RESEARCH IN DATA DRIVEN MATHEMATICAL MODELING

Author: Miles Kent
Advisors: Dr. Russell Milne, Dr. Hao Wang

January 5, 2024

THIS REPORT WAS SUBMITTED TO THE DEPARTMENT OF MATHEMATICAL AND STATISTICAL SCIENCES IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR MATH 497: READING IN MATHEMATICS

Abstract

Data Driven Modeling and Analysis of Methane Emissions From Mildred Lake
The mining of bitumen from the Athabasca oil sands makes up a large portion of Alberta's economy. During the separation process, trace amounts of diluent and bitumen are lost and released into tailings and end ponds. The interaction of these hydrocarbons with local bacteria in the ponds causes a large amount of methane, which is more than twenty-eight times as powerful as carbon dioxide at trapping heat, to be released into the atmosphere. In an attempt to predict methane emissions and understand what environmental factors influence methane concentration in the air from these ponds, we fit non-linear regression models and decision trees to a multi variable data set from the Mildred Lake air quality monitoring station. Our results are varied. We find that a decision tree has extremely low accuracy if we use the unmodified data set as training data. On the other hand, if we account for a base concentration of volatile methane, a small degree of measurement error, and bin our data accordingly, a decision tree can predict volatile methane concentration from Mildred Lake with over 90% accuracy. We also find using regression analysis that the five most important variables for predicting volatile methane concentration from Mildred Lake are the standard deviation of wind speed, relative humidity, air temperature, wind direction, and wind speed. Lastly, we compare five regression models but find that none of our models are very strong, signified by low R^2 scores and high AIC scores.

Data Driven Modeling and Prediction of Heatwaves and Cold Snaps
Periods of extreme temperature, such as heatwaves and cold snaps, can wreak havoc on infrastructure, agriculture, local ecosystems, and human life. Today's most up to date climate models and methods of weather forecasting are fairly accurate when it comes to predicting local weather several days in advance. The drawback is that these methods and models are very computationally and financially expensive to implement properly and are not always well suited to predict these extreme temperature events in advance. Viewing local climate through the lens of dynamical systems, we hypothesise that heatwaves and cold snaps can be described as bifurcations, and therefore should be preceded by early warning signals that can be used to predict these events. We test our hypothesis using the classifier from Bury et al. 2021 to see if the early warning signals preceding fold, transcritical, and Hopf bifurcations can be used to identify heatwaves and cold snaps in a large three hundred city local weather data set. We find that our classifier is able to identify heatwaves and cold snaps in the temperature data as mostly as fold and transcritical bifurcations. The drawback is that our model picks up many other temperature spikes and drops that are not classified as heat-weaves or cold snaps, limiting our model accuracy. Overall, our results act as a solid stepping stone to optimize of our model and modify of our methods for better results

Acknowledgements

I would like to first and for most thank my advisors. Thank you Russell for your constant feedback and encouragement throughout my time as your student. Thank you Hao, for consistently opening up doors and creating opportunities for me to continue contributing to ground breaking and meaningful research within the ILMEE. The experience and knowledge I have been able to gain from both the opportunities and feedback provided by you guys has been incredibly valuable. I am extremely grateful for it.

I would also like to thank by lab mates Binbing, Pablo, Shohel, Ilhem, and Liubov for the advice, feedback, conversations, and research productivity. You all are incredible researchers and it has been a pleasure to work with you all on these projects while I have been apart of the ILMEE.

Lastly, I would like to thank my parents Marilyn and Adam and siblings Chloe and Caleb. Thank you for always being there for me in my times of need and for the comical and wholesome phone calls. Without you guys, I would not have had the courage to go back to school to pursue my intrest.

Contents

DATA DRIVEN MODELING AND ANALYSIS OF METHANE EMISSIONS FROM MILDRED LAKE

1 Overview

The mining of bitumen from the Athabasca oil sands makes up a large portion of Alberta's economy, with the oil and gas sector making up over 21% of Alberta's annual GDP [6]. The extraction on bitumen takes place within the Athabasca oil (tar) sands [11] which involves excavating the sands and separating the bitumen from these sands with hot water and hydrocarbon biased diluent [5]. During this separation process, trace amounts of diluent and bitumen are lost and released into tailings and end ponds [5]. The interaction of these hydrocarbons with local bacteria in the ponds causes a large amount of methane, which is more than twenty-eight times as powerful as carbon dioxide at trapping heat, to be released into the atmosphere [3].

With climate change becoming a significant issue and Canadian legislation enforcing net carbon neutrality by 2050 [7], these emissions are very problematic and must be understood in order to be negated. We present a case study where we use air quality and weather data from Mildred Lake to try to understand what environmental factors influence methane emissions from an end pond. In addition, we also try to use machine learning to predict how much methane is emitted from Mildred Lake under certain environmental conditions. To complete these task we implement non parametric statistical test and regression analysis to understand what environmental conditions influence methane emissions and classification trees as a first attempt to build a machine learning model that can accurately predict emissions.

Our results are varied. We find using regression analysis and a non parametric statistical test that the five most important variables for predicting volatile methane concentration from Mildred Lake are the standard deviation of wind speed, relative humidity, air temperature, wind direction, and wind speed. We build five regression models using this knowledge and compare them, but find that all of our models are similar and non of them are very strong. This is signified by low R^2 scores and high AIC scores. Our predictive modeling results are notably stronger. We find that a classification tree has extremely low accuracy if the unmodified data set is used as training data. On the other hand, if we account for a base concentration of volatile methane, a small degree of measurement error, and bin our data accordingly, a classification tree can predict volatile methane concentration from Mildred Lake with over 90% accuracy.

Though we do make headway and build a solid foundation for future research with our initial attempts, we need to refine our models and methods to produce better results. This includes using regression trees instead of a classification trees for our machine learning modeling and using interaction terms and lower order terms in our regression analysis. If these changes are made, our analysis and predictive modeling will be much more successful and accurate.

2 Data

2.1 Generation and Filtering

Air quality monitoring data is taken from the Mildred Lake Air Quality Monitoring Station provided by the Alberta Data Warehouse (https://www.alberta.ca/alberta-air-data-warehouse). We assume that all measurements are accurate due to the level two status of the data, which means that it has gone through quality control checks implemented by the providing source. Data from the Mildred Lake Air Quality Monitoring Station is supplemented with data from the Mildred Lake Weather Station which has been provided by the Government of Alberta's historical climate data depository (https://acis.alberta.ca/acis/weather-data-viewer.jsp). The Mildred Lake Air Quality Monitoring Station is located about 1.5km from the Mildred Lake Weather Station, so we assume that what we supplement will maintain a high degree of accuracy due to the close proximity of the stations. The supplementing process is as follows. Both data set rows are matched by date and time. We start our filtering process by discarding all rows in the combined data set where there are no or invalid methane measurements. Both stations measure some of the same variables so we supplement the missing independent variable entries from the Mildred Lake Air Quality Monitoring Station with the Mildred Lake Weather Station data for these corresponding independent variables if it exist. This supplementary presses is not done for any repeated wind parameters that both stations measure as these parameters change over the distance of 1.51km significantly due to landscape features which can lead to inaccuracy. Once we have supplemented what we can, we drop all repeated independent variables (columns) in our combined data set, keeping the columns from the Mildred Lake Air Quality quality Monitoring Station. Finally, we drop any rows where there is not an entry for all variables within the row. This leaves us with our clean data set

2.2 Generation of the Ang_To_Stat_deg Variable

We generate a standardised variable, Ang_To_Stat_deg, that measures the the wind direction with zero being the direction from the center of Mildred Lake to the Mildred Lake Air Quality Monitoring Station. Ang_To_Stat_deg $\in (-180, 180]$, where the value of Ang_To_Stat_deg increases as you change direction from left to right. The calculation of this variable requires a transformation of the original wind direction variable entries in our data set which is done by the following steps.

1. First calculate amount of degrees that north needs to be shifted using

$$\theta = \arctan\left(\frac{lat1 - lat2}{long1 - 1long2}\right) - 90^\circ \tag{1}$$

   where θ is the amount of degrees that north needs to be shifted, $lat1$ and $long1$ are the coordinates of the Mildred Lake Air Quality Monitoring Station, $lat2$ and $long2$ are the coordinates of the center of Mildred lake, and the 90° shift account for the fact that zero degrees mathematically is directed exactly east on a standard map. Using the coordinates lat1 = 51.04978°, long2 = -111.5638°and lat2 = 57.052502°, long2 = -111.588333°, we get θ = -96.33634°.

2. Transform all wind direction values from [0°, 360°] to (-180°180°]. This is done my taking all coordinates that are greater then 180°and subtracting 360°from them.

3. Subtract θ from all wind directions calculated in step 2. This is equivalent to adding θ because our calculated θ is negative.

4. Finally, for any wind direction value calculated in step 3 that is greater then 180°, subtract 360°from it.

These new calculated wind direction values can be appended to there own column which is what we have done for Ang_To_Stat_deg.

2.3 Variables and Basic Properties

Our final data contains 24937 hourly measurements. These measurements are within the date range of November 26, 2019 to February 7, 2023. Our data set contains the following parameters which represent the environment around the Mildred Lake Air Quality Monitoring Station and the local weather around Mildred Lake.

- **Methane_ppm**: The dependent variable. This variable measures the average amount of methane detected in the air in parts per million over an hour period.
- **Outdoor_Air_Temp_C**: Measurements of the average Temperature in Celsius over an hour period
- **Relative_Humidity_Per**: Measurement of average humidity in percent over an hour period. $[0, 100]$
- **Wind_Direction_deg**: Measurement of the average wind direction in degrees relative to geographic north over an hour period. $[0, 360)$
- **Wind_Direction_Std_Dev**: The standard deviation of the wind direction over the hour long measurement period. This variable signifies how often the wind direction changes over an hour.
- **Wind_Speed_kmhr**: The measurement of the average wind speed in kilometers per hour over an hour period.
- **Wind_Speed_Std_Dev**: The standard deviation of the speed direction over the hour long measurement period. This variable signifies how often the wind speed changes during the hour measurement period.
- **Date_Time**; Signifies the date and time of the measurements in the format yyyy-mm-dd hh:mm:ss.
- **Precip_Amount_mm**: The measurement of the amount of precipitation that has fallen over the hour measurement period in millimeters.
- **Dist_To_Stat_km**: Is the distance to the station from the center of Mildred Lake in kilometers. Since this is a single station data set, all rows have a value of 1.51km.
- **Ang_To_Stat_deg**: Is the wind direction in degrees relative from the center of Mildred Lake to the Mildred Lake Air Quality Monitoring Station. $(-180, 180]$

3 Analysis of Monthly Emission Averages

3.1 Kruskal-Wallis Test

Due to the sample size and the non normality of our distribution of methane measurements (Figure 1), we decide to use a Kruskal-Wallis Test to analyse if there is a difference in emissions by month. The

Kruskal-Wallis Test is a non parametric statistical test used to see if there is a difference in the average value of a parameter between populations in a data set with non normal distributed populations [11]. The assumptions of this test are; all samples are random, all samples are independent of each other, all populations are relatively similar in shape (though not absolutely necessary), and the sample size of each group contains five or more individual measurements [11]. The null hypothesis, H_o, for this test is that there is not a difference in the average value of the parameter being tested between populations [11]. The alternative hypothesis, H_a, is that there is a difference in the average value of the parameter being tested between populations [11]. The way that this test is conducted is we first choose a significance level, preferably 0.05 or less. Next, assign each of our k samples a rank from lowest to highest and calculate the H critical value using

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(n+1) \tag{2}$$

where n is the total number of observations, R_j denotes the ranks for the sample data of group j, and n_j denotes the sum of the ranks for the sample data [11]. The H critical value follow a chi squared distribution with degrees of freedom equal to $k-1$, so we can use a chi squared table to identify the p value of our test [11]. If the p value is less then our significance level, we reject the null hypothesis [11]. On the other hand, if the p value greater then the significance level we choose, we cannot reject the null hypothesis [10].

3.2 Applying the Kruskal-Wallis Test

We conduct the test at a 0.05 significance level on our data using R 4.0.5. We obtain H = 4193.5 with 11 degree of freedom and a p value equal to $2.2 * 10^{-16}$. Since our p value is less then our significance level, we can reject H_o and conclude that there is a difference in average methane emission by month. With analysis of Figure 1, we see that emission is lower over the summer and higher over the winter.

4 Regression Analysis of Environmental Variables

We analyse the significance of each independent variable on methane emissions using regression analysis. We decided to use polynomial regressions to analyse this biased off of its ease of implementation and its flexibility. A polynomial regression can be defined as a function that estimates the value of dependent variable taking into account a set of independents. The slopes assigned to the regression model parameters are estimated by fitting the regression model to a data set using least squares [12]. The form of a polynomial regression with no interaction terms is

$$f(x_1, ..., x_n) = \sum_{i=0}^{k} \beta_{1,i} x_1^i + ... + \sum_{j=0}^{l} \beta_{n,j} x_n^j \tag{3}$$

where each β is the slope estimate for the each model parameter dictated by least squares, n signifies the number of independent variables in our regression (and often our data set), i and j is the number of degrees we decide to include our regression for each variable, and each x_n is an independent variable. The significance of each term is dictated by its β term, where we can use a basic statistical t-test to determine the terms significance [12].

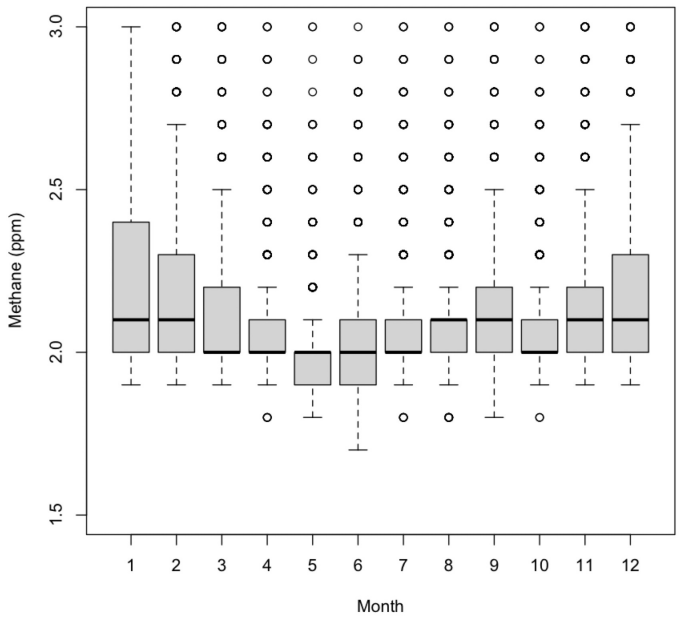


Figure 1: Methane Concentration (ppm) vs Month of Measurement

4.1 Building Regression Models; A Bottom Up Approach

We build our regression models using the steps as follows on our data-set. All test and parameters are calculated with R 4.0.5.

1. Transform each independent variable in our data set up to an order of ten. For each transformation, generate a new column for each variable transformation and add it to our data set.
2. Fit a simple linear regression to each independent variable (column) in the set and calculate the AIC score of the fitted model to determine variable significance. Once this is done, rank each model from lowest to highest where the most significant model has the lowest AIC score.
3. Calculate the Pearson's correlation coefficient between the variables/model and remove all highly correlated variables.
4. Fit full multivariate regression.
5. Remove less significant predictors if they exist

We decided to keep the ten most significant variables in our full model. This leaves our full model to be

$$f(x_{ws}, x_{rh}, x_T, x_a, x_{ssd}) = \beta_0 + \beta_1 x_{ssd} + \beta_2 x_{rh} + \beta_3 x_T + \beta_4 x_a + \beta_5 x_T^2 + \beta_6 x_a^2 + \beta_7 x_{ssd}^5 + \beta_8 x_{ws}^7 + \beta_6 x_T^7 + \beta_{10} x_a^7 \tag{4}$$

where x_{ssd} is the standard deviation of wind direction, x_{rh} represents relative humidity in percent, x_a represents the wind direction in degrees relative to the air quality monitoring station from the center of Mildred Lake, x_T is the air temperature in Celsius, and x_{ws} is the wind speed in kilometers per hours. From this we derive our additional models to compare which are

$$f(x_{ws}, x_{rh}, x_T, x_a, x_{ssd}) = \beta_0 + \beta_1 x_{ssd} + \beta_2 x_{rh} + \beta_3 x_T + \beta_4 x_a + \beta_8 x_{ws}^7 \tag{5}$$

$$f(x_{rh}, x_T, x_a, x_{ssd}) = \beta_0 + \beta_1 x_{ssd} + \beta_2 x_{rh} + \beta_3 x_T + \beta_4 x_a + \beta_6 x_a^2 + \beta_7 x_{ssd}^5 + \beta_{10} x_a^7 \tag{6}$$

$$f(x_{ws}, x_{rh}, x_T, x_a, x_{ssd}) = \beta_0 + \beta_1 x_{ssd} + \beta_2 x_{rh} + \beta_3 x_T + \beta_4 x_a + \beta_5 x_T^2 + \beta_6 x_a^2 + \beta_7 x_{ssd}^5 + \beta_8 x_{ws}^7 \tag{7}$$

$$f(x_{ws}, x_{rh}, x_T, x_a) = \beta_0 + \beta_2 x_{rh} + \beta_3 x_T + \beta_4 x_a + \beta_5 x_T^2 + \beta_6 x_a^2 + \beta_8 x_{ws}^7 + \beta_6 x_T^7 + \beta_{10} x_a^7 \tag{8}$$

where (5) is the model containing one of each individual parameter, (6) is the model containing parameters where each β has a significance less then 0.001, (7) is a model containing two or less of each parameter, and (8) is a model containing non calculated parameters (no standard deviation).

4.2 Model Assessment

We compare the models using AIC score, VIF score, RMSE, R^2 score, and adjusted R^2 score (Table 1). Below is a brief summary of how each one is calculated and how it helps us assess our models.

- **AIC**: The Akaike Information Criterion, or AIC, helps us determine which is the best model for a set of data when comparing models. It allows us to assess how well a model fits the data by taking into account the number of parameters in the model to prevent over fitting [15]. The lower the AIC score, the better the model is [15]. The AIC score of a model can be calculated with the following formula

  $$AIC = 2K - 2ln(L) \tag{9}$$

  where K is the number of parameters in a model and L is the log likelihood estimate of the model [15].

- **RMSE**: The Root Mean Square Error (RMSE) measures the difference between the true values of a data set and the predicted values of the model (residuals). It is a useful tool to assess how well a model fits a data set without accounting for the number of parameters in a model. The formula used to calculate (RMSE) is

  $$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(yest_i - yact_i)^2}{N - P}} \tag{10}$$

  where *yest* is the estimated parameter of the model for observation i, *yobs* is the observed value for observation i, N is the number of observations, and P is the number of parameter estimates, including the constant.

- **VIF** The Variance Inflation Factor (VIF) is a tool used to understand the degree of co-linearity of a models parameters. The smaller the VIF, the less co-linear the variables of a model are [8]. A models parameter is considered to have an acceptably low amount co-linearity if the parameter VIF is less then five [8]. The VIF can be calculated with the following formula

  $$VIF_i = \frac{1}{1 - R_i^2} \tag{11}$$

  where R_i^2 is the adjusted coefficient of determination for variable i in a regression model [8].

- R^2: The Coefficient Of Determination, or R^2 statistic, tells us how well the variation in the independent variables explains variation in the dependent variable. It can be used to dictate how well a model explains patterns in the data BUT not how well a model fits the data [12]. The R^2 ranges between zero and one. An R^2 score of one means that variation of the dependents is fully explained by the values of the independent while and R^2 score of zero states the contrary [12]. This means the higher the R^2 score, the better [12]. The R^2 score can be calculated by the following formula

$$R^2 = 1 - \frac{RSS}{TSS} \tag{12}$$

where RSS is the sum of squares of the model residuals and TSS is the total sum of squares of the model [12].

- **Adjusted** R^2: This can simply be described as the Coefficient Of Determination for higher order models. The adjusted R^2 statistic takes into account the number of model parameters and only raises its value if additional parameters within model are significant as apposed to the basic R^2 statistic, which will increase with the addition of parameters. This can be calculated with the following formula

$$R^2_{adj} = 1 - \frac{(1 - R^2)(N - 1)}{N - p - 1} \tag{13}$$

where R^2 is the un-adjusted coefficient of determination, R^2_{adj} is the adjusted coefficient of determination, N is the total sample seize, and p is the number of predictors [12].

4.3 Results and Analysis

We calculated all parameters for each of our models and we ended up and we get the following values (Table 1).

Model	AIC	RMSE	MAX Variable VIF	R^2	Adjusted R^2
3	8548.43	0.29	2.39	0.16	0.16
4	9905.85	0.29	1.84	0.15	0.15
5	8580.29	0.29	1.72	0.16	0.16
6	8548.43	0.29	2.39	0.16	0.16
7	9299.33	0.29	1.64	0.13	0.13

Table 1: Model statistics

All statistics indicate that (6) and (3) are the strongest models, but not by much as all of them have very similar statistical scores. Very little of the variation of the dependent parameter in each model is explained by the independent parameters, which is signified by a low R^2 and adjusted R^2 scores. This means that no model is very precise in predicting dependent variable values from the independents. Our models do have low co-linearity between independent variables, with the max variable VIF amongst all models being 2.39 which is less then 5. To make our models more effective, we need to do the following. First. we need to include interactions between the model parameters as methane transmission is directly effected by the interaction of several parameters in our data set. For example, humidity and sunlight together affect the transmission of methane as high humidity and high amounts of sunlight lead to the dissipation of volatile methane. Second, we need to include negative exponential terms in the regression model in addition to

the positive terms to account for negative relationships in the data. Finally, we don't need as high of an order of terms in our models as degree of ten is very steep and may not accurately represent the data. This especially may be true if we also include negative and interacting terms. Our regression model should represent data much better if these modifications are included, which will give us more significant results.

5 Predicting Emissions Using Classification Trees

A Classification Tree is a machine learning model that labels, records, and assigns variables into discrete classes [9]. These classes are build by recursive partitioning, which is an iterative process of splitting the data into partition and then splitting it up further on each of its branches [9]. The easiest way to build a model using this method is to feed an algorithm a set of training data, such as the CART algorithm, and then test it with a test data set.

5.1 Model Assessment

A classification tree can be simply assessed using a confusion matrix. A confusion matrix is a $n * n$ matrix that signifies whether a machine learning model has correctly classified an event [9]. The columns of the confusion matrix signify the correct classification of the event while the rows signify the classification of the model. This means that every entry on the matrix diagonal is a correct classification on an event while the entries not on the diagonal are incorrect classifications [9]. The accuracy of the model can then be dictated by

$$acc(A) = \frac{\sum_{i=1}^{n} a_{i,i}}{\sum_{j=1}^{n}(\sum_{i=1}^{n} a_{j,i})} \tag{14}$$

where $a_{j,i}$ and $a_{i,i}$ are elements of confusion matrix A [9].

5.2 Experimental Groups

We decided to test a classification trees effectiveness in predicting emissions by splitting up our data into six groups (Table 2). For each binned group (not RAW), we take into account that the air has a base concentration of methane [14]. To account for this, we consider every measurement less then 2.1ppm to be a base concentration and bin accordingly. For the raw data, the bins are of width 0.1ppm as this is the maximum precision of the measurements from the Mildred Lake Air Quality Monitoring Station.

Group	Test/Training Split	Bin Width
1	1/9	0.7ppm
2	1/6	0.7ppm
3	1/9	0.4ppm
4	1/6	0.4ppm
5	1/9	RAW
6	1/6	RAW

Table 2: Experimental setup

We train a total of six models for each group using ctree in R 4.0.5. where very the amount of training data the bin size fed to the model. This was to test models ability to make predictions with limited amount of training data as well as the precision and accuracy that the model can make predictions at. The overall goal was to generate a model with a very highest degree of accuracy that can operate effectively with limited amounts of training data.

5.3 Results and Analysis

We test each model on their respective test data sets and we get the following accuracy for each group (Table 3). We see that the models fit to the binned groups are significantly more accurate in predicting

Group	Model Accuracy
1	99.8%
2	98.6%
3	91.8%
4	92.1%
5	$> 0.1\%$
6	$> 0.1\%$

Table 3: Model accuracy

emissions within there assigned groups then the models trained with the raw data. The classification tree models maintain more then 90% accuracy if the bin width is at least 0.4ppm. We also see that the results are very similar for each training and test data split for each bin group. This shows that our model is fairly robust and can be fed less then 90% of the data to be accurate. We can conclude that if we take into account a base concentration methane in the air and bin our data with bin width 0.4ppm or more, our model will be able to make predictions with +90% accuracy.

When looking at the confusion matrices for the binned groups, we also see that even when our model miss-classifies an estimate, the model is not making "out of pocket" predictions. Every miss-classified prediction is within one bin of the the actual measurement (Table 4-7). We also see that the model is 100% accurate in predicting base concentration cases as well (Table 4-7). This means that there is a definite environmental parameter difference between measurements of base methane concentration and measurements of increased methane concentration. Though our binned models are highly accurate, we can most likely improve model accuracy with use of regression tree models instead of decision tree models. This is because a regression tree model assumes the output variable is continuous which allows for uncertainty in the prediction. Even though our outputs can be coincided discrete due to the 0.1ppm methane measurement precision of the Mildred Lake Air Quality Monitoring Station, all miss-classified predictions are close to their actual bins which means that if we do account for a small degree of uncertainty, model accuracy with most likely be greatly improved.

G1	A	B	C	D
A	2048	0	0	0
B	0	1889	0	0
C	0	48	100	0
D	0	0	10	25

Table 4: Group 1 confusion matrix

G2	A	B	C	D
A	1399	0	0	0
B	0	1251	0	0
C	0	27	70	0
D	0	0	7	16

Table 5: Group 2 confusion matrix

G3	A	B	C	D	E	F	G
A	2084	0	0	0	0	0	0
B	0	1652	0	0	0	0	0
C	0	237	0	48	0	0	0
D	0	0	0	78	0	0	0
E	0	0	0	22	0	10	0
F	0	0	0	0	0	14	0
G	0	0	0	0	0	11	0

Table 6: Group 3 confusion matrix

G4	A	B	C	D	E	F	G
A	1339	0	0	0	0	0	0
B	0	1076	0	0	0	0	0
C	0	175	0	27	0	0	0
D	0	0	0	59	0	0	0
E	0	0	0	11	0	7	0
F	0	0	0	0	0	9	0
G	0	0	0	0	0	7	0

Table 7: Group 4 confusion matrix

DATA DRIVEN MODELING AND PREDICTION OF HEATWAVES AND COLD SNAPS

6 Overview

Periods of extreme temperature, such as heatwaves and cold snaps, can wreak havoc on infrastructure, agriculture, local ecosystems, and human life [12]. Today's most up to date climate models and methods of weather forecasting are fairly accurate when it comes to predicting local weather several days in advance. The drawback is that these methods and models are very computationally and financially expensive to implement properly as they require huge numerical simulations, systems of partial differential equations, and satellite imaging [10]. In addition to this, these methods are not always well suited to predict these extreme temperature events due to their often rapid onset.

On the other hand, measuring climate parameters from ground biased observation stations is significantly cheaper and more widespread then satellite measurements, as many airports and cities have measurements stations [10]. This grounds biased observation station data is often publicly available and easy to access online. In addition to this, machine learning offers a computationally cheaper way to interpret and find patterns in large amounts of data quickly and efficiently. Viewing local climate through the lens of dynamical systems, we hypothesise that heatwaves and cold snaps can be described as bifurcations, and therefore should be preceded by early warning signals that can be used to predict these events [1].

We test our hypothesis using the LSTM classifier from Bury et al, 2021 [1] to see if the early warning signals preceding fold, transcritical, and Hopf bifurcations can be used to identify heatwaves and cold snaps in ground biased observation station temperature data from a large three hundred city data set. To elaborate, the classifier is tested on 500 point temperature data time series where we predict the transition at point 400 and point 450. We find that our classifier is able to identify heatwaves and cold snaps in the temperature data, mostly as fold and transcritical bifurcations. The drawback is that our model picks up many other temperature spikes and drops that are not classified as heat-weaves or cold snaps, limiting our model accuracy.

Overall these results are promising as they provide a stepping stone for optimising of our model and modifying of our methods for significantly higher success. This includes testing our model on 600 point time series, where we predict the bifurcation at point 500 with a single variable and multivariate data set. It also includes figuring out what other temperature transitions our LSTM model is identifying in our data as our classifier may actually be just as well suited to predict these other events as well. The end goal is to develop a model and method that can predict heatwaves and cold snaps in advance with an extremely high degree of accuracy.

7 Definitions

To identify the heatwaves and cold snaps in our data, we must define what heat waves and cold snap are, and we must must calculate the climate norms for each of our cities so these extreme temperature events can be identified.

7.1 Average Temperature Norm and the Warm and Cold Seasons

We define the average temperature norm for a geographical region as the average of all daily average temperature measurements. It can be calculated by

$$TN_{adv} = \frac{1}{N_{days}} \sum_{n=1}^{N_{days}} T_n \tag{15}$$

where TN_{adv} is the average temperature norm for a geographical location, T_n is the daily average temperature for day n, and N_{days} is the number of days being used for our average temperature norm calculation. To calculate the warm and cold season of a geographical region, we must calculate the average temperature norm for a geographical region by month. The warm season for a geographical region is defined as the five months with the highest Average Temperature Norm for that region. The cold season for a geographical region is defined as the five months with the lowest Average Temperature Norm for that region. For the sake of this study, we do not require that these months be consecutive.

7.2 Heatwaves and the Maximum Temperature Threshold

A heatwave can be defined as when the maximum daily temperature reaches above the maximum temperature threshold for three or more consecutive days [10]. We define the maximum temperature threshold for a geographical region as the upper 90% of all of the daily maximum temperature measurements during a geographical regions warm season.

7.3 Cold Snaps and the Minimum Temperature Threshold

A cold snap can be defined as when the minimum temperature in a 24 hour period (12:00pm - 11:59am) reaches below the minimum temperature threshold for three or more consecutive 24 hour periods (12:00pm - 11:59am). We define the minimum temperature threshold for a geographical region as the bottom 10% of all of the daily minimum measurements for that geographical regions cold season.

8 Data

8.1 Calculation of Norms

All norms are calculated using daily data from the NOAA (National Oceanic and Atmospheric Association) (https://www.ncei.noaa.gov/cdo-web/datasets). We use R 4.0.5 to organize our data and run our calculations. For all of our norm calculations, we only use data contained between January 1st, 1990 and December 31st, 2019.

8.1.1 Calculation of the Average Temperature Norms and the Warm and Cold Seasons

We calculate the average temperature norms by breaking up the data by city, then by month, and then applying (14). We output all values calculated for each city by month to a CSV file. This CSV file, "Average_Norms.csv", is contained in the Assignment 1 folder. From this, we calculate the warm and cold season for each city. For the cold season, we find the five months with the lowest average temperature norm by city and for the warm season, we find the five months with the highest average temperature norm by city.

8.1.2 Calculation of Maximum Temperature Threshold

We can use the calculation of the warm season for each city to calculate the maximum temperature threshold. For each city, take all of the daily maximum temperatures from the five warmest months (warm season) and organize them from least to greatest. Out of these temperatures, we take infimum of the top 90% of the temperatures. This is our maximum temperature norm. For each city, this calculation is listed under the "WARM_SZN" column in the "warm and cold season norm.csv" file. We have also done this by month. The calculation by city for each month is contained the "Maximum_Norms.csv" file.

8.1.3 Calculation of Minimum Temperature Threshold

We can use the calculation of the cold season for each city to calculate the minimum temperature threshold. For each city, take all of the daily minimum temperatures from the five coldest months (cold season) and organize them from least to greatest. Out of these temperatures, we take supremum of the bottom 10% of the temperatures. This is our minimum temperature norm. For each city, this calculation is listed under the "COLD_SZN" column in the "warm_and_cold_season norm.csv" file. We have also done this by month. The calculation by city for each month is contained the "Maximum_Norms.csv" file.

8.1.4 Using Max and Min Norms to Identify Heat Waves and Cold Snaps

We use our maximum temperature threshold and our minimum temperature threshold for each city's warm and cold season to identify heatwaves and cold snaps in our data. To identify a heatwave, we look for when the maximum daily temperature is above our maximum temperature threshold for three consecutive days. To identify a cold snap, we look for when the minimum temperature for a 24 hour periods (12:00pm-11:59am) fall below the minimum temperate threshold for three or more consecutive days. In summary, and without loss of generality for heatwaves and cold snaps, this done by creating an identification vector and iterating over each daily maximum (or minimum) measurement in chronological order. If the temperature in question is greater (or less then) our norm, append an element to the vector. If it is the opposite, clear the vector. If the vector has a length of three, append to your data set that we have identified a heatwave or cold snap over the last three consecutive days, including the day we are on. For every day where a heatwave has not been identified, append that there is no heatwave on those specified dates. We have done this method for both our hourly and daily data. All calculation outputs for our daily data are contained in "daily_data_w_hw.zip". Each file has two extra columns which are "is_hw" and "is_cs". For each row, if the entry is 1, then a heat wave or cold snap has been identified for that row/day/measurement. If the entry is zero, then no heatwave or cold snap have been identified. For our hourly data, this is done for each row of time series and is identified with zeros and ones in columns "HEATWAVE" and "COLDSNAP" in each "stationnumber_all_info.csv" file.

8.2 Cleaning and Organizing Time-series

8.2.1 Cleaning and Filtering Hourly Data

We start with a large amount hourly temperature data from the NOAA which included 300 global cities (https://www.ncei.noaa.gov/cdo-web/). We keep only the temperature data that passed all quality control checks and organize this data into 500 point time series where each measurement is exactly three hours apart. The data that is not within the Jan 1st, 1900 - Dec 31st, 2019 window is dropped so only the data that corresponds with the calculated climate norms is analysed. This leaves us with 401845 analysable time series from 127 global cities.

8.2.2 Experimental groups

For this experiment, we use the classifier from Bury et al, 2021 [1] which can identify the presence of early warning signals from Hopf, fold, and transcritical bifurcations from any selected point within a 500 point time series. Due to the flexibility of the classifier, we decide to create two main experimental groups. The first is bifurcation classified at point 400 and the second is bifurcation classified at point 450. After classification, the first group is left with 401845 analysable time series from 127 global cities while the second group is left with 372547 analysable time series from 116 global cities. This was due to us deciding to drop miss-classified data. We also decide to break our two main groups up into two subgroups (4 groups total) where we analyse the classified data with and without the Hopf classifier. This is because we do not believe that a Hopf bifurcation accurately represents a heatwave or cold snap, which would decrease our model accuracy.

9 Model Assessment

We assess our model on each of the four groups with the following.

9.1 ROC Curves

An ROC (Receiver Operating Characteristic) curve is a plot showing the performance of a classification model at all classification thresholds [4]. On the y-axis of plot there is the amount of true positive events (correctly classified) and on the x-axis of the plot there is the amount of false positive events (incorrectly classified) [4]. Most ROC curves have there axis's ranging from zero to one to signify the percent of predictions on a zero to one scale that are correct and incorrect [4]. The way that an ROC curve is generated is we calculate the model performance at many evenly distributed thresholds ranging from zero to one. At each calculated threshold, we plot a point on the ROC plot where the location of the point is dictated by the true positives and false positives dictated by the model at the specified threshold. When done with sufficient density, we can connect the points on the plot which generates our ROC curve.

9.2 AUC

A way that an ROC curve can be assessed is by calculating the ACU (area under the curve). The AUC measures the entire two dimensional area underneath the entire ROC curve [4]. This can be calculated in a verity of way but is mostly done by numerical integration using the points used to make the curve for the approximation [4]. The AUC ranges from zero to one, where the model is 100% correct if the area under the curve is 1 [4]. If the area under the curve is zero, then the model is 0% correct. If the model has an AUC > 0.9, it is considered to be strong. On the other hand, a model with an AUC < 0.6 can be considered as weak.

9.3 Assessment Parameters

In addition to the ROC curves and AUC, we assess our model using the following parameters. We represent the number of true positive events as TP, the number of true negative events as TN, the number of false negatives events classified by the model as FN, and the number of false positive events classified by the model as FN for each equation below.

- **Accuracy**: The accuracy of the model tells us the raw percentage of the classifications that the model correctly identities [2]. $Acc \in [0, 1]$ and can be calculated by

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{16}$$

where 0 is the lowest accuracy and 1 is the highest [2].

- **Precision**: The precision parameter identifies how well the model predicts the positive class [2]. $Pre \in [0, 1]$ and can be calculated by

$$Pre = \frac{TP}{TP + FP} \tag{17}$$

where 0 is the lowest precision and 1 is the highest [2].

- **Recall (Sensitivity)**: The recall measures how well the model identifies true positive events [2]. $Rec \in [0, 1]$ and can be calculated by

$$Rec = \frac{TP}{TP + FN} \tag{18}$$

where 0 means that the cannot identify true positive events and 1 means that the model correctly identifies all true positive events in the data [2].

- **F1-Score**: The F1-score is the weighted average score of the recall and precision [2]. $F1 \in [0, 1]$ and can be calculated by

$$F1 = \frac{2 * Pre * Rec}{Pre + Rec} \tag{19}$$

where 1 is the best performance and 0 is the worse [2].

- **Specificity**: Specificity measures the rate of true negatives that are identified correctly. $Spc \in [0, 1]$ and can be calculated by

$$Spc = \frac{TN}{FP + TN} \tag{20}$$

where 0 means that the model cannot identify true negatives in the data and 1 means that the model correctly identifies all true negatives in the data [2].

- **Prevalence**: Measures how often positive events occur in the data [2]. $Prv \in [0, 1]$ and can be calculated by

$$Prv = \frac{TP + FN}{TP + TN + FP + FN} \tag{21}$$

where 0 means that the data has no positive events while 1 means that the data has an extremely high amount of positive events [2].

10 Results and Discussion

We conduct the experiment on our four groups. We assess the model at thresholds 0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95, and 1.00 for all of our groups. In each table below, 400_w_h represents the group without the Hopf classifier where the bifurcation is estimated at point 400 within the 500 point time series, 400_wo_h represents the group wit the Hopf classifier where the bifurcation is estimated at point 400 within the 500 point time series, 450_w_h represents the group without the Hopf classifier where the bifurcation is estimated at point 450 within the 500 point time series, and 450_wo_h represents the group wit the Hopf classifier where the bifurcation is estimated at point 450 within the 500 point time series.

10.1 Model Accuracy

We assess the model accuracy for our four groups at each one of our selected thresholds using (1) (Table 8). There is an increase in model accuracy for each group as we increase threshold (Figure 1). We also see that the groups without the Hopf classifier are significantly more accurate then the groups with the Hopf classifier (Figure 2). The bifurcation at 450 group and the bifurcation at 400 groups have very similar accuracy when comparing there with and without Hopf groups (Figure 2). This suggest that a Hopf bifurcation does not represent heat waves or cold snaps which contributes to low model accuracy. It also suggest that there is some other temperature phenomena(s) that are being identified by the Hopf classifier in our data which suggest some other temperature phenomena can be accurately represented by a Hopf bifurcation. We need to investigate further to identify what these other temperature phenomena(s) are.

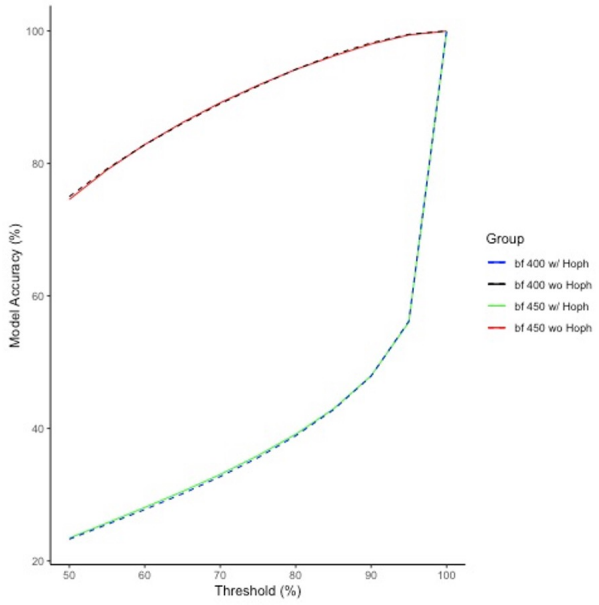


Figure 2: Model Accuracy

Thresholds	400_w_h	400_wo_h	450_w_h	450_wo_h
0.5	0.233	0.75	0.234	0.746
0.55	0.255	0.792	0.257	0.79
0.6	0.278	0.828	0.281	0.828
0.65	0.301	0.86	0.305	0.862
0.7	0.327	0.89	0.33	0.892
0.75	0.356	0.917	0.359	0.918
0.8	0.389	0.942	0.391	0.942
0.85	0.428	0.964	0.429	0.962
0.9	0.479	0.982	0.48	0.98
0.95	0.562	0.995	0.56	0.994
1.0	1	1	1	1

Table 8: Model Accuracy

10.2 F1 Score

We calculate the F1 score for each group at each one of our selected thresholds. We see that our overall F1 scores are very low with the maximum score being 0.094 and the lowest being 0.051 (Table 9). The NA values in the table are due to the NA values for the precision at the 100% threshold (will be discussion in the next section). We observe that the F1 score peaks for all groups at our lowest calculated threshold of 0.5. This suggest that our model either has a very low precision and recall, or precision or recall is significantly higher then its compliment in the formula (4). In our words, we observe that our model is not identifying true positives, true negatives, or both with a high degree of accuracy which is identified in sections 10.3 and 10.4.

Thresholds	400_w_h	400_wo_h	450_w_h	450_wo_h
0.5	0.093	0.077	0.095	0.086
0.55	0.093	0.074	0.095	0.086
0.6	0.092	0.073	0.094	0.085
0.65	0.092	0.072	0.094	0.086
0.7	0.092	0.067	0.094	0.086
0.75	0.091	0.067	0.093	0.082
0.8	0.091	0.066	0.092	0.076
0.85	0.091	0.065	0.091	0.06
0.9	0.09	0.058	0.091	0.056
0.95	0.088	0.051	0.092	0.09
1.0	NA	NA	NA	NA

Table 9: F1 scores

10.3 Precision

We calculate the model precision for each group at each of our selected thresholds. Overall, we observe that our model has a very low precision for all of our groups which ranges between 0.027 and 0.05 (Table

3). The NAs are caused by a zero in the denominator in the formula at threshold one. This will need to be further investigated before we get our next set of results for the project as it is very likely to be a mistake. We observe that the precision of the model is slightly higher at lower thresholds with the max calculated value being at a threshold of 0.5, 0.55, and 0.6 for most groups (Table 10). This suggest that our model is very bad at only identifying only true positives in the data, either by not being able to identify true positives at all or by identifying a bunch of false positive events as well. This implies that there are most likely other temperature spikes and drops in the data that can be represented at our three bifurcations in addition to just heatwaves and colds snaps. We will need assess the model recall (next section) to see if our model is actually identifying heatwaves and cold snaps in the data as well, or if it is solely identifying other temperature events.

Thresholds	400_w_h	400_wo_h	450_w_h	450_wo_h
0.5	0.049	0.04	0.05	0.045
0.55	0.049	0.038	0.05	0.045
0.6	0.048	0.038	0.049	0.044
0.65	0.048	0.037	0.049	0.045
0.7	0.048	0.035	0.049	0.045
0.75	0.048	0.034	0.049	0.043
0.8	0.048	0.034	0.048	0.04
0.85	0.047	0.033	0.048	0.031
0.9	0.047	0.03	0.048	0.029
0.95	0.046	0.027	0.048	0.049
1.0	NA	NA	NA	NA

Table 10: Precision

10.4 Recall

We calculate the model recall for each of our groups at each of our selected thresholds. Our model has very high recall of each of our groups excluding 450_wo_h for all thresholds excluding 1 (Table 11). For 450_wo_h, the model maintains a recall above 90% up to a threshold of 0.8 (Table 11). This suggest that our of our model is extremely good up to a threshold of 0.8 at identifying true positives in the data for all groups, which means that the model misses very little heatwaves and cold snaps at these thresholds. With the low precision identified in section 10.3, this implies that the model is picking up other temperature events that are not heatwaves or cold snaps. In addition to this, we also observe that the model overall has a higher recall for the groups including the Hopf bifurcations (Table 11). This implies that a small amount of heatwaves and cold snaps are actually identified by the Hopf classifier so without it, the model is prone to missing some of these events. We also observe that the recall decreases as we raise the threshold (Table 11). The cause of this is all heatwaves and cold snaps do not cause the model probability to be above 0.8. This leads to an increase in false negatives in these higher thresholds groups, and therefore a lower recall.

10.5 Specificity

We calculate the model specificity for each of our groups for each of our selected thresholds. The model specificity increases as we increase model threshold, with the groups without the Hopf classifier having a

Thresholds	400_w_h	400_wo_h	450_w_h	450_wo_h
0.5	0.999	0.985	0.998	0.981
0.55	0.999	0.982	0.998	0.975
0.6	0.999	0.979	0.998	0.97
0.65	0.999	0.972	0.998	0.966
0.7	0.999	0.965	0.998	0.955
0.75	0.998	0.954	0.997	0.94
0.8	0.998	0.929	0.997	0.907
0.85	0.998	0.892	0.996	0.83
0.9	0.997	0.792	0.995	0.692
0.95	0.996	0.484	0.994	0.536
1.0	0	0	0	0

Table 11: Recall

significantly higher specificity then the groups with the Hopf classifier (Table 12). This suggest that many of the false positives are caused by the Hopf bifurcation classifier which leads to the model having higher accuracy when the Hopf classifier is not included, which has been shown in section 10.1. The increase in specificity with the increase in model threshold suggest that the higher the threshold, the more likely the model is to properly identify a heat waves and cold snaps due to the decease in false positives identified by the model (5).

Thresholds	400_w_h	400_wo_h	450_w_h	450_wo_h
0.5	0.201	0.747	0.202	0.743
0.55	0.226	0.79	0.227	0.788
0.6	0.25	0.827	0.253	0.827
0.65	0.276	0.86	0.279	0.861
0.7	0.304	0.89	0.306	0.891
0.75	0.334	0.917	0.337	0.918
0.8	0.37	0.942	0.372	0.942
0.85	0.411	0.964	0.413	0.962
0.9	0.465	0.982	0.466	0.98
0.95	0.552	0.995	0.551	0.994
1.0	1	1	1	1

Table 12: Specificity

10.6 Prevalence

We calculate the prevalence for all of our group at all selected thresholds. We observe that the prevalence is very low for all groups, with the prevalence decreasing as we increase thresholds (Table 13). This suggest that that there is a very low abundance of heatwaves and cold snaps in the data, which is expected due to the definition of a heat wave and cold snap. It also suggest that there is a decrease in events as we raise model threshold which is also expected as the model is less likely to identify positive events if the

threshold is higher. Prevalence between all groups is very similar (Table 13). We will need to assess if and how much the model is skewed and or bias due to the low number of true positive events.

Thresholds	400_w_h	400_wo_h	450_w_h	450_wo_h
0.5	0.039	0.011	0.04	0.012
0.55	0.038	0.008	0.039	0.01
0.6	0.037	0.007	0.037	0.008
0.65	0.035	0.006	0.036	0.007
0.7	0.034	0.004	0.035	0.005
0.75	0.032	0.003	0.033	0.004
0.8	0.031	0.002	0.031	0.003
0.85	0.029	0.001	0.029	0.001
0.9	0.026	0.001	0.026	0.001
0.95	0.021	0	0.022	0.001
1.0	0	0	0	0

Table 13: Prevalence

10.7 AUC and ROC Curves

We generate ROC curves for each on of our thresholds (Figure 4 as example for all groups) and calculate the area under each ROC curve for each of our thresholds (Figure 3). We see that for the bifurcation at 450 group, the area under the curve ranges from 0.5335 to 0.5463 and is the same for both with and without Hopf (Figure 3). The bifurcation at 400 group, we see that both with and without Hopf groups have different are under there curves, with the without Hopf group having a slightly higher accuracy then the with Hopf group for all thresholds (Figure 3). For both groups, the area under the curve is constant for all thresholds (Figure 3). This suggest that our model has a low precision at identifying true positive events which is verified by our precision calculations (Section 3.3). This means that even though our model is quite successful at identifying heatwaves and colds snaps in the data, the model is also picking up other events in the data as well which has been mentioned in the above sections. This means we need to refine our method and possibly our definitions of heat waves and cold snaps to accurately identify these events in the data.

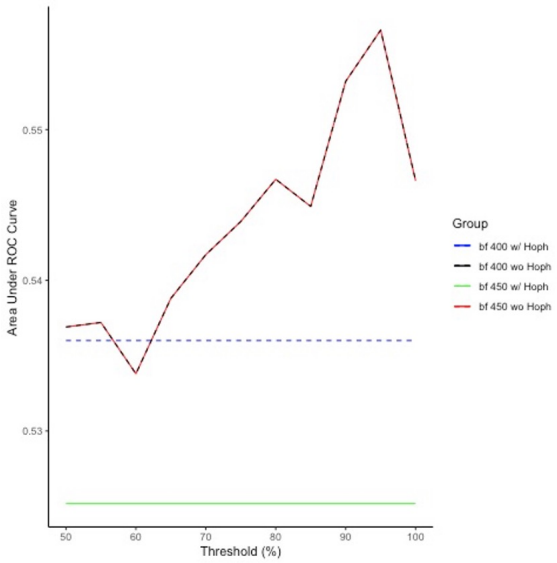
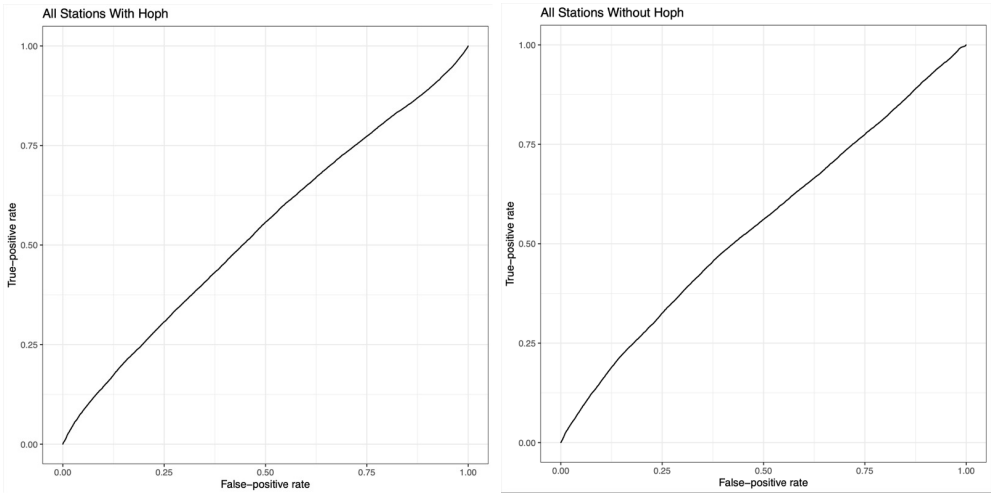


Figure 3: AUC



(a) (b)

Figure 4: ROC Curves

11 Future Research

Our results, though not perfect, are overall positive as they show us that our theory is not wrong signified by our model assessments. The main issue is that we are most likely identifying other weather phenomena(s) in addition to heatwaves and cold snaps. We can conclude from this that our results serve as solid ground

to eventually gain our intended results, which opens the door for future research to refine our method and make our model more successful in predicting only heatwaves and cold snaps. To improve our model and method, we plan to do the following.

- Test our the model on a 600 point temperature time series where we predict the bifurcation at point 500 for one hour and three hour gaps.
- Analyse multidimensional 600 point time series where we predict the bifurcation at point 500 for one hour and three hour gaps. Other dimensions include parameters such at pressure, wind speed, wind directions, and precipitation as independent variables. This is because extreme values other parameters are known to cause heat waves and cold snaps so there may be identifiable early warning signals within these other parameters the extreme temperature events.
- Analysing all three bifurcation classifiers separately for all groups and time series to see which is the most accurate, which picks up the most false positives, and which has the most false negatives.
- Analyse current results to see what is causing all of the false positives, especially for the Hopf classifier.
- Revisit and modify current definitions of heatwaves and cold snaps to improve accuracy.
- Preform lag window analysis to see how long after the predicted bifurcation does a heatwave or cold snap occur for true positive events.

  Due to what what we have achieved so far, we predict a high degree of success with our model if we investigate what is listed and make the changes necessary.

References

[1] Bury, T.M., Sujith R. I., Pavithran I., Scheffer M., Lenton T.M., Anand M., Bauch C.T. (2021). *Deep learning for early warning signals of tipping points.* Proceedings of the National Academy of Sciences of the United States of America. 118(39).

[2] DataTechNotes. (2019). *Precision, Recall, Specificity, Prevalence, Kappa, F1-score check with R.* DataTechNotes A blog about data science and machine learning https://www.datatechnotes.com/2019/02/accuracy-metrics-in-classification.html

[3] European Comission. 2023. *Methane emissions.* European Union

[4] Google For Developers. (2022). *Classification: ROC Curves and AUC.* Google https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc

[5] Government of Alberta. (2023). *Oil sands 101.* Government of Alberta https://www.alberta.ca/oil-sands-101: :text=Trucks%20move%20the%20oil%20sand,ponds%20where%20the%20sand%20settles.

[6] Government of Canada. (2023) *Alberta Sector Profile: Mining, Quarrying, and Oil and Gas.* Government of Canada https://www.jobbank.gc.ca/trend-analysis/job-market-reports/alberta/sectoral-profile-mining-oil-gas

[7] Government of Canada. (2023). *Net-zero emissions by 2050.* Government of Canada https://www.canada.ca/en/services/environment/weather/climatechange/climate-plan/net-zero-emissions-2050.html

[8] Department of Statistics. (2018). *10.7 Detecting Multicollinearity Using Variance Inflation Factors.* The Pennsylvania State University

[9] FrontlineSolvers inc. (2022). *Classification Trees.* FrontlineSolvers; A leader in analytics for spreadsheets and the web. https://www.solver.com/classification-tree

[10] Li, P., Yu, y., Huang, D., Wang, Z-H., Sharma,. A. (2022). *Regional Heatwave Prediction Using Graph Neural Network and Weather Station Data.* Geophysical Research Letters, 50(7)

[11] Wagner, G. (2022). *Notes-Topic 3 Inference For Several Populations.* University of Alberta

[12] Wagner, G. (2022). *Notes-Topic 5 Multiple Linear Regression.* University of Alberta

[13] Wang, C., Wang, Z.-H., Sun, L. (2020). *Early-warning signals for critical temperature transitions.* Geophysical Research Letters, 47

[14] Wood Buffalo Environmental Association. https://wbea.org/data/network-map-station-data/

[15] Zach. 2021, *What is Considered a Good AIC score*, Statology. https://www.statology.org/what-is-a-good-aic-value/