Time Series Analysis Comparing Climatic Averages and the Water Levels of Aquifers in Albany, NY and Queens, NY

Angela Rienzo[1], areinzo1@pride.hofstra.edu
Sally Louise Roscoe[1], sroscoe1@pride.hofstra.edu
Lauren Mahoney[1], lmahoney1@pride.hofstra.edu
Paul Weinstein[1], pweinstein2@pride.hofstra.edu
Katerina G. Tsakiri[2], ktsakiri@rider.edu
Charalambos Petrocheilos[3], charalambos.petrocheilos@postgrad.plymouth.ac.uk
Antonios Marsellos[1], antonios.e.marsellos@hofstra.edu

[1]Department of Geology, Environment and Sustainability, Hofstra University, Hempstead, NY
[2]Department of Information Systems and Supply Chain Management, College of Business Administration, Rider University, Lawrence, NJ
[3]School of Computing, Electronics, and Mathematics at Plymouth University, United Kingdom

---

1. Introduction

The topography of New York State varies by locality and this may affect the behavior of the watershed and subsequently the aquifers at different regions. This study analyzes two aquifers located approximately 265 km apart in New York State, as seen in Figure 1. The impact of atmospheric and hydrogeological factors on the water levels in Upstate New York has been modeled extensively in the past, including the implementation of GIS techniques to study flood hazards caused by ice jams on the Mohawk River in Upstate New York (Mahoney et al., 2018; Plitnick et al., 2018; Rienzo et al., 2018; Tsakiri et al., 2018; Marsellos & Tsakiri, 2015). The aquifers are reflective of the differing environmental conditions in the respective regions. Time series analysis are commonly utilized to compare the variables of groundwater recharge, as seen in recent studies of the US (Fu, 2019). This research specifically focuses on the aquifers in Queens and Albany which are impacted differently by the differing natural environments. Queens is located in the Long Island Watershed and is more densely populated with a flat elevation relative to Upstate New York. Albany is located in the Lower Hudson Watershed, and is more mountainous region while being less densely populated. This dynamic in Albany may cause variable behavior of the aquifer and subsequent groundwater levels to provoke flood events, to change runoff and precipitation, and recharge or discharge periodicity of the local aquifers, however, more of its groundwater is diverted to supply fresh drinking water in the New York metropolitan area. Groundwater levels in Upstate and subsequent flood occurrences may be potentially influenced by changes in water consumption in metro New York. Albany is located near the Mohawk Watershed, comprised of the Mohawk River, which flows into the Hudson River, and then finally drains into the Atlantic Ocean (Figure 1). Queens is located in the Long Island Watershed, which also drains directly into the Atlantic Ocean (Figure 1).

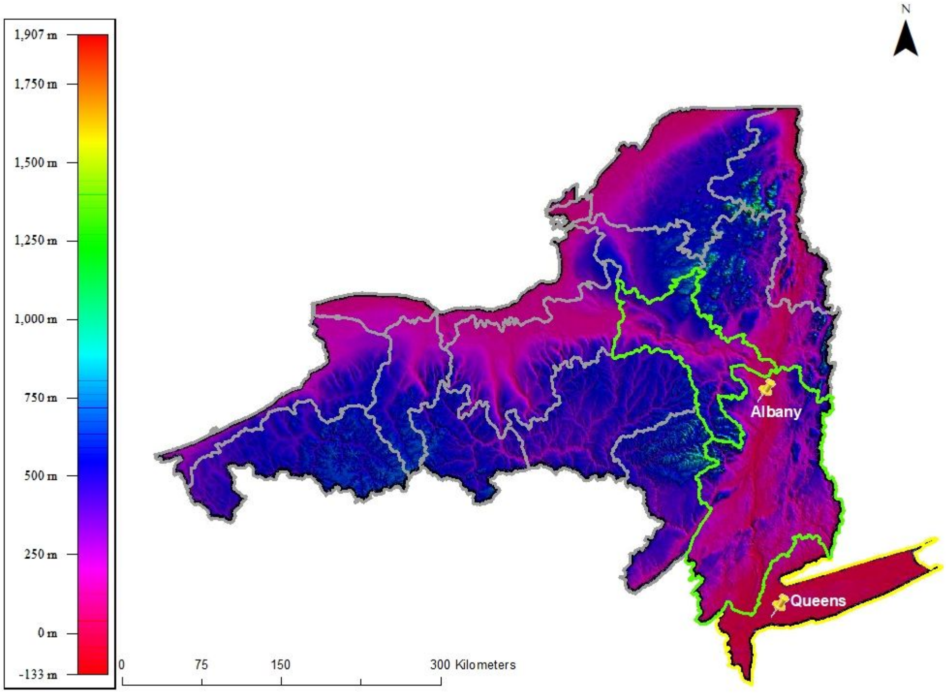

Figure 1: Watershed boundaries associated with Albany and Queens regions are shown on the map delineated by green and yellow color for the two locations, respectively. Watersheds are underlain by a digital elevation model (DEM) of New York State to better envision the contrasting landscape and proximity of the two locations.

2. Methodology

The weather data was collected from NOAA ranging from January 1st, 2010 to February 2nd, 2019. Separate datasets were created for both Queens and Albany, taken from JFK International Airport (https://www.ncdc.noaa.gov/cdo-web/datasets/LCD/stations/WBAN:94789/detail) and Albany International Airport (https://www.ncdc.noaa.gov/cdo-web/datasets/NORMAL_DLY/stations/GHCND:USW00014735/detail). The average daily variables derived from this dataset were precipitation, temperature, evaporation, wind speed, and tidal prediction. The groundwater levels of the two wells were collected from USGS and joined in a dataset of the same time range. The daily precipitation was measured in inches, temperature is measured in degrees fahrenheit, evaporation rate, wind speed in miles per hour, and tidal prediction in feet for daily highs and lows. The data were compiled, cleaned and post-processed in KNIME, which is an open source software for data analytics. Periodicities were retrieved by conducting spectral analysis in R using the time series analysis package (TSA package in R; Chan and Ripley, 2018) for Queens and Albany locations. A digital elevation model (DEM) of the study area, shown in Figure 2, was created in Google Earth and Global Mapper to compare the relative elevations of Queens and Albany and their surrounding regions.

The datasets were imported into the statistical analysis program SPSS for further statistical analysis such as intercollinearity between the independent variables on each linear regression model in Queens and Albany, respectively. Groundwater levels of each region areis dependent on the daily atmospheric variables derived from NOAA. A possible lag operator (in days) has been applied to the independent variables in respect to the groundwater in Queens and Albany that may improve the coefficient of determination derived by expression 1. Expression (1) is representative of the relationship between the variables in both Queens and Albany, NY.

$$GW = b_0 + b_1 X_{precipitation} + b_2 X_{temperature} + b_3 X_{evaporation} + b_4 X_{wind\ speed} + b_5 X_{tidal\ prediction} \qquad (1)$$

***Expression 1**:* Linear regression between groundwater (GW) and other atmospheric/tidal variables that has been applied in Queens and Albany station.

3. Results

The maximum value for the lag operator in Queen's groundwater was estimated to be 210-days, as it is shown in Figure 2 increasing the correlation by approximately four times. An addition improvement in correlation has been implemented by applying a lag operator of 82-days on the minimum tide values.

Using the spectral analysis, we determined two main periodicities for the groundwater in Queens and Albany regions, respectively (Table 1). For the Queens region, the two main periodicities are 250- and 375-days while for the Albany region, the two main periodicities are 375- and 100-days. Both regions present a 375-days period which is related with the seasonal component of the time series. In addition, the periodogram plots of both regions have been presented in Figure 3 showing the two high spectral density peaks for each location.

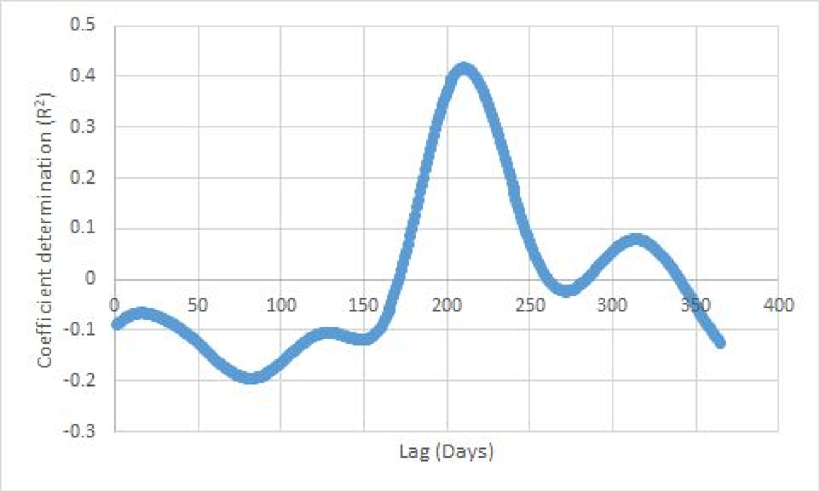

Figure 2: Coefficient determination (R^2) improvement between the groundwater by applying the lag operator for 365 days in groundwater from Queens in respect to groundwater from Albany.

Table 1: Main periodicities of the groundwater levels from the two wells located at Queens and Albany regions estimated using the spectral analysis method.

Queens

Frequency (cycles/time unit)	Spectral	Period (days)
0.004	1.64	250
0.00267	1.41	375

Albany

Frequency (cycles/time unit)	Spectral	Period (days)
0.00267	25	375
0.01	7.05	100

Applied the multiple linear regression model for both regions (Albany and Queens), we derive the best estimates for the coefficients and the coefficient of the determination in the model. The adjusted coefficient of determination for the Albany region is equal with 0.8075 while the adjusted coefficient of determination for the Queens region is equal with 0.5205 (Table 2). In addition, as it shown in Table 2, all the coefficients in the multiple linear regression model are statistically significant since the P-values are less than 0.05. Figure 4 also presents the real and predicted values of the long term component of the water discharge in Albany region.

Table 2. Linear Regression results for the groundwater well stations in Albany and Queens regions. The coefficient of determination, R_2, the standard error, the coefficients, and the p-values (less than 0.05) of the coefficients are presented in both linear regression models for the long-term component of the water discharge.

Albany - Long Term	**Coeff.**	**Std. Err.**	**p-value**
Wind speed	-0.0521	0.0031	0.00
Maximun DryBulb Temperature (lag of 8-days)	0.0032	0.0003	0.00
Precipitation (lag of 11-days)	-2.9911	0.0669	0.00
Minimum Tides (lag of 82-days)	0.2577	0.0099	0.00
Intercept	0.473	0.0399	0.00
Multiple R-Squared: 0.8081			
Adjusted R-Squared: 0.8075			

Queens - Long Term	**Coeff.**	**Std. Err.**	**p-value**
SeaLevel Pressure	-0.1997	0.0248	2.00E-15
Maximum DryBulb Temperature (lag of 90-days)	-0.0015	0.0001	0
Precipitation (lag of 16-days)	0.5383	0.0235	0
Intercept	6.0297	0.7429	1.11E-15
Multiple R-Squared: 0.5216			
Adjusted R-Squared: 0.5205			

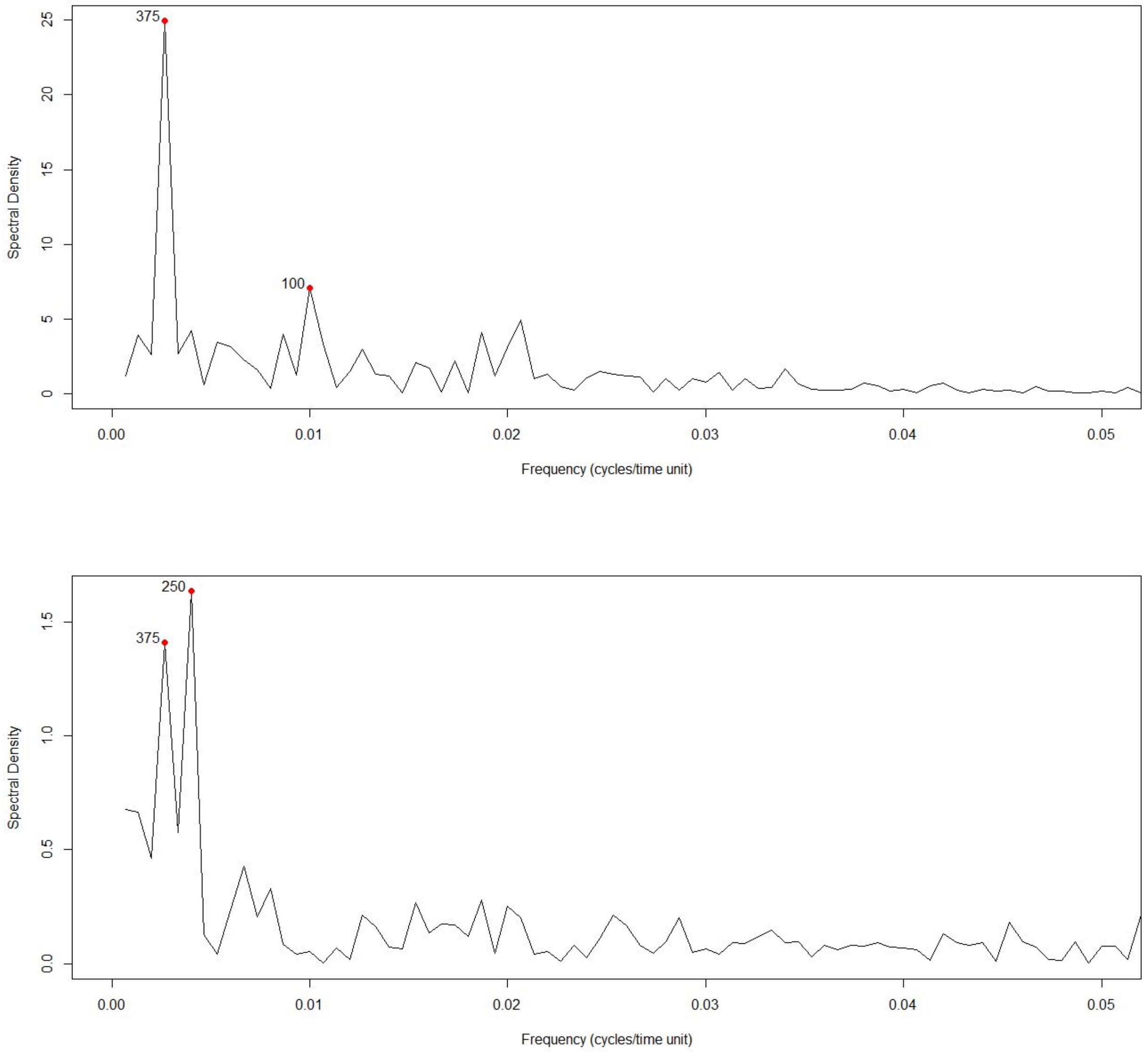

Figure 3. Periodograms showing the major cycles and associated high spectral density peaks of the groundwater levels from the well stations at Albany (upper) and Queens region (lower). Periodicities in days are showing with red bullet on each corresponding peak.

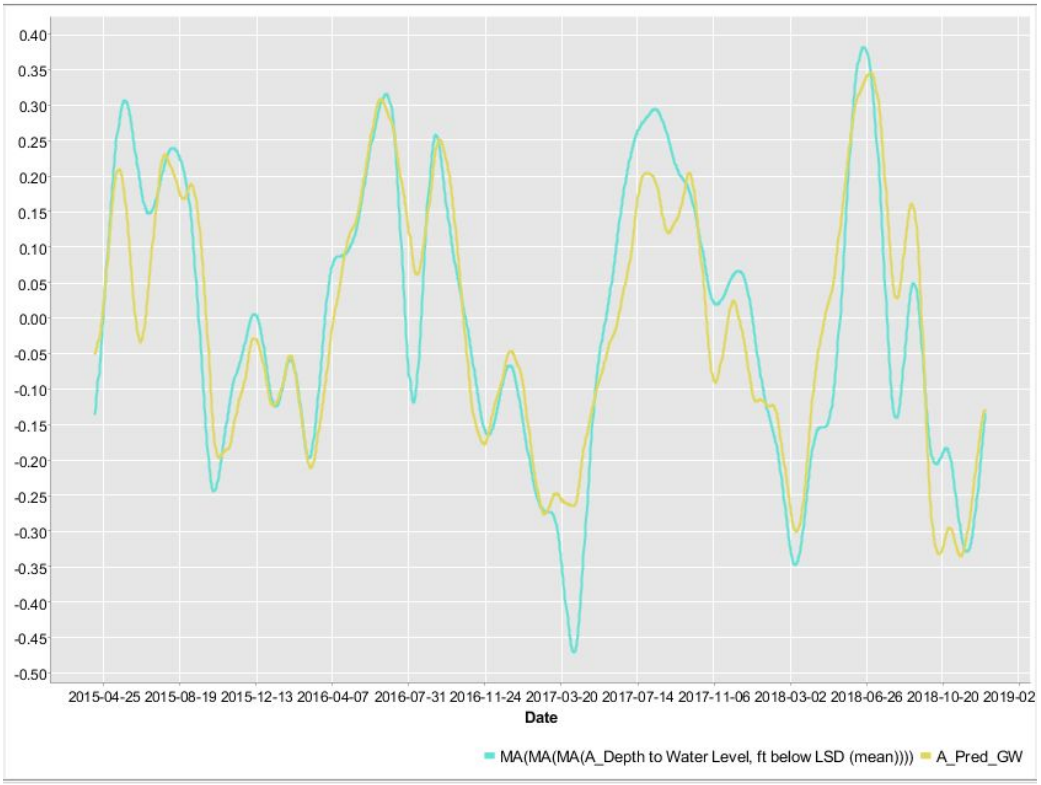

Figure 4. Linear regression model describing the detrended long-term component of the groundwater level (y-axis) from the well in Albany station against time (x-axis). Real values of the long-term component of the groundwater (blue) are compared to the predicted long-term component groundwater values (yellow).

4. Discussion

Albany located at the Lower Hudson watershed has a comparatively lower elevation than the surrounding area, resulting in larger water accumulation, especially being at a vicinity of a sizeable drainage area. As Albany is a suburban region and Queens is an urban region, there is a lower density of infrastructure in Albany. The human development of the natural terrain and associated water usage, and the relative smaller watershed area in Queens decrease the required groundwater major period (spectral density peak at 250 days) as it has been determined (Table 1) in comparison to the Albany larger watershed area and related longer period (375-days).

While Albany experiences more annual snowfall than Queens, colder prolonged temperatures and presence of snow extends the period to melt ice and therefore reach the aquifers. This could be a plausible explanation as to why the Albany watershed and related aquifer at the examined well requires more days for a period. At the contrary, Queens typically has a comparatively warmer climate with less snowfall than Albany, and with a smaller aquifer. These factors result in the Queens aquifer to be more readily rechargeable water source and more expandable than Albany due to the climate and precipitation of the region.

The Albany related aquifer is more affected by the regional tidal oscillations, as it's larger mass exposed above the sea level (relative to the Queens aquifer) results in the groundwater being more susceptible to the gravitational impact of the tides. Therefore, a possible correlation between the lunar cycles controlling the tides and the recharge rates of groundwater in the aquifer may be reasonable to apply such as the 82 days lag (that is four times the moon cycle). It is worthy to mention, that initially there is an inverse correlation between the tides and the depth to water (groundwater), meaning that the groundwater level actually increases and the aquifer recharges. After 82 days, which is the length of three lunar cycles, there is a stronger association but direct relationship rather than an inverse correlation, meaning that the groundwater depth will increase as the tides increase, so the amount of groundwater will decrease and the aquifer will discharge.

Both the Lower Hudson Watershed and Mohawk Watershed drain south towards Queens, it surpases the Long Island Watershed, and instead flows directly into the Atlantic Ocean. This situation provides an explanation as to why the Long Island Watershed does not experience quicker recharge rates, as it is not located in the water flow path of it's northern aquifers. One possible variable that our model didn't account for that could be affecting the groundwater of the Queens aquifer is the sea outflow, which is the groundwater discharged out into the ocean along the coastline. Since Queens is located so close to the Atlantic Ocean, a fair amount of groundwater could be lost to the Atlantic Ocean, and this variable is difficult to quantitatively measure and interpret. Additionally, baseflow from precipitation in Queens could discharge into the Atlantic as well, since the terrain is so flat that the runoff would flow into the ocean.

5. Conclusion

The groundwater fluctuations of the Queen's aquifer show a shorter period implying a faster recharge cycle than the Albany aquifer. Variables used to calculate Albany's groundwater produced a high R-Squared value. The lower R-Squared value produced from Queen's groundwater indicates that the variables did not account for enough of the possible impacts for Queens. This suggests that the groundwater in Queens is more affected by other values such as human impact, topography, size. sea outflow etc. when compared to the groundwater in Albany nevertheless that Queens and Albany have similar periodicities and we should expect similar prediction accuracy.

6. References

Fu, G., 2019, Attributing variations of temporal and spatial groundwater recharge: A statistical analysis of climatic and non-climatic factors Journal of hydrology Volume: 568, ISSN: 0022-1694 Online ISSN: 1879-2707

Chan K.S. and Ripley B., 2018, TSA: Time Series Analysis. R package version 1.2. https://CRAN.R-project.org/package=TSA

Mahoney, L., Roscoe, S.L., Marsellos, A.E., 2018, Reconstruction and flood simulation using GIS and Google Earth to determine the extent and damage of the January 14-15th 2018 ice jams on the Mohawk River in Schenectady, New York. Proceedings of the 2018 Mohawk Watershed Symposium, Schenectady, NY, ISBN: 978-1-939968-17-3, v.10:39-42.

Marsellos, A.E., Tsakiri, K.G., 2015. Flooding prediction in a large watershed: An example from Mohawk River in New York. Mohawk Watershed Symposium, March 2015, ISBN: 978-1-939968-05-0, p. 35-38.

Plitnick, T.A., Marsellos, A.E., Tsakiri, K.G., 2018. Time Series Regression for forecasting flood events in Schenectady, New York, Geosciences. 8(9), https://doi/10.3390/geosciences8090317.

Rienzo, A., Weinstein, P., Mecca, K., Marsellos, A.E., 2018, Multiple Flooding locations in Oneida County, NY in 2017: An approach to determine flood vulnerable sites using LiDAR in Geographical Information Systems (GIS) and flood simulations. Proceedings of the 2018 Mohawk Watershed Symposium, Schenectady, NY, ISBN: 978-1-939968-17-3, v.10:50-53.

Tsakiri, K., Marsellos, A.E., Kapetanakis, S., 2018. "Artificial Neural Network and Multiple Linear Regression for Flood Prediction in Mohawk River, New York." Water 10, no. 9: 1158. https://doi.org/10.3390/w10091158.