**Spatially and Statistically Analyzing the Freshwater-Saltwater Interface in the Floridan Aquifer in Charleston, Berkeley, and Dorchester Counties in South Carolina**

Mikala Randich April 20, 2016

**Research Question**

Saltwater intrusion is a documented phenomenon in many coastal freshwater aquifer systems, and has been documented along the Atlantic Coast in the United States (Barlow, 2003). Specifically in this study region, only one study has been conducted to document the extent of the saltwater intrusion into the Floridan aquifer in the tri-county region. This study was conducted by Park (1985) for the South Carolina Department of Natural Resources, previously named the South Carolina Water Resources Commission. The amount that the extent of intrusion into this aquifer in the study region has changed in the past 30 years, and the location of the freshwater-saltwater interface in the aquifer at present day are largely unknown. One portion of this project aims to answer the question of whether Chloride concentrations in the Floridan aquifer in Charleston, Berkeley, and Dorchester Counties in South Carolina can be predicted based on their spatial location. For this portion of the project, Park’s data published in his 1985 report is used. When updated data in the form of Chloride ion concentrations and updated GPS coordinates are collected in the course of the overall project, the same methods and analyses will be applied to that collected data, as well as comparative analyses between the two datasets.

Based on observed and studied properties of saltwater intrusion, it is expected that the location of the freshwater-saltwater interface in the Floridan aquifer has advanced further inland due to continued pumping of Floridan groundwater wells in the past 30 years (El Moujabber et al., 2006; Demirel, 2004; Paniconi et al., 2001; Bear et al., 1999; Narayan et al., 2003). Pumping of aquifer water has a documented effect of advancing the freshwater-saltwater interface further inland from the coast as water is removed from the aquifer, causing a drop in the surrounding potentiometric surface (Theis, 1938), and allowing saltwater from the coastal ocean to be drawn into the aquifer. Since the pumping of the Floridan aquifer has continued for anthropogenic purposes over the past 30 years since Park’s study was conducted, it is expected that the Floridan aquifer will see a similar encroachment of saltwater, and thus an increase in Chloride ion concentration. With regard to the analysis of Park’s data, it is expected that predicted values for Chloride concentration will be higher near the coast, with lower Chloride concentrations being predicted as the predicted locations are placed further and further inland.

**Methods**

First, since Park’s data was published in printed-only form, the data was typed into electronic form in Microsoft Excel. The data was in the form of identifying the tested well by SCGRID and county-name identification, identifying the aquifer system, sampling information, then the tested water quality parameter information. The data was imported into the R statistical program for analysis. In order to analyze the past data for spatial dependence, the location data (in the form of UTM East and North coordinates) is converted into a spatial points data frame, and a grid of the spatial UTM coordinates is created. The maximum distance that should be examined in analysis is calculated by finding the maximum distance between any two points, and dividing it by two. This is done to ensure that spatial patterns that may exist between only a few points of data aren’t examined, since more pairs of samples are present at smaller scales than at the extremes of the spatial grid (Bivand et al., 2008; Gelfand et al., 2010).

A Mantel test is conducted on the Chloride and x and y coordinate spatial distance data to examine potential indication of spatial dependence. The Mantel test essentially holds one of the distance matrices constant while the other is randomly shuffled. The logic behind the test being that if the null hypothesis of no correlation between the matrices is true, then random permutations of one of the matrices should be just as likely to provide a large of small coefficient, and the permutations ensure that the test isn’t relying on assumptions regarding the statistical distributions of the matrices (Sokal & Rohlf, 1995).

Next, linear models for the data are fit using generalized least squares. This method fits a linear model for the data using generalized least squares, where: the errors are allowed to be correlated and/or have unequal variances, individual observations are weighted, or to generalize OLS (Kariya & Kurata, 2004). The first gls model is the most basic model, with the ~ 1 defining a single constant predictor, leading to a spatially constant mean coefficient (Bivand et al., 2008). The model is then updated to include exponential spatial correlation with the UTM coordinates, and the third model incorporates a nugget. Variograms are created for each of these three models (seen in Figures 2 – 4). The spatial correlation of the data is modeled by the variogram instead of a correlogram or covariogram. These variograms plot semivariance as a function of distance. When R simply computes and plots the sample variogram, the program decides to ignore direction so that point pairs are merged on the basis of distance, not direction (Bivan et al., 2008). However, in this case, the calculated appropriate maximum distance is taken into consideration.

An analysis of variance (anova) is conducted to compare the gls models, and the AIC value is used to determine which model provides the best explanation of the data. The AIC value taken into consideration the trade-off between goodness of fit and model complexity, and thus is a useful statistic for model selection. However, it doesn’t test any individual model with any hypothesis tests (Burnham & Anderson, 2002). Another variogram is created to account for potential periodicity within the data (seen in Figure 6), and this periodic variogram is visually compared to the previously displayed three variograms.

Next, a different approach to analyzing the data is taken by creating a gstat object. The gstat function uses range parameters, in this case, the exponential model with sill and range parameters as discussed in Bivand et al. (2008). A variogrom is again created for this different model. Next, the krige function is used to create a grid of predicted kriged values for Chloride based on spatial location. A map of the kriging predicted values is displayed and visually examined, alongside a map of the variance. Finally, a non-parametric density map of Chloride concentrations was created to further visually examine the data.

**Results**

The first analysis of the data that was conducted was the Mantel test to examine potential evidence of spatial dependence. The Mantel test was conducted based on the spatial x and y distances of the location data of the tested well sites. The Mantel test results indicated that there may be some slight spatial dependence, however the presence of spatial dependence was not statistically significant based on the Mantel statistic value of 0.050 and a significance of 0.238. The visual display of the distance data confirmed this indication of only slight spatial dependence, and can be seen in Figure 1.

The next step of analysis involved the creation of the three gls models that were created for the data. The first model was a simple gls model, a summary of which provided an AIC value of 1027.741. The second model updated the first gls model to include an exponential spatial correlation structure involving the UTM coordinate locations of the well sites. A summary of this model provided an AIC value of 957.734 and a p-value of 0.372. The third model is identical to the second model, except added the inclusion of a nugget. A summary of this model provided an AIC value of 959.6755 and a p-value of 0.4145. The summaries of these models can be found in Tables 1 - 3. Variograms were created for each of the corresponding models (Figures 2 – 4). Based on the simple gls model having the highest AIC value, and the simple gls model variogram appearing to have the best fit for the data, it is determined that the gls model without an exponential spatial correlation structure and without a nugget was the best model for the data. This result is verified by the anova (seen in Table 4) used to compare the three models, with the AIC and p-values indicating that the gls model being the best model, with the gls and exponential spatial correlation model being the second best, and the gls and exponential spatial correlation model with nugget being the worst model. Because these variograms exhibited some periodic tendencies, a variogram with periodicity was created to see if this was a better fit for the data (Figure 6). However, upon visual examination, the periodic variogram did not fit the data better than the basic gls variogram.

Another method of analysis using the gstat package was used to analyze Park’s data. A gstat object and variogram were created to be used by the kriging analysis to create a spatial grid of predicted Chloride values. Based on the gstat object and variogram model, the krige function created a grid of predicted Chloride concentration values and their variances, while the ssplot function visually displayed this result, which can be seen in Figure 7. As Figure 7 demonstrates, the kriged grid of predicted values essentially grouped all predicted values into two distinct regions, instead of creating an interpolated surface with a large range of values as desired. The visual display of the variances makes more sense and aligns with expectations, in that the variance is larger around the edges of the spatial area, since there is less data in these areas, and less variance where there are more data locations closer together.

Since the kriging method did not align with expectations of making a smooth kriged map of predicted Chloride concentrations, a visual display heat map function was used to visually examine the data (seen in Figure 8). This visual display aligned with expectations in that it identified the area of higher Chloride concentrations near the coast, with decreasing concentrations as you move further inland. Although this is a non-parametric analysis and display, it successfully displayed levels of Chloride concentrations in a visual way to examine the data. The density map also followed similar lines as Park’s hand-drawn contours (seen in Figure 9), indicating this same pattern may likely be present when current data is collected.

**Interpretation**

The main takeaways from this analysis are that there is not significant statistical evidence of spatial dependence, and the ability to create a kriging grid of predicted Chloride concentrations based on spatial location was not successful. While the Mantel test and the created variograms indicate that there is some evidence of spatial dependence, these indications were not statistically significant. Additionally, the three general least squares models that were compared showed that the basic gls model without exponential spatial correlation or a nugget was the best model for the data. The failure of the kriging to produce the desired prediction map and the lack of spatial dependence are forms of common error where spatial correlation is difficult to infer from sample data because of their distribution, sample size, or spatial configuration (Bivand et al., 2008). Likely, the Chloride concentrations were too similar in a given region, and the spatial distribution of well sites was not conducive to this particular analysis. Going forward, higher resolution of well locations will be sampled and likely larger variations in Chloride concentrations will be observed, which will hopefully help correct these errors. Additionally, different methods of spatial analysis can be explored with the modern data collected in the course of this project.

**Figures and Tables**

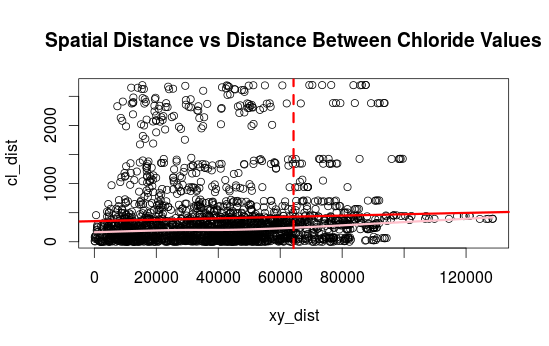


Figure : Plot of X and Y Coordinate distance versus distances between Chloride concentrations

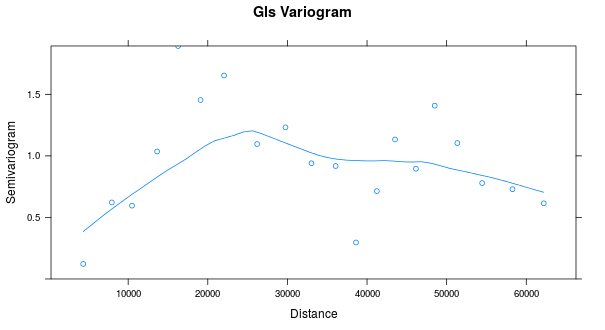


Figure : Basic GLS model variogram

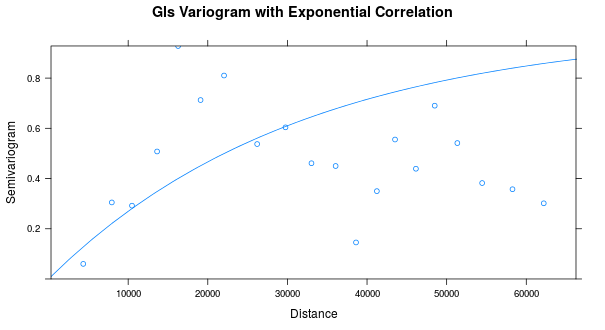


Figure : GLS model variogram with exponential spatial correlation component

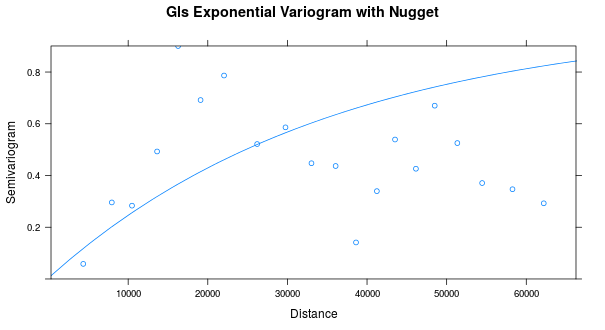


Figure : GLS model variogram with exponential spatial correlation component and nugget

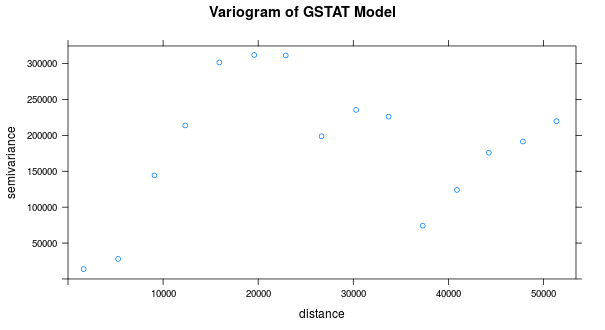


Figure : gstat variogram

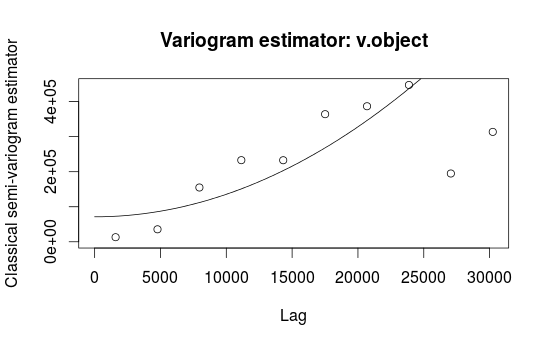


Figure : Variogram with periodicity

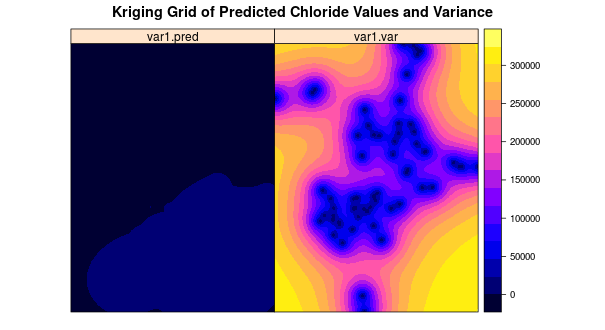


Figure 7: Kriging prediction grid for Chloride concentrations: right-variance of predicted values

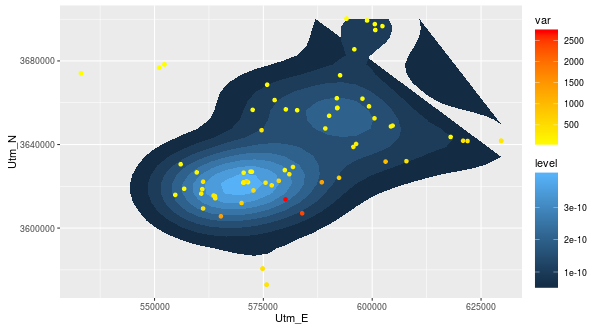


Figure 8: Density map of Chloride concentrations on the coordinate grid

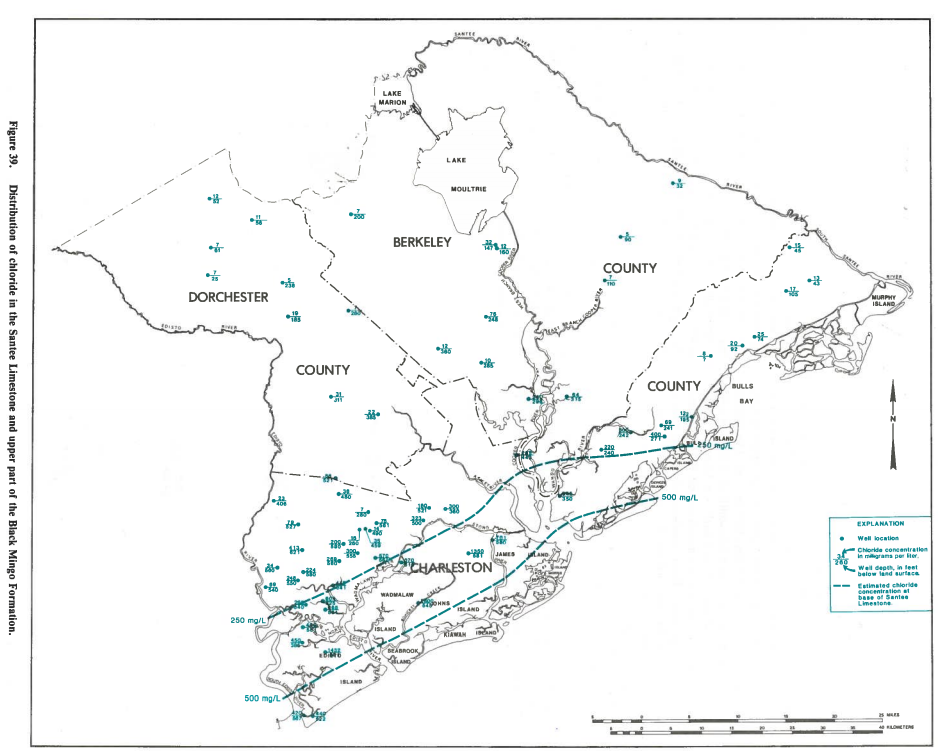


Figure 9: Distribution of chloride in Santee Limestone and upper part of Blank Mingo (Floridan) formation from Park, 1985

Table : Summary of basic GLS model

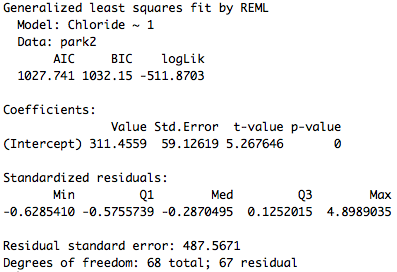


Table : Summary of GLS + Exponential Spatial Correlation

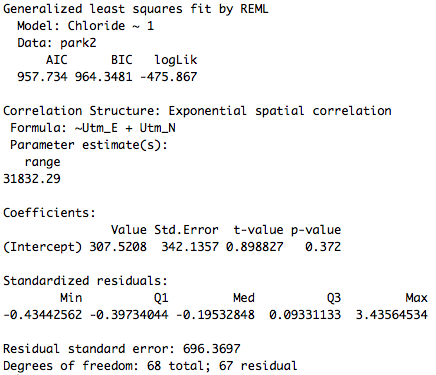


Table : Summary of GLS model + exponential spatial correlation + nugget

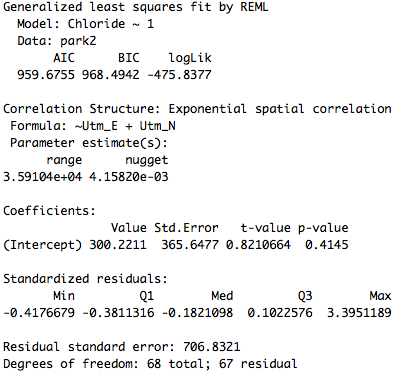
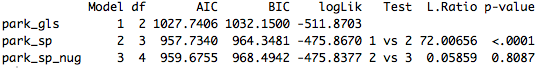


Table : ANOVA Comparison of Three GLS Models



**Literature Cited**

Barlow, P. M. 2003. Ground Water in Fresh Water-salt Water Environments of the Atlantic. Vol. 1262. U.S. Geological Survey.

Bear, J., Cheng, A. H. D., Sorek, S., Ouazar, D., & Herrera, I. (Eds.). 1999. Seawater intrusion in coastal aquifers: concepts, methods and practices (Vol. 14). Springer Science & Business Media.

Bivand, R. S., Pebesma, E. J., & Gomez-Rubio, V. (2008). Applied spatial data analysis with R. Springer.

Burnham, K. P.; Anderson, D. R. (2002), *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (2nd ed.), Springer-Verlag, [ISBN](https://en.wikipedia.org/wiki/International_Standard_Book_Number) [0-387-95364-7](https://en.wikipedia.org/wiki/Special:BookSources/0-387-95364-7).

Demirel, Z. 2004. The history and evaluation of saltwater intrusion into a coastal aquifer in Mersin, Turkey. *Journal of environmental management*,*70*(3), 275-282.

El Moujabber, M., Samra, B. B., Darwish, T., & Atallah, T. 2006. Comparison of different indicators for groundwater contamination by seawater intrusion on the Lebanese coast. *Water resources management*,*20*(2), 161-180.

Gelfand, A. E., Diggle, P., Guttorp, P., & Fuentes, M. (Eds.). (2010). *Handbook of spatial statistics*. CRC press.

Kariya, T., & Kurata, H. (2004). *Generalized least squares*. John Wiley & Sons.

Legendre, P. and Legendre, L. (2012) *Numerical Ecology*. 3rd English Edition. Elsevier.

Narayan, K. A., Schleeberger, C., Charlesworth, P. B., & Bistrow, K. L. 2003. Effects of groundwater pumping on saltwater intrusion in the lower Burdekin Delta, North Queensland. In *MODSIM 2003 International Congress on modelling and simulation* (Vol. 2, pp. 212-217).

Paniconi, C., Khlaifi, I., Lecca, G., Giacomelli, A., & Tarhouni, J. 2001. A modelling study of seawater intrusion in the Korba coastal plain, Tunisia. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*, *26*(4), 345-351.

Park, A. D. 1985. The ground-water resources of Charleston, Berkeley, and Dorchester Counties, South Carolina. South Carolina Water Resources Commission.

Sokal RR, Rohlf FJ (1995). *Biometry* (3rd ed.). New York: Freeman. pp. 813–819. [ISBN](https://en.wikipedia.org/wiki/International_Standard_Book_Number) [0-7167-2411-1](https://en.wikipedia.org/wiki/Special:BookSources/0-7167-2411-1).

Theis, C. V. 1938. The significance and nature of the cone of depression in ground-water bodies. *Economic Geology*, *33*(8): 889-902.