# Using sensor data to dynamically map large-scale models to site-scale forecasts: A case study using the National Water Model

**Kevin J. Fries**[1], **Branko Kerkez**[1]

[1]University of Michigan, Ann Arbor, MI, USA

**Key Points:**

- Approach to dynamically map large-scale model forecasts to site-scale prediction using local sensor data
- Successful case study using publicly-available outputs of National Water Model and 180 water level sensors
- Performance analysis, generalizability of approach to other data & models, and open-sourced software implementation

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2017WR022498

Corresponding author: Kevin J. Fries, kjfries@umich.edu

## Abstract

There has been an explosive growth in the ability to model large water systems. While these models are effective at routing water across massive scales, they do not yet forecast the street-level information desired by local decision makers. Simultaneously, the increasing affordability of sensors has made it possible for even small communities to measure the state of their watersheds. However, these real-time measurements are often not attached to a predictive model, thus making them less useful for applications like flood warnings. In this paper we ask the question: how can highly localized forecasts be generated by fusing site-scale sensor measurements with outputs from large-scale models? Rather than altering the larger physical model, our approach uses the outputs of the unmodified model as the inputs to a *dynamical system*. To evaluate the approach, a case study is carried out across the US state of Iowa using publicly-available measurements from over 180 water level sensors and outputs from the National Water Model. The approach performs well across a third of the studied sites, as quantified by a high normalized root mean squared error. A performance classification is carried out based on Principal Component Analysis and Random Forests. We discuss how these results will enable stakeholders with local measurements to quickly benefit from large-scale models without needing to run or modify the models themselves. The results are also placed into a broader sensor-placement context to provide guidance on how investments into local measurements can be made to maximize predictive benefits.


## 1 Introduction

As computational power has grown, so has the ability of hydrologists to model complex hydraulic and hydrologic systems [*Blöschl et al.*, 2014]. No longer limited to the study of single stream reaches or small watersheds, increasing access to supercomputers and graphical processing units (GPUs) is now enabling a new generation of massive models, some of which would have seemed infeasible even recently. Presently, one exciting example is the United States' National Water Model (NWM) which provides forecasts for nearly 2.7 million stream and river reaches across the continental US [*Office of Water Prediction*, 2017]. While very impressive in scale, the performance of the model across most of these locations has still to be evaluated. Beyond numerical modeling, a variety of studies have also highlighted the potential of big data in hydrology, wherein large quantities of data are analyzed to provide scientific insight and improve forecasting performance (e.g.

*Karandish and Šimůnek* [2016]; *Tiwari and Adamowski* [2015]; *Chang et al.* [2017]; *Demir and Krajewski* [2013]; *Gilles et al.* [2012]). As such, there is now an unprecedented opportunity to begin leveraging advances in computing and data science to explore a variety of large and complex water challenges.

Advances in computation have also been accompanied by improved access to real-time measurements. Wireless sensor networks have become much more affordable [*Jin et al.*, 2010] and cloud-based services are now readily available, even to small research groups (e.g. Amazon Web Services, Microsoft Azure, Google Cloud etc.). The open source hardware movement (e.g. *Wong and Kerkez* [2016]; *Gilles et al.* [2012]; *Bitella et al.* [2014]; *Bartos et al.* [2017]) is empowering many technological non-experts, such as decision makers and small research groups, who can now deploy their own sensors to measure a variety of water parameters in near real-time. This is allowing important, but limited, sources of data, such as USGS gauges, to be supplemented by a variety of smaller and stakeholder-relevant measurements.

These advances still do not appear to be ushering in new wave of water management. At the level of individual communities or cities, water managers seek answers to very practical and neighborhood-specific questions. For example, forecasting the water level at specific bridges or highway overpasses can help trigger flood alerts or dispatch emergency response personnel. Given their spatial extent and therefore low availability of relevant data, such as bathymetry and forcing data, large numerical models may not always be accurate at high resolutions, meaning that their forecasts may not be immediately useful to decision makers. Additionally, units and variables that are important to modelers (e.g. flow) may not be the units and variables that decision makers care about (e.g. water level under a bridge). Alternatively, sensor observation alone may only go so far. While making a direct measurement at any specific site may provide real-time information to decision makers, it does not provide a forecast or warning without a model. There is, however, an opportunity to fuse the forecasting benefits of large-scale models with the site-level accuracy offered by local measurements.

In this paper we ask the question: how can highly localized forecasts be generated by fusing site-scale sensor measurements with outputs from larger-scale physical models? Instead of increasing the complexity of the physical model or re-calibrating it to match the local measurement, our approach leaves the physical model unchanged and uses a dynam-

ical systems transformation to map the large-scale model outputs to site-scale conditions. To evaluate this approach, we carry out a case study in which water levels, as measured by a sensor, are predicted from modeled flows made by a publicly-available and large-scale physical model. This will illustrate how city managers and other stakeholders, who have access to local measurements, can quickly benefit from large-scale models without needing to run or modify the models themselves. Specifically, we apply this methodology to the outputs of the US National Water Model and a publicly-available data set of hourly water level observations, made by over 180 sensors across the entire US state of Iowa. Beyond evaluating predictive performance, a *Random Forest*-based classification analysis is also carried out to evaluate under which conditions the approach is expected to perform well. The paper concludes with a discussion on the generalizability of the approach and places the findings into into a broader context of making *big* models and data *useful* to stakeholders.

### 1.1 Background

To illustrate the challenges that may be faced when translating macro-modeled outputs to high-resolution local conditions, we begin by using the outputs from the US National Water Model (NWM) to predict water levels at sites of interest. The desire to predict water levels, rather than flow, is motivated by two factors. Firstly, water levels are necessary for local flood inundation mapping [*Gilles et al.*, 2012]. Secondly, and more importantly, local measurements of flow are expensive and rarely available. Water level sensors, on the other hand, are relatively inexpensive to deploy and maintain, making them a more realistically available data source [*Jin et al.*, 2010].

Given its spatial extent, the NWM assumes trapezoidal stream cross sections, which are derived from the National Hydrography Dataset [*US Geological Survey*, 2017]. A mapping of flows to heights for specific sites may thus not be directly evident, since each location will have its own nuanced topographic and hydraulic properties. As such, there is a motivation to discover how the outputs of this large numerical model can be translated to site-specific parameters that are not directly modeled. If a clear relationship can be established between the modeled flows and measured heights for any given location, the forecasts of the NWM could then be used to provide authorities with precise localized flood inundation maps. This would allow local water managers, who have access to their

own measurements and knowledge of local inundation elevations, to benefit directly from the expertise embedded in the larger NWM.

Traditionally, rating curves have been a primary tool for deriving stream flows from stage measurements and vice-versa [*Herschy*, 1999]. Reliable rating curves require a relatively long history of stage and discharge measurements. Measurement-constraint alternatives have been proposed (e.g. *Damangir and Abedini* [2014], *Aricò et al.* [2010]), but often only work under limited conditions. Secondly, even when a long history of observations is available, rating curves can have large uncertainties, particularly related to heteroscedasticity [*Petersen-Overlier*, 2004], extrapolation outside of the history [*Kuczera*, 1996], hysteresis [*Perumal et al.*, 2004], measurement error [*Westerberg et al.*, 2011], and backwater effects [*Hidayat et al.*, 2011]. Most importantly, however, in the context of our proposed problem, the flows are modeled rather than measured, which poses additional challenges when attempting to estimate site-specific water levels. Indeed, the NWM's utility in predicting water levels has recently been studied by other researchers on smaller scales. For instance, *Javaheri et al.* [2018] used an Ensemble Kalman Filter in conjunction with the Height Above Nearest Drainage (HAND) method to predict water levels from the NWM. This methodology should work well for locations where a rating curve can be developed using the NWM flow estimates.

To illustrate the challenge of deriving local height estimates from modeled flows, we compare the output of the NWM to two independent water level measurements made on small bridges in Iowa (Figure 1). For the first example (Figure 1a), it is qualitatively apparent that there is a strong relationship between the modeled flows and the measured heights. This is supported by a dynamical agreement between the two time series (Figure 1c), which align well temporally, with clear agreement of the hydrograph peaks, as well as a generally good agreement on the rates of the rising and falling limbs. This provides a reliable rating curve and makes a strong case that the flow forecasts of the model could be used to predict future heights. On the other hand, for the second example (Figure 1b), the relationship between modeled flows and measured height is not nearly as clear. While the presence of a rain storm is evident in each time series (Figure 1d), it is unclear how the dynamics of each variable are correlated. The modeled flow is temporally coarse and does not match the dynamics of the measured water level. Without a clear rating curve, it may seem difficult to establish a relationship between modeled forecasts and measured heights, which may limit the apparent utility of the modeled forecast to this specific site.

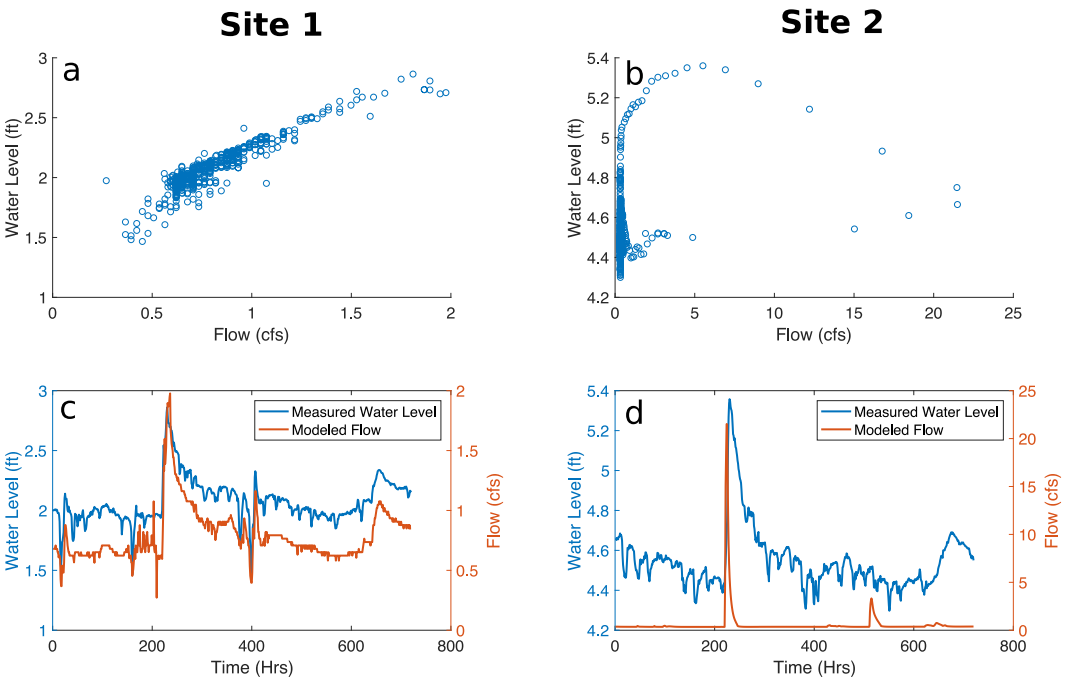



**Figure 1.** Measured water levels made by bridge sensors and modeled flows derived from the NWM for two example sites in the state of Iowa. The first example demonstrates a relatively strong relationship between modeled flows and measured water levels, while the second site does not.

When modeled flows do not directly align with local observations, one alternative is to directly assimilate the local measurements into the bigger model, thus improving its accuracy. Data assimilation is an established field in the hydrologic modeling community, relying on methods such as the Kalman Filter [*Beck*, 1987] or Particle Filter [*Moradkhani et al.*, 2005] to guide the model states toward the locally-measured values. In fact, the current version of the NWM performs a computationally low-cost form of data assimilation called Newtonian Nudging [*Hoke and Anthes*, 1976], whereby federal streamflow measurements from the United States Geological Survey (USGS) are used to "nudge" the model toward observed values. While the high quality and reliability of USGS gauges has been verified on many occasions (e.g. *U. S. Geological Survey and Koltun* [2015], *U. S. Geological Survey and Southard* [2013]), the number of gauges is limited compared to the scale and resolution of the NWM. As such, the NWM will benefit from assimilating alternative sources of information into its operation.

Expanding the coverage of the measurement network used by the NWM, such as measurements made by individual communities, poses a number of practical challenges

in the context of data assimilation. Firstly, given the sheer number of sensor manufacturers, deployment standards, and maintenance schedules, some sources of local data may be more reliable than others. Since measurement errors can propagate into the bigger model, assimilating local data thus poses questions regarding accuracy. Data will need to be approved and quality checks will be needed to ensure that any faulty sensors do not damage the model's integrity. Computational capacity will also need to be increased to ensure a growing number of assimilation points can be integrated. Given the sheer diversity of local water measurements and logistics associated with large-scale data assimilation, it is unclear when or if all of them will ever be ingested into the NWM. For those local water officials who do trust their own measurements, an alternative approach may still allow them to benefit from the existing forecasts offered by the NWM.

### 1.2 Approach and Contributions

Motivated by the challenges posed in the prior section, the major contribution of this paper is a computational approach by which independently-measured observations are combined with the output of a larger physical or numerical model to provide a dynamical forecast of local site conditions. In other words, historical model forecasts and independent historical measurements will be used to derive high-resolution and dynamical forecasts for a site of interest. The output will be an automated tool chain, which allows end-users to benefit from the expertise embedded in a large model without needing to update the model itself (Figure 2). Specifically, the approach will be evaluated by fusing outputs of the NWM and a large publicly-accessible stream sensor network in the state of Iowa [*Gilles et al.*, 2012]. Since these measurements have not been used in the NWM, they provide an independent data set for the evaluation of the approach. Practically, a successful demonstration of the approach will permit water managers, who may be inclined to invest into local measurements, to benefit directly from forecasts made by the NWM.

Since a one-to-one stage-discharge mapping is not possible for all sites (Figure 1b) our approach is based on dynamical systems theory [*Luenberger*, 1979]. Here, we treat the output of the physical model as the input to a *dynamical system*, with the idea that while the physical model may capture the general timing and magnitude of impulses, these outputs need to be mapped through a dynamical transfer function, to achieve agreement with measured values. Effectively, the approach will *learn* the response of a dynamical system, whose input is the physical model and output is the measured stage, and use it to

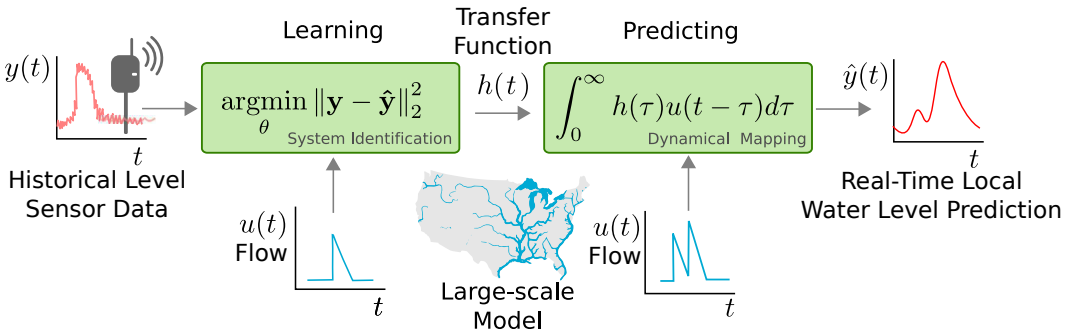



**Figure 2.** Conceptual diagram of dynamical mapping methodology. Historical measurements made by a sensor are used to *"learn"* a dynamical mapping between modeled flows and measured water levels. Once the parameters of the mapping are learned, water levels can then be predicted by dynamically transforming the modeled flows.

transform model forecasts to water level estimates. At a low-order level, this approach is analogous to *learning* a unit hydrograph [*Nash*, 1957], which have been used to map rainfall to flows (*Yang and Han* [2006], *Cluckie and Harpin* [1982]). However, simple single-order unit hydrographs are known to work mostly for smaller scale catchments [*Damangir and Abedini*, 2014]. Our approach addresses this limitation by expanding the order of the underlying system to be able to reflect more nuanced site-specific conditions.

The first part of this paper presents the theory, implementation and application of this approach to a large set of over 180 stream height observations. Secondly, we conduct a performance analysis which evaluates under which conditions the proposed approach will perform well. Given the sheer number of sites, each of which has a large number of physiographic features, a simple classification approach will not be adequate. Therefore, two analytical tools (principal component analysis and random forests) are used to determine which features explain when our approach can be used to reliably predict local conditions. The results of this analysis will provide a general sensor placement guide to help maximize the potential of mapping NWM output to local sites.

## 2 Methods

### 2.1 System Identification Theory

We frame the problem of mapping a physical model output $u(t)$ to a measured sensor value $y(t)$ as a transfer function operation, which can be represented in the time do-

main as a convolution with an impulse function $h(t)$ [*Luenberger*, 1979]:

$$y(t) = \int_{\tau=0}^{\infty} h(\tau)u(t-\tau)d\tau. \tag{1}$$

In our case study, the physical model output $u(t)$ represents the flow modeled by the NWM, while $y(t)$ are the height measurements made by a water level sensor at some location. The transfer function $h(t)$ can be converted to its frequency domain representation $H(S)$ using a Laplace transform:

$$\begin{aligned} Y(s) &= H(s)U(s) \\ H(s) &= \frac{Y(s)}{U(s)} \\ &= \frac{b_0 s^n + b_1 s^{n-1} + \cdots + b_{n-1}s + b_n}{s^n + a_1 s^{n-1} + \cdots + a_{n-1}s + a_n} \end{aligned} \tag{2}$$

where $(a_0, a_1, \cdots, a_n)$ and $(b_0, b_1, \cdots, b_n)$ are the $n^{th}$ order coefficients of the transfer function. More generally, the roots of the numerator's polynomial are known as the *zeros* and the roots of the denominator are known as the *poles* of the system. Since transfer functions are equivalent to systems of linear differential equations, an increase in the order of the system reflects the ability to represent more nuanced dynamics. Given a system order (i.e. number of poles and zeros), the goal is to learn the transfer function coefficients from prior measurement and modeled values, after which Equation 1 can be used to transform any future modeled flows to their corresponding heights. In the dynamical systems literature, this problem is broadly referred to as System Identification [*Luenberger*, 1979]. A common approach to learning the parameter $\theta := [a_1, \cdots, a_n, b_0, \cdots, b_n]$ of the model relies on the formulation

$$y(t) = \hat{y}(t, \mathbf{u}; \theta) + \epsilon(t, \theta) \tag{3}$$

where the measured output is a function of the predicted output $\hat{y}$ given parameter set $\theta$, which is corrupted by a noise term $\epsilon(t, \theta)$. Finding an estimate of the parameters $\hat{\theta}$ can be framed as an optimization problem that seeks to minimize the difference between modeled and observed values. Here we use the mean squared error as the loss function:

$$\begin{aligned} \hat{\theta}(\mathbf{y}, \mathbf{u}) &= \arg\min_{\theta} ||\mathbf{y} - \hat{\mathbf{y}}||_2^2 \\ &= \arg\min_{\theta} \sum_{t=1}^{n} (y(t) - \hat{y}(t, \mathbf{u}; \theta))^2 \end{aligned} \tag{4}$$

Our approach uses a Gauss-Newton method [*Bjorck*, 1996] to iteratively approach the minimum through the use of a gradient-based solver:

$$\theta^{(k+1)} = \theta^{(k)} - (\mathbf{J}^T\mathbf{J})^{-1}\mathbf{J}^T\epsilon(\theta^{(k)}) \tag{5}$$

where $\epsilon(\theta^{(k)}) = \mathbf{y} - \hat{\mathbf{y}}$ is a vector of the errors at iteration $k$, and $\mathbf{J}$ is the Jacobian matrix:

$$\mathbf{J} = \begin{pmatrix} \frac{\partial \epsilon_1(\theta^{(k)})}{\partial \theta_1^{(k)}} & \cdots & \frac{\partial \epsilon_1(\theta^{(k)})}{\partial \theta_n^{(k)}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \epsilon_m(\theta^{(k)})}{\partial \theta_1^{(k)}} & \cdots & \frac{\partial \epsilon_m(\theta^{(k)})}{\partial \theta_n^{(k)}} \end{pmatrix} \tag{6}$$

The Jacobian is a matrix of all the first-order partial derivatives of the error. Therefore, at each iteration, the parameterization of the transfer function model ($\hat{\theta} = [a_1, \cdots, a_n, b_0, \cdots, b_n]$) yields an estimated signal $\hat{\mathbf{y}}$ that approaches the true signal $\mathbf{y}$. Once $\theta$ is learned using time series of the inputs and outputs, forecasts can be made using Equation 1. A visual summary of the approach is provided in Figure 2.

### 2.2 Data sources and implementation

To promote transparency, reproducibility, and broader adoption by others, the authors have made all the formatted data, source code, and supplementary information available freely as an open source implementation on https://github.com/kLabUM/NWM/.

The approach was evaluated across two large data sources. These included the outputs of the US National Water Model, which served as the inputs $u(t)$ to our method. The second data set included 182 independently-measured (not assimilated into or used in the calibration of the NWM) streamgages across the state of Iowa, which represented the sensor measurements $y(t)$. The objective was to compare how well local water depths could be predicted by dynamically mapping the flows estimated flows by the NWM. Along with a summary of performance, an extensive analysis was also carried out using Principal Component Analysis [*Ouyang*, 2005] and Logit Boosted Random Forests [*Freund et al.*, 1996] to classify under which conditions the proposed approach may perform reliably.

#### *2.2.1 Data Source: The National Water Model*

The National Water Model (NWM) became operational in the fall of 2016, and is continuing to be developed by the Office of Water Prediction at NOAA. The NWM estimates flow for approximately 2.7 million stream reaches across the continental United

States. At its core, the NWM relies on large-scale Muskingum-Cunge routing, which is coupled with a gridded subsurface flow routing scheme [*Office of Water Prediction*, 2017]. The model is forced by rainfall from the Multi-Radar/Multi-Sensor System (MRMS) [*National Severe Storms Laboratory*, 2017] as well as a suite of model outputs ingested by WRF-Hydro [*Office of Water Prediction*, 2017]. Land surface processes, such as snowmelt, evapotranspiration, infiltration, and groudwater transfer, are simulated using Noah-MP [*Office of Water Prediction*, 2017]. Additionally, the NWM assimilates measurements from the national network of USGS streamgages. Given the continental scale of the model, a major appeal is that it routes flows from far away regions and covers locales that are often not captured by any other models. This should make it attractive for smaller communities seeking flash flood or streamflow forecasts but who may not have their own modeling resources. The NWM outputs hourly *nowcasts*, as well as 1-18 hour short-term forecasts, 0-10 days medium-term forecasts, and 0-30 days long-term forecasts [*Office of Water Prediction*, 2017]. Presently, modeled flows from the previous two days are freely available for download in NetCDF format on the National Centers for Environmental Prediction server (ftp://ftpprd.ncep.noaa.gov/pub/data/nccf/com/nwm). The NWM also provides an Analysis and Assimilation product which gives a three-hour *hindcast*. Because the NWM's forecasting ability is constantly being updated and improved, this paper uses this product to provide an upper bound baseline for our dynamical mapping approach.

#### *2.2.2 Data Source: The Iowa Flood Information System Sensors*

The Iowa Flood Center (IFC) was established in 2008 in response to the increasing frequency of flooding in the state [*Gilles et al.*, 2012]. One of their major initiatives was establishing the Iowa Flood Information System (IFIS), which provides real-time stream conditions and flood warning alerts [*Demir and Krajewski*, 2013; *Iowa Flood Center*, 2017]. IFIS ingests data from approximately 500 stream sensors, of which half are managed by the USGS and half are managed by the IFC (Figure 3). IFC gauges are primarily composed of bridge-mounted ultrasonic water level sensors, which transmit sub-hourly measurements across a wireless connection. Historical depth measurements are freely available on the IFIS website across a rolling 30 day window [*Iowa Flood Center*, 2017]. In this paper, we focus on the 220 bridge-mounted sensors that the IFC manages, since these sensors were not used in the calibration of the NWM. As such, they provide an independent validation data set for the proposed method.

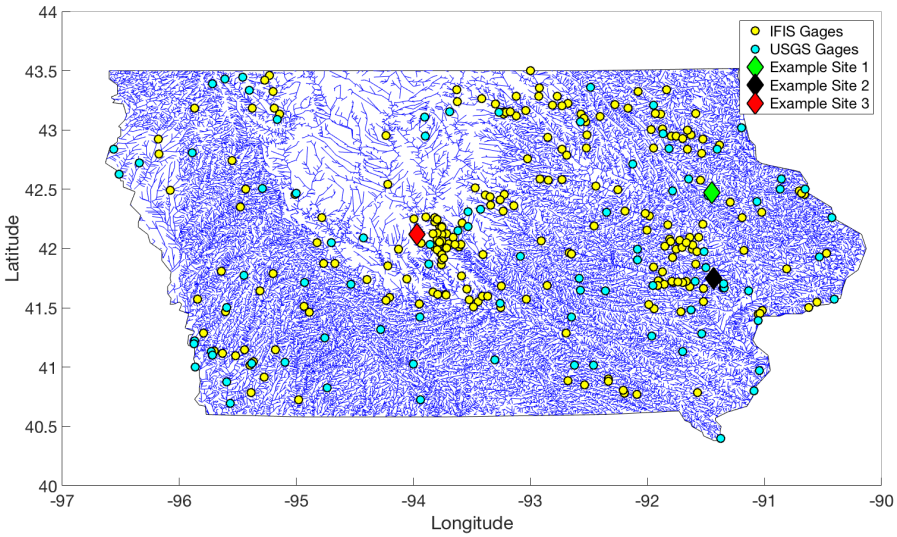



**Figure 3.** Visualization of the nearly 62,000 streams modeled by the NWM in the state of Iowa. USGS gages, which are assimilated into the NWM, are denoted as cyan circles. Locations of the IFIS water level sensors are denoted as yellow circles, with diamonds denoting the three example sites used in this paper.

### *2.2.3 Implementation*

Outputs from NWM and IFIS gauge measurements were recorded using an automated *Python* script on an hourly basis from October 2016 through May 2017 across the state of Iowa. IFIS gauge readings were logged in real-time as measurements became available. Out of the 220 candidate sites, 182 were co-located with outputs of the NWM and deemed to have a continuous record. For small data gaps (few missing points), linear interpolation was applied to create continuous time series. The NWM and IFIS time-series were linked by location, thereby providing individual model-measurement pairs that could be used in our dynamical mapping approach. Data from October to December were used to train our system identification approach, while data from March to May were used for validation. To reduce potential impacts of wintertime conditions (freezing, snow, and snowmelt), which may have influenced NWM outputs and gauge maintenance, data across January and February were not used in the analysis. Additionally, stationary of the physiographic features was assumed (e.g. it was assumed that erosion does not change stream width or roughness over the study period).

Prior to applying the system identification procedure, sensor data were linearly detrended to remove the impact of base flows, which was necessary to ensure that the transfer functions would decay to zero following a storm event. Since the complexity of the dynamical mapping was not known a priori, an ensemble of 14 different transfer functions was learned using the training data, with each mapping having varying numbers of poles and zeros (see Equation 2). These included all possible pole-zero pairings for first through fourth order systems ([0 poles, 1 zero],[1 pole, 1 zero],[0 poles, 2 zeros],...,[4 poles, 4 zeros]). This allowed for the average and upper-bound performance of the approach to be compared across mappings of varying complexities. The final software toolchain was implemented in `MATLAB`, using an implementation of the System Identification procedures from *Ljung* [1987]. For comparison, a standard regression rating curve procedure [*U.S. Geological Survey et al.*, 2010] was also implemented, whereby prior stage-discharge relationships (October-December) were used to predict future values (March-May). The normalized root mean squared error (nRMSE), which is equivalent to the Nash Sutcliffe Efficiency (NSE) [*Nash and Sutcliffe*, 1970], was used to evaluate performance:

$$E = 100\left(1 - \frac{||y - \hat{y}||}{||y - \bar{y}||}\right) \tag{7}$$

where $E$ is the NSE in percent, $y$ is the vector of observed water level, $\hat{y}$ is the vector of predicted water level, $\bar{y}$ is the mean of the observed water level , and $||\cdot||$ is the Euclidean norm [*Deza and Deza*, 2009]. For interpretation, a value of 100% would imply a perfect prediction of water levels, a value of 0% would imply a prediction that is as good as taking the historical average of water levels, and a value less than 0% indicates a performance more inadequate than taking the average. An NSE of 50% or above is generally considered the lower bound for a good predictor [*Krause et al.*, 2005], which is the threshold we adopted in this paper. The analysis considered the NSE of both the best ensemble member and the ensemble average.

### 2.3 Performance Classification

One major goal of this paper is to investigate under which conditions the proposed dynamical mapping approach will work well. Not all locations may benefit directly from our approach, even if investments into local sensors are made. Evaluating which features explain this behavior will be crucial to informing where investments into sensors should be made to maximally leverage the NWM. To classify the performance of our approach

under various physiographic conditions, we used a combination of Principal Component Analysis [*Hastie et al.*, 2001] and Random Forest Classifiers [*Freund et al.*, 1996].

The NWM is built on a number of distinct physiographic features from the National Hydrography Dataset (NHD) [*US Geological Survey*, 2017]. These include the channel bottom width, elevation, Manning's roughness, channel slope, and Strahler stream order [*Strahler*, 1957]. Some features, such as side slope, were available, but were were uniform throughout the study area and thus deemed uninformative. Some features served as proxies for others that would have been too labor intensive to derive (e.g. stream order and relative drainage area).We also calculated an additional feature, which captures the distance of a given water level sensor to the nearest USGS gauge. This will indicate whether our approach performs better near official NWM data assimilation locations. Overall, this provided six features that may be used to explain the performance of our dynamical mapping approach. For example, intuition would suggest that our approach would work well on larger rivers, where the NWM may be able to capture flow dynamics more accurately than in smaller, ungauged basins. This, however, has to be confirmed, especially given the array of other complex features that may explain performance.

Since some of the features analyzed in this study (e.g. stream order vs. bottom width) may exhibit collinearity or multicollinearity, they must be orthogonalized to maximize the ability to classify around them. Before the performance is classified, our approach used Principal Component Analysis (PCA) to shift the six dimensional feature space into an orthogonal subspace [*Hastie et al.*, 2001]. PCA changes the coordinates of the features, with the objective of finding a new set of features that are linear combinations of the original features. PCA initially determines the direction in which the greatest amount of variance lies, defines the first axis to align with that direction, and then iteratively re-orients subsequent axes such that each axis is aligned in the direction of next greatest variance. In doing so, the features are de-correlated and combined into composite principal axes that should maximize the ability to discover higher-dimensional hyperplanes that can be used during classification. In consideration of succinctness, a theoretical discussion of our PCA implementation is provided in Appendix A: .

Once the features that describe all of the 182 sensor locations were PCA-transformed, each of the sites was labeled based on performance of the dynamical mapping. The predictive performance was labeled in a binary sense, whereby sites with a maximum NSE

of 50% or greater were deemed to perform well (label 1), while any remaining sites were labeled as inadequate (label 0). The performance classification was then implemented as a *supervised learning* procedure, where the final classification seeks to predict how well the dynamical mapping approach will perform for a given set of features. While various classification algorithms exist, our approach used a statistical learning tool known as Logit Boosted Random Forests, or *Adaboost* with trees [*Freund et al.*, 1996].

*Adaboost* generates a large number of "weak learners" [*Hastie et al.*, 2001] in the form of small classification trees. A weak learner is a model that is only slightly better than randomly guessing [*Zhou*, 2012]. A classification tree partitions a feature space using a series of binary splits, resulting in a large number of labeled bins. For example, a simple tree may categorize the performance of our dynamical approach as either "high" or "low". These labels are assigned to training data, as described above. The tree may be trained to classify based on decision variables, such as stream order or elevation. The training data are then placed into bins, or leaves. The label of each bin is assigned based on majority voting, or how many "high" and "low" labels are inside each bin. The decision variables are used to improve binning. For example, all of the data may first be split up by elevation to create two bins. These bins may then be split further by stream order to reduce the label variance inside each bin. In this example, this would create only four bins. The low number of bins, and thus low complexity, of such a classifier is the reason why it is called a weak learner. Given their relative simplicity, these trees tend to have very low bias but also very high variance [*Hastie et al.*, 2001]. This can be addressed by generating an ensemble, or a *forest*, of many trees. Going a step further, we seek to ensure that each tree is developed to provide as much information gain as possible. In Logit Boosted Random Forests, each data entry (i.e. labeled row of our data matrix $\mathbf{T}$) is given an initial weight $w_i$. Then, as new trees are learned, the data entries are re-weighted so as to emphasize where the model is failing. That is, the final algorithm (Algorithm 1) ensures that misclassified data are stressed more in the learning of the next tree. In this implementation, $\mathbf{t}$ is the input data (i.e. a row of $\mathbf{T}$), $y$ is the observed data (1 for a site labeled as well-performing site, $-1$ for a bad site), $N$ is the number of observations, $M$ is the number of trees, $p_m$ is the probability output from tree $m$, and $H(z) := \mathbf{1}_{[z>0]}$ is the Heaviside step function [*Hastie et al.*, 2001]. The $M$ classification trees are learned in an iterative fashion.

1 **Initialize** $w_i = \frac{1}{N}, i = 1, 2, ..., N$;

2 **for** *m=1,2,...,M* **do**

3 Learn classification tree that outputs $p_m(\mathbf{t_i}) = P_w(y = 1|\mathbf{t_i}) \in [0, 1]$ with weights $w_i$;

4 Set $f_m(\mathbf{t_i}) \leftarrow \frac{1}{2} \log \frac{p_m(\mathbf{t_i})}{1-p_m(\mathbf{t_i})}$;

5 Set $w_i \leftarrow w_i \exp[-y_i f_m(\mathbf{t_i})], i = 1, 2, ..., N$, and renormalize such that $\sum_i w_i = 1$;

6 **end**

7 **Output** classifier as $H[\sum_{m=1}^{M} f_m(\mathbf{t})]$

**Algorithm 1:** Logit Boosted Random Forest

The logit function (line 4 of the algorithm) is used to re-weight the inputs (line 5). Because of the form of the logit function, much larger values exist closer to 0 and 1. The result is that if the data entry $\mathbf{t_i}$ is classified properly and with high probability, then $\exp[-y_i f_m(\mathbf{t_i})]$ in line 5 will trend towards zero. If it is classified improperly with high probability, then this term will approach infinity. This ensures the re-weighting will target poorly classified data on the next iteration and that properly classified data will be largely ignored. After learning all $M$ models, any new input $\mathbf{t}$ can be provided and, when summing over all $f_m(x)$ trees, a prediction can be made for whether a site will be a good candidate for our dynamical mapping approach. A good site will be one that sums to be greater than 0 and a bad site will sum to be less.

## 3 Results

### 3.1 Dynamical Mapping Performance

After training and applying the dynamical mapping (DM) procedure across all 182 sensor locations, predictions at approximately one-third of the sites (55/182) exceeded the desired 50% NSE threshold, while performance across 90 sites exhibited an NSE of at least 40%. The overall performance of the approach is summarized in Figure 4, showing that the DM procedure consistently performed better than a simple rating curve approach. Indeed, in all but 8 cases, water levels were predicted more accurately using the proposed DM approach compared to a regression between measured levels and NWM-modeled flows. The order (i.e. the number of poles and zeros) of the transfer functions that had the best performance was not consistent site-to-site. The order is analogous to the complexity of the transfer function, where higher order systems are capable of describing complex dy-

namics. However, complexity of the model in the frequency domain does not necessarily lead to improved results in the time domain since the accuracy depends on the amount of information that can be extracted from the system [*Rojas et al.*, 2010]. Here, information is tied to the number of frequencies present in the signal. Put simply, if the dynamics of the measured water level and NWM can be described by a small frequency range (which may be unique to each site), then making the transfer function more complex will lead to improvements in performance since the additional coefficients will be relatively small in magnitude. In many cases, higher-order poles and zeroes may actually cancel, thus effectively providing an implicitly low-order model. As such, higher complexity may not necessarily improve model performance. There were no site-specific physiographic features, which could be used to determine what this upper bound was in our case study.

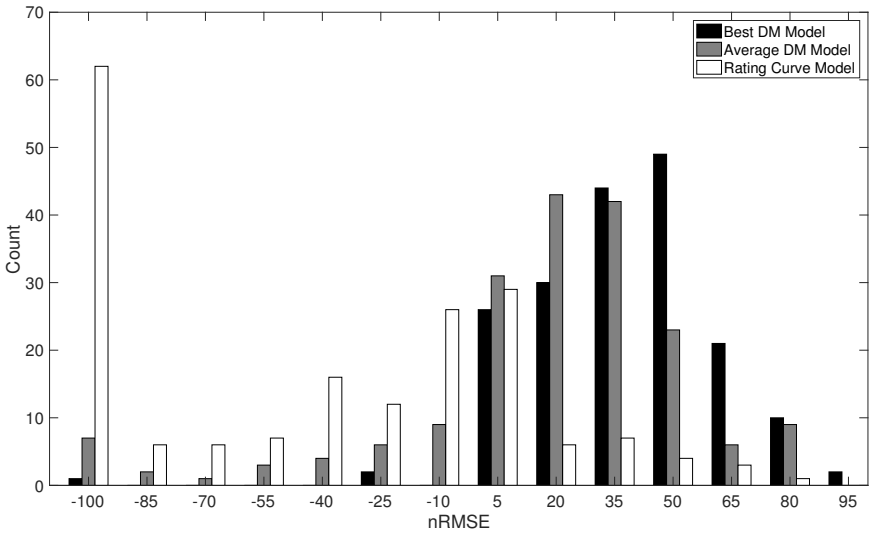
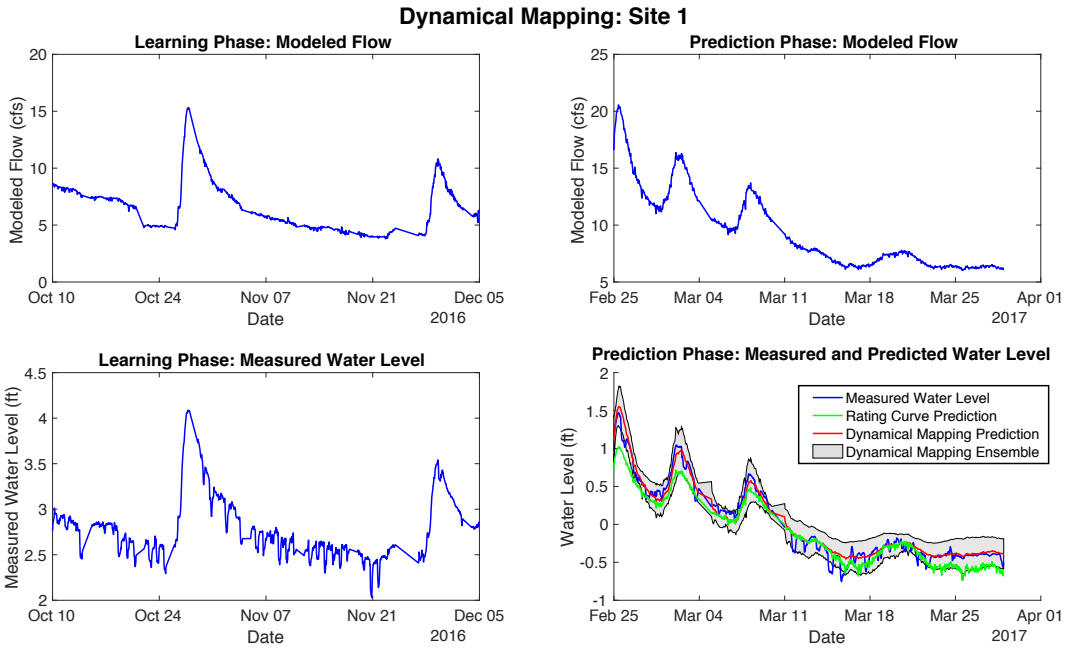



**Figure 4.** Histograms of prediction performance (nRMSE) evaluated across 182 sensor locations. A comparison is made between the best dynamical mapping (black), ensemble of dynamical mappings (gray), and a simple regression-based rating curve approach (white)

Given the sheer number of sites used in the analysis, this section will evaluate three locations in detail, while the remainder are plotted in the supplementary information. The three sites were selected to reflect three types of performance, as measured by NSE. The first is a location for which our DM approach provides a strong predictive performance, in large part due to a high correlation between the NWM predicted discharge and the observed stage. The second site exhibits strong predictive performance, despite the NWM

455 providing coarse outputs. The final example illustrates a case in which there is a limited
456 ability to predict observed heights from flows.

457 The first example demonstrates a case of strong predictive performance (Figure
458 5). The left column of the figure displays the training data, which includes the NWM
459 model outputs and measured water levels for the Fall of 2016. The right column shows
460 the NWM outputs and measured water levels for the Spring of 2017, as well as the wa-
461 ter level predictions made by our DM approach. Specifically, the bottom right panel is the
462 average prediction made by our approach across all 14 transfer functions (red line, with
463 gray area indicating variability within the ensemble) compared to the measured water lev-
464 els (blue line). Overall, our DM procedure performed well at this site, with an average
465 NSE close to 80%. Predictions of water levels at this site using a simpler regression-based
466 rating curve performed nearly as well, with an average NSE of 76%.




467 **Figure 5.** Dynamically mapping modeled flows to local water levels on site 1 (see Figure 3). Data used
468 to "learn" the mapping parameters are plotted in the left column, while the resulting mapping is applied to
469 future data in the right column. For this example site, the dynamical mapping performs relatively well (NSE
470 of 80%). A simple regression-based rating curve approach (not plotted) performs strongly as well, with an
471 NSE of 76%.

The second example (Figure 6) illustrates a case where a simple regression approach did not perform well (NSE of -4%). The modeled flows were quite impulsive and not representative of observed dynamics. However, when our DM approach was used, the results improved significantly, with an average NSE of over 50%.

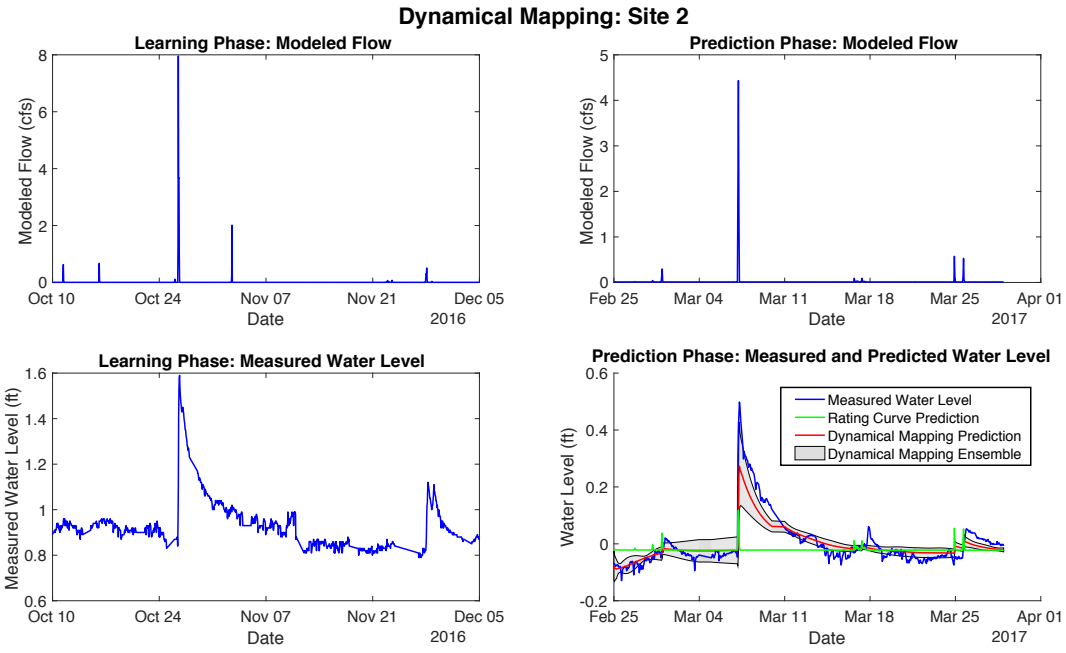
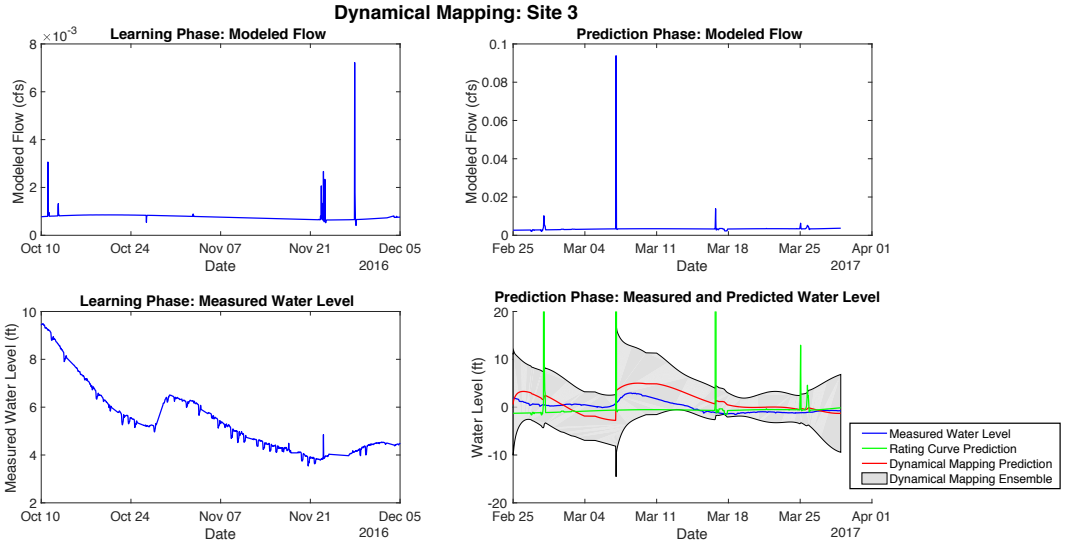



**Figure 6.** Dynamically mapping modeled flows to local water levels on site 2, following convention used in Figure 5. For this example site, the dynamical mapping performs relatively well (54% NSE), while a simple regression-based rating curve approach does not (−4% NSE).

Finally, the third example (Figure 7) illustrates a location at which no good predictive performance can be reached, regardless of the approach used. As evident in the figure, the measurements reflected a slowly changing system, while the NWM showed a series of rapid impulses. An average NSE of -145% was obtained using our DM approach, with only one of the 14 ensembled transfer functions showing a slightly favorable NSE (49%). The rating curve method was even more ineffective, with an NSE of -14,900%. Further, the rating curve method predicted water levels well outside the realm of possibility given its linear extrapolation (close to 6km water height).




487 **Figure 7.** Dynamically mapping modeled flows to local water levels on site 3, following convention in
488 Figure 5. For this example site, neither the dynamical mapping (0% NSE) nor regression-based rating curve
489 (NSE of $-14,900\%$) perform well.

490 ### 3.2 Performance Classification

491 The specific performance of the DM approach across sites 1-3 could be loosely de-
492 scribed by channel width, the stream order, and the distance to a USGS gage. Sites 1, 2
493 and 3 had channel widths of 20ft, 10ft, and 5ft; stream orders of 5, 3, and 1; and distance
494 to USGS gages of .2km, 2.2km, and 5.6km, respectively. More broadly, using the 50%
495 NSE criterion, 55 of the 182 sites were labeled as locations of *high* performance, while
496 127 were labeled as *low* performing, reflecting the ability of our DM approach to predict
497 flows from NWM outputs. These labels were then used to determine the combination of
498 physiographic characteristics that describes the conditions under which the DM approach
499 exhibits high performance. The normalized distributions of each physiographic feature,
500 split by performance criteria, is shown in Figure 8a. This normalization was performed
501 relative to all 64000 streams in Iowa, not just the 182 sites studied. Overall, little distinc-
502 tion was evident between high-performing and low-performing sites, with the distribution
503 of each physiographic feature showing similar means and variances. The distributions of
504 channel bottom width and channel slope showed the relatively largest discrepancy, sug-
505 gesting that sites at which the DM approach performed well had a larger stream width and
506 slope than lower-performing sites. However, the bounds on these distributions were not
507 sufficient to determine a consistent labeling.

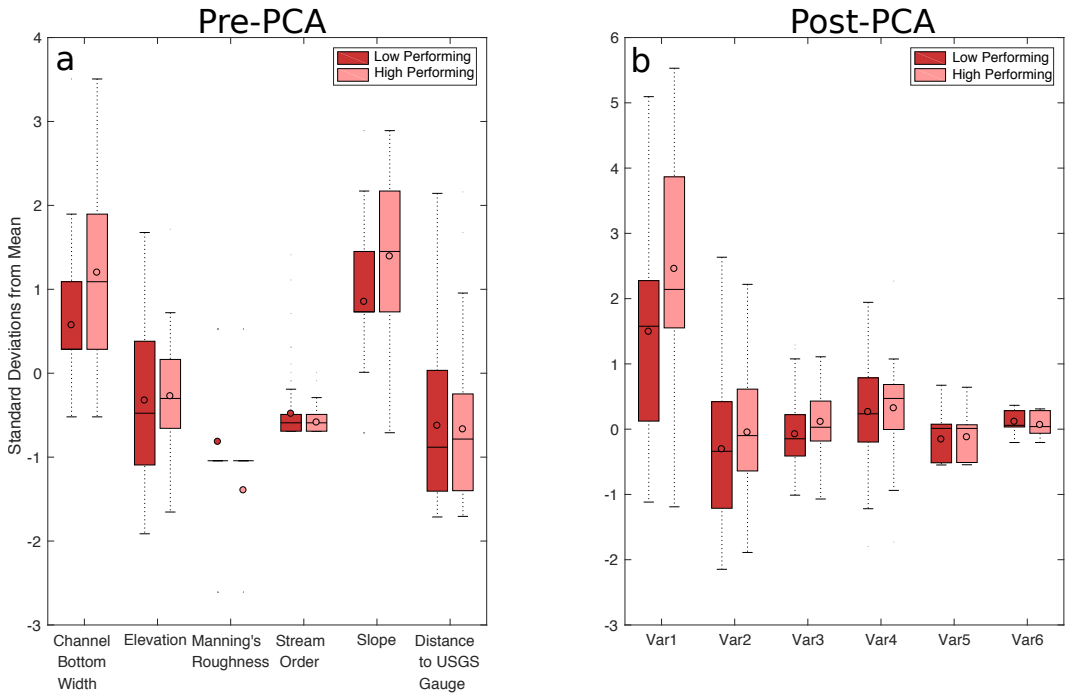



**Figure 8.** Boxplots representing the relative distribution of features, when split by the ability of a dynamical mapping to predict water levels from modeled flows. For any given feature, a clear difference between the two distributions would indicate that this feature describes a general condition for the dynamical approach to work well. a) This plot shows the splits based on stream physiographic features. It is not apparent in this figure that any features describe a general condition for the DM approach to work well. b) This plot shows the splits based on principal components (new variables 1-6). Here, the first principal components exhibits the strongest difference between the high and low performing sites, illustrating a potentially strong indicator of prediction performance.

Applying PCA to the physiographic features across the entire state of Iowa resulted in a $62000 \times 6$ data matrix. The resulting principal components are shown in Table 1. Each entry in a column of this table can be interpreted as the relative influence of a physiographic variable to a particular principal component. For example, considering the first principal component, which explains the greatest amount of variability in the physiographic data, it becomes apparent that the channel bottom width and the stream order both increase as the first principal component score increases. On the other hand, the Manning's roughness decreases as the principal component score increases. As such, if a stream reach in the data set has a large first component score, it will be relatively larger and smoother than other streams. Further, a number of the components exhibited oppos-

ing physiographic relationships. For example, for the second principal component, streams closer to a USGS gauge and located at a relatively higher elevation had relatively higher component values. For the fourth component, this relationship was reversed, as stream reaches at higher elevations and located further away from a USGS gauge tend to have higher component values. Similar interpretative examples could be provided for the other principal components.

**Table 1.** Principal Components resulting from applying PCA to features of 62,000 streams across the entire state of Iowa. $\sigma$ is the singular value associated with that component whose relative magnitude indicates the amount of variability the component explains in the data.

| Stream Feature | Comp. 1 $\sigma = 3.16$ | Comp. 2 $\sigma = 1.09$ | Comp. 3 $\sigma = 0.86$ | Comp. 4 $\sigma = 0.77$ | Comp. 5 $\sigma = 0.08$ | Comp. 6 $\sigma = 0.04$ |
|---|---|---|---|---|---|---|
| Bottom Width | 0.534 | 0.140 | 0.185 | -0.041 | 0.693 | -0.423 |
| Elevation | -0.198 | 0.648 | 0.182 | 0.712 | 0.016 | 0.000 |
| Manning's Roughness | -0.535 | -0.143 | -0.170 | 0.008 | 0.720 | 0.382 |
| Slope | -0.254 | -0.075 | 0.932 | -0.240 | -0.015 | 0.062 |
| Order | 0.545 | 0.141 | 0.105 | -0.004 | 0.023 | 0.819 |
| Proximity to USGS gage | -0.165 | 0.717 | -0.157 | -0.659 | -0.012 | 0.003 |

The performance of the DM approach, split by principal components, is shown in Figure 8b. Compared to splitting based on just physiographic features (Figure 8a), a more distinct clustering was evident for a few of the new variables. This is especially true for the first principle component, for which a larger component score generally corresponded with higher performance of the DM approach. While the other principal components did not exhibit as large of a discrepancy, the opposing physiographic relationships within each of their principal components, as noted above, suggested that application of a Logit Boosted Random Forest would enable effective classification.

To understand why Random Forests were considered as an appropriate next step, consider an example variable with a range of 0 to 1. Class 1 may have values of less than 0.25 or greater than 0.75, whereas class 2 may have a range of 0.25 to 0.75. The mean of the two classes would be indistinguishable, but the variance would be different. Therefore, one may conclude that this variable is not very useful for prediction. However, placing

two binary splits on that data (at 0.25 and 0.75) will yield a very strong classifier. Now, if one considers 6 variables in a high-dimensional space, finding similar split points through visual inspection would be difficult, if not impossible. Instead, all 6 can be ingested into the Logit Boosted Random Forest to leverage all potential partitions. If a variable provides no predictive power, then very few split points will occur on that variable and the performance should not be impacted.

After applying the Logit Boosted Random Forest algorithm (Algorithm 1), cross validation reflected a 75% accuracy in classifying whether the DM approach would work or not (Area under *receiver-operator curve* was 0.69). The resulting Random Forest model was then applied to all 62,000 PCA-transformed stream reaches in Iowa. The outputs were standardized on a scale from 0-1, indicating the probability that our DM algorithm would perform well at transforming NWM outputs to water levels. The final results are plotted for all stream reaches in Iowa in Figure 9, where the color blue is used to denote locations at which the DM approach is expected to perform well. It is important to note that this map covers many more streams than are measured by the 182 level sensors. As such, it should be interpreted as a map of potential future sensor sites. That is, placing a level sensor into any of the dark blue regions should correspond, on average, with a higher likelihood of successfully mapping NWM outputs to water levels using our DM approach.

## 4 Discussion

In lieu of recalibrating or expanding the complexity of a large numerical model, there may instead be immediate benefits to be gained by using sensor data to "learn" how larger-scale model outputs map to site-level conditions. To start, at approximately 30 of the 180 sites, a strong flow-to-height relationship already existed. Some of those sites were located close to USGS gages, which are assimilated into the NWM. For instance, Site 1 (Figure 5) is located only a few hundred meters from a USGS gage. Due to direct assimilation, the numerical model is likely to represent the nuanced flow dynamics more accurately at these locations, which leads to more reliable rating curves. In these instances, even a simple regression would have sufficed to predict local water levels. Naturally, our dynamical mapping approach performed well in all of these cases, too, since it can be generalized as a linear transformation [*Ljung*, 1987].

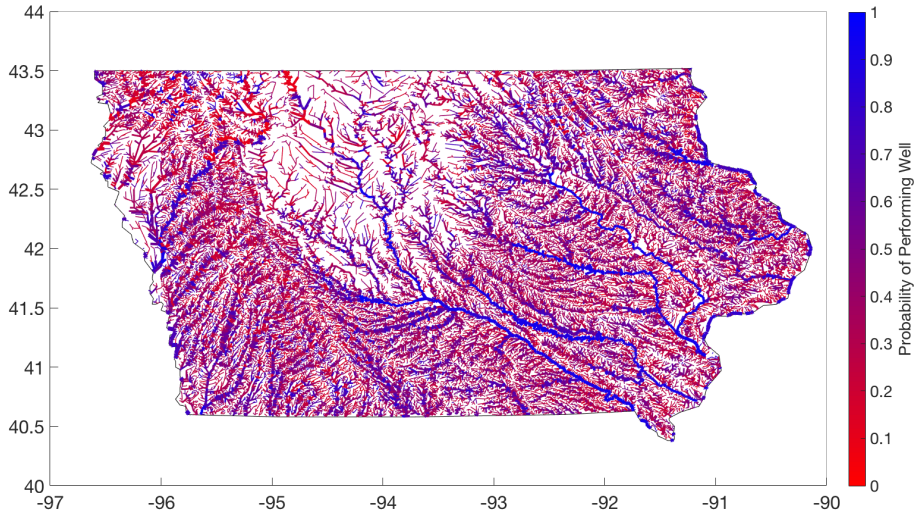



**Figure 9.** Map of site performance potential across the state of Iowa, showing a spectrum of locations where the dynamical mapping approach is expected to perform well in predicting local water levels from flows (blue) to those where it will likely not perform well (red).

While a simple regression may work in some cases, the number of instances where it can be used is fairly small. By comparing modeled flows from the NWM to measured water levels, our analysis demonstrated that these mappings are often not straightforward. Given the lack of a clear one-to-one mapping, a regression-based approach, or one that is based on simple physical equations, may not perform well because it does not account for the temporal transformation of the input signal. As such, a major benefit of our approach relates to its ability to make predictions when modeled values and local measurements do not exhibit a clear point-to-point relationship. This is particularly evident in cases where site-scale dynamics were accurately reconstructed despite the fact that large-scale NWM outputs appeared like a rapid set of impulses (Figure 6). To this end, a dynamical mapping, parameterized through system identification, shows promise as a general tool to transform modeled values to more accurate local predictions.

Our specific case study of the NWM reveals a number of generalizable requirements for the dynamical mapping to work well. Regardless of model- or site-specific dynamics, the modeled values and sensor measurements should generally agree in relative magnitude and timing. In other words, if the modeled flows show an increase over a period of time, a corresponding rise in water levels should be measured as well. This could occur irrespec-

tive of specific dynamical features. Namely, even if the modeled values appear as a set of sudden impulses, they can be adequately mapped to the more continuous in-situ sensor values if a sufficient level of agreement exists between the two times series. In the case of a hydrologic model, when using routing procedures like Muskingum-Cunge, particularly in headwater areas, it is not uncommon for flows to be modeled as "flashy" or as a series of brief spikes. While the physical model may not be designed to account for nuanced site-level dynamics, it may, in fact, be routing the mass of water correctly. In such cases, our approach can be used to represent these site-level dynamics by relying on the ability of the larger model to explain the underlying inputs. This is quite powerful, as it suggests that in many cases the site-level complexity can be explained without changing much, if anything, about the larger underlying numerical model. Rather, it may often be possible to rely on local sensor data to explain how modeled values are transformed to local observations.

Our classification analysis brings to bear under which conditions the DM approach may not perform well. In fact, at over two thirds of the evaluated sites our approach did not perform well in mapping NWM flows to local water levels, as quantified by the 50% NSE criterion. This may not necessarily be a limitation of the actual approach, but rather an indicator that the approach will improve as the physical model becomes more generally representative of local flows. In many cases, there was simply a general lack of temporal agreement between the numerical model and the measured data, with many instances of false positives and false negatives (e.g Figure 7). There were many instances during which the NWM predicted a change in flows, while no change in heights was ever measured. Similarly, many sensors measured storms that were never seen in the NWM. Naturally, our approach will not work under these conditions, since it requires changes in the inputs to be mapped to changes in outputs. Of course, our DM approach could benefit by including additional local data (e.g. independently-made rainfall measurements), but this increases its complexity, increases implementation overhead, and decreases its generalizability. This would defeat the original goal of simply relying on a publicly-available physical model that someone else updates and maintains. To that end, we expect that the performance of the DM approach will improve as the underlying physical model is improved, which is an ongoing and promising effort within the NWM community.

A number of insights, specific to the NWM, also emerged from our performance classification. Given the size, complexity, and collinearity of the data set, we illustrated

quickly that a simple classification of performance, based on individual physiographic features, does not provide much insight (Figure 8a). One take-away, though not strongly consistent, appears to be that our dynamical mapping performs well on larger streams and rivers. This should be intuitive, since the NWM would be expected to represent larger gauged rivers more effectively than smaller upstream headwater catchments. Furthermore, Muskingum-Cunge methods have been shown to work quite well in laboratory settings, but can introduce errors in field settings that, while negligible at small scales, can have major impacts as these errors propagate [*Perumal et al.*, 2009; *Sahoo*, 2013].

While the the application of PCA removed the challenge of using correlated features to explain the performance of the DM approach, the intuitive interpretation of principal components reaches a limit quickly. To that end, our application of Logit Boosted Random Forests allowed for the creation of a map that summarizes the expected performance of our approach across all 62,000 streams in Iowa (Figure 9). This visual representation provided an intuitive means by which to assess broader performance. As expected, our DM approach is expected to perform well across the major rivers in the state (thicker lines in map). Given their size, these streams are more likely to be instrumented by USGS gauges, meaning the NWM is more likely to accurately estimate flows. Many of the remaining streams on the performance map (Figure 9) showed roughly a 50% probability of successfully applying the dynamical mapping. Most of these were characterized by a mid-level stream order. These streams are likely more sensitive to local precipitation dynamics, which may not be captured by the MRMS precipitation product used by the NWM. At finer resolutions, the noise in the MRMS estimates may have a greater impact on the overall accuracy since feeding noisy observations into a non-linear model may amplify errors. As water is routed through the system, the spatiotemporal accuracy of the precipitation estimates likely has less of an impact as the overall volume is correct. This suggests that improved precipitation inputs have the potential to dramatically improve the accuracy of the NWM at higher resolutions, which should, in turn, improve the performance of our dynamical mapping.

Given its impressive extent, the NWM already shows great promise to provide high-resolution forecasts. Increasing the resolution, parameterization, and complexity of the underlying numerical model is one way of reaching the ultimate goal of high resolution local forecasts. Alternatively, as our case study demonstrated, the existing model may already be very strong in many locations, but its outputs just have to be mapped to site-specific

features using locally-available sensor data and a suitable mathematical transform. In other words, outputs from NWM, though still in their early stages, can be useful in estimating highly-local water levels *even now*. Nonetheless, our results may also provide a guide to help improve the numerical model. The map in Figure 9 intuitively conveys a general assessment of the performance of the underlying numerical model. Since the NWM is a relatively new model, it would be expected to initially perform well at larger scales. Even with this general trend, there are still lower-order streams on the map that suggest the possibility of successfully applying our DM approach. These red and purple regions on the map (0-50% chance of applying the dynamical transformation) may be of interest to modelers as locations at which the numerical model could be improved to reduce false positive and negative forecasts. Improving the model on these stream sections will likely also translate to better model performance on stream reaches that share similar physiographic or PCA-transformed features.

From a water management perspective, the benefits of our DM approach may already be realizable operationally. This is true for a number of already existing sensor locations, as well as potentially other similar streams on the map in Figure 9. A simple web-service application [*Wong and Kerkez*, 2016] could be written to extract NWM outputs and fuse them with local sensor data. If the dynamical mapping is reliable at this location, the site would benefit immediately from a localized water level forecast. Alternatively, if local measurements are not available, the map in Figure 9 could be used to deploy low-cost sensors at locations that maximize the probability of using the DM approach. Given the general structure and input data of the NWM across the US, we also anticipate that similar maps could be created for regions outside of Iowa by relying on the results from this study. For the approach to become operational, a moving training window may be needed to account for varying hydrologic regimes or seasons (e.g. spring vs summer). The effect of hydrological regimes was not evaluated as part of our case study due to the time required to log data from the NWM and level sensors. Fortunately, the training window needed for fitting the DM model is quite small compared to that of a rating curve, which will allow for effects of seasonality to be evaluated in the future.

## 5 Conclusions

In this paper, we provided a means by which outputs from a large-scale model can be fused with local sensor data to provide site-level forecasts. The novelty of the approach

relies on using the outputs of the physical models as the inputs into a dynamical mapping that *learns* what a specific sensor will measure. This is quite powerful, as it does not rely on the modification of the actual physical model or the direct assimilation of the sensor data, both of which would be infeasible for smaller communities. Instead, the approach is general, in that it can be directly repeated for any combination of sensor-model pairs. As such, the approach developed here could be applied directly without any modification of our open-source code. While the approach will not work under all conditions, it may already provide an immediate benefit to a large number of locations.

In the age of *Big Data in Hydrology* we contend that even models can be viewed as just one of many streams of data that will enable decision making. Overall, the approach of dynamically mapping outputs form large models to local sites may work for a number of models beyond just the NWM. The ability to use the approach with short data histories (e.g. only a few months of training data) makes it appealing for urban applications, where land use changes may occur rapidly and system re-identification may need to occur frequently. In such cases, our approach could be combined with popular urban water models, such as the stormwater management model (SWMM) to provide improved forecasts of urban flooding or sewer flows. More examples can be given, but we anticipate that the our data-driven approach could be generalized for many hydrologic and hydraulic models.

## A: Principal Component Analysis

The goal of PCA is to find the weighting vectors, or *principal components*, that yield linear combinations of the original feature space. We define $\mathbf{X} \in \mathbb{R}^{n \times d}$ as the data matrix with $n$ rows of observations and $d$ features, which in our case is populated with the physiographic features of the nearly 62,000 stream reaches in Iowa. Before PCA is applied, all input features also need to be standardized in magnitude to reduce impacts of overweighting some features over others [*Hastie et al.*, 2001]. By standardizing across each variable, one can consider the relative impacts of each more effectively.

To find the first principal component, $\mathbf{w}_1$, we find a unit vector that maximizes the variance of $\mathbf{X}$, that is:

$$\mathbf{w}_1 = \underset{||\mathbf{w}||=1}{\arg\max} ||\mathbf{X}\mathbf{w}||^2 = \arg\max \frac{\mathbf{w}^T \mathbf{X}^T \mathbf{X} \mathbf{w}}{\mathbf{w}^T \mathbf{w}} \tag{A.1}$$

This is a Rayleigh quotient [*Horn and Johnson*, 1990], and therefore the solution to this maximization problem is the largest eigenvector (i.e. the eigenvector of the largest eigen-

value) of $\mathbf{X}^T\mathbf{X}$. Each successive principal component is the next largest eigenvector of $\mathbf{X}^T\mathbf{X}$. Therefore, rather than solving iteratively for each principal component, it is possible to consider the singular value decomposition (Equation A.2) of the data matrix $\mathbf{X}$:

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{W}^T \tag{A.2}$$

$$\mathbf{X}^T\mathbf{X} = \mathbf{W}\Sigma\mathbf{U}^T\mathbf{U}\Sigma\mathbf{W}^T = \mathbf{W}\Sigma^2\mathbf{W}^T \tag{A.3}$$

As such, the eigenvectors of $\mathbf{X}^T\mathbf{X}$ are the rows of $\mathbf{W}$, meaning the principal components are the right singular vectors of our data matrix. Therefore, to place our data in an orthogonal feature space such that all the variables are de-correlated, the new data matrix, $\mathbf{T}$, is simply:

$$\mathbf{T} = \mathbf{X}\mathbf{W} \tag{A.4}$$

Using this matrix will lead to a more stable classification procedure, will reduce the likelihood of over fitting, and will enable more complex interactions between features to be captured [*Hastie et al.*, 2001].


## Acknowledgments

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 1256260. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We would also like to thank F. Salas for input regarding the National Water Model. We would also like to acknowledge the Iowa Flood Center and its Flood Information System, whose publicly available data made this research possible on such a large scale. All code may be found at https://github.com/kLabUM/NWM while all data can be found at *Fries and Kerkez* [2017].


## References


Aricò, C., G. Corato, T. Tucciarelli, M. B. Meftah, A. F. Petrillo, and M. Mossa (2010), Discharge estimation in open channels by means of water level hydrograph analysis, *Journal of Hydraulic Research*, *48*(5), 612–619, doi:10.1080/00221686.2010.507352.

Bartos, M. D., B. Wong, and B. Kerkez (2017), Open storm: a complete framework for sensing and control of urban watersheds, *Environ. Sci.: Water Res. Technol.*, pp. –, doi: 10.1039/C7EW00374A.

Beck, M. B. (1987), Water quality modeling: A review of the analysis of uncertainty, *Water Resources Research*, *23*(8), 1393–1442, doi:10.1029/WR023i008p01393.

Bitella, G., R. Rossi, R. Bochicchio, M. Perniola, and M. Amato (2014), A novel low-cost open-hardware platform for monitoring soil water content and multiple soil-air-vegetation parameters, *Sensors*, *14*(10), 19,639–19,659, doi:10.3390/s141019639.

Bjorck, A. (1996), *Numerical Methods for Least Squares Problems*, Society for Industrial and Applied Mathematics, doi:10.1137/1.9781611971484.

Blöschl, G., A. Bárdossy, D. Koutsoyiannis, Z. W. Kundzewicz, I. Littlewood, A. Montanari, and H. Savenije (2014), On the future of journal publications in hydrology, *Water Resources Research*, *50*(4), 2795–2797, doi:10.1002/2014WR015613.

Chang, N.-B., K. Bai, and C.-F. Chen (2017), Integrating multisensor satellite data merging and image reconstruction in support of machine learning for better water quality management, *Journal of Environmental Management*, *201*(Supplement C), 227 – 240, doi:https://doi.org/10.1016/j.jenvman.2017.06.045.

Cluckie, I., and R. Harpin (1982), A real-time simulator of the rainfall-runoff process, *Mathematics and Computers in Simulation*, *24*(2), 131 – 139, doi: http://dx.doi.org/10.1016/0378-4754(82)90095-7.

Damangir, H., and M. Abedini (2014), System identification and subsequent discharge estimation based on level data alone–Gradually varied flow condition, *Flow Measurement and Instrumentation*, *36*, 24 – 31, doi: http://dx.doi.org/10.1016/j.flowmeasinst.2014.01.002.

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

Deza, M., and E. Deza (2009), *Encyclopedia of Distances*, Encyclopedia of Distances, Springer Berlin Heidelberg.

Freund, Y., R. E. Schapire, et al. (1996), Experiments with a new boosting algorithm, in *International Conference on Machine Learning*, vol. 96, pp. 148–156.

Fries, K., and B. Kerkez (2017), Data for "Using sensor data to dynamically map large-scale models to site-scale forecasts: A case study using the National Water Model", doi: 10.7302/Z2PN93TZ.

Gilles, D., N. Young, H. Schroeder, J. Piotrowski, and Y.-J. Chang (2012), Inundation Mapping Initiatives of the Iowa Flood Center: Statewide Coverage and Detailed Urban

Flooding Analysis, *Water*, *4*(1), 85–106.

Hastie, T., R. Tibshirani, and J. Friedman (2001), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer series in statistics, Springer.

Herschy, R. (1999), *Hydrometry: Principles and Practice*, Wiley.

Hidayat, H., B. Vermeulen, M. Sassi, P. Torfs, and A. Hoitink (2011), Discharge estimation in a backwater affected meandering river, *Hydrology and Earth System Sciences*, *15*, 2717–2728.

Hoke, J. E., and R. A. Anthes (1976), The Initialization of Numerical Models by a Dynamic-Initialization Technique, *Monthly Weather Review*, *104*(12), 1551–1556, doi: 10.1175/1520-0493(1976)104<1551:TIONMB>2.0.CO;2.

Horn, R., and C. Johnson (1990), *Matrix Analysis*, Cambridge University Press.

Iowa Flood Center (2017), IFIS Web Service. http://ifis.iowafloodcenter.org/ifis/en/ws/.

Javaheri, A., M. Nabatian, E. Omranian, M. Babbar-Sebens, and S. J. Noh (2018), Merging real-time channel sensor networks with continental-scale hydrologic models: A data assimilation approach for improving accuracy in flood depth predictions, *Hydrology*, *5*(1), doi:10.3390/hydrology5010009.

Jin, N., R. Ma, Y. Lv, X. Lou, and Q. Wei (2010), A novel design of water environment monitoring system based on WSN, in *2010 International Conference On Computer Design and Applications*, vol. 2, pp. V2–593–V2–597, doi:10.1109/ICCDA.2010.5541305.

Karandish, F., and J. Šimůnek (2016), A comparison of numerical and machine-learning modeling of soil water content with limited input data, *Journal of Hydrology*, *543*(Part B), 892 – 909, doi:https://doi.org/10.1016/j.jhydrol.2016.11.007.

Krause, P., D. P. Boyle, and F. Bäse (2005), Comparison of different efficiency criteria for hydrological model assessment, *Advances in Geosciences*, *5*, 89–97.

Kuczera, G. (1996), Correlated rating curve error in flood frequency inference, *Water Resources Research*, *32*(7), 2119–2127, doi:10.1029/96WR00804.

Ljung, L. (1987), *System identification: theory for the user*, Prentice-Hall information and system sciences series, Prentice-Hall.

Luenberger, D. (1979), *Introduction to Dynamic Systems: Theory, Models, and Applications*, Wiley.

Moradkhani, H., K.-L. Hsu, H. Gupta, and S. Sorooshian (2005), Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter, *Water Resources Research*, *41*(5), n/a–n/a, doi:10.1029/2004WR003604,

w05012.

Nash, J. (1957), The form of the instantaneous unit hydrograph, *International Association of Hydrological Sciences Publications*, *45*, 114–121.

Nash, J., and J. Sutcliffe (1970), River flow forecasting through conceptual models part i: A discussion of principles, *Journal of Hydrology*, *10*(3), 282 – 290, doi: https://doi.org/10.1016/0022-1694(70)90255-6.

National Severe Storms Laboratory (2017), Multi-radar/multi-sensor system (mrms).

Office of Water Prediction (2017), The national water model.

Ouyang, Y. (2005), Evaluation of river water quality monitoring stations by principal component analysis, *Water Research*, *39*(12), 2621 – 2635, doi: https://doi.org/10.1016/j.watres.2005.04.024.

Perumal, M., K. B. Shrestha, and U. Chaube (2004), Reproduction of hysteresis in rating curves, *Journal of Hydraulic Engineering*, *130*(9), 870–878, doi:10.1061/(ASCE)0733-9429(2004)130:9(870).

Perumal, M., B. Sahoo, T. Moramarco, and S. Barbetta (2009), Multilinear Muskingum Method for Stage-Hydrograph Routing in Compound Channels, *Journal of Hydrologic Engineering*, *14*.

Petersen-Overlier, A. (2004), Accounting for heteroscedasticity in rating curve estimates, *Journal of Hydrology*, *292*(1), 173 – 181, doi: http://dx.doi.org/10.1016/j.jhydrol.2003.12.024.

Rojas, C., M. Barenthin Barenthin, J. Welsh, and H. Hjalmarsson (2010), The cost of complexity in system identification: Frequency function estimation of finite impulse response systems, *IEEE Transactions on Automatic Control*, *55*, 2298–2309.

Sahoo, B. (2013), Field application of the multilinear muskingum discharge routing method, *Water Resources Management*, *27*(5), 1193–1205, doi:10.1007/s11269-012-0228-5.

Strahler, A. N. (1957), Quantitative analysis of watershed geomorphology, *Eos, Transactions American Geophysical Union*, *38*(6), 913–920, doi:10.1029/TR038i006p00913.

Tiwari, M. K., and J. F. Adamowski (2015), Medium-term urban water demand forecasting with limited data using an ensemble wavelet–bootstrap machine-learning approach, *Journal of Water Resources Planning and Management*, *141*(2), 04014,053, doi:10.1061/(ASCE)WR.1943-5452.0000454.

U. S. Geological Survey, and G. F. Koltun (2015), An evaluation of the accuracy of modeled and computed streamflow time-series data for the ohio river at hannibal lock and dam and at a location upstream from sardis, ohio, *Tech. rep.*, Reston, VA, doi: 10.3133/ofr20151058.

U. S. Geological Survey, and R. E. Southard (2013), Computed statistics at streamgages, and methods for estimating low-flow frequency statistics and development of regional regression equations for estimating low-flow frequency statistics at ungaged locations in missouri, *Tech. rep.*, Reston, VA.

US Geological Survey (2017), Nhdplus high resolution national hydrography dataset watershed boundary dataset.

U.S. Geological Survey, D. P. Turnipseed, and V. B. Sauer (2010), Discharge measurements at gaging stations, *Tech. rep.*

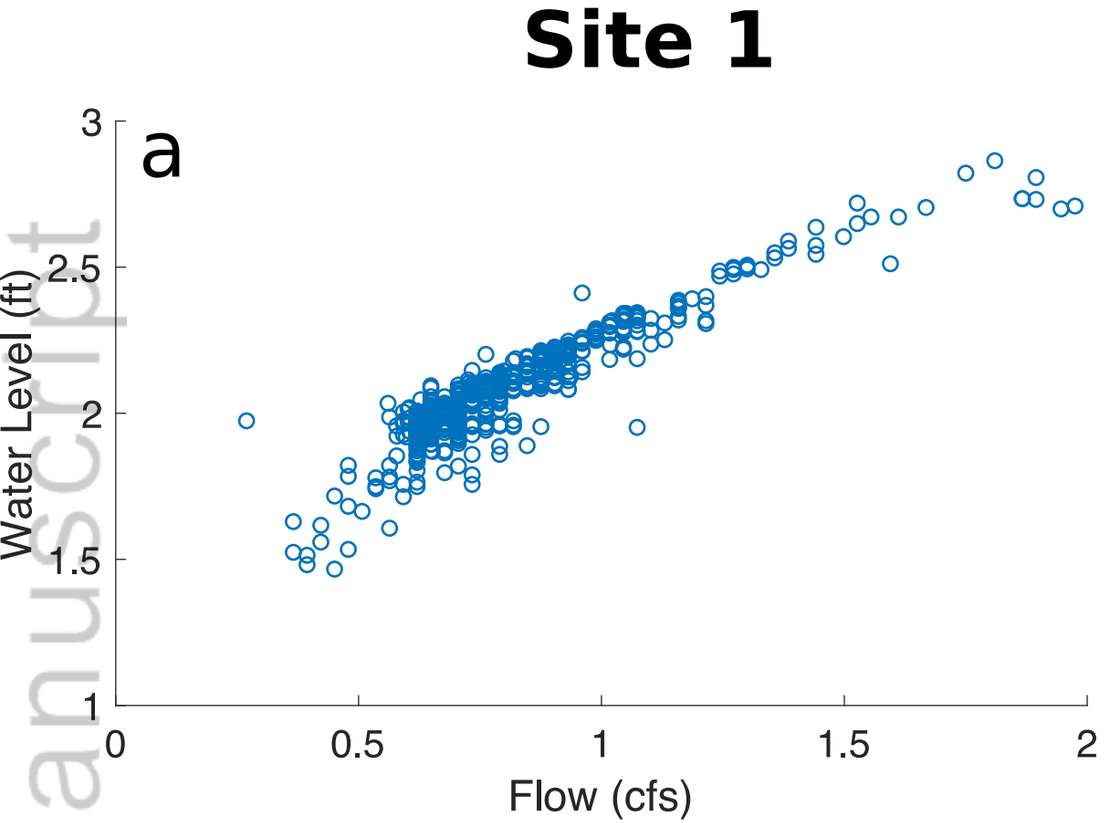
Westerberg, I., J.-L. Guerrero, J. Seibert, K. J. Beven, and S. Halldin (2011), Stage-discharge uncertainty derived with a non-stationary rating curve in the choluteca river, honduras, *Hydrological Processes*, *25*(4), 603–613, doi:10.1002/hyp.7848.

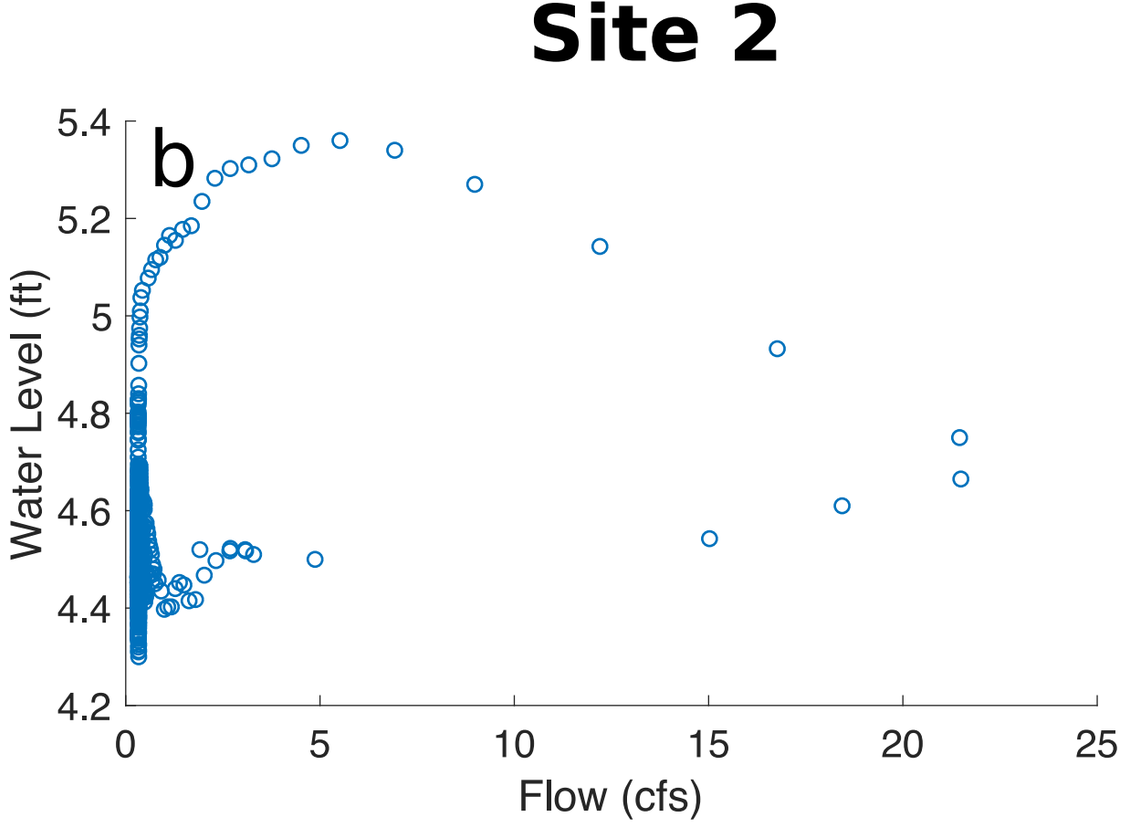
Wong, B. P., and B. Kerkez (2016), Real-time environmental sensor data: An application to water quality using web services, *Environmental Modelling & Software*, *84*, 505 – 517, doi:http://dx.doi.org/10.1016/j.envsoft.2016.07.020.

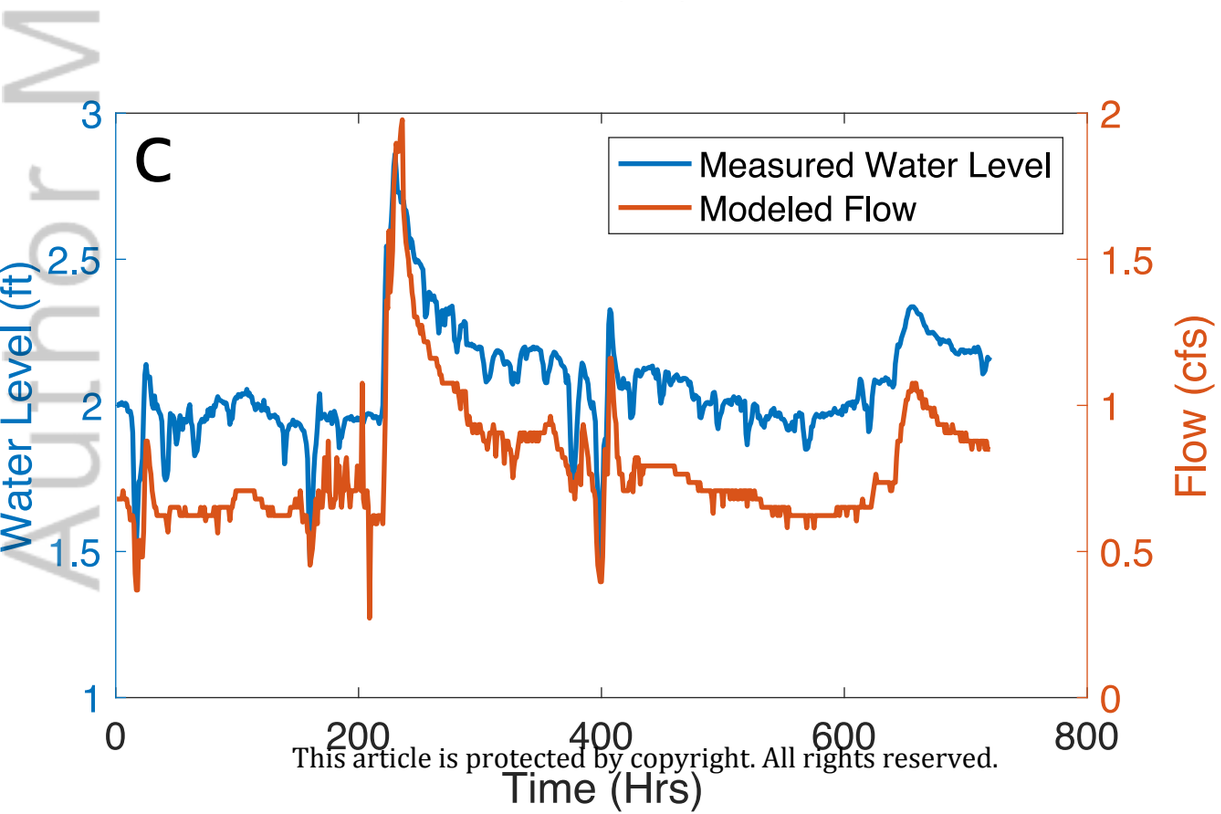
Yang, Z., and D. Han (2006), Derivation of unit hydrograph using a transfer function approach, *Water Resources Research*, *42*(1), n/a–n/a, doi:10.1029/2005WR004227, w01501.

Zhou, Z. (2012), *Ensemble Methods: Foundations and Algorithms*, Chapman & Hall/CRC Data Mining and Knowledge Discovery Series, Taylor & Francis.

**Figure 1.**





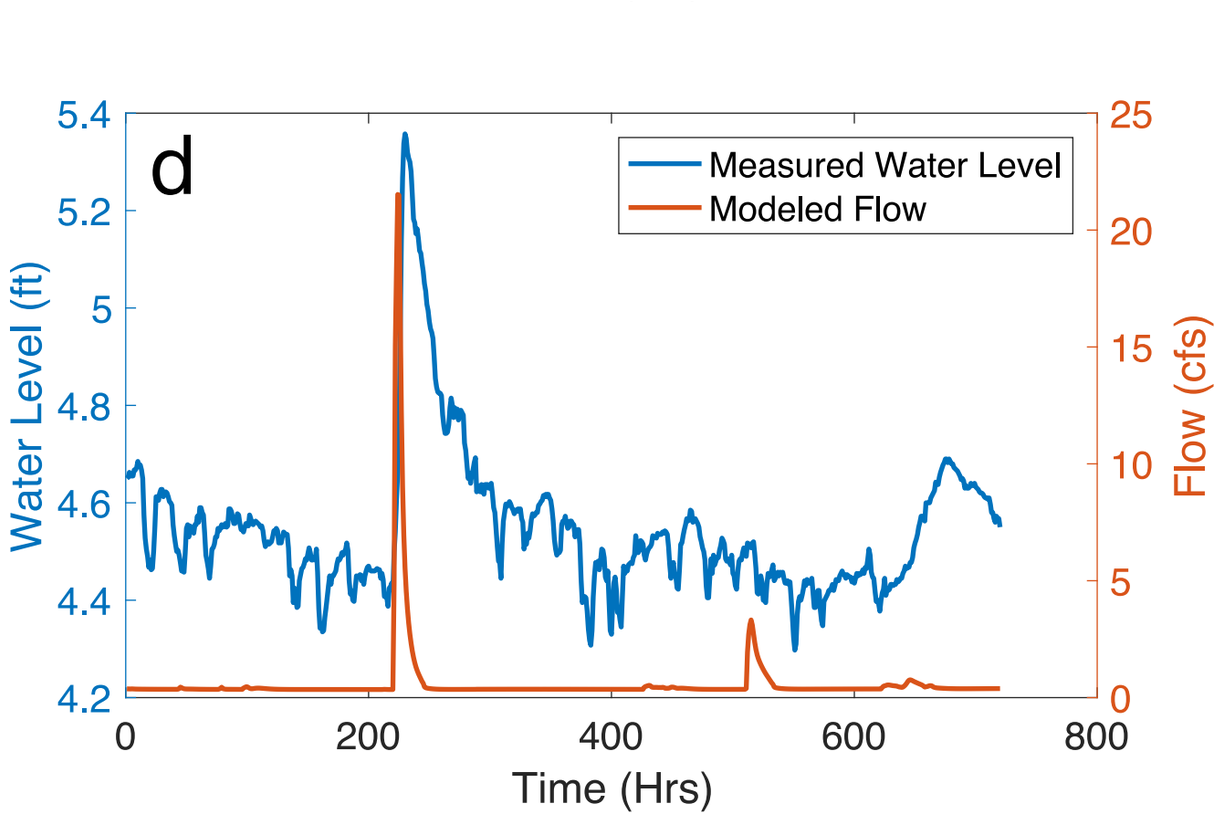



**Figure 2.**

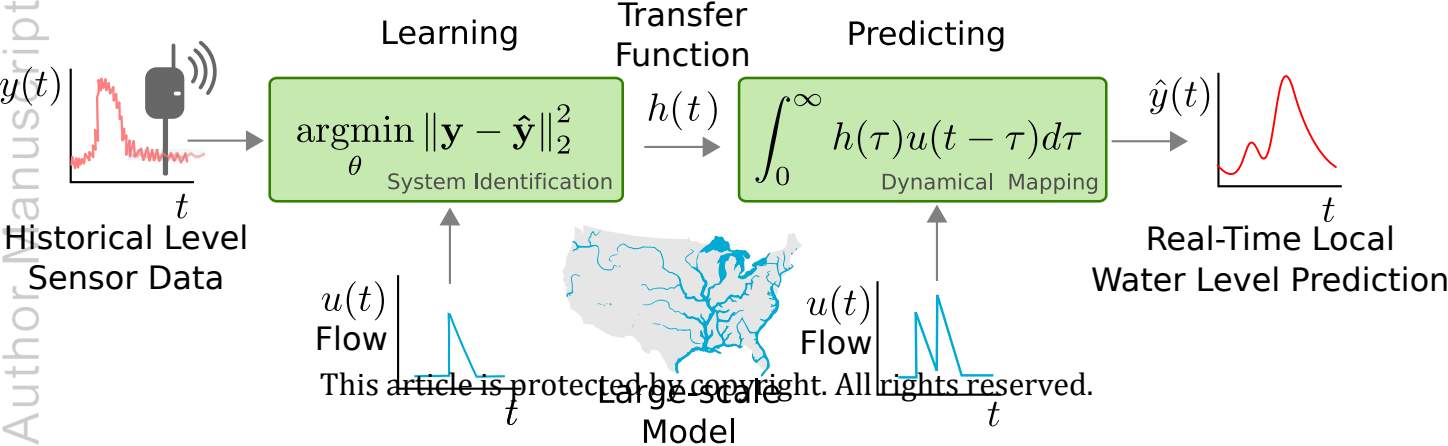


**Figure 3.**

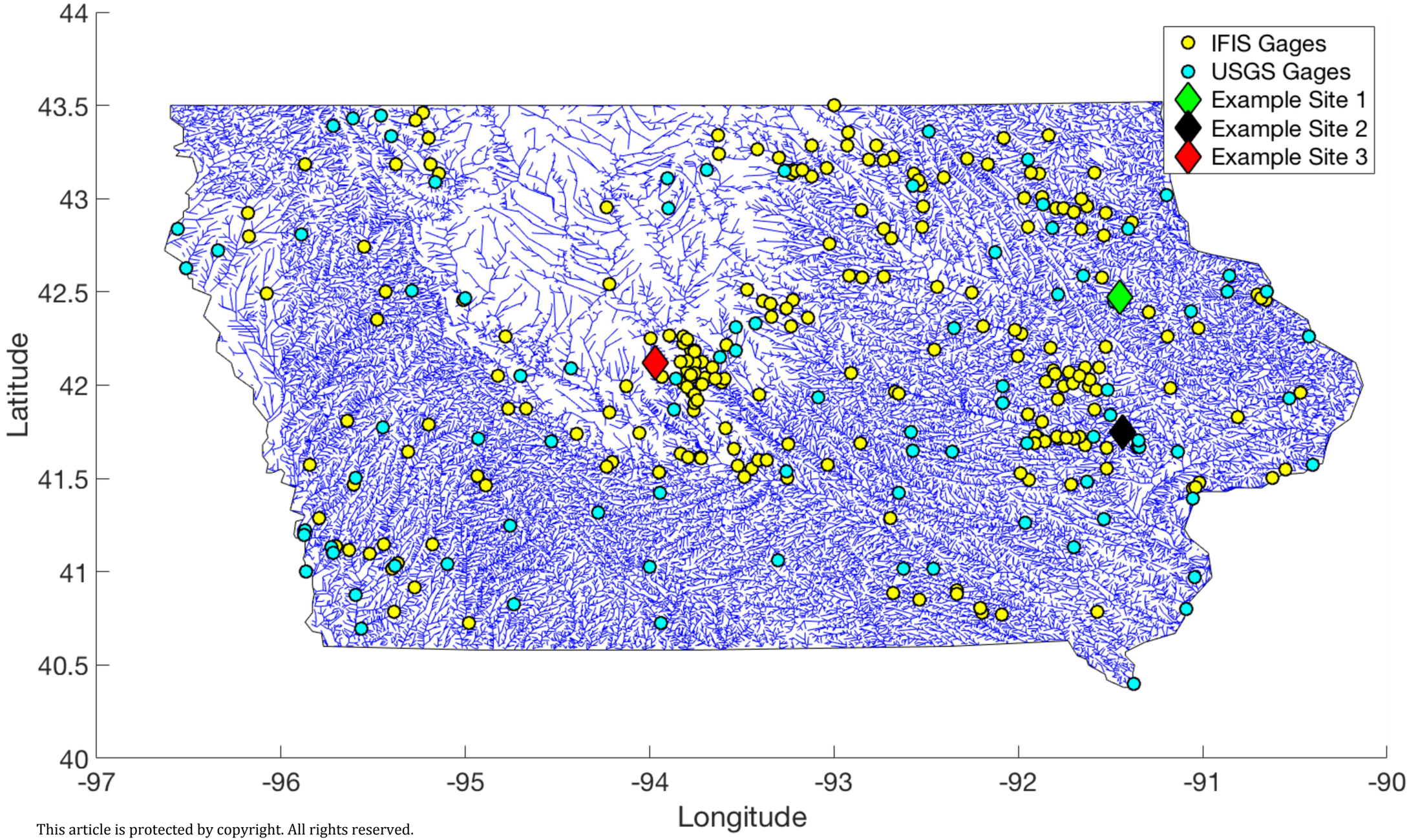


**Figure 4.**

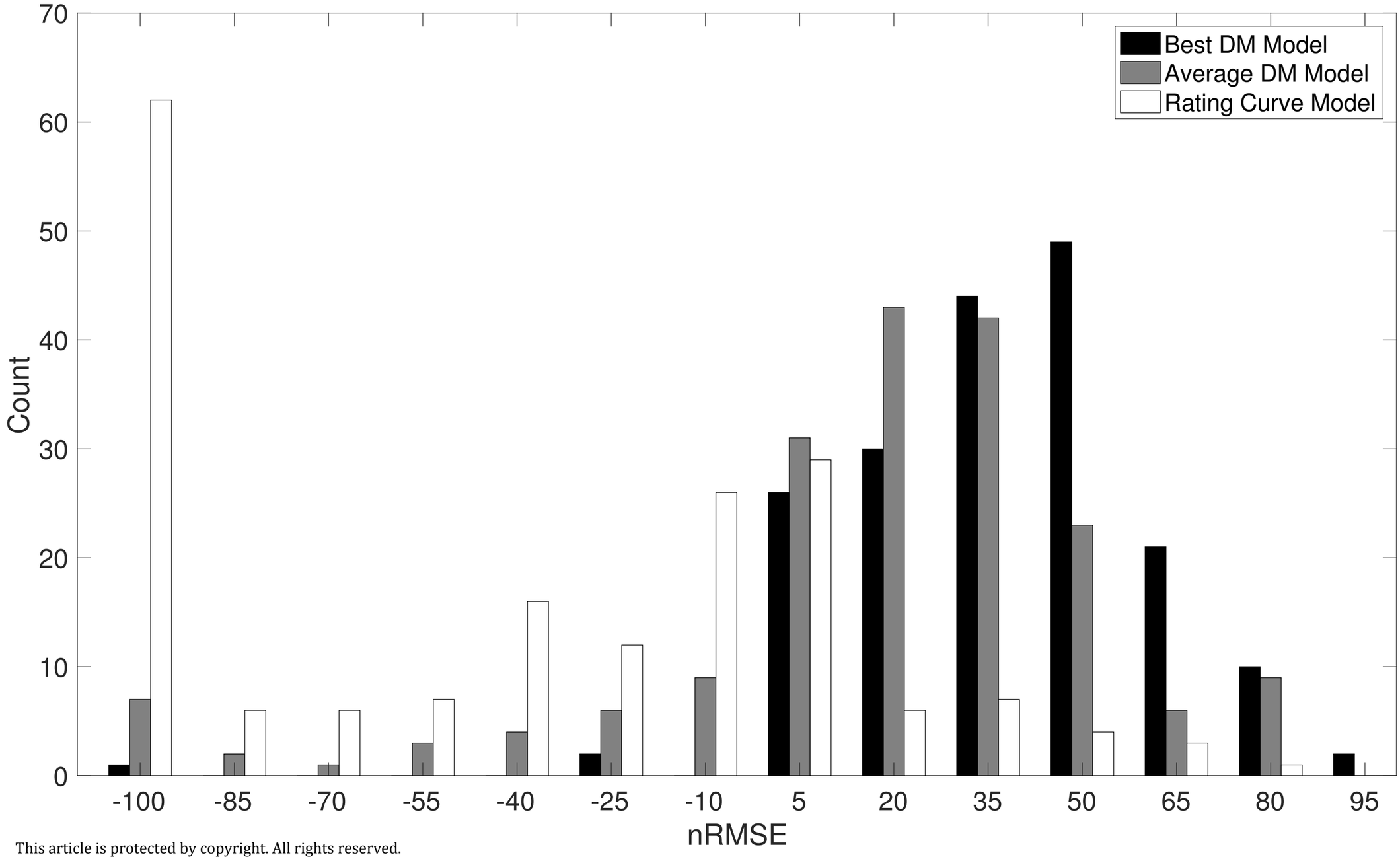
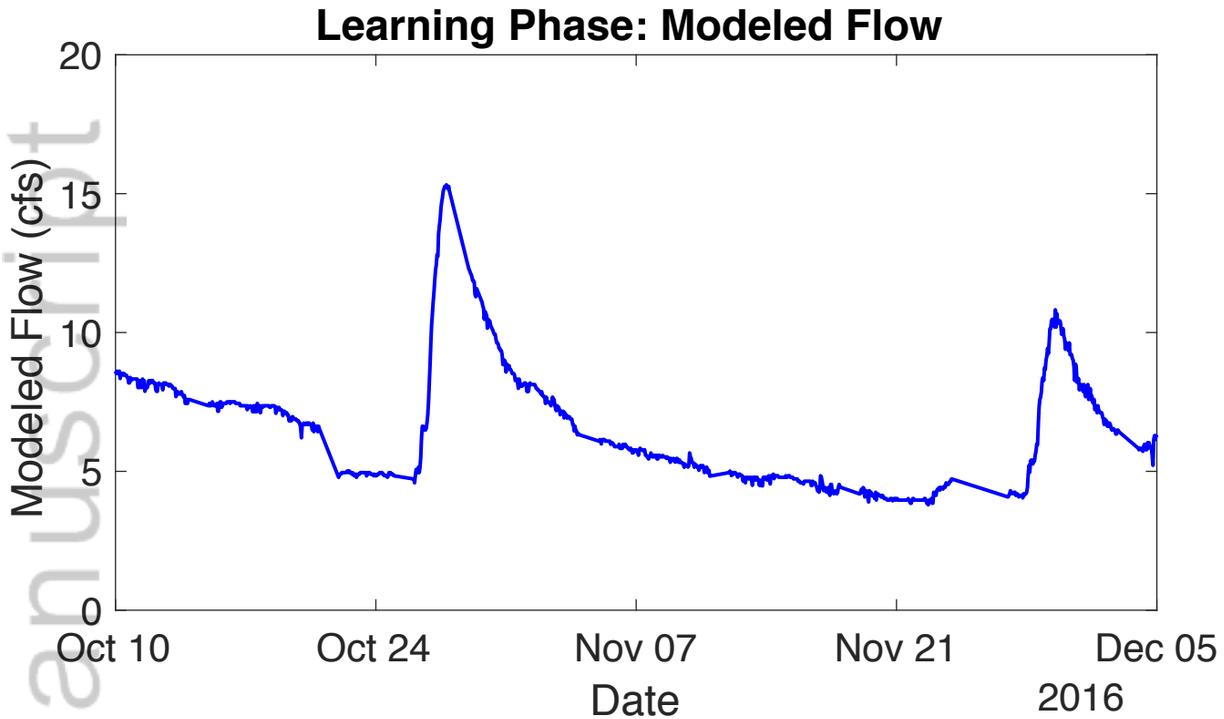
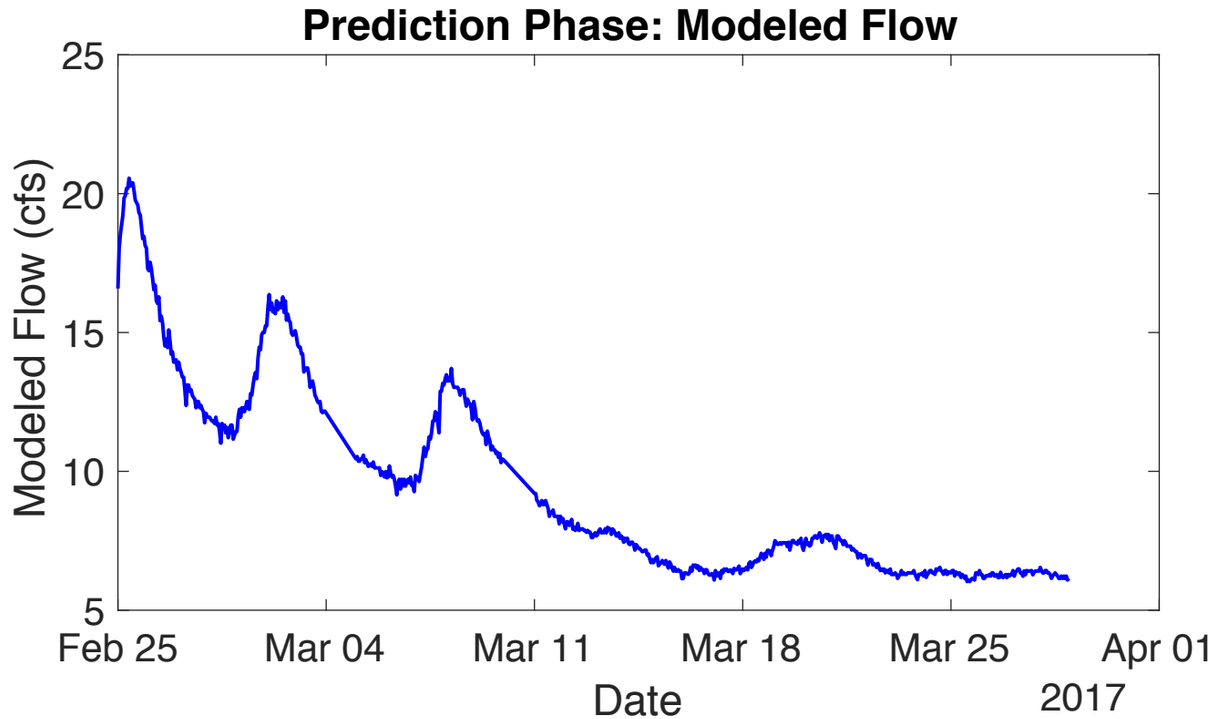
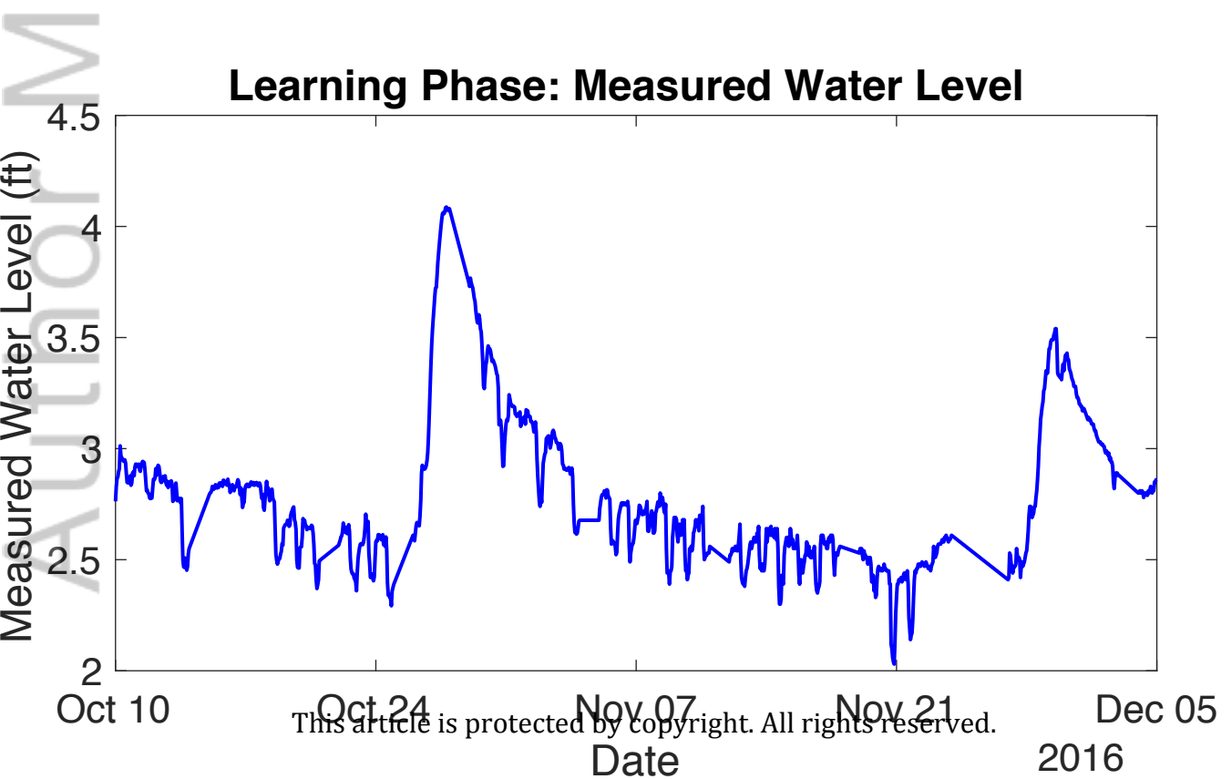


**Figure 5.**

# Dynamical Mapping: Site 1



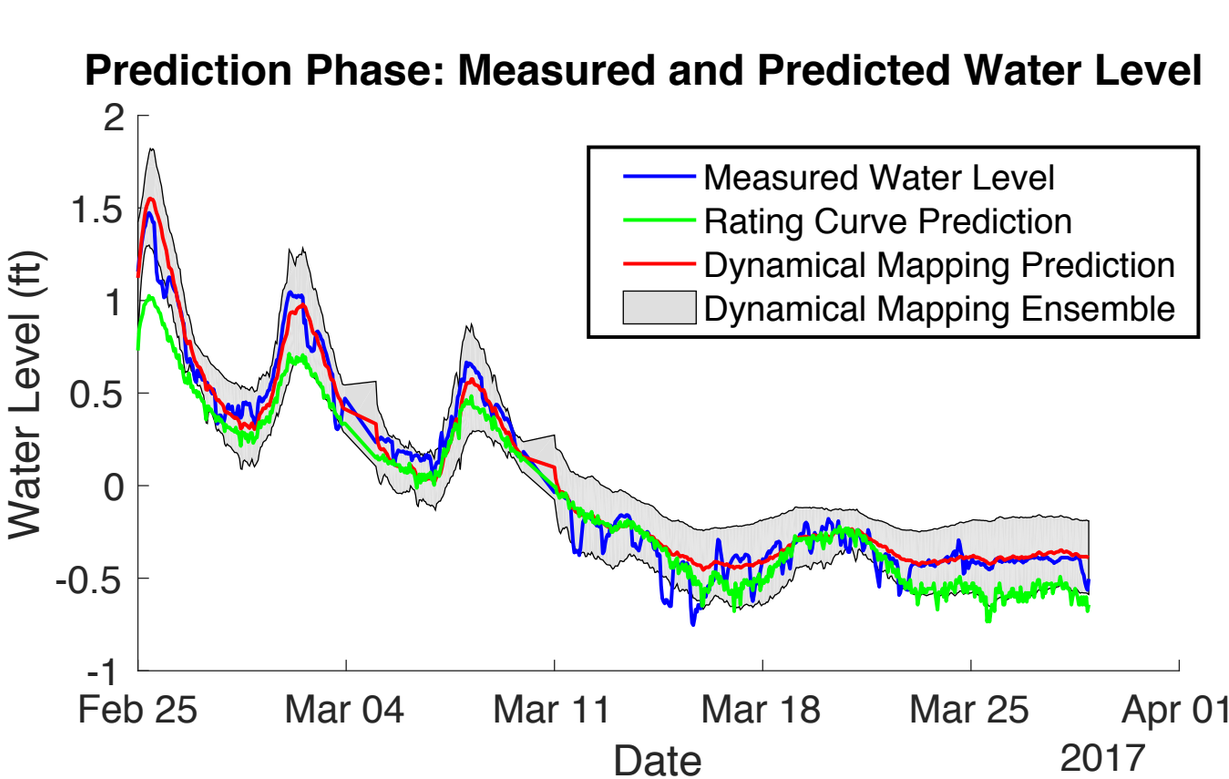
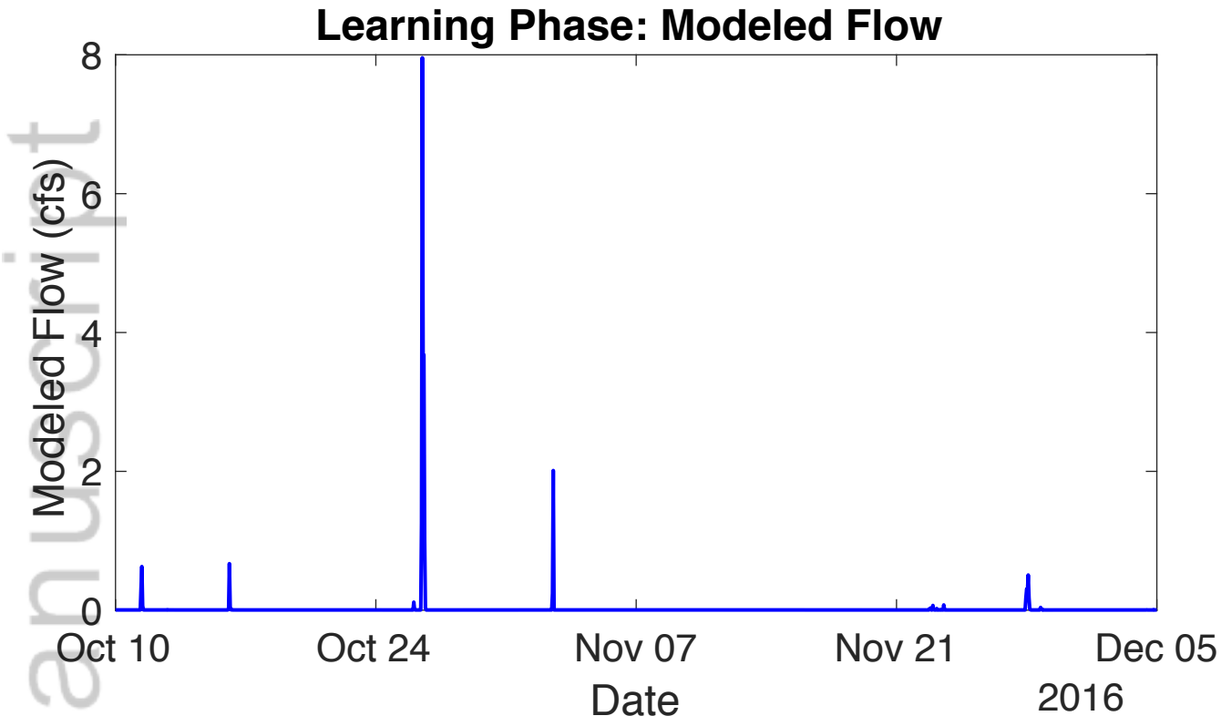
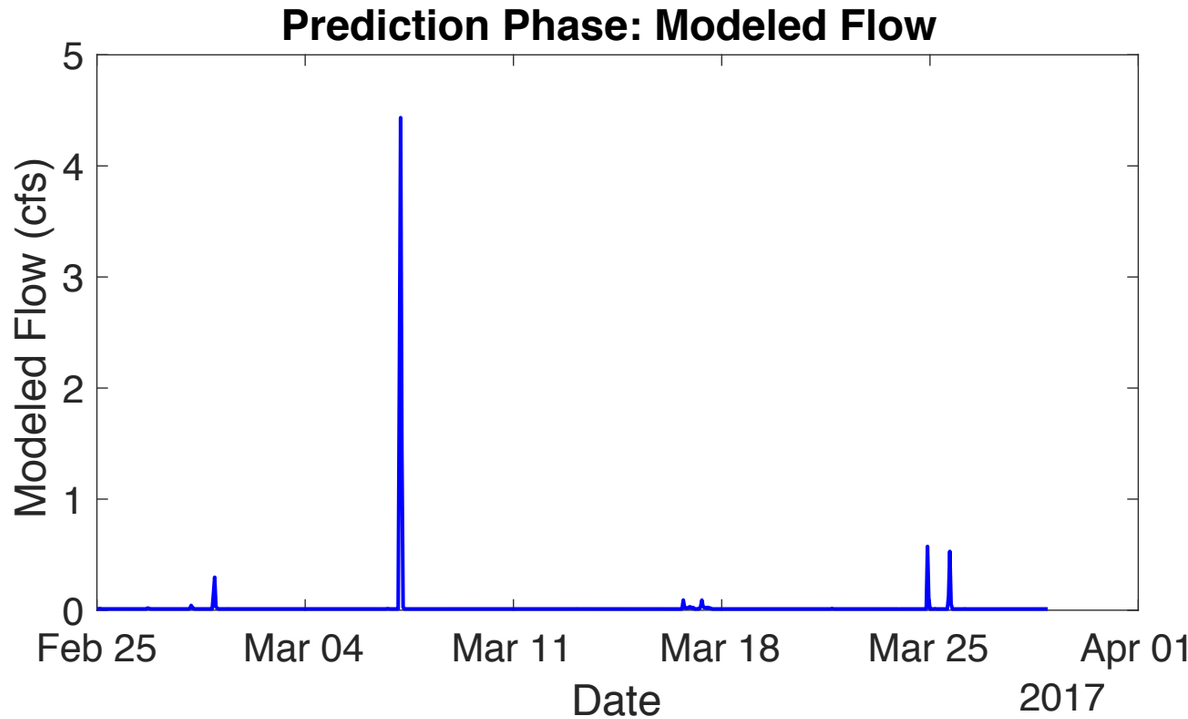
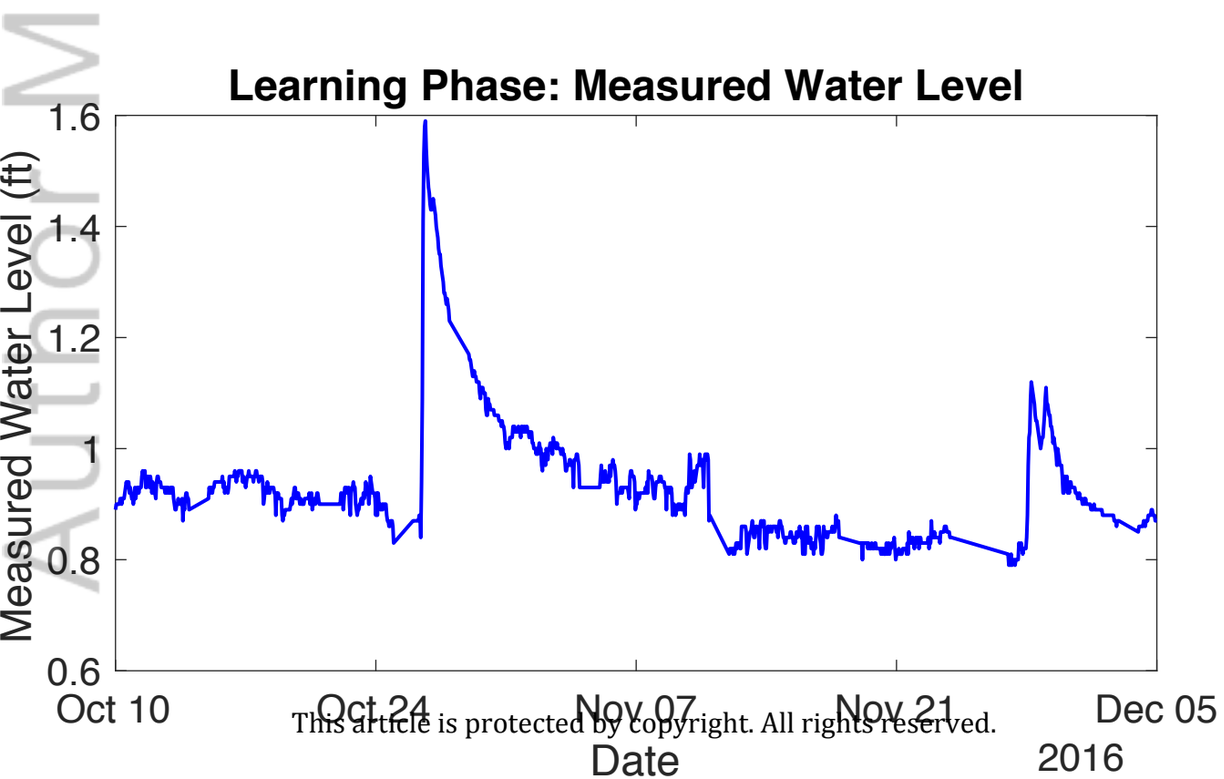



**Figure 6.**

# Dynamical Mapping: Site 2



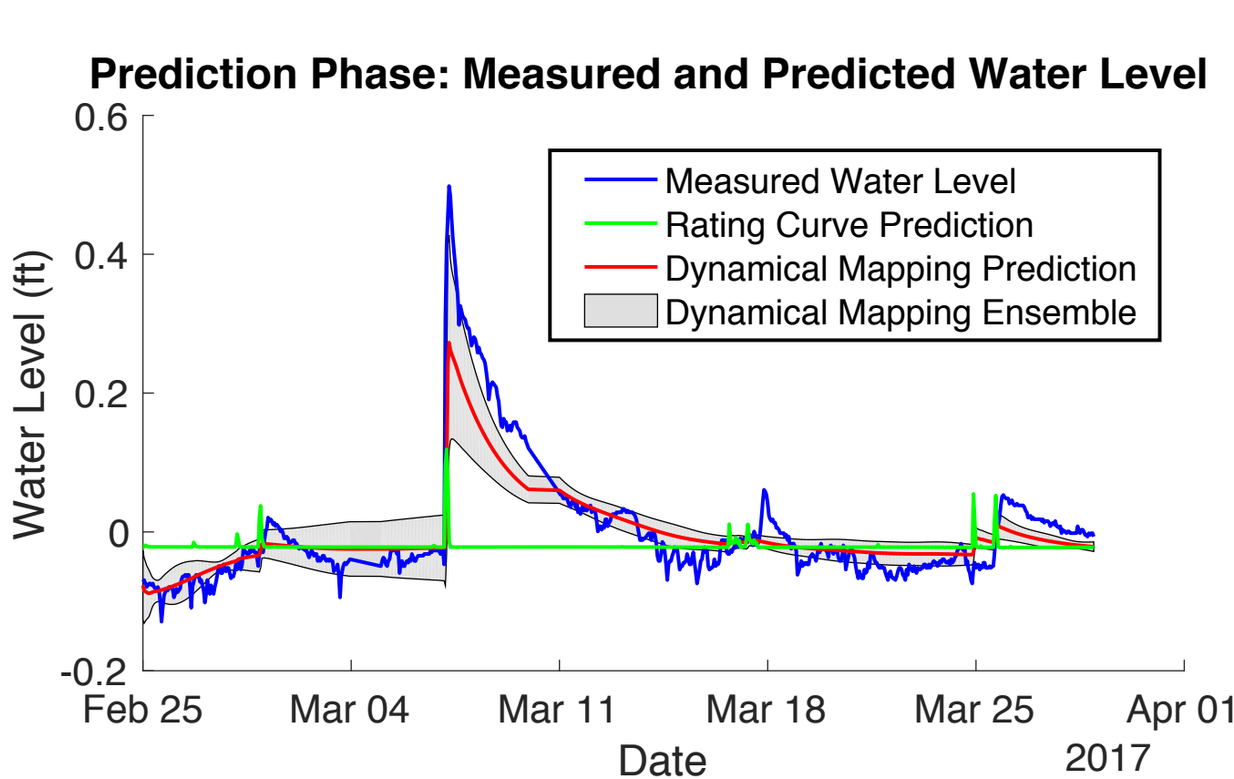
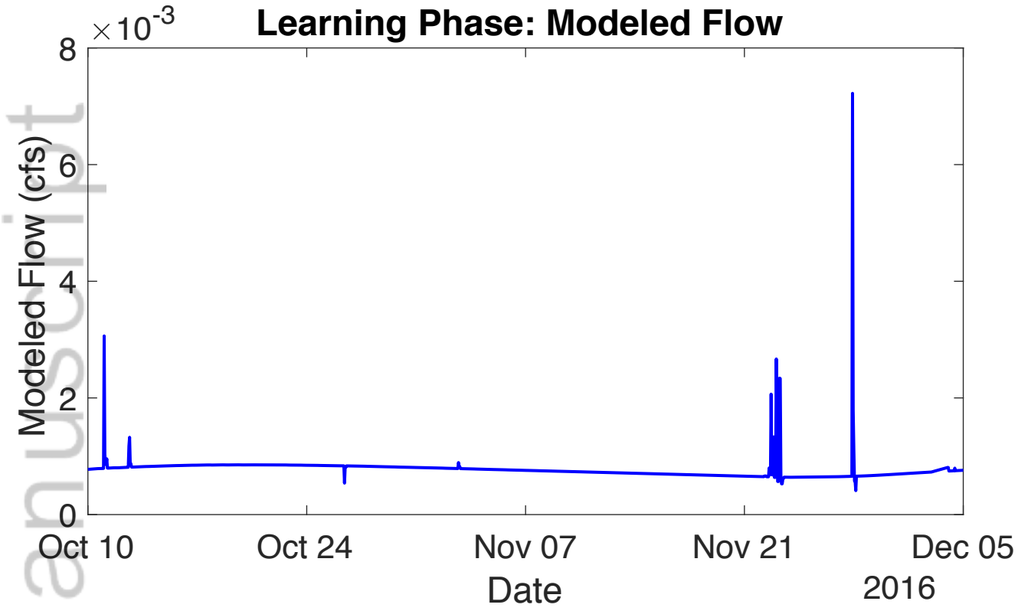
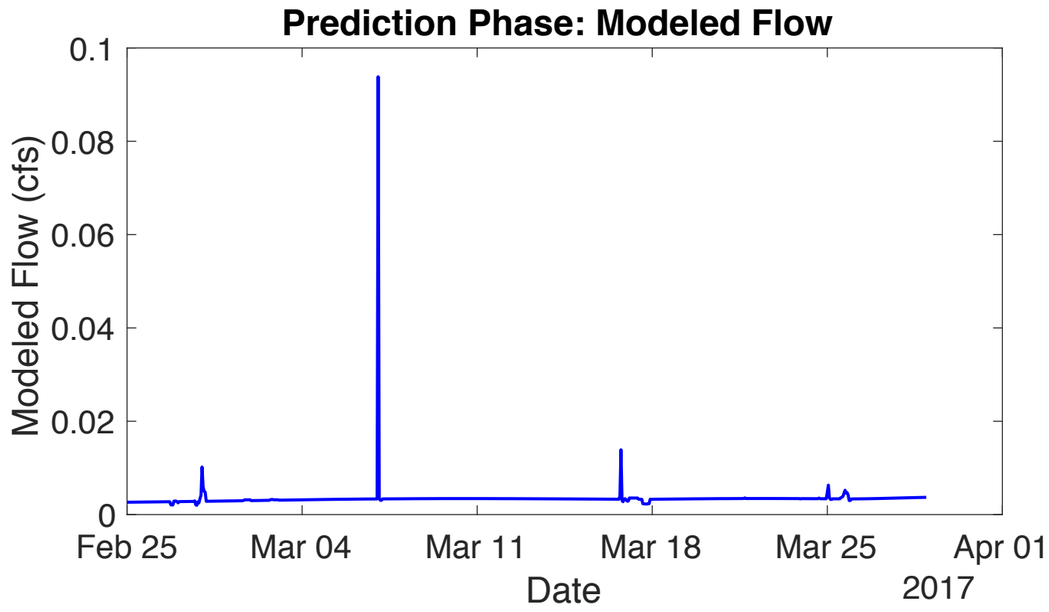
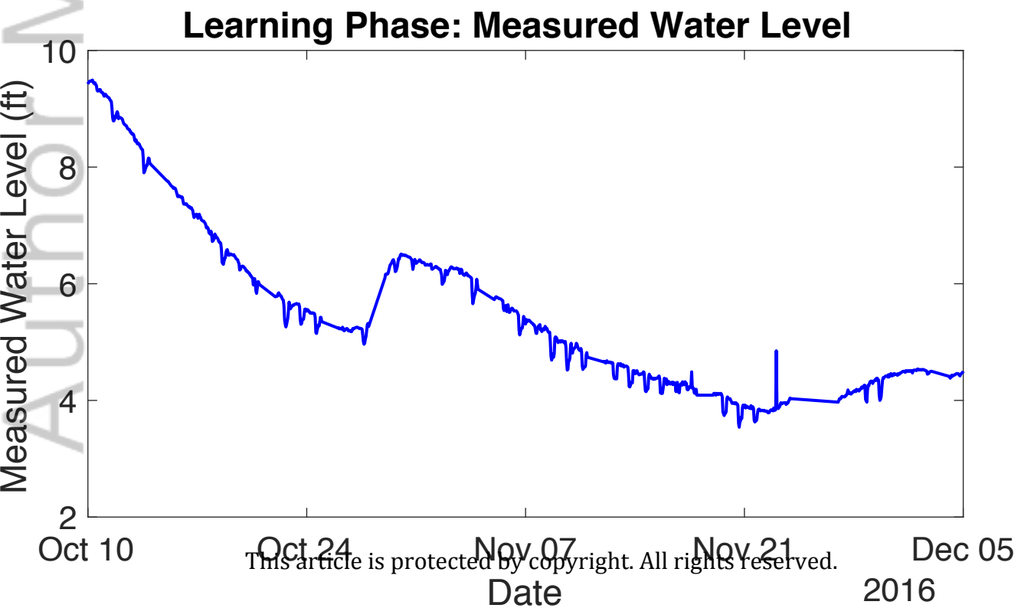



**Figure 7.**

# Dynamical Mapping: Site 3



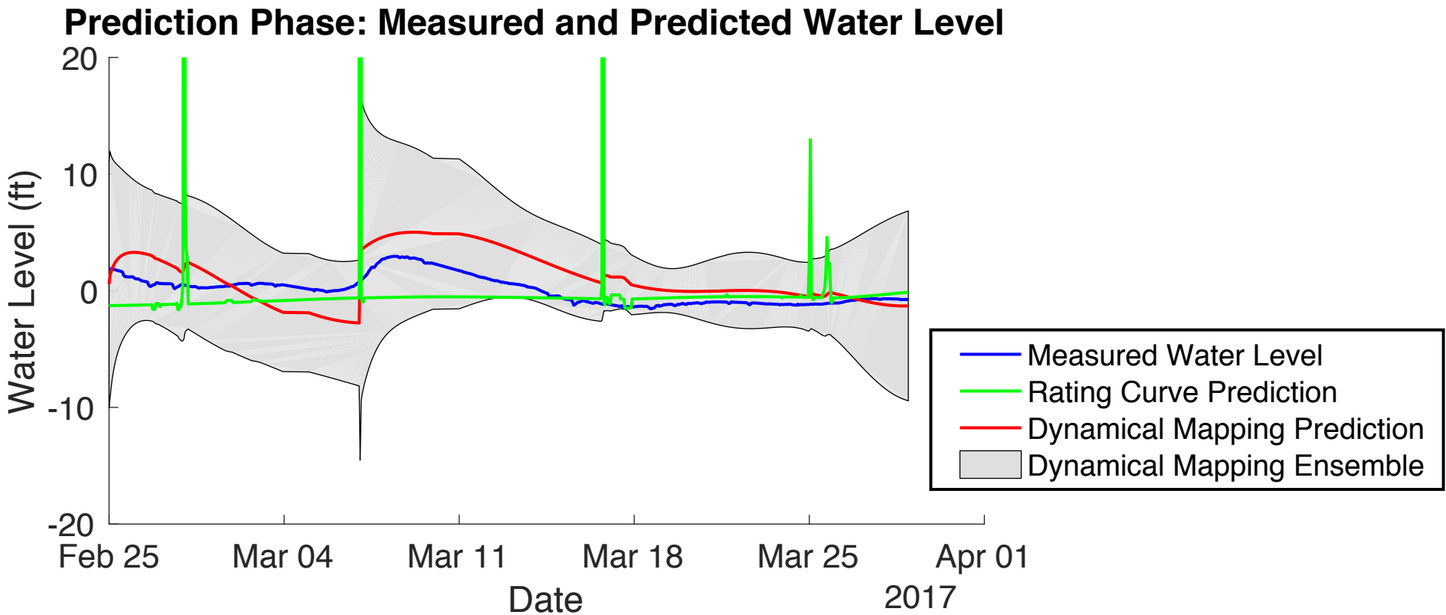
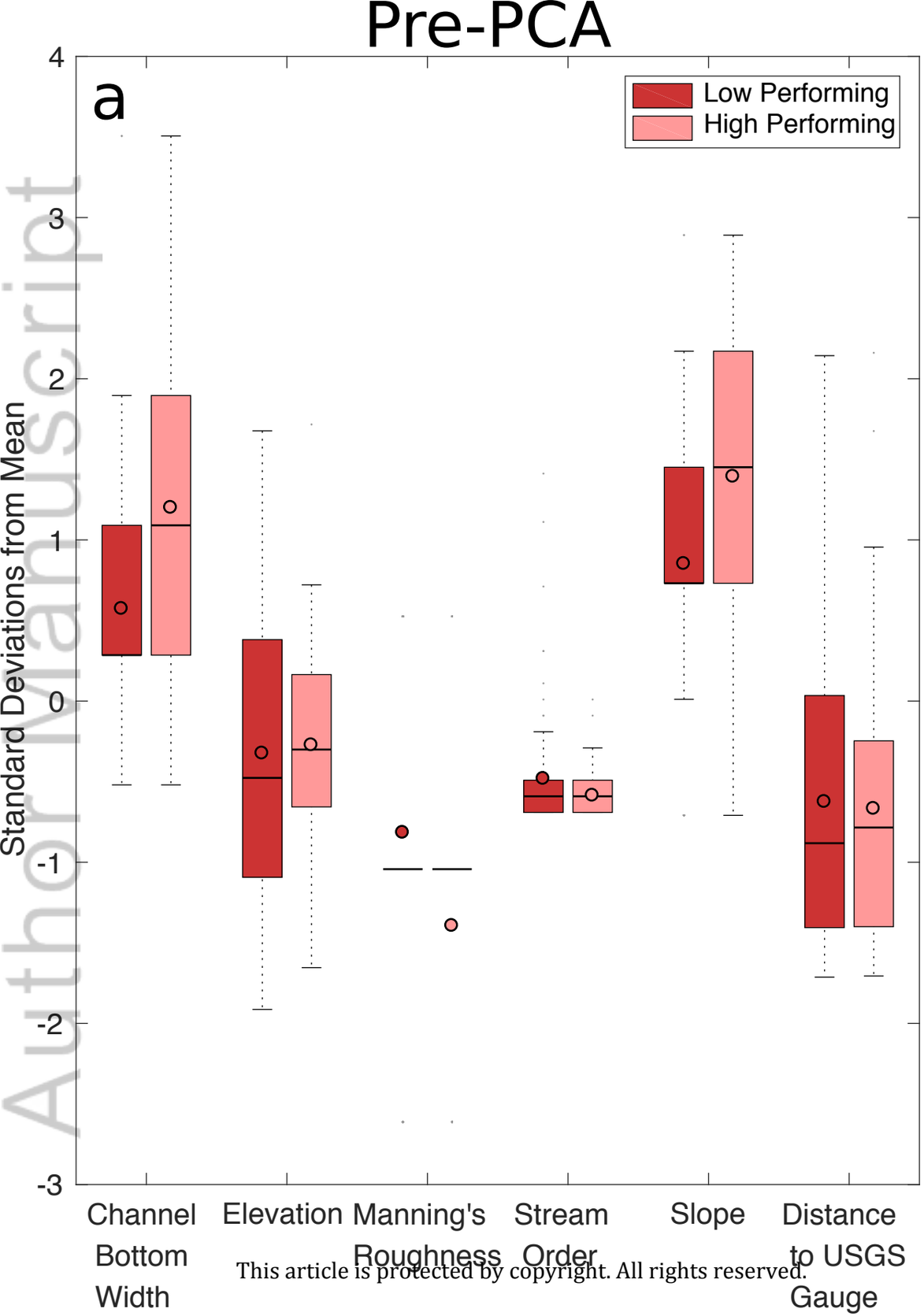



**Figure 8.**



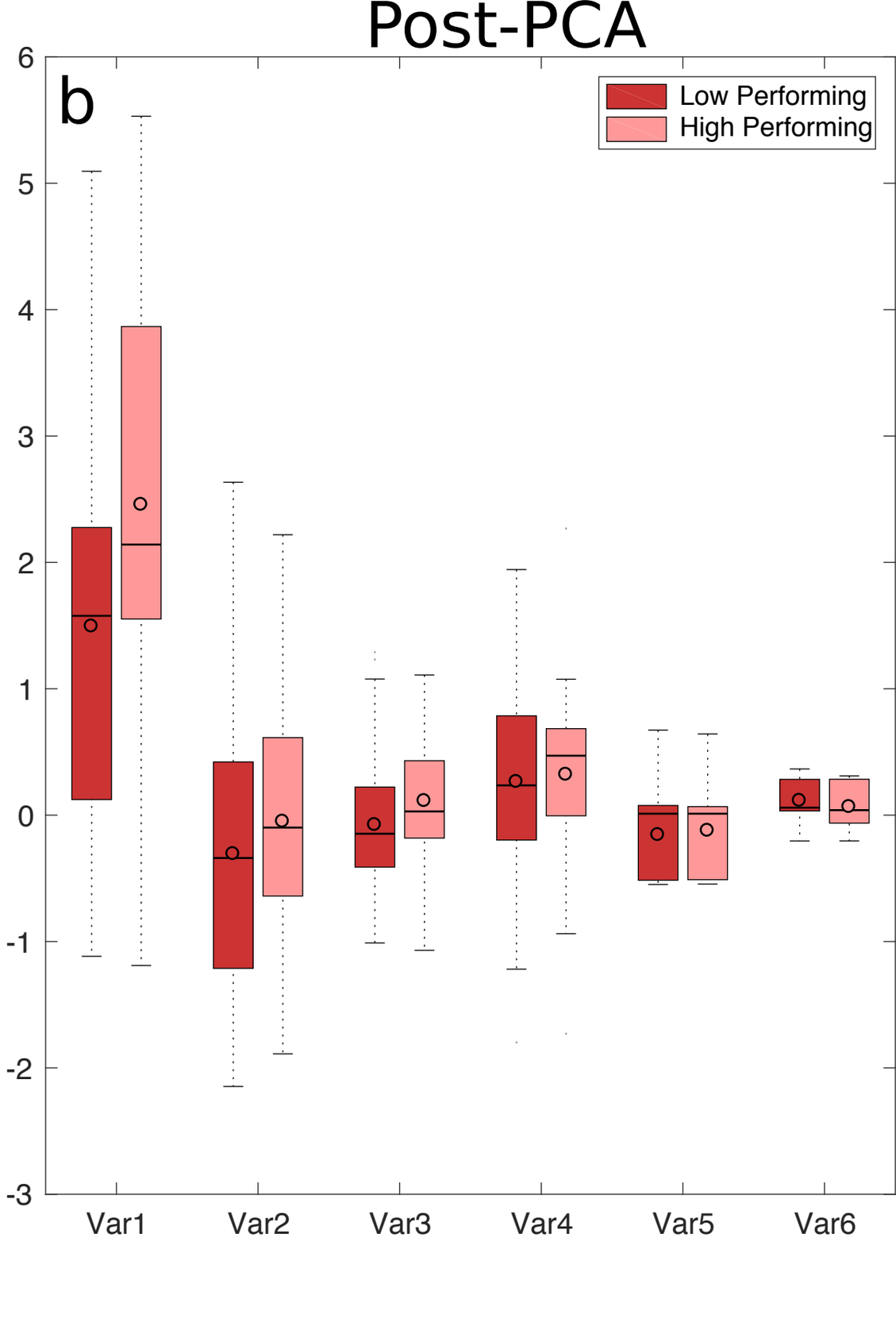
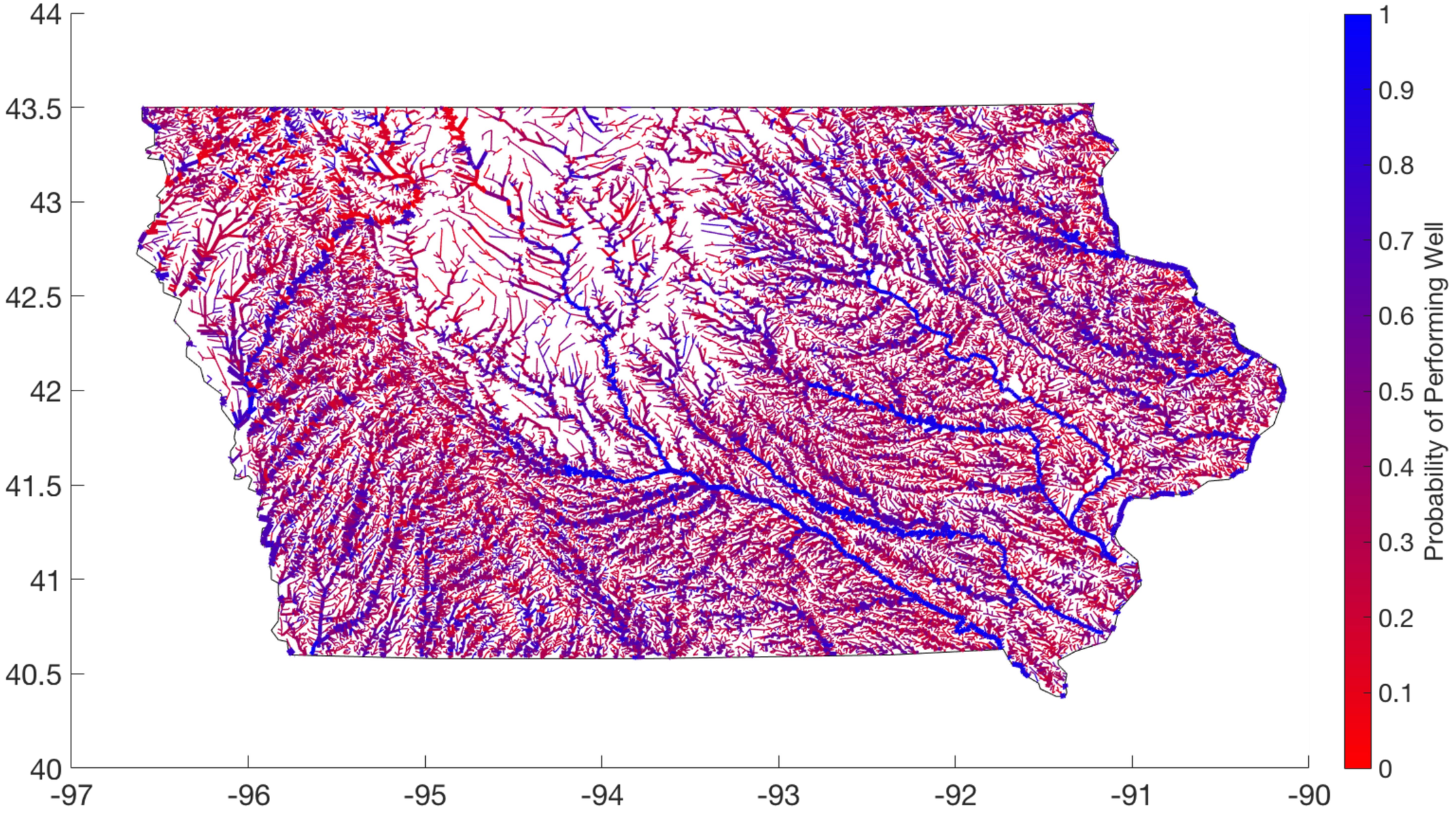
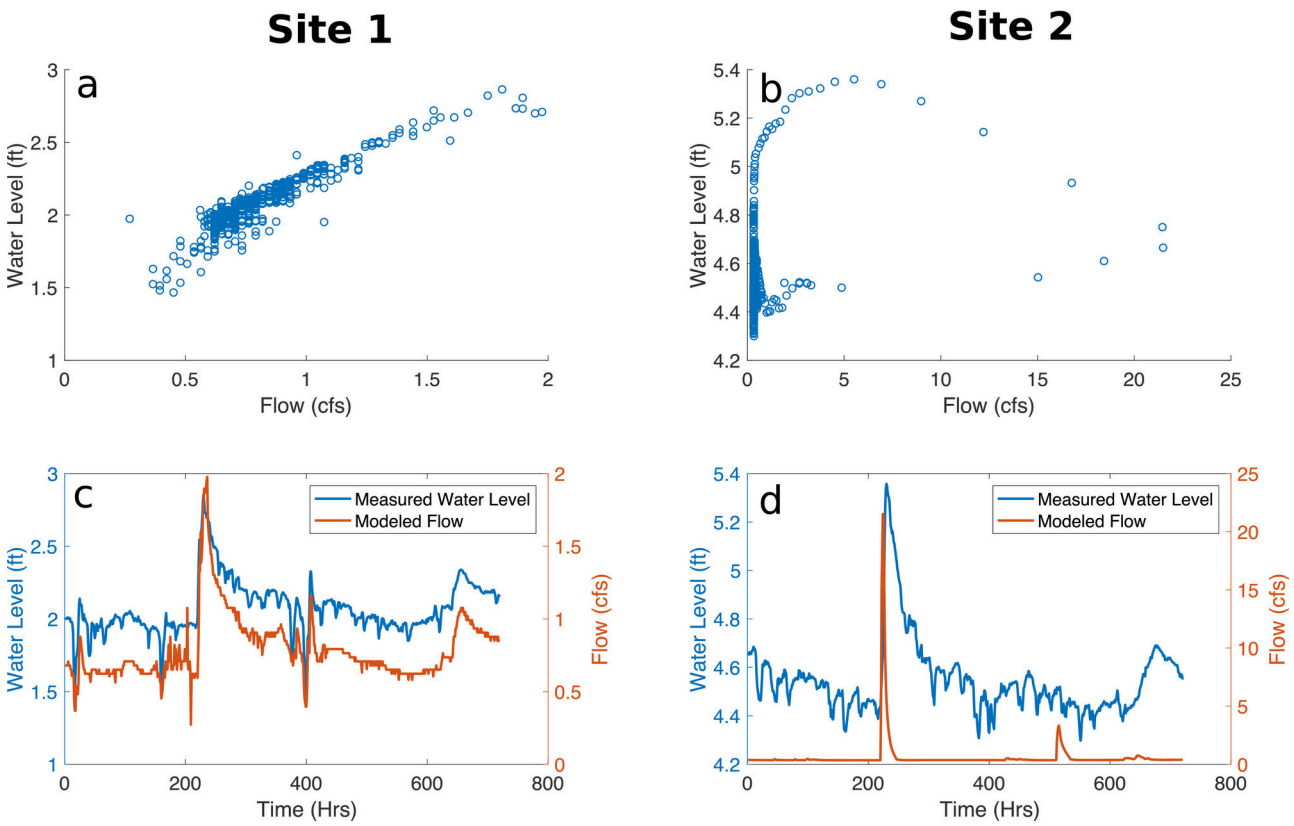
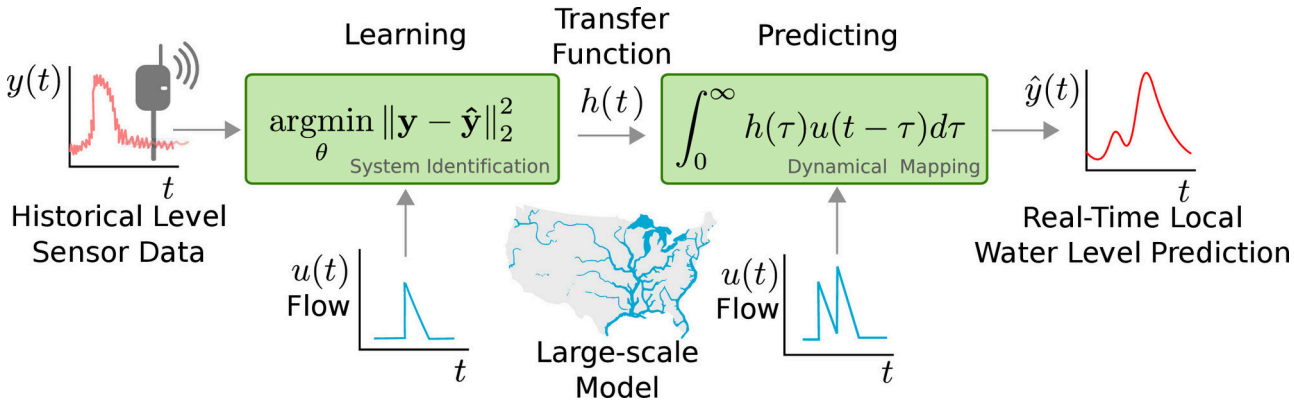



**Figure 9.**

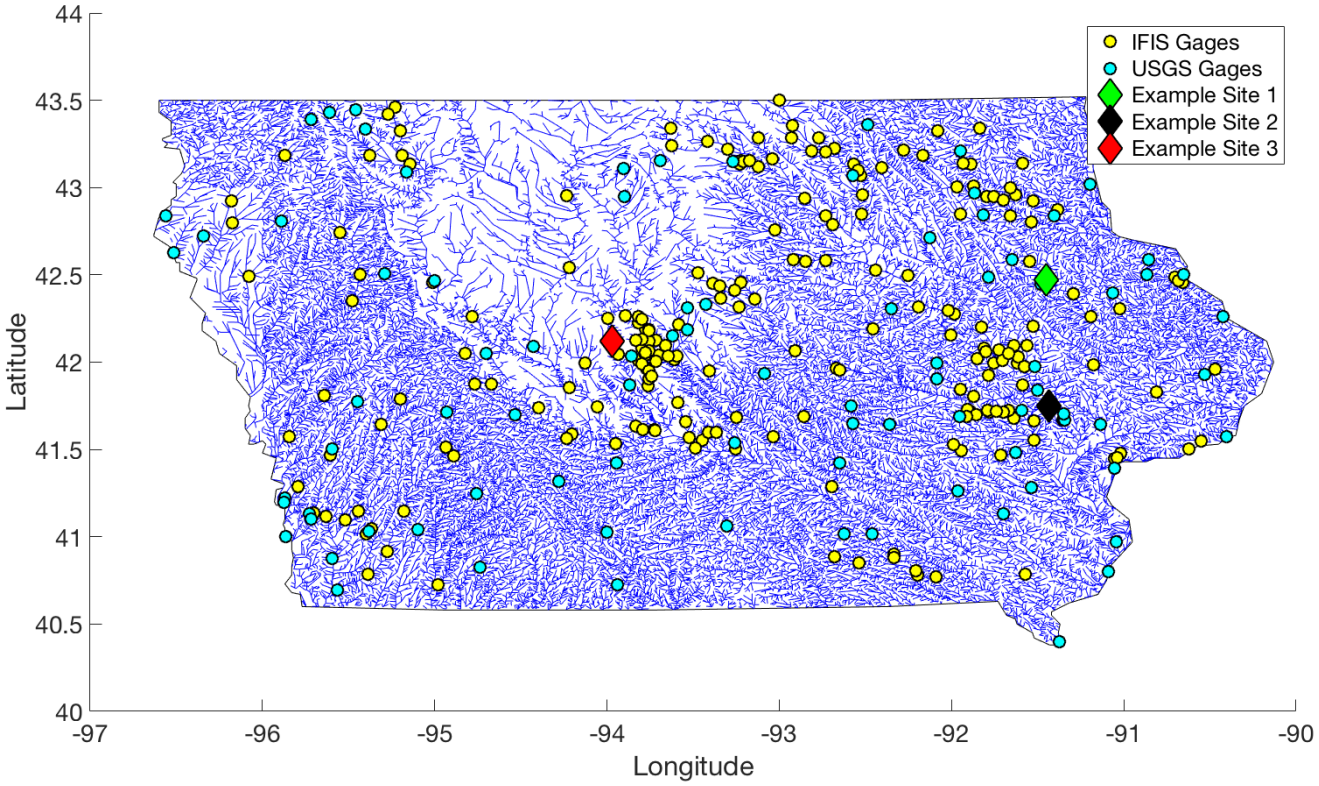




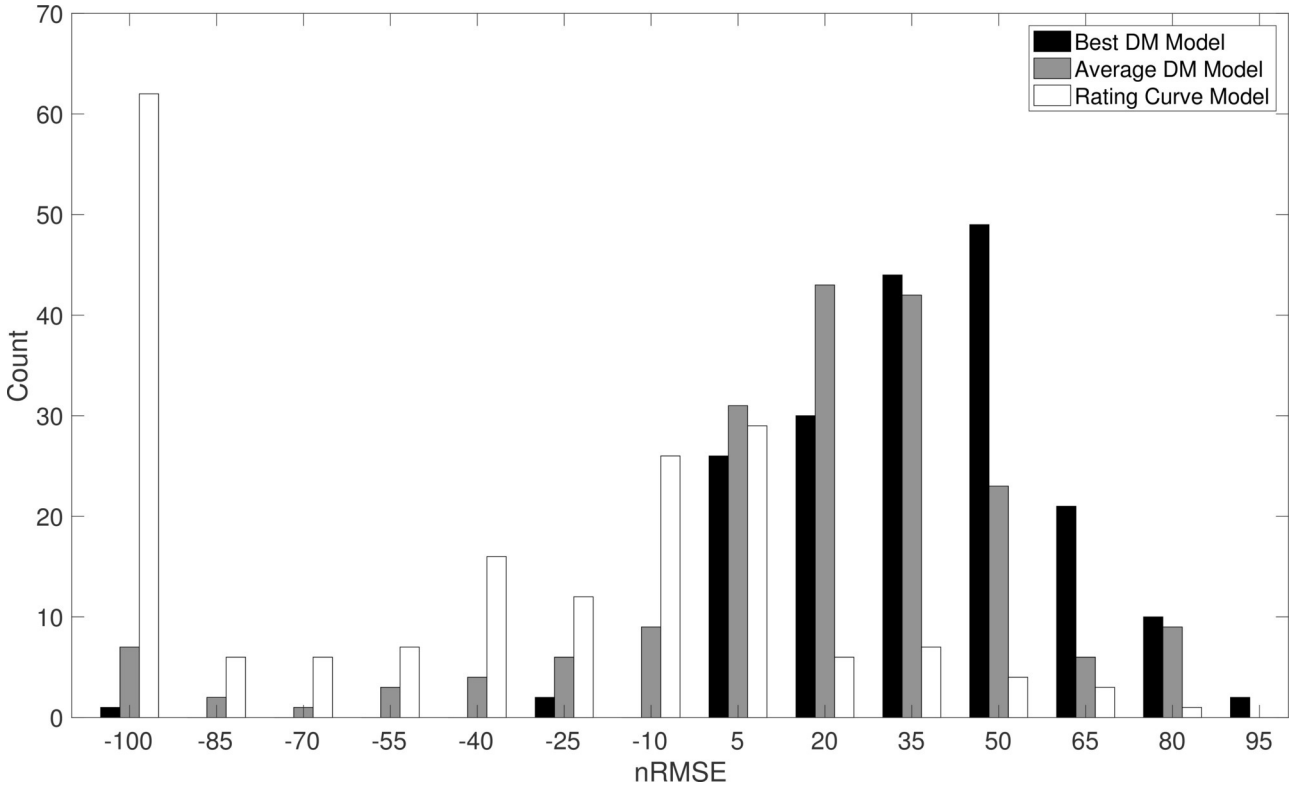

2017wr022498-f01-z-.eps



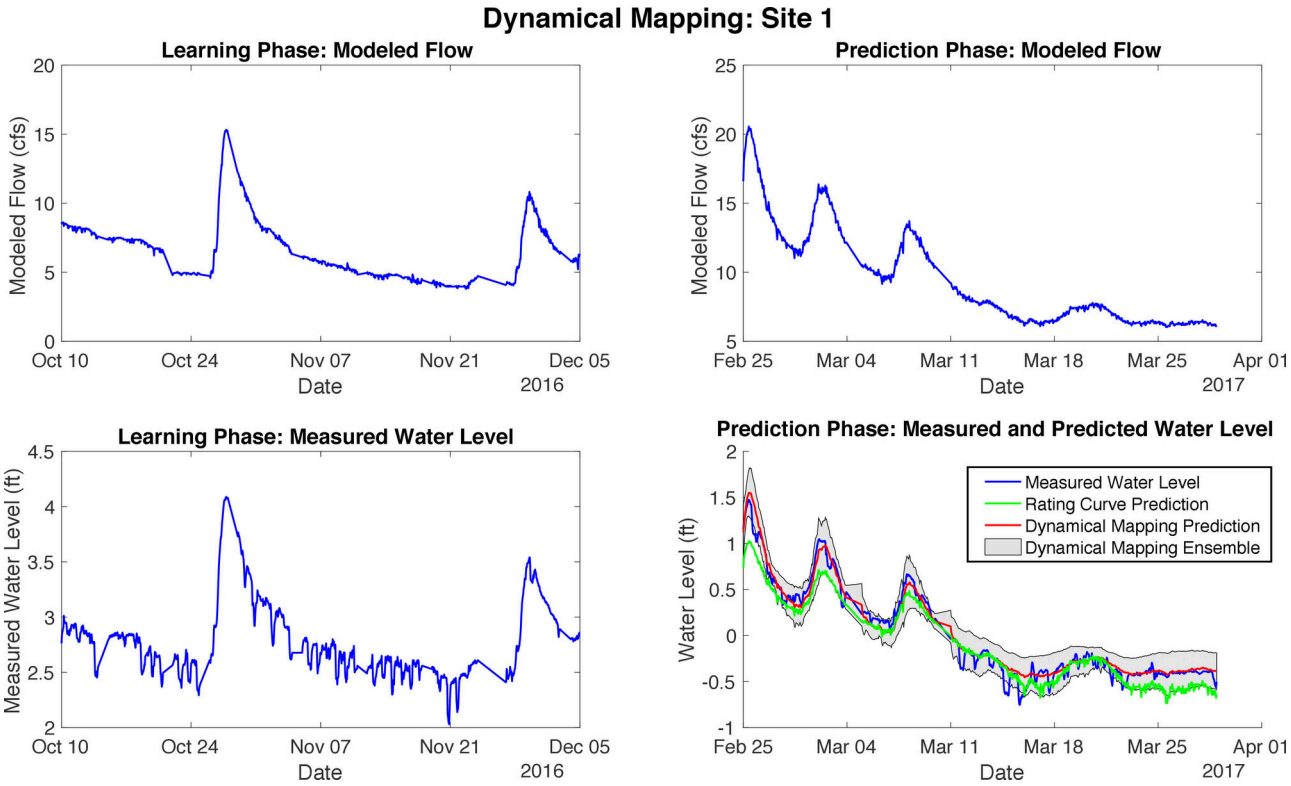

2017wr022498-f02-z-.eps



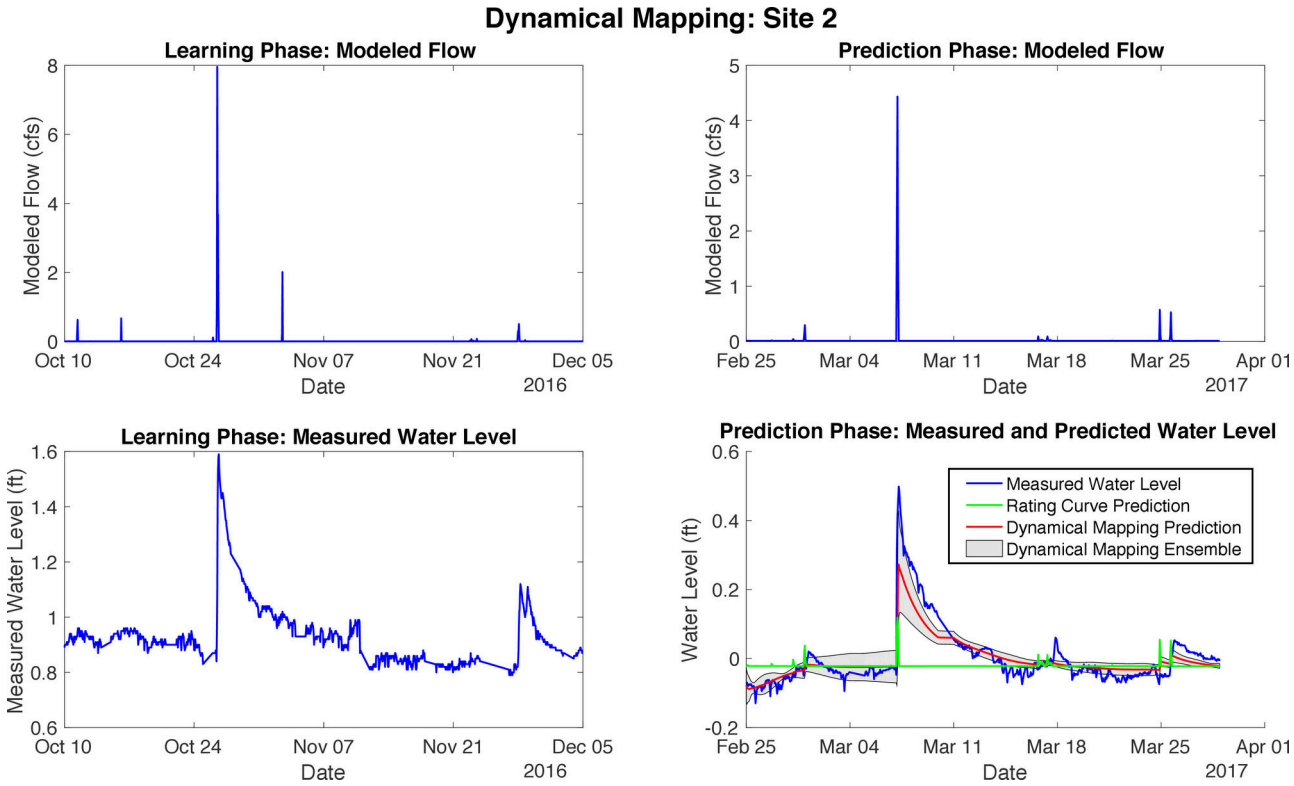

2017WR022498-f03-z-.png



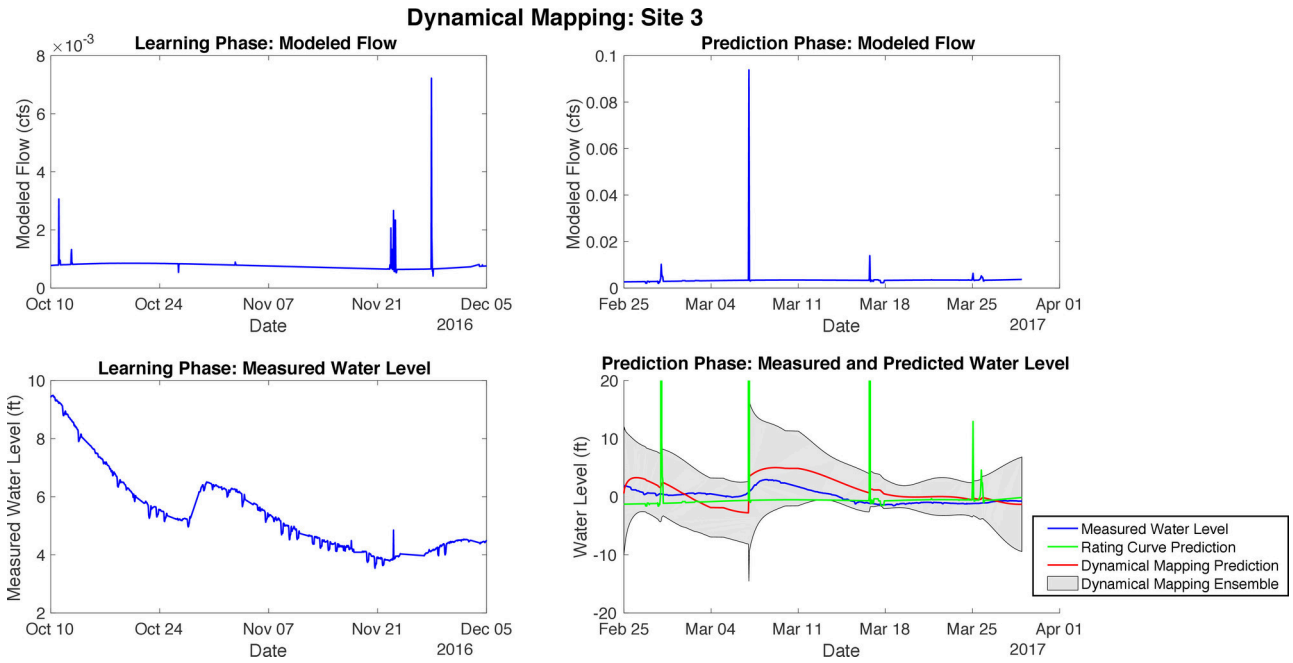

2017wr022498-f04-z-.eps



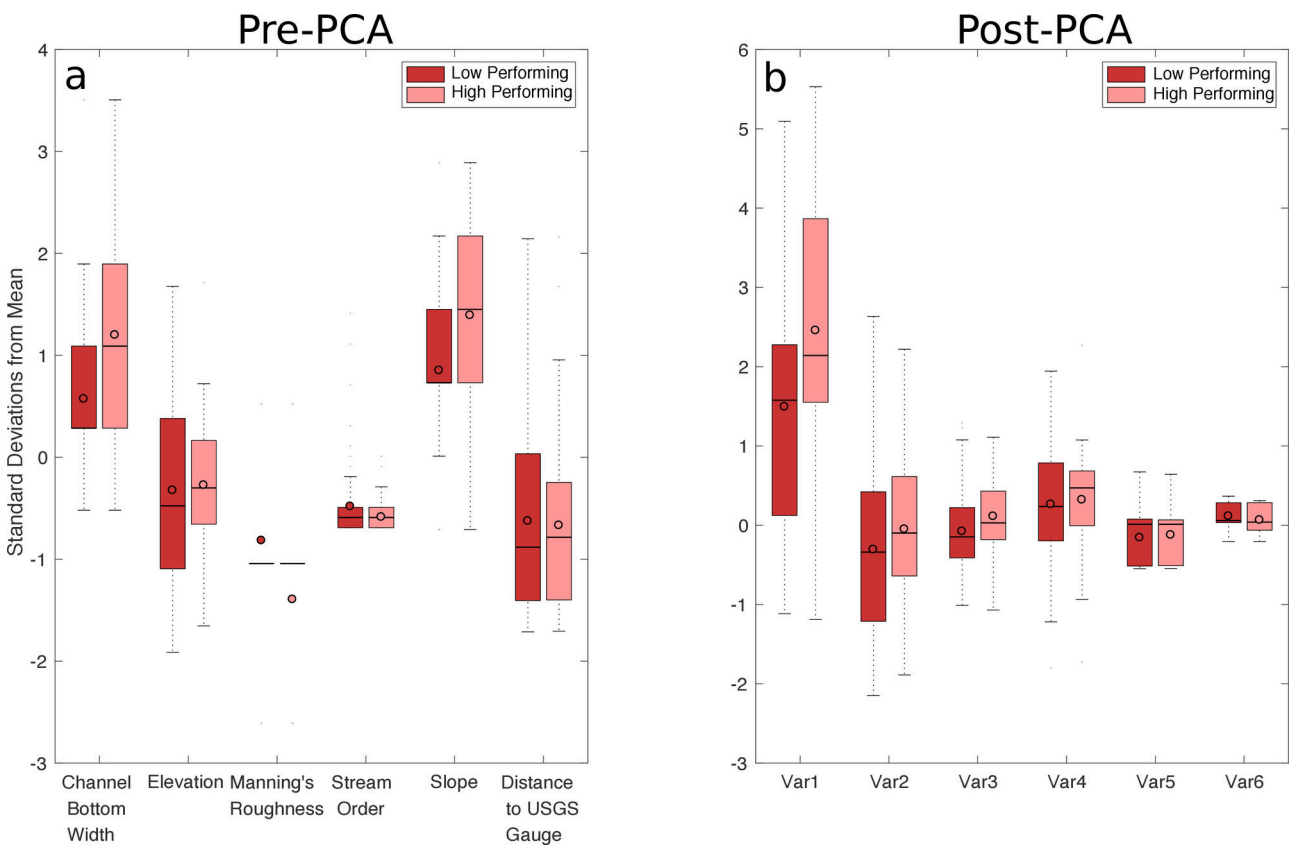

2017wr022498-f05-z-.eps



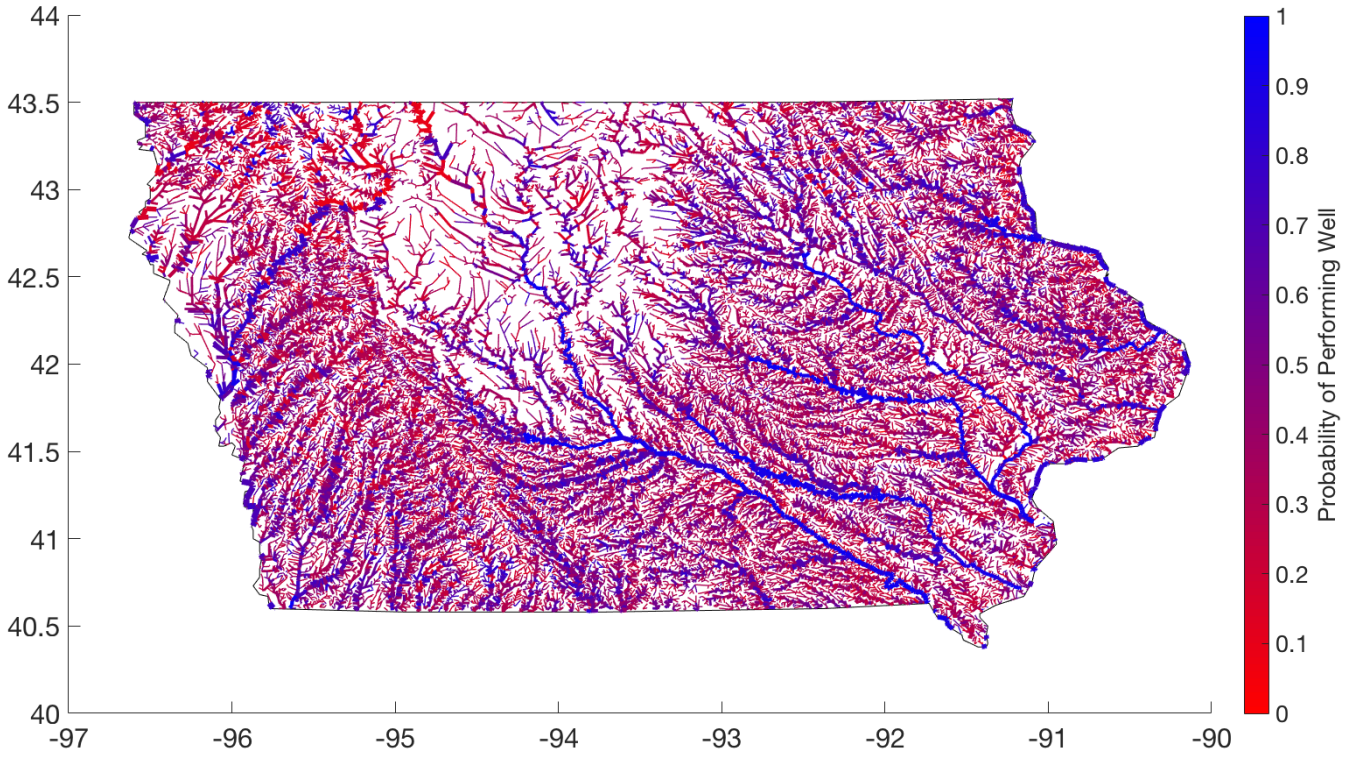

2017wr022498-f06-z-.eps




2017wr022498-f07-z-.eps




2017wr022498-f08-z-.eps




2017WR022498-f09-z-.png