ROTTERDAM SCHOOL OF MANAGEMENT
ERASMUS UNIVERSITY ROTTERDAM

MSC BUSINESS INFORMATION MANAGEMENT

MASTER THESIS

Collaborative noise measurement

The contribution of human sensors to the analysis of noise pollution in Rotterdam

Author:
Hugo KRIER
441418hk

Supervisor:
Dr. Jan VAN DALEN
Co-reader:
Rodrigo BELO

20-06-2016

Acknowledgements

It would have been quite complicated to write this thesis without the help and support of the kind and helpful people around me, to only some of whom it is possible to give particular mention here.

I would like to thank my supervisor, Jan van Dalen. In fact, as I did not have to write a Bachelor thesis during my studies in Switzerland, I was quite unsure about the entire process, but also the depth of relationship a coach was supposed to have with his thesis students. When I saw that Jan van Dalen was mentoring the thesis of more than 20 students, I felt a little worried. However, none of my doubts was actually accurate. Jan van Dalen allowed me enough autonomy to come up with my own topic, but at the same time provided me with the right amount of guidance that would make the thesis academically relevant, but also would enable me to pursue a strong learning effect in terms of data analysis. Hence, I would like to express a severe gratitude to him for having coached me along the way.

Another special thanks goes to my co-reader, Rodrigo Belo. Rodrigo's focus incorporated specific direction and feedback in terms of the analysis, which was not natural given his position. It was great to experience that I was able to benefit from a mentoring synergy from both professors, making the learning curve tremendously steep, but also providing me with new research ideas as well as guidance for many steps along the path.

I would also like to thank the DCMR, especially Henk Wolfert, for giving me the opportunity to work on this project by providing me with the necessary comparison data, but also for taking the time to investigate the report and to provide feedback.

Another thanks goes to Mark Verschuur, a GIS expert working at the Rotterdam municipality, without whom the graphical analysis in QGIS would have been extremely complicated and who, despite his very busy agenda, took the time to give a hand with QGIS.

Last but not least, I have to acknowledge the fact that several fellow students provided me with a decent amount of both technical and social support along the process. I would like to thank Jonathan Peene for taking the time to delve into my thesis, for giving regular feedback, but also for further suggestions in terms of the analysis in R. Moreover, I would like to thank Martin Kayser and Claude Zwicker, who showed expertise and technical skills that helped me to overcome several obstacles related to the analysis in R. Those people would really represent what you call helpful, and hopefully more interesting projects and learning exchanges with them will follow in the coming years.

Executive summary

Noise pollution has become a considerable challenge within most metropolitan areas and can lead to adverse health effects such as sleep disturbance, heart diseases or hypertensions. Important sources of noise pollution include aircraft and industry noise, and mostly road traffic. Generally, the EU prescribes the implementation of noise maps and noise action plans to agglomerations whose population exceeds 100,000 people. For these noise maps, agglomerations model the existing noise levels based through calculations with yearly averages. The traditional way of measuring noise includes several disadvantages: for one, it is not able to cover local variations, for another it is a very costly undertaking and levels are updated rarely, which can cause inaccuracies.

As a result, the method of collaborative noise measurement, where people act as noise sensors via their phones, has gained significant momentum and for this thesis, has been assessed in the realm of design science. The goal of design science research resides within the investigation of emergent properties caused by a symbiosis between technology and human behavior. The key to fulfill this research type is the construction or evaluation of an artifact, both by satisfying seven research evaluation guidelines, but also by following a design science research methodology. In the case of NoiseTube, one can conclude that the experiment complied with the majority of design science requirements, but that the latter in isolation does not sufficiently prepare the floor for a fruitful research framework. Design science has to be combined with the concepts of natural science and technology acceptance, in other words, it has increase its focus on the behavioral aspects of technology. As a result, this paper is able to provide incremental insight into the concept of design science.

The experiment's objective was to collect noise data via an application called NoiseTube, and to then graphically and statistically compare those data to the noise levels modeled by the Environmental Protection Agency (DCMR) in order to draw conclusions and recommendations for the analysis of noise pollution, but also to assess whether such a mobile application would be suitable for a large-scale implementation.

The reach of the experiment ended up moderate and data did not cover most of the Rotterdam area, also because the process of measuring noise turned out unwieldy. Mostly covered suburbs were Rotterdam centre and Kralingen. NoiseTube data were aggregated per 100 meter grids and then compared to modeled 2011 data at building level as well as 2007 data at contour level. The method comparison revealed noise level similarity between modeled and NoiseTube data for the majority of Rotterdam Centrum, but also highlighted differences in noise levels for several locations in Kralingen.

It turned out more accurate to compare aggregated NoiseTube data against modeled data at contour level, despite the fact that the two methods diverge in their data generation. Moreover, one could recognize a certain user pattern, which is that most people would either measure while in transit, or switch on the application at home and measure noise there, represented by a large amount of data points lying on top of each other at several residential locations in Kralingen.

Further, NoiseTube did not allow the tracing of noise levels back to their sources. This is because the application's functionalities were too inconvenient to add source information, but also due to the fact that it did not include a automatic source recognition specificity. A solution here could

be to add the technology of sound recognition into the application's functionalities, similar to what Shazam does for music. Thereby, one would not have to manually add noise information for one, but also enable a detailed insight into what really caused the noise level. Moreover, the application did not allow the inference about adverse health effects.

In order to gain further insight into user measurement behavior, a predictive model was created with the decision tree technique. To begin with, a logistic regression model was put in place to check for the significance of the predictors, followed by a binary the target variable as to whether a person that had already used NoiseTube would continue his or her endeavors or not. Results of the model highlighted that a majority of people were only predicted to continue measuring in case they were not annoyed by noise levels, in case they had measured at various locations and if there was a considerable time frame between their first and last measurement. Needless to say, the model advocates the conclusion that collaborative noise measurement highly depends on noise sensitivity, but also that it needs to include sufficient users to account for high-noise areas.

One major impediment to a large-scale implementation is the appropriate user base size. One absolutely needs a very large pool of volunteers to acquire enough data to sufficiently cover a city like Rotterdam, which was not the case in this experiment, primarily because people did not perceive the measurement as useful for them, but also due to the fact that the procedure of measuring was judged rather complicated. In order to have sufficiently detailed results, one would require an estimated 100 people per square kilometer to constantly measure noise through their mobile phones.

Therefore, in order to obtain sufficient data collectors, one needs to ameliorate the applications ease of use as well as give people a incentive, either financial or non-financial, to participate. Solutions here could be to either oblige municipality employees to measure via contract or to subsidize certain companies, such as flyer distributing agencies, so that their employees would include collaborative noise measurement into their daily work.

This study contributed to existing collaborative sensing and noise pollution literature. Noise pollution impacts any citizen and should be analyzed as precisely as possible. Public and people-centric noise measurement has the potential to constitute a viable complement to traditional noise measurement because it is able to provide urban planners with detailed and data-driven noise input, but also with an opportunity to cost-effectively leverage existing technology at the nexus of crowdsourced Web 2.0 tools. However, the implementation has to entail a sufficiently large user base, which can be achieved by obliging employees to use it, but also by partnering with companies whose job roles include significant people movement. Further, the utilized application used should permit source recognition features and the measurement process should be made as straightforward as possible. With a given volunteering effort, every resident has the opportunity to actively contribute to the reduction of noise levels and to thereby increase the quality of living in the medium-term.

Keywords: Noise pollution, NoiseTube, collaborative noise measurement, crowdsourcing, design science, method comparison, user acceptance, noise level reduction.

Contents

List of Tables **viii**

List of Figures **viii**

1 Introduction **1**

1.1 Research Background: Noise pollution 1

1.2 Participatory noise measurement 3

2 Research objectives **4**

3 Theoretical Background **6**

3.1 Noise Pollution: General 6

3.2 Sources of noise pollution 7

3.3 Effects of noise pollution 9

3.4 Traditional noise measurement 10

3.4.1 Context and definitions 10

3.4.2 Limitations of traditional measurement 11

3.5 Participatory noise measurement: The NoiseTube approach 12

3.5.1 Context 12

3.5.2 Benefits of participatory noise sensing 13

3.5.3 Challenges of participatory noise sensing 13

3.6 Comparison between modeled and NoiseTube data 13

3.7 Design science 16

3.7.1 General 16

3.7.2 Research evaluation criteria 16

3.8 Crowdsourcing 18
3.8.1 Context 18
3.8.2 The NoiseTube crowdsourcing component 19
3.8.3 Guidelines for practical crowdsourcing 20

4 Data and Methods **22**
4.1 Data source and population sampling 23
4.2 Data collection and equipment used 24
4.3 Data analysis and method comparison 27
4.3.1 Graphical inquiry 27
4.3.2 Statistical testing 31
4.4 Predictive model and user measurement behavior analysis 32
4.4.1 Model creation 32
4.4.2 Model evaluation 34
4.5 Design science: Research methodology 35

5 NoiseTube data description **36**
5.1 Context 36
5.2 Coverage 39
5.3 Outlier analysis and exclusion 40
5.4 Graphical visualization 42

6 Results **44**
6.1 Method comparison 44
6.1.1 Analysis with 2007 contour data 44
6.1.2 Analysis with 2011 building data 45

6.2 Measurement appraisal: User model 47
6.2.1 Model creation 48
6.2.2 Model interpretation 50
6.2.3 Model evaluation 51

7 Discussion **55**
7.1 Method comparison 55
7.2 User measurement model 57
7.3 Contribution to literature 59
7.3.1 Impact on sources and effects of noise pollution 59
7.3.2 Design Science 60
7.3.3 Technology acceptance 62

8 Conclusion **65**
8.1 Main findings 65
8.2 Managerial implications 66
8.3 Limitations and future research 67

Bibliography **69**

Appendices **73**

Appendix A Summary EU Directive 2002/49/EC **73**

Appendix B Leq Indicator **73**

Appendix C NoiseTube Analysis R script **73**

Appendix D Grid Construction R script **76**

Appendix E NoiseTube Map Construction R script **79**

Appendix F Predictive Model Creation R script **81**

Appendix G Predictive Model Evaluation R script **87**

Appendix H Summary Logistical Regression Model **91**

Appendix I Model Confusion Matrix Classification **92**

Appendix J Model Confusion Matrix Probabilities **93**

Appendix K AsRules Function R **93**

List of Tables

1 Benefits of participatory noise sensing 14

2 Types of volunteer groups 21

3 Different steps in the comparison between NoiseTube and DCMR data 30

4 Different steps in the creation of a user model in the NoiseTube context 33

5 Predictor variables and intuitions 49

List of Figures

1 Example of a noise map of Rotterdam 2

2 Thesis process framework 6

3 Comparison between modeled noise data and NoiseTube data 15

4 Design science research evaluation guidelines 17

5 Relationship among design framework features 18

6 Extract: individual NoiseTube Elog board 20

7 Example of a customized noise map 25

8 Example of NoiseTube data in .json format 26

9 2007 DCMR noise data at the contour level 28

10 2011 DCMR noise data at the building level 29

11 An example of a NoiseTube dataset 38

12 Rotterdam NoiseTube data coverage 39

13 Noise level distribution histogram 41

14 Noise level distribution boxplot 41

15 Illustration of a NoiseTube data map of Rotterdam 43

16 Rotterdam noise map with NoiseTube and contour data 44

17 Rotterdam noise map with NoiseTube and building data 46

18 Decision Tree for the user measurement behavior model 50

19 User model class distribution 52

20 Decision tree and logistical regression: ROC curves 53

21 The technology adoption lifecycle 63

22 The technology acceptance model 63

1 Introduction

1.1 Research Background: Noise pollution

The world has been experiencing extraordinary levels of urbanization over the last decade, with half of today's global population residing in cities, and municipalities face challenges like traffic congestion, environmental pollution or resource scarcity (Nam and Pardo, 2011). Cities that flourish in these settings are those that are, on one hand, progressively monitored by ever-pervasive computing and, on the other, are evolving as business hotspots through innovators and entrepreneurs, so-called smart cities (Kitchin, 2014).

Noise pollution is a somewhat underestimated polemic in the context of smart cities (hereafter SC) and quality of living. Since the era of industrial revolution, our daily lives, especially of those living in urban settings, have been raided by unsolicited and troublesome sounds (Hildebrand, 1970). In the context of noise at work, any noise above 85 decibels can be considered as harmful to human wellbeing and this level is constantly reached when exposing oneself to road traffic (Wolfert, 2015). Further, evidence has shown that any noise level above 42 dB is causing annoyance and that a level that tops 55 dB can lead to adverse health effects (Wolfert, 2015).

According to Maisonneuve, Stevens, Niessen, Hanappe and Steels (2009), elevated noise levels are considered a challenge in urban environments, affecting health, productivity and human behavior. Major sources of noise include traffic, aircraft noise, construction work, industry noise, as well as local events like festivals or sports games. Among noise-related health effects figure annoyance, sleep disturbances, but also hearing impairment or cardiovascular diseases, making noise a significant influence on quality of living (Theakston, 2011).

Noise exposure is becoming a considerable hazard to quality of living in the Netherlands, with 600 yearly deaths due to noise-related diseases (Stuijt, 2009). Additionally, one third of the population frequently suffers from sleep deprivation caused by noise and - due to population growth and urbanization - adverse health effects become more worrying every year (Stuijt, 2009).

Complaints about noise contact figure among the most frequent inhabitant complaints and it can be deduced that complaint frequency is rising with city size (Muzet, 2007). With the most regularly cited source being traffic (Muzet, 2007), it seems clear that areas that exhibit intense transit or congestion will always be at the origin of rather elevated noise levels.

Noise pollution has been deemed a major threat to human well being by the World Health Organization (WHO), leading to a reaction of the European Commission (EC), which stipulated a directive - the Environmental Noise Directive 2002/49/EC (END, Appendix A) - obliging major cities to establish a noise management policy by gathering realistic data on noise exposure and city decibel levels in order to then create local action plans (Maisonneuve et al., 2010).

The END is applicable to noise experienced in built-up areas, public parks and other sectors in agglomerations and, in an ideal scenario, aims to prescribe a common approach to prevent and reduce harmful effects caused by environmental noise exposure (Cox and Palou, 2002).

In this context, the European Union (EU) has required since 2012 that cities consisting of more than $100,000$ inhabitants draw noise distribution maps (Singh and Davar, 2004). A noise map

(Figure 1) is a graphic illustration of the sound level dispersal existing in a given area for a prescribed time frame (Maisonneuve et al., 2010). The idea is thereby to identify the areas that are mostly affected by intolerable noise and the percentage of inhabitants that suffers from excessive noise levels (Martin et al., 2011).

Figure 1: Example of a noise map of Rotterdam

Generally, those noise maps (Figure 1) used by both EU and national governments for action planning are primarily traced back to yearly averaged formulas such as Leq, calculated or modeled through standardized noise mapping methods (De Coensel and Botteldooren, 2014). Most cities have taken the approach to draw noise maps through propagation models that translate local measurements to wider boroughs (Maisonneuve, Stevens, Niessen, Hanappe and Steels, 2009). Occasional measurements are then made to validate the calculated noise levels, thereby increasing detail and accuracy of the noise maps (Wolfert, 2015).

However, the creation of noise maps and the collection of input data through traditional measurement is very costly and time-consuming, resulting in the fact that these maps are only updated after an extensive time period, such as five years in the UK for instance (Rana et al., 2015).

As a result, conventional measurement techniques can construct noise maps that can be used for major traffic infrastructure, but not to cover local dispersions, let alone unplanned or short-term noise pollution peaks such as construction. They are designed for strategic long-term planning, but not for a real-time and appropriately focused noise approach (De Coensel and Botteldooren, 2014) and thus constitute a shortcoming for ubiquitous noise measurement.

A recent example of insufficient information by average-based noise level simulations considers the complaints of inhabitants from "Landsinger Land" area (DutchNews, 2011): In spite of multiple resident complaints about excessive train noise, the minister of infrastructure would not put the sound reduction into his agenda, claiming noise levels would not trespass national averages. However, after having taken cross-check noise measurements, it turned out that noise levels were indeed above tolerated thresholds, obliging the minister to engage in noise abatement measures, together with the city council (DutchNews, 2011). Hence, this proves the need to seek for a more wide-ranging noise measurement portfolio, rather than through calculations made with yearly average-based simulations methods.

In other words, city authorities cannot not entirely identify the accurate degree of resident exposure to environmental noise, which somewhat undermines the intentions of being concordant to

the END. With the measurement deficiency, authorities are considering other, collaborative and crowdsourcing noise measurement approaches. The latter would enable them to take accurate and prompt action to address every noticeable local noise origin, such as vehicle or household noise(Maisonneuve et al., 2010). Resultantly, besides being fully compliant with the END, agglomerations would be able to matchlessly address noise pollution, descrying the opportunity to engender noise-related improvements in residential quality of living.

1.2 Participatory noise measurement

As simulation-based or modeled noise data do not always portray flawless noise exposure (Maisonneuve, Stevens, Niessen, Hanappe and Steels, 2009), other approaches that enable data-driven noise monitoring have to be identified. The establishment of a noise action plan is complex and cannot exclusively be implemented by city authorities, but requires the involvement of the general public in the form of responsibility sharing (Maisonneuve et al., 2010).

First steps included the use of wireless sensor networks in several areas, reflected for instance via the Katendrecht sound assessment study, which can be seen as a first form of participatory noise measurement (De Coensel and Botteldooren, 2014). A possible cutting-edge alternative for this collaborative approach is an application called NoiseTube. It allows its volunteers to gather data through their smartphone microphones that can then be used to for noise assessment analyses and to establish precise noise maps. Hitherto, measuring experiments with NoiseTube have been carried out in cities like Antwerp or Brussels (De Coensel and Botteldooren, 2014).

The NoiseTube approach, if endorsed by inhabitants, will permit city authorities to address questions that include high spatial variability as well as significant degrees of unpredictability (De Coensel and Botteldooren, 2014), issues that traditional noise measurement cannot tackle. Even more, NoiseTube will permit to construct different versions of noise maps as a function of location or time, thereby enabling a potential noise pattern recognition that cannot be derived through conventional measurement methods.

Following the NoiseTube approach, participating inhabitants would then proceed with day-to-day noise level measurement, which would give urban planners the opportunity to for instance reduce noise levels generated by traffic via periodic speed limits if an ample number of citizens in a high-traffic area would record excessive decibel levels.

Consequently, the ascription back to respective noise sources and the investigation of their influence on recorded data is also vital.To address this, the application uses a tagging feature that allows people to include comments into their measurement, making it straightforward to indicate the noise' source or other comments (Maisonneuve, Stevens, Niessen, Hanappe and Steels, 2009). Not only would the reduction of noise levels and annoyance increase Rotterdam's living standards, but also boost attractiveness of the city as a business and tourist harbor.

Notwithstanding, the NoiseTube approach, in order to be validated, has to fulfill academic requirements, in this case from Information Systems (IS) research. Moreover, the question as to whether individuals endorse participatory monitoring is also critical (Mousa et al., 2015).

2 Research objectives

This paper relates to the literature strand of IS and focuses on the impact of collaborative noise measurement on the analysis of noise pollution, an explorative study that can also be deemed academically relevant to Design Science (hereafter referred to as DS) and Information Technology (IT) research. The concept of DS, where the aim is to create and examine IT artifacts designed to solve precise organizational problems (von Alan et al., 2004) is academically relevant and applicable, because it establishes a valid foundation, methodology and evaluation procedure for this research, also in a SC context. In this context, the paper's underlying objective is to assess the impact of the NoiseTube approach within DS.

Rather than taking advantage of the hypothetico-deductive research method, this paper is considered an exploratory research paper, where the goal is to obtain intuition about a certain problem at hand in order to better understand the essence of a problem, mainly because little studies have been undertaken in the domain (Sekaran and Bougie, 2003). Sekaran and Bougie (2003, p.97) advance that "exploratory studies are also necessary when some facts are known, but more information is needed for developing a viable theoretical framework", which can be deemed accurate for the investigation about NoiseTube.

In this context, Drosatos et al. (2014) stress that further research studies in the field of noise pollution should be the application of collaborative noise measurement in real settings, giving this thesis academic and practitioner validation and perceiving opportunities within DS. This work, by comparing simulation-based to crowd-sourced noise sensing, adds incremental value to noise pollution literature.

The method of participatory and ubiquitous noise level measurement through human sensors and the subsequent comparison to noise data from conventional noise measurement constitute a novel approach in the domain of noise monitoring research. This technique has not been abundantly examined in real-life scenarios, except for a couple of short-term monitoring experiments (Drosatos et al., 2014). The intention thereby is to deepen intuition about noise dynamics, potentially illustrating conflicting findings about the degree of noise pollution. Crowdsourcing and other affordable sensing applications on mobile devices could portray noise levels differently than average calculations and should therefore be studied in a real-life setting. Consequently, the research question or problem statement can be formulated as follows:

How does the use of crowdsourced noise measurement, compared to traditional noise appraisal, affect the analysis of noise pollution in Rotterdam?

This paper's differentiating factor and objective reside within the experiment as to whether a collaborative noise measurement method can be granted technical validity, but also whether it can be accorded potential for a large-scale launch as auxiliary noise monitoring equipment.

In this context, the envisaged creation of a user model would enable the prediction of volunteer measurement activity, critical for a mass launch of the technology. One has to investigate how and whether people are going to use their phones as sensors in order to take a decision about a potential large-scale launch. Moreover, it could possibly facilitate the identification of specific measurement patterns, may it be as a function of user, location or noise level. This will permit a much more accurate insight into the dynamics of noise measurement and thereby create opportunities for detailed and data-driven urban planning.

Consequently, both the user model as well as the results from the method comparison are estimated to portray whether the implementation of a collaborative noise measurement technology can be considered conceivable. If successful, the experiment would permit to better assess the impact of noise on human annoyance. Moreover, the NoiseTube approach could be an opportunity to recognize behavioral, time-related and locational noise patterns. Subsequently, these patterns could be compared to existing municipality reduction and prevention programs and then used to align existing noise action plans, in a more hands-on and holistic manner than traditional methods do.

The NoiseTube approach can be of substantial managerial relevance for city authorities. The task of comparing modeled noise levels to NoiseTube data would allow for additional information about genuine noise levels, which is valuable to the Environmental Protection Agency (hereafter DCMR). Human noise sensing could then constitute the basis of a micro noise action plan, referring to more local noise exposure maps (De Coensel and Botteldooren, 2014).

Further, noise affects every inhabitant and needs to be evaluated as precisely as possible. The NoiseTube approach would hence enable a municipality to have more definite insight into recorded noise levels, which can then be used to take actions like night speed limits or mobile and temporary noise barriers, especially in the case of accidental or short-term noise exposure. Resulting action plans could even include the installment of temporary noise monitoring units in various suburbs as a means of further noise assessment (Wolfert, 2015).

If the NoiseTube approach would be successful, Rotterdam's municipality will then have a supplementary opportunity to reduce noise levels in the mid-term. Even more, the city will be able to collaborate with and empower its residents to actively contribute to achieving its ambitious SC initiatives, resulting in a co-evolving responsibility environment that would eventually lead to increased living standards. Consequently, this paper is able to seize the assessment of both academic and managerial relevance.

This paper will be organized as follows: Section 3 will present a theoretical background including noise pollution, both noise measurement methods, its prevailing sources and materialized effects of noise pollution. Moreover, it will lay the groundwork in terms of DS and crowdsourcing. Section 4 will cover the necessary methodological steps applied in the NoiseTube experiment, elaborating on the method comparison and on the predictive model creation.

Subsequently, Section 5 will emphasize the NoiseTube data description, before Section 6 will proceed through the analysis involving the comparison of modeled to NoiseTube data. Section 6 will also deal with the creation, interpretation and evaluation of a user model, assessing how likely people are to act as noise measurement volunteers.

Results and implications will the be debated in Section 7 before the paper will be concluded in Section 8, answering the research question and thus giving an overview as to whether the NoiseTube has the potential to be used as auxiliary noise measurement tool. The entire thesis process framework is illustrated in Figure 2 below.

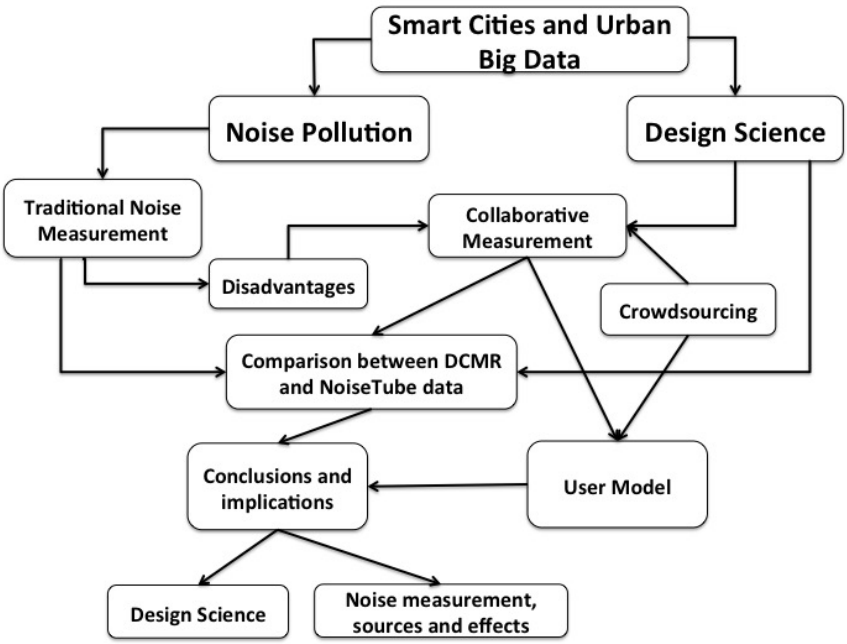

Figure 2: Thesis process framework

3 Theoretical Background

3.1 Noise Pollution: General

According to Sekaran and Bougie (2003), a good theoretical background has three functions: announce the study subject (1), highlight the problem (2) and compile previously completed work (3). One can initiate the theoretical background by differentiating between sound and noise. The former is created by any form of mechanical movement and diffused as a wave through air or other materials (Muzet, 2007). Further, sound is described through its mechanical energy and represented through energy-related units (Muzet, 2007).

Noise, however, reflects an unwanted form of sound, making it logical that sounds cannot be classified as noise, given the unique form of their physical features (Muzet, 2007). Generally speaking, one can consider noise as noticeable sound occurrence that may have opposing effects on people, and than can be perceived both physiologically and psychologically (Muzet, 2007). There is three basic factors embedded in every noise problem (Ouis, 2001): a sound origin, a transmission path, and a recipient. For the purpose of this paper, no difference will be made between sound and noise.

Muzet (2007) stresses that noise pollution is attained with the appearance of harmful or annoying levels of noise. Nowadays, Theakston (2011) considers noise pollution to be the third most harmful environmental pollution type and elevated noise levels are deemed a primordial source of annoyance in cities since the 1970s (Zannin et al., 2006). It is a topic of actual concern for municipalities, represented by multiple ongoing monitoring assessments.

Several theoretical studies and literature investigations have been undertaken in the field of noise pollution. Goines and Hagler (2007), but also Singh and Davar (2004), define noise, measured in decibels (dB), as an unwanted sound and consider environmental noise pollution as a form of air pollution. Further, it is stressed that noise pollution is more severe and widespread than ever before and that it will continue to soar due to population growth, urbanization and increases in car ownership, hence in mileage. According to Singh and Davar (2004), noise pollution originates from human activities and is becoming increasingly omnipresent.

According to Hildebrand (1970), three approaches have been identified to tackle excessive noise levels: Firstly, one can attempt to decrease noise at its source; secondly, as a second solution, one can try to shield sensitive dwellings or work locations like schools or hospitals; a third solution would be to cover or mask unwelcome noise with other acceptable, enjoyable sounds. Further, education, communication and sustained noise awareness could also contribute in reducing sound levels (Kam et al., 1994). Similarly, Ouis (2001) proposes that the best remedy to exposure to detrimental noise is to address the latter's emission at its source.

3.2 Sources of noise pollution

To be able to address noise pollution at its source (Hildebrand, 1970), one needs to be aware of its origins. As a further argument to validate the discussion about noise sources, one can stress guideline 7 of the DS research criteria, mentioned below in Section 3.7.2. The latter states that effective DS research must provide clear foundational contributions in terms of design foundations (von Alan et al., 2004). In this case, the various sources are part of the DS foundations and therefore need to be elaborated.

Sources of noise pollution are countless in numbers and variety, but four broad categories are worth underlining (Hildebrand, 1970): household appliances (1), industry and construction (2), traffic (3), and aircraft noise (4). Further, different noise sources may possess varying acoustical features, with some simply emitting a pure tone and others radiating a random sound with a recognized frequency range (Ouis, 2001). This paper will focus on (2), (3) and (4).

Salomons (2013) identifies road traffic noise as the major source of environmental sound and stresses that other types of noise include rail traffic noise, aircraft noise and industrial noise. In their study, they mainly focus on road traffic noise and stress that the latter shows considerable variations in cities. Noise levels are high near busy roads and low in shielded areas, which reflects the fact that traffic noise is dependent on traffic volumes. The latter in turn is closely related to infrastructure, particularly buildings (offices versus dwellings) and road networks.

According to noise maps in Rotterdam, the most important category of noise pollution in cities is road traffic (3) (Wolfert, 2015), even though passenger traffic per se is not automatically excessively loud. Car producers are manufacturing new car versions equipped with silencers, noiseless engines, and specifically designed quiet tire treads (Hildebrand, 1970).

Nevertheless, problems arise through the fact that cars are usually tested in ideal conditions, but in practice drive on surfaces that are less smooth than test locations, but also turn out to have other exhausts and wider tires (Wolfert, 2015). Additionally, sports cars or trucks with deficient mufflers produce heavy unexpected noise, with for instance the average truck driving at 95 km/h being twice as loud as a conventional car stream (Hildebrand, 1970).

According to Wolfert (2015), another dominant noise factor for automobiles is the contact of the rolling tire on the road. In fact, for cars, the engine is the major noise origin at speed levels around 35-40 km/h, but above this pace, it is the tire-road friction that appears to be the dominating noise aspect. The same reasoning holds for trucks at a speed level around 70km/h. As a result, one can conclude that in city centers and neighboring suburbs, it is the engine that predominates in the category of road traffic noise. The opposite however holds for highways and circumvention roads in and around the city, as higher speed levels induce noise coming from tire-road friction.

According to Ouis (2001), the driving method, including sudden accelerations and decelerations, can also cause increased traffic noise and thus high levels of public annoyance (Ouis, 2001). To remedy this, efficient urban planning could postulate speed limits and thereby engender smooth traffic flow, minimizing human annoyance due to traffic

In the category of industry and construction (2), it seems likely that workers in noisy surroundings would suffer from hearing impairment at an earlier age than other citizens. Scientists consider long-term exposure to noise levels above 80 decibels as a generally accepted reason for hearing loss. Even worse, temporary deafness can happen because of short-term exposure to noise rates between 100 and 125 decibels, with the ear being permanently damaged at 150 decibels, even with an extremely short-time contact (Hildebrand, 1970). Noises coming from industrial plants are a frequent source of significant noise (Muzet, 2007). Groundwork noise such as hammering, cranes or heavy trucks can be more or less regular or unexpected, with huge diversions in intensity, and can be transmitted across large spaces, making it a source that is generating high noise emissions (Muzet, 2007).

Aircraft traffic (4) research has been at nexus of noise research for three decades, with individual airplane noise having considerably decreased during this period, mostly due to changing engines (Muzet, 2007). However, due to an increase in the total amount of flights, aircraft and airport noise are considered more annoying than traffic or train noise (Shepherd et al., 2010).

Franssen et al. (2004) highlights that the most noticeable argument about airport noise is that its high frequency effects can happen during night-time. Surges in night traffic have often caused disagreements between airport authorities and inhabitants of surrounding suburbs, as people are worried about noise-related health effects but also about safety and other forms of pollution. Surges in aircraft noise levels affect a population's social structure, for instance by causing housing prices to decrease in locations next to large international airports, but also increase the use of medication taken for sleep and cardiovascular diseases (Franssen et al., 2004).

Most environments contain a combination of noise disturbances from different sources, with trains, airplanes and motor vehicles being a very frequent combination in metropolitan areas. As of today, there is not consensus on a model for measuring total annoyance from multiple noise sources (Goines and Hagler, 2007).

3.3 Effects of noise pollution

Theakston (2011) has stressed seven categories of negative health implications of noise pollution on humans. Goines and Hagler (2007) therefore consider as adverse health effects noise included hearing loss (NIHL), interferences with spoken communication, sleep disturbances, cardiovascular disorders, disturbances in mental health, impaired task performance and negative social behavior such as annoyance reactions. Ultimately, Theakston (2011) expresses the burden caused by environmental noise through disability-adjusted life years (DALYs).

As such, DALYs are the aggregated amount of potential life years lost due to untimely decease and the equivalent years of healthful life lost due to a state of poor health conditions. According to Theakston (2011), the DALY-related loss coming from noise pollution for all noise-related health effects ranges from 1 to 1.6 million DALYs in Western Europe (WE). Consequently, at least one million years are lost from traffic-related noise pollution in Western Europe every year.

Hildebrand (1970) stresses that elevated levels of noise pollution can lead to human annoyance and sleep disturbances, effects that negatively affect personal health and thus also quality of life. Adverse effects of excessive noise levels are interconnected and multivariate, although it may be complicated to show adverse effects of noise on communal and individual mental well-being (Hildebrand, 1970). The ability of humans to adapt to environmental corrosion further thwarts objective measurement of noise-related consequences; several humans eventually become neutral to some extent of several noises (Hildebrand, 1970), whereas others remain noise-sensitive at a very large scale (Wolfert, 2015).

Goines and Hagler (2007) state noise is not only considered a simple cause of annoyance, but can trigger strong adverse health and social effects. Even at intensities that do not directly damage hearing, noise can be subconsciously considered as an alarm signal, also in the case of an exposure during sleep. Another argument advanced by De Coensel and Botteldooren (2014) is that In this context, it seems clear that harassment caused by intense noise exposure can lead to more than annoyance and that municipalities or other authorities should attempt to minimize wide-ranging noise levels.

In this context, there are growing worries about the effect of noise on cardiovascular diseases, including hypertension and heart diseases like angina or myocardial infections (Theakston, 2011). In this context, there is evidence that traffic noise is positively correlated with heart diseases, including myocardial infarctions but also that there is a positive impact of both airport and traffic noise on the risk of high blood pressure (Theakston, 2011). Further, any exposure to noise levels above 65 dB increases the risk of diabetes and strokes (Sørensen et al., 2011). It is estimated that $61,000$ DALY years are lost for heart diseases caused by noise.

De Coensel and Botteldooren (2014) conducted a smart sound monitoring study in the area of Katendrecht and stresses that sleep turbulences figure among the most immediate adverse health effects, including remembered awakenings and problems to fall asleep. Further, the most disturbing location for noise pollution is someone's domicile, as noise enters uninvited, and remarking the sound is mostly sufficient to experience a certain annoyance. For annoyance-related health effects, the loss in DALYs amounts to $587,000$ in WE (Theakston, 2011).

Fyhri and Klæboe (2009) found out in a study that 24 million people in the European Union

are extremely irritated by road traffic noise, despite the fact that noise sensitivity can cause tremendous variations in annoyance, making it a mediator between noise exposure and annoyance. Causal relationships between annoyance and health are difficult to prove and annoyance is thereby a very subjective measure that can be found through specific, individual questionnaires. Human indicators of annoyance include headaches, pain in the stomach, dizziness or problems falling asleep (Fyhri and Klæboe, 2009).

Nevertheless, the model of Fyhri and Klæboe (2009) advocates that there are noteworthy, but small relations between increased noise levels and several annoyance features like sleeping problems, nervousness, tiredness, sore throat and headaches, underlining that the correlation between annoyance and noise can be measured to some extent and reveal valuable for generalizations in other settings.

Not only does noise pollution create adverse health effects, it can also engender economic effects. For one, one has to take into account that high-noise areas exhibit considerably lower real estate prices than do low-noise suburbs (Wolfert, 2015). Other economic effects can also be considered when it comes to aircraft traffic. In the case of Rotterdam The Hague Airport (RTHA), one has to reflect on whether externalities like environmental noise in surrounding outskirts would be deemed superior to economic, social and political value. However, the economic effects of noise pollution are outside the scope of this paper and will therefore not be taken into account for further analysis.

After having emphasized the most critical sources and effects of noise pollution, both the conventional and NoiseTube method will be presented in the following sections, with ultimately a comparison to reinforce and advocate the use of NoiseTube as an auxiliary noise sensing tool.

3.4 Traditional noise measurement

3.4.1 Context and definitions

Noise levels can be established by taking advantage of various rating methods, provided they adhere to international standards and the directives initiated by the European Commission (Santini et al., 2008), such as the END (summary in Appendix A). Given the distinction between noise and sound, this section highlights some of the important definitions in the field of regular noise pollution measurement.

For accurate sound measurement, Zannin et al. (2006) propose to take advantage of a weighting curve (dB (A)) in order to enhance frequencies to which human hearing is more reactive. As a result, noise levels represented by the weighting curve and rated in decibels (dB) can be used as adequate approximations of the human ear.

The basis and norm of recent noise research can be considered to be the END, which stipulates that as of 2012, "agglomerations with a population of over $100,000$ have to estimate the number of citizens exposed to average yearly sound levels of 55-75 dB, in bands of 5 dB, and over 75 dB, and this at 4 m above the ground on the most exposed facade of their home" (D'Hondt et al., 2013, p.3). Different rates are required for road, rail and air traffic, as well as for industrial sources, where only considerably congested roads, railroads and airports are taken into account (D'Hondt et al., 2013). Results must be visualized graphically and updated every 5 years.

In Europe, metropolises that are required to establish noise distribution maps base the calculation of traffic noise levels on detailed traffic and building data (Salomons and Pont, 2012). Noise distribution maps are being calculated through the "equivalent continuous sound pressure level LA,eq" (D'Hondt et al., 2013, p.3) and have to be made available to the public (Santini et al., 2008).

In fact noise levels are constantly increasing or decreasing over time, making a precise evaluation complicated (Zannin et al., 2006). This creates the need to facilitate things by defining sound level "as a continuous status that would produce the same effect on the human ear if compared to the actual noise observed, including all the variations" (Zannin et al., 2006, p.4). In this context, measured in decibel, Leq refers to the sound level of a constant source over a time period T that "has the same acoustic energy as the actual varying sound level pressure over the same interval" (Maisonneuve, Stevens, Niessen and Steels, 2009, p.3).

Leq is thus able to replace various sound variations with a single rate of noise level. What is more, the EU obliges municipalities to furnish noise data with regards to both the indicators LNIGHT and LDEN, representing average noise levels over the night only, for one, and over the entire day, for another (Santini et al., 2008).

3.4.2 Limitations of traditional measurement

Despite the fact that traditional noise measurement takes advantage of modeled data and uses measured data to cross-validate, one can consider the major impediment for accurate and wide-ranging noise measurement to be the cost of sensors or sound level meters (hereafter SLM). This includes the construction of modeled noise maps, but also sensor acquisition, installment and maintenance (Wolfert, 2015).

The DCMR is not able to install citywide monitoring sensors with its given budget and therefore welcomes auxiliary solutions to accurately assess noise levels in Rotterdam, but also to obtain further insight into the dynamics of noise. Similarly, the agency takes action based on EC-prescribed and long-term noise action plans and would welcome a procedure of data-driven micro action planning, often driven by the imagination of municipality governance (De Coensel and Botteldooren, 2014).

Maisonneuve et al. (2010) also stress the fact that conventional sound measurements include multiple disadvantages. In fact, next to propagation models, the measurement share is mainly carried out by officers installing SLM in several urban areas during a limited amount of time. This is usually done at a couple of locations of interest, such as close to roads or railways. As a result, this kind of mapping prerequisites considerable expertise, human resources and the use of expensive sound meters, often requiring investments not available in city budgets (Maisonneuve, Stevens, Niessen, Hanappe and Steels, 2009).

As a second drawback of traditional sound measurement, one can emphasize that static sensors are limited to outdoor noise monitoring, excluding the rather significant part of time that people spend inside buildings (Maisonneuve, Stevens, Niessen, Hanappe and Steels, 2009).

Furthermore, authorities use propagation models to create noise maps, using extrapolation of local monitoring to other areas. However, those models can be subject to data granularity, meaning that the data gathered at rare locations do not fulfill the requirements of high granularity

in both time and space, therefore producing an unspecified error margin that may result in incorrect deductions (Maisonneuve, Stevens, Niessen and Steels, 2009).

In this context, one can conclude that there is an emerging need for alternative or complementary sound measurement techniques. In this vein, the DCMR considers the approach of collaborative noise measurement to be a very promising solution, having shown itself cooperative for a deployment of participatory noise measurement in Rotterdam.

3.5 Participatory noise measurement: The NoiseTube approach

3.5.1 Context

In a world where billions of people travel with a variety of mobile sensing, computing and networking devices, one can witness the arrival of participatory noise sensing (Mousa et al., 2015). Further, Campbell et al. (2008) refer to people-centric sensing and attach it to the notion of public sensing, where data is shared for the greater public wellbeing. This paper will focus on public, people-centric participatory sensing, which has taken inspiration from and triumphed over wireless sensor networks (WSN) (Maisonneuve, Stevens, Niessen and Steels, 2009).

WSN have been considered a first version of urban sensing and geographical monitoring (Maisonneuve, Stevens, Niessen and Steels, 2009). Rather than taking advantage of limited and costly sensors, a WSN attempts to capture the benefits from high numbers of cheap and simplistic sensing devices that can be inserted into an ecosystem and enable real-time noise observing (Maisonneuve, Stevens, Niessen and Steels, 2009). As a result, WSNs have the capability to induce ameliorations for environmental noise measurement, since its subtle perceptions permit the elaboration of improved management plans in affected suburbs (Santini et al., 2008).

Nevertheless, it remains uncertain whether WSN will turn out successful in case of a large-scale deployment, because the sensors are static and outdoors, but also is there no citizen involvement in this methodology (Maisonneuve, Stevens, Niessen, Hanappe and Steels, 2009), inducing some of the same challenges mentioned with traditional noise measurement in Section 3.4.2.

In public participatory sensing however, volunteers capture noise data from adjacent locations using sensors integrated in their mobile devices (i.e. smartphones or tablets) in order to subsequently share the gathered data with a backend server, before the latter then processes the information to map, analyze or represent incidents of common interest (Mousa et al., 2015).

In their research paper, Maisonneuve, Stevens, Niessen and Steels (2009) have highlighted public participation as a means to address noise pollution and projects like NoiseTube, investigating how people-centric data collection can be used to create a low-cost, open platform to measure, annotate and localize noise pollution as it is perceived by the citizens themselves. In the sector of urban planning, there is a sustained evolution towards "participatory mapping" and new methods are being examined under the nomination of geographical information systems (GIS) (Maisonneuve, Stevens, Niessen, Hanappe and Steels, 2009).

Specifically, the NoiseTube approach stipulates to use mobile phones as noise sensing equipment and to actively empower citizens, rather to take advantage of a few static sensors (Maisonneuve, Stevens, Niessen, Hanappe and Steels, 2009). In the context of the growing influence of Web 2.0 and network effects, NoiseTube has the opportunity to lower barriers for environmen-

tal noise measurement by increasing the range of citizen participation (Maisonneuve, Stevens, Niessen, Hanappe and Steels, 2009). Further, most of the participatory sensing methods including NoiseTube take advantage of the client-server architecture. Three main parties can be highlighted in this system (Mousa et al., 2015): participants, a campaign administrator and an end user.

3.5.2 Benefits of participatory noise sensing

Participatory noise sensing originates from the sensing-as-a-service model (SaaS). Perera et al. (2014) describes SaaS as solution formed on Internet of Things (IoT) infrastructure and having the potential to address issues in smart cities. Several important benefits can be translated from SaaS to a participatory noise sensing approach, shown in Table 1 below.

3.5.3 Challenges of participatory noise sensing

According to Mousa et al. (2015), there is primarily an uncertainty about applied participant behaviors. Participatory systems are susceptible to both missing and malevolent volunteers, as well as flawed contributions. Among the most frequent assaults on participatory sensing systems figure (1) corruption attacks, stemming from dysfunctional mobile sensors, and (2) on-off attacks, where participants alternate between normal and adverse measurement behaviors (Mousa et al., 2015). A third problem can be a collusion attack, where multiple evil collectors act together to cause biased measurement data (Mousa et al., 2015).

Perera et al. (2014) stresses the challenge that sensing applications also need to guarantee and adhere to the highest security standards in order to ensure their trustworthiness. Therefore privacy protections and security certificates need to be introduced at several layers: they need to cover the technology layer, but also are they required to include the business and government level through strict legal terminology and conditions (Perera et al., 2014).

It also seems clear that the success of large-scale participatory noise sensing depends on social awareness and public technology acceptance (Davis, 1993). The model will stagger if people do not sufficiently embrace or trust collaborative sensing. Leadership roles are therefore required to proactively promote applications like NoiseTube to monitor and measure noise levels. Moreover, the developers need to put in place maximum user usability to ensure technology espousal of a wider society (Perera et al., 2014).

Another challenge remains with the technological quality of participatory sensing, where the question prevails as to whether today's cell phones are able to implement qualitative noise measurement. Although cell phone microphones have become more and more state-of-the-art, they do exhibit lower classification, meaning less accurate microphones (Wolfert, 2015). Further, as highlighted by Haklay (2010), crowdfunding technology needs to ensure positional accuracy in order to relate to the real value. Hence, a priority to implement the NoiseTube experiment will be to guarantee the technical validity of the used equipment.

3.6 Comparison between modeled and NoiseTube data

Continuing the investigation in Sections 3.5 and 3.4, the most critical difference between the NoiseTube approach and traditional measurement resides in the measurement method. Whereas traditional noise measurement is almost exclusively based on calculations and simula-

Table 1: Benefits of participatory noise sensing

Reusability and sharing	Following the logic of Internet of Things, the goal is to boost device communication and collaboration, requiring devices to be equipped with state-of-the-art sensing and communication capacities (Perera et al., 2014). With new devices available almost monthly, incentives are given to users to purchase the latest version, enabling high-quality noise sensing.
Integrated cloud computing	Collaborative noise sensing emulates the benefits of cloud computing models such as infrastructure-as-a-service, including scalability and availability of constantly reachable processing and storage resources. If there is a cost involved, users only need to pay for used data (Perera et al., 2014).
Decrease of data acquisition cost	Due to the data's collaborative and shared character the cost to acquire it will be significantly decreased stimulating further deployment (Perera et al., 2014). Technological advances further decline the mobile device manufacturing cost, making crowdsourced noise monitoring quite affordable, as most people do already possess a mobile phone for their own sake.
Collection of formerly inaccessible data	Participatory sensing allows sensor data gathering that was previously unreachable through non-interactive methods. As a result, the model also promotes empowerment, responsibility and feedback to enhance the user experience (Perera et al., 2014).
Opportunity for real-time decision and policy making	Collaborative noise mapping allows real-time data gathering from a variety of domains, hereby easing public decision-making
Applications and innovation	Due to the manageable data acquirement cost, increased amounts of citizens will have access to (previously unavailable) sensor data, which will fuel collaboration, innovation and creativity (Perera et al., 2014). Further, actors like government authorities or researchers will also be able to utilize those data to find solutions for various challenges within the SC concept, ranging from traffic to waste management.

tions, the NoiseTube method allows for a data-driven and measurement-based noise assessment procedure. A promising strength of NoiseTube is the opportunity for pattern identification. The NoiseTube approach gives the opportunity to assess noise dynamics via various characteristics - time, location, decibel level, individual - and can therefore conduct a much more specific noise analysis than mere model-based noise monitoring.

Furthermore, sensor movement is a key asset of NoiseTube: as people are in motion, so are their mobile phones including the noise sensors, compared to a limited amount of static sensors used as a means to cross-check and validate calculations and simulations. Hence, they are able to more efficiently collect data of local and micro-level noise events and to apply detailed neighborhood noise assessments. As a result, data acquisition cost, given a successful resident adoption of the NoiseTube approach, will significantly decrease, as people only need to register with NoiseTube to be able to download their data. Network effects would stimulate growing amounts of people to participate in collaborative noise data gathering, which would also increase the available quantity of data previously not available. In this context, the approach would also cause a shift from service-oriented architecture, representing a mere support function, to integrated cloud computing, supporting API analysis, data sharing or collaborative and IT-driven innovation.

Another valuable opportunity of the NoiseTube approach is that it can engender new innovations in the field of noise measurement. As the approach enables to identify several unexpected noise patterns, it could be the source of cutting-edge and innovative techniques and tools that would then enable new, technology-driven noise assessment procedures.

Additionally, the NoiseTube approach would also cause a great deal of citizen empowerment. The NoiseTube approach bolsters local, citizen-led action where participants create crowd-sourced noise maps in order to persuade local authorities on problematics that would usually require the time-consuming task of official data collection (Maisonneuve et al., 2010).

Moreover, the NoiseTube approach also permits indoor noise monitoring, which can turn out quite valuable as individuals spend a significant share of their time in indoor settings. Further, compared to traditional measurements, NoiseTube enables continuous noise recording, which is critical for short-term and prompt analysis and action planning. An overview of the main differences can be found in Figure 3.

Traditional Noise Measurement	Human Sensor noise monitoring
Service Oriented Architecture	Integrated cloud computing
Static sensing	Moving sensors
Outdoor measurement	Outdoor and indoor measurement
Expensive data acquisition	Affordable data acquisition cost
Limited in time	Continuous and real-time monitoring
Model- and simulation-based noise mapping	Measurement-based noise mapping
Rigid Measurement tools	Innovations and new applications
Macro & city level measurement	Detailed, local and neighbourhood level
Noise recording	Pattern identification

Figure 3: Comparison between modeled noise data and NoiseTube data

3.7 Design science

3.7.1 General

The NoiseTube experiment prevails at the nexus of people and technology. In order to receive academic validity, one needs to identify an accurate research pillar that would verify the process from a literature perspective. In other words, this paper needs to determine and emphasize a theoretical backbone for NoiseTube to academically validate the process.

The foundation to implement the NoiseTube approach resides within the field of Design Science (DS), a sub-category of IS research. Put in a simplistic way, DS investigates how things should be in order to achieve goals (von Alan et al., 2004). In the context of noise pollution, one can stipulate that DS endeavors to identify the ideal approach to address noise pollution, with reference to ensuring high living standards, but also by referring to the END and its goals.

DS conducts more than just a parallel investigation of technological or social systems, it focuses on examining the phenomena that arise when the two systems intermingle, making it a discipline that is at the crossroads of knowledge of properties of physical objects and the understanding of human behavior (Gregor and Jones, 2007). As the NoiseTube experiment can be deemed a symbiosis between technology (a mobile application) and human behavior (collaborative human noise sensing), it turns out as a noticeable ingredient to employ the idea of DS in a real-life context.

DS is per se a problem-solving procedure and the fundamental maxim of DS research is the comprehension and awareness of a design challenge, enabled by the procedure of building or judging an artifact (von Alan et al., 2004). The starting criterion to evaluate the polemic of noise measurement within DS research is therefore fulfilled. In order to complete the DS approach, this paper will specify both a DS research methodology as well as DS evaluation criteria for the NoiseTube experiment.

3.7.2 Research evaluation criteria

Not only is it useful and necessary to establish a methodology to proceed through an IS research experiment, one also eventually needs to evaluate the experiment's outcome from a DS perspective. In this context, seven design research criteria or guidelines have to be satisfied, depicted below in Figure 4. These criteria, which will be analyzed in the context of the NoiseTube approach, establish academic validation for this paper.

To initiate the evaluation, one can stress that guideline 1, stipulating that design research has to have an artifact in the form of model, construct, theory or instantiation, is satisfied. The NoiseTube approach engenders the application and use of an instantiation, the NoiseTube mobile application, which is then used to continuously monitor noise.

Thereafter, guideline 2 states the objective of a DS experiment, which is the creation of a technology-based solution to tackle relevant business problems. By having identified the application NoiseTube as enabler to deal with noise pollution, a critical issue for agglomerations, this paper is concordant to guideline 2 as well.

Guideline 3 states the design evaluation of the artifact, a procedure that will happen after the

Guideline	Description
Guideline 1: Design as an Artifact	Design-science research must produce a viable artifact in the form of a construct, a model, a method, or an instantiation.
Guideline 2: Problem Relevance	The objective of design-science research is to develop technology-based solutions to important and relevant business problems.
Guideline 3: Design Evaluation	The utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well-executed evaluation methods.
Guideline 4: Research Contributions	Effective design-science research must provide clear and verifiable contributions in the areas of the design artifact, design foundations, and/or design methodologies.
Guideline 5: Research Rigor	Design-science research relies upon the application of rigorous methods in both the construction and evaluation of the design artifact.
Guideline 6: Design as a Search Process	The search for an effective artifact requires utilizing available means to reach desired ends while satisfying laws in the problem environment.
Guideline 7: Communication of Research	Design-science research must be presented effectively both to technology-oriented as well as management-oriented audiences.

Figure 4: Design science research evaluation guidelines

noise data-gathering period, and therefore will be investigated posteriorly and discussed in Section 7.

Guideline 4 invokes the research contribution in the area of the design artifact. In this context, (Maisonneuve et al., 2010), (Maisonneuve, Stevens, Niessen, Hanappe and Steels, 2009), (Rana et al., 2015) and particularly (D'Hondt et al., 2013) have invested the procedure of participatory noise measurement through mobile phones and thus stress verifiable and clear research contributions on the NoiseTube approach.

Guideline 5 specifies the research consistency for both construction and evaluation of the design artifact. Accordingly, this criterion is somewhat linked to principles 3 and 4, and thereby can be, for one, assessed post-experiment and, for another, backed by relevant research contributions mentioned above. In fact, (Maisonneuve et al., 2010) delve into the construction of the NoiseTube application and thus provided required artifact research thoroughness.

Subsequently, guideline 6 requires the concordance to prevalent laws as well as the use of available methods to reach desired ends. This paper attempts to be consistent to the END and takes advantage of large-scale promotion mechanisms to proactively obtain sufficient data collectors, which is critical to accurate noise level assessments, to precise noise map construction, as well as to specific pattern recognition.

Guideline 7 refers to communication and stresses the requirement for appropriate presentation to both technology-based and management-based stakeholders. Results of this paper will be presented to both the city authority (i.e. management-based) and academic (i.e. technology-based) stakeholders after the submission deadline, and an appropriate process to convey the crucial arguments to these different audiences will be developed.

The NoiseTube approach indeed satisfies the seven DS criteria and, with the previously mentioned benefits, as well as the prevailing limitations of the traditional measurement approach can be identified as appropriate noise monitoring method. Furthermore, the NoiseTube experi-

ment's results will be discussed and evaluated in Section 7.3.2, with an emphasis on and implications for DS.

Nevertheless, Gregor and Jones (2007) stress that satisfying design research criteria through mastering relevant design theories or implementing state-of-the-art instantiations alone will not be sufficient. It is the combination of theories and artifacts that will be needed for human understanding and application (Figure 5). In this context, citizens establish theories (models, methods or constructs) to guide and understand the development of products in real-life scenarios.

In the context of noise pollution, it is the combination of NoiseTube (artifacts) and theories like collaborative noise sensing, together with individual citizen understanding that will be required to assess improvements in the analysis of noise pollution through human noise measurement.

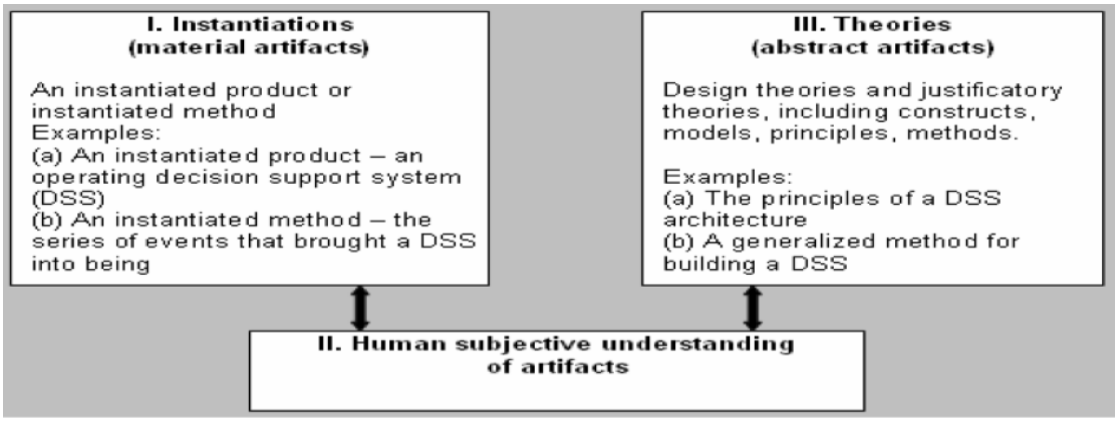

Figure 5: Relationship among design framework features

3.8 Crowdsourcing

3.8.1 Context

Albors et al. (2008) stresses that a notable share of human behavior comes from observing and modeling other human's behaviors, which do serve as code of conduct for upcoming action. Additionally, the growth of the Internet allowed users access to many websites and resources, especially in terms of social networking, which has enabled to solve many everyday problems, by outsourcing the problem to a certain group of people - a crowd - and by taking advantage of their "spare cycles" (Howe, 2006, p.1).

Crowdsourcing generally refers to the process of obtaining information from the crowd - the general public - of which many members carry communication equipment with them at any time in the form of mobile phones (Fienen and Lowry, 2012). Further, the process of allowing citizens to collect data through their phones has to potential to provide additional information unavailable by other, static methods, but also to raise awareness of scientific projects that eventually can significantly alter the public's quality of life (Fienen and Lowry, 2012). Therefore, the Web 2.0. procedures, described by network effects, openness and user contribution have caused to role of the public to evolve from passive information customer to active patrons (Maisonneuve et al., 2010). As such, the NoiseTube experiment, including the sharing of personal ex-

posure data is an opportunity to assess articulation between individualism and altruism in a real-life scenario Maisonneuve et al. (2010).

According to Heipke (2010), man citizens have become involved in the plotting of their everyday environment within the last five years, having thus taken a share in "geospatial crowdsourcing", made possible by rapid improvements in low-cost GPS-receivers, combined with a camera in smart phone or other mobile device. Additionally, one has experienced tremendous progress in collaborative methods, emerged with the progress of Web 2.0 tools like Wikipedia, as well as advances in communication links (Heipke, 2010). Even if citizen scientists lack formal expertise and training, produced "neo-geographical" maps turn out to be high quality, even without further map specifications and quality guarantee processes (Heipke, 2010).

Heipke (2010) further highlights that data creation by large amounts of volunteers, beginners as far as data gathering is concerned, is at the core of geospatial crowdsourcing. Subsequently, the crowd-sourced data is uploaded to and stored in a central platform, often a computer database. Thereafter, the task of automatic data processing is crucial to engender further feedback from measurement data (Heipke, 2010). Consequently, the NoiseTube approach is sufficiently consistent with the crowdsourcing literature, underlining the validation for this methodology.

3.8.2 The NoiseTube crowdsourcing component

In the past few years there has been a very rapid growth of interest in volunteered geographic information (VGI), a crowdsourcing feature in which the general public measures contributes geo-referenced data to websites where the data are aggregated into databases (Goodchild and Li, 2012).

Although individual contributors of crowdsourcing or Web 2.0. tools like NoiseTube usually have egoistic contribution motives at first, they eventually develop interpersonal connections that would then result in the emergence of, albeit weak at the beginning, opportunities of crowdsourcing (Maisonneuve et al., 2010). The possibility of sharing personal NoiseTube measurement information on a public profile includes the potential to transform user's motives from individual objectives to more crowdsourced and collective interests (Maisonneuve et al., 2010).

According to Goodchild and Li (2012), a critical factor of crowdsourcing resides within its importance for quality insurance for VGI, especially in cases where it is interpreted as the capacity to validate and correct the errors that an individual could commit. As an example, a software problem is more likely to be corrected if a multitude of engineers are trying to cope with it. Similarly, one can stress that the quality of NoiseTube data, both in terms of geometric exactitude and decibel level, is more probable to be accurate if several NoiseTube volunteers gather data, which can be considered a form a quality assertion (Goodchild and Li, 2012).

In order to fuel the crowdsourcing effect on NoiseTube, developers have included a user exposure board called "Elog" (Figure 6), representing the digital trace of a user's activity made public to the community (Maisonneuve et al., 2010). Hence, the idea is to support collaboration and interaction among noise gatherers to eventually achieve the desired outcome of crowdsourced citizen science, enabling residents to raise awareness about their living standards, which fuels democracy, transparency and shared environmental responsibility (Maisonneuve et al., 2010).

Tiwana (2013) stresses that another factor that reinforces the crowdsourcing theory resides in network effects. The more users there are, the more valuable the technology is deemed for an individual user. Again, the Wikipedia example serves it purpose to emphasize the concept: The more users contribute to the website, the more qualitatively well-written articles would be published (Bryant et al., 2005). In this context, the intention of NoiseTube developers is to design a tool that would boost network effects among volunteers, in order to obtain sufficient noise data to convince local authorities to grant the technology validity.

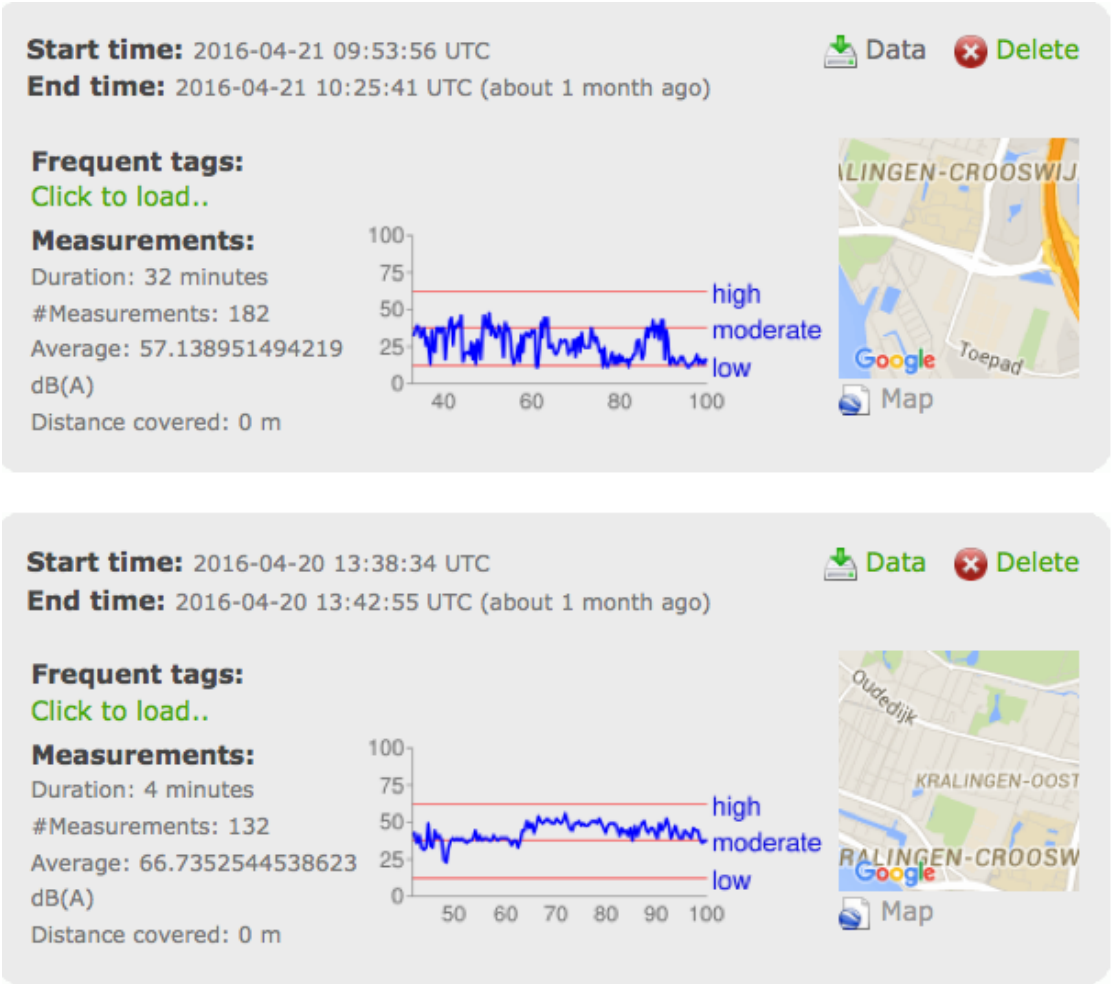

Figure 6: Extract: individual NoiseTube Elog board

3.8.3 Guidelines for practical crowdsourcing

A group that is doing collaborative measurement is very likely to have more in common than just an intention to contribute to mapping, namely social links (Heipke, 2010). To achieve a complete mapping process, it is thus critical to include members of various social settings. Whereas some participants might be more eager to implement valuable data gathering, others would be more energetic to report erroneous and omissible datasets, which will result in averaging out various volunteer features coming from several social environments (Heipke, 2010).

Heipke (2010) has identified several types of volunteer groups, all of which apply to the concept of collaborative noise data collection, and which are shown in Table 2.

Table 2: Types of volunteer groups

Data and map lovers	A small group of volunteers that creates highly valuable and trustworthy data, even willing to report erroneous data or measurement flaws.
Casual collectors	Casual data collectors partially intersect with data lovers but can be differentiated via their relatively moderate effort to record noise data. Further, casual collectors would rather input new data than report flaws.
Experts	Competent and leading data gatherers that are often active in organizations like traffic guiding or civil protection, involving a certain degree of mapping or data collection.
Media collectors	Citizens that can potentially constitute a very large share of the volunteers, but who are mostly one-off contributors guided by promotion or media campaigns.
Passive collectors	Any citizen carrying a device that has the capacity to passively gather data, often even without being aware of it.
Open collectors	People motivated by contributing cutting-edge public data, which currently hinders their collaboration with many public authorities.
Greedy collectors	Collectors who only participate because of money-based incentives or rewards.

As crowdsourcing results appear, one major concern is their quality. Heipke (2010) claims that it has to be guaranteed that results of a crowdsourcing data collection activity respect the same, standardized quality indicators (i.e. geometric accuracy...). This is the only solution to establish trust for crowdsourcing methods, especially when it comes to domains like noise pollution that can affect quality of life.

4 Data and Methods

This paper aims to gain insight into the dynamics of noise measurement in Rotterdam and to investigate whether a crowd-sourced noise measurement approach could be an appropriate auxiliary tool for noise measurement in cities. The subsequently elaborated methodology is based on three different process steps: The comparison between NoiseTube and modeled data (1), the construction of a model to analyze and predict user measurement behavior (2) and the discussion about implications of the NoiseTube experiment (3).

For the first step, the paper's method aimed to maximize citizen involvement in the collection of noise data through NoiseTube in a non-contrived setting, and for a three-month period. The more volunteers would carry out the NoiseTube approach, the more adequate and representative the experiment's results would be to compare the NoiseTube approach to the traditional method, but also investigate whether the approach can be granted validity as auxiliary pillar for future metropolitan noise measurement.

For the second step, the goal was to conduct an in-depth analysis of NoiseTube user behavior, potentially resulting in a predictive model. If the NoiseTube approach was to be considered a viable auxiliary for further noise measurement, it is clear that insight into aspects such as whether a person is likely to measure noise, how he or she would do it, for how long, where, and for what decibel levels are critical to guarantee the application's success in a future context.

Thirdly, the approach was to draw conclusions from the analysis, which attempted to satisfy the research objectives and to answer the research question. Recommendations were delivered for the inclusion of participatory noise sensing into future noise measurement. Further, the idea was to assess the implications for DS and to locate the analysis within its research foundation.

As promising as an analysis on quality of living through NoiseTube seemed, this remained out of scope with the given functionalities of the application. One was not capable to trace the gathered noise measurements back to their source as well as to conclude anything towards possible adverse heath effects. Therefore, the process of drawing conclusions of the impact of noise on quality of life remained unachievable with NoiseTube.

Despite the fact that action plans for noise level reduction should be based both on pervasive noise maps and existing "psychosocial" surveys (Cox and Palou, 2002) to stay consistent with the European Directive 2002/49/EC, the idea was to simply take advantage of certain individual survey insights in order to underline analysis outcomes. The received surveys were therefore used to back-up the analysis if applicable, provided their application remained valuable. The main results originated from resident survey responses about the impact of noise pollution on annoyance and sleep disturbances, giving rise to conclusions about losses in DALYs in Rijnmond, Rotterdam, Amsterdam and Utrecht.

A critical part of the paper resided within the sketching of the experiment as a valid research within the field of DS. Here, the methodology was to suggest a specified DS framework in the context of noise pollution and to then judge the latter's validity with the given DSRM and its evaluation guidelines. The DS framework, consistent with Gregor and Jones (2007), can have two different goals, either a methodology (1) or a product (2). This paper focused on goal (1) and covered the methodology, i.e. the comparison of two methods, as well as the investigation of user measurement behavior.

Consistent with von Alan et al. (2004), the design framework included four design artifacts: constructs, models, methods and instantiations. Instantiations refer to the material products that exist physically, as a piece of hardware or software, or as the series of physical actions (Gregor and Jones, 2007). This paper, similar to Gregor and Jones (2007), combined the first three into theories. Methods, constructs and models were considered necessary and useful prerequisites to guarantee a faultless functioning of a product in the case the NoiseTube application.

4.1 Data source and population sampling

A first step in pursuing a successful analysis consisted in gathering the relevant background information about noise pollution, traditional noise measurement but also the method. For this purpose, the collaboration with the DCMR turned out very fruitful due to multiple conversations with relevant employees, especially with Henk Wolfert, the policy officer of international and European affairs. Those interviews during regular meetings constituted a first step in gathering necessary background information on noise pollution and other related topics.

The success of the collaborative NoiseTube approach the subsequent analysis resided within a population's determinedness to volunteer to carry out collaborative noise measurement. In this vein, both the "Gemeente Rotterdam" and the DCMR supported promotion to maximize validity and city coverage. In this context, an article to present and boost the NoiseTube approach was included in the "Rotterdam Inside Out (RIO)" newsletter, which reached around 11.500 municipality employees.

As the goal was to obtain individual data from a maximum of suburbs to provide a realistic picture of noise levels, the procedure of convenience sampling was implemented. Data was thus collected by the ones that were the most rapidly and willingly available to participate in collaborative noise sensing (Sekaran and Bougie, 2003).

Consistent with Heipke (2010), in order to maximize the reach of the NoiseTube approach, the aim was to identify experts, open and casual collectors, as well as data lovers. Further, a major goal of crowd-sourced measurement was "the social experience to share a common (data collection) activity" (Heipke, 2010, p.552).

A method to do so was to conduct various individual conversations with citizens in order to obtain a small, but eventually valuable share of the afore-mentioned participant groups. Further, promotion presentations during university lectures were considered helpful to get people on board, as they enabled a very detailed and convincing project pitch. Follow-up reminder messages were then written in order to maximize efficiency and promotion output.

Clearly, it was important to inform volunteers about how to exactly proceed with the measurement, as the latter had follow several guidelines presented by the developers. Consequently, a user guide was distributed online and usage information had been given during the promotion presentations, attempting to achieve that people really knew how to use NoiseTube properly.

Moreover, in order to capture volunteers of other categories such as greedy or casual collectors, monetary incentives were included in the experiment, which has been taken care of by the DCMR, having added a voucher worth 150 Euros to be distributed among participants in a lottery procedure. In order to boost data collection, it was vital to allow to make it straightforward to contribute, without technological, legal or intellectual impediments (Heipke, 2010).

In this context, it was important to obtain a sufficiently large dataset, because of the fact that statistical averaging will be capable to counteract measurement flaws due to mobile phones, selection biases and incomplete data inputs. D'Hondt et al. (2013) have demonstrated that the calculation of noise averages can cope with random mistakes caused by wind or user conduct. In this vein, one can conclude that the larger the sample set, the higher the likelihood that errors differ, and hence that the average would compensate for these errors (D'Hondt et al., 2013).

It was also critical to choose volunteers depending on their location of domicile. In the case a person was residing in area with further opportunity to implement noise measurement, this individual was considered highly important in the measurement process, as he or she would be able to provide additional data, crucial for an accurate and complete noise map construction.

4.2 Data collection and equipment used

The collection of primary noise data was done through the application of NoiseTube, which users had to install on their mobile phones in order to use the latter as sensing devices. The mobile application was available for both IOS and Android phones, requiring phones to include GPS for localization as well as to support the "Java J2ME platform, including multimedia and localization extensions" (Maisonneuve, Stevens, Niessen, Hanappe and Steels, 2009, p.5) . Further, every volunteer needed to register on the website before starting to measure.

When measuring, the application visualizes the value of noise in real-time and decibels by associating a color to the respective health hazard (Maisonneuve, Stevens, Niessen, Hanappe and Steels, 2009). The following rationales apply in this context: green for <70 decibels (no risk), yellow for 70>noise<80 (be mindful) and red for >80 (serious risk).

Besides noise visualization, NoiseTube enables users to add information about the noise's origin or their respective level of annoyance, which can be quite valuable and meaningful for the creation of noise maps. Users are thus able to indicate the source of noise (i.e. traffic, construction...) and to establish an individual annoyance rating in the form of tags.

The experiment also tried to tackle the issue of lacking indoor placement. As noise pollution also considerably affects people that do spend their time indoors, NoiseTube gave volunteers the possibility describe their location using pre-specified tags (i.e. "Erasmus", "Ahoy") as an alternative to GPS-positioning (Maisonneuve, Stevens, Niessen, Hanappe and Steels, 2009).

An important step resided in the user privacy protection (Perera et al., 2014), which, in this case, was done by avoiding to use information that can link a measurement, a location and a name. One should be able to guarantee that NoiseTube would not disclose personal information about users. Although user information was not disclosed on the website, access to that sensitive information could have been obtained as a privilege from the developers. Nevertheless, the decision was taken to respect user privacy and therefore not take advantage of that kind of information.

Once a measurement activity had been closed, the data were uploaded onto the NoiseTube server and a confirmation mail was automatically sent to the user (Maisonneuve, Stevens, Niessen, Hanappe and Steels, 2009). Users were able to access and download their individual measurement data in JSON format (.json) and to contemplate them on noise maps with the support of Google Earth (Maisonneuve, Stevens, Niessen, Hanappe and Steels, 2009).

For the purpose of this paper, NoiseTube developers had granted a privilege to construct custom maps, which include all data collected in a specific area and time period. Similarly to individual measurement data, the customized data were available as .json, .kml and .shp files, and could be viewed on a map on the website, in Google Earth or downloaded for further analysis.

The visualization of customized maps as in Figure 7 was deemed handy, as it permitted to identify the areas that have experienced large-scale measurement activity, but also the ones where noise monitoring was scarce. An example of customized NoiseTube data is depicted in Figure 8.

Figure 7: Example of a customized noise map

The core focus of participatory noise measurement resided in personal exposure levels during day-to-day lives (D'Hondt et al., 2013). If this measurement form is to be taken seriously, one needed to evaluate the various types of equipment used, but also adhere to the same methods that modeled data are evaluated with (D'Hondt et al., 2013). Accordingly, to achieve adequate measurements of for instance traffic noise, equipment needed to take into account (1) the weighting curve (dB (A)), (2) direct read-out of sound levels in dB (A), (3) computation of LA,eq over random time periods, as well as (4) calibration (D'Hondt et al., 2013).

Calibration refers to a comparative process by which one device's readings are weighted against those of realistic reference devices, making it possible to identify systematic errors that can then be eliminated (D'Hondt et al., 2013). However, NoiseTube developers have not covered calibration for the most recently upgraded Android or IOS versions. As a result, non-calibrated phones might have recorded slightly different decibel levels.

Differences in recording primarily originated from discrepancies in microphone sensitivities or positioning (Santini et al., 2008). The need for calibration is mainly caused by the fact that microphones from mobile phones or other devices are designed for communication or music, but not for sensing tasks (Wolfert, 2015). Without clear calibration mobile devices may create data that is misleading or hardly useful (Maisonneuve, Stevens, Niessen, Hanappe and Steels, 2009).

```
[],"sum":0,"mean":null,"variance":-0.0,"stdev":0.0,"count":0,"maxdb":null,"mindb":null},
{"xmin":51.88685660531039,"xmax":51.88721617241111,"ymin":4.507222370361769,"ymax":4.507790055542654,"measures":
[],"sum":0,"mean":null,"variance":-0.0,"stdev":0.0,"count":0,"maxdb":null,"mindb":null},
{"xmin":51.88685660531039,"xmax":51.88721617241111,"ymin":4.507790055542654,"ymax":4.508357740723539,"measures":
[],"sum":0,"mean":null,"variance":-0.0,"stdev":0.0,"count":0,"maxdb":null,"mindb":null},
{"xmin":51.88685660531039,"xmax":51.88721617241111,"ymin":4.508357740723539,"ymax":4.508925425904423,"measures":
[],"sum":0,"mean":null,"variance":-0.0,"stdev":0.0,"count":0,"maxdb":null,"mindb":null},
{"xmin":51.88685660531039,"xmax":51.88721617241111,"ymin":4.508925425904423,"ymax":4.509493111085308,"measures":
[],"sum":0,"mean":null,"variance":-0.0,"stdev":0.0,"count":0,"maxdb":null,"mindb":null},
{"xmin":51.88685660531039,"xmax":51.88721617241111,"ymin":4.509493111085308,"ymax":4.510060796266193,"measures":
[],"sum":0,"mean":null,"variance":-0.0,"stdev":0.0,"count":0,"maxdb":null,"mindb":null},
{"xmin":51.88685660531039,"xmax":51.88721617241111,"ymin":4.510060796266193,"ymax":4.510628481447077,"measures":
[],"sum":0,"mean":null,"variance":-0.0,"stdev":0.0,"count":0,"maxdb":null,"mindb":null},
{"xmin":51.88685660531039,"xmax":51.88721617241111,"ymin":4.510628481447077,"ymax":4.511196166627962,"measures":
[],"sum":0,"mean":null,"variance":-0.0,"stdev":0.0,"count":0,"maxdb":null,"mindb":null},
{"xmin":51.88685660531039,"xmax":51.88721617241111,"ymin":4.511196166627962,"ymax":4.511763851808847,"measures":
[],"sum":0,"mean":null,"variance":-0.0,"stdev":0.0,"count":0,"maxdb":null,"mindb":null},
{"xmin":51.88685660531039,"xmax":51.88721617241111,"ymin":4.511763851808847,"ymax":4.512331536989731,"measures":
[],"sum":0,"mean":null,"variance":-0.0,"stdev":0.0,"count":0,"maxdb":null,"mindb":null},
{"xmin":51.88685660531039,"xmax":51.88721617241111,"ymin":4.512331536989731,"ymax":4.512899222170616,"measures":
[],"sum":0,"mean":null,"variance":-0.0,"stdev":0.0,"count":0,"maxdb":null,"mindb":null},
{"xmin":51.88685660531039,"xmax":51.88721617241111,"ymin":4.512899222170616,"ymax":4.513466907351501,"measures":
[],"sum":0,"mean":null,"variance":-0.0,"stdev":0.0,"count":0,"maxdb":null,"mindb":null},
{"xmin":51.88685660531039,"xmax":51.88721617241111,"ymin":4.513466907351501,"ymax":4.5140345925323855,"measures":
[],"sum":0,"mean":null,"variance":-0.0,"stdev":0.0,"count":0,"maxdb":null,"mindb":null},
{"xmin":51.88685660531039,"xmax":51.88721617241111,"ymin":4.5140345925323855,"ymax":4.51460227771327,"measures":
[],"sum":0,"mean":null,"variance":-0.0,"stdev":0.0,"count":0,"maxdb":null,"mindb":null},
{"xmin":51.88685660531039,"xmax":51.88721617241111,"ymin":4.51460227771327,"ymax":4.515169962894155,"measures":
[],"sum":0,"mean":null,"variance":-0.0,"stdev":0.0,"count":0,"maxdb":null,"mindb":null},
{"xmin":51.88685660531039,"xmax":51.88721617241111,"ymin":4.515169962894155,"ymax":4.5157376480750395,"measures":
[],"sum":0,"mean":null,"variance":-0.0,"stdev":0.0,"count":0,"maxdb":null,"mindb":null},
{"xmin":51.88685660531039,"xmax":51.88721617241111,"ymin":4.5157376480750395,"ymax":4.516305333255924,"measures":
```


Figure 8: Example of NoiseTube data in .json format

Prerequisites (1), (2) and (3) were covered, as NoiseTube employs an A-weighting filter to the recorded sound and then computes the equivalent sound level Leq, measured in dB (A) (D'Hondt et al., 2013). For criterion (4), every user had the opportunity to indicate his or her phone model used when proceeding with the registration (D'Hondt et al., 2013). This was critical, as the NoiseTube developers have implemented a calibration feature for a significant share of the most widespread smartphones in circulation and published the list on their website.

Hence, every user was able to trace the validity of individual measurements when cross-checking his or her phone model with the calibration database of NoiseTube. For several non-calibrated phones, calibration was done with the support of the DCMR, where the respective IOS and Android phones were cross-checked to a professional SLM - a class 1 SLM model - when used for measurement. In this context, the decision was taken to arrange meetings with the DCMR in several locations or neighborhoods, in which noise levels measured through NoiseTube were compared to the level captured by a professional SLM. This allowed to mostly satisfy criterion (4), although most users would already use a phone model calibrated by NoiseTube developers.

Consistent with the characteristics of a correlational study, the idea was to minimally interfere with the actual noise occurrence and to monitor its levels "as they appear", during day or night, and ensuring a merely marginal researcher intervention (Sekaran and Bougie, 2003). It was important to measure noise at identical locations during different times of the day in order to obtain an accurate representation of the actual noise level prevailing in a given area. Hence, a detailed outline of the task was required in order to obtain outstanding volunteering results.

4.3 Data analysis and method comparison

4.3.1 Graphical inquiry

The NoiseTube platform, in order to fuel insight and research, gave people access to its data via web API (Maisonneuve, Stevens, Niessen, Hanappe and Steels, 2009), which was then used to download and analyze the data. Given a successful API analysis, the results were then synthetized and presented graphically to enhance understanding.

For the analysis, the NoiseTube data were imported into the R tool, which was deemed appropriate for further analysis as well as for graphical visualization. The idea was to first proceed through the analysis with a sample dataset before applying the code to the final dataset.

Geometric accuracy, which had been deemed one of the challenges in Section 3.5.3, was delivered by taking advantage of a phone's built-in GPS and returned WGS84 coordinates (Maisonneuve et al., 2010). The phone location coordinates, according to Maisonneuve, Stevens, Niessen, Hanappe and Steels (2009), never vary by more than 30 meters and, given the current state of technology, are mostly considered an "exact science" (Rao and Minakakis, 2003, p.1).

Further, NoiseTube also takes advantage of a map matching algorithm which relies on GIS databases provided by municipalities or other crowdsourced services like OpenStreetMap (OSM) (Maisonneuve et al., 2010). Thereby, in order to account for possible GPS positioning flaws, NoiseTube matches every datapoint that is not situated on street to the nearest location that is (Maisonneuve et al., 2010). Given the fact the OSM is also a crowdsourced tool, its quality affirmation is confirmed for cities by the fact of having a multitude of contributors (Goodchild and Li, 2012). Consequently, despite the fact that errors can persist in metropolitan areas, the positional accuracy of NoiseTube can be deemed valid for the sake of this paper.

In parallel, two different noise datasets were received by the DCMR for the purpose of the method comparison.The idea here was to import the two DCMR datasets into QGIS, an open-source geographic information system (GIS). In QGIS, both DCMR and NoiseTube data were merged for visualization before exporting the combined attribute table into R for further statistical analysis. The entire process followed a step-by-step analysis, detailed in Table 3.

The first DCMR data set included road traffic noise data from 2007 in the form of contours, shown in Figure 9. By contours, one understands that noise levels were averaged over a certain grid location and then established at the edge of those grids, in other words, at contours. This dataset was used for a first method comparison.

Subsequently, the NoiseTube data were compared to a second dataset containing road traffic noise data at building level and at a 4-meter height (Figure 10). In this case, a much more precise overview of noise data was given at building level, with one decibel level being recorded behind a building and one in front of it. Both DCMR datasets were then used for visualization and statistical analysis in order to implement the comparison between the two data types.

The desired outcome of the method comparison was to provide recommendations and conclusions based on possible differences in noise levels gathered by graphical interpretations. However, visual observations alone are not enough to draw conclusions about the results of the method comparison (D'Hondt et al., 2013). Subsequently, the approach in order to present

Figure 9: 2007 DCMR noise data at the contour level

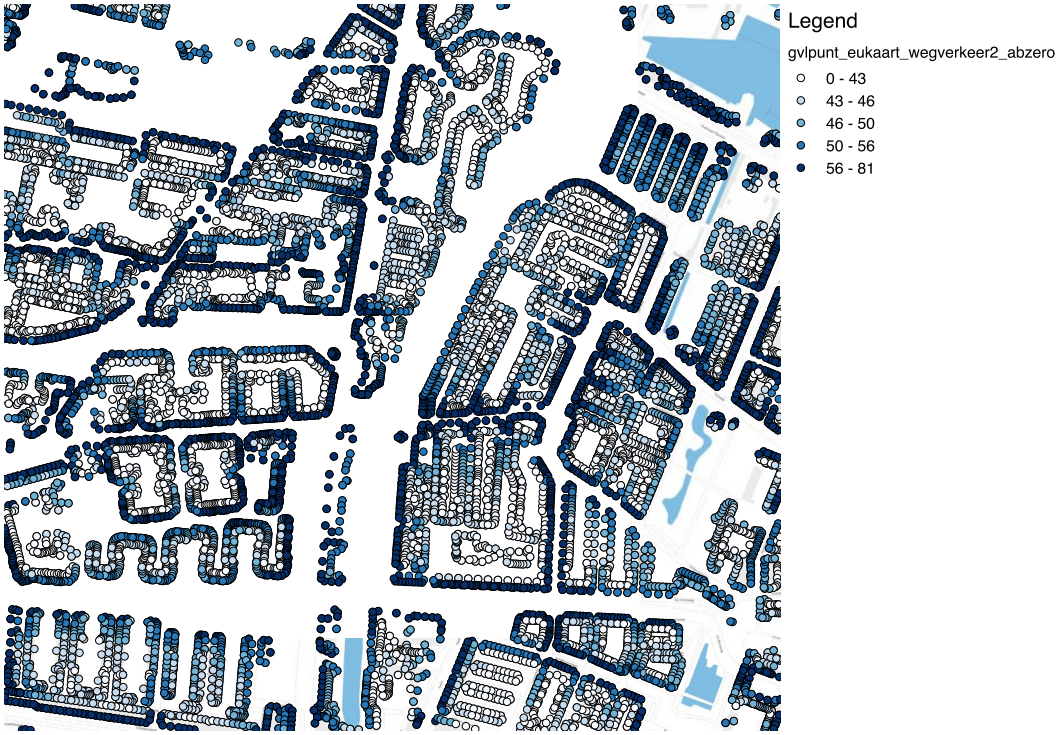

Figure 10: 2011 DCMR noise data at the building level

Table 3: Different steps in the comparison between NoiseTube and DCMR data

1. Data preparation	Data preparation included the import of NoiseTube data into R studio, the data cleaning and tidying such as the elimination of redundant variables, but also several test visualizations in order to investigate whether the data depict a realistic noise dynamics scenario. The idea here was to create a data frame in R that classifies the noise data by user, date, decibel level, coordinates, and classifies the noise levels into several categories. Further, the approach was to determine summary statistics such as average noise level or spread. As a single customized map obtained through the NoiseTube website did not cover the entire area where data had been gathered, the idea was to create multiple custom maps and to then ultimate combine them into one final data frame.
2. Outlier detection and exclusion	In order to implement a representative analysis, several types of noise data had to be excluded from analysis: Firstly, these included data where meteorological conditions were unfavorable, for instance with high windspeed or significant precipitation. Another possibility to exclude data is by looking at their respective tags, which could indicate that their was a disturbing noise factor that would permit exclusion. Moreover, one had to find out whether data could considered biased or flawed, or, in other words, had to be excluded because its noise levels are not in the L95-5 range.
3. Map construction	It was important to divide the city of Rotterdam into squares or grids for map construction, as this enabled the aggregation of noise data per grid and hence a more powerful visualization. The first attempt used grids of 100 by 100 meters, as shown in appendix D. In this context, it was thus critical to also sort the NoiseTube data by coordinates and to then insert the sorted data into the relevant grid or square. Further, for visualization enhancement, the data were aggregated and points in the coverage map created in R through ggmap (Kahle and Wickham, 2013), but also in the general noise map were noise level means, not individual measurements.
4. Importation of DCMR noise data	Import the noise data from the DCMR, available as .shp files for traffic noise. The first dataset included data from 2007, before carrying out the final procedure with more recent noise data. This step was done through QGIS.
5. DCMR map creation	In QGIS, construct the same or similar form of noise maps as with NoiseTube data. This was done by adding a first layer of .shp noise data into QGIS. Subsequently, another layer consisted of noise information obtained through the Publieke Dienstverlening Op de Kaart (pdok) plug-in, representation another way of adding supplementary noise data. Eventually, the data from NoiseTube were imported into QGIS and added onto the newly created map.
6. Map comparison and statistical testing	Comparison of the two datasets through means of visual map interpretation in order to identify potential differences in noise levels. Undertake further comparative tests through statistical methods.
7. Conclusions and recommendations	The identification of potential differences in noise levels between the two maps enabled a more realistic picture of noise levels in Rotterdam, giving the DCMR opportunities to modify action plans to more efficiently reduce, prevent or eliminate noise.

suggestions was therefore to conduct comparison and testing via statistical means in R in order to further interpret and underline result validity.

4.3.2 Statistical testing

D'Hondt et al. (2013) applied several descriptive statistical techniques to assess the accuracy of the noise maps, for instance did they compute confidence intervals over sample sets. Standard deviations (SD) would then indicate how much individuals within the sample would diverge from the sample mean. However, sound levels represent a highly variable parameter, which is why SD would be elevated anyway and therefore not give any insight into the quality of the sampling process nor the noise level accuracy (D'Hondt et al., 2013).

Two statistical tests were thus chosen to implement a statistical comparison. For one, the idea was to take advantage of the Spearman rank-order correlation, a nonparametric correlation test that measures the associative power between two variables (Artusi et al., 2002). As the test required the comparative variables to be either ordinal or interval, the test criteria were met, validating its application. The approach was to join the two data layers in QGIS and to then export the corresponding attribute table into R, where the Spearman correlation test was conducted.

The ordinal variable consisted in the noise level mean of each grid, whereas its counterpart, coming from DCMR data, was presented in the form of a noise level interval for the contour data category and a mean variable for the building data category. The correlation coefficient varies from $+1$ to -1, where a $+1$ outcome would imply perfect positive and a -1 perfect negative association between variables. The null-hypothesis would assume a zero correlation between two variables, i.e. a rho of 0.

For the Spearman correlation, the null-hypothesis would indicate that there would not be an association between the traditional and the crowdsourced noise measurement method, resulting in considerable differences in noise levels (Artusi et al., 2002). Hence, if the null-hypothesis could not be rejected, one can conclude that there is no statistical guarantee that differences in noise levels are not due to the measurement method, in other words, there is no certainty that two methods do not produce diverging results.

For another, the idea was to conduct the Wilcoxon signed rank test (Bridge and Sawilowsky, 1999). Noise levels coming from two datasets are paired by location, in this case by cellular grids, and as one would like to investigate whether noise levels are different depending on the measurement method, the usage of the non-parametric Wilcoxon signed rank test is appropriate. The assumptions of the Wilcoxon test, namely independently chosen pairs, data coming from same population, and ordinary numbers, are fulfilled (Bridge and Sawilowsky, 1999).

In the case of the Wilcoxon test, the null hypothesis consists in the fact that both methods show no differences in noise levels, i.e. that there is no change when opting for a collaborative noise measurement approach compared to simulation-based noise level calculations. As a significance level, 0.05 was chosen, and for a given number of observations, significance is determined by the fact whether the received W value is below a certain critical threshold. In the case the W value is larger than the threshold, the conclusion would be that it could not be excluded that the difference in noise levels could be due to the existence of two different noise measurement methods.

4.4 Predictive model and user measurement behavior analysis

It seemed valuable to implement a model that predicts future user measuring behavior, because the success of a large-scale launch of the NoiseTube approach as an auxiliary noise measurement tool, similarly to other information systems, clearly depends on how Rotterdam's residents embrace the technology. The idea was thus to build a model that would determine future user measurement behavior.

Zukerman and Albrecht (2001) stress that, generally speaking, statistical models deal with the utilization of a verified test sample in making predictions about unobserved dependent parameters. In the context of user modeling, the unknown variable represents an aspect of a user's future actions and user modeling thus attempts to conclude hidden information about a user from his available data by taking into account inference uncertainty (Zukerman and Albrecht, 2001). As a result, predictive statistical user models can depict anticipation of certain features of human behavior such as goals or actions (Zukerman and Albrecht, 2001).

A model of human conduct could be an enabler for reinforced human-machine systems. In the case of NoiseTube, if the mobile application, could recognize human behavior or even more, if it was able to anticipate it, NoiseTube could adjust its settings to better serve the needs of its users (Pentland and Liu, 1999), and therefore enable a smoother noise measurement process.

The concept of artificial intelligence has been able to produce several statistical methods and techniques that enable to establish a model to predict user behavior and two different approaches are generally used to perform the task of predictive modeling: content-based and collaborative (Zukerman and Albrecht, 2001). This paper will adopt the first approach, which stresses that an individual will depict a peculiar behavior under a given set of conditions and this behavior will be repeated under similar or equivalent conditions (Zukerman and Albrecht, 2001). In this vein, the behavior of users is predicted via his or her past conduct.

The idea of this paper was to initiate the model creation through a clear understanding of the dataset, which would then enable to determine the techniques and algorithms that can be applied in order to obtain insightful results. Examples of predictive modeling techniques within machine learning include decision trees, neural networks, as well as Bayesian networks (Zukerman and Albrecht, 2001). Among several techniques, the decision as to which one to apply will be communicated and argued in the following section.

A first test to assess the existence of patterns in user data was to implement a nearest-neighbor analysis, which would return a Z-score, pointing in the direction as to whether rejecting or accepting the null-hypothesis would be employed. In general, the existence of a very high (positive) or very low (negative) Z-score indicates that the data pattern would be to unusual to be coming from random occurrence.

4.4.1 Model creation

As one would like to gain insight as to whether NoiseTube can be launched as an auxiliary tool for noise measurement, the idea was to find out whether NoiseTube will be accepted by the large amount of residents of a city. Hence, the goal of the model was to predict whether people are likely to continue to measure noise through NoiseTube. Several steps prevailed with the creation of a predictive model, detailed in Table 4 and implemented in R (Appendix F).

Table 4: Different steps in the creation of a user model in the NoiseTube context

1. Data preparation	The data preparation included the importation of the initial dataset obtained through merging the various custom maps from the NoiseTube website. In fact, the initial dataset needed to be taken here, as the aim was to create the predictive model for people's measurement behavior, for which the fact as to whether people measured indoor noise did not matter. The same reasons were true for unfavorable meteorological conditions, as predicting whether people measure would also base its forecast on measurements taken during hostile weather periods.
2. Create target and predictor variables	In order to implement a representative model, one had to create a target variable as well as predictor variables. A very important concept was to take into account the time dimension in both the target and the predictor variables, which was done by establishing a break between the months in which measurement data were available. In fact, the model's target variable whether someone was likely to continue measuring had to be created with data after that break, whereas the predictors such as the number of locations or the mean noise level had to be implemented with monthly data before the time break.
3. Reduce table to one row per user	As the model's target and predictor variables were categorized by user, the idea for the model was to obtain a data frame with one row per user.
4. Holdout validation	For the assessment of model performance, it was critical to split the data into a training and test set, taking advantage of the former to create the model and of the latter to evaluate its process (Zukerman and Albrecht, 2001).
5. Run and evaluate model and interpret results	This step includes the definition of appropriate model evaluation criteria in R before then running the model and evaluating its performance.
6. Interpret results	This step attempts to interpret results in order to highlight potential findings that would permit the formulation of recommendations.
7. Discussions and conclusions	The drafting of recommendations and conclusions as to whether and how NoiseTube can be used as an auxiliary tool for noise measurement, giving the DCMR insight into how to initiate a potential large-scale deployment of NoiseTube. This step was done in in Section 7.2.

The techniques applied for the creation of the predictive model were decision tree and logistic regression. Decision trees emphasize rules to divide data into groups. In fact, they initiate its first split into branches coming from a root before eventually splitting into branches coming from nodes or other branches. The tree's leaves are then considered the final non-split groups (Koppius and Belo, 2015).

The chosen approach was to create a logistic regression model to check the significance of the predictors, then to run the model with a decision tree as well as through a logistic regression model in order to compare its performance against the one of the decision tree. Neville (1999) propose decision trees because they offer an outstanding opportunity to interpret the model's results. Further, they represent an intuitive method to understand the data structure and help to understand the results of other predictive models (Neville, 1999).

4.4.2 Model evaluation

The challenge with predictive models is generally to find out whether they are actually equipped with decent predictive power. In this context, the task is to determine whether the predictive model is applicable to other settings than the one used to build the model, in other words, to ensure the model is generalizable to other contexts (Koppius and Belo, 2015). In the case of NoiseTube, this means that one had to detect whether the forecast as to whether a user is likely to continue his measurement was not exclusively applicable to the cities that have been used to gather data to build the model, but also for other data. Scientifically put, the approach was to avoid model overfit, referring to the fact that a model should not only provide insight into the dataset that was used for its establishment (Koppius and Belo, 2015).

In this paper, several concepts were applied to verify model generalization for the decision tree. For one, it was made sure that the model did not grow a tree until its leaves are pure, which would be a sign of model overfit. To do so, the approach was to prune back the decision tree in case it would turn out to be too large, represented in the tree creation command in the R script in Appendix G .

Further, the use of model accuracy as an evaluation technique has been implemented. The idea was to apply holdout validation by splitting the data into training and test set. The former was then used to create both the logistic regression as well as the decision tree and the latter to evaluate their performances through holdout accuracy (Koppius and Belo, 2015).

The next step to evaluate decision tree and logistic regression model then consisted in creating a confusion matrix and in visualizing the performances through the receiver operating characteristic (ROC) curve as well as to calculate the area under the curve (AUC). The decision for ROC has been taken because it enables the organization and selection of classifiers contingent on their respective performance (Fawcett, 2006).

Both the ROC and AUC step were completed in R through the creation of specific functions and plotting procedures. This enabled to judge model quality, as well to assess the power of given predictions. Once the model was built and evaluated, its results were discussed and interpreted in Section 7.2.

4.5 Design science: Research methodology

Peffers et al. (2007) stipulate that every DS project, in order to have a fruitful result, requires a well-specified research methodology. There is considerable agreement among the most respected DS experts that six different tasks have to be completed to comply with the DS research methodology (DSRM) (Peffers et al., 2007). Hence, the NoiseTube experiment had to follow from a certain DS methodology and was then evaluated in the realm of DS in Section 7.3.2.

Activity one stressed the problem identification and motivation: the idea was to seek for specific research polemic in order to then prove the value of its potential solution. The justification inspires the researcher to promptly investigate the subject as well as to better understand the problem's rationale. In the case of NoiseTube, this had been done for the purpose of this paper, but also during multiple promotion presentations in order to persuade a maximum of volunteers to participate in the experiment.

Activity two concerned the definition of solution goals: here, the approach was to derive quantitative or qualitative objectives that assessed the feasibility and possibility of solutions. Examples include the description of how a product is expected to support solutions to several challenges not addressed beforehand. Broadly seen, the objective was to gain insight into noise dynamics in the city of Rotterdam. More specifically, the goal in this context was to assess whether a crowdsourced noise measurement campaign was able to constitute a powerful and valuable auxiliary to traditional noise measurement.

Activity three stated the design and development. In this context, the exerted rationale was to design an artifact, which can range from constructs, models, methods or instantiations, but also to determine the artifact's components and its architecture. In this context, this step had been delegated to other parties. Given this paper's rather strict timeline, the idea was to implement an examination of various noise measurement mobile applications in order to then choose the most straightforward and valuable one.

Activity four dealt with the demonstration. Speaking for itself, this activity prescribes a demonstration as to how the artifact works, for instance in a simulation, case study or experimentation. During the process of promotion, many demonstrations about how to use NoiseTube were conducted, in order to assist people in gaining the right momentum for the mobile application.

Activity five stressed the evaluation of the artifact. This activity has been explained in detail in Section 3.7.2. Lastly, activity six stipulated the communication of results, but also of the problem, its importance, its novelty and of the design. The audience for communication includes all relevant stakeholders such as politicians, urban planners, academics, researchers and measurement volunteers, and the communication will be done in August or September 2016.

5 NoiseTube data description

5.1 Context

The idea of this NoiseTube experiment is to verify to what extent a crowdsourced noise measurement through an application like NoiseTube could embody an auxiliary technique to complement traditional noise measurement described in Section 3.4, including a way to gain insight into noise origins as well as potential information about health-related noise effects.

In order to be able to validate the NoiseTube approach, one has to ensure NoiseTube measurements at given location are similar or almost identical in noise levels to measurements taken by professional SLM. To address this, measurements through NoiseTube and via a professional SLM were taken simultaneously at several locations in Kralingen on Wednesday, April 13, 2016. The result of this test confirmed the validity of the NoiseTube approach, as measurements differed only by 2-3 decibels, which averaged out when processing the data to the LA,eq indicator. Hence, this paper would acknowledge the similarity of NoiseTube to professionally recorded noise data, one of the challenges in section 3.5.3.

Furthermore, consistent with the crowdfunding concept in Section 3.8, one has to assure that the experiment's stakeholders, like DCMR representatives, develop a feeling of trust. In order to do so, results of a crowdsourcing experiment have to comply with the standardized quality indicators used for traditional noise measurement, for instance geometric accuracy. For one, this was already taken care of by verifying the noise level similarity during the cross-checks with a SLM. For another, this was tested through QGIS, as one had to apply a function that joins the NoiseTube data layer and the respective DCMR layer by location. Hence, without geometric accuracy, it would not have been possible to combine the layers.

The reach of NoiseTube turned out moderate and promotion insufficient to motivate a far-reaching amount of volunteers. Consequently, the data coverage of Rotterdam, as well as the predictive power of the user measurement model do remain marginal. The promotion of the experiment was able to reach data lovers, but not casual or greedy collectors and it did not yet reveal as a "social experience to share a common data activity" (Heipke, 2010, p.552). Reasons for this deficient measurement effort reside primarily within the absence of large-scale incentives, but also can be drawn back to the fact that people would frequently forget to measure.

The NoiseTube application, despite its straightforward usage, includes several shortages that need to be addressed before it can be taken advantage of in the case of large-scale launch of a crowd-sourced noise measurement approach. For one, the application is experiencing multiple crashes which, depending on the phone model, can cause a severe impediment to a sustained volunteering effort. Furthermore, the application, similar to many others, consumes a significant amount of phone battery, which can also impede many people's measurement activity.

For another, the application does not yet sufficiently encourage active information contribution and sharing. In fact, it allows users to add tags to their measurements, however, this is disturbing when simultaneously holding the phone to measure. Consequently, added tags are an exception, advocating a need to alter the functionalities to a more straightforward way to inform about location, noise source or annoyance. As a result, NoiseTube does neither empower its users or urban planners to profoundly dive into prevailing noise sources, nor does it enable

to establish noise-caused causalities between people's health disturbances and their annoyance. Nevertheless, data collection resulted in a somewhat large dataset used for further analysis and described in the upcoming sections of this paper.

An impediment to the NoiseTube data analysis resided within the fact that the API from the NoiseTube website was limited to the last 500 measurements. This caused issues for both map creation and data analysis in R. For the former, it resulted in the fact that the PHP tool could not be used to create the maps, as it has to take advantage of API credentials when loading customized maps in a browser. These credentials can only be retrieved via the NoiseTube API, which, limited to 500 measurements, would not result in a representative map.

For the data analysis, the API limit of 500 measurements would be largely insufficient to realize a profound noise analysis of Rotterdam, because of the fact that when measuring with NoiseTube, one measurement is taken every two seconds when using the application. Hence, 500 measurement would never suffice to conduct a complete analysis. In fact, a sample dataset would already consist of around $60,000$ data points, underlining the limitation of the NoiseTube API and hence of its use for this paper.

Due to the privilege given by NoiseTube developers to create customized maps including city-wide data access, this did not constitute a groundbreaking obstacle for this paper. The data of the customized maps was first imported into R in the form of a sample test set and provided fruitful ground for further analysis. A clear view of the dataset is given in Figure 11.

The NoiseTube data from Rotterdam showed maximum noise level of 128.51 dB, which is not surprising, given the fact that measurements were mostly taken in the city centre of a $600,000$ resident agglomeration. Minimum noise levels amounted to around 20 dB, which is going to be discussed in Section 5.3 below. Mean noise levels amounted to around 40 dB, primarily due to the fact that most volunteers would circulate around Kralingen, which is a rather low-noise area. Further, multiple measurements would have been taken indoors, where noise levels are considerably lower than outside.

However, as mostly explained in the next section, the coverage of NoiseTube data turned out to be marginal, resulting in the fact that the decision was taken to randomly generate supplementary data through an R function named "jitter". Doing so allowed a more widespread dataset to be used for analysis, map construction, as well as method comparison. As a result, the first dataset with genuine NoiseTube data was expanded by randomly generated jitter data and used as final data for visual representations, as well as method comparison.

Furthermore, a different approach was chosen to visualize the noise data and to create the noise maps. Whereas the first idea was to create a dynamic noise map that would depict characteristics of each data point (i.e. coordinates, noise level, date, user). However, this would not result in a very adequate representation, firstly because the dataset contains an excessive amount of data and secondly due to the fact that this is not advantageous for the comparison with noise maps created through DCMR data.

In fact, the approach eventually taken was to construct maps that were similar to ones established by the DCMR (see Figure 1). Hence, the solution in R is to create several layers that are then added onto each other. As a result, the maps created with NoiseTube data included a Google Maps layer for one, as well as a layer with the noise level information for another.

Filter

	loudness	date	user_id	lat	lng
1	60.167	2016-02-09	5072	51.90185	4.496280
2	60.892	2016-02-09	5072	51.90201	4.496317
3	61.007	2016-02-09	5072	51.90213	4.496355
4	62.207	2016-02-09	5072	51.90223	4.496379
5	62.494	2016-02-09	5072	51.90234	4.496399
6	61.793	2016-02-09	5072	51.90247	4.496436
7	61.194	2016-02-09	5072	51.90257	4.496452
8	64.348	2016-02-29	5072	51.90610	4.487807
9	63.821	2016-02-29	5072	51.90610	4.487807
10	62.988	2016-02-29	5072	51.90607	4.487876
11	64.247	2016-02-29	5072	51.90600	4.488007
12	67.667	2016-02-29	5072	51.90602	4.488054
13	65.366	2016-02-29	5072	51.90607	4.488022
14	67.697	2016-02-29	5072	51.90604	4.488119
15	68.264	2016-02-29	5072	51.90603	4.488144
16	69.934	2016-02-29	5072	51.90603	4.488144
17	72.229	2016-02-29	5072	51.90603	4.488144
18	67.932	2016-02-29	5072	51.90609	4.488145
19	68.962	2016-02-29	5072	51.90610	4.488151
20	69.913	2016-02-29	5072	51.90610	4.488151
21	70.632	2016-02-29	5072	51.90615	4.488138
22	65.510	2016-02-29	5072	51.90616	4.487736

Showing 1 to 22 of 125,278 entries

Figure 11: An example of a NoiseTube dataset

5.2 Coverage

In terms of region or city coverage, results showed that NoiseTube data did not canvas the entire Rotterdam area. Most measurements were taken in areas that consisted of high student residency like Kralingen, or transit movement, such as Rotterdam Centrum or Capelle aan den IJssel, as shown in Figure 12. Consequently, the NoiseTube approach is not able to implement a city-wide noise analysis at this point in time.

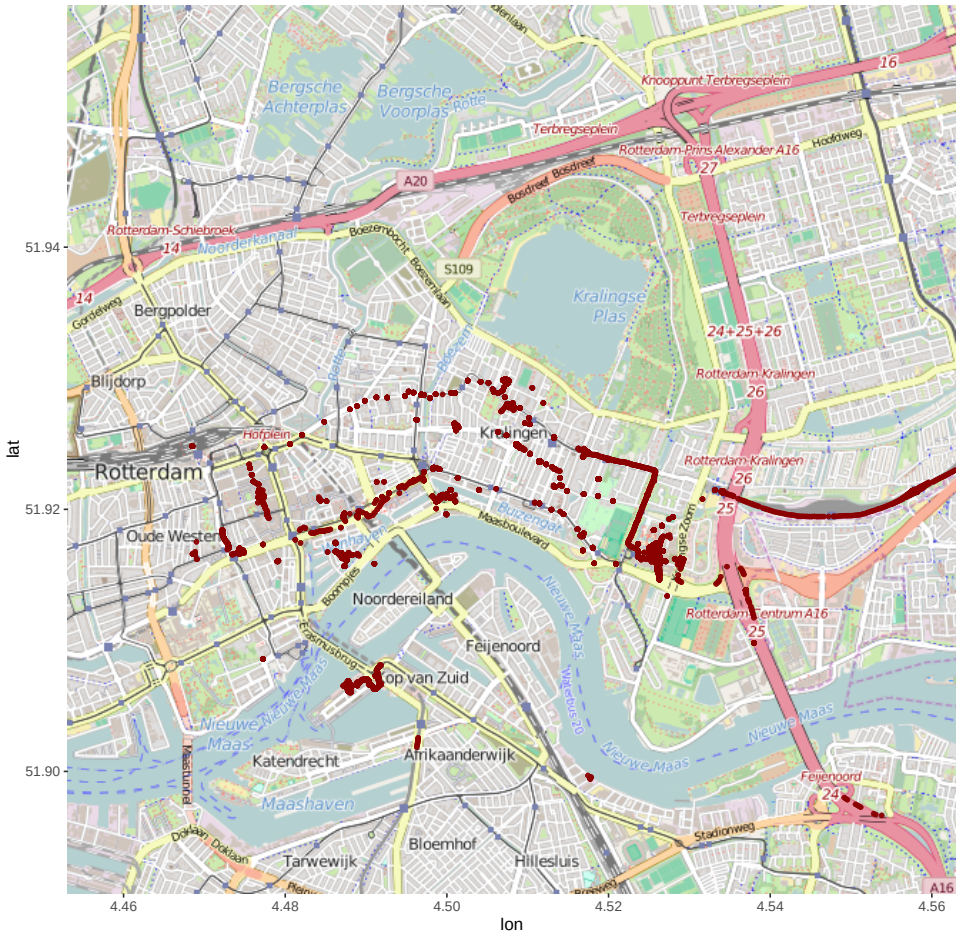

Figure 12: Rotterdam NoiseTube data coverage

As the customized map featured received by NoiseTube developers did not permit the compilation of map that would cover the entire area with existing data, several maps had to be constructed and their data imported into R. Subsequently, the established data frames were then merged in order to continue the analysis with one data frame covering all gathered data for city of Rotterdam.

In this context, the final dataset imported into R and prepared for the method comparison analysis consisted of 451, 004 individual measurements (jitter included) taken between January 15, 2016 and May 1, 2016. In total, 41 volunteers contributed in gathering noise data for Rotterdam.

Interestingly, it appears that there has not been one approach to measure noise. After several feedback conversations with volunteers, it turned out that some volunteers left their phone on the window ledge for quite a while to capture noise levels, while others measured almost ubiquitously and left the application switched on as long as the battery allowed it. Further, others primarily measured on their way to university or to work, thus only using it while in transit. Consequently, even though several user principles were specified at the beginning, there was a broad range of user measuring behaviors, which resulted in different types of noise levels, locations and measuring times during the day.

Further, 27804 individual measurements could not be traced back to a specific user. A possible explanation for this resides within the fact that several people would download the application, and initiate their measurement effort before having created a personalized account. Hence, those measurements were assigned a random user ID in order to take them into account for both the method comparison analysis and the measurement prediction model in following section.

As supplementary insight of the analysis, one can emphasize the fact that amount of NoiseTube data decreased over the prescribed measurement period. While February and March included a plethora of individual measurements, one could perceive a downturn of measurement activity in April and May, which basically leads to awareness about the existence of a certain user churn behavior. In fact, most people, would experience a honeymoon period (Roland, 1990) before decreasing their effort, which is an explanation for the decrease in noise data over the months.

5.3 Outlier analysis and exclusion

In the context of outlier exclusion, one has to mention that the steps below were implemented for the method comparison, but not for the implementation of the user model. To compare both NoiseTube and DCMR data, measurements gathered under unfavorable weather conditions have been excluded. This included rainy days, as well as days trespassing the threshold of a windspeed of 7miles per hour. In fact, in case of a wind protection shortage, NoiseTube can experience temporary systematic errors of up to 10 dB (D'Hondt et al., 2013). These conditions would not enable accurate noise level capturing and therefore had to be removed from the dataset. In R, these steps caused a decrease in data and resulted in the final dataset.

A second step consisted in removing the top five and lowest five percent of measurements. This task of obtaining the data representing the L95-5 [1] noise levels was necessary to implement a representative analysis, as those measurement data would very likely consist of biased noise levels, such as tagged noise-disturbing events or measurements caused by application malfunctioning or exaggeration (Wolfert, 2015).

Further, when addressing the L95-5 the histogram depicting the distribution of the measurement data was giving significantly surprising results, as shown in Figure 13. In fact, almost half of the measurement data resided within the 20-30 decibel range, which, according to DCMR employees, cannot be accurate, as any kind of measurement decibel level would be at least 35 dB and that therefore, the share below 30 dB should be excluded from the dataset. In order to double-check, a box-plot was created in R (Figure 14) which, depicted similar results than the histogram, namely the fact that a major part of the data reside within an interval of 20-30 dB.

[1] This refers to the noise data within a 95-percent interval of existing decibel levels.

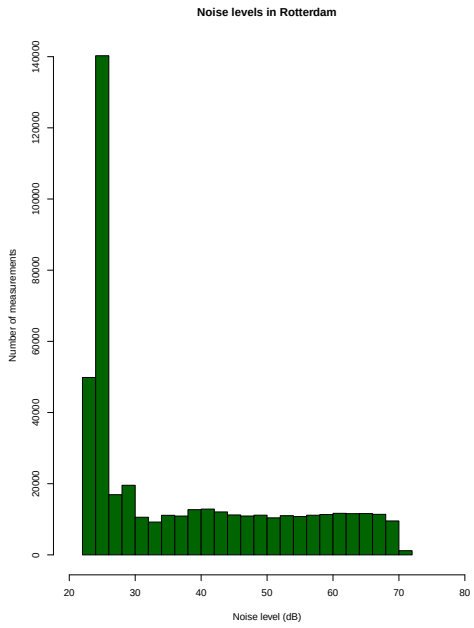

Figure 13: Noise level distribution histogram

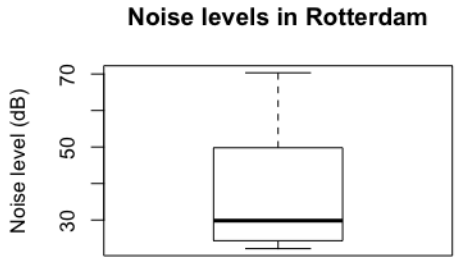

Figure 14: Noise level distribution boxplot

According to Wolfert (2015), a possible explanation for these numerous low-decibel data could be that many phones would include a noise background canceling feature. Phones, whose microphones are almost exclusively made for conversational purposes, i.e. voice sound, would thus consider other sources, such as traffic, as background noise. In order to guarantee a desirable conversation quality, phones would then cancel out or throttle other noise origins.

NoiseTube developers considered the issue of noise cancelling as a problematic feature due to implementation variations and little or no documentation by the manufacturers. However, when designing the application effort was made to reconcile noise levels when calibrating a particular phone with professional sound meters in an echo-free sound chamber. For newer IOS versions, there is the possibility to turn off the noise background cancelling. The comparison between NoiseTube and DCMR data has to be representative and adhere to EU-standards as much as possible and therefore, the decision to exclude measurements below 30 dB for the comparison of noise measurement methods can be deemed valid.

Nevertheless, there is a sound argumentation for the data at comparatively low decibel levels (<30 dB). In fact, one major advantage of the NoiseTube approach is that the method is able to capture indoor noise exposure. In this context, tests with several devices have been undertaken to check whether noise levels indoors would approximately amount to these levels and the results clearly confirmed that indoor noise exposure, due to the shielding from building walls, can be considered considerably lower than outdoor noise levels at street level. Therefore, the many measurements around 25 decibels as depicted in Figure 13 are considered representative for user noise measurement and are consequently taken into account when creating the predictive model for user behavior.

5.4 Graphical visualization

Eventually, it seems logical that there is no need for every single noise data point to represent a clear view of noise levels emanating from NoiseTube data. Hence, a resolution resided in representing aggregated data in grids of 100 by 100 meters (Appendix D) and to visualize those averaged data in a noise map, by plotting average points in varying colors depending on their noise level. A visualization of this NoiseTube map is given in Figure 15.

The inspection of the NoiseTube data map reveals that most grids in Kralingen depict a relatively calm and manageable level of noise pollution. The points in green color represent a noise level below 56 dB, and, as shown in Figure 13, most of these measurements reside in the interval of twenty to thirty dB. A modest share of the grids in Kralingen sketch average noise levels above 68 dB, which can lead to adverse health effects. Further, as any decibel level above 42 dB can cause human annoyance (Wolfert, 2015). Hence, one can conclude that, having canvassed data in Kralingen that has been averaged by grid and hence includes noise levels at various times of the day, most of the neighborhood exhibits considerably tolerable levels of noise pollution. In some of the area's noise grids however, the noise level could potentially be a source of annoyance or even adverse health effects.

In Rotterdam Centrum, the heat map of Figure 15 sketches a different picture: here, most grids exhibit rather high decibel levels, which is logical as these grids are located in an area of intense road traffic and human transit. The higher decibel levels in the center, passing the threshold of 86 dB, can have a quite disadvantageous impact on human wellbeing. However, as the share of residents living in this area is rather minuscule, the impact on health and annoyance of these elevated noise levels is governable.

As it also turned out, the visualization of raw data would not lead to different conclusion than the one with mean noise level data, but significantly inhibit the visual explanatory power of the representation. Hence, instead of implementing a graphical comparison with raw data, depicted in Figure 12, the decision was taken to take advantage of grid-aggregated NoiseTube data (Figure 15).

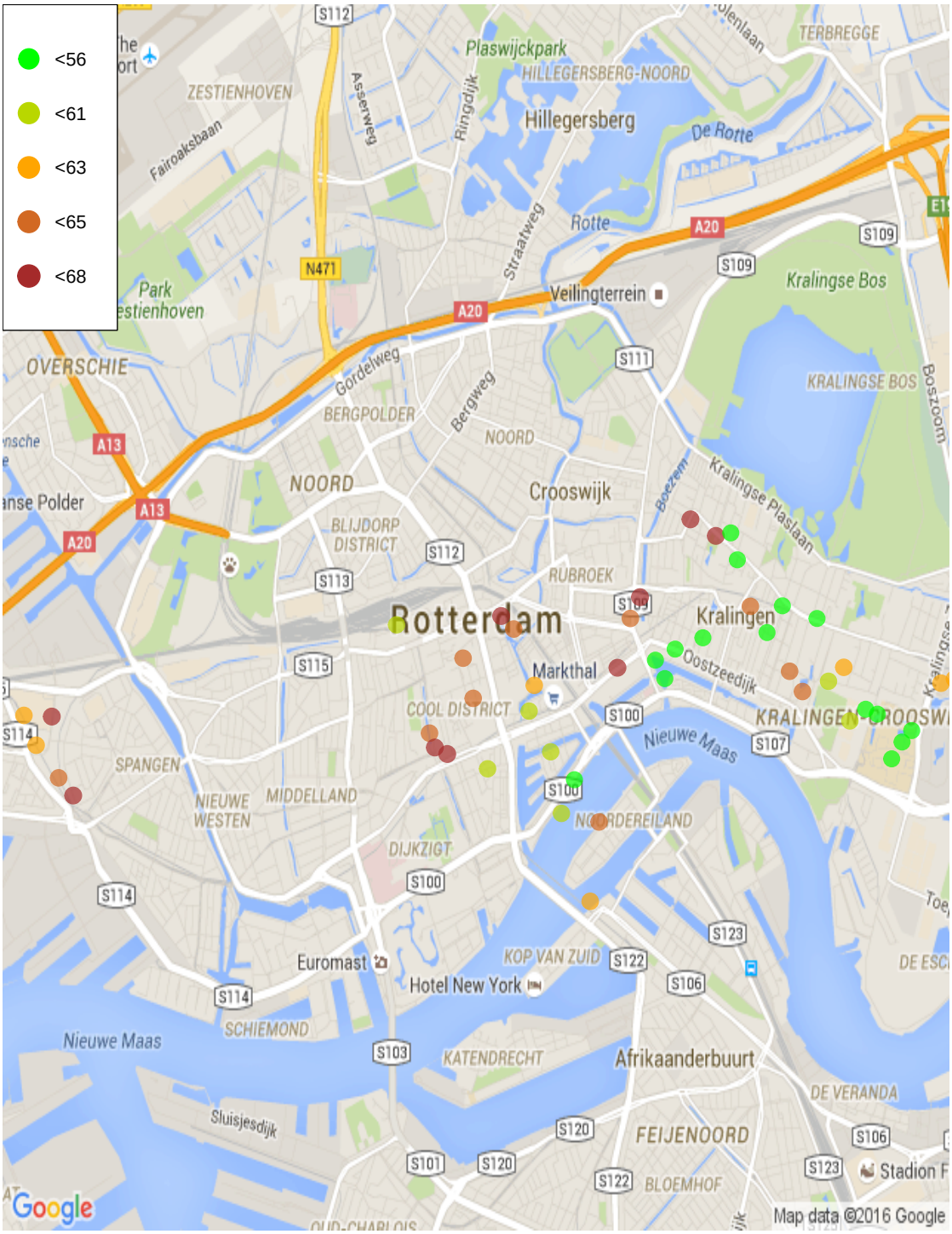

Figure 15: Illustration of a NoiseTube data map of Rotterdam

6 Results

6.1 Method comparison

6.1.1 Analysis with 2007 contour data

The map resulting including NoiseTube data and DCMR contour data from 2007, created in QGIS and visualized in Figure 16, consists of three different layers. A first one concerns the map of Rotterdam, similar to an OSM, and is available through the pdok plugin. Secondly, the DCMR data layer has been integrated, representing the 2007 noise levels coming from road traffic. And thirdly, the visualization includes the aggregated NoiseTube data by grids in the form of colored points, divided into five different noise level intervals.

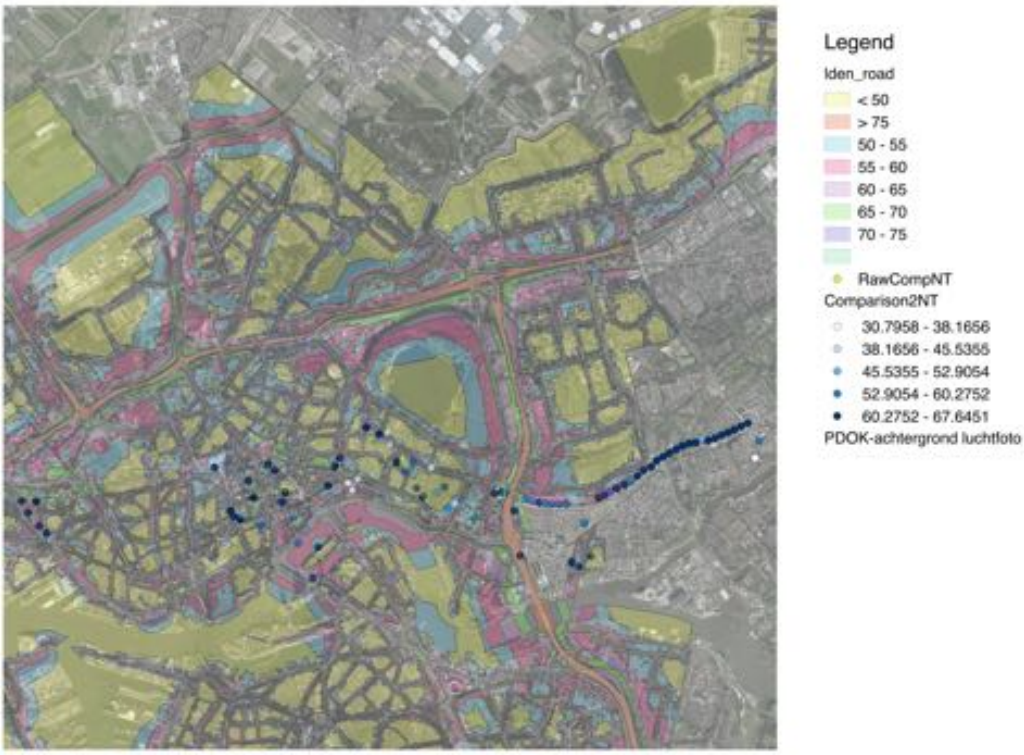

Figure 16: Rotterdam noise map with NoiseTube and contour data

The comparison between the two data types and measurement methods revealed that noise levels coming from NoiseTube data are not inherently higher than the ones established through average-based data modeling. For several categories, NoiseTube data have recorded significantly lower levels than DCMR data, for instance visualized by the blue squares. The latter, amounting to a mean of below or equal to 38 dB, exhibit lower levels than the green areas of the DCMR data, which represent noise levels ranging from 60 to 65 dB. A reason for this could be accidental user behavior, indoor measurement or just time.

As a counterpart, several NoiseTube grids, for instance the dark-green squares, display higher noise levels compared to DCMR data. Dark-green squares depict noise levels of 67 dB and more and can therefore be considered to be superior to most of the noise levels from DCMR layer areas they cover. One reason why NoiseTube data could be higher would be that they overestimate decibel levels due to either footsteps, or due to for instance biking sound while measuring in transit (D'Hondt et al., 2013).

Further, several NoiseTube squares depict identical noise levels than are shown by the DCMR data layer. This holds true for most almost every red-colored square, whose noise levels range around 60 dB, and are therefore very much in the range of both pink and light-green colored DCMR data segments.

Statistical tests were put in place in order to further compare the two methods. For the Spearman correlation, the result was 0.2332 for the dataset including NoiseTube averages and DCMR contours, which shows that the two datasets are not considerably correlated in terms of their respective noise level mean. As the obtained p-value of 0.05377 is larger the significance level of 0.05, the null-hypothesis, stipulating that there is no correlation between the two mean values, cannot be rejected. Thus, it cannot be excluded that both methods would result in different results.

Despite the fact that the two measurement methods express a certain degree of associative power, one can thus stress that there is no statistical guarantee that differences between noise levels measured through NoiseTube and noise levels modeled through traditional means would not be due to the fact of having two different noise mapping methods. By looking at the Spearman correlation, there is a certain statistical likeliness that the two methods could exhibit differences in noise levels.

In order to further compare the two methods, the Wilcoxon signed-rank test was used to assess mean noise level differences. Here, results showed a highly insignificant W value (W(88) = 2415). For a significance level of 0.05 and a given number of 88 observations, this clearly advocates the fact that the null hypothesis cannot be rejected, in other words, that there is a statistical probability that a noise level in a given location is not affected by the used method.

6.1.2 Analysis with 2011 building data

The comparison of NoiseTube data against 2011 DCMR building data from traffic noise, shown in Figure 17 includes several important messages:

Firstly, the noise levels from this type of DCMR data are captured both in front and behind a building. Logically, noise levels, depending on the location of their source, would exhibit different intensities behind a building than in front of it, which can also be observed with the difference in colors around one building in Figure 10. The NoiseTube data points however, depict an average noise level of a location grid of 100 meters meaning that they do not take into account the fact that one could be located in front or behind a building. In fact, NoiseTube data are recorded where people move or are, so their dB level depends on the location, but not whether they are situated exactly by a noise source or whether there is a building in between.

As a result, one has to acknowledge the fact that a single NoiseTube data point is less comparable to a point depicting a noise level at a building point than to average noise levels residing at

contours of locational grids. Therefore, the approach of comparing aggregated NoiseTube data to modeled data at contour level in the previous section is more valuable than comparing the former to modeled data at building level.

Secondly, it is important to notice that, given a plethora of building data in the city of Rotterdam, the decision was taken to create 100m buffers for the DCMR building data. The idea thereby was to only take into account the building data that would prevail at the locations were NoiseTube grids were available. This would enhance visibility but also allow a more straightforward performance of the QGIS program. Hence, the noise map created in Figure 17 only includes the building data that pertain to locations that also contain aggregated NoiseTube data.

Thirdly, similarly to the noise map with contour data, the decision was taken not to use raw NoiseTube data. This would enhance readability for one, but also improve visibility in order to permit an accurate graphical method comparison.

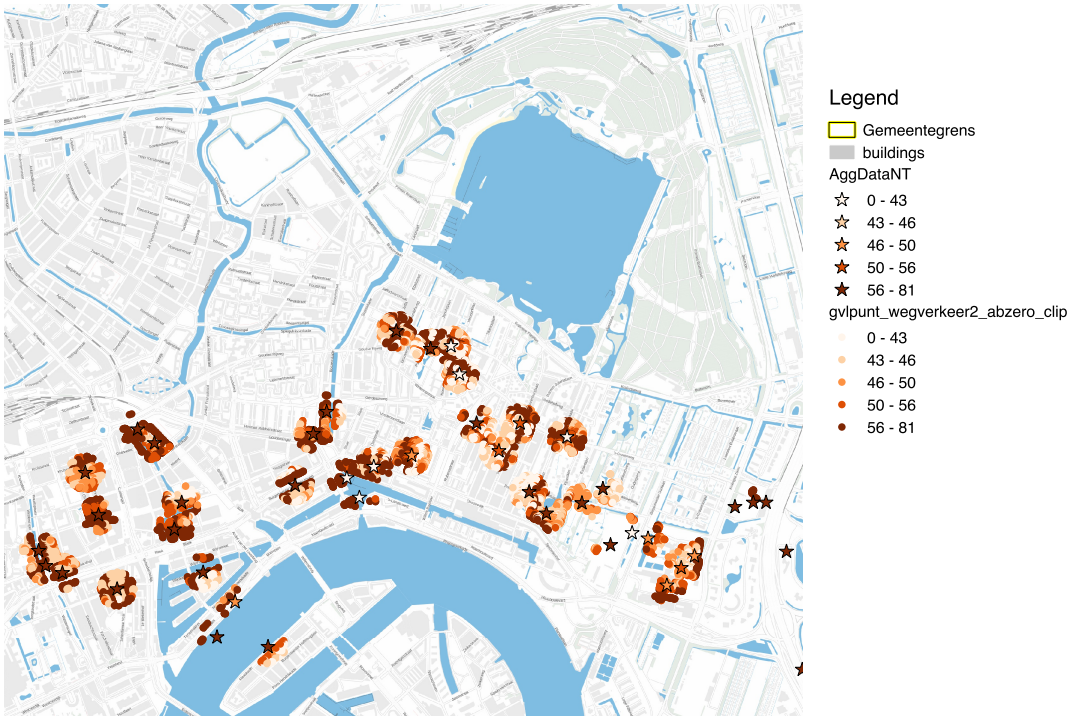

Figure 17: Rotterdam noise map with NoiseTube and building data

When visualizing parts of the noise map in Figure 17, one can observe that, for several spots, the aggregated NoiseTube data depict lower noise levels than the building level points surrounding them. This is for instance the case for both buffers on s'Gravensweg, whereas NoiseTube data exhibit higher noise levels than their circumventing blocks above and on Oostzeedijk. One reason could be that NoiseTube data would overestimate noise levels due to either disturbances by volunteers or the height they are collected in, which is 1 to 1.5 meters, whereas DCMR data would be modeled at a height of 4 meters (D'Hondt et al., 2013). For those groups of points,

the maps thus emphasize the fact that NoiseTube data depict a different picture than modeled data, which is also shown around Oostplein, where the noise level coming from NoiseTube is considerably lower than its DCMR counterpart. As such, one can also say that traffic noise is somewhat overestimated for those locations.

When the location of noise data is evolving towards Rotterdam city centre, the message is different: Most of the locations of NoiseTube data present the same noise level interval as their DCMR building counterpart. Examples of those occurrences include the data around Rotterdam Blaak, but also next to Coolsingel or Westsingel.

Similarly to the contour data, statistical tests were conducted to measure statistical dependence between the two noise level means. For the Spearman correlation, the result was 0.064, highlighting absent correlation between modeled and measured noise data. As a result, the differences are likely to be due to the existence of two different noise measurement methods. The p-value amounts to 0.65 and is therefore largely exceeding the 0.05 significance level, underlining the statistical incapability to reject the missing association between the two methods, in other words, to exclude the fact that differences in noise levels could be due to something else than the measurement method.

Moreover, a Wilcoxon signed-rank test was performed in the case of building data as well. Here, results confirmed the outcome from contour data, namely that noise levels in a given location remained unaffected by the choice of measurement method. In this case, the W(88) value amounted to 262 and therefore, with a p-value of 0.000174, is statistically significant can reject the null hypothesis that there this is no difference in average noise levels. In other words, the hypothesis that the two methods produce identical results is eliminated with a certainty of 95 per cent, which shows that there is no guarantee that both methods produce the same results.

6.2 Measurement appraisal: User model

A question raised with the information obtained through the method comparison is whether users are only measuring when they are annoyed by noise, or whether they measure noise while they are at home or traveling between two locations in the city. Hence, the mobile application leaves potential for further insight into user behavior, giving further validation to establish a model that would predict measurement behavior.

As mentioned above, only using NoiseTube to measure noise would hardly engender the identification of various users patterns. In order to fathom user patterns, one can initiate the process by analyzing their measurement behavior. In this context, the the first step to observe potential patterns was to implement a nearest-neighbor analysis (NNA) in QGIS. In the case of aggregated noise data and DCMR data of 2007, the analysis yielded a Z-score of -7.036, which clearly confirms that the null-hypothesis is to be rejected, reinforcing the fact that user measurement is not taking place as a random behavior, but rather at specific occasions and locations.

In order to check for consistency, the NNA was also implemented with the 2011 building data. Similarly to the 2007 contour data, one can observe a highly negative Z-score of -5.5356, emphasizing the conclusion that users do not move around randomly in order to measure noise.

One explanation for this result could be that many users are taking noise measurements on their way to work or university, represented in the itinerary-like distribution of points in Figure 16.

It seems clear that people would mostly measure noise while they are moving between two locations, rather than walking into a specific direction just to measure. Hence, the rejection of the null-hypothesis can be considered meaningful and advocates the fact that people measure "while on the go".

6.2.1 Model creation

As the experiment in Rotterdam did not include sufficient numbers of volunteers to successfully build a predictive model, user data coming from other cities had to be included. The approach here was to include user data from bigger cities such as New York, Rio de Janeiro, Sydney, Berlin or Brussels in order to receive sufficient numbers. As the variable as to whether users measured in a given month referred to the four months the NoiseTube experiment took place in Rotterdam, user data from other cities also had to come from the months January, February, March and April 2016. The customized map feature on the NoiseTube website permits the extraction of required datasets, enabling a sufficiently large user base to build a predictive model.

The cities mentioned above did not include sufficient numbers of users in the given measurement period either to create a credible model. As a response, the idea was to systemically perform searches regarding the measurement activity in various cities on the NoiseTube website. In fact, the website enables a person to check the latest measurement activity worldwide, as well as the total amount of measurements per city. Hence, once cities with sufficient measurement activity were identified, one had to create customized maps over several years, possible only for a 12-month period, and then check whether enough activity had taken place in this very time period. If this was the case, the dataset was adequate to be imported into R, prepared for and included into the predictive model (Appendix F).

A final approach in order to obtain sufficient users was to randomly allocate user IDs to the "NA" measurements in the dataset, and this for each city. As it was known that each individual measurement represented a two-second window, the idea was to allocate NA's into two or three-minute intervals, for instance in Rio, every two-minute measurement period would be given one user id and so forth. This step was consistent with the previous analysis, as it would appear that many volunteers would only measure for a very short time period. The aforementioned steps resulted in dataset with 2097 observations, which is not extremely large, but does allow for a certain degree of predictive power.

The NoiseTube data eventually obtained then had to be altered in R in order to be categorized by user, leading to the potential recognition of patterns. In fact, all the data received logically included only information about users that did actually measure noise, not about the ones that did not. Consequently, it was not possible to derive information about whether people would be likely to become noise sensing volunteers and thus the approach was to build a model determining whether people that had already taken noise measurements were likely to measure in the future as well.

Most importantly, to build a credible predictive model, it had to include a certain time function, meaning that data up to a certain point in time would be used to create the predictors, and data after this break point would be used to conceptualize the target variable. This was crucial, as otherwise the model would have exhibited the same data, categorized by user, to establish the target variable that were used to create the predictors. In other words, the months after the break point that were used to create the target variable had to be excluded from the dataset that

was used to create predictors such as the sum of locations or the mean decibel level per user.

Subsequently, the goal was to reduce the NoiseTube data, consisting of a multitude of rows per user, to one row per user. The model's predictor variables concern the timespan between first and last measurement, loudness mean and standard deviation in decibels,11 binary variables as to whether people have measured in a given month, as well as the sum of location grids a user has measured in. In total, for predictor and target variables, 14 months were taken into account: January-May 2016, January-April 2012, as well as August 2014, September 2014, June 2015, November 2015, and finally October 2011. The predictor variables are listed below in Table 5, together with the corresponding intuition.

The target variable, specified as "MeasuredFut", is a binary variable based on whether a user has measured in two months after the break point and returns yes if the given person is forecast to continue measuring in the future and no otherwise. The creation of the target and predictor variables is detailed in Appendix F.

Table 5: Predictor variables and intuitions

Timespan between first and last measurement	The intuition behind this is that in order to be likely to continue using a noise sensing app, one needs to have measured more than one day, but also would need to take on measuring after a certain "honeymoon period". Hence, someone that would have measured on sheer numbers of days that are spread out would be rather likely to continue his noise sensing activity.
Loudness standard deviation	In case of absent noise level standard deviation, a user is very likely to measure at the same time and location. However, if this occurs, that would mean he or she is not very likely to pursue a measurement activity, as this would include noise sensing at multiple locations and times of the day.
Months a user measured in	A user that only solely measures in a single month is not likely to extend his or her measurement efforts, as his measurement might be due to a single trial, or even a paid one-off effort to do so. However, in case the user shows repeated sensing efforts in multiple months, he or she is probable to pursue collaborative noise sensing.
Loudness mean	A user that would measure constantly elevated noise levels could lead to the assumption that he or she would be very likely to use the app exclusively in case of noise-related annoyance, which does not lead to a very representative measurement activity, nor to a conclusion that this user is very probable to continue measuring in the future.
Sum of location grids a user measured	A user that would only measure at one specific location grid - most probably around his domicile - could not be considered likely to continue measuring, as his effort remains at a rather small-scale level as he does not measure in multiple spaces or suburbs, but only in one specific grid.

In order to check the significance of the predictors, a logistical regression model was created, shown an Appendix H, and it turned out that four out of five predictors were significant on a 1percent level and one predictor on a 5 percent level. Hence, this constitutes a verification that the model's predictor are able to implement a qualitative forecast.

After including both predictors and target variable into the data frame, the decision was taken to firstly initiate the model throughput with the "rpart" package in R. In fact, this package enables

the creation of a decision tree and hence followed the rationale of using a decision tree as a modeling technique mentioned in Section 4.4.1.

6.2.2 Model interpretation

Decision trees pertain in a situation of uncertainty, which means that a 100 per cent correct decision does not exist (Neville, 1999). In this case, the degree of uncertainty is explained through responses, meaning that either a person is predicted to continue measuring (right bottom number of the node) or predicted to stop (left bottom number).

Once the model was plotted in R, graphical results were pictured and analyzed. In order to initiate the interpretation, the function "asRules" was applied in R (Appendix K). The nodes and branches of the decision tree, shown in Figure 18 with complexity eleven, give rise to the following results and will be discussed in Section 7.2:

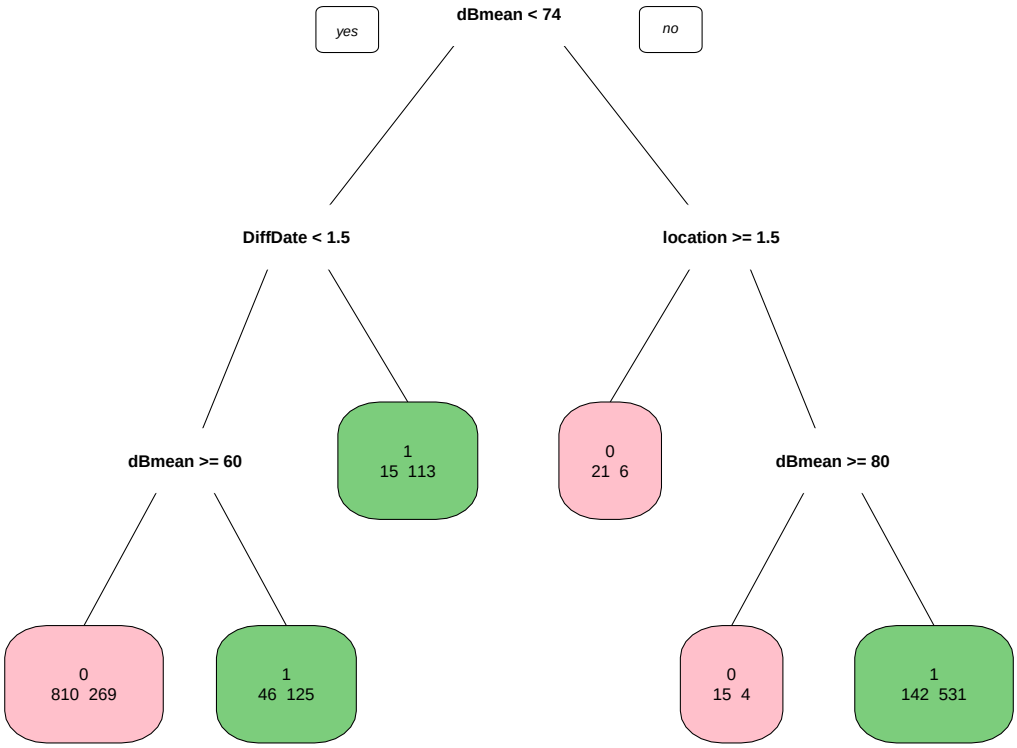

Figure 18: Decision Tree for the user measurement behavior model

For a given user, the model does predict that in case his measurement mean is below 74 dB, he or she will continue the measurement effort provided the difference between the first and last measurement is not below 1.5 days. In case it is below 1.5 days, the user will continue noise sensing if the average measurement noise level is below 60 dB, which covers rules 5, 15 and 9 of Appendix K.

Further, in case the user exhibits a mean decibel level higher than 74 dB, he will only continue measuring in case he has measured at more than 1.5, or two different location grids. If that is the case, than he or she will only persist if the average measurement noise level does not exceed 80 dB, covering rules 8, 6 and 14 of Appendix K .

The interpretation of the tree portrays a clear and logical user pattern, which is the fact that users are predicted to continue their measurement effort highly dependent on whether they are annoyed by noise or not. Even the case where are not annoyed by rather elevated noise levels (74 dB) does not predict unlimited noise measurement, as those users would also end their measurement activity at a certain tolerance level, which is shown on the left part of the tree in Figure 18. In case users are not annoyed by noise, then they also have to have measured in multiple locations in order to be predicted to remain noise measurement volunteers.

For the users that are annoyed by rather high noise levels, they would only be predicted to continue their measurement activity in low-noise areas, shown by the >=60 dB on the right bottom node in Figure 18. Further, they also have to have a decent timespan between their first and last measurement in order to be predicted to maintain collaborative noise monitoring.

Consequently, the kind of users that is predicted to maintain their measurement activity are thus primarily those who are not (highly) annoyed by noise, who have measured in multiple locations as well as with a within a rather large timespan. Those users are enjoying the NoiseTube approach as a "social experience" (Heipke, 2010, p.552), and most likely consider their achievements to contribute to a greater cause, positively impacting the quality of living in their society.

6.2.3 Model evaluation

At the evaluation level, one is usually concerned with determining what one would actually like to achieve with predictive modeling (Koppius and Belo, 2015), which in this case would be to gain insight into whether and how people would be forecast to measure in case a collaborative noise measurement application would be launched in Rotterdam.

The first important aspect concerns the fact that this model constitutes a binary classification tree. A classifier would divide instances into the ones that are predicted to continue (1) and the ones that are not (0), which is also why this model engenders a bimodal class distribution (Figure 19. One can see that the data are quite balanced in terms of their prediction, which is shown by the almost equivalent class surface area. Depending on the threshold one opts for, the model is able to classify the instances quite accurately.

As such, given the fact that the dataset was quite balanced, the first step to evaluate the model was accuracy (Appendix I). In this case, the result was 0.7842, which is acceptable, given the predictors' characteristics. One can say that, with the given predictors, the model forecasting power is deemed reasonable. However, in case of bifurcate decisions, given the budgetary constraints of a municipality, one has to ensure classifier reliability because of the skewed cost distribution (Marrocco et al., 2008). Misclassification errors can lead to the missed opportunity of increases in noise measurement volunteers, impacting the detail of noise level representation and therefore also the collaborative method's profit. In order to further evaluate the model's performance, the approach was thus to take advantage of AUC and the ROC curve, where the gist was to to compare the performance of the decision tree against the one of the logistic model,

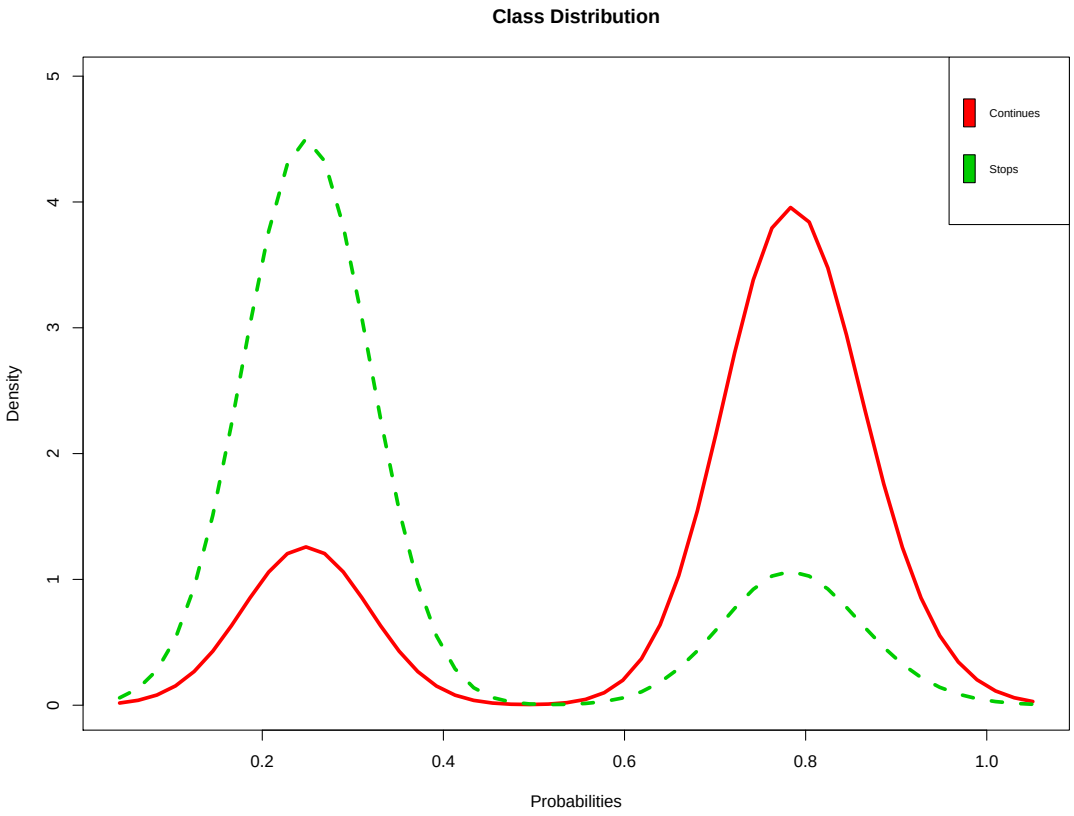

Figure 19: User model class distribution

as shown in Figure 20.

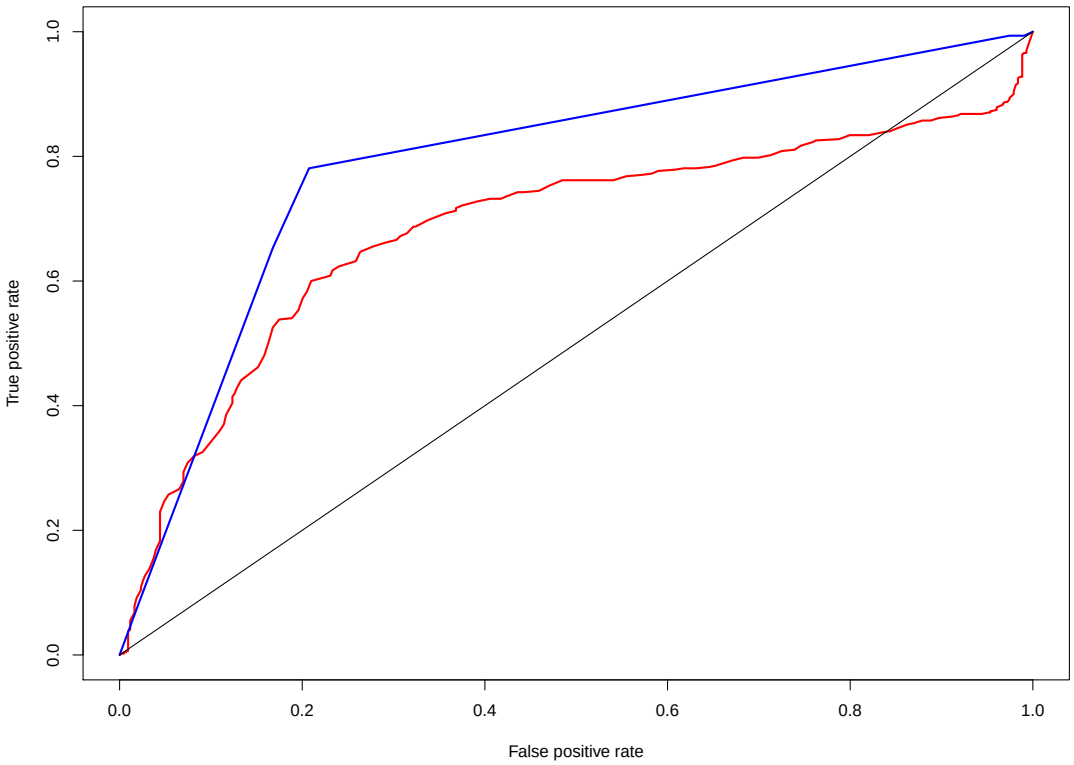

Figure 20: Decision tree and logistical regression: ROC curves

As one disposes of a binary classification tree in this case, the discrete classifier would have class labels (1 and 0) as outputs, and thereby produce one pair of false positive rate (FPR) and true positive rate (TPR), which would be represented by a single point in Figure 20 (Fawcett, 2006). However, the idea, as the follow-up decision as to whether or not implement a collaborative noise measurement tool would be how to incentivize people, one needs to establish a threshold above which people that would be predicted to continue measuring should be targeted. As such, people would then be targeted if the probability that they would maintain contributors was exceeding this threshold. This required an entire ROC curve, which then resulted in the creation of probability-based classifiers. In this context, every point in Figure 20 represents a confusion matrix with certain probability threshold as to whether a user is predicted to continue measuring or not.

For the given tree model, the AUC result is 0.784573, which represents the likelihood that the tree will rank a randomly opted positive instance higher than a negative instance (Koppius and Belo, 2015), which is not an unacceptable result. The AUC score for the logistic regression model amounts to 0.6851, which clearly underlines the superiority for the decision tree in this case. As the ROC depicts tradeoffs between TP and FP (Koppius and Belo, 2015), the tree model's predictive power can be deemed acceptable, as it does sufficiently exceed the power of

random guessing, which is 0.5. Similarly, the logistic regression also exceeds the power of random guessing until it reaches the specificity of 0.778. Consequently, one can conclude that the predictive power of the tree exceeds the one of the logistic regression model, which confirms the decision taken in Section 4.4.

However, with an AUC of 0.78 respectively 0.68, clearly there is room for amelioration with both model's respective predictive power. However, this is not surprising, as collaborative noise measurement behavior incorporates various - sometimes endogenous - complex relationships that go beyond the simplistic conceptualization this model takes advantage of when defining participatory noise sensing.

With the given model creation, the decision seems to perform relatively well, which means that it can be attributed predictive power when forecasting whether a person will be likely to maintain his or her collaborative noise measurement activity. The sometimes inevitable challenge with decision trees consists in overfitting. A model should be generalizable, in other words, it should be applicable to different settings and datasets (Koppius and Belo, 2015).

In the case of NoiseTube, the objective is to ensure the model also applies to other cities, but also to other crowdsourcing measurement tools, such as rain or C02 emission measurement tools. One can conclude that the model can be applied to other cities in a relatively straightforward manner, especially if the use NoiseTube data, but also otherwise, as the predictors include coordinates, days and months of measurement. These can be easily applied to other crowdsourcing tools, and the ones that are noise-specific, such as the average decibel level per user, should be able to be substituted by measures that capture other phenomena. As such, one has been able to somewhat mitigate the danger of model overfitting.

7 Discussion

7.1 Method comparison

Although one of the main challenges mentioned in Section 3.5.3 can indeed be deemed true, namely that the NoiseTube approach lacked sufficient volunteers and was thus not able to cover the majority of Rotterdam, it supplied interesting insights into noise dynamics in the city.

A first welcoming conclusion for the DCMR and urban planners concerns the fact that noise map comparison reveals that noise levels found through the crowdsourced approach are not excessively diverging from the ones established through propagation models, at least for the locations for which NoiseTube data exist. As a result, the visual comparison of different noise measurement methods reveals that both methods are returning mostly identical noise levels, confirming the results from D'Hondt et al. (2013).

A basic challenge for such a method comparison always resides in the decision for a comparison dataset. Consistent with Haklay (2010), one has to acknowledge that the comparison dataset, in this case DCMR data, is qualitatively more important and the closest to reality, making it possible to emphasize potential shortcoming of the NoiseTube method. Given the presumptions of Haklay (2010), the conclusion is that taking advantage of contour data was preferable compared to using building data, especially because NoiseTube data were averaged by grid. Building data exhibit different noise levels depending on the side of the noise source, and therefore the results from comparison to grid-averaged NoiseTube data are concluded to be less strong than to data that are averaged at zone contours.

Together with the decision for a comparison dataset, one also has take into account that DCMR data are based exclusively on road traffic noise and represented in Lden, which is a weighted average and a step away from the Leq indicator that NoiseTube decibel levels are computed in (D'Hondt et al., 2013). Whereas DCMR maps use daily averages at 4 meters above the ground and propagate traffic noise levels from larger streets into neighboring roads, NoiseTube maps on any kind of street noise measurement, taken at a height of 1 to 1.5 meters (D'Hondt et al., 2013). This indicator divergence is somewhat mitigated by the fact that most NoiseTube data have been collected during daytime, but also because most of NoiseTube's data would have traffic as source, making the use Lden with road traffic adequate (D'Hondt et al., 2013) and thereby providing validity for the method comparison.

Consistent with the mentioned benefits of participatory sensing in Section 3.5.2, NoiseTube is able to provide precise and data-driven acumen for locations that traditional measurement methods only supply simulation-based average noise levels. Thereby, the method contributes to accurate noise level determination, enhancing the general measurement process. The method comparison for instance ceritifies certain results from D'Hondt et al. (2013), as roads or intersections seem to be quite high-noise locations. The participatory method thereby offers insight into noise distribution dynamics such as peaks or temporary elevations from very specific locations, giving further information about user behavior and confirming two of the strengths stipulated in Section 3.6.

By contrast, and thereby somewhat opposing one of the strengths of participatory sensing advanced by Perera et al. (2014) in Table 1, the NoiseTube approach does not enable for real-time

noise monitoring yet, let alone real-time policy making. The measured data do take up a certain time period to be uploaded to the server, ranging from less than one hour to sometimes several hours. As a result, the idea of real-time noise measurement, at least with the given NoiseTube functionalities and the prevailing number of participants, has to be postponed.

According to Perera et al. (2014) in Section 3.5.3, a participatory sensing experiment needs to ensure user privacy. Therefore, a somewhat worrying result, although it has to be nuanced about its impact, is that one can actually access private user information. In fact, users have to provide a username for their registration, and there is several people that have chosen to provide their names as user names. Although the data are given a numerical user ID, one is able to trace this ID back to the user name, which, for the people that have used their actual name, allows to select all the measurements with a given ID. Given the pattern identification, one can then easily observe where this specific person lives, which constitutes a certain privacy breach. Therefore, it should made clear to volunteers that they should differentiate their user name from their real name in order to avoid this unethical intrusion.

A potential shortcoming of the NoiseTube method is the faction that malicious users cannot be pinpointed, which does mostly confirm one of the existing challenges introduced in Section 3.5.3. Whether a NoiseTube user really measures noise as it appears or whether he would just deliberately turn up his or her music to engender flawed measurements cannot be proved. Consistent to Mousa et al. (2015), the approach, similar to crowdsourcing applications in other settings, has to be to rely on trust here. In order to somewhat mitigate the issue, one would be able to exclude those flawed measurements through the L95-5 outlier exclusion implemented in Section 5, but also would those awry data average out with a sufficiently large dataset.

With reference to the pattern recognition advantage in Section 3.5.2, one is able to identify a certain user pattern that the latter seem to increasingly prefer to measure noise either at home or while in transit from A to B, which in spite of a certain vagueness, is preferable over traditional average and simulation-based noise monitoring. Thus, in the case of a large-scale launch, planners need to take into account that people do not move into a direction in order to measure noise, but rather measure noise when moving into a direction. That being said, planners need to ensure that there is enough people that transit in different locations in order to amass sufficient locational data.

Drawing back to Section 3.6, one can conclude that several other aspects need to be mentioned: Firstly, the fact that NoiseTube allows indoor noise measurement does not add value to the method comparison, simply because indoor data are inappropriate to assess against outdoor data, which is why they were not included into the method comparison (most of the indoor measurements exhibit lower decibel levels than their outdoor counterparts).

Another aspect of Section 3.6 concerns the fact that NoiseTube enables moving sensors. In fact, this factor can be deemed quite useful, as it enables the specific identification of noise levels on certain itineraries, for instance a person's transit from home to work. Hence, this allows to gain insight into noise levels on several popular itineraries, such as the ones shown in Figure 16.

The data acquisition cost with NoiseTube, as mentioned in Figure 3, is indeed very manageable, as one only needs to create a NoiseTube account in order to be able to download them from the website, which, compared to traditional data gathering costs, is highly valuable. Further, the fact that NoiseTube enables continuous monitoring throughout the day is a strong factor,

however the prevailing requirement for this strength resides within the fact that one needs to involve sufficient volunteers, which has not been the case.

The results coming from statistical testing produce mixed results. For aggregated NoiseTube data, the Spearman test cannot confirm the fact that there is no significant difference between noise levels from NoiseTube data and modeled data for both building and contour. The Wilcoxon test is able to reject his null-hypothesis in the case of building data but not for contours, meaning that for contour data, the measurement method does not impact the measured noise levels, but for building data, cannot exclude the possibility that differences in noise levels for a certain location could be due to the sensing method.

Generally speaking, there is thus no complete statistical insurance that the two measurement methods would lead to the same result, which complicates the assessment of the method comparison. However, this does not mean that NoiseTube as method cannot be deemed a valid tool for noise measurement in cities, simply because of the fact that the comparative data are modeled, which, as already mentioned in Section 1, does not prove their precision. However, as this is the method that is used in most EU cities, it has to be acknowledged that the noise levels coming from this data type are the reference point for statistical analysis.

To conclude, one can therefore stress that the NoiseTube method cannot be exclusively used for noise measurement, but has to be aligned with an available and reasonable long-term averaged model in the area in question, an argument that is also underlined by De Coensel and Botteldooren (2014). Further, a large-scale implementation has to involve sufficient users in order to be fruitful, which leaves the question as to whether and how people would really implement their measuring activity. This gives rise to the user model debate in the next section, discussing the implications about the decision tree's results in Figure 18.

7.2 User measurement model

The results form Section 6.2.2, emphasizing several interpretations from the user model, further assist the pattern recognition statement of Section 3.5.2. Despite the model's evaluation, one can conclude that most people would only measure if they were not annoyed by noise pollution.

The user model is thus able to complement the results from the method comparison, namely that people would not be significantly likely to take the role of noise measurement volunteer, unless they were given a financial incentive or obliged to do so. Hence, the model can be considered a novice step in noise measurement, confirming the innovation aspect in Figure 3.

Given the results in Section 6.2.2, on can stress that noise sensitivity also exerts an influence on individual perception of prevailing decibel levels (Shepherd et al., 2010). In this context, noise-sensitive individuals are more probable to pay attention to sound and to appraise it negatively. Moreover, noise-perceptive citizens are likely to exhibit reinforced emotional reactions to noise, and would consequently rate noise levels largely higher than people insensitive to sound (Shepherd et al., 2010). As a result, noise-sensitive people would be annoyed by more rapidly and hence remain less likely to continue their mapping process, a finding that has also been confirmed by the results of the user model in Figure 18.

Consistent with Roland (1990), people would sense noise during a couple of days or even weeks before promptly ending their honeymoon endeavors. They would measure a couple of times,

and then realize that this would not turn out to be an enjoyable activity, causing them to quit measuring. Therefore, the ones that are not highly annoyed (explained by the no for <74 dB and the subsequent no for >=80 in Figure 18) by noise after their honeymoon period and that have measured in multiple locations can thus be considered keen to continue noise mapping.

As a result, the model advocates the same conclusion as the method comparison: for one, a limited number of participants is not sufficient to obtain a clear prediction as to whether a citywide deployment of the NoiseTube approach would be an endeavor with a highly uncertain outcome or result in a success. The model also predicts that the majority of users would be highly uncertain to remain noise sensing volunteers in the medium- or long-term. Therefore, planners need to think about a way to either give people an incentive to participate, to continue, or to oblige certain parties to do so.

For another, one can stress that the amount of users that is likely to continue measuring noise after a honeymoon period is limited. In fact, the graphical method comparison in Section 3.6 has verified the existence of several high noise areas, especially in the city center (Figure 17). Therefore, as those areas exhibit high noise levels and given the fact that most people do not measure if they are annoyed, people living and traveling to those areas would be rather unlikely to pursue a measurement effort. This also advocates the need for a large amount of users to bridge the gap made by people that would not measure due to their noise-related annoyance.

Another important message to discuss here is the fact that this model bases its predictions on people that have already measured. Given the fact that annoyance plays such a critical role, one can not only conclude that people who had already measured would quit collaborative noise monitoring if they were annoyed, but also that a significant share of people would not even bother to begin, which has to be taken into account when promoting the approach or when deciding on how to incentivize residents.

In this vein, one has to establish an incentive model for people that have not started measuring, but also for the ones that are predicted to continue their activity. Hence, one needs to decide to establish a target threshold above which volunteers would be targeted, or offered an incentive to continue. The default threshold used to predict the binary classification outcomes is 0.5. However, one has to really take into account the goal of this measurement activity. In order to make collaborative noise measurement a success, one needs as many users as possible, As such, one needs to target a maximum number of people, which means the application's promotion and its demonstrations clearly have to reach a number of people that is close to the entire residential population to cover sufficient areas, potentially representing a threshold less than 0.5.

Further, to account for the fact that most people would not last as volunteers nor even begin, one needs to give them an incentive, which means that one needs to give a maximum of people an incentive. In this context, it seems clear that the threshold above which to target people with an incentive should be somewhat low, as one would then incentivize a maximum of TP instances, so people that would measure, which would result in more people that would actually measure, which is clearly the goal. When taking a look at the ROC curve in Figure 20, one should therefore choose a threshold that would represent a very high TP rate, and with it along a somewhat, depending on the cost of the incentives and the budget, manageable FP rate. As such, this model's classifier would then be deemed liberal (Fawcett, 2006), which is the way to go here, given the fact that the approach requires a significant amount of contributors.

In terms of costs and profit estimations, it seems pretty hard to estimate a given number value from targeting the right people. However, the profit should be considered in terms of volunteer numbers, so the more volunteers there would be, the more detailed collaborative noise measurement, and thus the higher the total profit from targeting people.

7.3 Contribution to literature

7.3.1 Impact on sources and effects of noise pollution

With the prevailing functionalities of the NoiseTube application, i.e. the tagging function, a clear investigation and description of noise sources is tremendously complex. In fact, given the three possibilities to address noise pollution, the most promising one is to address noise pollution at its source (Ouis, 2001). The process of tackling noise at its origin would clearly lead to a reduction of annoyance and thus ameliorate quality of living. However, the NoiseTube application does not enable a clear source recognition as described in Section 3.2. Based on NoiseTube functionalities, one simply does not know whether the source of the recorded noise level is road traffic, industry or other noise sources, which makes it problematic to undertake immediate action on noise reduction.

The only way to obtain insight into noise sources through NoiseTube is personal experience while measuring, or discussions with other volunteers. Those factors, similarly to as if one had conducted noise surveys, led to the adoption that road traffic can indeed be considered as the main source of noise pollution for the given locations and period, confirms the statements of Salomons (2013) in Section 3.2. In Kralingen, it is especially the emergence of delivery motorcycles that is causing short-term noise level peaks. Further, personal measurement activity also showed that the driving method was at the origin of several high-noise spikes, confirming Ouis (2001) in Section 3.2. Other than those experiences, the NoiseTube method revealed lacking insight into other noise origins such as aircraft or industry noise.

A suggestion to improve the functionalities of NoiseTube so that it would enable detailed source identification is to include a feature that applies automatic source pattern recognition, similarly to what the "Shazam" application does with music. While measuring, NoiseTube would then identify source patterns and thereby recognize what exact origin would belong to. This would also eliminate the inconvenient need to manually add noise source information that even alters the measurement as one touches the phone's surface.

Even though the experiment did not reveal as a superior method compared to traditional means of noise sensing, given its limited reach and the challenges to get volunteers on board, it also has the potential to have a positive impact on health. Noise-related annoyance is often measured in lost DALYs and in the Rijnmond area, and $2,630$ life years are lost each year due to road traffic, aircraft, rail and industry noise, representing a loss of around EUR 200 million, or EUR $78,500$ per year of life lost Theakston (2011). Both the DALYs as well as monetary repercussions have a tremendous impact on quality of living for both urban planners and residents.

The NoiseTube approach, if implemented on a large scale, could have the potential to mitigate these numbers due to the increased availability of detailed noise data. If a collaborative noise sensing method was able to provide location, source and time features along the decibel level, as NoiseTube could offer, this would depict a much more specific picture about the actual occurrence of noise pollution. As such, the countermeasures to noise pollution in Rotterdam would

be very much data-driven and not simulation-based.

Given the general complexity in assessing noise-related health effects, which was already established in Section 3.3, it is also rather problematic to draw other conclusions on health from the NoiseTube approach. Health is multidimensional and embodies not only illnesses and infirmity, but also wellbeing (Shepherd et al., 2010) and the adverse effects mentioned in Section 3.3 are not directly traceable back to NoiseTube.

Similarly to health effects, and as already mentioned in Section 3.3, economic effects of noise pollution cannot be traced back to NoiseTube and remain outside the scope of this paper.

Nevertheless, the similarity between NoiseTube data and modeled data allows to consider noise levels coming from the collaborative approach as acceptable, confirming that several areas in the city could be at the origin of noise-related annoyance. Hence, despite the complexity to draw conclusions on specific health effects, the method has confirmed the existence of areas that can potentially provoke severe annoyance or sleep disturbances, mediating conditions, which can decrease productivity and wellbeing, and thus also quality of life (Fyhri and Klæboe, 2009).

7.3.2 Design Science

DS focuses on either building or evaluating artifacts that serve a human purpose (March and Smith, 1995) and incorporates 4 different concepts: models, constructs, methods, and material products (von Alan et al., 2004). By judging the feasibility of crowdsourced noise measurement, this paper focuses on the evaluation of an artifact. Further, the paper covers all four DS concepts: A model to predict user behavior (1), the collaborative noise sensing method (2), the NoiseTube application as a material product (3) and the theoretical constructs of noise pollution such as its effects, sources, but also further explanations of the traditional method (4).

DS undertakings, stressing the goal to investigate how things should be to achieve objectives (von Alan et al., 2004), clearly have to include a technological aspect nowadays in order to obtain sufficient validation. The NoiseTube experiment, although relatively limited in geographical reach, has been able to stimulate both awareness and comprehension for the relevant stakeholders, ranging from DCMR employees to the municipality's urban innovators.

A DS undertaking, in order to be granted validity, has to follow a certain research methodology (DSRM): Having implemented the NoiseTube experiment, one can conclude that the demonstration and communication activities stressed by the DSRM have to be taken very seriously. In order to really guarantee a large-scale artifact implementation, the communication and demonstration activities have to go beyond a mere university promotion or a simple message in a newsletter. The promotion has to be conducted at multiple occasions and personally in order to to convince sufficient residents to participate.

One-time and mono-channel communication is not enough either. Communication has to occur repeatedly via multiple channels, especially personalized face-to-face and social media conversations. Looking at the research outcome, one can conclude that the pre-experiment communication could have been insufficient and focused on a too narrow environment.

With the given results in Section 7.2, one needs to take into account that for activity four of the DSRM, the process of promotion and demonstration, one also needs to take the people into

account that have completed measurement efforts already. In order to make them continue, one needs to regularly solicit - through demonstrations and communication - the ones that are predicted to continue measuring in order to achieve their future alignment with the procedure.

In order to entirely follow the DSRM, the artifact, in this case the NoiseTube application, would have had to be designed from scratch. However, time restrictions did not allow for this activity, but obliged to opt for an existing artifact, which was considered the best solution for the given context and requirements. Further, the process of designing an artifact, in this case building a mobile application, is a somewhat specialized activity, requiring expertise in the field of software applications. This expertise takes time to acquire, which is another explanation as to why activity 3 could not be exploited in integrality.

Consistent with von Alan et al. (2004), the NoiseTube experiment also has to be evaluated from a DS perspective, and had to satisfy seven different guidelines (figure 4) in order to be considered a relevant and valid DS research experiment. The NoiseTube experiment, within the given time constraint, covered all seven guidelines and can therefore be granted validity within DS. .

Nevertheless, its research rigor (guideline 5), compared to other DS tasks, can be classified as moderately stringent. Frankly judged, evaluation guideline 3, tackling the artifact and design evaluation, would be considered a moderately well-implemented in this case. For one, this is due to the overlooking of missing functionalities of NoiseTube, but for another, the research design of the method comparison, especially the investigation of further collaborative noise sensing applications, could have been expanded with other apps and thus ameliorated.

Further, one had to realize that, despite mostly fulfilling the research guidelines, the NoiseTube investigation experienced a rather poor analysis outcome, especially represented by the mixed statistical results, but also the limited volunteering effort. It turned out that most users would experience the previously mentioned honeymoon period before rapidly reducing their measurement effort. Therefore, DS, as ubiquitous it might be as a theoretical research foundation, does not guarantee a successful practical implementation of an artifact.

A possible research restriction that could have played a role for the limited research outcome is that neither methodology nor the evaluation criteria sufficiently emphasize the social aspect of the given setting. Despite the fact that, as mentioned in Section 3.7, the goal of DS is to assess emergent properties caused at the crossroads of technology and social systems, it needs to further expand its focus on the social aspect of technology in order to entirely cover a collaborative sensing experiment. Consequently, there has to be another research stream that needs to become merged with DS in order to obtain a sufficient research foundation. Hence, the idea is to combine DS with natural science for the academic justification of the analysis' outcome (March and Smith, 1995).

Natural science deals with conventional research in biological, social and behavioral domains. It is often represented as a compilation of two activities: discovery and justification. Discover can be deemed as the process of introducing new scientific claims: natural scientists conceptualize specific language that is used to characterize these phenomena, which are then applied in models or theories, "making claims about the nature of reality" (March and Smith, 1995, p.253).

Further, March and Smith (1995) describes justification as the actions that validate these claims. In the example of NoiseTube, natural science thus provides a research foundation as to how peo-

ple understand the use of technology in a given setting, i.e. how they socially deal and behave in the situation of participatory noise measurement. Further, natural science, in spite of the lacking conclusions drawn from NoiseTube on noise-related health effects, serves as groundwork for the latter, which can be considered a direction for further research.

While DS attempt to produce artifacts that serve human purposes, natural science thus seeks to understand the reality (March and Smith, 1995). Therefore, as technology-oriented as DS might be, it needs to be combined with natural science in order to result in a qualitative and valuable research foundation. The combination of the two research directions would lead to an adequate understanding of the prerequisites that have to be taken into account when evaluating an artifact like NoiseTube.

Importantly, as natural science is descriptive and explanatory in essence, artifacts created or evaluated in a DS context have to incorporate the former's instructions (March and Smith, 1995) in order to sense fruitful research output. Even more, DS research can be reinforced by in-depth understanding of natural phenomena. In this case, the approach taken from a DS perspective should be to analyze the behavioral dimension of the NoiseTube experiment, which means to gain insight into user attitude towards the artifact's technology, an aspect discussed in the following section.

The implemented research experiment, by building on the DS research framework and identifying a certain shortage that had to be bridged via natural science, has thus been able to marginally contribute to existing DS literature.

7.3.3 Technology acceptance

Drawing on the previous section, it is not exclusively the social and behavioral domain of natural science that bridges the gap between academic foundations and satisfying analysis outcomes. In order to really understand the depth of such an experiment, one has to go beyond a mere review of necessary or adequate research streams. In fact, given the technological dimension of the undertaking, one has to examine the social setting for a the prevailing technology, meaning one has to take into account how people are reacting to new technology and how their acceptance or tolerance evolves over time. Lack of user acceptance has been deemed a restriction to the successful implementation of IS for a while already, even though most IS eventually end up improving people's jobs or daily routines (Davis, 1993).

As with any kind of technology, there is an adoption lifecycle by its end users (Figure 21).The diffusion of the technology is initiated by a somewhat manageable share of users that adopt it for the sake of technology, rather than for the general utility the application would provide (Tiwana, 2013). Those so-called geeks are usually followed by early adopters, who then form approximately a sixth of a potential user base. In the case of NoiseTube, the lifecycle can therefore be considered to reside at its very beginning, and the participating volunteers can be estimated to about 3 per cent of a possible user base (Tiwana, 2013). Consequently, the adoption lifecycle is a clear explanation why the number of users has remained minimal for the experiment.

In the case of NoiseTube, it seems clear that, in order to hit a so-called utilization "chasm", a situation where one wave of adopters succeeds another in a way that 85 percent of potential users embrace the technology (Tiwana, 2013, p.52), NoiseTube needs more time. No new technology approach grew overnight and instantly hit its full base of potential users.

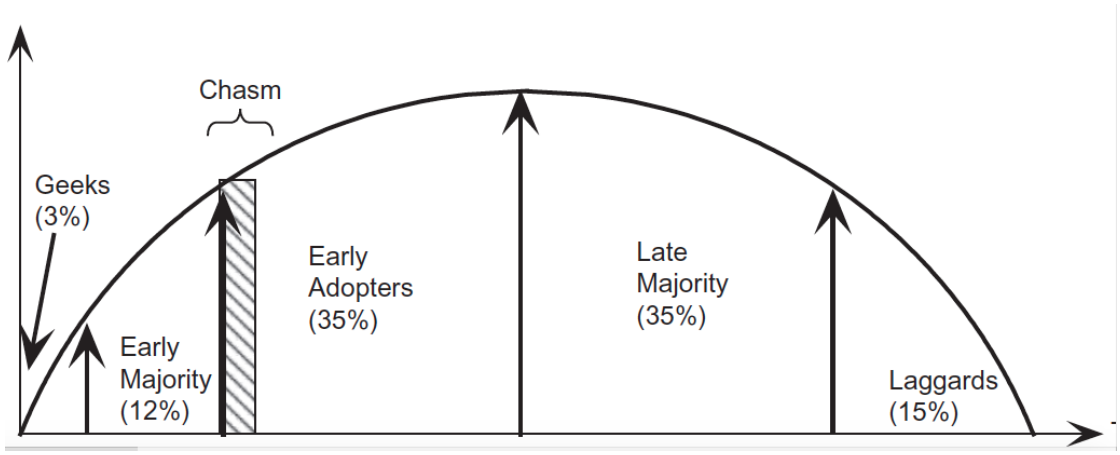

Figure 21: The technology adoption lifecycle

Another factor that determines the success of a technology prevails with its ease of use. The model to refer to in this case is the technology acceptance model (TAM), depicted in Figure 22.

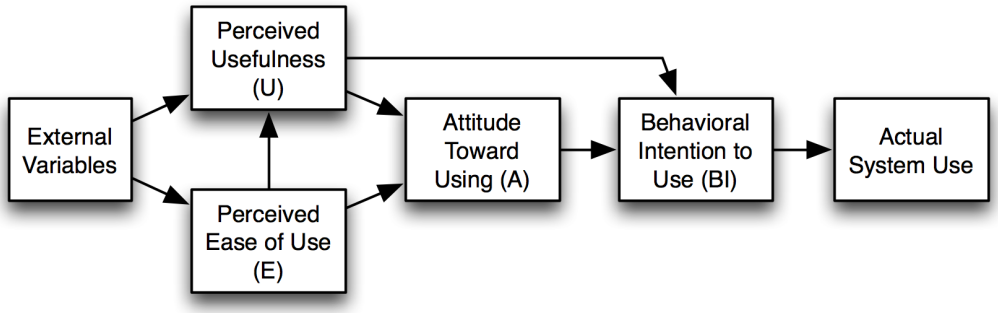

Figure 22: The technology acceptance model

As user initiate their measurement effort during a honeymoon period, they develop a certain attitude towards the NoiseTube method, which is influenced by the perceived usefulness and the perceived ease of use (Davis, 1993). This is where the method experiences a downturn. In fact, most users would not perceive the technology as very useful for their daily routines. Further, they would not realize a personal benefit in completing measurements, unless, as already mentioned, there was financial rewards involved or they were obliged to do so via contract.

The perceived ease of use causes a severe plunge in acceptance and attitude towards the NoiseTube technology, simply because measuring is a rather inconvenient task, which was also highlighted by the volunteers. Given the fact that, in order to obtain measurements qualitatively similar to modeled ones, one is required to actively hold the phone in one's hand to measure and should not leave it in the pocket, one can conclude that the method is rather cumbersome, to say the least. In this context, it is the method's measurement process, rather than its technology

that negatively impacts the perceived ease of use. Nevertheless, it reduces the general perceived ease of use of the mobile technology and thus constitutes a quite straightforward explanation as to why the method's tolerance can be described as paltry.

The perceived usefulness of a technology artifact can also be considered a motivation behind the success of several crowdsourcing elements (Davis, 1993). Drawing back on Section 3.8, positive examples of crowdsourcing procedures such as Wikipedia only underline the fact that there has to be an individual benefit for a NoiseTube volunteer in order to guarantee a large-scale launch, although those benefits might be traced back to different motivations (Bryant et al., 2005).

Within Wikipedia, people would contribute for the sake of the greater good, which clearly confirms the prevailing usefulness perception that partly guarantees positive attitude towards the technology (Bryant et al., 2005). With NoiseTube however, the relation between measurement and greater good has not been achieved yet. What is more, although participants are supposed to contribute for the greater good, the reality seems to be that they rather contribute to the profit of one single entity, the municipality. Therefore, there is an evident controversial in terms of crowdsourcing, consistent with Haklay (2010), that also adds to the argument of missing perceived usefulness.

With reference to Section 3.8, one cannot confirm that NoiseTube enables the creation of network effects. The issue here is that in order to contribute to such a platform, volunteers need to perceive an individual benefit (Maisonneuve et al., 2010), which cannot be deemed the case here. People are reluctant to contribute, especially because the fact that other people might have realized the benefit of doing so does not mean that they would. As such, the sharing benefit mentioned in Section 3.5.2 and intended by NoiseTube developers through their Elog (Section 3.8) cannot be considered accurate and constitutes another reason as to why the method has not seen enough volunteers at this point in time.

Needless to say, the prevailing stage of NoiseTube's technology lifecycle, together with perceived ease of use and usefulness, contain fairly high explanatory power as to why the experiment has not revealed itself as an ubiquitous noise measurement method at this given stage. One simply needs to provide volunteers with a personal benefit, as well as significant improvements in ease of use, without however reducing the quality of the data gathering itself.

The implemented research experiment, by building on the DS research framework and identifying a certain shortage that had to be bridged via natural science, has thus been able to marginally contribute to existing DS literature. Further, it has highlighted that technology acceptance, especially perceived usefulness and ease of use, have to be taken into account. One should, next to including social and behavioral aspects into the DS research methodology, carefully analyze a given technology in terms of the two mentioned aspects shown in Figure 22 in order to experience fruitful research outcomes.

8 Conclusion

8.1 Main findings

This study attempted to assess whether a collaborative noise measurement approach was able to complement information about noise pollution in Rotterdam by adequately sustaining existing noise measurement techniques. Noise pollution is an environmental phenomenon that affects sustainability and living standards and cannot be dealt with by policy makers exclusively, but does call for the involvement of citizens and the analysis of their behaviors (Maisonneuve, Stevens, Niessen, Hanappe and Steels, 2009).

NoiseTube does not indicate a clear direction as to whether modeled noise levels through traditional means are inferior or superior to the real experienced noise level in Rotterdam. This paper has shown that modeled and measured data are decidedly complicated to compare. Measurement differences between the two data types vary by neighborhood, but also depend on which type of modeled data they are compared to. Generally, the best way to compare NoiseTube data to modeled data is to use the latter in the form of contours, where aggregated noise levels are calculated at the edge of a certain area, which can be seen in Figure 9.

Additionally, NoiseTube does not enable an allocation of noise origins to individual measurements, nor does it enable the inference of potential effects of single noise data. As such, one cannot establish a definite statement as to whether the collaborative method would produce exact representations of the real noise levels experienced by citizens in Rotterdam. Nevertheless, the method enables insight into noise dynamics that modeled data do not provide and therefore represents a potential auxiliary to traditional noise measurement.

One clearly has to persuade sufficient users to participate in order to have fruitful outcomes for collaborative noise measurement. Without network effects, increased ease of use and rising perceived usefulness, people would simply remain reluctant to join. As such, one needs to find other ways to involve people in the procedure, such as by financially incentivizing them.

The results of the predictive model are clear: it shows a heavy dependency of future noise measurement on noise sensitivity and thus individual annoyance. People would simply not continue their measurement endeavors, except for a few technology lovers, if they would not be incentivized to do so, simply because they would not enjoy noise measurement while being (highly) sensitive to and annoyed by noise. Hence, as fruitful the idea of collaborative noise sensing might sound, the model also stresses that the only solution towards a large-scale implementation is to include a financial reward into its process, both for people that have already measured and for those that have not. As an example, one could include the activity into the contract of several municipality employees, such as parking ticket controllers or garbage men, but also to partner up with delivery firms or flyer distributing companies.

DS does provide a prosperous research framework to conduct the NoiseTube experiment. However, the sub-category of IS failed to amply recognize the social and behavioral aspect that is involved in the evaluation of artifacts. DS alone does not suffice as a research foundation to guarantee a fruitful research outcome, it should rather be combined with natural science and take several aspects of the technology acceptance model into account, for instance the impact of perceived ease of us and of perceived usefulness.

8.2 Managerial implications

The city of Rotterdam has undertaken ambitious programs to augur smart initiatives, and authorities are attempting to maximize the use of ICT innovations for urban development. Therefore, the NoiseTube approach certainly has the potential to achieve goal congruence with Rotterdam's SC resolutions. Not only do such collaborative and data-driven approaches provide residents with increased empowerment and responsibility, they also enable connectedness with other initiatives that attempt to ameliorate living standards, for instance air quality projects.

For Rotterdam's urban planners, the approach of using humans as mobile sensors is very attractive in a SC context: it can collect incredible amounts of data and it is a data-driven, and crowdsourced process that can bring in-depth insight about the city's noise distribution dynamics. To answer the research question of this paper, it does positively impact the analysis of noise pollution in Rotterdam. However, at this given point in time, the reach is too limited for the municipality to be able to take real action, let alone to utilize the approach to increase quality of living. Urban planners also absolutely need to ensure that a collaborative measurement tool adheres to the same quality standards as modeled data. For one, this means that the data need to guarantee geometrical veracity, but also do they have to be equivalent in terms of recorded noise levels to be successful.

In the case where NoiseTube does not include a straightforward feature that would consider the bookkeeping of noise origins, the link between individual noise measurement and its source obviously becomes convoluted and complicated to store for later analysis. What is worse, the application does not encourage a user to express any kind of reaction through its tagging function, as one has to type it while measuring, which is clearly inconvenient. Therefore, a clear noise source analysis is not possible with the given functionalities and therefore constitutes a significant impediment for further municipality action planning.

Drawing on Section 7.3.3, another possibility to expand the reach of NoiseTube would be to improve its ease of use. For an application like NoiseTube in order to perform qualitative measurements, the user guidelines prescribe to actively hold the phone at hand to measure. However, in order to make the measurement process more pleasant, one could attempt to connect the phone with a tiny microphone installed on someone's clothing. This would ensure the same measurement quality as if one was holding the phone in the hand, but would represent a significantly more enjoyable way of doing so.

In order to further improve NoiseTube's positional veracity, one could also, as already suggested by Maisonneuve, Stevens, Niessen, Hanappe and Steels (2009), add external GPS receivers via Bluetooth. However, this is rather a suggestion when a person is measuring at one place, as carrying the receiver around would not upgrade but rather worsen the inconvenient measurement process, contrasting the argument in the previous paragraph.

In order to have sufficiently detailed results, one would require an estimated 100 people per square kilometer to constantly measure noise through their mobile phones. To achieve this degree of resident involvement with application like NoiseTube, authorities need to think about a way to either give people an incentive to participate, or to oblige certain parties to do so.

At the municipality level, one could for instance partner with post delivery or advertising companies that employ flyer distributors, and persuade them to equip their agents with the tech-

nology. In this case, this would lead to a considerable increase in data gathering, as those individuals would not only measure noise while going from A to B, but they would follow nearby citizens to distribute their flyers, let alone into a majority of neighborhoods in order to insert flyers into mailboxes. Hence, the data collection could be amplified through this collaboration.

The NoiseTube approach definitely requires broad promotion and communication. Residents will not embrace the approach if they are simply not aware of it. Hence, urban planners should ensure a citywide promotion in order to guarantee successful large-scale usage. Strongly encouraged is also a clear user guide when launching a collaborative noise measurement tool. Users need to be exactly aware of how to use the equipment in order to maximize result quality.

Furthermore, there is a tradeoff between cost and return of the given approach. Traditional noise measurement occurs a certain cost and, although it remains complicated to estimate benefit values for a collaborative noise measurement approach, one has to also weigh its advantages against the incurred costs. Given the limited budget that usually has to be taken into account, the fact of including the task of measuring into certain municipality employees' duties would not cause additional costs, however subsidy-based partnerships with the aforementioned companies or incentives to persuade other volunteers could induce costs. As a result, urban planners, as with any innovation project, would have to gauge whether the value of such a project would exceed the encountered costs.

Not only does the approach imply a financial tradeoff, but also a bargain between measurement quality and individual comfort. The better the measurement quality, the more personal effort needs to be made, such as holding the phone at hand, trying to repeatedly measure at multiple locations during different times, and so on. Therefore, one has to weigh whether this personal commitment is not excessive, which is another argument as to why urban planners need to oblige people via contract or financially subsidize participants.

Another way of amplifying data collection would be to complement the NoiseTube approach with the installment of permanent sensors in several areas. Despite the fact that this would not entirely fall into collaborative noise monitoring, this decision could however seriously amplify data gathering. Here, one would also have to weigh the incurred costs of installing and maintaining those sensors against their potential value when capturing additional noise data.

In parallel with noise level recording, municipality planners should not ignore that the most evident remedy noise pollution is to tackle noise at its source (Hildebrand, 1970). Even though reduction of noise at its origin clearly benefits from data-driven measurement activities, several other ideas such as plastic roads could be valuable for urban planners in the short-term.

8.3 Limitations and future research

One of the clear caveats resided within the defective number of volunteers to implement the experiment. Even though several critical insights were obtained, results could have been amplified with a much larger amount of contributors. Therefore, for future collaborative research experiments, one clearly needs to ensure the participation of sufficient volunteers. In order to receive adequate results in a study like the one done through NoiseTube, an objective should be to involve around 100 people per square kilometer that would also permanently, not occasionally, measure noise or other conditions.

A somewhat anticipated limitation is that there is no 100 percent guarantee that collaborative noise sensing tools like NoiseTube would produce the same results as modeled data, making it hard to properly compare the two methods. Especially because NoiseTube uses a slightly different indicator then the one DCMR noise levels were represented in, the noise levels could possibly diverge by a few decibels. Moreover, the modeled data are established for a height of 4 meters, whereas most likely a majority of NoiseTube data would have been gathered at pedestrian level, which further complicated the comparison. One cannot eliminate the comparison uncertainty, especially because cell phone microphones have not the same quality as professional sound meters and might need to be calibrated in some cases. However, given the fact that modeled data would also not always be entirely accurate, this is a factor that has to be taken into account for this kind of public participatory experiment.

As already mentioned earlier, the limited functionalities of the NoiseTube application somewhat impeded a tracing of individual measurements to noise sources. Neither did the application permit the conclusion about potential adverse effects of individual measurements, which would have been important to implement a complete noise pollution analysis. It remains within the developer's hands to ensure that the missing functionalities will be added in order to make a serious alternative for a collaborative measurement tool.

Looking at future research directives, as most people would only measure in case they received a financial incentive, the construction of a reward system for a crowd-sourced sensing activity clearly remains crucial. Two categories of financial incentives would have to be designed: one for people who have measured before, taking advantage of the predictive model's insights, and that needs to convince people to initiate a measurement commitment.

In this context, a research idea might be to design and evaluate a gaming or competition experiment that would try to maximize participation. It seems logical that in order to make a collaborative sensing approach work, one needs to conduct experiments that would assess how people would be willing to participate in such a noise measurement method. This would build on the user behavior model of this paper and could be expanded through surveys, focus groups, and incentive-based experiments.

In order to refute the previously mentioned comparison uncertainty one would need to establish an official EU-prescribed comparison method that would then exactly specify how the two methods should be assessed against each other. This constitutes an interesting further research direction, especially because trends towards crowdsourcing appear in many different settings, making it trivial for the EU to prescribe an official methodology to compare crowdsourced and modeled environmental data.

Further, another research direction could be to think of a way to further develop the user measurement model used in this paper. Although decision trees represent a great way to identify possible measurement behaviors, they might be too simplistic. Even though a simple description produces accurate results, there might be other ways to describe them. Further, there could even other predictor variables that would play a significant role in the prediction of user measurement behavior. Hence, future research could build on this paper in order to either deepen the relationship between future measurement behavior and NoiseTube data, or to simply create a model that would predict how people that have never measured before would be assumed to act as collaborative measurement volunteers.

Bibliography

Albors, J., Ramos, J. C. and Hervas, J. L. (2008). New learning network paradigms: Communities of objectives, crowdsourcing, wikis and open source, *International Journal of Information Management* **28**(3): 194–202.

Artusi, R., Verderio, P. and Marubini, E. (2002). Bravais-pearson and spearman correlation coefficients: meaning, test of hypothesis and confidence interval, *Int J Biol Markers* **17**(2): 148–151.

Bridge, P. D. and Sawilowsky, S. S. (1999). Increasing physicians' awareness of the impact of statistics on research outcomes: comparative power of the t-test and wilcoxon rank-sum test in small samples applied research, *Journal of clinical epidemiology* **52**(3): 229–235.

Bryant, S. L., Forte, A. and Bruckman, A. (2005). Becoming wikipedian: transformation of participation in a collaborative online encyclopedia, *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*, ACM, pp. 1–10.

Campbell, A. T., Eisenman, S. B., Lane, N. D., Miluzzo, E., Peterson, R. A., Lu, H., Zheng, X., Musolesi, M., Fodor, K. and Ahn, G.-S. (2008). The rise of people-centric sensing, *Internet Computing, IEEE* **12**(4): 12–21.

Cox, P. and Palou, J. (2002). Directive 2002/49/ec of the european parliament and of the council of 25 june 2002 relating to the assessment and management of environmental noise, *Annex I, OJ* **189**(18.7): 2002.

Davis, F. D. (1993). User acceptance of information technology: system characteristics, user perceptions and behavioral impacts, *International journal of man-machine studies* **38**(3): 475–487.

De Coensel, B. and Botteldooren, D. (2014). Smart sound monitoring for sound event detection and characterization, *43rd International Congress on Noise Control Engineering (Inter-Noise 2014)*.

D'Hondt, E., Stevens, M. and Jacobs, A. (2013). Participatory noise mapping works! An evaluation of participatory sensing as an alternative to standard techniques for environmental monitoring, *Pervasive and Mobile Computing* **9**(5): 681–694.
URL: *http://linkinghub.elsevier.com/retrieve/pii/S1574119212001137*

Drosatos, G., Efraimidis, P. S., Athanasiadis, I. N., Stevens, M. and D'Hondt, E. (2014). Privacy-preserving computation of participatory noise maps in the cloud, *Journal of Systems and Software* **92**: 170–183.
URL: *http://www.sciencedirect.com/science/article/pii/S0164121214000430*

DutchNews (2011). High speed line makes too much noise.

Fawcett, T. (2006). An introduction to roc analysis, *Pattern recognition letters* **27**(8): 861–874.

Fienen, M. N. and Lowry, C. S. (2012). Social.Water—A crowdsourcing tool for environmental data acquisition, *Computers and Geosciences* **49**: 164–169.
URL: *http://linkinghub.elsevier.com/retrieve/pii/S0098300412002105*

Franssen, E. a. M., van Wiechen, C. M. a. G., Nagelkerke, N. J. D. and 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.

Fyhri, A. and Klæboe, R. (2009). Road traffic noise, sensitivity, annoyance and self-reported healthÑa structural equation model exercise, *Environment International* **35**(1): 91–97.

Goines, L. and Hagler, L. (2007). Noise pollution: a modern plague, *SOUTHERN MEDICAL JOURNAL-BIRMINGHAM ALABAMA-* **100**(3): 287.

Goodchild, M. F. and Li, L. (2012). Assuring the quality of volunteered geographic information, *Spatial statistics* **1**: 110–120.

Gregor, S. and Jones, D. (2007). The anatomy of a design theory, *Journal of the Association for Information Systems* **8**(5): 312.

Haklay, M. (2010). How good is volunteered geographical information? a comparative study of openstreetmap and ordnance survey datasets, *Environment and planning B: Planning and design* **37**(4): 682–703.

Heipke, C. (2010). Crowdsourcing geospatial data, *ISPRS Journal of Photogrammetry and Remote Sensing* **65**(6): 550–557.

Hildebrand, J. L. (1970). Noise pollution: An introduction to the problem and an outline for future legal research, *Columbia Law Review* **70**(4): 652–692.

Howe, J. (2006). The rise of crowdsourcing, *Wired magazine* **14**(6): 1–4.

Kahle, D. and Wickham, H. (2013). ggmap: Spatial visualization with ggplot2, *The R Journal* **5**(1): 144–161.

Kam, P., Kam, A. and Thompson, J. (1994). Noise pollution in the anaesthetic and intensive care environment, *Anaesthesia* **49**(11): 982–986.

Kitchin, R. (2014). The real-time city? big data and smart urbanism, *GeoJournal* **79**(1): 1–14.

Koppius, O. and Belo, R. (2015). *Big Data Management and Analytics: Session 4: Model Fit and Evaluation*, Erasmus University Rotterdam.

Maisonneuve, N., Stevens, M., Niessen, M. E., Hanappe, P. and Steels, L. (2009). Citizen noise pollution monitoring, *Proceedings of the 10th Annual International Conference on Digital Government Research: Social Networks: Making Connections between Citizens, Data and Government*, Digital Government Society of North America, pp. 96–103.

Maisonneuve, N., Stevens, M., Niessen, M. E. and Steels, L. (2009). Noisetube: Measuring and mapping noise pollution with mobile phones, *Information Technologies in Environmental Engineering*, Springer, pp. 215–228.

Maisonneuve, N., Stevens, M. and Ochab, B. (2010). Participatory noise pollution monitoring using mobile phones, *Information Polity* **15**(1, 2): 51–71.

March, S. T. and Smith, G. F. (1995). Design and natural science research on information technology, *Decision support systems* **15**(4): 251–266.

Marrocco, C., Duin, R. P. and Tortorella, F. (2008). Maximizing the area under the roc curve by pairwise feature combination, *Pattern Recognition* **41**(6): 1961–1974.

Martin, M., Tarrero, A. I., Machimbarrena, M., Gonzalez, J. and De Garibay, V. G. (2011). A methodology to study noise annoyance and to perform action plans follow up using as input an existing survey and noise map: Application to the city of málaga (spain), *Applied Acoustics* **72**(8): 495–504.

Mousa, H., Mokhtar, S. B., Hasan, O., Younes, O., Hadhoud, M. and Brunie, L. (2015). Trust management and reputation systems in mobile participatory sensing applications: A survey, *Computer Networks* **90**: 49–73.

Muzet, A. (2007). Environmental noise, sleep and health, *Sleep medicine reviews* **11**(2): 135–142.

Nam, T. and Pardo, T. A. (2011). Smart city as urban innovation: Focusing on management, policy, and context, *Proceedings of the 5th international conference on theory and practice of electronic governance*, ACM, pp. 185–194.

Neville, P. G. (1999). Decision trees for predictive modeling, *SAS Institute Inc* .

Ouis, D. (2001). Annoyance from road traffic noise: a review, *Journal of environmental psychology* **21**(1): 101–120.

Peffers, K., Tuunanen, T., Rothenberger, M. A. and Chatterjee, S. (2007). A design science research methodology for information systems research, *Journal of management information systems* **24**(3): 45–77.

Pentland, A. and Liu, A. (1999). Modeling and prediction of human behavior, *Neural computation* **11**(1): 229–242.

Perera, C., Zaslavsky, A., Christen, P. and Georgakopoulos, D. (2014). Sensing as a service model for smart cities supported by internet of things, *Transactions on Emerging Telecommunications Technologies* **25**(1): 81–93.

Rana, R., Chou, C. T., Bulusu, N., Kanhere, S. and Hu, W. (2015). Ear-phone: A context-aware noise mapping using smart phones, *Pervasive and Mobile Computing* **17**: 1–22.

Rao, B. and Minakakis, L. (2003). Evolution of mobile location-based services, *Communications of the ACM* **46**(12): 61–65.

Roland, C. (1990). Our love affair with new technology: Is the honeymoon over?, *Art education* **43**(3): 54–62.

Salomons, E. (2013). Qside action 4: demonstrations and scenarios, *QSIDE* **LIFE09 ENV**(August): 1–34.

Salomons, E. M. and Pont, M. B. (2012). Urban traffic noise and the relation to urban density, form, and traffic elasticity, *Landscape and Urban Planning* **108**(1): 2–16.

Santini, S., Ostermaier, B. and Vitaletti, A. (2008). First experiences using wireless sensor networks for noise pollution monitoring, *Proceedings of the workshop on Real-world wireless sensor networks*, ACM, pp. 61–65.

Sekaran, U. and Bougie, R. (2003). *Research Methods for Business: A Skill Building Approach. 2003*, John Willey and Sons, New York.

Shepherd, D., Welch, D., Dirks, K. N. and Mathews, R. (2010). Exploring the relationship between noise sensitivity, annoyance and health-related quality of life in a sample of adults exposed to environmental noise, *International journal of environmental research and public health* **7**(10): 3579–3594.

Singh, N. and Davar, S. (2004). Noise pollution-sources, effects and control, *J. Hum. Ecol* **16**(3): 181–187.

Sørensen, M., Hvidberg, M., Andersen, Z. J., Nordsborg, R. B., Lillelund, K. G., Jakobsen, J., Tjønneland, A., Overvad, K. and Raaschou-Nielsen, O. (2011). Road traffic noise and stroke: a prospective cohort study, *European heart journal* **32**(6): 737–744.

Stuijt, A. (2009). Noise pollution kills 600 dutch a year.
URL: *http://www.digitaljournal.com/article/267835*

Theakston, F. (2011). Burden of disease from environmental noise-Quantification of healthy life years lost in Europe. 2011, *World Health Organization* pp. 1–105.

Tiwana, A. (2013). *Platform ecosystems: aligning architecture, governance, and strategy*, Newnes.

von Alan, R. H., March, S. T., Park, J. and Ram, S. (2004). Design science in information systems research, *MIS quarterly* **28**(1): 75–105.

Wolfert, H. (2015). Noise pollution interview with the policy officer of international and european affairs, henk wolfert.

Zannin, P. H. T., Ferreira, A. M. C. and Szeremetta, B. (2006). Evaluation of noise pollution in urban parks., *Environmental monitoring and assessment* **118**(1-3): 423–433.

Zukerman, I. and Albrecht, D. W. (2001). Predictive statistical models for user modeling, *User Modeling and User-Adapted Interaction* **11**(1-2): 5–18.

Appendix A Summary EU Directive 2002/49/EC

- EU Directive 2002/49/EC on the management of environmental noise *(5)*; and
- the follow-up and further development of existing EU legislation relating to sources of noise such as motor vehicles, aircraft and railway rolling stock, and the provision of financial support to noise-related studies and research projects.

The European Parliament and Council adopted Directive 2002/49/EC of 25 June 2002, whose main aim is to provide a common basis for tackling noise problems across the EU. The underlying principles of the Directive are similar to those for other environment policy directives:

- monitoring the environmental problem by requiring competent authorities in Member States to produce strategic noise maps for major roads, railways, civil airports and urban agglomerations, based on harmonized noise indicators;
- informing and consulting the public about noise exposure, its effects and the measures considered to address noise, in line with the principles of the Aarhus Convention *(13)*;
- addressing local noise issues by requiring competent authorities to draw up action plans to reduce noise where necessary and maintain environmental noise quality where it is good (the Directive does not set any limit value nor does it prescribe the measures to be used in the action plans, which remain at the discretion of the competent authorities); and
- developing a long-term EU strategy, including objectives to reduce the number of people affected by noise and providing a framework for developing existing EU policy on noise reduction from sources.

Detailed information is available on the authorities responsible for implementing the Directive in Member States and on the agglomerations, major roads, railways and airports to be covered by the noise maps and action plans.

Appendix B Leq Indicator

$$L_{eq} = 10\log\frac{1}{T}\int_0^T \frac{p^2}{p_0^2}dt$$

Appendix C NoiseTube Analysis R script

```
#install.packages("jsonlite")
library(jsonlite)

#--- Input the json files from the different custom maps
Finaldata1 <- fromJSON("Kralingen.json")
Finaldata2 <- fromJSON("RottWest.json")
Finaldata3 <- fromJSON("Capelle.json")
Finaldata4 <- fromJSON("Centralmap.json")

#--- The trick is the flatten() function
#--- for all four datasets
```

```
names(flatten(Finaldata1))
flatten(Finaldata1)$measures
names(flatten(Finaldata2))
flatten(Finaldata2)$measures
names(flatten(Finaldata3))
flatten(Finaldata3)$measures
names(flatten(Finaldata4))
flatten(Finaldata4)$measures

#--- Application of the flatten() function
#--- within a do.call that rbind (row binds)
#--- the resulting data frames
dfMeasures1 <- do.call(rbind,flatten(Finaldata1)$measures)
dfMeasures2 <- do.call(rbind,flatten(Finaldata2)$measures)
dfMeasures3 <- do.call(rbind,flatten(Finaldata3)$measures)
dfMeasures4 <- do.call(rbind,flatten(Finaldata4)$measures)

#--- combine all the datasets into one
Finaldataset <- rbind(dfMeasures1, dfMeasures2, dfMeasures3, dfMeasures4)

#remove unnecessary columns & outlier analysis
Finaldataset$created_at <- NULL
Finaldataset$id <- NULL
Finaldataset$track_id <- NULL

#---change the variable loudness from chr to numeric
#--- use the read.csv whenever a new script is started
#write.csv(Finaldataset,"finaldataset.csv", row.names = FALSE)
Finaldataset <- read.csv("finaldataset.csv", header = TRUE)
Finaldataset$loudness_index<-NULL

#--- visusalizations of noise levels data (loudness)
#--- create a histogram
hist(FinalData$loudness, main = "Noise levels in Rotterdam", xlab = "Noise level (dB)",
      xlim = c(20,80), ylab = "Number of measurements", Ias=1,
      col= "darkgreen", border = "black")

#--- create a boxplot
boxplot(FinalData$loudness, main = "Noise levels in Rotterdam", ylab = "Noise level (dB)")

#--- identifying the number of users, 41 users
FinalData$user_id <- round(FinalData$user_id, digits = 0)
summary(as.factor(FinalData$user_id))

#---- get a better format for the dates in noisetube data
Finaldataset$made_at <- as.Date(Finaldataset$made_at)
Finaldataset$Date <- format(Finaldataset$made_at, format="%d %m %Y")
names(Finaldataset)[3] <- c("Date")

#--- rename columns in weather data set
x <- as.data.frame(Finalmeteodata_hourly)
names(x)[1:7] <- c('Region', 'Date','Time','Hourlywindspeed',
                   'MaxWindspeed','Rainfalltime', 'TotalhourlyRain')
```

```
#--- exclude all hours with heavy wind
rain <- which(x$Hourlywindspeed<70 )
x <- x[rain,]
wind <- which(x$MaxWindspeed <70)
x <- x[wind,]

#--- exclude all hours with rain
bad <- which(x$Rainfalltime < 0.5)
x <- x[bad,]

#--- save csv file of x
write.csv(x,"finalweatherdata.csv", row.names = FALSE)
x <- read.csv("finalweatherdata.csv", header = TRUE)

#--- check whether the finaldataset does include days with bad weather conditions
as.Date(as.character(x$Date), "%Y-%m-%d")
x$Date <- as.Date(as.character(x$Date), "%Y%m%d")
merge(Finaldataset, x, by="Date")

#--- exclude the top five and bottom five percent
tail(sort(Finaldataset$loudness), 6263) #top 5
L5 <- as.data.frame(tail(sort(Finaldataset$loudness,
                    decreasing = TRUE), 6263)) #bottom 5 percent
L95 <- as.data.frame(tail(sort(Finaldataset$loudness), 6263)) #top 5 percent
April.L95 <- subset(Finaldataset, Finaldataset$loudness < 70.33)
Finaldataset2 <- subset(April.L95, April.L95$loudness > 22.17)

#--- use the read.csv whenever a new script is started
#write.csv(Finaldataset2,"finaldataset2.csv", row.names = FALSE)
Finaldataset2 <- read.csv("finaldataset2.csv", header = TRUE)

#--- add some jitter to Finaldataset2
loudness <- jitter(FinalData$loudness, factor = 1.1)
user_id <- jitter(FinalData$user_id, factor = 2)
lat <- jitter(FinalData$lat, factor = 1.002)
lng <- jitter(FinalData$lng, factor = 1.002)
Noise <- as.data.frame(loudness)
user <- as.data.frame(user_id)
log <- as.data.frame(lng)
lat <- as.data.frame(lat)
Data <- cbind(loudness,lat,log,user_id)
FinalData <-rbind(FinalData, Data)
FinalData$user_id <- round(FinalData$user_id, digits = 0)

#remove unnecessary columns
FinalData$rangesver<-NULL
FinalData$rangeshor<-NULL

#--- save csv file for jitter data
#write.csv(FinalData, "FinaldatasetwithCells.csv", row.names=FALSE)
FinalData <- read.csv("FinaldatasetwithCells.csv", header = TRUE)
merge(FinalData, Finaldataset2, by="user_id")
```

```
#--- summary statistics by box of coordinates
#--- replace the april.map by april.final to obtain the L95-L5 dataset
#--- replace the same dataset for the grid function in the other script
library(plyr)
Cellavstats <- ddply(FinalData,~CellNr,summarise,mean=mean(loudness),
                     sd=sd(loudness), lat=mean(lat), long=mean(lng))
latlong <- paste(Cellavstats$lat, Cellavstats$lng, sep = ":")
#write.csv(Cellavstats, "Aggregatedata.csv", row.names=FALSE)

#--- read aggregated data QGIS attribute table into R
library(PBSmapping)
Mapcomp2 <- importShapefile(file.choose())
Mapcomp3 <- importShapefile(file.choose())
#write.csv(Mapcomp2, "comparisondata.csv", row.names=FALSE)
#write.csv(Mapcomp3, "comparisondata2.csv", row.names=FALSE)
Mapcomp2<- read.csv("comparisondata.csv", header = TRUE)
Mapcomp3<- read.csv("comparisondata2.csv", header = TRUE)

#--- Statistical analysis attribute table
#--- Spearman Rank correlation 2007 Data aggregated NT data
Mapcomp2$KLASSE <- as.numeric(Mapcomp2$KLASSE)
cor.test(Mapcomp2$mean, Mapcomp2$KLASSE, method = "spearman")
#--- Wilcoxon Test 2007 Data
wilcox.test(Mapcomp2$mean, Mapcomp2$KLASSE, paired = TRUE)

#--- Spearman Rank correlation 2011 Data aggregated NT data
Mapcomp3$meanVL_LDE <- as.numeric(Mapcomp3$meanVL_LDE)
cor.test(Mapcomp3$meanVL_LDE, Mapcomp3$mean, method = "spearman")
#--- Wilcoxon Test 2011 Data
wilcox.test(Mapcomp3$meanVL_LDE, Mapcomp3$mean, paired = TRUE)
```

Appendix D Grid Construction R script

```
#--- This script consists of 3 parts:
#--- a general part (for the functions etc.)
#--- then a part to create a grid
#--- then a part to add some info to your original
#--- measurement dataset, being: cell horizontal, cell vertical,
#--- and gridcell (which is a combination of cell hor and cell ver)
getwd()
#load library
library(plyr)

#Function: Apply function FUN to all combinations of
#arguments and append results to data frame of arguments
cmapply <- function(FUN, ..., MoreArgs = NULL, SIMPLIFY = TRUE,
                    USE.NAMES = TRUE)
{
  l <- expand.grid(..., stringsAsFactors=FALSE)
  r <- do.call(mapply, c(
    list(FUN=FUN, MoreArgs = MoreArgs, SIMPLIFY = SIMPLIFY,
```

```
         USE.NAMES = USE.NAMES),
    l
  ))
  if (is.matrix(r)) r <- t(r)
  cbind(l, r)
}

#--- PART: GRID

#--- LAT
earth.dist("51.861487,4.379368","51.994374,4.379368")
#14.79293
#conversion: 14.79293/(51.994374-51.861487)
#1lat = 111.3196 km
#0.01/111.3196 --> 100m = 0.0008983144
Lat100m <- 0.00008983144

#-- new borders
LatMin <- 51.875770324938564
LatMax <- 51.954485617801545

#--- LON
earth.dist("51.861487,4.379368","51.861487,4.601458")
#15.26803
#conversion 15.26803/(4.601458-4.379368)
#1lat = 68.74704 km
#0.1/68.74704 --> 10m = 0.0001454608 lon
Lon100m <- 0.001454608

#-- new borders
LonMin <- 4.2791748046875
LonMax <- 4.601923

#--- Now let's create the dataframes HorCelldf and VerCelldf
#-- first vertical / latitudes
VerCellCount <- round_any(((LatMax-LatMin) / Lat100m),
                          1, f = ceiling)
VerCelldf <- data.frame()
dynamicLat <- LatMin
for (i in 1:VerCellCount) {
  VerCelldf[i,"VerCell"] <- i
  VerCelldf[i,"VerCellLatMin"] <- dynamicLat
  dynamicLat <- dynamicLat + Lat100m
  VerCelldf[i,"VerCellLatMax"] <- dynamicLat
}
VerCelldf$VerCellLatMax<- VerCelldf$VerCellLatMax - 0.000000001
#-- then horizontal / longitudes
HorCellCount <- round_any(((LonMax-LonMin) / Lon100m),
                          1, f = ceiling)
HorCelldf <- data.frame()
dynamicLon <- LonMin
MOAR_LETTERS <- function(n=2) {
  n <- as.integer(n[1L])
```

```
  if(!is.finite(n) || n < 2)
    stop("'n' must be a length-1 integer >= 2")

  res <- vector("list", n)
  res[[1]] <- LETTERS
  for(i in 2:n)
    res[[i]] <- c(sapply(res[[i-1L]],
                function(y) paste0(y, LETTERS)))

  unlist(res)
}
LETTERSMOAR <- MOAR_LETTERS(3)
LetterVector <- LETTERSMOAR[1:VerCellCount]

for (i in 1:HorCellCount) {
  HorCelldf[i,"HorCell"] <- LetterVector[i]
  HorCelldf[i,"HorCellLonMin"] <- dynamicLon
  dynamicLon <- dynamicLon + Lon100m
  HorCelldf[i,"HorCellLonMax"] <- dynamicLon
}
HorCelldf$HorCellLonMax<- HorCelldf$HorCellLonMax - 0.000000000001
rm(dynamicLon,dynamicLat,i,HorCellCount,VerCellCount)


#--- Turn them into one Grid df
#--combine the unique identifiers of the two datasets
Grid <- cmapply(VerCell=VerCelldf$VerCell,
                HorCell=HorCelldf$HorCell, FUN=paste)
Grid <- rename(Grid, c(r="CellNr"))

Grid <- merge(Grid,HorCelldf,by.x=c("HorCell"),
              by.y=c("HorCell"))
Grid <- merge(Grid,VerCelldf,by.x=c("VerCell"),
              by.y=c("VerCell"))
Grid <- Grid[,c(3,1,2,4:7)]

write.csv(Grid, "/Users/hugokrier/Desktop/BIM MasterThesis
          /FinalThesis/RAnalysis/Scripts/Box100mRott.csv",
          row.names=FALSE)

### Part: combine GRID with Measurement-dataset

### ADD CellNr -  ###
#--- Load Grid created somewhere in calculate-distances.R
Grid <- read.csv("/Users/hugokrier/Desktop/BIM MasterThesis/
                  FinalThesis/RAnalysis/Scripts/Box100mRott.csv",
                  sep = ",", header = TRUE)
#testz <- FinalData[c(1,100,500,600,2600,2970),]

## !! need HorCelldf and VerCelldf from calculate-distances.R
#--first VerCell
#-- breaks for 'cut'
allranges=c(VerCelldf$VerCellLatMin[1],VerCelldf$VerCellLatMax)
```

```
VerCelldf$rangesver <- cut(VerCelldf$VerCellLatMin,allranges,
                           include.lowest=T)
FinalData$rangesver <-cut(FinalData$lat,allranges,include.lowest=T,
                          levels=levels(VerCelldf$rangesver))

FinalData<- join(Finaldataset,VerCelldf,by='rangesver')

#-- then HorCell
#-- breaks for 'cut'
allranges=c(HorCelldf$HorCellLonMin[1],HorCelldf$HorCellLonMax)
HorCelldf$rangeshor <- cut(HorCelldf$HorCellLonMin,
                           allranges,include.lowest=T)
FinalData$rangeshor <- cut(FinalData$lng,allranges,include.lowest=T,
                           levels=levels(HorCelldf$rangeshor))
FinalData <- join(FinalData,HorCelldf,by='rangeshor')
FinalData$CellNr <- paste(FinalData$VerCell,FinalData$HorCell)

#--- export!
#write.csv(FinalData, "FinaldatasetwithCells.csv", row.names=FALSE)
FinalData <- read.csv("FinaldatasetwithCells.csv", header = TRUE)
#--- delete redundant rows
FinalData$HorCellLonMin<-NULL
FinalData$HorCellLonMax<-NULL
FinalData$HorCell<-NULL
```

Appendix E NoiseTube Map Construction R script

```
#--- Install packages to draw maps
#install.packages("googleVis")
#install.packages("ggmap")
library(googleVis)
library(ggplot2)
library(ggmap)
library(RgoogleMaps)

#--- read new datasets
#Cellavstats dataset is with aggregated noise levels
FinalData <- read.csv("FinaldatasetwithCells.csv", header = TRUE)
#FinalData dataset is with raw measurements
Cellavstats <- read.csv("Aggregatedata.csv")

#--- Bubbled map
#--- a.get map
Roffa <-  GetMap(center= c(lon = 4.477732500000002, lat = 51.9244201),
                 zoom=13, destfile=file.path(tempdir (),"Rott.png"),
                 maptype = "mobile", SCALE = 1)

#--1b. Make bubble plots for noise, this one works
bubbleMap(Cellavstats ,coords =c("long","lat"), map = Roffa ,
          zcol =  "mean",
```

```
          key.entries= round(quantile (Cellavstats[,"mean"],(1:5) /5)))

#--- Map with the ggplot package
#--- Raw NoiseTube data and the get_map function (ggmap package)
#--- Get a map of the relevant locations of datapoints
map <- get_map(location = c(lon = mean(FinalData$lng),
                            lat = mean(FinalData$lat)),
               zoom = 14, scale = 2)

#Plot Map
ggmap(map) +
  geom_point(data = FinalData, aes(x = lng, y = lat, fill = "red", alpha = 0.8),
  size = 1, shape = 21) +
  guides(fill=FALSE, alpha=FALSE, size=FALSE)

#Map raw data without average location coordinates
#Kralingen coordinates:lon = 4.507963399999994, lat = 51.9246315
Rott <- get_map(location = c(lon = 4.526088299999969,
lat = 51.91815860000001),
                zoom = 13, crop = T, scale = "auto",
                color = "color", source = "osm")
lat <- FinalData$lat
long <- FinalData$lng
ggmap(Rott, extent = "panel", padding = 0) +
             geom_point(aes(x = long, y = lat),
             data = FinalData, alpha = .5,
             color="darkred", size = 1)

#--- Aggregated data map with the get_map function (ggmap package)
Rott <- get_map(location = c(lon = 4.526088299999969, lat = 51.91815860000001),
                zoom = 13, crop = T, scale = "auto",
                color = "color", source = "osm")
lat <- Cellavstats$lat
long <- Cellavstats$lng
ggmap(Rott, extent = "panel", padding = 0) +
  geom_point(aes(x = long, y = lat),
             data = Cellavstats, alpha = .5,
             color="darkred", size = 3) +
  ggtitle ("Data distribution Rotterdam")

#--- heat map with points coloured depending on their loudness
m <- qmap(location = c(4.51,51.92), data = statsApril, zoom = 13)
m + stat_bin2d(aes(x = long, y = lat, colour = mean), #, fill = mean
               size = 0.5, bins = 50, alpha = 0.7,
               data = Cellavstats)

m + stat_density2d(
  aes(x = long, y = lat, fill = ..level.., alpha = ..level..),
  size = 2, bins = 30, data = Cellavstats,
  geom = "polygon")
```

Appendix F Predictive Model Creation R script

```
#--- read dataset
Finaldataset <- read.csv("finaldataset.csv", header = TRUE)

#--- assign a user to the NA measurements
#--- replace NAs with random user_id
sum(is.na(Finaldataset))
Finaldataset$user_id[is.na(Finaldataset$user_id)] <- sample(1111:4444,
                     size=sum(is.na(Finaldataset$user_id)), replace = TRUE)
#--- count the number of users
summary(as.factor(Finaldataset$user_id))

#--- create a new column for the dates with only month and year
names(Finaldataset)[3] <- "date"
Finaldataset$date <- as.Date(Finaldataset$date)
Finaldataset$month <- format(Finaldataset$date, format="%Y-%m")

#--- write CSV file
#write.csv(Finaldataset, "Rotterdam0706.csv", row.names=FALSE)
Finaldataset <- read.csv("Rotterdam0706.csv", header = TRUE)

#--- identifying the number of users, 41 users
summary(as.factor(Finaldataset$user_id))

#--- Create a variable that returns 1 if people have measured
#--- in a month and 0 if they have not
#--- This dataset has its measurements in 2016, hence we will
#--- take May, April and March 2016 as target variables
Finaldataset$measuredOct11<- ifelse(Finaldataset$month == "2011-10", "1", "0")
Finaldataset$measuredJan12<- ifelse(Finaldataset$month == "2012-01", "1", "0")
Finaldataset$measuredFeb12<- ifelse(Finaldataset$month == "2012-02", "1", "0")
Finaldataset$measuredMar12<- ifelse(Finaldataset$month == "2012-03", "1", "0")
Finaldataset$measuredApr12<- ifelse(Finaldataset$month == "2012-04", "1", "0")
Finaldataset$measuredAug14<- ifelse(Finaldataset$month == "2014-08", "1", "0")
Finaldataset$measuredSep14<- ifelse(Finaldataset$month == "2014-09", "1", "0")
Finaldataset$measuredJun15<- ifelse(Finaldataset$month == "2015-06", "1", "0")
Finaldataset$measuredNov15<- ifelse(Finaldataset$month == "2015-11", "1", "0")
Finaldataset$measuredJan16<- ifelse(Finaldataset$month == "2016-01", "1", "0")
Finaldataset$measuredFeb16<- ifelse(Finaldataset$month == "2016-02", "1", "0")

#--- change the month variables from character to numeric
Finaldataset$measuredOct11 <- as.numeric(Finaldataset$measuredOct11)
Finaldataset$measuredJan12 <- as.numeric(Finaldataset$measuredJan12)
Finaldataset$measuredFeb12 <- as.numeric(Finaldataset$measuredFeb12)
Finaldataset$measuredMar12 <- as.numeric(Finaldataset$measuredMar12)
Finaldataset$measuredApr12 <- as.numeric(Finaldataset$measuredApr12)
Finaldataset$measuredAug14 <- as.numeric(Finaldataset$measuredAug14)
Finaldataset$measuredSep14 <- as.numeric(Finaldataset$measuredSep14)
Finaldataset$measuredJun15 <- as.numeric(Finaldataset$measuredJun15)
Finaldataset$measuredNov15 <- as.numeric(Finaldataset$measuredNov15)
Finaldataset$measuredJan16 <- as.numeric(Finaldataset$measuredJan16)
Finaldataset$measuredFeb16 <- as.numeric(Finaldataset$measuredFeb16)
```

```
#--- subset dataset with only X months
a <- Finaldataset[Finaldataset$measuredOct11 >0,]
b <- Finaldataset[Finaldataset$measuredJan12 >0,]
c <- Finaldataset[Finaldataset$measuredFeb12 >0,]
d <- Finaldataset[Finaldataset$measuredMar12 >0,]
e <- Finaldataset[Finaldataset$measuredApr12 >0,]
f <- Finaldataset[Finaldataset$measuredAug14 >0,]
h <- Finaldataset[Finaldataset$measuredSep14 >0,]
i <- Finaldataset[Finaldataset$measuredJun15 >0,]
j <- Finaldataset[Finaldataset$measuredNov15 >0,]
k <- Finaldataset[Finaldataset$measuredJan16 >0,]
g <- Finaldataset[Finaldataset$measuredFeb16 >0,]

#--- combine the various dfs and write CSV file
Finaldataset2 <- rbind(a,b,c,d,e,f,g,h,i,j,k)
write.csv(Finaldataset2, "XmonthsDf.csv", row.names=FALSE)
Finaldataset2 <- read.csv("XmonthsDf.csv", header = TRUE)
rm(a,b,c,d,e,f,g,h,i,j,k,l)


#--- find min and max for the dates
mins <-aggregate(Finaldataset2[ , c("date")], list(Finaldataset2$user_id) ,
                 function(x) min(as.character(x)) )
maxms <- aggregate(Finaldataset2[ , c("date")], list(Finaldataset2$user_id),
                   function(x) max(as.character(x)) )
Diff <- merge(maxms, mins, by="Group.1")

#--- create a loop to calculate the timespan
for(i in 1:nrow(Diff)) {
  Diff$Span[i] <- as.numeric(difftime(strptime(paste(Diff[i,2]),
                                               "%Y-%m-%d"),
                                      strptime(paste(Diff[i,3]),
                                               "%Y-%m-%d")))
}
#--- add names and round digits for the timespan variable
Diff$Span <- round(Diff$Span, digits = 0)
names(Diff)[1:4] <- c("user_id", "LastMeasure",
                      "FirstMeasure","DiffDates")

#--- merge initial training set with the max/min results
Finaldataset2 <- merge(Finaldataset2,Diff, by="user_id")
rm(maxms,mins)
rm(Diff)
#--- create a variable for loudness SD per user
library(plyr)
LoudnessSd <- ddply(Finaldataset2, .(user_id),
                  summarise, dBmean=mean(loudness), dBsd=sd(loudness))

#--- merge the two datasets together
Finaldataset2 <- merge(Finaldataset2, LoudnessSd, by="user_id")
rm(LoudnessSd)
```

```
#--- intermediary CSV file
write.csv(Finaldataset2, "RottUsers.csv", row.names=FALSE)
Finaldataset2 <- read.csv("RottUsers.csv", header = TRUE)

#--- include a latlong variable
Finaldataset2$latlong <- paste(round(Finaldataset2$lat, 2),
                              round(Finaldataset2$lng, 2), sep = ":")

#--- count the number of measurements per location per user
LatLongNumber <- ddply(Finaldataset2, c("user_id","latlong"),
                       summarize, random=sum(lng!="R analysis"))
LatLongNumberb <- ddply(LatLongNumber, c("user_id"),
                       summarize, locationsum=length(latlong))
Finaldataset2<- merge(Finaldataset2, LatLongNumberb, by="user_id")

#--- remove negligeable variables for further analysis
Finaldataset2$loudness<-NULL
Finaldataset2$loudness_index <- NULL
Finaldataset2$date <- NULL
Finaldataset2$lat<- NULL
Finaldataset2$lng <- NULL
Finaldataset2$latlong <- NULL
Finaldataset2$month <- NULL

#--- write CSV file for Finaldataset with location sums
write.csv(Finaldataset2, "IntermedRottModel.csv", row.names = FALSE)
Finaldataset2 <- read.csv("IntermedRottModel.csv", header = TRUE)
rm(LatLongNumberb)
rm(LatLongNumber)

#--- get one observation by user
Finaldataset2$user_id <- as.factor(Finaldataset2$user_id)
Predmodel <- unique(Finaldataset2[c("user_id","measuredOct11",
                           "measuredJan12","measuredFeb12", "measuredMar12",
                           "measuredApr12","measuredAug14", "measuredSep14",
                           "measuredJun15","measuredNov15","measuredJan16",
                           "measuredFeb16","LastMeasure","FirstMeasure",
                           "DiffDates","dBmean","dBsd","locationsum")])

#--- calculate the average/sum for each month per user
Newpredmodel1 <- aggregate(Predmodel$measuredOct11,
                          by=list(Predmodel$user_id), FUN=mean)
Newpredmodel2 <- aggregate(Predmodel$measuredJan12,
                          by=list(Predmodel$user_id), FUN=mean)
Newpredmodel3 <- aggregate(Predmodel$measuredFeb12,
                          by=list(Predmodel$user_id), FUN=mean)
Newpredmodel4 <- aggregate(Predmodel$measuredMar12,
                          by=list(Predmodel$user_id), FUN=mean)
Newpredmodel5 <- aggregate(Predmodel$measuredApr12,
                          by=list(Predmodel$user_id), FUN=mean)
Newpredmodel6 <- aggregate(Predmodel$measuredAug14,
                          by=list(Predmodel$user_id), FUN=mean)
Newpredmodel7 <- aggregate(Predmodel$measuredSep14,
```

```
                           by=list(Predmodel$user_id), FUN=mean)
Newpredmodel8 <- aggregate(Predmodel$measuredJun15,
                           by=list(Predmodel$user_id), FUN=mean)
Newpredmodel9 <- aggregate(Predmodel$measuredNov15,
                           by=list(Predmodel$user_id), FUN=mean)
Newpredmodel10 <- aggregate(Predmodel$measuredJan16,
                          by=list(Predmodel$user_id), FUN=mean)
Newpredmodel11 <- aggregate(Predmodel$measuredFeb16,
                            by=list(Predmodel$user_id), FUN=mean)
Predmodel$DiffDates <- as.numeric(Predmodel$DiffDates)

#--- remove user_id out of other month dataframes
User_id <- as.data.frame(Newpredmodel1$Group.1)
Newpredmodel1$Group.1 <- NULL
Newpredmodel2$Group.1 <- NULL
Newpredmodel3$Group.1 <- NULL
Newpredmodel4$Group.1 <- NULL
Newpredmodel5$Group.1 <- NULL
Newpredmodel6$Group.1 <- NULL
Newpredmodel7$Group.1 <- NULL
Newpredmodel8$Group.1 <- NULL
Newpredmodel9$Group.1 <- NULL
Newpredmodel10$Group.1 <- NULL
Newpredmodel11$Group.1 <- NULL

#--- merge the months together
Modeldf <- as.data.frame(cbind(Newpredmodel1, Newpredmodel2,
                               Newpredmodel3, Newpredmodel4,
                               Newpredmodel5, Newpredmodel6,
                               Newpredmodel7, Newpredmodel8,
                               Newpredmodel9, Newpredmodel10,
                               Newpredmodel11))
names(Modeldf) <- c("measuredOct11", "measuredJan12",
                    "measuredFeb12", "measuredMar12",
                    "measuredApr12", "measuredAug14",
                    "measuredSep14", "measuredJun15",
                    "measuredNov15", "measuredJan16",
                    "measuredFeb16")

#--- remove individual month dataframes
rm(Newpredmodel1)
rm(Newpredmodel2)
rm(Newpredmodel3)
rm(Newpredmodel4)
rm(Newpredmodel5)
rm(Newpredmodel6)
rm(Newpredmodel7)
rm(Newpredmodel8)
rm(Newpredmodel9)
rm(Newpredmodel10)
rm(Newpredmodel11)

#--- put 1 one into every cell that has a non-zero value in it
```

```
Modeldf$measuredOct11 <- ifelse(Modeldf$measuredOct11 >0, 1, 0)
Modeldf$measuredJan12 <- ifelse(Modeldf$measuredFeb12 >0, 1, 0)
Modeldf$measuredFeb12 <- ifelse(Modeldf$measuredApr12 >0, 1, 0)
Modeldf$measuredMar12 <- ifelse(Modeldf$measuredJan12 >0, 1, 0)
Modeldf$measuredApr12 <- ifelse(Modeldf$measuredMar12 >0, 1, 0)
Modeldf$measuredAug14 <- ifelse(Modeldf$measuredAug14 >0, 1, 0)
Modeldf$measuredSep14 <- ifelse(Modeldf$measuredSep14 >0, 1, 0)
Modeldf$measuredJun15 <- ifelse(Modeldf$measuredJun15 >0, 1, 0)
Modeldf$measuredNov15 <- ifelse(Modeldf$measuredNov15 >0, 1, 0)
Modeldf$measuredJan16 <- ifelse(Modeldf$measuredJan16 >0, 1, 0)
Modeldf$measuredFeb16 <- ifelse(Modeldf$measuredFeb16 >0, 1, 0)

#--- add users to Modeldf
Model2 <- cbind(Modeldf, User_id)
colnames(Model2) <- c("measuredOct11", "measuredJan12",
                      "measuredFeb12", "measuredMar12",
                      "measuredApr12", "measuredAug14",
                      "measuredSep14", "measuredJun15",
                      "measuredNov15", "measuredJan16",
                      "measuredFeb16", "user_id")
Predmodel2 <- unique(Predmodel[c("user_id","LastMeasure",
                                 "FirstMeasure","DiffDates","dBmean",
                                 "dBsd","locationsum")])
rm(Predmodel)
rm(Modeldf)

#--- merge Model2 and PredModel dfs, but
UserModel5 <- merge(Model2, Predmodel2, by="user_id")
UserModel5 <- unique(UserModel5, incomparables = FALSE)
rm(Model2)
rm(Predmodel2)
rm(User_id)

#write CSV file before target variable
#write.csv(UserModel5, "RottPred0706.csv", row.names=FALSE)
UserModel5 <- read.csv("RottPred0706.csv", header = TRUE)

#--- create a target variable
#--- read initial dataset
Finaldataset <- read.csv("Rotterdam0706.csv", header = TRUE)

#--- use the same analysis for the target variable as with the X variables
#--- check whether someone has measured in May, April or March 2016
Finaldataset$measuredMar16<- ifelse(Finaldataset$month == "2016-03", "1", "0")
Finaldataset$measuredApr16<-  ifelse(Finaldataset$month == "2016-04", "1", "0")
Finaldataset$measuredMay16<- ifelse(Finaldataset$month == "2016-05", "1", "0")
Finaldataset$measuredMar16 <- as.numeric(Finaldataset$measuredMar16)
Finaldataset$measuredApr16 <-  as.numeric(Finaldataset$measuredApr16)
Finaldataset$measuredMay16 <- as.numeric(Finaldataset$measuredMay16)

#--- remove unnecessary columns
Finaldataset$loudness_index <- NULL
Finaldataset$lat<-NULL
```

```
Finaldataset$lng<-NULL
Finaldataset$loudness<-NULL
Finaldataset$date<-NULL
Finaldataset$month<-NULL

#--- get one observation by user
FinaldatasetUnique <- unique(Finaldataset[c("user_id","measuredMar16",
                                            "measuredApr16","measuredMay16")])

#--- calculate the average/sum for each month per user
Newpredmodel1 <- aggregate(FinaldatasetUnique$measuredMar16,
                           by=list(FinaldatasetUnique$user_id), FUN=mean)
Newpredmodel2 <- aggregate(FinaldatasetUnique$measuredApr16,
                           by=list(FinaldatasetUnique$user_id), FUN=mean)
Newpredmodel3 <- aggregate(FinaldatasetUnique$measuredMay16,
                           by=list(FinaldatasetUnique$user_id), FUN=mean)

#--- remove user_id from individual dataframes and create one for it
User_id <- as.data.frame(Newpredmodel1$Group.1)
Newpredmodel1$Group.1 <- NULL
Newpredmodel2$Group.1 <- NULL
Newpredmodel3$Group.1 <- NULL

#--- add the y month into one dataframe
Modeldf2 <- as.data.frame(cbind(Newpredmodel1,
                                Newpredmodel2,Newpredmodel3))
names(Modeldf2) <- c("measuredMar16","measuredApr16","measuredMay16")
rm(Newpredmodel1)
rm(Newpredmodel2)
rm(Newpredmodel3)

#--- put 1 one into every cell that has a non-zero value in it
Modeldf2$measuredMar16 <- ifelse(Modeldf2$measuredMar16 >0, 1, 0)
Modeldf2$measuredApr16 <- ifelse(Modeldf2$measuredApr16 >0, 1, 0)
Modeldf2$measuredMay16 <- ifelse(Modeldf2$measuredMay16 >0, 1, 0)

#--- create a variable that calculates the sum per user of measured months
Modeldf2$measuredsum <- rowSums(Modeldf2)

#--- create target variable that checks whether a user is likely to continue
Modeldf2$measuredFut <- ifelse(Modeldf2$measuredsum >=1, 1, 0)

#--- combine user_id and modeldf and add names
Model3 <- cbind(Modeldf2, User_id)
colnames(Model3)<- c("measuredMar16","measuredApr16",
                     "measuredMay16","measuredsum",
                     "measuredFut","user_id")

#--- clean environment
rm(Modeldf2)
rm(User_id)
rm(FinaldatasetUnique)
```

```
#--- CSV file for target variable df
#write.csv(Model3, "RottUsersTV0706.csv", row.names=FALSE)
Model3<- read.csv("RottUsersTV0706.csv", header = TRUE)

#--- merge target variable df with predictors df
UserModel6 <- merge(UserModel5, Model3,by="user_id")

#--- CSV file for entire df
#write.csv(UserModel6, "RottFinal0706.csv", row.names=FALSE)
UserModel6<- read.csv("RottFinal0706.csv", header = TRUE)

#--- combine Tree2 and UserModel6
library(plyr)
Tree2 <- read.csv("Tree20706.csv", header = TRUE)
FinalTree <- rbind.fill(UserModel6, Tree2)

FinalTree$user_id<-as.numeric(FinalTree$user_id)

#--- create csv file for final Model dataset
#write.csv(FinalTree, "FinalTree0706.csv", row.names=FALSE)
FinalTree <- read.csv("FinalTree0706.csv", header = TRUE)


#--- create training and test set
colSums(is.na(FinalTrain))
colSums(is.na(FinalTest))
FinalTree$user_id[is.na(FinalTree$user_id)] <- mean(FinalTree$user_id,na.rm= TRUE)
nobservations <- nrow(FinalTree)
pTest <- 0.3
nTest <- round (0.3 * nobservations)
FinalTrain<- FinalTree [(1:( nobservations - nTest )) ,]
FinalTest <- FinalTree [ -(1:( nobservations - nTest )) ,]

testObservations <- sample (1: nobservations ,nTest ,
                            replace = FALSE )
testObservations <- sort ( testObservations )
FinalTrain <- FinalTree [- testObservations ,]
FinalTest <- FinalTree [ testObservations ,]

#--- create csv files for train and test set
write.csv(FinalTrain, "finalTrainset.csv", row.names=FALSE)
write.csv(FinalTest, "finalTestset.csv", row.names=FALSE)
FinalTest <- read.csv("finalTestset.csv", header = TRUE)
FinalTrain <- read.csv("finalTrainset.csv", header = TRUE)
```

Appendix G Predictive Model Evaluation R script

```
#--- Install necessary packages
library(aod)
library(Rcpp)
library(ggplot2)
library(rpart)
```

```
library(rpart.plot)
library(arules)
library(SnowballC)
library(e1071)
library(ROCR)
library(pROC)
library(rattle)
library(caret)
library(mlbench)
library(sm)

#--- read entire dataset, training and test set
FinalTree <- read.csv("FinalTree0706.csv", header = TRUE)
FinalTest <- read.csv("finalTestset.csv", header = TRUE)
FinalTrain <- read.csv("finalTrainset.csv", header = TRUE)

#--- check the share of positive outcomes
table(FinalTrain$measuredFut)

# Check missing values (alternatively, use complete.cases(dsHouses))
# No NAs present
colSums(is.na(FinalTree))
colSums(is.na(FinalTest))

#--- replace NA's in dBsd
FinalTrain$dBsd[is.na(FinalTrain$dBsd)] <- sample(1:11,
                                size=sum(is.na(FinalTrain$dBsd)), replace = TRUE)
FinalTest$dBsd[is.na(FinalTest$dBsd)] <- sample(1:11,
                                size=sum(is.na(FinalTest$dBsd)), replace = TRUE)

#--- Model 1: Estimate Logistic model:specify and estimate the model
# Specify the model
Mylogit <- measuredFut ~ dBmean + locationsum + DiffDates + dBsd

# Estimate the model with logistic regression
NTLogitModel <- glm(Mylogit, data=FinalTree, binomial(link="logit"))

# Summary of the results
summary(NTLogitModel)

#--- Logistic regression analysis: prediction
# Holdout sample prediction (use previously defined newHouses)
# The additional option se.fit gives the standard errors of the
# predicted values
predict(NTLogitModel, FinalTest, type=c("response")) # Probability

#--- Confusion matrix: simple example based on result logistic regression
# Retrieve original and predicted values of the dependent from the object
# NTLogitModel
yvalue <- FinalTest$measuredFut
ypred  <- predict.glm(NTLogitModel, FinalTest, type=c("response")) # Probability

# Turn predicted probability into predicted classification
```

```
ypred  <- as.numeric(ypred > 0.5)

# Confusion matrix
table(ypred,yvalue)
table(Predicted = ypred, Observed = yvalue)

# Accurracy
mean(yvalue == ypred) # Accuracy
sum(yvalue == 0)      # negatives
sum(yvalue == 1)      # positives

sum(ypred == 0)
sum(ypred == 1)

 #--- AUC
AUC <- performance(prediction(ypred, yvalue),
             measure = "auc")@y.values[[1]]

# The function accRates calculates the false positive rate
# and the true positive rate for given (globally defined)
# vectors of observed (yvalue) and predicted (ypred)
# classifications.
accRates <- function(s){
  FPR  <- sum((ypred > s)*(yvalue==0))/sum(yvalue==0)
  TPR  <- sum((ypred > s)*(yvalue==1))/sum(yvalue==1)
  return(c(FPR = FPR, TPR = TPR)) }

# Retrieve the original predicted probabilities
ypred  <- predict.glm(NTLogitModel, FinalTest, type="response")

# Examples
accRates(0.5)
accRates(0.2)

# Determine the associated classification for multiple threshold values.
# Before, the function accRates needs to be 'vectorized
# in order to be prepared for the multiple evaluations.
accRatesFun <- Vectorize(accRates)
accRatesVal <- accRatesFun(seq(0, 1, by = 0.005))

#--- Model 2: Decision tree with measuringCrit classifier
 rpart1 <-rpart(measuredFut~dBmean + locationsum
 + DiffDates + dBsd, data=FinalTrain, method = "class",
 parms = list(split = "information"), control=rpart.control(maxdepth=8))

#--- Plot the tree with the number of observations
 rpart.plot(rpart1,
                box.col =c("pink","palegreen3") [rpart1$frame$yval], extra= 1)

# Plot tree with probabilities and colors
 rpart.plot(rpart1,
                box.col =c("pink","palegreen3") [rpart1$frame$yval], extra = 4)
```

```
# Plot the tree as a set of rules
  asRules(rpart1)

  #--- Kernel plot
  dnsty <- density(yvalue)
  plot(dnsty, main="Kernel plot of class distribution",
       xlab="Probabilities", las = 1)
  polygon(dnsty, col = "orange")

#--- class distribution graph version 2
#1---create value labels
meas.f <- factor(yvaluetree, levels= c(1,0),
                  labels = c("Continues", "Stops"))

#2---plot densities
sm.density.compare(ypredtree, meas.f, xlab="Probabilities",
                   lwd=3.5)
title(main="Class Distribution")

#3---Add a legend (the color numbers start from 2 and go up)
legend("topright", levels(meas.f),
       fill=2+(0:nlevels(meas.f)), cex = 0.7)

#--- Classification trees: use tree for prediction
# Prediction: in sample probability prediction (holdout set)
 ypredtree <- predict(rpart1,FinalTest, type = "prob")[,2]
 yvaluetree <- FinalTest$measuredFut

 # Prediction: in sample class prediction (holdout set)
 # Classification probabilities
 ypredtree2 <- predict(rpart1,FinalTest, type = "class")
 ypredtree2 <- as.numeric(ypredtree2)
 yvaluetree <- FinalTest$measuredFut

 # Confusion matrix and Accuracy for Type class and type Prob
 xtab <- table(ypredtree2,yvaluetree)
 confusionMatrix(xtab)
 table(Predicted = ypredtree, Observed = yvaluetree)

 # Accurracy
 mean(yvalue == ypredtree2) # Accuracy
 sum(yvalue == 0)      # negatives
 sum(yvalue == 1)      # positives

 sum(ypredtree2 == 0)
 sum(ypredtree2 == 1)

 # True negatives: negatives that are predicted negative
 sum( (ypredtree2==0)*(yvalue==0))

 # False positives: negatives that are predicted positive
 sum( (ypredtree2==1)*(yvalue==0))
```

```
# False negatives: positives that are predicted negative
sum( (ypredtree2==0)*(yvalue==1))

# True positives: positives that are predicted positive
sum( (ypredtree2==1)*(yvalue==1))

# False positive rate, FPR (1-specificity)
FPR <- sum( (ypredtree2==1)*(yvalue==0) )/sum(yvalue==0)

# True positive rate, TPR (sensitivity)
TPR <- sum( (ypredtree2==1)*(yvalue==1) )/sum(yvalue==1)

 #--- AUC
AUCtree <- performance(prediction(ypredtree, yvaluetree), measure = "auc")@y.values[[1]]

#--- Confusion matrix: define function
#--- The function accRates calculates the false positive rate and the true positive rate
#--- for given (globally defined) vectors of observed (yvalue)
#--- and predicted (ypred) classifications.
accRates <- function(s){
  FPR  <- sum((ypredtree > s)*(yvaluetree==0))/sum(yvaluetree==0)
  TPR  <- sum((ypredtree > s)*(yvaluetree==1))/sum(yvaluetree==1)
  return(c(FPR = FPR, TPR = TPR))}

# Examples
accRates(0.5)

# For trees, also determine the associated classification for multiple threshold values.
# Results are stored in a separate vector
accRatesFunTree <- Vectorize(accRates)
accRatesValTree <- accRatesFunTree(seq(0,1,by=.005))

#-------------------------------------------------------------------------------
# ROC curve: Tree model and logit model based on in sample predictions
#-------------------------------------------------------------------------------

# Simple plot (first, the logit model results; then the
# tree results and finally the diagonal are depicted)
plot(accRatesVal[1,],accRatesVal[2,],
     col="red",lwd=2,type="l",
     xlab="False positive rate",
     ylab="True positive rate")
lines(accRatesValTree[1,],accRatesValTree[2,],
      col="blue",lwd=2,type="l")
lines(c(0,1),c(0,1))
```

Appendix H Summary Logistical Regression Model

```
Call:
glm(formula = Mylogit, family = binomial(link = "logit"), data = FinalTree)

Deviance Residuals:
```

```
    Min       1Q   Median       3Q      Max
-4.9579  -0.9195  -0.6979   1.0703   3.2383

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -9.12374    1.19234  -7.652 1.98e-14 ***
dBmean       0.12179    0.01560   7.805 5.95e-15 ***
locationsum -0.22319    0.09723  -2.295   0.0217 *
DiffDates    0.10506    0.02194   4.790 1.67e-06 ***
dBsd         0.10110    0.01427   7.086 1.38e-12 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05  0.1  1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1399.4  on 1019  degrees of freedom
Residual deviance: 1240.8  on 1015  degrees of freedom
  (1976 observations deleted due to missingness)
AIC: 1250.8

Number of Fisher Scoring iterations: 5
```

Appendix I Model Confusion Matrix Classification

```
Confusion Matrix and Statistics

          yvaluetree
ypredtree2   0   1
         0 339 104
         1  90 366

               Accuracy : 0.7842
                 95% CI : (0.7558, 0.8107)
    No Information Rate : 0.5228
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.5681
 Mcnemar's Test P-Value : 0.3506

            Sensitivity : 0.7902
            Specificity : 0.7787
         Pos Pred Value : 0.7652
         Neg Pred Value : 0.8026
             Prevalence : 0.4772
         Detection Rate : 0.3771
   Detection Prevalence : 0.4928
      Balanced Accuracy : 0.7845

       'Positive' Class : 0
```

Appendix J Model Confusion Matrix Probabilities

```
                   Observed
Predicted             0   1
  0.210526315789474   4   3
  0.249304911955514 329 100
  0.25                6   1
  0.730994152046784  17  60
  0.788690476190476  64 267
  0.8828125           9  39
```

Appendix K AsRules Function R

```
 asRules(rpart1)

 Rule number: 5 [measuredFut=1 cover=128 (6%) prob=0.88]
   dBmean< 73.85
   DiffDates>=1.5

 Rule number: 15 [measuredFut=1 cover=673 (32%) prob=0.79]
   dBmean>=73.85
   locationsum< 1.5
   dBmean< 80.38

 Rule number: 9 [measuredFut=1 cover=171 (8%) prob=0.73]
   dBmean< 73.85
   DiffDates< 1.5
   dBmean< 60.22

 Rule number: 8 [measuredFut=0 cover=1079 (51%) prob=0.25]
   dBmean< 73.85
   DiffDates< 1.5
   dBmean>=60.22

 Rule number: 6 [measuredFut=0 cover=27 (1%) prob=0.22]
   dBmean>=73.85
   locationsum>=1.5

 Rule number: 14 [measuredFut=0 cover=19 (1%) prob=0.21]
   dBmean>=73.85
   locationsum< 1.5
   dBmean>=80.38
```