University of Amsterdam

MSc Chemistry

Track: Analytical Chemistry

Master Thesis

Efficient compliancy monitoring: Comparison of both airborne and landside sniffing and spectrometric methods to provide direct control on the sulfur emission of ships.

---

by

Gerrit-Jan de Bruin, BSc

11065737

October 20, 2017

42 ECTS
January - October 2017

The Human Environment and Transport Inspectorate:
Jasper van Vliet, MSc

University of Amsterdam:
dr. Johan Westerhuis
dr. Wim Kok

Inspectie Leefomgeving en Transport
Ministerie van Infrastructuur en Milieu

Abstract

There is global interest in improving air quality to prevent large number of premature deaths. Especially in Northern Europe the emission of sulfur dioxides are of major concern. In order to enforce legislation, regular inspection is necessary.
The aim of this thesis was to compare different technologies which are able to remotely measure sulfur dioxide concentrations and to determine the emissions from an individual ship. Subsequently, measurements performed from these techniques were compared. Several ways to separate compliant and non-compliant ships were discussed. The compliance rate which these different approaches yielded were reported.
Four techniques which are able to measure fuel sulfur content, DOAS, LIDAR, NDIR and CRDS, were explained. The uncertainty at commonly encountered emission levels were merely reported in literature. Measurements from these techniques, which were provided by (governmental) agencies, showed a lack of validation with known fuel sulfur contents. The training and validation set were too small to allow supervised classification between ships sailing on compliant or non-compliant fuel. In order to estimate the number of ships running on compliant fuel, the two classes were modeled using a bimodal distribution with a normal and shifted lognormal mode using an EM-algorithm. In this way it was estimated that only 3 % was non-compliant. This value is low compared to high quality in-port measurements. It may be caused by several assumptions which were not met. For example, normality was required but not met.
In order to adequately perform remote sulfur controlling, more emphasis should be on cross validation. The few examples of cross validation which were present in the data, showed great discrepancies. The use of longer integration times will allow avoiding any mixing with other pollution sources in the neighborhood of the measured ship for future measurements. The use of helicopters, sample canisters or unmanned aerial vehicles can lower bias. In addition, duplicate set-ups can improve precision from measurements of remote sulfur fuel content.

Keywords

SECA control, IMO Annex VI compliance, data analysis, expectation maximization, sulfur dioxide, carbon dioxide

Samenvatting

Er is wereldwijde interesse om luchtkwaliteit te verbeteren, zodat vroegtijdige sterfte tegengegaan kan worden. Vooral in Noord-Europa is de emissie van zwaveldioxide vanuit scheepvaart een groot gezondheidsprobleem. Hoewel er strenge emissienormen zijn, laat de controle te wensen over.
Verschillende technologiën, die in staat zijn de zwaveldioxide emissie van schepen te meten, werden met elkaar vergeleken. Vervolgens werd de data geanalyseerd die met deze technieken waren geproduceerd. Verschillende manieren werden gebruikt om de schepen die niet voldoen aan de wetgeving te kunnen scheiden van degene die dat wel doen.
Vier verschillende technieken die het zwavelgehalte van brandstof van afstand kunnen meten werden uitgelegd, namelijk DOAS, LIDAR, NDIR en CRDS. De onzekerheid in de meting van het zwavelgehalte werd nauwelijks gerapporteerd voor gangbare emissies. Metingen van deze technieken, die werden uitgevoerd door (overheids)organisaties, waren erg gebrekkig gevalideerd met schepen die een bekende zwavelgehalte hadden. Er waren daardoor te weinig training en validatiewaarden om gecontroleerde classificatiemethoden te gebruiken. Deze classificatie tussen schepen die op legale of illegale brandstof varen werd daardoor bereikt door beide groepen schepen te modelleren. Een bimodale distributie met een normaal verdeelde modus en een verplaatste lognormaal verdeelde modus werd gebruikt. De parameters van deze distributie werden bepaald met behulp van een EM-algoritme. Op deze manier werd gevonden dat maar 3 % van de schepen zich niet aan de wetgeving hield wat betreft zwaveluitstoot. Deze waarde is laag vergeleken met metingen aan boord. Dit kan komen doordat aan verschillende aannames niet werd voldaan.
De zwaveluitstoot had beter gemeten kunnen worden als er meer nadruk was gelegd op kruisvalidatie. De enkele gevallen van kruisvalidatie die beschikbaar waren gedurende dit onderzoek, lieten grote discrepanties zien tussen de gemeten en gevalideerde waarden. Verder onderzoek zou zich moeten richten op het verlengen van de integratietijd en het voorkomen van het meten van zwaveldioxide en koolstofdioxide uit andere bronnen. Dit zou kunnen door op volle zee te gaan meten met behulp van drones of helikopters. In-duplo meten zou ook de precisie kunnen vergroten van de metingen op afstand van zwaveldioxide uitstoot.

Zoekwoorden

SECA, IMO Annex VI naleving, data analyse, expectation maximization, zwaveldioxide, koolstofdioxide

List of Figures

List of Tables

Contents

Chapter 1

Introduction

More than 90 % of global transport of commodities is performed by shipping. [1] Half a billion metric ton of goods are processed in Rotterdam, Europe's biggest port. [2] The large volumes of goods being processed and the associated amount of ship movements, impose an environmental problem. Shipping is a significant contributor to the emission of several pollutants. Main compounds emitted are carbon dioxide (CO_2), nitrogen oxides (NOx), carbon monoxide (CO), volatile organic compounds (VOCs), sulfur dioxide (SO_2), black carbon (BC) and particulate organic matter (POM). [3] While all these compounds influences on atmosphere and climate, especially SO_2 leads to serious human health issues. Corbett *et al.* reported 60,000 cardiopulmonary and lung cancer deaths per year globally through shipping-related emissions in 2007. [4] The clean air for Europe (CAFE) program estimated in 2005 that the exposure to air pollution through particulate matter (PM) led to an approximate 2-year reduction in statistical life expectancy in the Netherlands. [5]

1.1 Regulation

Legislative bodies recognized the imminent impact to the public health and has implemented legal restrictions on the sulfur content of marine fuels. A reduction of 82 % was achieved in the Netherlands since 1990. The maritime world followed later; the decrease started as late as 2005. Transport modes other than shipping did not have a significant share in the emission sulfur dioxide since 2005. The shipping sector contributed for 79 % on the total reduction of SO_2 since 2015, as shown in Figure 1.1. [6]
This reduction was achieved by the international maritime organization (IMO) through the prohibition of the use of high sulfuric fuels in the North Sea. The area marked in Figure 1.2 was designated as sulfur emission control area (SECA). The maximum allowed value of the fuel sulfur content (FSC) is only 0.1 wt.% in this area. This limit is only in force since 2015, as shown in Figure 1.3. In this figure is also shown that by 2020 the allowed FSC will drop from 3.5 wt.% to 0.5 wt.% in non-SECA. [7]
However, fuels with a low FSC come at a price, since these fuels needs additional processing to remove the sulfur. Kalli *et al.* calculated that ship transport expenses increased with 50 % to 70 % within the North Sea. [8] Hence the incentive to not comply with this legislation is great in economic sense.
To ensure a fair level playing field, the companies that do not comply need to be identified and punished. This will ensure maximum environmental gain from the legislation.

The regulation is part of MARPOL legislation in Annex VI. [9] In the Netherlands, it is the task of the 'Inspectie Leefomgeving en Transport' (human environment and transport inspectorate) (ILT) to enforce the legislation.

1.2 Monitoring

The most common solution to monitor compliance is to non-selective board ships while they are at berth. A fuel sample is then collected and subsequently analyzed in the lab. Legal actions will be taken if the FSC is more than 0.1 %(m/m). ILT uses handheld X-ray fluorescence (XRF) apparatus to check the FSC. If the apparatus reports a FSC above a certain threshold, the sample is sent to the lab for a more accurate analysis. The lab quantifies the FSC using International Organization for Standardization (ISO) 8217:2017. While this strategy will lower the amount of samples sent to the lab, there is still one important drawback. Only a few ships can be inspected. Some of the ships are visited because they were tagged by another country as possibly non-compliant. All other ships were selected at random.
The ILT recognized the need of monitoring a large amount of ships to ensure a level playing field. Inspection capacity is however limited. To make efficient use of the available resources, ILT would like to operate risk-based. [11] They requested Netherlands organization for applied scientific research (TNO) to perform remote measurements of the FSC. This allows them to determine compliance from much more ships than they could with on-board inspection.
The remote monitoring can either be performed stationary or from a moving position. The former uses the motion of the plume towards a fixed point along the coastline or upon a bridge. The moving platform can either be on an aircraft, unmanned aerial vehicle (UAV), helicopter or another ship. While expensive aircraft can sample much more ships in a given time, it is preferred over using a ship to measure the FSC from another ship. Aircraft can be used both unmanned and manned and rotary or fixed wing.
Earlier analysis of the measurements in this thesis showed compliance levels varying from 82 % to 99.4 %. [12] The range of these values is broad. It is clear that a uniform and logical view needs to be developed on how to report the compliance levels.

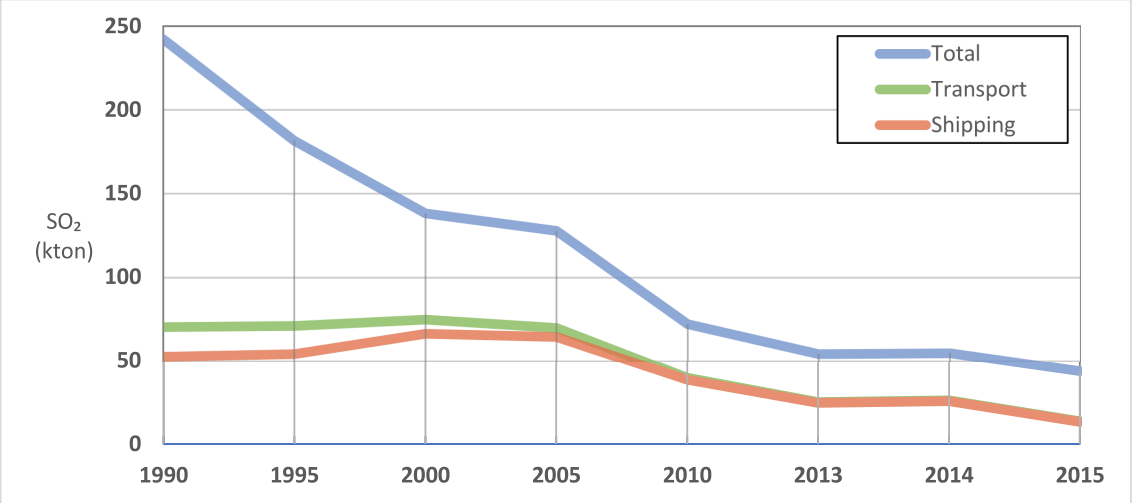

Figure 1.1: The emission of SO_2 by source over time. Data source: RIVM. [6]

Figure 1.2: SECA. Image courtesy of Dirk Jan Oostwoud Wijdenes and National Geographic [10].

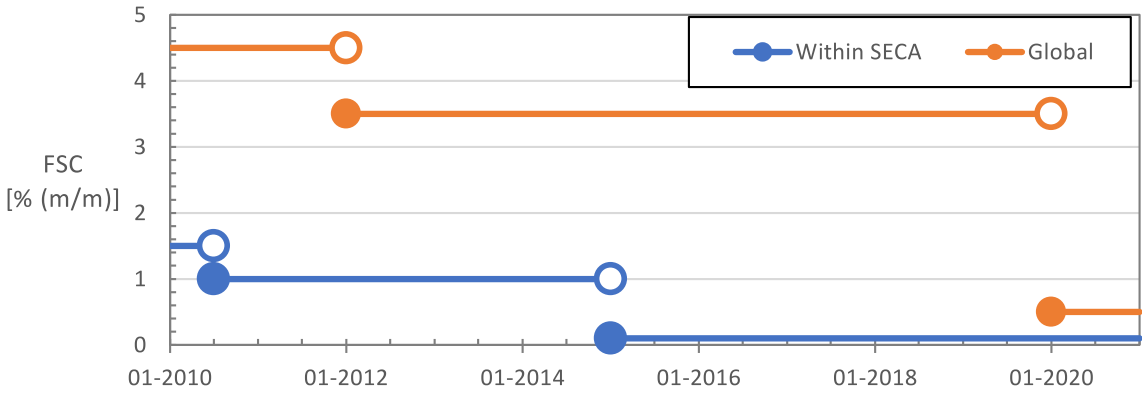

Figure 1.3: The development of regulation restricting maximum marine FSC.

1.3 Aim

A lot of time was already invested in the development of analytical techniques for compliance monitoring of sulfur dioxide. In many field campaigns these techniques were tested. Yet, a thorough review of all different instruments used by various organizations and an evaluation of all data produced by the campaigns is not available.
ILT has supported multiple initiatives in the development phase of remote sensing of marine FSC. An independent technology review is needed for future selection. Additionally, this research allows ILT to know what they should demand in further monitoring contracts.
Another major goal of the present research is the exploration of all measurements performed by ILT and other port state controls (PSCs). A large comparison on this scale was not performed earlier. The final aggregation of the campaigns and PSCs yield a total of more than 8000 measurements. A question regarding this data is, for example, what good decision criteria are, both for marking a ship non-compliant and for determining compliance rates in the sector. The former marking is to take enforcement action and the latter marking to calculate how much damage is caused by non-compliance ships, which could be used in determining how much employees should be put on sulfur regulation enforcement. Both classifications have their own assumptions and will result in different criteria.

1.4 Structure of this thesis

The review of all analytical instruments used to measure the data is given in Chapter 2. It contains also an introduction about how the theory how it is possible to measure remotely the sulfur contained within the bunker fuel. The third chapter will provide the methods necessary to understand the statistical tools applied in the Results chapter. This chapter will provide answers to questions like: how sure can we know that a ship is really non-compliant and how can we determine how many ships sail on non-compliant fuel? We will see that these questions are strongly related to classification problems, encountered more often in statistics. In the Discussions the pros and cons are given of the applied techniques. These considerations will lead to recommendations for further research. The Conclusion and Appendixes complete this thesis.

Chapter 2

Theory

This chapter will start with the concepts of measuring the FSC remotely. During the analysis of the data from the various (governmental) partners, it was clear that variation among measurement campaigns exists and are driven by several factors. One of the factors driving the variation among measurements was which operator conducted the measurements. In order to understand the differences between the data sources, it is valuable to explore which instruments are used in the sulfur compliance monitoring in Northern Europe.
A complete overview of all operators is shown in Table 2.1. All campaigns mentioned involved measuring ships where the emission could be resolved to a single ship. Most methods gave no information about the estimated uncertainty at 0.1 wt.%. Some methods were not intended to be used within the SECA. Other methods were not used before 2015, when the 0.1 wt.% limit came into force. These methods did report a uncertainty at other values for the FSC in some cases. However, it is difficult to convert a uncertainty to another value for the FSC. For example: TNO reported a relative standard deviation (RSD) of 21 % at 0.84 wt.% FSC. [13]. This cannot be converted to a RSD at 0.1 wt.%.
The following sections in this chapter will provide more information about how the instruments work. The optical instruments deduced the concentration of various compounds by looking how light that has traveled through the plume of a ship is modified. These instruments did not have to be located in the plume. This is in contrast with sniffing and sensing, where the apparatus had to be located in the plume. Tables were provided which summarize which instruments are used by which operator and what properties they have.

2.1 Measuring the FSC

The FSC is literally a measure of how much sulfur (S) is present in the fuel tank, regardless of the molecular structure. This can be determined by taking a fuel sample and analyze it according to ISO 8217:2017 using XRF. But the content of sulfur in fuel can also be measured remotely, by analyzing the plume. It is assumed that all sulfur in the fuel is converted to sulfur dioxide. This assumption is valid if the plume is sufficiently aged. If this is not the case, thus when a measurement of the plume is taken immediately after emission, SO_3, SO_4 and H_2SO_4 can also be expected apart from SO_2. Additionally Alföldy *et al.* reported conversion within the tank to H_2S. [14]
Furthermore 87 wt.% of the fuel corresponds to carbon, which is completely converted to

carbon dioxide. Using the correct molecular weights, equation 2.1 is obtained. [15]

$$\mathrm{FSC} = 0.232 \frac{\mathrm{SO_2}}{\mathrm{CO_2}} \tag{2.1}$$

Equation 2.1 is not influenced by the size of the ship, because bigger ships will both emit more sulfur dioxide and carbon dioxide. The ratio will remain also constant with higher wind speeds, since both gasses will be diluted equally. As long as both gasses will behave equally, the ratio applies. Since both molar masses are comparable, this will likely occur. More problematic are background sources which can cause heightened carbon dioxide and sulfur dioxide concentrations.

2.2 Optical instruments

The optical method is truly a remote method, since it is not necessary to position the equipment itself in any part of the exhaust plume. Optical systems are fast because there is no need to physically transport gasses through a system. However, soot in the plume can not be filtered. This soot can cause interference with the sulfur reading. Additionally, the optical method generally suffers more from uncertainties like wind and wave scattering. This is the reason that Chalmers uses the optical method only as a selection for further analysis with sniffing instruments.
There are three kinds of optical methods used in the remote determination of sulfur dioxide. Only DOAS and the UV-camera are actually in use since 2015. The other (LIDAR) deemed to be too inaccurate for the determination of sulfur dioxide and is not routinely used in northern Europe for IMO Annex VI compliance monitoring.
Note that calculating the FSC needs both the measurement of sulfur dioxide and carbon dioxide. The former can be measured by the optical methods, but none of the three methods is capable of measuring the carbon dioxide. Differential optical absorption spectroscopy (DOAS) is able to measure the nitrogen dioxide which can be used as an indication for the burned carbon dioxide. The other methods needs an estimation of or model for the burned fuel.
It is difficult to compare the precision in the sulfur dioxide measurements between the operators. But Prata reports typical emission values of $10\,\mathrm{g\,s^{-1}}$ to $100\,\mathrm{g\,s^{-1}}$, which results in a typical RSD of 5 % to 50 %. [16] A typical RSD for the measurements by Chalmers was 20 %. [17]
In the next three sections, the principles of the DOAS, ultra-violet (UV)-camera and light detection and ranging (LIDAR) will be explained. An overview of the optical instruments used by Bundesamt für Seeschifffahrt und Hydrographie (BSH)/ University of Bremen, Chalmers and Norwegian institute for air research (NILU) is given in Table 2.2.

Table 2.1: Operators able to monitor the FSC of ships.

Operator	Data Owner	Period	Area	Technique	N (ships) since 2015[1]	Platform	Information	Uncertainty @ 0.1 wt.% FSC
BSH	University of Bremen	2013 -	Hamburg	Optical Sniffing	> 1000	Fixed	[15,18–21,21–25]	15 % to 30 %[2]
Chalmers	Danish EPA MUMM Sweden	2008 -	Scandinavia USA Russia	Optical Sniffing	> 10000	Fixed Airborne	[14,18,26–36]	50 %[3]
DLR		2014	English Channel	Sniffing	0	Airborne	[37]	
Explicit		2014 -	Scandinavia	Sensing Sniffing		Airborne	[18,38]	13 %[3]
JRC		2009 -	Italy Netherlands	Sensing Sniffing		Fixed	[14,17,29]	
KINE	Trafi	2016 -	Finland	Sniffing	> 3000	Fixed	[18]	
Max Planck Institute Johannes Gutenberg University		2011 2012	Elbe	Sniffing	0	Van	[39,40]	
Metropolia University		2010 - 2011	Helsinki	Sniffing	0	Van	[41]	
NILU		2009	Rotterdam	Optical	0	Fixed Airborne	[16,29]	
NOAA		2002 2006 2010	USA	Sniffing	0	Airborne	[42–46]	
University of California		2007	USA	Sniffing	0		[47]	
University of Washington		1994 2000	USA Namibia	Sniffing	0	Airborne	[48,49]	
RIVM		2009	Rotterdam	Optical	0	Van	[14,29,50–52]	
TNO	ILT	2009 -	Rotterdam	Sniffing	> 2000	Fixed	[13,18,29,53]	

[1] Determined from measurements mentioned in publications. This number is thus a lower limit.
[2] All uncertainties were added up with the root of sum of squares method.
[3] Compared to a ship with a known FSC.

Table 2.2: Optical instrumentation used in the remote determination of the FSC of ships.

Operator	BSH/ University of Bremen	Chalmers	NILU
Technique	DOAS	DOAS	UV-camera
Spectrographs	2 Andor Shamrock SR-303i [54]	Andor Shamrock SR-303i [54] Andor Newton DU290N [55]	Hamamatsu
Detector	Princeton 1340/400-EMB CCD (UV) Andor iDus DV420-BU (Visible)		
Measurant	SO_2, NO_2	SO_2, NO_2	SO_2
t_{90}		1 s	< 10 ms
σ		20 ppb over 50 m	$5\,\mathrm{g\,s^{-1}}$
LOD	10 ppb over 50 m	1 ppm over 10 m	
QE			30 % at 300 nm
λ		294 nm to 324 nm (> 325 nm blocked)	200 nm to 600 nm ((308 ± 5) nm filter)
Range			5 km
Uncertainty		30 % to 45 %[1]	
Remarks	Needs active cooling	Needs active cooling	Limited accuracy

[1] All uncertainties were added up with the root of sum of squares method.

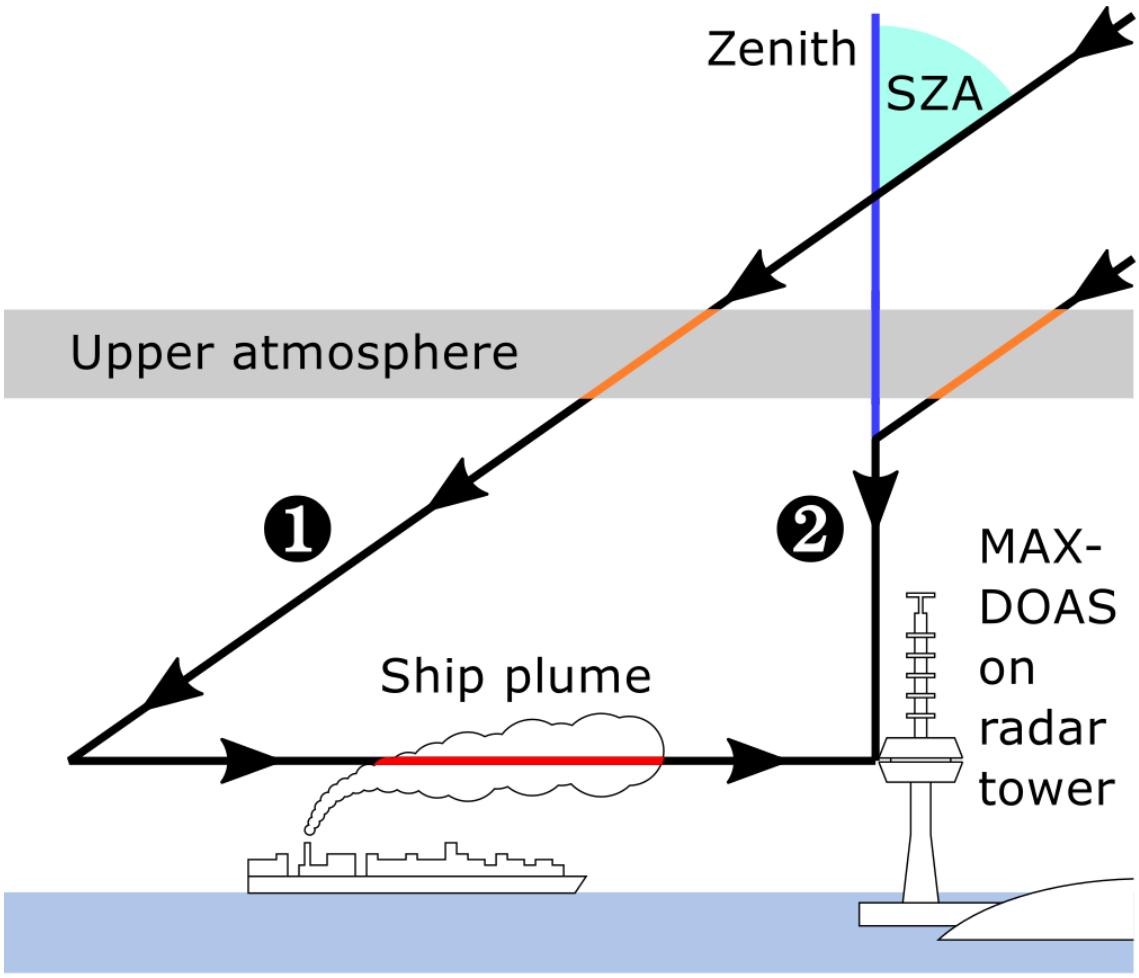

Figure 2.1: The multi-axis (MAX)-DOAS set-up used by BSH/ University of Bremen. Taken from [25].

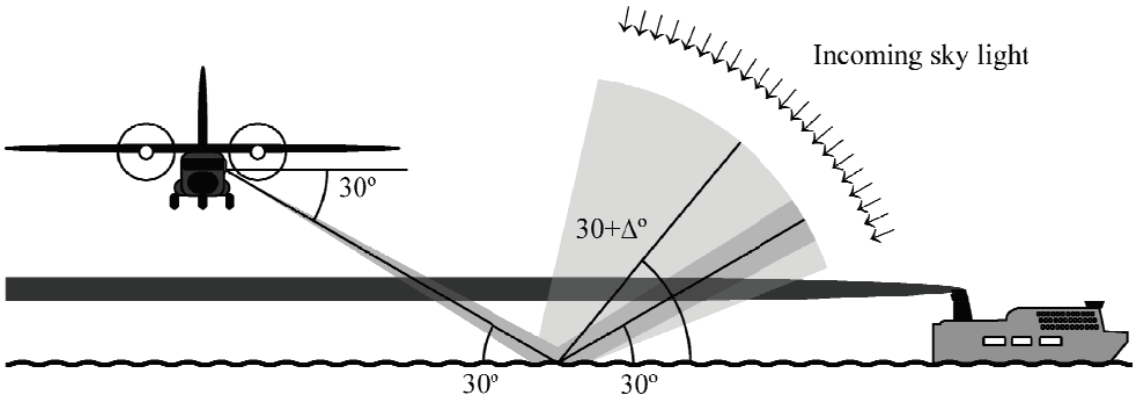

Figure 2.2: The DOAS set-up used by Chalmers. Taken from [28].

2.2.1 DOAS

DOAS use the difference in light spectra. The intensity as function of wavelength with and without the compounds of interest is measured. It is possible to use the sun as a light source, since only the differences in two light paths are measured. It obviously enables the measurement only at daylight. The molecule of interest will absorb partially the light in some specific wavelengths. The analysis is complicated by other trace gases, Rayleigh, Mie and Raman scattering and aerosol particles. These are corrected by measuring the spectra of all expected trace gases beforehand and subtract all slow varying functions $\sigma_i(\lambda)$. The specie of interest can than be fitted in the residual spectrum. Using the Lambert-Beer-law, the concentration can be calculated. A very clear explanation is found in reference [25].
Both BSH/ University of Bremen and Chalmers use the DOAS technique. The former uses a fixed position in Wedel, Germany. They use MAX-DOAS, with a set-up schematically shown in Figure 2.1. MAX-DOAS uses the difference between a light path traversing through the ship plume and a zenith sky reference. In this way the interference from sulfur dioxide and especially nitrogen dioxide in the upper atmosphere is prevented. [25]
Chalmers uses DOAS operated from a plane. It uses light reflected by waves, which transected the plume two times, see Figure 2.2. The waves scatter the light in various different directions, which cause an error term. [28]
Both BSH/ University of Bremen and Chalmers uses two spectrographs; one for measuring nitrogen dioxide and one for measuring sulfur dioxide. When the ratio SO_2/NO_2 is higher than 1, it is an indication that the ship is sailing with non-compliant fuel. Only Chalmers use this information to measure a ship more thoroughly with other sensors, to determine the FSC more accurately. [29]

2.2.2 UV-camera

The UV-camera combines the benefits of a spectrometer (with a high wavelength resolution and range) and the benefits of a camera (high spatial resolution using various pixels). The column density of sulfur dioxide can be calculated with an algorithm for each pixel and finally a total emission rate can be calculated, as shown in Figure 2.3. The UV-camera consumes less power than DOAS and LIDAR, since there is no active cooling needed. The camera can also be much faster than the other techniques. The camera can capture easily at 100 Hz. [16]
Prata (from the NILU) explored the technique but conclude that the technique can not reliable quantify sulfur dioxide emissions, mostly due to the presence of soot. [16] Particles in the plume are very efficient scatters, which makes it difficult to quantify accurate how much light is absorbed. [30] It is unknown whether the technique is actually applied now.

2.2.3 LIDAR

The LIDAR has been explored by the Dutch Dutch institute for public health and the environment (RIVM). It was operated from a truck. [51] The set-up consists of a laser beam and a spectrometer. The laser beam shoots two light pulses with a different wavelength (299.752 nm and 300.094 nm). From both pulses the reflection upon the molecules of interest is measured and timed. By directing the laser beam sequentially, a full 3D resolved sulfur dioxide emission pattern can be retrieved. A drawing of the set-up is shown in Figure 2.4.

The pilot was stopped because it was not possible to measure accurate emissions on low FSC and because only under the right wind, light and visibility conditions measurements were possible. Measurements during rain were also not possible. [52] Additionally, it was only possible to measure ships in proximity of land and not on full sea. Another drawback is that large errors are introduced by the need to know an accurate vertical wind profile, which can often not exactly be determined at the location of the ships plume that is being measured. [29]

2.3 Sensing instruments

Both sensing and sniffing instruments need to be located within the plume. The former are cheaper than the latter, because sensors make use of electronic resistance on a chip instead of more complicated techniques. The power consumption and low weight makes the sniffing instruments also more suitable for use in a unmanned aerial vehicles. However, the sensing instruments are described sparsely in scientific literature. Probably this is because the sniffing instruments are more accurate and precise.

2.3.1 Explicit

The company Explicit have patents pending on their technology. To the authors knowledge, Explicit is the only one that operates the sensing instruments. Their patents do not include detailed information about the instruments used. However, it is clear that several sensors are placed in a chamber inside the UAV. Figure 2.5 was taken from a patent filled by Knudsen. The air inlet can be shut by a relay, such that the chamber is a controlled, steady state environment. The UAV will allow to remain for long time in the plume of a ship, where the highest sulfur dioxide and carbon dioxide concentrations are present. [56,57] These conditions probably allow Explicit to make use of sensing instruments while having the same or better results than the sniffing instruments.

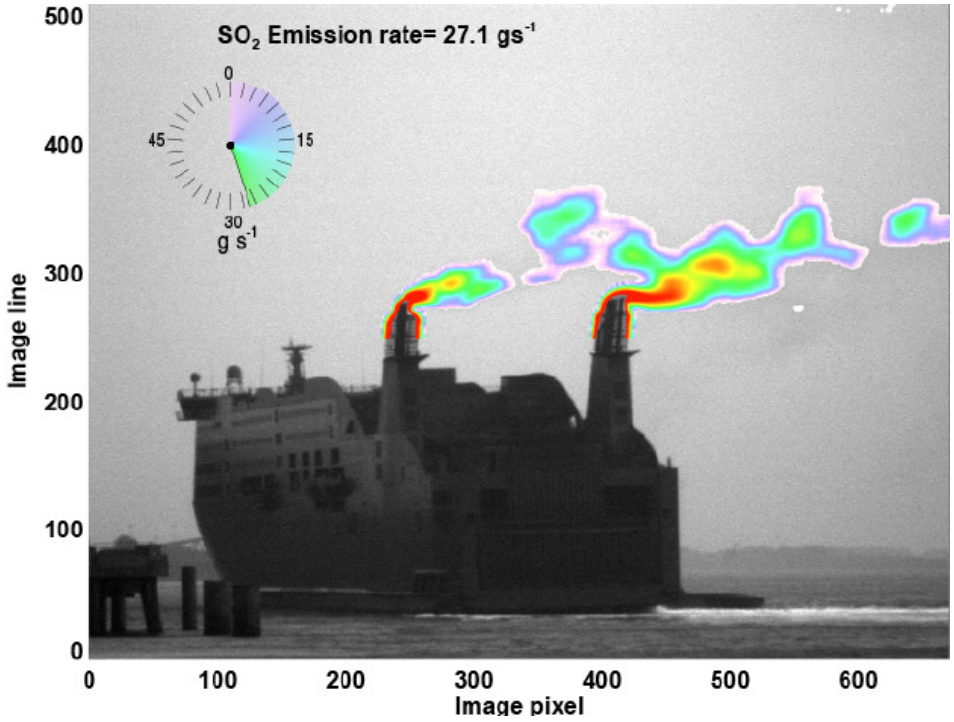

Figure 2.3: A photograph from the UV-camera with software for sulfur dioxide retrieval. Taken from [16].

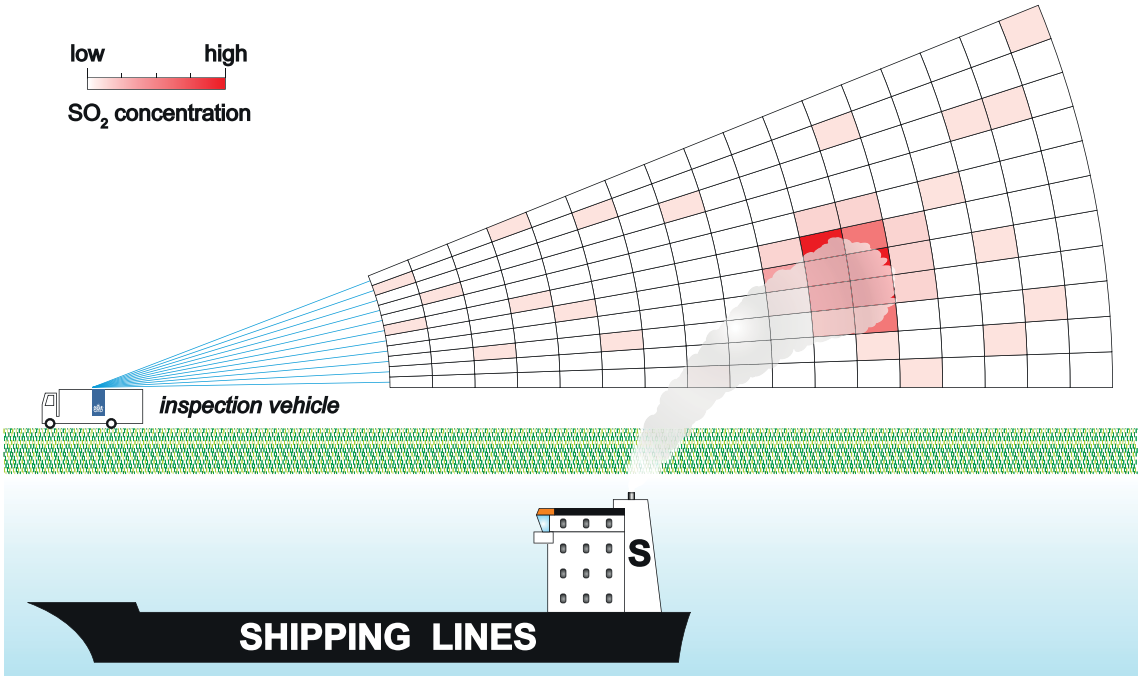

Figure 2.4: The set-up used for LIDAR by the RIVM. Taken from [51].

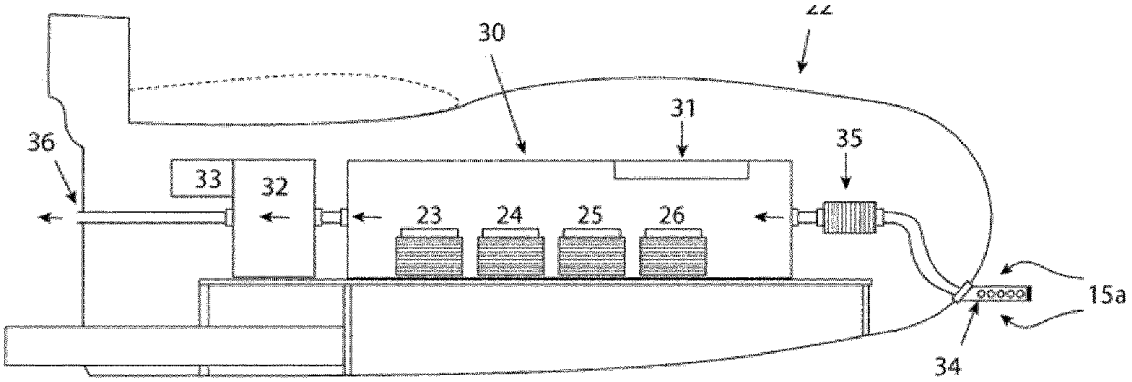

Figure 2.5: A UAV with sniffing instruments, used by Explicit. Taken from [57].

2.4 Sniffing instruments

Sniffing is the most used technique within the remote determination of the FSC. Common sniffing instruments sample the plume with help of air pumps and subsequently analyze it with techniques like fluorescence or infrared. It is possible to use filters to prevent soot interfering with the analysis, which result in repeatable and more accurate measurements. A major drawback of the sniffing instruments is that they need active pumping, which needs space and power.
Calibration can be performed by standard addition or standard replacement. Both zero air and background air can be used in these calibrations. Zero air can be produced in-situ or can be bought from suppliers.
The sniffing instruments can be used on all kind of platforms, except those having very limited space. The different platforms are shown in Table 2.1. For example, the ILT uses a fixed station near a waterway shown in Figure 2.6.
There are different techniques needed for the different species. The following sections will provide information about the measurements of sulfur dioxide, nitrogen oxides and carbon dioxide.

2.4.1 Sulfur dioxide

Sulfur dioxide can be measured with fluorescence. The sulfur dioxide molecule is excited by a laser pulse ($\lambda = 254\,\text{nm}$), which decays from an excited state to a ground state within a certain lifetime. By this decay a light pulse is emitted with a different wavelength ($\lambda = 330\,\text{nm}$) than the laser wavelength. The intensity of the light pulse is proportional to the amount of SO_2 molecules. However, VOCs and polycyclic aromatic hydrocarbons (PAHs) can interfere with the analysis. A hydrocarbon kicker removes these molecules by forced permeation through a tube wall. [58,59] An extensive manual about how the analysis should be carried out, can be find in European standard 14212 - Ambient air quality - Standard method for the measurement of the concentration of sulfur dioxide by ultraviolet fluorescence (EN 14212).
However, there are different apparatus on the market and several operators have added some own modifications. Some operators heat their tubes to prevent sticking of sulfur dioxide molecules to metal surfaces. [18] Secondly, the hydrocarbon kicker can be removed to increase sample flow, which decreases the limit of detection (LOD) and the response

Figure 2.6: Sniffing set-up used by ILT and operated by TNO, which is located by Rotterdam. Taken with permission from [38].

time. Orifices can be broadened to increase the sample flow, to decrease the LOD and the response time. [14] A complete overview of all sniffing instrumentation is given in Table 2.3.

2.4.2 Nitrogen oxide

Nitrogen oxides can cause environmental damage by contributing to acid rain and by enhancing surface ozone formation. The latter affects greenhouse warming. [60] Unlike sulfur dioxide, the concentration of nitrogen oxides is almost independent of the fuel used. The production is dependent on the type of motor used and especially on the peak temperatures of the combustion system. [61]
This makes nitrogen oxides interesting molecules to study. But the molecules interfere also with the sulfur dioxide measurement, just like the VOCs and PAHs. Measuring nitrogen oxides (NOx) allows to correct for the cross interference and hence the amount of sulfur dioxide can be measured more accurately. BSH/ University of Bremen corrects the SO_2 levels for 0.8 % of the NO reading. The reading is obtained from the Airpointer/ Horiba APNA-370. BSH/ University of Bremen has certified the measurement according to European standard 14211 - Ambient air quality - Standard method for the measurement of the concentration of nitrogen dioxide and nitrogen monoxide by chemiluminescence (EN 14211). [15]
The Horiba APNA-370 uses chemiluminescence to measure all nitrogen oxides. All NO is completely oxidized, whereby light ($h\upsilon$) is produced: $NO + O_3 \longrightarrow NO_2 + O_2 + h\nu$. This light intensity is proportional to the amount NO converted. In order to measure NO_x, all NO_x needs to be reduced to NO using platinum catalysts.

2.4.3 Carbon dioxide

There are two common techniques which can be used to measure the amount of carbon dioxide in a plume, as can be seen in Table 2.4. Cavity ring-down spectroscopy (CRDS) is used by Chalmers. It compromises a laser beam which is reflected multiple times in a cavity. According to the manual, a light path of over 20 km can be made. The decay constant τ is measured, which is inversely proportional to the concentration of carbon dioxide. The advantage of measuring the decay instead of the intensity, is that it does not suffer from differences in pulses. This result in much more stability and less need for calibrations. Another advantage is the low standard deviation, which can be as low as 200 ppb.
Another technique is non-dispersive infrared (NDIR). It uses the absorption of carbon dioxide in infrared light ($\lambda = \pm 2\,\mu m, 4\,\mu m, 15\,\mu m$) by application of the Lambert-Beer law. The main advantage of this technique is that it is possible to miniaturize this set-up to 23 cm × 16 cm × 8 cm.

2.5 Concentration methods

When sampling from an airplane, the sample time can be short. A remarkable solution to this problem is applied by the University of Washington. They used a Velostat grab-bag which consisted of 2.5 m^3 electrically conducting plastic, that could be filled in 12 s. Another solution is using a stronger pump, used by national oceanic and atmospheric administration (NOAA).

Table 2.3: Sniffing instrumentation used in the remote determination of sulfur dioxide emissions.

Operator	Model	Flow	t_{90}	σ	LOD	Pressure	Uncertainty	Corrected
BSH	Airpointer/ Horiba APSA-370 [62]	$0.7\,\mathrm{L\,min^{-1}}$	$< 2\,\mathrm{min}$	1 %	0.25 ppb			0.8 % NO
Chalmers	Thermo 43i-TLE 55[1]	$(5 \text{ to } 6)\,\mathrm{L\,min^{-1}}$	2 s	5 ppb	1 ppb	80 hPa	21 %[3]	1.5 % NO
DLR	Ion Trap MS		5 s				5 %	
Metropolia University	Thermo 43C-TL [63]	$< 1\,\mathrm{L\,min^{-1}}$	80 s	< 1 ppb	< 2 ppb		1 ppb/1 %	
JRC	Thermo 43C-TL [63][2]	$1.5\,\mathrm{L\,min^{-1}}$	15 s	< 1 ppb	< 2 ppb		1 ppb/1 %	
KINE			40 s					
NOAA	Thermo 43s			$0.35\,\mathrm{mL\,L^{-1}}$			10 %	
University of California	Thermo 43s			$0.35\,\mathrm{mL\,L^{-1}}$				
University of Washington	Thermo 43s			$0.2\,\mathrm{mL\,L^{-1}}$				
TNO	Teledyne-API 100E [58]	$0.65\,\mathrm{L\,min^{-1}}$	< 100 s	< 0.2 ppb	0.4 ppb	101 kPa		
Max Planck Institute	Airpointer [64]			(4 to 20) s	0.25 ppb	0.5 ppb		

[1] The hydrocarbon kicker was removed.
[2] The orifice was enlarged.
[3] All uncertainties were added up with the root of sum of squares method.

Table 2.4: Sniffing instrumentation used in the remote determination of carbon dioxide emissions.

Operator	Model	Technique	Flow	t_{90}	σ	LOD	Uncertainty	Pressure
Chalmers	Picaro G-2301M [65]	CRDS		0.5 s	< 100 ppb	200 ppb		
	LI-COR 7200 [66]	NDIR	(5 to 6) $\mathrm{L\,min^{-1}}$	1 s	200 ppb	200 ppb	7 %[1]	80 hPa
Max Planck Institute	LICOR 840A [67]	NDIR	$1\,\mathrm{L\,min^{-1}}$	1 s		1 ppm		
BSH/ University of Bremen	Airpointer/ LICOR 840A[67]	NDIR	$< 1\,\mathrm{L\,min^{-1}}$	1 s		200 ppb		
DLR	LICOR 6262 [68]	NDIR	$< 10\,\mathrm{L\,min^{-1}}$	1 s	$300\,\mathrm{nL\,L^{-1}}$	$300\,\mathrm{nL\,min^{-1}}$	$800\,\mathrm{nL\,L^{-1}}$	< 17 kPa
University of Washington	LICOR 6262 [68]	NDIR	$< 10\,\mathrm{L\,min^{-1}}$	1 s	$300\,\mathrm{nL\,L^{-1}}$	$300\,\mathrm{nL\,min^{-1}}$	$800\,\mathrm{nL\,L^{-1}}$	< 17 kPa
KINE	LICOR 820 [69]	NDIR	$< 1\,\mathrm{L\,min^{-1}}$	1 s		200 ppb		
JRC	LI-COR 7200 [66]	NDIR	$6\,\mathrm{L\,min^{-1}}$	1 s	70 ppb	200 ppb		< 200 kPa
TNO	LI-COR 7200 [66]	NDIR	$6\,\mathrm{L\,min^{-1}}$	1 s	70 ppb	200 ppb		< 200 kPa
NOAA	LI-COR 7200 [66]	NDIR	$6\,\mathrm{L\,min^{-1}}$	1 s	70 ppb	200 ppb		< 200 kPa
Metropolia University	Horiba VA 3100				3 %	2 ppm		

[1] All uncertainties were added up with the root of sum of squares method.

Chapter 3

Method

This chapter will explain the various statistics and software used in this research. It will provide the mathematical framework, which is subsequently used in the Results and Discussion. In section 3.2 an explanation is given of how classification can be performed with the use of the z-score. In the context of this z-score, type I and type II errors are explained and how these depend on each other. In order to perform power analysis, validation data is needed. A short elaboration on how this data can be obtained is given in Section 3.3. Finally, the expectation-maximization algorithm is explained, which allows to perform some power analysis without the need for training data. All the concepts explained in this chapter, are necessary to understand the results.

3.1 Software and data management

Partners of ILT were requested to share their data for research purposes. The data was provided in comma separated value (CSV) or Excel format. In `SAS` Software all data was aggregated. `SAS` Software [70] was chosen because this was the only relevant software package provided from ILT. A master table was created were each row was one measurement and all its columns were parameters. The data was made anonymous to allow working outside the company network with `R` [71]. An overview of the variety of parameters and observations can be found in Appendix 1.

3.2 Classification

3.2.1 Z-score

The normal distribution is a simple model for classification. In this model, each data point has its own normal distribution. This distribution is defined by the reported FSC and the standard deviation. It is a simple model but likely not suitable, since it has several assumptions. The model was however used to gain further understanding in the classification problem. In order to perform classification, two hypothesis were defined:

$$\begin{aligned} H_0 &: \text{The ship has a FSC of } 0.1\,\text{wt.\% or less} \\ H_1 &: \text{The ship has a higher FSC than } 0.1\,\text{wt.\%} \end{aligned} \tag{3.1}$$

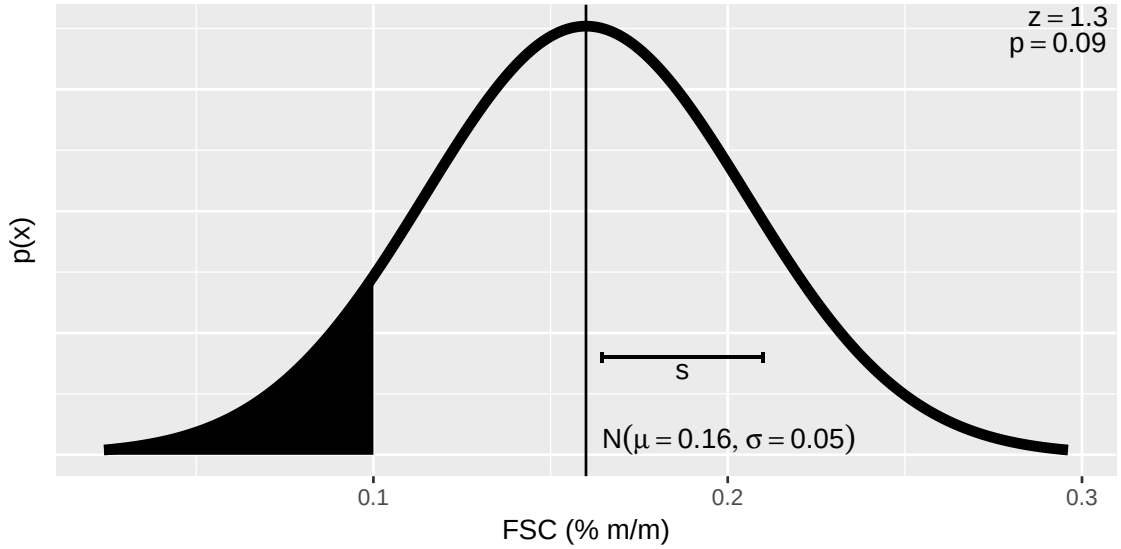

Figure 3.1: An example of a probability distribution of a measurement. The fraction which is filled under the curve is equal to the probability of obtaining this data or more extreme, given that H_0 is true.

The FSC is the value which was obtained in the case the measurement had no bias and if one could observed it infinity times. If both the $\bar{x}$ and s_x was known (where x is the measured FSC), a Z-score could be calculated, according to Equation 3.2. This equation was equal to the z-score in the case of $n = 1$. By using the t-test, it is assumed that x follows a normal distribution. This was certainly not the case, as multiple measurements would then have a significant probability of having negative FSC. This could not be the case, since a negative concentration of sulfur dioxide or carbon dioxide is impossible. See Figure 5.4 for a measurement with significant probability of having a negative FSC. However, by absence of other methods for classification, the t-test was used.

$$z = \frac{\bar{x} - \mu_0}{s_x/\sqrt{n}} \tag{3.2}$$

From the z-score, a probability can be calculated. The p-value is the probability that if H_0 is true, one will observe this data or more extreme. More extreme implies that the $\bar{x}$ would be more far off from $\mu_0 = 0.1$ and thus would be higher. If this p-value is sufficiently low (below α), H_0 should be rejected and thus H_1 is accepted. A measurement which was likely from a non-compliant ship is shown in Figure 3.1. It had only very little probability that this data (or more extreme) was obtained, given that H_0 is true.
It was assumed that the measurement had a normal probability density function with a given $\mu = \bar{x}$ and $\sigma = s_x$. The probability of being compliant is then given by the fraction of the function with FSC < 0.1.

3.2.2 Type I and II errors

The probability of rejecting H_0 while H_0 is true, also called the type I error, is equal to α. The other error, type II or β, is the probability that falsely H_0 is accepted (or better; not rejected) while H_1 is true. There is a trade-off between these errors; a lower threshold α

will result in a larger β. β is also depended on the effect size ($d = \frac{\mu_1 - \mu_0}{\sigma}$) which says how different the populations of all measurements given H_0 and H_1 are. β can be calculated using equation 3.3.

$$\begin{aligned} \beta \ &= p(X = x_{\text{crit}}) \text{ from } X \sim \mathcal{N}\left(\mu_{nc},\ \sigma_{nc}^2\right) \\ &\quad \text{where } \mu_{nc} \text{ and } \sigma_{nc} \text{ are true mean and standard deviation} \\ &\quad \text{from the population of FSC of non-compliant ships.} \end{aligned} \tag{3.3}$$

The positive classification rates are called specifity $(1 - \alpha)$ and power $(1 - \beta)$. The accuracy is the fraction of all data which is correctly classified. It can be calculated using equation 3.4.

$$\begin{aligned} \text{accuracy} \ &= \frac{n_c\ (1 - \alpha) + n_{nc}\ (1 - \beta)}{n_c + n_{nc}} \\ &\quad \text{where } n_c \text{ is the amount of compliant classified observations} \\ &\quad \text{and } n_{nc} \text{ is the amount of non-compliant classified observations} \end{aligned} \tag{3.4}$$

3.2.3 Critical value

A critical z-value can be calculated, which is the boundary between the two classes. Values $z_{\text{calculated}} > z_{\text{critical}}$ will be rejected as compliant, whereas all other values will not have sufficient prove to be rejected as compliant. The formula for the critical z-value is given in equation 3.5. The critical z-value is 1.645 for a one-tailed test with $\alpha = 0.05$.

$$z_{\text{critical}} = \Phi^{-1}(1 - \alpha) \text{ where } \Phi^{-1} \text{ is the inverse standard normal distribution} \tag{3.5}$$

Combining equations 3.3 and 3.5 gives the relation between β and α, which is given in equation 3.6. The effect on varying α and β is depended on the μ_{nc} and σ_{nc}.

$$\beta = p\left(X = \Phi^{-1}(1 - \alpha)\right) \text{ where } \Phi^{-1} \text{ is the inverse of } \mathcal{N}\left(\mu_{nc}, \sigma_{nc}\right). \tag{3.6}$$

3.3 Validation data

The objective is to differentiate between an observation from a ship sailing under compliant and non-compliant fuel. Training data should be data from which we know for sure to which group it belongs. In this case this data was obtained by ILT. The samples were taken from marine fuel, which were subsequently analyzed by either XRF or lab analysis. The allowed relative standard deviation of the laboratory used by ILT was only 5 %.[1] The cross reference was defined as a measurement on the same ship within two days.

3.4 Expectation-maximization algorithm

In order to be able to estimate the distributions of the FSC of the compliant and non-compliant ships, without the need of validated data, unsupervised learning was used. In the first stage, a model consisting of two normal distributions was used. It was assumed

[1] According to personal conversation with Jan-Willem Bil, inspector sampling team (ILT)

that the distribution separation explains the separation between the compliant and non-compliant ships. A more advanced assumption is a lognormal distribution. It was used because there are a lot of non-compliant ships just over the legal limit, which caused a long tail on the data. The lognormal could fit this tail.
A more thorough derivation is given in Appendix B.

3.4.1 Unimodal distributions

Maximum likelihood estimators (MLEs) for unimodal distributions are found by finding the parameters which maximize the (log-)likelihood. For a normal distribution ($\mathcal{N}(\mu, \sigma^2)$), the likelihood is given in equation 3.7 for the data $\boldsymbol{x}$. It is well known that maximizing this likelihood with respect to μ and σ, gives $\hat{\mu} = \bar{x} \equiv \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\hat{\sigma}^2 = \frac{1}{n}(\boldsymbol{x} - \bar{x})^T(\boldsymbol{x} - \bar{x})$.

$$p\left(\boldsymbol{X}|\mu, \sigma^2\right) = \prod_{i=1}^{N}\left[\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)\right] \tag{3.7}$$

3.4.2 Bimodal normal distribution

In order to model both classes with a normal distribution, a bimodal model was created. This model has five parameters (μ_1, σ_1, μ_2, σ_2 and $\boldsymbol{\pi}$). These parameters will provide information about what fraction is non-compliant and the variation between compliant and non-compliant groups. An example is given in Figure 3.2.
The log-likelihood of a bimodal normal distribution is given in Equation 3.8. Here $\boldsymbol{\pi}$ ($\{\pi_1, \pi_2\}$) are the prior probabilities to belong to either class 1 or 2 ($\pi_k = p(x \in \mathcal{C}_k)$) and thus $\sum_{k=1}^{2}\pi_k = 1$. Taking the derivative of the parameters is difficult. This will result in $\hat{\mu_1}$ depending on $\hat{\mu_2}$ and *vice versa*. The same is true for σ_1 and σ_2. This problem can be solved by creating responsibilities. This responsibility is the posterior probability (expectation value) of a point belonging to class k given the observation of the data. The probability is calculated using Bayes' theorem for each point and class, see equation 3.9. In the E-step, each point is softly assigned using the responsibilities. Then the likelihood is calculated for each class and point separately, see equation 3.10.
In the M-step, MLEs are calculated for each parameters. There is no problem anymore in calculating these, since there is only one term in the logarithm in equation 3.10. The MLEs are given in equation 3.11. With the updated parameters, the E-step can be performed again to calculate the updated responsibilities. These E- and M-steps should be iterated until convergence.

$$\log\ p\Big(\boldsymbol{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2\Big) = \sum_{i=1}^{N}\log\left[\sum_{k=1}^{2}\pi_k\,\mathcal{N}\Big(x_i|\mu_k, \sigma_k^2\Big)\right] \tag{3.8}$$

$$\begin{aligned}\gamma_{i,k} &\equiv p(x_i \in \mathcal{C}_k|\boldsymbol{x}) \\ \hat{\gamma_{i,k}} &= \frac{\overbrace{\pi_k}^{\text{prior}}\ \overbrace{\mathcal{N}\Big(x_i|\hat{\mu_k}, \hat{\sigma_k}^2\Big)}^{\text{likelihood ratio}}}{\underbrace{\pi_1\,\mathcal{N}\Big(x_i|\hat{\mu_1}, \hat{\sigma_1}^2\Big) + \pi_2\,\mathcal{N}\Big(x_i|\hat{\mu_2}, \hat{\sigma_2}^2\Big)}_{\text{evidence}}}\end{aligned} \tag{3.9}$$

$$\log\ p\Big(\boldsymbol{X}|\boldsymbol{\Gamma},\boldsymbol{\mu},\boldsymbol{\sigma}^2\Big) = \sum_{i=1}^{N}\sum_{k=1}^{2}\gamma_{i,k}\ \log\Big[\mathcal{N}\Big(x_i|\mu_k,\sigma_k^2\Big)\Big] \tag{3.10}$$

$$\begin{aligned}
\hat{\mu_1} &= \frac{\sum_{i=1}^{N}\gamma_{i,1}\ x_i}{\sum_{i=1}^{N}\gamma_{i,1}} \qquad & \hat{\sigma_1}^2 &= \frac{\sum_{i=1}^{N}\gamma_{i,1}\ (x_i-\hat{\mu_1})}{\sum_{i=1}^{N}\gamma_{i,1}} & & \\
\hat{\mu_2} &= \frac{\sum_{i=1}^{N}\gamma_{i,2}\ x_i}{\sum_{i=1}^{N}\gamma_{i,2}} \qquad & \hat{\sigma_2}^2 &= \frac{\sum_{i=1}^{N}\gamma_{i,2}\ (x_i-\hat{\mu_2})}{\sum_{i=1}^{N}\gamma_{i,2}} & \hat{\pi_k} &= \frac{\sum_{i=1}^{N}\hat{\gamma_{i,k}}}{N}
\end{aligned} \tag{3.11}$$

3.4.3 Bimodal distribution; normal and lognormal mode

The expectation-maximization (EM)-algorithm can be extended to a distribution with one normal mode and one lognormal mode. The lognormal mode was used because there was a right tail present just above the 0.1 wt.%. Other distributions could have been chosen. However, a lognormal mode allowed the incorporation in the EM-algorithm. Additionally, it provide a lower bound. This is logical for the ships sailing on non-compliant fuel.
The likelihood function of such a distribution is given in Equation 3.12. Using this likelihood, the responsibilities and MLEs can be calculated again. In contrast to a normal distribution with support $x \in (-\infty,\infty)$, the support of a lognormal distribution is $x \in (0,\infty)$. This means that either $x_i \leq 0$ should not occur for all i or the responsibility $\gamma_{i,2}$ should be 0 for each i where $x_i \leq 0$.

$$\log\ p\Big(\boldsymbol{X}|\boldsymbol{\Gamma},\boldsymbol{\mu},\boldsymbol{\sigma}^2\Big) = \sum_{i=1}^{N}\Big\{\gamma_{i,1}\log\mathcal{N}\left(x_i|\mu_1,\sigma_1^2\right) + \gamma_{i,2}\log\text{Lognormal}\left(x_i|\mu_2,\sigma_2^2\right)\Big\} \tag{3.12}$$

3.4.4 Bimodal distribution; normal and shifted lognormal mode

Data from non-compliant ships ($x_i \in \mathcal{C}_2$) have FSC > 0.1. The lognormal mode was shifted with 0.1, such that the support is $\boldsymbol{x} \in (0.1,\infty)$. An example of such a model is shown in Figure 3.3. The figure is much like Figure 3.2, but the non-compliant group is now modeled with a shifted log-normal mode.
The likelihood of the bimodal distribution is given in Equation 3.13. The responsibilities derived from the likelihood and the MLEs are given in Equation 3.14 and 3.15 respectively.

$$\log\ p\Big(\boldsymbol{X}|\boldsymbol{\Gamma},\boldsymbol{\mu},\boldsymbol{\sigma}^2\Big) = \sum_{i=1}^{N}\Big\{\gamma_{i,1}\log\Big[\mathcal{N}\left(x_i|\mu_1,\sigma_1^2\right)\Big] + \gamma_{i,2}\log\Big[\text{Lognormal}\left(x_i-0.1|\mu_2,\sigma_2^2\right)\Big]\Big\} \tag{3.13}$$

$$\begin{cases}
\hat{\gamma_{i,1}} &= \frac{\pi_1\ \mathcal{N}\left(x_i|\hat{\mu_1},\hat{\sigma_1}^2\right)}{\pi_1\ \mathcal{N}\left(x_i|\hat{\mu_1},\hat{\sigma_1}^2\right)+\pi_2\ \text{Lognormal}\left(x_i-0.1|\hat{\mu_2},\hat{\sigma_2}^2\right)} \text{ for } x_i > 0.1 \\
\hat{\gamma_{i,1}} &= 1 \text{ for } x_i \leq 0.1
\end{cases}$$
$$\begin{cases}
\hat{\gamma_{i,2}} &= \frac{\pi_2\ \text{Lognormal}\left(x_i-0.1|\hat{\mu_2},\hat{\sigma_2}^2\right)}{\pi_1\ \mathcal{N}\left(x_i|\hat{\mu_1},\hat{\sigma_1}^2\right)+\pi_2\ \text{Lognormal}\left(x_i-0.1|\hat{\mu_2},\hat{\sigma_2}^2\right)} \text{ for } x_i > 0.1 \\
\hat{\gamma_{i,2}} &= 0 \text{ for } x_i \leq 0.1
\end{cases} \tag{3.14}$$

$$\begin{aligned}
\hat{\mu_1} &= \frac{\sum_{i=1}^{N}\gamma_{i,1}\ x_i}{\sum_{i=1}^{N}\gamma_{i,1}} & \hat{\sigma_1}^2 &= \frac{\sum_{i=1}^{N}\gamma_{i,1}\ (x_i-\hat{\mu_1})}{\sum_{i=1}^{N}\gamma_{i,1}} & & \\
\hat{\mu_2} &= \frac{\sum_{i=1}^{N}\gamma_{i,2}\ \log(x_i-0.1)}{\sum_{i=1}^{N}\gamma_{i,2}} & \hat{\sigma_2}^2 &= \frac{\sum_{i=1}^{N}\gamma_{i,2}\ [\log(x_i-0.1)-\hat{\mu_2}]}{\sum_{i=1}^{N}\gamma_{i,2}} & \hat{\pi_k} &= \frac{\sum_{i=1}^{N}\hat{\gamma_{i,k}}}{N}
\end{aligned} \tag{3.15}$$

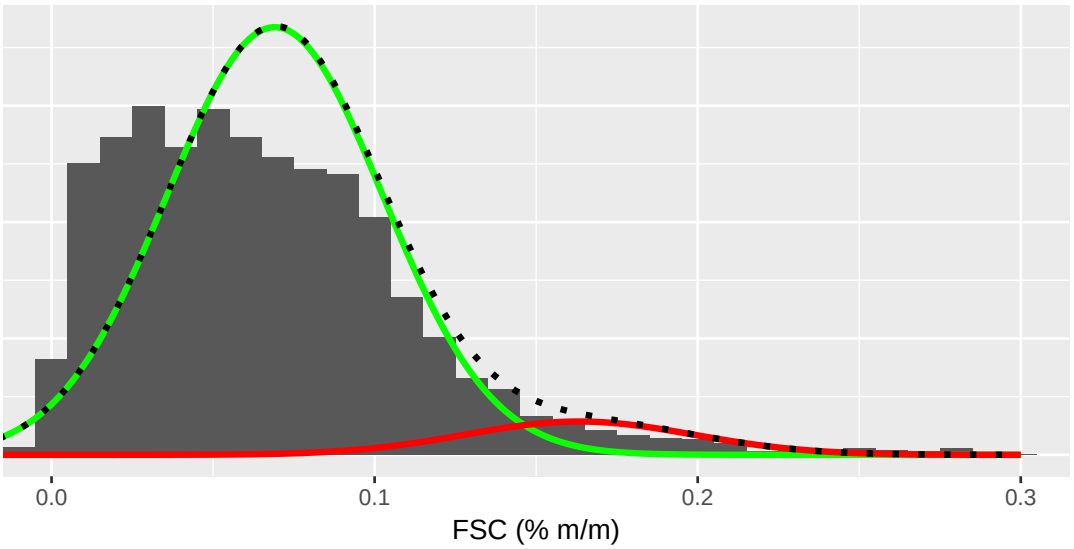

Figure 3.2: An example of a bimodal normal distribution. The two distributions (green and red line) can model the two classes of data, while the total distribution (dotted line) should fit the grey data. Data source: TNO

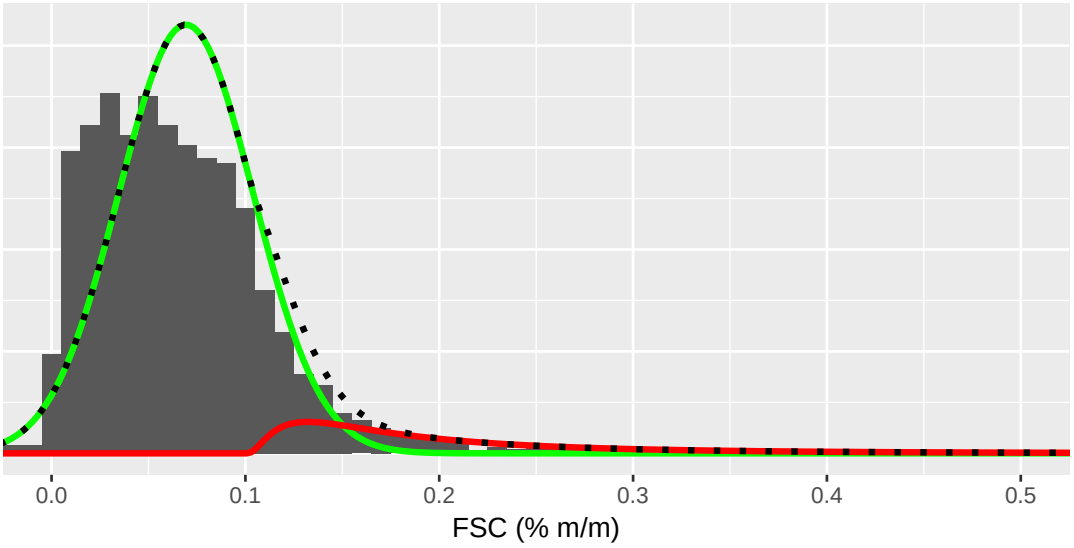

Figure 3.3: An example of a bimodal normal - shifted log-normal distribution. The two distributions (green and red line) can model the two classes of data, while the total distribution (dotted line) should fit the grey data. Data source: TNO

Chapter 4

Results

This chapter will provide the results obtained during this research by using the measurements from (partners of) ILT. This chapter will give information on how compliance rates can be determined and to what probability we can know that a ship is compliant given a certain measurement. Increasingly complex models will be used, in order to get a better understanding. The simplest model is the linear classification boundary. A bit more complex model is using the z-score in Section 4.3. The last section contains the most complex method which is the modeling of the two groups using an EM algorithm.

4.1 Measurements

In total 8049 measurements from at least 3197 ships were analyzed. The histogram of the fuel sulfur content is given in Figure 4.1. Note that the measurements originate from different apparatus and operators, as shown in Attachment A. In this paper, the label given by TNO was used as a reference. TNO used criteria including requirements about a minimum quality of the measurement in terms of ppb SO_2 and CO_2. TNO reported 9 % non-compliance, see Figure 4.2.
A problem is that some operators excluded outliers without further specifying criteria. Denmark reported more than half of the time a negative FSC, which is physically impossible. It was possible that also other operators excluded these measurements without reporting it either compliant or non-compliant. This would affect the compliance determination of all ships, as it is more likely that compliant ships would result in negative FSC measurement.
Another problem was that some reported FSC values could not be reproduced by the corresponding sulfur dioxide and carbon dioxide measurements. See Appendix F.

4.2 Linear decision boundary

A linear decision boundary is the most straightforward way to assign measurements either compliant or not. A certain value is chosen to be the decision boundary. All measurements below this boundary would be assigned to the compliant class and all points above will be assigned as non-compliant. A logical choice of the decision boundary would be 0.1 wt.%, since this is the legal limit where the hypotheses in Section 3.2.1 are based on. This would

result in 22 % non-compliance. In Figure 4.1 a histogram with all measurements is shown. The solid black line is the cumulative counts which can be used to easily obtain the results of a linear decision boundary.
The best way to evaluate whether the decision boundary is chosen correctly, is by comparing it with a validation set. When the decision boundary is set to 0.1 wt.% for both the remote measurements and laboratory results, an accuracy of $\frac{16}{30} = 53\,\%$ was found.

4.3 Z-score

The z-score is determined using the FSC and the standard deviation in that value. A common value chosen for α is 5 %. A histogram of all z-scores, the associated p-values and the effect of choosing this α is shown in Figure 4.3. Using $\alpha = 5\%$, the confusion matrix shown in Table 4.2 was calculated. The accuracy was 68 %. The relation between α, p-value, $\bar{x_{FSC}}$ and s_{FSC} is shown in Figure 4.4. In this figure, the blue line is the decision boundary based upon the z-score and $\alpha = 5\%$. The color and shape denotes the validated result, while the background color gives the probability obtained from the z-score at every spot. For example, TNO measured the most right point. A FSC of (0.53 ± 0.11) wt.% was measured. The marine fuel was labeled non-compliant, as one would expect with a high probability of having a FSC of more than 0.1 wt.%. However, the on-board inspection yielded a FSC of 0.068 wt.%, which is obviously compliant.
An equivalent figure was created for the complete database in Figure 4.5, where only 5 % is non-compliant. This time the color denotes the label given by TNO. For example, the point red triangle just above the label $\alpha = 0.05$ is a measurement which was reported non-compliant by TNO, had a p-value of 0.06 and thus was compliant using the z-score (since p-value $> \alpha$).

		Classification		
		Compliant	Non-compliant	Total
True value	Compliant	7	6	13
	Non-compliant	4	2	6
	Total	11	8	19

Table 4.1: Confusion matrix for linear decision boundary at 0.10 wt.% for the remote measurements.

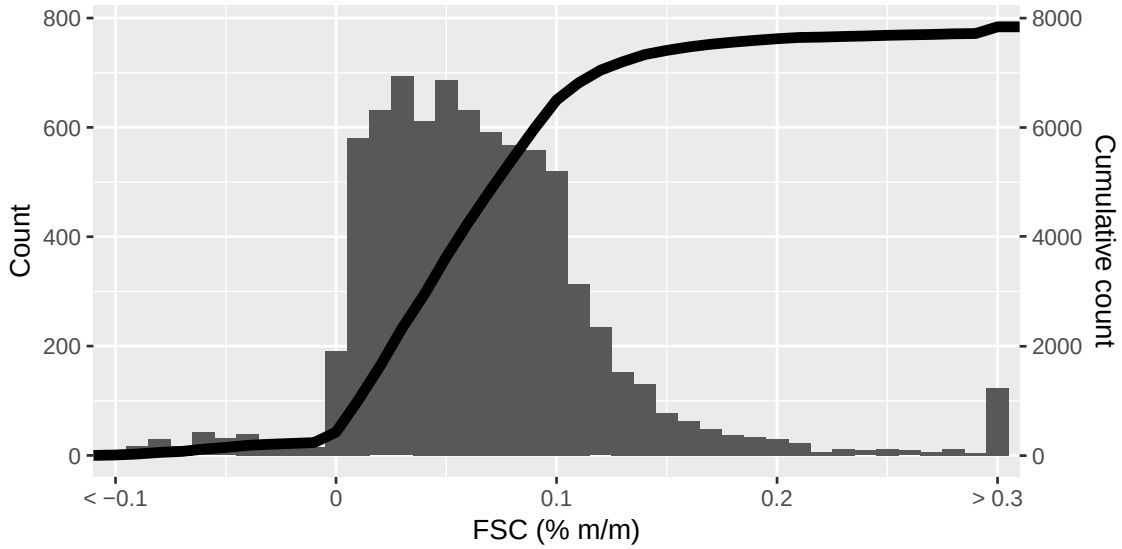

Figure 4.1: Histogram of FSC values. The black solid line is the cumulative count.

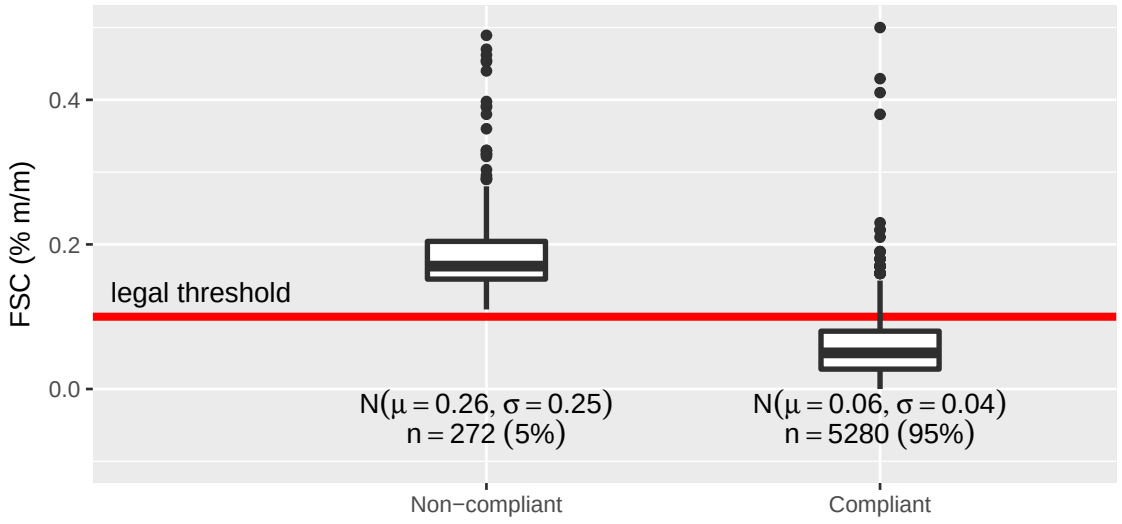

Figure 4.2: Box plot of compliance as determined by TNO. There were non-compliant measurements above the 0.5 wt.%. Data source: TNO

		Classification		
		Compliant	Non-compliant	Total
True value	Compliant	11	2	13
	Non-compliant	4	2	6
	Total	15	4	19

Table 4.2: Confusion matrix for classification using the z-score.

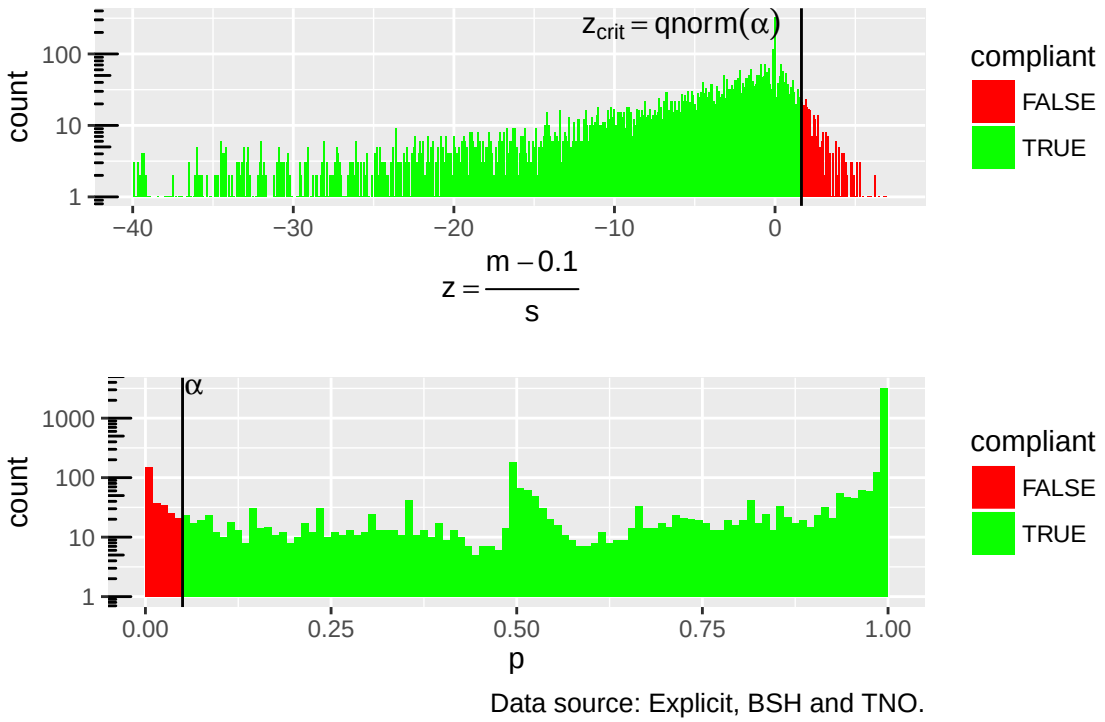

Figure 4.3: Classification using $\alpha = 0.05$. (Top) Histogram of all z-value. (Bottom) Histogram of the p-values, calculated from the z-score. This is the probability that that data point was obtained (or more extreme) given H_0

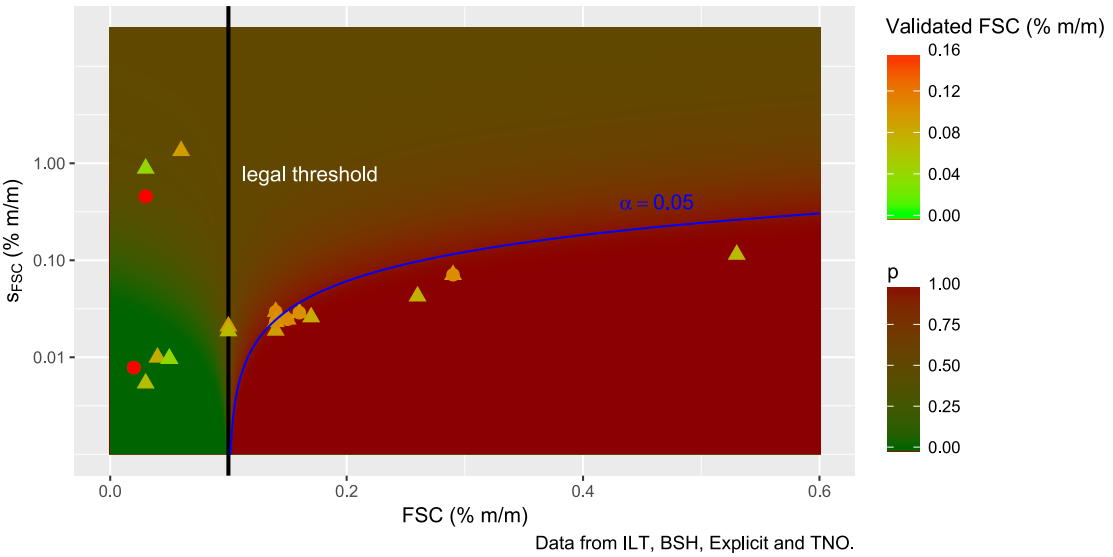

Figure 4.4: Cross-validation of measurements from Explicit and TNO with inspection results from ILT. The blue line gives the decision limit. The background color gives the probability of obtaining the data or more extreme given that a point is compliant. The shape of the point and the color gives what the value of the inspection results, where a triangle is below 0.1 wt.%, while a circle is below that value. The x- and y-axis denote the remote determination of the FSC and the associated standard error.

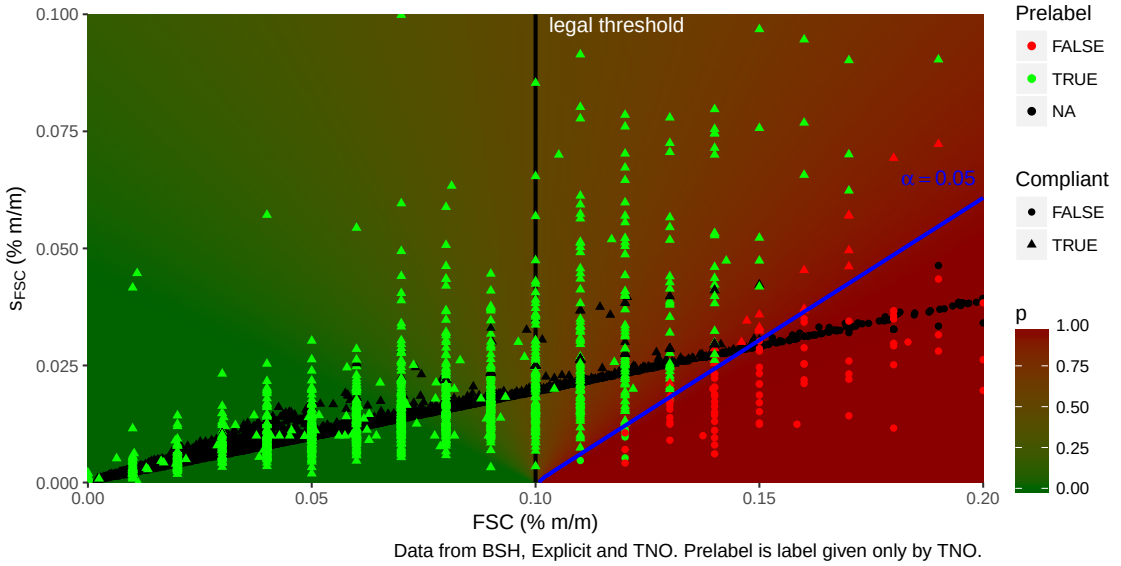

Figure 4.5: Relation between p-value, $x_{\bar{FSC}}$ and s_{FSC}. The blue line is the decision line.

4.4 EM algorithm

4.4.1 Bimodal normal distribution

In order to have an estimation of the errors encountered in the FSC values present, data modeling was applied. A bimodal distribution was used for the modeling, since there are two classes present in the data. At first, two normal modes were fitted. The initial parameters were used from the classification performed by TNO. They were used to plot Figure 3.2. The EM-algorithm did converge; in Figure C.1 the parameters are shown. The result is shown in Figure 4.6.

4.4.2 Bimodal distribution; normal and log-normal mode

In the TNO database, shown in Figure 4.7, it is clear that the non-compliant group (in red) is more log-normally distributed, since it has a longer tail. Therefor, the EM-algorithm was adjusted to accommodate a normal and a log-normal mode. The latter mode was shifted with 0.1 wt.%, to have support $(0.1, \infty)$. Using an EM-algorithm, a non-compliance rate of only 3 % was found as can be seen in Figure 4.8.

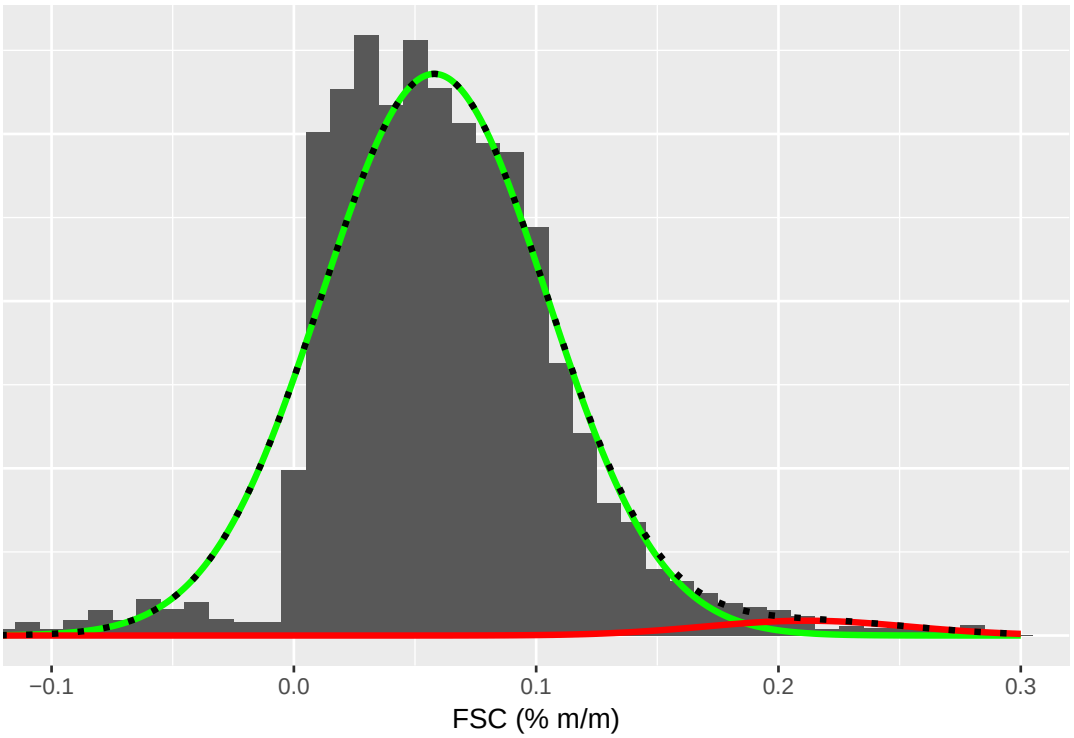

Figure 4.6: Result of fitting a bimodal normal distribution to the dataset. The green line denotes the compliant class, while the red denotes the non-compliant class. The sum is given by the dotted line.

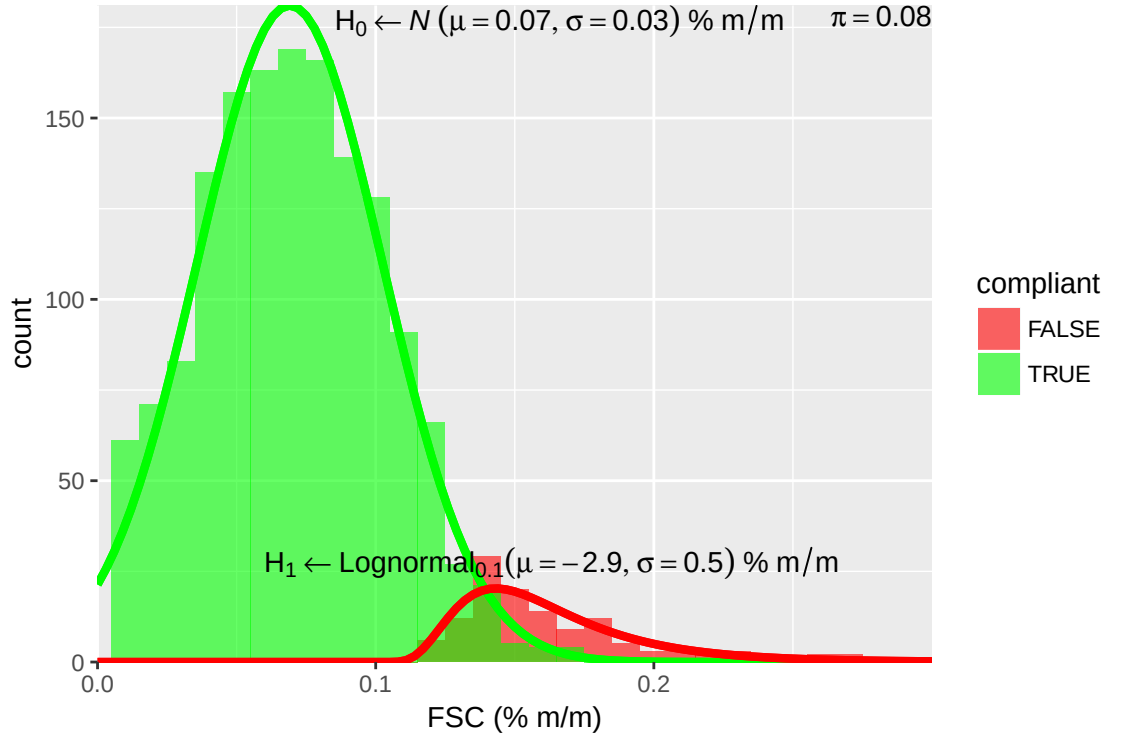

Figure 4.7: Data which was classified by TNO. To the compliant group a normal distribution was fitted, while for the non-compliant group a log-normal fit deemed to fit the best. A non-compliance rate of 8 % was found. Data obtained from TNO.

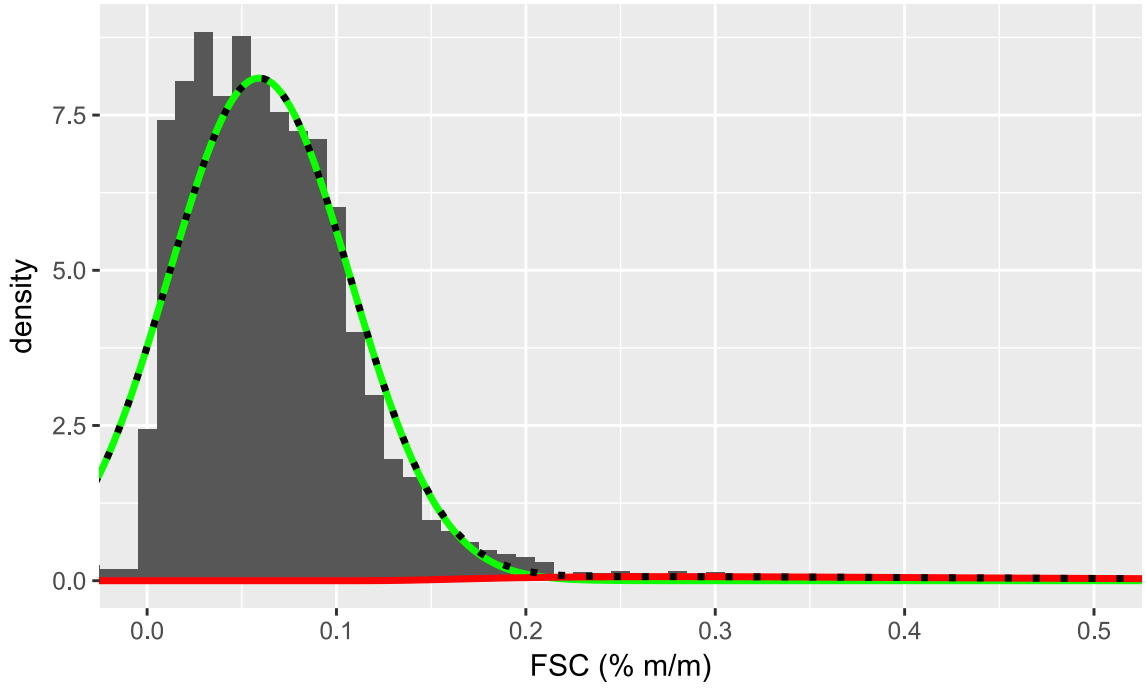

Figure 4.8: Bimodal distribution fitted to all data. The first mode is $\mathcal{N}\left(\boldsymbol{x}|\mu = 0.06, \sigma^2 = 0.0022\right)$, second mode is Log-normal $\left(\boldsymbol{x} - 0.1|\mu = -1.4, \sigma^2 = 0.6\right)$

Chapter 5

Discussion

This chapter will discuss the pros and cons of the applied techniques. It will start with evaluation on (the lack of) validation data. This notion will especially come back in the discussion of the results in the next two sections: Linear boundary classification and Z-score. The EM-algorithm offers an alternative approach to deal with the problem of lacking validation. The end of this chapter will provide discussion about how the measurements were performed.

5.1 Validation data

5.1.1 Validation against on-board inspection

Knowledge about the measurements, and the reliability of it, can be improved by expanding the validation set. Only 19 measurements with cross references of inspection data were available during this thesis. These 19 measurements are given in Appendix E. Figure 5.1 shows that in only $\frac{8}{293} = 2.7\,\%$ a ship was boarded when TNO sent the ILT an e-mail that a ship was non-compliant.[1] Most surprisingly is that only 1 out of 8 of these measurements was actually from an non-compliant ship.
One way too increase the number of measurements for the validation set is by sampling much faster. In the current system, after an alert from TNO it can take several weeks before the ship is boarded. Some delay might come because the ship is leaving the harbor, making it impossible to board the ship. But in many occasions the ship was however measured within two weeks, see Figure 5.2. This suggest that by coincidence the ship was measured and not because it was planned.

[1] These number was calculated using the data from ILT and TNO. Whether actually an e-mail was sent is unknown since TNO revoke or changed some measurements afterward. Another important remark is that the label was given originally by TNO. However it was not available for all measurements, so in this comparison a linear decision boundary at 0.12 wt.% was used. Additionally a match could only be made when both times an IMO number was reported, which was in about 91 % the case.

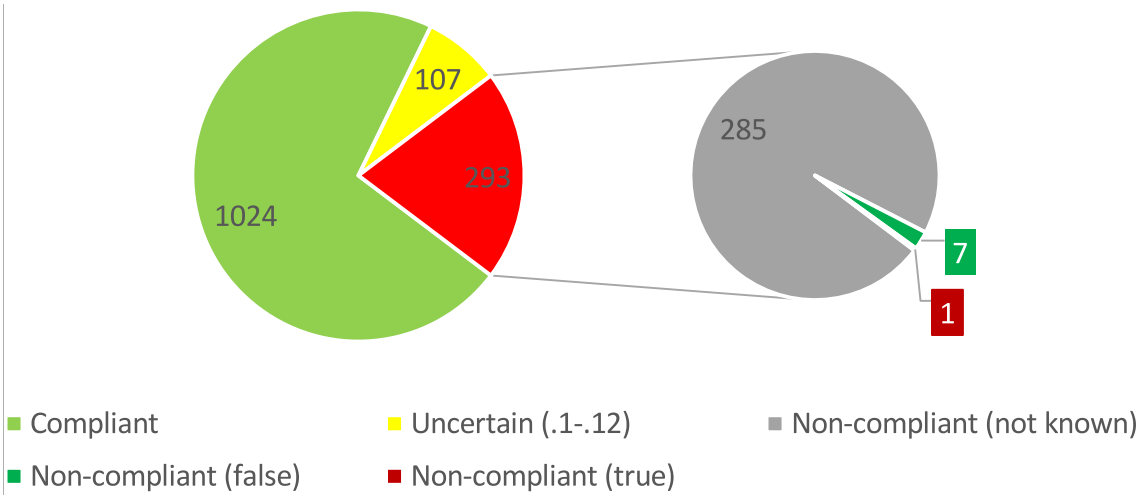

Figure 5.1: Left: The total number of measurements performed by TNO in 2016 and 2017, divided in the results of the measurements. Right: Zoom-in on the non-compliant measurements from TNO compared with the results of on-board inspection. Data from ILT and TNO.

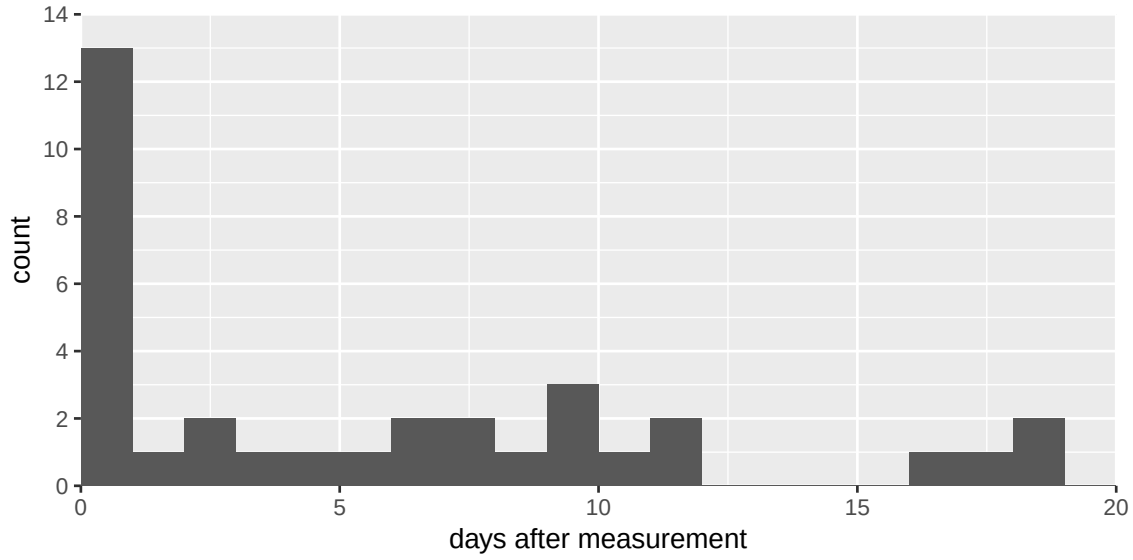

Figure 5.2: A histogram of the days ILT boarded the ship after a measurement from TNO. Data obtained from TNO and ILT.

5.1.2 Validation by cross comparison remote sensing

The amount of cross comparisons can also be increased by performing measurements with multiple instruments. The main advantage is the possibility to measure ten to twenty ships per hour in this way. The maximum amount of ships possible to measure with on-board inspection is two or three per day per person. The drawback of this approach is the difficulty to assess the bias and precision, since remote sensing has a low precision and high bias compared to an on-board inspection. A new campaign involving cross measurements is useful, but attention should be paid on how results should be interpret. Especially because the accuracy of the cross reference by remote sensing is much lower than a cross reference obtained by an on-board inspection. There is thus a trade-off for cross measurements between the amount of ships measured and the quality of these measurements.
An earlier approach concerning the comparison of multiple measurements by different techniques on individual ships was performed by Lööv *et al.* [29] This campaign was performed in 2009, when the maximum allowed FSC was 1.5 wt.%.

5.1.3 Validation against other information sources

It is useful to add data from other sources in future campaigns. All ships (above 5000 GT) have to report their carbon dioxide emissions. [72] These values can be useful to compare with the remote sensing measurements. Some ships are equipped with real time emission registration. If these ships will cooperate, this will allow both a high amount of ships measured and high quality of these measurements. Cooperation can be expected from ships involved in, for example, the Trident alliance. [73]

5.2 Type I and type II errors

The α should be chosen in such a way that it makes sense. This is the first problem, because what is a good value for α? An answer to this can only be given when the relation between α and β is known, so that the trade-off between type I and type II error can be chosen rationally. Three situations can be thought of:

- The α should be sufficient low if the remote measurement should be used in court. There is a strong preference not to prosecute ships that by nature of uncertainty are assumed non-compliant but are actually not. This is more important than improving the probability that a ship sailing with non-compliant fuel is caught.
- To get a best estimate of population behavior, α and β should be similar. There should be a balance between wrongly classified compliant and non-compliant ships.
- The sniffer is used as a screening method for PSC. A sufficient low β should be chosen in order to get all ships sailing with non-compliant marine fuel. More inspections are accepted as this will lower the probability that a non-compliant ship is not inspected.

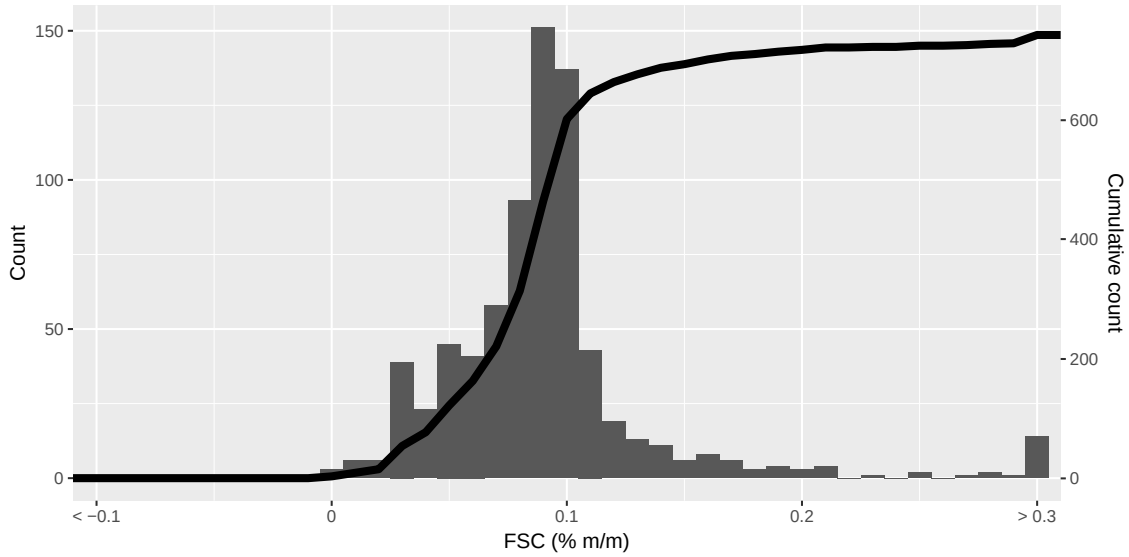

Figure 5.3: A histogram of all measurements performed by on-board inspection by ILT.

5.3 Linear boundary classification

The best linear boundary classification could be performed with the use of much validated data. Since, the data was only very limited available, it was not possible to train the model. Therefore, the data could only be used to calculate how well the classification took place.

Another problem is imposed by a low accuracy of the data. A quick look on the histogram on all data, Figure 4.1 on page 26, shows that there are much measurements with a FSC lower than 0.06 wt.%. If this was true, it is quite surprisingly, since sailing on such low sulfur fuel is much more expensive than sailing on 0.1 wt.%. A histogram with only measurements performed by an on-board inspection is shown in figure 5.3. The median of this distribution is 0.09 wt.%, which is higher than the median from all remote measurements (0.055 wt.%). Hence, lowering the bias and improving precision are important subjects for further research in the improvement of the analytical techniques.

With low precision it is impossible to separate to the compliant or non-compliant class only on the value of the FSC. A linear boundary classification would suffice if the precision would be low. Since this is not the case, a more complex model is used. In this complex model, added parameters like the reading for SO_2 and CO_2 could be added, together with there standard deviation.

5.4 Z-score

The z-score uses the information given by the standard deviation. The main assumption in this approach is that there is a normal (and thus symmetric) distributed error. However, the reported relative standard deviation was many times above the 100 %, see for an example Figure 5.4. Here the black area is proportional to the probability of having a negative FSC, which is not possible. Hence a symmetric probability distribution is very unlikely. Further investigation how the error should properly reported is inevitable for

calculating the probability that a ship is sailing with non-compliant fuel. An order of magnitude of this errors can be found by looking to a ship which runs likely always on the same quality fuel. This is true for the ship in Figure 5.5. The blue line is the linear regression line. Note the large variation which can be caused by a low precision. The measurement is stable over the six months.

5.5 EM

The lack of validation data stimulates the generation of a model which can simulate the variances without the need of validation data. These results can then be used in the power analysis. The best fit for the data gave a likelihood of only 3 %. This value is low compared to high quality in-port measurements, especially because measurements tends to have higher non-compliance levels at sea. All other methods gave much higher non-compliance rates. A reason why the non-compliance group might be under represented might be found in the very broad normal mode. Independent of initial parameters (within reasonable limits), the normal mode ended up being very broad. Maybe the measurements between 0 wt.% to 0.05 wt.% are indicative of the measurements where the conversion of SO_3 and SO_4 to SO_2 was not complete, as mentioned in Section 2.1. Or maybe these low reading was caused by washing out because of rain. Both will cause an underestimation of the SO_2 causing a lower apparent FSC than really is the case.

5.6 Analytical instruments

Precision can be improved with better instruments, but this can be limited by both costs and availability. Another strategy is by performing multiple (possibly independent) measurements. This can either be done by using a set-up in duplo (done by Explicit) or by applying two different techniques. A third possibility is to increase integration time. A helicopter or an UAV can stay easily one minute in a plume, whereas the airplane has at most a few seconds.
The precision can also be improved by measuring higher concentrations. In this thesis access to the raw data was merely available, but it is easy to understand that the concentrations of SO_2 and CO_2 are much higher just above the chimney. Optical methods measuring from several hundreds of meters suffer not only from much lower concentrations but also from much higher signal-to-noise ratios (SNRs). Especially the CO_2 background concentration can be heightened by other pollution sources, especially near the ports where the ground stations are often situated.
Another obvious problem by measurements performed at fixed stations is the possibility that shippers will avoid these spots. Or it is possible that ships shift to compliant fuel just before entering the areas which are controlled by fixed stations, since most ships are equipped with multiple fuel storage facilities.

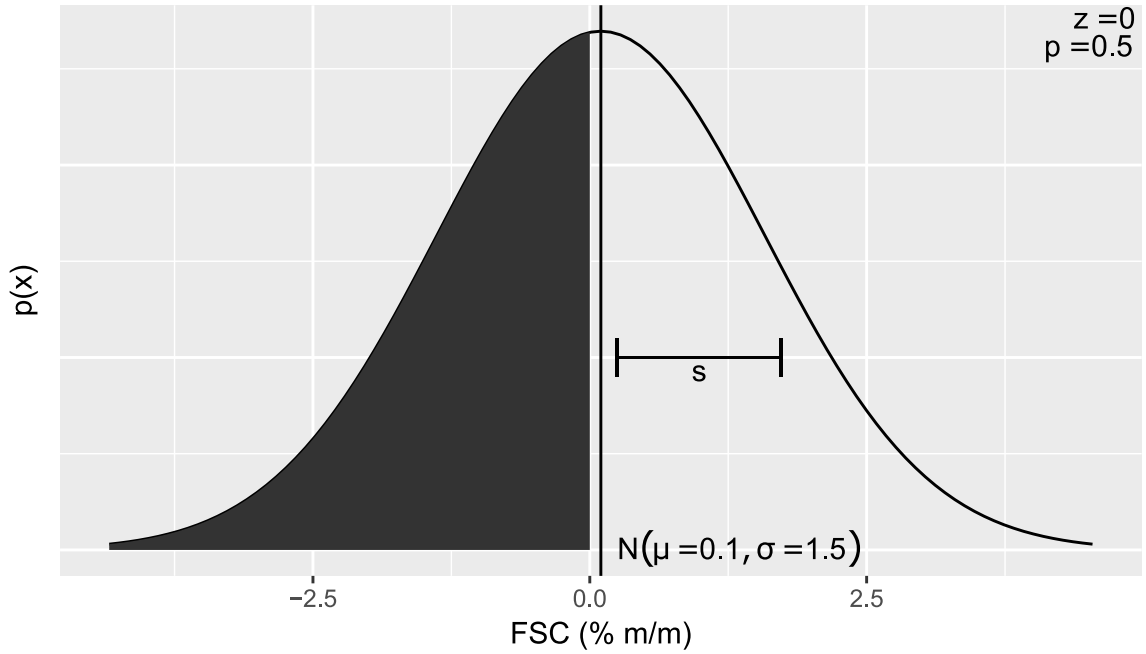

Figure 5.4: Some measurements had significant probability having a negative FSC if normality is assumed. Data source: TNO

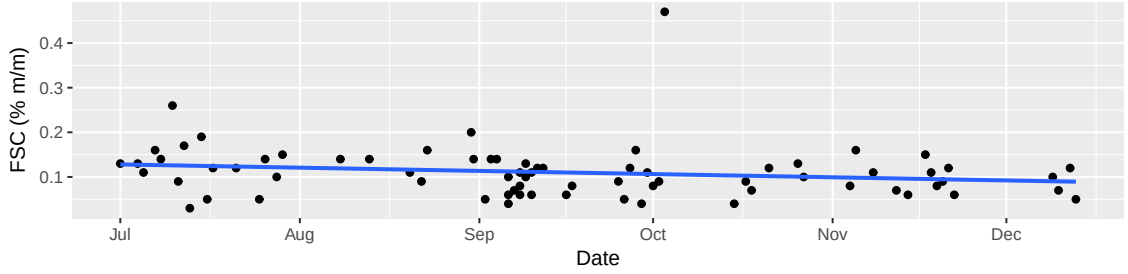

Figure 5.5: All measurements from the ship encountered the most times in the database. The ship was operating on a fixed route and likely used always the same fuel. Data from ILT and TNO.

Chapter 6

Conclusion

Data and measurement validation are of utmost importance when understanding the quality of remote sensing. Especially because the relative error in measurements can reach up to 100 %, classification of ships into compliant or non-compliant groups is difficult. Another struggle is the lack of validation through cross-reference measurements.
From an analytical perspective, this lack of validation makes it difficult to compare the different operators. Reported uncertainties are much too low to describe the difference between the few cross-reference measurements which were available. It is likely that a significant bias is present, both in the carbon dioxide measurements through varying background and in the sulfur dioxide measurements through washing out or other processes which remove sulfur.

6.1 Outlook

Future use of FSC measurements could be improved by measuring on full sea. This will prevent ships to avoid control and on full sea there are no other carbon dioxide emitters in the direct neighborhood. Another improvement would be the application of longer integration times or multiple measurements at the same time. This can lower the random error in the measurements and subsequently yield better results on determining whether the ship is compliant.
The data analysis after a campaign using such instruments will benefit from the lower uncertainties. But even better is to think a priori how many measurements should be made and which level of confidence are acceptable. The amount of measurements depend on the (expected) instrumental accuracy and which question needs to be answered.
A full Bayesian approach can result in a $p\left(\text{ship} \in \mathcal{C}_{\text{non-compliant}} | \mathcal{D}\right)$. Machine learning can possibly provide good priors given their AIS data and history to select ships which should be measured. In this way a double risk-driven inspection could be performed.
More research could probably lead to a future wherein non-compliant ships are directly fined after a measurement, with a fine depending on the sulfur content of the fuel. It can be just like speeding on the highway will result in a fine by a speed camera, with a fine depending on your speed. Who knows, maybe it is in the future possible to use satellite measurements to measure the FSC. In the analogy of the speed camera on the highway, this is like section control. Ships are regulated in that case over their complete journey. The fine could then depend on the climate and health effects imposed through

the higher FSC. It would be a perfect implementation of the principle of environmental law ‘the polluter pays’.

Bibliography

[1] V. Matthias, A. Aulinger, A. Backes, J. Bieser, B. Geyer, M. Quante, and M. Zeretzke, "The impact of shipping emissions on air pollution in the greater North Sea region-Part 2: Scenarios for 2030," *Atmospheric Chemistry and Physics*, vol. 16, no. 2, pp. 759–776, 2016.

[2] Port of Rotterdam, "Facts about the port of Rotterdam," 2016. [Online]. Available: https://www.portofrotterdam.com/nl/de-haven/feiten-en-cijfers-over-de-haven

[3] V. Eyring, I. S. Isaksen, T. Berntsen, W. J. Collins, J. J. Corbett, O. Endresen, R. G. Grainger, J. Moldanova, H. Schlager, and D. S. Stevenson, "Transport impacts on atmosphere and climate: Shipping," *Atmospheric Environment*, vol. 44, no. 37, pp. 4735–4771, 2010.

[4] J. J. Corbett, J. J. Winebrake, E. H. Green, P. Kasibhatla, V. Eyring, and A. Lauer, "Mortality from ship emissions: A global assessment," *Environmental Science and Technology*, vol. 41, no. 24, pp. 8512–8518, 2007.

[5] M. Amann, I. Bertok, J. Cofala, F. Gyarfas, C. Heyes, Z. Klimont, W. Schöpp, and W. Winiwarter, "Baseline Scenarios for the Clean Air for Europe (CAFE) Programme," European Commission, Laxenburg, Tech. Rep. 1, 2005.

[6] RIVM, "Emissieregistratie.nl." [Online]. Available: emissieregistratie.nl

[7] International Maritime Organization, "Sulphur oxides (SOx)." [Online]. Available: http://www.imo.org/en/OurWork/Environment/PollutionPrevention/AirPollution/Pages/Sulphur-oxides-(SOx)-âĂŞ-Regulation-14.aspx

[8] J. Kalli, S. Repka, and M. Alhosalo, "Estimating Costs and Benefits of Sulphur Content Limits in Ship Fuel," *International Journal of Sustainable Transportation*, vol. 9, no. 7, pp. 468–477, 2015. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/15568318.2013.808389

[9] International Maritime Organization, "Marpol Annex 6," 2005. [Online]. Available: http://www.imo.org/en/OurWork/Environment/PollutionPrevention/AirPollution/Pages/Air-Pollution.aspx

[10] National Geographic Society, "The World map: for the National Geographic Magazine." [Online]. Available: https://search.library.wisc.edu/catalog/9910195390902121

[11] Inspectie Leefomgeving en Transport, "Inspectieprogramma Vlaggenstaattoezicht meer effect en minder last," Inspectie Leefomgeving en Transport, Den Haag, Tech. Rep., 2014. [Online]. Available: https://www.ilent.nl/binaries/

ilt/documenten/publicaties/2014/06/25/inspectieprogramma-vlaggenstaattoezicht/Inspectieprogramma+Vlaggenstaattoezicht+2014.pdf

[12] D. Kubel, "Compliance Level at Different Sea Areas," Brussels, 2017. [Online]. Available: https://www.trafi.fi/filebank/a/1481613675/4dc191e52526ff2101d95f7143c8060e/23327-Compliance{_}rates{_}at{_}different{_}sea{_}areas{_}-{_}Dorte{_}Kubel.pdf

[13] J. Duyzer and H. Weststrate, "Meting van het zwavelgehalte van scheepsbrandstof door metingen in de pluim (een pilot study)," TNO, Utrecht, Tech. Rep., 2015.

[14] B. Alföldy, J. Balzani, and F. Lagler, "Remote sensing of ships' emissions of sulphur dioxide," 2011. [Online]. Available: http://ec.europa.eu/environment/air/transport/pdf/ships/Final-report.pdf

[15] L. Kattner, J. P. Burrows, A. Richter, S. Schmolke, A. Seyler, and F. Wittrock, "Monitoring compliance with sulfur content regulations of shipping fuel by in situ measurements of ship emissions," *Atmospheric Chemistry and Physics*, no. 15, pp. 10 087–10 092, 2015.

[16] A. J. Prata, "Measuring SO2 ship emissions with an ultraviolet imaging camera," *Atmospheric Measurement Techniques*, vol. 7, no. 5, 2014.

[17] B. Alföldy, J. B. Lööv, F. Lagler, J. Mellqvist, N. Berg, J. Beecken, H. Weststrate, J. Duyzer, L. Bencs, B. Horemans, F. Cavalli, J. P. Putaud, G. Janssens-Maenhout, A. P. Csordás, R. Van Grieken, A. Borowiak, and J. Hjorth, "Measurements of air pollution emission factors for marine transportation in SECA," *Atmospheric Measurement Techniques*, vol. 6, no. 7, pp. 1777–1791, 2013.

[18] J. Mellqvist, "Measurement techniques used within CompMon for fixed and airborne sulfur compliance monitoring," 2016. [Online]. Available: https://www.trafi.fi/filebank/a/1481613675/a425c23b2be7295f0ed86084d46d6ad3/23322-Measurement{_}tecniques{_}-{_}Johan{_}Mellqvist.pdf

[19] A. Seyler, "Retrieval of shipping emissions from MAX-DOAS measurements," Masterarbeit, Universität Bremen, 2014.

[20] F. Wittrock, "Einfluss von Schiffsemissionen in Nord- und Ostsee," Bremen, 2014. [Online]. Available: http://www.bsh.de/de/Das{_}BSH/Veranstaltungen/MUS/2014/Vortraege{_}24.{_}Meeresumwelt-Symposium{_}2014/Wittrock.pdf

[21] L. Kattner, B. Mathieu-üffing, M. Chirkov, J. Burrows, V. Matthias, E. Peters, A. Richter, A. Schönhardt, S. Schmolke, N. Theobald, S. Weigelt-krenz, and F. Wittrock, "MeSMarT âĂŞ Measurements of Shipping Emissions in the Marine Troposphere," p. 2641, 2013. [Online]. Available: http://www.doas-bremen.de/posters/egu{_}2013{_}kattner.pdfhttp://www.doas-bremen.de/posters/aq{_}2014{_}mathieu-ueffing.pdf

[22] A. Seyler, F. Wittrock, L. Kattner, B. Mathieu-üffing, E. Peters, A. Richter, S. Schmolke, N. Theobald, and J. P. Burrows, "Monitoring Shipping Emissions with MAX-DOAS Measurements of Reactive Trace Gases," p. 3618, 2015. [Online]. Available: http://www.doas-bremen.de/posters/egu{_}2014{_}wittrock.pdf

[23] L. Kattner, "MeSmarT âĂŞ Measurements of Shipping Emissions in the Marine Troposphere," Bremen, p. 28. [Online]. Available: http://www.mesmart.de/media/files/futoore{_}MeSmarT{_}public.pdf

[24] L. Kattner, B. Mathieu-üffing, A. Aulinger, J. Burrows, V. Matthias, D. Neumann, S. Schmolke, A. Seyler, N. Theobald, and F. Wittrock, "Monitoring Shipping Emissions with In-situ Measurements of Trace Gases," p. 2641, 2014. [Online]. Available: http://www.doas-bremen.de/posters/egu{_}2014{_}kattner.pdf

[25] A. Seyler, F. Wittrock, L. Kattner, B. Mathieu-Üffing, E. Peters, A. Richter, S. Schmolke, and J. P. Burrows, "Monitoring shipping emissions in the German Bight using MAX-DOAS measurements," *Atmospheric Chemistry and Physics Discussions*, no. 2, pp. 1–41, 2017. [Online]. Available: http://www.atmos-chem-phys-discuss.net/acp-2016-1153/

[26] J. Mellqvist, J. Beecken, and J. Ekholm, "Compliance measurements of ships from airborne and fixed stations," 2016. [Online]. Available: http://www.bonnagreement.org/site/assets/files/16918/chalmers{_}igps.pdf

[27] J. Mellqvist and N. Berg, "Identification of gross polluting ships," Göteborg, 2010. [Online]. Available: http://publications.lib.chalmers.se/records/fulltext/248358/local{_}248358.pdf

[28] N. Berg, J. Mellqvist, J.-P. Jalkanen, and J. Balzani, "Ship emissions of SO₂ and NO₂: DOAS measurements from airborne platforms," *Atmospheric Measurement Techniques*, vol. 5, no. 5, pp. 1085–1098, may 2012. [Online]. Available: http://www.atmos-meas-tech.net/5/1085/2012/

[29] J. M. Balzani Lööv, B. Alfoldy, L. F. L. Gast, J. Hjorth, F. Lagler, J. Mellqvist, J. Beecken, N. Berg, J. Duyzer, H. Westrate, D. P. J. Swart, A. J. C. Berkhout, J. P. Jalkanen, A. J. Prata, G. R. Van Der Hoff, and A. Borowiak, "Field test of available methods to measure remotely SOx and NOx emissions from ships," *Atmospheric Measurement Techniques*, vol. 7, no. 8, pp. 2597–2613, 2014.

[30] N. Berg, J. Mellqvist, J. Beecken, and J. Johansson, "Ship Emission Measurements by the Chalmers IGPS System during the Rotterdam campaign 2009," Göteborg, 2010. [Online]. Available: http://ec.europa.eu/environment/air/transport/pdf/ships/Appendix1.pdf

[31] J. Beecken, "Remote Measurements of Gas and Particulate Matter Emissions from Individual Ships," Thesis for the degree of doctor of philosophy, Chalmers University of Technology, 2015.

[32] J. Beecken, J. Mellqvist, K. Salo, J. Ekholm, and J.-P. Jalkanen, "Airborne emission measurements of SO2, NOx and particles from individual ships using a sniffer technique," *Atmospheric Measurement Techniques*, vol. 7, no. 7, pp. 1957–1968, jul 2014. [Online]. Available: http://www.atmos-meas-tech.net/7/1957/2014/

[33] J. Beecken, J. Mellqvist, K. Salo, J. Ekholm, J. P. Jalkanen, L. Johansson, V. Litvinenko, K. Volodin, and D. A. Frank-Kamenetsky, "Emission factors of SO2, NOx and particles from ships in Neva Bay from ground-based and helicopter-borne measurements and AIS-based modeling," *Atmospheric Chemistry and Physics*, vol. 15, no. 9, pp. 5229–5241, 2015.

[34] J. Mellqvist, "Compliance measurements of ships from airborne and fixed platforms within the identification of gross polluting ships project." [Online]. Available: http://www.bonnagreement.org/site/assets/files/16918/chalmers{_}igps.pdf

[35] J. Mellqvist, J. Ekholm, K. Salo, and J. Beecken, "Identification of Gross Polluting Ships To Promote a Level Playing Field Within the Shipping Sector," p. 39, 2014. [Online]. Available: http://publications.lib.chalmers.se/records/fulltext/214636/local{_}214636.pdf

[36] W. Van Roy, "Best Practices Airborne MARPOL Annex VI Monitoring," 2016. [Online]. Available: http://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetailDoc{&}id=29311{&}no=7

[37] H. Schlager, R. Baumann, M. Lichtenstern, A. Petzold, F. Arnold, M. Speidel, C. Gurk, and H. Fischer, "Aircraft-based Trace Gas Measurements in a Primary European Ship Corridor," in *Proceedings of the TAC-Conference, June 26 to 29, Oxford, UK*, Oxford, 2006, pp. 83–88.

[38] B. Knudsen and M. Berghout, "A Path to Reliable Compliance Detection and Enforcement âĂŞ Lessons from Rotterdam," Copenhagen, 2017.

[39] F. Drewnick, T. Böttger, S. L. Von Der Weiden-Reinmüller, S. R. Zorn, T. Klimach, J. Schneider, and S. Borrmann, "Design of a mobile aerosol research laboratory and data processing tools for effective stationary and mobile field measurements," *Atmospheric Measurement Techniques*, vol. 5, no. 6, pp. 1443–1457, 2012.

[40] J. M. Diesch, F. Drewnick, T. Klimach, and S. Borrmann, "Investigation of gaseous and particulate emissions from various marine vessel types measured on the banks of the Elbe in Northern Germany," *Atmospheric Chemistry and Physics*, vol. 13, no. 7, pp. 3603–3618, 2013.

[41] L. Pirjola, A. Pajunoja, J. Walden, J.-P. Jalkanen, T. Rönkkö, A. Kousa, and T. Koskentalo, "Mobile measurements of ship emissions in two harbour areas in Finland," *Atmospheric Measurement Techniques*, vol. 7, no. 1, pp. 149–161, jan 2014. [Online]. Available: http://www.atmos-meas-tech.net/7/149/2014/

[42] J. B. Nowak, D. D. Parrish, J. A. Neuman, J. S. Holloway, O. R. Cooper, T. B. Ryerson, J. K. Nicks, F. Flocke, J. M. Roberts, E. Atlas, J. A. de Gouw, S. Donnelly, E. Dunlea, G. Hübler, L. G. Huey, S. Schauffler, D. J. Tanner, C. Warneke, and F. C. Fehsenfeld, "Gas-phase chemical characteristics of Asian emission plumes observed during ITCT 2K2 over the eastern North Pacific Ocean," *Journal of Geophysical Research D: Atmospheres*, vol. 109, no. 23, pp. 1–18, 2004.

[43] D. A. Lack, J. J. Corbett, T. Onasch, B. Lerner, P. Massoli, P. K. Quinn, T. S. Bates, D. S. Covert, D. Coffman, B. Sierau, S. Herndon, J. Allan, T. Baynard, E. Lovejoy, A. R. Ravishankara, and E. Williams, "Particulate emissions from commercial shipping: Chemical, physical, and optical properties," *Journal of Geophysical Research Atmospheres*, vol. 114, no. 4, pp. 1–16, 2009.

[44] E. J. Williams, B. M. Lerner, P. C. Murphy, S. C. Herndon, and M. S. Zahniser, "Emissions of NOx, SO2, CO, and HCHO from commercial marine shipping during Texas Air Quality Study (TexAQS) 2006," *Journal of Geophysical Research*, vol. 114, no. D21, p. D21306, nov 2009. [Online]. Available: http://doi.wiley.com/10.1029/2009JD012094

[45] D. A. Lack, C. D. Cappa, J. Langridge, R. Bahreini, G. Buffaloe, C. Brock, K. Cerully, D. Coffman, K. Hayden, J. Holloway, B. Lerner, P. Massoli, S. M. Li, R. McLaren, A. M. Middlebrook, R. Moore, A. Nenes, I. Nuaaman, T. B. Onasch, J. Peischl, A. Perring, P. K. Quinn, T. Ryerson, J. P. Schwartz, R. Spackman, S. C. Wofsy, D. Worsnop, B. Xiang, and E. Williams, "Impact of fuel quality regulation and speed reductions on shipping emissions: Implications for climate and air quality," *Environmental Science and Technology*, vol. 45, no. 20, pp. 9052–9060, 2011.

[46] C. D. Cappa, E. J. Williams, D. A. Lack, G. M. Buffaloe, D. Coffman, K. L. Hayden, S. C. Herndon, B. M. Lerner, S. M. Li, P. Massoli, R. McLaren, I. Nuaaman, T. B. Onasch, and P. K. Quinn, "A case study into the measurement of ship emissions from plume intercepts of the NOAA ship Miller Freeman," *Atmospheric Chemistry and Physics*, vol. 14, no. 3, 2014.

[47] A. P. Ault, C. J. Gaston, Y. Wang, G. Dominguez, M. H. Thiemens, and K. A. Prather, "Characterization of the single particle mixing state of individual ship plume events measured at the port of Los Angeles," *Environmental Science and Technology*, vol. 44, no. 6, pp. 1954–1961, 2010.

[48] P. V. Hobbs, T. J. Garrett, R. J. Ferek, S. R. Strader, D. A. Hegg, G. M. Frick, W. A. Hoppel, R. F. Gasparovic, L. M. Russell, D. W. Johnson, C. O'Dowd, P. A. Durkee, K. E. Nielsen, and G. Innis, "Emissions from Ships with respect to Their Effects on Clouds," *Journal of the Atmospheric Sciences*, vol. 57, no. 16, pp. 2570–2590, 2000. [Online]. Available: http://journals.ametsoc.org/doi/abs/10.1175/1520-0469(2000)057{%}3C2570:EFSWRT{%}3E2.0.CO;2

[49] P. Sinha, P. V. Hobbs, R. J. Yokelson, T. J. Christian, T. W. Kirchstetter, and R. Bruintjes, "Emissions of trace gases and particles from two ships in the southern Atlantic Ocean," *Atmospheric Environment*, vol. 37, no. 15, pp. 2139–2148, 2003.

[50] M. Mooij, E. Gerlofs - Nijland, and D. Swart, "Zeeschepen: metingen van chemische stoffen in rookgassen en brandstoffen," Bilthoven, 2010. [Online]. Available: http://www.rivm.nl/dsresource?objectid=b271368e-6d3d-4c16-853c-77aa834d3927{&}type=org{&}disposition=inline

[51] A. Berkhout, D. Swart, G. van der Hoff, and J. Bergwerff, "Sulphur dioxide emissions of oceangoing vessels measured remotely with Lidar," Bilthoven, 2012. [Online]. Available: http://www.rivm.nl/bibliotheek/rapporten/609021119.pdf

[52] D. P. J. Swart, A. Berkhout, G. Van der Hoff, J. Bergwerf, and M. Broekman, "Zwaveldioxide-uitstoot van zeeschepen op afstand gemeten met lidar," Bilthoven, 2007. [Online]. Available: http://www.pbl.nl/sites/default/files/cms/publicaties/500045002.pdf

[53] J. Duyzer, K. Hollander, M. Voogt, H. Verhagen, H. Weststrate, A. Hensen, A. Kraai, and G. Kos, "Assessment of emissions of PM and NOx of sea going vessels by field measurements," 2006. [Online]. Available: http://publications.tno.nl/publication/102469/F9dj0x/2006-A-R0341.pdf

[54] Andor, "Andor Shamrock 303i." [Online]. Available: http://www.andor.com/spectrograph/shamrock-spectrograph-series/shamrock-303i

[55] ——, "Andor Newton 920." [Online]. Available: http://www.andor.com/scientific-cameras/newton-ccd-and-emccd-cameras/newton-920

[56] J. Knudsen, "A method and an unmanned aerial vehicle for determining emissions of a vessel," 2015.

[57] ——, "US20170003684A1.pdf," 2017.

[58] Teledyne, "Teledyne 100E," p. 388, 2011. [Online]. Available: http://www.teledyne-api.com/manuals/04515F{_}100E.pdf

[59] Thermo Scientific, "Thermo 43i TL," Franklin, 2015. [Online]. Available: https://www.thermofisher.com/order/catalog/product/43ITLE

[60] Ø. Endresen, E. Sørgard, J. Sundet, S. Dalsøren, I. Isaksen, and T. Berglen, "Emission from international sea transportation and environmental impact," *Journal of Geophysical Research*, vol. 108, no. D17, pp. 1–22, 2003.

[61] J. J. Corbett, P. S. Fischbeck, and S. N. Pandis, "Global nitrogen and sulfur inventories for oceangoing ships," *Journal of Geophysical Research*, vol. 104, no. D3, pp. 3457–3470, 1999.

[62] HORIBA, "Horiba APSA-370 Ambient Sulfur Dioxide Monitor." [Online]. Available: http://www.horiba.com/process-environmental/products/ambient/details/apsa-370-ambient-sulfur-dioxide-monitor-272/

[63] Thermo, "Thermo 43C," Franklin, 2004. [Online]. Available: http://www.thermo.com.cn/Resources/200802/productPDF{_}12267.pdf

[64] Airpointer, "Airpointer." [Online]. Available: https://www.airpointer.com/{#}sc-tabs-1495397994693

[65] Picarro, "PICARRO G2301 CRDS Analyzer CO2 CH4 H2O Measurements in Air," Picarro, Santa Clara, Tech. Rep., 2015. [Online]. Available: http://www.picarro.com/assets/docs/Data{_}Sheets/Data{_}Sheet{_}G2301{_}CRDS{_}Analyzer{_}for{_}CO2{_}CH4{_}H2O{_}in{_}Air{_}Data{_}Sheet.pdf

[66] LICOR, "LI-7200," Lincoln, pp. 14–17, 2001. [Online]. Available: https://www.licor.com/documents/k8b90wiwdhietarx04qd

[67] ——, "LI-840A," Lincoln. [Online]. Available: https://www.licor.com/documents/3n6u6pbsv3fafla3csyj

[68] ——, "LI-6262," Lincoln, 1996. [Online]. Available: http://envsupport.licor.com/index.jsp?menu=Gas{_}Analyzers{&}spec=LI-6262,Manuals

[69] ——, "LI-820," Lincoln. [Online]. Available: https://www.licor.com/env/products/gas{_}analysis/LI-820/

[70] SAS Institute, "SAS 9.2 Software," Cary, USA, 2017. [Online]. Available: http://support.sas.com/documentation/92/

[71] R Core Team, "R: A Language and Environment for Statistical Computing," Vienna, Austria, 2017. [Online]. Available: https://www.r-project.org/

[72] EMSA, "THETIS-MRV," 2017. [Online]. Available: https://mrv.emsa.europa.eu/{#}public

[73] Trident, "Trident Alliancy," 2017. [Online]. Available: http://www.tridentalliance.org/

Appendices

Appendix A

Parameters in database

The following table shows all parameters which are included in the database containing all measurements. For each agency (columns) the fraction of non-empty cells per parameters is shown. An italic *1* denotes that all entries in the database were provided with the corresponding parameter. At the last row, the absolute amount of entries for each agency is shown. The table was rendered using `Pandoc` in `R`.[71]

	BSH	ILT	TNO	MUMM	Explicit	DFDS-Maersk	Denmark	Average
Date	*1*	*1*	*1*	*1*	*1*	*1*	*1*	**1**
FSC	*1*	*1*	*1*	0.85	*1*	*1*	*1*	**0.97**
Instrument	0	*1*	0	0	0	0	0	**0.09**
Time	*1*	0	*1*	*1*	*1*	0.6	*1*	**0.91**
Longitude	0	0	0	0.98	*1*	0	*1*	**0.25**
Latitude	0	0	0	0.98	*1*	0	*1*	**0.25**
Length	*1*	0	0	0.99	*1*	0	*1*	**0.7**
Width	*1*	0	0	0.99	*1*	0	*1*	**0.7**
Destination	0.97	0	0.97	0.94	0.85	0	*1*	**0.87**
Speed	*1*	0	*1*	0.98	*1*	0	*1*	**0.9**
Course	*1*	0	*1*	0.98	0.94	0	*1*	**0.9**
Sign	*1*	0	0	0	0	0	0	**0.44**
FSC_SD	*1*	0	*1*	0	*1*	0	0	**0.69**
SO2	*1*	0	*1*	0	0	0	0	**0.65**
CO2	*1*	0	*1*	0	0	0	0	**0.65**
NO	*1*	0	*1*	0	0	0	0	**0.65**
NOx	*1*	0	0	0	0	0	0	**0.44**
O3	0.94	0	0	0	0	0	0	**0.42**
NO2	*1*	0	*1*	0	0	0	0	**0.65**
Plumetime	0.66	0	*1*	0	0	0	0	**0.5**
Airpressure	*1*	0	0	0	0	0	0	**0.44**
SO2_SD	*1*	0	0.97	0	0	0	0	**0.64**
Dewpoint	*1*	0	0	0	0	0	0	**0.44**
CO2_SD	*1*	0	0.97	0	0	0	0	**0.64**
NO_SD	*1*	0	0	0	0	0	0	**0.44**
NO2_SD	*1*	0	0	0	0	0	0	**0.44**
NOx_SD	*1*	0	0	0	0	0	0	**0.44**
O3_SD	0.94	0	0	0	0	0	0	**0.42**
Windgust	*1*	0	0	0	0	0	0	**0.44**
Temperature_housing	*1*	0	0	0	0	0	0	**0.44**
Precipitationtype	*1*	0	0	0	0	0	0	**0.44**
Rainintensity	*1*	0	0	0	0	0	0	**0.44**
Humidity	*1*	0	*1*	0	0	0	0	**0.65**
Temperature	*1*	0	*1*	0	0	0	0	**0.65**
Winddirection	*1*	0	*1*	0	0	0	0	**0.65**
Windspeed	*1*	0	*1*	0	0	0	0	**0.65**
Windchill	*1*	0	0	0	0	0	0	**0.44**
Comment	*1*	0	0	0	0.07	0	0	**0.44**
Platform	0	0	0	0	0	0	*1*	**0.04**
FSC_RSD	*1*	0	*1*	0	0.99	0	0	**0.69**
Distance	0	0	*1*	0	0	0	0	**0.21**
Pressure	0	0	*1*	0	0	0	0	**0.21**
Draught	0	0	*1*	0	0	0	0	**0.21**
Movement	0	0	0	0	*1*	0	0	**0.04**
Heading	0	0	0	0	0.99	0	0	**0.04**
NOx_Power	0	0	0	0	*1*	0	0	**0.04**
NOx_Power_SD	0	0	0	0	*1*	0	0	**0.04**
Compliant	0	0	*1*	0	0	0	0	**0.21**
Quality	0	0	0	0	0	0	*1*	**0.04**
Flag	0	0.1	*1*	0.97	*1*	0	0.99	**0.47**
Flag_long	0	0.1	*1*	0.97	*1*	0	0.99	**0.47**
Flag_short	0	0.1	*1*	0.97	*1*	0	0.99	**0.47**
Status	0.01	0	0	0	0	0	0	**0.01**
Status_string	0.01	0	0	0	0	0	0	**0.01**
Type	0.97	0.21	0.98	0.94	*1*	0	0.99	**0.9**
Type_string	0.97	0.21	0.98	0.94	*1*	0	0.99	**0.9**
ID	0.78	0.96	0.88	0.99	0.99	*1*	0.96	**0.87**
H2OBaseline	0	0	*1*	0	0	0	0	**0.21**
CO2Baseline	0	0	*1*	0	0	0	0	**0.21**
Manual	*1*	0	*1*	0	0	0	0	**0.65**
Total entries	**3564**	**743**	**1661**	**1390**	**327**	**10**	**354**	**8049**

Appendix B

MLE of bimodal distribution

A complete derivation is given from the MLEs of the bimodal distribution with a normal and a shifted lognormal mode, which is explained in section 3.4. The likelihood function from the model is:

$$\begin{cases} p\left(\boldsymbol{x}|\boldsymbol{a},\boldsymbol{b},\mu_1,\mu_2,\sigma_1,\sigma_2\right) = \prod_{i=1}^{N}\Big(a_i\ \mathcal{N}(x_i|\mu_1,\sigma_1) + b_i\ \text{Lognormal}(x_i-0.1|\mu_2,\sigma_2)\Big) \\ a_i + b_i = 1 \text{ for every } 1 \end{cases} \tag{B.1}$$

The maximum likelihood estimator of μ_1 can be find using:

$$\frac{\partial p\left(\boldsymbol{x}|\boldsymbol{a},\boldsymbol{b},\mu_1,\mu_2,\sigma_1,\sigma_2\right)}{\partial \mu_1} = 0 \tag{B.2}$$

Solving this equation is difficult because of the product in the likelihood. In order to get rid of this product, we can calculate the derivative of the log-likelihood because maximizing a likelihood is equivalent to maximizing a log-likelihood. The log-likelihood is given by:

$$\begin{aligned} &\log p\left(\boldsymbol{x}|\boldsymbol{a},\boldsymbol{b},\boldsymbol{\mu},\boldsymbol{\sigma}\right) \\ &\qquad = \log \prod_{i=1}^{N}\Big(a_i\ \mathcal{N}(x_i|\mu_1,\sigma_1) + b_i\ \text{Lognormal}(x_i-0.1|\mu_2,\sigma_2)\Big) \\ &\qquad = \sum_{i=1}^{N} \log\Big(a_i\ \mathcal{N}(x_i|\mu_1,\sigma_1) + b_i\ \text{Lognormal}(x_i-0.1|\mu_2,\sigma_2)\Big) \end{aligned} \tag{B.3}$$

Solving this equation is still problematic, since taking the derivative towards μ_1 will yield terms with μ_2 (and *vice versa*) while these parameters needs also to be optimized. To circumvent this problem each point is classified towards either compliant (and hence will contribute to the normal mode) or non-compliant (and hence will contribute to the shifted log-normal mode). This means that $\boldsymbol{a}$ and $\boldsymbol{b}$ are either 0 or 1. Equation B.3 can then be rewritten:

$$\begin{cases} \log p\left(\boldsymbol{x}|\boldsymbol{a},\boldsymbol{b},\boldsymbol{\mu},\boldsymbol{\sigma}\right) = \sum\limits_{i=1}^{N} \log\Big(\big(\mathcal{N}(x_i|\mu_1,\sigma_1)\big)^{a_i} + \big(\text{Lognormal}(x_i-0.1|\mu_2,\sigma_2)\big)^{b_i}\Big) \\ a_i + b_i = 1 \text{ for every } i \\ a_i, b_i \in \{0,1\} \end{cases} \tag{B.4}$$

And this equation can be simplified in such a way that only one term is present in the logarithm:

$$\log p\left(\boldsymbol{x}|\boldsymbol{a},\boldsymbol{b},\boldsymbol{\mu},\boldsymbol{\sigma}\right)=\sum_{i=1}^{N}a_i\log\mathcal{N}\left(x_i|\mu_1,\sigma_1\right)+\sum_{i=1}^{N}b_i\log\text{Lognormal}\left(x_i-0.1|\mu_2,\sigma_2\right) \quad \text{(B.5)}$$

When differentiating to μ_1 and σ_1 would use only the first term, while differentiating to μ_2 and σ_2 will need only the second.

B.1 Solving for the paramaters

B.1.1 MLE for the mean of the normal mode

$$\begin{aligned}\frac{\partial\log p\left(\boldsymbol{x}|\boldsymbol{a},\boldsymbol{b},\boldsymbol{\mu},\boldsymbol{\sigma}\right)}{\partial\mu_1}=&\frac{\partial\ \sum_{i=1}^{N}a_i\log\mathcal{N}\left(x_i|\mu_1,\sigma_1\right)}{\partial\mu_1}\\&=\sum_{i=1}^{N}\frac{\partial\ a_i\left(-\frac{1}{2}\log\left(2\pi\sigma_1^2\right)-\frac{\left(x_i-\mu_1\right)^2}{2\sigma_1^2}\right)}{\partial\mu_1}\\&=\sum_{i=1}^{N}a_i\frac{x_i-\mu_1}{\sigma_1^2}=0\end{aligned} \quad \text{(B.6)}$$

This last equality is equal to:

$$\sum_{i=1}^{N}a_ix_i=\sum_{i=1}^{N}a_i\mu_1^*\rightarrow\mu_1^*=\frac{\sum_{i=1}^{N}a_ix_i}{\sum_{i=1}^{N}a_i} \quad \text{(B.7)}$$

B.1.2 MLE for the standard deviation of the normal mode

$$\frac{\partial\log p\left(\boldsymbol{x}|\boldsymbol{a},\boldsymbol{b},\boldsymbol{\mu},\boldsymbol{\sigma}\right)}{\partial\sigma_1^2}=\sum_{i=1}^{N}a_i\left(-\frac{1}{2\sigma_1^2}+\frac{\left(x_i-\mu_1\right)^2}{2\sigma_1^4}\right)=0 \quad \text{(B.8)}$$

$$\sum_{i=1}^{N}a_i\left(-1+\frac{\left(x_i-\mu_1^2\right)^2}{\sigma_1^2}\right)=0 \quad \text{(B.9)}$$

$$\sum_{i=1}^{N}a_i=\sum_{i=1}^{N}a_i\frac{x_i-\mu_1}{\sigma_1^{2*}} \quad \text{(B.10)}$$

$$\sigma_1^{2*}=\frac{\sum_{i=1}^{N}a_i\left(x_i-\mu_1\right)^2}{\sum_{i=1}^{N}a_i} \quad \text{(B.11)}$$

B.1.3 MLE for the mean of the shifted lognormal mode

$$\begin{aligned}\frac{\partial \log p\left(\boldsymbol{x}|\boldsymbol{a},\boldsymbol{b},\boldsymbol{\mu},\boldsymbol{\sigma}\right)}{\partial \mu_2} &= \frac{\partial \sum_{i=1}^{N} b_i \log \text{Lognormal}\left(x_i - 0.1|\mu_2,\sigma_2^2\right)}{\partial \mu_2} \\ &= \sum_{i=1}^{N} \frac{\partial\, b_i \left(-\log(x_i-0.1) - \frac{1}{2}\log(2\sigma_2^2\pi) - \frac{(\log(x_i-0.1)-\mu_2)^2}{2\sigma_2^2}\right)}{\partial\, \mu_2} \\ &= \sum_{i=1}^{N} b_i \frac{\log(x_i-0.1)+\mu_2}{\sigma_2^2} = 0\end{aligned} \tag{B.12}$$

$$\mu_2^* = \frac{\sum_{i=1}^{N} b_i \log(x_i-0.1)}{\sum_{i=1}^{N} b_i} \tag{B.13}$$

B.1.4 MLE for the standard deviation of the shifted lognormal mode

$$\begin{aligned}\frac{\partial \log p\left(\boldsymbol{x}|\boldsymbol{a},\boldsymbol{b},\boldsymbol{\mu},\boldsymbol{\sigma}\right)}{\partial \sigma_2} &= \sum_{i=1}^{N} b_i \left(-\frac{1}{2\sigma_2^2} + \frac{\left(\log(x_i-0.1)-\mu_2\right)^2}{2\sigma^4}\right) \\ &= \sum_{i=1}^{N} b_i \left(\sigma_2^2 - \left(\log(x_i-0.1)\right)^2\right) \\ &= 0\end{aligned} \tag{B.14}$$

$$\sigma_2^{2*} = \frac{\sum_{i=1}^{N} b_i \left(\log(x_i-0.1)-\mu_2\right)^2}{\sum_{i=1}^{N} b_i} \tag{B.15}$$

B.2 Using responsibilities

Another way to interpret the MLEs for the mean of both modes is the following:

$$\begin{aligned}\mu_1^* &= \frac{\sum_{i\in\mathcal{C}_{\text{compliant}}} x_i}{N_{\text{compliant}}} & \sigma_1^{2*} &= \frac{\sum_{i\in\mathcal{C}_{\text{compliant}}} \left(x_i-\mu_1\right)^2}{N_{\text{compliant}}} \\ \mu_2^* &= \frac{\sum_{i\in\mathcal{C}_{\text{non-compliant}}} \log\left(x_i-0.1\right)}{N_{\text{non-compliant}}} & \sigma_2^{2*} &= \frac{\sum_{i\in\mathcal{C}_{\text{non-compliant}}} \left(\log\left(x_i-0.1\right)-\mu_2\right)^2}{N_{\text{non-compliant}}}\end{aligned} \tag{B.16}$$

Here $N_{\text{compliant}}$ denotes the total number of compliant points. It is clear that μ_1 and σ_1 are determined using points in the compliant class, while μ_2 and σ_2 are determined with the use of non-compliant points. All points i with $x_i \leq 0.1$ are assigned to the compliant class.

Instead of using a hard classification, the responsibilities given in equation 3.14 could be used. They are the posterior probability of a point belonging to either compliant or non-compliant class. In the case of $x_i < 0.1$ every point is classified to the normal mode, since the lognormal mode has support $x_i \in (0.1, \infty)$.

The maximum likelihood estimator becomes thus a weighted mean or standard deviation. Points with a high probability for the compliant class will influence μ_1 larger than points with a lower probability. The N_k in equation B.16 are now changed to the effective total points in a class, which is given for class k by $\sum_{i=1}^{N} \gamma_{i,k}$. Finally, the MLEs becomes equal to the ones given in Equation 3.15.

Appendix C

Parameters of the bimodal distribution during EM-algorithm

In Figure C.1 the parameters are shown as they are updated in each iteration of the EM-algorithm. After 50 iterations the parameters became stable.

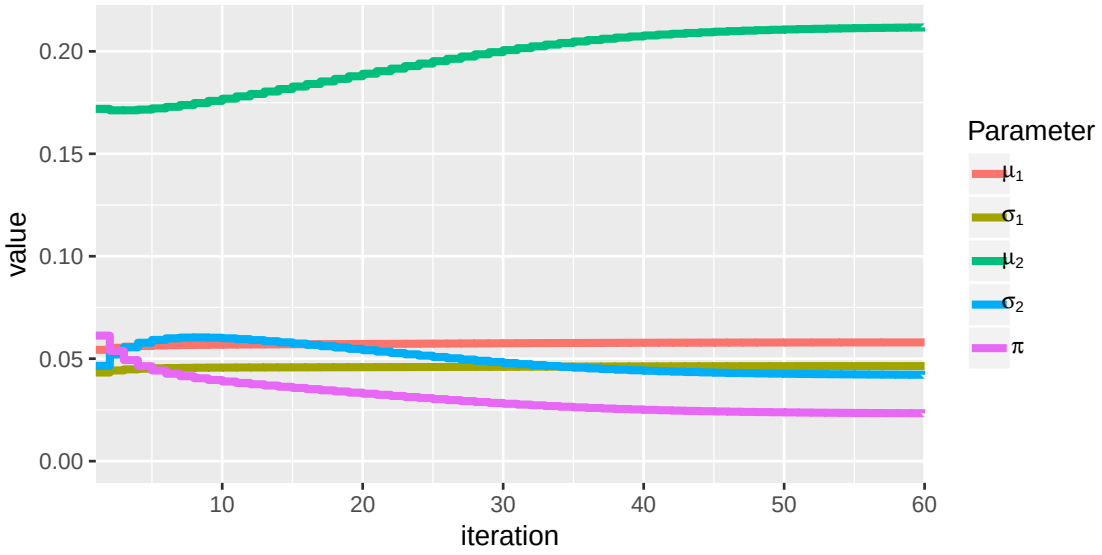

Figure C.1: The value of the parameters during the iterations of the EM-algorithm.

Appendix D

R Code for fitting the bimodal normal-lognormal distribution with EM

```
library(dplyr)
library(readxl)
library(ggplot2)
library(MASS)
Ships <- read_excel("~/Ships.xlsx")
Measurement <- read_excel("~/Measurements.xlsx", guess_max = 8049)
Measurement$Manual <- ((0.232 * Measurement$SO2) / Measurement$CO2)
compliant <- Measurement %>% tbl_df %>% filter(!is.na(Compliant)) %>%
  pull(Compliant) %>% as.logical()
FSC       <- Measurement %>% tbl_df %>% filter(!is.na(Compliant))
%>% pull(FSC)
FSC_SD    <- Measurement %>% tbl_df %>% filter(!is.na(Compliant))
%>% pull(FSC_SD)
agency    <- Measurement %>% tbl_df %>% filter(!is.na(Compliant))
%>% pull(Agency)

# Initial parameters
d <- data.frame(compliant, FSC, FSC_SD, agency)
i <- list(n = length(FSC), mu = mean(FSC), sd = sd(FSC),
n_c = sum(compliant), mu_c = mean(FSC[compliant]),
sd_c = sd(FSC[compliant]), n_nc = sum(!compliant),
mu_nc = mean(FSC[!compliant]), sd_nc = sd(FSC[!compliant]),
p_c = sum(compliant) / length(FSC), p_nc = sum(!compliant) / length(FSC),
max_c = max(FSC[compliant]), min_nc = min(FSC[!compliant]))
f  <- coef(fitdistr(d$FSC[!d$compliant] - .1, "log-normal"))
m2 <- f[['meanlog']]
s2 <- f[['sdlog']]
pi <- i$p_nc
o  <- .1 # Offset
m1 <- i$mu_c
```

```
s1 <- i$sd_c
rm(d, i, f)

# Used functions
dnorm2 <- function(x, m, s, f){f * dnorm(x, m, s)}
# Returns f * dnorm(x, m, s)
dlnorm2 <- function(x, m, s, o, f){
  x[x <= o] <- 0
  x[x > o] <- f * dlnorm(x[x > o] - o, m, s)
  return(x)} # Returns f * dlnorm(x, m, s) (and 0 if x<0)

ln2 <- function(x){
  x[x <= o] <- 0
  x[x > o] <- log(x[x > o])
  return(x)
} # Adapted in order to prevent warning from negative x in log(x)
E <- function(y, m1, s1, m2, s2, pi){
  a <- dlnorm2(y, m2, s2, o, pi)
  return(a / (a + dnorm2(y, m1, s1, 1 - pi)))} #Calculate responsibilities
plot <- function(m1, s1, m2, s2, pi){
  b <- 0
  e <- .5
  n <- 1000
  x <- data.frame(x = seq(from = b, to = e, length.out = n))
  tot <- function(x, m1, s1, m2, s2, pi, o){
    dnorm2(x, m1, s1, 1 - pi) + dlnorm2(x, m2, s2, o, pi)}

  ggplot(data.frame(d), aes(d)) +
    geom_histogram(aes(y = ..density..), binwidth = .01) +
    stat_function(aes(x), x, "line", fun = dnorm2, n = n, color = "green",
size = 1.2, args = list(m = m1, s = s1,  f = 1 - pi)) +
    stat_function(aes(x), x, "line", fun = dlnorm2, n = n, color = "red",
      size = 1.2, args = list(m = m2, s = s2, o = .1, f = pi)) +
    stat_function(aes(x), x, "line", fun = tot, n = n, linetype = "dotted",
      size = 1.2, args = list(m1 = m1, s1 = s1, m2 = m2, s2 = s2,
      pi = pi, o = o)) +
    coord_cartesian(c(b, e)) +
    xlab("FSC (% m/m)")
}

# Maximum likelihood estimators
ml_m1 <- function(y, g){sum((1 - g) * y) / sum(1 - g)}
ml_s1 <- function(y, g, m1){sqrt(sum((1 - g) * (y - m1)^2 / sum(1 - g)))}
ml_m2 <- function(y, g, o){sum(g * ln2(y - o)) / sum(g)}
ml_s2 <- function(y, g, m2, o){sqrt(sum(g * (ln2(y - o) - m2)^2) /sum(g))}

# Data import
d  <- Measurements %>% tbl_df %>% filter(FSC > 0) %>% pull(FSC)
```

```
# Iterate over EM-algorithm
print(plot(m1, s1, m2, s2, pi))
for (i in seq(from = 1, to = 5)){
  g <- E(d, m1, s1, m2, s2, pi)
  pi <- sum(g) / length(d)
  m1 <- ml_m1(d, g)
  s1 <- ml_s1(d, g, m1)
  m2 <- ml_m2(d, g, o)
  s2 <- ml_s2(d, g, m2, o)
  print(plot(m1, s1, m2, s2, pi))
}
```

Appendix E

Validation

Validation is required to know the robustness and accuracy of the used measurement set-ups. It can be performed by measuring a certified reference. Since in the case of fuel bunkers there is not really a certified reference available, good cross references were sought. In Table E.1, only cross references with ILT were used. This is because ILT boarded the ship, took a sample and analyzed it either directly with a handheld XRF or sent it to the laboratory. In the latter case also XRF is used. The laboratory had received an ISO 8217:2017 certification. The handheld XRF was validated using the same samples sent to the laboratory.
It happened sometimes that ILT measured multiple engines. These measurements are shown as duplicates.

Table E.1: Cross-validated measurements

Measurement by operator				On-board inspection	
Date	Operator	FSC (wt.%)	Label[1]	Date	FSC (wt.%)
2016-08-12	TNO	0.1±0.021	Compliant	2016-08-11	0.093
2016-11-17	TNO	0.06±1.34	Compliant	2016-11-16	0.092
2016-09-27	TNO	0.29±0.07	Non-compliant	2016-09-27	0.098
2016-09-11	Explicit	0.14±0.03		2016-09-13	0.102
2016-11-18	TNO	0.03±0.88	Compliant	2016-11-17	0.034
2016-11-22	TNO	0.17±0.03	Non-compliant	2016-11-23	0.079
2016-08-13	TNO	0.03±0.01	Compliant	2016-08-11	0.062
2016-09-10	TNO	0.02±0.01	Compliant	2016-09-13	0.15[2]
2016-09-10	Explicit	1.06±0.15			
2016-08-22	TNO	0.14±0.02	Non-compliant	2016-08-23	0.092
2016-11-19	TNO	0.26±0.04	Non-compliant	2016-11-21	0.074
2016-11-17	TNO	1±8	Non-compliant	2016-11-18	2.27
2016-07-20	TNO	0.53±0.11	Non-compliant	2016-07-21	0.068
2016-12-21	TNO	0.04±0.01	Compliant	2016-12-20	0.078
2016-11-17	TNO	0.03±0.46	Compliant	2016-11-18	0.16
2016-12-07	TNO	0.1±0.02	Compliant	2016-12-08	0.067
2016-08-20	TNO	0.15±0.03	Non-compliant	2016-08-22	0.101
2016-07-10	TNO	0.16±0.03	Non-compliant	2016-07-11	0.096[3]
2016-07-12	TNO	0.14±0.03	Non-compliant		
2016-11-19	TNO	0.05±0.01	Compliant	2016-11-17	0.036

[1] As given by the operator
[2] Same ship measured by different operators.
[3] Same ship measured two times.

Appendix F

Mismatch in calculated fuel sulfur content

Some operators such as BSH/ University of Bremen and TNO reported both the SO_2, CO_2 and FSC. While TNO reported values of FSC that were according to Equation 2.1, BSH/ University of Bremen did not. In the data of BSH/ University of Bremen, about 7 % of the measurements contained a negative SO_2. However, a negative FSC was reported no single time. A NOx correction could be a good explanation for this discrepancy. But Figure F.1 shows no relation between the mismatch between the calculated and reported FSC and the NO or NO_2 values.
Since the deviations are minor and occurs only under 0.1 wt.%, the deviations were ignored. Further research at BSH/ University of Bremen, could be useful.

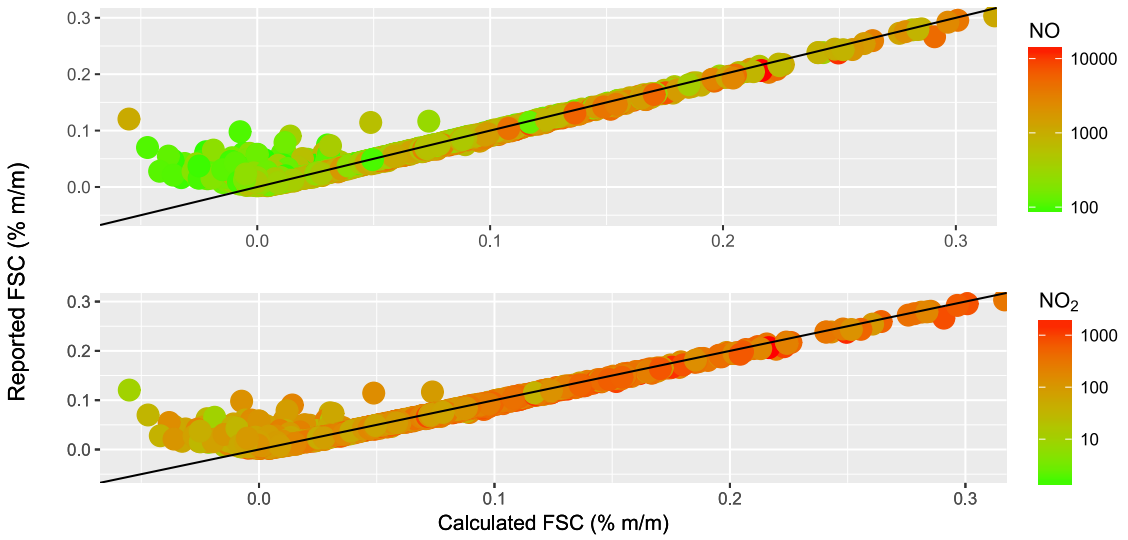

Figure F.1: Mismatch between reported and calculated FSC in the measurements from BSH. There is no clear relation with the reported NO, NO_2 values. Data source: BSH/ University of Bremen