# Diagnosis of Rainfall from Buoy Data

Matt Chmelewski[1], Joe Brown[1], Frederick Bingham[1], and Oksana Chkrebtii[2]

[1]UNC - Wilmington
[2]Ohio State University Dept. of Statistics

THE OHIO STATE UNIVERSITY

# Introduction and Research Questions

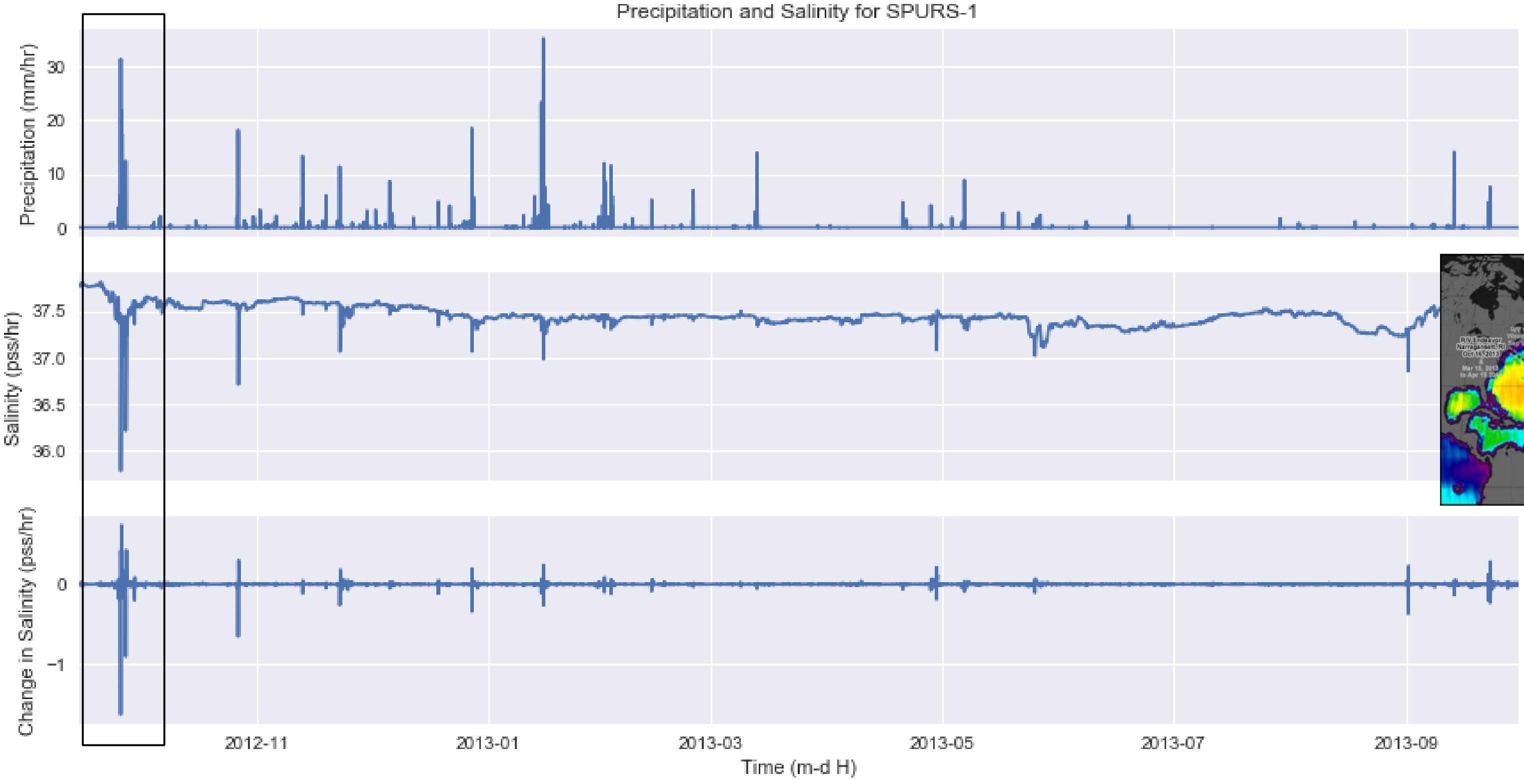
One motivating force behind satellite measurement of sea surface salinity (SSS) is the possibility of using it as a rain gauge (Lagerloef et al, 2008)

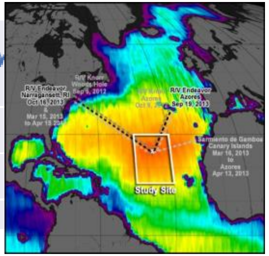
We explore this possibility here by attempting to use SSS, along with other variables to estimate values of rainfall, or rain/no-rain status

The variables being used are SSS, sea surface temperature (SST), precipitation, and wind speed collected by the central mooring during the SPURS-1 field campaign

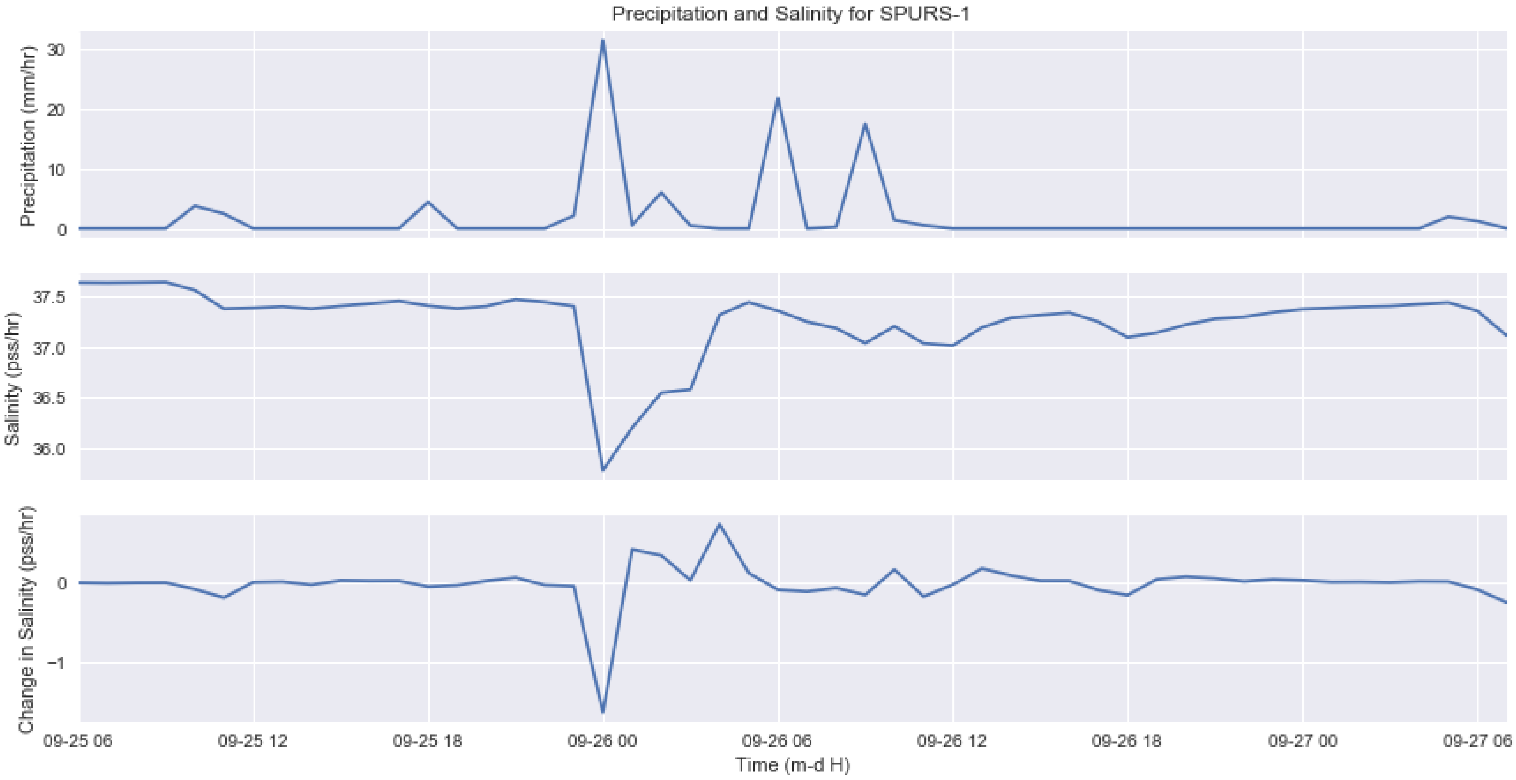
Supply et al. (2018) are doing some similar work, but on a larger scale

# Precipitation, Salinity, and Salinity Change for SPURS-1



# Precipitation, Salinity, and Salinity Change for SPURS-1



# Rain / no-rain

In order to classify precipitation as a binary function, we need to establish a threshold to separate rain events from non-rain events

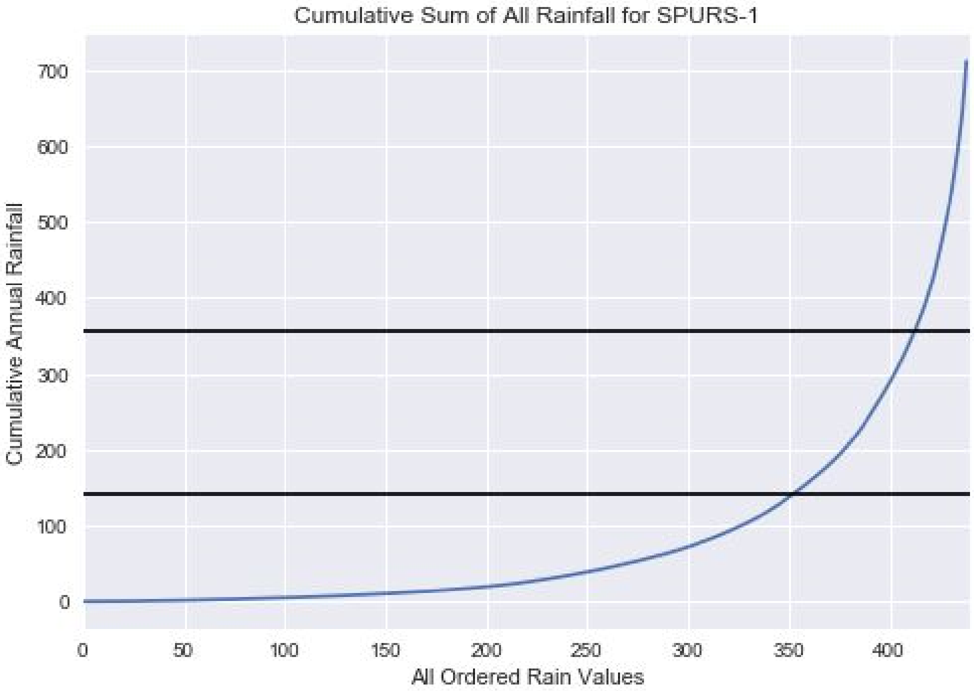
What rain rate do we need to be able to estimate?

# Methods

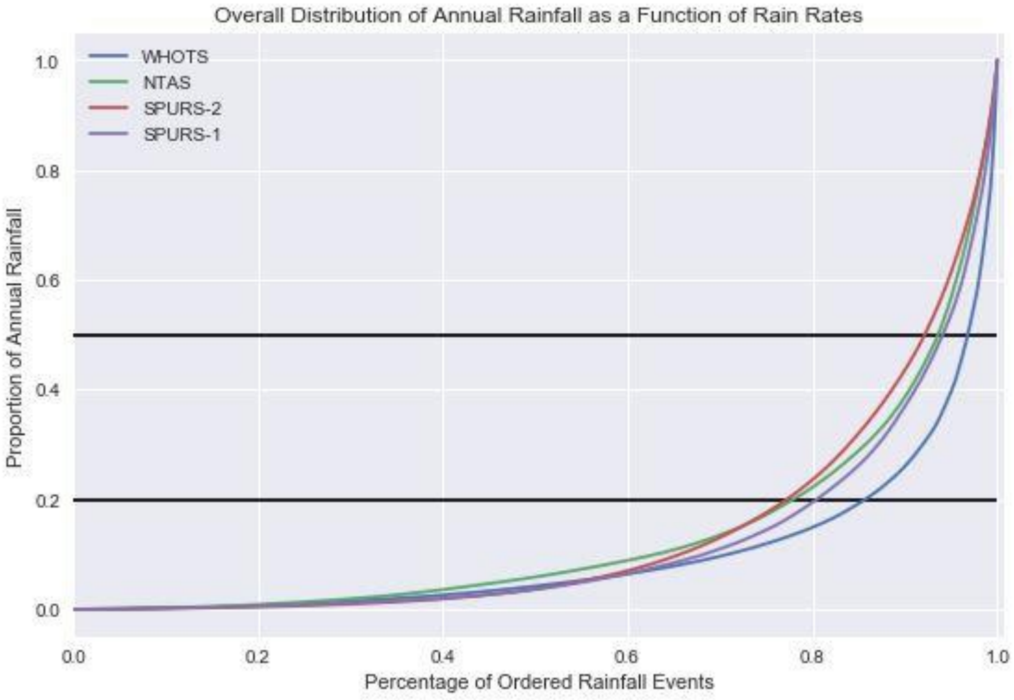
- Categorized rain events by presence of precipitation and ordered rainfall events
- Distribution of cumulative rainfall
- Univariate Correlation
- Selected Predictor variables

# Distribution of Annual Rainfall from SPURS-1

```
First Quantile of Annual Rainfall 0.1 mm/hr
 Median of Annual Rainfall Values 0.4 mm/hr
Third Quantile of Annual Rainfall 1.3 mm/hr
```



# Similar Rainfall Distributions from Other Buoys




```
Percentage of data that makes up 50% of SPURS-1 Annual Rainfall: 5.92 %
Percentage of data that makes up 50% of SPURS-2 Annual Rainfall: 7.86 %
Percentage of data that makes up 50% of NTAS Annual Rainfall: 6.45 %
Percentage of data that makes up 50% of WHOTS Annual Rainfall: 3.24 %
```

# Correlations

| | Surface Salinity | Surface Temp. | Wind Speed | 3 Hr Precip. | 6 Hr Precip. | Salinity Change | Wind Speed Change | Precipitation |
|---|---|---|---|---|---|---|---|---|
| **Salinity Change** | 0.125786 | 0.004978 | 0.017742 | -0.191359 | -0.022051 | 1.000000 | 0.066348 | -0.498071 |
| **Surface Salinity** | 1.000000 | 0.411407 | 0.128199 | -0.205460 | -0.189729 | 0.125786 | 0.002357 | -0.146640 |
| **Wind Speed Change** | 0.002357 | 0.008255 | 0.146354 | -0.010562 | 0.013694 | 0.066348 | 1.000000 | -0.082643 |
| **Surface Temp.** | 0.411407 | 1.000000 | 0.005372 | -0.010484 | -0.011724 | 0.004978 | 0.008255 | -0.006427 |
| **Wind Speed** | 0.128199 | 0.005372 | 1.000000 | 0.019348 | 0.048086 | 0.017742 | 0.146354 | 0.006914 |
| **6 Hr Precip.** | -0.189729 | -0.011724 | 0.048086 | 0.774195 | 1.000000 | -0.022051 | 0.013694 | 0.479483 |
| **3 Hr Precip.** | -0.205460 | -0.010484 | 0.019348 | 1.000000 | 0.774195 | -0.191359 | -0.010562 | 0.662495 |
| **Precipitation** | -0.146640 | -0.006427 | 0.006914 | 0.662495 | 0.479483 | -0.498071 | -0.082643 | 1.000000 |

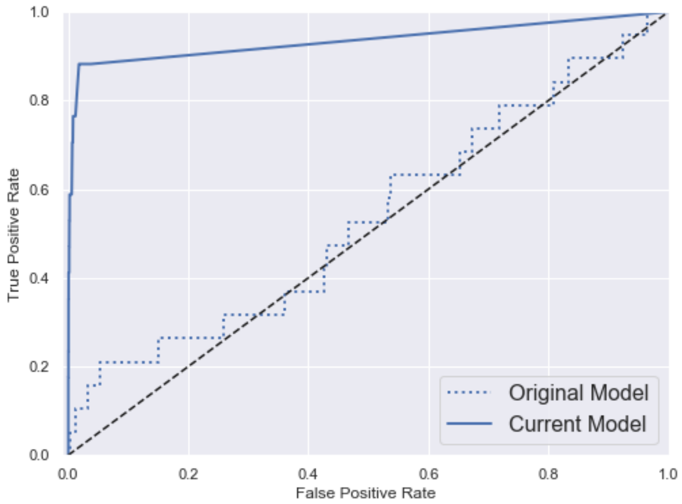
# Initial Exploratory Analysis

**Predictor Variable Selection**

- The optimal threshold for SPURS-1 was chosen to be the first precipitation rate that contributed to the greater 50% of the cumulative annual sum (6.4 mm/hr)
- Goal was to make a model of precipitation events as a function of an optimal number of variables to have maximum diagnostic ability and avoid overfitting
- Used Recursive Feature Elimination (RFE)
- Useful variables: wind speed, sea surface temperature, and the difference in salinity with respect to the previous hour

# Recursive Feature Elimination

- A form of backwards selection for models
- Fits logistic regression model on precipitation as a function of all available predictor variables
- Eliminates the variable that was least significant and repeats the process until a desired number of variables is reached




ROC curve

# Results and Conclusions

Cumulative rainfall follows similar exponential-type curve across a multitude of climates and seasons

- The greater 50% of every buoy’s cumulative annual sum falls in fewer than 8% of the rain events

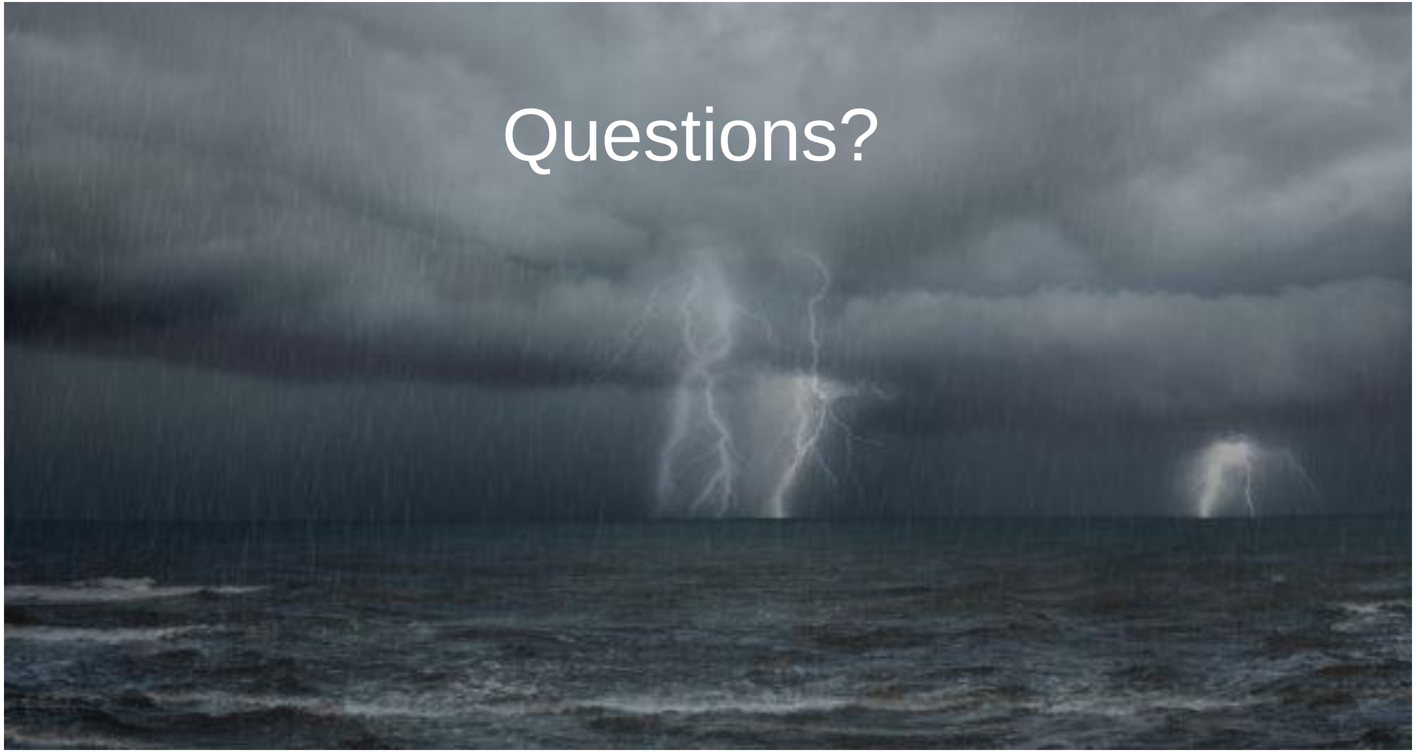
Our chosen model predicted all non-rain events but still struggled with false positives

- One difficulty in understanding the errors lies in the fact that salinity change, wind speed and temperature all vary across the set of false positives

# Going Forward

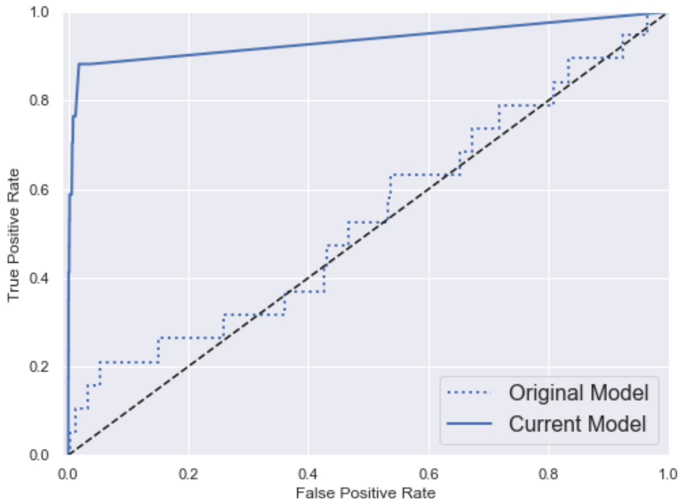
There are several potential directions for the future:

- Evaluating whether the threshold we have set (the first value that makes up the greater 50% of rainfall events) is actually the optimal one, or if a different/multiple thresholds should be created
- Splitting the data along the threshold and testing for a difference in means for salinity/salinity change or other factors (Tukey’s test)
- Introducing machine learning algorithms, such as tree-based models
    - These models have the power to handle large data set with higher dimensionality
    - They can be used for both classification and regression problems


Questions?

# Recursive Feature Elimination

- A form of backwards selection for models
- Fits logistic regression model on precipitation as a function of all available predictor variables
- Eliminates the variable that was least significant and repeats the process until a desired number of variables is reached




| | Actual Negative | Actual Positive |
|---|---|---|
| **Predicted Negative** | 2733 | 0 |
| **Predicted Positive** | 4 | 3 |