MID-AMERICA TRANSPORTATION CENTER

Report # MATC-UI: 145-3

Final Report
WBS: 25-1121-0005-145-3

# Real-Time Flood Forecasting for River Crossings - Phase III

**Witold Krajewski, PhD**
Rose & Joseph Summers Chair in Water Resources Engineering
Faculty Researcher Engineer, IIHR - Hydroscience & Engineering
Director, Iowa Flood Center
The University of Iowa

**Nicolas Velasquez, PhD**
Assistant Researcher, Iowa Flood Center
Department of Civil and Environmental Engineering
The University of Iowa

2022

A Cooperative Research Project sponsored by
U.S. Department of Transportation- Office of the Assistant Secretary for Research and Technology

# Real-Time Flood Forecasting for River Crossings – Phase III

Witold Krajewski, PhD, P.I.
Rose & Joseph Summers Chair in Water Resources Engineering
Faculty Research Engineer, IIHR - Hydroscience & Engineering
Director, Iowa Flood Center
The University of Iowa

Nicolas Velasquez, PhD, P.I.
Assistant Researcher, Iowa Flood Center
Department of Civil and Environmental Engineering
The University of Iowa

A Report on Research Sponsored by

Mid-America Transportation Center


University of Nebraska–Lincoln

August 2020

## Technical Report Documentation Page

| | | | |
|---|---|---|---|
| 1. Report No.<br>25-1121-0005-145-3 | 2. Government Accession No. | 3. Recipient's Catalog No. | |
| 4. Title and Subtitle<br>Real-Time Flood Forecasting for River Crossings | | 5. Report Date<br>August 2020 | |
| | | 6. Performing Organization Code | |
| 7. Author(s)<br>Witold Krajewski, PhD ORCID: 0000-0002-3477-9281<br>Nicolas Velasquez, PhD ORCID: 0000-0001-8207-5492 | | 8. Performing Organization Report No.<br>25-1121-0005-145-3 | |
| 9. Performing Organization Name and Address<br>Mid-America Transportation Center<br>Prem S. Paul Research Center at Whittier School<br>2200 Vine St.<br>Lincoln, NE 68583-0851 | | 10. Work Unit No. (TRAIS) | |
| | | 11. Contract or Grant No.<br>69A3551747107 | |
| 12. Sponsoring Agency Name and Address<br>Office of the Assistant Secretary for Research and Technology<br>1200 New Jersey Ave., SE<br>Washington, D.C. 20590 | | 13. Type of Report and Period Covered<br>Final Report, January 2019 – August 2020 | |
| | | 14. Sponsoring Agency Code<br>MATC TRB RiP No. 91994-49 | |
| 15. Supplementary Notes | | | |
| 16. Abstract<br>We have developed a generic prototype of a flood-forecasting model that is transferable to other locations around the Midwest to provide monitoring and forecasting flood potential at critical infrastructure points, such as bridges, where streamflow gauges are not available. A real-time web-based visualization platform to display the model predictions has been implemented. The platform will display the river network upstream from a point of interest and a time control slider that will allow exploring the evolution of flows everywhere in the network over the past several days, and about a week into the future. The model uses in-house developed radar-rainfall maps updated every 5 minutes with the spatial resolution of about 0.5 km currently covering the Iowa domain and extending some 100 km into the neighboring states. For the future rainfall, we use predictions for the National Weather Service High-Resolution Rapid Refresh (HRRR) forecasting system. The system provides hourly accumulation products for up to 20 hours ahead. Our system expands the forecasting capabilities of the current NWS by providing predictions at locations that have not been historically gauged. | | | |
| 17. Distribution Statement | | | |
| 18. Security Classif. (of this report)<br>Unclassified | 19. Security Classif. (of this page)<br>Unclassified | 20. No. of Pages<br>51 | 21. Price |

Table of Contents

List of Figures

## Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

Abstract

We have developed a generic prototype of a flood-forecasting model that is transferable to other locations around the Midwest to provide monitoring and forecasting flood potential at critical infrastructure points, such as bridges, where streamflow gauges are not available. Our efforts have centered around creating tools and protocols that would facilitate the implementation of the hydrological model in any of the four MATC states. The protocols include 1) a methodology to use existing regional data to determine the parameters in the runoff routing equation along the river network, 2) a diagnostic methodology to determine infiltration parameters controlling rainfall-runoff transformation, and 3) technology transfer between the University of Iowa and the University of Nebraska. In this phase, we focus our work in the implementation and validation of the methodologies developed during the last phase.

Chapter 1 Preliminaries: The Iowa Flood Center HLM hydrological model

The Iowa Flood Center hydrological model, Hillslope-Link Model (HLM), is a distributed hillslope-scale rainfall-runoff model that partitions Iowa into over three million individual control volumes following the landscape decomposition outlined in Mantilla and Gupta (2005). The model is parsimonious, meaning it uses ordinary differential equations to describe transport between adjacent control volumes. This characteristic reduces the computational resources needed by capturing the most essential features of the rainfall runoff transformation; it uses only a few parameters to obtain acceptable results. The model partitions the river network into river links (the portion of a river channel between two junctions of a river network) and the landscape into hillslopes (adjacent areas that drain into the links).


(a) (b)

**Figure 1.1** (a) illustration of landscape decomposition into hillslopes and decomposition of the river network into channel link and (b) vertical soil profile and control volumes included in the hydrological model

Mass conservation equations give rise to the system of coupled nonlinear ordinary differential equations that represent changes in the water storage in the hillslope surface ($s_{surf}$), top soil ($s_{tops}$), and deep soil ($s_{deeps}$) given by

$$\begin{cases} \dfrac{ds_{surf}(t)}{dt} = p(t) - q_{runoff}(t) - q_{infil}(t) - e_{surf}(t) & (1.1) \\ \dfrac{ds_{tops}(t)}{dt} = q_{infil}(t) - q_{percol}(t) - e_{tops}(t) & (1.2) \\ \dfrac{ds_{deeps}(t)}{dt} = q_{percol}(t) - q_{baseflow}(t) - e_{deeps}(t) & (1.3) \end{cases}$$

Fluxes in, across, and out of the vertical hillslope control volumes include precipitation $p(t)$, overland runoff $q_{runoff}(t)$, infiltration into the topsoil $q_{infil}$, percolation from the topsoil into the deeper soils $q_{percol}(t)$, baseflow into the channel $q_{baseflow}(t)$, and evaporation from the ponded, topsoil, and deep soil layers ($e_{surf}(t)$, $e_{tops}(t)$ and $e_{deeps}(t)$, respectively). The model assumes that percolation flux is a linear function of the amount of water stored at time $t$ in the topsoil $q_{percol}=k_{percol}\cdot s_{tops}$ and that the baseflow is a linear function of the water stored in deep soil $q_{baseflow}=k_{baseflow}\cdot s_{deeps}$. Overland runoff is a power function of the water stored on the hillslope surface (consistent with Manning's equation) given by

$$q_{runoff} = k_{runoff} s_{surf}^{1.67} \quad (1.4)$$

and infiltration is a nonlinear function of soil moisture content ($s_{tops}/T_{tops}$), where $T_{tops}$ is the thickness of the topsoil layer (i.e., A-horizon) and a linear function of hydraulic head $s_{surf}$ given by

$$q_{infil} = k_{dry}\left(1 - \frac{s_{tops}}{T_{tops}}\right)^{\phi} s_{surf} \quad (1.5)$$

where $k_{dry}$ corresponds to the case of dry soil and, similarly to $k_{runoff}$, $k_{percol}$, and $k_{baseflow}$ can be interpreted as the time constant (residence time) of the respective storage component. The hillslope area ($a_h$) for the elements in the distributed model is, on average, 0.05 km², and link length ($l_{link}$) is, on average, 400 m. Note that $a_h/(2l_{link})$ is the hillslope length. The exponent $\varphi$ is a nonlinearity introduced by the change in the potential matric of the soil column as soil moisture changes with time.

The HLM should be thought of as a modeling system rather than a single specific model. As the equations describing hillslope-scale processes are separated from the numerical solver, it is rather easy to explore different mathematical descriptions for water fluxes. For example, one can consider such simplifications as constant runoff coefficient or water transport velocity, or as an alternative, one can formulate these components based on the available physical characteristics.

Water transport through the river network is nonlinear and governs how channel links propagate flow through the river network. Formulated in the context of a mass conservation equation developed by Gupta and Waymire (1998), it uses the water velocity parameterization given by Mantilla (2007) as

$$\frac{dq_{link}(t)}{dt} = \frac{v_0 q_{link}^{\lambda_1}(t) A^{\lambda_2}}{(1-\lambda_1)l}\left[ a_h \left( k_{runoff} s_{surf}^{1.67}(t) + k_{baseflow} s_{deeps}(t) \right) - q_{link}(t) + q_1(t) + q_2(t) \right] \tag{1.6}$$

where $q_{link}$ is the discharge from the link at time $t$, $a_h$ is the total hillslope area draining to the link, $q_1(t)$ and $q_2(t)$ are the incoming flows of the upstream tributaries, $A$ is the upstream basin area, and $\lambda_1$, $\lambda_2$, and $v_0$ are global parameters of the water velocity component of the model and are set to 0.2, -0.1, and 0.3, respectively. The model can capture the main features of the

hydrographs including the maximum stage. Several studies used the model (e.g., Ayalew et al. 2014; Cunha et al. 2012). Krajewski et al. (2017) also discussed the model performance. The model is driven by radar-rainfall estimated from Level II NEXRAD data from seven WSR-88D weather radars covering the state of Iowa. The maps of rainfall intensity have spatial resolution of about 0.25 $km^2$ and are updated every five minutes. The algorithms are described in Krajewski et al. (2013) and Seo and Krajewski (2015).

An important aspect of our modeling approach is the avoidance of calibration. Instead, we rely on detailed information of the physical properties we model. This includes the topography, land use and land cover, soil properties, and details of the main forcing, i.e., precipitation. Comparing simulation results to streamflow observations across Iowa validates the model formulation and parameterization. Therefore, we can view the model as data-intensive and calibration-free when used in forecast-mode. This, in turn, implies that with more detailed, relevant, and accurate data, including model states and physical domain characterization as well as the driving inputs, the model will work better. The model is fully automatic in the sense that no corrections are applied to the model as it moves forward in time once initial and boundary conditions are imposed.

The model predicts the streamflow fluctuations associated with storm events over the catchment of interest using current observations of rainfall, and rainfall forecasts. The effect of storms on river ways is usually delayed a time ranging from days to weeks. Each point of interest in the landscape (bridge, culvert) can then be categorized according to the maximum warning time. The web interface will provide a visual tool to show when a particular location will be impacted, and it will provide an inundation map associated to the particular peak flow expected

for that location. Inundation maps are more effective tools in communicating the effects of flooding than crest stages at specific locations.

Chapter 2 Validation of a data approach for the model routing parameterization

During Phase II, we developed a methodology to estimate the routing parameters of HLM using observed flow records. In it, we estimated the routing parameters in relatively small watersheds in the state of Iowa by controlling the runoff. Then, we interpolated the routing parameters using a set of random forests. In Phase III, we further explored this approach and validated it using the records from 120 USGS gauges. Here we present a summary of the obtained results. A complete report of this work has been published in (Velásquez et al., 2022).

## 2.1 Background

Our first goal was to identify the routing parameters $v_0$ and $\lambda_1$ in multiple Iowa watersheds. We started by applying combinations of both parameters using a time-step runoff controlled (TRC) (see fig. 2.1) modeling scheme which uses streamflow records to estimate the correct runoff for fixed time steps. Using the TRC scheme, HLM routes the observed runoff volume into the channels.

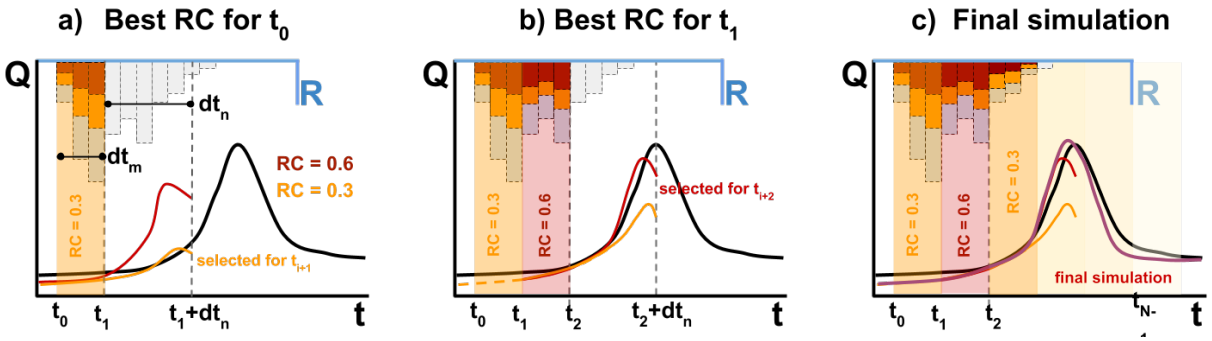


**Figure 2.1** Example for a single parametrization of the Time-step controlled runoff strategy (TRC). From left to right, the figure presents an example of how TRC adjusts the runoff coefficient (RC) to obtain the best possible streamflow $Q$ given a rainfall $R$ in a time step $dt_m$.

To apply the TRC strategy, we fixed the values that RC can take and the size of $dt_m$ and $dt_n$. We considered only events with a peak higher than the 25$^{th}$ percentile of the annual peak. The RC can take values between 0.0 and 0.9, with a step of 0.1. We set the modeling period $dt_m$ equal to three hours, and the no rainfall period $dt_n$ equal to six hours. Figure 2.2 presents an example of the results obtained for an event at Floyd River (Alton, IA). The figure shows the changes in the shape of the hydrograph, depending on the values of the parameters $v_0$ (rows) and $\lambda_1$ (columns). In this case, $v_0$ has more influence over the peak and $\lambda_1$ tends to shift the time. Moreover, their interplay produces the most significant changes. We applied the TRC strategy considering nine groups (or classes) of $v_0$ and $\lambda_1$ pairs. The classes correspond to all the possible combinations of $v_0$ taking the values of 0.2, 0.3, and 0.4, and $\lambda_1$ taking values of 0.15, 0.2, and 0.25.

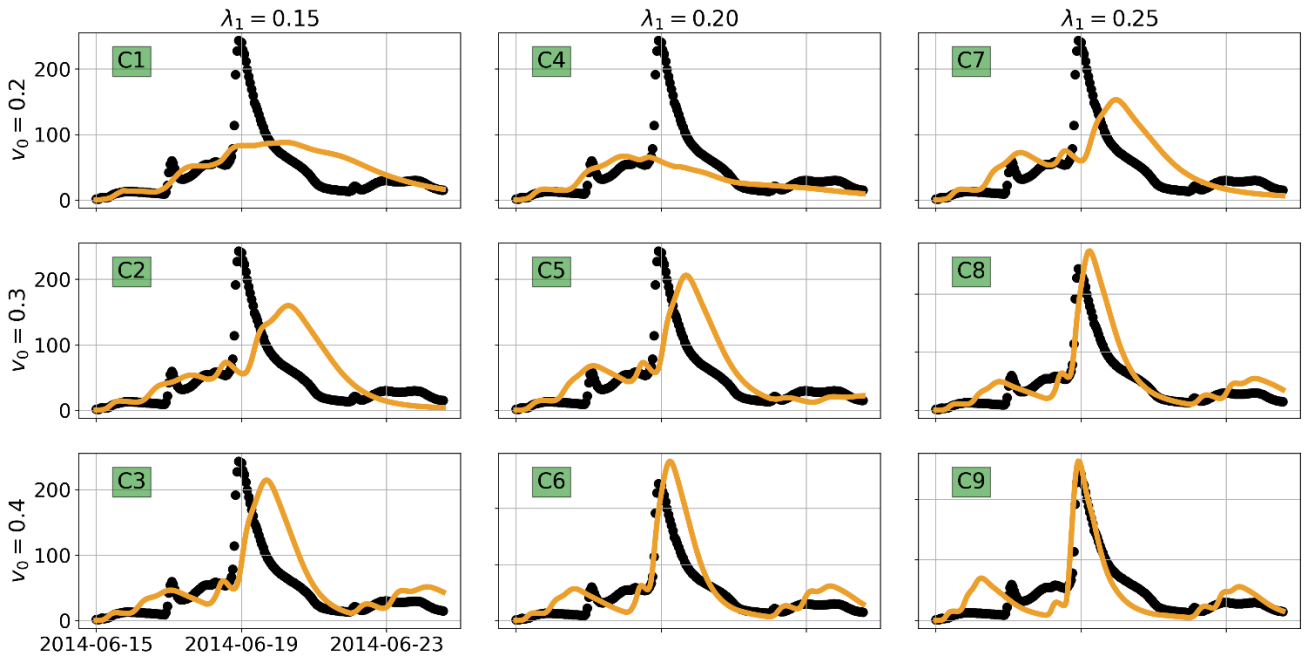


**Figure 2.2** Results of the TRC strategy applied for an event at Floyd River at Alton (USGS station 06600100). The frames correspond to the analyzed combinations of $\lambda_1$ (columns) and $v_0$ (rows).

We applied the TRC strategy to 48 small watersheds using nine different combinations of $v_0$ and $\lambda_1$ and several storm events (around 12 per watershed). For each watershed and event, we found the best parameter combination that maximizes the performance in terms of peak flow magnitude and timing (fig. 2.3a). Then, we obtained the predominant combination for each catchment or the local self-similarity case (fig. 2.3b). For this, we chose the $v_0$ - $\lambda_1$ pair with higher counts ranking first in the PMT index. In this procedure, we lost performance obtained at the best case by event. However, its peak flow estimation was still better than that of the open-loop case, which corresponds to one combination of $v_0$ and $\lambda_1$ for all the SW.

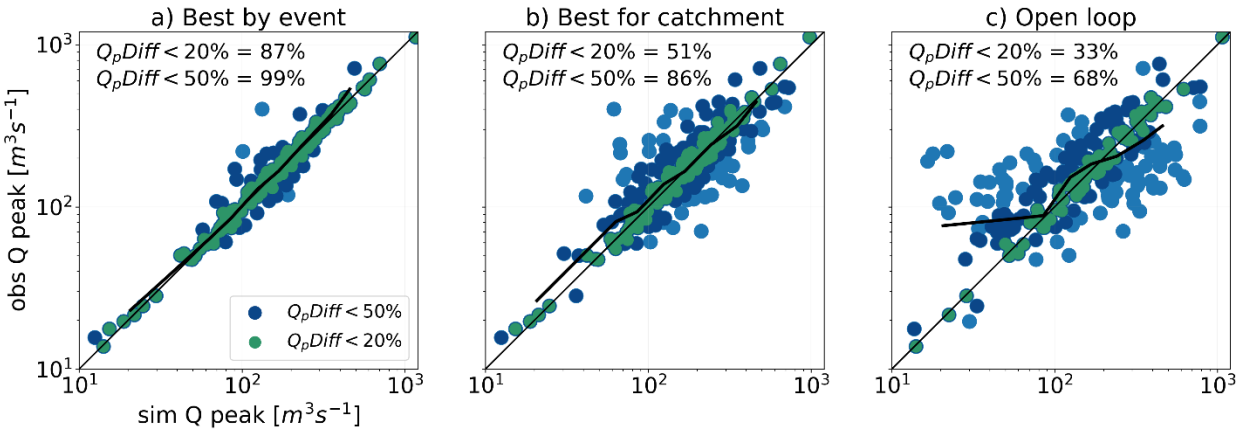


**Figure 2.3** Observed versus simulated peak flows for the training set events. The green dots correspond to peak flow differences lower than the 20%, the dark blue to differences lower than the 50%, the ligth blue to differences greater than 50%. The black line indicates the comparison of the simulated vs observed P50 for different magnitudes. a) Comparison of the result of the best $\lambda_1$ and $v_0$ combination for each event, b) Comparison of the best combination by watershed, and c) Results obtained with $\lambda_1 = 0.2$ and $v_0 = 0.33$.

Our results suggested that each analyzed watershed has a routing parameterization that varies from one event to the next (fig. 2.3a). We were able to capture a mean value of this oscillation with the best by catchment case, obtaining an overall good performance (fig. 2.3b). In the best by catchment case, about 88% of the cases produced a peak flow error below 50%, and

51% provided a peak flow error below 20%. The same percentages fell to 68% and 33% in the open-loop case. While the described results support the idea of the existence of a prevalent routing parametrization, the increased performance at best by event case (see legend in figure 2.3a) suggests that there were heterogeneities at the routing parameters of the channels due to changes on the hydraulic geometry.

## 2.2 Results

We interpolated the values of $v_0$ and $\lambda_1$ found for each basin using the ERF methodology. We trained and tested 400 RFs to obtain the random forest ensemble. Then, we selected the RFs with the best performance. We compared by couples three objective functions to determine the performance of each RF. At each comparison, we selected the realizations corresponding to the first ten Pareto fronts. We then chose the individuals corresponding to the intersection of the three Pareto fronts. After this procedure, we obtained an ERF composed of 74 individuals. The interpolation of $v_0$ and $\lambda_1$ for all the links of the region corresponded to the modal value of the ERF (fig. 2.4).

Our results showed a spatial distribution of the parameters that followed the landforms of the region (fig. 2.4). The map represents lower values for the routing parameters at the Des Moines Lobe, medium to high values at the Iowa Surface, medium at the Southern Iowa Drift, and high towards the Northwest Iowa Plains and the Loess Hills. The distribution of the parameters was linked to hydrological features used to train the Random Forest (fig. 2.4). In this process, the ERF gave more relevance to the slope of the channel ($S_0$), the travel time ($T_t$), the HUC level 8, and the upstream number of links ($N_l$).

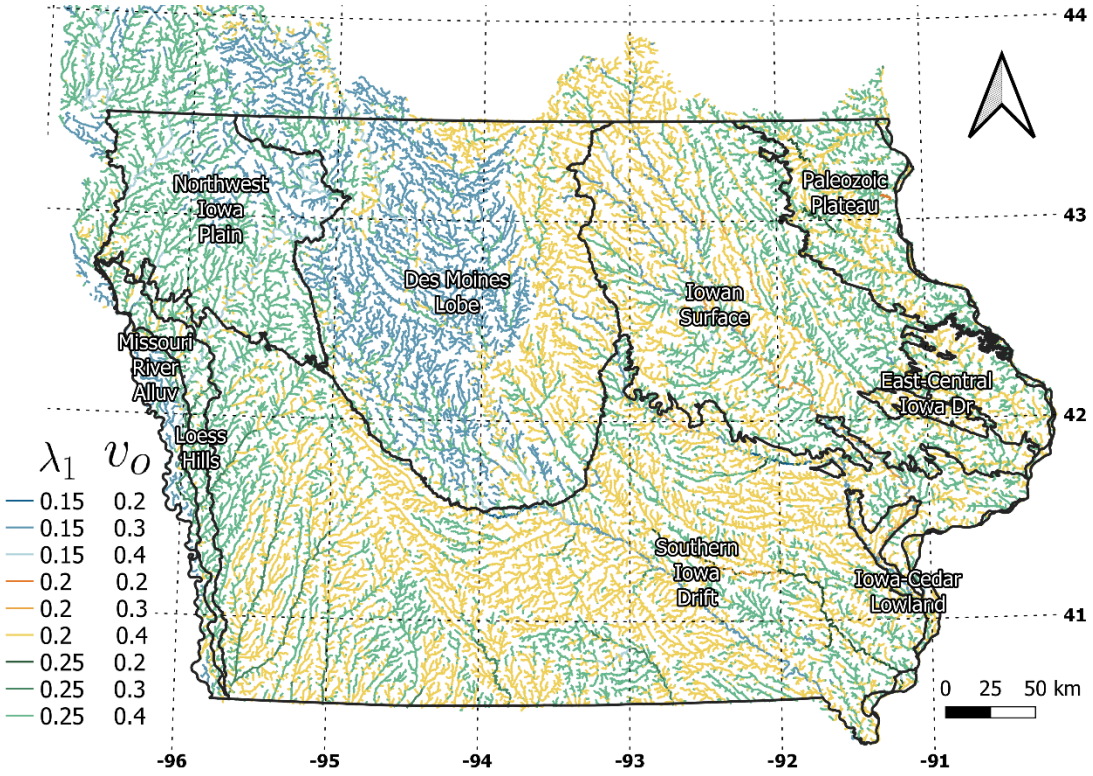


**Figure 2.4** Spatial distribution of $v_0$ and $\lambda_1$ obtained by the ERF interpolation.

According to figure 2.5a, there was an overall improvement; however, it may differ across scales. In figure 2.6, we present the distribution of the simulated and observed peak flows grouped by area intervals. Regardless of the setup, HLM tended to underestimate peak flows for small watersheds (below 1000 km$^2$). However, the setups exhibited differences in watersheds with larger areas. Between 1000 and 5000 $km^2$, the three models exhibited small differences at the mean and the maximum values. However, in this range the ERF maximum magnitude was more accurate. The open-loop setup tended to over-estimate peak flows for upstream areas larger than 5000 $km^2$. Contrary to the open loop, the HUCs approach tended to simulate accurate median peak flows but under-estimating the maximum values. On the other hand, with the ERF, we obtained accurate simulations on both the median and the maximum values. According to the

described results, our improvement was more evident over large watersheds (upstream areas above $1000km^2$).

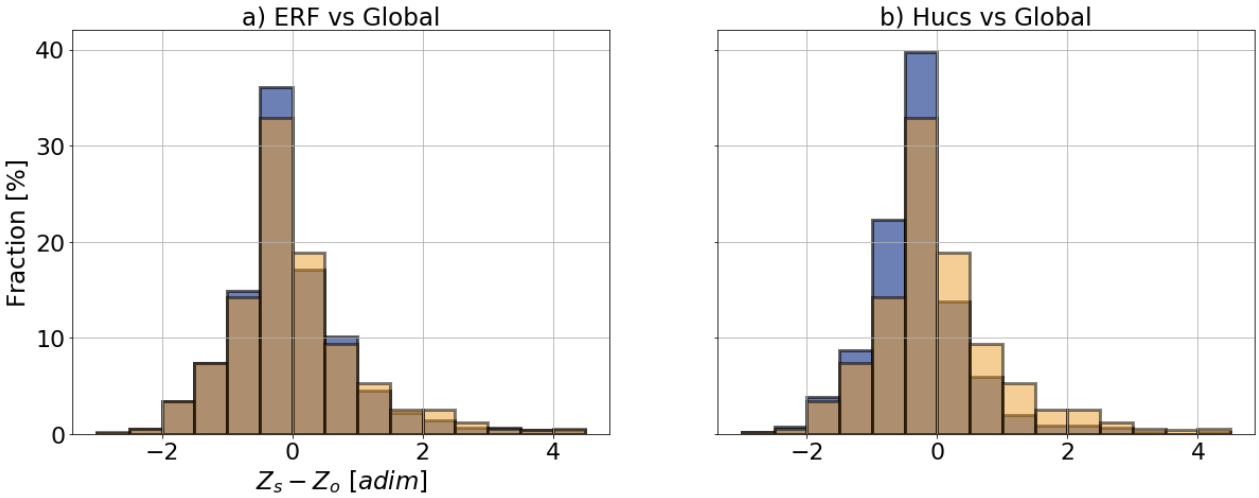


**Figure 2.5** Simulated vs Observed peaks. a) Open loop setup (orange) and the ERF setup (green). b) HUCs setup (orange) and ERF (green).

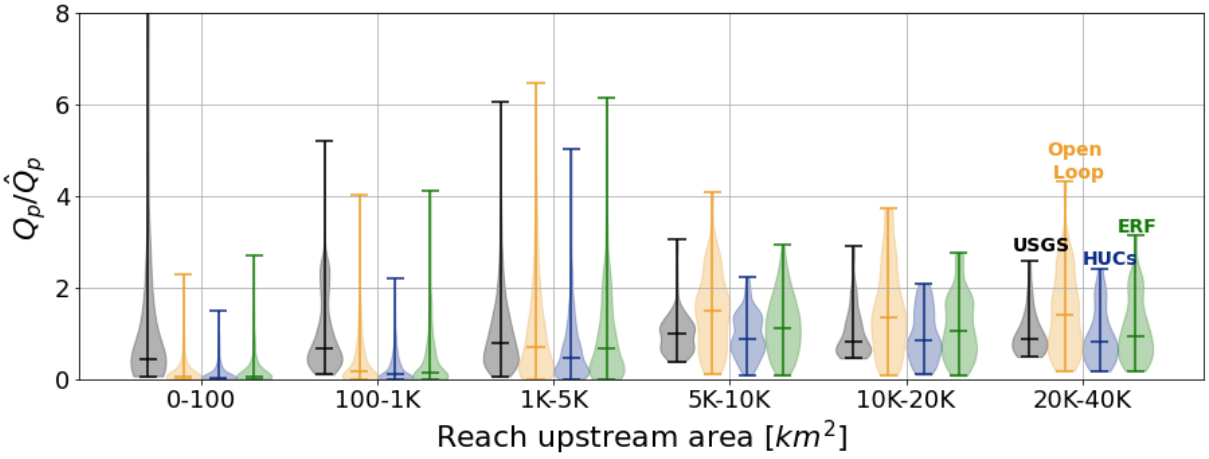


**Figure 2.6** Violins of the peak flow ratio variability in function of the upstream area. The black violins correspond to the observed data, the orange to the open-loop, the blue to the HUC steup, and the green to the ERF setup.

In addition to the peak flow magnitude, there was also a change at the simulated time to peak difference ($\Delta t$). To address this change, we computed the mean $\Delta t$ for each HLM setup at the USGS gauges and compared it with the watershed area (fig. 2.7a) and with the mean peak flow magnitude difference (fig. 2.7b). Additionally, we obtained the distribution of $\Delta t$ for each setup (fig. 2.7c). In all the cases, the time to peak difference decreased towards zero for large areas (fig. 2.7a). For upstream areas larger than $1000 km^2$, open-loop tended to simulate early peaks (negative $\Delta t$) more frequently than the ERF setup. On the other hand, the HUCs setup tended towards late peak estimations for all the upstream areas. In addition to the upstream area relationship, $\Delta t$ changed from early to late peaks when the peak flow difference went from over-estimations to under-estimations (fig. 2.7b). The described relationship was more dispersed in the open-loop case and more biased on the HUCs case. On the other hand, the ERF setup, corrected some of the observed bias improving both measurements at the time. The overall improvement of $\Delta t$ done by the ERF setup is observed in figure 2.7c.

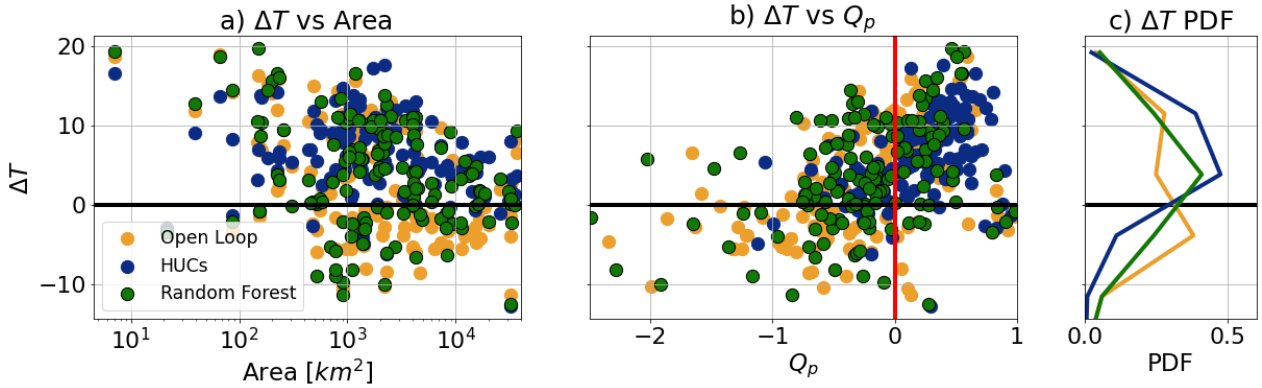


**Figure 2.7** Mean peak time difference by gauge. a) Comparison with the station area, b) Comparison with the mean peak flow difference by gauge, and c) overall relative time difference PDF. In the three plots the black line corresponds to the lowest time difference. In b, the red line indicates the lowest peak flow difference.

We present the KGE performance of the three setups in figure 2.8a. The KGE for the open-loop and ERF setups were similar, while the HUCs were smaller. Similar behavior was observed for streamflow gauges with upstream areas below 10,000 $km^2$ (fig. 2.8b). Alternatively, the KGE differences increased for large watersheds (fig. 2.8c), similar to the descriptions given at the peak flow differences in figure 2.6. We attributed the scale dependent differences to the aggregation process done by the streamflow network. The routing parameters seemed to amplify the errors and, consequently, the differences downstream. For upstream areas larger than 10,000 $km^2$ the HUCs KGE performance decreased, and there were more differences between the open-loop and the ERF setups. For large areas, ERF increased the cases with KGE values around 0.7, with some values around 0.9. Alternatively, the open-loop had a higher frequency around 0.8; however, it also increased the negative KGE count.

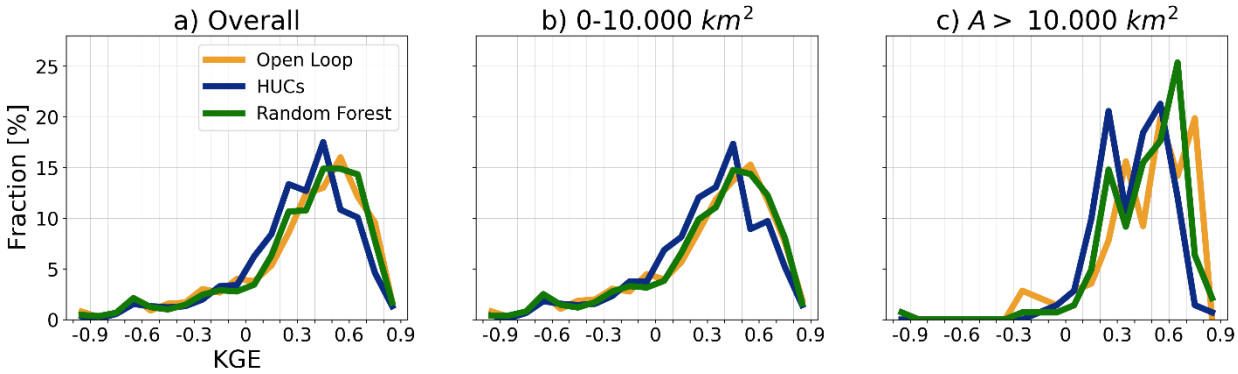


**Figure 2.8** KGE index distribution by reach area. a) Evaluation for all the USGS gages inside Iowa, b) evaluation for gages with areas between 0.1 and 10,000 km$^2$, and c) gages with upstream areas larger than 10,000 km$^2$.

We obtained some improvement at the estimation of peak flow magnitudes and timing using the ERF interpolation. Also, we obtained a spatial distribution of the routing parameters that seemed to be coherent with the landforms of the region. Compared with the open loop and

HUC parameterization, the ERF parametrization achieved a better representation of the peaks across scales. We attributed the improvement to the independent search of parameters through the TRC strategy and their posterior regionalization using a machine-learning approach. By searching parameters in the SW through the TRC, we corrected the hydrograph volume, which allowed us to perform a better identification of the routing parameters. Then, with the ERF, we performed a robust and non-linear interpolation of $v_0$ and $\lambda_1$ based on known hydrological features.

## 2.3 Conclusions

We developed a novel approach for parameter regionalization that found adequate values for routing parameters and then used an ensemble of random forest to interpolate them. We obtained the parameters using only small watersheds (SW) to reduce possible heterogeneities. Moreover, we applied a timestep-runoff controlled strategy over the SW to achieve an independent search of the routing parameters. We interpolated the found parameter values using two techniques: an ensemble of random forest (ERF) and nested HUCs. Then, we ran HLM between 2012 and 2018 using the open-loop setup and the ERF and HUCs interpolated fields. With the ERF interpolation, we obtained a parameters distribution that improved the model performance and followed the landscape of Iowa.

# Chapter 3 Improvements in Performance of the Hillslope Link Model in Iowa using a Non-linear Representation of Natural and Artificially Drained Subsurface Flows

Flood forecasts calculated using regional distributed hydrological models are becoming more common and relevant because they also provide information about internal watershed processes in large domains, along with predicted hydrographs for all streams in the river network. These forecasts are expected to be accurate at the region's ungauged watersheds (Samaniego et al., 2010) as a consequence of appropriate spatial representation of processes and parameters in the model.

Current hydrological models correctly identify many aspects of the streamflow hydrographs providing acceptable forecasts. However, they still struggle to reproduce the hydrograph recession. According to Mandeville (2016), modelers need to pay more attention to storm runoff's slow flow, which is a crucial component of the recession. For regional models, recession becomes more challenging because its non-linearity increases with the spatial scale (Chen & Krajewski, 2015; Clark et al., 2009; Harman et al., 2009). Landscape properties such as the topography, soil, and the stream network seem to be involved in the recession variability (Biswal & Marani, 2010; Shaw & Riha, 2012; Tallaksen, 1995). Additionally, human landscape and land use interventions, such as tile drainage, alter streamflow and its recession (Schilling et al., 2019; Schilling & Helmers, 2008).

## 3.1 Issues with the Hillslope Link Model (HLM) in Iowa

The Iowa Flood Center (IFC) produces flood forecasts for the state of Iowa using the Hillslope Link Model (HLM) (Mantilla and Gupta, 2005; Demir & Krajewski, 2013; Krajewski et al., 2017; Quintero et al., 2020). The operational HLM represents the hillslope subsurface flux using a linear-reservoir equation. According to Quintero et al. (2020), the current HLM

configuration accurately estimates peak flows and overall, has an acceptable performance in Iowa. However, the model has some limitations capturing the hydrograph recessions and the total runoff volume at some locations. The discrepancies between simulated and observed recessions are more notable in watersheds that are known to have been modified with tiling. Sample streamflow simulation results using the IFC HLM operational model for three Iowa watersheds are presented in figure 3.1a, b, and c (in red). The model's limitations are most evident in the watersheds located in the north and west regions of Iowa, where the model has low performance in terms of the Kling Gupta Efficiency (KGE) index (fig. 3.1d). We associate the model's poor performance in the region of north-central Iowa, known as the Des Moines Lobe, with the widespread use of artificial subsurface drainage (known as tile drains) in the region (Schilling & Helmers, 2008).

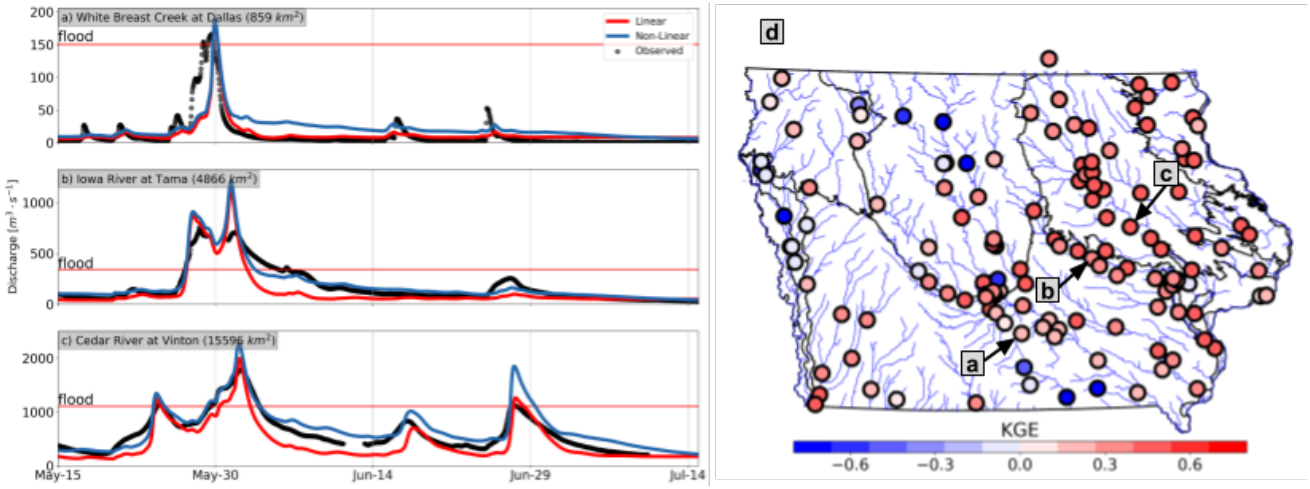


**Figure 3.1** a) Observed (black) and simulated streamflows by the linear (red) and the non-linear (blue) setups at three USGS gauged stations. b) Mean annual KGE performance of the HLM linear setup for Iowa between 2002 and 2018.

To address these issues, Fonley et al. (2021) developed a subsurface non-linear equation that can represent subsurface flow from hillslopes with different steepness and soil conductivities, as well as the presence of tile drainage. The blue lines in figure 3.1a to 3.1c show the resulting hydrographs using the non-linear equation with parameters corresponding to no tile and steepness of 2% (Fonley et al. 2021). Compared with the linear equation of the operational HLM, the non-linear equation tends to improve the total streamflow volume and the simulated recession shapes. However, we still observe discrepancies (fig. 1.1a and b) attributed to issues with parameter values and spatial representation of processes.

3.2 The diagnostic-prognostic approach

According to Clark et al. (2011), the development of a hydrological model is subject to the hypothesis-testing process. This process evaluates, rejects, and replaces model components. We performed a diagnostic-prognostic analysis of the model at 140 USGS gauges in Iowa to test the utility of the non-linear equation to represent the hillslope subsurface flux. In this case, we adapted the diagnostic-prognostic approach developed in studies on evapotranspiration (Allen et al., 2011; Kalma et al., 2008; Sur et al., 2020). Our diagnostic setups have simplified, spatially uniform parameter values while the prognostic scenarios use maps to determine parameter values. The diagnostic-prognostic approach offers complementary information about the model (Yilmaz et al., 2014) and the required independence to perform model comparisons (Crow et al., 2005).

According to Quintero et al., (2019), an insightful way to improve models starts with model performance verification followed by structure modification. We expanded on this approach by using the diagnostic-prognostic analysis to add tools to verify the model's processes and required parameters. In this paper, we first describe the HLM model and the equations

governing the hillslope processes. In the description, we include the operational linear equation and the non-linear equation to represent subsurface flux in the description. Then, we describe the diagnostic and prognostic setups. Finally, we compare the diagnostic and prognostic approach results at 140 USGS stations and analyze the parameters' influence on the model performance.

3.3 Model description

The Hillslope Link Model (HLM) represents the hydrological processes at the hillslope scale (fig. 3.2a and b) and routes the streamflow through the channel network (fig. 3.2c). At the hillslopes, HLM has three storages: ponded surface ($S_p\ [m]$), topsoil ($S_T\ [m]$), and subsurface storage ($S_s\ [m]$). The water from the ponded storage can either infiltrate the topsoil ($q_{pT}\ [m \cdot min^{-1}]$) or flow as runoff to the channel link ($q_{pL}\ [m \cdot min^{-1}]$). The water in the topsoil percolates ($q_{Ts}\ [m \cdot min^{-1}]$) to the soil storage. Finally, the water in the soil storage seeps into the channel link as subsurface runoff ($q_{sL}\ [m \cdot min^{-1}]$). Evaporation occurs from the three storages as a removal of volume from the model. Once in the river network, HLM transports the channel water ($q\ [m^3 \cdot s^{-1}]$) downstream. A detailed description of the hillslope and stream routing process can be found in Mantilla & Gupta (2005) and Quintero et al. (2020).

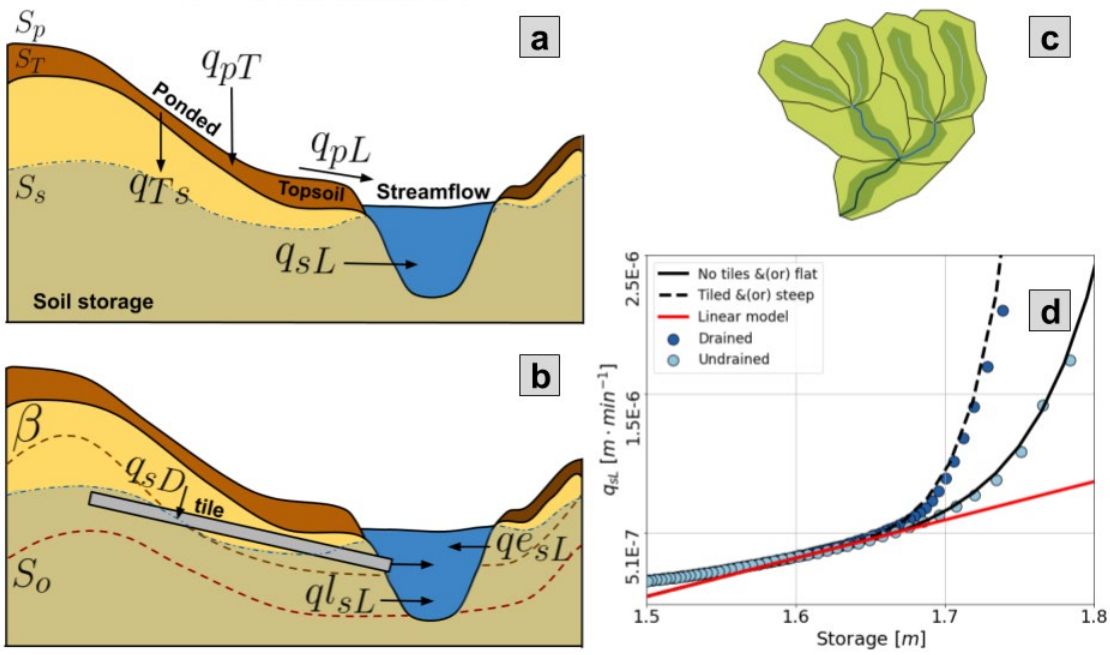


**Figure 3.2** Hillslope Link Model spatial discretization and schematic of the storages and processes represented at the hillslope scale: a) HLM hillslope process using the linear subsurface flux equation; b) hillslope process including the active layer ($\beta$), the exponential flux ($qe_{sL}$), and the tile drainage flux ($q_{sD}$); c) watershed decomposition into hillslopes and channel links; and d) functional form of the subsurface flux in the function of the soil storage ($S_s$) after Fonley et. al (2021).

The surface runoff, infiltration, and percolation rates are linked through the reference speed $v_r$ and the shape of the hillslope. Each hillslope has a parameter $k_2$ [$\text{min}^{-1}$] (eq. (3.1)) that depends on the hillslope link length ($L_i$ [$m$]) and area ($A_h$[$m^2$]), along with the reference velocity $v_r$. The parameter $k_2$ is the inverse of the runoff residence time in the hillslope. The runoff $q_{pL}$ and the infiltration $q_{pT}$ are linked to $k_2$ through equations (3.2) and (3.3), respectively. Also, the percolation rate $q_{Ts}$ is computed as a proportion of $k_2$, expressed by $k_i$. Usually, $k_i$ is 2% of $k_2$; however, its value may change depending on the soil and topographical properties.

$$k_2 = v_r \cdot \left(\frac{L_i}{A_h}\right) \cdot 60 \quad (3.1)$$

$$q_{pL} = k_2 \cdot s_p \quad (3.2)$$

$$q_{pT} = k_2 \cdot s_p \cdot 99 \cdot (1 - s_T/T_l)^3 \quad (3.3)$$

$$q_{Ts} = k_2 \cdot s_T \cdot k_i \quad (3.4)$$

The current HLM setup represents the subsurface flux to the channels ($q_{sL}$ $[m \cdot min^{-1}]$) with a linear equation (red line on figure 3.2d). The equation releases water to the channel at a rate $m$, when $S_s$ is greater than the no-flow threshold ($S_o$), as follows,

$$q_{sL} = m \cdot (s_s - S_o) \quad (3.5)$$

Fonley et al. (2021) developed a set of parameterizations for ordinary differential equations that adds a non-linear component to equation (3.5) when $S_s$ is above threshold storage. The following exponential equation (continuous line on figure 3.2d) is added to equation (3.5) if $S_s$ is greater than the activation threshold $\beta$ $[m]$,

$$qe_{sL} = \alpha(s_s - \beta)e^{17(s_s - \beta)} \quad (3.6)$$

where $\alpha$ is a parameter that depends on the hillslope properties, such as its steepness and the soil conductivity. Fonley et al. (2021) also developed an exponential equation that applies when the hillslope has tiles. The following equation (dashed line on figure 3.2d) is added when $S_s$ is greater than the tile relative depth $D_d$ $[m]$,

$$q_{sD} = c(s_s - D_d)e^{35(s_s - D_d)} \quad (3.7)$$

In the described scheme, subsurface flux becomes a set of equations that HLM activates, depending on the value of $S_s$ relative to the thresholds $S_o$, $\beta$, and $D_d$. The segmented subsurface runoff is as follows,

$$q_s L = \begin{cases} ql_{sL} & if \quad S_s < \beta \\ ql_{sL} + qe_{sL} & if \quad S_s > \beta \\ ql_{sL} + qe_{sL} + q_s D & if \quad S_s > D_d \end{cases} \tag{3.8}$$

The relative tile depth ($D_d$) is independent of $\beta$, so either could be larger depending on the tile configuration and the hillslope properties. Moreover, if there are no tiles, equation (3.8) is limited to its two first expressions. More details on the subsurface equation development can be found in Fonley et al. (2021).

3.4 Diagnostic and prognostic setups

We used both diagnostic and prognostic approaches to test the performance obtained using the non-linear equation. We used the river network for the state of Iowa derived from a DEM of 90m and decomposed into about 420,000 individual hillslopes, following the approach presented in Mantilla & Gupta (2005). The precipitation forcing corresponds to hourly Stage IV QPEs (Reed & Maidment, 1999, Lin, 2011). We forced the evapotranspiration using the mean annual monthly values derived from MODIS (Running et al., 2017) for the region.

Equation (3.8) offers a formulation for the subsurface flux that we want to validate on the Iowa domain. In this process, we can fix parameters uniformly over the space or distribute them spatially. A uniform setup assumes that each hillslope in the region uses the same model parameters, while a distributed setup assumes parameter variability as a function of landscape properties. Neither approach is without error because the parameters are only approximate, and

they could depend upon unknown factors that are variable in space. The fixed setup is unrealistic, and the distributed setup may be subject to spatial errors. However, both approaches are complementary. Fixed setups could help assess the ability of equation (3.8) to improve the accuracy of simulated streamflow fluctuations. In contrast, a distributed setup helps to validate the parameter description given by the map(s). Considering this, we used both approaches to validate the new $q_{sL}$ equation and to explore the limits of the so-called predefined setups. In the case of distributed parameters, we use the steepness of the hillslopes (fig. 3.3a) and the tiles localization according to the DNR (fig. 3.3b).

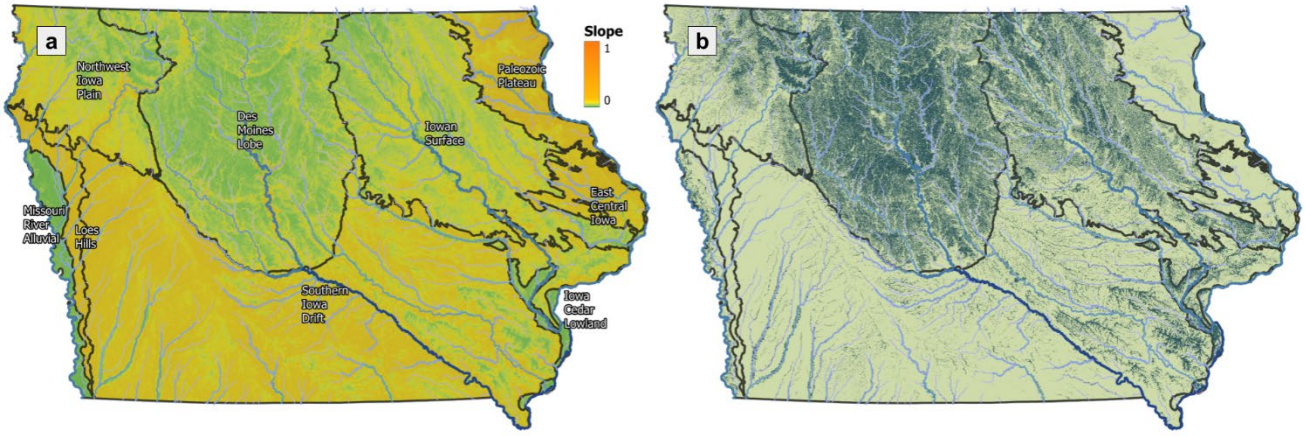


**Figure 3.3** Maps of the hillslope steepness (a) and tile drainage localization according to the Iowa DNR (b).

The model validation consists of comparing fixed (diagnostic) and distributed (prognostic) HLM setups (fig. 3.4). The diagnostic setup (fig. 3.4a) shows how different formulations could significantly improve the model over the region. On the other hand, the prognostic setups (fig. 3.4b) show the improvements and limitations derived from the application of "known" spatial variables.

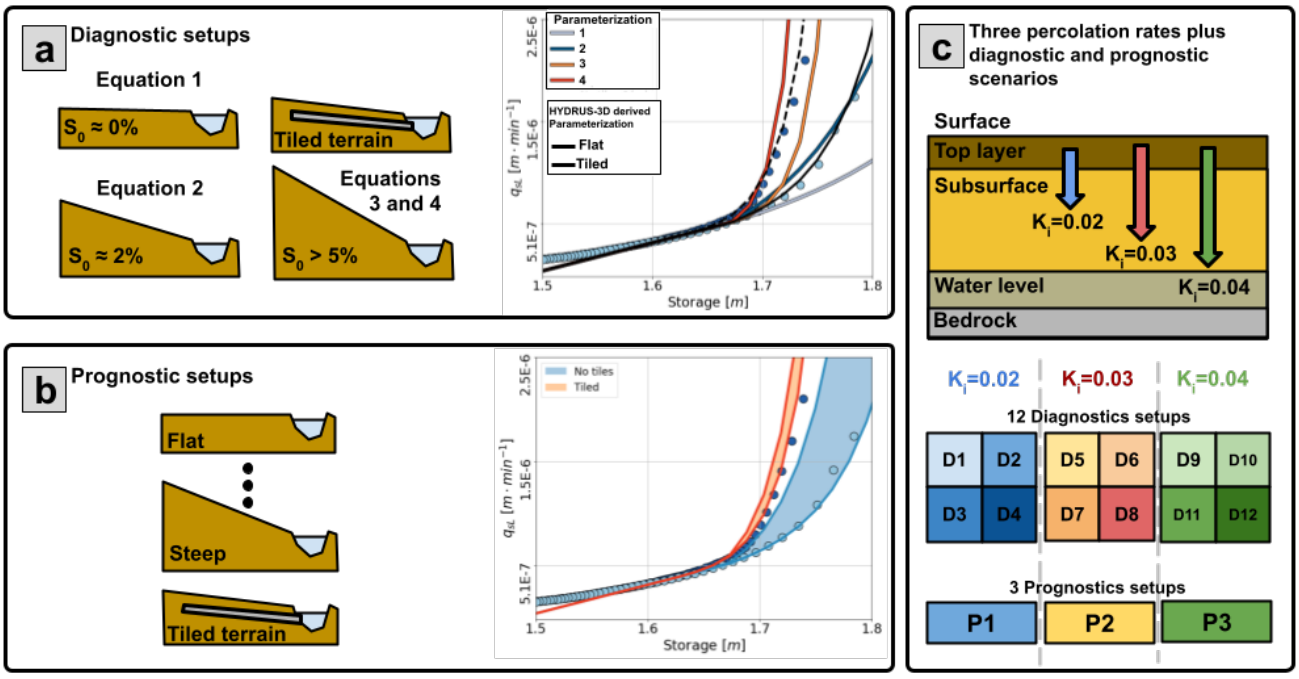


**Figure 3.4** Diagnostic and prognostic experiment setup: a) diagnostic case with four non-linear subsurface flux equations fixed for the domain of Iowa after Fonley et al (2021); b) prognostic case, with equations varying with the hillslopes steepness (blue scenario) and the presence of tiles (red scenario); and c) percolation rates fixed for the different scenarios. Their combination gives us 12 diagnostic scenarios and three prognostic scenarios.

*3.4.1 Diagnostic setups*

In the diagnostic setup (fig. 3.4a), we created four parametrizations of equation (3.8) for the Iowa domain. The parametrizations range from flat hillslopes (light blue line on figure 13a) to steep or tiled hillslopes (red line on figure 3.4a). By combining the four parameterizations and the three $k_i$ rates, we obtain 12 diagnostic scenarios (D1 to D12 in figure 3.4c). D1 to D4 use $k_i = 0.02$; D5 to D8 use $k_i = 0.03$; and D9 to D12 use $k_i = 0.04$.

*3.4.2 Prognostic setups*

In the prognostic setup, we distributed parameter values in function of the hillslopes steepness and the Iowa DNR map describing tiles presence (fig. 3.4b). According to Fonley et al. (2021), the parameter $\alpha$ of equation (3.6) can be explained by the hillslope steepness ($\gamma_h$) using a

linear equation. Using the following equation, we assigned $\alpha$ to each hillslope, obtaining functional forms that oscillate between the blue bands shown in figure 2.1b,

$$\alpha = \gamma_h(8.5 * 10^{-8}) + 9.48 * 10^{-7} \quad (3.9)$$

Additionally, we include equation (3.8) for tiled terrain following the tile distribution shown by the map in figure 3.4b. For the tile drainage equation, we use $c$ equal to $5.4 * 10^{-7}$ (see Fonley et al., 2021). Combined with the percolation rates $k_i$ of 0.02, 0.03, and 0.04, we developed the prognostic scenarios P1, P2, and P3, respectively (distributed setups in figure 3.4c).

3.5 Results

The KGE equation summarizes the correlation ($\gamma$), the mean value ($\mu$), and the deviation ($\sigma$). Our results suggest that the KGE performance depends heavily on the percolation rate ($k_i$). With $k_i = 0.02$ (first row of figure 3.5), all the non-linear setups tend to improve the linear model, with significant performance decrease in some events. Conversely, values of $k_i$ equal to 0.03 and 0.04 do not exhibit a significant KGE change (second and third rows of figure 3.5). Cases such as D5 and D11 exhibited a performance like the one obtained by the linear model. Other cases, such as D9, resulted in a general decrease in performance. D6, D8, and P2 exhibited a slight performance increase. The described results highlight the relevance of the percolation rate and the subsurface parameters. The comparison with the linear model shows that equation (3.8) can significantly improve the model performance, depending on the parameters.

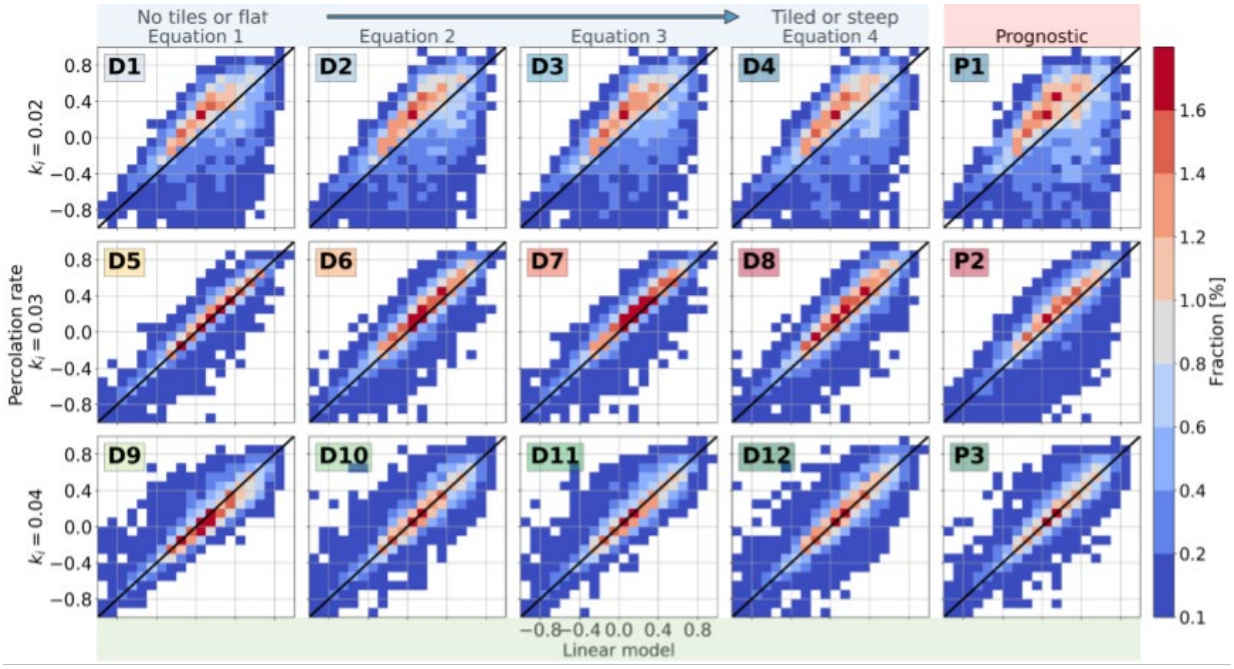


**Figure 3.5** Event-based KGEs comparison between the diagnostics setups and the linear model. Each row corresponds to a fixed percolation rate. Columns correspond to the four fixed equations. The color bar shows the percentage of events that fall at each bin of the 2D histogram.

Differences among the scenarios are highlighted when comparing the performance gauge by gauge. First, we choose the diagnostic (D) and prognostic (P) setup with the best performance at each gauge. For this, we used the KGE to select the setup outperforming the others at most of the events. In figure 3.6, we present the KGE distribution and the percentage of time chosen for each scenario. We found similarities between the diagnostic and prognostic chosen setups when grouped by the percolation rate values ($k_i$). D4 and P1 ($k_i = 0.02$) have a similar KGE distribution, as do D8 and P2 ($k_i = 0.03$) and the group that includes D9, D11, D12, and P3 ($k_i = 0.04$). The similarities among the described groups highlights the relevance of $k_i$. Moreover, some differences also highlight the relevance of the equation (3.8) parameters.

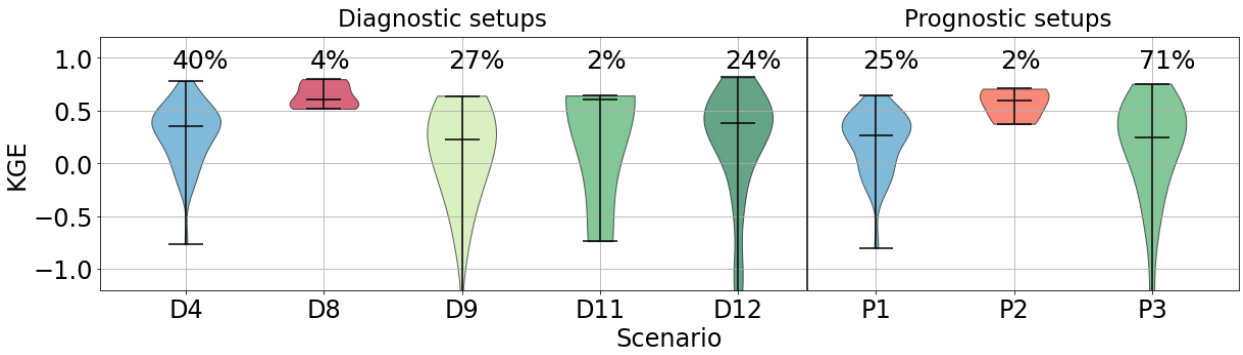


**Figure 3.6** Event-based KGE distribution for the selected scenarios at each station.

The results presented in figure 3.6 follow a spatial distribution. Figure 3.7 shows each USGS gauge colored by the diagnostic (fig. 3.7a) and prognostic (fig. 3.7b) setups with the best performance. In both cases, the percolation rate defines the spatial distribution. We can identify how the chosen setups (fig. 3.7) follow the Iowa landforms to some extent in the diagnostic case (see fig. 3.3a). Scenario D12 is recurrent over the Des Moines Lobe and the Northwest Iowa Plain. D9 recurs over the Missouri River Alluvial and Loess Hills landforms. D4 dominates over the Southern Iowa Drift area. In the remaining regions, we see a mix of scenarios. The spatial distribution is similar among the chosen prognostic scenarios (fig. 3.5b) and seems to be highly influenced by the percolation rates, represented here by tones of blue ($k_i = 0.02$), red ($k_i = 0.03$), and green ($k_i = 0.04$).

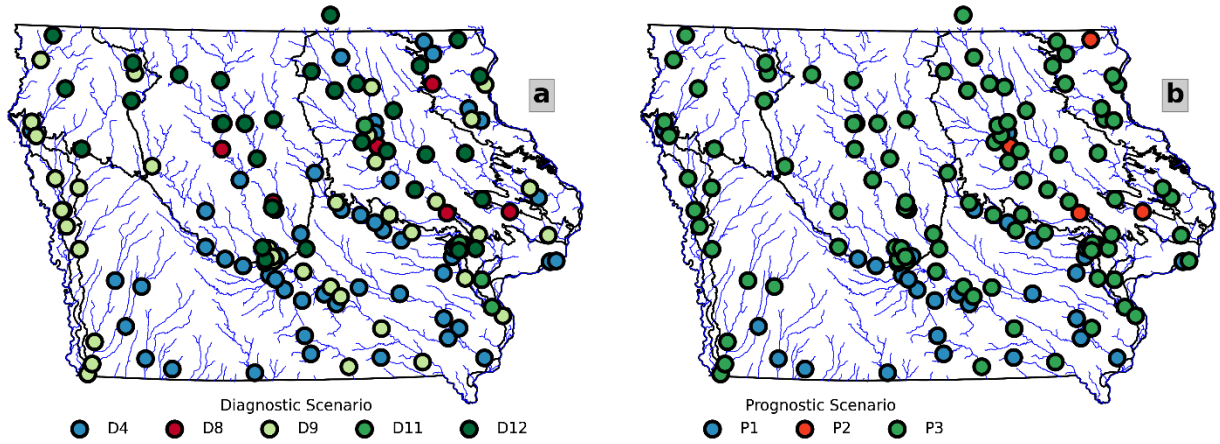


**Figure 3.7** Spatial distribution of the scenarios with best KGE performance at each USGS station: a) results obtained from the diagnostic scenarios; and b) results obtained from the prognostic scenarios. The green, red, and blue gauge colors correspond to the percolation rates of 0.02, 0.03, and 0.04, respectively.

According to figure 3.5, the chosen diagnostic and prognostic scenarios share percolation rates. However, differences exist in the spatial performance improvement distribution (fig. 3.6). Figure 3.6a and b show the diagnostic and prognostic scenarios of KGE improvement with respect to the linear model. With only two cases of negative KGE differences (red dots on figure 17a), the diagnostic scenarios outperform the linear model at almost all the USGS gauges. Alternatively, in the prognostic case (fig. 3.6b), the count of negative KGE differences increases to 13, while the number of gauges decreases where the improvement is more significant than 0.1 (yellow). We attribute the prognostic case performance decrease to the parameter's spatial distribution.

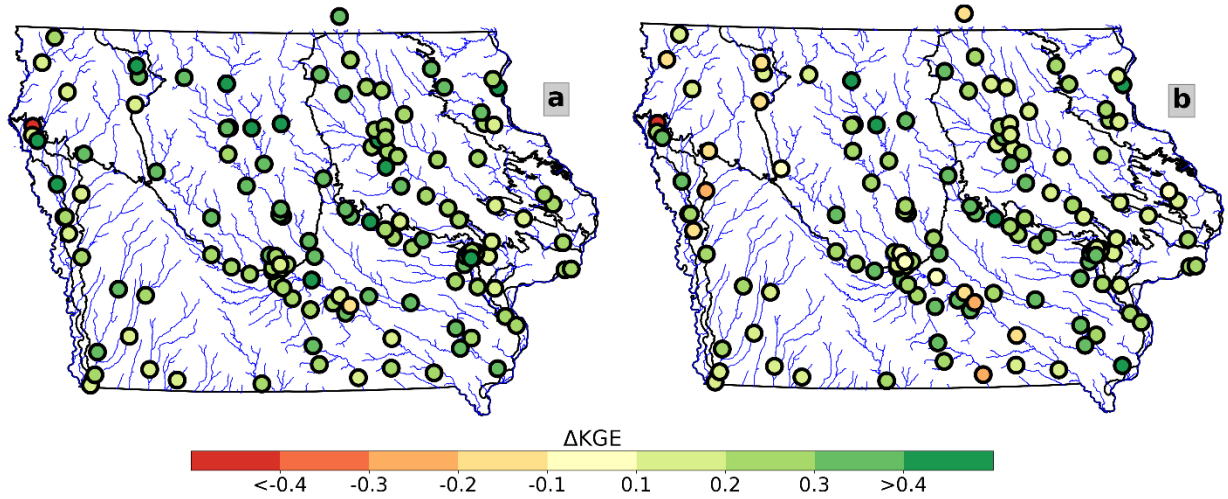


**Figure 3.8** Mean KGE spatial difference of the diagnostic and prognostic scenarios with respect to the linear model: a) diagnostic KGE minus linear model KGE; and b) prognostic KGE minus linear model KGE.

The prognostic scenario performance decrease happens mostly over the east and west regions of Iowa. The most significant decrease happens on the Northwest Iowa Plains landform (fig. 3.6b). In this region, the chosen diagnostic setups were D12 and D9 (fig. 3.5a), suggesting a mix between tiled terrain and flat hillslopes. Over the Southern Iowan Drift landform area, the $k_i$ value is the same for the diagnostic and prognostic scenarios. However, the prognostic scenario performance declines at several stations in this region. On the other hand, the Iowa Surface region exhibits more $k_i$ discrepancies between both scenarios, as well as more performance differences.

The described results suggest a level of heterogeneity in the parameters shown by the diagnostic and prognostic scenarios. This heterogeneity creates difficulties when choosing the most adequate regional parameterization for the model, regardless of whether it is fixed (diagnostic) or distributed (prognostic). To address this issue, we compare the KGE (upper diagonal in figure 3.7) and the mean ratio (lower diagonal in figure 18) of the chosen scenarios. According to figure 3.7, the KGE and mean ratio of scenarios D4 and P1 outperform almost all

the scenarios. Additionally, both scenarios have the highest percentage of events with KGE values above 0.4 (blue bars in figure 3.7 histograms). Compared with the linear model, D4 and P1 mean ratio correction is significant. In both plots (Linear-D4 and Linear-P1), there are almost no events where the linear setup outperformed D4 and P1.

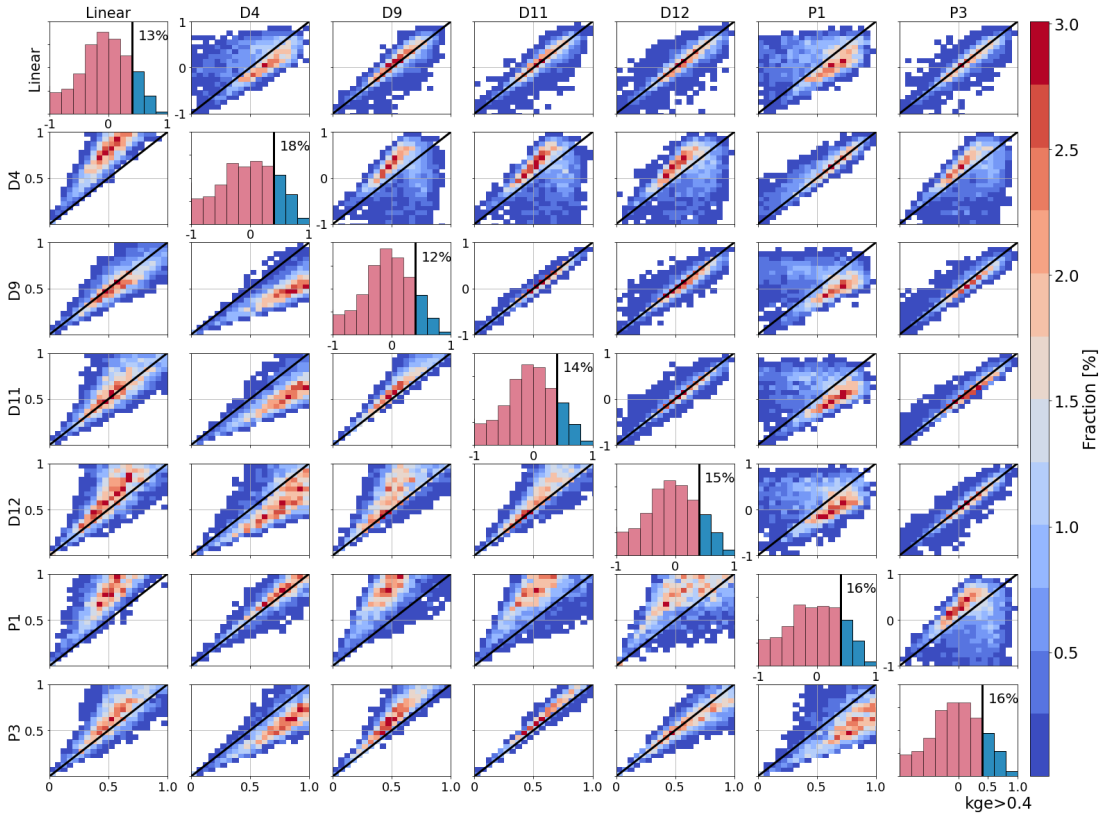


**Figure 3.9** Event-based KGE comparison of the diagnostic and prognostic dominant scenarios. Each row compares a scenario against the others. The upper diagonal panels correspond to the KGE histogram of the scenarios. Over the diagonal shows the KGE histogram of each setup coloring in blue the percentage of events with a KGE above 0.4. The lower diagonal compares the event based mean ratio error.

The scenarios D4 and P1 have the same $k_i$ (0.02) value; however, their subsurface parameters are different. The parameters of D4 are fixed for all the domains following line 4 of figure 2.1a. This parameterization represents highly conductive soils or the presence of tiles. On

the other hand, P1 parameters follow the hillslope steepness with equation (3.9), and the presence of tiles described by the map in figure 2.1b. The described differences in the parameters seem to develop slight dissimilarities in the performance. According to panel D4-P1 in figure 3.7, the KGE performance is similar in both, although D4 has a better performance in some events. Moreover, the panel P1-D4 shows that the mean ratio description of both setups is similar. Considering that D4 assumes tiles everywhere, our results suggest a high presence of tile-like signatures.

## 3.6 Conclusions

In the diagnostic setup, we implemented 12 fixed parameter scenarios, while in the prognostic setup, we distributed the parameters considering the hillslope steepness and presence of tiles. In both cases, we considered three fixed percolation rates. Results from this study indicate the following:

1. Compared with the linear equation, the exponential equation corrects the volume bias on the simulated streamflow. We attribute the correction to the active layer threshold on the exponential equation and the significant outflow increase once the storage is above this threshold. By contrast, in the linear equation, the water remains in the soil for extended periods because of the described absence of these processes.
2. Depending on the parameters, the exponential equation could improve the performance of HLM. We found that the exponential equation outperforms the linear equation for several parameter combinations with changes in the shape of the hydrograph, the simulated peaks, and the timing. We also found significant differences using different combinations of the equation parameters and the percolation rate.

3. The percolation rate plays a significant role in the representation of the subsurface flux from the described combinations. We found spatial coincidences in the percolation rates when choosing the best diagnostic and prognostic scenarios. Also, the percolation rate induces changes comparable with those produced by the exponential equation's parameters.
4. Determining the distributed parameters of HLM remains challenging. In this paper, we used the diagnostic and prognostic approach to analyze the parameters of HLM. The diagnostic approach assumes unknown conditions and fixed parameters over the space. On the other hand, the prognostic method is the more classical approach, in which the parameters are derived from maps of the landscape. In our experiments, the diagnostic setups tended to outperform the prognostic setups. Additionally, we found it hard to identify a link between the diagnostic and prognostic parameters and their respective performances.

We showed how a better representation of the processes and the correct parameters can improve a hydrological model. The improvement is supported by comparisons performed at 140 USGS gauges. Moreover, the differences between the diagnostic and prognostic setups suggest that identifying the parameters is still challenging. Despite the limitation of the number of gauges, the diagnostic approach reveals the parameters' potential spatial distribution.

Two main factors may explain the differences in parameters and performance between the diagnostic and prognostic setups: errors in the landscape description and unrepresented processes in HLM. Uncertainties exist in the tile localization maps; likewise, limitations exist in the representation of the average steepness at the hillslope scale. On the other hand, we have unrepresented processes in some regions of Iowa, such as potholes over the northwest and

agricultural terraces in the west. It is difficult to identify which one of these factors is more relevant to the implementation of a hydrological model. However, according to our results, the use of maps as landscape descriptors may lead to errors that are usually hidden in a posterior calibration process. Moreover, we found it hard to identify the errors caused by prescribed distributed parameters. Both issues could be addressed using diagnostic setups that help identify the uncertainties derived from the parameters and their possible regional distributions.

Chapter 4 Technology transfer.

In addition to the model improvements, we have also been working in transferring our technology to the University of Nebraska. The steps involve the back and front end of the platform. The back end includes the installation of the HLM model, watersheds delineation, HLM files configuration, acquisition of USGS streamflow information, and rainfall data processing. The front end includes the setup of the web page that deploys the HLM simulations results, the USGS data, and the metrics of the model performance. At the current stage, the setup has the watershed delineation, topology extraction, the USGS data, and the deployment of the web page. In figure 4.1 we present a snapshot of the setup done at the Elkhorn Watershed at Nebraska. In the current report, we describe in detail the steps followed starting from the data download to its deployment on the platform.

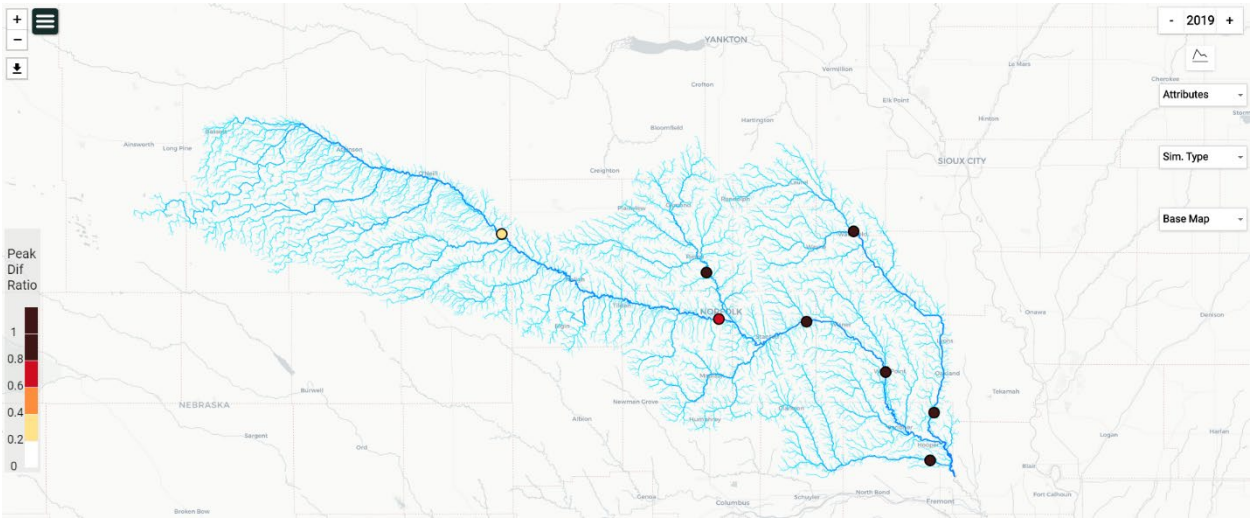


**Figure 4.1** Elkhorn Watershed setup using HydroVise software. URL: http://visualriver.net/hygis.html?config=http://s-iihr51.iihr.uiowa.edu/~nicolas/elkhorn.json

To setup the HLM model and the platform, we first performed the pre-processing of the data, and then, we configure HydroVise (Jadidoleslam et al., 2020) for the web platform (fig.

4.2). In the pre-processing we obtain the hydrological features for the region and downloaded the USGS streamflow gauge data. To obtain the hydrological features we start by downloading the rasters containing the DEM (Digital elevation model) for the region, and ended with vector layers describing the watershed, and the files required to run HLM. On the other hand, to download the USGS streamflow gauge data we use a Python script that uses the official USGS database. Then, we upload the obtained maps and time series to a Linux server. In the server, we configured HydroVise (https://github.com/njadid/HydroVisE) to show the results in an interactive web platform.

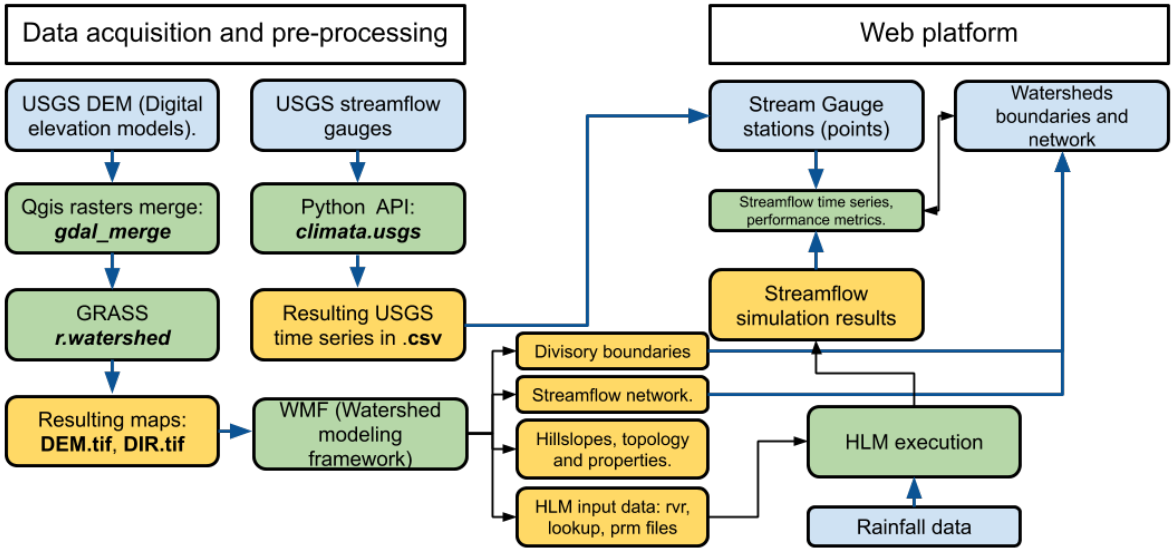


**Figure 4.2** Schematics of the steps followed for the setup of the Elkhorn watershed. Blue items indicate data elements, Green corresponds to software packages and programs, and Yellow corresponds to processed products.

## 4.1 Watershed delineation.

The shape of the watershed, its hillslopes, and its network depends mostly on the landscape and the topography of the region. A good descriptor of the landscape topography are

the DEMs (Digital elevation models). A DEM is a raster map that describes the elevation variability by cells of a certain resolution or scale. Depending on its size and the DEM pixel scale, a watershed could be contained in one or more DEM. For the current case, we downloaded the 1/3 arc-second DEM provided by the USGS at their web service TNM (The National Map viewer, url: https://viewer.nationalmap.gov/basic/). In total, we downloaded eight DEM maps each one with a resolution around 10 meters (see fig. 4.3).

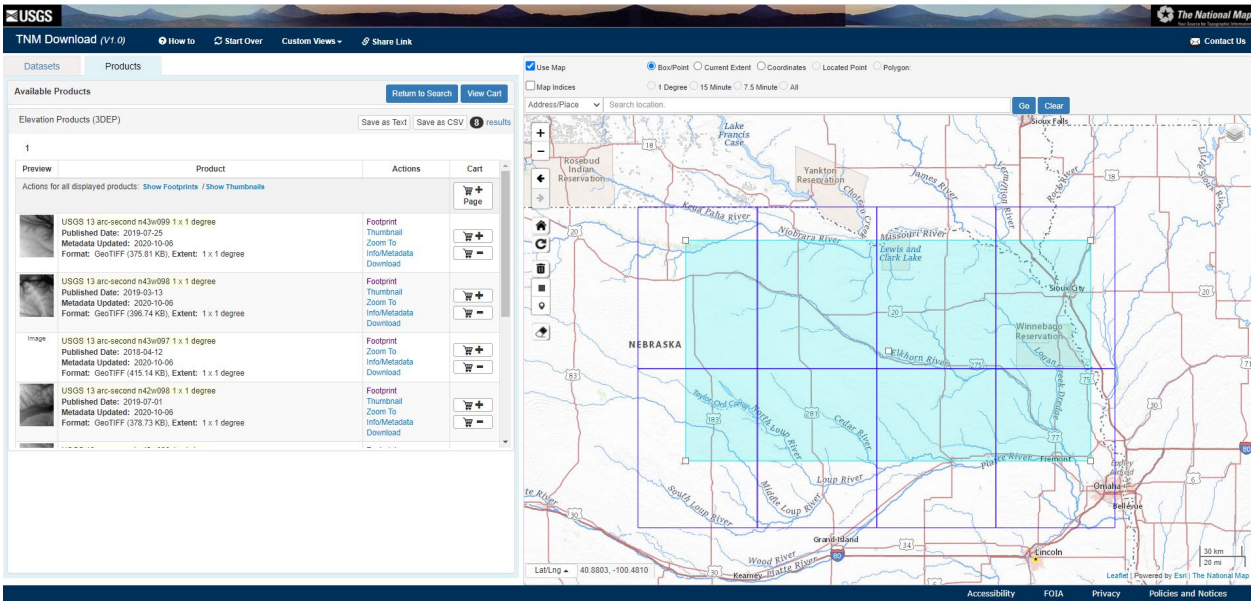


**Figure 4.3** USGS DEM download from TNM (The national map viewer). The blue regions correspond to the DEM downloaded to delineate the Elkhorn Watershed.

The DEMs default projection is NAD-83, this projection is not suited for the web platform and the rainfall products. After downloading all the data, we re-project the DEMs to WGS-86. To perform the re-projection, we use *gdalwarp* (Contributors, 2020). In this process we also change the pixel size of the DEMs. Although the procedure is done by GDAL, we

perform it using QGIS 3.12 (Project, 2020) in a Windows 10 system. However, the procedure can be easily done in a terminal using the following command:

```
gdalwarp -r near -of GTiff -ts 1205 0 original_dem.tif  processed_dem.tif
```

In the gdalwarp command **-r near** stands for the type of interpolation to be done for the reprojection. In this case it is nearest neighbor. The option **-of Gtiff** stands for the output format which in this case is GeoTiff. The option **-ts** determines if there is going to be a resample of the original raster. In this case, the 1205 is the new number of rows (originally 10812) and the number of columns. When both numbers are equal to zero, gdal reshapes the raster to fit the given non-zero value. In this case, with 1205 rows, we change the pixel size to 74m.

The reprojection process can also be done using the Qgis GUI. To access it the user must load the raster maps to Qgis and then go to Raster -> Projections -> Warp(Reproject). The program will pop-up a window with the options to select the map (or maps) to reproject, and several options (fig. 4.4).

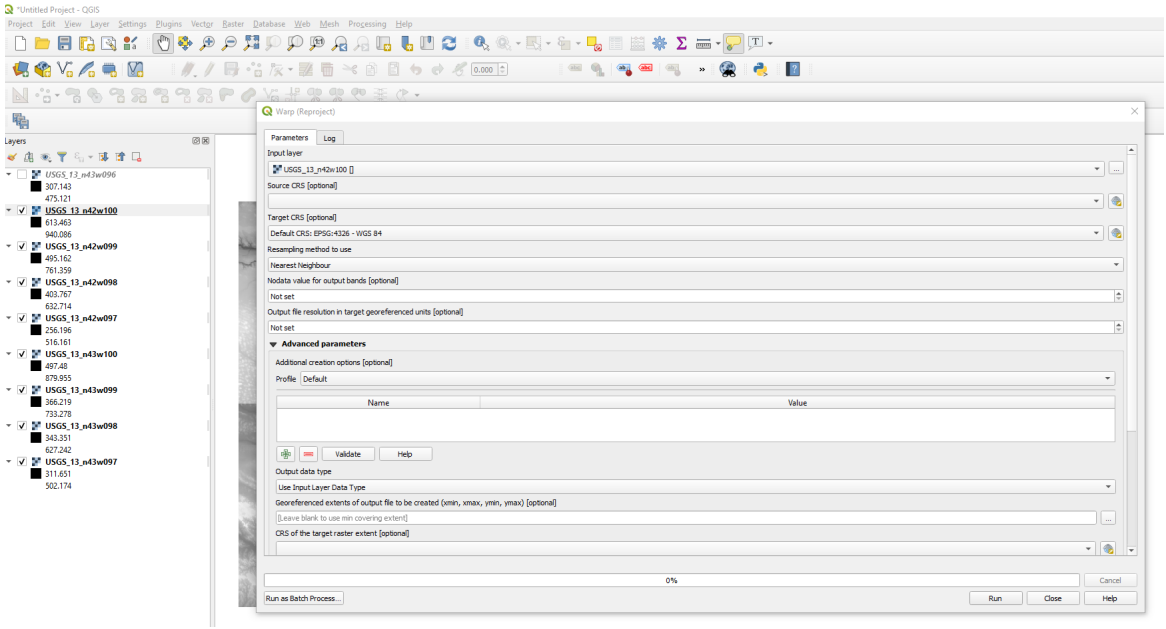


**Figure 4.4** Snapshot of the Qgis Warp(reproject) function used to change the DEMs projection.

After the reprojection we merge the DEMs into one large raster that contains the Elkhorn watershed. To perform the merge, we use the Merge tool in Qgis. To open this tool the user must go to Raster -> Miscellaneous -> Merge. In this process the user must select the input DEMs and the destination folder and name of the merged DEM.

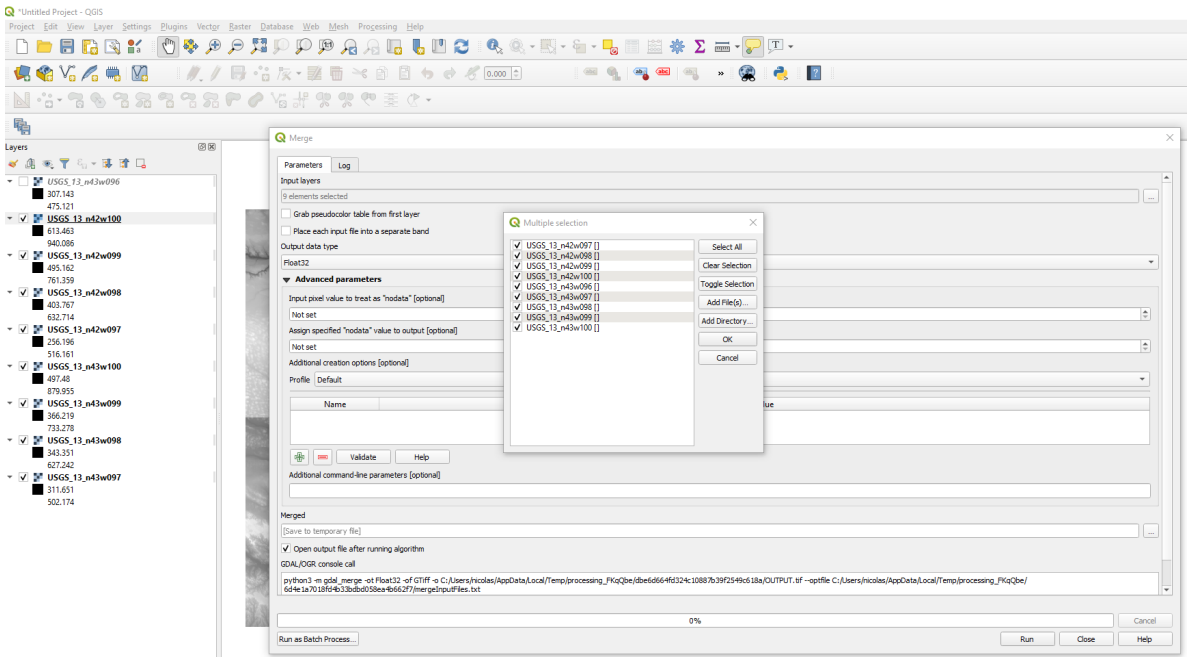


**Figure 4.5** Snapshot of the Qgis Merge function used to create the DEM that contains the Elkhorn watershed.

We proceed to process the merged DEM to obtain the flow direction map or DIR. In this case we use the function **r.watershed** from GRASS (Team, 2017) to process the DEM. The function uses the $A^T$ algorithm (Ehlschlaeger, 1989) to process the DEM obtaining several maps. To use r.watershed, the user must first setup a project on GRASS or use the function directly on Qgis through the toolbox option. Once the project is setup, the first step consists on importing the DEM map into GRASS. Then, the GRASS computational extent must be set to match the DEM extension. This procedure can be done once the map is loaded into the project. At the display panel the user must right click on the map and select the option “Set computational region from selected map”. Once this step is done, the user must select r.watershed function, which is found under Raster -> Hydrologic Modeling. The function will prompt a new window (fig. 4.6) containing several tabs. In this case we select the loaded DEM as the input elevation map on the Input tab, and we write down the name of the DIR map on the “Name of output drainage

direction raster map" on the Output tab. We then click the Run button, and after several seconds the DIR map is loaded. Finally we export the DIR map to a GeoTiff format by using the **r.out.gdal** GRASS function.

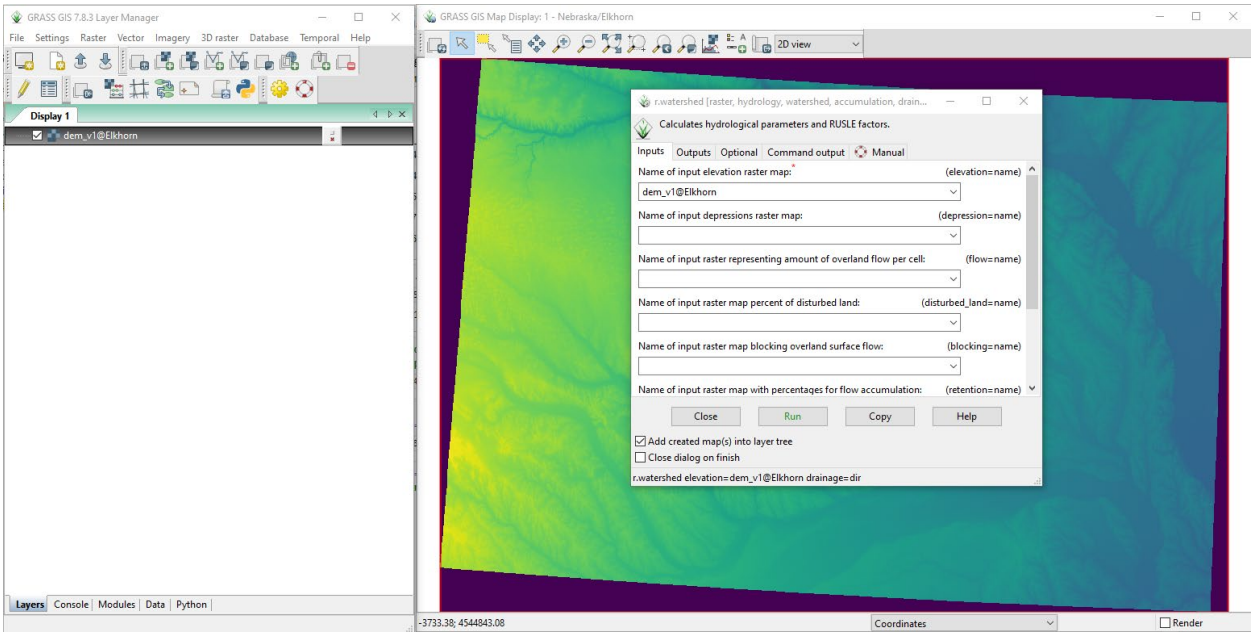


**Figure 4.6** Snapshot of the GRASS with the Merged DEM loaded and the r.watershed window.

Once we have the DIR map, we proceed to extract the watershed and its features. For this procedure we use the WMF (watershed modelling software). WMF (Velásquez et al., 2020) is an open Python toolbox that allows the delineation, interaction, and analysis of watersheds using a DEM and a DIR map. The toolbox could be found in the following github repository: https://github.com/nicolas998/WMF. WMF could be installed in a Linux machine through the terminal by cloning the github repository and installing the package with the python installer:

```
git clone  https://github.com/nicolas998/WMF
cd clone_dir/WMF/
python3 setup.py install –user
```

In windows the package could be installed using the Anaconda terminal and the MigWin64 package.

To work with WMF the user must use a Python script or terminal (like jupyter). Once in Python, the procedure starts by loading the DEM and DIR maps. Then, to obtain a more accurate result, the user can obtain the stream given by a particle that travels from any point inside the watershed to its outlet. Finally, the user must give the coordinates of the output of the watershed to obtain it. The described steps are as follows:

```python
# Import WMf into python
from wmf import wmf

#Load DEM and DIR maps
DEM,epsg = wmf.read_map_raster('DEM.tif', isDEMorDIR=True, dxp = 73.4, noDataP=-999)
DIR,epsg = wmf.read_map_raster('DIR.tif', isDEMorDIR=True, dxp = 73.4, isDIR=True, noDataP=-999)
DEM[DEM ==0] = -999

#Obtain a streamflow inside of the watershed that drains through the outlet.
Stream = wmf.Stream(x_stream, y_stream, DEM, DIR)

#Obtain the watershed element starting from its outlet.
Watershed = wmf.SimuBasin(x_out, y_out, DEM, DIR, umbral=170, stream=Stream)
```

In WMF the watershed is considered a Python class that has many properties and functions. Among them, there are functions to save the watershed boundary and its network as vector layers. Also, there are functions to write the files required by HLM for its execution. The following code writes the described maps and files:

```python
#Saves the network elements of the watershed.
Watershed.Save_Net2Map('network.shp',umbral=170, EPSG=int(epsg), Dict={'acum':Acum})

#Saves the boundary polygon of the watershed.
Watershed.Save_Basin2Map('Watershed_boundary.shp',dx = 77.46, EPSG=int(epsg))

#Obtains and writes the topological structure and parameters files required by HLM
Watershed.Transform_Basin2Asnych('HLM_files.rvr', lookup=True, prm=True)
```

The described procedure was followed to obtain the network and the boundary of the Elkhorn watershed shown in figure 1.1, and to obtain the boundary of the watersheds that correspond to the USGS gauges.

## 4.2 Stream gauge data

The setup of the USGS streamflow gauges involves the localization of the required gauges and its posterior download. First, we obtain the spatial localization of the streamflow gauges by downloading the **GageLoc.shp** vector layer from the USGS science base catalog (shortened url: https://bit.ly/3o9OG4P). Then, we select the gauges that are inside of the watershed boundary polygon obtained at the previous step. Finally, we downloaded the data using the Python package **climata**. The following example shows how the information was downloaded and saved as CSV (comma separated value) files.

```
import pandas as pd
from climata.usgs import InstantValueIO, DailyValueIO

#Read the instant value data for a given station between the initial and end dates.
data = InstantValueIO(date_initial, date_end, station = usgs_code, parameter = "00060")
#Extract the values and dates of the downloaded data
for series in data:
    flow = [r[0] for r in series.data]
    dates = [r[1] for r in series.data]
#Convert the data to a DataFrame
Q = pd.Series(flow, pd.to_datetime(dates, utc=True)) * 0.02832
Index = [d.replace(tzinfo = None) for d in Q.index]
Q.index = Index

#Rename the data column and save as a CSV file
Q.rename(columns={0:'Q'}, inplace=True)
Q.to_csv(dest_csv_file, float_format='%.2f', index_label='date')
```

## 4.3 Web platform setup

HydroVise supports the web platform putting together the vector and time series information. We setup HydroVise in a Linux machine by cloning it from its repository (https://github.com/hydrovise/HydroVisE) and adding it to the Apache web-server directory. Then, we upload the required information into a folder of the project. The folder contains a

**configure.json,** the **GeoJson** files of the vector layers, a **metric.csv** and **flood.csv** files, folders with the time series in csv format, and a folder with the boundary polygons of the watersheds. In figure 4.7 we show the current structure of the project.

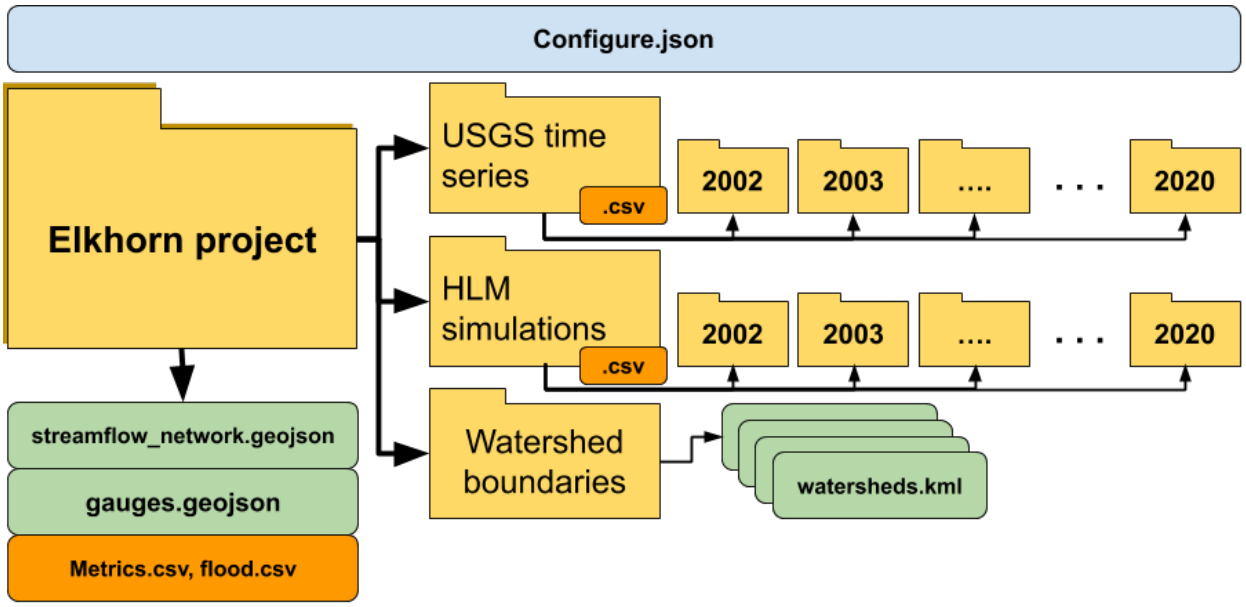


**Figure 4.7** Structure of the HydroVise setup for the Elkhorn watershed project. Yellow elements correspond to folders, green to GeoJson and kml files, the orange to CSV files, and blue to json files.

Each type of file has its role in the setup of HydroVise. The configure.json controls the layers and time series that are going to be shown on the platform. The GeoJson and kml files are the map layers present on the platform. The gauges.geojson is a MultiPoint file with a high level of interaction. This layer can change its colorbar in function of the metrics.csv file and shows the observed and simulated streamflow records. The streamflow_network.geojson and the kml of the watersheds are present for graphical reasons. The streamflow_network.geojson is loaded from the beginning and HydroVise colors the channel segments in function of the Horton orders described in the file. The kml of the network corresponds to the boundary polygons

corresponding to the gauges.geojson watersheds. Finally, time series are stored by product and by year in a CSV format. In the current case, we setup a folder for the USGS observations, and a folder for the HLM simulations. Nevertheless, the project could contain more folders with times series.

The connection among the data is given by an identifier that is present at the gauges.geojson, the metrics.csv, the name of the watersheds, and the name of the CSV time series files. The identifier could be the link ID of the channel segment corresponding to a USGS gauge or it could be the identification number of the station. In this case, we choose the USGS number as the identifier. The identifier number is present at the **metrics.csv** file as the **lid** column (fig. 4.8) and at the **gauges.geojson** (fig. 4.9).

```
year,lid,prod,usgs,area,mean,min,max
2008,6799000,usgs,06799000,7580,1.31,2.76,0.34
2009,6799000,usgs,06799000,7580,0.90,3.85,0.08
2010,6799000,usgs,06799000,7580,1.28,3.79,1.00
2011,6799000,usgs,06799000,7580,1.17,5.78,0.10
2012,6799000,usgs,06799000,7580,0.45,1.00,0.04
2013,6799000,usgs,06799000,7580,0.43,1.91,0.04
2014,6799000,usgs,06799000,7580,0.51,4.36,0.04
2015,6799000,usgs,06799000,7580,0.57,5.30,0.06
2016,6799000,usgs,06799000,7580,0.97,4.83,0.16
2017,6799000,usgs,06799000,7580,0.94,5.13,0.13
```

**Figure 4.8** Example of a metrics.csv file. The file must contain the columns year, lid, and prod, the remaining fields are optional and could contain statistics of the data or information about the model performance. Each row contains the data for a given year and place of interest.

```
{
"type": "FeatureCollection",
"name": "usgs_hydroanal",
"crs": { "type": "name", "properties": { "name": "urn:ogc:def:crs:OGC:1.
"features": [
{ "type": "Feature", "properties": {"lid": 6797500, "STAID": "06797500",
{ "type": "Feature", "properties": { "lid": "06799000", "STAID": "067990
{ "type": "Feature", "properties": { "lid": "06799100","STAID": "0679910
{ "type": "Feature", "properties": { "lid": "06799315","STAID": "0679931
{ "type": "Feature", "properties": { "lid": "06799350","STAID": "0679935
{ "type": "Feature", "properties": { "lid": "06799445","STAID": "0679944
{ "type": "Feature", "properties": { "lid": "06799500","STAID": "0679950
{ "type": "Feature", "properties": { "lid": "06800000","STAID": "0680000
]
}
```

**Figure 4.9** Example of a gauges.geojson file. The file can contain additional information about each point, however, it must have the lid field.

The web platform could be accessed once the files are on the server and the config.json file is set. To access the web platform the user must give HydroVise the path to the config.json file. The following command gives an example of how to perform this final task:

**http://**your-web-server-address/**HydroVisE.html?config=config.json**

# Chapter 5 Conclusions

An important aspect in providing a safe, efficient, and effective transportation system is anticipating natural hazards that can lead to road closures. Extreme floods can lead to bridge overtopping and/or compromising the structural integrity of river overpasses, including box culverts. The flood forecasting model and information system proposed here provides a tool to anticipate potential hazardous situations related to floods. It would allow time for the activation of action plans to minimize the impact on the overall transportation system. The forecasting model can be used in real time to anticipate floods and to look at past flooding scenarios to determine if all the actions taken were appropriate or can be improved. Our forecasting system will contribute to improving safety and minimizing risk associated with increasing multi-modal freight movements on the U.S. surface transportation system by *enhancing safety* and providing warning of potential road closures.

As part of this project, we have provided a prototype forecasting web platform with four specific innovations. 1) Forecasts at critical river/road intersections, 2) Spatial animated maps of flood evolution into the future, and 3) a measure of forecast accuracy at the newly incorporated forecast bridges. Our developments give us confidence that we can continue moving forward in developing a forecasting system that is transferable to other locations in the Midwest. As floods continue to be the most costly disaster in the nation, it becomes critical that tools are develop to better predict them.

References

Ayalew, T., Krajewski, W., Mantilla, R. (2013). Exploring the Effect of Reservoir Storage on Peak Discharge Frequency. J. Hydrol. Eng. 18, 1697–1708. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000721

Ayalew, T. B., Krajewski, W. F. & Mantilla, R. (2014), "Connecting the power-law scaling structure of peak-discharges to spatially variable rainfall and catchment physical properties," *Advances in Water Resources*, vol. 71, pp. 32–43.

Ayalew, T., Krajewski, W., Mantilla, R. (2015). Insights into Expected Changes in Regulated Flood Frequencies due to the Spatial Configuration of Flood Retention Ponds. J. Hydrol. Eng. 20, 4015010–4015010. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001173

Ayalew, T.B., Krajewski, W.F., Mantilla, R., Wright, D.B., Small, S.J. (2017). Effect of Spatially Distributed Small Dams on Flood Frequency: Insights from the Soap Creek Watershed. J. Hydrol. Eng. 22. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001513

Bhatt, Gopal et al. (2014) (https://pdfs.semanticscholar.org/715d/d10d9c3a356ef61c576bf1afb8f704bfd0ca.pdf)

Choi, C. C., Constantinescu, G., & Mantilla, R. (2015). Implementation of a Hydraulic Routing Model for Dendritic Networks with Offline Coupling to a Distributed Hydrological Model. Journal of Hydrologic Engineering.

Cunha, L.K., P.V. Mandapaka, W.F. Krajewski, R. Mantilla, and A.A. Bradley. (2012). Impact of radar rainfall error structure on estimated flood magnitude across scales: An investigation based on a parsimonious distributed hydrological model, Water Resources Research, 48(10), W10515.

Gupta VK, Waymire EC. (1998). Spatial Variability and Scale Invariance in Hydrologic Regionalization. In Scale Dependence and Scale Invariance in Hydrology, Sposito G

(ed.).Cambridge University Press: Cambridge; 88–135. DOI: 10.1017/CBO9780511551864.005

Krajewski, W.F., A. Kruger, S. Singh, B.-C. Seo and J.A. Smith, Hydro-NEXRAD-2: Real-time access to customized radar-rainfall for hydrologic applications, Journal of Hydroinformatics, 15(2), 580-590, 2013.

Krajewski, W. F., Ceynar, D., Demir, I., Goska, R., Kruger, A., Langel, C., Young, N. C. (2017). Real-time flood forecasting and information system for the state of Iowa. Bulletin of the American Meteorological Society, 98(3), 539–554. https://doi.org/10.1175/BAMS-D-15-00243.1

Laplante and Phillip. (2007). What Every Engineer Should Know About Software Engineering.

Mantilla, R. and V.K. Gupta, (2005). A GIS numerical framework to study the process basis of scaling statistics in river networks, IEEE Geoscience and Remote Sensing Letters, 2(4), 404-408.

Mantilla, R., Gupta, V. K. & Mesa, O. J. (2006).Role of coupled flow dynamics and real network Mantilla, R. (2007). Physical basis of statistical self-similarity in peak flows on random self-similar networks, PhD dissertation, University of Colorado, Boulder.

Mantilla, R. (2007). Physical basis of statistical self-similarity in peak flows on random self-similar networks, PhD dissertation, University of Colorado, Boulder.

Mantilla, R., Cunha, L. K., Krajewski, W. F., Small, S. J., Jay, L. O., Fonley, M. & Curtu, R. (2012), “Simulation of a Distributed Flood Control System using a Parallel Asynchronous Solver for Systems of ODEs,” Applied Simulation and Modeling, June, Napoli, Italy, http://dx.doi.org/10.2316/P.2012.776-042.

Moser, B. A., Gallus, Jr., W. A., & Mantilla, R. (2015). An Initial Assessment of Radar Data Assimilation on Warm Season Rainfall Forecasts for Use in Hydrologic Models. Weather and Forecasting, vol. 30, no. 6, pp. 1491-1520.

Newton, D.W., and Herrin, J.C. (1982). Assessment of commonly used methods of estimating flood frequency, Transportation Research Record, 896, 10-30.

NWS – National Weather Service. Climate Reports: August 2016 & summer 2016. [Online] 2016. http://www.weather.gov/dvn/Climate_Monthly_08_2016 (accessed Mar 21, 2017).

Seo, B.-C., W.F. Krajewski and L.K. Cunha, Impact of radar-rainfall uncertainty on hydrologic response for the eastern Iowa flood of 2008, Water Resources Research, 49, 2747–2764, 2013.

Small, S. J., Jay, L. O., Mantilla, R., Curtu, R., Cunha, L. K., Fonley, M., & Krajewski, W. F. (2013). An asynchronous solver for systems of ODEs linked by a directed tree structure. Advances in Water Resources, 53, 23–32. https://doi.org/10.1016/j.advwatres.2012.10.011

State of Iowa. (2009). https://www.legis.iowa.gov/docs/ico/section/466C.1.pdf

Allen, R. G., Pereira, L. S., Howell, T. A., & Jensen, M. E. (2011). Evapotranspiration information reporting: I. Factors governing measurement accuracy. *Agricultural Water Management*, *98*(6), 899–920. https://doi.org/10.1016/j.agwat.2010.12.015

Biswal, B., & Marani, M. (2010). Geomorphological origin of recession curves. *Geophysical Research Letters*, *37*(24), 1–5. https://doi.org/10.1029/2010GL045415

Chen, B., & Krajewski, W. F. (2015). Recession analysis across scales: The impact of both random and nonrandom spatial variability on aggregated hydrologic response. *Journal of Hydrology*, *523*, 97–106. https://doi.org/10.1016/j.jhydrol.2015.01.049

Clark, M P, Rupp, D. E., Woods, R. A., Tromp-van Meerveld, H. J., Peters, N. E., & Freer, J. E.

(2009). Consistency between hydrological models and field observations: linking processes at the hillslope scale to hydrological responses at the watershed scale. *Hydrological Processes*, *23*(2), 311–319. https://doi.org/https://doi.org/10.1002/hyp.7154

Clark, Martyn P., Kavetski, D., & Fenicia, F. (2011). Pursuing the method of multiple working hypotheses for hydrological modeling. *Water Resources Research*, *47*(9), 1–16. https://doi.org/10.1029/2010WR009827

Contributors, G. (2020). *Geospatial Data Abstraction software Library*.

Crow, W. T., Li, F., & Kustas, W. P. (2005). Intercomparison of spatially distributed models for predicting surface energy flux patterns during SMACEX. *Journal of Hydrometeorology*, *6*(6), 941–953. https://doi.org/10.1175/JHM468.1

Demir, I., & Krajewski, W. F. (2013). Towards an integrated Flood Information System: Centralized data access, analysis, and visualization. *Environmental Modelling and Software*, *50*, 77–84. https://doi.org/10.1016/j.envsoft.2013.08.009

Demir, I., Mantilla, R., Small, S., Krajewski, W. F., Riteau, P., Armstrong, P., Keahey, K., Seo, B., & Goska, R. (2015). *Hydroinformatics On The Cloud: Data Integration, Modeling And Information Communication For Flood Risk Management*. *February*, 1–5.

Ehlschlaeger, C. . (1989). Using the AT Search Algorithm to Develop Hydrologic Models from Digital Elevation Data. *Proceedings of International Geographic Information Systems (IGIS)*.

Fonley, M. R., Qiu, K., Velásquez, N., Haut, N. K., & Mantilla, R. (2021). Development and Evaluation of an ODE representation of 3D subsurface tile drainage flow using the HLM flood forecasting system. *Water Resources Research*, *57*(3). https://doi.org/10.1029/2020WR028177

Harman, C. J., Sivapalan, M., & Kumar, P. (2009). Power law catchment-scale recessions arising from heterogeneous linear small-scale dynamics. *Water Resources Research*, *45*(9), 1–13. https://doi.org/10.1029/2008WR007392

Jadidoleslam, N., Goska, R., Mantilla, R., & Krajewski, W. F. (2020). Hydrovise: A non-proprietary open-source software for hydrologic model and data visualization and evaluation. *Environmental Modelling and Software*, *134*, 104853. https://doi.org/10.1016/j.envsoft.2020.104853

Kalma, J. D., McVicar, T. R., & McCabe, M. F. (2008). Estimating land surface evaporation: A review of methods using remotely sensed surface temperature data. *Surveys in Geophysics*, *29*(4–5), 421–469. https://doi.org/10.1007/s10712-008-9037-z

Krajewski, W. F., Ceynar, D., Demir, I., Goska, R., Kruger, A., Langel, C., Mantilllla, R., Niemeier, J., Quintero, F., Seo, B. C., Smallll, S. J., Weber, L. J., & Young, N. C. (2017). Real-time flood forecasting and information system for the state of Iowa. *Bulletin of the American Meteorological Society*, *98*(3), 539–554. https://doi.org/10.1175/BAMS-D-15-00243.1

Mandeville, A. N. (2016). Insights gained from four component hydrograph separation. *Hydrology Research*, *47*(3), 606–618. https://doi.org/10.2166/nh.2016.061

Mantilla, R., & Gupta, V. K. (2005). A GIS Numerical Framework to Study the Process Basis of Scaling Statistics in River Networks. *October*, *2*(4), 404–408.

Project, O. S. G. F. (2020). *QGIS Geographic Information System*.

Quintero, F., Krajewski, W. F., Seo, B.-C., & Mantilla, R. (2019). Improvement and Evaluation of the Iowa Flood Center Hillslope Link Model (HLM) by Calibration-Free Approach. *Journal of Hydrology*, *584*(July 2019), 124686.

https://doi.org/10.1016/j.jhydrol.2020.124686

Quintero, F., Krajewski, W. F., Seo, B. C., & Mantilla, R. (2020a). Improvement and evaluation of the Iowa Flood Center Hillslope Link Model (HLM) by calibration-free approach. *Journal of Hydrology*, *584*(February), 124686. https://doi.org/10.1016/j.jhydrol.2020.124686

Quintero, F., Krajewski, W. F., Seo, B. C., & Mantilla, R. (2020b). Improvement and evaluation of the Iowa Flood Center Hillslope Link Model (HLM) by calibration-free approach. *Journal of Hydrology*, *584*. https://doi.org/10.1016/j.jhydrol.2020.124686

Reed, S. M., & Maidment, D. R. (1999). COORDINATE TRANSFORMATIONS FOR USING NEXRAD DATA IN GIS-BASED HYDROLOGIC MODELING. *Journal of Hydrologic Engineering*, *4*(April), 174–182.

Running, S., Mu, Q., & Zhao, M. (2017). *MOD16A2 MODIS/Terra Net Evapotranspiration 8-Day L4 Global 500m SIN Grid V006*. NASA EOSDIS Land Processes DAAC.

Samaniego, L., Kumar, R., & Attinger, S. (2010). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, *46*(5), 1–25. https://doi.org/10.1029/2008WR007327

Schilling, K. E., Gassman, P. W., Arenas-Amado, A., Jones, C. S., & Arnold, J. (2019). Quantifying the contribution of tile drainage to basin-scale water yield using analytical and numerical models. *Science of the Total Environment*, *657*, 297–309. https://doi.org/10.1016/j.scitotenv.2018.11.340

Schilling, K. E., & Helmers, M. (2008). *Effects of subsurface drainage tiles on streamflow in Iowa agricultural watersheds : Exploratory hydrograph analysis*. *4506*(May), 4497–4506. https://doi.org/10.1002/hyp

Shaw, S. B., & Riha, S. J. (2012). Examining individual recession events instead of a data cloud: Using a modified interpretation of dQ/dt-Q streamflow recession in glaciated watersheds to better inform models of low flow. *Journal of Hydrology*, *434–435*, 46–54. https://doi.org/10.1016/j.jhydrol.2012.02.034

Sur, C., Park, S. Y., Kim, J. S., & Lee, J. H. (2020). Prognostic and diagnostic assessment of hydrological drought using water and energy budget-based indices. *Journal of Hydrology*, *591*(September), 125549. https://doi.org/10.1016/j.jhydrol.2020.125549

Tallaksen, L. M. (1995). A review of baseflow recession analysis. *Journal of Hydrology*, *165*(1–4), 349–370. https://doi.org/10.1016/0022-1694(94)02540-R

Team, G. D. (2017). *Geographic Resources Analysis Support System (GRASS GIS) Software, Version 7.2*. Open Source Geospatial Foundation.

Velásquez, N., Hoyos, C. D., Vélez, J. I., & Zapata, E. (2020). Reconstructing the 2015 Salgar flash flood using radar retrievals and a conceptual modeling framework in an ungauged basin. *Hydrology and Earth System Sciences*, *24*(3), 1367–1392. https://doi.org/10.5194/hess-24-1367-2020

Velásquez, N., Mantilla, R., Krajewski, W., Quintero, F., & Zanchetta, A. D. L. (2022). Identification and Regionalization of Streamflow Routing Parameters Using Machine Learning for the HLM Hydrological Model in Iowa. *Journal of Advances in Modeling Earth Systems*, *14*(7), e2021MS002855.

Yilmaz, M. T., Anderson, M. C., Zaitchik, B., Hain, C. R., Crow, W. T., Ozdogan, M., Chun, J. A., & Evans, J. (2014). Comparison of prognostic and diagnostic surface flux modeling approaches over the Nile River basin. *Water Resources Research*, *50*(1), 386–408. https://doi.org/10.1002/2013WR014194