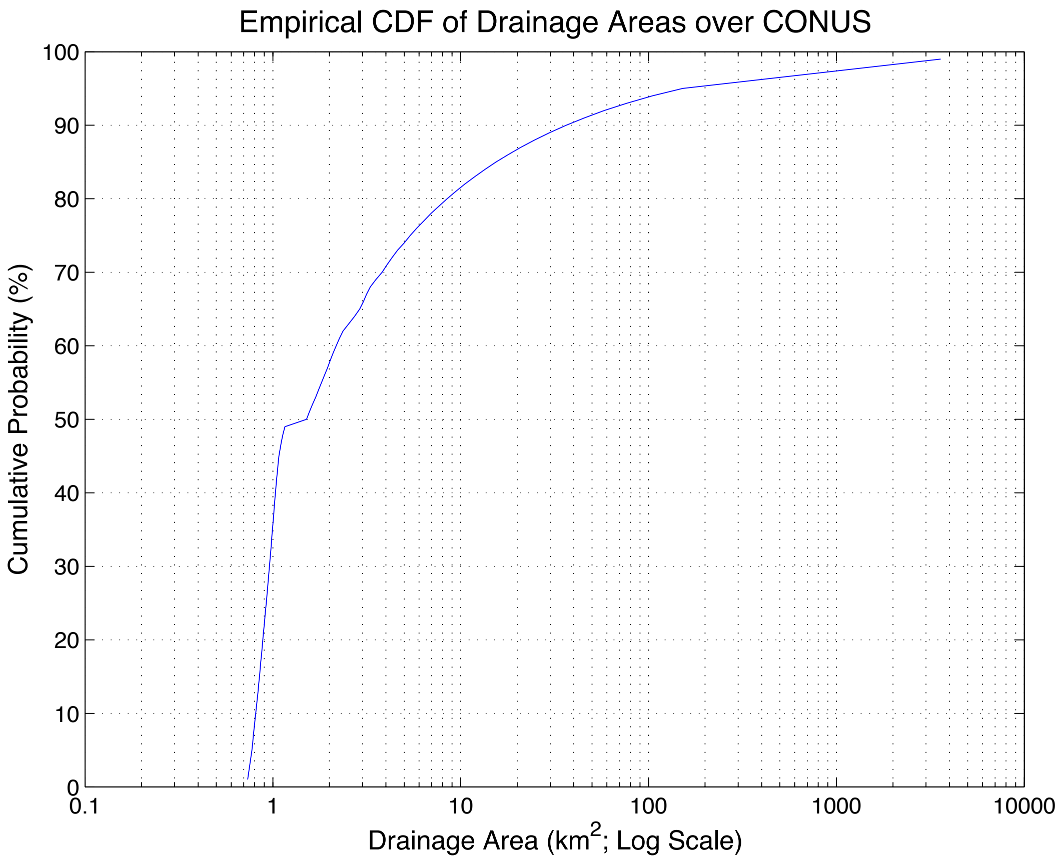

Figure 3: Cumulative distribution of drainage areas over CONUS, computed from the 1-km drainage area grid.

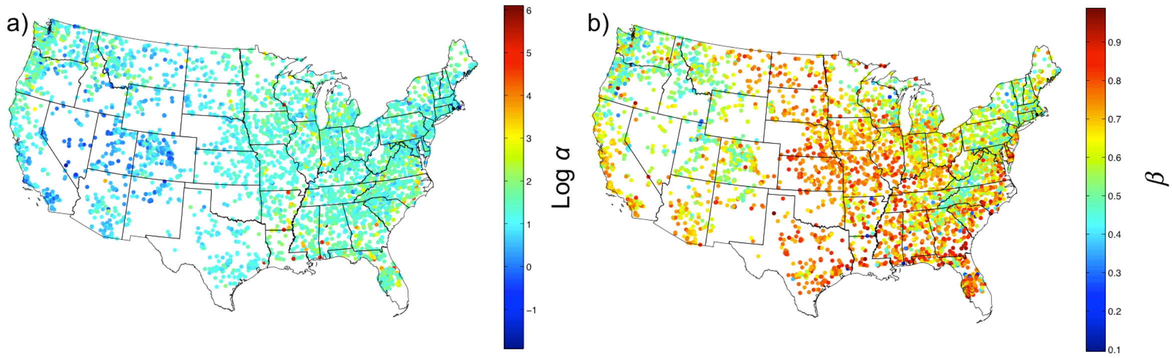

Figure 4: Spatial distribution of rating curve parameters for the catchments of the selected USGS stream gauges over the CONUS: a) α in log scale; and b) β.

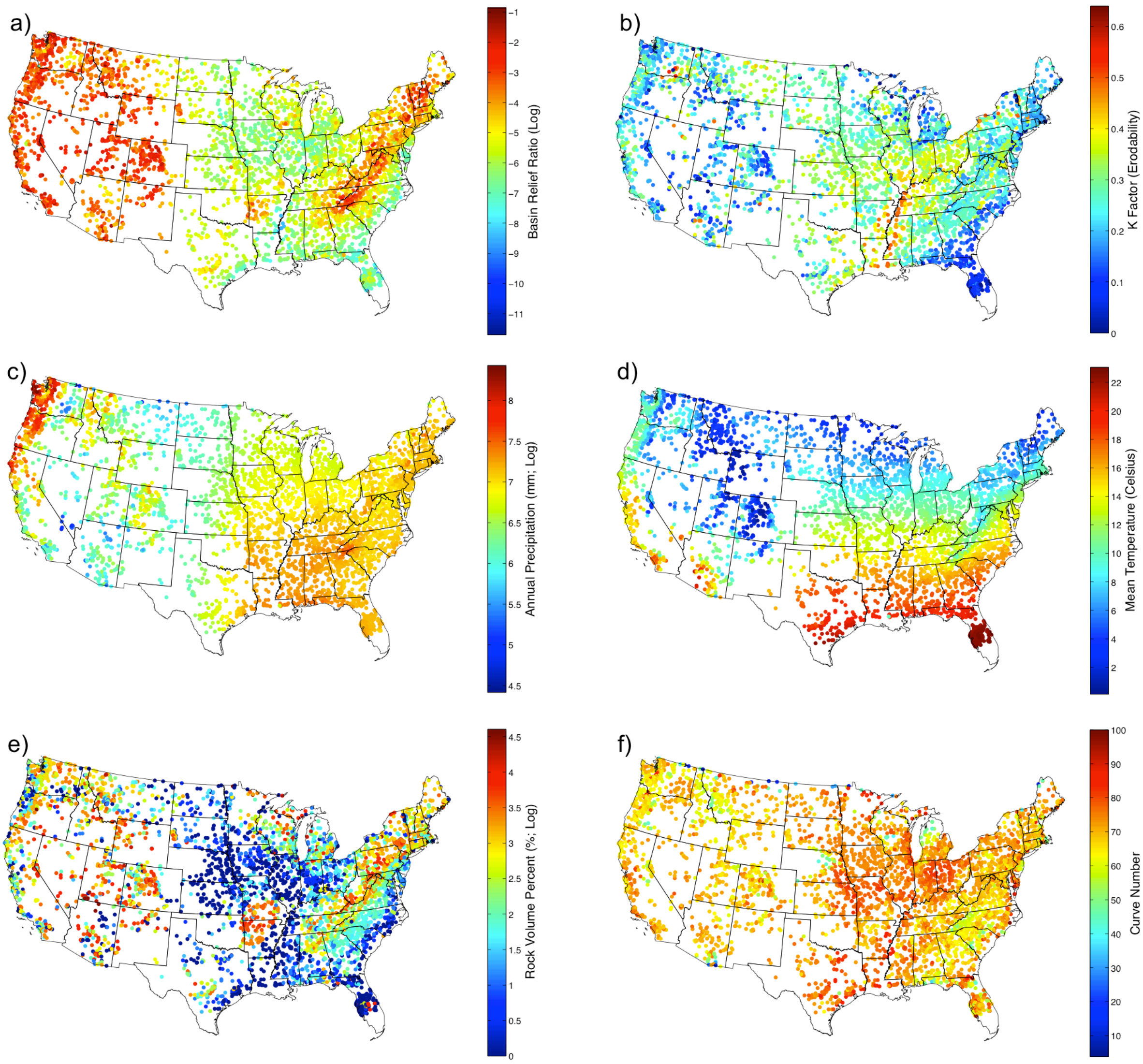

Figure 5: Sample of geospatial datasets used in the analysis of spatial variability of rating curve parameters: a) Relief ratio (log scale); b) K factor (Erodability); c) Mean annual precipitation (log scale; mm/year); d) Mean temperature (Celsius); e) Mean rock volume percent (log scale; %); and f) Runoff Curve Number.

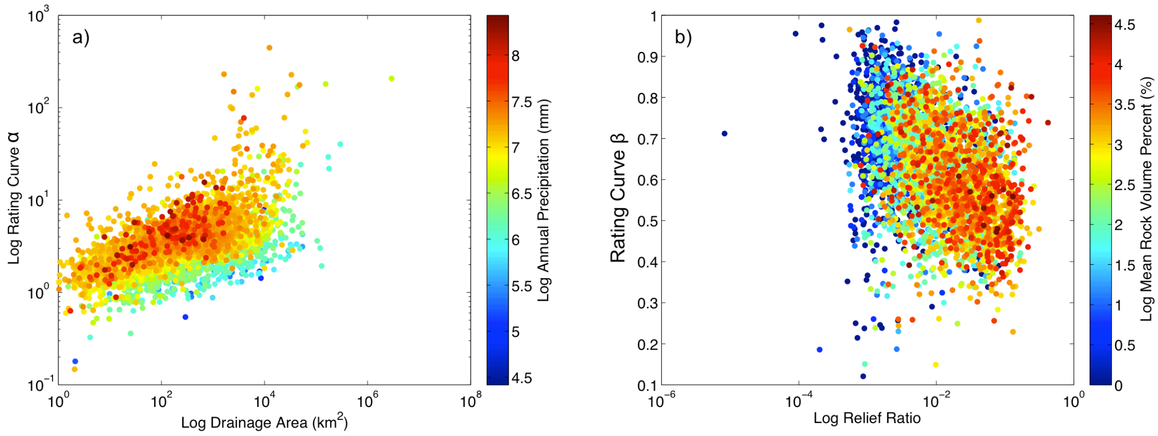

Figure 6: A sample of the obtained results from the analysis of associations of kinematic wave model parameters to geophysical variables.

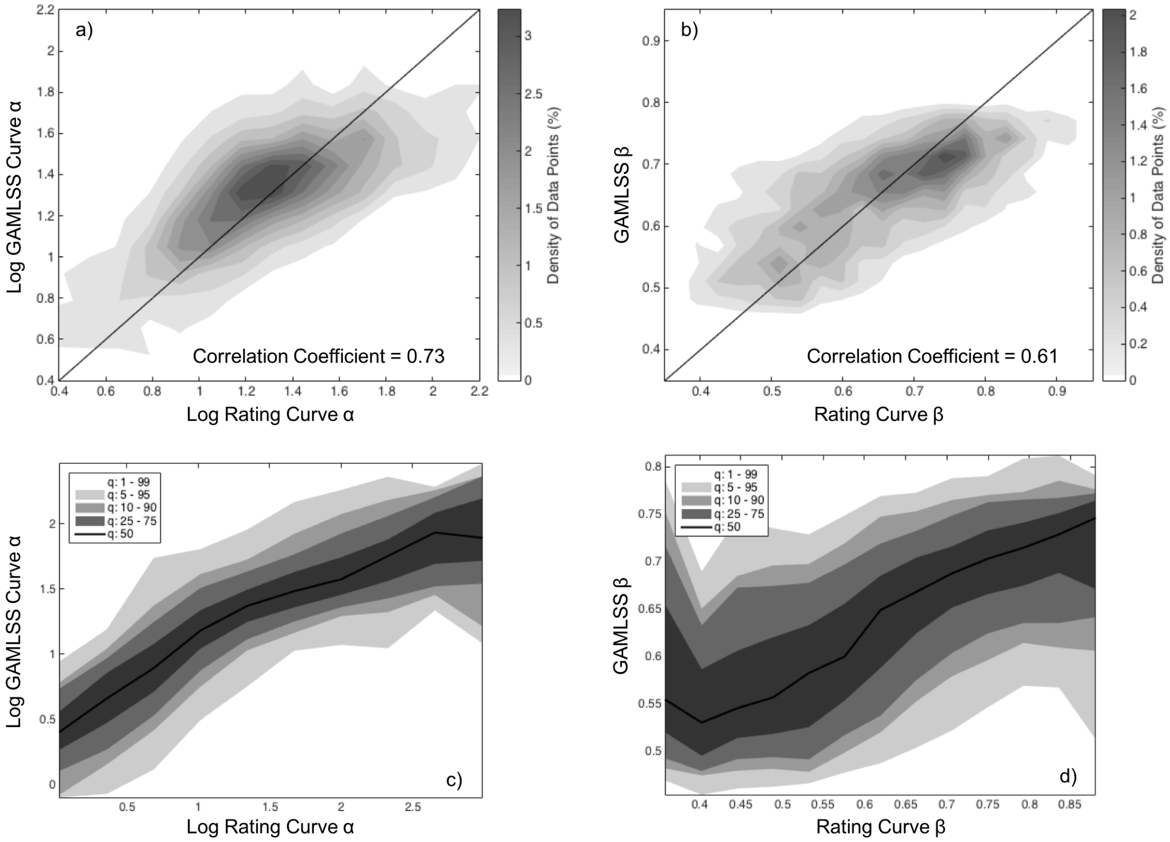

Figure 7: Evaluation of the *goodness-of-fit* of the GAMLSS model estimates of kinematic wave model parameters α and β : a) Scatter density plots of the reference rating curve parameter values and estimates produced with GAMLSS for α; b) Same as a) but for β; c) Conditional percentile plot of α estimates given reference rating curve parameter values; and d) Same as c) but for β. The 1-to-1 line and values of linear correlation coefficient are included for each fit in panels a) and b).

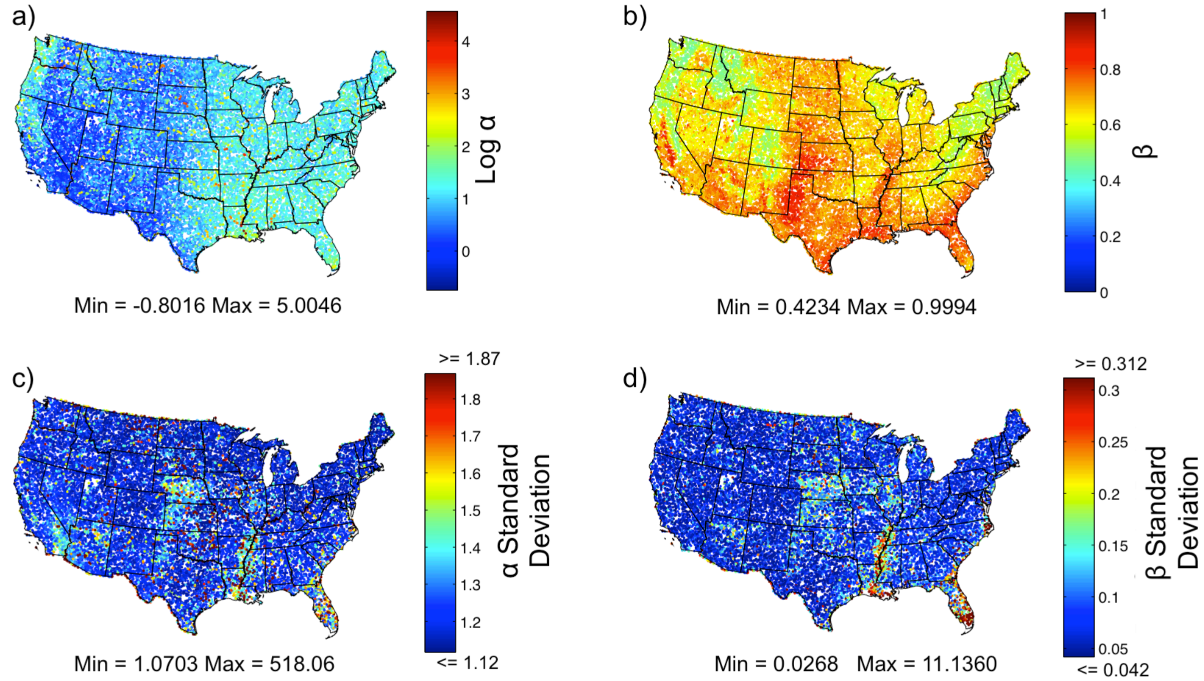

Figure 8: Samples of a) α *a-priori* estimates and b) β *a-priori* estimates, c) standard deviation of α *a-priori* estimates and d) standard deviation of β *a-priori* estimates. Standard deviation colormaps are stretched to 2% and 98% percentiles.

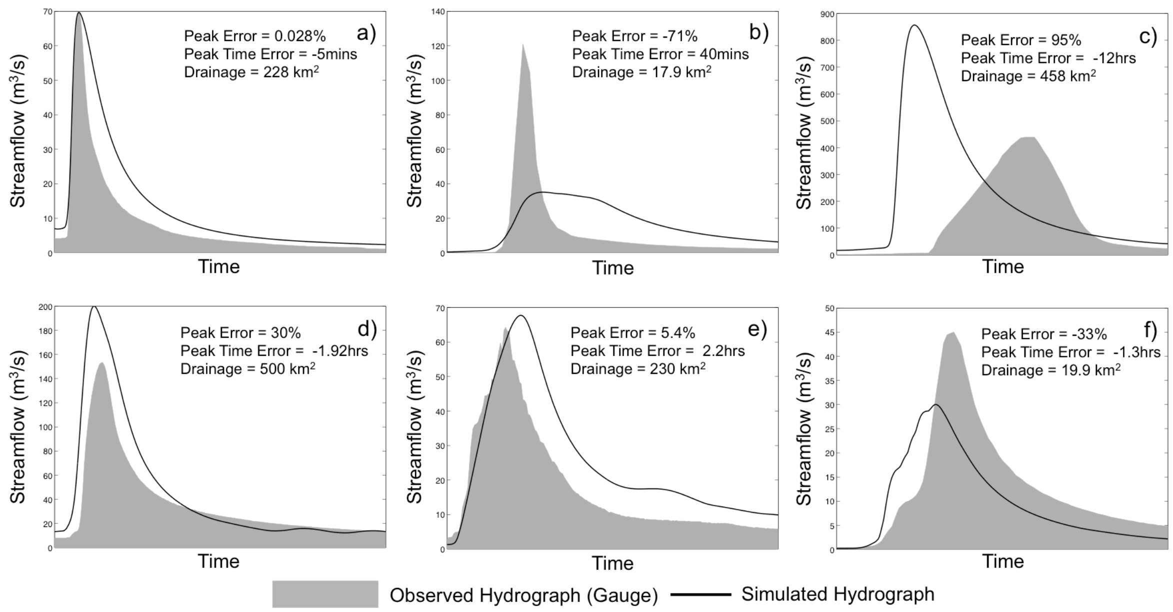

Figure 9: Sample hydrographs showing different simulated flow routing skill signatures. The hydrographs correspond to events occurred during September of 2009 on the Southeast of the United States: a) near perfect routing (Mississippi), b) late and low peak (Arkansas), c) early and high peak (Tennessee), d) early and high peak (Tennessee), e) late and high peak (Georgia) and f) early and low peak (near Atlanta, Georgia).

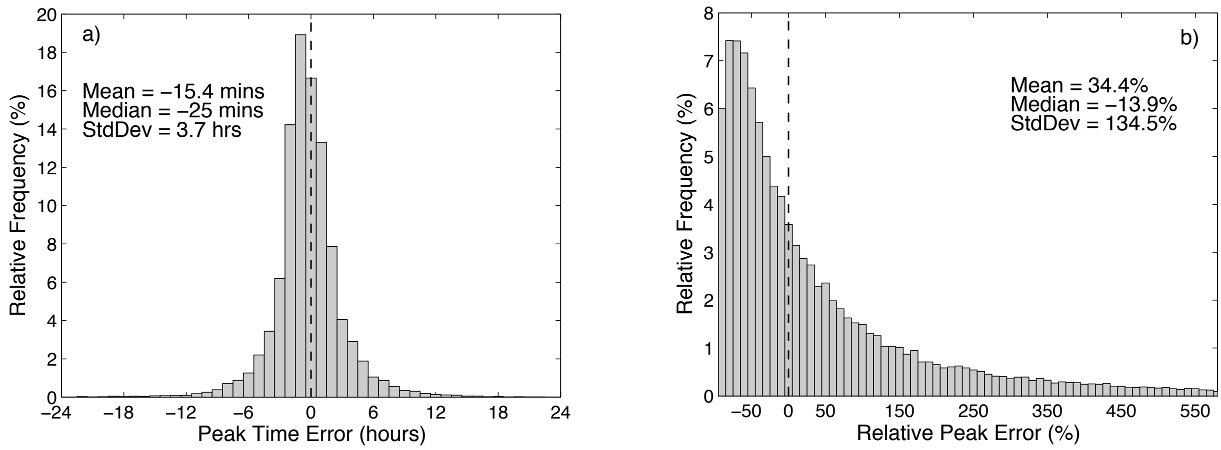

Figure 10: Histograms of the a) Peak Time Error (hours) and b) Relative Peak Error (%) for the approximately 47,563 events. Measures of location and scale are included for each case.

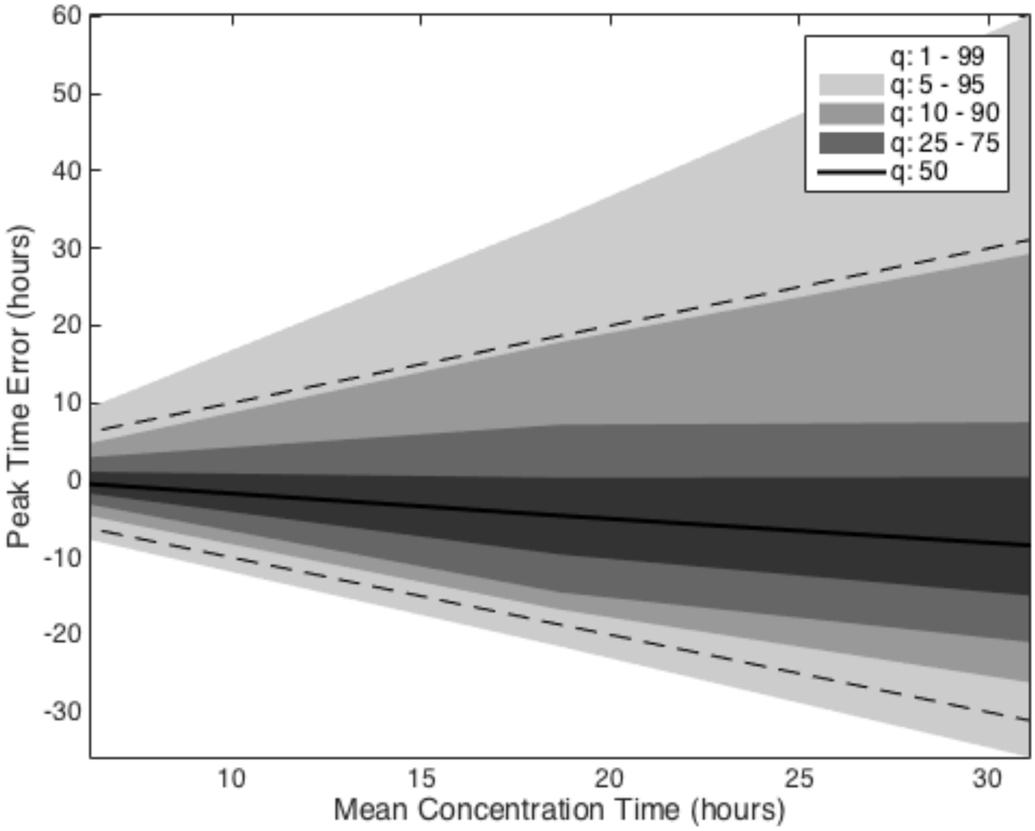

Figure 11: Quantile plot of conditional distributions of peak time errors with respect to the mean concentration time of the basins. The diagonal dashed lines show the direct (upper line) and indirect (lower line) 1-to-1 relationships. The gradients of gray area depict different distribution bounds: 1 – 99th percentiles, 5 – 95th percentiles, 10 – 90th percentiles and 25 – 75th percentiles. The black solid line represents the median of the conditional distributions.