PREDICTING GROUNDWATER LEVEL USING DATA MINING TECHNIQUES: A CASE STUDY OF THE EDWARDS AQUIFER

by

Lenée A. Dedeaux, B.S.

A thesis submitted to the Graduate Council of
Texas State University in partial fulfillment
of the requirements for the degree of
Master of Science
with a Major in Aquatic Resources
May 2017

Committee Members:

Benjamin F. Schwartz, Co-Chair

Yihong Yuan, Co-Chair

Ronald T. Green

FAIR USE AND AUTHOR'S PERMISSION STATEMENT

Fair Use

Duplication Permission

ACKNOWLEDGEMENTS

This work would not be possible without the support and guidance of my committee members, Drs. Schwartz, Yuan, and Green. In particular, I would like to thank Dr. Benjamin F. Schwartz who worked tirelessly to help me reach my goal of obtaining my master's degree. Thank you so much for everything you have taught me and everything you have done to make my dreams come true. And, I would also like to thank the Texas State Biology Department and the Texas State University Graduate College for generously supporting my research through funding. Lastly, I want to thank my wonderful, supportive family who traveled this challenging journey with me, although not always happily, but always selflessly. *Thank you*, Alvin, Mila, Sadie, and Ruby.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

I. INTRODUCTION

The hydro-geologically complex Edwards Aquifer supports multiple ecosystems and provides high quality groundwater to humans for recreation, household use, and industry. Humans rely on the spring-flows from the aquifer for recreational activities such as swimming, tubing, kayaking, and fishing. Groundwater from the Edwards Aquifer provides a reliable source of municipal water supply. Industries utilizing groundwater from the Edwards Aquifer include farming and ranching. The aquifer also supplies flows to important local springs and provides fresh water to fragile habitats supporting endangered species. Because human and natural ecosystem depend on this limited groundwater resources, it is important to predict how much groundwater is available. And, with so many users dependent upon the Edwards Aquifer groundwater, accurately predicting groundwater levels can improve planning and management of an increasingly overextended resource.

Predicting groundwater level in highly heterogeneous karst aquifers has proven difficult using traditional statistics-based models with limited datasets (Lindgren et al., 2004; Coppola et al., 2005; Szidarovszky et al., 2007; Chen et al., 2013; Wu & Zeng, 2013). And, the Edwards Aquifer's highly variable response to changes to the amount of water moving into or out of the system (i.e., recharge and discharge) require robust datasets representing a wide range of hydrologic responses to make consistent and accurate predictions of groundwater levels (Hensel and Hirsch, 2002; Ford & Williams, 2007; Drew & Goldscheider, 2007; Bear & Cheng, 2010; Krešic & Stevanovic, 2010). Recent technological advances have allowed for the aggregation of computer-based data collections of time-series datasets of a wide range of hydrologic parameters and

responses that are especially useful for computational modeling of complex systems, such as the Edwards Aquifer groundwater levels (Fayyad, Piatetsky-Shapiro & Smyth, 1996; Han & Gao, 2009; Hoffman at al., 2011; Esling &Agon, 2012; Paasche et al., 2014). Time-series datasets are embedded with patterns of information about the inherent relationships within a complex system and require specialized techniques, such as data mining. Due of the characteristics of hydrologic data, such as autocorrelation and high degrees of heterogeneity in both the spatial and temporal dimension, data mining techniques are particularly useful when constructing models to predict complex hydrologic systems (Hensel & Hirsch, 2002; Zhou & Li, 2011; O'Reilly et al., 2012; Sahoo & Jha, 2013). Data mining consist of tasks to extract patterns from datasets capable of describing the system, and then using those patterns to derive a mathematical approximation of a complex systems response to changes using parametric and non-parametric techniques (Miller, 2009; Sunitha & Reddy, 2014). Parametric techniques are used when you want to determine how much influence a finite number of physical parameters each contribute a response, while non-parametric techniques are used when you are not concerned about determining the most influential parameters but with predicting or simulating a response using parameters capable of describing the response (Hastie, Tibshirani & Friedman, 2009; Taneja et al., 2011). Unlike parametric prediction techniques, which seek to explicitly characterize a complex system typically with limited data, non-parametric data mining techniques do not require explicit characterization of the physical parameters influencing groundwater levels. In an alternative approach, non-parametric data mining techniques use the patterns embedded in time-series datasets of highly similar groundwater wells, considered high quality training data, to estimate, or

predict, other wells groundwater levels. Artificial neural network (ANN) is a non-parametric data mining techniques that processes information in a manner similar to a biological neural network, such as the human brain (Craven & Shavlik, 1997). ANNs are capable of making non-linear estimations, or predictions, of a complex systems response, such as groundwater levels, when trained with high quality training data (Han & Gao, 2009; Trichakis, Nikolos & Karatzas, 2011; Kalina, 2013). Because of this, ANNs have been found to work particularly well when predicting responses in a complex hydrologic system. And, for homogenous aquifer systems with similar hydrogeologic properties throughout, highly accurate predictions of groundwater levels can be made using ANNs (Daliakopoulos, Coulibaly, & Tsanis, 2005; Sirhan & Koch, 2012; Rakhshandehroo, Vaghefi & Aghbolaghi, 2012; Mohanty et al., 2013; Karthikeyan et al., 2013; Chitsazan, Rahmani & Neyamadpour, 2013).

Although, when ANNs were applied to highly heterogeneous hydrologic systems, prediction accuracies tend to suffer due to a lack of high quality training data representing a robust range of conditions and responses (Lallahem et al., 2005; Trichakis, Nikolos & Karatzas, 2009; Wu, 2010; O'Reilly et al., 2014). And, in complex systems such as the Edwards Aquifer, a spatially dependent and spatially heterogeneous karst aquifer, groundwater levels response to system changes vary greatly, creating additional challenges when identifying high quality training data (Trichakis, Nikolos & Karatzas, 2011). The following section provides a comparative overview of traditional numeric modeling and select data mining techniques, including ANNs, in the hydrologic sciences and discussion on current methods used to identify high quality training data.

Data mining techniques in hydrology

In recent literature, comparison of numerical and data mining techniques has been a common theme for researchers exploring the efficacy of applying data mining techniques to hydrologic systems (Szidarovszky et al., 2007; Mohanty et al., 2013) and several studies compared the efficacy of various data mining techniques (Taneja & Chauhan, 2011; Sahoo & Jha, 2013; Sunitha & Reddy, 2014). Terzi (2011 & 2012) compared the prediction capabilities of several data mining techniques to traditional prediction methods to estimate monthly river discharge, and then monthly rainfall, in Turkey and found that data mining techniques, including artificial neural network (ANN), produced comparable solutions much faster than traditional numerical hydrologic modeling methods. In 2013, Mohanty et al. conducted a comparative evaluation of MODFLOW, a numeric groundwater model, and ANNs to simulate weekly groundwater levels in India and found ANNs provided more accurate predictions than the MODFLOW for short term predictions. In Japan, Sahoo and Jha (2013) compared the data mining techniques of multi-linear regression (MLR) and ANN to predict monthly groundwater levels in 17 groundwater wells over the Konan groundwater basin of Kochi Prefecture. Using a correlation analysis to determine the physical parameters influencing groundwater levels, both MLR and ANN predicted with high accuracy, but the study found ANN predictions were in "better agreement" with observed groundwater levels because the ANNs ability to approximate the non-linearity of the system.

Because ANNs can be less time consuming and provide superior prediction accuracies over traditional modeling techniques when correctly applied, recent studies have focused on maximizing the robustness of ANN training data through parameter and algorithm optimization techniques to increase prediction accuracies, but with varying

results (Ratanamahatana & Keogh, 2004; Gaur et al., 2013; Gupta et al., 2014; Jha & Sahoo, 2015). Trichakis, Nikolos, and Karatzas (2009) used a differential equation for optimal selection of ANN parameters and found an improvement over empirical methods for selection of parameter inputs, which minimized training time and increased prediction accuracy. Then, Tapoglou et al. (2012) compared three different variations of the particle swarm optimization algorithm (PSO), an algorithm that moves 'particles' representing a candidate solution through a search space to attract like-values, aiming to increase the robustness of the ANNs training data. However, results showed the PSO algorithms could not sufficiently identify outliers in the data, thus resulting in poor prediction accuracies.

In most of these studies, ANNs were shown to have high prediction accuracies for relatively homogenous hydrologic systems, but prediction accuracy suffered with increasing hydrogeologic heterogeneity. Coppola et al. (2003), citing the results of a sensitivity analysis, noted the importance of identifying relevant input parameters, which are highly specific to the system being modeled, to increase the robustness of the training data. Lallahem et al. (2005), found that an "increase in the number of neighboring piezometers with different (hydrogeologic) features" prevented their test model from accurately predicting observed groundwater levels. Trichakis, Nikolos, and Karatzas (2011) applied ANNs to predict groundwater levels in the Edwards Aquifer using a correlation coefficient analysis to determine relevant input parameters. But, the study found the heterogeneity of the karst hydrogeology and a lack of pumping data hindered the ability of the ANN to capture the inherent relationships in the datasets during training, resulting in low prediction accuracy.

Some studies found that grouping data increased the robustness of the ANN training data, but only when grouping of the data based on similar responses to changes. Sreekanth et al. (2009), when predicting groundwater levels in a groundwater basin in India, randomly subset the data to reduce model complexity, incidentally reducing the heterogeneity within the training data, and saw an increase the ANNs prediction capabilities during the training period. However, while the test model fit was good, the ANN was not able to predict outliers in the data. And, when Chitsazan, Rahmani, and Neyamadpour (2013) applied ANNs to predict groundwater levels in Aghili Plain in southwest Iran, a relatively homogenous groundwater basin, their research showed that grouping groundwater wells based on similar hydrologic properties (such as groundwater depth, hydraulic conductivity, and transmissivity) produced far more accurate groundwater level predictions than when not grouped.

The results of these previous works show the importance of selecting highly representative ANN training data specific to the system being modeled, and illustrates some of the many methods used to increase the robustness of input data, from optimization techniques to manually grouping data. However, in those studies, parameter selection was mainly conducted in a trial and error approach during the ANN training phase instead of an explicit method to partition input data with the goal of identifying high quality training data. The lack of consensus on a straightforward methodology to identify highly robust ANN training data highlights the overall limitations of the current techniques used for data mining in the field of hydrology.

Research questions and objectives

The foundation of this study was guided by the success and failures of previous studies and aimed to improve ANN groundwater level prediction accuracies in the highly heterogeneous Edwards Aquifer. Results from past studies show that grouping data based on response similarity can improve prediction accuracy by reducing the heterogeneity within the training data. Therefore, this study first used the data mining technique of hierarchical clustering with the Dynamic Time Warping algorithm as a measure of similarity to facilitate the identification of highly representative ANN training data.

Dynamic time warping (DTW), a time-invariant distance algorithm, useful when modeling spatially and temporally complex systems, measures similarity, or distance, independent of time thereby identifying similar responses although those responses may not occur at the same time (Giorgino, 2009; Mori, Mendiburu & Lozno; 2015). DTW as a similarity measure to identify highly representative groundwater well time-series, used as ANN training data to model other wells groundwater levels, can increase groundwater level prediction accuracy by identifying outlier's groundwater levels present in the data. However, while hierarchical clustering using DTW is a robust method of identifying similarity between time-series that may not align, common in highly heterogeneous systems, no studies thus far have applied the data mining technique of hierarchical clustering to group highly heterogeneous groundwater level data using the DTW as a similarity measure.

Next, an ANN was constructed for each cluster of groundwater wells identified during the hierarchical clustering analysis. Using the highly representative clusters of groundwater wells as ANN training data, along with spring-flow and precipitation data, groundwater level predictions were made for each of the groundwater wells. These data

mining techniques, applied in a novel two-step approach, were selected with the specific goal of identifying a straight forward methodology for accurately predicting groundwater levels in highly heterogeneous systems, using the Edwards Aquifer as a case study.

The purpose of this research was to determine if the data mining techniques of hierarchical clustering using DTW to measure similarity and ANNs, applied in a two-step methodology, can predict with a high degree of accuracy groundwater levels of a highly dynamic karst groundwater system. This research has been designed to answers the following questions: 1.) To what degree can these data mining techniques accurately predict groundwater levels within the highly dynamic Edwards Aquifer using historical time-series datasets, and 2.) to what degree can these data mining techniques accurately train data to identify outliers within historical time-series datasets.

To answer question one, the accuracy of the two-step methodology was measured by calculating the difference between the ANN predictions and the observed groundwater levels, reported in terms of the coefficient of determination (R^2), a measure of overall model fit. To answer question 2, the average prediction error, or root mean square error (RMSE), was used to measure the degree to which the trained model could identify outliers in the data.

II. MATERIALS AND METHODS

A two-part methodology was explicitly designed with the goal of increasing groundwater level prediction accuracy in the Edwards Aquifer. Because ANNs can predict with a high degree of accuracy when trained with highly representative training data, identification of groups of wells that respond similarly across a range of hydrological conditions was considered integral to increasing ANN prediction accuracy. In step-one, hierarchical clustering, coupled with DTW as a measure of groundwater well similarity, was used to explore patterns in groundwater level time-series datasets from fourteen wells across a common interval of more than 7 years, and a range of hydrogeologic conditions including very wet and very dry periods. This technique was used to identify wells that respond similarly to a target well across all hydrologic conditions. Data from those wells were then used as ANN training data to predict groundwater levels of a target well. Wells that respond less similarly to the target well, or were similar under only certain hydrologic conditions, were rejected as ANN training data for a target well. In step-two, ANNs were constructed to predict daily observed groundwater levels for each of the fourteen wells using time-series data from wells that clustered with each target well, spring-flow, and precipitation as training data.

It is important to note that because this two-step methodology used non-parametric data mining techniques with the specific goal of increasing prediction accuracy, a comprehensive analysis of all possible physical parameters influencing groundwater levels was not necessary. Instead, careful consideration of select physical parameters thought to be capable of describing groundwater levels in the Edwards Aquifer guided the choice of parameters used as ANN training data.

Study Area

The Edwards Aquifer is a heterogeneous and highly productive freshwater karst aquifer in Central Texas (**Figure 1**) that has formed in Cretaceous carbonates deposited in a shallow sea. Karst aquifers are formed in highly soluble bedrock, such as limestones or dolomite, and dissolution enhances and enlarges an integrated and connected system of pores, fractures, and conduits that can quickly and efficiently transport groundwater throughout the system. The Edwards Aquifer, is considered a mature karst aquifer and contains multiple types of porosity ranging from primary granular porosity to large conduit porosity (including caves). Varying degrees of conduit connectivity result in interrelated changes in groundwater levels as a response to changes in groundwater storage (recharge, natural discharge, or pumping). In addition, as long-term hydrologic conditions transition between periods of extreme wet and dry, the degree to which conduits are interconnected can change dramatically as groundwater levels and hydraulic heads rise or fall, leaving some conduits dry and inactive, or even changing flow directions in others.

The Edwards Aquifer is 180 miles long and up to 40 miles wide and is generally oriented in a northeast-southwest direction (EAA). The aquifer is composed of three distinct hydrogeologic zones: the contributing, the recharge, and the artesian zones. The contributing zone drains 5,400 mi^2 of the central TX Hill Country to the northwest of the aquifer and serves as a catchment area for overland flows, feeding streams that eventually contribute to recharge as they flow across the Balcones Fault Zone. Most recharge occurs along the 1,250 mi^2 exposed, unconfined recharge zone where extensive faulting forms the Balcones Fault Zone (BFZ), which is lies between the contributing zone and the deep artesian zone. The BFZ also hosts the BFZ Aquifer, a portion of the Edwards Aquifer

which is under unconfined water-table conditions. In the confined artesian zone, several large springs discharge from the aquifer, including the largest and second largest springs in TX: Comal Springs and San Marcos Springs. The artesian zone is confined by the impermeable Del Rio Clay formation above and the Upper Glen Rose Limestone below. Each zone responds to system changes differently due to varying degrees of spatial dependency and spatial heterogeneity in aquifer properties. In the recharge zone, because it is unconfined, the water table rises and falls with changes in storage. But groundwater levels in the artesian zone, because it is confined and under atmospheric pressure, respond to changes in pressure. Further complexity arises from groundwater pumping from thousands of wells located throughout the aquifer. Pumping creates an inverted cone of drawdown concentric to the pumping well, which can change groundwater gradients and flows both around the pumping well and at surrounding wells.

Hydrologic datasets

Daily groundwater level data for fourteen wells were obtained from the Edwards Aquifer Authority (EAA), daily discharges from two springs from the United States Geological Survey (USGS), and daily precipitation values for nine climate stations from the National Oceanic and Atmospheric Administration (NOAA), were considered as training data for the ANNs. The 7-year study period was from April 1, 2004 to July 14, 2011. The range of precipitations and spring flow values for the 7-year study period were not as robust of a range as the entire period of record available. However, it was decided that the 7-year study period provided a robust range of well responses across hydrologic conditions and that variable conditions were more important to groundwater level prediction accuracy than precipitation and spring-flow. Other studies have used daily,

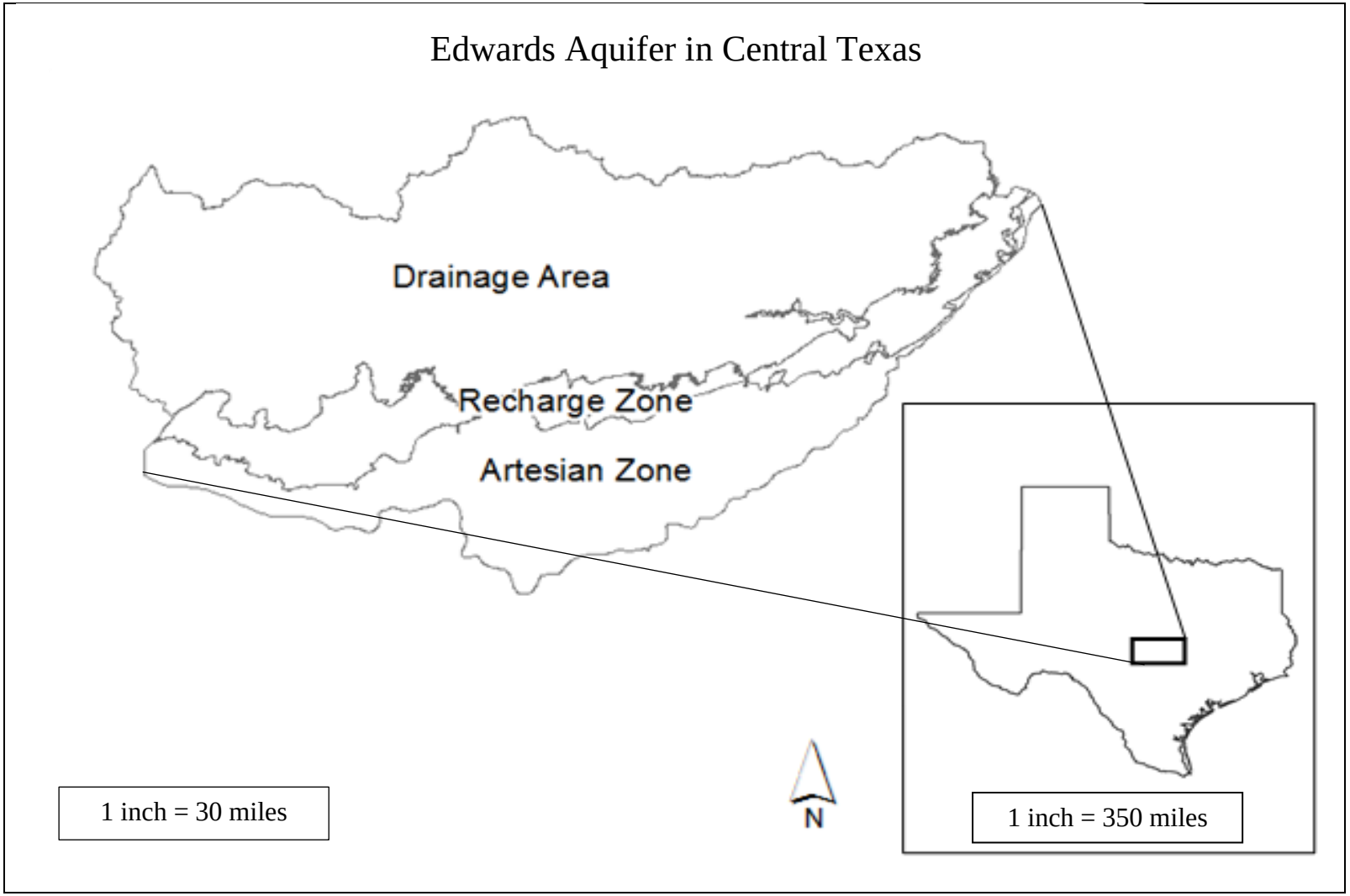

Figure 1. Edwards Aquifer in Central Texas.

weekly, or monthly resolutions when applying predictive ANNs with varying ranges of time-periods. Because daily datasets for the each of the input parameters chosen were available, a daily time resolution was used.

Groundwater level data

The data used in this study were selected from a larger set of data for 2683 wells for the period of 1918-2015, collected by the EAA, the Texas Water Development Board (TWDB), the USGS, and the San Antonio Water System (SAWS), that have been compiled and maintained in a database by the EAA. Groundwater level measurements for individual wells range from every fifteen minutes to daily, weekly, or monthly measurements. The data were filtered to select wells with the longest continuous concurrent time-periods with minimal missing data. For each well, the criterion for inclusion in the study was daily data with no more than 15% of the groundwater well dataset missing (training and testing datasets combined) for the longest time-period possible. Based on this criterion, fourteen wells with 7 years of continuous, concurrent daily groundwater levels were selected. The wells were placed in alphabetical order starting with the least amount of missing data, well A, to the most missing data, well N. For the records with missing data, linear interpolation was used to fill data gaps, some of which were several months long. A plot of the fourteen wells time-series serves to illustrate the heterogeneity present in the groundwater level datasets (**Figure 2)**. The wells chosen for this research were located both in the recharge and artesian zones, and across much of the length of the aquifer. Five of the fourteen wells, circled in **Figure 2**, were in the unconfined recharge zone or along the transition zone between the recharge

and artesian zones (wells C, K, M, N, and F), while the rest of the wells (wells, A, B, D, E, G, H, I, J, and L) were located in the confined artesian zone.

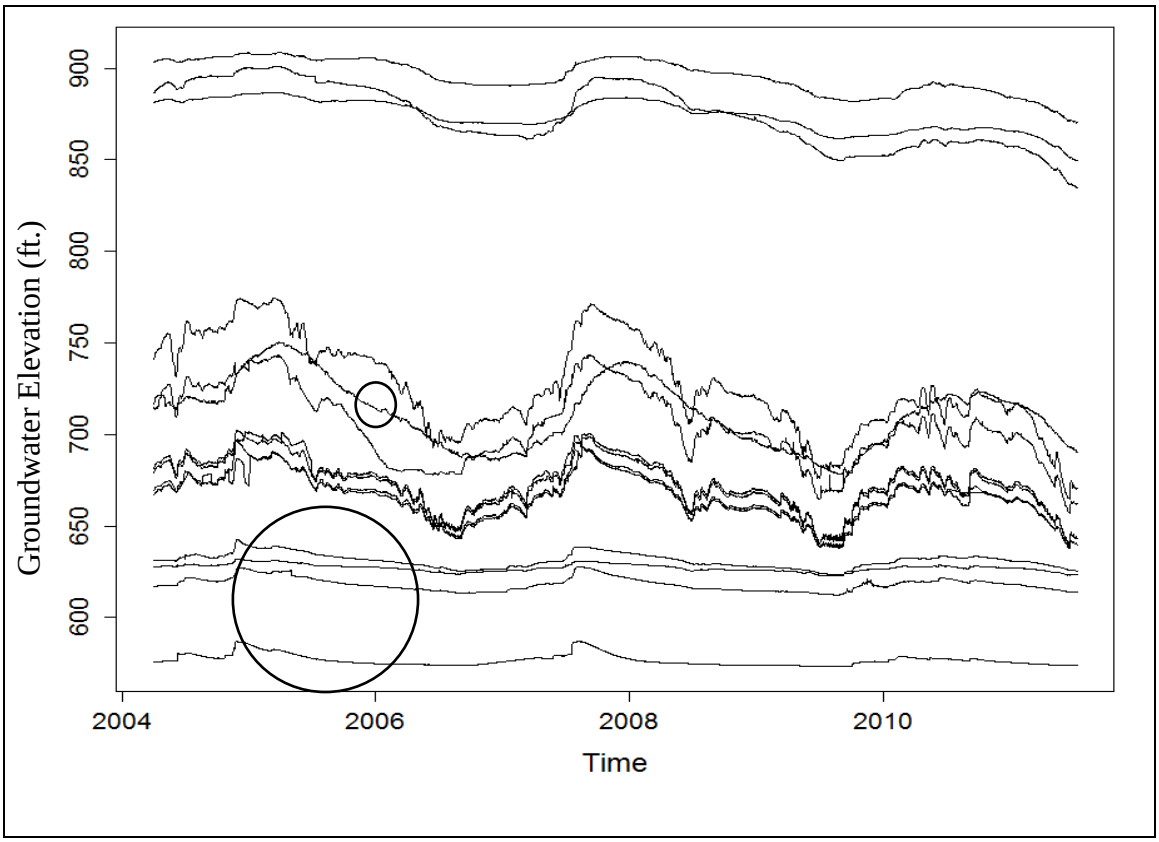

Figure 2. Raw time-series datasets of the fourteen wells.

Spring-flow

Daily spring-flow data is available for five springs located in the study area for the period of 1947-2012 (Comal, San Marcos, Leona, San Pedro, and San Antonio Springs). As hydrologic conditions change, the degree of hydrologic connectivity of each spring to the rest of the aquifer is highly variable. During times of severe drought, several of the springs have ceased flowing for extended periods (in some cases close to a decade), while at least one, San Marcos Spring, has flowed for the entire period of record.

Because most of the water naturally discharged from the aquifer flows from Comal Springs and San Marcos Springs (the first and second largest springs in TX, respectively), and both springs continuously flowed throughout the 7-year study period, only these two springs were selected as input training data for the ANNs. Discharge at the two springs is highly variable both within and between the springs, which further illustrates the spatial-temporal complexity of the Edwards Aquifer. A one-week example of raw data for San Marcos and Comal Springs illustrates that spring discharge can routinely vary by 1 to 2 cfs. per day; **Table 1**.

Precipitation

Precipitation from nine NOAA climate stations located across of the Edwards Aquifer region were used for ANN training data (figure 3). A 1-week sample of the raw data for each of the following nine stations is provided in **Table 2**: Brackettville (Brack), San Antonio International Airport (SAIA), New Braunfels (NB), Parade Ranch (PR), Kerrville (Kerr), Hondo, Canyon Dam (CD), Boerne, and Hunt.

Table 1. Example of 1 week of the spring flow data.

Date	San Marcos Springs (cfs.)	Comal Springs (cfs.)
10/18/2010	200	339
10/17/2010	302	340
10/16/2010	302	340
10/15/2010	203	341
10/14/2010	204	339
10/13/2010	204	340
10/12/2010	203	342

Table 2. Precipitation raw dataset. One-week precipitation values (inches per day) from each of the nine NOAA climate stations.

Date	SAIA	NB	Hunt	PR	Kerr	Hondo	CD	Brack	Boerne
10/18/2010	0	0	0	0	0	0	0.08	0	0
10/17/2010	0	0	0	0	0	0	0.24	0.01	0
10/16/2010	0	0.03	0	0	0	0	0	0	0
10/15/2010	0	0	0	0	0	0	0	0	0
10/14/2010	0.02	0	0	0	0	0	0.62	0	0
10/13/2010	0.06	0	0.24	0	0	0	1.91	0	0

Groundwater levels in the unconfined recharge zone are mostly controlled by precipitation and related recharge and have a range of less than 20 feet for the study period. In the confined zone, artesian where groundwater levels are also controlled by precipitation and recharge, in addition to pumping, changes in hydraulic head were over 100 feet for the study period. Because aquifer levels are highly dependent on precipitation and related recharge, precipitation data were considered to be important as ANN training data for wells located in the recharge zone and, to a lesser extent, the wells in the confined zone. However, since precipitation across the aquifer is highly variable and hydrologic conditions across the aquifer can change rapidly, often in just a few months, 30, 60 and 90-days moving averages of spatially averaged precipitation data were used as ANN training data instead of daily precipitation values from each individual station. Moving averages were calculated using the average daily precipitation value across all nine stations. The three moving average datasets were thought to provide the ANNs with important information representing antecedent hydrologic conditions across the aquifer, which was considered more relevant to groundwater levels than precipitation

values from each individual station. Locations for the 14 groundwater wells, the 9 NOAA climate stations, and the 2 springs are illustrated in **Figure 3**.

Step One – Hierarchical clustering using dynamic time warping (DTW)

A hierarchical clustering analysis of the fourteen wells was conducted to identify wells that respond similarly across a wide range of hydrologic conditions, which were considered highly representative ANN training data necessary for accurate groundwater level predictions. A common exploratory analysis technique of organizing clusters of data by ranking wells based on similarity, hierarchical clustering has been found to be especially informative when the goal is to identify outliers in highly heterogeneous systems, such as the Edwards Aquifer (Han & Gao, 2009; Hastie, Tibshirani & Friedman, 2009; Gupta et al., 2014).

Dynamic time warping (DTW) algorithm

Hierarchical clustering requires a matrix of the measured similarity of each well to each other well. The similarity, or distance, between two time-series is calculated by measuring the distance between signals, or response values, along the two time-lines. The algorithm selected to measure similarity between two time-series is an important consideration because the appropriate algorithm can reveal relationships other algorithms or humans may not be able to easily identify. Therefore, because the time-series of wells

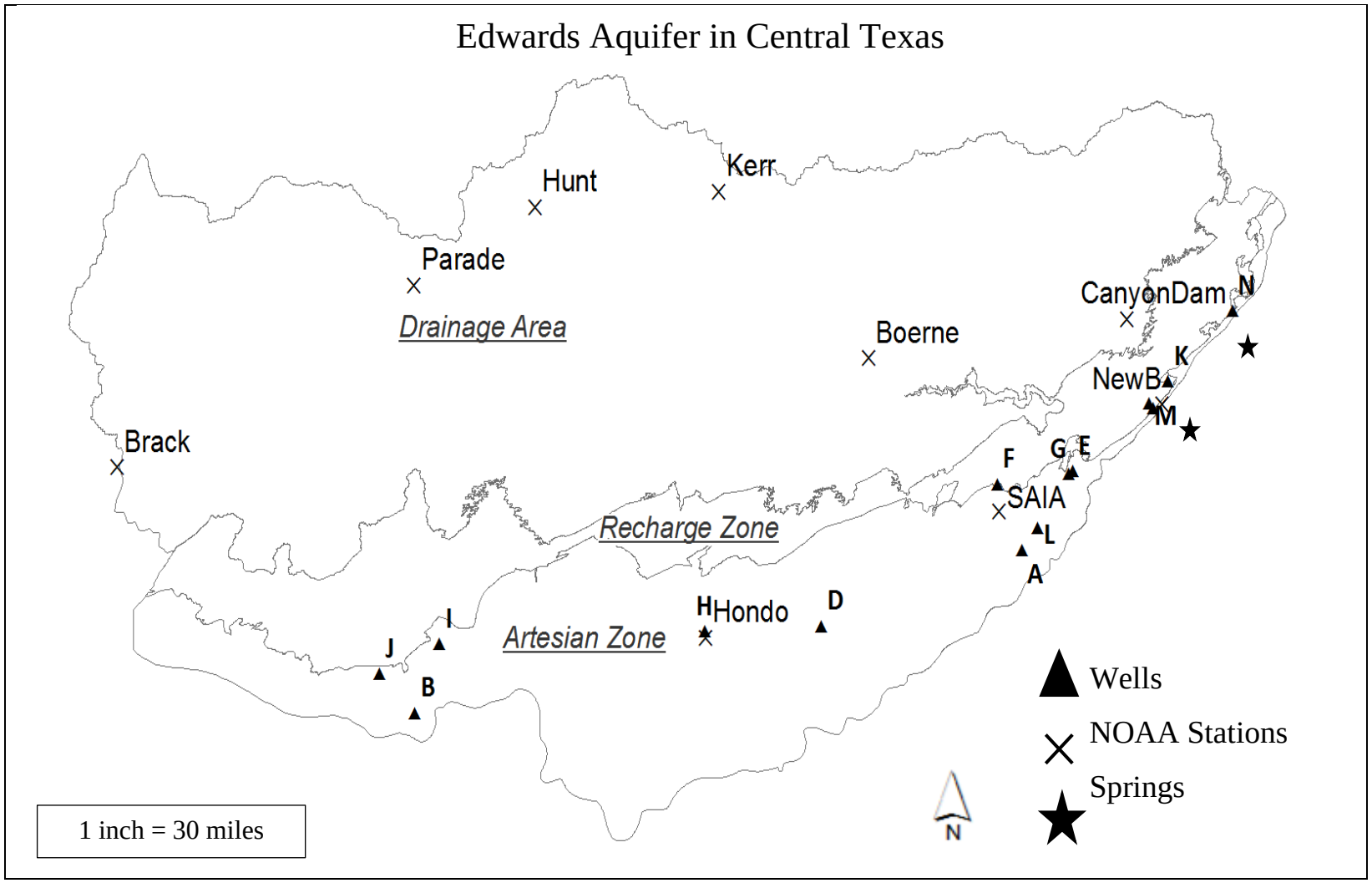

Figure 3. Location of hydrologic input parameters.

in the Edwards Aquifer frequently do not align because of time fluctuations, or lag times, between recharge or pumping and related groundwater level responses, the time-invariant

DTW, an algorithm that measures similarity independent of time, first originated in the field of automatic speech recognition when warping time-series was used to model time fluctuations of spoken words between two speech patterns (Sakoe & Chiba, 1978). Since then, research has repeatedly shown that measurements of similarity using DTW can identify the optimal alignment between two time-series more often than Euclidian distance, which simply measures vertical straight-line distance between two time-series (Ratanamahatana & Keogh, 2004; Han & Gao, 2009; Hastie, Tibshirani & Friedman, 2009).

For time-series that align (i.e., responses occur at the same time, but maybe not at the same magnitude), Euclidean, or straight-line distance, is sufficient. However, environmental parameter time-series datasets commonly do not align on a time line because of the spatial dependency and spatial heterogeneity of a system over time (i.e., response times and magnitudes vary). When comparing two time-series that may not align on a time line, DTW measures similarity of signals independent of time alignment, hence disregarding the constraints of time to find the optimal similarity between two time-series. **Figure 4** is an illustration of how DTW maps signals using a noisy sine and cosine wave as sample time-series. The independence from the constraints of time allows DTW to measure the response similarity of two wells that may geologically differ and/or have a lag time (the time it takes for groundwater to move from one well to another), thereby finding the optimal alignment between the two time-series by minimizing the distance between the two time-series. The lag time between two wells' responses to a

change in the system is represented by the difference of when two mapped signals occur in time (the x axis). **Figure 5** shows the optimal alignment of the noisy sine and cosine wave from **Figure 4** as calculated by DTW.

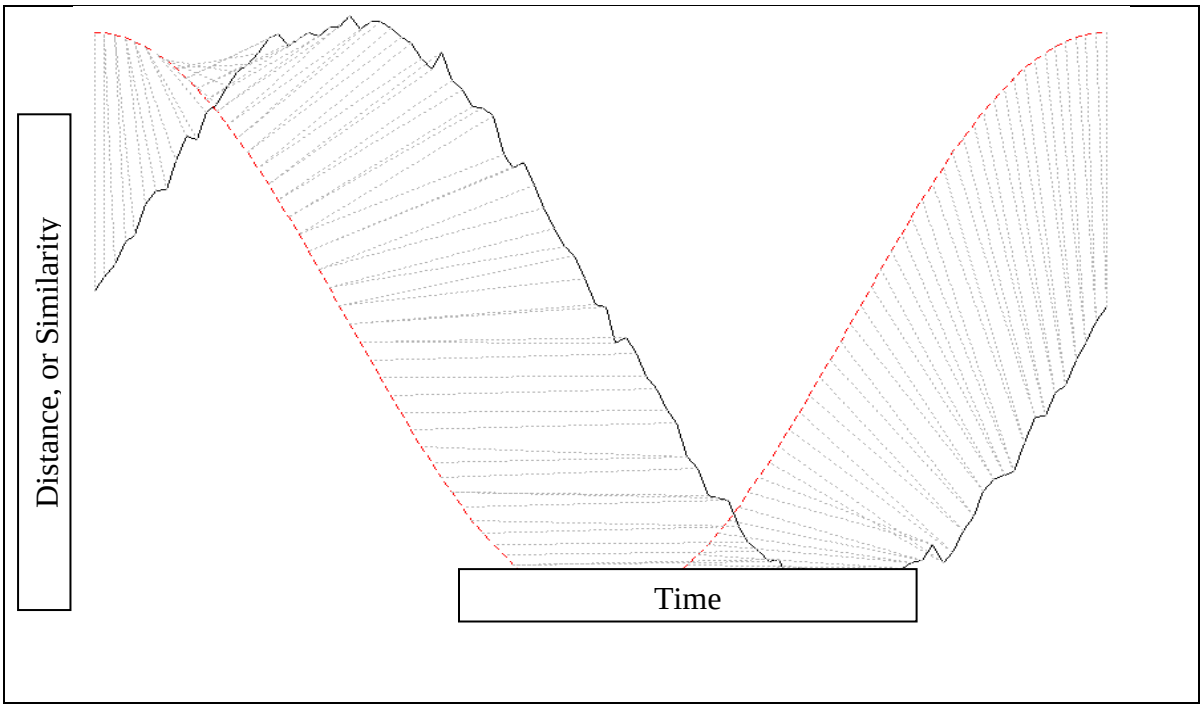

Figure 4. DTW mapped signals of a cosine and noisy sine wave.

Dividing the study period

With the goal of identifying which wells respond similarly as hydrologic conditions oscillate from extremely wet to extremely dry, the study period was divided into fifteen, 6-month time-periods to identify which wells responded similarly under each time-period's hydrologic conditions (**Figure 5**). Initially, the appropriate length of time, or resolution, for each time-period was unclear, so the hierarchical clustering analyses were conducted on time-periods of 3-month and 6-month, and on time-periods of varying lengths between 6-month to 2-years. To determine the best time resolution for the

clustering analysis, dendrograms of each time resolution were analyzed for patterns of wells that clustered and visually compared to the time-series of each corresponding time-period to determine if the clustering solutions correctly represented the patterns observed in the time-series (**Figure 6**).

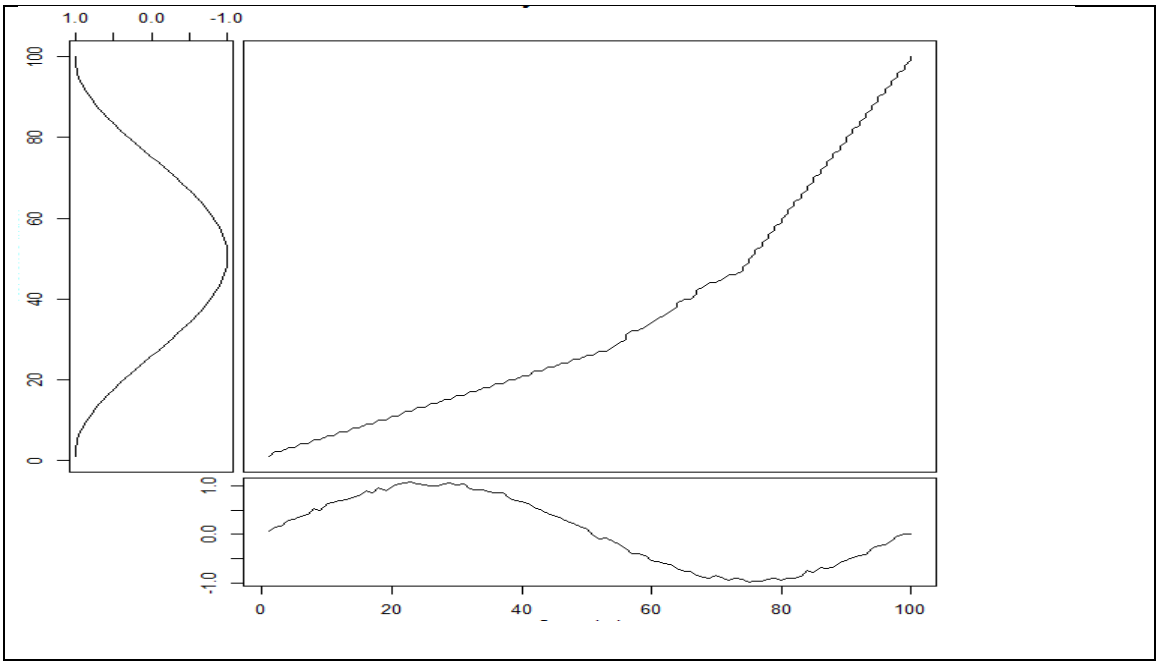

Figure 5. Optimal alignment between a noisy sine wave and a cosine wave.

The 3-month time-periods clustering solutions failed to follow the general patterns in the time-series. Because the DTW algorithm finds the optimal alignment independent of time, the 3-month time-periods were found to contain too little information for the hierarchical clustering analysis. The 6-month time-period clustering solutions reflected the patterns in the time-series for the time-periods the best. For the time-periods of varying lengths, the clustering solutions for the shorter time-periods (6-9 months) were approximately the same as the 6-month time-periods, but the clustering solutions for the longer time-periods (over a year) failed to cluster according to the

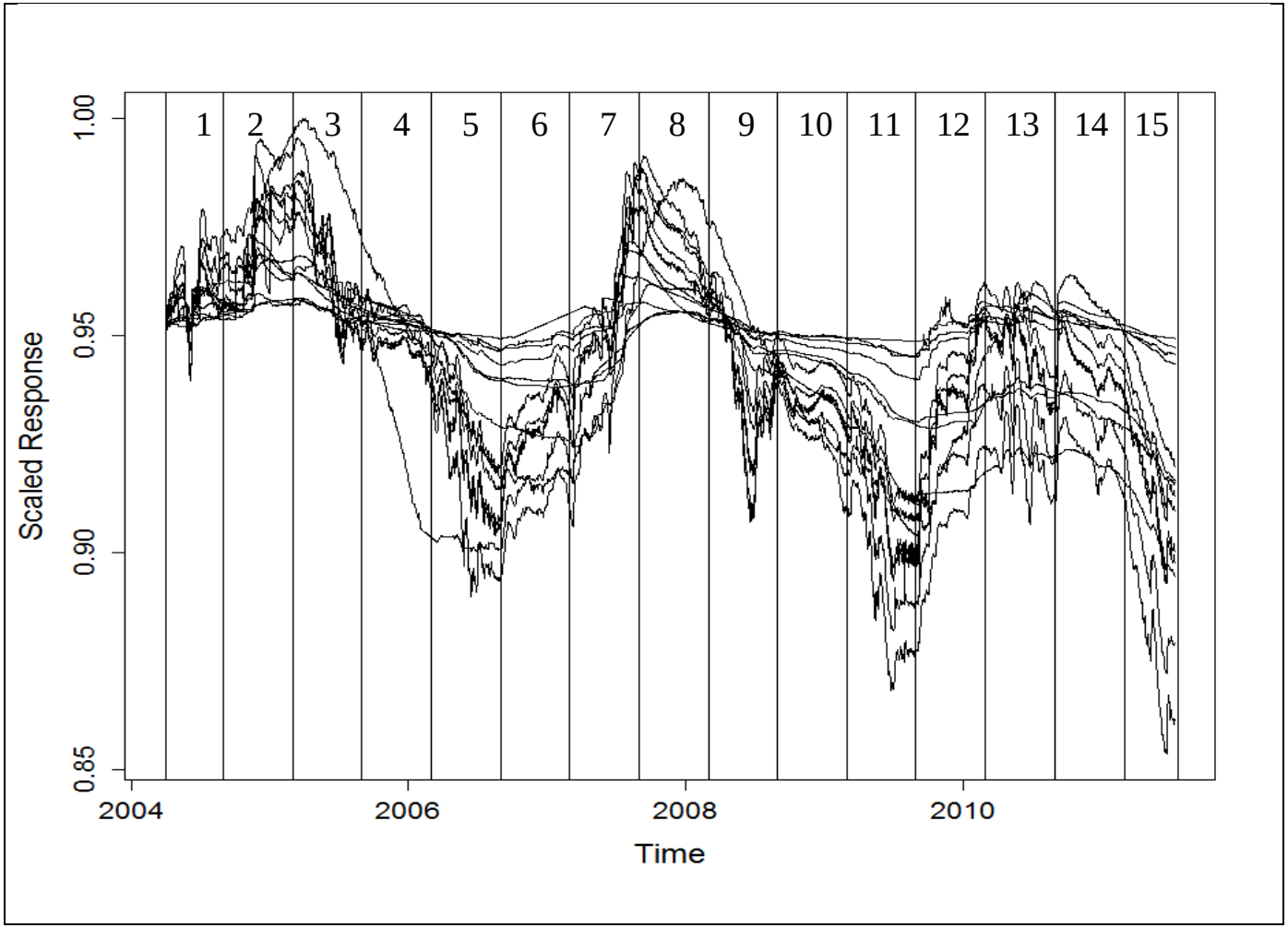

Figure 6. Study period divided into fifteen time-periods.

patterns present in the time-series for that time-period. Therefore, it was decided the 6-month time-periods were the appropriate time resolution for the hierarchical clustering analysis. The 6-month time-periods began March 1 and September 1 of each year in the study period, except for the first and last time-periods which were shorter periods. For each of the fifteen time-periods, the fourteen well time-series datasets were scaled [0,1] and standardized to begin at the same initial value. Then for each time-period, the well response similarity was measured using the DTW algorithm which provided fifteen similarity matrices for the hierarchical clustering analysis.

Identifying clusters of wells for each time-period

A hierarchical clustering analysis was conducted on each of the fifteen, 6-month time-periods. Results were inspected to determine the optimal number of clusters during each of the time-periods. In a hierarchical clustering analysis, the researcher can determine the optimal clusters based on their knowledge of the dataset using a dendrogram to visualize the results. A dendrogram is a tree-like graphic that is a useful tool to visually examine ranking patterns that are based on the similarity matrix as input (**Figure 6**). On the y-axis is the similarity at which new clusters were made, while the x-axis merely lists the clustered objects at equal distances and provides no information about a cluster's similarity.

The most telling feature of the dendrogram is the height of the vertical line merging two clusters or objects to form a single cluster. The robustness of the information present in a cluster is observed by examining the distance from where the cluster was formed to where the cluster merges with another cluster, i.e. the vertical distance between two adjacent horizontal lines. Longer vertical lines indicate greater

difference between clusters, while shorter vertical lines indicate less differences between clusters. In this study, a bottom-up approach, also referred to as an agglomerative approach, was used to join wells. Initially, each groundwater well is a single cluster at the bottom of the dendrogram. Then, in an iterative process, groundwater wells are clustered based on the quantitative DTW measure of similarity. When new clusters are formed, the two closest clusters are joined by a horizontal line, indicating the distance at which the cluster was formed along the y-axis, with individual wells initially having a distance of 0. When comparing, and interpreting multiple dendrogram, such as in this study, it was useful to set an arbitrary threshold on the y-axis to guide selecting the optimal number of clusters. For this study, the arbitrary threshold was set at 1.0 which can be seen in **Figure 7**. It is important to note that this threshold is subjective and must be determined by each researcher based on their own knowledge of the study area and datasets.

Identifying clusters of wells across hydrologic conditions

Selecting the optimal number of clusters for each time-period's set of solutions was a task essential to identifying which wells responded similarly across the fifteen time-periods (considered a cluster of wells) and assisted in determining whether outlier wells (wells that did not cluster with others) are present among the well datasets. The following criteria were used to determine which wells clustered across hydrologic conditions and which wells were outliers: 1) a well that failed to cluster with at least one other well across at least twelve of the fifteen time-periods was considered an outlier well and not considered for ANN training data, and 2) wells that repeatedly clustered together for at least twelve or more of the time-periods were determined to be a cluster of wells.

Using these criteria, each of the fourteen wells were either assigned to a cluster or determined to be an outlier.

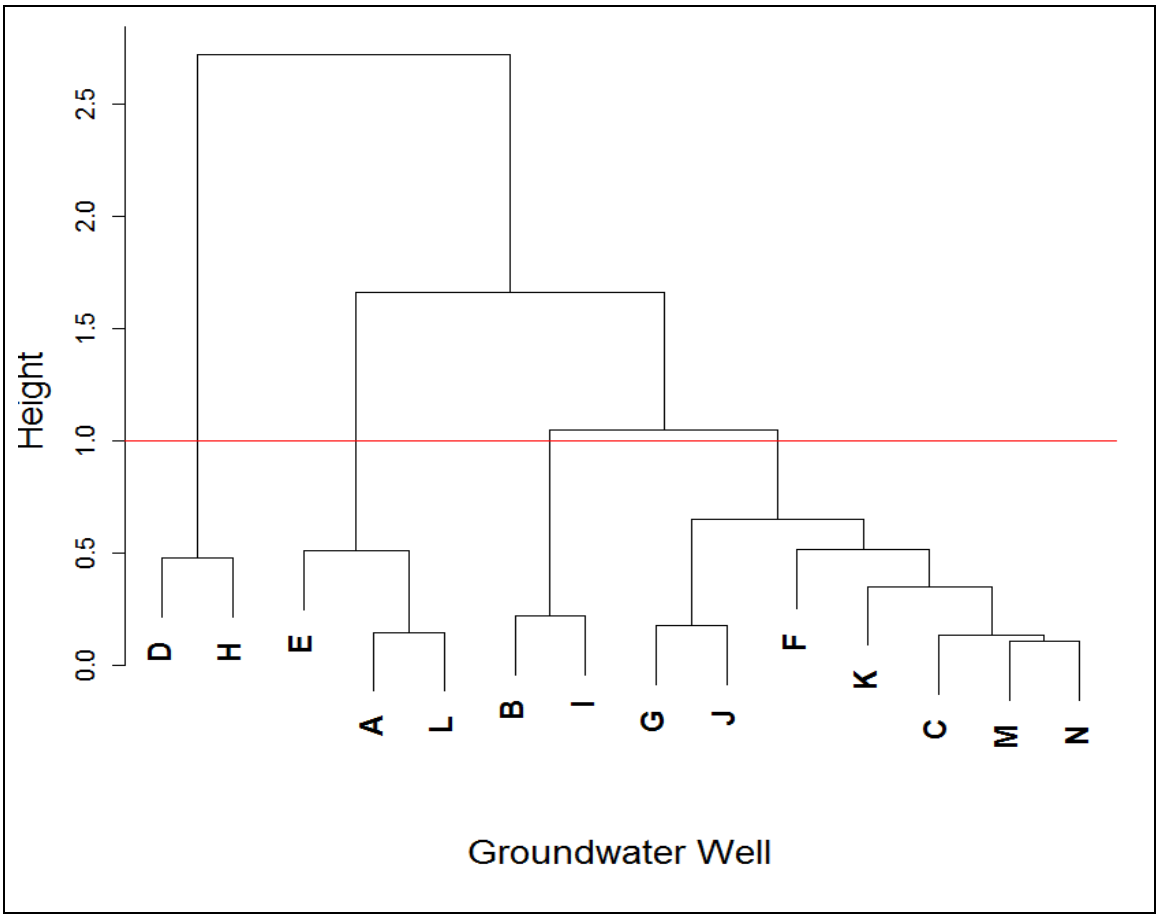

Figure 7. Dendrogram to visualize clustering solutions.

Step Two – Artificial neural networks

In step-two, artificial neural networks (ANNs) were constructed for each of the wells that clustered with another well, or wells, across hydrologic conditions using the criteria provided in section 2.3.3.

ANN architecture

ANNs architecture consists of an input layer(s), a hidden layer(s), and an output layer(s). The ANNs in this study were each constructed with a single input layer, two hidden layers, and single output layer. ANN architecture also refers to the direction information flows during training. Because the emphasis of this study was on discovering high quality ANN training data to predict groundwater levels, the successes of previous studies guided the selection of the feed-forward, back-propagation configuration, a universal approximator commonly used to simulate and predict within complex systems, particularly useful for predictive ANNs (Hastie, Tibshirani, and Friedman 2009). In a feed-forward/back-propagation architecture, as input training data is fed into the hidden layer, the error of the fit is iteratively subtracted from each prediction attempt and the resulting output is reintroduced to the ANN as new input.

Predicted groundwater levels are iteratively calculated by an activation function located in neurons connecting each layer that calculates the weight of each input to obtain the best model fit. In an ANN, neurons are basically a series of interconnected activation functions aggregated to create a neural network. As each neuron recursively recalculates new weights, the predictions are further constrained to an increasingly smaller range of values, effectively reducing the error from the previous attempt to fit the data. ANN training is improved by iteratively constraining and adjusting the weights. Because ANNs are commonly used to model non-linear systems, Sigmoid activation function, valuable for representing highly heterogeneous, non-linear relationships, are often used with ANNs to present non-linearity to the predicted output and was therefore used in this study (Taneja and Chauhan 2011). **Figure 8** is an illustration of the structure of an ANN which includes the input, hidden, and output layers, and final calculated weights of each input. It

should be noted that, because ANNs are non-parametric, input weights cannot be meaningfully interpreted.

ANN training data

ANN construction included selecting the optimal physical parameters for each well. All possible ANN training data included the physical parameter datasets (30, 60, and 90-day moving averages of precipitation and spring flow from San Marcos and Comal Springs), and the groundwater wells that clustered together. All the training datasets were scaled to [0,1] to prepare the data for ANN input. It was not immediately apparent whether all precipitation and spring-flow datasets selected for this study were useful ANN training data. Therefore, all physical parameters for precipitation and spring flow were initially included as training data and in a trial and error process during the training phase the decision of whether to include specific spring-flow and/or precipitation datasets as ANN training data was guided by whether exclusion of the respective physical parameter dataset resulted in an increase in groundwater level prediction accuracy. However, inclusion of a specific well dataset as ANN training data was pre-determined during the hierarchical clustering analysis where the wells used for training data were identified.

ANN training and testing

The 7-year study period was separated into two datasets for the purposes of training and testing: an approximately 6-year test dataset, April 1, 2004 to September 1, 2010, and a 1-year testing dataset, September 2, 2010 to July 14, 2011. Approximately 80% of the data was used for ANN training, while 20% was reserved for ANN testing; a common data ratio used with predictive ANNs. During the testing phase of the ANN, the

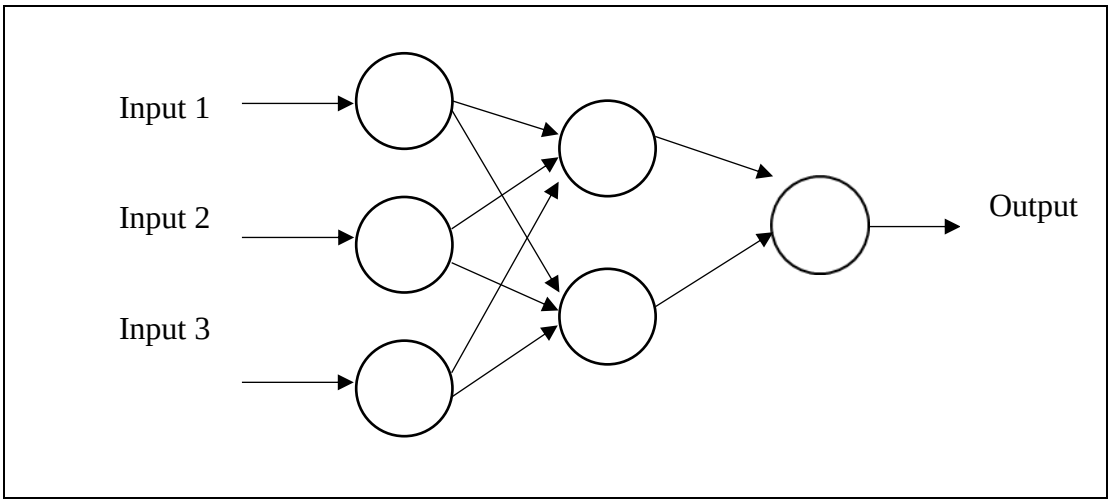

Figure 8. Example ANN architecture.

test portion of the data was used to measure the ANNs ability to accurately predict groundwater level in the target groundwater well by comparing the ANNs predictions to observed groundwater levels. How well the ANN models predict the test data was determined by inspecting the coefficient of determination, R^2, a statistical indicator of model fit. The degree to which the trained ANN could accurately identify outlier groundwater levels was measured by calculating the average prediction error, or root mean square error (RMSE).

III. RESULTS AND DISCUSSION

Clusters of wells responding similarly across hydrologic conditions

A hierarchical clustering analysis was conducted on each of the fifteen, 6-month time-periods to identify clusters of wells for ANN training data. The results of the hierarchical clustering were fifteen dendrograms visualizing each of the fifteen-time-period's set of solutions (**Figures 9-23**). First, the solution sets of each dendrogram were confirmed by comparing the clustering solutions to the respective time-series (**Figures 24-38**) to ensure the hierarchical clustering analysis correctly clustered wells that responded similarly. The dendrograms solution sets appeared to correctly cluster of wells most of the time, except time-period 2 (**Figure 10**) where the set of solutions show low similarity for all the wells. Inspection of the time-series for time-period 2 (**Figure 25**) reveal the data is poor quality as a result of missing and corrupted data. Several of the time-periods had similar data quality issues, such as time-period 3 (figure 10, c) where well L had corrupted data and well D had missing data that had been linearly interpolated, and in time-period 5 (**Figure 28**) well D again had missing data that had been linearly interpolated. However, when comparing the dendrograms of the solution to the time-series of those time-periods, the set clustering solutions were considered to represent the overall patterns of the time-period. Therefore, only time-period 2 clustering solutions were not considered in the hierarchical clustering analysis because the clustering solutions were incomprehensible. Because time-period 2 was rejected from the hierarchical clustering analysis, the criteria for a well to be considered a member of a cluster was adjusted to require a well to cluster with the other members only eleven of the remaining fourteen time-periods.

Next, the dendrograms of the remaining fourteen time-periods were inspected to determine which wells responded similarly by interpreting the height at which a well, or wells, joined another well, or wells, to create a new cluster. An arbitrary threshold, set at a height of 1.0 on the y-axis of each of the fourteen dendrograms, was used as a guideline to compare which wells responded similar across each time-period. Wells that repeatedly clustered together at a low height were considered a cluster of very similar wells and thought to be within similar hydrogeologic settings across all hydrologic conditions, although whether the wells were hydrologically connected could not be determined. And, wells that repeatedly clustered but at varying heights were thought to have different hydrogeologic settings across some portion of the hydrologic conditions. As each of the fourteen time-periods dendrogram were inspected, a generalized clustering pattern emerged that showed four distinct groups of wells cluster together at least eleven of the fourteen time-periods, the criteria set forth in section 2.3.2 for wells to qualify as part of a cluster: wells A, E, G, and L (Cluster 1; **Figure 39**), wells C, K, M, and N (Cluster 2; **Figure 40**), wells B, J, and I (Cluster 3; **Figure 41**), and wells H and D (Cluster 4; **Figure 42**). However, well F failed to cluster with a cluster of wells consistently enough to be considered part of a cluster and was therefore rejected as potential ANN training data.

Cluster 1: wells A, E, G, and L

Inspection of the height of the fourteen dendrograms (**Figures 9-23**), excluding time-period 2, show that Cluster 1 wells (circled) have highly similar responses to changes during all time-periods, except time-periods 4, 10, and 14 (**Figures 12, 18, 22**) where at least one of the four wells in Cluster 1 failed to cluster with the other wells. A

closer inspection of the time-series of the wells in Cluster 1 during time-period 4 (**Figure 27**) showed well E's response is similar to wells A, G, and L during the first half of the time-period, however in the last half of the time-period, well E responded more similarly to wells I, K, M, and N. Although well E clustered with Cluster 2 wells during time-period 4, it was determined that because well E clustered with wells in Cluster 1 for the other thirteen time-periods, and the general shape of well E time-series was more similar to the wells in Cluster 1, well E was considered a member of Cluster 1. In time-period 10, wells E and G clustered together separately than well A and L, which clustered together. But because during this time-period each of the wells clustered with at least one other member of Cluster 1, the clustering solutions for time-period 10 did not affect the inclusion of wells A, E, G, and L as members of Cluster 1. The last time-period where Cluster 1 wells did not cluster together was in time-period 14, where well G did not cluster with wells A, E, and L. Inspection of the time-series for time-period 14 shows that well G had poor quality and missing data. Because during the portion of the time-period the data were not poor quality well G responded similarly to wells A, E, and L, well G was determined to correctly belong with Cluster 1 wells during time-period 14.

During each of the three time-periods where Cluster 1 wells did not all cluster together, hydrologic conditions were getting dryer (**Figure 39**). This indicated that as groundwater levels lowered due to drier conditions, the heterogeneity between the wells in Cluster 1 increased. The wells in cluster 1 are located in the confined artesian zone where groundwater levels are mainly influenced by pressure and groundwater levels can vary greatly. Because of this, heterogeneity of the wells' clustering patterns during drier hydrologic conditions were thought to be the result of changing hydrogeologic properties

as groundwater levels lower in response to drier conditions. However, because during the majority of the time-periods Cluster 1 wells clustered together tightly, thus having a low similarity value, the heterogeneity of the wells responses during dryer times was considered very minimal.

Cluster 2: wells C, K, M, and N

Cluster 2 wells (rectangle) clustered together across most of the clustering time-periods excepts time-period 3, 4 and 12 (**Figures 12-13, 20**). During time-period 3, wells C, K and M clustered together while well N clustered with wells A, E, G, and L. Inspection of the time-series for time-period 3 (**Figure 26**) shows that well N's response to changes is not similar to wells A, E, G, and L, and is actually more similar to wells in Cluster 2, wells C, K and M. Therefore, the results for well N for time-period 3 were not considered when well N was assigned to Cluster 2. During time-period 4, well C did not cluster with Cluster 2 wells, instead clustering with wells B and J. Visual inspection of the time-series for time-period 4 (**Figure 27**) showed that well C was more similar to Cluster 2 wells and the results were not considered when assigning well C to Cluster 2. Then, in time-period 12, Cluster 2 wells split into two clusters: wells C and N clustered together and wells K and M clustered together. Inspection of the time-series for time-period 12 (**Figure 35**) shows that well K appears to have corrupted data and a time-series that is very different than well M's time-series. This highlights the reason the clustering solutions should be compared to the time-series of each time-period to confirm the clustering solutions correctly cluster wells based on response. DTW algorithm measures similarity between two time-series based on the distance between the two time-series, and because of this if two time-series have the same mean response, even though the two

time-series have very different shapes, the wells will have a lower overall similarity value than if the wells had different mean responses. This can lead wells which do not have similar response values to have low overall similarity values, such as the case with wells K and M. In time-period 12, because the mean response of well K and M time-series was more similar, the wells were clustered together even though the shape of the two time-series of the two wells were not similar. Because well K appeared to have poor quality data for time-period 12, the clustering solutions of well K for this time-period were not considered. And, since well M was clustered with well K because the wells had similar mean responses and were not actually similar, well M clustering solutions were also not considered for time-period 12. However, the clustering solutions for wells C and N were used for time-period 12 since the clustering solutions for wells C and N matched the time-series for time-period 12.

The wells in Cluster 2 clustered together tightly most of the time-periods, even during the time-periods when all four wells did not cluster such as in time-periods 3, 4, and 12. Because the well in Cluster 2 are located in the unconfined recharge zone of the aquifer where groundwater levels are mainly influenced by changes in precipitation (instead of pressure like the wells in Cluster 1, 3, and 4), the range of groundwater levels for the wells in Cluster 2 vary little across the 6-year study period. And, since the clustering analysis compared the responses of all fourteen wells, most of which have a much larger range of groundwater levels because they are in the pressure-controlled, confined artesian zone, Cluster 2 wells repeatedly formed into a cluster well below the guideline threshold of 1.0.

Cluster 3: wells B, I, and J

The hierarchical clustering dendrograms (**Figures 9-23**) show that of the fourteen-time-period considered, wells B, I, and J (triangle) clustered together each time-period except time-periods 4, 5, and 8 where well I did not cluster with wells B and J each time. In time-period 4 and 5, hydrologic conditions were transitioning from dry to very dry conditions and in time-period 8 hydrologic conditions were transitioning from wet to dry. Inspection of time-series of time-periods 4, 5, and 8 (**Figures 13-14, 31**) shows wells B and J responses had less magnitude than well I's response. Well B, I, and J are in the confined artesian zone where groundwater levels mainly respond to pressure and vary greatly. In addition, wells in Cluster 3 are known to be under the influence of the Knippa Gap, a hydrogeologic barrier inhibiting groundwater movement during certain hydrogeologic conditions. Further, well I is known to be hydrologically connected to the Frio watershed and experiencing groundwater-surface water interactions that wells B and J do not. The Knippa Gap and interflow from the Frio watershed is thought to be one reason well I respond with more magnitude than wells B and J during certain hydrologic conditions (Ronald T. Green, personal communication, September 2016). This can be seen in the time-series of the time-period 4 and 5, where well I groundwater levels lower more than wells B and J during dry conditions, and again in the time-series for time-period 8 where well I groundwater levels rise more in response to wetter changes.

Cluster 4: wells D and H

Wells D and H (rectangle with curved corners) repeatedly cluster tightly across most time-period (**Figures 9-23**). But, during time-periods 4, 5, and 8, wells D and H did not cluster together (**Figures 24-38**). During time-period 4, well D clustered with well F

(**Figure 12**) and inspection of the time-series of time-period 4 (**Figure 27**) shows well D and F indeed had more similar responses than the other wells during the last half of the time-period where groundwater levels in well D and F declined more quickly than the other wells. However, during the first half of the time-period, well D had a large amount of missing data that had been linearly interpolated and may have affected the accuracy of the rest of the time-periods responses. And, during time-period 5, well D again had large amounts of missing data (**Figure 28**). Therefore, well D clustering solutions for time-period 4 and 5 not used in the hierarchical clustering analysis.

During time-period 8, a very wet period that was getting drier quickly, wells D and H clustered together loosely. Inspection of the time-series of time-period 8 (**Figure 31**) shows Wells D and H have a similar response, but well H responses were more dramatic. Well D and H were thought to be hydrologically connected to each other based on the wells proximity to each other in the aquifer and the general shape of the well's time-series. The wells are located in the confined artesian zone of the aquifer where groundwater levels for wells D and H ranged less than 20 feet during the study period. For wells that are hydrologically connected, groundwater levels can respond quickly to change as conduits of varying sizes transport large volumes of groundwater between wells quickly, depending on their degree of connectivity. Therefore, the loose clustering of well D and H was thought to be because well H may have a higher connectivity to other surrounding wells and therefore receive preferential groundwater flow from other wells, while well D may have received less.

Discussion

The patterns in the solutions of the fourteen dendrograms show that for most of the wells, regardless of whether the well was confined or not, as hydrologic conditions change so does their response and to a varying degree. Certain clusters of wells responses remained highly similar across hydrologic conditions, such as Cluster 1 (wells A, E, G, and L) and Cluster 2 (wells C, K, M, and N), while the degree to which other clusters of wells exhibited similar responses changed across hydrologic conditions, such as Cluster 3 (wells B, I, and J) and Cluster 4 (wells D and H). For Cluster 1, the plotted raw time-series of the wells shows the high similarity of the wells responses across most hydrological conditions (**Figure 39**). The plotted raw time-series of the wells in Cluster 2 appear to indicate that the wells have less similarity (**Figure 40**). However, this is because the y-axis of the plot has a much smaller range of values since Cluster 2 wells are located in the unconfined recharge zone of the aquifer and the wells had highly similar. Clusters 3 and 4 wells had highly similar responses under certain hydrologic conditions, then less similarity under different hydrologic conditions, reflected in the raw time-series of each the clusters of wells (**Figures 41-42**). During certain time-periods, some wells showed very little similarity to the cluster of wells it most often joins. This was attributed to possible changes in the hydrogeologic properties of a well as hydrologic conditions oscillate from very wet to very dry (causing groundwater levels rise and lower) and changes in a wells hydrologic connectedness to other wells in the aquifer.

It is known that the hydrogeological connectivity of the Edwards Aquifer changes across hydrological conditions. A recent study by Land et al. (2011) reported a change in preferential flow during times of drought that allowed groundwater to bypass the San Marcos springs and instead contribute to the Barton Springs flow, which was initially

observed in groundwater level datasets. In this study, the result of the hierarchical clustering analyses shows that when hydrologic conditions were transitioning between wet to dry, all the clusters of wells had more heterogeneity between well responses. For example, each of the four clusters had wells that did not fully cluster during time-period 4, 5, 8, or 10, time-periods when hydrologic conditions were transitioning to drier conditions. The results of a hierarchical clustering analysis, such as this analysis which clusters well responses across hydrologic conditions, can be used to confirm the suspected changing preferential flow paths in the Edwards Aquifer by comparing what is known about the aquifer to the dendrogram solutions across hydrologic conditions.

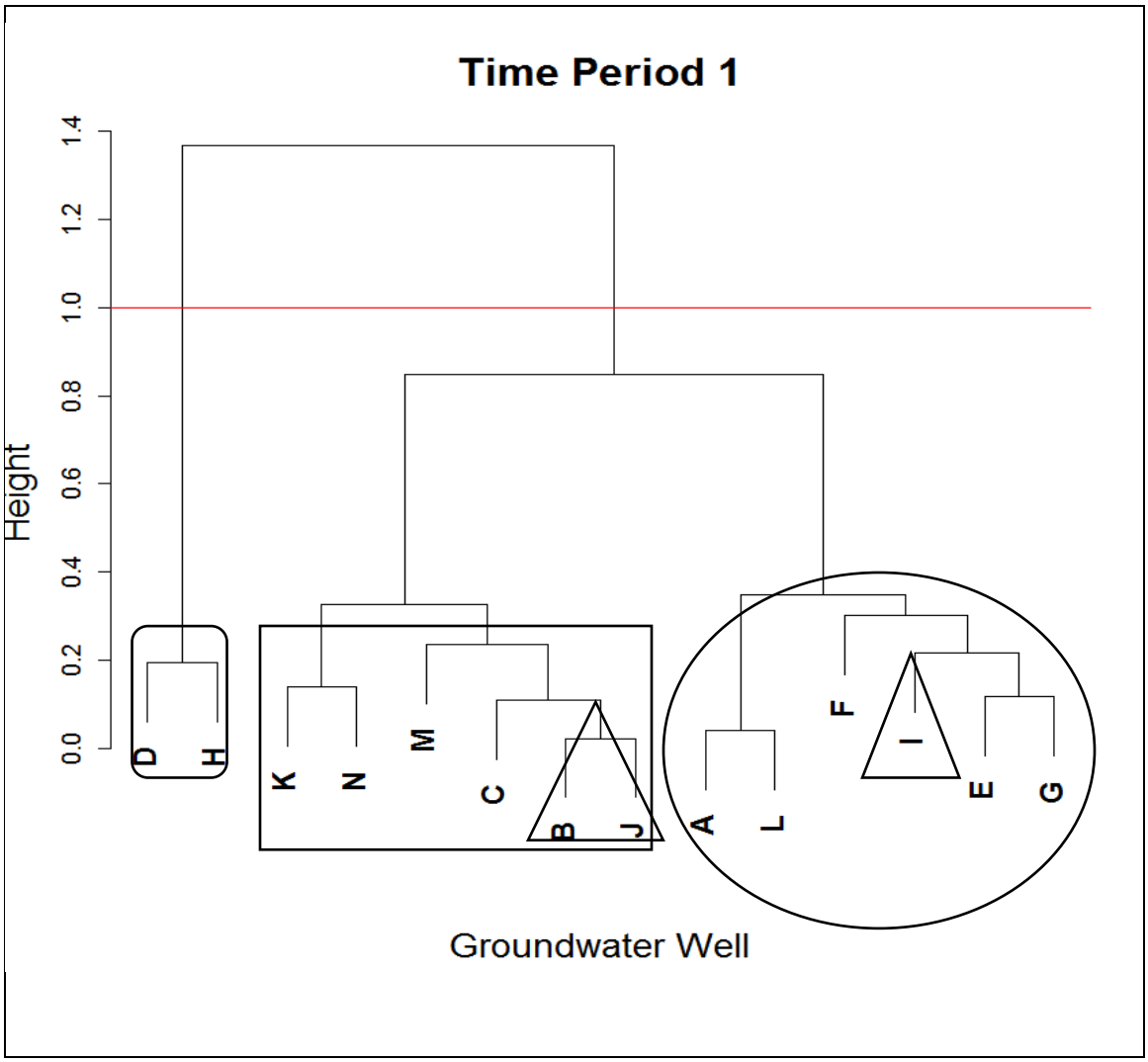

Figure 9. Dendrogram for time-period 1.

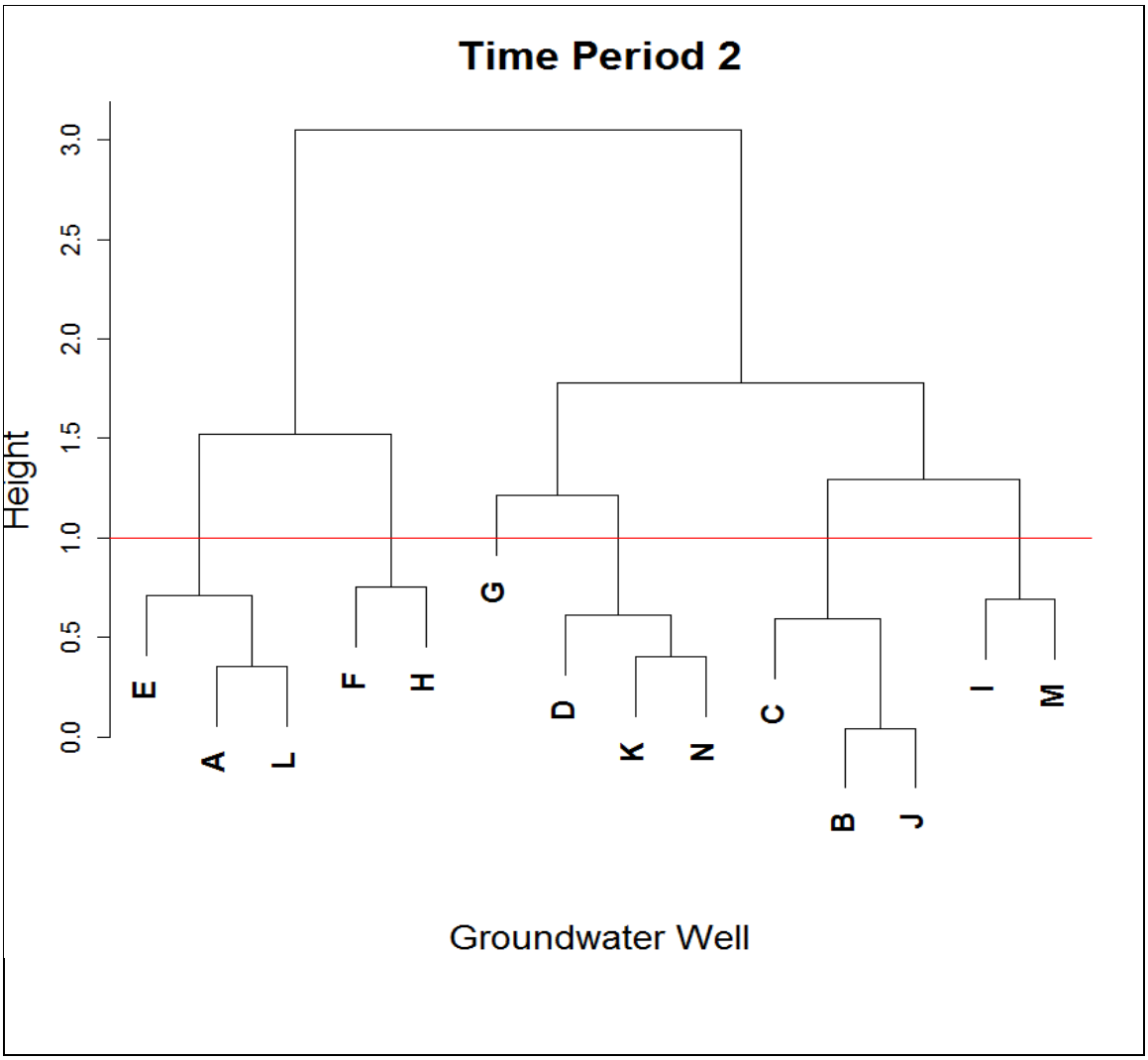

Figure 10. Dendrogram for time-period 2.

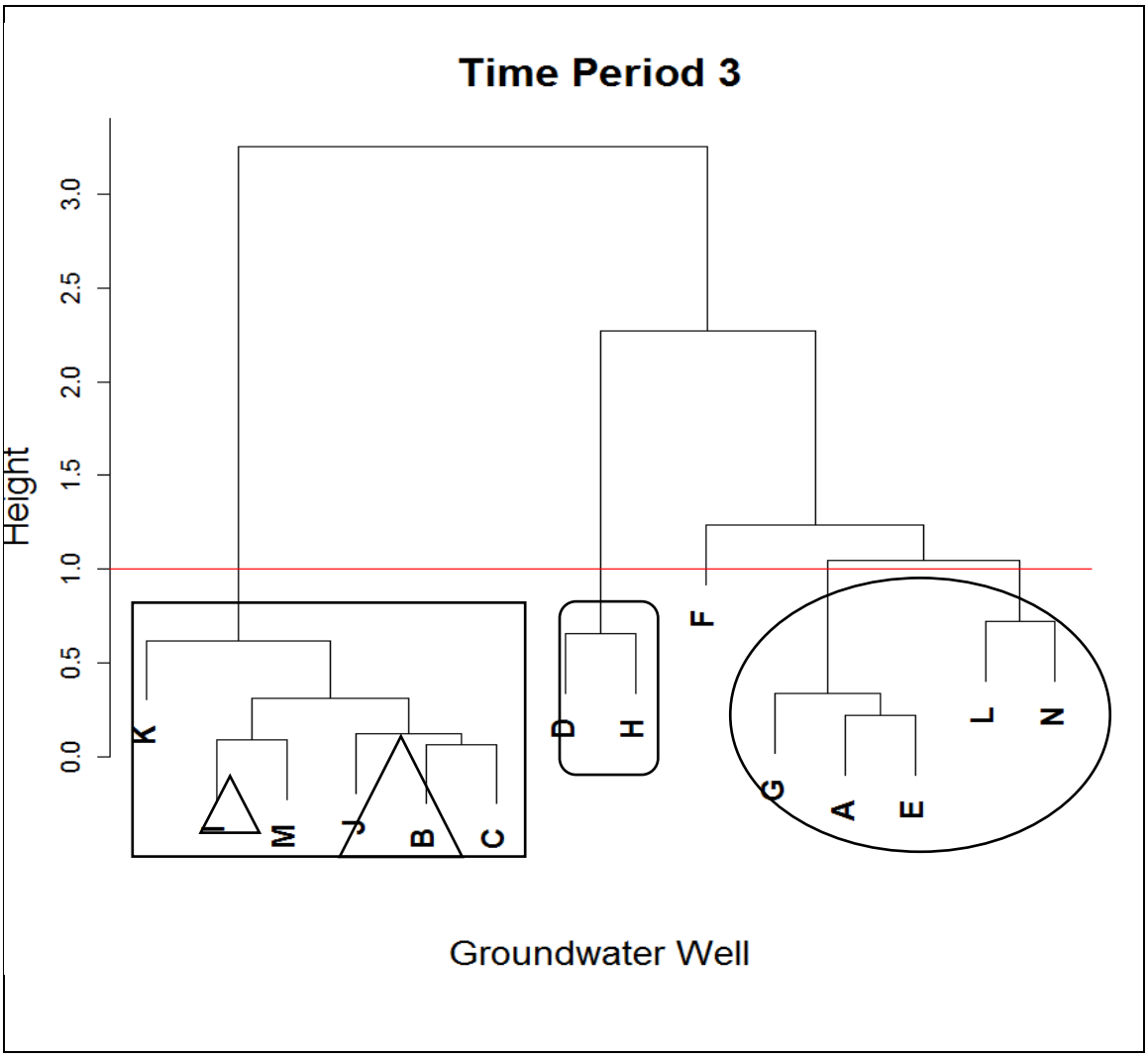

Figure 11. Dendrogram for time-period 3.

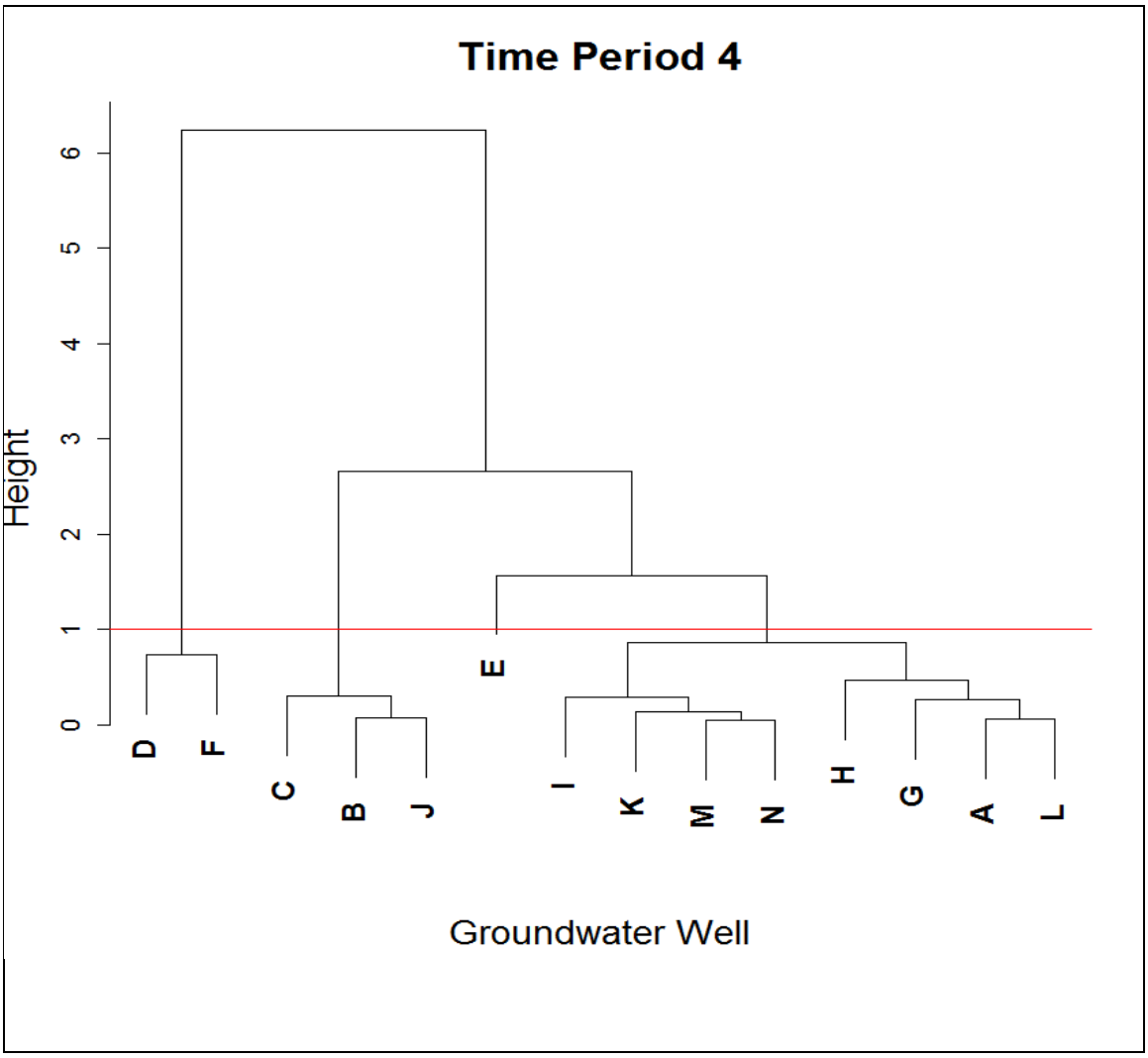

Figure 12. Dendrogram for time-period 4.

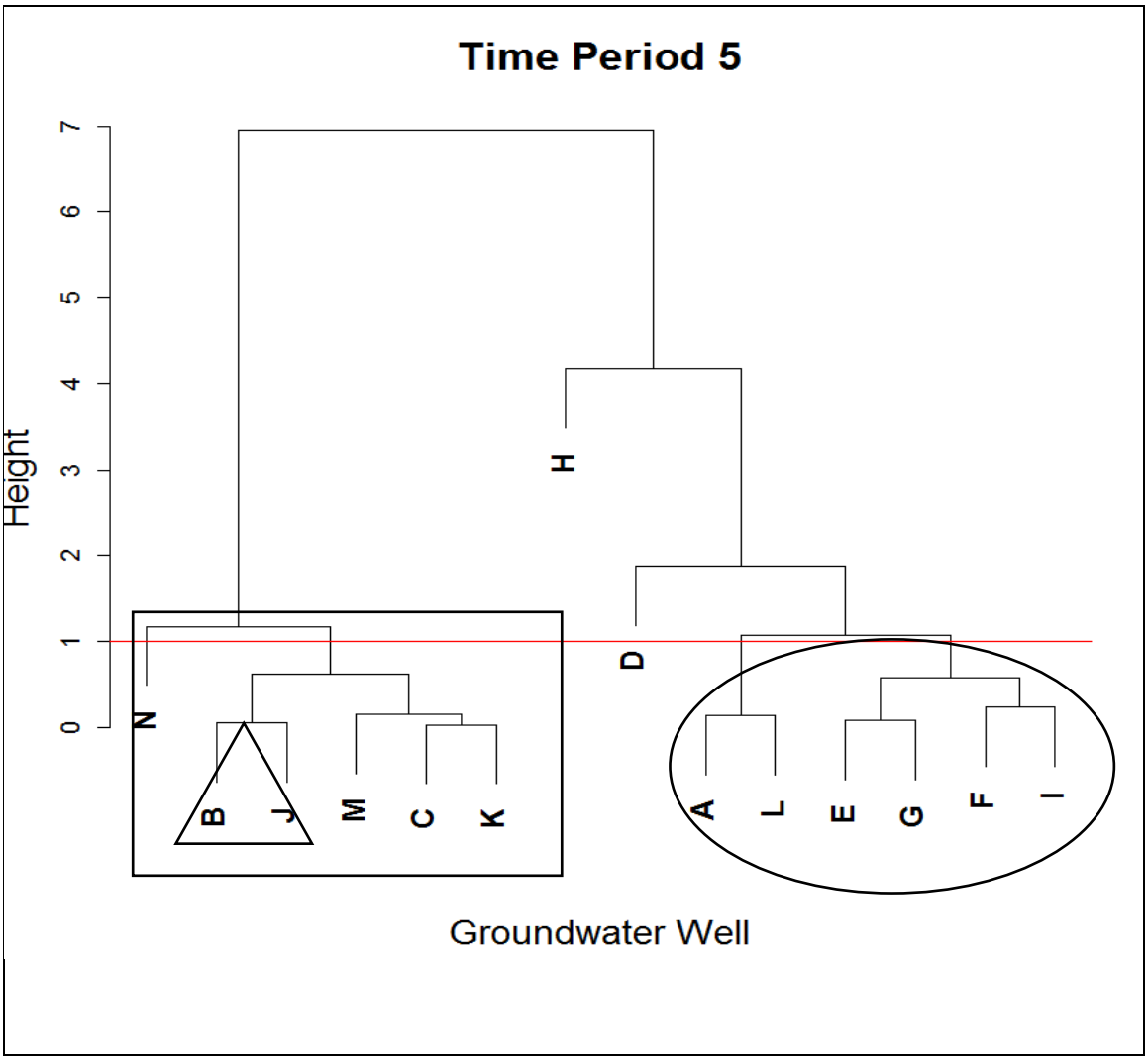

Figure 13. Dendrogram for time-period 5.

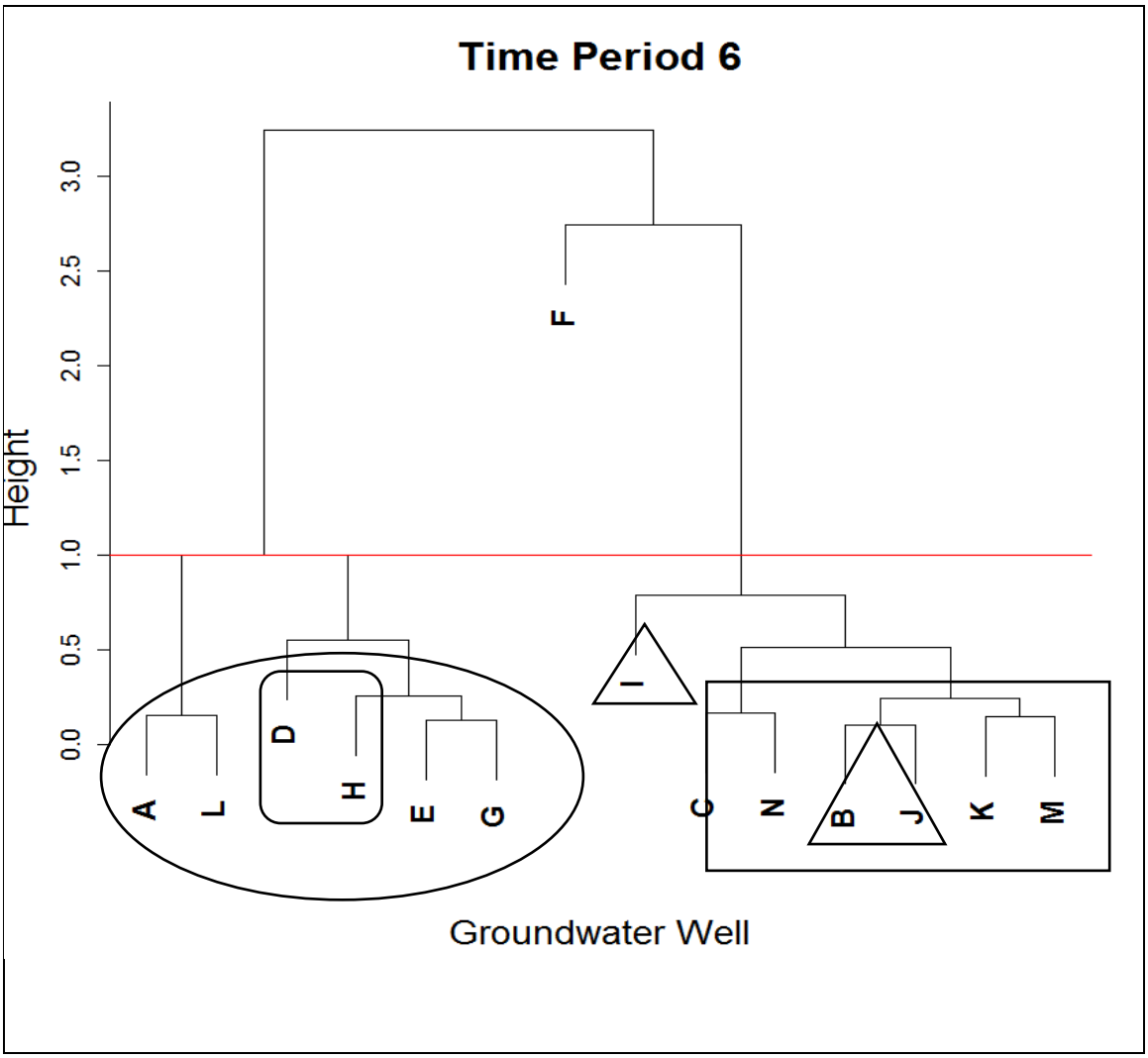

Figure 14. Dendrogram for time-period 6.

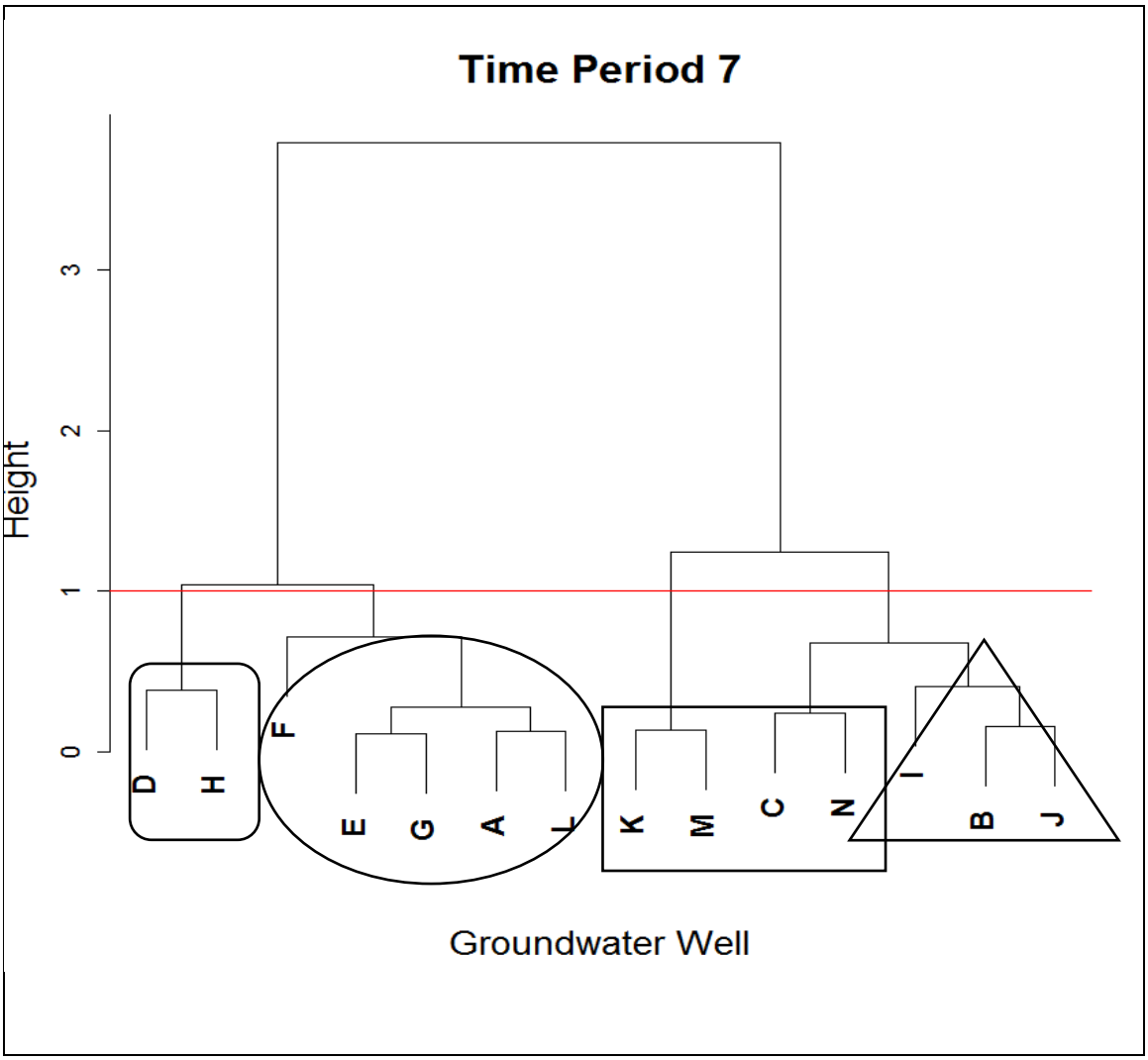

Figure 15. Dendrogram for time-period 7.

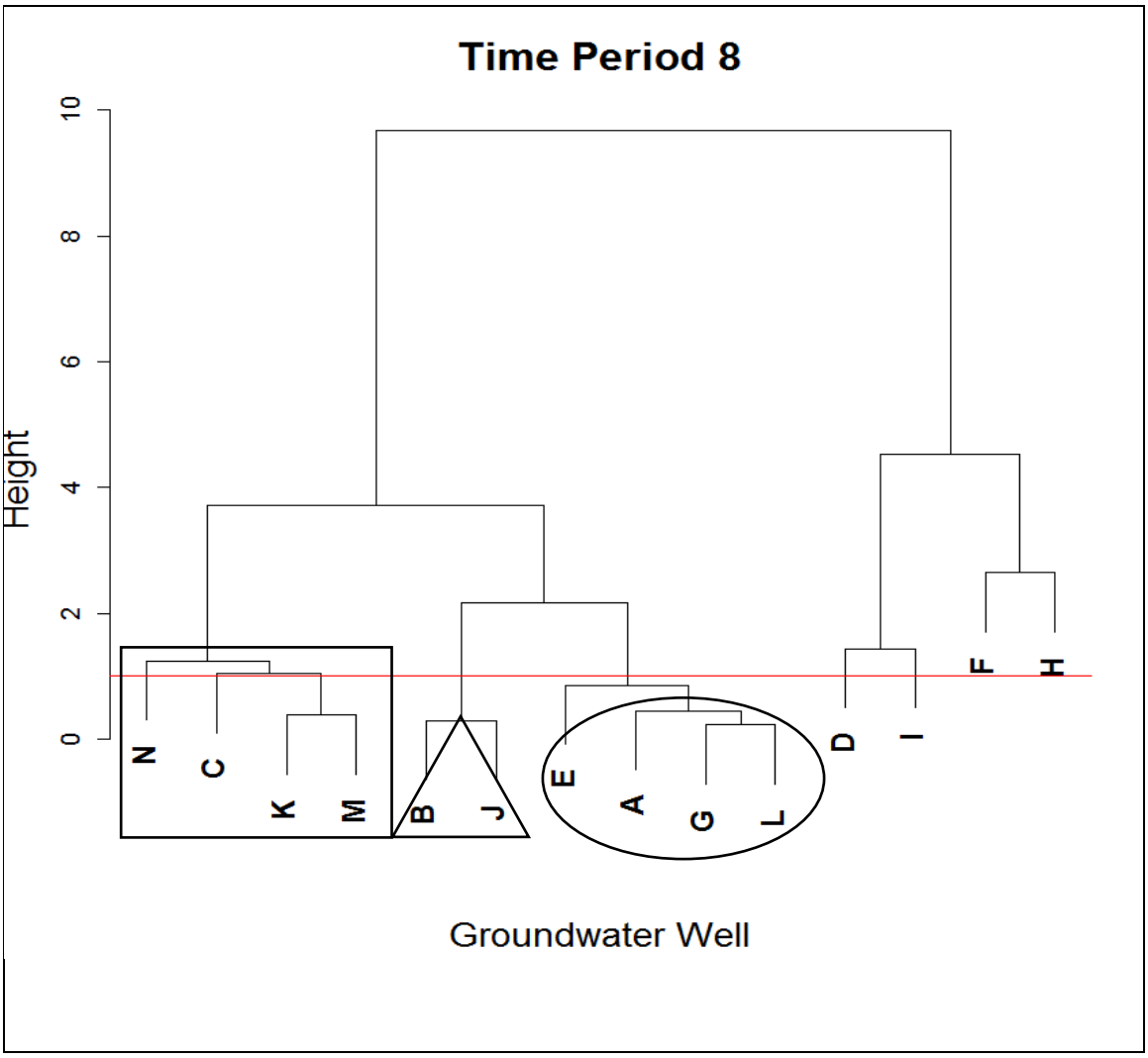

Figure 16. Dendrogram for time-period 8.

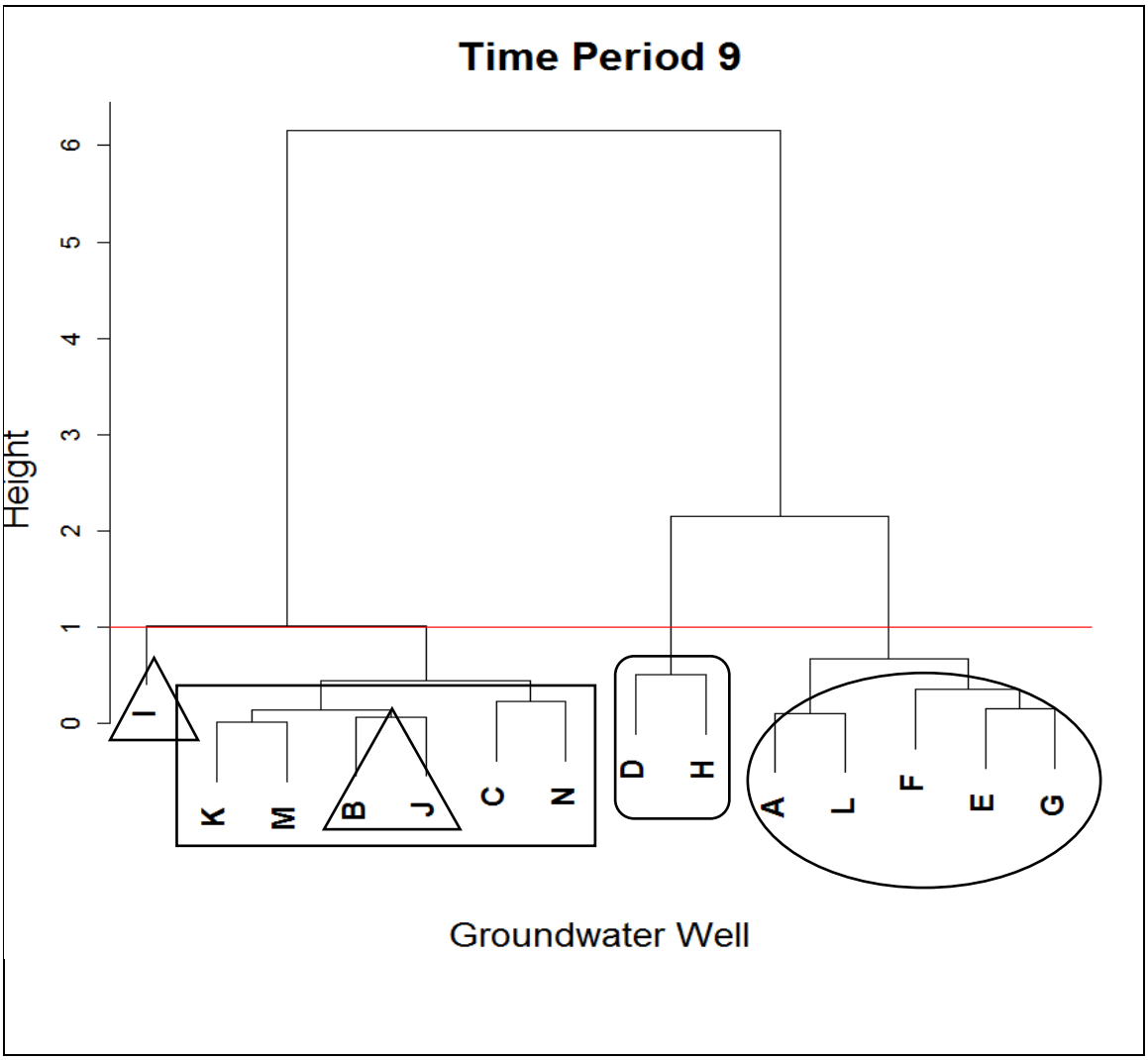

Figure 17. Dendrogram for time-period 9.

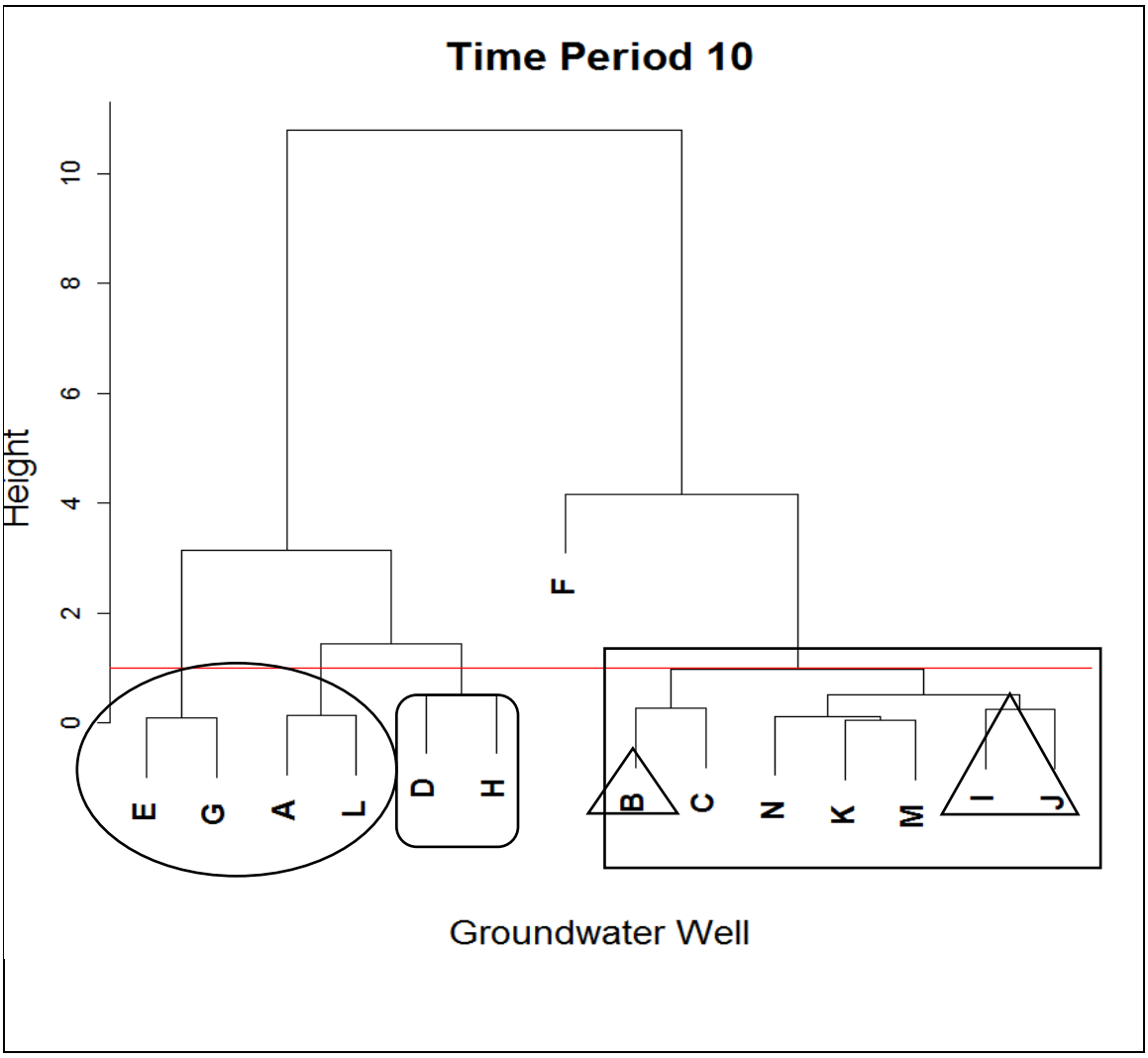

Figure 18. Dendrogram for time-period 10.

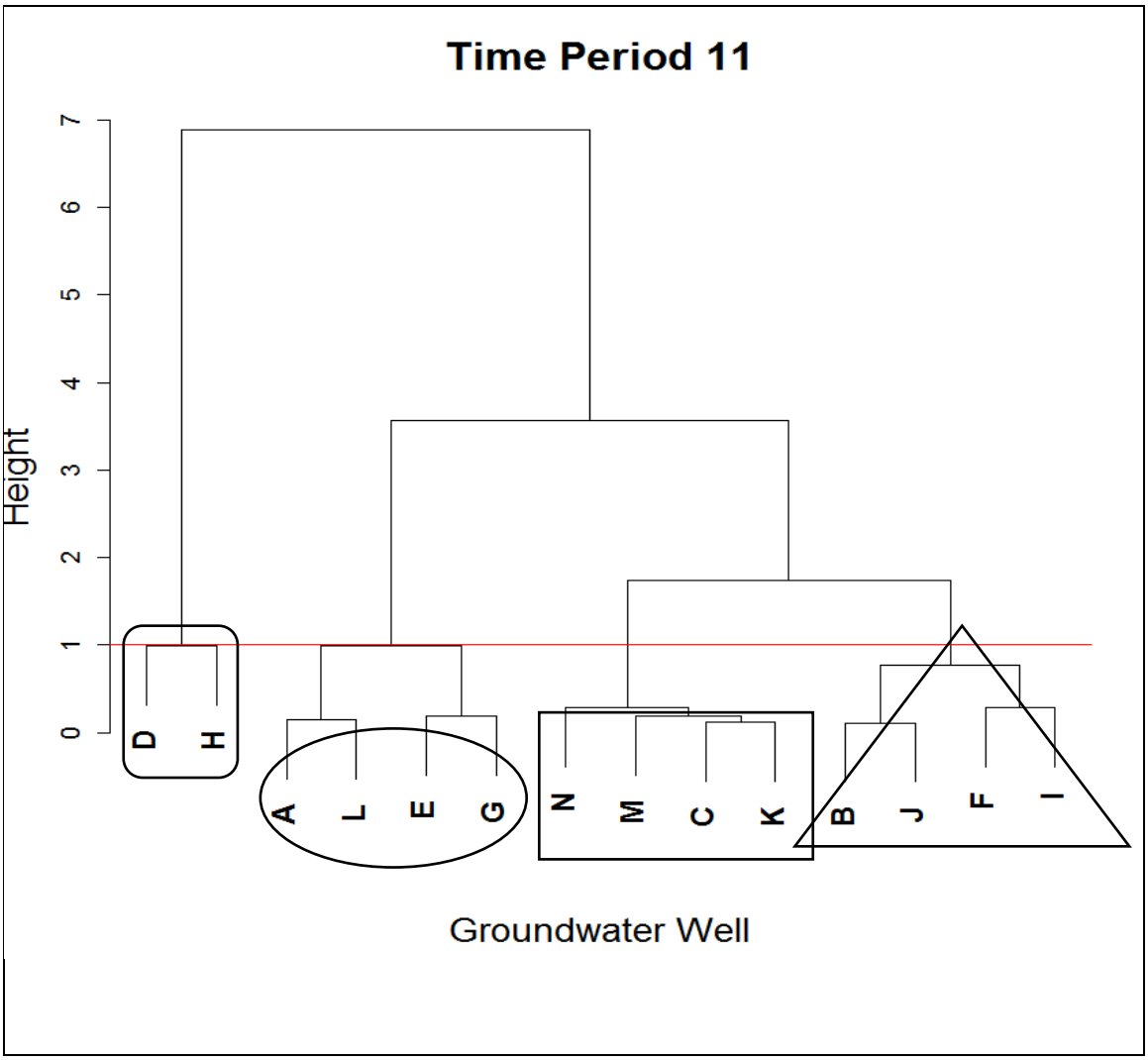

Figure 19. Dendrogram for time-period 11.

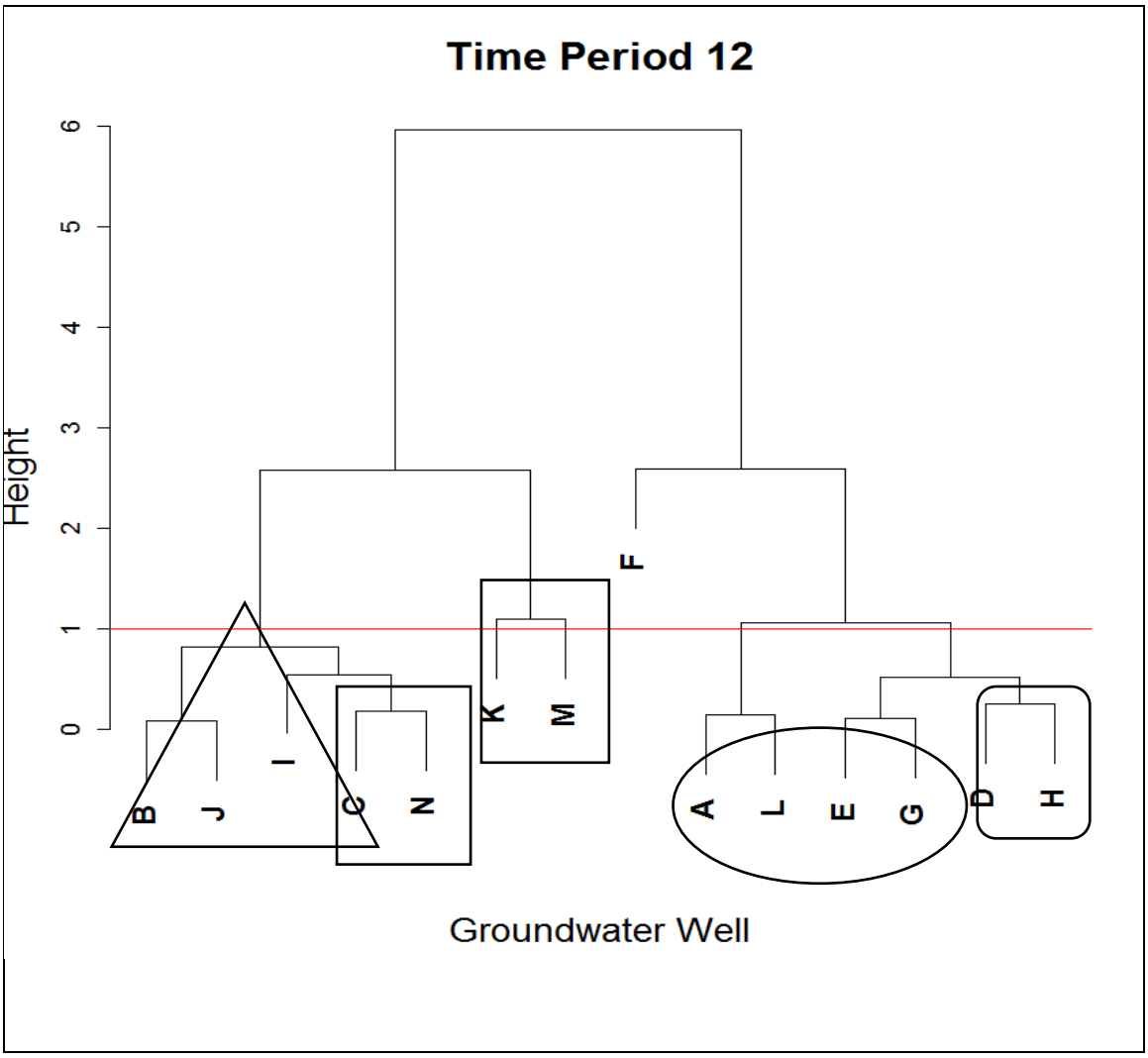

Figure 20. Dendrogram for time-period 12.

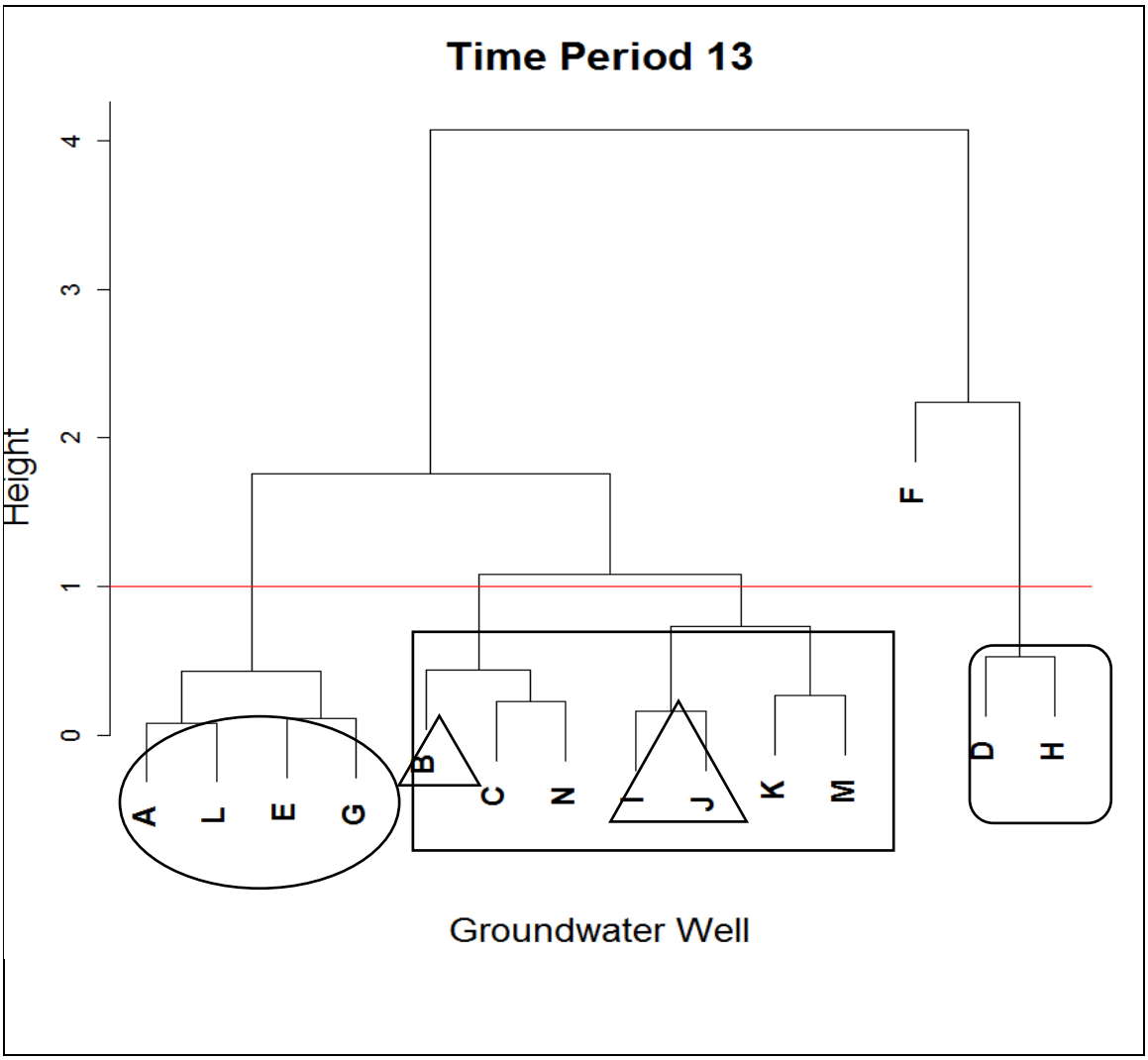

Figure 21. Dendrogram for time-period 13.

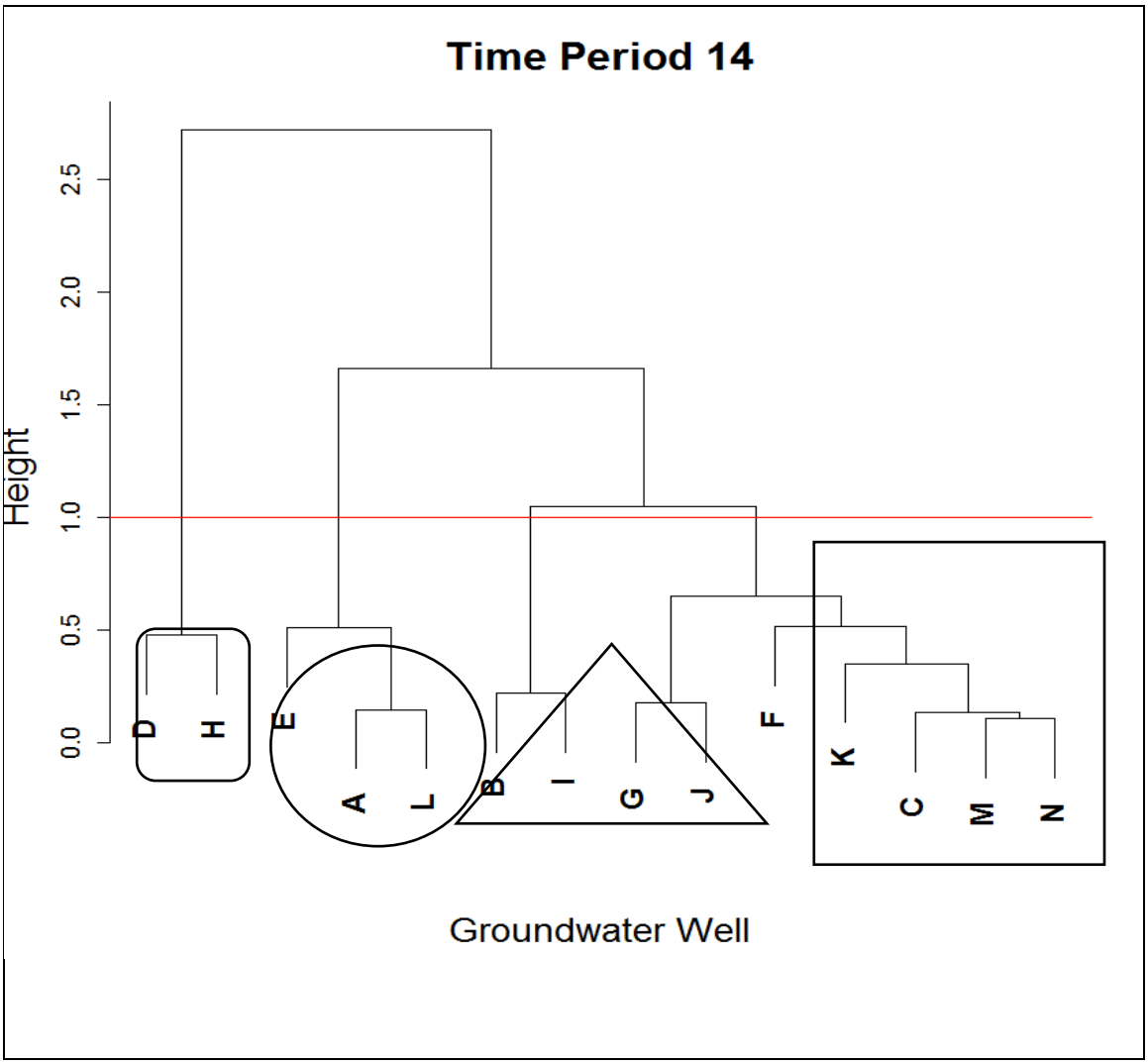

Figure 22. Dendrogram for time-period 14.

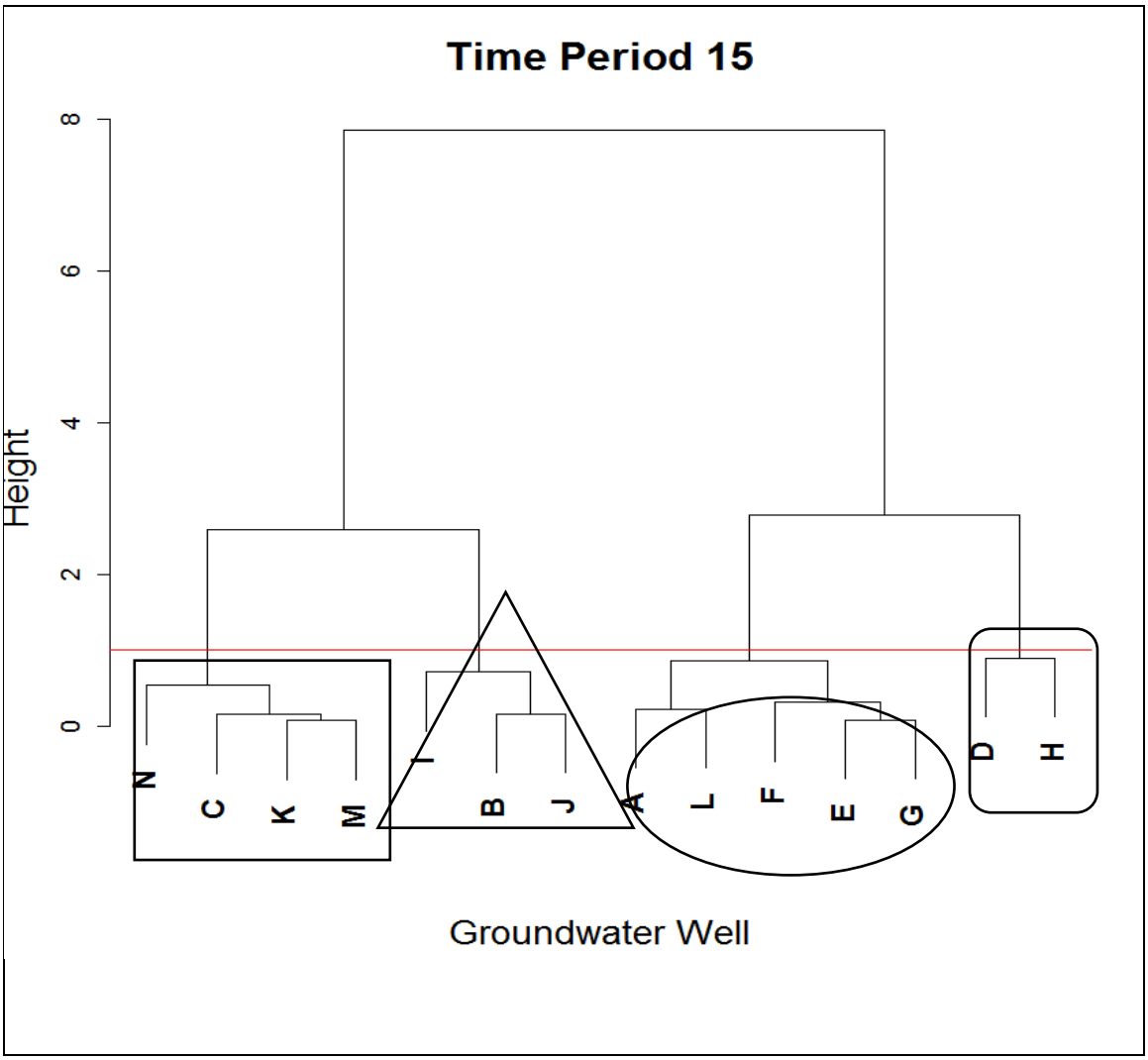

Figure 23. Dendrogram for time-period 15.

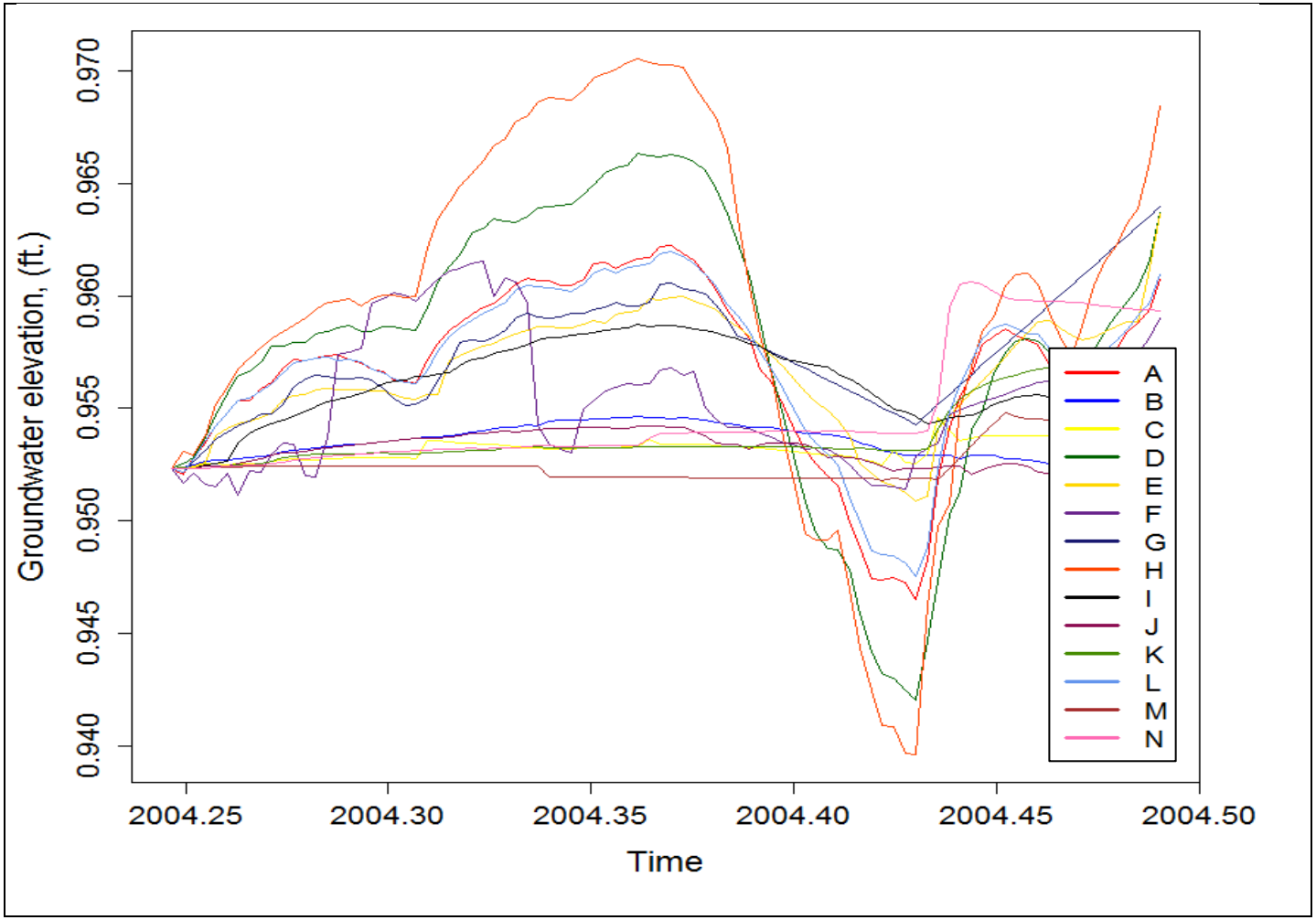

Figure 24. Time-series of the fourteen wells during time-period 1.

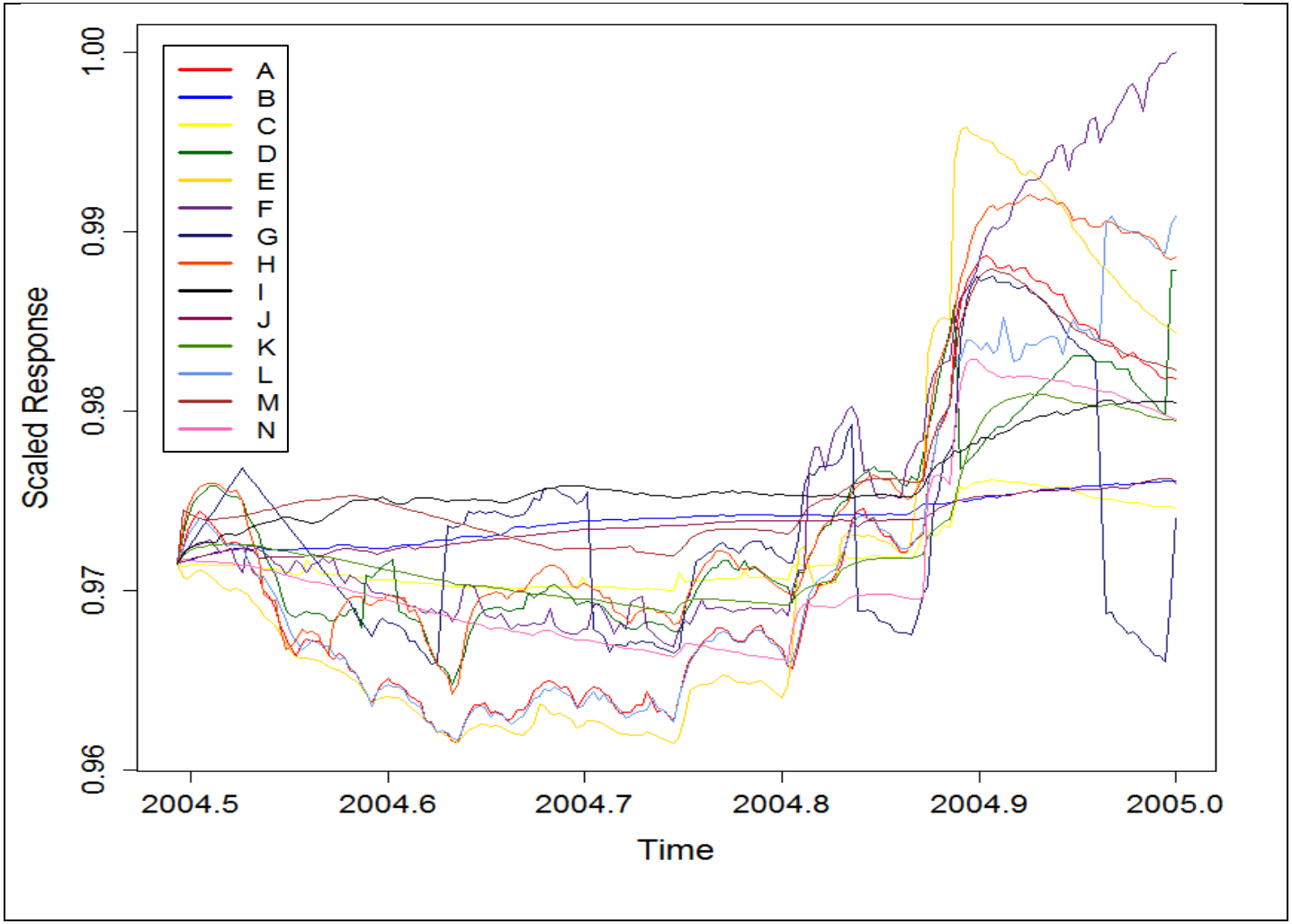

Figure 25. Time-series of the fourteen wells during time-period 2.

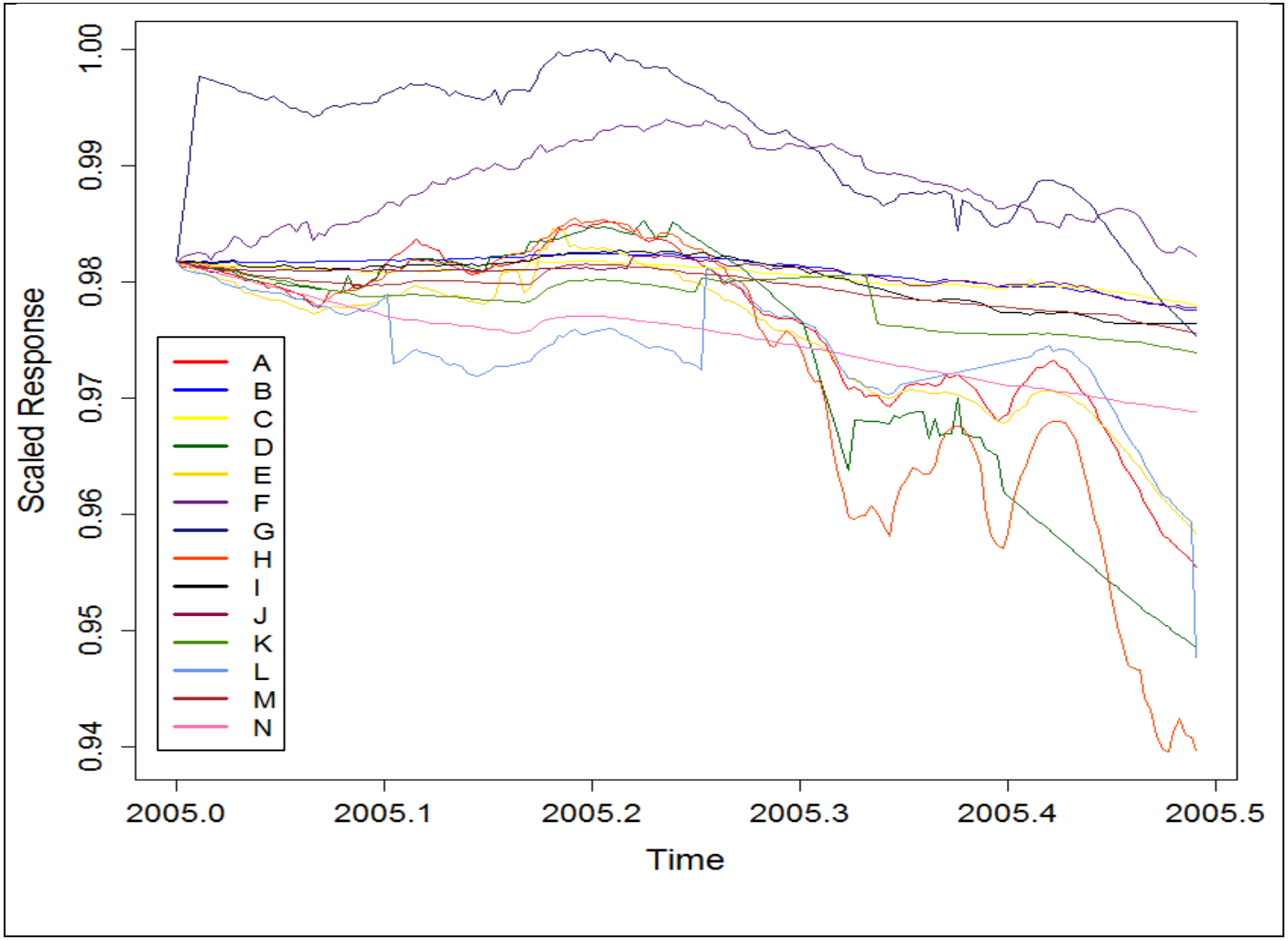

Figure 26. Time-series of the fourteen wells during time-period 3.

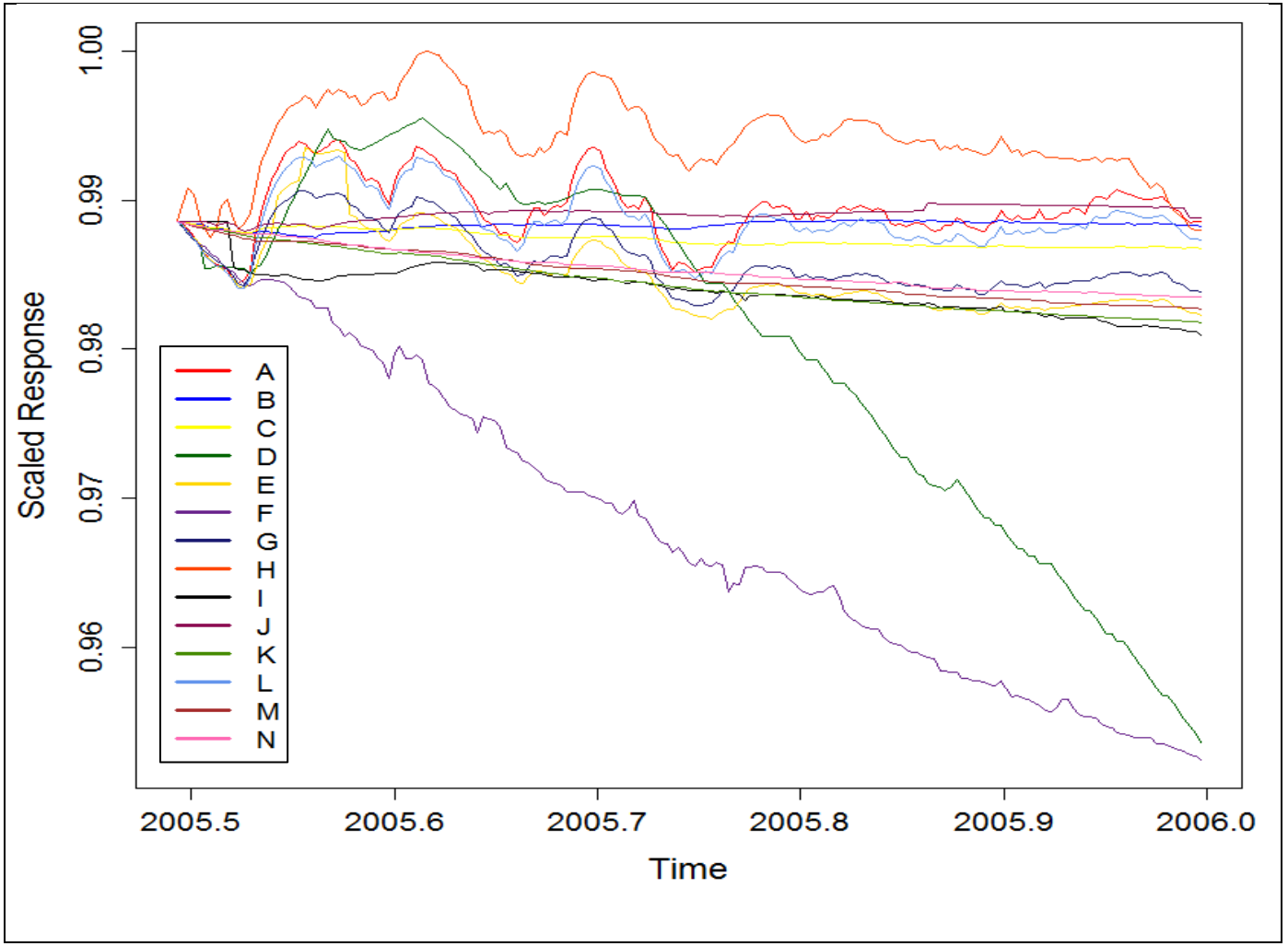

Figure 27. Time-series of the fourteen wells during time-period 4.

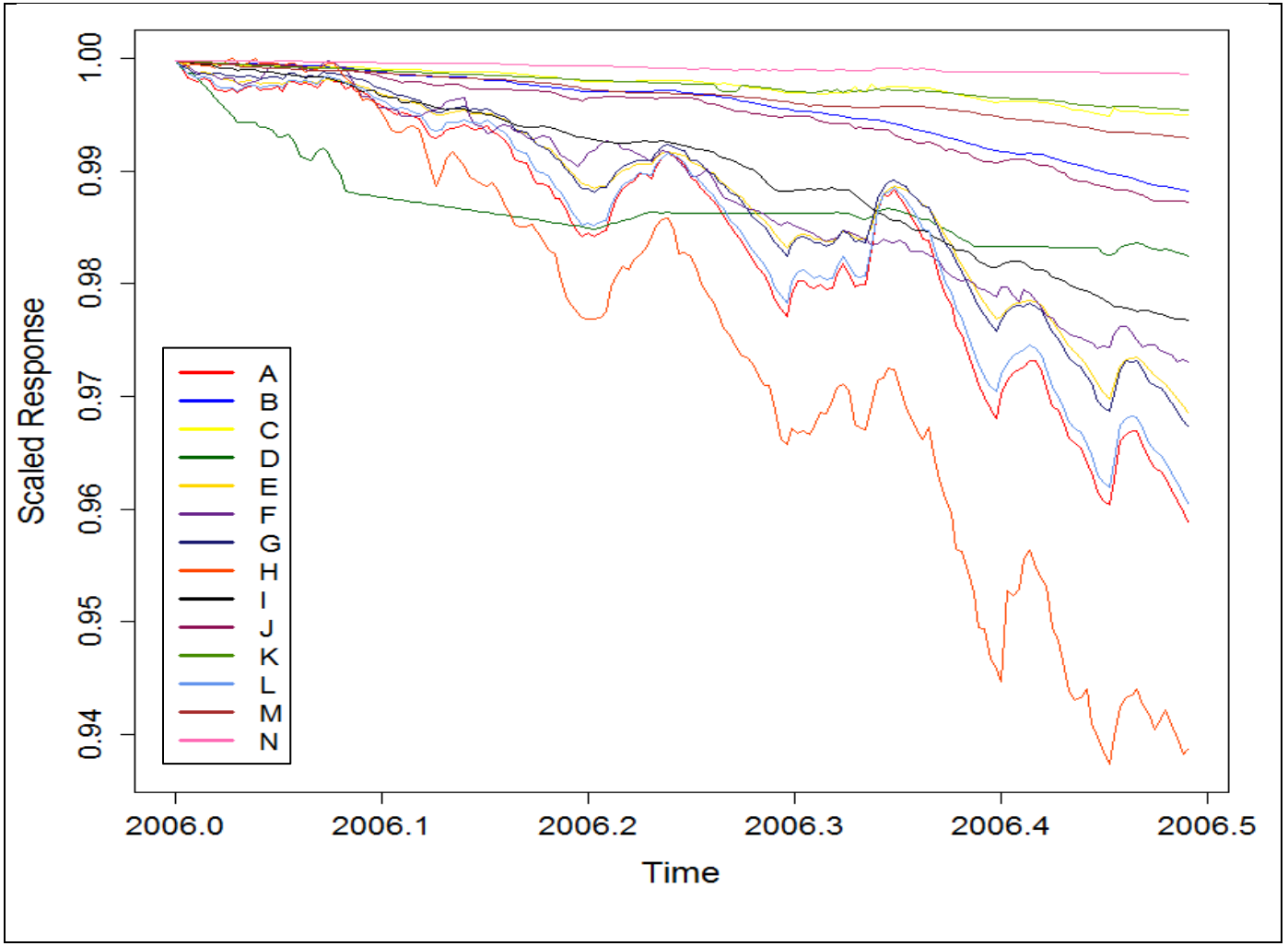

Figure 28. Time-series of the fourteen wells during time-period 5.

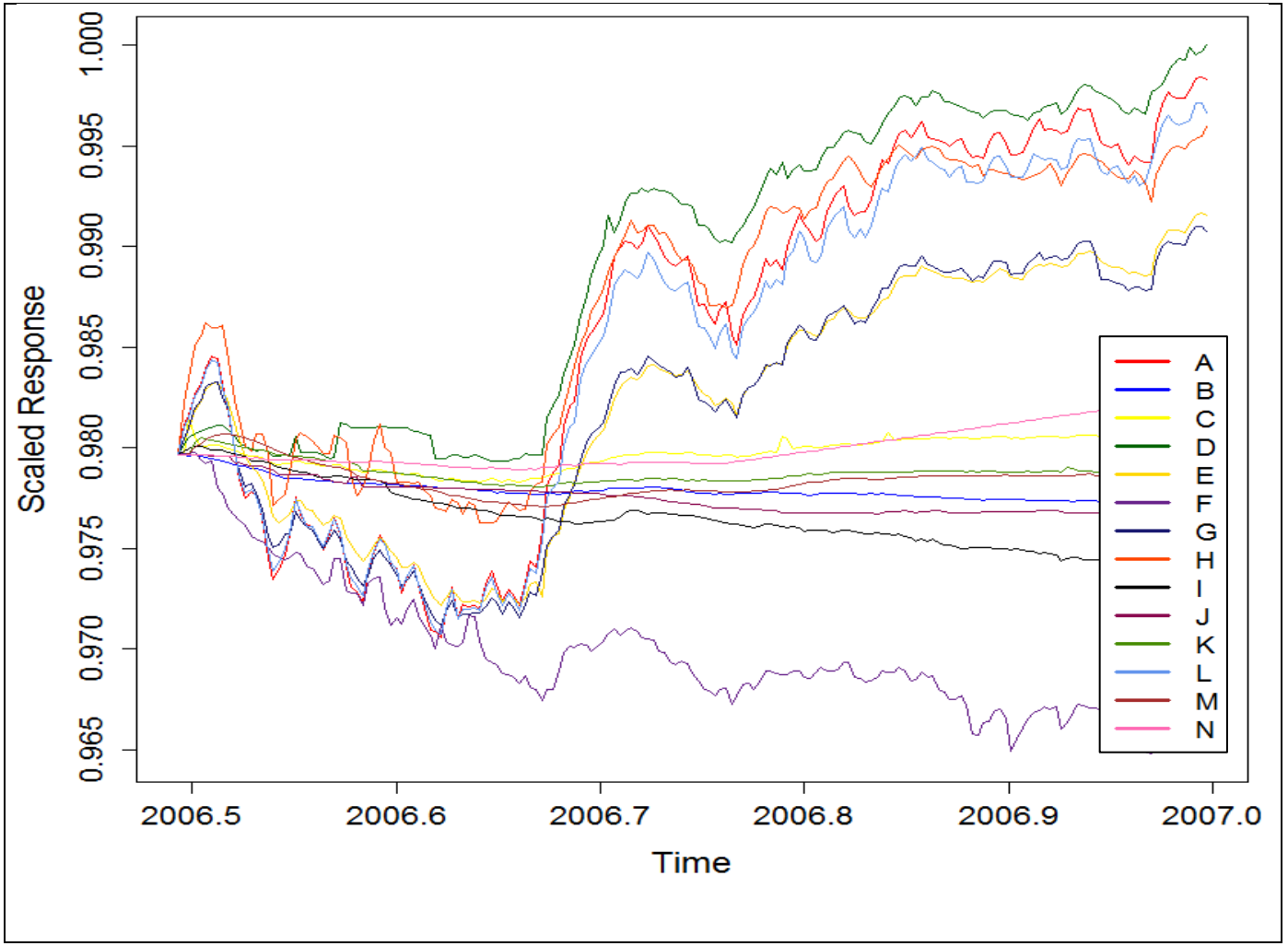

Figure 29. Time-series of the fourteen wells during time-period 6.

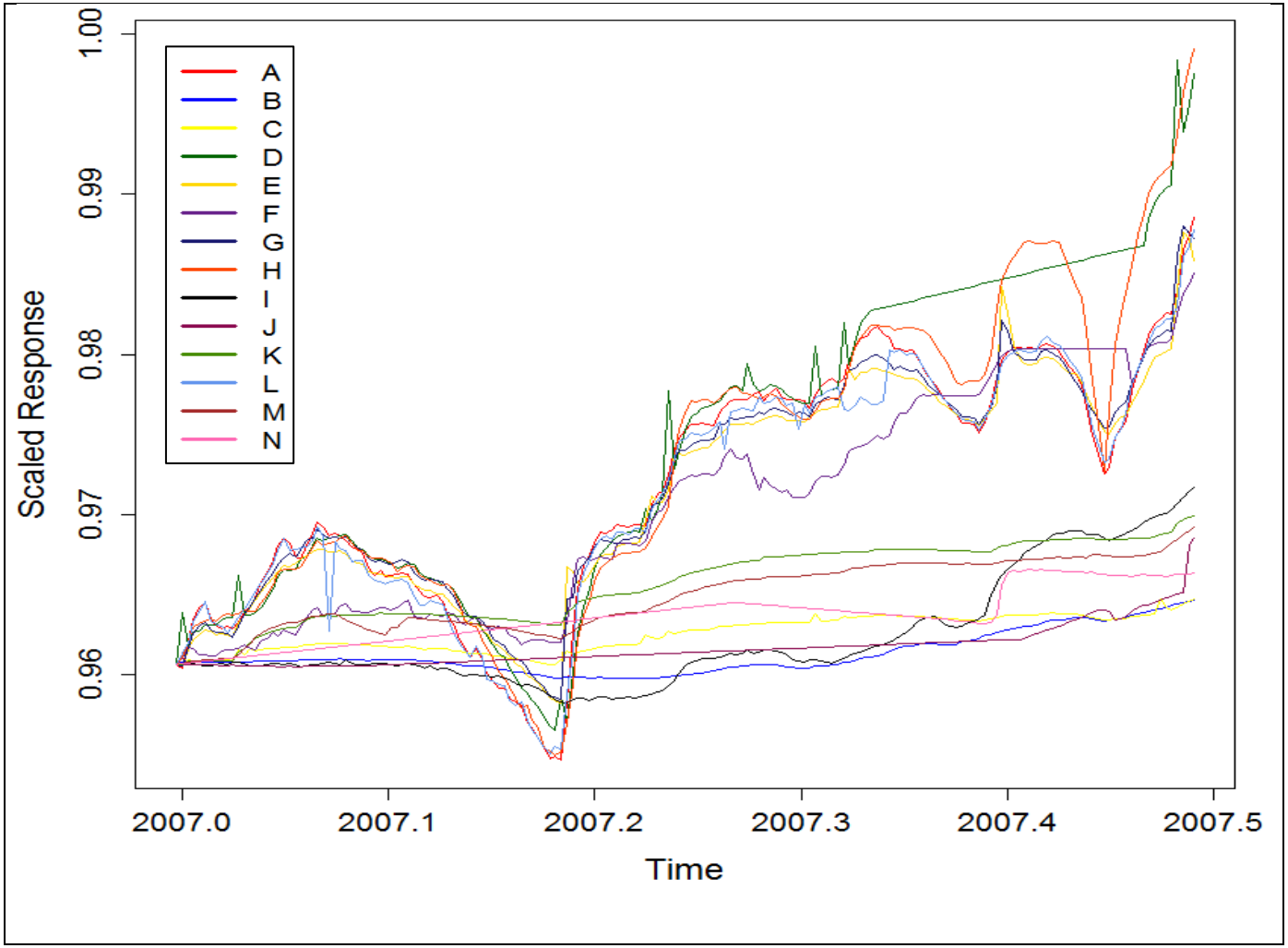

Figure 30. Time-series of the fourteen wells during time-period 7.

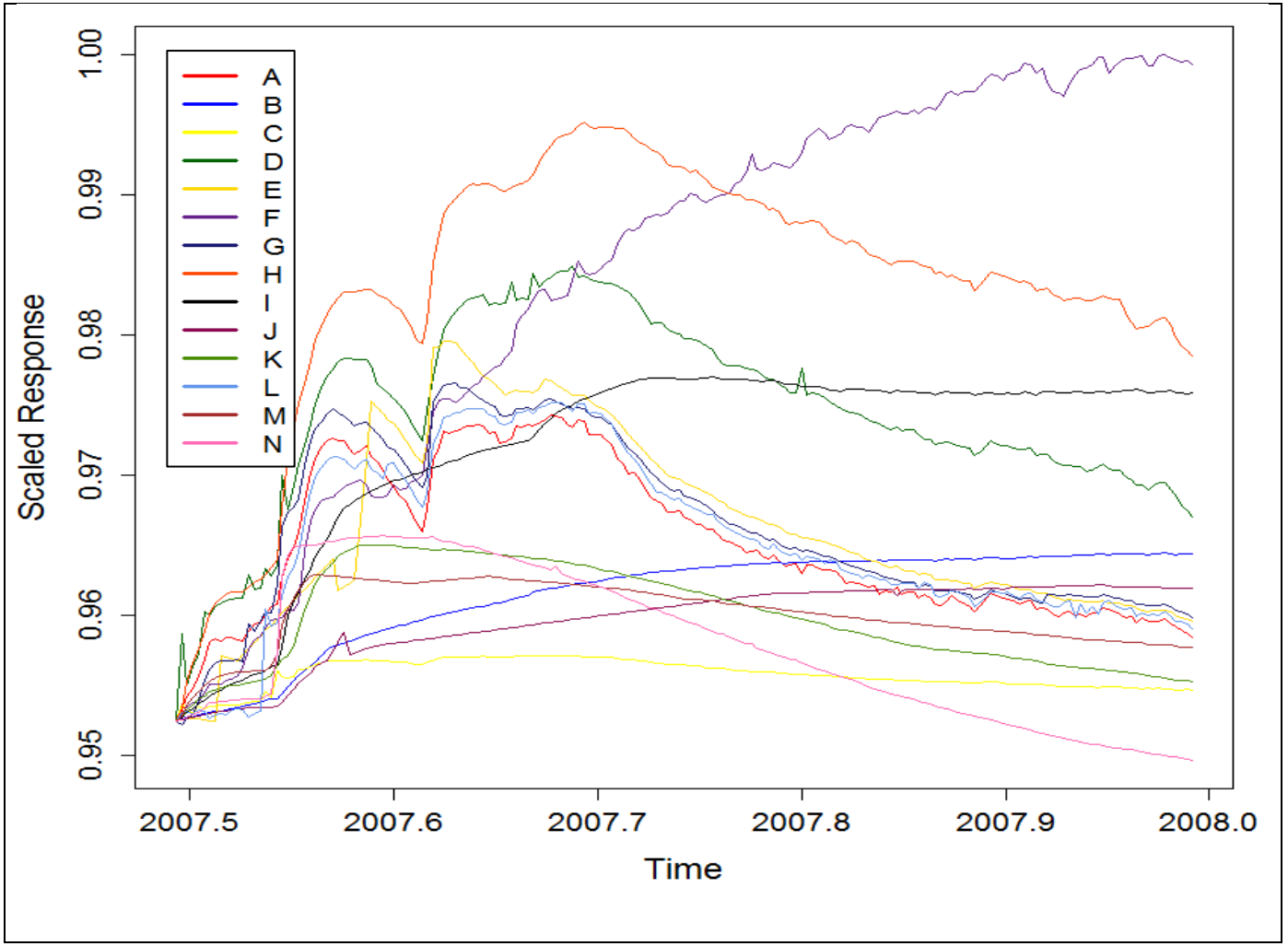

Figure 31. Time-series of the fourteen wells during time-period 8.

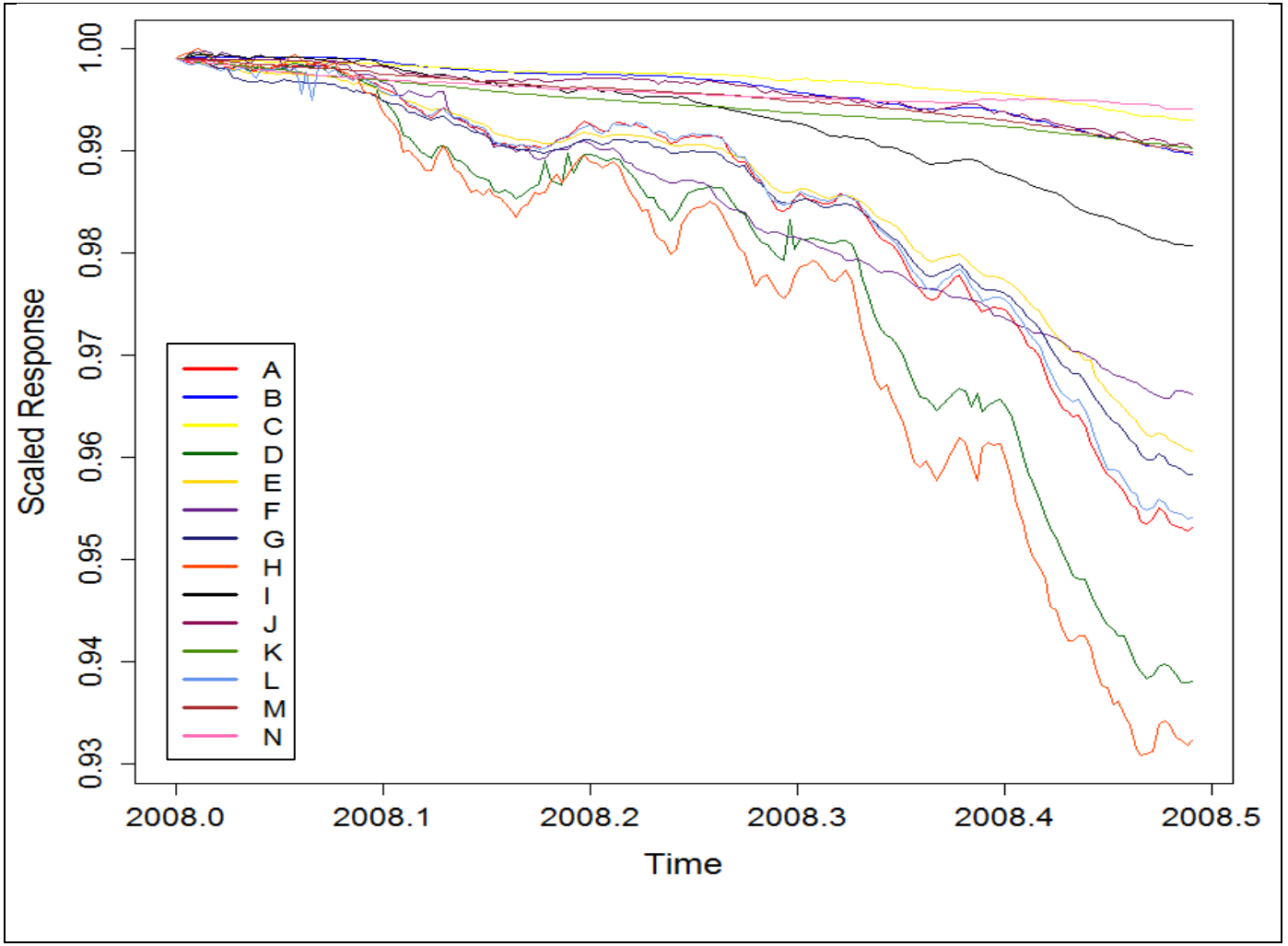

Figure 32. Time-series of the fourteen wells during time-period 9.

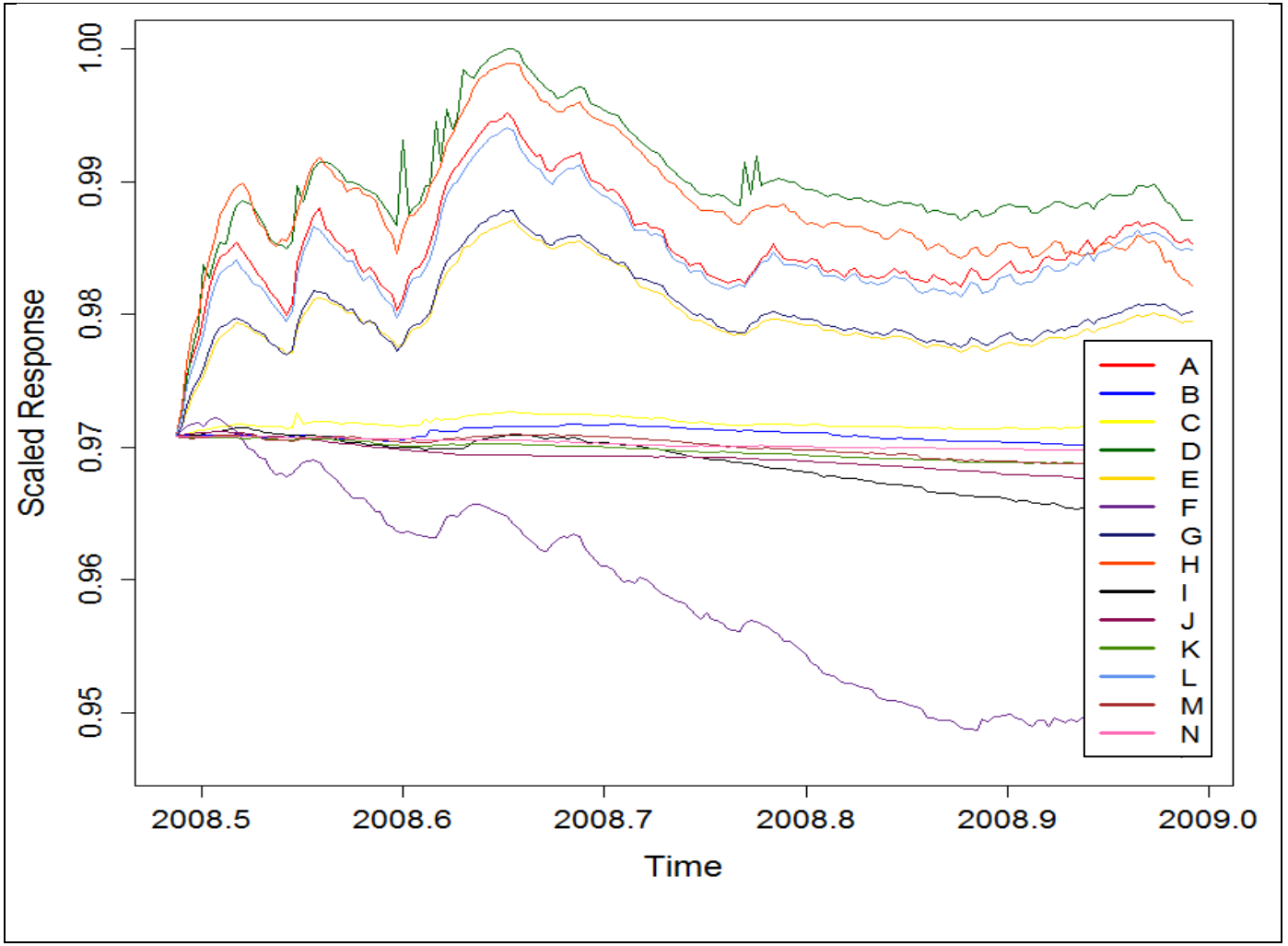

Figure 33. Time-series of the fourteen wells during time-period 10.

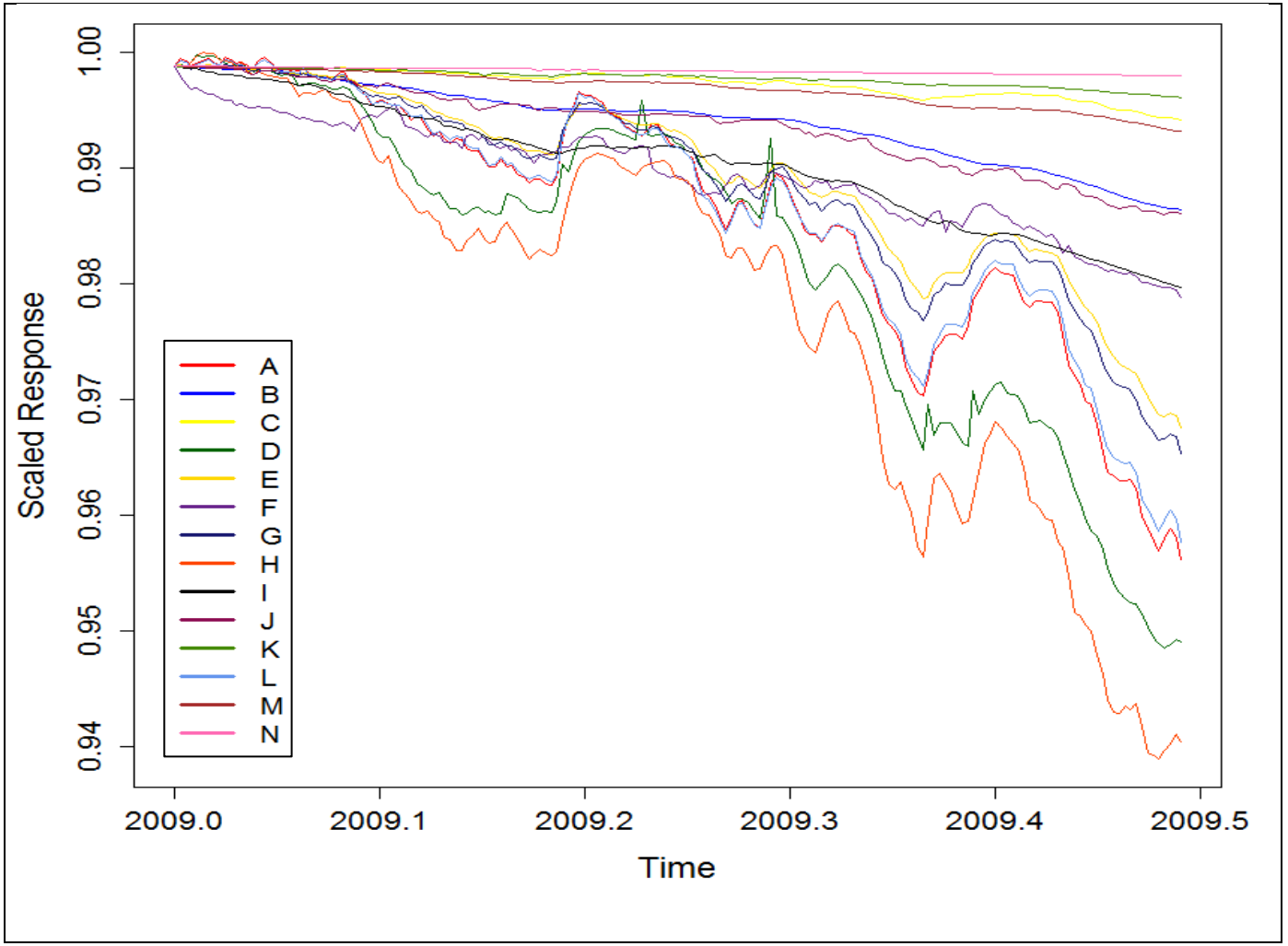

Figure 34. Time-series of the fourteen wells during time-period 11.

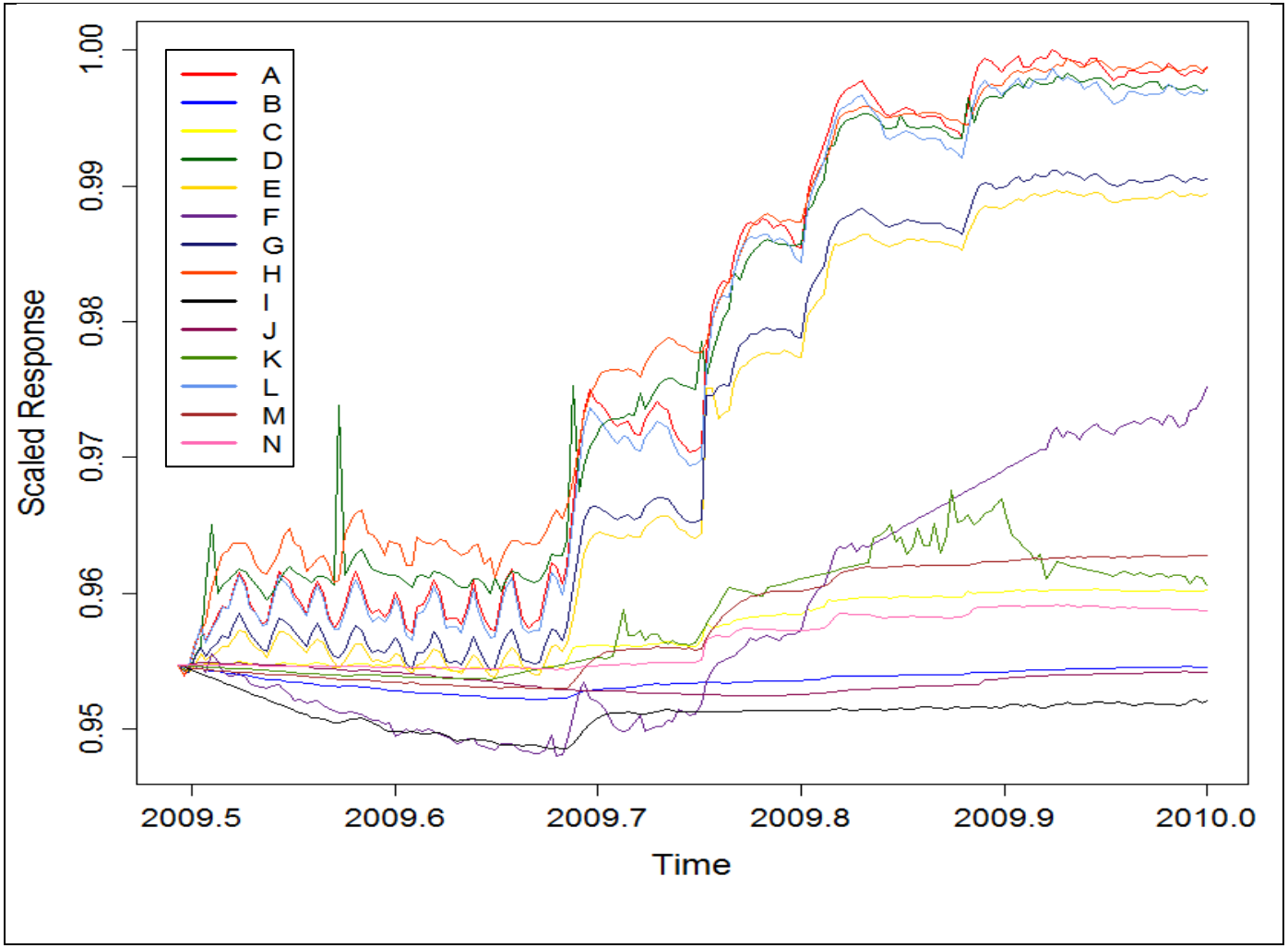

Figure 35. Time-series of the fourteen wells during time-period 12.

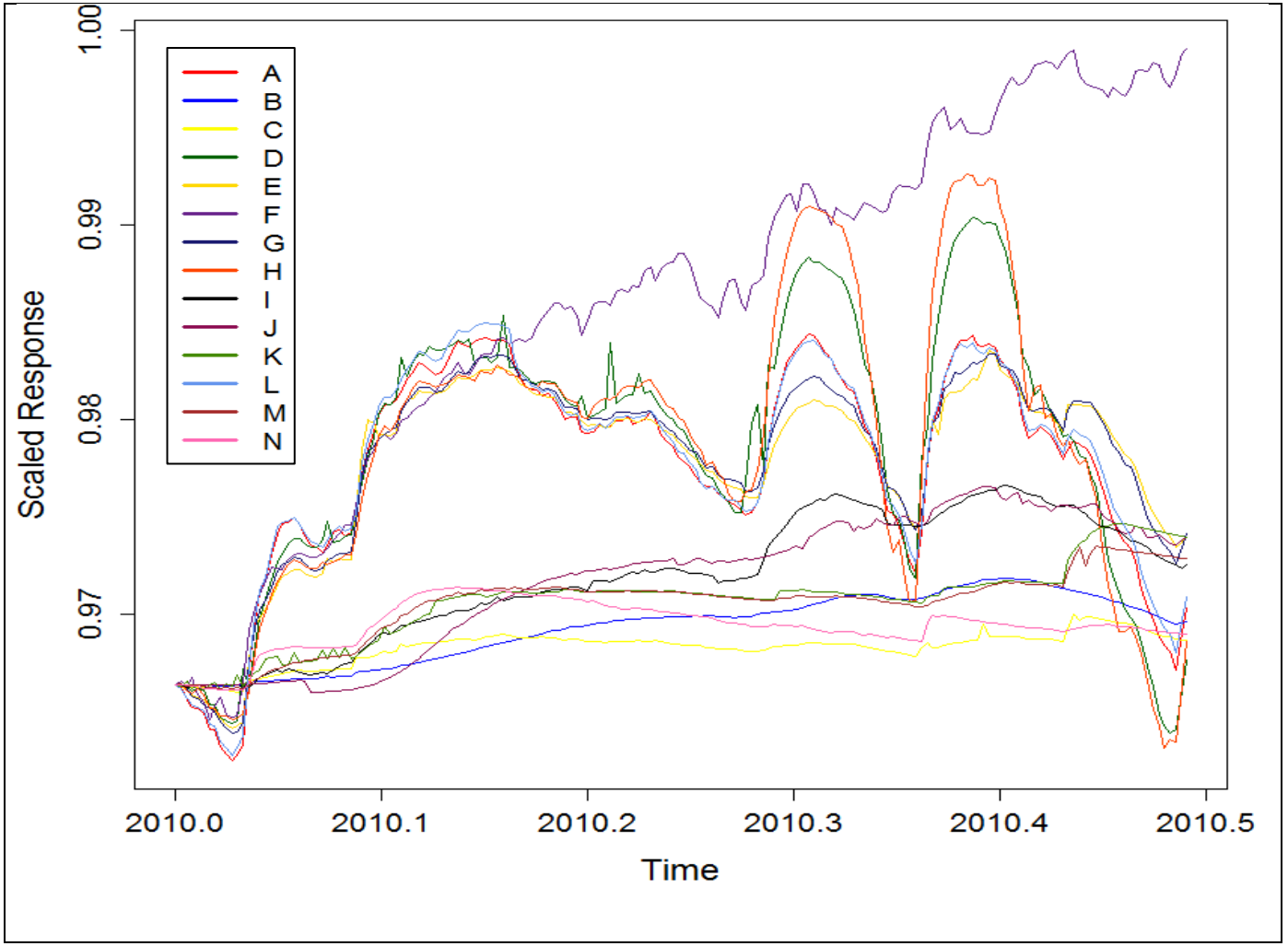

Figure 36. Time-series of the fourteen wells during time-period 13.

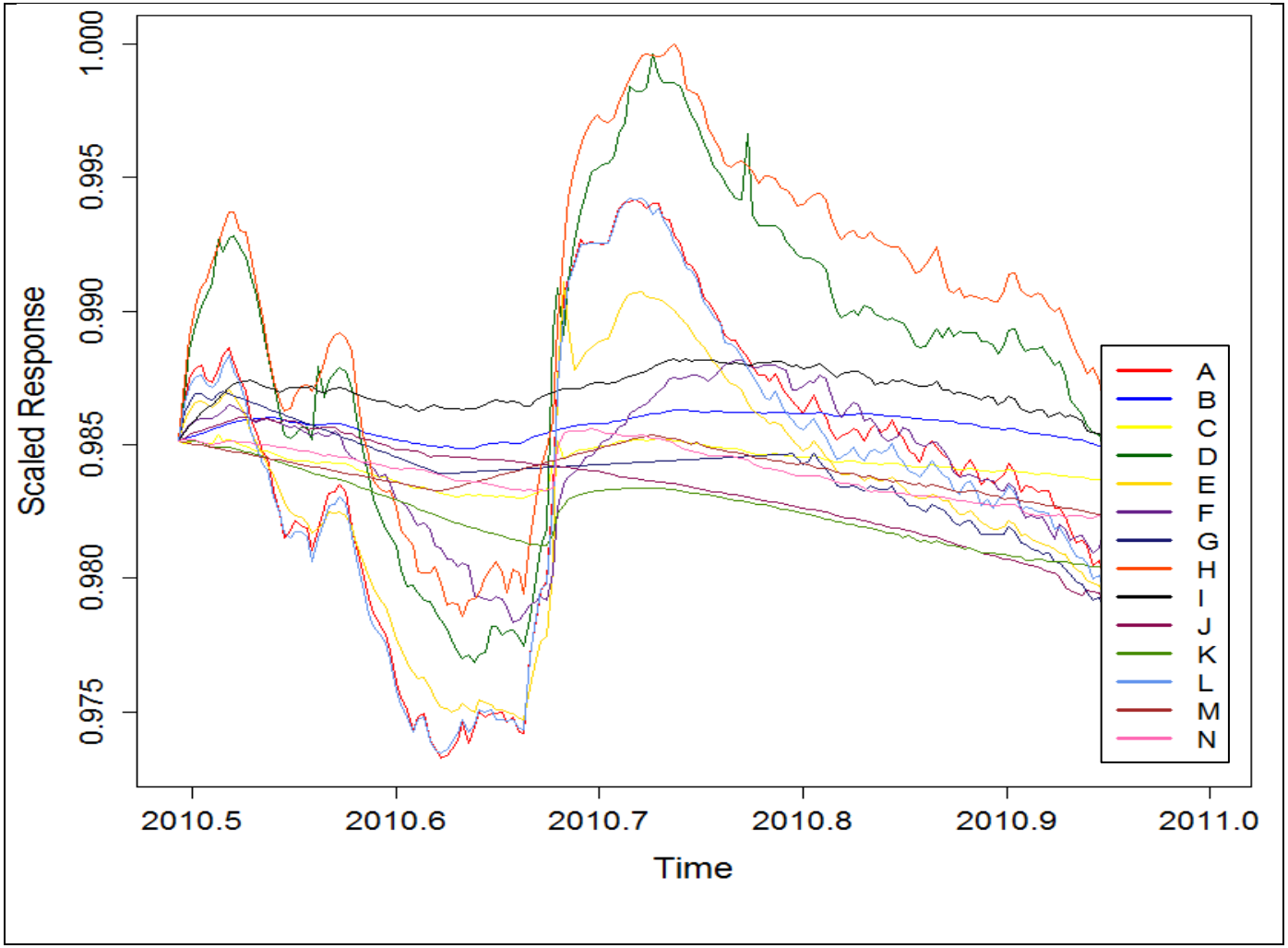

Figure 37. Time-series of the fourteen wells during time-period 14.

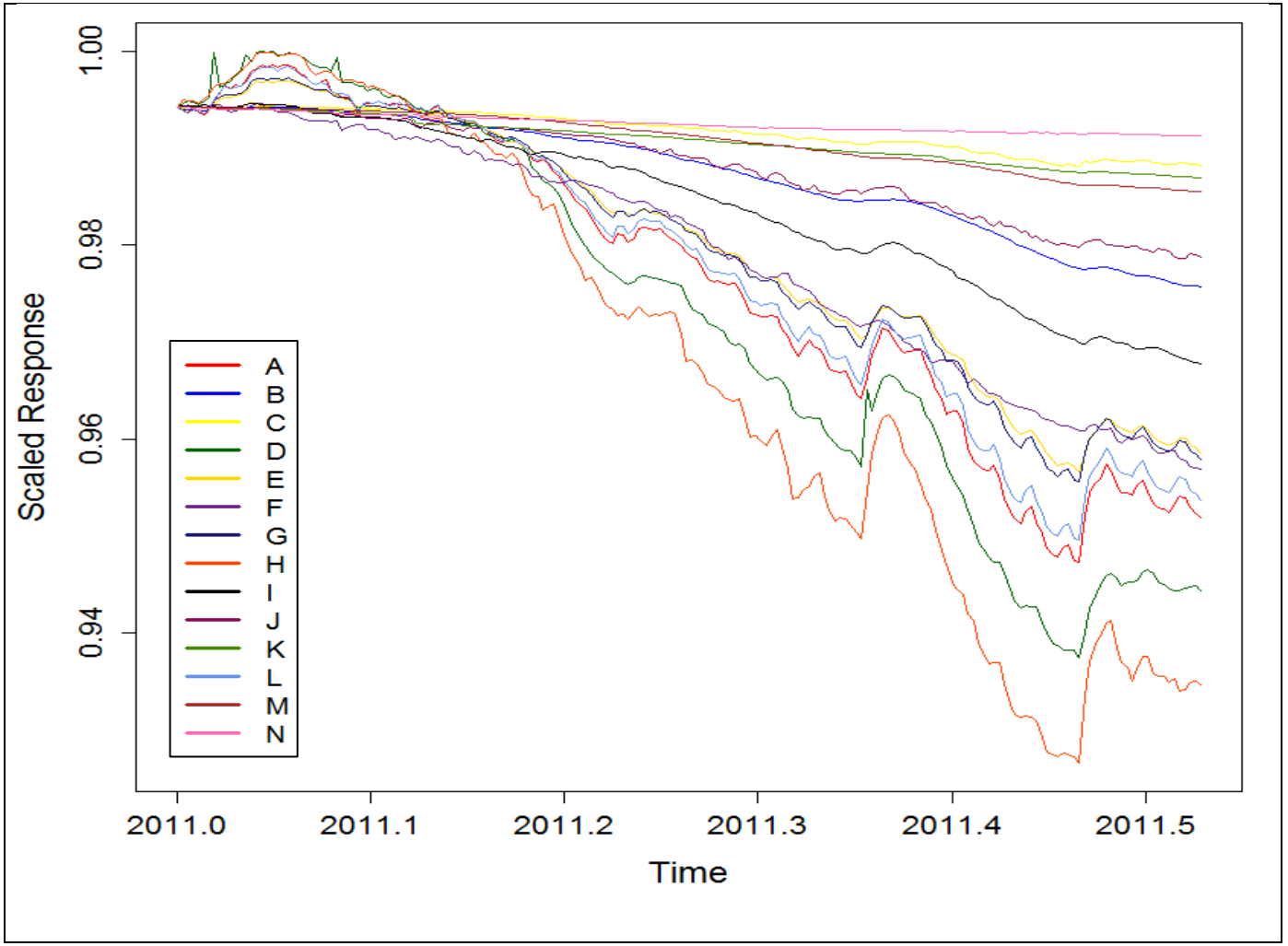

Figure 38. Time-series of the fourteen wells during time-period 15.

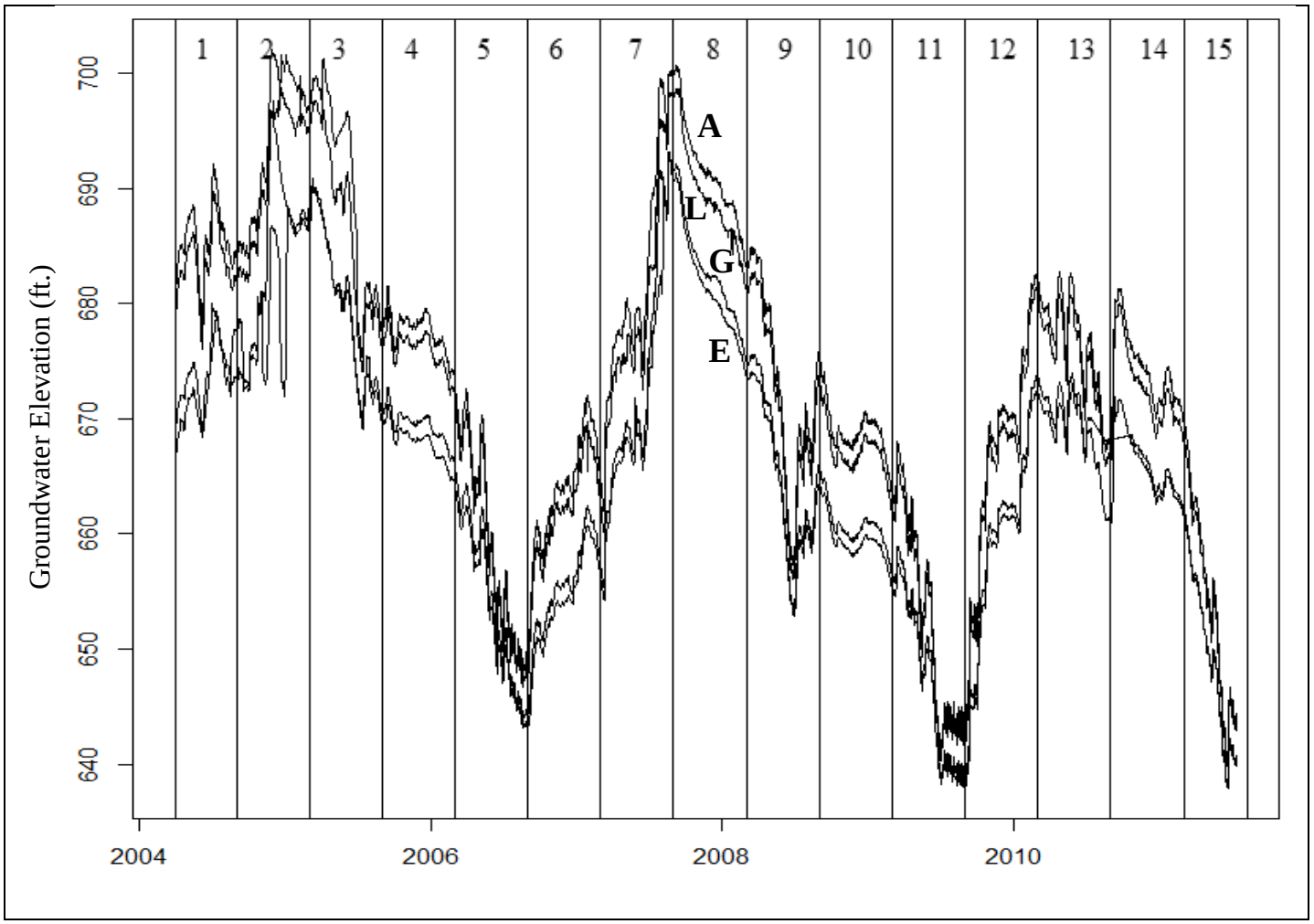

Figure 39. Cluster 1: Wells A, E, G, and L.

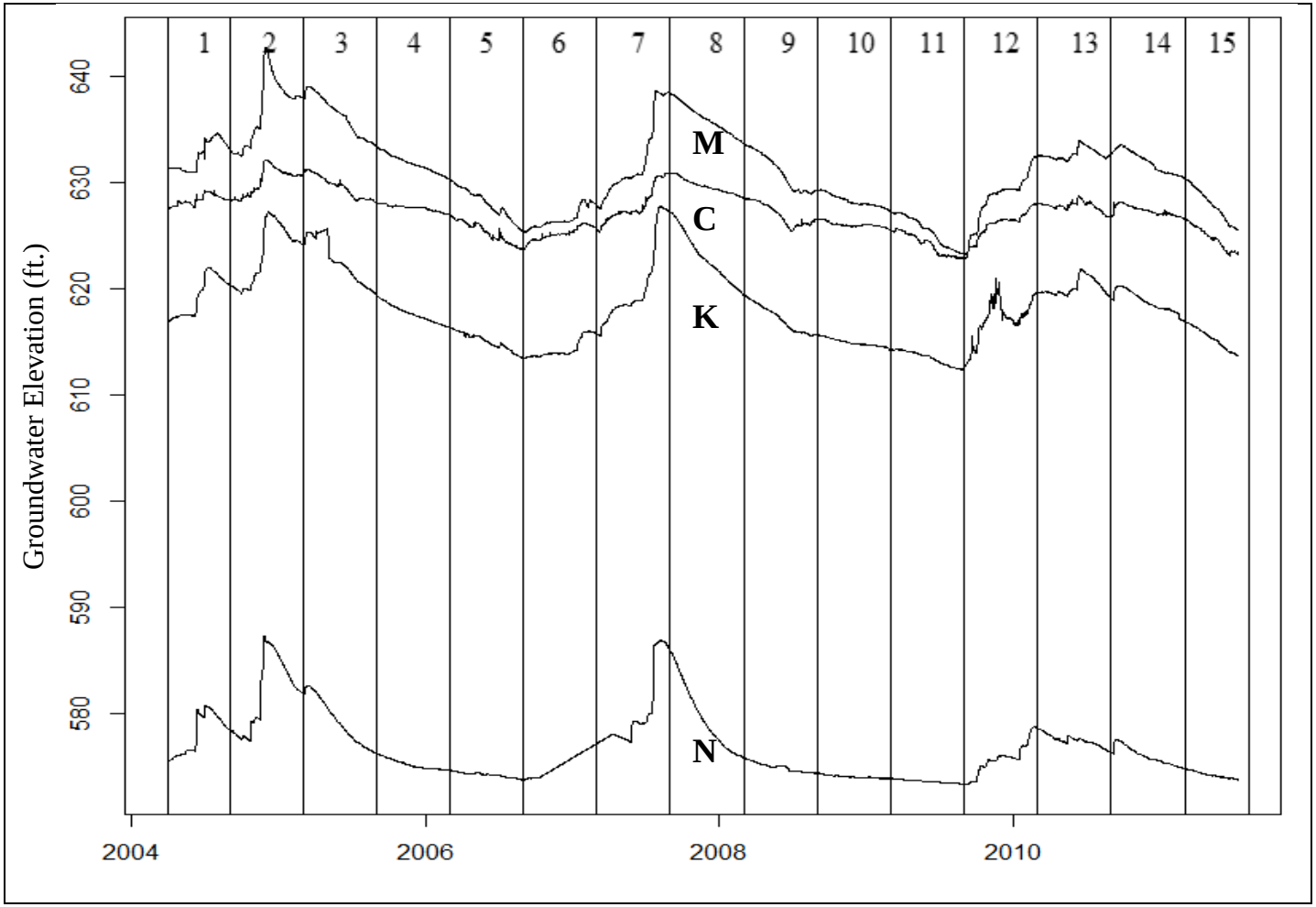

Figure 40. Cluster 2: Wells C, K, M, and N.

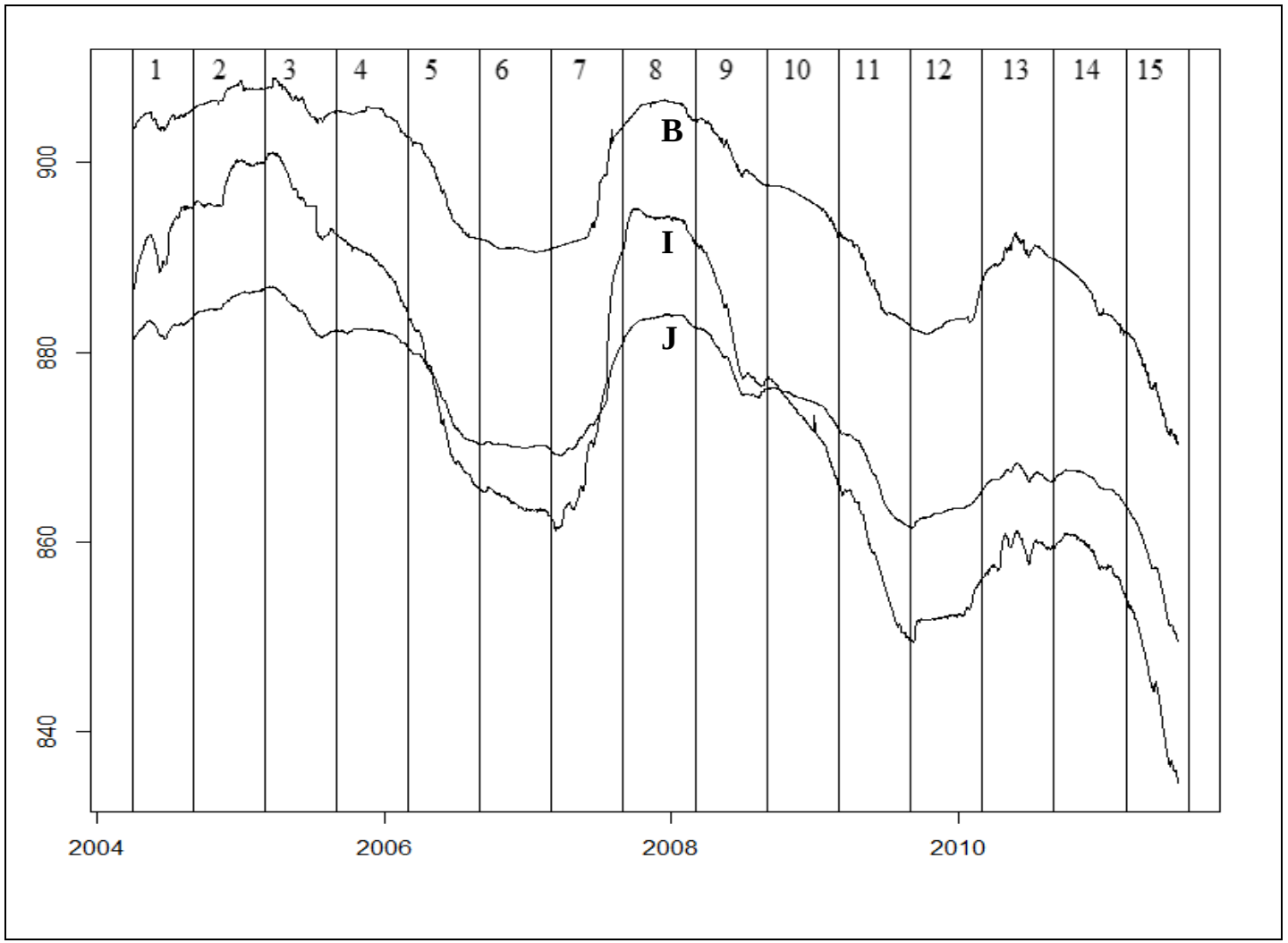

Figure 41. Cluster 3: Wells B, J, and I.

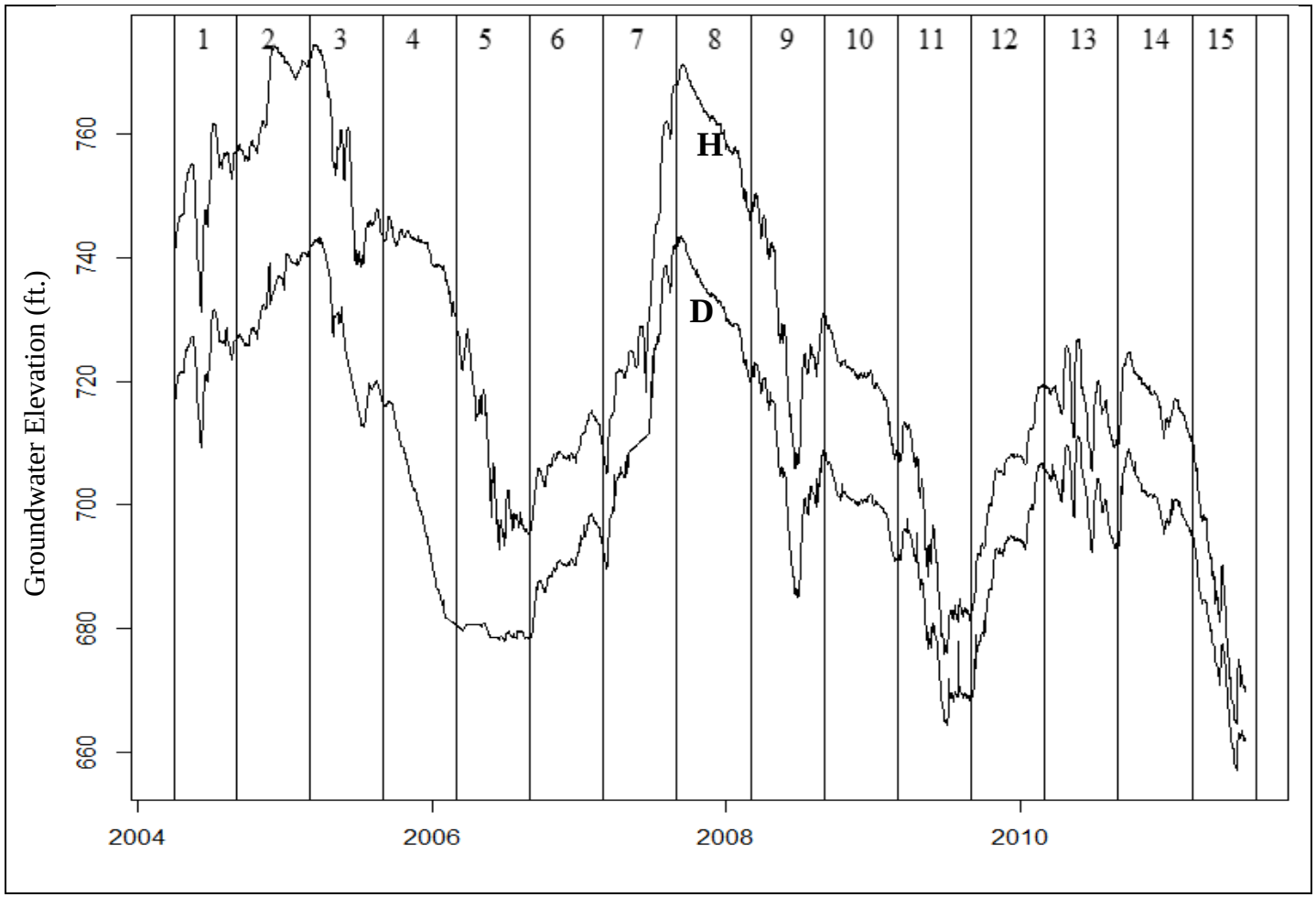

Figure 42. Cluster 4: Wells D and H.

ANN groundwater level prediction accuracy

During the hierarchical clustering analysis, four clusters of wells were identified for ANN training data. ANNs were constructed for each well in the study, except for well F because it failed to cluster with other wells for at least eleven of the fourteen time-periods. For each ANN constructed, the possible ANN training data included wells it clustered with, 30, 60 and 90-day moving averages of precipitation and spring flows for San Marcos and Comal Springs. Optimal ANN training data and the 1-year test period predictions for each well were calculated, as well as statistical assessments of the ANNs prediction accuracy (R^2 and RMSE); **Table 3.**

Cluster 1 ANNs

ANNs for Cluster 1 included wells A, E, G, and L. During the training process for well A, various combinations of the possible training data were used as input (**Figures 43-49)**. Initially, spring flow was included as training data, the ANN fit the data relatively poorly (ANN predictions crossed the observed and were misaligned) and had a low R^2 value (**Figure 43**). This may be because discharge from San Marcos and Comal Springs is influenced by larger-scale hydrogeologic relationships and environmental drivers across the aquifer, resulting in an inability to accurately predict Cluster 1 groundwater levels. The model inputs providing the best fit for well A were wells E, G, and L, and 30, 60, and 90-day moving averages of precipitation (**Figure 44**). During the testing period, the ANN groundwater level predictions fit the observed with a R^2 of 99.94% and a RMSE of 0.3985 feet. When moving averages of precipitation were removed and only wells E, G, and L were used as training data, model fit was high, R^2 = 99.82%, but the RMSE of 1.091 feet was consistently higher than when the moving averages of

precipitation were included as training data (**Figure 45**). This was considered an indication the ANN needed information about the recent past groundwater levels to calibrate to the predictions to initial groundwater conditions. Then, because well G was known to have data issues (missing and corrupted data during the training and testing time-periods), an ANN was trained without well G to determine if training improved. When trained without well G, ANN prediction accuracy (R^2 was 99.90% and RMSE was 1.054 feet) did not improve over the best model (**Figure 46**). While the prediction accuracy appears high during the first part of the testing period, prediction accuracy gradually decreased until the end of the testing period indicating that well G was highly similar to well A toward the end of the testing period. It was thought that more accurate predictions were made when the ANN was trained with wells E, G, and L because each of the three wells contribute important information about observed well A groundwater levels. And, although well G contains missing and corrupted data, the best ANN constructed for well A (**Figure 44**) included well G, highlighting the flexibility of ANNs prediction capabilities, and the fact that even though an input may not appear to be high quality, it may still contain valuable information.

ANNs were also constructed for wells E and L. During the test period the ANN predicted groundwater levels with high accuracy, with a R^2 of 99.75% and RMSE of 0.6849 feet for well E (**Figure 47**) and R^2 of 99.75% and RMSE of 0.5715 feet for well L (**Figure 49**), which was expected since Cluster 1 wells clustered together tightly (were highly similar) across all hydrologic conditions during the hierarchical clustering analysis. During the test period for well G, data were missing and estimated using a linearly interpolated to fill the data gaps. Therefore, the true accuracy of the ANNs

predictions could not be measured because of the inaccuracy of the interpolated data. However, other than the period with missing data, it can be visually seen in **Figure 48** that ANN predictions are very close to observed groundwater levels. The predictions for well G were included in the results of this study to illustrate how accurate predictions for missing groundwater levels using ANNs can be made. The predictions made for the missing data during the test period can be inserted into the original dataset in place of the linear interpolated data, resulting in a much higher quality dataset.

Cluster 2 ANNs

Cluster 2 consisted of the wells C, K, M, and N. When ANNs were constructed for the wells in Cluster 2, the model fit and RMSE were high during the testing period for each of the wells (**Figures 50-53**). However, closer inspection of Cluster 2 ANN test predictions compared to observed groundwater levels in the test period show that, although R^2 and RMSE seem to indicate highly accurate predictions, the predictions do not smoothly follow the observed. Cluster 2 wells are in the unconfined zone of the Edwards Aquifer and groundwater levels range only 20 feet across all four wells. Because of the small range of groundwater levels, R^2 is high and RMSE is very low for each of the wells' test predictions even though the predicted and observed did not align well. The misalignment is thought to be a result of low quality training and testing data. The wells were ranked alphabetically based on the amount of missing data and Cluster 2, wells K, M, and N contain the highest amount of missing data of the four clusters. As seen in **Figure 51**, well K appears to have corrupted data during the middle portion of the test period which then caused low prediction accuracies for wells C, M, and M (**Figures 50, 52-53**).

Cluster 3 ANNs

ANNs were constructed for Cluster 3, wells B, I, and J, located in the confined artesian zone and in an area where the hydrogeology is known to be relatively complex. During certain hydrologic conditions, well I is connected to the neighboring Frio watershed and well I's degree of connectivity to other wells in the aquifer is controlled by the Knippa Gap, a groundwater divide that disconnects well I from wells B and J when groundwater levels are lower than the divide. During the hierarchical clustering analysis, well B and J clustered in most of the fourteen-time-periods. However, well I failed to cluster with well B and J during time-periods 4, 5, and 8, time-periods when hydrologic conditions were transitioning from wet to dry (**Figures 27-28, 31**). The effect of the Knippa Gap on groundwater levels can be seen in the time-series of well B, I, and J (**Figure 41**). During times when groundwater levels are relatively high, wells B, I, and J respond similarly. But, when transitioning between wetter to dryer times, when groundwater levels are lower, well I's response is more pronounced than well B and J because well I is under the influence of the Knippa Gap and exchanges flows with the Frio watershed. Because of this, it was at first unclear whether to include well I in Cluster 3 ANN trainings data. To answer this, ANNs were constructed for well B and J with well I included as training data, and compared to the ANNs trained without well I. The most accurate predictions for wells B and J were made when ANN training included well I. Using the input training data of well J and I, and a 90-day of moving average of precipitation, well B ANN test prediction accuracy was R^2 of 90.05% and RMSE of 2.156 feet (**Figure 54**). ANN test prediction accuracy for well J, trained on well B and I, and a 90-day moving average of precipitation, was R^2 of 91.58% and RMSE of 2.125 feet for well J (**Figure 56**). It was thought that including well I in the training process for both

wells B and J contributed important information to the ANNs that was not available using just well B or just well J as training data. This was confirmed when training the ANN to predict well I groundwater levels (**Figure 55**). When both well B and J were used as training data, along with a 90-day moving average of precipitation, ANN prediction accuracy was even better for well I than well B and J, with R^2 of 99.65% and RMSE of 1.507 feet. For each of the wells in Cluster 3, ANN prediction accuracy could improve in future work, if additional wells are identified as possible training data.

Cluster 4 ANNs

Lastly, ANNs were constructed for Cluster 4 which contained wells D and H. The ANN test predictions for both well D and H were poor because of poor data quality. Although during the test period the model fit was high, well D had a R^2 of 98.53% and well H had a R^2 of 99.98%, the RMSE was considerably high, 9.291 feet and 5.293 feet respectively (**Figures 57-58**). Because predicted groundwater levels for both well H and well D were displaced by approximately the same value across the test period, similar to the test prediction displacement seen for well A in Cluster 1 without 30, 60 and 90-day moving averages of precipitation as training data, it is thought that yet-to-be-identified physical parameters could improve ANN prediction accuracy for wells D and H.

Table 3. Statistical measurements of the ANNs prediction accuracy.

Cluster	Well	Inputs	R^2	RMSE (ft.)
1	A	Wells E, G, L, & 30, 60, 90-days moving averages of precipitation	99.94%	0.3985
	E	Wells A, G, L, & 30, 60, 90-days moving averages of precipitation	99.75%	0.6849
	G	Wells A, E, L, & 30, 60, 90-days moving averages of precipitation	95.20%	0.2139
	L	Wells A, E, G, & 30, 60, 90-days moving averages of precipitation	99.75%	0.5715
2	C	Wells K, M, N, & 90-days moving averages of precipitation	99.54%	0.3229
	K	Wells M, N, & 90-days moving averages of precipitation	98.78%	0.4303
	M	Wells C, K, N, & 90-days moving averages of precipitation	98.46%	0.3615
	N	Wells C, K, M, & 90-days moving averages of precipitation	94.51%	0.5969
3	B	Wells I, J, & 90-days moving averages of precipitation	99.05%	2.156
	I	Wells B, J, & 60, 90-days moving averages of precipitation	99.65%	1.507
	J	Wells B, I, & 90-days moving averages of precipitation	91.58%	2.125
4	D	Wells H & 90-days moving averages of precipitation	99.98%	5.293
	H	Wells D & 90-days moving averages of precipitation	98.53%	9.291

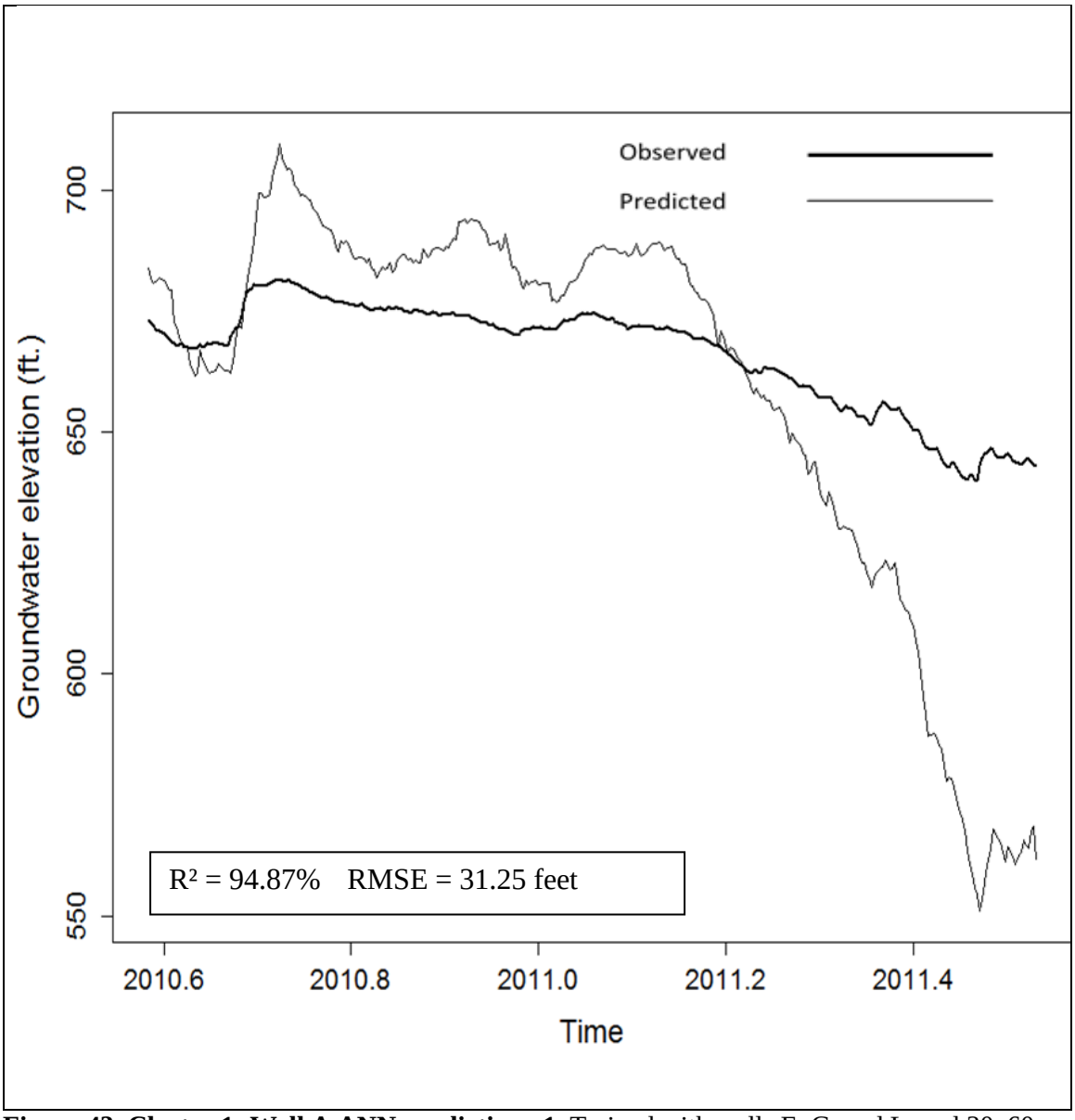

Figure 43. Cluster 1: Well A ANN predictions-1. Trained with wells E, G, and L, and 30, 60, and 90-day moving averages of precipitation, and spring flow from San Marcos and Comal Springs as input.

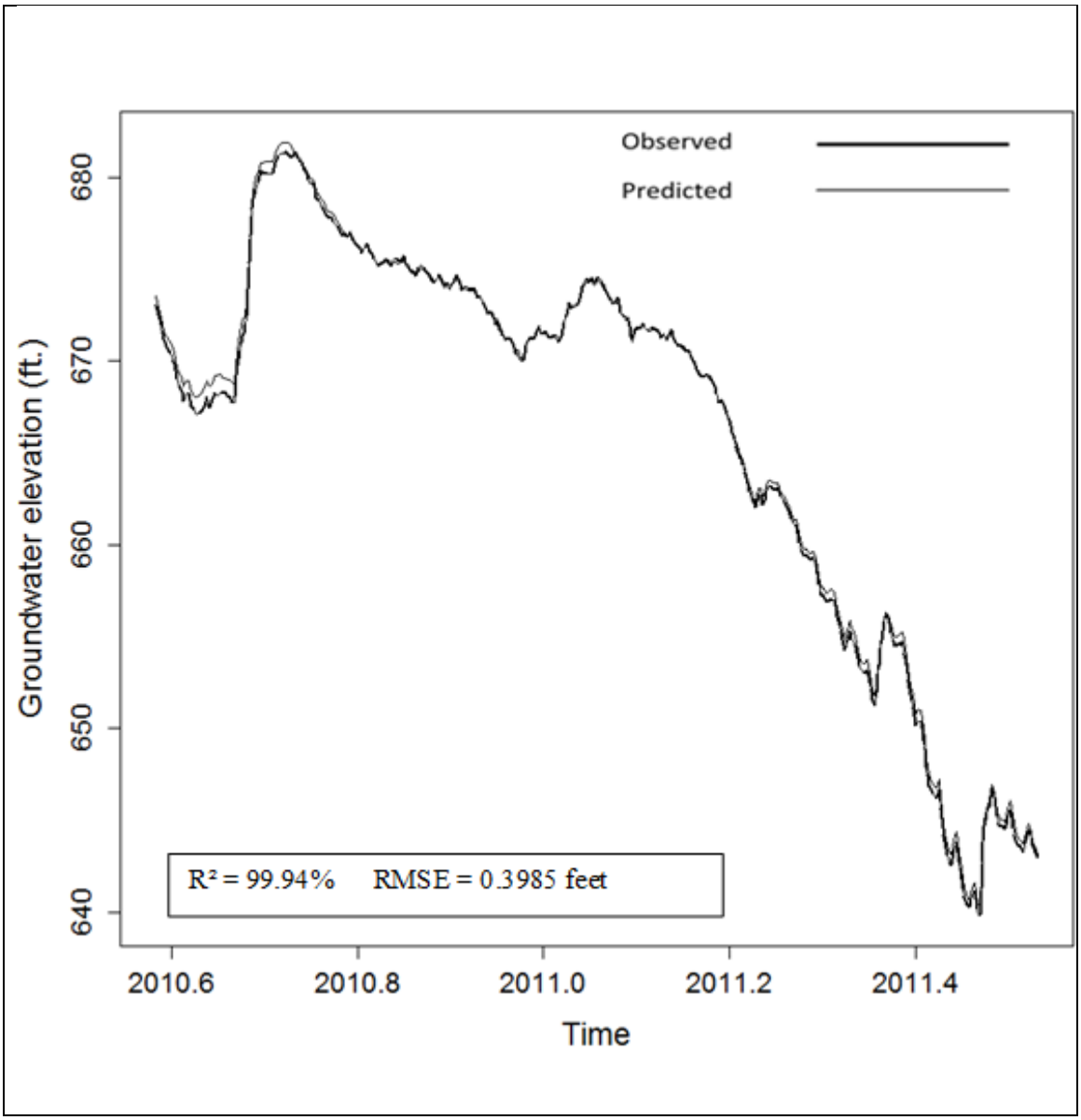

Figure 44. Cluster 1: Well A ANN predictions-2. Trained with wells E, G, and L, and 30, 60, and 90-day moving averages of precipitation as input.

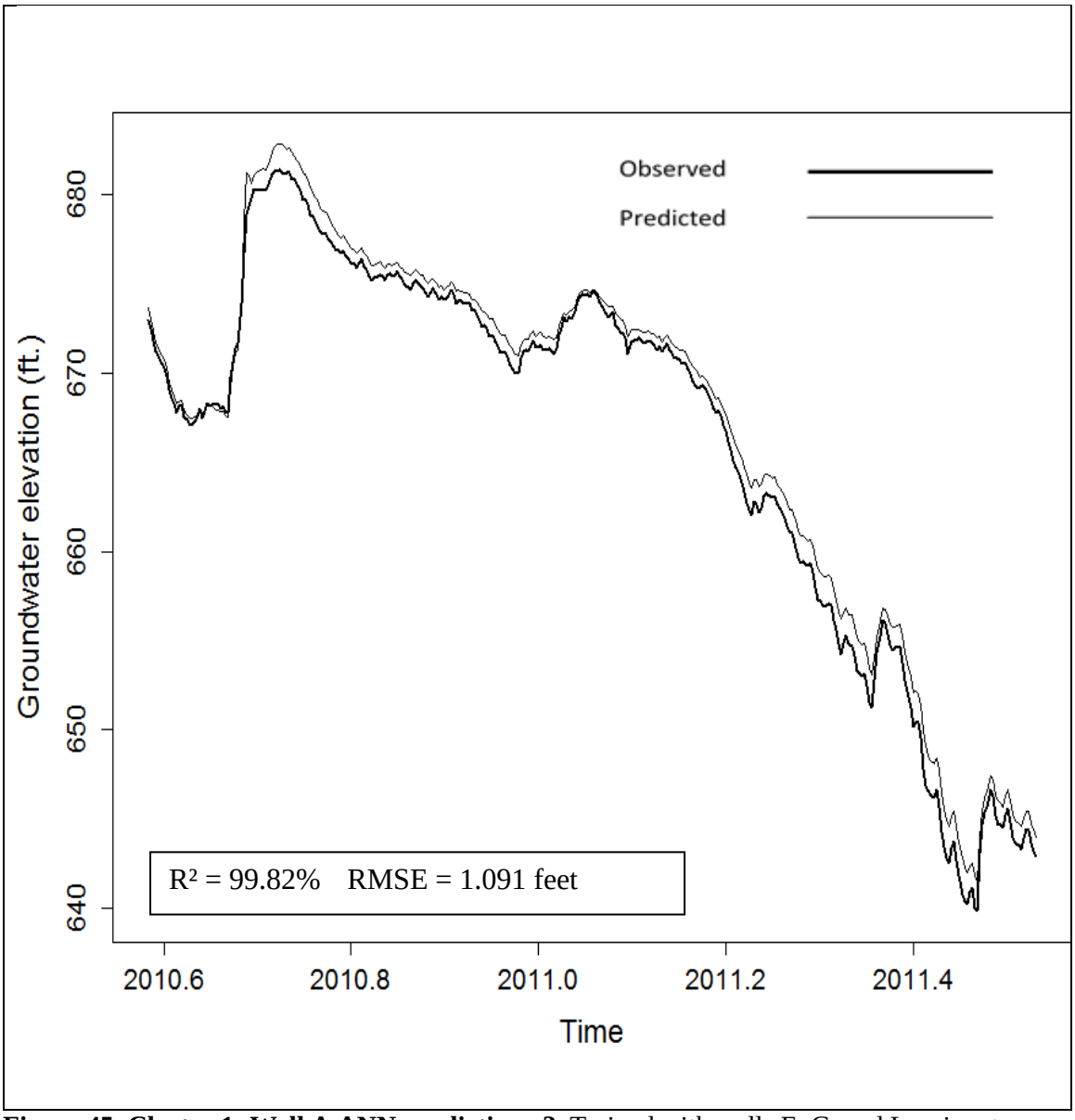

Figure 45. Cluster 1: Well A ANN predictions-3. Trained with wells E, G, and L as input.

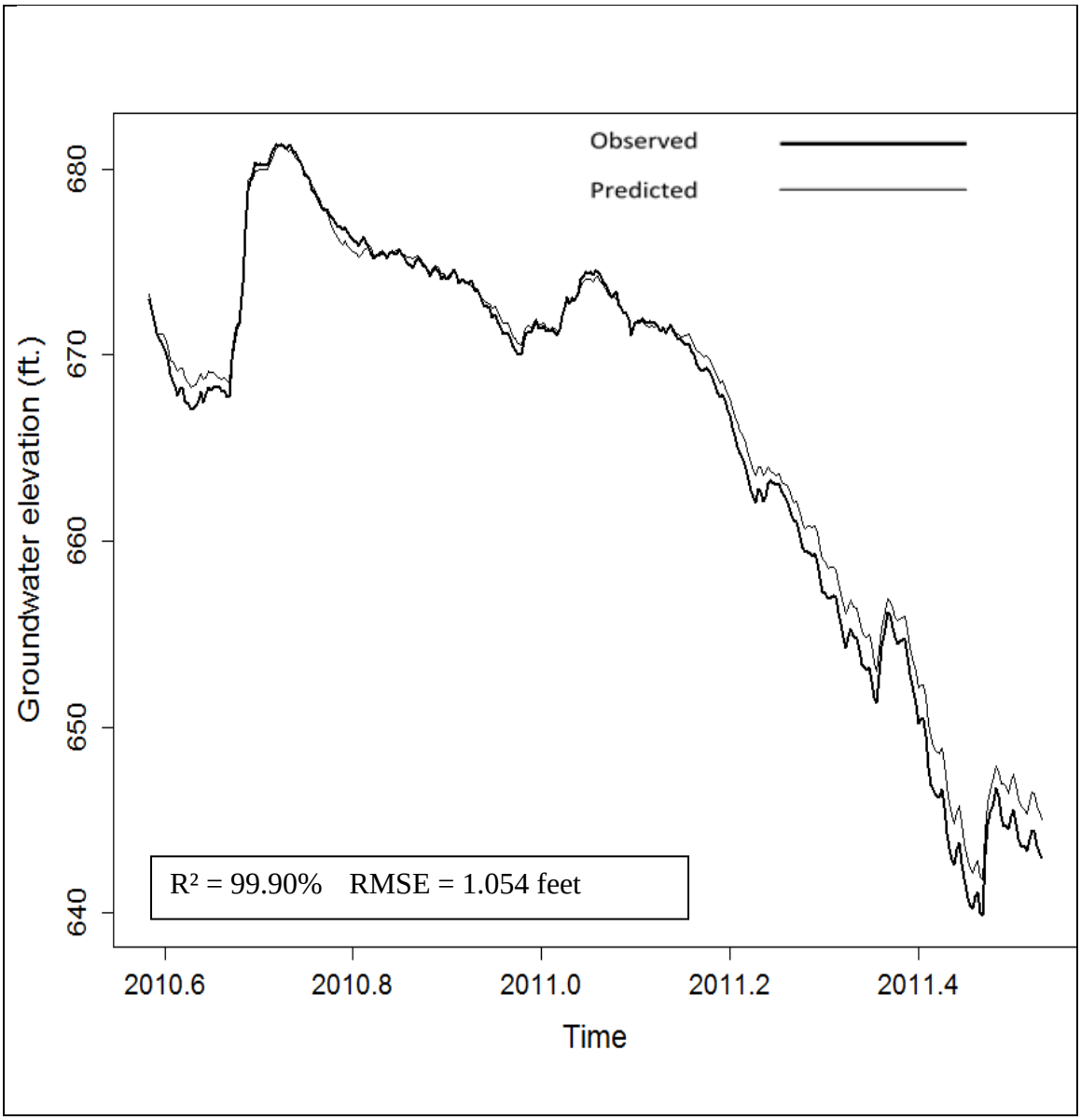

Figure 46. Cluster 1: Well A ANN predictions-4. Trained with wells E and L, and 30, 60, and 90-day moving averages of precipitation as input.

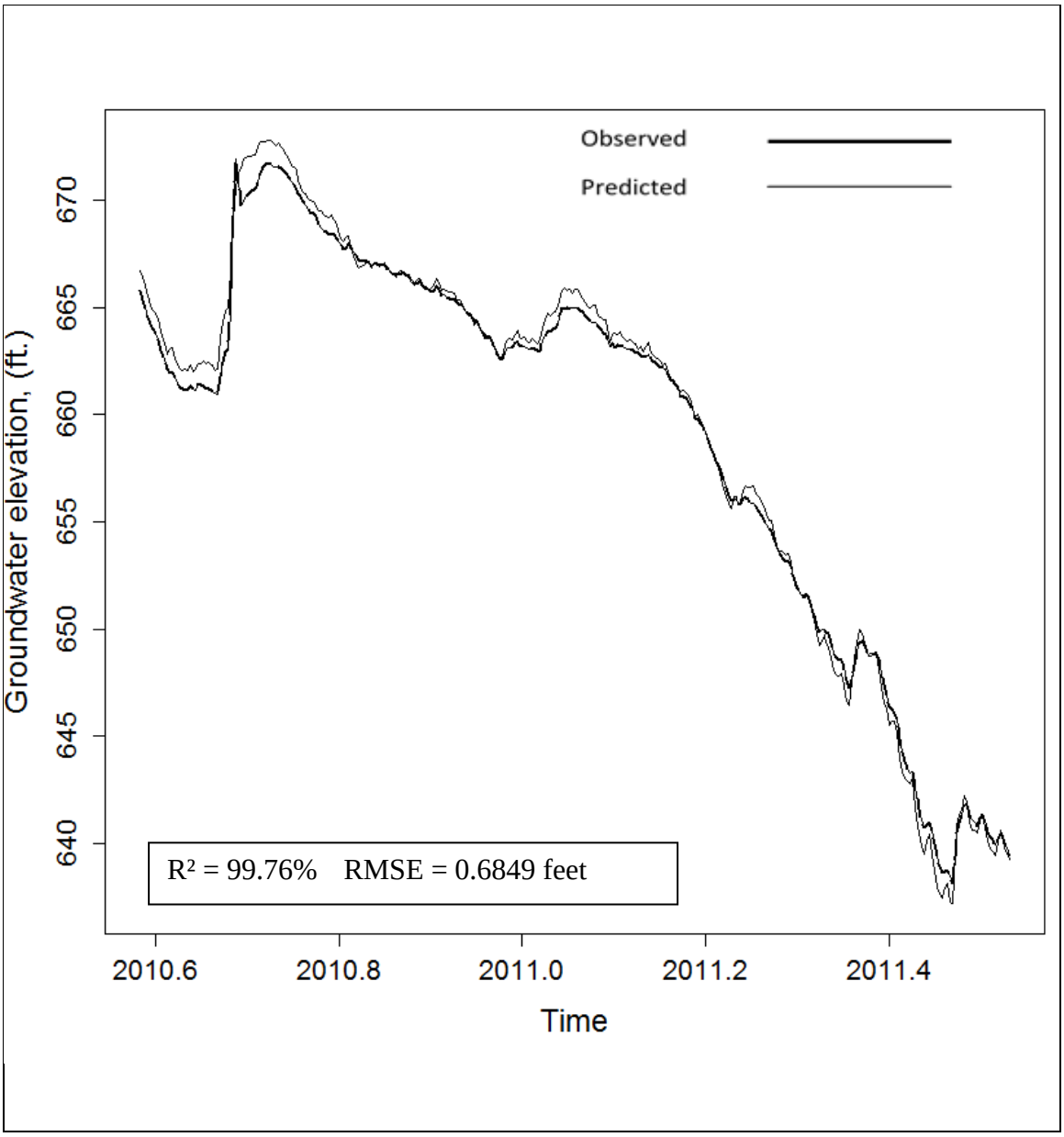

Figure 47. Cluster 1: Well E ANN predictions. Trained with wells A, G, and L, and 30, 60, and 90-day moving averages of precipitation as input.

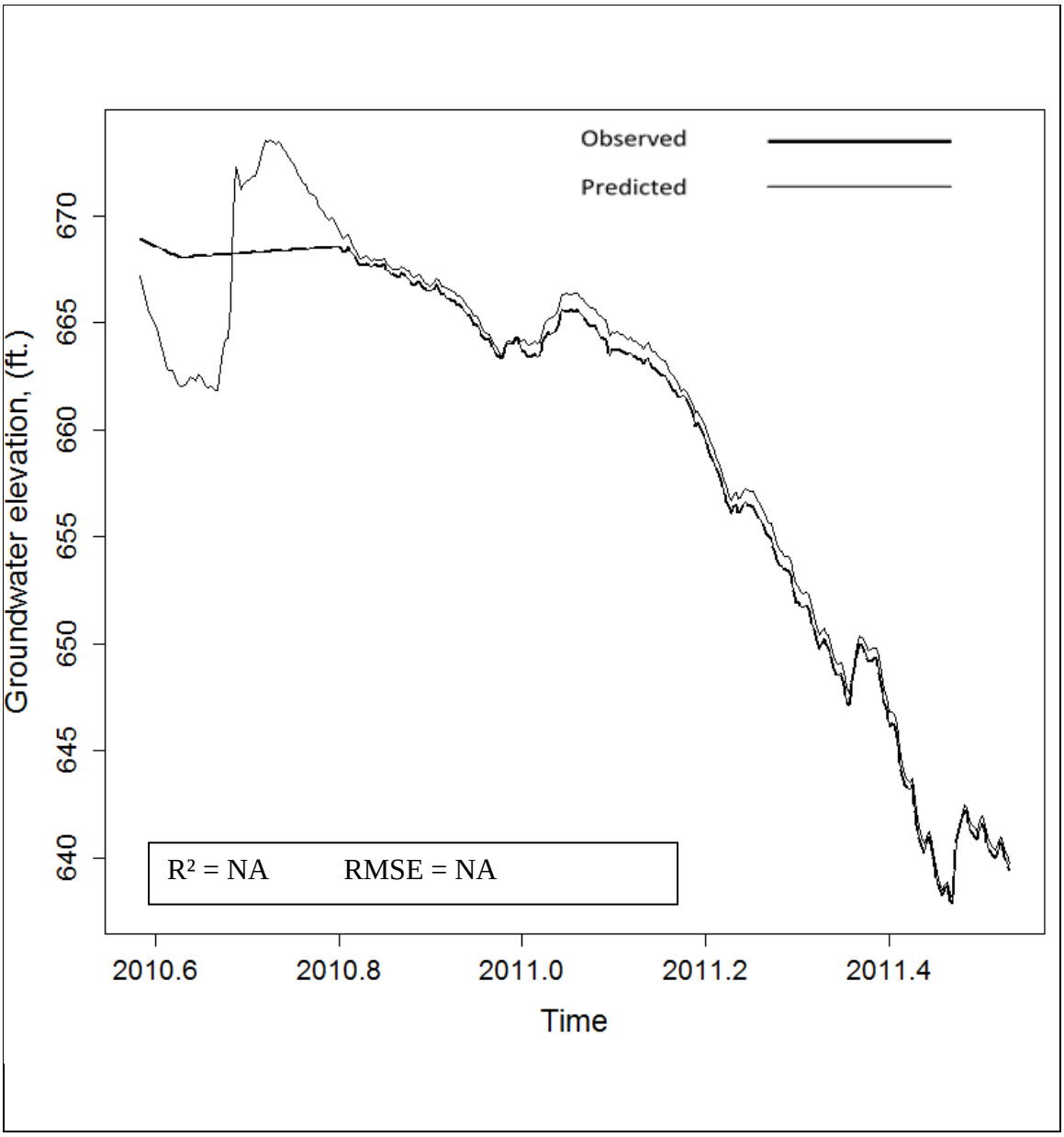

Figure 48. Cluster 1: Well G ANN predictions. Trained with wells A, E, and L, and 30, 60, and 90-day moving averages of precipitation as input.

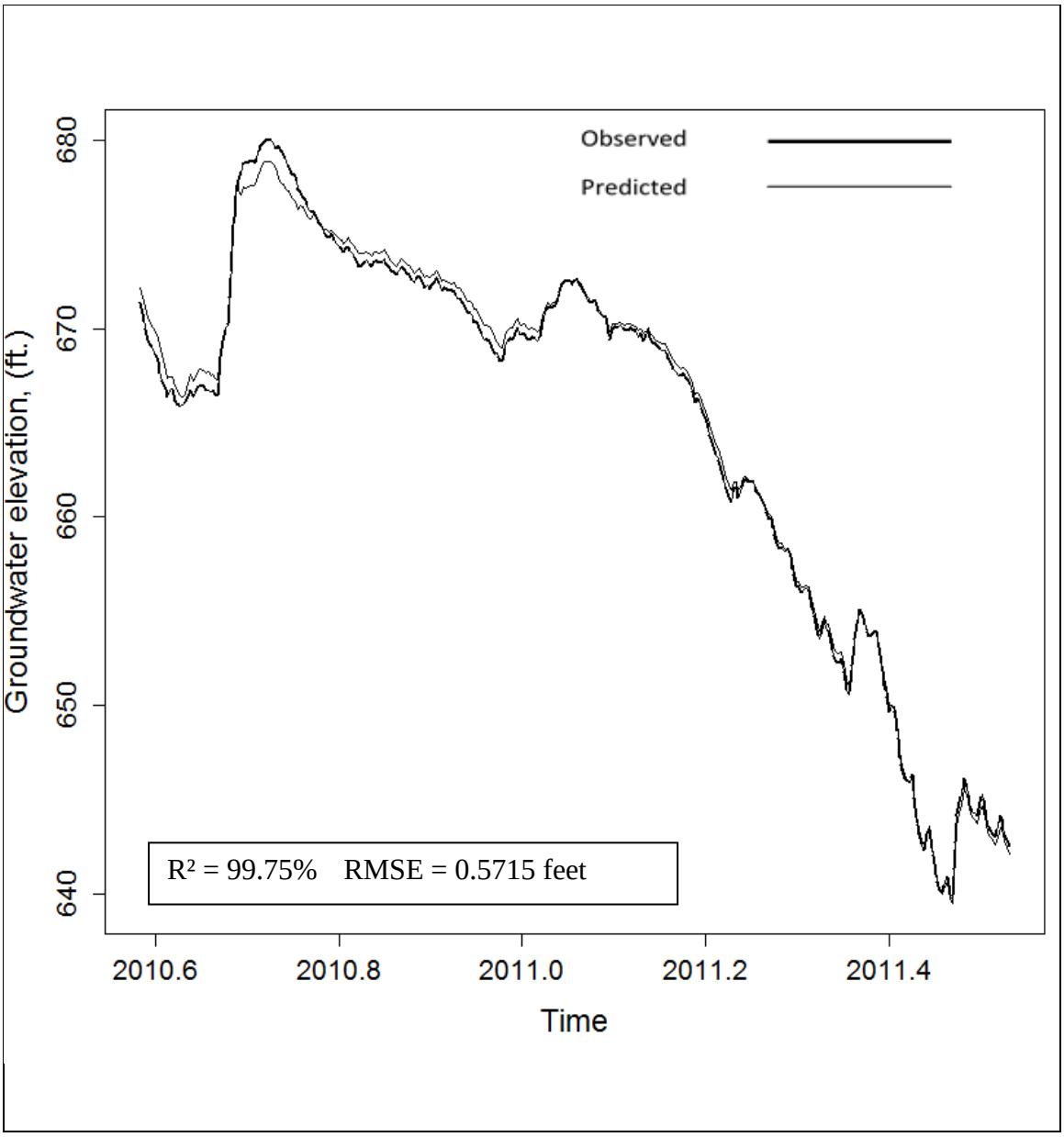

Figure 49. Cluster 1: Well L ANN predictions. Trained with wells E, G, and L, and 30, 60, and 90-day moving averages of precipitation as input.

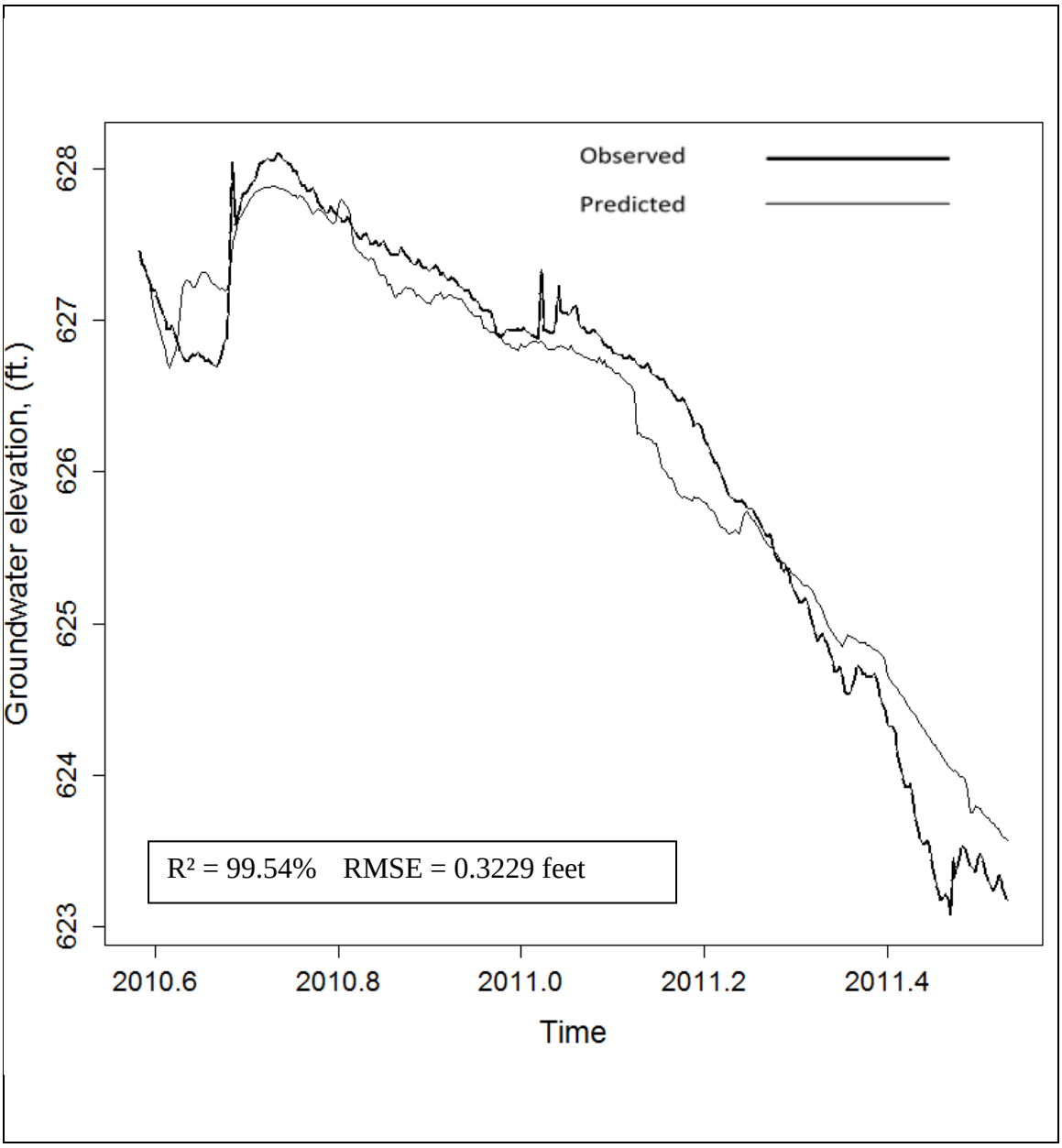

Figure 50. Cluster 2: Well C ANN predictions. Trained with wells K, M, and N, and a 90-day moving average of precipitation as input.

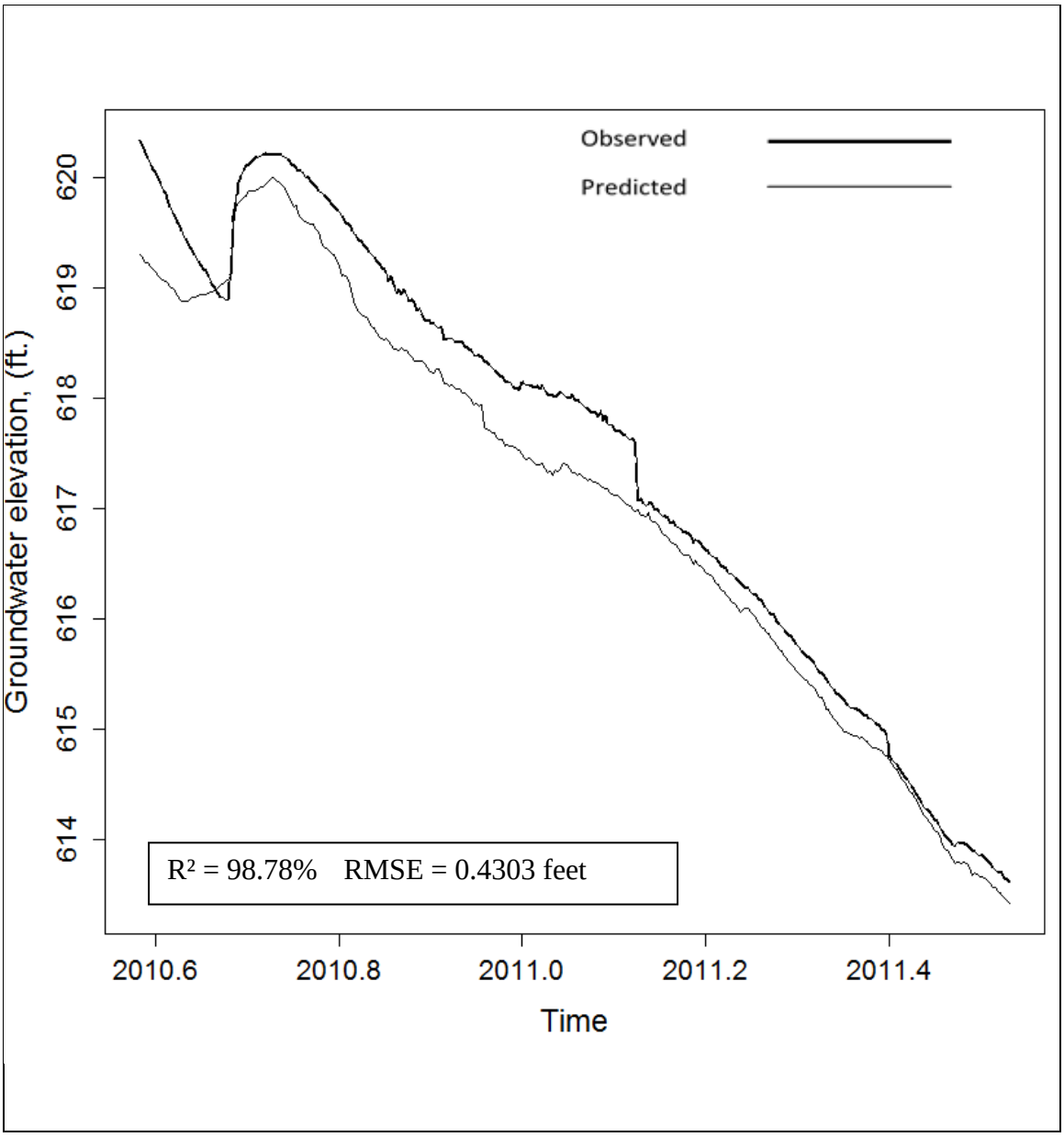

Figure 51. Cluster 2: Well K ANN predictions. Trained with wells M and N, and a 90-day moving average of precipitation as input.

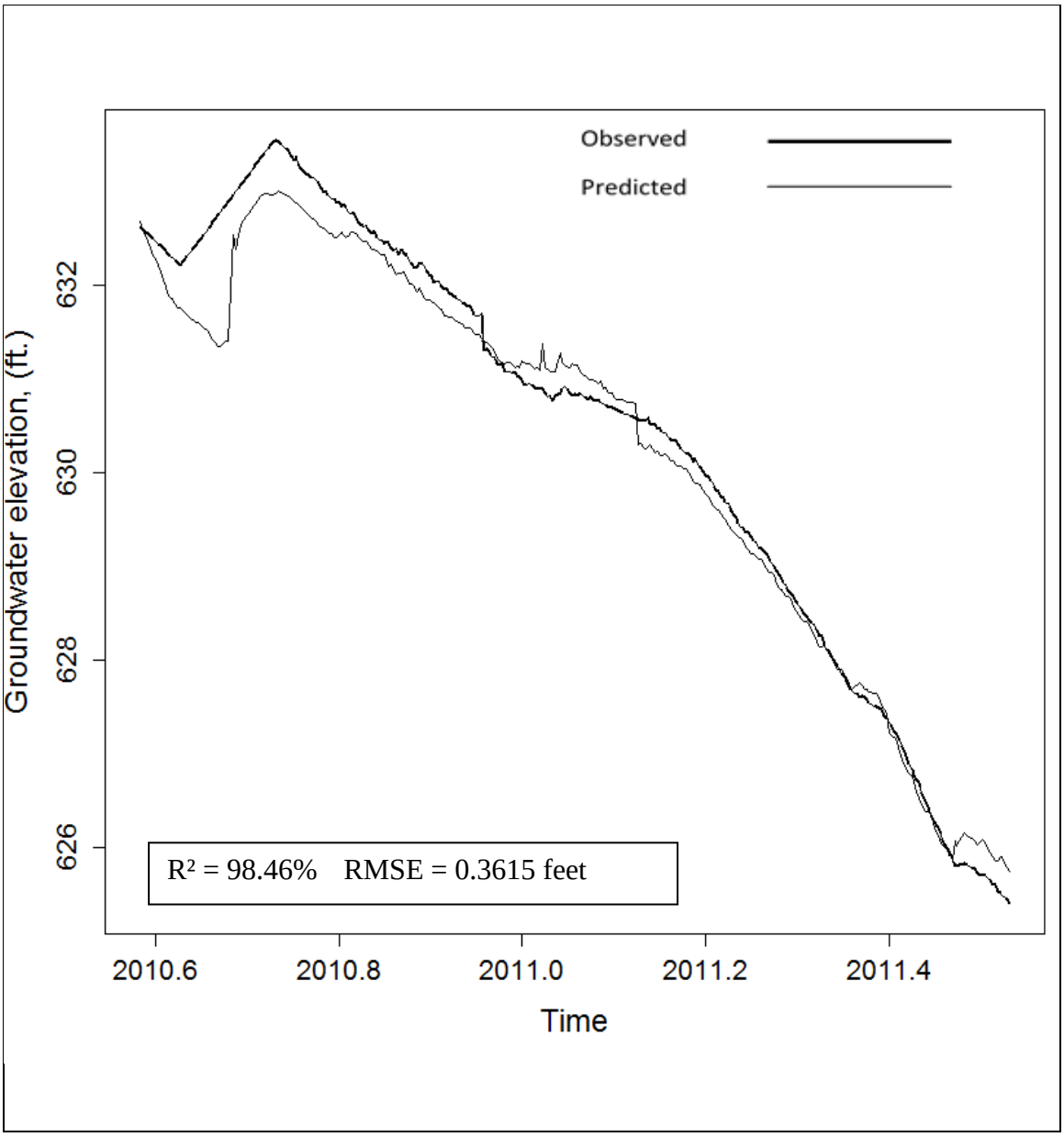

Figure 52. Cluster 2: Well M ANN predictions. Trained with wells C, K, and N, and a 90-day moving average of precipitation as input.

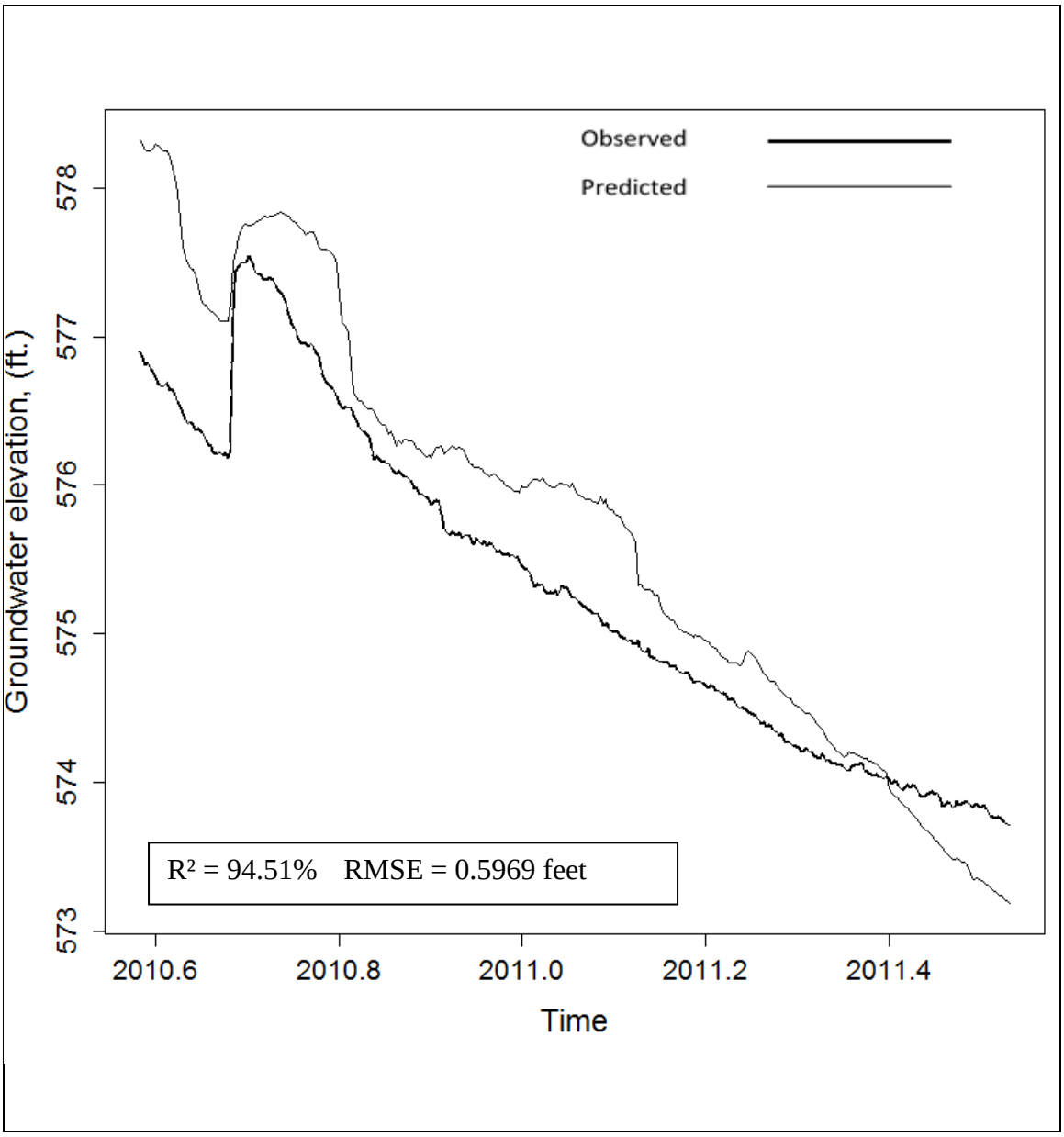

Figure 53. Cluster 2: Well N ANN predictions. Trained with wells M, K, and N, and 90-day moving average of precipitation as input.

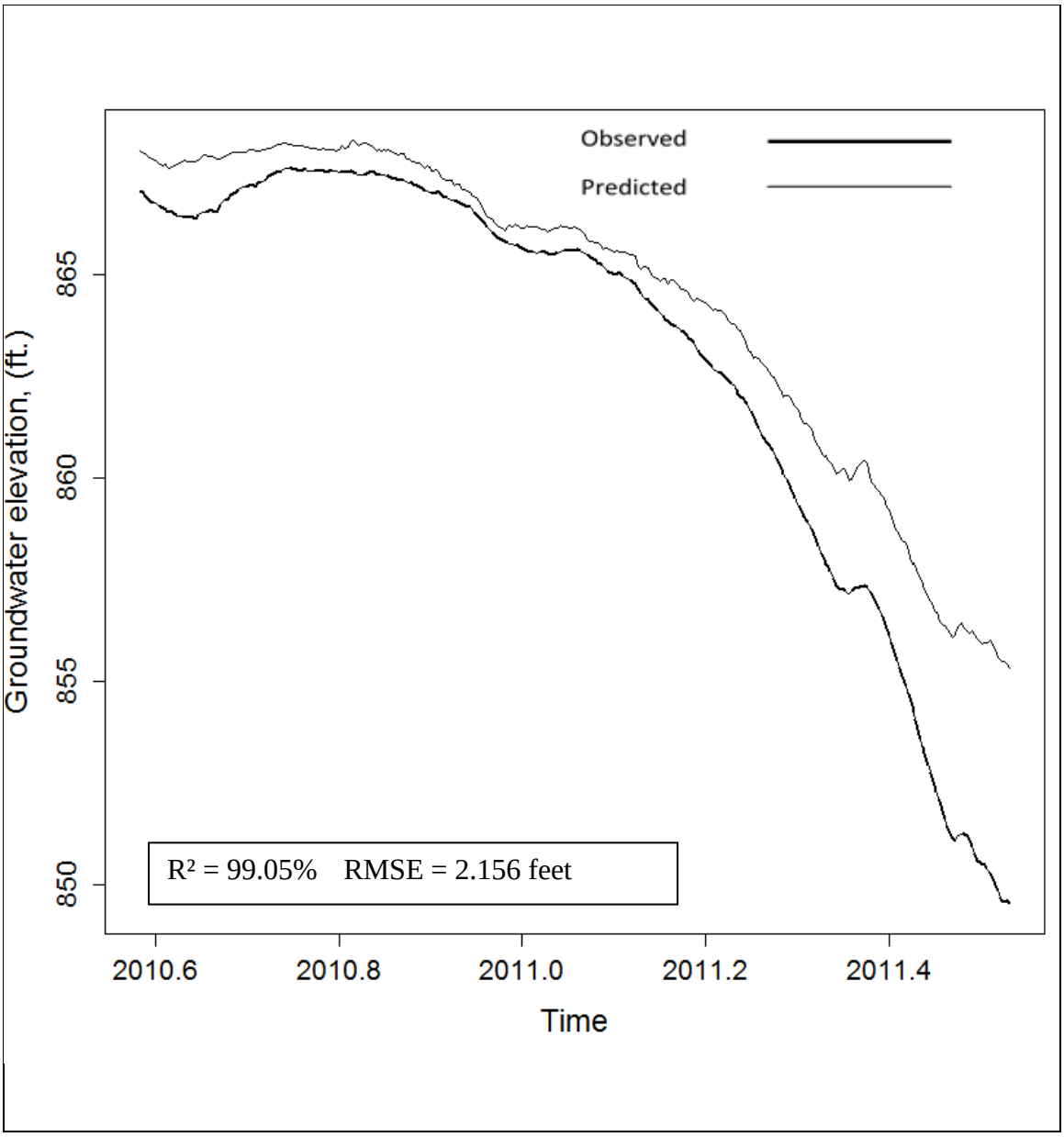

Figure 54. Cluster 3: Well B ANN predictions. Trained with wells I and J, and a 90-day moving average of precipitation as input.

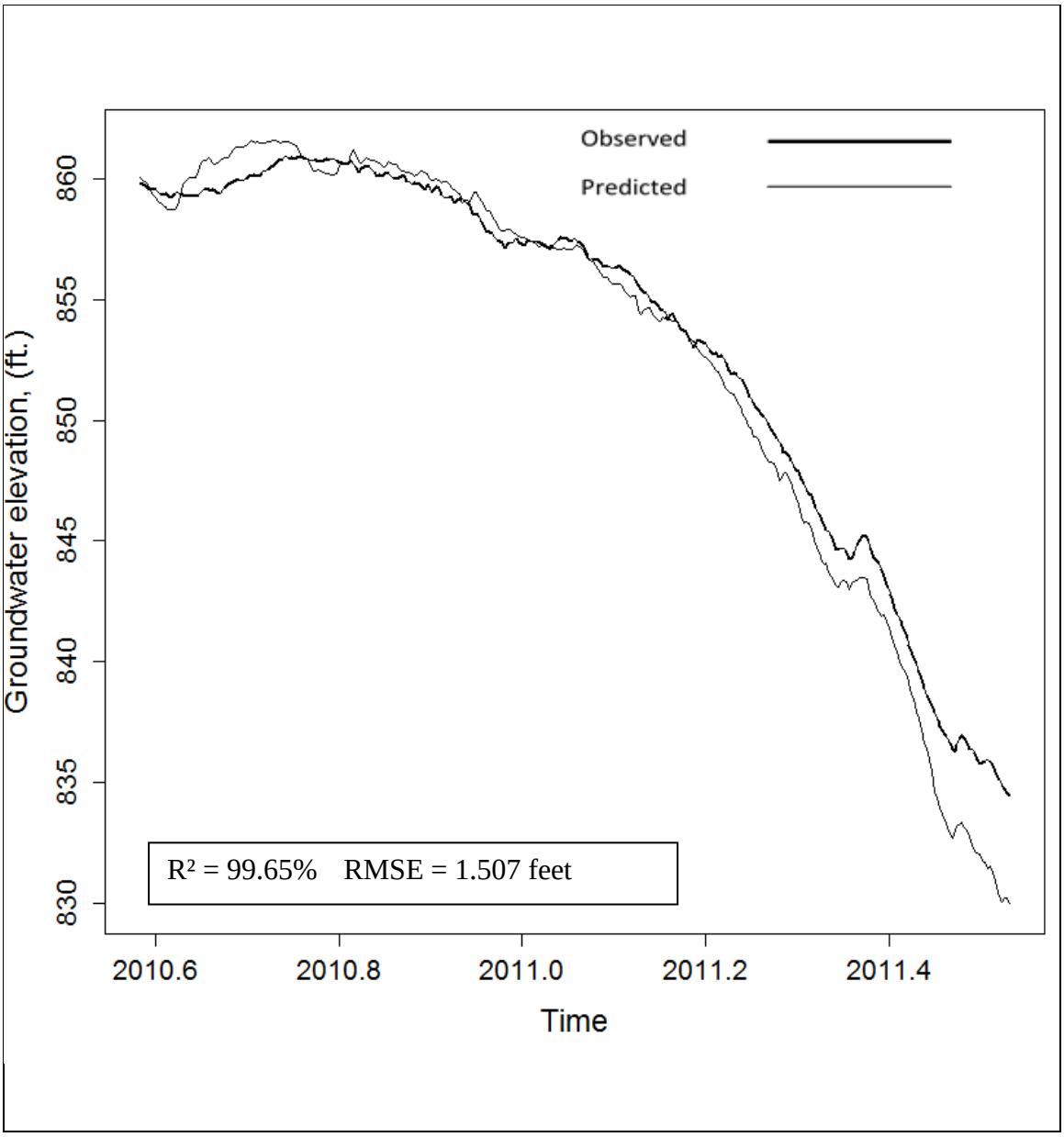

Figure 55. Cluster 3: Well I ANN predictions. Trained with wells B and J, and a 90-day moving average of precipitation as input.

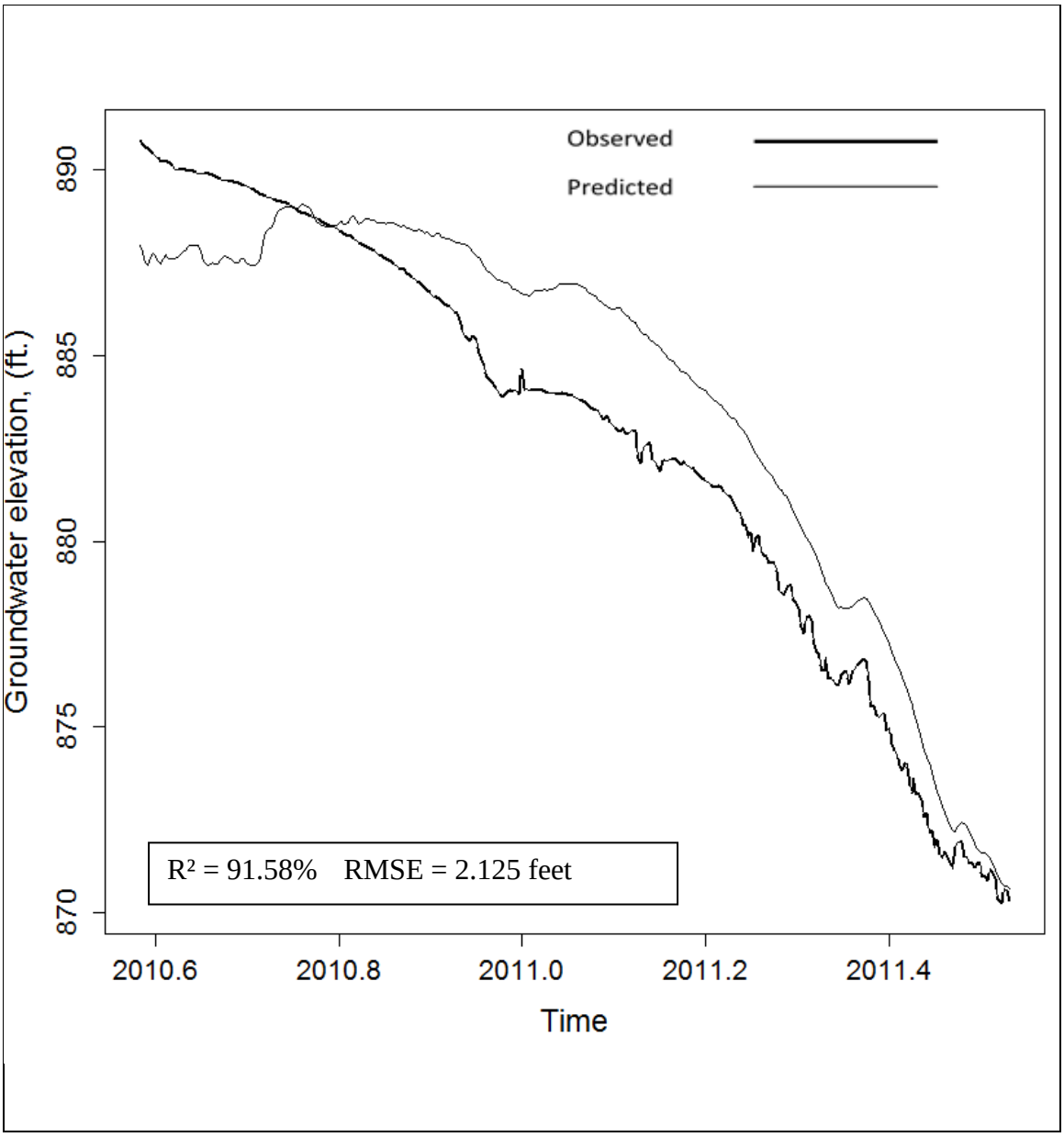

Figure 56. Cluster 3: Well J ANN predictions. Trained with wells B and I, and a 90-day moving average of precipitation as input.

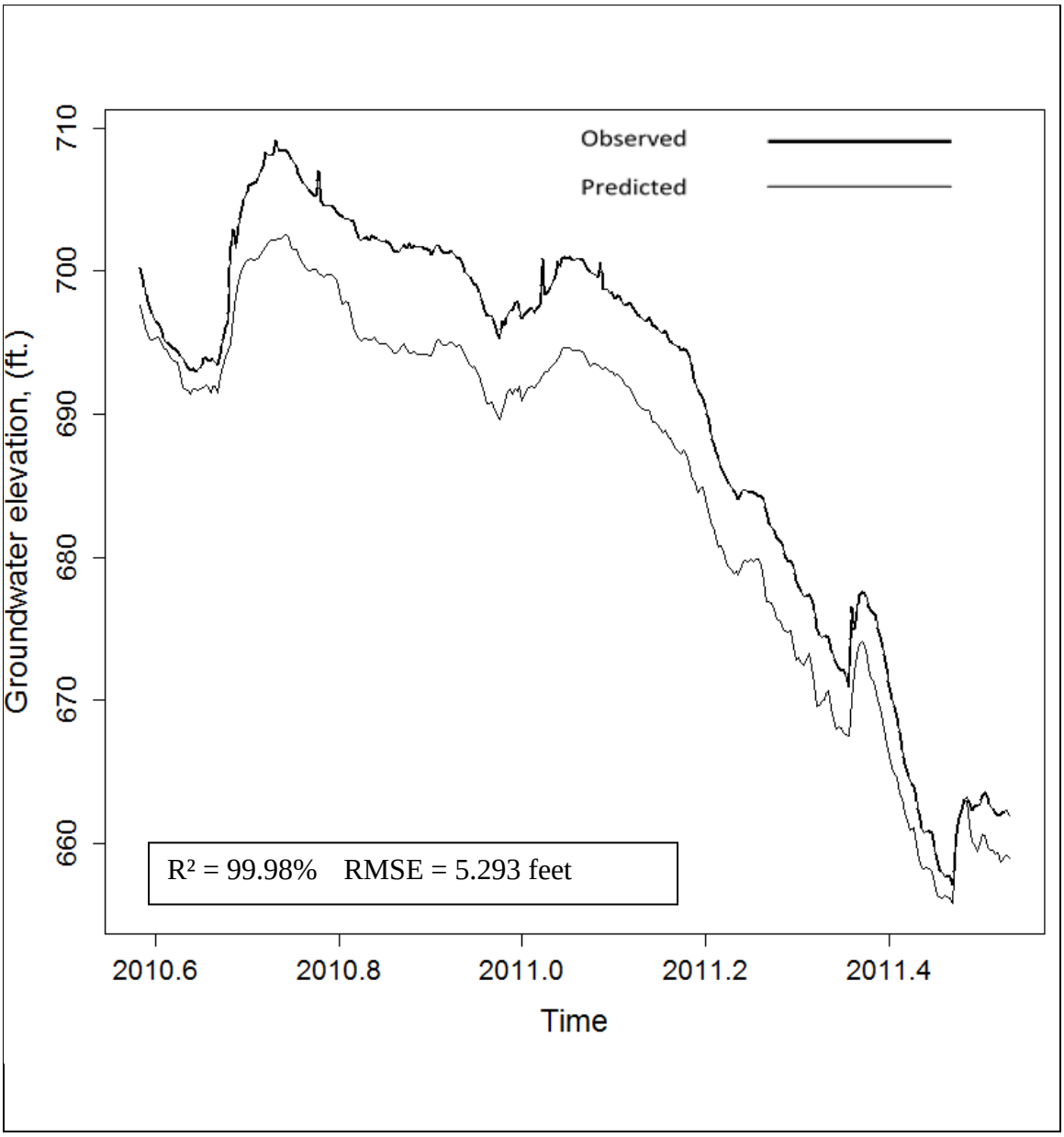

Figure 57. Cluster 4: Well H ANN predictions. Trained with Well D and 90-day moving average precipitation as input.

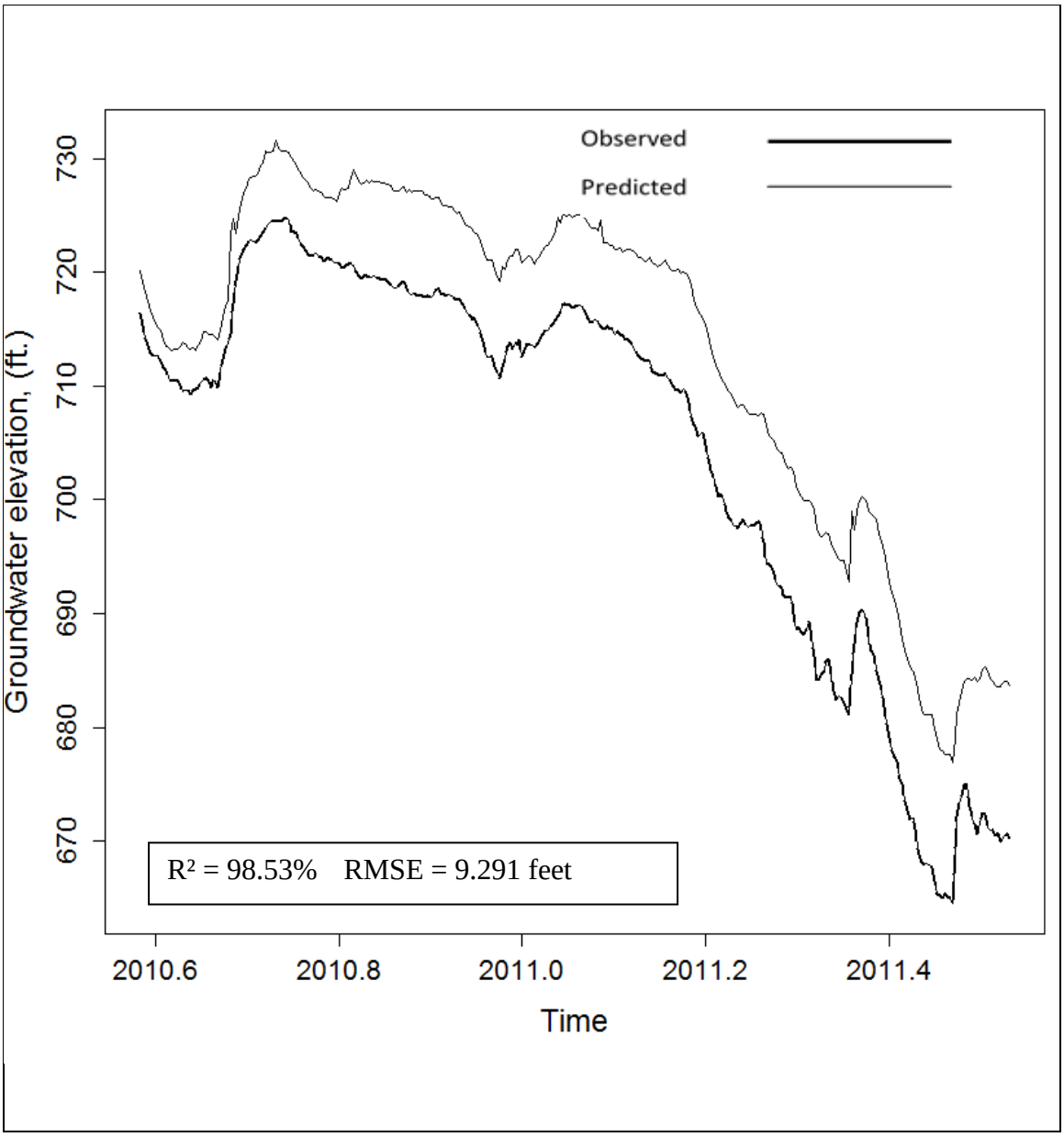

Figure 58. Cluster 4: Well D ANN predictions. Trained with well H and a 90-day moving average of precipitation as input.

IV. CONCLUSION

The complex hydrogeology of the Edwards Aquifer result in a wide range of observed groundwater levels that change as hydrologic conditions oscillate from wet to dry conditions and back, challenging accurate predictions groundwater levels. Further complexity is introduced from thousands of pumping wells which affect nearby wells groundwater levels to a varying degree. To address these issues when predicting groundwater levels in the Edwards Aquifer, in step-one of this study, the data mining technique of hierarchical clustering was used to identify wells that respond similarly under all hydrologic conditions with the aim of providing highly robust data for ANN prediction. And, because of the spatial and temporal heterogeneity of the Edwards Aquifer, the DTW algorithm was used to match well groundwater levels although they may not have occurred at the same time. It was determined that the DTW algorithm revealed well response similarities that may not have been as easily identified if simply grouped according to geographic location or hydrologic properties, for example Cluster 3 (wells B, I, and J) and Cluster 4 (wells D and H). In Cluster 3, although the magnitude of well I's response was more pronounced than wells B and J when oscillating between hydrologic conditions, the wells in Cluster 3 wells clustered together for most of the hydrologic time-periods. When wells B and J did not cluster with well I (time-periods 4, 5, and 8) hydrologic conditions were getting drier (**Figure 41**). And, in Cluster 4, wells D and H clustered through all considered hydrologic time-periods (time-period 4 and 5 for well D were discarded due to poor data quality) except for time-period 8 when hydrologic conditions were getting drier.

Next, in step-two, the data mining technique of ANN were used to predict groundwater levels for each well using the other wells that clustered with it, 30, 60, and 90-day moving averages of precipitation and spring flows. The degree of similarity varied between each cluster, and as expected, the clusters with wells that responded the most similarly across all hydrologic conditions provided the highest quality ANN training data, such as Cluster 1. Cluster 1 ANNs predicted groundwater levels for each of the wells in the cluster with a high degree of accuracy even though well G was missing data (**Figures 44-49)**. However, the inclusion of spring flow during ANN training resulted in a poor model fit where the predictions crossed, and were poorly aligned with, the observed (**Figure 43**). But, ANN prediction accuracies were best when moving averages of precipitation were used as training data (**Figures 44-45**). In Cluster 2, although wells were highly similar and moving averages of precipitation were used as input, ANN predictions suffered due to poor quality data (**Figures 50-53**) which inhibited ANN training. The ANNs for wells in Clusters 3 resulted in lower ANN prediction accuracies because wells B, I, and J were less similar across hydrologic conditions (**Figures 54-56**). The ANNs for Cluster 4, although the RMSE was high, the R^2 was very high indicating that predictions could improve with higher quality training data such as other groundwater wells (**Figures 54-56**).

The Earth Sciences have traditionally been slow to integrate data mining techniques to model complex systems; however, this research shows that data mining techniques are effective tools to explore and model physical relationships within complex systems using environmental time-series data. The results of the ANNs highlight the importance of using high quality training data when training ANNs to predict in a highly

heterogeneous system. This study shows that using hierarchical clustering with DTW to identify high quality training data can improve ANN prediction accuracies. This methodology can be used to accurately predict missing values in hydrologic datasets (**Figure 48**), such precipitation, discharge, and spring flow datasets. Areas of future research include determining whether using ANNs to predict Cluster 2 missing and corrupted data can smooth ANN test predictions and determining whether inclusion of additional training data, such as additional wells or physical parameters, improves prediction accuracies for the wells in Cluster 3 and 4.

REFERENCES

Bear, J., & A. H. Cheng. 2010. Modeling groundwater flow and contaminant transport.

Chen, X., Y. F. Zhang, Y. Y. Zhou, & Z. C. Zhang. 2013. Analysis of hydrogeological parameters and numerical modeling groundwater in a karst watershed, southwest China. *Carbonites and Evaporates* 28:89-94.

Chitsazan, M., G. Rahmani, and A. Neyamadpour. 2013. Groundwater level simulation using artificial neural network: a case study from Aghili plain, urban area of Gotvand, south-west Iran. Geopersia 3:35-46.

Coppola Jr., E. A., A. J. Rana, V. W. Uhl, M. M. Poulton, & F. Szidarovszky. 2005. A Neural Network Model for Predicting Aquifer Water Level Elevations. *Ground Water* 43:231-241.

Coppola, E., F. Szidarovszky, M. Poulton, & E. Charles. 2003. Artificial Neural Network Approach for Predicting Transient Water Levels in a Multilayered Groundwater System under Variable State, Pumping, and Climate Conditions. *Journal of Hydrologic Engineering* 8:348.

Craven, M. W., & J. W. Shavlik. 1997. Using neural networks for data mining. *Future Generation Computer Systems* 13:211-229.

Daliakopoulos, I. N., P. Coulibaly, & I. K. Tsanis. 2005. Groundwater level forecasting using artificial neural networks. *Journal of Hydrology* 309:229-240.

Drew, D., & N. Goldscheider. 2007. Methods in Karst hydrogeology. London; New York: Taylor & Francis, c2007.

Esling, P., & C. Agon. 2012. Time-Series Data Mining. *ACM Computing Surveys* 45 (12):1-12.

Fayyad, U., G. Piatetsky-Shapiro, & P. Smyth. 1996. From Data Mining to Knowledge Discovery: an overview. In *Advances in Knowledge Discovery and Data Mining,* ed. U. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy, 1-34. Menlo Park, CA: AAAI Press.

Ford, D., & P. W. Williams. 2007. *Karst hydrogeology and geomorphology [electronic resource] / Derek Ford and Paul Williams.* Chichester, England; a Hoboken, NJ: John Wiley & Sons, c2007; Rev. ed.].

Gaur, S., S. Ch, D. Graillot, B. Chahar, & D. Kumar. 2013. Application of Artificial Neural Networks and Particle Swarm Optimization for the Management of Groundwater Resources. *Water Resources Management* 27:927-941.

Giorgino, T. 2009. Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package. *Journal of Statistical Software*.

Gupta, M., J. Gao, C. C. Aggarwal, & J. Han. 2014. *Outlier detection for temporal data / Manish Gupta, Jing Gao, Charu Aggarwal, Jiawei Han.* San Rafael, California (1537 Fourth Street, San Rafael, CA 94901 USA): Morgan & Claypool, 2014.

Han, J., & J. Gao. 2009. Research Challenges for Data Mining in Science and Engineering. In *Next Generation in Data Mining,* ed. H. Kargupta, J. Han, P. S. Yu, R. Motwani & V. Kumar, 1-11. Chapman & Hill.

Hastie, T., R. Tibshirani, & J. H. Friedman. 2009. *The elements of statistical learning: data mining, inference, and prediction / Trevor Hastie, Robert Tibshirani, Jerome Friedman.*New York: Springer, c2009; 2nd ed.

Hensel, D. R., & R. M. Hirsch. 2002. *Statistical Methods in Water Resources.* U.S. Geological Survey.

Hoffman, F. M., J. W. Larson, R. T. Mills, B. J. Brooks, A. R. Ganguly, W. W. Hargrove, J. Huang, J. Kumar, & R. R. Vatsavai. 2011. Data Mining in Earth System Science (DMESS 2011). *Procedia Computer Science* 4:1450-1455.

Jha, M. K., & S. Sahoo. 2015. Efficacy of neural network and genetic algorithm techniques in simulating spatio-temporal fluctuations of groundwater. *Hydrological Processes* 29:671-691.

Kalina, J. 2013. Highly Robust Methods in Data Mining. *Serbian Journal of Management* 8:9-24.

Karthikeyan, L., D. Kumar, D. Graillot, & S. Gaur. 2013. Prediction of Ground Water Levels in the Uplands of a Tropical Coastal Riparian Wetland using Artificial Neural Networks. *Water Resources Management* 27:871-883.

Krešic, N., & Z. Stevanovic. 2010. *Groundwater hydrology of springs: engineering, theory, management, and sustainability.* Burlington, MA: Butterworth-Heinemann, c2010.

Lallahem, S., J. Mania, A. Hani, & Y. Najjar. 2005. On the use of neural networks to evaluate groundwater levels in fractured media. *Journal of Hydrology* 307:92-111.

Lindgren, R.J., A.R. Dutton, S.D. Hovorka, S.R.H. Worthington, & S. Painter. 2004. *Conceptualization and Simulation of the Edwards Aquifer, San Antonio Region, Texas.* U.S. Geological Survey Scientific Investigations Report 2004. U.S. Geological Survey, Report Number, 5277.

Land, F. L., B. B. Hunt, B. A. Smith, & P. J. Lemonds. 2011. Hydrologic Connectivity in the Edwards Awuifer between Sna Marcos Springs and Barton Springs during 2009 Drought Conditions. *Texas Water Journal* 2:39-53.

Miller, H. 2009. Geographic data mining and knowledge discovery: An overview. In H. J. Miller, & J. Han (Eds.), Geographic Data Mining and Knowledge Discovery (Second Edition): 3-32. London: CRC Press.

Mohanty, S., M. K. Jha, A. Kumar, & D. K. Panda. 2013. Comparative evaluation of numerical model and artificial neural network for simulating groundwater flow in Kathajodi–Surua Inter-basin of Odisha, India. *Journal of Hydrology* 495:38-51.

Mori, U., A. Mendiburu, & J. Lozno A. 2015. Distance Measures for Time-series in R: The TSdist Package.

O'Reilly, A. M., E. A. Roehl, P. A. Conrads, and R. C. Daamen. 2012. *Neural network analysis of the effects of the rainfall variability and groundwater usage of the hydrologic system in a karst terrain, Central Florida, USA.* Paper presented at: The 10th Int. Conf. on Hydroscience and Engineering, November 4-7, ICHE.

O'Reilly, A. M., E. A. Roehl, C. Jr. P. A., R. C. Daamen, & M. D. Petkewich. 2014. *Simulation of the Effects of Rainfall and Groundwater Use on Historical Lake Water Levels, Groundwater Levels, and Spring-flows in Central Florida.* U.S. Geological Survey Scientific Investigations Report 2014. U.S. Geological Survey, Report Number, 5032.

Paasche, H., D. Eberle, S. Das, A. Cooper, P. Debba, P. Dietrich, N. Dudeni-Thlone, C. Gläßer, A. Kijko, A. Knobloch, A. Lausch, U. Meyer, A. Smit, E. Stettler, & U. Werban. 2014. Are Earth Sciences lagging behind in data integration methodologies? *Environmental Earth Sciences* 71:1997-2003.

Rakhshandehroo, G., M. Vaghefi, & M. Aghbolaghi. 2012. Forecasting Groundwater Level in Shiraz Plain Using Artificial Neural Networks. *Arabian Journal for Science & Engineering (Springer Science & Business Media B.V.)* 37:1871-1883.

Ratanamahatana, C., & E. J. Keogh. 2004. *Making Time-Series Classification More Accurate Using Learned Constraints.* Paper presented at: Proceedings of the Fourth SIAM International Conference on Data Mining, April 22-24, University of Trier: Schloss Dagstuhl - Leibniz Center for Informatics, pp. 11-22.

Rose, P. R. (1972). *Edwards Group, surface and subsurface, central Texas*. Austin: University of Texas Bureau of Economic Geology Report of Investigations 74.

Sahoo, S. & M. K. Jha. 2013. Groundwater-level prediction using multiple linear regression and artificial neural network techniques: a comparative assessment. *Hydrogeology Journal* 21:1865-1887.

Sakoe, H. & S. Chiba. 1978. Dynamic Programming Algorithm Optimization for Spoken Word Recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 1:43-49.

Sirhan, H., & M. Koch. 2012. *Prediction of Dynamic Groundwater Levels in the Gaza Coastal Aquifer, South Palestine, using Artificial Neural Networks.* Paper presented at: First International Colloquium REZAS12, November 13-16, pp. 1-26.

Sreekanth, P. D., N. Geethanjali, P. D. Sreedevi, S. Ahmed, N. R. Kumar, & P. D. K. Jayanthi. 2009. Forecasting groundwater level using artificial neural networks. *Current Science (00113891)* 96:933-939.

Sunitha, G., & A. Mohan Reddy. 2014. Mining Frequent Patterns from Spatio-Temporal Data Sets: a Survey. *Journal of Theoretical & Applied Information Technology* 68:265-274.

Szidarovszky, F., E. A. Coppola, J. Long, A. D. Hall, & M. M. Poulton. 2007. A Hybrid Artificial Neural Network-Numerical Model for Ground Water Problems. *Ground Water* 45:590-600.

Taneja, A., & R. K. Chauhan. 2011. A Theoretical Framework for Comparison of Data Mining Techniques. *International Journal of Advanced Research in Computer Science* 2:28-33.

Tapoglou, E., I. C. Trichakis, Z. Dokou, I. K. Nikolos, & G. P. Karatzas. 2014. Groundwater-level forecasting under climate change scenarios using an artificial neural network trained with particle swarm optimization. *Hydrological Sciences Journal/Journal des Sciences Hydrologiques* 59:1225-1239.

Terzi, Ö. 2011. Monthly River Flow Forecasting by Data Mining Process. In *Knowledge-Oriented Applications in Data Mining*, ed. K. Funatsu, 127-134. InTech, 2011-01-21.

Terzi, Ö. 2012. Monthly Rainfall Estimation Using Data-Mining Process. *Applied Computational Intelligence & Soft Computing*1-6.

Trichakis, I. C., I. K. Nikolos, & G. P. Karatzas. 2009. Optimal selection of artificial neural network parameters for the prediction of a karstic aquifer's response. *Hydrological Processes*.

Trichakis, I. C., I. K. Nikolos, & G. P. Karatzas. 2011. Artificial Neural Network (ANN) Based Modeling for Karstic Groundwater Level Simulation. *Water Resources Management* 25:1143-1152.

Wu, J. 2010. Hydrological predictions using data-driven models coupled with data preprocessing techniques.

Wu, J., & X. Zeng. 2013. Review of the uncertainty analysis of groundwater numerical simulation. *Chinese Science Bulletin* 58:3044-3052.

Zhou, Y., & W. Li. 2011. A review of regional groundwater flow modeling. *Geoscience Frontiers* 2:205-214.