

Fusing Large Datasets and Models to Improve Understanding of Hydrologic and Hydraulic Processes

by

Kevin Fries

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Civil Engineering)
in the University of Michigan
2018

Doctoral Committee:

Professor Branko Kerkez, Chair
Dr. Andrew Gronewold, NOAA GLERL
Professor Seth Guikema
Professor Jerome Lynch
Professor Jeffrey Scruggs

Kevin Fries
kjfries@umich.edu
ORCID iD: 0000-0002-6573-6356

©Kevin Fries 2018

This dissertation is dedicated to my wonderful wife Danielle who was by my side almost the entire way and to my family who gave me the opportunities that enabled me to get to where I am today.

ACKNOWLEDGMENTS

First, I would like to thank my advisor, Branko Kerkez, for his mentorship and guidance throughout this process. He treated me as a peer over the last five years, and I could not have asked for a better advisor.

I would also like to thank the rest of my committee, Dr. Jerry Lynch (CEE), Dr. Jeff Scruggs (CEE), Dr. Drew Gronewold (NOAA GLERL), and Dr. Seth Guikema (IOE). Dr. Lynch and Dr. Scruggs provided invaluable insight to my work from a civil systems perspective. My coursework with Dr. Guikema and subsequent meetings enabled me to grow in my ability to apply data science techniques. And Dr. Gronewold's perspectives on hydrologic problems enabled my work to have a greater impact on the scientific community.

I would also like to acknowledge all the people who helped guide my research along the way. This includes NOAA GLERL, particularly Greg Lang, and the Iowa Flood Center and its Flood Information System for providing access to all their data products. I am also grateful to John Lenters, Chris Spence, Peter Blanken, and Jay Austin for helpful discussions regarding my Great Lakes research. I would also like to thank Fernando Salas, Dave Gochis, and Laura Read for input regarding the National Water Model.

I would also like to thank the other graduate students within CEE, particularly the Real-Time Water Systems Lab, who helped talk through problems, provide different perspectives, and make my experience in Ann Arbor a memorable one.

I would also like to thank the Erb Family Foundation, the Graham Institute Sustainability Fellows Program, the University of Michigan Water Center, and the Cooperative Institute for Limnology and Ecosystems Research for supporting this research at various points.

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.

TABLE OF CONTENTS

Dedication	ii
Acknowledgments	iii
List of Figures	v
List of Tables	vi
List of Appendices	vii
Abstract	viii
 Chapter	
1 Introduction	1
1.1 Integrating Large and Noisy Sensor Datasets with Numerical Models	2
1.2 Using sensor data to dynamically map large-scale models to site-scale forecasts	5
1.3 Model reduction for real-time change detection across complex water systems	7
2 Integrating Large and Noisy Sensor Datasets with Numerical Models	9
2.1 Introduction	9
2.1.1 Motivation: The Great Lakes	10
2.1.2 Existing Models and Data	11
2.1.3 An untapped data source	12
2.2 Materials and Methods	15
2.2.1 Spatial Estimation	15
2.2.2 Implementation	20
2.3 Results	25
2.3.1 Comparison of Vessel Measurements to Physical Models	26
2.3.2 Assimilation Performance	26
2.3.3 Final Data Product	29
2.4 Discussion	32
2.4.1 Comparison of Vessel Measurements to Physical Models	32
2.4.2 Assimilation Performance	34
2.4.3 Final Data Product	36
2.5 Conclusion	39

3	Using sensor data to dynamically map large-scale models to site-scale forecasts	41
3.1	Introduction	41
3.1.1	Background	43
3.1.2	Approach and Contributions	46
3.2	Methods	48
3.2.1	System Identification Theory	48
3.2.2	Data sources and implementation	50
3.2.3	Performance Classification	54
3.3	Results	58
3.3.1	Dynamical Mapping Performance	58
3.3.2	Performance Classification	62
3.4	Discussion	66
3.5	Conclusions	70
4	Model reduction for real-time change detection across complex water systems	72
4.1	Introduction	72
4.1.1	Background	73
4.1.2	Contribution	74
4.2	Methods	75
4.2.1	Initial Model Fit	76
4.2.2	Detecting System Change	80
4.2.3	Implementation	82
4.3	Results	84
4.3.1	Fit and Computational Complexity	84
4.3.2	Damage Detection	86
4.4	Discussion	92
4.5	Conclusion	97
5	Thesis Conclusion	98
5.1	Summary of Discoveries	98
5.2	Future Directions	99
	Bibliography	101
	Appendices	109

LIST OF FIGURES

1.1	Chapter 2 Methodology. Informing an estimate of a hydrometeorological process on the Great Lakes $f(\mathbf{X})$ by mapping inputs from a state of the art model \mathbf{X} to co-located non-standard observations \mathbf{y}	4
1.2	Chapter 3 methodology. Historical measurements made by a sensor are used to <i>learn</i> a dynamical mapping between modeled and measured values. Once the parameters of the mapping are learned, predictions can be made by dynamically transforming the modeled estimates.	6
1.3	Chapter 4 methodology. A linearized model is identified from sensor data and an iterative search is performed to determine changes in the systems. This enables the localization of damage across the system.	8
2.1	Comparison of ship measurement densities to the stationary observations. Red color indicates few measurements, yellow corresponds with locations which contain over 100 measurements, while blue indicates no available ship measurements.	13
2.2	Probabilistic Methodology. Inform an estimate of the process $f(\mathbf{X})$ by matching inputs \mathbf{X} to co-located ship observations \mathbf{y}	16
2.3	Seasonal Mean Absolute Error for Lake Surface Temperature, comparing the measured value from a ship to the co-located physically modeled estimate, averaged over each instance at that grid cell.	22
2.4	Seasonal Mean Absolute Error for Wind Speed. Interpretation follows from Figure 2.3	23
2.5	Seasonal Mean Absolute Error for Air Temperature. Interpretation follows from Figure 2.3	24
2.6	Comparing air temperature predicted by the physical model (left panel) and the GP approach (right panel) to ship observations that where not used to in training the probabilistic approach. For visualization purposes the data have been binned into 0.1 degree Celsius bins along the x-axis (ship observations) and color-coded as a histogram along the y-axis (model estimate). The color corresponds to the relative density of the data in that bin. In a very good model the highest density of points will fall along the 45-degree line.	27
2.7	Comparing lake surface temperature predicted by the physical model (left panel) and the GP approach (right panel) to ship observations that where not used in training of the GP model.	28

2.8	Comparing wind speeds predicted by the physical model (left panel) and the GP approach (right pane) to ship observations that where not used in training of the GP model.	30
2.9	Winter 2007 comparison of physical models to the probabilistic method. The first row contains the seasonally averaged physical model estimates for air temperature (first column), lake surface temperature (second column), and wind speed (third column). The second row contains the GP model’s estimate, seasonally averaged across the Winter of 2007. The third row displays the average difference between the physical model and GP approach. The fourth row displays the average predictive standard deviation of the GP model for each grid cell. Please see the data archive to generate more detailed maps.	31
2.10	Summer 2013 comparison of physical models to the probabilistic method. Interpretation follows that of Figure 2.9.	32
3.1	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.	45
3.2	Conceptual diagram of dynamical mapping methodology. Historical measurements made by a sensor are used to “ <i>learn</i> ” 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.	47
3.3	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 chapter.	52
3.4	Logit Boosted Random Forest	57
3.5	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 correlation-based rating curve approach (white)	58
3.6	Dynamically mapping modeled flows to local water levels on site 1 (see Figure 3.3). Data used to “ <i>learn</i> ” the mapping parameters are plotted in the left column, while the resulting mapping is applied to future data in the right column. For this example site, the dynamical mapping performs relatively well (nRMSE of 80%). A simple regression-based rating curve approach (not plotted) performs strongly as well, with an nRMSE of 76%.	60
3.7	Dynamically mapping modeled flows to local water levels on site 2, following convention used in Figure 3.6. For this example site, the dynamical mapping performs relatively well (54% nRMSE), while a simple regression-based rating curve approach does not (−4% nRMSE).	61

3.8	Dynamically mapping modeled flows to local water levels on site 3, following convention in Figure 3.6. For this example site, neither the dynamical mapping (0% nRMSE) or regression-based rating curve (nRMSE of $-14,900\%$ nRMSE) perform well.	62
3.9	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.	63
3.10	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).	65
4.1	A simplified abstraction of a complex system is generated through system identification from a sensor observation. Then, when an anomalous signal is detected, an iterative search can be performed quickly to probabilistically identify likely locations for damage.	76
4.2	Node i where the flow passing through the node at the next time step (x_i) is a function of the runoff ($b_{im}u_m$), the flow from upstream ($a_{ki}x_k$), the flow going downstream ($a_{ij}x_j$), and the flow staying at the node ($a_{ii}x_i$)	77
4.3	Change Detection Algorithm	82
4.4	The stormwater network to be analyzed in this chapter. This particular network has an area of nearly $5km^2$. Here, the SWMM model is overlaid on a map of the area.	83
4.5	The computational time required to fit the initial state space model as a function of the number of nodes in the network. The model is fit in linear time, with some variation in convergence time due to random initialization of the parameters	85
4.6	Time Series of modeled vs observed flows. The linearized model abstraction is capable of capturing the dynamics both before and after a pipe is damaged.	86
4.7	The computational time required to iterate over each node in a pipe blockage scenario and identify the damaged location.	87

4.8	Example performance on a simple network. The green link represents the broken pipe, the green outlined node represents the node with the highest likelihood of damage, and the color of the node shows the relative likelihood as a measure of the normalized RMSE, with negative NRMSE values being set to zero. The more red a node, the higher the likelihood that it may be damaged, according to the proposed approach. Circles represent junctions while squares represent storage nodes.	88
4.9	The algorithm is able to detect a pipe blockage with high confidence. The damaged pipe is very close to the outfall and contributes the majority of the flow.	88
4.10	The algorithm is able to detect the pipe blockage when each branch of the system is contributing similar magnitudes of flow. Note that while many candidates were selected, the most likely candidate was the correct node. . .	89
4.11	The algorithm is again able to detect the pipe blockage when each branch of the system is contributing similar magnitudes of flow.	89
4.12	The algorithm does not identify the correct pipe because the contributing flow and time of concentration from that pipe is quite similar to the identified node.	90
4.13	The algorithm does not identify the correct pipe because the contributing flow from the pipe is relatively small, with the majority of the flow entering the network from the top left branch. Therefore, all the terminal nodes are highlighted as candidate sites.	91
4.14	The algorithm does not identify the correct pipe because it is too far from the outfall and passes through too many storage nodes, resulting in non-linear dynamics at the outfall.	91
A.1	Comparison of ship measurement densities to the stationary observations of temperature. Red color indicates few measurements, yellow corresponds with locations which contain over 100 measurements, while blue indicates no available ship measurements.	110
A.2	Comparison of ship measurement densities to the stationary observations of surface temperature. Red color indicates few measurements, yellow corresponds with locations which contain over 100 measurements, while blue indicates no available ship measurements.	111
A.3	Comparison of ship measurement densities to the stationary observations of wind speed. Red color indicates few measurements, yellow corresponds with locations which contain over 100 measurements, while blue indicates no available ship measurements.	112
A.4	Winter 2006	113
A.5	Spring 2006	114
A.6	Summer 2006	115
A.7	Fall 2006	116
A.8	Spring 2007	117
A.9	Summer 2007	118
A.10	Fall 2007	119

A.11	Winter 2008	120
A.12	Spring 2008	121
A.13	Summer 2008	122
A.14	Fall 2008	123
A.15	Winter 2009	124
A.16	Spring 2009	125
A.17	Summer 2009	126
A.18	Fall 2009	127
A.19	Winter 2010	128
A.20	Spring 2010	129
A.21	Summer 2010	130
A.22	Fall 2010	131
A.23	Winter 2011	132
A.24	Spring 2011	133
A.25	Summer 2011	134
A.26	Fall 2011	135
A.27	Winter 2012	136
A.28	Spring 2012	137
A.29	Summer 2012	138
A.30	Fall 2012	139
A.31	Winter 2013	140
A.32	Spring 2013	141
A.33	Fall 2013	142
A.34	Winter 2014	143
A.35	Spring 2014	144
A.36	Summer 2014	145
A.37	Fall 2014	146
B.1	Site BEARCRK01	147
B.2	Site BEAVER01	148
B.3	Site BEAVER02	148
B.4	Site BEAVER03	149
B.5	Site BEAVERCRK01	149
B.6	Site BEVRCRK01	150
B.7	Site BEVRCRK02	150
B.8	Site BFFLOCR01	151
B.9	Site BFFLOCR02	151
B.10	Site BLCKHWK01	152
B.11	Site BNNTTCR01	152
B.12	Site BOYERRV01	153
B.13	Site BOYERRV02	153
B.14	Site BRCHFLD01	154
B.15	Site BRCKCMP01	154
B.16	Site BRSHYCR01	155
B.17	Site BVRCR01	155

B.18	Site CAMPCR01	156
B.19	Site CEDARRV02	156
B.20	Site CEDARRV04	157
B.21	Site CEDARRV05	157
B.22	Site CHQSTCR01	158
B.23	Site CHQSTCR02	158
B.24	Site CHSLSCR01	159
B.25	Site CLRCRK01	159
B.26	Site CLRCRK03	160
B.27	Site CLRCRKS01	160
B.28	Site CLRCRKW01	161
B.29	Site COONCR01	161
B.30	Site CRANECR02	162
B.31	Site CTFSHCR01	162
B.32	Site DEEPCRK01	163
B.33	Site DEERCR01	163
B.34	Site DRNGDTCH01	164
B.35	Site DRYCR01	164
B.36	Site DRYCR02	165
B.37	Site DRYCRK02	165
B.38	Site DRYRNCRK01	166
B.39	Site DSMNSRV01	166
B.40	Site DSMNSRV03	167
B.41	Site DSMNSRV04	167
B.42	Site DUCKCR01	168
B.43	Site EFDSMNS01	168
B.44	Site EINDNCR01	169
B.45	Site ELKRN01	169
B.46	Site ENDWY01	170
B.47	Site ENISH01	170
B.48	Site ENISH02	171
B.49	Site ENISH03	171
B.50	Site ENISH04	172
B.51	Site FLOYDRV02	172
B.52	Site FOURMLE01	173
B.53	Site FOURMLE02	173
B.54	Site FOURMLE03	174
B.55	Site GENEVACR01	174
B.56	Site GRANCRK01	175
B.57	Site GRNGRCR01	175
B.58	Site HOOSRCR01	176
B.59	Site HRGVCR01	176
B.60	Site INDCR01	177
B.61	Site INDCR04	177

B.62	Site INDIAN01	178
B.63	Site INDIAN02	178
B.64	Site IOWARV01	179
B.65	Site KEGCR01	179
B.66	Site LAKECR01	180
B.67	Site LTLCEDAR01	180
B.68	Site LTLCEDAR02	181
B.69	Site LTLSTIOUX01	181
B.70	Site LTLSTIOUX02	182
B.71	Site LTLSTIOUX03	182
B.72	Site LTLSTIOUX04	183
B.73	Site LTLSTIOUX05	183
B.74	Site LTLSOAP01	184
B.75	Site LTLTKY01	184
B.76	Site LTLTKY02	185
B.77	Site LTLWAPSI01	185
B.78	Site LTLWLNT01	186
B.79	Site MADCRK01	186
B.80	Site MADCRK02	187
B.81	Site MAPLERV01	187
B.82	Site MAPLERV02	188
B.83	Site MCHKNCK01	188
B.84	Site MDCR01	189
B.85	Site MDDLRCN01	189
B.86	Site MDDLRCN02	190
B.87	Site MDDLRCN03	190
B.88	Site MQKTARV01	191
B.89	Site MQKTARV02	191
B.90	Site MQKTARV04	192
B.91	Site MRGNCR01	192
B.92	Site MSQUTOCR01	193
B.93	Site MSQUTOCR02	193
B.94	Site MUDCR01	194
B.95	Site MUDCRK01	194
B.96	Site NENGLSH01	195
B.97	Site NFCTFSH01	195
B.98	Site NFMQKTA01	196
B.99	Site NRCCNRV01	196
B.100	Site NRCCNRV02	197
B.101	Site NRCCNRV03	197
B.102	Site NSKNK01	198
B.103	Site NWLNTCR01	198
B.104	Site NWLNTCR02	199
B.105	Site OCHYDNRV01	199

B.106	Site OTTERCR01	200
B.107	Site OTTRCRK01	200
B.108	Site OTTRCRK02	201
B.109	Site OTTRCRK03	201
B.110	Site OTTRCRK04	202
B.111	Site PRAIRIECRK01	202
B.112	Site PRICECR02	203
B.113	Site RAPIDCR01	203
B.114	Site RAPIDTRB01	204
B.115	Site RAVENCR01	204
B.116	Site RBRTSCR01	205
B.117	Site RCCNRV01	205
B.118	Site RLSTNCR01	206
B.119	Site ROCKCR01	206
B.120	Site ROCKCR02	207
B.121	Site SBVRCRK01	207
B.122	Site SCHQSTCR01	208
B.123	Site SENGLSH01	208
B.124	Site SFIOWA01	209
B.125	Site SFIOWA03	209
B.126	Site SHLRK01	210
B.127	Site SHLRK02	210
B.128	Site SHLRK03	211
B.129	Site SHLRK04	211
B.130	Site SHLRK05	212
B.131	Site SLOUGHCR01	212
B.132	Site SLVRCR01	213
B.133	Site SOAPCR01	213
B.134	Site SOAPCR03	214
B.135	Site SPRINGCR01	214
B.136	Site SPRNGCRK01	215
B.137	Site SQWCR01	215
B.138	Site SQWCR02	216
B.139	Site SQWCR03	216
B.140	Site SQWCR04	217
B.141	Site SRCCNRV01	217
B.142	Site SSKNK01	218
B.143	Site SSKNK02	218
B.144	Site SSKNK03	219
B.145	Site SSKNK04	219
B.146	Site SUGARCR01	220
B.147	Site TIPTNCR01	220
B.148	Site TIPTNCR02	221
B.149	Site TKYRV01	221

B.150	Site TKYRV03	222
B.151	Site TRBLSM01	222
B.152	Site UIWARV01	223
B.153	Site VLGARV01	223
B.154	Site VLGARV02	224
B.155	Site VLGARV03	224
B.156	Site WALNUTCR01	225
B.157	Site WAPSI02	225
B.158	Site WAPSI03	226
B.159	Site WBFLOYDRV01	226
B.160	Site WFKLTSIOUX01	227
B.161	Site WFLTSIOUX01	227
B.162	Site WLLWCR02	228
B.163	Site WLNTCR01	228
B.164	Site WLNTCRK01	229
B.165	Site WNBGORV03	229
B.166	Site WNBGORV04	230
B.167	Site WNBGORV06	230
B.168	Site WNISH01	231
B.169	Site WNISH02	231
B.170	Site WNISH03	232
B.171	Site YELLOWRV01	232

LIST OF TABLES

2.1	Comparing the performance of the proposed approach to physical models in predicting predicting ship observations that were not used in the training procedure. Quantified as the percent reduction in mean difference and variance (averaged over all 45 models for each lake-variable pair).	29
3.1	Principal Components resulting from applying PCA to features of 62,000 streams across the entire state of Iowa. σ is the singular value associated with that component whose relative magnitude indicates the amount of variability the component explains in the data.	64

LIST OF APPENDICES

A Supporting Information for Big Ship Data: Using Vessel Measurements to Improve Estimates of Temperature and Wind Speed on the Great Lakes . .	109
B Supporting Information for Using in-situ measurements to dynamically map large-scale models to site-scale forecasts: A case study using the National Water Model	147

ABSTRACT

Global water systems are being stressed by aging infrastructure, climate change, and resource withdrawals. The ability to model large water systems has attempted to keep pace with these challenges, with modern water models now operating effectively across massive spatial and temporal scales. Despite great advances in numerical modeling, many decision makers require high-resolution information that large-scale models do not yet provide. Simultaneously, the affordability and ease of use of sensing platforms has improved dramatically, enabling even small communities and research groups to deploy observation networks. Unfortunately, these real-time measurements are often not attached to a physical or numerical model, which prevents their use in predictive applications. To that end, this dissertation poses the question: how can the domain knowledge embedded in large-scale models be fused with new forms of sensor data to improve understanding of hydrologic and hydraulic processes? Three primary issues currently prevent this question from being answered. First, many datasets are irregular or noisy, making integration with models difficult. This dissertation addresses this issue by providing a methodology for integrating non-standard and distributed measurements with large numerical models. The approach is applied to an unprecedented data set of over one million ship observations across the Great Lakes to generate new insights about distributed hydrometeorological processes. Second, the scales across which water models operate do not often match the scales at which we measure. This dissertation addresses this issue by providing a methodology for dynamically mapping large-scale model outputs to site-scale forecasts. The approach is applied to flood forecasting across the entire state of Iowa, where nearly two hundred sensors are fused with the US National Water Model. Third, since many numerical models of water systems are often heavily parameterized, it is difficult to determine how to update these models when novel sources of sensor data emerge. This dissertation addresses this issue by providing a methodology for abstracting simple models from complex water networks

to enable efficient detection and localization of change. The approach will underpin a real-time asset management methodology for stormwater systems. Ultimately, this dissertation seeks to contribute to the emergence of *Big Data Hydrology* by discovering opportunities in data-driven water modeling that will be enabled by systems engineering and data science.

CHAPTER 1

Introduction

The goal of this dissertation is to advance the knowledge of water systems by allowing large hydrologic and hydraulic models to be fused with new forms of sensor data. This will improve the fundamental understanding of hydrologic processes and inform better water management. Around the globe, water systems are being stressed by aging infrastructure, climate change, and resource withdrawals [1]. The ability to model large water systems has attempted to keep pace with these challenges. While new models are effective at representing hydrologic and hydraulic processes across massive scales, they do not yet provide the fine-resolution insights desired by local decision makers [2].

Simultaneously, the increasing affordability of sensors has made it possible for even small communities and research groups to monitor their water systems [3]. However, these real-time measurements are often not attached to a physical or numerical model, limiting their ability to be used in studies of physical processes or forecasting applications. In this dissertation I ask the question: *How can the domain knowledge embedded in large-scale models be fused with new forms of sensor data to improve understanding of hydrologic and hydraulic processes?* At this time, a number of fundamental knowledge gaps make it difficult to answer this question, including:

1. Many sensor datasets are irregular or noisy, which challenges integration with models using traditional data assimilation approaches.

2. The scales across which numerical models operate do not often match the scales at which sensor observations are made.
3. Many numerical models of complex water systems are heavily parameterized. Therefore, when novel sources of sensor data emerge, it is difficult to determine how to update these models efficiently and effectively.

This dissertation will bridge these knowledge gaps and discover emerging opportunities in data-driven water modeling by using techniques from systems engineering and data science. Specifically, The contributions of this dissertation are:

Chapter 2 A probabilistic methodology for integrating noisy, irregular observations into large-scale water models. The approach will be applied to a massive dataset of over one million ship observations collected across the Great Lakes.

Chapter 3 A methodology for dynamically mapping large-scale model outputs to site-scale forecasts, applied to flood forecasting across the entire state of Iowa.

Chapter 4 An approach for deriving and identifying linear models from complex water systems, with the aim of enabling real-time asset management in urban watersheds.

1.1 Integrating Large and Noisy Sensor Datasets with Numerical Models

The sheer size of many natural water systems limits the ability of in-situ sensor networks to resolve spatiotemporal variability of underlying hydrologic processes. Therefore, hydrologists often rely on physically-based models to estimate hydrologic processes [2]. As sensing platforms have become more affordable, many novel data sources have begun to emerge. Many of these datasets are irregularly formatted, inconsistently measured, and noisy [4]. However, many of these measurements have become available at locations that

are traditionally difficult to observe, thereby providing potentially valuable insights. The question then becomes: how can these non-traditional measurements be used to improve hydrometeorological estimates across large water systems? The first chapter of this dissertation will investigate how large sets of distributed sensor data can be integrated with models using probabilistic methods, in particular Gaussian Process Regression. The approach will be evaluated by integrating over one million new ship-based observations across the Great Lakes with a large numerical water model.

The Great Lakes watershed has an unprecedented water to land ratio. As such, evaporation is an unusually large component of the water balance [5]. Understanding the spatiotemporal distribution of evaporation will enable more static water levels, saving the local economy hundreds of millions of dollars due to increased shipping revenue [6]. The Great Lakes Coastal Forecasting System (GLCFS), the primary regional water model, numerically estimates hydrometeorological conditions across the Great Lakes, but it is calibrated almost exclusively on near-shore or terrestrial observation stations. Therefore, open-water dynamics are highly uncertain. Recently, mobile observations from ship-based weather stations have emerged, which provide observations across regions of the lakes that have never been measured before.

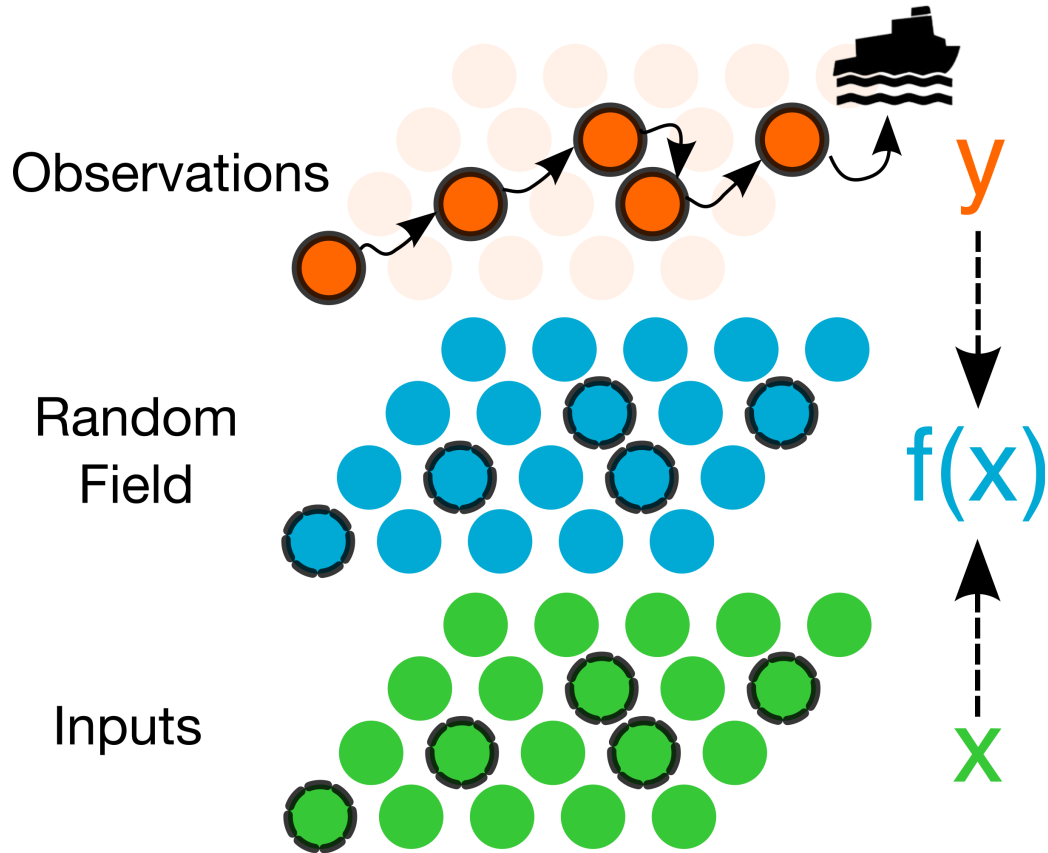


Figure 1.1: Chapter 2 Methodology. Informing an estimate of a hydrometeorological process on the Great Lakes $f(\mathbf{X})$ by mapping inputs from a state of the art model \mathbf{X} to co-located non-standard observations \mathbf{y} .

Given the noisy nature of the observations, direct ingestion of mobile sensor data through traditional assimilation and filtering approaches becomes impractical [7]. For a given ship, without knowing the measurement height, reporting methods, and maintenance schedules, a single data point carries little perceived information. However, by formulating a new approach around probabilistic processes, this chapter will demonstrate how these ship observations can be fused with large regional water models to derive updated spatial estimates of various hydrometeorological processes (Figure 1.1).

This research will have two primary contributions. Firstly, this research provides an approach by which to integrate noisy, irregular observations into large-scale models to generate new insights about hydrologic processes. This is especially useful when traditional sensing platforms are difficult or impossible to deploy. Secondly, this research results in a

new openly available data product consisting of nine years of gridded hourly estimates of wind speed, air temperature, surface temperature, and relative humidity – the four primary variables for estimating evaporation using bulk methods. This will be the first gridded database for the Great Lakes that integrates over a million open water observations.

1.2 Using sensor data to dynamically map large-scale models to site-scale forecasts

Computational capacity has grown rapidly over the last few decades, enabling massive-scale hydrologic models. However, when modeling across continental or global scales, the amount of data available for calibration is comparatively small [8]. This results in coarse model dynamics at the local scale, which limits the utility of these models to local decision makers. Therefore, there is a need to downscale complex, state-of-the-art large-scale models to the site level so decision makers can have actionable information [9]. While many local sensor data are available, they are often not used in the calibration of the larger model. The question then becomes: how can highly-localized forecasts be generated by fusing site-scale sensor measurements with outputs from larger-scale physical models?

The US National Water Model (NWM), which recently became operational, models 2.7 million stream reaches across the entire continental US with the ultimate goal of providing street-level flood forecasts. While this ambitious model routes water effectively on large rivers, many streams in headwater areas are modeled coarsely and do not match real-world dynamics. At the same time, sensing platforms are becoming cheaper and easier to deploy. A great example is given in the state of Iowa, where nearly 200 stream gages are maintained by the Iowa Flood Center (IFC). For decision makers in areas such as Iowa, early warning flood forecasting systems are extremely important. However, with the NWM not being able to resolve stream dynamics at the site-level, its present utility appears limited. Discovering a way to fuse NWM forecasts with locally-available sensor observations thus becomes a

motivation for providing potentially life-saving forecasts.

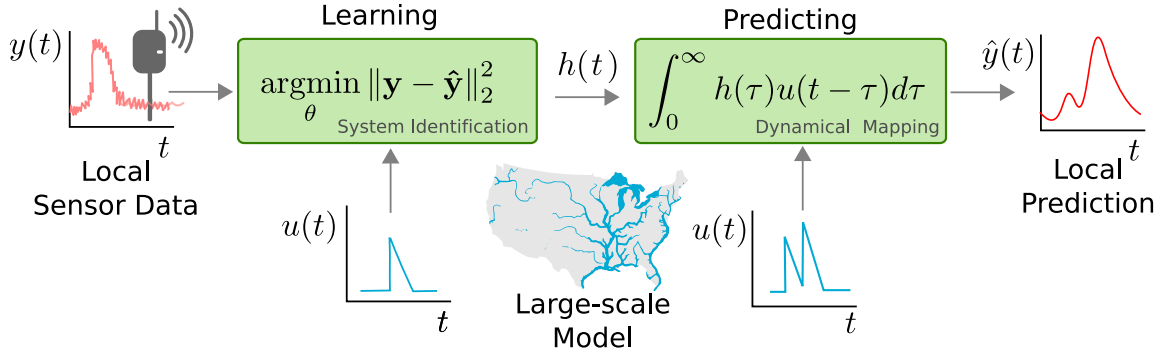


Figure 1.2: Chapter 3 methodology. Historical measurements made by a sensor are used to *learn* a dynamical mapping between modeled and measured values. Once the parameters of the mapping are learned, predictions can be made by dynamically transforming the modeled estimates.

By pairing measured water levels with coarsely modeled flows, this research will demonstrate how to learn a dynamical mapping based on system identification. The outputs of a large numerical model will be used as inputs into a dynamical system, which will predict stream heights that are measured by a sensor. Once the parameters of the dynamical mapping are learned, future forecasts of site-scale conditions will then be generated (Figure 1.2). This methodology will be assessed by using nearly 200 water level sensors across Iowa. A performance classification, based on principal component analysis and boosted random forests, then will also evaluate under which conditions the methodology can be expected to perform well.

This chapter will have two primary contributions. Firstly, this research proposes a general methodology for dynamically downscaling large-scale, coarse hydrological models to site-scale forecasts. This will allow highly-localized forecasts to be derived from existing models and sensor data, to the immediate benefit of local decision makers. Secondly, an extensive site-classification will highlight where this methodology can be used effectively, which is immediately useful to decision makers seeking to forecast local flooding or select locations to place new sensors.

1.3 Model reduction for real-time change detection across complex water systems

Detecting changes across hydraulic and hydrologic systems is often difficult due to the large number of states and parameters that underpin their dynamics [10]. At the system scale, changes may often be detected by sensors, but the exact source of the change is often unclear because the sensors are measuring a signal comprised of many additive subsystems. Numerical models can be re-tuned to determine the source of the change, but this can be difficult due to their highly structured or parameterized nature. Determining where and why a change occurred can require an exhaustive search over millions of possible parameterizations. To that end, the contribution of this chapter is a methodology by which a simpler linearized model, with much fewer parameters, is abstracted from a networked water system, identified through sensor data, and then used to detect where the water system may have changed (Figure 1.3).

Detecting and locating system changes is of great importance as the world's population continues to concentrate in urban areas [11]. Much of the water infrastructure in the United States' and other developed countries' urban areas are approaching or have exceeded their design lives [12]. One major, costly challenge facing city managers is system maintenance or asset management. Pipes often clog or break and basins fill up with sediment, and these events are rarely identified in real-time. This compromises system performance and can lead to flooding and water quality impairments. To address these concerns, many cities make use of the EPA's Stormwater Management Model (SWMM) for their management and planning of stormwater assets. However, this numerical model is highly parameterized. Therefore, when an anomaly is detected, it is difficult to determine where a pipe may be blocked or a pond may need dredging without running many simulations or blanketing the network in sensors. Computationally efficient methods that do not require many observations are needed to enable improved asset management, especially in the real-time

setting.

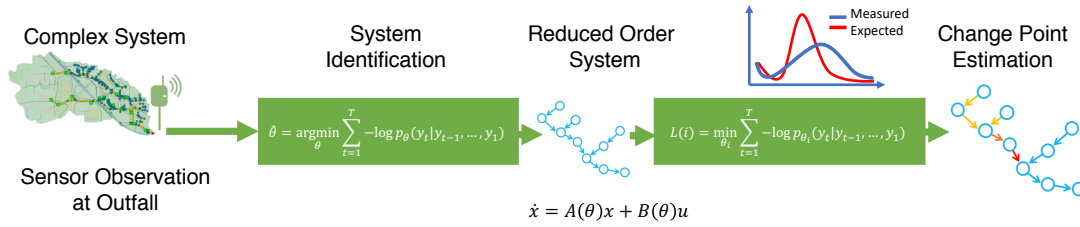


Figure 1.3: Chapter 4 methodology. A linearized model is identified from sensor data and an iterative search is performed to determine changes in the systems. This enables the localization of damage across the system.

The goal of this chapter is to establish a foundation for real-time asset management systems – one that is computationally efficient and applicable when real-time data is spatially sparse. For this case study, it will be assumed that a single sensor at the outlet of an urban watershed is measuring flow. Based solely on the networked connectivity of the stormwater system (only storage nodes and links), a linear system of differential equations can be assembled with a relatively small number of parameters compared to a traditional model. Then, using system identification, a sparse state space representation of the watershed can be parameterized. When an anomalous signal is detected at the outlet, an efficient search can be carried out over the linearized model. This search will use maximum likelihood estimation, enabling the probabilistic identification of sites that may have changed in the system. Using these likelihood scores, sources of damage across the system can be proposed.

This chapter has two primary contributions. Firstly, this research will provide a formal approach for abstracting simplified models from complex networked water systems with the ability to support damage detection. Secondly, it will result in a decision support tool for real-time asset management, which is urgently needed by municipalities to manage urban watersheds and infrastructure.

CHAPTER 2

Integrating Large and Noisy Sensor Datasets with Numerical Models

2.1 Introduction

Advances in sensing are transforming the measurement and understanding of water resources. However, site access and resource constraints still challenge the ability of in-situ observatories to resolve the spatiotemporal variability of many hydrologic systems. This is particularly true across large surface water bodies, such as many of Earth's large lakes, where strong seasonality and the sheer size of study areas limit the permanent and spatially dense deployment of observing platforms [13].

To that end, we ask the question: How can new and non-traditional sensor measurements, such as those made by volunteer ship captains, be used to improve hydrometeorological estimates across large surface water systems? This question is answered through the analysis of one of the largest such data sets: an unprecedented collection of approximately one million unique measurements made by ships on the North American Great Lakes over 2006-2014. The contribution of this chapter is a flexible probabilistic framework which can be used to distribute ship measurements, or any other general sets of irregular or Lagrangian point measurements, into contiguous gridded datasets. The performance of this probabilistic framework is assessed through the development of a new ship-based spatial

data product of surface water temperature, air temperature, and wind speed across the Laurentian Great Lakes. The generality and flexibility of this approach enables it to be applied to many similar data sets and will be of use to those seeking to merge large collections of measurements with other sources of data, such as physical models, satellite products, and even data collected by drones [14]. Direct applications could include, but are not limited to, the assimilation of similar ship data across other large surface water systems, or the use of stationary sensor network data in applications such as snowpack estimation or flood forecasting.

2.1.1 Motivation: The Great Lakes

The North American Great Lakes comprise a vast hydrologic network whose daily hydroclimate impacts over 30 million Canadian and American residents in the region [15]. Collectively, the Great Lakes basin accounts for 90 percent of the United States' and 20 percent of the world's fresh surface water supplies, while simultaneously housing one-third of both the United States' population and Gross State Product [16]. Only a small number of other freshwater bodies, such as Lake Baikal or the African Great Lakes, compare in size, making the North American Great Lakes one of the largest sources of surface fresh water on the planet. This underscores the need to urgently improve understanding of the Great Lakes water budget in an uncertain climate [17, 18]. Many operational and research models have been developed to study and predict hydroclimatic conditions on the Lakes, ranging from water levels using Net Basin Supplies [19] to harmful algal blooms using HAB Tracker [20]. Lake level models, in particular, have been of interest due to their impact on shipping, agriculture, power production, recreation, and real estate [21].

Surface water covers over a third of the total watershed area; no other large basin in the world has a comparable land-to-water surface area ratio [13]. As such, an understanding of the energy fluxes at the air-lake interface is critical toward assessing the potential impacts of climate change on the Great Lakes water balance. It is suspected that

evaporation has been the major contributor to recent lake level fluctuations, a hypothesis presently being tested through an expansion of evaporation measurements throughout the Great Lakes Basin [13]. In general, data availability presents one of the major challenges to hydrometeorological studies on the Great Lakes, as existing in-situ measurements are limited to seasonal buoys and a small number of permanent stations. A number of studies have demonstrated remotely sensed data's utility in calibration and validation of models [22]. While satellite observations are showing great promise to help fill these observational gaps [23, 24], their calibration and downscaling could nonetheless still benefit significantly from ground-based measurements. Without additional overlake measurements, especially away from shore at hard-to-reach lake centers, it becomes difficult to confidently assess the impacts of climate on the water balance of the Great Lakes.

2.1.2 Existing Models and Data

A large number of observational stations and buoys on the Great Lakes are maintained by the National Data Buoy Center (NDBC) and equipped with common hydrometeorological sensors, including those measuring wind speed, wind direction, air temperature, atmospheric pressure, and significant wave height (Figure 2.1). There is also a smaller contingent of buoys and lighthouse stations that are equipped with advanced instrumentation for specialized studies, such as those related to energy fluxes [25, 26]. All of these data are freely available on NOAA's NDBC website (ndbc.noaa.gov). Unfortunately, many of these measurements are not available throughout the entire year, making it difficult to study many phenomena during the late fall and winter, a period during which energy fluxes such as evaporation are known to be at their peak [27]. Buoys are generally deployed only during the summer and fall months (May through November) to avoid issues with ice floes and heavy seas, and they are generally moored near-shore for ease of maintenance and avoidance of shipping traffic, while permanent stations are often located on islands or shallow reefs and thus limited by bathymetry.

Questions that cannot be addressed via measurements alone have been addressed via models, which simulate over lake conditions at locations that are not instrumented. Physical models require accurate observations for calibration and validation. Since over-lake monitoring networks are sparse when compared to terrestrial observations, the ability to validate off-shore dynamics is very limited. Some approaches address this by distributing terrestrial observations using combined data sources such as remote sensing and in-situ measurements [28, 29, 16], while others use regional climate models to capture the over-lake dynamics [30].

The National Digital Forecast Database (NDFD) is one of the primary forcing inputs used for physical models of hydrometeorological variables across the Great Lakes. It is a regionally-focused model that deterministically outputs meteorological estimates based on measurements from regional weather stations [31]. One model that uses the NDFD for forcing is the Great Lakes Coastal Forecasting System, which models a suite of variables not covered by the NDFD such as the temperature profile of the lakes, significant wave height, and other variables of interest to researchers and stakeholders [32]. As with the NDFD, the GLCFS is a deterministic model underpinned by assumptions about overlake conditions, for which measurements are unavailable. Given this lack of measurements, it becomes difficult to provide error estimates on the performance of these physical models. As such, there is significant consensus that more over-lake observations are needed to better address fundamental questions underpinning the short- and long-term variability of hydrometeorological phenomena across the Great Lakes basin [13].

2.1.3 An untapped data source

Many ships on the Great Lakes are equipped with meteorological instruments for navigational purposes. Much of this data feeds into the Volunteer Observing Ships (VOS) program which has been archiving Great Lakes data from shipping, research, and coast guard vessels since 1987. These data are freely available through NOAA's CoastWatch

system. The primary goal of the global VOS program is to provide data to organizations invested in understanding weather and climate, particularly as they relate to commerce, extreme events, and the safety of life at sea. Many of the maritime routes traversed by these ships are active throughout the entire year and pass through locations that are not measured by the existing buoy network, including the center of each lake. Over the past decade, the data comprise over 400,000 unique ship reports, each measuring from one to ten variables. Depending on the vessel, these may include air temperature, water temperature, wind speed, wind direction, dew point, significant wave height, wave period, cloud cover, solar radiation, and/or barometric pressure. A coverage map summarizing these measurements is shown in in Figure 2.1.

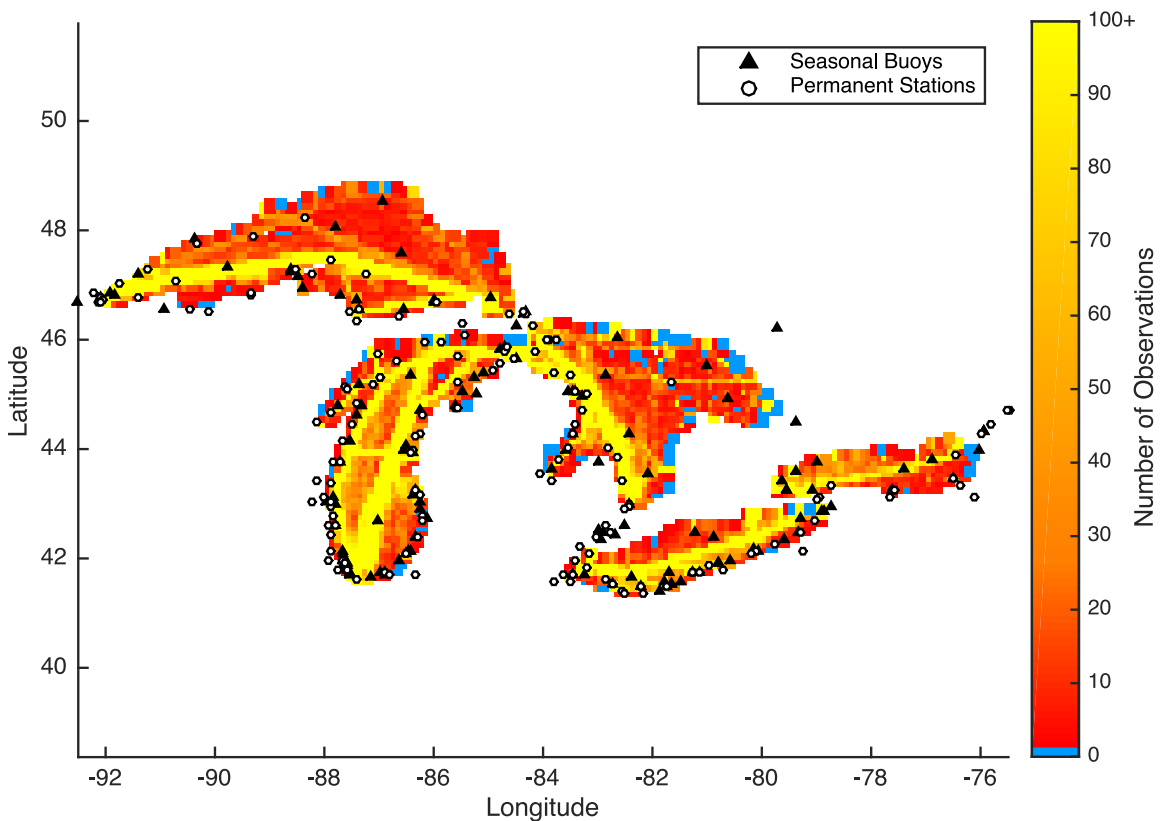


Figure 2.1: Comparison of ship measurement densities to the stationary observations. Red color indicates few measurements, yellow corresponds with locations which contain over 100 measurements, while blue indicates no available ship measurements.

The VOS program has been used in a number of oceanic studies, primarily in the North

Atlantic and North Pacific oceans [33]. [34] used a semivariogram method to determine the spatial variability and uncertainty in ship observations. This methodology relies solely on observations and does not use physical models to inform an initial estimate of the state, thus requiring many observations to capture both the observation noise and the spatial variability. In another oceanic study, [35] used VOS data to correct subjective observations of swell, significant wave height, and wind speed to improve hindcasts of wind waves in the North Atlantic and North Pacific. To the author's knowledge, the vessel data have not, however, been comprehensively used on the Great Lakes as part of a rigorous probabilistic framework.

While vessel measurements provide valuable information at locations that would otherwise remain unobserved, these measurements still inherently comprise a collection of point observations, which must be distributed throughout space and time to improve the understanding of fine-grained, lake-scale phenomena. Given the variety of existing VOS measurements, their non-stationary nature, delayed reporting, as well as the number, complexity, and variety of physical models across the Great Lakes, a traditional on-line data assimilation approach (e.g. [36]) quickly becomes impractical. Furthermore, many spatial data products, such as those measured by satellites, do not depend on a physical model but could still benefit from being updated by spatially-distributed ground observations. To that end, this chapter presents a much more general and flexible framework, based on probabilistic processes, to address all these challenges and fuse the variety of data sources. The approach uses prior information provided by a physical model, a combination of physical models, or other spatial data products and then spatially distributes information from measurements, such as those made by ships, to improve estimates of system states. This not only enables us to make predictions at unobserved locations, but also provides invaluable variance estimates that can be used to characterize the uncertainty of these predictions. Further, the implementation does not require intimate knowledge of a physical model, access to model source code, or the need to re-run an underlying physical model, which often

poses a major obstacle in many research projects.

2.2 Materials and Methods

2.2.1 Spatial Estimation

Our estimation problem is framed by the need to condition a field $f(\mathbf{x}) : \mathbb{R}^d \rightarrow \mathbb{R}$, which maps a d -dimensional input vector \mathbf{x} to a phenomenon $f(\mathbf{x})$ of interest, based on a set of observations y with zero-mean Gaussian measurement error $\varepsilon \sim \mathbb{N}(0, \sigma_n^2)$. Formally:

$$y = f(\mathbf{x}) + \varepsilon \tag{2.1}$$

Represented graphically (Figure 2.2), each nonlinear mapping of the field $f(\mathbf{x})$ is a process of interest, such as wind speed or water temperature. The dimension of the input vector \mathbf{x} is governed by features that are known to affect the process of interest. They can include, but are not limited to, latitudes, longitudes, estimates made by physical models (e.g. NDFD or GLCFS), or time. Therefore the goal is to fuse the observations from the boats (top of Figure 2.2, y) with the model estimates (bottom of Figure 2.2, \mathbf{x}) and yield a new field (middle of Figure 2.2, $f(\mathbf{x})$). The parameterization of the mapping function $f(\mathbf{x})$ is not known *a priori* and must be conditioned or *learned* using the set of sparse and noisy ship observations y . This mapping may depend on highly nonlinear and unknown relationships, which limits the use of common estimation methods that assume explicit mappings, such as linear regression or optimal interpolation [37]. Once the mapping is learned, it can be used to derive estimates at locations where ship observations are not available.

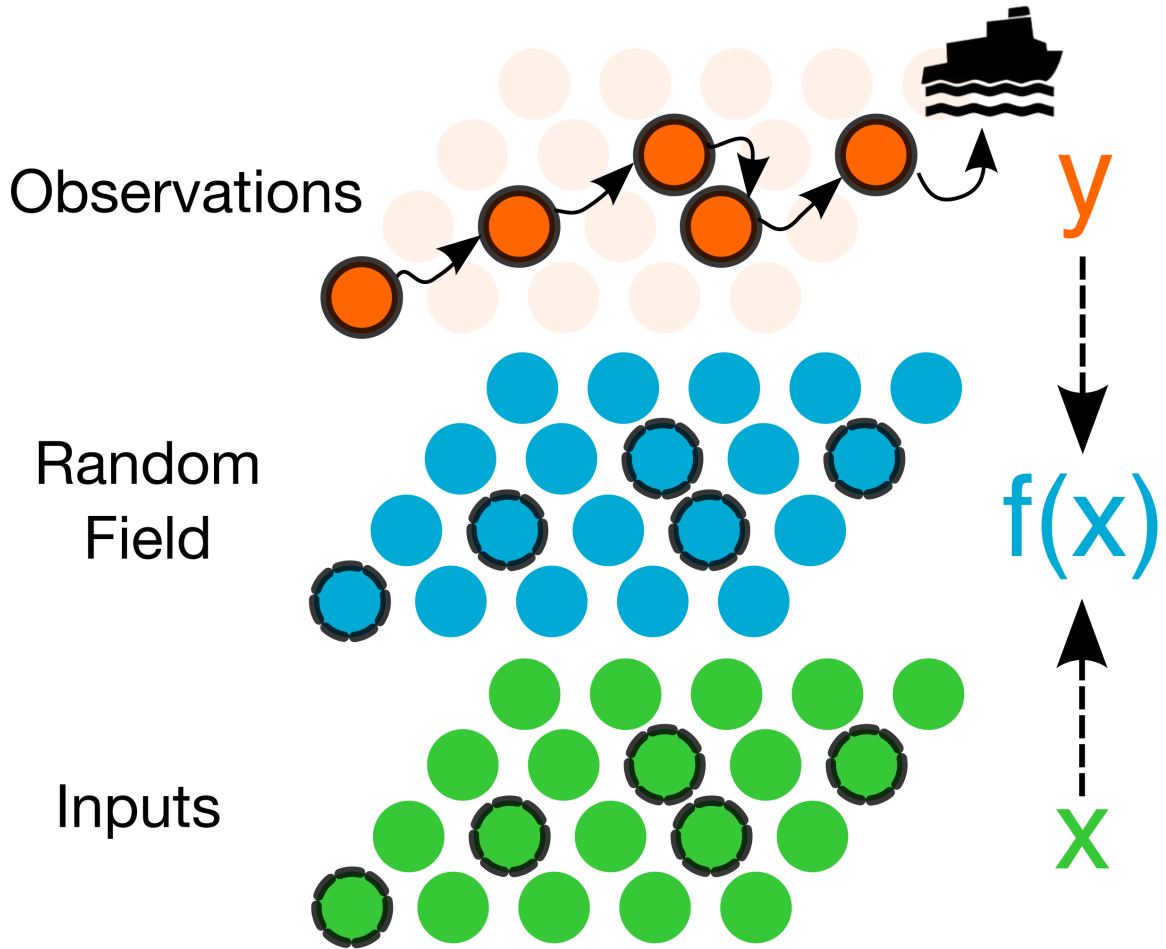


Figure 2.2: Probabilistic Methodology. Inform an estimate of the process $f(\mathbf{X})$ by matching inputs \mathbf{X} to co-located ship observations y .

Probabilistically, an infinite number of mappings may characterize the input-output relationship. As such, rather than finding an explicit relation, it is more desirable to characterize the distribution over these mappings and generate a random field onto which the data can be projected to derive new estimates (Figure 2.2). This field can be fully described by its mean and covariance functions:

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})] \quad (2.2)$$

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))] \quad (2.3)$$

By assuming a zero mean probabilistic process, these equations form a *Gaussian Process* (GP) [37], which is fully characterized by its covariance, or *kernel*, function $k(\cdot, \cdot)$. A zero-mean process can, for example, be obtained by subtracting the ship measurement from a co-located prediction of a physical model. In this case, the resulting probabilistic process becomes analogous to bias-correcting that can also be used to characterize the uncertainty of the underlying physical model.

It can be shown that rather than assuming an explicit mapping for $f(\mathbf{x})$, a much more powerful regression framework can be obtained by focusing instead on choosing a kernel $k(\cdot, \cdot)$ and learning its hyperparameters. Hyperparameters are simply the parameters of the kernel function and are called such to distinguish them from the parameters of the underlying model (in this case, $y = f(\mathbf{x}) + \varepsilon$). An example covariance function for two inputs \mathbf{x}_p and \mathbf{x}_q is given by the squared exponential kernel:

$$k(\mathbf{x}_p, \mathbf{x}_q) = \exp \left(-1/2(\mathbf{x}_p - \mathbf{x}_q)^T \begin{bmatrix} l_1^{-2} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & l_d^{-2} \end{bmatrix} (\mathbf{x}_p - \mathbf{x}_q) \right) + \sigma_n^2 \delta_{pq} \quad (2.4)$$

with hyperparameters σ_n^2 and l , where σ_n^2 is the measurement noise associated with the n -th observation, l is a vector of the characteristic length scales l_i of each input feature, and δ_{pq} is the Dirac delta function. In this parameterization the length scales are spatially analogous to the radius of influence of a measurement. When individual length scales are used in lieu of a single length scale, the relative magnitude of each length scale indicates the relative importance of each input to the predictive model. Therefore this kernel yields a method by which to infer the relevance of individual inputs, which in turn helps provide insight about the output of interest. For example, a relatively shorter length scale for a given feature in \mathbf{x} would indicate that this input variable may be more informative in explaining the output y , while a relatively longer length scale suggests that a feature could

be entirely eliminated from future analyses to improve computational efficiency.

The use of kernels not only provides significant computational benefits but enables a much richer set of basis functions upon which the process can be trained. For example, it can be shown that the choice of some kernels, such as the squared exponential kernel, is equivalent to regression across an infinite set of basis functions, compared to a limited set that would be obtained using classical linear regression [37].

The task, then, is to *learn* the hyperparameters of the kernel given a set of observed data, in this case the ship measurements. Using the n observations from the ships as the output ($\mathbf{y} \in \mathbb{R}^n$) and a matrix of co-located features (latitude, longitude, physical model forecast, etc.) as the input $\mathbf{X} \in \mathbb{R}^{n \times d}$, a predictive GP model can be *learned* for each hydrometeorological variable (air temperature, wind speed, etc.) by maximizing the marginal likelihood:

$$p(\mathbf{y}|\mathbf{X}) = \int p(\mathbf{y}|f(\mathbf{X}), \mathbf{X})p(f(\mathbf{X})|\mathbf{X})d f(\mathbf{X}) \quad (2.5)$$

where the probability distribution of \mathbf{y} , given $f(\mathbf{X})$ (the posterior), is normally distributed around $f(\mathbf{X})$ with some measurement noise σ_n^2 ($p(\mathbf{y}|f(\mathbf{X})) \sim \mathbb{N}(f(\mathbf{X}), \sigma_n^2 I)$, where I is the identity matrix) and the probability of $f(\mathbf{X})$, given an input \mathbf{X} (the likelihood), is normally distributed around a mean function and a covariance function that must be learned ($p(f(\mathbf{X})|\mathbf{X}) \sim \mathbb{N}(m(\mathbf{X}), \mathbf{K})$ where $\mathbf{K} \in \mathbb{R}^{n \times n}$ with $\mathbf{K}_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$). It is important to note that the only assumption is that the noise ε of the ship measurements is normally distributed, which is the case for many real-world sensors, and given a Gaussian likelihood, the resulting posterior distribution is then Gaussian. The kernel is learned by maximizing this marginal likelihood with respect to the kernel hyperparameters. Given the normally distributed noise assumption, which can readily be justified given the measurement error exhibited by real-world sensors, a closed form solution can be obtained for the above integral [37]. Because the logarithm function is monotonic, the log of the marginal

likelihood will achieve its maxima at the same points as the marginal likelihood function. The solution to Equation 2.5 is an exponential function, so taking the logarithm eliminates the exponential while also making any multiplicative terms additive. This enables us to find the hyperparameters that minimize the negative log-marginal-likelihood (maximize the positive log-marginal-likelihood) in a computationally efficient manner though the use of a gradient descent optimization algorithm. The log-marginal-likelihood is given by:

$$\log p(\mathbf{y}|\mathbf{X}) = -\frac{1}{2}\mathbf{y}^T\mathbf{K}^{-1}\mathbf{y} - \frac{1}{2}\log|\mathbf{K}| - \frac{n}{2}\log 2\pi \quad (2.6)$$

Once the kernel hyperparameters have been *learned*, the joint distribution can be used to spatially distribute the ship measurements and derive variance (uncertainty) estimates across the study area:

$$\begin{bmatrix} \mathbf{y} \\ f(\mathbf{X}') \end{bmatrix} \sim \mathbb{N} \left(m(\mathbf{X}), \begin{bmatrix} \mathbf{K} + \sigma_n^2 \mathbf{I} & \mathbf{K}' \\ \mathbf{K}'^T & \mathbf{K}'' \end{bmatrix} \right) \quad (2.7)$$

where \mathbf{X} are the inputs co-located with observations \mathbf{y} , $\mathbf{X}' \in \mathbb{R}^{m \times d}$ is a matrix of inputs for the m locations where measurements are not available, $\mathbf{K}' \in \mathbb{R}^{n \times m}$ with $\mathbf{K}'_{ij} = k(\mathbf{x}_i, \mathbf{x}'_j)$ is the covariance between the n observed locations and the m unobserved locations, $\mathbf{K}'^T \in \mathbb{R}^{m \times n}$ is the transpose of \mathbf{K}' , and $\mathbf{K}'' \in \mathbb{R}^{m \times m}$ is the covariance matrix between each of the unobserved locations. The above distribution can then be used to arrive at a set of predictive equations that can be used to make estimates at all unobserved locations:

$$f(\mathbf{X}') = \mathbf{K}'^T(\mathbf{K} + \sigma_n^2\mathbf{I})^{-1}\mathbf{y} \quad (2.8)$$

$$\text{cov}(f(\mathbf{X}')) = \mathbf{K}'' - \mathbf{K}'^T(\mathbf{K} + \sigma_n^2\mathbf{I})^{-1}\mathbf{K}' \quad (2.9)$$

2.2.2 Implementation

The proposed method was used to spatially distribute ship measurements of wind speed, air temperature and lake surface temperature from 2006 to 2014 across the Great Lakes. The choice to focus on these variables was motivated by their large role in lake evaporation and the broader lake energy balance [27]. While the fundamental mechanisms and interaction between these variables are beyond the scope of this chapter, the resulting spatial data sets are expected to be of significant value in future studies of lake hydroclimate. This chapter focuses primarily on the evaluation of the proposed probabilistic method. Estimates of wind speed and air temperature, co-located spatiotemporally with measurements made by the ships, were retrieved from the NDFD, which provides outputs on an hourly basis across a 5km grid. The NDFD generates a regionally-focused model at each of the 122 weather forecasting offices nationwide, with a dozen of these located in the Great Lakes region [31]. The NDFD also interpolates or smooths model outputs at boundaries between weather forecasting offices. lake surface temperature estimates were obtained from the GLCFS since the NDFD does not model this variable. The starting year of the analysis was chosen because the GLCFS model became operational in 2006. While these physical models were chosen due to regional popularity, the methods presented herein can readily be repeated using other physical models or combinations of models. Furthermore, the use of two distinct physical models in this study also highlights the flexibility of the proposed framework, showing that very limited overhead is required to apply the method to different physical models.

To determine if there are benefits to be gained from the proposed probabilistic method, an initial assessment was first carried out to determine how well physical models alone perform across the study region. A mean absolute error (MAE) analysis was used to compare physical model outputs to the ship measurements. The physical models were resampled via averaging to reflect the coarser 0.1 degree resolution of the ship reports. The MAE was calculated for any grid cell in which one or more ship measurements were available. No

time averaging was performed. The ship measurements were simply compared to the outputs of their corresponding model grid cell. The results were used to determine any spatial or seasonal biases that may be present in the physical model, treating the ship observations as the true value of the underlying hydrometeorological variable.

An initial visual inspection revealed relatively consistent spatial MAE patterns during different seasons of each year. As a result, before training the GP algorithm, the ship data were separated into seasons (January-March winters, April-June springs, July-September summers, October-December falls). The choice to bin the ship data into seasons was motivated by computational complexity. Initially, a set of GP models was evaluated in which time was used as one of the input variables. This provided no discernible improvement over the seasonally binned approach. This choice to bin is justified because binning into seasons implicitly captures the temporal variation as seen in the MAE plots (Figures 2.3, 2.4, and 2.5). More practically, the assimilation of all of the measurements into one probabilistic model would have come at a high computational cost, so learning separate models for each season enables faster compute times and more temporally focused models.

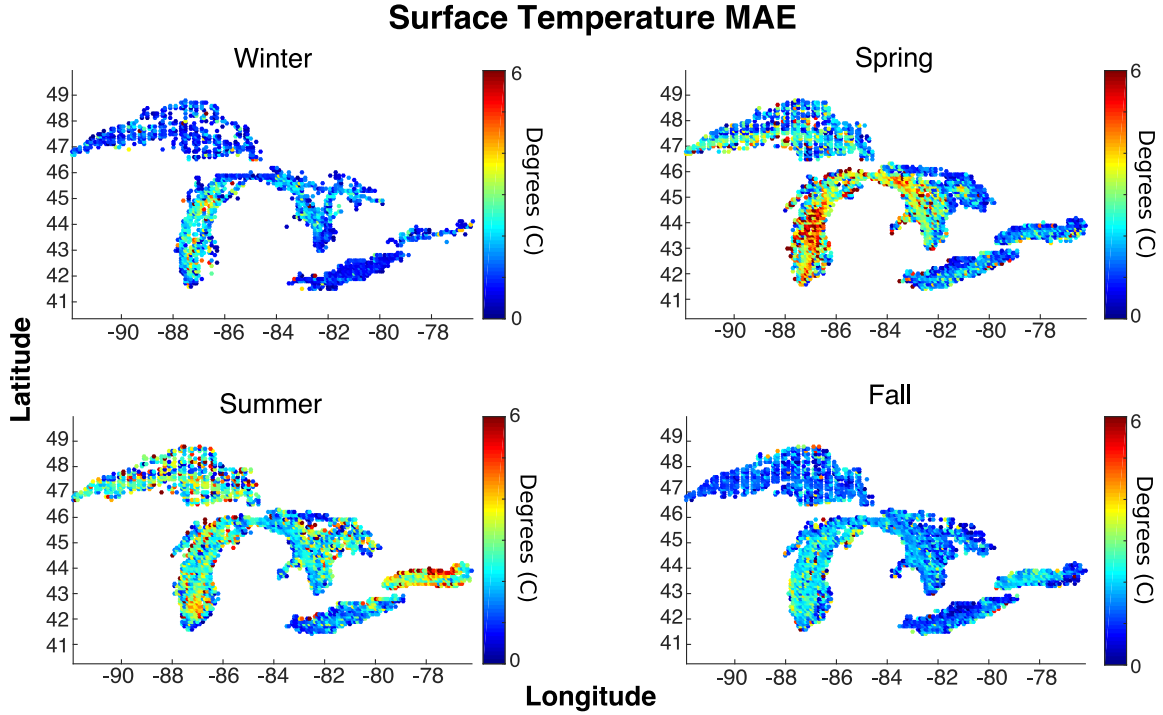


Figure 2.3: Seasonal Mean Absolute Error for Lake Surface Temperature, comparing the measured value from a ship to the co-located physically modeled estimate, averaged over each instance at that grid cell.

The inputs, \mathbf{X} , to the GP framework are defined as the latitude, longitude, and physical model estimate from either the NDFD (air temperature and wind speed) or GLCFS (lake surface temperature). To satisfy the zero mean GP assumption, the output, y , is defined as the difference between the physical model estimate (wind speed, air temperature, SST) and co-located ship observation. A squared exponential kernel was used with a unique length-scale for each input. The choice of kernel was based on the ability to carry out automatic relevance detection (ARD), which, as mentioned previously, would permit for length scales to be ranked, thus providing insight into which inputs were the most informative for predicting the final output. A shorter length scale suggests that a feature is more important to the prediction than other features. For instance, one might expect that the physically modeled estimates of air temperature would be more important in explaining ship observations than the location of the ship. As such this feature's length would should be shorter than those for latitude and longitude. Additionally, the choice to use a squared

exponential kernel was motivated by its radial basis. This encodes the notion that the covariance between two points is related to the distance between the points. Other kernels were also tested (such as the rational quadratic), but these yielded either lower or similar likelihoods for all seasons and variables. Therefore, the squared exponential kernel is used for all models to maintain consistency across the dataset.

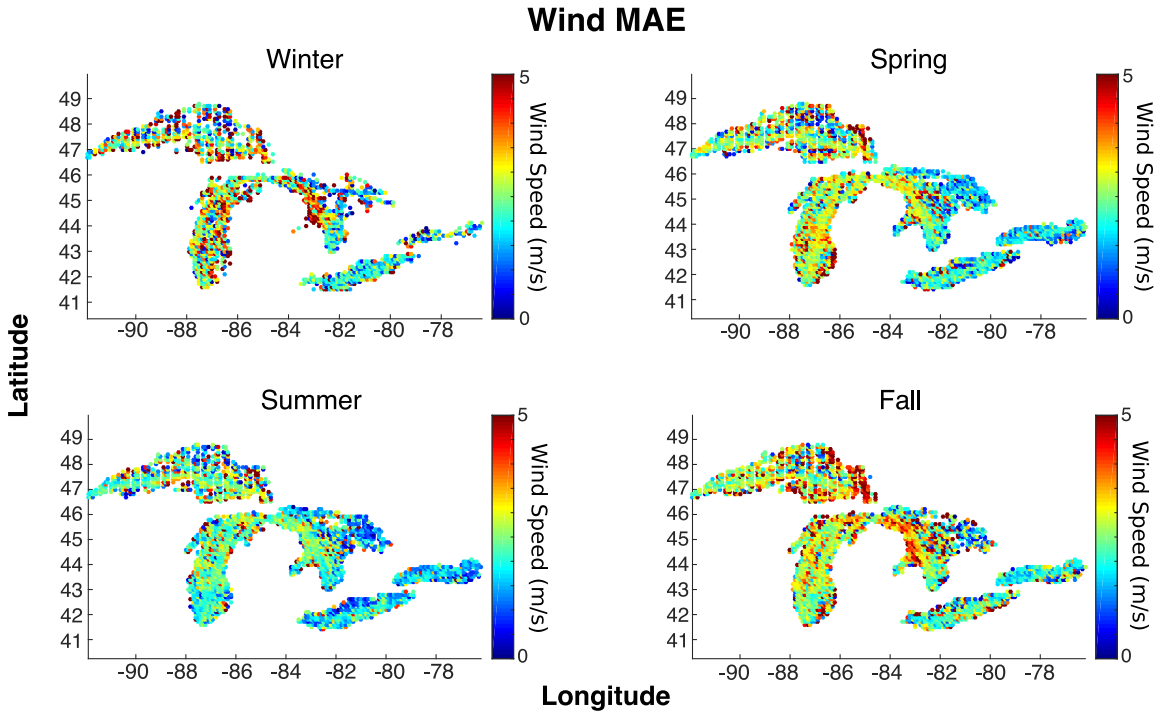


Figure 2.4: Seasonal Mean Absolute Error for Wind Speed. Interpretation follows from Figure 2.3

Next, the input space was normalized to reduce any adverse scaling impacts on the minimization algorithm and to allow for the comparison of the relative importance of each length scale parameter. Without this scaling, a relative comparison of length scale (and thus the relative importance of an input) would be difficult. Scaling also creates a more spherical search space for the likelihood search algorithm. If no scaling was performed, the surface would be more elliptical in shape and the gradient descent may take steps in suboptimal directions, leading to increased computational time [38].

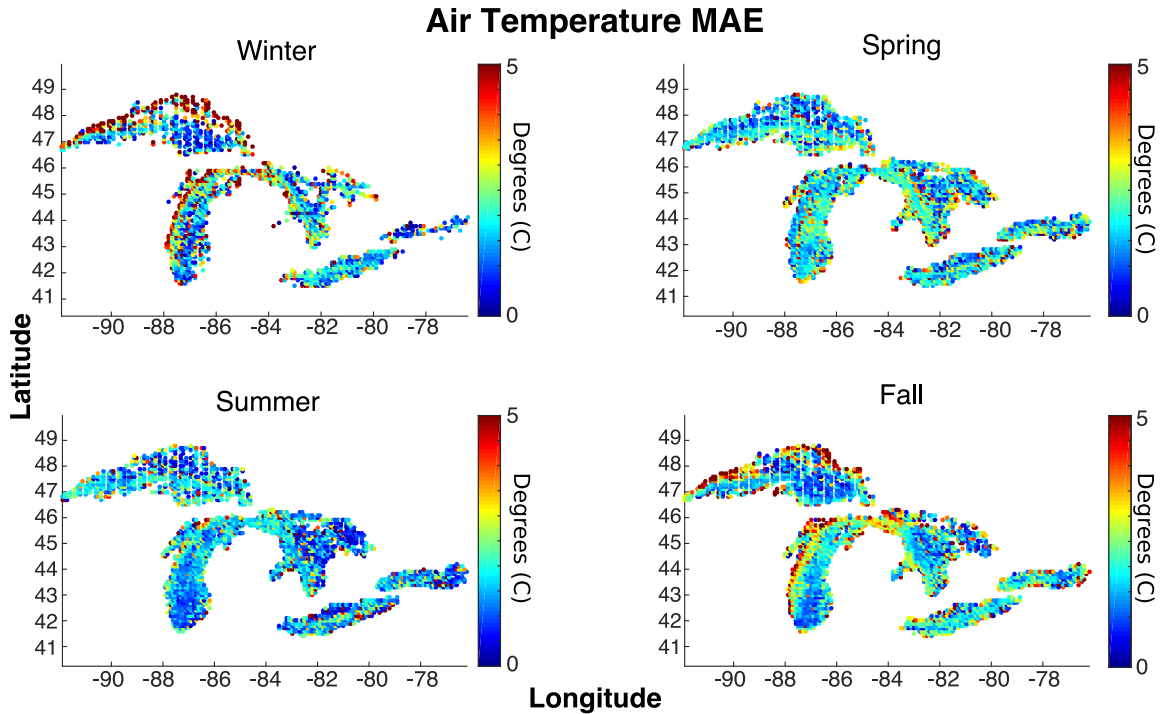


Figure 2.5: Seasonal Mean Absolute Error for Air Temperature. Interpretation follows from Figure 2.3

For each season of ship data, half of the input-output pairs were randomly split to form the training data sets and the minimization algorithm was executed five times. Multiple random restart conditions were carried out to ensure overfitting was not an issue and to reduce the effects of potential local minima. The kernel hyperparameters resulting in the largest maximum marginal likelihood (minimum negative marginal likelihood) were deemed representative of the “best” achievable GP model. The remaining half of the input-output pairs (the testing set) was then used to validate the accuracy of the GP model. While a larger training set could have been used (training sets comprised of 80-90% of observations are not uncommon in machine learning), the choice to use only half the data for training was motivated by a desire to limit overfitting and to determine if the algorithm can perform well even if observations are sparse. Once each algorithm was trained, it was used to make a prediction of the testing data. Given the computational demand of the GP framework, the analysis was executed on the high performance computing cluster at

the University of Michigan, allowing for a parallelization of the required 540 algorithm executions (4 seasons x 3 variables x 5 lakes x 9 years = 540 models).

It is important to note that aside from minimally removing physically impossible values (e.g. negative wind speed) an explicit quality assurance/control (QA/QC) preprocessing procedure was not carried out on the ship observations because the GP framework accounts for such measurements during the training and forecasting step. In fact, the subjective removal of any ship data from the training set can bias the final model and lead to an inaccurate characterization of the uncertainty. This is one of the added benefits of using a probabilistic approach, such as the one presented here: given a sufficient number of training points, any outliers will fall in the tails of the distribution.

Once the “best” GP models (those with the highest marginal likelihood) were selected and validated for each season, a full GP regression using all of the available observations was then carried out to derive hourly estimates (3-hourly for SST) of all hydrometeorological variables across all lakes at 1/10 degree spatiotemporal resolution, using latitude, longitude, and physically modeled values as the inputs to the GP. The result is a spatially distributed data product of lake surface temperature, air temperature, and wind speeds from 2006 to 2014.

2.3 Results

Due to the large size of the data set, the following section provides a summary of the overall analysis. To facilitate transparency and motivate adoption of the proposed methods, all of the input data, the source code of the entire implementation, as well as any resulting data products and additional figures are available on a public web repository (<https://goo.gl/rfGPpt>).

2.3.1 Comparison of Vessel Measurements to Physical Models

Very notable spatial and temporal differences between the physical models and the vessel measurements were evident. For example, lake surface temperature MAE (Figure 2.3) was larger during the summer for all lakes except Lake Michigan, which was highest in the spring. Spatial biases were also evident, showing that the physical model generally differed more from the vessel measurements in open water rather than near-shore. Wind speed predictions (Figure 2.4) made by the physical models on Lake Erie and Lake Ontario agreed with the vessel measurements the most, while the Upper Great Lakes (Superior, Michigan, and Huron) showed less agreement. In general, the MAE was highest off-shore, though some near-shore locations saw large MAE as well (e.g. the junction of Michigan, Huron, and Superior.) MAE patterns for air temperature (Figure 2.5), resembled those seen with lake surface temperature (Figure 2.3). Summer showed the most agreement between the physical models and ship observations, while winter and fall showed higher MAE. When the MAE was averaged over all lakes and seasons, the physical models differed from the vessel measurements on average by 2.39 m/s for windspeed, 1.49° C for air temperature and 1.82° C for SST.

2.3.2 Assimilation Performance

The proposed GP framework, when calibrated on a randomized subset of seasonal vessel observations (training data), outperformed the physical models in predicting the remaining set of vessel observations (testing data), both spatially and temporally. A comparison of the probabilistic algorithm to the physical model is shown in Figures 2.6, 2.7, and 2.8. On the left of each figure, a randomly sampled subset of all of the ship observations (testing data) is compared to co-located physical model estimates. The observations not plotted (training data) are then used to generate the covariance matrix \mathbf{K} (Equation 2.8) and make new estimates. The testing observations are then compared to the probabilistic prediction

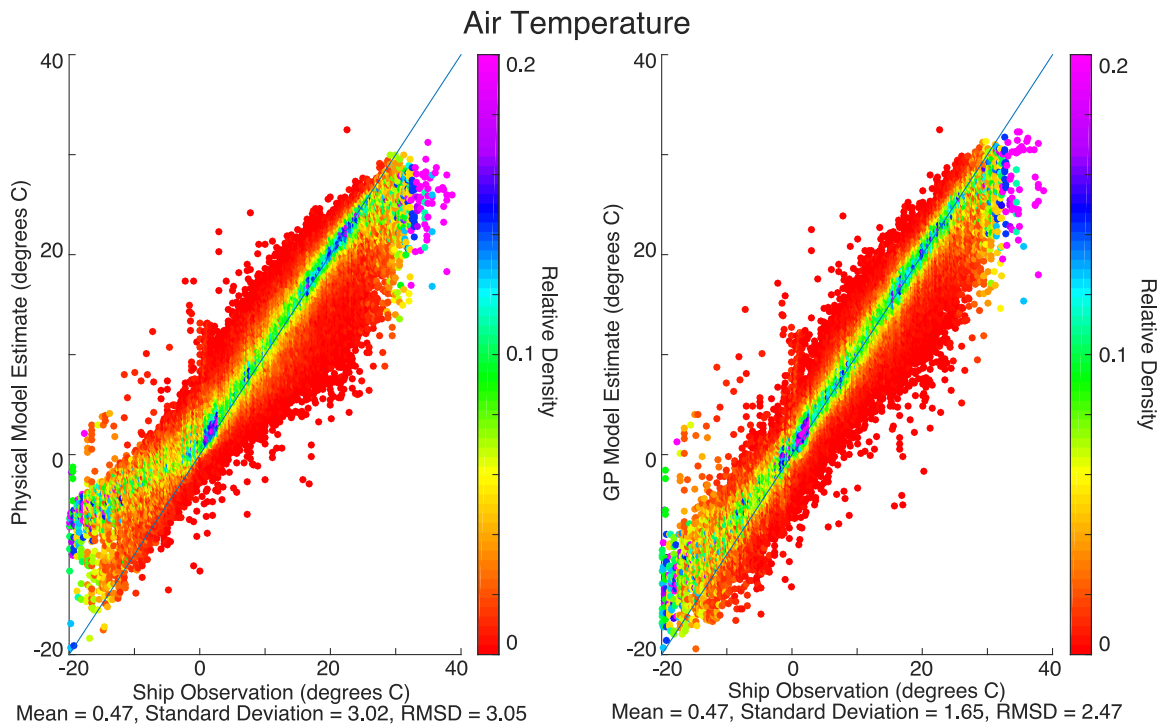


Figure 2.6: Comparing air temperature predicted by the physical model (left panel) and the GP approach (right panel) to ship observations that were not used to train the probabilistic approach. For visualization purposes the data have been binned into 0.1 degree Celsius bins along the x-axis (ship observations) and color-coded as a histogram along the y-axis (model estimate). The color corresponds to the relative density of the data in that bin. In a very good model the highest density of points will fall along the 45-degree line.

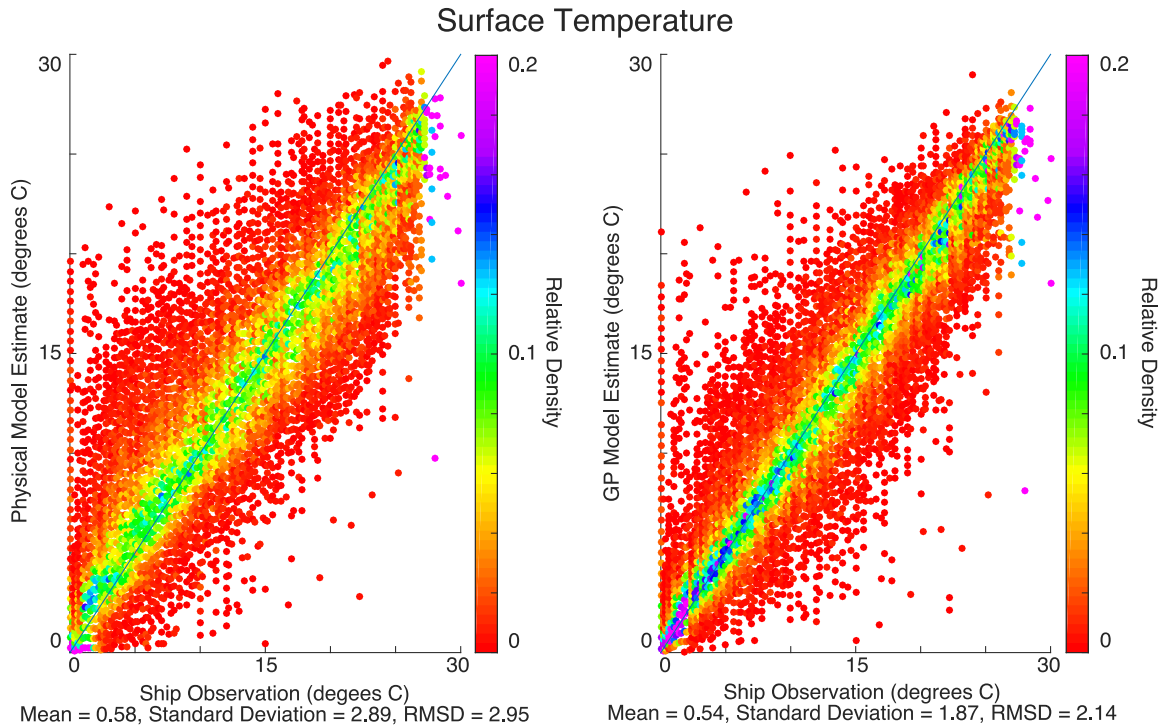


Figure 2.7: Comparing lake surface temperature predicted by the physical model (left panel) and the GP approach (right panel) to ship observations that were not used in training of the GP model.

on the right of each figure. A notable bias was exhibited when evaluating the ability of the physical models to predict vessel measurements. In general, the physical model tended to underestimate air temperature at upper extremes and overestimate at lower extremes (Figure 2.6). A similar tendency was seen with lake surface temperature, where the physical model over-predicted lake surface temperature at lower extremes while under-predicting lake surface temperature at higher extremes (Figure 2.7). For wind speed, a significant mean difference and large variance were also evident when comparing the physical model to the ship measurements (Figure 2.8). Overall, the root mean square difference (RMSD) between the physical model and actual observations for air temperature, lake surface temperature, and wind speed, were 3.06, 2.96, and 3.38, respectively.

When comparing the ability of the GP algorithm to predict the same vessel observations (right plots in Figures 2.6, 2.7, and 2.8), there was an improvement in the mean, variance, and the RMSD of the prediction residuals. For each variable, the GP approach reduced the

biases seen with the physical model at the upper and lower extremes. When compared to the physical model’s performance in predicting air temperatures, the GP approach reduced the mean difference from 0.49°C to 0.46°C, the standard deviation of the difference from 3.02°C to 1.63°C, and the RMSD from 3.06°C to 2.47°C. For lake surface temperature, the mean difference was reduced from 0.58 to 0.52°C, the standard deviation was reduced from 2.9 to 1.83°C, and the RMSD was reduced from 2.96 to 2.14°C. For wind speeds (Figure 2.8), the mean difference was reduced from -0.91 m/s to -0.83 m/s, the standard deviation from 3.25 m/s to 1.52 m/s, and the RMSD from 3.38 to 2.82. Further, for each case of air temperature, lake surface temperature, and wind speed, the pairwise t-test of the residuals between the GP-based outputs and the physical models indicated that the two data sets are significantly different from each other to an α -level of essentially zero (machine epsilon of 2×10^{-16}). Further results are summarized in Table 2.1, where the performance of the GP-based approach in predicting vessel data is compared to the physical model (quantified as percent improvement in the mean difference and variance of the difference).

Table 2.1: Comparing the performance of the proposed approach to physical models in predicting predicting ship observations that were not used in the training procedure. Quantified as the percent reduction in mean difference and variance (averaged over all 45 models for each lake-variable pair).

	Temperature		Wind Speed		SST	
	Mean	Variance	Mean	Variance	Mean	Variance
Erie	65.20	17.60	0.76	17.62	70.61	16.81
Huron	60.92	14.01	77.37	10.97	79.74	16.44
Michigan	70.09	14.89	56.85	6.36	78.13	20.34
Ontario	49.09	14.22	47.18	-7.70	48.13	19.93
Superior	76.65	10.56	80.76	4.23	6.50	17.89

2.3.3 Final Data Product

Upon validating the ability of the GP-based assimilation technique to predict the testing data, all of the vessel observation were combined into a complete data set and used to refine the parametrization of the GP. These final GPs were then used to distribute all of the

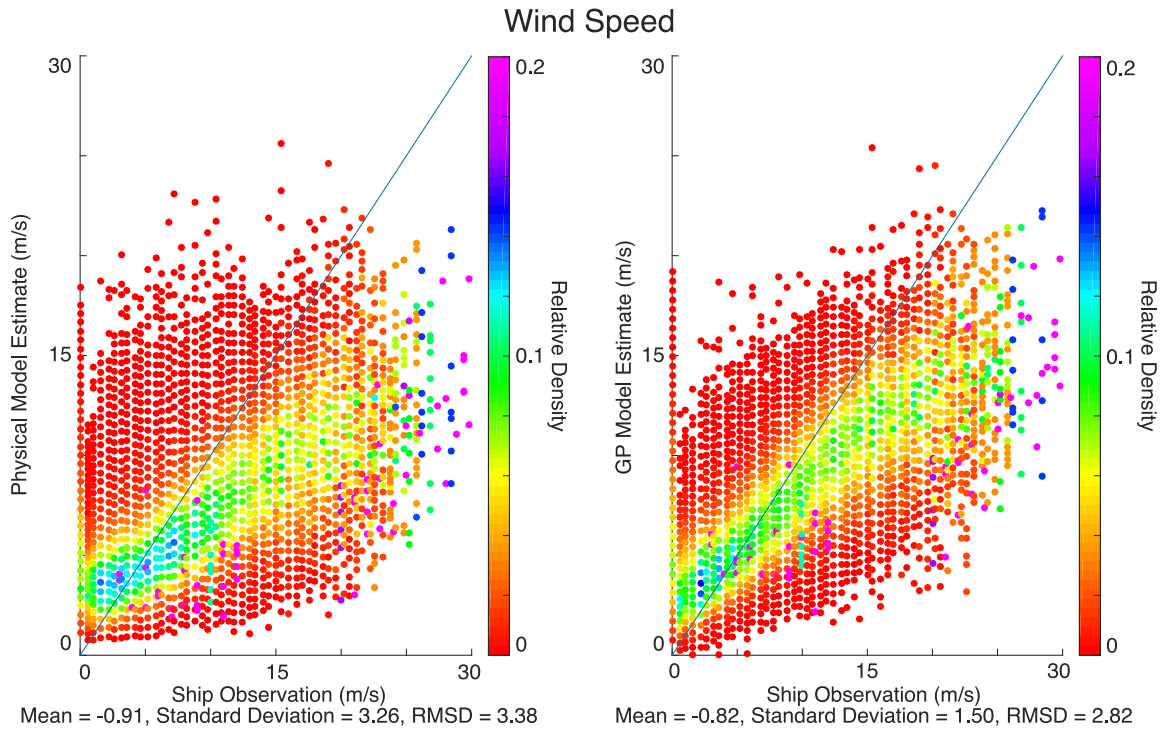


Figure 2.8: Comparing wind speeds predicted by the physical model (left panel) and the GP approach (right pane) to ship observations that were not used in training of the GP model.

ship observations, yielding a final data product for air temperature, wind speed, and lake surface temperature at 1/10 degree spatial and hourly temporal resolution. The run time of the gradient descent training algorithm increased with each season due to an increase in the number of available ship observations (the number of operations required to execute the algorithm increases as a cube of the number of data points [37]). The gradient descent training procedure converged in under an hour for earlier years (2006-2009), while it required upwards of 48 hours to perform all five restarts of the procedure for later years. In most cases, the five randomized gradient descent restarts resulted in nearly identical local minima and hyperparameters. For most variables and seasons, the hyperparameter with the shortest length-scale corresponded with the input of the physical model, implying that the physical model, rather than location in space, was deemed as the most relevant explanatory variable of the vessel observations' covariance. The predictive variance of the GP model was also larger in the earlier years of the analysis.

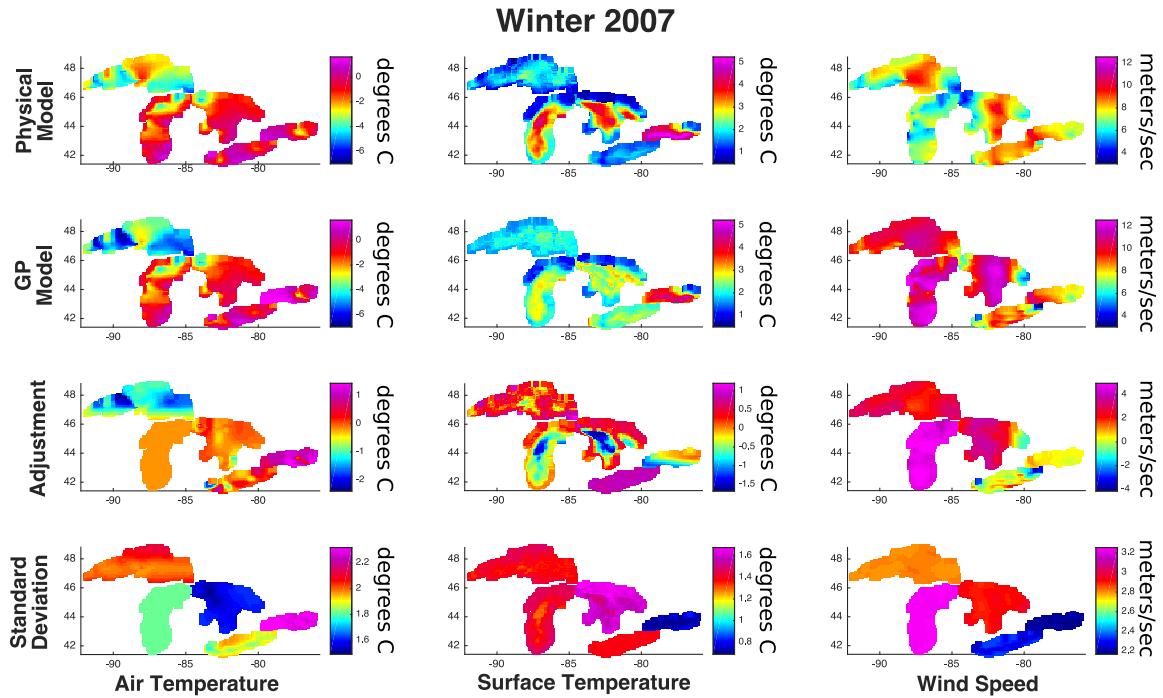


Figure 2.9: Winter 2007 comparison of physical models to the probabilistic method. The first row contains the seasonally averaged physical model estimates for air temperature (first column), lake surface temperature (second column), and wind speed (third column). The second row contains the GP model’s estimate, seasonally averaged across the Winter of 2007. The third row displays the average difference between the physical model and GP approach. The fourth row displays the average predictive standard deviation of the GP model for each grid cell. Please see the data archive to generate more detailed maps.

Seasonal averages for Winter 2007 and Summer 2013 are presented in Figures 2.9 and 2.10, comparing the predictions made by the physical models to that of the GP algorithm while also displaying the uncertainty of the GP algorithm’s predictions (variance in Equation 2.9). To provide insight into how the size of the training set can impact the final model, these two seasons were chosen because they contained the relatively least (Winter 2007) and most (Summer 2013) number of ship observations. All other seasons across the study period are plotted in the supplementary information. For winter 2007, there were only 891 air temperature observations, 543 lake surface temperature observations, and 913 wind speed observations. For summer 2013, there were 21,632 air temperature observations, 11,642 lake surface temperature observations, and 21,411 wind speed observations. With approximately twenty times the observations in 2013 compared to 2007, the GP in 2013

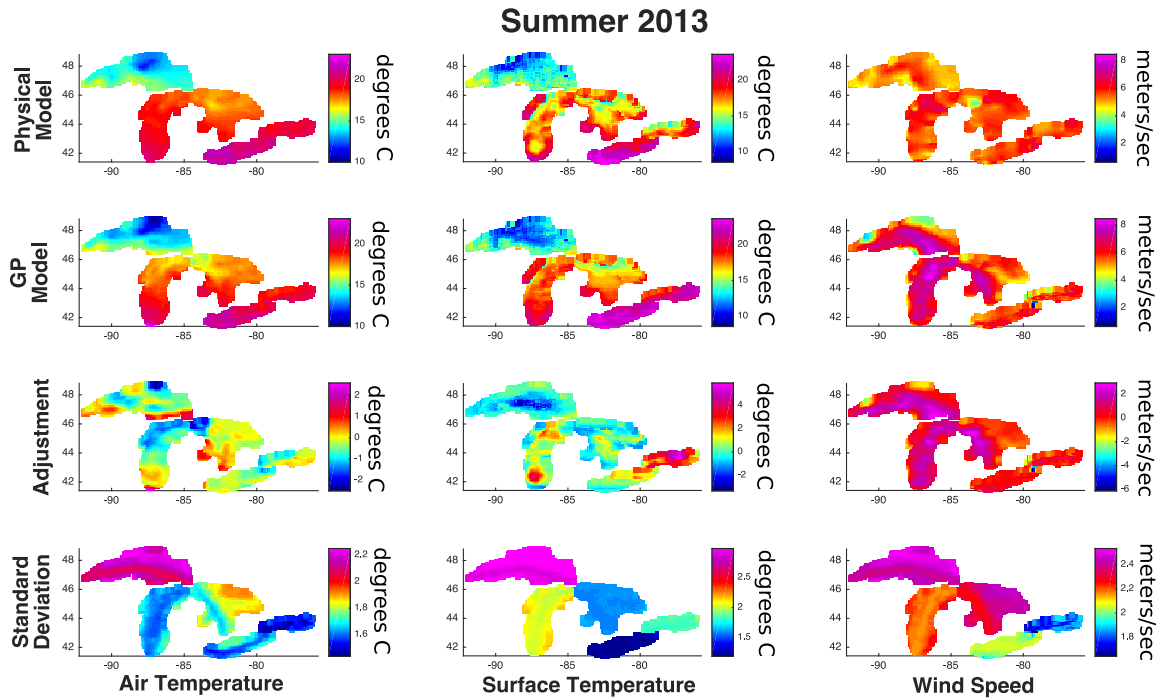


Figure 2.10: Summer 2013 comparison of physical models to the probabilistic method. Interpretation follows that of Figure 2.9.

made higher resolution adjustments. For example, in the Winter of 2007 the GP algorithm performed, on average, uniform adjustment across Lake Michigan (-0.22°C on average), with an average variance of about 1.8°C . For the summer 2013, much finer scale adjustments were made across many regions of the lake, while lower uncertainties aligned with the ship trajectories. Compared to the physical model, the final GP data product showed, on average, a -0.19°C difference in air temperature, a 0.99°C difference in lake surface temperature, and a -0.05 m/s difference in wind speed.

2.4 Discussion

2.4.1 Comparison of Vessel Measurements to Physical Models

Limited patterns in the spatial error structure of the physical models (Figures 2.3, 2.4 and 2.5) suggest that a simple one-to-one comparison may not shed much light onto which

specific factors govern the performance of the physical models. This, along with the noise of the vessel observations and significant complexity of the physical models, makes it difficult to use the ship measurements directly to highlight times or locations where the models need improvement, which may, in part, also explain why these ship observations have seen limited use across the Great Lakes. While the difference between physical models and vessel measurements did correlate in numerous instances with distance from shore (larger difference further away from shore), this pattern was not consistent through all seasons and lakes. A lack of correlation with other factors, such as bathymetry or location of buoys, thus makes it difficult to provide informed conclusions on the observed difference between the ship data and the physical models.

In many instances, MAE was less variable in locations that aligned with ship trajectories, suggesting that the availability of more measurements in those locations led to more accurate estimates of the spatial bias in physical models. In other words, lake regions that were visited by fewer ships only provided few measurements with which to calculate MAE, thus leading to more variability of MAE in those locations. This, however, was not always true across all lakes and seasons and suggest that the MAE is governed by more than just spatial factors.

Similarly, the lack of temporal consistency in the error pattern also suggests that the performance of the physical models may at times be biased by seasonal parameterizations, which are impacted by buoy data availability. While some lakes (e.g. surface temperature on Lake Michigan, Figure 2.3) did exhibit pronounced seasonal differences between the physical models and ship measurements, such patterns or their magnitude were not consistent year-to-year. Additionally, it is worth noting that Lake Erie most likely does not experience drastic MAE seasonality due to the greater density of buoys available for calibration of the physical models as well as the considerably greater amount of resources put into the operational model for this lake [32].

Additionally, the observed error structure may in fact be stochastic and change over

time, as is evident in the varying performance of the physical model during different seasons. Capturing this variability may be difficult using a physical model alone, which is why the proposed GP framework provides a powerful alternative to learning, and correcting for, these seasonal patterns. The one notable consistency in MAE between the physical model and vessel observations was evident through a bias of the physical model at the extremes (Figures 2.6, 2.7, and 2.8). For example, given that the physical model consistently over-predicts vessel air temperature at the lower extremes, the outputs of the physical model can be adjusted at the lower extremes to be more representative of what the ships are measuring. Such an adjustment should not, however, be made haphazardly or without statistical certainty, as these biases may be more pronounced across different locations and at different times. This further supports the use of a holistic statistical estimation framework.

2.4.2 Assimilation Performance

Recall that the feature inputs to this model (\mathbf{x}) are chosen based on the belief that they influence the variable of interest. Therefore, by choosing some initial estimate from a physical model as one of the features, the implied belief is that the models do a good job of explaining physical processes on the lakes, but that they could be improved by embedding new information based on spatiotemporal biases (latitude, longitude, and season) and ship observations. Therefore, the estimates that are generated should remove some of the predictive variance from the initial estimate. When compared to the physical model, the performance of the GP algorithm in predicting ship measurements of air temperature (Figure 2.6) and lake surface temperature (Figure 2.7) suggests that the probabilistically-based approach does indeed provide a robust and reliable framework by which to distribute vessel measurements through the entire spatial extent of the study area. When the hyperparameters are learned, the final GP makes use of all available data sources, fusing the outputs of the physical models (and, implicitly, the buoy data used for calibration) with the additional information provided by the vessel observations. This is true not only across space

and time, but also across the range of estimates of the underlying physical model. For example, when estimating temperature, the GP algorithm takes advantage of regions where the physical model performs well (middle temperature range), while correcting the more pronounced differences at the lower and upper temperature extremes. As such, the major benefit of the GP approach is not necessarily just its ability to reduce error in the overall predictions, but rather to improve the predictive variance. By shrinking and quantifying the “error bars,” the approach thus explains more of the variability in the ship observations compared to the physical models.

While still significant, the ability of the GP algorithm to predict wind speeds is accompanied by a nuanced statistical point (Figure 2.8). Much of the variability evident when comparing vessel observations to those of the physical model is likely due to the height at which these measurements were made. While the NDFD forcing model explicitly outputs surface wind speeds (which drives wave heights), the measurement heights on each individual vessel are variable and unknown. It is reasonable to assume that the wind measurement heights on these vessels are above the height used by the physical model, which would explain the consistent underestimation by the physical model when compared to the ship observations. As such, it is important to view the GP-derived wind product as a spatially averaged representation of ship measurements, rather than a prediction of wind speed at a known height. Probabilistically, it represents the wind speed at an average, but unknown, measurement height. The predictions made by the physical model are still however very important in providing a prior estimate that can be used to spatially distribute the ship measurements, as evident in the performance of the GP estimates during evaluation of the testing data (Figure 2.8). Given the variability in measured wind speeds, it is also important to note that the GP did not seek to overtune the final model or drastically reduce its variance. Rather, the resulting variance or “error bars” on the windspeed predictions are an indication that, as intended, the likelihood function used in the training procedure struck a balance between data quality and model performance. Until the actual measure-

ment heights are known or are estimated via independent data (they are not presently being recorded by the VOS program), the GP wind product can be used to analyze spatial variability and trends, rather than specific estimates.

2.4.3 Final Data Product

Unlike a simple seasonal MAE analysis (Figures 2.3, 2.4 and 2.5), the GP-derived data product can be used to provide some insights into the factors governing the biases of the physical models. To that end, the difference between the GP-derived data and the physical models (third row, Figures 2.9 and 2.10) defines the magnitude to which the physical models would need to be adjusted at any point in time to reflect what a ship would observe. While these GP predictions give us additional understanding of the biases in the physical models, a major benefit of this approach relates to the ability to provide probabilistic estimates of uncertainty (last row, Figures 2.9 and 2.10). As would be expected, many of the lower levels of uncertainty coincide with regions that are traversed by the ships most frequently. However, many of the lower levels of uncertainty also overlapped with locations of the stationary buoy network, implying that the use of the buoys in the calibration of the physical model may have provided a level of consistency that improved the confidence of the GP predictions.

An increase in the availability of vessel measurements also improved the confidence and spatial resolution of the GP data product. In earlier years when fewer observations were reported, the algorithm adjusted the physical models at coarser resolutions (Figure 2.9). In later years, as more vessels joined the VOS program, the adjustments were made at much finer resolutions and resulted in lower predictive variances (Figure 2.10). As such, the robustness and certainty of the approach are expected to improve as more measurements are used. That said, there will likely be a point of diminishing returns, after which more measurements only marginally improve the performance of the GP algorithm. There were almost twenty-fold more observations in the summer of 2013 when compared to win-

ter 2007, but this did not result in a similar magnitude GP performance gain. In fact, most seasonal models yielded similar predictive performance after being trained on only a few thousand ship observations (as opposed to the over 20,000 in summer 2013).

This suggests that at some point the performance of the algorithm likely begins to be limited by the noise (uncertainty) of the sensors. Therefore, what may be most important is not to have as many measurements as possible, but to have enough measurements at representative locations. While a mathematically rigorous analysis of what constitutes “enough” measurements may be carried out formally using information theoretic measures (see the work of Krause and Guestrin, for example [39]), performance will likely vary based on the application and quality of underlying sensors. It may also be determined empirically on a case-by-case basis. Most of the later seasons in this study would converge to a model close to the full GP (i.e. one trained on all the available data) when trained on about 7000 observations for each variable of interest. Reducing the size of the training set down three-fold would reduce computation times by approximately 27 times (roughly 96%) [37]. In this study, it took a day to learn the hyperparameters when trained on nearly 21,000 data points for all three variables. It would then be possible to reduce the computational time to under an hour with a smaller training set, without significantly sacrificing performance. With computation times at this level, it would be possible to update the model at regular intervals (e.g. overnight each day) in a semi-online fashion using a moving window for training and updating the GP model.

The use of the variance estimates is very important when interpreting the final data product, as any outputs need to be weighed against predictive uncertainty. In many instances, locations lacking observation (buoy or ship) often had the highest error bars. The GP captures this by minimally adjusting the output in those locations, thus placing more weight on the output of the physical models. This implicitly captures the intuition that the physical model presents the best estimate of a process at locations or during times that measurements may not be available. This is most evident in the Georgian Bay (Eastern

Lake Huron), where limited ship and buoy measurements are made and therefore adjustments to air temperature are minimal, but error bars are much higher than the rest of the lake (first column, Figure 2.10). While this further emphasizes the need for more spatially representative measurements, it also highlights the value of the vessel measurements and benefits of the GP framework, which can provide confidence bounds on all its projections. These uncertainty estimates can be used to selectively identify the most informative regions of the lake, by enabling a quantitative tool to recommend measurement locations for new buoy networks (e.g. near-optimal sensor placement [40]). Beyond this, it would also be possible to leverage these uncertainty maps in more complex experimental settings. For instance, if an experimental vessel plans to collect measurements across the lakes, the most informative routes may be planned using the uncertainty maps generated by the GP product.

There are likely many other factors governing the discrepancies between the physical models and the new GP data product, suggesting that the error structure of the physical models may be stochastic in nature or impacted by regionally-specific physical model calibrations. While beyond the scope of this chapter, a more exhaustive analysis of the physical models could be carried out in the future to inform the implementation of the physical models. Generally speaking, the use of the proposed GP framework could thus also aid as a tool to modelers seeking to identify potential sources of model bias.

To reiterate – the goal of the proposed method is not to improve a physical model or expand it with data assimilating capabilities. Rather, this method seeks to find a general means by which to fuse multiple data sources to generate a combined data product. These sources of data could include multiple heterogeneous inputs, thus enabling significant flexibility. For instance, instead of using only the outputs of one physical model, it is possible to combine multiple modeled estimates or remotely sensed data products. This approach would then not only learn which of these inputs best explains the in-situ observations, but also where they perform best. This could, for example, lead to a final data product that

relies on one model to explain one part of a spatiotemporal domain, while using another to explain different regions. This creates a final data product that is weighted in response to those inputs that most consistently explain a phenomenon of interest.

It should be noted that a traditional data assimilation approach, such as a Kalman filter, may also have performed well at improving predictions when compared to the physical model outputs. However, the need to implement such an approach across two physical models, one of which is driven by the other, would have come at a significant implementation overhead. This would require the source code and the computational capacity to execute both physical models, which is beyond the scope and feasibility of many scientific studies. Rather, this approach provides a flexible means by which to extract just the readily-available output from these models, as well as many other data sources, and fuse the collective information into one final spatial data product.

In this chapter, used Gaussian likelihood with a squared exponential kernel is used. Mathematically, this approach becomes analogous to kriging, which has readily been adopted in the geostatistical community. The framework presented here, however, is extensible to many other likelihood and covariance functions, such as those used for classification rather than regression. This should afford additional flexibility in addressing a variety of hydrologic and water resource problems that are underpinned by heterogeneous observations and data sources.

2.5 Conclusion

This chapter introduced a probabilistic method by which to spatially distribute large quantities of ship measurements across surface water systems. Evaluated through cross-validation, this approach integrated the vessel measurements with operational physical models to generate a new spatial data product. While physical models alone may not readily accommodate the kind of measurements made by ships, the probabilistic method presented herein

offers a rigorous and flexible approach to combine the hydrometeorological expertise embedded in these models with the added benefit of independent distributed measurements. This approach is computationally demanding, but offers a means by which to begin ingesting large and unconventional data sources into studies of large water systems. The approach can be applied to many other water systems where irregular data are available. Its flexibility permits it to be informed by a combination of distributed and temporally irregular data sources, as well as a variety of spatial data inputs, such as physical models, satellite data or other new data sources such as those from drones. For example, remotely sensed data could readily be used as an input into the framework, with the objective of generating a ship-corrected satellite data product. Since the method does not set a bound on the number of inputs that are used, its only cost relates to computational complexity, which will increase with the number of input data features. An added benefit of the approach also involves the enabling of more complex tasks, such as the design of expanded measurement networks through the use of variance estimates.

An analysis of the final Great Lakes data product suggests that the availability of measurements across the Great Lakes will continue to play a large role in the confidence with which these large surface water systems can be studied and modeled. The ability to incorporate new sources of data could significantly improve understanding of these systems in an uncertain climate. For instance, the consistent over-estimation of air temperatures by the physical model at lower extremes could result in a smaller predicted temperature gradient between the surface and the air during peak evaporation in late fall. Without the knowledge provided by additional data, such as those used in this study, this may result in the underestimation of evaporative fluxes and therefore the over-prediction of water levels. While beyond the scope of this chapter, future studies will investigate these impacts, and Gaussian Process Regression is one tool that can enable the community to do so.

CHAPTER 3

Using sensor data to dynamically map large-scale models to site-scale forecasts

3.1 Introduction

As computational power has grown, so has hydrologists' ability to model complex hydraulic and hydrologic systems [41]. No longer limited to the study of single stream reaches or small watersheds, increasing access to supercomputers 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 [42]. 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. [43, 44, 45]). 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 [46] 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. [3, 47, 48, 49]) 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, 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 chapter 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, this approach leaves the physical model unchanged and uses a dynamical systems transformation to map the large-scale model outputs to site-scale conditions. To evaluate this approach, a case study is carried out 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 need-

ing to run or modify the models themselves. Specifically, this methodology is applied 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 chapter 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.

3.1.1 Background

To illustrate the challenges that may be faced when translating macro-modeled outputs to hyper-local conditions, the outputs from the US National Water Model (NWM) are used 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 relevant to local flood inundation mapping [47]. 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 [46].

Given its spatial extent, the NWM assumes trapezoidal stream cross sections, which are derived from the National Hydrography Dataset [50]. 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 [51]. Reliable rating curves require a relatively long history of stage and discharge measurements. Measurement-constraint alternatives have been proposed (e.g. [52], [53]), 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 [54], extrapolation outside of the history [55], hysteresis [56], and measurement error [57]. Most importantly, however, in the context of the proposed problem, the flows are modeled rather than measured, which poses additional challenges when attempting to estimate site-specific water levels.

To illustrate the challenge of deriving local height estimates from modeled flows, the output of the NWM is compared to two independent water level measurements made on small bridges in Iowa (Figure 3.1). For the first example (Figure 3.1a), 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 3.1c), 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 3.1b), 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 3.1d), it is unclear how the dynamics of each variable are correlated. 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.

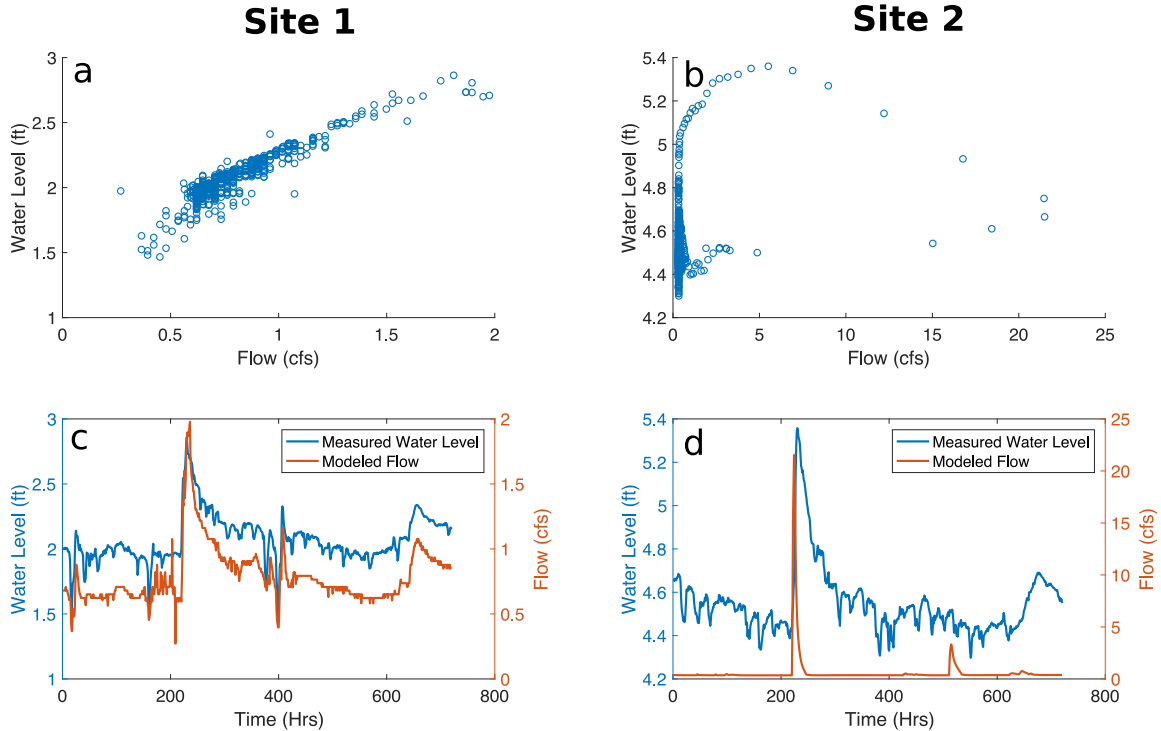


Figure 3.1: Measured water levels made by bridge sensors and modeled flows derived from the NWM for two example sites in the state of Iowa. The first example demonstrates a relatively strong relationship between modeled flows and measured water levels, while the second site does not.

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 [58] or Particle Filter [4] 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, 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. [59], [60]), 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 mea-

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.

3.1.2 Approach and Contributions

Motivated by the challenges posed in the prior section, the major contribution of this chapter 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, but perhaps coarse, model without needing to update the model itself (Figure 3.2). 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 [47]. Since these measurements have not been used in the calibration of 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.

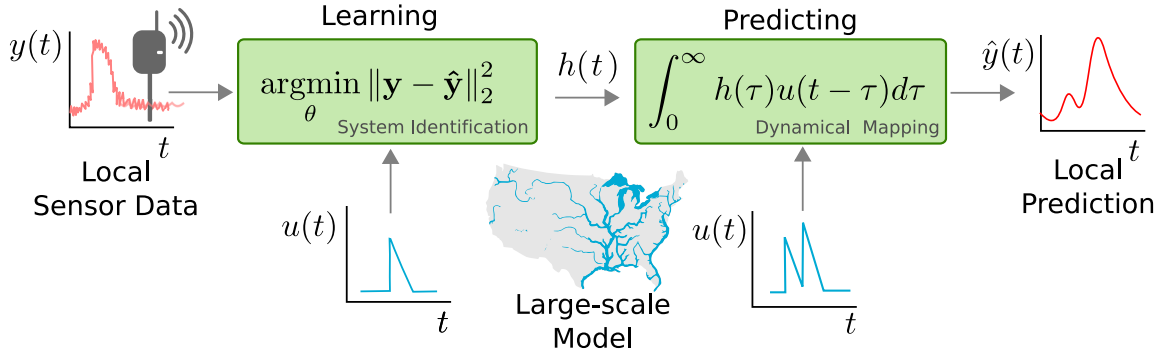


Figure 3.2: Conceptual diagram of dynamical mapping methodology. Historical measurements made by a sensor are used to “learn” a dynamical mapping between modeled flows and measured water levels. Once the parameters of the mapping are learned, water levels can then be predicted by dynamically transforming the modeled flows.

Since a one-to-one stage-discharge mapping is not possible for all sites (Figure 3.1b), this approach is based on dynamical systems theory [61]. Here, the output of the physical model is treated 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 transform model forecasts to water level estimates. At a low-order level, this approach is analogous to *learning* a unit hydrograph [62], which have been used to map rainfall to flows ([63], [64]). However, simple single-order unit hydrographs are known to work mostly for smaller scale catchments [52]. This 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 chapter presents the theory, implementation and application of this approach to a large set of over 180 stream height observations. Secondly, a performance analysis is conducted 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 this 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.

3.2 Methods

3.2.1 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 domain as a convolution with an impulse function $h(t)$ [61]:

$$y(t) = \int_{\tau=0}^{\infty} h(\tau)u(t - \tau)d\tau. \quad (3.1)$$

In this 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:

$$\begin{aligned} Y(s) &= H(s)U(s) \\ H(s) &= \frac{Y(s)}{U(s)} \\ &= \frac{b_0s^n + b_1s^{n-1} + \dots + b_{n-1}s + b_n}{s^n + a_1s^{n-1} + \dots + a_{n-1}s + a_n} \end{aligned} \quad (3.2)$$

where (a_0, a_1, \dots, a_n) and (b_0, b_1, \dots, 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 3.1 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 [61]. A common approach to learning the parameter $\theta := [a_1, \dots, a_n, b_0, \dots, b_n]$ of the model relies on the formulation

$$y(t) = \hat{y}(t, \mathbf{u}; \theta) + \epsilon(t, \theta) \quad (3.3)$$

where the measured output is a function of the predicted output \hat{y} given parameter set θ , 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 the mean squared error is used as the loss function:

$$\begin{aligned} \hat{\theta}(\mathbf{y}, \mathbf{u}) &= \arg \min_{\theta} \|\mathbf{y} - \hat{\mathbf{y}}\|_2^2 \\ &= \arg \min_{\theta} \sum_{t=1}^n (y(t) - \hat{y}(t, \mathbf{u}; \theta))^2 \end{aligned} \quad (3.4)$$

This approach uses a Gauss-Newton method [65] to iteratively approach the minimum through the use of a gradient-based solver:

$$\theta^{(k+1)} = \theta^{(k)} - (\mathbf{J}^T \mathbf{J})^{-1} \mathbf{J}^T \epsilon(\theta^{(k)}) \quad (3.5)$$

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:

$$\mathbf{J} = \begin{pmatrix} \frac{\partial \epsilon_1(\theta^{(k)})}{\partial \theta_1^{(k)}} & \dots & \frac{\partial \epsilon_1(\theta^{(k)})}{\partial \theta_n^{(k)}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \epsilon_m(\theta^{(k)})}{\partial \theta_1^{(k)}} & \dots & \frac{\partial \epsilon_m(\theta^{(k)})}{\partial \theta_n^{(k)}} \end{pmatrix} \quad (3.6)$$

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 ($\hat{\theta} = [a_1, \dots, a_n, b_0, \dots, b_n]$) yields an estimated signal \hat{y} that approaches the true signal y . Once θ is learned using time series of the inputs and outputs, forecasts can be made using Equation 3.1. A visual summary of the approach is provided in Figure 3.2.

3.2.2 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 this 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 made by the NWM. Along with a summary of performance, an extensive analysis was also carried out using Principal Component Analysis [66] and Logit Boosted Random Forests [67] to classify under which conditions the proposed approach may perform reliably.

3.2.2.1 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 [42]. The model is forced by rainfall from the Multi-Radar/Multi-Sensor System (MRMS) [68] as well as a suite of models ingested by WRF-Hydro [42]. Additionally, it 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 [42]. 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 chapter uses this product to provide an upper bound baseline for the dynamical mapping approach.

3.2.2.2 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 [47]. One of their major initiatives was establishing the Iowa Flood Information System (IFIS), which provides real-time stream conditions and flood warning alerts [69]. IFIS ingests data from approximately 500 stream sensors, of which half are managed by the USGS and half are managed by the IFC (Figure 3.3). IFC

gauges are primarily composed of bridge-mounted ultrasonic depth 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. In this chapter, the focus is 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.

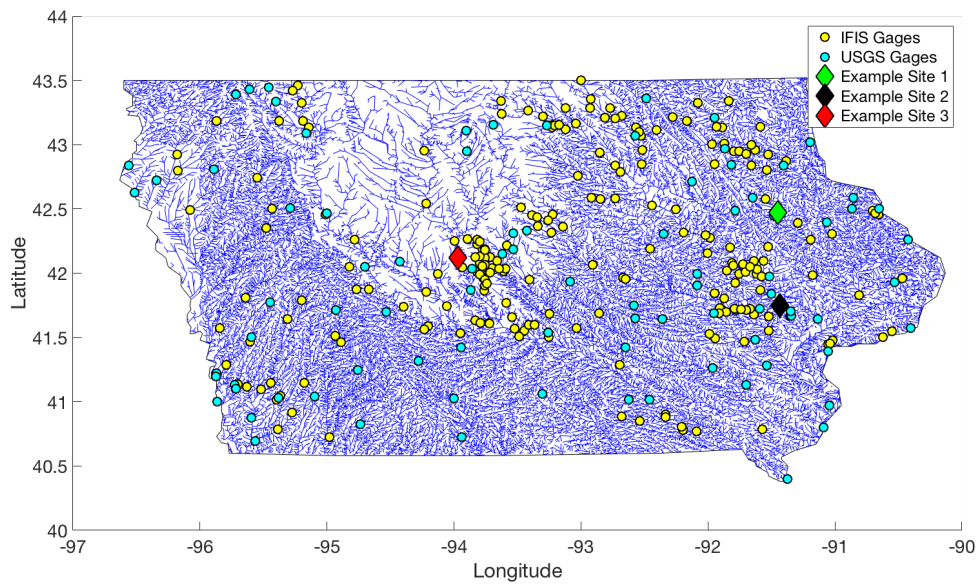


Figure 3.3: Visualization of the nearly 62,000 streams modeled by the NWM in the state of Iowa. USGS gages, which are assimilated into the NWM, are denoted as cyan circles. Locations of the IFIS water level sensors are denoted as yellow circles, with diamonds denoting the three example sites used in this chapter.

3.2.2.3 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 timeseries were linked

by location, thereby providing individual model-measurement pairs that could be used in the dynamical mapping approach. Data from October to December were used to train the 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.

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. These included all possible pole-zero pairings for first through fourth order systems ([0,1],[1,1],[0,2]..., [3,4],[4,4]). 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 [70]. For comparison, a standard regression rating curve procedure [71] 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 [72], was used to evaluate performance:

$$E = 100 \left(1 - \frac{\|y - \hat{y}\|}{\|y - \bar{y}\|} \right) \quad (3.7)$$

where E is the nRMSE 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 [73]. 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 nRMSE of 50% or above is generally considered the lower bound for a good predictor [74], which is the threshold adopted in this chapter. The analysis considered the nRMSE of both the best ensemble member and the ensemble average.

3.2.3 Performance Classification

One major goal of this chapter is to investigate under which conditions the proposed dynamical mapping approach will work well. Not all locations may benefit directly from this 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 the proposed approach under various physiographic conditions, a combined approach of Principal Component Analysis [75] and Random Forest Classifiers [67] is used.

The NWM is built on a number of physiographic features from the National Hydrography Dataset (NHD) [50]. These include the channel bottom width, elevation, Manning's roughness, channel slope, and Strahler stream order [76]. An additional feature was also considered, which captures the distance of a given water level sensor to the nearest USGS gauge. This will indicate whether this approach performs better near official NWM data assimilation locations. Overall, this provided six features that may be used to explain the performance of the dynamical mapping approach. For example, intuition would suggest that the 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, this approach used Principal Component Analysis (PCA) to shift the six dimensional feature space into an orthogonal subspace [75]. 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-orientes 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.

The goal of PCA is to find the weighting vectors, or *principal components*, that yield linear combinations of the original feature space. $\mathbf{X} \in \mathbb{R}^{n \times d}$ is the data matrix with n rows of observations and d features, which in this 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 [75]. By standardizing across each variable, one can consider the relative impacts of each more effectively.

To find the first principal component, \mathbf{w}_1 , a unit vector that maximizes the variance of \mathbf{X} must be found. That is:

$$\mathbf{w}_1 = \arg \max_{\|\mathbf{w}\|=1} \|\mathbf{X}\mathbf{w}\|^2 = \arg \max \frac{\mathbf{w}^T \mathbf{X}^T \mathbf{X} \mathbf{w}}{\mathbf{w}^T \mathbf{w}} \quad (3.8)$$

This is a Rayleigh quotient [77], and therefore the solution to this maximization problem is the largest eigenvector (i.e. the eigenvector of the largest eigenvalue) 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 3.9) of the data matrix \mathbf{X} :

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{W}^T \quad (3.9)$$

$$\mathbf{X}^T\mathbf{X} = \mathbf{W}\Sigma\mathbf{U}^T\mathbf{U}\Sigma\mathbf{W}^T = \mathbf{W}\Sigma^2\mathbf{W}^T \quad (3.10)$$

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 the data matrix. Therefore, to place the data in an orthogonal feature space such that all the variables are de-correlated, the new data matrix, \mathbf{T} , is simply:

$$\mathbf{T} = \mathbf{X}\mathbf{W} \quad (3.11)$$

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 [75].

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 nRMSE 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, this approach used a statistical learning tool known as Logit Boosted Random Forests, or *Adaboost* with trees [67].

Adaboost generates a large number of “weak learners” [75] in the form of small classification trees. A weak learner is a model that is only slightly better than randomly guessing [78]. The tree partitions a feature space using a series of binary splits, resulting in a large

number of labeled bins. Given their relative simplicity, these trees tend to have very low bias but high variance [75]. This can be addressed by generating an ensemble, or a *forest*, of many trees. Going a step further, each tree should be developed to provide as much information gain as possible. In Logit Boosted Random Forests, each data entry (i.e. labeled row of the data matrix \mathbf{T}) is given an initial weight w_i . Then, as new trees are learned, the data entries are reweighted 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), and $H(z) := \mathbf{1}_{[z>0]}$ is the Heaviside step function [75]. M classification trees are learned in an iterative fashion.

```

1 Initialize  $w_i = \frac{1}{N}, i = 1, 2, \dots, N;$ 
2 for  $m=1, 2, \dots, M$  do
3   | Learn classification tree that outputs  $p_m(\mathbf{t}) = P_w(y = 1|\mathbf{t}) \in [0, 1]$  with weights
   |  $w_i;$ 
4   | Set  $f_m(\mathbf{t}) \leftarrow \frac{1}{2} \log \frac{p_m(\mathbf{t})}{1-p_m(\mathbf{t})};$ 
5   | Set  $w_i \leftarrow w_i \exp[-y_i f_m(\mathbf{t}_i)], i = 1, 2, \dots, N,$  and renormalize such that
   |  $\sum_i w_i = 1;$ 
6 end
7 Output classifier as  $H[\sum_{m=1}^M f_m(\mathbf{t})]$ 

```

Figure 3.4: 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 the 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.

3.3 Results

3.3.1 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% nRMSE threshold, while performance across 90 sites exhibited an nRMSE of at least 40%. The overall performance of the approach is summarized in Figure 3.5, 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.

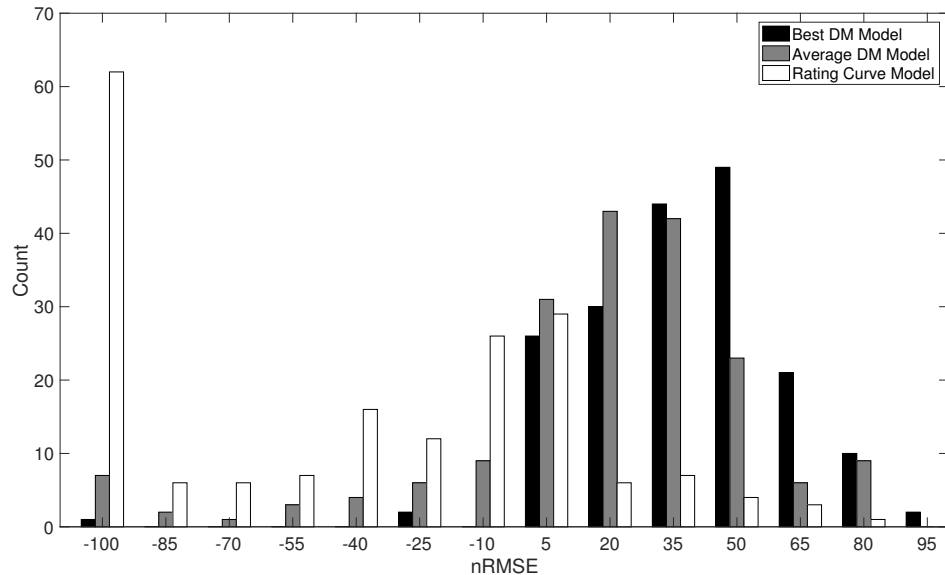


Figure 3.5: Histograms of prediction performance (nRMSE) evaluated across 182 sensor locations. A comparison is made between the best dynamical mapping (black), ensemble of dynamical mappings (gray), and a simple correlation-based rating curve approach (white)

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. The first is a location for which the 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 providing coarse outputs. The final example illustrates a case in which there is a limited ability to predict observed heights from flows.

The first example demonstrates a case of strong predictive performance (Figure 3.6). The left column of the figure displays the training data, which include the NWM model outputs and measured water levels for the Fall of 2016. The right column shows the NWM outputs and measured water levels for the Spring of 2017, as well as the water level predictions made by the DM approach. Specifically, the bottom right panel is the average prediction made by the approach across all 14 transfer functions (red line, with gray area indicating variability within the ensemble) compared to the measured water levels (blue line). Overall, the DM procedure performed well at this site, with an average nRMSE close to 80%. While not plotted, predictions of water levels at this site using a simpler regression-based rating curve performed nearly as well, with an average nRMSE of 76%.

Dynamical Mapping: Site 1

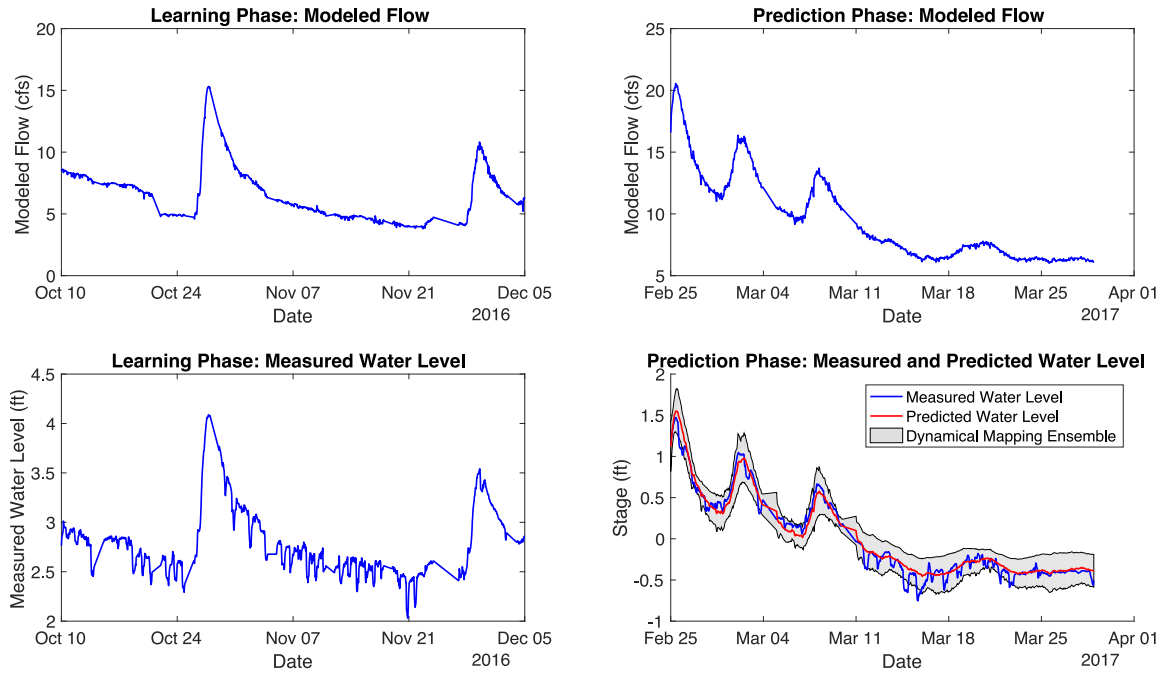


Figure 3.6: Dynamically mapping modeled flows to local water levels on site 1 (see Figure 3.3). Data used to “learn” the mapping parameters are plotted in the left column, while the resulting mapping is applied to future data in the right column. For this example site, the dynamical mapping performs relatively well (nRMSE of 80%). A simple regression-based rating curve approach (not plotted) performs strongly as well, with an nRMSE of 76%.

The second example (Figure 3.7) illustrates a case where a simple regression approach did not perform well, largely because modeled flows and measured water levels did not correlate (nRMSE of -4%). The modeled flows were quite impulsive and not representative of observed dynamics. However, when the DM approach was used, the results improved significantly, with an average nRMSE of over 50%.

Dynamical Mapping: Site 2

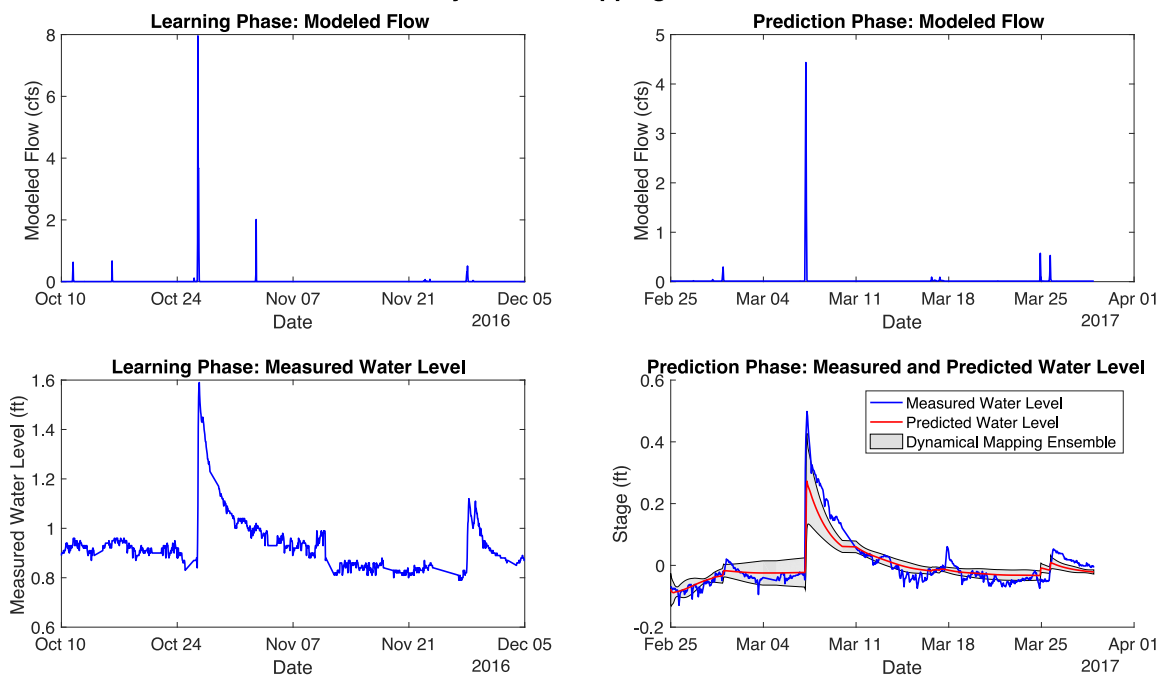


Figure 3.7: Dynamically mapping modeled flows to local water levels on site 2, following convention used in Figure 3.6. For this example site, the dynamical mapping performs relatively well (54% nRMSE), while a simple regression-based rating curve approach does not (-4% nRMSE).

Finally, the third example (Figure 3.8) 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 nRMSE of -145% was obtained using the DM approach, with only one of the 14 ensembled transfer functions showing a slightly favorable nRMSE (49%). The rating curve method was even more ineffective, with an nRMSE of -14,900%.

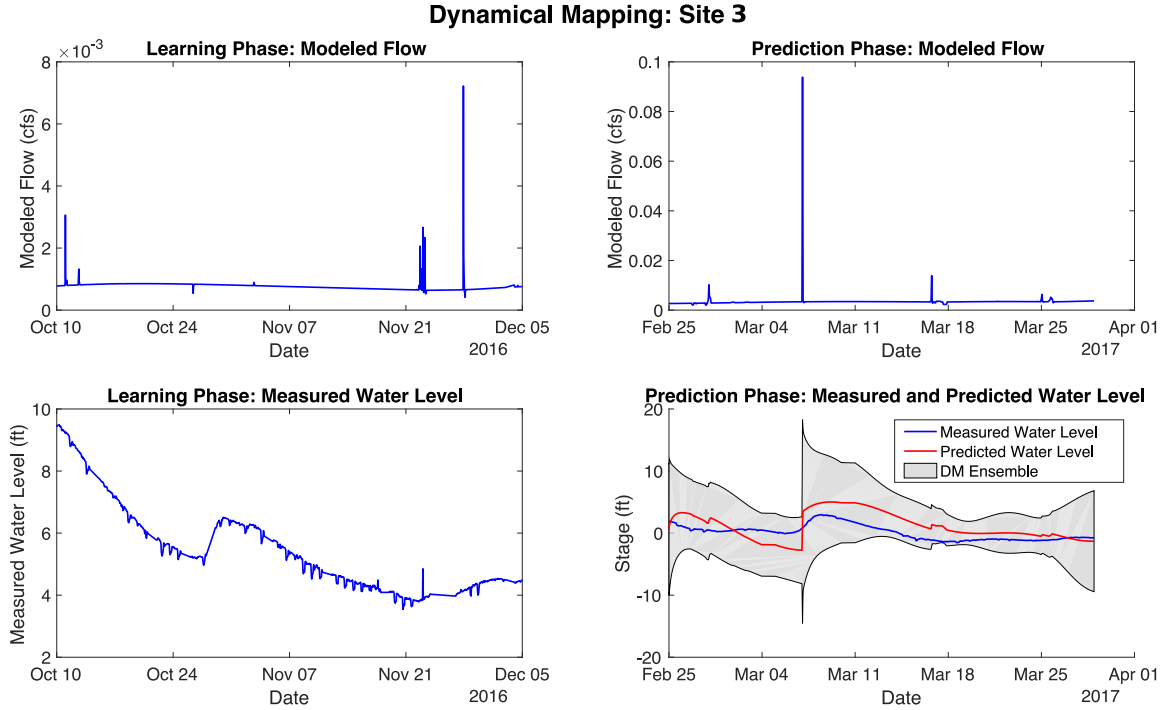


Figure 3.8: Dynamically mapping modeled flows to local water levels on site 3, following convention in Figure 3.6. For this example site, neither the dynamical mapping (0% nRMSE) or regression-based rating curve (nRMSE of -14 , 900% nRMSE) perform well.

3.3.2 Performance Classification

Using the 50% nRMSE criterion, 55 of the 182 sites were labeled as locations of *high* performance, while 127 were labeled as *low* performing, reflecting the ability of the DM approach to predict flows from NWM outputs. These labels were then used to determine the combination of physiographic characteristics that describes the conditions under which the DM approach exhibits high performance. The normalized distributions of each physiographic feature, split by performance criteria, is shown in Figure 3.9a. Overall, little distinction was evident between high-performing and low-performing sites, with the distribution of each physiographic feature showing similar means and variances. The distributions of channel bottom width and channel slope showed the relatively largest discrepancy, suggesting that sites at which the DM approach performed well had a larger stream width and slope than lower-performing sites. However, the bounds on these distributions were

not sufficient to determine a consistent labeling.

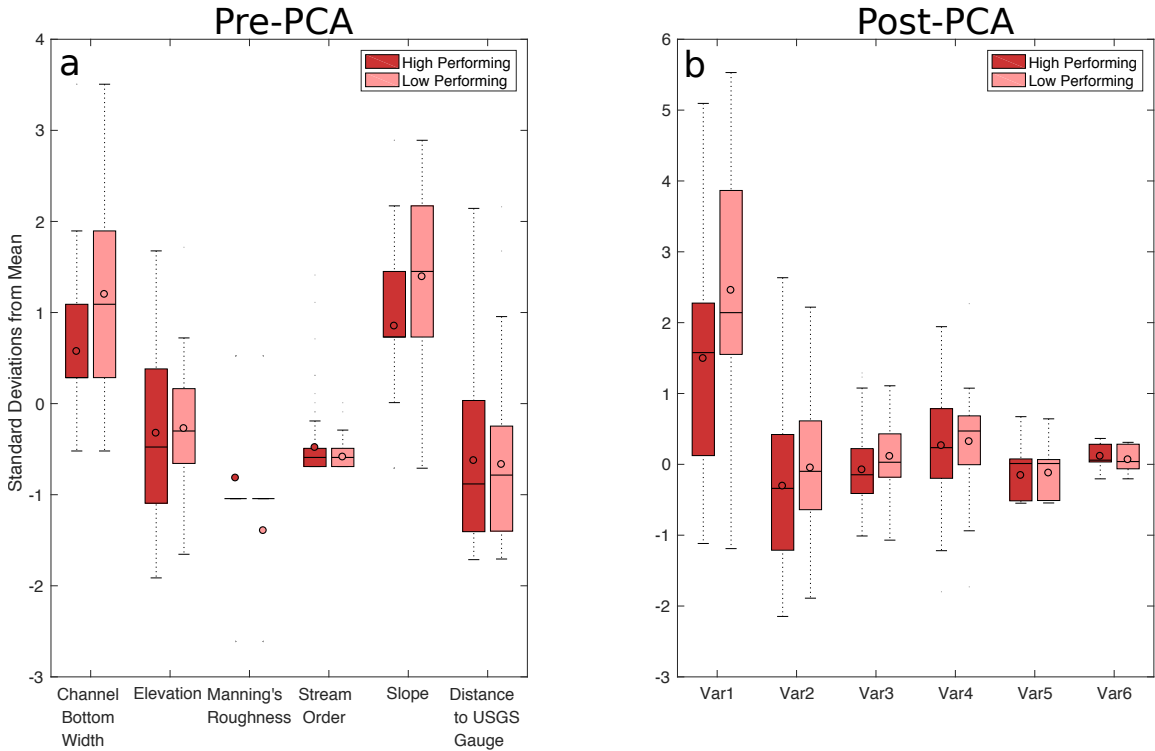


Figure 3.9: Boxplots representing the relative distribution of features, when split by the ability of a dynamical mapping to predict water levels from modeled flows. For any given feature, a clear difference between the two distributions would indicate that this feature describes a general condition 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 62000×6 data matrix. The resulting principal components are shown in Table 3.1. 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. Similar interpretative examples could be provided for the other principal components. Further, a number of the components exhibited opposing physiographic relationships. For example, for the second principal component, streams closer to a USGS gauge and located at 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.

Table 3.1: Principal Components resulting from applying PCA to features of 62,000 streams across the entire state of Iowa. σ 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 principle components, is shown in Figure 3.9b. Compared to splitting based on just physiographic features (Figure 3.9a), 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.

After applying the Logit Boosted Random Forest algorithm (Algorithm 3.4), cross val-

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 the 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 3.10, 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 the DM approach.

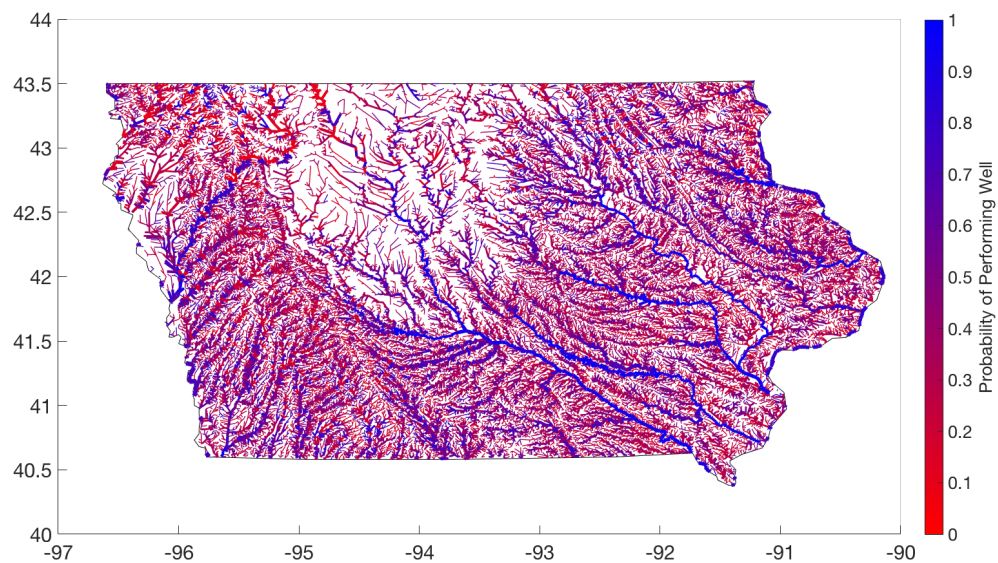


Figure 3.10: Map of site performance potential across the state of Iowa, showing a spectrum of locations where the dynamical mapping approach is expected to perform well in predicting local water levels from flows (blue) to those where it will likely not perform well (red).

3.4 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. 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, the dynamical mapping approach performed well in all of these cases, too, since it can be generalized as a linear transformation [70].

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, this 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 this 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 3.7). 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 irrespective 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, this 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, this approach did not perform well in mapping NWM flows to local water levels, as quantified by the 50% nRMSE 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 3.8). 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, this approach will not work under these conditions, since it requires changes in the inputs to be mapped to changes in outputs. Of course, this DM approach could benefit by including additional local data (e.g. rainfall), 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, the performance of the DM approach is expected to 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 the performance classification. Given the size, complexity, and collinearity of the data set, it is clear that a simple classification of performance, based on individual physiographic features, does not provide much insight (Figure 3.9a). One take-away, though not strongly consistent, appears to be that the 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 [79, 80].

While 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, the application of Logit Boosted Random Forests allowed for the creation of a map that summarizes the expected performance of the approach across all 62,000 streams in Iowa (Figure 3.10). This visual representation provided an intuitive means by which to assess broader performance. As expected, the 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 3.10) 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. 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 the dynamical mapping.

Given its impressive extent and recent operationalization, 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 *hyper-resolution* forecasts. Alternatively, as this 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. Nonetheless, these results may also provide a guide to help improve the numerical model. The map in Figure 3.10 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 the 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 the 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 3.10. A simple web-service application [3] 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 3.10 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, similar maps could be created for regions outside of Iowa by relying on the results from this study.

3.5 Conclusions

In this chapter, a means by which outputs from a large-scale model can be fused with local sensor data to provide site-level forecasts is provided. 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 the 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*, 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 from 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, this 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 the data-driven approach could be generalized for many hydrologic and hydraulic models.

CHAPTER 4

Model reduction for real-time change detection across complex water systems

4.1 Introduction

Much of the urban water infrastructure across the world is approaching or has exceeded its design life [12]. In the particular case of urban stormwater systems, it is not uncommon for clogged pipes to remain undetected or for basins to fill with sediment. This compromises system performance and can lead to dangerous flooding and water quality impairments [81]. As such, it has become imperative to detect subtle changes in aging urban water infrastructure before they become bigger problems.

To that end, *asset management* has risen to prominence as a data-driven concept to enable better maintenance of urban water infrastructure systems [82]. In most cases, however, *asset management* is just now taking its first and simple first step – namely, documenting where infrastructure assets are, but not tracking if they are performing as intended. For municipalities that take it a step further, inventories are also supplemented with more advanced inspection schedules in which the age, criticality and repair cost of the asset are used as factors to inform predictive maintenance [82]. This, however, has limitations since unexpected problems can often arise during periods between inspections. In many cases, by the time a problem is detected, a significant amount of damage may have already been

incurred.

In the age of adaptive water management, novel solutions are needed to more rapidly detect critical changes in infrastructure [11]. This brings to bear the notion of real-time asset management, where streaming data will be used to continuously update knowledge of changing urban water systems. The steep reduction in cost of modern wireless sensing technologies is now empowering even small communities to begin measuring water flows, water quality, and meteorological variables [49]. This promises to enable the detection of problems as they occur, thus allowing for rapid maintenance and emergency response.

4.1.1 Background

While real-time infrastructure measurements are on the rise, it may be unreasonable to assume that every single infrastructure asset will be measured. Instead, sensors will likely only be located at major points of interest or will be placed to maximize system-level coverage [81]. Sensor data may easily reveal that *something* may have changed or been damaged, but locating *where* the change occurred is difficult using data alone if the entire system is not instrumented. As such, streaming sensor data will need to be supplemented with effective models and fast computational methods to map real-time system changes to potential causes.

The complexity of the built environment makes changes difficult to detect, especially when popular numerical models are used. Most water systems are difficult to model efficiently, with even simple models having thousands or even millions of parameters needing calibration. For instance, the Stormwater Management Model (SWMM) includes parameters for conduit size, storage node volumes, storage curves, subcatchment runoff, and infiltration, to name a few [83]. As a result, even a simple 20 node system can have well over 1000 parameters. *The curse of dimensionality* thus makes locating a change in the system very time consuming [84], requiring computationally complex re-calibration algorithms and intimate knowledge of the system.

Some approaches have been proposed to automate damage detection, particularly in the fields of water distribution systems and structural health monitoring. Methods such as error domain model falsification [85], vulnerability analysis [86], and graph theoretic approaches [87] can automate the detection of faults, but are highly computationally complex, requiring an exhaustive simulation of every possible damage scenario. Alternatively, a number of real-time approaches have been proposed, able to detect a damaging event but not localize it [88, 89]. Efficient localization algorithms generally require a long time history to train on [90, 91]. Most importantly, however, many of these methods require very well-observed systems with many sensors.

4.1.2 Contribution

This chapter asks the question: *how can changes in partially-observed urban stormwater or sewers systems be quickly detected and located?* Answering this question will allow for downstream changes, as detected by one sensor, to be mapped to upstream causes. This approach abstracts a stormwater system as a linear graphical model, which is parameterized with sensor data at the outlet of an urban watershed. This computationally-efficient abstraction is rapidly-identifiable, which means that it can be quickly re-parameterized to spatially localize changes in the stormwater system. The specific contributions of this chapter are:

- A computationally efficient methodology, based on *System identification* and *Kalman Filtering*, that parameterizes a linear model of a storm water system using only knowledge of infrastructure connectivity and a single sensor observation at the outlet of a stormwater system
- A likelihood-based probabilistic approach that uses the linearized model to estimate and localize potential damage in the entire system

- A physically-based simulation of the approach, with a case study focused in detecting and localizing clogged stormwater pipes in an urban watershed
- A discussion of the performance and limitations, seeking to specifically identify the conditions and scale across which the methodology should be applied

The outcome of this chapter builds toward a larger goal of enabling real-time asset management, which will transform emerging forms of streaming sensor data into actionable insights for city water managers.

4.2 Methods

Our methodology for damage detection in stormwater systems is shown in Figure 4.1. The urban watershed is represented as multi-input, single-output linear state-space system where the only observable state is the flow at the outlet of the storm water network. The inputs are the subcatchments' rainfall runoff, which are assumed to be known. The outlet of the watershed is measured by a flow sensor. It is assumed that only network connectivity (nodes and pipes in the network) is known, and no other information is needed. Using the sensor measurements, a maximum likelihood estimation approach is used to estimate the parameters of the model, which is very computationally efficient given the linear model assumption. Given a new rainfall event, if the measurements do not match the flows predicted by the model, an efficient search is then performed to determine which parameter had the highest likelihood of change. Simulations yielding the highest likelihood scores then indicate where the damage is most likely located. As an initial step, this chapter only focuses on damage in the form of complete pipe blockages, with other types of damage to be investigated in the future.

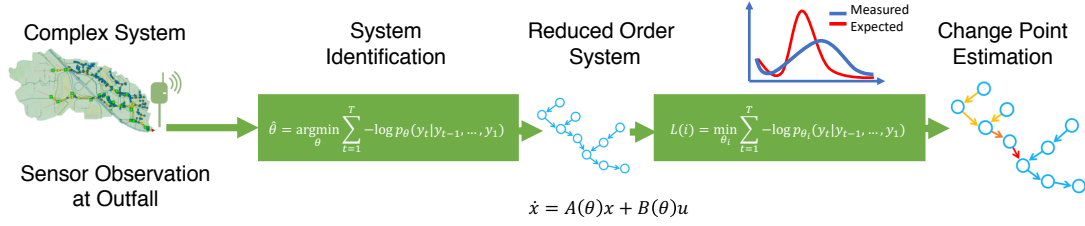


Figure 4.1: A simplified abstraction of a complex system is generated through system identification from a sensor observation. Then, when an anomalous signal is detected, an iterative search can be performed quickly to probabilistically identify likely locations for damage.

4.2.1 Initial Model Fit

Here, a stormwater system is represented as a directed graph of connected nodes (pipes and junctions). A larger system can be structured as a series of N discrete-time linear differential equations, where N is the number of nodes in the system:

$$\begin{aligned}
 x_i(t+1) &= a_{ii}x_i(t) + a_{ki}x_k(t) - a_{ji}x_j(t) + b_iu_k(t) \\
 y(t) &= x_N(t)
 \end{aligned} \tag{4.1}$$

where $x_i(t)$ is the volume of water passing through node i at time t , y is the sensor measurement at the outfall, and $u_k(t)$ is the runoff from subcatchment k at time t . The volume of water passing through a node at any time is a function of the runoff, outflow, and inflow for that node at the previous time step. This is shown graphically in Figure 4.2. The coefficients a and b are the parameters, which are to be estimated using the downstream sensor data. For a junction node, the a_{ii} parameter is simply zero, whereas for a storage node, the a_{ii} parameter will be on the interval $[0, 1]$. This is because the state at the current time step is dependent on the state at the previous time step. Structuring these equations in matrix form yields:

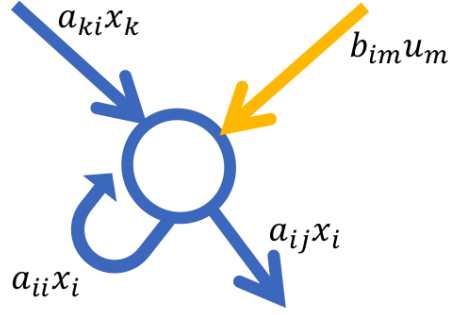


Figure 4.2: Node i where the flow passing through the node at the next time step (x_i) is a function of the runoff ($b_{im}u_m$), the flow from upstream ($a_{ki}x_k$), the flow going downstream ($a_{ij}x_i$), and the flow staying at the node ($a_{ii}x_i$)

$$\mathbf{x}(t+1) = \begin{bmatrix} a_{11} & 0 & 0 \cdots 0 & a_{1j} & 0 \cdots 0 \\ 0 & a_{22} & 0 \cdots 0 & a_{2j} & 0 \cdots 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & a_{Nj} & 0 \cdots 0 & a_{NN} \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} b_{1k} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & b_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & b_{Nk} & 0 \end{bmatrix} \mathbf{u}(t)$$

$$\mathbf{y}(t) = \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix} \mathbf{x}(t) \quad (4.2)$$

We simplify this notation as:

$$\mathbf{x}(t+1) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t)$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) \quad (4.3)$$

Note that by coupling the differential equations in this form, the $-a_{ij}$ parameter in Figure 4.2 (the flow to the downstream node) disappears, reducing the number of parameters needing estimation. This results in a very sparse discrete time state-space system. The problem can then be formulated in the context of a Maximum Likelihood system identi-

fication framework to identify the parameters $\theta = [a_{11}, a_{12}, \dots, a_{NN}, b_{11}, b_{12}, \dots, b_{NK}]$ that define the system given a known topology. This procedure is explained in detail in [92].

The joint probability density of the observations is:

$$p(y(T)|\theta) = p(y(0)|\theta) \prod_{t=1}^T p(y(t)|y(t-1), \dots, y(1); \theta) \quad (4.4)$$

Defining the likelihood function as $\mathcal{L}(\theta) = -\log p(y(T)|\theta)$ and assuming Gaussian noise on the observations (runoff and sensor measurements, with unknown variance q_k and o , respectively), yields:

$$\mathcal{L}(\theta) = \sum_{t=1}^T \log \det \Lambda_t(\theta) + \epsilon_t(\theta)^\top \Lambda_t(\theta)^{-1} \epsilon_t(\theta) \quad (4.5)$$

$$\epsilon_t(\theta) = y_t - \hat{y}_{t|t-1}(\theta) \quad (4.6)$$

$$\hat{y}_{t|t-1}(\theta) := \mathbb{E}[y_t | y_{t-1}, \theta] \quad (4.7)$$

$$\Lambda_t(\theta) := \mathbb{E}[\epsilon_t(\theta) \epsilon_t(\theta)^\top | y_{t-1}, \theta] \quad (4.8)$$

To determine these quantities, a time-varying Kalman filter framework is used, yielding:

$$\hat{y}_{t|t-1}(\theta) = C\hat{x}_{t|t-1}(\theta) \quad (4.9)$$

$$\hat{x}_{t+1|t}(\theta) = A(\theta)\hat{x}_{t|t-1}(\theta) + B(\theta)\mathbf{u}_t + K(\theta)\epsilon_t(\theta) \quad (4.10)$$

$$K(\theta) = (A(\theta)P_{t|t-1}(\theta)C^\top)\Lambda_t^{-1}(\theta) \quad (4.11)$$

$$P_{t+1|t} = A(\theta)P_{t|t-1}(\theta)A(\theta)^\top + B(\theta)Q(\theta)B(\theta)^\top - K(\theta)\Lambda_t(\theta)K(\theta)^\top \quad (4.12)$$

$$\Lambda_t(\theta) = CP_{t|t-1}(\theta)C^\top + o(\theta) \quad (4.13)$$

where \hat{y} is the predicted signal, \hat{x} is the predicted state, Q is a diagonal matrix of the observation noise ($q_1 \dots q_k$) for each subcatchment k , K is the Kalman gain, P is the covariance of the state estimate, and Λ is the innovations (the covariance of the residual). Taking the gradient of the likelihood function with respect to θ , via the following recursive terms, yields:

$$\frac{\partial \mathcal{L}}{\partial \theta} = \sum_{t=1}^T \text{tr} \left(\Lambda_t^{-1} \frac{\partial \Lambda_t}{\partial \theta} \right) + 2 \frac{\partial \epsilon_t^\top}{\partial \theta} \Lambda_t^{-1} \epsilon_t - \epsilon_t^\top \Lambda_t^{-1} \frac{\partial \Lambda_t}{\partial \theta} \Lambda_t^{-1} \epsilon_t \quad (4.14)$$

$$\frac{\partial \epsilon_t}{\partial \theta} = -C \frac{\partial \hat{x}_{t|t-1}}{\partial \theta} \quad (4.15)$$

$$\frac{\partial \hat{x}_{t+1|t}}{\partial \theta} = \frac{\partial A}{\partial \theta} \hat{x}_{t|t-1} + A \frac{\partial \hat{x}_{t|t-1}}{\partial \theta} + \frac{\partial B}{\partial \theta} \mathbf{u}_t + \frac{\partial K_t}{\partial \theta} \epsilon_t + K_t \frac{\partial \epsilon_t}{\partial \theta} \quad (4.16)$$

$$\frac{\partial K_t}{\partial \theta} = \left(\frac{\partial A}{\partial \theta} P_{t|t-1} C^\top + A \frac{\partial P_{t|t-1}}{\partial \theta} C^\top - K_t \frac{\partial \Lambda_t}{\partial \theta} \right) \Lambda_t^{-1} \quad (4.17)$$

$$\begin{aligned} \frac{\partial P_{t+1|t}}{\partial \theta} = & \frac{\partial A}{\partial \theta} P_{t|t-1} A^\top + A \frac{\partial P_{t|t-1}}{\partial \theta} A^\top + A P_{t|t-1} \frac{\partial A^\top}{\partial \theta} + \frac{\partial B}{\partial \theta} Q B^\top + B \frac{\partial Q}{\partial \theta} B^\top + \\ & B Q \frac{\partial B^\top}{\partial \theta} - \frac{\partial K_t}{\partial \theta} \Lambda_t K_t^\top - K_t \frac{\partial \Lambda_t}{\partial \theta} K_t^\top - K_t \Lambda_t \frac{\partial K_t^\top}{\partial \theta} \end{aligned} \quad (4.18)$$

$$\frac{\partial \Lambda_t}{\partial \theta} = C \frac{\partial P_{t|t-1}}{\partial \theta} C^\top + \frac{\partial O}{\partial \theta} \quad (4.19)$$

To determine the minimum of Equation 4.5, a Gauss-Newton approach is used, where the update for θ is:

$$\theta_{i+1} = \theta_i + H_i \frac{\partial \mathcal{L}}{\partial \theta} \quad (4.20)$$

$$H_i = (J^\top J)^{-1} J^\top \quad (4.21)$$

$$J_{tk} = \frac{\partial \epsilon_t}{\partial \theta_k} \quad (4.22)$$

4.2.2 Detecting System Change

Once the initial state-space model is parameterized, it can then be used to predict flows at the outlet. For damage detection, the likelihood function $\mathcal{L}(\theta)$ can be modified slightly:

$$\mathcal{L}(\theta_i) = -\log p(y(T)|\theta_i) \quad (4.23)$$

where θ_i is the parameter subset associated with node i . With this formulation, when an anomalous signal is detected and suspected, an iterative search over all the nodes in the system can be quickly performed to identify the node with the highest likelihood of having a pipe blockage. Focusing on a pipe blockage in a single output system provides two benefits for the analysis. First, a single outfall will have a likelihood equivalent to the mean squared error (MSE, with scaling). Second, when performing the iterative search for possible damage locations, instead of implementing a complete system identification procedure (as outlined in the previous section), each of the parameters in θ_i can simply be set to zero to indicate damage (pipe blockage). Therefore, when performing the search, the algorithm simply changes θ_i to zero, feeds the input signal into the modified state-space, and checks the MSE of the signal against the anomalous signal. The result is a highly efficient algorithm that takes mere seconds to analyze a large system (Algorithm 4.3).

The output of this procedure is not just one candidate site, at which damage is suspected, but a list of possible damage locations, ordered by likelihood. This is an important feature since many stormwater systems contain elements that are very similar (storage volumes or travel times). This means that a very similar change in the outflow signal could be caused by multiple sites. As such, it becomes important to reduce the search space to the most likely damage locations, rather than just one point of interest. Nonetheless, it is expected that one site will generally be returned as the location of highest likelihood of damage.

It should be noted that if the damage scenario is only a partial blockage, then the system identification approach outlined in Section 4.2.1 would need to be implemented iteratively for each parameter set θ_i . While this type of damage is not considered in this chapter, the search could nonetheless be carried out over only a small number of parameters and be

performed very efficiently.

```

1 Fit Initial Model
2 Initialize check =  $\infty$ ;
3 for  $m=1,2,\dots,10$  do
4   Randomly initialize  $\theta = [a_{11}, a_{12}, \dots, a_{NN}, b_{11}, b_{12}, \dots, b_{NK}]$ ;
5   Estimate  $\theta_m = \operatorname{argmin} \mathcal{L}(\theta_m)$ ;
6   if  $\mathcal{L}(\theta_m) < \text{check}$  then
7     Set check  $\leftarrow \mathcal{L}(\theta_m)$ ;
8     Set  $\hat{\theta} \leftarrow \theta_m$ ;
9   end
10 end

11 Search
12 for node in nodes do
13   Set  $\theta_{node} \leftarrow [0 \dots 0]$ ;
14    $l_{node} = \mathcal{L}(\theta_{node})$ 
15 end
16 Output Damaged Node =  $\min(l_{node})$ 

```

Figure 4.3: Change Detection Algorithm

4.2.3 Implementation

The water system being analyzed in this chapter is a real-world stormwater system in the Midwestern United States. (Figure 4.4). Given the size of the watershed (5km^2), a number of sub-networks (up to 30 nodes) are delineated and analyzed. In this case study, the sensor is located at the outlet of each sub network. For evaluation purposes, the sensor readings are simulated using a SWMM model of the system, which provides physically realistic flow values that could be expected in the field. Furthermore, the SWMM model is used to simulate pipe blockages, which would not be possible in the field. A series of scaled rain events from May of 2016 are used to force the SWMM model and the “virtual” sensor readings are used in the analysis. These precipitation events are scaled to be roughly equivalent to a 1-year storm. This ensures that the storm events were significant, but not enough to fully flood the system. All simulations are carried out in MATLAB using the

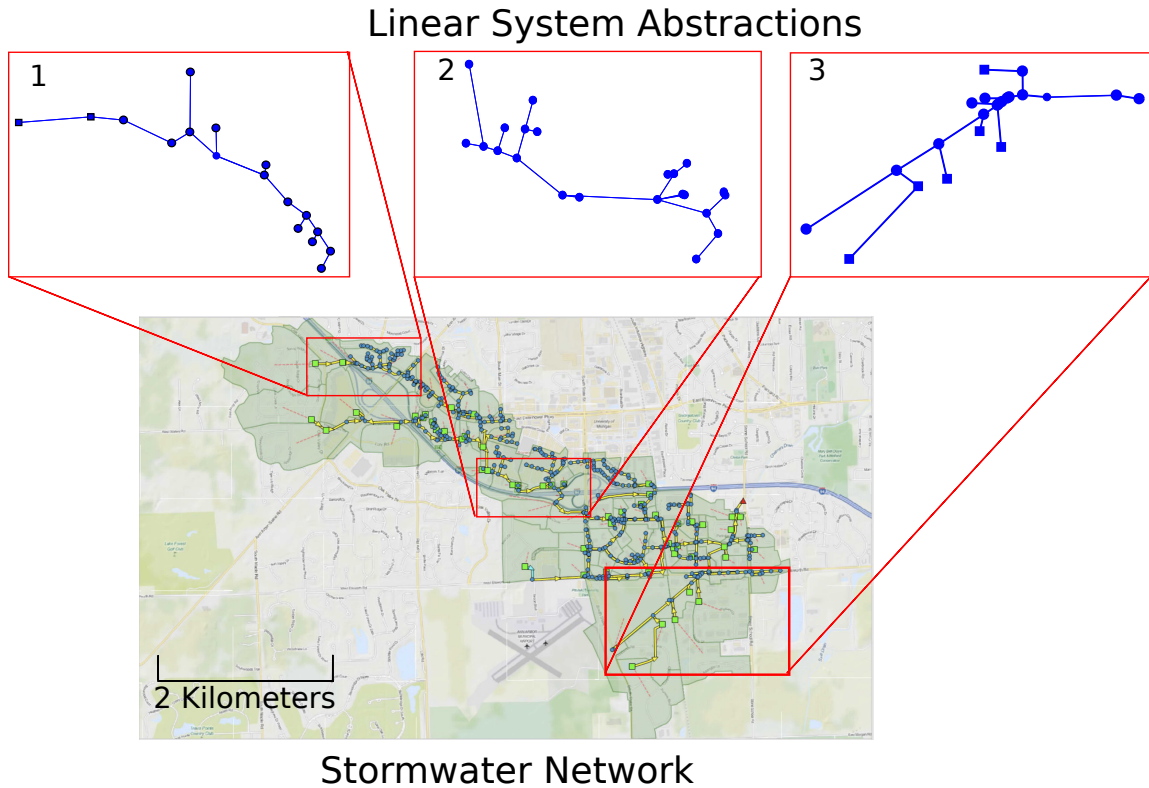


Figure 4.4: The stormwater network to be analyzed in this chapter. This particular network has an area of nearly $5km^2$. Here, the SWMM model is overlaid on a map of the area.

MatSWMM toolbox [93].

Once the time series of the inputs and output of each systems are simulated, the initial model is fit. The initial parameterization of the state space is randomized, and 10 random restarts are used to ensure that the algorithm does not converge on a local minimum. The random restart resulting in the best likelihood is then used as the initial model. The computational time is also noted for performance evaluation. For reference, the full approach was implemented on a laptop.

After fitting the initial linear model, pipes in the SWMM model are then “blocked,” one at a time, by setting their cross-sectional area to zero. A new simulation is run to generate the anomalous sensor signal at the outfall. With the new time series of the output, the likelihood search is performed by iteratively setting the parameters for a given node to zero, feeding the inputs into the state space model, and generating a modeled flow at the outfall.

Here, the likelihood is simply the MSE between the modeled flow and the simulated “true” flow. All of the sub networks are then analyzed across a series of damage scenarios and the performance of the approach is analyzed to determine how well it is able to detect the location of the damages. The performance of the approach is also broadly evaluated across a number of network topology types (sizes of the system, volume distributions, etc.) to determine if the performance can be explained through features of the network.

The likelihood of damage is evaluated visually by plotting the network and color-coding each node based on the likelihood it would yield the measured flow at the outfall. For interpretability, the likelihood is expressed using normalized RMSE, which is equivalent to the Nash Sutcliffe Efficiency:

$$E = 1 - \frac{\|\mathbf{y} - \hat{\mathbf{y}}\|}{\|\mathbf{y} - \bar{\mathbf{y}}\|} \quad (4.24)$$

where $\|\cdot\|$ is the 2-norm, \mathbf{y} is the observed signal, $\hat{\mathbf{y}}$ is the estimated signal, and $\bar{\mathbf{y}}$ is the mean of the observed signal. This yields a score between negative infinity and 1, where a value equal to zero indicates that the signal has a fit equivalent to just taking the mean value of the signal. A strong NRMSE is considered those above 0.7.

4.3 Results

4.3.1 Fit and Computational Complexity

The method proposed in this chapter is very computationally efficient compared to iteratively searching the parameters in a large scale hydrodynamic model. A network of 90 nodes (Figure 4.5) can be fit with a simplified state-space model in less than 30 minutes. This fit is purely based on the topology of the network and the training data for the system (the inputs from runoff and the output at the outfall as measured by a sensor). Further, fitting the models is done in linear time, with some small variability due to some random

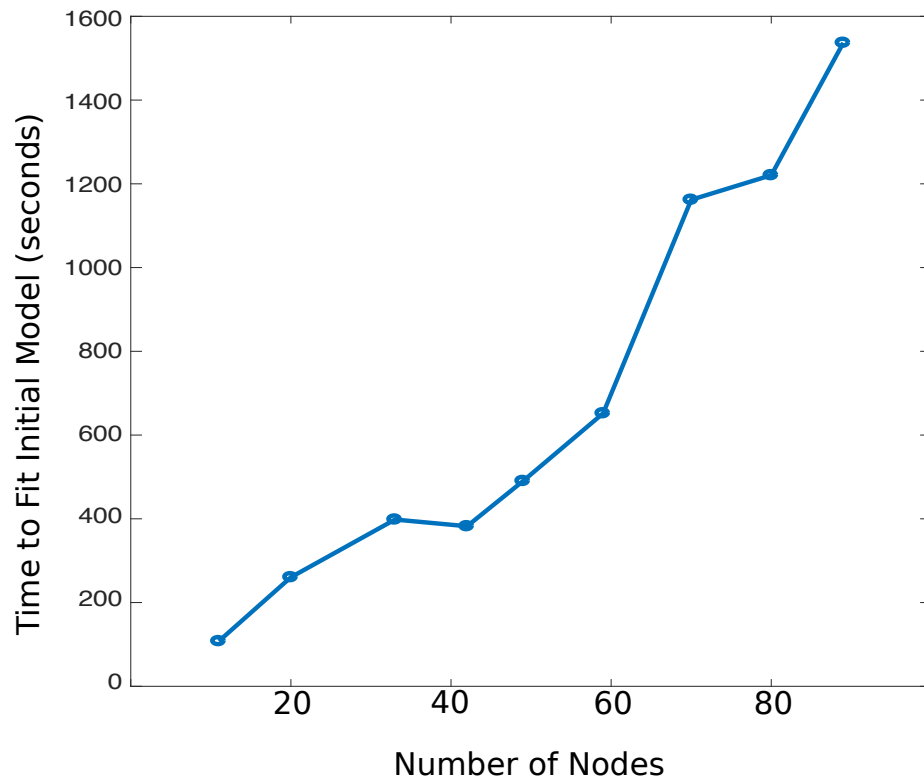


Figure 4.5: The computational time required to fit the initial state space model as a function of the number of nodes in the network. The model is fit in linear time, with some variation in convergence time due to random initialization of the parameters

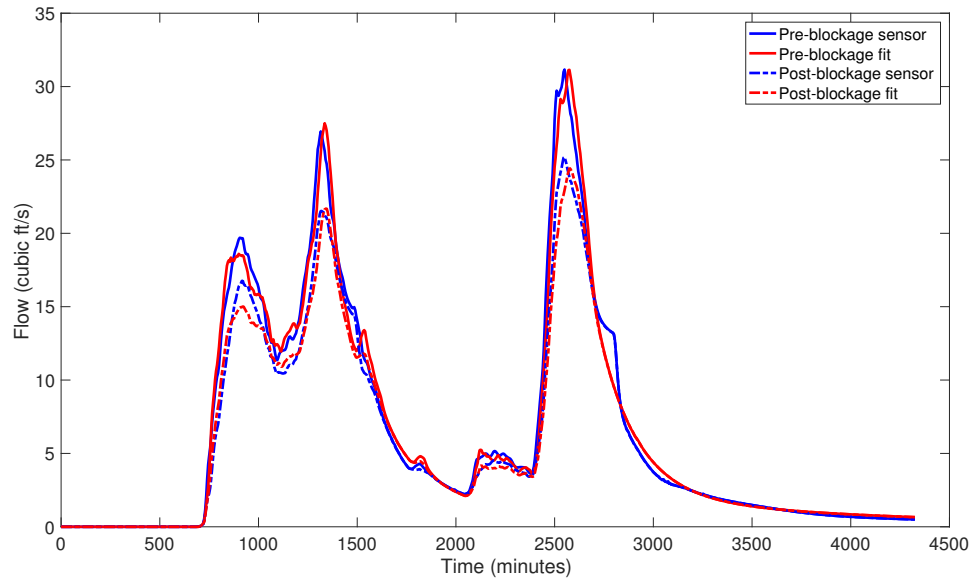


Figure 4.6: Time Series of modeled vs observed flows. The linearized model abstraction is capable of capturing the dynamics both before and after a pipe is damaged.

restarts converging faster than others.

Overall, the linear model is able to represent the flow dynamics across the studied system with a large degree of agreement (Figure 4.6). This is true for both the undamaged systems, as well as in cases when a pipe is damaged (linear model parameter set to zero). While model fit was not the intended goal of this work, this nonetheless validates the ability of simpler liner models to represented complex stormwater dynamics.

The search algorithm executes nearly instantaneously on a laptop (Figure 4.7). This is primarily due to the fact that the focus is on pipe blockages, and therefore no optimization procedure is needed to fit the parameters θ_i .

4.3.2 Damage Detection

Damage detection is evaluated across a number of different scenarios. Here, a set of examples is selected, including some that are able to localize the damage and some that do not. As an initial proof of concept, a small test subcatchment covering just under half a square

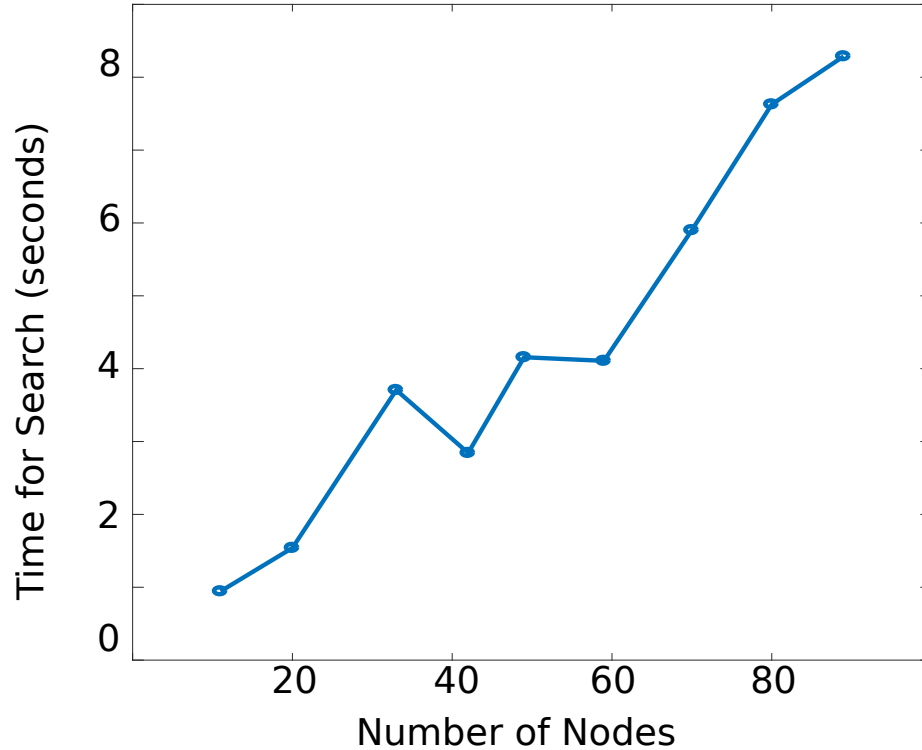


Figure 4.7: The computational time required to iterate over each node in a pipe blockage scenario and identify the damaged location.

mile (northwest corner in Figure 4.4) is first analyzed. The results are shown in Figure 4.8, where a main pipe was blocked and the algorithm was able to identify the proper node attached to the pipe. It is important to note that other candidate nodes were identified as well, but that the correctly identified node had the largest likelihood score. This is a relatively simple network with one main trunk line, so more complex networks must be analyzed.

In Figure 4.9, a complex network with multiple storage nodes and two main branches is tested. The algorithm is able to identify the broken pipe with a likelihood significantly larger than any of the other candidate nodes. In Figures 4.10 and 4.11, the networks comprise of two main branches that intersect near or at the outfall. In each of these cases as well, the algorithm was able to detect the broken pipe, even though many other locations had similarly large likelihoods.

However, there are scenarios where this algorithm does not identify damage correctly. The first scenario in which it cannot work is relatively simple (Figure 4.12). The pipe that

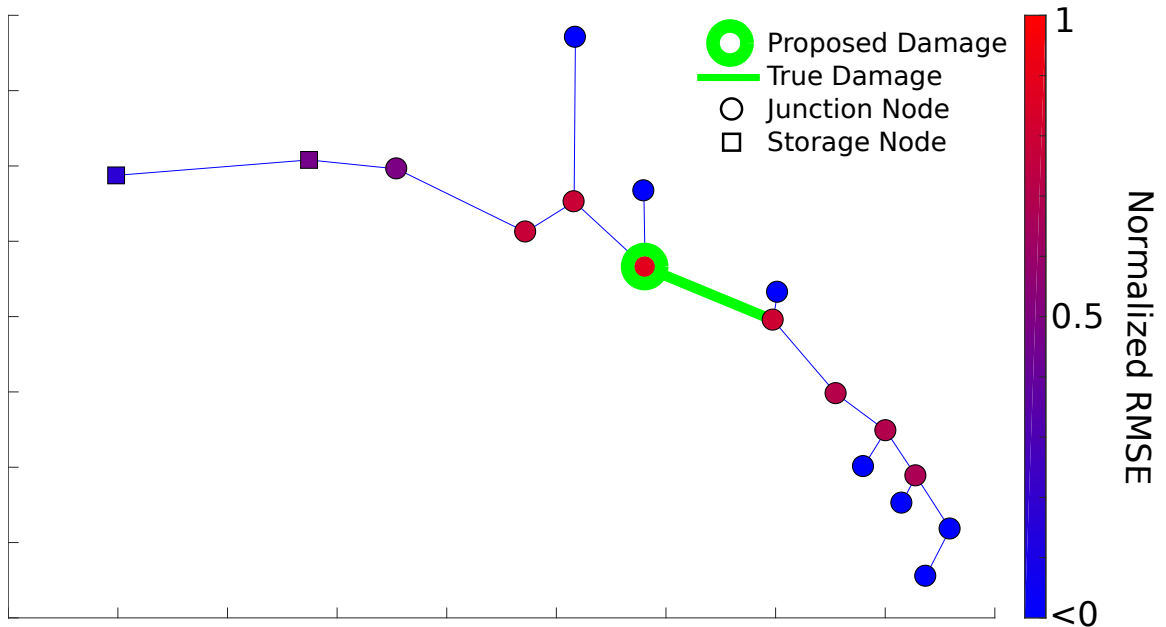


Figure 4.8: Example performance on a simple network. The green link represents the broken pipe, the green outlined node represents the node with the highest likelihood of damage, and the color of the node shows the relative likelihood as a measure of the normalized RMSE, with negative NRMSE values being set to zero. The more red a node, the higher the likelihood that it may be damaged, according to the proposed approach. Circles represent junctions while squares represent storage nodes.

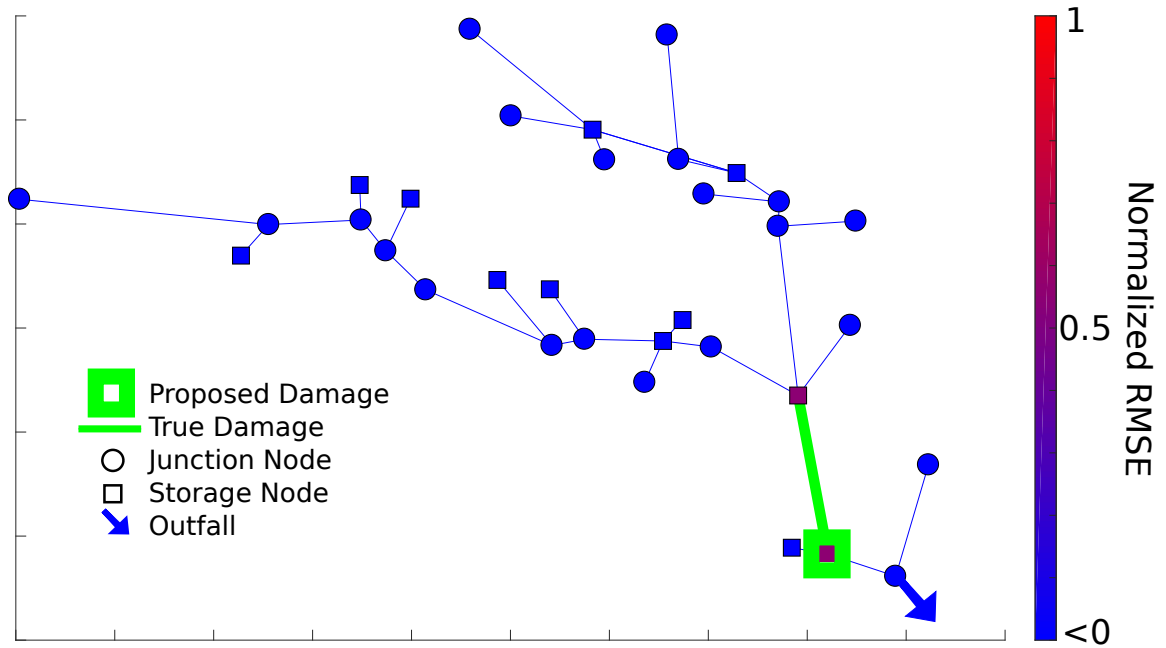


Figure 4.9: The algorithm is able to detect a pipe blockage with high confidence. The damaged pipe is very close to the outfall and contributes the majority of the flow.

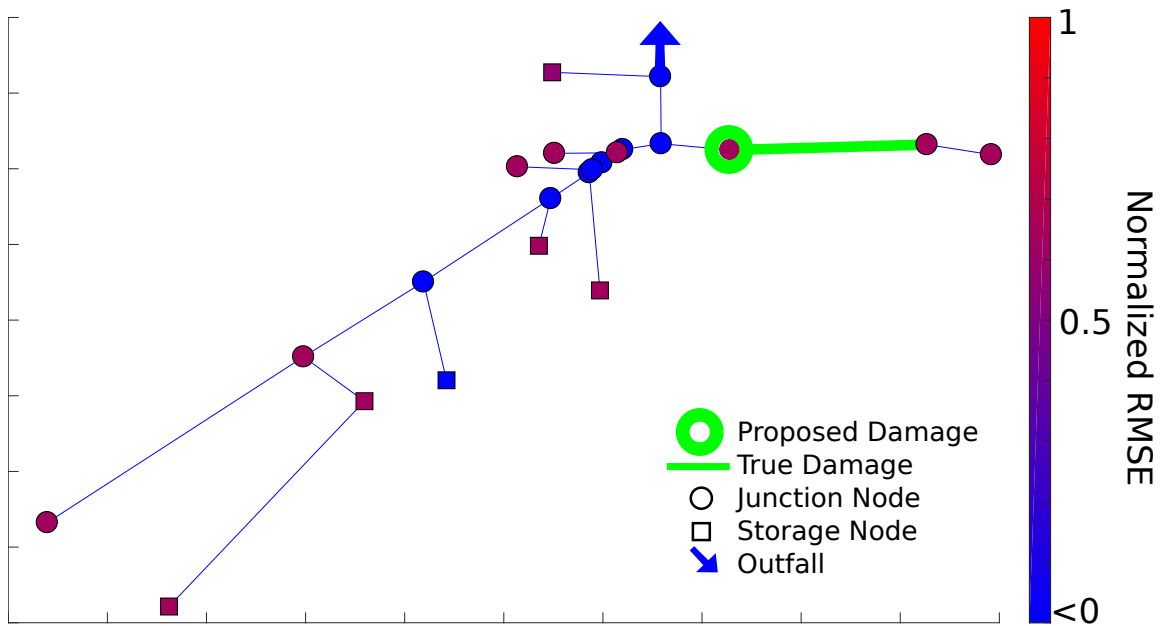


Figure 4.10: The algorithm is able to detect the pipe blockage when each branch of the system is contributing similar magnitudes of flow. Note that while many candidates were selected, the most likely candidate was the correct node.

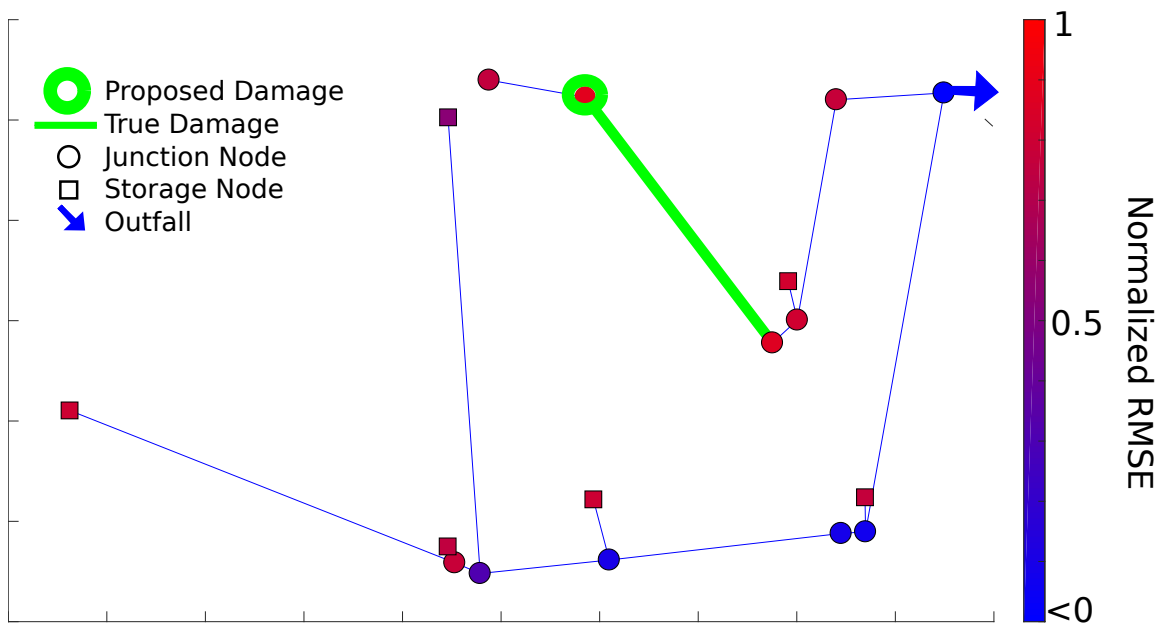


Figure 4.11: The algorithm is again able to detect the pipe blockage when each branch of the system is contributing similar magnitudes of flow.

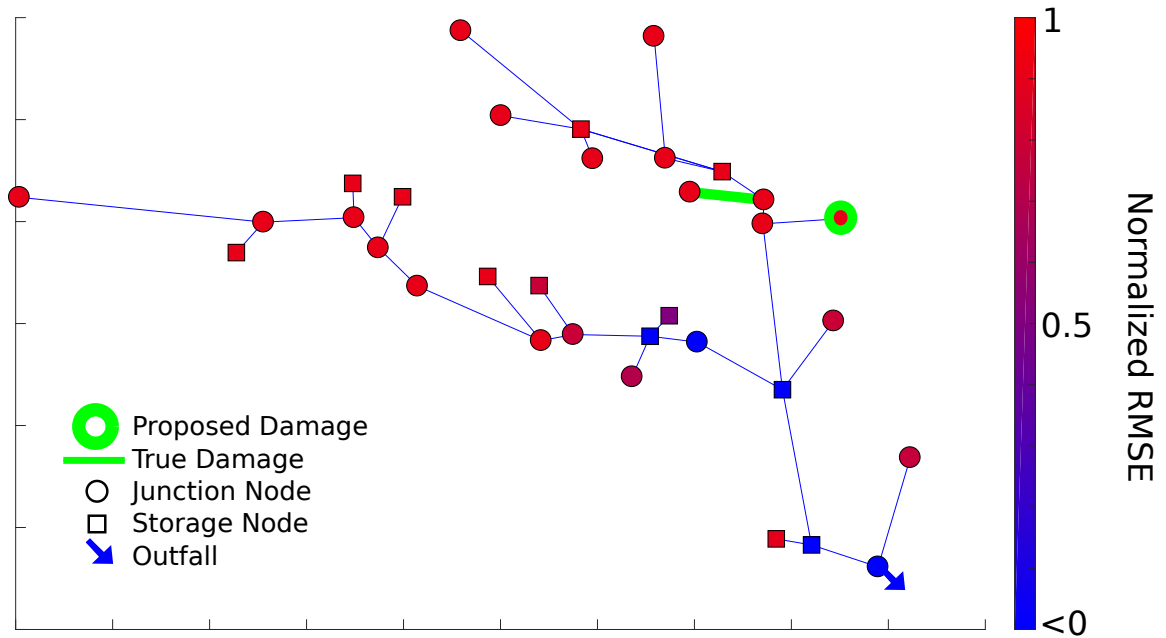


Figure 4.12: The algorithm does not identify the correct pipe because the contributing flow and time of concentration from that pipe is quite similar to the identified node.

was broken and the pipe that was identified by the algorithm each have similar volumes of water and travel times to the outfall, making it difficult to separate their contributing dynamics at the outlet of the system.

A second major scenario where this algorithm does not perform well is when the flow contribution from the broken pipe is small compared to the overall flow being measured at the outfall. This is evident in Figure 4.13. The flow that would normally pass through the broken pipe is approximately $2ft^3/s$ at its peak, while the flow passing through the outfall is approximately $100ft^3/s$. As such, the algorithm selects all the terminal nodes as the likely candidates because their RMSE values will all be only slightly different from the initial model.

The third major scenario where the algorithm does not perform well is when linearity does not sufficiently describe actual dynamics. This happens in two ways. First, when the water passes through too many segments of the system before reaching the outfall. The result is that the flow at the outfall can be represented similarly by a number of different parameterizations. Therefore, the initial model may not be accurately representing real-

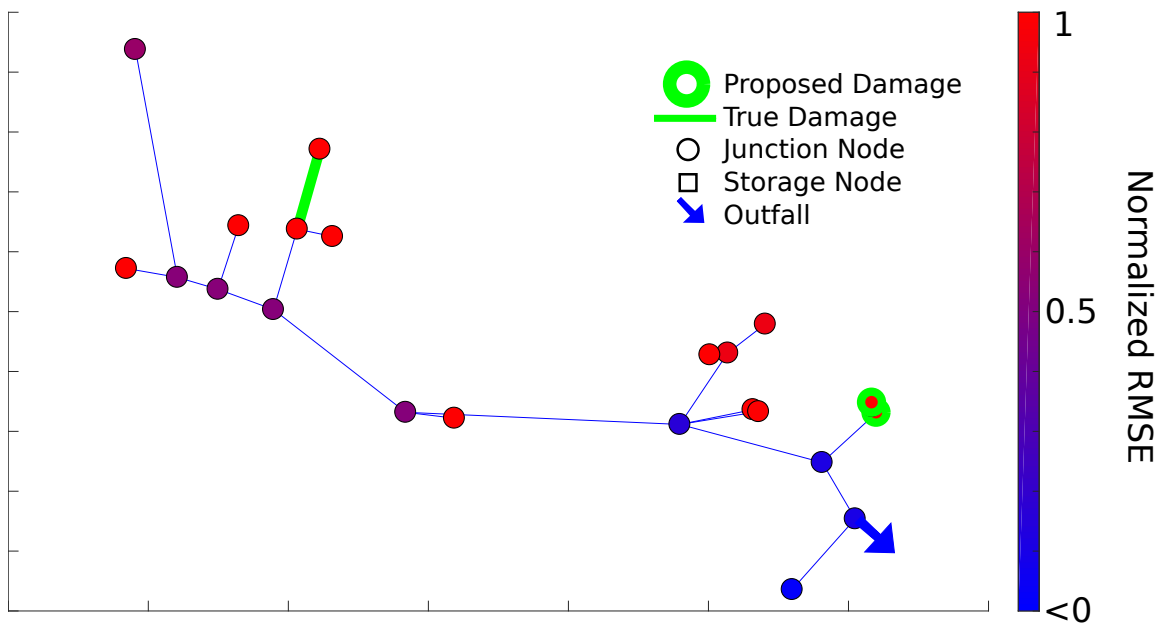


Figure 4.13: The algorithm does not identify the correct pipe because the contributing flow from the pipe is relatively small, with the majority of the flow entering the network from the top left branch. Therefore, all the terminal nodes are highlighted as candidate sites.

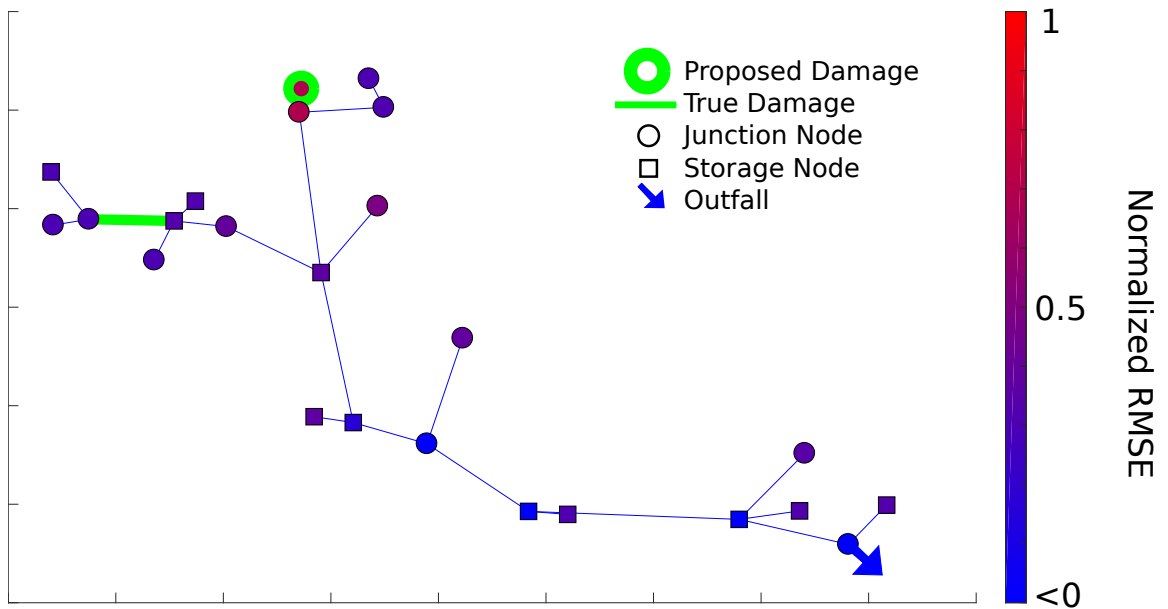


Figure 4.14: The algorithm does not identify the correct pipe because it is too far from the outfall and passes through too many storage nodes, resulting in non-linear dynamics at the outfall.

world dynamics. It is observed that linearity is also challenged at storage nodes, which have highly non-linear dynamics. The greater the volume of water in a storage node, the more rapidly it will discharge, which makes it difficult to capture the dynamics at a storage node using a single parameter. In Figure 4.14, the broken pipe is far upstream and the flow of water passes through many storage nodes (represented by squares in figure). As a result, none of the possible damage scenarios are given a very high likelihood.

4.4 Discussion

Currently, when an underlying process changes in a hydraulic or hydrologic model, detecting that change is a time consuming process, requiring many simulations and, often, intimate knowledge of the system being modeled. The method proposed herein is very efficient at identifying system changes because it leverages the directed graph topology of many water networks. With search simulations taking less than a second for networks as large as 100 nodes, it is possible to efficiently iterate on not just single blockage scenarios, but potentially for multi-blockage scenarios as well (Figure 4.7). This should make it very suitable for real-time applications.

While efficiency is important for enabling real-time asset management, the accuracy must still be strong. Currently, the method is reliable for a subset of scenarios, which can be generalized to provide guidelines for sensor deployment. These scenarios are tied to the *identifiability* of the system [94, 95]. For a system to be uniquely identifiable a number of requirements must be met to ensure that there can only be a single parameterization that defines the system. First, the system must be sufficiently excited. That is, the inputs must be large enough to perturb the system components. Therefore, small rain events will not be sufficient for detecting system changes. This is why larger rain events, roughly 1-year storms, were used for this analysis.

The second is centered around the parameters themselves. For a system to be globally

identifiable, the Jacobian of its Markov Parameter Matrix must be full rank [94], where the Markov Parameter Matrix is:

$$G(\theta) = \begin{bmatrix} D(\theta) \\ C(\theta)B(\theta) \\ C(\theta)A(\theta)B(\theta) \\ \vdots \\ C(\theta)A(\theta)^{2n-1}B(\theta) \end{bmatrix} \quad (4.25)$$

where A , B , and C are the matrices as outlined in Equation 4.3 and $D = [0 \dots 0] \in \mathbb{R}^m$ where m is the number of inputs (subcatchments). Unfortunately, due to the sparse nature of this problem with a single output and few terms on the diagonal of the A matrix, the large majority of stormwater systems will not be globally identifiable. This means that there will be multiple parameterizations of the initial model fit that can yield identical results at the outflow of the system as measured by a sensor.

However, there is the potential to determine if subsystems of the overall system are uniquely identifiable. For instance, the work detailed in [96] describes how *compartments* of systems can be identified. Compartments are essentially subsystems that are each input/output connectable. This test can be used to determine if simplifications and subdivisions into compartments can be made to yield an identifiable system. Then, the algorithm outlined herein can be used to determine which “compartments” are likely to have undergone some change as opposed to finding one specific node. At this time, however, there is no heuristic for optimally determining the best compartmentalization of the system. This is because much of the research to date has focused on systems with only a few components, and therefore an exhaustive search can be done efficiently. More research is necessary to enable the efficient compartmentalization of systems larger than a few nodes.

In general, this chapter is able to identify a few requirements for an identifiable system based on the simulations presented. The first requirement for successful identification is

heterogeneity of the stormwater system. For example, two pipes with similar contributing volumes and times of concentration will be indistinguishable from each other at the outfall. Alternatively, two pipes with similar times of concentration but different contributing volumes will still be distinguishable by the algorithm. If a network has homogeneous elements, the algorithm will be effective at proposing multiple candidates, instead of just one. This ties back into the requirement that the Jacobian of the Markov parameter matrix should be full rank. If we have a system that is nearly full rank, then subsystems of the overall network should be identifiable. This should still help decision makers because it reduces the number of nodes that need to be inspected for damage. The algorithm will work best at locating the exact damage when the network is heterogeneous so that different pipes are distinct.

The algorithm also does not perform strongly when flows at damaged locations are relatively small compared to the outflow. This is tied to the *sufficient excitation* requirement. Simply put, if a small pipe breaks upstream, it is hard to detect this change at the outflow to begin with. If the blocked and unblocked signals are very similar, then the algorithm will have trouble distinguishing smaller upstream nodes from each other. While this is difficult to avoid, there is promise to being able to distinguish the “true” broken link from the others. As illustrated in Figures 4.10 and 4.11, the identified nodes did not have significantly higher likelihoods than other nodes within the system. However, the proper damage location was still identified. This was a consistent trend. Despite a relatively large number of likely candidates, the most likely candidate was consistently the node that was actually damaged. This, of course, depends on having an initial fit for the state-space model that well represents the real-world system dynamics.

Additionally, if a network is too large, it becomes difficult to identify broken pipes due to the complexity of flows, again tying back to the full rank requirement of $G'(\theta)$. Each time the flow passes through a node, the state space transforms the flow. In these cases, a challenge arises related to the uniqueness of the linear model fit. As the number of

flow transfers increases, a wider array of potentially equivalent parameterizations becomes possible for representing the output signal. Identifying which of those parameterizations is correct is not necessarily possible with only a single observation at the outflow. Therefore, when a pipe is blocked, there can be a number of locations in the network that may match the dynamics similarly. More research must be done to determine how large the model may be before uniqueness becomes an issue, and whether constraint-based optimization may enable unique parameterizations to arise. Here, networks of around 30 nodes or fewer are detectable for damage.

Finally, linear-time-variant and nonlinear storage nodes pose a challenge to the linear-time-invariant representation. In a stormwater network there may often be underground cisterns, ponds, swales. Most of these storage nodes behave non-linearly. Larger volumes of water tend to discharge at a faster rate, many nodes have multiple outlets for various heights of water, and some can even have passive or active controls. To address this challenge, a number of modifications can be made. If a storage node has multiple outlets, then indicator functions are needed. That is, the parameterization is a function of the magnitude of x_i being above or below a certain threshold. Other storage nodes are designed to have a certain retention time, so the parameterization is a function of the time since the flow entered the pond. These nodes may be linear, but are not time-invariant. Work must be done to evaluate how these situations can best be represented and estimated using linear models. This process most likely will involve adding further states to the model. For instance, if representing a storage node requires knowledge of the previous m time steps, then m more states would need to be added to the model and parameterized. Finding a simple yet robust way to model these non-linear processes will go a long way towards making this approach operationally viable.

While there are a number of issues to still address, this methodology does show significant promise. It has proven capable of detecting pipe blockages on smaller heterogeneous networks (Figures 4.9, 4.10, 4.11), and it has done so without the need for a hydrodynamic

model. This is an important distinction. If a municipality already has accurate data for runoff and outflow, or are capable of simulating it accurately using SCS or other methods, then this method can represent the system effectively. The downside, in the specific case of this current implementation, is that the method can currently only detect *new* damage. However, if a municipality had a well-calibrated stormwater model at some point in its history, then this method could also be used to detect changes to the system in the past, not just for real-time detection.

Another advantage of this method is that it probabilistically locates candidates. Even if it does not always identify one single problem location, the method poses a great first step towards providing decision makers with information that would otherwise be difficult to systematically detect. As the initial model fit becomes more representative of real-world dynamics, the detectability of the blocked pipe will increase. Yet, as is evident in Figure 4.11, even with multiple candidates, the algorithm detects the proper pipe on many occasions among a group viable locations.

A further benefit of this methodology is that it can provide decision makers with a lower bound sensor placement strategy. The method can reliably identify blocked pipes with low network depth and high contributing flow ratio, so with this knowledge a municipality can begin to deploy sensors to partition the network into many subnetworks of approximately 30 nodes. As the method is improved, the spacing of these sensors, and therefore the size of the subnetworks, will likely be able to be increased. With an expanded formulation, multiple sensors could also be used to help refine the damage estimates by eliminating unlikely candidates.

Finally, while the case study presented here focused on real-time asset management, the method may be directly applicable to any situation where a complex model can be represented by a directed graph. This may one day include other water systems such as stream networks, sewer systems, and possibly even integrated models, where there are many submodels feeding into a larger forecast. Representing each of these submodels by

a small parameter set can guide the modeler towards understanding where changes may be occurring.

4.5 Conclusion

This work provides a valuable first step to demonstrating the utility of simplified model abstractions for damage localization in stormwater systems. Presently, real-time asset management is limited by the ability to either deploy large observation networks or the ability to compute many simulations rapidly. This method, while still in its early stages, will enable municipalities to perform real-time asset management with limited sensing resources. Future work will be necessary to address non-linearities at storage nodes and the level of heterogeneity needed to enable unique and accurate representations of real-world dynamics.

CHAPTER 5

Thesis Conclusion

5.1 Summary of Discoveries

The goal of this dissertation was to advance the ability to understand water systems by enabling the fusion of large hydrologic and hydraulic models with new forms of sensor data. To that end, a number of fundamental discoveries were made in each chapter:

Chapter 2: I discovered that noisy, irregular observations could be fused with state of the art models to reveal dynamics that were being modeled insufficiently through numerical modeling alone. This work will allow operational models to ingest a greater wealth of calibration data so that future forecasts can better capture hydrometeorological dynamics and improve decision making.

Chapter 3: I discovered that coarse forecasts from large scale models could be dynamically mapped to site-scale estimates by coupling with sensor observations. This work will allow decision makers to deploy sensors and rapidly couple their measurements with a large-scale model. This will save time and money while also providing much more useful and timely information.

Chapter 4: I discovered that simplified model abstractions can be used to detect where a hydrologic or hydraulic process may have changed in near real-time. This will

enable real-time asset management without the need re-calibrate large models or maintain very large sensor networks.

5.2 Future Directions

This dissertation made a number of foundational discoveries that should be investigated further to support the emergence of *Big Data Hydrology*. In Chapter 2, much of the work was performed in a post-processed setting, months to years after the observations were made. To incorporate non-standard sensor measurements into operational forecasts, more research should be conducted to create an on-line machine learning platform. Additionally, while the method itself will be valuable to the scientific community, more research should be done to analyze the specific results from the case study on the Great Lakes. Given the time period of 2006-2014, this dataset provides the scientific community with valuable estimates of overlake dynamics during the time period in which the Great Lakes rebounded from the El Niño as well as experienced to the “polar vortex”. Further, there is a consensus within the community that wind speeds will continue to increase on the Great Lakes due to rising air temperatures, but there is evidence to suggest in this dataset that this may not be the case.

For the research performed in Chapter 3, the method was based on a causal relationship between the forecasted flow and the observed water level. But because these two variables are coupled, introducing a non-causal relationship may yield improved results. The major limitation, however, was the assumption of accurate, near perfect, forecasts. As such, more work is needed to investigate the role of uncertain forecasts and weather inputs. As understanding improves of how the methodology behaves in other cities and locales, this method may be used for sensor placement strategies or even enable forecasting of stream heights at unaged locations.

For the research performed in Chapter 4, a significant amount of future work is needed

before the method can become operational. As stated previously, the method works across a specific set of network types. Research must be conducted to effectively and efficiently model non-linear processes within the systems. Further, improved constraints on the initial model fit should limit the issue of non-uniqueness caused by the “curse of dimensionality”. Once these primary issues are addressed, more research should then be conducted to understand at what scales this method can work effectively, which will aid in sensor placement .

As novel sources of water data continue to emerge, improved tools and methods are needed to properly combine these data with knowledge embedded in more classical hydrologic and hydraulic models. By extracting information from non-standard data sources and coupling it with the domain expertise embedded in large-scale models, the field of hydrology will be able to move forward into the era of *Big Data*. The work in this dissertation only scratches the surface of what may be learned by applying machine learning and systems theory to hydrology, but much research remains to be conducted to support this emerging field.

BIBLIOGRAPHY

- [1] Vörösmarty, C. J., Green, P., Salisbury, J., and Lammers, R. B., “Global Water Resources: Vulnerability from Climate Change and Population Growth,” *Science*, Vol. 289, No. 5477, 2000, pp. 284–288.
- [2] Wood, E. F., Roundy, J. K., Troy, T. J., van Beek, L. P. H., Bierkens, M. F. P., Blyth, E., de Roo, A., Dll, P., Ek, M., Famiglietti, J., Gochis, D., van de Giesen, N., Houser, P., Jaff, P. R., Kollet, S., Lehner, B., Lettenmaier, D. P., Peters-Lidard, C., Sivapalan, M., Sheffield, J., Wade, A., and Whitehead, P., “Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth’s terrestrial water,” *Water Resources Research*, Vol. 47, No. 5, 2011, pp. n/a–n/a, W05301.
- [3] Wong, B. P. and Kerkez, B., “Real-time environmental sensor data: An application to water quality using web services,” *Environmental Modelling & Software*, Vol. 84, 2016, pp. 505 – 517.
- [4] Moradkhani, H., Hsu, K.-L., Gupta, H., and Sorooshian, S., “Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter,” *Water Resources Research*, Vol. 41, No. 5, 2005, pp. n/a–n/a, W05012.
- [5] Gronewold, A. D. and Stow, C. A., “Unprecedented Seasonal Water Level Dynamics on One of the Earth’s Largest Lakes,” *Bulletin of the American Meteorological Society*, Vol. 95, No. 1, 2014, pp. 15–17.
- [6] Marchand, D., Sanderson, M., Howe, D., and Alpaugh, C., “Climatic change and great lakes levels the impact on shipping,” *Climatic Change*, Vol. 12, No. 2, Apr 1988, pp. 107–133.
- [7] Mazzoleni, M., Alfonso, L., and Solomatine, D., “Influence of spatial distribution of sensors and observation accuracy on the assimilation of distributed streamflow data in hydrological modelling,” *Hydrological Sciences Journal*, Vol. 62, No. 3, 2017, pp. 389–407.
- [8] Kauffeldt, A., Wetterhall, F., Pappenberger, F., Salamon, P., and Thielen, J., “Technical review of large-scale hydrological models for implementation in operational flood forecasting schemes on continental level,” *Environmental Modelling & Software*, Vol. 75, 2016, pp. 68 – 76.

- [9] Wood, A. W., Leung, L. R., Sridhar, V., and Lettenmaier, D. P., “Hydrologic Implications of Dynamical and Statistical Approaches to Downscaling Climate Model Outputs,” *Climatic Change*, Vol. 62, No. 1, Jan 2004, pp. 189–216.
- [10] Wong, H., Hu, B., Ip, W., and Xia, J., “Change-point analysis of hydrological time series using grey relational method,” *Journal of Hydrology*, Vol. 324, No. 1, 2006, pp. 323 – 338.
- [11] Larsen, T. A., Hoffmann, S., Lüthi, C., Truffer, B., and Maurer, M., “Emerging solutions to the water challenges of an urbanizing world,” *Science*, Vol. 352, No. 6288, 2016, pp. 928–933.
- [12] Kessler, R., “Stormwater Strategies: Cities Prepare Aging Infrastructure for Climate Change,” *Environmental Health Perspectives*, Vol. 119, 2011, pp. A514–A519.
- [13] Gronewold, A. D. and Stow, C. A., “Water Loss from the Great Lakes,” *Science*, 2014, pp. 1084–1085.
- [14] Tauro, F., Porfiri, M., and Grimaldi, S., “Surface flow measurements from drones,” *Journal of Hydrology*, 2016, pp. 240–245.
- [15] NOAA GLERL, “About Our Great Lakes: Great Lakes Basin Facts,” .
- [16] Desai, A. R., Austin, J. A., Bennington, V., and McKinley, G. A., “Stronger winds over a large lake in response to weakening air-to-lake temperature gradient,” *Nature Geoscience*, 2009, pp. 855–858.
- [17] Stow, C. and Lamon, E., “Lake Superior Water Level Fluctuation and Climatic Factors: A Dynamic Linear Model Analysis,” *Journal of Great Lakes Research*, 2010, pp. 172–178.
- [18] Spence, C., Blanken, P., Hedstrom, N., and Lenters, J., “Evaporation from Lake Superior: 1. Physical controls and processes,” *Journal of Great Lakes Research*, 2011.
- [19] US Army Corps of Engineers, Detroit District, “Monthly Bulletin of Great Lakes Water Levels,” .
- [20] Anderson, E., Rowe, M., Xu, J., Zhang, A., Lang, G., Kelley, J., and Stumpf, R., “Using the Next-Generation Great Lakes Operational Forecasting System (GLOFS) to Predict Harmful Algal Bloom (HAB) Transport with the HAB Tracker,” *The 33rd Conference on Environmental Information Processing Technologies*, Seattle, WA, 2017.
- [21] International Upper Great Lakes Study, “Lake Superior Regulation: Addressing Uncertainty in Upper Great Lakes Water Levels,” Tech. rep., International Joint Commission, 2012.

- [22] Zhao, L., Jin, J., Wang, S., and Ek, M., “Integration of remote-sensing data with WRF to improve lake-effect precipitation simulations over the Great Lakes region,” *Journal of Geophysical Research: Atmospheres*, 2012.
- [23] Voss, K. A., Famiglietti, J. S., Lo, M., de Linage, C., Rodell, M., and Swenson, S. C., “Groundwater depletion in the Middle East from GRACE with implications for transboundary water management in the Tigris-Euphrates-Western Iran region,” *Water Resources Research*, Vol. 49, No. 2, 2013, pp. 904–914.
- [24] de Paiva, R. C. D., Buarque, D. C., Collischonn, W., Bonnet, M.-P., Frappart, F., Calmant, S., and Bulhes Mendes, C. A., “Large-scale hydrologic and hydrodynamic modeling of the Amazon River basin,” *Water Resources Research*, Vol. 49, No. 3, 2013, pp. 1226–1243.
- [25] Laird, N. F. and Kristovich, D. A., “Variations of Sensible and Latent Heat Fluxes from a Great Lakes Buoy and Associated Synoptic Weather Patterns,” *Journal of Hydrometeorology*, Vol. 3, 2002, pp. 3–12.
- [26] Spence, C., Blanken, P., Lenters, J., and Hedstrom, N., “The Importance of Spring and Autumn Atmospheric Conditions for the Evaporation Regime of Lake Superior,” *Journal of Hydrometeorology*, Vol. 14, 2013, pp. 1647–1658.
- [27] Cleave, K. V., Lenters, J. D., Wang, J., and Verhamme, E. M., “A regime shift in Lake Superior ice cover, evaporation, and water temperature following the warm El Niño winter of 1997-1998,” *Limnology and Oceanography*, 2014, pp. 1889–1898.
- [28] Watkins, D., Li, H., and Cowden, J., “Adjustment of radar-based precipitation estimates for Great Lakes hydrologic modeling,” *Journal of Hydrologic Engineering*, 2007, pp. 298–305.
- [29] Schwab, D. J. and Morton, J. A., “Estimation of Overlake Wind Speed from Overland Wind Speed: A Comparison of Three Methods,” *Journal of Great Lakes Research*, 1984, pp. 68–72.
- [30] Holman, K., Gronewold, A., Notaro, M., and Zarrin, A., “Improving historical precipitation estimates over the Lake Superior basin,” *Geophysical Research Letters*, 2012.
- [31] Glahn, H. R. and Ruth, D. P., “The New Digital Forecast Database of the National Weather Service,” *Bulletin of the American Meteorological Society*, 2003.
- [32] Chu, P. Y., Kelley, J. G., Mott, G. V., Zhang, A., and Lang, G. A., “Development, implementation, and skill assessment of the NOAA/NOS Great Lakes Operational Forecast System,” *Ocean Dynamics*, 2011, pp. 1305–1316.
- [33] Kent, E. C., Berry, D. I., Woodruff, S. D., and Taylor, P. K., “Voluntary Observing Ships: A Vital Marine Observing System in Decline,” *CLIVAR Exchanges*, Vol. 11, 2006, pp. 20–21.

- [34] Kent, E. C., Challenor, P. G., and Taylor, P. K., “A Statistical Determination of the Random Observational Errors Present in Voluntary Observing Ships Meteorological Reports,” *Journal of Atmospheric and Oceanic Technology*, Vol. 16, 1999, pp. 905–914.
- [35] Gulev, S. K., Grigorieva, V., Sterl, A., and Woolf, D., “Assessment of the reliability of wave observations from voluntary observing ships: Insights from the validation of a global wind wave climatology based on voluntary observing ship data,” *Journal of Geophysical Research*, Vol. 108, 2003.
- [36] Dong, J., Steele-Dunne, S. C., Ochsner, T. E., and van de Giesen, N., “Estimating soil moisture and soil thermal and hydraulic properties by assimilating soil temperatures using a particle batch smoother,” *Advances in Water Resources*, Vol. 91, 2016, pp. 104 – 116.
- [37] Rasmussen, C. and Williams, C., *Gaussian Processes for Machine Learning*, MIT Press, 2006.
- [38] Grus, J., *Data Science from Scratch*, O’Reilly, 2015.
- [39] Krause, A., Singh, A., and Guestrin, C., “Near-Optimal Sensor Placements in Gaussian Processes: Theory, Efficient Algorithms and Empirical Studies,” *Journal of Machine Learning Research*, 2008, pp. 235–284.
- [40] Welch, S. C., Kerkez, B., Bales, R. C., Glaser, S. D., Rittger, K., and Rice, R. R., “Sensor placement strategies for snow water equivalent (SWE) estimation in the American River basin,” *Water Resources Research*, Vol. 49, 2013, pp. 891–903.
- [41] Blöschl, G., Bárdossy, A., Koutsoyiannis, D., Kundzewicz, Z. W., Littlewood, I., Montanari, A., and Savenije, H., “On the future of journal publications in hydrology,” *Water Resources Research*, Vol. 50, No. 4, 2014, pp. 2795–2797.
- [42] Office of Water Prediction, “The National Water Model,” 2017.
- [43] Karandish, F. and Šimůnek, J., “A comparison of numerical and machine-learning modeling of soil water content with limited input data,” *Journal of Hydrology*, Vol. 543, No. Part B, 2016, pp. 892 – 909.
- [44] Tiwari, M. K. and Adamowski, J. F., “Medium-Term Urban Water Demand Forecasting with Limited Data Using an Ensemble Wavelet-Bootstrap Machine-Learning Approach,” *Journal of Water Resources Planning and Management*, Vol. 141, No. 2, 2015, pp. 04014053.
- [45] Chang, N.-B., Bai, K., and Chen, C.-F., “Integrating multisensor satellite data merging and image reconstruction in support of machine learning for better water quality management,” *Journal of Environmental Management*, Vol. 201, No. Supplement C, 2017, pp. 227 – 240.

- [46] Jin, N., Ma, R., Lv, Y., Lou, X., and Wei, Q., “A novel design of water environment monitoring system based on WSN,” *2010 International Conference On Computer Design and Applications*, Vol. 2, June 2010, pp. V2–593–V2–597.
- [47] Gilles, D., Young, N., Schroeder, H., Piotrowski, J., and Chang, Y.-J., “Inundation Mapping Initiatives of the Iowa Flood Center: Statewide Coverage and Detailed Urban Flooding Analysis,” *Water*, Vol. 4, No. 1, 2012, pp. 85–106.
- [48] Bitella, G., Rossi, R., Bochicchio, R., Perniola, M., and Amato, M., “A Novel Low-Cost Open-Hardware Platform for Monitoring Soil Water Content and Multiple Soil-Air-Vegetation Parameters,” *Sensors*, Vol. 14, No. 10, 2014, pp. 19639–19659.
- [49] Bartos, M. D., Wong, B., and Kerkez, B., “Open storm: a complete framework for sensing and control of urban watersheds,” *Environ. Sci.: Water Res. Technol.*, 2017, pp. –.
- [50] US Geological Survey, “NHDPlus High Resolution National Hydrography Dataset Watershed Boundary Dataset,” 2017.
- [51] Herschy, R., *Hydrometry: Principles and Practice*, Wiley, 1999.
- [52] Damangir, H. and Abedini, M., “System identification and subsequent discharge estimation based on level data alone—Gradually varied flow condition,” *Flow Measurement and Instrumentation*, Vol. 36, 2014, pp. 24 – 31.
- [53] Aricò, C., Corato, G., Tucciarelli, T., Meftah, M. B., Petrillo, A. F., and Mossa, M., “Discharge estimation in open channels by means of water level hydrograph analysis,” *Journal of Hydraulic Research*, Vol. 48, No. 5, 2010, pp. 612–619.
- [54] Petersen-Overlier, A., “Accounting for heteroscedasticity in rating curve estimates,” *Journal of Hydrology*, Vol. 292, No. 1, 2004, pp. 173 – 181.
- [55] Kuczera, G., “Correlated Rating Curve Error in Flood Frequency Inference,” *Water Resources Research*, Vol. 32, No. 7, 1996, pp. 2119–2127.
- [56] Perumal, M., Shrestha, K. B., and Chaube, U., “Reproduction of Hysteresis in Rating Curves,” *Journal of Hydraulic Engineering*, Vol. 130, No. 9, 2004, pp. 870–878.
- [57] Westerberg, I., Guerrero, J.-L., Seibert, J., Beven, K. J., and Halldin, S., “Stage-discharge uncertainty derived with a non-stationary rating curve in the Choluteca River, Honduras,” *Hydrological Processes*, Vol. 25, No. 4, 2011, pp. 603–613.
- [58] Beck, M. B., “Water quality modeling: A review of the analysis of uncertainty,” *Water Resources Research*, Vol. 23, No. 8, 1987, pp. 1393–1442.
- [59] U. S. Geological Survey and Koltun, G. F., “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, 2015.

- [60] U. S. Geological Survey and Southard, R. E., “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, 2013.
- [61] Luenberger, D., *Introduction to Dynamic Systems: Theory, Models, and Applications*, Wiley, 1979.
- [62] Nash, J., “The Form of the Instantaneous Unit Hydrograph,” *International Association of Hydrological Sciences Publications*, Vol. 45, 1957, pp. 114–121.
- [63] Yang, Z. and Han, D., “Derivation of unit hydrograph using a transfer function approach,” *Water Resources Research*, Vol. 42, No. 1, 2006, pp. n/a–n/a, W01501.
- [64] Cluckie, I. and Harpin, R., “A real-time simulator of the rainfall-runoff process,” *Mathematics and Computers in Simulation*, Vol. 24, No. 2, 1982, pp. 131 – 139.
- [65] Bjorck, A., *Numerical Methods for Least Squares Problems*, Society for Industrial and Applied Mathematics, 1996.
- [66] Ouyang, Y., “Evaluation of river water quality monitoring stations by principal component analysis,” *Water Research*, Vol. 39, No. 12, 2005, pp. 2621 – 2635.
- [67] Freund, Y., Schapire, R. E., et al., “Experiments with a new boosting algorithm,” *International Conference on Machine Learning*, Vol. 96, 1996, pp. 148–156.
- [68] National Severe Storms Laboratory, “MULTI-RADAR/MULTI-SENSOR SYSTEM (MRMS),” 2017.
- [69] Demir, I. and Krajewski, W. F., “Towards an integrated Flood Information System: Centralized data access, analysis, and visualization,” *Environmental Modelling & Software*, Vol. 50, 2013, pp. 77–84.
- [70] Ljung, L., *System identification: theory for the user*, Prentice-Hall information and system sciences series, Prentice-Hall, 1987.
- [71] U.S. Geological Survey, Turnipseed, D. P., and Sauer, V. B., “Discharge measurements at gaging stations,” Tech. rep., 2010.
- [72] Nash, J. and Sutcliffe, J., “River flow forecasting through conceptual models part I: A discussion of principles,” *Journal of Hydrology*, Vol. 10, No. 3, 1970, pp. 282 – 290.
- [73] Deza, M. and Deza, E., *Encyclopedia of Distances*, Encyclopedia of Distances, Springer Berlin Heidelberg, 2009.
- [74] Krause, P., Boyle, D. P., and Bäse, F., “Comparison of different efficiency criteria for hydrological model assessment,” *Advances in Geosciences*, Vol. 5, Dec. 2005, pp. 89–97.

- [75] Hastie, T., Tibshirani, R., and Friedman, J., *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer series in statistics, Springer, 2001.
- [76] Strahler, A. N., “Quantitative analysis of watershed geomorphology,” *Eos, Transactions American Geophysical Union*, Vol. 38, No. 6, 1957, pp. 913–920.
- [77] Horn, R. and Johnson, C., *Matrix Analysis*, Cambridge University Press, 1990.
- [78] Zhou, Z., *Ensemble Methods: Foundations and Algorithms*, Chapman & Hall/CRC Data Mining and Knowledge Discovery Series, Taylor & Francis, 2012.
- [79] Perumal, M., Sahoo, B., Moramarco, T., and Barbetta, S., “Multilinear Muskingum Method for Stage-Hydrograph Routing in Compound Channels,” *Journal of Hydrologic Engineering*, Vol. 14, 2009.
- [80] Sahoo, B., “Field Application of the Multilinear Muskingum Discharge Routing Method,” *Water Resources Management*, Vol. 27, No. 5, Mar 2013, pp. 1193–1205.
- [81] Barbosa, A., Fernandes, J., and David, L., “Key issues for sustainable urban stormwater management,” *Water Research*, Vol. 46, No. 20, 2012, pp. 6787 – 6798, Special Issue on Stormwater in urban areas.
- [82] Grigg, N., *Water, Wastewater, and Stormwater Infrastructure Management, Second Edition*, Taylor & Francis, 2012.
- [83] EPA Office of Research and Development, *Storm Water Management Model User’s Manual Version 5.1*, United States Environmental Protection Agency, 2015.
- [84] Barco, J., Wong, K. M., and Stenstrom, M. K., “Automatic Calibration of the U.S. EPA SWMM Model for a Large Urban Catchment,” *Journal of Hydraulic Engineering*, Vol. 134, No. 4, 2008, pp. 466–474.
- [85] Goulet, J.-A., Coutu, S., and Smith, I. F., “Model falsification diagnosis and sensor placement for leak detection in pressurized pipe networks,” *Advanced Engineering Informatics*, Vol. 27, No. 2, 2013, pp. 261 – 269.
- [86] Zhang, C., Wang, Y., Li, Y., and Ding, W., “Vulnerability Analysis of Urban Drainage Systems: Tree vs. Loop Networks,” *Sustainability*, Vol. 9, No. 3, 2017.
- [87] Candelieri, A., Conti, D., and Archetti, F., “A Graph based Analysis of Leak Localization in Urban Water Networks,” *Procedia Engineering*, Vol. 70, 2014, pp. 228 – 237, 12th International Conference on Computing and Control for the Water Industry, CCWI2013.
- [88] Sumer, D., Gonzalez, J., and Lansley, K., “Real-Time Detection of Sanitary Sewer Overflows Using Neural Networks and Time Series Analysis,” *Journal of Environmental Engineering*, Vol. 133, No. 4, 2007, pp. 353–363.

- [89] Ives, A. R. and Dakos, V., “Detecting dynamical changes in nonlinear time series using locally linear state-space models,” *Ecosphere*, Vol. 3, No. 6, 2012, pp. 1–15, art58.
- [90] Taeb, A., Reager, J. T., Turmon, M., and Chandrasekaran, V., “A Statistical Graphical Model of the California Reservoir System,” *Water Resources Research*, Vol. 53, No. 11, 2017, pp. 9721–9739.
- [91] Vazquez, J., Bellefleur, D., Gilbert, D., and Grandjean, B., “Real time control of a combined sewer network using graph theory,” *Water Science and Technology*, Vol. 36, No. 5, 1997, pp. 301–308.
- [92] Zamani, M., Ninness, B., and Agero, J. C., “On Identification of Networked Systems with Time-invariant Topology,” *IFAC-PapersOnLine*, Vol. 48, No. 28, 2015, pp. 1184 – 1189, 17th IFAC Symposium on System Identification SYSID 2015.
- [93] Riao-Briceo, G., Ramirez-Jaime, A., Barreiro-Gomez, J., Quijano, N., and Ocampo-Martinez, C., “Co-simulation for the design of controllers in urban drainage systems,” *2015 IEEE 2nd Colombian Conference on Automatic Control (CCAC)*, Oct 2015, pp. 1–6.
- [94] Grewal, M. and Glover, K., “Identifiability of linear and nonlinear dynamical systems,” *IEEE Transactions on Automatic Control*, Vol. 21, No. 6, Dec 1976, pp. 833–837.
- [95] Glad, S. T. and Ljung, L., “Model structure identifiability and persistence of excitation,” *29th IEEE Conference on Decision and Control*, Dec 1990, pp. 3236–3240 vol.6.
- [96] Cobelli, C., Lepschy, A., and Jacur, G., “Identifiability of compartmental systems and related structural properties,” *Mathematical Biosciences*, Vol. 44, No. 1, 1979, pp. 1 – 18.

APPENDIX A

Supporting Information for Big Ship Data: Using Vessel Measurements to Improve Estimates of Temperature and Wind Speed on the Great Lakes

This section presents figures similar to Figure 2.1 in the main article, but broken down by variable measured (Figures S1-S3).

This section also presents figures similar to Figures 2.9 and 2.10 in the main article, but for every season in the dataset (Figures S4-S39).

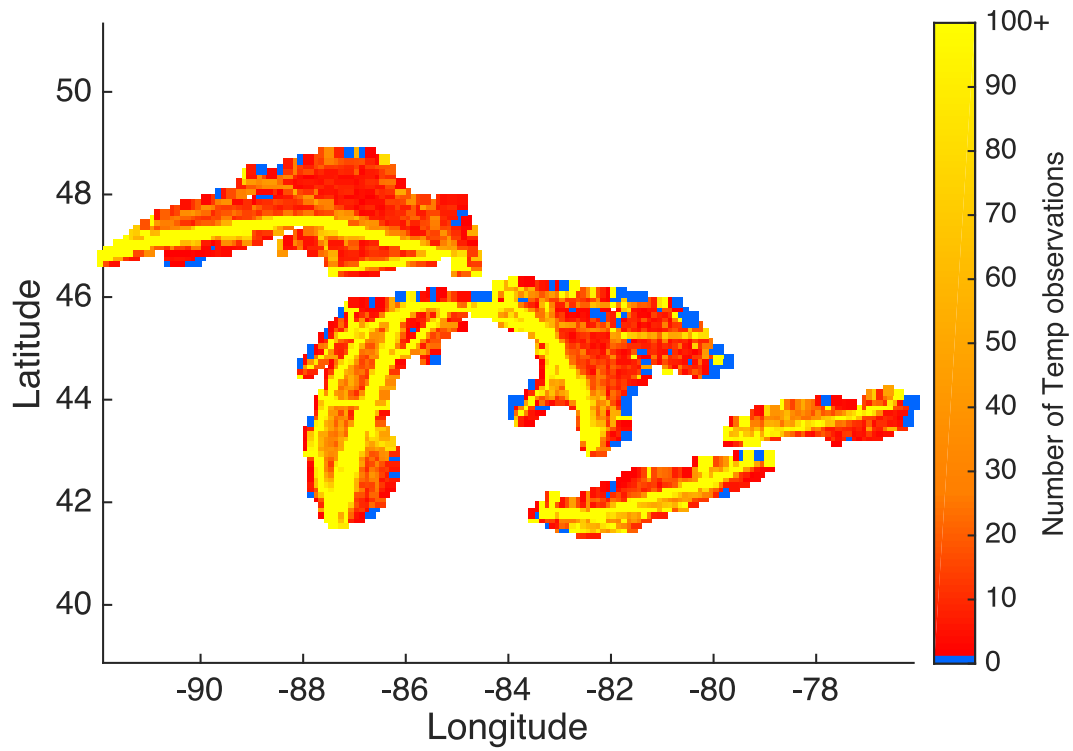


Figure A.1: Comparison of ship measurement densities to the stationary observations of temperature. Red color indicates few measurements, yellow corresponds with locations which contain over 100 measurements, while blue indicates no available ship measurements.

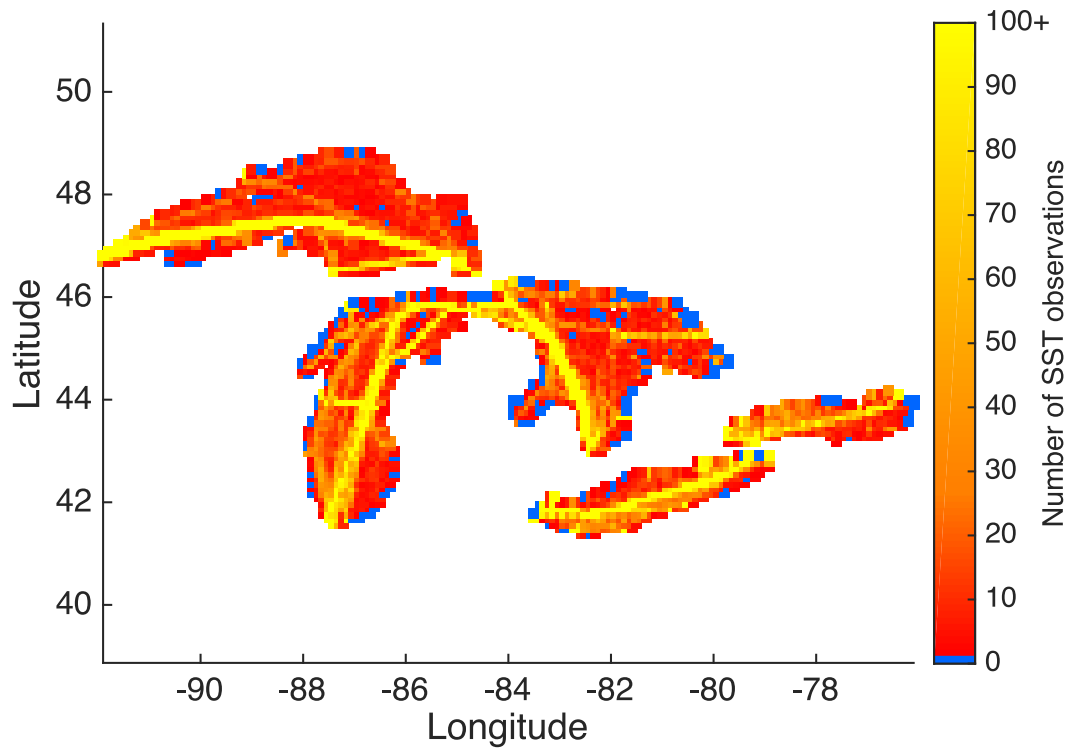


Figure A.2: Comparison of ship measurement densities to the stationary observations of surface temperature. Red color indicates few measurements, yellow corresponds with locations which contain over 100 measurements, while blue indicates no available ship measurements.

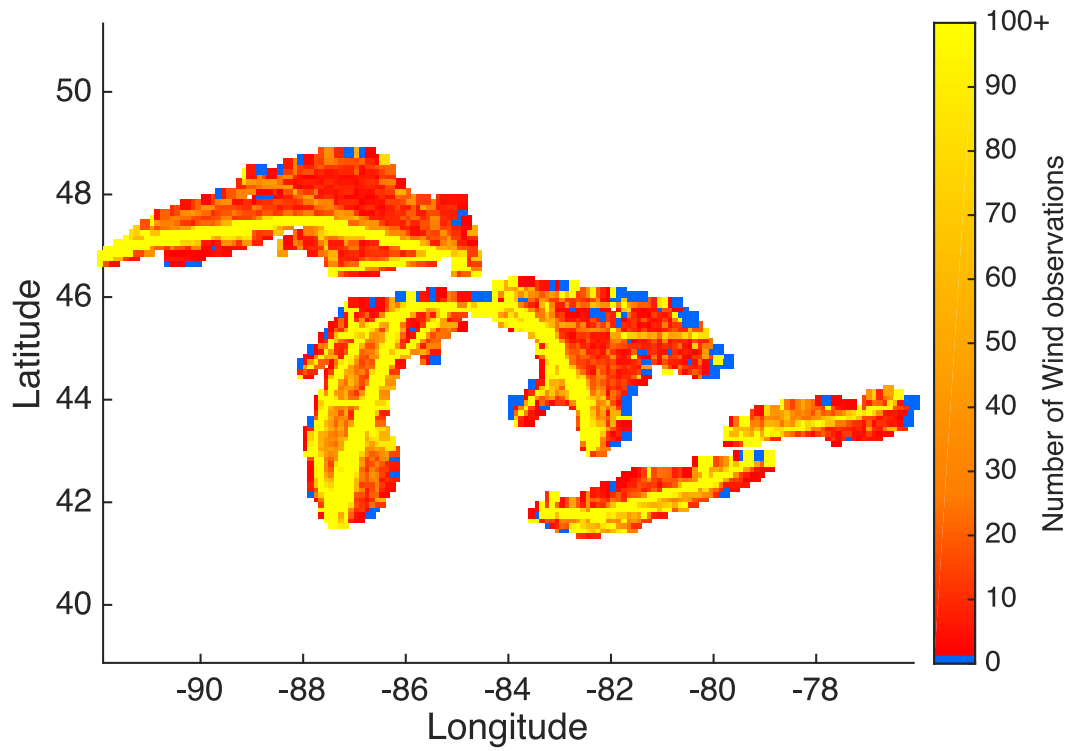


Figure A.3: Comparison of ship measurement densities to the stationary observations of wind speed. Red color indicates few measurements, yellow corresponds with locations which contain over 100 measurements, while blue indicates no available ship measurements.

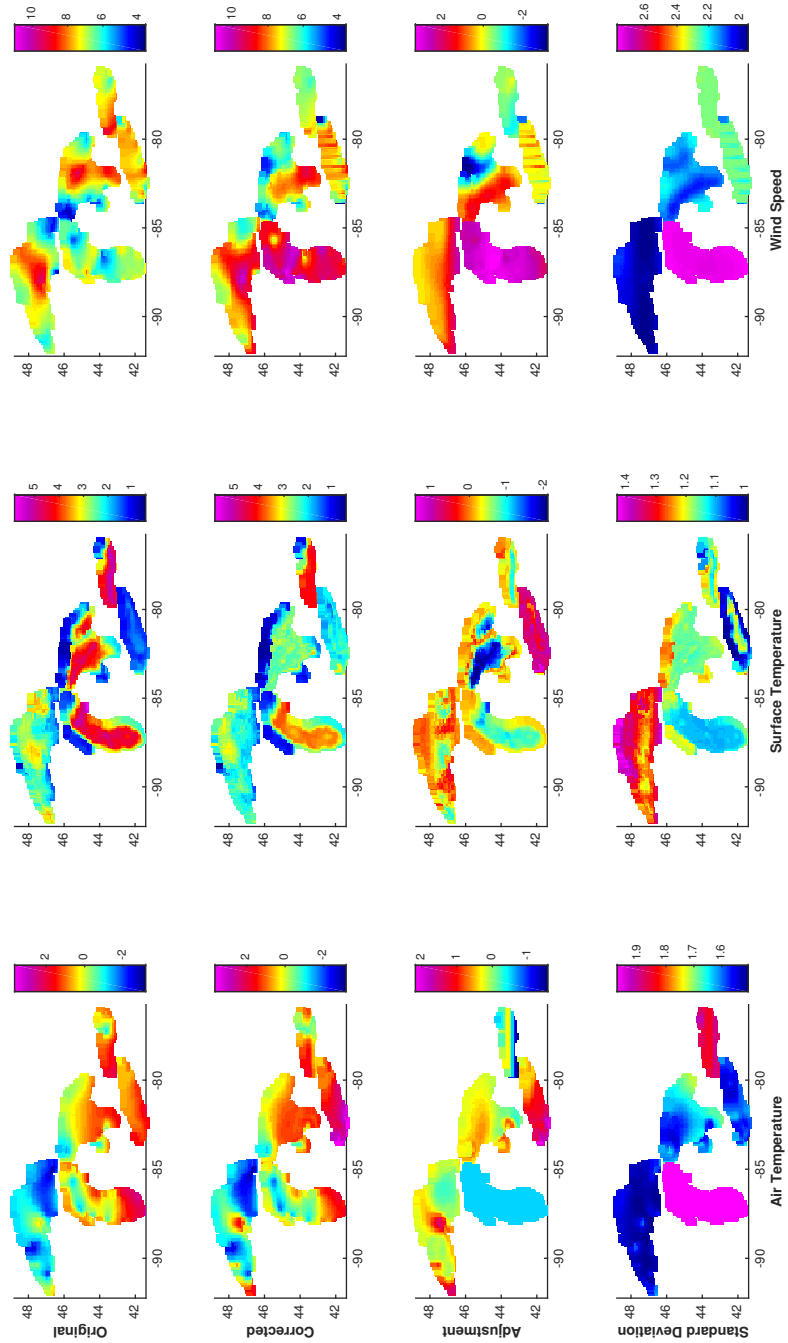


Figure A.4: Winter 2006

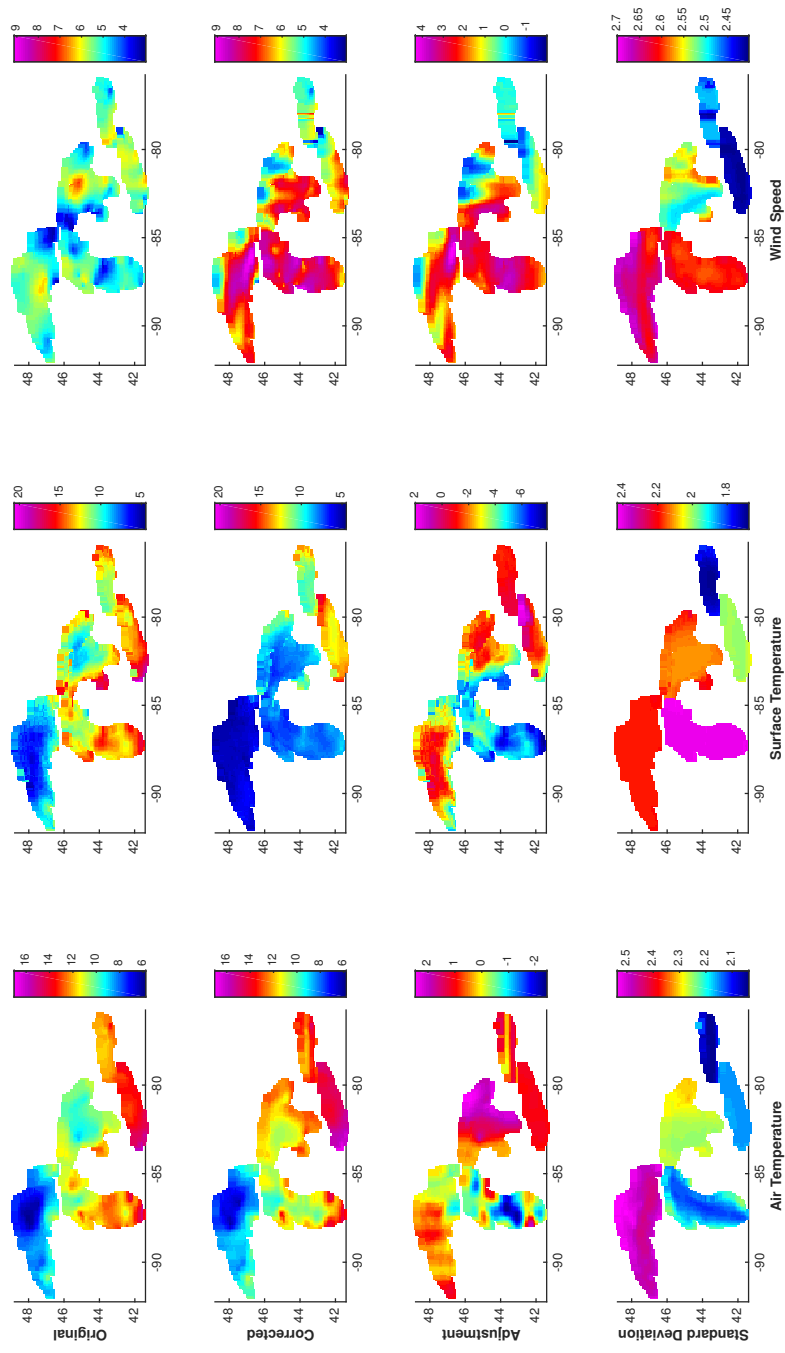


Figure A.5: Spring 2006

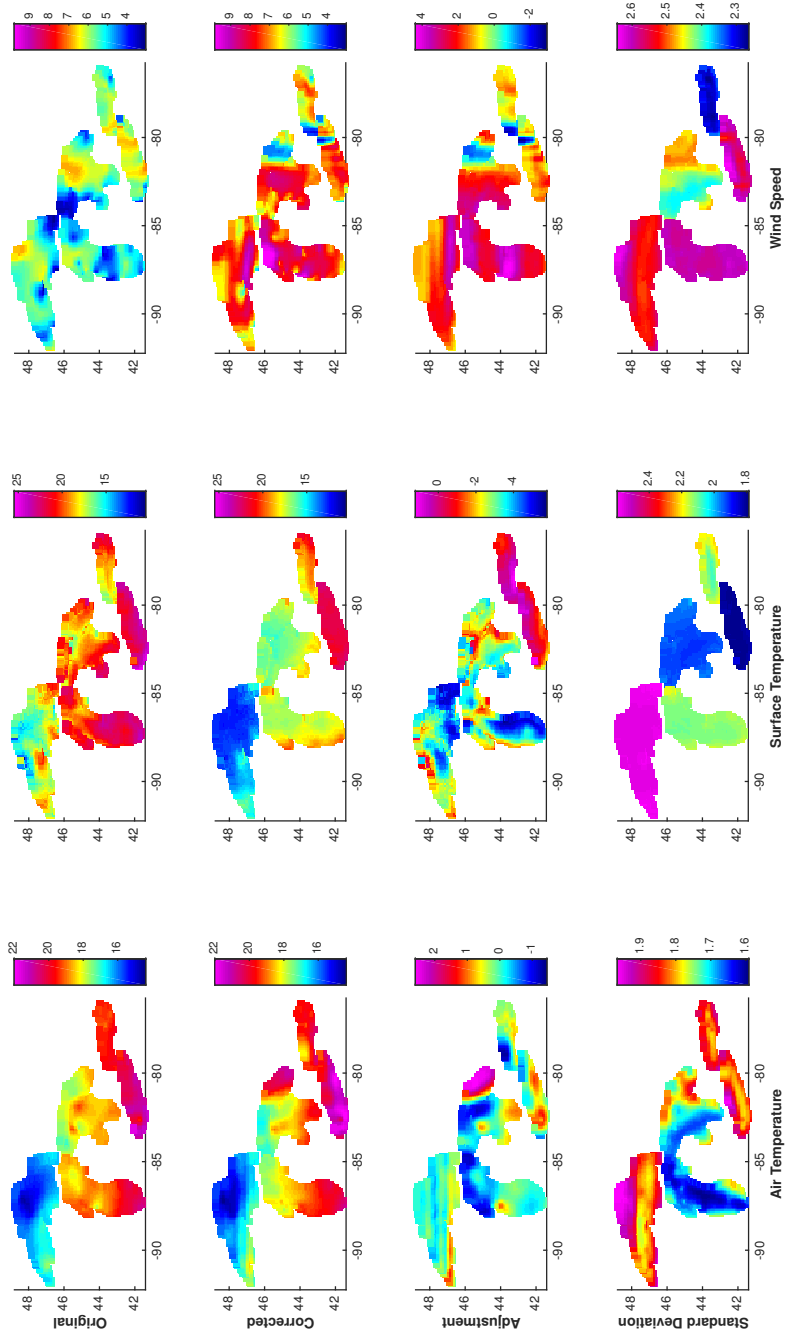


Figure A.6: Summer 2006

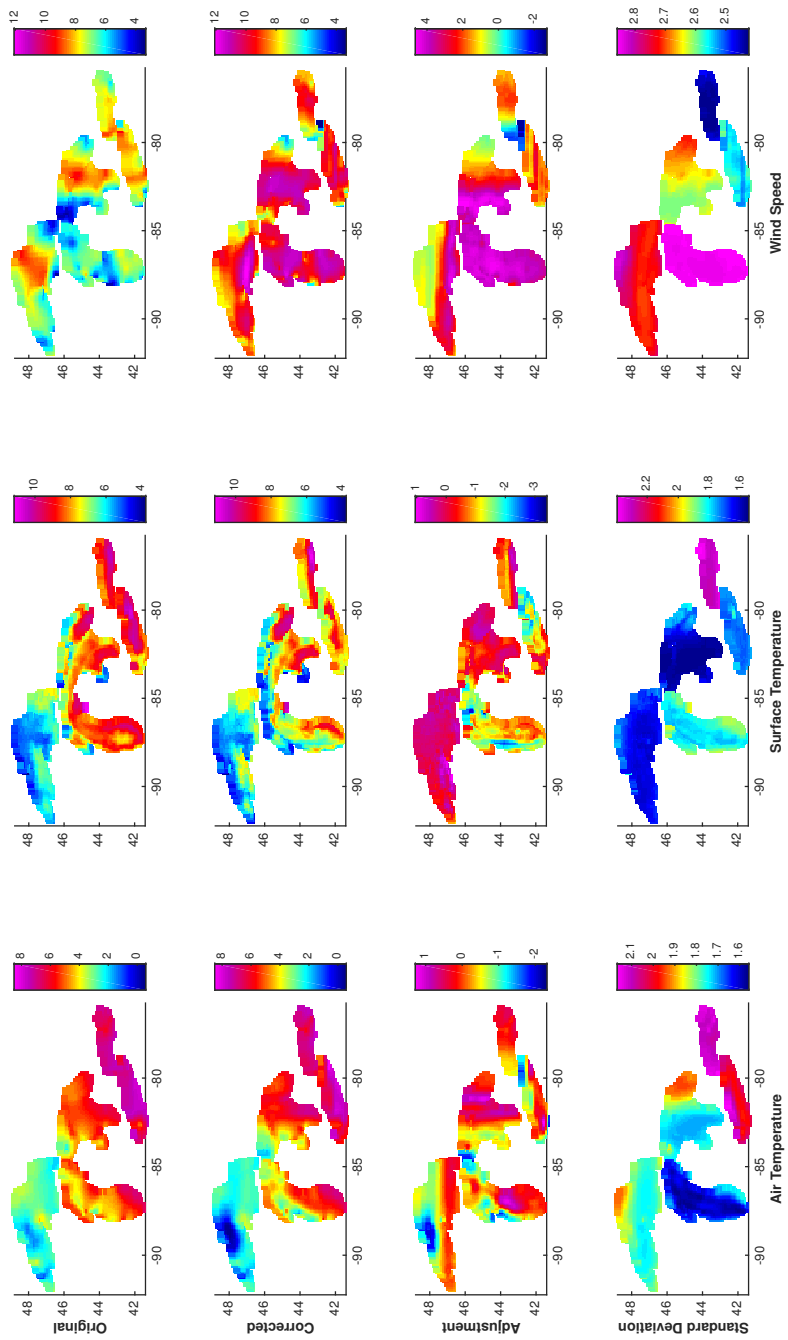


Figure A.7: Fall 2006

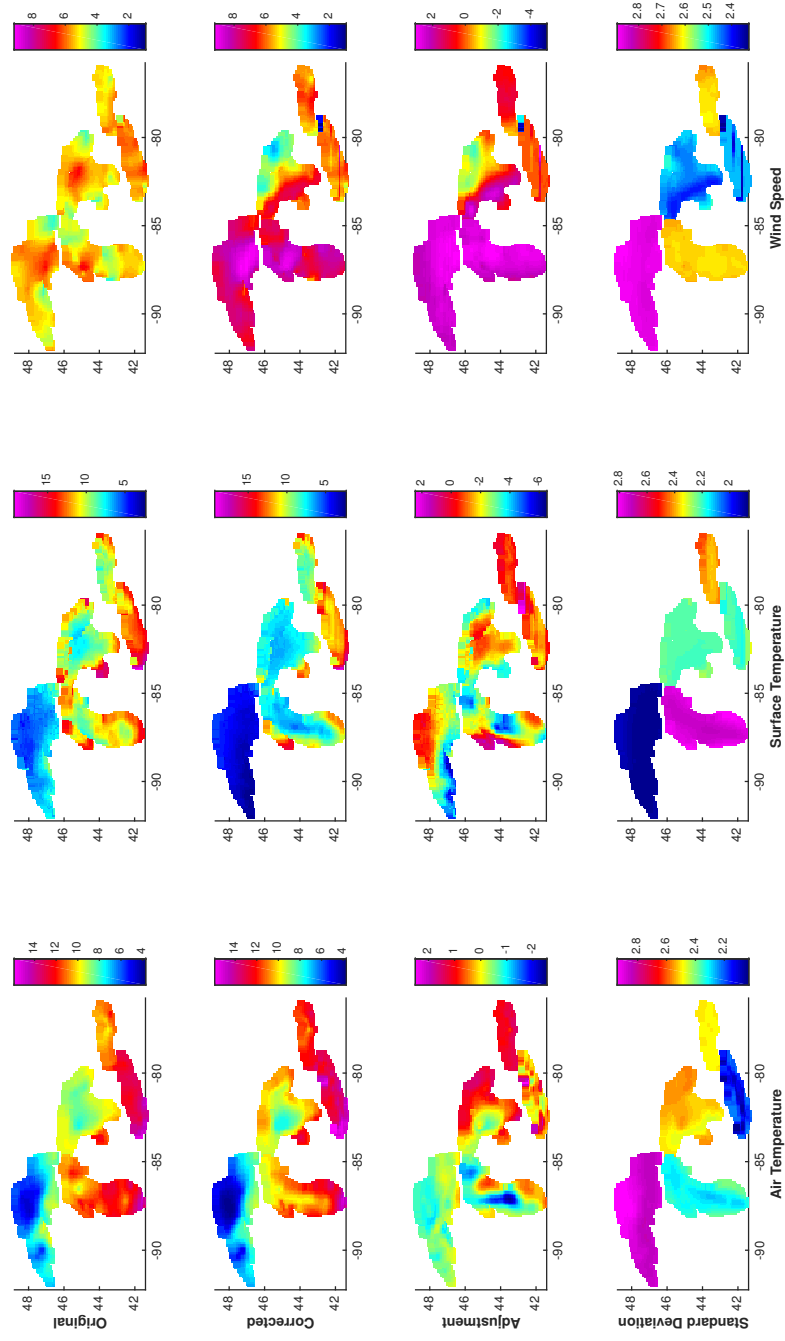


Figure A.8: Spring 2007

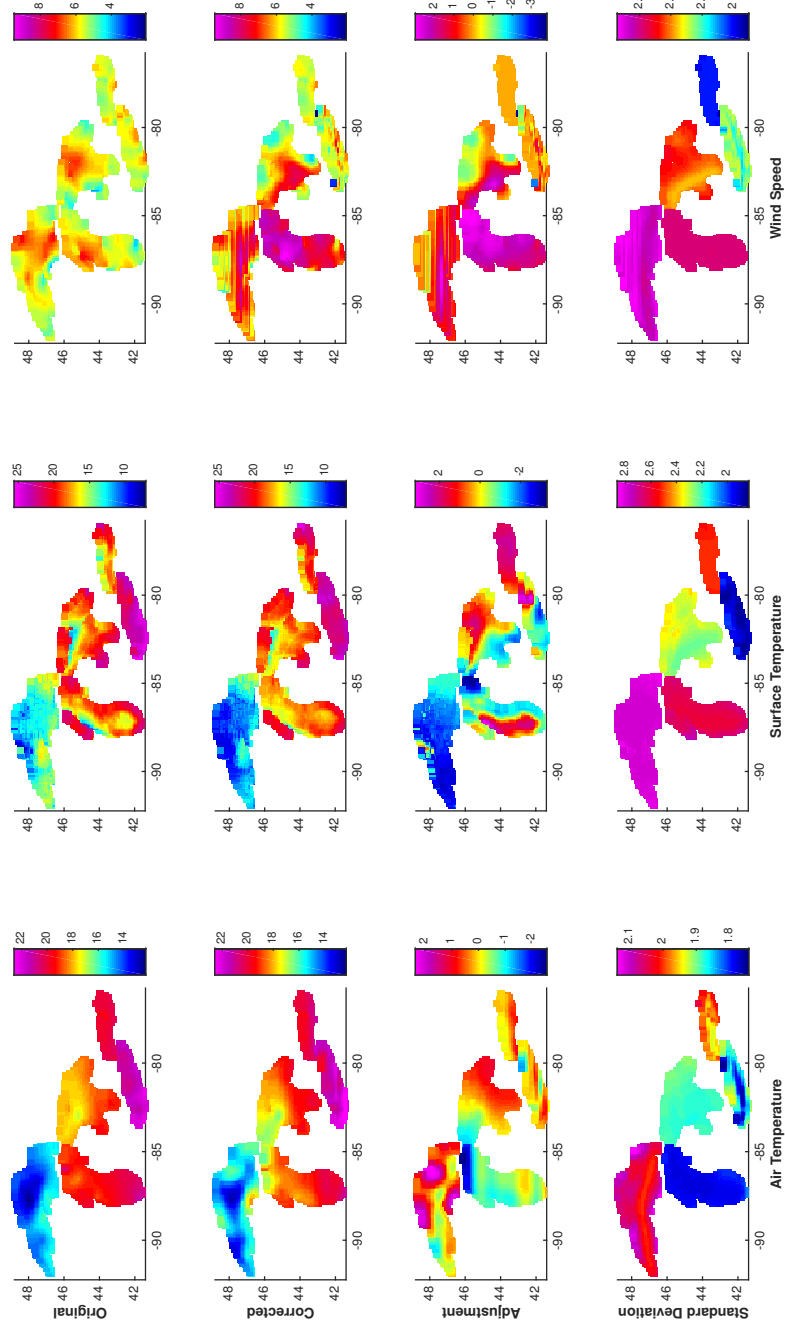


Figure A.9: Summer 2007

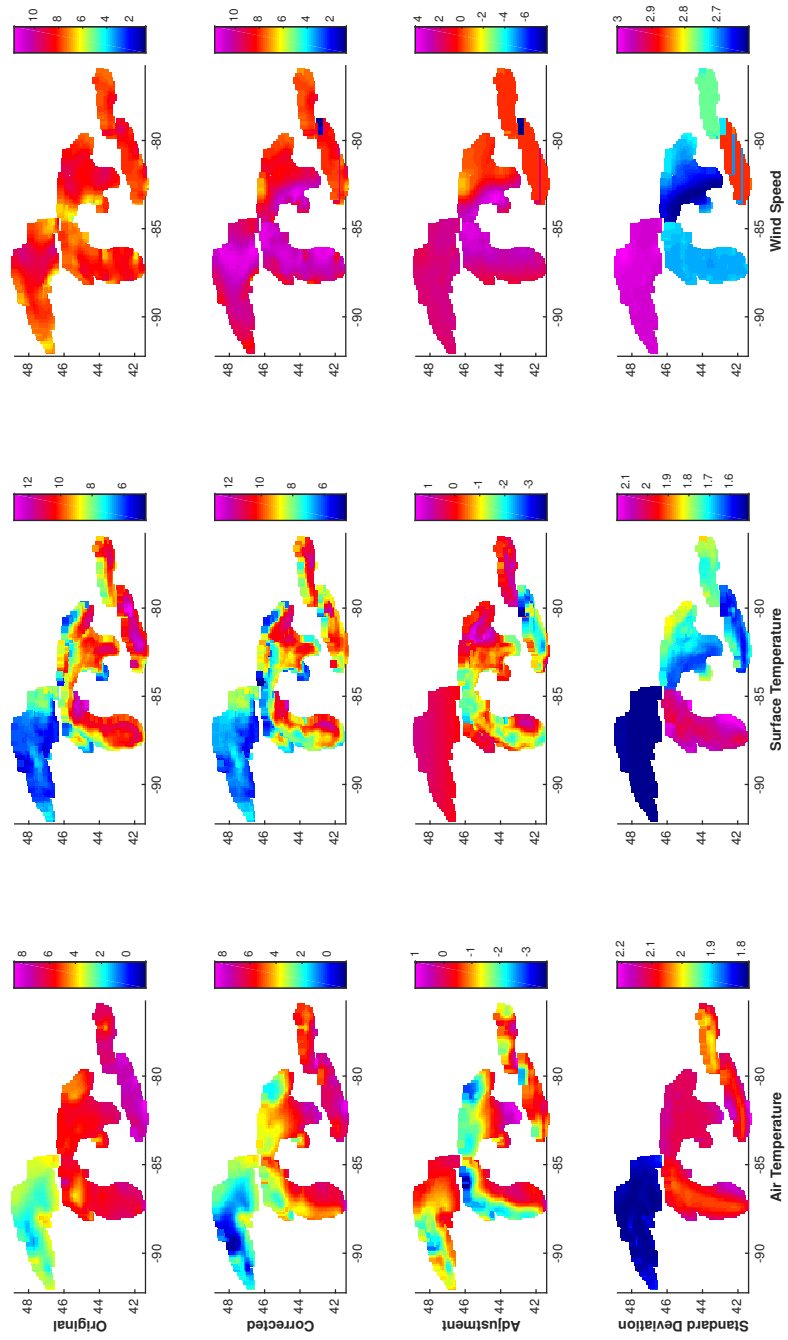


Figure A.10: Fall 2007

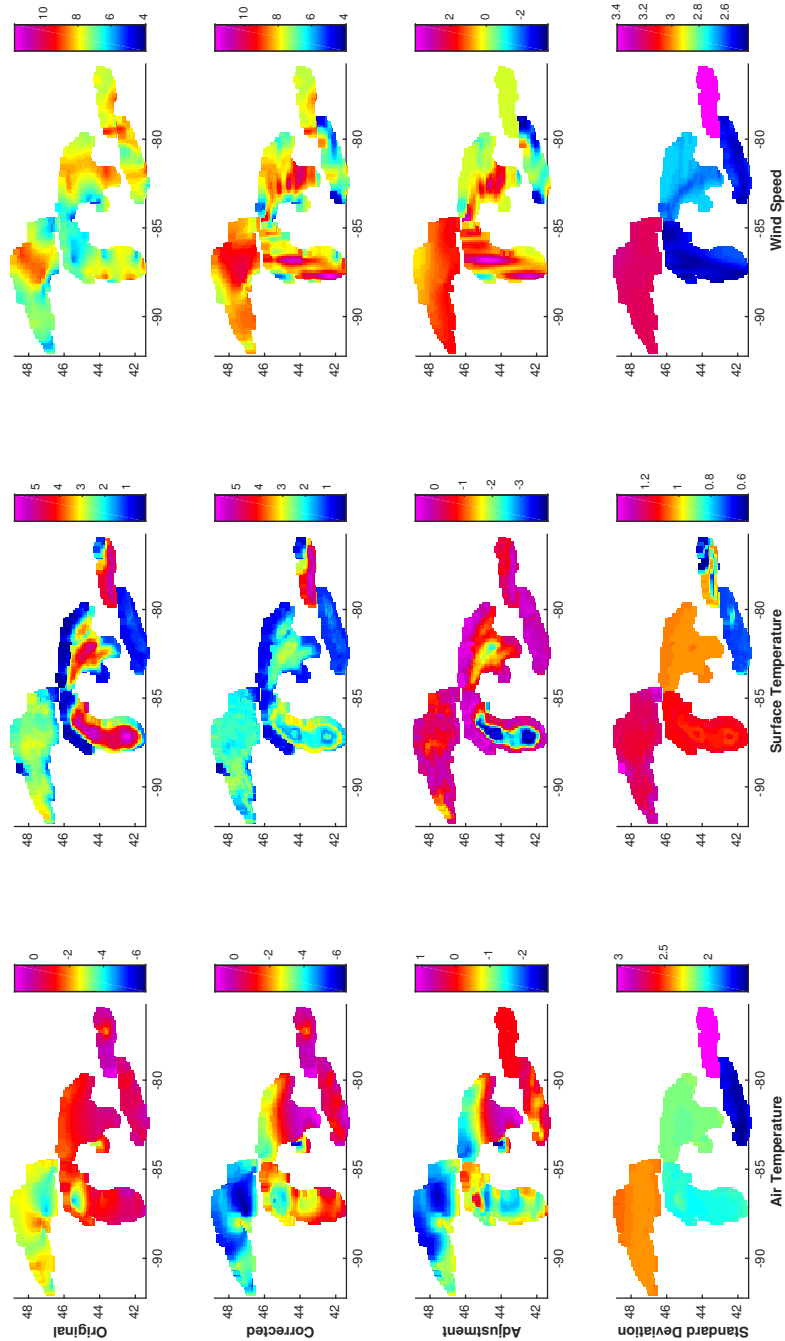


Figure A.11: Winter 2008

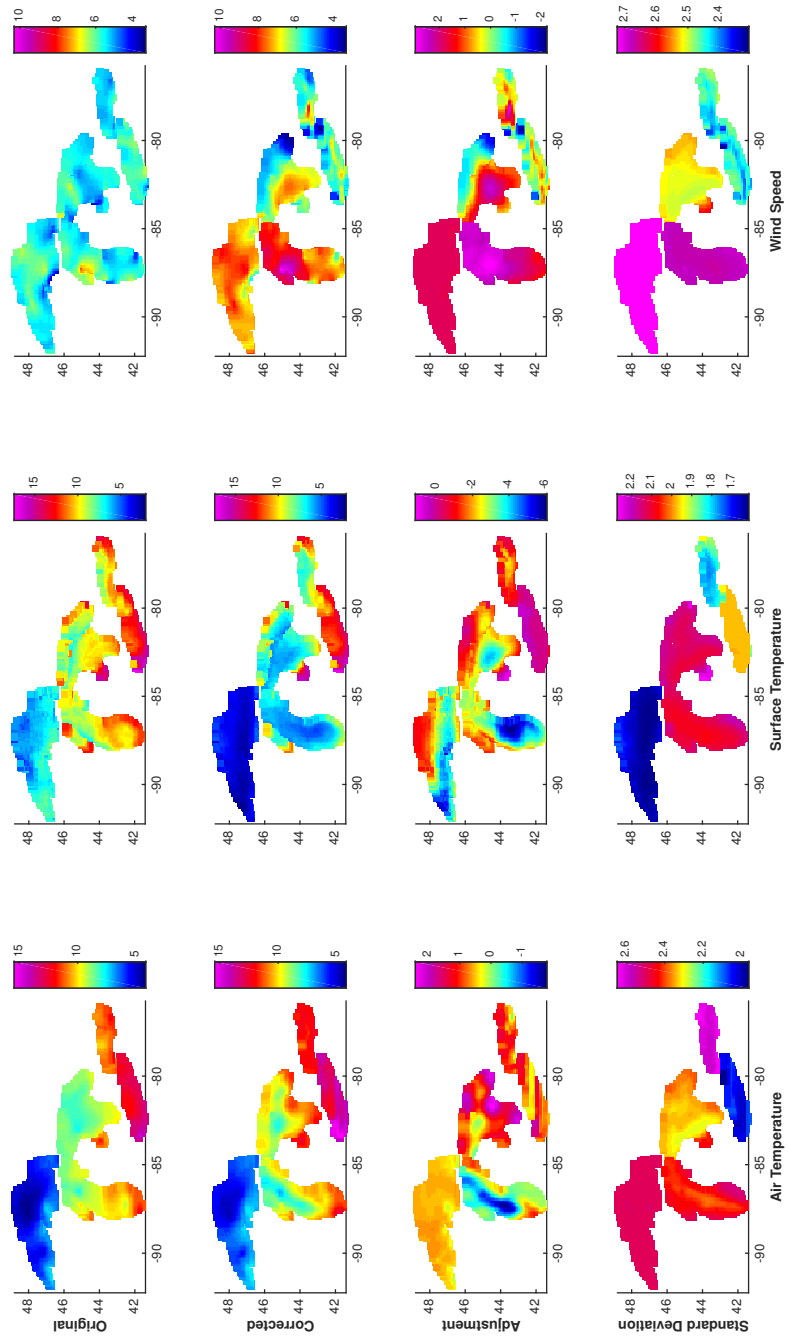


Figure A.12: Spring 2008

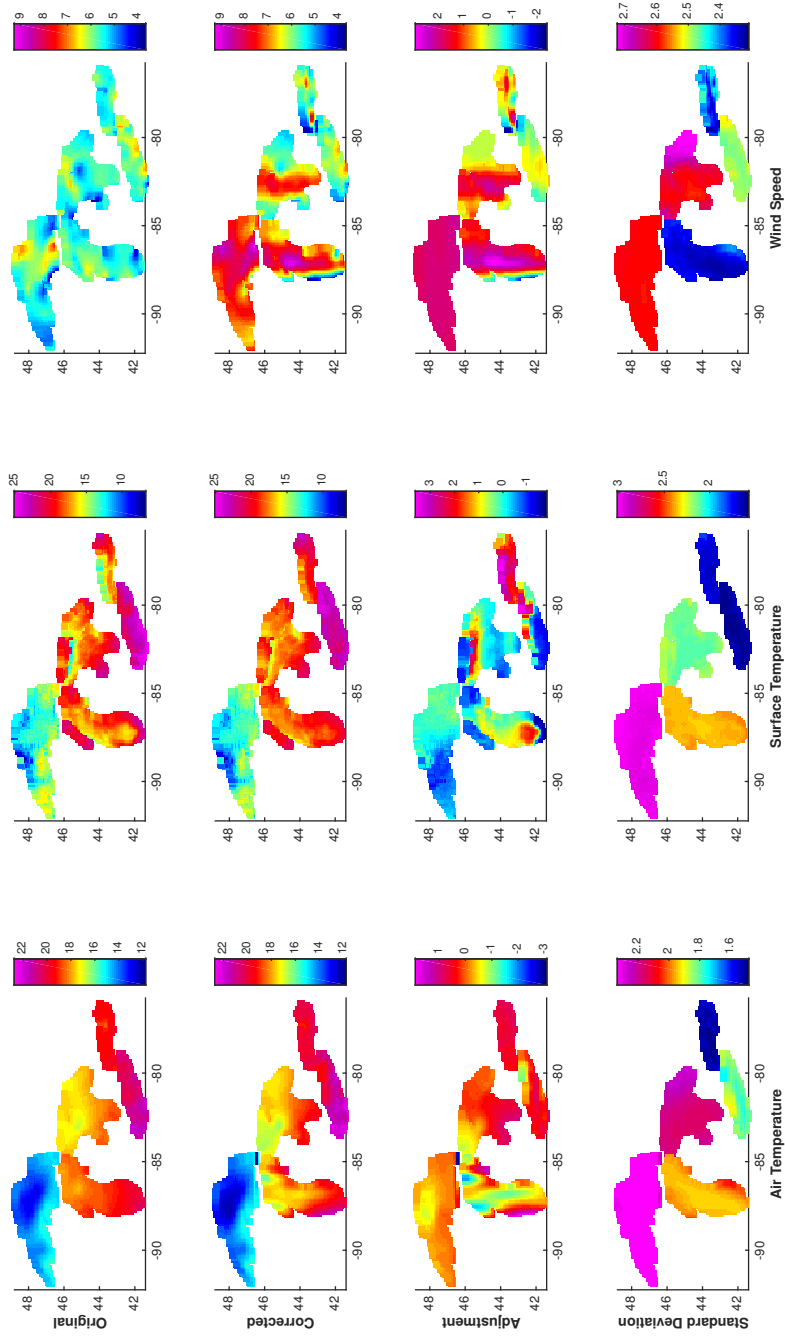


Figure A.13: Summer 2008

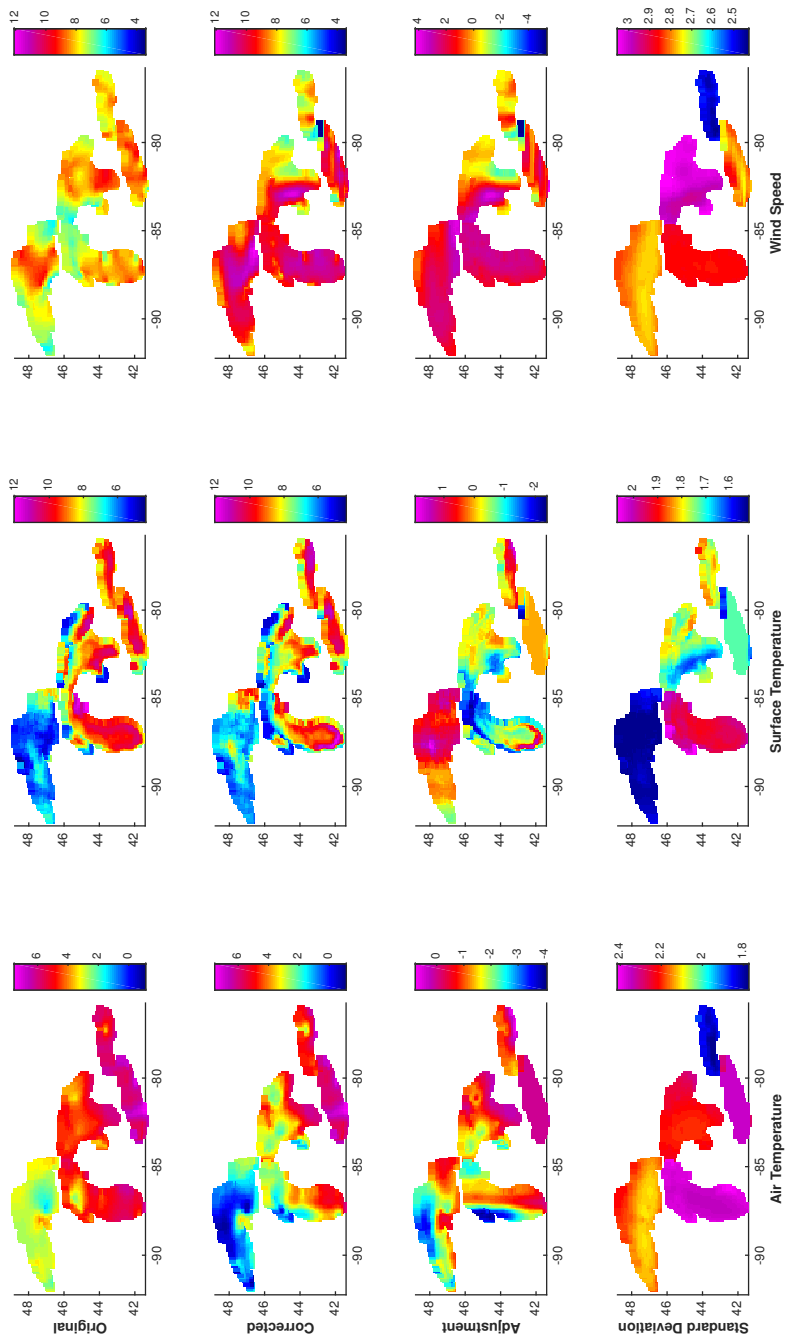


Figure A.14: Fall 2008

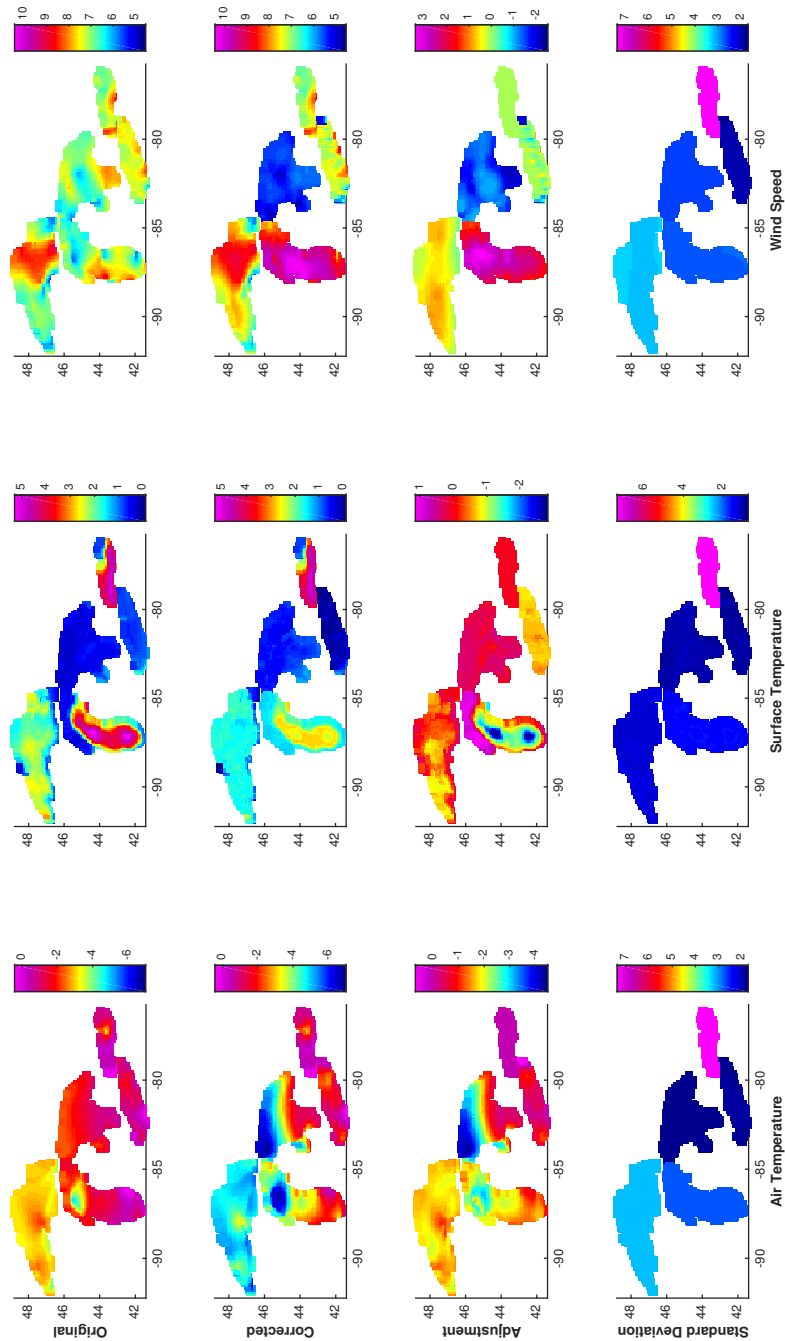


Figure A.15: Winter 2009

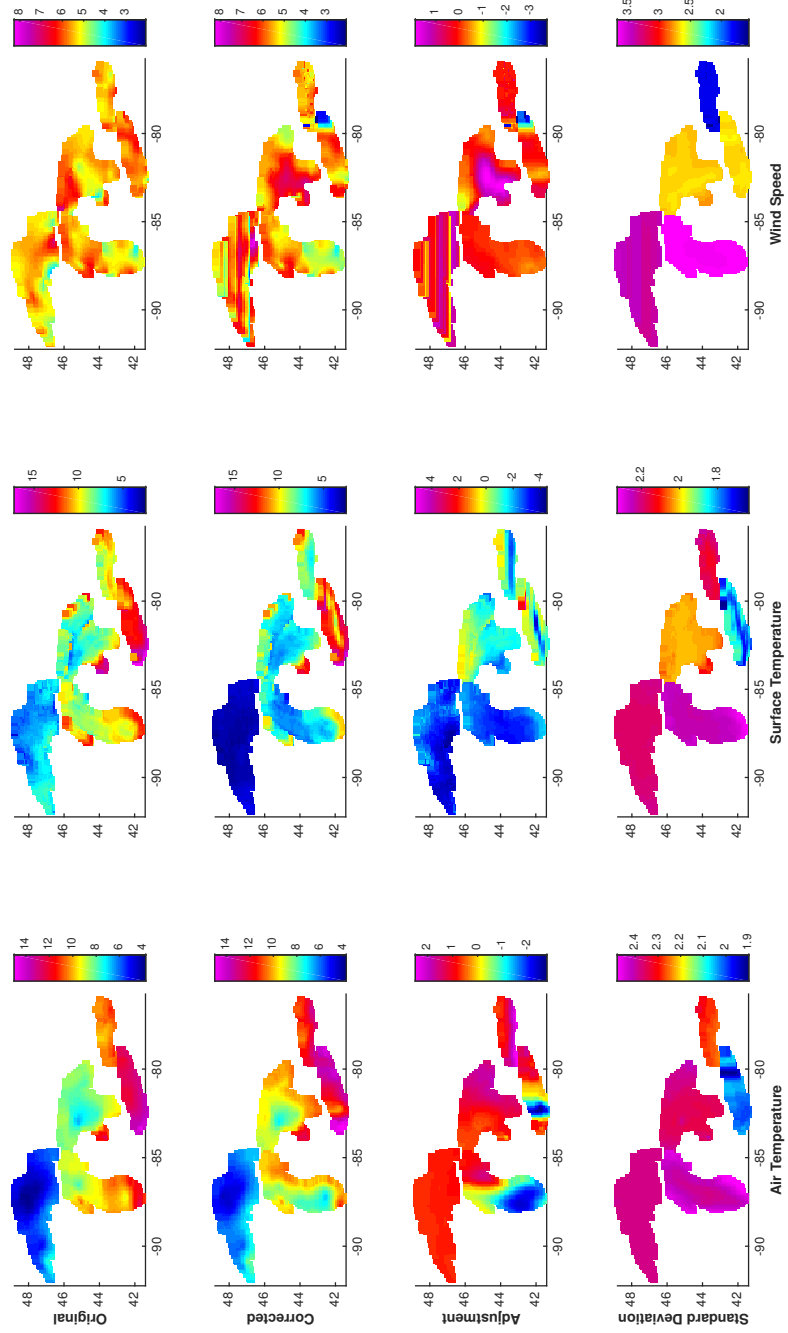


Figure A.16: Spring 2009

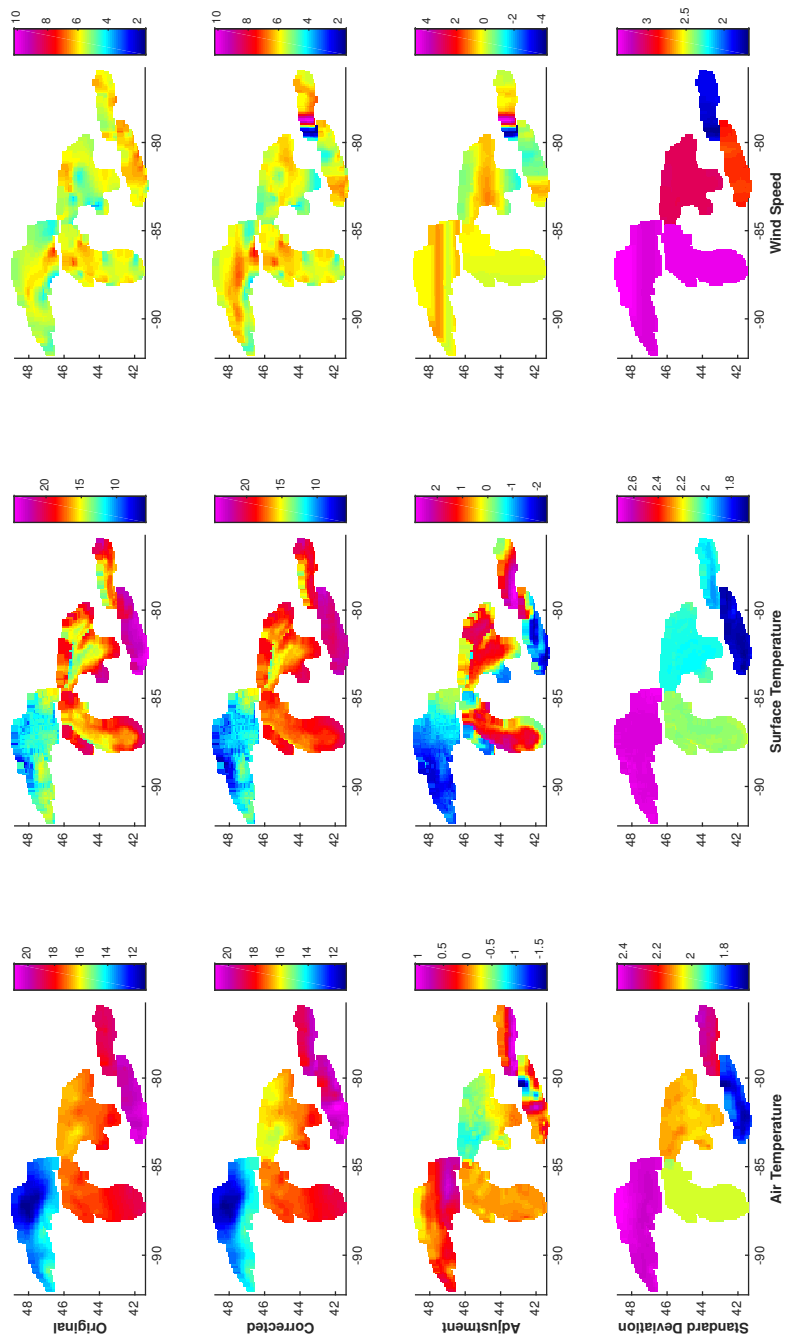


Figure A.17: Summer 2009

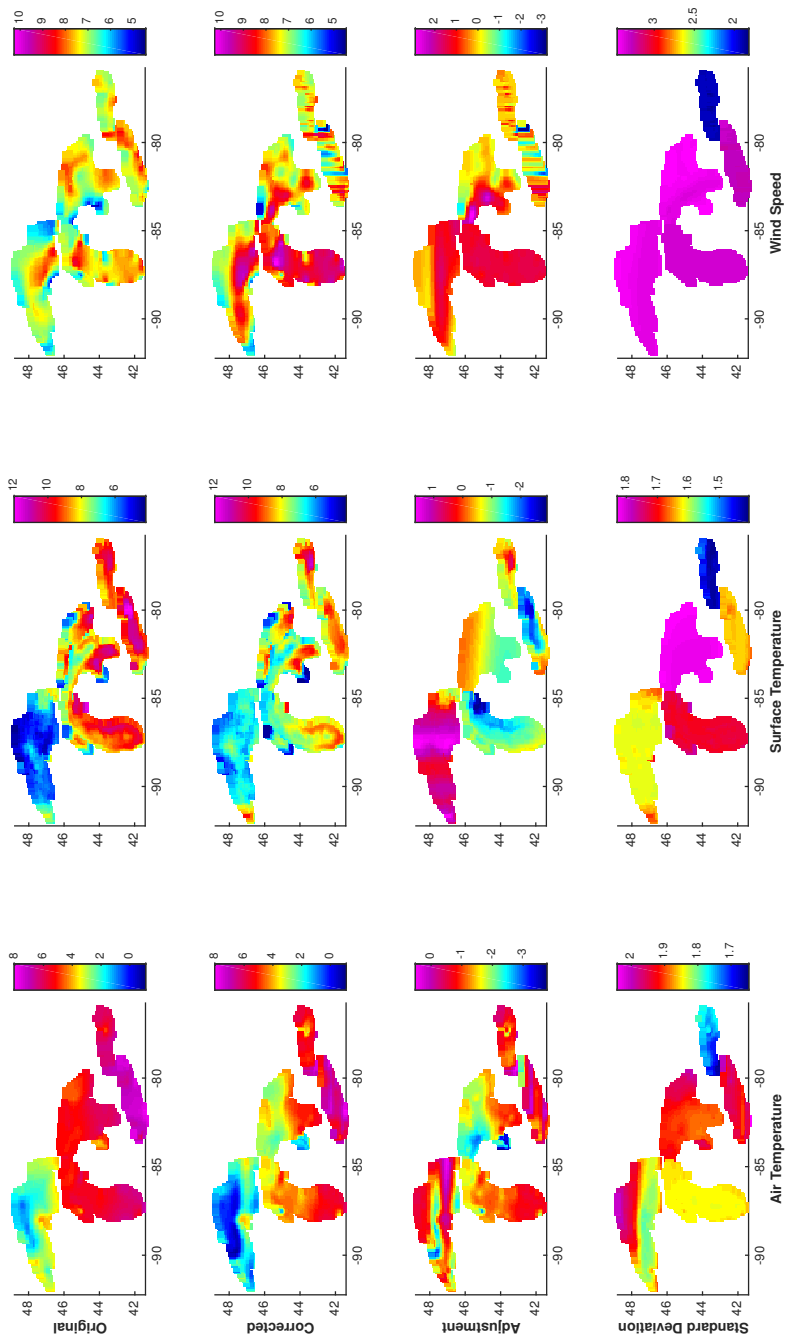


Figure A.18: Fall 2009

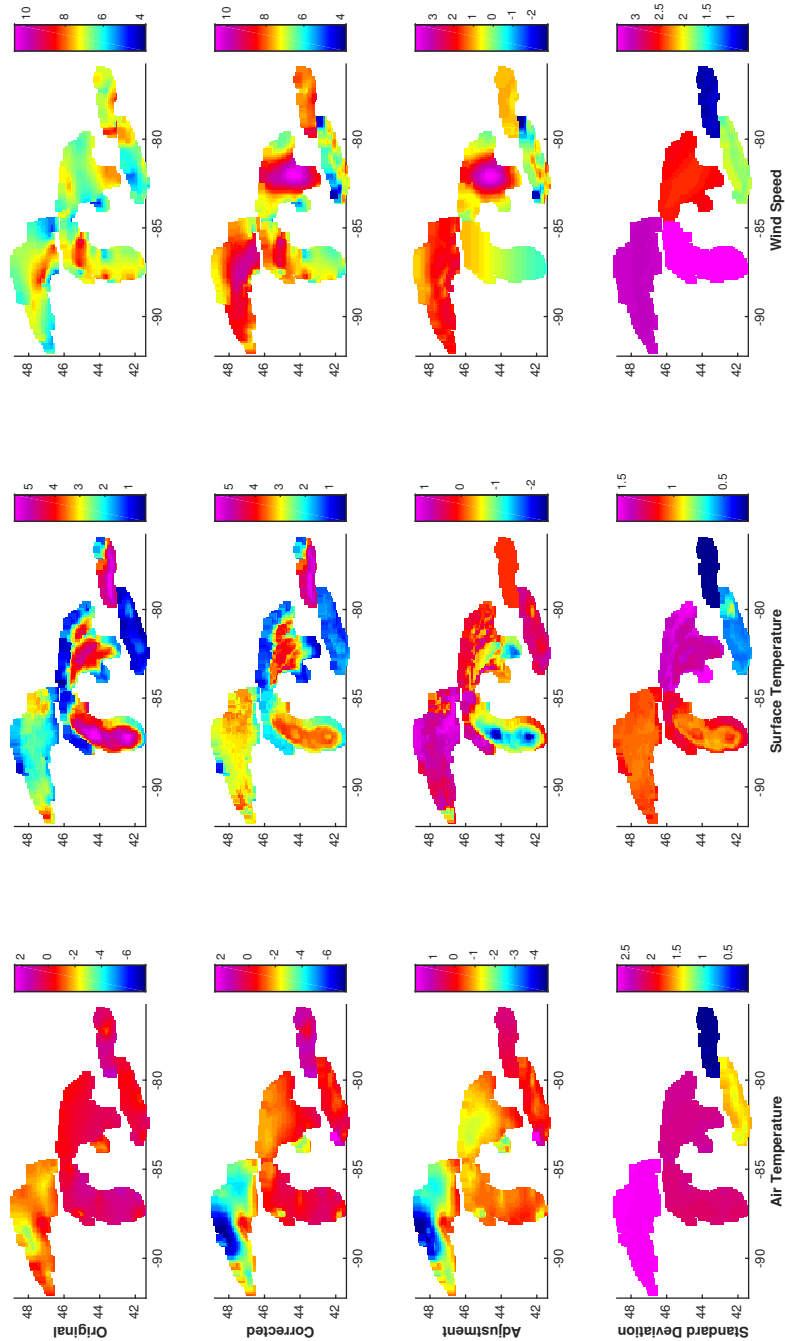


Figure A.19: Winter 2010

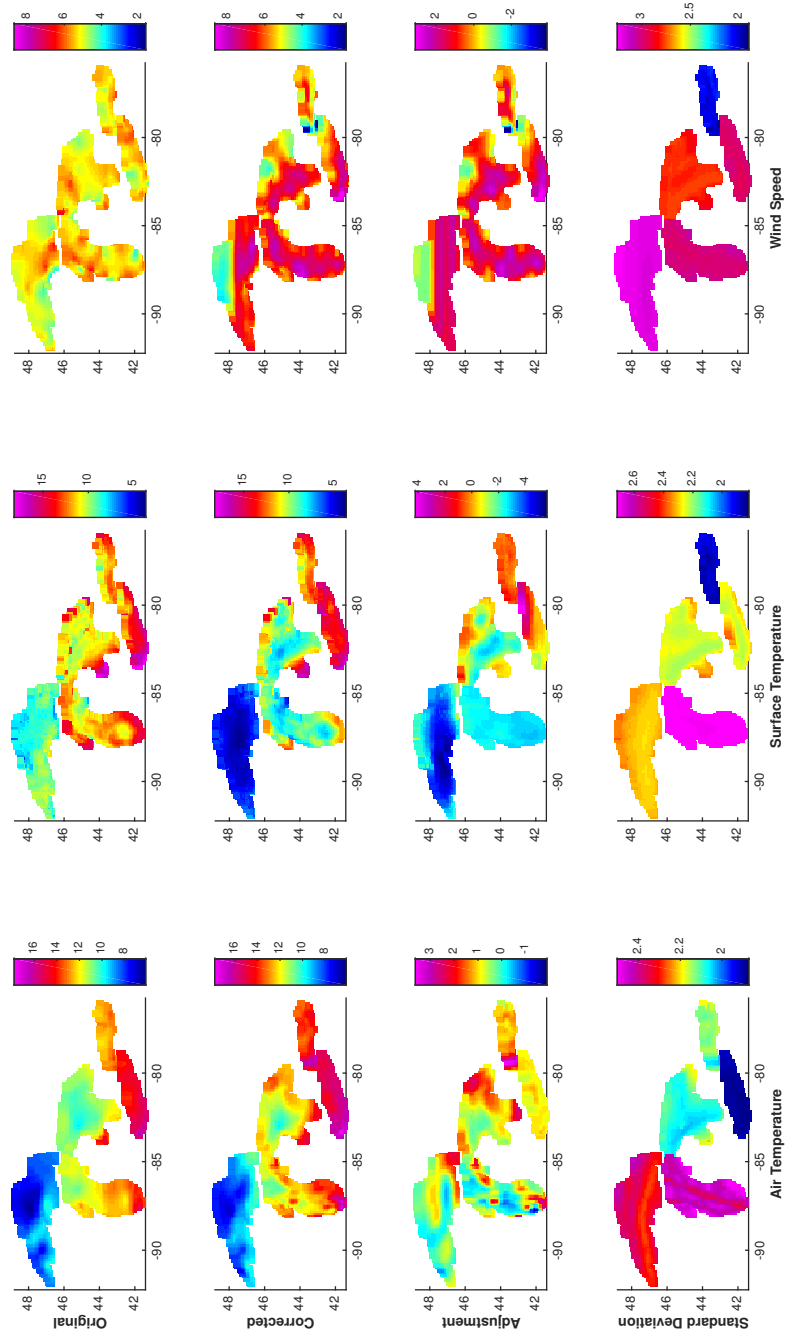


Figure A.20: Spring 2010

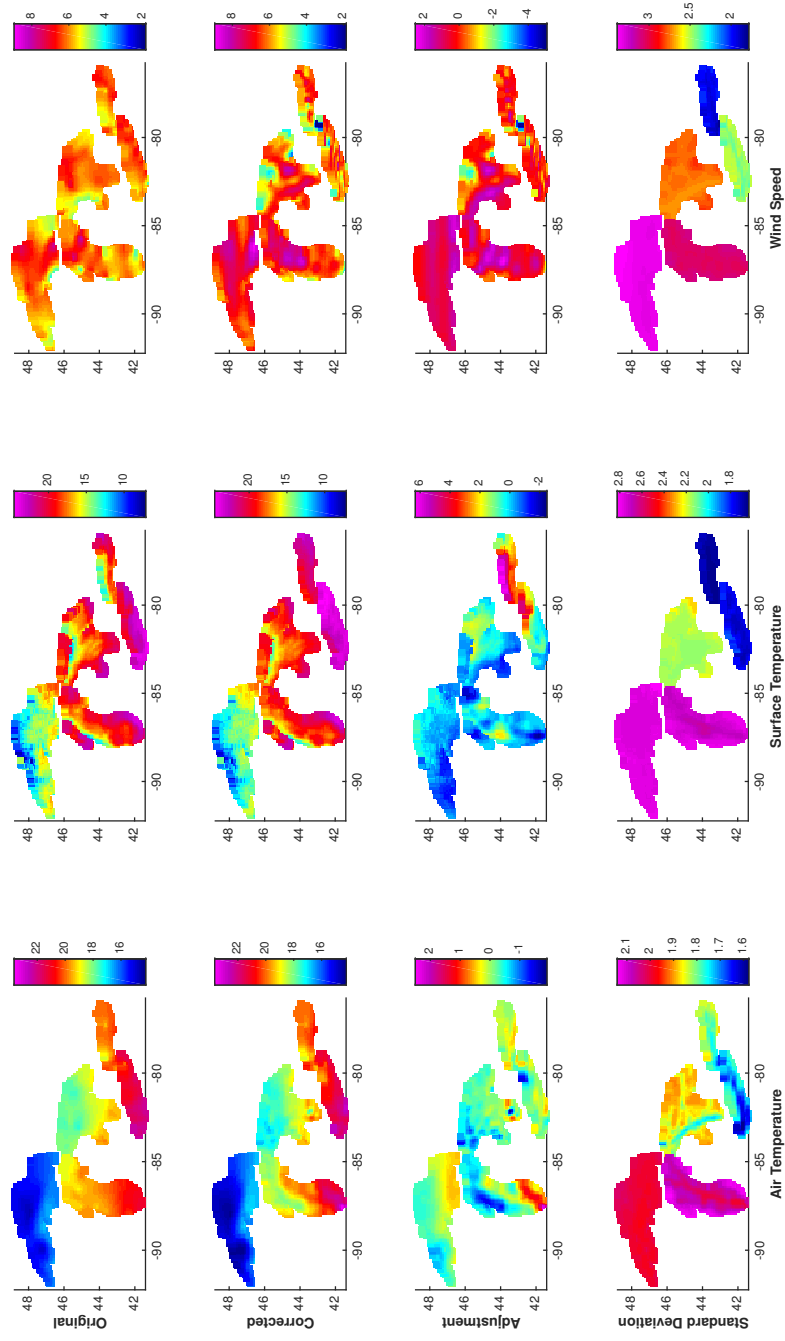


Figure A.21: Summer 2010

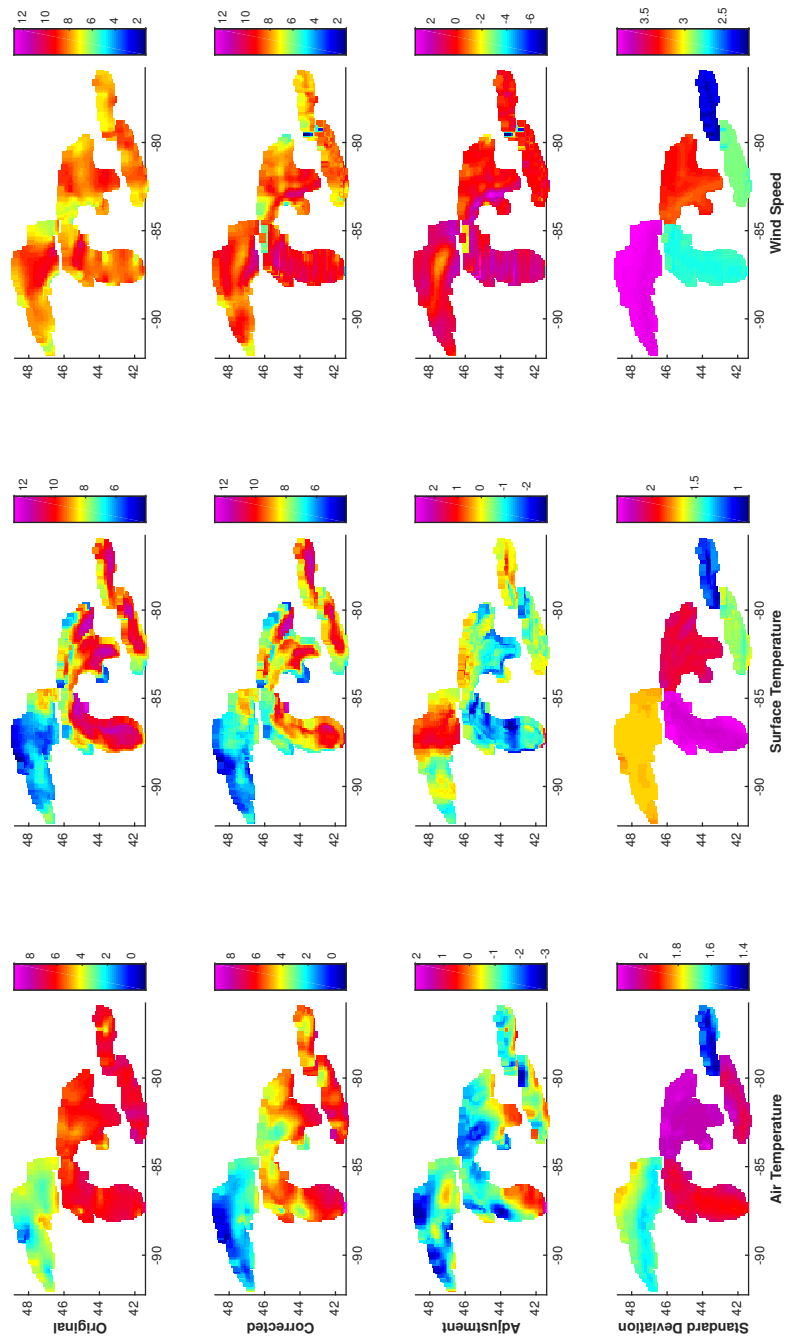


Figure A.22: Fall 2010

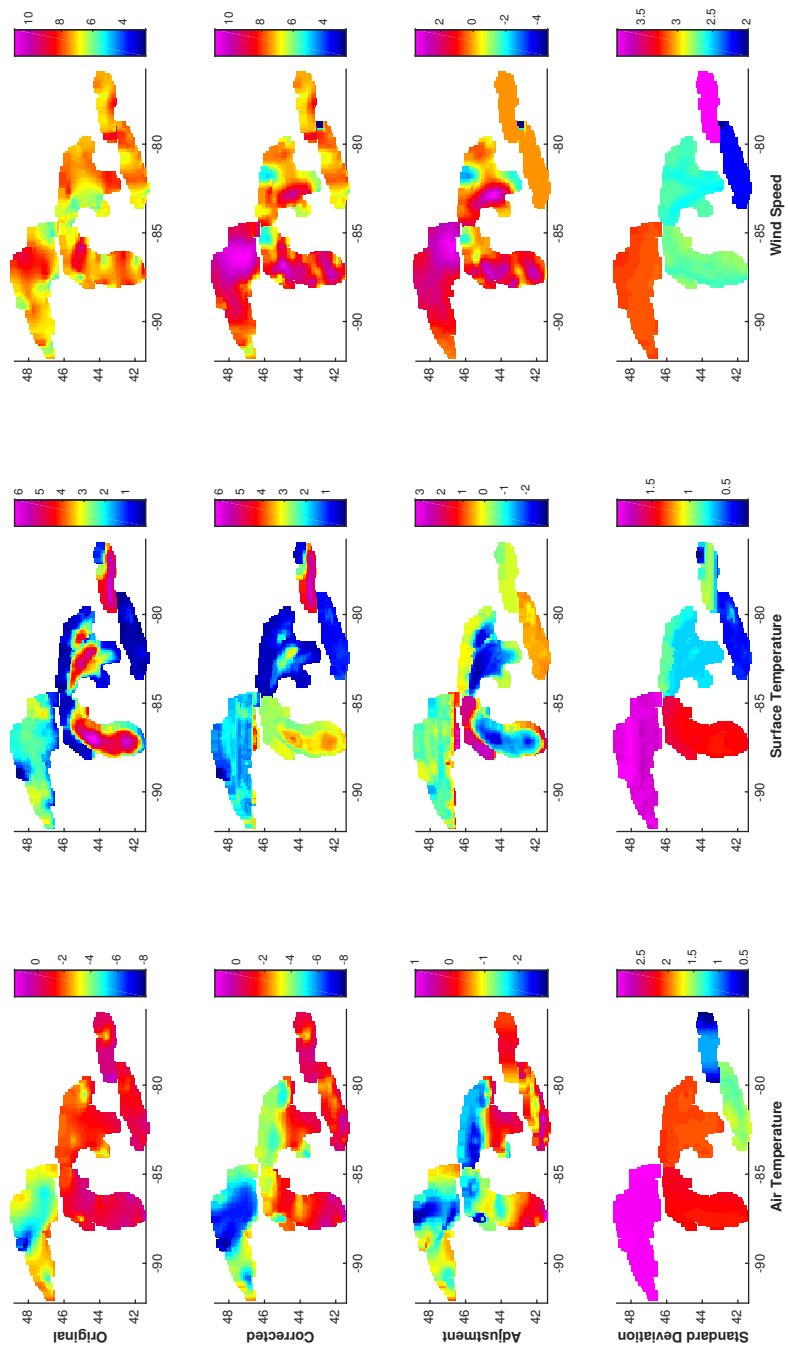


Figure A.23: Winter 2011

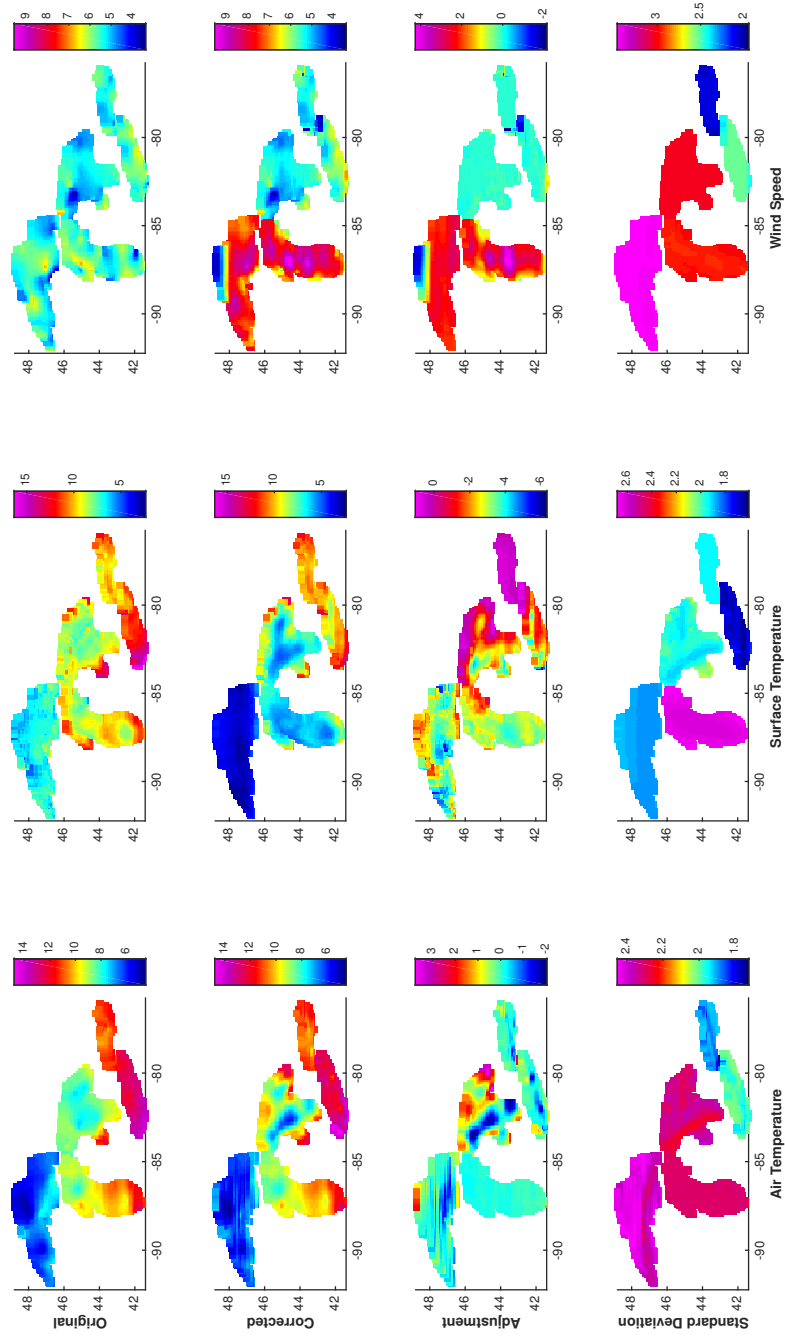


Figure A.24: Spring 2011

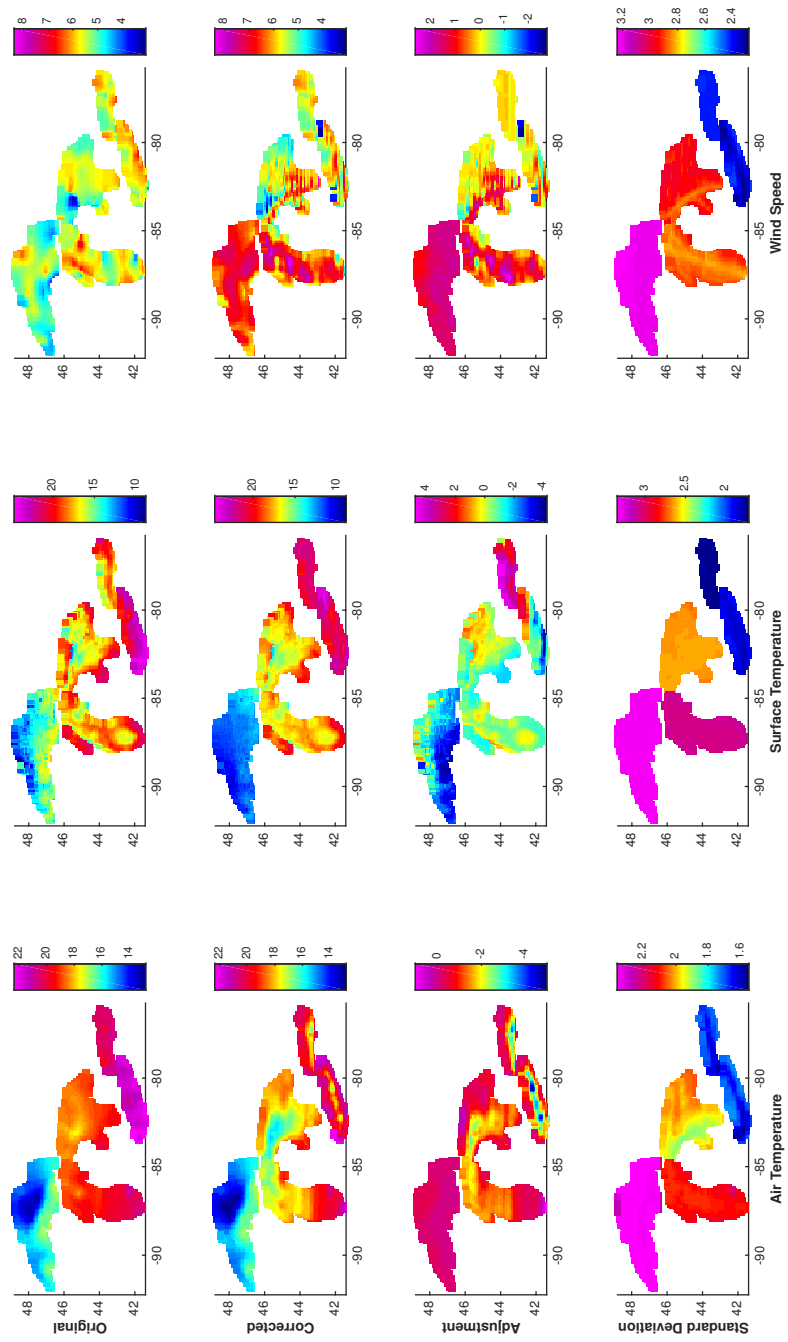


Figure A.25: Summer 2011

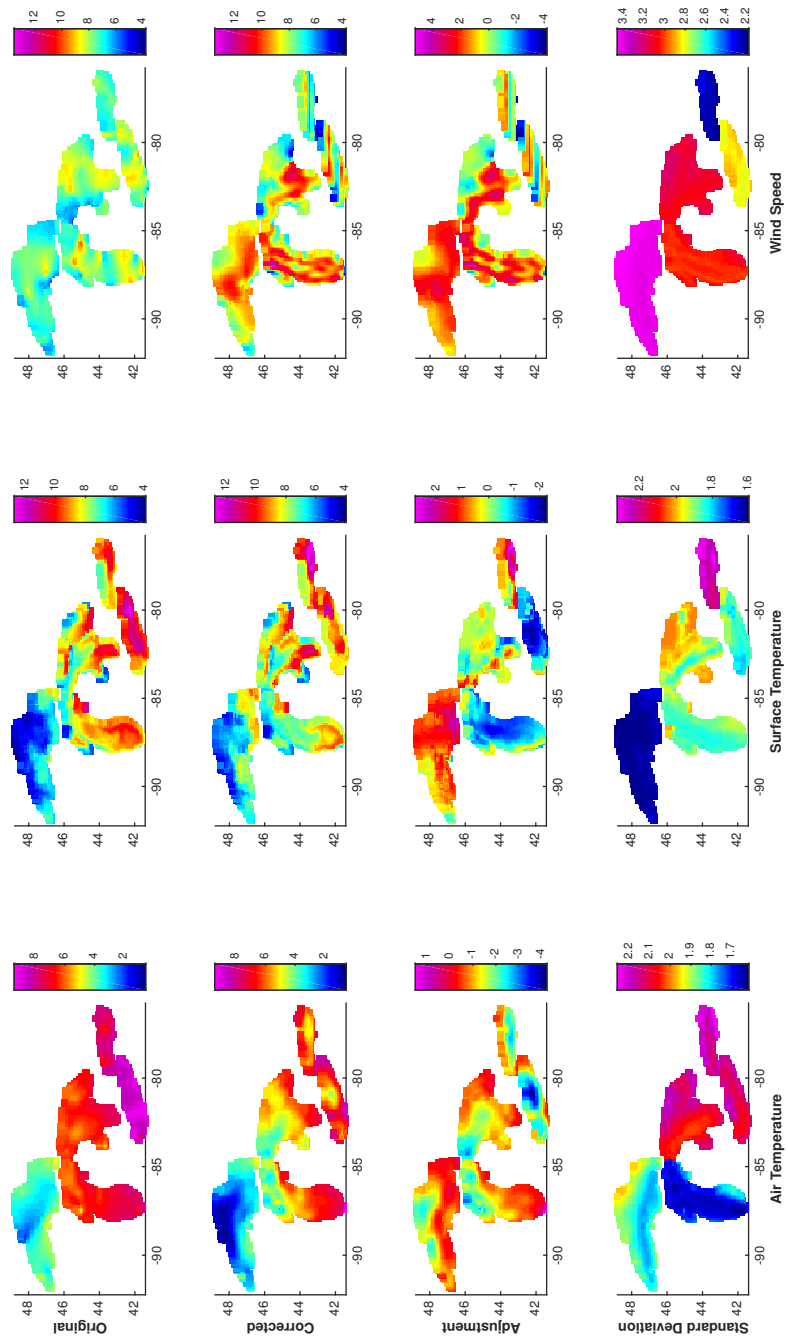


Figure A.26: Fall 2011

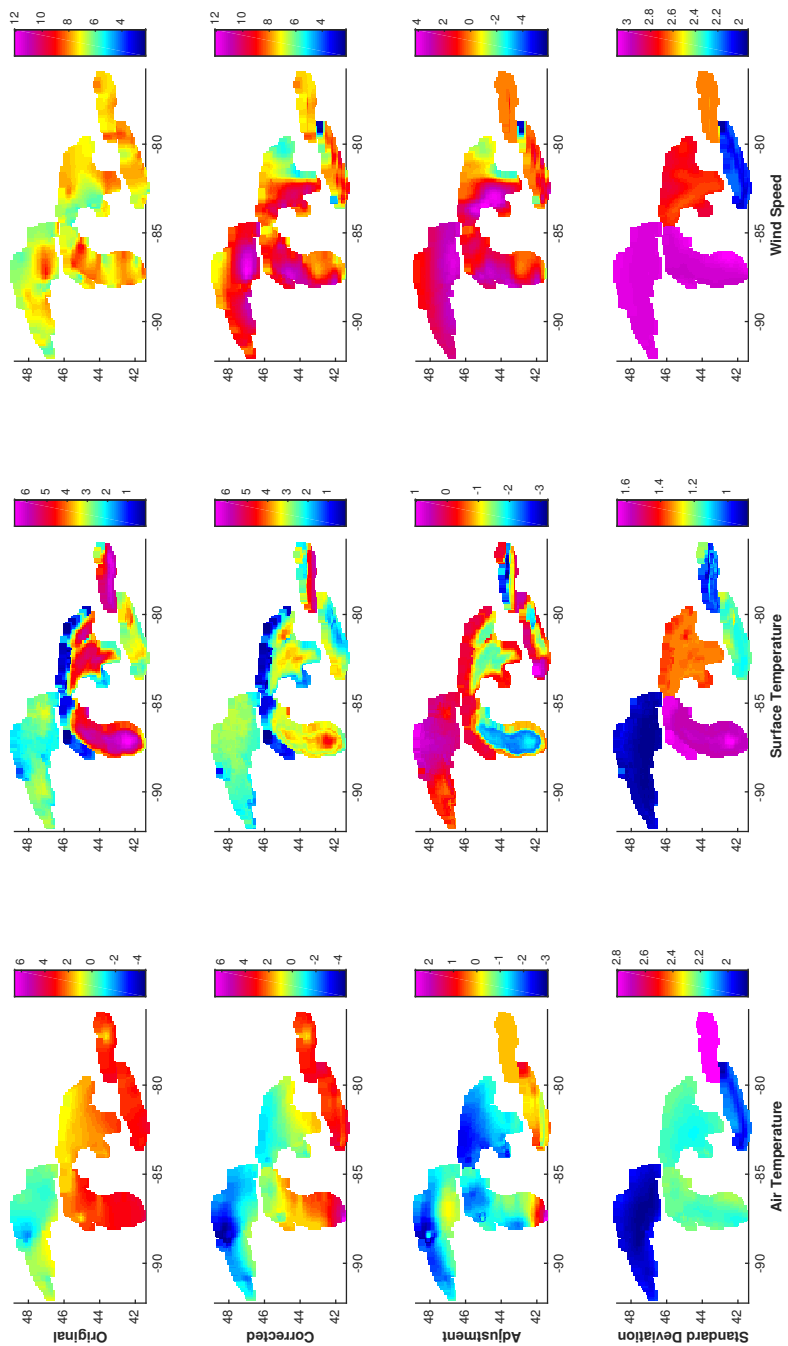


Figure A.27: Winter 2012

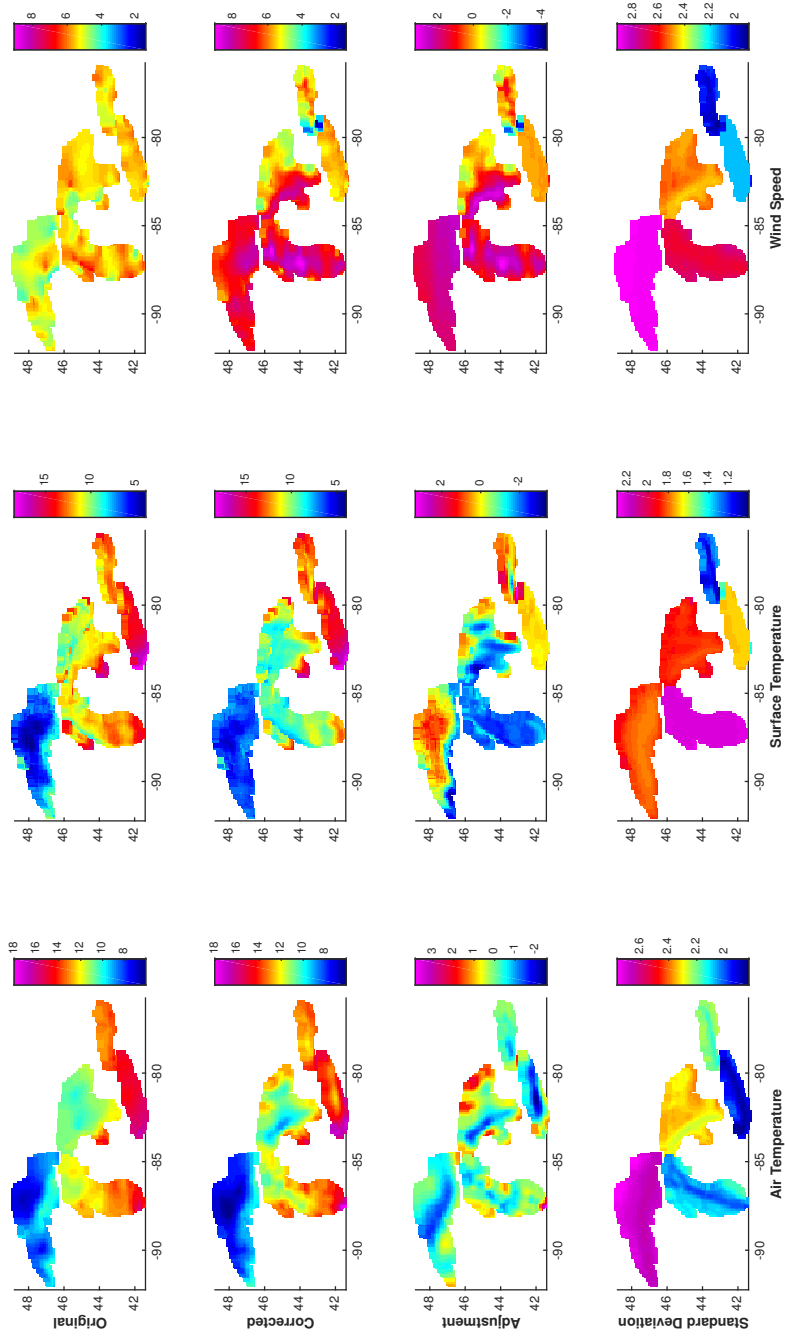


Figure A.28: Spring 2012

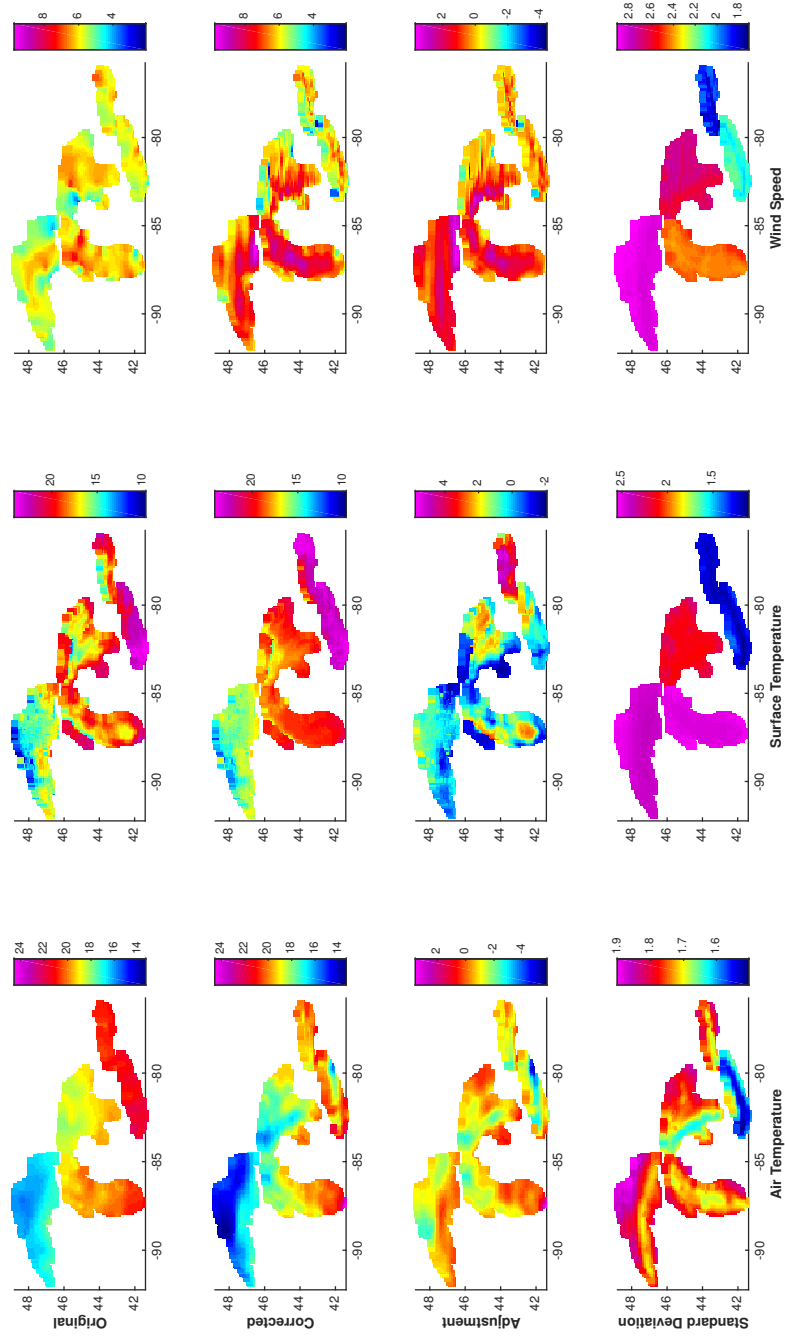


Figure A.29: Summer 2012

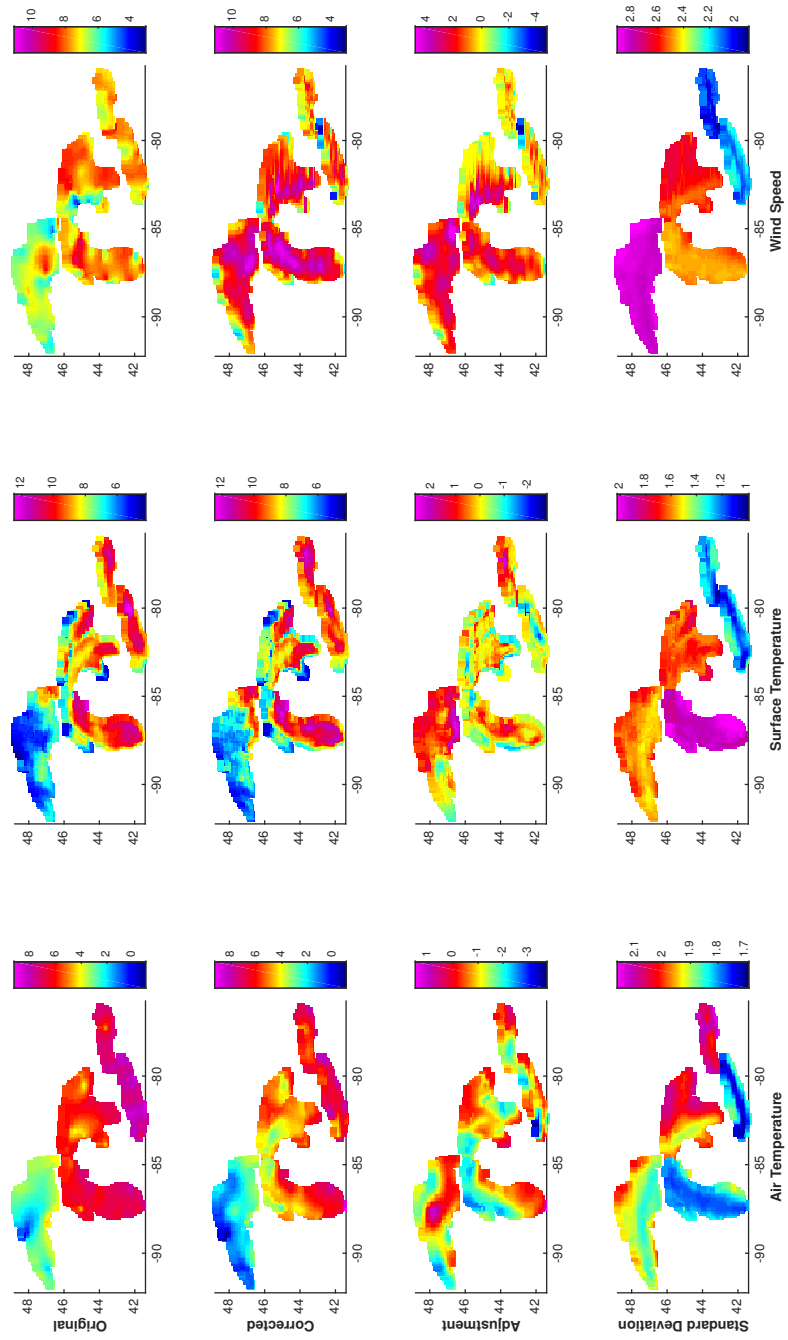


Figure A.30: Fall 2012

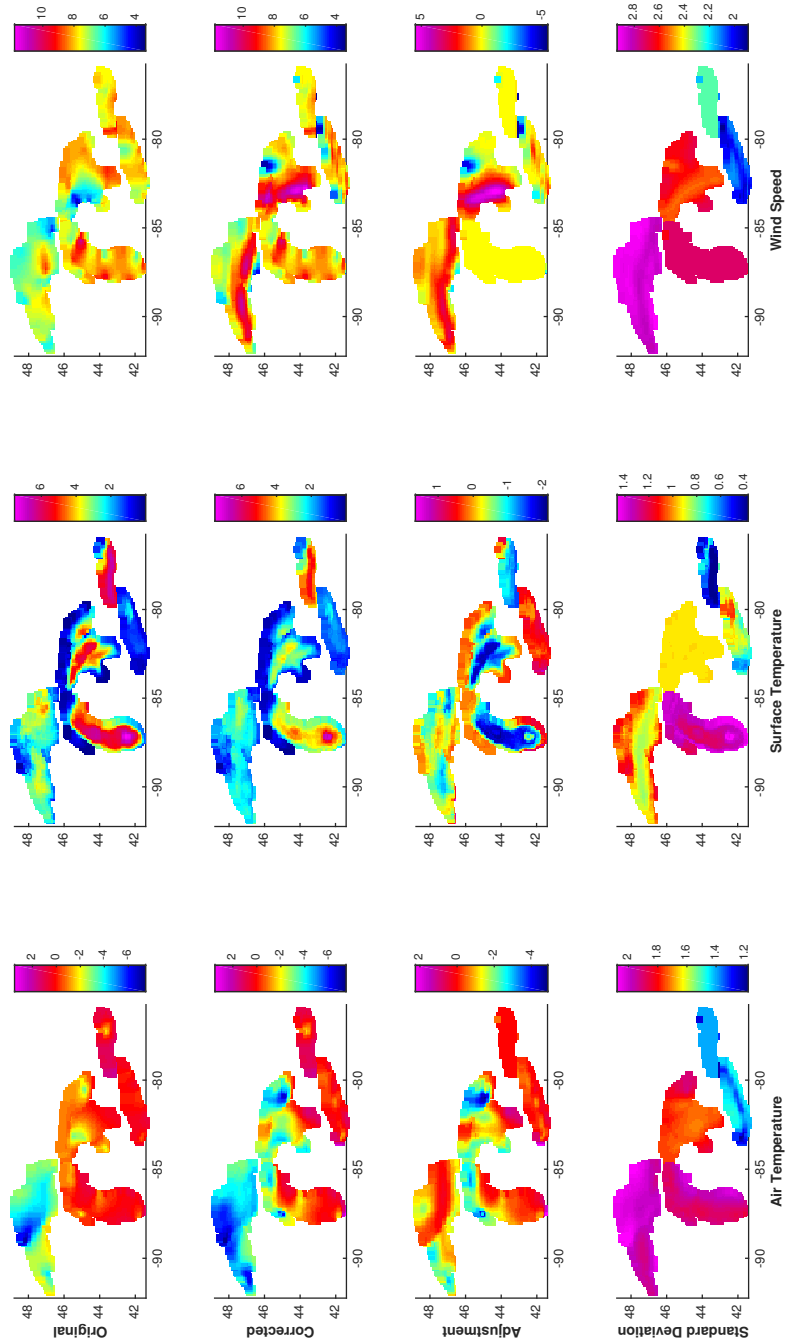


Figure A.31: Winter 2013

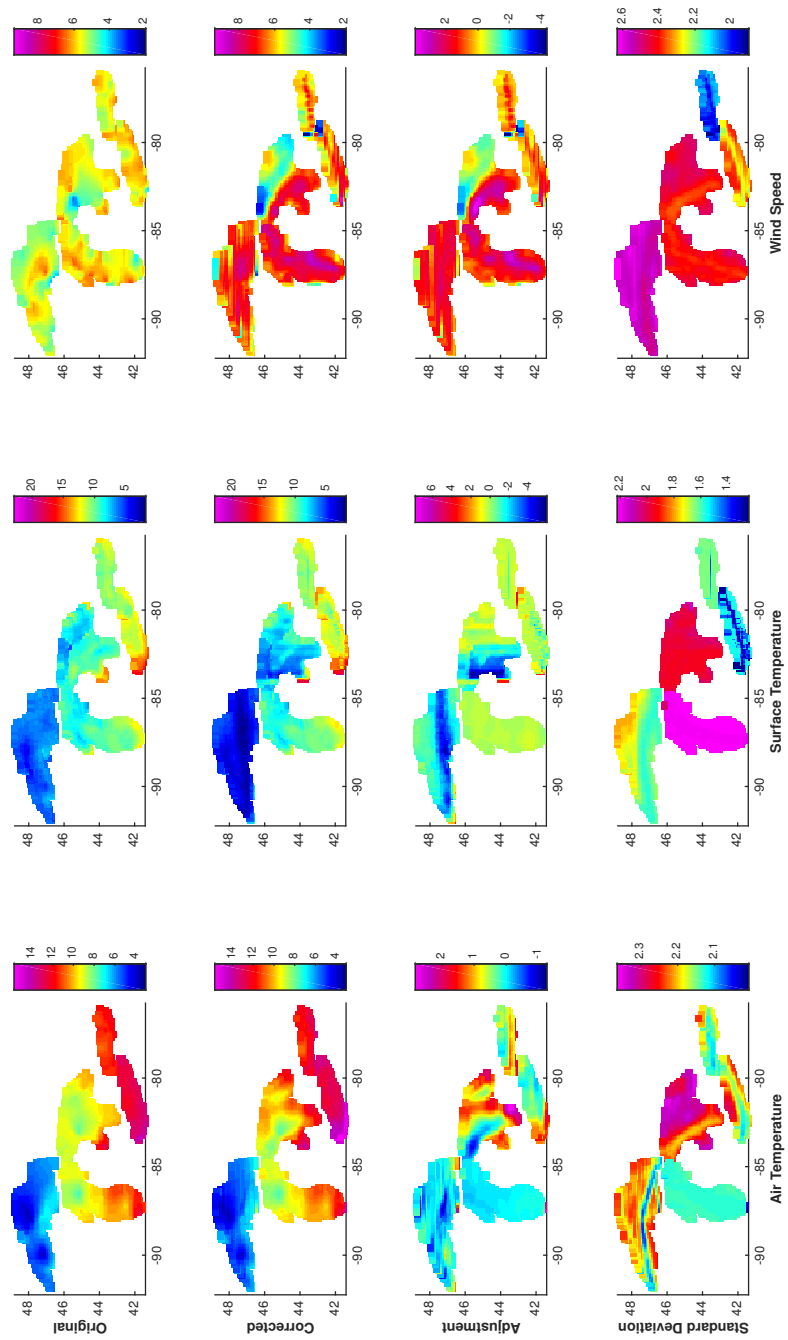


Figure A.32: Spring 2013

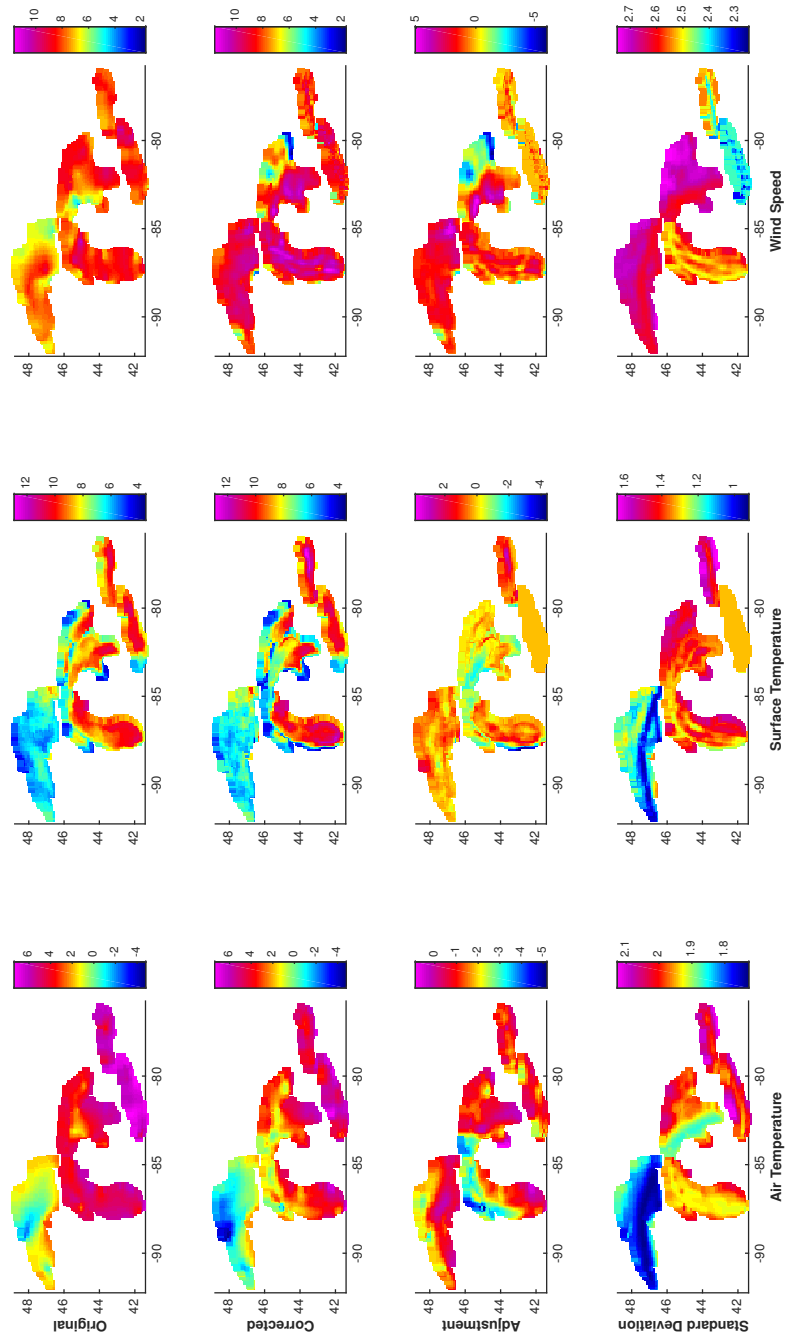


Figure A.33: Fall 2013

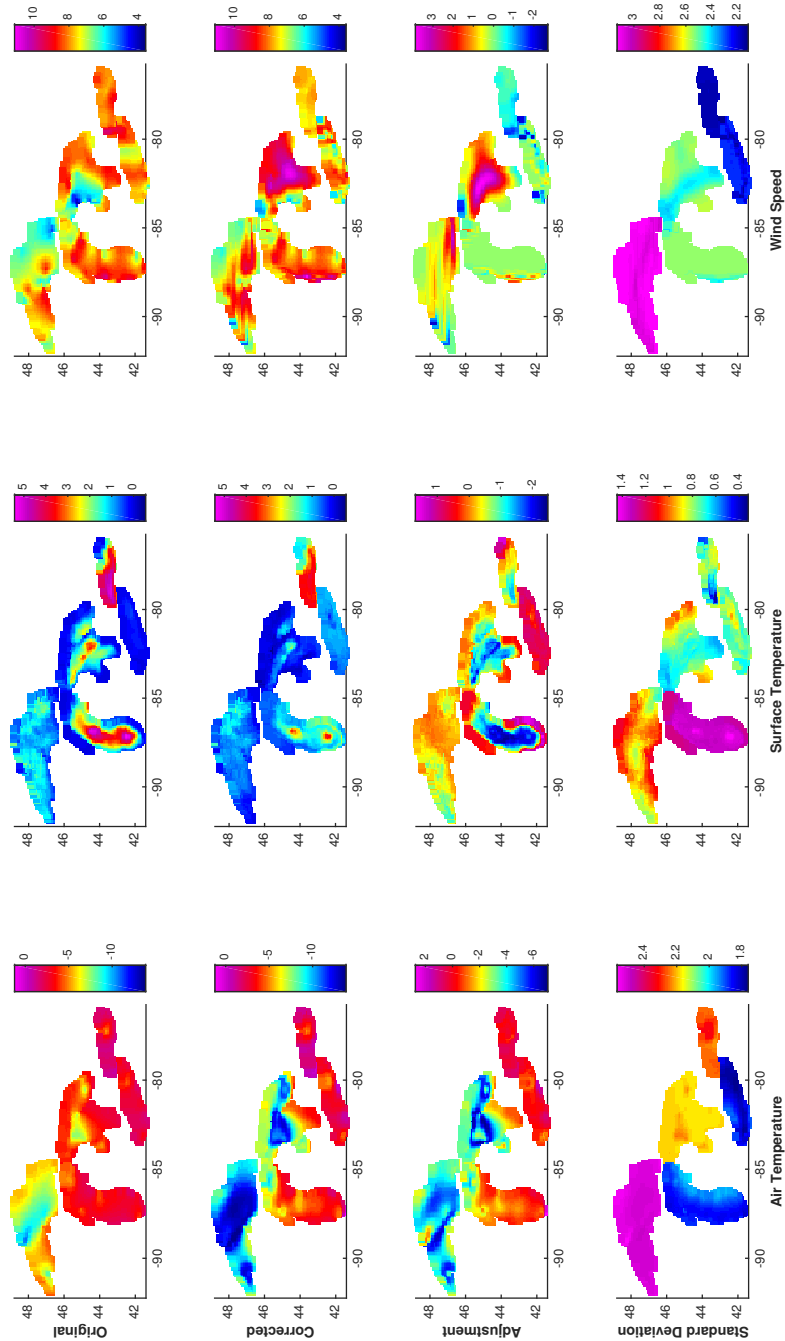


Figure A.34: Winter 2014

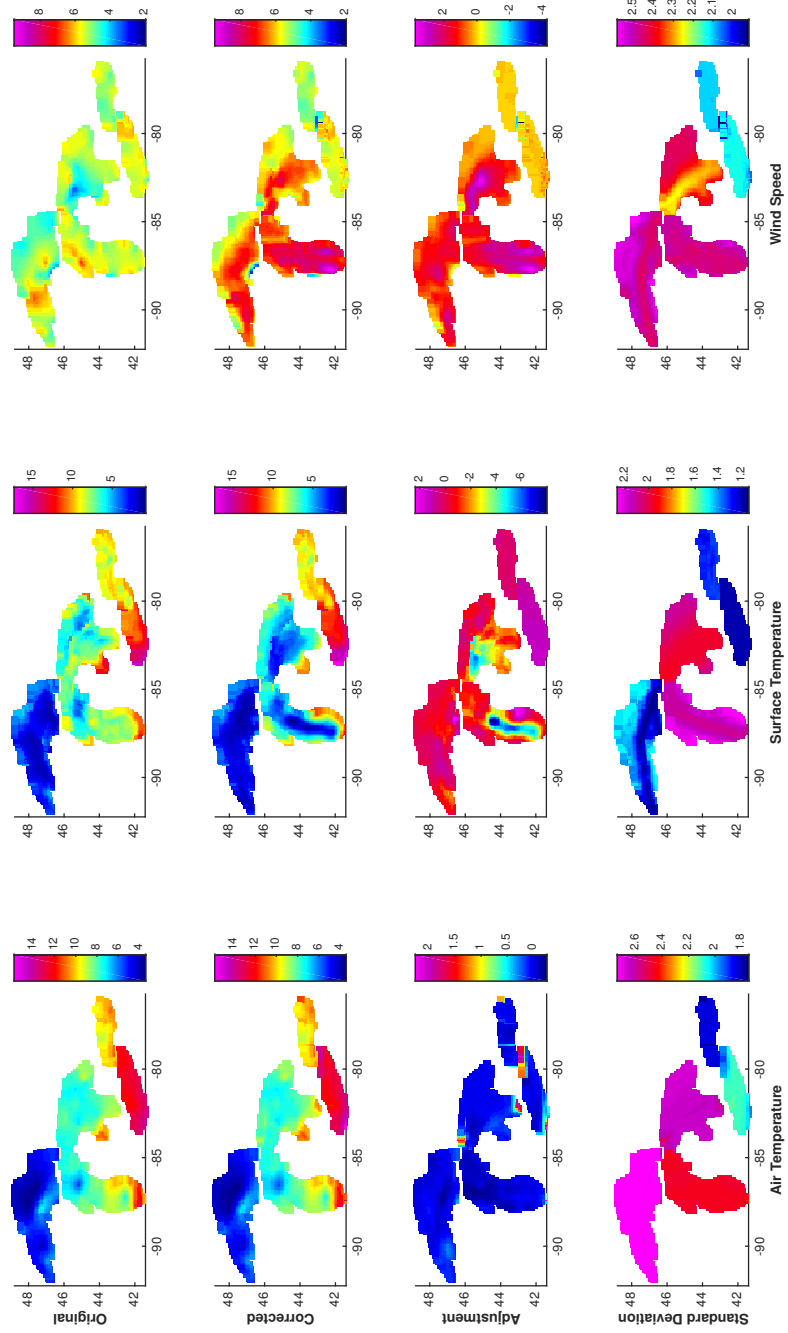


Figure A.35: Spring 2014

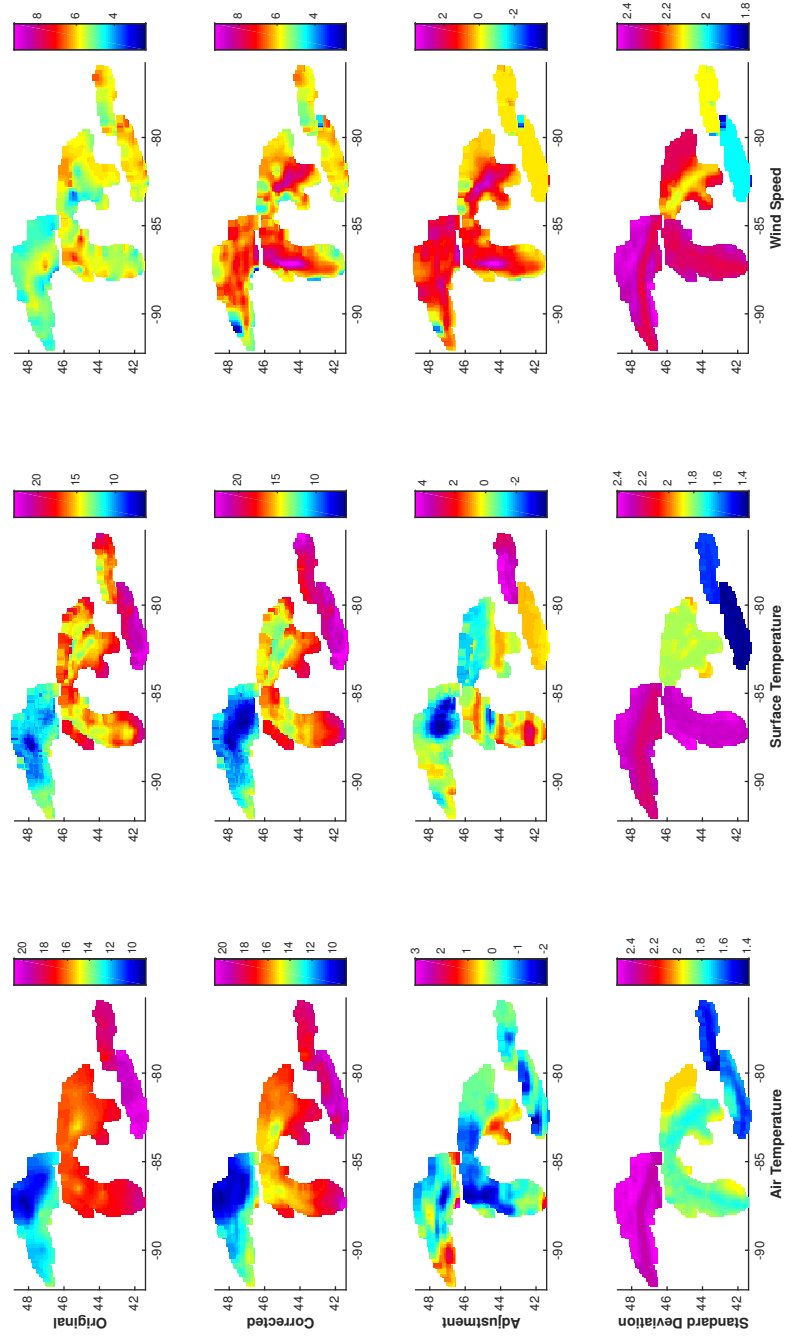


Figure A.36: Summer 2014

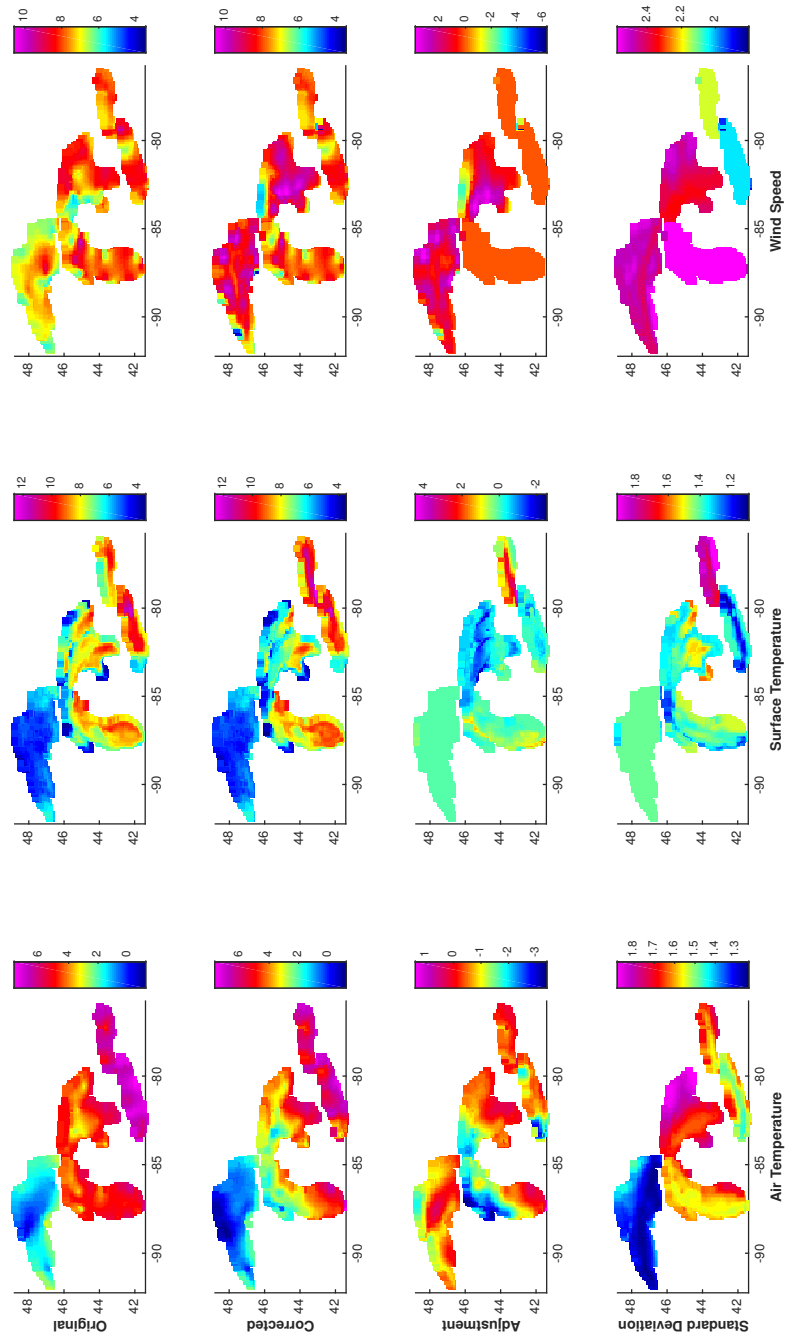


Figure A.37: Fall 2014

APPENDIX B

Supporting Information for Using in-situ measurements to dynamically map large-scale models to site-scale forecasts: A case study using the National Water Model

This file provides figures comparable to Figures 3.6, 3.7, and 3.8 for all the other sites studied.

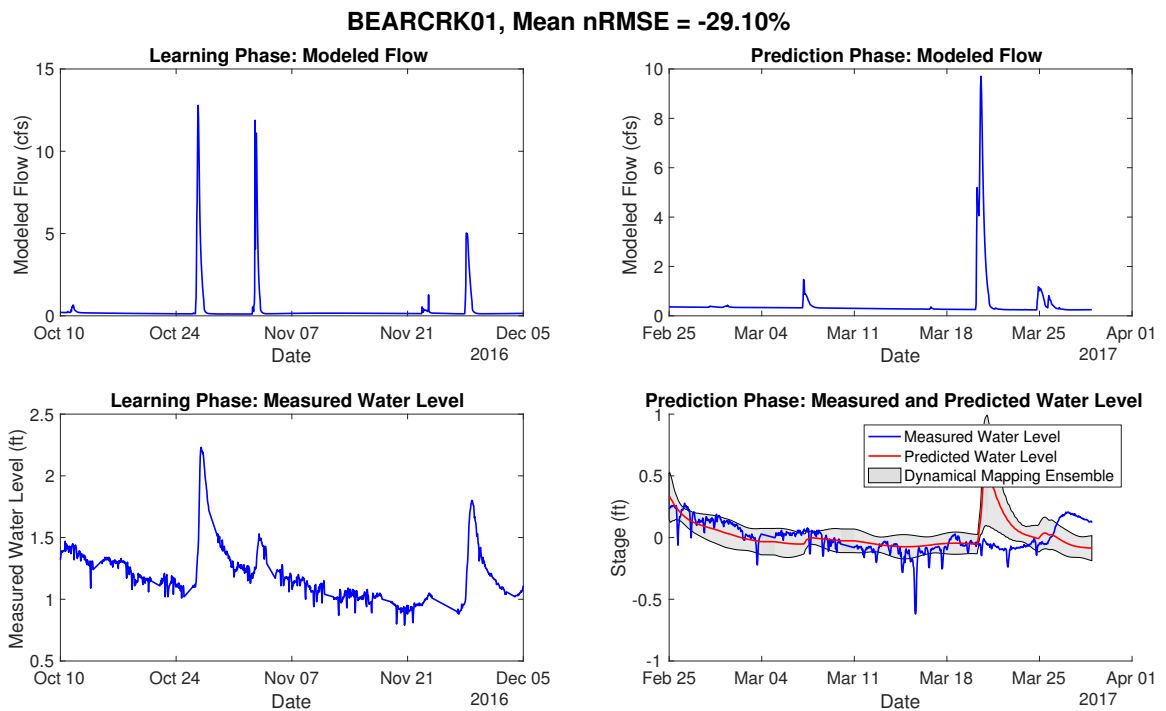


Figure B.1: Site BEARCRK01

BEAVER01, Mean nRMSE = 39.94%

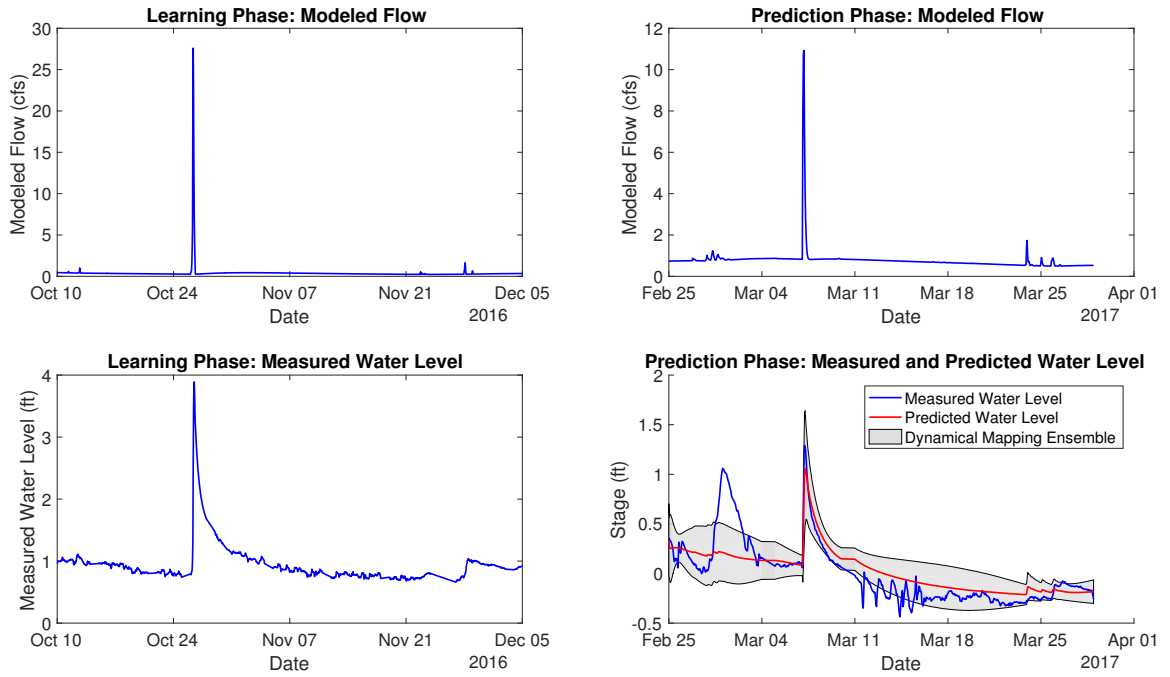


Figure B.2: Site BEAVER01

BEAVER02, Mean nRMSE = 39.69%

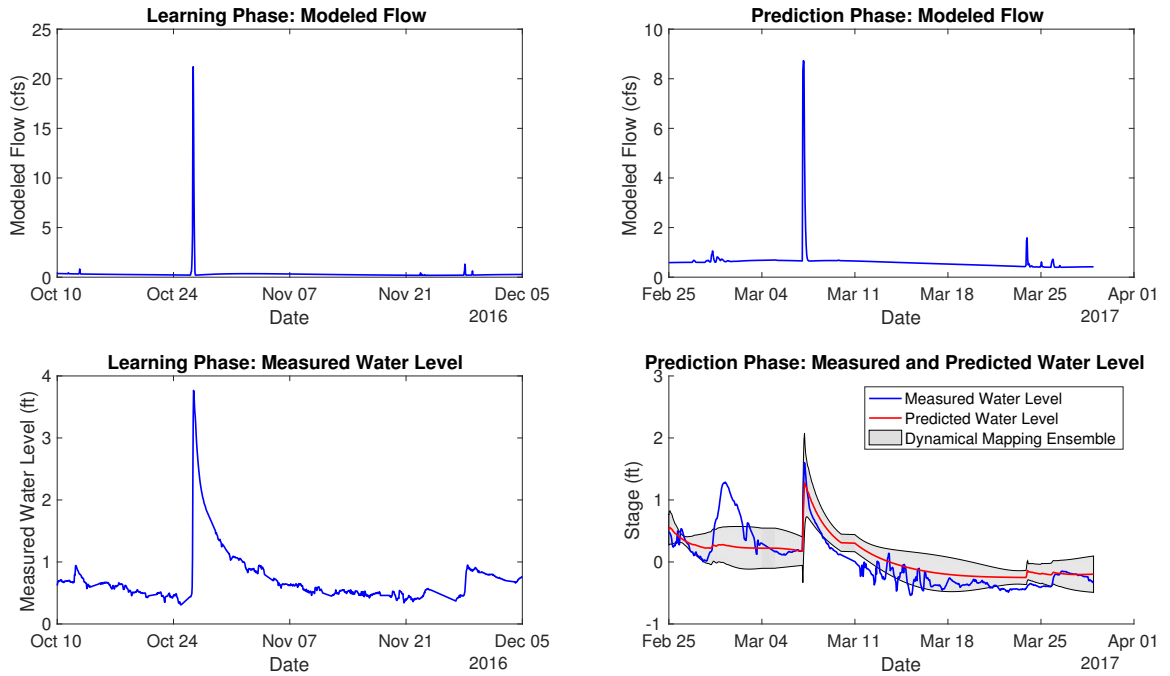


Figure B.3: Site BEAVER02

BEAVER03, Mean nRMSE = 41.10%

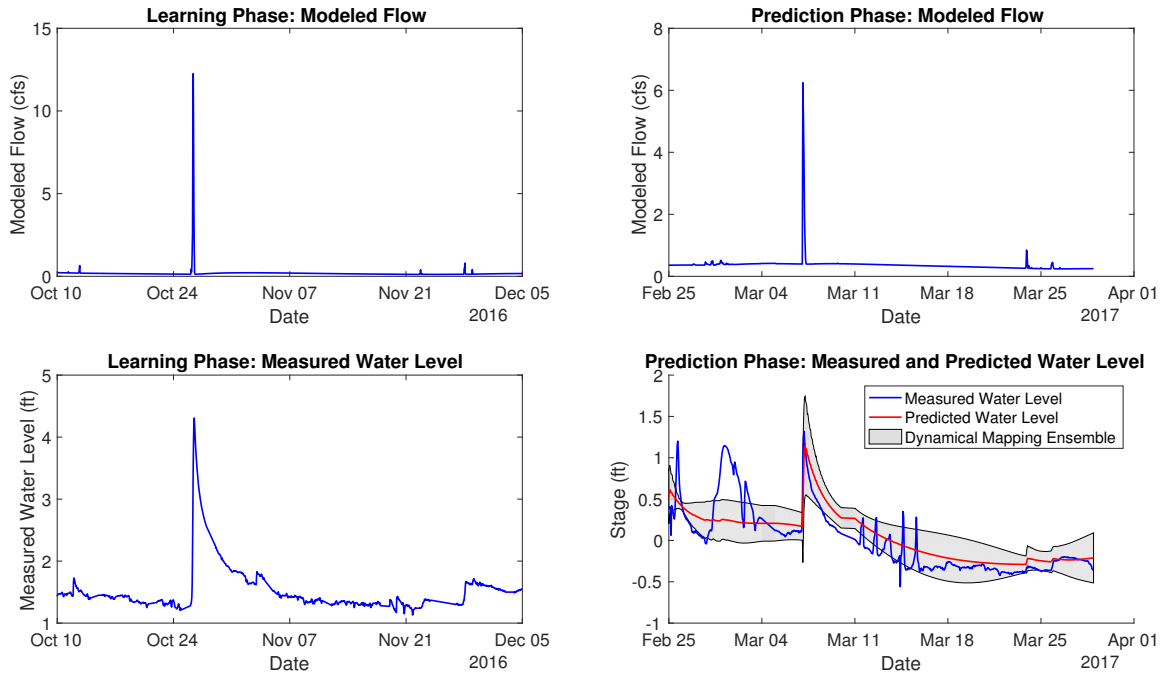


Figure B.4: Site BEAVER03

BEAVERCRK01, Mean nRMSE = -10.12%

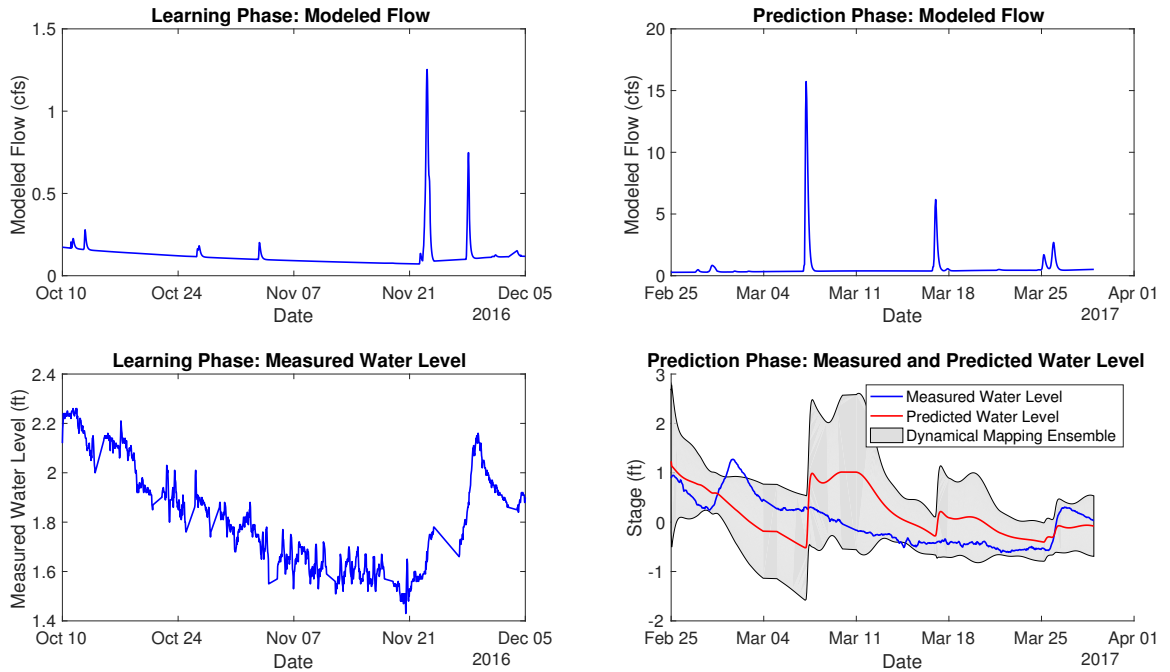


Figure B.5: Site BEAVERCRK01

BEVRCRK01, Mean nRMSE = 40.47%

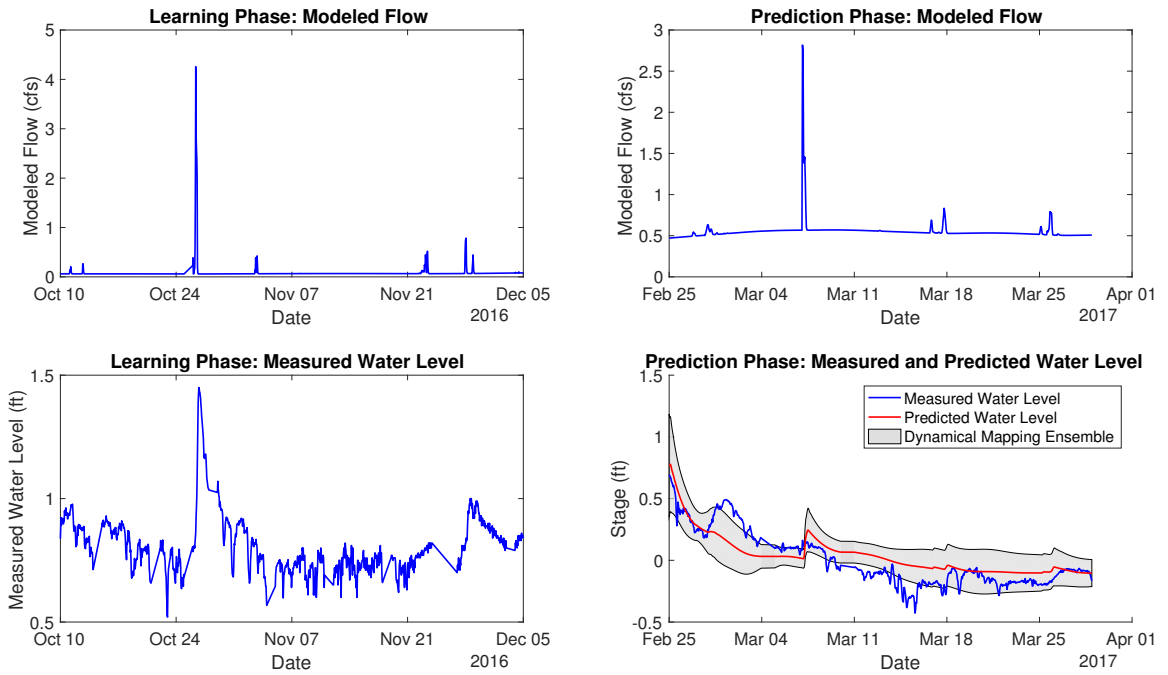


Figure B.6: Site BEVRCRK01

BEVRCRK02, Mean nRMSE = 22.32%

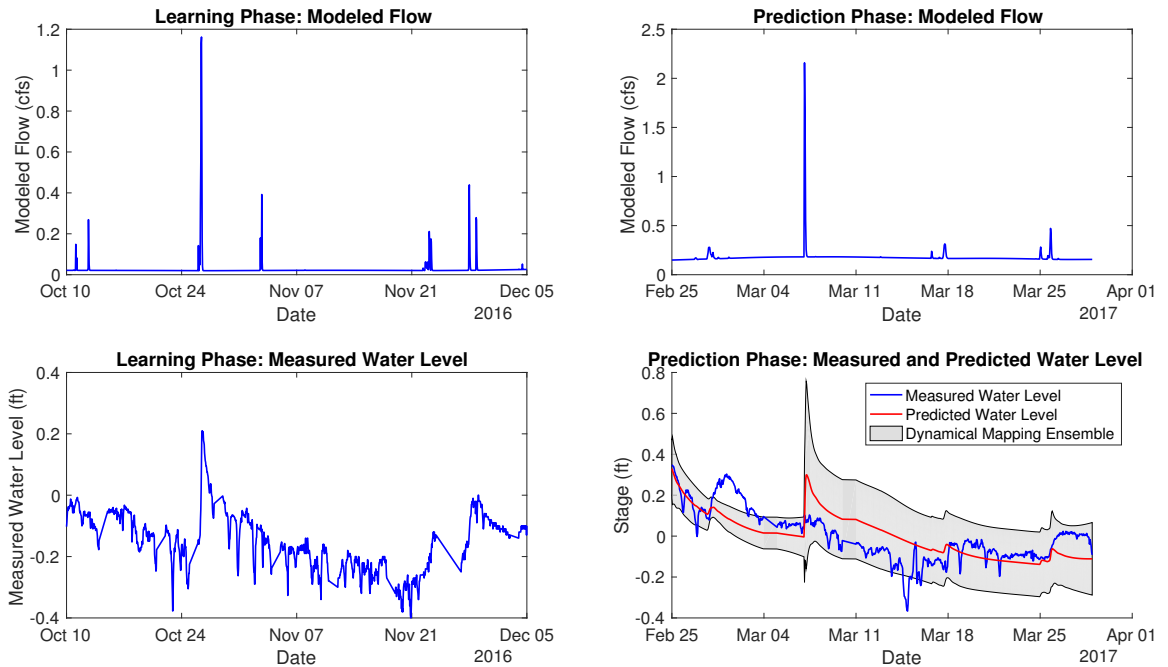


Figure B.7: Site BEVRCRK02

BFFLOCR01, Mean nRMSE = 1.54%

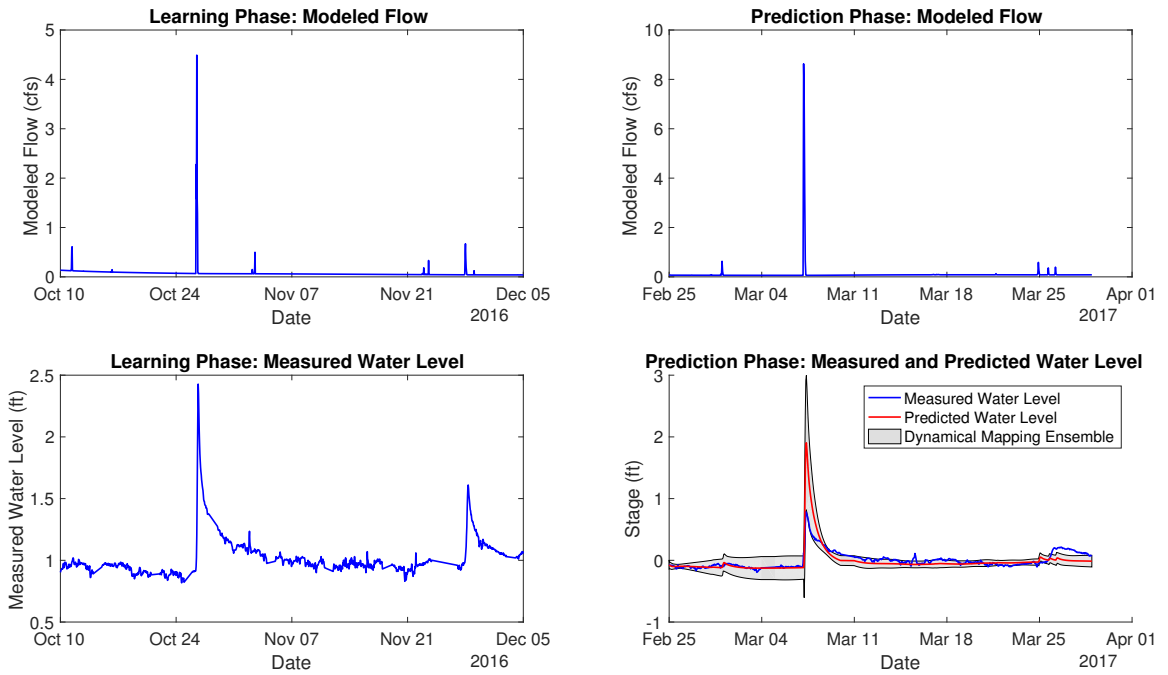


Figure B.8: Site BFFLOCR01

BFFLOCR02, Mean nRMSE = -9.80%

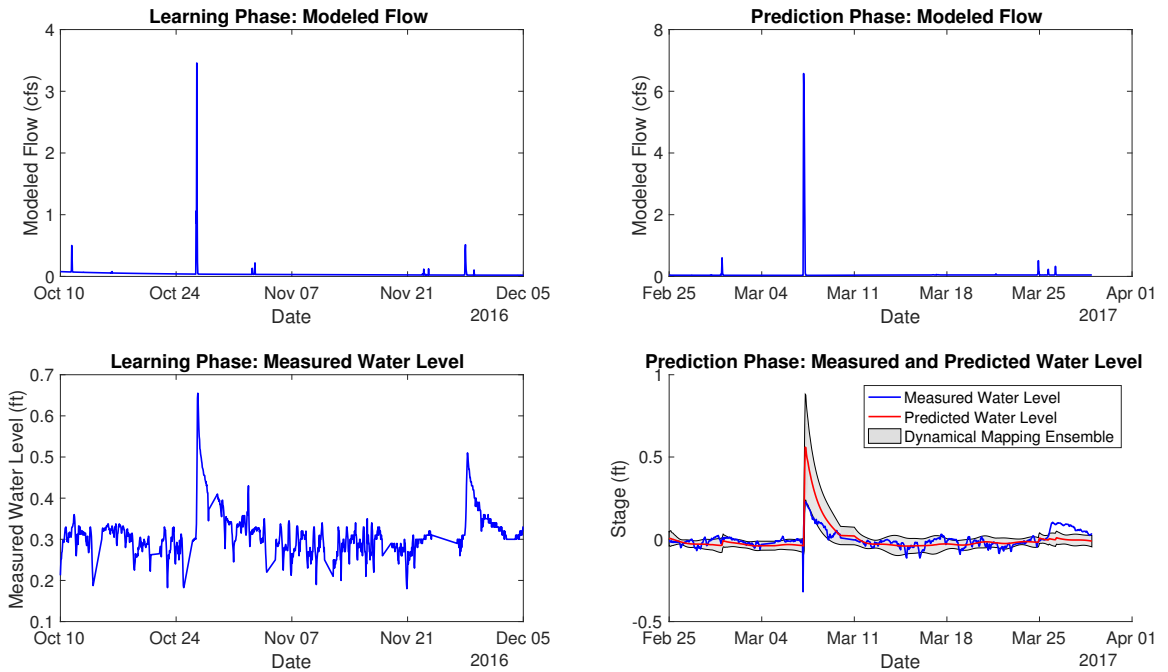


Figure B.9: Site BFFLOCR02

BLCKHWK01, Mean nRMSE = 11.21%

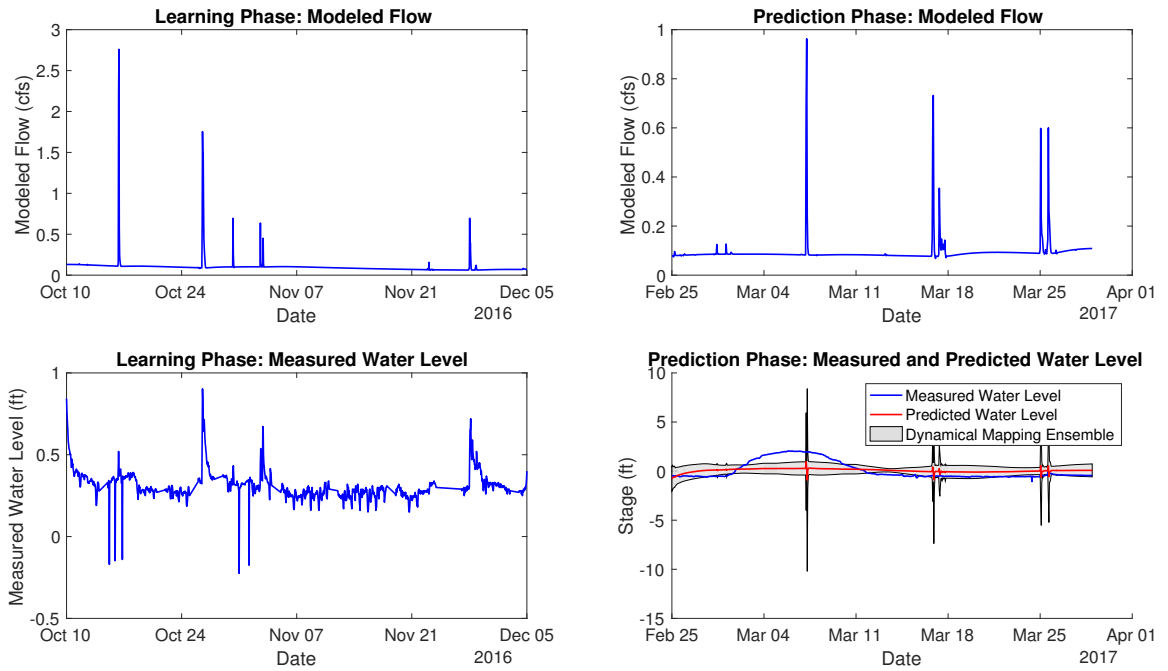


Figure B.10: Site BLCKHWK01

BNNTTCR01, Mean nRMSE = -11.91%

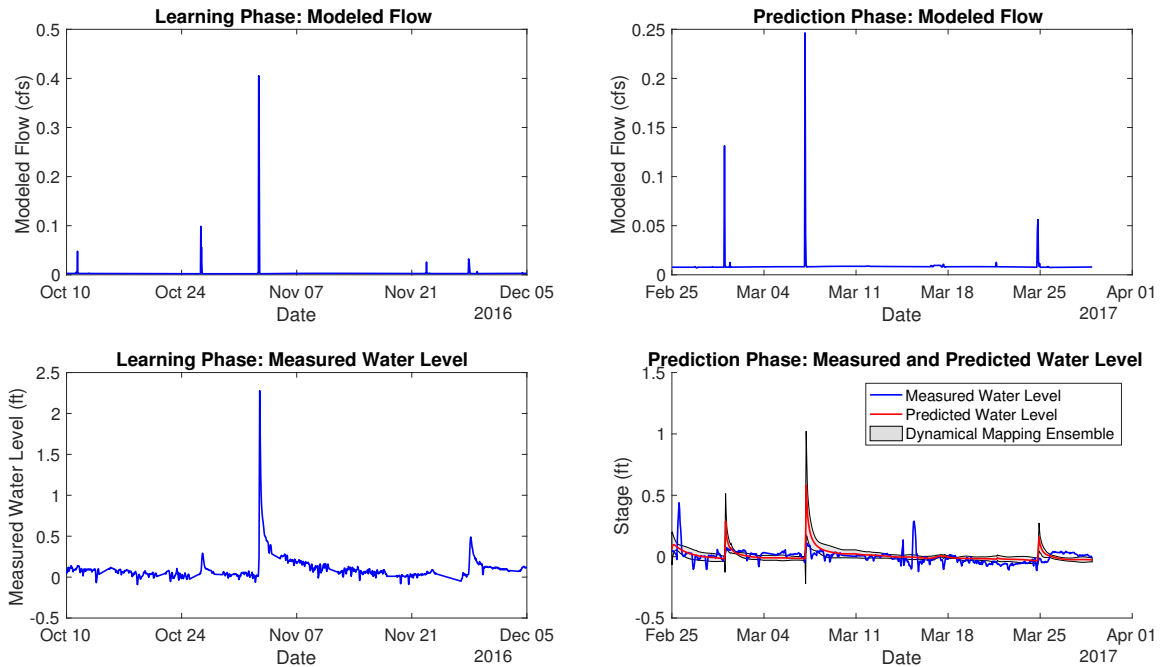


Figure B.11: Site BNNTTCR01

BOYERRV01, Mean nRMSE = 38.35%

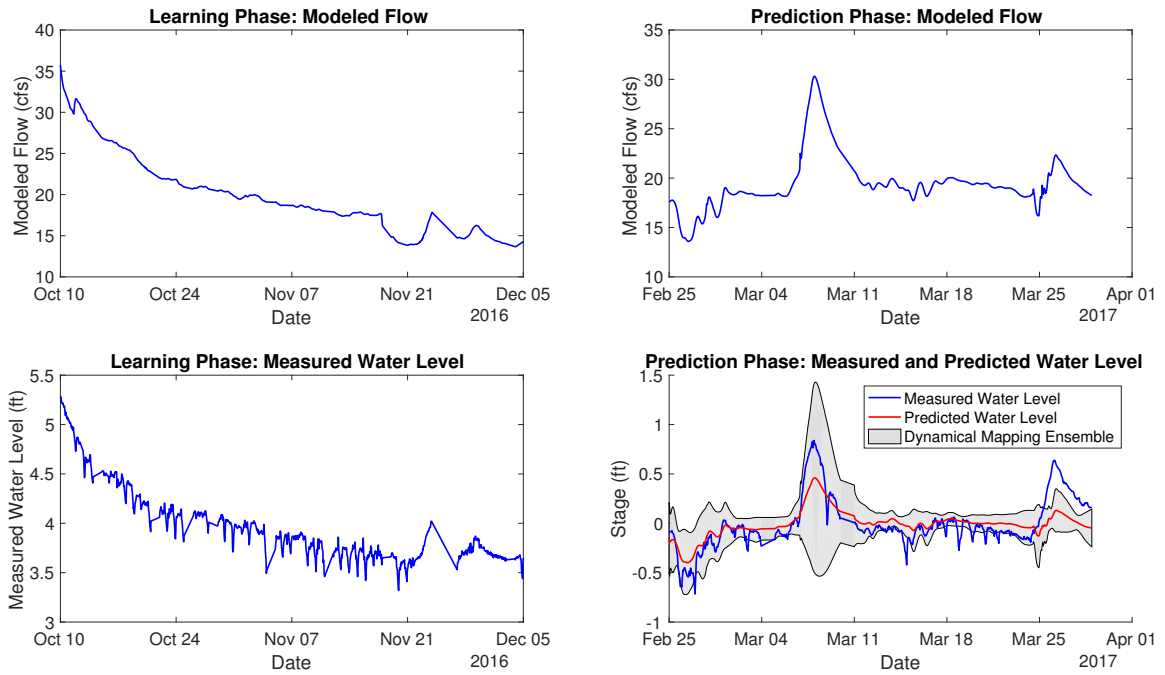


Figure B.12: Site BOYERRV01

BOYERRV02, Mean nRMSE = 35.17%

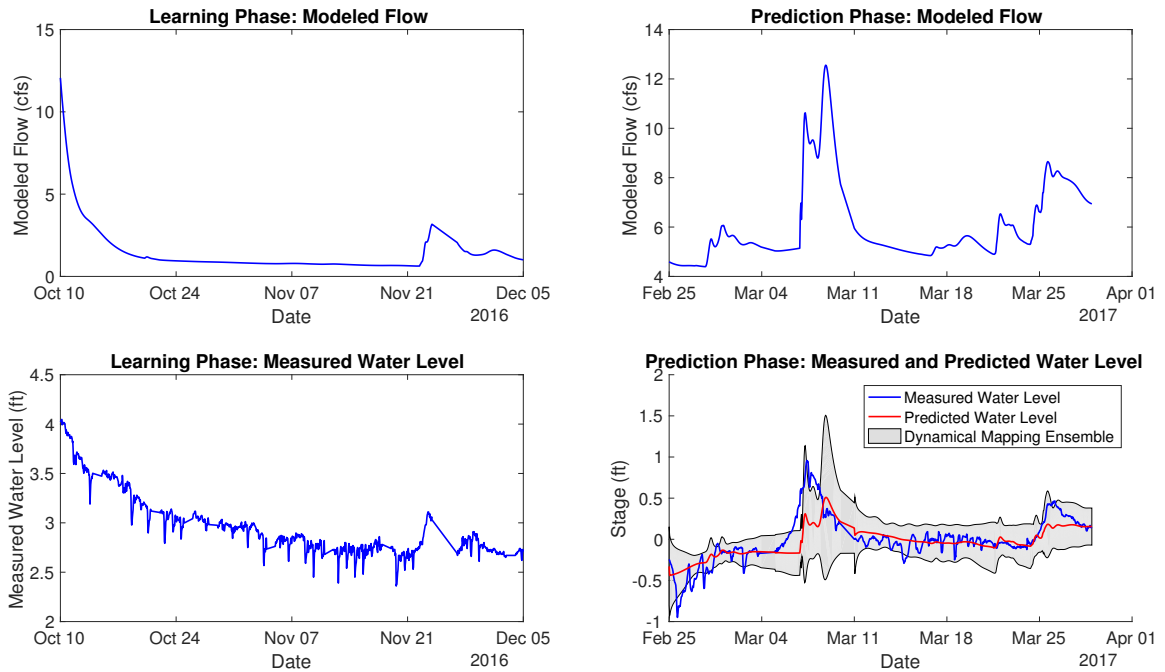


Figure B.13: Site BOYERRV02

BRCHFLD01, Mean nRMSE = -214.83%

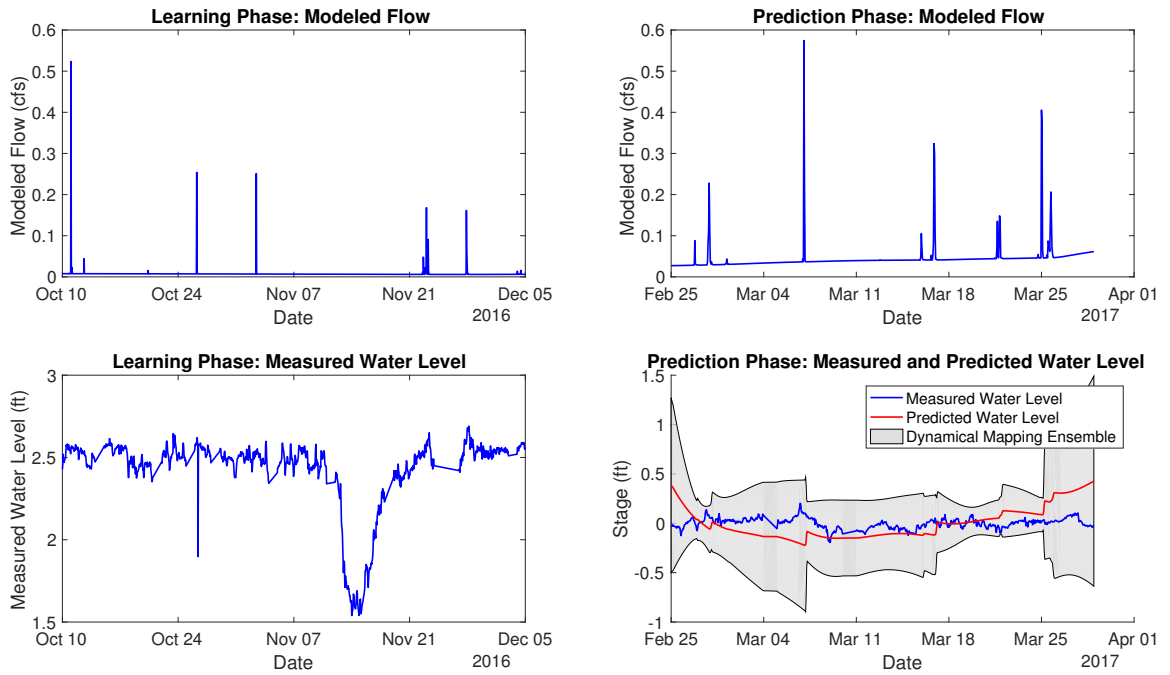


Figure B.14: Site BRCHFLD01

BRCKCMP01, Mean nRMSE = 19.23%

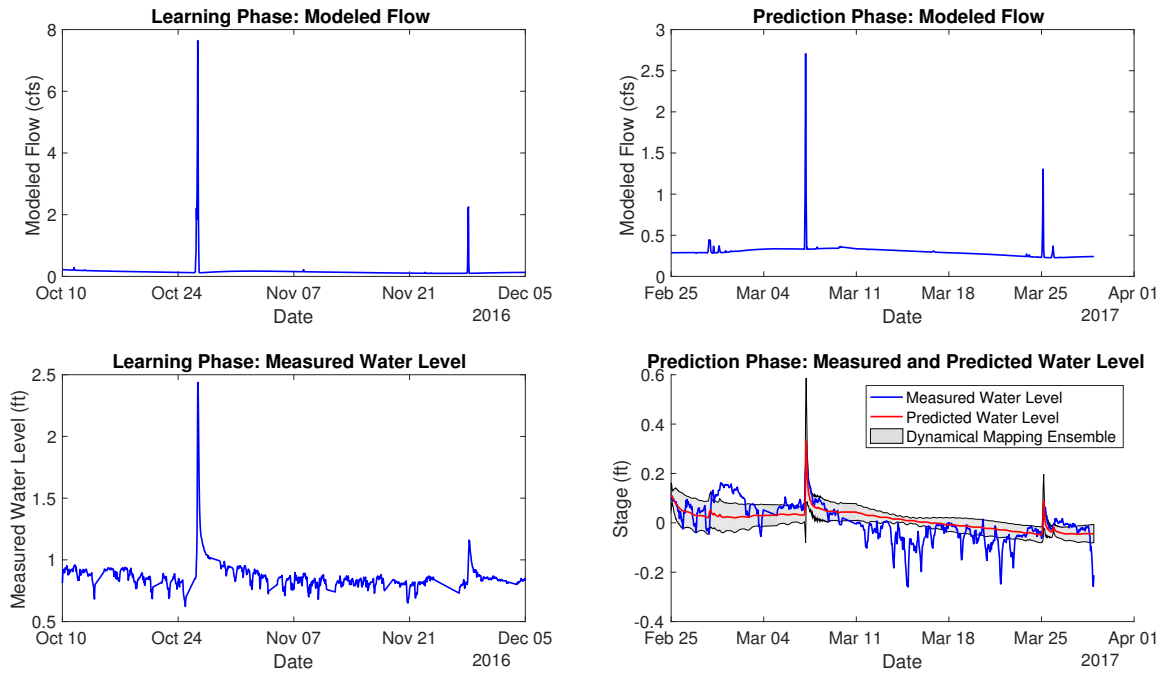


Figure B.15: Site BRCKCMP01

BRSHYCR01, Mean nRMSE = -23.72%

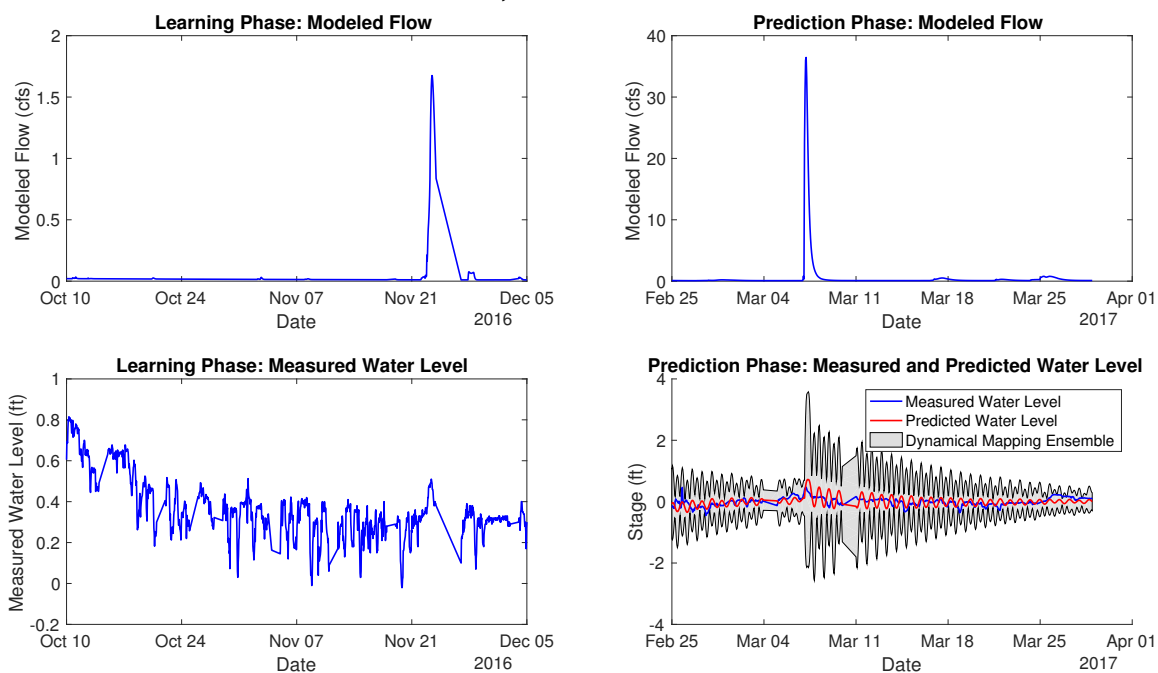


Figure B.16: Site BRSHYCR01

BVRCR01, Mean nRMSE = 49.22%

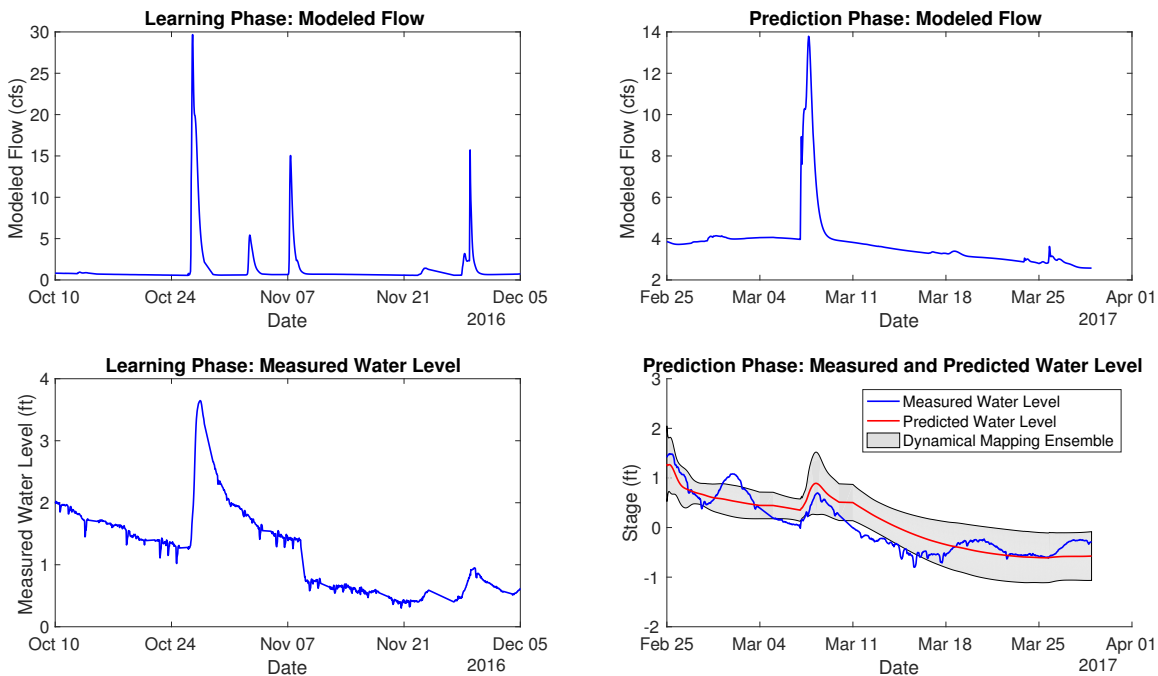


Figure B.17: Site BVRCR01

CAMPCR01, Mean nRMSE = 13.70%

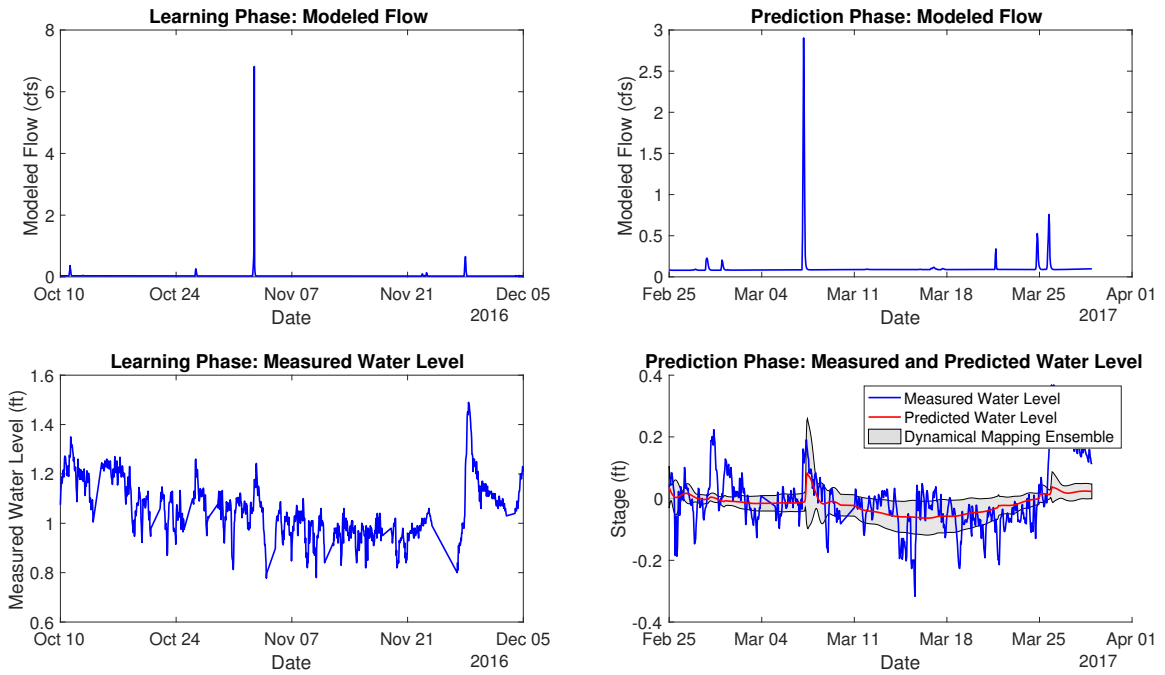


Figure B.18: Site CAMPCR01

CEDARRV02, Mean nRMSE = 68.78%

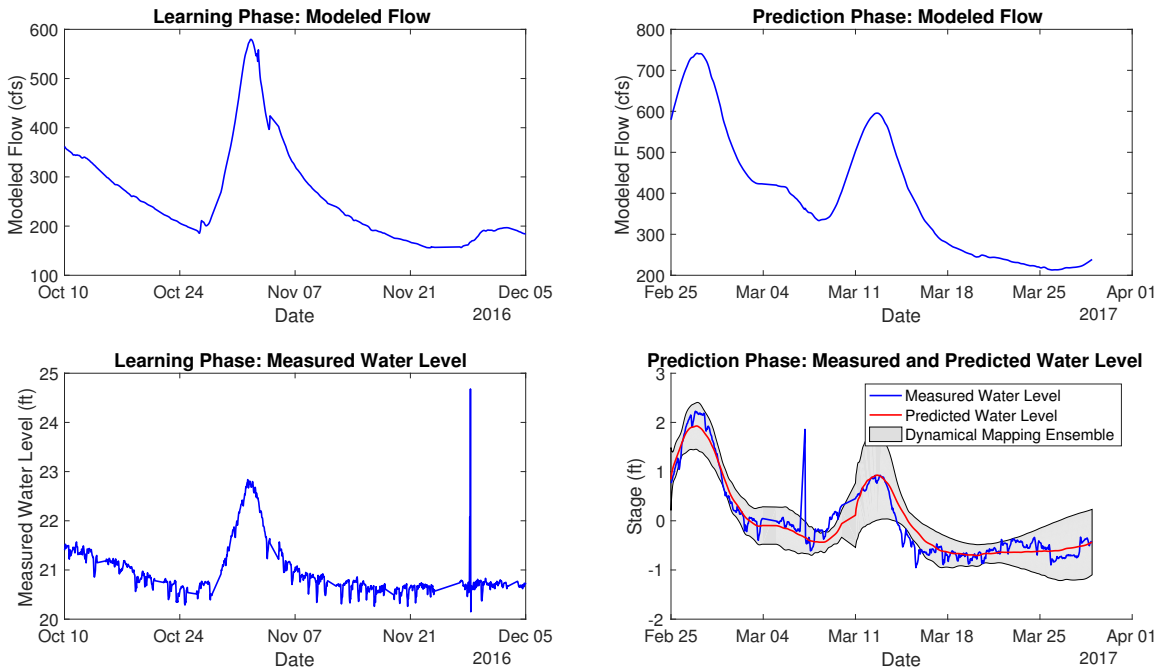


Figure B.19: Site CEDARRV02

CEDARRV04, Mean nRMSE = 33.32%

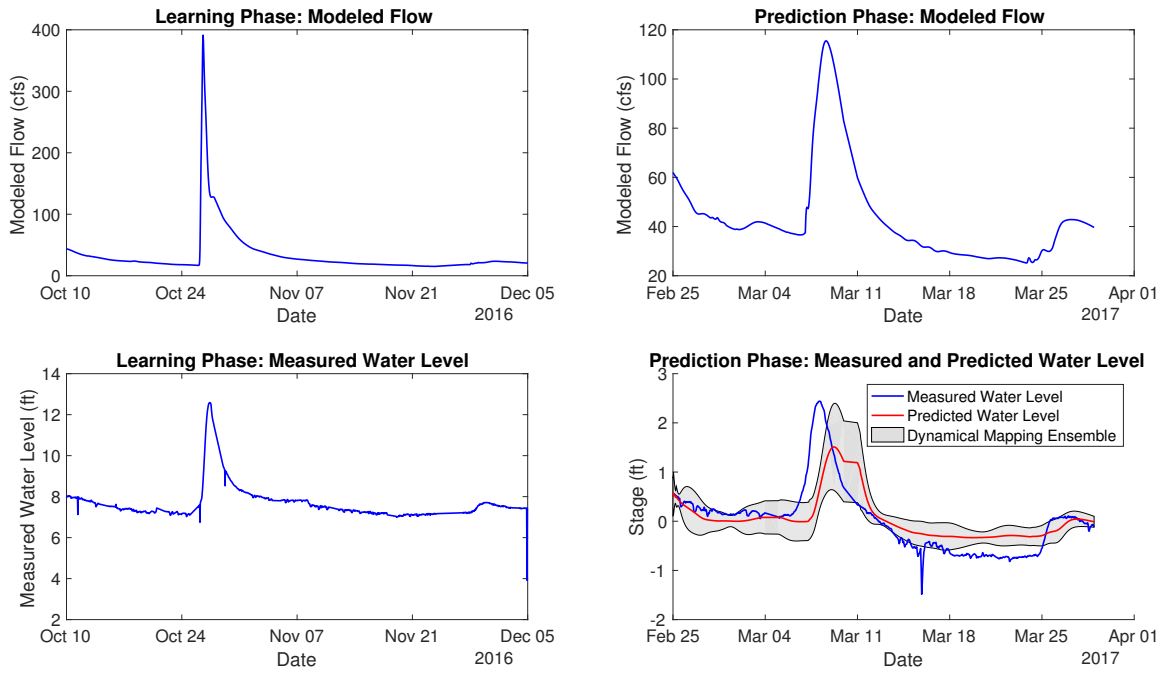


Figure B.20: Site CEDARRV04

CEDARRV05, Mean nRMSE = 48.69%

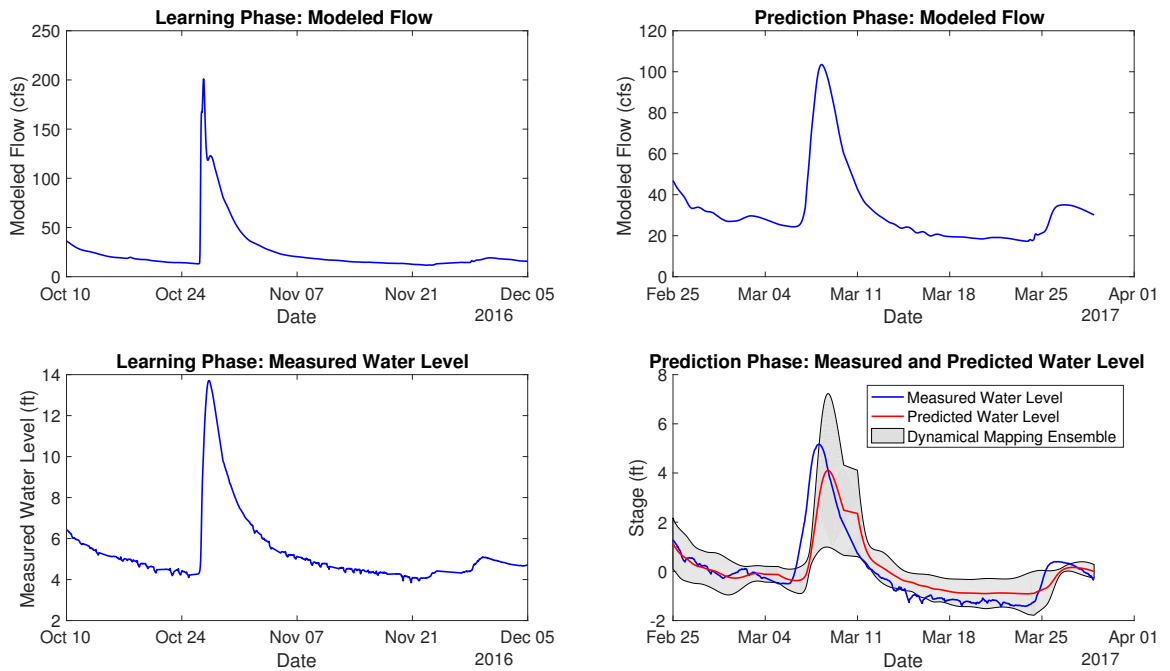


Figure B.21: Site CEDARRV05

CHQSTCR01, Mean nRMSE = 9.14%

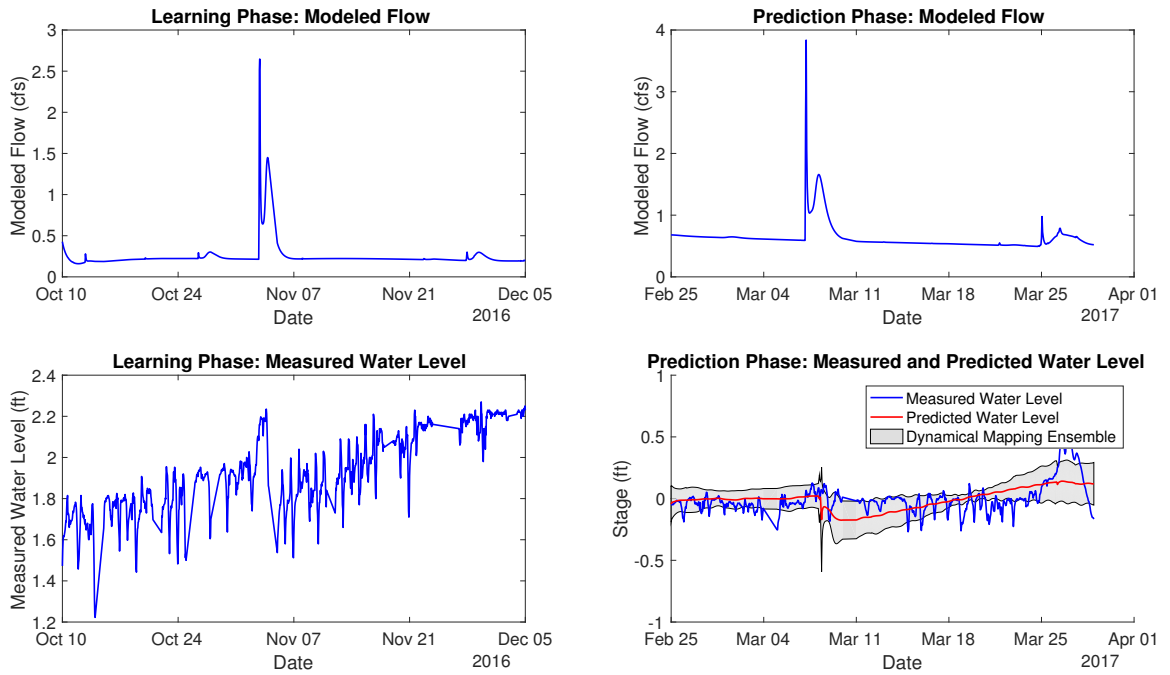


Figure B.22: Site CHQSTCR01

CHQSTCR02, Mean nRMSE = 12.31%

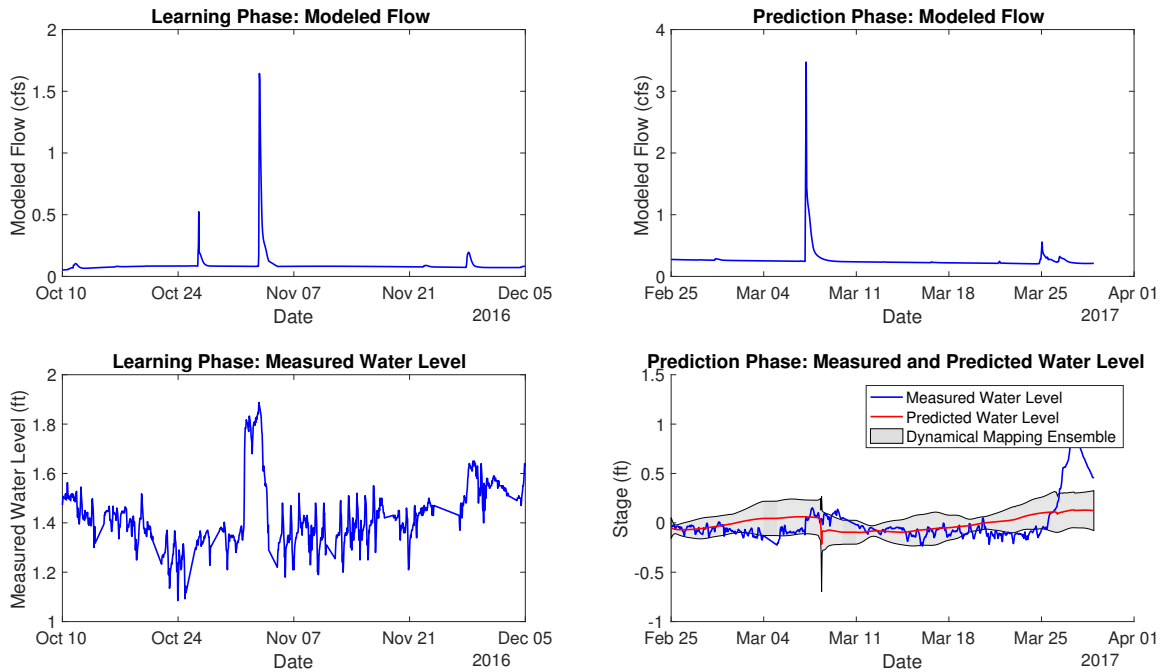


Figure B.23: Site CHQSTCR02

CHSLSCR01, Mean nRMSE = 21.53%

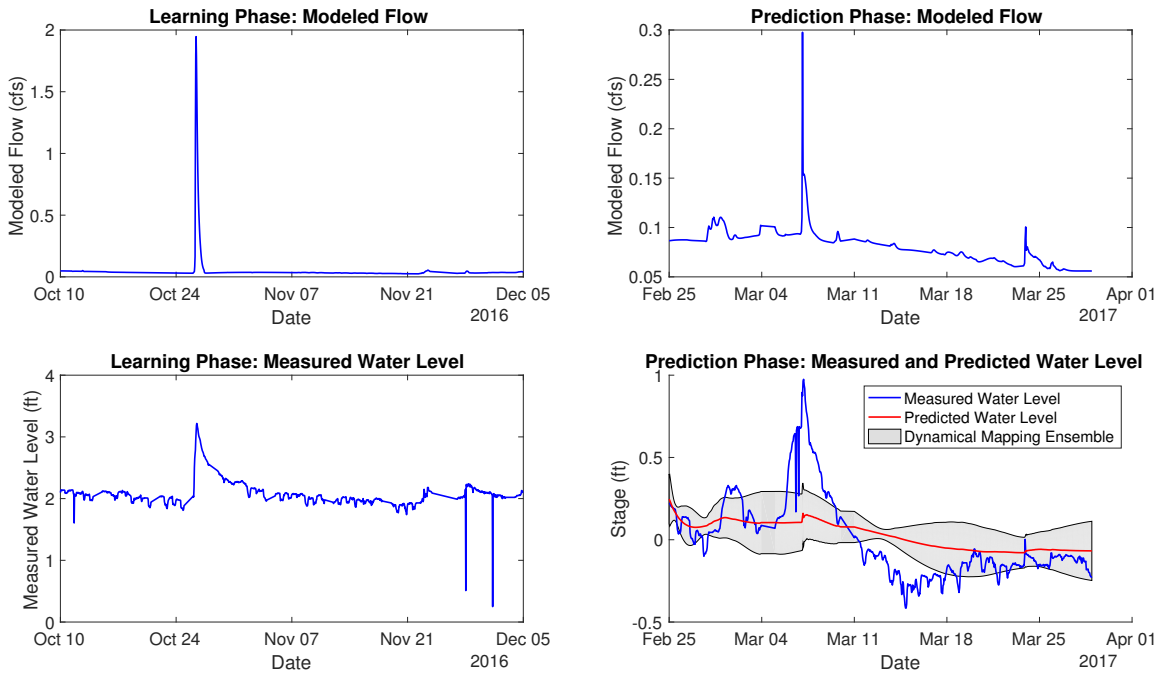


Figure B.24: Site CHSLSCR01

CLRCRK01, Mean nRMSE = 55.69%

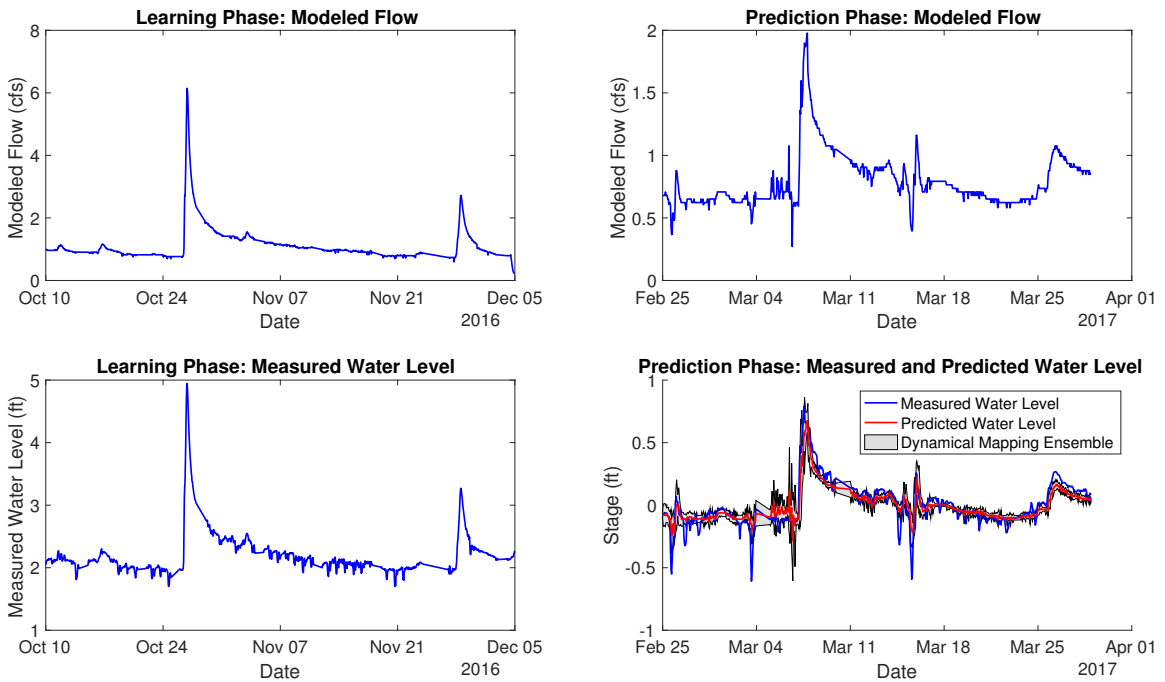


Figure B.25: Site CLRCRK01

CLRCRK03, Mean nRMSE = 21.31%

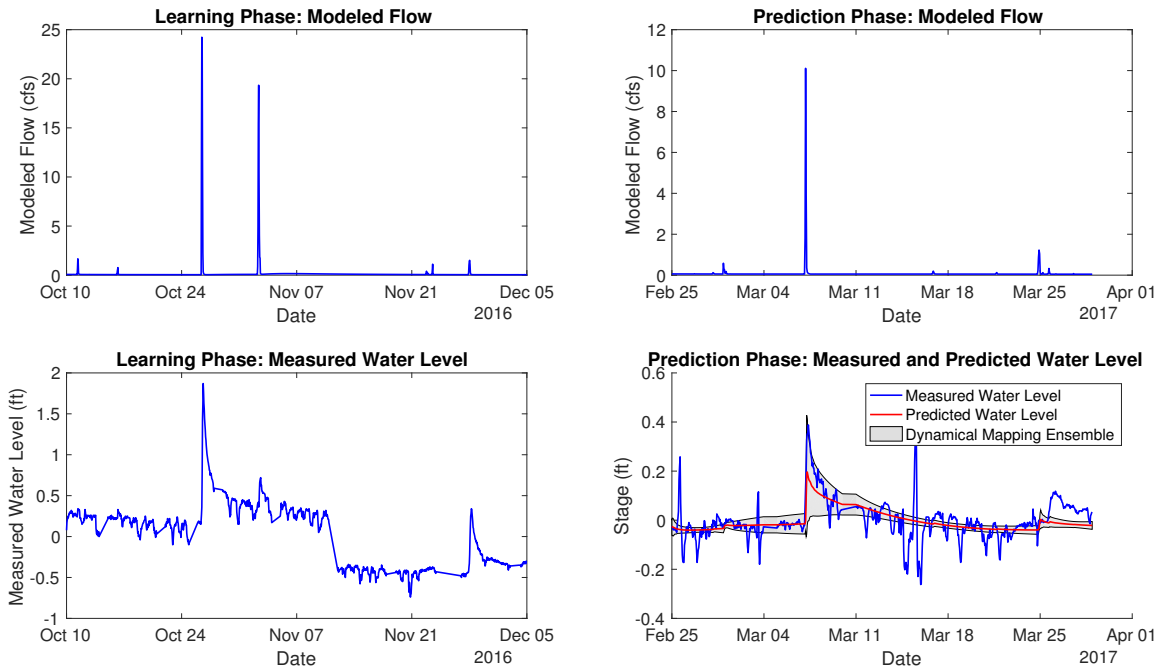


Figure B.26: Site CLRCRK03

CLRCRKS01, Mean nRMSE = 39.07%

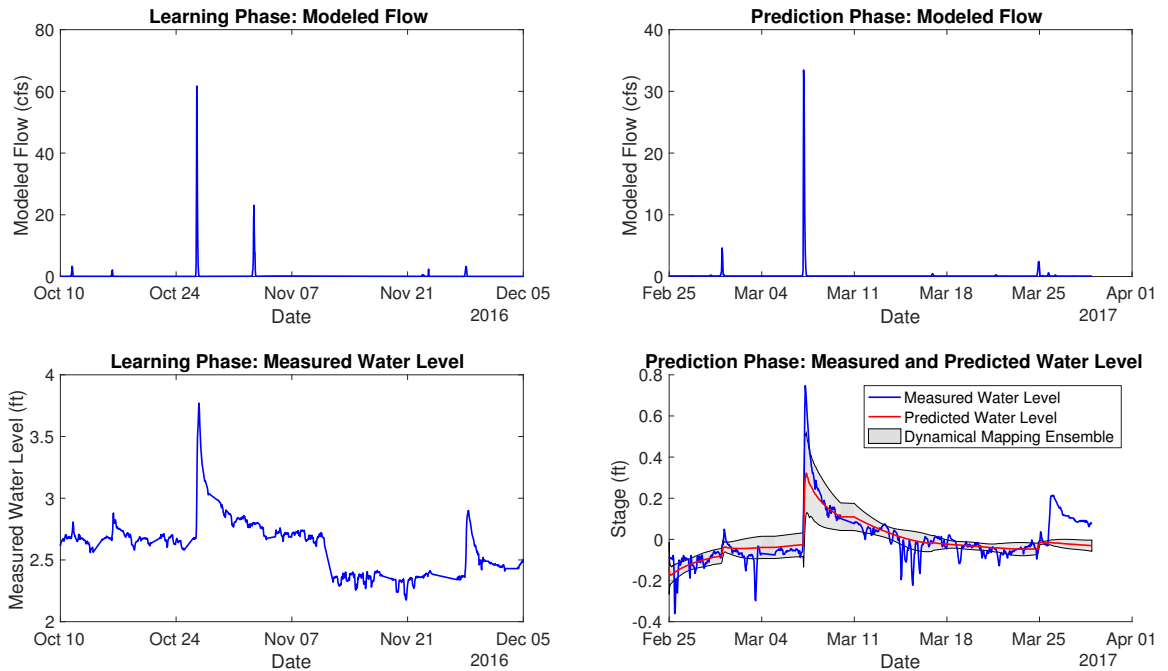


Figure B.27: Site CLRCRKS01

CLRCRKW01, Mean nRMSE = 25.45%

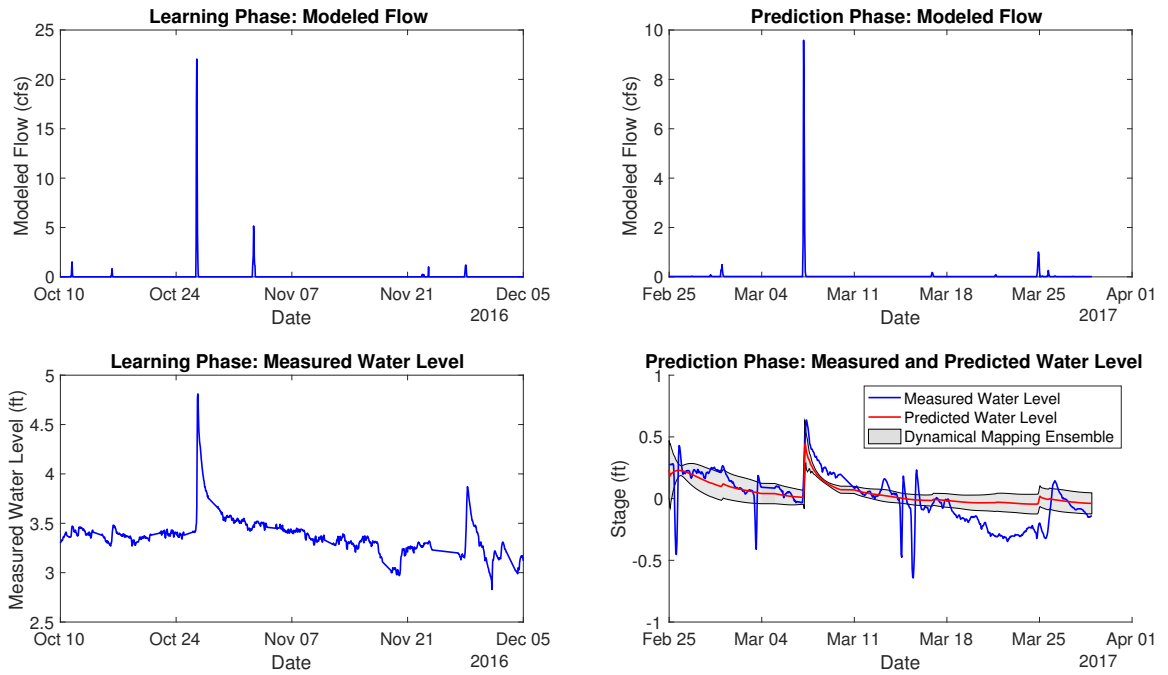


Figure B.28: Site CLRCRKW01

COONCR01, Mean nRMSE = 38.22%

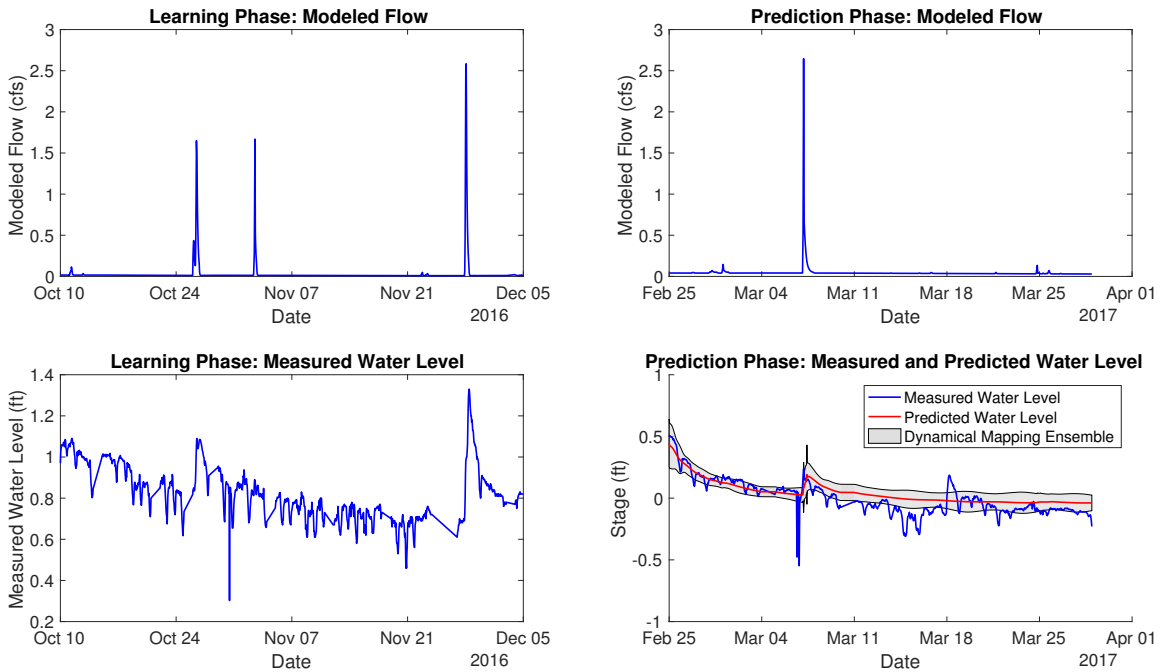


Figure B.29: Site COONCR01

CRANECR02, Mean nRMSE = 38.28%

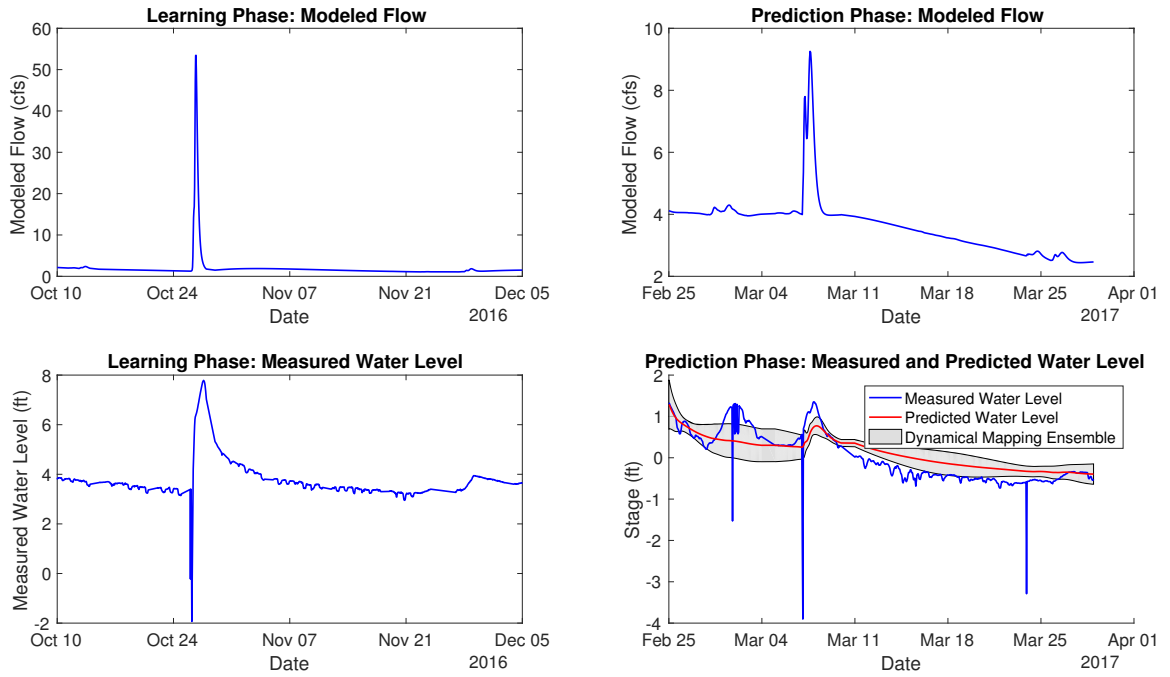


Figure B.30: Site CRANECR02

CTFSHCR01, Mean nRMSE = 24.70%

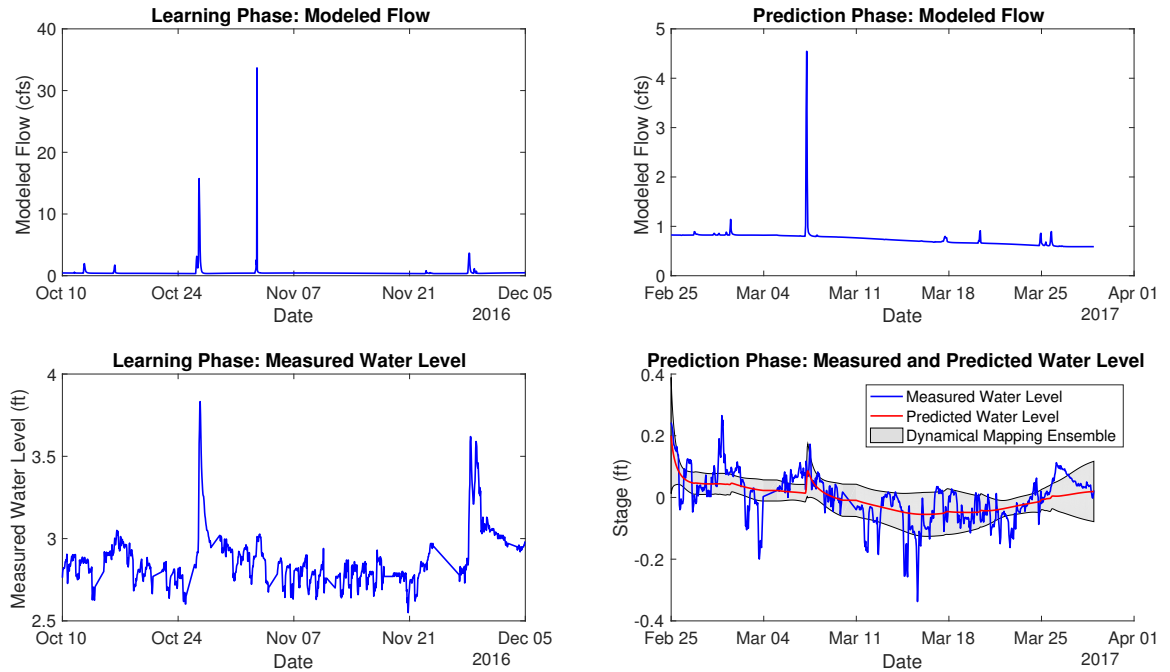


Figure B.31: Site CTFSHCR01

DEEPCRK01, Mean nRMSE = 38.48%

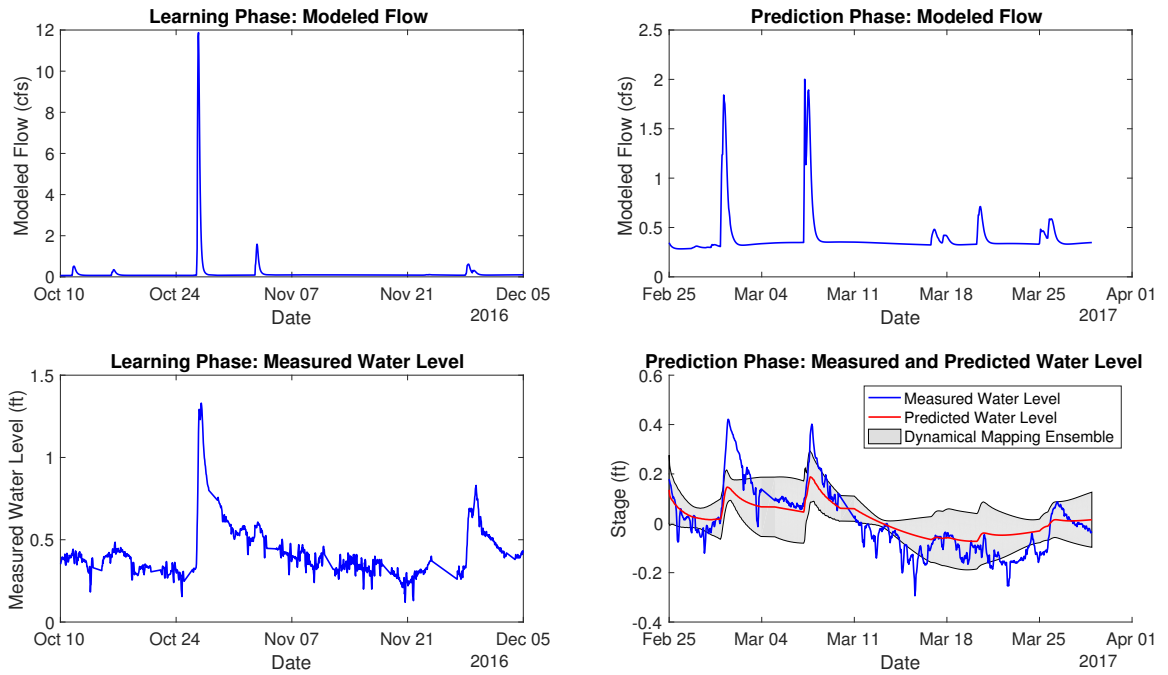


Figure B.32: Site DEEPCRK01

DEERCRO1, Mean nRMSE = 24.25%

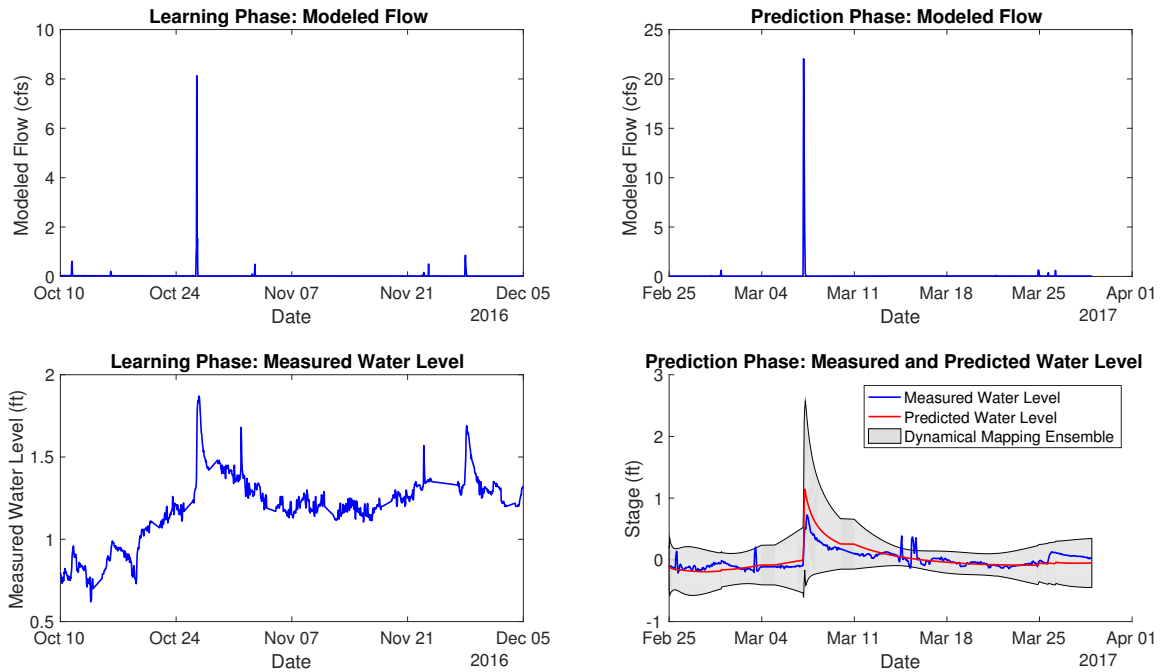


Figure B.33: Site DEERCRO1

DRNGDTCH01, Mean nRMSE = 34.06%

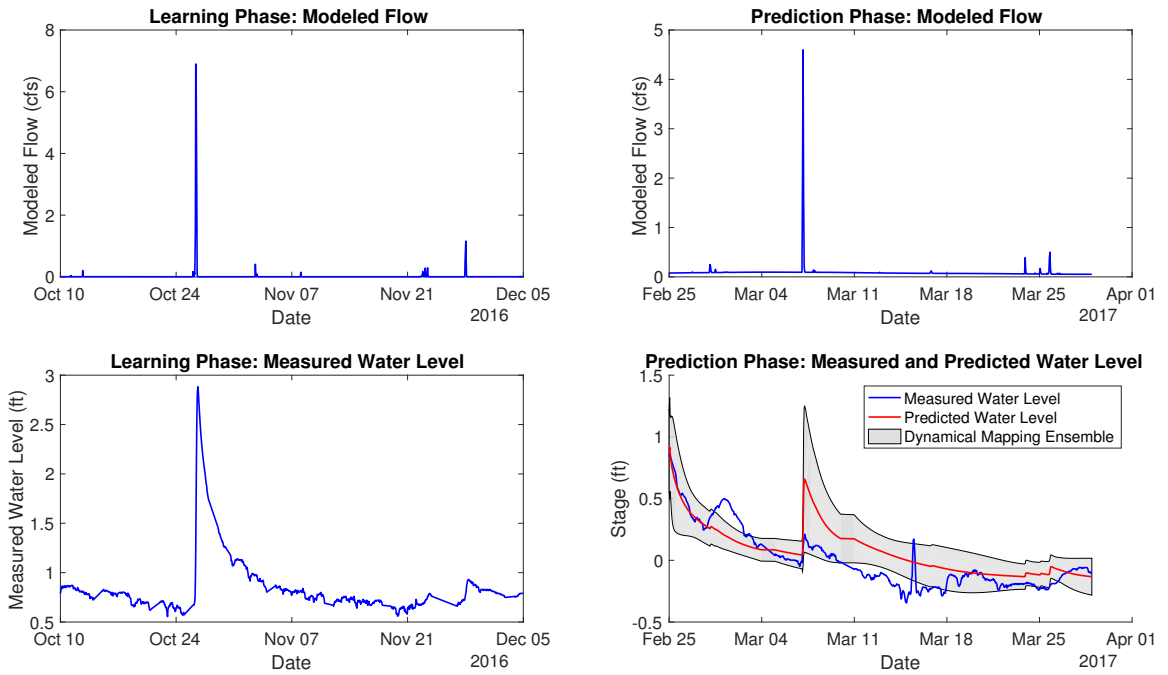


Figure B.34: Site DRNGDTCH01

DRYCR01, Mean nRMSE = 33.09%

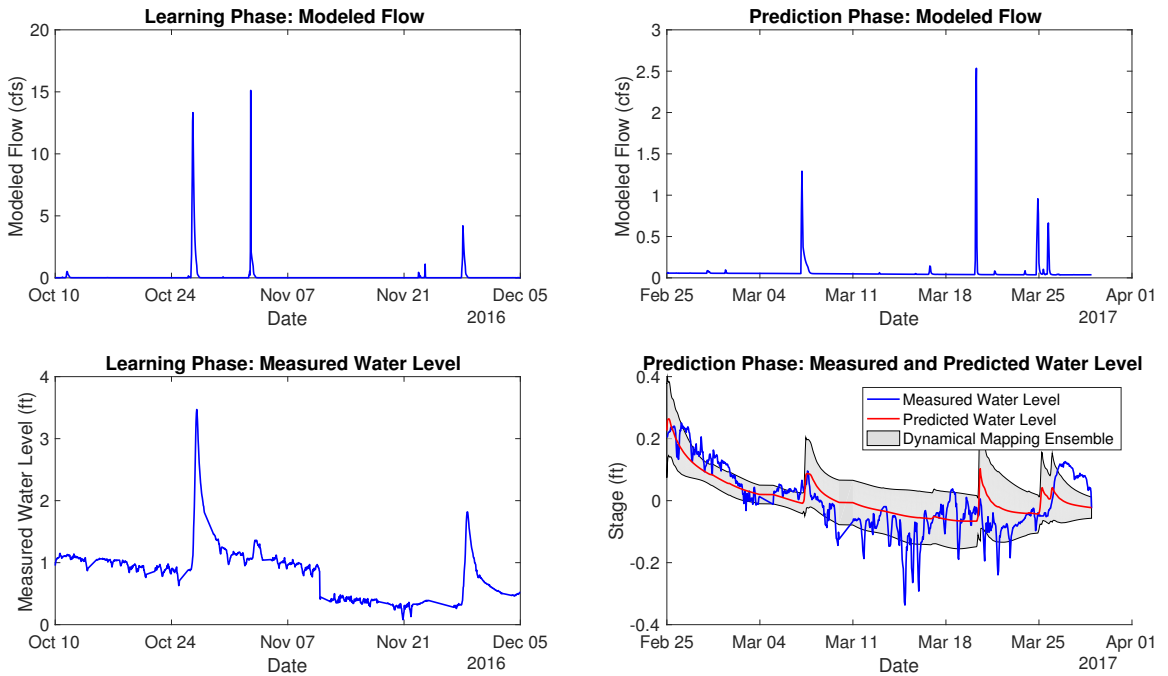


Figure B.35: Site DRYCR01

DRYCR02, Mean nRMSE = 10.16%

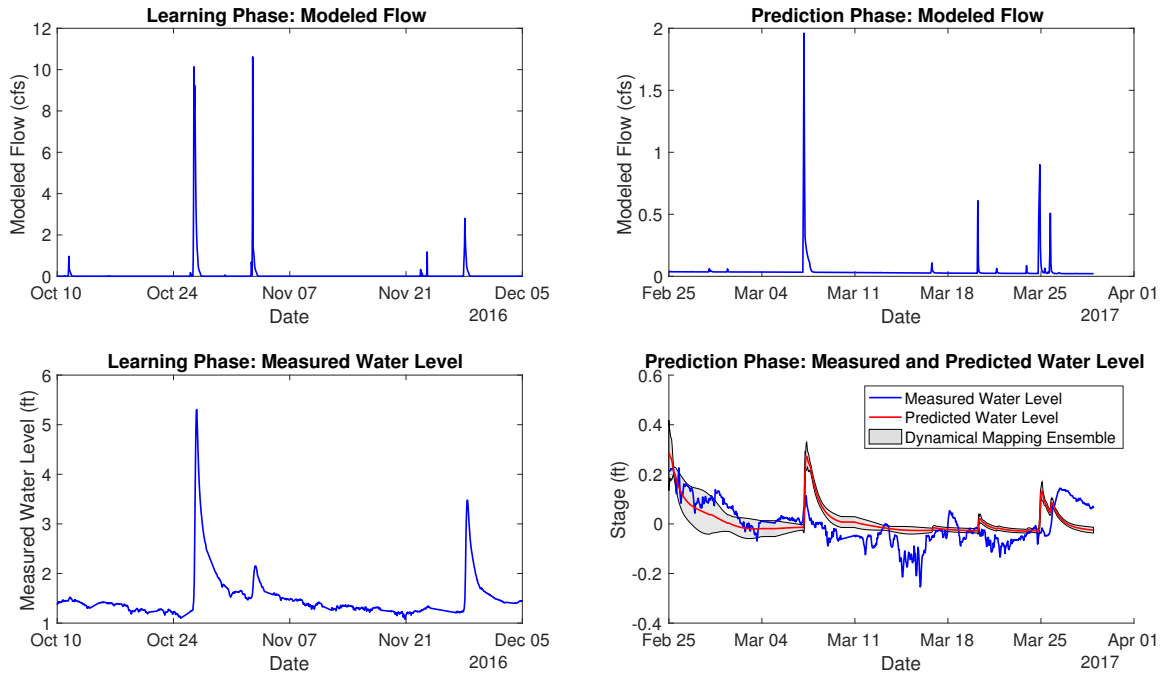


Figure B.36: Site DRYCR02

DRYCRK02, Mean nRMSE = 42.95%

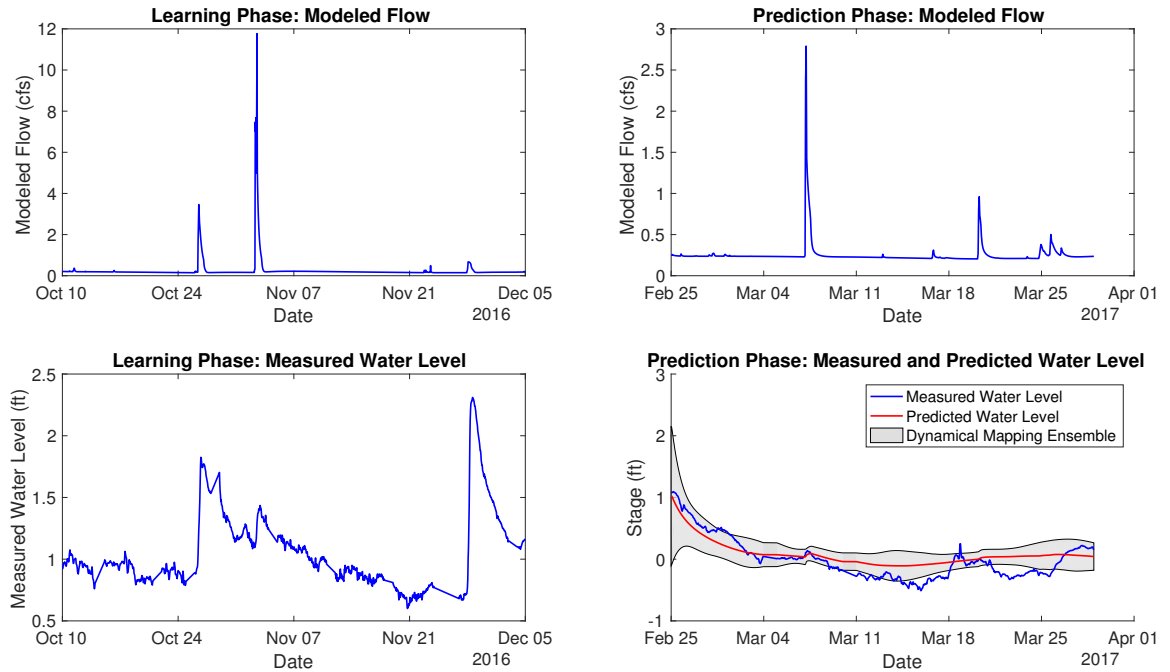


Figure B.37: Site DRYCRK02

DRYRNCRK01, Mean nRMSE = 32.58%

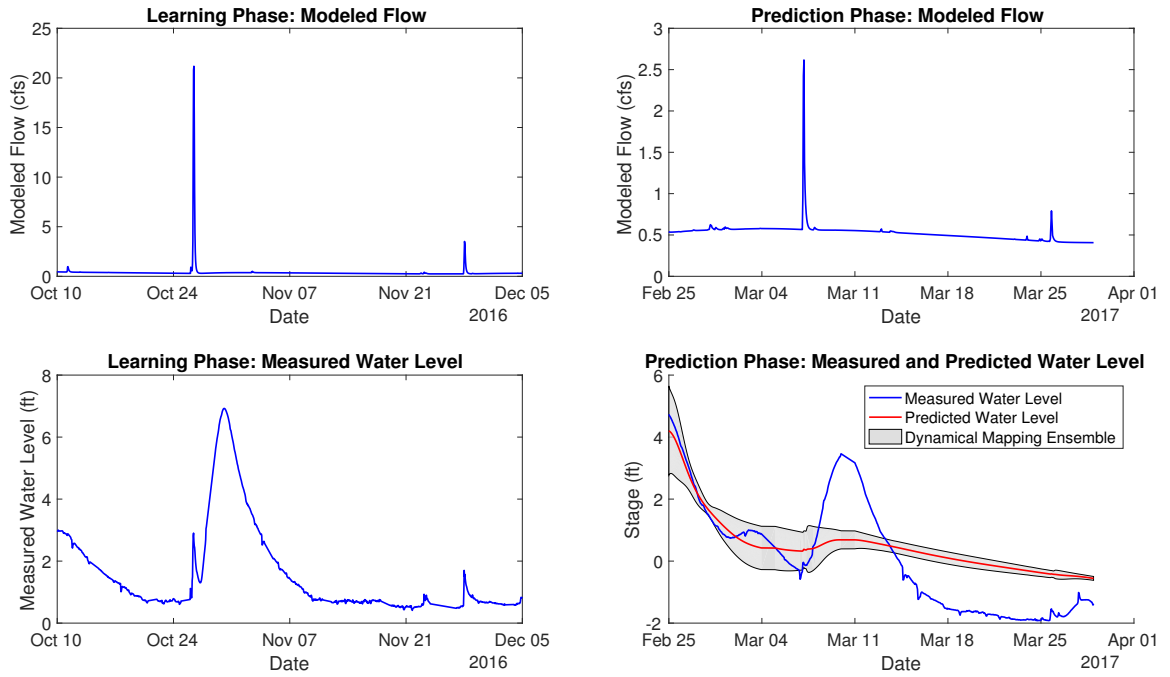


Figure B.38: Site DRYRNCRK01

DSMNSRV01, Mean nRMSE = 75.58%

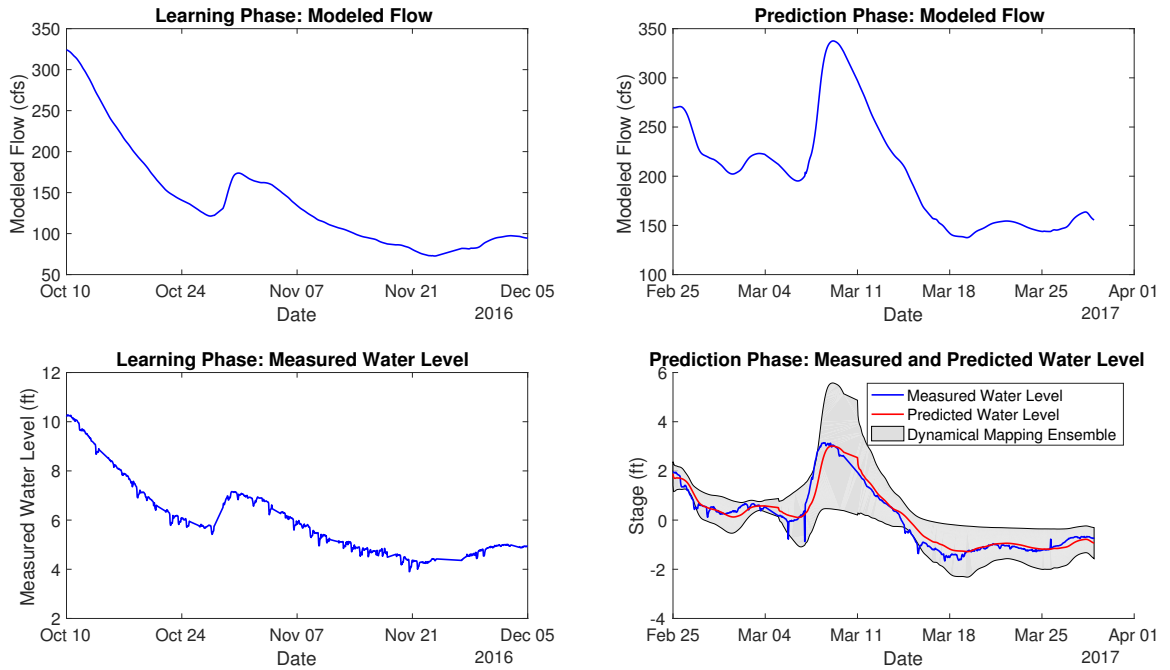


Figure B.39: Site DSMNSRV01

DSMNSRV03, Mean nRMSE = 61.63%

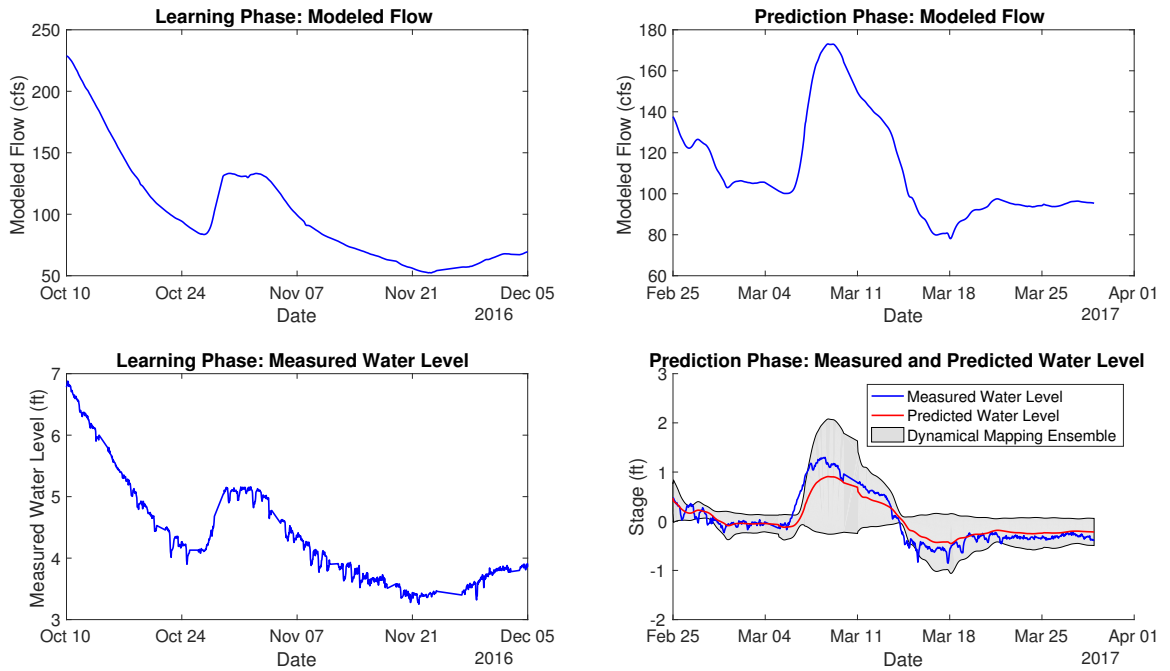


Figure B.40: Site DSMNSRV03

DSMNSRV04, Mean nRMSE = 87.21%

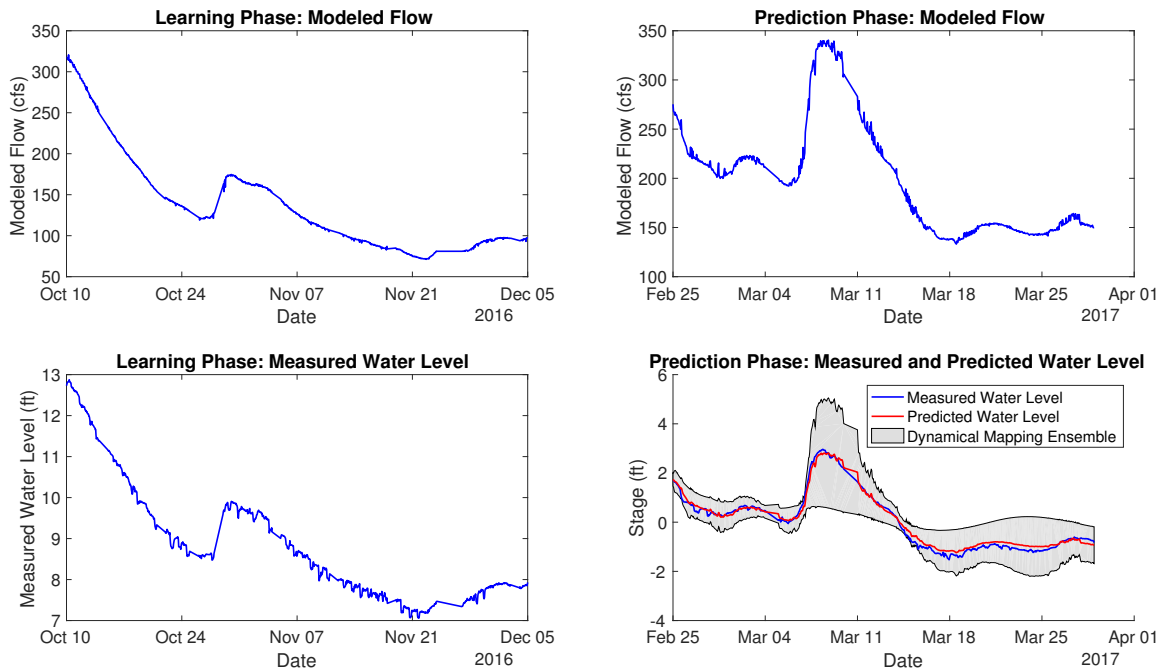


Figure B.41: Site DSMNSRV04

DUCKCR01, Mean nRMSE = 17.63%

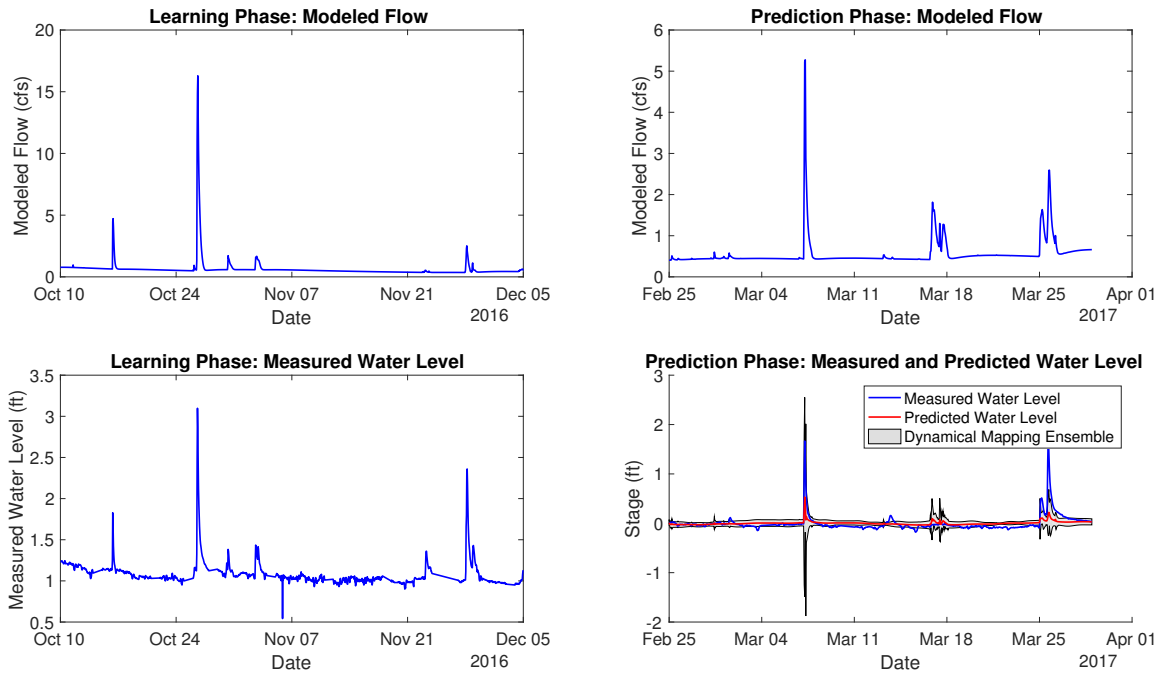


Figure B.42: Site DUCKCR01

EFDSMNS01, Mean nRMSE = 50.79%

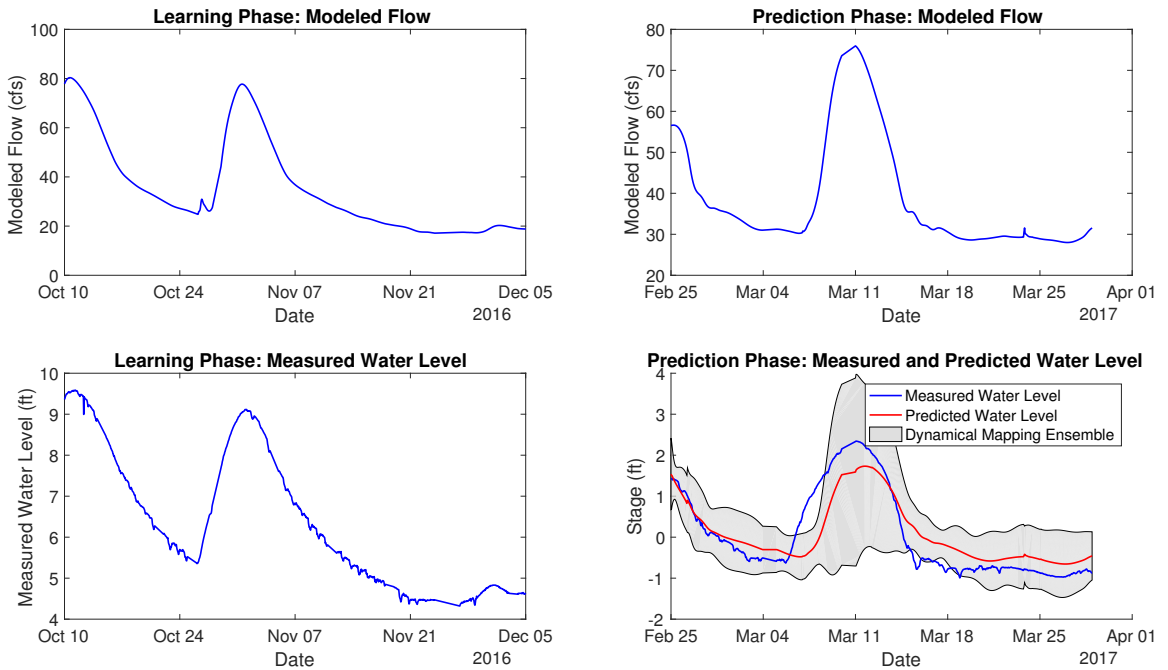


Figure B.43: Site EFDSMNS01

EINDNCR01, Mean nRMSE = 57.68%

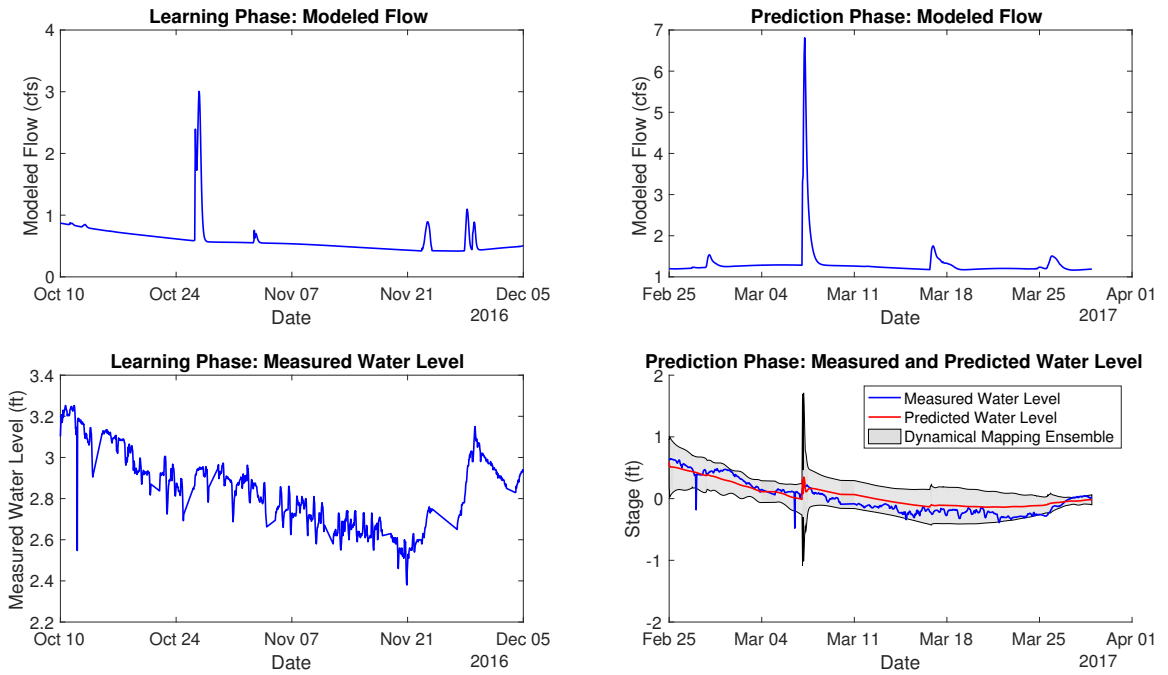


Figure B.44: Site EINDNCR01

ELKRN01, Mean nRMSE = -37.83%

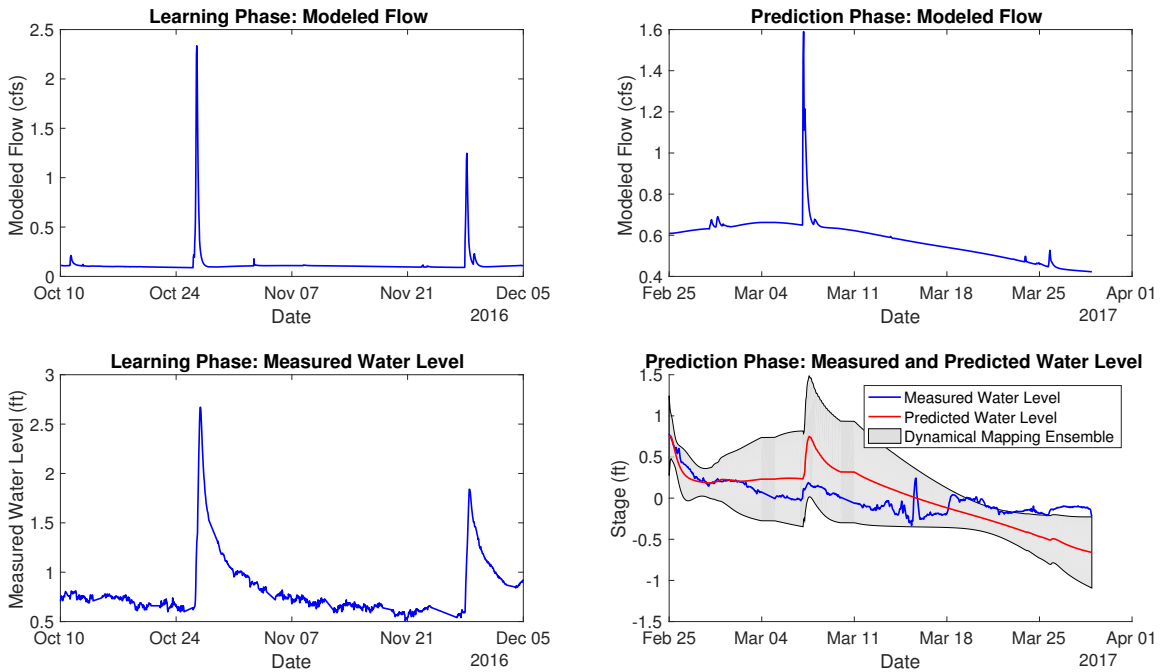


Figure B.45: Site ELKRN01

ENDWY01, Mean nRMSE = 4.42%

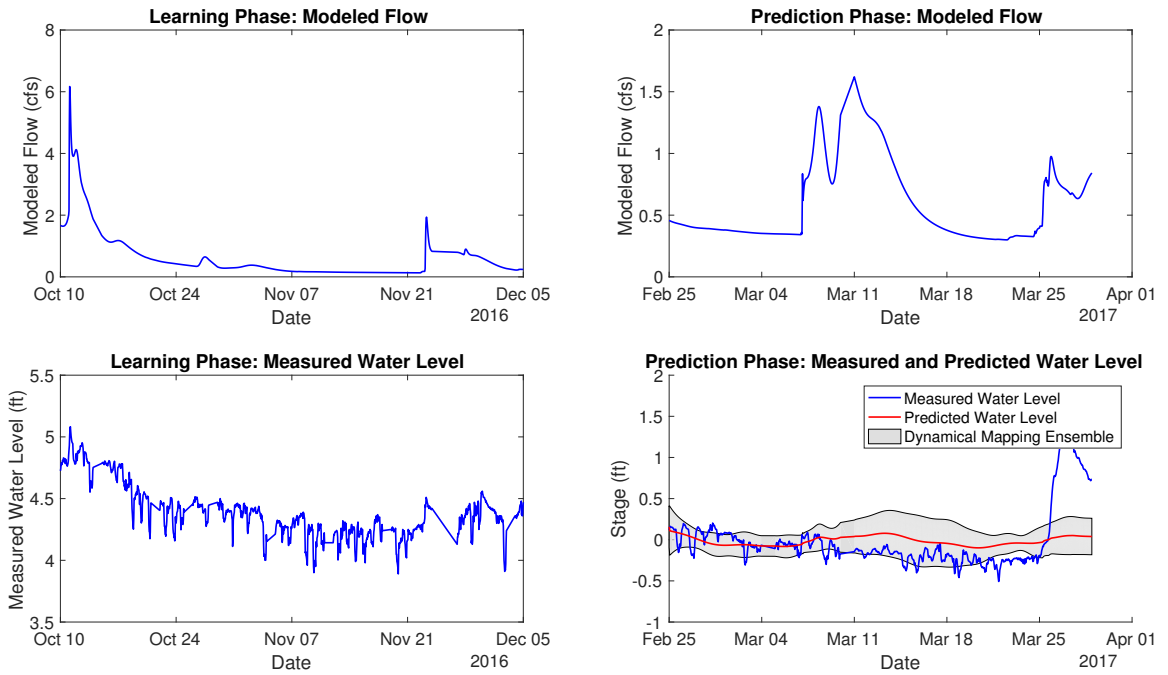


Figure B.46: Site ENDWY01

ENISH01, Mean nRMSE = 26.28%

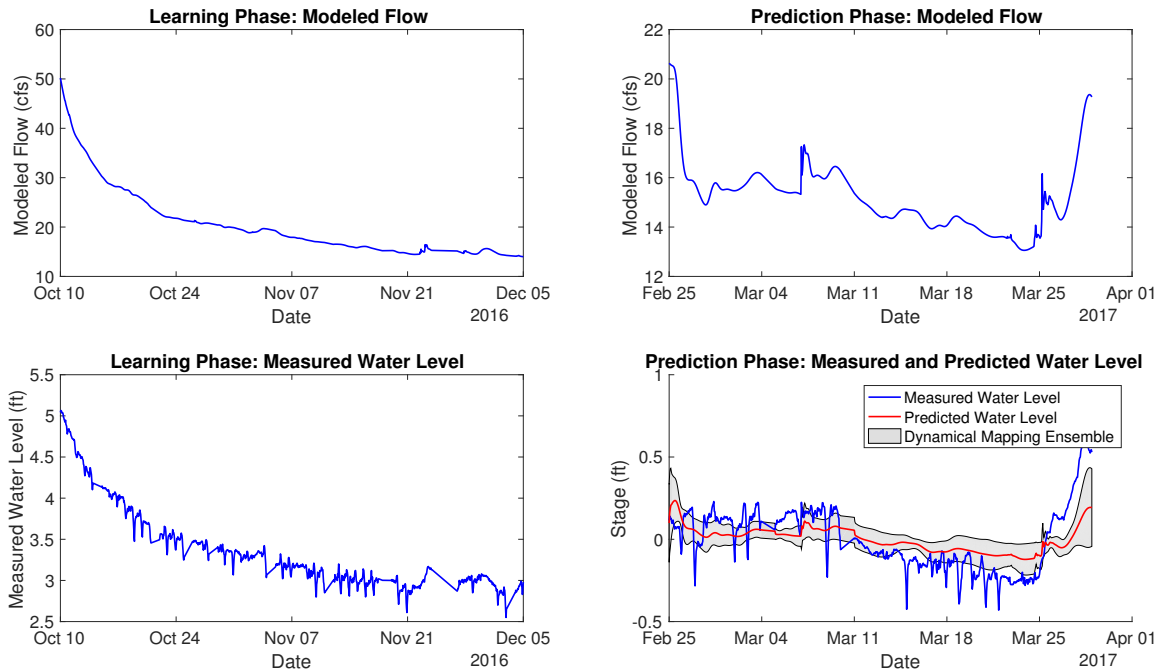


Figure B.47: Site ENISH01

ENISH02, Mean nRMSE = 36.28%

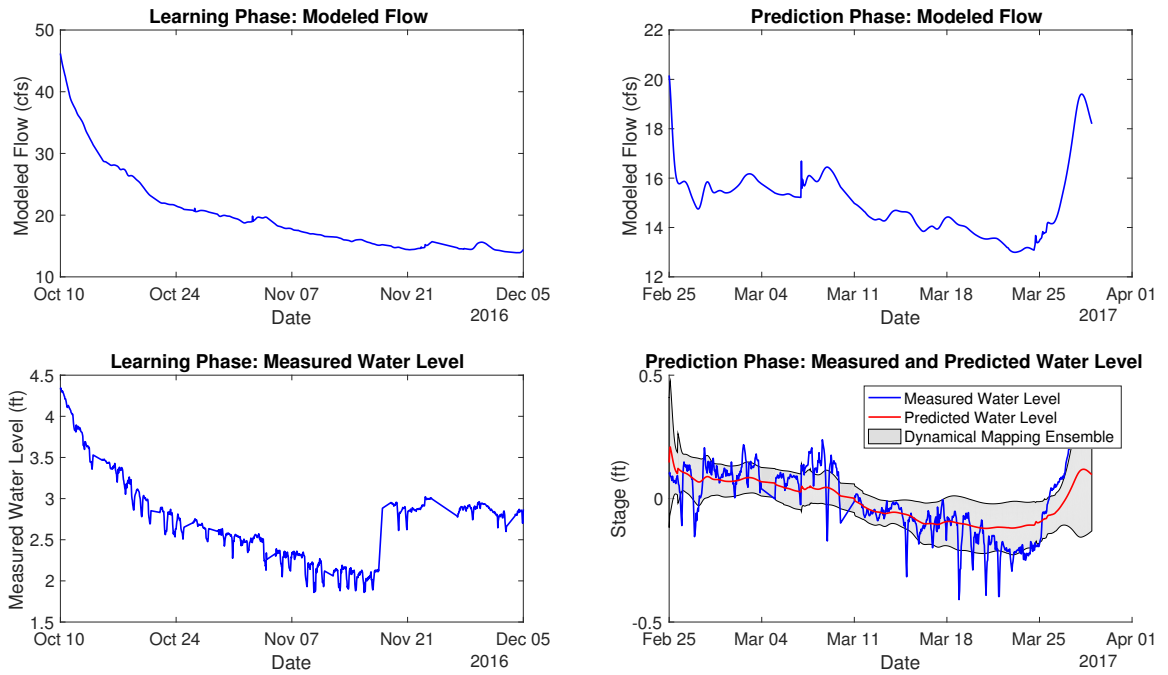


Figure B.48: Site ENISH02

ENISH03, Mean nRMSE = 19.78%

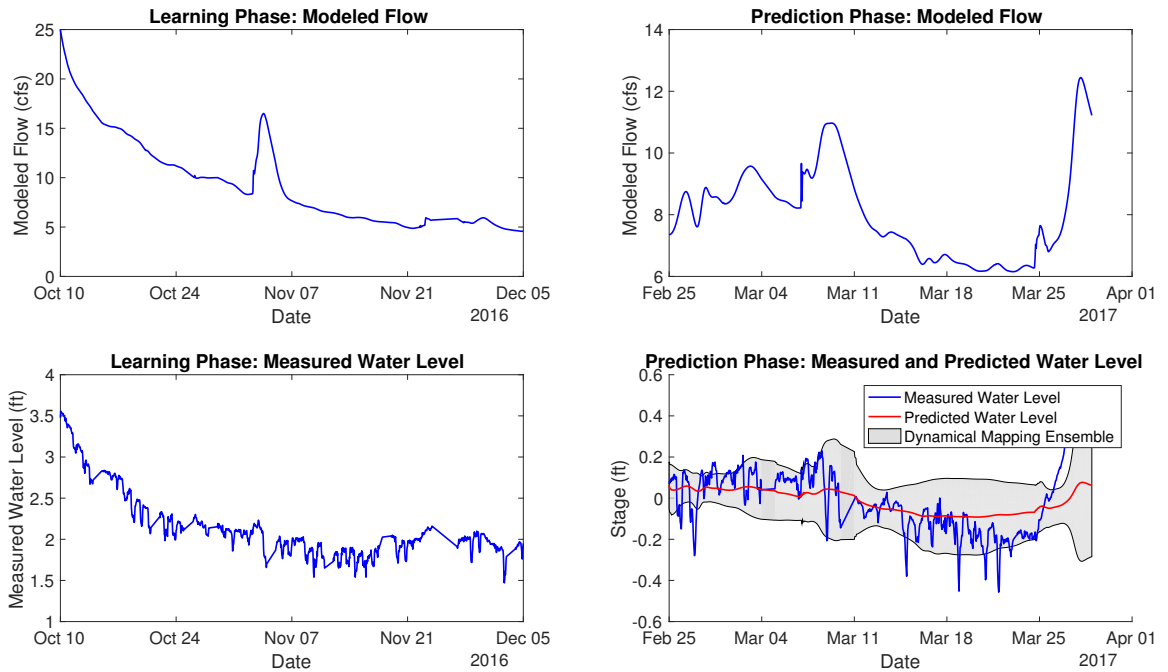


Figure B.49: Site ENISH03

ENISH04, Mean nRMSE = 19.26%

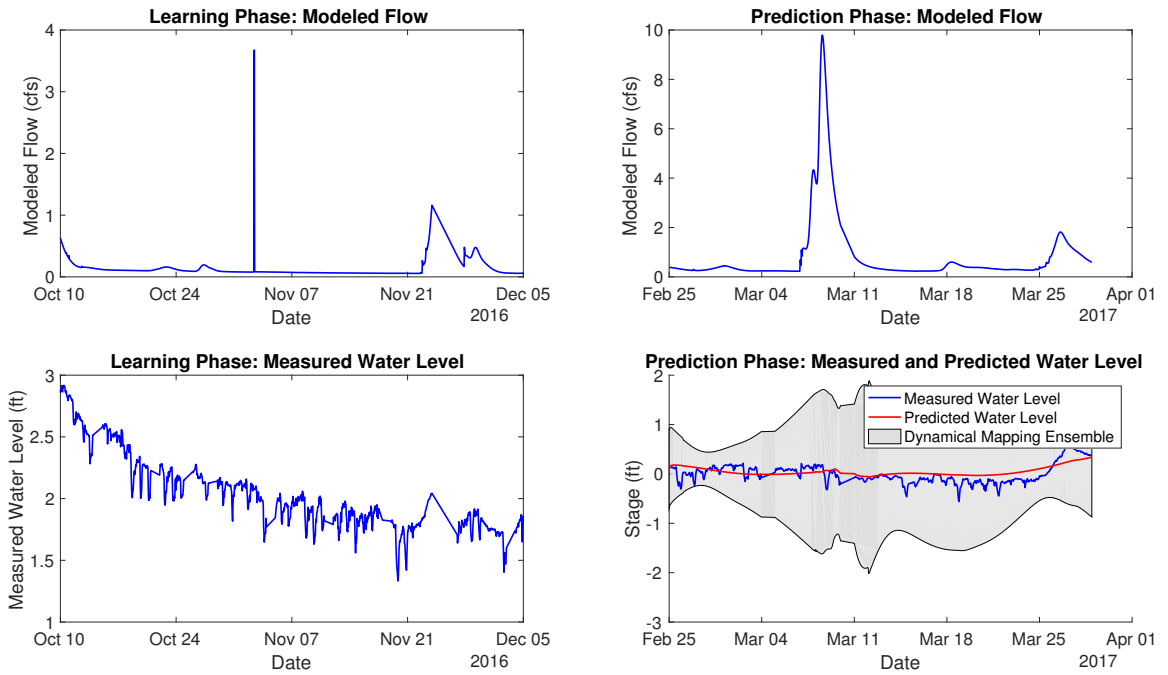


Figure B.50: Site ENISH04

FLOYDRV02, Mean nRMSE = -125.02%

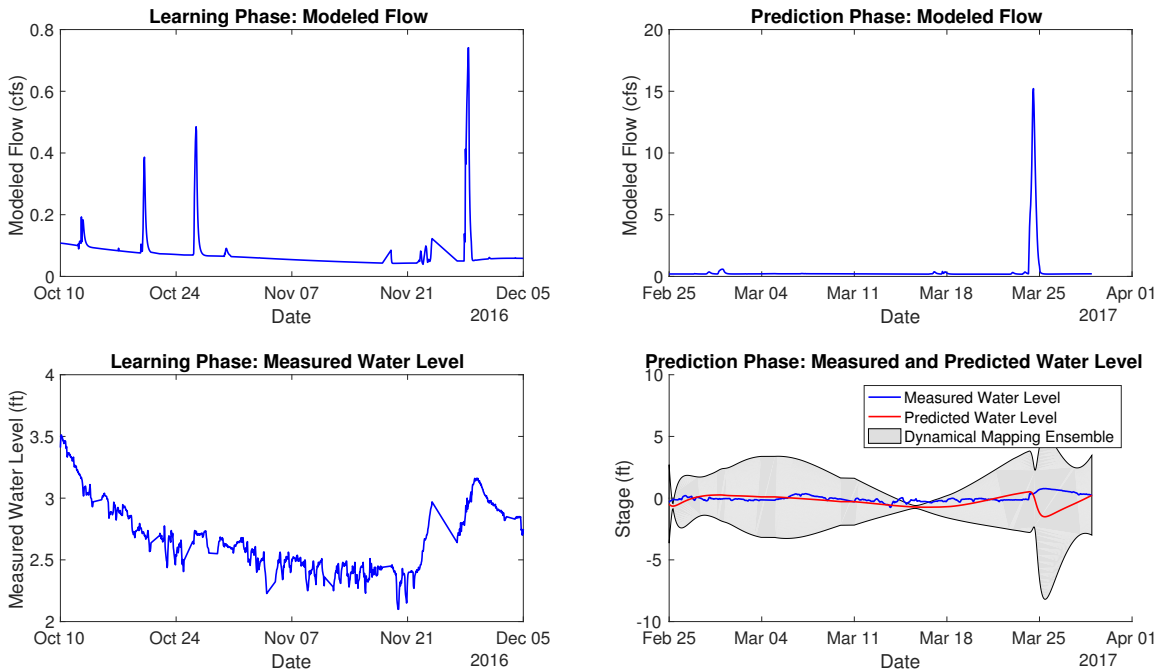


Figure B.51: Site FLOYDRV02

FOURMLE01, Mean nRMSE = 34.26%

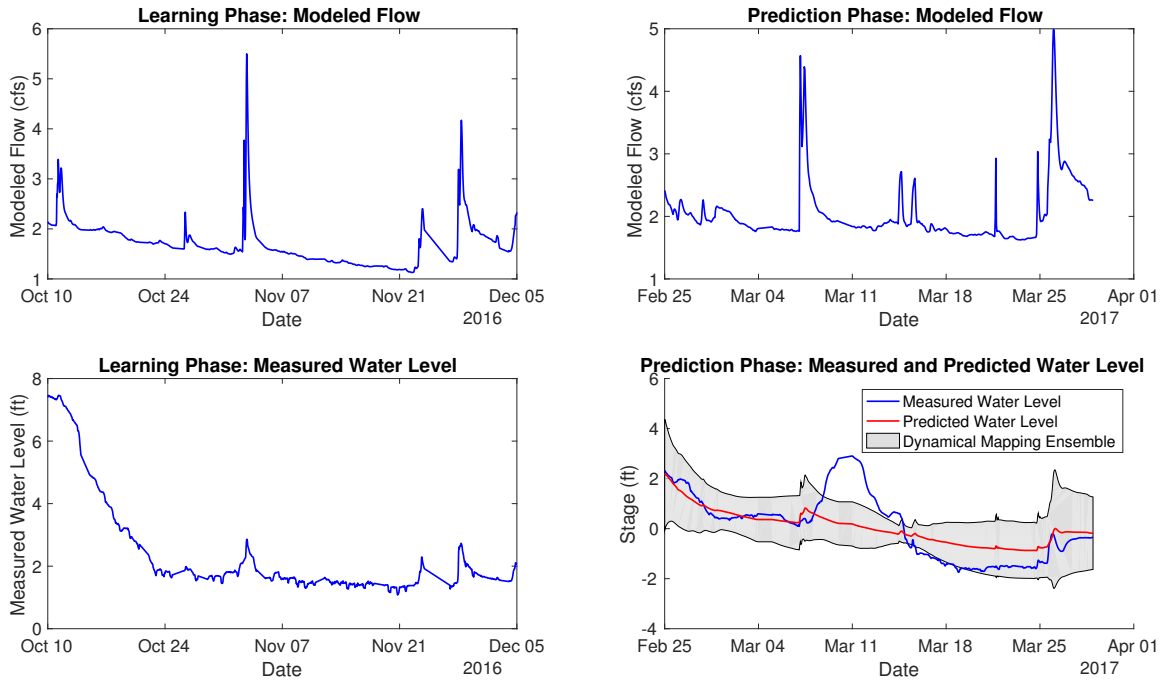


Figure B.52: Site FOURMLE01

FOURMLE02, Mean nRMSE = 55.71%

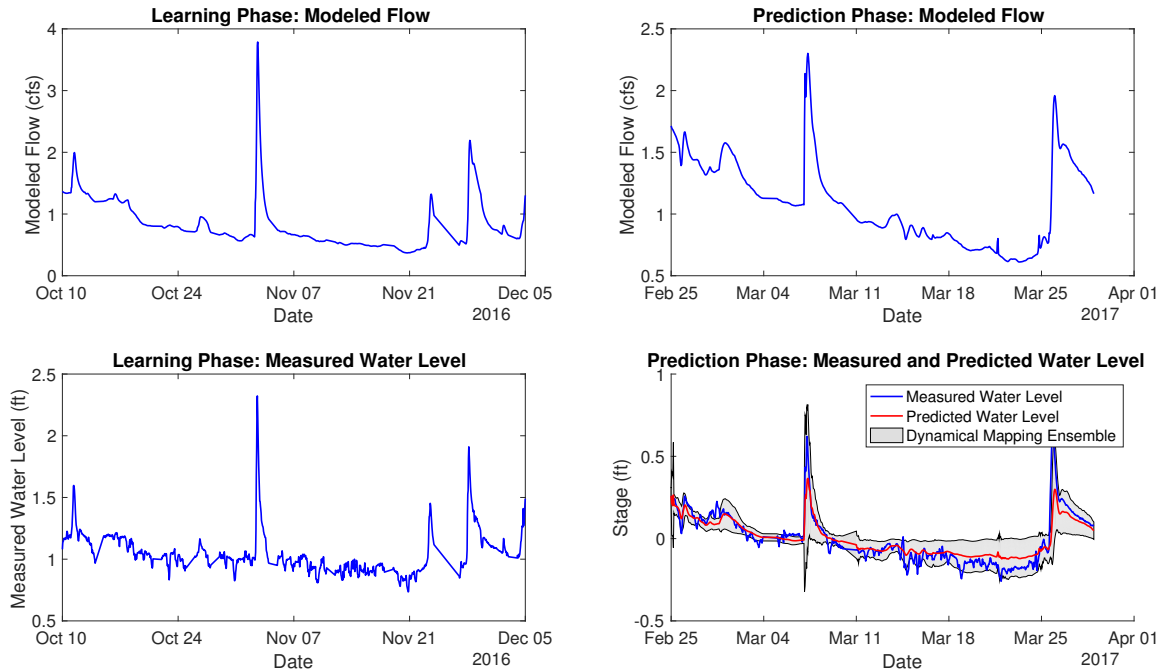


Figure B.53: Site FOURMLE02

FOURMLE03, Mean nRMSE = 35.13%

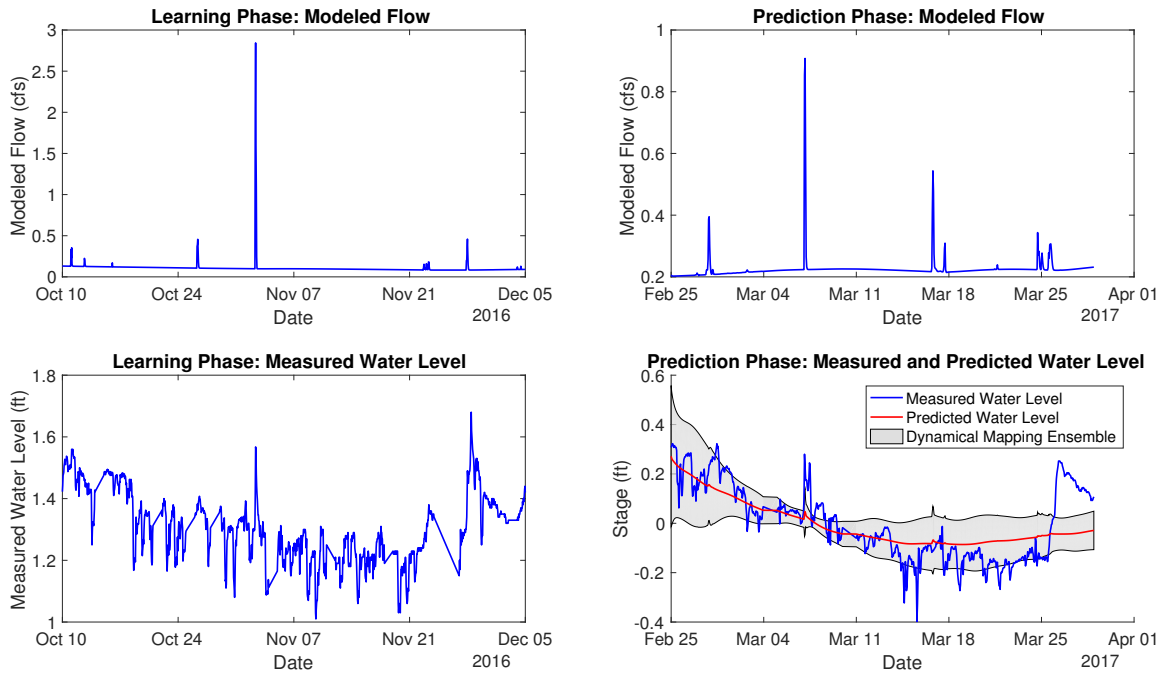


Figure B.54: Site FOURMLE03

GENEVACR01, Mean nRMSE = 8.62%

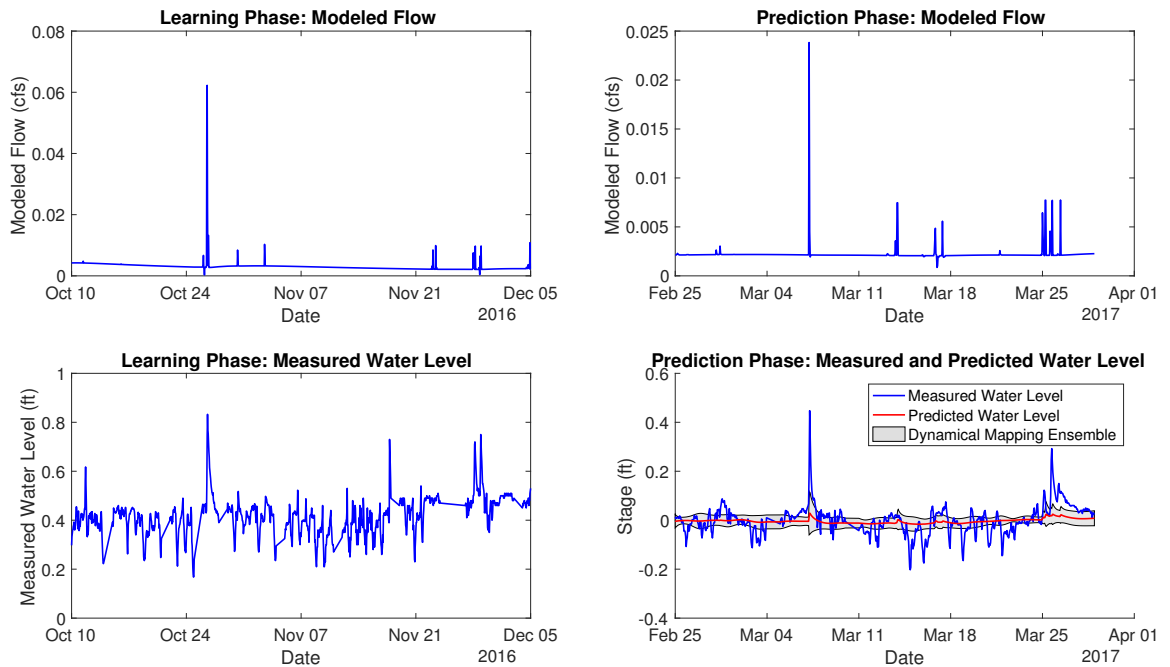


Figure B.55: Site GENEVACR01

GRANCRK01, Mean nRMSE = 44.53%

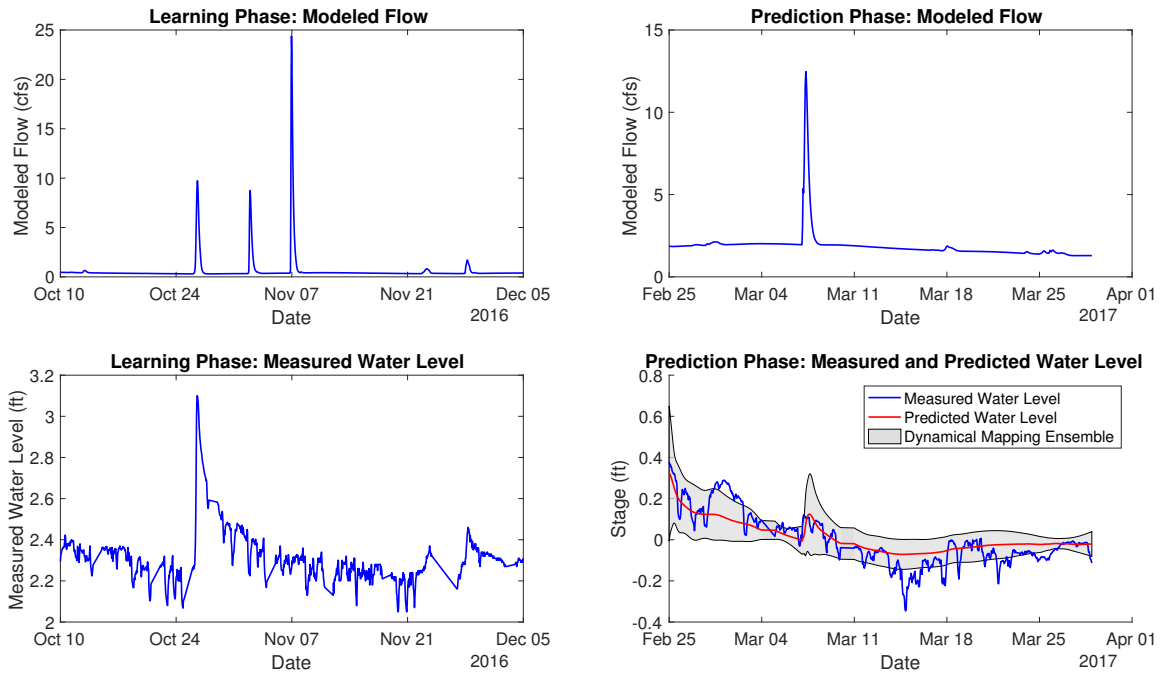


Figure B.56: Site GRANCRK01

GRNGRCR01, Mean nRMSE = 25.80%

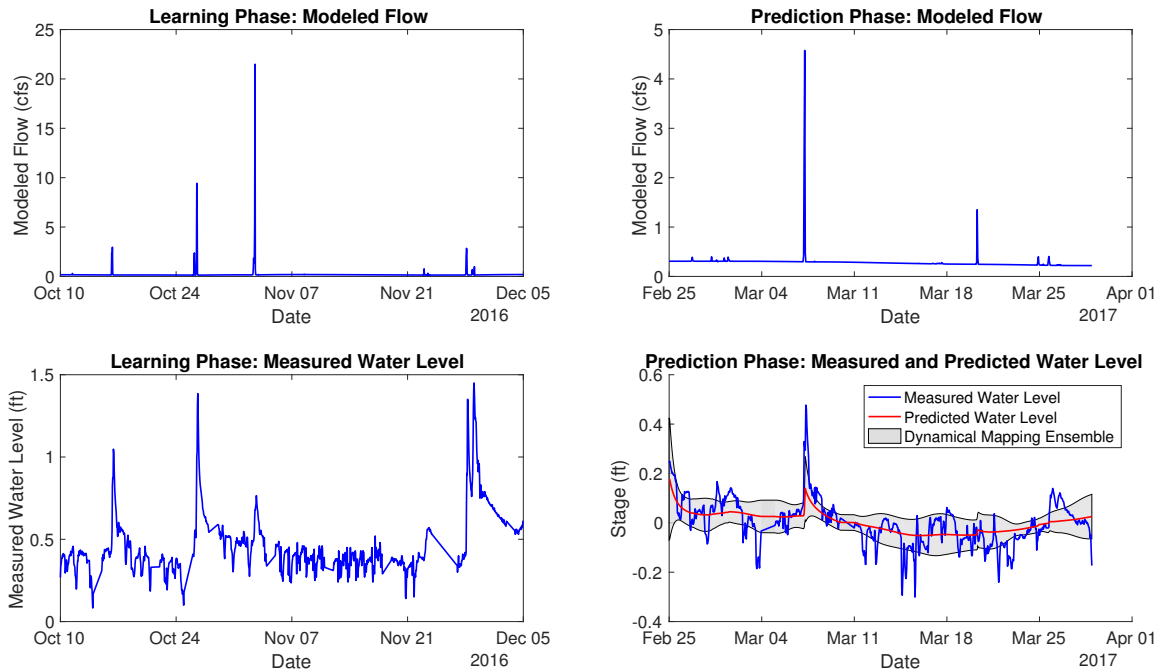


Figure B.57: Site GRNGRCR01

HOOSRCR01, Mean nRMSE = 6.16%

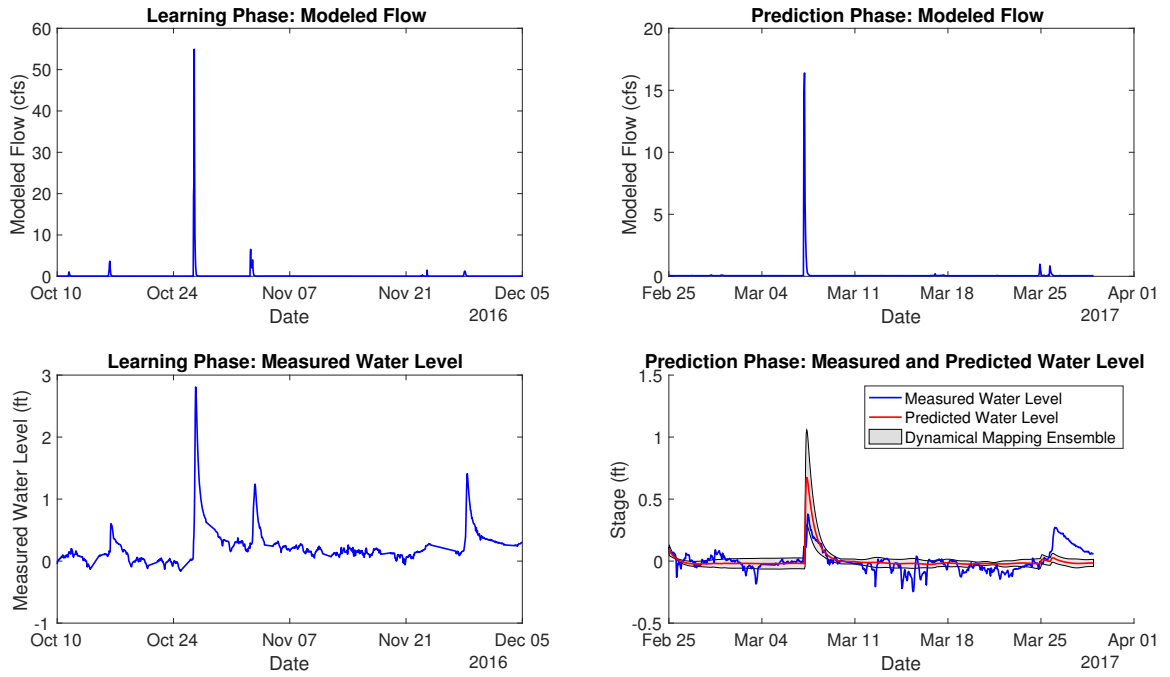


Figure B.58: Site HOOSRCR01

HRGVCR01, Mean nRMSE = 40.17%

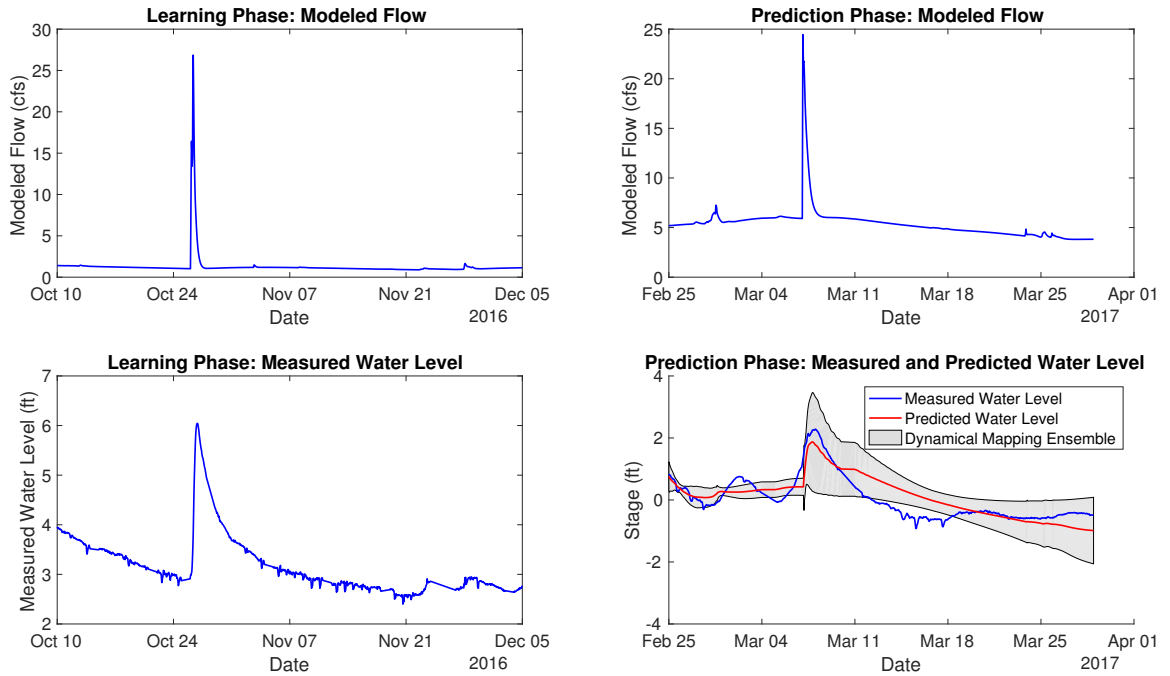


Figure B.59: Site HRGVCR01

INDCR01, Mean nRMSE = 21.79%

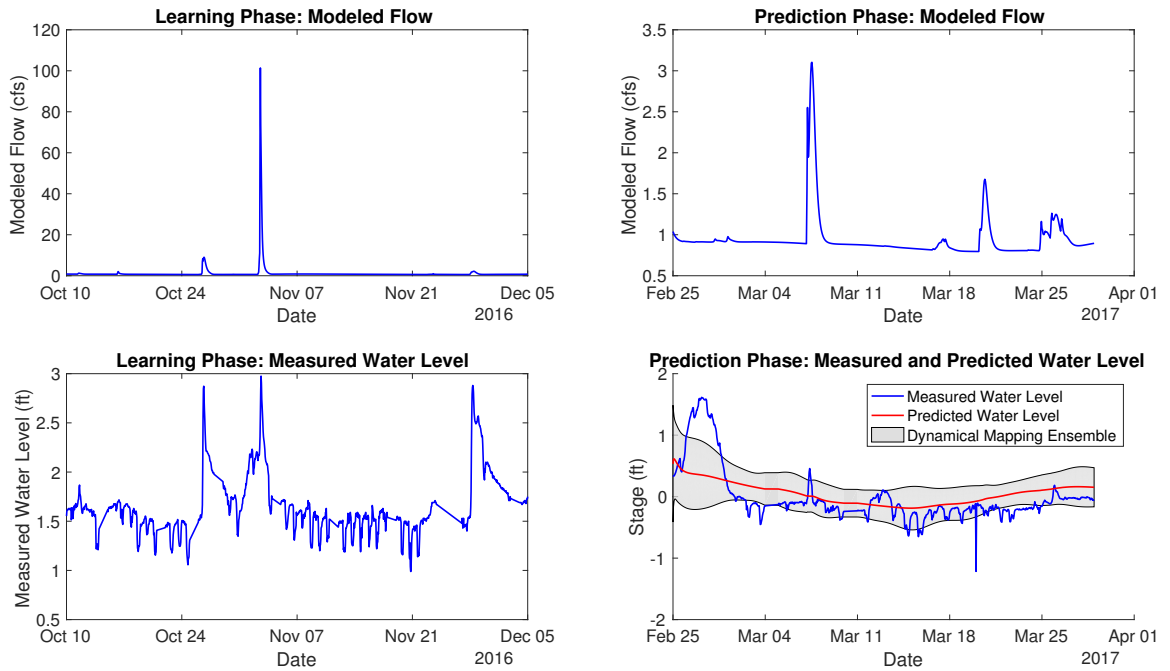


Figure B.60: Site INDCR01

INDCR04, Mean nRMSE = 32.25%

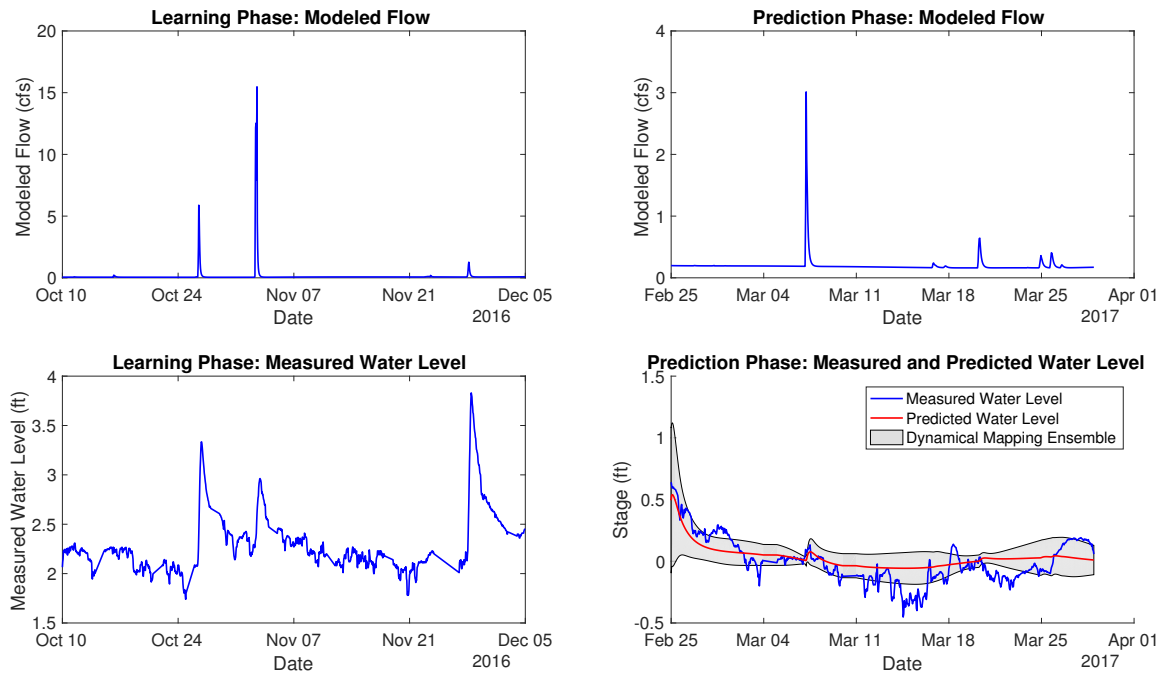


Figure B.61: Site INDCR04

INDIAN01, Mean nRMSE = -1.96%

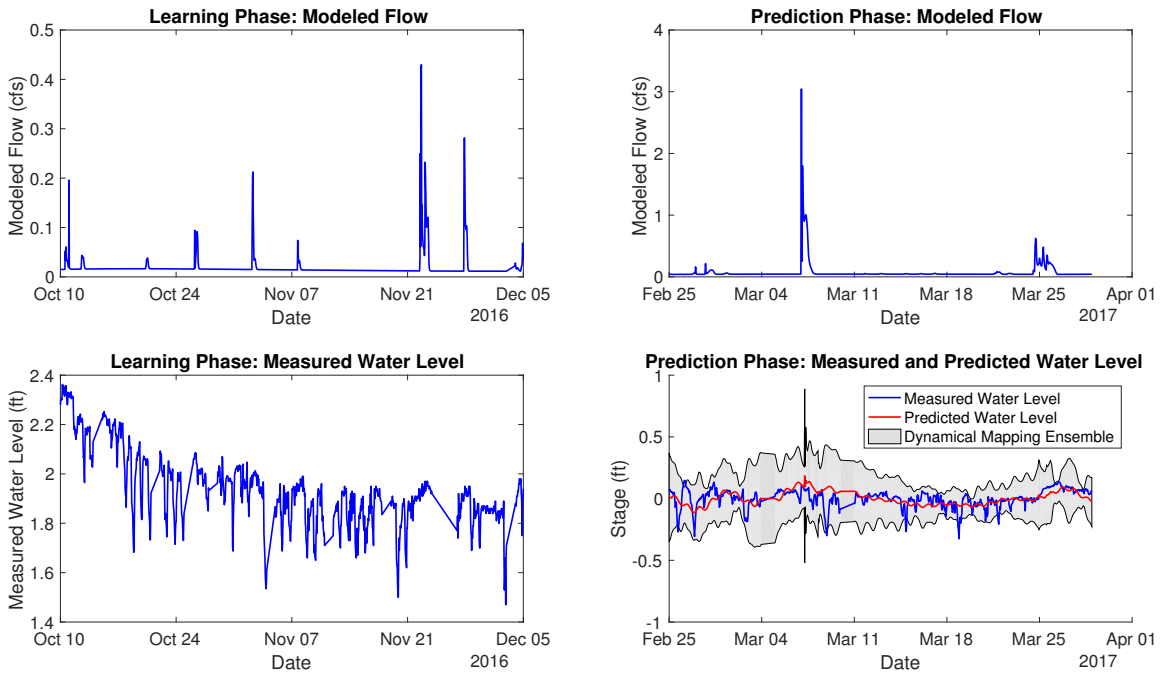


Figure B.62: Site INDIAN01

INDIAN02, Mean nRMSE = -7.18%

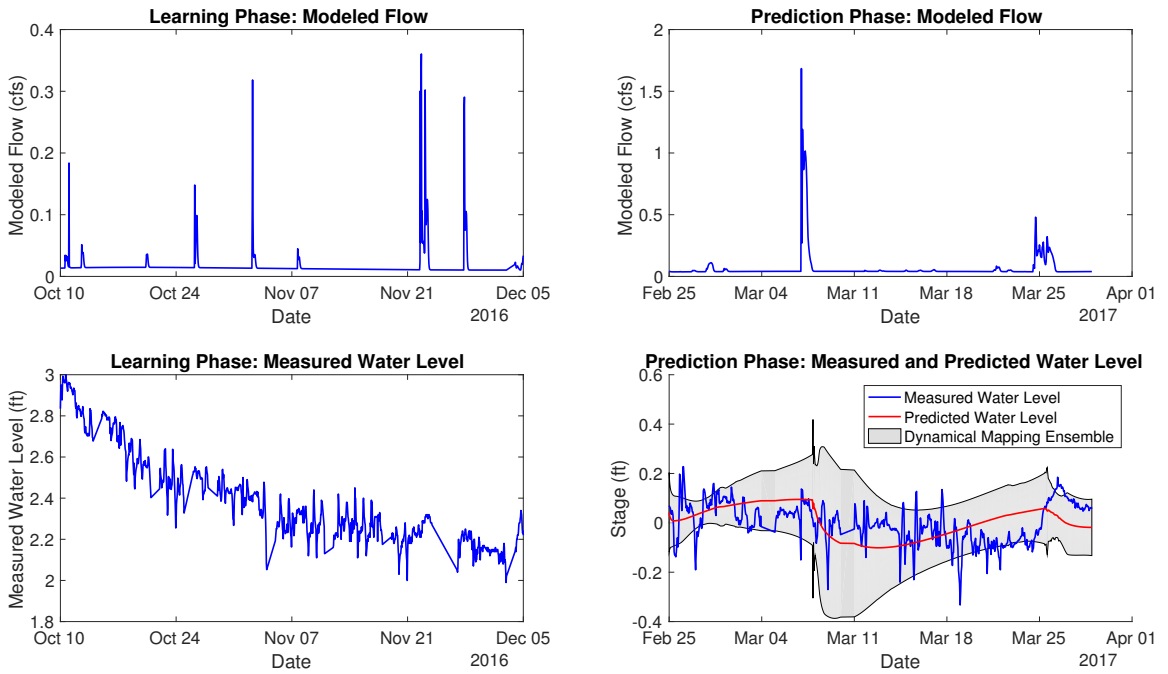


Figure B.63: Site INDIAN02

IOWARV01, Mean nRMSE = 69.02%

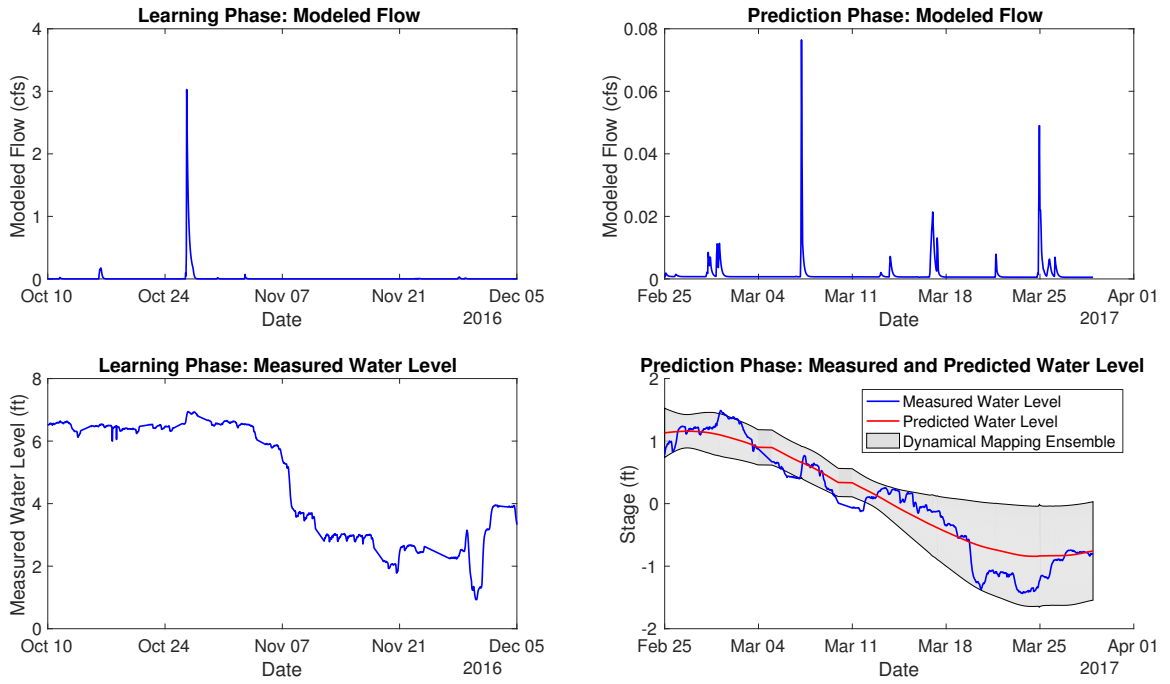


Figure B.64: Site IOWARV01

KEGCR01, Mean nRMSE = -69.92%

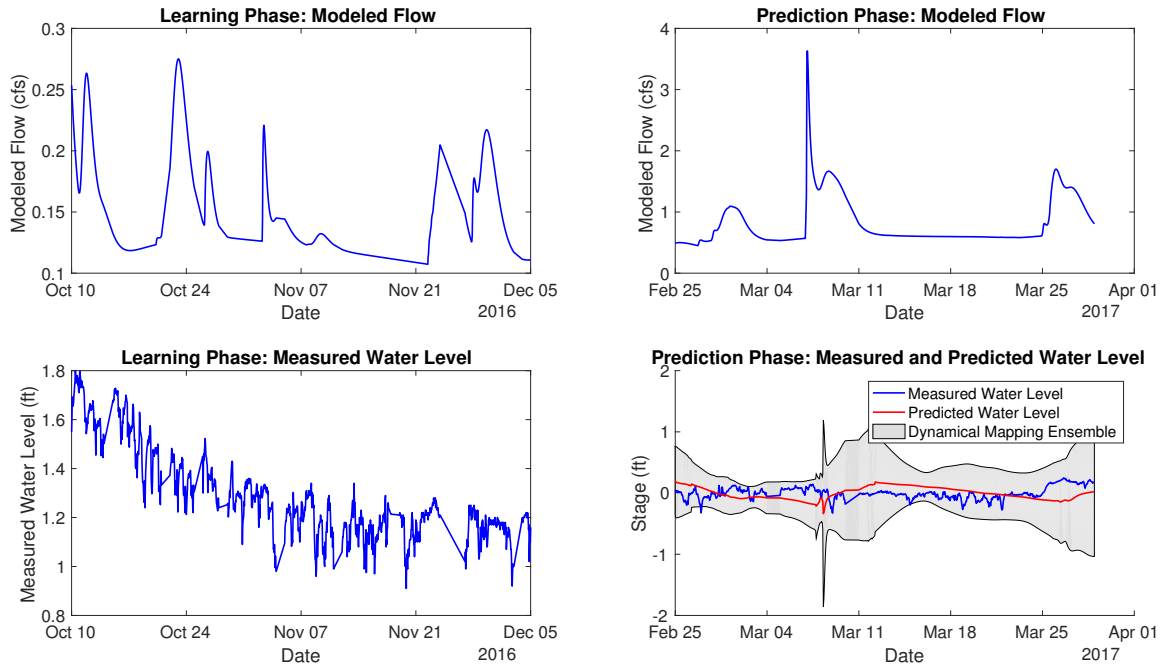


Figure B.65: Site KEGCR01

LAKECR01, Mean nRMSE = 23.51%

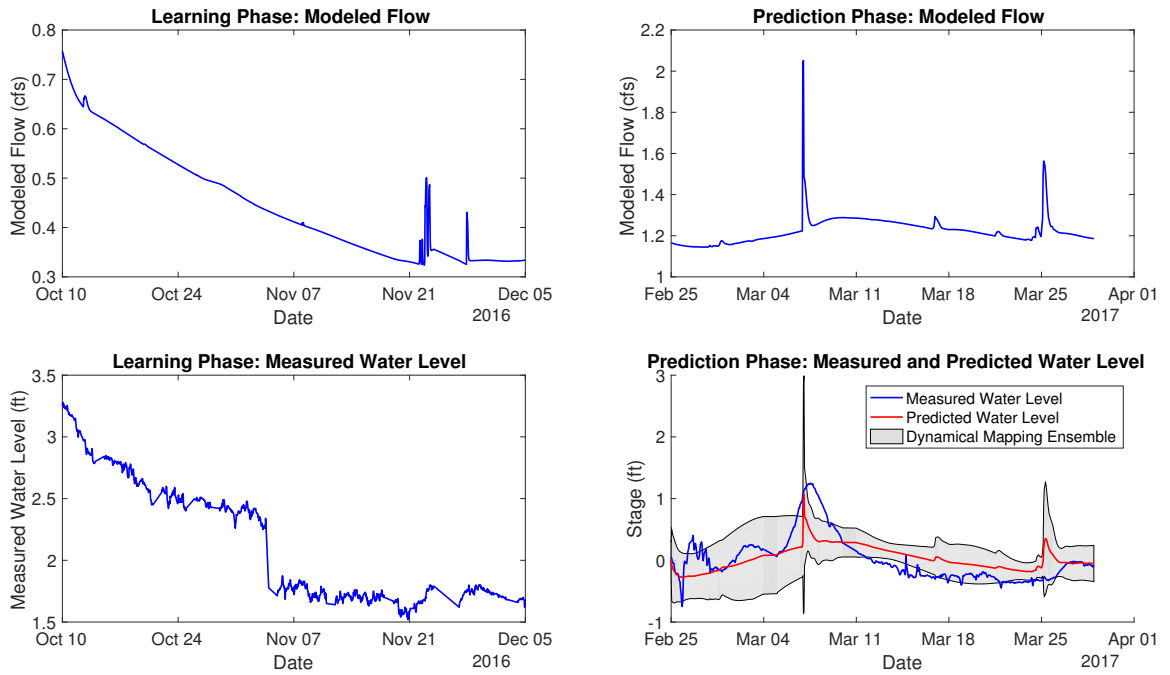


Figure B.66: Site LAKECR01

LTLCEDAR01, Mean nRMSE = 83.70%

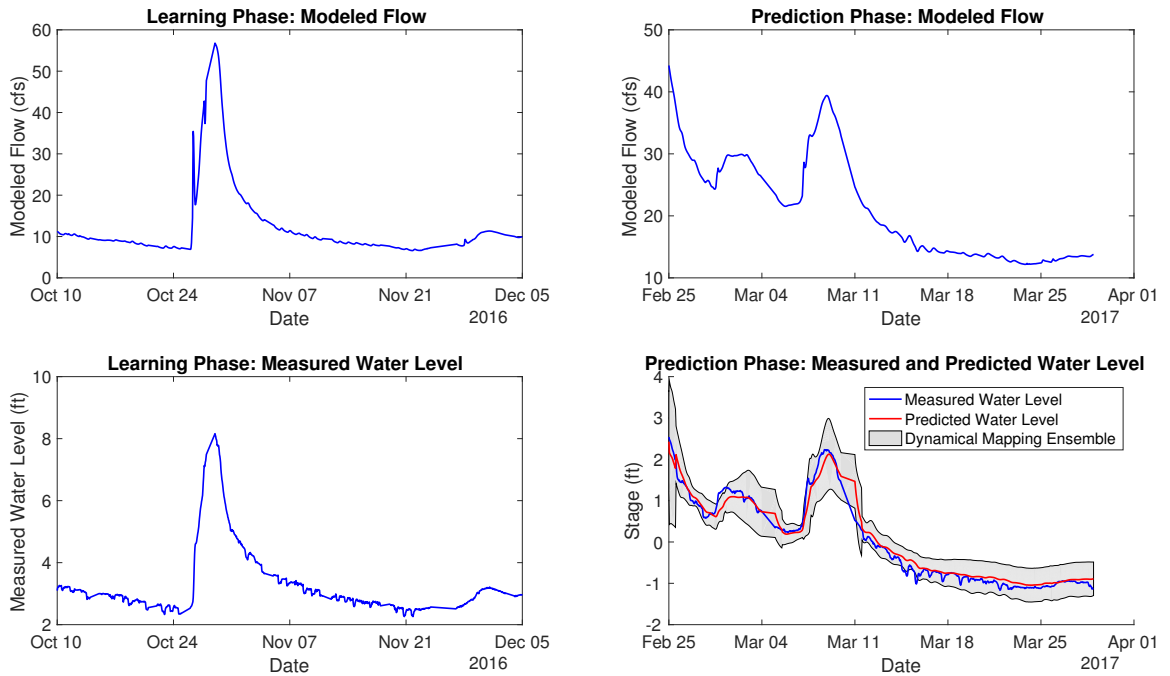


Figure B.67: Site LTLCEDAR01

LTLCEDAR02, Mean nRMSE = 38.40%

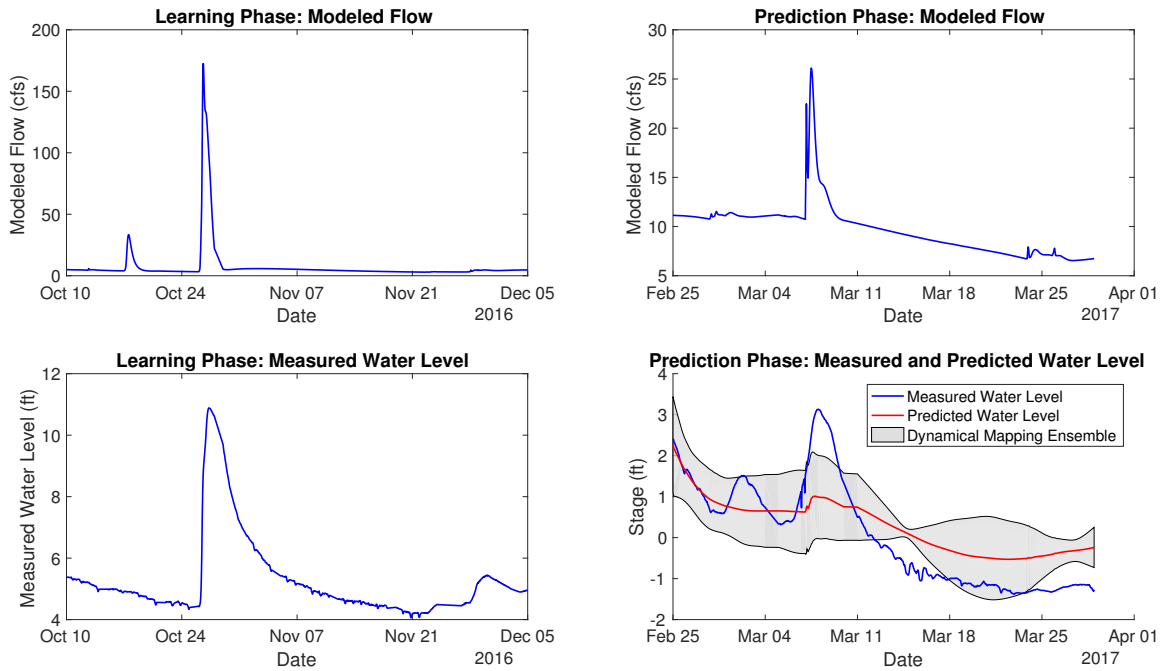


Figure B.68: Site LTLCEDAR02

LTLSTIOUX01, Mean nRMSE = 11.63%

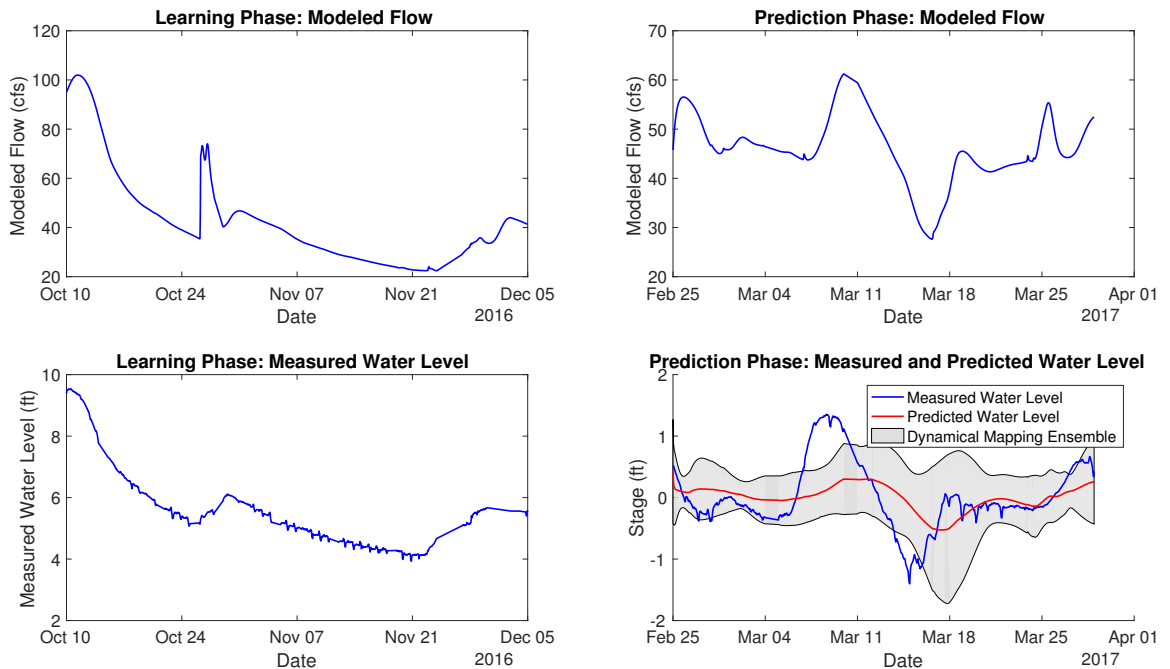


Figure B.69: Site LTLSTIOUX01

LTLSTIOUX02, Mean nRMSE = 6.55%

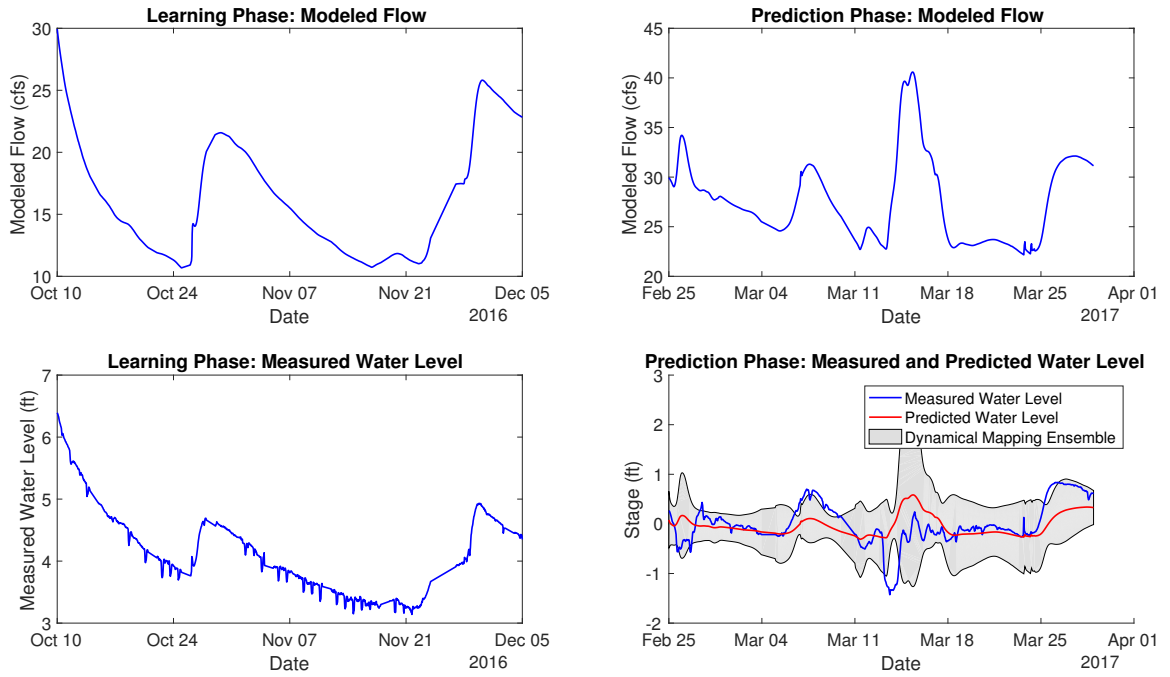


Figure B.70: Site LTLSTIOUX02

LTLSTIOUX03, Mean nRMSE = 23.88%

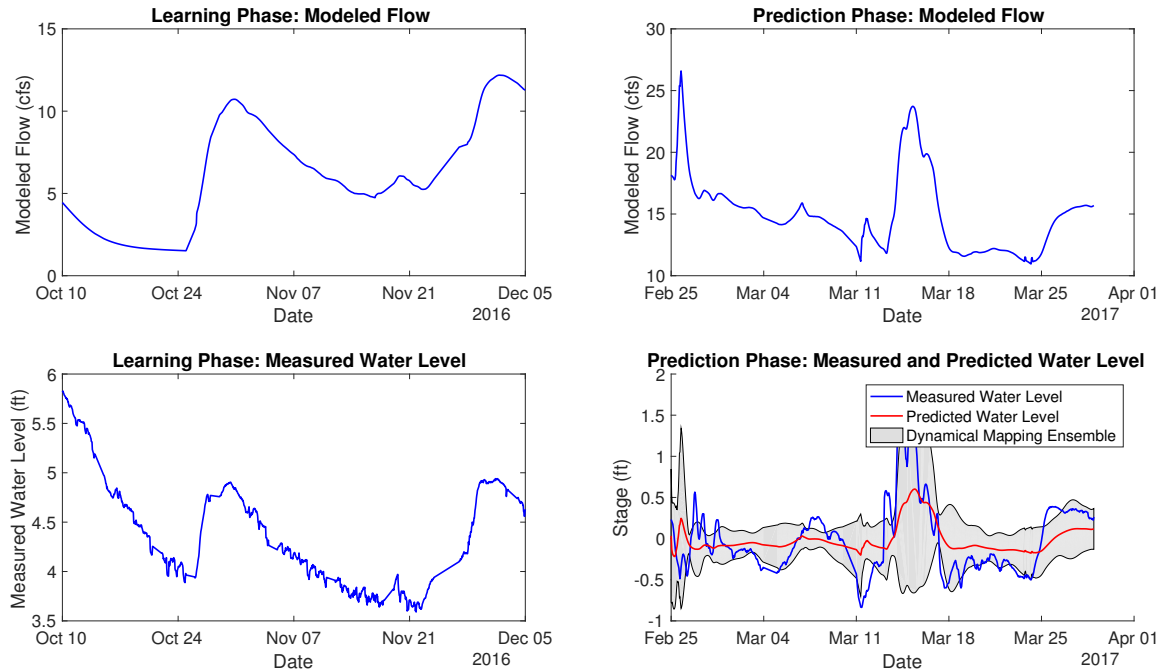


Figure B.71: Site LTLSTIOUX03

LTLSTIOUX04, Mean nRMSE = 17.24%

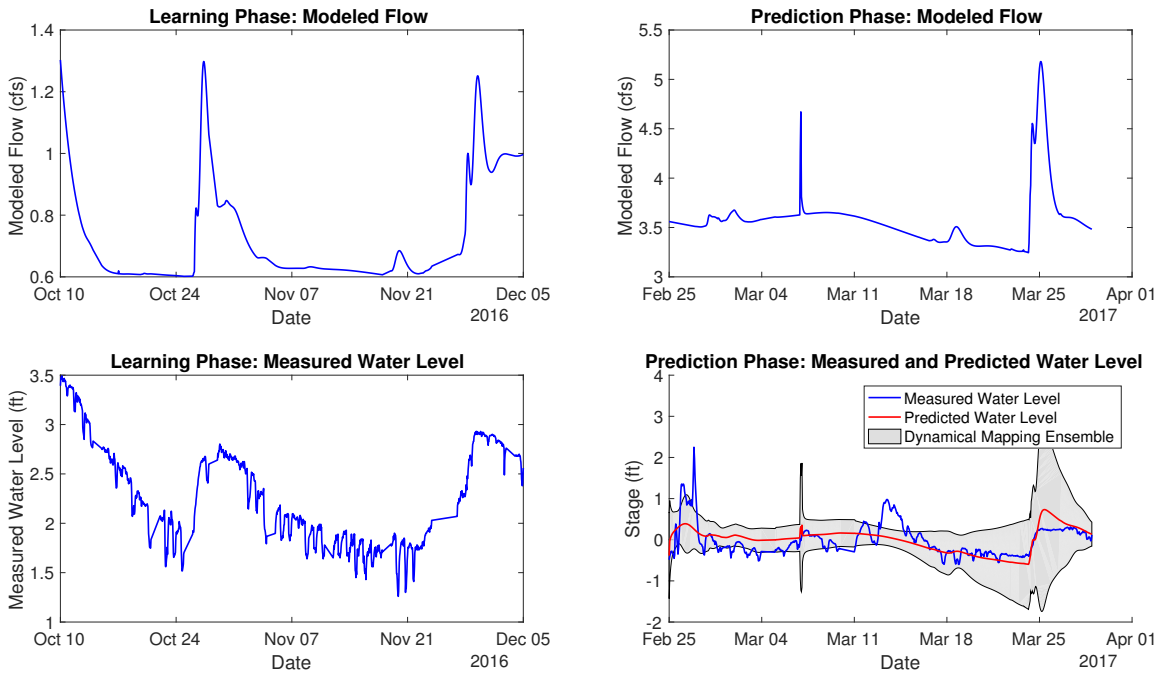


Figure B.72: Site LTLSTIOUX04

LTLSTIOUX05, Mean nRMSE = -6.43%

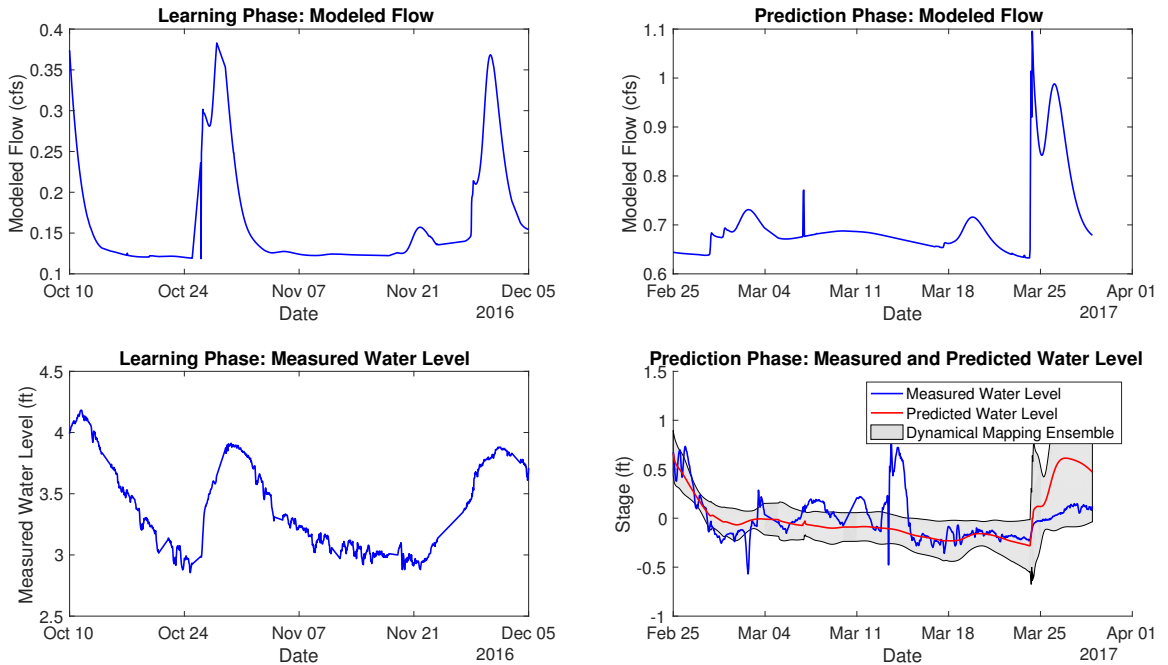


Figure B.73: Site LTLSTIOUX05

LTLSOAP01, Mean nRMSE = -79.39%

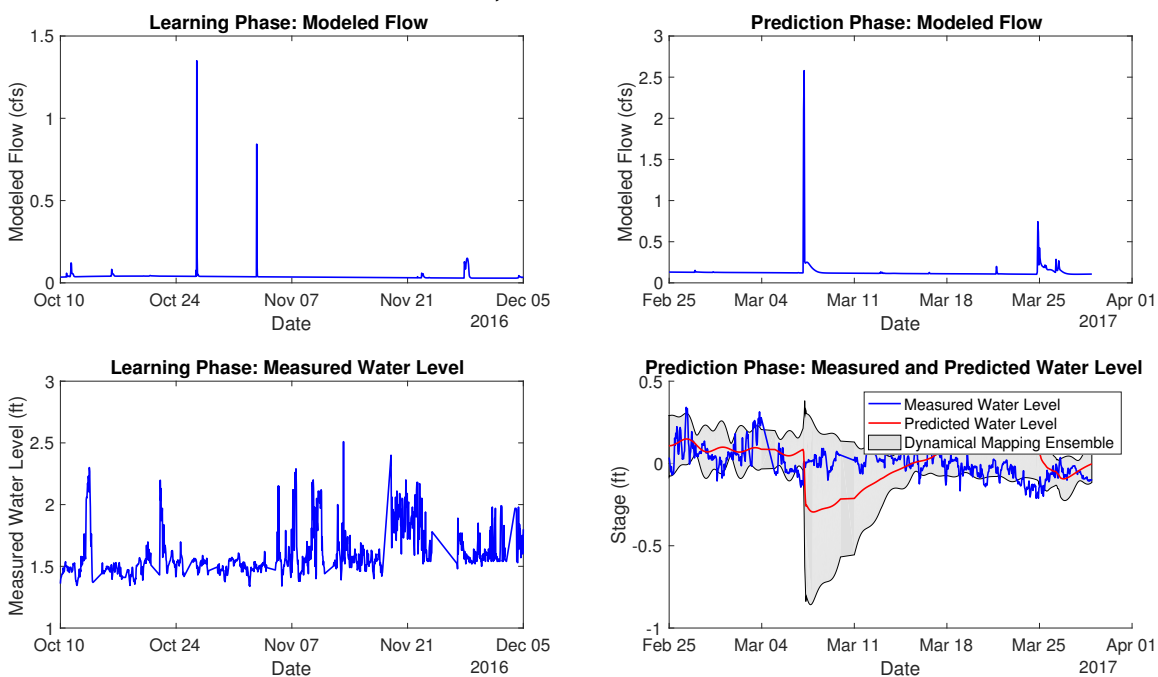


Figure B.74: Site LTLSOAP01

LTLTKY01, Mean nRMSE = 48.48%

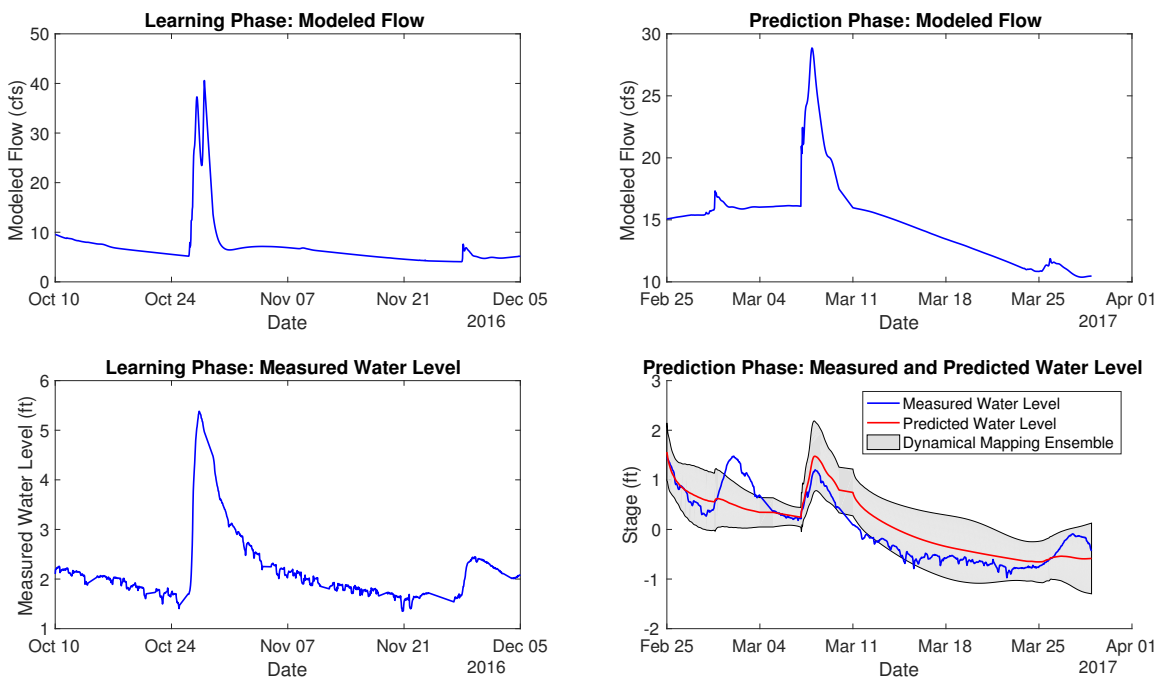


Figure B.75: Site LTLTKY01

LTLTKY02, Mean nRMSE = 36.92%

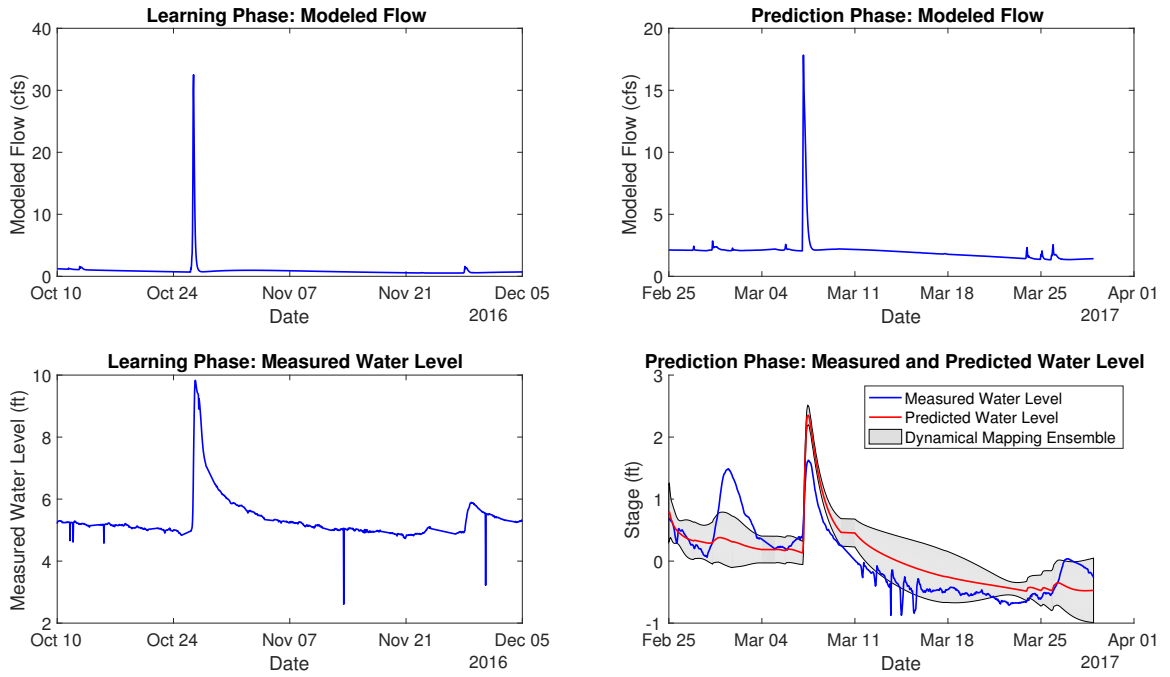


Figure B.76: Site LTLTKY02

LTLWAPSI01, Mean nRMSE = 44.05%

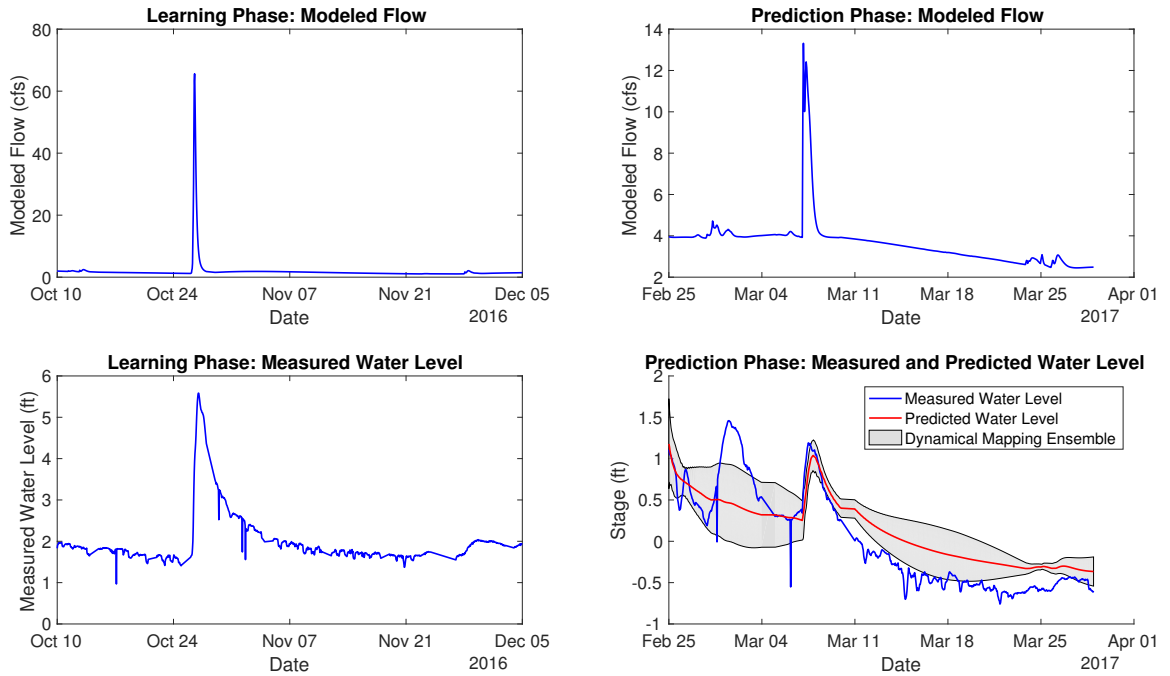


Figure B.77: Site LTLWAPSI01

LTLWLNT01, Mean nRMSE = 5.89%

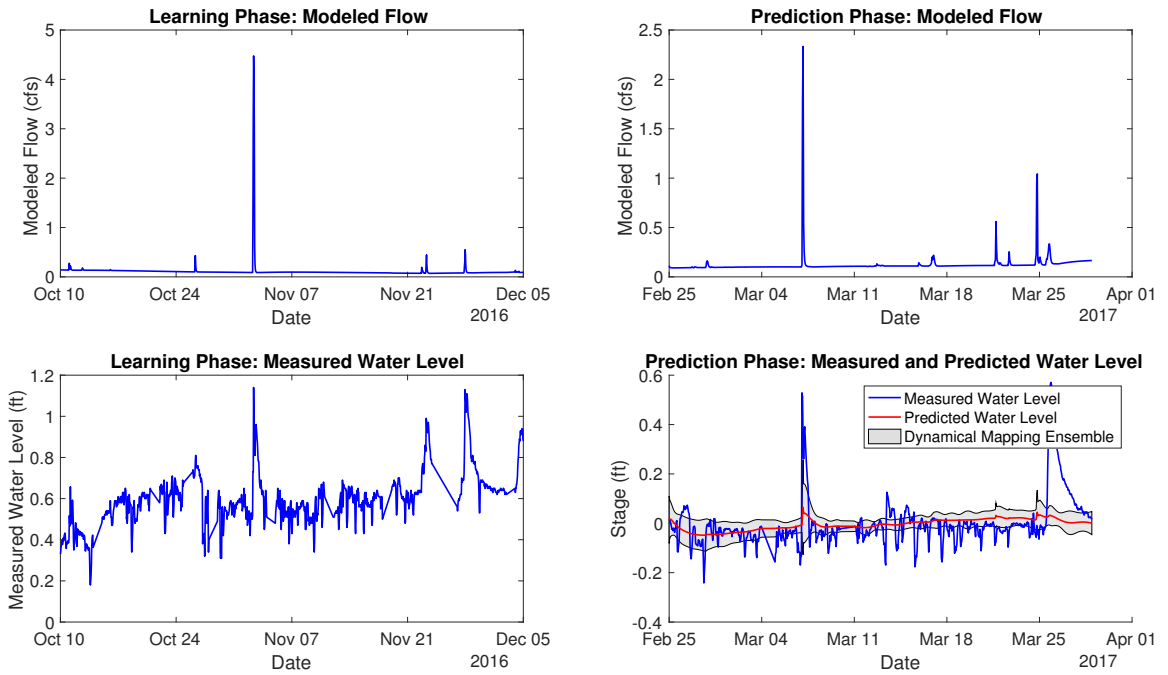


Figure B.78: Site LTLWLNT01

MADCRK01, Mean nRMSE = 5.93%

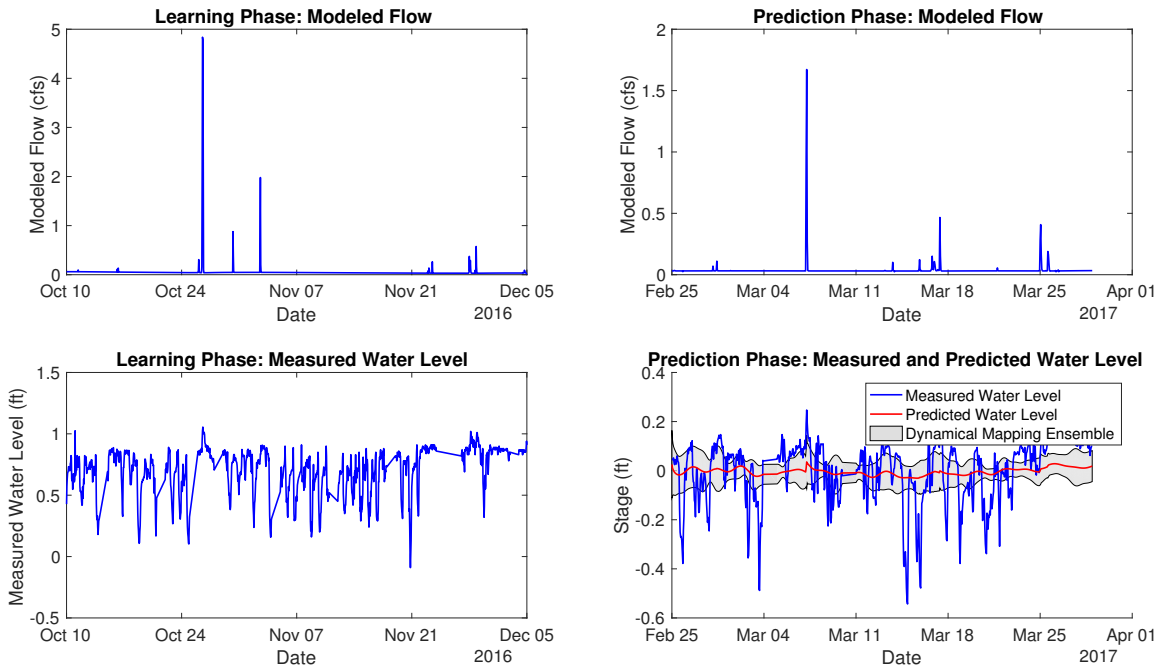


Figure B.79: Site MADCRK01

MADCRK02, Mean nRMSE = 11.69%

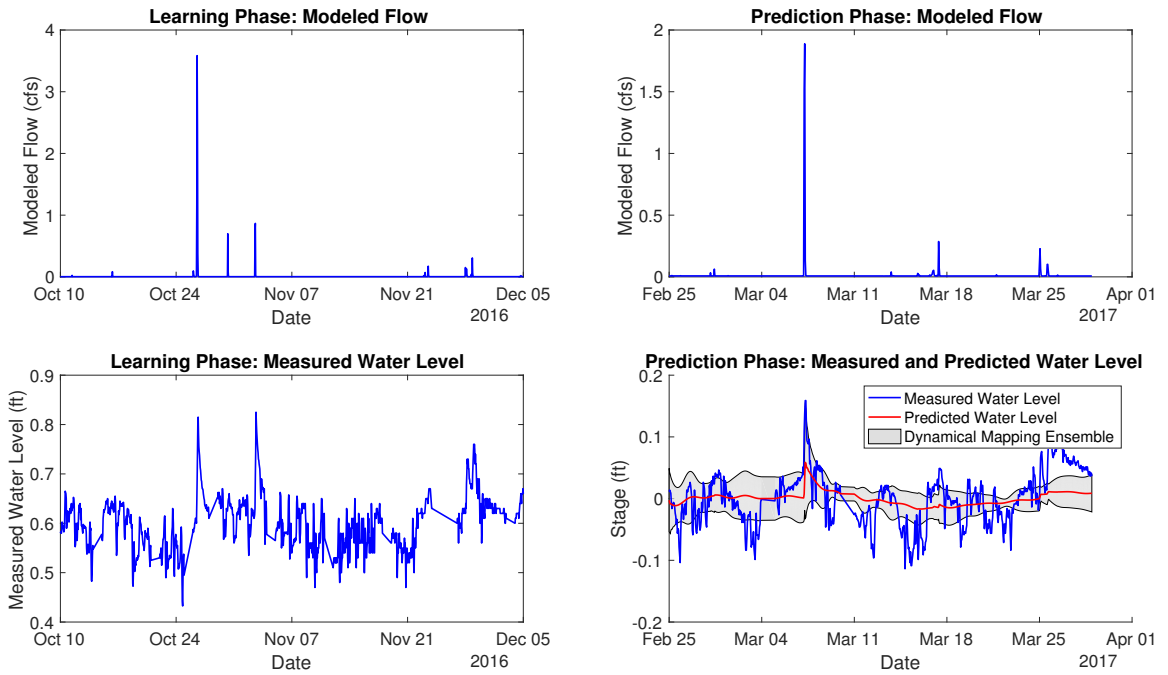


Figure B.80: Site MADCRK02

MAPLERV01, Mean nRMSE = 23.54%

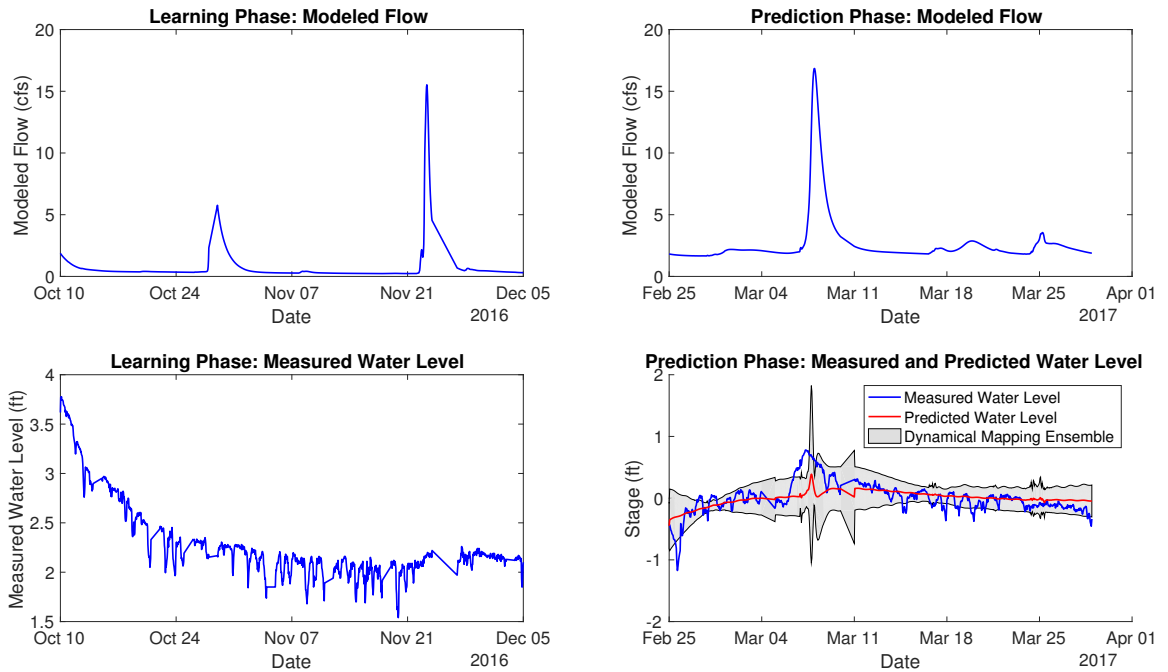


Figure B.81: Site MAPLERV01

MAPLERV02, Mean nRMSE = 25.86%

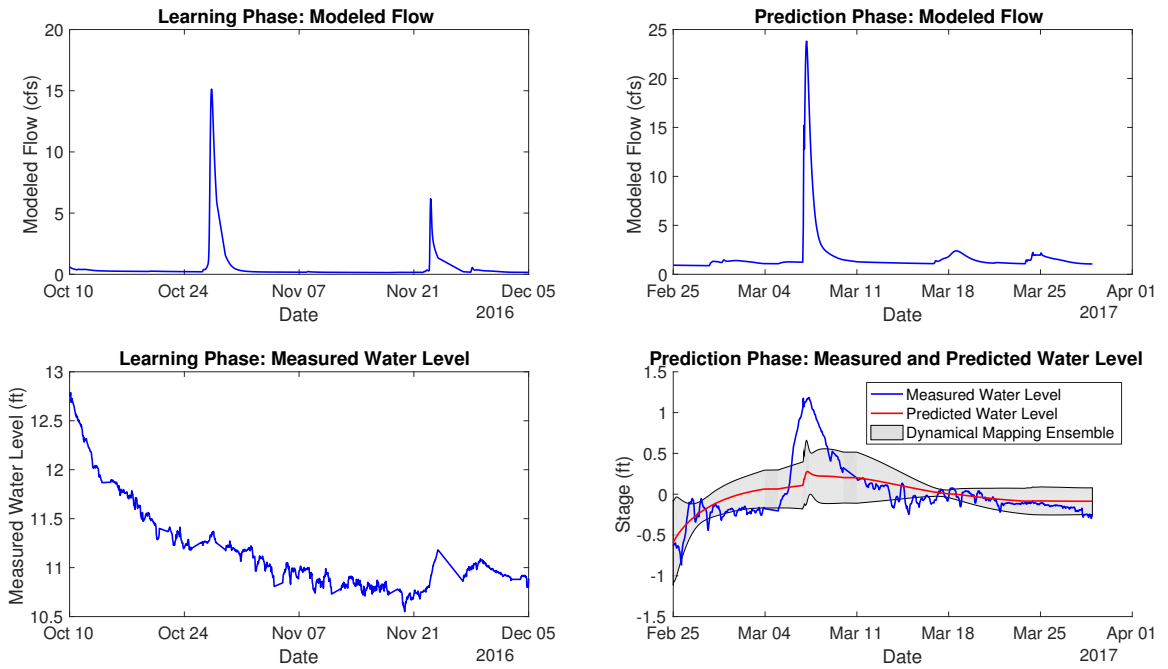


Figure B.82: Site MAPLERV02

MCHKNCK01, Mean nRMSE = 12.96%

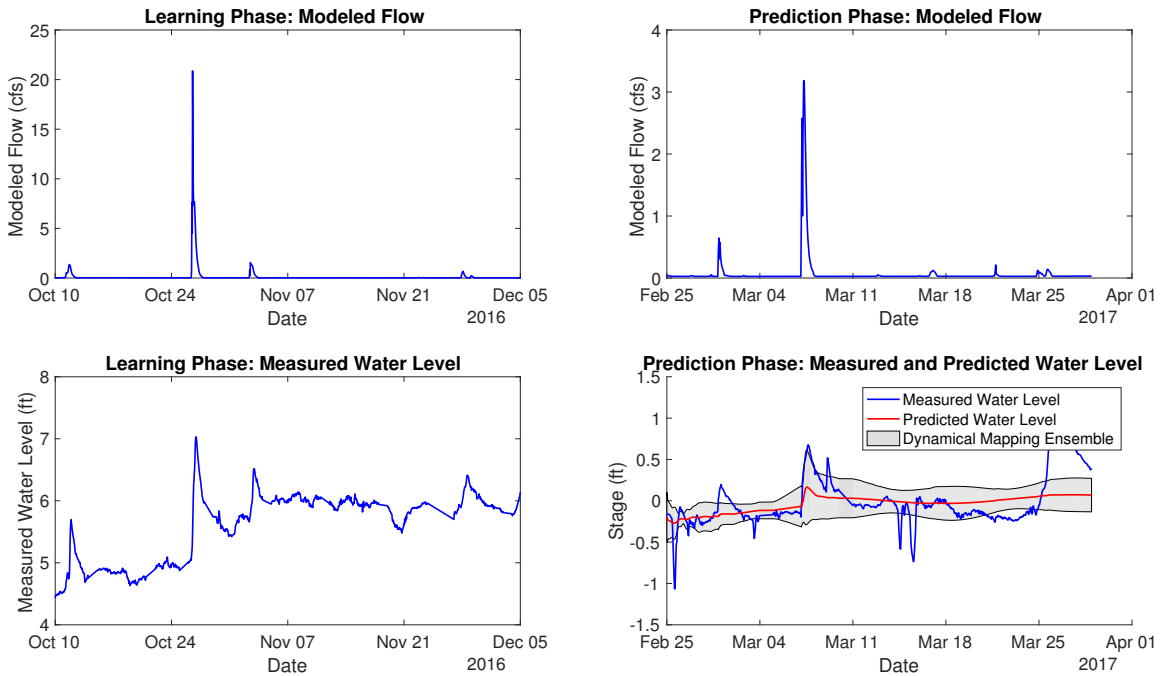


Figure B.83: Site MCHKNCK01

MDCRK01, Mean nRMSE = 7.97%

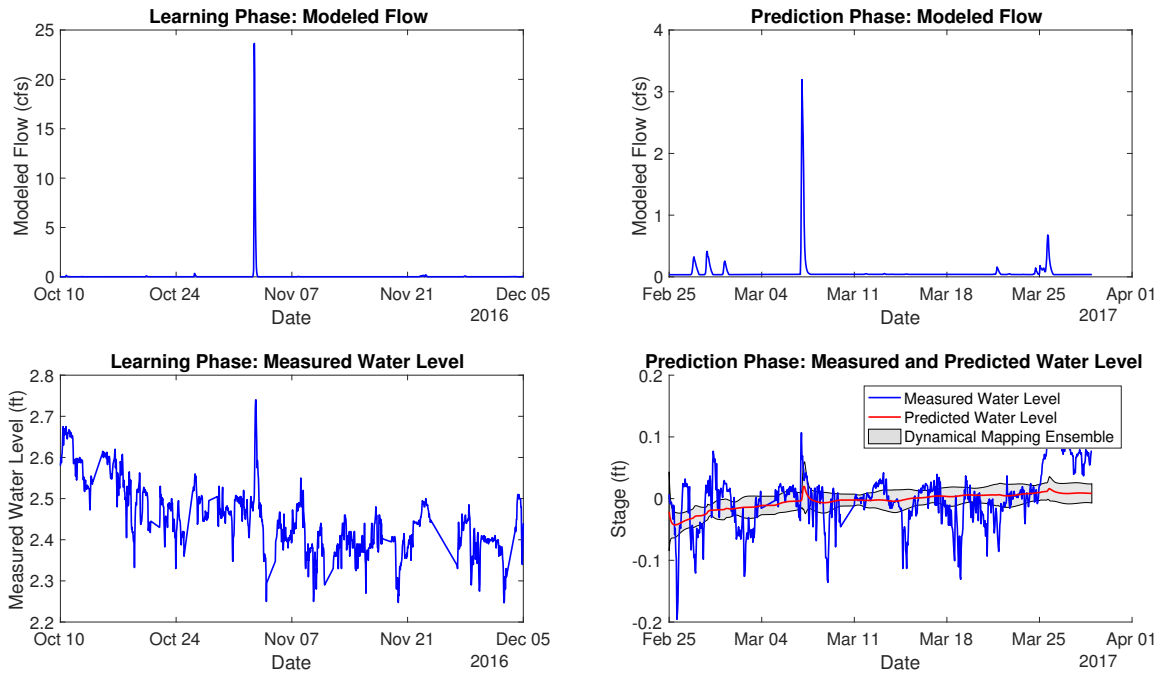


Figure B.84: Site MDCRK01

MDDLRCN01, Mean nRMSE = 6.76%

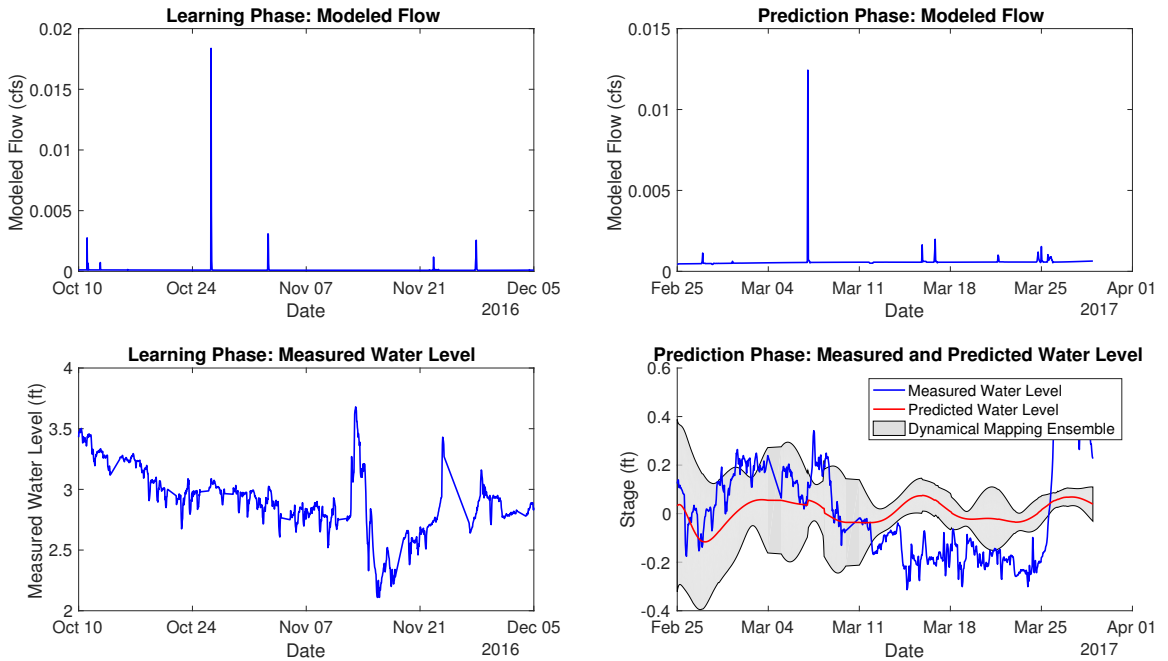


Figure B.85: Site MDDLRCN01

MDDLRCN02, Mean nRMSE = -101.90%

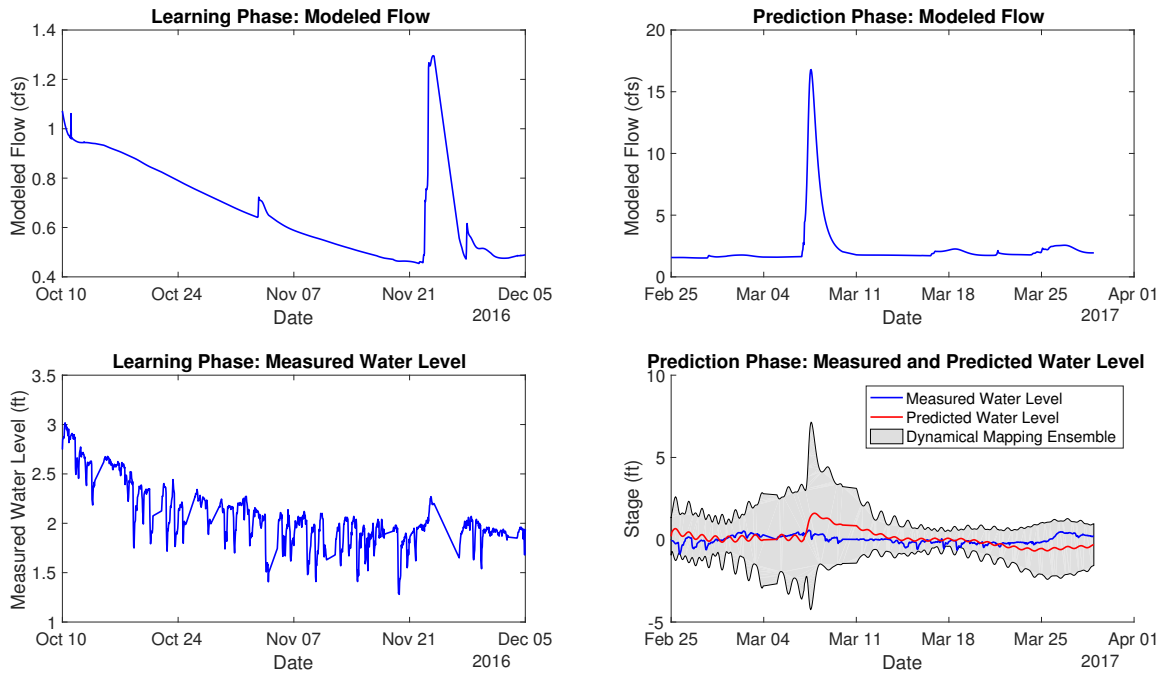


Figure B.86: Site MDDLRCN02

MDDLRCN03, Mean nRMSE = 8.74%

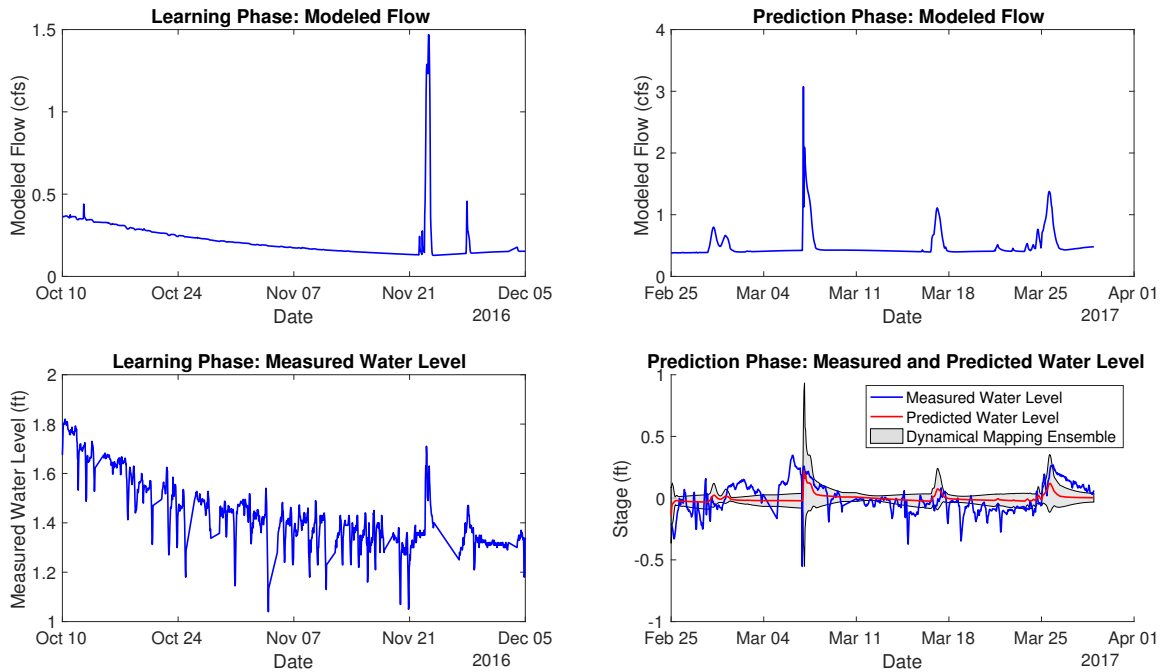


Figure B.87: Site MDDLRCN03

MQKTARV01, Mean nRMSE = 47.91%

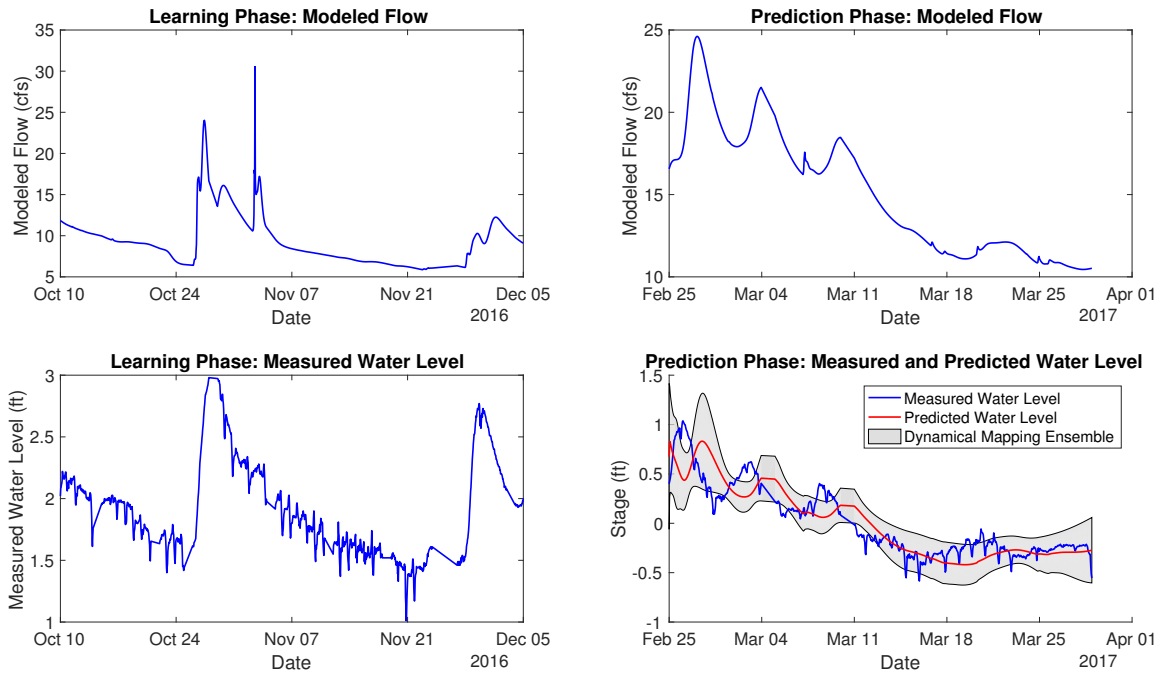


Figure B.88: Site MQKTARV01

MQKTARV02, Mean nRMSE = 38.32%

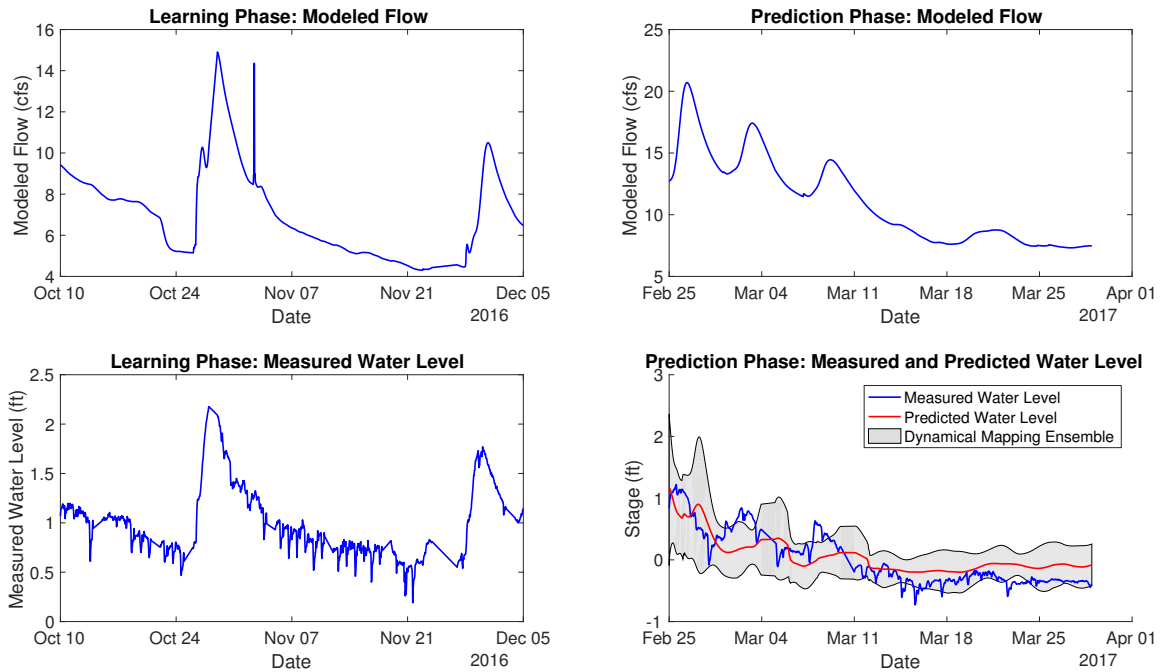


Figure B.89: Site MQKTARV02

MQKTARV04, Mean nRMSE = 0.04%

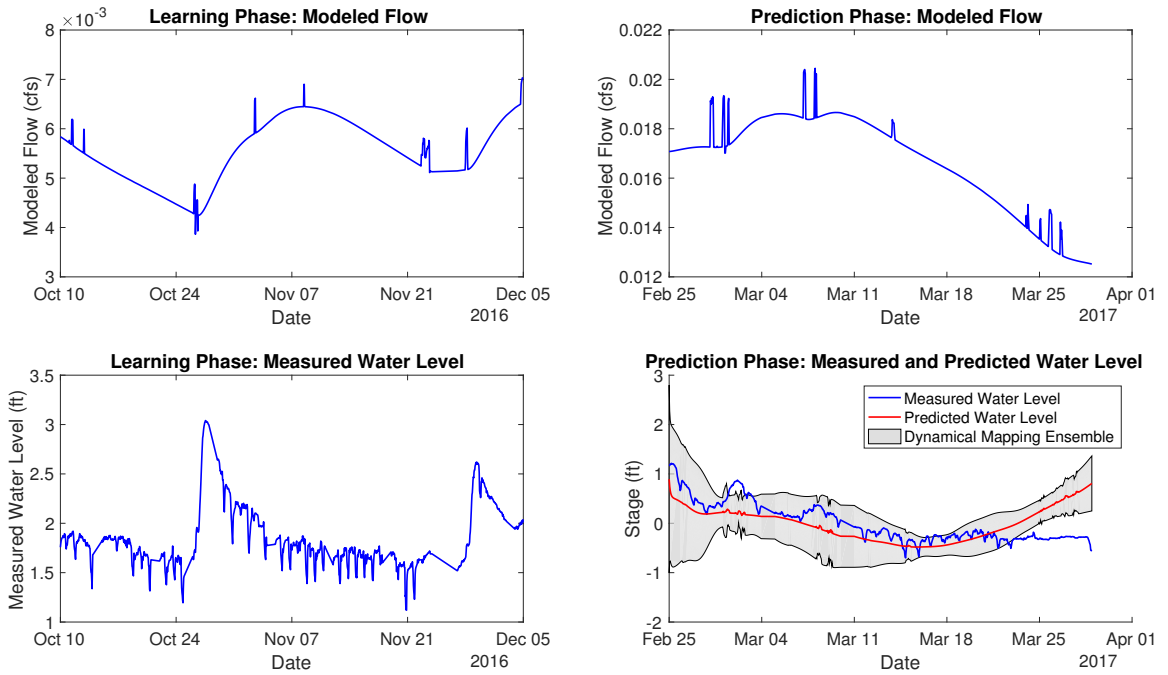


Figure B.90: Site MQKTARV04

MRGNCR01, Mean nRMSE = 22.15%

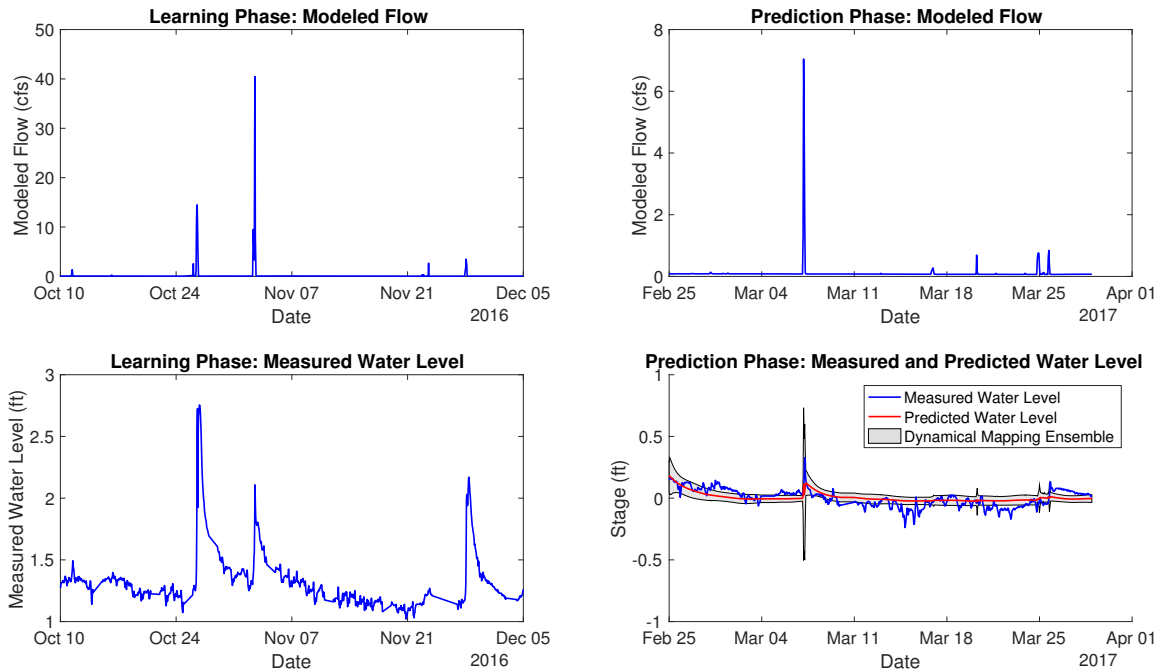


Figure B.91: Site MRGNCR01

MSQUTOCR01, Mean nRMSE = -30.20%

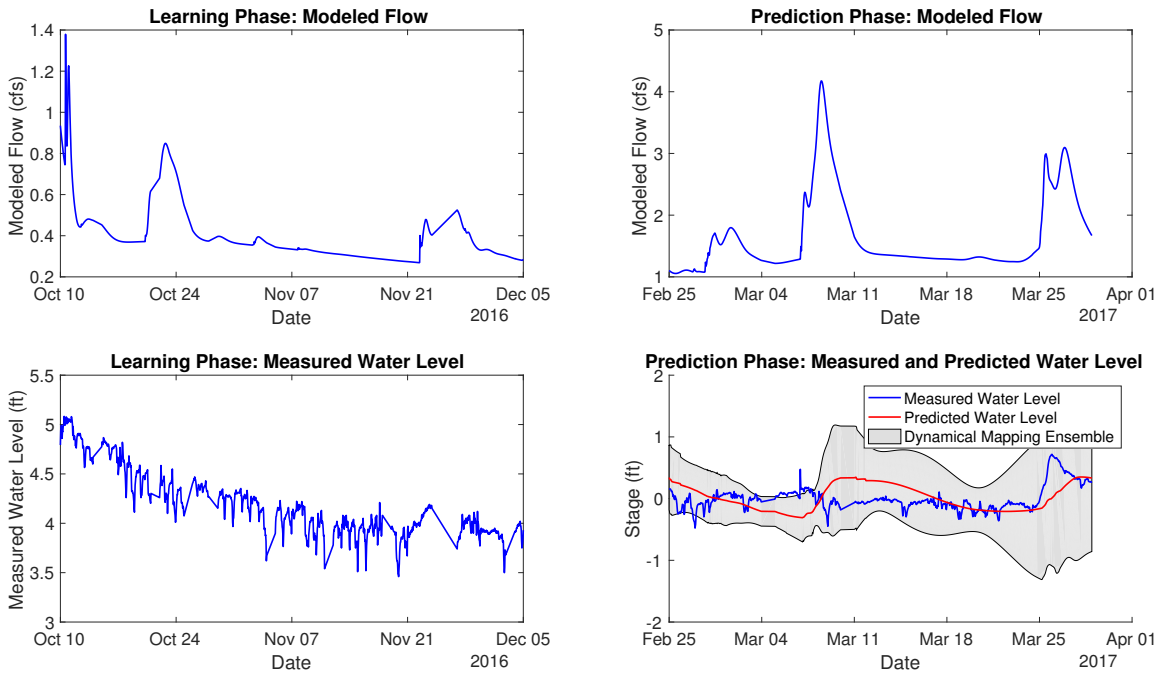


Figure B.92: Site MSQUTOCR01

MSQUTOCR02, Mean nRMSE = 27.54%

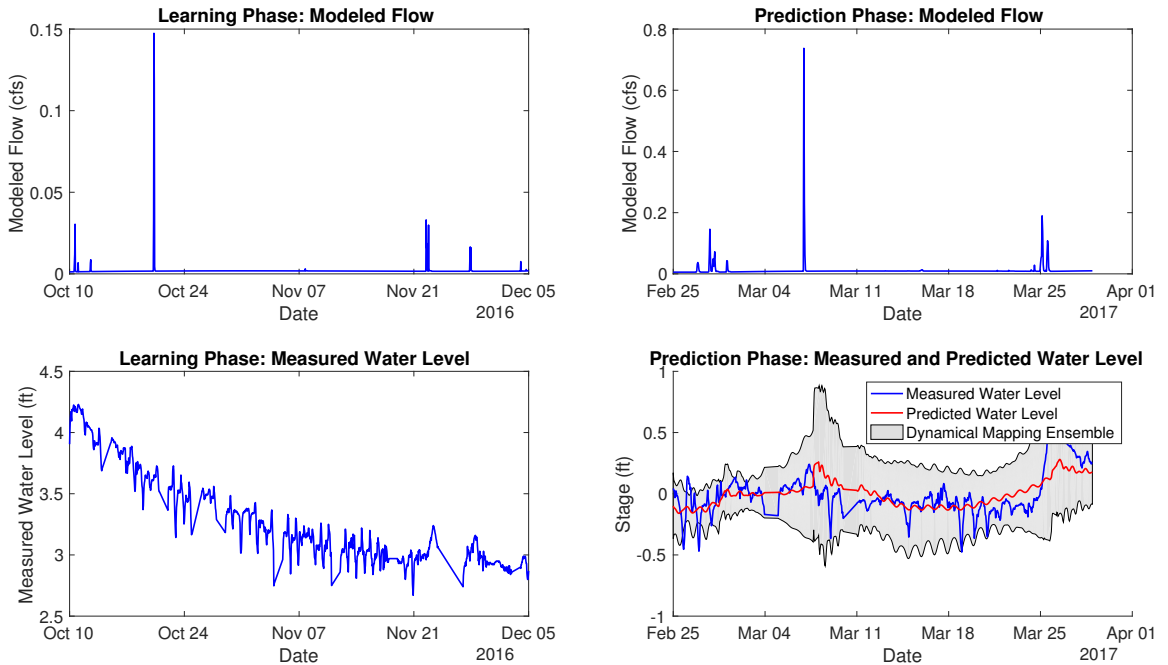


Figure B.93: Site MSQUTOCR02

MUDCR01, Mean nRMSE = -80.43%

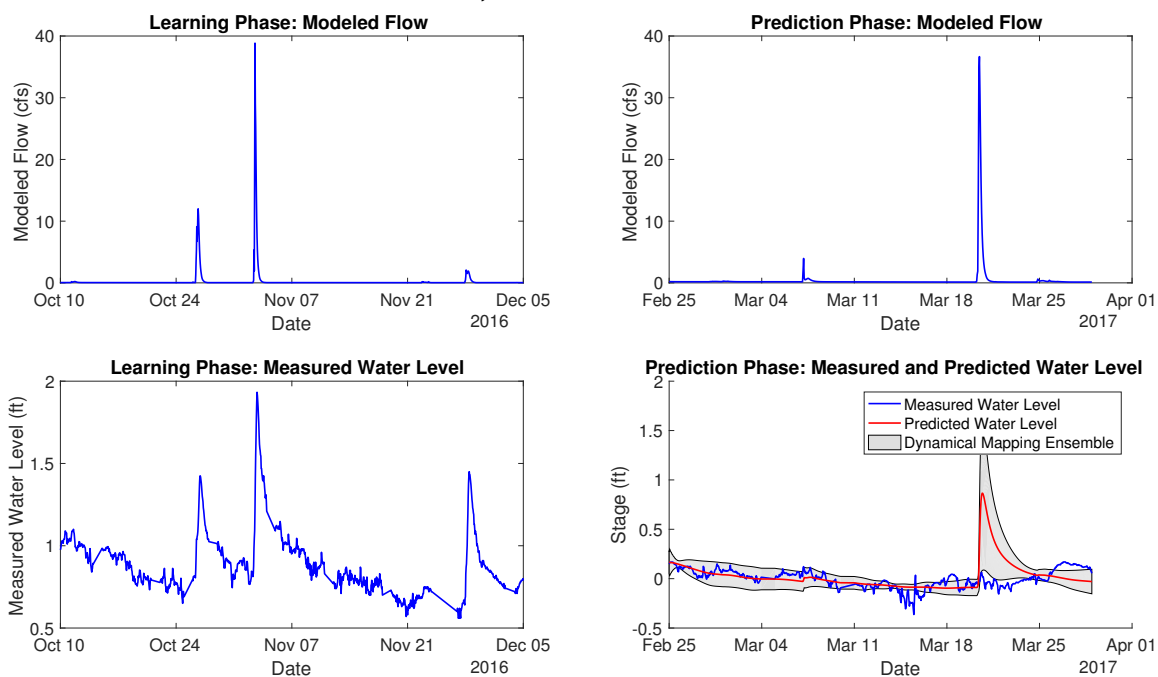


Figure B.94: Site MUDCR01

MUDCRK01, Mean nRMSE = 6.83%

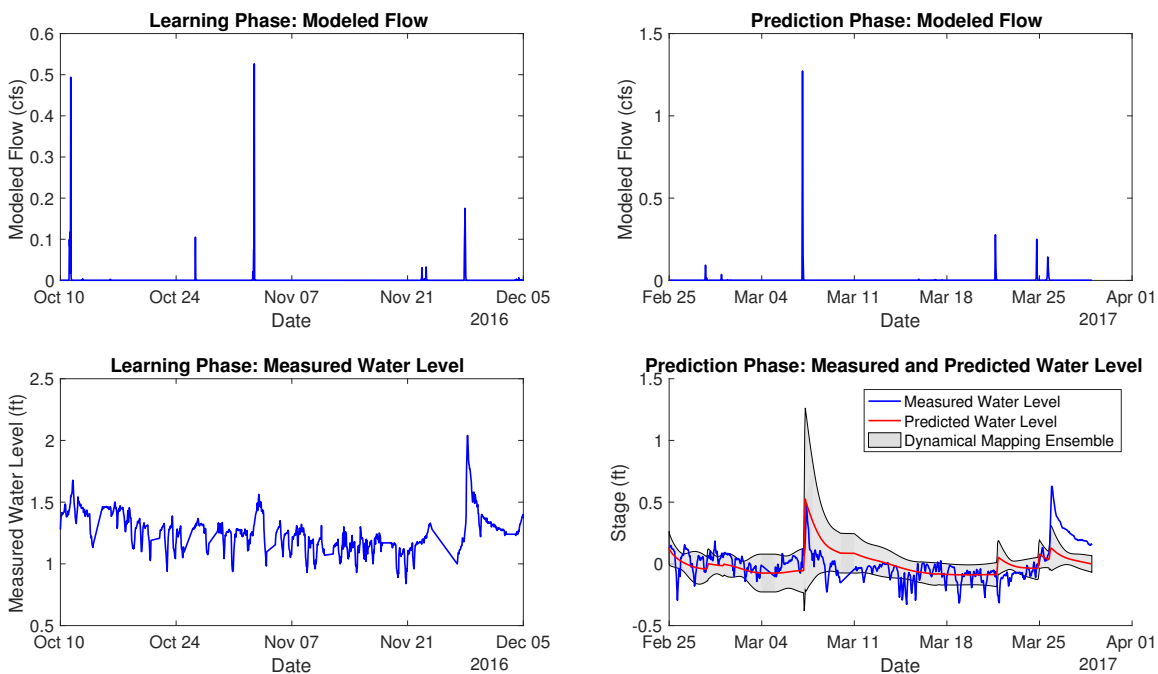


Figure B.95: Site MUDCRK01

NENGLSH01, Mean nRMSE = 17.29%

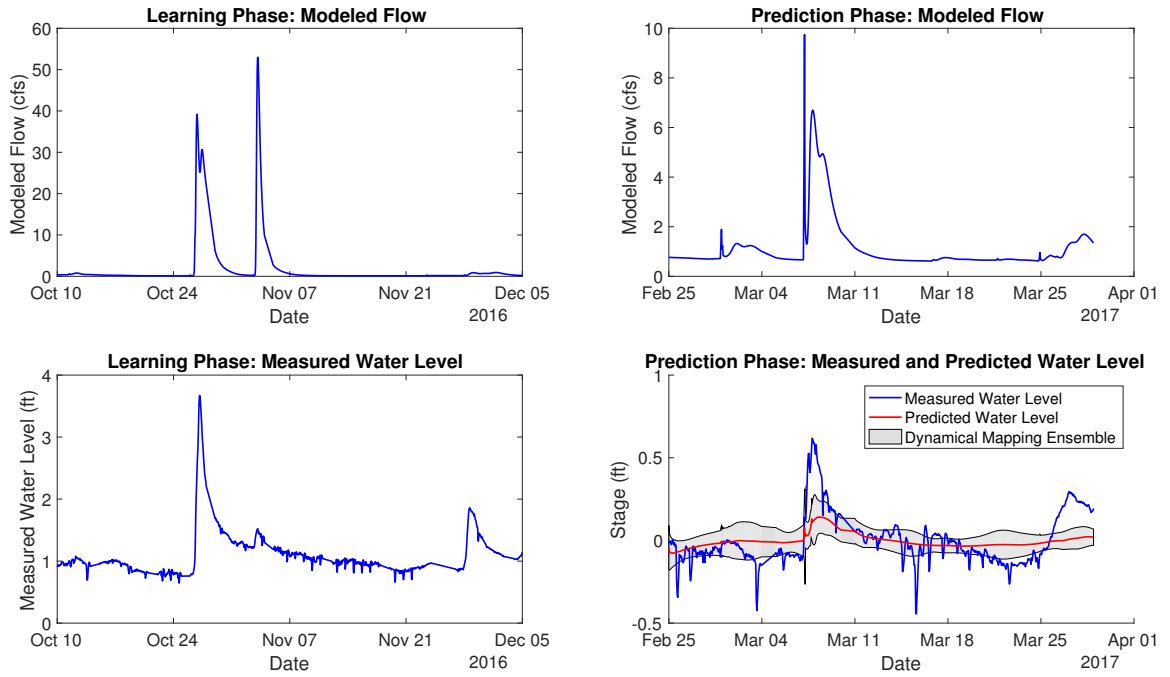


Figure B.96: Site NENGLSH01

NFCTFSH01, Mean nRMSE = 5.26%

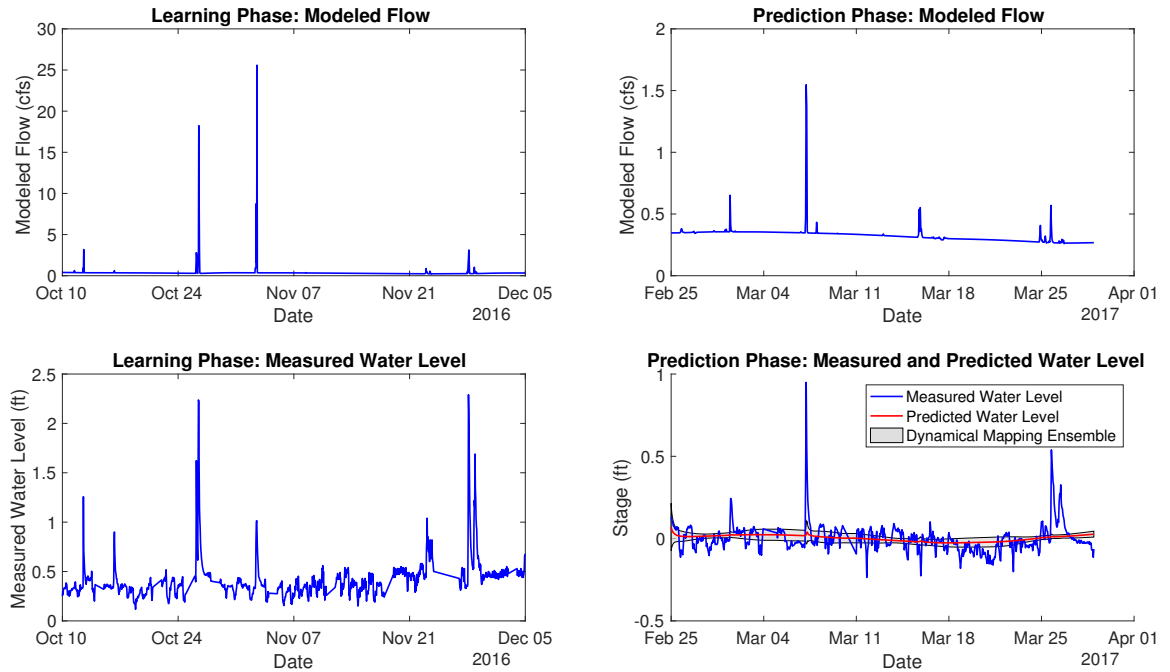


Figure B.97: Site NFCTFSH01

NFMQKTA01, Mean nRMSE = 47.27%

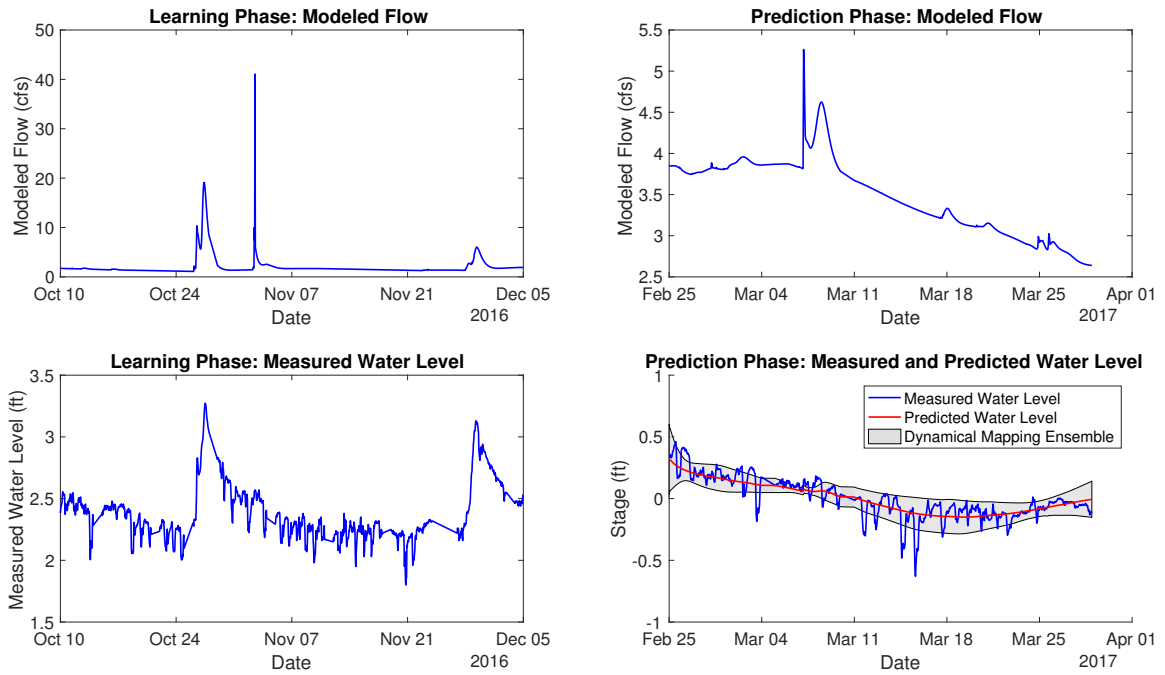


Figure B.98: Site NFMQKTA01

NRCCNRV01, Mean nRMSE = 56.86%

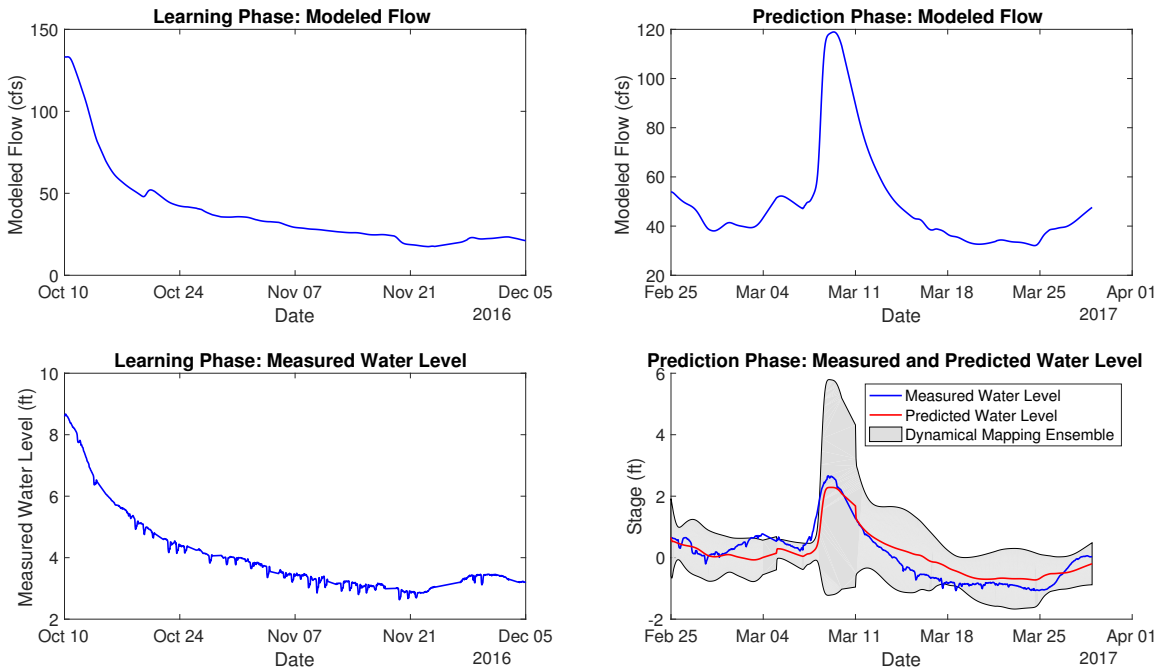


Figure B.99: Site NRCCNRV01

NRCCNRV02, Mean nRMSE = -4.54%

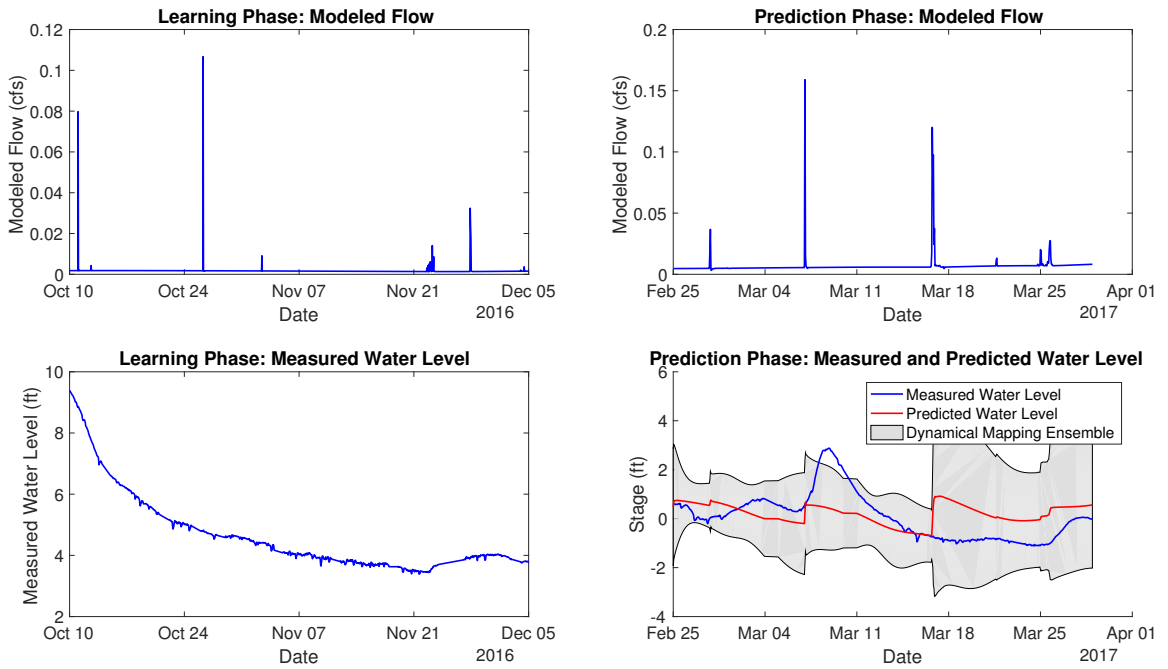


Figure B.100: Site NRCCNRV02

NRCCNRV03, Mean nRMSE = 18.35%

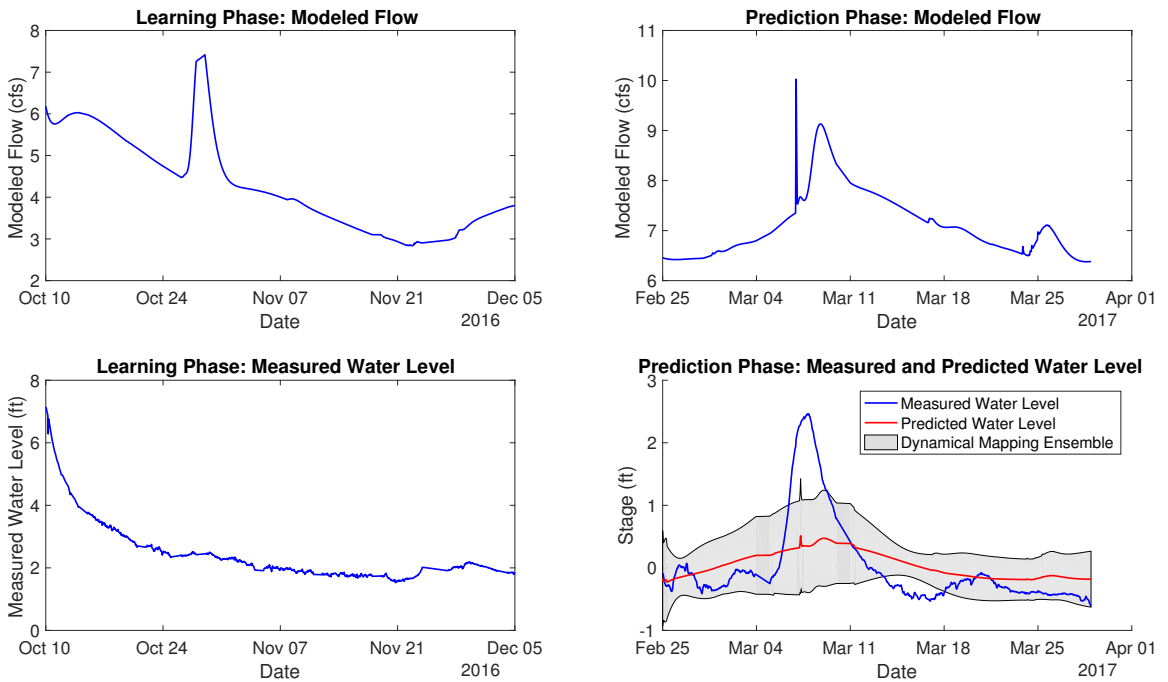


Figure B.101: Site NRCCNRV03

NSKNK01, Mean nRMSE = 0.88%

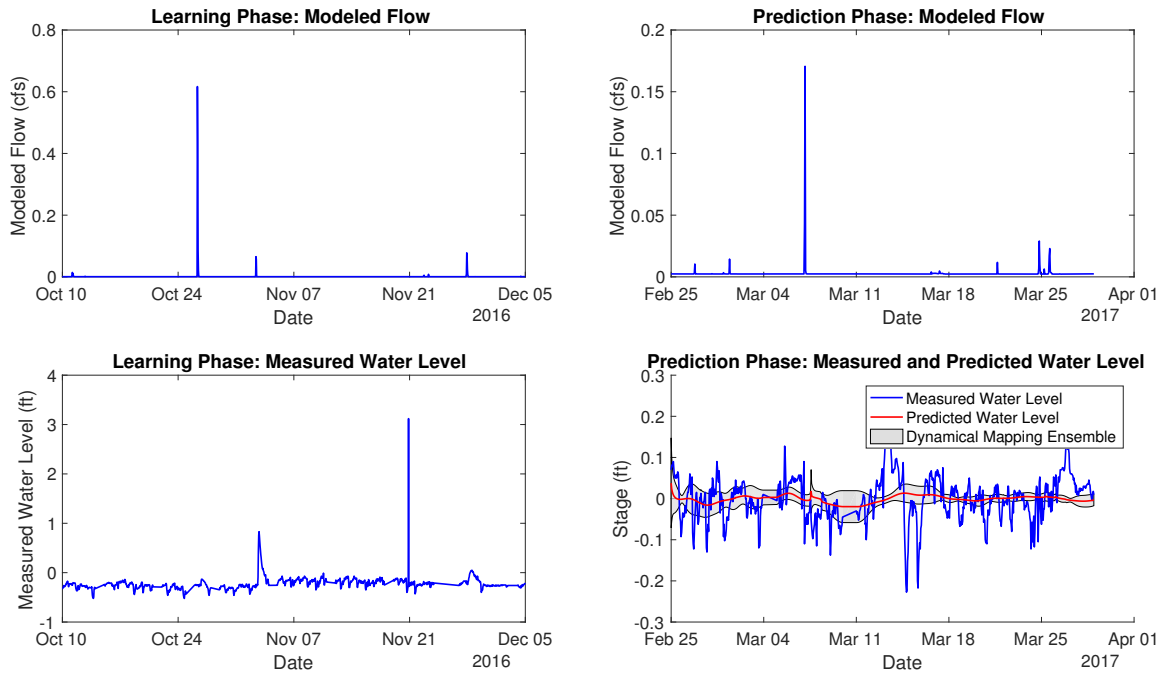


Figure B.102: Site NSKNK01

NWLNTCR01, Mean nRMSE = 5.98%

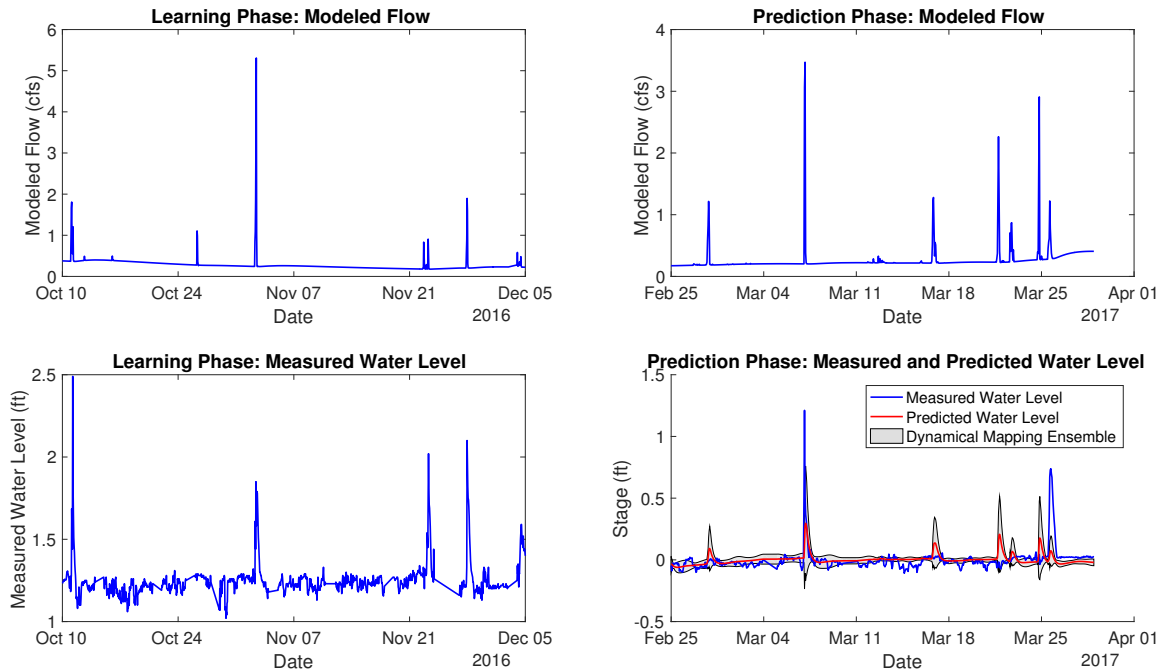


Figure B.103: Site NWLNTCR01

NWLNTCR02, Mean nRMSE = 19.85%

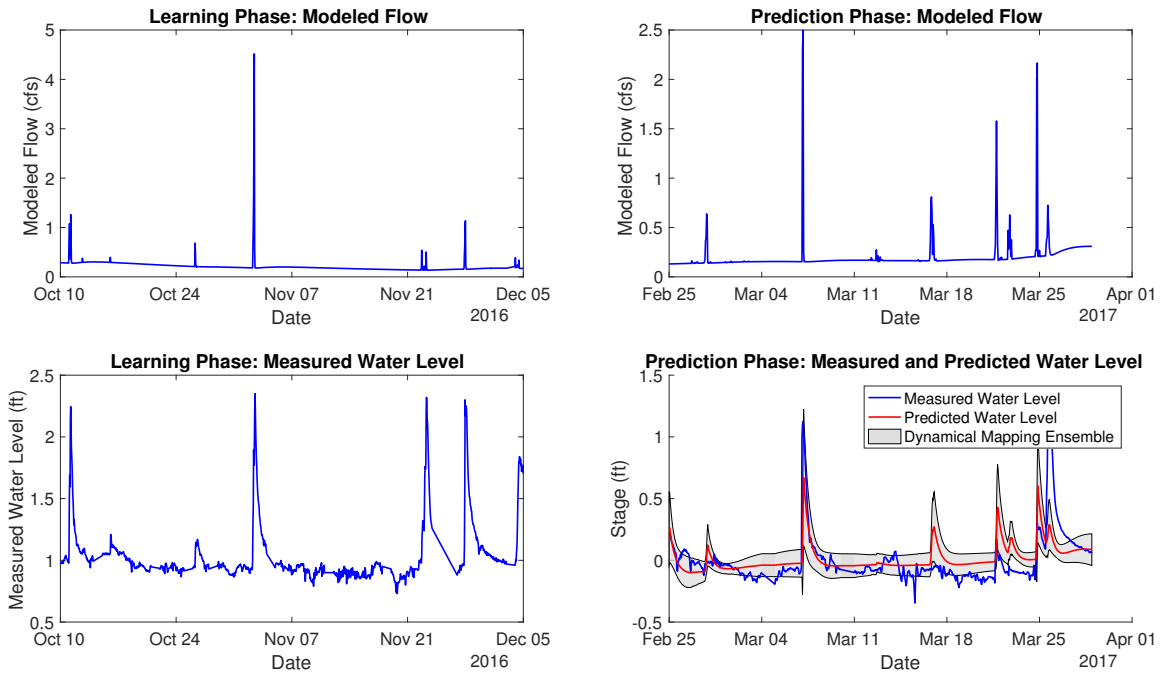


Figure B.104: Site NWLNTCR02

OCHYDNRV01, Mean nRMSE = -41.84%

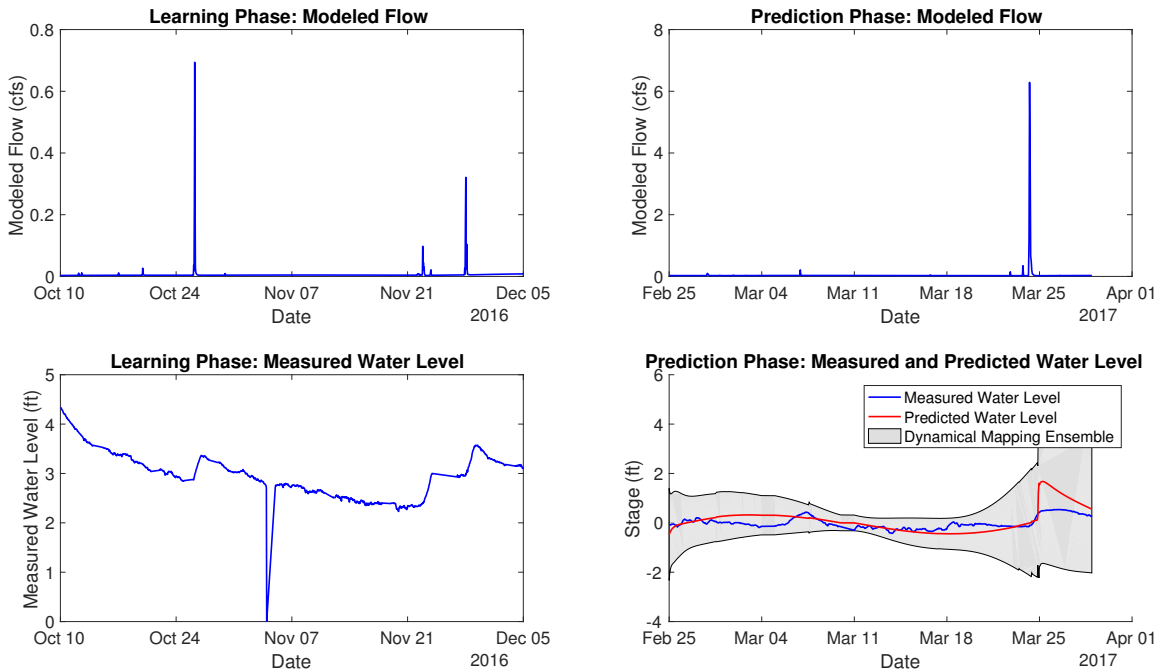


Figure B.105: Site OCHYDNRV01

OTTERCR01, Mean nRMSE = 44.73%

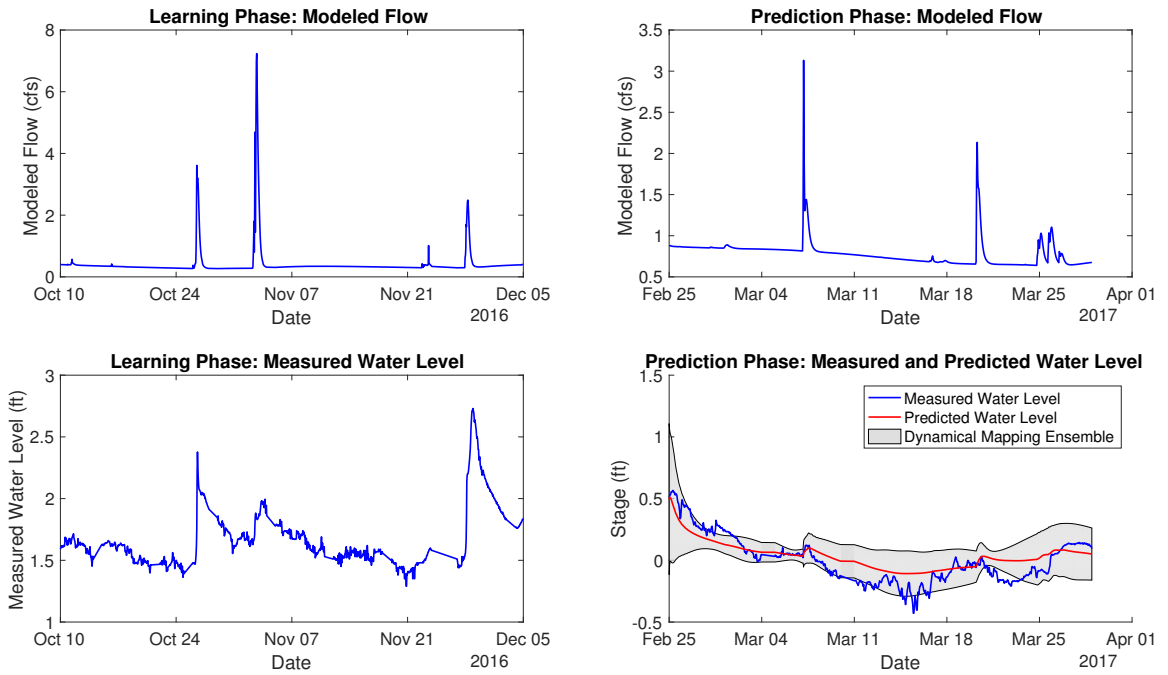


Figure B.106: Site OTTERCR01

OTTRCRK01, Mean nRMSE = 31.04%

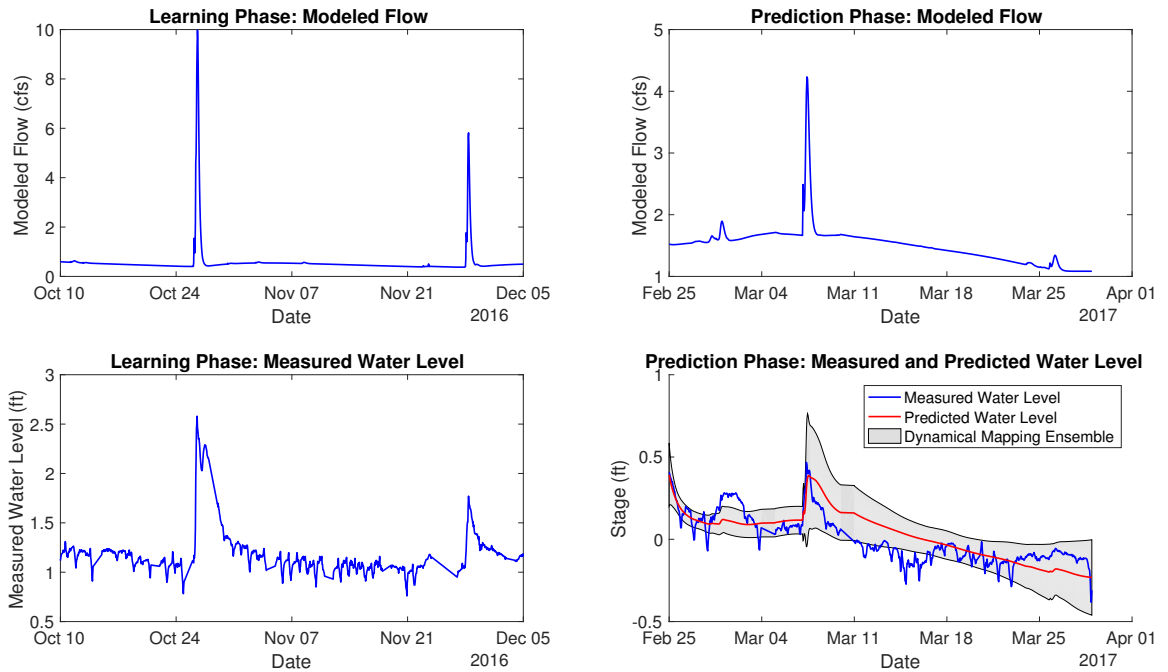


Figure B.107: Site OTTRCRK01

OTTRCRK02, Mean nRMSE = 42.35%

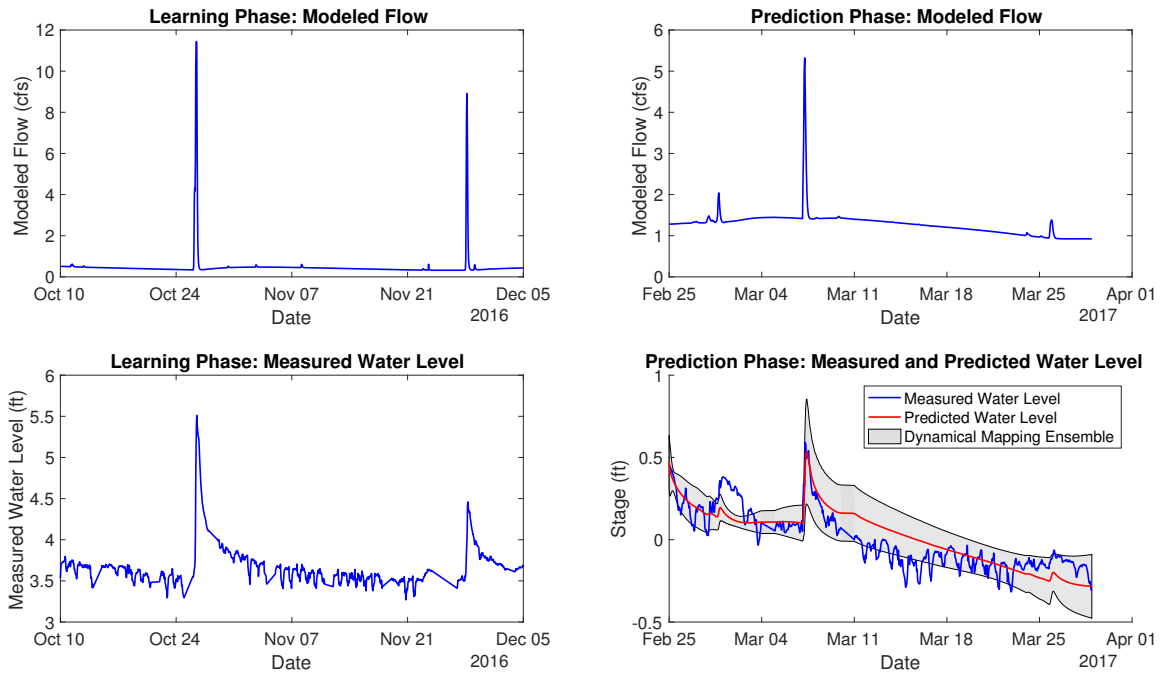


Figure B.108: Site OTTRCRK02

OTTRCRK03, Mean nRMSE = 37.40%

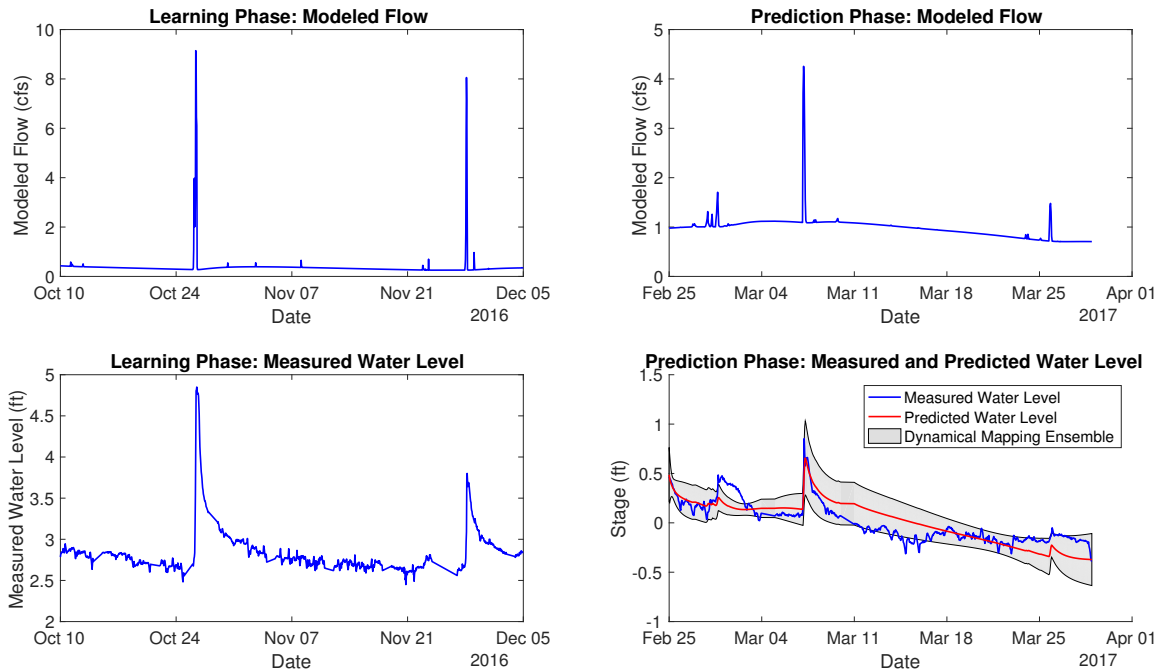


Figure B.109: Site OTTRCRK03

OTTRCRK04, Mean nRMSE = 4.56%

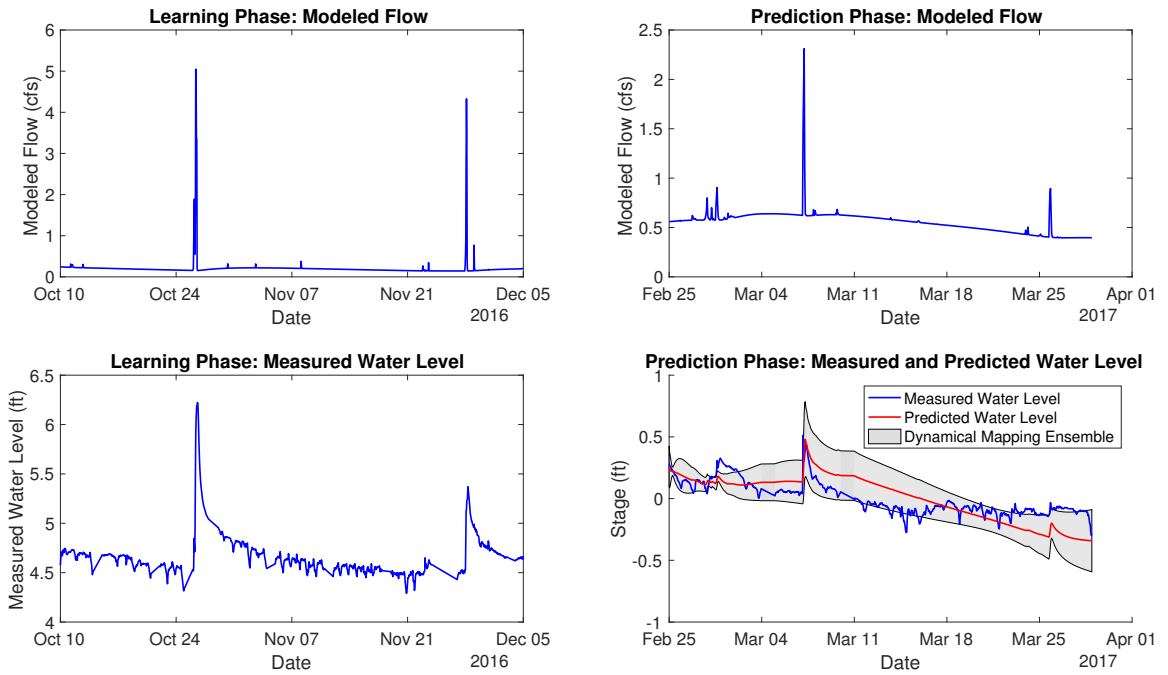


Figure B.110: Site OTTRCRK04

PRAIRIECRK01, Mean nRMSE = 15.64%

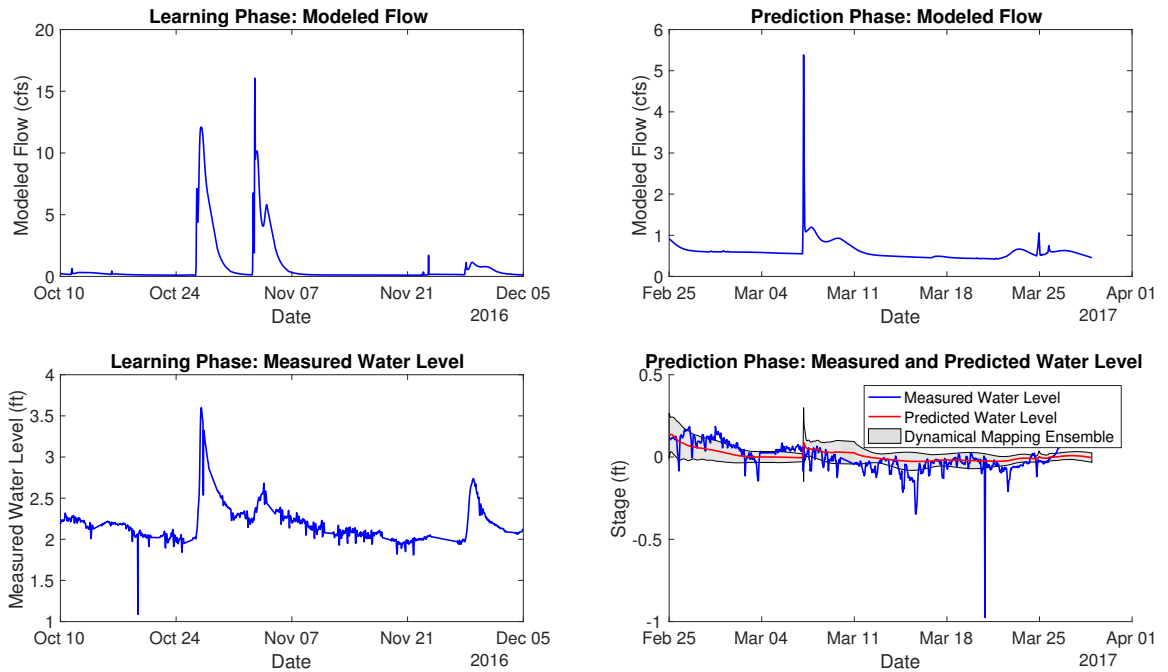


Figure B.111: Site PRAIRIECRK01

PRICECR02, Mean nRMSE = 9.16%

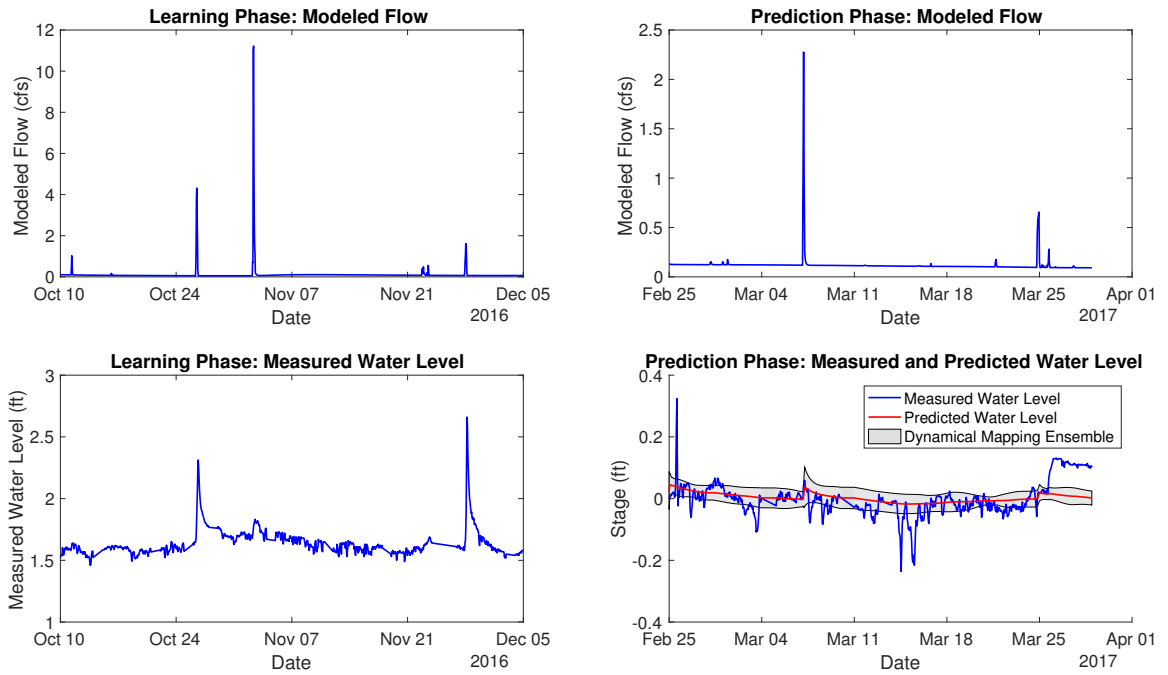


Figure B.112: Site PRICECR02

RAPIDCR01, Mean nRMSE = 28.11%

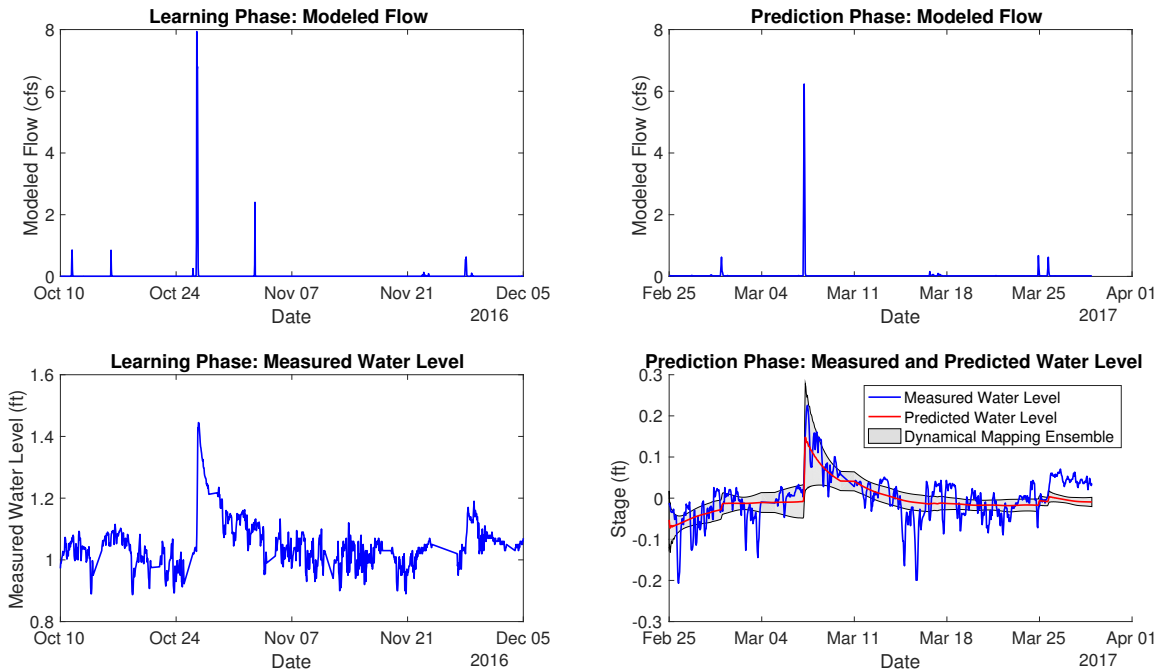


Figure B.113: Site RAPIDCR01

RAPIDTRB01, Mean nRMSE = 53.90%

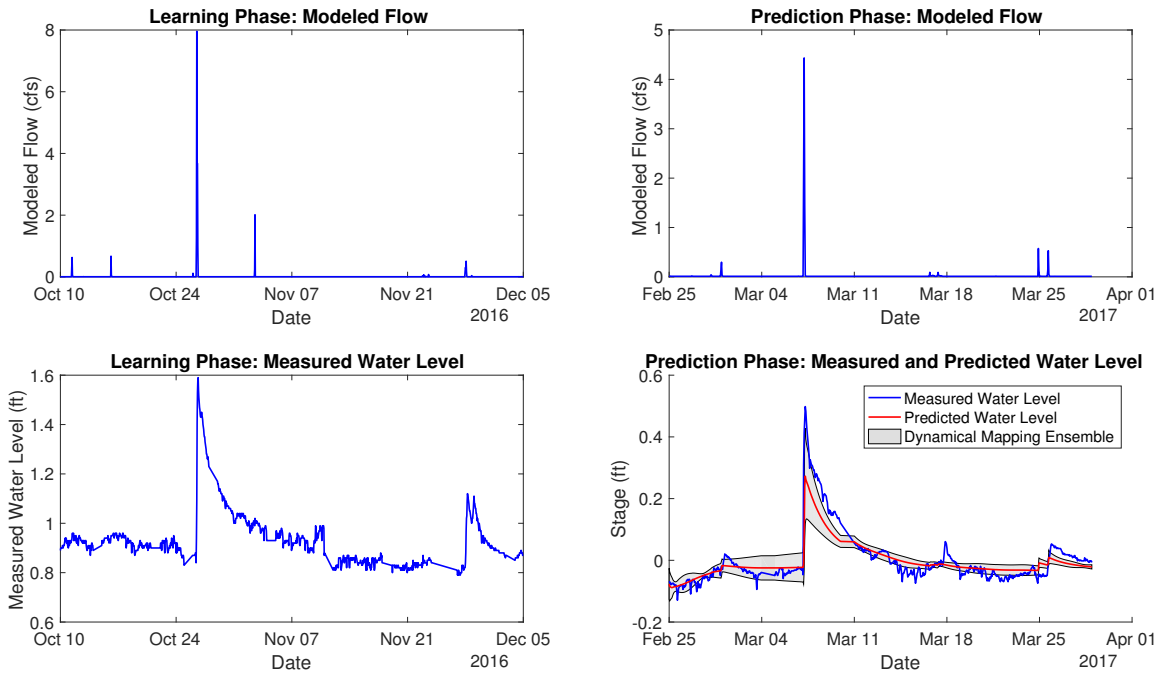


Figure B.114: Site RAPIDTRB01

RAVENCR01, Mean nRMSE = -3.24%

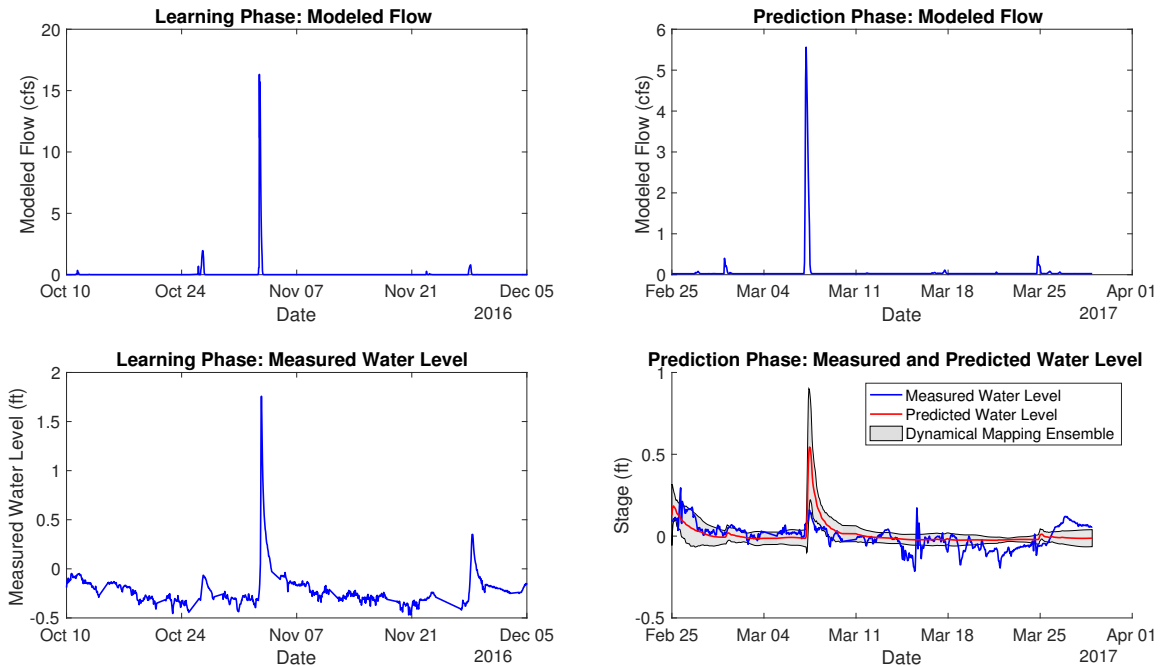


Figure B.115: Site RAVENCR01

RVRTSCR01, Mean nRMSE = 12.81%

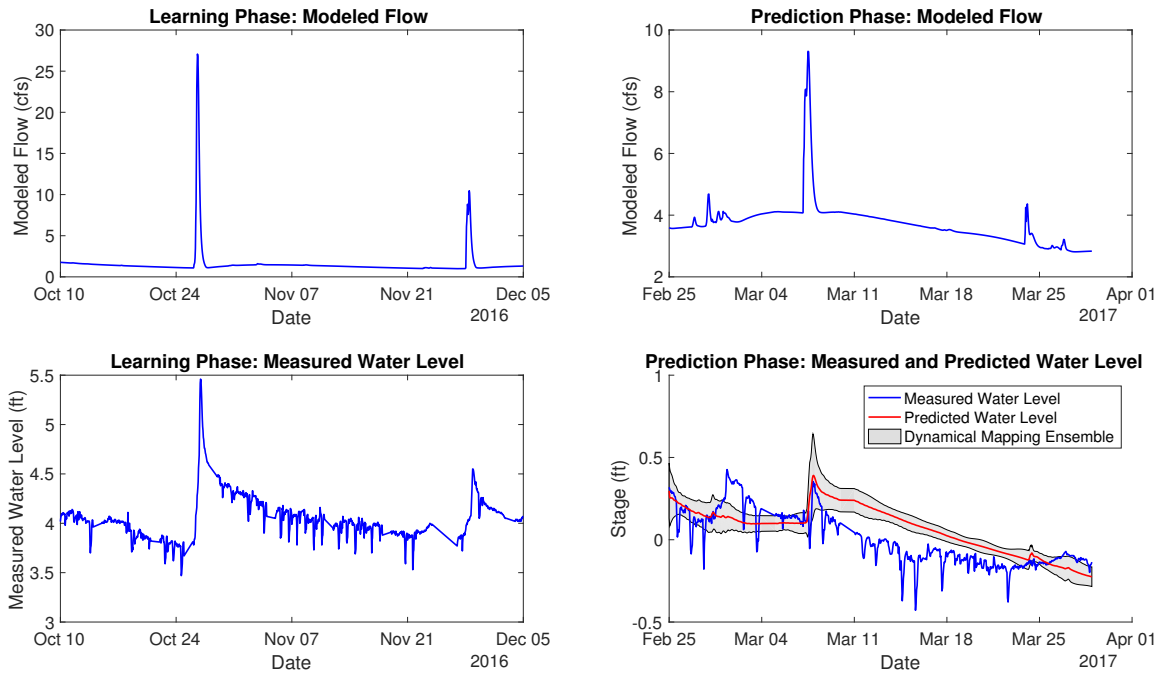


Figure B.116: Site RVRTSCR01

RCCNRV01, Mean nRMSE = 77.56%

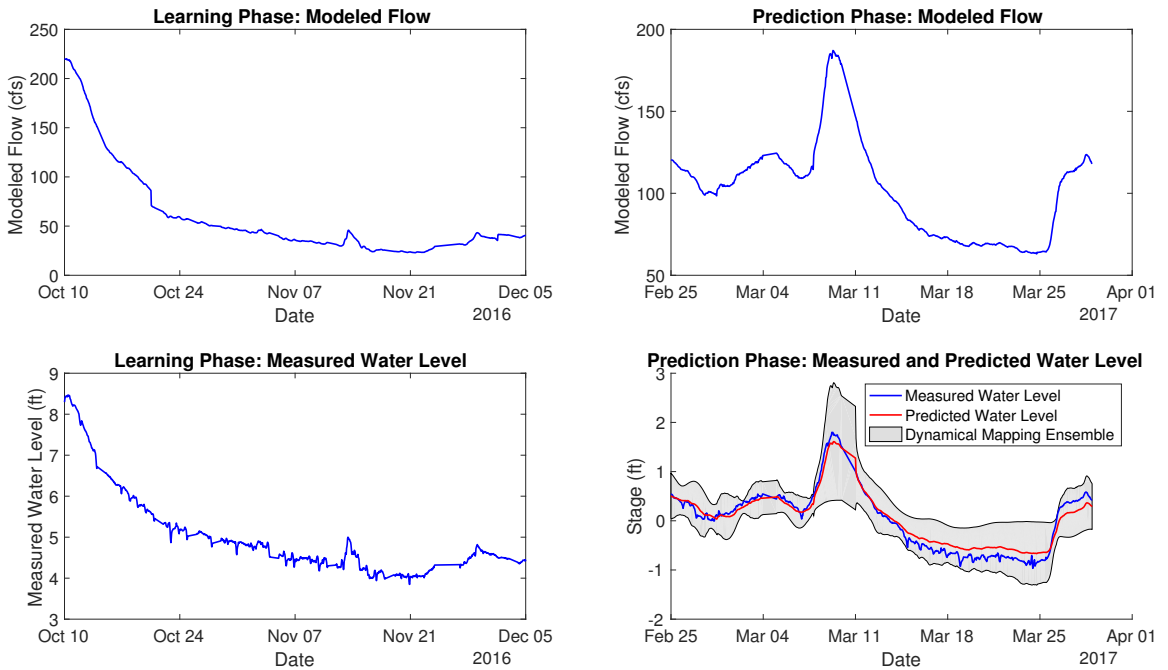


Figure B.117: Site RCCNRV01

RLSTNCR01, Mean nRMSE = 42.24%

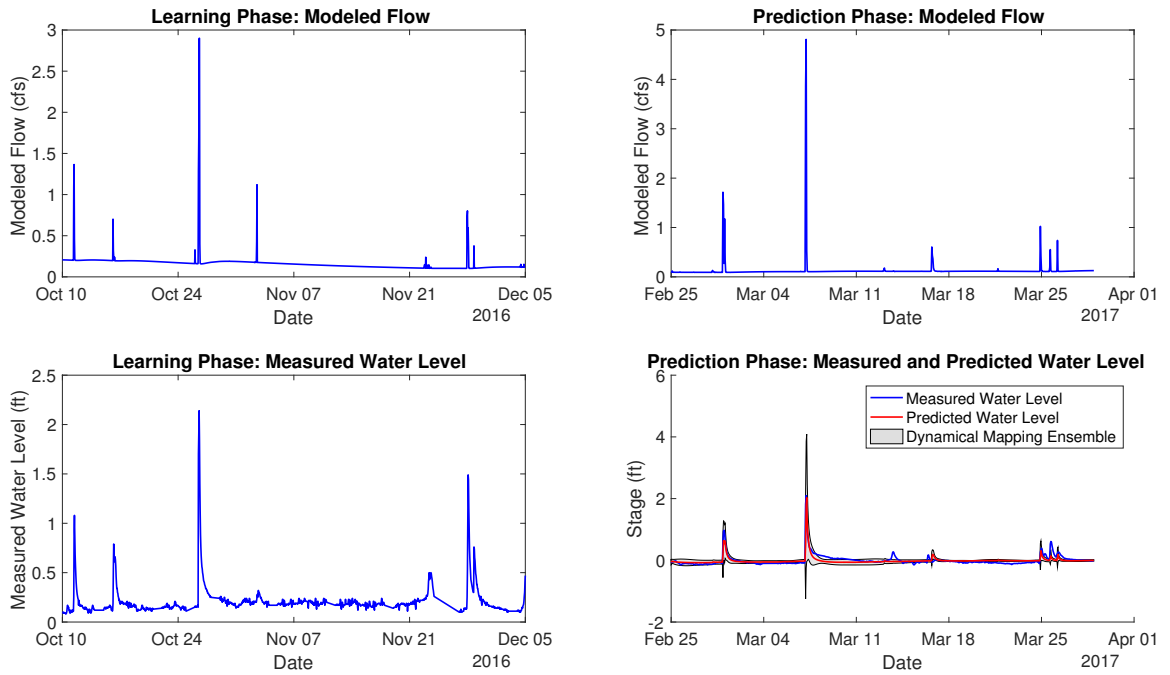


Figure B.118: Site RLSTNCR01

ROCKCR01, Mean nRMSE = 25.62%

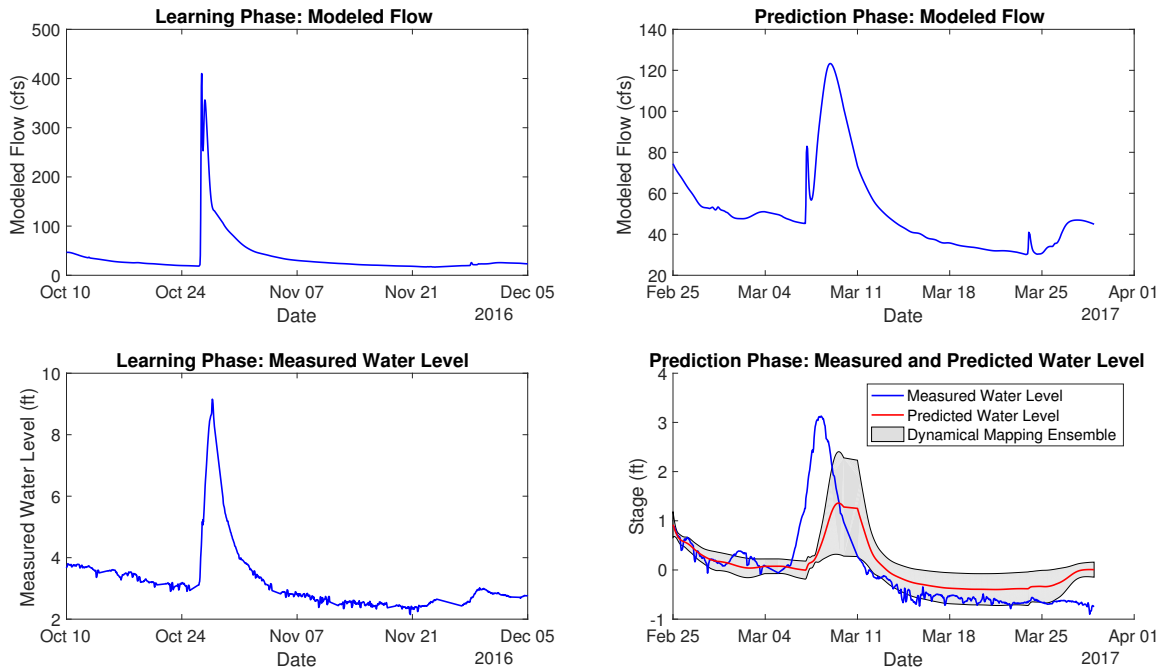


Figure B.119: Site ROCKCR01

ROCKCR02, Mean nRMSE = 37.18%

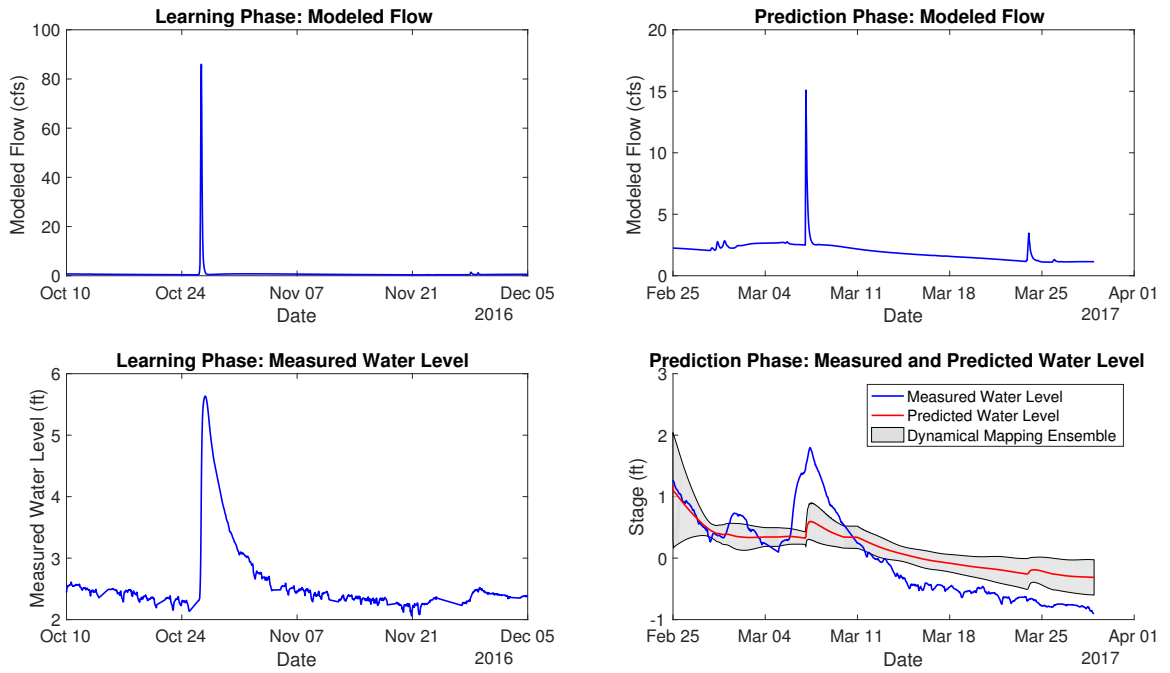


Figure B.120: Site ROCKCR02

SBVRCRK01, Mean nRMSE = 36.36%

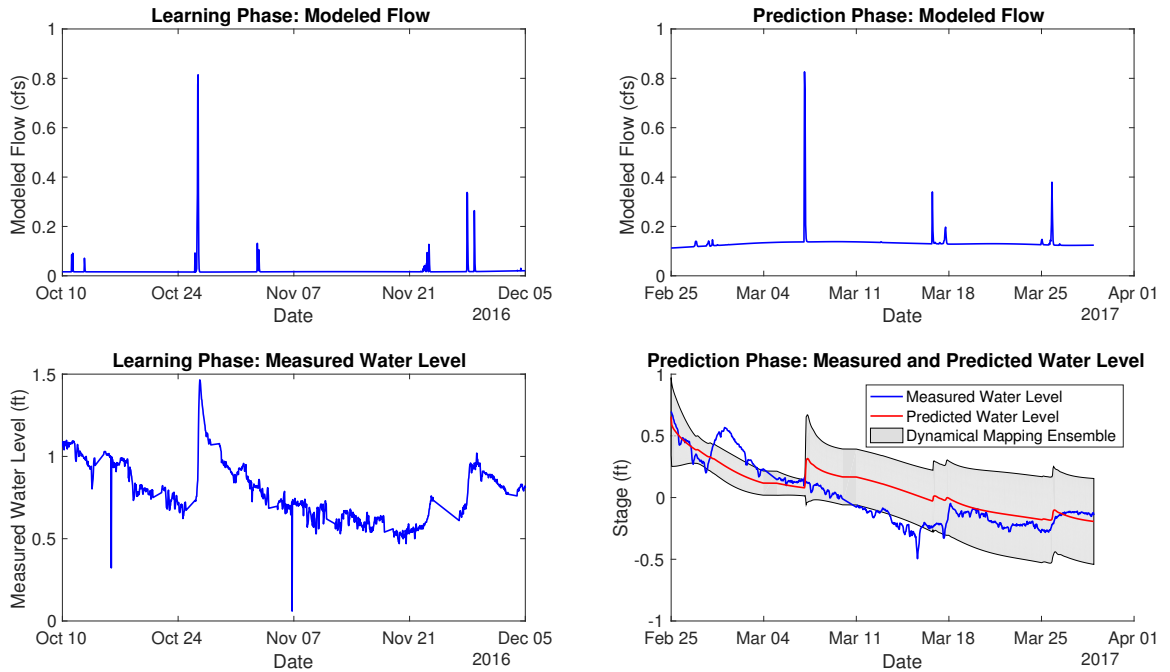


Figure B.121: Site SBVRCRK01

SCHQSTCR01, Mean nRMSE = 8.06%

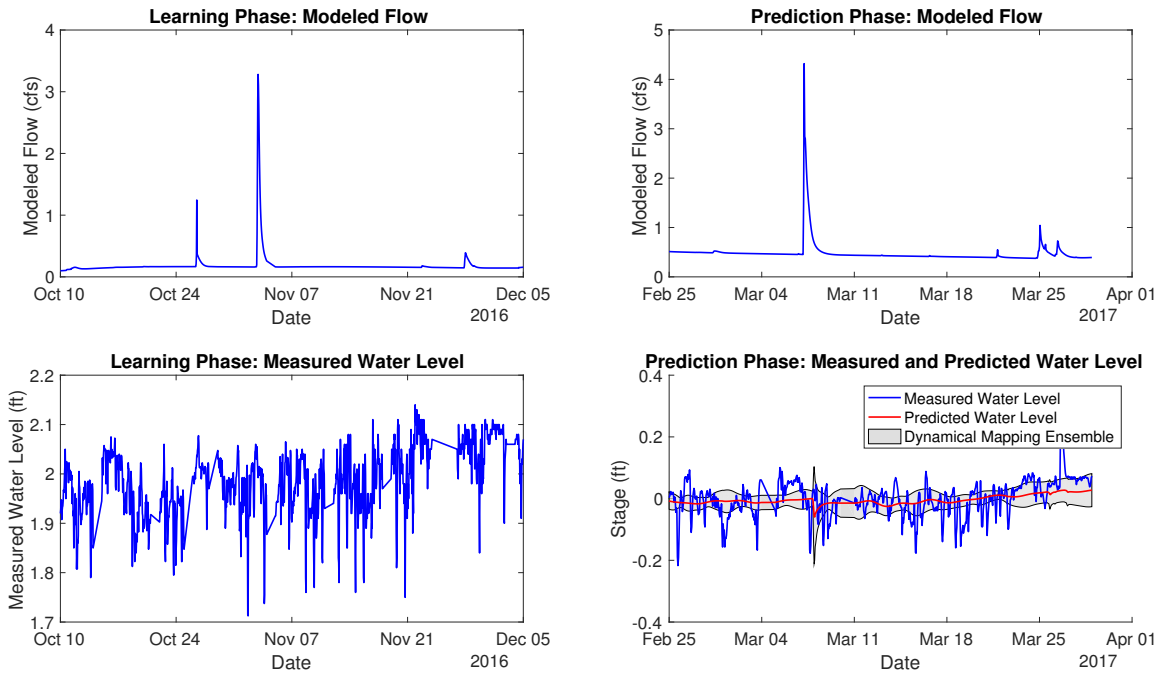


Figure B.122: Site SCHQSTCR01

SENGLSH01, Mean nRMSE = 22.75%

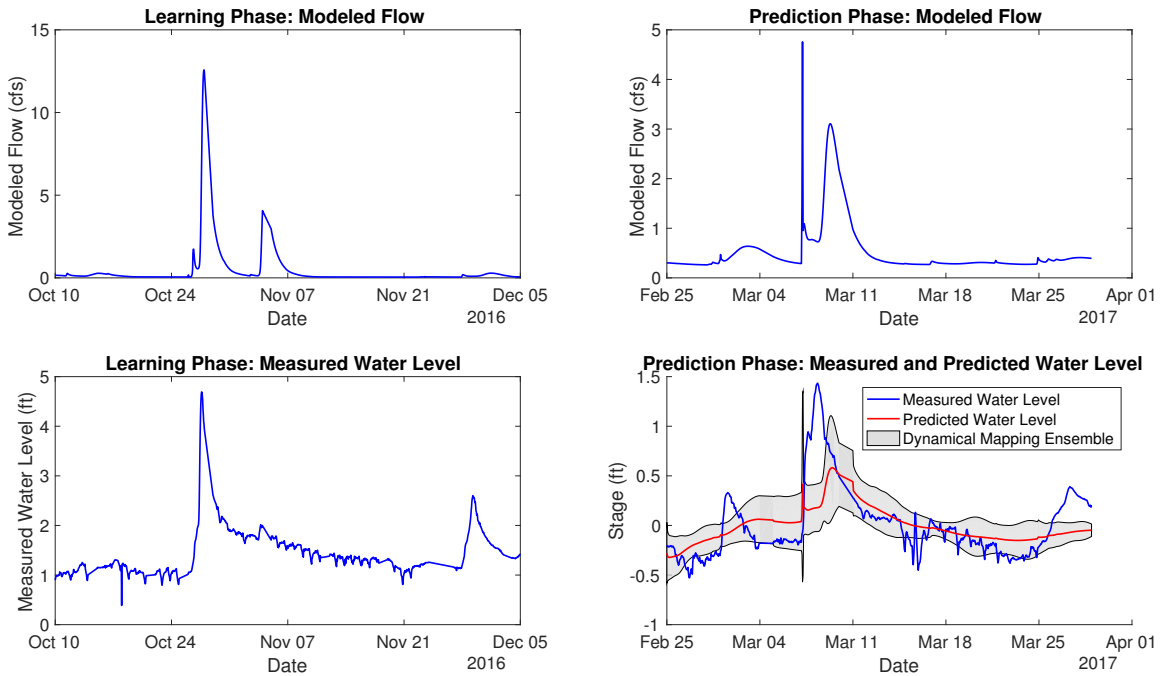


Figure B.123: Site SENGLSH01

SFIOWA01, Mean nRMSE = -59.45%

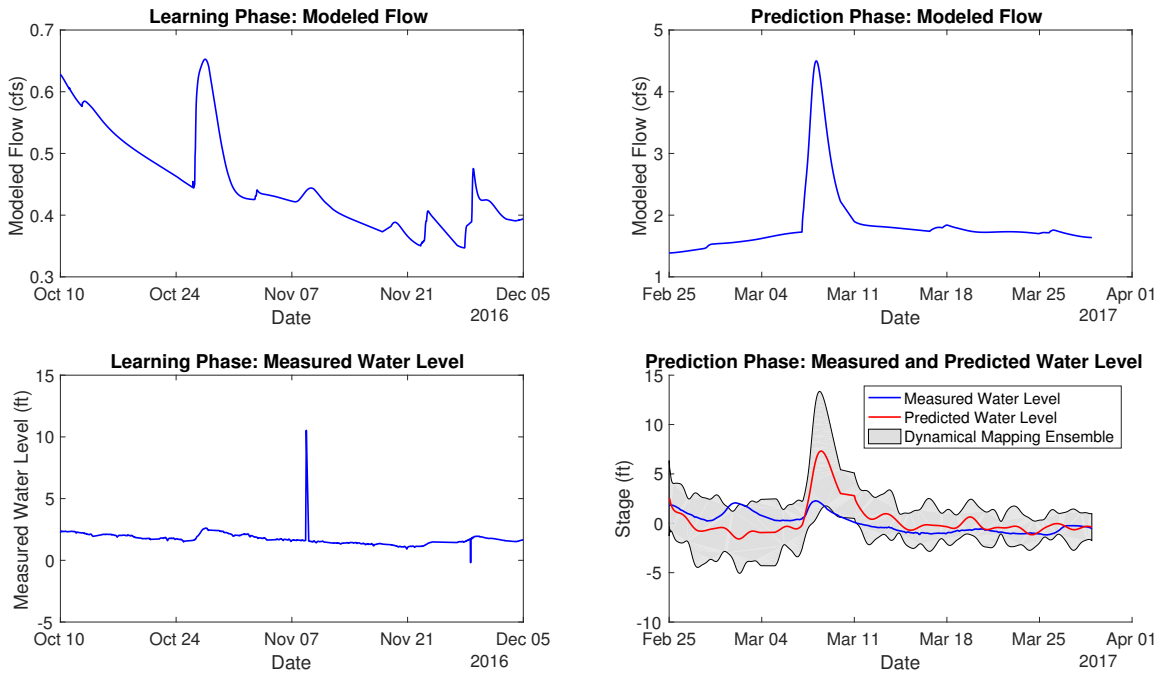


Figure B.124: Site SFIOWA01

SFIOWA03, Mean nRMSE = -3.95%

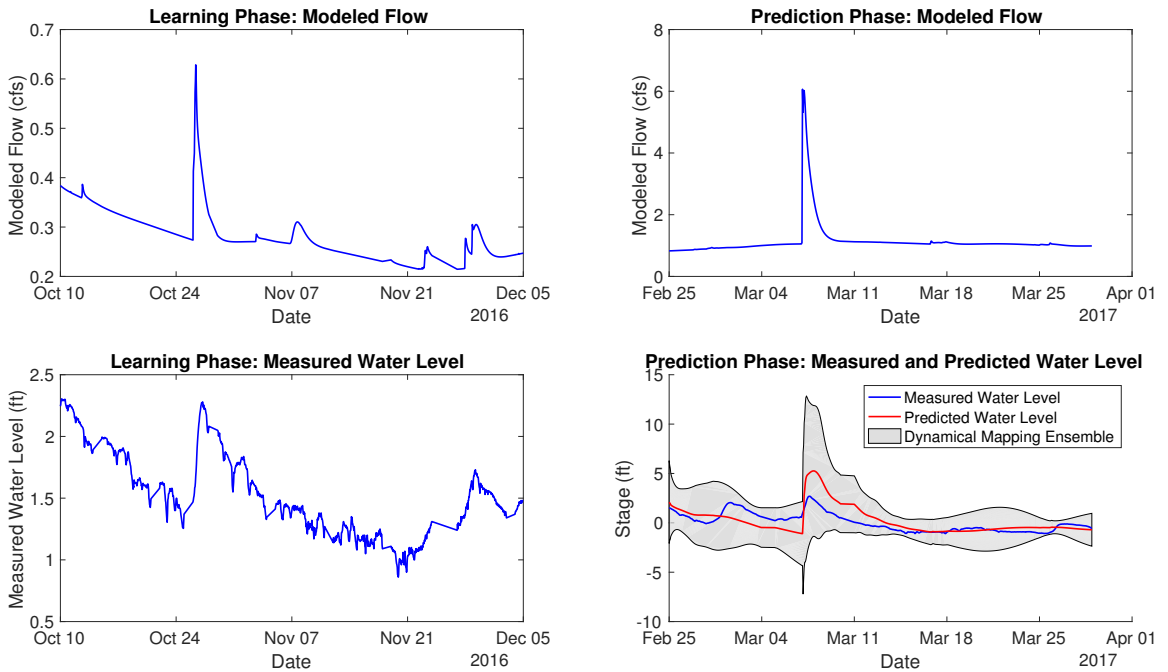


Figure B.125: Site SFIOWA03

SHLRK01, Mean nRMSE = -46.67%

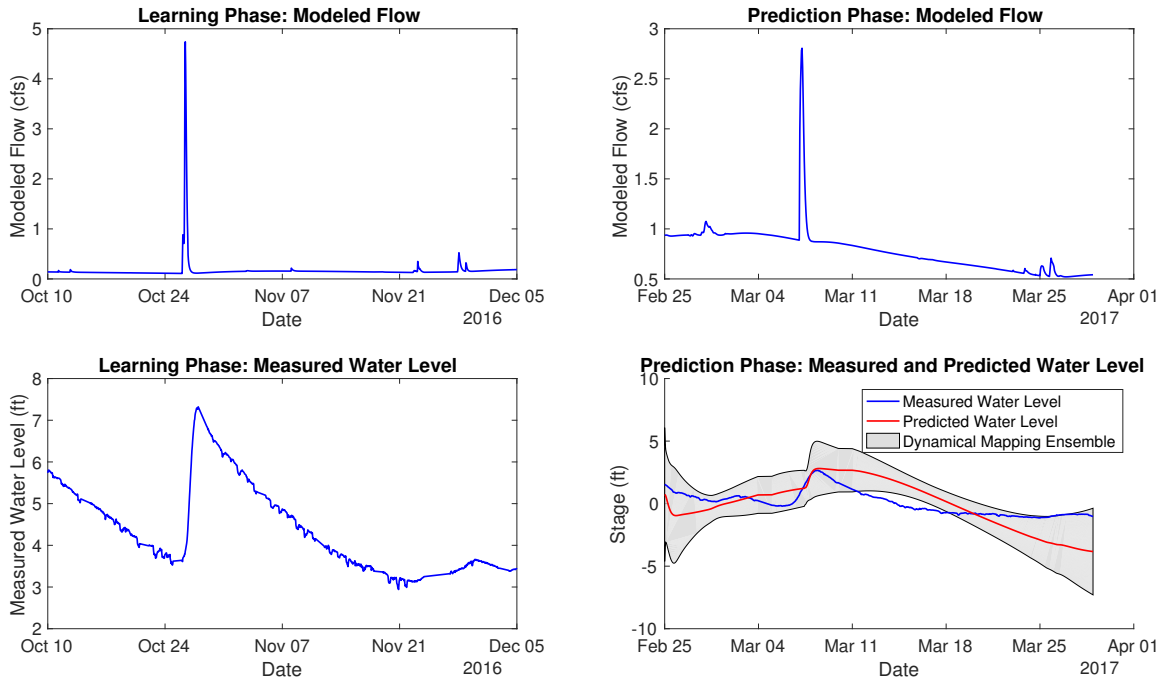


Figure B.126: Site SHLRK01

SHLRK02, Mean nRMSE = 46.71%

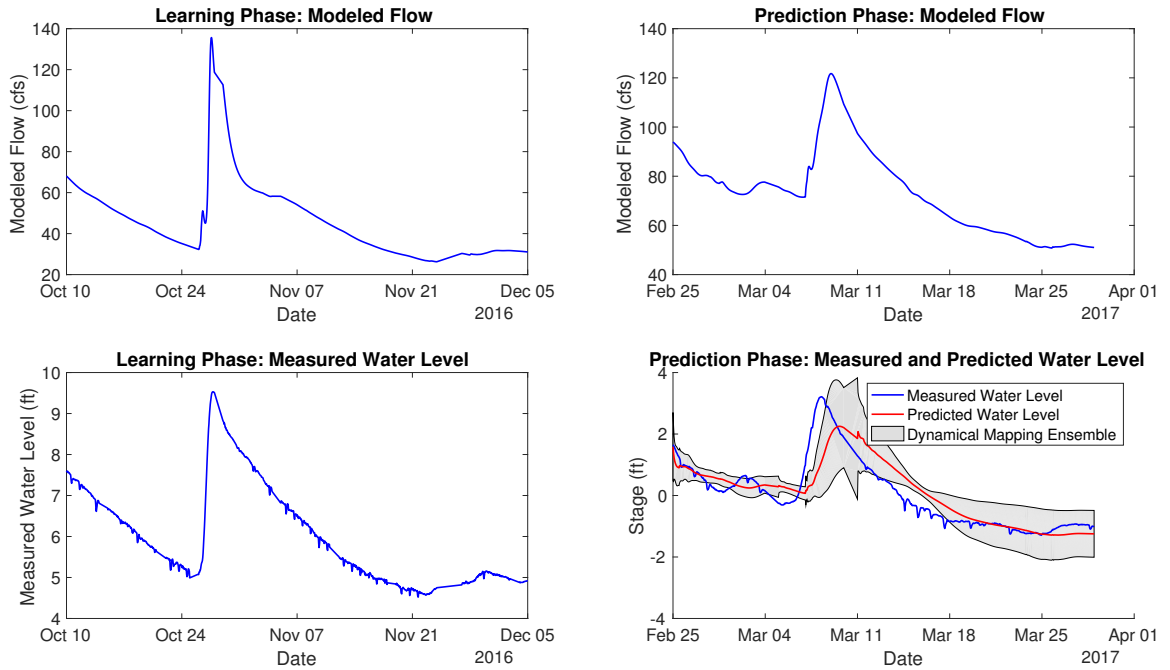


Figure B.127: Site SHLRK02

SHLRK03, Mean nRMSE = 52.52%

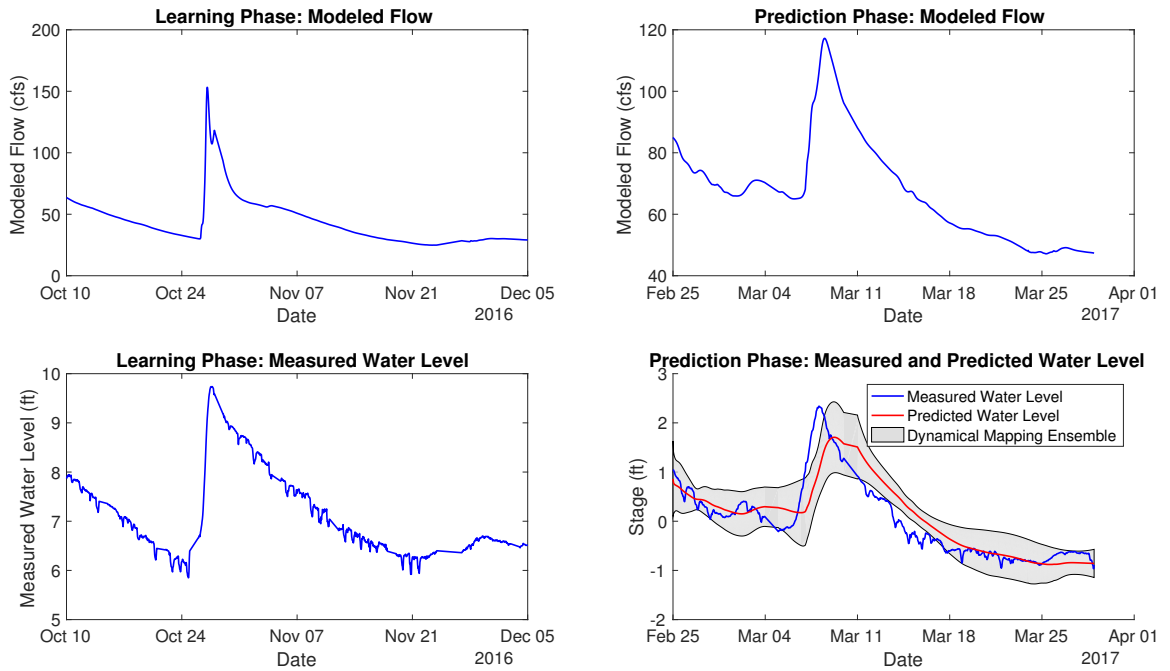


Figure B.128: Site SHLRK03

SHLRK04, Mean nRMSE = 40.12%

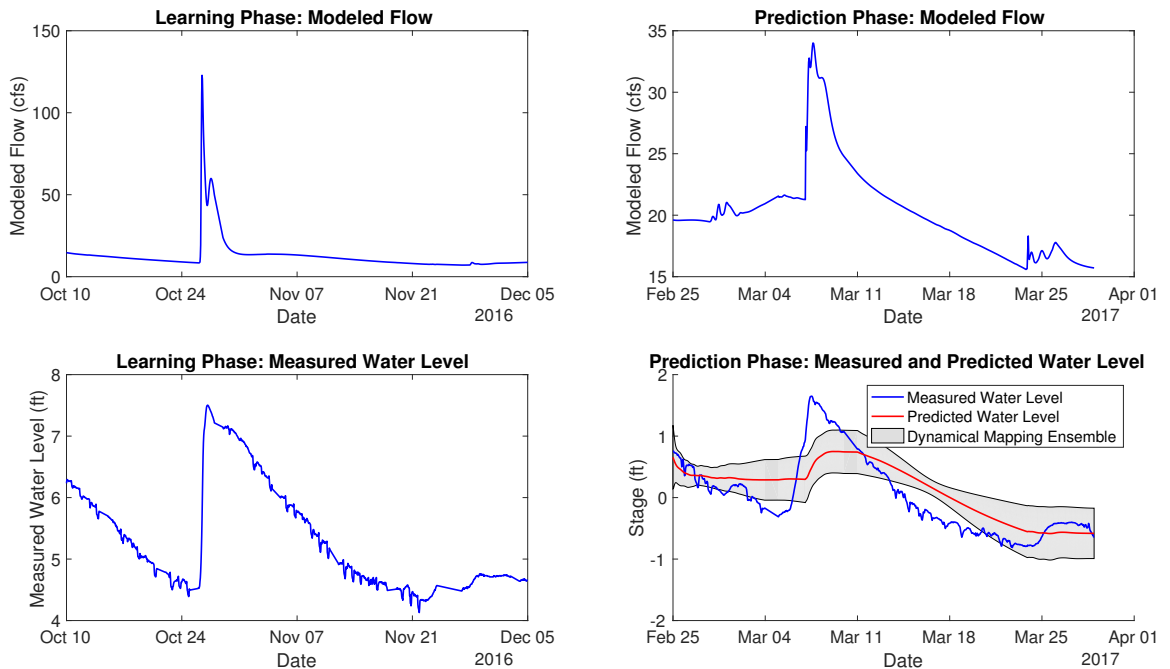


Figure B.129: Site SHLRK04

SHLRK05, Mean nRMSE = 41.48%

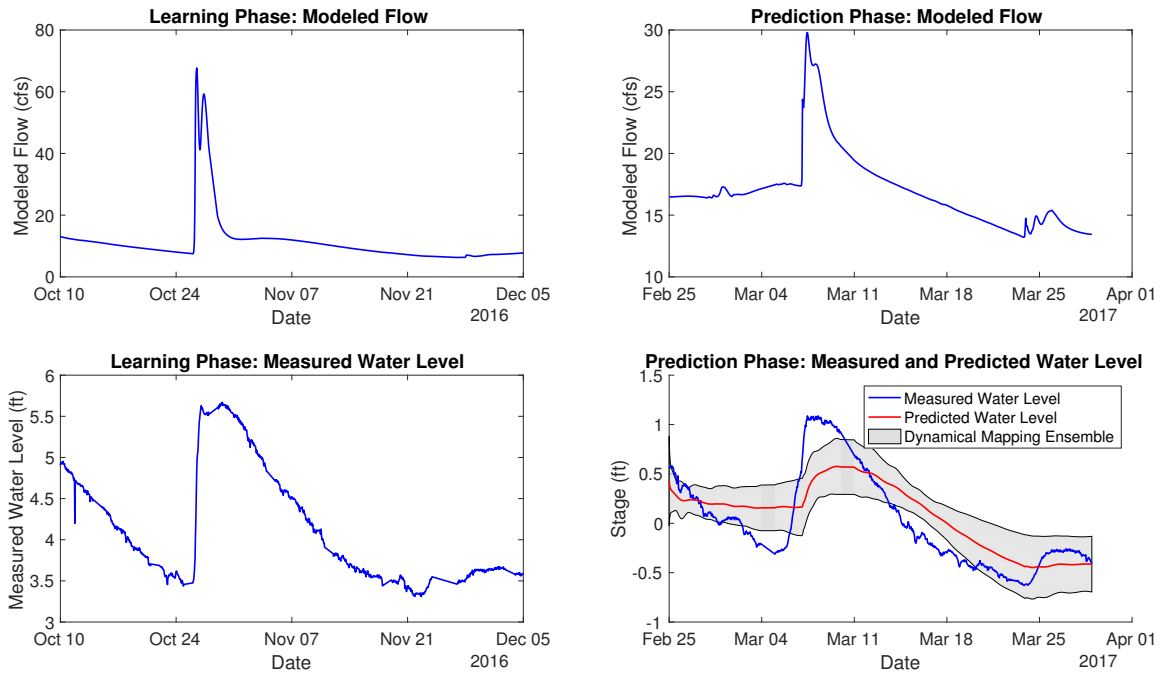


Figure B.130: Site SHLRK05

SLOUGHCR01, Mean nRMSE = 46.53%

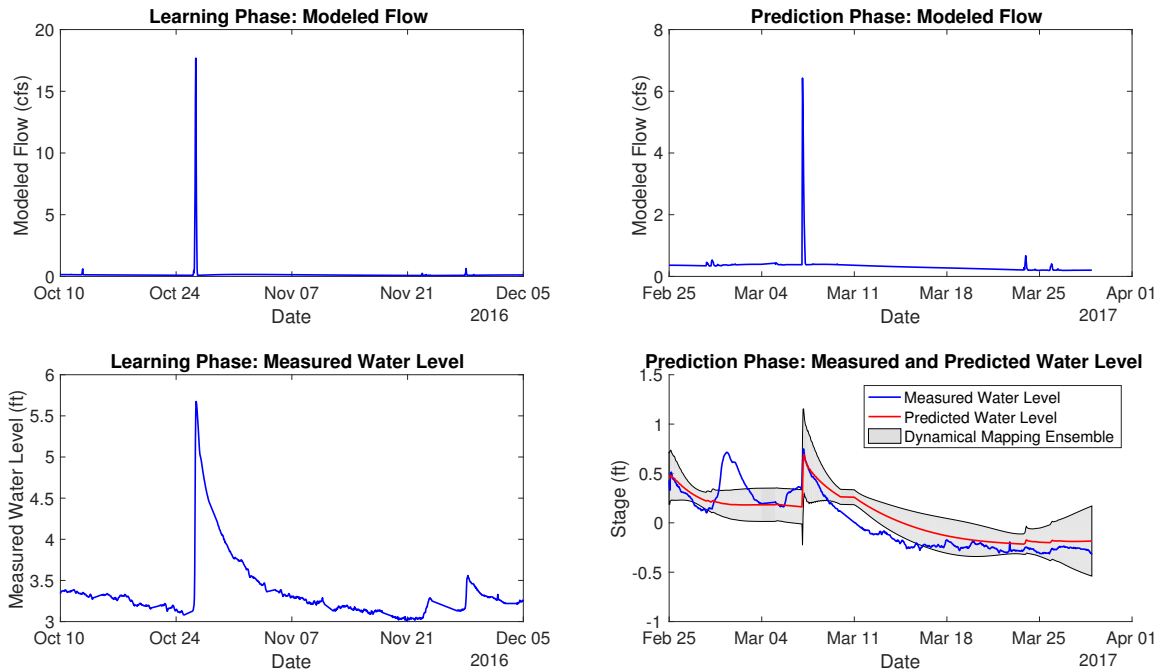


Figure B.131: Site SLOUGHCR01

SLVRCR01, Mean nRMSE = -34.44%

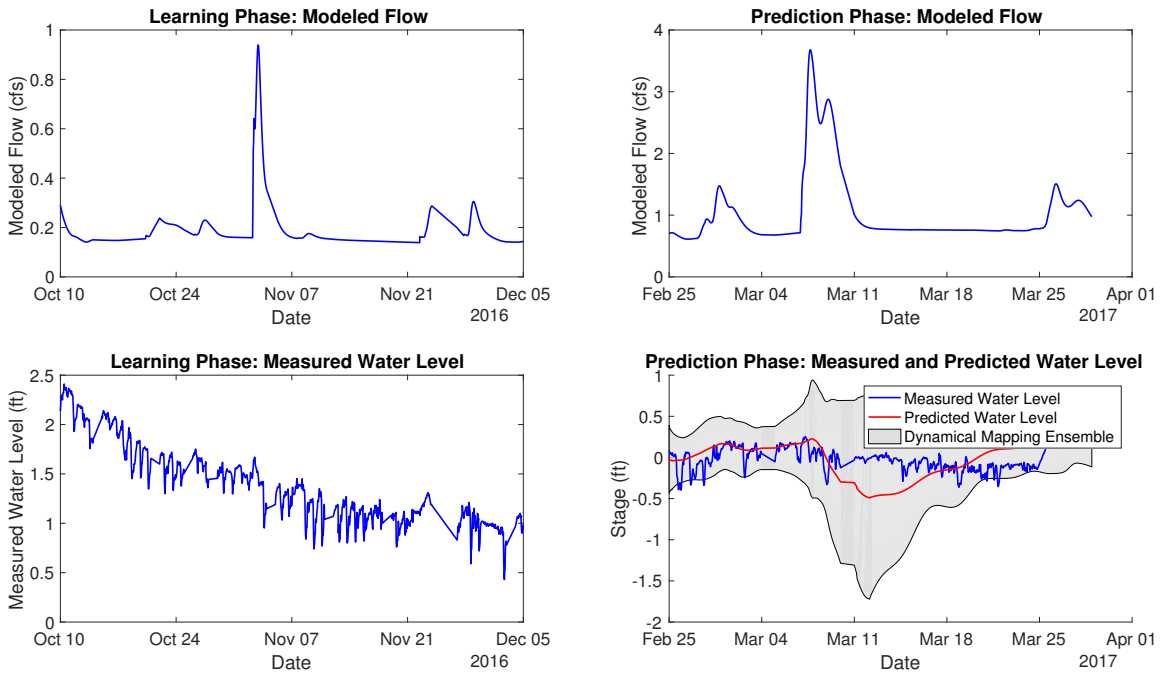


Figure B.132: Site SLVRCR01

SOAPCR01, Mean nRMSE = 12.50%

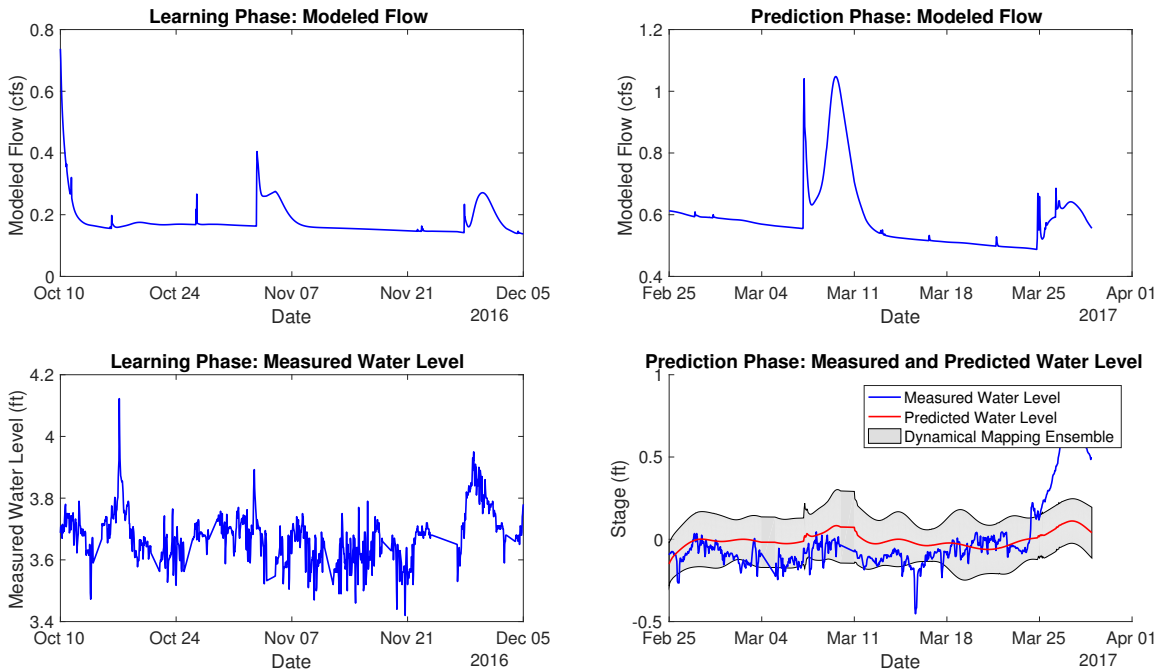


Figure B.133: Site SOAPCR01

SOAPCR03, Mean nRMSE = -57.26%

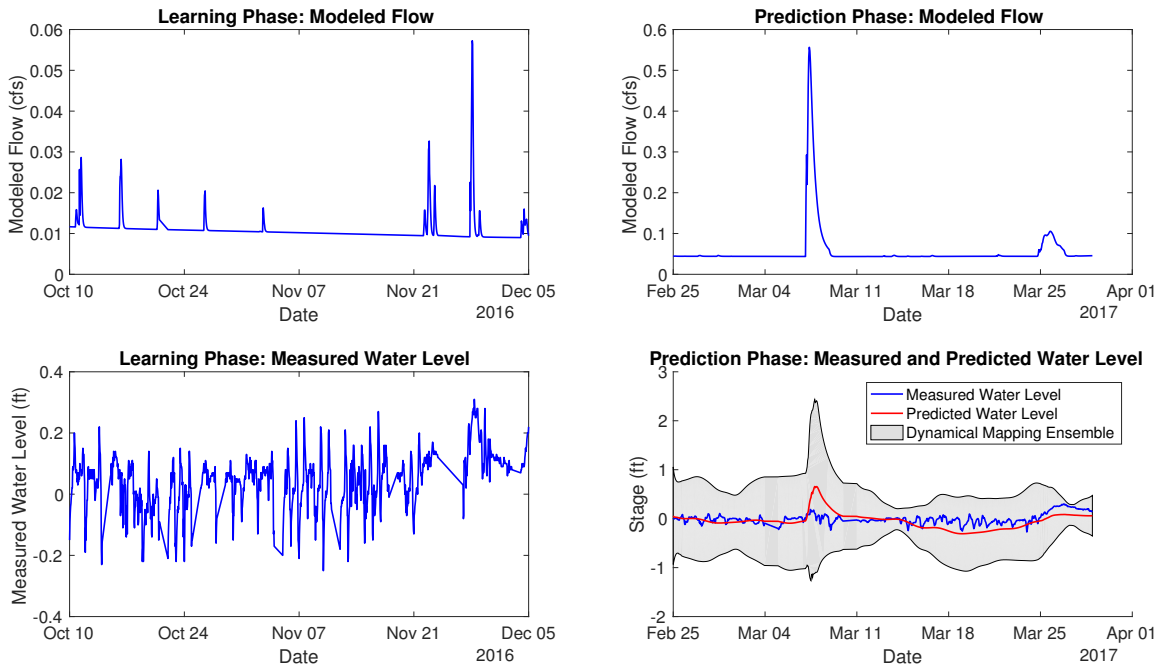


Figure B.134: Site SOAPCR03

SPRINGCR01, Mean nRMSE = 18.54%

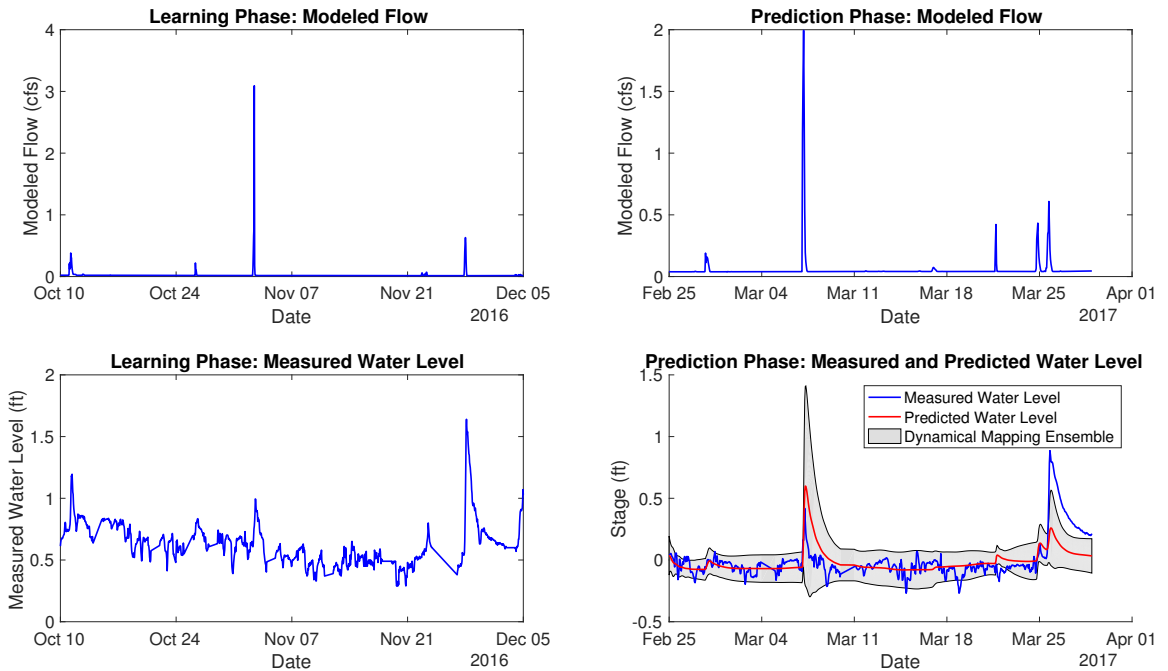


Figure B.135: Site SPRINGCR01

SPRNGCRK01, Mean nRMSE = 29.67%

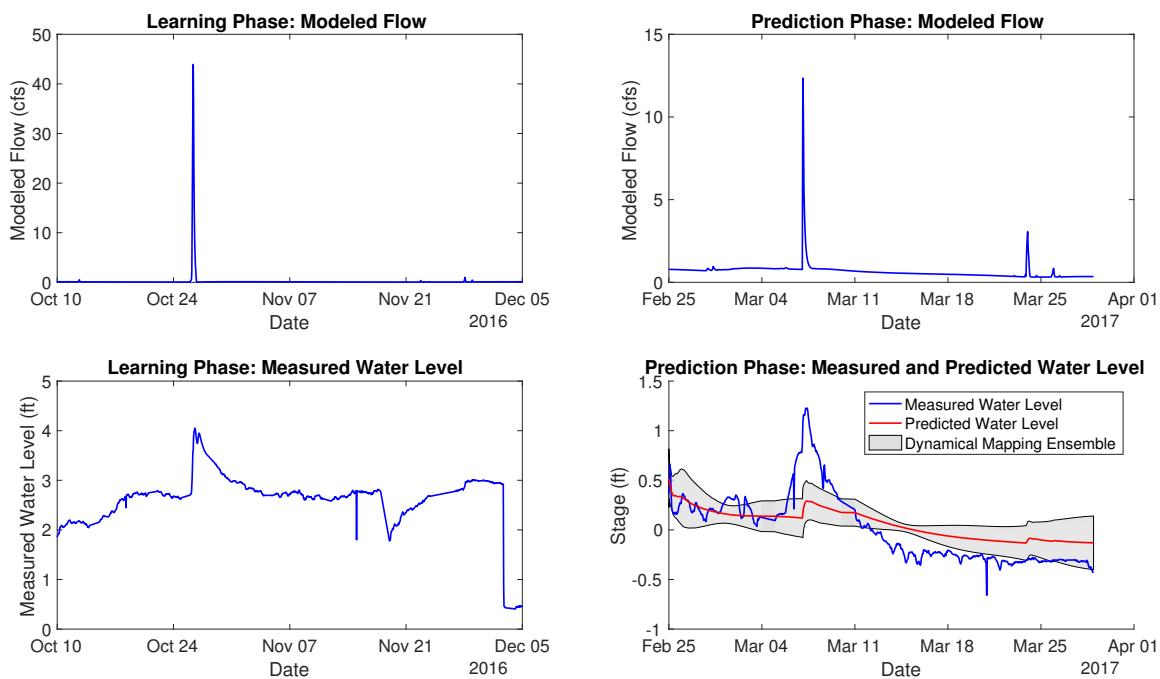


Figure B.136: Site SPRNGCRK01

SQWCR01, Mean nRMSE = 72.74%

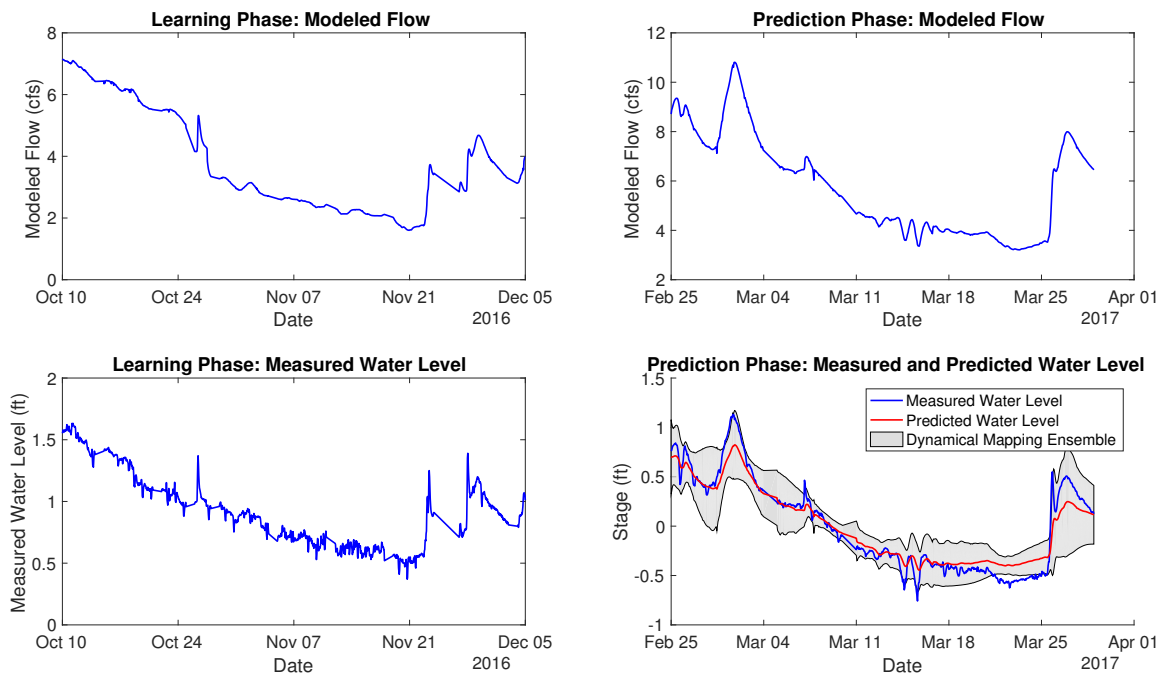


Figure B.137: Site SQWCR01

SQWCR02, Mean nRMSE = 48.24%

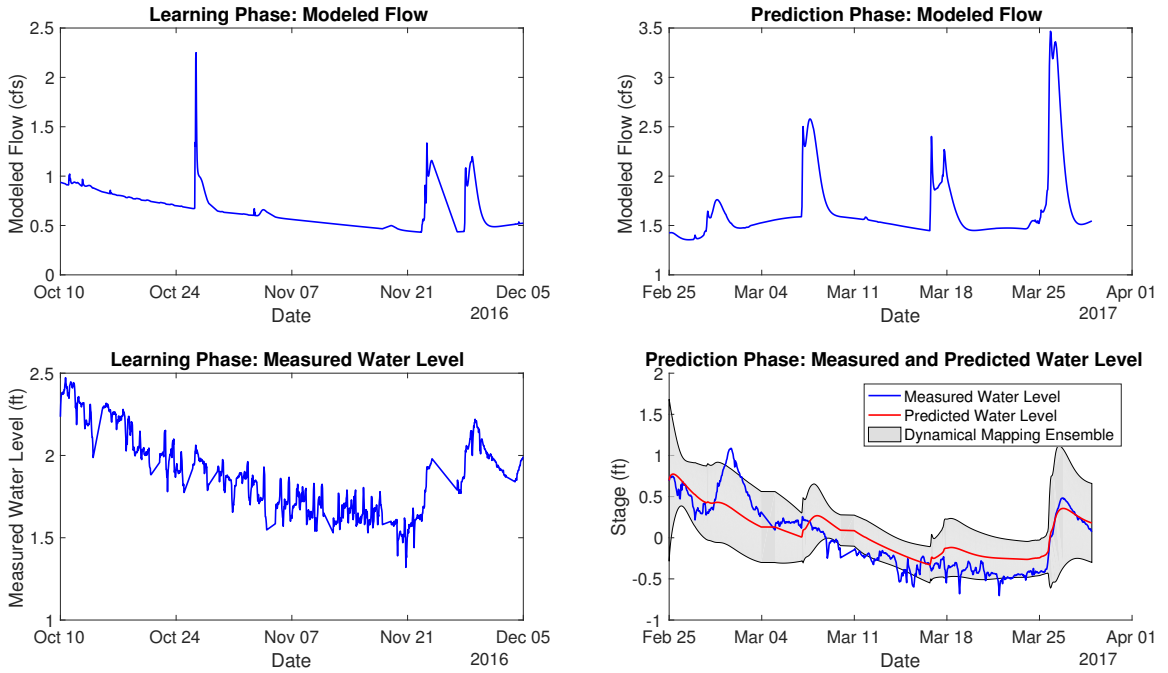


Figure B.138: Site SQWCR02

SQWCR03, Mean nRMSE = 46.32%

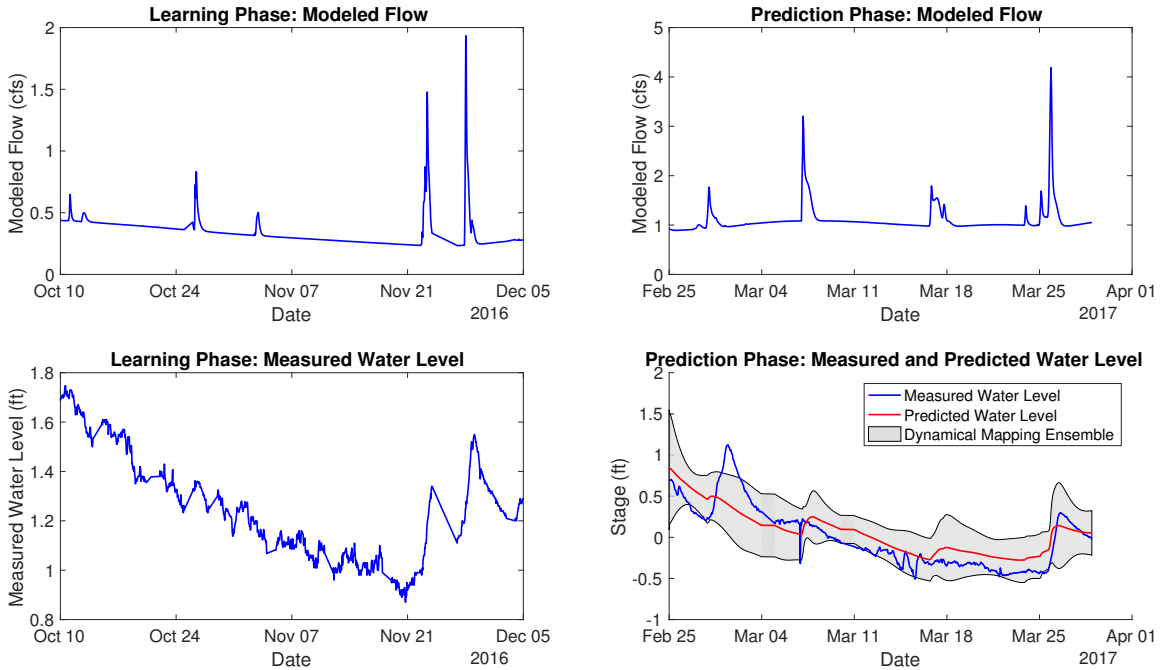


Figure B.139: Site SQWCR03

SQWCR04, Mean nRMSE = 31.29%

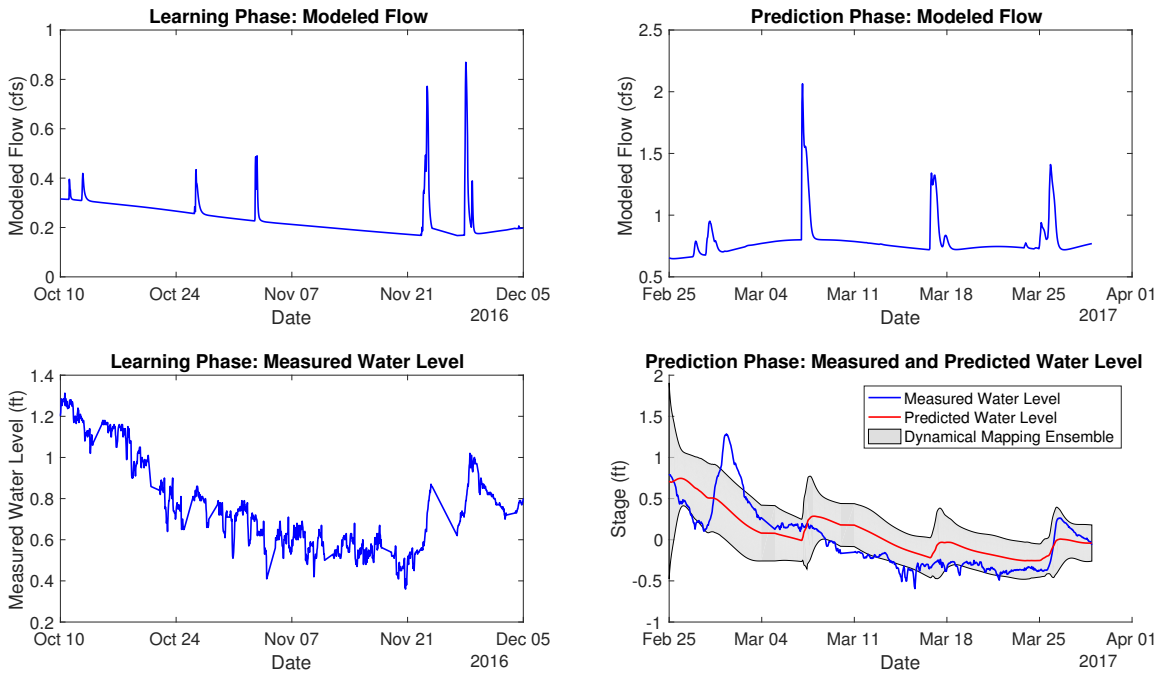


Figure B.140: Site SQWCR04

SRCCNRV01, Mean nRMSE = 13.90%

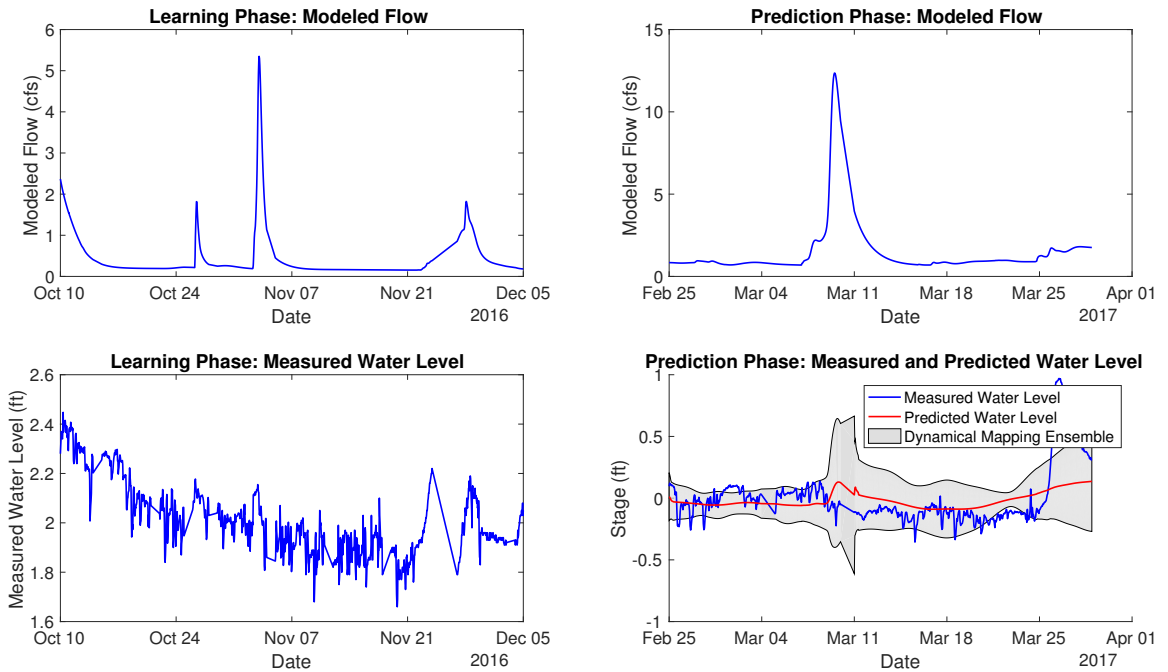


Figure B.141: Site SRCCNRV01

SSKNK01, Mean nRMSE = 67.86%

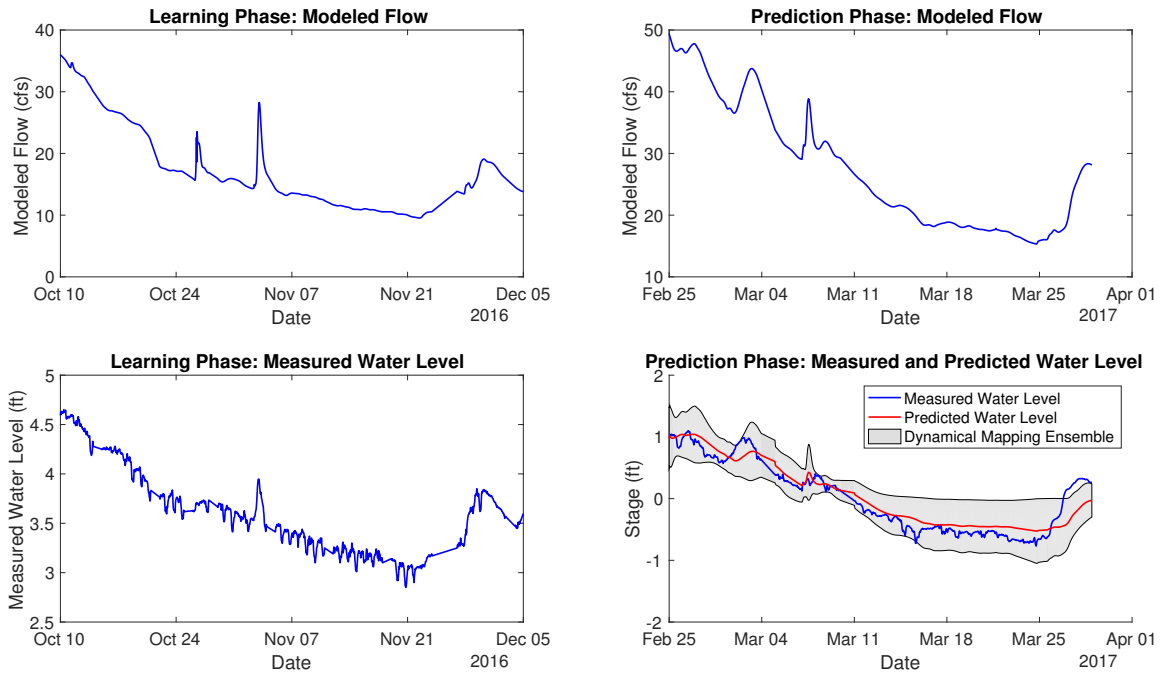


Figure B.142: Site SSKNK01

SSKNK02, Mean nRMSE = 78.34%

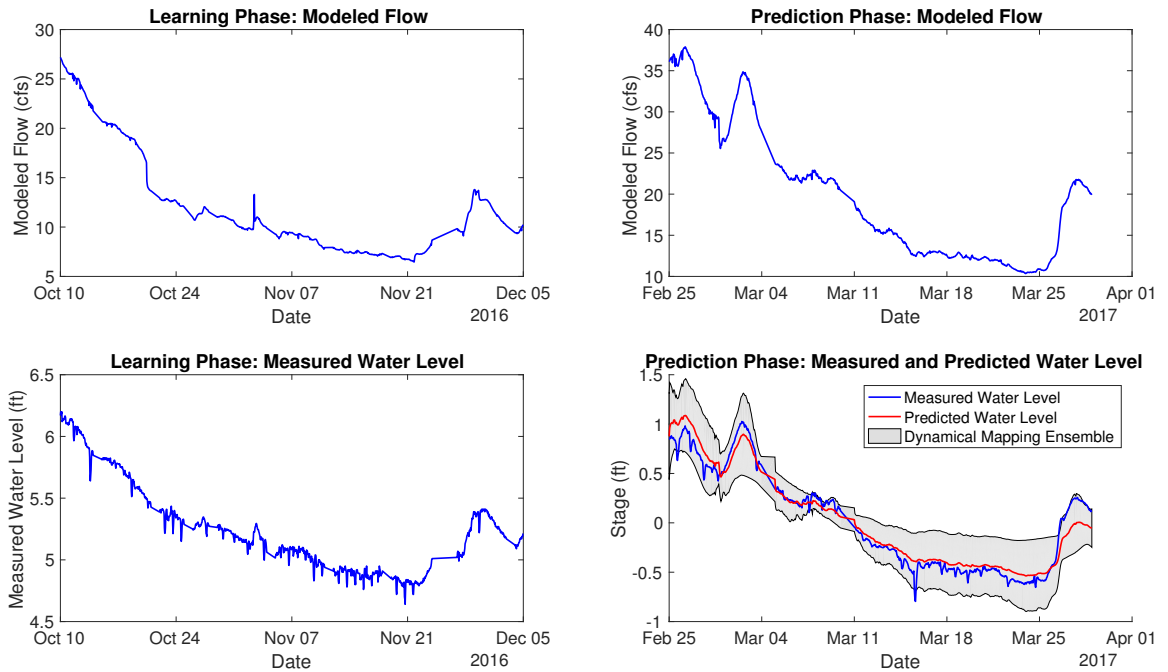


Figure B.143: Site SSKNK02

SSKNK03, Mean nRMSE = 74.86%

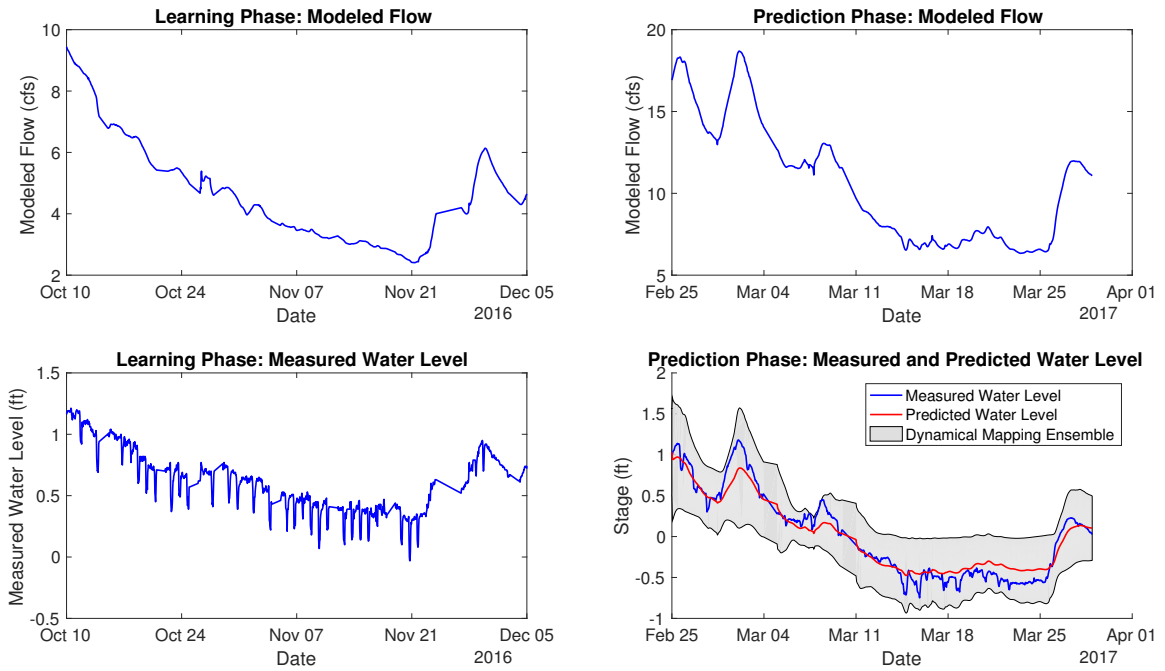


Figure B.144: Site SSKNK03

SSKNK04, Mean nRMSE = 28.59%

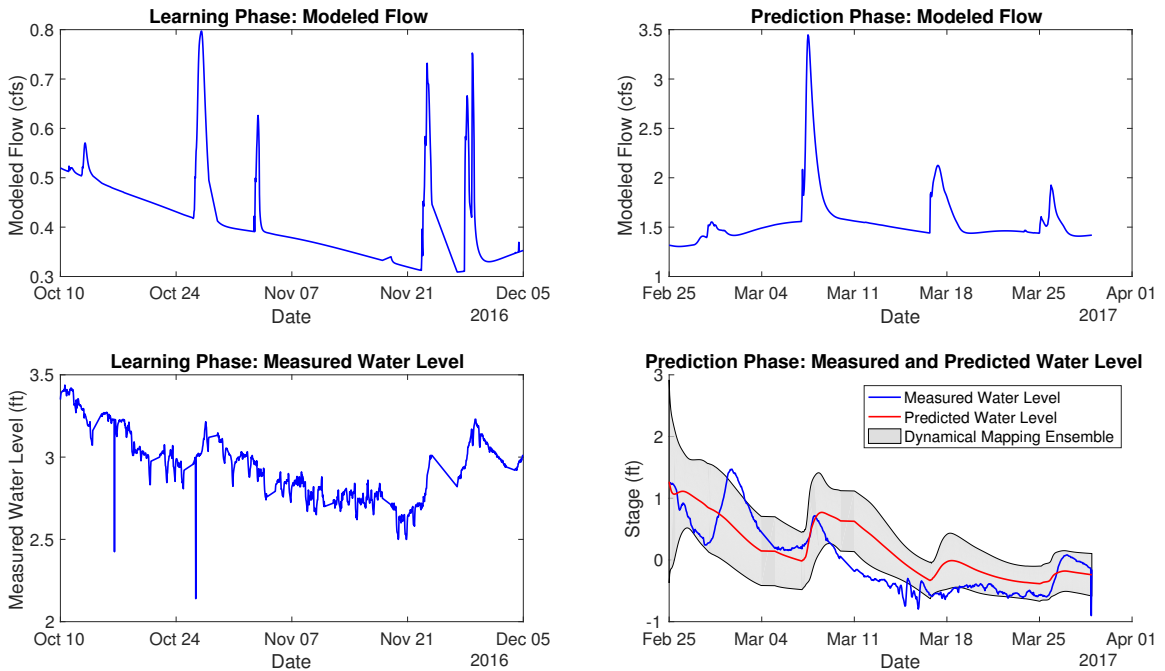


Figure B.145: Site SSKNK04

SUGARCR01, Mean nRMSE = 10.25%

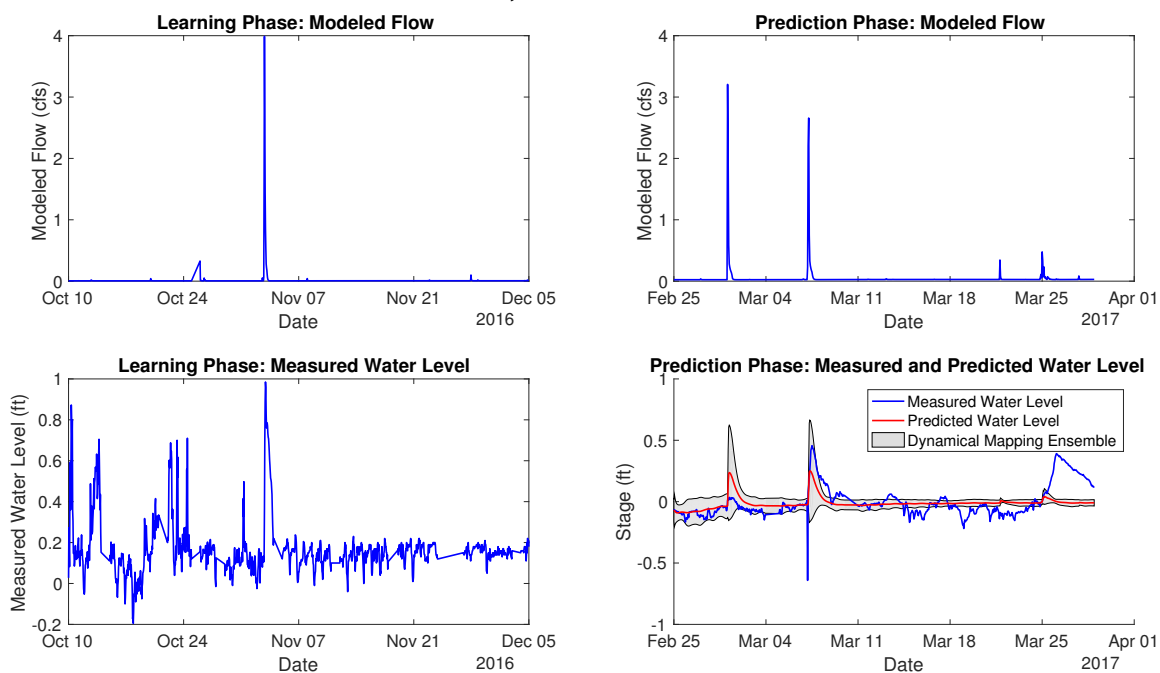


Figure B.146: Site SUGARCR01

TIPTNCR01, Mean nRMSE = 66.89%

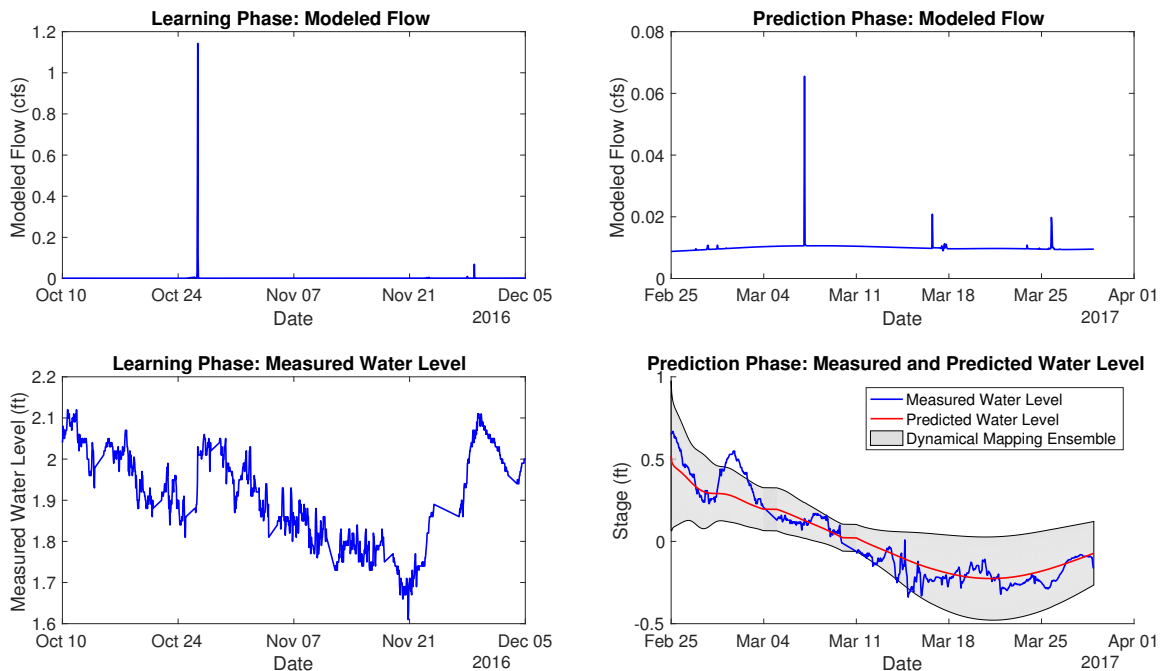


Figure B.147: Site TIPTNCR01

TIPTNCR02, Mean nRMSE = 30.14%

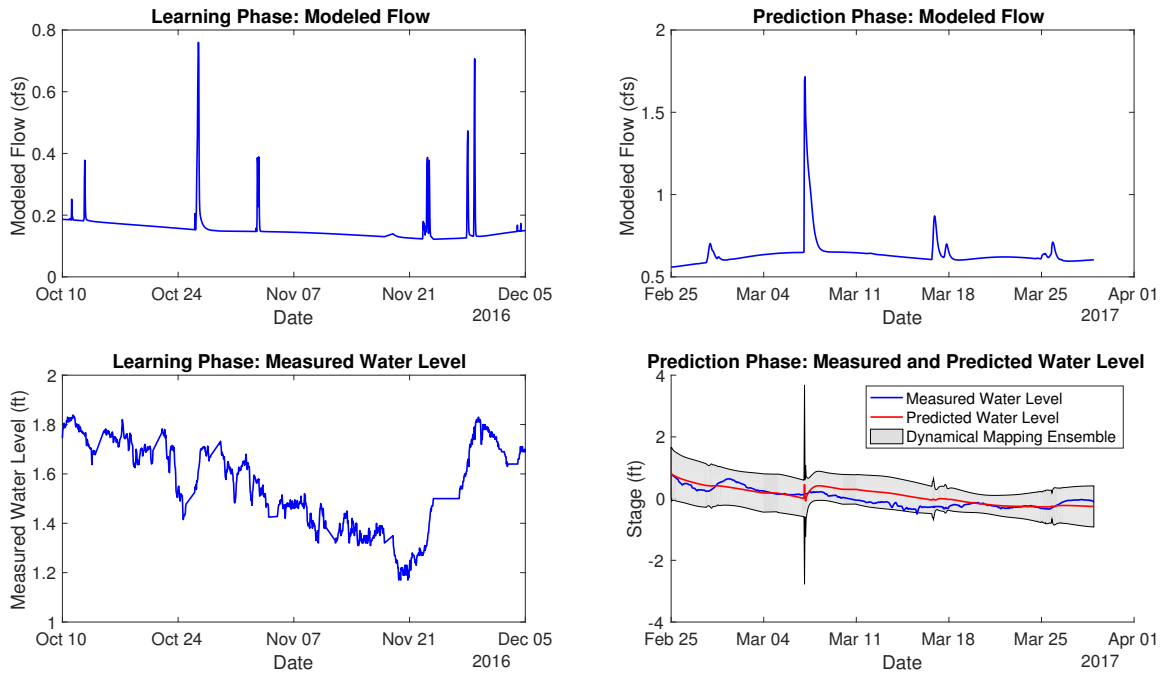


Figure B.148: Site TIPTNCR02

TKYRV01, Mean nRMSE = 46.98%

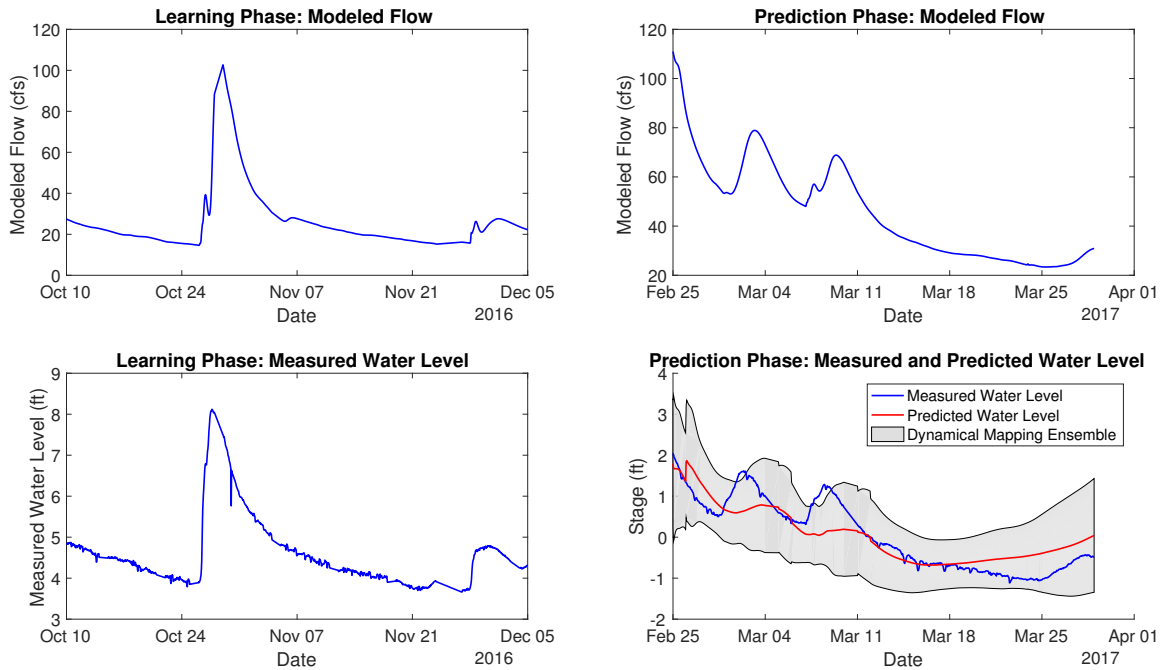


Figure B.149: Site TKYRV01

TKYRV03, Mean nRMSE = 40.68%

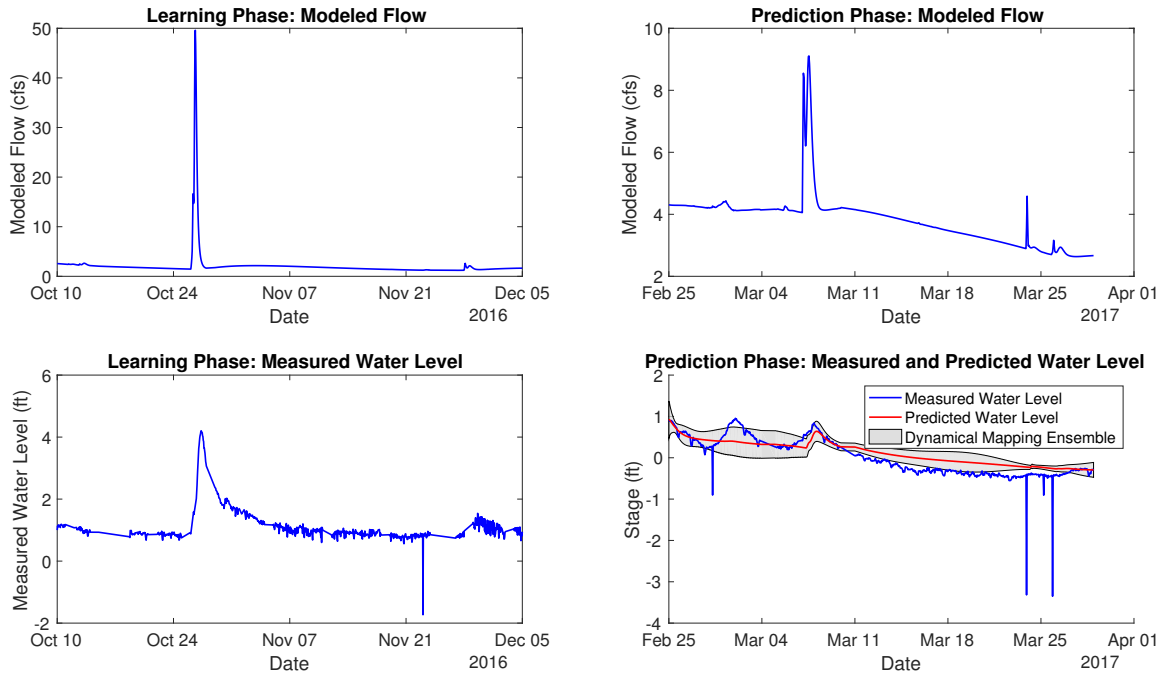


Figure B.150: Site TKYRV03

TRBLSM01, Mean nRMSE = 8.44%

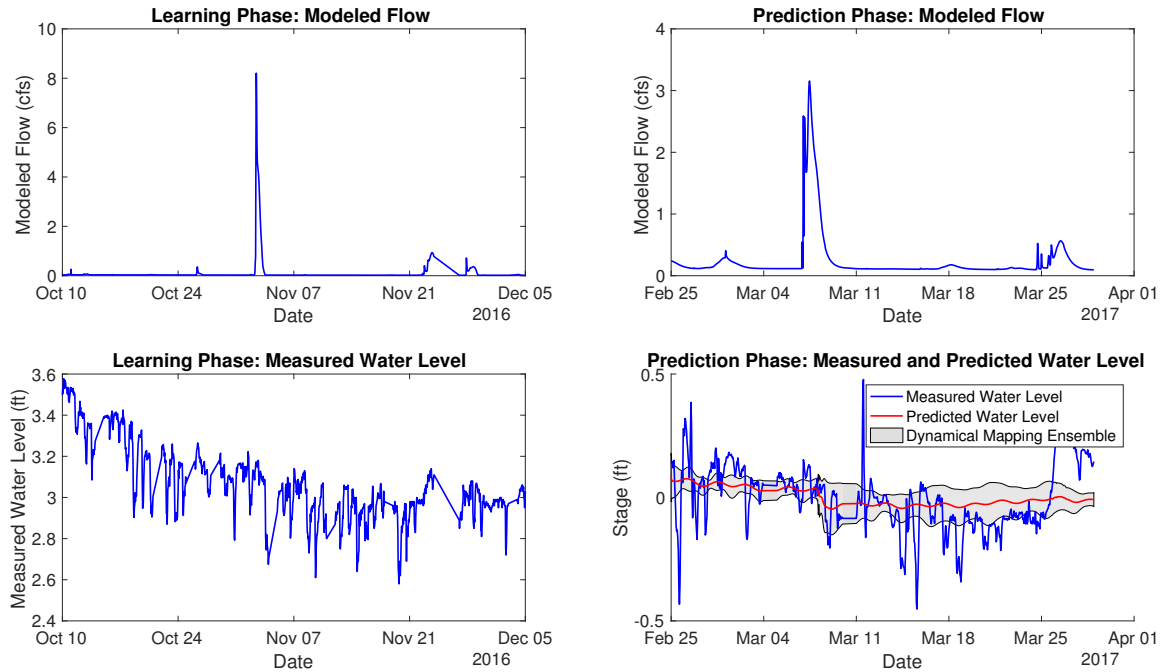


Figure B.151: Site TRBLSM01

UIWARV01, Mean nRMSE = 80.70%

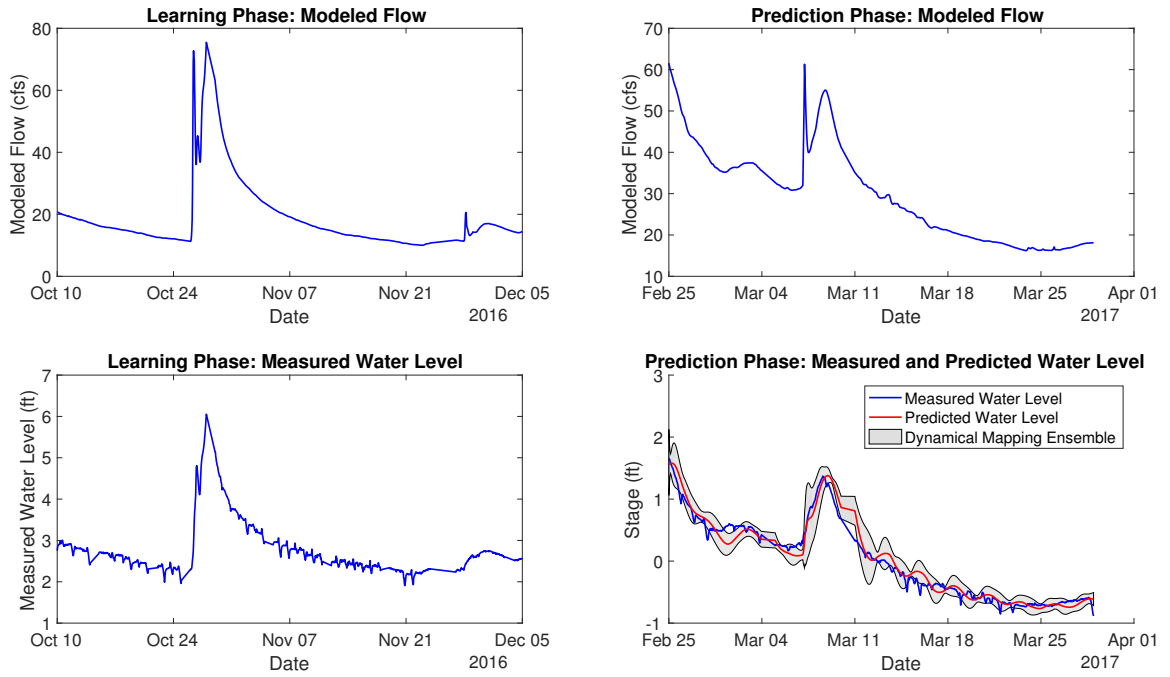


Figure B.152: Site UIWARV01

VLGARV01, Mean nRMSE = 25.68%

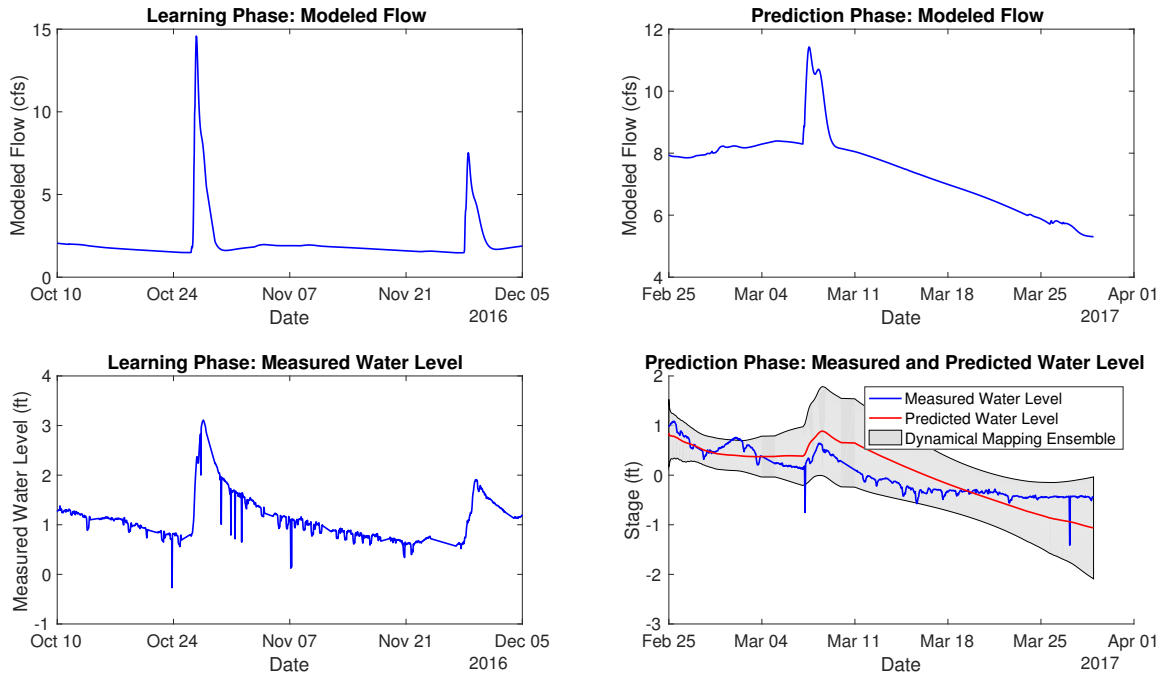


Figure B.153: Site VLGARV01

VLGARV02, Mean nRMSE = 2.02%

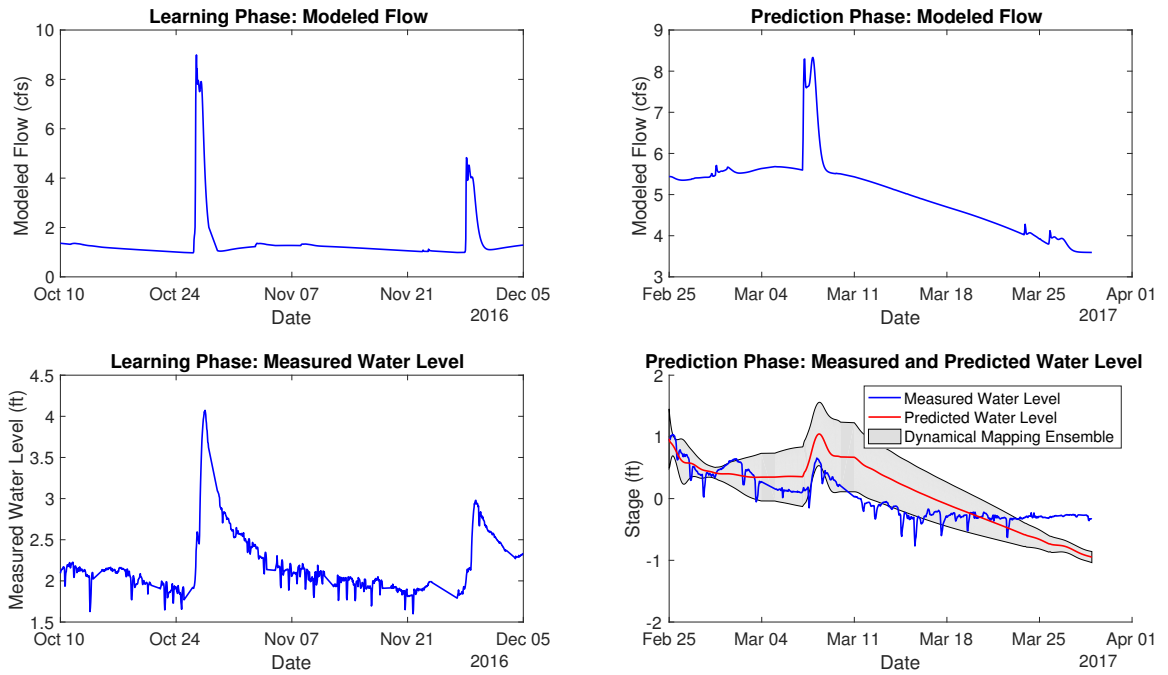


Figure B.154: Site VLGARV02

VLGARV03, Mean nRMSE = -18.34%

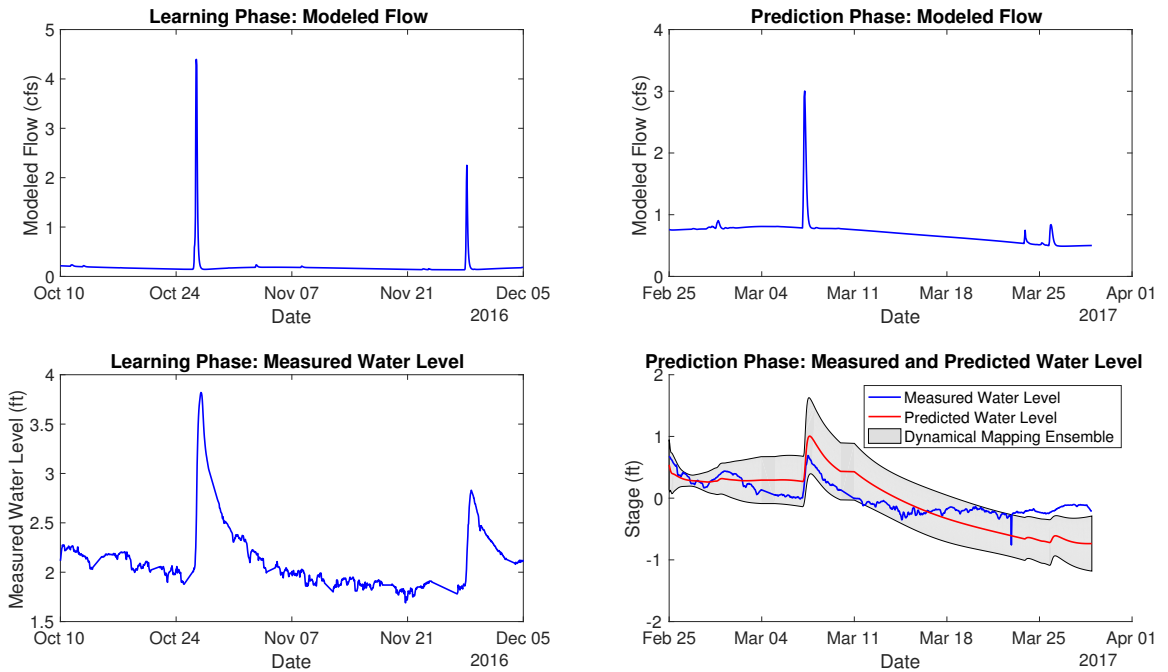


Figure B.155: Site VLGARV03

WALNUTCR01, Mean nRMSE = 26.86%

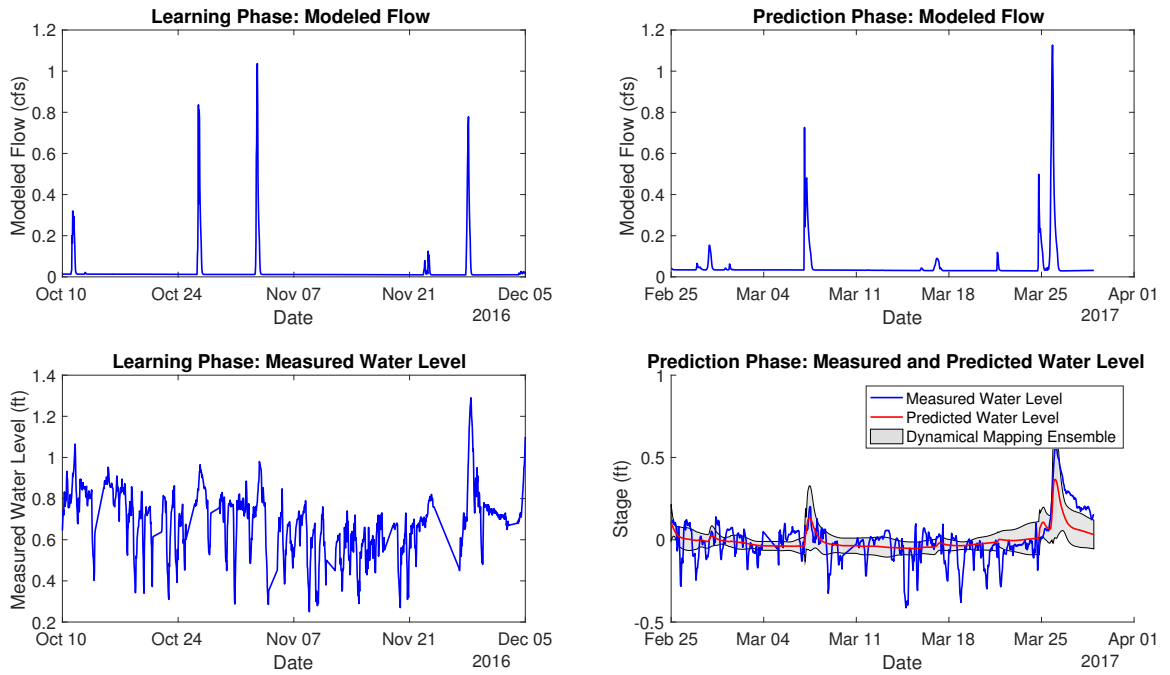


Figure B.156: Site WALNUTCR01

WAPSI02, Mean nRMSE = 19.65%

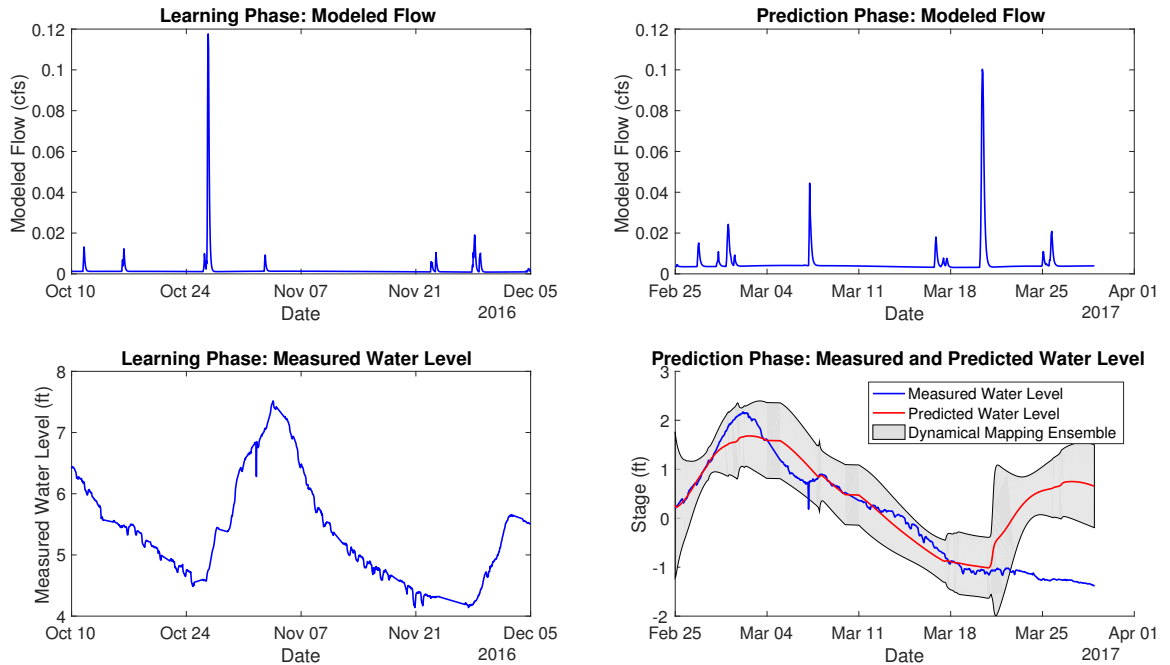


Figure B.157: Site WAPSI02

WAPSI03, Mean nRMSE = 37.47%

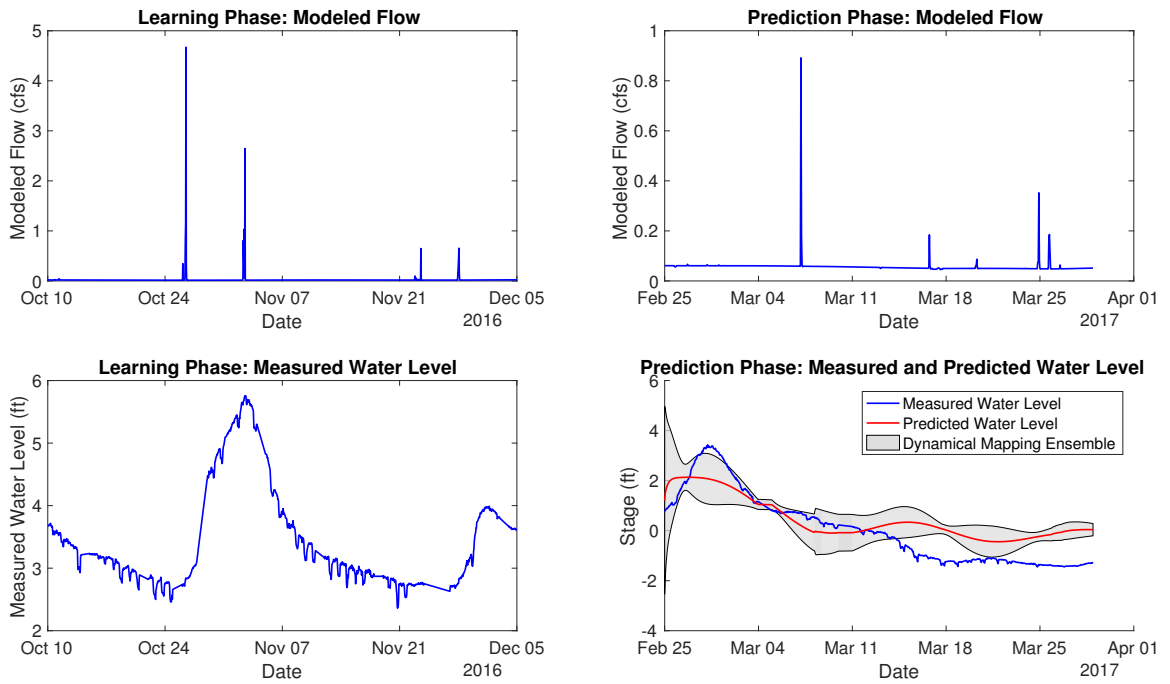


Figure B.158: Site WAPSI03

WBFLOYDRV01, Mean nRMSE = -186.49%

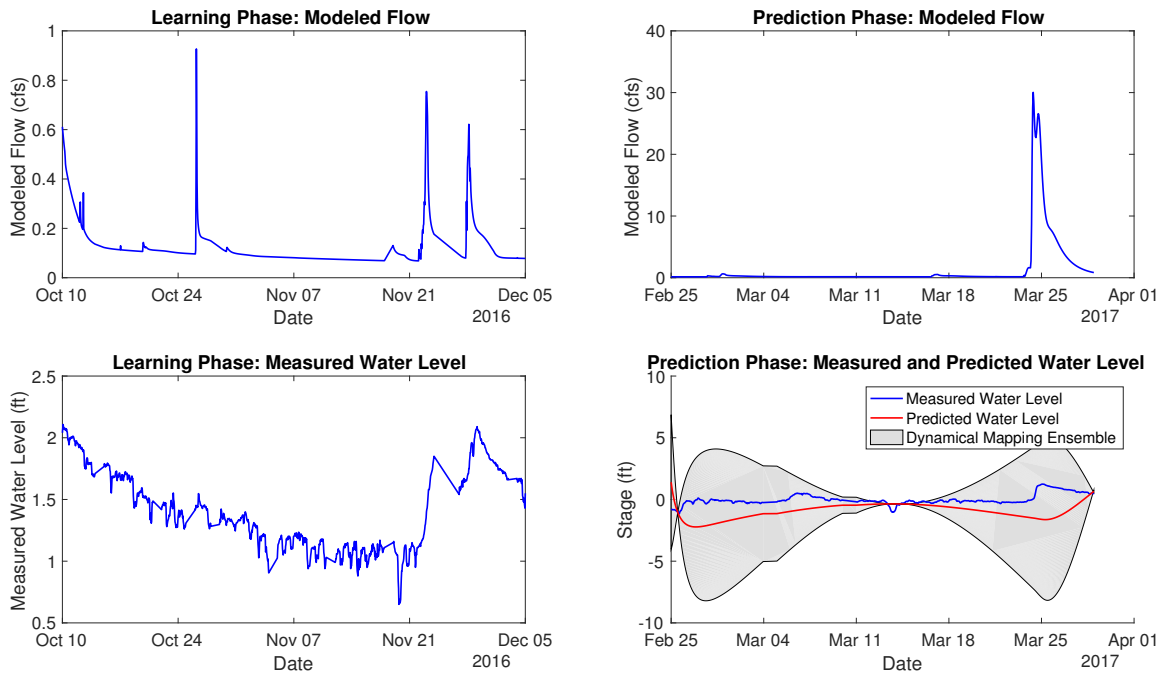


Figure B.159: Site WBFLOYDRV01

WFKLTSIOUX01, Mean nRMSE = 36.07%

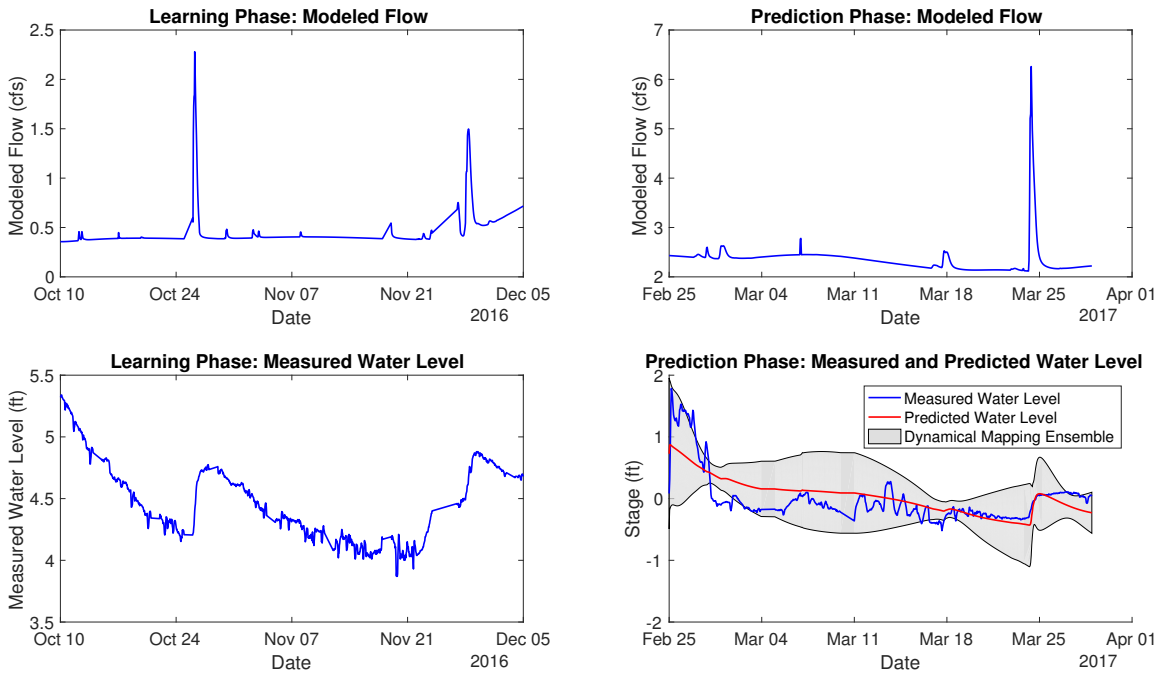


Figure B.160: Site WFKLTSIOUX01

WFLTSIOUX01, Mean nRMSE = 22.98%

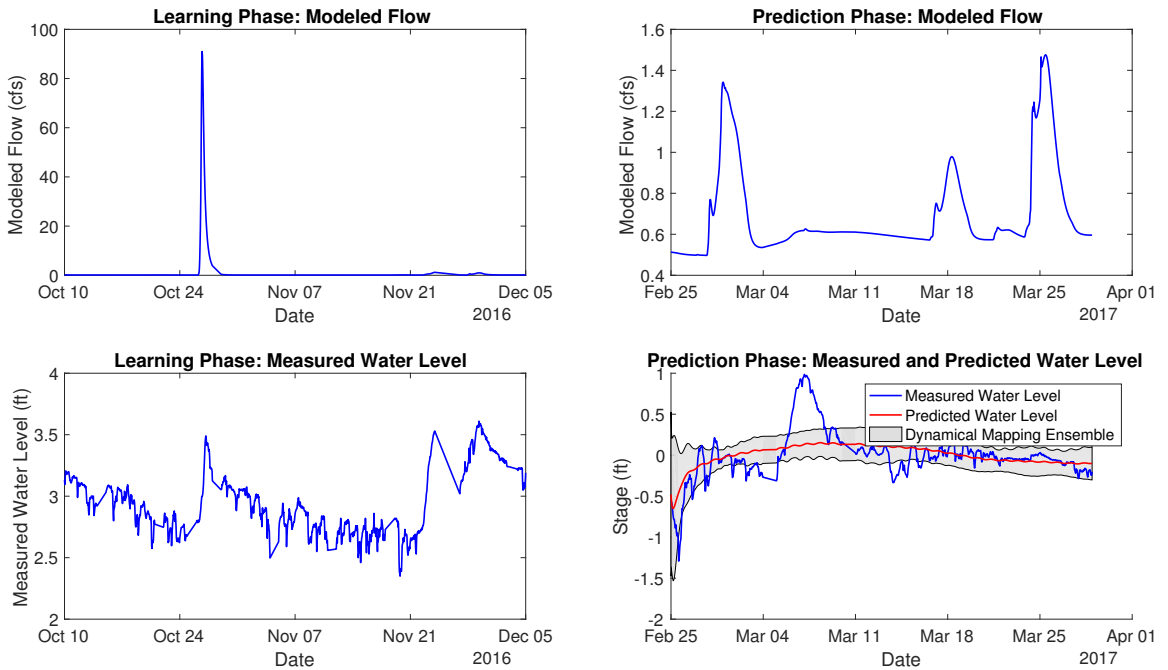


Figure B.161: Site WFLTSIOUX01

WLLWCR02, Mean nRMSE = 37.69%

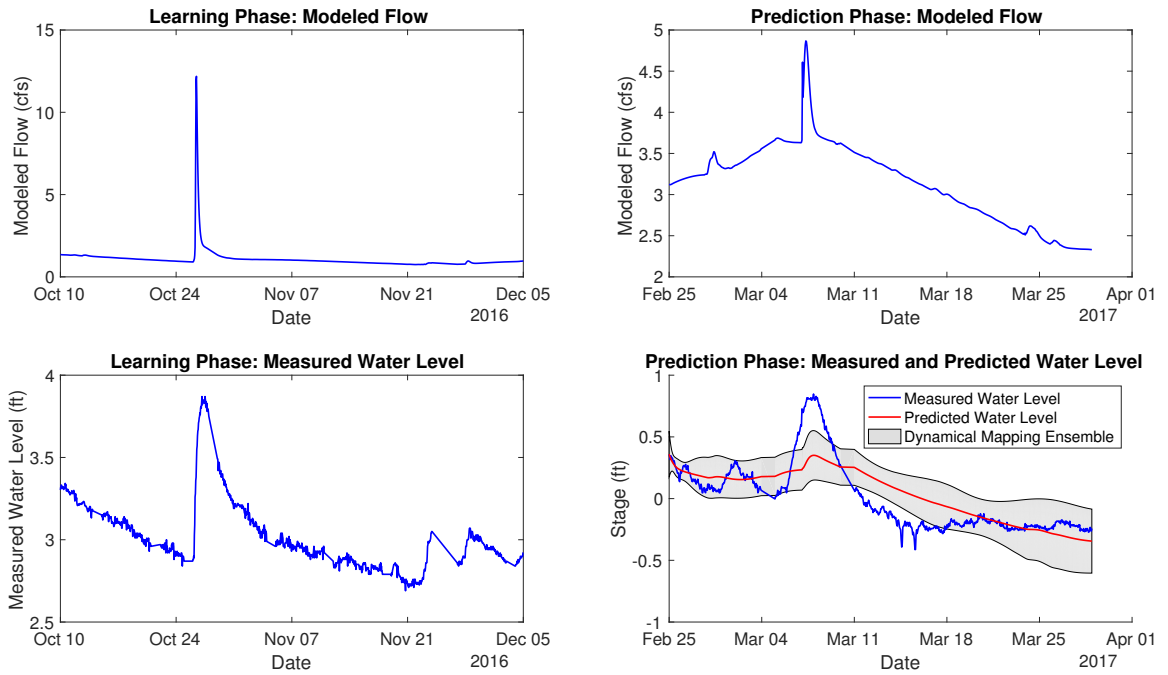


Figure B.162: Site WLLWCR02

WLNTCR01, Mean nRMSE = 5.28%

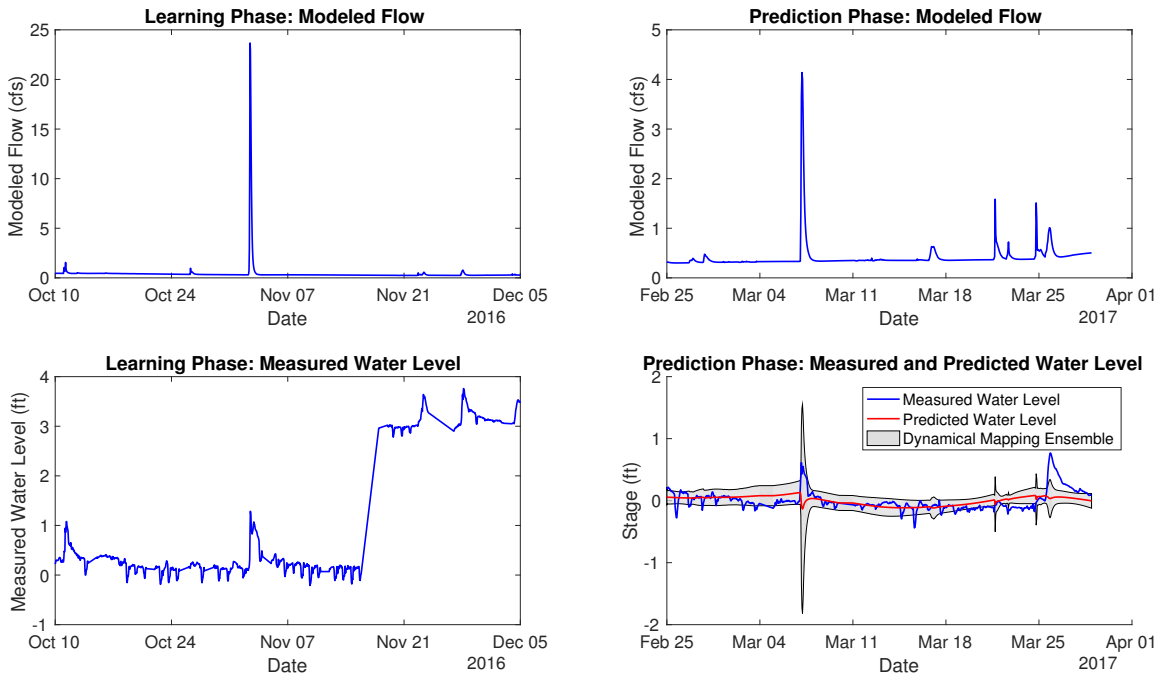


Figure B.163: Site WLNTCR01

WLNTRK01, Mean nRMSE = 50.63%

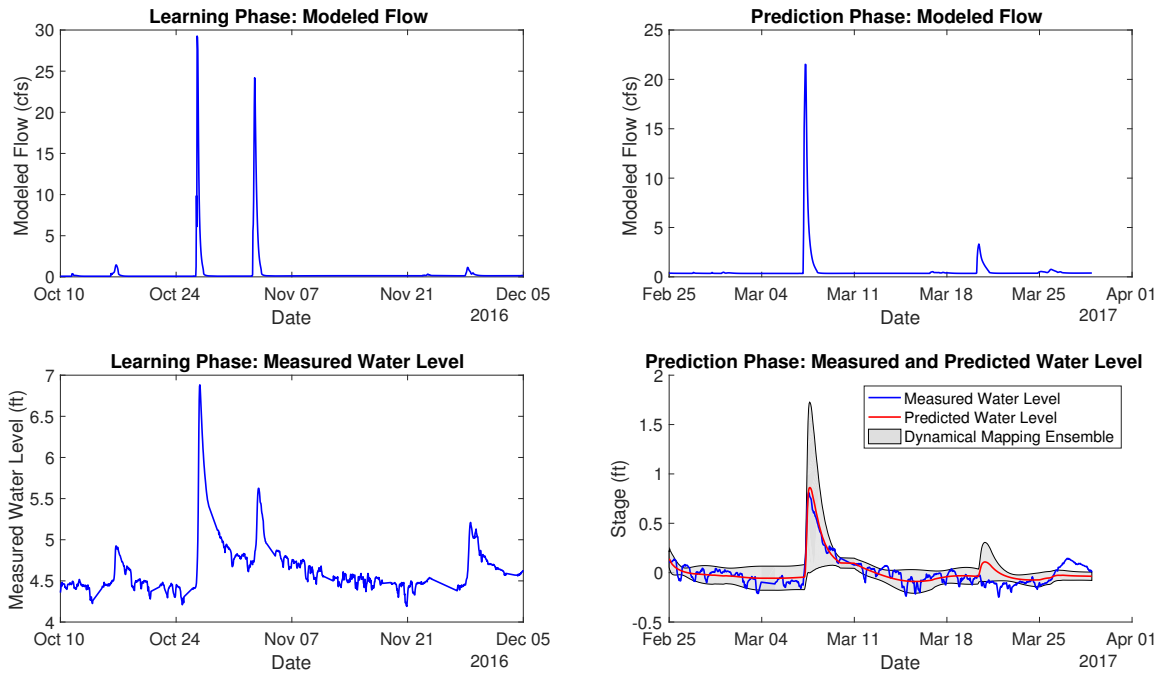


Figure B.164: Site WLNTRK01

WNBGORV03, Mean nRMSE = 22.27%

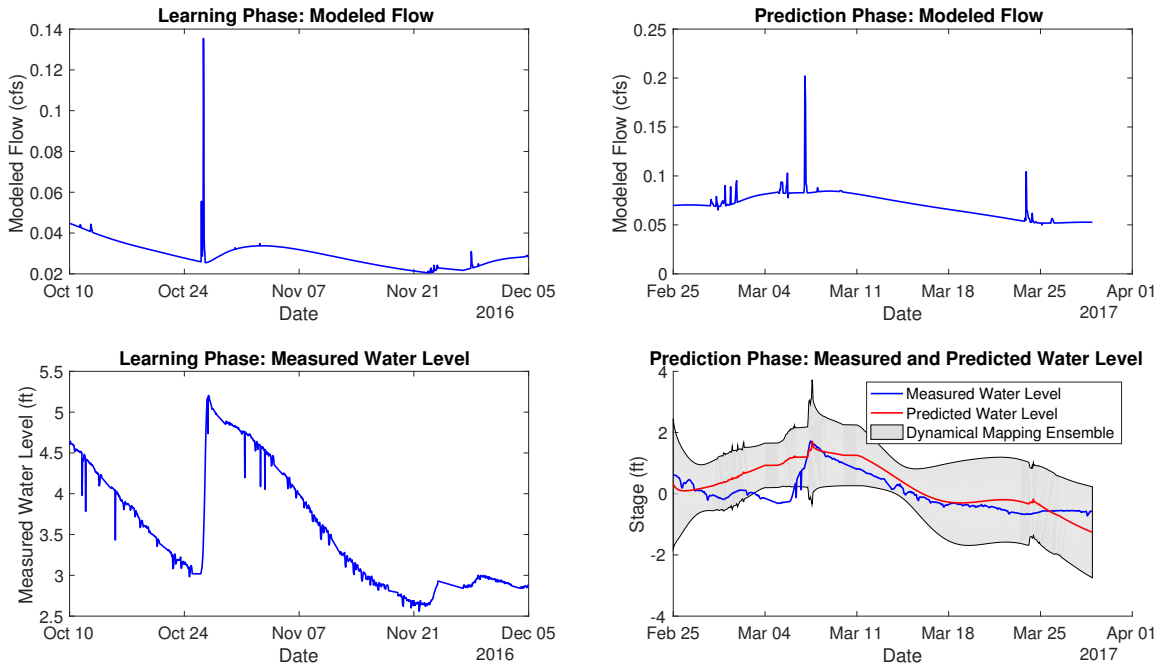


Figure B.165: Site WNBGORV03

WNBGORV04, Mean nRMSE = 22.78%

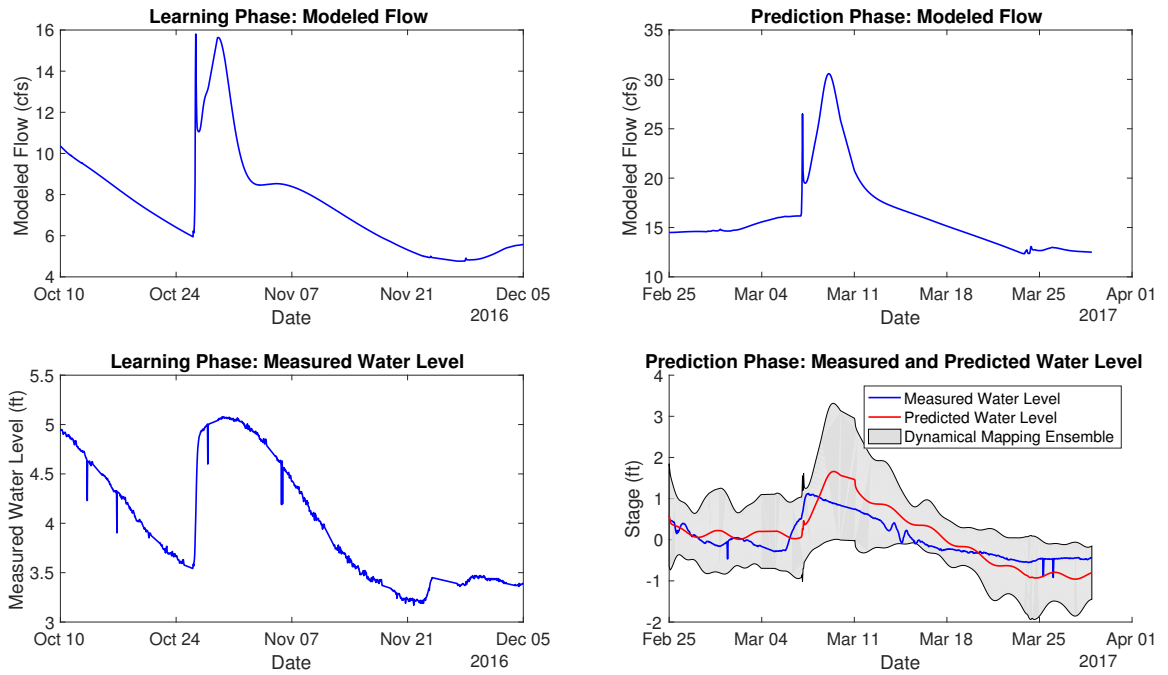


Figure B.166: Site WNBGORV04

WNBGORV06, Mean nRMSE = -271.19%

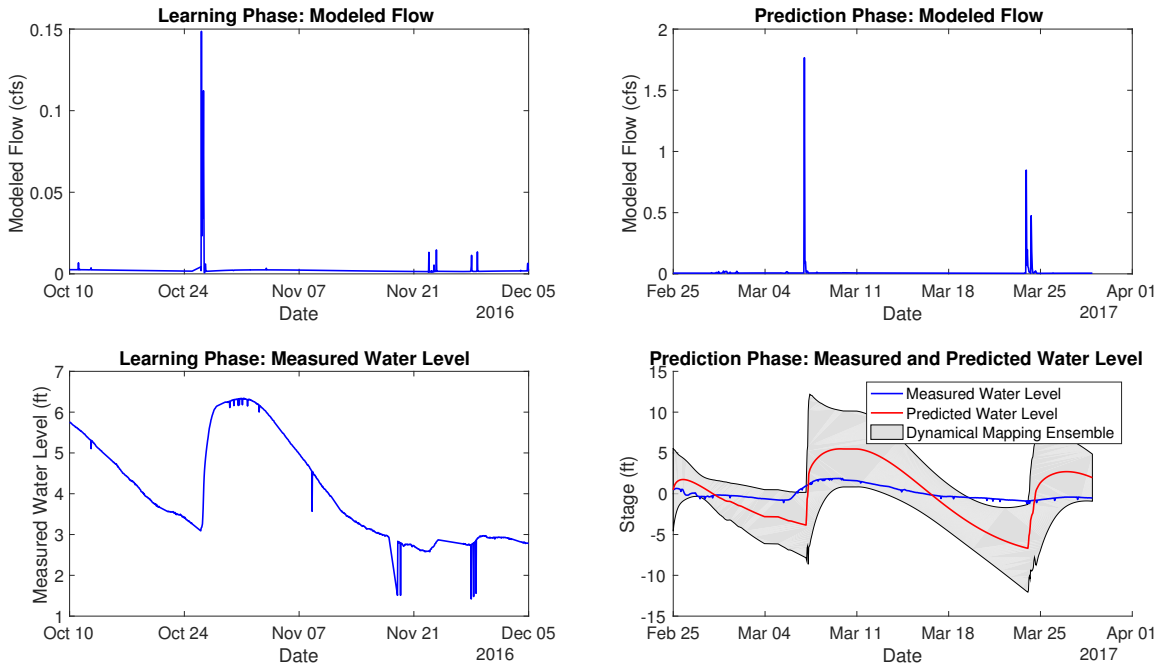


Figure B.167: Site WNBGORV06

WNISH01, Mean nRMSE = 6.07%

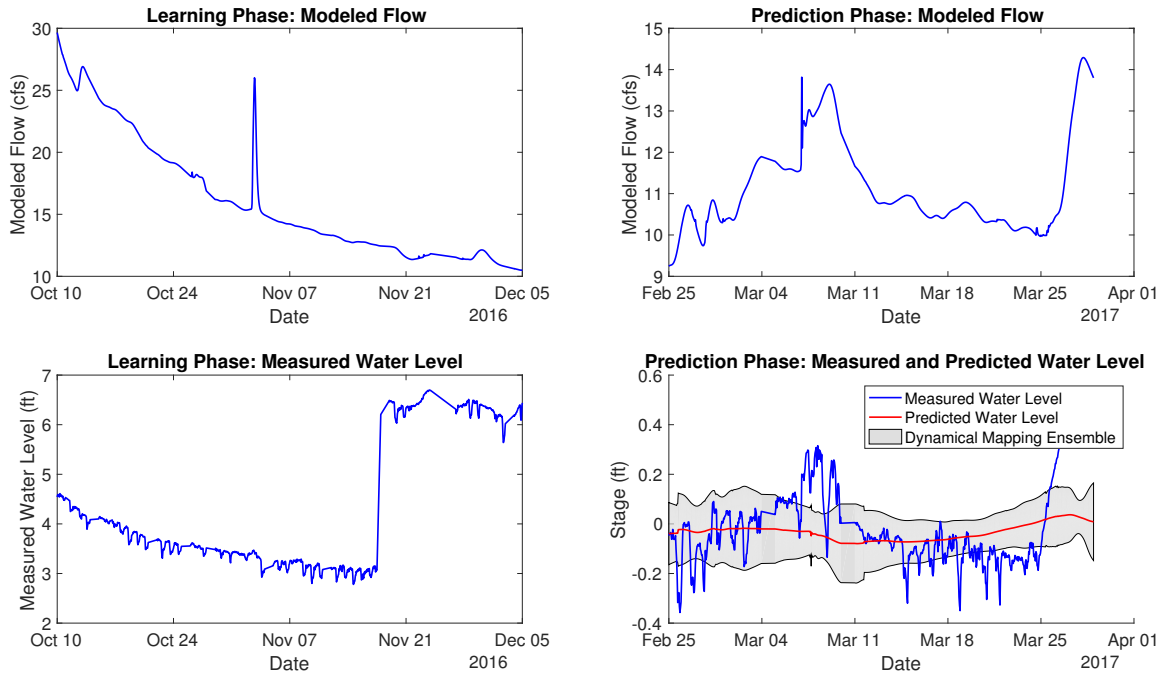


Figure B.168: Site WNISH01

WNISH02, Mean nRMSE = -17.75%

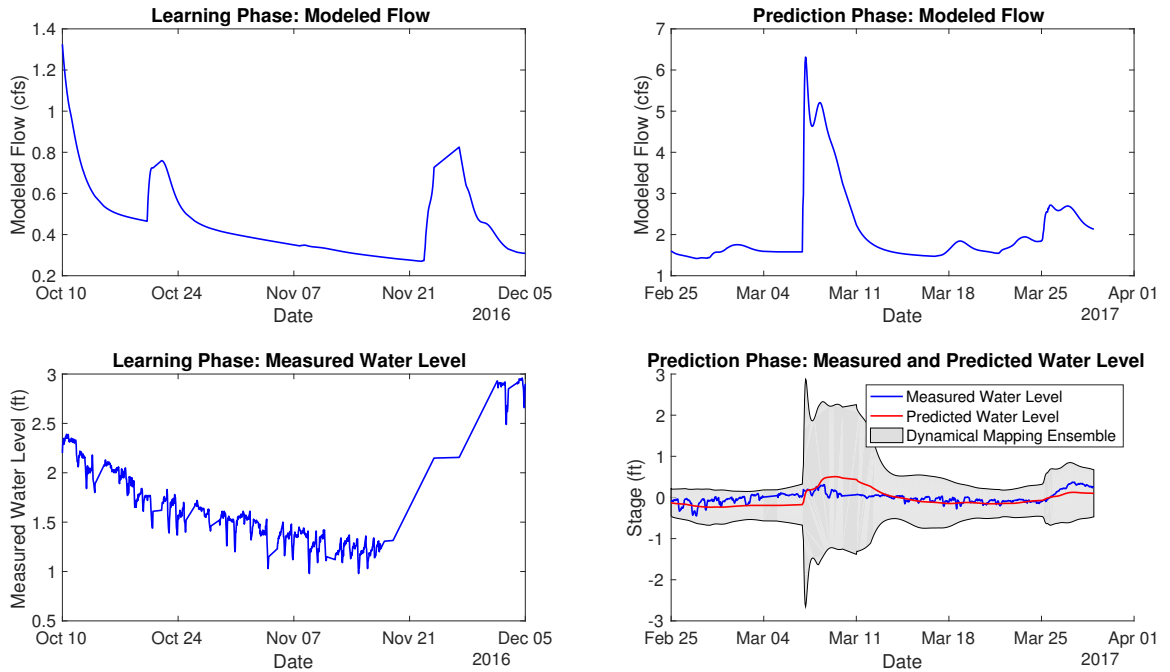


Figure B.169: Site WNISH02

WNISH03, Mean nRMSE = -26.71%

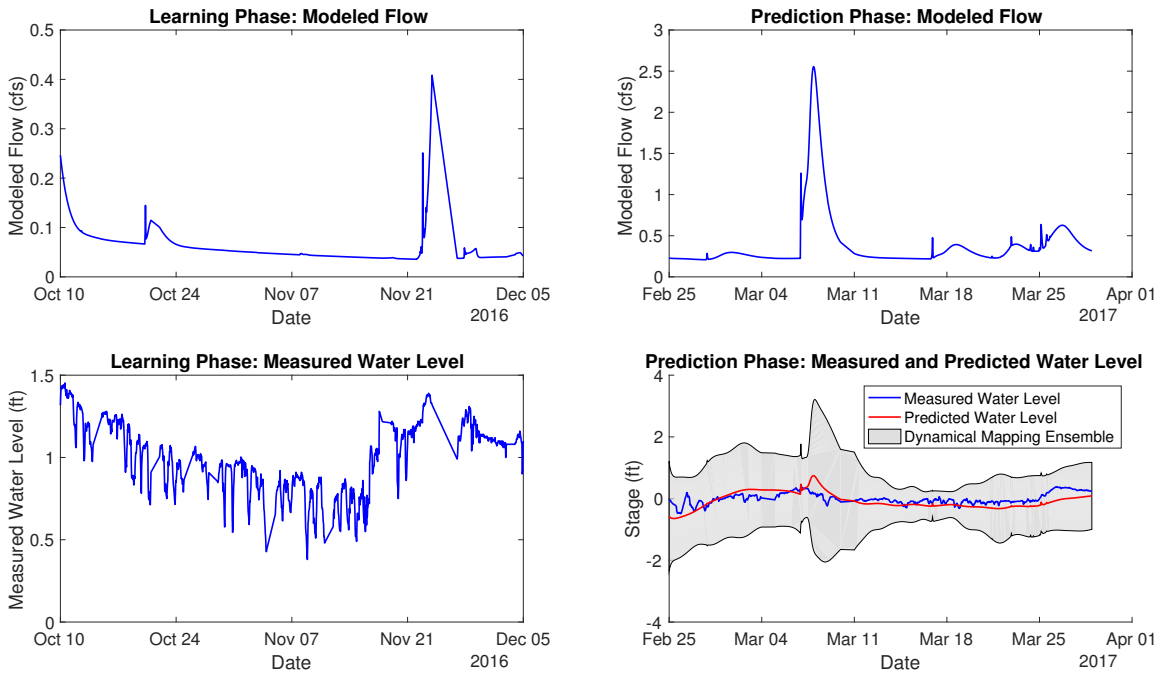


Figure B.170: Site WNISH03

YELLOWRV01, Mean nRMSE = 36.01%

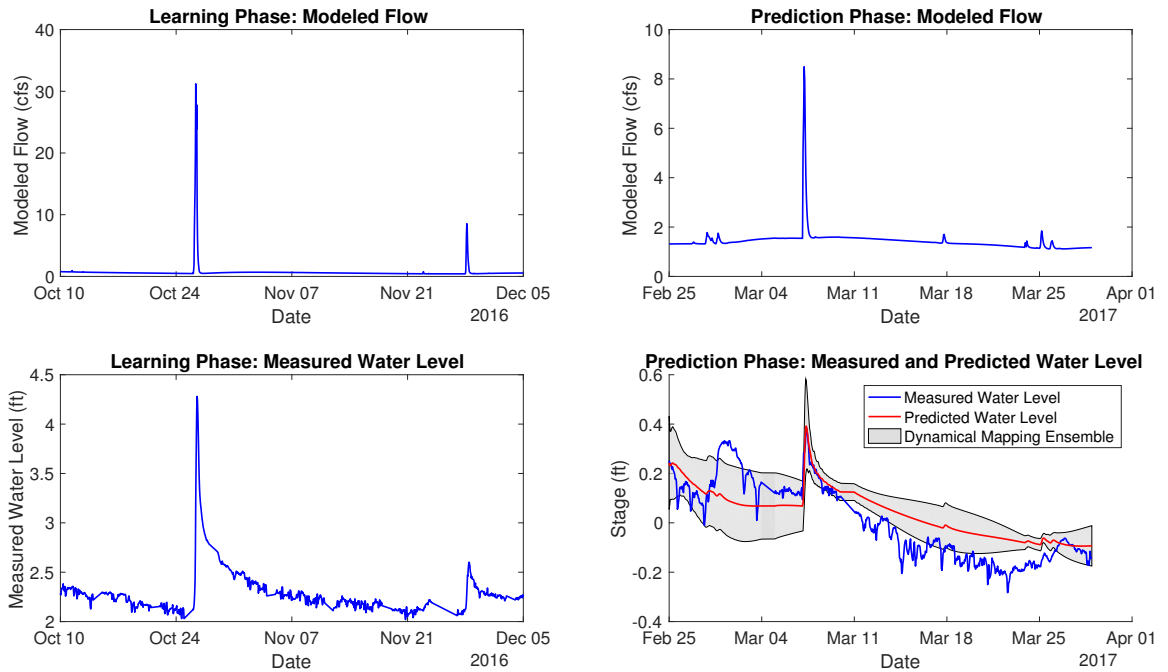


Figure B.171: Site YELLOWRV01