Tampereen yliopisto

Konsta Peltola

ANALYSING THE EFFECTS OF WEATHER TO REGIONAL BUS TRAFFIC IN TAMPERE REGION

Faculty of Information Technology and Communication Sciences
M. Sc. Thesis
April 2019

ABSTRACT

Konsta Peltola: Analysing the effects of weather to regional bus traffic in Tampere region
M. Sc. Thesis
Tampere University
MDP in Software Development

April 2019

The topic of this thesis is to study how weather is affecting regional bus traffic in Tampere region. Study focuses on evaluating the direct effects of air temperature and snowfall. Delays of regional bus traffic were evaluated by using real time location data and timetable schedule data from regional buses in Tampere region. Based on this information, link travel times were counted for each stop-to-stop trip on the route of the bus. This approach also enabled the evaluation of between which stops the biggest delays occur and during which time of the day. The weather conditions were evaluated by using open weather observation data from Finnish Meteorological Institute and different ranges of air temperature and precipitation were categorized. Link travel times during good driving conditions were used as a baseline when compared to link travel times observed during more challenging conditions. Effect of weather was also evaluated during different times of the day and week.

It is expected that there is a correlation between adverse weather and longer link travel times in bus traffic.

- Hypothesis 1a: Adverse weather causes longer link travel times for the bus traffic
- Hypothesis 1b: Effects of adverse weather to link travel times are more severe in specific stop-to-stop links

First objective is to evaluate Hypothesis 1a and to find out how clear and big the correlation between severe weather and changes in bus link travel times are. Objective of the study is also to try to find out which components of the weather are affecting the link travel times more (air temperature or precipitation).

If hypothesis 1a is proven to be true, the next object will be to try to examine if delays are affected by weather more in some specific places or areas of the region.

Key words and terms: weather, congestion, probe vehicle network.

Contents

Contents ... iii
1. Introduction ... 1
2. Measuring road traffic ... 2
   2.1. Automated traffic sensor networks ... 3
      2.1.1. Fixed sensors ... 3
      2.1.2. Mobile sensing ... 4
      2.1.3. Vehicular sensors ... 5
   2.2. Buses as probe vehicle network ... 6
3. Previous work discussing the effects of weather to traffic ... 8
   3.1. Effects in urban areas ... 9
   3.2. Effects in non-urban areas ... 10
   3.3. Evaluating the results of previous work ... 12
   3.4. Indirect effects of weather ... 14
4. Data sources used in this study ... 15
   4.1. Bus data ... 15
      4.1.1. Tampere public transport SIRI interface ... 16
      4.1.2. General transit feed specification (GTFS) ... 17
      4.1.3. Journeys API ... 18
   4.2. Weather data ... 20
   4.3. Map data ... 21
5. Technologies used in this study ... 22
   5.1. Data collection and pre-processing ... 22
   5.2. Database ... 22
   5.3. Data processing ... 22
   5.4. Data presentation ... 23
6. Data collection and analysation system ... 23
   6.1. Data collection and pre-processing ... 23
      6.1.1. Bus data ... 24
      6.1.2. Weather data ... 27
   6.2. Database structure ... 28
   6.3. Data processing and presentation ... 29
7. Results ... 31
   7.1. Identifying the recurrent delays during different times of the day on different days of the week ... 32
   7.2. Identifying the weather categories ... 35
   7.3. Overall effects of air temperature to link travel times ... 37

7.4. Overall effects of snowfall to link travel times 39
7.5. Links where the effects of snowfall to link travel times are the biggest 41
8. Conclusions 45
References 46

1. Introduction

The ever-increasing amount of traffic and resulting congestion in cities have large social and economic impacts in modern world. Lots of valuable time and fuel is wasted while people are standing in traffic jams and predicting of traveling times is harder. Therefore, congestion causes stress for the commuters, increases air pollution and increases the wear and tear of vehicles. The ways to tackle the issues are either to increase the capacity of the traffic system, decrease the amount of traffic, or to make the traffic flow more efficient on the existing traffic network. Increasing the traffic system capacity, especially in cities, is expensive and transportation infrastructure often takes up a lot of the space that otherwise could be used for something else. So as a result, there is a great need for solving the congestions issues by reducing the demand and by using the existing traffic system more efficiently.

This is where the public transportation and modern intelligent traffic systems (ITS) step in. Public transportation increases the amount of people that can be moved through the traffic system. It is easy to understand that 30 people and their private cars take up much more room on the streets and on the parking lots than 30 people travelling with a regional bus or with a tram. Intelligent traffic systems can both increase the throughput of the traffic network and encourage people to use correct forms of public transportation on correct time of the day, depending on their destination. As we know, most of the issues with traffic systems come up during the rush hours so one of the best solutions to traffic congestion is to try to make rush hour peaks less steep. Additionally, if the passenger information systems, which are part of the intelligent traffic system, succeed in making the travel of passengers in public transportation easy and fast, passengers are much more likely to use the public transport also in the future. [Bacon *et al.*, 2011]

During recent years, the increasing worry about the climate warming has also become one of the key factors why we should improve our public transportation systems and encourage people to use the services of public transportation more. The key factor in achieving this goal is to develop more intelligent traffic systems which encourage people to make the more ecological decision of using public transportation instead of commuting with their private cars.

The most common effects of congestion are experienced during recurrent delays caused by rush hour traffic, but weather is known to be one of the reasons which might cause non-recurrent congestion and issues in the traffic network. Research by Shrank and Lomax [2001] suggests that adverse weather causes 27% of all non-recurrent traffic congestion and according to the Highway Capacity Manual [Transportation Research Board, 2000], rainfall reduces speeds on highways by 2 – 17% and snowfall by 3 – 35% depending on the intensity of precipitation.

The effect of weather can be direct, by making people drive more slowly, or indirect, e.g. by people choosing different means of transportation or by increasing the amount of accidents which can then cause even more congestion. According to Andrey *et al.* [2001], precipitation increases the risk of accidents from 50 to 100 percent. The direct effect of weather to delays in regional bus traffic is what this study is focusing on trying to evaluate. This objective is pursued by recording bus link travel times between consecutive bus stops from live location data provided from Tampere regional buses during different weather circumstances. By comparing the link travel times during different weather circumstances, it is also possible to evaluate if delays occur more often in some specific areas of the road network. Link travel times are also recorded based on the day of the week and time of day. This enables also the evaluation of the day of the week or time of day when the biggest delays occur.

Paula Syrjärinne made her doctoral thesis in 2016 about analysis of traffic data from buses in Tampere region. During the writing of her thesis, Syrjärinne acted also as a supervisor for my bachelor's thesis. This is one of the reasons why I also became interested in the topic of bus location data. Combining bus travel time data and weather data was one of the suggestions for future research topics by Syrjärinne. Therefore, her doctoral thesis and its results are in multiple occasions acting as a base for this study.

Next, in Chapter 2 I explain the basics and scientific background of how traffic can be measured and monitored. After that, in Chapter 3 I review the previous scientific work which has been examining the relation between weather circumstances and road traffic. In Chapter 4 I describe the sources of bus location and weather observation data which were used in this study. Chapter 5 gives an overview of the software development technologies used in this study. Chapter 6 focuses on describing the automated data collection and analysation system which was developed during this study. In Chapter 7 I present the results of the study and finally in Chapter 8 I evaluate the overall success and value of the study.

2. Measuring road traffic

There are multiple ways of measuring traffic amounts and fluency and each of these ways has its strengths and weaknesses. The oldest way, which is still being used today, is by using manual calculations from personnel counting traffic amounts on the ground [Toth *et al.,* 2013]. This is of course quite an accurate way of counting the traffic amounts and gives also the possibility of counting many different types of traffic, for example pedestrians and cyclists, but there is a possibility for human error in the counting. Also, it is easy to understand that in order to get 24/7 real-time information about the status of the traffic in the entire city or area, automated sensors and systems have to be developed instead of relying on manual counting.

2.1. Automated traffic sensor networks

Today, there are multiple different technologies which can be used to automatically measure traffic. Biggest difference comes from the placement of the measurement devices, as they can be fixed and installed to the road infrastructure (e.g. on the side of the road), they can be placed onboard the vehicle itself or they can be carried by passengers who are using the traffic network. Next, I will explain some of the most common systems in more detail.

2.1.1. Fixed sensors

Traditionally automated traffic sensor systems have been based on fixed sensors. Today there are multiple different kinds of fixed sensors which are using different technologies to measure the traffic.

Most common type of fixed sensors are inductive loops which are installed to the road pavement. Single inductive loop can only detect presence and size of a car on one lane from the change the car makes to the electro-magnetic field of the loop. However, when this kind of loops are installed as a pair, it is possible to determine also the direction and speed of the vehicle. However, it is not possible to identify a specific vehicle from a group of vehicles with a loop detector, so for example counting of link travel times for a specific vehicle is not possible with loop detectors. In Finland, inductive loops are the common way to detect if there is a vehicle approaching or already waiting in the intersection. Often information about different intersections is also combined so traffic lights in different intersections work in cooperation with each other to improve the performance of the traffic system. Inductive loops are quite expensive to install and difficult to maintain, as repairs can't be made without disrupting the traffic. Also, inductive loops might not detect motorcycles as their magnetic mass is much smaller. In addition to usage of inductive loops to detect vehicles in traffic light controlled intersections, loop detectors are often used in monitoring traffic speeds and volumes on freeways. Also, some implementations of automatic traffic and speed monitoring by the police can be based on inductive loops. [Bacon *et al.*, 2011]

One alternative for inductive loops is detecting traffic volumes from traffic cameras. With a single camera, it is possible to detect speed and direction of vehicles from multiple lanes [Bacon *et al.*, 2011]. Camera can also detect the size of the vehicle and with multiple cameras and automatic number plate recognition system it is also possible to count travel times for individual vehicles between two points in the traffic system [Tsapakis *et al.,* 2013]. Camera based systems are easier to maintain than inductive loops, but they are more vulnerable for external circumstances e.g. snow blocking the lens.

Another commonly used type of fixed sensor is Remote Traffic Microwave Sensor (RTMS). RTMS sensors are mounted on the side of the road and therefore also easier to install and maintain without disrupting the traffic than loop detectors. A RTMS sensor is transmitting microwave beams towards the road and can detect movement from the beams which are reflected back to the sensor from objects within the target area. This enables the RTMS sensor to monitor traffic volume, occupancy and speeds from multiple lanes at the same time. According to study by Yu and Prevedouros [2013] the speed measurement accuracy of RTMS can be up to 95% and therefore is higher than the accuracy of a single loop detector. [Ma et al. 2015]

All types of fixed traffic sensors have some limitations in the information they can provide. As the name already says, the sensors are fixed and therefore can provide information only from the roads where sensors have been installed. Sensors of this kind are quite expensive to install and therefore the coverage of the sensor network is not often very extensive. The high price is also the reason why fixed sensors are more often used for monitoring of highway traffic instead of traffic flow on urban areas.

For example in Finland, Finnish Transport Agency has about 500 active fixed automatic traffic measuring stations, also known as LAM (Liikenteen Automattinen Mittausasema) -stations. The LAM -station consists of two inductive loops installed to a single road lane. With this setup, it is possible to detect the direction, speed and length/size of the vehicle. Most of the sensors are placed on highways and the information from LAM stations is open to public but wasn't used in this study. [Finnish Transport Agency, 2018]

The inductive loop sensor data from some of the intersections in Tampere region is also available to the public. From this information it is possible to get the traffic volumes and length of the queues on intersections at a given time. However, data from these sensors wasn't used in this study, as it doesn't give information about the effects of weather to travel times. [ITS factory, 2018]

2.1.2. Mobile sensing

Mobile devices that are carried by the passengers themselves can also be used as data sources for traffic monitoring. Different levels of mobile sensing have been identified, based on the involvement of the person carrying the mobile device. Some systems require manual input from the passenger, some require special device or on smartphone special permissions or software to be installed on the device and some systems might work even without the user of the device necessarily knowing about it.

This kind of mobile sensing is usually based on the GPS/GNSS receivers of the passenger's mobile device. However, some systems might also use accelerometers or WLAN or cellular radios of the device to better determine the location of the device. [Syrjärinne, 2016]

Good example of an application that uses also manual input from the users is mobile application called Waze. The user selects the address where he or she wants to drive to and the application will suggest the fastest route based on traffic status data provided by other users of the application. Some of the data, e.g. driving time and route, is reported automatically, but the user can also manually add information about road incidents, e.g. an accident, or even police speed traps, so other users can avoid using the affected route. [Waze, 2019]

Good example of an application that can collect data even without the average user even knowing about it is Google Maps, as traffic status information shown in Google Maps is based on data from mobile sensors (usually smartphones) carried by the passengers. This kind of data collection is called crowdsensing. Many of the people who carry their smartphones might not even know that they are sharing their location data to instances like Google. Therefore, the justification of this kind of data collection has often been questioned. However, the data which is available by crowdsensing gives much more detailed and extensive data regarding the status of the road network when compared to other types of traffic sensor networks. [Bacon *et al.*, 2011]

Mobile sensing is in many ways better for monitoring of traffic in urban areas than fixed sensors. Mobile sensing is relatively new and cheap way of measuring traffic as nowadays most people are carrying smartphones with them wherever they go. However, there are still some open questions regarding the privacy of location data which is being collected from personal devices of passengers without them necessarily knowing about it.

2.1.3. Vehicular sensors

Real time location data can also be collected from wireless devices which are installed directly to the vehicles. Today most of new cars contain devices of this kind already from the factory, but usually the location information collected from these devices is not available to the public, the infrastructure or to other vehicles. The vision for the future is that the vehicles and the infrastructure could communicate with each other, in a standardized manner, to find the most efficient ways to direct traffic through the entire traffic network as fluently as possible [Parrando and Donoso, 2015].

As the data from sensors which are installed to cars on the factory is not available, separate wireless real-time location devices have been installed to vehicles to detect and report the status and location of the vehicle in traffic network. Often devices of this kind are installed to cars owned by companies, e.g. buses, taxis, ambulances and delivery vehicles. Therefore, these location transmitting devices are usually developed for different use cases than monitoring of the traffic flow. Vehicles which have been equipped with real time location devices are usually called probe vehicles or floating cars and together they are called a probe vehicle fleet. Most common use cases for vehicular sensors are e.g. if the vehicle gets stolen or taxi driver sends an emergency signal and therefore the update frequency of location data is often fairly low (e.g. once per minute). Some studies have tried to tackle the issue of using low frequency location data for traffic monitoring by developing models to estimate travel times for links also when distance between location pings is longer than the travel link itself [Jenelius and Koutsopoulos, 2013]. Data which has been collected from different kinds of probe vehicles also have their own characteristics which need to be considered when analysing the data. For example, taxis and buses often have special lanes is cities, so location data gathered from these vehicles might not actually tell the whole story about status of traffic for private vehicles. Therefore, it is important to evaluate the assumptions about what can actually be evaluated based on the data that has been gathered from the probe vehicle fleet.

For different uses cases of the data which has been gathered from vehicular sensors, different meta data would also improve the usability of the data. For example, for traffic flow research it would be good to not only know the location of the vehicle but also how many people there are onboard. This would enable for example the studying of which bus stops are most popular during which times of the day. Many buses have this kind of passenger counting systems installed but at least from Tampere, the passenger counting information is not available to the public. Also, to be able to detect the arrivals and departures to bus stops from bus location data more precisely, data from the opening and closing of the bus's doors would be useful.

There have been studies about using taxi fleet as a probe vehicle network but data from taxis has its own specialities. Taxis are moving freely around the road network and they are not following predefined routes or schedules. This makes the coverage of probe vehicle data recorded from taxis very extensive but harder to analyse, for example when compared to data collected from bus fleet. Often the biggest issue with location data collected from taxis has been the low frequency of data. Also, if it is not possible to see from the data during which times the taxi had a customer, it is not easy to detect only from the location data of the taxi whether it is waiting for a customer or standing in a traffic jam with customer onboard. [Kuhns *et al*., 2011] [Jenelius and Koutsopoulos, 2013]

There have been lot of studies where regional buses have been fitted with devices which are sending real time location data. This kind of location data has also been used in this study and therefore the usage of buses as probe vehicles has been discussed in more detail in chapter 2.2

2.2. Buses as probe vehicle network

Buses as probe vehicles also have their own characteristics. They are following a predefined route and a schedule, and this makes it easy to observe if bus is running late from schedule. On the other hand, because buses are following a route, probe vehicle data from buses might not cover as much of the road network as data from taxis or private cars could. Also, some metadata is needed in order to know which bus is running on which route.

Syrjärinne [2016] has already studied real time bus location data from same real-time interface which was used in this study. Based on the location data, she monitored the performance of the public transport and the traffic fluency in overall.

To detect the combination of time, location and bus line which are regularly associated with delays, Syrjärinne applied frequent item set mining to the bus location data. Based on this information a concept of adaptive, data driven bus schedules was introduced which are taking these recurring delays into account when presenting the bus departure times to the user. With this system, the user can download the data driven bus schedule for a bus stop of his or her choice. Data driven bus schedule presents the forecasted departure time of each bus which is leaving from the selected bus stop, based on the historical data from recurring delays. User is also informed with an estimated accuracy of the forecast based on the variance of the historical data. [Ajoissa pysäkillä, 2015]

Regarding the performance of public transportation, Syrjärinne also examined the effects of traffic light prioritisation for buses which were running late from schedule. The prioritisation system can detect if a bus which is running late from schedule is approaching the intersection and can change the sequence of traffic lights so that the bus doesn't have to wait for the traffic lights to change for as long as it would without the prioritisation. Syrjärinne had recorded bus location data both before and after the traffic light prioritisation system was taken into use in Tampere. The result was that in most of the intersections the effect of prioritisation system was positive. This means that buses which were running late, could better catch up their schedule between the stops.

Syrjärinne also studied how bus location data could be used to monitor overall traffic fluency in the city. For this purpose, she presented the concept of link travel times, which was also used in this study to find the differences in travel times during different weather conditions. In her study, history data of link travel times was used to generate link travel time profiles. From these profiles it was possible to identify the regular delays associated with specific bus lines. Syrjärinne used these link travel time profiles to create a real time incident detection system. By comparing the link travel time profiles and real time location data, the system was able to detect abnormal and significant delays and report them in few minutes after an incident had happened. [Syrjärinne, 2016]

Uno et al. [2009] used bus probe data to evaluation of travel time variability and the level of service of roads. For travel time variability study, they used bus probe data from Hirakata city, Japan. Bus location data wasn't provided in real time but instead it could be uploaded from onboard system of the bus every time when the bus returned from service back to the depot. Therefore, as the frequency of the location data was not limited by the bandwidth of mobile data connection, the frequency of the location data was good (1 second). As there was no metadata attached to the location data, they also had to create a supplemental survey for an investigator who then recorded the positions of the bus stops along the route. They also recorded the travel times for same routes with a private car. After subtracting the time spent on bus stops from the travel times of the bus, their result was that the travel time of the bus was nearly the same as the travel time with a private car.

For the level of service study, Uno et al. [2009] evaluated two things: The average travel time spent on distance of 1 kilometre and the variation of the total travel time of the route. On the area, from which they had collected their bus probe data from, new highways were opened in 2003. This enabled them to compare the results from 2002 and 2003 to see whether the opening of the new highways might have affected the performance of the traffic network. Their result was that the average travel time for 1 kilometre was shorter in 2003 than in 2002 and also the variance of travel times had decreased. So, in overall, they concluded that most probably this improvement in the level of service of the road network was due to the opening of the new highways.

Location data from buses has even been used to detect the cycles of fixed time traffic signals. Fayazi et al. [2015] studied low frequency bus location data in San Francisco, USA, to estimate the cycles in which the city's fixed traffic lights operate. Their study was successful, and this kind of information can be used for example to improve fuel efficiency by telling the bus driver to slow down for red lights even before the light has changed from green.

3. Previous work discussing the effects of weather to traffic

There are a lot of earlier studies regarding the non-recurring effects of adverse weather to the traffic performance. These studies can roughly be divided into two categories: Effects of weather in urban areas and effects of weather in non-urban areas. First, I will examine the effects of adverse weather in urban areas on Chapter 3.1 and then I will examine the effects of adverse weather on more remote areas on Chapter 3.2. The results from research that are easily comparable are examined in Chapter 3.3. I have also shortly examined the previous work regarding the indirect effects of weather to traffic in Chapter 3.4.

Many of the previous studies are more concerned about the maximum traffic flow during different weather circumstances because they are focused on trying to understand the weather-based congestion. Traffic flow means the number of vehicles passing a certain point in the traffic network for example per hour. However, as my study is focusing more on understanding the differences in travel times during different weather circumstances, I have focused on gathering the results regarding the changes in speed, travel time and other variables which are more relevant to this study.

There have also been previous studies regarding the effects of darkness and daylight to traffic, but as the results have been varying and as the counting of this effect is more complex when the length of the day was changing rapidly during the data collection period, this effect hasn't been taken into account in the results of this study. [Transportation Research Board, 2000] [Jägerbrand and Sjöbergh, 2016]

3.1. Effects in urban areas

Tsapakis et al. [2013] studied the effects of rain, snow and air temperature to travel times in the Greater London area by using fixed, camera-based sensors and automatic number plate recognition. They gathered traffic data from three 2-hour periods on weekdays during the morning, afternoon and the evening. Weather data was available from multiple weather stations once per hour. First part of their research was to find out how big radius they could use from weather stations to reliably estimate the amount of precipitation on the road from which they are measuring the travel times. By comparing differences in precipitation amounts between different weather stations, they decided to use a value of 3 kilometres as the radius. This means that only travel links within this radius from nearest weather station were considered as valid for the examination. Result of their research was that the temperature had little or no effect in the travel times. The effect of rain was somewhat related to the amount of rainfall as light rain caused delays from 0.1% to 2.1%, moderate rain delays from 1.5% to 3.8% and heavy rain delays from 4.0% to 6.0%. The effect of snow was also greatly dependent on the amount of snowfall as delay caused by light snow was from 5.5% to 7.6% and by heavy snow from 7.4% up to 11.4%. Tsapakis et al. [2013] also noted that the magnitude of the effect to travel times was smaller in the central London and inner London area than in the outer London area.

Xu *et al.* [2013] examined effects of weather to road network performance in Haizhu district, China. Traffic data was gathered from more than 20,000 taxis with location data sampling rate between 20 and 120 seconds. Additionally, data from fixed sensors in some of the intersections on the selected routes was used. For measurement of weather, they used only one weather station located near the Sun Yat-Sen university. However, the longest distance from the weather station to the road being monitored was only about 2 kilometres. To get comparable results for rainy and non-rainy situations, they recorded data as a set of rainy and non-rainy days from the same day of the week and with maximum difference of two weeks between the two dates. They got clear results that precipitation decreased the performance of the road network. Additionally, they noted that in their research the effect was more significant during the evening rush hour than during the morning.

Smith *et al.* [2004] studied the impact of different intensities of rainfall to urban freeway traffic capacity and operating speeds in Virginia, United States. Traffic data was collected from two urban freeway links with 2 -minute intervals. These two freeway links were selected for investigation because they both were within 4,8 kilometre (3 mile) radius from the weather station which was used for the study. Weather station was located at the Norfolk International Airport and precipitation data was available on an hourly basis. Weather and traffic data from August 1999 to July 2000 was used in the study. One of the main goals of the study was to set exact values for categorizing of light and heavy rainfall as this wasn't always clearly stated in earlier studies which they had examined. Their result for the categorisation was that rainfall was considered light when precipitation was between 0,254 and 6,35 millimetres per hour and heavy when rainfall was more than 6,35 millimetres per hour. Results regarding the effects of rainfall to operating speeds was that light rain caused average operating speed decreases from 4,84% to 6,11% and heavy rain caused operating speed decreases from 5,43% to 6,41%.

3.2. Effects in non-urban areas

Chung [2012] examined effects of rain and snowfall to traffic volumes and speeds on freeways in Korea, based on data from year 2008. According to his results, the effect of rainfall was clearly affected by its intensity. He also concluded that during the whole year, rainfall of less than 1 millimetre per hour caused an average of 1071 vehicle hours of non-recurrent congestion per freeway kilometre while rainfall of more than 30 millimetres per hour caused an average of 2282 vehicle hours of non-recurrent congestion per freeway kilometre. He also noticed that the effect of rainfall to non-recurrent congestion was biggest during the afternoon peak hours (16:30 – 20:00) and smallest during the night (20:00 – 07:00). Chung applied the same logic into studying of the effects of snowfall to traffic volumes and speeds. The effect of snowfall during different times of the day was nearly the same as with rainfall but the effect to average vehicle hours of non-recurrent congestion per kilometre was bit different with snowfall. The average non-recurrent congestion was rising only until the amount of snowfall exceeded 25 millimetres per hour. After that the average non-recurrent congestion per kilometre started to decrease. According to Chung, this was most likely due to reduced traffic demand and suggest that people will start to avoid travelling once the snowfall gets so heavy. The amounts of precipitation observed by Chung are very extreme when compared to precipitation amount observed in this study.

Ibrahim and Hall [1994] studied the effects of adverse weather to freeway traffic in Mississauga, Ontario. Traffic data was collected from two double loop detectors with 30 second frequency and weather observations were retrieved from weather station located in Pearson International Airport. To get more data including both rainfall and snow, data from October, November and December 1990 and January and February 1991 were selected. When analysing the traffic data on clear, rainy and snowy days, Ibrahim and Hall noticed that the differences between days with similar weather conditions were smaller during clear weather conditions than during rainy and snowy conditions. Therefore, they decided to use data from days with clear weather as one base dataset and compare that to different rainy and snowy days to get the minimum and maximum effects of rainfall and snowfall. Their result was that during the maximum observed traffic flow, heavy rain caused speed decrease of 10% and heavy snow decrease of 45% in traffic speeds.

Kyte et al. [2001] investigated the effects of adverse weather conditions to free-flow speeds on rural Interstate freeways in south-east Idaho. Free-flow speed is defined as the mean speed of passenger cars in low to moderate traffic volumes in prevailing traffic conditions. As a difference to previous studies, they examined also the effects of fog and high winds in addition to wet and snowy pavement. Another special feature of this study is that the data has been gathered from a relatively long period of time compared to other studies as it was collected from 1996 to 2000. Their results suggest that reduced visibility has only minor effect to passenger vehicle speeds until the visibility decreases to less than 300 meters. After this point, the speeds start decreasing significantly. The decreasing of visibility can be caused for example by fog, heavy rainfall or heavy snowfall. According to their results, also high winds seem to start affecting the passenger vehicle speeds after wind speed exceeds 6,7 meters per second. However, high winds also seem to increase the variance of passenger vehicle speeds which could be explained by gusty wind, different types of vehicles reacting differently to wind or different drivers reacting differently to high wind conditions. Regarding the effects of wet or snowy pavement, their result was that wet pavement reduced the free-flow speeds by 9,5 kilometres per hour and snowy pavement reduced free-flow speeds by 16,4 kilometres per hour.

Chin et al. [2004] estimated the yearly weather-related capacity reductions and weather-related delays in the entire USA based on data from 1999. They used estimates for reduction of capacity and speed based on previous studies. During light rain, they estimated that speed reduction would be 10% on all types of roads. During heavy rain, speed reduction was estimated to be between 10% and 25% based on the type of the road. Light snowfall was estimated to reduce speeds by 15% on freeways and 13% on smaller, arterial roads. For heavy snow, speed reduction was estimated to be 38% on freeways and 25% on arterial roads. The result of the estimation was that during the whole year, all weather-related events caused capacity reduction of 20,9 billion vehicles on freeways and principal arterial roads. This reduction in capacity was estimated to have caused 330,1 million vehicle-hours of delays. If these estimates are even remotely correct, it is easy to understand why it is important to try to better understand the effects of adverse weather to traffic performance as so much more time is spent in traffic due to adverse weather.

3.3. Evaluating the results of previous work

As is easily noticeable from previous chapters, there have been much more studies investigating the effects of weather to traffic in freeway settings than in the urban traffic network. This is because most of the studies are based on traffic data from fixed sensors, like loop detectors, which are more often installed to freeways. Also, many of the weather stations used in previous studies have been located for example on airports, which are usually located bit further away from city centres. Traffic performance studies which are using mobile sensing or vehicular sensors are all quite new and not many studies regarding the effects of adverse weather to traffic performance has yet been done based on data from these mobile sensors. This is also one of the main motives for this study.

Results from previous studies which can be considered to have estimated similar conditions regarding the effect of rainfall to traffic speeds and delays, have been collected to Table 1. In overall, the results about the size of the effect differ based on the intensity of the rainfall and the time of day. However, all of the studies have come to a conclusion that rainfall has an effect on the speeds and travel times. Most of the studies also noted that the intensity of the rainfall greatly affects the size of the effect. Some of the studies also took into account the difference in effect on different days of the week and times of the day.

Research	Environment	Amount / Effect
Tsapakis *et al.* [2013] Automatic number plate recognition data from more than 380 travel links	Urban / Greater London, UK	0 – 0,25 mm/h / 0,1-2,1% 0,25 – 6,35 mm/h / 1,5-3,8% > 6,35 mm/h / 4,0-6,0% (increase in travel time depending on time of day)
Akin *et al.* [2011] Remote Traffic Microwave Sensor (RTMS) data from two highway corridors	Urban freeway / Istanbul, Turkey	Speed decrease of 8-12% depending on the measurement point
Xu *et al.* [2013] From GPS receivers installed in over 20 000 taxis	Urban / Guangzhou, Haizhu district	Speed decrease of 4,4-15,6% depending on the time of day
Chung [2012] Weather and traffic data from 2064 freeway links	Freeway / 25 major freeways in Korea	Effect to amount of delay per vehicle-hour per kilometre is dependent on the amount of rainfall. Biggest effect during the

		afternoon peak hours.
Ibrahim and Hall [2004] Traffic data from double loop detectors	Freeway / Mississauga, Ontario, Canada	Light rain caused speed decrease of 2 km/h. Heavy rain caused speed decrease from 5 to 10 km/h.
Kyte *et al.* [2001] Traffic and weather data from road side sensors and loop detectors	Freeway / I-84 south-eastern Idaho, USA	Rainfall caused speed decrease of 9,5 km/h
Smith *et al.* [2004] Traffic data from urban freeways close to weather station located on an airport	Urban / Hampton Roads, Virginia, USA	0,254 - 6,35 mm/h / 4,84% - 6,11% > 6,35 mm/h / 5,43% - 6,41% (decrease in average operating speed compared to situation when rain was less than 0,254 mm/h)

Table 1. Effect of rainfall to traffic from some of the previous studies

Results from previous studies which can be considered to have estimated similar conditions regarding the effect of snowfall to traffic speeds and delays, have been collected to Table 2. Again, all of the studies came to a conclusion that snowfall has an effect on the performance of the traffic system. The effect of the intensity of snowfall, as well as the time of the day, were again mentioned in many of these studies. Bit surprisingly, some of the studies discovered that speeds might even increase in some situations during snowfall. All of the studies that discovered this result were explaining it with reduced traffic amounts. For example, Chung [2012] noticed that when the intensity of the snowfall increased, the traffic speeds were decreasing until the traffic amount started decreasing. After that, traffic speeds started to increase again.

There are some differences between the results of some of the studies as some suggest that even heavy snow fall has smaller effects than heavy rainfall has in some of the other studies. However, if effects of rainfall and snowfall were investigated in the same study, the result was always that with same intensity (low or heavy) snowfall always had bigger effect than rainfall. Bigger effect of the snowfall is easily understandable as snow affects the traction of the pavement more than rain does. Also, the effect of snowfall is most likely affected by the bigger reduction in visibility caused by snowfall. The effect of snowfall is also most likely very dependent on the location and time of the year because if drivers are not used to driving in snowy conditions or the vehicles don’t have winter tyres, the effect will most likely by significantly bigger.

Research	Environment	Amount / Effect
Tsapakis *et al.* [2013] Automatic number plate recognition data from more than 380 travel links	Urban / Greater London, UK	0 – 2,5 mm/h / 5,5-7,6% > 2,5 mm/h / 7,4-11,4% (increase in travel time depending on time of day)
Akin *et al.* [2011] Remote Traffic Microwave Sensor (RTMS) data from two highway corridors	Urban freeway / Istanbul, Turkey	Speed increase of 4-5% most likely due to traffic volume decrease of 65-66%
Chung [2012] Weather and traffic data from 2064 freeway links	Freeway / 25 major freeways in Korea	Effect to speed is dependent on the amount of rainfall, until it snows so much that people start avoiding travelling. Biggest effect during the afternoon peak hours.
Ibrahim and Hall [2004] Traffic data from double loop detectors	Freeway / Mississauga, Ontario, Canada	Light snow caused speed decrease of 3 km/h. Heavy snow caused speed decrease from 38 to 50 km/h.
Kyte *et al.* [2001] Traffic and weather data from road side sensors and loop detectors	Freeway / I-84 south-eastern Idaho, USA	Speed decrease of 16,4 km/h

Table 2. Effect of snow to traffic from some of the previous studies

3.4. Indirect effects of weather

There has also been a lot of studies about how weather affects traffic indirectly. Saneinejad *et al.* [2011] have studied how weather affects commuter’s choices of how to get to work in Toronto. Their results suggest that usage of bicycle decreases when temperature is below 15 ºC and walking becomes less common when temperature drops below 5 ºC. This increases the amount of people going to work by private car or by public transit, which is in the case of buses, competing mostly from the same road resource as private cars.

Singhal et al. [2014] studied effects of weather to subway ridership in New York City during years 2010-2011. They discovered that the weather affects subway ridership much more during the weekends than it does during the work days. They also noticed that in New York City, during a heavy snowfall people are more likely to leave their own cars at home and use the subway instead.

In overall after reading about many studies how weather can affect the traffic indirectly, it is clear that understanding and predicting of indirect effects of weather is not possible in the scope of this study. The indirect effects are also greatly affected by day of the week, time of the day and what alternative means of travel are available and if those means compete from the same road resources. Therefore, I have decided to leave out the evaluation of indirect effects of weather from this study. Of course, all this kind of indirect effects will affect the results of this study. However, I expect the indirect effects of weather to be much smaller than the direct effects, so I assume that the objective of studying the direct effects won't be affected too much by the indirect effects.

4. Data sources used in this study

Both bus location data and weather data are of high quality and are available with high frequency from Tampere region. Therefore, Tampere region is the ideal environment for studying the effects of weather to regional bus traffic. Both traffic and weather data sources are open to public, so no special permissions were needed to conduct this study.

4.1. Bus data

In this study, I have used the real time location data from regional buses operating in Tampere area. As buses have their own characteristics of how they operate in traffic, this data is used only to evaluate the delays of the buses from which the location data has been available. Real time location information from buses operating in Tampere is freely available to the public. Location data of the regional buses is also used by the traffic control system to give priority in traffic light -controlled intersections to buses which are running late from schedule [Syrjärinne, 2016].

City of Tampere has promoted the importance of opening public data for companies and individuals to develop new services and applications. Lot of different datasets and interfaces have been opened to the public, ranging from datasets regarding the building stock of the city to data concerning the availability and expenses of the public healthcare. To empower development of new smart traffic, City of Tampere has established a community called ITS Factory to encourage communication between different parties and empower the development of better ITS solutions in Tampere. This community consists of public organizations, universities, companies and individuals who are working together to create new and better solutions based on open data. Opening of the access to public transportation data creates new opportunities for companies and one of the targets of ITS Factory is to make the solutions scalable and standardized so they could be used anywhere in the world.

All open data provided by the city of Tampere has been licenced with a special licence developed by the city of Tampere. It requires that any use of open data provided by the city of Tampere should be mentioned with a copyright referring to the city of Tampere. It grants the right for the user to copy and share the data commercially or non-commercially and as a part of an application. Additionally, the city of Tampere doesn't take any responsibility regarding the availability or correctness of the data or responsibility regarding any harm caused by the usage or disruptions in the availability of the data. [City of Tampere, 2019]

In this study, regional buses are considered as "probe vehicle network" and no other sensors or data sources were used in this research to collect traffic data. An assumption had to be made that the bus location data and its meta-data and the schedule information, received from the API is valid. However, if the results suggest so, in later phase of the study there might be a need to also evaluate the reliability of the bus location data.

In Tampere, regional buses have different schedule for summer and winter periods. For example, during writing of this thesis the local bus traffic operator, TKL (Tampereen Kaupunkiliikenne Liikelaitos) has published a schedule for winter period which is valid from 22.10.2018 to 02.06.2019. This gives the operator the possibility to make some changes and adjustments to bus schedules for different seasonal weather circumstances, but by splitting the year into only two periods it doesn't allow very accurate adjustments according to seasonal weather. Obviously changing of the schedules more often would require more attention from the passengers as they would have to check the new schedules more often.

4.1.1. Tampere public transport SIRI interface

SIRI (Standard Interface for Real-time Information) is a technical standard developed by CEN (European Committee for Standardization) for communication of real-time information related to public transportation in Europe. However, SIRI interface standard is being used by multiple operators around the world. The SIRI standard contains lot of different features, for example planned timetable data, real-time arrivals and departures of the vehicles to stops and the real-time location of the vehicles [SIRI, 2019].

In Tampere, SIRI interface is provided by the ITS Factory. Only real-time location data of the vehicles is provided through SIRI, and this interface has been available since September 2013. It offers real-time data from all regional buses which have been fitted with GPS transceivers and are currently performing a service (driving on a route with passengers onboard). Update frequency of the real-time location data is once per second. Only current status of the bus fleet is available from the SIRI interface, so it is not possible to get historical data. Therefore, the real time data must be stored by the user in order to analyse any patterns in the behaviour of the buses. According to the standard, data from SIRI interface is available as XML but in Tampere, also JSON variant is available. Currently, the direct usage of Tampere SIRI interface is considered deprecated, as new, more extensive APIs have been developed on top of the data which is available from the SIRI interface. [ITS Factory, 2019]

4.1.2. General transit feed specification (GTFS)

When studying the behaviour of buses, we need to have static information about the operators, routes, departures and locations of the bus stops, in addition to the live location data gathered from bus probe vehicle network, to be able to estimate the delays. From Tampere region, this kind of static timetable information is available in GTFS - format.

General transit feed specification (formerly known as Google Transit Feed Specification) has been developed by Google and it has reached a de facto status in sharing of static transit data. [Ferris *et al.*, 2009]. However, the specification isn't very strict and therefore some implementations may vary from each other. GTFS is usually published as a single zip -file, composed of 6 required and 0-7 optional comma separated text files, describing different attributes of the transit system. Table 3 describes briefly the content of each possible text file, shows which ones are mandatory according to the specification and which are available from Tampere region. [GTFS, 2018]

File	Required	Tampere	Description
agency.txt	X	X	One or more transit agencies that provide the data in this feed
stops.txt	X	X	Individual locations where vehicles pick up or drop off passengers
routes.txt	X	X	Transit routes
trips.txt	X	X	Trips for each route. A trip is a sequence of two or more stops that occurs at specific time.
stop_times.txt	X	X	Times when a vehicle is planned to arrive at and depart from individual stops for each trip
calendar.txt	X	X	Validity period for services and during which weekdays they are performed
calendar_dates.txt		X	Exceptions for services defined in calendar.txt
fare_attributes.txt			Fare information for a transit organization's routes

fare_rules.txt			Rules for applying fare information for a transit organization's routes
shapes.txt		X	Rules for drawing lines on a map to represent a transit organization's routes
frequencies.txt			Headway (time between trips) for routes with variable frequency of service
transfers.txt		X	Rules for making connections at transfer points between routes
feed_info.txt			Additional information about the feed itself, including publisher, version, and expiration information

Table 3. General transit feed specification. [GTFS, 2018]

In Tampere, GTFS -feed is provided by the ITS Factory. The GTFS -feed for the regional buses has only one agency (Tampereen joukkoliikenne) as there are no other operators for regional bus traffic in Tampere. The feed is updated with an interval of one to two months and the size of the feed is relatively small, as the biggest file (stop_times.txt) has about 740 000 lines on the feed which was released 3rd of January 2019. In comparison, the same file in the GTFS -feed containing bus and train routes and timetables from Sweden, has 5,8 million lines.

4.1.3. Journeys API

From Tampere region, all bus data used to conduct this study is currently available from single public API, called Journeys API, which is provided by the ITS Factory. This API combines data from multiple different sources and offers it to developers and clients through HTTP REST API in JSON -format. Different types of objects are available in a hierarchy and API enables the user to perform searches on the data. Live location data of the vehicles is collected from Tampere public transport SIRI interface, described in chapter 4.1.1 and static timetable, route and stop data is collected from the GTFS -feed, described in chapter 4.1.2. The structure of the Journeys API is presented in figure 1.

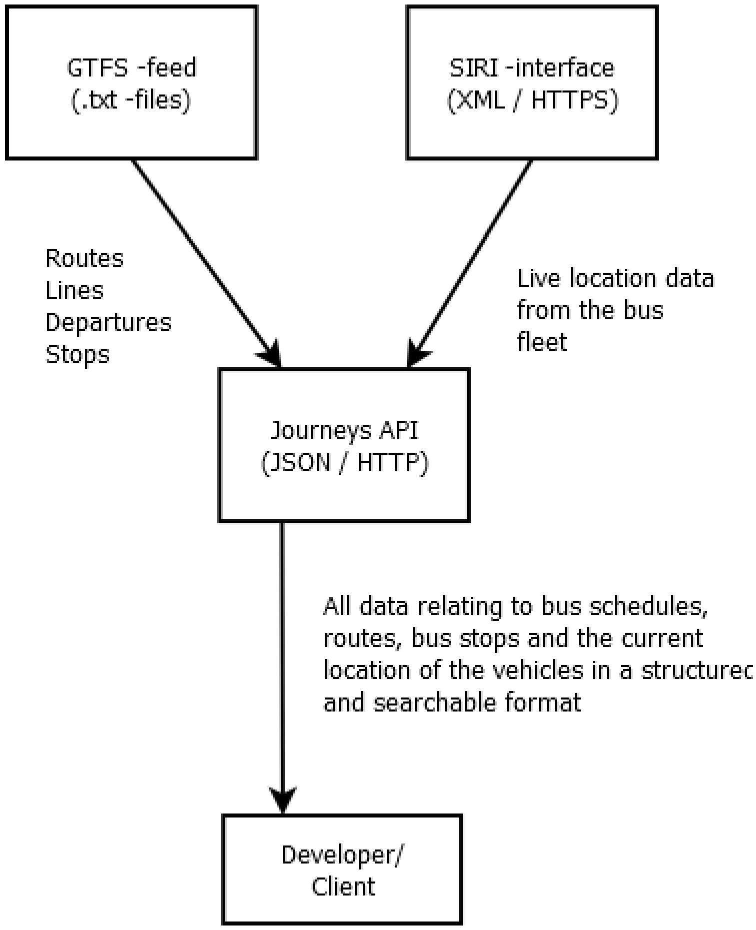

Figure 1. Overview of the Journeys API.

Journeys API combines the data from these two interfaces into single API and maps the timetable and route data directly to the live location data received from the buses. This means that the developer or client doesn't have to perform the complex and laborious task of combining the data from GTFS to live location data from the bus fleet. Access to Journeys API doesn't require any kind of identification and the API is well documented on the ITS Factory website.

In comparison to for example Syrjärinne [2016], who had to manually combine the live bus location data from the SIRI interface and the static timetable and bus stop data from the GTFS -feed, currently it is possible to get all this data from Tampere through this single interface. In the beginning of this study, I also started by manually combining the real-time location data and timetable data. However, after spending some time studying the SIRI and GTFS interfaces and after I heard that they are considered as deprecated, I decided to use the better supported Journeys API instead, as it contains all the same information as the SIRI and GTFS -interfaces. Of course, using of this API adds a possibility of additional noise and faults in the data as it is processed more before it is delivered to the client. However, after using both solutions, I came to a conclusion that using the JourneysAPI was better option as it was easier to use, and I didn't find issues with the quality of the data. [Journeys API, 2019]

4.2. Weather data

Weather data which was used in this study is collected from the open data API of Finnish Meteorological Institute (FMI). Weather observation data from FMI has been opened as a part of Finnish government's open data policy and the EU INSPIRE directive. FMI offers a great collection of different datasets containing for example observation data from, weather stations and road weather stations, water level stations, wave buoys and weather radar images. Also, historical weather observation data since 1959 and also data from national forecast models is available. Data is available in XML format and the access requires registration. Creative Commons Attribution 4.0 International license is used which means that data can be stored and used as long as appropriate credit is given to the provider of data and a link to the license is available for the user. [FMI Open data final report, 2013]

Name of the station	Latitude	Longitude
vt3 Lempäälä Kulju R	61.40841	23.76504
vt12 Tampere R	61.49754	23.89794
vt12 Tre Paasikiventie R	61.50770	23.69954
vt3 Pirkkala R	61.45264	23.63602
vt3 Lempäälä Kulju Opt	61.40841	23.76504
vt12 Tre Paasikiventie Opt	61.50770	23.69954
vt12 Tampere Opt	61.49754	23.89794
vt3 Pirkkala Opt 12	61.45264	23.63602
yt3495 Rautaharkko Opt	61.47738	23.77428
yt3495 Rautaharkko R	61.47738	23.77428

Table 4. Road weather stations which were used in the study and their location.

Observation data from road weather stations in Tampere region were used in this study. The road weather stations are presented in Table 4. Observation data from weather stations in Tampere region would also have been available from Tampere region, but the decision to use only data from road weather stations was made because of following reasons:

- With a radius of 4 kilometres, the observation data from road weather stations covers most of the area under examination.
- Observation data from road weather stations is available more frequently and changes in the intensity of precipitation are available more precisely.

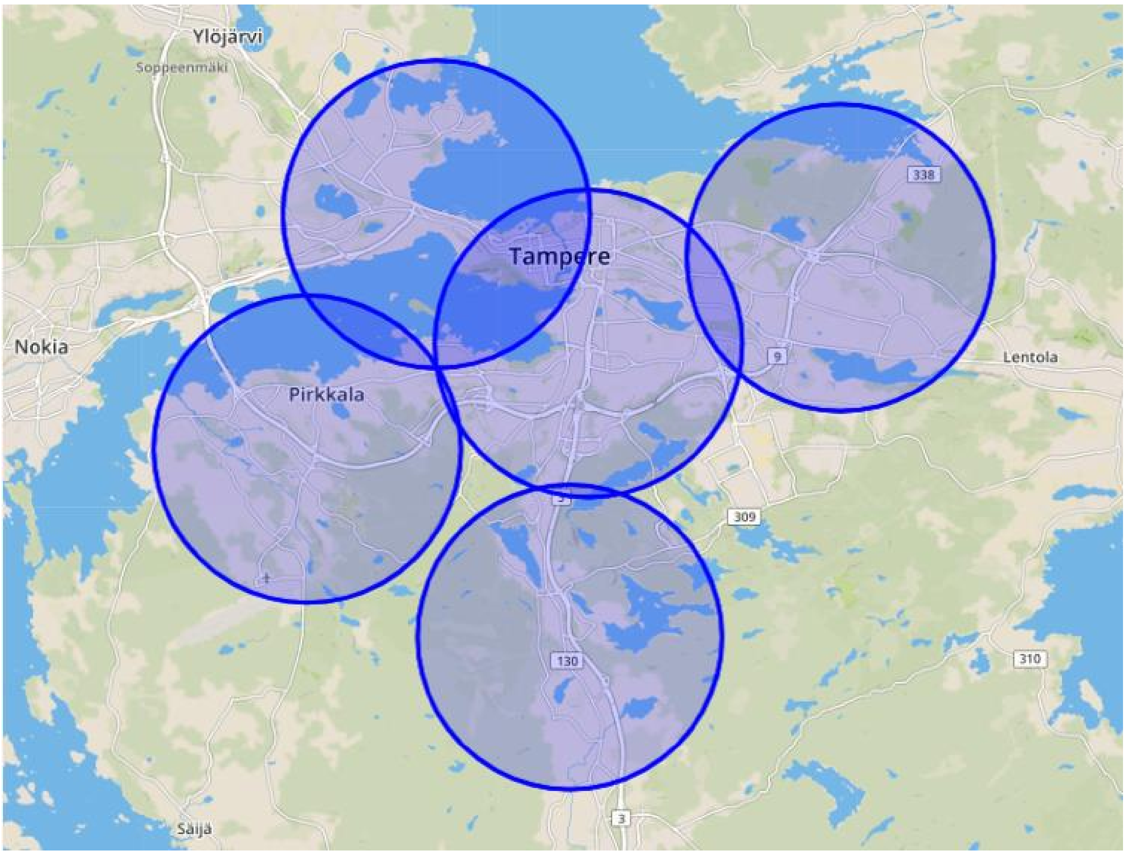

Figure 2. The coverage of road weather stations in Tampere region, with radius of 4 kilometres.

Interval for observation data from road weather stations is 5 minutes and following data was saved to the database from each observation: identification of the road weather station, timestamp of the observation, air temperature, precipitation amount during last 10 minutes, wind speed, gust speed, wind direction, relative humidity, dev point and visibility. However, not all these values were always available from each road weather station. Station identification, timestamp of the observation, air temperature and precipitation amount during last 10 minutes were the only values used in this study.

4.3. Map data

Throughout the entire process of this study, map data was used for debugging of the data collection and presentation system and for visualisation of the results. Map pictures attached to this thesis are screenshots from this map data. An open-source JavaScript library called Leaflet was used to easily visualize data and results on a map. Map data itself was provided by OpenStreetMap -project [2019], which is a collaborative effort aiming to collect and share map data free of charge. MapBox [2019] was used as a hosting service for the map data.

5. Technologies used in this study

Learning to use new software development technologies and improving my software development skills was one of the objectives of this thesis. Therefore, in this chapter I will describe the technologies used to conduct this study and the reasons why I have decided to use these technologies.

5.1. Data collection and pre-processing

The data collection and pre-processing system was implemented by using Java. Java is one of the most popular programming languages and in this case, I decided to use it because I wanted to improve my skills. Additionally, I decided to use Maven as a build automation and dependency management tool because I need it in my current job and wanted to strengthen my knowledge. Also, some external libraries were used, for example log4j for logging. I also added few very simple unit tests, but due to lack of time, I couldn't implement the unit tests that I would have liked to implement for the pre-processing system.

5.2. Database

In the beginning of the study, I decided to use MySQL as the database where to store all pre-processed data gathered from the buses and weather. MySQL is very commonly used open source relational database management system. I decided to use MySQL because I haven't used it before, but I am familiar with SQL -databases as I have used PostgreSQL before. However, during the process of the study, I decided to change to using MariaDB instead of MySQL. MariaDB is a community developed and commercially supported fork of MySQL, so the differences aren't very big. The reason why I decided to change to MariaDB during the study, was that since late December 2017, MariaDB has had a function to query for medians from the data. Getting the median is possible also with MySQL, but the structure of the query is much more complex. Additionally, the performance of the querying for medians is better in MariaDB.

5.3. Data processing

Data processing system was implemented by using NodeJS. NodeJS is an open-source, cross-platform JavaScript run-time environment, which runs JavaScript outside of a browser. I decided to use NodeJS because I didn't have much knowledge from it before hand, but as it is very commonly used, I knew it would be able to do what I wanted to achieve. In overall, my experience from JavaScript wasn't very extensive, so this way I could strengthen my knowledge of JavaScript both while making the server- and client-side systems. NodeJS was used as a backend system for processing the data which was collected to database before sending it to the data presentation system, running in the browser.

5.4. Data presentation

As the presentation of the data was done on a browser, HTML, CSS and JavaScript were used to present the results. I decided to use these technologies because I wanted to learn more about frontend development and this solution made it easy for me to check the status of the data collection system whenever I wanted and wherever I wanted. I decided to make web -based data presentation system also because there are a lot of easy and free to use JavaScript libraries for data presentation. Especially presenting of data on map and as charts was important in this use case. An open-source JavaScript library called Leaflet was used to easily visualize map related data and results on a map. Community developed JavaScript library called Chartist was used to visualise data and results as charts.

6. Data collection and analysation system

Figure 3 presents the overview of the automated data collection and analysation system which was developed during this study. The objective was to create an automated, real-time system for collection and processing of the data and for presentation of the results. One of the original requirements for the system was also that it would have to be able to collect data from long periods of time.

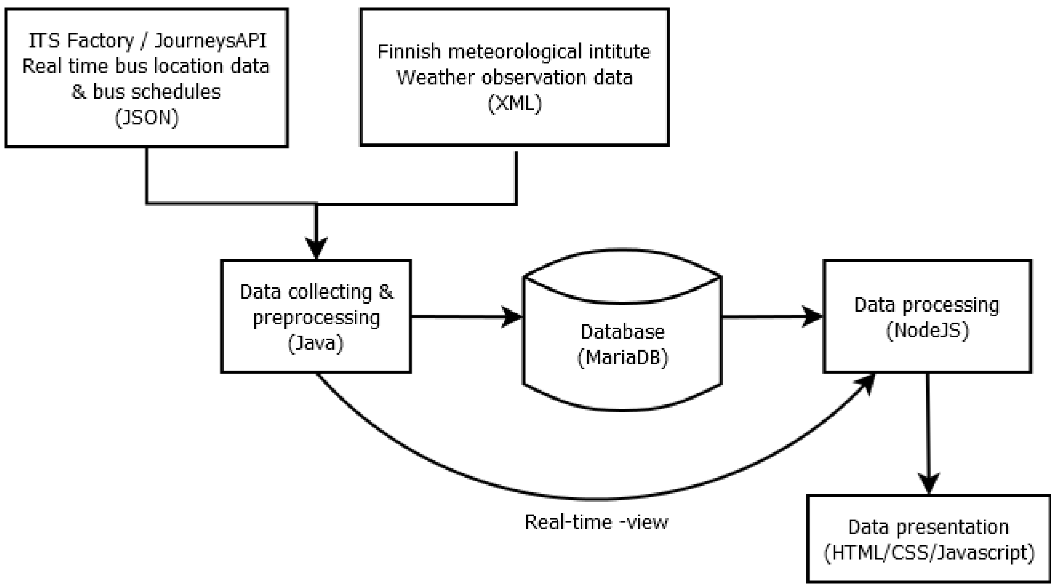

Figure 3. Data collection and analysing system overview

6.1. Data collection and pre-processing

Data collection and pre-processing system was implemented with Java and the pre-processed data was saved to a MariaDB -database. During testing, the system was running in a separate test environment, but for the collection of the actual results, which would be analysed in more detail, the system was running on a cloud hosted virtual machine. The virtual machine was hosted by DigitalOcean.

In multiple occasions, the pre-processing of data required the system to calculate distances between two GPS -points. This happened for example when the system had to count the distance from the current location of the bus to the bus stops along the route and when evaluating whether the midpoint of a travel link was close enough to a weather station. The Haversine Formula [The Haversine Formula, 2019] was used for the calculation of distances between two points:

$$a = sin^2\left(\frac{\Delta\varphi}{2}\right) + \cos\varphi_1 \times \cos\varphi_2 \times sin^2\left(\frac{\Delta\lambda}{2}\right)$$
$$c = 2 \times \arctan\left(\frac{\sqrt{a}}{\sqrt{(1-a)}}\right)$$
$$d \approx R \times c$$

Where $\Delta\varphi$ is latitude difference between points 1 and 2 in radians, $\Delta\lambda$ is longitude difference between points 1 and 2 in radians, φ_1 is the latitude of point 1 in radians, φ_2 is the latitude of point 2 in radians, d is the approximated distance between points 1 and 2 in radians and R is radius of the earth (mean radius = 6,371km).

As also pointed out by Syrjärinne [2016] there are multiple different options how to calculate the distances between two GPS points, which all would have given relatively accurate answers in this use case. Therefore, she ended up using more simple equation on her study to increase the performance of data processing. However, as the pre-processing system developed in this study was running real-time, the calculation of more complex equations wasn't a performance issue and therefore I decided to use the Haversine formula instead. The Haversine formula is the most accurate of the options which don't take the altitude changes and elliptic shape of the earth into account. This solution gives accurate enough distance measurements in the context of this study, where distances are relatively short, and the altitude changes are small.

During the journey of a bus, weather observation data and bus data were collected separately. Bus data and weather observation data were linked to each other and the results written to the database when it had been 15 minutes since the last location data was received from the bus. Last location data is usually received when the bus arrives to the terminus stop. Following chapters will describe the data collection systems for bus and weather observation data.

6.1.1. Bus data

Because real-time bus location data is updated once every second, this means that the data collection system has to poll the Journeys API real-time data interface once a second. This data contains only the data related to the current status of the bus on the route which it is currently running on. Therefore, the data collection system also has to request for the route and stop related data separately, at least once, for each bus service which was returned from the real-time interface.

The biggest issue concerning the collection of real-time bus data, was the amount of data that should be collected. If all data, received from the real-time interface, would have been saved in uncompressed format it would have required about 1,6 gigabytes of disk space per day. In addition to real-time data, also the data regarding routes, stops and schedules would have to be saved. As one of the requirements for the system was for it to be able to stay up and running for long periods of time, it quickly became clear that it wasn't possible to collect and store all data which was available from the bus fleet.

This meant that only data which was valuable towards the objective of the study should be saved to the database. In case of evaluating the link travel times, the most important information from the bus location data are the timestamps of the arrivals and departures to bus stops. Therefore, only the arrivals and departures to bus stops were recorded.

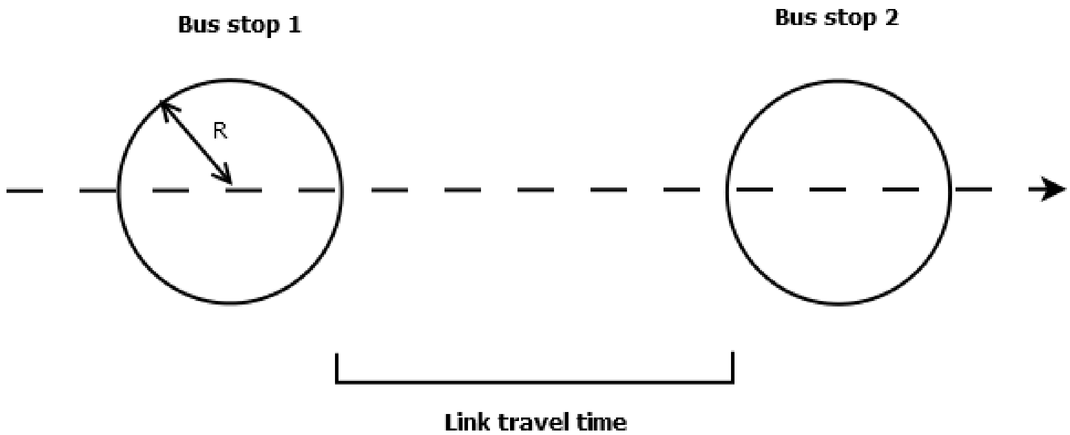

Figure 4. Counting of the link travel time. R value of 50 metres was used in this study.

To detect if bus was on a bus stop, distance from the current location of the bus was calculated to the location of the bus stop, reported by the Journeys API. Bus was considered to have arrived at a bus stop when distance to the stop was less than 50 metres and the bus hadn't arrived at this bus stop before during its journey. Bus was considered to have departed from the bus stop when distance to the bus stop was more than 50 metres and bus had arrived at the bus stop earlier. During the development of the data collection system, different values for R were tested. With smaller values for R, the arrival accuracy to the bus stops started to decrease quickly. Meaning that according to the bus location data and stop location data, buses never went close enough to the bus stop, so that the system would have detected the arrival and departure. 50 metres may seem quite high number as radius to detect if the bus is at a bus stop or not. For example, if there is a traffic light -controlled intersection right next to the bus stop, the system might detect the arrival to a bus stop bit earlier or the departure bit later than when then arrival or departure has actually happened. In order to detect the arrivals and departures to bus stops more accurately, more meta data, e.g. the signal from the door release would be needed.

On the other hand, with bigger values for R the variation to link travel times, produced by bus sometimes stopping to the bus stop to pick up or drop off passengers and sometimes only driving past it, are made smaller. Meaning that if bus has enough distance to accelerate from the bus stop before the link travel time counting is started, link travel times between the times when bus has stopped and the times when bus has only passed the bus stop, are more comparable. Same applies when approaching a bus stop, counting of the link travel time should be stopped before the bus has to start deaccelerating if it is going to stop at the bus stop. As the objective of this study is to investigate the differences in link travel times produced by adverse weather, bigger value for R is better. Therefore, the concept of arrival and departure to a bus stop is bit different from what one could expect, as the bus stops are considered as timepoints along the route of the bus and not really as places to pick up and drop off passengers. This study is aiming to exclude the effect to link travel times produced by passengers boarding or alighting the bus.

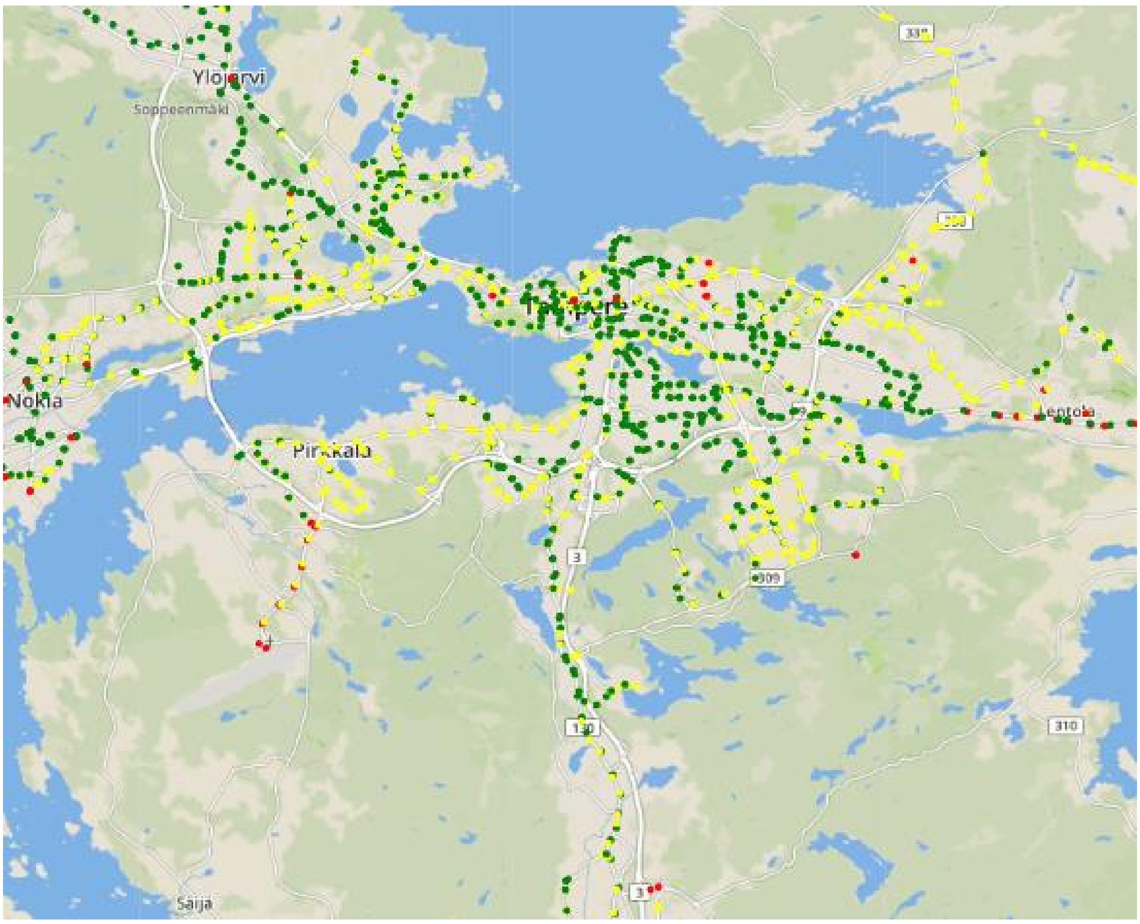

Figure 6. Arrival accuracy of the buses to bus stops.

Figure 6 presents the arrival accuracy of the buses to bus stops during the evaluation period with R value of 50 metres. Green dot means that the bus, which should have visited the bus stop, has visited the bus stop with over 95% accuracy. Yellow means that the arrival accuracy has been between 95% and 80% and red means that the arrival accuracy has been less than 80%.

There are number of different possibilities why the arrival and departure from a bus stop might not have been recorded:

- Connection issue between the bus and the Journeys API
- Connection issue between the data collection system and Journeys API
- Incorrect position data for the bus stop
- Bus has driven through a different route than planned and therefore circling around the bus stop
- Bus service got cancelled before the bus arrived at the bus stop

Arrival accuracy to the bus stops is important factor in the case of this study, because otherwise it is not possible to count the link travel time for the link. However, as visible from figure 6, the arrival accuracy to the bus stops has been high with R value of 50 metres.

6.1.2. Weather data

It would have been possible to query also for historical data from the FMI open data API, but as the bus data had to be collected by querying the JourneysAPI real-time, same solution was used in the collection of road weather observation data. However, in difference to the one second interval used in bus data collection, weather observation data was queried only once every 5 minutes as the observation interval for road weather station observations is 5 minutes. Also, in difference to bus data, all weather observation data was saved also to the database as the amount of weather observation data is considerably smaller than the amount of bus location data. Therefore, the amount and handling of weather observation data didn’t introduce the same challenges as the bus location data did.

Most important values from the road weather observations were air temperature and the amount of precipitation from the past 10 minutes. The ability to get the air temperature and precipitation amount of past 10 minutes, once every 5 minutes from each road weather station gives very precise data regarding the weather. Especially, when compared to weather data used in the past studies, where the precipitation amounts have usually been available on an hourly basis. The high frequency of weather observation data and using of multiple road weather stations gives much more precise data about the status of precipitation on a specific area or stretch of road. Heavy rainfall and heavy snowfall can often affect only a small area of the road network and only for short periods of time. Therefore, precise data about the amount of precipitation is especially important when precipitation is so severe that it affects traffic speeds by reducing the visibility. Often this level of precipitation doesn't last very long and the data about the intensity of the short rain or snowfall could be lost if the precipitation data would be available on an hourly basis.

6.2. Database structure

The structure and content of the database was designed so that the responsibility of filtering the results would be on the data processing and presentation system. The data collection and pre-processing system would only record the data and save it to the database in easily usable format. This way it would be possible to later define and experiment with different limiting values, for example for the distance from weather station to the midpoint of the link and how close to the time of link observation the weather observation would have to be in order for link observation to be counted as valid.

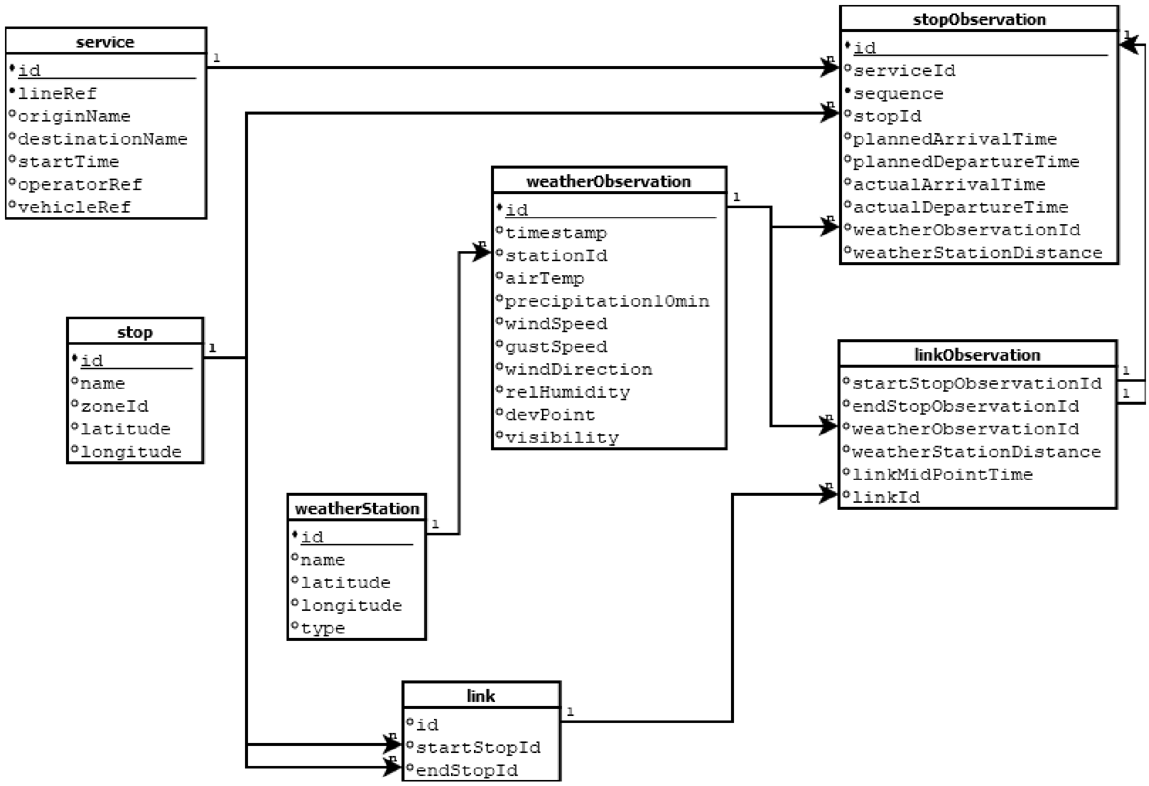

Figure 7. Overview of the database structure.

Figure 7 presents the structure of the database. From each bus service, the data related to the service, stops and links between stops was saved to the database already after the first observation received from the bus service. This way the same stops and links between stops could be referred to from each bus service which used the same stops and same links. Journeys API doesn't have a system to indicate whether the service ended successfully or if the connection to the bus was somehow interrupted. Therefore, the pre-processing system considered the bus service to have ended once it hadn't received the status data from the bus service for 15 minutes. At this point, the pre-processing system evaluated the stop and link observations which had been made during the process of the bus service and linked the stop and link observations to the nearest weather observation from the nearest road weather station. From road weather station observations, all data was saved to the database.

As is visible from figure 7, there is some recurrent data, and data which would be possible to count from other values in the database. This kind of data are for example the distances to nearest weather stations and mid time point for the link. This kind of data was saved in the database like this, so it would be easier and faster to process the results in the data processing and presentation system.

The database also contains some extra data which wasn't used in this study, but which would already enable for example evaluating of the effect of wind or visibility, measured from the road weather stations, to the link travel times. It would also be possible to investigate if some specific bus lines are more likely to suffer from delays in different weather circumstances.

6.3. Data processing and presentation

For the final processing and presentation of the data, a web -based processing and presentation system was developed. Processing of the data was done with NodeJS on the server side of the system and the presentation was done using JavaScript on the client side. The purpose of this system was to count and visualize some key values and results from the data which had been stored into the database. In addition to analysing the results from the database, the system enables the user to see live view of the locations of the buses. This enabled the debugging and evaluation of the data already during the data collection. Additionally, some of these results and visualizations could already be used as they are for the final evaluation of the results of the study. However, the biggest benefit from the system came in the form of making the debugging and development easier. The visual presentation of the data also enabled more detailed examination of the quality of data which was available from the Journeys API and from the FMI open data interface.

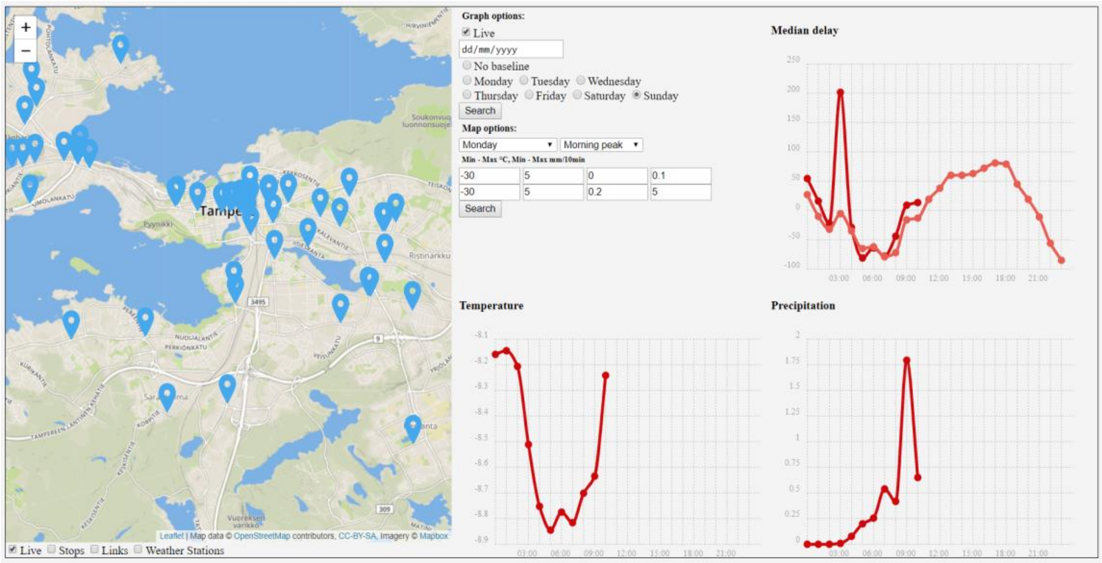

Figure 8. Overview of the data presentation system

Figure 8 presents the overview of the data presentation system. On the left side of the view is a map, on which various information can be visualised:

- Live location of all buses which are currently in operation
- Locations of the bus stops and their arrival accuracy
- Links between the bus stops
- Location and coverage of the road weather stations
- Results for a link travel time comparison during different weather circumstances

The right side of the view consist from several components: options for map and graph data, temperature graph, arrival delay graph and precipitation graph.

First, on top of the options view, there are the options which can be used to retrieve data from a specific day to the graphs. By default, the data shown on the graphs is from the current day. User can also select a baseline to be shown on the arrival delay graph, to compare the delays of the current day or from the day he searched, to the hourly median delays on a specified day of the week. Next on the options view are the options related to comparing of links travel times during different days of the week, time of the day and different weather circumstances. Below the options view, is the temperature graph, which shows either the hourly average temperatures measured today or during the day which the user has searched for.

On the right side of the temperature graph is the hourly average precipitation graph, which works the same way as the temperature graph. In the right upper corner of the view is the median arrival delay graph. Just as the temperature and precipitation graph, this graph can also show the values either from the current day or from the day which the user has searched for. In addition to this, the user can select a baseline graph from a day of a week, to compare the median delays during different days.

7. Results

The data for analyzation of final results was collected during winter 2019, from Saturday 5th of January to Friday 8th of February (5 weeks). During the observation period there were no special days (for example national holidays) which would have had to be excluded from the evaluation. Originally the plan was to collect data from much longer period of time but due to late changes in the data collection system, the data collection period was shortened. This also meant that as the data collection happened in January, all precipitation came down in the form of snow during the data collection and therefore the effects of rainfall couldn't be evaluated. The daily temperatures measured from Härmälä weather station in Tampere during the data collection were between 2.0°C and -26.1°C and the biggest daily precipitation amount was 7.5 millimetres. The daily weather observations from weather station in Härmälä, Tampere during data collection period are presented in table 5.

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
					5.1 -1.4°C to -7.4°C 0.0mm	6.1 -1.3°C to -5.8°C 0.1mm
7.1 1.5°C to -1.7°C 1.5mm	8.1 -0.4°C to -1.5°C 4.5mm	9.1 -1.5°C to -4.8°C 0.0mm	10.1 -4.6°C to -8.5°C 0.8mm	11.1 1.8°C to -6.2°C 0.0mm	12.1 2.0°C to -9.7°C 0.0mm	13.1 -2.3°C to -13.1°C 5.5mm
14.1 -1.9°C to -7.9°C 0.9mm	15.1 -4.7°C to -13.0°C 0.1mm	16.1 -4.2°C to -9.5°C 4.1mm	17.1 -6.8°C to -9.8°C 1.4mm	18.1 -8.2°C to -16.5°C 0.0mm	19.1 -3.9°C to -20.1°C 2.5mm	20.1 -3.9°C to -14.1°C 0.0mm
21.1 -12.3°C to -24.2°C 0.0mm	22.1 -8.9°C to -26.1°C 0.4mm	23.1 -1.7°C to -8.9°C 0.4mm	24.1 -3.0°C to -15.8°C 0.1mm	25.1 -3.4°C to -12.2°C 0.9mm	26.1 -3.4°C to -10.4°C 0.8mm	27.1 -10.4°C to -23.9°C 0.0mm
28.1 -10.7°C to -24.6°C 2.8mm	29.1 -5.3°C to -10.8°C 6.0mm	30.1 -6.8°C to -8.8°C 0.4mm	31.1 -7.5°C to -14.8°C 0.0mm	1.2 -5.9°C to -10.3°C 0.4mm	2.2 -6.3°C to -12.5°C 2.9mm	3.2 -3.2°C to -8.5°C 7.5mm
4.2 -0.1°C to -8.8°C 1.6mm	5.2 0.6°C to -7.6°C 5.6mm	6.2 0.4°C to -12.5°C 2.2mm	7.2 -0.6°C to -7.7°C 1.3mm	8.2 1.7°C to -1.0°C 1.2mm		

Table 5. Daily weather during the data collection period (Tampere Härmälä).

During January, it is also very dark in Tampere as the length of the day was only 5 hours and 48 minutes in the beginning of the data collection and 8 hours and 25 minutes in the end of the data collection. At the time of data collection, there were also lot of road constructions on going in the city centre of Tampere as new tram network was under construction at the time of the data collection. However, the effect of darkness or the change in the length of the day nor the effect of road constructions were taken into account in the evaluation of the results.

During the 5 weeks of data collection, data from 3 million bus arrival events to bus stops were recorded. The counting of the link travel time was possible for 2,9 million link travel events from 2463 links between 2158 bus stops.

Total congestion is the sum of recurrent and non-recurrent congestion. The effects of recurrent congestion and delays were ruled out from the results of this study by always comparing the results from the same time of the day on a same type of day (working day or weekend). Therefore, the travel time changes evaluated in this study are caused by non-recurrent events like accidents, road constructions, special events or adverse weather. To be able to evaluate the effects which are caused by weather, an assumption had to be made that other types of non-recurrent congestion occurred with same possibility during different weather circumstances. However, as discussed earlier, according to previous studies this is not always true as for example accidents do happen more often during adverse weather. Evaluation of the size of this effect would have been very complicated and was therefore left out of the scope if this study. Therefore, the assumption that congestion caused by other non-recurrent events would happen with the same possibility during different weather circumstances will add some noise to the results of this study.

7.1. Identifying the recurrent delays during different times of the day on different days of the week

Syrjärinne has studied the regular delays of bus traffic in Tampere region in her doctoral thesis. She compared the planned and actual arrival times to bus stops and identified the bus lines, hours of the day and areas of the city where buses were regularly running late from schedule. According to the results of her study, the biggest regular delays occurred in the morning between 8 and 9, and during afternoon between 15 and 18. [Syrjärinne, 2016]

In Tampere, also a local newspaper has made a study in which they examined the timing of the morning and afternoon traffic peaks in Tampere based on the data from LAM -sensors of Finnish Transport Agency. According to the newspaper’s results, in most places the morning traffic peak extends approximately from 07:00 to 09:00 and afternoon traffic peak from 15:00 to 17:00 on the highways in and out of Tampere [Aamulehti, 2017].

In order to investigate the recurrent delays associated with regional bus traffic in Tampere, I compared the planned and actual arrival times of the buses on different hours of the day and days of the week. By evaluating the recurrent delays, it would be easier to identify the times of the day on different days of the week when the recurrent delays are similar. The times when recurrent delays are similar, could later be compared against each other when trying to detect the effects of additional non-recurrent delays caused by adverse weather. If the result set would have been bigger, it would have been possible to simply compare the link travel times on same hours of the day on same days of the week, but categorization of the different times of the week allowed the usage of bigger samples from the result set.

When studying the delays of arrival to bus stop, the scheduled arrival time to bus stop was compared to the actual arrival time according to the live bus location data. The arrival to the bus stop was evaluated as described in chapter 6.1.1. Especially from the observations which were made during night time, it was noticed during the study that the Journeys API might return incorrect planned arrival times to bus stops. For example, if bus was actually detected to have arrived at a bus stop at 2019-01-11 03:56:54, the planned arrival time from Journeys API was 2019-01-10 18:40:30. This might have happened due to issues in Journeys API or for example due to the bus being assigned to a wrong route. In order to reduce the effect of this kind of invalid outlier values, medians were used instead of averages when counting the difference between planned arrival and actual arrival times to bus stops. Additionally, all stop arrivals which have happened over ±6 hours from the planned arrival time were not taken into consideration when retrieving the median values. To further optimize the results, this value could still have been decreased for example to 2 hours, as usually the routes and intervals between buses running on the same route are relatively small in the case of regional bus traffic.

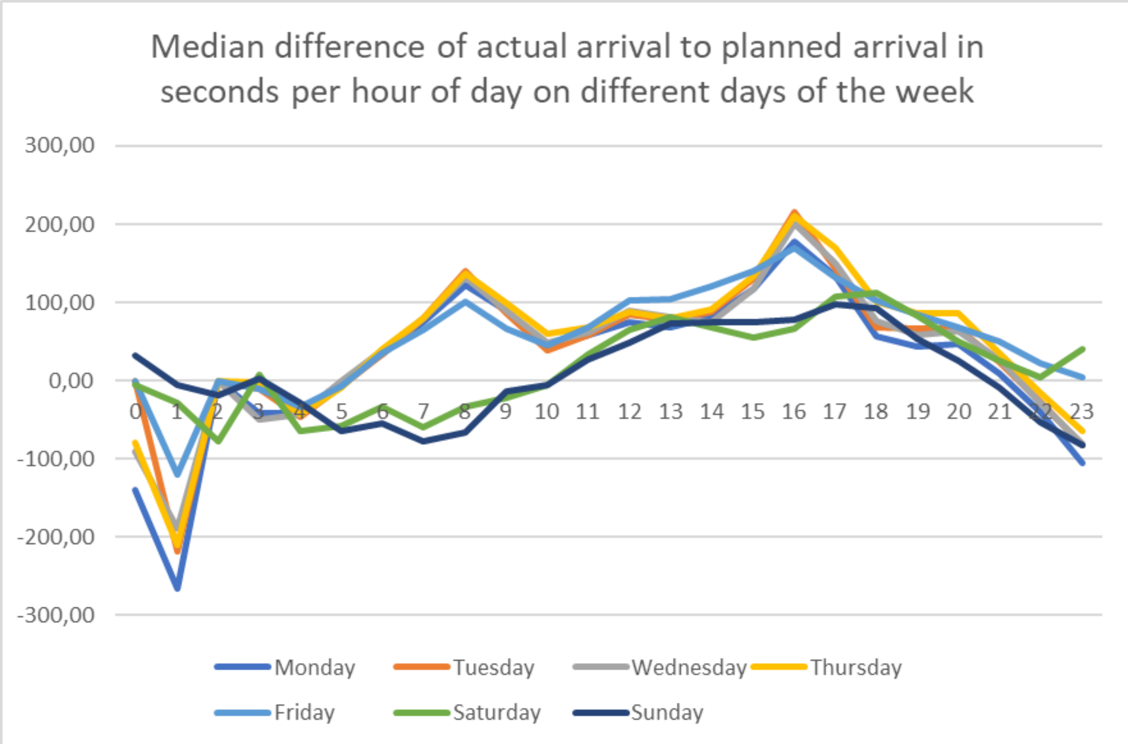

Figure 9. Median difference between planned arrival time to bus stop and the actual arrival per hour of day

Figure 9 presents the results of the study examining the recurrent delays during different times of the day on different days of the week. The results clearly indicate that the daily profile of the delays is very similar from Monday to Friday and on Saturday and Sunday. During the working days, the biggest difference seems to be that during Fridays, the recurrent delays during afternoon peak hours are starting bit sooner and they are also smaller during their highest peak. The result gives clear indication that link travel times during working days and weekends cannot be compared to each other. Therefore, also non-recurrent delays caused by weather should be evaluated separately for working days and weekends. From these results, it is also possible to detect different times of the days which should be compared separately when examining the non-recurrent delays caused by weather. The results regarding the daily recurrent delays from working days are very similar to what Aamulehti observed in their study based on the loop detectors.

Based on results by Syrjärinne, study made by Aamulehti and the results of my median arrival difference study, I decided to use the categories described in table 6 for evaluation of the effects of adverse weather during different times of the day and days of the week in my data processing.

Working days		
	Morning peak	07:00 – 09:00
	Noon	09:00 – 15:00
	Afternoon peak	15:00 – 17:00
	Evening	17:00 – 21:00
	Night	21:00 – 07:00
Weekends		
	Day	07:00 – 21:00
	Night	21:00 – 07:00

Table 6. Categorisation of times of day and days of week

7.2. Identifying the weather categories

To be able to compare the link travel times during different weather circumstances, the weather observations have to be categorized. As the data collection period was quite short, it has to be made sure that the categories are defined so that each of the categories includes enough observations as too small result set in one of the categories could make the results invalid. The results will also get more biased if the frequency of link travel times during the weather category are low.

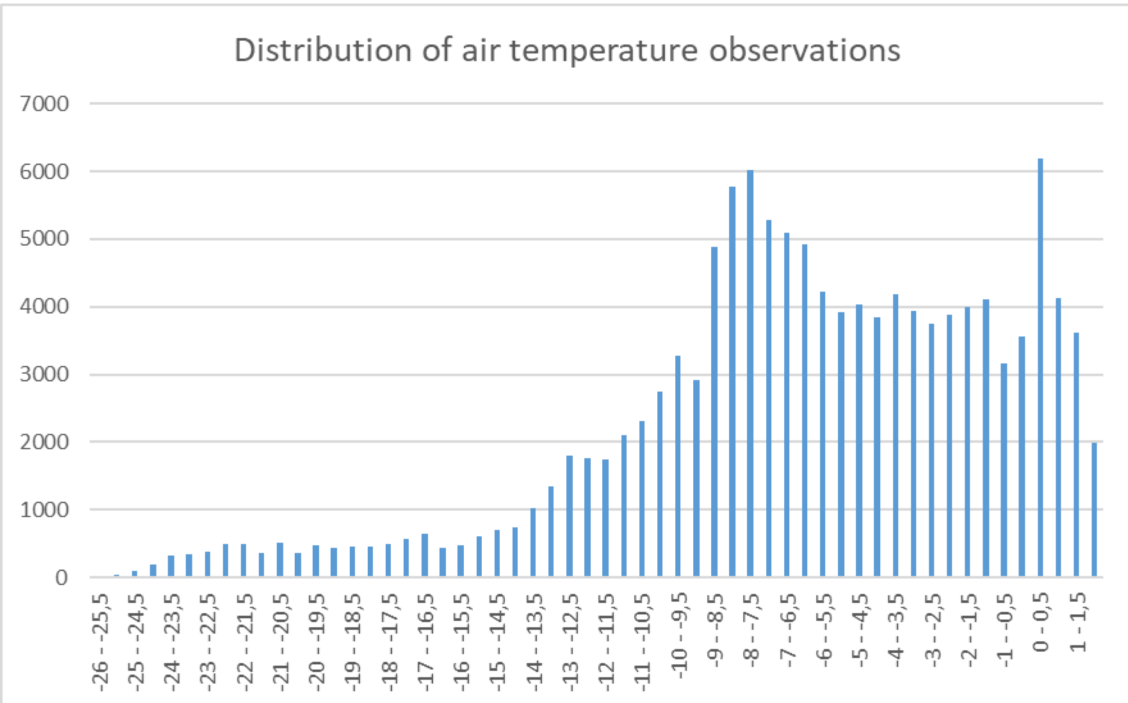

Figure 10. Distribution of air temperature observations per 0.5°C measured during the data collection period.

Figure 10 presents the distribution of air temperature from all weather observations which were made during the data collection period. I decided to divide the weather observations into three categories based on air temperature. This way it would be possible to more clearly see if there would be a difference in the link travel times during higher and lower temperatures as the medium temperatures wouldn't get to even out the results. To divide the weather observations based on the air temperature into three categories with all categories having high frequencies, values described in table 7 were used.

Category	Air temperature range	Frequency
High	> -4,0°C	23339
Medium	-4,0°C >= and > -10,0°C	46104
Low	<= -10°C	23489

Table 7. Categories for evaluating the effects of air temperature.

Same categorization has to be made also for the precipitation. In this case I also decided to divide the weather observations into three categories. This would enable the investigation whether the effect of precipitation to link travel times would be greater when the precipitation was more intense. As majority of the weather observations didn't measure any precipitation, the non-rainy situation was logical decision to be selected as one of the three categories. The observations which had measured some amount of precipitation were divided so that the category with high precipitation amount would still have high enough frequency so the lack of link travel time observations wouldn't affect the result.

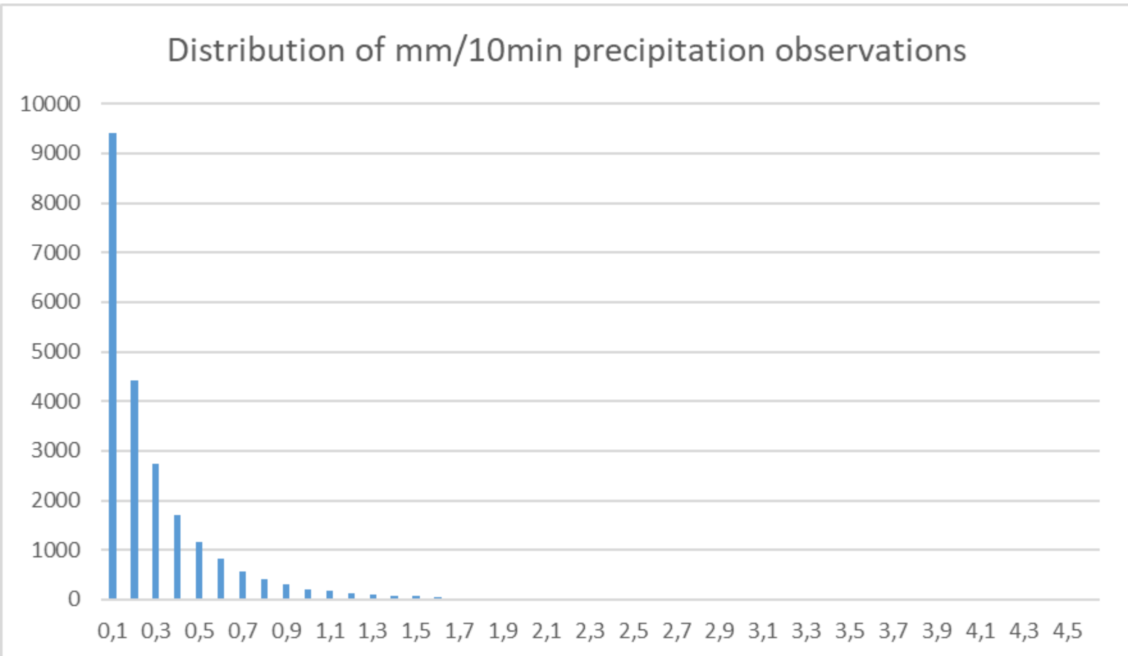

Figure 11. Distribution of precipitation observations per 0.1 mm/10min measured during data collection period.

Figure 11 presents the distribution of weather observations which included some precipitation. Precipitation of 0.0mm/10min was left out from the diagram to make the distribution of observations which included precipitation more visible, as 82.4% of the weather observations during the data collection were from situations without any precipitation. Table 8 presents the categorization used for evaluating the effect of precipitation to link travel times.

Category	Precipitation range	Frequency
High	> 0.5mm/10min	1898
Low	0.5mm/10min >= and > 0.0mm/10min	14019
No precipitation	= 0.0mm/10min	77015

Table 8. Categories for evaluating the effects of precipitation.

7.3. Overall effects of air temperature to link travel times

To evaluate the effects of air temperature to link travel times, variance analysis was performed for link travel times during different air temperatures on different times of the week. As the distances of the links are not know, comparison of average travel speeds during different circumstances couldn't be made. Instead, the link travel times on same links were compared to the median link travel time on the same link during high temperature circumstances.

In order for the link travel time to be counted as valid for evaluation, the midpoint of the link (GPS midpoint between start and end bus stops) had to be closer than 4000 meters from nearest weather station. Additionally, there had to be weather observation from this weather station within ±5 minutes from the mid time of the link for the link travel time observation to be valid for evaluation. Also, the frequency of valid link travel times from the link during the evaluated weather circumstances had to be at least 5, so links with low sample size couldn't bias the result.

After counting the percentual difference of the valid link travel times compared to the median travel time of the same link during high temperature circumstances, also some of the clearly outlier link travel time values were removed by filtering the results were the travel time was less than 20% or higher than 400% compared to the median travel time during high temperature circumstances. This filtering was required as there were some outlier values which were clearly due to either an issue with the data collection and pre-processing or due to some exceptional events like accidents. These outlier values could have had proportionally big effect in the results of the study. The direct effects of weather are expected to be much smaller than the effects which were represented by the outlier values.

Day of Week	Time of Day	Temperature Category	N	Mean	95% Confidence Interval for Mean	
					Lower bound	Upper bound
Working day	Morning	Low	60838	1.05	1.05	1.05
		Medium	118174	1.04	1.04	1.05
		High	61586	1.05	1.05	1.06
	Noon	Low	143554	1.03	1.03	1.03
		Medium	284467	1.04	1.04	1.04
		High	176807	1.05	1.05	1.05
	Afternoon	Low	71894	1.04	1.04	1.04
		Medium	171179	1.05	1.05	1.05
		High	124689	1.06	1.05	1.06
	Evening	Low	60434	1.02	1.02	1.03
		Medium	149432	1.04	1.04	1.04
		High	98102	1.04	1.04	1.04
	Night	Low	48731	1.03	1.03	1.03
		Medium	113998	1.03	1.03	1.04
		High	61925	1.04	1.04	1.04
Weekend	Day	Low	108721	1.04	1.03	1.04
		Medium	195598	1.06	1.06	1.06
		High	76790	1.04	1.04	1.05
	Night	Low	16144	1.03	1.03	1.03
		Medium	27608	1.03	1.03	1.04
		High	10608	1.03	1.03	1.03

Table 9. Results of the variance analysis for the effect of air temperature to link travel times

Table 9 presents the results of the variance analysis for the effect of air temperature to link travel times. According to the One-Way ANOVA test, during all times of the week the means of the groups differed from each other except during weekend nights, when the differences between groups weren't statistically meaningful.

According to results in all categories, the mean travel times during high temperature circumstances were never smaller than during low temperature during the same time of the week. So, in overall it seems like the near zero temperature (2,0°C to -4,0°C) would predict slightly longer link travel times when compared to cold weather (-26°C to 10°C). However, the differences are relatively small, ranging from 2 to 0 percent depending on the time of the week. Bit surprisingly, during daytime on weekends it seems like medium air temperature would suggest 3 to 1% longer link travel than during the same time of the week under high and low temperature circumstances. In overall it seems like the effects of air temperature to link travel times are small.

The results of this study are in line with the results of previous studies as they haven't found big correlation between air temperature and traffic delays either. Of course, it would have been interesting to collect link travel times from longer period of time to see whether higher temperatures would affect the link travel times somehow as all temperatures observed during this study were relatively low.

As the effects of air temperature to link travel times were small, the areas or links were the effects of air temperature are biggest weren't investigated.

7.4. Overall effects of snowfall to link travel times

To evaluate the effects of precipitation to link travel times, variance analysis was performed for link travel times during different amounts of precipitation on different times of the week. Same study method and evaluation criteria were used for validation of the link travel times as was used when evaluating the link travel time differences during different air temperatures. The link travel times on same links were compared to the median link travel time on the same link during non-rainy circumstances.

Day of Week	Time of Day	Snowfall Category	N	Mean	95% Confidence Interval for Mean	
					Lower bound	Upper bound
Working day	Morning	Heavy Rain	14331	1.09	1.09	1.10
		Light Rain	49162	1.08	1.08	1.08
		No Rain	177209	1.05	1.05	1.05
	Noon	Heavy Rain	12800	1.05	1.04	1.05
		Light Rain	112867	1.05	1.05	1.05
		No Rain	479130	1.05	1.05	1.05
	Afternoon	Heavy Rain	6473	1.05	1.05	1.06
		Light Rain	59428	1.08	1.07	1.08
		No Rain	301821	1.05	1.05	1.06
	Evening	Heavy Rain	4619	1.06	1.06	1.07
		Light Rain	60458	1.04	1.04	1.05
		No Rain	242979	1.04	1.04	1.04

	Night	Heavy Rain	9326	1.06	1.06	1.07
		Light Rain	37555	1.04	1.03	1.04
		No Rain	178010	1.04	1.04	1.04
Weekend	Day	Heavy Rain	9129	1.05	1.04	1.05
		Light Rain	53508	1.05	1.05	1.06
		No Rain	318936	1.04	1.04	1.04
	Night	Heavy Rain	553	1.01	.99	1.03
		Light Rain	6409	1.03	1.02	1.03
		No Rain	51410	1.03	1.03	1.03

Table 10. Results of the variance analysis for the effect of precipitation to link travel times

Table 10 presents the results of the variance analysis for the effect of precipitation to link travel times. According to the One-Way ANOVA test, during all times of the week the means of the groups differed from each other except during weekend nights, when the variances of different groups differ from each other so that the result of the ANOVA test can't be considered valid. This was most likely caused by the small sample of link travel times from heavy rain conditions during weekend nights (553).

According to the results, the effect of precipitation to link travel times differs during different days of the week and times of the day. The biggest effect was observed during morning peak hours on working days when mean link travel times were 4 to 5% longer during heavy snowfall than during non-rainy circumstances. During evenings and night time on working days the effect was also noticeable as link travel times during heavy snowfall were 2 to 3% higher than during non-rainy circumstances. Surprisingly, during working day afternoons, the effect of heavy snow fall was not visible while light snowfall caused 2 to 3% longer link travel times. During noon on working days and during daytime on weekends precipitation seemed to not have an impact on the link travel times. In overall it seems like snowfall has bigger effects on link travel times than air temperature. The effect of the intensity of snowfall seems to be bit contradictory as during working day afternoon peak hours, light snowfall had bigger effects to link travel times than heavy snow fall.

The results of this study indicate much smaller effects to travel times during snowfall than what many of the previous studies have observed. However, as reported for example by Chung [2012] and Tsapakis *et al.* [2013], the time of the day seems to have an influence on the size of the effect which snowfall has for travel times.

7.5. Links where the effects of snowfall to link travel times are the biggest

To be able to evaluate where the effects of snowfall to link travel times are the biggest, a data visualisation tool was developed. The tool enables visual comparison of link travel times during different weather circumstances during the same time of the week.

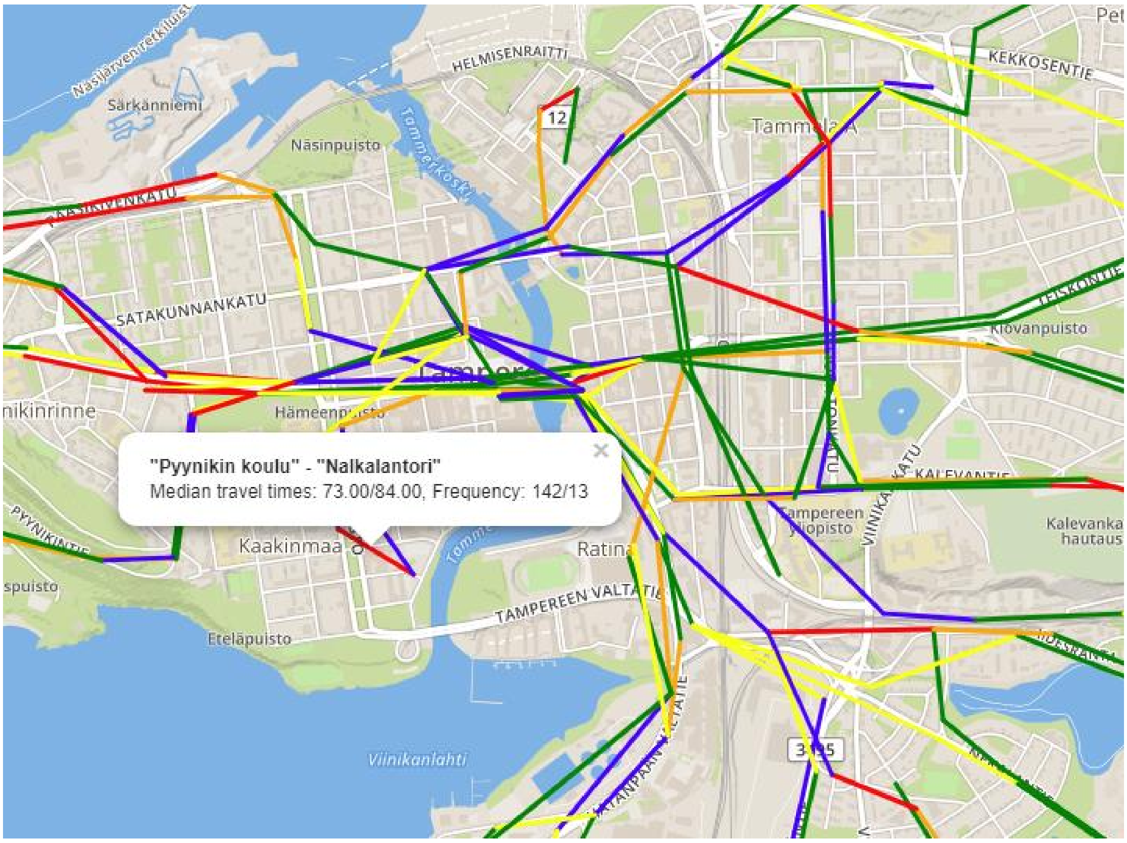

Figure 12. Visualization tool for evaluation of links which are affected more by the weather

Figure 12 presents the visualization tool developed for evaluation of links which are affected more by the weather. The tool enables closer examination of individual links by reporting the median link travel times during two weather categories which are currently being compared and the frequencies of link travel time observations during both weather categories. The visualisation tool uses the same categories for different times of the week and for different weather circumstances as was used in the evaluation of overall effects of weather for link travel times. Based on the results about the overall effects of snowfall to link travel times I decided to present the data visualisations from only the most interesting cases.

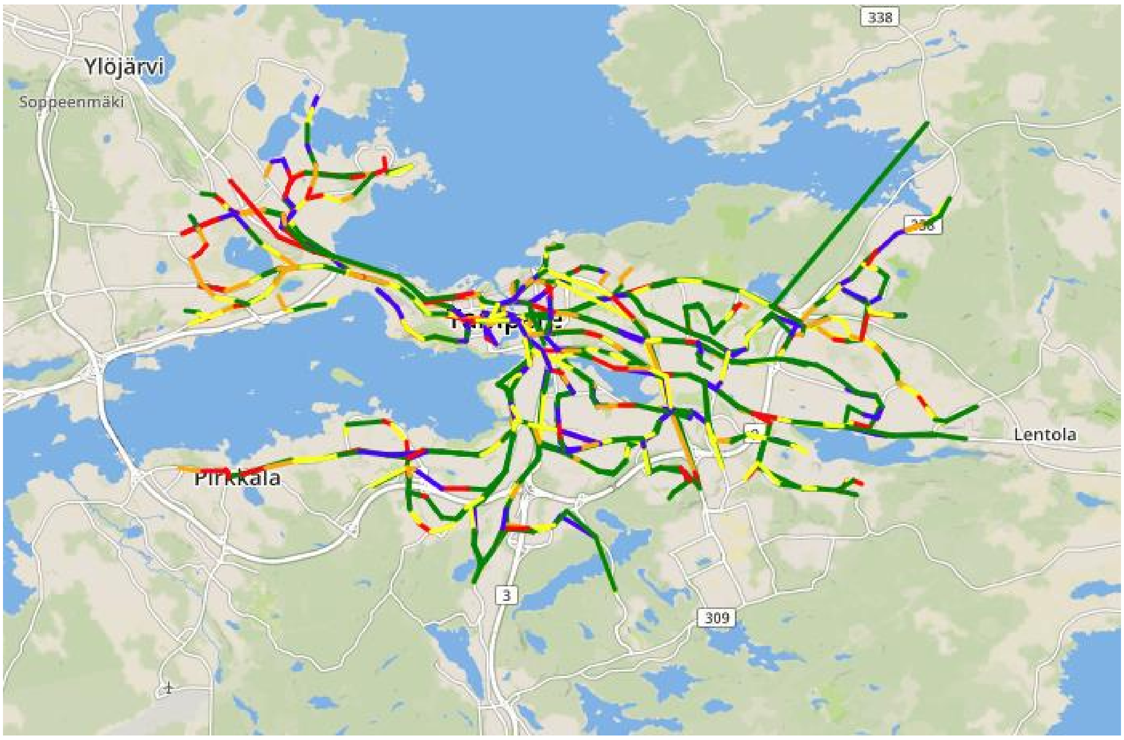

Figure 13. Comparison of median link travel times during non-rainy and heavy snowfall circumstances on working day mornings.

Colour	Difference of median link travel time	Number of links
Red	>= +15%	100
Orange	+15% > and >= +10%	110
Yellow	+10% > and >= +5%	176
Green	+5% > and > -5%	387
Blue	<= -5%	119

Table 11. Distribution of differences in median link travel times during non-rainy and heavy snowfall circumstances on working day mornings.

Figure 13 presents the visual presentation of median link travel times differences during non-rainy and heavy snowfall circumstances on working day mornings. Table 11 presents the distribution of differences in median link travel times during the same conditions. By indicating that more links had longer median travel times during heavy snowfall than during non-rainy conditions, the result of the visualization suggests the same as the result of the study of overall effects of snowfall to link travel times. However, it is not easy to identify areas which would have more links with higher link travel times during heavy snowfall. The links with higher travel times are spread out somewhat evenly around the area under investigation. This suggest that heavy snowfall doesn't cause large-scale congestion to big areas but instead affects only smaller areas like individual intersections.

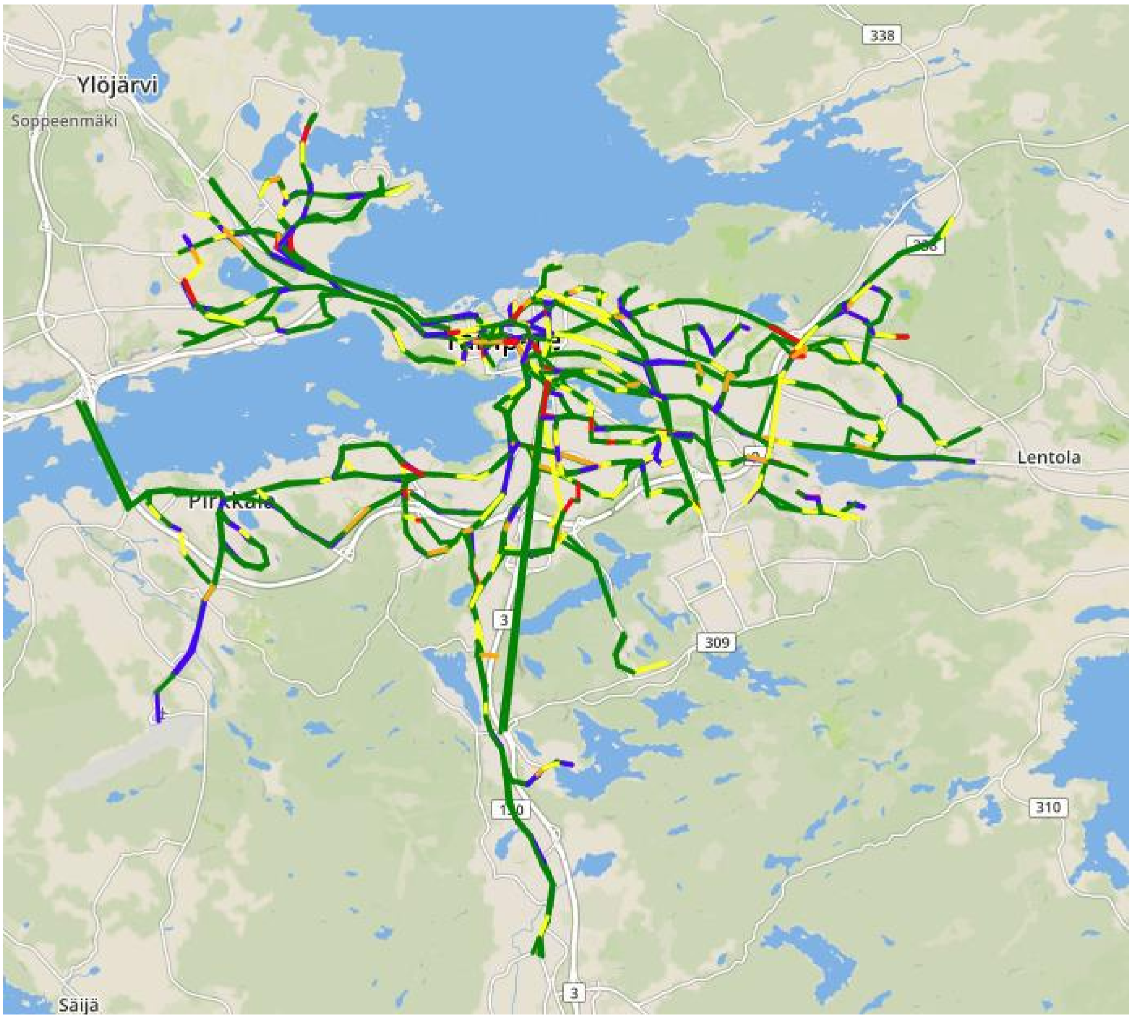

Figure 14. Comparison of median link travel times during non-rainy and light snowfall circumstances on working day afternoons.

Colour	Difference of median link travel time	Number of links
Red	>= +15%	39
Orange	+15% > and >= +10%	58
Yellow	+10% > and >= +5%	210
Green	+5% > and > -5%	770
Blue	<= -5%	146

Table 12. Distribution of differences in median link travel times during non-rainy and light snowfall circumstances on working day afternoons.

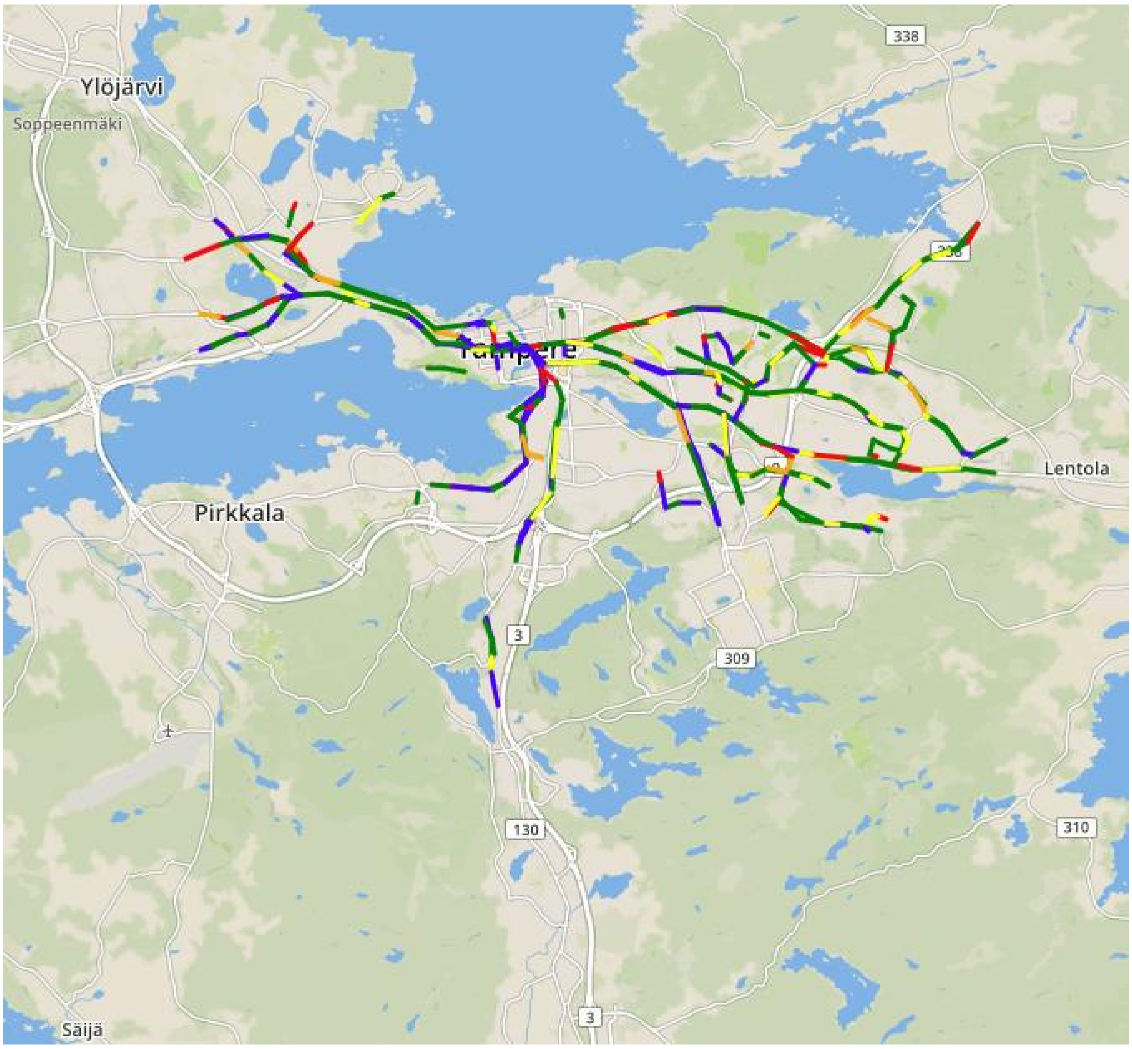

Figure 15. Comparison of median link travel times during non-rainy and heavy snowfall circumstances on working day afternoons.

Colour	Difference of median link travel time	Number of links
Red	>= +15%	45
Orange	+15% > and >= +10%	32
Yellow	+10% > and >= +5%	57
Green	+5% > and > -5%	229
Blue	<= -5%	95

Table 13. Distribution of differences in median link travel times during non-rainy and heavy snowfall circumstances on working day afternoons.

Figures 14 and 15 present the visual presentation of median link travel time differences during non-rainy and light snowfall circumstances and during non-rainy and heavy snowfall circumstances on working day afternoons. Tables 12 and 13 present the distribution of differences in median link travel times during the same conditions. According to the results of the study of overall effects of snowfall to link travel times light snowfall had bigger effects to link travel times during working day afternoons than heavy snowfall. However, as is clearly visible from the visualization of the results, there were observations from much smaller amount of links during heavy snowfall circumstances. Therefore, the amount of link travel time observations as well as the areas from which the observations are might have affected the result of the study of overall effects of snowfall to link travel times. If no link travel time observations were made during heavy snowfall from the areas where snowfall affects the link travel times more, this could have affected the results of the study.

However, in the extent of this study it is not possible to evaluate the size of this effect. Therefore, it would be important to gather link travel times form longer period of time, so travel time data from all weather conditions would also be available from the whole area which is under examination.

8. Conclusions

The results of the study were roughly in line with the results of previous studies. However, the size of the effect of snowfall to link travel times was smaller than reported by most of the previous studies.

The length of the data collection period should have been longer in order to get higher frequencies for link travel times during adverse weather from all parts of the area which was under examination. Most likely data should be gathered from multiple years, as for example heavy snowfalls might not happen at all during all the different times of the week, which were evaluated in this study, during the course of one winter. With a longer data collection period it would have also been possible to evaluate the effects of rainfall and the effects of air temperature with a broader temperature range.

To better understand the effects of weather to link travel times, more variables than just the air temperature and precipitation should be taken into consideration. In overall the status of the road should be considered in more detail as the current weather doesn't always tell about the status of the road and how fast vehicles can drive on it. For example, in the case of snowfall the build-up of snow on the roads often affects the traffic for some time also after the snowfall has stopped. For this purpose, there already are some public API:s available which could be used to get the information about which road have been cleared from snow and which not. Also, the effect of temperature to travel times is most likely more complex than the approach which was used in this study. For example, if the data collection system would be able to detect possibly icy road conditions when temperature goes below 0°C after rainfall, the system would have better understanding of the factors which are actually affecting the travel times.

References

Aamulehti. 2017. Retrieved from https://www.aamulehti.fi/uutiset/nama-ovat-pahimmat-ruuhkahetket-tampereen-tarkeimmilla-sisaantulovaylilla-saada-matka-aikaa-vartilla-ja-saastat-aikaa-yksi-paikka-kannattaa-kiertaa-ruuhka-aikaan-200640173

Singhal, A., Kamga, C., Yazici, A. 2014. *Impact of weather on urban transit ridership.* Transportation Research Part A 69 (2014) 379–391.

Ajoissa pysakilla. 2015. Retrieved from http://ajoissa.pysakilla.fi

Akin, D., Sisiopiku, V, Skabardonis, A. 2011. *Impacts of Weather on Traffic Flow Characteristics of Urban Freeways in Istanbul.* 6th International Symposium on Highway Capacity and Quality of Service Stockholm, Sweden June 28 – July 1.

Andrey, J., Mills, B., and Vandermolen, J. 2001. *Weather information and road safety.* Institute for catastrophic loss reduction paper series 15. University of western Ontario, London, Ontario, Canada.

Bacon J., Bejan A.I., Beresford A.R., Evans D., Gibbens R.J., Moody K. 2011. *Using Real-Time Road Traffic Data to Evaluate Congestion*. In: Jones C.B., Lloyd J.L. (eds) Dependable and Historic Computing. Lecture Notes in Computer Science, vol 6875. Springer, Berlin, Heidelberg

Chin, S.M., Franzese, O., Greene, D.L., Hwang, H.L., Gibson, R.C., 2004. *Temporary Losses of Highway Capacity and Impacts on Performance: Phase 2.* Oak Ridge National Laboratory, Oak Ridge, Tennessee.

City of Tampere, 2019, Retreived from https://www.tampere.fi/tampereen-kaupunki/tietoa-tampereesta/avoin-data/avoin-data-lisenssi.html

Fayazi, A., Vahidi, A., Mahler, G., Winckler, A. 2015. *Traffic Signal Phase and Timing Estimation From Low-Frequency Transit Bus Data*. IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, VOL. 16, NO. 1.

Ferris, B., Watkins, K., Borning, A. 2009. *OneBusAway: A transit traveller information system.* Proceedings of Mobicase 2009. San Diego, CA, USA, October 26-29.

Finnish Transport Agency. 2018. Retrieved from https://www.liikennevirasto.fi/avoindata/tietoaineistot/lam-tiedot#.W-12WegzZPb

GTFS. 2018. Retrieved from http://code.google.com/transit/spec/transit_feed_specification.html

Ibrahim, A., Hall, F., 1994. *Effect of adverse weather conditions on speed–flow–occupancy relationships.* Transportation Research Record: Journal of the Transportation Research Board 1457,184–191.

ITS Factory. 2018. Retrieved from http://wiki.itsfactory.fi/index.php/Tampere_traffic_lights_API

Jenelius, E., Koutsopoulos, H. 2013. *Travel time estimation for urban road networks using low frequency probe vehicle data*. Transportation Research Part B 53 (2013) 64–81.

Jägerbrand, A., Sjöbergh, J. 2016. *Effects of weather conditions, light conditions, and road lighting on vehicle speed.* SpringerOpen.

Journeys API. 2019. Retrieved from http://wiki.itsfactory.fi/index.php/Journeys_API

Kuhns, G., Ebendt, R., Wagner, P., Sohr, A., Brockfeld, E. 2011. *Self Evaluation of Floating Car Data based on Travel Times from actual Vehicle Trajectories*. German Aerospace Center, Institute of Transportation Systems (DLR-TS).

Kyte, M., Khatib,Z., Shannon,P., Kitchener,F. 2001. *Effect of weather on free-flow speed.* Transportation Research Record: Journal of the Transportation Research Board 1776,60–68.

Ma, X., Tao, Z., Wang, Y., Yu, H., Wang, Y. 2015. *Long short-term memory neural network for traffic speed prediction using remote microwave sensor data,* Transportation Research Part C: Emerging Technologies Volume 54, May 2015, Pages 187-197

Mapbox. 2019. Retrieved from https://www.mapbox.com/

OpenStreetMap. 2019. Retrieved from https://www.openstreetmap.org/

Parrando, N., Donoso, Y. 2015. *Congestion Based Mechanism for Route Discovery in a V2I-V2V System Applying Smart Devices and IoT.* Sensors 2015, 15, 7768-7806.

Syrjärinne, P. 2016. *Urban Traffic Analysis with Bus Location Data.* Doctoral Dissertation, School of Information Sciences, University of Tampere.

Saneinejad, S., Roorda, M., Kennedy, C. 2012. *Modelling the impact of weather conditions on active transportation travel behaviour.* Transportation Research Part D 17 (2012) 129–137

Shrank, D., Lomax,T. 2001. *The 2001 Urban Mobility Report*. Texas Transportation Institute, Texas A&M University, College Station, Texas.

SIRI. 2019. Retrieved from http://www.transmodel-cen.eu/standards/siri/

Smith, B.L., Byrne,K.G., Copperman,R.B., Hennessy,S.M., Goodall,N.J. 2004. *An investigation into the impact of rainfall on freeway traffic flow.* In: Proceedings of the TRB 83rd Annual Meeting, Transportation Research Board, Washington, DC.

FMI Open data final report. 2013. Retrieved from https://helda.helsinki.fi/bitstream/handle/10138/42401/2013nro6.pdf

The Haversine Formula. 2019. Retrieved from http://www.longitudestore.com/haversine-formula.html

Toth, C., Suh, W., Elango, V., Sadana, R., Guin, A., Hunter, M., Guensler, R. 2013. *Tablet-Based Traffic Counting Application Designed to Minimize Human Error.* Transportation Research Record: Journal of the Transportation Research Board. 2339.

Transportation Research Board. 2000. *Highway Capacity Manual*. National Research Council

Tsapakis, I., Cheng, T., Bolbol, A. 2013. *Impact of weather on marcoscopic urban travel times.* Journal of Transport Geography, Volume 28, April 2013, Pages 204-211

Uno, N., Kurauchi, F., Tamura, H., & Iida, Y. 2009. *Using bus probe data for analysis of travel time variability.* Journal of Intelligent Transportation Systems: Technology, Planning and Operations, 2-15.

Waze. 2019. Retrieved from https://www.waze.com/en

Xu, F., He, Z., Sha, Z., Zhuang, L., Sun, W. 2013. *Assessing the impact of rainfall to traffic operation of urban road network.* Proceedings of 13th COTA International Conference of Transportation Professionals. Guangzhou, China

Chung, Y. 2012. *Assessment of non-recurrent congestion caused by precipitation using archived weather and traffic flow data.* Transport Policy, Volume 19, Issue 1, January 2012, Pages 167-173.

Yu, X., Prevedouros, P. 2013. *Performance and challenges in utilizing non-intrusive sensors for traffic data collection*. Advances in Remote Sensing, Vol. 2 No. 2, 2013, pp. 45-50.