Creating a bias in inspection data

Exploring the medium- to long-term effects of data-driven risk-based regulation

16 August 2019

Master thesis

MSc Engineering and Policy Analysis

Ivo Sedee

Faculty of Technology, Policy and Management

TU Delft, the Netherlands

Creating a bias in inspection data

Exploring the medium- to long-term effects of data-driven risk-based regulation

Master thesis submitted to Delft University of Technology

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in **Engineering and Policy Analysis**

Faculty of Technology, Policy and Management

by

Ivo Rikardo Sedee

Student number: 4396944

To be defended in public on August 30, 2019

Graduation committee

Chairperson	: Dr. S. Cunningham,	Policy Analysis
First Supervisor	: Dr. H.G. van der Voort,	Organisation & Governance
External Supervisor	: Ir. T.J.P. Booijink,	Dutch Authority on Food and Safety
External Supervisor	: Dr. E.E. van der Vaart,	Dutch Authority on Food and Safety

Executive Summary

Monitoring organisations are under pressure to prevent incidents, perform quality control and assess the risks and the integrity of firms and projects in a wide range of sectors. These tasks require a lot of capacity from monitoring organisations which means their resources have to be allocated efficiently. These organisations are currently in a transition phase towards risk-based inspections using data models and algorithms in order to increase their efficiency and transparency.

The use of data by monitoring organisations can have several benefits. The first sought benefit of data-driven inspections is an increase in efficiency. The use of data science has the potential to identify companies with higher risks of violating the rules in order to better allocate inspectors and other resources. Next to this, working data-driven allows monitoring organisations to justify their inspection process to the outside world. This way they can objectively show which steps were taken for choosing inspection locations. This is needed when they are questioned by governmental organisations or the public. Lastly, working data-driven creates the possibility for monitoring organisations to send their inspectors based on outcomes achieved by their inspections, in order to more efficiently allocate their resources.

The use of data by monitoring organisations, however, can also have negative effects in the medium- to long-term. Most importantly, their data will get biased when they mostly collect data from companies the risk models initially identified as having a higher risk of violating. This bias will create several issues in the long-term. Firstly, the monitoring organisations will have a skewed overview of their sector, leaning towards a specific group of companies their models identified as risky in early stages. Secondly, their perceived efficiency and compliance rates will be wrongly assessed. This can lead to a false sense of efficiency and undetected malpractice. Thirdly, their data will not have the quality needed to be used for justification and self-assessment. This created bias could also lead to a decrease in the predictive power of their models and a slower reaction to changes in risk indicators or the inspected companies.

Next to the creation of a bias in inspection data, performing data-driven inspections can increase the strategic behaviour from the involved companies. When monitoring organisations use transparent and structured models to select their inspection targets, companies could use this to avoid inspections or be better prepared. To add to these potential problems, data-driven risk-based inspections could become unfair or even discriminatory when the data from monitoring organisations gets biased.

The objective of this study is to explore the medium- to long-term effects of the use of data models and algorithms for risk-based regulation. The main research question this study will answer is:

How should monitoring organisations use data-driven risk-based inspections in order to utilize the long-term positive effects of the use of data while minimizing their long-term negative effects?

In order to study the medium- to long-term effects of data-driven inspections an agent-based model was created in NetLogo. Combinations of policy options were tested under a wide range of possible future scenarios. This study showed how these models can be used for research

about inspections and monitoring organisations. We confirmed the potential increase in efficiency the use of data can offer for monitoring organisations. At the same time, however, we confirm the concerns from other studies by showing the long-term risks of data-driven inspections. This study provides evidence for claims about the long-term effects of data-driven monitoring created with the use of an agent-based model and advanced analysis techniques.

This study generated numerous insights about the medium- to long-term effects of data-driven risk-based regulation. The effects of data-driven inspections vary widely based on the sector they are performed in. The sectors' most defining variables are the number of companies, the number of available inspectors and the incentives for companies to violate. Our results show that data-driven inspections have the most potential in large sectors with a relatively low overall compliance rate.

The results of the performed experiments show that the use of data for inspections has potential, but should hardly be considered without the addition of random inspections. Performing all inspections fully data-driven will have a large negative impact on all of the previously mentioned long-term issues, especially when they are performed with no or very few additional random inspections. They will reduce the monitoring organisations' data quality and overview of their sectors and remove the narrative needed for justification and self-learning.

Our research shows there is a point, or range, where the percentage of data-driven inspections starts to have a large negative impact on the data quality. Before this, the negative long-term effects are relatively small. The positive effects of data-driven inspections are more linear and do not seem to suddenly increase after a certain percentage of inspections is performed with the use of data.

The use of data by monitoring organisations with the goal to improve efficiency is dangerous. As mentioned, the use of data has several long-term benefits. However, increasing efficiency should not be the main goal for data-driven regulation. Our study shows the negative long-term effects of having efficiency as the main target. Working data-driven is best not combined with performing fewer inspections. The combination of fewer inspections and data-driven risk-based regulation has a large negative impact on the bias in the collected data and the overview monitoring organisations have of their sector.

This study also found an increase in strategic behaviour when performing data-driven inspections. Our model showed an increase in strategic behaviour even when the transparency of monitoring organisations is not taken into account. Additionally, our results indicate that monitoring organisations more often overestimate than underestimate the compliance rates of companies, but both can happen, especially when performing a high percentage of inspections with the use of data.

Contents

1. Introduction

Non-compliance with regulations is causing a lot of damage to society (Cliff, Wall-Parker, Cliff, & Wall-Parker, 2017). When this is brought to light the media often speaks about government failure or social service organisations that allowed for criminal behaviour. Monitoring organisations are held responsible for preventing these kinds of activities. These organisations are also responsible for quality control and assessing the risks and the integrity of firms and projects. Despite having these organisations who verify that strict regulations are being followed there still is much financial malpractice and fraudulent behaviour, even with many other bystanders around as well.

This can mostly be assigned to the lack of resources from monitoring organisations (Shimshack, 2014). They have a limited number of inspectors and an almost endless number of possible offenders. It is impossible for them to address all risks with the same amount of manpower. This is where they have to make informed choices about their focus of attention. Traditionally these choices are made by expert inspectors based on experience and instinct. With the current amount of available data these monitoring organisation could make more well-founded choices and reason about them more structurally. Data can be used to support the experts in choosing where to go, help understand why experts do what they do and translate their tacit knowledge to data or take on an even stronger role and tell the inspectors where to look (Provost & Fawcett, 2013). This will also allow the monitoring organisations to justify their behaviour to the outside world and explain why they perhaps missed some cases of fraud. This can be achieved with the use of data by creating a more structured and transparent way of performing inspections. One can argue that companies with insufficient compliance should get inspected more often than companies with good compliance.

The use of data in this kind of risk assessment, however, has consequences in the medium- to long-term. The data owned and used by the monitoring organisations will get biased if they add data collected through inspections selected by the same data. This biased data can lead to a skewed overview of the sector leaning towards companies that were identified as risky in early stages. This can lead to a false sense of efficiency and undetected malpractice in other places. Monitoring organisations might get locked in on two or three risk factors and miss out on other indicators or fail to respond to changes within the sector.

To add to this problem, companies could intentionally “clean up” these risk indicators if they are known in an attempt to avoid inspections. When monitoring organisations use clear and structured models to make their selections in order to justify their choices to the outside world this could become a big problem where malicious companies have a lot of knowledge about the way the monitoring organisations choose their targets. The main research question this study will answer is:

How should monitoring organisations use data-driven risk-based inspections in order to utilize the long-term positive effects of the use of data while minimizing their long-term negative effects?

1.1 Thesis Outline

Chapter 2 of this thesis will provide a review of the available literature on this topic and will use this to identify the knowledge gaps. From these knowledge gaps a research question will be formulated. This question and sub questions that need to be answered to answer the main question will be provided in chapter 3. Chapter 4 will show a research overview and explain the used methodology. In chapter 5 relevant theory will be provided which is used during the modelling phase and to argue for certain choices and recommendations. Chapter 6 shows the conceptual model and explains the most important design choices and their impact on the model. The model implementation is provided in chapter 7. Verification and validation are discussed in chapter 8. In chapter 9 the performed experiments are given and argued for. The next chapter, chapter 10, will discuss the results generated by the experiments. Conclusions and recommendations are given in chapter 11. The last two chapters discuss further research and reflect back on the use of the created model.

2. Literature Review

To identify the core concepts and main lack of knowledge in this research area a literature study was performed. The first step in this process was to find as much literature as possible by using keywords such as: risk, data-driven, model, regulation, inspection, monitoring and knowledge. The first set of papers was found using Scopus and Google Scholar. After a filtering process the snowball method was used together with suggestions by reference managers, Scopus and Google Scholar to find even more literature.

The main use for this initial literature study is to find where knowledge is lacking and to place this study within its scientific context. The literature that is provided here, however, is also used in further stages of this research paper, including the modelling phase, for recommendations and the reflection. Additional resources were added to support the modelling process which can be found in chapter 5 and 6.

2.1 Data-driven Decision-making

Organisations, including monitoring organisations, have realized there is much to be gained by hiring data scientists. But what exactly is data science and what does a data scientist do? According to Provost and Fawcett (2013) data science is a concept intertwined with other important concepts such as big data and data-driven decision-making. They argue that defining the boundaries of data science is important in order to understand its relationships to other important concepts and to explain what data science has to offer. In their paper they explain that the data scientists try to view business problems from a data perspective and that data visualization methods are vital in their success. They expand on this by saying that a data scientist combined with intuition, creativity, common sense and knowledge of an application can provide practitioners with structure and principles to help them solve their problems in a systematic way. Data-driven decision-making relates to this by making decisions based on data analysis rather than intuition alone.

Our study tries to further explore the use of data by practitioners and the effects it can have on the medium- to long-term. If monitoring organisations want to use intuition, creativity and knowledge of an application when working with data it is important to first create knowledge and get insights about how the system they are working in can behave when they start working with data-driven inspections. This paper will study the medium- to long-term effects in order to facilitate this and make the use of data more effective, while at the same time keeping the current workflow of inspectors in mind. New insights and an overview of the consequences of data-driven monitoring could also help in getting more support from the inspectors.

2.2 Benefits of Data-driven Inspections

Because of their limited capacity and very large list of potential malpractioners, monitoring organisations have to be smart in where they perform their inspections in order to get the most out of their resources (Shimshack, 2014). Multiple researchers show that the use of data and risk models for decision-making can improve this efficiency (Divan, 2018; Hutter, 2005). Hutter (2005) adds to this that risk models can be used to objectively show the steps that were taken to manage risks. This allows monitoring organisations to increase their transparency and opens up possibilities for justification of their inspection choices to the outside world. For them this is an important opportunity because of the political and legal risks they are facing. The increase in efficiency and ability to justify are the main arguments for data-driven inspections by monitoring organisations. Therefore, it is important we further explore these positive effects and see if the sought increases in efficiency described in this literature are realistic.

The collection and use of data also creates the option for outcome-based regulation instead of assigning a number of inspection hours as mentioned by an expert from the Dutch Authority on Food and Safety (NVWA). This way a certain amount of inspections can be performed with a defined goal and the number of assigned inspectors can be argued for and dynamically changed. Carrying out inspections based on outcome requires that monitoring organisations have knowledge about the state of the sector and the results of their inspections. This means they will need clear and unbiased data from which they can extract this knowledge and should try to avoid working from a source containing only the riskier organisations. Our study will look at this potential bias monitoring organisations can have about the compliance of the companies in their sector and how this influences the ability to work based on outcomes.

Research further supports the idea of improving the efficiency of regulators by stating that regulatory policy that involve consistent inspections that include a cooperative or educational component in the industry is the only significant way to tackle corporate offending (Schell-Busey, Simpson, Rorie, & Alper, 2016).

Baldwin and Cave (1999) support the claim that greater transparency can be reached with the use of statistical and accounting techniques. They also support the idea of higher efficiency and fairness by better coordination instead of by semi-random inspections performed by inspectors without the use of data.

2.3 Creation of a Data Bias

Measuring this mentioned efficiency, however, is very difficult in the inspection and monitoring sector (Burnett, Carthey, & Vincent, 2013). This is an even larger issue when the same data is used for this reflection as for the risk-analysis. This will create a bias in the data source. The data might show “good” results but will fail to warn inspectors when they are missing things that are not in the data. The creation and effects of this bias and the difference in observed and real efficiency needs to be further explored.

This is a big issue when data is collected during inspections. Adding only data from inspections means the monitoring organisations will have less data or no data at all about companies that were initially not identified as risky by the risk models (Jacobusse & Veenman, 2016). This will give monitoring organisations a skewed view of their efficiency, thinking they are more efficient than they might be in reality.

When the data owned by the monitoring organisations gets supplemented with inspection data collected during data-driven inspections it will become difficult to get a neutral view on the system as a whole. In this case the monitoring organisations will get their own view of efficiency that differs from the system's perspective. The creation of a bias in the data of monitoring organisation will also greatly reduce their possibilities for justification by means of data. Human and political factors also add bias to this data, this is further explored in the next subchapter.

2.4 Combining Inspector Knowledge with Risk-models

Sparrow (2000) coins the idea of a "new craftsmanship" where regulators have an overview of risks which can show them which businesses are more likely to misbehave. The inspectors will have to interpret a list of risk factors, which is different from their current way of working and might prove difficult and counterintuitive. This is a mix between fully data-driven inspections where the data tells the inspectors where to go, random inspections and risk-based inspections based on tacit knowledge and expertise from the inspectors. This mix will require support and cooperation from the inspectors to start working with data. Models can help in gaining this support from inspectors and opening up a discussion. As mentioned earlier, our study will try to further explore the use of data by practitioners and the effects it has on the medium- to long-term.

This mismatch between data scientists and inspectors can perhaps best be described by Lipsky's (2010) famous notion of street level bureaucrats. These civil servants, the inspectors in this case, who have regular direct contact with citizens often use discretionary decision-making and have a certain amount of freedom from their management. The data scientists, however, are not in direct contact with citizens and have more of a management role where they develop processes to select and evaluate inspections. The combination of these perspectives can prove difficult but very useful and even necessary according to Alexander and Judy (1988), which state domain-specific knowledge is needed to efficiently use strategic knowledge. This difference relates to the often-discussed notion of tacit versus explicit information (Albino, Garavelli, & Gorgoglione, 2004). This shows the use of data science in inspections does not mean the inspectors can be replaced which is also argued by Brown and Duguid (2000).

Baldwin and Black (2016) state that regulators make their choices not only based on facts but also on political and ethical norms. They divide this in two clearly different methods of defining risk (risk-based and problem-based) and explain the importance of understanding why regulators make the choices they make. Ezingeard, Leigh and Chandler-Wilde (2000) argue with the help of a case study that models and data can help structure this decision-making process and transition between tacit and explicit knowledge. Nonaka, Takeuchi and Umemoto (1996) have implemented a framework for creating organisational knowledge from tacit knowledge that for example the inspectors have. This framework also tries to transfer knowledge the other way around but this does, however, not work well with big data and therefore different solutions are needed here.

This literature can be used as the basis for data science in monitoring organisations. However, it does not look further into the long-term effects of data science for these organisations and the impact on the quality of available data. This literature study shows that there is more literature available on the relation of inspectors and data science than there is on the use of data for inspections in general. While this literature is important, it is currently equally important that the long effects of the use of data are explored, even for closing the mentioned gap between inspectors and data scientists. It also highlights how inspectors already add a bias based on their own knowledge to the inspections they perform and the data that is collected.

2.5 Medium- to Long-term Effects of Implementation

The mismatch between data science and inspectors can have an effect on the collection and use of data by monitoring organisations, as it shows the data cannot be used for completely objective decision-making. How monitoring organisations choose to implement the use of risk models will have a large impact on the long-term quality of their data.

Human and political factors will influence this process and also affect which data is collected. This increases the uncertainty about the long-term effects of the use of data science and the collection of data. The aforementioned literature also highlights the need for a balance between data-driven risk-analysis and the traditional approach which is more problem-based. Literature about this balance and the impact it has within the inspection sector and on the quality of the data owned by monitoring organisations is lacking.

The balance between risk-based and random inspections also plays an important role in ensuring high quality data. Literature is still lacking here as well. Monitoring organisations should perform random inspections next to their data-driven inspections to fight some of the aforementioned issues, such as the creation of a bias in their own data which skews their attention and limits the available data suited for feedback and reflection (Jacobusse & Veenman, 2016). Knowledge about this balance is needed to ensure good practice, especially when monitoring organisations are under much pressure to perform. How many random inspections should they perform and how do these impact the other outcomes the monitoring organisations are interested in? A balance needs to be found between the positive and negative long-term effects of data-driven inspections.

2.6 Additional Issues and Missing Knowledge

Creating a risk model that facilitates the combination of inspector knowledge with that of data comes with issues and trade-offs. A trade-off will have to be made between an understandable, interpretable model that can fit many situations or a more complex model that includes complex behaviours better but might be difficult to interpret or apply to every case (Myung, 2000; Raick, Soetaert, & Grégoire, 2006). For this trade-off, knowledge about the effects of using a risk model on the quality of decision-making is required, which leads us back to the issue of measuring efficiency with a biased dataset.

Another issue relates to the transparency of the model. When a risk model is used as discussed earlier, to justify choices made by monitoring organisations, it has to be an open, structured and easy to understand model. Other actors might abuse this and try to “game” the system if the algorithm is known (de Laat, 2017). The effects and dynamics of this issue should be identified to ensure good decision-making in the long-term. By modelling the behaviour of inspected companies we can get insights about their strategic behaviour.

According to Coglianese and Sapir (2017) the use of a risk model will require much more regulation and guidance to avoid ad hoc decision-making as they show in many cases. This also increases the existing bias in the data even more. More research is needed to help structure this required regulation and guidance. This study will provide knowledge about the effects of possible data-driven strategies which is required to create a structured way of working with data.

A bias in the data of monitoring organisation can also warrant a critical look at the types of modelling techniques they are using for their risk assessment, since some are better suited to deal with biased data than others (C. Liu, White, & Newell, 2013). Some modelling techniques could, for example, only look at the data where malpractice is identified and deal with it as such in order to negate the negative effects of a bias. This does require insights about the size of the bias and the available information, which is currently missing. This study seeks to contribute in this matter by identifying how large potential biases can get for different ways of data-driven monitoring.

There is, however, also literature that mentions another domain of issues that arises with this use of data. The issue of profiling and predictive policing is one that is much discussed in literature (Ferguson, 2012; Jackson & Bekerian, 1997). Hajian, Bonchi and Castillo (2016) add to this the statement that models can discriminate even if they were not created with this intent. This happens when they use a discriminatory data source. When monitoring organisations only inspect "risky" businesses their source will become biased and discriminatory issues will be even more problematic. Will companies receive fair treatment? Some companies could have to deal with relatively more of the administrative burdens for no fair reasons. This, combined with the aforementioned issues, calls for a study on the medium- to long-term effects of data-driven decision-making by monitoring organisations.

3. Lack of Insight and Research Questions

This literature review shows that this is an active field of research trying to merge with practice at the moment. Numerous issues have been mentioned in literature and solutions have been proposed. The perfect solution or an implementation is, however, often still lacking. A linking factor in the found literature is that a biased data source will prevent the monitoring organisations from achieving their goals for various reasons. This issue can become an important issue for monitoring organisations that collect their data from their own inspections, which has not been addressed in literature before. Potential medium- to long-term effects discussed in the literature are:

- An increase in the efficiency of inspections;
- The ability to send inspectors based on outcomes;
- The ability to objectively show which steps were taken in selection procedures;
- The creation of a bias in the data;
  - Limited access to data for reflection and justification;
  - Skewed view of efficiency and compliance rates;
  - Decrease in predictive power;
  - Slower reaction to changes in risk indicators;
- An increase in strategic behaviour from involved companies;
- Unfair / discriminatory inspections, or fairer inspections compared to risk-based inspections without the use of data.

Therefore, the main research question will be:

How should monitoring organisations use data-driven risk-based inspections in order to utilize the long-term positive effects of the use of data while minimizing their long-term negative effects?

Sub questions that need to be answered in order to answer this main question are:

- What is the effect of data-driven inspections on the quality of the used data?
- How should random inspections be used in relation to fully data-driven risk-based supervision in order to utilize the positive effects of the use of data while minimizing the long-term negative effects?
- How much does the efficiency perceived by monitoring organisations differentiate from the real efficiency?
- How fair are data-driven inspections to the companies involved regarding administrative burdens?
- What is the effect of transparency on the strategic behaviour of companies?
- How can inspectors be directed based on an objective outcome instead of input (hours) and what will be the consequences of this?

3.1 Dutch Authority on Food and Safety

To answer these questions this study will be performed at the data science department of an inspection agency, the Dutch Authority on Food and Safety (NVWA). For this study a model will be developed for exploring the medium- to long-term effects which are explained in the previous chapter. The model and its results will be verified using a real-life case and expert opinions within this data science department at the NVWA. Doing this research within a data science team within a monitoring organisation ensures the scope is relevant and all important concepts and relations are included. The NVWA is a good fit for this study because they have experience with data-driven inspections and are now starting to think about the longer-term effects this will have. They work with a separate data science department and the tools that are developed there will support a wide range of sectors. Each of these sectors is very different. This will increase the ability for us to generalise our findings outside of this study as multiple sectors are being monitored here.

3.2 Scientific Relevance

How to combine expert knowledge with data science is being widely studied and researchers show there is room for improvement and this is still a difficult task which we can contribute to. The long-term effects of this way of working and the use of data for selecting inspection targets, however, has not been studied before. This is where this study will have the largest impact. Inspection games, which will be explained in the next chapter, have been studied for a long time. The use of data science by inspection agencies is a new addition to these games. This has a large impact on the game and has never been studied before. This study also looks at these games over a longer period of time. The implementation of an inspection game in an agent-based model can also offer new insights. This has been done before but with very simple rules. This research seeks to expand the strategic behaviour of both the monitoring organisations and companies in order to generate new knowledge. The use of exploratory modelling and analysis (EMA) is also new in this field of research and offers tools to work with uncertainties in the model and its outcomes. Next to the use of the exploratory modelling, this study uses a variance-based sensitivity analysis method in order to partition the relevance of certain system aspects and policy options. This is not yet standard practice and offers some additional benefits, next to the one factor at a time sensitivity analysis, which are explained in chapter 9. Furthermore, we will be testing hypotheses mentioned in the literature and see if these effects indeed occur under our assumptions.

4. Research Approach

In this chapter the selected research approach will be presented and argued for. First an overview of the approach will be given. After this the pros and cons of this type of research will be explained. Lastly, an additional benefit of this approach, further uses for the created model will be given.

4.1 Agent-based Modelling Combined with Game Theory

The identified questions seem to be approachable with the use of an agent-based model (ABM). An ABM can be used to explore the medium- to long-term effects of different data-driven policy implementations on the various aspects in the system such as the quality of the data and the (perceived) efficiency of inspectors. Building such a model on game theory can provide insights about the relation and interaction between inspectors and businesses. The next chapter will elaborate more on the use of game theory for the creation of our model.

The modelling phase of this study involves the use of literature and experts to gain knowledge about the system. This part can already provide insights which can help us in answering the research questions by providing a structured way of inventorying available knowledge into a model. After the model has been created, experiments can be performed to gain additional insights about the different policy options and included model variables.

ABM is used to model interactions between autonomous agents in an effort to research the system as a whole. It is often applied where traditional methods have to make strong assumptions and simplifications to offer solutions, making them distant from the real world (Smojver, 2012). In this case we are looking at inspectors and companies which are individual and autonomous agents driven by their own characteristics, environment and behaviour. This autonomous behaviour, however, leads to behaviour that is visible on the system’s level.

In this study we are interested in the effect of the use of data by inspectors over time. An agent-based model will be used to simulate how the available data can help inspectors with their decisions and the effect that that has on their efficiency and the quality of the data in the long-term.

Game theory is a good starting point for research about interactions and strategic decision-making (Smojver, 2012). In this case we are facing a specific class of game theory called inspection games (Andreozzii, 2010; Taylor & Avenhaus, 2010). In these inspection games the inspectors wish to stimulate good behaviour, and even more so wish to punish bad behaviour. The inspectees (companies) prefer to show bad behaviour if they do not get inspected. In case of an inspection they prefer to show good behaviour. In this game the inspectors have limited capacity. This means they cannot detect everything and this explains why they prefer to see bad behaviour during their inspections. In this game both participants are incentivised to continuously change their strategy. Modelling this with ABM allows us to add many inspectors and businesses with their own characteristics and include other real-world system aspects, such as information asymmetry and time dynamics, to stay closer to reality.

4.2 Model Description

The agent-based model will consist of a number of inspectors with limited capacity and a number of companies. The inspectors will estimate which companies have the highest chance of violation based on the available data. The companies will make a trade-off between their payoffs including an estimation of the chance they will get inspected, based on previous inspections.

The data that is available to inspectors for their risk-assessment will be updated after their visits. This means the data will only contain information about the companies that are visited. At the same time the companies have certain visible characteristics that make them more likely to violate. The data will show which characteristics lead to more violation and the inspector will act accordingly, therefore the inspection chance of a company that has not yet been visited can still be increased by the use of data. The companies have multiple different characteristics so it is possible for the data-driven inspectors to miss other signs of violation in this way and get stuck on only one indicator.

After the model has been built different scenarios and experiments are created. This way, we can explore the effects of variables such as the ratio between random and data-driven inspections on the data bias under different assumptions and scenarios.

With this simulation study we hope to provide insights related to the identified sub-questions in order to give an overview of the effects of the use of data for choosing inspection locations and to provide an answer to the main question. Additionally, the created model can support strategic decision-making and be used for creating a discussion between inspectors and management.

4.3 The Added Value of Modelling

Modelling and simulations as tools seem to be very applicable and useful for the identified questions. This is because we are interested in long-term effects and behaviour of different policy options. We wish to get insights about the development of existing systems into the future. Modelling such systems allows us to explore possible future scenarios and evaluate them. The modelling process allows for stakeholder engagement which ensures the study stays relevant and offers useful and convincing insights. This design process combines knowledge of experts, the modeller's environment and literature and feeds back gained insights to these areas. Creating a model for answering these research question has the additional benefit of being able to support decision-making. This way better grounded real-world decisions can be made regarding the implementation of data-driven strategies by showing decision-makers possible outcomes and trade-offs.

4.4 Pros and Cons of Agent-based Modelling

When deciding what kind of modelling technique to use it is important to look at what the techniques are good at and what they cannot do. Agent-based modelling was chosen as an approach because of several factors. Firstly, agent-based models are easy to grasp, even for those who are unfamiliar with the approach (Dam, Nikolic, & Lukszo, 2012). When stakeholders are involved, agent-based modelling allows for easy communication about the simulation model by including natural concepts. Agents and their relatively simple rules are easy to explain even though they can generate complex behaviour. Each agent represents a real entity in the real world. These agents are modelled in such a way as to behave in the same ways the real entities do in the real world. This will lead to recognizable behaviour for stakeholders which makes the model and approach easy to grasp.

A second benefit of agent-based models is that they can be quite transparent when supporting decision-makers. Decision-makers themselves can come up with strategies and interact with the model and verify if the model shows plausible behaviour. They can “play” with the different parameter settings and even learn by doing so, in contrast to modelling approaches that generate a set of outcomes in a black box.

A last advantage of agent-based modelling as an approach is its ability to deal with complexity. The real world is very complex, consisting of many interacting non-linear parts. Agent-based modelling allows us to model these interactions of social and technical systems in a natural way. This allows us to create an adaptive system where agents learn from themselves and the environment around them in a way that closely resembles the real-world dynamics.

Because agent-based modelling combines both social and technical elements, it requires the modeller to be multidisciplinary as well. The quality of the model depends on the modeller’s ability to combine complex phenomena and write them down in simple ways a computer can work with. This policy-relevant style of modelling simulations thus requires knowledge about many different system aspects and actor relations, but also about how to use the model for policy advice. When working with stakeholders it can become tempting to oversell one’s model. Experienced modellers are aware of the limitations and assumptions in their models. However, involved stakeholders often look for short and precise predictions about the future. Here the modeller has to be careful not to be tempted by the desires of their problem owner, finding a balance between too simple and too complicated answers while keeping the limitations in mind, in an attempt to give relevant policy advice.

4.5 Model Uses for Monitoring Organisations

The model used for this study is created with multiple purposes in mind. The first and most important one is to explore the long-term effects of data-driven monitoring as they are described in literature and mentioned by experts. With this model we will try to gain insights about the long-term effects of the use of data for risk-based inspections and generate scientific knowledge. This is done using various analysis techniques and by testing specific experiments and scenarios. By doing this we are hoping to find insights related to data-driven monitoring that can help us in supporting decision-makers when choosing their inspection strategies, with their data quality in mind.

Next to this, the modelling process and results will highlight where additional research is needed to better understand the system. In doing this we generate additional questions for further research which can be of added value to the monitoring industry. Thirdly, the model can help decision-makers by showing trade-offs between their outcomes of interest and show how their choices might influence these trade-offs. Creating an agent-based model allows stakeholders to interact with the model themselves to get a feel for the different complex relations between all the variables in the model. This can lastly also help structure their conversations with inspectors to ease a potential transition towards more data-driven inspections. Visually showing how strategies might affect outcomes can be convincing and offers a framework for discussion.

In this study the focus is on finding scientific insights that can at the same time help monitoring organisations with setting an effective and robust strategy with the long-term effects in mind. We will explore the impact of each of the variables on the outcomes of interest, analyse the effect of certain strategies under different scenarios and look at the trade-offs that have to be made between the positive and negative long-term effects of the use of data for risk-based inspections.

5. Theoretical Background

In this chapter additional relevant theory will be discussed. When creating a simulation model, the modeller has to be careful when making assumptions. The theory in this chapter will be used to support the model and the modelling process when making important choices and assumptions.

5.1 Game Theory

Game theory has been applied in the monitoring and inspection industry for a long time. Game theory is the study of mathematical models of strategic interaction between rational decision-makers (Kalai, 2004). It is used in order to describe, predict, explain and even prescribe behaviour using mathematical equilibria.

5.1.1 Inspection Games

Inspection games are a specific class of game theory games that have been widely used as a model for crime deterrence (Andreozzii, 2010). They are mathematical models of the situation where an inspector verifies wheter or not another agent adheres to the rules. In this situation the inspector prefers to catch misbehaving agents and the inspectee prefers to adhere to the rules when they get inspected and prefers not to adhere to the rules when they do not get inspected. Inspection games raised awareness when they were used to show that higher fines would not lead to lower crime rates, under certain assumptions of rationality. Instead inspectors would be inclined to inspect less, making them more efficient but not lowering the crime rate (Rauhut, 2015). These games are also used to get insights about the efficiency and the number of required inspectors. They help in finding a balance between a couple of very efficient inspectors and a larger number of less efficient inspectors with a lower overall crime rate (Taylor & Avenhaus, 2010). Conclusions drawn from inspection games are often strong and controversial, such as that of the effect of fines and bonuses on crime rate (Nosenzo, Offerman, Sefton, & Van Der Veen, 2014).

In this study inspection games and their payoff matrices are used as a framework to study the interaction between inspectors and companies and to develop these interactions into a computational simulation. Variables such as the height of a potential fine, the costs of compliance and estimated chances to get inspected are described by these theories and used in our simulation model.

5.1.2 Differences with Inspection Games

This subchapter explains why and how our study differentiates from the classic inspector games.

Traditionally inspection games are used to calculate and predict how many inspectors are needed to reach a certain balance or efficiency (Taylor & Avenhaus, 2010). The main focus of this research is, however, not on the capacity of the monitoring organisations but rather on the use of data and how this influences the other variables and mechanisms. The number of inspectors is still varied during different scenarios and tests to see how it interacts with the other factors and influences the results of the model, but the main focus is not on finding an optimal or efficient number of inspectors. During most of the model runs the number of inspectors will be set in line with a real-world case in an effort to create a balance with which the effect of the other variables and policy options can be evaluated.

Avenhaus (2004) showed that inspections games do not capture some real-world elements such as moral and more complex strategic behaviour. Since we are exploring the effects of the use of data over a longer time period, dynamic strategic behaviour is included with learning agents. The agents also have several reasons to violate and do not act fully rationally. This is an attempt to create a model that is closer to the real world than the traditional inspection games.

Avenhaus (2004) also showed how different assumptions in an inspector game will lead to different strategies and results. For this reason, it is important we are clear about our assumptions under which our model results hold. In an agent-based model we can simulate the effect of different assumptions and see how they impact our results. Andreozzi (2010) shows how strict some of the assumptions made in a traditional inspection game are. Firstly, the game must be played simultaneously, meaning each player chooses without knowing the choice made by the other player. Secondly, the players both know each other’s payoffs and are rational enough to calculate their equilibrium probabilities. This study will try to relieve some of these restrictions on the played inspection games.

The concept of leadership, where the inspector announces their strategy and commits to it is also applied in inspection games. This is a powerful concept in these games where inspectors can surprisingly gain an advantage and a higher payoff by playing first. This concept does, however, not hold for games with more than two players which we are exploring in this study (von Stengel & Zamir, 2010).

Perc & Szolnoki (2015) argue for a complex systems approach for studying inspector games. They say it is in particular because the complex system’s behaviour cannot be inferred from the simple actions of individual actors. They explain that the emergent perceived phenomena can be associated with complex social and biological systems. This claim is supported by other sources showing that malpractice is far from being uniformly distributed across space and time (Alves, Ribeiro, Lenzi, & Mendes, 2013; Picoli, Castillo-Mussot, Ribeiro, Lenzi, & Mendes, 2015). They argue that crime is recurrent and does not linearly scale with fines. The agent-based model used in this study strives to include these concepts and generate behaviour closer to the more complex real world by including different decision algorithms and a changing environment. This study will hopefully support the claim that these complex system approaches are useful for studying inspector games.

5.2 Principal-agent Theory

Concepts from principal-agent theory can also be used in the monitoring industry. The added value of principal-agent theory in comparison with agent-based theory lies in the relation between the inspectors and the inspected, which principal-agent theory offers additional knowledge about. Information asymmetry and misaligned incentives lead to monitoring and compliance costs. Companies might be “free-riding” when monitoring organisations simply expect them to comply with the rules. Principal-agent theory provides a framework for working with concepts like these and transaction costs without the use of more sophisticated game theory (Laffont & Martimort, 2018). These concepts help us with looking at real-world behaviour and translating it into a model. One could argue that the use of data reduces the principal-agent problem by reducing the information asymmetry. The focus of this research is broader than the principal-agent problem and will therefore just be using this theory to better implement some model mechanisms. Principal-agent theory is also too hierarchical in its nature to be used directly in inspector games. Information asymmetry is used in this study as the companies and inspectors will have a different view of the compliance level and have to make their own predictions about the behaviour of others. Principal-agent theory also explains why monitoring organisations in our model prefer to see non-compliant behaviour during their inspections. This is because of the monitoring costs they have to make. These monitoring costs are only efficiently spent if they result in a higher compliance rate and inspecting non-compliant companies gives inspectors a better chance of increasing the overall compliance rates.

5.3 Implications of Data-driven Decision-making

Issues with data analytics and data-driven decision-making have been widely studied before. These studies are about issues like privacy and data protection. Issues about the enabling role data-driven solutions offer as techno-regulation orders have not been studied as much (Bayamlıoğlu & Leenes, 2018). Techno-regulation refers to the intentional influencing of behaviour by embedding norms into technological systems (Bayamlıoğlu & Leenes, 2018). Currently these practices are growing fast and data-driven monitoring is an example.

Regulating, monitoring and influencing behaviour with the use of data has been around for a while. However, the way legislation is now being transformed to executable code is what is new (Bayamlıoğlu & Leenes, 2018). This is the case when algorithms start telling inspectors where to perform their inspections.

Bayamlıoğlu & Leenes (2018) state that these practices can lead to unjust discrimination. They warn against the legitimization of the “power of the code” in order to protect notions such as non-discrimination, fairness and privacy. Encompassing narratives are needed to go from correlations to causal attributes (Bollier, 2010). Our study will report the effect that the use of data has on the types of companies that get inspected. These insights can be further used to argue if data-driven monitoring can be seen as fairer or discriminatory. The creation or destruction of an encompassing narrative when working with data that is mentioned in this literature is one of the possible long-term effects that this study will provide insights about.

The rule of law requires the capability to contest decisions (Waldron, 2010). This requires that a person must be aware of the automated data-driven process and be able to foresee and understand its impacts (Routledge, 2013). This could be another factor forcing monitoring organisations to work more transparently and increase the potential for strategic behaviour for companies. We are aware of the potential risks of increasing the transparency of monitoring organisations and will consider these when drawing conclusions and making recommendations. These risks are, however, difficult to model explicitly and are for now left out of the model implementation.

6. Conceptual Model

In this chapter the conceptual model will be explained. This is the result of a modelling phase where knowledge from literature and experts were combined into a conceptual model. This chapter will try to explain all the relevant variables and relations of the model used for this study. The implementation of this conceptual model as a computational model in NetLogo is given in the next chapter.

The conceptual model combines insights from the discussed literature and will be explained in detail in the subchapters below. Our model assumes companies are able to change their strategy for the next inspection round based on their inspection history, as mentioned studies show a company's estimated chance to get inspected has an impact on their strategy (Baldwin & Cave, 1999; Kagan & Scholz, 2013; Schell-Busey et al., 2016). Inspectors will prefer to see non-compliant behaviour when they inspect companies, as described in inspection games theory and principal-agent theory (Andreozzii, 2010; Rauhut, 2015; Smojver, 2012). This is why they will search for high-risk companies with low compliance rates when working data driven. Here we assume there are different types of violators that violate at different points, for different reasons, based on a study by Kagan & Scholz (2013). When inspectors visit relatively more high-risk companies, they will collect relatively more data on those companies and an accurate prediction about the rest of the companies will be more difficult to make (Jacobusse & Veenman, 2016). Since companies adjust their strategy over the years, the collected data by the inspectors will also change in the long-term. This conceptual model with all its relations will be explained in more detail below. The model will be implemented and experiments will be performed in order to explore the medium- to long-term effects of data-driven risk-based regulation.

6.1 Time

Each run will simulate 100 rounds of inspections. Every timestep in the model will represent one round of inspections which means that that is the smallest time interval in the model and nothing is calculated between inspections.

6.2 Agents

The model consists of two groups of active agents interacting with each other: Companies and Inspectors. The individual inspectors collectively represent the larger agent of the monitoring organisation. In reality the monitoring organisation assigns the inspectors locations to visit. In the model this is done by the observer. The "orders" are given in policies set before each experiment.

The inspectors in this model have no individual memory but instead add knowledge to their shared data source, owned by the monitoring organisations. Because they do not have an individual memory or their own strategies but are controlled by the higher entity, the monitoring organisation, and can only perform one inspection per timestep, the number of inspectors is equal to the number of performed inspections. This means that if our model shows the effects of an increase on the number of inspectors this could also be interpreted as an increase in the number of inspections, or an increase in the inspection capacity of monitoring organisations.

6.2.1 Why and When do Companies Violate?

In order to model the behaviour of companies under supervision it is important to understand why and when companies break the rules. Kagan & Scholz (1980) distinguish between three groups of corporate offenders. These typologies were studied to understand and model the company agents. In addition to these three possible offenders, an additional group of companies is modelled, which will always show compliance with the rules, disregarding the size of a potential fine and other factors as these were mentioned by domain experts at the NVWA.

Amoral Calculators

The first group of possible offenders in the model are the amoral calculators. These companies base their strategy on their estimated profit. They will weigh their costs of compliance against a potential fine and act accordingly. Within this group, companies break the rules at different profit points. Each company has a propensity to violate. This is a number between 1 and 100. The higher the number, the more likely a company is to break the rules. At 100 it means the amoral calculator will stop complying with the rules as soon as the costs of compliance are higher than the estimated amount they will have to pay in fines. At a propensity to violate of 50 these companies will start breaking the rules when they estimate it to be twice as profitable compared to complying.

To make this profit estimation the companies estimate the chance that they will get inspected. We assume here that the companies have information on the number of inspectors and their capacity. When no other information is available to them, they believe the chance that they get inspected to be equal to the number of available inspectors divided by the number of companies. When a company gets caught, it will increase its estimated chance to get inspected. Step by step this returns back to the initial estimation when they are not getting inspected. If a company has not been inspected for a while, while it is at its initial expectation, it will reduce their expectation each year, increasing the chance they will break the rules.

Political Citizens

The second group Kagan and Scholtz identified are the political citizens. Political citizens, or principled objectors, seek to deliberately violate the rules they find immoral, regardless of their profit and a potential fine. Kagan and Scholz suggest negotiation as an enforcement method for these violations. Because the model assumes enforcement is done using deterrence, in the form of fines, these political citizens are assumed to show behaviour similar to the amoral calculators. Upon inspection they will increase their estimated chance to get inspected, decreasing the chance they will violate the rules. One could argue that this is the same behaviour they would show when they would have been negotiated with, only here the height of the fine and the costs of compliance would have influenced their strategy less. This is why they are represented by companies with a very high propensity to violate, making them less sensitive to fines, instead of being a completely separate entity.

Organisational Incompetent

The third and last group of potential offenders are the organisational incompetent. These companies start breaking the rules when the rules get too complicated for them. Complexity of the rules is assumed to be included in the costs of compliance variable. This means that for this study they show very similar behaviour as the two other groups where they violate based on the costs of compliance and their inspection history, which could be interpreted as educational visits. They have a relatively lower propensity to violate and are more sensitive to changes in the costs of compliance. These companies exist in the model when the parameters that allow companies with a low propensity to violate still show non-compliant behaviour. This is the case when there is a low number of companies that always comply with the rules and the minimum expectation to get inspected of companies is also low. This makes these companies less sensitive to financial incentives but more sensitive to inspection history.

Unique Characteristics

Each of these companies is unique and has different characteristics. The aforementioned propensity to violate is one of them. This variable is linked to a couple of other characteristics. These generic company characteristics are created after the propensity to violate is set and could be interpreted and/or replaced by any real-world characteristics such as company size or location. The generic characteristics have a correlation with the propensity to violate. This correlation is set at the start of the simulation. The inspectors try to find this correlation in order to best predict the propensity to violate of companies. The propensity to violate also has a random part which represents other possible reasons to violate or comply.

6.2.2 Where do Monitoring Organisations Send Their Inspectors?

Monitoring organisations are assumed to use different strategies for selecting their inspection locations. One of their options is to perform data-driven inspections. These will be based on information they have about the companies. The inspectors learn information about companies when they visit them. Based on the characteristics and the compliance of the inspected companies, the inspectors estimate the correlation between the characteristics and the propensity to violate. Using these correlations, the inspectors then make an estimate of each company’s propensity to violate. Inspectors combine this with their knowledge about the compliance of companies to select their targets when working data-driven.

A compliance score is generated based on the observed compliance of a company and the range of the estimated propensities to violate as to make sure the scores are measured on the same scale. Before running the simulation, the weight of each of these scores can be set. In this compliance score, companies which have never been inspected are set to have a compliance score equal to the average of the inspected companies.

The other option the monitoring organisation has is to send its inspectors to random locations. In real life inspectors might choose locations based on instinct and their own tacit knowledge. The model created for this study can represent this third option by setting the selecting process to non-random with a high weight to the estimated propensity to violate. This way the expertise inspectors offer is taken into account in the model. Because the tacit knowledge from inspectors is difficult to quantify in a model, this way of implementing their behaviour might undervalue their ability to identify high risk companies.

During the modelling phase, simulations showed inspectors get stuck when inspecting solely on the propensity to violate of companies. In reality, as discussed in the literature review, inspectors use more elaborate strategies which are very difficult to study, let alone model. This is why, during this study, we will compare data-driven inspections against random inspections in order to gain insights on the effects of the use of data. Explicitly modelling the expertise of inspectors and their own way of selecting inspection targets would require strong assumptions which would greatly reduce the legitimacy of our results. The expertise of inspectors is still assumed for our model in their ability to estimate the propensity to violate of companies upon inspection.

Based on the selected policy options the inspectors have additional ways to select their targets which are discussed later.

6.3 Additional Model Concepts

As shown in the previous subchapter assumptions have to be made about the way the agents behave in our model. In this process we try to simplify the real world while trying to keep as much of the perceived real-world behaviour, and possible future behaviour, in the model. Each extra variable we add comes with added uncertainty in the model. For this reason the model is kept as simple as possible while still being able to produce real-world behaviour and the effect of each uncertainty is carefully evaluated. Several additional design choices and assumptions have been made. In this subsection design choices and assumptions will be explained. The assumptions defined below hold for the general model parameter values within the specified uncertainty range without any additional policies added.

6.3.1 Model Framework

At the start of the modelling process four key variables for modelling an inspection game as an agent-based model were identified. These variables are:

- Number of inspectors;
- Number of companies;
- Costs of compliance;
- Height of a potential fine.

These variables together already capture the essence of an inspection game where a potential impact of a change in these values is very clear. By making the companies estimate the chance that they will get inspected they can calculate expected profits. The companies compare their estimated profit against their costs of compliance divided by their propensity to violate as explained earlier in this chapter. These variables are the main levers when initialising the model with an expected compliance rate. The number of inspectors and companies have a very clear relation with the percentage of companies that will be inspected and therefore with the compliance rate as well. The number of inspectors is, however, a limited resource for monitoring organisations.

It is to be noted that the companies and the monitoring organisation do not have a budget or keep track of their costs in any way. Every inspection round the companies set their strategy independently from a build-up wealth pool or debt. They do, however, take into account their own inspection and compliance history. The way the costs of compliance and a fine influence the decision-making of some companies has been discussed in the previous subchapter.

In order to study the effect of the use of data by monitoring organisations these four variables are not enough and additional concepts were added. These are explained in the next two subchapters.

6.3.2 Adding Data Science to Inspections Games

Unique characteristics have been given to each company in order to differentiate their behaviour and to make it possible for the inspectors to identify companies which are more likely to violate by making use of collected data to turn them into risk indicators.

The correlation of these characteristics with the propensity to violate is estimated based on visits by the inspectors. In the first timestep the inspectors use the correlation given by the variable “initial correlation estimation”. This is because the inspectors have no available data at this point. This initial correlation does not have a large influence on the behaviour of the model since it is only used in the first timestep. Nonetheless, an initial correlation estimation is needed in order to make the model run. After the first round of inspections the NetLogo extension “stats” is used to calculate the correlation between each characteristic and the propensity to violate of the companies. This calculation is performed only on the characteristics and the propensity to violate of the companies inspected during that timestep. Here it is assumed that inspectors are able to use their expertise to estimate the propensity to violate of a company when visiting. This can be seen as the inspector’s tacit knowledge and ability to rather quickly estimate what kind of company they are visiting. The characteristics in the model have no real-world values or meaning but could be replaced by any of the real-world characteristics companies have, such as their location, size or revenue.

The calculated correlations are then used to estimate the propensity to violate of all companies for the next round of inspections. It is assumed the general characteristics of the companies are visible to and known by all the inspectors. This estimation is made by using the highest correlating factors first, similar to a decision tree where a complex problem gets broken up into simpler factors, starting with the factor that adds the most amount of predictive power (Sharma, Bhargava, & Mathuria, 2013). This continues until all five factors are used or their estimated correlation is 1 or higher. At this point, using more factors with a lower correlation would decrease their accuracy. The factors get multiplied with their respective characteristic and these outcomes are added. At the end, this sum is divided by the sum of the used factors, giving the inspectors an estimation of the propensity to violate based on the correlation they calculated earlier.

This is a simplified version of the real-world data science behind inspections and aims to replicate its behaviour in simple code without the use of more advanced software and data science. This implementation allows the monitoring organisation to make use of a couple of characteristics which they think have a strong correlation and possibly miss out on other correlations when they choose to work fully data-driven, which is similar to the real world.

This process is executed at every inspection round and the correlation estimations of previous years are updated. This way, changes in these estimations by the inspectors show up more quickly and the relevant causes are identified more easily. This also allows the inspectors to react more quickly to potential changes in the environment such as changes in the correlation factors or the companies.

The compliance history of companies, however, is not reset every round and weighs more heavily than the estimation of the propensity to violate. The perceived compliance of companies here is measured as a percentage of the number of visits a company showed compliant behaviour. This compliance is one of the most important things monitoring organisations learn about companies during inspections. In the model used for this study this compliance is only calculated when companies get visited. When the inspectors have no information about the compliance of a company, they use the average compliance of the other companies for their risk assessment. This ensures every company has an estimated propensity to violate and an estimated compliance score.

When a company does not get inspected for a while their compliance score in the inspectors' risk assessment also gets reset to the average. If this is not the case, it is possible for offenders to get one lucky inspection early on in the simulation and enjoy their 100% perceived compliance rate for the rest of it. It is implemented the other way around as well. If a company shows compliant behaviour a number of times in a row, the inspectors will reset their compliance score in their risk assessment to the average. This ensures inspectors do not get stuck on companies violating only one time and then showing compliant behaviour for the rest of the model run. Their observed compliance scores do not get reset, only the values used for the risk estimations take on the average during the years they are not inspected.

The data quality is measured as the sum of the absolute values of the inspectors' estimations of the correlation between the characteristics and the propensity to violate minus the real correlations. This tells us how much they are off from the real correlations. A larger deviation can occur when inspectors are not distributed evenly and only focus on companies with certain characteristics. In addition, the inspectors' perceived compliance is compared with the simulated compliance to see how much their sector overview is skewed. We do this for both absolute numbers, to see if in some scenarios their perception is better or worse, as well as for fluctuations and patterns.

6.3.3 Strategic Behaviour

The companies also have some additional variables to help improve their strategic decision-making. These variables are used in their estimation of the chance that they get inspected and are:

- Expected inspection multiplier;
- Time to reset expectation after getting caught;
- Percentage expectation decrease;
- Minimum expectation percentage.

All of the companies use the same mechanism for strategic behaviour under all model scenarios. There are different ways companies can show strategic behaviour. In this model, strategic behaviour refers to the choice companies make whether they comply with the rules or not and how advanced the decision making is to reach that decision. The level with which they change their strategy based on their expectations depends on their propensity to violate, which is different for each company, and the incentives for violating. Both of these can be varied during experiments. The level of strategic behaviour the companies show can then be measured by how often the companies change their strategy and how often they choose the "correct" strategy from their perspective. The way each of the companies' expectations are calculated is explained below. The four added variables are required for the updating of their estimations based on their inspection history.

Expected Inspection Multiplier

A company's estimation of the chance that they will get inspected goes up whenever they get caught. This is because they learn that the inspectors now know they are a riskier company and at the same time they notice they get inspected (Schell-Busey et al., 2016). If they get caught again their expectation grows exponentially, multiplying with a factor every time they get caught. This factor is the uncertainty called "expected inspection multiplier". This factor increases a company's estimation of the chance they get inspected every time they get inspected. This way the companies show they have a memory of the times they get inspected and their estimation can take many values depending on how often and when they got inspected.

Time to Reset Expectation After Getting Caught

If they do not get caught again their expectation immediately starts resetting back to their initial estimation over multiple steps. This number of steps is called the "time to reset expectation after getting caught". This uncertainty is a fixed number of inspection-rounds it takes the companies to reset their current estimation back to the default estimation of the number of inspectors divided by the number of companies. Resetting this gradually ensures companies also gradually start violating again at different timesteps. If the companies would suddenly forget their increased estimations this would lead to strange behaviour where large groups of companies choose the same strategies independent of their propensity to violate. This would also not fully take into account the impact of an inspection, which could be larger or smaller depending on the "expected inspection multiplier" uncertainty. The "time to reset expectation after getting caught" uncertainty represents how long the impact of an inspection lasts, not how big it is. Together these two uncertainties create different inspection impacts on the companies. When a company gets inspected while it is "resetting" its expectations, their estimation of the chance they get inspected increases again from their current estimation with the factor described before. The value of this multiplier and the speed with which their expectations go back to normal are unknown and experiments are performed for different values.

Percentage Expectation Decrease

The "percentage expectation decrease" uncertainty is the level with which a company's expectations decrease each round after not being inspected. This decrease is separate from the gradual reset explained in the previous paragraph. This decrease only happens when companies are at or below the initial estimations of the number of inspectors divided by the number of companies. This uncertainty decreases their estimation by a certain percentage every timestep they do not get inspected. This percentage is unknown and experiments are performed to see the impact for different values. If a company gets inspected while having a lower than default estimation, they will immediately use the default estimation for the next round. In this case the expected inspection multiplier does not matter, unless it would have put them at a higher estimated chance to get inspected.

Minimum Expectation Percentage
Their expectation level can only drop to a certain percentage of their initial estimation to prevent companies with a lower propensity to violate from violating. Because these are important financial decisions with legal consequences, we can assume there is a minimum expectation level companies will stay at, at which they are unwilling to risk malpractice, especially since these are the companies with lower propensities to violate in the first place. The fact that there is a group of companies that always comply, which is the group of companies with the lowest propensity to violate, also prevents this, but in case this group is set to be very small this minimum expectation level will be used. The "always comply" uncertainty functions more as a lower bound in extreme imbalance in the height of the fine and costs of compliance. We assume that even when the financial incentives to violate are very large, a small percentage will still comply with the rules. The minimum expectation percentage, however, does not do this. This uncertainty is there to make sure companies with a lower propensity to violate will comply with the rules in more balanced scenarios.

This combination of assumptions leads to company behaviour that shows a clear and recognizable relation with their propensity to violate. Inspectors should be able to identify companies with a high propensity to violate using data they collect during their visits. These uncertainties were added during the modelling phase while experimenting with the companies' behaviour. Literature discussed in previous chapters and experts at the NVWA described how companies would behave. By adding these uncertainties this behaviour was recreated in our model and we got insights about factors that play a role in deciding company behaviour. This was mostly done using the tracking of single agents to see how they behave. Multi-agent tracking highlighted the abrupt changes in strategy of groups of companies and showed the need for more gradual decreases in the estimation to get inspected by companies in order to showcase their unique propensities to violate. Preventing the inspectors from getting stuck was also an insightful modelling exercise. This was done by using the normalised average compliance rates of companies when they have never been inspected or have not been inspected for a while. Without this, the inspectors would get stuck on a company that showed non-compliant behaviour once, or a group of "high risk" companies, even though they showed compliant behaviour. This also prompted us to use the compliance history of the companies in the risk analysis of the inspectors. This made the inspectors show much more realistic behaviour where they actually strategically target the companies instead of just the ones with the highest propensity to violate. Because we are unsure about the weight of the compliance history versus the estimation of the propensity to violate, an additional uncertainty, "compliance data weight" was added as a policy option. This allows monitoring organisations to set strategies based more on the compliance history or more on the estimated risk indicators.

7. Model Implementation

NetLogo was chosen as our modelling environment because of its ease of use and its inherent ability to generate a visual and interactive object-based simulation model. For the implementation and documentation of the model the ODD protocol is used (Grimm et al., 2006). The basic idea of this protocol is to structure information about the model in order to give readers a quick overview of the model's focus, resolution and complexity. The following subchapters below follow the steps described and examples provided by the ODD protocol. This chapter makes it possible for readers to implement this model in their own object-oriented programming language.

7.1 Model Purpose

The purpose of this model is to explore the long-term effects of different data-driven monitoring policies under a range of possible scenarios and assumptions.

7.2 State Variables

The model consists of two active types of agents and their environment. Inspectors are sent by monitoring organisations and work either randomly or based on a risk analysis performed on the data they collected through inspections. The data-driven inspections can be performed based on the compliance history of companies, risk indicators identified by their risk models or a combination of those. The inspectors have an observed compliance level. The total compliance level is tracked on the environment level. The risk estimation used by the inspectors is based on the data they have collected during their visits. This contains information about the compliance of companies as well as their propensity to violate. The inspectors also have knowledge about the average compliance rate they observed and remember how long ago each of the companies was inspected. When working data-driven the inspectors use all the information they have and try to catch non-compliant companies.

The companies have varying general characteristics that are partly random and partly correlated with their propensity to violate. They make an estimation of the chance that they will get inspected. This estimation in combination with their propensity to violate determines their strategy each round. Their strategy can change every round. Their estimation is based on their inspection history. When they get inspected their estimation gets multiplied with an owned variable. Their estimation also decreases when they do not get inspected based on other variables they own. These are the time to reset their expectations, which lowers their expectations gradually, the percentage of expectation decrease after their estimations have been reset and the minimum percentage of estimation.

There are additional variables in the environment that influence the behaviour of the agents. The number of inspectors and the numbers of companies can be set each experiment. The height of a potential fine and the costs of compliance are also set on the environment level. These variables determine how likely companies are to violate. Different propensities to violate and minimum compliance percentages can be set to simulate different kinds of violators.

7.3 Simplified Model Narrative

Setup:

1. Set initial values
2. Create inspectors
3. Create companies
    a. set propensity to violate
    b. set characteristics based on propensity to violate and correlation
    c. set initial estimated chance to get inspected

Go:

1. Companies choose their strategy
    a. estimate chance to get inspected
    b. choose strategy
2. Inspectors select companies to inspect
    a. estimate propensity to violate
    b. estimate chance to violate
    c. select companies with highest chance to violate
    d. select random companies
3. Inspectors perform their inspections
    a. give fines
    b. update compliance history
4. Update estimated correlations
5. Update statistics

7.4 Design Concepts

Emergence: Inspection patterns emerge from the combination of data-driven inspections and estimations to get inspected by the companies. These are simple rules that guide one agent's behaviour but lead to patterns visible on the systems level. Especially when sorting the companies by their propensity to violate this behaviour becomes visible. Inspectors can get stuck inspecting a certain group of companies or switching between multiple groups.

Adaptation: Companies adapt their strategies when they notice they get inspected more often than initially estimated. Inspectors adapt their targets based on the data they collect during their inspections.

Objectives: Companies' objectives can differ from each other. Their propensity to violate sets their objectives. Some companies only care for financial incentives and try to choose the most profitable strategy each round. Other companies care more about the rules and need a higher incentive for violating or will always comply.

Prediction: Companies try to predict future events based on the number of available inspectors per company and their memory of times they got inspected. When they get inspected often, they will predict they will get inspected more. When they do not get inspected, they predict to get inspected less.

Sensing: Inspectors are assumed to know, by "sensing", the propensity to violate of companies upon visiting and share this with the rest of the inspectors, representing the monitoring organisation as a whole. The general characteristics of companies are assumed to be known to the inspectors. The companies “sense” how many inspectors there are in total and how many companies there are.

Interaction: Inspectors share all knowledge between each other as they are assumed to be part of the larger monitoring organisation. Companies do not share information between each other. Inspectors and companies interact when companies get inspected. Based on this interaction the inspectors update their data and the companies recalculate their strategy for the next round.

Stochasticity: The model contains randomness. The most obvious place where randomness plays a role in the model is when inspectors choose to inspect randomly and not based on data. Other parts where the model contains randomness is in the setup of the model when creating the companies and setting their propensity to violate. This randomness makes each company unique and behave a little differently based on their incentives. The characteristics the companies get also have a random part depending on the height of the correlation of these characteristics with the propensity to violate. This makes it so the inspectors cannot perfectly estimate the propensity to violate of the companies.

Collectives: The inspectors are representing a larger entity, the monitoring organisation.

Observation: During the modelling and exploration phase the companies were sorted on their propensity to violate and showed either red or green based on their chosen strategy. For the experiments several outcomes of interest on the system’s level were recorded. These are the data quality, observed compliance, total compliance and the difference between the observed and total compliance.

7.5 Input Data and Variable Ranges

In order to run experiments and explore the effects of data-driven inspections parameter ranges were identified. These ranges and the values for the reference scenario can be found in the table below (Table 6.1). These ranges were kept as large as possible to fully capture all possible combinations the model allows. The model initially starts without observed compliances rates or inspections histories. This is why the model requires a warm-up period to generate its behaviour. Starting the model with initial data would prevent us from getting insights about the first stages of data-driven monitoring where there is no data available yet.

The number of agents in the model has a minimum value needed for the model to run. The number of agents has a large influence on the speed with which the model can run experiments, for this reason, and because it is mostly the ratio of these two that matters and not the absolute numbers, the maximums of these are kept relatively low. For the fine and costs of compliance ranges were identified where all possible ratios were possible. This means it allows for cases where it is extremely profitable to violate as well as cases where there is no profit to be gained at all. The maximum value for the fine is higher than the one for the costs of compliance to more often get realistic scenarios where there is an incentive to stick to the rules when getting inspected. For the other uncertainties the ranges were tested to see where changes to these parameters still had impact on the results.

The range of the percentage of companies that always comply with the rules is kept relatively small compared to the other variables because it can overshadow the other variables when it takes on an (unrealistically) high value. Sectors where there is very little malpractice are still considered during the experiments because of the higher range of the minimum expectation companies will have. This means that companies in these well-behaving sectors will not start misbehaving even when not getting inspected for a long time, which is similar to the percentage of companies that always complies variable. The percentage always comply variable is more of a bottom boundary for these kinds of companies as explained in the previous chapter.

The values for the reference scenario are based on personal communication with experts. The specific values below are based on a conversation with a catering expert from the department expertise at the NVWA. Catering is a sector with a relatively large amount of companies, at about 120.000. Within this group of companies about 20.000 companies get inspected, these inspections are performed at roughly 15.000 different companies. The compliance rate here is 60% and inspectors go back to non-compliant companies for second inspections or companies have to prove themselves they are complying with regulation after the inspections. A large part of the non-compliant companies in this sector are unknowingly violating the rules because they do not know them. There are, however, also companies that knowingly make financial trade-offs when choosing their strategy. It was also mentioned that companies need to be inspected structurally to have an impact on their compliance. In this sector they try to perform 10% random and 90% risk-based inspections. The reference values in the table below were created with this sector in mind to make sure they are representative for a coherent validated sector where the ratios of the parameters are at least roughly related to those in the real world. This way we can say get insights about the impact of each of the variables and our policy options on a real-world-like scenario.

The numbers and ratios given by the expert were directly used for the reference scenario. The level of strategic behaviour was adjusted with the uncertainties representing strategic behaviour to match the described behaviour. For this and the types of violators it was important we had a reference compliance rate so we roughly knew what the model outcomes should look like under this scenario with a fixed policy. Using this expert knowledge as input for the model immediately showed recognizable behaviour and resulted in outcomes that were given by the expert as well. Where knowledge seemed to be lacking, the variables ranges were kept as wide as possible to explore the effects of these variables. Other scenarios are randomly generated using the identified ranges.

Table 6. 1 Parameter ranges and reference values

Variable Name	**Reference Scenario**	**Sample Range**
Number of Companies	240	5 – 500
Number of Inspectors	40	5 – 100
Fine	120	20 – 200
Costs of Compliance	40	5 – 100
Percentage Data-driven	90	0 – 100
Compliance Data Weight	3.0	0 – 5
Percentage Expectation Decrease	0.02	0.01 – 0.05
Percentage Always Comply	30	20 – 40
Minimum Percentage Average Expectation	70	0 – 70
Expected Inspection Multiplier After Caught:	1.2	1.1 – 2
Number of Inspections to Build Trust	5	0 – 10
Memory Compliance Years	5	0 – 10
Time to Reset Inspection Expectation After Caught	5	1 – 10

8. Verification and Validation

This chapter will explain the verification and validation process performed for the created model. First the verification process will be given along with the results. After this the validation of this model will be discussed.

8.1 Verification

Verification is about making sure our conceptual model has been translated to code correctly. The results of the verification steps can be found in Appendix B. The verification of this model is done by going through the four verification phases as described by Dam, Nikolic and Lukszo (2012):

1. Tracking Agent Behaviour
2. Single-agent Testing
3. Interaction Testing in a Minimal Model
4. Multi-agent Testing

8.1.1 Tracking Agent Behaviour

The first performed step was to track agent behaviour. In order to verify the correct working of the model, relevant output variables were selected and monitored. The input, states and output of individual agents are recorded for both the agent and the internal processes performed by the agent. During the implementation phase print statements were also actively used to monitor internal agent states. During the modelling phase issues with the calculation of the compliance rates used for the risk estimation were found. These should have returned to the average compliance rates when they do not get inspected for a while. This was fixed and during the verification step no unexpected behaviour was shown here.

8.1.2 Single-agent Testing

The second verification step was the testing of a single agent. During this step theoretical predictions were made of how an agent will behave under normal inputs and these hypotheses were then tested. After this, we tried “breaking” the agent to find parameter values where the agents behaved in an unexpected way or broke altogether. These unexpected errors do not necessarily mean an implementation mistake was made but can also be used to establish the limitations of the computer code and/or model. It is important to know these parameter ranges when experimenting with the model. No errors were found during this validation step. During the modelling phase, single-agent testing has helped with the modelling of the behaviour of companies. When inspecting a single company, it became clear why certain variables were needed in order to make them show natural behaviour, these uncertainties are discussed in chapter 6.

8.1.3 Interaction Testing in a Minimal Model

After testing a single agent, the third step is to test agent interaction in a minimal model. For this step, the same tests as for the single agent will be used but instead the model is set up to run with the minimum number of agents necessary to run. Here we confirm if the basic agent interactions happen as they are described in the model narrative. This test showed the monitored compliance rates could be higher than 100% when every company gets inspected.

This had to do with an issue where the compliance rates were being calculated after the first timestep but were divided by the total number of timesteps. This issue has been resolved. Next to this, this verification step showed the companies were always compliant because of the large number of inspectors per company. This makes sense and confirms their default risk estimation works.

8.1.4 Multi-agent Testing

The last and fourth verification step performed was multi-agent testing. In this step the same tests are performed again but for the whole model looking at predicted behavioural patterns and behaviour. Two additional tests are performed during this step. The first being variability testing. Because of the chaotic nature of agent-based models, we should take care to uncover any interaction-order artefacts by exploring the variability in the output space. By running many replications of the same model setup this variability has been examined across a number of output variables. This is done by looking at the standard deviation, variance and skewness. The second additional test is to test the timeline sanity. During this step the output of default model runs was checked by stakeholders for unexpected behaviour. If weird behaviour shows up that cannot be explained by reasoning through the model logic, it might indicate an implementation error. The model showed no unexpected behaviour during this last step.

8.2 Validation

Validation is concerned with the question *"did we build the right thing?"* (Dam et al., 2012). Is the model we designed capable of helping us answer our research questions? Validation ensures the outcomes of our study are relevant to the problem owner. This can best be seen as a continuous process during the modelling phase instead of a one-time check at the end of the process. Here we follow the method of Dam et al (2012) who deviate from the traditional view that model validation is only about the accuracy with which it represents real-world systems. There is currently not a real-world system available to easily compare results because we are doing explorative research and try to answer "what if" types of questions (Louie & Carley, 2008). We cannot compare our simulated behaviour to real behaviour if there is no real system available for comparison. Instead we focused on whether our model is useful and convincing in explaining how possible futures might come about. The true outcomes of this model are insights about the system we try to simulate, and not the outcome numbers the model generates. Therefore, we try to validate if these insights are valuable to the problem owner and help us answer our research questions.

Some other studies have also mentioned the difficulty of validating these types of agent-based models (Liu, 2011; Xiang, Kennedy, Madey, & Cabaniss, 2005). They argue that these models can still be verified by making predictions about their output behaviour and then testing if the model is able to produce these results. For this study this has already been done as part of the verification of the model (appendix B). Liu (2011) argues the researcher should identify a set of stylized facts he/she is interested in explaining with a model. After this the researcher should build the model in a way that keeps the description about the agents and rules as close as possible to empirical and experimental evidence and then see if the model shows the desired behaviour. Our study implemented simple agent rules grounded in literature, as discussed in chapter 6. With this model we try to explain possible future behaviour that was mentioned by literature and experts, as explained in the first chapters of this study. Our model seems to be capable of explaining and recreating this described behaviour in the form of the creation of a data bias, and showing the described behaviour of inspectors and the inspected, and is therefore validated for its designed purposes (Liu, 2011). Xiang et al (2005) support this way of validation. Additionally, they present *animation* as a way of allowing domain experts to validate an agent-based model. This requires built-in features from simulation software for visualisation and agent-tracking, which are included in NetLogo.

This way of validating, through animation with experts, was also part of our validation as this helps us in validating if our generated insights are valuable to the problem owner. The initial problem and model description have been discussed several times with the problem owner at the start of this study. Our model does not use real data and could therefore not be validated with the use of real-world data. Domain experts and problem owners were involved during the modelling process and discussed the behaviour of agents, the patterns they show and the application of this model in the real world. An interactive presentation of the model has been given for a group of eight data-scientists and the departmental head of knowledge and research for the NVWA. During this presentation possible model behaviour was demonstrated, such as the effects of the use of data for inspections, and discussed. This discussion was quite a lively one where the animated behaviour was clearly recognizable. The discussion was often about additional factors that could be included to expand the scope of the research. The experts could be convinced that many of these were in the end not required to perform this study. Patterns the model showed were clearly recognized by the experts and all the key mentioned factors were included in the model, while keeping it as simple as possible. These experts do not work in a specific sector but have much knowledge and experience with monitoring and data science in a wide range of sectors. This increases the ability of this validation to generalise across multiple sectors and scenarios.

Additionally, a session was had with a sector-specific strategic domain expert. During this session the potential use of this model for his sector was discussed. The expert understood the assumptions that were made for the behaviour of the agents and could therefor provide us with input parameters for a reference scenario which resembles his sector. The expert additionally provided us with system behaviour the model should be able to generate. This included model output numbers such as the observed compliance rates with a certain number of inspectors, but also behaviour from companies. For example, companies often needed more than one visit in order to change their strategy. Our model was able to replicate scenarios described by both this case expert during personal communication as well as the possible scenarios and identified issues in literature mentioned in earlier chapters. Therefore, we conclude this model is validated for its designed purposes.

9. Experimental Design

Experiments are used to evaluate how different combinations of uncertainties and policies can affect the outcomes of interest. Because running experiments takes time and computing power it is not realistic to run every possible combination. Therefore, we design sets of experiments to efficiently cover a wide space of possible parameter combinations in an attempt to capture all possible model behaviour. This way we can effectively explore strategies for monitoring organisations under different scenarios and assumptions. Next to this we design our experiments in a way that enables us to evaluate the effects of each parameter on the outcomes of interest. Individual policies can also be tested under a range of possible scenarios. This will give us insight in the performance of a policy under uncertain conditions. Each model run contains randomness. By running the model many times and replicating the experiments we try to make sure we get reproducible behaviour from which we can learn. This prevents us from drawing conclusion based on a single run which only holds for very specific parameter values or even specific random values.

The experiments performed during this research will be explained in sections below. First, the software used to execute the experiments will be explained. Secondly the outcomes of interest will be discussed. Thirdly, the performed Sobol global sensitivity analysis will be explained. Lastly, open exploration with the use of the ema workbench will be discussed.

9.1 Using Python to Run Experiments

The performed experiments were set up and controlled using Python. NetLogo does offer an integrated way of sweeping parameters and running many replications at the same time. Using Python, however, we are able to use multiple different samplers to sample our input space in a more efficient way and immediately get back the results for visualisation and analysis. The library used for interacting with NetLogo from Python is called pyNetLogo (Jaxa-Rozen & Kwakkel, 2018). This library is needed in order to use and access NetLogo models from Python and run our sampled experiments.

The SALib library (Herman & Usher, 2017) was used to sample and analyse suitable experiments for a Sobol global sensitivity analysis. This library contains Python implementations of the most commonly used sensitivity analysis methods. This library allows us to calculate the effects of model inputs and exogenous factors on our outcomes of interest. This sensitivity analysis and its sampling will be explained later this chapter.

Furthermore, a package called ipyparallel was used to run multiple simulations on multiple processor cores at the same time in order to run the experiments faster. The EMA Workbench (Kwakkel, 2017) was also used to design and perform experiments. The EMA Workbench offers tools for supporting exploratory modelling in various modelling environments, including NetLogo. This was used to perform the open exploration of the model and explore policy effects under a wide range of uncertainties. The workbench contains tools for setting up experiments, sampling input spaces, performing experiments and analysing and visualising the results. Next to the open exploration, the EMA Workbench also offers the option for directed search, where one actively searches for parameter combinations which lead to desired results or to find worst case scenarios by the use of optimisation algorithms.

9.2 Outcomes of Interest

Four outcomes of interest were identified to be used for the experiments. Exploring these outcomes together with the input parameters can help us to answer our research questions. These four were chosen as they give a good overview of the identified issues of data quality versus compliance.

Total Compliance
The first outcome of interest is called the Total Compliance. This represents the compliance level of the sector as a whole from the perspective of the all-knowing observer. In the real world it is impossible to measure the actual compliance level and sampling and calculations are used to make an estimation of this level. Because this is a simulation, we are able to observe the "actual" compliance levels of the companies from an independent observer perspective. This is calculated by taking the companies that choose to comply as a strategy and then dividing them by the total number of companies. In the real world these values can be incorrect and malpractice can be missed. A higher total compliance level is favourable for monitoring organisations since this is one of their main targets.

Observed Compliance
The second outcome of interest is closely related to the first and is called the Observed Compliance. This is the compliance level as observed by the inspectors in our model. This is calculated by taking all companies who were inspected in a certain timestep while having a compliant strategy and dividing them by the total number of companies that were inspected during that timestep. This represents the view of the overall compliance the inspectors have of the companies. This compliance level is closer to the numbers we get in the real world. Monitoring organisations wish to have a high level of compliance but at the same time they try to mostly visit bad companies. When they have a high level of observed compliance this is only good when their inspected locations are representative of the whole sector. It could be possible that all companies that are not visited do not comply with the rules. At the same time, it could be true that monitoring organisations mostly inspect non-compliant companies because they are actively searching for them. This is why it is important to compare this with the Total Compliance as it does not say too much on its own. The Observed Compliance is still taken as an outcome of interest because it allows us see where the difference in the total and observed compliance come from and if it is a positive or negative difference.

Absolute Observed Compliance – Total Compliance
The third outcome of interest is used to monitor the gap between the Total Compliance and Observed Compliance and is called "Absolute Observed Compliance – Total Compliance". As explained in the previous section this is an important outcome that indicates how skewed the overview of a monitoring organisation is. This outcome is calculated by taking the absolute value of the Observed Compliance minus the Total Compliance. This gives us the difference between these two outcomes. A large difference indicates the inspectors are not representatively inspecting companies and this could be seen as a warning when monitoring organisations want to say something about the compliance level of a sector. A high difference does not have to be a bad thing since monitoring organisations are actively looking for violators and prefer to inspect them instead of companies that comply with the rules. This is why it is important to look at this as a warning and also compare it side by side with the observed and total compliance to see where the difference is coming from.

Data Quality
The last outcome of interest used to answer our research question is the data quality. This is an important outcome for answering our research questions. It represents the quality of the data the monitoring organisations possess after doing their inspections. Quality here means the accuracy with which the data can be used to say something about the correlation between the characteristics and propensity to violate of companies. The inspectors learn about the correlation between characteristics and the propensity to violate when visiting companies. When all companies get visited the inspectors will have a good estimation of the correlations. When the estimations of the inspectors differentiate from the actual correlations in the model this is seen as working with lower quality data. This can be the case when inspectors only visit a small group or specific subset of companies. The data quality is calculated by taking the absolute value of the estimated correlations minus the actual modelled correlations. A lower data quality indicates the monitoring organisations ability to evaluate risk indicators is getting worse. Combining this with the compliance outcomes will let us evaluate how biased the data used by monitoring organisations can become under different scenarios and policy options.

Fair inspections
To answer our question about the fairness of data-driven monitoring to the involved companies we could look at the percentage of inspections at specific groups of companies. This, however, does not give us a clear answer on how fair inspections are and might even be misleading as fairness is not just about a number of inspections. A high number of inspections, and thus higher administrative burdens, can be argued for as fair when a company is showing less compliant behaviour. The number of inspections at companies with a high propensity to violate is monitored in the NetLogo interface (Appendix C) and used to rationalise and think about certain outcomes and behaviour we see. Adding it to the outcomes of interest, however, would clutter the results since there is already much too process, analyse and most importantly visualise. Interesting scenarios were recreated using the NetLogo interface to help explain the results by for example looking at these mentioned distributions of inspections. They mostly show linear behaviour with little added value to our current set of experiments. For these reasons it was chosen to leave them out of the outcomes of interests for the global experiments.

9.3 Uncertainties

The model consists of policies and uncertainties. These uncertainty variables have a range within which they can assume a value. Each of these uncertainties and their ranges are explained in chapter 6 and 7. A combination of these uncertainties together we call a scenario. Each uncertainty has its own effects on the outcomes of the model and influences the effects of policies on the outcomes. For the experiments we split them into two different groups. This is done because some uncertainties can be known beforehand or estimated by monitoring organisations depending on the sector they monitor. This first group we call “sector specific uncertainties”. They differ from sector to sector but are mostly stable within a sector. Because we try to say something about data-driven inspections in general and not for a specific sector, they are still uncertainties. However, because they can be known by monitoring organisations and they have a large impact on the outcomes of interest we chose to keep them fixed or within a small range for a part of our experiments to test the effects of the other uncertainties and the policies in given sector using a sensitivity analysis. For this we used the reference values given in chapter 6. Results from this part of the analysis do in principle only hold for the reference scenarios used in this research. It is, however, likely the uncertainties would affect the outcomes in the same way under different real-world scenarios. The sector specific uncertainties are:

- Number of Inspectors;
- Number of Companies;
- Fine Amount;
- Costs of Compliance.

Experiments where the full range of these uncertainties were varied were also performed. These results hold for all possible parameter combinations within the ranges identified in chapter 6.

The various correlations between the general characteristics and the propensity to violate of companies are not varied during these results. A first exploration with these characteristics shows that they do not have a very large impact on the results (Appendix G). Especially when randomly sampling them independently they do not add much value to our experiments because one or two characteristics can be good enough for inspectors to make a decent estimation. Because there are five different correlations, sampling them all would add up to a lot of extra model runs and time if they were added in with all the other variables. Experiments with these correlations should be setup separately in a way different hypothesis can be tested. More on experiments with these correlations can be found in the reflection (Chapter 13).

9.4 Policy Options

The policy option that is explored for this paper is the use of data-driven monitoring, implemented as explained in chapter 6 and 7. For the experiments this is a combination of the two variables, Percentage Data-driven and Compliance Data Weight. and their ranges which are given in table 6.1. Since the percentage of data-driven inspections can be varied from 0 to 100 it is important to test many options, especially when combining it with the Compliance Data Weight. Each of these combinations could be a viable policy option for monitoring organisations. With the designed experiments, which are explained below, we can explore the effects of these policies on the outcomes of interest under a wide range of possible scenarios. We will also explore how this effect is affected by the uncertainties in the model. It should be noted that we compare this way of data-driven risk-based inspections against random inspections, risk-based inspections without the use of data are not explicitly modelled and do therefore not come back in the experiments.

9.5 Run Length

The model run length is 100 timesteps. After some exploration and testing with the model parameters this seemed to be the longest time the model needed to stabilise after changing values mid-run or when starting the model without any compliance history. This means each experiment runs for 100 rounds of inspections. This is quite a large number of rounds which does increase the time needed to calculate the results. However, during the modelling and testing phase some interesting patterns were shown to develop after longer runs and to prevent missing out on these results 100 timesteps was chosen.

9.6 Sobol Global Sensitivity Analsysis

Sensitivity analysis is used to explore how a model's output can be apportioned to its different input variables. Variance-based sensitivity analysis, also called Sobol (named after Ilya M. Sobol) is a tool for global sensitivity analysis. It decomposes the output variance of a model to fractions which can be linked to inputs or sets of input. This means that for example, we can find that 80% of the variance in the model output is caused by one input while 10 percent is cause by the second input and the remain 10% due to interaction between these two inputs. We use a variance-based sensitivity analysis because they are able to measure sensitivity across the whole input space, can deal with nonlinear behaviour and are very useful for measuring interaction effects between different inputs (Saltelli & Annoni, 2010). Compared to the traditional method of changing one factor at a time, this offers the benefit of being able to analyse interaction effects between variables.

The sensitivity of the model to each of the input variables is given by the sensitivity index. The creators of the SALib library (Herman & Usher, 2017) list three forms of sensitivity indices. These forms are directly taken from their website and are:

1. First-order indices: measures the contribution to the output variance by a single model input alone.
2. Second-order indices: measures the contribution to the output variance caused by the interaction of two model inputs.
3. Total-order index: measures the contribution to the output variance caused by a model input, including both its first-order effects (the input varying alone) and all higher-order interactions.

Using the Python library SALib it is possible to perform a sensitivity analysis on our NetLogo model. SALib is responsible for generating the model inputs and then computing the Sobol sensitivity indices using the model's output. The first step in this sensitivity analysis would be to determine the model inputs and their sample range, which has been done in chapter 6. After this the Saltelli sampler is used to generate our inputs. This sampler is a revised version of Sobol's sequence for sampling for a Sobol sensitivity analysis (Herman & Usher, 2017). Using ipyparallel the model then executes using the sampled input parameters in parallel and returns the results. The results will be analysed and visualised also with the help of SALib.

For this paper we performed this procedure for two different parameter ranges. All of these experiments include sampling of the random seed to see how much of the output can be addressed to randomness. The first analysis was performed using all uncertainties and policies with their respective ranges.

The second analysis was performed while leaving all the sector specific uncertainties at their default values and only sampling the remaining uncertainties and policy options as explained above. This is in order to determine the effect of all input variables without them being overshadowed by the other inputs which are in many cases less uncertain than the ranges we test them for in the first analysis. Next to this, the second analysis is also performed while keeping the policy options fixed at different values just to determine the impact of certain uncertainties when a policy option is chosen to see what uncertainties matter most and might require additional knowledge.

9.7 Open Exploration

Next to the sensitivity analysis an open exploration of the model was performed using the EMA Workbench. For this exploration many combinations of policy and uncertainty options were sampled using Latin hypercube sampling. This exploration can show the many possible outcomes the model is able to generate. The model outcomes are visualised and sorted by policy option to show how the percentage of data-driven inspection influences the possible range of outcomes under the given uncertainties. Using pairs plots, trade-offs between the outcomes of interest can be explored for each of the different policy options. For this visualisation the first timestep of each model run has been removed. This open exploration was also performed with both the whole range for the uncertainties and the fixed sector specific variables to get better insights about the effect of our policy options.

10. Results

In this chapter the results of the experiments will be presented and further explained. All of the results can be found in Appendix D - H. This chapter will give observations about the results and show some of them ordered by the experiments they were gathered from. When looking at the results it can be useful to keep in mind what our desired values for the outcomes of interest are, which were explained in the previous chapter. In short, monitoring organisations would like to increase the total compliance, while keeping their resources constant, while at the same time not giving up too much on their data quality and their overview of the sector.

10.1 Reference Scenario

The first results to look at from our experiments are the effects of data-driven inspections on the reference scenario. These results can be found in Appendix D - H. The figure below (Figure 10.1) shows the effects of the percentage of data-driven inspections on each of the outcomes of interest for the reference scenario. The variance in the results in this figure that cannot be explained by the percentage of data-driven inspections can either be attributed to the compliance data weight or randomness.

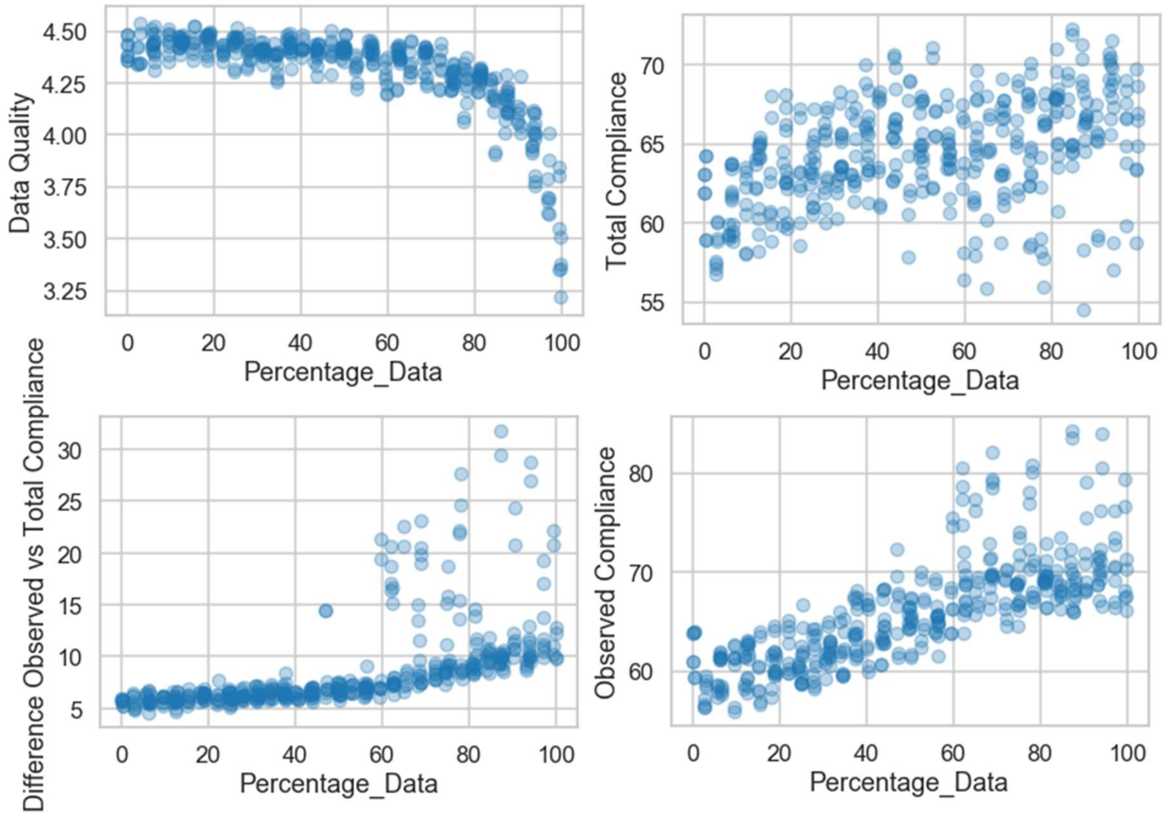

Figure 10. 1 Effect of the percentage of data-driven inspections on the outcomes of interest (reference scenario)

The figure above (Figure 10.1) shows that the percentage of data-driven inspections has a strong negative correlation with the data quality, especially at the higher ends. The Sobol analysis shows that 85% of the variance in the data quality for the reference scenario can be explained by the percentage of data-driven inspections (Appendix G). Interesting here is the drop off point the data quality seems to have. This suggests there is an optimum amount of data-driven inspections where the data quality is not influenced as much yet.

10.1.1 Effect data-driven Inspections on Compliance for Reference Scenario

When looking at the effect of the amount of data-driven inspections on the total compliance a Pearson correlation coefficient of 0.397 was found. This is quite a strong positive correlation. The figure above, however, shows that at the higher percentages of data-driven inspections there is also a chance they will actually lead to a lower compliance rate. Here the compliance data weight had a similar contribution to the variance and a similar positive correlation, meaning more emphasis on compliance history opposed to other risk indicators lead to higher compliance rates when working data-driven.

The observed compliance has a high Pearson correlation coefficient of 0.769 with the percentage of data-driven inspections. This means the observed compliance increases more when increasing the amount of data-driven inspection than the total compliance, meaning the difference in observed vs total compliance will also increase, leading to a skewed overview of the sector for monitoring organisations. Both the percentage of data-driven inspections and the compliance data weight have a positive correlation with the total compliance (Appendix F). Meaning more data-driven inspections with emphasis on compliance history could lead to a better overall compliance in the sector. Note that these are only the first set of experiments and only hold for the reference scenario.

10.1.2 Explanation and Interpretation of the Effects of Data-driven Inspections

This first set of results lead us to the question where the sudden decrease in data quality came from when inspecting data-driven. When exploring this in more depth the model showed that the risk indicators that had a low correlation or were not used for selecting companies were estimated much worse than those that were used. The highly correlation characteristics were estimated close to correct.

Another thing that can be noticed when running the model with high percentages of data-driven inspections is that this will lead to more strategic behaviour from the companies. Under these circumstances companies with higher propensities to violate will often end up violating less than companies with lower propensities to violates because they get inspected more often. This way data-driven inspections can cause monitoring organisations to miss a group of violators if they are not identified as being of high enough risk. This group can then start showing non-compliant behaviour when they are not inspected for a while. Even when the data models are not transparent the companies showed more strategic behaviour, avoiding inspections more often than with random inspections. This can be explained by the fact that the increase in the number of inspections at “high risk” companies will influence their behaviour. The companies will learn this and adapt to it. This showed to be bad for the inspectors’ overview of the sector.

10.1.3 Increasing Incentives for Malpractice

When increasing the costs of compliance during a model run, the first thing that could happen is an increase in malpractice, as expected. Against expectation, however, the compliance could afterwards stabilise at a higher level. This can be explained by the fact that inspectors learn about the compliance history of companies and use this to select their next targets. When it is very attractive for companies to violate, the amoral calculators will be easier to be picked out by the inspectors. When working only with random inspections this does not happen. It should be noted that this way of baiting companies to violate in order to catch them only has a chance of succeeding when there are enough inspectors available to deal with the violators.

An additional result of this is that it has the potential to make inspectors fairer. When there are many offenses inspections will be fairer since inspectors can achieve higher catch rates. This means they will inspect the violating companies more often and leave the non-violators alone. When there is a very low amount of malpractice and the inspectors still work data-driven, the inspectors will still inspect the “higher risk” companies even when they are not offending. This could be argued for as unfair inspections since there is no proof that they are actually at higher risk. When there is little malpractice the risk models will not have accurate enough data to say anything meaningful here. This is another reason the model showed in favour of weighing more heavily on the compliance history when working data-driven compared to the identified risk indicators. This does, however, require that monitoring organisations build up their data. Random inspections could be argued for as more fair or less fair depending on if you believe “high risk“ companies should get inspected more, even if they do show the same level of compliance as other companies.

10.2 Variance in the Outcomes of Interest

As explained in the previous chapter the first step of the Sobol global sensitivity analysis was performed using all variables varied over their identified ranges. These ranges were given in chapter 6. The histogram below (Figure 10.2) shows the variance in the outcomes of interest during this experiment. What is immediately noticeable here is the compliance rate being 100% during a large part of the model runs. This can be mostly explained by the high attribution of the sector specific variables to the compliance rates (Table D.1). These four variables together are accountable for over 90% of the variance in the outcomes. This can explain the observed behaviour rather easily by following the model logic. When there are enough inspectors to inspect close to all companies or the fine and costs or compliance are not balanced this will have a large impact on the compliance rates. This can lead to a compliance rate of 100% which will also cause the observed compliance to be 100% and the difference between the total and observed compliance to be 0. These values for the number of inspectors, the number of companies and the fact that some companies make a consideration between the fine and costs of compliance are known within specific sectors and therefore we also ran these experiments with these numbers at fixed values (Figure 10.3).

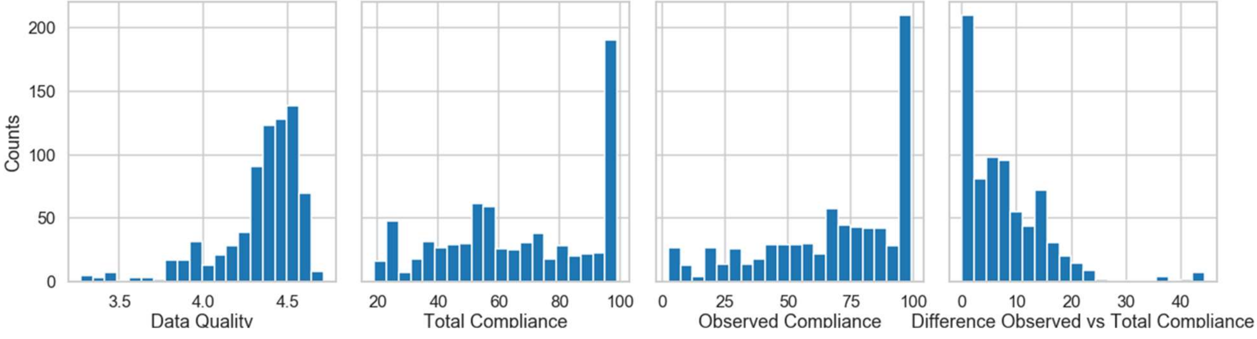

Figure 10. 2 Histogram of all outcomes of interest

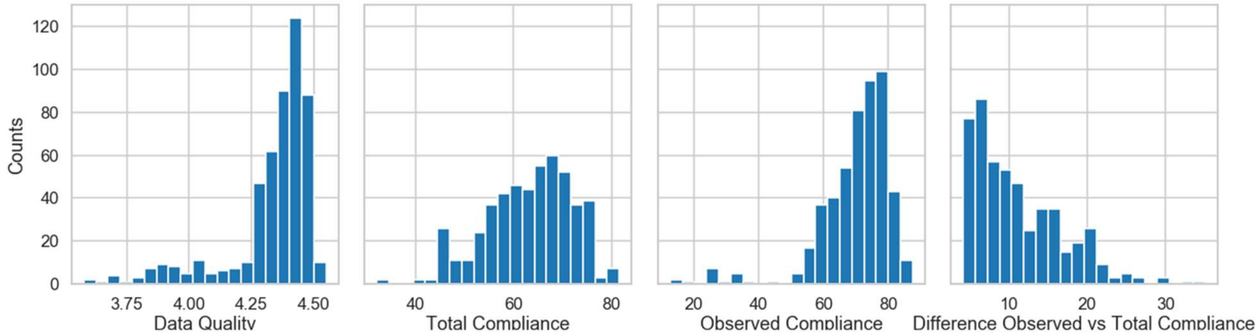

Figure 10. 3 Histogram of outcomes of interest with sector specific variables fixed

The difference between the two sets of experiments shows very clearly in these figures, with the second set having much less overall variance in the output parameters than the first.

Interesting is the fact that even though the sector specific variables account for more than 90% of the variance in the compliance rates, the percentage of data-driven inspection still has a large impact, of about 30%, on the quality of the data and the observed vs total compliance variables. This highlights the risks that come with data-driven inspections. Even when they do not increase the total compliance or the efficiency of monitoring organisations, they can still bias the data used by monitoring organisations. This impact of the percentage of data-driven inspections on the data quality is mostly an interaction effect with the number of inspectors and the strategic behaviour of companies.

The rest of the variance in the data quality, and the largest part, can be assigned to the number of inspectors when taking all uncertainties into considerations. The figure (Figure 10.4) below shows how sending low numbers of inspectors can have a large impact on the quality of the data. Here the effect also reduces when the number of inspectors increases past a certain point. This point is closely related to the range set for the number of companies. Allowing a larger range there would mean more inspectors are needed to secure higher quality data.

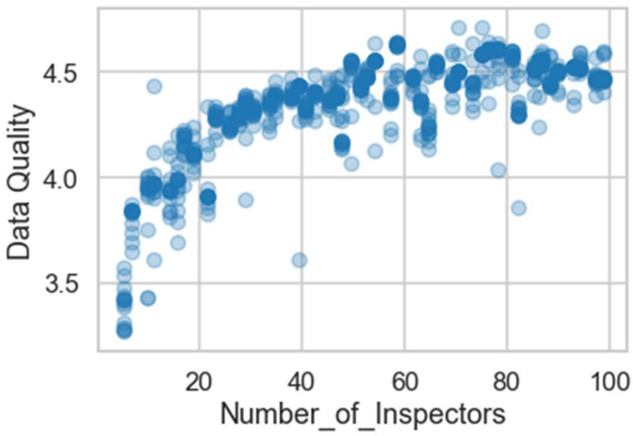

Figure 10. 4 Effect of the number of inspectors on the data quality including all uncertainties

The remaining variance in the observed vs total compliance is also mostly caused by interaction effects between the number of inspectors and the other variables, which makes sense. When monitoring organisations send a small number of inspectors, they can only visit a small part of the companies and their overview can get skewed heavily.

10.2.1 Random Seed

The global sensitivity analysis also showed the random seed always had an influence on the data quality. The other outcomes of interest were mostly unaffected by the random seed. This can be explained by the fact that the measurement of the data quality is quite a volatile one which can change quickly between timesteps. Only the last timestep of the results is analysed for the Sobol indices which causes the data quality variable to have some spread. The other variables, however, also affect the data quality and often times even more so than the random seed.

10.3 Sensitivity Analysis with Fixed Sector Specific Variables

In this subchapter the results from the Sobol sensitivity analysis without the sector specific variables will be given. During these experiments the policy options and other uncertainties were still being sampled over their full ranges as identified in chapter 6. Similar to the experiments with the reference scenario these experiments showed a large impact on the data quality when performing more than 80% of the inspections data-driven (Appendix F). Interesting is that the other uncertainties did not impact the data quality that much. They did, however, have an impact on the compliance rates, as can been sin in the figure below (Figure 10.5).

The time it takes for companies to reset their expectations after an inspection, so how long the impact of an inspection lasts for a company, has the largest impact on the overall compliance rates. Also, the effect of the inspections, and how much they increased the companies' estimations for that time had a large influence on the compliance rates. This shows us the importance of research on the effects of inspections on a company for different sectors and also on how to increase this impact as it can greatly help with compliance. These impacts could be caused by learning effects, financial incentives or deterrence effects. These experiments again suggest that data-driven inspections have a large impact on the outcomes about data quality, while the other factors influence mostly influence the other outcomes of interest, suggesting the amount of data-driven inspections should be carefully tuned to the desired level of data quality since that is the most sensitive outcome of interest.

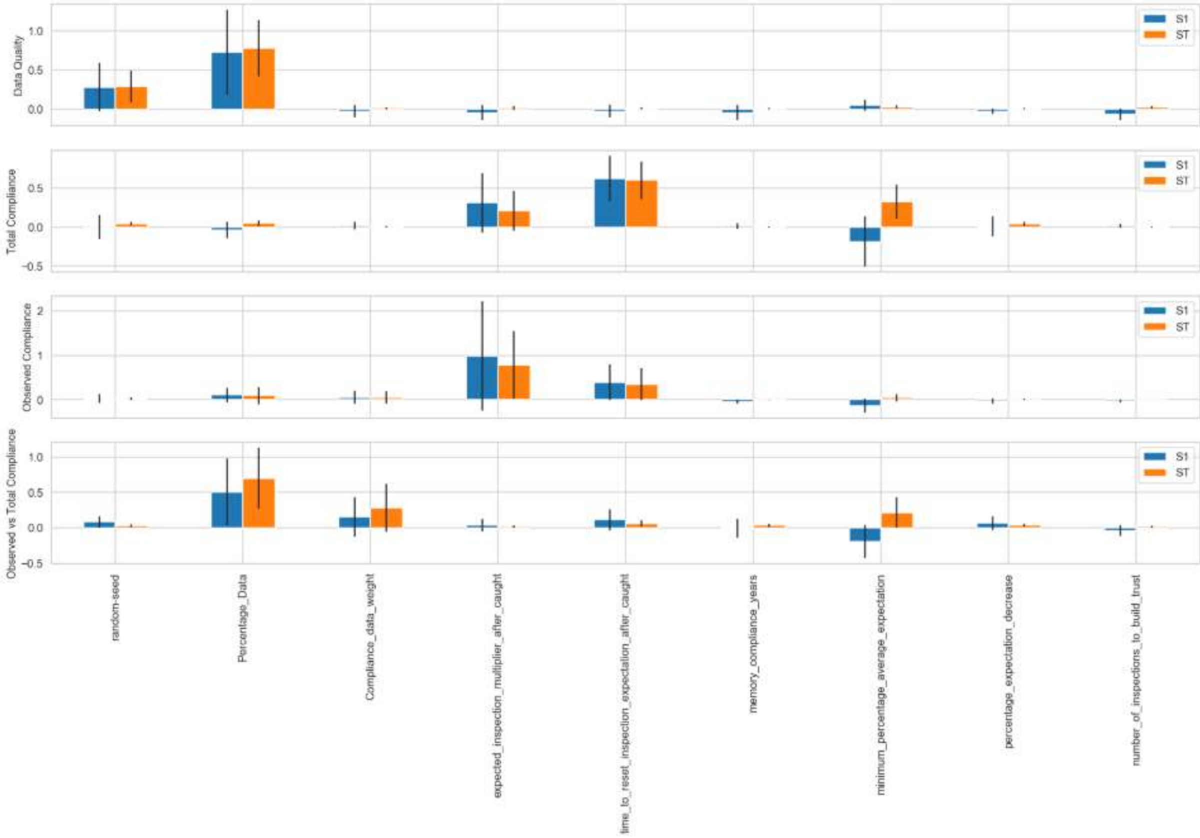

Figure 10. 5 Plot Sobol fixed sector specific uncertainties with confidence intervals

When removing the policies from this experiment the same results showed. The duration of the effect of an inspection on the companies is the most important variable for compliance rates in these experiments. The data quality is mostly random when the policy options are fixed, and is thus not influenced by any of the other uncertainties.

When keeping everything fixed except for the two policy options, we can see both have a similar impact on the outcomes of interest. However, the compliance data weight has a negative impact below a certain value. This means that when the monitoring organisations in the model did no use the compliance history to estimates risk, but based their risk estimations purely on the characteristics of the companies and the estimated correlations, they were off worse. The data quality is not impacted by the amount of compliance history used for the data-driven risk estimation. Again, the amount of data-driven inspections strongly correlates with the data quality at high percentages (Appendix F).

10.4 Trade-offs Between the Outcomes of Interest

The figure below (Figure 10.6) shows the trade-offs between the outcomes of interest for ten different percentages of data-driven inspections between 0 and 100. The light blue dots are 100% random inspections and the dark blue are fully data-driven. The yellow dots are 89% data-driven. There are 7 other policy options in this plot that are mostly hidden below the others. This again highlights the fact that there is a point or range where the long-term effects of data-driven inspections increase very strongly. Before this point, the effects on the trade-offs between the outcomes of interests are smaller. This plot shows a couple of things. The first thing it shows is that when working fully data-driven the outcomes can take much more extreme values and the spread is larger. This especially shows in the data quality, which is worse in many scenarios compared to working fully random and in the difference between the total and observed compliance. It also visualises trade-offs between different outcomes of interest. It shows how the observed compliance relates pretty closely to the total compliance except for the two highest data-driven inspection rates. The mismatch between observed and total compliance has the potential to get larger when the observed compliance is either on the high end or low end and is much worse when working with really high percentages of data-driven inspections. It also shows the compliance rates are more often over estimated than underestimated, but both can happen. The last notable observation is that there seems to be less correlation between the difference in observed and total compliance and the data quality than one would initially expect. Almost all of the correlation there can be appointed to the percentage of data-driven inspections, which has a large impact on both.

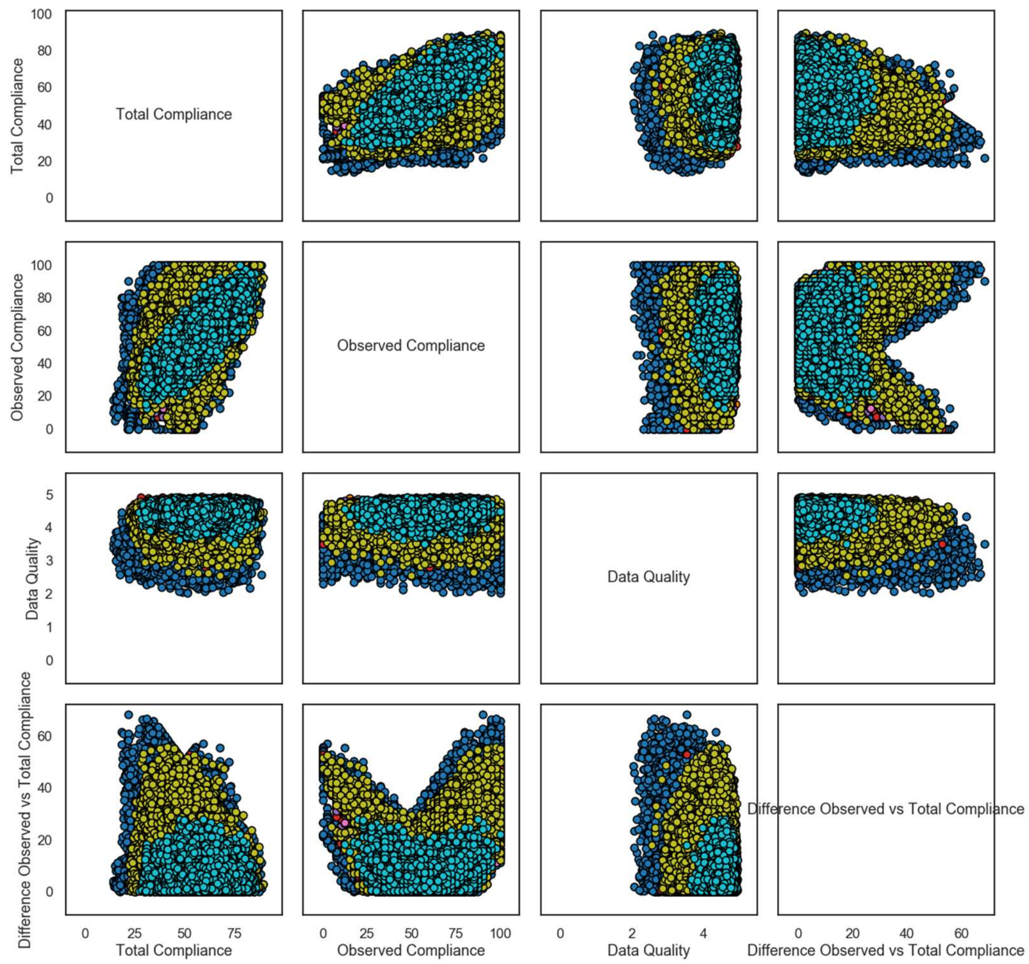

Figure 10. 6 Pairs Plot open exploration with 10 policy options

11. Conclusions and Recommendations

The main research question this study is trying to answer is:

How should monitoring organisations use data-driven risk-based inspections in order to utilize the long-term positive effects of the use of data while minimizing their long-term negative effects?

In order to answer this question several sub questions were created that contribute to answering the main question. In this section we will start by providing the gained insights for each of these questions before combining them and answering the main research question.

What is the effect of data-driven inspections on the quality of the used data?
The quality of data is one of the main long-term issues monitoring organisations face when working with risk-based data-driven decision-making. In this research this is defined as the ability of the data available to monitoring organisations to estimate risk factors and the level of bias their data contains. The results show that when monitoring organisations only inspect high risk companies, their data quality will decrease severely, both the risk estimations and the perceived compliance rates will be observed wrongly. This will be the case when data models send inspectors to companies that are estimated to have a higher risk of offending.

Furthermore, the results show that the quality of data is a very sensitive factor compared to other outcomes of interest such as the compliance rates. Data-driven inspections have the potential to increase efficiency and/or compliance, but at the same time the quality of data collected during inspections might decrease more and outweigh the positive effects. Especially when working close to fully data-driven, where the data tells the inspectors which locations they should visit, the data quality suffers. A small amount of random inspections can significantly increase the quality of the data collected by the inspectors, contributing to less bias, preventing monitoring organisations from getting stuck on a certain subset of companies.

Another effect data-driven inspections can have when they aim to improve efficiency is that monitoring organisations might choose to send fewer inspectors when they are able to achieve the same compliance rates while adding the use of data analysis to their work. However, our results show that fewer inspections can also severely decrease the quality of data and the bias in observed compliance rates. This can be extra dangerous in combination with data-driven inspections, where there is already a data bias in the selection of the companies, impacting the data quality as explained above. Therefore, the effect of data-driven inspections should preferably not be a decrease in the number of inspectors.

The last effect of data-driven inspections on the quality of data this study showed was that there are different kinds of data quality issues. We discussed how data-driven inspection can lead to biased data and a skewed overview of the sector. Next to this our results showed that the data can not only be biased towards high risk companies but also towards certain characteristics. This study found the effect of data-driven monitoring on the data quality to foremostly affect the estimation of characteristics that were not initially used for the risk-analysis. Company characteristics with a low correlation with their propensity to violate, or initially low estimated correlation, will be underrepresented when inspecting data-driven. When these initial estimations are wrong, or the correlations change over time, data analysts might not pick up the importance of these characteristics since most of the inspections will be performed at companies that score high on other characteristics. Even when monitoring organisations build a new risk model after every inspection round, they will be slow to pick up changes in these low correlation, or lowly estimated correlation, characteristics.

How should random inspections be used in relation to fully data-driven risk-based supervision in order to utilize the positive effects of the use of data while minimizing the long-term negative effects?

Our model showed that random inspections play an important role in safeguarding data quality and preventing a data bias. Full data-driven monitoring will result in a much worse overview of the sector compared to having a small amount of inspections performed at random. The first random inspections will add a lot to the compliance overview. When adding more random inspections it will reach a point where the added value to the observer bias is very small. This study showed similar results for the effect of random inspections on the quality of the data.

In conclusion, a small percentage of random inspection can greatly increase both the data quality and reduce the mismatch between the observed and total compliance. The positive effect of data-driven inspections on the compliance rate does not seem to increase significantly at this percentage and is steady before and after this point. This suggests there is an optimum percentage of random inspections where monitoring organisations can have a large part of the positive effects on the compliance rate without completely ignoring their data quality or compliance overview. Because this point, or range, exists, random inspections should be used to find this point while keeping the focus on data quality and a potential bias instead of improvements in compliance. Finding an optimal solution would be to find the point where the data quality starts to get significantly worse. Our experiments show the improvements in compliance are much more linear. The initial results of this study suggest this optimum percentage could be around 10 to 20%, depending on different assumptions and sectors. One of the factors that seems to impact this optimal point is the initial compliance rates of a sector. In a low compliance rate sector with relatively many companies, the drop off point for the data quality seems to be at a higher percentage of data-driven inspections. At the same time, however, as mentioned before, in sectors with low amounts of inspectors the data quality can already suffer. This would suggest that large sectors, both in number of companies and inspectors, with low compliance rates allows for the highest percentage of data-driven monitoring. More research on this is recommended.

How much does the efficiency perceived by monitoring organisations differentiate from the real efficiency?

Inspection efficiency is a concept that is very hard to define and measure even in simulations. Instead, this study looked at the compliance rates our agents observed versus the total simulated compliance rate in an effort to say something about the perceived versus real efficiency. The number of inspectors in this model is equal to the number of inspections and could be seen as the capacity of a monitoring organisation. The performed experiments show that the difference between the observed and the total compliance has a potential to be three times as large when working fully data-driven compared to performing fully random inspections (a maximum of 20 percentage points compared to 65 percentage points). Furthermore, the results highlight that a really high or low observed compliance is more likely to be false, and with a bigger mismatch, than average or expected observed compliance rates. The difference between the perceived and real compliance does not seem to have a strong direct relation with the data quality as one could expect. This study showed that the observable correlation is mostly caused by a shared influencer, the percentage of data-driven inspections.

Data-driven inspections have an impact on the mismatch between the observed and real compliance rates. Since these compliance rates are used to estimate efficiency, data-driven inspections will also affect these estimations. Depending on the strategy of monitoring organisations this impact could both over or underestimate the compliance. Our experiments show that more often the compliance rates are overestimated by monitoring organisations, but both can happen. The results of this study show the importance of the monitoring organisations' strategy here. Policy such as an extra inspection after a year when a company has been caught, or leaving them alone for one or two years when they show good behaviour will have a large effect on the observed compliance rates. The level of strategic behaviour of companies is the next largest factor on the observed versus total compliance rates. Our results show that they mostly effect the total compliance, the perceived compliance will be largely unaffected, thus increasing the mismatch. The use of data increase both the strictness of the policies, leaving less room for random inspections as well as the level of strategic behaviour, as this study showed, thereby increasing the issue of observed versus real efficiency.

How fair are data-driven inspections to the companies involved regarding administrative burdens?

Conclusions about the fairness of data-driven inspections are difficult to make and should go further than distributions about which companies are inspected. What this study did show is that high incentives for malpractice increases the ability of monitoring organisations and in particular that of data analysts to filter out high-risk companies. When these incentives are high enough or when the strategic behaviour of companies is limited, inspectors can achieve a higher catch rate and at the same time pay less visits to compliant companies. This could be interpreted as "fair" inspections since the non-compliant companies get inspected more than the compliant ones which could be argued for as fair. It has also been argued that structural use of data also improves fair treatment by removing bias introduced by the inspectors during their inspections. The use of data could, however, also lead to less fair inspections if the data source gets more biased over the years.

Whenever the rate of malpractice is very low and monitoring organisations choose to perform their inspections based on their data, this could lead to arguably unfair inspections where companies with the same compliance rates get inspected way more than others. This can happen when the compliance in general is very high, when models cannot accurately predict where inspections are needed and still select "high-risk" companies, which score slightly higher on their risk indicators, which could be inaccurate because of the lack of data on non-compliant companies.

What is the effect of transparency on the strategic behaviour of companies?

Working with data could improve justification by monitoring organisations. Disclosure of risk models and inspection data to the outside world is mentioned as one of the benefits of working data-driven. At the same time, however, companies could misuse a more transparent way of inspection target selection. The experiments performed for this study showed an increase in strategic behaviour from companies when monitoring organisations choose to work data-driven compared to random. This is even without the added effect of transparent algorithms. This increase in strategic behaviour, even without modelling the effect of transparent models, is bad for the compliance overview of monitoring organisations. This is even worse when combined with the use of risk-based data-driven inspections, which caused the increase in the first place. It should be noted that this increase in strategic behaviour is in comparison with truly random inspections. When monitoring organisations inspect without the use of data, they can still work based on risks identified by inspectors. This way of risk-based inspections might already lead to a higher rate of strategic behaviour and thus the difference when working fully data-driven might be smaller than our results show here.

How can inspectors be directed based on an objective outcome instead of input (hours) and what will be the consequences of this?

A model extension to implement outcome-based direction of inspectors still has to be made. However, the current model shows the large impact the number of inspectors has on the quality of data. This means that when data-driven inspections are used for increasing efficiency and thereby potentially reducing the number of required inspectors, monitoring organisations should be extra careful as the combination of these two are destructive for the quality of data collected by inspectors. Desired outcomes should potentially include something like high quality of data or a minimum number of random inspections or inspections at compliant companies.

How should monitoring organisations use data-driven risk-based inspections in order to utilize the long-term positive effects of the use of data while minimizing their long-term negative effects?

The main research question can be answered by taking the insights about the long-term use of data-driven inspections together and use them as a basis for our recommendations. First additional insights that did not fit any of the above sub questions will be given. After that we conclude by giving recommendations on how to deal with the medium- to long-term effects of risk-based decision-making with the use of data, answering our initial research question.

The first additional insight gotten from this study that can help answer our research question is that there are sector specific variables, which are unrelated to the strategic use of data by monitoring organisations, but have a very large impact when looking at the outcomes of interest. Knowing these sector characteristics is an important first step in making effective data-driven policy. After these variables are known and fixed, the effects of a potential strategy become clearer. These variables include the number of inspectors, the number of companies, the height of a potential fine and the costs of compliance. They show that in sectors with low compliance rates much more is to be gained by working data-driven. Larger improvements in compliance rates were observed here. At the same time the negative effect of the loss in data quality by working data-driven was smaller in sectors with a low compliance rate. In sectors with already high compliance rates, working data-driven still has a positive effect on the compliance rate, but it is smaller. The quality of data, however, gets significantly worse inspecting data-driven in sectors with high compliance rates since inspectors will only try to visit a, non-representative, relatively small group of non-compliant companies.

Secondly, the results of the analysis show the value of data about the compliance history of companies. When monitoring organisations have access to good data about historical compliance of the companies under supervision, this improves their ability to perform data-driven inspections. This again highlights the need for monitoring organisations to keep any potential biases in their data as small as possible and compensate with random inspections. Otherwise the quality of compliance rates will suffer and the power of the data models will decrease, while many of the negative long-term effects we discussed earlier will still be developed.

Recommendations

- The first recommendation to monitoring organisations that follows from the results of this study is to never work completely data-driven. Sending inspectors 100% based on data models will have a large negative impact on the perceived compliance and data quality in the medium- to long-term. At the same time, it removes the narrative needed for understanding their risk models and justification to the outside world. These issues will at the same time make it very difficult to learn from performed inspections and are potentially creating unfair treatment to involved companies. By performing a "small amount" of random inspections these issues are largely resolved. The number of this small amount of inspections depends on the sector which is being supervised and the strategy of the monitoring organisations and requires further research. Random inspections do not take away much from the ability to justify inspection targets when they are performed next to data-driven risk-based inspection instead of risk-based inspections without the use of data.
- The second recommendation is to be very careful about estimating compliance rates when working with data collected through data-driven inspections. Monitoring organisations should strive to keep this data separate from data collected from random inspections and at least be aware of the potential bias that exists in their data.
- An increase in data-driven inspections will increase the amount of strategic behaviour from companies. Both the increase in data collected through data-driven inspections and the increase in strategic behaviour will negatively impact data quality. For this reason, it is important for monitoring organisations to learn about the strategic behaviour of the companies in their sector, and how they react to inspections, both when working with and without data, transparent and closed. This knowledge is required in order to find an optimal amount of data-driven inspections. While studying the behaviour of companies, monitoring organisations might even find ways to improve the effectiveness of their inspections, making the effects on company bigger or last longer, which showed to be of great importance to the overall compliance levels of a sector.
- The use of data-driven inspections does have a positive effect on the total compliance, even in the long-term. However, as reported before, at a certain percentage of data-driven inspections the negative effects will increase very rapidly. For this reason, monitoring organisations should fine-tune their level of data-driven inspections based on their desired quality of data, overview of the sector or level of justification, with their compliance and efficiency goals as secondary objectives. When these goals cannot be met with a reasonable amount of random inspections it might be better to look at alternative solutions to data-driven inspections, for example, by increasing the number of inspectors.
- When observed compliance rates are exceptionally high or low while most of the inspections are performed with the use of data, monitoring organisations should be extra aware of the potential bias their data shows and / or the strategic behaviour their inspections are causing.
- Monitoring organisations could best first implement data-driven risk-based inspections in sectors which currently have a relatively low observed compliance rate. This increases the potential increase in effectivity and at the same time limits the negative

effects of the creation of a data bias. It would be even better if this is a sector where compliance rates of the companies are already available and there is a decent number of inspectors available.

- Observed compliance rates also depend greatly on other policy executed by monitoring organisations. They should be aware, and potentially compensate for this when estimating compliance levels. This includes policy such as extra inspections after a negative inspection and not inspecting companies for a while after a positive one.

12. Further Model Use and Research

The model developed during this study has helped us gain insights about the long-term effects of data-driven inspections both during the model development phase and by running experiments and analysing the results. Next to this, the development and use of the model also generated questions for further research and has additional purposes that will be discussed in this chapter.

12.1 Additional Model Use

The most obvious further use of the created NetLogo model is explore more strategic policy options related to data-driven monitoring. Monitoring organisations can develop plans and evaluate them using the model. Some of these strategic options might require additional modelling to extent the boundaries of the model. Some potential extensions will be discussed later.

Additionally, the NetLogo model can be used during conversation with inspectors to help with the transition to a more data-driven approach. The NetLogo interface can be used to help structure the conversation with the involved parties. Showing inspectors the impact the use of data can have on the system as a whole and on the behaviour of companies can be convincing and open up the discussion with their strategic managers. This way, management and inspectors can see the effect certain ways of working can have on their desired outcomes of interest and how it might influence their work. This can help in convincing the right people. It could also be used to show the fact that trade-offs need to be made and the system is more complex than one might initially suspect.

12.2 Further Research

The experiments performed during this study unveiled some additional important research needed to optimally perform inspections using data. Firstly, it was shown that monitoring organisations could benefit from information about strategic behaviour from companies in their specific sector for making effective data-driven policy. Next to this, the model showed the importance of the effect of inspections on the companies involved. A stronger and longer lasting effect can significantly improve the overall compliance rates of a sector. This can be done through both detterence and learning effects. Many researchers have already explored this topic. What is still lacking is knowledge from the monitoring organisations themselves on what kind of companies they are dealing with and how they will behave when being exposed to data-driven regulation.

This study showed the existence of an optimal percentage, or range, of data-driven inspections. Further research is needed to find more accurate values for this optimum. Research should also be conducted on how this optimum is influenced by different policy options, different types of violators and sector specific variables. The model created for this study could provide a good starting point for this kind of study. Our results suggest that the initial compliance rate of a sector has a large impact on this optimal percentage.

13. Reflection

13.1 Scientific Relevance

Monitoring organisations have been working risk-based for a long time now. Most of the time, however, they are not yet making use of the available data in order to estimate their risks. Many have seen the potential benefits the use of data can offer for these organisations. At the same time multiple risks are mentioned. These risks are, however, unexplored. The same holds for the potential positive long-term effects of data-driven inspections. This study listed these long-term effects and went further in exploring them, generating insights about how they are created and giving recommendations on how to deal with them.

The modelling approach used for this study proved useful for the exploration of long-term effects of data-driven monitoring. Agent-based models have been used before to study inspection games but not with the added use of data by the inspectors. This study showed what this approach is capable of in this research field and suggests additional model extensions and further exploration with these types of complex models. Our model was capable of recreating results created with the use of inspection games found in literature. The model can be used to search for an optimum number of inspectors under fixed assumptions, similar to how inspection games are traditionally often used, which is not within the scope of this study. With our agent-based implementation of an inspection game we were able to add dynamic behaviour to both the inspectors and inspectees, changing their behaviour over time depending on their inspection history.

With this study we confirmed and showed some of the long-term effects mentioned during the literature study and expanded on these. This study demonstrates the issue of the creation of a bias in inspections data and shows inspectors might get locked in on a couple of companies under some assumptions and scenarios. The notion that the only way to effectively tackle corporate fraud is with consistent inspections (Schell-Busey et al., 2016) was also highlighted during this research. Using data for finding the companies is only the first step, after this, our model showed structural inspections are needed to improve compliance, which was also confirmed by an expert during personal communications.

As our literature study showed, it is difficult to estimate how successful data-driven inspections will be and how they can increase effectiveness. Even with the use of our model it is difficult to validate how effective a risk-model can be in the real world because many factors have to be considered. We confirmed the fact that adding unbiased data to a biased data source can already greatly mitigate the long-term issues mentioned in literature. This was shown before by Jacobusse and Veenman (2016). Our study shows and confirms the importance of adding random data to the collected data.

As Sparrow (2000) describes, it is likely an intermediate step will first exist in a potential transition towards data-driven inspections. In this scenario where inspectors evaluate a list of risk indicators provided by data analysts our results and the mentioned long-term effects still hold, only to a possible lesser extent. A bias by inspectors could, however, increase the bias in collected data even more. More research on this way of working is needed. As a lot of literature argues for, the collaboration between data scientists and inspectors is important and will have a, currently unknown, effect on the bias of the collected data. Our study highlighted this and provided additional recommendations in order to better recognize a data bias, how to limit the bias and insights on how to interpret collected data. This allows inspectors to better collaborate with data scientists and identify potential risks themselves and feed this back to the data scientists.

13.2 Model Improvements

This study was performed within a 25-week period. During these weeks, choices had to be made on what to include and what to leave out for further research. Below we will discuss some additional improvements that could be made to further explore this research topic.

13.2.1 Possible Model Extensions

In the real world a large part of the work of monitoring organisations is based on alerts. They inspect companies based on alerts from both internal and external sources. This could be a useful addition to the model in order to see how they affect the capacity to work data-driven under different circumstances.

The use of external data sources is also not considered during this study. The addition of external sources could mitigate the bias created in inspection data by adding unbiased data. The use of these sources could still lead to a biased collection of data but would reduce the feedback loop where the data gets supplemented with only the inspection data. External data can however also be biased. This can be extra risky since monitoring organisations have less knowledge about the possible biases these data contain compared to their own inspection data.

The effects of different types of violators can also be extended in this model. Currently, the combination of propensity to violate with the ratio between the fine and the costs of compliance can generate many possible violators as they are described in literature. For the experiments in this study, the distribution of the propensity to violate was kept fixed but the other variables were varied. Explicit modelling of each of these violators would require the modelling of different measures or punishments and their effects on each of these violators. As this was a study focused on the use of data, and not on the compliance and strategic behaviour of different types of companies, this was not currently the scope, this could, however, still be interesting for further research. This way insights could possibly be gained about each of these types of violators and how they influence the outcomes and therefore the sector and desired strategy.

Data-driven Risk-based Inspections versus Risk-based Inspections Without Data
Risk-based inspections without the use of data is also something that is currently happening in the real world which is not implemented in this model version. The current way of risk-based inspecting could, as mentioned before, already add a bias to the data and improve effectivity but does not help with increasing the ability to learn from inspections or justifying them for monitoring organisations. This concept of risk-based inspections without the use of data is one that proved difficult to model because it is difficult to quantify and put the tacit knowledge of inspectors in to a model. Exploring this further, however, might lead to some additional insights and can help compare risk-based inspections with and without the use of data.

Next to the increase in the ability to justify inspections and learn from them, the use of data also allows monitoring organisations to store knowledge. When working with tacit knowledge from inspectors, monitoring organisations depend on these inspectors for their knowledge which is difficult to transfer, whereas the use of data allows them to store this knowledge effectively and easily transfer it between many data scientists. The use of data can also increase fairness according to Baldwin and Cave (1999), when inspections become more structured instead of semi-random.

The bias in collected data that the risk-based inspections without the use of data are creating at the moment currently only leads to a small part of the identified issues as mentioned above because their next inspection rounds are unaffected by this collected data. When monitoring organisations, however, start using this information for their risk analysis these issues may get bigger when they have to work with a skewed overview of the compliance of companies in their sector from the last years. The bias effects of risk-based inspections become more extreme when adding data, for both the positive and negative effects, with the use of data having some extra advantages over the use of tacit knowledge, if their predictive power is the same.

13.2.2 Additional Strategies for Monitoring Organisations

The option to come back to a company after they show non-compliant behaviour during an inspection is in the model. Because of the large impact this has on the measured efficiency and it being outside the scope of data related inspections, this has not been used in the performed experiments. For future research, it is useful to see how this affects the outcomes of interest and how monitoring organisations can prevent this from creating a bias in their data. The same thing can be said about monitoring organisations leaving compliant companies alone for a fixed period of time after a positive inspection. The model contains both this option, and the option for inspectors to learn which companies are more compliant using data analysis which results in similar behaviour. For this reason, and the aforementioned reason of having a large effect on the observed efficiency, only the second option was used during this study. Both of these policy options bring additional complexity in strategic behaviour and what inspectors would want to see during inspections.

Because this is an agent-based model that runs over a longer time, created in order to explore long-term effects, dynamic strategies can be added. It could be useful to see for example, the effect it has when monitoring organisations first do a couple of random inspections before starting to work data-driven. Another example of a dynamic strategy that could be further explored is that of outcomes-based inspections. Where the amount of inspections every round is based on the desired observed outcomes. They could also base their amount of data-driven inspections on other factors to test more complex strategies that might be more robust. This way one could also more extensively study the effect of sending inspectors based on outcomes.

For this study one reference scenario was used based on a real-world case with their policy options and parameterisation in mind. For further research it is interesting to contrast this to other cases and compare more in depth how the different policy options perform under the specific cases instead of the generalisation that was made in this study. Using the performed experiments, we were able to get insights about specific sector variables and how they influence each other. However, no direct comparisons or sector specific recommendations could be given based on these results.

13.2.3 Additional Outcomes of Interest

The experiments of this study were performed using four outcomes of interest. Additional indicators could be added to perform an analysis on. The model already contains variables about the distribution of inspections for example but many more options are thinkable.

13.2.4 Exploring the Effect of Correlations and Predictive Power

The correlations between the characteristics and the propensity to violate of companies that were implemented in this model were not extensively experimented with during this study. They were sampled to see their effect on outcomes of interest for the reference scenario. Next to this short analysis, more in depth exploration of these correlations is possible when more time and computing power allows for it. Our quick analysis for example showed that higher correlating factors lead to a better risk estimation by the inspectors. A hypothesis is that this can lead to a larger bias, when this estimation is used by the inspectors (lower compliance data weight). When working only with random inspections these correlations do not matter. Different combinations of these correlation factors can prove worth studying. They can represent different levels of power a risk model has and might result in interesting behaviour from the inspectors when working data-driven. For this study we assumed there was at least some kind of correlation between some of the factors to be able to explore the effects of data-driven risk-based inspections.

An alternative option is to use real-world risk models implemented in R and connect them to our model using RNetLogo as is discussed in the next chapter. The effects of tools and analysis that knowingly try to work with biased data also has to be further explored. This might mitigate some of the long-term risks our model showed and reduce the amount of required random inspections.

13.3 Model Reflection

Next to the list of assumptions given in Appendix A, there are some additional issues and concession had to be made. The first one is the way the monitoring organisations use data to choose their targets. Initially we tried implementing risk models in R and connecting them to NetLogo through the RNetLogo extension. Because of some unresolved issues that took too much time to solve, it was decided to implement a hand-made solution in NetLogo directly. It would be interesting to look at how the use of a Bayesian belief network or decision tree implemented in R would affect the results found in this study.

The EMA Workbench used during this study offers more exploration tools than the ones that were used during this study. These tools could still prove to be of added value for creating additional experiments. The first would be performing a PRIM analysis to explore where (un)desired outcomes come from and which combinations of input parameters allow these to occur. Directed search is another tool the EMA Workbench offers to optimize over both the policy options or uncertainties. This could be a very useful first way of further exploring an optimum and/or robust percentage of data-driven inspections.

An issue we ran into using the EMA Workbench with NetLogo was that the workbench did not recognize the returned outcomes of the experiments as a time series but instead as 100 different experiments for some of the tools. This limited the options for this study given the time, but could be addressed and adjusted for future research.

In general, most of the experiments performed for this study looked only at the outcomes of interest at the last timestep. Patterns about the development of the outcomes of interest can be of added value, especially since this is a study about longer term effects. Therefore, many experiments have been recreated manually in the NetLogo interface or ran 1 by 1 through PyNetLogo to explore these patterns and explain the behaviour. Additional research, with the use of tools that analyse these patterns in a more designed experimental way could perhaps provide additional insights.

13.4 Connection with EPA Master's Programme

This study has been executed in partial fulfilment of the Engineering and Policy Analysis master's programme. This programme focusses on supporting decision-makers in grand challenges in the public domain while taking into account the socio-economic and political environment in which they are embedded. This study used EPA methods for analysing and exploring the long-term effects of the use of data by monitoring organisations. An agent-based model was created in an attempt to fully capture the system's complexity and all its dynamics. Using this model, relevant policy recommendations were made to decision-makers facing challenges in the monitoring industry. A multi-actor perspective was used in order to give relevant advice. Pressure from outside governmental organisations and the public, and the need for the ability to justify inspection choices, play a large part in the use of data by monitoring organisations, which is taken into account during this study. The intricate relation inspectors have within a monitoring organisation has also been analysed in an attempt to create support for potential policy options. Different possible perspectives of companies and the complexity of the system in which monitoring organisations operate have been modelled in order to generate relevant policy advice.

Bibliography

Albino, V., Garavelli, A. C., & Gorgoglione, M. (2004). Organization and technology in knowledge transfer. *Benchmarking*, *11*(6), 584–600.

Alexander, P. A., & Judy, J. E. (2008). The Interaction of Domain-Specific and Strategic Knowledge in Academic Performance. *Review of Educational Research*, *58*(4), 375–404.

Alves, L. G. A., Ribeiro, H. V., Lenzi, E. K., & Mendes, R. S. (2013). Distance to the Scaling Law: A Useful Approach for Unveiling Relationships between Crime and Urban Metrics. *PLoS ONE*, *8*(8), e69580.

Andreozzii, L. (2010). Inspection games with long-run inspectors. *European Journal of Applied Mathematics*, *21*(4–5), 441–458.

Baldwin, R., & Black, J. (2016). Driving priorities in risk-based regulation: What's the problem? *Journal of Law and Society*, *43*(4), 565–595.

Baldwin, R., & Cave, M. (1999). *Understanding Regulation: Theory, Strategy, and Practice*. Oxford University Press.

Bayamlıoğlu, E., & Leenes, R. (2018). The 'rule of law' implications of data-driven decisionmaking: A techno-regulatory perspective. *Law, Innovation and Technology*, *10*(2), 295–313.

Bollier, D. (2010). *The Promise and Peril of Big Data*. *Communications and Society Program*.

Brown, J. S., & Duguid, P. (2000). Balancing Act: Capturing Knowledge Without Killing It. *Harvard Business Review*, *78*(3), 73–80, 212.

Burnett, S., Carthey, J., & Vincent, C. A. (2013). The Measurement and Monitoring of Safety. *The Health Foundation*.

Cliff, G., Wall-Parker, A., Cliff, G., & Wall-Parker, A. (2017). Statistical Analysis of White-Collar Crime. *Oxford Research Encyclopedia of Criminology and Criminal Justice*.

Coglianese, C., & Sapir, A. (2017). Risk and Regulatory Calibration: WTO Compliance Review of the US Dolphin-Safe Tuna Labeling Regime. *World Trade Review*, *16*(2), 327–348.

Dam, K. H. Van, Nikolic, I., & Lukszo, Z. (2012). *Agent-based modelling of socio-technical systems*. Springer.

de Laat, P. B. (2017). Big data and algorithmic decision-making. *ACM SIGCAS Computers and Society*, *47*(3), 39–53.

Divan, M. J. (2018). Data-driven decision making. In *2017 International Conference on Infocom Technologies and Unmanned Systems: Trends and Future Directions, ICTUS 2017* (Vol. 2018-Janua, pp. 50–56).

Ezingeard, J.-N., Leigh, S., & Chandler-Wilde, R. (2000). Knowledge Management at Ernst & Young UK: Getting Value Through Knowledge Flows. *Social Epistemology*, *15*(3), 201–213.

Ferguson, A. G. (2012). Predictive Policing and the Future of Reasonable Suspicion. *SSRN Electronic Journal*, *62*.

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., … DeAngelis, D. L. (2006). A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*, *198*(1–2), 115–126.

Hajian, S., Bonchi, F., & Castillo, C. (2016). Algorithmic Bias: From Discrimination Discovery to Fairness-aware Data Mining. *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2125–2126.

Herman, J., & Usher, W. (2017). SALib: An open-source Python library for Sensitivity Analysis. *The Journal of Open Source Software*, *2*(9), 97.

Hutter, B. M. (2005). *The Attractions of Risk-based Regulation: accounting for the emergence of risk ideas in regulation*. *CARR Discussion Paper*.

Jackson, J. L., & Bekerian, D. A. (Eds.). (1997). *Offender profiling: Theory, research and practice. Offender profiling: Theory, research and practice.* Hoboken, NJ, US: John Wiley & Sons Inc.

Jacobusse, G., & Veenman, C. (2016). On selection bias with imbalanced classes. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 9956 LNAI, pp. 325–340).
Jaxa-Rozen, M., & Kwakkel, J. H. (2018). Pynetlogo: Linking netlogo with python. *JASSS*, *21*(2), 4.
Kagan, R. A., & Scholz, J. T. (2013). The “Criminology of the Corporation” and Regulatory Enforcement Strategies. In *Enforcing Regulation* (pp. 67–95). Wiesbaden: VS Verlag für Sozialwissenschaften.
Kalai, E. (2004). Game theory: Analysis of conflict. *Games and Economic Behavior*, *3*(3), 387–391.
Kwakkel, J. H. (2017). The Exploratory Modeling Workbench: An open source toolkit for exploratory modeling, scenario discovery, and (multi-objective) robust decision making. *Environmental Modelling and Software*, *96*, 239–250.
Laffont, J.-J., & Martimort, D. (2018). *The Theory of Incentives. The Theory of Incentives*. Princeton University Press.
Lipsky, M. (2010). *Street-Level Bureaucracy: Dilemmas of the Individual in Public Service, 30th Anniversary Expanded Edition*. Russell Sage Foundation.
Liu, C., White, M., & Newell, G. (2013). Selecting thresholds for the prediction of species occurrence with presence-only data. *Journal of Biogeography*, *40*(4), 778–789.
Liu, F. C. S. (2011). Validation and agent-based modeling: A practice of contrasting simulation results with empirical data. *New Mathematics and Natural Computation*, *07*(03), 515–542.
Louie, M. A., & Carley, K. M. (2008). Balancing the criticisms: Validating multi-agent models of social systems. *Simulation Modelling Practice and Theory*, *16*(2), 242–256.
Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology*, *44*(1), 190–204.
Nonaka, I., & Toyama, R. (1998). A theory of organizational knowledge creation. *Knowledge Creation Diffusion*, *11*(7/8), 25472–25472.
Nosenzo, D., Offerman, T., Sefton, M., & Van Der Veen, A. (2014). Encouraging compliance: Bonuses versus fines in inspection games. *Journal of Law, Economics, and Organization*, *30*(3), 623–648.
Perc, M., & Szolnoki, A. (2015). A double-edged sword: Benefits and pitfalls of heterogeneous punishment in evolutionary inspection games. *Scientific Reports*, *5*(1), 11027.
Picoli, S., Castillo-Mussot, M. del, Ribeiro, H. V., Lenzi, E. K., & Mendes, R. S. (2015). Universal bursty behaviour in human violent conflicts. *Scientific Reports*, *4*(1), 4773.
Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data*, *1*(1), 51–59.
Raick, C., Soetaert, K., & Grégoire, M. (2006). Model complexity and performance: How far can we simplify? *Progress in Oceanography*, *70*(1), 27–57.
Rauhut, H. (2015). Stronger inspection incentives, less crime? Further experimental evidence on inspection games. *Rationality and Society*, *27*(4), 414–454.
Routledge. (2013). Privacy, Due Process and the Computational Turn: The Philosophy of Law Meets the Philosophy of Technology (Hardback) - Routledge.
Saltelli, A., & Annoni, P. (2010). How to avoid a perfunctory sensitivity analysis. *Environmental Modelling and Software*, *25*(12), 1508–1517.
Schell-Busey, N., Simpson, S. S., Rorie, M., & Alper, M. (2016). What Works?: A Systematic Review of Corporate Crime Deterrence Schell-Busey et al. Corporate Crime Deterrence. *Criminology and Public Policy*, *15*(2), 387–416.
Sharma, G., Bhargava, R., & Mathuria, M. (2013). Decision_Tree_Analysis_on_J48_Algorithm. *International Journal of Advanced Research inComputer Science and Software Engineering*, *3*(6), 1114–1119.
Shimshack, J. (2014). The Economics of Environmental Monitoring and Enforcement. *Ssrn*.
Smojver, S. (2012). Analysis of Banking Supervision via Inspection Game and. *Central European Conference on Information and Intelligent Systems*, 355–361.

Sparrow, M. K. (2000). *The Regulatory Craft: Controlling Risks, Solving Problems & Managing Compliance*. Brookings Institution Press.
Taylor, P., & Avenhaus, R. (2010). Applications of inspection games. *Mathematical Modelling and Analysis*, *9*(November 2014), 37–41.
von Stengel, B., & Zamir, S. (2010). Leadership games with convex strategy sets. *Games and Economic Behavior*, *69*(2), 446–457.
Waldron, J. (2010). *The Rule of Law and the Importance of Procedure*. *SSRN*.
Xiang, X., Kennedy, R., Madey, G., & Cabaniss, S. (2005). Verification and Validation of Agent-based Scientific Simulation Models. In *Proceedings of Spring Simulation Multiconference* (p. 11).

Appendix A – Assumptions and Limitations

This appendix contains a list of assumptions and a list of limitations. These needed to be made in order to model our problem. They do have an impact on the outcomes of our study and it is important to make them transparent and keep them in mind.

List of assumptions

- There is a correlation between the characteristics and the propensity to violate of companies.
- The company characteristics are visible to the outside world. On inspection their compliance and propensity to violate are learned.
- The inspectors have an initial estimation of the correlation between the characteristics and the propensity to violate.
- The inspectors have the legal authority to fine the companies.
- Companies expect to get inspected more often after getting inspected.
- Companies do not want to get inspected when they do not comply with the rules.
- Companies in general prefer to show compliant behaviour when getting inspected. Some of them still prefer to not comply when it is still profitable to them.
- Companies are aware of the limited capacity of the inspectors and use this in their risk assessment.
- The compliance history and estimated propensity to violate are both used in the risk assessment of inspectors when working data-driven, the weight of each can be adjusted. The compliance history is normalised to fit the propensity to violate scale of the companies.
- When a company has not yet been visited, the inspectors assume its compliance to be equal to the average of the visited companies.
- Inspectors strive to increase the compliance of the sector by catching malpractice and visiting compliant companies less.
- Compliance is remembered for an adjustable number of years. If a company has not been inspected for this amount of years, their compliance score will reset to the average.

List of Limitations

- There is a fixed number of inspectors and companies which does not change during a simulation run.
- Inspectors can only visit one location per timestep and every inspection requires one inspector.
- The model starts without information about the compliance of companies.
- The model does not take into account false positives or negatives.
- The data estimated propensity to violate can exceed 100, while the actual propensity to violate of companies is capped at 100. The absolute value of this estimation does not matter since it is only being compared to the other estimations and the compliance history is normalised to fit this.
- This model does not take into account a company's size in relation to their power over inspection agencies because of economic interests or other factors.
- The correlation between the company characteristics and their propensity to violate can only be positive, negative estimations from inspectors will be set to 0.
- Companies and inspectors do not have a budget. Fines do not impact the companies in the model and the inspector gain nothing from giving fines.

Appendix B – Verification

In this appendix the results of the verification steps are given. The list below shows the hypothesis tested and verified during the verification steps.

- A company’s expected chance to get inspected should stay between 0 and 1. **Confirmed**
- A company’s not inspected counter should increase when they do not get inspected and reset after an inspection. **Confirmed**
- The observed and total compliance rates should be between 0 and 100. **Fixed**
- Observed compliance + observed malpractice should add up to 1. **Confirmed**
- If a company is not inspected for longer than “memory compliance years”, their compliance statistics should be set to -1 and its normalised compliance should be set to the average of all the companies until they get inspected again. **Confirmed**
- A company’s normalised compliance should be between de highest and lowest compliance score. **Confirmed**
- A company’s number of times getting caught + number of compliant inspections should add up to their total amount of inspections. **Confirmed**
- A company’s estimation to get inspected should increase when they get inspected and decrease when they do not get inspected for a while. **Confirmed**
- A company’s propensity to violate should be between 0 and 100. **Confirmed**
- There should never be more than one inspector at the locations of a company. **Confirmed**
- Companies should be sorted on their propensity to violate from top left corner to the bottom right corner. **Confirmed**
- The sum of received fines should be equal to the sum of given fines. **Confirmed**
- The companies with the highest propensity to violate should be the first to not comply if nobody is inspected. **Confirmed**
- Companies with a propensity to violate lower than the threshold should never violate, even when not inspected. **Confirmed**
- The actual correlation of the companies’ characteristics with the propensity to violate should be close to the chosen correlation during setup. **Confirmed**
- If the percentage always comply is set to 100, every company should comply. **Confirmed**
- The data quality variable should always stay between 0 and 5. **Confirmed**
- Companies set their strategy based on their estimations in comparison with the propensity to violate. **Confirmed**
- Stats extension in NetLogo correctly calculates correlation factors of the observed companies every timestep. **Confirmed**
- Every companies changes strategy if they get inspected every year while the costs of compliance are lower than the fine. **Confirmed**

Timeline Sanity Test

During the modelling phase of the model, its behaviour was continuously checked by means of output monitors. These model runs been on some occasions also been presented to data scientists at the NVWA. This way it was checked if the model behaved in the way it was expected to when changes were applied. A screenshot of these output monitors is shown below (Figure B.1). During the implementation phase these helped in quickly noticing and resolving unexpected behaviour. After the implementation phase no unexpected behaviour has been encountered.

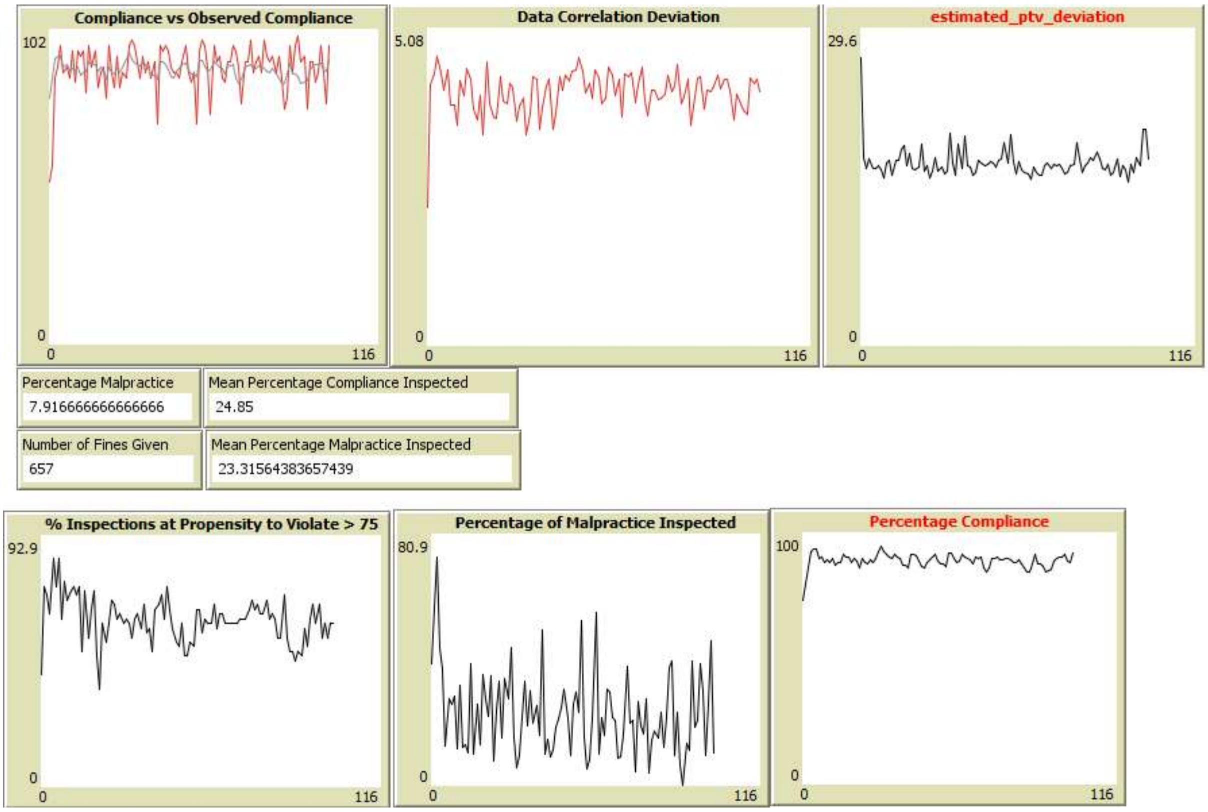

Figure B. 1 Timeline sanity test

Appendix C – NetLogo Interface

The figure below shows the NetLogo interface developed and used for this study. On the left it contains all the sliders for the policy levers and the uncertainties. On the right are the output monitors used to track the behaviour and the outcomes of the model.

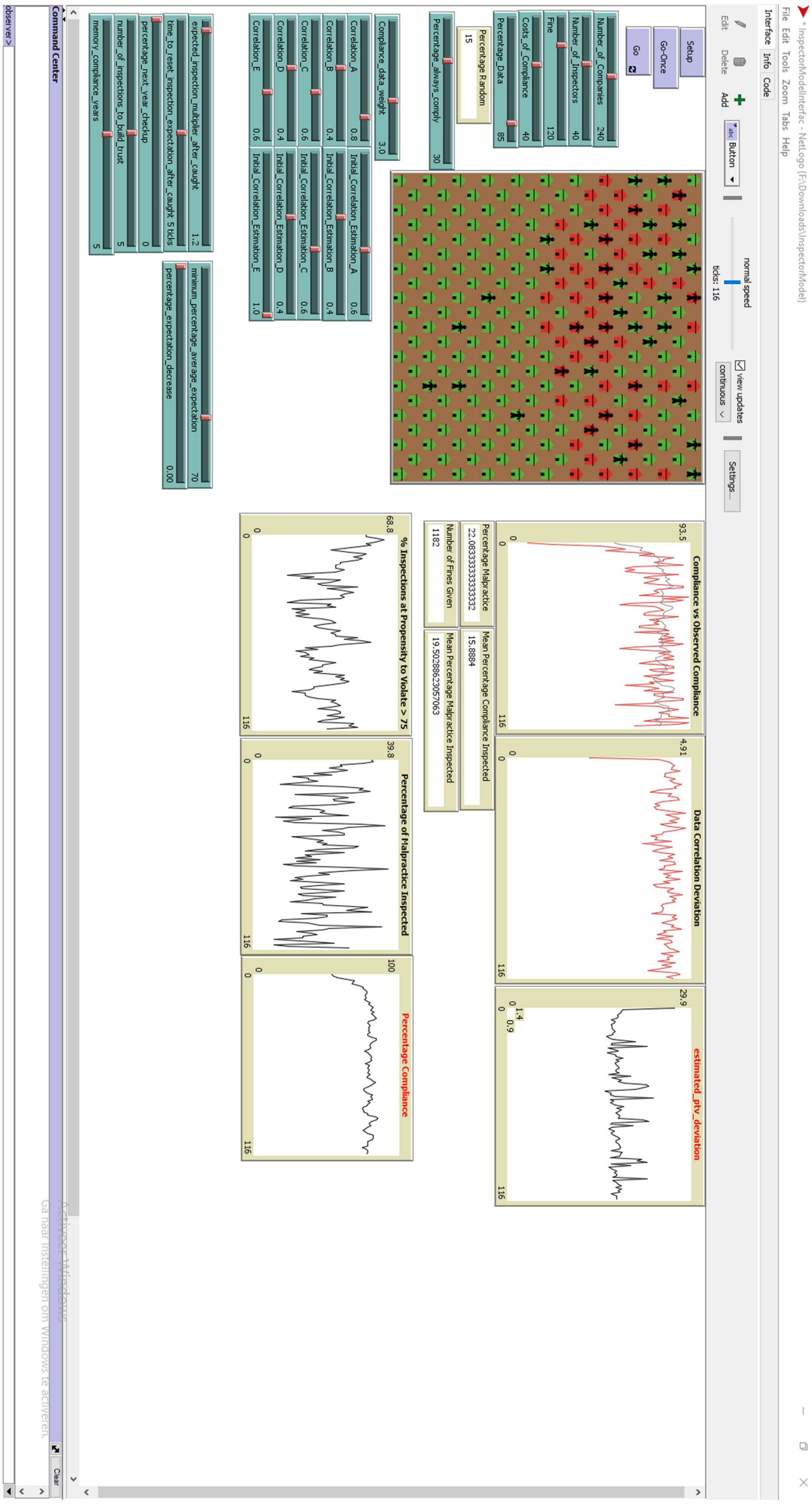

Figure C. 1 NetLogo interface

Appendix D – Results – Sensitivity Analysis Histograms

In the coming appendices all results are presented. They are ordered by visualisation / analysis in order to make an easier comparison between the performed experiments. Insights gained from these results are discussed in chapter 10. Within the ordering by visualisation they are again ordered by the performed experiments, beginning with the effect of the policy options for the reference scenario, after that the global sensitivity analysis sampling all variables and lastly the sensitivity analysis leaving out the sector specific variables.

The first presented results below are histograms of the outcomes of interest for each of the experiments. These show how large the spread of each of the outcomes is for each of the experiments. These histograms are useful when for seeing how much variance in the outcomes we are trying to attribute to the input variables. They serve as an overview before getting to the more in-depth analysis. Larger spreads mean they are more sensitive to the uncertainties or policy options.

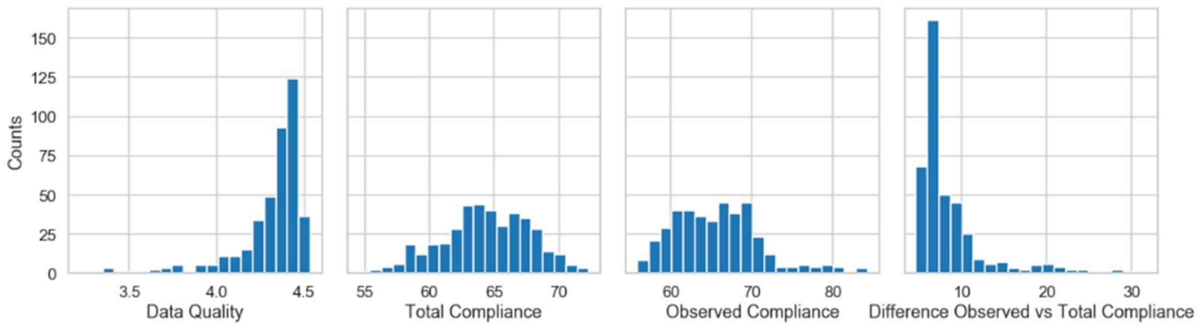

Figure D. 1 Histogram policy options for reference scenario

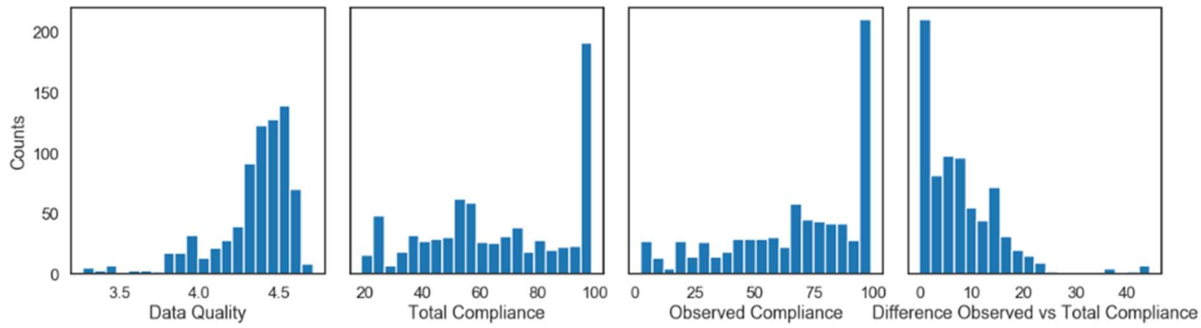

Figure D. 2 Histogram policy options for refence scenario

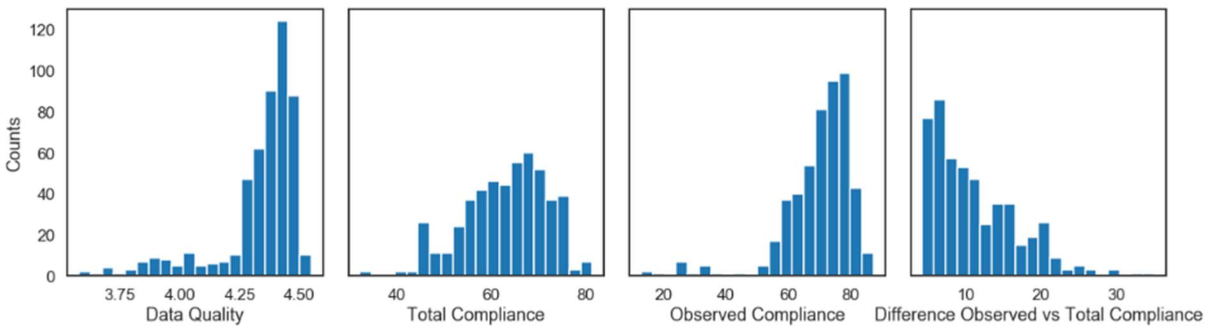

Figure D. 3 Histogram policy options without sector specific variables

Appendix E – Results – Plots with Confidence Intervals

The figures below show the estimated first and total-order indices with their estimated confidence intervals (shown as error bars). The first order indices, shown in blue, indicate how much each variable directly impacts one of the outcome variables, without any interaction effects with other variables. The total-order indices, shown in orange, show the total impact of each variable on the outcomes, including second and higher-order interactions effects. These indices are presented in tables at the end of the appendix (Appendix G). The interaction effects between the variables are better visualised using circle plots (Appendix F).

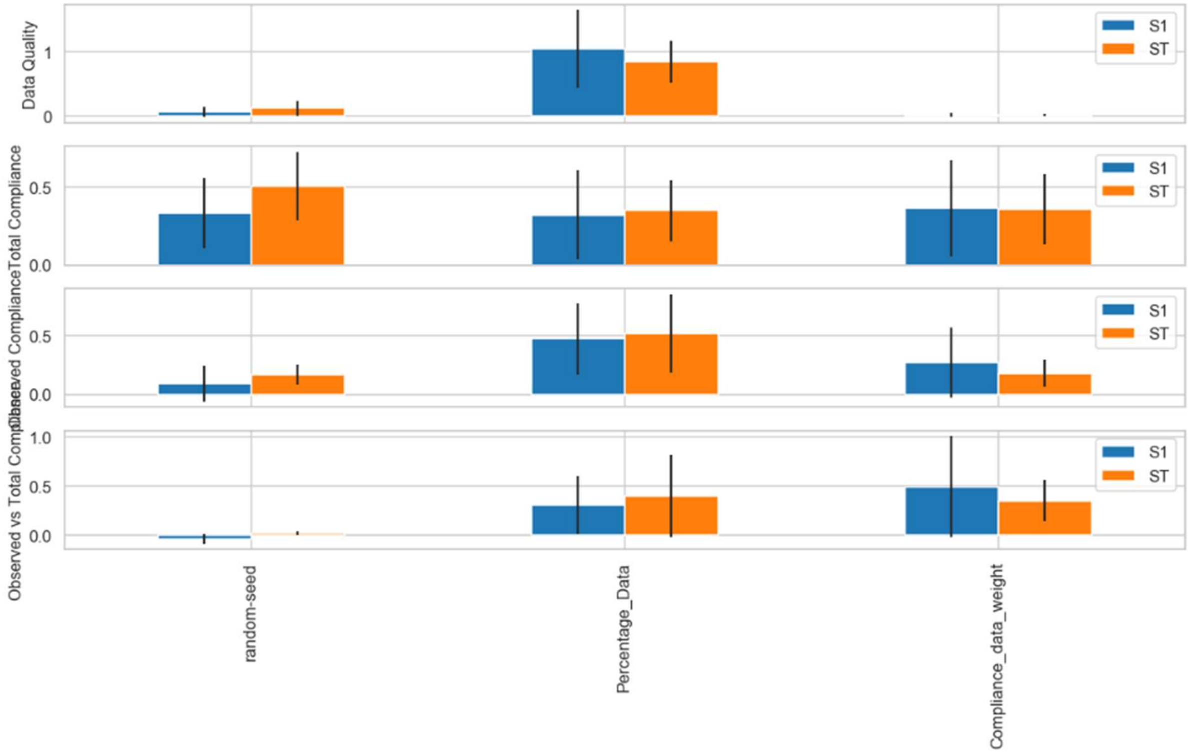

Figure E. 1 Plot Sobol indices with confidence intervals reference scenario

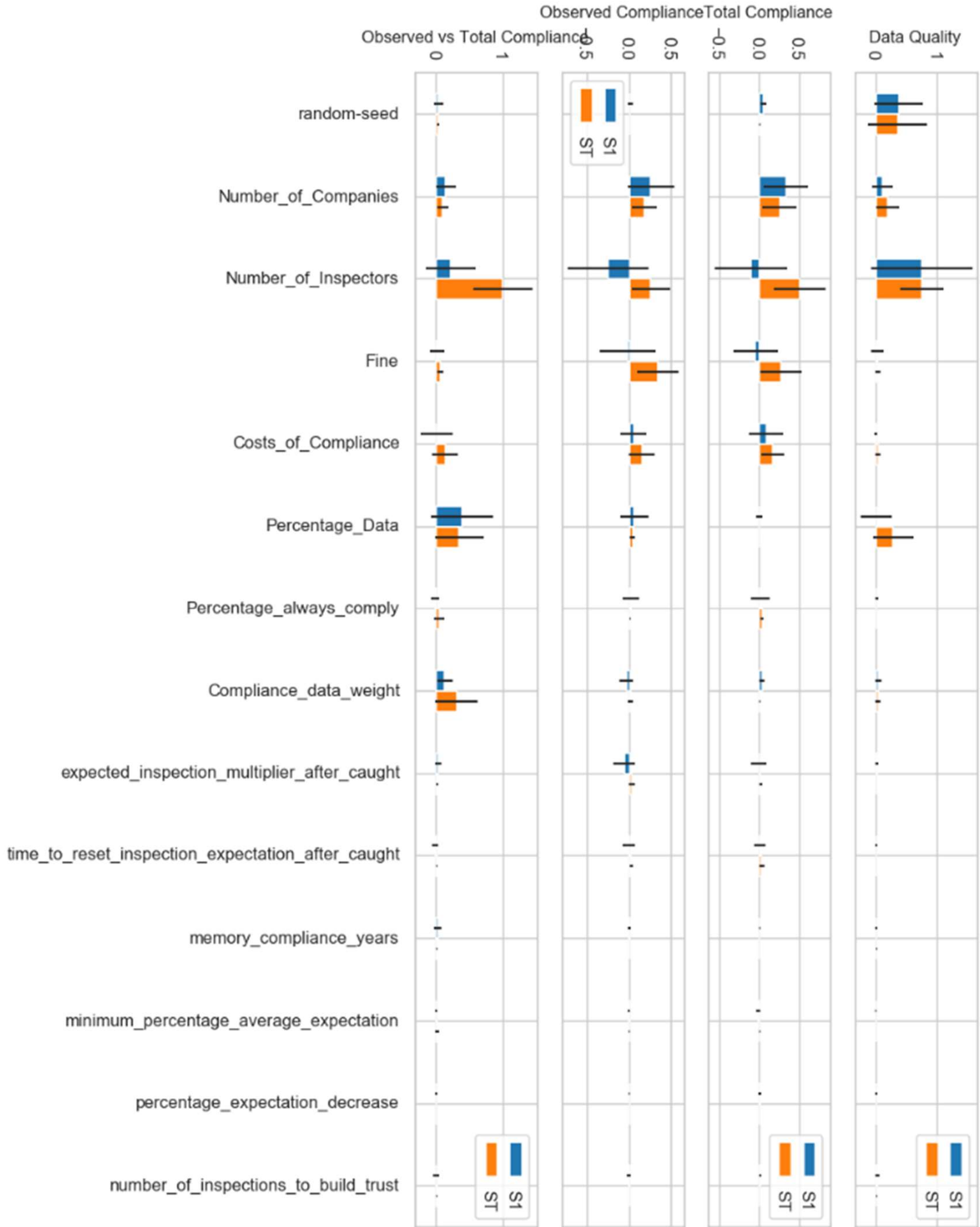

Figure E. 2 Plot Sobol indices with confidence intervals all variables

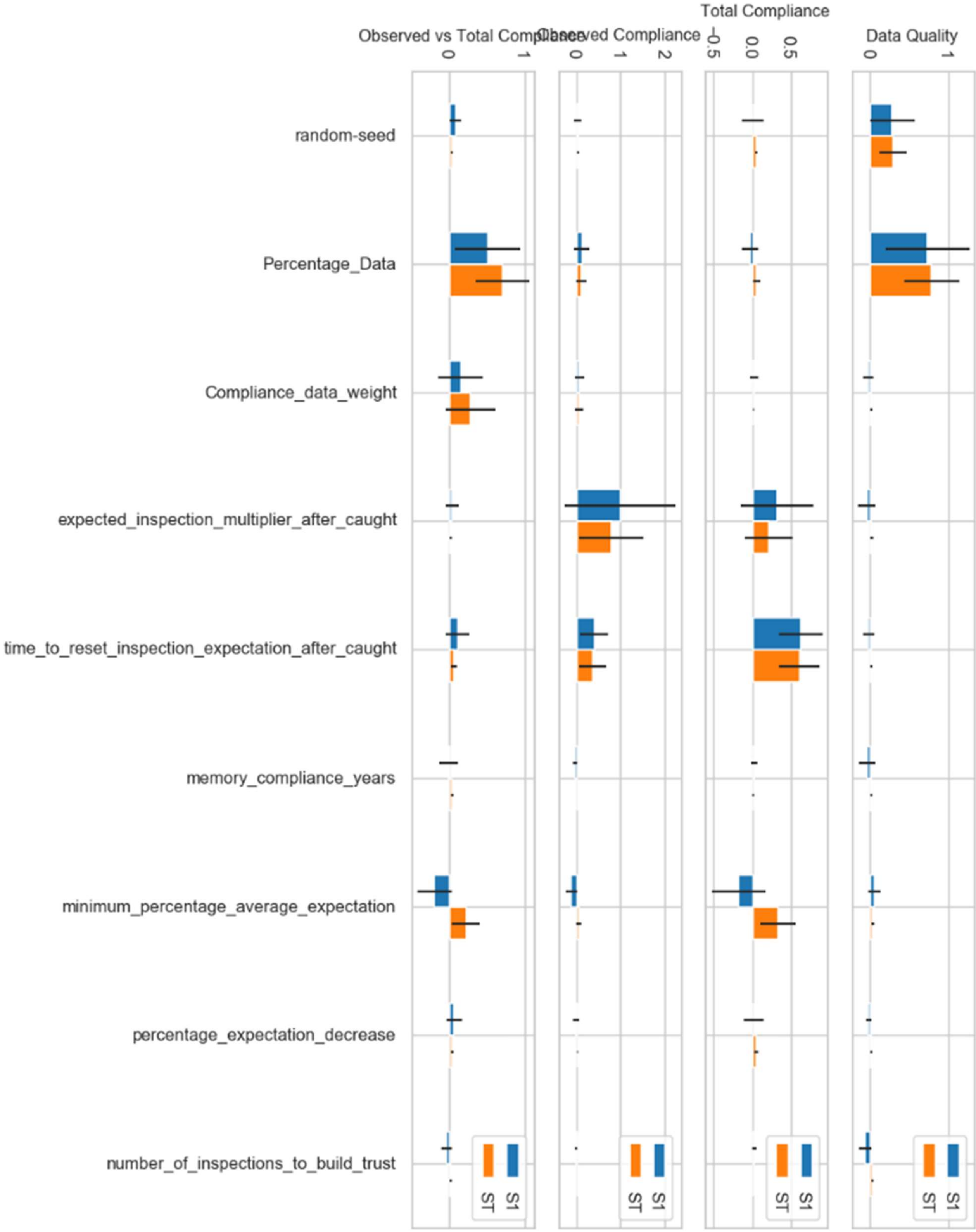

Figure E. 3 Plot Sobol indices with confidence intervals without sector specific variables

Appendix F – Results – Regression Plots

Below are bivariate scatter plots used to visualize the relationships between each input parameter and the outcomes of interest. This way positive or negatives correlations can be identified. A Pearson correlation factor is drawn through the plot to indicate the correlation.

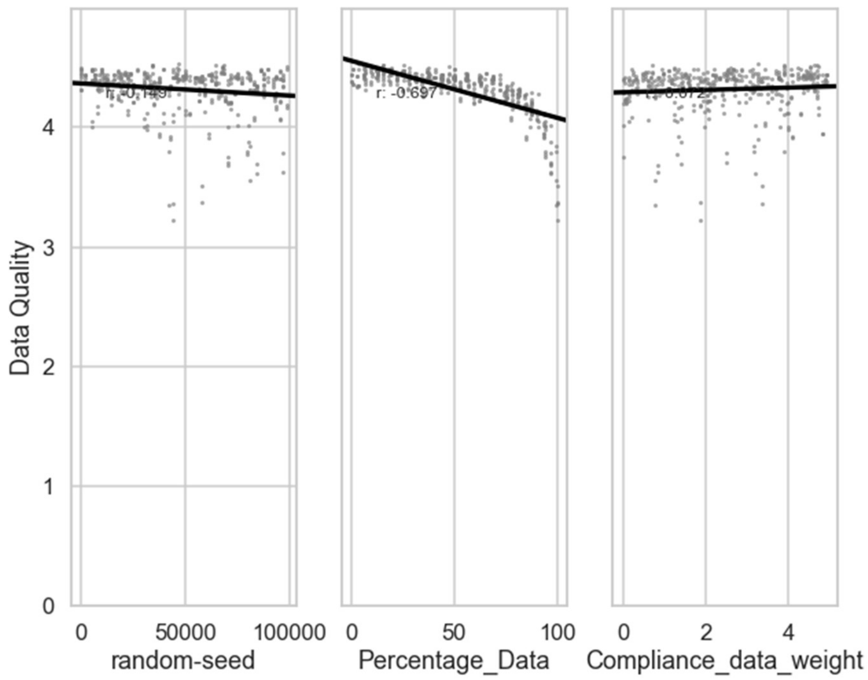

Figure F. 1 Bivariate scatter plot data quality reference scenario

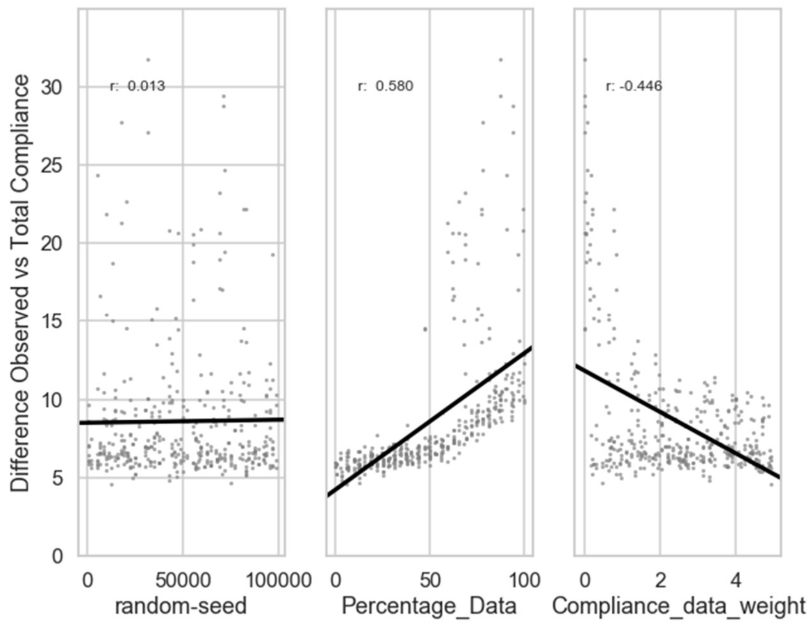

Figure F. 2 Bivariate scatter plot difference observed vs total compliance reference scenario

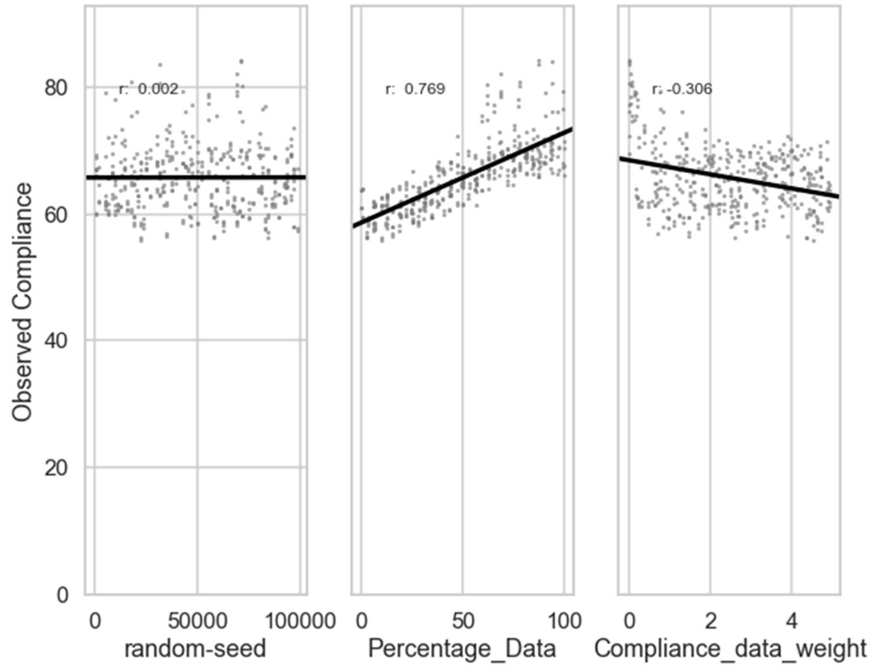

Figure F. 3 Bivariate scatter plot observed compliance reference scenario

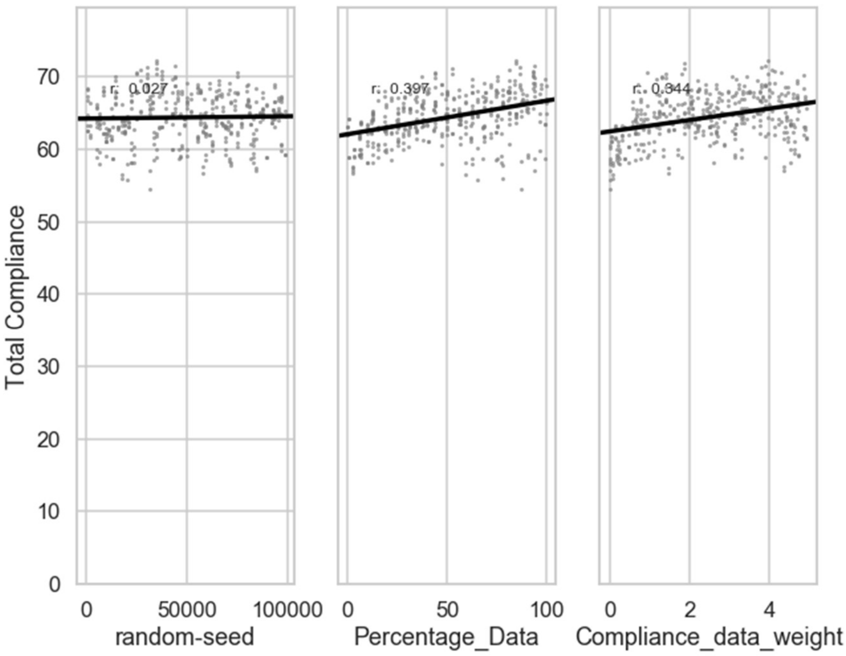

Figure F. 4 Bivariate scatter plot total compliance reference scenario

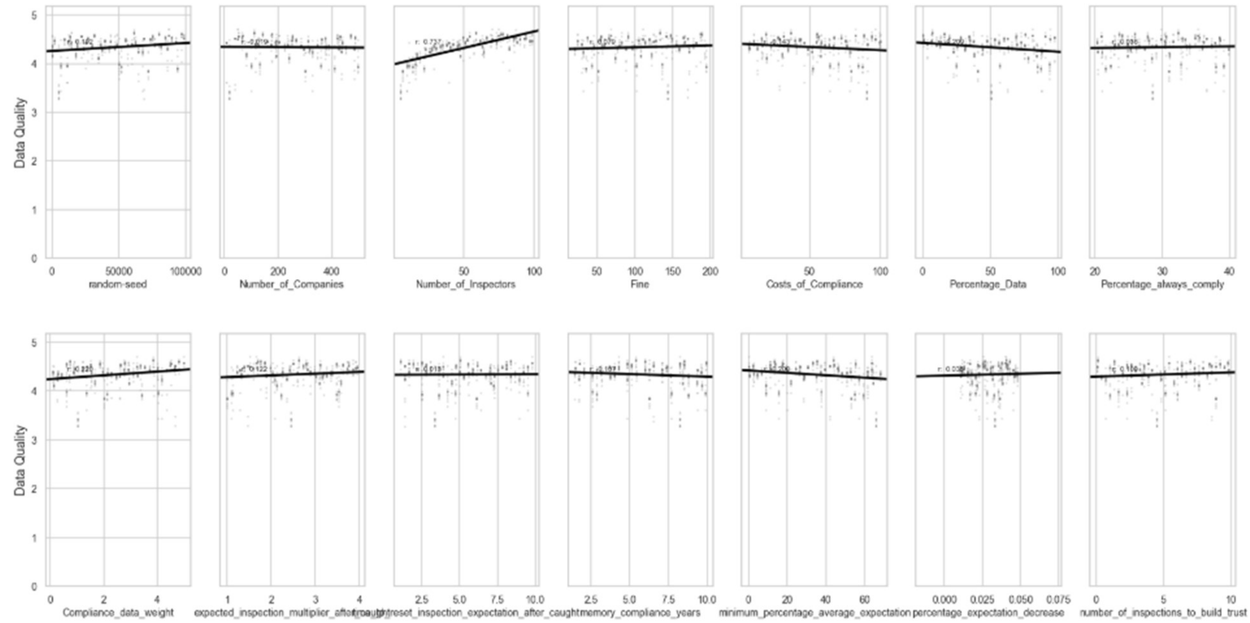

Figure F. 5 Bivariate scatter plot data quality all variables

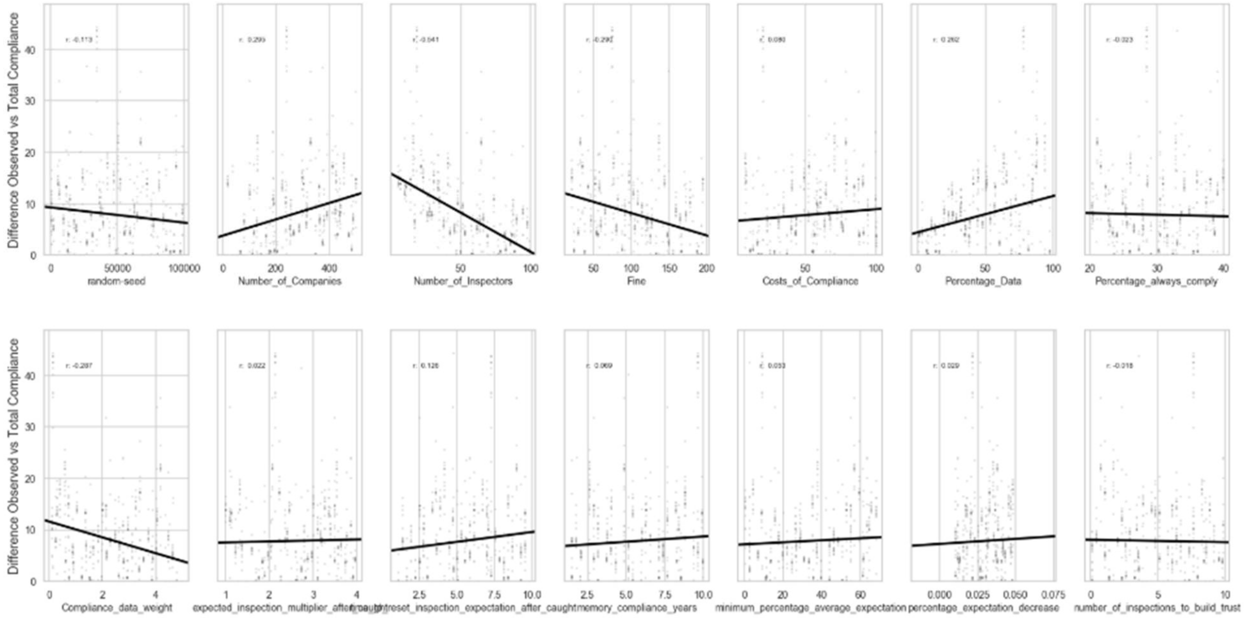

Figure F. 6 Bivariate scatter plot difference observed vs total compliance all variables

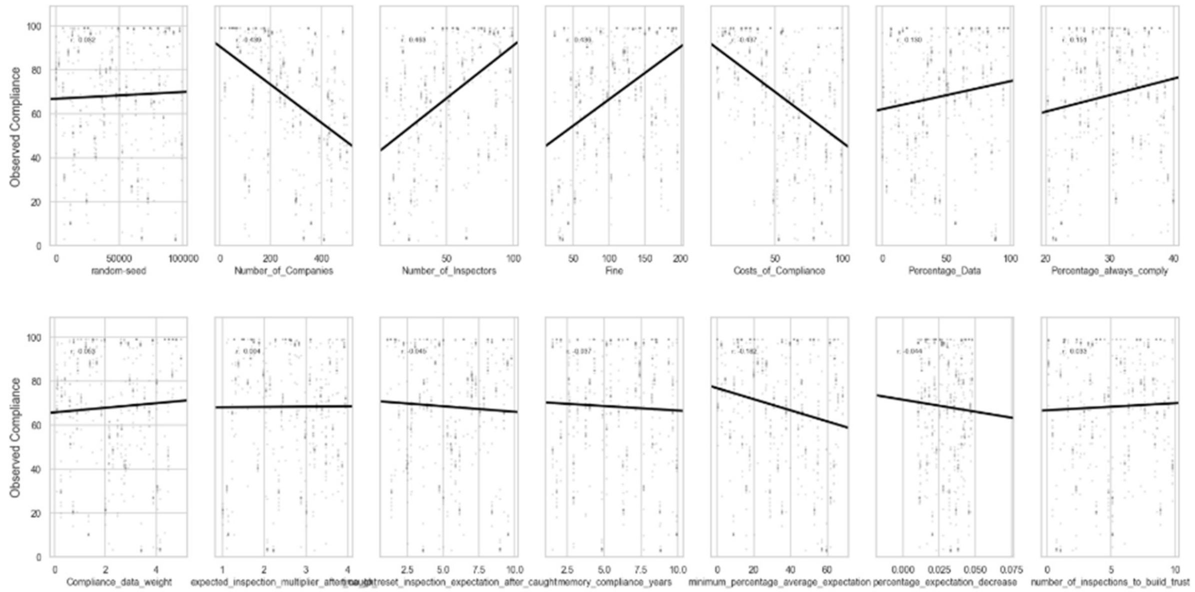

Figure F. 7 Bivariate scatter plot observed compliance all variables

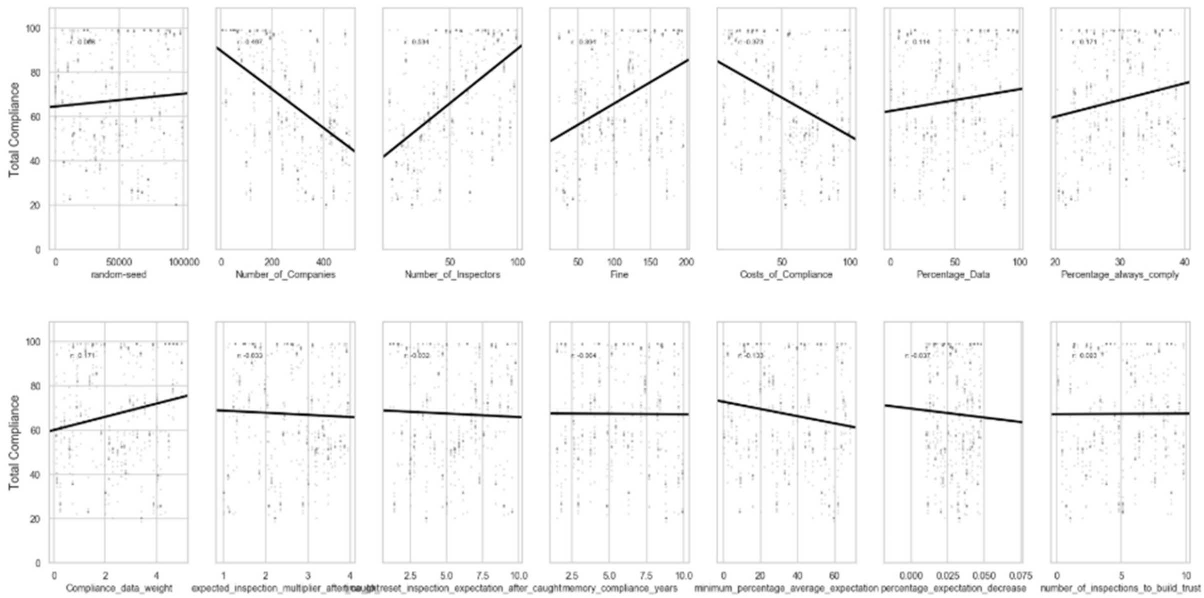

Figure F. 8 Bivariate scatter plot total compliance all variables

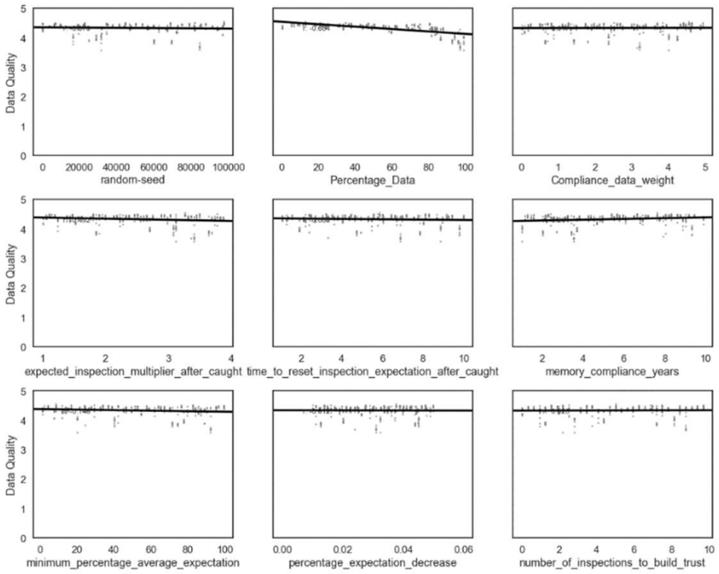

Figure F. 9 Bivariate scatter plot data quality without sector specific variables

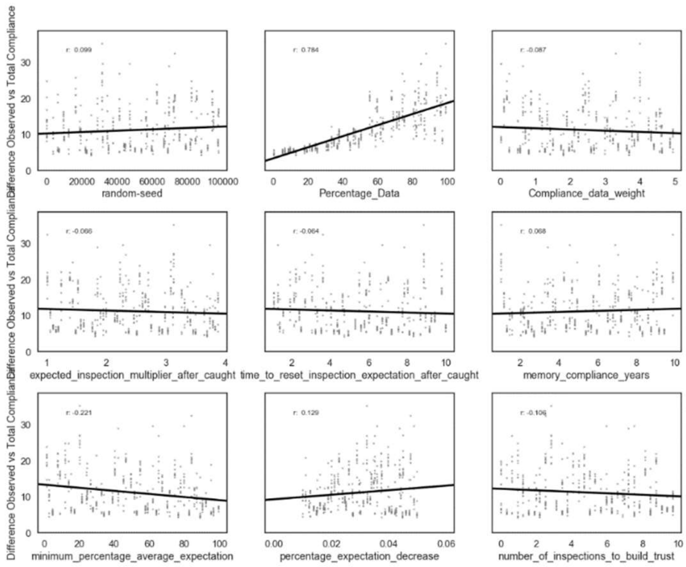

Figure F. 10 Bivariate scatter plot difference observed vs total compliance without sector specific variables

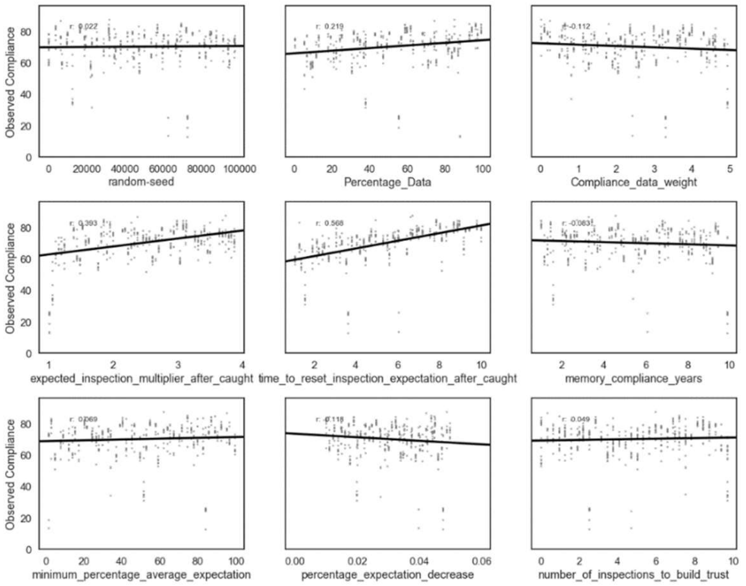

Figure F. 11 Bivariate scatter plot observed compliance without sector specific variables

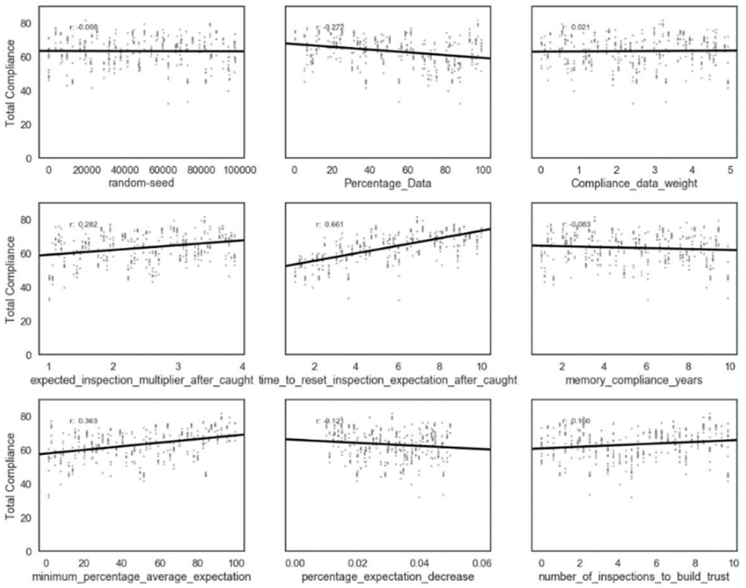

Figure F. 12 Bivariate scatter plot total compliance without sector specific variables

Appendix G – Results – Tables with Sobol Indices

The tables below give for each experiment the first and total-order Sobol indices of each of the variables for every outcome of interest with their confidence interval.

Table G. 1 Sobol indices reference scenario

Data Quality	ST	ST_conf	S1	S1_conf
random-seed	0.125387	0.112352	0.065810	0.080044
Percentage_Data	0.843551	0.321441	1.045168	0.607919
Compliance_data_weight	0.020529	0.019451	0.022878	0.037082

Total Compliance	ST	ST_conf	S1	S1_conf
random-seed	0.508511	0.220041	0.333013	0.228388
Percentage_Data	0.351247	0.197980	0.324000	0.286842
Compliance_data_weight	0.359070	0.227293	0.365477	0.312877

Difference Compliance	ST	ST_conf	S1	S1_conf
random-seed	0.022749	0.019172	-0.036001	0.051954
Percentage_Data	0.399626	0.415165	0.307118	0.292596
Compliance_data_weight	0.352613	0.211686	0.493972	0.515560

Observed Compliance	ST	ST_conf	S1	S1_conf
random-seed	0.167182	0.086282	0.092100	0.154279
Percentage_Data	0.519024	0.332258	0.469270	0.301186
Compliance_data_weight	0.178685	0.113541	0.271605	0.296045

Table G. 2 Sobol indices data quality all variables

Data Quality	ST	ST_conf	S1	S1_conf
random-seed	0.348394	0.557112	0.372080	0.504539
Number_of_Companies	0.186038	0.196332	0.105966	0.167192
Number_of_Inspectors	0.745353	0.412329	0.742629	0.942679
Fine	0.024733	0.037995	0.018963	0.078647
Costs_of_Compliance	0.026200	0.038423	-0.011416	0.022863
Percentage_Data	0.276711	0.335623	0.002727	0.200871
Percentage_always_comply	0.002802	0.002988	0.003188	0.020439
Compliance_data_weight	0.029589	0.037678	0.028758	0.048071
expected_inspection_multiplier_after_caught	0.002022	0.001549	0.010105	0.020856
time_to_reset_inspection_expectation_after_caught	0.002434	0.002224	-0.001014	0.018232
memory_compliance_years	0.008006	0.008499	0.000591	0.015971
minimum_percentage_average_expectation	0.000424	0.000731	-0.006117	0.011735
percentage_expectation_decrease	0.000803	0.001001	0.000128	0.011605
number_of_inspections_to_build_trust	0.004002	0.004174	0.010037	0.030361

Table G. 3 Sobol indices difference observed vs total compliance all variables

Difference Compliance	ST	ST_conf	S1	S1_conf
random-seed	0.025091	0.018960	0.037310	0.057727
Number_of_Companies	0.096858	0.105010	0.144687	0.154315
Number_of_Inspectors	0.992381	0.519363	0.209436	0.330335
Fine	0.061610	0.046584	0.015400	0.097193
Costs_of_Compliance	0.133796	0.228975	0.006352	0.206015
Percentage_Data	0.341197	0.348296	0.380389	0.468908
Percentage_always_comply	0.041059	0.082620	-0.017299	0.060365
Compliance_data_weight	0.303099	0.330956	0.126863	0.135565
expected_inspection_multiplier_after_caught	0.014047	0.025237	0.027302	0.053488
time_to_reset_inspection_expectation_after_caught	0.009263	0.010741	-0.014675	0.058334
memory_compliance_years	0.008259	0.006938	0.025068	0.063839
minimum_percentage_average_expectation	0.010570	0.025697	-0.003857	0.014485
percentage_expectation_decrease	0.001185	0.002258	0.001326	0.014866
number_of_inspections_to_build_trust	0.007112	0.011397	-0.003414	0.043797

Table G. 4 Sobol indices observed compliance all variables

Observed Compliance	ST	ST_conf	S1	S1_conf
random-seed	0.002173	0.002318	0.011041	0.026622
Number_of_Companies	0.175963	0.146083	0.255559	0.288310
Number_of_Inspectors	0.255595	0.198496	-0.252799	0.426430
Fine	0.339294	0.254165	-0.021107	0.317466
Costs_of_Compliance	0.146876	0.217612	0.050297	0.180690
Percentage_Data	0.037996	0.030284	0.057809	0.135425
Percentage_always_comply	0.013729	0.010514	0.016980	0.078665
Compliance_data_weight	0.012039	0.024258	-0.035977	0.069557
expected_inspection_multiplier_after_caught	0.026436	0.027695	-0.061695	0.113069
time_to_reset_inspection_expectation_after_caught	0.019888	0.025816	-0.006224	0.061185
memory_compliance_years	0.000538	0.000444	-0.002991	0.018183
minimum_percentage_average_expectation	0.001071	0.002127	-0.005236	0.014860
percentage_expectation_decrease	0.000157	0.000272	0.002218	0.010362
number_of_inspections_to_build_trust	0.001039	0.001197	-0.007945	0.029426

Table G. 5 Sobol indices total compliance all variables

Total Compliance	ST	ST_conf	S1	S1_conf
random-seed	0.006753	0.009344	0.049124	0.059845
Number_of_Companies	0.255173	0.213078	0.333349	0.276178
Number_of_Inspectors	0.509238	0.338245	-0.107361	0.506028
Fine	0.276272	0.285487	-0.047377	0.267965
Costs_of_Compliance	0.169152	0.140874	0.088324	0.269620
Percentage_Data	0.002669	0.001908	-0.003406	0.035581
Percentage_always_comply	0.033512	0.018839	0.013746	0.111015
Compliance_data_weight	0.006327	0.010434	0.031710	0.048193
expected_inspection_multiplier_after_caught	0.017064	0.017945	-0.004577	0.092958
time_to_reset_inspection_expectation_after_caught	0.028767	0.026702	0.006729	0.078598
memory_compliance_years	0.000084	0.000080	0.001808	0.004213
minimum_percentage_average_expectation	0.003894	0.007811	-0.010644	0.026770
percentage_expectation_decrease	0.000552	0.000505	0.008049	0.015638
number_of_inspections_to_build_trust	0.000385	0.000392	0.009589	0.011359

Table G. 6 Sobol indices data quality without sector specific variables

Data Quality	ST	ST_conf	S1	S1_conf
random-seed	0.289579	0.251811	0.280741	0.321287
Percentage_Data	0.781746	0.341489	0.729864	0.540660
Compliance_data_weight	0.011574	0.032665	-0.026109	0.103746
expected_inspection_multiplier_after_caught	0.016487	0.027248	-0.045852	0.096356
time_to_reset_inspection_expectation_after_caught	0.009705	0.023135	-0.024424	0.072256
memory_compliance_years	0.010560	0.013317	-0.044889	0.080378
minimum_percentage_average_expectation	0.026217	0.022905	0.052238	0.058620
percentage_expectation_decrease	0.007486	0.018685	-0.026693	0.041474
number_of_inspections_to_build_trust	0.022123	0.023542	-0.068129	0.096921

Table G. 7 Sobol indices difference observed vs total compliance without sector specific variables

Difference Compliance	ST	ST_conf	S1	S1_conf
random-seed	0.028296	0.014796	0.084779	0.077318
Percentage_Data	0.698204	0.386844	0.507445	0.405881
Compliance_data_weight	0.281909	0.374492	0.153320	0.315430
expected_inspection_multiplier_after_caught	0.021187	0.015014	0.038955	0.087395
time_to_reset_inspection_expectation_after_caught	0.060429	0.048090	0.113735	0.149471
memory_compliance_years	0.037021	0.022341	-0.005045	0.125412
minimum_percentage_average_expectation	0.218079	0.205229	-0.193431	0.226434
percentage_expectation_decrease	0.037191	0.017355	0.066699	0.099620
number_of_inspections_to_build_trust	0.017309	0.015258	-0.036482	0.073583

Table G. 8 Sobol indices observed compliance without sector specific variables

Observed Compliance	ST	ST_conf	S1	S1_conf
random-seed	0.024796	0.023181	0.024875	0.091833
Percentage_Data	0.098709	0.113942	0.110908	0.170058
Compliance_data_weight	0.053679	0.126976	0.058981	0.139094
expected_inspection_multiplier_after_caught	0.780289	0.691999	0.980268	1.158475
time_to_reset_inspection_expectation_after_caught	0.352884	0.289696	0.392021	0.341701
memory_compliance_years	0.003439	0.003807	-0.044850	0.034159
minimum_percentage_average_expectation	0.047386	0.050216	-0.127428	0.118325
percentage_expectation_decrease	0.008079	0.010599	-0.021192	0.066271
number_of_inspections_to_build_trust	0.001674	0.001857	-0.019465	0.026922

Table G. 9 Sobol indices total compliance without sector specific variables

Total Compliance	ST	ST_conf	S1	S1_conf
random-seed	0.043423	0.019028	0.001468	0.139362
Percentage_Data	0.051461	0.040946	-0.035168	0.096938
Compliance_data_weight	0.009989	0.008551	0.019055	0.053210
expected_inspection_multiplier_after_caught	0.207144	0.258867	0.314324	0.392244
time_to_reset_inspection_expectation_after_caught	0.605262	0.247781	0.625807	0.322559
memory_compliance_years	0.004970	0.005608	0.016067	0.035444
minimum_percentage_average_expectation	0.327326	0.237029	-0.182931	0.349850
percentage_expectation_decrease	0.043134	0.023125	0.008685	0.138290
number_of_inspections_to_build_trust	0.003643	0.003687	0.015740	0.024129

Table G. 10 Sobol indices data quality and total compliance with correlations

Data Quality	ST	ST_conf	S1	S1_conf
random-seed	0.335003	0.175590	0.304819	0.270140
Percentage_Data	0.571758	0.403565	0.656371	0.368074
Compliance_data_weight	0.014464	0.011199	0.037320	0.074548
Correlation_A	0.148000	0.074287	0.216894	0.234908
Correlation_B	0.109473	0.061445	-0.022692	0.190851
Correlation_C	0.080974	0.057817	-0.077838	0.196436
Correlation_D	0.090422	0.050119	0.166834	0.219743
Correlation_E	0.084092	0.046734	0.112489	0.173274

Total Compliance	ST	ST_conf	S1	S1_conf
random-seed	0.951506	0.510672	0.793096	0.456195
Percentage_Data	0.328858	0.273879	0.063979	0.406625
Compliance_data_weight	0.165941	0.191674	0.015533	0.105069
Correlation_A	0.032495	0.018702	0.015178	0.114567
Correlation_B	0.047548	0.034542	0.105665	0.156011
Correlation_C	0.044362	0.025632	0.028941	0.101851
Correlation_D	0.037909	0.025361	0.093922	0.097854
Correlation_E	0.037913	0.021206	0.004459	0.097502

Table G. 11 Sobol indices difference observed vs total compliance and observed compliance with correlations

Difference Compliance	ST	ST_conf	S1	S1_conf
random-seed	0.093796	0.043729	0.059235	0.160834
Percentage_Data	0.968252	0.779205	0.563486	0.557326
Compliance_data_weight	0.532195	0.581938	0.048752	0.170165
Correlation_A	0.071813	0.042890	0.001783	0.122931
Correlation_B	0.050896	0.019215	0.102860	0.089770
Correlation_C	0.164570	0.156718	0.002638	0.162792
Correlation_D	0.165569	0.078686	-0.020953	0.135888
Correlation_E	0.115447	0.047434	0.044099	0.187791

Observed Compliance	ST	ST_conf	S1	S1_conf
random-seed	0.566990	0.406909	0.269686	0.404191
Percentage_Data	0.808096	0.566438	0.438677	0.385384
Compliance_data_weight	0.125740	0.126502	0.047920	0.123060
Correlation_A	0.022045	0.029750	0.001082	0.056974
Correlation_B	0.019543	0.014681	0.055323	0.073559
Correlation_C	0.038154	0.050278	-0.008747	0.071244
Correlation_D	0.048865	0.054989	-0.011901	0.053037
Correlation_E	0.035273	0.041800	-0.006595	0.054793

Appendix H – Results – Circle Plots

The circle plots below help with visualising the second order interaction effects between the variables. The lines between the circles indicates these interaction effects. The sizes of the circles represent the first and total-order Sobol indices.

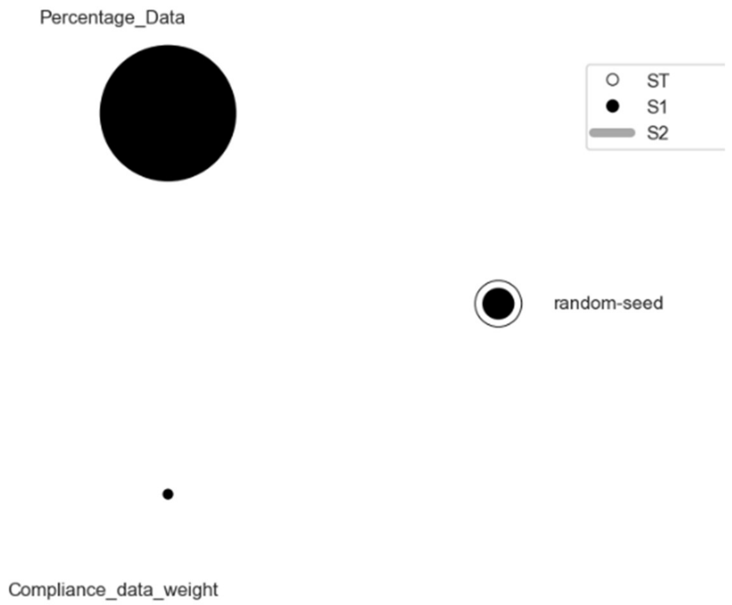

Figure H. 1 Circle plot data quality reference scenario

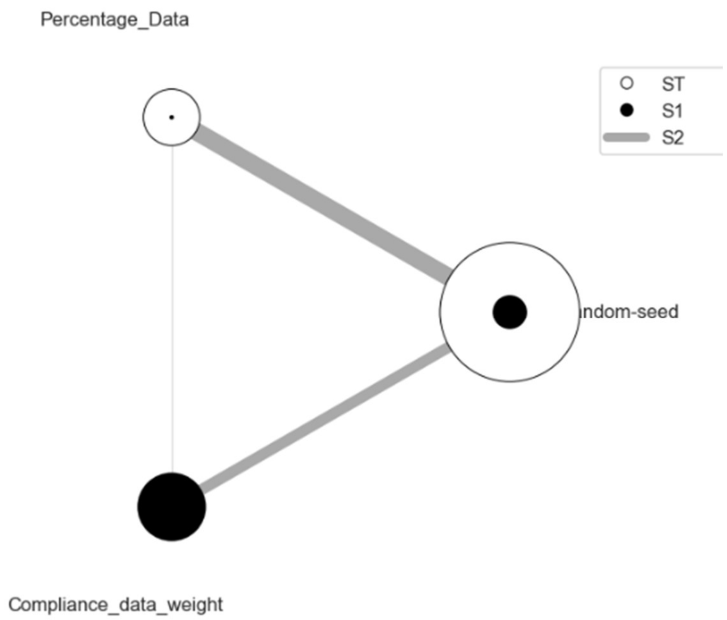

Figure H. 2 Circle plot total compliance reference scenario

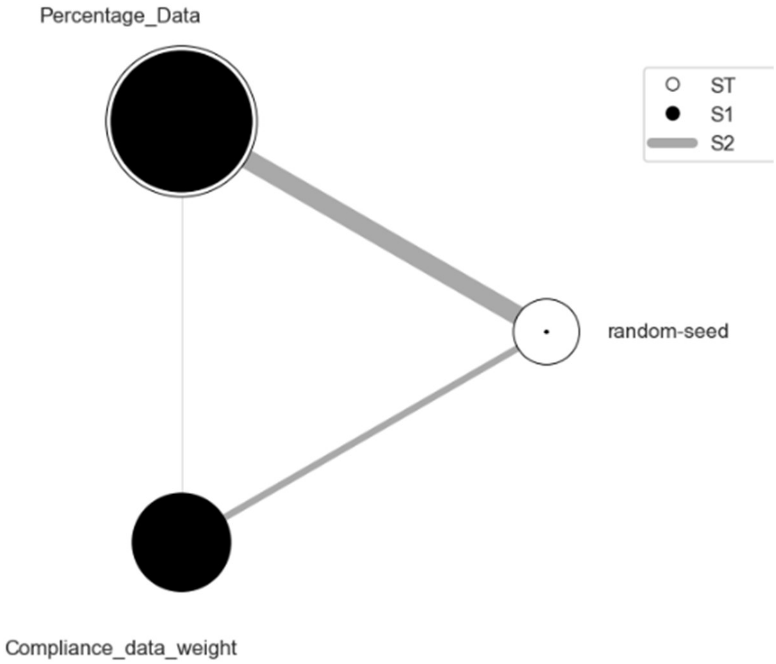

Figure H. 3 Circle plot observed compliance reference scenario

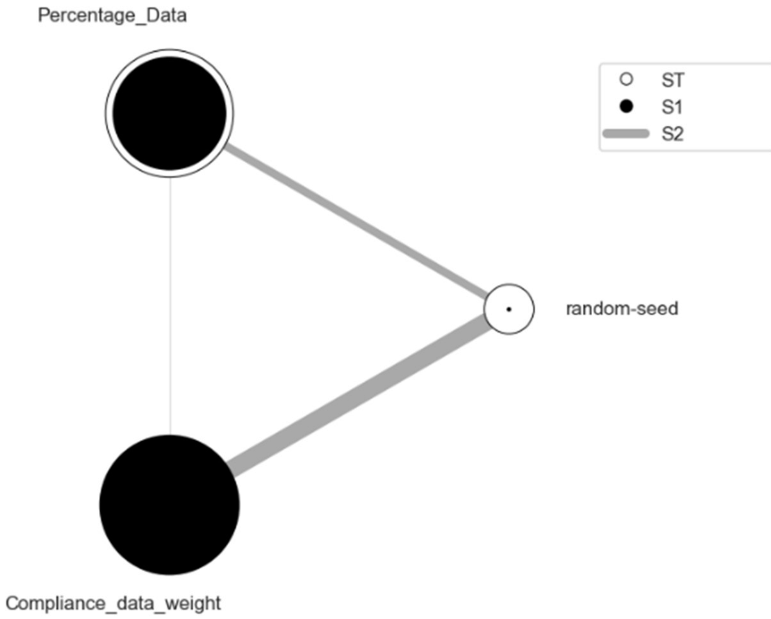

Figure H. 4 Circle plot difference observed vs total compliance reference scenario

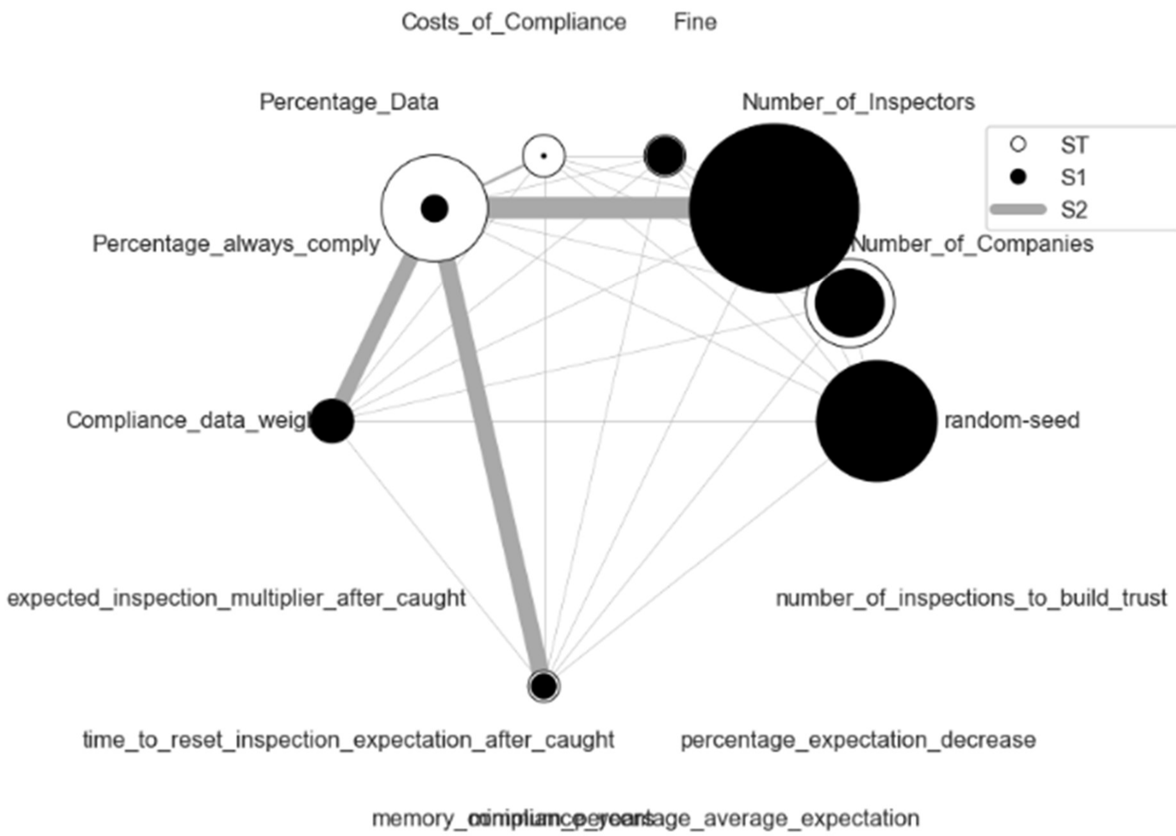

Figure H. 5 Circle plot data quality all variables

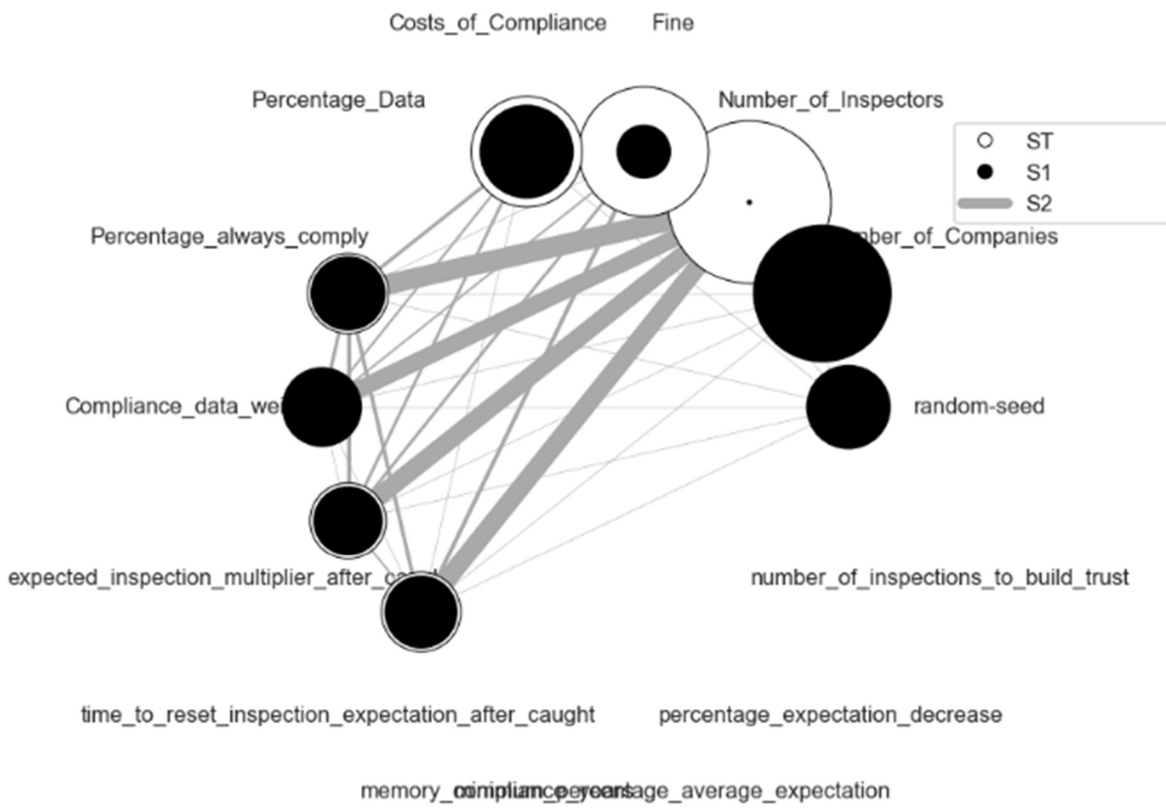

Figure H. 6 Circle plot total compliance all variables

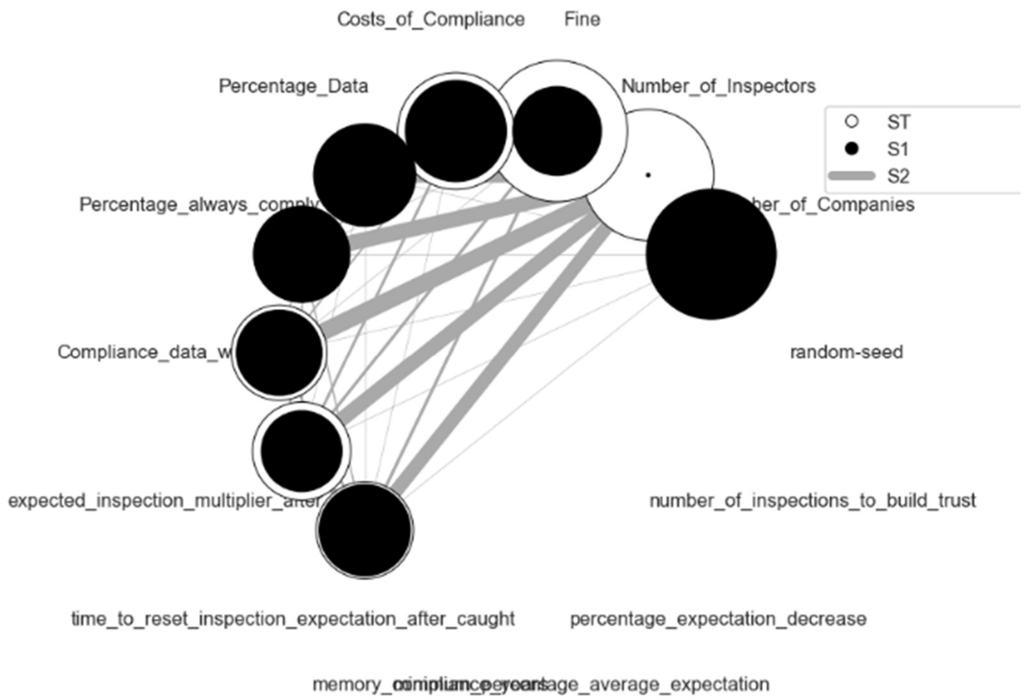

Figure H. 7 Circle plot observed compliance all variables

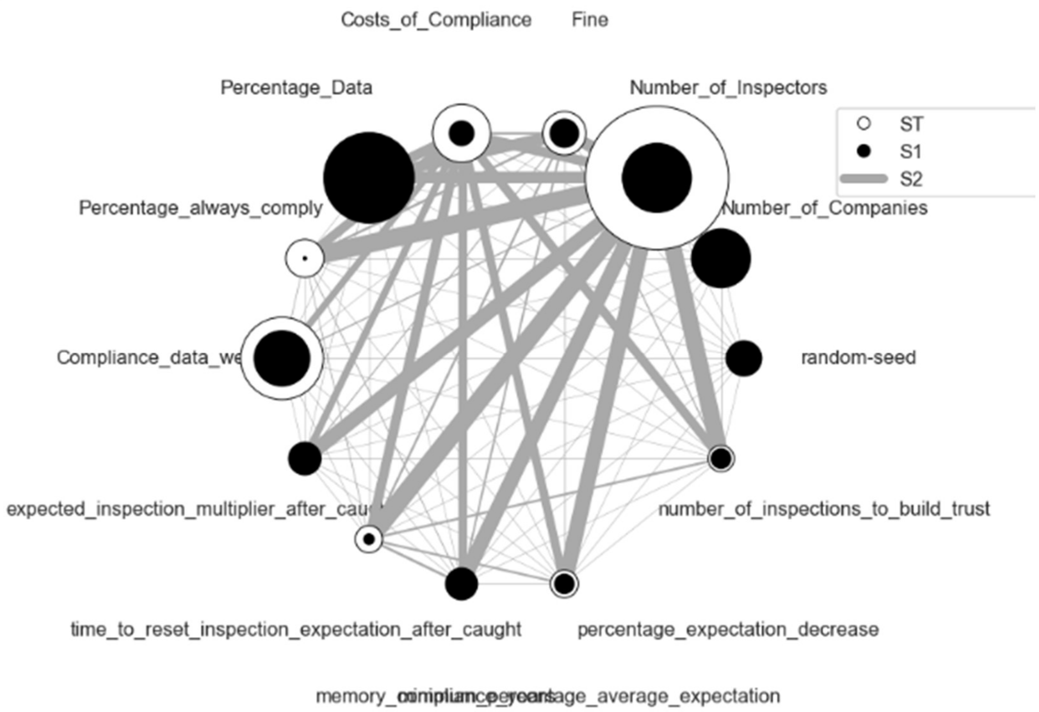

Figure H. 8 Circle plot difference observed vs total compliance all variables

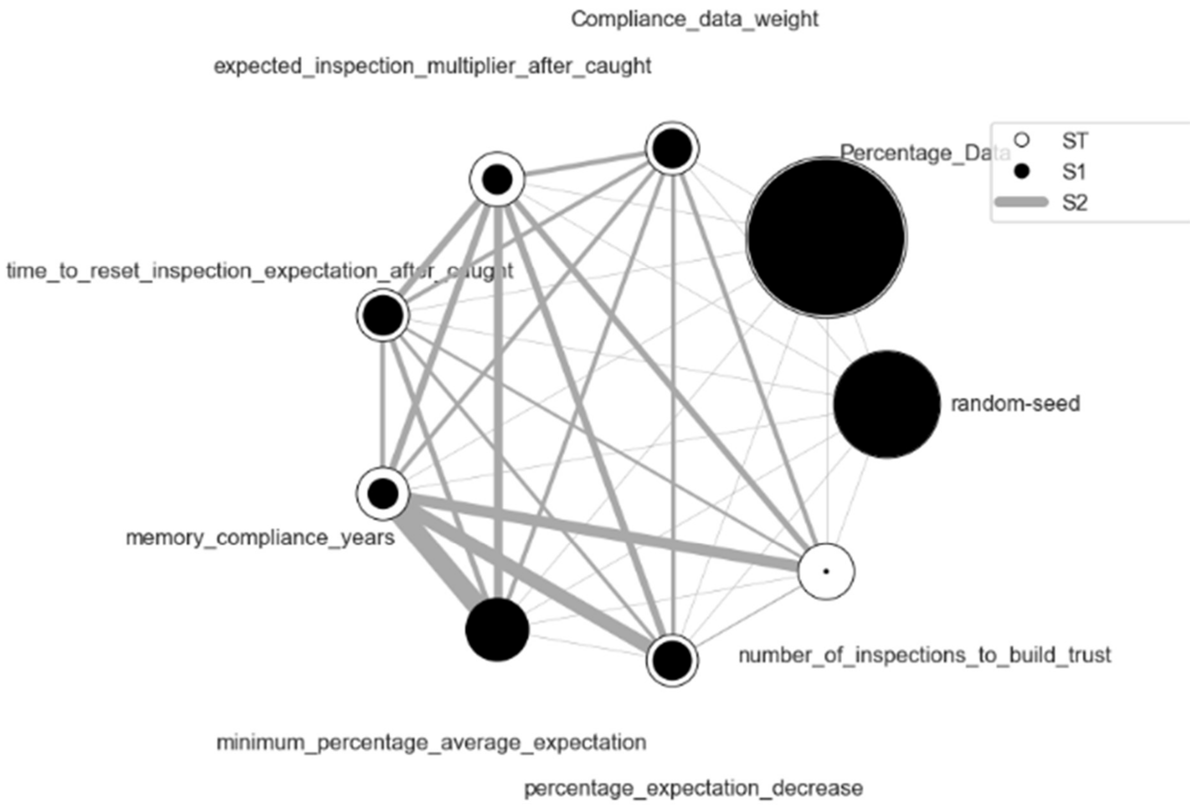

Figure H. 9 Circle plot data quality without sector specific variables

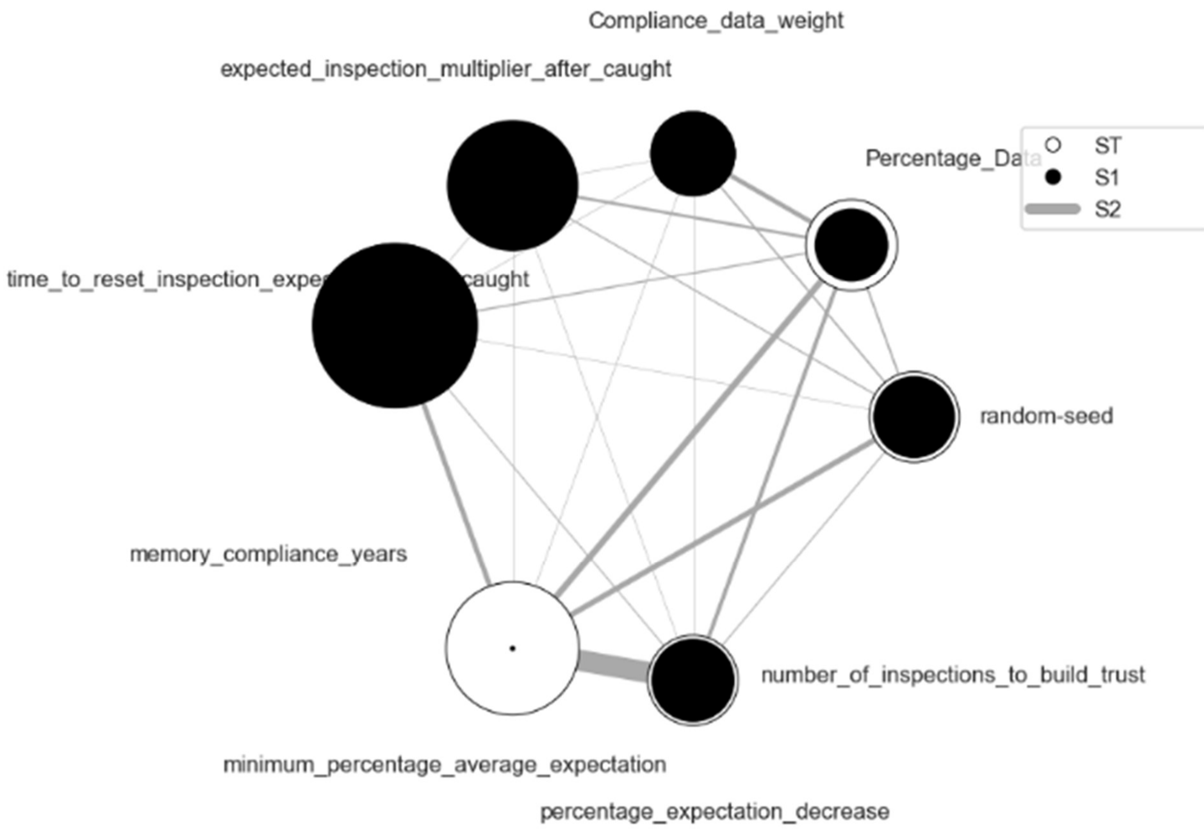

Figure H. 10 Circle plot total compliance without sector specific variables

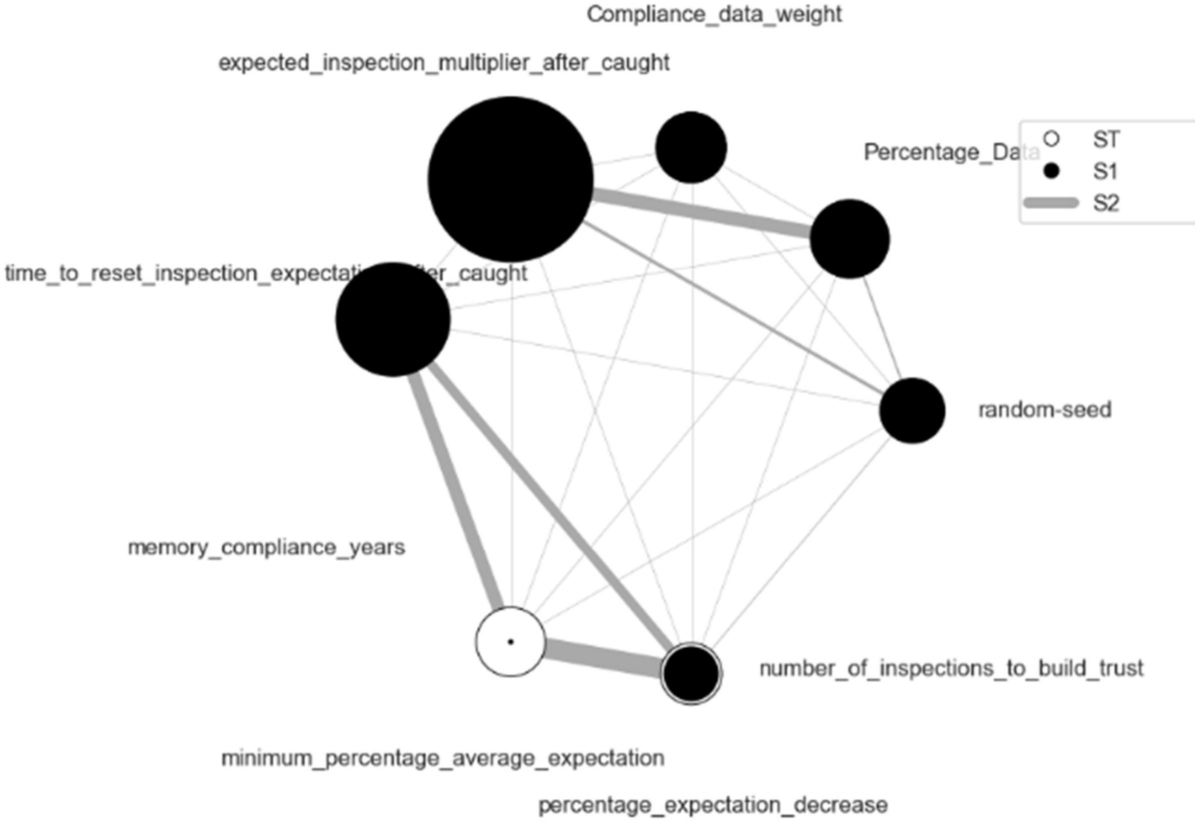

Figure H. 11 Circle plot observed compliance without sector specific variables

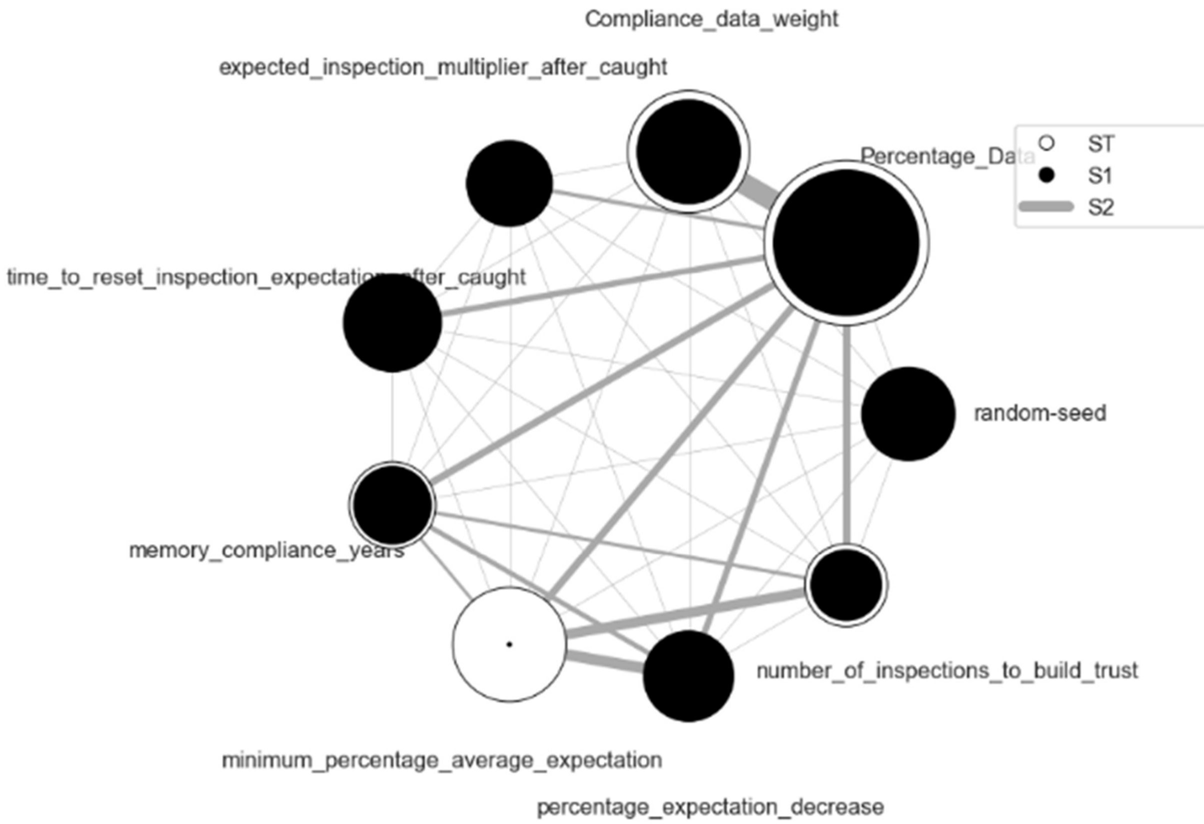

Figure H. 12 Circle plot difference observed vs total compliance without sector specific variables