Anomaly Detection: Wednesday Night Household Global Power Usage

Group 15

Daniel Tavaszi
Stefano Macri
Khizr Pardhan
Darien Flamont
Anderi Hristea

Objectives:

- Decide on a time window that was consistent across all test sets

- General Data exploration on different time windows and seasonality

- Explain point anomalies using max and min ranges in training set and relating them to moving window averages of test sets

- Explain contextual anomalies using HMMs and normalized log likelihood calculations

Fig. 1.1 – Wednesday Morning MSE Calculations

	Mean Error Test 1	Mean Error Test 2	Mean Error Test 3	Mean Error Test 4	Mean Error Test 5	STD Error Test 1	STD Error Test 2	STD Error Test 3	STD Error Test 4	STD Error Test 5
Spring - Season 1	0.287	0.370	0.287	12.943	13.230	0.2558	0.2262	0.2558	5.3846	5.5031
Summer - Season 2	0.128	0.139	0.128	6.666	6.581	0.0952	0.1078	0.0952	3.6175	3.5694
Fall - Season 3	0.385	0.376	0.385	16.507	16.488	0.1782	0.1693	0.1782	5.2776	5.2382
Winter - Season 4	0.176	0.237	0.176	14.279	14.398	0.1462	0.1447	0.1462	5.1911	5.4687

Fig. 1.2 – Wednesday Night MSE Calculations

	Mean Error Test 1	Mean Error Test 2	Mean Error Test 3	Mean Error Test 4	Mean Error Test 5	STD Error Test 1	STD Error Test 2	STD Error Test 3	STD Error Test 4	STD Error Test 5
Spring - Season 1	0.210	0.231	0.210	10.238	9.830	0.1093	0.0992	0.1093	3.9950	3.7116
Summer - Season 2	0.246	0.272	0.246	8.413	8.183	0.1125	0.1262	0.1125	4.6450	4.5540
Fall - Season 3	0.139	0.124	0.139	8.733	8.372	0.0972	0.0825	0.0972	4.8271	4.4512
Winter - Season 4	0.185	0.411	0.185	16.468	16.741	0.1265	0.1564	0.1265	6.9111	7.0086

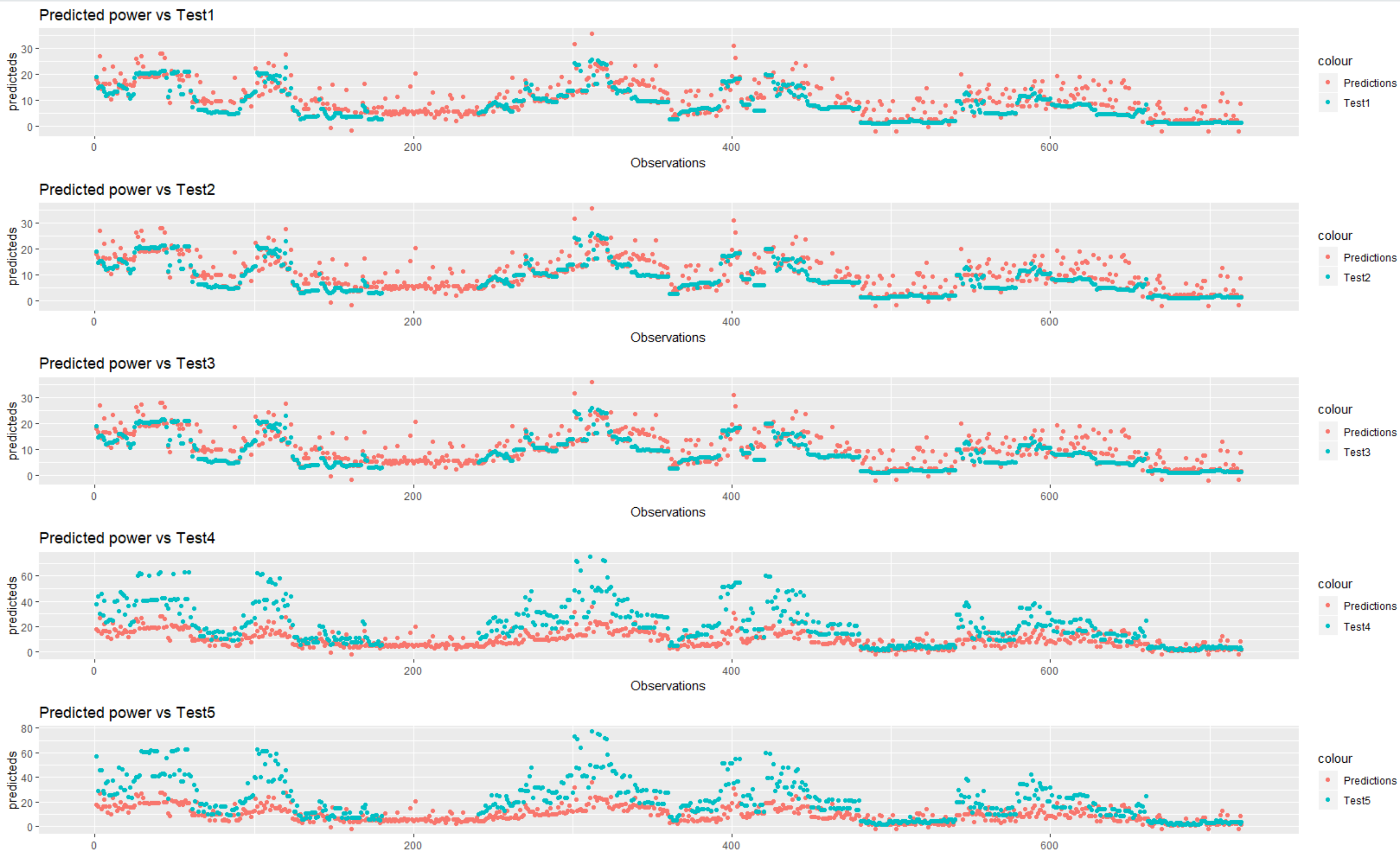
Fig 1.3 - Correlation between all observations when time was aggregated hourly

	Global_active_power	Global_reactive_power	Voltage	Global_intensity	Sub_metering_1	Sub_metering_2	Sub_metering_3
Global_active_power	1.0000000	0.2121330	-0.3073528	0.8092017	0.3691985	0.3467608	0.5165714
Global_reactive_power	0.2121330	1.0000000	-0.1544859	0.3346676	0.3541653	0.2567012	0.1124776
Voltage	-0.3073528	-0.1544859	1.0000000	-0.3979771	-0.2099719	-0.1710845	-0.2991657
Global_intensity	0.8092017	0.3346676	-0.3979771	1.0000000	0.5056849	0.4598049	0.6772855
Sub_metering_1	0.3691985	0.3541653	-0.2099719	0.5056849	1.0000000	0.1237429	0.2042420
Sub_metering_2	0.3467608	0.2567012	-0.1710845	0.4598049	0.1237429	1.0000000	0.1372597
Sub_metering_3	0.5165714	0.1124776	-0.2991657	0.6772855	0.2042420	0.1372597	1.0000000

Fig 1.4 - Correlation between all observations when time was aggregated daily

	Global_active_power	Global_reactive_power	Voltage	Global_intensity	Sub_metering_1	Sub_metering_2	Sub_metering_3
Global_active_power	1.00000000	-0.09430221	0.15414334	0.86313850	0.39828051	0.38089763	0.60700967
Global_reactive_power	-0.09430221	1.00000000	-0.06300499	0.05077379	0.33947151	0.17856335	0.04455324
Voltage	0.15414334	-0.06300499	1.00000000	0.08135940	-0.07240053	-0.07435111	0.13252725
Global_intensity	0.86313850	0.05077379	0.08135940	1.00000000	0.54520350	0.48912452	0.73306878
Sub_metering_1	0.39828051	0.33947151	-0.07240053	0.54520350	1.00000000	0.24212045	0.32911920
Sub_metering_2	0.38089763	0.17856335	-0.07435111	0.48912452	0.24212045	1.00000000	0.22734383
Sub_metering_3	0.60700967	0.04455324	0.13252725	0.73306878	0.32911920	0.22734383	1.00000000

Fig. 1.5 - Predicted values against recorded values for Global Intensity



Point Anomalies

• Shift from Phase 1 to only Wednesday Nights (mainly summer) to explain relationship between test sets

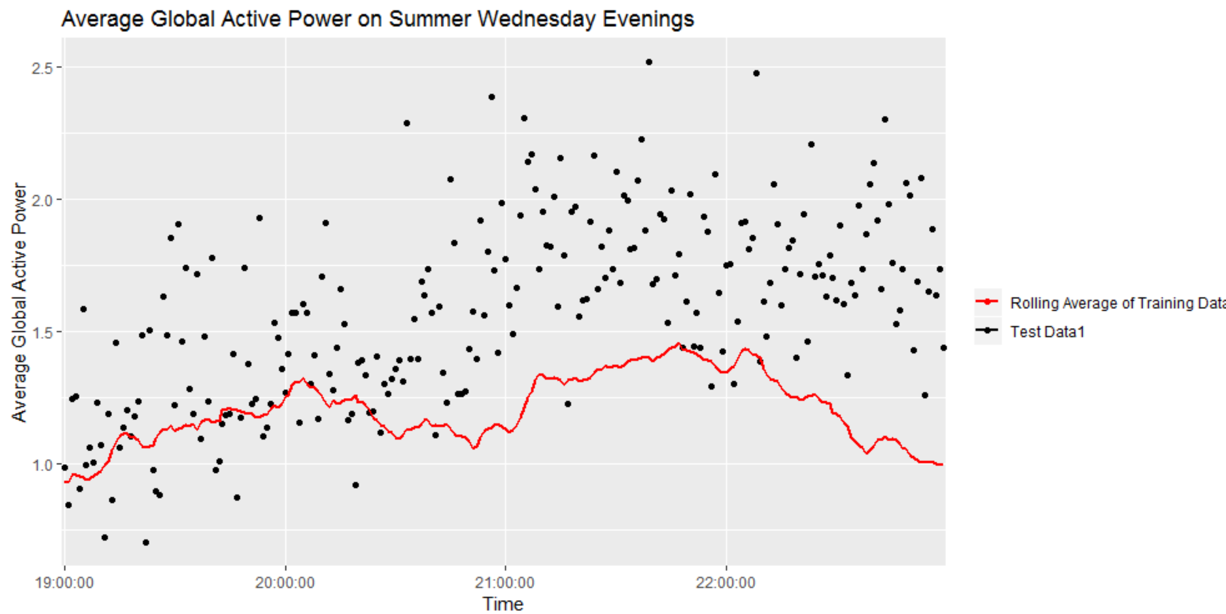
• Used a maximum and minimum bound based on the total training set on Wednesday Nights during each season

Fig. 1.6 - Training set GAP Max and Min Wednesday nights

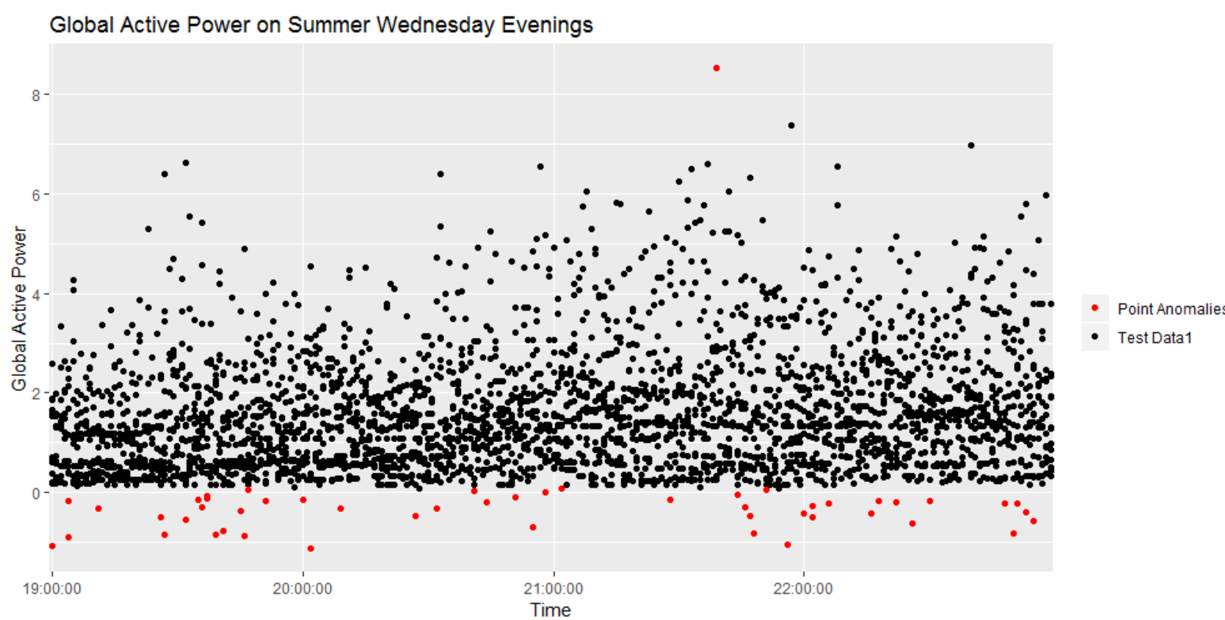
	Max Global Active Power	Min Global Active Power
Spring - Season 1	8.06	0.184
Summer - Season 2	7.436	0.08
Fall - Season 3	7.03	0.188
Winter - Season 4	8.974	0.204

Using these maximum and minimum bounds we tested each test set for Wednesday nights during the summer to determine point anomalies in each data set

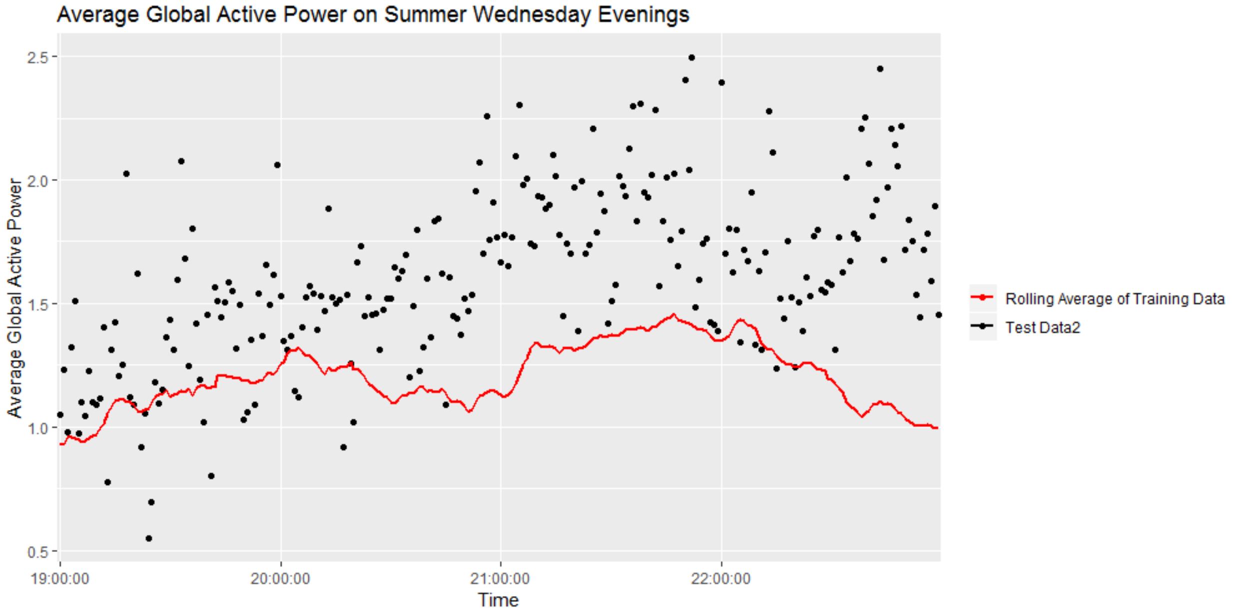
Test set #1 averages vs training set rolling average



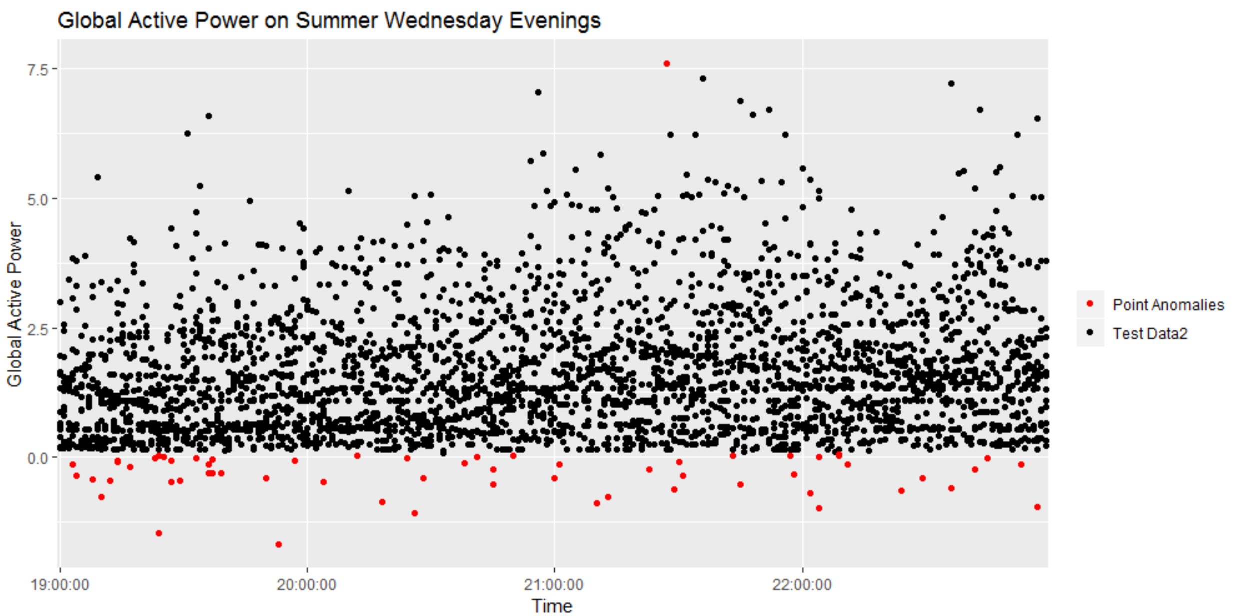
Test set #1 Data with point anomalies



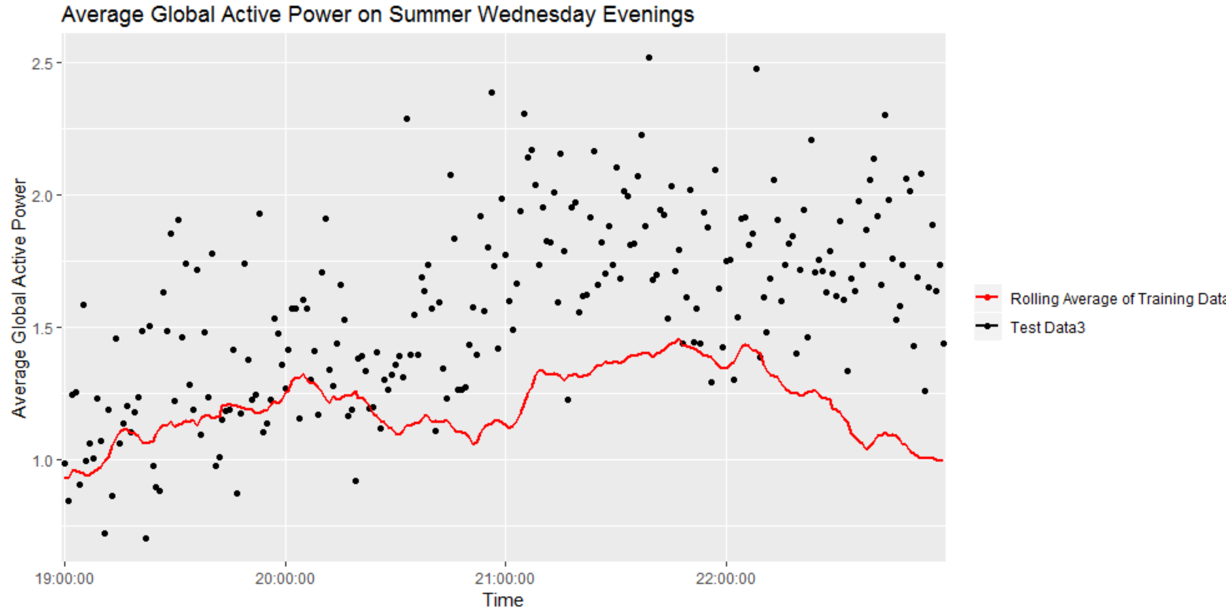
Test set #2 averages vs training set rolling average



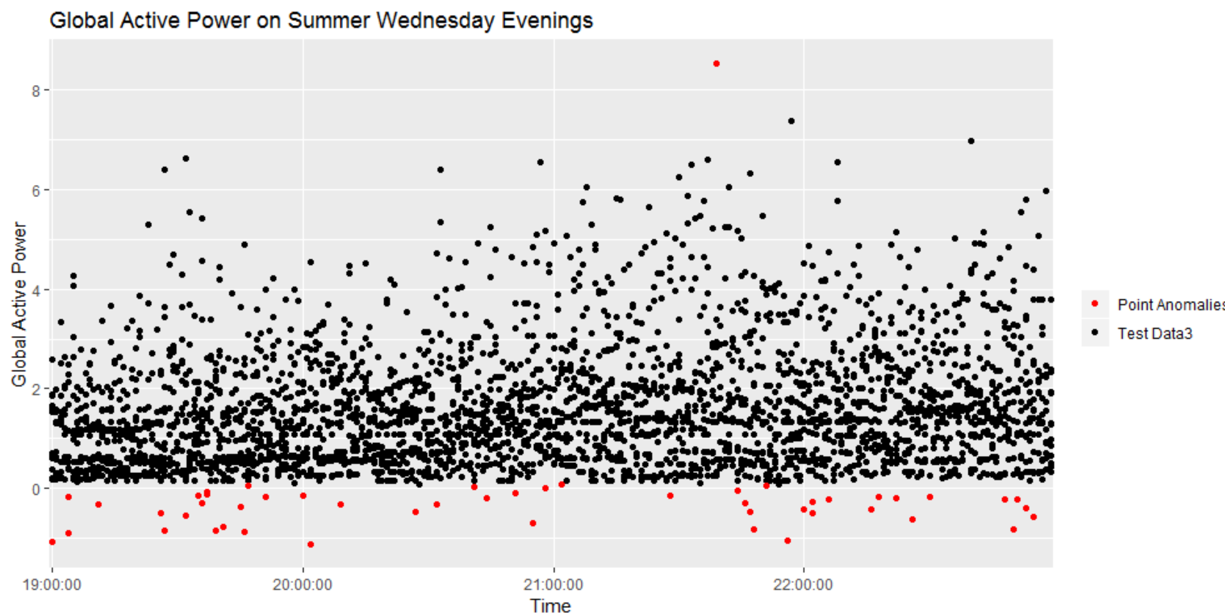
Test set #2 Data with point anomalies



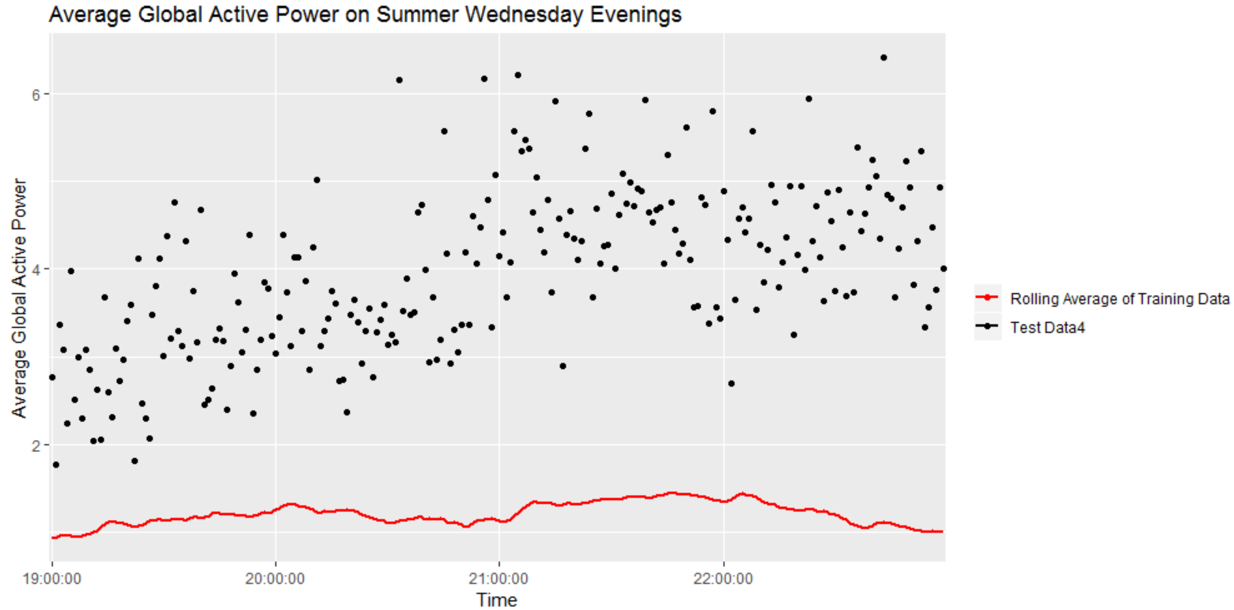
Test set #3 averages vs training set rolling average



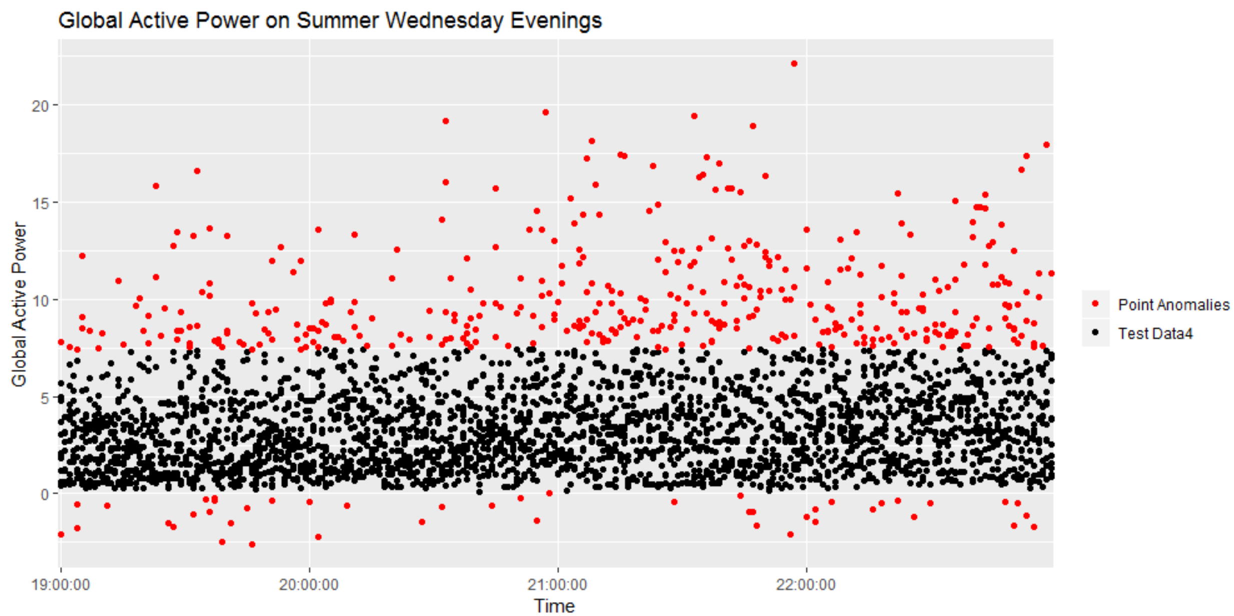
Test set #3 Data with point anomalies



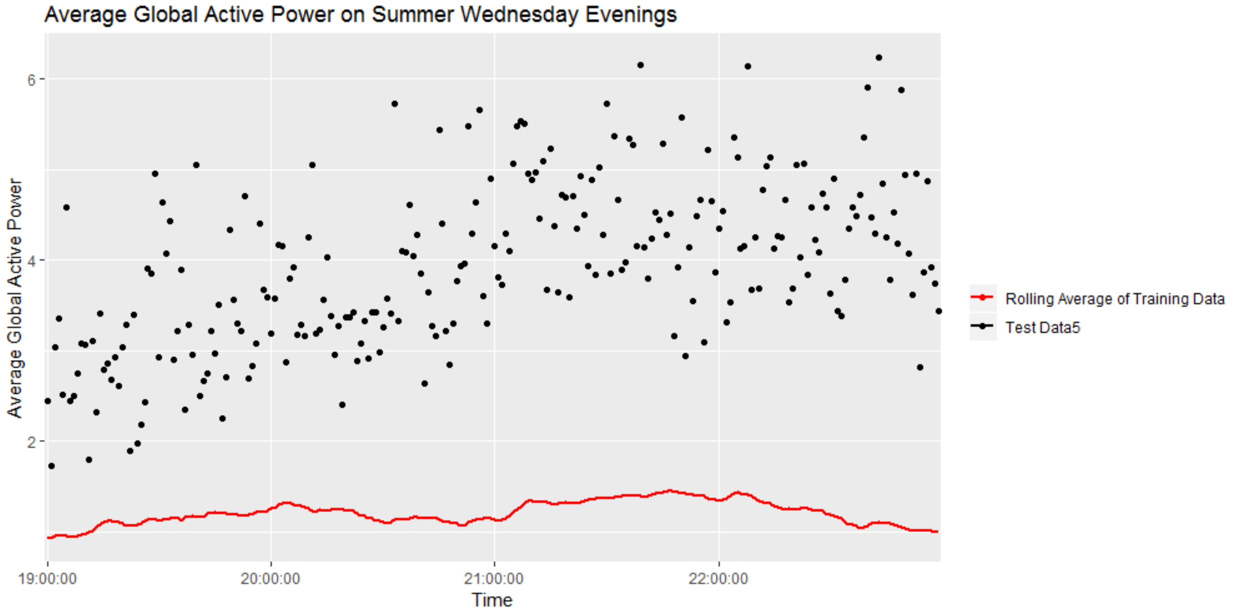
Test set #4 averages vs training set rolling average



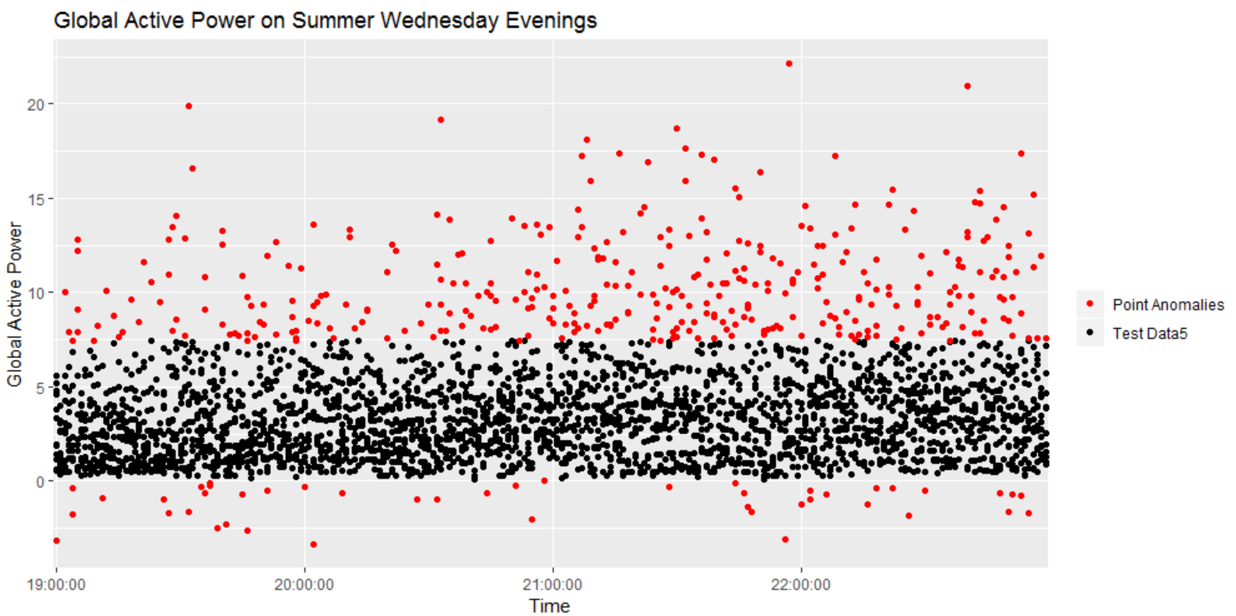
Test set #4 Data with point anomalies



Test set #5 averages vs training set rolling average

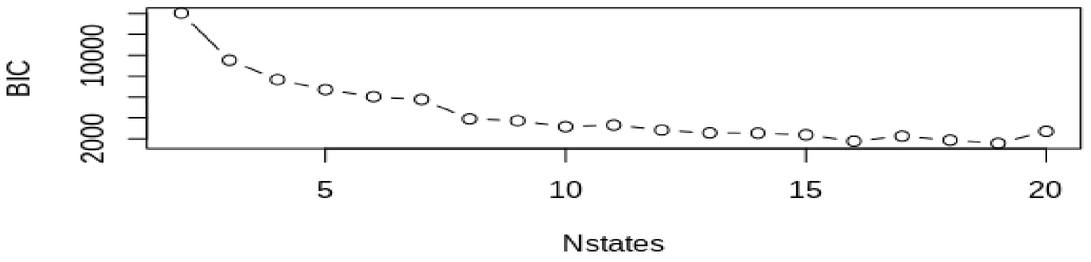


Test set #5 Data with point anomalies



Contextual Anomalies

Training data BIC per Nstates



Normalized log likelihoods using 39 observations for training data and 13 observations per training set

	Training set (training original)	Test set #1	Test set #2	Test set #3	Test set #4	Test set #5
Normalized Log Likelihood	185.6229 (556.8686)	-5208.521	-5538.565	-5208.823	-16194.67	-15983.83

Key Findings

• Test sets #1 and #3 included exactly the same data points for Wednesday nights during the summer when doing point anomalies and we assume one is just a copy based on error calculations and contextual anomalies

• Test sets #4 and #5 were highly anomalous with very high values compared to the expected behaviour of the training set and had very high log likelihoods

• Global active power and Global reactive power had the closest correlation constants which we expected because they are a function of one another

Future applications

• Using a live max and min point anomaly detection techniques could potentially give a first response to malicious or unexpected behaviour

• Contextual anomaly detection can be used to determine normal behaviours across long spans of time to make better reporting of power consumption in the past when compared to the present.