Predicting aircraft noise complaints directed at Schiphol Airport using multivariate linear regression

Vera Willemijn de Jong
Anr. 424036

THESIS SUBMITTED IN PARTIAL FULFILMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN COMMUNICATION AND INFORMATION SCIENCES, MASTER TRACK DATA SCIENCE: BUSINESS AND GOVERNANCE,
AT THE FACULTY OF HUMANITIES
OF TILBURG UNIVERSITY

Thesis committee:

Dr. Grzegorz Chrupala
Dr. ing. Sander Bakkes

Tilburg University
Faculty of Humanities
Department of Communication and Information Sciences
Tilburg, the Netherlands
July, 2016

Preface

My internship at Heijmans B.V. has taught me more about data science than any Iris dataset probably could. I am grateful to have been given the opportunity to write my thesis at Heijmans B.V. and for the organizations that provided data: the Royal Netherlands Meteorological Institute (KNMI), Intermediator Residents Schiphol (BAS) and Schiphol Airport.

I would like to start by thanking my colleagues at Heijmans B.V. Firstly I want to thank Gerard de Leede, for the interesting discussions about the future of data science in business. Secondly, I wish to thank Nynke van der Lei, whose determination and persistence were a great help during the time that we were trying to obtain data. Thirdly, I would like to thank my colleagues and classmates at Heijmans B.V. for sharing the frustrations when code wasn't working and for sharing the joy when it was. Finally, many thanks go out to my supervisor Jordi Overeem. His enthusiasm, belief in the project, creativity and advice have been of invaluable help, for which I am very grateful.

Special and many thanks go out to my supervisor Grzegorz Chrupala, whose expertise and willingness to answer my questions have been very helpful and much appreciated. I am also grateful for the freedom that I was given in choosing a topic and creating a dataset.

I'm thankful to all my classmates, many of whom I have become good friends with over the past year. I want to thank Muriel Versteeg for her support in particular.

I would also like to thank my friends, especially Marion Passchier and everyone of Supernova. Last but not least, many thanks go out to my family for cheering me on and for their moral support.

Abstract

Aircraft noise is a problem to people's health and wellbeing. It is also an issue for airports, because noise complaints limit them in their wishes to expand. Previous studies have often focused on minimizing aircraft noise or on relating aircraft noise with complaints. This study takes a different approach to aircraft noise complaints and asks the question: to what extent can the number of complaints be predicted per day and per runway used for arrival or departure at Schiphol Airport? Another question that is explored is: which features have the most predictive power in predicting the number of complaints per runway, per day? The data that is used in this study are from independent organizations BAS (Residents Intermediator Schiphol) and KNMI (Royal Netherlands Meteorological Institute) and from Schiphol Airport. Multivariate linear regression is applied to the data. Model 2, which includes all the independent features except for the number of flights, shows the best evaluation scores (r-squared = 0.8920, RMSE = 25.2798, MAE = 12.3657). It is recommended that more research is done and more complex models are used in order to obtain better predictions of the number of complaints.

Contents

1. Introduction

Over the past few decades, commercial airplane flights have expanded immensely. Recent forecasts estimate that the demand of commercial flights will continue to increase with five percent per year for at least twenty years (He, Wollersheim, Locke & Waitz, 2014). Despite potential economic benefits, a rise in air traffic also poses challenges. These problems can primarily be found in both environmental and public health domains. One of the most pressing problems are health concerns as a result of all the extra aircraft noise (Berry & Sanchez, 2014). Research in this field has shown that air traffic noise has a significant negative impact on the health of residents who live close to an airport (Morrel, Taylor & Lyle, 1997). For instance, persons living in residential neighborhoods around airports experience more stress, are less productive and are more prone to depression (Goines & Hagler, 2007; Hjortebjerg et al., 2015; Correira, Peters, Levy, Melly & Dominici, 2013).

People have often complained against increased noise levels. In fact, it is one of the most protested topics by the public (Berglund & Lindvall, 1995). Based on these protests and because of the adverse health effects of noise pollution, governments and other regulatory organizations such as the International Civil Aviation Organization (ICAO) have made an effort to reduce aircraft noise (Filippone, 2014). The ICAO has installed a noise abatement program in 2001 that advocates a 'balanced' approach in reducing aircraft noise. They aim to do this by encouraging airline companies to produce aircrafts that generate less noise and by introducing laws about proper management of the airplane runways (Zaporozhets, Tokarev & Attenborough, 2011). Models of noise prediction have been adapted to these new international aviation standards. The focus of these models has been on predicting levels of aircraft noise, which will be elaborated upon in section 2 (Netsajov, 2011).

Instead of predicting noise levels, researchers have also started to investigate complaints that are related to aircraft noise. Individuals who are exposed to aircraft noise often feel annoyed or stressed (Lim, Kim, Hong & Lee, 2008). They express their discontent by reporting complaints to the airport or to an independent organization. Although several studies have indicated that this is an important issue, not much attention has been paid to exploring complaints and people's complaint behavior (Hume et al., 2003; Maziul, Job & Vogt, 2005). Previous research on this topic has demonstrated that there are many different findings on the reasons why people complain. Ultimately, it is not always clear why or when people complain about noise levels or what makes a person complain about noise, but it can have much influence on an airport's noise abatement strategies (Wiechen, Franssen, de Jong & Lebret, 2002).

1.1 Research questions

Since the focus of the literature in the field of aircraft noise has been on finding associations between aircraft noise (prediction), annoyance and complaints, this paper takes an alternative approach. It uses multivariate linear regression to *predict* the number of complaints that are reported per runway, per day and per arrival or departure usage at Schiphol Airport. There are six airplane runways at Schiphol Airport. Each of these runways consists of two routes that airplanes can use to arrive on or to departure from. The complaints will be predicted per each of these arrival or departure routes, instead of using the entire runway. This will give a more accurate view of the complaints and whether arrival or departure that per airplane runway affects the number of complaints.

Instead of doing analyses from a statistical perspective, this study takes a data science approach. This means that multivariate linear regression is used to do predictions instead of making statistical inferences. Furthermore, data from several sources is retrieved to be able to extract information from them, which is a fundamental element of data science (Provost & Fawcett, 2013). This could yield more insight in when individuals are more likely to complain and how the usage of runway affect the number of complaints. Airport Schiphol is used as an example in this study, because they have a coherent system in place that registers the complaints of citizens in the Netherlands and connects them to the runway. Compared to similar airports, they also have one of the highest levels of complaint behavior (Wiechen et al., 2002; Hulshof & Noyon, 1997). Hence, there is reliable data available about the number of complaints per runway and per day. This study also uses data that is obtained from Schiphol and from open source websites such as the Royal Netherlands Meteorological Institute (KNMI).

Airport expansion has been difficult in the past because of the complaint behavior of Dutch citizens (Wiechen et al., 2002). The annoyance that people experience around the airports and in the rest of the Netherlands as a consequence of the increase in the number of flights has therefore also been topic of public debate. The problem of focusing too much on the noise levels themselves, is that it does not necessarily affect the complaints. People may for instance complain because they are stressed or because of other reasons that do not relate to aircraft noise itself. Yet, it is ultimately the complaints and public protests that could really make a difference in an airport's policy, apart from the (inter)national aircraft noise regulations (Netjasov, 2011). Therefore this study aims to focus on predicting the complaints and aims to answer the following questions:

1. To what extent can the number of complaints be predicted per runway and per day at Schiphol Airport?
2. Which features have the most predictive power in predicting the number of complaints per runway, per day?

In order to obtain the best possible accuracy score, this paper makes a few assumptions, which are further elaborated upon in section 3.

1.2 Practical relevance

Receiving too many complaints could be detrimental to an airport (Bergland & Lindvall, 1995). After all, receiving (too) many protests may lead to constraints for airports in terms of the number of flights that they can employ or it could limit potential plans of expansion. Hence, having a better understanding of when people complain and about which airplane, could be very insightful. It enables data driven decision making, of which the main benefits in this case are twofold.

First of all, being able to predict the number of complaints per runway can yield more information about when it would be most convenient to do tasks such as maintenance of the airplane runways. Every airplane runway needs to be repaired at set times each year, which makes them unavailable for usage. Consequently, other runways have to use more of their capacity in order to still be able to account for all the flights. If it is found that the number of flights on a particular runway highly affects the amount of complaints, then maintenance tasks can be planned differently. Having a better understanding of how many complaints the airport receives can cause the maintenance schedule be done in such a way that it helps reduce complaints.

Secondly, when Schiphol airport knows when noise complaints are going to occur because at a certain day there may for example be more flights, then this can be taken into account when the flights are planned. Of course, there are many different factors that need to be taken into account when flights are scheduled. The importance of the complaints should therefore be weighed against other tasks, such as making sure that all the flights can be managed. Nevertheless, it can still be important to take into account. Ultimately, when complaints are reduced, airports could generate more commercial flights and possibly expansion that will result in increased profits.

1.3 Academic relevance

Much research has been done on predicting aircraft generated noise levels and on how aircraft noise affects people's health. Yet, very few publications can be found in the literature that focuses on the number of complaints that airlines receive about the noise. Even the few studies that have been conducted on complaints are from over a decade ago (Hume et al., 2003; Maziul et al., 2005).Yet it is still relevant to investigate aircraft noise complaints. Complaints could yield more information about the complainant behavior of people and what factors could influence the number of complaints. It should be noted that in this study, complaints are defined as concrete results of the annoyance that people experience which could therefore be used as a measure to assess people's feelings of annoyance (Hume, Terranova & Thomas, 2002). Since not much research has been done on complaints yet, this study could already potentially add to the body of literature.

What makes this study unique within the body of research in this field is that it uses multivariate linear regression to predict the number of complaints. This has not been done in this area of study before as the focus has always been on finding associations between levels of noise, annoyance and the effect of personal characteristics on complaints. The models that have been used before have not always been properly described in the literature, so not much is known about the methods that have been employed in understanding complaints and noise levels (Wiechen, et al., 2002).

Moreover, this study uses data from various sources to predict the number of complaints per runway, per day and per arrival or departure. This also makes it unique, because it takes into account features that have not been linked with complaints before (i.e. weather data and whether any maintenance was done on a runway). By using data from various sources and creating a dataset, predicting complaints with multivariate linear regression, this study has a unique contribution to make to the body of literature in this field.

1.4 Outline

This study aims to answer the research question by first describing the current literature and models that have been used more in depth. This theoretical framework also investigates the scope of this paper in more detail in section 2. After this section, the methods are described in section 3. This includes an overview of the dataset (section 3.2) and the analyses that are going to be used (section 3.5). Following the methods section, the results are reported and evaluated in section 4. These results are going to be discussed in section 5. Finally, the conclusion (section 6) gives an overview of the main findings and gives an answer to the problem statements.

2. Theoretical framework

Aircraft noise became a much-discussed topic in scientific literature once airports started to expand in the 1960s and 1970s. Aircraft noise is usually not considered as inherently harmful. Yet, a well-known study that was conducted by Meecham and Shaw (1979) demonstrated that higher mortality rates in neighborhoods around Los Angeles Airport were found due to noise pollution. Since then, subsequent studies have been critical about this conclusion, arguing that mortality rates increased because of other factors unrelated to noise (Frerichs, Beeman &

Coulson, 1980). Nevertheless, this study prompted research about the consequences of noise pollution. Despite disagreements and discussions in many studies about what exactly these health effects are and how severe their impact is, there is a consensus among researchers that aircraft noise has serious negative effects on people's health (Morrell, Taylor & Lyle, 1997; Kaltenbach, Maschke & Klinke, 2008). For instance, Franssen, Staatsen and Lebret (2002) conclude in their research about aircraft noise at Schiphol Airport and public health that being exposed to aircraft noise can affect the health of individuals who are exposed to it in a negative manner. Noise pollution is important and its effects should not be underestimated. It is not exactly clear which health effects are most problematic, although annoyance comes up most often as there is a low barrier to experiencing it (Lim, Kim, Hong & Lee, 2008). This study adopts the definition of annoyance that was stated in a study by Babisch et al., (2009): 'annoyance is a term used in general for all negative feelings such as disturbance, dissatisfaction, displeasure, irritation and nuisance'.

To reduce noise levels so airports can expand, research about aircraft noise has usually focused on the levels of noise and annoyance and the prediction thereof (Filippone, 2014; Steele, 2001). One of the first and most-cited studies about predicting noise disturbance was done by Hazard (1971). This research has become one of the prominent studies in literature on aircraft noise and annoyance, because Hazard recognized that both noise levels and personal characteristics could play a role in how much annoyance is experienced. His goal was to predict the level of annoyance based on the two aforementioned characteristics. He conducted interviews with individuals who lived nearby various larger airports in the USA and used an 'annoyance measure' index that was first developed by the National Opinion Research center in 1952 (Hazard, 1971). Noise levels were recorded by tape and were measured in perceived noise decibel (PNdB) before any of the interviews took place. Several social-psychological features were also selected, based on how well they predict annoyance. In addition, Hazard used multiple classification analysis to examine correlations between each predictor variable and other predictors (Hazard, 1971). He found that noise measures do not necessarily improve the prediction of annoyance. More specifically, he concluded that noise itself is a poor indicator of how much annoyance individuals experience. He concluded that other factors such as personality and demographics are much more important indicators of how people experience and respond to noise pollution (Hazard, 1971).

Despite these insights, however, there are also two main drawbacks to Hazard's (1971) study. First of all, the noise that was recorded spanned over a period of three months and occurred before any of the interviews were held. This discrepancy may affect the outcomes of the results in either a positive or negative manner. For example, there may have been much higher noise levels in the months that the noise was recorded or vice versa. The outcome of the study may not be completely representative of how people experienced aircraft noise. Secondly, the aircraft industry changes rapidly. Many new technologies are developed each year to create airplanes that generate less noise. Simultaneously, noise metrics that measure noise pollution have become more advanced. Nowadays, technology has become more precise at recording noise levels and reducing noise at different stages of the production of an aircraft. Since Hazard's study is relatively old and the necessary technologies were not developed yet, he was unable to sample aircraft noise continuously (Hazard, 1971). Measuring aircraft noise is in itself a difficult task, which some researchers consider impossible (Steele, 2001). There are many different methods to analyze this and there is no general agreement about what the best practice is for aircraft noise (Steele, 2001). Therefore, results of studies that were done with measuring noise should be carefully evaluated (Steele, 2001).

One of the limitations of this study and similar ones that have been conducted since Hazard (1971) is that exploring levels of noise, annoyance and personal characteristics excludes another important feature: complaints. Airports usually have a system in place that enables individuals to file complaints about the levels of noise that they experience (Netjasov, 2011; Fidell & Howe, 1998). Researchers have proposed to put more emphasis on investigating these aircraft noise complaints (Fidell & Howe, 1998). The reasons for doing this is because it could give a better understanding of why complaints are reported and what could affect the number of complaints. Specifically, it could have a number of practical reasons such as using complaint data to assess the usefulness of noise mitigation measures (Fidell & Howe, 1998). However, research on this topic has not nearly been as extensive as studies on the relationship between noise and annoyance (Berglund & Lindvall, 1995; Maziul et al., 2005). For example, Uphard, Maughan, Raper and Thomas (2003) propound the view that not much attention has been paid to the scientific investigation of complaint data. This premise also been supported by other researchers, who also found that qualitative analyses in regards to complaint data has not been extensive enough (Maziul et al., 2005).

A study that does involve complaint data in relation to aircraft noise was carried out by researchers at Manchester Metropolitan University (Hume, Martin, Thomas & Terranova, 2003). This research distinguished itself from other studies by combining noise levels and complaints. The reason for doing so was to investigate whether there was a link between aircraft noise levels and complaints that were reported by citizens. The complaint data was collected from Manchester Airport Community Relations Department to which people could express their discontent in form of complaints. Once people made a complaint, their address, gender, time-of-day and nature of complaint were all reported as well. The researchers retrieved complaints from July-December 1998 and July-December 1999. Manchester Airport has a monitoring system in place that measures the levels of noise at nine locations around the airport (Hume, Martin, Thomas & Terranova, 2003). Five of these monitors were placed at places within 6.5 kilometers from the airport. This takes all the fluctuating noise levels into account and creates an average. The researchers also included four other monitors that were positioned in densely populated neighborhoods close to Manchester Airport (Hume et al., 2003). Only half of the noise complaints that were received were matched with the corresponding noise recordings, because many complaints were not specifically associated with a particular aircraft. Another obstacle that arose in this study is that there were problems with the recordings. Hence, not all noise events were registered.

The researchers associated the noise levels and complaints with each other via the monitoring system, though they do not exactly describe how they carried out their analyses (Hume et al., 2003). By doing so they found that found that the number of complaints increased as the noise levels increased as well. Moreover, they concluded that the main reason for complaints directed at Manchester Airport are because of aircraft noise. They also discovered that there are large individual differences between how people experience annoyance and when people decide to complain (Hume et al., 2003). Whether people decide to complain depends for a large part on factors such as assertiveness and whether individuals feel that their complaints matter. Some individuals would complain very often (i.e. several times per day) and were classified as 'serial complainers'. The reasons for why they complain much was not known, but it did not influence the main findings of the study (Hume et al., 2003). Despite the personal characteristics that play a key role in the complaints that Manchester Airport received, this study emphasizes that ultimately there is a stronger relationship between aircraft noise and complaints.

One of the studies that is not in accordance with aforementioned conclusion was carried out by Maziul, Job and Vogt (2005). Their study consists of a literature review, which discusses the studies that have been carried out on this topic. Their main premise is that the complaints that people report do not accurately reflect noise levels. One of the arguments they put forth to support this claim is that people experience more noise annoyance when they do not expect to be exposed to higher noise levels (Maziul, Job & Vogt, 2005; Hatfield, Job, Carter, Peploe, Taylor and Morrell, 2001). For example, a new air route was introduced at Sydney and Vancouver airports that would require airplanes to fly over areas that were previously relatively noise free. When the airplanes started to fly in these areas, the airports received many complaints. There was not an objective difference between the noise that was experienced in the neighborhoods around the new air route and the regular aircraft noise. However, people had not expected the new aircraft generated noise levels so there was a high increase in the number of complaints. It could be suggested that once individuals are used to noise, they are less likely to complain. Furthermore, the researchers point to other features that have much more effect on why individuals complain instead of the noise levels themselves (Maziul, Job & Vogt, 2005). For instance, people who consider themselves more sensitive to noise are much more likely to complain. Moreover, other factors such as having noise insulated windows and even one's political preferences could play a role in whether someone is more likely to complain (Maziul, Job & Vogt, 2005; Wirth, Brink & Schirz, 2003). These are examples of features that could be important in determining the number of complaints that an airport will receive. Since this study only consists of a literature review, the methods and analyses were only briefly touched upon. Even by examining the studies that were referenced more in depth, the methods of analyses were still not properly elaborated upon and in some cases even lacking (Maziul, Job & Vogt, 2005; Wirth, Brink & Schirz, 2003).

A study that does incorporate analyses of noise levels and complaints was conducted by Wiechen, Franssen, de Jong and Lebret (2002). They are researchers at the National Institute for Public Health in the Netherlands. The goal of their study was to explore the relationship between aircraft noise exposure of citizens living around Schiphol airport and the prevalence of complaints. In the Netherlands there had been much political discussion about airport expansion (Wiechen et al., 2002). As a result, the researchers were ordered by the Dutch Ministry of Housing, Spatial Planning and the Environment to investigate whether expansion would be feasible. For their analyses, they used complaint data that was registered by the Environment Advisory Committee Schiphol (EACS). This is an independent committee where Dutch citizens could file their complaints via telephone or in writing, 24 hours a day. Their name and address is also recorded and people who complain anonymously were not taken into account in this research. Additionally, the researchers mailed their questionnaire to 30000 randomly selected individuals of which 11812 were returned. This survey consisted of questions about how individuals experience annoyance, their personal characteristics, self-perceived health and what type of actions they have previously taken against aircraft noise (Wiechen et al., 2002). Furthermore, data of noise levels were also included in the research. The noise levels were calculated by the National Aerospace Laboratory, which was done in a way that was according to the Dutch Aviation Act (Wiechen et al., 2002). In order to associate complaint data to the aircraft noise levels, the postal codes of the individuals who complained were combined with the noise levels that were calculated by the National Aerospace Laboratory. Afterwards, they were combined with Geographic Information System (GIS) data, so it was possible to recognize where the complaints were coming from and what the noise levels were in that neighborhood.

A constraint of this study is that no reference is given to model that was used, how exactly the survey data was analyzed and that no information is given about the calculations that were done by the National Aerospace Laboratory. Another limitation is that the people who responded to the questionnaire could not be properly linked with the persons who filed a complaint because of privacy reasons. The researchers were able to circumvent this limitation somewhat by combining the complaint data with the levels of aircraft noise through the GIS system. In their analyses, they found that there is an ‘exposure-response relation’ as the number of complaints increase when noise levels become higher as well (Wiechen et al., 2002). However at higher levels of noise, feelings of annoyance started to decline. The authors attributed this to sound insulation that people will have added to their houses if the noise would have become too loud. Yet they stressed that the relations between noise annoyance and complaints may not paint a complete picture because the complaints do not give an accurate depiction of the actual response of individuals to noise pollutions (Wiechen et al., 2002). Even though many people may be annoyed by aircraft noise, less than a fifth of the people who were severely annoyed said that they had complained about the aircraft noise. Hence, only few people took the trouble of complaining even when aircraft noise really bothers them. In addition, the authors of this study found that there are other factors that influence complaint behavior. They mention determinants such as fear for an aircraft crash or concern about health and environment as other reasons for people to complain. Hence, this research concludes that even though noise levels affect complaints to a certain extent, most people do not report complaints. When individuals do file a complaint, it is not just because they are annoyed with the aircraft nose, but also because they could be afraid of an airplane crash or because they are worried about their health.

2.1 Research gaps

As can be inferred from aforementioned studies, it is not exactly clear what causes people to complain and how important aircraft noise level is. Either it is argued that noise levels affect the complaints that people make or, as is demonstrated in other research, other features could play a role in people’s complaint behavior. What the studies have in common is that the researchers have tried to associate noise levels with complaints and have sometimes combined them with personal characteristics. This approach has shown to be useful to a certain extent though it has also introduced some difficulties. For example, it is not always possible to relate complaints to survey data or associate the complaints to the noise levels. The research has also shown that there is still much ambiguity about what motivates people to complain when they experience aircraft noise. It is still useful to obtain more insight in why people complain, as airports will continue to expand. Therefore, this study takes a different approach to complaint data.

Instead of focusing on associations between complaint data and noise annoyance, this study predicts the number of complaints per day, per airplane runway and per arrival or departure at Schiphol Airport. By predicting the number of complaints it is possible to circumvent the problems around associating the complaints with the noise levels. Moreover, prediction of complaints enables airports understand complaint behavior better and to create measures that could reduce the complaints. This study makes use of multivariate linear regression to make these predictions. This in particular is a unique aspect because no predictive models have yet been applied in any of the previous studies about complaints. Yet it can yield interesting insights, especially as organizations and airports are starting to accumulate more data because of technological advanced methods. By mining that data, information could be obtained that could improve decision making, as data-driven decisions usually give more reliable outcomes (Shmueli, Patel & Bruce, 2010).

Furthermore, questionnaire data and noise data measurements could be somewhat problematic, as previous studies have indicated. Especially noise measurement is tricky, because there are many possibilities different possibilities in measuring it. For instance, one could choose to measure noise when an airplane is departing, or measure noise in nearby neighborhoods, choose various measuring noise systems, etc. Therefore, this study is going to emphasize other features that could be relevant in making predictions about the number of complaints. The data set that is used for this research consists of a combination of open source data and data that was obtained via Schiphol Airport and BAS (Residents Intermediator Schiphol). In brief, the dataset consists of, but is not limited to, the following features: wind, average temperature, maximum takeoff weight, number of flights, etc. More information about the dataset can be found in section *3.2* and Appendix A. This is also a unique aspect of this study, because the features in the dataset have not yet been related to complaints in previous studies. By focusing on other factors besides noise and personal characteristics, this study hopes to find features that could also play a role in the number of complaints that are reported.

3. Methods

3.1 Outline

This section gives an overview of the steps that were taken to investigate the research questions. This study predicts the number of complaints per runway, per day, by using multivariate linear regression. Section 3.2.1, 3.2.2 and 3.2.3 give a description of the various data sources and explains which assumptions were made in the process of obtaining this data. Section 3.2.4 and 3.2.5 discuss merging the data and the dataset itself. Section 3.3 shows more information about exploratory data analysis that was applied to the dataset. Section 3.4 elaborates on the pre-processing techniques that were applied to the dataset. The subsequent section, section 3.5, describes the chosen method of analysis in more detail and the evaluation metrics that are chosen to assess performance of the models.

3.2 Dataset description

3.2.1 Residents Intermediator Schiphol (BAS)

There are no datasets publicly available for the purpose of predicting complaints. A new dataset is therefore developed for this study. The data is obtained from various sources. Complaint data was retrieved from Residents Intermediator Schiphol (BAS) which is an independent organization where individuals can file their complaints via their website: www.bezoekbas.nl. The number of complaints per day, per runway and whether the plane arrived or departed from a particular runway were selected for analysis. This data was available on the website, but not directly downloadable. Permission from the organization was given to put the data into an Excel file. The date of the complaint data ranges from January 1st, 2008 until March 1st, 2016. The choice for this date is further elaborated upon in section 3.2.4. This resulted in 25697 rows and four features: date, runway, arrival or departure and number of complaints, an example of which is given in table 1 and a description of the data is given in Appendix B. Other data that was available on the BAS website consisted of maintenance schedules of airplane runways since 2010. When an airplane runway could not be used because of maintenance, other runways need to process more flights. This could result in an increase in complaints addressed to that particular runway. The maintenance dates were put into an Excel file, with a binary outcome for maintenance: zero for when there was no maintenance on a specific runway and one if there was. The file contains two features: date and maintenance. It consists of 475 rows, an example of which is shown in table 2. A description of the features is available in Appendix C. The dates should be viewed as an estimator for when maintenance took place, because actual maintenance often deviates from this schedule (BAS, 2016).

Table 1. Example of the complaint data that was retrieved from BAS

	Date	**Runway**	**Arrival / Departure**	**Complaints**
1	01-01-08	18L	D	83
2	01-01-08	09	D	274
3	01-01-08	06	A	40
4	01-01-08	24	D	55
5	01-01-08	18R	A	556

Table 2. Example of the maintenance data that was retrieved from BAS

	Date	Maintenance
1	24-05-10	1
2	25-05-10	1
3	31-05-10	1
4	01-06-10	1
5	02-06-10	0

3.2.2 KNMI Data

Data from the Royal Netherlands Meteorological Institute (KNMI) was retrieved because weather conditions could play a role in the usage of runways. For example, a lot of rain or snow may cause a specific runway to temporarily shut down, which could reduce the complaints for that particular runway (BAS, 2016). The KNMI is the Dutch national weather institute that records and stores data of all the weather stations in the Netherlands. This includes a weather station that is stationed at Schiphol airport. The data is publicly available on their website (www.knmi.nl) in .txt format. The file contains 41 features, such as average temperature, wind vector speed, average amount of precipitation. It contains 23803 rows starting from January 1st, 1953. For an example of the data see table 3 and for a full list of the features and a description, see Appendix D.

Table 3. Example of KNMI data (only four of the 41 features are shown here, for a full list, see Appendix D.

	Date	Wind direction average	Wind speed average	Average temperature	Sunshine duration
1	01-01-2008	104	35	2.4	0.7
2	02-01-2008	82	57	-0.5	6.5
3	03-01-2008	93	64	0.4	1.9
4	04-01-2008	132	62	3.4	0
5	05-01-2008	194	68	6.6	0.6

3.2.3 Schiphol Airport data

Based on the assumption that the number of flights and the maximum takeoff weight of an airplane affect the amount of noise, these features could also be useful in predicting the number of complaints. This data is not publicly available, so it was provided by Schiphol Airport. The time span ranges from January 1st, 2008 - March 1st, 2016 and there are 622029 rows in total. The features that are included are: Date, Arrival or departure, ICAO type, Description, Runway code, maximum takeoff weight (MTOW) and number of flights. ICAO type is the code of the airplane and description is the name of the type of the airplane. The feature maximum takeoff weight is the amount of weight in tons that a specific aircraft is allowed to have at departure or arrival.

Hence, the feature MTOW in the Schiphol Airport depends solely on the type of airplane and *not* on the number of flights. This means that if the ICAO type of an airplane is A319, as is the case in row one and row two in table 4, the MTOW will also be the same for these two rows. Even though the second row shows that there were nine flights that day from that runway, the MTOW does not change. The feature MTOW is therefore somewhat meaningless because it only indicates the MTOW of the type of the airplane, instead of how much MTOW there actually was on a runway. The feature MTOW could play an important role in the amount of noise, because higher weights could mean that more noise is produced on a specific runway. Another feature was therefore created that shows the sum of the MTOW. This was done by multiplying the number of flights with the MTOW. This gives a better accurate view of the maximum takeoff weight that a runway experienced on a day. This can be seen in table 5, where the new feature MTOW sum is created and the 'old' feature MTOW is removed from the dataset.

Table 4. Example of the Schiphol data.

	Date	**Arrival / Departure**	**Run-way Code**	**ICAO type**	**Description**	**MTOW**	**Nr. of flights**
1	01-01-08	D	09	A319	Airbus A319-1	77	1
2	01-01-08	D	24	A319	Airbus A319-1	77	9
3	01-01-08	D	18L	B738	B737-800 WING	80	3
4	01-01-08	A	06	B738	B737-800 WING	80	5
5	01-01-08	A	18R	B742	B747-400F	380	1

Table 5. Example of the Schiphol data with the MTOW sum and without the MTOW.

	Date	Arrival/ Departure	Run-way Code	ICAO type	Description	MTOW sum	Nr. of flights
1	01-01-08	D	09	A319	Airbus A319-1	77	1
2	01-01-08	D	24	A319	Airbus A319-1	693	9
3	01-01-08	D	18L	B738	B737-800 WING	240	3
4	01-01-08	A	06	B738	B737-800 WING	400	5
5	01-01-08	A	18R	B742	B747-400F	380	1

3.2.4 Merging the data

The chosen time span for the new dataset was January 1st, 2008 until March 1st, 2016. This range was chosen because it corresponds with all of the data that was available, except for the maintenance data from BAS. This resulted in some missing values for the maintenance data, which is further discussed in section 3.4.1.

Before the Schiphol data was merged with any of the other datasets, the format of the Schiphol Airport data needed to be changed. In the Schiphol Airport data there could be one to 500 rows for each day in the dataset. This is because there are many different combinations of runway and arrival or departure. For example, on January 1st there could be 20 rows which all have the same runway and arrival or departure but they have different airplanes arriving on each runway. An example of this is provided in table 6 where the first three rows have the same date, same runway code and they are all departures. However, all the other features are different. The problem of having so many rows with the same values for runway, date and arrival or departure is that it does not give a good overview of the total amount of complaints. Since this study is predicting the number of complaints per day, per runway and per arrival or departure, the format of the data will therefore need to be changed. For each day, only one row per day, runway and arrival or departure combination was made. Since there are 12 different runways and it can be either arrival or departure, this means that there are 24 rows per day in total (see appendix E). When changing the format, the number of flight for a certain date, runway and arrival or departure combination were added so that the number stayed accurate, which is shown in table 7. Table 6 shows that the first three rows that have the same date, arrival or departure and runway combination. However, these three rows will need to be reduced to one row that shows that exact date, arrival or departure and runway combination. The MTOW sum and number of flights of table 6 are all added together for the same date, runway and arrival or departure combination so that only one row in table 7 remains with the correct amount of MTOW sum and number of flights.

Table 6. Example of the Schiphol Airport data to show the format.

	Date	Arrival/ Departure	Run-way Code	ICAO type	Description	MTOW sum	Nr. of flights
1	01-01-08	D	09	A319	Airbus A319-1	77	1
2	01-01-08	D	09	A319	Airbus A319-1	693	9
3	01-01-08	D	09	B738	B737-800 WING	240	3
4	01-01-08	A	06	B738	B737-800 WING	400	5
5	01-01-08	A	06	B742	B747-400F	380	1

Table 7. Example of how the dataset looks like after the format was changed.

	Date	Arrival/ Departure	Runway code	MTOW sum	Nr. of flights
1	01-01-08	D	09	1010	13
2	01-01-08	A	06	780	6

By changing the format, the ICAO type and the description of the airplanes had to be removed, since it was not possible to accurately sum up all the ICAO types and descriptions per day, runway and arrival or departure. However, because the MTOW takeoff weight and number of flights have been summed up, this should make up for the loss of these features. Especially because the MTOW sum already shows which aircrafts were used (because the weights are different for each airplane). More importantly, by changing the format it is possible to do analyses of the day and the usage of the runway per arrival and departure. Once the format was changed, the KNMI data was merged with the Schiphol data. After that, the complaints and maintenance data from BAS were also added. An example of how the actual dataset looks like could be viewed in Appendix E.

3.2.5 Extra features and dataset description

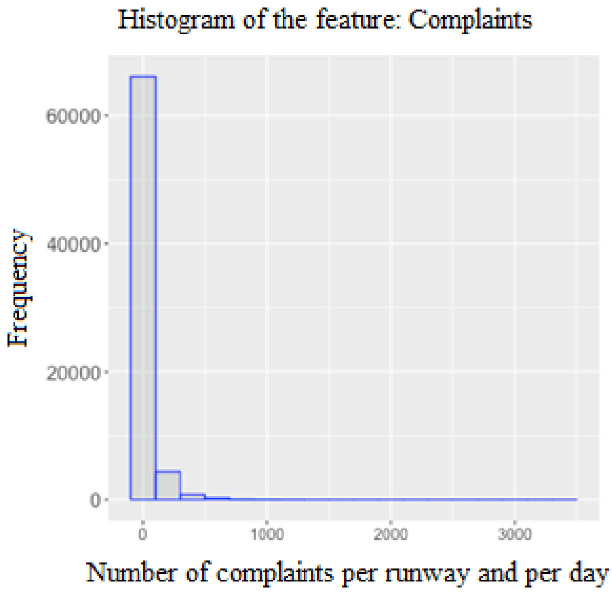
New features that were added to the dataset are 'Weekday' and 'Season'. There could be more complaints in summer as it is one of the busiest times of the year at Schiphol Airport (BAS, 2015). Furthermore, individuals may be more inclined to complain during the weekends, because they will be home more often and have more time to report a complaint. Thus, the feature weekday is also included in the dataset. Consequently, the entire dataset contains 38 features, including the target feature and 54122 rows. Of these features, five are categorical and 33 are continuous (see appendix A). To gain a visual understanding of the runways at Schiphol Airport and which code belongs to which runway, this can be seen in Appendix F.

3.3 Exploratory data analysis

In order to obtain a better understanding of the data and its underlying patterns, the dataset is further explored by using R, which is a programming environment to do statistical computing with (Rproject, 2016). A histogram of the dependent feature, complaints, can be found in figure 1, which shows that this feature is very much skewed to the right. The blue lines show that there are data points on the far right of the x-axis but that they are very small compared to smaller number of complaints that occurred more than 60000 times. This is expected, as there will only be few complaints each day. Yet, highly skewed features could potentially be more difficult in terms of analysis, so it will be logarithmically transformed. This is further discussed in section 3.4.1.

Descriptive statistics of the dependent feature confirm that it is skewed, as 50 percent of the data show that the number of complaints were between zero and ten. Yet the maximum value is 3343 and the mean value is 26.74. This demonstrates that there are many outliers. The standard deviation is also rather high: 82.75, which means that there is much variation in the frequency of complaints. The median is zero, which is probably because there are many days that no complaints were reported. Other features that showed much skewness are MTOW sum and number of flights, which are shown in Appendix G. The other features showed varying distributions, though none of them were normally distributed.

Figure 1. Histogram of feature: complaints. This shows that the feature is skewed to the right.



Before any of the other continuous features were examined, a correlation matrix of all the continuous features was made to find out what kind of relationships exist among the independent features and in relations with the dependent variable. The independent features that showed the highest correlation with the dependent feature were the maximum takeoff weight sum, which has Pearson correlation of 0.637 and the number of flights, which yielded a Pearson correlation of 0.6422. Since the correlation is difficult to examine in a regular scatterplot, figure 2 shows the features on a log-log scale which demonstrates the moderate relationship between the two features. Some of the independent features show strong relationships with each other, because they have a Pearson correlation that is higher than 0.70. For a full list of these features, see Appendix H. Multicollinearity does not necessarily have to be a problem in a dataset but it could affect the interpretability of the model's coefficients.

Figure 2. Scatterplot of the correlation between number of flights and complaints on a log-log scale

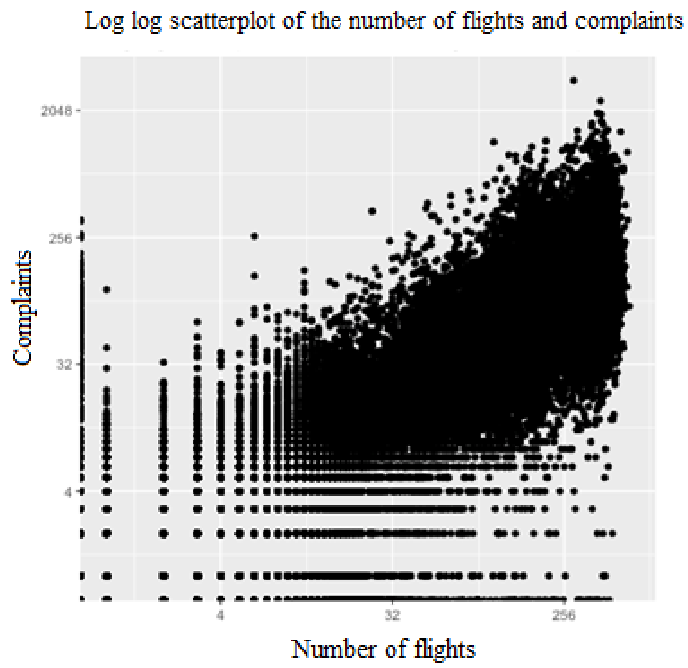
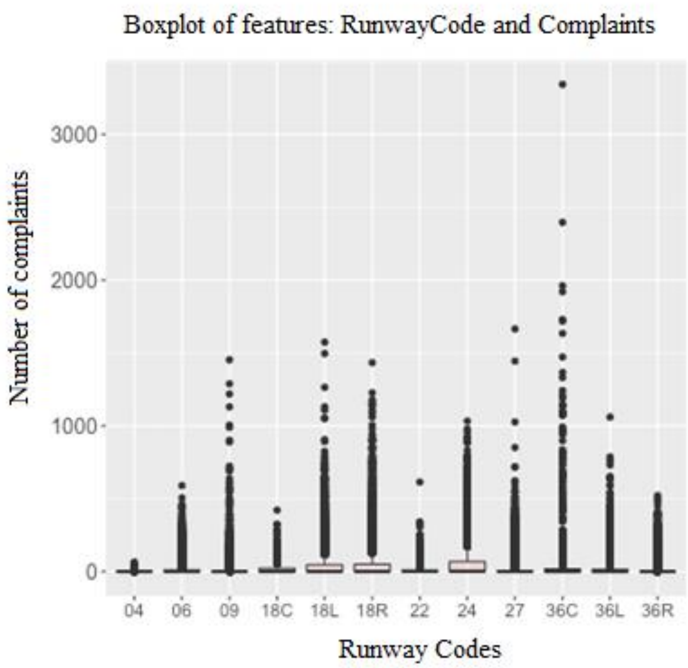


Figure 3. Boxplot of the relation between the separate runways and complaints



Since the complaints are predicted per runway used for arrival or departure, it could be interesting to examine how arrival and departure affect the number of complaints. As can be inferred from figure 3, both runway number 36C and 09 demonstrate that there are many outliers. Runway 24 probably receives most of the complaints as the mean is also much higher than the other runways except for runway 22. Figure 4 shows that there are many more outliers for departing airplanes than for arriving ones. Apart from that, arrival and departure show similar boxplots.

3.4 Pre-processing

3.4.1 Missing values

The complaint data only consists of reported complaints. This may seem obvious, but it means that there were many dates that were not included in this dataset because there were no reported complaints on that day. Instead of showing this as a missing value in the dataset, the choice was made to report these dates in the dataset as zero complaints. This is done because no instances on a specific date means that there were no complaints.

The feature number of flights contains data points of all the flights. This means that for this dataset there were also no instances reported for when there were no flights. Sometimes a certain runway is not used because of maintenance, or an airplane does not depart from a certain runway because of weather conditions. Thus, this data is not missing from the dataset but it is also not reported because it only shows data from when there were flights. For the dates that no flights were reported, the values were set at zero.

The feature MTOW sum only contains values when there was a flight on a certain day. When there were no flights, there was also not a maximum takeoff weight. No data points in this case also means that the runway remained unused and that there was no MTOW sum. Therefore, these values are zero as well.

A dataset that does contain missing values is the maintenance data. This dataset only contains instances from 2010 onwards. This means that for the years 2008 and 2009 there was no data available. Thus, these years are reported as missing values (NA's).

3.4.1 Logarithmic transformation

As was mentioned before, some of the features are highly skewed. A problem of skewed features is that it can make it more difficult to find patterns in the dataset (Feng et al., 2014). Logarithmic transformation is a popular method to reduce skewness. Logarithmic transformations were therefore applied to the features in the dataset that are highly skewed, because this could help improve the feature's distributions. These features are: complaints, MTOW sum and number of flights. Thus, these features will henceforth be referred to as complaints transformed, MTOW sum transformed and number of flights transformed so no confusion could arise about whether the feature was transformed or not. The other features in the dataset did not show as much skewness and were therefore not transformed. Since there are many zeros in the dataset, an extra constant needs to be added to the features before they are transformed (Carroll & Ruppert, 1988). The equation for logarithmic transformation with base 2 and the constant, c, is as follows:

$$f(x_i) = \log_2(x_i + c)$$

Where x_i is the feature that is going to be transformed and f is the transformation function. Often, c becomes a one or another constant value. The c in this dataset was set at one, because it is standard score. The descriptive statistics of the logarithmic transformed features can be found in table 8, which shows that because of the logarithmic transformation, the larger data points become much smaller whereas the smaller data points become slightly larger. The feature MTOW sum and MTOW transformed especially shows this very well. Even though the standard deviation of this feature is 9086.64 and much higher than the standard deviations of the other features, the logarithmic transformation transforms it and causes it to be much closer to the other features. This makes the distribution of the features more normal. Therefore, the logarithmically transformed features were kept in the dataset and the other non-transformed versions were removed.

Table 8. Descriptive statistics of features with and without logarithmic transformation.

	Mean	**Standard deviation**	**Median**	**Min**	**Max**
Complaints. transformed	1.213	1.819	0	0	6.681
Nr.flights. transformed	1.496	2.161	0	0	6.339
MTOWsum. transformed	3.119	4.133	0	0	10.810
Complaints	26.74	82.753	0	0	3343
Nr. Flights	49.40	108.001	0	0	565
MTOWsum	4092	9086.64	0	0	49300

3.5 Description of experimental procedure

3.5.1 Multivariate linear regression

The dataset is divided in three different parts, a train set (60%), validation set (20%) and test set (20%) which is according to traditional machine learning standards which has the purpose of preventing the models from overfitting (Shmueli, Patel & Bruce, 2010). The train set is used to fit the model. This should give objective evaluation scores (which are further discussed in the subsequent section, 3.5.2), because the test set has not been used for any model training or parameter tuning. Thus, the predictive performance of the classifier on this dataset is assessed by using the test set.

Predicting the number of complaints per runway used for arrival or departure is a regression task. One of the most basic algorithms to solve this regression task is multivariate linear regression. Multivariate regression goes a step further than linear regression by introducing two or more independent variables, which gives the equation:

$$Yi = \beta 0 + \sum_{i=1}^{n} \beta_i X_i + e_i$$

Where Y_i is the dependent variable, which is complaints transformed in this case. $\beta 0$ Is the intercept and β_i is the regression coefficient and X_i is the independent variables. When finding out which line would best fit the data, the fitted values of the multiple linear regression line should be as close as possible to actual observed data points. In this case, e_i is the difference between the actual dependent variable and the predicted dependent variable, and is also known as the residual error (Field, 2005).

The first model that is build is the intercept model, model 0. The intercept model predicts the mean of the dependent feature while all the independent features are zero. A baseline is necessary to compare the performance of each model against. The second model, model 1, consists of all the independent features. As was discussed before in section 3.3, there is some multicollinearity present in the dataset. Since this makes it more difficult to correctly interpret the coefficients of the features in the model, each feature is tested separately. For instance, one of the models could include all the independent features, with the exception of the feature maintenance. In total, 37 models are tested because there are 37 independent features. Only the model without the feature number of flights transformed showed interesting results. The results of this model are discussed in section four and five. The evaluation scores of the other models are presented in Appendix I.

Table 9. The models that will be tested and evaluated.

Models	Selected features
Model 0	Intercept model, features are set at zero.
Model 1	All independent features
Model 2	All independent features without the feature nr.flights.transformed

3.5.2 Evaluation criteria

To assess whether a model is able to predict new data points, several performance evaluation metrics are employed. One of the evaluation metrics that is used is r-squared. R-squared is a statistical measure that shows how well a model can account for the variance in the dependent feature y. The r-squared scores range from zero to one, with one being the highest. The equation for r-squared is:

$$r^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}$$

Where $\hat{y}_i$ are the predicted values and $\bar{y}_i$ is the mean. $\hat{y}$ - $\bar{y}$ takes the distance from the predicted values to the mean. y_i stands for the actual value of the dependent features. The mean is subtracted from the actual features. Then $(\hat{y}_i - \bar{y}_i)$ and $(y_i - \bar{y}_i)$ are squared and the values are summed by $\sum_{i=1}^{n}$.

The mean absolute error (MAE) is also used as a performance evaluation metric, as this performance metric could explain more about the accuracy of the model. The equation of MAE is:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} abs(y_i - \hat{y}_i)$$

Where the sum of the absolute differences between the predicted complaints and the actual complaints are divided by all of the data points. y_i stands for the actual values, whereas $\hat{y}_i$ represents the predicted values. The absolute difference of these values are summed by $\sum_{i=1}^{n}$. This number is divided by all the data observations with $\frac{1}{n}$.

Another measure that is used to assess the accuracy of the model is the root mean squared error (RMSE). The RMSE is more sensitive to larger errors and variance in the error compared to the MAE, because the MAE focuses on the mean error (Chai & Draxler, 2014). If there is much variability in the errors and the size of the errors is large, the RMSE score increases as well. It is therefore better to use RMSE with errors that show a normal distribution, compared to the MAE. Ultimately, the RMSE score should preferably as low as possible. The equation of RMSE is:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Where the square root is taken from the average squared difference between the predicted complaints and the actual complaints. In this equation, the difference between the predicted value, $\hat{y}_i$ and the actual values, y_i are squared. The sum, $\sum_{i=1}^{n}$ is taken of these values and divided by the number of data points, $\frac{1}{n}$, to obtain the average squared difference. Then, the square root is taken from this difference. Since some of the features were logarithmically transformed in order to reduce skewness, the logarithmic transformation will be undone after the predictions are made. This makes it easier to interpret the RMSE and MAE scores.

4. Results

4.1 Outline

This section reports the results that were obtained from the models. First, the evaluation results of the analyses are presented and briefly explained. After that, the coefficients of the model are examined and shown in tables 9 and 10 and figure 4.

4.2 Evaluation of results

The purpose of this study is to predict the number of complaints per day and per runway used for arrival or departure and to find out the predictive power of each of the independent features. The focus of previous research has been on associating noise levels and demographics, so this study takes an alternative approach by predicting the number of complaints based on a variety of features instead. The evaluation scores of all the models are shown in table 9. The evaluation results of all models that were applied to the dataset are shown in Appendix I.

Both RMSE and MAE employ the same unit scale as the dependent feature complaints. Consequently, both RMSE and MAE can be compared with other. The RMSE score on the test set of model 0 is 3.1564. The MAE score of the model's test set is 1.751 which is somewhat better than the RMSE score. Since RMSE penalizes larger values, and because there were many outliers in the dataset, this is to be expected (Chai & Draxler, 2014). Both the RMSE and MAE scores are relatively low, so there is probably not much variation in the size of the errors.

Model 1, which includes all the independent features, yields an r-squared score that is relatively high ($r^2 = 0.9200$). The score is slightly higher than on the model's train set and validation set. This could indicate that the model is not complex enough and that it is under fitting. The RMSE score of model 1's test set is 36.1949. This is much higher than the RMSE score of the baseline model, model 0, with a difference of 32.9459 (36.0864 - 3.1405). It is likely therefore that model 1, despite its high r-squared score, actually does not predict very well. The high score for r-squared could perhaps be a result of the many independent features that are included in the model. Each independent feature could potentially artificially increase the score of r-squared, which will therefore yield a higher r-squared although it does not improve predictions of the values. The MAE score of model 1 is 15.6534. This MAE score means that there is an average difference of around 16 complaints between the actual complaints and the predicted number of complaints. The difference between the MAE of the test set and the train set is 0.3452. Such smaller differences between the train, validation and test set could indicate that the model will obtain consistent scores on new data. It does not necessarily mean that it will obtain good prediction scores, just that the MAE scores will be relatively constant. Compared to the baseline model, the MAE score of this model is also much worse, with a difference of 13.9019 (15.9986 - 15.6534). Since the MAE takes the average of the error size between the actual and the predicted values, this score could indicate that there were many large errors.

Model 2 consists of all the independent features, but without the number of flights. This model showed the biggest differences with model 1. Just like model 1, the evaluation scores are better on the test set than on the train and validation set. The score of r-squared is 0.8920, compared to 0.9200 of model 1. This is a little bit lower, but still very close to the 0.9200 of model 1. The differences are better noticeable when observing the RMSE and MAE score. Both the scores have decreased, which is positive. Especially the RMSE score has lowered much, since it went from 36.0864 in model 1 to 25.2798 in this model. This indicates that the feature number of flights is responsible for much of the variation in the size of the errors, especially because RMSE penalizes outliers. This is further exemplified with the MAE score. The MAE for model 2 is also lower than it is for model 1, but the difference between the two is only 3.2877 (15.653 - 12.3657). By excluding the number of flights transformed, the average difference between the actual values and the values that are predicted by this model decreases with roughly three complaints. Even though model 1 has a better r-squared score, this model would be preferred because it has better accuracy than model 1.

This was the only model that showed some differences in terms of evaluation scores with model 1. All the other independent features demonstrate much smaller effects on the number of complaints. It is therefore likely that from all the independent features, the number of flights transformed has the best predictive power. To answer the second research question more in depth, table 10 provides insight in the five strongest regression coefficients of model 1. A graphical representation of all the coefficients of model 1 can be found in Appendix J. Table 10 shows that the runway feature has the second strongest coefficients scores after the feature number of flights transformed. Runway is a discrete feature that consists of various levels. The coefficients therefore need to be interpreted differently than the continuous features. In this case, Runway 4 is used as the default level and the other coefficients show how much they change in comparison with Runway 4. It is understandable that the feature runway has better scores as the complaints are described in the dataset per usage of a particular runway and per arrival and departure. Furthermore, some runways such as 18L are used very often and have to process many flights on a day, especially when maintenance is planned. Since the number of flights is indicative of the number of complaints, it is also more likely that runways that are used more often will receive more complaints. As can be inferred from Appendix J, MTOW sum does not have a very strong coefficient. This is not as expected because MTOW sum has a moderate relationship with the dependent feature and a strong relationship with the number of flights.

Table 9. Results of multivariate linear regression models

	r^2	*RMSE*	*MAE*
Model 0 train	0	3.1564	1.7515
Model 0 validation	0	3.1596	1.7555
Model 0 test	0	3.1405	1.7372
Model 1 train	0.9170	36.9731	15.9986
Model 1 validation	0.9185	37.2690	16.1167
Model 1 test	0.9200	36.0864	15.6534

Model 2 train	0.8865	25.4326	12.4541
Model 2 validation	0.8864	25.3871	12.4563
Model 2 test	0.8920	25.2798	12.3657

Table 10. Coefficient score of the strongest coefficient scores of model 1.

Feature	Coefficient score
nr.flights.transformed	0.8082
RunwayCode18L	0.3628
RunwayCode27	0.3099
RunwayCode09	0.2323
RunwayCode09	0.2325

5. Discussion

5.1 Outline

This section explores the results and limitations of this study. The goal of this research was to predict the number of complaints per day and per runway used for arrival or departure. Moreover, this study aimed to find out which features are the best predictors of the number of complaints. Since previous research has mostly focused on associating noise and personal characteristics data, this study focused on making predictions using multivariate linear regression instead.

5.2 Discussion of results

Model 1, which includes all the independent features of the dataset, showed the highest r-squared score, which is 0.9200. The RMSE and MAE scores, however, were relatively high with a RMSE of 36.0864 and a MAE of 15.6534 respectively. Despite its high r-squared score, this model therefore does not predict very well. Especially compared with the baseline, the scores of the RMSE and MAE are much worse. One of the explanations for such high RMSE and MAE scores is that there is much variation in the difference between the observed values and the predicted values.

Model 2 does not include the feature number of flights transformed and shows that the performance of the r-squared is a bit worse. However, RMSE and MAE perform somewhat better with a score of 25.2798 for RMSE and MAE has a score of 12.3657. Without the feature number of flights transformed, the predictions for MAE are better with an average of three complaints. The number of flights transformed probably has much variation in the size of its values, because removing it has a positive impact on the RMSE and MAE. Since removing number of flights transformed increases yields better accuracy this model is therefore better than model 1. The improved accuracy outweighs the smaller r-squared score and shows that it is better at predicting new data than model 1. It should be noted, however, that the scores are still much higher than the baseline. Unfortunately, the model is still not very good at predicting the number of complaints.

Since the correlation matrix showed that there existed multicollinearity between some of the independent features (see Appendix H), various models were tested with each excluding one of the features. Almost all of the features showed results that were similar to model 1. Except for model 2, which did not include the number of flights transformed. It was expected that excluding MTOW sum transformed would have more of an effect on the evaluation scores. Especially because this feature showed a strong relationship with the number of flights and because it has a moderately positive relationship with the dependent feature. A possible explanation could be that this feature will only have more effect on the evaluation scores when it is used in combination with for instance the feature number of flights transformed or the runway. On its own, however, it is not as influential as expected. In general, the independent features have not shown much predictive power and has actually made the accuracy of the model worse than the baseline.

5.3 Limitations

One of the limitations of this study is that it makes use of complaints that have been reported by people. Often, people experience annoyance or stress because of noise, but they decide not to do anything about it (Wirth, Brink & Schirz, 2003). This could be due to a number of reasons, but previous research has suggested that people do not file complaints because they feel that doing so is useless. There are also individuals who complain very often and who are known as "serial complainers" (Hume et al., 2003). This is an issue that researchers have experienced before when investigating complaints (Hume et al., 2003). Of the complaint data, 71 percent of the people who reported a complaint in 2015 was a serial complainant (BAS, 2016). The percentages of complainants in previous years were much less and around 30 - 40 percent (BAS, 2016). With the data that was available, however, it was not possible to distinguish the complainants. This could potentially have introduced some bias in the dataset and it gives a less accurate picture of how individuals in general experience noise.

Another limitation in terms of data is that the number of features is restricted. This study could probably have benefited from other features such as the asphalt of the runways, airplane movements and other things that could affect aircraft noise and other reasons why people complain. In particular, the date feature is also limited because it only contains all the complaints per day. Since people will probably complain more in the evenings than during the day, or late at night when they are unable to sleep because of the aircraft noise, it could be useful to predict the number of complaints per hour. It has not been possible to use this type of data, so this could be investigated upon in future research.

Furthermore, this study is limited in the models that it uses. The models that have been applied to the dataset are relatively simple ones. Multivariate linear regression is among the more basic models and no parameters were tuned. The lack of complexity showed in the evaluation scores, because the models were under fitting and the RMSE and MAE scores were much lower than the baseline. Better results could be achieved by tuning parameters or by using a different model altogether

5.4 Implications

Previous research in this field does not include any predictions on the number of complaints. This study took a novel approach by predicting the number of complaints using multivariate linear regression, instead of associating complaints and noise with each other (Wirth, Brink, Schierz, 2003). The outcome of this study shows that predicting the number of complaints has not resulted in very good evaluation scores. Even though the r-squared is scoring relatively well on model 1 with a score of 0.9203, the RMSE and MAE are much higher than the baseline scores. Therefore, this study was not able to obtain good accuracy scores.

Furthermore, this study has made a start with creating a new dataset from various data sources to predict the number of complaints. One of the novel features of this study is that instead of relying on levels of noise and personal characteristics, other features were included in the dataset as well. This has brought some limitations, as mentioned in the section 5.3, but it also brings new insights. Some of the features do not have as much of an effect as was hoped for, such as the weather features, but it still shows that the number of flights and the runways have some predictive power. This research has also shown that people complain mostly because of noise levels. After all, the number of flights and runway are all related to the amount of noise that people experience. In this sense it relates back to the research that was carried out by Hume et al. (2003), who argued that noise was the main factor of people's reasons to complain. Other researchers focused on other features, by stating that personal characteristics or being sensitive to noise are more important when investigating aircraft noise complaints (Wirth, Brink & Schirz, 2003). Considering these differences in research by researchers about the importance of features relating to the number of complaints this study could thus contribute to the existing body of literature by showing that noise features, but also maintenance or date could play a role in predicting the number of complaints.

For further research it is therefore recommended that instead of focusing on personal characteristics and complaints, the scope of the features could be extended even further/ to extent the scope of the type of features even further. It is possible that there are many other features that could also be important and affect the predictive power in a positive manner. It would therefore be interesting to include these and other features in further research.

This study has made a start with creating a new dataset from various data sources to predict the number of complaints. Despite the lower evaluation scores, it could still be valuable to continue with predicting the number of complaints for further research. Using complex models or tuning the parameters of the model on the validation set could yield other interesting insights about the features and the data. Once the data is improved upon, predicting the number of complaints could still be useful for a number of stakeholders, such as airline companies or governments.

6. Conclusion

This section presents a conclusion to the research questions. Moreover, the results are placed within the context of the existing research and future recommendations are given. The research questions that this study aimed to answer were:

1. To what extent can the number of complaints be predicted per runway, per day and per arrival or departure at Schiphol Airport?
2. Which features have the most predictive power in predicting the number of complaints per runway, per day?

The first research question was answered by creating a dataset that contains features that could affect the number of complaints. Data was received from Intermediator Residents Schiphol

(BAS), Royal Netherlands Meteorological Institute (KNMI) and from Schiphol Airport. They were all put together in one dataset. In order to obtain the best possible predictive scores and to reduce the effects of multicollinearity in the dataset, several models were made. The best results were obtained with model 2, which includes all the independent features except the feature number of flights transformed. Even though this model's r-squared score is slightly lower than model 1's r-squared score, the RMSE and MAE are much better. Therefore, this model has the best accuracy in predicting new data. The other models that were tested did not show much differences with model 1 However, the predictions of model 2 are still worse than the predictions that are made by the baseline model.

For future research it is therefore recommended that a model could be used that is more complex, especially because the models were under fitting. Multivariate linear regression is a relatively simply model and there is probably much predictive power to be gained by using other models or by tuning parameters on the validation set. Since this is the first study to predict the number of complaints, it would be very interesting to see how other models could improve the predictions that have been made in this study. In addition, predicting the number of complaints has also much practical value as predicting the number of complaints could for instance help airports rearrange their maintenance or flight schedules.

The second research question has been answered by testing the different models and by examining the regression coefficients of model 1 in a more detailed way. The number of flights and the runway are the best predictors of the number of complaints. Previous research has focused on noise and on personal demographics in order to associate noise with complaints. This study has shown that other features relating to noise have an effect on the number of complaints. It is recommended for further research that a wide variety of features are tested out and that other more advanced models are used to predict the number of complaints. Ultimately, this could yield a better picture of why people complain about aircraft noise.

7. References

Babisch, W., Houthuijs, D., Pershagen, g.,Cadum, E., Katsouanni, K., Velonakis, M., Dudley, M., Marohn, H., Swart, W., Breugelmans, O., Blum, G., Selander, J., Vigna-alianti, F.,Pisai, S., Haralaidis, A., Dimakopoulou, K., Zachos, I., Jarup, L. (2009). Annoyance due to aircraft noise has increased over the years - Results of the HYENA study. *Environment International, 3*, 1169-1176.

BAS. (2016). Bas Bewoners Aanspreekpunt Schiphol. Retrieved from: www.bezoekbas.nl.

Berglund, B. & Lindvall, T. (1995). Community Noise. *Archives of the Center for Sensory Research, 2*(1), 1 – 195.

Berry, B. & Sanchez, D. (2014). The economic and social value of aircraft noise effects: a critical review of the state of the art. 11th International Congress on Noise as a public health problem (ICBEN), Nara, Japan. Retrieved from: http://5.196.80.71/~uecnaeu/wp-content/uploads/2015/04/Berry-B-SanchezD072014.pdf

Correia, A. W., Peters, J. L., Levy, J. I., Melly, S., & Dominici, F. (2013). Residential exposure to aircraft noise and hospital admissions for cardiovascular diseases: multi-airport retrospective study. http://dx.doi.org/10.1136/bmj.f5561

Fidell, S. & Howe, R. (1998). *Use of airport noise complaint files to improve understanding of community response to aircraft noise* (Report No NASA/CR-1998 - 207650). Retrieved from National Aeronautics and space administration, Langley Research Center.

Filippone, A. (2014). Aircraft noise prediction. *Progress in Aerospace Sciences*, *68*, 27-63.

Franssen, E.A.M., Staatsen, B.A.M., Lebret, E. (2002). Assessing health consequences in an environmental impact assessment, the Case of Schiphol. *Environmental Impact Assesment Review, 22*, 633-653.

Franssen, E. A. M, van Wiechen, C.M.A.G., Nagelkerke, N. & Lebret, E. (2004). Aircraft noise around a large international airport and its impact on general health and medication use. *Occupational and Environmental Medicine, 61*(5), 405-413.

Frerichs, R.R., Beeman, B.L. & Coulson, A.H. (1980). Los Angeles airport noise and mortality -- faulty analysis and public policy. *American Journal of Public Health, 70*(4), 357-362.

Goines, L. & Hagler, L. (2007). Noise Pollution: a modern plague. *Southern Medical Journal, 100*(3), 287-294.

Hatfield, J., Job, R.F.S., Carter, N.L., Peploe, P., Taylor, R. & Morrell, S. (2001). The influence of psychological factors on self-reported physiological effects of noise. *Noise and Health, 3*(10), 1-13.

Hazard, W.R. (1971). Predictions of noise disturbance near large airports. *Journal of Sound and Vibration, 15*(4), 425-445.

He, Q., Wollersheim, C., Locke, M., Waitz, I. (2014). Estimation of the global impacts of aviation-related noise using an income-based approach. *Air Transportation and the Environment, 34*, 85-101.

Hjortebjerg, D., Andersen, A.M.N., Christensen, J., Ketzel, M., Raaschou-ielsen, O., Sunyer, J., Julvez, J., Forns, J., Sørensen, M. (2015). Exposure to road traffic noise and behavioral problems in 7-year-old-children: a cohort study. E*nvironmental Health Perspectives, 124*(2), 228 – 234. http://dx.doi.org/ 10.1289/ehp.1409430

Hume, K., Terranova, D., Thomas, C. (2002). Complaints and annoyance caused by aircraft operations: temporal patterns and individual bias. *Noise & Health, 4*(15), 45-55.

Hume, K., Martin, G., Thomas, C. & Terranova, D. (2003). Complaints caused by aircraft operations: an assessment of annoyance by noise level and time of day. *Journal of Air Transport Management, 9*(3), 153 - 160.

Kaltenbach, M., Maschke, C. & Klinke, R. Health Consequences of Aircraft Noise. *Deutsche Arzteblatt International, 105*(31-32), 548-556.

Lim, C., Kim, J., Hong, J., Lee, S. (2008). Effect of background noise levels on community annoyance from aircraft noise. *The Journal of the Acoustical Society of America, 123*(2), 766-71.

Maziul, M., Job, R.F., Vogt, J. (2005). Complaint data as an index of annoyance -- theoretical and methodological issues. *Noise and Health, 7*(28), 17-27.

Meecham, W.C. & Shaw, N. (1979). Effects of jet noise on mortality rates. *British Journal of Audiology, 13*(3), 77-80.

Meecham, W.C. & Shaw, N.A. (1979). Increase in mortality rates due to aircraft noise. *Schriftenreihe des Vereins fur Wasser-, Boden- und Lufthygiene, 88*, 428-441.

Morrell, S., Taylor, R., Lyle, D. (1997). A review of health effects of aircraft noise. *Australian and New Zealand Journal of Public Health, 21*(2), 221-236.

Netsajov, F. (2012). Contemporary Measures for noise reduction in airport surroundings. *Applied Acoustics, 73*(10), 1076 - 1085.

Provost, F. & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data, 1*(1), 51-59

The R Project for Statistical Computing. (n.d.). Retrieved July 12, 2016, from: https://www.r-project.org/

Upham, P., Maughan, J., Raper, D., Thomas, C. (2003). Towards Sustainable Aviation. *Earthscan Publications Ltd: New York.*

Shmueli, G., Patel, N.R., Bruce, P.C. (2010). Data Mining for Business Intelligence. *John Wiley & Sons: Hoboken: New Jersey.*

Steele, C. (2001). A critical review of some traffic noise prediction models. *Applied Acoustics, 62*, 271-287.

Suau-Sanchez, P., Pallares-Barbera, M. & Paul, V. (2011). Incorporating annoyance in airport environmental policy: noise, societal response and community participation. *Journal of Transport Geography, 19*(2), 275 - 284.

Wiechen, C. M. A.G van., Franssen, E.A.M., Jong, R.G. de, Lebret, E. (2002). Aircraft noise exposure from schiphol airport: a relation with complainants. *Noise and Health, 5*(17), 23-34.

Wirth, K., Brink, M., Schierz, C. Aircraft noise annoyance around airport Zurich – Kloten. Institute of Hygiene and Applied Physiology, Zurich, Switzerland.

Zaporozhets, O., Tokarev, V., Attenborough, K. (2011). *Aircraft Noise: Assessment Prediction and* Control. Boca Raton: CRC Press.

8. Appendices

8.1 Appendix A

List of the descriptions of all the feature in the dataset

	Features	**Feature type**	**Description**
1	Date	Categorical	Ranges from 01-01-2008 to 01-03-2016.
2	RunwayCode	Categorical	Consists of 24 codes for parts of the main airplane runways. These codes are: 04, 06, 09, 18C, 18L, 18R, 22, 24, 27, 36C, 36L and 36R.
3	AD	Categorical	Arrival or departure of an airplane
4	Complaints	Continuous	The number of complaints that were received
5	WindDirectionAv	Continuous	The average wind direction in degrees.
6	VectorWindSpeedAv	Continuous	Vector mean wind speed in 0.1 m/s.
7	WindspeedAv	Continuous	The vector mean wind speed in 0.1 m/s. This is the average of the wind speed calculated by using vectors, i.e. average wind speed of 10 m/s from the northwest.
8	HighWindspeedAv	Continuous	The highest average wind speed in 0.1 m/s.
9	LowWindspeedAv	Continuous	The lowest wind speed average in 0.1 m/s.
10	MaxWindGust	Continuous	The highest recorded wind gust in 0.1 m/s.
11	TempAv	Continuous	The average temperature in 0.1 degrees Celsius per day.
12	MinTemp	Continuous	The minimum temperature in 0.1. degrees Celsius per day.

13	MaxTemp	Continuous	The maximum temperature in 0.1 degrees Celsius per day.
14	MinTemp10cm	Continuous	The recorded minimum temperature at 10 cm above the ground in 0.1 degrees Celsius.
15	SunshineDur	Continuous	Duration of sunshine, measured in 0.1 hour. When the amount of sunshine was less than 0.05, the duration of sunshine was recorded as -1.
16	MaxPotSunshine	Continuous	The maximum potential amount of sunshine which is calculated in percentages.
17	GlobalRad	Continuous	The amount of global radiation that occurred on a day in J/cm2.
18	PrecipDur	Continuous	The amount of time that there was precipitation on a day in 0.1 hour.
19	PrecipAmount	Continuous	The amount of precipitation that occurred on a day in 0.1 mm. When there was less than 0.05 mm precipitation, the amount of precipitation was recorded as -1.
20	HighestPrecip	Continuous	The highest amount of precipitation that fell in an hour in 0.1 mm. When there was less than 0.05 mm, precipitation was recorded as -1.
21	PressureAv	Continuous	The daily average air-pressure which is which is expressed in sea level pressure in 0.1 hPa.*
22	MaxPressure	Continuous	The highest recorded level of air-pressure which is expressed in sea level pressure in 0.1 hPa*.
23	MinPressure	Continuous	The lowest recorded level of air-pressure which is expressed in sea level pressure in 0.1 hPa*.
24	MinVisibility	Continuous	The minimum amount of visibility. It is recorded in categories: 0: < 100m, 1 : < 100 -200m, 2: < 200-

			300 …, 81: 35-40 km, the highest category is 89: > 70 km.
25	MaxVisibility	Continuous	The maximum amount of visibility. It is recorded in categories: 0: < 100m, 1 : < 100 -200m, 2: < 200-300 …, 81: 35-40 km, the highest category is 89: > 70 km.
26	CloudAv	Continuous	The amount of cloud cover per day which is recorded on a scale of one to nine, with one the lowest and nine the highest. Nine means that the sky is completely covered in clouds and therefore invisible.
27	HumidityAv	Continuous	The maximum amount of humidity per day which is recorded in percentages
28	MaxHumidity	Continuous	The maximum amount of humidity in the atmosphere in percentages.
29	MinHumidity	Continuous	The minimum amount of humidity in the atmosphere in percentages.
30	Evap	Continuous	Potential amount of evapotranspiration (which is both transpiration from trees and evaporation from ocean's added together) in 0.1 mm.
31	MTOWsum	Continuous	Maximum takeoff weight sum consists of the maximum takeoff weight multiplied by the number of flights
32	Nr.flights	Continuous	The number of flights per runway code and per arrival or departure.
33	Weekday	Categorical	The day of the week.
34	Season	Categorical	Whether it was spring, summer, fall or winter
35	Degrees	Continuous	The wind direction (vectors) of the runways.
36	Length	Continuous	The length of the runways in meter

37	Weekend	Continuous	Whether it was weekend or not, outcomes are 0 and 1
38	Maintenance	Continuous	Whether there was planned maintenance on that date

* hPa stands for hectopascal.

Once the features: complaints, MTOW sum and number of flights were logarithmically transformed, the names of the features were changed as well and non-transformed features were removed from the dataset:

Feature	**Feature type**	**Description**
Complaints.transformed	Continuous	Logarithmically transformed complaints.
MTOWsum.transformed	Continuous	Logarithmically transformed MTOWsum
Nr.flights.transformed	Continuous	Logarithmically transformed Nr.flights

8.2 Appendix B

List of the features of the complaint data.

Feature	Description
Date	The date of this dataset ranges from 01-01-2008 to 01-03-2016
Runway	Runway consists of codes: 04, 06, 09, 18C, 18L, 18R, 22, 24, 27, 36C, 36L, 36R.
AD	Whether the airplane arrived or departed from a runway.
Complaints	The number of complaints per day, runway and arrival or departure.

8.3 Appendix C

List of the features of the maintenance data

Feature	Description
Date	The date of this dataset ranges from 24-05-2010 to 11-10-2015.
Maintenance	This feature shows whether there was maintenance on a certain day, with 1 for maintenance and 0 if there was not maintenance. Since the dataset only consists of the dates when there was scheduled maintenance, this feature consists completely of 1's.

8.4 Appendix D

List of all the features and a description of the KNMI data.

Features	Description
WindDirectionAv	Average of the wind direction in degrees. Consists of: 360= north, 90 = east, 180 = south, 270 = west.
WindspeedAv	The vector mean wind speed in 0.1 m/s. This is the average of the wind speed calculated by using vectors, i.e. Average wind speed of 10 m/s from the North West.
HighWindspeedAv	The highest average wind speed in 0.1 m/s.
LowWindspeedAv	The lowest wind speed average in 0.1 m/s.
MaxWindGust	The highest recorded wind gust in 0.1 m/s.
TempAv	The average temperature in 0.1 degrees Celsius per day.
MinTemp	The minimum temperature in 0.1 degrees Celsius per day.
MaxTemp	The maximum temperature in 0.1 degrees Celsius per day.
MinTemp10cm	The recorded minimum temperature at 10 cm above the ground in 0.1 degrees Celsius.
SunshineDur	Duration of sunshine, measured in 0.1 hour. When the amount of sunshine was less than 0.05, the duration of sunshine was recorded as -1.
MaxPotSunshine	The maximum potential amount of sunshine which is calculated in percentages.
GlobalRad	The amount of global radiation that occurred on a day in J/cm2.
PrecipDur	The amount of time that there was precipitation on a day in 0.1 hour.
PrecipAmount	The amount of precipitation that occurred on a day in 0.1 mm. When there was less

	than 0.05 mm precipitation, the amount of precipitation was recorded as -1.
HighestPrecip	The highest amount of precipitation that fell in an hour in 0.1 mm. When there was less than 0.05 mm, precipitation was recorded as -1.
PressureAv	The daily average air-pressure which is which is expressed in sea level pressure in 0.1 hPa*.
MaxPressure	The highest recorded level of air-pressure which is expressed in sea level pressure in 0.1 hPa*.
MinPressure	The lowest recorded level of air-pressure which is expressed in sea level pressure in 0.1 hPa*.
MinVisibility	The minimum amount of visibility. It is recorded in categories: 0: < 100m, 1 : < 100 -200m, 2: < 200-300 …, 81: 35-40 km, the highest category is 89: > 70 km.
MaxVisibility	The maximum amount of visibility. It is recorded in categories: 0: < 100m, 1 : < 100 -200m, 2: < 200-300 …, 81: 35-40 km, the highest category is 89: > 70 km.
CloudAv	The amount of cloud cover per day which is recorded on a scale of one to nine, with one the lowest and nine the highest. Nine means that the sky is completely covered in clouds and therefore invisible.
HumidityAv	The maximum amount of humidity per day which is recorded in percentages.
MaxHumidity	The maximum amount of humidity in the atmosphere in percentages.
MinHumidity	The minimum amount of humidity in the atmosphere in percentages.
Evap	Potential amount of evapotranspiration (which is both transpiration from trees and

	evaporation from ocean's added together) in 0.1. mm.
VectorWindSpeedav	Vector mean wind speed in 0.1 m/s.

* hPa stands for hectopascal.

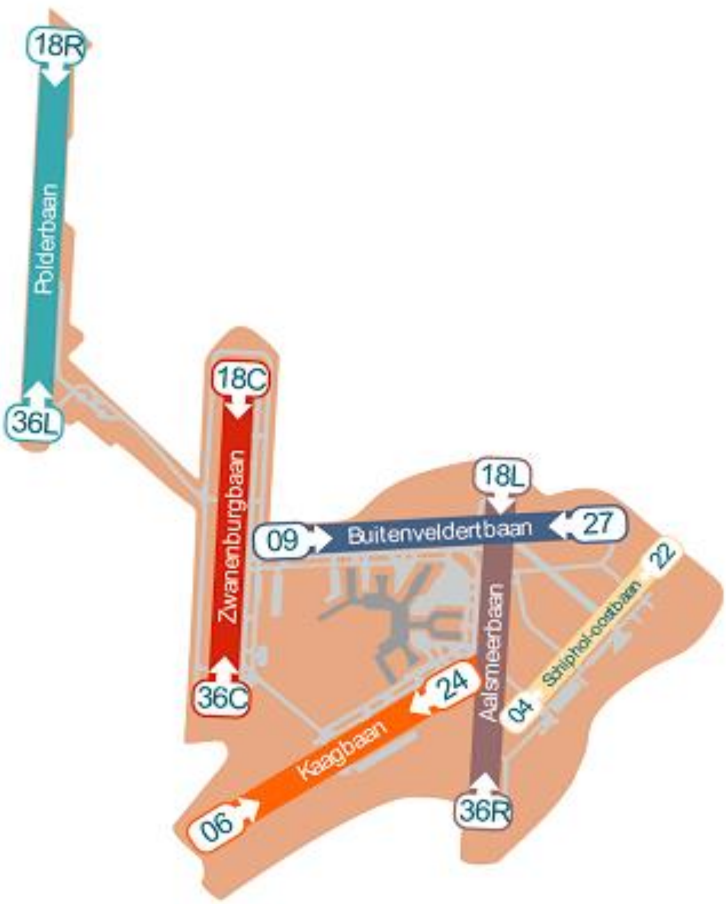
8.5 Appendix E

Example of the dataset with the new format.

	Date	Runway	AD	Nr. of Flights	MTOW sum	Complaints
1	01-01-08	04	A	0	0	0
2	01-01-08	04	D	0	0	0
3	01-01-08	06	A	119	10752	40
4	01-01-08	06	D	0	0	0
5	01-01-08	09	A	0	0	0
6	01-01-08	09	D	84	7243	274
7	01-01-08	18C	A	15	1110	0
8	01-01-08	18C	D	0	0	0
9	01-01-08	18L	A	0	0	0
10	01-01-08	18L	D	33	2444	83
11	01-01-08	18R	A	225	16890	556
12	01-01-08	18R	D	0	0	0
13	01-01-08	22	A	7	472	0
14	01-01-08	22	D	5	353	0
15	01-01-08	24	A	0	0	0
16	01-01-08	24	D	218	17108	55
17	01-01-08	27	A	2	148	0
18	01-01-08	27	D	0	0	0
19	01-01-08	36C	A	0	0	0
20	01-01-08	36C	D	0	0	0
21	01-01-08	36L	A	0	0	0
22	01-01-08	36L	D	68	5853	72
23	01-01-08	36R	A	0	0	0
24	01-01-08	36R	D	0	0	0
25	02-01-08	04	A	356	7	0

8.6 Appendix F

Runway Code consists of the runway codes of each runway at Schiphol. This image gives a graphical representation of the runway codes of the runways at Schiphol Airport (BAS, 2016).



8.7 Appendix G

Histograms of features: maximum takeoff weight sum, maximum takeoff weight average and number of flights

Figure 1. Histogram of the total maximum takeoff weight (MTOW sum).

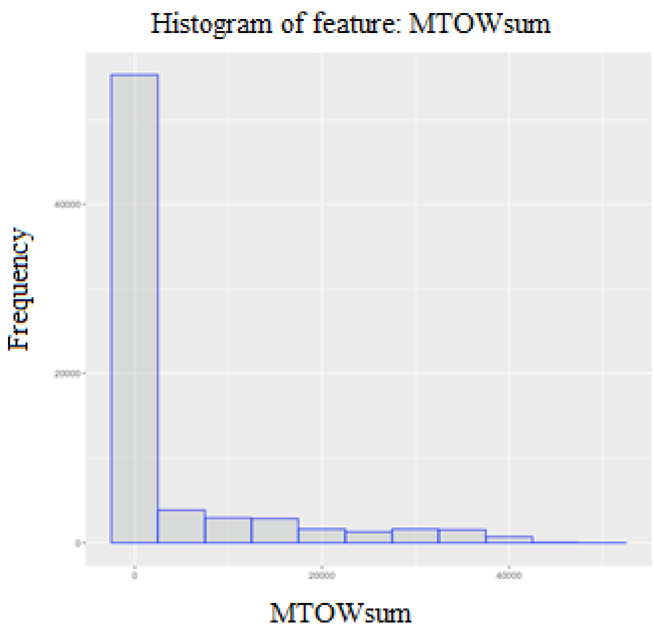
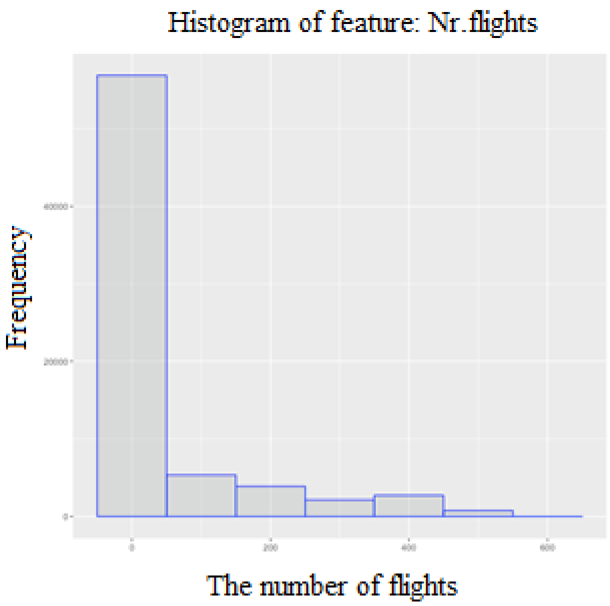


Figure 3. Histogram of the number of flights.



8.8 Appendix H

Tables of the features that showed multicollinearity (Pearson correlation > 0.70). The features that did not show a strong relationship among each other were left blank.

	WindspeedAv	**LowWindSpeed - Av**
VectorWindspeed-Av	0.9654	0.8494
HighWindspeedAv	0.9215	
LowWindspeedAv	0.8582	
MaxWindGust	0.8873	
MinTemp10cm		
WindspeedAv		0.8582

	MinTemp-10cm	**Precip-Amount**	**Pressure Av**	**Cloud Av**
TempAv	0.8916			
MinTemp	0.9746			
MaxTemp	0.8063			
PrecipDur		0.7551		
HighestPrecip		0.8517		
SunshineDuration				0.8538
MaxPotSunshine				0.8757
MaxPressure			0.9744	
MinPressure			0.9763	

	Complaints. transformed	Nr.flights. transformed	MTOWsum. transformed
Complaints. transformed			
Nr.flights. transformed	0.9548		
MTOWsum. transformed	0.9305	0.9780	

8.9 Appendix I

Results of all the models that were tried. First there is a list of a description of which feature was excluded from which model. After that, all the results are shown. The model that excludes the number of flights is not included in this list, because it has already been discussed in section 4 and 5 of this research.

Model	Without feature:
1	MTOW.sum.transformed
2	Weekend
3	Length
4	Degrees
5	Season
6	Weekday
7	Evap
8	MinHumidity
9	MaxHumidity
10	HumidityAv
11	CloudAv
12	MaxVisibility
13	MinVisibility
14	MinPressure
15	MaxPressure
16	PressureAv
17	HighestPrecip
18	PrecipAmount
19	PrecipDur
20	GlobalEad
21	MaxPotSunshine
22	SunshineDur

23	MinTemp10cm
24	MaxTemp
25	MinTemp
26	TempAv
27	MaxWindGust
28	LowWindspeedAv
29	HighWindspeedAv
30	WindspeedAv
31	VectprWindspeedAv
32	WindDirectionAv
33	AD
34	RunwayCode
35	Maintenance
36	Date

The evaluation scores of the models:

	r^2	***RMSE***	***MAE***
Model 1 train	0.9179	36.9329	15.9875
Model 1 validation	0.9185	36.9392	16.0257
Model 1 test	0.9200	36.1949	15.6829
Model 2 train	0.9170	36.9731	15.9986
Model 2 validation	0.9185	37.2690	16.1167
Model 2 test	0.9200	36.0864	15.6534

Model 3 train	0.9178	36.8483	15.9622
Model 3 validation	0.9175	37.7467	16.3061
Model 2 test	0.9186	35.9918	15.5770
Model 4 trainset	0.9170	36.9709	15.9984
Model 4 validation	0.9185	37.2559	16.1143
Model 4 test	0.9199	36.0926	15.6539
Model 5 train	0.9169	36.9513	15.9940
Model 5 validation	0.9185	37.2265	16.1087
Model 5 test	0.9198	36.0332	15.6428
Model 6 trainset	0.9169	36.9687	15.9974
Model 6 validation	0.9185	37.2594	16.1143
Model 6 test	0.9200	36.0872	15.6535
Model 7 train	0.9169	36.9687	15.9974

Model 7 validation	0.9185	37.2594	16.1143
Model 7 test	0.9200	36.0835	15.6535

Model 8 train	0.9170	36.9732	15.9986
Model 8 validation	0.9185	37.27016	16.1166
Model 8 test	0.9200	36.0835	15.6520

Model 9 Train	0.9171	36.9724	15.9985
Model 9 validation	0.9195	37.2637	16.1156
Model 9 test	0.9200	36.0942	15.6547

Model 10 trainset	0.9170	36.9740	15.9987
Model 10 validation	0.9185	37.2693	16.1167
Model 10 test	0.9200	36.0920	15.6533

Model 11 train	0.9170	36.9729	15.9985
Model 11 validation	0.9185	37.2510	16.1170
Model 11 test	0.9200	36.0836	15.6527

Model 12 trainset	0.9170	36.9687	15.9974
Model 12 validation	0.9185	37.2594	16.1143
Model 12 test	0.9200	36.0872	15.6535

	r^2	*RMSE*	*MAE*
Model 13 train	0.9170	36.9720	15.9984
Model 13 validation	0.9185	37.2450	16.1122
Model 13 test	0.9200	36.0538	15.6460
Model 14 train	0.9170	36.9713	15.9980
Model 14 validation	0.9185	37.2690	16.1167
Model 14 test	0.9200	36.0853	15.6532

Model 15 train	0.9170	36.9691	15.9970
Model 15 validation	0.9185	37.2731	16.1175
Model 15 test	0.9200	36.0875	15.6534

Model 16 trainset	0.9170	36.9715	15.9981
Model 16 validation	0.9185	37.2729	16.1173
Model 16 test	0.9200	36.0864	15.6534
Model 17 Train	0.9170	36.9725	15.9983
Model 17 validation	0.9185	37.2763	16.1177
Model 17 test	0.91200	36.0864	15.6534
Model 18 trainset	0.9169	36.9737	15.9986
Model 18 validation	0.9185	37.2725	16.1169
Model 18 test	0.9200	36.0865	15.6534
Model 19 train	0.9169	36.9732	15.9986
Model 19 validation	0.9185	37.2702	16.1169
Model 19 test	0.9200	36.0858	15.6533

Model 20 train	0.9170	36.9708	15.9973
Model 20 validation	0.9185	37.2616	16.1143
Model 20 test	0.9200	36.0891	15.6534
Model 21 Train	0.9170	36.9715	15.9984
Model 21 validation	0.9175	37.2701	16.1170
Model 21 test	0.9200	36.0855	15.6532
Model 22 trainset	**0.9170**	**36.9732**	**15.9986**
Model 22 validation	**0.9185**	**37.2690**	**16.1167**
Model 22 test	**0.9200**	**36.0864**	**15.6534**
Model 23 train	0.9170	36.9729	15.9985
Model 23 validation	0.9185	37.2633	16.1157
Model 23 test	0.9200	36.0846	15.6532
Model 24	0.9170	36.9662	15.9971

trainset			
Model 24 validation	0.9185	37.2581	16.1139
Model 24 test	0.9200	36.0818	15.6527
	r^2	*RMSE*	*MAE*
Model 25 train	0.9170	36.9720	15.9984
Model 25 validation	0.9185	37.2664	16.1163
Model 25 test	0.9200	36.0837	15.6529
Model 26 train	0.9170	36.9734	15.9987
Model 26 validation	0.9185	37.2664	16.1159
Model 26 test	0.9200	36.0856	15.6532

Model 27 train	0.9170	36.9744	15.9988
Model 27 validation	0.9185	37.2754	16.1180
Model 27 test	0.9200	36.0862	15.6533
Model 28 trainset	**0.9170**	**36.9773**	**15.9993**
Model 28	**0.9185**	**37.2759**	**16.1181**

validation			
Model 28 test	**0.9200**	**36.0865**	**15.6534**
Model 29 Train	0.9170	36.9731	15.9986
Model 29 validation	0.9185	37.2759	16.1181
Model 29 test	0.91200	36.0865	15.6534
Model 30 trainset	0.9169	36.9551	15.9949
Model 30 validation	0.9185	37.2243	16.1075
Model 30 test	0.9200	36.0743	15.6509
Model 31 train	0.9200	36.9497	15.9940
Model 31 validation	0.9185	37.2215	16.1075
Model 31 test	0.9200	36.0731	15.6509
Model 32 train	0.9170	36.9610	15.9958
Model 32 validation	0.9185	37.2644	16.1151
Model 32 test	0.9200	36.0779	15.6520

Model 33 Train	0.9170	36.9793	15.9996
Model 33 validation	0.9175	37.2940	16.1199
Model 33 test	0.9200	36.1000	15.6558
Model 34 trainset	**0.9146**	**37.7603**	**16.0985**
Model 34 validation	**0.9160**	**37.8917**	**16.1816**
Model 34 test	**0.9171**	**36.6713**	**15.7035**
Model 35 train	0.9170	36.9708	15.9982
Model 35 validation	0.9185	37.2680	16.1165
Model 35 test	0.9200	36.0774	15.6520
Model 36 trainset	0.9167	36.9717	15.9948
Model 36 validation	0.9183	37.2801	16.1150
Model 36 test	0.9198	36.0894	15.6512MTOW

8.10 Appendix J

A graphical representation of the coefficients of model 1.