2017 Student Research Case Study Challenge
Society of Actuaries

Team Name: Cardinal Direction
Ball State University

April 7, 2017

Tim Hoblin, Cody Kocher, Alex Bolles, Jacob Eckert

Table of Contents

Executive Summary

In response to the Akua Coastal Commission's 2017 Coastal Act, a team of actuarial consultants was assembled to identify a land use plan for Akua's 20 undeveloped coastal zones. Our mission is to determine a development plan for each of the 20 undeveloped zones while adhering to the commission's rules, objectives, and mission and balancing competing interests of industry stakeholders.

Using data provided by the client, three major analyses were made: modeling sea level data, data imputation for missing sea level data, and ranking zones for development suitability.

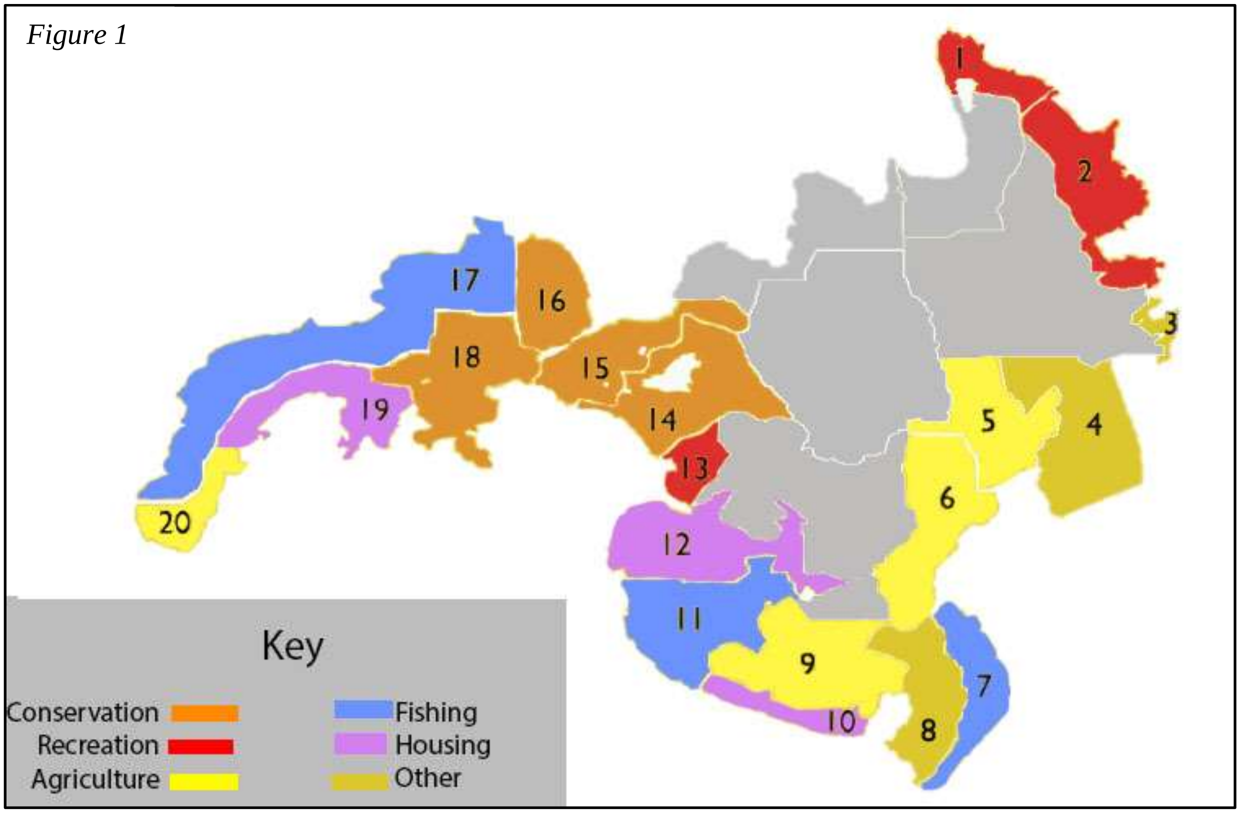
First, 5-year sea level projections of all 20 zones were calculated to determine if any zones were unsuitable for recreation or housing development due to flooding risks. After observing upward and seasonality trends in the sea level data, we applied a triple exponential smoothing model for our forecasts. The advantage of this model is it bases its projections on 3 different parameters involving general, seasonality, and annual trends.

Because much of the prior sea level data was missing, a data imputation method was implemented to complete the data. This allowed us to confidently apply our 5-year projections. We averaged multiple imputations to eliminate bias and account for outliers in any single imputation.

Having a better idea of each zone's flood risk, we could now begin to designate land use for each zone. To determine development suitability for each zone, we created two contrasting ranking systems of our own design. Using two systems allowed us to create different outlooks for zone designation; opening discussion of pros and cons for every designation.

Figure 1 shows our recommendation for the best land use plan for each of the 20 zones.

Figure 1



Data Modeling

Forecasting

To forecast water levels, a model needed to be established using the data provided. Regardless of the model implemented, we need to be sure to capture all relevant factors affecting sea level changes: overall global rising of tides, annual seasonal trends, and overall yearly rise (Sea Level Rise). Figure 2 shows a historic graph of Zone 1's water gauge levels from the last 10 years. After careful inspection, we noted recurring annual "W" shapes; large peaks occurring in the early spring and late fall and usually a smaller spike in mid-summer. Our model would need to produce forecasts that closely match these reoccurring shapes.

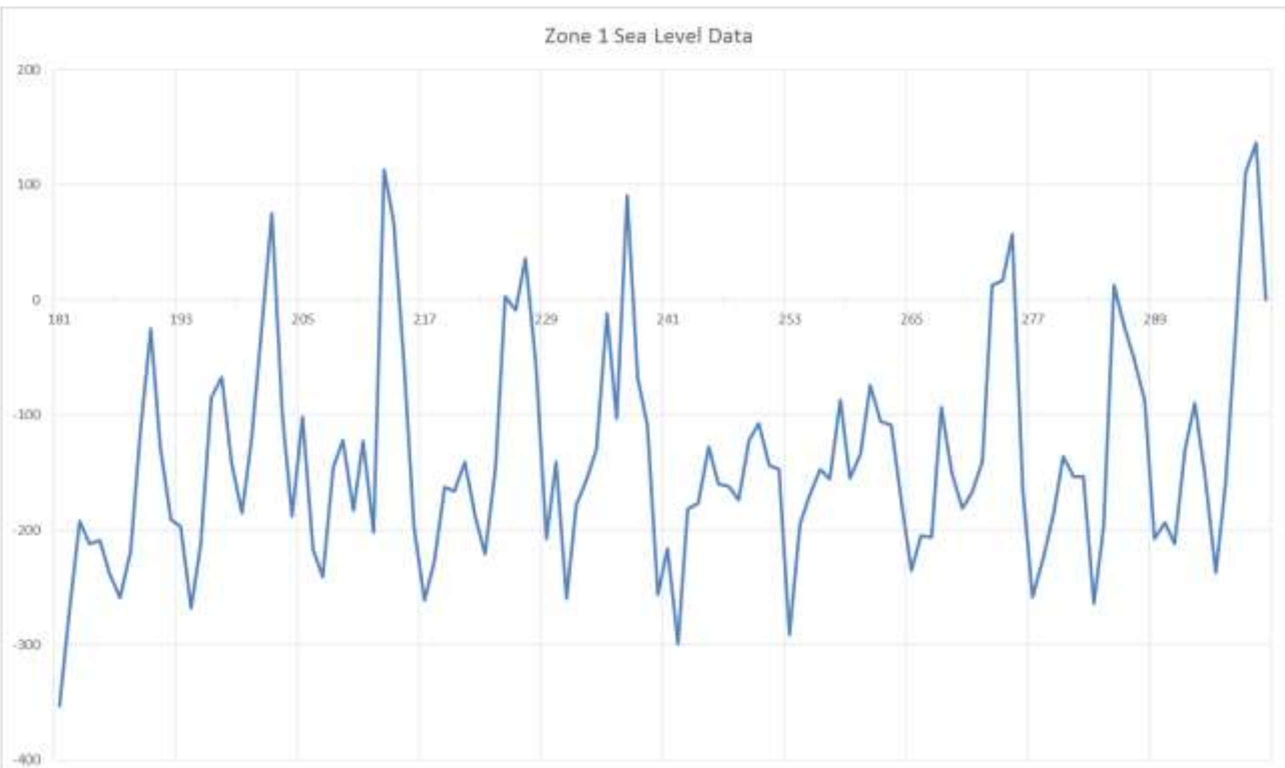


Figure 2: This visual representation of the last 10 years of Zone 1's sea level data shows distinct seasonality. The horizontal axis marks every twelfth month.

After trial and error, we found that Winters' Triple Exponential Smoothing model best fit our data. On its most basic level, exponential smoothing determines the future point using a weighted average of the point that came before and the point produced by the exponential smoothing algorithm. The point from the algorithm is computed from a weighted average of the data point and the algorithm produced point (Nau). This chain of points produces a geometric progression-like model. In Winters' Triple Exponential Smoothing model, we use three windows of exponential smoothing (Hyngman): one for seasonality trend, one for trends related to weather phenomenon related to water rise, and one for the annual trend. From fitting our model, we produce three parameters that we can use to forecast future values.

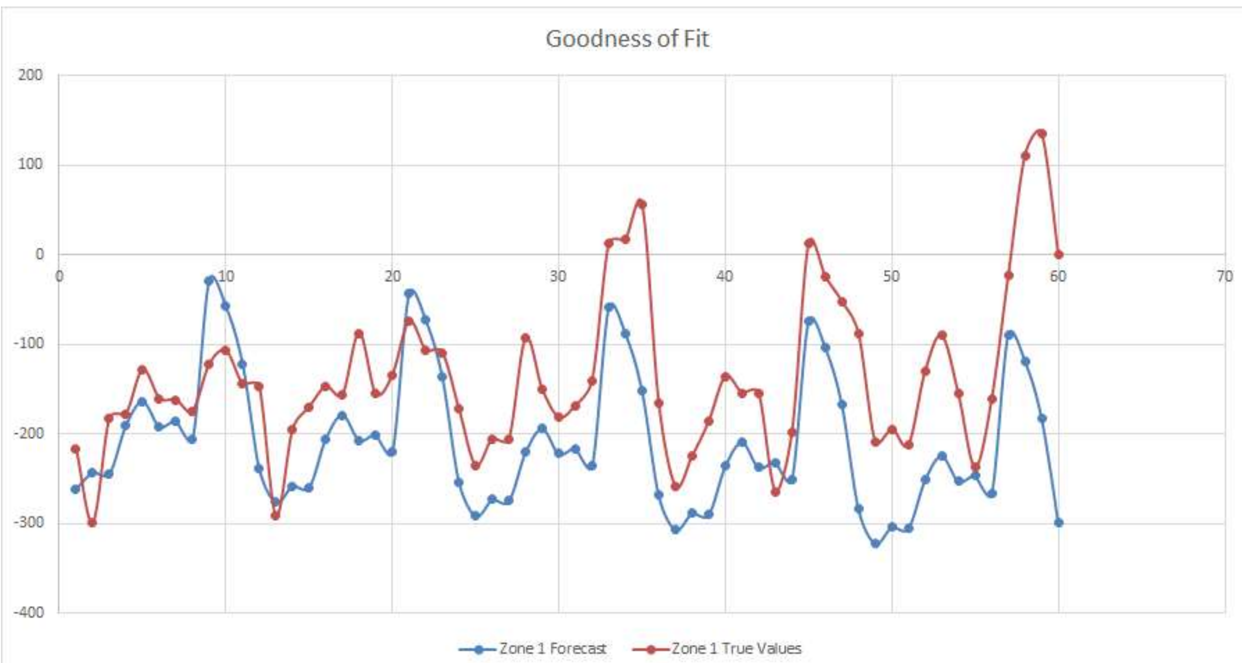


Figure 3: A 5-year forecast produced from a model created from January 1992 to December 2011 in Zone 1 compared to the given values from January 2012 to December 2016 in Zone 1. The projection is most accurate early on. As time advances the model begins to noticeably deviate from the true values. Because the forecast maintains shape, we relied on large confidence intervals in our predictions.

To test the goodness of fit of our model, we produced a forecast for a zone we already knew the values for. Figure 3 above shows a 5-year forecast produced from a model created from January 1992 to December 2011 in Zone 1 compared to the given values from January 2012 to

December 2016 in Zone 1. This shows Winters' model to be relevant to our data. Compared between different transformations and different models, the Winters' Triple Exponential Smoothing model produced the lowest variance. To ensure we make accurate assumptions, we computed 80% and 95% confidence intervals based on the models point estimates, pictured below. These confidence intervals were significant factors in making zone decisions.

Data Imputation

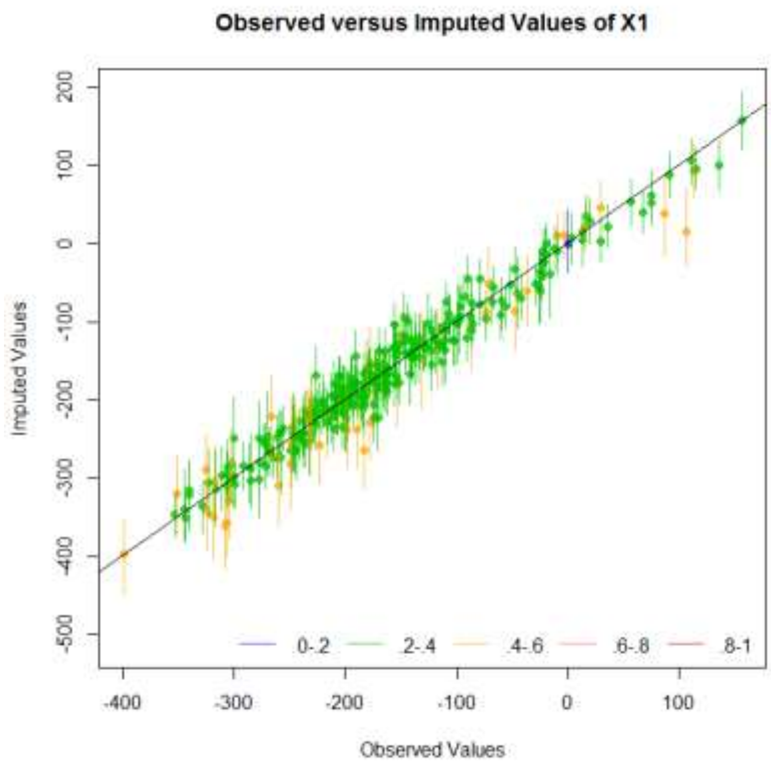
To correctly forecast future water levels for each zone, we needed to rely on the past water zone levels. For a few zones, like Zone 1, there is no issue with the data; however, multiple zones are missing readings. Rather than sacrificing accuracy in our forecasts, we explored methods of imputation to recover the missing data.

We assume our data is missing at random and our data as a whole follows multivariate normal properties. Further explanation of these assumptions can be found in the appendix.

Our chosen method of imputation is the Amelia II method. Because our data set exhibits multivariate normal properties, we built a model that utilized all zones present. Determining maximum likelihood estimators for every zone(Barbour), establishes parameters for the model. We apply this model multiple times to the months where data is missing and solve for the missing points. Rather than choosing one set of imputed values over another, we averaged the values together to avoid bias. Our team explored correlations between the number of imputations and variations in said imputations; running up to twenty-five imputations for every zone and analyzing their variability. We chose to run ten imputations for every zone; more imputations after that yielded similar results.

Although we met all the assumptions for our method of imputation, we were still curious if our imputations resembled actual values. To check our assumptions, we imputed over values we already had. Zone 1 made an ideal candidate because all data values are present. Figure 4 below shows a graph comparing actual values to those imputed. Colors indicate the percentage of values that fell outside a 90% confidence interval. For majority of readings, between 20% to 40% of imputation lie outside this interval. To mitigate this bias, we rely on our averaging method for all imputations.

Figure 4: A plot of observed values against imputed values. The colored lines represent a 90% confidence interval around the true observed value. This represents an ideal imputation diagnostic as the confidence intervals align with the black line indicating a perfect one to one ratio of observed versus imputed values.



Ranking

To examine development suitability for each zone, we invented two different ranking systems to help us allocate each zone. Using two systems helped us eliminate possible bias in any one system and gave us two different allocation suggestions to compare and contrast. The first system used a linear combination of the relevant characteristics for each possible zone use, applied scaling constants for each characteristic to ensure that each characteristic had an equal impact, and determined a final suitability score for each development type. The second looked at the relevant characteristics for each development type, ranked them in ascending order from 1 through 20 using Excel's RANK function, and averaged each feature's rank to get a zone suitability score for each type. Because some development types valued certain zones equally, we developed a method to compare zone suitability scores across development types. An example of this can be seen in Figure 5 below with the full rating systems being found in the Appendix. While many zones were clearly suited for one development type, others were not so. Keeping in mind our sea level forecasts and the rules and objectives set by the Coastal Act, we were able to compare the conclusions made by each system and come to our designation conclusions.

Rank Within Individual Development Types					
Coastal Zone	Conservation	Recreation	Agriculture	Housing	Fishing
15	5	5	14	13	7

Development Score					
Coastal Zone	Conservation	Recreation	Agriculture	Housing	Fishing
15	6.50	7.67	14.00	11.00	9.00

Rank Across All Development Types					
Coastal Zone	Conservation	Recreation	Agriculture	Housing	Fishing
15	16	28	79	55	36

Figure 5: Looking at the individual ranks for the development type, we noticed that Conservation and Recreation both valued Zone 15 as the 5th most suitable zone; more suitable than valued by the others. However, it can be seen looking at the Development Score table that this does not imply the same score for each development type. Ranking development types across all others for all zones gave us further insight to zone suitability.

Data Limitations

Absent zone sea level data was one of the limitations we faced. We struggled with not only deciding which method of imputation would be best, but also whether we could even compare zones to each other. Since individual zone readings were based off the change in the final reading on December of 2016, readings were not necessarily the same across all zones. Our method of imputation would have to be focus on the changes between zone readings rather than the individual numbers.

Another data limitation we faced was the average altitude measurement 100 meters inland from sea level. Considering altitude will vary throughout a zone, basing our projections solely off this average may not be completely indicative of a zone's flood risk. It may have been useful to have data about the minimum and maximum altitudes within each zone. With this information we would have been able to determine if there were any low-lying areas within a zone that could be at risk of flooding.

Designations

We designated Zones 14, 15, 16, and 18 for development as conservation areas. Conservation areas need to have significant Akua duck populations and sizable wetlands areas, especially wetland areas that are in danger of migration due to sea level rise. These zones fit these criteria very well. Our sea level forecasts show that these zones are particularly vulnerable to sea level rise above the current elevation measured 100 meters from the shore. These zones also contain 43% of the island's duck population which make them very important areas for environmental protection.

Zones 1, 2, and 13 were designated for development as recreation areas. These zones have a good mix of both grassland and forest areas, along with significant portions of coastline that we do not project to be in danger of sea level rise within the next five years. In particular, Zones 1 and 2 have a very even mix of grassland and forest areas along with 33.6 kilometers of coastline and will combine to form a very nice recreation area on the northeast side of the island. Zone 13 borders the conservation area formed in Zone 14 which would serve as a gateway to the rest of the conservation areas in that section of the island.

Zones 5, 6, 9, and 20 are designated for agriculture development. Agricultural development values large areas of grassland that can be converted to farmland and high levels of soil organic material to help crops grow. These four zones fit these characteristics very well, with all zones having at least 70% grassland and fairly high levels of soil organic matter. This further verifies the suitability of this land for healthy crops.

We designated zones 7, 11, and 17 for fishing development. Fishing development values large amounts of coastline for the development of fisheries along with low exploitation rates of existing snapper exploitation. The 62 kilometers of coastline designated for fishing zones allows considerable space for development fisheries. While Zone 7 may have a snapper exploitation rate that is greater than desired, this exploitation rate is largely counteracted by the very low exploitation rates seen in Zones 11 and 17.

Zones 10, 12, and 19 are designated for housing development. Our 5-year sea level forecasts do not indicate any danger of flooding in these zones, yet residents will still be able to enjoy plenty of coastline for beachfront development. Along with coastline, housing development also demands large amounts of grassland area. These zones contain 16% of the available coastline on the island and also have considerable amounts of grassland, which makes them very good candidates for housing development.

We designated Zones 3, 4, and 8 for use in economic development other than the five specific areas outlined by the council. They would have been marginal choices for use in any of the other development areas, so we decided instead to leave them for use for other development. While they do not fit well into any of the categories that we were given, they will be useful for some other form of development on the island.

Conclusion

After examining the data presented by the Coastal Commission, our team was able to designate each of twenty zones to fit the requirements for the Coastal Act for the six possible zone designations. We designated each zone based on two ranking systems. The ranking systems used a linear combination model and Excel's RANK formula. Each ranking system utilized the different characteristics required for zone designation. To complete the ranking process our team needed to fill the missing data in *Monthly Sea Level* data to allow us to create a five year sea level prediction. The missing data was filled using the Amelia II imputation method and the five year sea level prediction was created using the Winters' Triple Exponential smoothing. Lastly, our team was able to complete each designation with the five year sea level prediction. If the two ranking systems deemed multiple types of development, our team discussed the possible trade-offs for each zone upon final designations. Based on the information we were given, our team feels that these designations are the best use for each of the development zones on Akua Island.

Appendix

Below are the tables for the ranking system created with linear combinations. The tables are concerned with each zone's development score, their ranking within each development type, and their ranking against all development types respectively.

Development Score					
Coastal Zone	Conservation	Recreation	Agriculture	Fishing	Housing
1	-0.818	1.831	-1.133	-1.973	0.617
2	-0.818	2.373	-0.926	-1.073	1.431
3	-0.984	-0.764	-1.080	-1.845	-0.045
4	-0.071	-0.221	-0.677	-0.550	-0.867
5	-0.403	-1.383	0.293	-1.088	-1.264
6	-0.636	-0.296	1.128	-0.305	0.245
7	0.044	-0.566	-1.126	0.297	-0.294
8	-0.277	-0.229	-1.487	0.010	-0.774
9	-0.864	-1.145	0.111	-0.323	-0.709
10	-0.279	0.172	0.762	0.322	0.540
11	-0.568	0.891	0.071	0.507	-0.077
12	-0.458	-0.168	0.294	0.273	0.038
13	-0.488	0.241	-0.483	-0.109	-0.686
14	0.066	-0.445	-0.666	0.072	-1.412
15	0.538	-0.347	-0.652	0.398	-0.806
16	1.343	-0.934	-0.031	0.453	-0.762
17	2.987	1.391	1.760	2.381	2.142
18	1.786	-1.113	0.768	0.876	-0.198
19	0.291	0.063	1.220	1.091	1.758
20	-0.390	0.649	1.853	0.584	1.122

Rank Within Individual Development Types					
Coastal Zone	Conservation	Recreation	Agriculture	Fishing	Housing
1	17	2	19	20	5
2	17	1	16	17	3
3	20	16	17	19	9
4	8	10	15	16	18
5	12	20	8	18	19
6	16	12	4	14	7
7	7	15	18	9	12
8	9	11	20	12	16
9	19	19	9	15	14
10	10	7	6	8	6
11	15	4	10	5	10
12	13	9	7	10	8
13	14	6	12	13	13
14	6	14	14	11	20
15	4	13	13	7	17
16	3	17	11	6	15
17	1	3	2	1	1
18	2	18	5	3	11
19	5	8	3	2	2
20	11	5	1	4	4

Rank Within All Development Types					
Coastal Zone	Conservation	Recreation	Agriculture	Fishing	Housing
1	81	6	93	100	22
2	81	3	85	88	10
3	87	78	89	99	47
4	48	53	74	68	84
5	63	96	32	90	95
6	71	58	14	59	35
7	43	69	92	30	57
8	55	54	98	45	79
9	83	94	38	60	76
10	56	37	20	29	24
11	70	17	40	26	49
12	65	51	31	34	44
13	67	36	66	50	75
14	41	64	73	39	97
15	25	61	72	28	80
16	12	86	46	27	77
17	1	11	8	2	4
18	7	91	19	18	52
19	33	42	13	16	9
20	62	21	5	23	15

Below are the tables for the ranking system created by average ranks of zone features. The tables are concerned with each zone's development score, their ranking within each development type, and their ranking against all development types respectively.

Development Score					
Coastal Zone	Conservation	Recreation	Agriculture	Housing	Fishing
1	11.50	7.33	17.00	12.00	13.50
2	11.50	5.00	16.00	9.00	10.50
3	8.50	14.33	11.00	10.00	17.50
4	10.50	7.33	10.50	12.00	10.50
5	14.50	14.33	9.50	14.50	17.50
6	14.50	10.33	5.00	8.50	13.50
7	8.00	10.67	14.50	8.00	9.50
8	10.50	12.33	17.00	14.50	14.00
9	16.00	12.67	7.50	10.50	16.00
10	6.00	14.00	7.50	9.00	10.00
11	12.00	11.67	11.00	16.00	7.50
12	13.00	12.33	9.50	10.50	9.50
13	13.50	11.33	14.00	16.50	12.00
14	11.50	7.67	15.50	17.50	12.50
15	6.50	7.67	14.00	11.00	9.00
16	2.50	12.33	11.00	10.00	7.00
17	4.50	5.33	4.00	3.50	1.50
18	2.00	12.67	7.00	7.50	5.50
19	8.50	8.00	4.50	2.50	3.50
20	6.50	11.00	3.50	7.00	6.50

Rank Within Individual Development Types					
Coastal Zone	Conservation	Recreation	Agriculture	Housing	Fishing
1	12	3	19	14	15
2	12	1	18	7	11
3	8	19	11	9	19
4	10	3	10	14	11
5	18	19	8	16	19
6	18	8	4	6	15
7	7	9	16	5	8
8	10	13	19	16	17
9	20	16	6	11	18
10	4	18	6	7	10
11	15	12	11	18	6
12	16	13	8	11	8
13	17	11	14	19	13
14	12	5	17	20	14
15	5	5	14	13	7
16	2	13	11	9	5
17	3	2	2	2	1
18	1	16	5	4	3
19	8	7	3	1	2
20	5	10	1	3	4

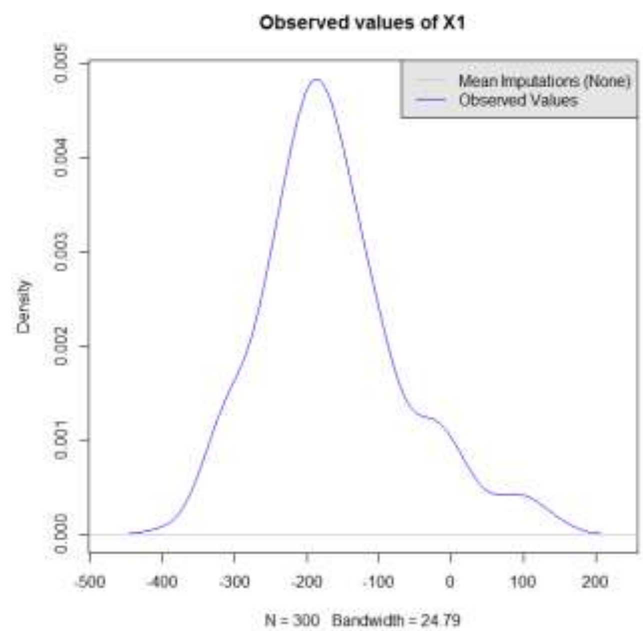
Rank Across All Development Types					
Coastal Zone	Conservation	Recreation	Agriculture	Housing	Fishing
1	61	22	96	65	76
2	61	11	91	36	47
3	33	83	55	43	98
4	47	22	47	65	47
5	85	83	39	85	98
6	85	46	11	33	76
7	30	54	85	30	39
8	47	69	96	85	79
9	91	73	24	47	91
10	15	79	24	36	43
11	65	64	55	91	24
12	75	69	39	47	39
13	76	60	79	95	65
14	61	28	90	98	72
15	16	28	79	55	36
16	3	69	55	43	19
17	9	13	8	5	1
18	2	73	19	24	14
19	33	30	9	3	5
20	16	55	5	19	16

Imputation Assumptions

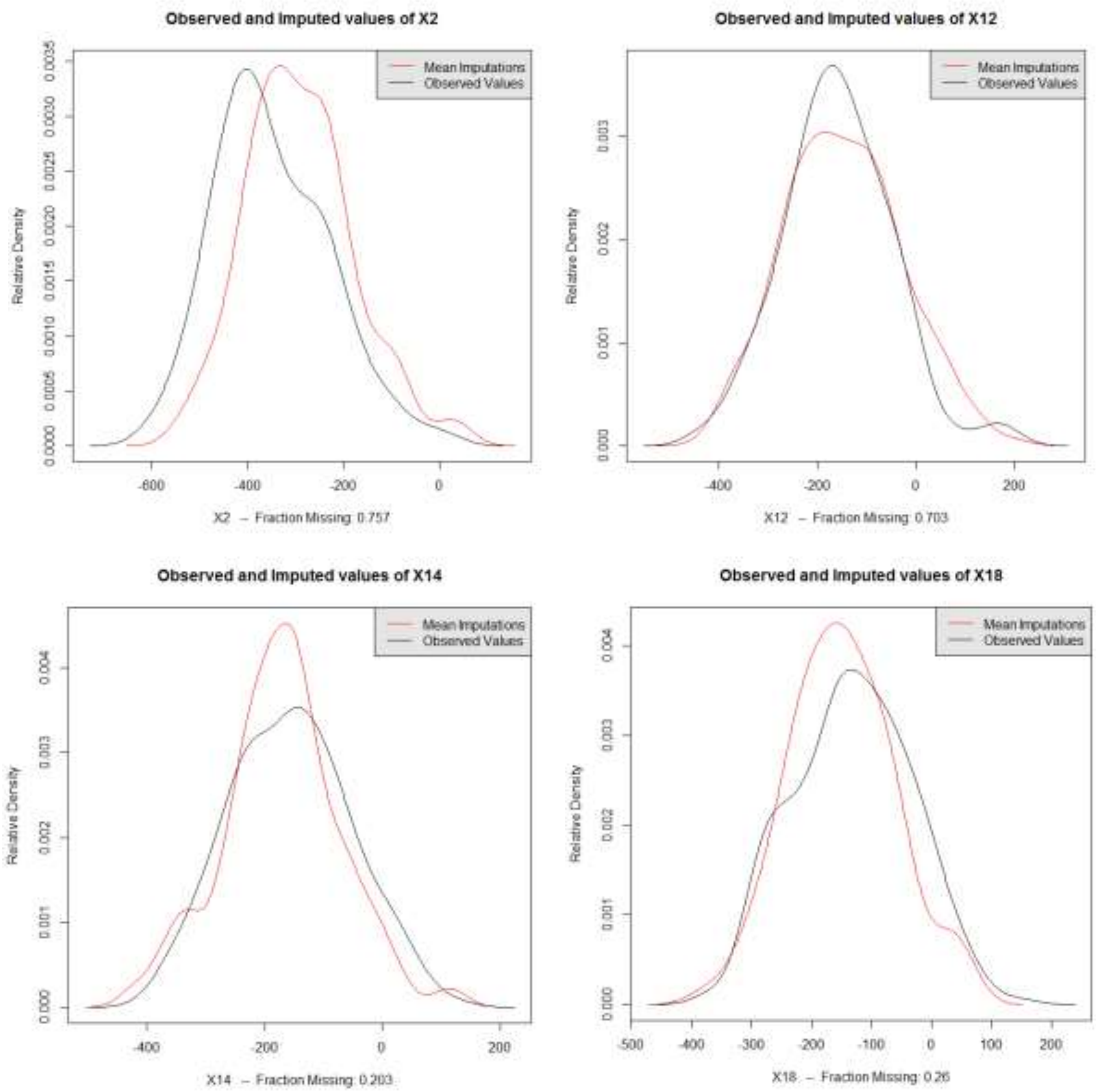
Whenever methods of imputation are used, we need to assume our data is missing at random (Honaker 4). Reflecting back on the data, one could argue that data is not missing at random in the individual zones themselves. If a water level gauge was not installed at that location one month, there is a high probability that it will remain uninstalled next month. On the contrary, if we assume that missing data is random between zones, this fits closer to our assumption. If one zone is missing data due to a missing or malfunctioning gauge, that has no effect on another zone's gauge reading.

Based on the method we chose, one other assumption was required: our complete data needed to be mathematically normally distributed across all zones (Honaker 4). Figure 6 below shows a density plot of the gauge readings from Zone 1. As the figure shows, our reading exhibits a normal bell-curve shape. Similar plots on the next page show the same reoccurring bell curve shape appearing in other zones. Zones' water gauge readings change in a similar manner; therefore, in our dataset can be characterized as a multivariate normal distribution.

Figure 6: This density plot observes the distribution of Zone 1 data points split into ranges of approximately 25 mm. Due to Zone 1's completeness, there are no imputed values represented in this plot.



The following graphs show bell shaped curves in some of the following imputed zone level data. We looked to make sure the relative heights and shapes of the curves were similar. This indicates the imputed values are good estimations for our data.



Attached below is code ran in R used for imputation and forecasting.

```
#Imputation
install.packages("Amelia")
install.packages("Zelig")
install.packages("Rcpp")
library(Zelig)
library(Rcpp)
library(Amelia)
AmeliaView()
#Load in .csv with all missing values replaced with NA
#Right Click Month Category as a time series and press “impute!” button

#Forecasts
install.packages("forecast")
library(forecast)

#Goodness of Fit for Zone 1
z1goodnessts <- ts(z1goodness, start = c(1992,1), end = c(2011, 12), frequency = 12)
z1goodnessfit <- HoltWinters(z1goodnessts)
show(z1goodnessfit)
forecast.HoltWinters(z1goodnessfit)
forecastz1goodness<- forecast(z1goodnessfit,60)
plot(z1goodnessfit)
plot(forecast(z1goodnessfit,60))
show(forecastz1goodness)

#Zone Forecasts
#ZONE 1
z1allts <- ts(z1all, start = c(1992,1), end = c(2016, 12), frequency = 12)
z1allfit <- HoltWinters(z1allts)
show(z1allfit)
```

```
forecast.HoltWinters(z1allfit)
forecastz1<- forecast(z1allfit,60)
plot(z1allfit)
plot(forecast(z1allfit,60))
show(forecastz1)
#ZONE 2
z2allts <- ts(z2ave, start = c(1992,1), end = c(2016, 12), frequency = 12)
show(z2allts)
plot(z2allts)
z2allfit <- HoltWinters(z2allts)
forecastz2<- forecast(z2allfit,60)
show(forecastz2)
plot(forecast(z2allfit,60))
#ZONE 3
z3allts <- ts(z3ave, start = c(1992,1), end = c(2016, 12), frequency = 12)
show(z3allts)
plot(z3allts)
z3allfit <- HoltWinters(z3allts)
forecastz3<- forecast(z3allfit,60)
show(forecastz3)
plot(forecast(z3allfit,60))
#ZONE6
z6allts <- ts(z6ave, start = c(1992,1), end = c(2016, 12), frequency = 12)
show(z6allts)
plot(z6allts)
z6allfit <- HoltWinters(z6allts)
show(z6allfit)
forecastz6<- forecast(z6allfit,60)
show(forecastz6)
plot(forecast(z6allfit,60))
#Zone 7
```

```
z7allts <- ts(z7ave, start = c(1992,1), end = c(2016, 12), frequency = 12)
show(z7allts)
plot(z7allts)
z7allfit <- HoltWinters(z7allts)
show (z7allfit)
forecastz7<- forecast(z7allfit,60)
show(forecastz7)
plot(forecast(z7allfit,60))
#Zone 9
z9allts <- ts(z9ave, start = c(1992,1), end = c(2016, 12), frequency = 12)
show(z9allts)
plot(z9allts)
z9allfit <- HoltWinters(z9allts)
show (z9allfit)
forecastz9<- forecast(z9allfit,60)
show(forecastz9)
plot(forecast(z9allfit,60))
#ZONE10
z10allts <- ts(z10ave, start = c(1992,1), end = c(2016, 12), frequency = 12)
show(z10allts)
plot(z10allts)
z10allfit <- HoltWinters(z10allts)
show(z10allfit)
forecastz10<- forecast(z10allfit,60)
show(forecastz10)
plot(forecast(z10allfit,60))
#zone13
z13allts <- ts(z13ave, start = c(1992,1), end = c(2016, 12), frequency = 12)
show(z13allts)
plot(z13allts)
z13allfit <- HoltWinters(z13allts)
```

```
show(z13allfit)
forecastz13<- forecast(z13allfit,60)
show(forecastz13)
plot(forecast(z13allfit,60))
#zone12
z12allts <- ts(z12ave, start = c(1992,1), end = c(2016, 12), frequency = 12)
show(z12allts)
plot(z12allts)
z12allfit <- HoltWinters(z12allts)
show(z12allfit)
forecastz12<- forecast(z12allfit,60)
show(forecastz12)
plot(forecast(z12allfit,60))
#zone14
z14allts <- ts(z14ave, start = c(1992,1), end = c(2016, 12), frequency = 12)
show(z14allts)
plot(z14allts)
z14allfit <- HoltWinters(z10allts)
show(z14allfit)
forecastz14<- forecast(z14allfit,60)
show(forecastz14)
plot(forecast(z14allfit,60))
#zone15
z15allts <- ts(z15ave, start = c(1992,1), end = c(2016, 12), frequency = 12)
show(z15allts)
plot(z15allts)
z15allfit <- HoltWinters(z15allts)
show(z15allfit)
forecastz15<- forecast(z15allfit,60)
show(forecastz15)
plot(forecast(z15allfit,60))
```

```
#zone17
z17allts <- ts(z17ave, start = c(1992,1), end = c(2016, 12), frequency = 12)
show(z17allts)
plot(z17allts)
z17allfit <- HoltWinters(z17allts)
show(z17allfit)
forecastz17<- forecast(z17allfit,60)
show(forecastz17)
plot(forecast(z17allfit,60))
#zone18
z18allts <- ts(z18ave, start = c(1992,1), end = c(2016, 12), frequency = 12)
show(z18allts)
plot(z18allts)
z18allfit <- HoltWinters(z18allts)
show(z18allfit)
forecastz18<- forecast(z18allfit,60)
show(forecastz18)
plot(forecast(z18allfit,60))
#ZONE19
z19allts <- ts(z19ave, start = c(1992,1), end = c(2016, 12), frequency = 12)
show(z19allts)
plot(z19allts)
z19allfit <- HoltWinters(z19allts)
forecastz19<- forecast(z19allfit,60)
show(forecastz19)
plot(forecast(z19allfit,60))
#zone 20
z20allts <- ts(z20ave, start = c(1992,1), end = c(2016, 12), frequency = 12)
show(z20allts)
plot(z20allts)
z20allfit <- HoltWinters(z20allts)
```

```
show(z20allfit)
forecastz20<- forecast(z20allfit,60)
show(forecastz20)
plot(forecast(z20allfit,60))
```

Bibliography

Barbour, R., Ph.D (Writer). (2013, May 31). Making Statistics Accessible - Multiple Imputation As An Approach to Missing Data [Video file]. Retrieved March 30, 2017, from https://youtu.be/xqQ1bLLe3B4

Honaker, J., King, G., & Blackwell, M. (2011). Amelia II: A program for missing data. Journal of statistical software, 45(7), 1-47.

Hyndman, R. J., & Athanasopoulos, G. (2014). Forecasting: Principles and Practive. OTexts.

Nau, R. (2017). Moving average and exponential smoothing models. People.duke.edu. Retrieved 30 March 2017, from http://people.duke.edu/~rnau/411avg.htm

Sea Level Rise. (2017). Nationalgeographic.com. Retrieved 30 March 2017, from http://www.nationalgeographic.com/environment/global-warming/sea-level-rise/