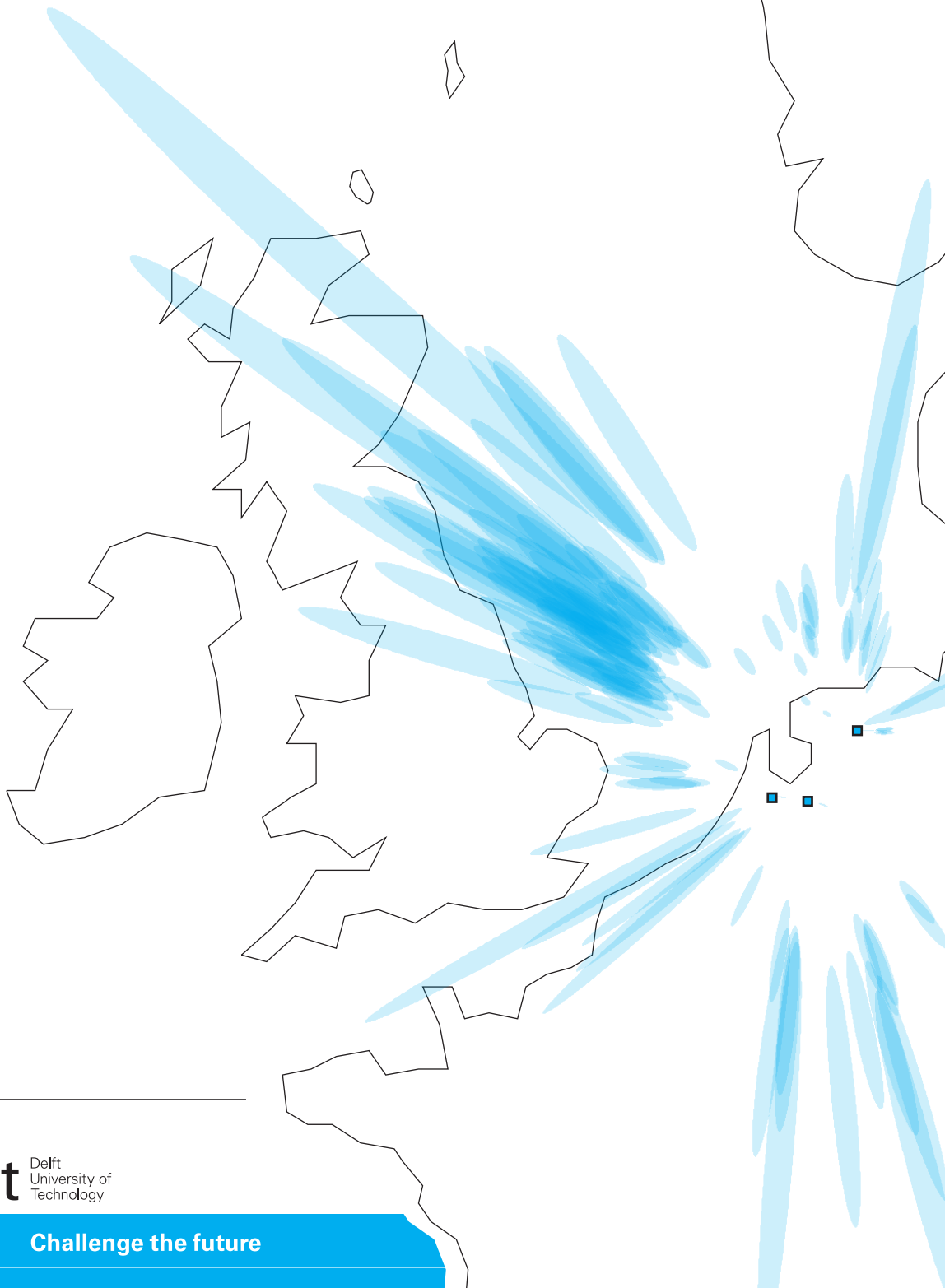
Infrasound source location

An automated method to estimate the location of infrasound sources
as detected by microbarometer array recordings
in the Netherlands over 2007

Sven A. Briels



TU Delft Delft University of Technology

Challenge the future

Cover illustration: 2σ error ellipses for the best unbiased estimate of source locations, estimated for events detected by De Bilt infrasound array (DBN), Deelen infrasound array (DIA) and Exloo infrasound array (EXL) in the period from January 1 until September 3, 2007.

MSc. Geomatics Graduation Thesis

Infrasound source location

An automated method to estimate the location of infrasound sources
as detected by microbarometer array recordings
in the Netherlands over 2007

Graduation thesis by Sven Andreas Briels, January 2010
Contact email address: s.a.briels@gmx.net

Graduation Professor (DUT): Prof. dr. Dick G. Simons
Supervisor (DUT/KNMI): Dr. Läslo G. Evers
Internship supervisor (CEA): Dr. Alexis Le Pichon
Co-reader (DUT): Dr. ir. Hans van der Marel

Delft University of Technology (DUT)
Department of Earth Observation and Space Systems
Acoustic Remote Sensing
Faculty of Aerospace Engineering
The Netherlands

Koninklijk Nederlands Meteorologisch Instituut (KNMI)
Department Climate and Seismology
Division Seismology
Address P.O. Box 201
3730 AE De Bilt
The Netherlands

Commissariat à l'Énergie Atomique (CEA) DAM/DIF
F-91297 Arpajon Cedex
France

Abstract

Infrasound source location is very closely related to the two main applications in infrasound research. The current major drive for infrasound research is the Comprehensive Nuclear-Test-Ban Treaty (CTBT) which aims at detecting and locating nuclear tests. Thus, source location itself is the goal. For the major scientific application, use of infrasound to study the state of the atmosphere, the location of the source is needed as input. The source location and the state of the atmosphere determine what will be recorded. Roughly said, once the source location is known, one can start studying the influence of the atmosphere between the source and receiver. Although multiple examples can be given where the source location is estimated for a single clearly distinguished event (Evers & Haak, 2007) (Brown, 2002) there is no standard procedure to locate the hundreds of indistinguishable events an infrasound array detects every day. Therefore, an automated method to estimate the location of infrasound sources is developed. This method is applied to recordings of three infrasound arrays in the Netherlands operated by the Koninklijk Nederlands Meteorologisch Instituut (KNMI) and three additional arrays located in Germany and France. The estimated source locations are validated in twofold. First, by a comparison against locations made by independent infrasound arrays and processing as performed at the Commissariat à l'Énergie Atomique (CEA). Second, the number of located events is compared against the number of events that will be generated for random data.

A method was developed around a three step approach. First the recordings of each infrasound array are processed separately. This results in a list of infrasound detections for each array. The detection time and back azimuth (i.e. the direction of the signal in the horizontal plane, measured clockwise from north) of the detections are used in step two to combine detections at multiple arrays that could originate from the same event into a single event. For each event the mean back azimuth is calculated from its detections at each array. In the final step the location of the events is estimated by crossing the mean back azimuths between all arrays. The location is described in latitude and longitude and is estimated for a spherical earth.

Extensive filtering is applied to clean the data from as much noise and signals from local sources as possible. Filters are applied on frequency, apparent velocity, Fisher ratio, back azimuth and celerity. Due to the geometry of the De Bilt infrasound array (DBN) and EXL detections were limited to a frequency band of 1 to 10 Hz, coinciding with the frequency band which is expected for manmade sources. 92% of all detections were filtered out based on the set criteria.

Avoiding the combination of unrelated detections into false events was found to be a major subject. These events are caused by detections which have by chance a back azimuth and detection time that allow them to originate from the same event, but are in fact unrelated. These infrasound events did not occur in reality and are therefore described as false. The major cause for these false events is the large number of detections from nearby sources, which are detected only locally. It is expected that the majority of all detections over the used frequency band (1-10Hz) can be ascribed to local sources such as traffic at tunnels and overpasses, local industry or close-by military practice. The detections at the DBN infrasound array show this clearly as half of all detections were received over 4 small azimuth bands pointing to overpasses of roads and small tunnels. Also at DIA, EXL and the IGADE infrasound array near Bremen, Germany, large numbers of detections were found over small back azimuth bands. Exact local sources are often difficult to find from maps and aerial imagery and their signals are usually indistinguishable from signals that travelled from a distant source. At DBN and

IGADE signals from local sources were found to be so dominating that it was chosen to remove all detections over certain back azimuths.

The developed procedure was applied on detections over a frequency band of 1 to 10 Hz made by three infrasound arrays located in the Netherlands and to the detections of three more arrays located in France (1 array) and Germany (2 arrays). Locations of the events were found dominantly on the western part of the North Sea. A comparison with the locations of gas flaring showed a significant difference in location. It is therefore expected that most detections originate from a different source. The events could be related to military jet exercise (sonic booms) or the detonation of explosives (e.g. clearing of naval mines), but this could not be verified due to the lack of ground truth data. Sources that were found over land showed a strong correlation with military presence (e.g. practice grounds and military airfields). Almost all events occurred during weekdays (99%) and between 7 am and 10 pm local time (98%), which strongly suggests that signals from manmade sources are dominant over the used frequency band. 16% of the events were recorded for over 10 minutes, which suggests that the source emitted multiple signals over time. For the combination of the three KNMI infrasound arrays 141 events were located over 160 days. These events were detected by all 3 of the arrays. For the combination of 6 arrays 57 events were located which were detected by 5 or 6 of the arrays.

The estimated infrasound source locations are validated in twofold. First, the locations as found by the three KNMI arrays over the North Sea are compared against locations made by independent infrasound arrays and processing as performed at CEA. The event location map from CEA shows a comparable source region as the KNMI event map. It was possible to indicate 9 events which appear in both lists within 2 minutes and would generate a back azimuth at the KNMI arrays within 10 degrees of each other. These events form 21% of the CEA events and about 12% of the KNMI events. Second, the number of located events is compared against the number of events that will be generated for random data. As infrasound events are found for only a very small portion of all detections, random data can be used to evaluate the number of false events that is generated by the applied method. The random data is generated such that the time and azimuth distribution of the original detection data is preserved. By inserting random data into the location process it was found that 6% of the detections made by all three KNMI arrays can be contributed to the combination of unrelated detections. For all six arrays none of the detections made by six arrays and 18% of the detections made by five out of six arrays can be contributed to false events.

Generally it can be concluded that it proved possible to develop an automated method to estimate the geographic location of infrasound sources. The developed method was succesfully applied to infrasound recordings of infrasound arrays in the Netherlands, Germany and France. Event location allows for better identification of infrasound sources and can be used to distinguish detections of far field sources from local sources. Using the source location one can potentially evaluate signal travel paths through the atmosphere, which will contribute to the use of infrasound for atmospheric studies.

Preface

There is more to sound than what reaches the ear. Distant rumble from the oceans, deep hums from the aurora borealis and sudden cracks from exploding meteors are heard if one could expand human hearing to frequencies below the 20Hz boundary that we are used to. When the challenge to convert from listening to seeking the source locations was set to me I accepted, and I found myself exploring this unheard world as a master's thesis.

A challenge always starts with great expectations, what was there to be found? Will some surprising sources turn up? And although still far away in research, could the results be used for something useful, can we sound the atmosphere, complement measurements from weather balloons and measure even beyond the altitudes balloons can reach? But then already after a few days, as for any modern science it seems, the main challenge is found blinking at the computer screen and a long road winding through if then and else dooms up.

One of the main values I found during this research was the value of sharing, showing and writing down results, thoughts and methods. At these moments of self reflection, forcing yourself to explain every detail and discussing the results with others the results are tested and often give away the signs on which improvements can be made and flaws have to be fixed. The most obvious example of this are the words before you, of which every single one needed thought and was often changed more than once. Many of these words forced changes in filters, redrawing of figures and debugging parts of code, flaws which were uncovered only when the explaining words turned into unexplainable twists. At pub talks, birthdays and the occasional colloquium presentation I found myself explaining my research, sharpening the words until the story became understandable for anyone who pointed the question. These conversations were often important for the drive to continue and regaining the sense that this research would be useful, more than to obtain a grade for the final test of my studies. I would like to thank all of those who have asked these questions, I often found them most welcome.

On the topic of sharing information, I much enjoyed the exchange with CEA near Paris. The change of scenery, comments on my work and simple differences in the point of view sharpened my senses. The two months I worked there were probably the two most productive months of the past year. I much enjoyed living in Paris and eagerly explored it. I had a good time with colleagues by day and with friends and family in the weekends and evenings.

Now after a year of research and many lines of code the path comes to an end. It is at today's marvelous snow-clad winter day that I find myself typing for me the final, and for the reader the first lines of this adventure. I would like to thank anyone who has helped me in the process. Special thanks to Läslo Evers for letting his enthusiasm get me into this adventure, Jelle de Plaa for his help on the array processing software and variance estimation, the NERIES program for supporting the exchange with CEA, Jocelyn Guilbert for his great support in practical manners and both Jocelyn and especially Alexis le Pichon for their interest in my work and useful comments and questions during my stay at CEA.

Sven Andreas Briels
Delft, December 2009

Contents

Abstract iii

Preface v

1 Introduction 1
1.1 Motivation and research objective . . . 1
1.2 Research methodology . . . 2
1.3 Thesis outline . . . 2

2 Introduction to infrasound 3

3 Processing Method 9
3.1 Array processing . . . 10
3.1.1 Signal direction and receiver location . . . 10
3.1.2 Frequency-wave number analysis . . . 12
3.1.3 Fisher ratio . . . 13
3.1.4 Estimate variance of back azimuth and apparent velocity . . . 14
3.1.5 Array processing block scheme . . . 16
3.2 Combine multiple arrays . . . 17
3.2.1 Select common events between arrays . . . 17
3.2.2 Group events . . . 20
3.3 Estimate source location . . . 21
3.3.1 Cross bearing on the planar surface . . . 22
3.3.2 Cross bearing on the sphere . . . 25
3.4 Applied filters . . . 29
3.4.1 Frequency range determined by array specifications . . . 31
3.4.2 Frequency range determined by the source of interest . . . 37
3.4.3 Apparent velocity . . . 38
3.4.4 Fisher ratio . . . 40
3.4.5 Back azimuth . . . 42
3.4.6 Celerity . . . 45

4 Results KNMI Infrasound arrays 47
4.1 Source location . . . 47
4.1.1 Location from two arrays . . . 47
4.1.2 Location from first detections . . . 47
4.1.3 Location from three arrays and multiple detections per array . . . 50
4.2 Statistics on results . . . 52
4.2.1 Back azimuth . . . 52
4.2.2 Time and date . . . 52
4.2.3 Signal frequency . . . 52
4.2.4 Detections per event . . . 54
4.2.5 Signal duration . . . 54
4.2.6 Celerity . . . 56
4.3 Origin of sources . . . 59

4.3.1 Sources over land . . . 59
4.3.2 Sources over sea . . . 59

5 Results six infrasound arrays **63**
5.1 Detections per array . . . 63
5.1.1 The PMCC correlator . . . 63
5.1.2 Detection characteristics . . . 65
5.1.3 Additional filter on consistency and back azimuth . . . 67
5.2 Source location . . . 70
5.3 Statistics on results . . . 73
5.3.1 Back azimuth . . . 73
5.3.2 Time and date . . . 73
5.3.3 Detections per event . . . 74
5.3.4 Signal duration . . . 74

6 Validation **77**
6.1 Comparison of KNMI locations with CEA locations . . . 77
6.2 Events caused by false combination of events . . . 79

7 Conclusions and recommendations **85**

Bibliography **90**

List of Tables

2.1 Number of elements and aperture of the infrasound arrays at DBN DIA EXL 3
2.2 Frequency of known infrasound sources 4

3.1 Tolerances on back azimuth and celerity 18
3.2 Distance differences between WGS84, spherical and plane surface distance 21
3.3 Percentage of detections removed by filters 29
3.4 Theoretical usable frequency boundaries for DBN, DIA, EXL 34
3.5 Applied frequency range for DBN, DIA, EXL 37
3.6 False detection probability as a function of the Fisher threshold 41
3.7 Missed detection probability as a function of snr and Fisher threshold 41
3.8 Filtered back azimuths for local sources DBN 43
3.9 Celerity regimes as a function of atmospheric layer 46
3.10 Celerity regimes for fast and conventional stratopause ducted waves 46

4.1 Standard deviation and 95 % error for DBN, DIA, EXL 51
4.2 The number of events counted for DBN, DIA, EXL 54
4.3 Description of military presence found close to infrasound events 61

5.1 Location, elements and aperture of DBN, DIA, EXL, FLERS, IGADE, I26DE 63
5.2 Number of detections at DBN, DIA, EXL, FLERS, IGADE, I26DE 65
5.3 Back azimuth filter for IGADE 70
5.4 Standard deviation and 95 % error for DBN, DIA, EXL, FLERS, IGADE, I26DE 70
5.5 Number of events as detected by DBN, DIA, EXL, FLERS, IGADE, I26DE 74

6.1 False combinations of detections that propagate into events for DBN, DIA, EXL 80
6.2 False combinations of events that propagate into events for all six arrays, consistency more than 8 81
6.3 False events as a function of Fisher ratio for DBN, DIA and EXL 82
6.4 False events as a function of Fisher ratio for 6 infrasound arrays 83

List of Figures

2.1 Location of DBN, DIA, EXL . . . 5
2.2 Example of atmospheric signal propagation based on ray tracing. . . . 6
2.3 Back azimuth and frequency spectrum for detections at DBN in 2007 . . . 7
2.4 Infrasound propagation under variable atmosphere conditions . . . 8

3.1 Flow chart of processing . . . 9
3.2 Incident angle and back azimuth . . . 11
3.3 Slowness grid and fitted Lorenz profiles . . . 14
3.4 Block scheme software for array processing . . . 16
3.5 One month of back azimuths of detections at DBN, DIA, EXL . . . 17
3.6 Search window to match detections between multiple arrays . . . 18
3.7 Search windows in the 315 degree direction from DBN . . . 19
3.8 Cross bearing on the planar surface . . . 22
3.9 Cross bearing on the sphere . . . 25
3.10 Effect of filtering on back azimuth and frequency contect of detections at DBN, DIA and EXL 30
3.11 Fisher values over a slowness grid for a synthetic signal from directly above the array . . . 33
3.12 Central F-distribution, typical for noise and two non-central F-distributions with F=1.5 and F=3. 33
3.13 Probability of correct detection of a F=2 signal as a function of frequency for DBN, DIA, EXL 34
3.14 Array response and probability of correct detection of a 2.5 Hz signal for DBN, DIA and EXL . 35
3.15 Array response and probability of correct detection of a 10 Hz signal for DBN, DIA and EXL . 36
3.16 Logarithmic wind speed profile . . . 39
3.17 Wind speed distribution at de Bilt and Deelen in 2007 . . . 39
3.18 Histograms of apparent velocity of detections at DBN, DIA and EXL over 2007 . . . 39
3.19 Histogram of the Fisher ratio of detections at DBN, DIA and EXL over 2007 . . . 40
3.20 Missed event and false alarms from the Fisher threshold . . . 41
3.21 Polar histogram of back azimuth, local sources at De Bilt . . . 42
3.22 Frequency spectra of tunnels near DBN . . . 43
3.23 Histogram of back azimuth of detections at DBN, DIA, EXL over 2007 . . . 43
3.24 Polar histogram of back azimuth, local sources at Deelen . . . 44
3.25 Polar histogram of back azimuth, local sources at Exloo . . . 44

4.1 Best unbiased estimates of source locations from combinations of 2 arrays for DBN, DIA, EXL 48
4.2 Precision of locations constructed by first detections . . . 49
4.3 Histogram of the standard deviation on back azimuth . . . 50
4.4 Histogram of azimuthal adjustment of detections within an event . . . 51
4.5 Best unbiased estimates of source locations of events detected by DBN, DIA, EXL . . . 51
4.6 Histogram on azimuthal adjustment for location . . . 53
4.7 Distribution of events over time and date . . . 53
4.8 Frequency spectra of events detected by three arrays . . . 53
4.9 Number of detections grouped into one event . . . 55
4.10 Recorded signal duration histogram . . . 55
4.11 Signal duration map . . . 55
4.12 Estimate of minimal and maximal celerity . . . 57
4.13 Celerity estimation and errors . . . 57

4.14 Histograms on timing and location errors on the celerity estimates 58
4.15 Map of identified atmospheric returns 58
4.16 Map of military presence found close to infrasound events 60
4.17 Locations of infrasound sources compared with gas flare locations 62

5.1 Layout and location of FLERS, IGADE and I26DE 64
5.2 Example of the PMCC correlator 65
5.3 Signal duration of detections at FLERS, IGADE, I26DE 66
5.4 Histogram of frequency, apparent velocity and consistency for FLERS, IGADE and I26DE . . . 66
5.5 Standard deviation on back azimuth for DBN, EXL, FLERS, IGADE, I26DE 67
5.6 Polarhistogram of backazimuth, FLERS 68
5.7 Polarhistogram of backazimuth, IGADE 68
5.8 Polarhistogram of backazimuth, I26DE 69
5.9 Source locations FLERS, IGADE, I26DE without azimuth filter 69
5.10 histograms of the azimuthal adjustment towards the mean azimuth per event 71
5.11 Best unbiased estimate of source locations of events detected by six infrasound arrays 71
5.12 Infrasound event locations as a function of the detection array 72
5.13 Histogram on azimuthal adjustment for location 73
5.14 Distribution of events over time and date 73
5.15 Number of detections grouped into one event 75
5.16 Signal duration histogram 75
5.17 Signal duration map 75

6.1 Comparison between KNMI and CEA locations on the North Sea 78
6.2 Histogram of difference in event time and azimuth between KNMI and CEA events 79
6.3 Source locations found for random detections for DBN, DIA, EXL 80
6.4 Source locations found for random detections for DBN, DIA, EXL, FLERS, IGADE, I26DE . . 81
6.5 Graph of false events generated at different minimum F-ratio for the KNMI arrays 82
6.6 Graph of false events generated at different minimum F-ratio for all 6 arrays 83

List of Abbreviations

CEA Commissariat à l'Énergie Atomique

CTBT Comprehensive Nuclear-Test-Ban Treaty

DBN De Bilt infrasound array

DIA Deelen infrasound array

DMSP Defense Meteorological Satellite Program

EXL Exloo infrasound array

FLERS Flers infrasound array, France

IGADE IGADE infrasound array near Bremen, Germany

I26DE CTBT infrasound station 26, Germany

KNMI Koninklijk Nederlands Meteorologisch Instituut

MF Magnification Factor

NERIES Network of Research Infrastructures for European Seismology

OLS Operational Linescan System

PMCC the Progressive Multichannel Correlation method

RAF Royal Air Force

Chapter 1

Introduction

The three sections of this Chapter describe the motivation and research objective, the methodology used to reach this objective and the outline for the rest of the report.

1.1 Motivation and research objective

Infrasound source location is very closely related to the two main applications in infrasound research. The current major drive for infrasound research is the Comprehensive Nuclear-Test-Ban Treaty (CTBT) which aims at detecting and locating nuclear tests. Thus, source location itself is the goal. For the major scientific application, use of infrasound to study the state of the atmosphere, the location of the source is needed as input. The source location and the state of the atmosphere determine what will be recorded. Roughly said, once the source location is known, one can start studying the influence of the atmosphere between the source and receiver. Although multiple examples can be given where the source location is estimated for a single clearly distinguished event (Evers & Haak, 2007) (Brown, 2002) there is no standard procedure to locate the hundreds of indistinguishable events an infrasound array detects every day.

This need for source localization forms the research objective:

To develop an automated method to estimate the location of infrasound sources

At the start of this research only a single publication was present which presented source locations obtained by an automatic procedure (Le Pichon *et al.*, 2008). The infrasound receiver arrays in the Netherlands, which are operated by the Koninklijk Nederlands Meteorologisch Instituut (KNMI), were not included in the previous research. Although the above mentioned publication showed probable locations of sources over central Europe, there are no independent methods to check these locations. Ground truth data is available for only a handful events. Most infrasound arrays record numerous signals of which the source is unknown. Automated source location can give this information, and the resulting knowledge on origin and travel path can give an important impulse to current infrasound research. Therefore, it was decided that a method for automated estimation of source location should be developed at KNMI for their infrasound arrays. In contrary to the method as presented by Le Pichon et al. celerity (i.e. signal propagation velocity) should not be assumed to be constant. Instead the range of celerities observed for infrasound signals as introduced by Revelle & Whitaker (1996) should be transparently incorporated. The celerity of a signal gives information on the atmospheric layers through which the signal has travelled. Dependent on the state of the atmosphere and the distance from the source one or more arrivals can be detected. Each detection has a specific celerity determined by the atmospheric layer through which it travelled. In reverse, the celerities and number of arrivals can be used to infer the state of the atmosphere. Incorporating the proposed range of celerities allows us to use infrasound for meteorological purposes. These observations form two requirements on the method developed in the presented research:

- The method should be usable for KNMI infrasound arrays
- The method should be based on a range of physically possible celerities

As an infrasonic signal travels through the atmosphere, the state of the atmosphere influences what is detected at the infrasound array. This influence should be corrected for to obtain the true source location. An atmospheric model could be used to estimate this influence. Unfortunately, especially at high altitudes, the knowledge of the state of the atmosphere is limited. In fact, one would like to use the infrasound detections to improve the atmospheric model. This two sided relation deserves undivided attention. Therefore all results presented in this research are estimated while allowing atmospheric influence, but are not adjusted to an atmospheric model.

1.2 Research methodology

The presented research is carried out in three major steps. First per array recordings are processed separately. Second, the recordings at multiple arrays are related and these sets are used to get a source location. In the third step the locations are validated against locations obtained by independent research. These three steps are explained in detail below.

Process the per array recordings and apply appropriate filters on the results. The first part of the research is carried out at KNMI. KNMI operates five infrasound arrays in the Netherlands of which three offer continuous measurements from January 1st to September 3rd, 2007. The automated method for infrasound source location is developed using these recordings as a basis. This not only provides the input to develop and debug the method, it also ensures that the requirement is met which states that the method should be usable for KNMI infrasound arrays. The existing experience on processing and interpreting the generated data at KNMI is used to get detection lists for each array which contain among other signal properties the back azimuth (i.e. horizontal direction of arrival) and recording time of each detection. During and after processing several filters have to be applied to exclude noise and local sources as much as possible.

Combine detections of multiple arrays into events and estimate the source location for these events. In contrast with the first part which is based on existing knowledge and -partly- existing processing code the second step forms a complete new addition to the KNMI tool kit. Therefore, the results presented in this study will focus on source location as obtained in this step. In designing the method the second requirement is taken into account which states that the method should be based on a range of physically possible celerities. The resulting event list may contain multiple recordings of a single event which should then be grouped to a single event.

Validation of the located events The third part concentrates on validating both the occurrence of events and the obtained source locations. The probability that all events are true events is evaluated by processing a random data set. The source locations are compared to locations as determined independently by the previously mentioned method of Le Pichon *et al.* (2008). This part is carried out during an internship at Commissariat à l'Énergie Atomique (CEA) in France where these locations were obtained. Next to a straightforward comparison the opportunity is taken to use the location algorithm, as developed in part two, to three more infrasound arrays located in France and Germany. This exchange of knowledge and data was made possible by the NERIES program.

1.3 Thesis outline

For readers new in the infrasound research a short introduction to infrasound is given in Chapter 2. Chapter 3 presents the processing method which is illustrated by a flow chart in Figure 3.1. This Chapter covers the theory for the first two steps of the research methodology. Chapter 4 shows the resulting source locations when applying this method to the KNMI infrasound arrays, giving the results of the first two steps of the methodology. Chapter 5 shows the source locations which are found using both the three KNMI infrasound arrays and three more arrays located in France and Germany. In Chapter 6 the third step of the methodology is described in which a comparison is made between the locations obtained by the KNMI infrasound arrays and the locations as determined independently by the method of (Le Pichon *et al.*, 2008). Conclusions and recommendations are presented in Chapter 7.

Chapter 2

Introduction to infrasound

This chapter provides a brief summary of the information needed to understand the presented research by those unfamiliar to the topic of infrasound. First the signal properties in terms of frequency, amplitude and velocity are introduced. Second the known sources of infrasound are presented. Third the concept of measuring infrasound using infrasound arrays is introduced. In the fourth paragraph more insight is given in the atmospheric propagation of infrasound signals.

Sound with a frequency between 0.002 to 20 Hz is considered as infrasound. The upper frequency boundary of infrasound is determined by the minimal frequency which can be registered by the human ear at 20 Hz. The naming of infrasound is analogue to infrared light, which has a frequency lower than what can registered by the human eye. The lower frequency boundary of infrasound is usually considered at 0.002 Hz (Evers, 2008). This boundary can be visualized by comparing the wavelength of the signal in comparison with the thickness of the atmosphere. At 293 K an acoustic wave of 0.002 Hz has a wavelength of 172 km (approximation for an ideal gas). The main restoring force of such a wave is gravity (Gossard & Hooke, 1975) and therefore we speak of gravity waves instead of acoustic waves. An infrasound signal travels with a velocity equal to that of audible sound which is a function of temperature and wind speed (Gossard & Hooke, 1975). The typical amplitude of a detected infrasound signal is in the same order as audible sound at tens to hundreds of Pascal (Evers, 2008).

To create a signal with frequencies below 20 Hz a displacement of a large volume of air is needed, therefore sources of infrasound are generally large and highly energetic. This is illustrated by the summary of sources and their typical frequencies, given in Table 2.2. The low frequency of an infrasound signal causes low attenuation, which means that a strong signal can keep enough energy to be detected after travelling vast distances. As shown by the Krakatau explosion a very strong signal is capable of travelling around the globe several times (Symons, 1888). Especially for higher frequencies one should not forget that local sources such as traffic-tunnel interaction can be an important source of infrasound. As the name already reveals, these sources have not enough energy to travel large distances, but will be detected locally. In this research we try to filter out as much local sources as possible and locate sources in the far field, located tens to hundreds of kilometres away from the measurement location.

The proven method of detecting infrasound is by applying infrasound receiver arrays, or shortly infrasound arrays. An array consists of multiple receivers capable of recording low frequency sound. Each receiver has a different horizontal position which results in a different signal arrival time, determined by the incoming direction of the signal. This time difference between recordings is used to find the originating direction of the signal,

array	elements	aperture
DBN	6	80 m
DIA	16	1500 m
EXL	6	250 m

Table 2.1: Number of elements and aperture of the infrasound arrays DBN, DIA and EXL.

frequency	source
14 - 35 Hz	African elephant (*Loxodonta africana*)
1 - 10 Hz	military activity (sonic booms, rockets, helicopters) mining (open pit mining, gas flares) lightning avalanches meteors (many 10 times a year)
0.1 - 5 Hz	volcanoes
0.1 - 0.5 Hz	1 kT test (explosion) meteors (less than 5 times a year) microbaroms (interaction of ocean waves) severe weathersystems
< 0.1 Hz	large volcanoes mountain associated waves Aurora large nuclear test (explosion) large meteor (1 a year)

Table 2.2: Frequency of known infrasound sources (Evers, 2008)

see Section 3.1. In the Netherlands, KNMI operates a varying number of infrasound arrays, depending on the research performed. On average five infrasound arrays are operated and three of these provide continuous recordings since 2006. These three arrays provide the most important data source for this research and are located in de Bilt (DBN), Exloo (EXL) and Deelen (DIA), Figure 2.1. The aperture and number of elements vary between the arrays and are shown in Table 2.1. A yearly record of infrasound detections made by De Bilt infrasound array is shown in Figure 2.3. This figure shows 114742 events, or an average of one detection every five minutes.

The sound speed is inhomogeneous over the atmosphere due to varying temperature and wind speed. Snell's law (Rees, 2001) shows that where the velocity is not homogeneous, a travelling wave will deviate from linear (i.e. straight line) propagation. Once the varying sound speed over the atmosphere is known, the signal's travel path can be predicted using raytracing methods. An efficient method for raytracing is the tau-p method as presented by Garcés *et al.* (1998). Johnson (2003) gives a very nice illustration of the influence of temperature and wind on infrasound propagation which is included in Figure 2.4. For illustrative purposes a raytrace made by the tau-p method is shown for a synthetic signal travelling from Lyon to de Bilt, Figure 2.2. Under normal conditions raytraces like this produce deviations in back azimuth around 3 degrees and propagation velocities in the range of 220 to 350m/s.

Due to the atmospheric influence on the sound speed multiple definitions of sound speed are introduced in infrasound research. Four definitions are used in this report and therefore explained in more detail. Perhaps the most obvious sound speed is the local sound speed, which is valid only for a single point. The local sound speed can be approximated by the sound speed for an ideal gas (c_{ideal}). If the local wind vector is taken into account it is called the effective sound speed (c_{eff}). In array processing it is common to use the speed at which a signal travels over the infrasound array, which is denoted by the apparent velocity (c_{app}). In a common array configuration all receivers are placed in the horizontal plane. For such an array a signal which travels in horizontal direction will have an apparent velocity which is equal to the effective sound speed. A signal from the zenith is received simultaneously at all receivers. The signal then moves over the array with infinite velocity and has infinite apparent velocity. During the travel from source to receiver the signal will travel at varying speed according to the effective sound speed. The total travel time of the signal divided by the distance between source and receiver is denoted by the celerity (c).

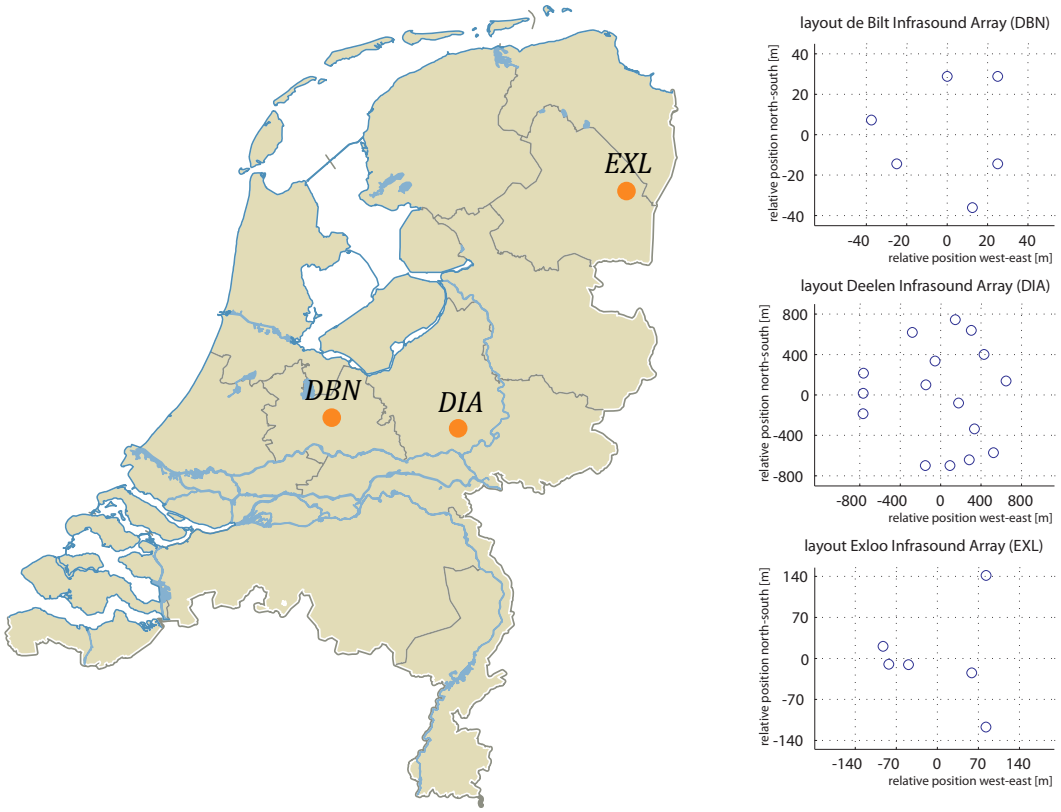


Figure 2.1: Location of infrasound arrays at de Bilt (DBN), Deelen (DIA) and Exloo (EXL). The relative position of the array receivers is shown on the right hand side.

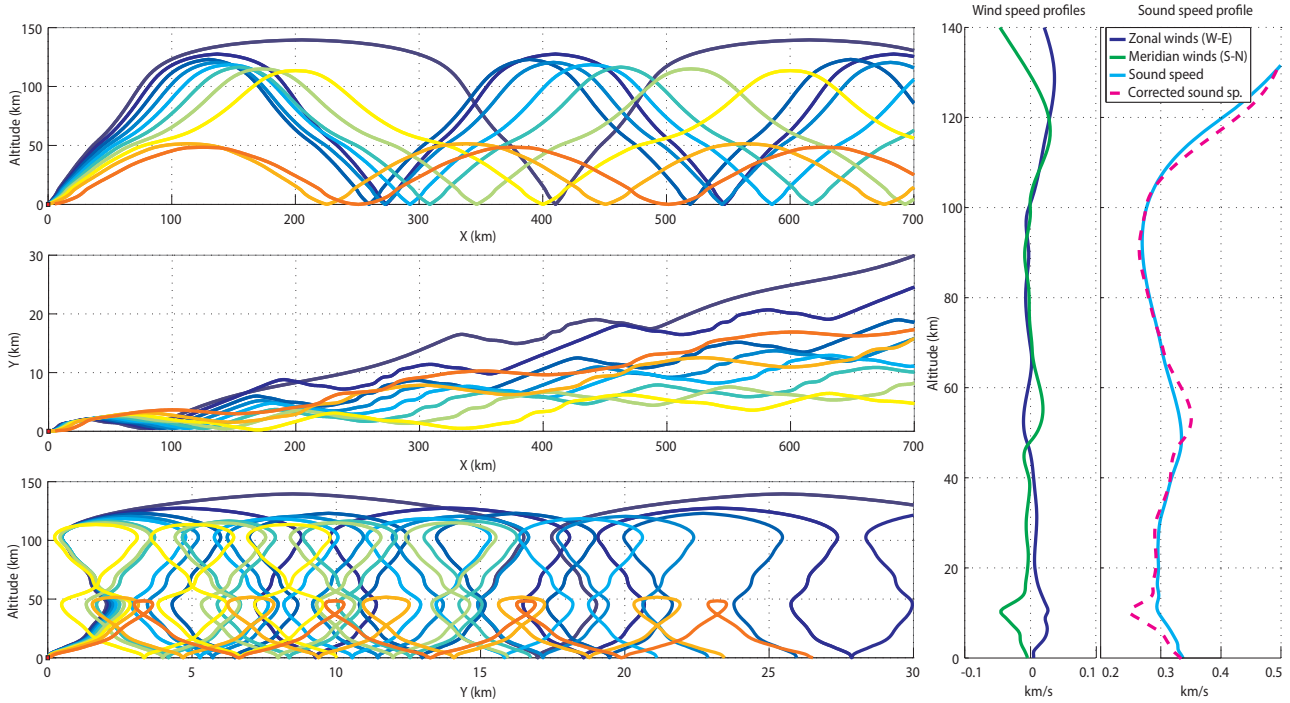


Figure 2.2: Example of atmospheric signal propagation based on ray tracing. Signal propagation is modelled by calculating the propagation of directed signals shot at different angles. The propagation is calculated using the tau-p method (Garcés *et al.*, 1998). On the left hand side the ray traces are shown from a side, top and frontal view. On the right hand side the zonal and meridional wind speed and the sound speed are shown as a function of altitude. The profiles are taken for the Netherlands at April 18, 2007. A synthetic source is set at 0.5km altitude and 700km south of the receiver, comparable to a source at Lyon and the infrasound array at de Bilt. From the source ten rays are shot at different directions in the vertical plane, each ray is shown at a different colour. As the signal travels from south to north, the meridian wind is added to the sound speed profile to form a corrected sound speed profile. The variations in sound speed over altitude causes the ray paths to turn in the vertical plane. Stratospheric returns (turning at 50km) and thermospheric returns (above 100km) are visible. These correspond to the increasing sound speed at 50 km and above 100 km. When the source direction is changed, the contributions of the zonal and meridian winds to the corrected sound speed change accordingly (i.e. inner product with direction). For this temperature profile the positive contribution of the meridian wind is crucial to obtain the sound speed increase needed for stratospheric returns. These returns therefore disappear for sources in other directions. Different crosswinds at different altitudes causes each ray to deviate from its original path in the horizontal plane. For north-south propagation the crosswind is equal to the zonal wind.

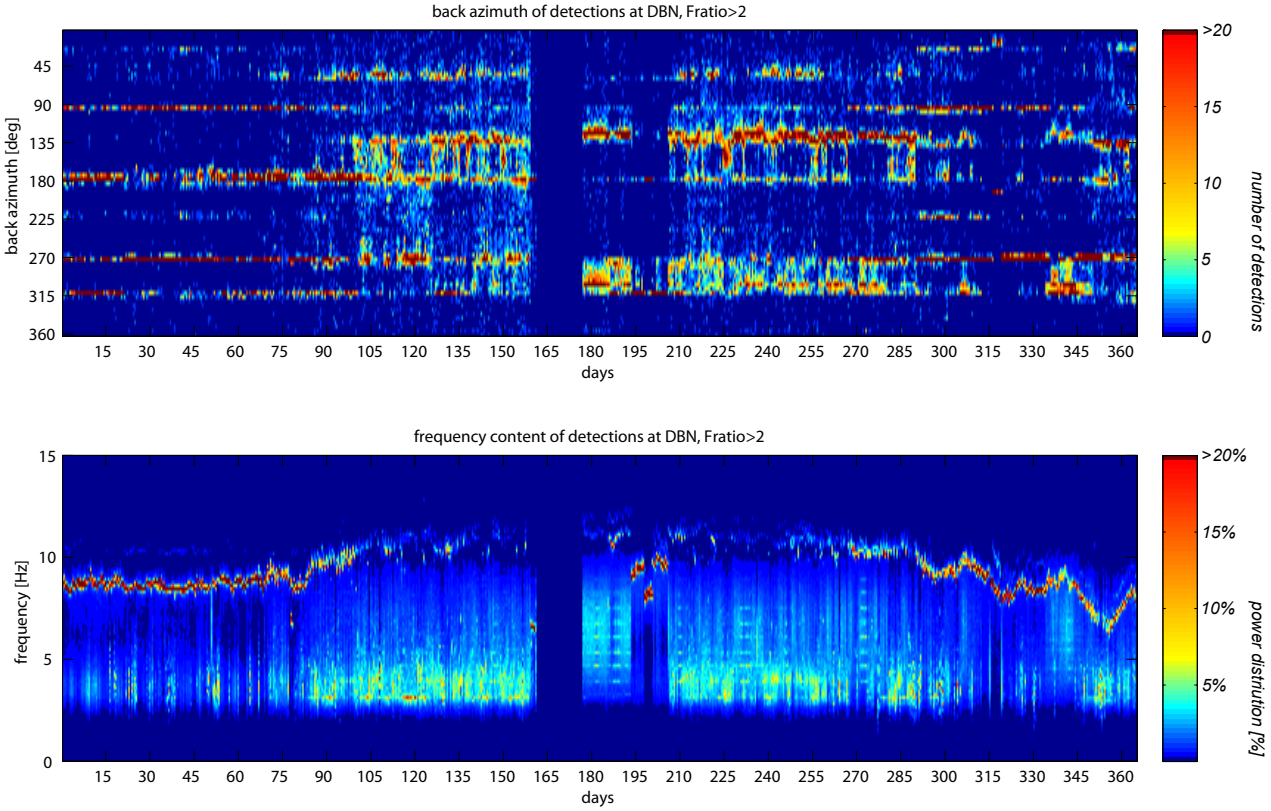


Figure 2.3: The back azimuth and frequency spectrum for all detections with a Fisher ratio larger than 2 (i.e. signal to noise ratio larger than 0.4) at the de Bilt infrasound array over the year 2007. The back azimuth gives the horizontal direction of the signal in degrees from north. All detections are binned over 12 hour by 5 degree bins and the number of detections is shown in the graph. A large number of detections arrives from a distinct set of back azimuths. These detections are expected to be related to local sources. The distribution of detections is not evenly spread over the year, which can be caused by changing atmospheric conditions. A gap in the data is visible around day 165. The frequency spectrum shows the frequency distribution for the detected signals binned over 12 hour bins. The signals are separated over 96 frequency bands and the percentage contribution of each frequency band is shown in the graph. As input the recorded signals are filtered such that only the signals between 1 and 9.9 Hz are used. A strong high frequency signal is present which increases above the 10 Hz boundary in summer and therefore disappears from the detection list. This relatively high frequency phenomenon is possibly related to local sources but the seasonal variation in frequency is not yet explained.

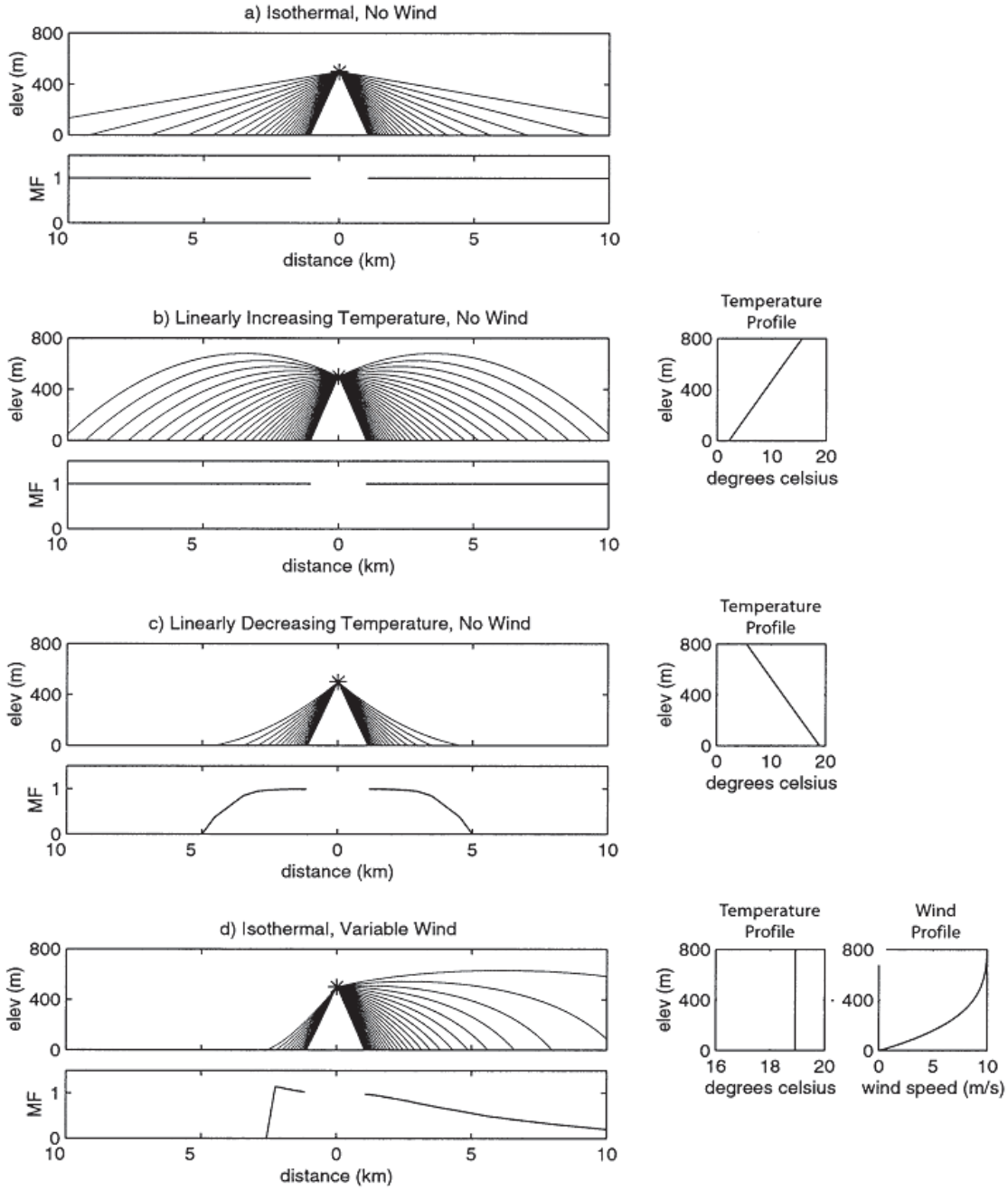


Figure 2.4: Figure taken from (Johnson, 2003). Infrasound focusing at close offsets. Dependence of raypaths on variable atmospheric conditions: (a) homogeneous temperature structure with no wind, (b) temperature inversion with no wind, (c) normal temperature gradient with no wind, and (d) isothermal atmosphere and wind increasing with altitude. Acoustic rays are drawn at 1° increments. For all four scenarios the Magnification Factor (MF) is shown for each raypath. The MF is defined here as the square root of the ratio of the recorded pressure amplitude, to the expected pressure amplitude for an isotropic acoustic source radiating spherically into a homogenous atmosphere. An absence of raypaths impacting the ground shows a region where no signal will be received (i.e. a shadow zone).

Chapter 3

Processing Method

The process from individual infrasound recordings to possible locations of infrasound sources is explained in this chapter. Processing is preformed in three stages. The stages are shown in a flowchart, Figure 3.1, which indicates the explaining sections for each stage. First the recordings of each array are processed separately, Section 3.1. This results in a detection list per array which gives various statistics for each detected signal. Of these statistics the detection time and back azimuth (i.e. the horizontal direction of the signal) are required in stage two to combine detections at multiple arrays that could originate from the same event, Section 3.2. These detections are combined into events. In the final stage the location of each event is determined by combining the back azimuths of all detections within the event, Section 3.3. Filters are applied at various stages in the processing scheme. All filters are discussed in Section 3.4.

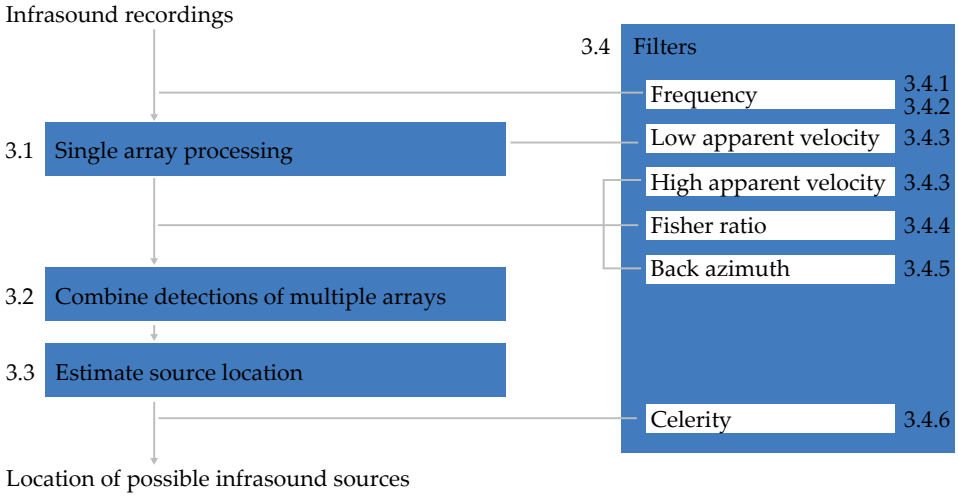


Figure 3.1: Flow chart of processing. The numbers give the sections in which the stage is explained in detail.

3.1 Array processing

The detection of infrasound signals is two folded: it has to be determined if a signal exists and from which direction it has arrived. The two questions are solved simultaneously by determining the probability that a signal is received from a range of directions. As input small time blocks of the recordings of multiple receivers within a single infrasound array are used. Each receiver has a different location in the array which results in a different signal arrival time at each receiver, determined by the incoming direction of the signal. This relation is shown in Section 3.1.1. The same principle is used to determine the probability of a signal over a range of directions in frequency-wavenumber analysis, Section 3.1.2. The probability is described by the Fisher ratio which is given in more detail in Section 3.1.3. For each time block the direction with maximal probability is selected. If this probability exceeds a set threshold (corresponding to a Fisher ratio larger than 2 for this research) it is saved as a detection. For each detection the theory of frequency-wavenumber analysis is used to give the variance on the direction, Section 3.1.4. The number of statistics saved per detection is variable, but a typical detection list contains the detection time, Fisher statistic and the two angles that form the incoming direction: back azimuth and incident angle. A block scheme of the software used for the single array processing is given in Section 3.1.5.

3.1.1 Signal direction and receiver location

The time shift between signal recordings at multiple receivers is a function of the receiver location and the incoming direction of the signal, Equation (3.1). The incoming direction is described by a two dimensional slowness vector $\vec{p}$ which is the inverse of the signal's velocity vector.

The infrasonic signal emerges from a source with a known or unknown location, both in space and time. Due to the small array aperture (in the order of hundred metre) relative to the expected travel distances (in the order of ten to hundred kilometre) the signal travels over the array as a plane wave (i.e. at the same angle at each receiver). With this assumption we can estimate three parameters which give information on the source location: two angles for the incoming direction (back azimuth and incident angle, Figure 3.2) and arrival time.

For every receiver of the array one observation is added to solve the back azimuth, incident angle and arrival time. It is therefore already clear that we need at least three receivers (three knowns) to solve these parameters (three unknowns). With the assumption of infinite distance, the signal will arrive as a plane wave and thus arrives from the same direction at every receiver position. Once one 'reference' receiver has recorded the arrival time, every next receiver records the same signal shifted in time and possibly corrupted with noise. This time shift can now, through the sound speed, be related to the position of this receiver with respect to the reference receiver. The recorded signal can be expressed as:

$$y(\vec{r},t) = \sum_{\omega} A(\omega)e^{i\omega(t-\vec{p}\cdot\vec{r})+\phi(\omega)} + y_0 + \delta_{receiver}(t) + \delta_{medium}(t)) \tag{3.1}$$

where y is the observed signal for the receiver position $\vec{r}$ and time t. Noise is added as a receiver noise parameter $\delta_{receiver}(t)$ and noise added by medium differences between receivers $\delta_{medium}(t)$. The total signal is the summation of multiple frequency (ω) contributions with amplitudes $A(\omega)$ and phase shifts $\phi(\omega)$. The phase changes over time by $i\omega t$. With respect to a reference point at location $(0,0,0)$, the receiver position causes a time shift of the signal. Slowness is defined as the inverse of velocity (e.g. s/m), therefore the dot product of the signal's slowness vector $\vec{p}$ and the receiver position $\vec{r}$ gives the time shift of the signal with respect to this reference point. If a constant DC component is present in the signal this is expressed by y_0.

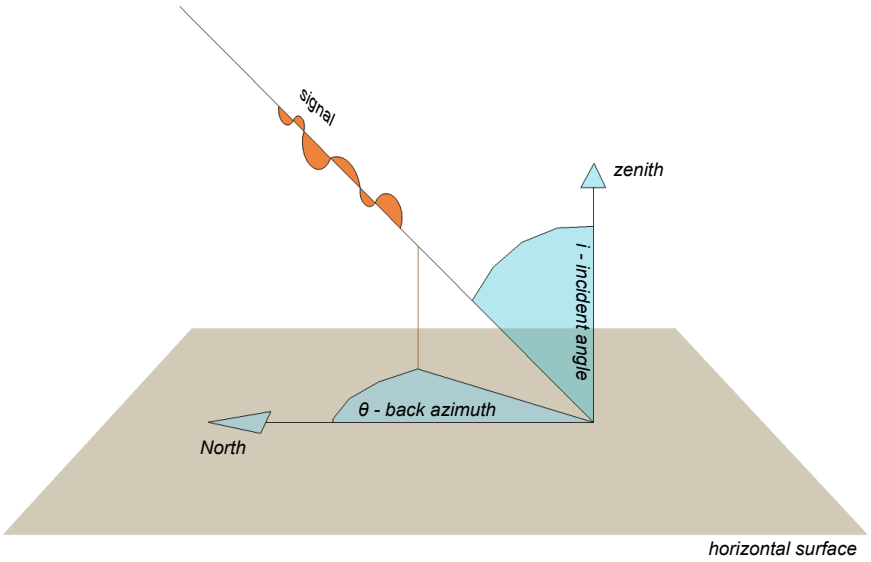


Figure 3.2: The direction of an incoming signal is described in two angles: the back azimuth and the incident angle. The back azimuth is measured with respect to the local north direction. Incident angle is the deviation from the zenith (i.e. perpendicular to the horizontal).

In a traditional configuration all receivers are placed in the same horizontal plane. The horizontal receiver position is therefore described by a two dimensional vector $\vec{r} = [x, y]$. The previously introduced slowness vector will be defined in the same coordinate system: $\vec{p} = [p_x, p_y]$. The slowness vector gives the incoming direction of a signal, but it can be more convenient to describe this direction in two angles: back azimuth θ and incident angle i, as shown in Figure 3.2. This transformation can be made by

$$\theta = \arctan\left(\frac{p_x}{p_y}\right) \tag{3.2}$$

$$i = \arcsin\left(c\,|\vec{p}|\right) \tag{3.3}$$

where $|\vec{p}|$ is the length of the slowness vector and c is the local sound speed. If the local sound speed is unknown, it is more convenient to describe the incident angle as the apparent velocity (Rost & Thomas, 2002):

$$c_{app} = \frac{1}{|\vec{p}|} \tag{3.4}$$

Slowness can be exchanged with wavenumber $\vec{k}$ and frequency ω using:

$$\vec{p} = \frac{\vec{k}}{\omega} \tag{3.5}$$

3.1.2 Frequency-wave number analysis

Frequency-wave number analysis gives the power of a signal at a certain frequency over a range of wave numbers. In infrasound research it is common to replace the wavenumber by the related slowness vector. The signal power is converted to a Fisher ratio which is related to the signal to noise ratio. For each time block of recordings the Fisher ratio's are calculated over a grid of slowness vectors. The maximum Fisher ratio is compared to a threshold. This threshold can be set according to the probability on a false detection that one wishes to allow, where it is assumed that noise has a central Fisher distribution. If the threshold is exceeded the detection is saved. By selecting the maximum Fisher ratio in the slowness grid, the slowness which has the highest probability of containing a signal is selected. As shown in Section 3.1.1 the slowness lets us calculate the back azimuth of the signal, the arrival time is given by the time of the inserted block of recordings.

For frequency-wave number analysis the recorded signals have to be separated in a number of frequency bands. Separation in frequency bands can be done in the time domain by bandpass filtering or in the frequency domain by Fourier analysis. In this section the signals will be described in the frequency domain. Consequently, time shifts between received signals are described by phase shifts.

When a signal arrives at the array each receiver records the signal at a different phase. The phase difference between recordings can be calculated from the frequency, the receiver locations and a yet to determine slowness of the signal, Equation (3.1). Frequency and location are known and for a range of slowness values the phase differences are removed from the recordings and the signals are summed. The summed signal with highest power will correspond to the slowness of the signal that moved over the array. The slowness range is defined in two dimensions forming a square grid of slownesses. The summed signal power estimated at each slowness is given by the frequency-wavenumber power spectral (Smart & Flinn, 1971):

$$P(\omega, \vec{p}) = \left| \sum_{n=1}^{N} G(\omega, \vec{r_n}) e^{-i\omega \vec{p} \cdot \vec{r_n}} \right|^2 \tag{3.6}$$

where the recorded signal at receiver n is given as $G(\omega, \vec{r_n}) = A_n(\omega) e^{i\alpha(\omega)}$. The phase delay expected for slowness $\vec{p}$ is set equal for each receiver by $e^{-i\omega \vec{p} \cdot \vec{r_n}}$, a value that has to be calculated once for every receiver and slowness, and can then be applied to all recorded signals. All phase shifted signals are summed and squared. Aligned signals will amplify each other where misaligned signals will weaken each other. The alignment generating the highest power will most likely correspond to the direction of the incoming signal.

If one does not know if there is a signal, the power of a signal is only meaningful in comparison with the power of the noise. This is incorporated by calculating the Fisher ratio for each slowness vector in the slowness grid. The Fisher ratio gives the ratio between the power of the signal and the remaining recorded power. This ratio forms a predictable indication for the presence of a signal. By selecting the maximum Fisher ratio in the slowness grid, the direction which has the highest probability of containing a signal is selected. The Fisher ratio is defined as the estimated signal power divided by the estimated noise power (Shumway, 1971) (Evers, 2008):

$$P_s(\omega, \vec{p}) = \left| \frac{1}{N} \sum_{n=1}^{N} G(\omega, \vec{r_n}) e^{-i\omega \vec{p} \cdot \vec{r_n}} \right|^2 \tag{3.7}$$

$$P_t(\omega) = \frac{1}{N} \sum_{n=1}^{N} |G(\omega, \vec{r_n})|^2 \tag{3.8}$$

$$F(\omega, \vec{p}) = \frac{P_s(\omega, \vec{p})}{P_t(\omega) - P_s(\omega, \vec{p})} (N-1) \tag{3.9}$$

$$\tag{3.10}$$

The noise power is defined as the total recorded power P_t minus the power attributed to the signal P_s. For more information on the Fisher ratio see section 3.1.3. An example of Fisher ratio's calculated over a slowness grid is shown in Figure 3.3.

3.1.3 Fisher ratio

The Fisher ratio gives an indication for the existence of a signal in multiple recordings with unknown signal and noise. It is related to the signal to noise ratio by Equation (3.18). This section derives the Fisher ratio and follows (Melton & Bailey, 1957).

The mean or expected value of a random recording is given as

$$E(x) = \bar{x} = \frac{1}{NT} \sum^{T} \sum^{N} x_{nt} \tag{3.11}$$

Where N is the number of channels or receivers, T is the number of samples taken at each receiver and x_{nt} are the recorded values at channel n and time t.

The average channel voltage or apparent signal is formed by the average over all channels:

$$\bar{x}_t = \frac{1}{N} \sum^{N} x_{nt} \tag{3.12}$$

The variance of the apparent signal with respect to the mean recorded value is called between sub-sample variance:

$${V_t}^2 = \frac{N}{T-1} \sum^{T} (\bar{x}_t - \bar{x})^2 \tag{3.13}$$

The variance of all recordings with respect to the mean recorded value is called within sub-sample variance:

$${V_n}^2 = \frac{1}{T(N-1)} \sum^{T} \sum^{N} (x_{nt} - \bar{x}_t)^2 \tag{3.14}$$

The Fisher ratio is defined as the ratio of sub-sample variance with within sub-sample variance:

$$F = \frac{{V_t}^2}{{V_n}^2} \tag{3.15}$$

The estimated signal power is defined by (Melton & Bailey, 1957): "that variance due to the deviation not ascribable to the common cause". It is formed by the apparent signal power minus the estimated noise power, devided by the number of receivers:

$${P_S}^2 = \frac{{V_t}^2}{N} - \frac{{V_n}^2}{N} \tag{3.16}$$

The estimated noise power:

$${P_N}^2 = {V_n}^2 \tag{3.17}$$

The signal to noise ratio is then given by:

$$\mathrm{SNR}^2 = \frac{{P_S}^2}{{P_N}^2} = \frac{1}{N} \left(\frac{{V_t}^2}{{V_n}^2} - 1 \right) = \frac{1}{N} (F - 1) \tag{3.18}$$

3.1.4 Estimate variance of back azimuth and apparent velocity

The back azimuth and apparent velocity of an incoming signal were determined by the point of maximum Fisher ratio over a slowness grid, Section 3.1.2. Because the Fisher ratio's over the entire slowness grid are known this can be used to determine the precision at which the position of this single point is determined, expressed in the variance of back azimuth and apparent velocity.

To limit the computation time needed to obtain the variance the slowness grid is approximated by a single strip of slowness in horizontal and vertical direction. Over these two strips an approximation of the distribution is made using a two dimensional Lorenz function:

$$y(x) = \frac{A}{1 + \left(\frac{x-B}{C}\right)^2} \tag{3.19}$$

where A, B and C give the normalisation, centre and width of the Lorenz profile. The Lorenz function is fitted to the slowness grid strip using the Levenberg-Marquardt method (Press *et al.*, 1992) which provides the covariance matrix for the fitted parameters A, B and C describing the Lorenz function. The variance on the central peak position B of the fit is used as estimate for the variance on the true central peak position. The choice for a horizontal and vertical strip now results in the variance on slowness $[\sigma_{B_{hor}}, \sigma_{B_{ver}}] = [\sigma_{p_x}, \sigma_{p_y}]$. The Lorenz fit is illustrated in Figure 3.3.

The Levenberg-Marquardt method uses the variances of the calculated F-ratio of each gridpoint to come to a fit. These variances can be calculated from the F-distribution using (Weisstein & Eric, 2002)

$$\sigma^2 = \frac{v_2^2[\lambda^2 + (2v_1 + 4)\lambda + v_1(v_1 + 2)]}{v_1^2(v_2 - 4)(v_2 - 2)} \tag{3.20}$$

The F-distribution is described by two degrees of freedom and a non-centrality parameter; $F(v_1, v_2, \lambda)$. The first degree of freedom is given by the number of samples in the measurement $v_1 = T - 1$. The second degree of freedom is given by the number of samples and the number of receivers $v_2 = T * (N - 1)$. The non-centrality parameter is a function of the Fisher ratio and v_1, $\lambda = (F - 1) * v_1$ (Teunissen *et al.*, 2006).

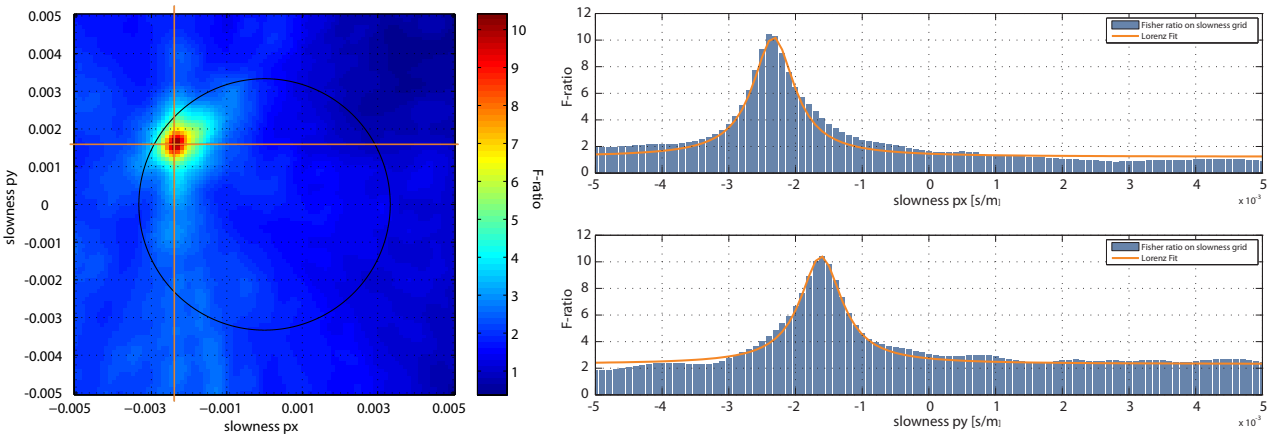


Figure 3.3: On the left Fisher ratio's over a slowness grid during the arrival of a signal recorded at DIA are shown. The signal contained energy between 0 and 2 Hz. The black circle shows the boundary of apparent velocity which is expected to be above 320m/s. A dark red spot is shown in the upper-left giving the slowness generating the largest Fisher ratio. The signal is therefore estimated to arrive from a North-Western direction with an apparent velocity of 353m/s. The two orange bars show the location of the horizontal and vertical strips used to fit a Lorenz profile, as shown on the right. The Levenberg-Marquardt method (Press *et al.*, 1992) results in a fitted profile, shown in orange, and gives the variance on the position of the peak.

To estimate the variance on back azimuth and apparent velocity we take the first order Taylor approximation for the variance of the two-dimensional function $\underline{z} = g(\underline{x}_1, \underline{x}_2)$ (Teunissen *et al.*, 2006):

$$\sigma_z^2 = \left(\frac{\partial g\left(\bar{x}\right)}{\partial x_1}\right)^2 \sigma_{x_1}^2 + \left(\frac{\partial g(\bar{x})}{\partial x_2}\right)^2 \sigma_{x_2}^2 + 2\rho\sigma_{x_1}\sigma_{x_2}\frac{\partial g(\bar{x})}{\partial x_1}\frac{\partial g(\bar{x})}{\partial x_2} \tag{3.21}$$

The third term from Equation (3.21) describes the covariance of $\underline{x}_1$ and $\underline{x}_2$. When $\underline{x}_1$ and $\underline{x}_2$ are orthogonal as is the case with p_x and p_y the covariance becomes zero. With the first two terms of Equation (3.21) and $\underline{z} = g(\underline{x}_1, \underline{x}_2)$ set to $\theta = \arctan(p_x/p_y)$ (Equation (3.2)) the variance on back azimuth is:

$$\sigma_\theta^2 = \left(\frac{\frac{1}{p_y}}{1 + \left(\frac{p_x}{p_y}\right)^2}\right)^2 \sigma_{px}^2 + \left(\frac{p_x}{p_x^2 + p_y^2}\right)^2 \sigma_{py}^2 \tag{3.22}$$

Now set $\underline{z} = g(\underline{x}_1, \underline{x}_2)$ to $c_{app} = 1/\sqrt{p_x^2 + p_y^2}$ (Equation (3.4)) and the variance on apparent velocity is:

$$\sigma_{app}^2 = \frac{p_x^2}{(p_x^2 + p_y^2)^3}\sigma_{px}^2 + \frac{p_y^2}{(p_x^2 + p_y^2)^3}\sigma_{py}^2 \tag{3.23}$$

3.1.5 Array processing block scheme

A block scheme of the software used for array processing is shown in Figure 3.4. The processing software is build from a combination of existing KNMI software code and was written in Fortran and Perl.

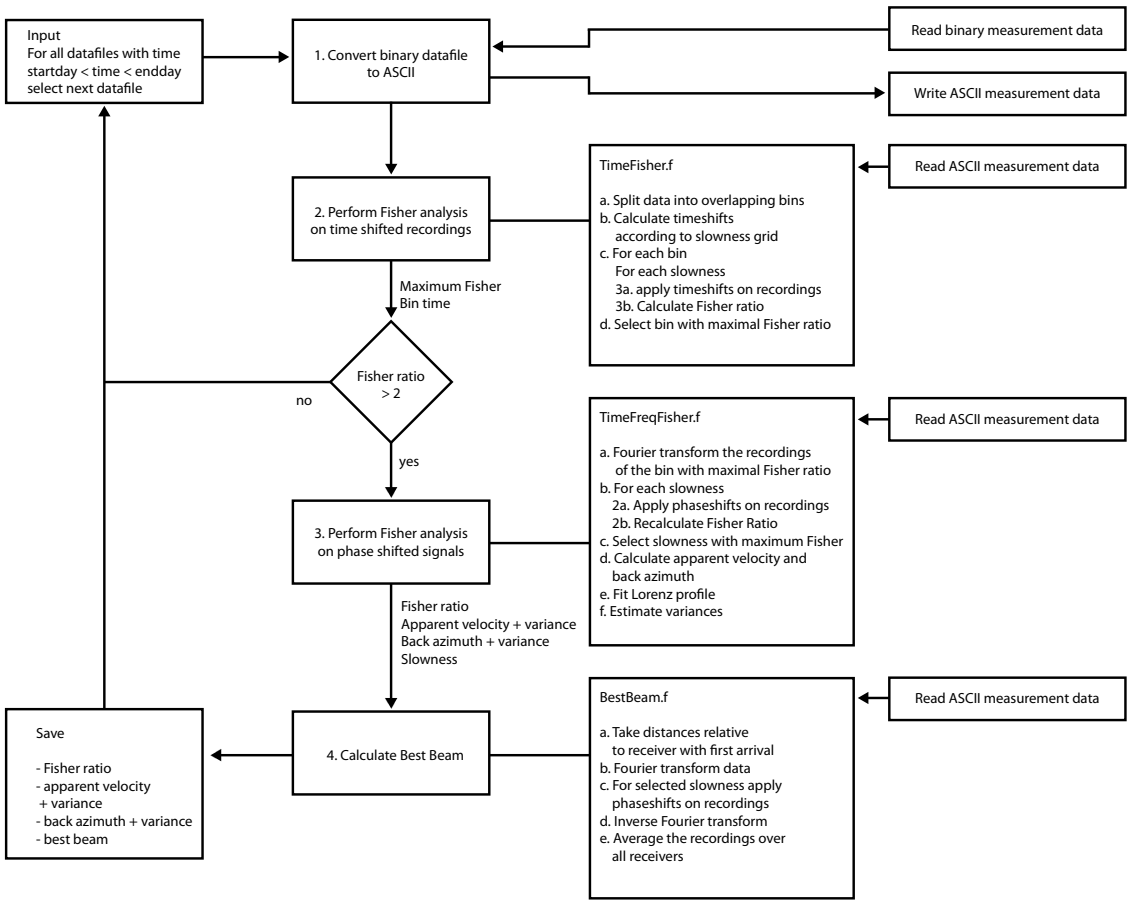


Figure 3.4: Block scheme of the software used for array processing. The scheme is runned for each datafile in the selected period. Input are the chosen days and years that should be processed from array data given in binary files. 1. A single binary data file is selected and converted to ascii. 2. The datafile is split into overlapping time bins of 6.4 seconds. For each bin the Fisher ratio's are calculated over a slowness grid where the shift in signal between receivers is applied in the time domain. The bin which generates the maximum Fisher ratio is selected for further processing. The procedure in the time domain is considerably faster than a procedure where the signals are converted to the frequency domain. However, due to a limited time resolution of the recorded signals (40 Hz) the time domain procedure provides limited accuracy. Therefore, if a signal with a Fisher ratio larger than two is found, the procedure is repeated in the frequency domain in step 3, but only for the bin which gave a maximum Fisher ratio in the time domain. The signals in this bin are converted to the frequency domain, phase shifts are applied and the signals are converted back to the time domain to calculate the Fisher ratio. It is possible to perform this entire step in the frequency domain as presented in section 3.1.2 but it was chosen to use the existing and proven KNMI software. The slowness with a maximum Fisher is selected and the corresponding back azimuth and apparent velocity are calculated. In this step also the Lorenz fit as presented in section 3.1.4 is made and the variances on back azimuth and apparent velocity are estimated. 4. A best beam signal is obtained by taking the average of the phase shifted signals of all receivers. Finally, the detection is saved and the next binary file is processed. In the current procedure only the detection with largest Fisher ratio is saved for each binary data file of 256 seconds.

3.2 Combine multiple arrays

Individual array processing resulted in detection lists for each array. In this section a method is presented to select possible common events between these detection lists based on the detection time and back azimuth. Tolerances are applied on the back azimuth and celerity to accommodate atmospheric influences. Afterwards multiple detection pairs which are close in back azimuth and time are merged into events.

3.2.1 Select common events between arrays

On average, the infrasound arrays used in this research generate one detection every five minutes. Only a few of these detections originate from events that are detected at multiple arrays. This makes it almost impossible to spot common detections directly from the detection list, see Figure 3.5. Instead, an automated process is developed to use the recorded back azimuth and detection time to check if it is physically possible to link two or more detections. One array is selected as primary array and for each detection of this array a search is made to link detections at the other arrays. For a chosen set of source distances the expected arrival time and back azimuth at the other arrays is calculated. The detection lists are searched for detections matching these timings and azimuths. If one or more detections are found they are saved as one event. To accommodate atmospheric influences fixed tolerances are applied on the back azimuth and celerity. The tolerances allow for the maximal deviation from straight line propagation that is expected. This deviation is different for each detection and can only be estimated from an atmospheric model. The search process is illustrated in Figure 3.6. The search windows used in this research for the 315° direction are shown in Figure 3.7. The applied tolerances are given in Table 3.1.

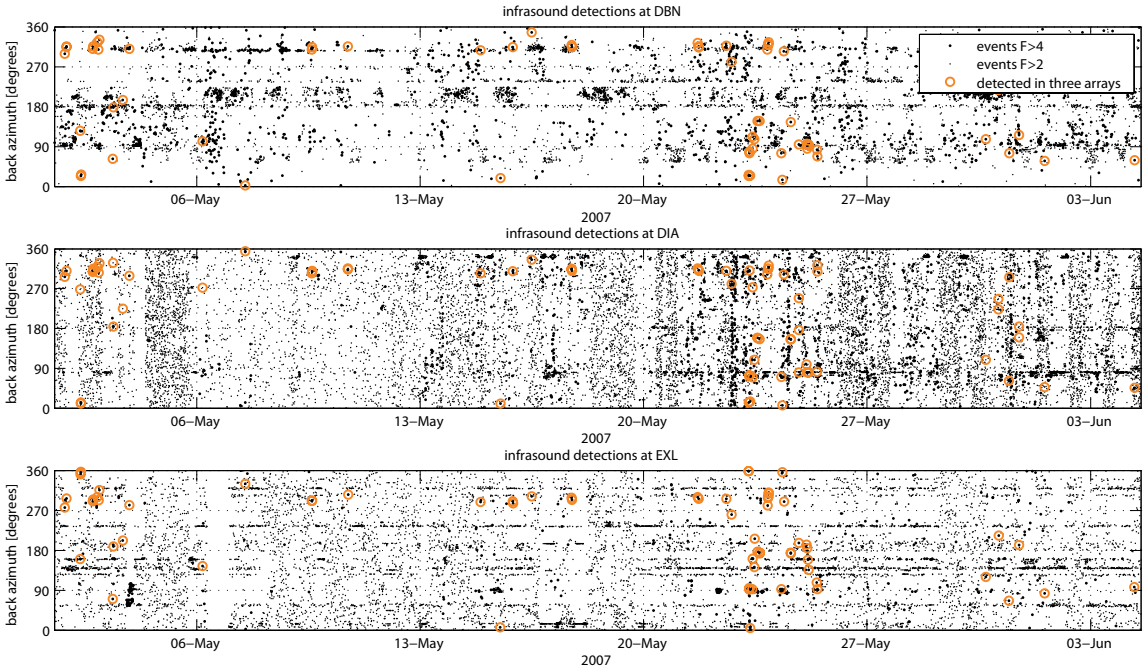


Figure 3.5: The back azimuth for one month of detections at the de De Bilt infrasound array (DBN), Deelen infrasound array (DIA) and Exloo infrasound array (EXL). Orange circles show which detections are selected as possible common events between the three arrays. Each circled detection is related to the circled detections at roughly the same time in the other two plot windows (i.e. directly above or below), not to circled detections in the same plot window. Due to the large amount of detections it is very difficult to see correlation between the detections at the three arrays without estimating back azimuths and arrival times.

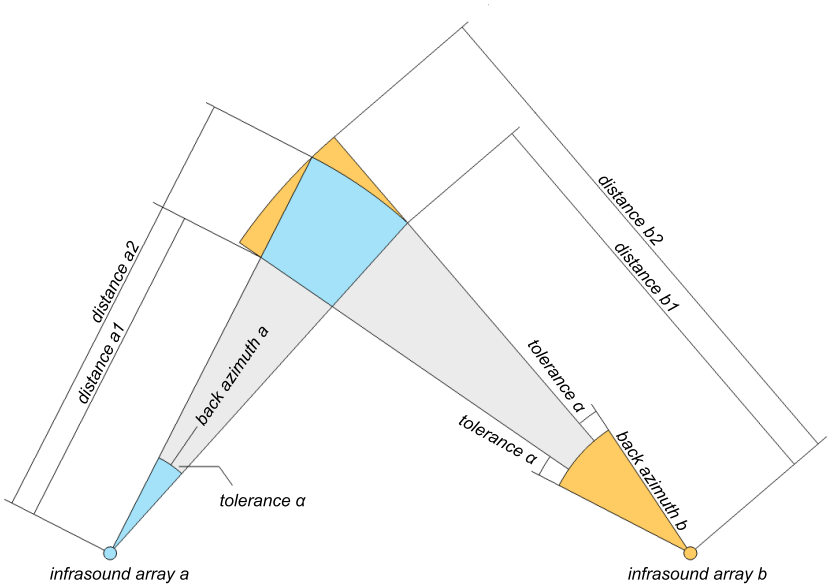


Figure 3.6: Illustration of a search window to match detections between multiple arrays. From the primary infrasound array a a search angle is defined by back azimuth a of a detection at this array plus and minus the azimuthal tolerance α. For two subsequent distances $a1$ and $a2$ from the set of source distances the four corner points of the search window are calculated. From these four points a back azimuth range at the secondary infrasound array b is obtained. Again the tolerance is added to form the minimal and maximal back azimuth b. These values will be used to search in the detection list of array b. Minimal and maximal distances b to array b are selected. All four distances ($a1$, $a2$, $b1$, $b2$) are combined with the minimal and maximal allowable celerity to calculate a minimal and maximal arrival time at secondary array b.

The horizontal propagation speed of the signal between source and receiver (i.e. celerity) is used to calculate the signal arrival time:

$$t_b = t_a + d_a/c_a - d_b/c_b \tag{3.24}$$

where t_a is the arrival time at the primary array, t_b at the secondary array, d_a is the distance from the source to the primary array, c_a the celerity between the source and the primary array, d_b and c_b are the distance and celerity between the source and the secondary array. The arrival time is calculated for all combinations of minimum and maximum distances for both a and b (distance $a1$, $a2$, $b1$, $b2$ in Figure 3.6) and the minimal and maximal allowed celerity, Table 3.1. The maximal and minimal value of these eight arrival times are used as search boundaries to search the detection list of the secondary array. All distances and angles are calculated on a spherical earth using the inverse problem of geodetics (the equations are not needed here, but are given in Section 3.3.2).

	minimal value	maximal value
azimuth	-10°	+10°
celerity	220 m/s	360 m/s

Table 3.1: Tolerances on back azimuth and range of celerities. The tolerances allow for the atmospheric deviation from straight line propagation and are fixed for all detections. The deviation differs between each detection and can only be estimated from an atmospheric model. The tolerances are set at the maximal deviation that is expected. By doing so, one allows multiple detections of a single event which experienced a different atmospheric deviation to be registered as one event. The range of allowed celerities is taken from literature (Revelle & Whitaker, 1996; Evers & Haak, 2007; Ceplecha *et al.*, 1998). Azimuthal deviations up to 7 degrees from the straight line direction were observed by Evers & Haak (2007). The azimuthal deviation is assumed to be within 10 degrees and found to rarely exceed 5 degrees for locations estimated without atmospheric correction in this research.

Due to the fact that infrasound arrays have a fixed geographical location the search windows for each back azimuth have to be determined only once. All detections are rounded to the nearest back azimuth for which the search windows are already available. The introduced round off error should be sufficiently small with respect to the azimuth tolerance. For this research the search windows are calculated for each degree from 1 to 360, resulting in a maximum of half a degree round off error on 10 degree tolerance. Because the search windows only have to be calculated once, the matching process is very fast. An entire year of measurements (ten thousands of detections) at two arrays can be compared in half a minute on a laptop pc (Intel Core2Duo 2.0Ghz, year 2007). Another advantage of using match windows is the fact that one can easily choose which errors or atmospheric influence on back azimuth and celerity are allowed.

It should be noted that using this method one can only combine detections of events which are detected at the primary array. When using more than two arrays, and when it is desired to detect events which are not detected by the primary array but only by the secondary arrays, multiple runs should be made, each time changing the primary array. Once all combinations are made, each event will be saved a multiple number of times, once for every array which has detected the event. This redundancy is removed in Section 3.2.2.

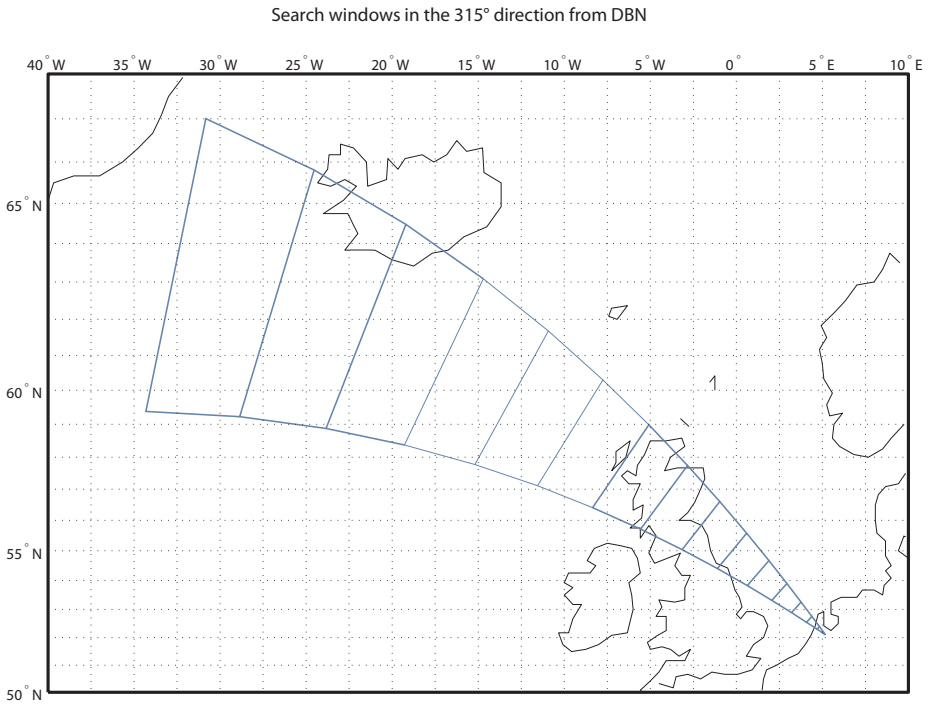


Figure 3.7: Example of search windows in the 315 degree direction from the de Bilt infrasound array as used during this research. The set of source distances is defined as a function $d = x^2$ where d ranges from 0 to 256000 m with nonlinear intervals created by increasing x from 0 to 1600 in steps of 100. This nonlinear distance vector is chosen to have the advantage of small search windows near the arrays located in the Netherlands and large search windows in the far field where such a resolution is not necessary. The window coordinates are calculated on a spherical earth, which causes the line of windows to follow the curvature of the earth.

3.2.2 Group events

In the first part of this section detections from multiple arrays which could be traced to a possible common event are linked. If sources generated multiple atmospheric returns, or emitted over a longer timespan, this process results in multiple events for a single source. To combine these events to one event a search procedure is needed. In the case of signals that share a common travel path it would be logical to use the variances on back azimuth as a statistic to discriminate between related and unrelated events. However it is expected that one can receive multiple atmospheric returns for a single event, in which each return has a different travel path through the atmosphere. To accommodate this the variance which is valid for a single detection cannot be used. Therefore, detections are combined based on fixed tolerances on back azimuth and detection time, similar to the tolerances used in Section 3.2.1. The use of fixed tolerances results in a predictable procedure and gives the observer clear inside on why events are merged.

The tolerances used for this combination procedure can be tweaked for a specific purpose or dataset. During this research a tolerance on back azimuth of 10 degrees and a tolerance on time of 20 minutes are selected. The azimuthal tolerance coincides with the tolerance used in Section 3.2.1. The timing tolerance will result in multiple events for continuous sources that emit for over 20 minutes. Because this is known, it can be identified manually in the results. For two detections to be merged into one event it is required that the differences in detection time and back azimuth at each array are within these tolerances. When multiple events can be merged, the two events which are closest in back azimuth are merged first and further detections are added based on the mean back azimuth of the detections already present in the event.

Additionally, an association number is added to each event. This number is shared between events for which not all array detections are within the tolerances, but the tolerances are met at at least one array. Events sharing the same association number show at least at one array detections that probably originate from the same source. At some other arrays the detections for these events are not found to be related. This suggests that only one of these events combines detections of a single source while the other events combine the detections of this source at one array with unrelated detections at other arrays. The association number can now be used to pick one of these events as most probable true, the pick will be made based on celerity in Section 3.4.6.

For each event the question now rises on how to combine the information of the individual detections. In this research, for source location only the back azimuth is used, therefore only the combination of this parameter is discussed. Le Pichon *et al.* (2008) suggests to select the detection combination which gave a location resulting in a celerity closest to 300m/s. This is an elegant method to distinguish detections that are valid for this event from unrelated detections but it will remove information of multiple detections. Arrowsmith *et al.* (2008) applies the suggestion of Brown *et al.* (2002) to flag the first arrival of a source and then uses this first arrival for source location. This selection also removes information of multiple arrivals. Most important, both selection criteria are not based on physical knowledge on which arrival one can use best for source location. Evers & Haak (2007) gave a clear example of a single impulsive event which resulted in multiple valid detections. Under the influence of strong winds in the upper atmosphere the detections suffered from azimuthal deviations, varying from 1 to 7 degrees. Travel paths which spend a longer time in the high wind region suffered from larger azimuthal deviations. The different phases were distinguished by different celerities and incident angles, the deviation on azimuth could be predicted by raytracing. To come to the most accurate location one would like to use the detection with the smallest azimuthal deviation. Unfortunately it is not possible to select this best detection on the basis of detection parameters or celerity estimates only. Therefore it is opted to weigh all detections at a single array and within an event at equal importance, meaning that the average back azimuth at each array is used for source location. The estimated variances on the back azimuth (Section 3.1.4) are only valid for a single detection, not for the assembly of multiple returns which are biased by an atmospheric azimuth deviation. After an initial position is obtained, one can try to identify different atmospheric returns. If the modelled returns agrees with the actual detections the back azimuth can be adjusted for the atmospheric influence. This adjustment is not applied in this research.

3.3 Estimate source location

In this section a method is derived to come to a best unbiased estimate for the source location and source time. First the different approaches for such a multidimensional problem are given, then the solution is worked out for the planar case which can be used for short distances (less than 5000 km) and finally for the spherical case which should be used over larger distances. Models are shown to estimate source location, source time and celerity. From the given methods, the spherical method to estimate the source location will be applied in this research. Input for this estimation are the means of the back azimuths of all detections within an event for each array.

Following (Bakker *et al.*, 1995) we differentiate between three methods for computations on the earth's surface.

1. computation in three dimensional space, E^3
2. computation on a ellipsoid or sphere, R^2
3. computation on a plane surface, E^2

The three dimensional case requires us to set extra conditions for the solutions to lie on the earth's surface. In the other two cases this is already solved for as these are the only available coordinates. They are therefore found to be more practical. When considering very large distances the planar surface cannot cope with the curvature of the earth which will hamper us to come to an accurate cross bearing. When source and receiver are more than 45 degrees apart (about 5000 km), the straight line distance starts to deviate more than 10 percent of the ellipsoidal distance. It is well known that large infrasonic events can travel these distances. For shorter distances, as will be treated mostly by this research, a plane surface could be sufficient. To illustrate the difference over an expected distance for this research the distance between de Bilt and Edinburgh is chosen as a benchmark, the results are shown in Table 3.2.

For the ellipsoid, positioning from angles and distances cannot be solved analytically, but has to be solved iteratively by using differential equations, see Bakker *et al.* (1995) for details or Vincenty (1975) for solutions to the direct and inverse problems. A method based on differential equations makes solving the cross-bearing problem much more complicated and it is therefore omitted. The remaining two methods, for a planar surface and for the sphere, are worked out in the rest of this section.

reference surface	g.c. distance	error	percentage
WGS84 Ellipsoid	694945.5 m	0 m	0 %
Spherical with local earth radius	693998.2 m	947.3 m	0.14 %
Spherical with mean earth radius	693258.6 m	1689.9 m	0.24 %
Plane surface distance	692589.8 m	2355.7 m	0.34 %

Table 3.2: Differences in distance between de Bilt (52.100,5.176)° and Edinburgh (55.950,-3.187)° with the great cirle distance over the WGS84 ellipsoid as reference value. Values are calculated on a sphere with reduced radius for de Bilt, a sphere with the earth mean radius, and the plane surface distance. The plane surface distance is taken as the distance on the plane tangent to the earth's surface at de Bilt. The second point is taken at the projection of the point on the ellipsoid on this plane. This would be a logical choice for a planar calculation as a cross bearing based on angles remains valid for this projection.

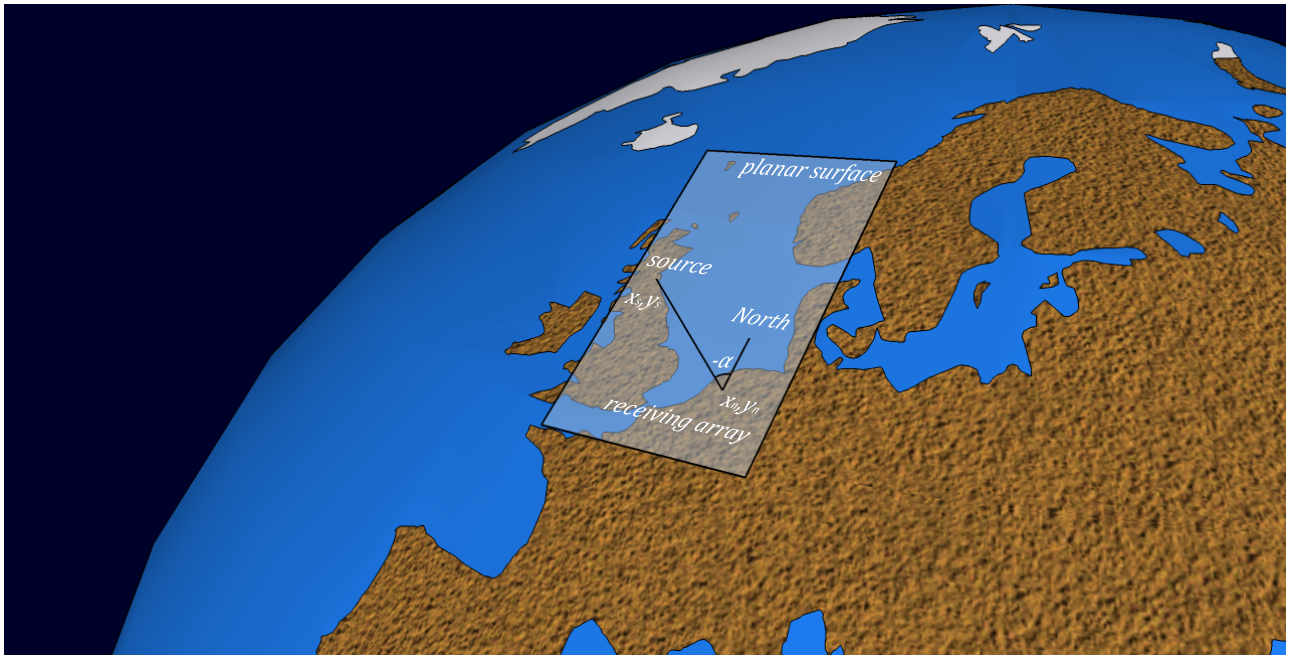


Figure 3.8: A planar surface is projected onto the earths surface. Note that the back azimuth at the receiver is displayed as a negative angle (clockwise is positive).

3.3.1 Cross bearing on the planar surface

A-model of source location

The model of observations gives the observations (back azimuth and detection time) as a function of the source location (x_s and y_s) and event time (t_s)is given by:

$$E\begin{bmatrix}\alpha_n \\ t_n \\ \vdots\end{bmatrix} = A\left(x_s, y_s, c, t_s\right) = \begin{bmatrix}\arctan\left(\frac{x_s - x_n}{y_s - y_n}\right) \\ \frac{1}{c}\sqrt{\left(x_n - x_s\right)^2 + \left(y_n - y_s\right)^2} + t_s \\ \vdots\end{bmatrix} \tag{3.25}$$

On the left side one can find the observations back azimuth (α_n) and detection time (t_n) at array n for N arrays. On the right hand side these are related to the source location (x_s, y_s), array location (x_n, y_n), the source time (t_s) and the celerity (c). This model can be extended by adding extra lines for as many as arrays at which the event is detected. Detection at at least two arrays is needed to solve for the unknown parameters.

The presented A model makes three assumptions: The source location in latitude and longitude, the source time, and the celerity are assumed equal for all detections within an event. For each assumption cases can be identified for which this does not hold. For one of the expected infrasound sources in the chosen frequency domain, sonic booms, a varying source location is expected. An aircraft in supersonic flight behaves as a line source over the supersonic flight path. The assumption of equal celerity is doubtful as each detected signal has a different atmospheric path and therefore has travelled at different celerities. The range in celerity can be rather large, from 220 m/s to 360 m/s (Evers & Haak, 2007). The source time can vary for each detection when a source emits a signal over a longer timespan. Two important sources that are expected, sonic booms and gas flares, both emit signals over a longer timespan. If one falsely assumes a single estimate for celerity and source time the source locations are found to move to the far field. For a network of arrays which are globally in the same direction from the source, a far field solution for the location will smooth out celerity differences by generating a large total travel time. E.g., for a source on the North Sea and infrasound arrays in the Netherlands the solution will move away from the Netherlands, towards Greenland.

Because the assumptions on equal source time and celerity were found to be invalid for the used dataset it is decided to abandon the estimation of celerity and source time in this step and focus on source location only. This requires only the assumption of constant source location. Although this not the case for sonic booms it is assumed that the variation in source location is relatively small with respect to the distance between the source and array. The model is simplified to estimate only latitude and longitude from back azimuth observations:

$$E\begin{bmatrix}\alpha_n \\ \vdots\end{bmatrix} = A\left(x_s, y_s, c, t_s\right) = \begin{bmatrix}\arctan\left(\frac{x_s - x_n}{y_s - y_n}\right) \\ \vdots\end{bmatrix} \tag{3.26}$$

Linearised A-model of source location

As the A-model is not linear, it cannot be solved directly. Therefore, a linearised A-model is constructed of the form $E(\Delta y) = \partial_x A(x^0)\Delta x$. This model will be used to adjust an initial estimate of the source location and time to the least-squares estimate. The linearised A-model assuming constant source latitude, longitude, time and celerity is:

$$E\begin{bmatrix}\Delta\alpha_n \\ \Delta t_n \\ \vdots\end{bmatrix} = \begin{bmatrix}\frac{\partial\alpha_n}{\partial x_s} & \frac{\partial\alpha_n}{\partial y_s} & 0 & 0 \\ \frac{\partial t_n}{\partial x_s} & \frac{\partial t_n}{\partial y_s} & \frac{\partial t_n}{\partial c} & 1 \\ \vdots & \vdots & \vdots & \vdots\end{bmatrix}\begin{bmatrix}\Delta x_s \\ \Delta y_s \\ \Delta c \\ \Delta t_s\end{bmatrix} \tag{3.27}$$

The linearised A-model assuming constant source latitude and longitude is:

$$E\begin{bmatrix}\Delta\alpha_n \\ \vdots\end{bmatrix} = \begin{bmatrix}\frac{\partial\alpha_n}{\partial x_s} & \frac{\partial\alpha_n}{\partial y_s} \\ \vdots & \vdots\end{bmatrix}\begin{bmatrix}\Delta x_s \\ \Delta y_s\end{bmatrix} \tag{3.28}$$

As input for both linearised A-models the equations of the model of observation (3.27) are differentiated towards source parameters x_s, y_s and c:

$$\frac{\partial\alpha_n}{\partial x_s} = \frac{1}{1 + \left(\frac{x_s - x_n}{y_s - y_n}\right)^2}\,\frac{1}{y_s - y_n} \tag{3.29}$$

$$\frac{\partial\alpha_n}{\partial y_s} = \frac{1}{1 + \left(\frac{x_s - x_n}{y_s - y_n}\right)^2}\,\frac{x_s - x_n}{y_s - y_n} \tag{3.30}$$

$$\frac{\partial t_n}{\partial x_s} = \frac{1}{c}\frac{x_s - x_n}{\sqrt{\left(x_s - x_n\right)^2 + \left(y_s - y_n\right)^2}} \tag{3.31}$$

$$\frac{\partial t_n}{\partial y_s} = \frac{1}{c}\frac{x_s - x_n}{\sqrt{\left(y_s - y_n\right)^2 + \left(y_s - y_n\right)^2}} \tag{3.32}$$

$$\frac{\partial t_n}{\partial c} = -\frac{1}{c^2}\sqrt{\left(x_n - x_s\right)^2 + \left(y_n - y_s\right)^2} \tag{3.33}$$

Least squares iteration

Finally the least squares estimator of the source location is given by (Teunissen, 1995):

$$\Delta y = y - A\left(x^i\right) \tag{3.34}$$

$$Q_x = \left[\partial_x A\left(x^i\right) * Q_y^{-1} \partial_x A\left(x^i\right)\right]^{-1} \tag{3.35}$$

$$x^{i+1} = x^i + Q_x \partial_x A\left(x^i\right) * Q_y^{-1} \Delta y \tag{3.36}$$

The obtained estimate x^{i+1} can be put back into Equation (3.34) such that the solution converges to the actual least-squares estimate $\hat{x}$. For the initial estimate x^0 we take the average coordinates of the matching windows used while matching the detections of the different arrays (described in Section 3.2). The initial estimate of the source time can be taken as the average of the record times minus their corresponding travel distance divided by the expected sound speed. The three equations are looped i times, until $x^{i+1} - x^i$ becomes sufficiently small. For the least squares estimate, it is sufficient to set the weight matrix $W = Q_y^{-1}$ in Equations (3.35) and (3.36) to the identity matrix I_{2n}. This will give all observations an equal weight of one. For a best unbiased estimate the weight matrix is chosen equal to the inverse of the variance matrix of the vector of observables (Teunissen, 1995) $W = Q_y^{-1}$.

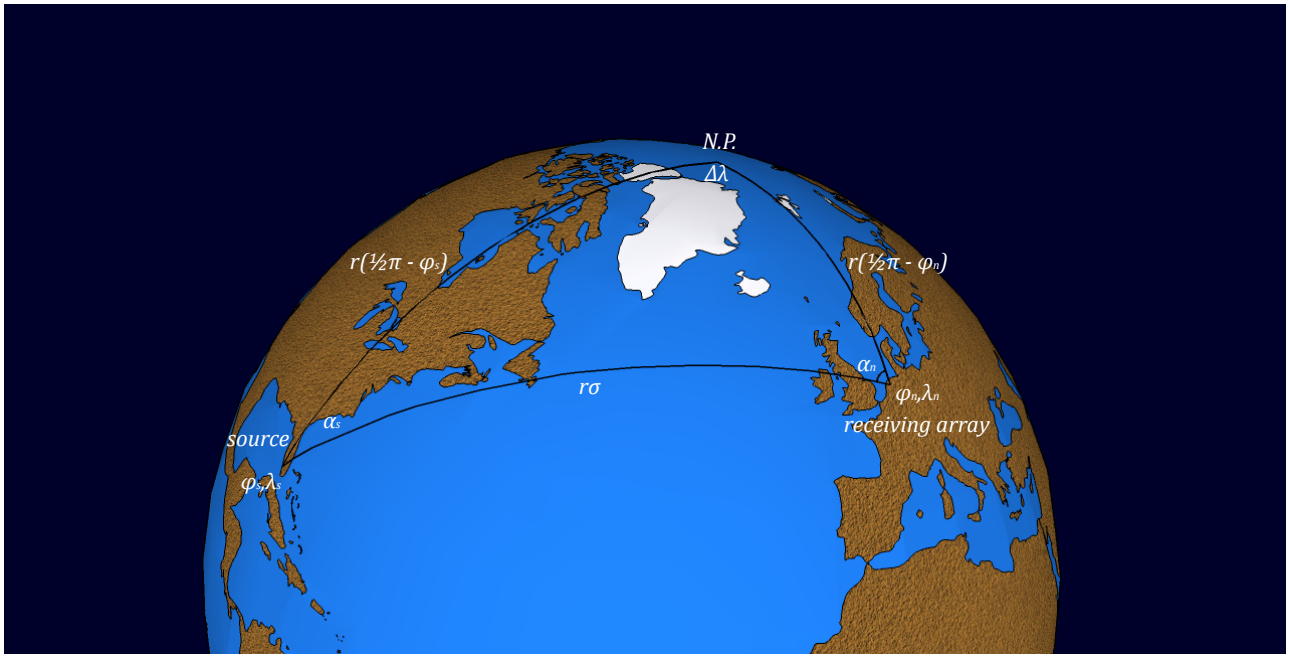


Figure 3.9: A spherical triangle displayed on the earth's surface. The triangle is formed by the three points of the source, the receiving array and the north pole. The triangle sides touching the north pole are given as a function of the latitude φ and the earth's radius r. The distance between source and receiver is given by the arc length in radians (σ) times the earth radius. The angle at the north pole is the difference in longitude between the source and receiver ($\Delta\lambda = \lambda_s - \lambda_n$). The two other angles are the azimuths at the source (α_s) and at the receiver (α_n). Note that the values of α_n and $\lambda_s - \lambda_n$ will be negative in the given triangle (angles from $-\pi$ to $+\pi$).

3.3.2 Cross bearing on the sphere

We will first derive some basic spherical relations after which the cross bearing problem will be solved by least-squares iteration of a linearised a-model, similar to the planar case.

Spherical relations

The spherical relations are illustrated by the spherical triangle, Figure 3.9. Similar to the sine and cosine rules for planar triangles, sine (Equation (3.37)) and cosine rules (Equation (3.38) and (3.39)) (Bakker *et al.*, 1995) exist for spherical triangles:

$$\frac{\sin a}{\sin A} = \frac{\sin b}{\sin B} = \frac{\sin c}{\sin C} \tag{3.37}$$

$$\cos a = \cos b\ \cos c + \sin b\ \sin c\ \cos A \tag{3.38}$$

$$\cos A = -\cos B\ \cos C + \sin B\ \sin C\ \cos a \tag{3.39}$$

Here A, B and C depict the angles and a, b and c the opposite sides of the triangle as shown in Figure 3.9. The sine and cosine rules allow us to solve the direct and inverse problem discussed in the next two sections.

The direct problem

Computation of a point coordinate and related azimuth from a given starting coordinate, azimuth and distance is commonly referred to as the direct problem in geodetics. The problem is solved by the sine and cosine rule, Equation (3.37) and (3.38). From Figure 3.9 we set angle A at the shown receiver array n, angle B at the source and angle C at the north pole such that these values become $A = \alpha_n$, $B = \alpha_s$ and $C = \lambda_s - \lambda_n = \Delta\lambda$. With an earth radius r, the length of the sides become $a = r(\frac{\pi}{2} - \varphi_s)$, $b = r(\frac{\pi}{2} - \varphi_n)$ and c is the distance between source and receiver $r\sigma$. The entire triangle is scaled down by dividing r from all lengths for simplicity. This turns Equation (3.37), and (3.38) into:

$$\frac{\cos\varphi_s}{\sin\alpha_n} = \frac{\cos\varphi_n}{\sin\alpha_s} = \frac{\cos\sigma}{\sin\Delta\lambda} \tag{3.40}$$

$$\sin\varphi_s = \sin\varphi_n\ \cos\sigma + \cos\varphi_n\ \cos\sigma\ \cos\alpha_n \tag{3.41}$$

The latitude of the source is now given by the cosine rule (3.41) as (3.42), the longitude is given by the sine rule (3.40) as (3.43) (Bakker *et al.*, 1995):

$$\varphi_s = \arcsin\left(\sin\varphi_n\ \cos\sigma + \cos\varphi_n\ \sin\sigma\ \cos\alpha_n\right) \tag{3.42}$$

$$\Delta\lambda = \arcsin\left(\frac{\sin\sigma\ \sin\alpha_n}{\cos\varphi_s}\right) \tag{3.43}$$

The inverse problem

Computation of azimuths and arc length of a great circle between two point coordinates is commonly referred to as the inverse problem in geodetics. In the cosine rule, Equation 3.38, we set $a = \sigma$, $b = \frac{\pi}{2} - \varphi_s$, $b = \frac{\pi}{2} - \varphi_n$ and $A = \Delta\lambda$. This gives the most direct solution for the inverse problem (Bakker *et al.*, 1995):

$$\sigma = \arccos\left(\cos\varphi_n\ \cos\varphi_s\ \cos\left(\lambda_s - \lambda_n\right) + \sin\varphi_n\ \sin\varphi_s\right) \tag{3.44}$$

This method is found to give rounding problems if the points are close or almost antipodal (errors in the order of centimetres for double precision). If this is undesirable, the special case of the Vincenty formula for the sphere should be used (Vincenty, 1975):

$$\sigma = \arctan\left[\frac{\sqrt{\left(\cos\varphi_n\ \sin\left(\lambda_n - \lambda_s\right)\right)^2 + \left(sin\varphi_n\ \cos\varphi_s - \cos\varphi_n\ \sin\varphi_s\ \cos\left(\lambda_n - \lambda_s\right)\right)^2}}{\sin\varphi_n\ \sin\varphi_s + \cos\varphi_n\ \cos\varphi_s\ \cos\left(\lambda_n - \lambda_s\right)}\right] \tag{3.45}$$

For this study the precision of the spherical cosine rule is sufficient. From the arc length the detection time can be calculated as the source time plus the travel time, which is the arc length divided by the celerity:

$$t_n = t_s + \frac{r}{c}\sigma \tag{3.46}$$

For the azimuth α we make use of the law of sines for spherical trigonometry (Equation (3.37)) where we set $a = \alpha_n$, $A = \frac{\pi}{2} - \varphi_s$, $b = \Delta\lambda$ and $B = r\sigma$ to get (Bakker *et al.*, 1995):

$$\alpha_n = \arcsin\left(\frac{\cos\varphi_s\ \sin\left(\lambda_s - \lambda_n\right)}{\sin\sigma}\right) \tag{3.47}$$

Quadrant adjustments should be made, i.e. when $\varphi_n > \varphi_s$ subtract the result from respectively $-\pi$ and $+\pi$ for a negative or positive α. Furthermore this formula can have inaccuracies near an azimuth of 90°. To overcome these problems it can be more convenient to eliminate σ and use (Snyder, 1987):

$$\alpha_n = \arctan\left(\frac{\cos\varphi_s\ \sin\left(\lambda_s - \lambda_n\right)}{\cos\varphi_n\ \sin\varphi_s - \sin\varphi_n\ \cos\varphi_s\ \cos\left(\lambda_n - \lambda_s\right)}\right) \tag{3.48}$$

A-model of source location

The model of observations gives the observations (back azimuth and detection time) as a function of the source location (φ_s and λ_s) and event time (t_s):

$$E\begin{bmatrix}\alpha_n \\ t_n \\ \vdots\end{bmatrix} = A\left(\varphi_s, \lambda_s, c, t_s\right) \tag{3.49}$$

On the left side one can find the observations back azimuth (α_n) and detection time (t_n) at array n for N arrays. On the right hand side these are related to the source location (φ_s, λ_s), the source time (t_s) and the celerity (c). This model can be extended by adding extra lines for as many as arrays at which the event is detected. Detection at at least two arrays is needed to solve for the unknown parameters.

The presented A model makes three assumptions: The source location in latitude and longitude, the source time, and the celerity are assumed equal for all detections within an event. For each assumption cases can be identified for which this does not hold. For one of the expected infrasound sources in the chosen frequency domain, sonic booms, a varying source location is expected. An aircraft in supersonic flight behaves as a line source over the supersonic flight path. The assumption of equal celerity is doubtful as each detected signal has a different atmospheric path and therefore has travelled at different celerities. The range in celerity can be rather large, from 220 m/s to 360 m/s (Evers & Haak, 2007). The source time can vary for each detection when a source emits a signal over a longer timespan. Two important sources that are expected, sonic booms and gas flares, both emit signals over a longer timespan. If one falsely assumes a single estimate for celerity and source time the source locations are found to move to the far field. For a network of arrays which are globally in the same direction from the source, a far field solution for the location will smooth out celerity differences by generating a large total travel time. E.g., for a source on the North Sea and infrasound arrays in the Netherlands the solution will move away from the Netherlands, towards Greenland.

Because the assumptions on equal source time and celerity were found to be invalid for the used dataset it is decided to abandon the estimation of celerity and source time in this step and focus on source location only. This requires only the assumption of constant source location. Although this not the case for sonic booms it is assumed that the variation in source location is relatively small with respect to the distance between the source and array. The model is simplified to estimate only latitude and longitude from back azimuth observations:

$$E\begin{bmatrix}\alpha_n \\ \vdots\end{bmatrix} = A\left(\varphi_s, \lambda_s\right) \tag{3.50}$$

Linearised A-model of source location

As in the planar case, the A-model is not linear and it cannot be solved directly. Therefore, a linearised A-model is constructed of the form $E(\Delta y) = \partial_x A(x^0)\Delta x$. This model will be used to adjust an initial estimate of the source parameters to the least-squares estimate. The linearised A-model assuming constant source latitude, longitude, time and celerity is given by:

$$E\begin{bmatrix}\Delta\alpha_n \\ \Delta t_n \\ \vdots\end{bmatrix} = \begin{bmatrix}\frac{\partial\alpha_n}{\partial\varphi} & \frac{\partial\alpha_n}{\partial\lambda} & 0 & 0 \\ \frac{\partial t_n}{\partial\varphi} & \frac{\partial t_n}{\partial\lambda} & \frac{\partial t_n}{\partial c} & 1 \\ \vdots & \vdots & \vdots & \vdots\end{bmatrix}\begin{bmatrix}\Delta\varphi_s \\ \Delta\lambda_s \\ \Delta c \\ \Delta t_s\end{bmatrix} \tag{3.51}$$

The A-model assuming constant source latitude and longitude is given by:

$$E\begin{bmatrix}\Delta\alpha_n \\ \vdots\end{bmatrix} = \begin{bmatrix}\frac{\partial\alpha_n}{\partial\varphi} & \frac{\partial\alpha_n}{\partial\lambda} \\ \vdots & \vdots\end{bmatrix}\begin{bmatrix}\Delta\varphi \\ \Delta\lambda\end{bmatrix} \tag{3.52}$$

This is the A-model that is used for the locations estimated in this research. Celerity and source time are estimated in a separate step, using the source location estimate as input.

As input for both linearised A-models the equations for the inverse problem are differentiated towards φ, λ and c:

$$\frac{\partial \alpha_n}{\partial \varphi_s} = \frac{\cos\varphi_n \sin(\lambda_s - \lambda_n)}{\cos^2\varphi_n \cos^2\varphi_s (1 + \cos^2(\lambda_n - \lambda_s)) + 2\cos\varphi_n \cos\varphi_s \sin\varphi_n \sin\varphi_s \cos(\lambda_n - \lambda_s) - \cos^2\lambda_n - \cos^2\varphi_s} \tag{3.53}$$

$$\frac{\partial \alpha_n}{\partial \lambda_s} = \frac{\sin\varphi_n \cos^2\varphi_s - \cos\varphi_n \cos\varphi_s \sin\varphi_s \cos(\lambda_s - \lambda_n)}{\cos^2\varphi_n \cos^2\varphi_s (1 + \cos^2(\lambda_n - \lambda_s)) + 2\cos\varphi_n \cos\varphi_s \sin\varphi_n \sin\varphi_s \cos(\lambda_n - \lambda_s) - \cos^2\lambda_n - \cos^2\varphi_s} \tag{3.54}$$

$$\frac{\partial t_n}{\partial \varphi_s} = -\frac{r}{c} \frac{-\cos\varphi_n \ \sin\varphi_s \ \cos(\lambda_n - \lambda_s) + \sin\varphi_n \ \sin\varphi_s}{\sqrt{1 - (\cos\varphi_n \ \cos\varphi_s \ \cos(\lambda_n - \lambda_s) + \sin\varphi_n \sin\varphi_s)^2}} \tag{3.55}$$

$$\frac{\partial t_n}{\partial \lambda_s} = -\frac{r}{c} \frac{\cos\varphi_n \ \cos\varphi_s \ \sin(\lambda_n - \lambda_s)}{\sqrt{1 - (\cos\varphi_n \ \cos\varphi_s \ \cos(\lambda_n - \lambda_s) + \sin\varphi_n \sin\varphi_s)^2}} \tag{3.56}$$

$$\frac{\partial t_n}{\partial c} = -\frac{r}{c^2}\,\sigma \tag{3.57}$$

Least squares iteration

Finally the least squares estimator of the source location is given by (Teunissen, 1995):

$$\Delta y = y - A\left(x^i\right) \tag{3.34}$$

$$Q_x = \left[\partial_x A\left(x^i\right)^* Q_y^{-1} \partial_x A\left(x^i\right)\right]^{-1} \tag{3.35}$$

$$x^{i+1} = x^i + Q_x \partial_x A\left(x^i\right)^* Q_y^{-1} \Delta y \tag{3.36}$$

The obtained estimate x^{i+1} can be put back into Equation (3.34) such that the solution converges to the actual least-squares estimate $\hat{x}$. For the initial estimate x^0 we take the average coordinates of the matching windows used while matching the detections of the different arrays (described in Section 3.2). The initial estimate of the source time can be taken as the average of the record times minus their corresponding travel distance divided by the expected sound speed. The three equations are looped i times, until $x^{i+1} - x^i$ becomes sufficiently small. For the least squares estimate, it is sufficient to set the weight matrix $W = Q_y^{-1}$ in Equations (3.35) and (3.36) to the identity matrix I_{2n}. This will give all observations an equal weight of one. For a best unbiased estimate the weight matrix is chosen equal to the inverse of the variance matrix of the vector of observables (Teunissen, 1995) $W = Q_y^{-1}$. In the presented research the event location is estimated on the basis of the back azimuths of all detections, possibly from different atmospheric returns, that can be related to this single event. The events do not contain enough detections to come to a reliable variance on back azimuth per event. Therefore, a single variance on back azimuth is estimated for each array from the back azimuth distributions of multiple detections within events. This variance describes the variation in back azimuth over multiple detections of a single event.

3.4 Applied filters

Four filtering moments are included in the processing of array detections, see the flowchart in Figure 3.1. First the individual receiver recordings are restricted to the same frequency band, Section 3.4.1 and 3.4.2. Second, during frequency-wavenumber analysis the recorded detections are restricted to apparent velocities above 319 m/s, Section 3.4.3. After these two steps a list of infrasound detections over this frequencyband and bounded to these apparent velocities is obtained. A third filtering step is applied to clean this list from further noise and local sources. This filter is based on a maximum apparent velocity, Section 3.4.3, Fisher ratio, Section 3.4.4 and on back azimuth, Section 3.4.5. After the source locations are estimated a final filter is applied to check on the celerity, Section 3.4.6. An overview of the back azimuth and frequency content of signals over 2007 at DBN, DIA and EXL are shown before and after filtering in figure 3.10. The percentage of detections filtered out is given in Table 3.3.

	DBN	DIA	EXL
maximum apparent velocity	0.4%	4.2%	48.5%
Fisher ratio	71.8%	90.4%	47.2%
back azimuth	53.8%	0.0%	0.0%
total	92.9%	91.0%	90.6%

Table 3.3: Percentage of detections removed by the applied filters. The percentage of the detections that would be removed when applying a single filter (shown left) are given for each array. The combined effect of all filters is shown as 'total'. The effect of filters on frequency and minimum apparent velocity are not shown as these do not simply discard detections but are integrated in the processing. No filter on back azimuth is applied for DIA and EXL.

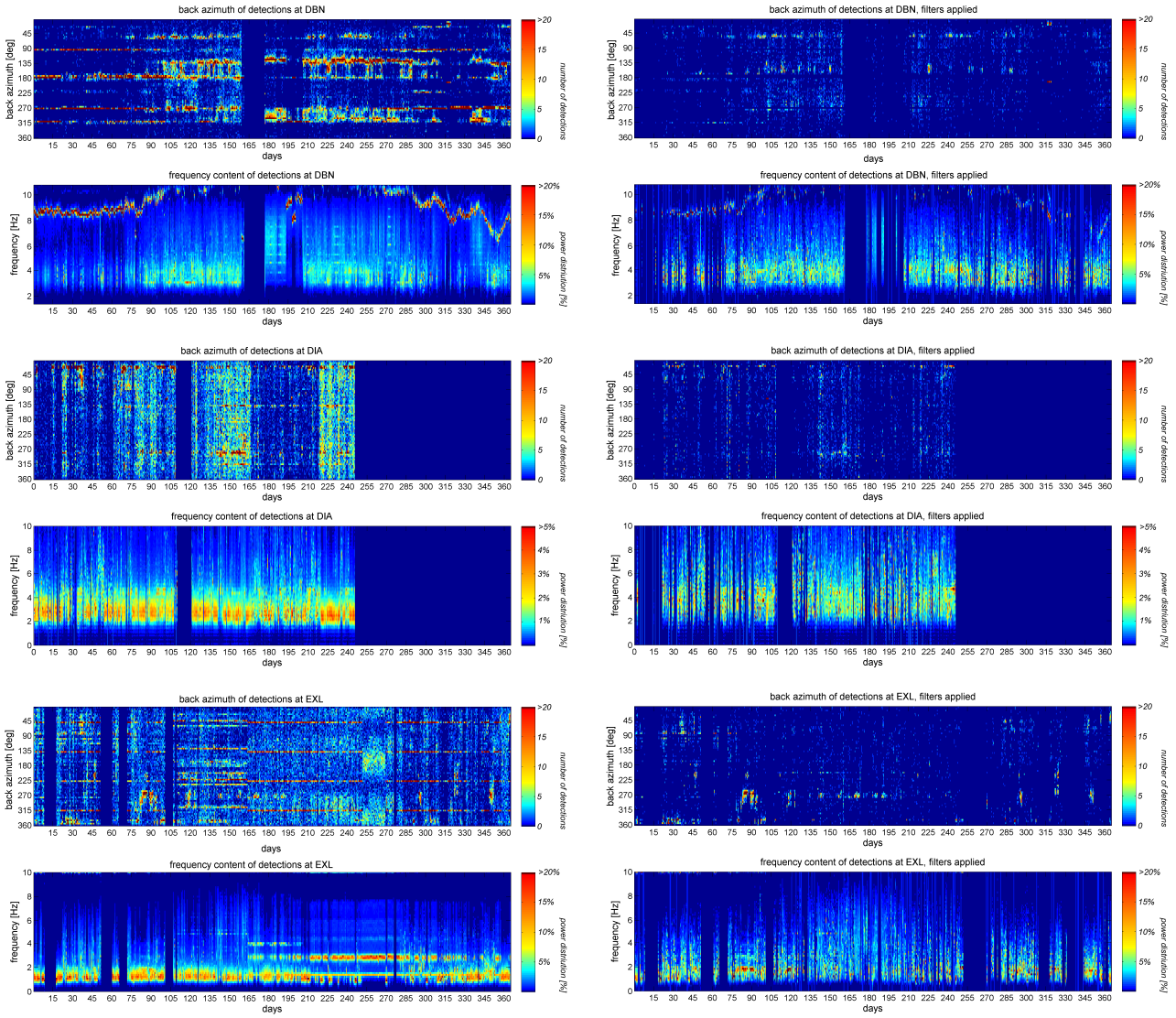


Figure 3.10: Back azimuth and frequency content of signals detected at DBN, DIA and EXL over 2007. On the left hand side the detections before filtering are plotted, on the right hand side the detections are plotted which remain after filtering on maximum apparent velocity, Fisher ratio and back azimuth. For DBN, in the back azimuth plot small bands are visible where the amount of detections is significantly higher than in other directions. Four of these directions could be associated with local sources, see Section 3.4.5. The peaks at these back azimuths are removed after filtering. Also for DBN a strong peak in the frequency spectrum is visible. The frequency of this peak seams to increase during the summer. This relatively high frequency phenomenon is possibly related to local sources and is strongly reduced after filtering. At DIA the distribution of the detections is much more homogeneous. Also at this array some peaks are visible in the back azimuth which are effectively removed by the applied filters. The frequency content of detections at DIA seems to increase slightly after filtering. The peaks in the back azimuth of the detections at EXL are associated with high apparent velocity, see Section 3.4.3, and are removed after filtering. From day 165 multiple bands are visible in the frequency spectrum. These are not visible in the filtered frequency spectrum. For all arrays the total number of detections has reduced by more than 90% after filtering.

3.4.1 Frequency range determined by array specifications

Although the microbarometers that form the receivers of the KNMI infrasound arrays are capable of recording from 0.002 to 50 Hz (Evers, 2008), which is well beyond the frequency range of interest, the geometry of the array can be a limiting factor in determining the direction of the incoming signal. Using the frequency domain Fisher analysis a method is devised to obtain the frequency range over which we can clearly determine the incoming direction of a signal. The probability of correct detection as a function of frequency is calculated for the DBN, DIA and EXL arrays (Figure 3.13). A 95% probability of correct detection is chosen as boundary values for the usable frequency range for each array (Table 3.4).

As stated in Section 3.1.2 the Fisher statistic in the frequency domain is given by:

$$P_s(\omega,\vec{p}) = \left| \frac{1}{N} \sum_{n=1}^{N} G(\omega,\vec{r_n}) e^{-i\omega\vec{p}\cdot\vec{r_n}} \right|^2 \tag{3.8}$$

$$P_t(\omega) = \frac{1}{N} \sum_{n=1}^{N} |G(\omega,\vec{r_n})|^2 \tag{3.9}$$

$$F(\omega,\vec{p}) = \frac{P_s(\omega,\vec{p})}{P_t(\omega) - P_s(\omega,\vec{p})}\ (N-1) \tag{3.10}$$

The signal power P_s is that part of the total received power P_t that is coherent with slowness vector $\vec{p}$. $G(\omega,\vec{r_n})$ is the received signal for each array n out of N arrays at frequency w and array location $\vec{r_n}$. Now, the signal power P_s is set to one for the slowness vector corresponding to the signal's back azimuth and incident angle. The total received power P_t then consists of this coherent part of unity and an incoherent noise contribution P_N. For simplicity, the signal is assumed to have an incident angle of zero. This means that the signal comes directly from above the array, time shifts of the signal between receivers are zero and the slowness vector is zero. For a signal power of one and an incident angle of zero Equations (3.8), (3.9), (3.10) reduce to:

$$P_s(\omega,\vec{p}) = \left| \frac{1}{N} \sum_{n=1}^{N} e^{-i\omega\vec{p}\cdot\vec{r}} \right|^2 \tag{3.58}$$

$$P_t(\omega) = 1 + P_N \tag{3.59}$$

$$F(\omega,\vec{p}) = \frac{\left| \frac{1}{N} \sum_{n=1}^{N} e^{-i\omega\vec{p}\cdot\vec{r}} \right|^2}{1 + P_N - \left| \frac{1}{N} \sum_{n=1}^{N} e^{-i\omega\vec{p}\cdot\vec{r}} \right|^2}\ (N-1) \tag{3.60}$$

Using the array response function

$$R(\omega,\vec{p}) = \left| \frac{1}{N} \sum_{n=1}^{N} e^{-i\omega\vec{p}\cdot\vec{r_n}} \right|^2 \tag{3.61}$$

the Fisher statistic can be conveniently obtained for the entire slowness grid:

$$F(\omega,\vec{p}) = \frac{R(\omega,\vec{p})}{1 + P_N - R(\omega,\vec{p})}\ (N-1) \tag{3.62}$$

In analogy to the array response function, Equation 3.62 gives a "Fisher response"for a signal of zero incident angle.

We consider a signal on a single frequency and a given Fisher ratio. The noise power is determined by the response at zero slowness which corresponds to the previously chosen incident angle of zero. The signal power P_S now becomes unity and the noise power P_N is given from Equation (3.10) as a function of the F-ratio F and the number of receivers in the array N:

$$P_N = \frac{N-1}{F} + 1 \tag{3.63}$$

The Fisher response for a 2.5Hz signal with an F-ratio of 2 is calculated for the infrasound array in Exloo and shown in Figure 3.11. This Figure shows the theoretical Fisher statistic for a signal coming directly from above the array at different slowness vectors.

To give a guideline to what detections contain valuable information we state that for each detection the obtained incident direction is required to be within 10 degrees of the true direction. This is in line with the azimuthal tolerance set in Section 3.2. Equation (3.2) shows that a 10 degree deviation in direction corresponds to a deviation in slowness up to 0.0005 s/m. For the previously created response the true slowness vector is at the point $(0, 0)s/m$. The alternative hypothesis is set to contain all other slowness vectors with a horizontal velocity of 300 m/s to infinity. Both boundaries are shown as black circles in Figure 3.11.

Each Fisher value in Figure 3.11 corresponds to a probability density function from which the actual measurement will be drawn. Following (Evers, 2008), this probability density function has a F-distribution with two degrees of freedom and a non-centrality parameter; $F(v_1, v_2, \lambda)$. The first degree of freedom v_1 is given by the number of samples in the measurement: $v_1 = T - 1$. The second degree of freedom v_2 is given by the number of samples and the number of receivers: $v_2 = T(N - 1)$. The non-centrality parameter λ is a function of the Fisher ratio from Figure 3.11 and v_1: $\lambda = (F - 1) * v_1$. An example of the probability distributions for noise and an Fisher ratio of 1.5 an 3 is shown in Figure 3.12.

In frequency-wave number analysis, the maximal Fisher value over the slowness grid is selected and this point determines the measured incident angle and back azimuth. The realized value of each point in the slowness grid is drawn from the F-distribution determined by the true F-ratio as shown by the Fisher response. The probability P that the selected maximum lies inside the previously defined 10 degree margin is therefore given by the probability that the maximum realized value inside the peak is larger than all realized values outside the peak. The maximum value inside the peak is governed by the sum of the probabilities for each point to give a certain Fisher value (F) times the probability for all other points in the peak to be smaller or equal. This gives the following relation:

$$P\left(\left(x_{\text{inside}}\right)_{\max} = F\right) = \sum_{n=1}^{N_{\text{inside}}} \left(P(x_n = F) \cdot \prod_{m=1}^{N_{\text{inside}}} P\left(x_{m \neq n} \leq F\right) \right) \tag{3.64}$$

$$P\left(x_{\text{outside}} \leq F\right) = \sum_{n=1}^{N_{\text{outside}}} \left(\int_0^F P(x_n = F)\ dF \right) \tag{3.65}$$

$$P(\text{peak correctly recognized}) = P\left(\left(x_{\text{inside}}\right)_{\max} = F\right) \times P\left(x_{\text{outside}} \leq F\right) \tag{3.66}$$

Where N_{inside}, N_{outside} give the number of points in- and outside of the peak, x is the realized Fisher ratio for this point and F is the overall maximum Fisher ratio that is realized inside the peak. These relations are applied to the array response of three infrasound arrays in the Netherlands; de Bilt (DBN), Deelen (DIA) and Exloo (EXL). As the results will improve with increasing Fisher ratio the worst case is selected, corresponding to the minimum Fisher ratio that was required for a detection to be saved. This is a Fisher ratio of 2, Section 3.1.5. Using the statement that 95% of all registrations should be within the 10 degree interval a minimum and maximum frequency is obtained. These boundaries are given in Table 3.4 and Figure 3.13. The lower bound is determined by the width of the main lobe in the response which can be seen for DBN in Figure 3.14. The upper bound is determined by the occurrence of aliasing which can be seen for EXL in Figure 3.15.

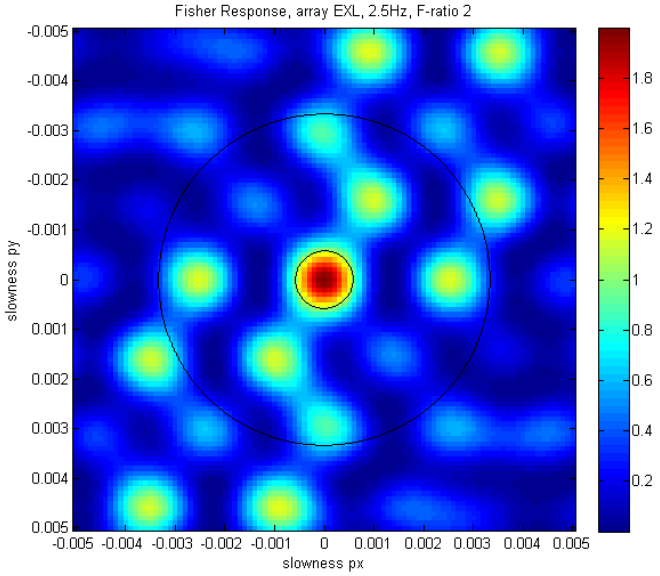


Figure 3.11: Fisher values for a slowness grid, calculated for a 2.5 Hz signal with a F-ratio of 2 and incident angle zero. Noise is incoherent for all slowness vectors. The slowness vector is directly related to the direction of the incoming signal. The top of the figure corresponds to the north, left to the west. The signal arrives from the zenith, resulting in a peak at the centre. The small black circle contains all slowness vectors corresponding to this true direction and a maximum of 10 degrees deviation. The large black circle contains all slowness vectors with a sound velocity above 300 m/s. Each point in the grid gives the F-ratio calculated on the recorded signal for this specific direction. As the signal arrives from above the array the central point gives the largest F-ratio of 2. Side lobes are visible as local maxima.

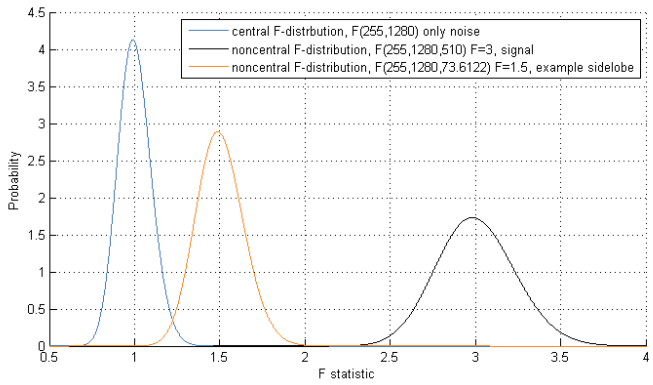


Figure 3.12: Central F-distribution, typical for noise and two non-central F-distributions with F=1.5 and F=3. Each plotted line corresponds to a recorded signal. Due to noise on the signal the F-ratio is not deterministic, but drawn from the shown distribution. The vertical axis shows the probability that the determined F-ratio has a value as given on the horizontal axis.

array	lower limit [Hz]	upper limit [Hz]
DBN	2.83	13.27
DIA	0.16	> 100
EXL	0.89	8.79

Table 3.4: Frequency boundaries for the de Bilt (DBN), Deelen (DIA) and Exloo (EXL) infrasound arrays. Within these boundaries 95% of the detections have a direction within 10 degrees of the correct direction. The signal has a Fisher ratio of 2

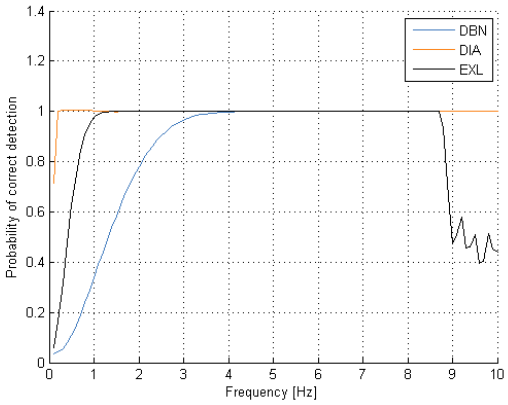


Figure 3.13: Probability of correct detection of a beam for the de Bilt (DBN), Deelen (DIA) and Exloo (EXL) infrasound arrays. A detection is counted as correct when the incident angle is within 10 degrees of the true direction. The applied signal has a Fisher ratio of 2

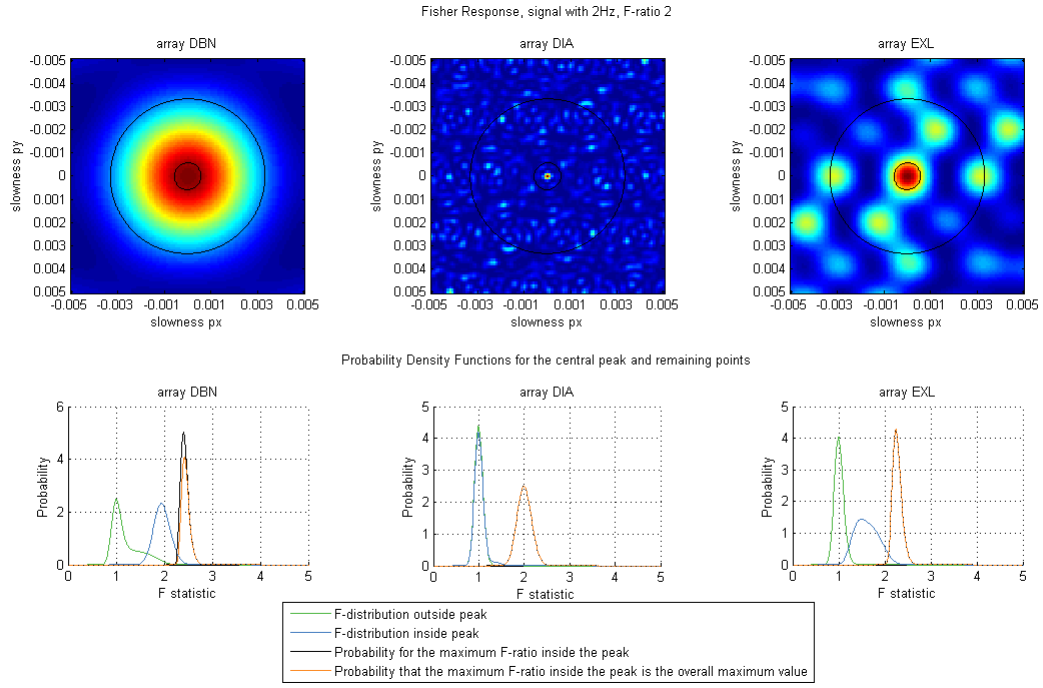


Figure 3.14: Response for a synthetic signal of 2.5 Hz, F-ratio of 2 and incident angle zero for DBN, DIA and EXL. The low frequency signal of 2.5 Hz is chosen to show the difference with a signal of 10 Hz shown in Figure 3.15. Top: Array response in Fisher ratio. The small black circle contains all slowness vectors corresponding to a 10 degree region around the true direction of the signal (zenith). The large black circle contains all slowness vectors for sound velocities above 300 m/s. Bottom: Probability density functions, i.e. the probability that the F-ratio on the horizontal axis is registered. Outside the peak refers to the area within the large black circle excluding the small black circle. Inside the peak refers to the area within the small black circle. Probabilities are also plotted for the value of the maximum F-ratio within the peak and the probability that this value is the maximal value of all values within the 300 m/s boundary. For DBN, the total probability that the maximum F-ratio inside the peak is also the overall maximum value is less than 1 (compare with the probability density curve for the maximal F-ratio inside the peak which does have a total probability of one). This shows that for this frequency it is probable that some registrations will give a direction more than 10 degrees from the true value. In fact, this is the case for more than 5% probability and therefore a 2.5 Hz signal is rejected for DBN, see Table 3.4. For DIA and EXL the probability curves for the maximum F-ratio inside the peak and the overall maximum value do coincide. This shows that the maximum value inside the peak will be the overall maximum value and the signal's incoming direction is determined correctly. For DIA the distribution inside the peak looks very similar to that outside the peak. As can be seen in the array response, a large area in the defined peak behaves similar to the values outside the peak, which gives a similar distribution. A very small area within the defined peak gives a high response, which results in the tail trailing to the right of the main peak. EXL shows a distinct probability curve for the area inside the defined peak with respect to the area outside the peak.

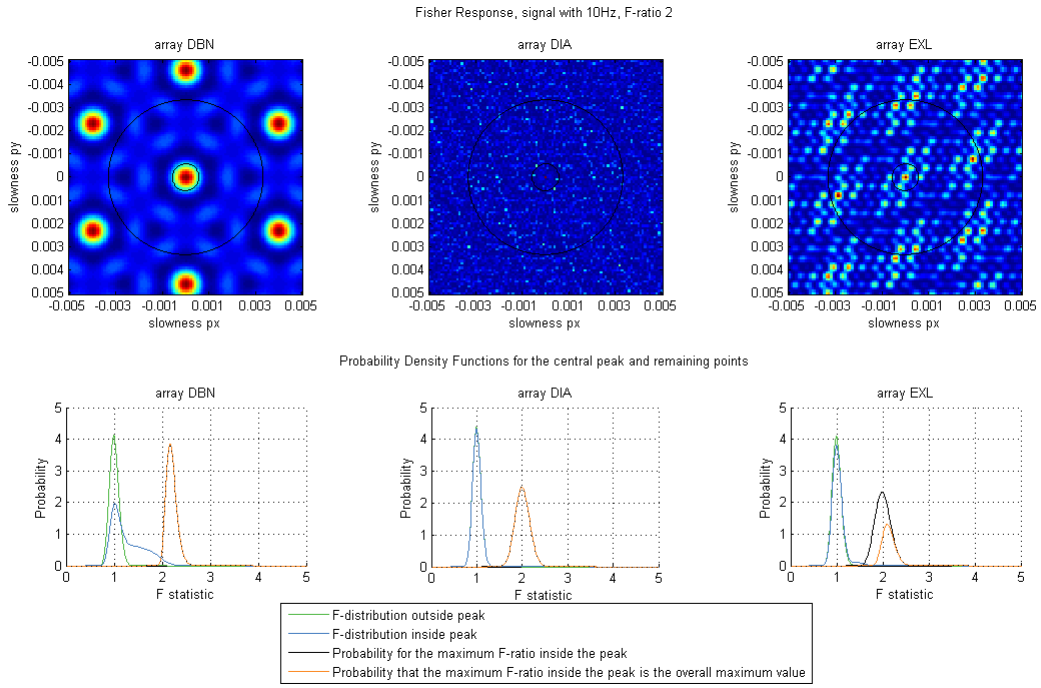


Figure 3.15: Response for a synthetic signal of 10 Hz, F-ratio of 2 and incident angle zero for DBN, DIA and EXL. This figure is equal to that of Figure 3.14 but shows a higher frequency signal. Top: Array response in Fisher ratio. Bottom: Probability that the F-ratio on the horizontal axis is registered. Outside the peak refers to the area within the large black circle excluding the small black circle. Inside the peak refers to the area within the small black circle. Probabilities are also plotted for the value of the maximum F-ratio within the peak and the probability that this value is the maximal value of all values within the 300 m/s boundary. For EXL, the total probability that the maximum F-ratio inside the peak is also the overall maximum value is less than 1 (compare with the probability density curve for the maximal F-ratio inside the peak which does have a total probability of one). This shows that for this frequency it is probable that some registrations will give a direction more than 10 degrees from the true value. This is caused by the existence of side lobes (i.e. secondary peaks) in the array response, an effect known as aliasing. For this frequency it is probable that some registrations will be given at these secondary peaks and will give a wrong signal direction.

3.4.2 Frequency range determined by the source of interest

Apart from the technical possibilities of the used arrays, the selected frequency range is a choice to select the sources one would like to investigate. All three KNMI infrasound arrays which are used in this research are capable of detecting signals between 2.8 and 8.8 Hz, Section 3.4.1. This frequency region corresponds to that of man made sources, Table 2.2. It is undesirable to detect different sources at the individual arrays and then try to combine these to common events. Therefore the frequency range determined by array specification is further limited to the frequency range of man made sources, between 1 and 10 Hz. The resulting frequency range is presented in Table 3.5. This selection is implemented by applying a butterworth filter on the recorded signals, before they are inserted into frequency wavenumber analysis.

array	lower limit [Hz]	upper limit [Hz]
DBN	2.83	10.00
DIA	1.00	10.00
EXL	1.00	8.79

Table 3.5: Frequency range applied for the de Bilt (DBN), Deelen (DIA) and Exloo (EXL) infrasound arrays. This is a combination of the 95% correct direction boundaries as shown in Table 3.4 and the frequency range associated with man made sources , Table 2.2. Only man made sources are concidered as these sources can be detected by all three arrays. It is undesirable to detect different sources at the individual arrays and then try to combine these to common events.

3.4.3 Apparent velocity

The second filter is based on the apparent velocity of the signal. The lower boundary of the apparent velocity is applied during Frequency-Wavenumber analysis by selecting only detections above a minimum apparent velocity. The upper boundary of the apparent velocity is applied after frequency-wavenumber analysis and is therefore included in the third filtering step. A sound speed ranging from 325 m/s to 355 m/s, a maximal wind speed of 5 m/s and an incident angle ranging from 10 to 90 degrees is applied. This results in a lower and upper boundary of 320 m/s and 2044 m/s for the apparent velocity.

An acoustic signal will travel over the array at a velocity determined by the sine of the incident angle and the sound speed at the local conditions. The local sound speed is approximated by the sound speed for an ideal gas combined with the local wind vector (Gossard & Hooke, 1975):

$$c_{eff} = \sqrt{\gamma RT} + \vec{n} \cdot \vec{u} \tag{3.67}$$

The adiabatic index γ is taken to be the ideal diatomic gas value 1.4000. R is the ratio of the molar gas constant with the molar mass in kilograms per mole which is 287.09 for air. The wind contribution is formed by the inner product of the direction of the incoming signal $\vec{n}$ and the wind speed vector $\vec{u}$. For a range of temperatures T from 263 K to 313 K (-10 degree Celsius to +40 degree Celsius) the sound speed varies from 325 m/s to 355 m/s. According to the logarithmic wind speed profile (Figure 3.16), the wind at 1 meter altitude will most probably be less than 40% of the wind at 10 m. The measured wind speed distribution at 10 m altitude over 2007 is shown in Figure 3.17. Directly at the surface, where the infrasound detections are made, the wind speed will be even less. Over 2007 the potential wind at 10 meter exceeded the 12 m/s in only 0.25% of the full year of hourly measurements for de Bilt and in 0.61% of the hourly measurements for Deelen. For Exloo, no data was available. The maximum expected wind speed at the surface is taken at 40% of this, which is 5 m/s. Taking both minimum temperature and maximum wind into account, the sound speed is not expected to decrease below 320 m/s.

The relation between incident angle and apparent velocity was discussed in Section 3.1.1:

$$c_{app} = \frac{c}{\sin i} \tag{3.68}$$

At an incident angle of 90 degrees sound is travelling parallel to the surface. In this case, the apparent velocity is equal to the local sound speed. At 0 degrees the sound arrives from above the array and will be detected by all receivers simultaneously. The apparent velocity becomes infinity. If the incident angle becomes small it also becomes difficult to determine a correct azimuth direction which is needed to make a cross bearing of the source. A large amount of detections of very high apparent velocity is visible in the histogram of Figure 3.18 for Exloo and Deelen. Due to the limited resolution of the slowness grid, which relates apparent velocity to back azimuth, these high velocities result in back azimuths of 45, 135, 225 and 315 degrees. This effect is clearly visible in the back azimuth plot of Figure 3.10 for Exloo. To remove this unwanted effect a threshold is set for a maximum apparent velocity which corresponds to an incident angle of 10 degrees. With a sound speed ranging from 325 m/s to 355 m/s and an incident angle ranging from 10 to 90 degrees we now have an acceptable apparent velocity region of 319 m/s to 2044 m/s. This removes the data spike at 14000 m/s. The lower boundary of 319 m/s also removes a substantial amount of detections. These low apparent velocities could be related to the detection of sidelobes on the slowness grid during frequency-wavenumber analysis.

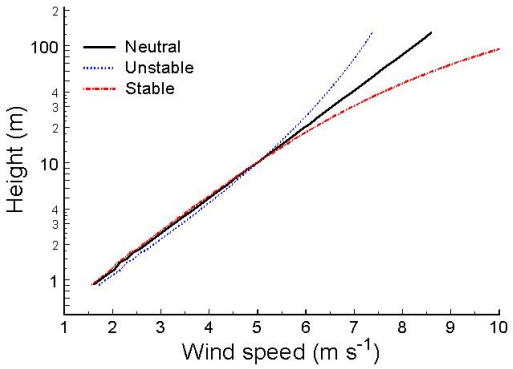


Figure 3.16: Logarithmic wind speed profile from the surface (0m) up to 100m (KNMI, 2008).

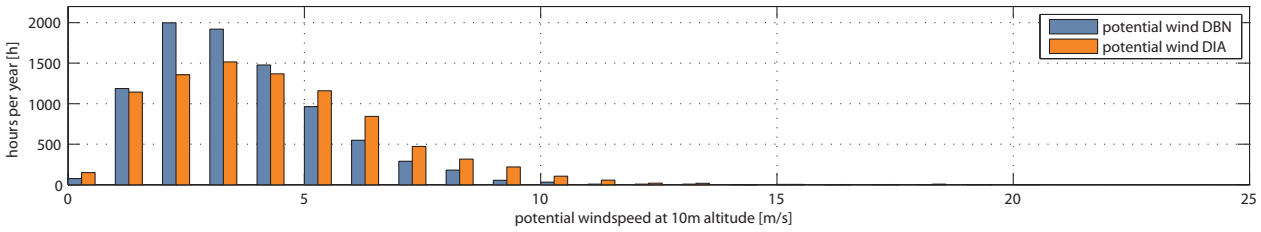


Figure 3.17: Wind speed distribution at de Bilt and Deelen in 2007. Histogram based on one per hour measurements, corrected to the wind speed at 10 m height over open land with roughness length 0.03 meter (KNMI, 2007). For Exloo, no data was available.

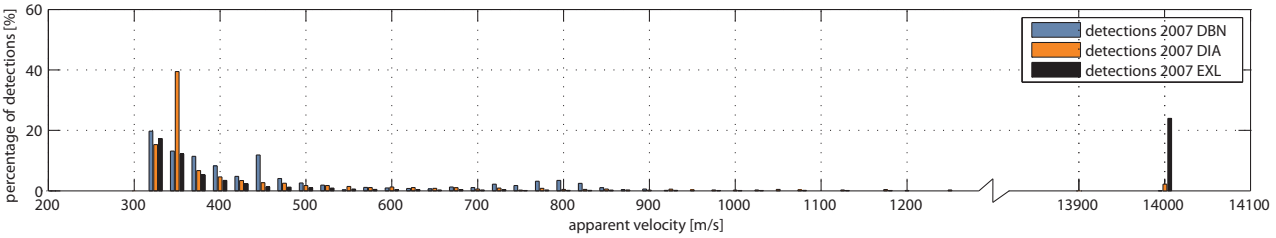


Figure 3.18: Histograms of apparent velocity of detections at DBN, DIA and EXL over 2007. A distinct peak is visible at 14000 m/s which is related to signals with very small incident angles. During frequency-wavenumber analysis only detections with an apparent velocity larger than 319 m/s were selected, therefore lower apparent velocities are not present in this histogram. All detections have a Fisher ratio larger than 2, filters on back azimuth are not applied.

3.4.4 Fisher ratio

A very direct way of cleaning the list of detections is by using the Fisher ratio, which is directly related to the signal to noise ratio. Although only a low threshold is needed (less than $F = 2$) to remove almost all noise it is chosen to set the threshold at a higher $F = 4$. By increasing the required Fisher ratio the research is focused on locating more powerful sources. This reduces the amount of detections considerably which is important in reducing the amount of uncorrelated detections that will be combined into events.

The Fisher ratio is directly related to the signal to noise ratio, Equation (3.18), therefore the probability that a detection is signal and not noise increases with the Fisher ratio. Noise is expected to have a central F-distribution around $F = 1$ with two degrees of freedom $F(v_1, v_2)$ (Evers, 2008). Signal can be described by an F-distribution with two degrees of freedom and a non-centrality parameter; $F(v_1, v_2, \lambda)$. When a threshold is set for the Fisher-ratio one can calculate the probability for false detections, Table 3.6, and the probability for missed detections, Table 3.7, see also Figure 3.20. For the used number of samples per measurement ($T = 256$) the amount of false detections is very low for all thresholds above $F = 1.3$. One should be careful not to choose this threshold to low as the probability on false detections at $F = 1.3$ one still generates 24800 false detections a year (time bins of 6.4 seconds times 2 because of overlapping bins). At a threshold of $F = 1.5$ this is reduced to 28 detections a year and at a threshold of $F = 1.7$ only one false detection is made every 66 years.

All detections with a Fisher ratio larger than 2 were recorded at DBN, DIA and EXL. From the number of false detections this threshold seems to be conservative. For an array of 6 elements, a threshold of $F = 2$ records signals with a signal to noise power ratio above 1/6. There are two reasons two raise the minimum required Fisher ratio. First, a signal from a source in the far field which has at arrival low energy left is less likely to be observed at multiple infrasound arrays than if this source emitted a more powerful signal. Second, when searching the detection lists for common events, the possibility exists that detections originating from events which are not related are combined to a single event. The number of these random combinations is directly proportional to the total amount of detections that are made. If a low threshold is accepted, the large amount of detections could cause a large amount of random combinations which will then obscure the powerful events. The amount of events resulting from random combinations will be assessed in detail in Section 6.2. As this is a first research in source detection from KNMI infrasound arrays, it is decided to see what sources can be located from stronger detections first, before seeking the limits of the method. This research will therefore be focussed on signal detections with a F-ratio larger than 4. These signals give a signal to noise power ratio of 0.5 at DBN and EXL and of 0.19 at DIA (Equation (3.18)). The number of detections decreases rapidly with increasing Fisher ratio, Figure 3.19.

Although it seems less likely for a weaker signal to be observed at multiple infrasound arrays, one should be aware of the fact that strong detections are not necessarily detectable at multiple arrays. The Fisher ratio or signal to noise ratio cannot distinguish between signals from far field sources which we would like to detect and near field sources which we do not like to detect. For example, a truck passing a tunnel can create a coherent signal with limited energy. This signal may not have enough energy to travel large distances but locally it can have the same amplitude as a higher energy source in far field, of which most of the energy has already dissipated during its atmospheric journey. As the used dataset has a high amount of detections of local sources these are removed using a different filter based on back azimuth, Section 3.4.5.

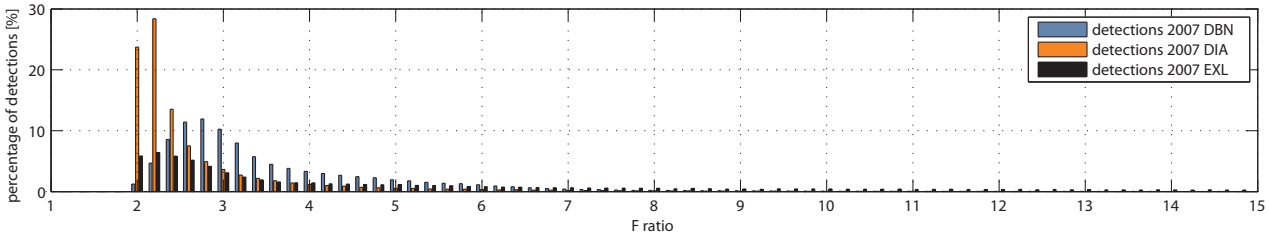


Figure 3.19: Histogram of the Fisher ratio of detections at DBN, DIA and EXL over 2007. Only detections with Fisher ratio larger than 2 were saved. Filters on maximum apparent velocity and back azimuth are not applied.

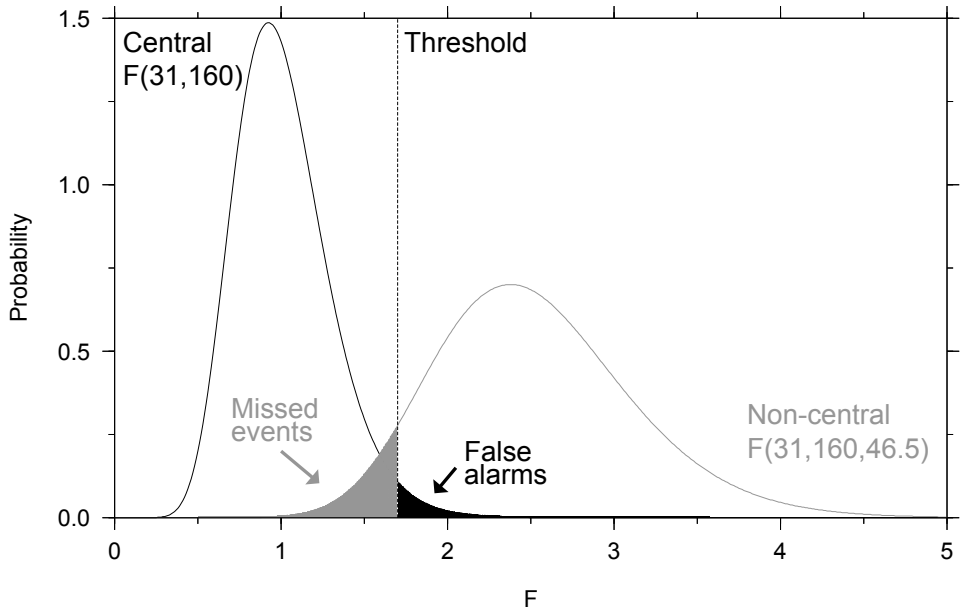


Figure 3.20: The central $F(v_1; v2)$ and non-central F-distribution $F(v_1; v_2; \lambda_{nc})$, for $T = 32$, $N = 6$ and $snr = 1.5$ as function of F (Evers, 2008). The missed event and false alarm rate are indicated by the shaded areas for a chosen detection threshold of $F = 1.7$. The first degree of freedom v_1 is given by the number of samples in the measurement: $v_1 = T - 1$. The second degree of freedom v_2 is given by the number of samples and the number of receivers: $v_2 = T(N - 1)$. The non-centrality parameter λ is a function of the detected Fisher ratio and v_1: $\lambda = (F - 1) * v_1$. For the used number of samples ($T = 256$) the normal distribution will be much more peaked, see Figure 3.12.

Threshold (F-ratio)	1.0	1.1	1.2	1.3	1.4	1.5
false detection probability	0.49	0.16	0.03	0.00	0.00	0.00

Table 3.6: False detection probability as a function of signal to noise ratio and Fisher ratio threshold. Values are calculated for an array of 6 receivers.

Threshold	Signal to noise ratio of signal				
(F-ratio)	1	2	3	4	5
1	0.00	0.00	0.00	0.00	0.00
2	0.51	0.00	0.00	0.00	0.00
3	1.00	0.50	0.00	0.00	0.00
4	1.00	1.00	0.50	0.00	0.00
5	1.00	1.00	1.00	0.50	0.00
6	1.00	1.00	1.00	1.00	0.50
7	1.00	1.00	1.00	1.00	0.99

Table 3.7: Missed detection probability as a function of signal to noise ratio and Fisher ratio threshold. Values are calculated for an array of 6 receivers and 256 samples per measurement.

3.4.5 Back azimuth

Several back azimuths where the amount of detections is significantly higher than in other directions were visible in Figure 3.10. Polar histograms plotted on top of a map show a direct relation of these spikes with spatially fixed sources as tunnels, Figure 3.21, 3.24 and 3.25. It is highly undesirable to have a large amount of detections from local sources in the data as these could cause false cross bearings with unrelated detections of the other arrays. Because of the very high number of detections from local sources at DBN it is decided to remove all detections over the back azimuths given in Table 3.8 at this array.

The effect of false cross bearings can be easily imagined if we think of a continuous source near one array. For a series of random detection at a second array we will try to make cross bearings with the continuous source, resulting in a line of false locations in the extend of the line between the first receiver and the continuous source.

In Figure 3.22 the frequency spectra of the back azimuths as found to be related to local sources are shown. Over these back azimuths the contribution of the frequencies above 8 Hz is larger than over the back azimuths where no local sources were identified. This could be an indication to lower the boundary of the frequency filter which would reduce the effect of these local sources. However, the local sources also show signal power at lower frequencies. Filtering on back azimuth will therefore remain necessary. To keep a consistent approach at all three arrays it is chosen to keep the frequency band from 1 to 10 Hz.

Figure 3.21: Polar histogram of the back azimuth of detections at de Bilt. All detections have Fisher ratio larger than 4 and are filtered on apparent velocity. Bins of 5 degrees width are used. Note that bins containing more than 1000 detections are limited to a length corresponding to 1000 detections. The bins from 45 to 55 degrees contain 3255 detections, from 90 to 95 degrees 13133 detections, from 185 to 190 degrees 2456 detections and from 265 to 270 degrees 3063 detections. A clear correlation is visible between the number of detections and four local sources. At 50 degrees back azimuth one finds a bicycle tunnel under the N237, at 92 degrees an overpass of the N225 over the A28, at 185 degrees a bicycle tunnel under the A28 and at 270 degrees an overpass of the A27 over the N237. The 5 story building of KNMI is located between 0 and 45 degrees back azimuth effectively blocking detections in this direction. Map source: OpenStreetMap (2009).

local source	from [deg]	to [deg]
bicycle tunnel N237	47	51
bicycle tunnel A28	181	194
N225/A28 overpass	88	95
A27/N237 overpass	268	272

Table 3.8: Filtered back azimuths and the identified local sources for the de Bilt infrasound array.

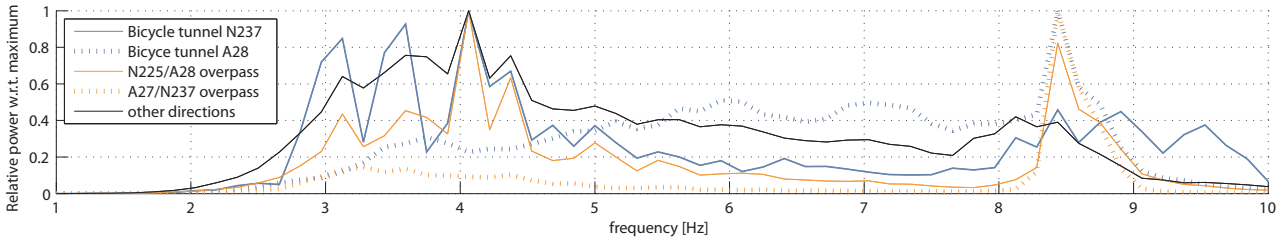


Figure 3.22: Average frequency spectra of detections over the back azimuths as found to be related to local sources near DBN. The back azimuth are specified as shown in table 3.8. The black line shows the average frequency spectrum of detections over all other back azimuths. The local sources show a high frequency contribution above 8 Hz which is much smaller over other back azimuths. The power is normalized to the maximum power.

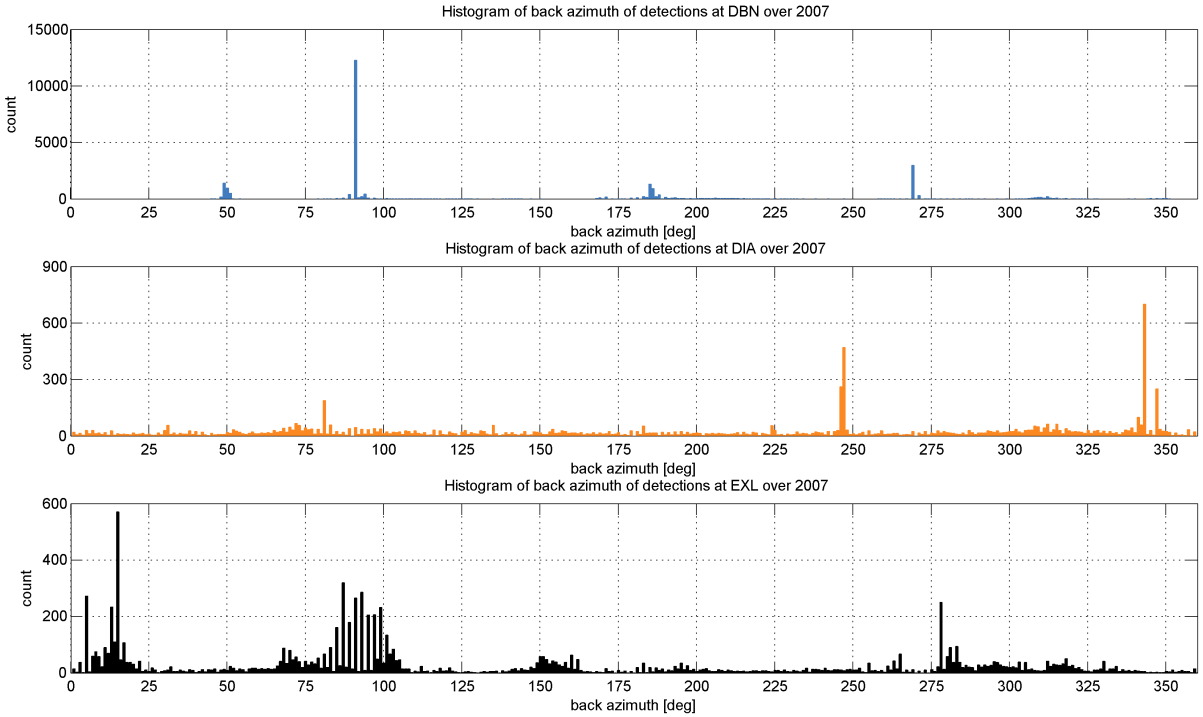


Figure 3.23: Histogram of the back azimuth of detections at de Bilt, Deelen and Exloo over 2007. For DBN, a very large number of detections is visible around 50, 92, 185 and 270 degrees which are related to local sources, see also Figure 3.21. For DIA, two distinct peaks are visible. These peaks did not show a relation with possible local sources visible in Figure 3.25 and are therefore not filtered. As DIA is located on a air force base these directions could be related to military activity. For EXL, the peak at 15 degrees is probably related to a local source, Figure 3.25. Due to the moderate amount of detections in this direction (about 2 per day) this peak is not removed from the dataset. In the band from 75 to 100 degrees alternating low and high detection numbers are visible, caused by the resolution of the slowness grid used in frequency-wavenumber analysis. All detections have Fisher ratio larger than 4 and are filtered on apparent velocity.

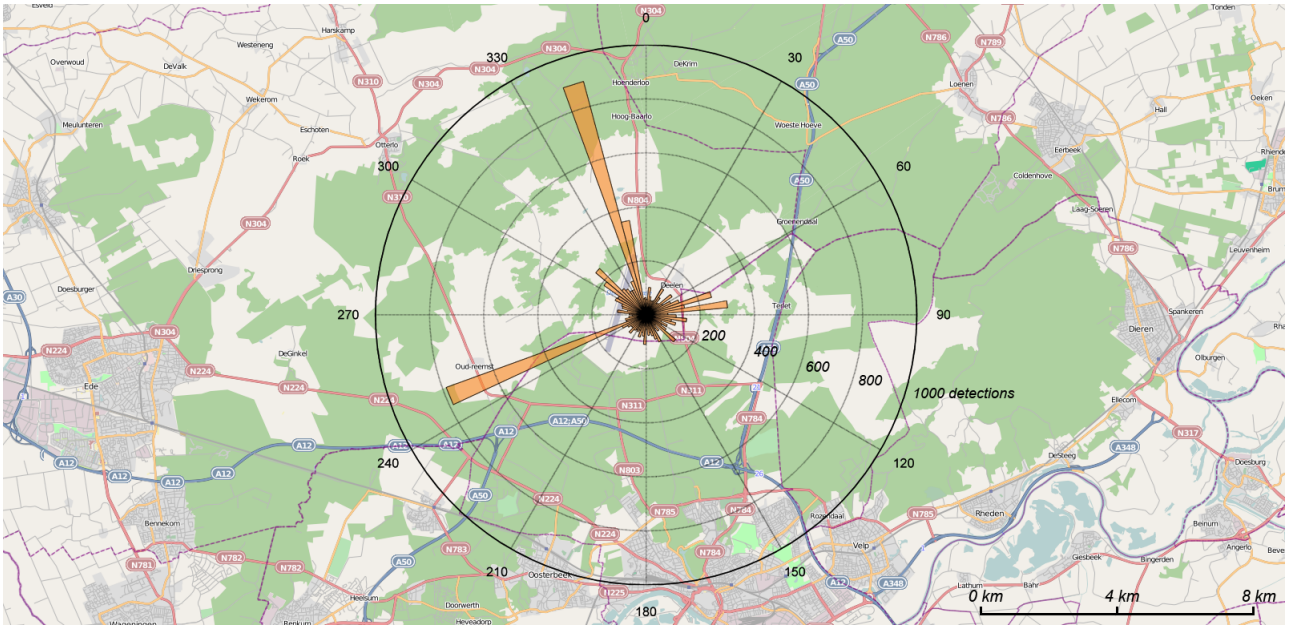


Figure 3.24: Polar histogram of the back azimuth of detections at Deelen. All detections have Fisher ratio larger than 4 and are filtered on apparent velocity. Bins of 5 degrees width are used. Two directions with significantly higher amount of detections are visible. These directions could not be directly related to local sources. The Deelen infrasound array is located on an air force base. The high amount of detections could be related to regular shooting exercises or helicopter or aircraft operation. Map source: OpenStreetMap (2009).

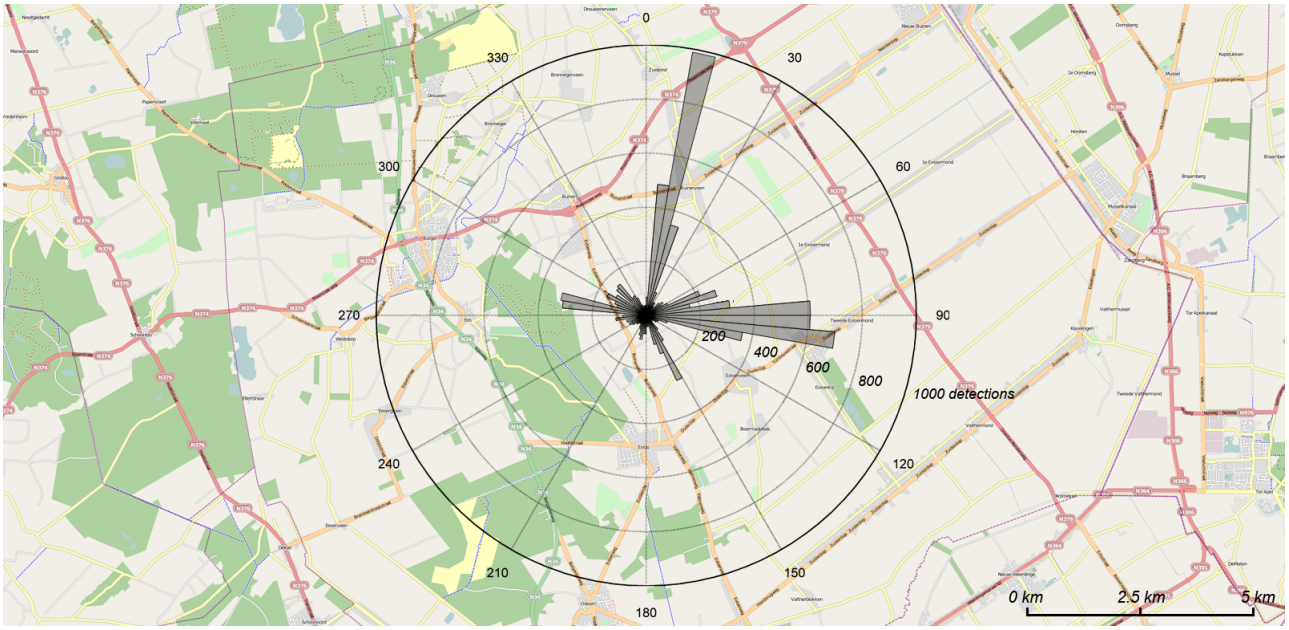


Figure 3.25: Polar histogram of the back azimuth of detections at Exloo. All detections have Fisher ratio larger than 4 and are filtered on apparent velocity. Bins of 5 degrees width are used. A large number of detections at 15 degrees back azimuth could be caused by the overpass of the N374 over the N379 which is located 6.3 km from the infrasound array. Map source: OpenStreetMap (2009).

3.4.6 Celerity

Celerity is the signal propagation velocity between source and receiver. The ratio of detection time difference and source distance difference at two arrays can be used to estimate the celerity (Equation (3.69)). The celerity of an infrasound signal gives information on through which atmospheric layer it has travelled, Table 3.9 and 3.10. The celerity was used to define the search windows in Section 3.2. After grouping and source location a second check is made to check if the celerity requirements still hold. As the source location is only an estimate but is used to calculate the celerity this filter is not set very strict, the celerity is required to be within 0 - 1000m/s. All of the events found in this research proved to fulfil this requirement, the filter is therefore more a check on the results. Events which share detections or are very close to each other in time and back azimuth were labelled as associated in Section 3.2. The celerity estimate of these associated events is used to select the most likely event for plotting.

Once a source location is estimated, source receiver distances can be calculated and an estimate of the celerity can be made from each pair of detections at two different arrays:

$$c = \frac{d_1 - d_2}{t_1 - t_2} \tag{3.69}$$

Where d_1 and d_2 are the distances from the source to the first and second array and t_1 and t_2 are the detection times at the first and second array. For more than two arrays all combinations can be made and the results combined. When there are multiple detections per array within a single event this poses a question, which of the detection pairs should be used to estimate the celerity? Three options are discussed.

First option. If all arrays would detect the same number of phases from a source one could easily calculate the celerity for all first arrivals (i.e. first arriving phase) between the arrays, for all second arrivals, etcetera. However, the signal propagation paths through the atmosphere can be such that not all phases are received at all arrays. Furthermore, even if a phase is detected at a certain array it could remain undetected due to a second stronger detection in the same time window of 256 seconds (Section 3.1.5).

Second option. A more robust method is to take all detection combinations and calculate the corresponding celerities. Now all celerities are checked to see if they can be categorized to one of the arrival types of Table 3.9. Celerities which do not correspond to these ranges are omitted. This method is very suitable to produce an overview of the different phases that are possibly detected over a large number of detections. A good algorithm will try to extract the phases in the same order as we expect them to arrive, fastest phases first. Each detection should only be categorized to one phase, as the different celerities do not permit them to arrive simultaneously. This method could result in the case that none of the celerities are valid for an event because of inaccuracy in the location estimate or detection time.

Third option. The most robust method, is to calculate a minimum and maximum celerity within the time range from the first to the last detection at all arrays. This range will always contain the true celerity for the estimated source location. A combination of multiple arrivals will almost always result in a range of celerities which is larger than the range that is expected to be physically possible. One should take care of the asymptotic behaviour of Equation (3.69) near $t_1 - t_2 = 0$. The maximum celerity can be set to infinity if we have both negative and a positive celerities. The minimum celerity is given by the minimum positive celerity. As negative celerities are not physically possible they are omitted from the results.

For the check on celerity the third option is chosen, as this will generate one minimum and one maximum celerity value which can be compared with the allowed celerity range of 0 to 1000 m/s. For the choice between associated events also the third option is applied, but now the event with celerity closest to the range of 220 to 360 m/s as shown in Table 3.9 and 3.10 is selected. In detailed analysis, it is easy to recall the omitted associated events. The second and third method are both used to investigate the celerities of the located events in Section 4.2.6.

type	abbreviation	Signal velocity
Lamb-surface wave	I_{Lm}	0.33 - 0.34 km/s
Tropopause ducted	I_w	0.30 - 0.32 km/s
Stratopause ducted	I_s	0.28 - 0.31 km/s
Ionosphere ducted	I_t	0.22 - 0.24 km/s

Table 3.9: Celerity regimes as a function of atmospheric layer (Revelle & Whitaker, 1996). Ionosphere ducted phases are often referred to as thermospheric returns (Evers & Haak, 2007; Ceplecha *et al.*, 1998; Arrowsmith *et al.*, 2008). Abbreviations as introduced by Brown (Brown *et al.*, 2002).

type	abbreviation	Signal velocity
conventional stratopause ducted	I_s	0.25 - 0.31 km/s
fast stratopause ducted	I_{sf}	0.31 - 0.36 km/s

Table 3.10: Celerity regimes for fast and conventional stratopause ducted waves as proposed by Evers (Evers & Haak, 2007)

Chapter 4

Results KNMI Infrasound arrays

The first direct result from the processing scheme as described in Chapter 3 are the estimated locations of sources as detected by the three KNMI infrasound arrays at De Bilt, Deelen and Exloo. In this Chapter an investigation is made into the characteristics of these results. In the first section, the resulting event locations and the precision of these locations is given. In the second section, various statistics of the properties of the events are shown. In the third section, the origin of the found events is investigated.

4.1 Source location

The events that are generated using the procedure presented in Chapter 3 can be detected by 2 or 3 arrays and can contain multiple detections per array. Event locations are estimated in latitude and longitude on a spherical earth. Locations are estimated using the back azimuths of the detections at each array. If multiple detections exist at one array, the average of the back azimuths of these detections is used. The location problem requires detections at at least two arrays to come to a location. Location using all information will be build up step by step. Firstly, the locations obtained when combining two out of three arrays are shown. Secondly, locations are estimated using all three arrays, but of the multiple detections per array only the first detection is used. Thirdly, the information of all three arrays and all detections per event are used to estimate the source locations shown in Figure 4.5.

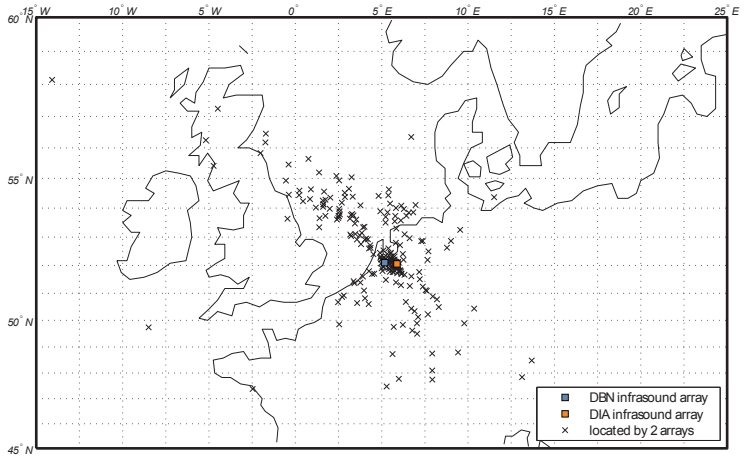
4.1.1 Location from two arrays

Before integrating three arrays, the results which can be obtained by two arrays are shown in Figure 4.1. Large numbers of events are located on the North Sea and close to the arrays. A significant number of the events that are detected by only two arrays are expected to be false. These false events are caused by detections which have by chance a back azimuth and detection time that allow them to originate from the same event, but are in fact unrelated. These infrasound events did not occur in reality and are therefore described as false. The influence of this effect is quantified after processing, during validation (Chapter 6). For now it is sufficient to know that this effect is dominant when using two out of three arrays, but is strongly reduced when each event is required to be detected by all three arrays. In the latter case 4% of all events are expected to be false (Chapter 6). The rest of this chapter will therefore discuss only the events which are detected by all three arrays.

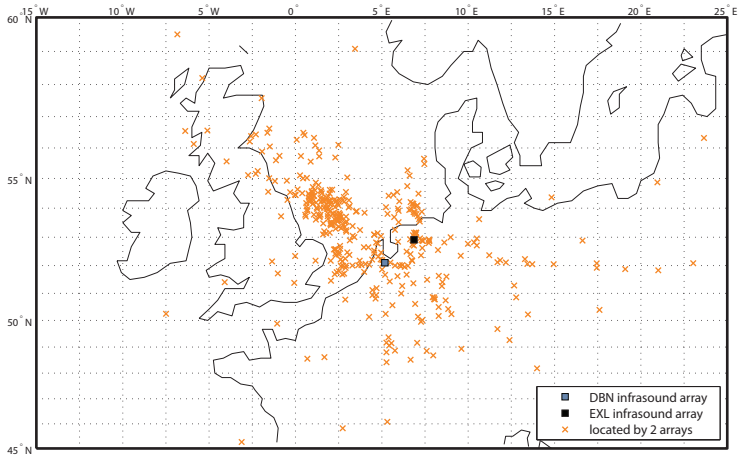
4.1.2 Location from first detections

The combination of multiple detections into events influences the estimated location and the precision of the estimation. To illustrate this effect, the location and precision will first be estimated using only one detection for each array per event. Following Arrowsmith *et al.* (2008) and Brown *et al.* (2002) the first arriving detection is used. The location is based on the back azimuth of the first detections. The precision of the location is obtained by propagating the standard deviation on the back azimuths to the location estimate.

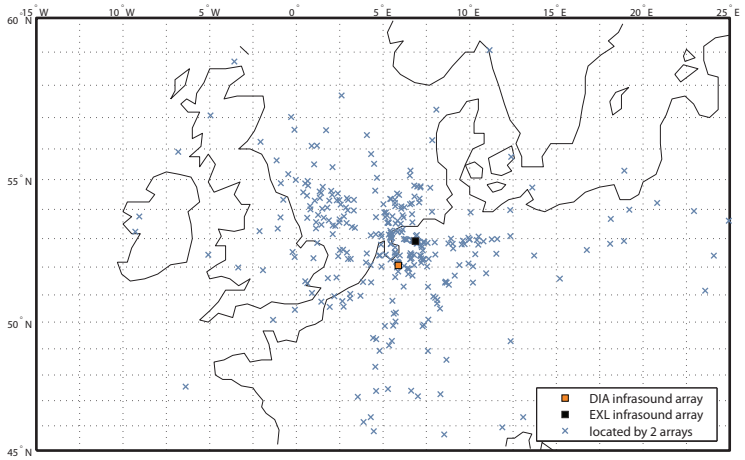
The standard deviation on the back azimuth of each detection was estimated by fitting a Lorenz profile on the Fisher detector's slowness grid (Section 3.1.4). A histogram of the standard deviations on back azimuth is



(a) Source locations of events detected by DBN and DIA, 236 locations found



(b) Source locations of events detected by DBN and EXL, 392 locations found



(c) Source locations of events detected by DIA and EXL, 336 locations found

Figure 4.1: Best unbiased estimates of source locations from infrasound arrays at de Bilt and Deelen (a), de Bilt and Exloo (b) and Deelen and Exloo (c). For Deelen detections from January 1 until the September 3 are used, while for de Bilt and Exloo the entire year 2007 is used. Consequently, the DBN-DIA and DIA-EXL plot both show locations until September 3 and the DBN-EXL plot until December 31. The combination DBN-EXL gives 259 locations in the period from January 1 to September 3.

shown in Figure 4.3. From this figure the detections at DIA seem to have a much higher standard deviation than at the other two arrays. This is opposite from what would be expected as DIA has a much larger aperture (1500 m for DIA, 80 m for DBN, 250 m for EXL) and more receivers (16 for DIA, 6 for DBN and EXL). Because of this large aperture the peak of high Fisher values on the slowness grid could become very small, up to only one pixel (this effect is visible in Figure 3.15). The smaller peak at DIA shows that DIA should be able to determine the back azimuth of a signal with a lower standard deviation than DBN and EXL. However, when the peak is contained in only a few pixels, the Lorenz fit was found to become unreliable, resulting in wrongly estimated standard deviations. As the locations estimated on the first detections serve an illustrative purpose only, a quick fix is applied. For all detections at DIA a fixed standard deviation of 0.3 degrees is assumed. This value is taken at the lower end of the standard deviation distribution that is observed for DBN and EXL.

By inserting the variances on back azimuth as weights into the least square iteration a best unbiased estimate of the event location is obtained. When the iteration converges to a solution, Equation (3.35) gives the variance-covariance matrix on the estimated location. The eigenvalues and eigenvectors of the variance-covariance matrix are used to calculate an error ellipse at a distance of two times the length of the standard deviation (i.e. 2σ), which corresponds to a 95% error ellipse for a Gaussian distribution. Both the location and the error ellipse at 2σ are plotted in Figure 4.2. This figure shows the location constructed by a specific combination of detections. A different location and error ellipse will be obtained for each combination of detections within an event.

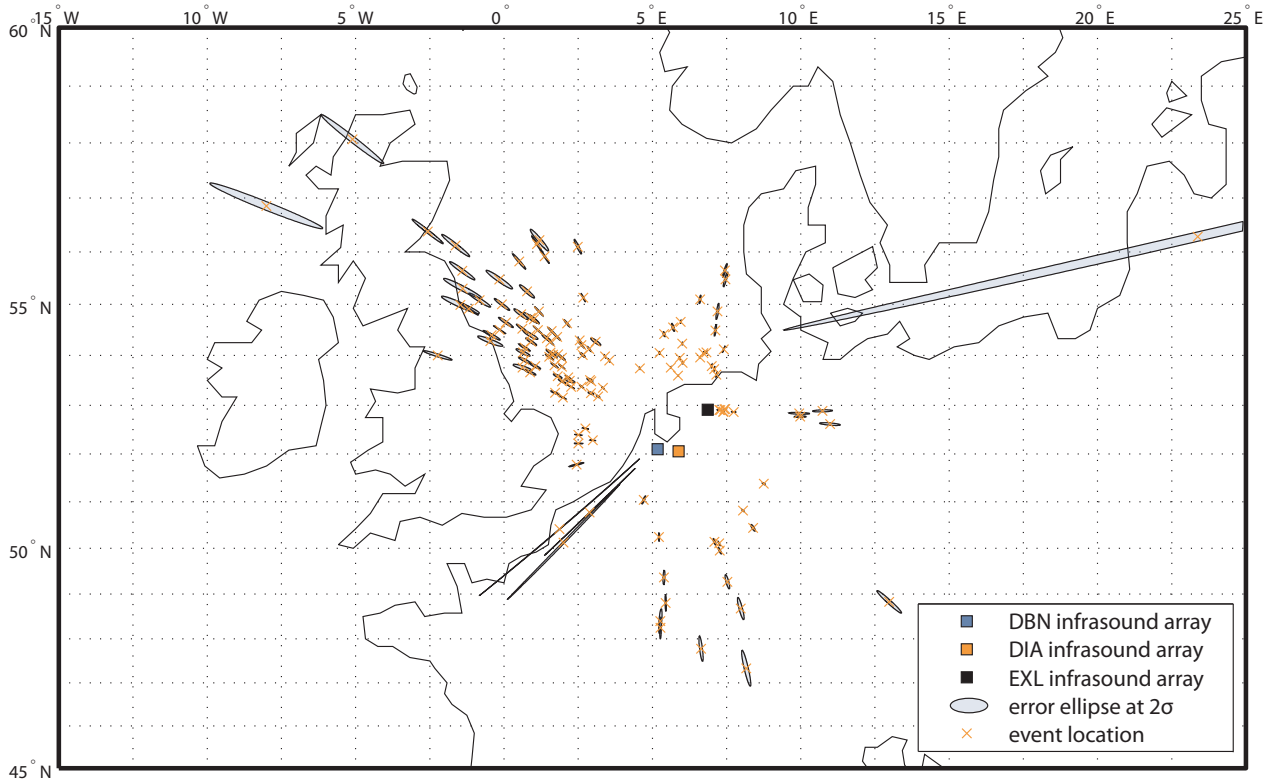


Figure 4.2: Locations as obtained by combining the first arriving detections of each event at each array. The estimated variances on back azimuth for each detection are propagated into 2σ precision error ellipses for this combination of detections. Estimates of the variance on back azimuth were not available for DIA, therefore this plot is made using a constant variance for DIA. This variance is assumed at $0.09\,\mathrm{deg}^2$ for the construction of this plot, this value should however be estimated for each detection for a proper plot of the precision. Error ellipses become larger when locations are further away from the arrays. The arrays are located roughly in a North-Eastern direction. Events that are located in a North-Eastern or South-Western direction therefore have similar back azimuths at the three arrays. As the locations are estimated by crossing these back azimuths, the distance to these points is difficult to estimate. This is visible from the elongated error ellipses of the locations near $(50.5^o, 2.5^o)$. Error ellipses are not corrected for the spherical earth, visible for the error ellipse near $(56^o, 23^o$

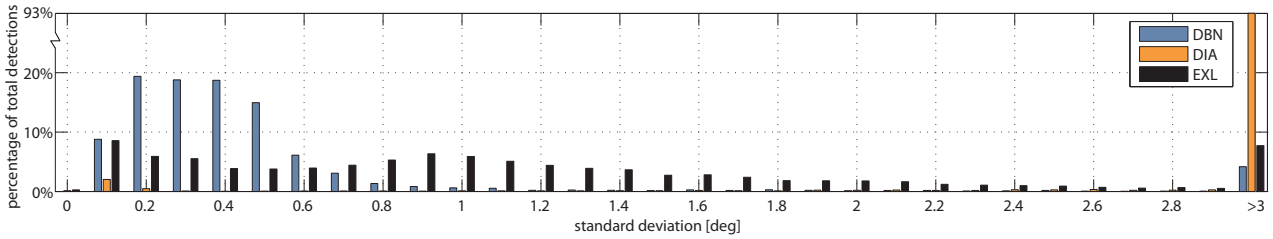


Figure 4.3: Histogram of the estimated standard deviation on back azimuth for all detections at DBN, DIA and EXL. 93% of the detections at DIA show a standard deviation larger than 3 degrees, where it is expected that they are smaller than the standard deviations for DBN and EXL. It is expected that the standard deviations for DIA are wrongly estimated. When comparing the standard deviations of individual detections at DBN and EXL, the standard deviations for EXL seem to be larger than for DBN. This is opposite from what would be expected as EXL has a larger aperture (240 m) than DBN (80 m) and the same amount of receivers (6). This can be caused by the influence of sidelobes or different frequency content of the made detections. Figure 3.13 showed that EXL suffers more from sidelobes than DBN, which can cause the fitted Lorenz function to have a broader peak, resulting in a larger estimate for the standard deviation.

4.1.3 Location from three arrays and multiple detections per array

To come to the most reliable location estimate, information of all three arrays and all detections within each event are used. Using three in stead of two arrays strongly increases the probability that the event actually occurred. Using all detections within an event in stead of only the first detection allows to make use of the information of multiple arrivals. The back azimuth of the first detection can be less accurate than the back azimuth of one of the other detections within this event. Under the influence of wind (usually strong winds in the upper atmosphere) each detection experiences a different azimuthal deviation, depending on the signal's travel path. To come to the most accurate location one would like to use the detection with the smallest azimuthal deviation. Unfortunately, it is not possible to select this best detection directly because the travel path of the signal is (yet) unknown. Therefore it is opted to weigh all detections at equal importance. The average back azimuth of all detections within an event, at each array, is used for source location.

The estimated variances on the detection's back azimuth are only valid for a single detection, not for the assembly of multiple returns which are each biased by a different azimuth deviation. In stead, the difference between the back azimuths of the detections and the average back azimuth of the event, is used to estimate the variance on back azimuth. This variance contains both the precision at which the back azimuth is determined, and the difference between the back azimuths caused by the atmospheric bias. Not all events consist of multiple detections and the amount of detections within each event is very limited. This makes an estimate of the variance for each individual event very unreliable. The total amount of events containing more than one detection do give enough information to give a distribution of the back azimuth for the average event. The variance on the back azimuth for each individual event is assumed to be equal to this average event. A histogram of the back azimuth distribution is given in Figure 4.4. From this distribution the standard deviation (σ) on back azimuth is estimated for DBN, DIA and EXL as 2.1, 1.6 and 1.6 degrees respectively. These values are larger than the average precision of the back azimuth of individual detections.

The event locations are estimated from least square iteration (Section 3.3.2) using the average of the back azimuths within an event as input. By inserting the variances on this average back azimuth as weights into the least square iteration a best unbiased estimate of the event location is obtained. When the iteration converges to a solution, Equation (3.35) gives the variance-covariance matrix on the estimated location. The eigenvalues and eigenvectors of the variance-covariance matrix are used to calculate an error ellipse at a distance of two times the length of the standard deviation (i.e. 2σ), which corresponds to a 95% error ellipse for a Gaussian distribution. This agrees with the 95% error of the distributions of back azimuths within events, which is close to 2σ, Table 4.1. Both the location and the error ellipse at 2σ are plotted in Figure 4.2.

It should be noted that the error ellipses only includes the variation which is recorded within the events, and can still contain a bias valid for all detections within the event. This is illustrated with an example. In this example, one event generates four detections of which the back azimuths deviate from the straight line direction

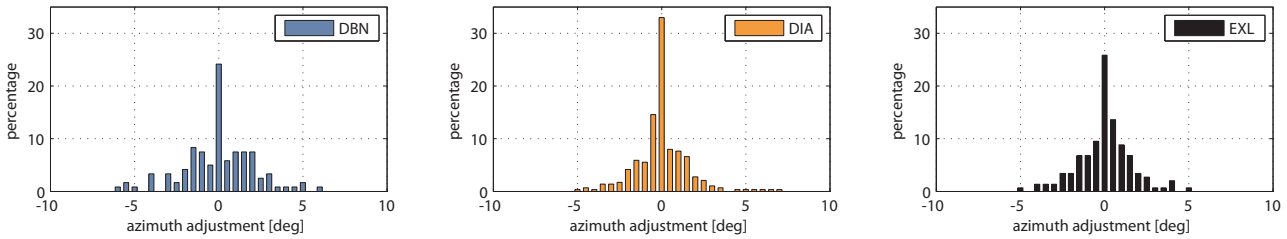


Figure 4.4: Histogram of the back azimuth distribution within all events. The distribution is obtained by taking the difference of each detection with the average azimuth of the corresponding event. Events which consisted of only one detection do not show a distribution and are therefore omitted from the histogram. 98% of the individual detection azimuths differ less than 5 degrees from the average azimuth of the event.

array	DBN	DIA	EXL
standard deviation (σ) [deg]	2.1	1.6	1.6
95% error [deg]	4.8 (2.3σ)	3.4 (2.1σ)	3.7 (2.3σ)

Table 4.1: Standard deviation and 95 % error for DBN, DIA and EXL. Values are calculated from the histograms of the azimuthal adjustment as shown in Figure 4.4.

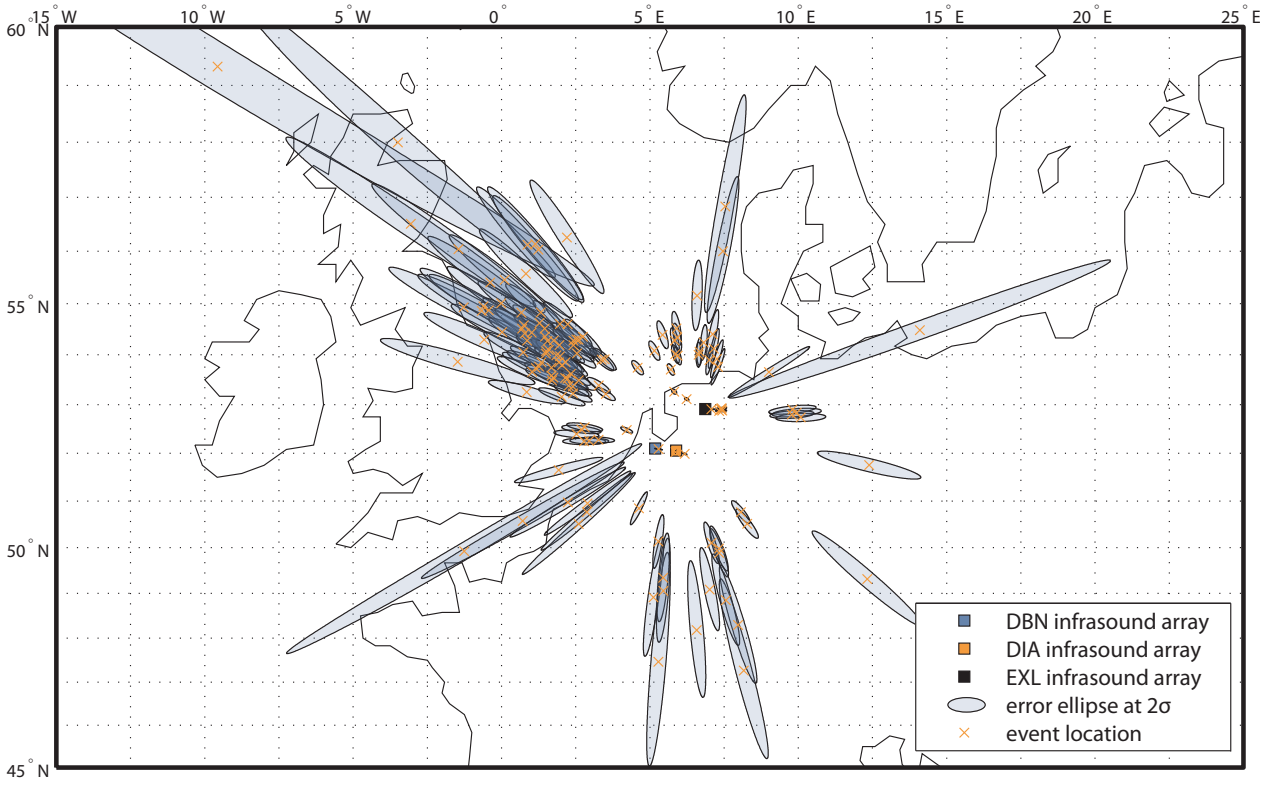


Figure 4.5: Best unbiased estimate of source locations of events detected by infrasound arrays located near de Bilt, Deelen and Exloo. Only events which were detected by all three arrays are shown. 141 events were located and occurred in the period from January 1 until September 3. Excluding days without data (see Figure 3.10) this is 0.9 events per day. Locations are given in orange with precision shown as 2σ error ellipses in transparent blue. As locations are estimated using back azimuth the direction with respect to the arrays is known much better than the distance. Locations are shifted with respect to the location obtained by first detections (Figure 4.2) and are expected to be more accurate because of the use of information of multiple detections of the same event. The error ellipses are larger as they include the range in azimuths of the multiple detections.

due to wind. These deviations can for example be 4, 5, 6 and 7 degrees. The range of these offsets (3 degrees) was modelled by the 95% error. However, the average bias of 5.5 degrees from straight line propagation is not known, and thus not included in the location estimate and precision. If the atmospheric conditions are known, this bias can be modelled by raytracing, but therefore a location estimate is needed first. In this research this location estimate is given, the correction of the atmospheric bias is recommended for further research.

4.2 Statistics on results

Various event properties are shown in more detail in this section. The event properties are described in terms of the adjustment of the average back azimuth, the distribution of the events over day, week and year, the frequency content of the events, the number of detections within events and the recorded event durations. In the final subsection the celerity will be estimated on the basis of the estimated source location and the detection times. This is compared to the celerities that are expected for different atmospheric returns from literature. All statistics are based on the events detected by all three arrays.

4.2.1 Back azimuth

In the location procedure the mean azimuths per event are adjusted for each array in such a way that the azimuths at the three arrays point to one location. A histogram of the required azimuth adjustments is shown in Figure 4.6. On average, the adjustment on the observations at EXL is smaller than the adjustment at DBN and DIA. This can be explained when one looks at the location of the infrasound arrays with respect to the main source area. The location of a source over the North Sea will change only slightly if the back azimuth at EXL is biased by 1 degree. If the back azimuth at EXL is fixed, then the back azimuths at DBN and DIA will be adjusted less than 1 degree to compensate for this change. On the contrary, if the back azimuth at either DBN or DIA is biased by 1 degree and the location is cross-located from these two arrays then the location will move a much larger distance. If the back azimuths at DBN and DIA are fixed, then this will cause an adjustment of more than 1 degree to the back azimuth at EXL. A least square adjustment of the back azimuths (as is applied) minimizes the required adjustment for all three back azimuths and will therefore prefer the smaller adjustment to the back azimuth at DBN and DIA. This geometric effect causes the observations at EXL to be adjusted less than those at DBN and DIA.

4.2.2 Time and date

The distribution of the events over the year, the days of the week and the hours of the day is shown in Figure 4.7. The distribution over the week shows that 99% of the events were detected during working days, only 1% (2 events) on Saturday and none on Sunday. The distribution over the day shows that 98% of the events were detected between 7 am and 10 pm, 2 events were detected before 7 am and one event after 10 pm. The two events on Saturday are located in the east on the map of Figure 4.5 at $(51.3°, 16.9°)$ and $(51.8°, 12.4°)$. All events near $(53°, 10°)$ were detected in the early morning, three between 7 and 8 am and one between 8 and 9 am. The distribution over the week and day strongly suggest that the events are related to human activity.

4.2.3 Signal frequency

The frequencies of the events that could be detected at three arrays are slightly lower than the frequency spectrum of all detections, see Figure 4.8. This is expected as sources in the far field will lose more higher frequency content due to attenuation than sources that are located close by. As the signals which are related to events originate from the same sources one would expect to receive similar frequency spectra for all three arrays. This is however not observed and the difference in frequency content is not yet explained. The difference in frequency content can explain the difference in the back azimuth precision for single detections as shown in Figure 4.3. A signal with higher frequency content generates a sharper peak for the Fisher ratios on the slowness grid, which resulting in a smaller standard deviation. Figure 4.8 shows that DBN is subjected to relatively more higher frequency signals than EXL. This could be a cause for the smaller standard deviation of DBN with respect to EXL.

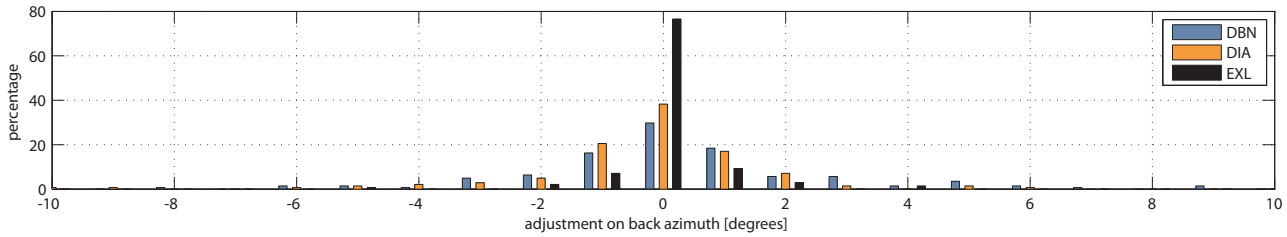


Figure 4.6: Histogram of the required azimuth adjustments for location. For each event the mean azimuth per array is calculated. This mean azimuth is adjusted such that it points to the estimated source location. This adjustment is shown in the histogram. Only the adjustment for events detected by three arrays are shown.

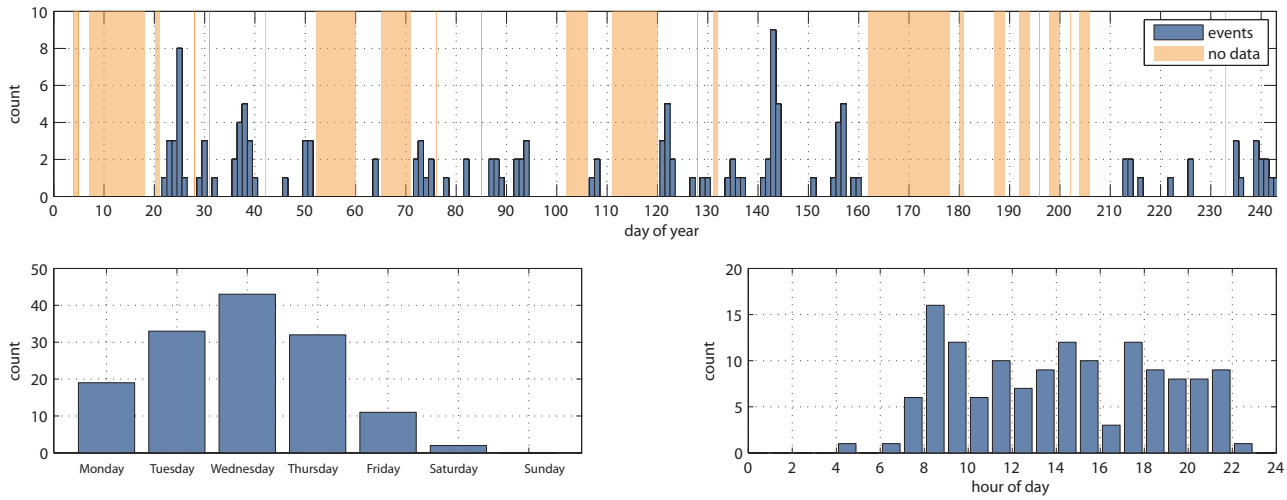


Figure 4.7: Distribution over time and date of events detected by all three arrays. Top: distribution of events over the year 2007. Days where no data was available for one of the three arrays are shown in orange. Lower left: distribution of events over the week. Lower right: distribution of events over the day.

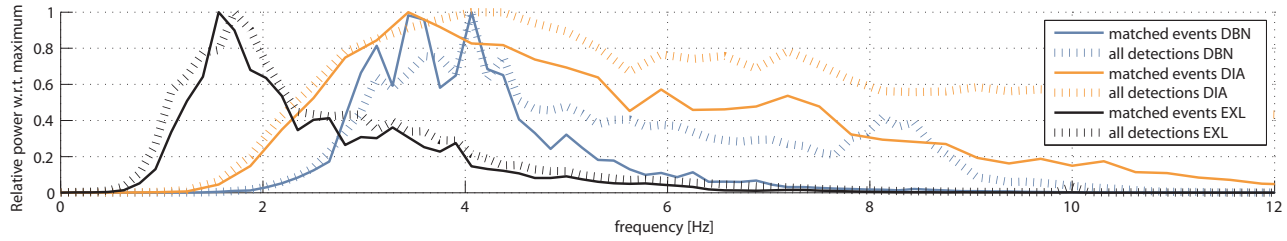


Figure 4.8: Average frequency spectra for all events detected by three arrays are shown as continuous lines. The average frequency spectra of all detections at the arrays are shown as dashed lines. The latter show more energy over higher frequencies for DBN and DIA. For EXL the two spectra are very similar. The signals of the individual array receivers were filtered on frequency before processing, Section 3.4.2. The shown frequency spectra are calculated from the best beam signal which is formed by the average signal over all receivers, corrected for the time difference between the receivers. For DBN and EXL the best beam contains the same frequencies as the individual receiver signals (1 to 8.8Hz and 2.8 to 10Hz respectively). For DIA the best beam develops a higher frequency content than what is allowed by the filtered recordings of the individual receivers (1 to 10Hz). This high frequency content is most probably caused by interference between lower frequency recordings of the 16 receivers of this array. For the events detected by three arrays the frequency spectrum of DIA is much closer to the input frequency band.

4.2.4 Detections per event

Only about three percent of all detections propagate into the 141 events that are detected by all three arrays (Table 4.2). Although this number is almost tripled for the unreliable two array detections this still shows that the vast majority of the detections cannot be related to events which are detected at multiple arrays. It is therefore expected that most detections are caused by local events. If one looks at the number of detections in the histogram of Figure 4.9 it becomes clear that slightly less than half of all events contain multiple detections per event. This shows the importance of grouping multiple detections into events.

array	DBN	DIA	EXL	total
number of detections	8112	8606	9798	26516
events detected by 2 arrays	508	469	547	762
detections propagated into 2 array events	656 (8.1%)	822 (9.6%)	757 (7.7%)	2235 (8.4%)
events detected by 3 arrays	141	141	141	141
detections propagated into 3 array events	209 (2.6%)	343 (4.0%)	228 (2.3%)	780 (2.9%)

Table 4.2: The number of events counted for DBN, DIA, EXL and the total number of registered events. The number of detected events is roughly the same for the three arrays, which was expected as also the number of detections at the three arrays is comparable. About 3% of the detections generate events dat are detected by all three arrays. One event can be detected multiple times by each array.

4.2.5 Signal duration

By subtracting the time of the first detections in an event from the time of the last detection (both at the same array) the recorded signal duration can be calculated. An overview of the signal duration of all identified events is shown as a histogram in Figure 4.10 and on a map in Figure 4.11. For a source at a distance of 350 km (halfway de Bilt - Edinburgh) generating a signal at the maximum expected celerity (360 m/s, Table 3.10) and the minimum expected celerity (220 m/s, Table 3.9) a signal duration of 10 minutes is obtained. This is the extreme case as for most events no detections are expected for both the maximum and minimum celerity and most events are located within this 350 km distance. Therefore it is expected that for the events that are recorded for over 10 minutes the multiple detections are not (only) caused by different atmospheric returns but because the source emitted multiple signals over time. Sources emitting multiple signals over time can be found to be continuous sources, such as gas flares or for the duration of supersonic flight of a jet fighter. The multiple signals can also be caused by multiple events which occur close in time and location, such as the detonation of multiple explosives or multiple jet fighters joining in a practice scenario. The locations of sources with a median duration of over 10 minutes are spread both over land and sea. Near (53^oN,10^oE) four events are located very close together of which three have a signal duration of over 10 minutes.

A remark should be made on the influence of the processing procedure on the results. The Fisher detector which is used to flag detections is fed with time blocks of 256 seconds for DBN and EXL and 64 seconds for DIA. These time blocks are subdivided into bins of 6.4 seconds. For each bin the Fisher ratio is determined but only the bin with maximum Fisher ratio is saved as a detection. A single detection therefore has a length of 6.4 seconds, but no secondary detections are possible within the time block of either 256 or 64 seconds. This influences the determination of the exact signal duration, but is of less importance if one looks at the signal duration with a resolution higher than 256 seconds, as used for Figures 4.10 and 4.11. The shorter time blocks used for DIA explains the larger number of events that was shown for DIA in Figure 4.9.

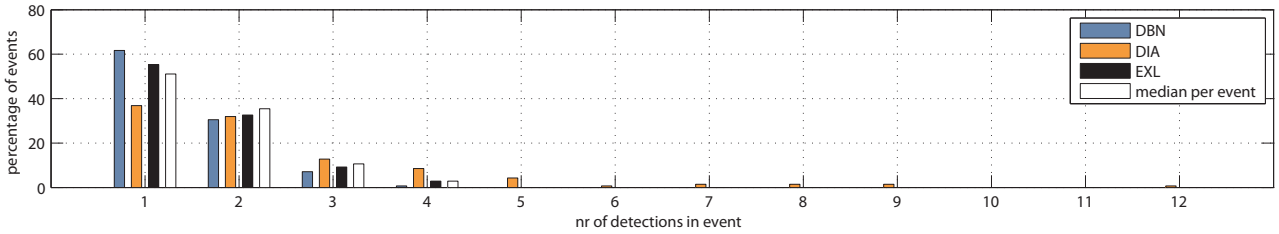


Figure 4.9: Histogram of the number of detections that are grouped into a single event. The histogram shows the number of detections at each array and the median of these three numbers for each event. Most events consist of only one detection per array. DIA shows more detections for each event.

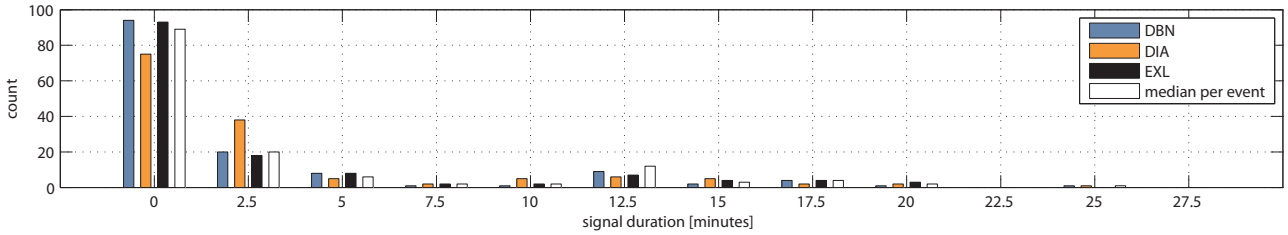


Figure 4.10: Histogram of the recorded signal duration at the individual arrays. Short signal detections of only one detection (i.e. 6.4 seconds) make up the majority of the detected events. For DBN this were 87 detections or 62% of the events, for DIA 52 (37%) and for EXL 78 which are 55% of the events. The median of the signal durations per event follows the signal duration of the three individual arrays. This indicates that the events of long duration were registered for this timespan at more than one arrays, and it is therefore more probable that the source indeed emitted energy throughout this timespan.

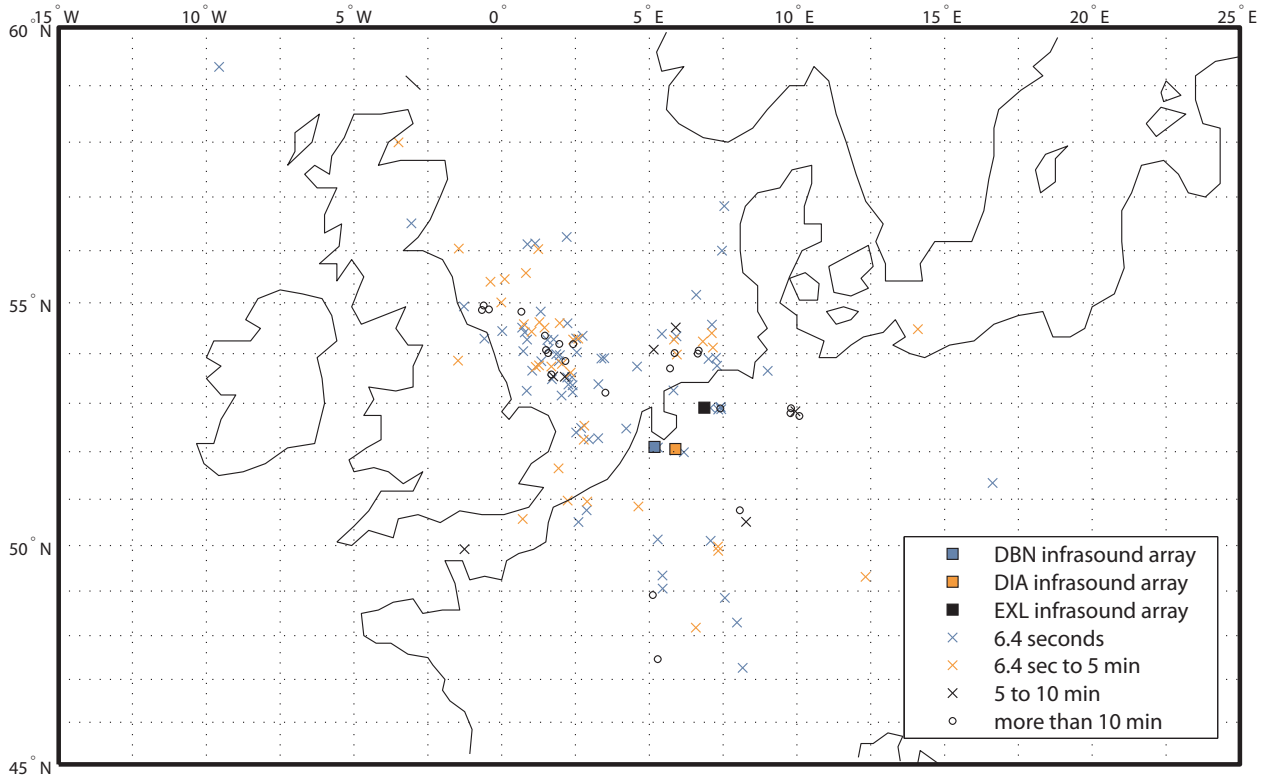


Figure 4.11: Event locations found by three arrays colour coded to the median signal duration of the event. The signal duration can differ at each array, for this map it is chosen to display the median of these three durations. The minimal signal duration that can be distinguished is 6.4 seconds.

4.2.6 Celerity

Because a single event can consist of multiple detections, care should be taken on which receive times are used for celerity estimation, see also Section 3.4.6. An impulsive source can result in multiple returns, refracted from different atmospheric layers, each with a different celerity. Three difficulties in estimating the celerity of multiple returns are identified. First, not all of these returns have to be detected at all arrays. Second, the source does not have to be impulsive, it can generate multiple signals over a longer timespan. Third, as discussed in the previous section on signal duration, only one return within each block of 256 seconds (DBN and EXL) or 64 seconds (DIA) is saved. Thus from the returns that follow each other up within this time span only the strongest return (i.e. with the highest Fisher ratio) is registered.

The celerity of the results will be estimated in twofold. First, the time range from the first to the last detection at all arrays is used to calculate a minimum and maximum celerity. This gives a straightforward result on which one can distinguish correct from false event locations. This information was used to filter results on celerity at the end of processing, Section 3.4.6 and to distinguish between multiple associated events as discussed in the second paragraph of this chapter. The celerity regions are shown in Figure 4.12. In 51% of the events consisting of multiple detections the celerity range is larger than 1000 m/s and in 46% of the events the maximum celerity becomes infinity. Therefore this method is of no use if one would like to use the celerity to investigate which atmospheric returns are detected.

To investigate the possibility of categorizing detections into different atmospheric returns, a second approach is used. For this all individual detections between a single event are combined, and the resulting celerities are calculated. All combinations within the expected celerity regime (220 - 360 m/s, Table 3.9) are plotted in Figure 4.13. Error bars are added for the precision on timing and position. The distribution of these errors are given in Figure 4.14. The error from location is defined as the maximum error which occurs within the 95% error ellipse as shown in Figure 4.5. The error from timing is caused by the length of the timewindow of the Fisher detector which is 6.4 seconds. The arrival time is therefore assumed to have an error with a maximum of 6.4 seconds. As can be seen on the right hand side of this graph, the different celerity regimes as introduced by Revelle & Whitaker (1996) are partly overlapping and relatively small in comparison with the expected errors. This makes it very difficult to make a reliable categorization of the detection pairs to these categories. To make this approach more feasible only three categories of detections are made. For the first category the regime of thermospheric returns is taken from 220 m/s (Revelle & Whitaker, 1996) to the lower boundary of conventional stratopause ducted returns at 250 m/s (Evers & Haak, 2007). The label for thermospheric returns is used for this category: I_t. The second category is formed by conventional stratopause ducted returns and runs from 250 m/s until 300 m/s. The label for stratopause ducted returns is used: I_s. From 300 m/s multiple possible returns overlap each other in celerity. Therefore a mixed category is created running from 300 m/s to 360 m/s. This category contains the regions of tropopause ducted returns, a small part of the stratopause ducted returns as given by Revelle & Whitaker (1996) and the stratopause ducted forerunners as observed by Evers & Haak (2007). This category contains all fast returns and is labelled as I_f. Figure 4.13 shows a continuous distribution of the celerities of the three regions and the error bars often span more than one category. Having this uncertainty in mind, the three categories are used to make only an initial investigation on which returns are expect to be present in the results. For 35% of the events combinations could be made which indicate thermospheric returns, for 58% there were combinations which indicate stratospheric returns and for 82% of the events fast returns were indicated. The events and the indicated categories are plotted on a map in Figure 4.15. In this map thermospheric and stratospheric returns are identified for locations which are very close to the arrays. This is not expected, as these type of returns only occur after a certain distance, over which the atmosphere was able to reflect the signal back to the earth's surface. This shadow zone is usually considered to be around 200 km (Evers, 2008). Both this discrepancy and the large errors that were observed show that the investigation of celerities from only detection times and location estimates is not advisable. To make a reliable investigation on the different returns in an event one should compare the detections against synthetic results obtained by ray tracing.

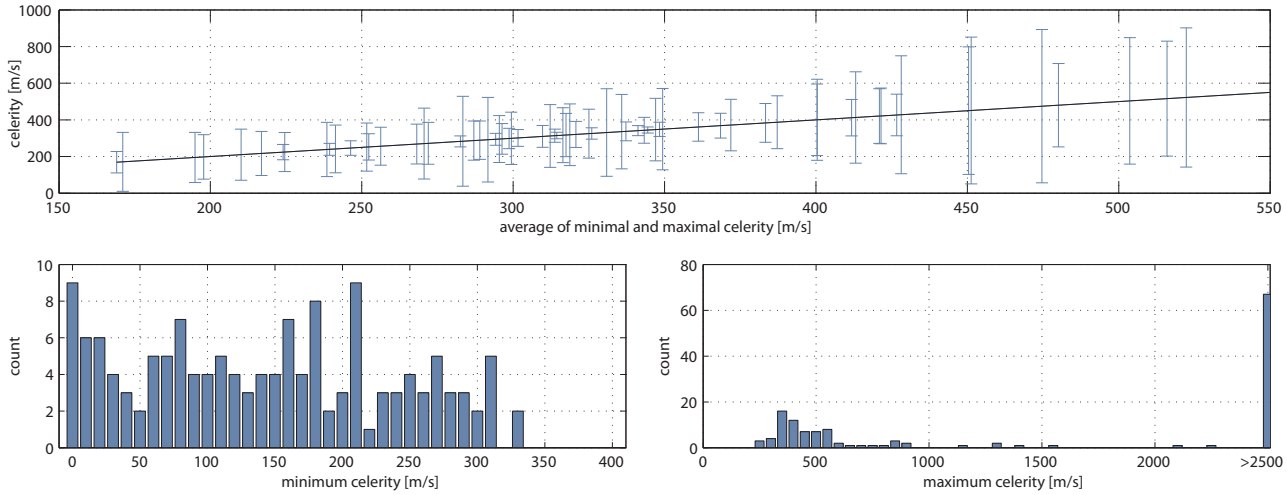


Figure 4.12: Top figure: events with a maximum and minimum celerity within the range of 0 to 1000 m/s. Lower left figure: Histogram of the minimal celerity per event. Lower right figure: Histogram of the maximal celerity per event. Celerities are calculated using the minimum and maximum time difference between arrays, generated by comparing the time range from the first to the last detection within an event between all arrays. This is divided by the difference in source distance between the arrays to come to a minimal and maximal celerity. Only 49% of the total events is represented in the top figure. Why the rest of the events are outside of the display range becomes clear from the histogram of maximum celerity. The histogram shows that 52% of the maximal celerities is larger than 1000 m/s and 47% is larger than 2500 m/s. For 46% of the detections an infinite maximal velocity is indicated. An infinite celerity is generated when a detection at an array further away from the source is made at the same time, or before a detection is made at an array closer to the source. If only negative celerities are generated the event is removed by the filter on celerity (Section 3.4.6). When the time range generates both positive and negative celerities, the maximum possible celerity for this time range becomes infinite.

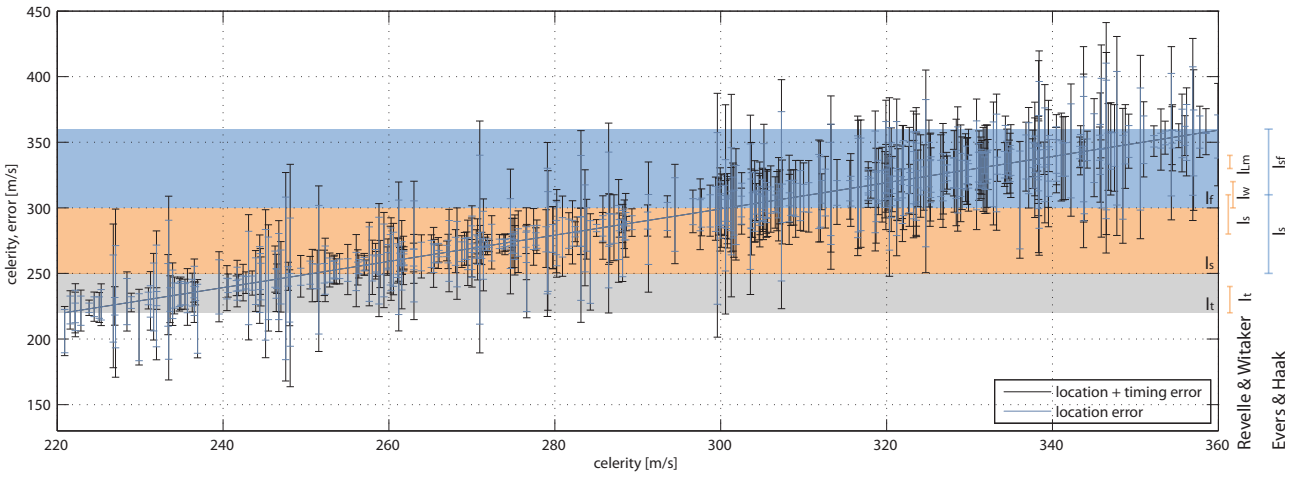


Figure 4.13: Celerity estimates on combinations of detections between two of the three arrays. Celerities are obtained by combining all possible pairs of individual detections between two stations within a single event. Only combinations with celerities within the expected celerity regime (220 - 360 m/s) are shown. For displaying purposes the most unreliable celerities with a total error larger than 100 m/s are removed. This is 11.3% of the total amount of pairs within the expected celerity regime. Error bars show the errors on celerity caused by an error in location (blue) and added to this the error in timing (black). The distribution of these errors is shown in Figure 4.14, of which the caption describes the calculation of the errors. The black, orange and blue bands show the regions for three categories of returns which correspond to thermospheric, stratospheric and a mix of fast returns. These categories are a simplification from the categories used by Revelle & Whitaker (1996) and Evers & Haak (2007), shown on the right hand side. The color bands correspond to the markers on the map of Figure 4.15.

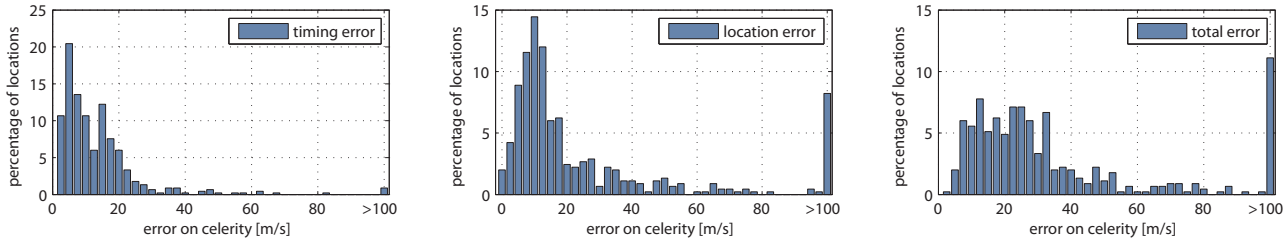


Figure 4.14: Histograms on timing and location errors on the celerity estimates. The error from location is the maximum error which occurs within the 95% error ellipse as shown in Figure 4.5. The error from timing is caused by the length of the timewindow of the Fisher detector which is 6.4 seconds. Therefore, the arrival time is assumed to have a maximal error of 6.4 seconds. The average error in celerity caused by the timing inaccuracy is 14 m/s with a 95% error 36 m/s. The average celerity error caused by the location inaccuracy is 34 m/s with a 95% error 169 m/s.

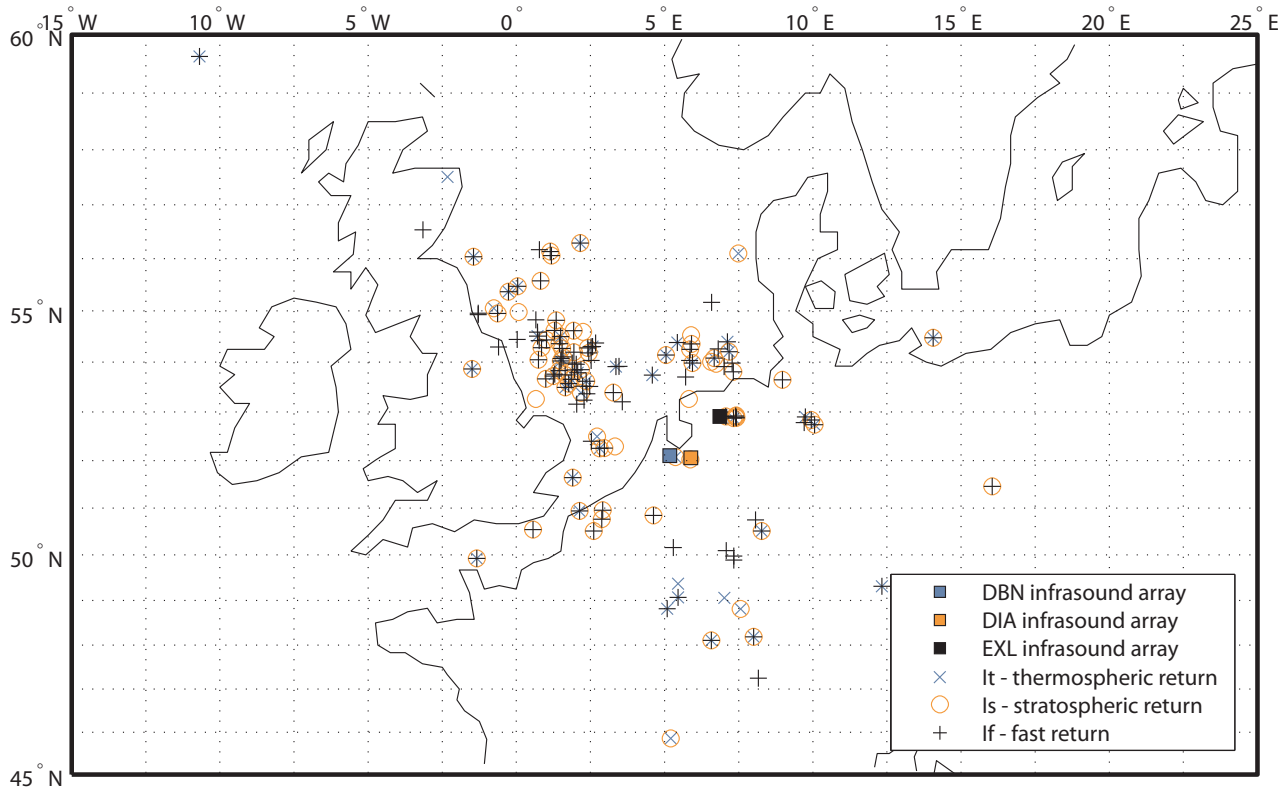


Figure 4.15: Event locations found by three arrays colour coded to the estimate of the type of atmospheric return. Detection combinations with a celerity of 220 - 250 m/s are flagged as thermospheric returns, 250 - 300 m/s as stratospheric returns and 300 - 360 m/s as fast returns. Fast returns can consist of Lamb-surface waves, tropopause ducted, stratopause ducted and stratopause ducted forerunners. For displaying purposes the most unreliable celerities with a total error larger than 100 m/s are removed. This is 11.3% of the celerity estimates within the expected celerity regime. The continuous distribution over the three categories and the size of the errors, Figure 4.13, indicate that this map should only be used as a first indication on which event generates which kind of returns. To make a reliable investigation on the different returns in an event it is advised to compare the detections against synthetic results obtained by ray tracing.

4.3 Origin of sources

In this section an explanation is sought for the estimated source locations. In Section 4.2 de properties of the located events showed two important clues for what sources to look for. Firstly, it was shown that 99% of the events were detected during working days and 98% of the events were detected between 7 am and 10 pm. This strongly suggests that most events are related to human activity. Secondly, 16% of the events had a signal duration longer than 10 minutes. This duration cannot be explained by multiple atmospheric returns and therefore had to emit multiple signals over time. Sources emitting multiple signals over time can be found to be continuous sources, such as gas flares or for the duration of supersonic flight of a jet fighter. The multiple signals can also be caused by multiple events which occur close in time and location, such as the detonation of multiple explosives or multiple jet fighters joining in a practice scenario. Events with long signal duration are found both over land and over sea. For sources that are found over land one can use aerial imagery to find potential sources. For the sources over sea this is not available. Both regions are therefore discussed separatelty.

4.3.1 Sources over land

To investigating the estimated infrasound locations on aerial imagery a web-based viewer was developed in cooperation with ESRI (2009). A significant number of locations was found to be located close to military bases or practice grounds. This was not assumed beforehand, but the infrasound event locations quickly turned into a map locating camouflaged bunkers, practice grounds and military airstrips. An overview of the identified locations is given in Table 4.3 and Figure 4.16.

4.3.2 Sources over sea

Infrasound sources over sea can be expected from military jet practice (sonic booms), the detonation of explosives (e.g. clearing of naval mines) and gas flaring at oil and gas production facilities. Records of military practice or controlled detonations were not available for this research. As most locations are found over the western part of the North Sea information of the United Kingdom's air force, the Royal Air Force (RAF), should be most valuable to investigate the influence of sonic booms. It is likely that this information can be obtained through the freedom of information law, however it was assessed that the costs for providing data on all supersonic flights over the North Sea in 2007 would exceed the appropriate limit set for this law (Royal Air Force, 2009).

NGDC presented estimates of gas flaring using data collected by the U.S. Air Force Defense Meteorological Satellite Program (DMSP) Operational Linescan System (OLS) (Elvidge *et al.*, 2009). The OLS collects single band low light images of visible and near infrared during the night. Gas flares were identified as circles of light, bright in the center and dim at the outer edges. As these features are much larger than the flares, it was expected that the OLS detects the lit up sky surrounding active flares. NGDC identified gas flaring features in a time series of annual night time light composites extending from 1995 through 2006. The identified gas flares are compared to the estimated source locations in Figure 4.17. Most gas flares are identified in the centre of the North Sea where no infrasound events are found. Most events are located on the western part of the North Sea where only a few gas flaring locations were identified. If these events were caused by gas flaring one would also expect to find events in the centre part of the North Sea. The majority of the located infrasound events is therefore expected to be related to other sources than gas flaring.

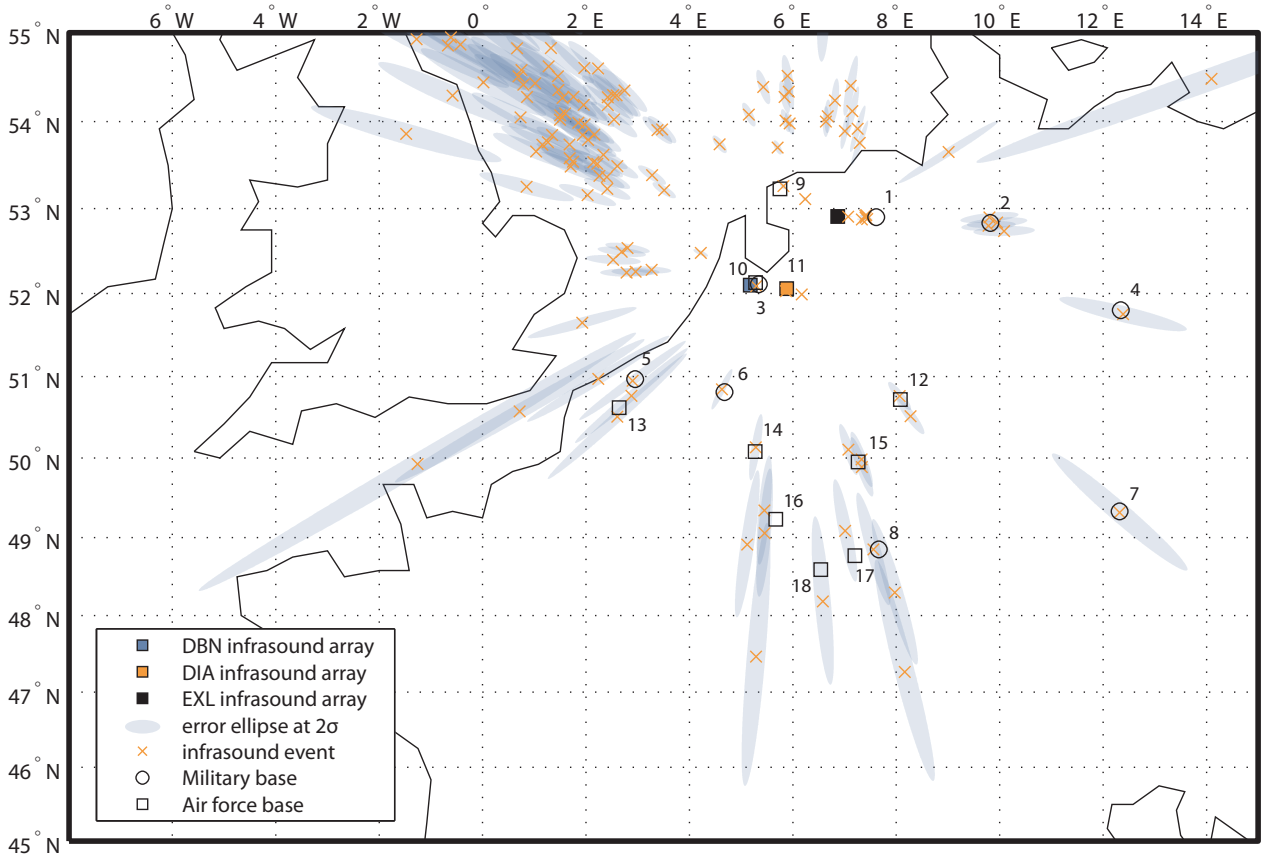


Figure 4.16: Map of military presence found close to infrasound events in aerial imagery. This is not a complete overview of all military or airforce bases over western Europe. As military bases and airfields are spread all over western Europe one should be careful in stating correlation for each single event. However, for the majority of infrasound events over land, military presence could be found at very short distance. Therefore it is expected that military activity and the found infrasound events are correlated.

type	number	lat	lon	description
military	1	52.90	7.61	Lorup (ammunition depot)
	2	52.82	9.81	Bergen-Hohne (practice ground)
	3	52.11	5.34	Leusderheide (practice ground)
	4	51.80	12.34	Oranienbaum (practice ground)
	5	50.97	2.95	Bos van Houthulst (demining base)
	6	50.81	4.68	Kwartier Meerdaalbos
	7	49.33	12.32	Pissau/Neunburg (shooting range)
	8	48.85	7.66	camouflaged barracks (unknown)
airbase	9	53.23	5.75	Leeuwarden
	10	52.13	5.28	Soesterberg
	11	52.06	5.87	Deelen
	12	50.72	8.08	Siegerland
	13	50.62	2.64	de Merville - Calonne
	14	50.08	5.27	St. Hubert
	15	49.95	7.26	Frankfurt Hahn
	16	49.23	5.67	Étain - Rouvres
	17	48.77	7.20	Phalsbourg - Bourscheid
	18	48.59	6.54	Luneville - Chenevières

Table 4.3: Description of military presence found close to infrasound events in aerial imagery. The numbers correspond to the numbers in the map of Figure 4.16.

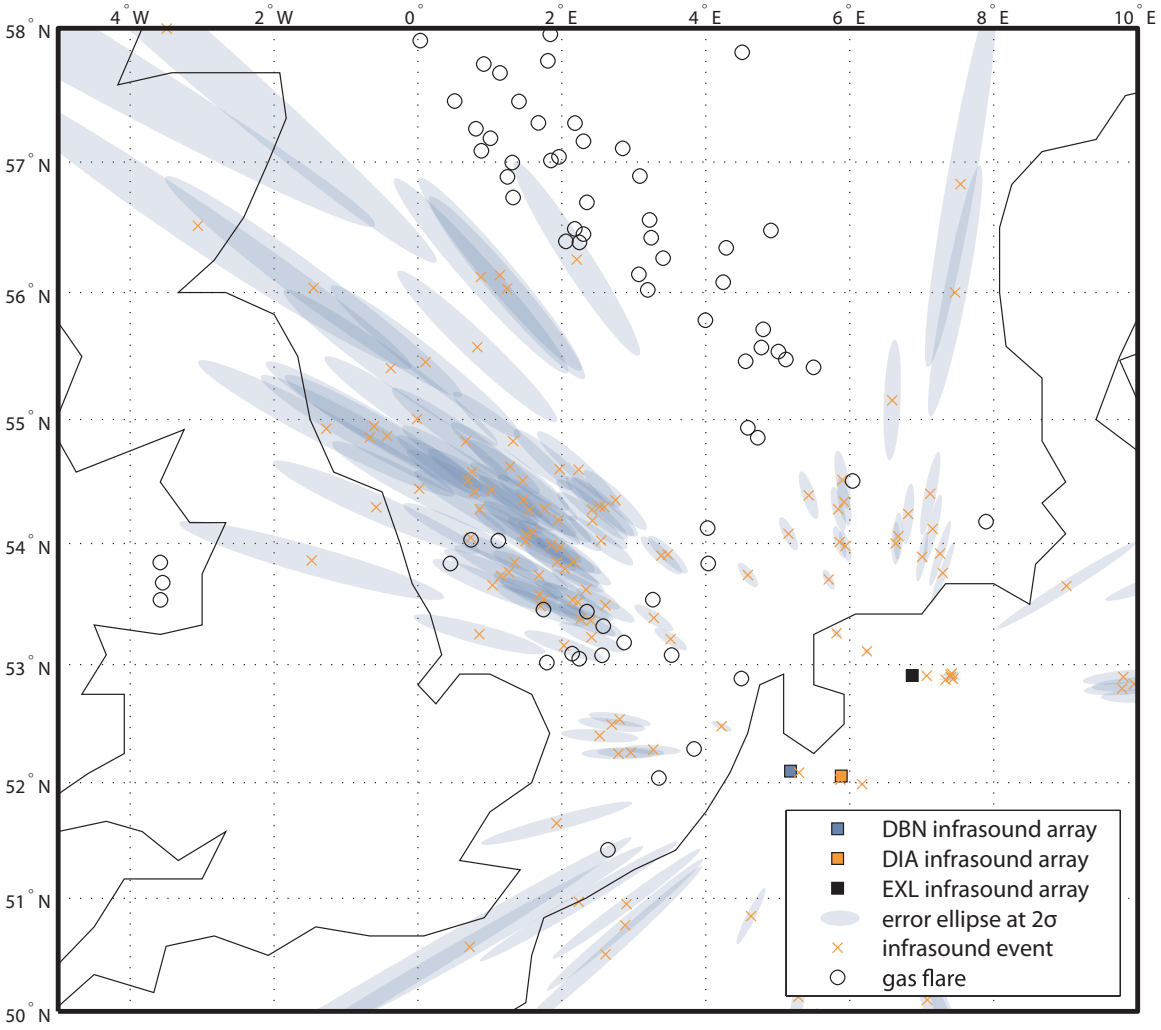


Figure 4.17: Locations of infrasound sources compared with gas flare locations presented by NGDC (Elvidge *et al.*, 2009). Most infrasound locations are estimated in different regions of the North Sea than gas flare locations.

Chapter 5

Results six infrasound arrays

Data of three more infrasound arrays located in Germany and France was made available as part of the NERIES program. The location and layout of the three added arrays can be seen in Figure 5.1 and are labelled FLERS, IGADE and I26DE. The detection lists for these arrays were already filtered on frequency and apparent velocity. Extra filtering is applied on consistency for all six arrays and on back azimuth for the detections of IGADE. Except for the individual array processing steps the processing steps are equal to what was used for the KNMI infrasound arrays. Using these six arrays 11 events where identified which were detected by all 6 arrays and 53 which were detected by 5 out of 6 arrays.

5.1 Detections per array

The processing procedure on the three added arrays is different from the approach presented for KNMI arrays. Therefore a short summary of the processing procedure and a investigation of the detection characteristics are given in this section. Based on the detection characteristics an extra filter on back azimuth is applied for the IGADE infrasound array near Bremen, Germany.

5.1.1 The PMCC correlator

In stead of the Fisher detector the procedure used to process the recordings is based on the the Progressive Multichannel Correlation method (PMCC) (Cansi, 1995). PMCC is based on a search process in which one starts with a small subset of array sensors. The recordings of more sensors are added to this set but only if the consistency between the recordings increases by doing so. An in-depth discussion of the differences between PMCC and the Fisher detector is outside the scope of this research, the interested reader could look into the comparison of the two methods by Caljé (2005) for more detail. To give the reader some insight in the differences one can expect the procedure to select signal from noise will be explained.

In PMCC, a measure of consistency is used to distinguish signal from noise, analogue to the Fisher ratio for the KNMI arrays. Because of the nature of the PMCC search process the consistency of the detection is not

infrasound array	latitude	longitude	elements	aperture
DBN	52.10^{o}N	5.18^{o}E	6	80m
DIA	52.06^{o}N	5.88^{o}E	16	1500m
EXL	52.91^{o}N	6.87^{o}E	6	250m
FLERS	48.76^{o}N	0.48^{o}W	4	2700m
IGADE	53.26^{o}N	8.69^{o}E	4	800m
I26DE	48.85^{o}N	13.71^{o}E	5	2600m

Table 5.1: Location, elements and aperture of DBN, DIA, EXL, FLERS, IGADE, I26DE

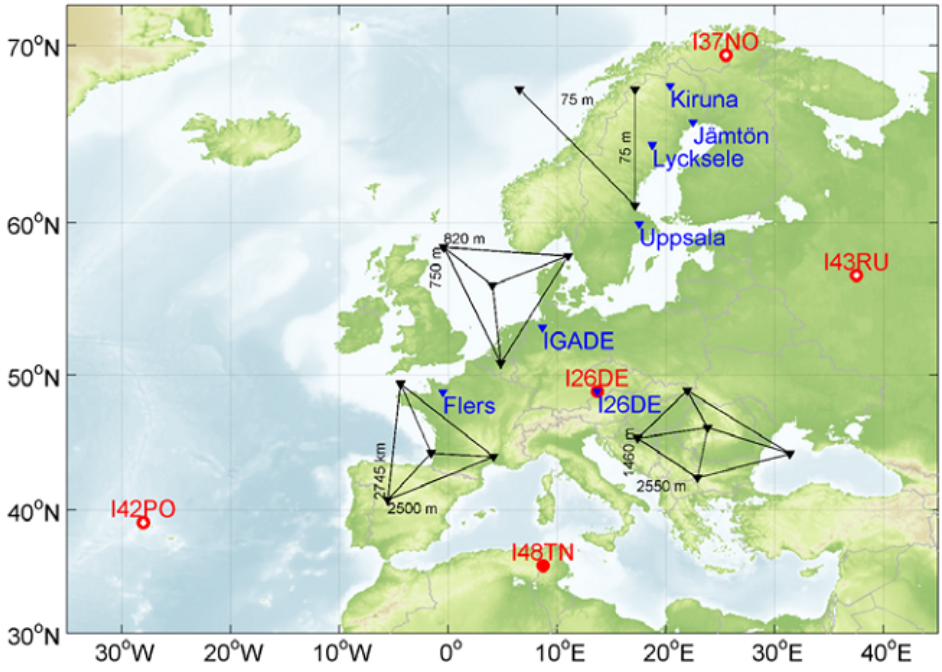


Figure 5.1: Layout and location of FLERS, IGADE and I26DE (Le Pichon *et al.*, 2008).

a function of the signal to noise ratio, as is the case for the Fisher ratio. Consistency is defined as the mean quadratic residual of the closure relations (Cansi & Klinger, 1997), using the notation of Caljé (2005):

$$C_n = \sqrt{\frac{1}{A}\sum_{l=1}^{A} r_{ijk_l}^2} \tag{5.1}$$

$$A = \frac{n(n-1)(n-2)}{6} \tag{5.2}$$

$$r_{ijk} = \Delta t_{ij} + \Delta t_{jk} + \Delta t_{ki} \tag{5.3}$$

Where C_n is the consistency. A gives the number of possible independent triangles from a sub array with n elements. For each possible triangle of sensor combinations, the closure relation r_{ijk} is calculated. The closure relation gives the closing error between the sensor recordings. The time differences Δt between these recordings form a closed loop over the triangle of sensors. If no noise is present the sum of time delays becomes zero, resulting in infinite consistency. A larger closing error gives a smaller consistency. The time differences are obtained by the correlation function applied on the two sensor recordings.

Consistency is calculated for signal subsets bounded in frequency and time. If the consistency is larger than a certain value, in this case a consistency of 6, the subset is saved. For each subset the combination of the array sensor locations and the timing differences give an apparent velocity and back azimuth, similar to the method used for the KNMI arrays, Section 3.1.1. The subsets are grouped into detections by means of a nearest-neighbour search in the time / frequency / azimuth / velocity domain. A weighted Euclidian distance is used to connect close-enough points to form a single detection, see Figure 5.2.

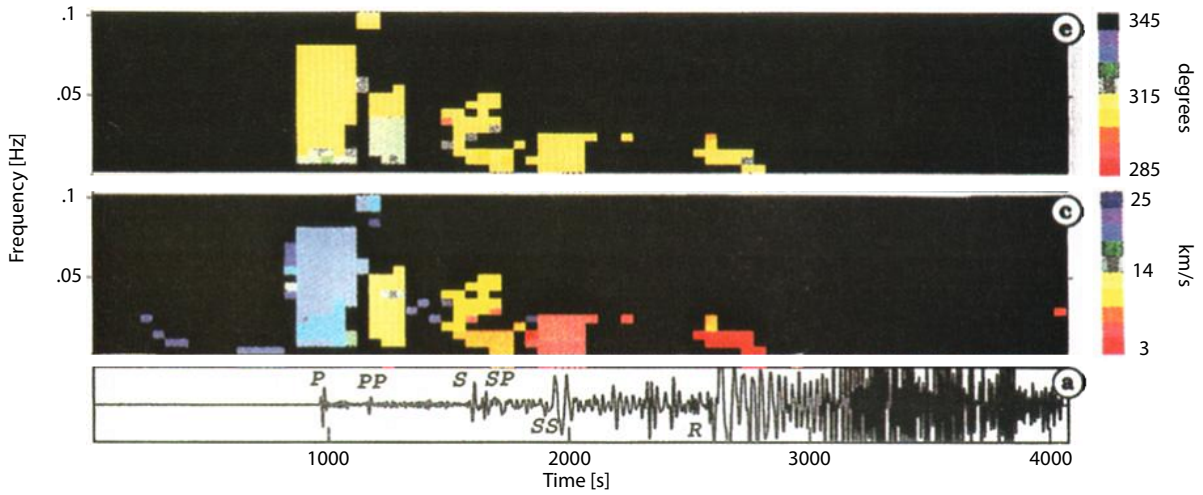


Figure 5.2: Example of a detection result from the PMCC correlator taken from Cansi (1995). The results for a teleseismic event recorded on the LDG long-period network are shown. The lower plot (*a*) gives the waveform of the signal on which multiple phases are labelled. The plots above show the apparent velocity (*c*) and the azimuth (*e*) as a function of time and frequency. The phases P, PP, S/SP, SS and R are received with equal back azimuth and different velocity, i.e. incident angle. The amount of equal coloured pixels is a result of the duration and frequency bandwidth of each phase. A stronger detection results in more pixels.

5.1.2 Detection characteristics

The detection characteristics will be discussed by means of the total number of detections, the detection length, apparent velocity, frequency, consistency and standard deviation on back azimuth for all detections.

Number of detections

The total number of detections at all six arrays are shown in Table 5.2. The number of detections at IGADE is found to be a factor 4 to 7 higher than at the other arrays. It is expected that these very high number of detections will cause high numbers of false events, created by combining unrelated detections (this will be shown in Section 6.2). After applying filters on consistency and back azimuth the number of detections per array become comparable.

array	number of detections	after filtering
DBN	4695	
DIA	5719	
EXL	6165	
FLERS	8364	3318
IGADE	32914	6663
I26DE	7476	4539

Table 5.2: The number of detections at all six arrays from January 1st 2007 until June 12 2007. During this period detections were available from all six arrays. Without the additional filtering, the number of detections at FLERS, IGADE and I26DE are considerably higher than the number of detections at the three KNMI arrays. IGADE shows a very large number of detections compared to the other arrays, a total number of 32914 over 163 days, or 200 events per day.

Detection length

The detection time length for the FLERS, IGADE and I26DE is shown in Figure 5.3. In the processing procedure for the KNMI arrays a fixed time window length was used over which a single detection could be made. The procedure applied for the PMCC detector adds multiple small correlation windows bounded in frequency and time. This allows long signals to be captured in one detection, up to 600 seconds in the used dataset.

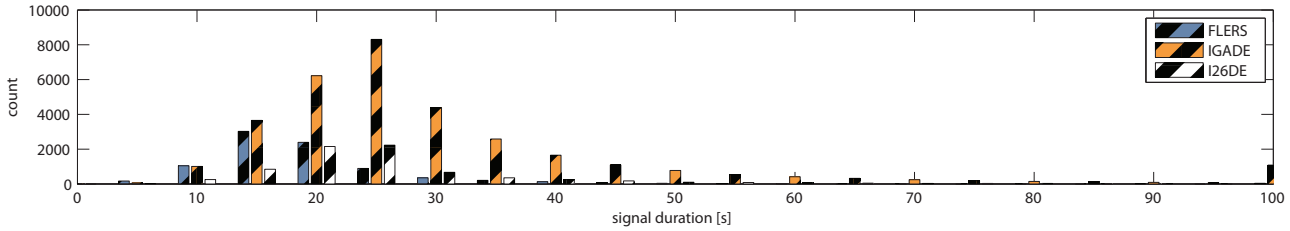


Figure 5.3: Histogram of signal duration of the detections at FLERS, IGADE, I26DE. The signal duration of the individual detections are shown, these individual detections will be grouped into events in further processing.

Apparent velocity, frequency and consistency

The datasets of FLERS, IGADE and I26DE that were made available for this research had been filtered on apparent velocity, frequency and consistency. The boundaries of these filters become clear from the histogram of these parameters, Figure 5.4. An additional filter on the consistency was applied to reduce the amount of events from false combinations of detections and to give all six arrays a comparable number of detections.

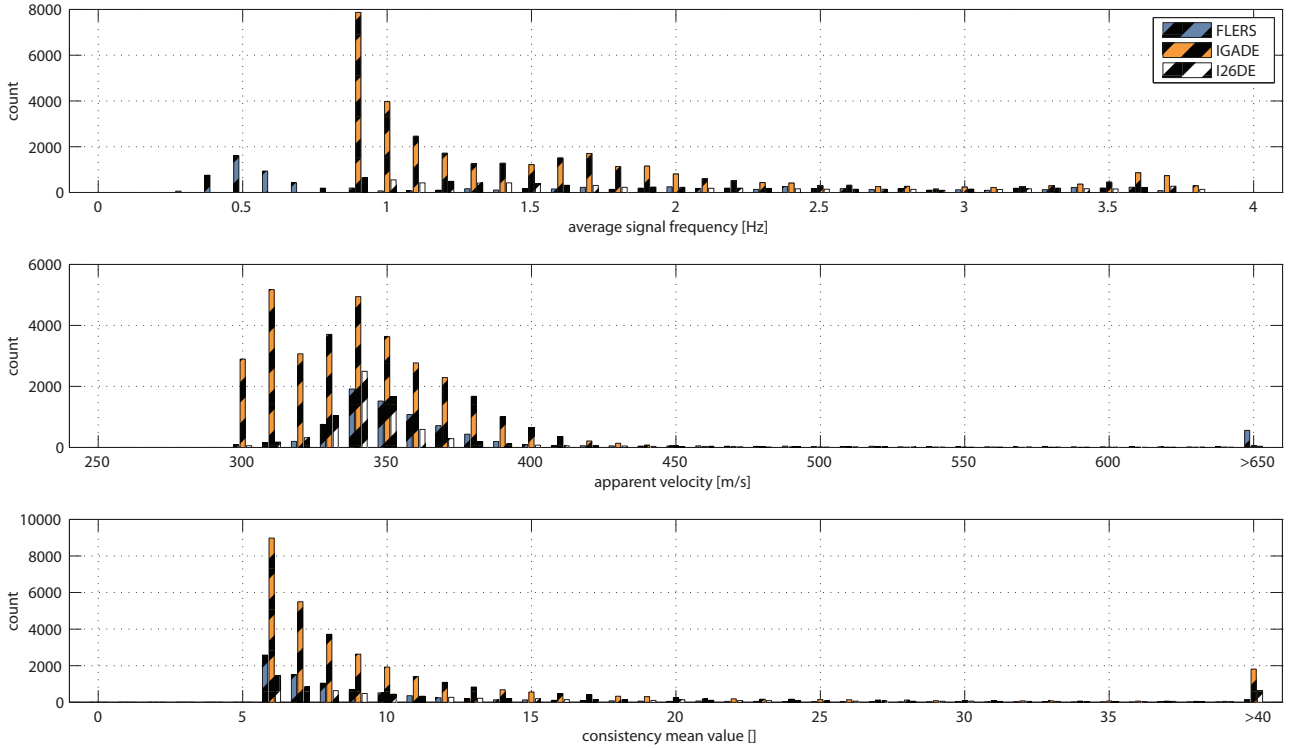


Figure 5.4: From top to bottom, histograms of frequency, apparent velocity and consistency. All histograms are given for FLERS, IGADE and I26DE. The upper plot shows a permitted frequency range of 0.295 Hz - 3.805 Hz for FLERS and 0.865 Hz - 3.835 Hz for IGADE and I26DE. For the apparent velocity it is shown that all detections below 300 m/s were removed. The minimal consistency in the original detection lists is a consistency of 6, this is changed to a minimum of 8 for this research.

Standard deviation on back azimuth

The arrays in Germany and France have larger apertures than the KNMI arrays, see Table 5.1. This would theoretically increase the accuracy of the estimated signal back azimuth. This effect is stronger for signals with longer wavelengths, i.e. lower frequencies. The standard deviation on back azimuth is plotted for all six arrays in Figure 5.5. From the array aperture sizes one would expect FLERS and I26DE to estimate the back azimuth with smallest standard deviations, followed by IGADE and then DBN and EXL. The histogram however shows lower standard deviations for DBN and comparable standard deviations for EXL, FLERS, IGADE and I26DE. The lower standard deviation of DBN in comparison with EXL could be due to DBN's higher frequency detections (Figure 4.8) and lower susceptibility to sidelobes (Figure3.15). For the comparison between DBN and EXL on the one hand and FLERS, IGADE and I26DE on the other there are four arguments in favour of a larger standard deviation at the latter arrays. First, the detections at the KNMI infrasound arrays had frequencies up to 10 Hz which is considerably higher than the upper frequency bound of 3.8 Hz as used for the three other arrays. The back azimuth of these higher frequency signals can be determined with a higher accuracy as the wavelength becomes smaller in comparison with the array aperture. Second, the Fisher detector as used for the KNMI arrays is known to solve the back azimuth with a lower standard deviation than the PMCC detector, especially for frequencies above 1.5 Hz. Caljé (2005) showed standard deviations for detections with PMCC at a factor 2-10 higher than detections made by Fisher on the same dataset. Third, the PMCC correlator allows a longer signal duration for each detection than the time windows of 6.4 and 3.2 seconds as used for the Fisher detector. This could allow multiple phases which arrive close in time but with slightly different back azimuths to be recorded as one detection. Fourth, the three added arrays have a lower number of elements, which increases the influence of noise. It should be noted that the method to obtain the standard deviation for FLERS, IGADE and I26DE is different from the method used for DBN and EXL.

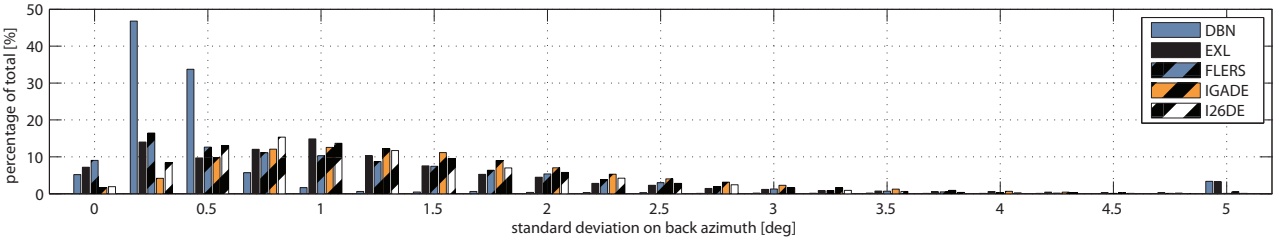


Figure 5.5: Standard deviation on back azimuth for DBN, EXL, FLERS, IGADE and I26DE. DIA is not shown as the estimation of the standard deviation for this array was found to be unreliable, see Figure 4.3. The detections at DBN show the smallest standard deviation.

5.1.3 Additional filter on consistency and back azimuth

To reduce the amount of low energy local detections the minimum required consistency is risen to 8. This reduces the amount of false combinations and allows us to focus on the more powerful detections. As with the KNMI infrasound arrays, a check is made on the distribution of the back azimuth of all detections to check for strong local sources. Back azimuth histograms are shown in Figure 5.6, 5.7 and 5.8. IGADE shows a very large number of detections in South and South-West direction. After a first processing run including these detections a large number of events were located in these directions, see Figure 5.9. As the locations form a line in stead of a point or region it is expected that the events are a result of false correlation of detections between the arrays. This unwanted effect is most probably caused by a local source, such as the local sources that were found for the de Bilt infrasound array, Section 3.4.5. As with DBN the most effective method to remove these local sources is by removing all detections over these back azimuths. The selected azimuths to remove are shown in Table 5.3. For FLERS and I26DE no back azimuth filter is applied.

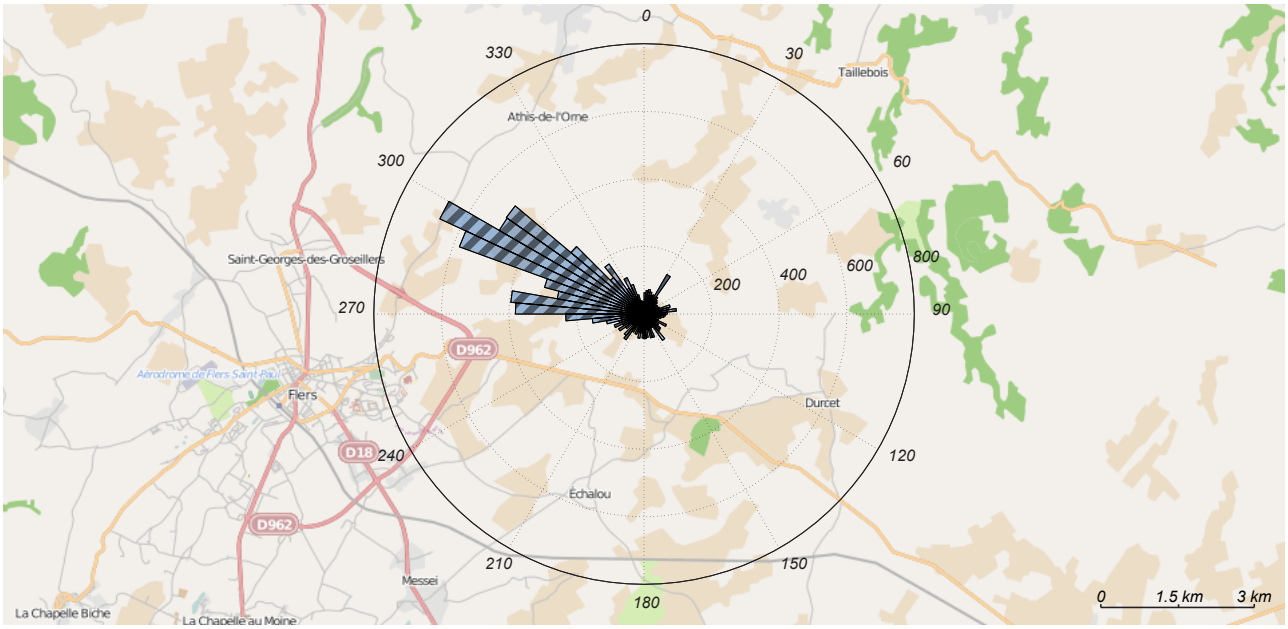


Figure 5.6: Polar histogram of the back azimuth of detections at FLERS. The majority of the events is detected in the North-West direction, towards the Atlantic Ocean.

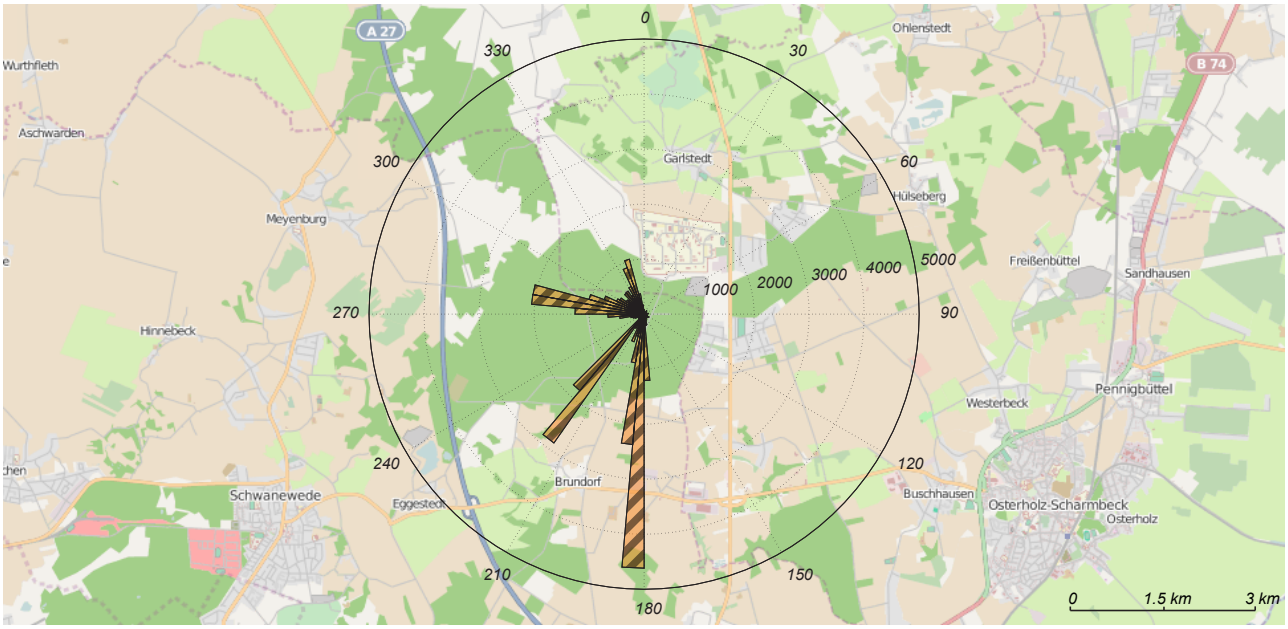


Figure 5.7: Polar histogram of the back azimuth of detections at IGADE. Almost no detections are made in eastern direction. The directions in North and Western direction can be explained by sources over the North Sea, 80 km in this direction. The very large number of detections in South and South-West directions are not yet explained but are expected to be related to local sources.

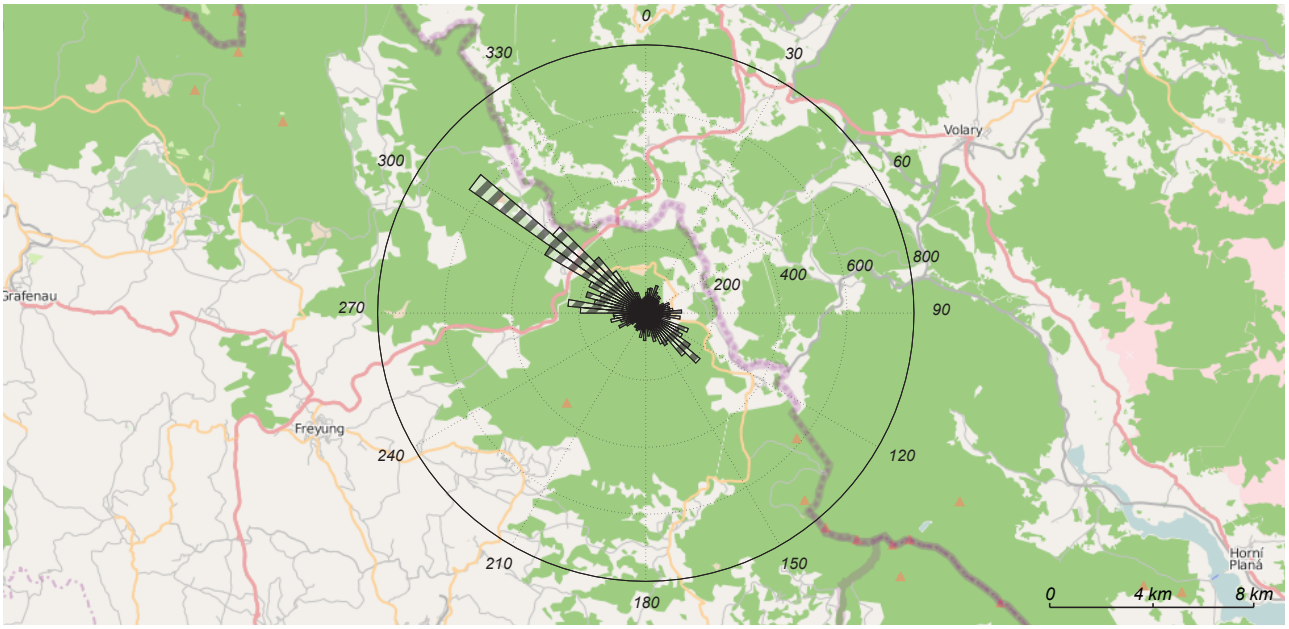


Figure 5.8: Polar histogram of the back azimuth of detections at I26DE. The majority of the detections is made in the North West direction, which is towards central Germany or further, the North Sea.

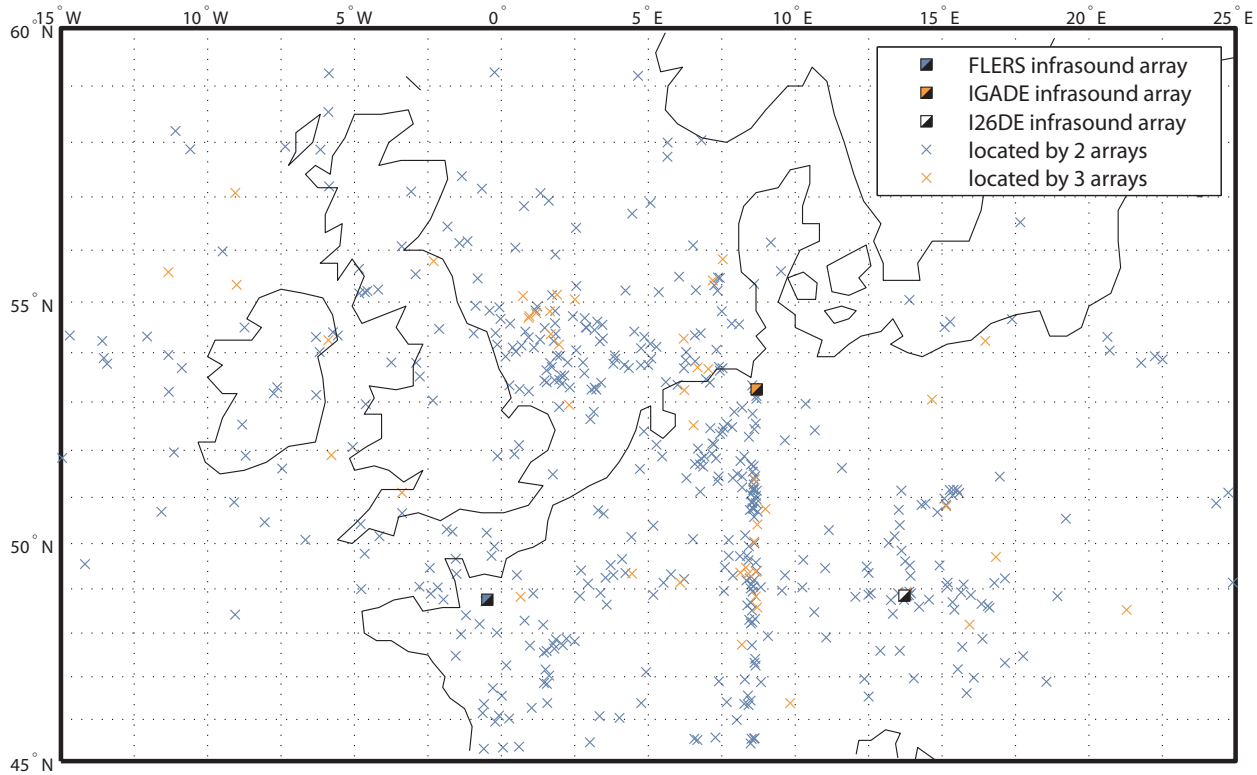


Figure 5.9: Source locations as found when combining FLERS, IGADE and I26DE without applying an azimuth filter on IGADE. From IGADE two distinct lines of detections are visible in the South and South-West direction. These are considered to be artefacts caused by false correlation of detections of a local source close to IGADE with detections at the other arrays.

removed azimuth	number of detections	percentage of detections
175 - 206 deg	10085	30.6 %
212 - 226 deg	4930	15.0 %

Table 5.3: Azimuths over which detections are removed for IGADE. Two directions are removed in South (175 - 206 deg) and South-West (212 - 226 deg) direction. This filter removes 46 % of the total number of detections. The number of detections is taken before applying the additional filter on consistency.

5.2 Source location

The results in this chapter are obtained similar to the source locations as found by three infrasound arrays and multiple detections per array in Chapter 4. The event locations are estimated in latitude and longitude on a spherical earth. The estimate is made using a least square iteration procedure (Section 3.3) which uses the average back azimuth, of all detections within the event, at each array, as input. By inserting the variances on this average back azimuth as weights into the least square iteration a best unbiased estimate of the event location is obtained. Similar to the procedure followed in Chapter 4, the difference between the back azimuths of the detections and the average back azimuth of the event, is used to estimate the variance (σ^2) on back azimuth. This variance contains both the precision at which the back azimuth is determined, and the difference between the back azimuths caused by the atmosphere. A histogram of the difference between the back azimuths of the detections and the average back azimuth of the event is shown for each array in Figure 5.10. From these distributions the standard deviation (σ) and 95% error on back azimuth is estimated and shown in Table 5.4. The standard deviation for DBN, DIA and EXL is increased by a factor 1.2 to 1.7 compared to the values found in Section 4.1. This could be caused by the distance from the arrays to the detected infrasound sources. As the six arrays are spread out over Europe it is expected that a larger part of the detected sources is located over this larger area and less sources that are close to the Netherlands are detected. The sources close to the Netherlands could remain undetected by the arrays in France and Germany. Therefore, the average travel path between the source and the array will be longer. A longer travel path can cause larger azimuthal deviation due to wind. The distribution of the location of events over Europe as a consequence of the used array becomes clear if one maps only the events which are detected by this particular array. These maps are shown for all six arrays in Figure 5.12. These maps clearly show that FLERS and I26DE detect less events over the North Sea. To increase the reliability of the events that are discussed in detail, the rest of this chapter will describe only the events which are detected by five or six out of the six arrays.

When the source location is estimated by the least square iteration procedure, Equation (3.35) gives the variance-covariance matrix on the estimated location. The eigenvalues and eigenvectors of the variance-covariance matrix are used to calculate an error ellipse at a distance of two times the length of the standard deviation (i.e. 2σ), which corresponds to a 95% error ellipse for a Gaussian distribution. This agrees with the 95% error of the distributions of back azimuths within events, which is close to 2σ, Table 5.4. The location and the error ellipse at 2σ are plotted in Figure 4.2 for the events that are detected by five or six out of six arrays.

array	DBN	DIA	EXL	FLERS	IGADE	I26DE
standard deviation (σ) [deg]	2.5	2.0	2.7	2.5	3.7	2.6
95% error [deg]	4.8 (1.9σ)	4.6 (2.3σ)	6.0 (2.2σ)	5.3 (2.1σ)	7.8 (2.1σ)	5.4 (2.1σ)

Table 5.4: Standard deviation and 95 % error for DBN, DIA, EXL, FLERS, IGADE and I26DE. Values are calculated from the histograms of the azimuthal adjustment as shown in Figure 5.10.

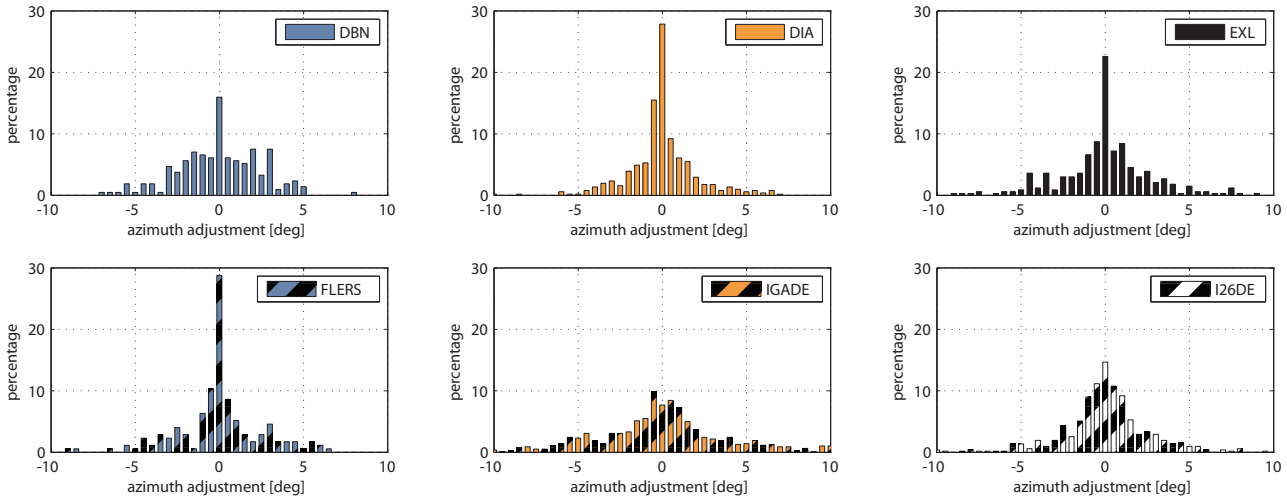


Figure 5.10: Histograms of the azimuthal adjustment towards the mean azimuth per event. Events of only one detection are not included in the distribution. Only events which were detected by at least four arrays were taken. Comparable distributions were observed when taking only the events detected by at least five arrays.

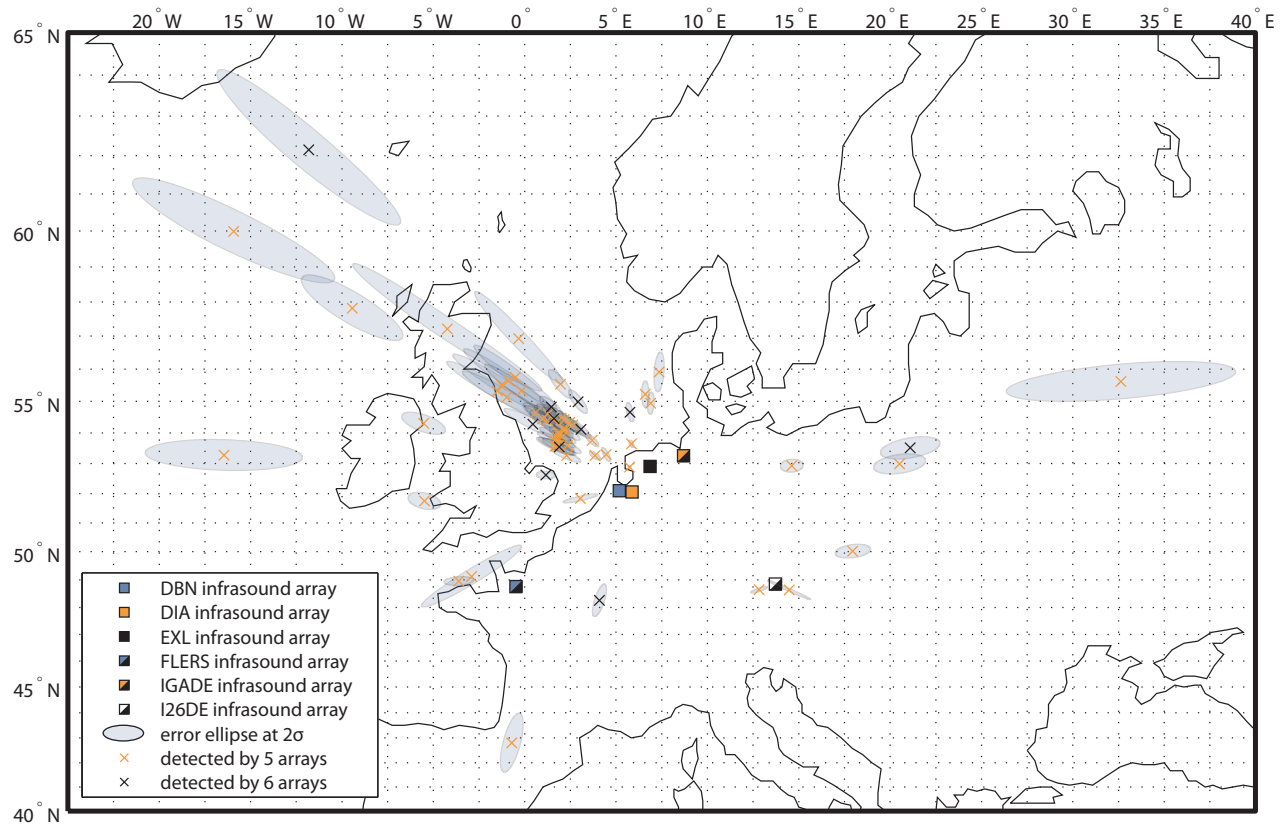


Figure 5.11: Best unbiased estimate of source locations of events detected by infrasound arrays DBN, DIA, EXL, FLERS, IGADE and I26DE. Only the events which were detected by five of six out of the six arrays are shown. 64 events were located and occurred in the period from January 1 until September 3. This is 0.4 events per day. Locations are indicated as crosses with precision shown as 2σ error ellipses in transparent blue. Locations are estimated from the average back azimuth of each event, at each array. The error ellipses are estimated from the back azimuth distribution between multiple detections within an event.

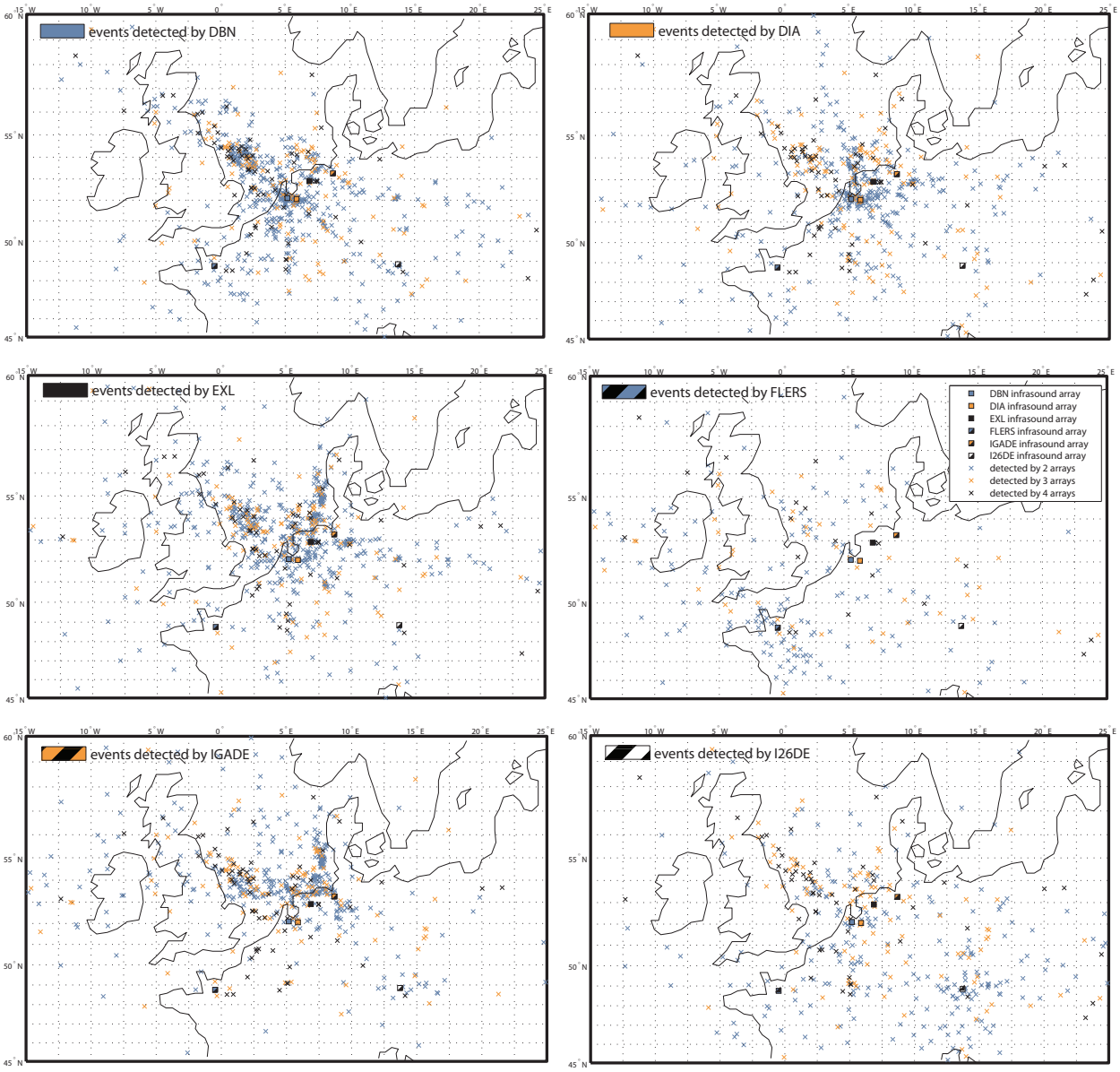


Figure 5.12: Infrasound event locations. The six maps show the events which were detected by DBN, DIA, EXL, FLERS, IGADE and I26DE respectively. The colour code used for each of these arrays is shown in the top left, the map's legend is shown in the middle-right map. Only events which were detected by the colour coded array and could be combined with detections from 1, 2 or 3 other arrays are shown. The events detected by five and six arrays are similar for at least five of the maps. These events are shown separately in Figure 5.11. As expected, the two arrays located furthest from the North Sea (FLERS and I26DE) show the least events on the North Sea. The events detected by 2 and 3 arrays are expected to be dominated by false events. These false events are caused by detections which have by chance a back azimuth and detection time that allow them to originate from the same event, but are in fact unrelated. These infrasound events did not occur in reality and are therefore described as false. Although the individual events are unreliable, the distribution of the events does show us in what region, and in what direction from the arrays, we can expect to find infrasound sources.

5.3 Statistics on results

Various event properties are shown in more detail in this section. The event properties are described in terms of the adjustment of the average back azimuth, the distribution of the events over day, week and year, the number of detections within events and the recorded event durations. Only the events detected by five and six arrays are described.

5.3.1 Back azimuth

In the location procedure the mean azimuths per event are adjusted for each array in such a way that the azimuths at the three arrays point to one location. A histogram of the required azimuth adjustments is shown in Figure 5.13. The average azimuth adjustment is comparable for each array. When using only the three KNMI arrays a smaller adjustment was observed for EXL, this effect has disappeared. It was expected that this smaller adjustment was caused by the geometry of the three KNMI arrays, which has changed due to the addition of FLERS, IGADE and I26DE. The average adjustment has also increased. This can be caused by the average distance between the arrays and the located events which has increased with respect to Chapter 4 but can also be caused by the use of more arrays and a change of geometry of the arrays.

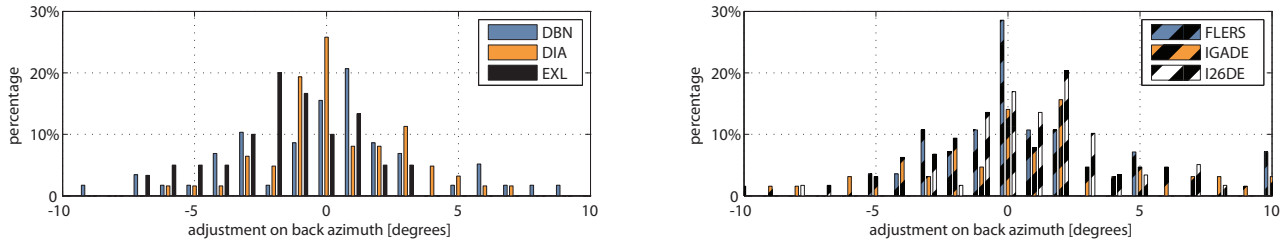


Figure 5.13: Histogram of the required azimuth adjustments for location. For each event the mean azimuth per array is calculated. This mean azimuth is adjusted such that it points to the estimated source location. This adjustment is shown in the histogram.

5.3.2 Time and date

The distribution of the events over the year, the days of the week and the hours of the day is shown in Figure 4.7. The distribution over the week and day strongly suggest that the events are related to human activity.

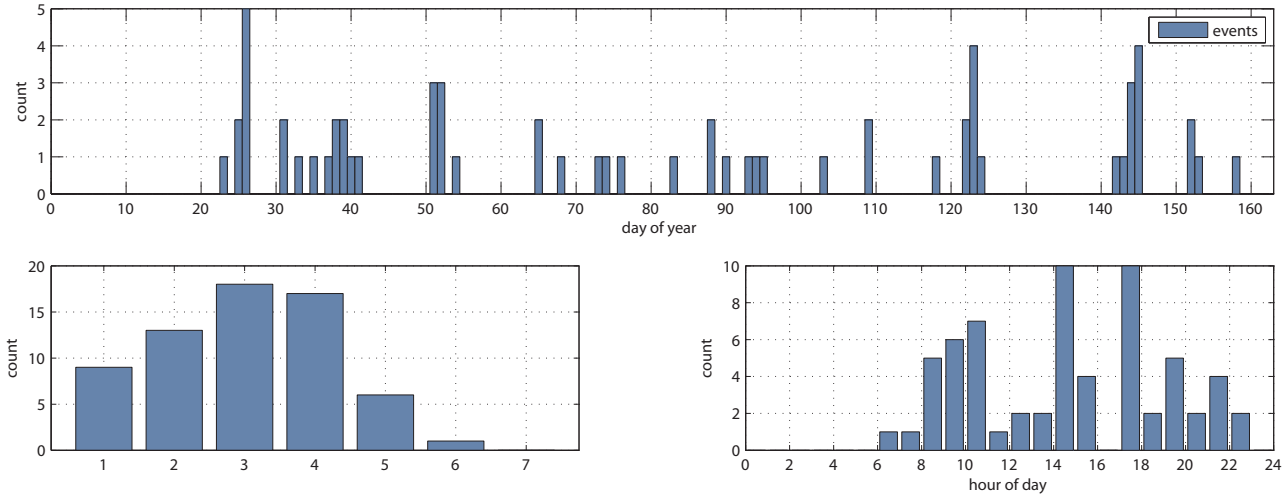


Figure 5.14: Top: distribution of events over the year 2007. Lower left: distribution of events over the week. Lower right: distribution of events over the day. 98% of the events were detected during working days, only 2% (1 event) on Saturday and none on Sunday. 95% of the events were detected between 7 am and 10 pm, one event was detected before 7 am and two events after 10 pm.

5.3.3 Detections per event

The number of detections per event are given in Table 5.5 and as a histogram in Figure 5.15. FLERS and I26DE detected the lowest number of events which is in line with what would be expected from the total number of detections, Table 5.2. These two arrays are located furthest from the four other arrays which are more clustered near the main source area over the North Sea.

array	DBN	DIA	EXL	FLERS	IGADE	I26DE	total
events detected by 2 arrays	594	470	604	254	492	322	1368
events detected by 3 arrays	140	144	143	73	141	127	256
events detected by 4 arrays	76	72	68	29	75	64	96
events detected by 5 arrays	47	51	49	17	53	48	53
events detected by 6 arrays	11	11	11	11	11	11	11
total number of events	868	748	875	384	772	572	1784

Table 5.5: The number of events as detected by DBN, DIA, EXL, FLERS, IGADE and I26DE. As expected, the number of events detected by only two of the six arrays is the largest and this number decreases for events which are detected by more arrays. There are 11 events which are detected by all 6 arrays. DBN and EXL show the largest number of detected events, with a very high number of events which are only detected by one other array. A sharp decline in events is visible between events detected by 2 arrays and those detected by 3 and 4 arrays. The amount of events which can be ascribed to false combinations of detections decreases when the event is detected by more arrays.

5.3.4 Signal duration

By subtracting the time of the first detections in an event from the time of the last detection an estimate for the signal duration can be calculated. An overview of the signal duration of all identified events is shown as a histogram in Figure 5.16 and on a map in Figure 5.17. Short signal detections of only less than 2.5 minutes form about one third of the detected events. IGADE and I26DE show less short duration detections than the other arrays. The median of the signal durations per event follows the signal duration of the individual arrays. This indicates that the events of long duration were registered for this timespan at more than one arrays, and it is therefore more probable that the source indeed emitted a signal throughout this timespan.

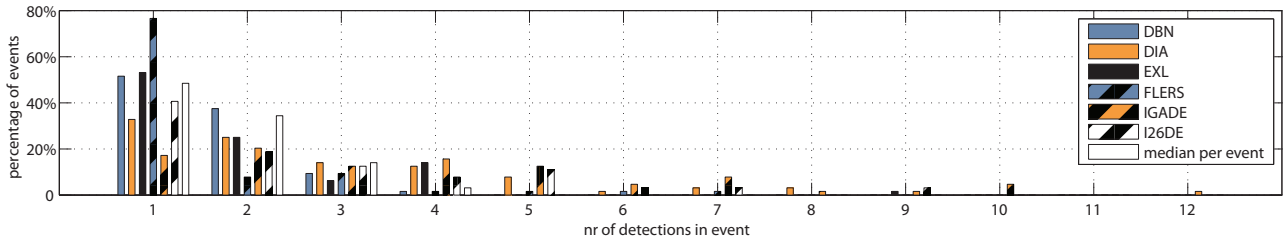


Figure 5.15: Histogram of the number of detections that are grouped into a single event. The histogram shows the number of detections at each array and the median of these three numbers for each event. Although IGADE shows up to twelve detections per event the median of the number of detections is four at maximum. Most events consist of only one detection per array.

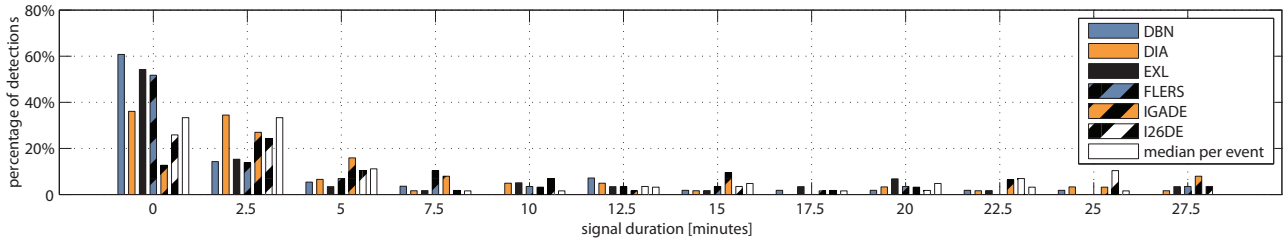


Figure 5.16: Histogram of the signal duration at the individual arrays for all events which were registered at five and six arrays.

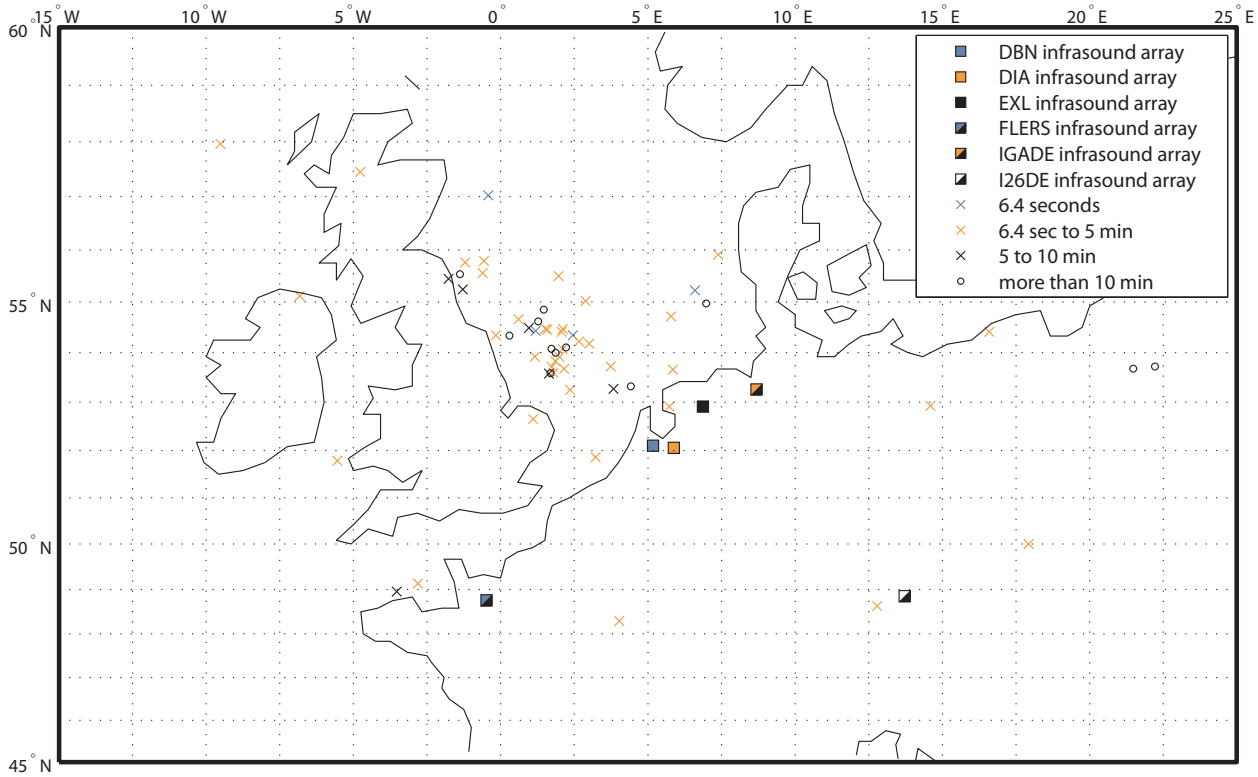


Figure 5.17: Event locations found by five and six arrays colour coded to the median signal duration of the event. The signal duration can differ at each array, for this map it is chosen to display the median of these three durations, see also Figure 5.16.

Chapter 6

Validation

The estimated infrasound source locations are validated on their existence in twofold. First, the locations as found by the three KNMI arrays over the North Sea will be compared against locations made by independent infrasound arrays and processing as performed at CEA. The event list from CEA shows a comparable source region as the KNMI event list. It was possible to indicate 9 events which appear in both lists within 2 minutes of each other and would generate a back azimuth at the KNMI arrays within 10 degrees of the detected back azimuths. These events form 21% of the CEA events and about 12% of the KNMI events. Second, the number of located events is compared against the number of events that will be located for random detections. The random detections are generated such that the time and azimuth distribution of the original data is preserved. By inserting random data into the location process it was found that 6% of the detections made by all three KNMI arrays can be contributed to random combination of detections. For all six arrays approximately 0% of the detections made by six arrays and 18% of the detections made by five out of six arrays can be contributed to random detections.

6.1 Comparison of KNMI locations with CEA locations

The infrasound research group at CEA located infrasound events from detections made by seven infrasound arrays located in France, Germany and Sweden, the location and geometry of the arrays are shown in Figure 5.1. The events were signalled and located using the method presented by Le Pichon *et al.* (2008). The method of Le Pichon creates an event for every combination of detections which can be combined such that the location estimate generates a celerity within 30% from 300 m/s. This forms a celerity region of 210 to 390 m/s as opposed to the region of 220 to 360 m/s as used in the research presented in this report. Multiple detections at one array are not combined, only the single detections which form a location generating a celerity closest to 300 m/s are used. No tolerances on back azimuth are defined. Estimates on the location accuracy or the probability on false events are not available. Because the decision if two detections could be related to one event is made after the location has been estimated, this method is potentially more computational intensive than a method based on search windows, such as presented in this report.

As part of the NERIES program the locations from the CEA research were made available for the entire year 2007 over the North Sea area. These locations will be referred to as the CEA event list. The area of the North Sea is selected as a rectangular box from $(52^o, -1^o)$ to $(57^o, 7^o)$. The KNMI event locations are compared to these locations. In the overlapping period (January 1 until September 3) and rectangular region the CEA event list gave 42 events and the KNMI event list gave 78 events. A plot of all locations over 2007, Figure 6.1, shows a comparable source area concentrating on the South-Western part of the region with almost no events located in the North-East.

When comparing the estimated source time 7 event pairs are found with an overlapping source time. The locations of 6 out of these 7 events are such that the back azimuth from the KNMI arrays towards the CEA locations is within 10 degrees of the observed azimuths at the KNMI arrays. For one event pair the back azimuth

from the KNMI arrays towards the CEA location deviated up to 130 degrees from the actual detections. These two events are therefore expected to be unrelated. In Figure 6.2 a histogram of the time and azimuth deviation is shown for the events which occur within one hour of each other and where the back azimuth deviates less than 20 degrees. Only one of these events shows a azimuth difference of more than 10 degrees, and this occurs at only one of the arrays. This observation supports the assumption that the shown events are indeed related and not random. For random combinations one would expect an equal distribution over the 40 degree region. The KNMI and CEA events that occurred within 2 minutes and within 10 degree back azimuth from the KNMI arrays are shown in Figure 6.1 by connected black circles. If one compares the locations as found from the first detections of an event as presented in Section 4.1.2 only five common events are found which comply to this criteria. Also a larger back azimuth deviation was found if the histogram of Figure 6.2 is reproduced for these events. This supports the choice to use the information of all detections within an event.

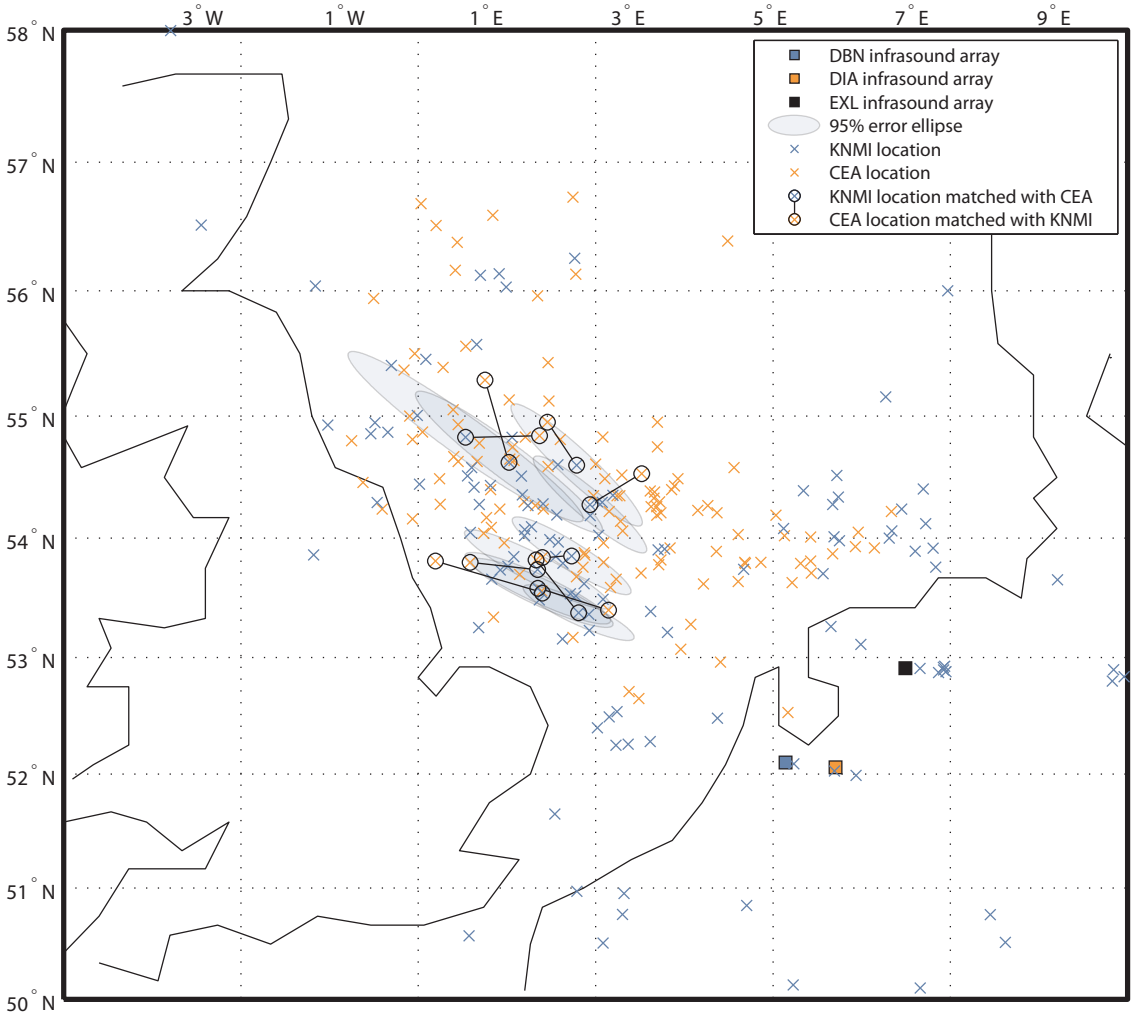


Figure 6.1: Comparison between KNMI and CEA locations on the North Sea. CEA events are shown for the region from (52^o, -1^o) to (57^o, 7^o). The events which are detected by KNMI are shown in blue and the CEA events are shown in orange. The error ellipses as estimated in Chapter 4 are shown in light blue but only for the events which also occurred in the CEA event list. These events are indicated by connected black circles. The combinations are made for all events which have estimated source times within two minutes at both lists and where the back azimuth of the detections at KNMI arrays does not deviate more than 10 degrees from the azimuth from the array towards the CEA location. Only two of the connected CEA events are located within the error ellipse of the KNMI location. It is undetermined whether this is caused by inaccuracy of the KNMI locations or by inaccuracy of the CEA locations.

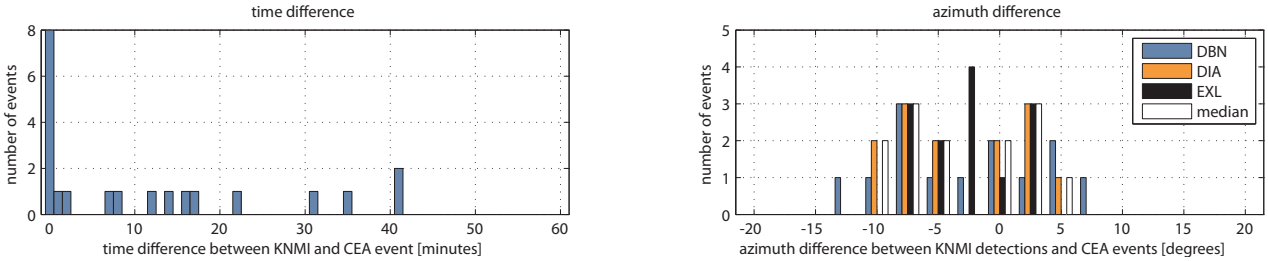


Figure 6.2: Histogram of the difference in event time and azimuth between KNMI and CEA events. Combination of events are made for all events which have estimated source times within one hour at both lists and where the back azimuth of the detections at KNMI arrays does not deviate more than 20 degrees from the azimuth from these arrays towards the CEA location. Only a single events shows an azimuth difference of more than 10 degrees, and this occurs at only one of the arrays. The CEA locations would generate a back azimuth at the KNMI arrays which is generally smaller, i.e. counter-clockwise rotated from the detections.

6.2 Events caused by false combination of events

Detections at multiple arrays are grouped into events if the back azimuth and detection time of these events allows them to originate from a common source, see Section 3.2. This leaves the possibility for random detections, which have by chance the correct back azimuth and detection time, to be registered as an event. The amount of detections that can be contributed to these false combinations can be evaluated by inserting random detection files into the location procedure. In Chapter 4 an average of only 1 out of 40 array detections was registered as an event by the three arrays. This supports the assumption that the data is dominated by detections of local sources. These detections are not random distributed and therefore the result would be biased if the time and back azimuth for this test would be made completely random. The detections are not evenly distributed over time as more detections are made during weekdays and daytime than in the weekends and at night. As shown in the polar histograms in Section 3.4.5 and 5.1.3 the azimuth is also unevenly distributed. To take this into account the time and back azimuth of the actual detections are taken, but the back azimuths are randomly distributed over the detection times. Then to make also the detection time random a random time between 0 and 1800 seconds is added to all individual detection times. The resulting number of events and the percentage with respect to the events as detected by the original data are given in Table 6.1 for the combination of the three KNMI arrays and in Table 6.2 for the combination of all six arrays. The location of the random events are plotted in Figure 6.3 and 6.4.

There are two methods to reduce random combinations of signals emitted by low energy local sources. The first method is to increase the minimum number of arrays at which the signal must be detected. A random combination at two arrays is much more probable than at three arrays. This effect is clearly visible in the probabilities shown in Table 6.1 and 6.2. For the second method the minimum required signal to noise ratio is increased. This has two effects. First, if one assumes that most local sources have only limited energy, than it is expected that signals arriving at higher energy than signals from local sources are easily recognizable. Second, in Section 3.4.4 it was argued that a signal from a source in the far field which has at arrival low energy left is less likely to be observed at multiple infrasound arrays than if this source emitted a more powerful signal. The influence of the signal to noise ratio is tested by increasing the minimum required signal to noise ratio and then evaluating the amount of detections that can be attributed to random combinations. For the three KNMI infrasound arrays the minimum required signal to noise ratio is controlled by the minimum required F-ratio, which is directly related to the signal to noise ratio, Equation (3.18). The random contribution and number of events for the recorded dataset are given in Figure 6.5 and Table 6.3. For the three additional arrays (FLERS, IGADE and I26DE) not the F-ratio but the consistency was used to accept or reject events. In Figure 6.6 and Table 6.4 the effect of changing the minimum required consistency for FLERS, IGADE and I26DE is shown for the combination of six arrays.

From the comparison of the results for all six arrays and the three KNMI arrays it becomes clear that increasing the number of arrays can also cause an increase in false events. The arrays that one wishes to use should therefore be chosen wisely. It is also possible to search for events over all available arrays, but then displaying only the events which are detected by a subset of these arrays. In this way the advantage of using all available information is combined with a lower amount of false events. After a subset has been chosen, the signal to noise ratio can be tuned such that an acceptable amount of false events is generated.

	DBN	DIA	EXL	total
events detected by 2 arrays	298 (57%)	380 (79%)	383 (69%)	537 (69%)
events detected by 3 arrays	5 (4%)	5 (4%)	5 (4%)	5 (4%)
total number of events:	303 (46%)	385 (62%)	387 (56%)	541 (59%)

Table 6.1: The amount of false combinations of detections that propagate into events for the combination of DBN, DIA and EXL. The number of events are the average over 5 runs on the original detections with random redistributed back azimuths. The percentages shown give the difference compared to the amount of events as generated by the original detections. Based on this comparison, the detections made by only two arrays are found to be unreliable.

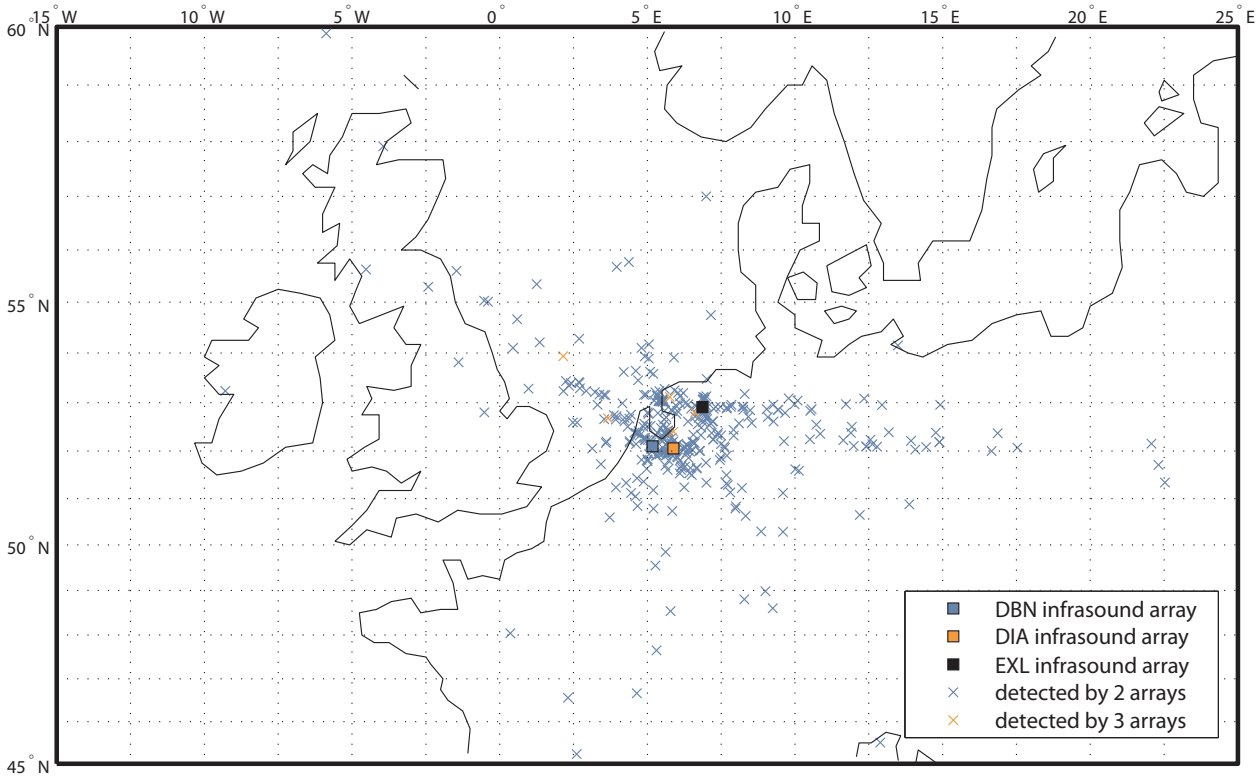


Figure 6.3: Source locations as found for random detections for DBN, DIA and EXL. A high concentration of events is found near the infrasound arrays, similar to what was found for the events detected by 2 out of 3 arrays for the recorded detections.

	DBN	DIA	EXL	FLERS	IGADE	I26DE	total
events detected by 2 arrays	82%	97%	88%	157%	80%	153%	98%
events detected by 3 arrays	87%	97%	86%	155%	85%	141%	105%
events detected by 4 arrays	45%	56%	40%	87%	45%	59%	53%
events detected by 5 arrays	12%	17%	15%	32%	18%	20%	18%
events detected by 6 arrays	0%	0%	0%	0%	0%	0%	0%
total number of events:	76%	86%	80%	141%	74%	128%	95%

Table 6.2: The amount of false combinations of events that propagate into events for the combination of DBN, DIA, EXL, FLERS, IGADE and I26DE with a consistency larger than 8. The number of events are the average over 5 runs on the original detections with random redistributed back azimuths. The values for the individual loops deviated up to 20% from the shown values. The percentages shown give the difference compared to the amount of events as generated by the original detections. Based on this comparison, the detections made by only two, three and four arrays are found to be unreliable.

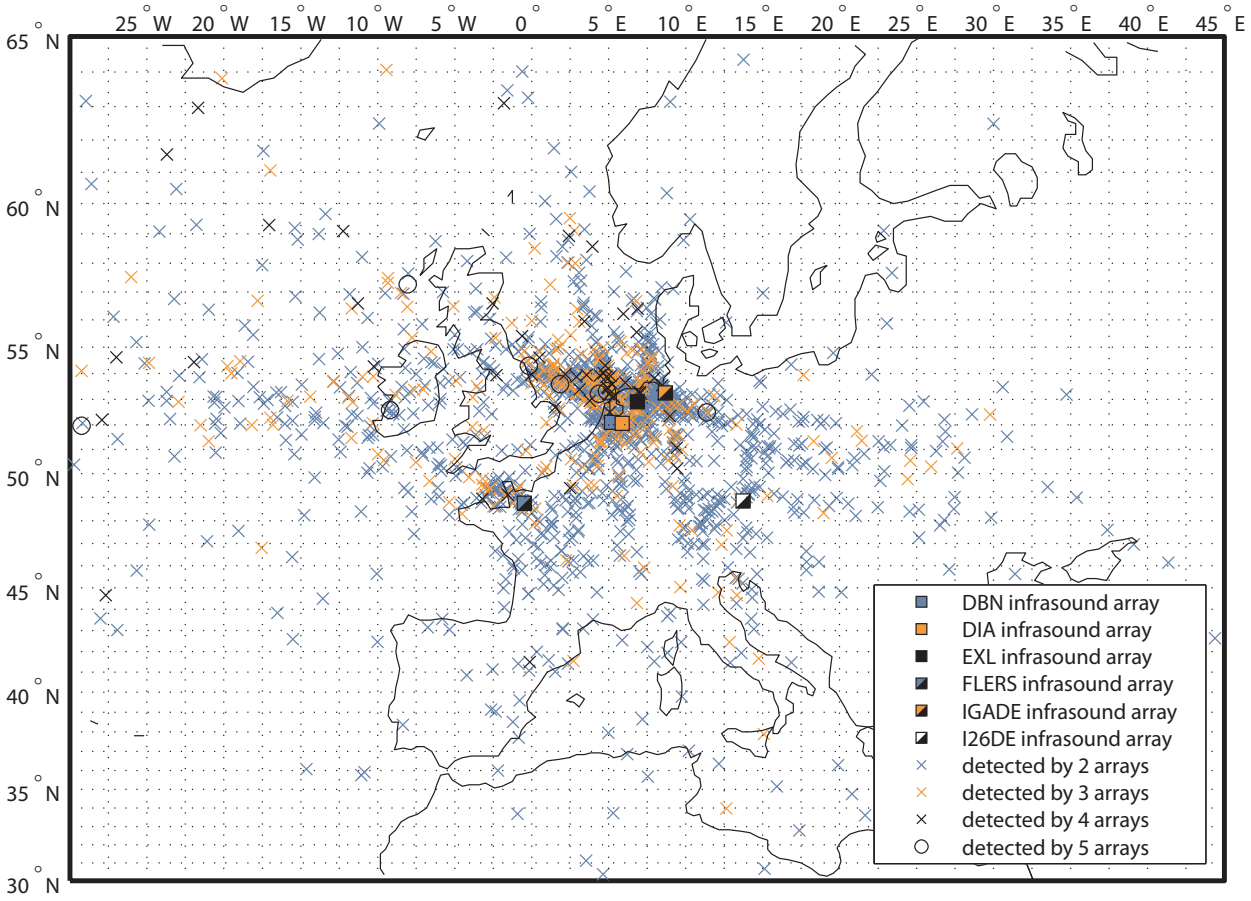


Figure 6.4: Source locations as found for random detections for DBN, DIA, EXL, FLERS, IGADE and I26DE. Most false sources are found near the four closest located arrays, DBN, DIA, EXL and IGADE. Several lines of false events are shown, these lines usually point in the direction of an active local source.

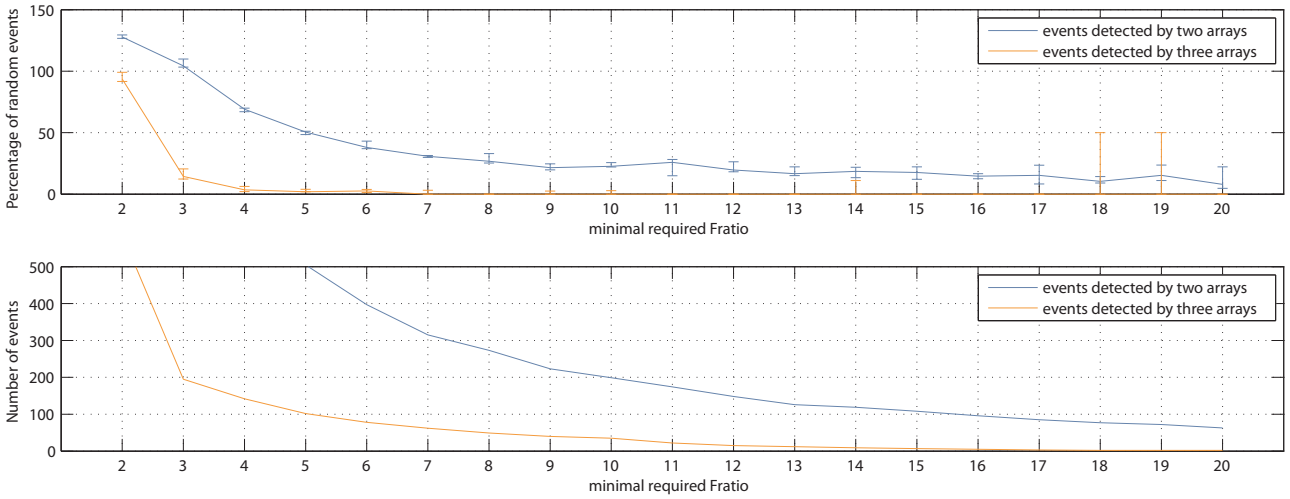


Figure 6.5: Top: graph of the contribution of false events, generated by random data, given for different minimum F-ratio's, for the three KNMI arrays. Bottom: Number of detections as generated from the original dataset. The graph is based on the values presented in Table 6.3. The shown percentage is the median value for 5 random datasets, error bars show the minimum and maximum percentage of these 5 datasets.

	Events detected by 2 arrays				Events detected by 3 arrays			
	Events	Random contribution			Events	Random contribution		
F-ratio		median	minimum	maximum		median	minimum	maximum
2	6055	128%	127%	129%	586	94%	92%	99%
3	1531	104%	103%	110%	195	14%	12%	21%
4	778	69%	67%	70%	141	4%	2%	6%
5	505	50%	49%	51%	102	2%	2%	4%
6	397	38%	37%	43%	78	3%	1%	4%
7	315	31%	30%	31%	62	0%	0%	3%
8	273	27%	25%	33%	49	0%	0%	0%
9	223	22%	20%	25%	40	0%	0%	3%
10	199	23%	22%	26%	35	0%	0%	3%
11	174	26%	15%	28%	22	0%	0%	0%
12	148	20%	18%	26%	15	0%	0%	0%
13	126	17%	15%	22%	12	0%	0%	0%
14	119	18%	13%	22%	9	0%	0%	11%
15	108	18%	12%	22%	7	0%	0%	0%
16	96	15%	13%	17%	5	0%	0%	0%
17	85	15%	8%	24%	3	0%	0%	0%
18	77	10%	9%	14%	2	0%	0%	50%
19	72	15%	11%	24%	2	0%	0%	50%
20	63	8%	5%	22%	2	0%	0%	0%

Table 6.3: Number of events and the random contribution as a function of Fisher ratio for DBN, DIA and EXL. The number of events is calculated from the recorded dataset which is filtered on a minimum required Fisher ratio as shown in the left column. For each minimum Fisher ratio five random datasets are generated by redistributing the detections randomly over the original detection times and adding a random time between 0 and 1800 seconds. These datasets are processed similar to the original dataset. For each 5 datasets at equal minimum Fisher ratio the median, minimum and maximum ratio of the found events with respect to the original number of events is displayed. Notice the maximum contribution of random generated events of 50% for the events found by three arrays at a minimum Fisher ratio of 18 and 19. This number is caused by the fact that only 2 events were located in the original dataset and one of the random datasets generated a single event.

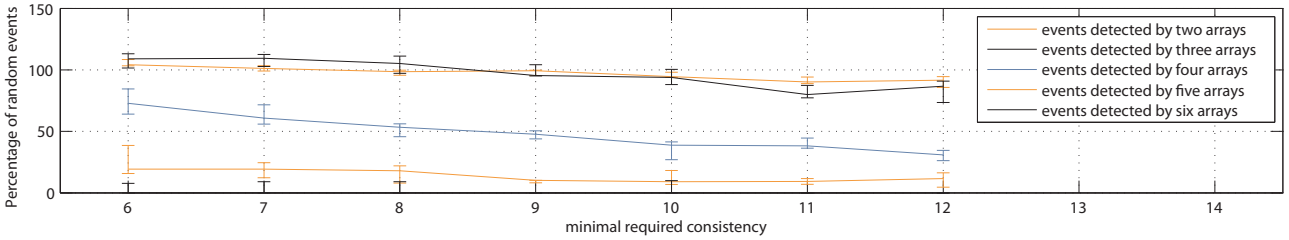


Figure 6.6: Graph of the random contribution for the events found for different minimum consistency for all 6 infrasound arrays. The graph is based on the values presented in Table 6.4. The shown percentage is the median value for 5 random datasets, error bars show the minimum and maximum percentage of these 5 datsets. At DBN, DIA and EXL the minimum required F-ratio is kept at 4.

Minimum consistency	6	7	8	9	10	11	12
Events detected by 2 arrays	1696	1639	1580	1493	1434	1401	1360
Random contribution 2 arr.	104%	101%	98%	99%	94%	90%	92%
Events detected by 3 arrays	335	295	285	264	259	255	241
Random contribution 3 arr.	109%	109%	105%	95%	94%	80%	87%
Events detected by 4 arrays	103	102	105	107	111	110	107
Random contribution 4 arr.	73%	61%	53%	48%	39%	38%	31%
Events detected by 5 arrays	57	57	50	49	44	43	43
Random contribution 5 arr.	19%	19%	18%	10%	9%	9%	12%
Events detected by 6 arrays	13	11	11	10	10	10	10
Random contribution 6 arr.	0%	0%	0%	0%	0%	0%	0%

Table 6.4: Number of events and the random contribution as a function of Fisher ratio for DBN, DIA, EXL, FLERS, IGADE and I26DE. The number of events is calculated from the recorded dataset which is filtered on a minimum required consistency as shown in the upper row. For each minimum consistency five random datasets are generated by redistributing the detections randomly over the detection times and adding a random time between 0 and 1800 seconds. These datasets are processed similar to the original dataset. For each set of 5 datasets at equal minimum consistency the median, minimum and maximum ratio of the found events with respect to the original number of events is displayed. At DBN, DIA and EXL the minimum required F-ratio is kept at 4. Therefore the random generated events shown in Table 6.3 at an F-ratio of 4 are also present in this table. The amount of events and the random contribution decreases slowly by increasing the minimum required consistency. It is recommended to test the detections at FLERS, IGADE and I26DE on F-ratio in the future to make them comparable with the KNMI infrasound arrays.

Chapter 7

Conclusions and recommendations

Generally it can be concluded that it proved possible to develop an automated method to estimate the geographic location of infrasound sources. The developed method was succesfully applied to infrasound recordings of infrasound arrays in the Netherlands, Germany and France. Source location allows for better identification of infrasound sources and can be used to distinguish detections of far field sources from local sources. Using the source location one can potentially evaluate signal travel paths through the atmosphere, which could bring the actual use of infrasound for atmospheric studies closer to our possibilities. More detailed conclusions will be given on the infrasound source locations found by this procedure, the detection of local sources and on the causes for and avoidance of the creation of false events by combining unrelated detections. After these three paragraphs recommendations are given for further research.

The developed procedure was applied on detections over a frequency band of 1 to 10 Hz made by three infrasound arrays located in the Netherlands and to the detections of three more arrays located in France (1 array) and Germany (2 arrays). Locations of the events were found dominantly on the western part of the North Sea. A comparison with the locations of gas flaring showed a large difference in location. It is therefore expected that most detections originate from a different source. The events could be related to military jet exercise (sonic booms) or the detonation of explosives (e.g. clearing of naval mines), but this could not be verified due to the lack of ground truth data. The events show no correlation with the locations expected for gas flaring. Sources that were found over land showed a strong correlation with military presence (e.g. practice grounds and military airfields). 16% of the events could be recorded for over 10 minutes, suggesting that the source emitted multiple signals over time. Almost all events occurred during weekdays (99%) and between 7 am and 10 pm local time (98%), which strongly suggests that signals from manmade sources are dominant over the used frequency band. The uncertainties on source location and detection time were found to make direct estimates of signal celerity from these parameters unreliable. The expected errors are found to be too large to reliably distinguish different atmosperic returns from the celerity estimate. If one wishes to distinguish different atmospheric returns it is therefore advised to compare the made detections with modeled detections made by ray tracing.

Half of all detections at the De Bilt infrasound array (DBN) were received over 4 small azimuth bands pointing to overpasses of roads and small tunnels. Also at Deelen infrasound array (DIA), Exloo infrasound array (EXL) and the IGADE infrasound array near Bremen, Germany, very large numbers of detections were found over small back azimuth bands. Exact local sources are often difficult to find from maps and aerial imagery and their signals are usually indistinguishable from signals that travelled from a distant source. It is expected that the majority of all detections over the used frequency band (1 - 10 Hz) can be ascribed to local sources such as traffic at tunnels and overpasses, local industry or military practice.

Avoiding the combination of unrelated detections into false events was found to be a major subject. These events are caused by detections which have by chance a back azimuth and detection time that allow them to originate from the same event, but are in fact unrelated. These infrasound events did not occur in reality and are therefore described as false. The amount of false events was found to increase if the number of detections per array increases and if the number of used arrays is increased. The amount of false events was found to decrease if the number of arrays by which an event is detected increases and if the signal to noise ratio of the detections within an event increases. Filtering was adjusted such that the number of false events generated by the used datasets was acceptable (varying from 0 to 18%, dependent on the number of arrays used). To control the amount of false events, a four step approach is advised. First, the removal of detections from local sources should be optimized. Second, only the detections made by arrays close to the investigated source area should be used to form events. Third, the minimum required signal to noise ratio can be adjusted to control the amount of false detections that is allowed. Fourth, related detections at other arrays are added for the accepted events only. By adding information of other arrays after event identification the number of false events is not increased by using more arrays.

Recommendations:

- The next major step in improving the source location estimates is by modelling the atmospheric propagation. It is only a small step to model the atmospheric propagation by ray tracing. This propagation can be modelled from the estimated source locations to the known array locations. The estimated source locations are expected to be close enough to the true locations to make these propagations valid. The crucial step lies in a robust method to compare the modelled propagation paths to the made detections. If this link can be made, the source location estimate can be adjusted to include the modelled atmospheric propagation. The modelled propagation will predict the azimuthal deviation caused by the atmosphere and the azimuth of the detection can be adjusted by this amount to point more accurately to the source location.
- Even without adjusting the source location it would be insightful to see whether multiple detections within events can be explained by the predicted atmospheric returns from an atmospheric model. If the detections show resemblance with the model the deviation between model and observation can be used to adjust the atmospheric model.
- The events detected by three arrays for DBN, DIA and EXL gave most energy in the frequency range from 1 to 6 Hz. For DBN it was observed that detections that could be related to local sources had considerable high frequency content above 6 Hz. It is therefore recommended to lower the used frequency bandwidth from 1 - 10 Hz to 1 - 6 Hz.
- For a large area search for infrasound sources, such as by combining the 6 infrasound arrays, it is advised to use a requirement for each event to be detected by the closest infrasound arrays. By setting this requirement the amount of false events, caused by the combination of unrelated detections, is decreased.
- Ground truth information can give important insight in the capabilities of the developed method and support better validation. Records of military jet practice and gas flaring can be very valuable to explore detection capabilities. Incidental events which are well documented both in space and time can be used to validate the estimated event time and location.

Some general remarks to solve limitations in the current processing procedure:

- Improve the standard deviation estimate on back azimuth for single detections at DIA. The standard deviation on back azimuth is estimated by fitting a Lorenz profile over the Fisher ratios calculated on a slowness grid. When the peak becomes too small (i.e. a few pixels) the fit becomes unreliable. This can be overcome by fitting a Lorenz profile over Fisher ratios calculated over a higher resolution slowness grid around the peak position. This higher resolution grid shows the peak over more pixels which should make it possible to fit a Lorenz function.
- In current processing, only the single detection with maximal F-ratio of each time window is selected. It is however possible that multiple atmospheric returns arrive within one time window of 256 seconds. To overcome this, one could save all detections above a certain F-ratio threshold within this time window. To reduce computation time this can be applied as post processing on the time windows corresponding to identified events.
- Investigate the possibilities of compensating for temporary variations in noise to get more reliable detections. Promising results were shown by Arrowsmith *et al.* (2008), by which the F-ratio was scaled to the average time window F-distribution, assuming that the average time window F-distribution is noise dominated.
- It is recommended to lower the lower frequency boundary for DBN from 2.8 Hz to 1 Hz. Although in Section 3.4.1 it was shown that it is not possible to make a precise azimuth estimate for such low frequency signals, the lower frequency information can improve detections which also have a higher frequency content.

Bibliography

Arrowsmith, S.J., Whitaker, R., Taylor, S.R., Burlacu, R., Stump, B., Hedlin, M., Randall, G., Hayward, C., & ReVelle, D. 2008. Regional monitoring of infrasound events using multiple arrays: application to Utah and Washington State. *Geophysical Journal International*, **175**, 291300.

Bakker, G., de Munck, J.C., & Strang van Hees, G.L. 1995. *Radio positioning at sea. geodetic survey computations, least squares adjustment.* rev. ed. edn. Delft : Universiteitsdrukkerij TU, 1995.

Brown, D.J., Katz, C.N., Le Bras, R., Flanagan, M.P., Wang, J., & Gault, A.K. 2002. Infrasonic signal detection and source location at the Prototype International Data Centre. *Pure and Applied Geophysics*, **159**, 10811125.

Brown, P.G. 2002. Multi-station infrasonic observations of two large bolides: signal interpretation and implications for monitoring of atmospheric explosions. *Geophysical Research Letters*, **29**, 1–6.

Caljé, L. 2005. *Exploring boundaries of the Fisher and PMCC signal-detectors using infrasound signals.* M.Phil. thesis, Universiteit Utrecht.

Cansi, Y. 1995. An automatic seismic event processing for detection and location: The PMCC method. *Geophysical Research Letters*, **22**, 1021–1024.

Cansi, Y., & Klinger, Y. 1997. An automated data processing method for mini-arrays. *Newsletter of the European–Mediterranean Seismological Centre*, **11**, 2–4.

Ceplecha, Z., Borovička, J.Í., Elford, W.G., Revelle, D.O., Hawkes, R.L., Porubčan, V.Í., & Šimek, M. 1998. Meteor phenomena and bodies. *Space Science Reviews*, **84**, 327471.

Elvidge, C.D., Kimberly, B., Tuttle, B., Ziskin, D., Tilottama, G., Zhizhin, M., & Pack, D. 2009 (august). *Improving Satellite Data Estimation of Gas Flaring Volumes.* year two final report to the GGFR. NOAA National Geophysical Data Center, Cooperative Institute for Research in the Environmental Sciences (CIRES), Space Research Institute Russian Academy of Science, The Aerospace Corporation.

ESRI. 2009. *Flexviewer for infrasound sources.* http://arcgis93.esri.nl/FlexViewer_Infrasound/index.html.

Evers, L.G. 2008. *The inaudible symphony: on the detection and source identification of atmospheric infrasound.* Ph.D. thesis, Delft University of Technology.

Evers, L.G., & Haak, H.W. 2007. Infrasonic forerunners: Exceptionally fast acoustic phases. *Geophysical Research Letters*, **34**, L10806.

Garcés, M.A., Hansen, R.A., & Lindquist, K.G. 1998. Traveltimes for infrasonic waves propagating in a stratified atmosphere. *Geophysical Journal International*, **135**, 255–263.

Gossard, E.E., & Hooke, W.H. 1975. *Waves in the atmosphere: Atmospheric infrasound and gravity waves - Their generation and propagation.* Vol. 2. Elsevier Scientific Publishing Company.

Johnson, J.B. 2003. Generation and propagation of infrasonic airwaves from volcanic explosions. *Journal of Volcanology and Geothermal Research*, **121**, 1–14.

KNMI. 2007. *potential wind speed measurements.* `http://www.knmi.nl/klimatologie/onderzoeksgegevens/potentiele_wind/`.

KNMI. 2008. *logarithmic wind speed profile.* `http://www.knmi.nl/samenw/hydra/faq/profile.htm`.

Le Pichon, A., Vergoz, J., Herry, P., & Ceranna, L. 2008. Analyzing the detection capability of infrasound arrays in Central Europe. *Journal of Geophysical Research*, **113**(D12115).

Melton, B.S., & Bailey, L.F. 1957. Multiple Signal Correlators. *Geophysics*, **22**, 565–588.

OpenStreetMap. 2009. *Mapnik map.* http://www.openstreetmap.org.

Press, W.H., Flannery, B.P., Teukolsky, S.A., & Vetterling, W.T. 1992. *Numerical Recipes in FORTRAN 77: The Art of Scientific Computing (v. 1).* 2 edn. Cambridge University Press.

Rees, W.G. 2001. *Physical Principles of Remote Sensing.* 2 revised edn. Cambridge University Press.

Revelle, D.O., & Whitaker, R.W. (eds). 1996 (July). *Lamb Waves from Airborne Explosion Sources: Viscous Effects and Comparison to Ducted Acoustic Arrivals.* Ecole Centrale de Lyon, France. 7th Long Range Sound Propagation Symposium.

Rost, S., & Thomas, C. 2002. Array seismology: Methods and applications. *Rev. Geophys*, **40**(3), 1008.

Royal Air Force, Headquarters Air Command. 2009 (June). *Freedom of Information Act 2000, Request 13478.* personal communication.

Shumway, R.H. 1971. On detecting a signal in N stationarily correlated noise series. *Technometrics*, **10**, 523–534.

Smart, E., & Flinn, E.A. 1971. Fast Frequency-Wavenumber Analysis and Fisher Signal Detection in Real-Time Infrasonic Array Data Processing. *Geophysical Journal International*, **26**, 279–284.

Snyder, J.P. 1987. *Map projections - a working manual.* 1395 edn. U.S. Geological Survey Professional Paper.

Symons, G.J. 1888. *The eruption of Krakatoa and subsequent phenomena.* Trübner & Co., London.

Teunissen, P.J.G. 1995. *Adjustment theory, an introduction.* first edition edn. Series on Mathematical Geodesy and Positioning. Leeghwaterstraat 42, 2628 CA Delft, The Netherlands: VSSD.

Teunissen, P.J.G., Simons, D.G., & Tiberius, C.C.J.M. 2006. *Probability and Observation Theory.* Department of Earth Observation and Space systems (DEOS). Lecture notes AE2-E01.

Vincenty, T. 1975. Direct And Inverse Solutions Of Geodesics On The Ellipsoid With Application Of Nested Equations. *Survey Review*, **22**, 88–93.

Weisstein, & Eric, W. 2002. *Noncentral F-Distribution.* From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/NoncentralF-Distribution.html.