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

Nicolás Velasquez, Ricardo Mantilla, Witold F. Krajewski, and Morgan Fonley

Iowa Flood Center, University of Iowa, Iowa City, IA, 52245

To be submitted to JAMES

Corresponding author: Nicolas Velasquez

## Abstract

This evaluates the potential for a newly proposed non-linear subsurface flux equation to improve the performance of the hydrological Hillslope Link Model (HLM). The equation contains parameters that are functionally related to the hillslope steepness and the presence of tile drainage. As a result, the equation allows a better representation of hydrograph recession curves, hydrograph timing, and total runoff volume. The authors explore the new parameterization's potential by comparing a set of diagnostic and prognostic setups in HLM. In the diagnostic approach, they configure 12 different scenarios with spatially uniform parameters over the state of Iowa. In the prognostic case, they use information from topographical maps and known locations of tile drainage to distribute parameter values. To assess performance improvements, they compare simulation results to streamflow observations during a 17-year period (2002–2018) at 140 U.S. Geological Survey (USGS) gauging stations. The operational setup of the HLM model used at the Iowa Flood Center (IFC) serves as a benchmark to quantify overall model improvement. In particular, the new equation provides better representation of recession curves and the total streamflow volumes. However, when comparing the diagnostic and prognostic setups, the authors find discrepancies in the spatial distribution of hillslope scale parameters. The results suggest that more work is required when using maps of physical attributes to parameterize hydrological models. The findings also demonstrate that the diagnostic approach is a useful strategy to evaluate models and assess changes in their formulations.

## Plain Language Summary

In the current work, we test the accuracy of a hydrological model using a novel equation that represents the subsurface flow from hillslopes to channels. The equation has into account the terrain steepness and the presence of tile drainage. We test the model in the state of Iowa comparing its results with records from about 140 USGS streamflow gauges between 2002 and 2018. We set up the model using constant and distributed parameters. In the distributed case, we use a digital elevation map to describe the steepness and a map describing the tiles drainage localization in Iowa. We found that regardless of the setup the novel equation improves the model performance. The fixed-distributed approach gave us information about the strengths and weaknesses of the model and its inputs. According to our results, distributed variables do not necessarily improve the model performance suggesting that more work is required to find the adequate spatial parameter distribution for hydrological models.

## Key findings

- The use of a non-linear equation to represent subsurface flux improves the hydrological model performance at multiple scales.
- Parameters derived from maps that describe the landscape variability do not necessarily help the model to improve its performance.
- The diagnostic-prognostic approach helps to identify the limits of a non-linear equation using distributed information.

# 1 Introduction

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).

## 1.1 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 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 Kling Gupta Efficiency (KGE) index (Figure 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).

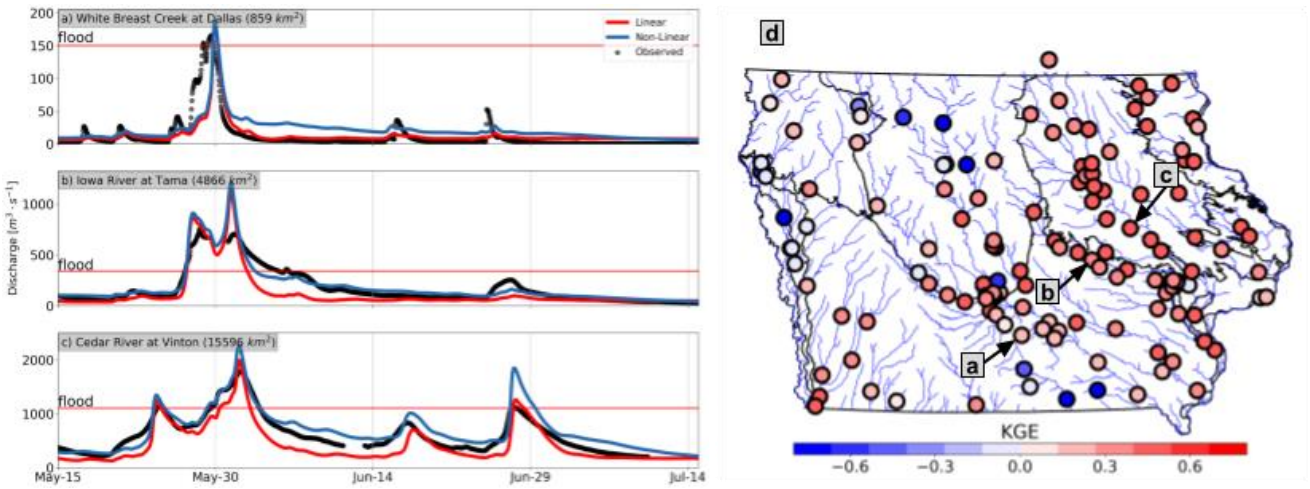


Figure 1. a) Observed (black) and simulated streamflows by the linear (red) and the non-linear (blue) setups at three USGS gauged stations. b) Mean annual KGE performance of the HLM linear setup for Iowa between 2002 and 2018.

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 1a to 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 (Figures 1a and b) attributed to issues with parameter values and spatial representation of processes.

## 1.2 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.

# 2 Methodology and Data

## 2.1 Model description

The Hillslope Link Model (HLM) represents the hydrological processes at the hillslope scale (Figure 2a and b) and routes the streamflow through the channel network (Figure 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).

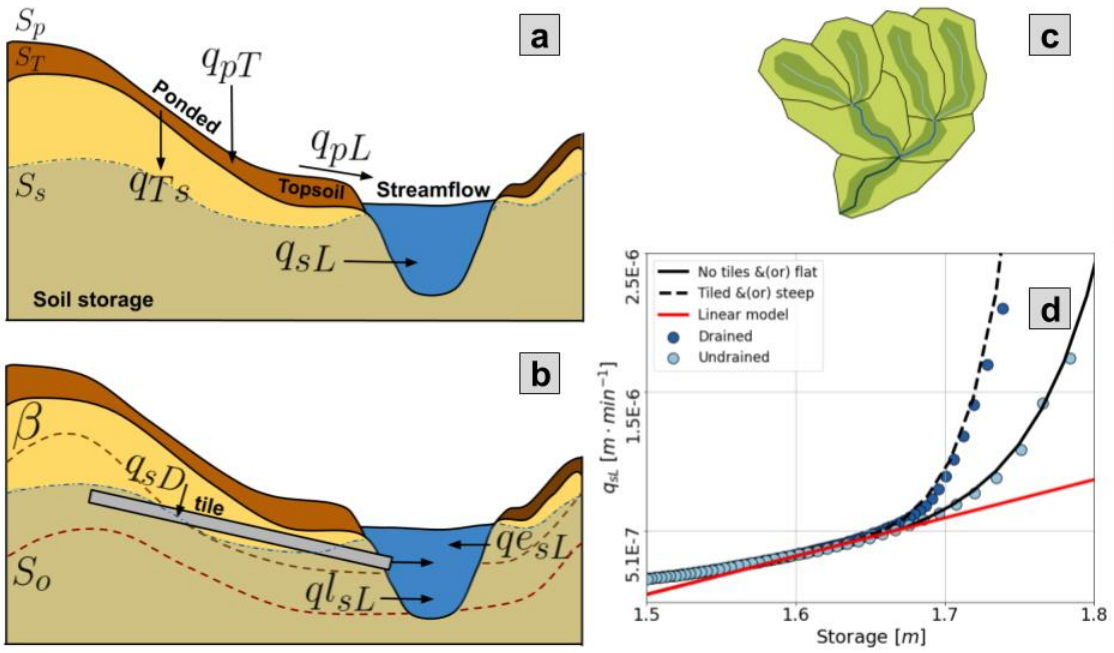


Figure 2. Hillslope Link Model spatial discretization and schematic of the storages and processes represented at the hillslope scale: a) HLM hillslope process using the linear subsurface flux equation; b) hillslope process including the active layer ($\beta$), the exponential flux ($qe_{sL}$), and the 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 $k_2$ [$\mathrm{min}^{-1}$] (Equation (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 (2) and (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.

$$k_2 = v_r \cdot \left(\frac{L_i}{A_h}\right) \cdot 60 \quad (1)$$

$$q_{pL} = k_2 \cdot s_p \quad (2)$$

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

$$q_{Ts} = k_2 \cdot s_T \cdot k_i \quad (4)$$

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 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,

$$q_{sL} = m \cdot (s_s - S_o) \quad (5)$$

Fonley et al. (2021) developed a set of parameterizations for ordinary differential equations that adds a non-linear component to Equation (5) when $S_s$ is above threshold storage. The following

exponential equation (continuous line on Figure 2d) is added to Equation (5) if $S_s$ is greater than the activation threshold $\beta$ $[m]$,

$$qe_{sL} = \alpha(s_s - \beta)e^{17(s_s-\beta)} \quad (6)$$

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 2d) is added when $S_s$ is greater than the tile relative depth $D_d$ $[m]$,

$$q_{sD} = c(s_s - D_d)e^{35(s_s-D_d)} \quad (7)$$

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,

$$q_sL = \begin{cases} ql_{sL} & if \quad S_s < \beta \\ ql_{sL} + qe_{sL} & if \quad S_s > \beta \\ ql_{sL} + qe_{sL} + q_sD & if \quad S_s > D_d \end{cases} \quad (8)$$

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 (8) is limited to its two first expressions. More details on the subsurface equation development can be found in Fonley et al. (2021).

## 2.2 Model setup and data

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 (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 (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 distribute parameters case, we use the steepness of the hillslopes (Figure 3a) and the tiles localization according to the DNR (Figure 3b).

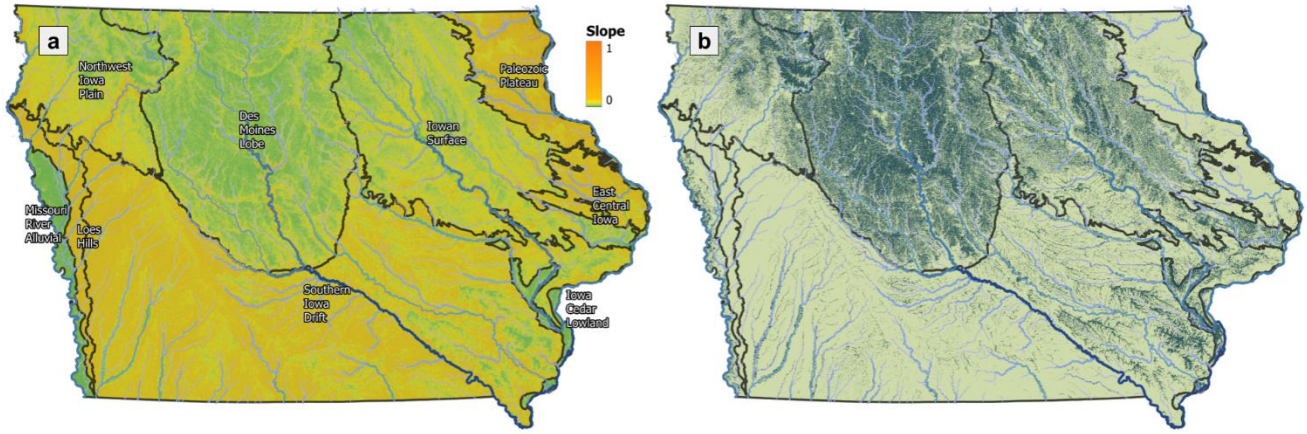


Figure 3. Maps of the hillslope steepness (a) and tile drainage localization according to the Iowa DNR (b).

The model validation consists of comparing fixed (diagnostic) and distributed (prognostic) HLM setups (Figure 4). The diagnostic setup (Figure 4a) shows how different formulations could significantly improve the model over the region. On the other hand, the prognostic setups (Figure 4b) show the improvements and limitations derived from the application of "known" spatial variables.

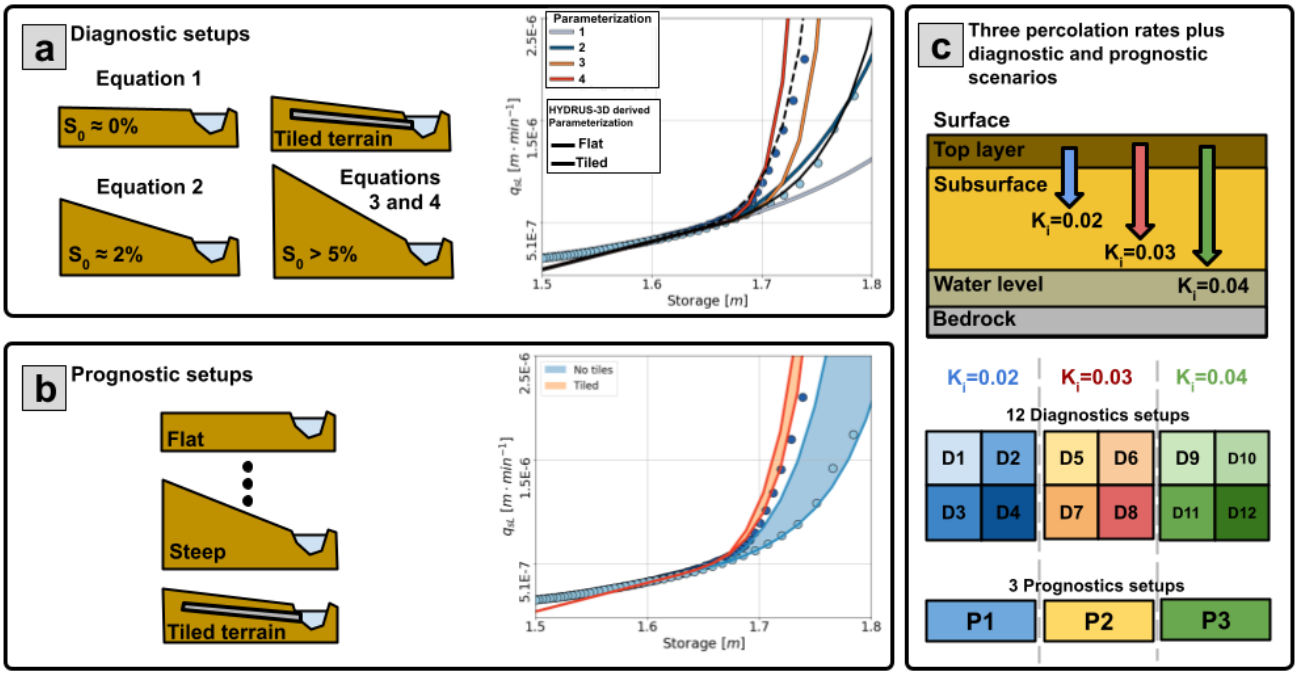


Figure 4. Diagnostic and prognostic experiment setup: a) diagnostic case with four non-linear subsurface flux equations fixed for the domain of Iowa after Fonley et al (2021); b) prognostic case, with equations varying with the hillslopes steepness (blue scenario) and the presence of tiles (red scenario); and c) percolation rates fixed for the different scenarios. Their combination gives us 12 diagnostic scenarios and three prognostic scenarios.

The formulation of Equation (8) relies on the percolation rate because the non-linear formulation depends upon the amount of water in the subsurface storage. The described dependence increases the relevance of the percolation parameter ($k_i$). The distribution of $k_i$ can be derived from maps of the soil profile properties. However, using an additional map may increase the errors affecting the comparison of both setups. For this reason, we choose to fix three different percolation rates for the diagnostic and prognostic setups (Figure 4c).

### 2.2.1 Diagnostic setups

In the diagnostic setup (Figure 4a), we created four parametrizations of Equation (8) for the Iowa domain. The parametrizations range from flat hillslopes (light blue line on Figure 4a) to steep or tiled hillslopes (red line on Figure 4a). By combining the four parameterizations and the three $k_i$ rates, we obtain 12 diagnostic scenarios (D1 to D12 in Figure 4c). D1 to D4 use $k_i = 0.02$; D5 to D8 use $k_i = 0.03$; and D9 to D12 use $k_i = 0.04$.

### 2.2.2 Prognostic setups

In the prognostic setup, we distributed parameter values in function of the hillslopes steepness and the Iowa DNR map describing tiles presence (Figure 4b). According to Fonley et al. (2021), the parameter $\alpha$ of Equation (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 4b,

$$\alpha = \gamma_h(8.5 * 10^{-8}) + 9.48 * 10^{-7} \quad (9)$$

Additionally, we include Equation (8) for tiled terrain following the tile distribution shown by the map in Figure 4b. For the tile drainage equation, we use $c$ to equals $5.4 * 10^{-7}$ (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 4c).

# 3 Results and Discussion

## 3.1 Insights from a diagnostic-prognostic approach

The diagnostic and prognostic setups produced significant differences between the model outputs. In Figure 5, we present the simulated hydrographs at three watersheds simulated by the diagnostic scenario D4 in blue and the prognostic scenario P1 in red. In this case, the diagnostic setup assumes that all the hillslopes have tiles or are steep. On the other hand, the prognostic setup assumes tiles only at some hillslopes and that the parameter $\alpha$ of Equation (6) varies with the steepness. In these three cases, the diagnostic (or fixed) setup produces a longer recession curve than the one obtained by the prognostic setup. The diagnostic case has a better match on the Iowa River at Tama (Figure 5b), while the prognostic setup exhibits a better match on the White Breast Creek (Figure 5a) and at the Cedar River (Figure 5c). Figure 5 gives a brief description of the expected differences between the setups. Also, it shows that Equation (8) can improve the streamflow representation, given the correct set of parameters.

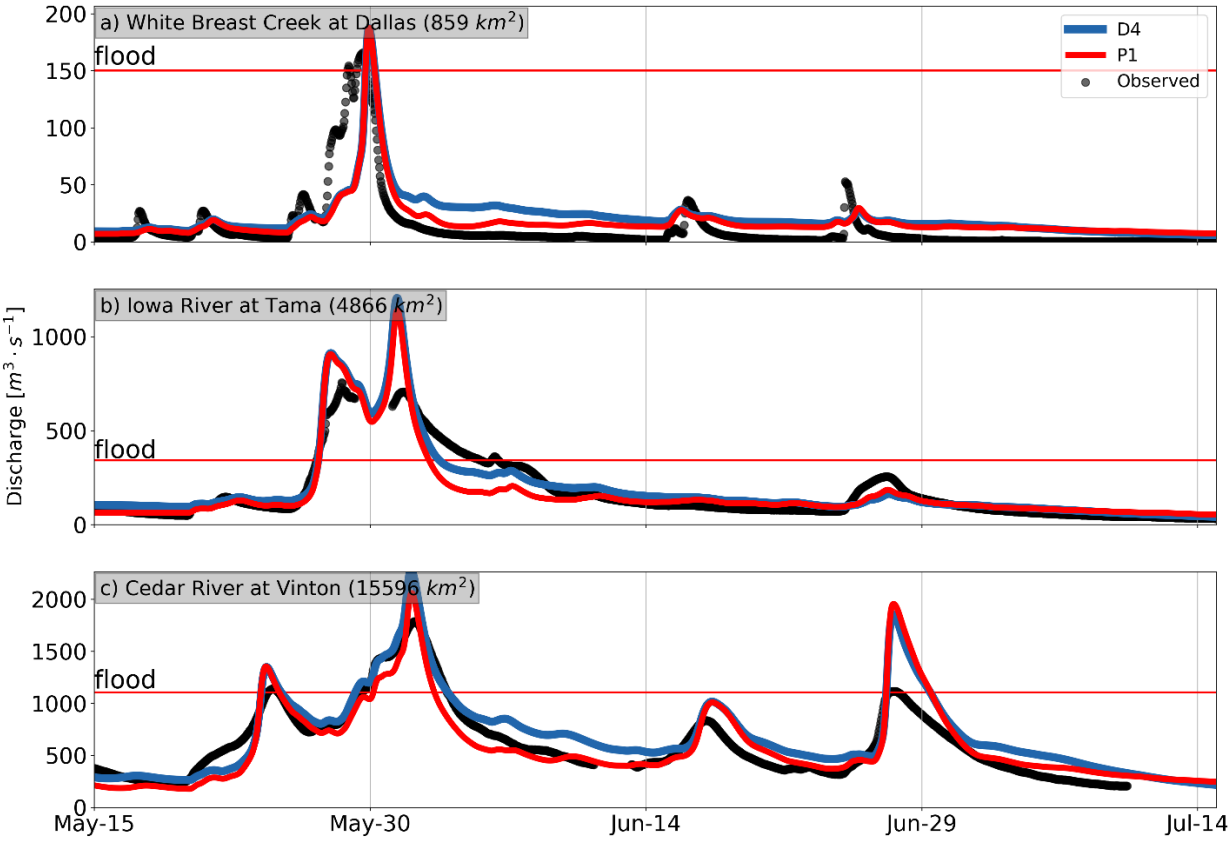


Figure 5. Diagnostic (blue) and prognostic (red) simulations done for the year 2013. Results are compared with the USGS observations (black).

According to Figure 5, the non-linear model can produce a good representation of the hydrograph falling limb and early recession, depending on the parameters. Considering the described sensitivity, we compare the event-based KGE (Equation (10)) of the non-linear setups and the linear model (Figure 6). 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 6), 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 6). Cases such as D5 and D11 exhibited a performance like the one obtained by the linear model. Other cases, such as D9, tended toward 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 (8) can significantly improve the model performance, depending on the parameters.

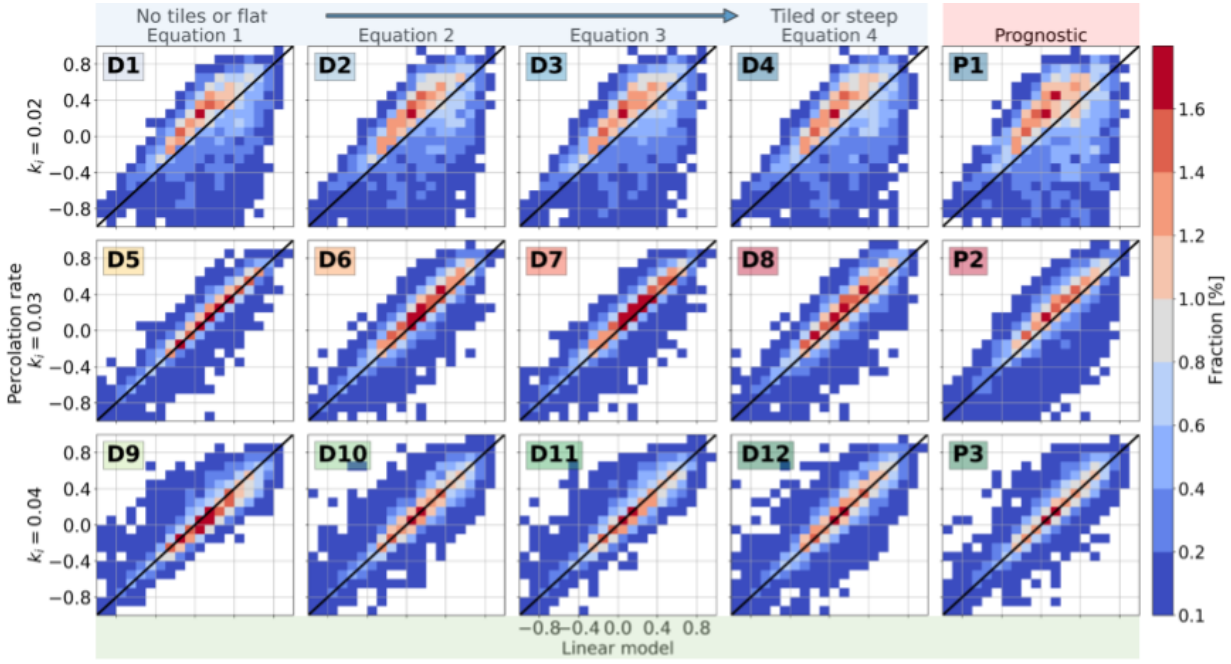


Figure 6. Event-based KGEs comparison between the diagnostics setups and the linear model. Each row corresponds to a fixed percolation rate. Columns correspond to the four fixed equations. The color bar shows the percentage of events that fall at each bin of the 2D histogram.

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 7, we present the KGE distribution and the percentage of time each scenario was chosen. 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 (8) parameters.

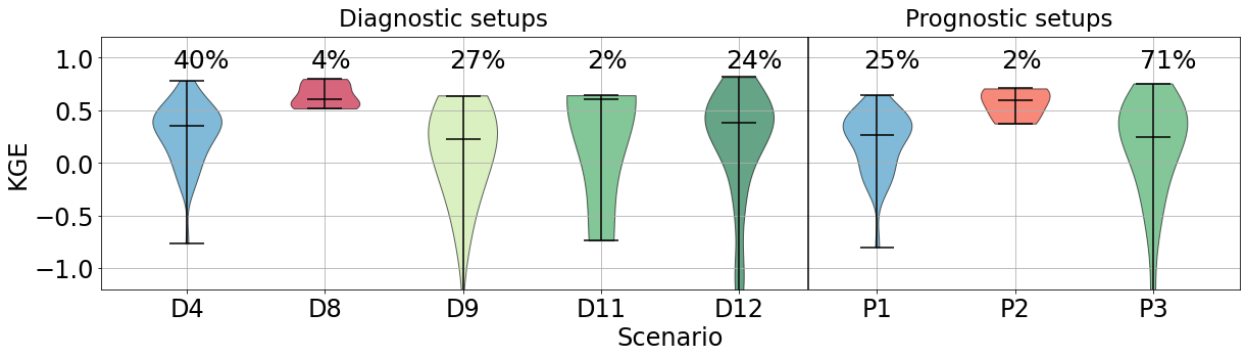


Figure 7. Event-based KGE distribution for the selected scenarios at each station.

The results presented in Figure 7 follow a spatial distribution. Figure 8 shows each USGS gauge colored by the diagnostic (Figure 8a) and prognostic (Figure 8b) setups with the best performance. In both cases, the percolation rate defines the spatial distribution. We can identify how the chosen setups (Figure 8) follow the Iowa landforms to some extent in the diagnostic case (see Figure 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 (Figure 8b) 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 ($k_i = 0.04$).

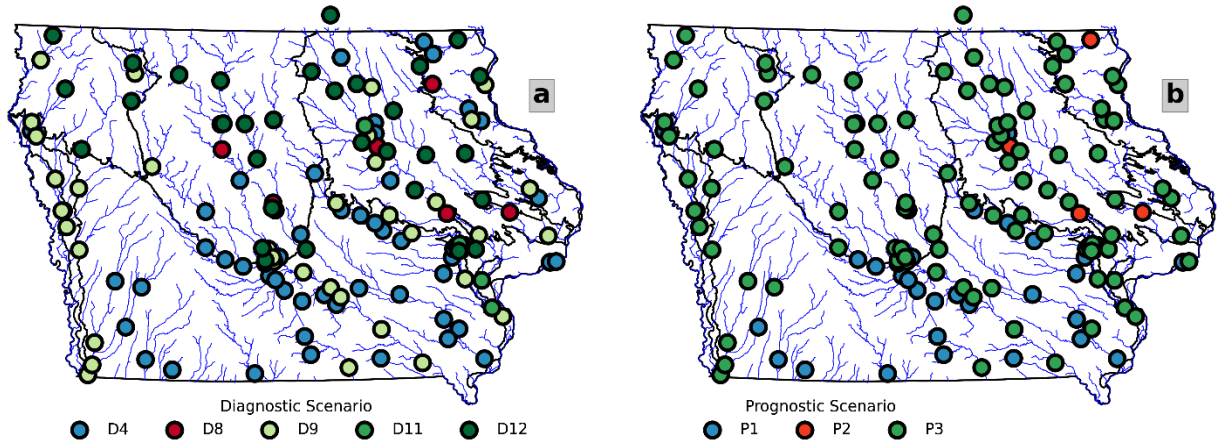


Figure 8. Spatial distribution of the scenarios with best KGE performance at each USGS station: a) results obtained from the diagnostic scenarios; and b) results obtained from the prognostic scenarios. The green, red, and blue gauge colors correspond to the percolation rates of 0.02, 0.03, and 0.04, respectively.

According to Figure 8, the chosen diagnostic and prognostic scenarios share percolation rates. However, differences exist in the spatial performance improvement distribution (Figure 9). Figure 9a 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 9a), the diagnostic scenarios outperform the linear model at almost all the USGS gauges. Alternatively, in the prognostic case (Figure 9 b), 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.

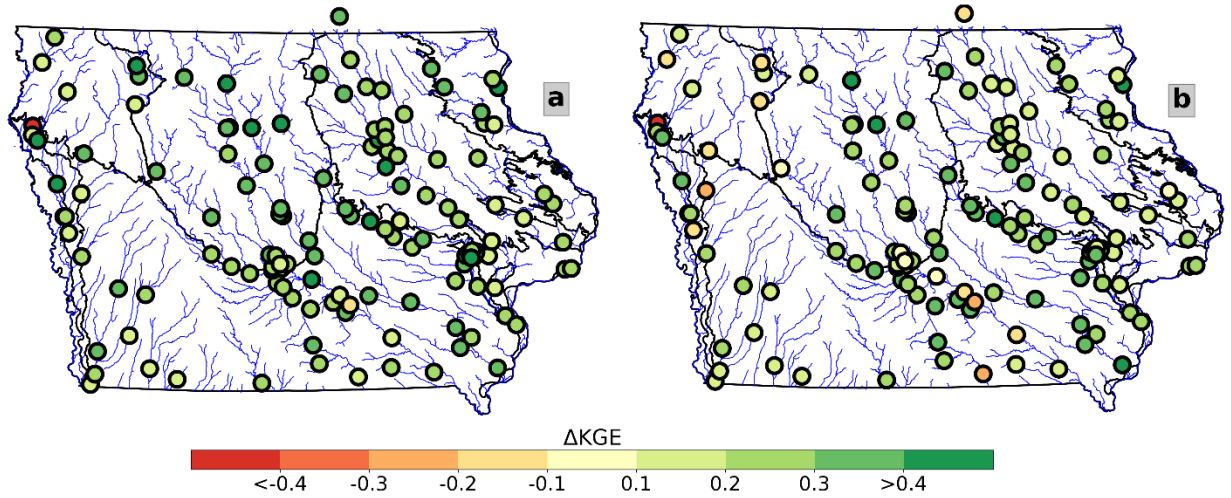


Figure 9. Mean KGE spatial difference of the diagnostic and prognostic scenarios with respect to the linear model: a) diagnostic KGE minus linear model KGE; and b) prognostic KGE minus linear model KGE.

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 (Figure 9b). In this region, the chosen diagnostic setups were D12 and D9 (Figure 8 a), 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 11) and the mean ratio (lower diagonal in Figure 11) of the chosen scenarios. According to Figure 11, 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 11 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.

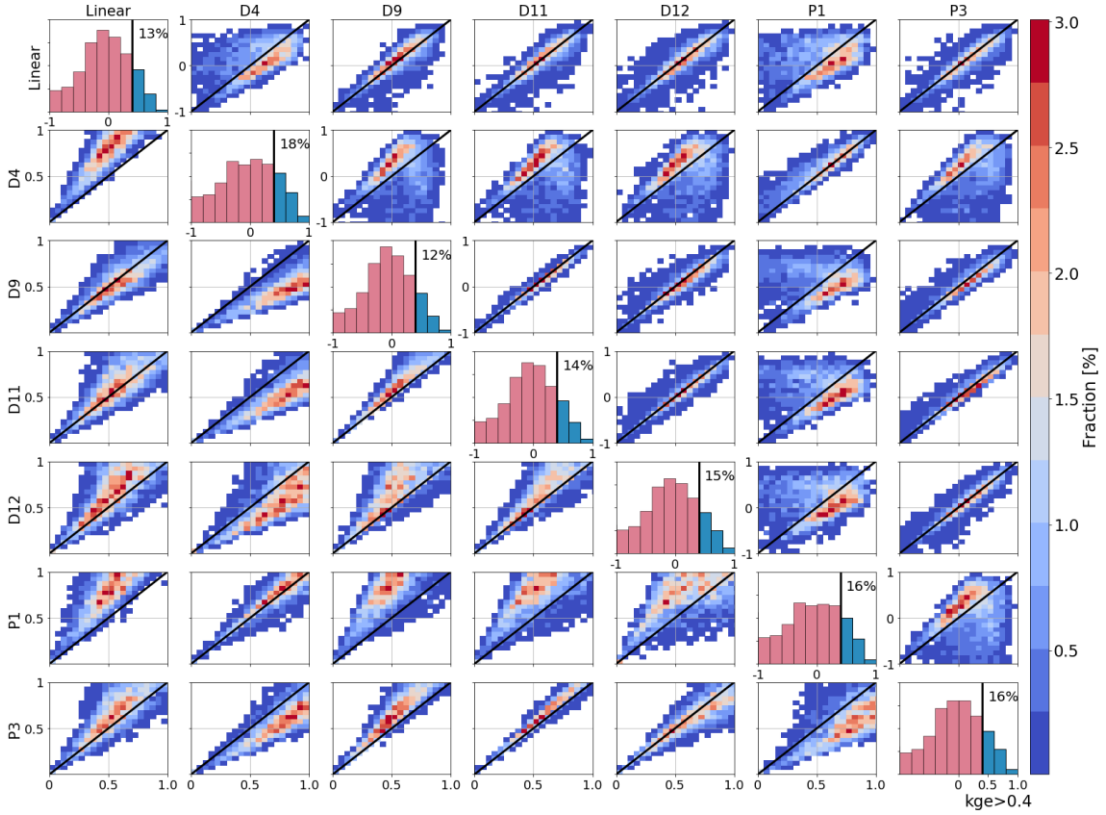


Figure 10. Event-based KGE comparison of the diagnostic and prognostic dominant scenarios. Each row compares a scenario against the others. The upper diagonal panels correspond to the KGE histogram of the scenarios. Over the diagonal shows the KGE histogram of each setup coloring in blue the percentage 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 4a. This parameterization represents highly conductive soils or the presence of tiles. On the other hand, P1 parameters follow the hillslope steepness with Equation (9), and the presence of tiles described by the map in Figure 4b. The described differences in the parameters seem to develop slight dissimilarities in the performance. According to panel D4-P1 in Figure 11, 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.

### 3.2 Extended metrics

According to the diagnostic and prognostic KGE comparisons, the performance differences between the two scenarios are relatively small. However, the KGE is subject to three parameters that do not necessarily reflect all the relevant changes in the simulated streamflows. With this in mind, we also compared the NSE (Nash Sutcliffe efficiency), the hit rate, and the lags (Figure 11 a, b, and c, respectively). The NSE contrasts the simulated data prediction skill with the mean value of the observations. NSE below 0 indicates that the mean value performs better than the model, and an NSE of 1 indicates a perfect simulation. The hits rate is the percentage of time that observations and simulations share during floods. A hit rate of zero corresponds to missing all the floods, and a hit rate of one corresponds to a perfect match. The lag is a measure of the displacements applied over the simulated data to maximize the correlation. We make hourly displacements from -48h to 48h. Negative lag values correspond to cases in which the simulated

data exhibit responses earlier than the observed, and positive values correspond to the opposite behavior. A simulated series with good performance must have lags near zero.

Results from Figure 11 shows that D4 and P1 have some similarities and some relevant differences. Regarding the NSE (Figure 11a), D4 has a slightly better score. Regarding the hits rate (Figure 11b), it is hard to tell which scenario has a better performance. P1 has a higher fraction of hit rates approaching one, but it also has higher frequencies at some lower intervals. The number of lags is also similar (Figure 11c). Nevertheless, P1 tends toward negative lag values more than D4 does, representing more frequent early peak estimations.

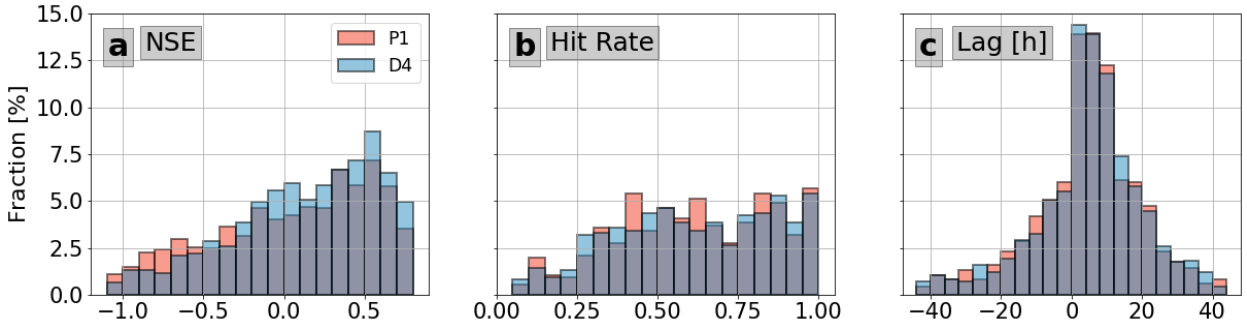


Figure 11. Histograms of performance metrics for the scenarios D4 and P1. Panel a compares the NSE. Panel b compares the Hit Rate considering a Hit when the simulated and observed streamflow are above the flood level of the gauge. Panel c compares the Lags (in hours) required to maximize the correlation between observations and simulations.

In addition to the indexes, we compare the simulated peaks of the chosen diagnostic and prognostic scenarios. Because the gauged watersheds areas range between 40 and 18,000 $km^2$, we performed a scale-independent comparison. To obtain scale-independent peaks ($Z$), we divided the peaks $Q_p$ $[m^3 \cdot s^{-1}]$ by the mean annual peak $\overline{Q_p}$ $[m^3 \cdot s^{-1}]$. Then, for each event of each link, we computed the difference between the simulated ($Z_s$) and the observed ($Z_o$) standardized peaks.

The peak flow estimation of D4 and P1 exhibited a similar performance, with D4 being superior. The D4 scenario reaches a fraction of 32% for differences near zero (Figure 12), while in P1 this value drops to 28% (Figure 12). Also, P1 has a higher fraction of errors equal or greater than one.

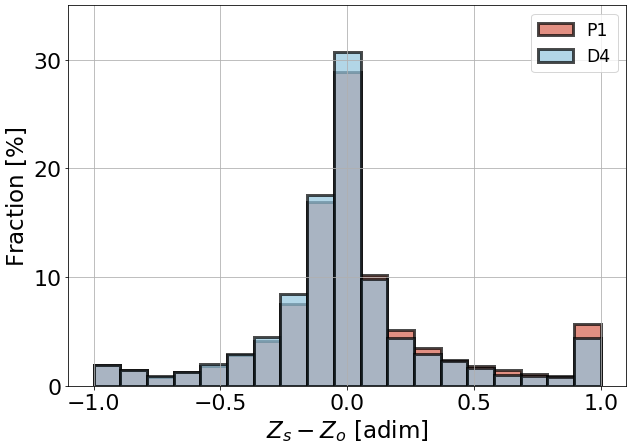


Figure 12. Histograms of the standardized peak flows difference for the D4 (blue) and P1(red) scenarios.

We expected the diagnostic superiority because in the prognostic case, we impose restrictions based on maps. The resulting differences among simulations emphasize the parameters' relevance and the need for their correct representation. Contrasted with the diagnostic scenarios, the prognostic scenarios tend to reduce the performance. The differences between D4 and P1 suggest that the landscape descriptors could have errors that lead to decreases in the modeling performance. Also, our results suggest that there may be more tiles than the ones represented by the map in Figure 4b.

### 3.3 Analysis of parameter values

The diagnostic and prognostic scenarios offer different ways to determine the values of parameters in space. In the diagnostic cases, we identified the best fixed-parameter combination for each gauged watershed. In the prognostic cases, we pre-defined a set of distributed parameters based on available information. In a previous step, we defined the best diagnostic and prognostic setup for each gauge (Figure 8a and b, respectively). According to our results, a spatial distribution of the parameters seems to be explained by $k_i$ and the parameters of Equation (8). Additionally, the chosen diagnostic scenarios outperform the chosen prognostic ones (Figure 9a and b). In some gauges, the performance differences are small; however, in others, the difference is relatively large. This is an interesting result because the only difference between both cases is the Equation (8) parameterization. Considering the described performance differences, we explore in more detail how they are related to the parameterizations of the diagnostic and prognostic setups.

We compare the Equation (8) setup for the diagnostic and prognostic scenarios to evaluate whether variations in the parameters explain the observed performance differences. We made the comparison at each gauged watershed. For the comparison, the prognostic setup has a set of curves $q_{sL}(P)$ for a given watershed (light blue lines in Figure 13), and there is one diagnostic curve $q_{sL}(D)$ for the same watershed (dark blue line in Figure 13). Using Equation (11), we compare $q_{sL}(D)$ with the $50^{\text{th}}$ percentile of $q_{sL}(P)$ for storages between 1.6 and 1.705$[m]$ (green region in Figure 13).

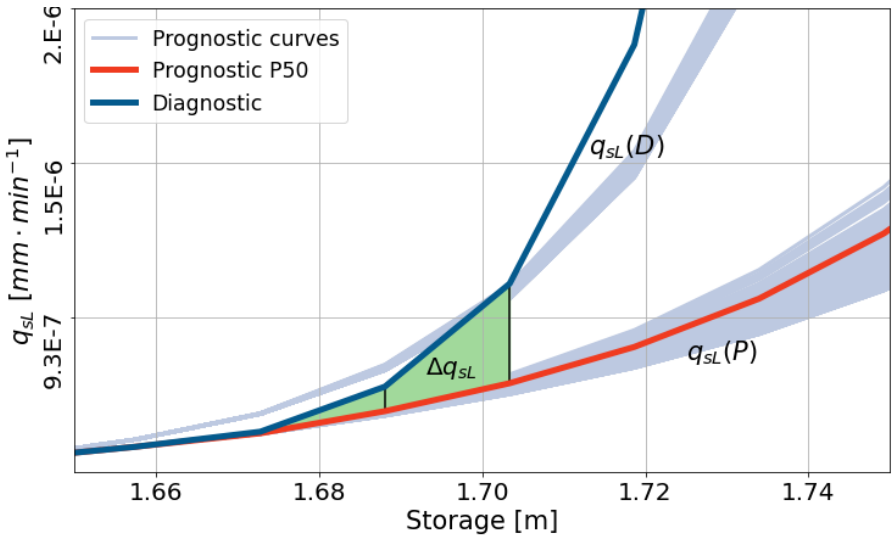


Figure 13. Example of the $q_{sL}$ parameterization comparison between the diagnostic and prognostic scenarios. The light blue lines are to the $q_{sL}$ curves of the prognostic scenario for a given watershed. The red line corresponds to the 50$^{th}$ percentile of the prognostic $q_{sL}$ curves. The dark blue line is the $q_{sL}$ curve of the diagnostic scenario. We used the green region to perform a comparison between the scenarios.

$$\Delta q_{sL} = \frac{1}{N} \cdot \sum_{i}^{N} \frac{q_{sL}(D)_i - P50(q_{sL}(P))_i}{q_{sL}(D)_i} \quad (11)$$

According to Figure 14, the differences of the parameters ($\Delta q_{sL}$) hardly explains the performance differences between the diagnostic and the prognostic scenarios. In some cases, low absolute values of $\Delta q_{sL}$ are linked to low KGE differences. However, the described behavior does not apply for large absolute values of $\Delta q_{sL}$. Figure 14 shows many watersheds in which the differences of the absolute parameters are larger than 10% (x-axis), while the KGE absolute differences are low. Also, some cases with low absolute $\Delta q_{sL}$ exhibit large KGE differences. On the other hand, according to the colors of Figure 14 (non-absolute $\Delta q_{sL}$), positive values of $\Delta q_{sL}$ are related to low KGE differences; and negative values of $\Delta q_{sL}$ correspond to high KGE differences.

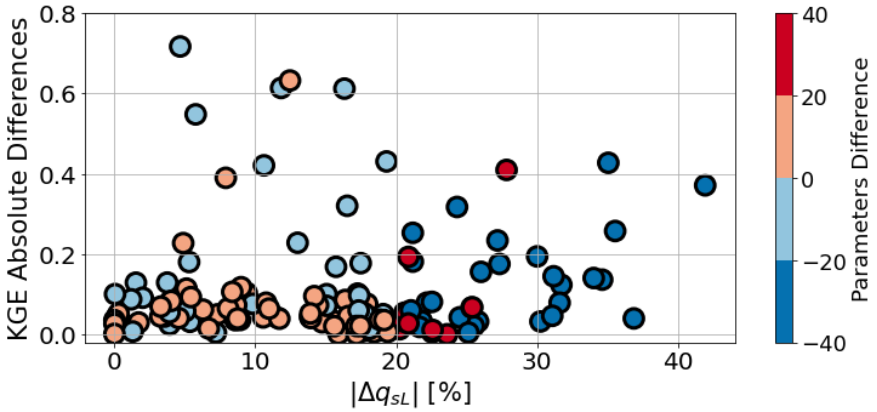


Figure 14. Parameters' absolute differences vs. KGE differences at the USGS gauges. The colors correspond to non-absolute differences in the parameters.

We also compared the $\Delta q_{sL}$ and the KGE differences in space. According to Figure 15, the coincidences between the KGE and $\Delta q_{sL}$ do not show a strong regional pattern. We observe some similarities only in the Des Moines Lobe and the Northwest Iowa Plains regions. In both cases, some significant KGE differences match with large absolute $\Delta q_{sL}$ values. Also, there is a match between low KGE and $\Delta q_{sL}$ differences in the Iowan Surface region, with some exceptions.

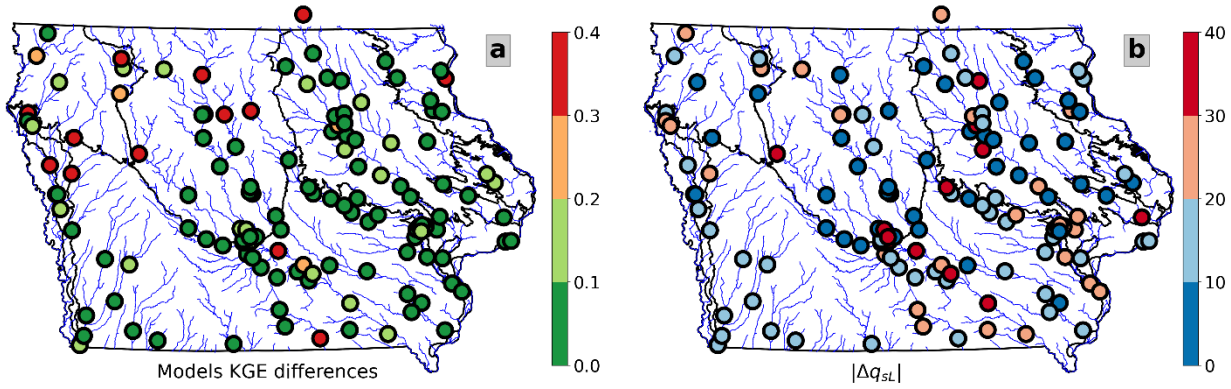


Figure 15. Maps showing KGE differences (a), and $\Delta q_{sL}$ (b), in USGS gauges.

435 It is hard to establish a relationship between the diagnostic and prognostic parameters and their
436 performance differences. We attribute this lack of correlation to the model's non-linear
437 transformations at the hillslope scale and throughout the network. It is expected that parameters
438 would alter the model's output. However, our results show that a pre-defined distribution of the
439 parameters could lead to modeling errors that are hard to identify.

440 # 4 Conclusions

441 The Iowa Flood Center (IFC) has been making operational flood forecasts for the state of Iowa
442 since 2010. IFC forecasters use the hydrological Hillslope Link Model (HLM), along with rainfall
443 data. HLM has been accurate in forecasting streamflow fluctuations at several scales (Quintero et
444 al., 2020). However, the model has limitations in its representation of the recession curve, and it
445 underestimates the total streamflow volume. Moreover, the limitations seem to increase in a tiled
446 landscape. Fonley et al. (2021) attributed these limitations to the linear equation HLM uses to
447 represent the subsurface flux and the lack of an equation representing tiled terrain. To address this
448 issue, Fonley et al. (2021) developed an exponential equation that can be parameterized to
449 represent the function of the hillslope steepness and the presence of tiles.

450 This paper evaluated the exponential equation proposed by Fonley et al. (2021), which represents
451 subsurface hillslope-link interaction in HLM. The equation can represent hillslopes with and
452 without tile drainage. We performed the equation evaluation at 140 USGS gauges in Iowa. The
453 analysis used hourly records between 2002 and 2018. In the evaluation, we compared the
454 exponential equation with a linear equation. The comparison used a diagnostic and a prognostic

approach to set up the parameters. 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.

In the current work, 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 related to 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 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.

## Acknowledgements

This work was completed with support from the Iowa Flood Center, MID-American Transportation Center (Grant number: 69A3551747107), the Iowa Highway Research Board and Iowa Department of Transportation (Contract number:TR-699).

## Data

The data supporting the results of this work can be found in the following repository:

```
https://doi.org/10.5281/zenodo.4796661
```

## References

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

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

Fonley, M., Qiu, K., Velasquez, N., Haut, N. ., & Mantilla, R. I. (2021). Development and Evaluation of an ODE Representation of 3D Subsurface Tile Drainage Flow Using the HLM Flood Forecasting System. *Water Resource 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

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.

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:

573 Using a modified interpretation of dQ/dt-Q streamflow recession in glaciated watersheds to
574 better inform models of low flow. *Journal of Hydrology*, *434–435*, 46–54.
575 https://doi.org/10.1016/j.jhydrol.2012.02.034

576 Sur, C., Park, S. Y., Kim, J. S., & Lee, J. H. (2020). Prognostic and diagnostic assessment of
577 hydrological drought using water and energy budget-based indices. *Journal of Hydrology*,
578 *591*(September), 125549. https://doi.org/10.1016/j.jhydrol.2020.125549

579 Tallaksen, L. M. (1995). A review of baseflow recession analysis. *Journal of Hydrology*, *165*(1–
580 4), 349–370. https://doi.org/10.1016/0022-1694(94)02540-R

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

585

586