Institute of Technology Blanchardstown
Ireland

Masters of Science

Traffic Prediction and Analysis using a Big Data and Visualisation Approach

Author:
Declan McHugh

Supervisor:
Laura Keyes

A thesis submitted in partial fulfilment of the requirements for the degree of Masters of Science

in the field of

Computer Science
Business Intelligence and Data Mining

September 2014

Declaration of Authorship

I, Declan McHugh, declare that this thesis titled, 'Traffic Prediction and Analysis using a Big Data and Visualisation Approach' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: __________________________________

Date: __________________________________

"Thanks to my solid academic training, today I can apply my knowledge of Data Science in both the Academic and Industry fields"

Declan McHugh

Abstract

This thesis is an approach of using big data, visualisation and data mining techniques to predict and analyse traffic. The objective is to understand traffic patterns in Dublin City. The data was captured from open data sources, Dublinked, Wunderground and Twitter.

With the aid of Python's Sklearn Kit, Google Maps and MongoDB a scalable solution was implemented to identify the roads that are impacted by adverse weather conditions, amoung other causes for poor traffic conditions and which regression models best predict areas of the city.

Seasonality and trends found that traffic pattern on weekends differ from business days. Peak time also differ spatially. Peak times for traffic is not the same time and can vary from the expected time 8-9am to late evening 8-9pm for inbound traffic.

The ARIMA model was heavily used as a forecasting model. Using Ordinary least squares regression a clear pattern was shown to that lagging the day by 3, and using a lagged week 1 would be best suited for a regression model.

The importance of using all spatial neighbours for the prediction model was insignificant and the highest correlated neighbour was used measured by Principle Component Analysis.

Weather conditions also cause diverse traffic patterns. With high temperature it was found the travel increases around national parks and city centre. This indicates people are likely to leave the home to socialise, go for walks or shopping. Wet conditions the impact is more evenly spread.

Traffic tweets with geographical position stored was unsuccessful for determining the location of a traffic incident or the cause. This was due to users do not immediately tweet updates at the time of the occurring event.

The end result was an high performance web application that produces a analytical dashboard providing traffic prediction and analysis using historical traffic and social media data.

Acknowledgements

Thanks to Laura Keyes for the guidance in writing this paper, Geraldine Grey for putting together a well structured course and Marcus Hoffemann for the experience in the field of Data Mining.

Contents

List of Figures

List of Tables

Abbreviations

STT	**S**tandard **T**ravel **T**ime
OL	**O**bserved **L**ocation
P	**P**arameters ID
MSE	**M**ean **E**rror
R2	**R**2 (coefficient of determination) regression score function
EVS	**E**xplained **V**arience **S**core
MAE	**M**ean **A**bsolute **E**rror
Q	**Q**uantile 80

For my wife for being patient with me during my studies

Chapter 1

Introduction

1.1 Overview

The aim of this thesis is to analyse traffic patterns for an urban city, Dublin and to provide a visual dashboard for analysing traffic for a Smarter City. The initial data sets used vary from remote sensed data and social media information from open-data sources. One of the challenges this paper achieve is the Big Data four V's (volume, velocity, variety and veracity), see section 1.5.

Smart Cities is an initiative that has been adopted in many European cities. Smart Cities goal is an enabler for better planning, social and infrastructure management. Examples of cities using this includes Dublin, Lyon, Amsterdam and Barcelona.

Many European cities including Dublin have an active open data programme. Although there ongoing issues around privacy laws there are still many open data portals available online to the general public. The Dublin City Council make over 250 data sets available [3].

Wireless sensor networks is a technology which has played a massive role enabling a Smarter City. Dublin along with many other cities is using this technology to gather data related to traffic. The objective is to have a complete infrastructure that enable the monitoring of traffic behaviours so decisions on city development can be made in a smarter way. Variables such as weather conditions and seasonality may be able to improve decision on road network design.

1.2 Project Objectives

Objectives of this work are:

- To obtain and store historical traffic, related weather data to build and generate a generic prediction model.
- To obtain and store twitter data and design an approach to provide further analysis for traffic related events.
- To create a analytics dashboard demonstrates traffic patterns from data mining techniques, prediction models and twitter analysis.

1.3 Research Question

Can open data and social media be used to predict and analyse traffic as part of a smarter city?

1.4 Methodology Outline

Research was conducted to identify traditional methods of traffic prediction and analysis in the area forecasting and spatial data mining. This was used to identify problems that researchers had developing analyses. Further research was to form concept how to approach an analysis of traditional traffic prediction method with historical data and social media. Data Mining CRISP DM methodology was used throughout the project. CRISP DM is an industry standard for data mining. This methodology played a key role appropriate techniques and tools.

1.5 Big Data Background

Throughout the thesis techniques where employed in all phases of development to handle the four V's of Big Data: volume, velocity, variety and veracity.

- **Volume** of traffic data is a challenge that is overcome using **MapReduce**. By grouping related data together that allowed the database system perform searching efficiently through another mechanism called **Indexing**.

- **Velocity** of twitter data for this system was read in at real-time. Again **MapReduce** and **Indexing** was used to process and store the data.
- **Variety** of the data sources traffic, weather and twitter contain data types such as timestamps, geo-spatial, strings and integers. The database system called MongoDB was used and catered for these needs.
- **Veracity** in this case is the storing the data in preparation for analysis.

The technique for overcoming in big data systems *divide and conquer*[4]. MongoDB is a NoSQL the considers the challenges of the four V's[5] and allows users implement a design in such a way that data can be stored and retrieved efficiently.
NoSQL is lightweight Big Data database. When implementing a solution the database design creating indexes is crucial. These indexes allow for the system to divide a collection into segments. In the background the database the collections are being chunked and an index table is then generated for mapping data and its location on the file system. For instance in figure 1.1, a user collection on the file system is chunked based on the index criteria **score**. This leads to fast retrieval of the collection as the database knows to only search chunked files containing relevant data.

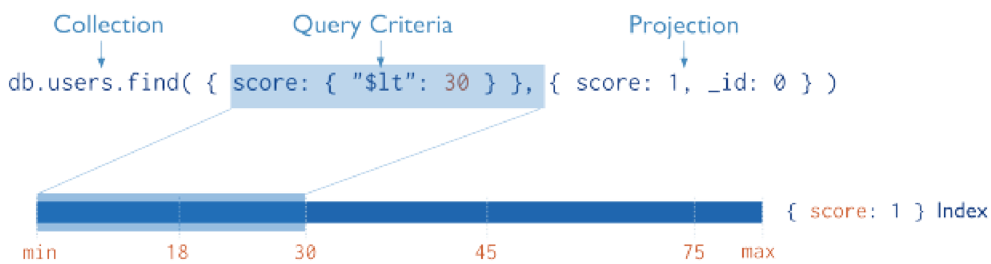


FIGURE 1.1: MongoDB internal divide and conquer approach [5]

For time-series collection of data, such as this paper provides, the data is indexed based on the timestamp and or spatial location.

1.6 Application Tooling

In this work there are a number of aspects that needed to be considered. There a lot of data preprocessing, natural language processing, mathematical algorithms, twitter integration, spatial data mining, visualization, web application, storage of structured and unstructured data. Python provides many of the tools necessary for data and scientific processing 1.1.

Table 1.1: Main Python Modules

Package	*Description*
SciPy	Scientific Algorithms and Methods
NumPy	Number Manipulation
Django	Web Application
TwitterAPI	Twitter Integration
PyMongo	MongoDB NoSQL Integration

1.7 Structure of this thesis

The aim of this work, as described within this chapter, is to explore the open data from Dublinked, Wunderground and Twitter using big data and data mining techniques which include methods such as visualisation.
Chapter 2 explains the different components for traffic prediction, big data and smart city.
Chapter 3 gives in greater detail the information the open data sources make available.
Chapter 4 is the collection and exploration the data while integrating the *divide and conquer* approach.
Chapter 5 contain the development of the generic traffic prediction model and an approach for text classification analysis of twitter.
Chapter 6 outlines the results and conclusions of the investigation as well analytics dashboard outcome.

Chapter 2

Literature Review

Much research around traffic patterns in road networks in a city limited to small number of roads and/or limited size if time series [some reference]. In this paper there is an objective if recognising how the many different algorithms perform. Dublin City offers a an opportunity to avail of showing the contrasting roads.

2.1 Introduction

This section starts with a detailed review of traffic prediction and analyses. The focus of the review is to identify different methods and techniques researchers have applied in algorithms, analysis with social media and big data. The following will contain the aspects of the reading that are deemed most relevant to the paper.

2.2 At The Beginning

Remote sensory system is the most common method for monitoring traffic. The resulting data is in the form of traffic volumes rather than speed or travel time. Travel time is estimation is down to its most common method Kalman Filtering. Kalman filtering, one of the most advanced methods in modern control theory. This method was initially proposed in 1960 by Kalman R.E. Stephanedes (1983) compares two very well established methods for predicting traffic flow and volume taken from the Kalman Filter theory and the other is UTCS-2 (Urban traffic control system) [6]. The paper explains the mathematical applications mostly deployed today in calculating speed and travel time in Urban Traffic Control. An evolution of techniques are provided to give the reader some background on prediction methods then follows that with a detailed analyses of

UTCS-2 using average prediction error and average error. As result many wireless sensor networks that are installed in cities are measure volume. Algorithms based on Kalman Theory for state space control measures volume to calculate travel time. These calculations a not 100% accurate but is a very common technique which it algorithms has been modified and improved over a long period of time and has been accepted as the best way of measuring travel time. The reason for measure volume and not travel time is to account for traffic signals and vehicles not completing routes.

2.3 Forecasting Time Series

Autoregressive Integrated Moving Average **(ARIMA)** is the most common approaches taken for forecasting travel time. In 1983 [2] outlines the variations of the ARIMA that can be seen in 2.2 .

Research into traffic prediction is a common use case around a time series problem. Autoregressive Integrated Moving Average (ARIMA) and Neural Networks are algorithms the appear to perform best in this area. For example in 2008, Dehuai Zeng et el explores the variations of the linear model ARIMA and non-linear Neural Network [7] and in 2010 claims support vector regression model (SVR) has been widely used to solve non-linear time series problems [8].

The models are modified to cater for the randomness of the so called unknown factors that effect traffic. This is also known as ARIMA-GARCH. GARCH is algorithms and models that account for the errors. Some of the random factors have been investigated such as weather and road incidents [9–11].

The core of most traffic prediction analysis is with time-series data model built from historical data as discussed in by Stephanedes (1983) [12].

A study in 2008 Dehuai Zeng et el compares the artificial neural network, ARIMA, and a hybrid model ANN-ARIMA, see figure 2.1 and table 2.1.

Dehuai Zeng et el parametrises ANN with the ARMA model (BPNN) and the hybrid model is an extension of BPNN by using its predictions of error terms for the ARIMA model [7].

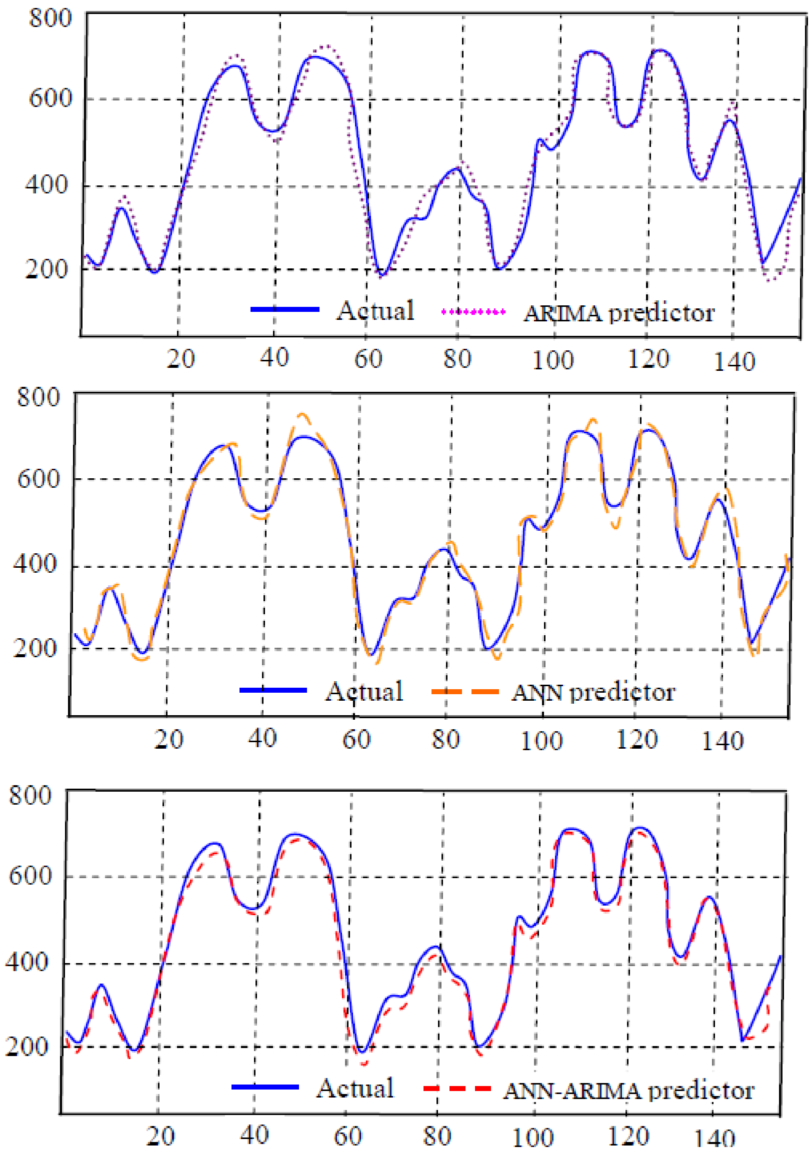


FIGURE 2.1: Comparison of predicted travel flow with different models

TABLE 2.1: Performance comparison of different predictor

Predictor	rmerr%	marerr%	rmsrerr%
ARIMA	0.92	4.26	12.44
BPNN	0.89	3.94	11.64
Hybrid	0.58	2.34	5.68

As technology has improved, roads networks have got better and car safety has improved. The number of road incidents decreased and historical data can be accessed easier making predictability of traffic delay more accessible to research [2]. With this the ARIMA models have evolved. V. Gavirangaswamy et el takes ARIMA and variations of the model.

TABLE 2.2: Summary of ARIMA variations [2]

SARIMA	Seasonal ARIMA	Good for data with short range recurring pattern
FARIMA	Fractional ARIMA	Considers recurring pattern over long ranges
MARIMA/ARIMAX	Multivariate ARIMA	Includes other time series as dependent variable
k-factor GARIMA	Gegenbauer Polynomials ARIMA	Accounts for both the short-range and longrange dependencies considering different K data frequencies
Switching ARIMA	Different ARIMA models are fitted	Apply different ARIMA for different characteristic

The historical data from Metro Detroit was aggregated hourly from the years 2009 to 2011. Initial time series chart showed the presence of seasonal data which was ideal for SARIMA. The scoring mechanism used was root mean squared error (RMSE). Using SARIMA the performance of the test improved by 5 % over ARIMA. ARIMA-GARCH model's predicted result is improved by 40% 2.3. The application of this model can be used both for short time traffic prediction and offline. The generation of the model is computationally expensive. The use of some modern big data techniques and technologies would be of great benefit to such implementation [4].

TABLE 2.3: Comparison of performance using RMSE of ARIMA, SARIMA, and ARIMA-GARCH with historical values (HV) on Downtown Street (DS), State Highway (SH), Interstate Highway (IH) [2]

HV	DS	SH	IH	DS	SH	IH	DS	SH	IH
300	367.29	375.55	143.68	346.52	346.67	140.91	212.82	251.86	89.57
500	374.42	340.71	133.92	361.1	316.53	126.52	214.09	207.26	86.08
800	339.82	355.63	142.42	346.53	347.3	142.33	214.56	207.3	88.33

2.4 Effects of Weather on Traffic

In an effort to improve traffic prediction accuracy, much research has been done to add variables to historical traffic data such a weather conditions. There is little doubt that weather conditions are correlated in some manner to traffic times and volume. According to Stephen Dunne and Bidisha Ghosh in 2013 *"Rainfall influences traffic conditions and, in turn, traffic volume in urban arterials"*. Therefore when possible the data model should include weather variables when building a prediction algorithm for traffic conditions. Stephen Dunne and Bidisha Ghosh show that using stationary version of Discrete Wavelet Transform (DWT) called SWT for a forecasting model can show correlation between the traffic volume and weather conditions outperforming Artificial Neural Network for the same tests. The studies build a traffic volume data built on Kalman Filtering. A structure of SWT is used to create a weather neurowavelet traffic forecasting system. The neurowavelet (SWT) prediction algorithm is proposed for forecast hourly traffic flow while also accounting for rainfalls levels. The study uses the wavelet form where other research uses variations of moving average. Ideally comparisons between moving average of the rather other wavelet forms would be more ideal. The study shows that rainfall has an impact on traffic flow and that an algorithm as results in figure below display [11].

Table 2.4: Comparison of wet and dry conditions on traffic flow

TCS 106 SWT-ACNN Model		
Overall MAPE	Dry Period	Wet Period
9.0936	10.6463	4.4362
TCS 106 Standard-ANN Model		
Overall MAPE	Dry Period	Wet Period
14.1061	16.5664	6.7254
TCS 125 SWT-ACNN Model		
Overall MAPE	Dry Period	Wet Period
8.0082	9.9116	2.2979
TCS 125 Standard-ANN Model		
Overall MAPE	Dry Period	Wet Period
13.3406	15.9555	5.4958

The result do not take into account the seasonality or trends of the traffic data. Traffic volumes differ on days of the week and times of the day. Keay and Simmonds investigated the influence of weather variables with road volume in 2004 [9]. The authors make a big effort in comparing trend and seasonality data in analysing results of basic regressions models. They split day-time and night-time data in understanding traffic volume and compare it to daily data. They also compares a multitude trend separation i.e. separate each day Monday-Friday and Saturday/Sunday and include school and public holidays etc. They found that Rainfall plays the big influence in traffic volume. High rainfall and colder weather decreases traffic at night-time and cooler months but highlight day-time volume stays the same with weather conditions. It is suggested the reason for this is that people need to travel to work and schools where optional activities decrease with harsher weather conditions. The study shows that there is a correlation between cool and wet weather and traffic volume. Traffic volume decreases in cool wet conditions. On week days the reduction in traffic volume is minimal at 1% compared to the 17% reduction on Sundays, showing that the necessity for people to get to work or similar activities is great. The analyses was all done using stepwise standard linear regression against season, weekly trend traffic volume data and weather variables [9].

2.5 Spatial Techniques

In a study Zhang (2012) uses a method of traffic clustering to group road points that are spatially and time related. This is a way of reducing the amount of computation of necessary. Neural Network was the proposed prediction mechanism. They mention future investigation is needed for improvements in accuracy but the main focus of the exercise was to provide the clustering approach. They propose their own *online traffic clustering algorithm* by clustering combination road point of similar dynamics. This is certainly a good consideration for a option avoid high computation cost in a bid data solution. The clustering algorithm is compared against Bayesian Neural Networks [13].

Road in networks are correlated both spatially and temporally. Roads volumes influence the travel times of its neighbours. Upstream bare an obvious significance and distant roads are insignificant. A number of models have been tested to improve the predictive value of traffic volume. In many cases traditional forecasting models have been used including Holt Winters and Multivariate Structural Time Series. In 2012 Yousef-Awwad Daraghmi et el compared Naïve Bayes Regression against the forecasting models. The proposed method used a series of lags tested over a number of different time's intervals using stepwise forward elimination in adding the number of variables to be included into the model until the differences irrelevant [14].

2.6 Social Media

Endarnoto et al wrote a paper on "Traffic Condition Information Extraction & Visualization from Social Media Twitter for Android Mobile Application" (2011). The research devised a model using text data mining techniques to extract traffic events in Jakarta. The experiments used tweets from a "TMC Polda Metro Jaya", the Twitter of National Traffic Management Center of Indonesia. Twitter account which suggested that the data extract conformed to a semi-structured text. In this case Part of Speech tagging played a pivotal role in the results. The main cause of disruption of result was due to the prediction of location is a 'From' location or a 'To' location. The experiment did however use a simple model that could be used for not just traffic event where it extract date/time, location to/from and condition. The source of tweet data is reliant on the quality of information from the user. In this case it is the national body the reports on metropolitan information. This is realistic situation for most known cities and in turn makes the study relevant in many cases. The study used the sequential order of the part of speech names, see figure 2.2, based on the POS dictionary in table 2.5.

```
AT NP V NP AJ V
AT NP V NP N AJ
AT NP V NP N AJ V
```

FIGURE 2.2: Part of speech sentence pattern

TABLE 2.5: Part of Speech Names

POS	POS	Name	Example
1	AJ	Adjective	Ramai (crowded), Macet (jammed)
2	AT	Adjective	Time 06:50
3	AV	Adverb	Sangat (highly)
4	CJ	Conjunction	Dan (and), Lalu (then)
5	N	Noun	Lalin (traffic), Arus (stream)
6	NP	Noun	Place Pondok Indah, Bintaro
7	P	Preposition	Di (at), Ke (to), Dari (from)
8	V	Verb	Merayap (crawling), Terjadi (happening)

The main obstacle in this research is that tweets that do not conform to these rules which they call, Out of Rules and Out of Vocabulary and it is handled by using a POS "indicator". The results of the simplified rules in figure 1 are at best 70% from the tested run in the experiment. [15]

Alternatively to Sri Krisna Endarnoto et al approach, Bei Pan et el look into using TF-IDF approach to against classified traffic related tweets. The tweets themselves are not necessarily traffic related but also of social events that may have an impact on traffic. In the study one such event determined was a wedding event exhibition.

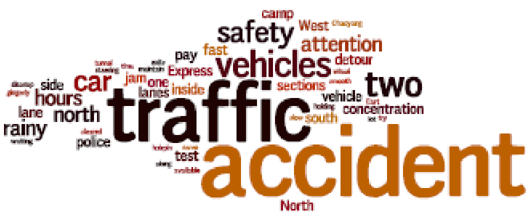


FIGURE 2.3: Traffic word cloud, event features

FIGURE 2.4: Wedding word cloud, event features

The research also builds a corpus based on a predefined source the Beijing Transportation Bureau to extract features relevant to traffic. The research provides limited details on how tweets on the event in 2.3 were extract or if any document tokenizing implementation where used. The tweets were retrieved from an area where traffic anomalies reported by local authorities but no traffic incidents had occurred. [10]

2.7 Big Data

Big Data is data that is so large and complex where it becomes problematic. The problems largely focus on the four V's of Big Data [16].

- Volume
  Back in the year 2000 a PC might have had 10 gigabytes of storage. Social Media sites such as Twitter and Facebook consumes 500 terabytes a day.

- Velocity
  This mostly relates to the capturing of real-time data at high speed. In particular Twitter is a good example of real-time data monitoring. As well as consuming real-times messages from users, they are exposing APIs that allow the public leverage on this data. Internally Twitter is consuming an quickly process as much or even more than 500 terrabytes of data.

- Variety
  Big Data needs to be able handle a variety of data type such as spatial attributes, graphic, audio and video, and unstructured text. Traditional RDBMS were designed to handle smaller volumes of structured data.

- Veracity
  Is a more recent adage of the V's. Is the term for using the data analysis for decision making, problem solving and knowledge outcomes. With this ensuring data quality.

Traditional database systems are designed to operate on a single machine. This provides a limitation to the scalability of the solution as capacity is finite. The use of application and development practices have become agile, as production have evolved onto the cloud for multi-tenet user base the database needs to grow horizontally the more users there are using the system. Big Data databases, such as MongoDB, solve these problems and provide companies with the means to create tremendous business value [5].

2.7.1 Map Reduce

Map Reduce is a technique that plays a massive role in the volume and velocity of big data. Map Reduce is elastic scalable, promotes efficiency and high availability [17, 18].

In some of the works mentioned in this paper it has a common problem with detailing with large volumes data from traffic observations and twitter data [2, 7, 10]. In recent year the term *Big Data* has come into fruition. Vinay Gavirangaswamy et el [2] mentions with regards the tests *took around 220 computational hours to run these experiments* on a machine with 8 gigabytes of RAM. Big Data is data that is so large and complex where it becomes problematic. Brito et al proposed an approach called *StreamMapReduce* a task that is considered a Big Data problem. The characteristics of Big Data are mainly Volume, Variety, Velocity and Veracity[19, 20]. The research claims that its mechanism can *allow a hundred fold improvement in response time* and *a ten-fold per node throughput increase in comparison to Hadoop*. The concept behind the approach acts as an improvement on Event Stream Process and MapReduce by filtering out data that is not relevant or considered duplicates. Mostly the improvements in performance are down to aggregation of data which inevitably lose some data or mapping documents that contain the same class data. The study does highlight that MapReduce and/or Event Stream Processing does not answer all Big Data problems and *StreamMapReduce* is a solution to some use cases [21]. However in 2013, Duckwon Chung et el apply big data technology Hadoop and HBase to analyse real-time traffic collisions from a number of different sources, traffic information, social sites, mobile phone GPS signals. One terrabyte of data was extracted from these sources over a ten year period. The solution involves multiple data nodes for consume the observation data distributed by a master data node , see figure 2.5. The master node decides which of the nodes to send the observation based on an index. In this case location called *detectors* was the deemed the mostly fitting index. The nodes then map and reduce the date which is a way of aggregating the data for further analyses. Once the aggregation is processed algorithm can be generated, see figures 2.6 2.7 [1]

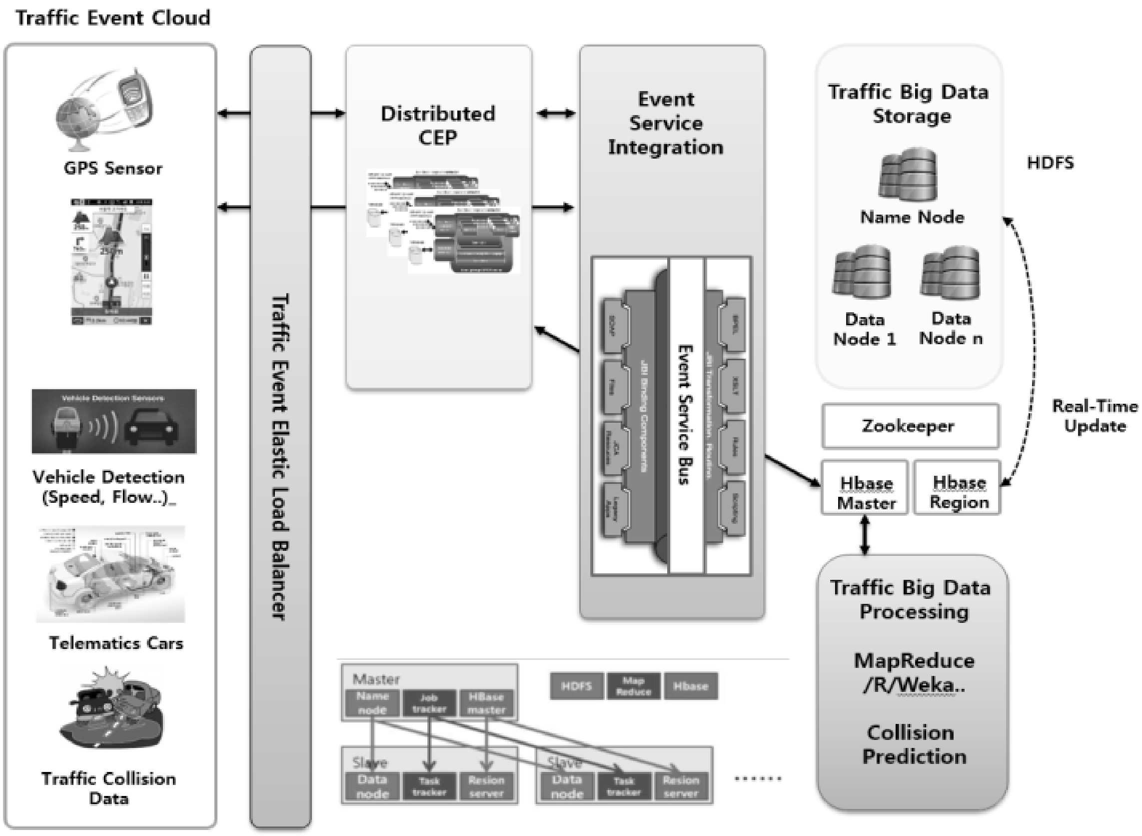


FIGURE 2.5: Proposed architecture for Big Data solution [1]

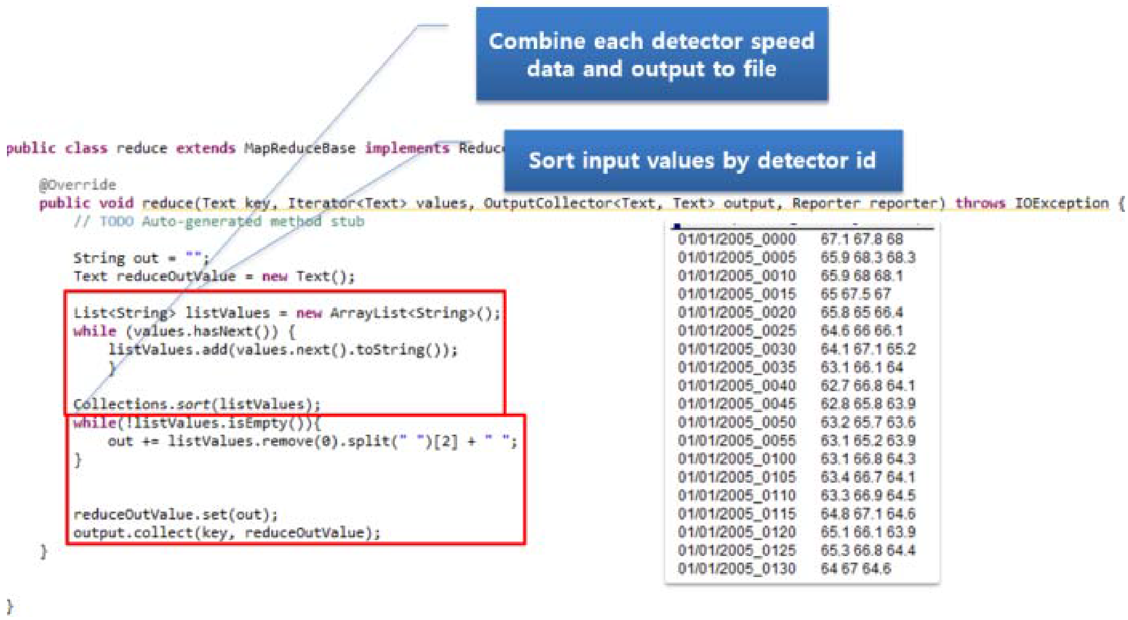


FIGURE 2.6: Map and Reduce [1]

From	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	1.98E-05	0	0	0	0	1.73E-05	0	0	0	0	0	0	0	0	0
10	0	2.24E-05	5.60E-06	6.31E-06	0	2.90E-05	0	3.44E-05	2.09E-05	0	0	1.88E-05	2.77E-05	0	0	0
15	0	4.91E-05	0	7.12E-06	5.71E-06	6.46E-06	3.31E-06	7.26E-06	9.70E-06	0	1.44E-05	0	0	2.07E-05	0	0
20	0	0	5.71E-06	4.61E-06	6.03E-06	4.02E-06	5.64E-06	3.85E-06	1.07E-05	6.89E-06	0	5.01E-06	0	3.01E-06	0	0
25	0	0	0	0	3.41E-06	3.91E-06	4.52E-06	2.81E-06	6.42E-06	2.45E-06	8.45E-06	1.91E-06	5.94E-07	5.32E-06	2.34E-06	0
30	0	0	0	1.95E-06	5.22E-06	5.97E-06	4.63E-06	3.35E-06	3.29E-06	3.46E-06	4.13E-06	2.31E-06	3.50E-06	2.78E-06	2.26E-06	0
35	0	0	0	8.28E-06	6.31E-06	4.73E-06	3.74E-06	2.52E-06	2.62E-06	1.37E-06	3.36E-06	1.00E-06	2.86E-06	2.10E-06	2.80E-06	0
40	0	0	0	8.78E-06	5.91E-06	5.88E-06	3.06E-06	4.38E-06	5.60E-06	4.05E-06	2.24E-06	2.34E-06	2.18E-06	1.70E-06	2.95E-06	0
45	0	0	0	1.46E-05	6.70E-06	4.97E-06	2.46E-06	4.91E-06	3.72E-06	2.28E-06	2.01E-06	1.56E-06	1.55E-06	1.98E-06	4.12E-06	0
50	0	0	0	0	2.71E-06	5.85E-06	6.94E-06	2.12E-06	5.45E-06	1.59E-06	1.60E-06	5.64E-07	5.14E-07	0	1.09E-06	0
55	0	0	0	1.23E-05	7.53E-06	5.02E-06	1.29E-06	2.65E-06	3.83E-06	2.09E-06	8.87E-07	2.24E-07	5.13E-07	5.19E-07	4.10E-07	0
60	0	0	2.36E-05	0	4.58E-06	1.13E-06	4.36E-06	3.93E-06	7.18E-07	1.41E-06	7.69E-07	4.32E-07	3.24E-07	2.98E-07	3.91E-07	0
65	0	0	0	1.67E-05	3.84E-06	1.95E-06	9.51E-07	1.69E-06	1.79E-06	1.36E-06	1.90E-07	2.57E-07	2.84E-07	2.39E-07	1.94E-07	0
70	0	0	0	0	0	2.11E-05	5.27E-06	0	0	0	0	3.46E-07	3.28E-07	3.08E-07	3.10E-07	0
75	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

FIGURE 2.7: Collision Prediction [1]

It is generally understood that analysing streaming Twitter streams is the volume of data consumed by an application. With a large Twitter data set McCreadie et al experiments with a *divide and conquer* technique to efficiently scale big data streams at estimated thousands of tweets per second [4]. McCreadie seem certain that MapReduce and traditional DBMS are not well suited for real-time Twitter streaming processing, especially DBMS where it uses a 'store-then-process' method for dealing with data. The experiment uses a platform called *Storm*, which is now part of the Hadoop stack to handle real-time streams of data. It works but releasing short batches of streaming data to different nodes. Within its Event Detection Topology it uses algorithm for clustering data that are similar using a Distributed Lexical Key Partitioning (DLKP) to cluster data documents in groups. DLKP is term for Storm to implement a Local Cosine Distance calculation. The paper gives a good approach for handling unstructured twitter data with unknown key attributes. [4]

2.7.2 Analytical dashboards

With Big Data is not all about writing and reading data. It is necessary to provide analytical views of data. It is difficult for users to read volumes of data. Analytic dashboard is technique for display analytical information. In 2013 Kristopher Reese et al explain the importance of using visual dashboards for analysing large amounts of data [22]. Examples are provided to show how best use colours and spatial information, see figures 2.8 and 2.9

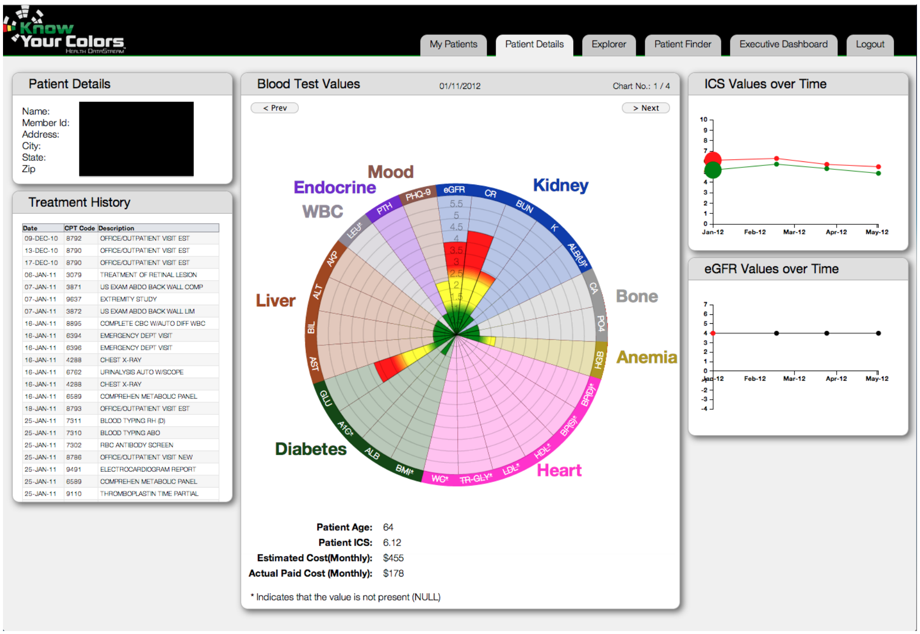


FIGURE 2.8: Big data Color Visualisation
[22]

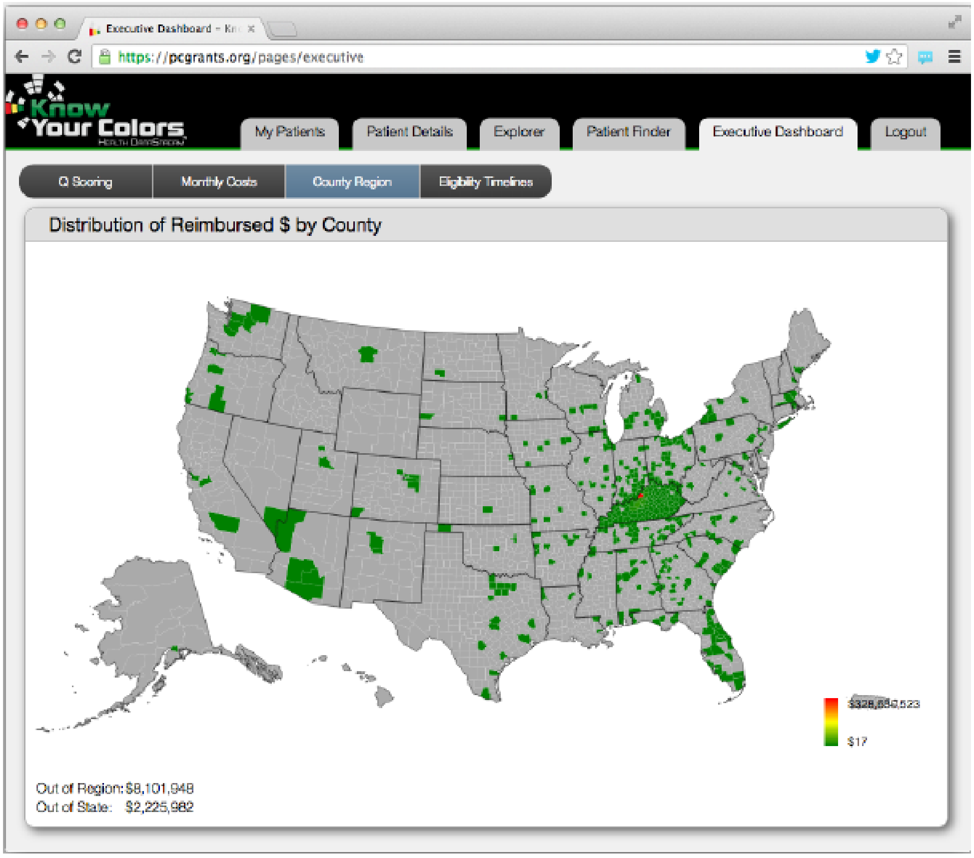


FIGURE 2.9: Big data Spatial Visualisation
[22]

2.7.3 NoSQL

The trend to use NoSQL databases in place of relational databases have increased in certain use cases. Often requirements of data model can change frequently. In 2013, Luís A. Bastião Silva et al explain that document-based databases do not have the limitation of RDBMS databases [23]. The demonstrate that Lucene, MongoDB and CoucheDB have high performance levels.

2.8 Algorithms

Linear and non-linear algorithms have been used for forecasting time series. Dr. Vincent Granville in 2014 describes linear regression algorithms and is summerized in list 2.8 [24].

- Linear regression
  Is the oldest regression model. Sensitive to over-fitting and outliers.
- Logistic (Poisson or Cox) regression
  Often used in clinical trials, scoring and fraud detection and is considered.
- Ridge regression
  Regression with constraints on the coefficients. Not as sensitive to over-fitting as the Linear regression model.
- Lasso regression
  Same as Ridge except it automatically uses variable reduction.
- Logic regression
  Sets all the variables to binary. Can be more robust than logistic regression. Often used in fraud detection.
- Bayesian regression
  Assumes prior knowledge of the coefficients. Flexible compared to linear regression and the error must contain a normal distribution.
- Logistic regression
  Compares the relationship between a dependent variable and one or more independent variables. Is analogous to linear regression.

Other Non-linear algorithms as mentioned in section 2.3 can be used in regression models as Support Vector Regression as long as the kernal is set to RBF by Monte Carlo approximation of its Fourier transform while Stochastic Gradient Descent (SGD) can be used as an Artificial Neural Net when using back-propagation. As Neural Net is know to perform slowly SGD performs well for large scale learning [25, 26].

The Python Sci-Py kit provide algorithm for regression, classification, clustering and dimension reduction, see figure 2.10.

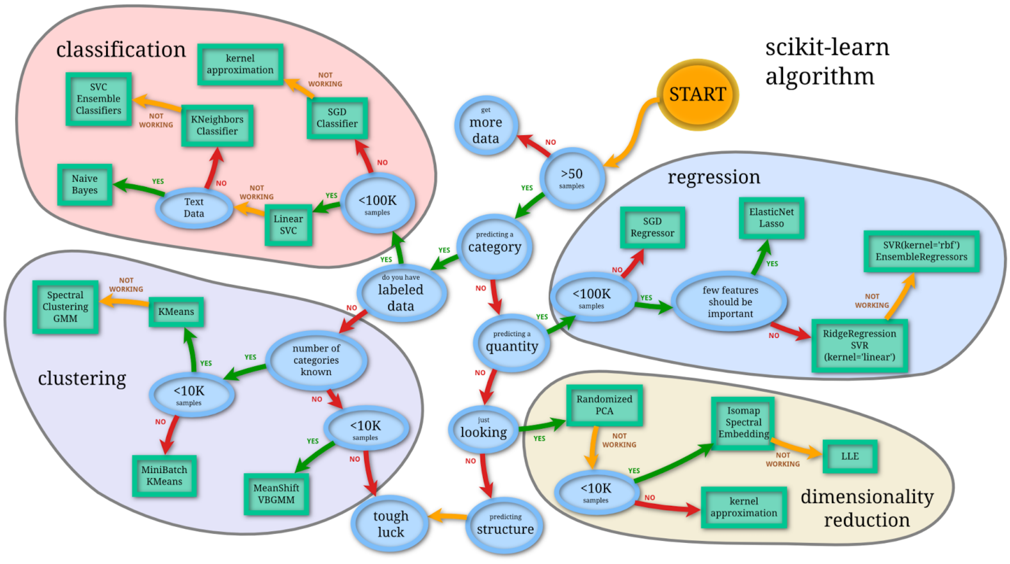


FIGURE 2.10: Sci-Py Algorithms
[26]

Some of the linear algorithms mentioned perform well with handling of noise with many parameters available to manipulate the coefficients through techniques as normalisation and setting boundaries to the independent variable known as upper and lower bound limits 2.8.

2.9 Conclusion

State space control is the underlying method for measuring traffic volume and speed. Kalman Theory is the most prominently in modern day algorithms and is widely used in traffic monitoring systems. These methods do not guarantee absolute accuracy but is the most widely used method in traffic estimation systems. Based on the data generated from the monitoring systems forecasting methods are implemented to form traffic prediction. Many of the model used for prediction are ARIMA and GARCH. GARCH is used to manage noise in the traffic data sets. For the purpose of the exercise algorithms from Python SciPy will be compare. Python SciPy has a wide variety of Linear models, some that handle noise in different ways and only limited options Non Linear algorithm. Perceptron will be used for Artificial Neural Net and Support Vector Regression (SVR).

Chapter 3

Data Understanding

3.1 Introduction

This section describes the information made available from the open data sources Dublinked, Wunderground and Twitter prior to data collection. The objective of chapter is to provide a background on the information available.

3.2 Traffic Data Sets

TRIPS data is comprised of three datasets which is made available through an open data website DubLinked [3]. Journey times are provided some of the main routes across Dublin City. Each route consists of a number of links, each link is a pair of geo-referenced traffic control sites. DubLinked distribute maps that can be imported into Open Street Map and Google Maps known as shapefiles and KML files, see figure 3.2. With this the locations of the traffic control sites marked by yellow pins in in google maps along designated routes. The purpose of the traffic control sites is to monitor traffic volume using sensors.

FIGURE 3.1: DubLinked Google Map

3.2.1 Traffic Data

Historical traffic data is stored via the DubLinked TRIPS Archive Directory [3]. The archive is a HTTP directory-list of binary zip files. Each binary zip file contains one day of historical observation data in a CSV format 3.2.1. The file name is marked with the date in the format *day-YYMMDD.csv.bz2*

```
Timestamp, Route, Link, Direction, STT, AccSTT, TCS1, TCS2
----------------------------------------------------------------
20131213-1339, 4, 3, 2,  18, 168,125,470
20131213-1339, 4, 4, 1,  35, 204,125,667
20131213-1339, 4, 4, 2,  43, 211,667,125
20131213-1339, 4, 5, 1,  22, 226,667,422
20131213-1339, 4, 5, 2,  19, 230,422,667
20131213-1339, 4, 6, 1,  49, 275,422,151
20131213-1339, 4, 6, 2,  37, 267,151,422
```

LISTING 3.1: File day-20131213.csv.bz2$day-20131213.csv line 501052-501059

TABLE 3.1: Observation Attributes

Attribute	*Description*
Timestamp	Date time of observation YYYYMMDD-HHmm
Route	Road with 1 or more observed links
Link	Segment of road between 2 control sites
Direction	Direction of flow of traffic

3.2.2 Junction Data

DubLinked provide junction data which relate to the Traffic Control Sites. Each observation recorded hold identifiers about the two traffic control sites called TCS1 and TCS2 as seen in listing 3.2.1. The details are provided in three formats, CSV, KML, and Shape file the example privided in junctions.csv 3.2.2. Between the traffic control sites TCS1 and TCS2 is the value of travel time is estimated and in further references in the research will be know as Observed Location (OL).

TABLE 3.2: Junction Attributes

Attribute	*Description*
SiteID	Relates to the identifier TCS1 and TCS2
X	Longitude [Irish Grid (IG; EPSG:29902) Coordinate Value]
Y	Latitude [Irish Grid (IG; EPSG:29902) Coordinate Value]
Location	Name of Road

```
SiteID, X,  Y,  Location
______________________________________________________________

125,  312666, 236290, NAVAN RD NEPHIN RD
126,  315128, 233640, BULL ALLEY ST NICHOLAS STREET BRIDE ROAD
127,  314842, 235872, NORTH CIRCULAR ROAD CABRA ROAD
128,  315942, 235743, NORTH CIRCULAR ROAD BELVEDERE ROAD
129,  316239, 235647, NORTH CIRCULAR ROAD FITZGIBBON ST
130,  313378, 234887, NCR INFIRMARY RD
131,  313112, 239100, NORTH RD MELLOWES RD
132,  317281, 235971, NORTH STRAND RD EAST WALL RD
```

LISTING 3.2: Example data junctions.csv

3.2.3 Routes Data

DubLinked provide route data which relate to a section of road made up of a number of links. A link is the length of road between to Traffic Control Sites (TCS1 and TCS1) and mentioned above. Each observation recorded hold identifiers route and link, see 3.2.1. The details are provided in three formats, CSV, KML, and Shapefile with the routes.csv listed in 3.2.3

Attribute	*Description*
Route	Stretch of road being monitored
Link	Segment of the Route
Direction	Direction of traffic along link
TCS1	Control point for traffic entering link
TCS1	Control point for traffic exiting link
WKT	Irish Grid Coordinates

TABLE 3.3: Route Attributes

```
Route,Link,Direction,TCS1,TCS2,WKT
1,1,1,6006,2031,LINESTRING(321909 228333 comma 321106 228863)
1,1,2,2031,6006,LINESTRING(321106 228863 comma 321909 228333)
1,2,1,2031,6003,LINESTRING(321106 228863 comma 320545 229272)
1,2,2,6003,2031,LINESTRING(320545 229272 comma 321106 228863)
1,3,1,6003,6008,LINESTRING(320545 229272 comma 320380 227100)
1,3,2,6008,6003,LINESTRING(320380 227100 comma 320545 229272)
1,4,1,6008,1125,LINESTRING(320380 227100 comma 319684 229203)
1,4,2,1125,6008,LINESTRING(319684 229203 comma 320380 227100)
```

LISTING 3.3: Example data routes.csv

3.3 Weather Data

Much research has shown the impact weather conditions has had on traffic[9]. This section will cover the weather data extraction process. The weather data itself is taken from an open source website Weather Underground [27]. Wunderground is a provider of weather station data. The weather stations are owned by the general public. It this paper three stations are select as a source of weather data, see appendix C for a list of stations. Weather conditions move and change over time. Wet weather conditions can be specific to a small area at one time and not guaranteed the greater Dublin area with experience wet condition all at once. For example rain takes time to travel. This means that rain effects traffic at different times and locations. The weather station are located in the North, West and South Dublin 3.3.

Id	*Location*
ICODUBLI2	Lucan, Co Dublin West
ILEINSTE8	Blackrock, Dublin 8, South
IDUBLINC2	Artane, Dublin 5, North

TABLE 3.4: Weather Stations

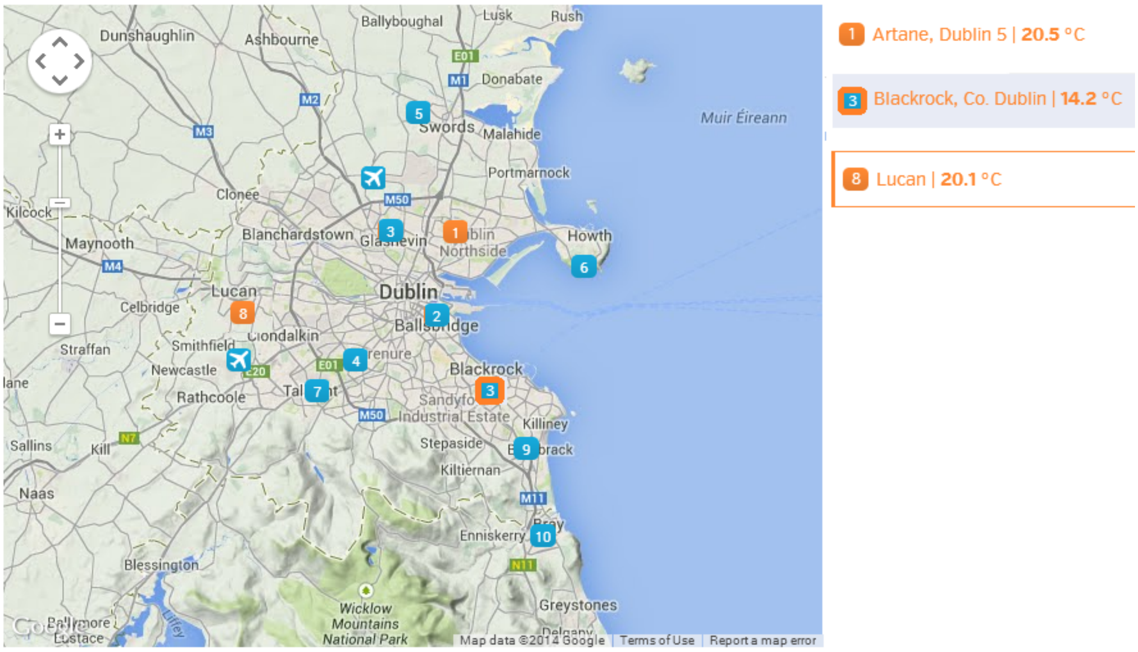
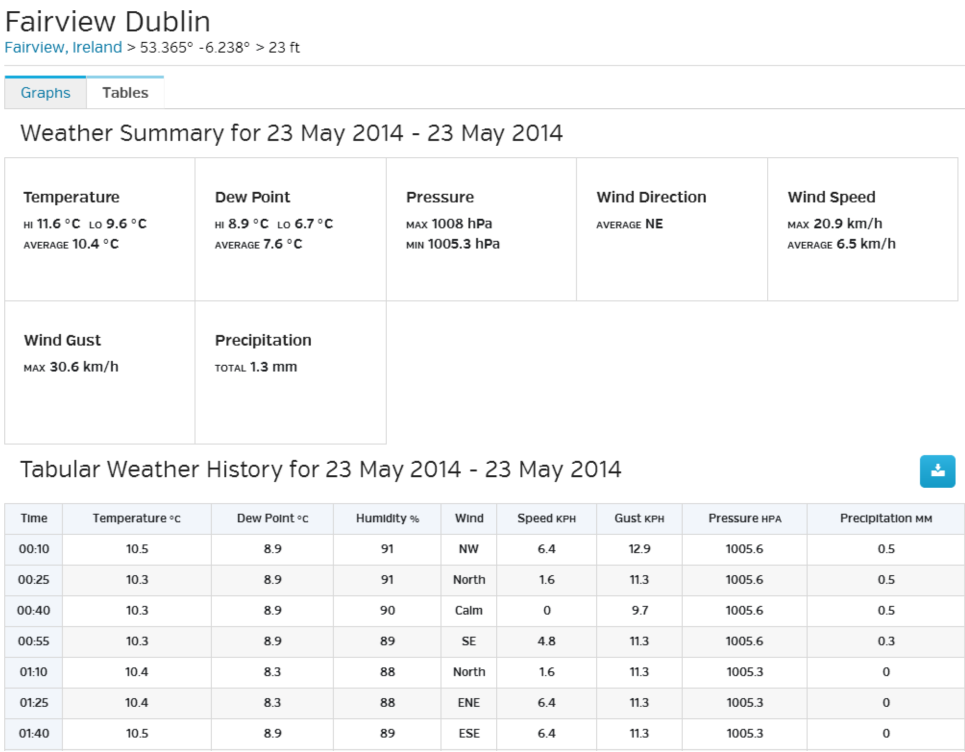


FIGURE 3.2: Open Data Weather Stations Dublin

The number icons in figure 3.3 are weather stations available on the Wunderground website provide access to historical weather of each day and location separately via the web [27]. Highlighted in orange are the weather station capture for the purpose of the paper.

http://www.wunderground.com/personal-weather-station/dashboard?ID=IDUBLINF2#history/data/s20140423/e20140423/mtoday



Fairview Dublin
Fairview, Ireland > 53.365° -6.238° > 23 ft

Graphs Tables

Weather Summary for 23 May 2014 - 23 May 2014

Temperature	Dew Point	Pressure	Wind Direction	Wind Speed
HI 11.6 °C LO 9.6 °C AVERAGE 10.4 °C	HI 8.9 °C LO 6.7 °C AVERAGE 7.6 °C	MAX 1008 hPa MIN 1005.3 hPa	AVERAGE NE	MAX 20.9 km/h AVERAGE 6.5 km/h

Wind Gust	Precipitation
MAX 30.6 km/h	TOTAL 1.3 mm

Tabular Weather History for 23 May 2014 - 23 May 2014

Time	Temperature °C	Dew Point °C	Humidity %	Wind	Speed KPH	Gust KPH	Pressure HPA	Precipitation MM
00:10	10.5	8.9	91	NW	6.4	12.9	1005.6	0.5
00:25	10.3	8.9	91	North	1.6	11.3	1005.6	0.5
00:40	10.3	8.9	90	Calm	0	9.7	1005.6	0.5
00:55	10.3	8.9	89	SE	4.8	11.3	1005.6	0.3
01:10	10.4	8.3	88	North	1.6	11.3	1005.3	0
01:25	10.4	8.3	88	ENE	6.4	11.3	1005.3	0
01:40	10.5	8.9	89	ESE	6.4	11.3	1005.3	0

FIGURE 3.3: Wunderground Day View

The complete data set can be accessed in CSV format using the parameters described in table 3.5 with the URL below for any particular day. The URL has the identifier *ICODUBLI2* which is related to Fairview. By passing the parameters *day=21*, *month=05* and *year=2014* means the data returned is for 21st May 2014 in Fairview.

http://www.wunderground.com/weatherstation/WXDailyHistory.asp?ID=ICODUBLI2&day=21&year=2014&month=05&format=1

Parameter	*Description*
ID	Identifier for weather station
day	Day of month
year	Year
month	Month

TABLE 3.5: Weather URL Parameters

Wunderground also provide its own forecast information for the area. This is potential useful for building the non lagged weather data with the predictive model, figure 3.4.

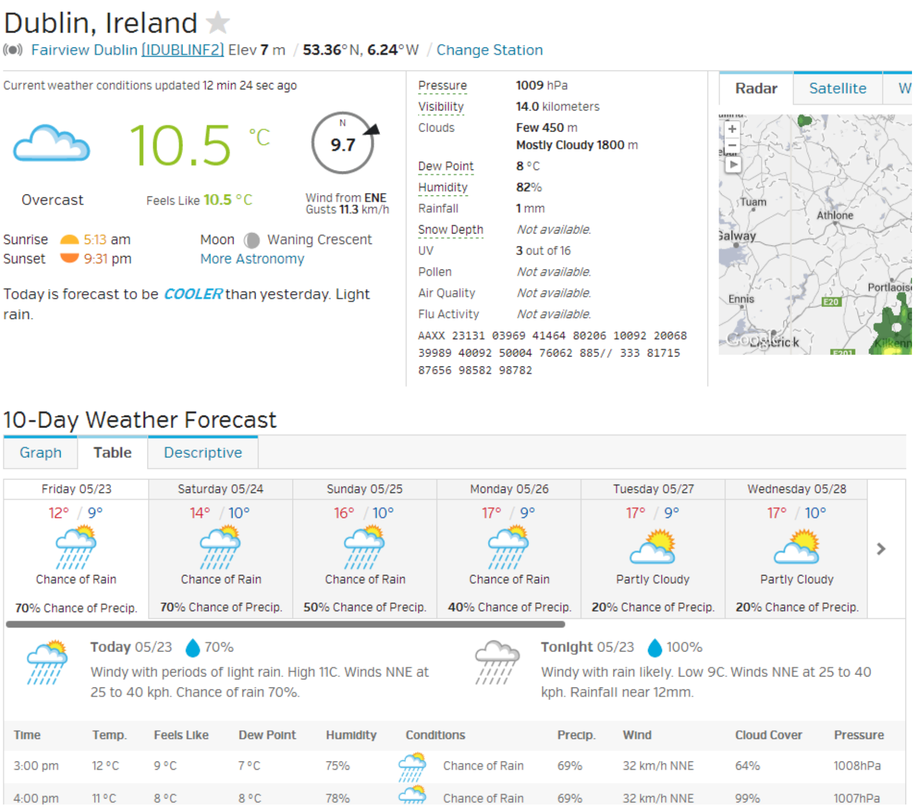


FIGURE 3.4: Wunderground Forecast View

http://www.wunderground.com/weather-forecast/IE/Dublin.htm

3.4 Twitter Data

Twitter provides an API for searching historical data based on different types of filters on attribute and/or search streaming tweets. Using the technique of capturing data from a provider of traffic related tweets to classify live streaming tweets. Known providers of traffic data such as AA Roadwatch 3.5 are a public server to provide traffic information updates.

3.4.1 Twitter User Timeline Traffic Data

The tweets from AA Road Watch and Live Drive providers do not carry any geographical information other the text data. Where the live stream carry the geographical information i.e. co-ordinates. But the text data may not provide the location. The object here is to relate traffic related tweets to the geographical position within the tweet. The difference between the two sources is that the tweets from Live Drive are re-tweets from the general public and the tweets from AA Road Watch are tweets delivered as a National Service. Although the AA Road Watch and Live Drive tweets are Traffic related it does not guarantee that all tweet contain traffic information 4.

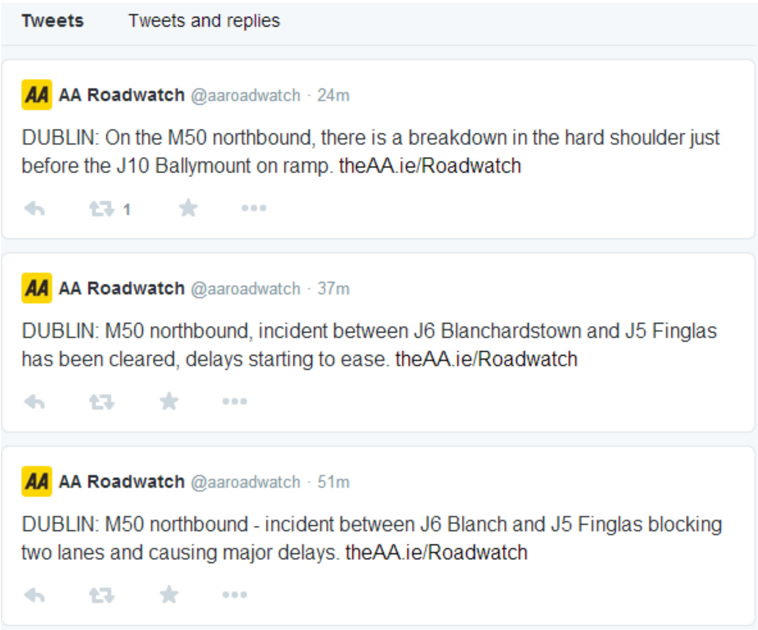


Figure 3.5: AA Roadwatch Tweets

3.4.2 Twitter Streaming Data

The Twitter Streaming API provides a service for capturing data live Tweets in real-time. The Twitter API has geographical parameters the allows for targeting specific areas around the world. In this research Dublin will be targeted, see figure 3.6. The sample of a tweet from the twitter real-time service contain user, geographic and the text, see listing 3.4.2.

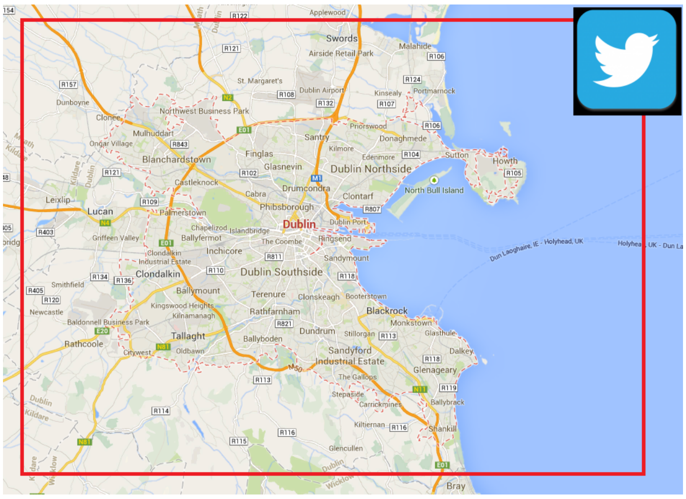


FIGURE 3.6: Monitering of Twitter Stream


```
{
    "_id" : "456013499119190016",
    "user_contributors_enabled" : "False",
    "user_notifications" : "None",
    "user_default_profile" : "False",
    "timestamp" : ISODate("2014-04-15T11:17:53.000Z"),
    "item_id" : "456013499119190016",
    "user_default_profile_image" : "False",
    "geo" : "{'type': 'Point', 'coordinates': [52.9731794, -6.0478466]}",
    "date" : "2014-04-15 11:17:53 ",
    "user_statuses_count" : "30038",
    "retweeted" : "False",
    "user_profile_background_image_url_https" : "https://abs.twimg.com/
    images/themes/theme10/bg.gif",
    "place_id" : "577c65949ddabbc9",
    "user_url" : "None",
    "user_follow_request_sent" : "None",
    "user_profile_sidebar_fill_color" : "F6FFD1",
    "user_is_translation_enabled" : "False",
    "user_profile_background_image_url" : "http://abs.twimg.com/images/
    themes/theme10/bg.gif",
    "user_listed_count" : "15",
    "user_id" : "398974774",
```

```
"user_profile_background_color" : "382D8A",
"user_profile_image_url_https" : "https://pbs.twimg.com/profile_images/3623831526/fe8cefe137693b064bba63cad65403cc_normal.jpeg",
"coordinates" : "{'type': 'Point', 'coordinates': [-6.0478466, 52.9731794]}",
"user_id_str" : "398974774",
"user_profile_background_tile" : "True",
"user_name" : "Alistair ",
"user_is_translator" : "False",
"user_verified" : "False",
"place_full_name" : "Wicklow",
"user_location" : "Wicklow, Ireland",
"user_created_at" : "Wed Oct 26 20:21:46 +0000 2011",
"user_geo_enabled" : "True",
"source" : "<a href=\"http://twitter.com/download/android\" rel=\"nofollow\">Twitter for Android</a>",
"place_contained_within" : "[]",
"place_attributes" : "{}",
"user_time_zone" : "Dublin",
"user_friends_count" : "1663",
"place_place_type" : "city",
"user_profile_link_color" : "FF0000",
"user_profile_sidebar_border_color" : "000000",
"place_name" : "Wicklow",
"user_profile_banner_url" : "https://pbs.twimg.com/profile_banners/398974774/1397201272",
"user_favourites_count" : "11458",
"user_screen_name" : "Al_toMyFriends",
"user_utc_offset" : "3600",
"user_profile_text_color" : "333333",
"text" : "@vitaminsludge You probably teach Chinese students, so you would have some insights. @guardian @whithernow",
"user_protected" : "False",
"user_lang" : "en",
"place_country_code" : "IE",
"user_followers_count" : "843",
"user_profile_use_background_image" : "True",
"user_description" : "Irish.        Friendly.      Apathetic activist.  In a time of universal deceit telling the truth is a revolutionary act.  - George Orwell",
"user_profile_image_url" : "http://pbs.twimg.com/profile_images/3623831526/fe8cefe137693b064bba63cad65403cc_normal.jpeg",
"user_following" : "None",
"place_country" : "Ireland",
"place_url" : "https://api.twitter.com/1.1/geo/id/577c65949ddabbc9.json",
```

```
    "place_bounding_box" : "{'type': 'Polygon', 'coordinates':
    [[[-6.791799, 52.682057], [-6.791799, 53.2338548], [-5.9988317,
    53.2338548], [-5.9988317, 52.682057]]]}"
}
```

Listing 3.4: Example data routes.csv

3.4.3 Twitter Summary

The objective of using the two sources for tweets is that the user specific tweets can be used to extract features that are related to the subject matter of traffic tweets and use a similarity measure with associated rule learning to match traffic related tweets from the real-time data which contains the geographic location which is not part of the user-timeline data.

Chapter 4

Data Collection and Exploration

4.1 Introduction

In this section it is explained how big data techniques are used to store the data in an unstructured database called MongoDB. There is an emphases on *divide and conquer*. This database along Python Pandas module is used to provide faster searching queries of large document store with indexing. Using Pythons Pandas is a time series module for manipulating time series data. Pandas is has feature that aggregate and manipulate time series data for the purpose of Map Reduce which is used to move the reduced collection of data back into the NoSQL database. In the following sections the Traffic and Weather is merged for a data model that can be used for regression models. Twitter data is reduce to contain only the attribute important for traffic analysis.

4.2 Data Collection

The data for this research consists of three fundamental areas traffic, weather and twitter data. All the data can be obtained through web and open data on-line sources. In the following sections the paper will discuss the techniques used to collect and store all the data for the three areas. The techniques include web scraping, data manipulation, data quality, database storage and performance.

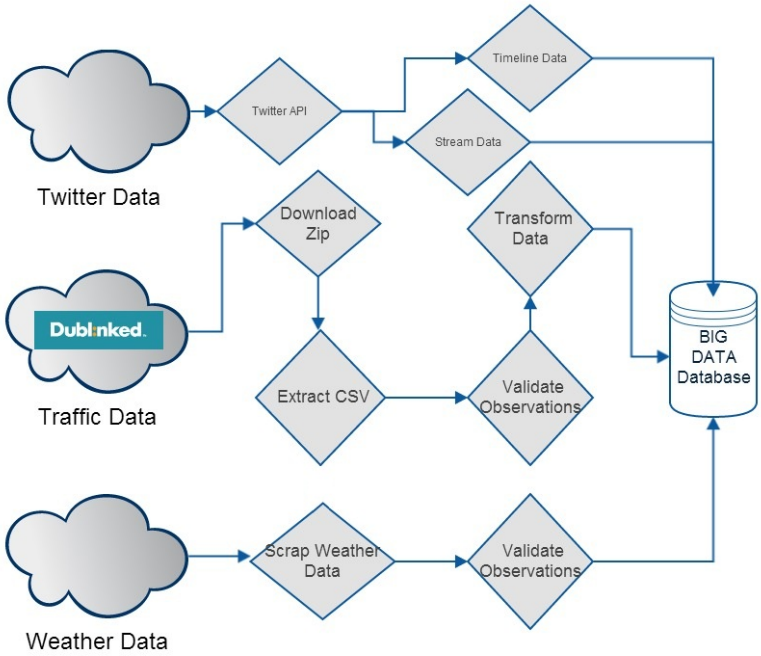


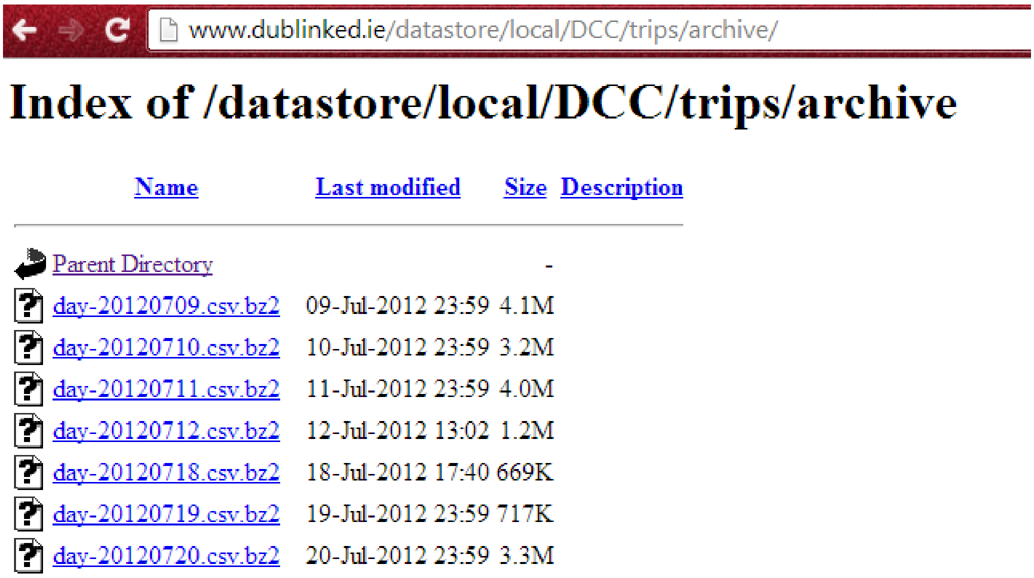
FIGURE 4.1: Extraction process of traffic observations

4.2.1 Traffic Data Extraction

The data store of traffic observations is known as TRIPS is maintained in an DubLinked Archive [3]. For each day of data the a Comma Separated Value (CSV) file is generated and placed into an BZIP2 Compressed file that can be access via an individual links [3]. The data needs validate prior to storage. A concern for the storage is the volume of data. A Big Data approach is necessary for making the data accessible and for fast queries to retrieve the data for analysis.

4.2.2 Traffic Web Scraping

Web scraping is a term used for extracting data from the web. DubLinked contains a list of the most currently available BZIP2 Compressed files, see figure 4.2 [3]. Using Python, a list of archive files are read and files are downloaded and stored into a temporary archives directory 4.2.2.



www.dublinked.ie/datastore/local/DCC/trips/archive/

Index of /datastore/local/DCC/trips/archive

Name	Last modified	Size	Description
Parent Directory		-	
day-20120709.csv.bz2	09-Jul-2012 23:59	4.1M	
day-20120710.csv.bz2	10-Jul-2012 23:59	3.2M	
day-20120711.csv.bz2	11-Jul-2012 23:59	4.0M	
day-20120712.csv.bz2	12-Jul-2012 13:02	1.2M	
day-20120718.csv.bz2	18-Jul-2012 17:40	669K	
day-20120719.csv.bz2	19-Jul-2012 23:59	717K	
day-20120720.csv.bz2	20-Jul-2012 23:59	3.3M	

FIGURE 4.2: Traffic observation archive

```python
        # check if file already exists on disk
        if os.path.isfile(outputFile) is not True:
            print("Skipping " + archiveFile)
            print("Downloading ",archiveFile)
            http_pool = urllib3.connection_from_url(archiveFile)
            rcsv = http_pool.urlopen('GET',archiveFile)
            # save data to disk
            output = open(outputFile,'wb')

            output.write(rcsv.data)
            output.close()



        if os.path.isfile("extracted/" + filename + ".csv") is not True:
            zfobj = bz2.BZ2File(outputFile, 'rb')
```

LISTING 4.1: Download Archive File

4.2.3 Traffic Indexing and Map Reduce

Once the raw data has been store locally the data needs to be store in the database so it can be read efficiently. In a traditional solution the data set of observations are inserted *as is* in a structure database table with defined columns. In listing 4.2.3 is an example of CSV records. Each archive file contains observations which amount above 3.3mb in file size and have between 900,000 and 1,000,000 records. RDBMS can limit database or table size and or even record count size.

The approach in consider the **divide and conquer** technique to organise the observations for fast retrieval. Each document in a mongoDB collection groups the observations based on its time-stamp attributes day and hour as well as the spatial location/direction, see figure 4.3.

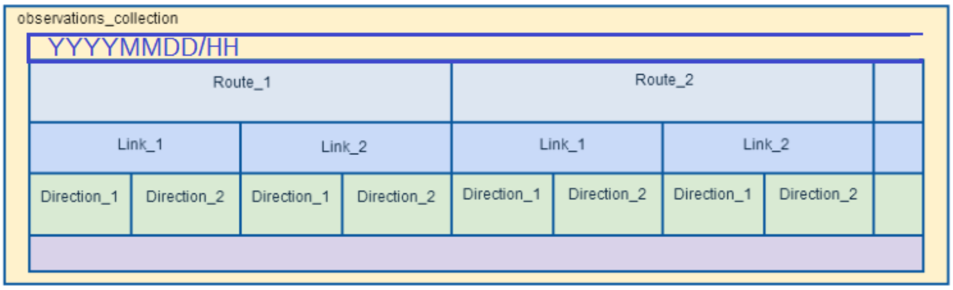


FIGURE 4.3: High Level Observation collection indexed

```
20140125-0232, 5, 5, 2,  22, 186,459,458
20140125-0232, 5, 6, 1,  87, 277,459,405
20140125-0232, 5, 6, 2,  86, 272,405,459
20140125-0232, 6, 1, 1,  24,  24,405,150
20140125-0232, 6, 1, 2,  30,  30,150,405
20140125-0232, 6, 2, 1,  22,  46,150,148
```

Listing 4.2: File day-20140125.csv.bz2$day-20140125.csv

```
{
    "_id" : "17/5/2/20140202/17",
    "hour" : "17",
    "link" : "5",
    "day" : "20140202",
    "direction" : "2",
    "item" : [
        {
            "date" : ISODate("2014-02-02T17:00:00.000Z"),
            "stt" : 30
        },
        {
            "date" : ISODate("2014-02-02T17:01:00.000Z"),
            "stt" : 30
        },
        .....
    ],
    "route" : "17"
}
```

Listing 4.3: MongoDB observation collection Reduced

As a result the data is transformed using **Map Reduce** from the CSV 4.2.3 to the NoSQL JSON format 4.2.3. The attributes in listing 4.2.3 **"_id", "hour", "link", "day", "direction", "route"** are all there to facilitate queries to the databases and are also indexes in the collection. The **"items"** is the attribute that contain all observations for the time and spatial location.

4.2.4 Weather Data Extraction

The weather data does not impose the same level of volume as traffic or twitter data. This task therefore does not impose the level of performance issues that a larger dataset can produce. In section 3.3 it is discussed where the data is available. Also it is available in CSV format. The CSV of historical weather data for each day is accessible. Each weather collection item contains all the recorded observation for that day and location.

The collection index is set to the day and location value, see figure 4.4. The **"items"** is the attribute that contain all weather observations for the time and spatial location.

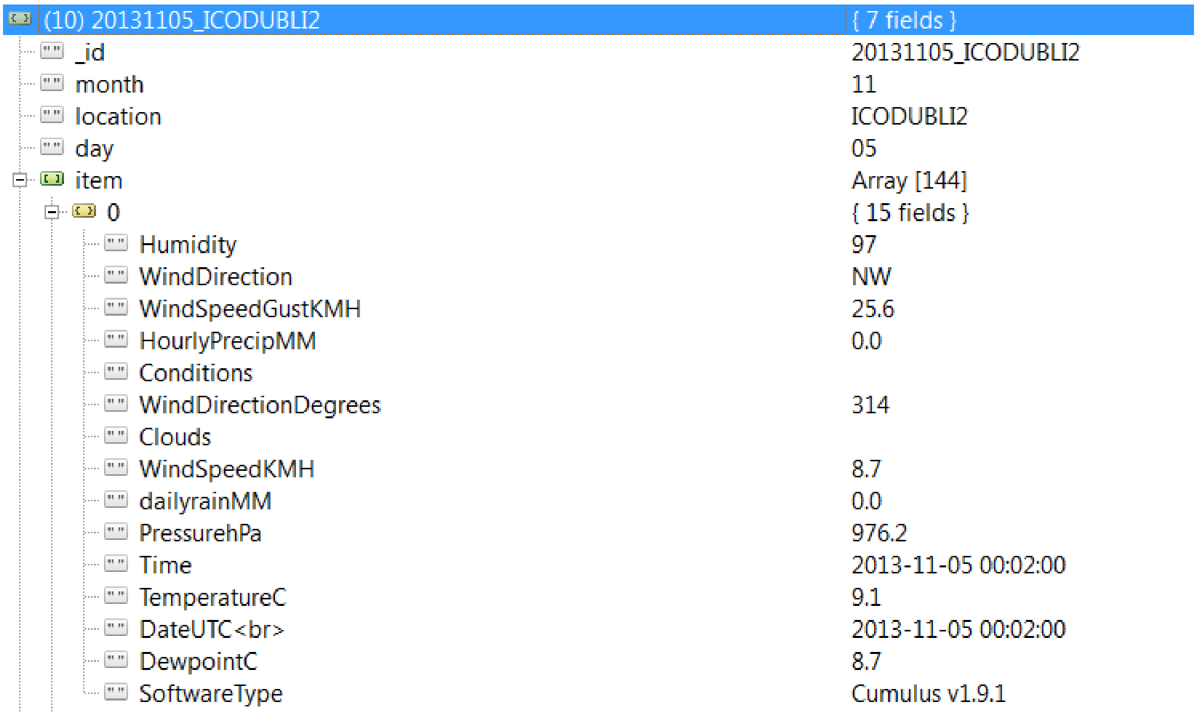


FIGURE 4.4: Weather Collection Document

TABLE 4.1: Weather Collection Attributes

Attribute	*Description*
_id	Index generated from date and location
month	Month of weather observations in item set
year	Month of weather observations in item set
location	Id of the weather station location
date	Date of weather observations in item set
item	item set containing weather observation

Table 4.2: Weather Collection Attributes

Attribute	*Description*
Humidity	Humidity
WindDirection	Wind Direction
WindSpeedGustKMH	Wind Speed Gust KMH
HourlyPrecipMM	Hourly Precip MM
Conditions	Conditions
WindDirectionDegrees	Wind Direction Degrees
Clouds	Clouds
WindSpeedKMH	WindSpeed KMH
dailyrainMM	Daily rain MM
PressurehPa	PressurehPa
Time	Time
TemperatureC	TemperatureC
DateUTC¡br¿	Date UTC
DewpointC	Dew point C
SoftwareType	Software Type

4.2.5 Twitter API Data Extraction

As mentioned in Chapter 3 Twitter provides a number of methods for collecting tweets. For the purpose of this study tweets are collected based on streaming tweets which can be filtered by geographical location and specific users. The purpose of collecting the user specific tweets is to act as training data. The trained algorithm is to be applied to the streamed data to capture Real-Time traffic related tweets. Each tweet within the Twitter is capture in real-time through Python with the module Twitter API http service, see figure **??**. User timeline searched tweet does not contain the geographical location of a tweet where the streaming tweets can be filtered based on the tweet attribute. Unlike the traffic observation section 4.2.1 the tweets already come in the JSON format ready MongoDB. Once the data read from the TwitterAPI then it can be stored into the database. An exercise is still necessary to reduce the number of attributes. This exercise is to improve the performance for running queries faster.

4.2.5.1 Geographical Referenced Tweets

Capturing Twitter data is an example of a Big Data problem. Each tweet can have up to 60 attributes of information. In this paper the geographical area being capture is Dublin with the co-ordinates of *'-7,51,-5,54'*, see figure 4.5.

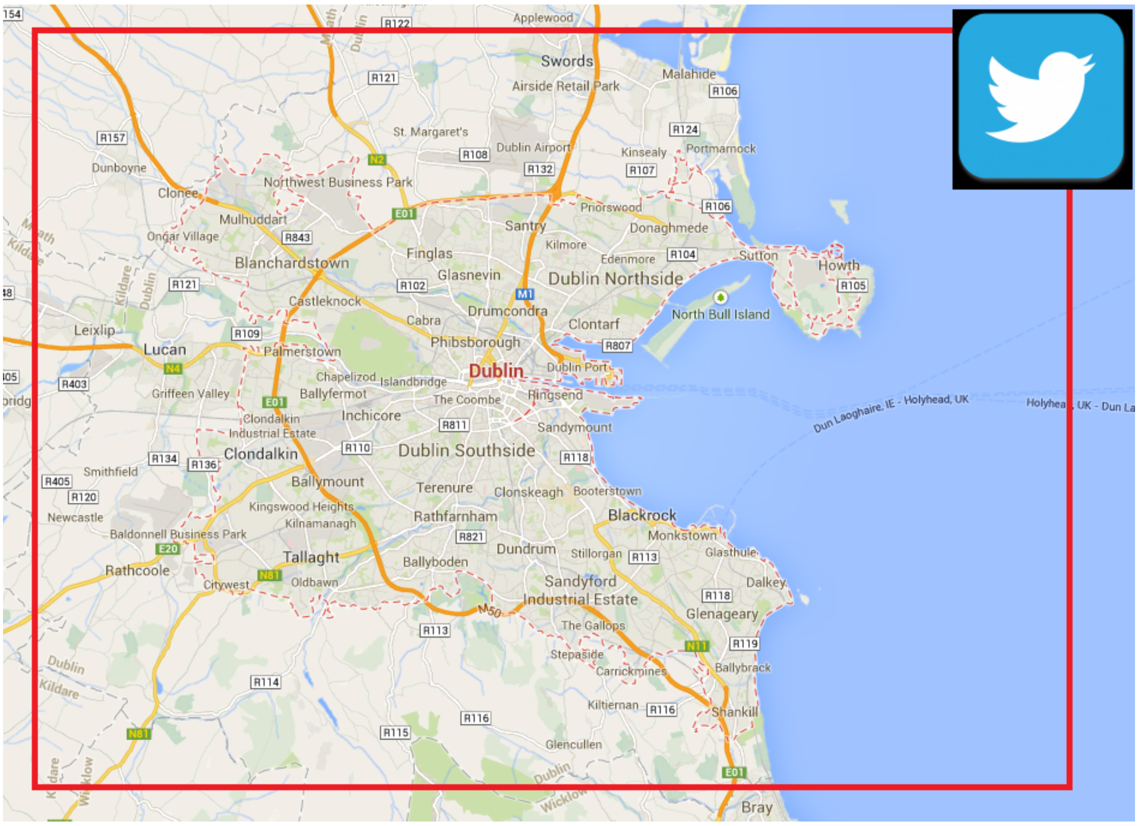


FIGURE 4.5: Twitter API filter by location

4.2.5.2 User Timeline Tweets

Tweets from **#AARoadWatch** timeline contain tweets on the domain of Traffic. There is no spatial information on these tweet other that what is in the text itself. The tweets from the time line are used to generate features that can be used in a classification of real-time tweets.

4.2.5.3 Tweet Map Reduce

A single tweet contains a lot of data attributes. This research is not performing analysis on the tweet user or the relationship between user and the text. The objective is to do text analyses for user timeline for traffic features for extraction of traffic tweets from Twitter Streaming API as gathered in appendix D . Listing 4.2.5.3 is a small view of some of the attribute that make a tweet. Using Map Reduce the tweet is reduced to 4.2.5.3. In the same way of observations tweets are aggregated with into a collection item based on the date and hour *YYYYMMDD/HH*, see listing 4.2.5.3. The example 4.2.5.3 reflect the only tweet within the mapped reduced into a one collection item. Most collection will range from 1,000 - 2,000 tweets into a single collection item or tweets for that hour.


```
{
    "_id" : ObjectId("534bfab9c009e418f4c742a1"),
    "user_profile_sidebar_fill_color" : "EFEFEF",
    "user_created_at" : "Sun Apr 04 23:58:55 +0000 2010",
    ...
    "text" : "@FunStarsGoLive @haven @DonifordOwners Hahahahahaha!!!! Didn't even know #Stanboardman did a #WorldCupSong! You proper made me chuckle!",
    "place_place_type" : "admin",
    ...
    "coordinates" : "{'type': 'Point', 'coordinates': [-3.36751463, 50.61453593]}",
    "user_description" : "London born Luton raised Devon based comedian coming to a town near you!",
    ....
    "user_lang" : "en",
    "user_followers_count" : "1594",
    "user_default_profile_image" : "False"
}
```


LISTING 4.4: Twitter Tweet


```
{
    "_id" : "2014/04/18/09",
    "items_id" : "2014/04/18/09",
    "item" : [
        {
            "hour" : "09",
            "parent_id" : "2014/04/18/09",
            "item_id" : "457081081184157696",
            "coordinates" : "{'type': 'Point', 'coordinates': [-6.25743524, 53.36674358]}",
            "date" : "2014-04-18 09:59:59 ",
            "text" : "@neilmbriscoe thankfully I have a green floor which is now an exceptionally well nourished green floor #willieverlearn",
            "place_place_type" : "city",
            "timestamp" : ISODate("2014-04-18T09:59:59.000Z"),
            "geo" : "{'type': 'Point', 'coordinates': [53.36674358, -6.25743524]}"
        },
        ....
    ]
}
```


LISTING 4.5: Twitter Tweet Attribute Map

4.2.6 Collection Result

As a result of the collection exercise all records of observations attained and accessible with no data loss. Each traffic related observation can be re-generated back to its full form. The same applies to Twitter tweets. Tweets have been reduced into another collection removing unnecessary attributes for fast analytic queries. Each record can be mapped back to its original form. This may be useful if further investigation is required for a twitter user.

4.3 Data Exploration

The data exploration section will provide a detailed description of the data collected in the section 4.2 along with visualisations. Using visualisation through Google Maps and Python is used to identify quality issues that is not feasible filters through a manual process or using standard tools. The objective of this section is to generate a generic data model that regression algorithms could be applied to each of the observed locations **(OL)**. The main purpose of using a generic model is that testing a best fit model all the **OL**s is not a practical. By analysing features through principle component analysis, histograms and correlation matrices allows for a better general understanding of the data and create a data model that each individual dataset can using.

4.3.1 Exploring Traffic

In this section the traffic observation is discussed in details exploring different distribution of values, aggregation of the data, varied seasonality of data and meta data mostly using visualisation.

A complexity of analysing the observations in a big data is the volume of information. Some of the detail is easier to comprehend using visualisation with detailed maps. Listing 4.3.1 details a total of 47 routes across the urban road network. A route can have up to 25 observed links going in 2 directions.

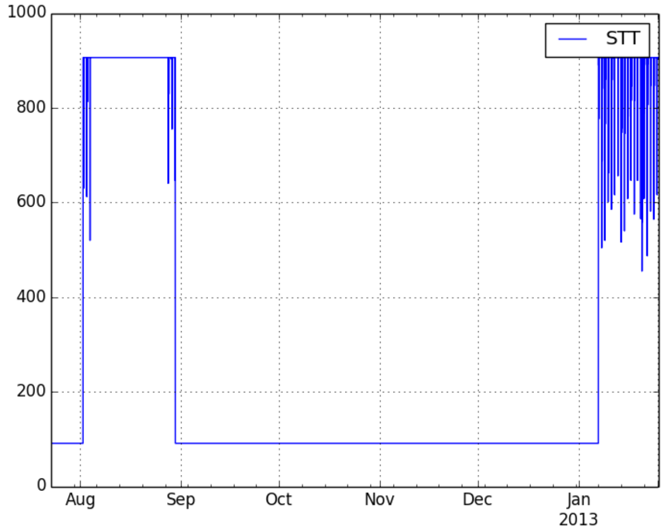
```
            direction       links      routes
count             698         698         698
min                 1           1           1
max                 2          25          47
```

Listing 4.6: Junction Meta Data

Table 4.3 is showing distribution of values from 23/07/2012 to 19/04/2014 23:50. The table provides the number of observation sample used *count*, standard deviation *std*, minimum and maximum value, and the quantile distribution. Quantile distribution at .50 (%50) is the average value and .80 (%80) is the average of the top %20 percent of values. 128 of the 698 observed locations have a count of 26834. The observed locations considered invalid for the analysis. The volume of data missing is too large to perform any imputation of missing values. Some of the observed location with count values of 26834 have been found to be duplicate locations of other location with a more complete observation count. As a result the number of complete observation data sets are 578. In figure 4.6 shows a view of travel time observations for location 10/8/2.

TABLE 4.3: Samples Distribution of Travel Time Values in Seconds

id	*count*	*std*	*min*	*max*	*.20*	*.40*	*.60*	*.80*
1/1/1	91584	62	117	1045	118	127	133	159
1/1/2	91584	8	59	411	60	61	62	63
1/10/1	91584	51	7	773	7	7	7	99
1/10/2	91584	29	7	482	15	19	29	53
1/11/1	91584	22	18	242	18	18	18	27
1/11/2	91584	24	18	242	18	18	18	31
1/12/1	91584	11	9	130	10	10	11	16
1/12/2	91584	17	9	103	9	9	18	33
1/13/1	91584	17	5	263	12	23	35	39
1/13/2	91584	17	5	263	5	5	5	29
1/14/1	*26834*	*2*	*5*	*86*	*10*	*11*	*11*	*11*



10/8/2

FIGURE 4.6: Location has 26834 observations

4.3.1.1 Travel Time Data Sets

Figure 4.7 shows that the full data sets have gaps in the data. The gaps are consistent across all links. The pattern in the data are still in tact. The tests on the data set are reduced to the date within the purple rectangle in the figure 4.7.

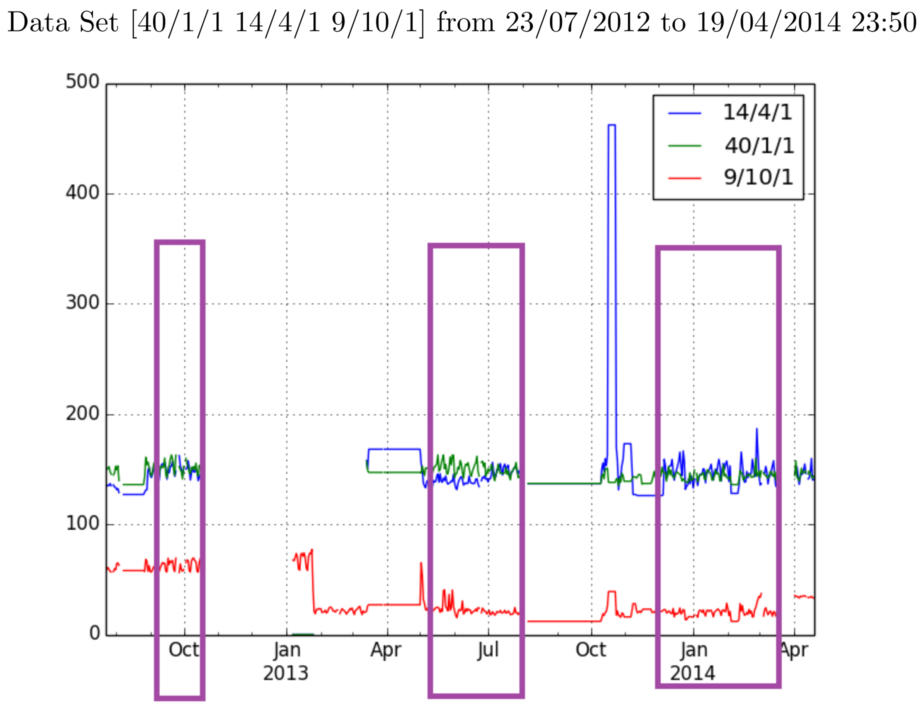


Figure 4.7: Data Sets from 23/07/2012 to 19/04/2014 23:50

The date range from September/October in 2012, May/July in 2013, Jan/May 2014. Due to this some seasonality test such as monthly, weekly is not possible. Figure 4.8 represent a daily mean data set values for the observed location *40/1/1*.

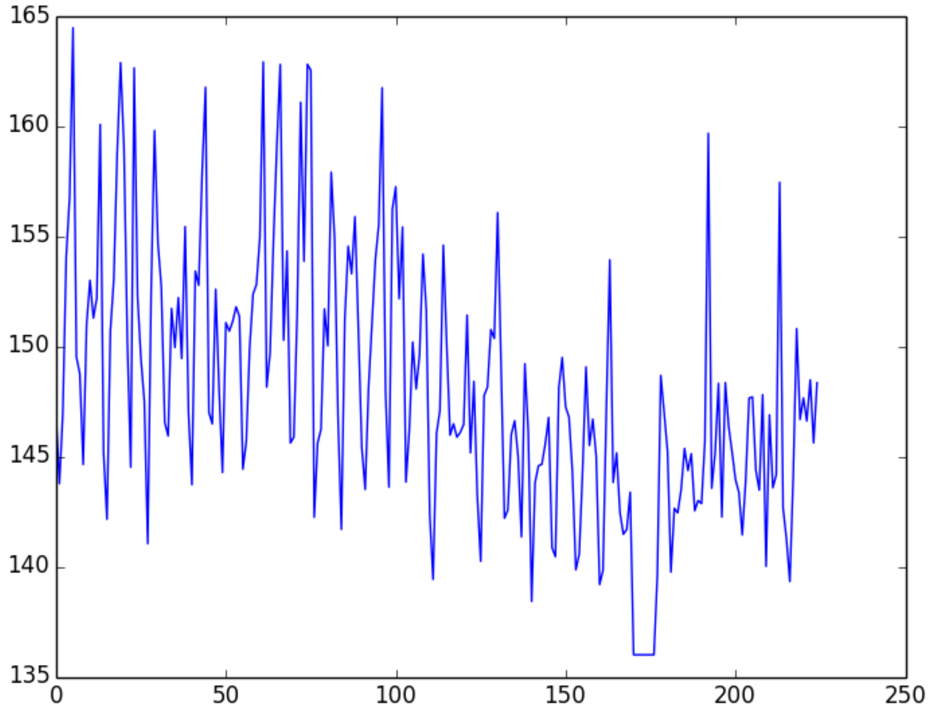
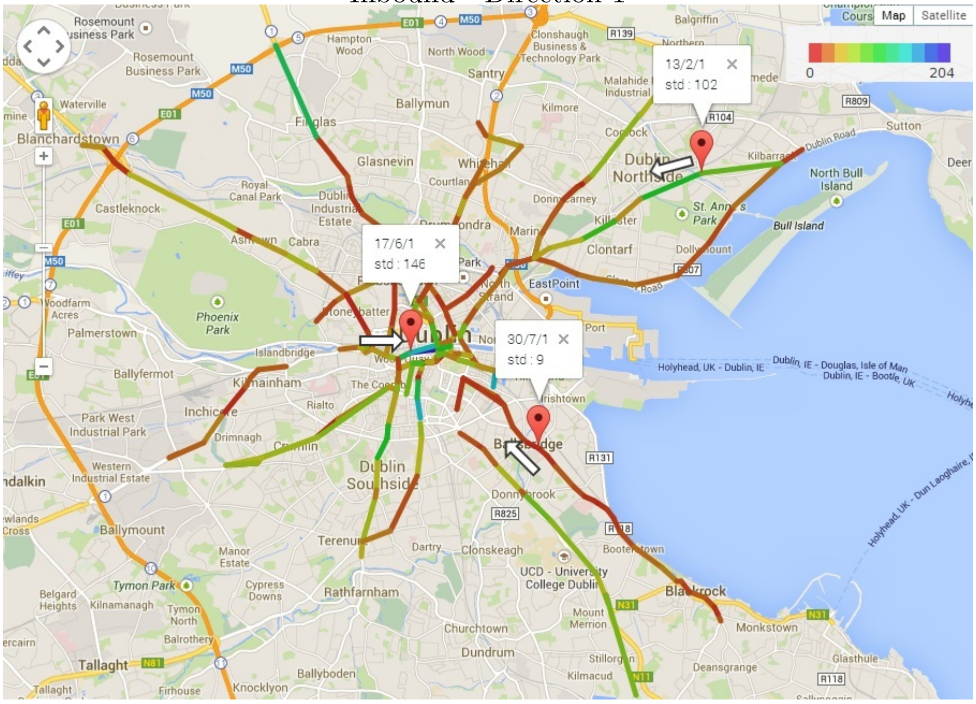


FIGURE 4.8: Daily Mean of 40/1/1

4.3.1.2 Standard Deviation of Travel Time

This is true for observed locations that are on the same links. In table 4.3 a sample of meta data shows the complexity filtering useful information. Using visualisation the information spatial elements of the data helps identify similarities between observed locations. Figure 4.9 demonstrates the standard deviation of travel times in Dublin from 23/07/2012 to 19/04/2014 23:50. Standard deviation can be considered a way of measuring the volatility [28]. A range of colour from **Red to Green to Blue** reflects the standard deviation from **Low to Medium to High**. Junctions **30/7/1, 13/2/1, 17/6/1** are examples of these categories with values of **9, 102, 146** respectively. Base on the same standard deviation scale in outbound direction 2 of 0 to 204 junctions **30/4/2, 10/7/2, 16/2/2** are examples of these categories with values of **13, 102, 192**.

Inbound - Direction 1



Outbound - Direction 2

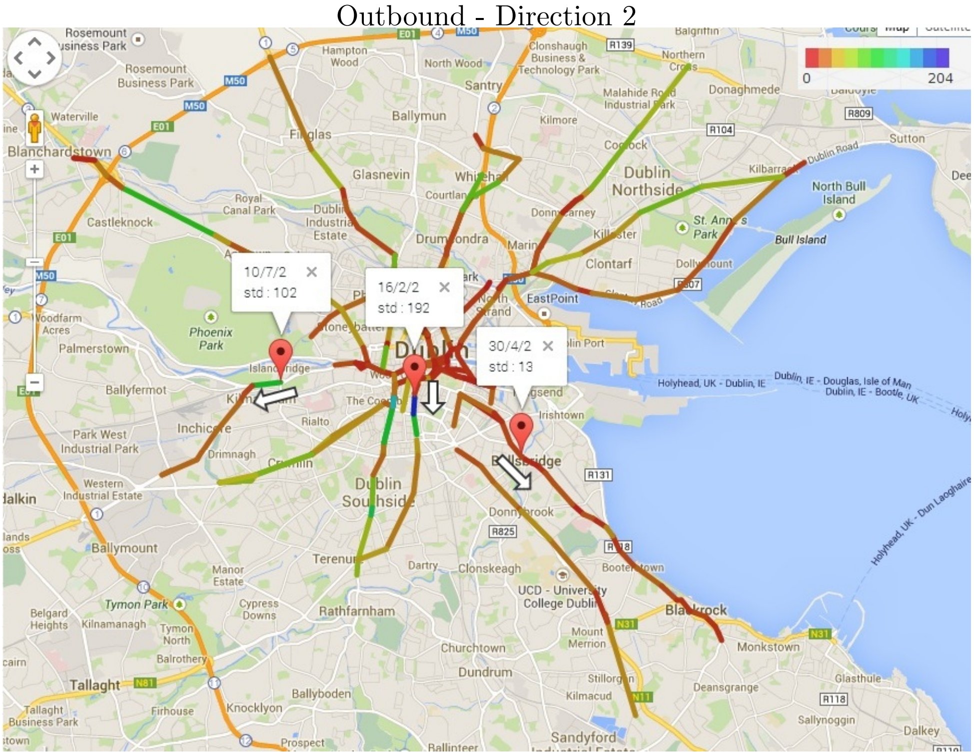
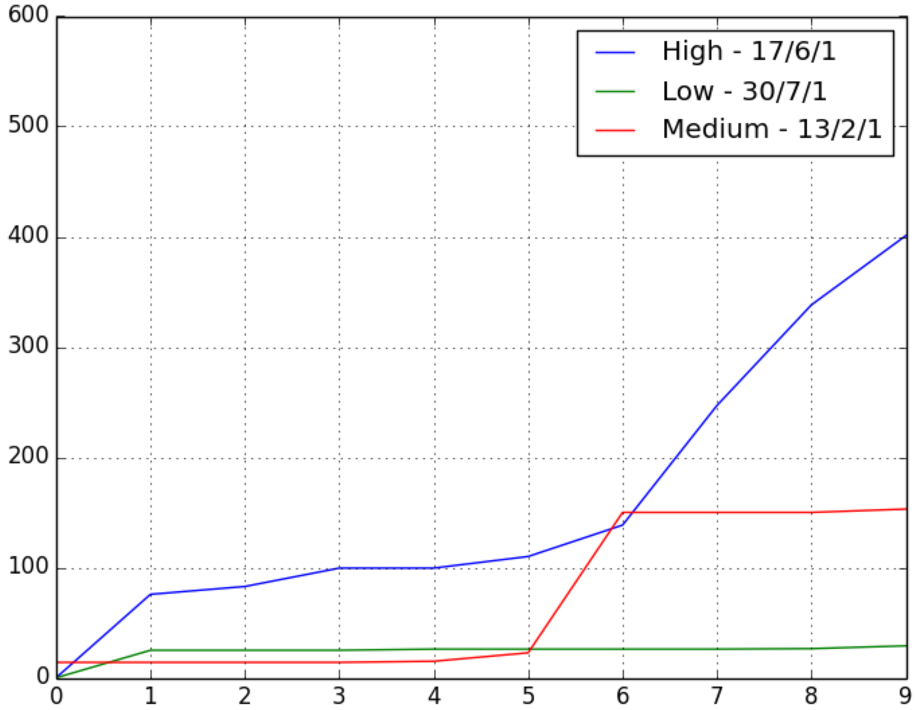


FIGURE 4.9: Inbound and Outbound Traffic Observations

Most standard deviations (STD) are in the low to medium value ranges. The volatility of observed junctions does not change much between inbound direction 1 traffic and outbound direction 2 traffic. Outbound at Aungier Street 16/2/2 demonstrates that it clearly the most volatile. This means that the route link is the most unpredictable observed location in Dublin City. Dublin city centre Inbound shows a more medium to high STD opposed to outbound demonstrate low STD, see figures 4.10

Inbound - Direction 1



Outbound - Direction 2

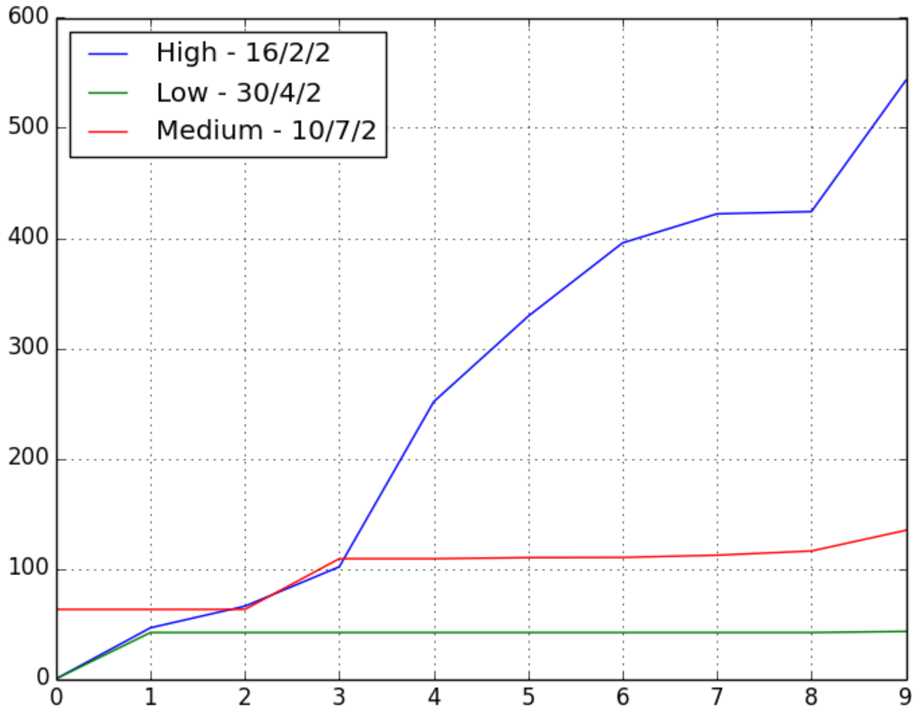


FIGURE 4.10: Inbound and Outbound, Low - Medium - High

4.3.1.3 Seasonality of Travel Time

Seasonality are commonly observed in quarterly and monthly time series, with multiple overlying seasonality occurring in weekly, daily and hourly data [29].

Fluctuations over segmented periods of time are commonly observed in research in order to relate patterns in data to time [29]. Often these periods are in the form for yearly, seasonally (Summer/Winter/Spring/Autumn), monthly, weekly, daily, and hourly. Stephen Dunne and Bidisha Ghosh [11] and Sri Krisna Endarnoto et al [15] have both researched the seasonality differences of peak time traffic and also the differences in week-day and week-end traffic. Sri Krisna Endarnoto et al assumed a peak time of 9am. In this section this assumption is explored.

The method of understanding peak times comes from using quantile percentage. Quantile 0.80 is used as a measure to identify high value points in the data. Quantile is similar to the mean value. The mean is equal to quantile of point 0.50. In figure 4.10 the x-axis provides the quantile distribution. The axis at point 7 is quantile 0.80. To calculate the peak hours of a road section the hour that return the highest quantile at 0.80 is used. As a result the data shows that not all roads. In table 4.4 shows that it is incorrect to assume the inbound traffic peak times are from early morning between 7am and 10am. The result shows the *14/4/1* road is busy between the 14th hour and the 16th, ie 2pm to 6pm. Both Road *14/4/1* and *40/1/1* are both a similar distance from the city centre but have different peak hours, see 4.5. In general during peak times appear to be outside working hours before 10am and after 4pm. On weekends peak hours are highest around 12am and before midnight. Outbound in much of the peaks times is in the late evening after 5pm working hours. It could be considered unusual that peak hours are between 9pm and 11pm. This could be down to many late evening events that people drive too or grocery shopping.

Table 4.4: Sample Weekday Inbound Peak hours

id	*hour (value) from highest*
14/4/1	15 (170.57) 14 (168.93) 16 (168.0)
1/7/1	21 (126.4) 20 (126.4) 19 (126.4)
17/6/1	17 (401.0) 9 (401.0) 16 (401.0)
40/1/1	8 (186.16) 7 (173.08) 9 (161.62)
9/10/1	8 (65.73) 23 (56.0) 22 (56.0)
6/6/1	8 (71.0) 7 (70.81) 9 (70.99)
43/2/1	9 (61.2) 8 (57.78) 7 (56.19)

Table 4.5: Peak Times

Sample Weekend Inbound Peak hours

id	*hour (value) from highest*
14/1/1	12 (139.0) 13 (139.0) 14 (138.07)
1/7/1	20 (126.4) 19 (126.4) 18 (126.4)
17/6/1	13 (401.0) 14 (401.0) 15 (401.0)
40/1/1	13 (149.8) 12 (148.5) 14 (147.78)
9/10/1	23 (56.0) 22 (56.0) 21 (56.0)
6/6/1	9 (63.4) 10 (62.63) 11 (57.0)
43/2/1	23 (47.0) 22 (47.0) 21 (47.0)

Sample Weekday Outbound Peak hours

id	*hour (value) from highest*
14/4/2	17 (239.77) 15 (236.66) 16 (232.6)
1/7/2	23 (62.0) 22 (62.0) 21 (62.0)
17/6/2	23 (73.0) 22 (73.0) 21 (73.0)
40/1/2	17 (362.7) 16 (345.62) 18 (337.0)
9/10/2	23 (57.0) 22 (57.0) 21 (57.0)
6/6/2	23 (7.0) 22 (7.0) 21 (7.0)
43/3/2	17 (30.86) 18 (29.28) 8 (28.74)

Sample Weekend Outbound Peak hours

id	*hour (value) from highest*
14/4/2	14 (232.43) 15 (226.5) 13 (226.47)
1/7/2	23 (62.0) 22 (62.0) 21 (62.0)
17/6/2	23 (73.0) 22 (73.0) 21 (73.0)
40/1/2	14 (309.3) 13 (301.68) 15 (298.49)
9/10/2	23 (57.0) 22 (57.0) 21 (57.0)
6/6/2	23 (7.0) 22 (7.0) 21 (7.0)
43/2/2	18 (53.6) 17 (52.1) 14 (52.09)

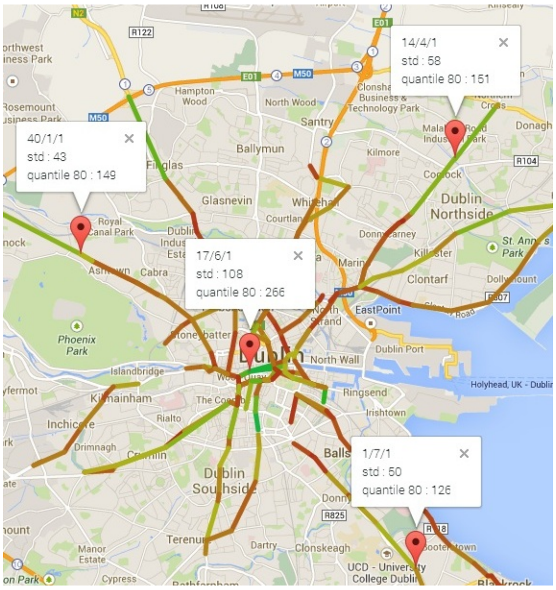


Figure 4.11: Peak Hours Inbound

4.3.1.4 Exploring Traffic Result

Much of the remote sensor data available is corrupt and configured incorrectly leading duplication of roads and large numbers of observations missing. Patterns in the data are still available as seen in 4.8. Some roads have little change in its traffic travel times. Other roads show large deviations in times and identifying volatility on roads was simplified using Google Maps. Seasonality was identified by separating weekday and weekend data. By listing the peak times it is proven that in table 4.5 that each observed location has different characteristics. In section 5 Model Selection will identify some of the characteristics that influence an observed location.

4.3.2 Exploring Weather

For this research three different weather stations has been chosen to build a model for traffic analyses. Each station is records data at random interval mostly between 5 and 10 minutes. There is no guarantee on the time of the weather observation is recorded. The three data sets are comprised of up to 100,000 observations. For exploration the data sets are compared to each other with different forms of aggregation. Initially some detail of the full data sets are provided. Other data sets will use aggregated seasonal data sets such as daily, hour and different time ranges. At this point we know from other research that weather does effect road conditions [11]. As this study is exploring the dynamics of roads over a large geographical location. It is required to know that not all roads are under the same whether condition at one time or even receive the same level of condition.

The weather variables associated with rainfall and temperature are discussed in greater details in terms of seasonality and reasons behind dimension reduction.

4.3.2.1 Weather Data Set

In listing 4.3.2.1 the description of the data shows more often the conditions dry and the normal temperature ranges from 7 to 15 averaging 11 for a normal giving day. The maximum hourly rainfall is at 2539.7mm. This is either an outlier or a sign of poor data quality.

The correlation between HourlyPercipMM and DailyPercipMM is very high. One of these attributes is a candidate for attribute reduction. DailyPercipMM is derived from an accumulation HourlyPercipMM.

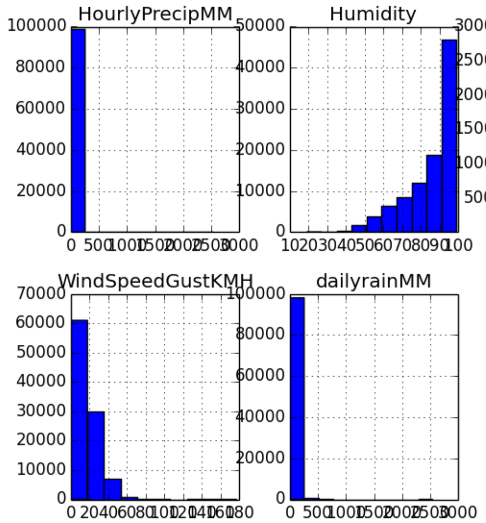


FIGURE 4.12: Weather Histograms, Lucan, Co. Dublin

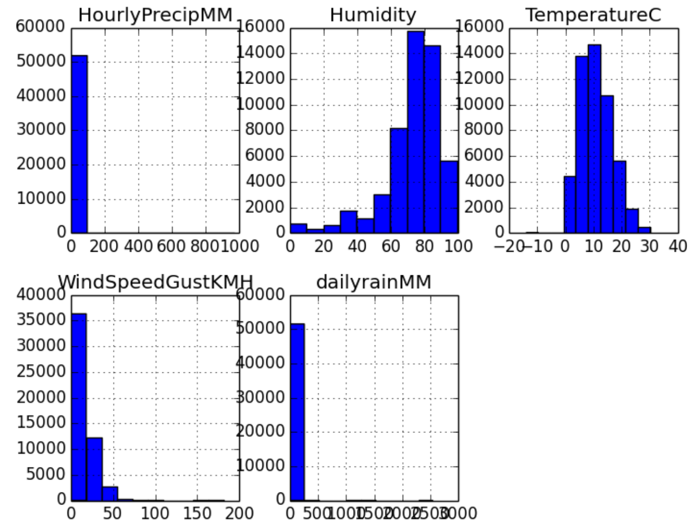


FIGURE 4.13: Weather Histograms for Blackrock, Dublin 8

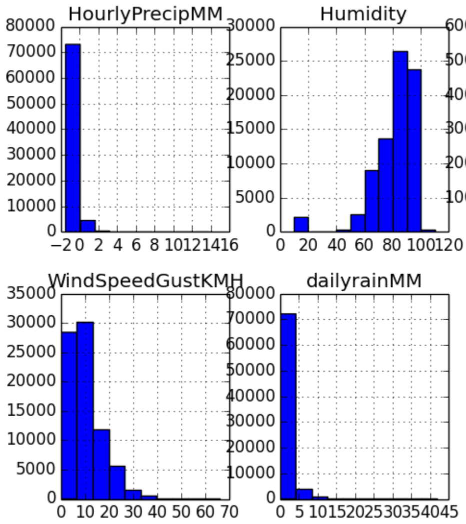


FIGURE 4.14: Weather Histograms for Artane, Dublin 5

```
                HourlyPrecipMM          Humidity         TemperatureC   \
count                98,666.00         98,666.00            98,666.00
mean                      0.21             85.41                11.34
std                      16.34             14.10                 5.52
min                       0.00             19.00                −3.30
25%                       0.00             77.00                 7.20
50%                       0.00             90.00                11.00
75%                       0.00             98.00                15.00
max                   2,539.70             99.00                34.50

              WindSpeedGustKMH        dailyrainMM
count                98,666.00         98,666.00
mean                     14.63              3.61
std                      13.41             34.75
min                       0.00              0.00
25%                       0.00              0.00
50%                      13.40              0.00
75%                      23.30              0.50
max                     177.00          2,539.70
```

LISTING 4.7: Weather Data Set for ICODUBLI2, Lucan

```
                          HourlyPrecipMM          Humidity   \
HourlyPrecipMM                    266.87              0.29
Humidity                            0.29            198.91
TemperatureC                        0.35            −38.60
WindSpeedGustKMH                   −1.18            −32.36
dailyrainMM                       266.74            −24.19
```

```

                   TemperatureC   WindSpeedGustKMH   \
HourlyPrecipMM             0.35              -1.18
Humidity                 -38.60             -32.36
TemperatureC              30.49              13.57
WindSpeedGustKMH          13.57             179.90
dailyrainMM                0.41             -16.85

                    dailyrainMM
HourlyPrecipMM           266.74
Humidity                 -24.19
TemperatureC               0.41
WindSpeedGustKMH         -16.85
dailyrainMM            1,207.90
```

LISTING 4.8: Weather Correlation Matrix for ICODUBLI2, Lucan

4.3.2.2 Daily Aggregation Precipitation

In figure 4.15, using the daily mean of rain it clearly demonstrates that the levels of rain vary according to geographic location. This show a significant difference between weather conditions with each weather station.

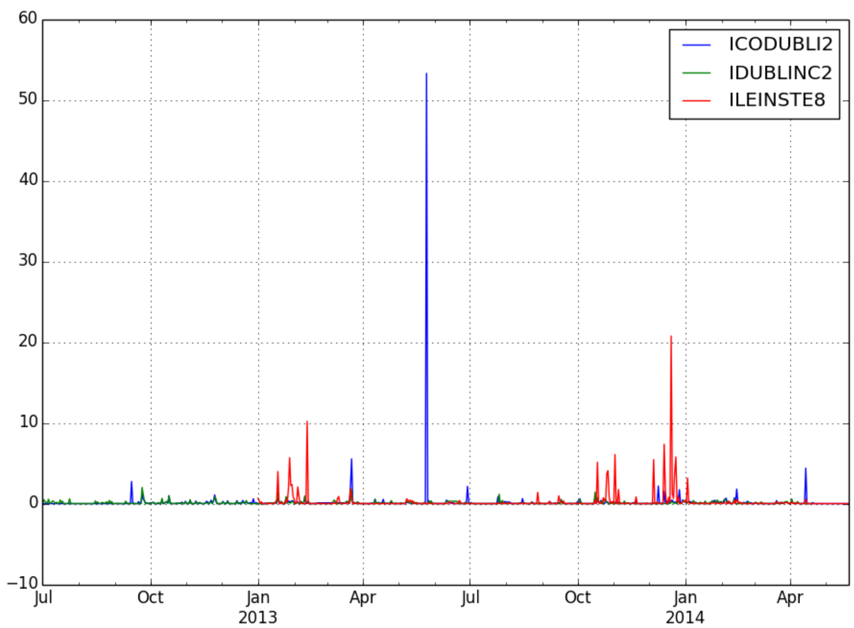


FIGURE 4.15: Weather Stations Rain Daily Mean

4.3.2.3 Hourly Aggregation Precipitation

By focusing in on a smaller sample time frame it is evident that the rainfall closing aligned. Still the graph proves the data has some level of correlation but does not align exactly 4.6. In figure 4.16 on the 27th of January it even demonstrates that has rained in North Dublin (*ICODUBLI2*) but yet the rain did not follow into West or South Dublin.

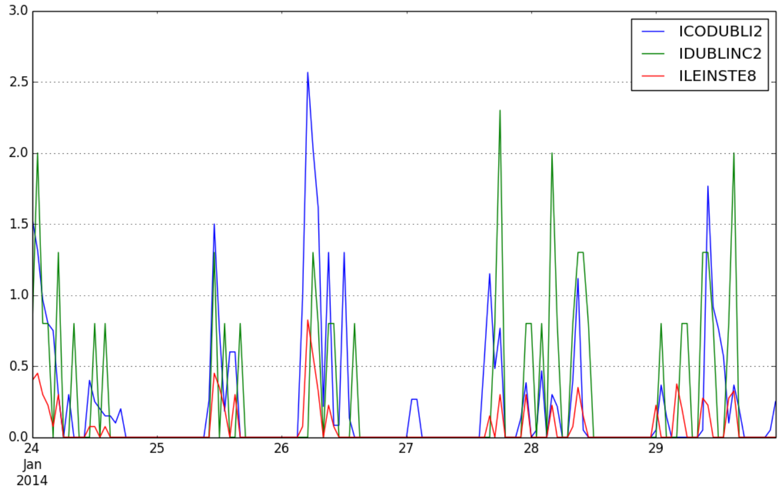


Figure 4.16: Weather Stations Rain Hourly Mean January 24th-29th

	ICODUBLI2	*IDUBLINC2*	*ILEINSTE8*
ICODUBLI2	1.000000	0.435804	0.754284
IDUBLINC2	0.435804	1.000000	0.587087
ILEINSTE8	0.754284	0.587087	1.000000

Table 4.6: Weather Stations Correlation Linear Regression on HourlyPrecipMM in Jan 24th - 29th

4.3.2.4 Daily Aggregation Temperature

In figure 4.18, using the daily mean of temp it clearly demonstrates that the levels of rain vary according to geographic location. Unlike the precipitation the temperature shows an element of seasonality as discussed by [11]. In IDUBLINC2, Artane Dublin 5 there is erroneous values. In July 2013 values differentiate by nearly 400 degrees Celsius. In table 4.7 the correlation is higher than the precipitation.

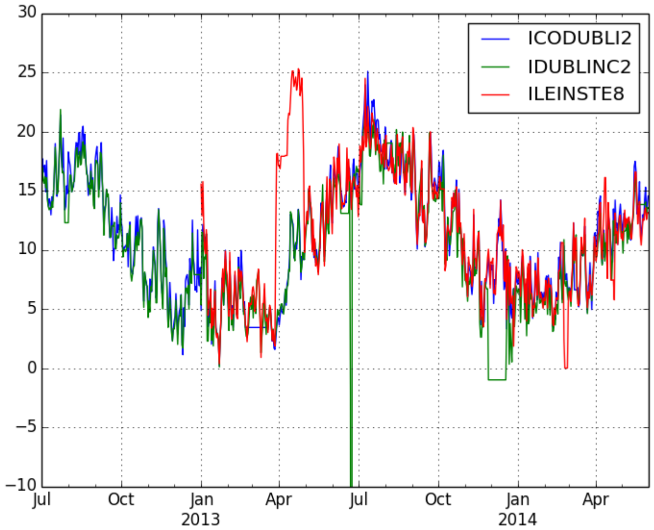


Figure 4.17: Weather Stations Temperature Daily Mean

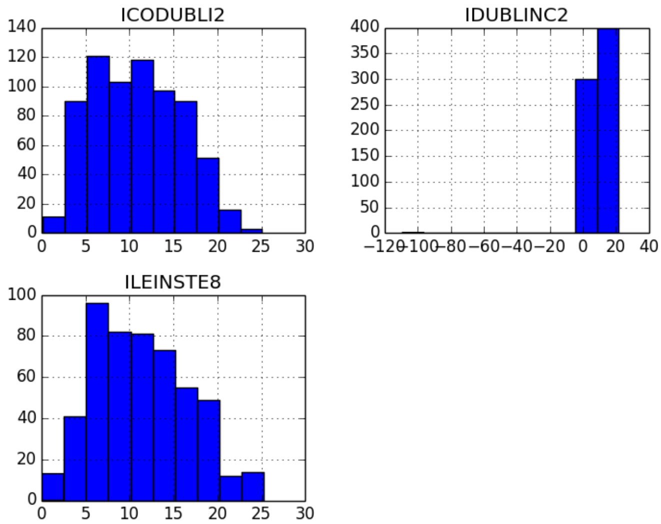


Figure 4.18: Weather Stations Temperature Daily Correlation

	ICODUBLI2	*IDUBLINC2*	*ILEINSTE8*
ICODUBLI2	1.000000	0.672347	0.774402
IDUBLINC2	0.672347	1.000000	0.502557
ILEINSTE8	0.774402	0.502557	1.000000

Table 4.7: Weather Stations Correlation Linear Regression on TemperatureC

4.3.2.5 Weather Exploration Result

As a result of the weather exploration if the graphs show that there a clear indication the reasoning behind using different weather stations is positive. The weather stations demonstrate different patterns in the data. Therefore it is reasonable to believe that the effects on a weather station is spatially related to a road. It has been proven by Kevin Keay and Ian Simmonds in 2004 [11] the there is a relationship between weather condition and traffic patterns. Using correlation techniques the dimension of data can be reduced. Precipitation variable HourlyPrecipMM is the best measure for determining wet and dry conditions and TemperatureC is the best variable for cool and hot conditions. It is not necessary to have dailyrainMM and Himidity to fit any model.

4.3.3 Exploring Twitter

There are two twitter data sets, user timeline data set which will be used to aid the extract the classification of traffic related tweets from real-time tweets. The objective of this is to classify tweets with traffic features from *AA Road Watch* tweets.

Using the AA Road Watch data set the tweets are analysed and split into word tokens. The token analyser will filter words out that provide better results. These options are explored in the model selection 5. The tokens are used as features for scoring geographically referenced tweets as traffic related.

The process of correlating the traffic tweets to traffic observations is a visual mechanism. The tweet data set available does not span across the same time frame as the traffic observations. The geographically referenced tweets data ranges from 2014/04/15 to 2014/04/26. The AA Road Watch user timeline data set is a sample taken from 2013/12/15 to 2014/04/26.

4.3.3.1 User Timeline Tweets

The AA Road produce tweets that are communicated to the public as a service to the state with information contain traffic news. The AA Road Watch corpus contains 5267 tweets. Using tokeniser parameters 4.3.3.1 a TF-IDF vector 4.3.3.1 represents the sample features in traffic related tweets. The result accumulates 4866 total features. Using 4.19 the features are visualised. This method reveals immediate influential key features and provides an overview of the vocabulary used to provide traffic information.

```

TfidfVectorizer(
        analyzer='word',
        token_pattern=r'[a-z]{4,}',
        use_idf=True,
        strip_accents='unicode',
        sublinear_tf=False)
```

Listing 4.9: Python Word Tokenizer

The tokenizer parameters allows for each tweet to be broken into token features. Not all parameter are utilized in the exploration stage. Further investigation into N-Grams and Frequency Ranges are considered in the Chapter Model Selection 5.

Parameter	Options	Description
analyzer	'word', 'char', 'char_wb'	types of tokens, words or characters
token_pattern	r'[a-z]4,' only contains letters a-z length 4	Regular expression denoting what constitutes a "token"
strip_accents	'ascii', 'unicode', None	Remove accents during the preprocessing step. 'ascii' is a fast method that only works on characters that have an direct ASCII mapping. 'unicode' is a slightly slower method that works on any characters. None (default) does nothing.

Table 4.8: Tf-idf Tokenizer

The word cloud 4.19 provides an insight into the traffic vocabulary. The TF-IDF scores the features in the corpus and the word cloud displays the most prominent features. The features can be categorised into three fundamental areas. Word associated with Traffic, Location and Punctuated Twitter Words.

Traffic features words from the word cloud are **traffic, debris, volume, broken, overturned**. The Traffic feature in the thesis will be used to score or categorise the geographically referenced tweets.

Location features are not an indicator of a tweets being traffic related but do indicate the location of where a traffic information is referring too, i.e. **waterford, monastervin, bray, knock**.

Punctuated Twitter Words are features associated with HashTag, Users and URLs, i.e. **#AARW, Peterbowles and http://t.co/YEsG6RDQW3**. The vectorisor mechanism transforms the URL from http://t.co/YEsG6RDQW3 to YEsG6RDQW3.

Figure 4.19: AA Road Watch Cloud

```
affect              0.259152401722
cleared             0.42853734037
closures            0.42853734037
collision             0.42853734037
limerick            0.140959969996
monastervin           0.245823461838
msfrugalone           0.21506648088
newtownmountkennedy  0.289172991764
```

```
northbound         0.312560130227
ofzigkql        0.26570466618
outbound        0.232460185377
overturned         0.312560130227
quay         0.312560130227
quays           0.152360240527
qvbpcgxuj       0.312560130227
removed         0.312560130227
report          0.4472135955
southbound         0.4472135955
tara         0.301011047979
there           0.301011047979
this         0.404731899373
traffic         0.344059090764
volume          0.404731899373
```

LISTING 4.10: Word TF-IDF Vector Sample

4.3.3.2 Conclusion

As a result the word cloud makes it easier to understand the data compared to the word vector listing 4.3.3.1. In chapter 5 the scoring mechanism will compare the geographical referenced tweets using the different categories of features such as traffic, location and punctuated twitter words. Some tweets updates do not contain traffic information. Often updates contain only thank you messages such as "
@HamillsRecovery Thanks for that" and "
@Clareokeeffe19 Thanks for the heads up Clare. http://t.co/qvBPcggRh8". The data set contains 5270 tweets. The TF-IDF Vectorizer is not effected by such tweets.

4.3.3.3 Geographical Referenced Tweets

The purpose of this section is to explore geographically referenced tweets. Using Google Maps the number of tweets is visualised, see 4.20. The figure represents all tweets for the date and hour of 2014/04/18 9:00pm. Google Maps provide a clustering mechanism the groups the spatially related tweets together.

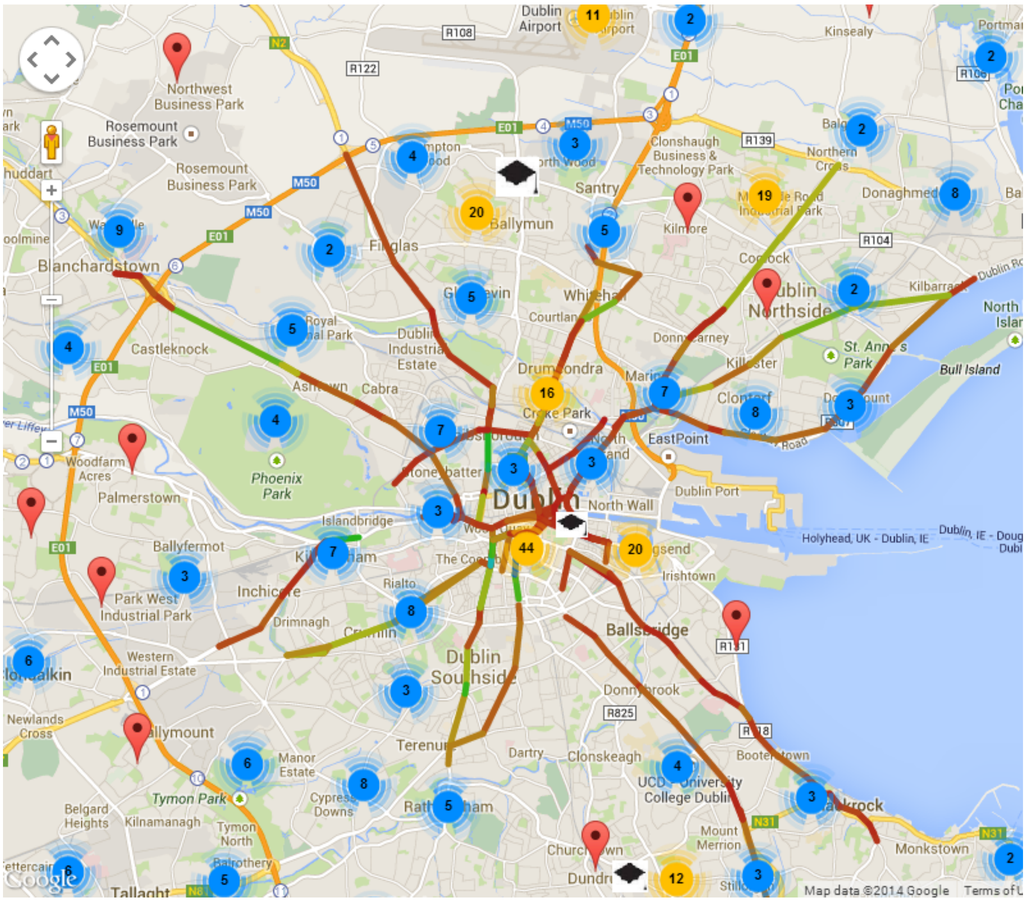


FIGURE 4.20: Geographical Referenced Tweets on 2014/04/18 9:00pm

Without the traffic scoring mechanism in place is no apparent relationship between the tweets from potentially high volume traffic in **??**. The Hat icons represent universities in Dublin, Trinity College in the city centre, Dublin City University in the north of Dublin and University College Dublin in south of Dublin. The high volume yellow clusters correlate with the universities of Dublin, one yellow cluster in Dublin Airport and one on the Malahide Road. In chapter 5 the relationship between extremely busy junctions along with traffic tweets will be analysed.

In figure 4.21 junction 16/2/1 of Wexford St and Kevin St shows that this particular junction is a volatile junction as discussed in section 4.3.1. The peak time for junction 16/2/1 during week days is the hours of 18:00 and 17:00. Using the method of extracting peak times in section 4.3.1 and searching tweets that contain any of the words from the word cloud 4.19 **traffic, debris, volume, broken, overturned** proves that traffic tweets exist within the data set within the maximum distance of 1km from the junction.

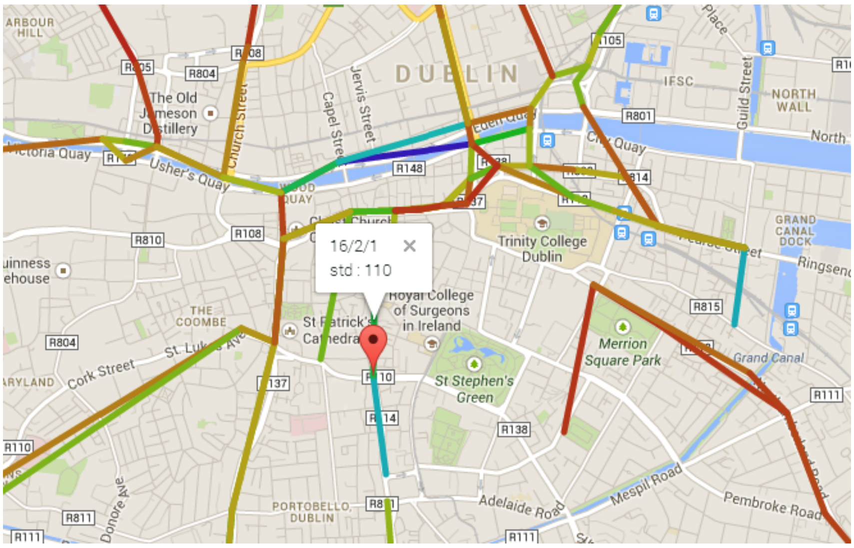


FIGURE 4.21: Junction Wexford St and Kevin St

Based on the tweets containing **traffic, debris, volume, broken, overturned** the listings show mixed results matching traffic related tweets using those terms. **Debris and overturned** contain no results, **traffic** contained 4 results which are related to traffic and **volume, broken** both returned results not related to traffic.

```
It took almost 20 minutes to get from Parnell to trinity the traffic is
    ridiculous today wtf
Today I may cry as a commuter #protest #sittingonabusforoverahr #traffic #
    Dublin
Last 67 and the bus driver wouldnt open the door ten yards from the stop
    for someone who'd missed it. We were stuck in traffic. Awful stuff.
Jesus. Junkie bleeding from his face and playing Frogger with Capel St and
    Quays traffic then lunging and grabbing at tourists. Shocking.
```

LISTING 4.11: Tweets with 'traffic' 2014/04/25 18:00-18:59pm

```
Someone is playing rock/heavy metal on max volume. Guess I'll have to wait
    a lil bit longer then before I can study #sorryimnotsorry
```

LISTING 4.12: Tweets with 'volume' 2014/04/25 18:00-18:59pm

```
@towerdublin pls tell us what limit is on #unbrokenuntied ? @delorentos
    fans wanna help each other out. Wanna tell ppl if we can't buy \&gt;1
```

LISTING 4.13: Tweets with 'broken' 2014/04/25 18:00-18:59pm

4.3.3.4 Twitter Conclusion

As a result the geographical referenced tweets contain traffic related tweets. The most words with that scored highest in the TF-IDF result does not ensure to produce tweets closer related to traffic using the mechanism of a single word search. Further investigation on this is in the Model Section chapter 5. The result of the **traffic** word search resulted in traffic tweets. In both cases locations mentioned in the tweets can reflect the location of where the tweet was broadcast 4.22.

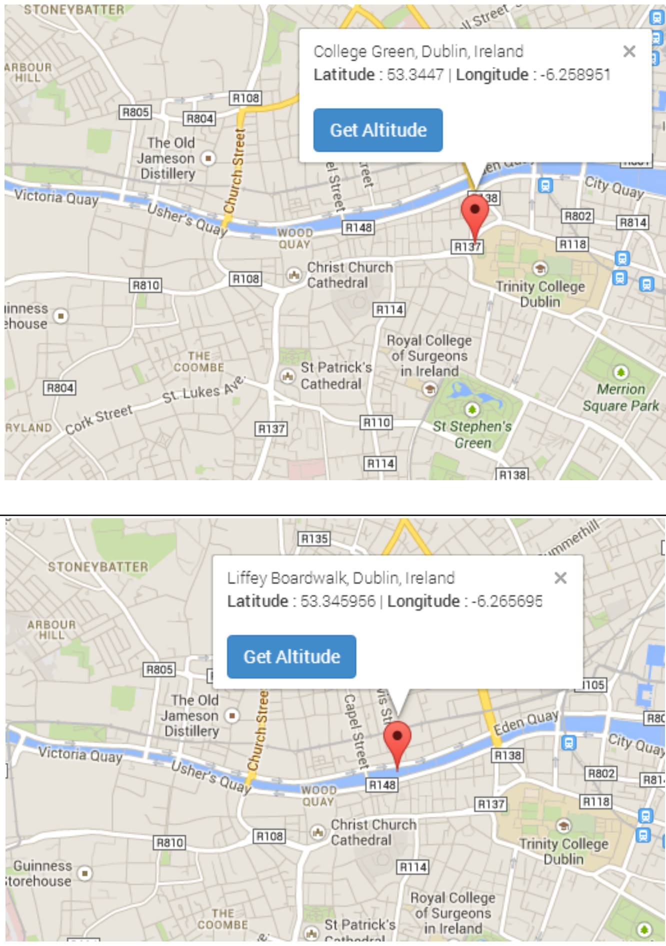


FIGURE 4.22: Tweet From Location 2014/04/25 18:00-18:59pm

Chapter 5

Model Selection

5.1 Introduction

In this chapter, we will present data models for prediction based on features travel time observations, spatial neighbours and weather data. The features used are determined by analysing and reducing the necessary ARIMA attributes for building a generic model that will fit all observed locations **OL**s. To create a generic model features a selected by analysing the correlation using data mining techniques and visualisation.

5.1.1 Standard Travel Time (STT) Model Selection

In this section the focus is mainly on prediction of peak times during weekdays. Three types of traffic volatility identified in the data exploration section analysing distribution of observation values categorised as Low to Medium to High 4.3.1.2. These junctions cover the categories Low to Medium to High. the daily seasonality for observations on peak times during business days Monday to Friday 4.3.1.3. Other seasonality and trends are ignored with the volume of quality data being limited 4.6. Therefore there is no attempt to find trends for monthly or quarterly means.

TABLE 5.1: Correlation coefficients Matrix Colour Coded Summary

	Lag -1	Lag -2	Lag -3	Lag -4	Lag -5
>= **0.5**	50%	39%	34%	29%	30%
>**0** and <**0.5**	49%	59%	63%	68%	69%
<**0**	0.2%	1.0%	1.5%	2.0%	1.0%

Lag 1	Lag 2	Lag 3	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7	Lag 8	Lag 9	Lag 10	Lag 11	Lag 12	Lag 13	Lag 14
0.97	0.77	0.72	0.71	0.71	0.70	0.70	0.69	0.69	0.69	0.69	0.68	0.67	0.67	0.67
0.47	0.27	0.27	0.24	0.28	0.02	0.00	-0.08	-0.02	0.03	-0.07	-0.01	-0.04	0.08	0.22
0.97	0.94	0.91	0.89	0.86	0.84	0.82	0.80	0.78	0.76	0.75	0.73	0.72	0.70	0.69
0.34	0.20	0.17	0.20	0.16	0.14	0.04	0.06	0.15	0.15	0.05	-0.03	0.01	0.03	0.18
0.95	0.44	0.45	0.43	0.46	0.48	0.51	0.49	0.49	0.54	0.59	0.60	0.54	0.54	0.55
0.97	0.94	0.91	0.88	0.86	0.83	0.81	0.79	0.77	0.76	0.74	0.72	0.71	0.69	0.68
0.68	0.67	0.66	0.63	0.55	0.62	0.61	0.54	0.55	0.62	0.49	0.53	0.53	0.50	0.44
0.97	0.74	0.70	0.69	0.68	0.68	0.69	0.68	0.67	0.67	0.67	0.67	0.66	0.65	0.65
0.36	0.25	0.25	0.19	0.25	0.03	-0.02	-0.06	-0.04	0.13	-0.05	-0.13	-0.09	-0.01	0.21
0.40	0.31	0.29	0.26	0.37	0.16	-0.02	0.00	0.10	0.24	0.07	0.01	-0.01	0.13	0.33
0.97	0.94	0.91	0.89	0.86	0.84	0.82	0.81	0.79	0.77	0.76	0.74	0.73	0.71	0.70
0.97	0.94	0.91	0.89	0.86	0.84	0.82	0.80	0.78	0.76	0.74	0.73	0.71	0.70	0.69
0.75	0.21	0.20	0.16	0.19	0.40	0.56	0.47	0.31	0.31	0.45	0.60	0.48	0.34	0.25
0.41	0.23	0.22	0.32	0.44	0.16	0.05	-0.03	0.10	0.30	0.08	0.00	0.02	0.13	0.39
0.74	0.63	0.66	0.57	0.53	0.43	0.47	0.38	0.31	0.37	0.33	0.28	0.30	0.32	0.38
0.97	0.89	0.85	0.83	0.82	0.80	0.79	0.78	0.76	0.76	0.75	0.75	0.73	0.73	0.71
0.97	0.94	0.91	0.88	0.86	0.83	0.81	0.79	0.77	0.76	0.74	0.72	0.71	0.69	0.68
0.96	0.62	0.59	0.61	0.61	0.61	0.62	0.61	0.60	0.60	0.60	0.60	0.59	0.59	0.58
0.76	0.78	0.76	0.70	0.75	0.70	0.73	0.71	0.68	0.74	0.65	0.70	0.72	0.69	0.72
0.66	0.42	0.37	0.36	0.35	0.36	0.40	0.38	0.37	0.37	0.39	0.41	0.40	0.37	0.36
0.45	0.27	0.23	0.31	0.44	0.31	0.07	0.09	0.19	0.20	0.08	-0.07	-0.05	0.08	0.16
0.81	0.81	0.80	0.78	0.74	0.73	0.74	0.68	0.68	0.66	0.66	0.61	0.62	0.59	0.57
0.97	0.94	0.91	0.88	0.86	0.84	0.81	0.79	0.78	0.76	0.74	0.72	0.71	0.70	0.68
0.04	0.12	0.00	-0.01	0.30	0.16	0.03	0.00	-0.20	0.14	0.02	0.01	0.00	-0.04	0.01
0.34	0.17	0.11	0.22	0.07	0.13	0.08	-0.09	0.00	0.04	0.04	-0.14	-0.07	0.08	0.14
0.97	0.94	0.91	0.88	0.86	0.83	0.81	0.79	0.77	0.76	0.74	0.72	0.71	0.69	0.68
0.37	0.25	0.26	0.30	0.28	0.21	0.12	0.13	0.21	0.20	0.16	0.17	0.08	0.15	0.13
0.18	0.32	0.33	0.14	0.33	0.15	0.20	0.19	0.00	0.16	0.05	0.03	0.17	0.00	0.12
0.43	0.43	0.41	0.37	0.43	0.35	0.35	0.34	0.28	0.39	0.30	0.33	0.34	0.28	0.42
0.37	0.18	0.20	0.27	0.50	0.22	0.06	0.08	0.14	0.31	0.07	0.00	0.01	0.08	0.23
0.42	0.19	0.29	0.25	0.32	0.16	-0.02	-0.07	0.01	0.09	-0.07	-0.24	-0.22	-0.17	-0.07
0.98	0.89	0.84	0.81	0.80	0.80	0.80	0.79	0.78	0.77	0.76	0.75	0.74	0.74	0.73
0.02	0.05	0.11	0.02	0.09	0.02	0.30	0.05	0.01	0.09	0.00	0.01	0.04	-0.01	0.02
0.24	0.28	0.00	-0.09	-0.38	0.00	-0.04	-0.10	-0.46	-0.22	-0.24	-0.04	-0.26	-0.29	-0.55
0.54	0.57	0.51	0.39	0.55	0.43	0.51	0.52	0.39	0.47	0.37	0.43	0.50	0.40	0.47
0.96	0.90	0.86	0.84	0.83	0.82	0.81	0.80	0.81	0.81	0.81	0.79	0.79	0.79	0.79
0.39	0.28	0.35	0.25	0.35	0.05	0.04	0.05	-0.02	0.17	0.00	0.07	0.06	0.06	0.26
0.96	0.74	0.70	0.69	0.69	0.68	0.68	0.68	0.67	0.67	0.67	0.67	0.68	0.69	0.70
0.39	0.31	0.30	0.24	0.30	0.13	0.24	0.32	0.26	0.30	0.16	0.13	0.20	0.18	0.20
0.68	0.54	0.58	0.47	0.55	0.50	0.47	0.46	0.37	0.39	0.43	0.38	0.46	0.43	0.38
0.27	0.17	0.20	0.19	0.30	0.13	0.11	-0.04	-0.04	0.14	-0.06	-0.06	-0.10	-0.09	0.05
0.66	0.52	0.48	0.36	0.25	0.09	0.04	-0.01	-0.04	-0.03	-0.02	0.04	0.12	0.13	0.17
0.97	0.94	0.91	0.88	0.86	0.83	0.81	0.79	0.77	0.76	0.74	0.72	0.71	0.69	0.68
0.96	-0.02	-0.19	-0.31	-0.34	-0.36	-0.37	-0.33	-0.32	-0.32	-0.36	-0.39	-0.40	-0.40	-0.40
0.62	0.68	0.58	0.62	0.62	0.57	0.61	0.57	0.59	0.59	0.57	0.53	0.51	0.57	0.53
0.99	0.97	0.96	0.95	0.94	0.92	0.91	0.90	0.89	0.88	0.86	0.85	0.84	0.83	0.82

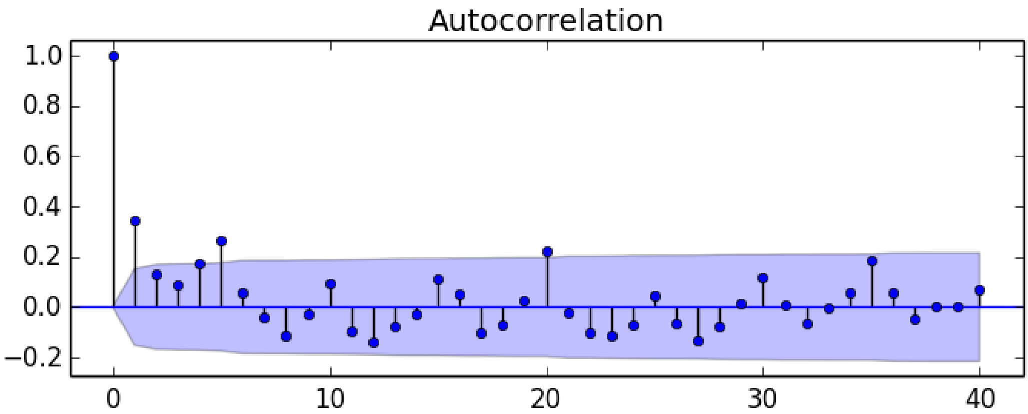
FIGURE 5.1: Correlation coefficients Matrix Colour Coded

Figure 5.1 represents the mean peak time daily mean correlation of observed locations. The first row is of junction **1/13/2** daily mean for the peak hour 10pm. The columns are the correlation of the daily lag i.e, Lag 1 is equal to -1 day, Lag 2 is -2 days and so on. The colour range demonstrates the correlation to the Standard Travel Time (STT) from positive value 1 correlation is closer Green to negative value -1 correlation Red and similarly yellow is closer to 0 is little or no correlation.

The algorithm for the correlation is based on the relationship between the correlation coefficient matrix, 'P', and the covariance matrix, 'C', is

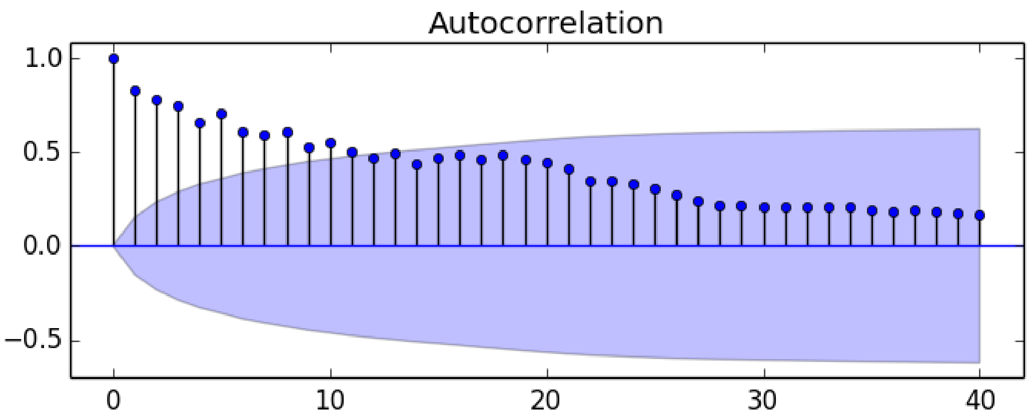
$$P_{ij} = \frac{C_{ij}}{\sqrt{C_{ii} * C_{jj}}}$$

The results show there mostly a correlation with the Lag of -1 and gradual deterioration the more distant the lag becomes. In some cases there is a slight increase in correlation Lag -4 which is a week in business days terms. The figures [5.2,5.3,5.4] also clarify the daily and weekly correlation. These correlograms represent High, Medium and Low volatility.



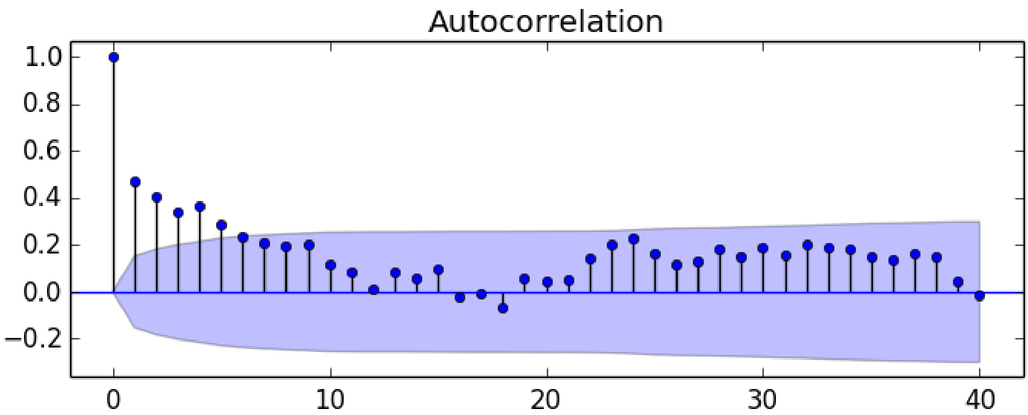
Ordinary Least Squares	
No. Observations	180
Intercept	40.9404
Adj. R-squared	0.347
Prob (F-statistic)	2.70
Lag 1,2,3,4,5,6	0.39, 0.16, 0.01, 0.18, 0.04, -0.07
Cond. No.	2.75

FIGURE 5.2: Auto Correlation of most volatile time 30/7/1 ['8:00', '8:59']



Ordinary Least Squares	
No. Observations	180
Intercept	40.9404
Adj. R-squared	0.347
Prob (F-statistic)	2.70
Lag 1,2,3,4,5,6	0.39, 0.16, 0.01, 0.18, 0.04, -0.07
Cond. No.	2.75

FIGURE 5.3: Auto Correlation of most volatile time 13/2/1 ['8:00', '8:59']



Ordinary Least Squares	
No. Observations	180
Intercept	40.9404
Adj. R-squared	0.347
Prob (F-statistic)	2.70
Lag 1,2,3,4,5,6	0.39, 0.16, 0.01, 0.18, 0.04, -0.07
Cond. No.	2.75

FIGURE 5.4: Auto Correlation of most volatile time 17/6/1 ['8:00', '8:59']

5.1.1.1 Standard Travel Time (STT) Conclusion

As a result data models for each observed location above 0.5 exists for 50% and is positive for another 49% for lagged -1 day. This percentage slowly decreases further down the historical path with a slight increase in lag -5 which represent lag -1 week. Therefore it may be beneficial to maintain the lag -1 week in the data model. With the correlogram in Figure 5.2 demonstrates a clear correlation between lag week -1. Figures [5.3 , 5.4] do not contain such correlation. In the section 5.1.5 a comparison of exponential moving average and historical lagged data is evaluated.

5.1.2 Weather Model Selection

In this section the STT correlation with the weather conditions of rain (dailyrainMM) and temperature (TemperatureC). The exploration section has identified these attributes 4.3.2.1 as the variables suitable for weather prediction. The same correlation metric is used for weather that was used in the STT modelling section 5.1.1 which is the correlation coefficient matrix and the covariance matrix. The figure [5.55.6] are visualisations using Google Maps of the correlation of rainfall and figures [5.55.6] are visualisations of the correlation of temperature. The visualisations circles on the map provides the strength of the correlation by the size of the circle and the colour representing the weather station it has the strongest correlation too. The circles are overlaying its related observed location. Negative correlation is apparent in the map when the circle has an opacity of 0.5 making it half transparent. The range values for correlation is between -1 and 1. On the maps the circle represent the station with the correlation value furthest from 0 whether it is negative or positive. The yellow circles provide the scale and transparency as an example of the value representation.

For example, if green Blackrock is value -0.3 and the red Artane value is 2.9, then the map will display the value -0.3 represented by the green BlackRock.

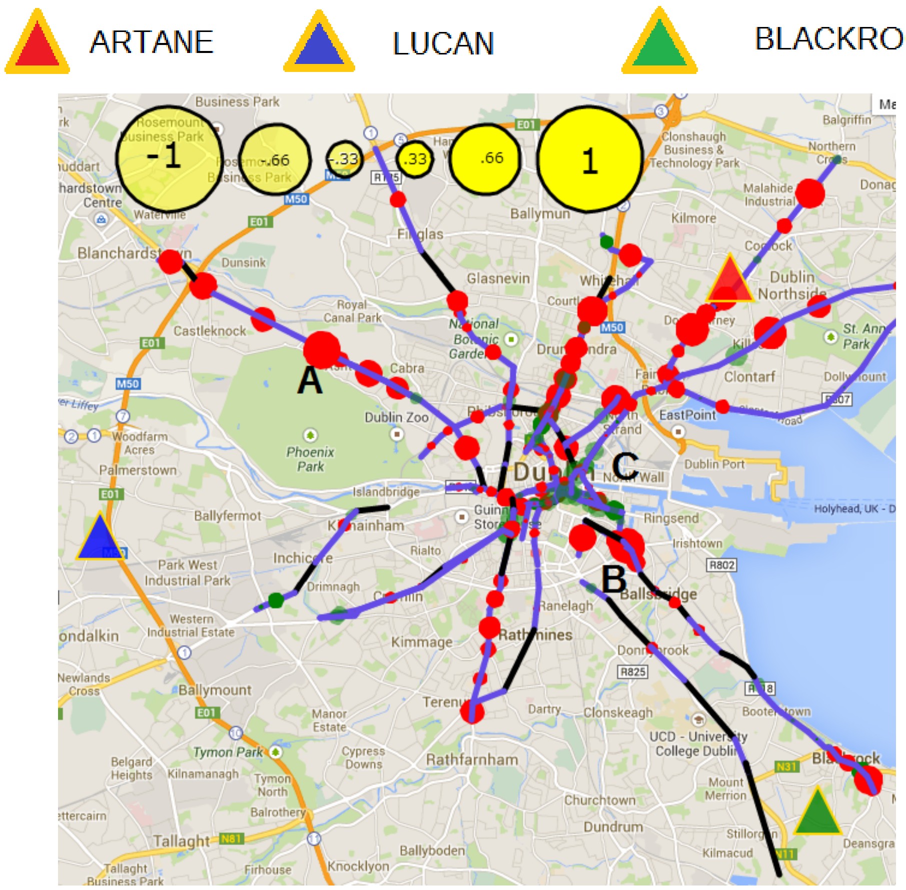


FIGURE 5.5: Correlation of Rain Map Direction Inbound Peak Times

In figure 5.5 the strongest rain correlation for peak times inbound is dominated by the weather station IDUBLINC2 in Artane Dublin 5. The correlation is also mostly positive. The circle **A** at location 4/1/1 and **B** at location 30/2/1 are highly affected by rain. The dailyRainMM correlation for 4/1/1 is 0.36 and the correlation for 30/2/1 is 0.34. Many of the other correlations are less significant and contradictory to adjacent locations where the correlation is positive. This is demonstrated by area at **C** where the correlations are also in green for weather station ILEINSTE8, Blackrock Dublin 4, see table 5.1.2. Both streches of road along the coastline see little or no correlations.

Location	Correlation
27/2/1	-0.22
28/8/1	0.19

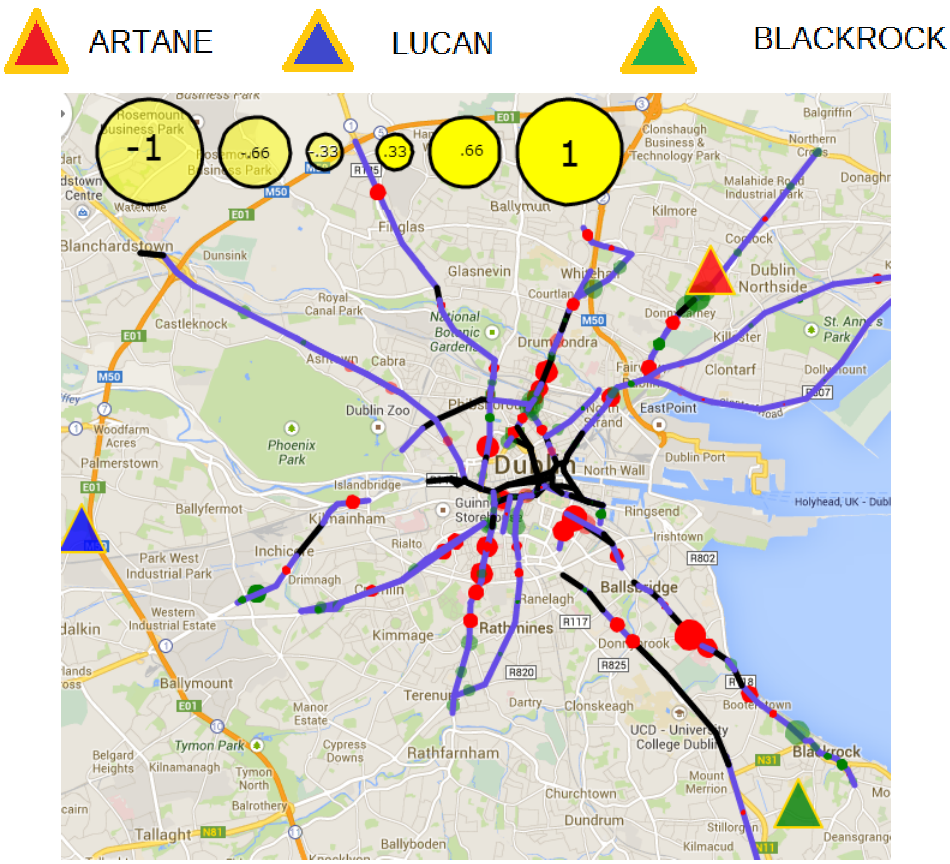


Figure 5.6: Correlation of Rain Map Direction Outbound Peak Times

In figure 5.6 the strongest rain correlation for peak times outbound is spread between the weather station IDUBLINC2 in Artane Dublin 5 and ILEINSTE8, Blackrock Dublin 4. Weather ICODUBLI2 in Lucan is not the main influence at any point. The peak outbound results are not as strong as the impact as the peak inbound. The most likely reason for this is that the weather has a stronger impact in the morning times. Peak times during the business days Monday - Friday are mostly at the hours 8am and 9am with some exceptions see table 4.4 compared to outbound peak times which is mostly after 2pm, see table 2.4.

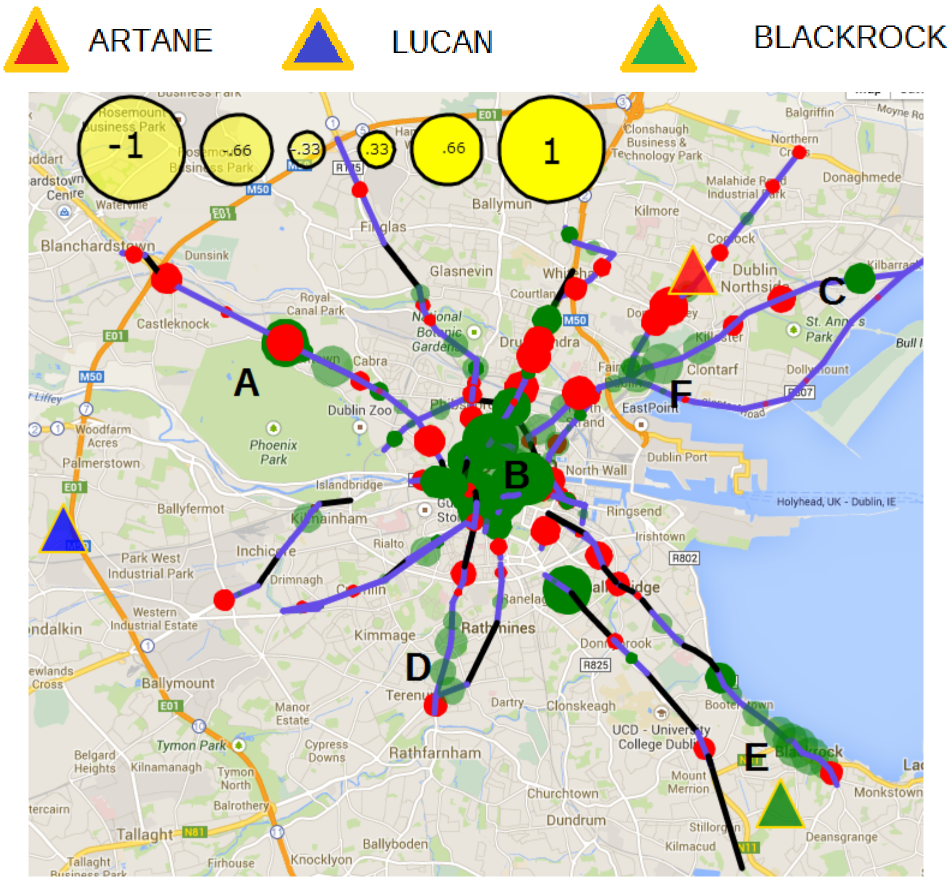


FIGURE 5.7: Correlation of Temperature Map Direction Inbound Peak Times

In figure 5.7 the strongest Temperature correlation for peak times inbound is spread between the weather station IDUBLINC2 in Artane Dublin 5 and ILEINSTE8, Blackrock Dublin 4. Weather ICODUBLI2 in Lucan is not the main influence at any point. When the temperature is high and is warm the map appears to have a positive correlation in the national parks Phoenix Park and St. Annes labeled **A, C** in the figure. The city centre has a largely positive influence. The indicates that when the temperature is warm the traffic volume increase to both the nation parks and the city centre. The suburban areas on the other hand traffic increases when the weather is cold as marked on the map at **D, E, F**. This indicates people are more likely to use their cars as transport in cold weather.

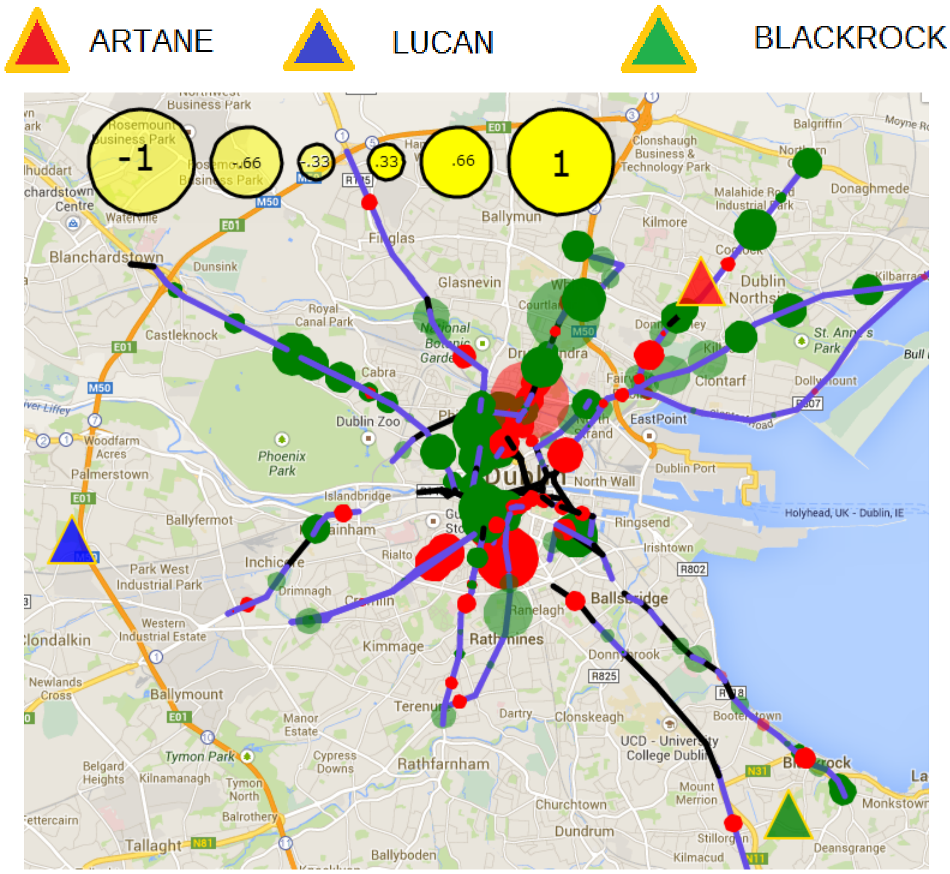


FIGURE 5.8: Correlation of Temperature Map Direction Outbound Peak Times

In figure 5.8 the Peak Outbound correlation map is similar to the Peak Inbound correlation map.

TABLE 5.2: Correlation coefficients Rain and Temperature Peak Times

	Lucan		Blackrock		Artane	
Range	Rain	Temp	Rain	Temp	Rain	Temp
>= **1** & >= **0.66**	0%	0%	0%	0%	0%	0%
>= **0.33** & <= **0.66**	5%	0%	7%	2%	8%	8%
>= **0.0** & <= **0.33**	71%	63%	40%	58%	65%	59%
<= **0.0** & >= **-0.33**	28%	31%	60%	31%	34%	31%
<= **-0.33** & >= **-0.66**	0%	1%	3%	0%	2%	8%
>= **-1** & <= **-0.66**	0%	0%	0%	0%	0%	0%

5.1.3 Weather Model Selection Conclusion

In table 5.2 the correlation of weather variables to the observed location are on the low range between -0.33 to 0.33. This indicates that not much impact on weather to the locations. Any location lower than -0.33 or above 0.33 implies the further investigation on the quality of the road or design of the road network may need further investigation.

The benefit of having Artane and Blackrock weather variables has significant importance to predicting STT and will used as features in the prediction model.

5.1.4 Spatial Model Selection

In spatial model selection the focus is determining the best predictive features for the standard travel time (STT). The number of attributes vary between different locations. The number of neighbours using the vertex method may can range from 1 to 39, excluding itself. Vertex is the point of where two lines meet 5.9. The figure shows 9/10/1 Inbound neighbours as white arrows and Outbound arrows in bright blue. The spatial modelling compares the features using both Inbound and Outbound together and then comparing Inbound locations with Inbound neighbours or Outbound with Outbound neighbours and the highest correlated neighbour. Using all neighbours as features for the predictive model is considered non feasible in this paper due to the complexity of differences in locations with the number of neighbours. The sample subset matrix in figure 5.10 illustrates this complexity. For the scoring of spatial models correlation based on the relationship between the correlation coefficient matrix, 'P', and the covariance matrix, 'C' [26], is:

$$P_{ij} = \frac{C_{ij}}{\sqrt{C_{ii} * C_{jj}}}$$

This correlation score is the primary scoring factor for each location. Each data will only contain peak times of each location during business days Monday to Friday.

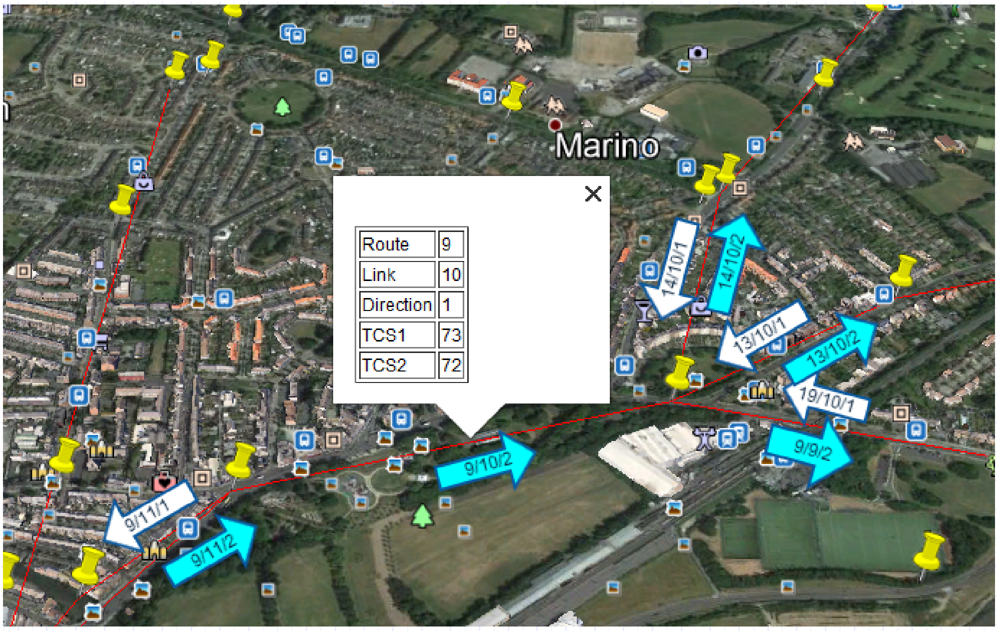
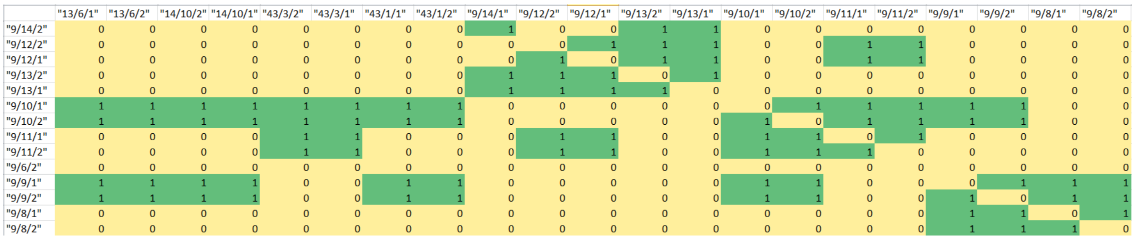


FIGURE 5.9: Vertex neighbouring with no filter



	"13/6/1"	"13/6/2"	"14/10/2"	"14/10/1"	"43/3/2"	"43/3/1"	"43/1/1"	"43/1/2"	"9/14/1"	"9/12/2"	"9/12/1"	"9/13/2"	"9/13/1"	"9/10/1"	"9/10/2"	"9/11/1"	"9/11/2"	"9/9/1"	"9/9/2"	"9/8/1"	"9/8/2"
"9/14/2"	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0
"9/12/2"	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	0	0	0	0
"9/12/1"	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	1	1	0	0	0	0
"9/13/2"	0	0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0
"9/13/1"	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0
"9/10/1"	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	1	1	1	1	0	0
"9/10/2"	1	1	1	1	1	1	1	1	0	0	0	0	0	1	0	1	1	1	1	0	0
"9/11/1"	0	0	0	0	1	1	0	0	0	1	1	0	0	1	1	0	1	0	0	0	0
"9/11/2"	0	0	0	0	1	1	0	0	0	1	1	0	0	1	1	1	0	0	0	0	0
"9/6/2"	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
"9/9/1"	1	1	1	1	0	0	1	1	0	0	0	0	0	1	1	0	0	0	1	1	1
"9/9/2"	1	1	1	1	0	0	1	1	0	0	0	0	0	1	1	0	0	1	0	1	1
"9/8/1"	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1
"9/8/2"	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0

FIGURE 5.10: Sample vertex neighbouring with matrix

Any location neighbours will contain one or more neighbouring locations. The STT for each neighbour converted to a single feature using the mean of each set of neighbours. As described in section 5.1.1 historical data available is only available after 1 day. The correlation measure will be compared to a minimum of one day lagged. The section also has proven that lag one day is highest scoring correlation. with this the spatial correlation be measured against the same historical distance measure. In the figure 5.11 the sample of the results is described in table 5.3.

TABLE 5.3: Description sample spatial correlation result figure

Label	**Description**
itself-1	The location with lagged STT of one day
inout-1	Inbound & Outbound neighbours with lagged STT of one day
samedir-1	Directionally similar neighbours with lagged STT of one day
oppdir-1	Directionally opposite neighbours with lagged STT of one day
max-1	Highest correlated neighbour lagged STT of one day value

Evaluating the sample correlation result in figure 5.11 each location varies in the strength of the correlation from the STT. inout-1, oppdir-1 and samedir-1 show little variance in values. These variables also show minor differences to the STT lagged -1.

Max-1 highlifhted in blue 5.11 variable is the most likely to pose the significant difference. Column **13/5/1** at **max-1** the value is **-0.16** while **15/12/1** at **max-1** the value is **0.98**.

	itself-1	inout-1	max-1	oppdir-1	samedir-1
"46/6/2"	0.37	0.34	0.26	0.37	0.30
"14/10/1"	0.81	0.82	0.81	0.81	0.84
"16/2/2"	0.84	0.83	0.83	0.84	0.82
"16/2/1"	0.48	0.45	0.58	0.48	0.41
"14/10/2"	0.65	0.63	0.64	0.65	0.61
"16/1/1"	0.25	0.31	0.27	0.25	0.41
"16/1/2"	0.36	0.41	0.64	0.36	0.49
"43/3/1"	0.58	0.60	0.61	0.58	0.61
"43/3/2"	0.82	0.80	0.74	0.82	0.77
"22/9/1"	0.79	0.80	0.74	0.79	0.81
"13/6/2"	0.32	0.29	0.05	0.32	0.25
"22/2/1"	0.34	0.29	0.05	0.34	0.21
"46/4/2"	0.17	0.19	0.23	0.17	0.21
"13/4/1"	0.39	0.39	0.47	0.39	0.38
"13/4/2"	0.73	0.72	0.63	0.73	0.70
"1/6/1"	0.77	0.76	0.67	0.77	0.74
"1/6/2"	0.41	0.46	0.56	0.41	0.51
"13/5/1"	0.50	0.44	-0.16	0.50	0.37
"1/5/2"	0.13	0.18	0.23	0.13	0.24
"1/5/1"	0.50	0.47	0.41	0.50	0.41
"13/5/2"	0.46	0.43	0.36	0.46	0.36
"6/1/1"	0.64	0.63	0.13	0.64	0.61
"5/2/2"	0.48	0.48	0.75	0.48	0.48
"15/12/1"	0.67	0.60	0.98	0.67	0.51

FIGURE 5.11: Sample spatial correlation result

TABLE 5.4: Overview of Spatial Correlation Results

	itself-1		inout-1	max-1	oppdir-1	samedir-1
$>=$ **0.5** & $<=$ **1**	33.96%		31.55%	33.96%	33.96%	35.03%
$>=$ **0.0** & $<=$ **0.5**	62.30%		65.24%	53.48%	62.83%	59.89%
$>=$ **0.0** & $<=$ **0.5**	3.21%		3.21%	12.30%	3.21%	4.81%

5.1.4.1 Spatial Model Selection Conclusion

The table 5.4 is the overview of the correlation result scores. It demonstrates the each data model works with similar effect as a potential variable to best predict STT. The correlations above 0.5 are above 30%. The benefit of having each variable as a feature in a prediction has been proven that it may have little significant difference with compared to the SST lagged -1. The one exception to this is the variable max-1. In the prediction model section 5.1.5 the **max-1** (strongest correlation neighbour) will be used as a feature. The **inout-1, oppdir-1, samedir-1** not be tested as features.

5.1.5 Prediction Model Fitting

In this section the prediction algorithms are applied to the dataset comprised of the features discussed in the Standard Travel Time Selection 5.1.1, Weather Model Selection 5.1.2, Spatial Model Selection 5.1.4. The objective of this section is the generate the best fit model for estimating next day STT. In this a description of the datasets are provided,

To estimate the best fit model a scoring mechanism is needed to build a comparison of algorithms for each observed location.

5.1.5.1 Predictive Datasets

Spatial data and Weather data variables account for some of the noise associated with the STT value. The lagged values provide a measure forecasting. Each 563 observed locations is a unique dataset comprised of the same features described in 5.5. A sample dataset is in Appendix E.

TABLE 5.5: Description sample spatial correlation result figure

Feature	**Description**
STT	**The predictive label lagged STT -1 day**
S_MAX	Highest Correlated neighbour
W_DLR	Artane Rainfall in Millimetres
W_DLT	Artane Temperature in Celcius
W_ILR	Blackrock Rainfall in Millimetres
W_ILT	Blackrock Temperature in Celcius
STT1	lagged STT -1 day
STT2	lagged STT -2 day
STT3	lagged STT -3 day
STT5	lagged STT -5 day or -1 week

To avoid over fitting cross validation is used split the dataset into training and testing datasets. The test size set to 30% of the full datasets resulting into a 93 to 40 split. The dataset contain 138 samples. The missing 5 samples are due to the removal of missing values from the moving average mechanism.

5.1.5.2 Prediction Algorithms

The algorithms chosen are to handle the noise that is not accounted for in the current features in the datasets. The STT prediction values are can be highly volatile (see section 4.3.1.3) and weather and spatial features can account for some of the noise, see table

5.6. The algortihm are part of the Sklearn toolkit within Python. Linear regression is one of oldest algorithm. Both Bayesian Ridge and Online Passive Aggressive Regressor are modern linear algorithm that account for noise by setting bounds on the residuals. While Support Vector Regression is a popular non-linear algorithm.

TABLE 5.6: Estimation algorithms

Algorithm	**Parameters**
LinearRegression	Fit Intercept , Normalise Coeffients
LinearRegression	Default
BayesianRidge	Default
Online Passive Aggressive Regressor	Default
Support Vector Regression	kernel rbf

5.1.5.3 Evaluating estimator performance

To avoid over fitting cross validation is used to split the data sets in training and testing data set. The samples are split to the ratio *7:3*. Due to the limited size of the dataset some over fit still occurs.

The python Sklearn scipy kit provides approaches to evaluate the quality of estimation algorithm for Regression models [26].

Regression models scoring mechanisms known as Mean Absolute Error (MAE), Mean Squared Error (MSE), Regression Coefficient score (R2), Explained Variance Score (EVS) are all available scoring mechanisms in Sklearn.

MAE is used for scoring the best prediction algorithm. MAE score measure is best when closer to *0*. EVS best score is *1* and worst is *0* and provide a easier mechanism for comparison to other **OL** scores.

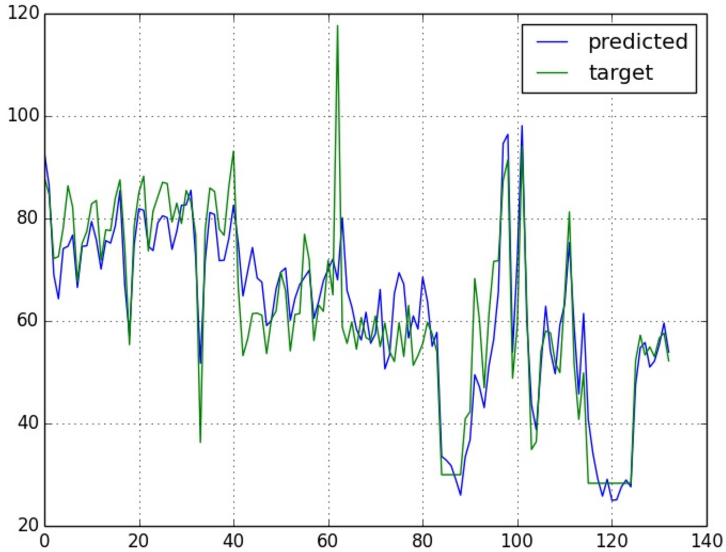
5.1.5.4 Prediction Results

As result it was achievable to build a generic model that would match a good level of accuracy for each observed location **OL**. Appendix F provides detailed results while the result are plotted in figures 5.12 5.14 5.13 5.15 give a overview of the accuracy.

- Linear regression
  Ordinary Least Square Linear Regression scores the best prediction algorithm in the majority of cases. This is due to the standard distribution (STD) of the values having a value of less than *1*. STD of value less than *1* occurs on *39* of the *563*

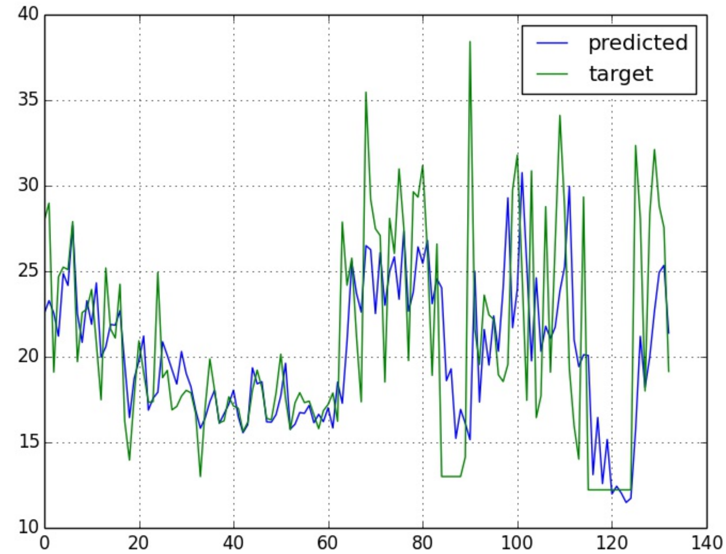
predictors. For the other Linear Regression algorithms the perform best when the parameters are set to *fit intercept is true* and *normalize is true.*

FIGURE 5.12: Linear Regression Fit intercept and Normalise 14/3/2



- Support Vector Regression
  (SVR) can be used when the kernal is linear. This algorithm also performs well. SVR is a type of Support Vector Machine algorithm that can be fit for linear models.

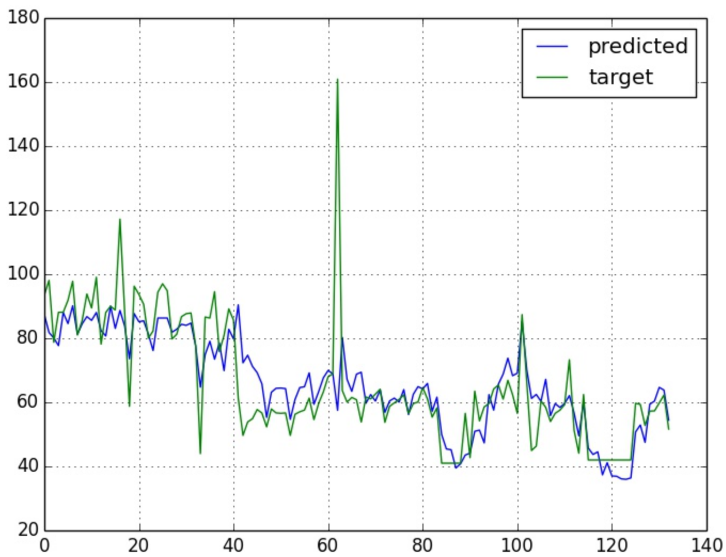
FIGURE 5.13: Support Vector Machine Regression 30/20/1



- Bayesian Ridge Regression

  Bayesian Ridge assumes gaussian process and is similar to ARD Regression without the normalisation of parameters [26].

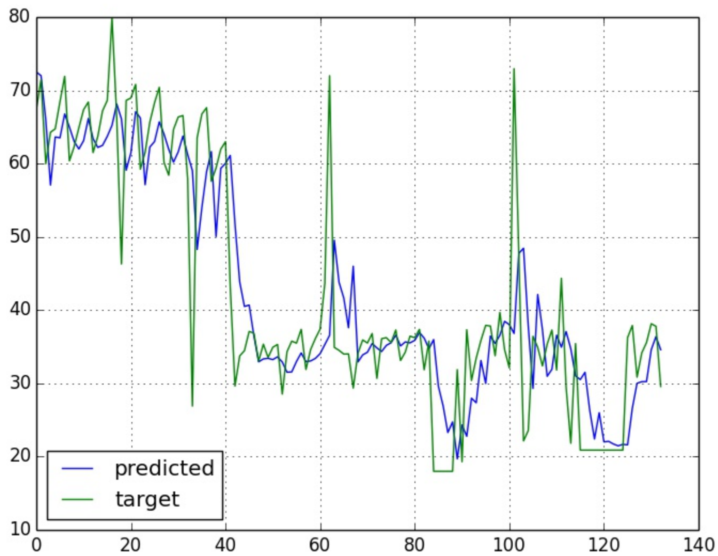
FIGURE 5.14: Bayesian Ridge Regression 31/5/2



- Online Passive Aggressor Regression

  A variation of the SGD Regressor the Online Passive-Aggressive [26, 30].

FIGURE 5.15: Online Passive Aggressor Regression 18/6/1



As a result the figure 5.16 the spatial distribution of algorithms in Dublin at off-peak time with inbound direction. This highlights the high use of the linear model of Bayesian Ridge.

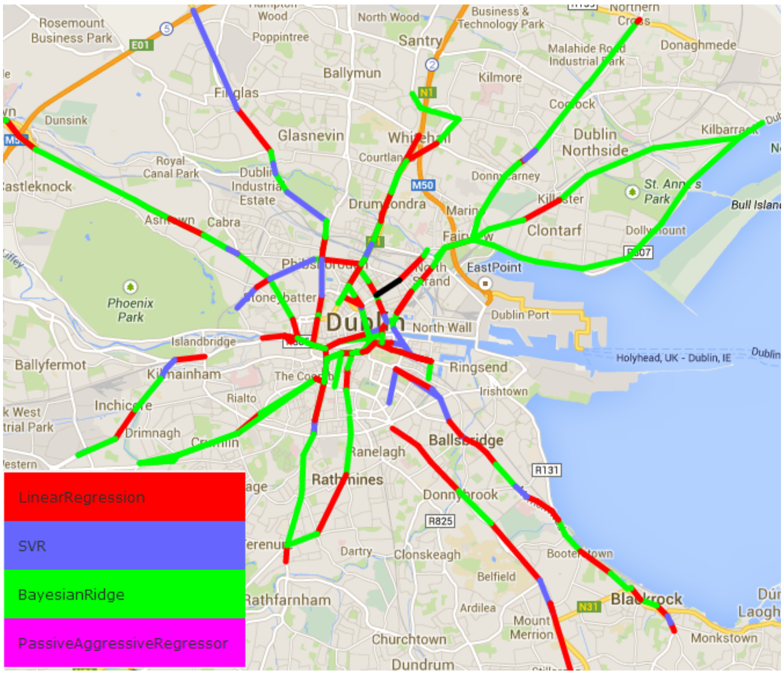


FIGURE 5.16: Off-peak and Inbound Algorithm Map

5.2 Twitter Traffic Modelling

The objective of this section is to analyse the approach of using the traffic domain tweets to extract tweets from real-time data that is related to the traffic domain.

- Passive Aggressive Classifier

Tokeniser

```
TfidfVectorizer(
        analyzer='word',token_pattern=r'[a-z]{3,}',
        use_idf=True, strip_accents='unicode',
        sublinear_tf=True,  max_df=0.95,  min_df=0.05,
        stop_words='english')
```

Result

```
\lablel{algorithm_tweetresult}
        precision      recall      f1-score    support     samples
```

```

          Traffic        1.00      0.31      0.48      5000
      Non-Traffic        0.59      1.00      0.74      5000

      avg / total        0.80      0.66      0.61      10000
```

Passive Aggressive Classifier is an example of one of the algorithms used to classify the real-time traffic tweets. Using TfidfVectorizer in Python tweets are tokenised to fit the predictive model for the classifier. The result contain the result of using samples taken from the traffic domain and real-time data. As expected tweets from the traffic domain is 100%. For a successful result for extracting traffic tweets from real-time data must not be 100%. Table 5.7 show the result of real-time tweets classified as traffic related.

Table 5.7: Real-time Tweets Classified as Traffic

Accuracy	**Text**
True Positive	No better way to start your day with a car crash, and then forgetting about the banana in my pocket going through security...
False Positive	We're gonna crash vine if we keep doing this
False Positive	Lyndsay Lohan looks like a car crash.... She is wrote off #ChattyMan
True Positive	I bloody hate waiting #delays http://t.co/Yh55PrfQK3
True Positive	There's after been a crash outside my estate, 3 fire trucks and 3 ambulances

5.2.1 Twitter Conclusion and Analysis

Not enough variations of tokenising, feature selection and algorithm have been tested. Part of Speech would be useful to identify celebrities or place name that may improve the elimination of the false positives. As a proof of concept the classification approach worked as designed. The real-time traffic tweet could be used to provide further analysis on traffic delays. Results section 6.2 provides more details such implementation.

Chapter 6

Results and Conclusions

6.1 Big Data

As a result of using NoSQL to overcome the challenges of the four V's the approach stored volumes of data that on a single machine RDMS system would of been problematic 6.1.

Table 6.1: Volumes of Data

Data Source	*Items*	*No. of Documents*
Traffic Observations	501,402,840	8,356,714
Real-time Tweets	3,048,310	116
User Tweets	5,267	5,267
Weather Records	229,311	2,103

Throughout the sections of data collection, exploration and modelling techniques of *divide and conquer* allowing for the data to be aggregated to the point 563 prediction models were created for peak and off-peak times, see chapters 4 and 5. The work was done on a single machine but the database system and the indexing mechanism provide scalability that allow adding more machines to that database for velocity and volume storage.

6.2 Visualisation

Google Maps, JQuery and Pythons web framework Django was heavily utilised to create an application that was not only for exploration spatial data but analysing the patterns of traffic such as volatility, analyse past events and predict traffic. The resulting application provides functionality for visual analysis on, volatility of roads, the effect of weather on locations with respect to travel times and performing analysis on locations for a specific time, see figure 6.1.

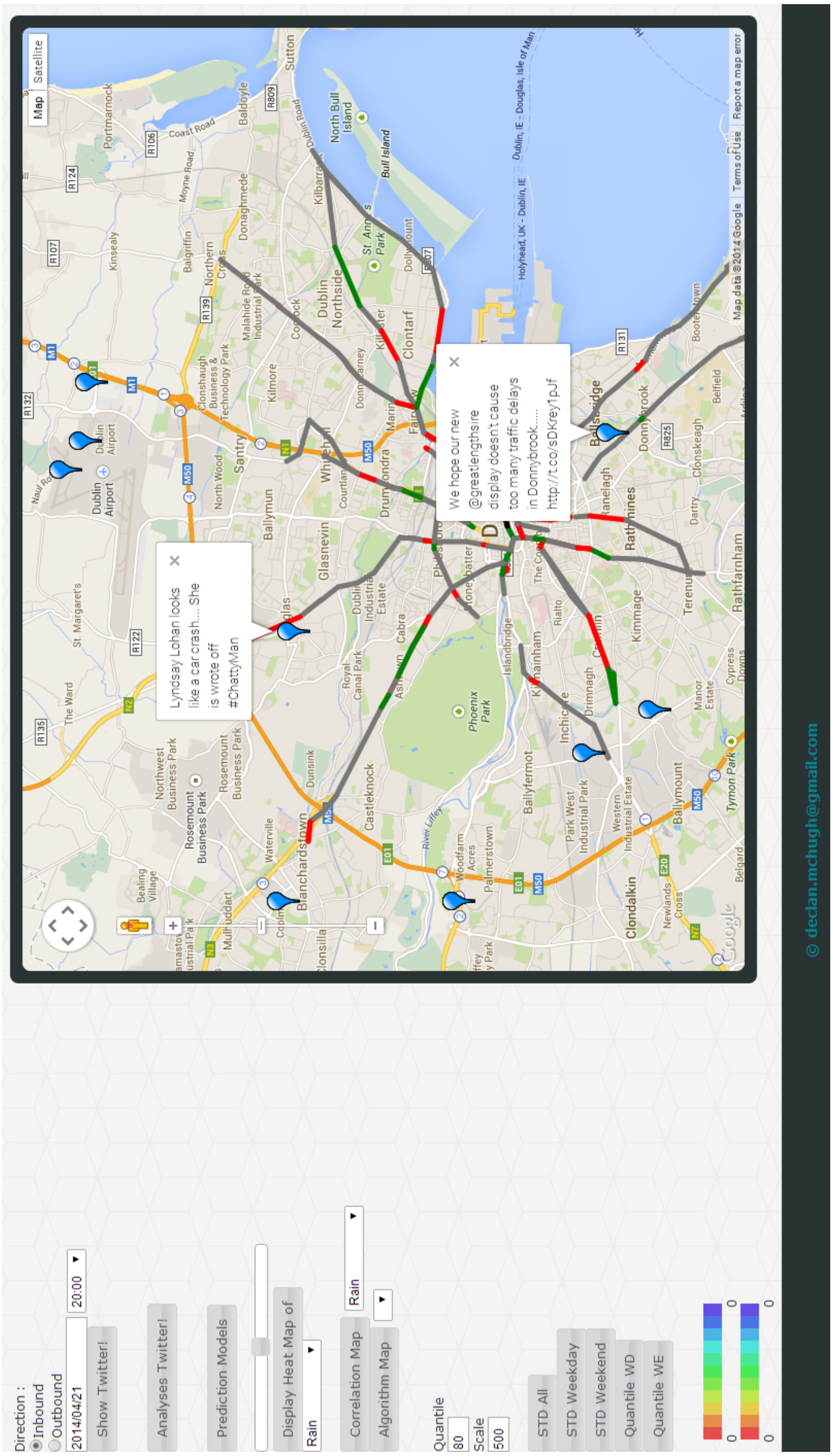


FIGURE 6.1: Dashboard Analysis for 21/04/2014 8pm to 9pm

6.3 Traffic Analysis

6.3.1 Seasonality

Measuring all aspects of seasonality came as a challenge due to the lack of quality data. This is a common problem when dealing with open data. Annual, quarterly and monthly trend analysis could not be performed. The work concentrated mostly on weekly and daily trends. The difference in peak traffic times of weekend and weekday was clear. Week-end peak times usually centred on mid-day and week-day peak time was in the morning or evening time from 2pm on.

6.3.2 Weather

Trends in the traffic patterns based on weather were identified. It was demonstrated that high temperature people were more likely to travel into the city centre and public parks as travel times would spike when this happened. The impact with rainfall was with an increase in travel time near small villages at Castleknock, Raheny, Drumcondra among others. This could indicate the village cannot handle the increase in traffic flow while people are more likely to drive than walk in the rain to their shopping. This may indicate that people will travel short distances to a local village when it is raining but when it is dry and warm people may travel to do there shopping in the city centre.

6.3.3 Prediction Model

The final prediction model became a hybrid of SARIMA and Multivariate ARIMA. It may be considered the linear outperformed the non-linear algorithms for prediction accuracy. It also needs to be considered that the data model designed was a generic model that would fit all observed locations. One of the limitations was the an artificial neural network could not be implemented as part of Python SciPy Toolkit. Therefore only one non-linear model was compared to four linear algorithms. Due to the number of overall samples tested there may be a case that over-fitting affected some results. Some algorithms benefited due to the lack of volatility in the standard deviation.

6.3.4 Analytics Dashboard

In figure 6.1 the dashboard contains a some false positives for the time **21/04/2014 8pm to 9pm**. The approach of obtaining tweets from a specific domain to classify real-time data feasible. It this classifying tweets using more intelligent classification models is needed. In figure 6.1 it is reasonable to see why *"Lyndey Lohan looks like a car crash.. she is wrote off #ChattyMan"* is classified as a traffic related tweet. At the same time *"We hope our new display doesn't cause too many delays in Donnybrook .... http://t.co/sDKrey1pJf"*.

6.4 Future Work

For future due to the data quality the number of samples was limited. Seasonality comparisons were not tested to account for school holidays over the summer or winter breaks. In the visualisation of prediction results the observed locations did not apply any sort of key performance indicator (KPI). The current state applies higher or lower than predicted by colouring the prediction red or green. A method for using a colour range to allow the reader understand the scale the prediction deviates from the actual result.

More evaluation on the algorithms, tokenising mechanisms and scoring of the Twitter traffic classification is necessary to improve the quality of the result. Text mining techniques such as Stop Word removal and Part-of-Speech would likely help to improve the classification.

Bibliography

[1] Duckwon Chung et el. Road traffic big data collision analysis processing framework. *IEEE*, 2013.

[2] Vinay Gavirangaswamy et el. Assessment of arima-based prediction techniques for road-traffic volume. 2013.

[3] DubLinked. Trips data, 2012-2014. URL http://www.dublinked.ie/datastore/datasets/dataset-215.php.

[4] McCreadie et al. Scalable distributed event detection for twitter. *IEEE*, 2013.

[5] MongoDB. Big data explained @ONLINE, 2014. URL http://www.mongodb.com/big-data-explained.

[6] Kalman R. E. A new approach to linear filtering and prediction problems. *IEEE*, 1960.

[7] Dehuai Zeng et el. Short term traffic flow prediction using hybrid arima and ann models. 2008.

[8] Wei-Chiang Hong et al. Seasonal adjustment in a svr with chaotic simulated annealing algorithm traffic flow forecasting model. 2010.

[9] Kevin Keay and Ian Simmonds. The association of rainfall and other weather variables with road traffic volume in melbourne australia. 2004.

[10] Bei Pan et el. Crowd sensing of traffic anomalies based on human mobility and social media. *IEEE*, 2013.

[11] Stephen Dunne and Bidisha Ghosh. Weather adaptive traffic prediction using neurowavelet models. 2013.

[12] Yorgos J. Stephanedes. Dynamic prediction of traffic volume through kalman filtering theory rescue. 1983.

[13] Bowu Zhang et al. Traffic clustering and online traffic prediction in vehicle networks. *A Social Influence Perspective*, 2012.

[14] Yousef-Awwad Daraghmi et el. Space-time multivariate negative binomial regression for urban short-term traffic volume prediction. *IEEE*, 2012.

[15] Sri Krisna Endarnoto et al. Traffic condition information extraction & visualization from social media twitter for android mobile application. *IEEE*, 2011.

[16] Amit Sheth. Transforming big data into smart data: Deriving value via harnessing volume, variety, and velocity using semantic techniques and technologies. 2014.

[17] Shweta Pandey and Dr.Vrinda Tokekar. Prominence of mapreduce in big data processing. 2014.

[18] Howard Gobioff Sanjay Ghemawat and Shun-Tak Leung. The google file system, October, 2003.

[19] Duckwon Chung et el. United nations global pulse, 2012, big data for development: Challenges & opportunities. May 2012.

[20] Executive Office of the President. Office of science and technology policy. *IEEE*, 2012.

[21] Brito et al. Scalable and low-latency data processing with streammapreduce. *IEEE*, 2011.

[22] Pricilla Hancock Kristopher Reese, Russell Bessette. Knowyourcolors: Visual dashboards for blood metrics and healthcare analytics. *IEEE*, 2013.

[23] Carlos Costa Luís Bastiao Silva Louis Beroud and José Luis Oliveira. Medical imaging archiving: a comparison between several nosql solutions. 2013.

[24] Dr. Vincent Granville. Developing analytic talent: Becoming a data scientist. *Wiley; 1st edition*, 2014.

[25] Gong Jun et al. Forecasting urban traffic flow by svr. 2013.

[26] Machine Learning in Python Scikit-learn. Scikit-learn developers, 2010 - 2014. URL `http://scikit-learn.org/`.

[27] Wunderground. Wunderground, 2012 - 2014. URL `http://www.wunderground.com`.

[28] Daniel J Tulloch. A garch analysis of the determinants of increased volatility of returns in the european energy utilities sector since liberalisation. *IEEE*, 2012.

[29] Rohit Dhawan Sven F. Crone. Forecasting seasonal time series with neural networks: A sensitivity analysis of architecture parameters. *IEEE*, 2007.

[30] et al Koby Crammer. Online passive-aggressive algorithms. 2006.

Appendix A

Appendix Traffic Web Crawl

```
import bz2
import os

import attrdict
import urllib3
from datetime import datetime
from bs4 import BeautifulSoup

# check for extraction directories existence
 if not os.path.isdir('downloaded'):
     os.makedirs('downloaded')

 if not os.path.isdir('extracted'):
     os.makedirs('extracted')

def persistTrafficData(a):
    #if os.path.isfile("extracted/" + a):
     filename = a
     date = filename[4:12]

     argv = attrdict.AttrDict()
     argv.filename = a
     argv.date = date
    #del GisConvert
    #if "201310" in a:
     import smartcity.module.TransformTrafficData as tranformData
     import smartcity.module.WeatherCrawlExtract as wc

     date = datetime.strptime(date,"%Y%m%d")
    w = wc.crawlWeatherForDate(date)
     g = tranformData.processfile(argv)
```

```
    del g
    del w
    del argv

http = urllib3.PoolManager()
archiveDirectory = "http://www.dublinked.ie/datastore/local/DCC/trips/archive/"
http_pool = urllib3.connection_from_url(archiveDirectory)
stream = http_pool.urlopen('GET',archiveDirectory)
text = stream.data


# retrieve list of URLs from the webservers
soup = BeautifulSoup(text)
res = soup.find_all('a',href=True)
links = []
for n in res:
    if "." in n['href']:
        links.append(n['href'])

# only parse urls
for filename in links:
    if '.' in filename:

        # download the file
        archiveFile = archiveDirectory + filename
        outputFile = "downloaded/" + filename
        print("Start ",datetime.now())
        # check if file already exists on disk
        if os.path.isfile(outputFile) is not True:
            print("Skipping " + archiveFile)
            print("Downloading ",archiveFile)
            http_pool = urllib3.connection_from_url(archiveFile)
            rcsv = http_pool.urlopen('GET',archiveFile)
            # save data to disk
            output = open(outputFile,'wb')

            output.write(rcsv.data)
            output.close()


        if os.path.isfile("extracted/" + filename + ".csv") is not True:
            zfobj = bz2.BZ2File(outputFile, 'rb')
            try:
                #save extracted file
                f = open("extracted/" + filename + ".csv",'wb')
                f.write(zfobj.read())
                f.close()
```

```
            except (RuntimeError, TypeError, NameError) as error:
                print("Error: ",error)
            finally:
                zfobj.close()

            persistTrafficData(filename + ".csv")
        print("End ",datetime.now())
        exit()
```

Listing A.1: Traffic Observation Extraction

Appendix B

Appendix Transform Traffic Data

```
'''
Created on 3 Dec 2013

@author: declan
'''

import csv, os
from xml.etree import ElementTree
from datetime import datetime

from attrdict import AttrDict
import attrdict

from pymongo import Connection as mongoConn

connection = mongoConn('mongodb://localhost:27017/')
db = connection.traffic
collection = db.observation
collectionj = db.junctions

def processfile(argv):
    argv.timeframe = AttrDict()
    argv.mapset = {}
    if len(Junction.items.values()) == 0:
        csvreadjunctions(argv)
    if False:
        for value in Junction.items.values() :
            print(value)
    if True:
        csvreaddata(argv)

```

```
    return
listInvalid = {'1','22'}
def isValidObserbvation(value):
    if (value["route"] in listInvalid):
        return False
    return True
def formatdate(date_string):
    # expect string date format is "%Y%m%d-%H%M"
    result = ""
    try :
        result = datetime.strptime(date_string, "%Y%m%d-%H%M")
    except (RuntimeError, TypeError, NameError, ValueError) as error:
        print("Error format date:",date_string,error)
    return result


def csvreaddata(argv):
    #print argv.filename
    datefromfile = r"observation/" + argv.filename[4:12]
    print(datefromfile)
    result = os.path.isdir(datefromfile)
    argv.juncs={}
    if result is False:
        with open("extracted/" + argv.filename) as csvfile:
            print("getting extracted/" + argv.filename)
            csvreader = csv.reader(csvfile, delimiter=',')
            for i,row in enumerate(csvreader):
                # Validate row, a row must contain 8 values
                if len(row) > 7:
                    try :
                        dateOfObservation = formatdate(row[0].strip())
                        dayOfObservation = dateOfObservation.strftime('%Y%m
    %d')
                        hourOfObservation = dateOfObservation.strftime('%H'
    )
                        item = {'stt':int(row[4].strip()),
                                    'date': dateOfObservation}
                        observation = {'day': dayOfObservation,
                               'route': row[1].strip(),
                               'link': row[2].strip(),
                               'hour': hourOfObservation,
                               'direction':row[3].strip(),
                               'item':[]}
                    # Two junctions make up a link
                        junction1 = j.find(row[6])
                        junction2 = j.find(row[7])

```

```
                        # If junctions exist in observation transform row for DB
                        if (junction1 and junction2):
                            observation['_id'] = makeObservationId(observation)
                            jid = makeJunctionId(observation)
                            argv.juncs[jid]={
                                    "_id":jid,
                                    "junction1":junction1,
                                    "junction2":junction2
                                    }
                            keys = argv.timeframe.keys()
                            if not observation['_id'] in keys:
                                argv.timeframe[observation['_id']] = {}
                            keys = argv.timeframe[observation['_id']].keys()
                            if not hourOfObservation in keys:
                                argv.timeframe[observation['_id']][hourOfObservation] = observation
                            argv.timeframe[observation['_id']][hourOfObservation]['item'].append(item)
                    except (AttributeError, RuntimeError, TypeError, NameError, ValueError) as error:
                        db.observation_errors.insert({
                                "item":row,
                                "filename":file_name,
                                "linenumber":i
                            });
                        print("ERROR:",row,error)

            for o in argv.timeframe:
                print ("Hour > " + o)
                for o1 in argv.timeframe[o]:
                    collection.insert(argv.timeframe[o][o1])


def csvreadjunctions(argv):
    parser = ElementTree.XMLParser()
    tree = ElementTree.parse("junctions.kml",parser)
    sites = tree.iterfind(".//*[@name=\"SiteID\"]")
    points = tree.findall('.//{http://www.opengis.net/kml/2.2}coordinates')
    my={}
    loc={}

    p = points
    s = sites
    for i,x in enumerate(s):
        my[i] = x.text
```

```
    for i,p in enumerate(points):
        loc[my[i]] = p.text

    with open("junctions.csv") as csvfile:
        csvreader = csv.reader(csvfile, delimiter=',')
        for y,row in enumerate(csvreader):
            if (y > 0):
                if (validatecsvjunctions(row)):
                    jun = {'point':loc[row[0].strip()],'id': row[0].strip(), 'lon': row[1].strip()[:8], 'lat': row[2].strip()[:8],'desc':row[3].strip()}
                    j.add(jun);

class Junction:
    items = dict()

    def __init__(self, *entries):
        self.__dict__.update(entries)

    def add(self, item):
        self.items[item['id']] = item

    def find(self,genid):
        if (self.items.get(genid)):
            return self.items[genid]
        return ''


def makeObservationId(item):
    return item["route"] + "/" + item["link"]  + "/" + item["direction"] + "/" + item["day"]+ "/" + item["hour"]

def makeJunctionId(item):
    return item["route"] + "/" + item["link"]  + "/" + item["direction"]


def validatecsvjunctions(argv):
    result = True;
    for w in argv:
        if (len(w) == 0):
            result = False
    if (not (argv[0].isdigit())):
        result = False
    if (not (argv[1].isdigit())):
        result = False
    if (not (argv[2].isdigit())):
        result = False
```

```
    return result

j = Junction()
filename = "day-20140413.csv.bz2.csv"
filepath = ""
date = ''
find_errors = True
if __name__ == "__main__":

    date = filename[4:12]
    if find_errors:
        for root, _, files in os.walk("extracted/"):
            for file_name in files:
                with open("extracted/" + file_name) as csvfile:
                    csvreader = csv.reader(csvfile, delimiter=',')
                    for i, row in enumerate(csvreader):
                        if not len(row) > 7:
                            item = {
                                "item":row,
                                "filename":file_name,
                                "linenumber":i,
                                "_id": file_name + "." + str(i)
                            }
                            db.observation_errors.insert(item)


    else:
        argv = attrdict.AttrDict()
        argv.filename = filename
        argv.filepath = filepath
        argv.date = date

        processfile(argv)
```

Listing B.1: Transform Traffic Data

Appendix C

Appendix Available Weather Stations

```
Lucan, Co. Dublin
Tallaght, Dublin
Templogue / Terenure, Dublin
Irish Climate Analysis \& Research Units, Maynooth
Ballygall Dublin, Glasnevin
Fairview Dublin, Fairview
Ballsbridge, Dublin City, Dublin
Artane, Dublin 5, Dublin
Blackrock, Co. Dublin, Dublin
Swords West, Swords
Dunshaughlin, Meath
Newtown, Enfield/Kilcock
Naas
Cherrywood, Loughlinstown
Earlscliffe, Ceanchor Road, Baily
Ardmore Park, Bray
Southern Cross, Bray
```

Listing C.1: Station Location

Appendix D

Appendix Sample Tweet


```
/* 0 */
{
    "_id" : ObjectId("534bfab9c009e418f4c742a1"),
    "user_profile_sidebar_fill_color" : "EFEFEF",
    "user_created_at" : "Sun Apr 04 23:58:55 +0000 2010",
    "place_country_code" : "GB",
    "user_screen_name" : "noelbrodie",
    "place_country" : "United Kingdom",
    "user_following" : "None",
    "place_full_name" : "South West, United Kingdom",
    "user_id_str" : "129638809",
    "place_contained_within" : "[]",
    "user_profile_background_image_url_https" : "https://abs.twimg.com/images/themes/theme14/bg.gif",
    "user_listed_count" : "2",
    "user_notifications" : "None",
    "user_geo_enabled" : "True",
    "user_profile_use_background_image" : "True",
    "text" : "@FunStarsGoLive @haven @DonifordOwners Hahahahahaha!!!! Didn't even know #Stanboardman did a #WorldCupSong! You proper made me chuckle!",
    "place_place_type" : "admin",
    "user_profile_text_color" : "333333",
    "user_time_zone" : "Amsterdam",
    "user_profile_link_color" : "009999",
    "user_profile_background_tile" : "True",
    "place_attributes" : "{}",
    "geo" : "{'type': 'Point', 'coordinates': [50.61453593, -3.36751463]}",
    "user_profile_sidebar_border_color" : "EEEEEE",
    "user_is_translator" : "False",
    "user_follow_request_sent" : "None",
    "user_is_translation_enabled" : "False",
```

```
    "user_profile_background_color" : "131516",
    "user_utc_offset" : "7200",
    "user_statuses_count" : "5180",
    "source" : "<a href=\"http://twitter.com/download/iphone\" rel=\"
    nofollow\">Twitter for iPhone</a>",
    "retweeted" : "False",
    "user_id" : "129638809",
    "user_profile_image_url" : "http://pbs.twimg.com/profile_images
    /378800000818236216/8a89de53bccefc0b948e4e450badcdb0_normal.jpeg",
    "coordinates" : "{'type': 'Point', 'coordinates': [-3.36751463,
    50.61453593]}",
    "user_description" : "London born Luton raised Devon based comedian
    coming to a town near you!",
    "user_friends_count" : "389",
    "user_protected" : "False",
    "user_contributors_enabled" : "False",
    "user_favourites_count" : "357",
    "place_name" : "South West",
    "user_profile_banner_url" : "https://pbs.twimg.com/profile_banners
    /129638809/1395276984",
    "item_id" : "455725112806109184",
    "user_default_profile" : "False",
    "user_name" : "Noel Brodie",
    "user_lang" : "en",
    "user_verified" : "False",
    "place_bounding_box" : "{'type': 'Polygon', 'coordinates':
    [[[-6.36850399906372, 49.8824720005481], [-6.36850399906372,
    52.1125420344225], [-1.48573420014269, 52.1125420344225],
    [-1.48573420014269, 49.8824720005481]]]}",
    "user_url" : "http://noelbrodie.co.uk",
    "user_profile_background_image_url" : "http://abs.twimg.com/images/
    themes/theme14/bg.gif",
    "date" : "2014-04-14 16:11:53 ",
    "user_location" : "UK",
    "user_profile_image_url_https" : "https://pbs.twimg.com/profile_images
    /378800000818236216/8a89de53bccefc0b948e4e450badcdb0_normal.jpeg",
    "place_url" : "https://api.twitter.com/1.1/geo/id/25d3e991f5637f5a.json
    ",
    "place_id" : "25d3e991f5637f5a",
    "user_followers_count" : "1594",
    "user_default_profile_image" : "False"
}
```


LISTING D.1: Twitter Tweet

Appendix E

Appendix Sample Predictive Model Dataset

	STT	S_MAX	W_DLR	W_DLT	W_ILR	W_ILT	STT1	STT2	STT3	STT5
2013-05-01	5	38.973611	0.000000	10.831597	0.000000	11.497569	NaN	NaN	NaN	NaN
2013-05-02	5	34.350000	0.000000	11.705903	0.000000	12.140278	5	NaN	NaN	NaN
2013-05-03	5	21.384896	0.205787	13.464236	0.077083	13.631481	5	5	NaN	NaN
2013-05-06	5	10.715972	0.652778	13.986343	0.275000	15.386806	5	5	5	NaN
2013-05-07	5	19.422917	0.455556	13.637153	0.002083	13.815278	5	5	5	NaN
2013-05-08	5	22.364583	5.806250	12.172222	4.323611	12.646528	5	5	5	5
2013-05-09	5	22.875000	1.543056	10.089236	3.502778	9.858333	5	5	5	5
2013-05-10	5	21.099769	1.777894	10.548264	3.964120	10.719213	5	5	5	5
2013-05-13	5	19.524923	0.553125	9.101042	3.639583	8.686806	5	5	5	5
2013-05-14	5	20.222917	0.849306	9.489931	0.068750	9.275694	5	5	5	5
2013-05-15	5	21.413889	0.191667	8.951042	0.487500	8.841667	5	5	5	5
2013-05-16	5	22.884722	0.592014	8.094097	1.811111	7.587153	5	5	5	5
2013-05-17	5	21.626878	0.056481	11.177623	0.008333	10.876157	5	5	5	5
2013-05-20	5	18.498611	0.000000	13.567708	0.000000	13.515278	5	5	5	5
2013-05-21	5	20.224788	0.000000	12.850694	0.000000	12.968056	5	5	5	5
2013-05-22	5	20.553472	0.000000	11.768403	0.000000	11.588194	5	5	5	5
2013-05-23	5	23.536111	0.000000	9.120833	1.251389	8.735417	5	5	5	5
2013-05-24	5	20.678125	0.000000	12.107407	0.408333	12.755324	5	5	5	5
2013-05-27	5	19.045833	4.547917	11.776389	1.595139	12.193750	5	5	5	5
2013-05-28	5	20.153935	0.048264	11.343750	0.913194	11.843056	5	5	5	5
2013-05-29	5	21.164120	7.879861	14.161806	1.751389	13.998611	5	5	5	5
2013-05-30	5	22.543056	0.000000	15.786111	0.052083	15.630556	5	5	5	5
2013-05-31	5	19.311574	0.000000	13.487153	0.000000	13.359606	5	5	5	5
2013-06-03	5	8.069444	0.000000	16.055035	0.000000	16.570139	5	5	5	5
2013-06-04	5	18.061806	0.000000	15.369213	0.000000	15.095833	5	5	5	5
2013-06-05	5	19.560417	0.000000	15.418403	0.000000	15.077431	5	5	5	5
2013-06-06	5	21.563194	0.000000	14.996528	0.000000	13.861111	5	5	5	5
2013-06-07	5	19.650746	0.000000	16.196373	0.000000	16.412500	5	5	5	5
2013-06-10	5	19.034028	0.000000	15.457292	0.000000	15.065278	5	5	5	5

	STT	S_MAX	W_DLR	W_DLT	W_ILR	W_ILT	STT1	STT2	STT3	STT5
2013-06-11	5	20.134182	3.006944	14.630903	0.517361	14.635417	5	5	5	5
2013-06-12	5	21.285417	4.300000	17.450000	0.106250	13.736806	5	5	5	5
2013-06-13	5	22.168056	4.300000	17.450000	0.797917	12.989583	5	5	5	5
2013-06-14	5	21.605093	4.300000	17.450000	0.713657	13.682639	5	5	5	5
2013-06-17	5	18.145139	4.300000	17.450000	1.138889	13.500000	5	5	5	5
2013-06-18	5	19.556944	4.300000	17.450000	0.687500	16.736111	5	5	5	5
2013-06-19	5	21.003472	4.300000	17.450000	0.000000	16.008681	5	5	5	5
2013-06-20	5	23.333449	4.300000	17.450000	0.000000	16.413542	5	5	5	5
2013-06-21	5	20.255748	0.603704	-31.971547	1.626852	14.903935	5	5	5	5
2013-06-24	5	6.800000	0.000000	14.156713	1.661111	14.504861	5	5	5	5
2013-06-25	5	20.634028	0.000000	15.788889	0.000000	15.843056	5	5	5	5
2013-06-26	5	20.784028	0.000000	16.929051	0.000000	16.327083	5	5	5	5
2013-06-27	5	24.990278	1.828819	15.039931	0.559028	15.297801	5	5	5	5
2013-06-28	5	21.338426	0.089815	16.366705	0.226273	16.415972	5	5	5	5
2013-07-01	5	17.003819	0.000000	14.552431	0.000000	15.107986	5	5	5	5
2013-07-02	5	19.076620	1.886111	14.994097	1.338194	14.650347	5	5	5	5
2013-07-03	5	20.690972	0.000000	13.800694	0.125000	16.150000	5	5	5	5
2013-07-04	5	26.537365	0.000000	13.800000	0.000000	17.711806	5	5	5	5
2013-07-05	5	18.922222	0.000000	16.455517	0.000000	19.224769	5	5	5	5
2013-07-08	5	18.454167	0.000000	18.242708	0.000000	19.184722	5	5	5	5
2013-07-09	5	20.689815	0.000000	21.413194	0.000000	21.848611	5	5	5	5
2013-07-10	5	21.802238	0.000000	19.523495	0.000000	18.691667	5	5	5	5
2013-07-11	5	22.138194	0.000000	19.486806	0.000000	19.165972	5	5	5	5
2013-07-12	5	21.061188	0.000000	19.657350	0.000000	20.338194	5	5	5	5
2013-07-15	5	22.817361	0.000000	17.922569	0.000000	18.534028	5	5	5	5
2013-07-16	5	22.650926	0.000000	19.148958	0.000000	19.428472	5	5	5	5
2013-07-17	5	24.311806	0.000000	20.868287	0.000000	21.561806	5	5	5	5
2013-07-18	5	22.900000	0.000000	17.300347	0.000000	20.723611	5	5	5	5
2013-07-19	5	19.593236	0.000000	19.054398	0.000000	19.789815	5	5	5	5
2013-07-22	5	19.251389	0.000000	17.962847	0.000000	19.514583	5	5	5	5
2013-07-23	5	21.291722	0.000000	18.065972	0.000000	20.190972	5	5	5	5
2013-07-24	5	23.432639	0.000000	18.834606	0.000000	19.148611	5	5	5	5
2013-07-25	5	21.595139	2.902083	18.318403	1.297222	18.183333	5	5	5	5
2013-07-26	5	21.021296	1.825116	16.488021	0.127778	17.928704	5	5	5	5
2013-07-29	5	19.758160	1.287153	16.463889	0.005556	18.084722	5	5	5	5
2013-07-30	5	23.236111	0.000000	15.502431	0.000000	18.769444	5	5	5	5
2013-07-31	5	22.743519	2.058333	16.517708	0.838194	16.965278	5	5	5	5
2013-08-01	5	26.365355	0.208333	18.445833	0.434722	19.381944	5	5	5	5
2013-08-02	5	32.023090	1.125000	18.041705	0.357407	17.890278	5	5	5	5
2014-01-06	5	24.361111	1.116667	8.933333	0.493750	10.007639	5	5	5	5
2014-01-07	5	24.945139	0.000000	8.012500	0.297917	9.430556	5	5	5	5
2014-01-08	5	26.715278	1.025000	6.808333	72.299306	7.202778	5	5	5	5
2014-01-09	5	23.507330	1.245833	4.166667	0.291667	5.576389	5	5	5	5
2014-01-10	5	23.393801	0.594444	5.391667	29.748765	6.408758	5	5	5	5
2014-01-13	5	24.531944	0.500000	3.737500	0.502083	7.751736	5	5	5	5
2014-01-14	5	26.965278	1.520833	3.975000	0.302778	5.407639	5	5	5	5
2014-01-15	5	24.893056	0.366667	8.279167	0.156250	9.059722	5	5	5	5

	STT	S_MAX	W_DLR	W_DLT	W_ILR	W_ILT	STT1	STT2	STT3	STT5
2014-01-16	5	28.747222	0.000000	5.820833	0.097917	6.369792	5	5	5	5
2014-01-17	5	25.299067	1.336111	4.956944	0.404167	5.286574	5	5	5	5
2014-01-20	5	23.360725	0.000000	5.500000	0.000000	5.113194	5	5	5	5
2014-01-21	5	26.540432	0.000000	5.500000	0.000000	7.983333	5	5	5	5
2014-01-22	5	26.245293	0.166667	5.320833	0.000000	7.435417	5	5	5	5
2014-01-23	5	29.443056	0.700000	5.116667	0.229167	6.156944	5	5	5	5
2014-01-24	5	27.554051	3.188889	6.987500	1.287037	7.662731	5	5	5	5
2014-01-27	5	25.691601	0.766667	5.575000	0.321528	6.399306	5	5	5	5
2014-01-28	5	26.396991	5.458333	5.837500	0.706944	6.018750	5	5	5	5
2014-01-29	5	30.307639	4.900000	5.512500	1.420139	5.683333	5	5	5	5
2014-01-30	5	31.197396	0.000000	5.125000	0.000000	5.045139	5	5	5	5
2014-01-31	5	31.717785	1.363194	5.880324	0.696528	6.348611	5	5	5	5
2014-02-03	5	27.411420	6.133333	7.062500	1.570139	7.032639	5	5	5	5
2014-02-04	5	6.000000	1.962500	4.533333	0.412500	5.079167	5	5	5	5
2014-02-05	5	6.000000	4.466667	7.479167	1.540972	7.607639	5	5	5	5
2014-02-06	5	6.000000	0.000000	6.033333	0.000000	6.705556	5	5	5	5
2014-02-07	5	6.000000	0.659722	5.669444	0.272917	6.162616	5	5	5	5
2014-02-10	5	6.000000	0.000000	3.087500	0.000000	3.778472	5	5	5	5
2014-02-11	5	12.629861	2.054167	3.145833	0.831944	3.594444	5	5	5	5
2014-02-12	5	38.114931	5.916667	4.679167	5.809722	5.070833	5	5	5	5
2014-02-13	5	41.287153	0.000000	3.662500	0.125000	4.534722	5	5	5	5
2014-02-14	5	29.001260	0.000000	5.106944	3.793750	5.863657	5	5	5	5
2014-02-17	5	20.136497	0.000000	8.083333	1.781944	8.259722	5	5	5	5
2014-02-18	5	25.347762	0.375000	7.529167	0.097917	8.227778	5	5	5	5
2014-02-19	5	27.758140	0.000000	8.716667	0.000000	9.227083	5	5	5	5
2014-02-20	5	29.678318	0.654167	7.533333	0.000000	10.600000	5	5	5	5
2014-02-21	5	43.407465	0.188889	7.831944	0.000000	8.722106	5	5	5	5
2014-02-24	5	43.259576	3.612500	7.183333	0.000000	0.000000	5	5	5	5
2014-02-25	5	26.971238	0.333333	6.916667	0.000000	0.000000	5	5	5	5
2014-02-26	5	24.465278	0.000000	7.383333	0.000000	0.000000	5	5	5	5
2014-02-27	5	27.145602	0.000000	6.325000	0.000000	0.000000	5	5	5	5
2014-02-28	5	31.970660	0.693056	5.997222	0.310417	4.868056	5	5	5	5
2014-03-03	5	16.786535	1.500000	5.054167	1.025000	5.638889	5	5	5	5
2014-03-04	5	14.888889	0.000000	6.720833	0.000000	7.588889	5	5	5	5
2014-03-05	5	22.270525	0.000000	7.975000	0.000000	8.591667	5	5	5	5
2014-03-06	5	31.330015	0.354167	11.533333	0.016667	12.283333	5	5	5	5
2014-03-07	5	28.842065	1.512500	10.170833	0.821296	10.305093	5	5	5	5
2014-03-10	5	25.277006	0.000000	6.491667	0.000000	6.859722	5	5	5	5
2014-03-11	5	26.520139	0.000000	5.950000	0.187500	5.431944	5	5	5	5
2014-03-12	5	39.160298	0.000000	5.412500	0.000000	5.257639	5	5	5	5
2014-03-13	5	47.664120	0.000000	6.287500	0.000000	5.866667	5	5	5	5
2014-03-14	5	24.991037	0.000000	9.512500	0.000000	9.732253	5	5	5	5
2014-03-17	5	14.202778	0.000000	9.366667	0.000000	9.692361	5	5	5	5
2014-03-18	5	23.365278	0.000000	10.112500	0.000000	10.688889	5	5	5	5
2014-03-19	5	5.777778	0.000000	11.291667	0.000000	12.380208	5	5	5	5
2014-03-20	5	5.777778	0.000000	9.183333	2.925000	8.881250	5	5	5	5
2014-03-21	5	5.777778	0.672222	5.726157	0.555556	6.009954	5	5	5	5

	STT	S_MAX	W_DLR	W_DLT	W_ILR	W_ILT	STT1	STT2	STT3	STT5
2014-03-24	5	5.777778	0.899306	6.179861	1.213194	5.893750	5	5	5	5
2014-03-25	5	5.777778	0.866667	7.458333	0.622917	7.868750	5	5	5	5
2014-03-26	5	5.777778	0.000000	6.850000	0.131250	6.631250	5	5	5	5
2014-03-27	5	5.777778	3.291667	5.733333	1.529167	6.160417	5	5	5	5
2014-03-28	5	5.777778	2.529167	8.847222	1.575926	8.799306	5	5	5	5
2014-03-31	5	5.777778	0.145139	10.481944	0.167361	10.445833	5	5	5	5
2014-04-01	5	5.777778	0.772222	9.354167	0.200000	9.036806	5	5	5	5
2014-04-02	5	24.635571	6.550000	9.512500	2.935417	9.733333	5	5	5	5
2014-04-03	5	28.111265	1.800000	10.475000	0.471528	11.245833	5	5	5	5
2014-04-04	5	27.945454	0.495833	12.004167	0.137500	12.987269	5	5	5	5
2014-04-07	5	27.530710	2.125000	9.683333	1.331944	10.074306	5	5	5	5
2014-04-08	5	30.761111	0.000000	8.712500	0.000000	9.214583	5	5	5	5
2014-04-09	5	32.834819	0.000000	11.312500	0.000000	11.703472	5	5	5	5
2014-04-10	5	33.797454	0.000000	11.391667	0.000000	13.192361	5	5	5	5
2014-04-11	5	26.751260	0.000000	9.944444	0.038889	15.141088	5	5	5	5

Appendix F

Appendix Prediction Algorithm Results

P = Parameter from table
MSE = Mean Squared Error
R2 = R2
EVS = Explained Varience Score
STD = Standard Deviation
Q = Quantile @ 80

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'14/10/2'	BayesianRidge	5	6.54	0.58	1.88	0.59	4.6	56.6
'14/1/2'	LinearRegression	0	455.14	0.31	13.76	0.32	26.7	101.7
'14/1/1'	LinearRegression	0	304.8	0.45	11.88	0.47	22.3	91.3
'14/2/1'	BayesianRidge	5	44.95	0.19	5.21	0.2	8.2	163.6
'14/2/2'	LinearRegression	0	187.86	0.1	8.12	0.18	11.1	166.7
'14/8/2'	BayesianRidge	5	210.12	0.05	8.3	0.08	13.2	78.5
'14/8/1'	BayesianRidge	5	24.68	0.74	3.11	0.75	11.0	63.6
'14/9/2'	LinearRegression	1	2.73	0.76	1.28	0.76	3.5	51.2
'14/9/1'	BayesianRidge	5	23.26	0.22	3.45	0.22	5.8	58.8
'14/3/1'	BayesianRidge	5	622.89	0.3	18.44	0.3	30.6	118.3
'14/3/2'	LinearRegression	1	212.93	0.38	10.35	0.39	18.4	80.5
'14/4/2'	BayesianRidge	5	292.43	0.32	11.48	0.34	18.9	172.6
'14/4/1'	BayesianRidge	5	120.26	0.24	6.54	0.28	9.7	149.2
'14/5/2'	BayesianRidge	5	190.26	0.55	9.35	0.57	18.8	68.9
'14/5/1'	LinearRegression	1	94.83	0.43	6.57	0.43	14.3	65.0

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'14/6/1'	BayesianRidge	5	197.02	0.51	9.07	0.51	22.3	83.3
'14/6/2'	LinearRegression	0	354.12	0.43	10.75	0.44	23.7	78.2
'14/7/1'	BayesianRidge	5	295.9	0.59	10.79	0.59	27.7	113.5
'14/7/2'	LinearRegression	1	196.47	0.49	8.53	0.52	20.6	89.7
'13/3/2'	BayesianRidge	5	5.75	0.98	1.77	0.98	16.5	93.7
'13/1/1'	BayesianRidge	5	376.29	0.87	9.87	0.87	54.8	131.0
'13/1/2'	BayesianRidge	5	369.39	0.88	9.71	0.88	55.9	133.6
'13/6/2'	LinearRegression	0	1527.8	0.6	18.85	0.64	54.8	137.8
'13/6/1'	BayesianRidge	5	32.01	0.98	2.23	0.98	40.7	123.0
'13/5/1'	BayesianRidge	5	31.89	0.98	3.61	0.98	44.7	113.3
'13/5/2'	LinearRegression	0	41.33	0.98	3.98	0.98	45.7	114.7
'13/4/1'	LinearRegression	1	2621.59	0.17	30.44	0.26	43.7	130.9
'13/4/2'	LinearRegression	1	9.0	0.94	2.08	0.94	12.7	72.5
'13/3/1'	BayesianRidge	5	4.26	0.97	1.35	0.97	13.3	85.8
'13/2/2'	BayesianRidge	5	90.01	0.98	5.87	0.98	62.4	149.8
'13/2/1'	BayesianRidge	5	92.02	0.98	6.42	0.98	63.7	152.0
'12/2/2'	BayesianRidge	5	3.5	0.93	0.99	0.93	7.6	72.7
'12/3/2'	SVR	3	87.16	0.03	3.2	0.04	6.2	54.8
'12/1/1'	BayesianRidge	5	10.28	0.8	2.04	0.81	7.9	25.2
'12/1/2'	BayesianRidge	5	0.01	0.98	0.05	0.98	0.9	9.0
'12/2/1'	LinearRegression	0	2.36	0.96	1.14	0.96	8.1	74.1
'11/4/1'	BayesianRidge	5	107.76	0.18	6.37	0.18	10.9	43.3
'11/1/2'	BayesianRidge	5	22.16	0.02	3.38	0.03	4.4	67.5
'11/1/1'	SVR	3	0.03	0.03	0.13	0.04	0.3	59.0
'11/3/1'	SVR	3	72.56	-0.01	6.68	-0.01	8.1	74.3
'11/2/1'	LinearRegression	0	60.41	0.1	5.25	0.1	7.3	32.0
'11/5/2'	BayesianRidge	5	51.07	0.22	3.63	0.24	7.2	87.6
'10/4/1'	BayesianRidge	5	0.65	-0.09	0.54	-0.09	0.8	81.0
'10/5/2'	LinearRegression	1	0.03	-0.55	0.15	-0.49	0.2	15.0
'10/5/1'	SVR	3	108.42	0.01	7.33	0.01	9.6	39.6
'10/6/1'	SVR	3	13.79	0.0	2.74	0.0	3.6	61.5
'10/6/2'	SVR	3	3.48	-0.12	1.25	-0.1	1.6	59.0
'10/2/1'	BayesianRidge	5	22.27	0.04	3.62	0.05	5.1	78.4
'10/2/2'	LinearRegression	1	27.29	0.15	4.23	0.22	5.5	79.8
'10/1/2'	LinearRegression	0	38.25	0.1	4.8	0.1	6.3	46.8
'10/1/1'	SVR	3	0.86	-0.14	0.66	-0.12	1.0	38.9
'10/3/2'	LinearRegression	0	41.15	0.19	4.55	0.2	6.6	24.3
'10/7/2'	BayesianRidge	5	206.2	0.25	11.29	0.26	16.4	143.4

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'1/5/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	16.0
'1/8/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	9.0
'1/8/1'	BayesianRidge	5	142.04	0.16	8.35	0.17	13.2	51.6
'1/6/1'	SVR	3	405.19	0.0	14.38	0.01	19.6	90.9
'1/6/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	17.0
'1/7/1'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	16.0
'1/7/2'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	16.0
'1/9/2'	SVR	3	78.29	-0.02	5.88	-0.02	9.2	38.8
'1/9/1'	BayesianRidge	5	56.17	0.16	4.76	0.18	8.4	33.9
'15/10/1'	BayesianRidge	5	166.44	0.55	8.85	0.55	19.1	88.2
'15/1/1'	LinearRegression	1	276.21	0.49	11.03	0.49	23.7	93.4
'15/3/1'	BayesianRidge	5	70.24	0.61	6.36	0.61	13.8	67.1
'15/2/1'	BayesianRidge	5	200.9	0.35	10.72	0.35	18.7	136.9
'15/5/1'	BayesianRidge	5	406.31	0.51	14.51	0.51	30.7	125.6
'15/4/2'	BayesianRidge	5	336.41	0.33	14.54	0.33	23.3	102.9
'15/4/1'	BayesianRidge	5	227.17	0.37	11.54	0.37	19.8	91.1
'15/7/1'	BayesianRidge	5	380.47	0.06	15.34	0.06	21.8	136.2
'15/7/2'	BayesianRidge	5	4576.51	0.05	36.59	0.08	53.3	204.2
'15/6/2'	BayesianRidge	5	1386.52	0.15	25.3	0.16	43.2	162.5
'15/9/2'	LinearRegression	1	76.72	0.45	6.0	0.47	13.4	65.7
'15/9/1'	LinearRegression	1	828.94	0.19	18.52	0.22	31.0	120.5
'15/6/1'	LinearRegression	0	1666.73	0.18	24.19	0.2	46.5	175.4
'15/8/2'	BayesianRidge	5	244.16	0.03	11.84	0.06	18.2	88.2
'15/8/1'	SVR	3	634.09	0.0	18.63	0.0	26.8	112.7
'15/11/2'	LinearRegression	0	799.72	0.51	19.11	0.57	41.2	153.0
'15/11/1'	BayesianRidge	5	328.04	0.17	13.72	0.2	20.7	82.8
'15/12/1'	BayesianRidge	5	237.88	0.19	9.75	0.2	18.4	85.5
'15/12/2'	LinearRegression	0	512.45	0.7	17.26	0.71	40.5	140.3
'14/10/1'	BayesianRidge	5	130.33	0.06	7.71	0.07	11.5	71.6
'11/5/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	72.0
'11/4/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	7.0
'11/3/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	54.0
'11/2/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	14.0
'10/4/2'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	79.0
'10/7/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	108.0
'10/3/1'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	4.0
'1/5/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	31.0
'9/7/2'	BayesianRidge	5	6.72	0.68	1.06	0.69	3.7	18.0

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'7/4/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	36.0
'7/5/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	4.0
'7/6/2'	LinearRegression	0	0.0	0.0	0.01	0.0	0.0	58.0
'7/12/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	22.0
'6/7/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	22.0
'6/6/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	7.0
'6/5/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	29.0
'5/2/1'	LinearRegression	0	0.0	1.0	0.0	1.0	0.0	45.0
'46/5/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	25.0
'46/4/2'	LinearRegression	0	16.7	0.07	1.87	0.11	2.9	15.9
'45/1/2'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	32.0
'40/4/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	15.0
'40/3/1'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	30.0
'39/2/1'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	30.0
'36/4/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	55.0
'35/1/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	21.0
'35/9/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	13.0
'35/15/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	14.0
'35/3/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	26.0
'34/8/2'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	22.0
'34/1/2'	SVR	3	0.39	0.42	0.4	0.44	0.9	15.6
'34/6/2'	SVR	3	0.31	0.4	0.37	0.41	0.8	11.4
'34/7/2'	SVR	3	0.8	0.4	0.56	0.41	1.3	18.2
'34/4/2'	SVR	3	0.64	0.43	0.51	0.45	1.2	19.1
'34/2/2'	SVR	3	0.54	0.4	0.46	0.4	1.0	15.8
'31/4/1'	LinearRegression	1	1178.99	0.2	26.17	0.23	35.0	127.3
'31/4/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	28.0
'31/1/1'	LinearRegression	0	503.45	0.04	16.59	0.1	20.1	73.0
'31/1/2'	SVR	3	87.12	-0.04	6.06	-0.02	8.5	34.7
'31/7/2'	BayesianRidge	5	5.88	0.91	1.51	0.92	8.6	51.9
'31/7/1'	BayesianRidge	5	37.85	0.53	4.02	0.54	9.1	47.0
'31/2/2'	BayesianRidge	5	19.3	0.2	3.18	0.2	5.2	112.1
'31/3/1'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	16.0
'30/11/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	11.0
'30/11/1'	SVR	3	59.17	-0.06	4.57	0.0	5.7	22.4
'30/10/2'	BayesianRidge	5	2.19	-0.47	0.91	-0.46	1.2	21.6
'30/10/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	19.0
'30/13/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	16.0

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'30/13/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	16.0
'30/19/2'	LinearRegression	1	0.39	0.14	0.4	0.19	0.6	48.3
'30/19/1'	BayesianRidge	5	71.54	0.09	5.9	0.1	8.9	63.9
'30/18/2'	SVR	3	0.15	-0.07	0.27	-0.02	0.3	75.8
'30/18/1'	LinearRegression	1	3.31	-0.01	1.3	0.09	1.6	78.1
'30/9/1'	BayesianRidge	5	6.88	-0.05	0.96	0.01	1.6	63.6
'30/9/2'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	62.0
'30/8/1'	LinearRegression	0	0.0	1.0	0.0	1.0	0.0	24.0
'30/8/2'	SVR	3	25.32	-0.29	2.94	-0.13	5.3	30.0
'30/2/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	24.0
'30/2/1'	SVR	3	2.07	-0.41	0.97	-0.35	1.7	27.5
'30/3/2'	BayesianRidge	5	0.69	-0.23	0.57	-0.19	0.7	47.8
'30/1/2'	SVR	3	0.14	-0.02	0.24	-0.01	0.9	82.4
'30/6/1'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	16.0
'30/7/1'	LinearRegression	1	1.45	-0.07	0.78	-0.04	1.1	27.7
'30/5/1'	SVR	3	0.34	-0.36	0.41	-0.33	0.7	43.5
'30/15/2'	SVR	3	40.9	-0.12	5.2	-0.05	7.1	51.9
'30/17/1'	BayesianRidge	5	0.78	-0.13	0.64	-0.05	0.8	48.6
'30/16/1'	BayesianRidge	5	0.36	-0.41	0.46	-0.32	0.5	32.7
'30/12/2'	BayesianRidge	5	27.31	-0.08	4.44	-0.08	5.5	33.3
'30/24/2'	LinearRegression	1	10.18	0.03	1.98	0.07	2.9	20.3
'30/24/1'	SVR	3	1.67	0.35	0.76	0.37	1.6	17.1
'30/25/2'	LinearRegression	0	0.05	0.18	0.15	0.32	0.2	26.5
'30/25/1'	LinearRegression	0	84.21	0.07	3.46	0.1	6.2	32.5
'30/17/2'	BayesianRidge	5	2.25	-0.14	0.79	-0.05	1.4	48.8
'30/20/2'	LinearRegression	1	7.21	0.16	1.88	0.22	2.8	19.1
'30/20/1'	BayesianRidge	5	34.79	0.06	4.26	0.09	6.0	27.3
'30/21/2'	LinearRegression	0	1.12	-0.26	0.77	-0.12	0.9	30.3
'30/21/1'	LinearRegression	1	3.56	0.05	1.39	0.05	1.9	32.3
'30/22/1'	BayesianRidge	5	16.58	0.11	2.75	0.13	3.9	51.8
'30/22/2'	SVR	3	0.81	-0.18	0.53	-0.13	0.8	47.7
'30/23/1'	LinearRegression	1	24.07	0.11	2.67	0.13	4.1	29.6
'30/23/2'	BayesianRidge	5	0.12	-0.07	0.26	-0.07	0.5	29.5
'30/3/1'	SVR	3	0.07	0.62	0.19	0.65	0.7	46.8
'3/1/1'	SVR	3	7.2	-0.15	2.07	-0.07	3.0	23.9
'3/1/2'	LinearRegression	1	42.47	0.56	3.1	0.57	8.2	18.8
'3/2/2'	SVR	3	1.62	0.12	0.71	0.13	2.3	31.5
'3/3/1'	SVR	3	7.11	0.03	1.94	0.05	2.5	69.3

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'29/2/1'	BayesianRidge	5	7645.9	0.07	54.99	0.12	78.1	238.5
'29/2/2'	SVR	3	14.79	-0.09	2.87	-0.04	4.7	66.2
'28/8/1'	LinearRegression	0	273.19	0.05	11.76	0.06	17.5	80.9
'28/5/1'	SVR	3	24.32	-0.04	3.44	-0.02	5.5	50.1
'28/4/1'	BayesianRidge	5	900.03	0.01	22.27	0.01	28.2	114.5
'28/7/1'	SVR	3	310.88	-0.01	11.87	0.0	18.2	81.4
'28/7/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	16.0
'28/6/1'	SVR	3	45.99	-0.02	4.41	-0.02	6.8	49.1
'28/1/1'	LinearRegression	1	108.45	0.15	7.36	0.17	11.4	86.3
'28/3/1'	SVR	3	472.37	-0.01	15.87	-0.01	22.5	96.9
'28/2/1'	BayesianRidge	5	1446.23	0.15	22.64	0.19	31.5	109.5
'27/3/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	14.0
'27/10/1'	LinearRegression	1	1409.0	0.15	22.93	0.21	30.2	103.5
'27/5/1'	BayesianRidge	5	592.58	0.53	16.62	0.55	35.8	127.1
'27/2/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	10.0
'27/2/1'	LinearRegression	0	398.92	0.54	13.55	0.56	28.8	109.9
'27/1/1'	LinearRegression	0	5983.9	-0.05	34.54	0.02	46.9	102.0
'27/8/1'	BayesianRidge	5	901.96	0.42	19.79	0.45	36.9	125.5
'27/8/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	15.0
'27/9/1'	BayesianRidge	5	471.62	0.39	15.46	0.39	27.1	109.2
'27/9/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	11.0
'27/6/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	7.0
'27/6/1'	LinearRegression	0	94.47	0.78	6.19	0.79	21.8	94.7
'27/7/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	13.0
'27/7/1'	LinearRegression	0	748.3	0.27	18.51	0.31	27.8	108.0
'27/4/1'	LinearRegression	0	1959.98	0.34	29.5	0.39	50.6	152.3
'27/3/1'	LinearRegression	1	675.35	0.57	19.42	0.59	36.2	151.9
'27/4/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	20.0
'27/5/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	12.0
'26/3/1'	LinearRegression	0	83.01	0.61	7.4	0.62	13.1	85.0
'26/2/1'	LinearRegression	0	106.44	0.2	6.92	0.25	10.7	53.9
'26/1/1'	SVR	3	25.09	-0.03	3.83	-0.03	5.8	63.2
'25/1/1'	BayesianRidge	5	105.46	0.2	6.72	0.23	11.7	65.3
'25/2/1'	LinearRegression	0	703.53	0.47	16.66	0.48	35.9	103.0
'25/3/1'	SVR	3	117.52	-0.1	7.65	0.01	12.7	75.1
'25/4/1'	BayesianRidge	5	236.02	-0.03	10.2	0.0	12.5	73.4
'22/7/2'	LinearRegression	1	3.06	0.24	1.24	0.28	2.0	34.3
'22/7/1'	BayesianRidge	5	0.37	0.51	0.45	0.51	1.0	32.5

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'22/6/2'	LinearRegression	0	0.07	0.71	0.16	0.71	0.5	33.0
'22/6/1'	BayesianRidge	5	3.67	0.54	1.44	0.54	3.1	39.4
'22/5/1'	LinearRegression	1	0.09	-0.16	0.21	-0.15	0.3	98.9
'22/5/2'	LinearRegression	1	2847.3	0.04	14.87	0.07	31.7	109.4
'22/4/1'	BayesianRidge	5	7.79	-0.09	2.36	-0.07	3.3	74.4
'22/4/2'	SVR	3	194.18	0.0	6.38	0.01	8.8	80.8
'22/3/2'	BayesianRidge	5	5.81	0.25	1.84	0.26	3.4	101.2
'22/3/1'	BayesianRidge	5	72.93	0.27	6.22	0.27	10.5	122.3
'22/14/1'	BayesianRidge	5	4.71	0.48	1.26	0.48	3.2	48.0
'22/2/2'	LinearRegression	0	0.17	0.89	0.25	0.9	1.3	34.1
'22/2/1'	BayesianRidge	5	1.52	0.78	0.91	0.79	2.8	37.5
'22/9/1'	BayesianRidge	5	4.47	0.81	1.42	0.81	5.1	26.1
'22/9/2'	BayesianRidge	5	19.11	0.64	2.97	0.64	7.8	34.5
'22/8/1'	BayesianRidge	5	1.2	0.69	0.77	0.69	2.1	61.7
'22/8/2'	LinearRegression	0	0.02	0.43	0.09	0.43	0.2	57.3
'22/14/2'	LinearRegression	1	0.07	0.25	0.18	0.26	0.3	41.9
'22/12/2'	LinearRegression	1	0.08	0.7	0.16	0.7	0.5	16.1
'22/12/1'	BayesianRidge	5	1.81	0.64	0.89	0.65	2.6	21.7
'22/11/2'	BayesianRidge	5	1.04	0.71	0.78	0.72	2.0	25.3
'22/13/2'	LinearRegression	1	3.93	0.47	1.4	0.48	2.9	17.4
'22/13/1'	LinearRegression	0	1.74	0.75	0.95	0.77	3.1	17.5
'22/10/1'	LinearRegression	0	0.11	0.13	0.3	0.14	0.3	51.7
'22/10/2'	BayesianRidge	5	2.56	0.06	0.63	0.07	1.1	53.2
'22/11/1'	BayesianRidge	5	0.81	0.26	0.53	0.26	1.0	23.4
'21/3/1'	BayesianRidge	5	40.28	0.49	4.76	0.49	9.8	43.3
'21/3/2'	BayesianRidge	5	366.26	0.74	11.36	0.76	39.9	264.9
'21/2/1'	BayesianRidge	5	20.35	-0.08	3.8	-0.04	5.9	39.0
'21/2/2'	LinearRegression	1	606.24	0.86	11.5	0.86	68.9	441.5
'21/4/2'	BayesianRidge	5	4631.37	0.39	40.12	0.4	86.6	446.1
'21/4/1'	BayesianRidge	5	2.24	0.63	0.99	0.63	2.6	38.8
'21/5/2'	BayesianRidge	5	491.59	0.78	10.45	0.79	49.6	308.2
'21/5/1'	BayesianRidge	5	5.1	0.45	1.65	0.45	2.8	26.0
'20/4/1'	LinearRegression	0	6.54	0.28	1.66	0.31	3.1	35.8
'20/5/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	11.0
'20/5/1'	BayesianRidge	5	7.93	0.7	1.79	0.7	4.6	17.9
'20/4/2'	LinearRegression	0	0.0	1.0	0.0	1.0	0.0	33.0
'20/6/1'	BayesianRidge	5	7.45	0.76	1.86	0.76	5.1	23.7
'20/6/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	15.0

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'20/1/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	13.0
'20/1/1'	BayesianRidge	5	26.44	0.57	4.06	0.57	8.5	36.9
'20/3/1'	LinearRegression	0	0.02	0.82	0.1	0.82	0.4	22.8
'20/3/2'	LinearRegression	0	0.0	1.0	0.0	1.0	0.0	22.0
'20/2/1'	BayesianRidge	5	7.23	0.5	1.93	0.51	4.5	30.3
'20/2/2'	LinearRegression	0	0.0	1.0	0.0	1.0	0.0	22.0
'19/1/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	10.0
'19/4/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	16.0
'19/1/1'	BayesianRidge	5	261.0	0.44	11.42	0.5	22.2	91.8
'19/4/1'	BayesianRidge	5	1098.36	0.15	20.87	0.29	33.7	118.7
'19/3/1'	BayesianRidge	5	224.91	0.43	9.78	0.46	20.7	79.8
'19/3/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	10.0
'19/2/1'	BayesianRidge	5	196.21	0.74	9.86	0.75	30.6	111.9
'19/2/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	16.0
'18/4/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	31.0
'18/4/1'	BayesianRidge	5	1299.01	0.07	15.98	0.09	28.2	95.9
'18/5/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	39.0
'18/5/1'	BayesianRidge	5	356.86	0.3	10.33	0.31	19.1	86.7
'18/6/1'	LinearRegression	0	141.53	0.49	8.1	0.49	16.8	64.1
'18/6/2'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	14.0
'18/7/1'	LinearRegression	0	250.93	0.58	11.16	0.58	26.1	90.4
'18/7/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	15.0
'18/1/2'	LinearRegression	0	0.0	1.0	0.0	1.0	0.0	32.0
'18/1/1'	BayesianRidge	5	1785.77	0.2	27.21	0.2	42.0	187.5
'18/8/2'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	52.0
'18/8/1'	LinearRegression	0	9.73	0.88	1.81	0.89	9.8	73.6
'18/2/1'	BayesianRidge	5	3063.09	0.03	37.45	0.04	51.4	253.3
'18/2/2'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	72.0
'18/3/1'	LinearRegression	1	168.46	0.42	7.85	0.44	14.1	60.1
'18/3/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	36.0
'17/4/1'	LinearRegression	0	247.11	0.56	10.98	0.56	20.2	47.8
'17/4/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	27.0
'17/3/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	39.0
'17/3/1'	BayesianRidge	5	74.58	0.47	6.64	0.47	11.9	69.5
'17/2/2'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	16.0
'17/2/1'	LinearRegression	0	594.17	0.37	15.78	0.39	33.3	136.2
'17/1/1'	LinearRegression	0	605.09	0.14	15.61	0.18	28.9	116.3
'17/1/2'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	12.0

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'17/7/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	34.0
'17/7/1'	SVR	3	33.41	-0.12	4.36	-0.06	7.4	45.0
'17/6/2'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	73.0
'17/6/1'	LinearRegression	0	4824.8	0.25	44.85	0.35	73.5	275.6
'17/5/1'	LinearRegression	1	1819.68	0.55	27.66	0.55	59.5	192.5
'17/5/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	30.0
'16/1/1'	BayesianRidge	5	2230.55	0.47	36.55	0.58	65.7	249.7
'16/6/2'	BayesianRidge	5	185.43	0.67	9.57	0.67	22.9	76.4
'16/7/2'	BayesianRidge	5	259.05	0.16	11.04	0.22	19.2	86.1
'16/4/1'	BayesianRidge	5	44.48	0.79	5.09	0.79	14.8	59.7
'16/4/2'	LinearRegression	1	0.05	0.4	0.16	0.41	0.3	23.6
'16/5/2'	LinearRegression	0	355.96	0.37	13.6	0.4	23.3	112.0
'16/7/1'	BayesianRidge	5	268.88	0.26	11.0	0.29	21.2	102.7
'16/2/2'	BayesianRidge	5	8579.06	0.18	67.99	0.29	93.2	322.9
'16/6/1'	SVR	3	7.09	0.19	1.87	0.24	3.1	20.0
'16/1/2'	BayesianRidge	5	358.7	0.93	12.23	0.93	72.3	224.8
'16/5/1'	BayesianRidge	5	317.79	0.13	12.58	0.2	17.0	140.7
'16/2/1'	LinearRegression	1	1377.39	0.45	26.4	0.47	50.5	168.1
'16/3/2'	BayesianRidge	5	311.62	0.16	11.94	0.26	19.2	77.7
'16/3/1'	LinearRegression	0	0.13	0.06	0.22	0.08	0.6	9.8
'15/10/2'	LinearRegression	0	178.19	0.78	7.05	0.79	26.6	95.4
'15/1/2'	LinearRegression	1	855.65	0.32	18.49	0.36	33.1	104.7
'15/3/2'	LinearRegression	1	3050.87	0.12	35.32	0.15	56.4	193.5
'15/2/2'	LinearRegression	1	160.93	0.55	10.38	0.58	19.2	142.5
'15/5/2'	BayesianRidge	5	37.89	0.52	4.64	0.53	10.2	68.6
'12/3/1'	LogisticRegression	6	0.0	1.0	0.0	1.0	3.3	52.0
'1/12/1'	LinearRegression	0	6.63	0.18	0.93	0.21	1.9	11.9
'1/12/2'	BayesianRidge	5	10.68	0.66	2.12	0.67	5.9	21.5
'1/11/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	18.0
'1/4/2'	BayesianRidge	5	86.93	0.55	5.77	0.58	14.8	47.0
'1/4/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	32.0
'1/11/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	18.0
'1/10/2'	LinearRegression	1	52.17	0.11	4.44	0.14	7.7	26.5
'1/10/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	7.0
'1/13/1'	LinearRegression	0	34.99	0.45	3.63	0.47	8.2	27.2
'1/13/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	5.0
'9/2/1'	BayesianRidge	5	114.7	0.7	4.1	0.7	15.4	40.8
'9/2/2'	BayesianRidge	5	114.48	0.7	4.12	0.7	15.5	40.0

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'9/6/1'	BayesianRidge	5	47.89	0.46	4.31	0.49	8.1	26.1
'9/3/1'	BayesianRidge	5	32.46	0.7	2.29	0.71	8.3	38.0
'9/3/2'	BayesianRidge	5	33.7	0.69	2.45	0.7	8.3	37.9
'9/8/2'	LinearRegression	1	17.73	0.27	2.54	0.32	4.5	20.0
'9/8/1'	BayesianRidge	5	13.39	0.48	2.13	0.5	4.4	15.2
'9/9/2'	LinearRegression	0	50.94	0.31	4.99	0.32	10.4	38.6
'9/9/1'	BayesianRidge	5	14.25	0.73	1.68	0.73	6.8	19.5
'9/6/2'	BayesianRidge	5	36.06	0.7	2.4	0.7	8.7	7.0
'9/11/2'	BayesianRidge	5	155.78	0.04	8.01	0.13	10.5	40.7
'9/11/1'	BayesianRidge	5	58.64	0.29	5.74	0.29	8.1	38.9
'9/10/2'	BayesianRidge	5	13.44	0.69	1.54	0.69	5.2	11.4
'9/10/1'	LinearRegression	0	222.2	0.04	5.93	0.1	9.7	23.9
'9/13/1'	BayesianRidge	5	11.49	0.69	1.44	0.69	4.8	12.9
'9/13/2'	BayesianRidge	5	13.53	0.68	1.55	0.69	5.2	9.5
'9/12/1'	LinearRegression	1	59.71	0.39	4.61	0.39	8.4	24.0
'9/12/2'	LinearRegression	0	535.41	0.59	11.11	0.62	25.9	29.4
'9/14/2'	BayesianRidge	5	99.66	0.0	7.1	0.0	9.6	45.5
'9/14/1'	LinearRegression	0	18.98	0.08	2.23	0.1	3.1	24.6
'9/1/2'	BayesianRidge	5	18.6	0.7	1.68	0.71	6.3	14.0
'9/1/1'	BayesianRidge	5	16.71	0.69	1.99	0.7	5.9	20.1
'9/7/1'	BayesianRidge	5	29.11	0.22	3.42	0.24	5.3	29.4
'9/4/2'	BayesianRidge	5	46.99	0.7	2.65	0.7	9.9	32.0
'9/4/1'	BayesianRidge	5	43.44	0.7	2.87	0.71	9.7	35.8
'9/5/2'	BayesianRidge	5	81.66	0.7	3.5	0.7	13.1	22.0
'9/5/1'	BayesianRidge	5	81.66	0.7	3.5	0.7	13.1	22.0
'8/3/1'	BayesianRidge	5	8.84	0.13	2.39	0.15	3.6	44.5
'8/3/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	36.0
'8/2/1'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	58.0
'8/2/2'	SVR	3	367.62	-0.02	15.69	0.0	20.2	107.3
'8/1/2'	SVR	3	61.29	0.0	6.17	0.01	8.7	54.8
'8/1/1'	SVR	3	287.45	-0.01	12.18	0.0	14.4	63.3
'8/5/2'	BayesianRidge	5	186.69	0.59	9.97	0.63	20.7	59.4
'8/5/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	7.0
'8/4/2'	BayesianRidge	5	177.79	0.06	10.26	0.08	13.4	54.1
'8/4/1'	SVR	3	1.68	0.07	0.83	0.09	1.5	18.8
'7/8/1'	BayesianRidge	5	81.59	0.08	6.71	0.09	8.6	44.0
'7/8/2'	BayesianRidge	5	700.41	0.08	18.6	0.09	27.9	97.1
'7/14/1'	SVR	3	28.35	-0.08	3.56	-0.05	4.6	25.6

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'7/10/1'	BayesianRidge	5	72.29	0.06	6.1	0.07	8.1	40.7
'7/9/1'	SVR	3	5.65	0.04	1.62	0.05	2.4	30.6
'7/9/2'	BayesianRidge	5	118.21	0.0	8.63	0.0	11.1	52.7
'7/1/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	43.0
'7/1/2'	BayesianRidge	5	945.0	0.16	18.74	0.17	34.1	124.3
'7/2/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	14.0
'7/2/1'	BayesianRidge	5	170.86	0.33	9.62	0.34	15.7	59.8
'7/3/2'	LinearRegression	0	1.59	0.39	1.0	0.41	1.7	27.1
'7/3/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	22.0
'7/4/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	36.0
'7/5/1'	BayesianRidge	5	59.5	0.3	4.77	0.3	7.8	33.4
'7/6/1'	BayesianRidge	5	0.94	0.06	0.8	0.07	1.1	61.0
'7/7/2'	BayesianRidge	5	777.53	0.05	13.36	0.1	17.9	52.8
'7/7/1'	BayesianRidge	5	4.98	0.16	1.57	0.16	3.1	36.8
'7/14/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	14.0
'7/13/2'	BayesianRidge	5	175.03	0.33	8.6	0.34	15.6	49.1
'7/13/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	14.0
'7/12/1'	BayesianRidge	5	13.65	0.13	2.69	0.14	4.0	33.0
'7/11/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	18.0
'7/11/2'	SVR	3	40.39	-0.04	4.92	-0.04	7.3	38.3
'7/10/2'	BayesianRidge	5	434.13	0.1	14.9	0.11	21.3	82.5
'6/4/1'	SVR	3	3.49	0.02	0.91	0.06	1.4	74.9
'6/4/2'	LinearRegression	0	909.36	0.21	15.47	0.23	23.8	115.7
'6/3/2'	BayesianRidge	5	51.62	0.03	4.5	0.05	6.2	31.0
'6/3/1'	LinearRegression	1	32.93	0.05	4.21	0.07	5.6	32.3
'6/1/1'	BayesianRidge	5	5.45	-0.01	1.72	0.02	2.1	28.3
'6/1/2'	BayesianRidge	5	1793.76	0.05	28.97	0.11	37.5	122.9
'6/2/2'	BayesianRidge	5	192.87	-0.05	10.45	-0.05	14.3	60.2
'6/2/1'	BayesianRidge	5	334.18	-0.02	5.08	0.0	10.4	29.6
'6/7/2'	LinearRegression	0	792.89	0.25	21.22	0.27	31.2	108.9
'6/6/1'	BayesianRidge	5	53.91	0.31	4.83	0.32	8.5	32.2
'6/5/2'	BayesianRidge	5	63.13	0.2	5.67	0.21	8.3	48.0
'5/4/2'	LinearRegression	1	275.54	0.02	7.59	0.06	10.3	27.4
'5/4/1'	SVR	3	2.64	0.09	1.27	0.09	2.0	19.7
'5/6/1'	SVR	3	0.5	0.03	0.5	0.06	0.7	88.5
'5/6/2'	LinearRegression	0	7.99	0.34	1.46	0.37	3.0	91.0
'5/1/2'	BayesianRidge	5	29.19	-0.02	4.11	-0.02	5.7	103.4
'5/3/1'	BayesianRidge	5	21.39	0.05	3.76	0.08	4.3	25.7

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'5/3/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	14.0
'5/1/1'	SVR	3	2186.01	0.0	33.97	0.0	38.9	178.7
'5/5/2'	LinearRegression	1	1012.65	0.03	9.79	0.06	18.2	28.7
'5/5/1'	BayesianRidge	5	0.53	0.03	0.59	0.04	0.8	25.0
'5/2/2'	BayesianRidge	5	1.22	-0.14	0.83	-0.14	1.2	47.9
'46/3/1'	SVR	3	0.11	-0.68	0.23	-0.65	0.3	20.1
'46/3/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	19.0
'46/5/1'	BayesianRidge	5	9.7	0.65	2.19	0.65	5.2	38.8
'46/4/1'	LinearRegression	0	0.0	1.0	0.0	1.0	0.0	11.0
'46/6/1'	BayesianRidge	5	3.38	0.55	1.48	0.56	2.7	46.5
'46/6/2'	BayesianRidge	5	12.17	-0.02	1.83	0.03	2.6	45.1
'45/1/1'	BayesianRidge	5	2.15	0.68	1.12	0.69	2.7	31.6
'43/1/2'	LinearRegression	1	0.09	0.0	0.24	0.02	0.3	44.0
'43/1/1'	BayesianRidge	5	0.66	0.0	0.56	0.02	0.8	45.3
'43/3/1'	BayesianRidge	5	0.82	-0.12	0.7	-0.08	0.9	27.5
'43/3/2'	LinearRegression	1	0.22	0.16	0.34	0.17	0.5	26.6
'43/2/1'	BayesianRidge	5	1.57	0.52	0.96	0.52	1.8	49.4
'43/2/2'	LinearRegression	0	159.44	-0.01	5.09	0.07	7.6	53.2
'42/1/2'	SVR	3	72.32	-0.01	5.29	0.0	7.5	42.4
'42/2/1'	LinearRegression	0	139.33	0.0	7.83	0.12	8.4	47.6
'42/2/2'	BayesianRidge	5	1.37	-0.03	0.73	0.0	1.1	32.4
'42/1/1'	BayesianRidge	5	5.0	0.16	1.58	0.16	2.3	30.4
'42/3/1'	BayesianRidge	5	37.47	0.82	3.25	0.83	16.0	73.8
'42/3/2'	LinearRegression	1	26.42	0.0	3.18	0.04	3.7	48.6
'41/2/2'	BayesianRidge	5	33.52	0.66	4.27	0.66	10.0	45.1
'41/2/1'	BayesianRidge	5	1078.67	0.15	20.73	0.21	34.5	130.6
'41/3/2'	LinearRegression	1	773.37	0.12	19.93	0.14	30.7	138.8
'41/3/1'	LinearRegression	1	82.71	0.0	6.18	0.0	9.8	71.3
'41/4/1'	LinearRegression	1	278.04	0.41	9.79	0.45	22.0	87.5
'41/1/1'	BayesianRidge	5	44.65	0.73	4.31	0.74	13.6	53.4
'41/1/2'	LinearRegression	1	504.16	0.31	14.68	0.34	24.3	83.7
'41/6/1'	SVR	3	311.61	-0.01	10.68	0.0	14.9	61.1
'41/7/2'	BayesianRidge	5	1178.22	0.13	20.04	0.17	31.1	117.5
'41/7/1'	SVR	3	44.68	-0.03	4.94	-0.02	7.4	45.7
'41/4/2'	LinearRegression	1	339.2	0.04	11.28	0.1	15.9	63.1
'41/5/1'	LinearRegression	1	238.24	0.3	9.51	0.32	18.5	82.6
'41/5/2'	LinearRegression	1	377.66	0.2	11.92	0.24	21.2	83.7
'41/6/2'	BayesianRidge	5	395.13	0.07	12.48	0.07	19.6	80.6

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'40/4/1'	BayesianRidge	5	373.31	0.12	12.1	0.15	17.2	64.7
'40/1/1'	BayesianRidge	5	34.12	0.2	4.44	0.27	6.1	151.5
'40/3/2'	SVR	3	183.56	-0.01	7.51	-0.01	10.9	57.7
'40/2/2'	LinearRegression	1	69.67	0.22	5.56	0.27	8.4	38.3
'40/2/1'	LinearRegression	1	4.43	0.12	1.38	0.13	2.4	22.6
'40/1/2'	LinearRegression	0	878.57	0.41	20.36	0.44	33.7	225.6
'4/5/1'	SVR	3	3.13	-0.03	1.18	-0.03	2.3	23.3
'4/1/2'	BayesianRidge	5	337.34	0.0	4.84	0.0	13.0	109.2
'4/1/1'	BayesianRidge	5	5.64	0.87	1.51	0.87	6.7	119.9
'4/6/2'	LinearRegression	0	1869.87	0.29	21.19	0.32	35.4	81.3
'4/5/2'	BayesianRidge	5	100.13	0.04	3.16	0.06	6.2	22.1
'4/4/2'	BayesianRidge	5	147.02	0.18	8.52	0.21	11.1	57.4
'4/4/1'	BayesianRidge	5	161.01	-0.01	9.9	0.0	12.9	63.5
'4/3/1'	LinearRegression	0	1089.36	0.08	16.16	0.14	20.8	42.7
'4/3/2'	BayesianRidge	5	2.57	0.63	1.04	0.66	2.8	24.0
'4/2/1'	BayesianRidge	5	51.53	0.05	5.7	0.06	7.3	55.7
'4/2/2'	BayesianRidge	5	28.48	0.11	3.71	0.14	5.2	51.9
'4/9/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	18.0
'4/9/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	18.0
'4/8/2'	LinearRegression	0	1.13	0.82	0.37	0.84	2.7	12.9
'4/8/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	7.0
'4/7/1'	BayesianRidge	5	0.1	0.88	0.16	0.89	1.0	74.1
'4/7/2'	BayesianRidge	5	33.52	0.36	3.81	0.36	6.9	87.6
'4/6/1'	BayesianRidge	5	27.23	0.07	3.41	0.08	4.7	46.5
'39/1/1'	BayesianRidge	5	1.54	0.79	0.85	0.79	2.8	21.5
'39/1/2'	LinearRegression	1	21.48	0.59	3.31	0.59	8.0	35.2
'39/3/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	15.0
'39/3/1'	BayesianRidge	5	164.41	0.26	8.7	0.27	15.2	59.2
'39/2/2'	BayesianRidge	5	87.66	0.1	6.44	0.1	9.7	52.7
'38/2/2'	BayesianRidge	5	128.38	0.32	7.09	0.37	15.1	64.4
'38/2/1'	BayesianRidge	5	123.1	0.33	6.36	0.35	14.6	62.5
'38/6/1'	BayesianRidge	5	265.01	0.84	10.69	0.84	43.8	144.7
'38/4/1'	BayesianRidge	5	103.35	0.63	7.35	0.63	17.5	113.4
'38/4/2'	BayesianRidge	5	852.67	0.27	20.06	0.32	31.1	145.2
'38/5/1'	LinearRegression	0	0.03	0.83	0.11	0.83	0.4	43.9
'38/5/2'	BayesianRidge	5	72.56	0.96	5.14	0.96	45.4	138.5
'38/3/2'	LinearRegression	0	0.03	0.73	0.11	0.73	0.3	44.7
'38/3/1'	BayesianRidge	5	648.52	0.45	16.69	0.46	31.0	125.2

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'38/1/1'	LinearRegression	1	1086.99	0.2	18.31	0.22	31.3	92.2
'38/1/2'	LinearRegression	1	526.09	0.27	12.08	0.3	24.6	83.0
'38/6/2'	BayesianRidge	5	281.93	0.71	11.78	0.71	32.6	123.2
'36/3/1'	BayesianRidge	5	31.93	-0.02	4.11	-0.01	6.3	76.7
'36/3/2'	BayesianRidge	5	329.65	0.0	10.56	0.08	15.5	105.1
'36/4/2'	BayesianRidge	5	71.63	0.11	6.13	0.13	9.3	81.1
'36/5/2'	BayesianRidge	5	15.86	0.7	2.48	0.7	7.3	98.1
'36/5/1'	BayesianRidge	5	181.64	0.06	9.9	0.06	13.5	117.2
'36/6/1'	BayesianRidge	5	573.42	0.18	15.59	0.2	24.9	107.9
'36/6/2'	BayesianRidge	5	496.2	-0.01	10.74	0.0	16.5	67.1
'35/9/1'	BayesianRidge	5	177.58	0.65	9.04	0.67	19.9	66.7
'35/8/1'	BayesianRidge	5	29.44	0.73	3.27	0.75	11.3	77.0
'35/8/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	5.0
'35/1/1'	SVR	3	837.25	0.0	20.63	0.0	28.6	113.7
'35/10/1'	LinearRegression	0	492.69	0.08	15.29	0.13	23.9	98.0
'35/10/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	11.0
'35/11/1'	LinearRegression	0	95.6	0.36	6.24	0.36	11.9	45.5
'35/11/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	15.0
'35/14/1'	SVR	3	189.03	-0.01	9.42	-0.01	13.7	60.7
'35/14/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	10.0
'35/7/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	7.0
'35/15/1'	LinearRegression	0	434.82	0.23	14.08	0.25	22.6	78.6
'35/7/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	7.0
'35/6/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	6.0
'35/6/1'	BayesianRidge	5	124.07	0.29	6.96	0.29	14.8	82.3
'35/5/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	10.0
'35/5/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	10.0
'35/4/1'	BayesianRidge	5	352.08	-0.06	14.08	0.03	18.7	79.8
'35/4/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	13.0
'35/12/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	18.0
'35/12/1'	SVR	3	367.95	-0.02	12.9	0.0	19.8	86.8
'35/3/1'	BayesianRidge	5	167.5	0.51	8.69	0.52	18.3	82.7
'35/13/2'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	19.0
'35/13/1'	SVR	3	44.04	-0.03	4.71	-0.02	6.8	44.5
'35/2/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	11.0
'35/2/1'	LinearRegression	1	497.96	0.25	13.84	0.28	26.3	114.8
'34/1/1'	LinearRegression	0	671.44	0.01	16.08	0.08	20.1	63.8
'34/8/1'	LinearRegression	0	1334.32	0.28	25.75	0.35	43.4	181.2

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'34/9/1'	BayesianRidge	5	232.64	0.13	11.19	0.17	15.2	125.2
'34/9/2'	SVR	3	17.83	0.34	2.97	0.36	5.8	96.0
'34/6/1'	BayesianRidge	5	360.64	0.35	14.04	0.38	23.7	98.2
'34/7/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	16.0
'34/4/1'	BayesianRidge	5	1329.61	0.41	25.51	0.44	44.6	161.5
'34/5/1'	BayesianRidge	5	515.87	0.28	16.7	0.32	25.1	102.9
'34/5/2'	SVR	3	1.01	0.38	0.61	0.39	1.4	22.4
'34/2/1'	BayesianRidge	5	881.07	0.35	20.46	0.37	33.8	117.9
'34/3/2'	SVR	3	1.24	0.38	0.68	0.38	1.6	25.7
'34/3/1'	LinearRegression	0	672.63	0.02	10.25	0.04	16.4	47.0
'33/5/2'	BayesianRidge	5	517.4	0.49	14.18	0.49	29.2	121.0
'33/1/2'	SVR	3	204.38	-0.01	8.82	-0.01	12.9	48.4
'33/1/1'	LinearRegression	0	493.47	0.1	16.57	0.16	21.0	79.8
'33/3/1'	BayesianRidge	5	285.0	0.09	11.28	0.12	14.7	126.8
'33/3/2'	BayesianRidge	5	853.2	0.19	18.18	0.22	26.0	149.0
'33/2/1'	BayesianRidge	5	630.72	0.06	19.13	0.09	24.0	159.6
'33/2/2'	BayesianRidge	5	28.15	0.22	3.85	0.23	6.3	116.1
'33/4/2'	BayesianRidge	5	485.51	0.39	13.9	0.4	27.9	105.6
'33/4/1'	BayesianRidge	5	402.27	0.38	12.09	0.38	25.7	92.6
'33/5/1'	BayesianRidge	5	843.16	0.18	19.51	0.21	27.3	121.2
'31/3/2'	LinearRegression	0	329.22	0.24	12.62	0.27	18.9	70.2
'31/6/2'	BayesianRidge	5	485.49	0.39	13.9	0.4	27.9	105.6
'31/6/1'	BayesianRidge	5	405.21	0.51	12.46	0.52	28.6	109.4
'31/5/1'	BayesianRidge	5	285.0	0.09	11.28	0.12	14.7	126.8
'31/5/2'	BayesianRidge	5	853.21	0.19	18.18	0.22	26.0	149.0
'31/2/1'	BayesianRidge	5	11.82	0.68	1.9	0.68	6.2	106.4
'30/1/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	82.0
'30/6/2'	BayesianRidge	5	0.22	0.01	0.32	0.07	0.4	16.9
'30/7/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	25.0
'30/4/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	42.0
'30/5/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	42.0
'30/15/1'	LinearRegression	0	0.0	1.0	0.0	1.0	0.0	35.0
'30/14/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	24.0
'30/14/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	24.0
'30/16/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	31.0
'30/12/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	23.0
'3/2/1'	LinearRegression	0	0.0	1.0	0.0	1.0	0.0	28.0
'3/3/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	63.0

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'28/8/2'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	32.0
'28/3/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	30.0
'28/5/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	30.0
'28/4/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	30.0
'28/6/2'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	25.0
'28/1/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	57.0
'28/2/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	19.0
'27/10/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	34.0
'27/1/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	44.0
'26/3/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	48.0
'26/2/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	6.0
'26/1/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	48.0
'25/2/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	25.0
'25/3/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	7.0
'25/4/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	44.0
'25/1/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	25.0

Table F.2: Peak Algorithm Results

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'1/13/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	5.0
'8/5/2'	BayesianRidge	2	679.46	0.27	19.79	0.33	28.9	76.6
'8/1/1'	SVR	6	8482.33	-0.01	78.47	0.0	88.3	235.9
'8/1/2'	SVR	6	4317.35	-0.02	51.51	0.0	74.2	197.7
'8/2/2'	BayesianRidge	2	8181.3	0.05	80.17	0.05	99.1	248.6
'8/3/1'	BayesianRidge	2	545.47	-0.02	18.51	-0.02	23.3	73.4
'9/5/1'	BayesianRidge	2	73.14	0.73	2.9	0.74	13.2	22.0
'9/5/2'	BayesianRidge	2	73.28	0.73	2.91	0.74	13.2	22.0
'9/4/1'	LinearRegression	1	35.97	0.77	2.7	0.78	10.1	33.0
'9/4/2'	BayesianRidge	2	41.73	0.74	2.17	0.74	10.0	32.0
'9/7/1'	BayesianRidge	2	1730.01	0.04	35.59	0.04	45.1	123.4
'9/1/1'	BayesianRidge	2	128.67	0.25	8.52	0.27	14.1	49.1
'9/1/2'	PassiveAggressiveRegressor	3	14.66	0.77	1.41	0.77	6.3	14.0
'9/14/1'	BayesianRidge	2	46.95	-0.03	1.88	0.02	3.9	19.4
'9/14/2'	BayesianRidge	2	2905.59	0.14	42.54	0.14	58.7	168.3
'9/12/2'	LinearRegression	0	1313.05	0.28	25.52	0.29	34.6	80.0
'9/12/1'	LinearRegression	0	807.2	0.16	19.27	0.17	27.3	80.0
'9/13/2'	BayesianRidge	2	12.13	0.71	1.4	0.72	5.2	10.0

Table F.2: Peak Algorithm Results

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'17/4/1'	BayesianRidge	2	5202.11	0.42	62.01	0.43	97.6	237.1
'18/3/1'	BayesianRidge	2	4803.38	0.14	46.0	0.15	67.0	132.8
'18/2/1'	BayesianRidge	2	3063.09	0.03	37.45	0.04	51.4	253.3
'18/8/1'	BayesianRidge	2	328.03	0.76	9.82	0.8	39.7	137.0
'18/1/1'	BayesianRidge	2	6952.7	0.19	64.41	0.2	75.6	306.0
'18/7/1'	LinearRegression	0	1821.05	0.15	29.39	0.15	48.0	148.0
'18/6/1'	BayesianRidge	2	1490.9	-0.03	26.94	-0.03	40.0	142.6
'18/5/1'	BayesianRidge	2	2143.29	0.23	34.49	0.24	58.6	207.4
'18/4/1'	SVR	6	8739.74	-0.02	80.83	0.0	95.7	293.9
'19/2/1'	BayesianRidge	2	619.29	0.59	16.44	0.59	42.6	152.8
'19/3/1'	SVR	6	1168.55	0.0	26.42	0.0	32.7	109.8
'19/4/1'	BayesianRidge	2	3018.02	-0.11	39.77	0.01	55.0	199.4
'19/1/1'	BayesianRidge	2	1178.63	0.11	22.38	0.13	35.1	128.9
'20/2/1'	BayesianRidge	2	217.22	-0.16	8.87	-0.15	13.1	35.0
'20/3/1'	LinearRegression	0	3.76	0.85	0.87	0.86	5.4	33.8
'20/1/1'	BayesianRidge	2	220.09	0.27	10.98	0.29	18.3	54.4
'20/6/1'	BayesianRidge	2	25.66	0.38	4.24	0.49	8.5	27.6
'20/5/1'	LinearRegression	0	17.23	0.47	3.22	0.57	7.9	17.0
'20/4/1'	LinearRegression	0	11.9	0.32	2.15	0.32	3.5	37.5
'21/5/1'	LinearRegression	0	34.54	0.58	4.16	0.6	8.6	31.4
'21/5/2'	BayesianRidge	2	259.53	0.86	9.22	0.87	45.3	309.0
'21/4/1'	PassiveAggressiveRegressor	3	50.79	0.41	5.45	0.43	9.6	50.5
'21/4/2'	SVR	6	8980.95	-0.03	82.43	0.0	98.9	479.9
'21/2/2'	BayesianRidge	2	555.38	0.86	13.45	0.87	65.4	445.0
'21/2/1'	BayesianRidge	2	2938.61	-0.03	45.58	-0.02	69.0	115.8
'21/3/2'	BayesianRidge	2	302.52	0.79	10.5	0.81	40.0	269.0
'21/3/1'	SVR	6	469.03	-0.07	16.5	0.0	24.9	75.0
'22/11/1'	BayesianRidge	2	14.01	0.52	2.96	0.53	6.1	33.8
'22/10/2'	PassiveAggressiveRegressor	3	5.43	0.36	1.53	0.37	2.9	56.1
'22/10/1'	LinearRegression	0	0.18	0.07	0.4	0.1	0.4	52.0
'22/13/1'	BayesianRidge	2	25.26	0.53	3.63	0.57	8.5	26.6
'22/13/2'	BayesianRidge	2	49.49	0.41	5.56	0.43	10.3	35.1
'22/11/2'	LinearRegression	0	28.63	0.54	4.5	0.55	8.2	36.6
'22/12/1'	BayesianRidge	2	43.25	0.36	5.37	0.36	9.3	35.1
'22/12/2'	LinearRegression	1	1.64	0.79	0.75	0.79	3.0	21.3
'22/14/2'	BayesianRidge	2	0.69	0.21	0.6	0.23	1.1	43.2

Table F.2: Peak Algorithm Results

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'22/8/2'	BayesianRidge	2	0.09	0.56	0.19	0.56	0.5	58.0
'22/8/1'	BayesianRidge	2	23.75	0.49	3.52	0.51	7.6	73.3
'22/9/2'	BayesianRidge	2	134.6	0.62	9.3	0.63	20.2	63.4
'22/9/1'	BayesianRidge	2	39.69	0.48	4.64	0.49	10.7	39.2
'22/2/1'	PassiveAggressiveRegressor	3	48.59	0.45	5.58	0.45	10.9	52.3
'22/2/2'	LinearRegression	1	3.1	0.87	0.76	0.87	5.2	42.4
'22/14/1'	PassiveAggressiveRegressor	3	30.82	0.21	3.01	0.21	6.0	51.7
'22/3/1'	BayesianRidge	2	1563.22	0.33	32.17	0.34	53.6	203.0
'22/3/2'	PassiveAggressiveRegressor	3	1012.59	0.45	21.52	0.45	52.5	173.1
'22/4/2'	LinearRegression	1	1073.42	0.08	22.67	0.12	31.6	108.0
'22/4/1'	SVR	6	1292.14	-0.25	25.77	-0.01	47.6	132.5
'22/5/2'	LinearRegression	1	6546.34	0.45	50.5	0.47	87.3	173.8
'22/5/1'	BayesianRidge	2	0.42	-0.12	0.46	-0.12	0.6	99.7
'22/6/1'	BayesianRidge	2	134.51	0.34	9.13	0.35	15.8	63.3
'22/6/2'	BayesianRidge	2	0.02	0.88	0.11	0.89	0.4	32.9
'22/7/1'	BayesianRidge	2	11.91	0.19	2.56	0.19	4.0	36.9
'22/7/2'	BayesianRidge	2	37.41	0.07	4.9	0.08	6.1	42.9
'25/4/1'	SVR	6	3332.11	-0.07	41.04	0.0	60.7	134.8
'25/3/1'	SVR	6	295.26	-0.12	14.88	0.01	19.6	96.2
'25/2/1'	BayesianRidge	2	5698.57	0.13	56.12	0.13	79.4	208.0
'25/1/1'	SVR	6	575.28	-0.03	16.39	0.0	27.8	110.7
'26/1/1'	SVR	6	421.02	-0.13	11.0	-0.03	22.2	82.6
'26/2/1'	PassiveAggressiveRegressor	3	267.52	0.3	10.71	0.31	20.5	86.4
'26/3/1'	BayesianRidge	2	4495.81	0.05	49.26	0.05	75.9	160.0
'27/3/1'	BayesianRidge	2	1549.17	0.3	27.94	0.3	44.9	210.0
'27/4/1'	SVR	6	8520.71	-0.14	79.3	0.0	91.3	300.0
'27/7/1'	SVR	6	2914.88	0.0	46.41	0.0	54.8	200.8
'27/6/1'	BayesianRidge	2	201.73	0.59	8.44	0.61	24.8	102.0
'27/9/1'	SVR	6	2111.9	0.0	40.27	0.0	45.1	171.0
'27/8/1'	BayesianRidge	2	4643.79	0.3	50.75	0.35	80.9	228.4
'27/1/1'	BayesianRidge	2	32733.01	-0.23	143.1	-0.07	143.4	313.6
'27/2/1'	SVR	6	1245.0	-0.02	28.75	0.0	37.3	145.0
'27/5/1'	LinearRegression	0	1477.93	0.28	20.67	0.28	47.4	177.0
'27/10/1'	SVR	6	13437.77	0.0	99.6	0.0	116.2	364.9
'28/2/1'	SVR	6	6522.39	-0.06	71.85	0.0	82.3	260.2
'28/3/1'	BayesianRidge	2	12961.59	-0.04	91.03	-0.04	119.7	331.4

Table F.2: Peak Algorithm Results

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'28/1/1'	BayesianRidge	2	2275.8	0.1	38.12	0.12	56.0	160.0
'28/6/1'	SVR	6	922.25	-0.04	25.24	0.0	37.2	110.6
'28/7/1'	SVR	6	4849.99	-0.01	58.45	0.0	72.2	221.6
'28/4/1'	BayesianRidge	2	9416.35	0.03	74.41	0.04	107.0	301.1
'28/5/1'	BayesianRidge	2	245.07	-0.1	12.18	-0.09	16.7	75.3
'28/8/1'	SVR	6	5505.28	-0.04	56.79	0.0	76.2	197.1
'29/2/2'	SVR	6	631.14	-0.14	15.14	0.0	34.3	89.0
'29/2/1'	SVR	6	46146.67	-0.01	177.77	0.0	192.7	540.5
'3/3/1'	SVR	6	622.09	0.0	21.35	0.0	26.5	115.3
'3/2/2'	SVR	6	11.1	0.06	2.66	0.06	14.6	37.2
'3/1/2'	PassiveAggressiveRegressor	3	48.69	0.18	3.86	0.21	7.8	17.7
'3/1/1'	SVR	6	125.08	0.0	8.84	0.0	17.9	44.8
'30/3/1'	PassiveAggressiveRegressor	3	0.05	0.64	0.15	0.66	0.4	47.0
'30/23/2'	BayesianRidge	2	3.86	0.03	1.33	0.04	2.3	31.9
'30/23/1'	SVR	6	2077.68	-0.11	14.6	0.0	50.3	33.3
'30/22/2'	BayesianRidge	2	11.21	-0.01	2.38	0.0	4.7	49.7
'30/22/1'	LinearRegression	0	0.15	0.12	0.33	0.13	0.4	47.0
'30/21/1'	LinearRegression	0	255.33	0.24	8.93	0.25	15.8	49.4
'30/21/2'	BayesianRidge	2	11.4	0.06	2.9	0.06	3.8	37.3
'30/20/1'	BayesianRidge	2	2180.31	0.06	36.62	0.06	54.3	119.0
'30/20/2'	SVR	6	80.98	0.0	6.01	0.0	8.1	33.7
'30/17/2'	BayesianRidge	2	208.82	-0.03	8.86	0.01	13.7	58.7
'30/25/1'	BayesianRidge	2	214.26	0.26	10.94	0.29	18.3	51.6
'30/25/2'	LinearRegression	1	0.09	0.16	0.23	0.16	0.4	27.1
'30/24/1'	LinearRegression	0	175.28	0.68	7.03	0.69	25.5	68.7
'30/24/2'	SVR	6	52.41	-0.09	5.39	-0.05	7.6	34.3
'30/12/2'	BayesianRidge	2	5314.88	-0.18	62.87	-0.18	69.0	155.9
'30/16/1'	BayesianRidge	2	5.73	0.2	1.79	0.21	2.8	37.3
'30/17/1'	BayesianRidge	2	141.86	-0.05	8.41	-0.03	11.5	60.8
'30/15/2'	SVR	6	4812.18	-0.08	55.45	0.0	72.1	184.0
'30/5/1'	BayesianRidge	2	16.25	0.03	3.32	0.06	4.1	51.3
'30/7/1'	SVR	6	52.73	-0.01	6.2	0.01	7.7	43.6
'30/6/2'	BayesianRidge	2	0.9	0.03	0.41	0.05	0.8	17.0
'30/1/2'	LinearRegression	0	0.12	0.38	0.31	0.43	0.5	83.0
'30/3/2'	BayesianRidge	2	66.64	-0.07	5.91	-0.06	8.4	57.1
'30/2/1'	SVR	6	95.79	-0.16	5.97	-0.04	10.7	39.6

Table F.2: Peak Algorithm Results

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'30/8/2'	SVR	6	321.17	-0.24	11.01	0.01	27.1	56.2
'30/9/1'	BayesianRidge	2	85.3	-0.02	5.64	0.01	8.3	69.3
'30/18/1'	PassiveAggressiveRegressor	3	485.42	0.12	14.35	0.14	22.3	92.1
'30/18/2'	BayesianRidge	2	0.59	-0.13	0.55	0.04	1.7	76.1
'30/19/1'	BayesianRidge	2	5285.28	0.16	57.82	0.17	88.8	214.0
'30/19/2'	BayesianRidge	2	3.27	0.04	1.56	0.04	2.3	51.5
'30/10/2'	SVR	6	285.77	-0.22	8.17	-0.05	13.3	32.7
'30/11/1'	SVR	6	2395.88	-0.03	41.84	0.0	46.5	116.2
'31/2/1'	LinearRegression	0	0.17	0.91	0.25	0.91	1.4	98.7
'31/2/2'	BayesianRidge	2	478.04	0.21	17.4	0.21	26.1	162.3
'31/7/1'	LinearRegression	1	71.63	0.36	2.96	0.36	10.7	53.0
'31/7/2'	BayesianRidge	2	8.79	0.86	1.82	0.86	8.1	53.0
'31/1/2'	SVR	6	1678.76	-0.04	32.86	-0.01	41.3	123.5
'31/1/1'	SVR	6	2496.66	-0.05	43.41	0.0	49.1	159.0
'31/4/1'	SVR	6	16178.37	-0.01	107.56	0.0	127.7	368.7
'31/5/2'	BayesianRidge	2	3471.84	-0.08	42.55	-0.08	58.7	254.0
'31/5/1'	SVR	6	4009.88	-0.01	51.79	0.0	60.1	226.6
'31/6/1'	BayesianRidge	2	1587.09	0.16	30.74	0.18	42.4	155.0
'31/6/2'	LinearRegression	1	1252.33	0.21	15.02	0.21	41.9	155.0
'31/3/2'	SVR	6	2316.55	-0.03	41.31	-0.01	46.9	154.5
'33/5/1'	SVR	6	12532.97	-0.03	88.54	0.0	107.4	301.3
'33/4/1'	BayesianRidge	2	1466.85	0.25	26.37	0.25	43.6	151.5
'33/4/2'	LinearRegression	1	1252.33	0.21	15.02	0.21	41.9	155.0
'33/2/2'	BayesianRidge	2	671.64	0.23	21.25	0.24	31.2	176.2
'33/2/1'	BayesianRidge	2	31802.71	-0.09	138.51	-0.09	176.2	442.0
'33/3/2'	BayesianRidge	2	3471.84	-0.08	42.55	-0.08	58.7	254.0
'33/3/1'	SVR	6	4009.88	-0.01	51.79	0.0	60.1	226.6
'33/1/1'	SVR	6	2344.03	-0.08	41.64	0.0	47.4	159.0
'33/1/2'	SVR	6	2045.4	-0.01	37.55	-0.01	46.1	146.6
'33/5/2'	BayesianRidge	2	1897.58	0.01	28.04	0.01	47.5	182.5
'34/3/1'	BayesianRidge	2	4137.28	-0.08	43.78	-0.07	50.3	111.5
'34/3/2'	BayesianRidge	2	610.47	0.2	17.78	0.34	31.8	79.2
'34/2/1'	BayesianRidge	2	3484.94	0.02	48.02	0.03	54.6	213.6
'34/2/2'	BayesianRidge	2	244.57	0.2	11.22	0.33	20.0	49.3
'34/5/2'	BayesianRidge	2	483.39	0.2	15.96	0.33	28.2	69.5
'34/5/1'	BayesianRidge	2	1984.33	0.12	33.46	0.13	48.2	162.2

Table F.2: Peak Algorithm Results

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'34/4/2'	BayesianRidge	2	291.06	0.27	12.14	0.4	22.9	59.2
'34/4/1'	BayesianRidge	2	4353.45	0.16	46.22	0.17	72.1	255.1
'34/7/2'	BayesianRidge	2	377.54	0.22	13.8	0.35	25.1	60.9
'34/6/1'	PassiveAggressiveRegressor	3	1477.05	0.06	29.14	0.14	38.8	146.3
'34/6/2'	BayesianRidge	2	120.66	0.23	7.86	0.35	14.3	35.7
'34/9/2'	BayesianRidge	2	8659.61	0.29	65.42	0.42	125.5	319.1
'34/9/1'	BayesianRidge	2	2902.82	0.02	39.75	0.03	61.2	208.9
'34/8/1'	BayesianRidge	2	2724.79	0.11	38.2	0.16	61.3	219.0
'34/1/2'	BayesianRidge	2	164.66	0.23	9.28	0.37	17.1	44.8
'34/1/1'	SVR	6	4005.21	-0.12	54.0	0.0	56.3	165.4
'35/2/1'	SVR	6	1508.69	-0.06	29.62	0.02	40.9	161.0
'35/13/1'	SVR	6	1617.14	-0.09	33.7	0.0	43.8	126.5
'35/3/1'	LinearRegression	1	4895.83	-0.07	50.05	-0.07	61.3	159.7
'35/4/1'	BayesianRidge	2	644.97	0.17	21.61	0.18	33.8	99.5
'35/6/1'	BayesianRidge	2	156.65	0.43	5.24	0.46	18.1	93.0
'35/15/1'	SVR	6	3572.56	-0.04	51.39	0.0	57.9	178.6
'35/14/1'	BayesianRidge	2	1416.51	0.02	28.92	0.03	38.8	131.8
'35/11/1'	BayesianRidge	2	771.72	0.13	20.31	0.16	25.1	72.6
'35/10/1'	SVR	6	2045.14	-0.22	37.9	0.0	44.2	164.6
'35/1/1'	SVR	6	6251.72	-0.02	63.0	0.0	78.1	235.7
'35/8/1'	BayesianRidge	2	43.41	0.68	3.75	0.76	12.3	80.7
'35/9/1'	BayesianRidge	2	1250.98	0.39	27.43	0.4	44.8	131.5
'36/6/2'	SVR	6	6077.7	-0.02	60.03	0.0	81.3	216.7
'36/6/1'	BayesianRidge	2	15165.89	0.1	103.91	0.15	130.0	388.2
'36/5/1'	BayesianRidge	2	10711.8	0.23	86.82	0.23	123.3	383.0
'36/5/2'	BayesianRidge	2	227.07	0.43	11.81	0.43	21.4	126.1
'36/4/2'	SVR	6	1002.8	-0.01	23.37	0.0	36.6	145.9
'36/3/2'	BayesianRidge	2	1490.71	-0.01	30.34	0.01	42.3	190.3
'36/3/1'	BayesianRidge	2	2085.92	0.2	34.17	0.2	54.9	164.4
'38/6/2'	BayesianRidge	2	2403.44	0.45	34.77	0.45	70.3	227.8
'38/1/2'	SVR	6	3142.99	-0.03	48.4	0.0	54.8	156.7
'38/1/1'	BayesianRidge	2	6820.59	-0.04	64.43	-0.03	77.6	186.6
'38/3/1'	BayesianRidge	2	2793.96	0.44	42.83	0.45	73.3	226.1
'38/3/2'	LinearRegression	0	0.1	0.53	0.19	0.54	0.5	45.0
'38/5/2'	BayesianRidge	2	646.46	0.91	16.71	0.91	90.6	240.3
'38/5/1'	SVR	6	0.09	0.58	0.22	0.58	0.5	44.0

Table F.2: Peak Algorithm Results

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'38/4/2'	BayesianRidge	2	4652.76	0.32	54.32	0.34	84.2	298.7
'38/4/1'	BayesianRidge	2	4248.05	0.19	49.78	0.21	76.8	241.5
'38/6/1'	BayesianRidge	2	1728.66	0.68	27.3	0.69	79.6	236.3
'38/2/1'	LinearRegression	1	316.78	0.16	6.72	0.18	21.5	83.0
'38/2/2'	LinearRegression	1	235.33	0.46	7.7	0.46	21.6	83.0
'39/2/2'	SVR	6	2674.95	-0.02	33.61	0.0	53.4	123.1
'39/3/1'	SVR	6	1408.03	-0.02	28.73	-0.01	38.1	110.5
'39/1/2'	BayesianRidge	2	840.97	0.11	18.08	0.12	31.8	66.7
'39/1/1'	BayesianRidge	2	36.16	0.72	4.35	0.73	12.1	34.4
'4/6/1'	BayesianRidge	2	572.93	-0.07	18.18	-0.02	21.8	80.8
'4/7/2'	BayesianRidge	2	919.85	0.18	20.05	0.18	29.8	110.3
'4/7/1'	LinearRegression	1	9.32	0.79	1.13	0.81	7.2	87.4
'4/8/2'	SVR	6	0.09	0.58	0.23	0.58	0.5	8.0
'4/2/2'	LinearRegression	0	146.71	0.0	9.94	0.1	14.0	68.2
'4/2/1'	BayesianRidge	2	3151.26	0.02	41.19	0.03	63.9	172.2
'4/3/2'	SVR	6	3.61	0.02	0.74	0.03	1.2	19.2
'4/3/1'	BayesianRidge	2	4998.05	0.17	51.87	0.18	75.0	186.1
'4/4/1'	BayesianRidge	2	7507.27	0.02	71.46	0.02	84.8	235.4
'4/4/2'	BayesianRidge	2	2217.02	0.01	36.77	0.05	46.8	143.3
'4/5/2'	LinearRegression	0	235.94	0.08	9.49	0.08	15.3	23.2
'4/6/2'	BayesianRidge	2	5722.07	0.04	56.15	0.09	75.3	127.6
'4/1/1'	LinearRegression	0	194.86	0.83	7.97	0.83	36.3	150.2
'4/1/2'	SVR	6	0.95	0.08	0.62	0.09	7.2	109.0
'4/5/1'	SVR	6	56.35	-0.11	5.53	0.0	8.0	38.5
'40/1/2'	BayesianRidge	2	9790.79	0.0	73.34	0.01	90.3	380.6
'40/2/1'	BayesianRidge	2	69.21	0.34	6.5	0.34	11.5	43.1
'40/2/2'	BayesianRidge	2	981.57	-0.01	19.75	0.0	32.2	83.6
'40/3/2'	SVR	6	2836.03	-0.01	36.45	0.0	55.7	147.6
'40/1/1'	BayesianRidge	2	1849.22	-0.04	30.7	0.0	39.5	204.5
'40/4/1'	SVR	6	1158.26	-0.07	24.35	0.0	35.7	116.9
'41/6/2'	SVR	6	1831.23	0.0	32.08	0.0	46.3	164.0
'41/5/2'	SVR	6	1357.42	-0.05	32.32	0.0	38.5	141.5
'41/5/1'	BayesianRidge	2	1312.41	0.06	26.73	0.08	38.3	142.0
'41/4/2'	SVR	6	1336.21	-0.16	30.8	0.0	35.9	131.7
'41/7/1'	BayesianRidge	2	1533.95	0.01	31.1	0.01	42.9	118.5
'41/7/2'	SVR	6	2812.31	-0.01	36.65	0.0	54.7	206.2

Table F.2: Peak Algorithm Results

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'41/6/1'	SVR	6	2825.62	0.0	43.77	0.0	53.7	176.9
'41/1/2'	PassiveAggressiveRegressor	3	1652.27	0.29	25.24	0.29	46.2	151.6
'41/1/1'	BayesianRidge	2	464.23	0.48	16.49	0.48	31.3	89.5
'41/4/1'	BayesianRidge	2	1068.03	0.0	18.04	0.0	34.8	128.6
'41/3/1'	BayesianRidge	2	831.4	0.13	21.93	0.13	34.5	125.9
'41/3/2'	BayesianRidge	2	5433.08	-0.04	57.21	-0.04	79.6	267.4
'41/2/1'	SVR	6	5079.44	-0.01	52.36	0.0	74.6	260.6
'41/2/2'	BayesianRidge	2	2370.64	0.09	38.15	0.1	55.6	106.0
'42/3/2'	SVR	6	215.55	-0.08	11.36	-0.02	14.7	78.2
'42/3/1'	LinearRegression	1	15.51	0.24	2.68	0.29	5.2	43.3
'42/1/1'	BayesianRidge	2	25.1	0.21	4.1	0.21	5.7	39.2
'42/2/2'	SVR	6	19.15	-0.07	3.57	-0.05	4.7	41.9
'42/2/1'	SVR	6	448.11	-0.01	18.19	0.0	21.1	83.8
'42/1/2'	SVR	6	544.62	-0.06	16.13	0.0	27.6	88.7
'43/2/2'	LinearRegression	1	394.34	0.04	14.98	0.05	17.0	81.1
'43/2/1'	PassiveAggressiveRegressor	3	161.42	0.43	7.94	0.44	18.7	72.6
'43/3/2'	BayesianRidge	2	10.13	0.02	2.74	0.03	2.9	32.0
'43/3/1'	SVR	6	30.05	0.0	4.6	0.0	5.7	38.1
'43/1/1'	BayesianRidge	2	17.46	0.05	3.47	0.06	4.4	52.9
'43/1/2'	BayesianRidge	2	3.9	0.01	1.52	0.02	2.2	46.5
'45/1/1'	BayesianRidge	2	110.22	0.34	7.23	0.38	13.2	42.7
'46/6/2'	BayesianRidge	2	139.24	-0.01	8.81	0.01	10.7	64.4
'46/6/1'	BayesianRidge	2	330.38	0.37	15.48	0.39	24.3	90.9
'46/4/2'	LinearRegression	0	47.49	0.21	5.54	0.21	8.1	29.3
'46/5/1'	BayesianRidge	2	200.16	0.38	10.66	0.39	18.8	60.3
'46/3/1'	BayesianRidge	2	12.67	-0.07	2.04	-0.05	3.5	21.7
'5/2/2'	BayesianRidge	2	243.74	-0.01	11.49	0.01	16.0	66.5
'5/5/1'	SVR	6	17.11	-0.01	3.48	0.04	4.3	33.5
'5/5/2'	BayesianRidge	2	2125.39	-0.03	17.71	-0.01	27.5	50.5
'5/1/1'	SVR	6	53757.74	-0.06	210.89	0.0	225.2	632.9
'5/3/1'	BayesianRidge	2	1997.86	0.05	35.84	0.06	42.7	116.7
'5/1/2'	BayesianRidge	2	4168.68	0.0	50.45	0.0	77.7	258.4
'5/6/2'	PassiveAggressiveRegressor	3	161.33	0.34	6.46	0.34	15.7	94.6
'5/6/1'	LinearRegression	1	67.38	0.15	5.69	0.18	7.7	101.5
'5/4/1'	BayesianRidge	2	81.27	0.05	7.24	0.05	9.6	37.4
'5/4/2'	PassiveAggressiveRegressor	3	2634.48	0.05	39.77	0.21	55.3	134.9

Table F.2: Peak Algorithm Results

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'6/5/2'	BayesianRidge	2	1960.72	0.13	35.12	0.14	47.8	113.7
'6/6/1'	BayesianRidge	2	820.85	0.21	22.3	0.22	31.0	94.9
'6/7/2'	SVR	6	6694.84	-0.03	67.15	0.0	79.2	232.3
'6/2/1'	BayesianRidge	2	1013.49	0.02	23.84	0.02	30.9	73.9
'6/2/2'	SVR	6	2066.05	-0.03	37.49	0.0	46.3	139.5
'6/1/2'	SVR	6	11062.35	0.0	88.07	0.0	105.8	320.2
'6/1/1'	SVR	6	67.51	-0.05	6.38	0.0	8.7	43.7
'6/3/1'	BayesianRidge	2	246.67	0.04	12.31	0.07	15.3	55.7
'6/3/2'	BayesianRidge	2	422.88	-0.19	15.45	-0.19	19.8	69.3
'6/4/2'	BayesianRidge	2	4208.15	0.68	50.95	0.68	118.3	337.4
'6/4/1'	BayesianRidge	2	51.97	0.08	5.84	0.08	8.0	85.4
'7/10/2'	BayesianRidge	2	3386.69	0.02	44.9	0.02	61.7	185.5
'7/11/2'	SVR	6	2013.17	-0.01	36.7	-0.01	50.0	126.9
'7/12/1'	BayesianRidge	2	485.39	0.02	16.38	0.03	21.1	66.7
'7/13/2'	BayesianRidge	2	2781.85	0.03	39.49	0.05	57.2	138.4
'7/7/1'	BayesianRidge	2	158.53	0.08	10.37	0.12	14.7	66.2
'7/7/2'	BayesianRidge	2	2353.34	0.03	38.32	0.03	46.6	133.7
'7/6/1'	LinearRegression	1	112.46	0.05	7.98	0.05	11.4	78.0
'7/5/1'	PassiveAggressiveRegressor	3	131.65	0.46	5.54	0.49	14.9	53.0
'7/3/2'	PassiveAggressiveRegressor	3	2.07	-0.15	0.74	0.01	1.0	23.6
'7/2/1'	BayesianRidge	2	3944.81	0.11	49.9	0.12	64.2	187.8
'7/1/2'	BayesianRidge	2	6004.83	0.04	62.88	0.04	87.8	247.3
'7/9/2'	SVR	6	4804.82	-0.02	53.56	0.0	77.2	190.8
'7/9/1'	BayesianRidge	2	48.48	0.2	5.43	0.23	8.2	43.7
'7/10/1'	BayesianRidge	2	1844.94	0.15	36.99	0.16	49.4	142.8
'7/14/1'	BayesianRidge	2	255.48	0.02	11.86	0.03	16.9	45.0
'7/8/2'	SVR	6	9336.05	0.0	80.15	0.0	99.7	302.8
'7/8/1'	BayesianRidge	2	3723.16	0.05	49.05	0.05	67.7	189.4
'8/4/1'	BayesianRidge	2	29.92	0.01	3.96	0.09	8.4	30.3
'8/4/2'	BayesianRidge	2	4741.1	0.03	56.0	0.04	69.5	203.2
'9/7/2'	BayesianRidge	2	6.25	0.71	0.91	0.72	3.7	18.0
'9/13/1'	BayesianRidge	2	11.53	0.73	1.31	0.75	5.3	10.0
'9/10/2'	BayesianRidge	2	11.94	0.72	1.34	0.73	5.2	12.0
'9/11/1'	BayesianRidge	2	1166.4	0.27	25.2	0.3	46.0	127.9
'9/11/2'	BayesianRidge	2	2204.45	-0.01	37.76	0.0	54.8	128.0
'9/6/2'	BayesianRidge	2	32.27	0.73	2.04	0.74	8.7	7.0

Table F.2: Peak Algorithm Results

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'9/9/1'	SVR	6	102.48	0.0	6.67	0.0	18.3	37.5
'9/9/2'	BayesianRidge	2	4221.43	0.04	57.83	0.05	68.0	187.4
'9/8/1'	BayesianRidge	2	12.0	0.66	1.48	0.7	4.8	8.1
'9/8/2'	BayesianRidge	2	37.11	0.18	3.16	0.35	5.0	11.0
'9/3/1'	LinearRegression	1	27.81	0.75	2.2	0.76	8.4	37.0
'9/6/1'	PassiveAggressiveRegressor	3	45.83	0.63	3.28	0.66	8.9	8.9
'9/2/2'	PassiveAggressiveRegressor	3	92.39	0.76	3.19	0.76	15.6	40.0
'9/2/1'	PassiveAggressiveRegressor	3	94.59	0.75	3.19	0.75	15.5	41.0
'1/13/1'	SVR	5	380.22	0.01	16.07	0.01	20.3	60.8
'1/10/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	7.0
'1/10/2'	SVR	5	740.87	-0.02	22.71	-0.01	28.5	79.7
'1/11/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	18.0
'1/4/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	32.0
'1/4/2'	BayesianRidge	2	1636.35	0.41	26.64	0.46	55.8	126.7
'1/11/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	18.0
'1/9/1'	BayesianRidge	2	503.14	0.23	17.61	0.23	26.7	85.1
'1/9/2'	SVR	5	1509.83	-0.1	31.25	-0.01	38.8	114.6
'1/7/2'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	16.0
'1/7/1'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	16.0
'1/6/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	17.0
'1/6/1'	BayesianRidge	2	15259.29	-0.02	96.62	-0.01	120.3	330.1
'1/8/1'	BayesianRidge	2	1405.99	0.18	29.64	0.18	40.3	130.0
'1/8/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	9.0
'1/5/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	31.0
'1/5/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	16.0
'1/12/2'	BayesianRidge	2	141.33	0.46	7.18	0.48	17.0	44.0
'1/12/1'	LinearRegression	0	6.1	0.36	1.33	0.36	2.7	13.5
'10/7/2'	BayesianRidge	2	26178.29	0.12	126.53	0.12	155.0	415.8
'10/3/2'	BayesianRidge	2	110.47	-0.11	4.21	0.0	7.0	7.4
'10/3/1'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	4.0
'10/1/1'	BayesianRidge	2	37.29	0.01	4.39	0.05	8.0	46.9
'10/1/2'	BayesianRidge	2	1486.99	-0.15	30.5	-0.13	48.7	95.1
'10/7/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	108.0
'10/2/2'	LinearRegression	1	449.48	0.15	17.36	0.16	28.4	119.1
'10/2/1'	BayesianRidge	2	393.61	0.02	16.36	0.03	25.2	94.9
'10/6/2'	SVR	5	305.37	0.0	9.74	0.0	12.0	77.5

TABLE F.2: Peak Algorithm Results

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'10/6/1'	BayesianRidge	2	962.33	0.01	23.75	0.03	31.5	117.2
'10/5/1'	BayesianRidge	2	1494.25	0.07	28.52	0.07	40.8	101.5
'10/5/2'	LinearRegression	0	0.03	-0.41	0.08	-0.3	0.2	15.0
'10/4/1'	BayesianRidge	2	47.74	0.0	2.59	0.0	7.0	81.6
'10/4/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	79.0
'11/5/2'	LinearRegression	1	386.34	0.41	14.56	0.41	36.9	135.7
'11/2/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	14.0
'11/2/1'	BayesianRidge	2	1150.13	-0.02	25.99	0.02	30.4	87.2
'11/3/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	54.0
'11/3/1'	SVR	5	910.51	-0.01	25.14	0.0	34.7	127.5
'11/1/1'	SVR	5	0.23	-0.1	0.26	-0.07	0.6	59.1
'11/1/2'	SVR	5	2320.75	-0.02	40.35	0.0	55.4	161.3
'11/4/1'	LinearRegression	1	806.19	0.24	19.19	0.26	31.9	107.0
'11/4/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	7.0
'11/5/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	72.0
'12/2/1'	BayesianRidge	2	31.04	0.56	3.68	0.56	9.5	75.6
'12/1/2'	LinearRegression	0	0.0	0.99	0.04	1.0	0.9	9.0
'12/1/1'	BayesianRidge	2	222.5	0.78	7.97	0.78	35.6	89.5
'12/3/1'	LogisticRegression	3	0.0	1.0	0.0	1.0	3.3	52.0
'12/3/2'	SVR	5	479.86	-0.15	16.8	0.0	28.4	98.0
'12/2/2'	BayesianRidge	2	0.76	0.99	0.61	0.99	8.3	73.0
'13/2/1'	BayesianRidge	2	264.43	0.95	9.8	0.95	75.6	162.8
'13/2/2'	SVR	5	4998.17	-0.35	39.77	0.04	63.5	151.5
'13/3/1'	SVR	5	715.75	-0.25	13.32	0.01	32.8	93.0
'13/4/2'	LinearRegression	1	2688.22	0.28	36.08	0.28	55.2	115.1
'13/4/1'	LinearRegression	0	19428.37	0.12	92.42	0.12	154.7	347.3
'13/5/2'	BayesianRidge	2	186.23	0.92	10.34	0.93	52.3	130.0
'13/5/1'	BayesianRidge	2	840.18	0.56	19.68	0.66	56.2	128.3
'13/6/1'	BayesianRidge	2	5.41	1.0	1.81	1.0	40.9	123.2
'13/6/2'	LinearRegression	0	1078.78	0.75	21.13	0.76	60.3	162.6
'13/1/2'	BayesianRidge	2	343.99	0.89	9.68	0.89	58.9	135.3
'13/1/1'	LinearRegression	0	513.27	0.83	12.81	0.84	57.1	138.2
'13/3/2'	BayesianRidge	2	72.99	0.85	4.84	0.86	23.3	100.7
'14/7/2'	LinearRegression	1	2018.56	0.25	33.01	0.27	57.0	166.5
'14/7/1'	BayesianRidge	2	8306.51	0.06	71.94	0.06	98.3	283.0
'14/6/2'	LinearRegression	0	1120.8	0.37	25.32	0.38	43.9	131.1

Table F.2: Peak Algorithm Results

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'14/6/1'	BayesianRidge	2	2832.11	-0.12	39.32	-0.12	53.4	149.4
'14/5/1'	SVR	5	2508.62	0.0	35.81	0.0	59.9	143.8
'14/5/2'	LinearRegression	1	694.7	0.43	19.98	0.44	36.3	112.7
'14/4/1'	LinearRegression	0	319.55	0.03	14.26	0.03	17.6	172.7
'14/4/2'	SVR	5	1651.26	0.0	31.44	0.0	43.0	244.7
'14/3/2'	BayesianRidge	2	1982.82	0.18	31.76	0.2	55.4	180.0
'14/3/1'	BayesianRidge	2	4822.19	0.1	53.87	0.1	78.2	224.5
'14/9/1'	BayesianRidge	2	1338.35	0.12	24.88	0.13	36.8	102.1
'14/9/2'	BayesianRidge	2	74.03	0.57	5.73	0.57	13.3	70.5
'14/8/1'	BayesianRidge	2	853.13	0.49	18.07	0.53	49.2	134.5
'14/8/2'	BayesianRidge	2	1517.53	-0.05	28.93	-0.05	39.9	153.6
'14/2/2'	SVR	5	3637.97	-0.2	46.58	0.0	52.2	256.9
'14/2/1'	BayesianRidge	2	841.23	0.24	22.91	0.24	36.7	220.0
'14/1/1'	BayesianRidge	2	3121.18	0.03	41.71	0.05	59.2	188.1
'14/1/2'	SVR	5	4414.31	-0.08	45.32	0.0	66.4	218.4
'14/10/2'	BayesianRidge	2	153.95	0.2	9.91	0.24	15.3	82.2
'14/10/1'	BayesianRidge	2	8619.12	0.09	68.55	0.09	105.9	241.8
'15/12/2'	LinearRegression	0	1397.04	0.48	29.18	0.49	59.9	183.1
'15/12/1'	BayesianRidge	2	2605.66	0.06	36.05	0.06	53.8	170.6
'15/11/1'	BayesianRidge	2	986.06	0.18	25.16	0.22	41.9	102.7
'15/11/2'	BayesianRidge	2	1615.76	0.24	29.65	0.28	52.0	178.0
'15/8/1'	SVR	5	6710.56	-0.08	63.91	0.0	80.2	236.9
'15/8/2'	BayesianRidge	2	9299.59	0.19	76.86	0.2	109.1	287.8
'15/6/1'	SVR	5	18875.94	-0.01	112.29	0.0	137.0	423.3
'15/9/1'	SVR	5	4893.24	-0.04	60.25	0.0	69.1	229.2
'15/9/2'	BayesianRidge	2	518.87	0.31	17.29	0.31	29.5	112.3
'15/6/2'	SVR	5	11790.72	-0.2	74.83	0.0	107.5	329.2
'15/7/2'	BayesianRidge	2	43873.2	0.08	189.01	0.1	223.5	629.1
'15/7/1'	SVR	5	14110.85	-0.02	101.48	0.0	124.2	383.8
'15/4/1'	BayesianRidge	2	19291.04	0.16	111.82	0.17	151.3	390.5
'15/4/2'	SVR	5	14819.43	-0.06	104.72	0.0	125.5	331.7
'15/5/1'	SVR	5	13406.8	-0.04	105.19	0.0	121.5	338.4
'15/5/2'	BayesianRidge	2	442.05	0.16	15.24	0.21	26.1	94.7
'15/2/2'	BayesianRidge	2	6417.07	0.03	59.37	0.09	84.3	231.8
'15/2/1'	LinearRegression	1	30545.62	-0.05	117.51	-0.05	176.3	314.4
'15/3/2'	SVR	5	16859.5	0.0	97.93	0.0	133.7	439.2

Table F.2: Peak Algorithm Results

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'15/3/1'	SVR	5	9847.74	-0.07	66.9	0.0	98.9	201.2
'15/1/1'	BayesianRidge	2	2569.93	0.43	42.5	0.44	73.9	216.4
'15/1/2'	BayesianRidge	2	9272.46	0.13	71.23	0.13	108.8	312.7
'15/10/1'	SVR	5	7262.82	-0.12	65.0	0.0	72.4	171.1
'15/10/2'	LinearRegression	0	247.42	0.81	10.6	0.81	35.1	117.4
'16/3/1'	LinearRegression	1	35.4	0.24	3.59	0.25	7.2	23.0
'16/3/2'	SVR	5	771.03	-0.02	20.72	0.03	31.0	125.0
'16/2/1'	BayesianRidge	2	4474.33	0.26	53.19	0.27	88.5	255.5
'16/5/1'	BayesianRidge	2	733.04	0.06	19.98	0.15	26.5	167.4
'16/1/2'	BayesianRidge	2	8751.7	0.58	73.11	0.58	151.9	440.8
'16/6/1'	BayesianRidge	2	768.79	0.19	16.23	0.25	31.3	77.1
'16/2/2'	SVR	5	22855.13	-0.03	124.88	0.0	159.2	561.1
'16/7/1'	SVR	5	1050.46	0.02	26.68	0.02	33.0	133.0
'16/5/2'	SVR	5	1366.93	-0.03	27.92	0.0	36.1	141.1
'16/4/2'	BayesianRidge	2	14.13	0.38	2.54	0.42	5.0	33.4
'16/4/1'	BayesianRidge	2	145.34	0.58	9.8	0.58	19.6	63.1
'16/7/2'	SVR	5	666.72	-0.02	18.42	0.02	27.7	113.5
'16/6/2'	LinearRegression	0	788.33	0.26	22.23	0.27	36.7	113.6
'16/1/1'	BayesianRidge	2	21168.51	0.18	113.23	0.29	168.4	498.0
'17/5/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	30.0
'17/5/1'	BayesianRidge	2	10690.59	0.18	78.64	0.18	117.2	341.6
'17/6/1'	SVR	5	18007.68	-0.03	102.41	0.0	142.8	449.8
'17/6/2'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	73.0
'17/7/1'	SVR	5	2675.22	-0.12	35.78	0.0	65.6	133.9
'17/7/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	34.0
'17/1/2'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	12.0
'17/1/1'	BayesianRidge	2	2062.34	0.09	29.53	0.11	49.9	181.0
'17/2/1'	BayesianRidge	2	3848.1	0.1	42.02	0.11	67.1	240.0
'17/2/2'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	16.0
'17/3/1'	BayesianRidge	2	4750.44	0.0	50.84	0.01	63.2	144.8
'17/3/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	39.0
'17/4/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	27.0
'18/3/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	36.0
'18/2/2'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	72.0
'18/8/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	52.0
'18/1/2'	LinearRegression	0	0.0	1.0	0.0	1.0	0.0	32.0

Table F.2: Peak Algorithm Results

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'18/7/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	15.0
'18/6/2'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	14.0
'18/5/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	39.0
'18/4/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	31.0
'19/2/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	16.0
'19/3/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	10.0
'19/4/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	16.0
'19/1/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	10.0
'20/2/2'	LinearRegression	0	0.0	1.0	0.0	1.0	0.0	22.0
'20/3/2'	LinearRegression	0	0.0	1.0	0.0	1.0	0.0	22.0
'20/1/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	13.0
'20/6/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	15.0
'20/4/2'	LinearRegression	0	0.0	1.0	0.0	1.0	0.0	33.0
'20/5/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	11.0
'25/1/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	25.0
'25/4/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	44.0
'25/3/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	7.0
'25/2/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	25.0
'26/1/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	48.0
'26/2/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	6.0
'26/3/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	48.0
'27/5/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	12.0
'27/4/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	20.0
'27/7/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	13.0
'27/6/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	7.0
'27/9/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	11.0
'27/8/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	15.0
'27/1/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	44.0
'27/2/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	10.0
'27/10/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	34.0
'27/3/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	14.0
'28/2/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	19.0
'28/1/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	57.0
'28/6/2'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	25.0
'28/7/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	16.0
'28/4/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	30.0

Table F.2: Peak Algorithm Results

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'28/5/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	30.0
'28/3/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	30.0
'28/8/2'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	32.0
'3/3/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	63.0
'3/2/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	28.0
'30/12/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	23.0
'30/16/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	31.0
'30/14/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	24.0
'30/14/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	24.0
'30/15/1'	LinearRegression	0	0.0	1.0	0.0	1.0	0.0	35.0
'30/5/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	42.0
'30/4/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	42.0
'30/7/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	25.0
'30/6/1'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	16.0
'30/1/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	82.0
'30/2/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	24.0
'30/8/1'	LinearRegression	0	0.0	1.0	0.0	1.0	0.0	24.0
'30/9/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	62.0
'30/13/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	16.0
'30/13/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	16.0
'30/10/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	19.0
'30/11/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	11.0
'31/3/1'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	16.0
'31/4/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	28.0
'34/7/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	16.0
'34/8/2'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	22.0
'35/2/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	11.0
'35/13/2'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	19.0
'35/3/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	26.0
'35/12/1'	SVR	6	6460.76	0.0	68.45	0.0	83.6	259.7
'35/12/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	18.0
'35/4/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	13.0
'35/5/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	10.0
'35/5/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	10.0
'35/6/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	6.0
'35/7/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	7.0

Table F.2: Peak Algorithm Results

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'35/15/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	14.0
'35/7/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	7.0
'35/14/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	10.0
'35/11/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	15.0
'35/10/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	11.0
'35/8/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	5.0
'35/9/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	13.0
'35/1/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	21.0
'36/4/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	55.0
'39/2/1'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	30.0
'39/3/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	15.0
'4/8/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	7.0
'4/9/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	18.0
'4/9/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	18.0
'40/3/1'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	30.0
'40/4/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	15.0
'45/1/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	32.0
'46/4/1'	LinearRegression	0	0.0	1.0	0.0	1.0	0.0	11.0
'46/5/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	25.0
'46/3/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	19.0
'5/2/1'	LinearRegression	0	0.0	1.0	0.0	1.0	0.0	45.0
'5/3/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	14.0
'6/5/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	29.0
'6/6/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	7.0
'6/7/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	22.0
'7/11/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	18.0
'7/12/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	22.0
'7/13/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	14.0
'7/14/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	14.0
'7/6/2'	LinearRegression	0	0.0	0.0	0.02	0.0	0.0	58.0
'7/5/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	4.0
'7/4/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	36.0
'7/4/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	36.0
'7/3/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	22.0
'7/2/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	14.0
'7/1/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	43.0

Table F.2: Peak Algorithm Results

Location	Algorithm	P	MSE	R2	EVS	MAE	STD	Q
'8/5/1'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	7.0
'8/2/1'	LinearRegression	0	0.0	0.0	0.0	1.0	0.0	58.0
'8/3/2'	LinearRegression	0	0.0	0.0	0.0	0.0	0.0	36.0
'9/10/1'	BayesianRidge	2	419.8	0.11	15.59	0.11	20.6	60.8
'9/3/2'	BayesianRidge	2	30.17	0.73	1.85	0.74	8.5	37.0