Exploring Persistence in Stream Flow Forecasting

Ganesh Raj Ghimire and Witold F. Krajewski

Civil and Environmental Engineering (Ghimire, Krajewski), The University of Iowa, Iowa City, Iowa, USA (Correspondence to Ghimire: ganesh-ghimire@uiowa.edu)

This work has been accepted to Journal of the American Water Resources Association for publication. The JAWRA does not guarantee that the copy provided here is an accurate copy of the final published work.

Research Impact Statement: The **improved understanding** of forecasting through persistence and its dependence on basin scale and geometric properties of the river network helps using it for benchmarking model skill evaluations.

Abstract: In this study, the authors explore three persistence approaches in streamflow forecasting motivated by the need for forecasting model skill evaluation. The authors use stream flow observations with 15 minutes resolution from the year 2008 to 2017 at 140 U.S. Geological Survey (USGS) streamflow gauges monitoring the streams and rivers over the State of Iowa. The spatial scale of the basins ranges from about 7 km^2 to 37,000 km^2. The study explores three approaches: simple persistence, gradient persistence and anomaly persistence. The study shows that persistence forecasts skill has strong dependence on basin scales and weaker but non-negligible dependence on geometric properties of the river network for a given basin. Among the three approaches explored, anomaly persistence shows highest skill especially for small basins, under about 500 km^2. The anomaly persistence can serve as a benchmark for model evaluations considering the effect of basin scales and geometric properties of river network of the basin. This study further reiterates that persistence forecasts are hard-to-beat methods for larger basin scales at short to medium forecast range.

Key Terms: Streamflow Forecast Verification, Persistence, Width Function, Anomaly Persistence

INTRODUCTION

Persistence is a simple concept that says – "tomorrow will be like today". It exploits the notion of 'memory' of the analyzed system (Mittermaier, 2008). In this paper, we explore the concept as it applies to streamflow forecasting. Persistence-based predictions can serve as reference in measuring the skill of operational streamflow forecasting systems. The accuracy of real-time streamflow forecasts is of utmost importance for emergency response and water management agencies. In the United States, the National Weather Service (NWS) employs persistence forecasts as reference for the performance evaluation of their streamflow forecasts (see Gutierrez et al., 2006; NOAA and NWS, 2017).

We explore three approaches to persistence-based forecasting as well as their spatial scale dependence using data from 140 stream gauges in Iowa. Our study is motivated by the need for skill evaluation of hydrological model based forecasts and persistence is a convenient reference. We focus on short-term persistence or memory associated with streamflows. In addition, we explore the variability of persistence measures in relation to basin area and the geometric properties of river network (e.g., Iturbe and Rinaldo, 1997; Lashermes and Foufoula-Georgiou, 2007; Moussa, 2008). We explore several geometric descriptors to identify connections to the persistence characteristics of the system.

Several reference forecasts have been studied and implemented operationally around the world; these include simple persistence, streamflow climatology, a hydrological model fed with zero rainfall and a hydrological model fed with an ensemble of resampled historical rainfall (Bennett et al., 2013). Bennett et al. (2013) argued that using the streamflow observations for evaluating the skill of real-time streamflow forecasts is a rigorous test considering that rainfall forecasts have their own considerable uncertainties. Bennett et al. (2013) suggest that the use of persistence as reference can therefore be a better alternative for evaluating future estimates of stream flows at short to medium range.

Hydrologic systems possess some inherent unpredictability in its coupling with the weather and climate system. There is some likelihood of repetition at seasonal or even daily time scales. Under such conditions, the accuracy of forecasts may be higher (Mittermaier, 2008). Mittermaier (2008) points out that despite making it harder to beat, persistence as a reference forecast adds credibility to the actual forecast verification process. Other authors, den Dool (2007) and Fraedrich and Ziehmann-Schlumbohm (1994) concluded that only forecasts with skill better than persistence can handle the forecast of the time derivative, which is the essence of forecasting. For hydrologic predictions, the use of short-term persistence can be a better alternative than using climatology forecasts (Bennett et al. 2013; Pappenberger et al. 2008; den Dool, 2007). However, the degree of persistence shows variation with location, time of the year and the system studied (den Dool, 2007; Wu and Dickinson, 2005). In particular, den Dool (2007) indicated that the persistence forecast skill would be high for smaller lead times, if the process being forecast possesses a long time-scale. It is for this reason that persistence is a hard-to-beat method for many complex models found in literature. A recent study by Palash et al. (2018) illustrates this point.

Surprisingly, there is still a dearth of sufficient documentation on the quantitative evaluation of various versions of persistence forecasts of streamflow and their connection to hydrologic variables. Earlier studies related to streamflow variability and persistence have been in the context of developing statistical forecasting models and identifying hydrologically homogeneous regions (Vogel et al., 1998). Some authors associated the presence of persistence with the carryover storage of water in lakes, and below the land surface (Pagano and Garen, 2005).

Pagano and Garen (2005) studied observed trends in inter-annual streamflow variability and persistence in western U.S. and discussed comprehensively their implications for water management and seasonal forecasting. Welles and Sorooshian(2008) and Zalenski et al. (2017) used persistence to evaluate the skill of the operational NWS streamflow forecasts. Zalenski et al. (2017) explored the importance of river stage forecast verification process against the persistence forecasts. Their specific contribution was towards identifying the dependence of these forecast skills against certain hydrologic characteristics of the basins. For example, they revealed the strong dependence of forecast skill on the number of stations available upstream of the station being evaluated. The dependence of these skills with respect to drainage area and travel time was not as clear.

Though NWS uses persistence as reference forecasts for evaluating the skill of the model predictions, their own comprehensive evaluation of persistence is still lacking. The quantitative assessment for skill of persistence forecasts can provide the context for NWS to decide upon using them as reference at certain spatial scales and lead-times. This paper is aimed at exploring one specific objective: how skilled are persistence forecasts as function of river basin scale and geometric properties of river network?

METHODS

A collection of persistence forecast approaches can be found in the literature. These include simple persistence, anomaly persistence, and damped persistence among others (den Dool, 2007; Wu and Dickinson, 2005). Here, we explore three approaches: simple persistence, gradient persistence and anomaly persistence in streamflow forecasting.

If t is the time of measurement for streamflow observation X_t, the simple persistence forecast for lead-time Δt is given by,

$$X(t + \Delta t) = X_t \quad (1)$$

where the forecasts $X(t+\Delta t)$ are compared with observations X_t. One obvious advantage of this approach is its ability to predict peaks accurately, but delayed by the lead-time Δt.

For anomaly persistence, we compute anomalies as

$$X'(t) = X_t - clim(t) \quad (2)$$

$$X'(t + \Delta t) = X(t + \Delta t) - clim(t + \Delta t) \quad (3)$$

Then, $X'(t)$ is the anomaly forecast for $X'(t+\Delta t)$, where $clim(t)$ represents the climatology at time t. In this study, the climatology is based on the seventeen-year record for all stations. We computed it as seventeen-year average of streamflows at time t of each year. The forecast $X(t+\Delta t)$ is then given by,

$$X(t + \Delta t) = X_t - clim(t) + clim(t + \Delta t) \quad (4)$$

When climatology remains the same over Δt, anomaly persistence and simple persistence boils down to same forecasts. The skill of persistence forecast by (4), when measured in terms of correlation, is same as the auto-correlation in the long run (den Dool, 2007).

Using the gradient persistence forecasts approach, the forecast for lead-time Δt is given by,

$$X(t + \Delta t) = \frac{X_t - X(t - \Delta t)}{t - (t - \Delta t)} [(t + \Delta t) - (t - \Delta t)] + X(t - \Delta t) \quad (5)$$

which is equivalent to

$$X(t + \Delta t) = X(t - \Delta t) + 2 \times [X_t - X(t - \Delta t)] \quad (6)$$

The authors expect that gradient persistence forecasts can capture the rising or falling limbs of hydrographs for short lead-times.

To evaluate the skill of the persistence forecast, we report the performance measure called Kling-Gupta Efficiency (KGE). The metric can be decomposed into three components, i.e. correlation (r), variance ratio (α) and bias (β) (see Gupta et al., 2009). The ideal value of KGE is equal to 1 and is given by,

$$KGE = 1 - ED \tag{7}$$

where ED is given by,

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

where $\alpha = \frac{\sigma_s}{\sigma_o}$, and $\beta = \frac{\mu_s}{\mu_o}$, σ_s is the variance of persistence forecasts, σ_o is the variance of observations, μ_s is the mean of persistence forecasts, and μ_o is the observations. An attractive feature of the KGE metric is its simplicity of interpretation of each term. In addition, we report the skill in terms of timing of forecasts, and percent peak difference (PPD). Timing of forecasts is computed as the shift in the time needed with reference to observations to maximize the cross-correlation between the two. The PPD is computed as,

$$PPD = \frac{(peak_{sim} - peak_{obs})}{peak_{obs}} \times 100 \tag{8}$$

where $peak_{sim}$ and $peak_{obs}$ are the peaks of forecasts and observation respectively.

We implemented our study in the domain of the State of Iowa, which is bounded by the Mississippi and Missouri Rivers. Some 140 USGS streamflow gauges monitor the streams and rivers and provide the data we used in the study. Figure 1 presents the spatial distribution of these stations over the state. We used entire available record of streamflow observations with 15-minutes temporal resolutions from the year 2002 to 2018 in our analysis. Since this note is just an illustration of certain concepts, we deemed a period of seventeen years to be enough to capture the potential influence of seasonality or inter-annual variability in the streamflow dynamics. The spatial scale of the basins considered in this study ranges from about 7 km^2 to 37,000 km^2. This wide range of scales allows us to capture the spatial scale dependence of persistence forecast skill. About 65% of the area of the state drains to the Mississippi River while about 35% of the area drains to the Missouri River (Ghimire et al., 2018). The North-eastern part of the state is characterized by high-reliefs with deeply carved terrain, narrow valleys and relatively higher slope of the streams while the rest of the state is characterized by low-reliefs with relatively milder stream slope. Land use is predominantly agricultural with row crops.

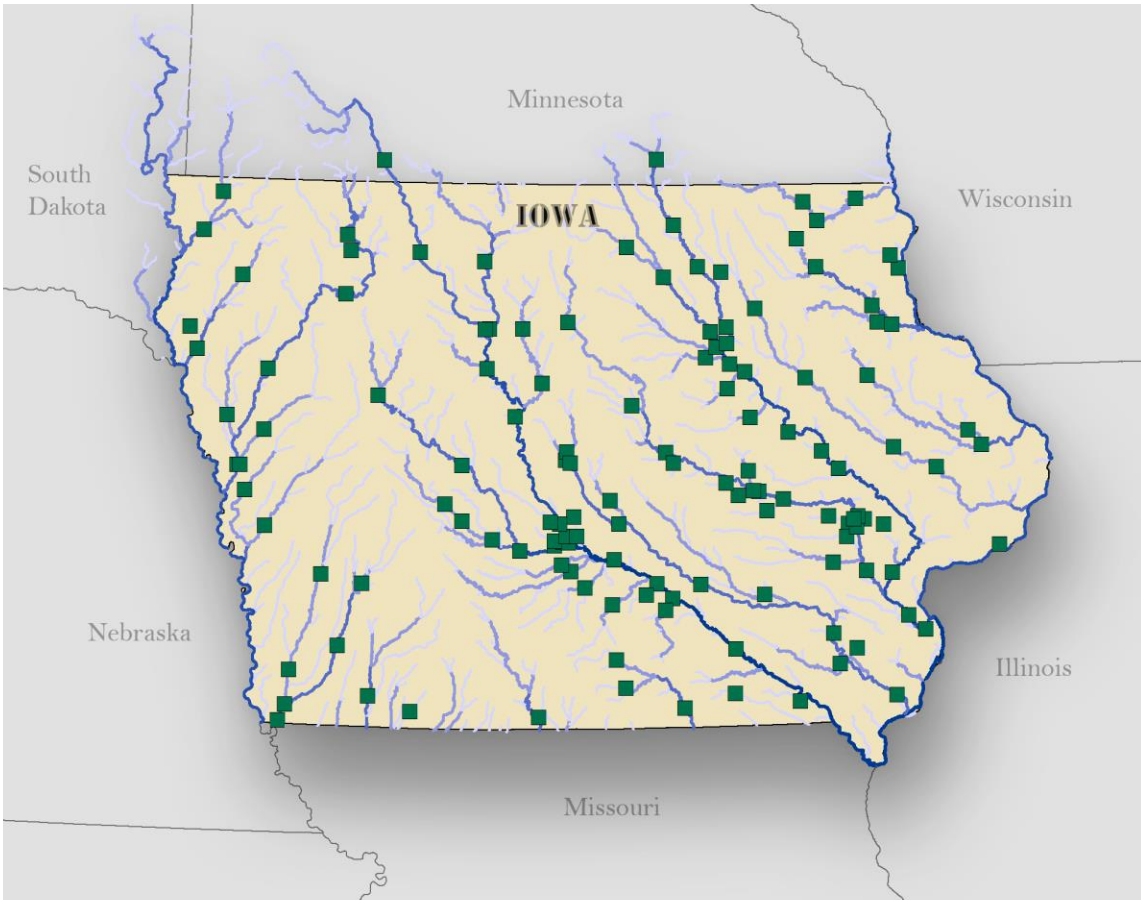

Figure 1: USGS stream gauge stations over the state of Iowa used in the analysis

RESULTS AND DISCUSSION

In this section, we present some key results from our persistence forecast evaluation at 140 USGS stations presented in Figure 1. Our goal here is to evaluate the performance of persistence forecasts for a range of lead times across spatial and temporal scales and to present comparative analysis of the three forecast approaches discussed in Section 2.

In this analysis, we mainly focus on the KGE as a measure of persistence forecast skill. Since for simple persistence by construction α=1 and β=1, the KGE is fully determined by the autocorrelation of the streamflow time series. The method does not modify the peaks, but all forecasts are delayed by the lag Δt. The anomaly persistence shows very similar behavior. The values of α and β are very close to unity, and consequently, KGE is mostly explained by the correlation. In other words, short-term persistence of streamflow is extensively dependent on its autocorrelation and is indicative of how good the forecasts are.

It is interesting to explore the performance of the persistence methods for different lead times. In Figure 2, we present the variability of KGE measures for different lead-times across basin scales based on simple persistence approach (1). We evaluated persistence forecasts up to lead-time of five days with an increment of 15 minutes. That is, we computed metrics for 480 such increments. We present here plots for some pertinent lead times from practical perspective, in connection with NWS's short to medium range forecasts. Three key features emerge out of subplots in Figure 2. First, the persistence skill depicted through KGE is clearly basin-scale dependent. As the basin size increases especially above 10,000 km^2, the KGE metric decreases at much slower rate. The forecast skill is virtually asymptotic to unity up to a lead-time of 1 day for the basin scales greater than 10,000 km^2. Second, the persistence forecasts skill is lead-time dependent. It is not surprising considering the potential decorrelation with longer lead-times. However, deeming the short to medium range streamflow forecasts issued by the NWS, it is apparent that simple persistence demonstrates KGE > 0.5 at more than 50% of our USGS monitoring stations.

Another interesting feature is the conditional variability in the persistence forecasts. For example, let us look at the three-day lead-time forecast skill. A large variability emerges in persistence measure for similar basin sizes, especially in the range of 100 km^2 to 10,000 km^2. Studies (e.g. Ayalew and Krajewski, 2017; Perez et al., 2018) show that river network geometry plays an integral role in explaining the difference in streamflow dynamics in addition to basin scales. We explore the connection of this conditional variability in persistence forecast skill to river network properties in our discussion later.

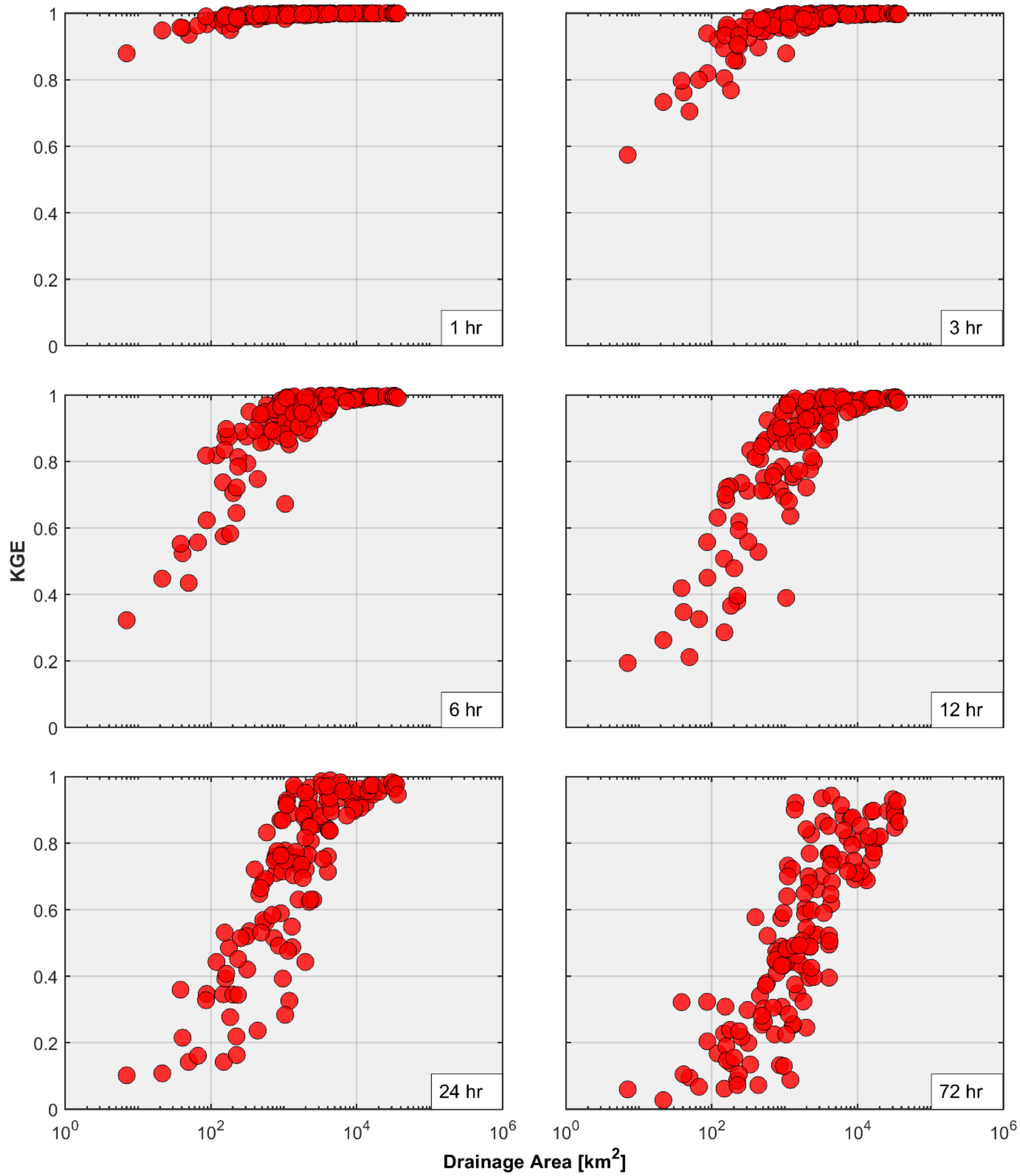

Figure 2: Variability of Kling-Gupta Efficiency (KGE) based on simple persistence approach across basin scales for different lead-times

The timing of simple persistence forecasts is, by construction, same as the lead-time. For anomaly and gradient persistence, timing is virtually close to the lead-time, however, some stations show an improvement over simple persistence. We attribute the improvement for these stations to the improvement in the correlation for anomaly persistence forecasts, as depicted by KGE. By construction, PPD is zero for simple persistence. The variability in PPD for anomaly and gradient persistence grows with lead-times while reduces with the basin scales. However, the variability is much smaller for anomaly persistence in comparison to gradient persistence. The median value of PPD ranges from 0 to 8% at 1 and 72 hours, respectively, for anomaly persistence. For gradient persistence, it ranges from 4 to 92% at 1 and 72 hours respectively. The timing and PPD skills show similar pattern of dependence with basin scales as KGE.

In terms of spatial dependence, both anomaly and gradient persistence methods demonstrate similar pattern for skill. Therefore, in Figure 3 we only present the one-to-one comparison between them. Clearly, the anomaly persistence shows better skill at all lead-times systematically, especially for smaller basin scales. The clear benefit with the anomaly persistence is that it accounts for the climatology into forecasts (see 4). For gradient persistence, more stations show further decay in skill with lead-times. This is somewhat surprising as one would expect it to work better at rising or falling limbs of hydrographs with no penalty for long inter event periods. However, its skill deteriorates when we forecast for longer lead-times (see 5) because of larger gradient proportional to longer interval.

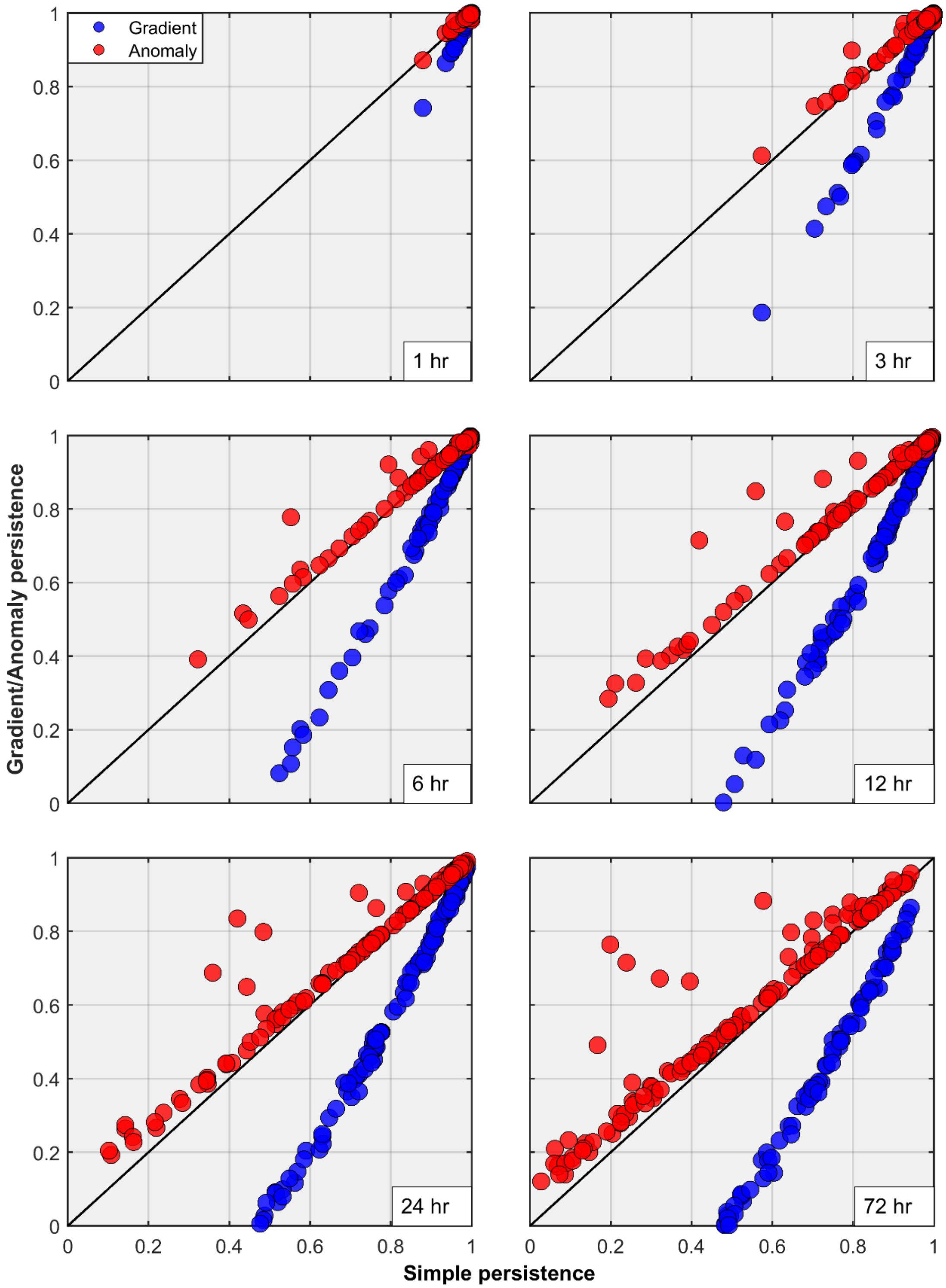

Figure 3: One-to-one comparison of KGE between different approaches of persistence for different lead-times. Blue dots represent comparison between gradient and simple persistence, while red dots represent the comparison between anomaly and simple persistence.

Figure 4 illustrates the evolution of persistence skill with lead-times across basin scales. For simplicity, we present here the basin scales that are different by an order of magnitude. There is a clear indication that the decay rate of skill is much slower for larger basins compared to smaller ones. Top two panels show that the decay is much faster at smaller lead- times and it takes longer for skill to decay from there onwards for smaller basins up to about 500 km^2. Another interesting result is that the discrepancy between the skills of simple and anomaly persistence methods is significant for basins up to about 500 km^2.

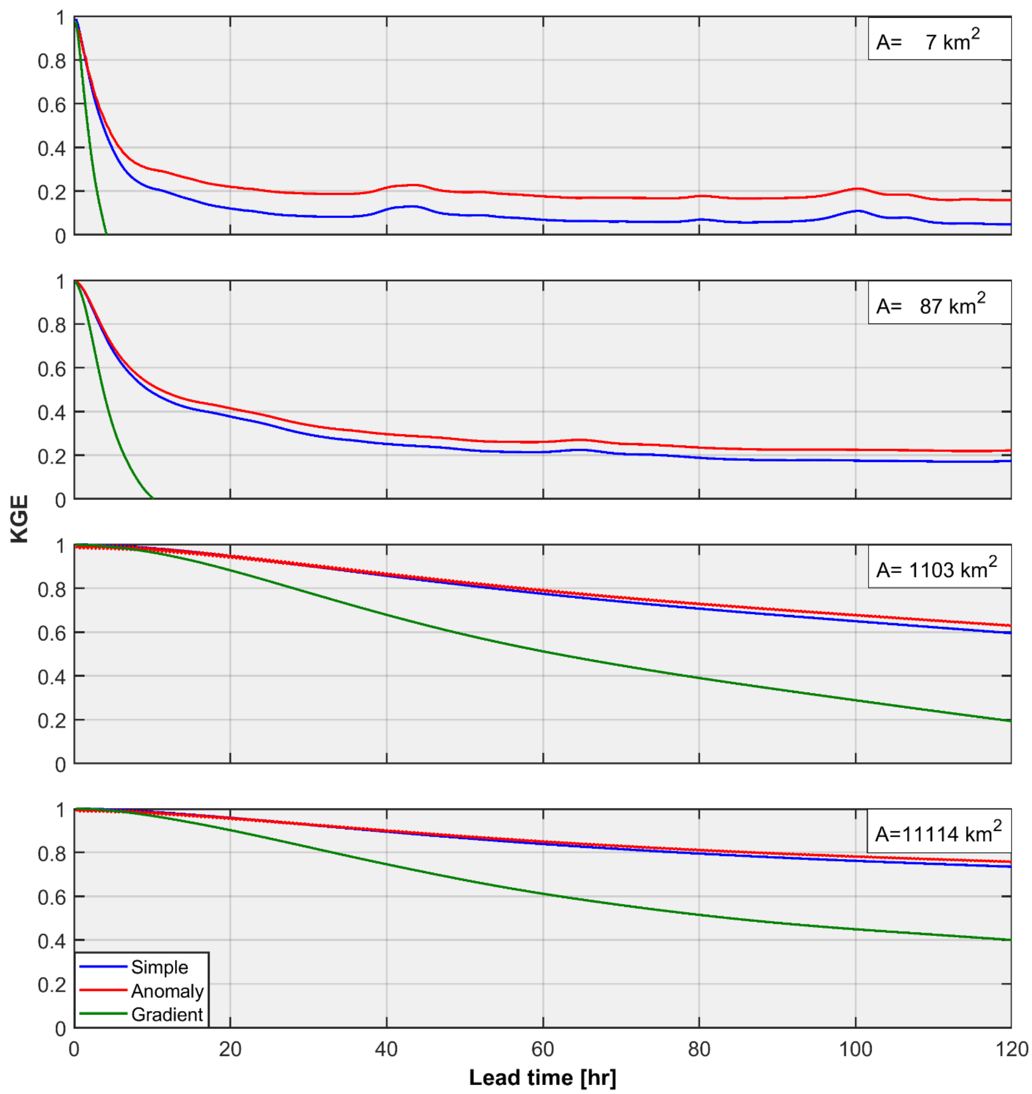

Figure 4: Comparative evolution of KGE measure across basin scales. Blue, red and green lines correspond to the simple, anomaly and gradient persistence methods respectively.

In addition to basin scales, we explore another explanatory variable associated with width function (see Rodriguez-Iturbe and Rinaldo, 1997) and refer to it as width function descriptor (WFD) (see Perez et al., 2018). It is a geometric property of the river-network showing the distribution of distance of stream-link to the basin outlet. As Perez et al. (2018) and Ayalew and Krajewski (2017) showed, basin scales show a strong relationship with WFD. Therefore, to avoid potential collinearity issues with drainage area, we explored the variability of persistence skill with the WFD at site and WFD at region (refer Perez et al., 2018 for details). Figure 5 shows the evolution of explanatory power of non-linear regression models depicted in terms of RMSE and R^2 for simple persistence forecast skills using two models. First model uses drainage area (A) as explanatory variable, while the second model uses both A and WFD as explanatory variables. We show in Figure 5 the comparison between these two models over lead-times considering normalized D_B (WFD at region), which denotes the normalized base of the width function, as other explanatory variable. Among several WFDs we explored, the normalized D_B showed better performance in combination with basin scale. The exponents of the model are positive implying that the persistence skill shows increase for larger basin scales with larger values of WFDs. Figure 5 clearly shows that as normalized D_B i.e. the time of concentration increases, the persistence skill also shows increase (see R^2 and RMSE) conditional on the same basin scale. Its soft implication is that persistence forecast skill depends not only on basin size but also on the basin shape. For example, an elongated basin will have higher persistence skill than a fat basin with the same size. As lead-times of basins get larger than the time of concentration of basins, they start showing the decline in the forecast skill. Physically, it indicates the persistence of streamflow until the system is triggered by the new event or the memory of the system is altered. Therefore, we can still achieve KGE > 0.8 for basins greater than 10,000 km^2 for a lead-time up to three days.

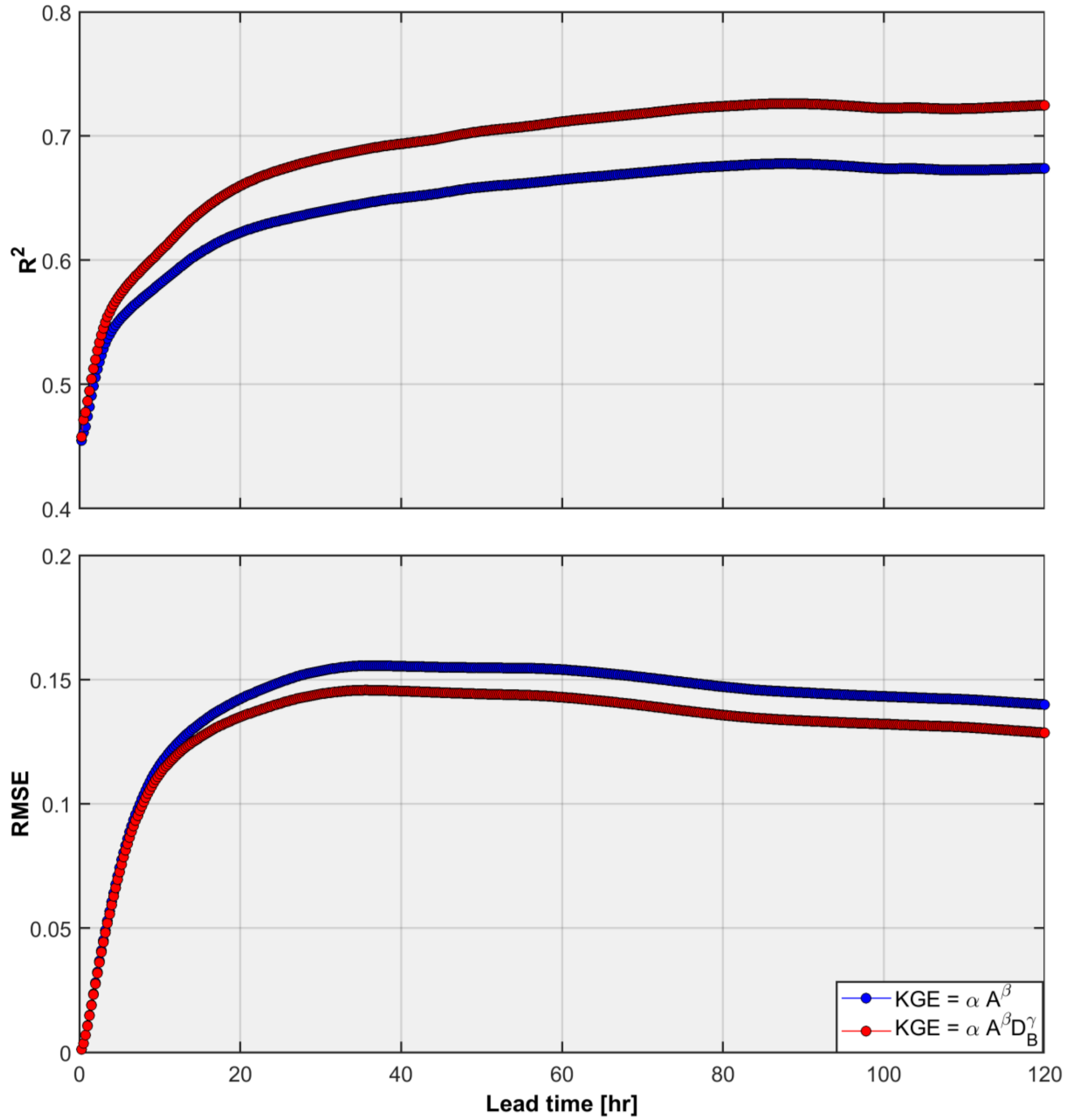

Figure 5: Comparison of explanatory power of regression models for persistence skills. Both R^2 and $RMSE$ are for models $KGE = \alpha A^{\beta}$ and $KGE = \alpha A^{\beta} D_B{}^{\gamma}$ for different lead-times.

Our results in Figure 5 illustrate that network characteristics depicted by WFDs can explain some variability of persistence forecast skill with A. However, there is still an unexplained variability in skill, which can arise from other potential factors such as basin scale heterogeneity, presence of regulating structures, lakes, and ponds, agricultural patterns, basin and channel slopes among others, which affect the memory of the system.

Our results on persistence forecasts skill demonstrate that there could be a wide range of its applicability as reference forecasts. As the NWS has been using it for some time in evaluating their real time streamflow forecast skill, we showed here that the use of anomaly persistence could enhance our reference forecast skill for smaller basins thereby also motivating the need to enhance our model forecast skills. One could also explore other approaches such as damped persistence for comparative analysis, which is beyond the scope of this study.

Another important implication is its potential use in the data assimilation. Our analysis clearly indicates that both simple and anomaly persistence skills are hard-to-beat for any hydrologic forecasting models in short to medium range for large basin scales. Therefore, forecasting models should assimilate these persistence forecasts, or observations, to enhance their skill. The point of this paper is not to suggest replacing models with temporal persistence. The authors acknowledge the importance of hydrologic models in the context of real-time streamflow forecasting in medium-long range.

SUMMARY AND CONCLUSIONS

We explore streamflow persistence for the state of Iowa at 140 USGS basins. Our study shows the variability of streamflow persistence forecasts skill depicted in terms of KGE, across basin scales and lead-times. Some of our key findings are as follows:

- Persistence forecasts skill show strong dependence on basin scales. It also depends on network characteristics of the basin. The three-persistence approaches explored *viz.* simple, anomaly and gradient show similar patterns of dependence. Among the three, anomaly persistence shows higher skill especially for basin scales up to about 500 km^2. Hence, it can prove to be better reference forecasts for evaluating skill of real time streamflow forecasts for basins of the order of such scales.

- This study further reiterates the fact that persistence forecasts are hard to beat method for larger basin scales with long timescale processes at short to medium forecast range.

While the results of this simple study may seem trivial, we hope that it will serve to illuminate the intuitively expected behavior with quantitative data. Surprisingly, we could find very little discussion pertaining to persistence forecasting in the hydrologic literature. Our results can serve as a benchmark for model evaluations; they should also be useful for the issues of data assimilation as well as development of data-based models through the ever increasing popularity of AI techniques.

Though we performed persistence analysis in Iowa, we believe that the forecast skill will show similar dependence on basin scales and network properties if extended to other parts of the U.S. Mountains and snow dominant regions, highly regulated rivers and basins with storages among others might demonstrate different degree of persistence, which is open to further research.

We neglected to discuss the effects of streamflow measurement errors as the USGS data are said to be accurate within 5%. The consideration of the errors becomes more important if one would like to consider the trade-offs between benefits of installing additional inexpensive stream gauges (to increase the river network coverage) and the errors due to synthetically derived rating curves (Quintero et al., 2018). We intend to extend the analysis to spatial aspects and hope to report our findings in the future.

ACKNOWLEDGEMENTS

The authors acknowledge the Iowa Flood Center at the University of Iowa for providing financial support for this work. We are thankful for the insightful comments and editorial suggestions from three anonymous reviewers. The second author acknowledges the support of the Rose & Joseph Summers endowment.

LITERATURE CITED

Ayalew, T.B. and W.F. Krajewski, 2017. Effect of River Network Geometry on Flood Frequency: A Tale of Two Watersheds in Iowa. *Journal of Hydrologic Engineering* 22(8):06017004. DOI: 10.1061/(ASCE)HE.1943-5584.0001544.

Bennett, J.C., D.E. Robertson, D.L. Shrestha, and Q.J. Wang, 2013. Selecting Reference Streamflow Forecasts to Demonstrate the Performance of NWP-Forced Streamflow Forecasts. *20th International Congress on Modelling and Simulation* (December):2611–2617.

den Dool, H. van, 2007. Empirical Methods in Short-Term Climate Prediction. Oxford University Press.

Fraedrich, K. and C. Ziehmann-Schlumbohm, 1994. Predictability Experiments with Persistence Forecasts in a Red-Noise Atmosphere. DOI: 10.1002/qj.49712051608.

Ghimire, G.R., W.F. Krajewski, and R. Mantilla, 2018. A Power Law Model for River Flow Velocity in Iowa Basins. *Journal of the American Water Resources Association* 54(5). DOI: 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. *Journal of Hydrology* 377(1–2):80–91. DOI: 10.1016/j.jhydrol.2009.08.003.

Gutierrez, C.M., J. Conrad C. Lautenbacher, and N.W.S.D.L. Johnson, 2006. National Weather Service River Forecast Verification Plan. *Baseline* (October).

Lashermes, B. and E. Foufoula-Georgiou, 2007. Area and Width Functions of River Networks: New Results on Multifractal Properties. *Water Resources Research* 43(9):1–19. DOI:

10.1029/2006WR005329.

Mittermaier, M.P., 2008. The Potential Impact of Using Persistence as a Reference Forecast on Perceived Forecast Skill. *Weather and Forecasting* 23(5):1022–1031. DOI: 10.1175/2008WAF2007037.1.

Moussa, R., 2008. What Controls the Width Function Shape, and Can It Be Used for Channel Network Comparison and Regionalization? *Water Resources Research* 44(8):1–19. DOI: 10.1029/2007WR006118.

NOAA and NWS, 2017. National Weather Service Instruction 10-1601.

Pagano, T. and D. Garen, 2005. A Recent Increase in Western U.S. Streamflow Variability and Persistence. *Journal of Hydrometeorology* 6(2):173–179. DOI: 10.1175/JHM410.1.

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. *Journal of Hydrometeorology* 19(1):201–225. DOI: 10.1175/jhm-d-16-0202.1.

Pappenberger, F., K. Scipal, and R. Buizza, 2008. Hydrological Aspects of Meteorological Verification. *Atmospheric Science Letters* 9(2):43–52. DOI: 10.1002/asl.171.

Perez, G., R. Mantilla, and W.F. Krajewski, 2018. The Influence of Spatial Variability of Width Functions on Regional Peak Flow Regressions. *Water Resources Research* 54(10):7651–7669. DOI: 10.1029/2018WR023509.

Quintero, F., W.F. Krajewski, M. Muste, M. Rojas, G. Perez, S.J. Johnson, A.N. Anderson, T.J. Honemuller, W. Cappuccio, and J. Zogg, 2018. Non-Conventional Ratings for Discharge Estimation with Consideration of Key Uncertainty Sources, Submitted To. *Journal of Hydrologic Engineering*.

Rodriguez-Iturbe, I. and A. Rinaldo, 1997. Fractal River Basins: Chance and Self-Organization.

Vogel, R.M., Y. Tsai, and J.F. Limbrunner, 1998. The Regional Persistence and Variability of Annual Streamflow in the United States. *Water Resources Research* 34(12):3445–3459. DOI: 10.1029/98WR02523.

Welles, E. and S. Sorooshian, 2008. Scientific Verification of Deterministic River Stage Forecasts. *Journal of Hydrometeorology* 10(2):507–520. DOI: 10.1175/2008jhm1022.1.

Wu, W. and R.E. Dickinson, 2005. Warm-Season Rainfall Variability over the U.S. Great Plains and Its Correlation with Evapotranspiration in a Climate Simulation. *Geophysical Research Letters* 32(17):215. DOI: 10.1029/2005GL023422.

Zalenski, G., W.F. Krajewski, F. Quintero, P. Restrepo, and S. Buan, 2017. Analysis of National Weather Service Stage Forecast Errors. *Weather and Forecasting* 32(4):1441–1465. DOI: 10.1175/WAF-D-16-0219.1.