Streamflow Forecasting without Models

Witold F. Krajewski, Ganesh R. Ghimire, and Felipe Quintero

Iowa Flood Center and IIHR-Hydroscience & Engineering, The University of Iowa, Iowa City, Iowa 52242.

*Corresponding author: Witold F. Krajewski, witold-krajewski@uiowa.edu

Abstract

The authors explore simple concepts of persistence in streamflow forecasting based on the real-time streamflow observations from the years 2002 to 2018 at 140 U.S. Geological Survey (USGS) streamflow gauges in Iowa. The spatial scale of the basins ranges from about 7 km^2 to 37,000 km^2. Motivated by the need for evaluating the skill of real-time streamflow forecasting systems, the authors perform quantitative skill assessment of different persistence schemes across spatial scales and lead-times. They show that skill in temporal persistence forecasting has a strong dependence on basin size, and a weaker, but non-negligible, dependence on geometric properties of the river networks in the basins. Building on results from this temporal persistence, they extend the streamflow persistence forecasting to space through flow-connected river networks. The approach simply assumes that streamflow at a station in space will persist to another station which is flow-connected; these are referred to as *pure spatial persistence forecasts* (PSPF). The authors show that skill of PSPF of streamflow is strongly dependent on the monitored vs. predicted basin area-ratio and lead-times, and weakly related to the downstream flow distance between stations. River network topology shows some effect on the hydrograph timing and timing of the peaks, depending on the stream gauge configuration. The study shows that the skill depicted in terms of Kling-Gupta efficiency (KGE) > 0.5 can be achieved for basin area ratio > 0.6 and lead-time up to three days. The authors discuss the implications of their findings for assessment and improvements of rainfall-runoff models, data assimilation schemes, and stream gauging network design.

1 Introduction

In this paper we demonstrate, in a systematic way, that persistence is a hard-to-beat streamflow forecasting method (Palash et al. 2018), a fact well-known to operational forecasters. We limit our considerations to real-time forecasting in space and time in a river network. River networks aggregate water flow that originates from the transformation of rainfall and/or snowmelt to runoff. Additional recognition of the importance of river networks stems from the fact that many communities reside by streams and rivers and thus require streamflow forecasts on a regular basis for a number of applications ranging from flood protection to water resource management.

Our study is motivated by ten years of experience of the Iowa Flood Center (IFC) providing operational real-time streamflow forecasts for the citizens of Iowa (Demir and Krajewski 2013; Krajewski et al. 2017). The IFC has accomplished this in a traditional way, i.e. using a state-wide implementation of a distributed rainfall-runoff model. The IFC model is referred to as the Hillslope-Link Model (HLM) and has its origins in landscape decomposition discussed by Mantilla and Gupta (2005). The HLM has been presented and its performance discussed in several prior studies, including Cunha et al. (2012), Ayalew et al. (2014), ElSaadani et al. (2018), and Quintero et al. (2018). In addition, the National Weather Service, via its North Central and Central River Forecasts Centers issues “official” streamflow forecasts for a number of locations throughout the state of Iowa (e.g. Zalenski et al. 2017).

In this study, we take a step back and ask the question: how well could someone forecast streamflow without a rainfall-runoff model? What would be the skill and limitations of such forecasts? We address these questions using the approach of simple persistence applied both in space and in time. We explore a range of forecast lead times, from hours to days ahead, and characterize their skill according to several popular metrics used in evaluation of hydrologic models. Our study domain is the entire state of Iowa but without considering the border rivers, i.e. the Mississippi and the Missouri Rivers. Iowa has a dense drainage network of streams and rivers that feed its border rivers. The state is home to some 1,000 communities located near streams and rivers and is frequented by severe storms and the resultant floods. The U.S. Geological Survey (USGS) operates and collects streamflow data in real time at some 140 stream gauges in Iowa and the IFC has deployed 250 stage-only sensors (Kruger et al. 2016; Krajewski et al. 2017).

Streamflow forecasting in space without using hydrologic models is little studied, and limited efforts have been documented in the literature. Some recent probabilistic approaches (Botter et al. 2008; Müller et al. 2014) include a physically based mechanism for hydrologic predictions in sparsely gauged basins under changing climatic conditions (Doulatyari et al. 2015; Müller and Thompson 2016; Dralle et al. 2016). However, the spatial cross-correlations among multiple basin outlets have not been studied in this framework (Betterle et al. 2017). Specifically, geo-statistical methods have been effectively employed to explore spatial patterns in streamflows accounting for the river network topology (e.g. Betterle *et al.*, 2017; Castiglioni *et al.*, 2011; Laaha *et al.*, 2014; Skøien *et al.*, 2006). Spatial streamflow correlation has been proven to be a better indicator in transferring the flow attributes to ungauged outlets from gauged outlets in comparison to proximity (Betterle et al. 2017). Notwithstanding the progress achieved over the last few years in spatial streamflow predictions, further efforts on exploiting the spatial and temporal persistence of streamflow in the context of streamflow forecasting is needed and this study follows in that direction.

We organize this paper as follows. In the next section, Experimental Domain and Data, we present the data used in our investigations and describe Iowa and its major river basins. In the Methods section, we describe the experimental setup for our analysis and discuss the performance metrics used to characterize the skill of the persistence forecasting. The Results section documents our findings reported as a function of spatial scale characteristics. We also discuss the connection to the river network geometry depicted in terms of width function. We end with a discussion of implications of our study for operational forecasting systems, hydrologic and hydraulic models, river basin gauging network design as well as data assimilation and Artificial Intelligence (AI) methods.

2 Experimental Domain and Data

We explore the forecasting skill of simple persistence in both time and space. Our domain of interest is the state of Iowa and its interior rivers (Figure 1). Iowa is bordered by the Mississippi River on the east and the Missouri River on the west side. The interior rivers all discharge to these two large rivers with about two-thirds of the state's 150,000 km^2 area draining to the Mississippi. Most drainage basins of those interior rivers are contained within the administrative borders of the state, but a small portion drains parts of southern Minnesota and South Dakota.

The climate of Iowa is characterized by marked seasonal variations due to its latitude and interior continental location. During the warm season, the moist air coming from the Gulf of Mexico produces a summer rainfall maximum; during the cold season, flow of dry Canadian air causes cold and dry winters. The air masses moving across the western United States throughout the year produce mild and dry weather.

The landscape is predominantly agricultural, characterized by low relief, fertile soils, high drainage density and thus numerous rivers (e.g., Prior 1991). Much of the state is covered by tile drains, thus changing the hydrologic regime of the region (e.g., Schilling et al. 2019). The combination of the characteristics of its landscape, together with significant rainfall during the summer months makes the state prone to the occurrence of floods (e.g., Villarini et al. 2011a, 2013; Mallakpour and Villarini 2015; Villarini et al. 2011b).

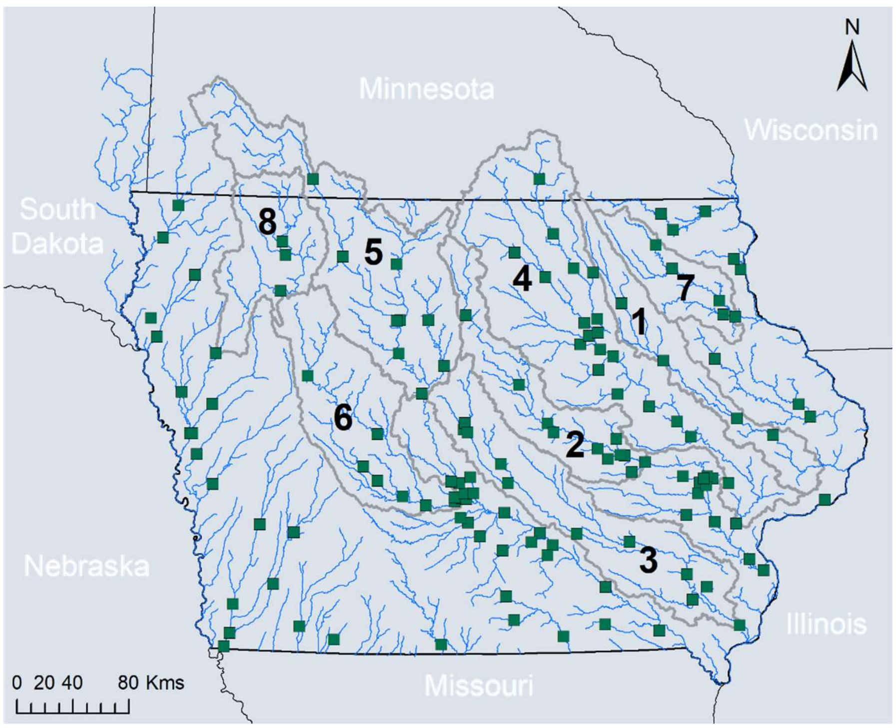

Figure 1: USGS stream gauges over the state of Iowa used in this study. Numbered labels show the river basins with multiple gauges thus enabling assessment of skill of persistence forecasting in space.

The U.S. Geological Survey currently operates about 140 stream gauges in the state (Figure 1), which provide 15-minute resolution stage data that can be converted to discharge using well-maintained and frequently updated rating curves. The USGS record of 15-minute data has only

17 years but we claim that this is sufficient to illustrate the objectives of our study. Though the Iowa Flood Center has added 250 stage-only gauges (Kruger et al. 2016), we will not use the data in this study. Figure 2 shows the basin size histogram at the outlet of these two sets of gauges. It is clear that the IFC emphasizes monitoring smaller basins than the USGS.

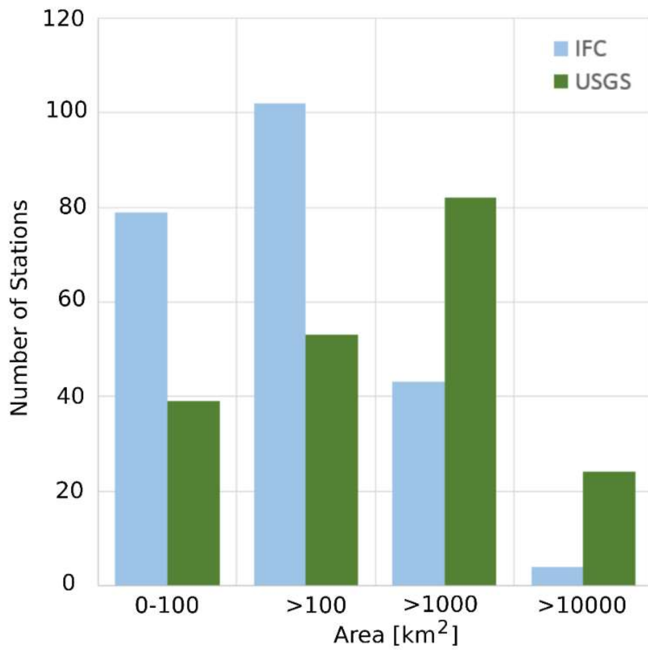

Figure 2: Distribution of the IFC (blue) and the USGS (green) stream gauges over the state of Iowa. It is apparent that most USGS gauges monitor streamflow at larger basin scales.

In addition to the discharge data, we will make use of the water flow velocity data collected by the USGS. The velocity data are synthetized into a simple power-law model documented by Ghimire et al. (2018). In the later part of the paper, we will refer to the model simulation results for the Iowa domain. The model, called Hillslope Link Model (HLM) is briefly presented in Krajewski et al. (2017) and several other publications (e.g., Quintero et al. 2016; ElSaadani et al. 2018; Cunha et al. 2012; Ayalew et al. 2014b; Small et al. 2013). A notable point of the model's use in the operation of the IFC is that the model has not been calibrated in the classic meaning of the practice.

3 Methods

3.1 Experimental Design

We investigated the streamflow forecasting performance of a simple persistence approach in time and in space. We considered a range of spatial scales (~10-10,000 km^2) as well as a range of the forecast lead times (15 minutes to 5 days). Let us denote locations where we would like to forecast streamflow with index i and the stations which data we use to make the forecast as j. In the time domain, the streamflow Q forecasting equation becomes simply

$$Q_i(i, t_0 + \Delta t) = Q_i(i, t_0) \qquad (1)$$

where, the time of issuing the forecast is t_0 and Δt is the forecast lead time, Δt=[0, Δt_{max}]. Here, we simply repeat the main findings documented in Ghimire and Krajewski (2019). These authors investigated three different persistence approaches for Δt ranging between 15 minutes and five days, but here we present the results of the simplest one only, i.e. that the future will be the same as the present, as given by (1). In Figure 3, we illustrate the result from this approach with 15-minutes of streamflow data from 2002 to 2018 at about 140 USGS stations. It is clear that for small scale basins, where the streamflow variability is large and the basin response time short, the persistence forecasting quickly loses skill. For large basins, the skill persists for considerable forecast lead time.

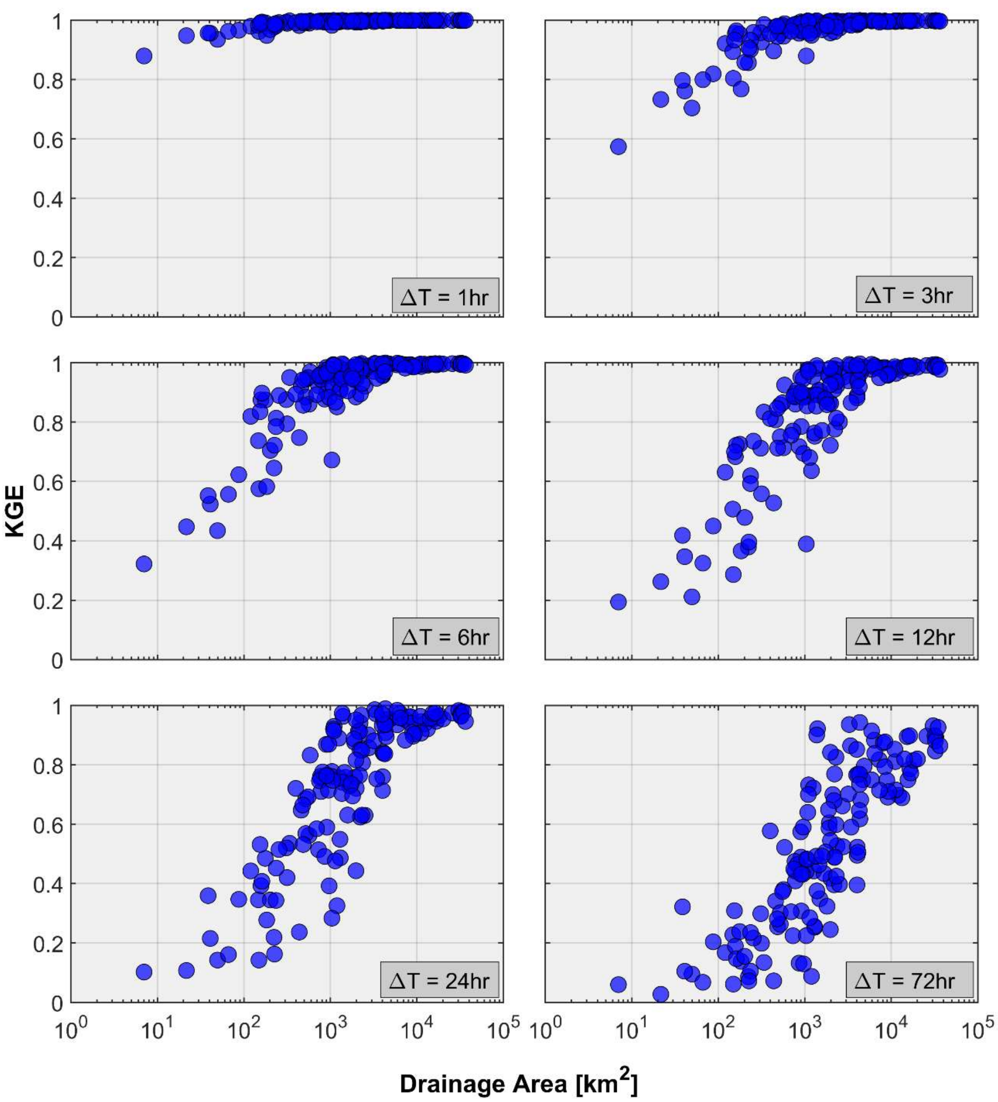

Figure 3: Variability of Kling-Gupta Efficiency based on simple persistence forecasting approach across basin scales for different lead-times (after Ghimire and Krajewski 2019). The forecast skill in terms of KGE shows clear dependence on basin scales and forecast time horizon.

In space, to be fully parallel to the investigation in the time domain, we first investigated scenarios where time is not involved. Here, however, the 140 USGS stream gauges must be stratified not only by basin size but also by basin membership. We only considered gauges that belong to the same basin and thus are flow-connected in either downstream or upstream direction (but not both). Therefore, we excluded consideration of hydroclimatic similarity and land use or ecosystem classification (e.g., Schilling et al. 2015). For example, a basin in western Iowa is excluded from considerations for persistence forecasting in eastern Iowa.

In Figure 1, we show the major basins in the state and in Table 1 we provide details on USGS gauges in each. Now, within a basin we can consider all potential forecasting vs. evaluation scenarios. For example, consider the situation of a supposed basin with five gauges as numbered in Figure 4. First, consider downstream direction only. Hypothetically, one could apply the persistence forecasting paradigm using data from Gauge 1 to forecast streamflow at Gauge 5. The results might not be very good as Gauge 1 misses the contributions from a large portion of the basin and major tributaries upstream from Gauge 5. Similarly, one could evaluate the forecasting performance at Gauge 5 using data from Gauges 2, 3, and 4. Or, one could evaluate the spatial persistence at Gauge 3 using data from Gauges 1 or 2 but not 4 or 5. Gauge 5 is flow-connected with Gauge 3 but in the upstream direction. Gauge 4 is not flow-connected with Gauge 3 as the connection would require downstream and upstream direction for water travel.

Table 1: Basin characteristics with linear regression-based model parameters for KGE

ID	Basin	Area [km^2]	Number of Stations	Number of Stations/km^2	α	β	R^2	RMSE
1	Wapsipinicon River	6,000	4	0.67	-0.45	1.44	0.94	0.11
2	Iowa River	7,300	11	1.51	-0.39	1.41	0.97	0.09
3	Skunk River	11,300	11	0.97	-0.43	1.44	0.97	0.08
4	Cedar River	20,100	20	1.00	-0.44	1.43	0.98	0.07
5	Des Moines River	14,300	9	0.63	-0.61	1.60	0.96	0.10
6	Raccoon River	9,400	9	0.96	-0.41	1.39	0.96	0.09
7	Turkey River	4,100	5	1.22	-0.38	1.39	0.97	0.08
8	Little Sioux River	6,300	4	0.63	-0.41	1.40	0.98	0.08

* $KGE = \alpha + \beta A_r$

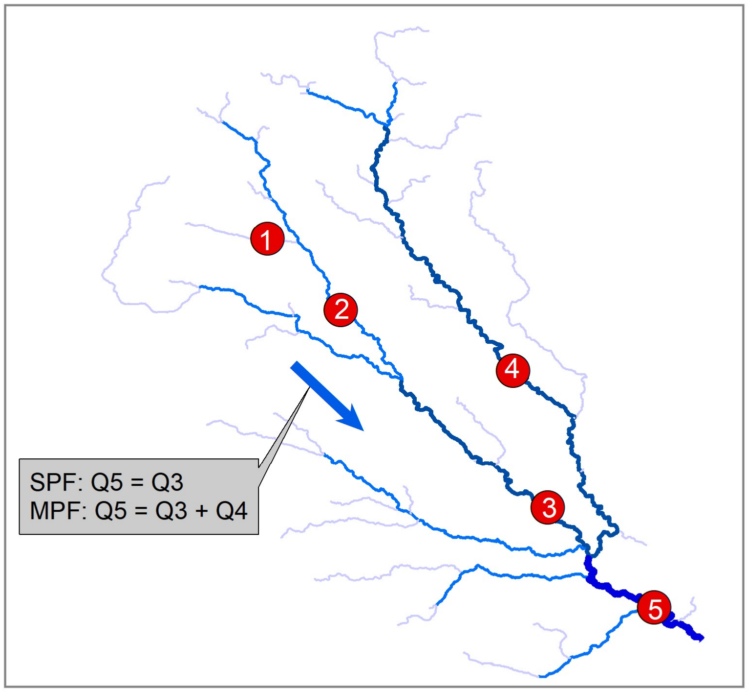

Figure 4: Demonstration of experimental setup for the spatial persistence forecasting. Arrows show the direction for streamflow forecast in a flow connected river network. The schematic illustrates Single Point Forecasts and Multiple Point Forecasts. In the upstream direction, only SPF is possible.

Similarly, if we consider the upstream direction, one could make forecast at Gauge 1 using data from Gauge 5. Full analogy follows for all other pairing. In summary, for this example basin, we have seven pairs to consider in each direction. These pairs of the forecast and evaluation points will allow us to evaluate persistence forecasting in space as function of spatial scale. We will refer to these forecasting scenarios as single point forecasts (SPF).

The above discussion is valid for the forecast lead time equal to zero, i.e. streamflow observed now at location i is the same as the forecast at location j. We can extend the same construction to other lead times $\Delta t = 0, \ldots, \Delta t_{max}$. The forecast at location i for the lead time Δt

is the same as the flow now at the stream gauge *j*. In terms of the spatial arrangement between the forecast and evaluation points, nothing changes i.e.

$$Q_i(i, t_0 + \Delta t) = Q_j(j, t_0) \qquad \text{for} \quad \Delta t = 0, \dots, \Delta t_{max} \tag{2}$$

Now, let us expand the forecasting setup. Following the same example in Figure 4, it is clear that if one adds data from Gauge 3 and Gauge 4, the forecasts at Gauge 5 should improve, perhaps considerably. However, forecasting at Gauge 3 does not benefit from considering data from Gauge 1 and Gauge 2 together. For the example basin in Figure 4, there is only one multi point forecast (MPF) combination possible. However, for larger basins with rich river networks and dense system of gauges, there could be several to consider. Note, that in the upstream direction there are no MPFs possible. As earlier, considering different forecast lead times does not change the spatial arrangements between the forecast and evaluation points. Mathematically,

$$Q_i(i, t_0 + \Delta t) = \sum_{j=1}^{n} Q_j(j, t_0) \qquad \text{for} \quad \Delta t = 0, \dots, \Delta t_{max} \tag{3}$$

where *n* is the number of stream gauges that constitute an MPF.

Next, let us add a simple travel time consideration. Since it is well established that in most places the direction of water movement is easy to predict, we can add water flow velocity and river distance information to the persistence forecasts. The simplest choice for water flow velocity is a constant value. Mathematically, for a SPF

$$Q_i(i, t_0 + \Delta t) = Q_j\left(j, t_0 + \frac{d}{v}\right) \qquad \text{for} \quad \Delta t = 0, \dots, \Delta t_{max} \tag{4}$$

where *v* is a constant water flow velocity and *d* is the distance along the river network between locations *i* and *j*. Recognizing that water flow velocity is a non-linear function of discharge and other channel parameters (e.g., Ghimire et al. 2018; Mantilla 2007; Ayalew et al. 2014a,b), we limit our consideration to the constant velocity case to keep the setup simple.

As river networks display a wide range of shapes (topologies) that may affect hydrologic response (e.g., Ayalew and Krajewski 2017; Perez et al. 2018), it is interesting to check their effect on the forecasting skill of persistence (also see Ghimire and Krajewski 2019). To accomplish this, we consider the width function of the basins defined by the existing USGS gauges. Basin width function can be interpreted as a distribution of stream segments (distances) with respect to the outlet of the basin (e.g., Rodriguez-Iturbe and Rinaldo 1997). If one assumes constant water flow velocity, the distribution of distances is equivalent to the distribution of travel times. Consider the width functions of the Cedar River at Cedar Rapids and the Iowa River at Marengo (Figure 5). We can see that the maximum of the travel time for the Cedar River basin is located some 200 km away from the outlet at Cedar Rapids. For the Iowa River, the maximum is just upstream from Marengo, some 50 km from the outlet. Note that in Figure 5, the travel time between the forecast point and the evaluation point (basin outlet) is the same. In addition, the base of the width function is also similar (~350 km). However, the forecast point of the Cedar basin entails not only the peak of the width function, but also a larger upstream drainage area.

These differences in the shape of the basin may influence the collective performance of persistence-based forecasting, especially for the MPF scenarios. However, since considering the entire distribution is inconvenient, we examined several simple geometric descriptors of width functions (e.g., Perez et al. 2018) such as base of the width function(D_B), centroid of the width function ($\bar{d}$), mass of the width function (M), and distance to the peak of the width function (d_{Wmax}). Perez et al. (2018) and Ghimire and Krajewski (2019) pointed out that normalized form of D_B explains the most variability in performance metrics. Whether the forecasting scenario is SPF or MPF, we define relationships between forecast and evaluation points in pairs. The ratio of D_B explains the same variability as normalized $D_{B,}$ which we discuss later in detail.

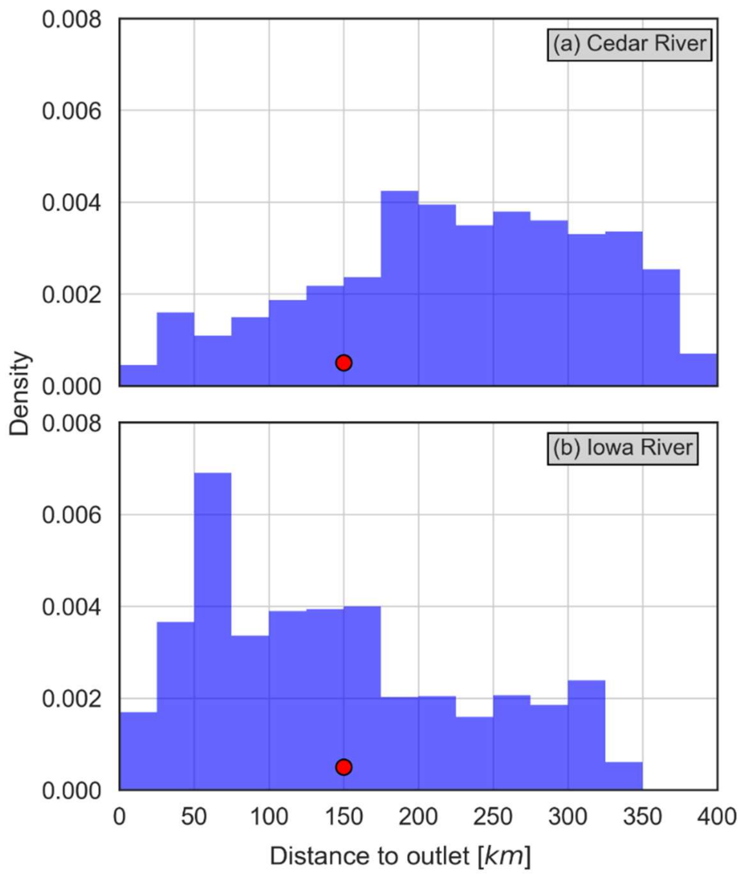

Figure 5: Width function depicting normalized histogram (density) of distance of each link in the river network to the basin outlet. (a) Width function for the Cedar River at Cedar Rapids and (b) Width function for the Iowa River at Marengo. The red dot represents the Single Forecast Point with the evaluation site being the basin outlet.

3.2 Evaluation Metrics

We evaluated persistence derived streamflow forecasts in space and time against USGS streamflow observations at evaluation points. We computed forecast skills in terms of several statistical metrics. The major performance skill measures we report here are Kling-Gupta Efficiency (*KGE*), non-parametric *KGE*, mean absolute error (*MAE*), normalized *MAE* (*nMAE*), hydrograph timing (T_H), timing of the annual peak (T_P), and percent peak difference (*PPD*). The *KGE* metric has three components: Pearson's correlation (r), variance ratio (α) and mean ratio (β) (see Gupta et al. 2009). The ideal value of *KGE* is equal to 1.

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \quad (5)$$

where $\alpha = \frac{\sigma_s}{\sigma_o}$, and $\beta = \frac{\mu_s}{\mu_o}$, σ_s is the variance of forecasts, σ_o is the variance of observations, μ_s is the mean of forecasts, and μ_o is the mean of observations. As an example, if the variance ratio α and the mean ration β are close to one, the *KGE* is dominated by the correlation coefficient r. The frontier of KGE for good/bad forecasts relative to the mean of the observations could be considered as -0.41 (Knoben et al. 2019). A modification was proposed by Pool et al. (2018) to get away with the implicit assumption of linearity of data, normality of data and the absence of outliers in the formulation of *KGE* by Gupta et al. (2009). It comprises non-parametric components, i.e. Spearman rank correlation and normalized flow duration curve while the mean ration (multiplicative bias) component stays the same. The modified measure is given by

$$KGE_{NP} = 1 - \sqrt{(r_s-1)^2 + (\alpha_{NP}-1)^2 + (\beta-1)^2} \quad (6)$$

where r_s is the Spearman rank correlation and α_{NP} is the normalized flow duration curve. For further details, readers are referred to Pool et al. (2018).

Likewise, we report *MAE* for stage forecasts associated with streamflow forecasts derived using rating curves and is given by,

$$MAE = \frac{1}{N}|h_{\boldsymbol{s}} - h_o| \tag{7}$$

where h_s and h_o are forecast and observed stage, respectively, and N is the total number of time-series nodes. Also, it is interesting for streamflow forecasters to consider stage forecast error in terms of actual values depicted in terms of median (Δh_{median}) i.e.

$$\Delta h_{median} = median(h_s - h_{\boldsymbol{o}}) \tag{8}$$

The normalized *MAE* (*nMAE*) for streamflow forecasts is computed similar to (7) but normalized by upstream drainage area. Hydrograph timing (T_H) is associated with the cross-correlation between the entire observed and forecast time-series. We computed it as the number of hours a forecast time-series needs to be shifted so that the cross-correlation coefficient is maximized. A positive value indicates delay in the timing of hydrograph while a negative value indicates early timing of forecast relative to the observed hydrograph. Timing of the peak T_P is simply the difference in time at which the peak of forecast occurs relative to the peak of observations.

Next, we computed PPD as,

$$PPD = \frac{peak_{\boldsymbol{sim}} - peak_{\boldsymbol{obs}}}{peak_{\boldsymbol{obs}}} \times 100 \tag{9}$$

where, $peak_{sim}$ and $peak_{obs}$ are the peaks of forecasts and observation respectively. We limit the consideration only to annual peaks (based on the observations).

4 Results

4.1 Purely Temporal Persistence Forecasts

While Ghimire and Krajewski (2019) provide a comprehensive characterization of three persistence schemes, here we present their key result as it will be useful for our discussion of the spatio-temporal persistence. In Figure 3, we show the KGE statistic for different forecast lead times up to three days ahead organized as a function of basin drainage area. It is interesting to note that for one-day ahead forecast for basins of the size of about 1000 km^2, this simple persistence method skill is about 0.8, equivalent to a very good model (for a discussion of what is very good, see the analysis in Knoben et al. (2019). As Ghimire and Krajewski (2019) show, this skill can be even somewhat increased by incorporating simple climatology information.

4.2 Purely Spatial Persistence Forecasts (PSPF): Downstream Direction

In this scenario, the point of interest, e.g. a community, is downstream from the forecast point. In this study, we assume we have a stream gauge at the forecast location and can use the data to evaluate the quality of the forecast that used data from a point upstream. For example, consider the Cedar River basin and take streamflow data at Waverly, Iowa (Gauge '2' in Figure 6) and use them as the forecast for Cedar Rapids (Gauge '5' in Figure 6). The performance is not very good: the KGE is -0.1 with components of correlation of 0.49, the mean ratio of 0.25, and the variance ratio of 0.38. However, if we now use data at Waterloo, Iowa (Gauge '3' in Figure 6) and use them to make forecasts for Cedar Rapids, the skill increases dramatically: the KGE goes up to 0.7. Now, we summarize the performance results for all possible combinations of the Single Point Forecasts (SPFs) in Figure 6 (see red dots) organizing them as function of the upstream drainage area ratio, hereafter referred to as monitored area fraction, A_r (e.g. gauge at

Waterloo monitors drainage of about 0.8 of the basin area at Cedar Rapids). It is evident that using data from stations that capture over 70% of the basin of interest, i.e. $A_r = 0.7$, persistence-based forecasting performs very well (KGE > 0.5). The results shown here for the Cedar River basin, indicate strong relationship of the KGE measure and the monitored area fraction.

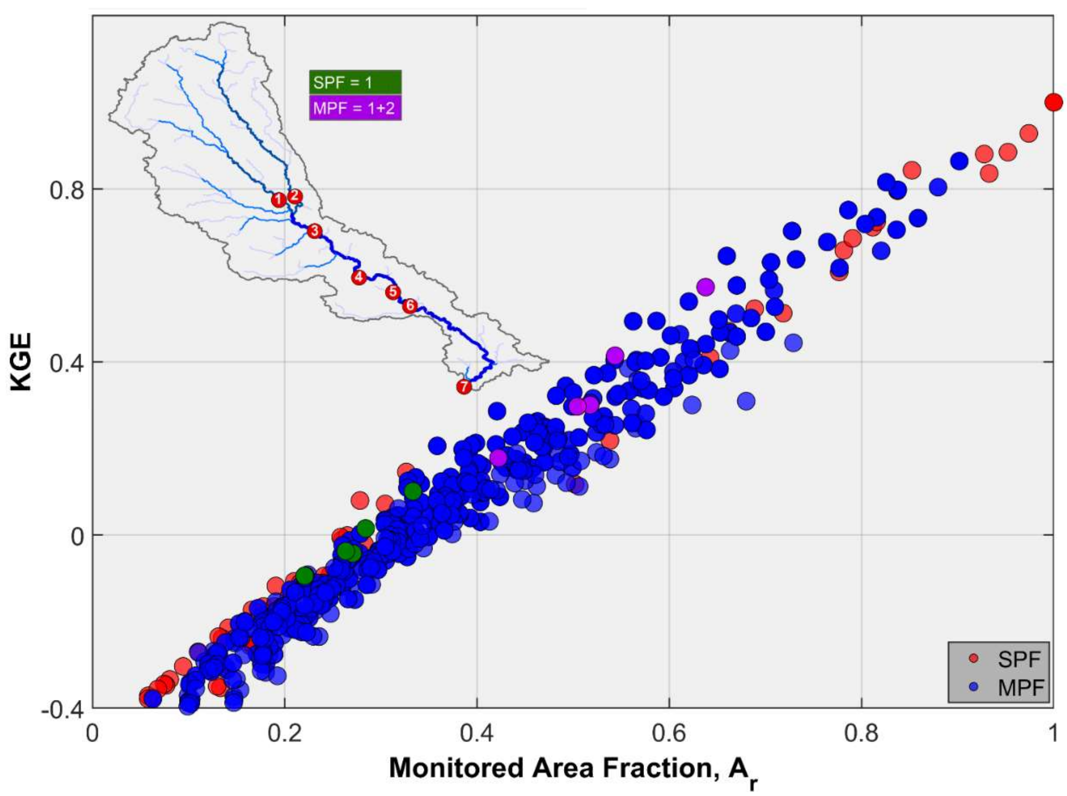

Figure 6: Relationship of the KGE metric with upstream drainage area ratio, A_r for the Cedar River basin for the year 2016. Every dot in this figure implicitly represents a pair of forecast and evaluation points. Red dots show SPF performance while blue dots represent MPF forecast performance. As a demonstration of improvement in performance of MPF based forecast from SPF based forecast consider the green and magenta points. The green dots are evaluation points 3 to 7 using SPF at forecast point 1. The magenta dots are the same evaluation points upon addition of forecast point 2 to 1 showing a dramatic increase in the KGE metric.

The performance increases if we combine data coming from different upstream branches of the river network (MPFs). A simple demonstration of how MPFs contribute to performance increase is shown by blue dots in Figure 6. Here, green dots with forecast issued from Gauge 1 transforms to magenta dots upon addition of forecast from Gauge 2. The benefit does not apply to all stations as there are many locations that have only one station upstream. However, it is interesting to note that the performance based on MPFs also follows the strong relationship with A_r.

The above results are based on data for 2016. A natural question arises whether the same performance stays in other years. To illustrate the year-to-year variability, we have explored data for the 17-year period, 2002-2018. In Figure 7, we show violin plots depicting distribution of KGE over this period for eight major basins (refer Figure 1 and Table 1) in Iowa. It is clear that both the median KGE (red line) and overall distribution display the same signature as in Figure 6. A linear-regression based predictive model with goodness-of-fit measures are presented in Table 1. Clearly, R^2 and $RMSE$ measures corroborates the strong relationship of KGE with A_r. Note that the intercept, α and slope parameter, β of the predictive models show very small variability across basins except the Des Moines River. We speculate that this difference could be attributed to the distinct hydrogeology of this basin much of which is in the Des Moines Lobe characterized with pothole connectivity augmented by an extensive tile drainage network.

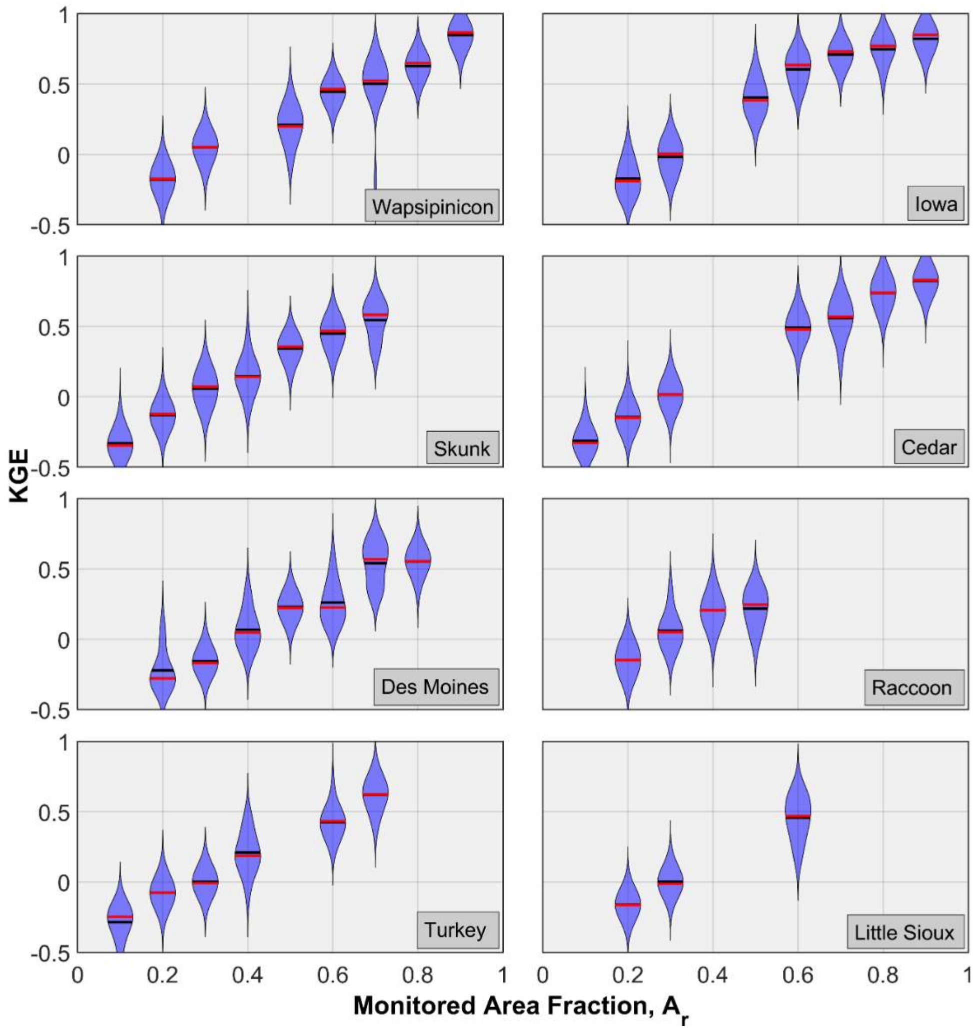

Figure 7: Violin plots showing inter-annual variability of KGE metric with upstream drainage across basins in Iowa. Each violin plot shows distribution of KGE over the period of 2002–2018. The red line and black line show median and mean, respectively, of the KGE over the years.

In Figure 8, we show relationship of KGE and its decomposition components with A_r for the pooled data from all basins for the year 2016. As discussed above, KGE statistic is explained strongly by A_r. Among its components, both bias and variance ratio show similar signature with A_r while correlation does not show as strong relationship. It is interesting to explore if river network geometry in terms of width function can explain the variability in KGE and its components. It is clear from Figure 9 that a width function descriptor of D_B ratio does not explain the variability of KGE and its components as strongly as A_r. However, note that correlation is explained almost equally well by the D_B ratio. If this result is interpreted in conjunction with Figure 5, larger variability in correlation with A_r can also be explained by D_B ratio. This, in turn, has particular implication for T_H and T_P, which we discuss later.

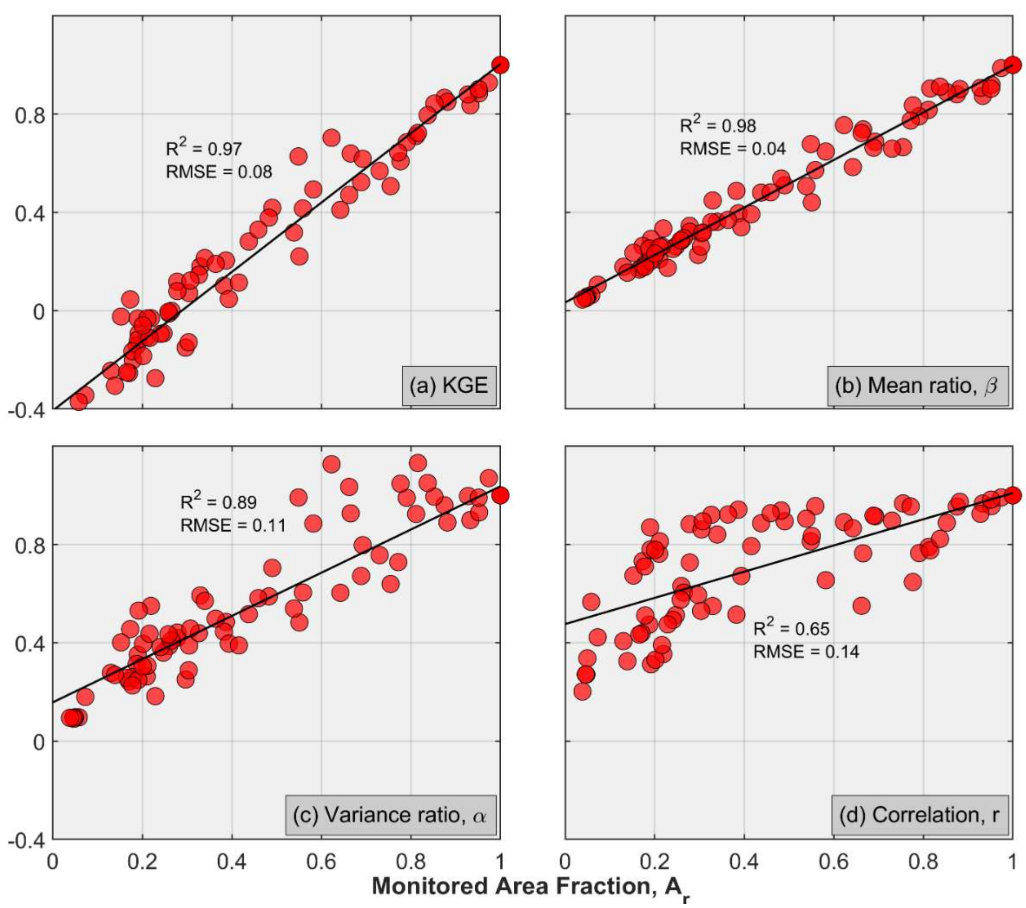

Figure 8: Relationships of KGE metric decomposition components with upstream drainage area ratio for pooled data from all basins in Iowa for the year 2016. KGE (a) shows strong linear dependence with A_r. The mean ratio (b), variance ratio (c) and correlation coefficient (d) show decreasing degree of dependence on A_r.

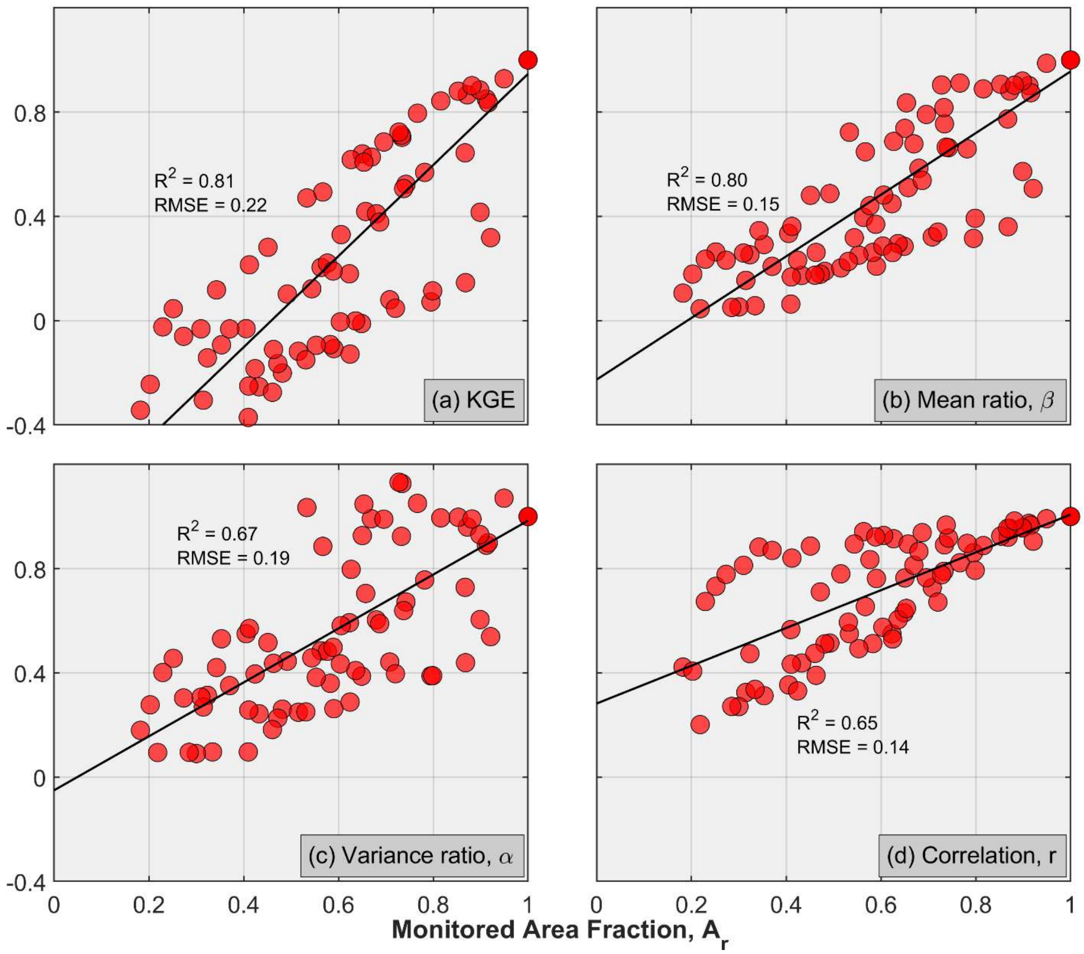

Figure 9: Relationships of KGE metric decomposition components with base of the width function, D_B ratio, for pooled data from all basins in Iowa for the year 2016. The dependence of KGE and its components with D_B ratio is not as strong as with A_r (see Figure 8). Note that, the correlation is explained similarly by both D_B ratio and A_r.

While KGE may not be the ultimate performance criterion for the operational forecasters, its regular behavior with the monitored area fraction provides high level of confidence in the streamflow forecasting. However, for many applications, including flood forecasting, the variable of most interest is river stage. In Figure 10, we evaluate our persistence forecasting in terms of stage errors. We use the rating curves provided by the USGS to convert the propagated discharge to stage. As the relationship between discharge and stage is influenced by the local hydraulic properties of the river channel, and because stage is measured with respect to an arbitrary datum, we expressed the results in terms of stage error in units of length [m]. In Figure 10, we show the distribution of median of the stage errors obtained over a period of seventeen years across eight basins in Iowa. The year-to-year variability is high, in particular, for cases of small A_r. Somewhat surprisingly, the median stage forecast errors are small for A_r larger than 0.7. For $A_r > 0.7$, median stage forecast errors are within 30 cm (one foot) for all basins indicating the potential of PSPF, which model-based forecasts should improve upon.

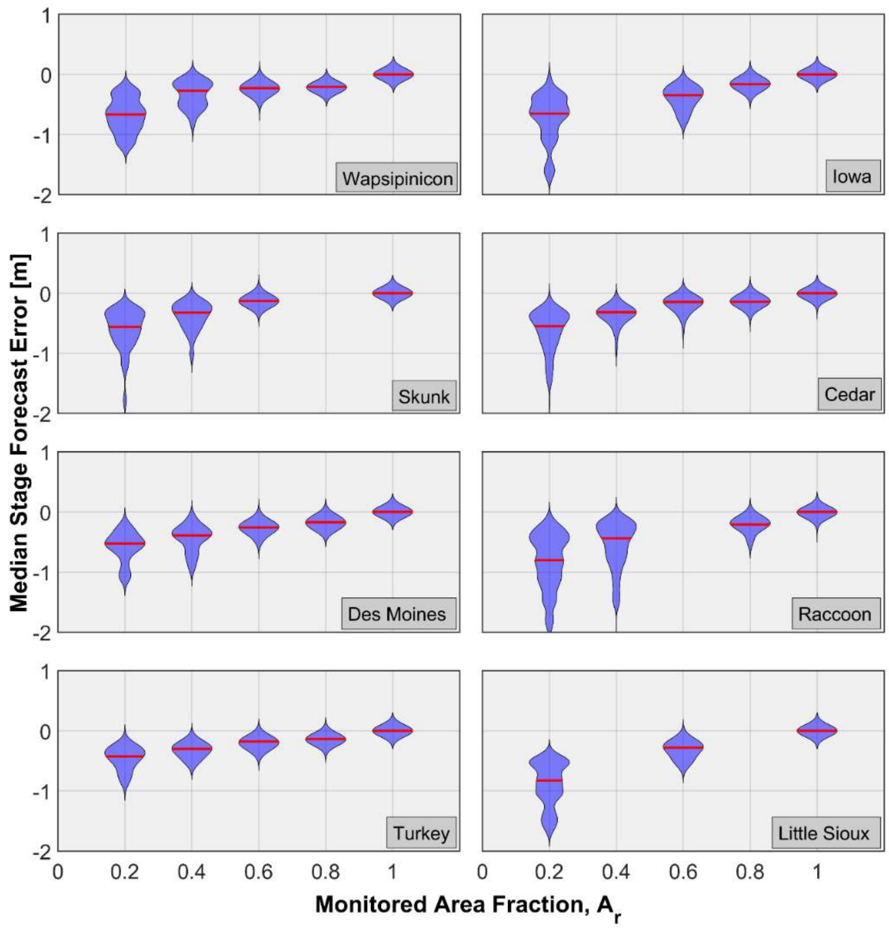

Figure 10: Violin plots showing inter-annual variability of median stage forecast error with monitored area fraction, A_r across basins in Iowa. Each violin plot shows distribution of stage forecast error over the period of 2002–2018. The red line shows the median of stage errors over the years.

In addition to the skill measures discussed above, operational forecasters are interested in other statistics while making streamflow forecasting decisions. In Figure 11, we show variability of *nMAE*, *PPD*, T_H and T_P over the seventeen-year period with A_r for the Cedar River basin. The variability and median values of *nMAE* and *PPD* show similar signature as KGE with A_r. These relationships of T_H and T_P with A_r are similar but weaker. This weaker relationship is partly explained by the variability of correlation explained by the D_B ratio (also see Figure 9).

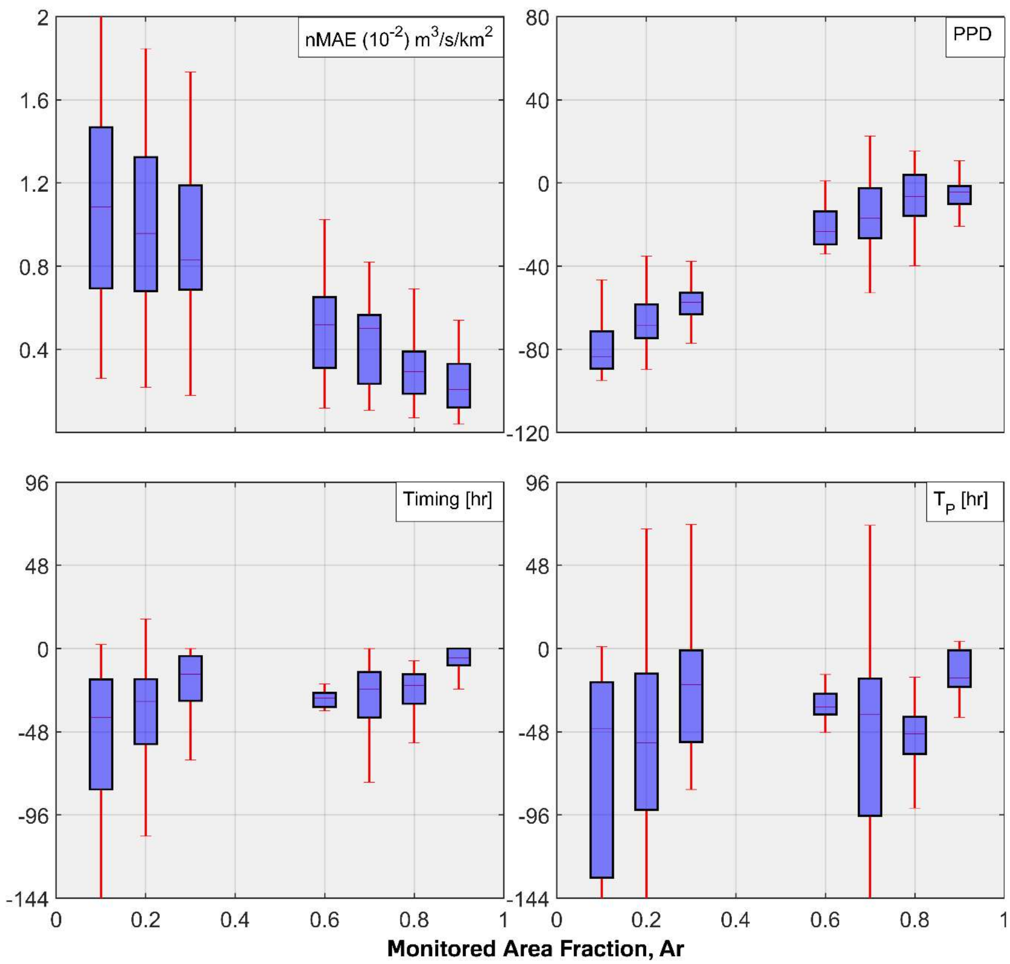

Figure 11: Box plots showing inter-annual variability of forecast skill measures normalized MAE (*nMAE*), percent peak difference (*PPD*), hydrograph timing (T_H) and timing of the peaks (T_P) with monitored area fraction, A_r for the Cedar River basin. The *nMAE* and *PPD* show strong dependence on A_r while the signature of T_H and T_P with A_r is not as strong.

4.3 Purely Spatial Persistence Forecasts (PSPF): Upstream Direction

In this scenario, the point of interest, e.g. a community or a water intake or a waste water discharge, is upstream from the forecast point. The results are similar but, surprisingly, worse than in the downstream direction, at least in terms of the KGE statistic (Figure 12). This lack of symmetry between the downstream and upstream direction can be easily explained by the construction of the KGE metric. While the correlation coefficient between the predicted and observed series is immune to the upstream/downstream translation, the other terms are ratios. In the downstream direction the predicted values are systematically lower than the observed (by construction of the forecasting scheme), the mean ratio is always less than one. For the upstream direction, the ratio has the inverse value. The variance ratio behaves similarly. As a result, the KGE values are generally lower for the upstream case and not better than the mean of the observations, especially for the A_r below 0.5. Note that we kept the ratio constraint by the larger scale gauge, although it is inconsistent with its original definition (forecasted vs. monitored area).

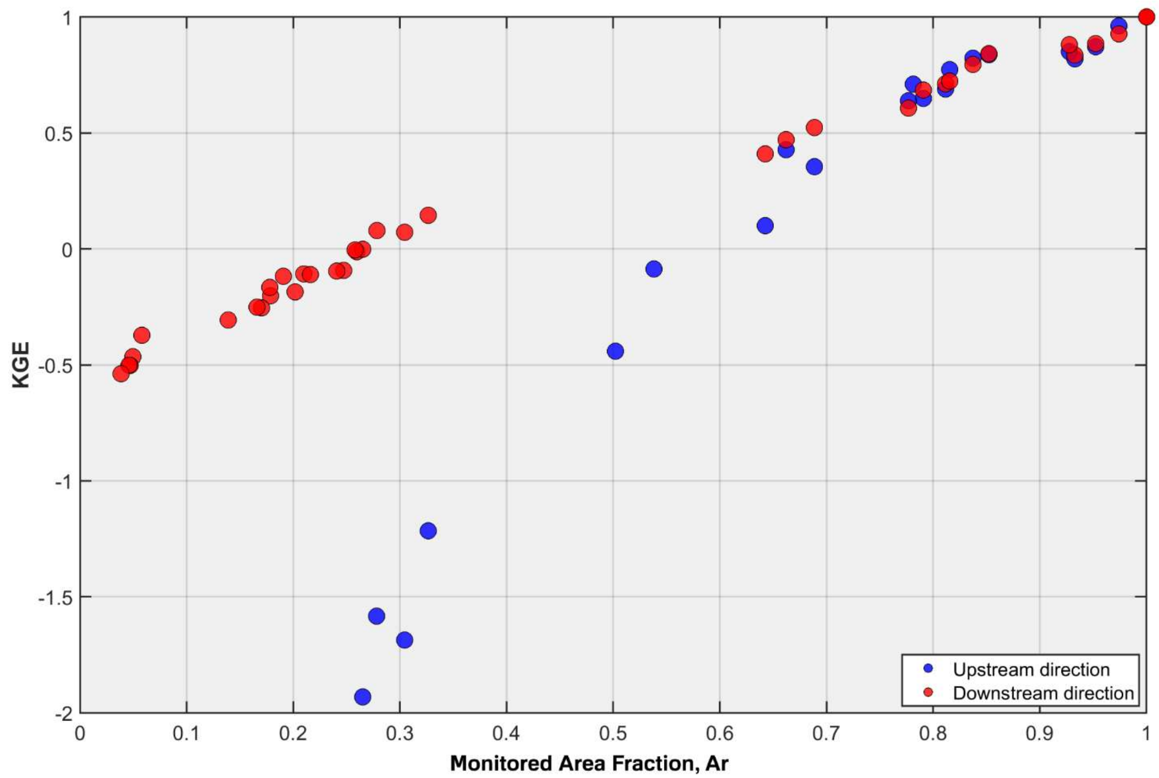

Figure 12: Effect of forecast direction in the river network on the relationship between KGE and A_r. With upstream direction of persistence forecast, the persistence forecasts lose skill faster with A_r.

4.4 Spatio-temporal Persistence Forecasts

Here, we consider forecasting scenarios with lead-time, $\Delta t > 0$ to the purely spatial persistence forecasts (PSPF) (see equation 2 and 3). In the same framework discussed in 4.3, we investigate the two-dimensional variability of performance metrics, i.e. with monitored area fraction, A_r and the lead-time, Δt. In Figure 13, we show contour plots depicting the spatio-temporal persistence forecast skill in terms of non-parametric *KGE* for the Cedar River basin. Since these contour plots are empirically produced based on a limited sample size, lack of smoothness of the contours is expected. As a demonstration, we show here three basin scales in top to bottom panels (Figures 13a-13c) which represent the SPF points used to issue forecast at locations downstream. It is interesting to note that, for all scales at $\Delta t = 3$ days and $A_r = 0.6$,

KGE_{NP} is larger than 0.5. Following from our discussion in 4.3, one can fit regression models with A_r and Δt as explanatory variables giving rise to a contour plot that works for all basin scales and can be used by forecasters for model evaluation. We have developed similar results (not shown) in terms of the KGE statistic.

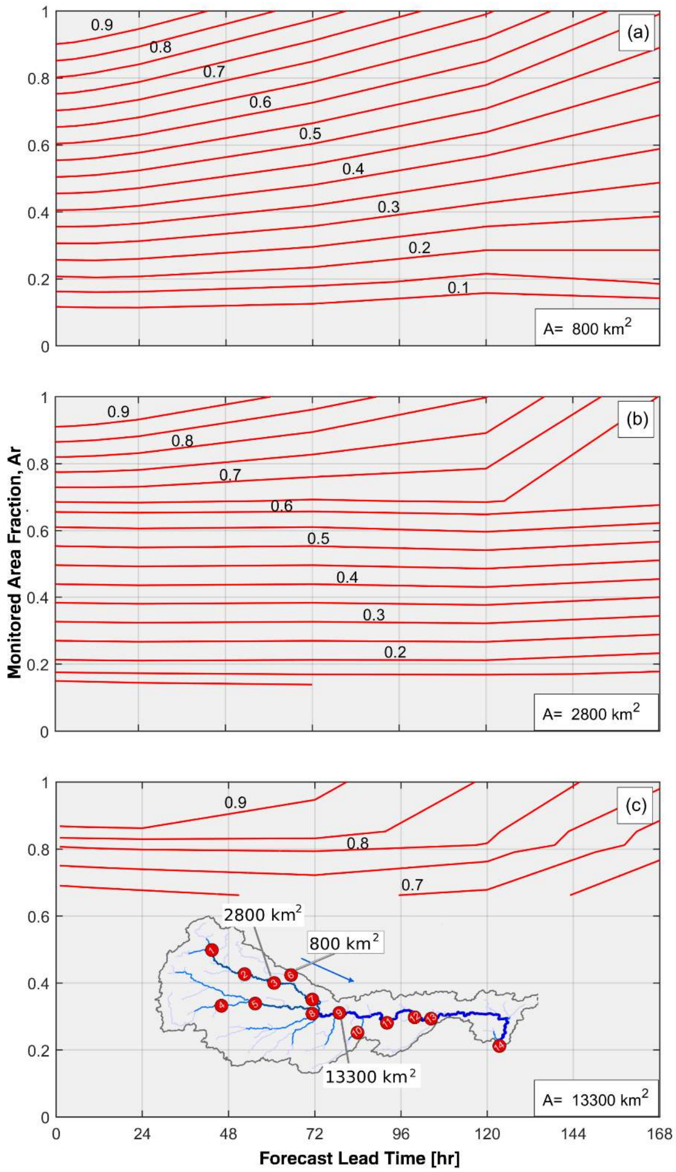

Figure 13: Contour plots depicting the spatio-temporal persistence forecast skill in terms of non-parametric *KGE* (see contour labels) for the Cedar River basin. Three basin scales from top to bottom panels (a-c) represent the SPF points used to issue forecast at locations downstream. For example, for all scales at lead time of three days and A_r of 0.6, KGE_{NP} is larger than 0.5.

4.5 Adding Velocity Effect

Staying true to the principle of extreme simplicity of this investigation, we would like to explore the effect of downstream water transport to improve the forecasting skill. Intuitively, it seems reasonable to expect that since water flows downstream, we should be able to improve the forecasting by applying the proper time offset. Since the distance along the river is known, the other piece of information we need is the water flow velocity. Assuming a constant velocity, the question becomes what value should we use? We explored the empirical distribution of streamflow velocity based on USGS gauges in Iowa. We investigated over 100,000 data points for the velocity distribution. The observed velocity in Iowa's streams and rivers ranges from 0.2 to over 2 m/s. We investigated the velocities that correspond to each 10^{th} quantile of the distribution. It turns out that the best choice, in terms of maximizing the KGE, is a velocity significantly higher than the median or the mean but not the very highest velocity.

In Figure 14, we show the dependence of the KGE and its components on the chosen constant velocity (each 10^{th} quantile of the distribution) for the Cedar River basin. We show the components of the KGE because the metric itself shows little sensitivity with respect to velocity. In the Cedar River basin, we have a large number of USGS gauges, some added following the 2008 extreme flood (Smith et al. 2013). This results in a large number of pairs of gauges and thus a wide range of the A_r. In Figure 14, we show (in red dots) all possible SPF pairs. Those with small values of A_r result in poor performance in terms of KGE. To facilitate the analysis, we selected two pairs: one with a small value of A_r (~0.04) shown in green, and second one with high (but not highest) A_r (~0.81) shown in blue. The corresponding color horizontal lines show the KGE for the purely spatial persistence forecasting (PSPF). Recall that "purely spatial" refers to the case of Δt = 0.

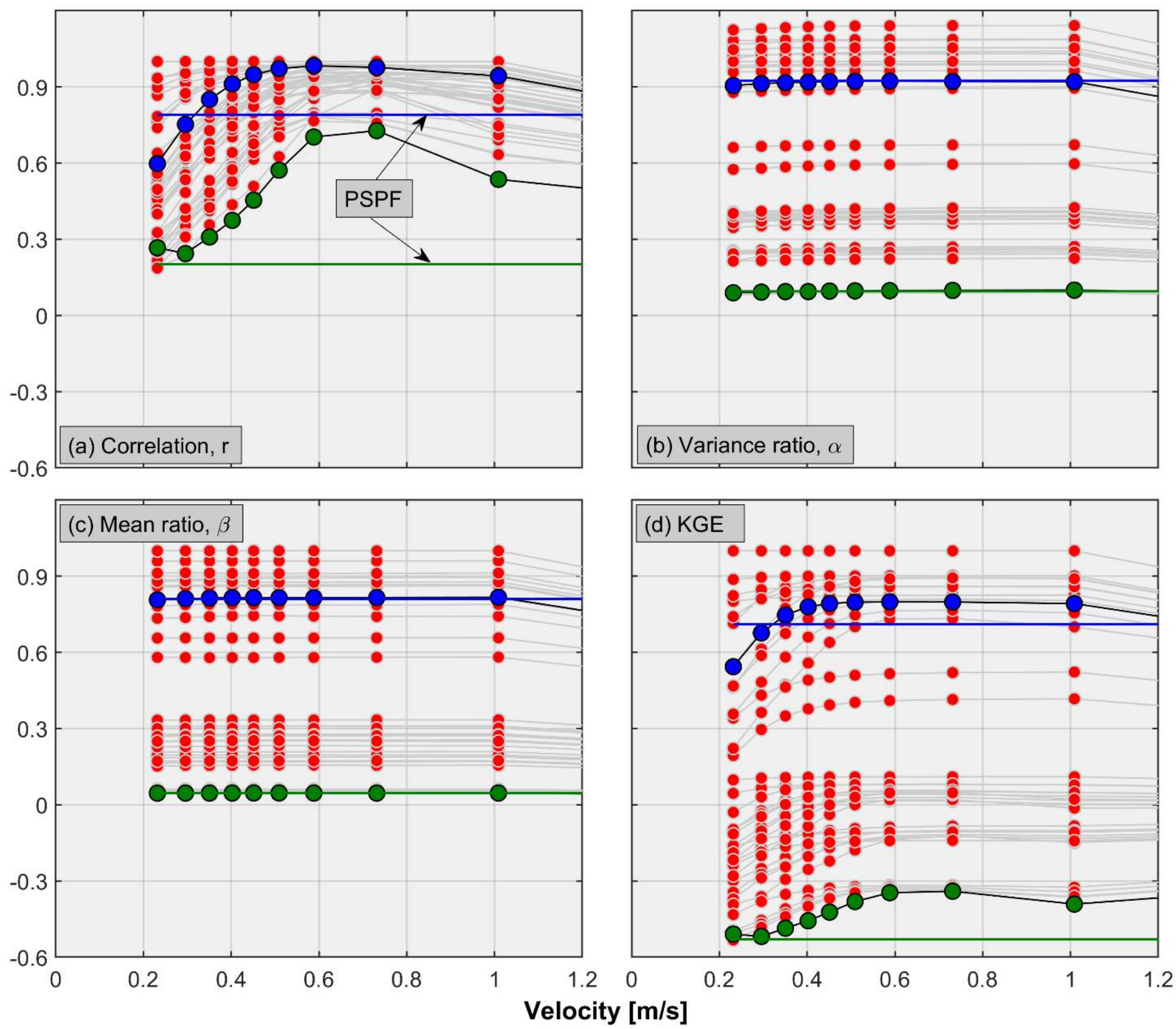

Figure 14: Effect of constant velocity routing on Pure SPF forecasts skill for the Cedar River basin. The panels show effect of constant velocity on each component of KGE. The velocity axis shows the constant velocity used for routing PSPF forecasts while vertical axis shows KGE components conditional on a given constant velocity. Each grey line connecting the red dots represents a value of A_r. Green and blue horizontal lines show the corresponding PSPF forecast skills for smaller (A_r=0.04) and larger (A_r=0.81) area ratios. The green and blue lines with dots show forecast skills for respective area ratios with the addition of constant routing velocity.

Several aspects of the problem become clear from the plots. First, the sensitivity of the KGE is dominated by the mean and the variance ratios. These two components show no sensitivity to the velocity as the same hydrographs are passed between the two gauges, just some faster and some slower. The sensitivity is in the correlation coefficient. Increasing the velocity improves the correlation but only to a point. Further increase of the velocity will compromise the correlation. Also, the sensitivity is higher for the small A_r. The gauge pairs characterized by high A_r, show negligible impact of the velocity. This implies that the forecasting performance is dominated by the A_r. However, for some values of A_r, including velocity from the mid-range of the observed values, improves the forecasting performance, again, mostly in terms of the correlation.

Deeming the results found for the spatio-temporal persistence forecasts in 4.4, we expect that similar results will hold upon addition of constant velocity. The effect of constant velocity on other statistics, such as *nMAE*, *PPD*, T_H and T_P, results directly from the above results. The values of T_H and T_P depend mainly on the correlation.

4.6 Closing Comments

We hope that readers find the results of this simply constructed study interesting and useful. We demonstrate that data-only based real-time streamflow forecasting has a high level of skill for well-instrumented basins. Our results could be useful in several ways. We establish a performance reference level that more elaborate approaches, such as distributed hydrologic or AI based models should be expected to improve on. This reference is location-dependent based on the configuration of the stream gauges. One simple demonstration of this is in Figure 15. Here, we show the KGE along the river network in the Cedar River basin obtained by using the linear relationship from Figure 6. The skill changes as the distance from the gauges increases (as the monitored area fraction changes as well). For each segment of the river network (or a link in the HLM nomenclature), we find the maximum skill for upstream or downstream (SPF or MPF) options. Adding IFC sensors would expand the useful skill of the persistence (assuming "perfect" rating curves).

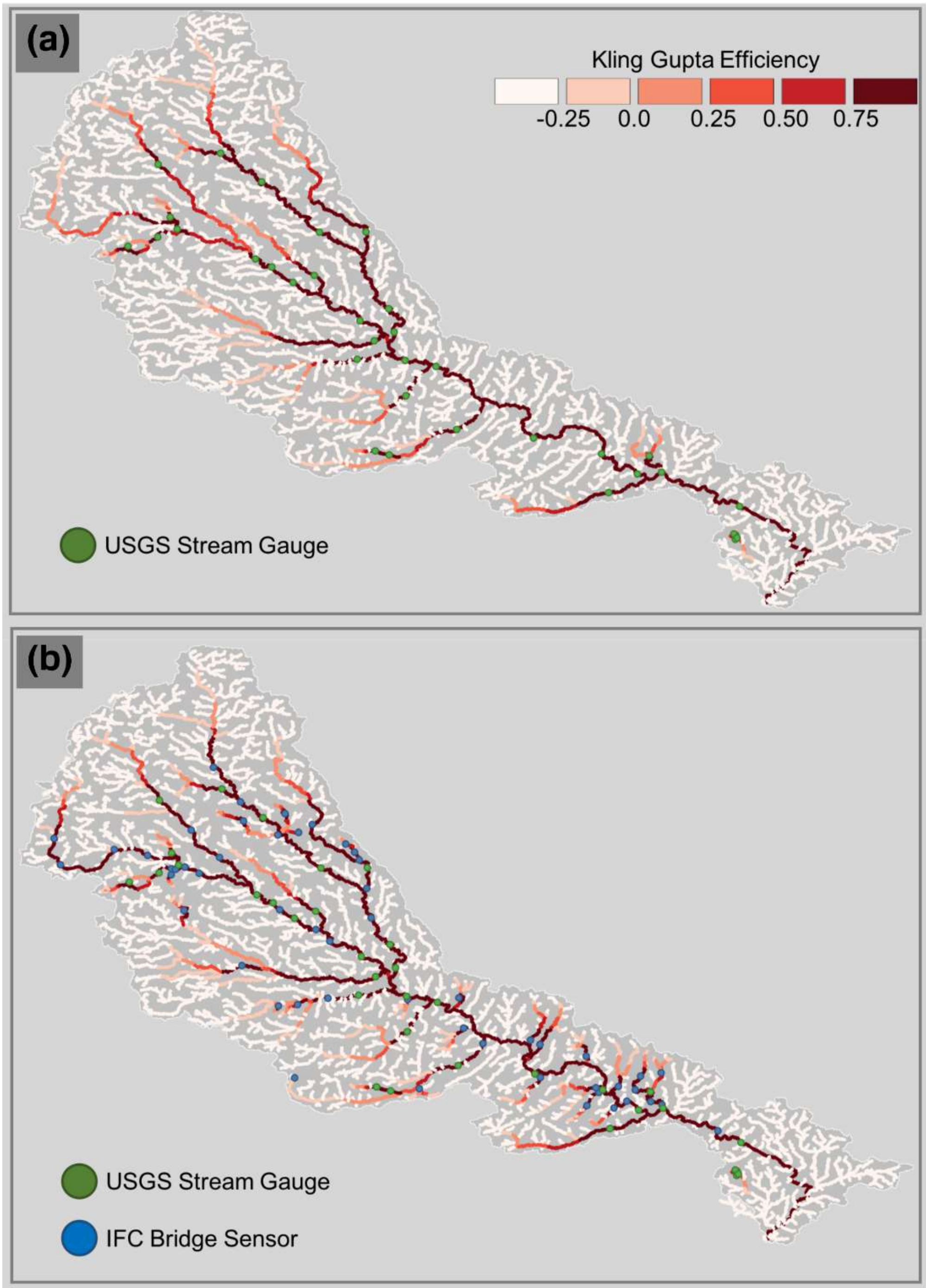

Figure 15: Demonstration of Pure SPF skill in terms of KGE in the river network using the USGS only (a), and USGS and IFC bridge sensors (b) as forecasting locations for each link. Addition of IFC bridge sensor in the network shows the potential for enhanced coverage of forecasts and improved the corresponding skill in the river network. The plot assumes perfect rating curves for both USGS and IFC gauge location.

The above results could serve as a basis for streamflow gauging network design as they clearly guide placing additional stream gauges to increase the A_r. They can also help with the assessment of the quality of synthetic rating curves developed for stage-only gauges (Kruger et al. 2016).

We foresee several extensions of this work in the direction of systematic assessment of the forecasting skill of hydraulic and hydrologic models. For example, we have demonstrated that it is difficult to find a constant velocity that improves upon a purely spatial approach. Would hydraulic models of open channel flow lead to better results? They require additional information in the form of channel geometric and hydraulic (roughness) properties. While the cross sections and slopes are straightforward to obtain, parametric description of roughness is subject to considerable uncertainty.

Finally, while comparing the PSPF skill to that of highly distributed (physically based) hydrologic models is outside the scope of this paper, we include an example in Figure 16. It shows the KGE of an open loop HLM model as well as the same model with a simple streamflow data assimilation scheme. The scheme is a replacement of the initial conditions (the HML is a system of ordinary differential equations) and thus directly comparable to the PSPF. The scheme does not modify system states or parameters upstream of the streamflow gauge. The results show better forecast skill of the HLM in terms of KGE for smaller basin scales and for longer lead times. For example, from lead-time of one day to five days the forecast skill of the HLM model starts bettering the performance of PSPF. With the addition of the simple data assimilation framework, the HLM forecast skill is much enhanced over PSPF, as expected. The HLM simulation assumes “perfect” rainfall forecast (i.e. limited only to the observational and estimation errors). In reality, the rainfall forecasts are subject to considerable uncertainty, especially for the longer lead times (e.g., Seo et al. 2018).

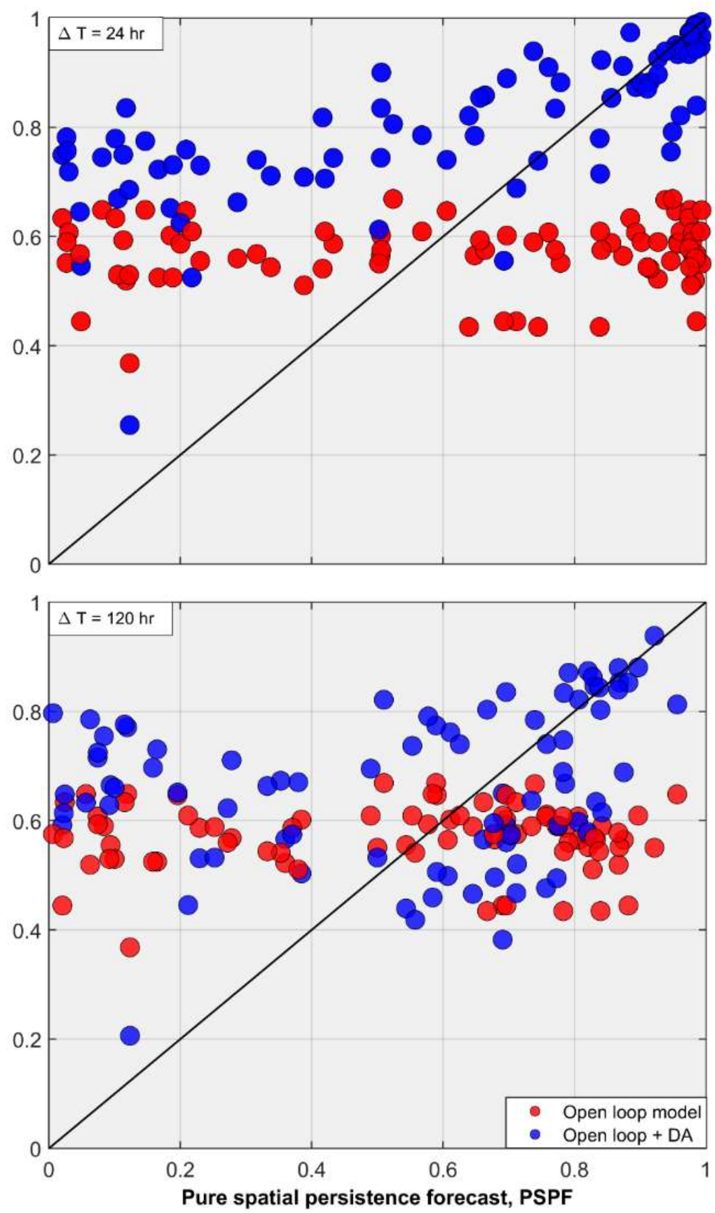

Figure 16: One-to-one comparison of a hydrologic model forecast skill with the PSPF. Red dots show the comparison of KGE of the open loop HLM model with PSPF while blue dots show the comparison of KGE of simple data assimilated open loop model with PSPF. This scatterplot shows a potential of enhanced forecast skill upon adding modeling complexity to PSPF.

Acknowledgements

This study was funded by the Iowa Flood Center of the University of Iowa. The first author also acknowledges partial support from the Rose & Joseph Summers endowment. The authors are grateful to many colleagues at the IFC who facilitated the study by providing observational and computational support.

References

Ayalew, T. B., and W. F. Krajewski, 2017: Effect of river network geometry on flood frequency : A tale of two watersheds in Iowa. *J. Hydrol. Eng.*, **22**, 1–7, https://doi.org/10.1093/nar/gkn1071.

Ayalew, T. B., W. F. Krajewski, and R. Mantilla, 2014a: Connecting the power-law scaling structure of peak-discharges to spatially variable rainfall and catchment physical properties. *Adv. Water Resour.*, **71**, 32–43, https://doi.org/10.1016/j.advwatres.2014.05.009.

——, ——, ——, and S. J. Small, 2014b: Exploring the effects of hillslope-channel link dynamics and excess rainfall properties on the scaling structure of peak-discharge. *Adv. Water Resour.*, **64**, 9–20, https://doi.org/10.1016/j.advwatres.2013.11.010.

Betterle, A., M. Schirmer, and G. Botter, 2017: Characterizing the spatial correlation of daily streamflows. *Water Resour. Res.*, **53**, 1646–1663, https://doi.org/10.1002/2016WR019195.

Botter, G., S. Zanardo, A. Porporato, I. Rodriguez-Iturbe, and A. Rinaldo, 2008: Ecohydrological model of flow duration curves and annual minima. *Water Resour. Res.*, **44**, 1–12, https://doi.org/10.1029/2008WR006814.

Castiglioni, S., A. Castellarin, A. Montanari, J. O. Skøien, G. Laaha, and G. Blöschl, 2011: Smooth regional estimation of low-flow indices: Physiographical space based interpolation and top-kriging. *Hydrol. Earth Syst. Sci.*, **15**, 715–727, https://doi.org/10.5194/hess-15-715-2011.

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 Resour. Res.*, **48**, 1–22, https://doi.org/10.1029/2012WR012138.

Demir, I., and W. F. Krajewski, 2013: Towards an integrated Flood Information System: Centralized data access, analysis, and visualization. *Environ. Model. Softw.*, **50**, 77–84, https://doi.org/10.1016/j.envsoft.2013.08.009.

Doulatyari, B., A. Betterle, S. Basso, B. Biswal, M. Schirmer, and G. Botter, 2015: Predicting streamflow distributions and flow duration curves from landscape and climate. *Adv. Water Resour.*, **83**, 285–298, https://doi.org/10.1016/j.advwatres.2015.06.013.

Dralle, D. N., N. J. Karst, and S. E. Thompson, 2016: Dry season streamflow persistence in seasonal climates. *Water Resour. Res.*, **52**, 90–107, https://doi.org/10.1002/2015WR017752.

ElSaadani, M., W. F. Krajewski, R. Goska, and M. B. Smith, 2018: An investigation of errors in distributed models' stream discharge prediction due to channel routing. *J. Am. Water Resour. Assoc.*, **54**, 742–751, https://doi.org/10.1111/1752-1688.12627.

Ghimire, G. R., and W. F. Krajewski, 2019: Exploring persistence in stream flow forecasting (accepted for publication). *J. Am. Water Resour. Assoc.*,.

Ghimire, G. R., W. F. Krajewski, and R. Mantilla, 2018: A power law model for river flow velocity in Iowa basins. *J. Am. Water Resour. Assoc.*, **54**, https://doi.org/10.1111/1752-1688.12665.

Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez, 2009: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.*, **377**, 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003.

Knoben, W. J. M., J. E. Freer, and R. A. Woods, 2019: Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores. *Hydrol. Earth Syst. Sci. Discuss.*, 1–7, https://doi.org/10.5194/hess-2019-327.

Krajewski, W. F., and Coauthors, 2017: Real-time flood forecasting and information system for the state of Iowa. *Bull. Am. Meteorol. Soc.*, **98**, 539–554, https://doi.org/10.1175/BAMS-D-15-00243.1.

Kruger, A., W. F. Krajewski, J. J. Niemeier, D. L. Ceynar, and R. Goska, 2016: Bridge-mounted river stage sensors (BMRSS). *IEEE Access*, **4**, 8948–8966, https://doi.org/10.1109/ACCESS.2016.2631172.

Laaha, G., J. O. Skøien, and G. Blöschl, 2014: Spatial prediction on river networks: Comparison of top-kriging with regional regression. *Hydrol. Process.*, **28**, 315–324, https://doi.org/10.1002/hyp.9578.

Mallakpour, I., and G. Villarini, 2015: The changing nature of flooding across the central United States. *Nat. Clim. Chang.*, **5**, 250–254, https://doi.org/10.1038/nclimate2516.

Mantilla, R., 2007: Physical basis of statistical scaling in peak flows and stream flow hydrographs for topologic and spatially embedded random self-similar channel networks. University of Colorado at Boulder, .

——, and V. K. Gupta, 2005: A GIS numerical framework to study the process basis of scaling statistics in river networks. *IEEE Geosci. Remote Sens. Lett.*, **2**, 404–408, https://doi.org/10.1109/LGRS.2005.853571.

Müller, M. F., and S. E. Thompson, 2016: Comparing statistical and process-based flow duration curve models in ungauged basins and changing rain regimes. *Hydrol. Earth Syst. Sci.*, **20**, 669–683, https://doi.org/10.5194/hess-20-669-2016.

Müller, M. F., D. N. Dralle, and S. E. Thompson, 2014: Analytical model for flow duration curves in seasonally dry climates. *Water Resour. Res.*, **50**, 5510–5531, https://doi.org/10.1002/2014WR015301.

Palash, W., Y. Jiang, A. S. Akanda, D. L. Small, A. Nozari, and S. Islam, 2018: A streamflow and water level forecasting model for the Ganges, Brahmaputra, and Meghna Rivers with requisite simplicity. *J. Hydrometeorol.*, **19**, 201–225, https://doi.org/10.1175/jhm-d-16-0202.1.

Perez, G., R. Mantilla, and W. F. Krajewski, 2018: The influence of spatial variability of width functions on regional peak flow regressions. *Water Resour. Res.*, **54**, 7651–7669, https://doi.org/10.1029/2018WR023509.

Pool, S., M. Vis, and J. Seibert, 2018: Evaluating model performance: towards a non-parametric variant of the Kling-Gupta efficiency. *Hydrol. Sci. J.*, **63**, 1941–1953, https://doi.org/10.1080/02626667.2018.1552002.

Prior, J. C., 1991: *Landforms of Iowa*. University of Iowa Press,.

Quintero, F., W. F. Krajewski, R. Mantilla, S. Small, and B.-C. Seo, 2016: A spatial–dynamical framework for evaluation of satellite rainfall products for flood prediction. *J. Hydrometeorol.*, **17**, 2137–2154, https://doi.org/10.1175/JHM-D-15-0195.1.

——, R. Mantilla, C. Anderson, D. Claman, and W. Krajewski, 2018: Assessment of changes in flood frequency due to the effects of climate change: Implications for engineering design. *Hydrology*, **5**, https://doi.org/10.3390/hydrology5010019.

Rodriguez-Iturbe, I., and A. Rinaldo, 1997: *Fractal river basins: chance and self-organization*.

Schilling, K. E., C. F. Wolter, and E. McLellan, 2015: Agro-hydrologic landscapes in the Upper Mississippi and Ohio River Basins. *Environ. Manage.*, **55**, 646–656, https://doi.org/10.1007/s00267-014-0420-x.

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

Seo, B. C., and Coauthors, 2018: Comprehensive evaluation of the IFloodS radar rainfall products for hydrologic applications. *J. Hydrometeorol.*, **19**, 1793–1813, https://doi.org/JHM-D-18-0080.1.

Skøien, J. O., R. Merz, and G. Blöschl, 2006: Top-kriging - geostatistics on stream networks. *Hydrol. Earth Syst. Sci.*, **10**, 277–287, https://doi.org/10.5194/hess-10-277-2006.

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

Smith, J. A., M. L. Baeck, G. Villarini, D. B. Wright, and W. Krajewski, 2013: Extreme flood response: The june 2008 flooding in Iowa. *J. Hydrometeorol.*, **14**, 1810–1825, https://doi.org/10.1175/JHM-D-12-0191.1.

Villarini, G., J. A. Smith, M. L. Baeck, and W. F. Krajewski, 2011a: Examining flood frequency distributions in the midwest U.S. *J. Am. Water Resour. Assoc.*, **47**, 447–463, https://doi.org/10.1111/j.1752-1688.2011.00540.x.

——, ——, ——, R. Vitolo, D. B. Stephenson, and W. F. Krajewski, 2011b: On the frequency of heavy rainfall for the Midwest of the United States. *J. Hydrol.*, **400**, 103–120, https://doi.org/10.1016/j.jhydrol.2011.01.027.

——, ——, and G. A. Vecchi, 2013: Changing frequency of heavy rainfall over the central United States. *J. Clim.*, **26**, 351–357, https://doi.org/10.1175/JCLI-D-12-00043.1.

Zalenski, G., W. F. Krajewski, F. Quintero, P. Restrepo, and S. Buan, 2017: Analysis of national weather service stage forecast errors. *Weather Forecast.*, **32**, 1441–1465, https://doi.org/10.1175/WAF-D-16-0219.1.