

Anomaly Detection in Time Series Data

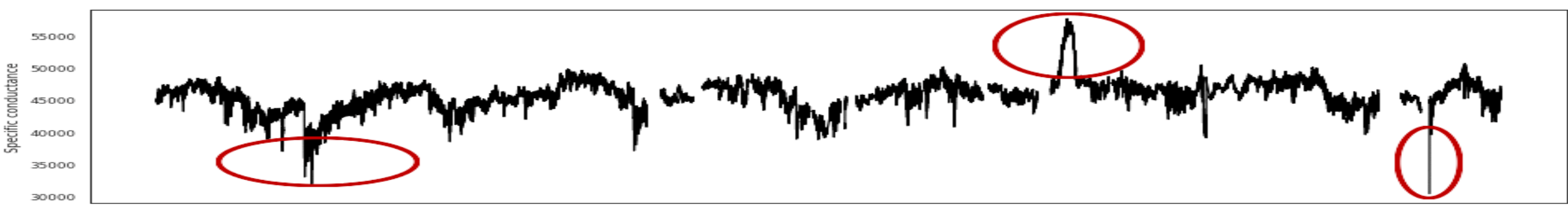
Applied Study in Water Quality Variables

Reem Elmahdi¹; Willie Brink¹; Josefine Wilms²
¹Department of Mathematical Sciences, Stellenbosch University, ²CSIR



Time Series Data & Anomaly Detection

Time Series Data (TSD): is a set of observations indexed in time order. Time adds more information to the data but it also adds complexity.



Specific conductance time series measurements and possible anomalies.

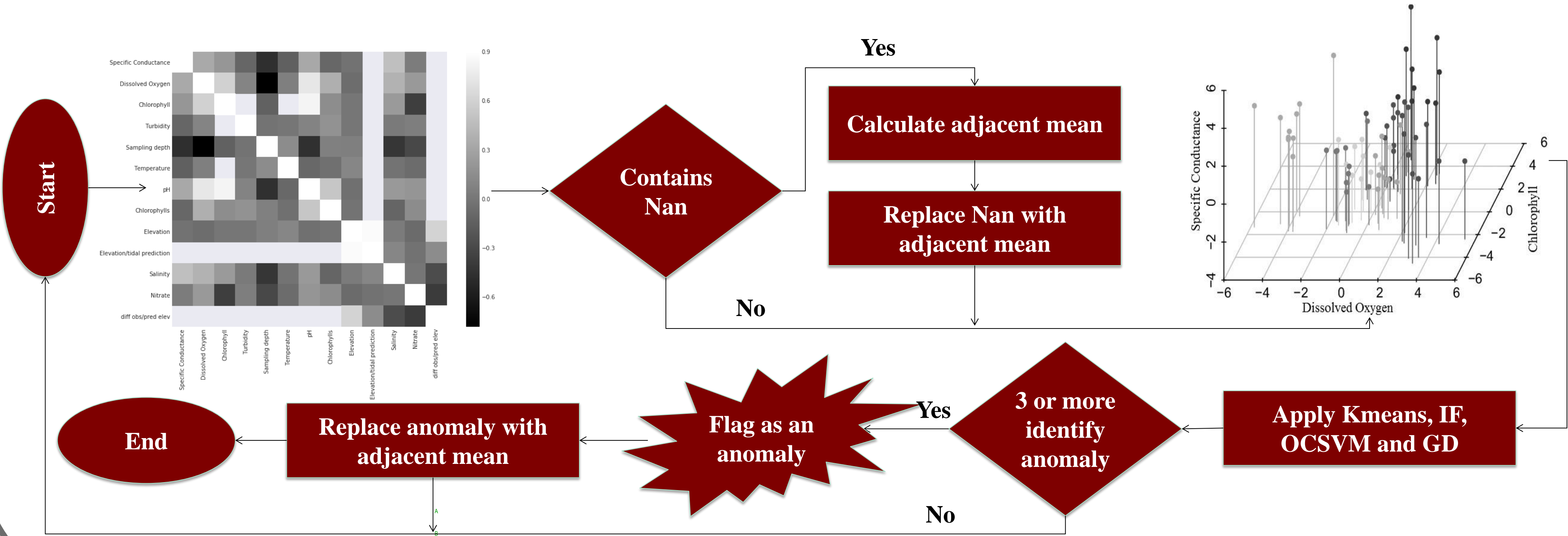
Anomaly Detection: is the process of identifying or classifying deviating observations from the normal ones.

Problem: How to identify observation x_t that deviates from a time series data $X=\{x_1, x_2, \dots, x_n\}$ behaviour, where n is the number of observations; Does removing anomalies increase the accuracy of the prediction models?

Applications: medical diagnosis, network intrusion detection systems, stock market analysis, financial posts, earth science, etc.

Dataset & Methodology

Water quality variables measurements was obtained the United States Geological Survey (USGS) website. Hog island measurements were selected for this experiment. Measurements include specific conductance (SC), dissolved oxygen (DO) and chlorophyll (CL). The data is available in six minute time intervals and Ranges from 2010 to 2018. This data is the input of the models.

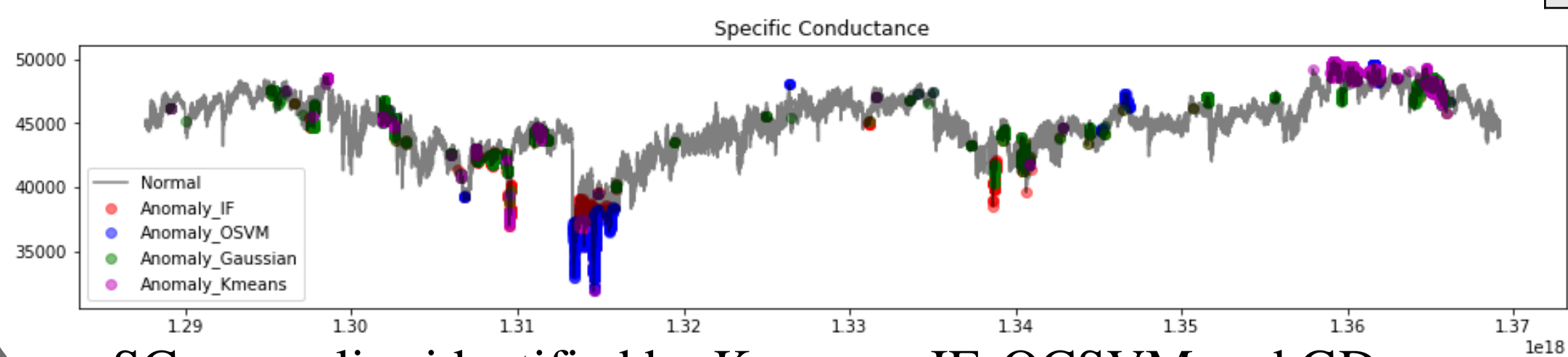


Anomaly Detection Methods & Results

K means, Isolation Forests (IF), One Class Support Vector Machine (OCSVM) and Gaussian Distribution (GD) were used to detect anomalies.

Multi-Layer Perceptron (MLP) was used in prediction process, observations was reduced to 24/day, 18 observations where used to predict the following 6 ones. MLP has 2 layers (10:5), used Relu activation function and Adam solver.

| Var | Mean | Obs | Missin g% | MLP accuracy (EVS) | | |
|-----|-------|------|--------------|--------------------|----------------|----------------------|
| | | | | Deleted nan | Filled gaps | Removed anomalies |
| SC | 45037 | 5148 | 2.3 | 59.7 | 58.4 | 63.5 |
| DO | 8.85 | 5238 | 0.6 | 88 | 88.7 | 87.1 |
| CL | 13.7 | 4404 | 16.4 | 50.2 | 54.8 | 59.8 |



SC anomalies identified by K means, IF, OCSVM and GD.

Explained Variance Score (EVS) value after applying MLP.

Scan to Get More Results!

