**Average Well Depth By Month:**

**Time Series Analysis**

Time Series

Dr. Susan Simmons

Reported by: Orange Team 5

Dave Hiltbrand, Liz Mao, Michael Yuan, Melissa Sandahl, Ashwin Bhat

**Summary**

As a group of analysts, we are responsible for creating time series models for predicting the level of water (in feet) in the well, F-179, located near Miami, FL. Our goal is to create a stationary time series which can later be used in a forecasting model such as Autoregressive Integrated Moving Averages (ARIMA). Based on the results from the Augmented Dickey-Fuller (ADF) test with Lag 0 through 2, a significant level of evidence suggests the well is stationary.

Before testing that a unit root existed, also known as nonstationary, we had to deal with the seasonality that was found during STL decomposition. We tested the strength of the season and adjusted the series based on the results of the test.

Looking ahead we will be developing an ARIMA model to forecast future depths of the well. Had we not first confirmed stationarity in our well, the results of the ARIMA model would have been unreliable.

**Background**

Our dataset is a sample of 253,925 measurements of depth in feet collected from October 1, 2007, until June 4, 2018, in intervals of fifteen minutes. We were given information on the well depth on the original scale of well feet and an adjusted measurement of well feet. We were recommended to use the adjusted variable which uses a reduction to account for a change in the standard height of the well. We reduced the dataset to 129 observations, including 123 in the training and 6 in the validation dataset by aggregating our quarter-hourly data into monthly data and computed the average depth for each month. There were 955 missing observations in our original time series data. Since the data does not contain any months that are completely missing, we used the average well depth for a month as our variable instead of imputing the missing values.

During the data cleaning, we discovered there are negative values in the well depth variable. Certain forecasting models, such as the Holt-Winter’s multiplicative model, do not allow negative values and thus would cause trouble when fitting the model. To fix this problem, we shifted the data by adding 1 to each of the observation.

We ran an STL decomposition to determine whether seasonal and trend components were present. Based on the results there was an annual seasonal component but virtually no trend. In order to better forecast the well depth in the future using ARIMA model, the time series needs to be stationary by modeling trend, seasonality and correlation structure. We implemented the ADF test to determine if the time series is stationary.

**Results**

The first step to achieve stationarity was to remove the seasonal component that had been discovered during STL decomposition. We used the forecast R package to measure the strength of the seasonal component. The outcome exceeded the threshold of the test indicating to take a seasonal difference. After creating the differenced time series, we measured the seasonal strength a second time. The result was below the test’s threshold indicating there was no longer a seasonal component.

To test for stationarity of the time series we ran the Augmented Dickey-Fuller (ADF) test. The results gave us a p-value below a level of significance at 5% which suggested there was enough evidence to support the alternative hypothesis that the time series was stationary.

The graph of the time series after achieving stationarity is in Figure 1. Following this, Figure 2 displays the results from running the ADF test on the well times series data. Using a lag 0, lag 1, and lag 2 we found that the well data is stationary as all of the p-values are well below a threshold for a level of significance at 5%.



Figure 1. Stationary Time Series after removing seasonal differences.

|  |  |  |  |
| --- | --- | --- | --- |
| **Test Statistic** | **P-value** | **Lag** | **Result** |
| -5.36 | 0.01 | 1 | stationary |
| -4.31 | 0.01 | 2 | stationary |
| -5.26 | 0.01 | 3 | stationary |

Figure 2. Results from Augmented Dickey-Fuller Test in R

**Conclusion**

The Augmented Dickey-Fuller test allows us to determine whether we need to modify our dataset by taking differences in order to build a more accurate ARIMA model, therefore yielding more accurate forecasts.

Achieving stationarity is important to forecast modeling and has to meet three criteria:

1. Constant mean across the time series
2. Stable variance across the time series
3. Constant autocorrelation across the time series

Without meeting these criteria an ARMA model could have different accuracy at different sections through the time series.

We were able to achieve stationarity in the well time series by first removing the seasonal component by taking seasonal differencing. Then, the stationarity of the time series was confirmed by using the Augmented Dickey-Fuller test.

Looking ahead we would like to build a model a model using ARIMA. The goal would be the ability to predict future depths of the well with high accuracy and interpretability.

**Appendix**

1. Information about the well gathered at <https://www.ncdc.noaa.gov/>
2. Seasonal difference equivalent to
3. The more negative a test statistic from the ADF test, the more the alternative hypothesis is support
4. The threshold for testing seasonal strength greater than 0.64 via following method using the *nsdiffs* function