Report: ARIMA Forecasting of Well G-1260 Water Height

TEAM LEAD: JAKE LASKY

TEAM MEMBERS: ABIGAIL DEXTER-BOONE, JD KENT, CHRIS NOBBLITT

SUMMARY

To assist the US Department of Agriculture's program on Water, Sustainability, and Climate in the prediction of hourly well water height levels, we developed a seasonal ARIMA model to forecast a week-long period for Well G-1260, located outside of Deerfield Beach, Florida. This involved using well water height data aggregated hourly over a span of approximately 10 years, from October 1, 2007, through June 8, 2018. Following the Hyndman approach, we assumed a deterministic seasonality and fit trigonometric functions to our time series, and an ARIMA(12,1,2) was fit with a MAPE of 0.054 and an MAE of 0.34. Moving forward, we suggest using additional sources of information on local rainfall and tide levels.

METHODOLOGY AND RESULTS

We identified an ARIMA(12,1,2) model fit with ten sine and cosine functions to best predict well water height levels. This model shows a Mean Absolute Percent Error (MAPE) of 0.054 and a Mean Absolute Error (MAE) of 0.34.

Our dataset analysis started with approximately 113,000 observations of well water height, recorded at least every hour, often in 15-minute intervals. This series was aggregated up to hourly intervals by averaging all incremented time steps within the hour. At this point, there were approximately 500 missing hourly observations, which can be seen highlighted in red in *Figure 1*. These missing hourly data points were imputed with linear interpolation. Our complete data began October 1, 2007 and was recorded through June 8, 2018. We subsetted the final week of observations to use as a validation dataset in which to test our forecasted model for accuracy. All model performance measures that are reported are derived from the validation dataset.

**FIGURE 1:** WELL WATER HEIGHT OVER TIME WITH MISSING VALUES INDICATED BY RED BARS.
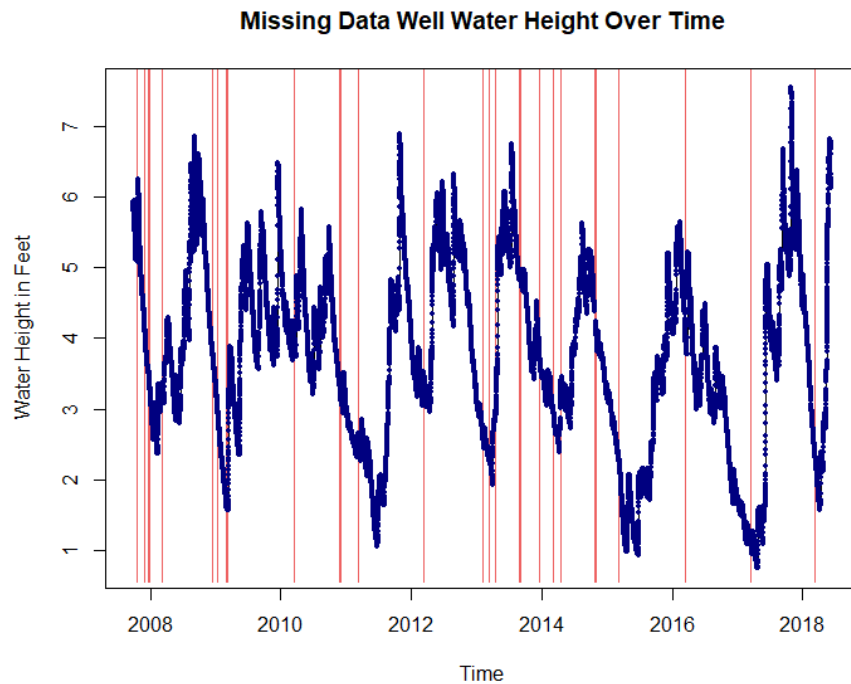
Based on the initial decomposition of our time series, seen in *Figure 2*, we determined an annual seasonal component to be appropriate. Because of the long seasonal period, we assumed deterministic seasonality in a Hyndman-like approach. To account for that seasonality, we used ten sine and cosine functions to model the data over a 365.25-day period (8766 hours). After taking differences of lag one to account for a random walk, a Dickey-Fuller test indicated stationarity of the series.

FIGURE 2: INITIAL DECOMPOSITION OF TIME SERIES WITH AN ANNUAL SEASONAL COMPONENT. TITLE: STL DECOMPOSITION OF DATA
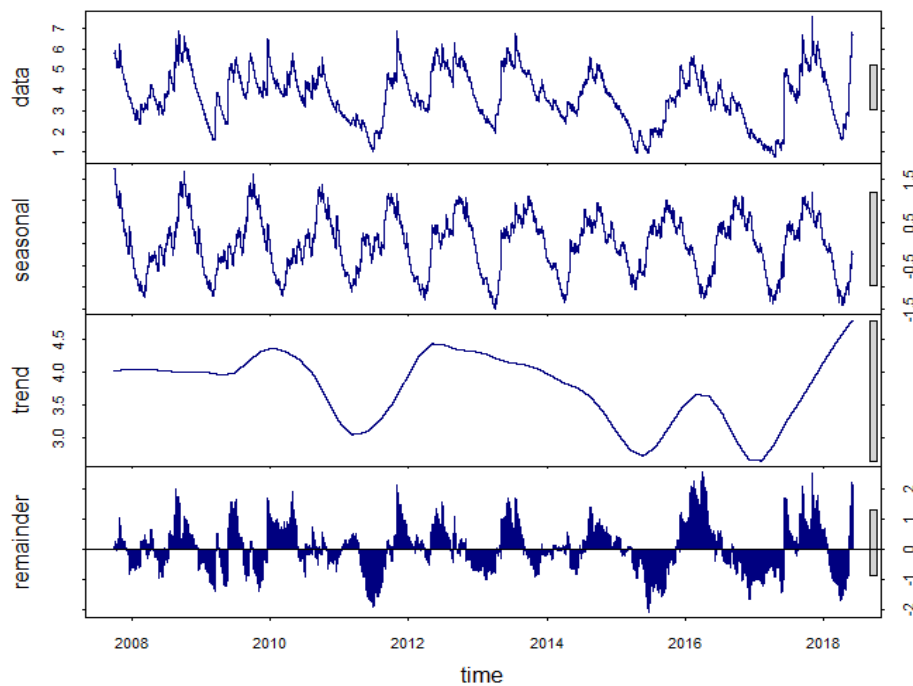
After iteratively fitting various AR and MA terms to the ARIMA model and examining the Ljung-Box test for white noise, seen in *Figure 3*, in order to model as much of the correlation structure as possible, we settled on an ARIMA(12,1,2) model. Predicted well water height levels for the final week of our dataset are shown in *Figure 4* in red, with the actual water height levels shown in black. We can see that the while there is some unexplained variation within the actual water height values, our predicted levels do align with the observed peak of our data.

FIGURE 3: LJUNG-BOX TEST FOR WHITE NOISE, HIGHER BARS INDICATE WHITE NOISE.
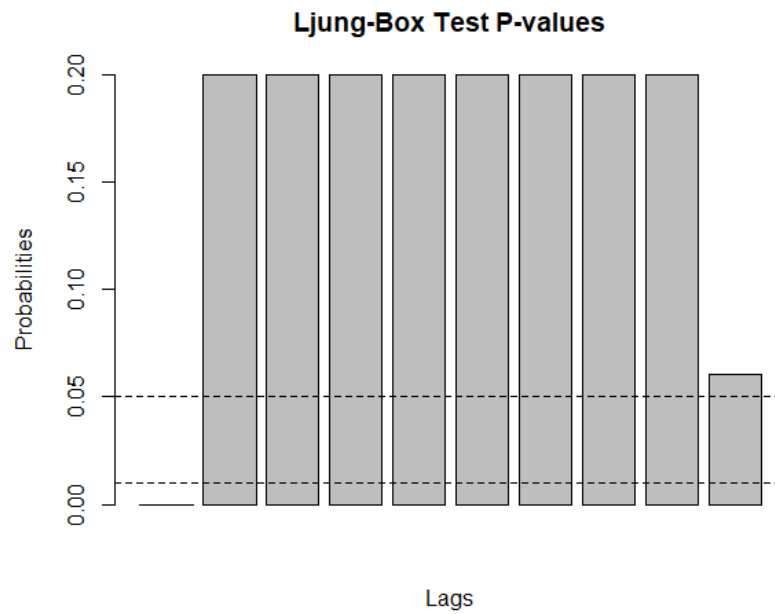
FIGURE 4: ONE-WEEK PREDICTION OF WELL WATER HEIGHT. THE RED LINE SHOWS FORECASTED WELL HEIGHT WITH ACTUAL LEVELS SHOWN IN BLACK. FORECASTED PERIOD FROM JUNE 1, 2018 12:00PM TO JUNE 8, 11:00AM.
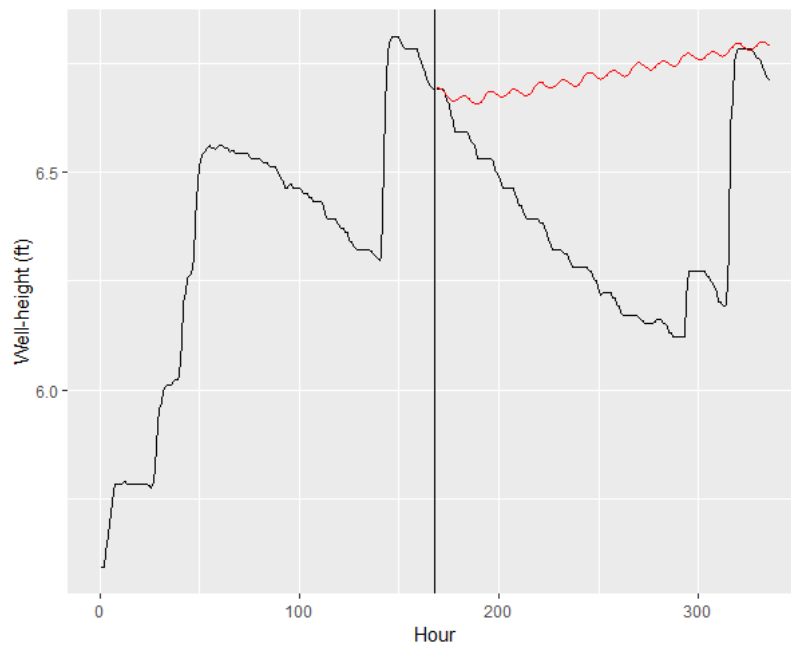
CONCLUSION AND RECOMMENDATIONS

Developing a model to predict hourly well water height levels can assist the US Department of Agriculture's program on Water, Sustainability, and Climate in forecasting water levels. This can allow for a more efficient use of organizational resources and can inform future water availability. Our analysis indicates a deterministic seasonal pattern and trend due to a random walk. An ARIMA(12,1,2) was ultimately decided to appropriately maximize white noise and allow for a reasonable forecast of hourly well height levels.

We acknowledge that within a complicated system such as well water levels, our predictive power is likely limited based on the amount of information available. It is likely that other factors influencing water systems could offer important insights into well-water height, such as regional rainfall or even tidal patterns, as this well is located along the coastline of Florida. Including additional water source information in an ARIMAX model could improve predictive accuracy and allow for a more comprehensive picture of future well water levels.