Reconstructing illicit supply chains with sparse data

A simulation approach

van Schilt, Isabelle M.

DOI
10.4233/uuid:eaa5058e-f2b8-499d-94f7-e2711d931908

Publication date
2025

Document Version
Final published version

Citation (APA)
van Schilt, I. M. (2025). *Reconstructing illicit supply chains with sparse data: A simulation approach*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:eaa5058e-f2b8-499d-94f7-e2711d931908

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Reconstructing illicit supply chains with sparse data

A simulation approach

Isabelle M. van Schilt

RECONSTRUCTING ILLICIT SUPPLY CHAINS WITH SPARSE DATA

A SIMULATION APPROACH

RECONSTRUCTING ILLICIT SUPPLY CHAINS WITH SPARSE DATA

A SIMULATION APPROACH

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus prof. dr. ir. T.H.J.J. van der Hagen,
Chair of the Board for Doctorates
to be defended publicly on
Friday 17 January 2025 at 15:00 o'clock

by

Isabelle Maria VAN SCHILT

Master of Science in Engineering and Policy Analyis,
Delft University of Technology, the Netherlands,
born in Nieuwerkerk aan den IJssel, the Netherlands.

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof. dr. ir. A. Verbraeck,	Delft University of Technology, promotor
Prof. dr. ir. J. H. Kwakkel,	Delft University of Technology, promotor

Independent members:

Prof. dr. M. E. Warnier,	Delft University of Technology
Prof. dr. M. T. J. Spaan,	Delft University of Technology
Prof. dr. L. I. Shelley,	George Mason University
Dr. ir. M. Winkenbach,	Massachusetts Institute of Technology
Prof. dr. ir. L. A. Tavasszy,	Delft University of Technology, reserve member

Additional members:

Dr. ir. J. P. Mense,	Politie

Dr. ir. J. P. Mense contributed significantly to the completion of the dissertation.

This research was fully funded by Nationale Politielab Artifical Intelligence.

TRAIL Thesis Series no. T2025/2, the Netherlands Research School TRAIL

TRAIL, P.O. BOX 5017, 2600 GA Delft, The Netherlands
E-mail: info@rsTRAIL.nl

ISBN 978-90-5584-358-9

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

Printed in the Netherlands by Haveka. Cover by Marta Ríos.

Have a little faith,
and if that doesn't work,
have a lot of mimosas.

Blair Waldorf

CONTENTS

ACKNOWLEDGEMENTS

After walking the halls of Technology, Policy, and Management (TPM) for 11 years, this dissertation marks the end of my journey at Delft University of Technology. Throughout these years, TPM has started to feel like a home – a place where I could walk down the halls and have a chat with any student, professor, or any other colleague. It was a true pleasure to be educated and to work with these amazing people at TPM. For TPM to be a place of like-minded people and my safe haven.

First, I would like to express my gratitude to my promotor Alexander Verbraeck who has been part of this journey for over 9 years. You have taught me so much during the course of the years in all the projects, and really helped me grow on a professional and personal level. You were always open for a chat to support me – no matter what the topic was. Your everlasting wisdom and many innovative ideas are an inspiration for me and the rest of the academic community.

Second, I would like to express my gratitude to my promotor Jan Kwakkel who was a great counterweight. Your knowledge on deep uncertainty research, practical mindset, critical thinking, and writing has been essential for the PhD. I cannot get the ABT-structure out of my head, and I am always wondering whether I should be writing. Also, you made the conferences and other events/trips a lot of fun – making sure that PhDs were drinking enough is important.

Third, I would like to thank my external supervisor Jelte Mense who has been a valuable part of the Police side of the PhD. You guided me around the Police organisation, even in times of COVID, and helped me with finding my place here. You always checked to ensure I was still enjoying the PhD process.

I would also like to thank the committee members for their interest in my PhD research. First, Louise Shelley for her critical feedback on the topic of counterfeit illicit trade, and her kindness during my research visit at TraCCC. I remember staying awake until 2:30 am to listen to your talk on the OC24, which resulted in a great collaboration between TraCCC and TU Delft. I would like to thank Matthias Winkenbach for this support during my PhD and our longstanding collaboration which I highly value. It is great to work with you, and you always make me feel welcome. Next, I would like to thank Matthijs Spaan for his critical feedback on my PhD dissertation and his enthusiasm for the project. Last, I would like to thank Martijn Warnier for his invaluable support during the lunches, drinks, and end-of-the-day talks.

The many collaborations and visits during my PhD were invaluable. I would like to thank TraCCC for sharing their knowledge on illicit trade, and for the amazing conversations during my visits. In particular, I would like to thank Layla, Edward, and Julia for our fruitful collaboration. Also, a special thanks to Elisa and Camilla for organizing DISCC together. I would like to thank MIT Center of Transportation & Logistics for the inspiring visits and the warm welcome every time I visited. Also, I would like

to thank Rita for our wonderful collaboration and her advice on a professional and personal level.

I want to extend my gratitude to all my colleagues at TPM for making the office such a comfortable place, and taking the time to have a coffee or a drink. In particular, this PhD journey would not have been the same without my colleagues at the Policy Analysis Section at TPM and all the social events. A special thanks to all the PhDs at our section (Alessandro, Ashta, Brayton, Damla, Ignasi, Joos, Liz, Nely, Omar, Palock, Ruth, Thorid and more) for having a lot of laughs at the Coffee Star, Thursday drinks, and many more events.

I would like to thank my fellow PhD council members: Anna, Josephine, Nienke, Javanshir, Jessie, Steven, Vittorio, Vladimir. Working together to represent PhDs, support their well-being, and organize events for all PhDs was much fun. Janine, thank you for all the hard work on the Graduate School side and the great collaboration.

This dissertation also marks the end of my PhD work at the National Police AI Lab. In particular, I would like to thank Theo van der Plas, Ron Boelsma, Marius Kok, and Bas Testerink for believing in my research and making it possible. Next, I would like to thank TWO for the fun times on Mondays in Nieuwegein. Especially, Tamar Kastelein who has been part of my journey since the beginning, a great colleague and friend, and the many great memories with working together at the office at Rotterdam. A special thanks to the "Woerden" squad (Elize, Jelte, Jeroen, Judith, Mark, Michiel) for the many fun times, especially during COVID and beyond. I will never forget the conversations about The Bachelor, and our visits to the McDonald's. I would also like to thank Max and Dennis, who were part of the TU Delft NPAI lab, for their amazing drone research.

I want to thank all my colleagues at the Police who have helped me and supported me during my PhD. In particular, those who have helped with supervising Master students and with spreading the message of my research (Erik, Esther, Paul, Tom, Mario, Wout, Ryan, Robert, Thom, and more). A special thanks to Ron van den Bosch for this great collaboration and the inspiring conversations.

The next gratitude goes to my Master students: Lucas, Roos, Bruno, Ryan, Veronica, Hanna, Hillary, Ewout, and Thomas. You have taught me so much about how to be a supervisor, and about the amazing research you have been doing. It was a great pleasure to work with you all.

This journey would not have reached an end without my amazing paranymphs and friends: Karoline and Irene. Karoline, thank you for the many coffees and walks in Rotterdam during COVID, the TRAIL conferences, the trip to Mexico, many beers, and talks about our PhD and life in general. Irene, my PhD bestie, I really could not have done this PhD without you. Whether it was an in-depth discussion on the PhDs, support with supervisors, being angry with the HPC, or talking about what to wear to the conference, you were always there for me. I cannot express how much our everyday chats have meant to me, and how much they helped me to end this PhD. It was amazing to be on trips with you – Mexico, spending 24/7 in Marina Bays Sands in Singapore or having beers in London. Spending so much time together that even our supervisors were mixing us up – Dutch Twins. I want to thank you so much, and I am so grateful to call you one of my best friends after this journey.

I would like to express my gratitude to all my friends who were always up for a coffee, drinks, and a chat. Their friendship has been crucial during my PhD – in moments of stress or just to relax. The friends that have been there since the beginning of my studies at TPM, EPA, Curius, Laga, Rotterdam, Boston, and many more - it was amazing having you around. A special thanks to Carlie and Laurette for keeping up with me as roommates during the beginning of my PhD and surviving COVID-19 together.

Next, I would like to thank my family for their everlasting support and trust in my PhD. This PhD journey would not have been the same without your unconditional love. Last but not least, I would like to thank my love, Ron. You always expressed much interest in my PhD, even though you had no clue what I was talking about. Thank you for patiently listening to my concerns, cooking me amazing meals, and always taking care of me. Forever grateful to have you in my life.

Isabelle Maria van Schilt
Rotterdam, October 2024

SUMMARY

The COVID-19 pandemic led to a steep rise in the worldwide demand for Personal Protective Equipment (PPE) such as face masks, respirators, gloves, and goggles. PPE can be divided into two categories: medical and non-medical. Medical PPE is certified and typically comes with a higher price and profit margin, making it an attractive counterfeiting target for fraud-involved organizations, i.e., legitimate companies engaged in fraud. As a result, a significant number of organizations engaged in fraud entered the market after the initial stages of COVID-19, trying to sell non-medical PPE as medical. However, detecting counterfeit PPE has been challenging for law enforcement as (1) criminals take advantage of legitimate supply chains to mask their counterfeits, also known as piggybacking, (2) the legitimate supply chains for PPE to battle COVID-19 were partly new as well, so there was little historical data, and (3) fraud-involved organizations obfuscate their data as much as possible. Thus, detecting and effectively intervening in this largely invisible supply chain is difficult for law enforcement.

The counterfeit PPE case is just one example where supply chain visibility is of the utmost importance. Supply chain visibility means the ability to track parts, components, or products in transit from supplier to customer, addressing the actors' capability to monitor and trace the movement of goods with accurate and timely information. Even in this digital era, the data required to improve supply chain visibility, such as data on demand, inventory levels, processing times of a manufacturer, and transportation times, is often sparse due to the actors' reluctance to share information. This data sparseness leads to uncertainties about the operation within a supply chain (e.g., inventory, travel times) as well as about the overall structural composition and geographical boundaries (e.g., number and location of the actors). Illicit supply chains, in particular, suffer from limited information and a high level of uncertainty, making it challenging to effectively disrupt this criminal supply chain. In recent years, many studies in the supply chain field have focused on a new generation of information and communication technology systems to collect data for improving supply chain visibility. However, gathering more data using, for example, collaboration between actors in the supply chain is not always possible, especially in the case of an illicit supply chain.

An often-used method for gaining an understanding of a supply chain is simulation modeling, which is used to get insight into the behavior of complex systems, study relations over time, and explore future ("what-if") scenarios. Model calibration, i.e., the process of tuning and estimating the simulation model parameters using observed data to match the real system, is essential. However, research on calibrating supply chain simulation models in cases where the available data is sparse, is still lacking. Additionally, most research on model calibration in logistics is primarily aimed at estimating the parameters (i.e., parametric uncertainty) rather than find-

ing the best fit for the model structure (i.e., structural uncertainty). When calibrating a simulation model with sparse data, there is a large variety of plausible simulation models that could explain the sparse observations about the real-world supply chain, i.e., equifinality. Only focusing on one model for analysis could, therefore, lead to a "poor" understanding of the system and, hence, ineffective interventions in the real world. Thus, a diverse ensemble of models is needed to analyze the robustness of interventions in such cases, rather than relying on a single model. Yet, it is harder to find such an ensemble. Research on how to generate a diverse ensemble of reconstructions of a supply chain that could be used for identifying robust intervention is lacking. Therefore, this dissertation addresses the following research question:

How to generate a diverse ensemble of reconstructions of a supply chain, in cases where the available data is sparse?

Throughout this dissertation, we use a simulation calibration approach in combination with a ground truth set-up. For this, we develop a ground truth discrete event simulation model of a stylized counterfeit PPE supply chain as case study. We extract data from this model, systematically vary the degree of data sparseness, and assess the extent to which various model calibration techniques can still reconstruct the underlying supply chain. Four steps are taken to answer the main research question.

First, we provide a classification of data sparseness for supply chain visibility. A literature review is conducted on data sparseness and supply chain visibility, and a quantitative analysis is performed to assess the impact of data sparseness on supply chain visibility. Based on a review of supply chain visibility and data quality literature, we propose to characterize data sparseness as a lack of data quality across the entire supply chain, where data sparseness can be classified into three dimensions: noise, bias, and missing values. The quantitative analysis relies on a stylized simulation model of a moderately complex illicit supply chain. We use scenarios involving actors with different data perspectives in the supply chain, either supply or demand-oriented, to evaluate the combined effect of the individual dimensions of data sparseness. Results show that when a data sparseness of 90% is applied, supply chain visibility reduces to 52% for noise, to 65% for bias, and to 32% for missing values. The scenarios also show that companies with a supply-oriented view typically have a higher supply chain visibility than those with a demand-oriented view. The classification and assessment offer valuable insights for improving data quality and for enhancing supply chain visibility.

Second, we analyze the extent to which various model calibration techniques can identify the underlying parameters of a supply chain model when increasing the degree of data sparseness. In this step, we investigate a subset of model calibration techniques that are promising for handling sparse data: Approximate Bayesian Computing and Genetic Algorithms. We evaluate the quality-of-fit of these two model calibration techniques and a reference technique, Powell's Method, with a counterfeit PPE simulation model, given a systematic increase in missing values. The results demonstrate that these techniques are suitable for calibrating the parameters of a linear supply chain model with randomly missing values. This step offers a first insight

into the quality-of-fit for the model calibration techniques in the case of sparse data with parametric uncertainty.

Third, we evaluate the quality-of-fit for the model calibration techniques for reconstructing the underlying structure of a supply chain, i.e., for structural uncertainty, when the available data is sparse. Calibration methods for simulation models of illicit supply chains typically have to deal both with sparse data, and with a partially unknown structure of the supply chain. We evaluate the quality-of-fit of a reference technique, Powell's Method, and three model calibration techniques that have shown promise in the case of sparse data: Approximate Bayesian Computing, Bayesian Optimization, and Genetic Algorithms. We parameterize structural uncertainty using the System Entity Structure approach. The results demonstrate that Bayesian Optimization and Genetic Algorithms are suitable for reconstructing the underlying structure of an illicit supply chain for a varying degree of data sparseness. Both techniques identify a structurally diverse set of optimal solutions that fit with the sparse data. For a comprehensive understanding of the supply chain structures, or graphs, approximating the ground truth, we propose to combine the results of the two techniques. This step indicates that future research should focus on developing a combined algorithm and on incorporating solution diversity.

Fourth, we assess the potential of the Quality Diversity algorithm for generating a diverse ensemble of supply chain simulation models (i.e., solution diversity) in the case of sparse data, for both parametric and structural uncertainty. When calibrating a simulation model, there is a large variety of plausible simulation models that could explain the sparse observations about the real-world supply chain. A novel approach for generating this diverse set of plausible models is the Quality Diversity algorithm. The results show that the Quality Diversity algorithm is able to generate a diverse ensemble of supply chain models, including the ground truth. As expected, the Quality Diversity algorithm successfully identifies the structure of the ground truth most frequently at 0% of data sparseness. When more data sparseness is present, the Quality Diversity algorithm is prone to overfitting in more complex structures. We also highlight the importance of gathering information on the upstream supply chain to accurately reconstruct the counterfeit PPE supply chain.

In conclusion, this dissertation offers a first insight into generating a diverse ensemble of reconstructions of a supply chain in the case of sparse data using a simulation approach. We highlight three main scientific contributions. First, this dissertation is the first study that systematically varies the degree of data sparseness to evaluate the impact of the various model calibration techniques. Although the exact degree of data sparseness is often unknown in real life, this research gives a scientific insight into the impact of the degree of data sparseness on supply chain visibility and supply chain modeling. This set-up allows us to first theoretically assess the quality-of-fit of model calibration techniques before applying them in real life. Second, this dissertation fills the research gap of calibrating the structure of the supply chain simulation model in addition to fitting its parameters. Especially in the case of illicit supply chains, the structural composition and the geographical locations are of importance for decision-making. This research proposes a tool for creating various structures of a supply chain simulation model for model calibration, and presents metrics to com-

pare these structures. Third, the results identify model calibration techniques that are suitable for accurately reconstructing a supply chain characterized by sparse data, for both parametric and structural uncertainty. However, the techniques often overfit to more complex supply chain structures, or graphs, than the true underlying supply chain. Additionally, when calibrating a simulation model with sparse data, diversity should be included in terms of the use of multiple seeds, the combination of multiple techniques, and the use of solution diversity.

For supply chain practitioners and decision-makers, this dissertation presents three main contributions to practice. First, this dissertation offers insight into data sparseness for supply chain visibility and modeling. It is valuable for supply chain practitioners and decision-makers to have an understanding of how to cope with the sparseness of currently available data and how this sparseness impacts supply chain visibility. Second, this research highlights the importance of gathering information on the upstream supply chain. Third, a contribution of this dissertation to practice is that supply chain practitioners should recognize that there is not a single model for a supply chain when the available data is sparse, but there could be multiple models. This could help in making more robust decisions on, for example, interventions that are effective for disrupting an illicit supply chain. The next crucial step for future research is to evaluate the effectiveness of this diverse ensemble of reconstructions of supply chains for identifying these robust interventions, both theoretically and in practice.

SAMENVATTING

De COVID-19 pandemie leidde tot een sterke stijging van de wereldwijde vraag naar *Personal Protective Equipment* (PPE) zoals mondkapjes, handschoenen en veiligheidsbrillen. PPE kan worden onderverdeeld in twee categorieën: medisch en niet-medisch. Medische PPE zijn gecertificeerd en hebben meestal een hogere prijs en winstmarge, waardoor ze een aantrekkelijk doelwit zijn voor frauduleuze organisation (legitieme bedrijven die fraude plegen) om ze te vervalsen. Als gevolg hiervan kwam er een aanzienlijk aantal frauduleuze organisaties op de markt gedurende COVID-19, die probeerden niet-medische PPE te verkopen als medisch. Voor handhaving is het opsporen van vervalste PPE een grote uitdaging omdat (1) criminelen gebruik maken van legitieme goederenketens, ofwel *supply chains*, om hun vervalsingen te maskeren, ook bekend als *piggybacking*, (2) veel legitieme supply chains voor PPE ten tijde van COVID-19 waren ook nieuw, waardoor er weinig historische gegevens beschikbaar waren, en (3) frauduleuze organisaties maskeren of manipuleren hun gegevens zoveel mogelijk om onder de radar te blijven. Daarom is het voor handhavingsorganisaties zoals de Politie moeilijk om inzicht te krijgen in deze grotendeels onzichtbare supply chain en om effectief in te grijpen.

De casus omtrent vervalste PPE is slecht één voorbeeld waarbij de zichtbaarheid van de supply chain van groot belang is. De zichtbaarheid van de supply chain, ofwel *supply chain visibility*, richt zich op het kunnen traceren van onderdelen, componenten of producten die worden vervoerd van leverancier naar klant. Hierbij gaat het om in welke mate de actoren de verplaatsingen van goederen kunnen monitoren met accurate en actuele informatie. Zelfs in het huidige digitale tijdperk zijn de gegevens die nodig zijn om de zichtbaarheid van de supply chain te verbeteren vaak schaars. Denk hierbij aan informatie zoals vraag-aanbod gegevens, voorraadniveaus, verwerkingstijdens van een fabrikant en transporttijden. Een reden is dat actoren in de supply chain terughoudend kunnen zijn met het delen van data. Dit gebrek aan informatie leidt tot onzekerheden over de operatie van de supply chain (bijvoorbeeld voorraadhoogtes, transporttijden), en over de structurele samenstelling en geografische karakteristieken van de supply chain (bijvoorbeeld het aantal en de locatie van actoren). Met name bij illegale supply chains is er sprake van beperkte informatie en een hoge mate van onzekerheid, waardoor het effectief verstoren van deze illegale supply chains een grote uitdaging is. In de afgelopen jaren heeft onderzoek in het supply chain domein zich voornamelijk gericht op een nieuwe generatie informatie- en communicatietechnologiesystemen om gegevens te verzamelen voor het verbeteren van de *supply chain visibility*. Het verzamelen van meer gegevens met behulp van, bijvoorbeeld samenwerking tussen actoren in de keten, is echter niet altijd vanzelfsprekend, zeker in het geval van illegale supply chains.

Een veelgebruikte methode om de supply chain te modelleren is simulatie, een techniek om inzicht te krijgen in het gedrag van complexe systemen, het herkennen van verbanden, en het verkennen van toekomst (*"what-if"*) scenario's. Modelkalibratie, het proces van het schatten van de parameters van een simulatiemodel met behulp van observaties om het gedrag van het model af te stemmen op de werkelijkheid, is essentieel. Onderzoek naar het kalibreren van supply chain simulatiemodellen in situaties waar de beschikbare data schaars is, ontbreekt echter nog. Daarnaast heeft het meeste onderzoek naar modelkalibratie in de logistiek voornamelijk betrekking op het kalibreren van de parameters (parametrische onzekerheid) en niet op het kalibreren van de modelstructuur (structurele onzekerheid). Bij het kalibreren van een simulatiemodel met schaarse data, is er sprake van equifinaliteit. Dit betekent dat er meerdere plausibele simulatiemodellen zijn die de schaarse observaties van de werkelijke supply chain kunnen verklaren. Slechts één model analyseren kan leiden tot een verkeerd beeld van het systeem en daardoor tot mogelijk misplaatste interventies in de echte wereld. Er is dus een gevarieerd ensemble van modellen nodig om de robuustheid van interventies in dergelijke gevallen te analyseren, in plaats van gebruik te maken van één model. Een ensemble is echter moeilijker te vinden. Onderzoek naar het genereren van een divers ensemble van reconstructies van een supply chain die gebruikt kunnen worden voor het identificeren van robuuste interventies ontbreekt. Daarom richt dit proefschrift zich op de volgende onderzoeksvraag:

Hoe genereren we een divers ensemble van reconstructies van een supply chain in situaties waar de beschikbare data schaars is?

In dit proefschrift gebruiken we een simulatie-kalibratie methode in combinatie met een *ground truth* analyse. Hiervoor ontwikkelen we een *ground truth* simulatiemodel van een nagebootste supply chain van vervalste PPE als case study. We gebruiken een *discrete event* simulatie model. We extraheren data uit dit model, variëren systematisch de mate van de schaarste van de data en beoordelen in hoeverre verschillende modelkalibratietechnieken nog steeds de onderliggende supply chain kunnen reconstrueren. Vier stappen worden doorlopen om de onderzoeksvraag te beantwoorden.

Als eerste classificeren we dataschaarste voor *supply chain visibility*. Er wordt een literatuurstudie gedaan omtrent dataschaarste en de zichtbaarheid van supply chains. Vervolgens wordt er een kwantitatieve analyse uitgevoerd om de impact van de dataschaarste op *supply chain visibility* te meten. Op basis van de literatuurstudie, karakteriseren we dataschaarste als een gebrek aan datakwaliteit in de hele supply chain, waarbij dataschaarste kan worden onderverdeeld in drie dimensies: *noise*, *bias* en missende datapunten. De kwantitatieve analyse is gebaseerd op een nagebootst simulatiemodel van een redelijk complexe illegale supply chain. Scenario's worden gebruikt om het gecombineerde effect van de afzonderlijke dimensies te evalueren vanuit actoren met verschillende perspectieven in de supply chain, zowel aanbod- als vraaggericht. De resultaten laten zien dat bij een dataschaarste van 90% de zichtbaarheid van de supply chain afneemt tot 52 % voor *noise*, tot 65% voor *bias* en tot 32% voor missende datapunten. Verder laten de scenario's zien dat bedrijven met een aanbodgeoriënteerde visie (begin van de supply chain) doorgaans een hogere zicht-

baarheid van de supply chain hebben dan bedrijven met een vraaggeoriënteerde visie (eind van de supply chain). De classificatie en impact analyse bieden waardevolle inzichten voor het verbeteren van de dataschaarste en voor het verbeteren van de *supply chain visibility*. Deze classificatie wordt in de rest van het onderzoek gebruikt.

Ten tweede analyseren we de mate waarin verschillende modelkalibratietechnieken de parameters van een supply chain model correct kunnen identificeren bij een toenemende mate van dataschaarste. In deze stap onderzoeken we een subset van modelkalibratietechnieken die veelbelovend zijn met betrekking tot het omgaan met schaarse data: *Genetic Algorithms* en *Bayesian Optimization*. We analyseren de *quality-of-fit* van deze twee modelkalibratietechnieken en een referentietechniek, *Powell's Method*, voor een simulatiemodel over vervalste PPE, gegeven een systematische toename in missende datapunten. De resultaten tonen aan dat deze technieken geschikt zijn voor het kalibreren van de parameters van een lineair supply chain model met willekeurig missende datapunten. Deze stap biedt een eerste inzicht in de kwaliteit van de modelkalibratietechnieken in het geval van schaarse data en parametrische onzekerheid.

Ten derde testen we de *quality-of-fit* van de modelkalibratietechnieken voor het reconstrueren van de onderliggende structuur van een supply chain, ofwel voor structurele onzekerheid, wanneer de beschikbare data schaars is. Voor het simuleren van illegale supply chains moet de modelkalibratie kunnen omgaan met schaarse data. Daarnaast hebben deze supply chains te maken met structurele onzekerheid. Dit betekent dat de modelkalibratietechnieken de modelstructuur moeten identificeren. We analyseren de *quality-of-fit* van een referentietechniek, *Powell's Method*, en drie modelkalibratie technieken die veelbelovend zijn voor het omgaan van schaarse data: *Approximate Bayesian Computing*, *Bayesian Optimization*, en *Genetic Algorithms*. Hiervoor gebruiken we een simulatiemodel van een nagebootste supply chain van vervalste PPE als *ground truth*. We extraheren data uit dit model en variëren systematisch de kwaliteit van de data op het gebied van *noise*, *bias*, en missende datapunten. We formaliseren structurele onzekerheid met behulp van *System Entity Structures*. De resultaten tonen aan dat *Bayesian Optimization* en *Genetic Algorithms* geschikt zijn voor het reconstrueren van de onderliggende structuur van een illegale supply chain, gegeven een variërende mate van schaarsheid van de data. Beide technieken identificeren een diverse set van optimale oplossingen die goed overeenkomen met de schaarse data. Voor een alomvattend beeld van supply chains die de *ground truth* benaderen, raden wij aan om de resultaten van de twee technieken te combineren. Vervolgonderzoek moet zich richten op het ontwikkelen van een gecombineerd algoritme en op het integreren van diversiteit in de oplossingen.

Ten vierde beoordelen we de potentie voor het *Quality Diversity* algoritme voor het genereren van een divers ensemble van supply chain simulatiemodellen (diversiteit in de oplossingen) in het geval van schaarse data, voor zowel parametrische als structurele onzekerheid. Bij het kalibreren van simulatiemodellen is er een grote diversiteit van plausibele supply chains die de schaarse observaties van de werkelijke supply chain kunnen verklaren. Een relatief onbekende methode voor het genereren van een diverse set van mogelijke simulatiemodellen is het *Quality Diversity* al-

goritme. De resultaten laten zien dat het *Quality Diversity* algoritme in staat is om een divers ensemble van supply chain simulatiemodellen te genereren, waaronder de *ground truth*. Zoals verwacht vindt het *Quality Diversity* algoritme de structuur van de *ground truth* het vaakst bij 0% dataschaarste. Wanneer de data schaarser wordt, is het *Quality Diversity* algoritme gevoelig voor *overfitting* van complexere structuren. We benadrukken hier nogmaals het belang van het verzamelen van informatie over de *upstream* supply chain om de vervalste PPE supply chain te kunnen reconstrueren.

Concluderend, dit proefschrift biedt als eerste inzicht in het generen van een divers ensemble van reconstructies van een supply chain in het geval van dataschaarse met behulp van simulate. Dit proefschrift is het eerste onderzoek is dat systematisch de mate van dataschaarste varieert om de impact van de verschillende modelkalibratietechnieken te evalueren. Hoewel de exacte hoeveelheid dataschaarste in de praktijk vaak onbekend is, geeft dit onderzoek wetenschappelijk inzicht in de impact van dataschaarste op de zichtbaarheid van de supply chain en supply chain modellering. Deze opzet stelt ons in staat om de kwaliteit van modelkalibratietechnieken theoretisch te beoordelen voordat ze in de praktijk worden toegepast. Daarnaast behandelt dit proefschrift het kalibreren van de structuur van een supply chain simulatiemodel, als toevoeging op het schatten van de parameters. Vooral voor illegale supply chains zijn de structurele samenstelling en de geografische locaties van belang voor besluitvorming. Daarom presenteert dit onderzoek een methode voor het creëren van verschillende structuren van een supply chain simulatiemodel, en presenteert het metrieken om deze structuren te vergelijken. Een andere wetenschappelijke bijdrage is dat de resultaten laten zien dat enkele modelkalibratie technieken geschikt zijn voor het accuraat reconstrueren van een supply chain die gekenmerkt wordt door schaarse data, voor zowel parametrische als structurele onzekerheid. De technieken hebben last van *overfitten* van complexere supply chains dan de echte onderliggende supply chain. Bij het kalibreren van een simulatiemodel met schaarse data moet rekening gehouden worden met diversiteit in de vorm van het gebruik van meerdere *seeds*, de combinatie van meerdere technieken, en de oplossingsdiversiteit.

Voor supply chain professionals en besluitvormers biedt dit proefschrift inzicht in het concept van dataschaarste voor de *supply chain visibility* en modellering van supply chains. Het is waardevol voor supply chain professionals en besluitvormers om inzicht te hebben in hoe om te gaan met de schaarste van de beschikbare data en hoe dit de zichtbaarheid van de supply chain beïnvloedt. Daarnaast benadrukt dit onderzoek het belang van het verzamelen van informatie over de *upstream* supply chain. Een andere bijdrage aan de praktijk is dat supply chain professionals moeten realiseren dat er niet één mogelijke beschrijving is van een supply chain wanneer de beschikbare data schaars is, maar dat er meerdere mogelijke modellen zijn. Dit kan helpen om robuustere beslissingen te nemen over, bijvoorbeeld, interventies die effectief zijn voor het verstoren van een illegale supply chain. De volgende cruciale stap voor toekomstig onderzoek is het evalueren van de effectiviteit van een divers ensemble van reconstructies van supply chains voor het identificeren van robuuste interventies, zowel theoretisch als in de praktijk.

1

INTRODUCTION

1.1. BACKGROUND

In today's unpredictable and changing world, ensuring that the right quantity of a product is at the right place at the right time is increasingly challenging. Disruptions such as the outbreak of a global pandemic (2019), the blockage of the Suez Canal (2021), the global semiconductor shortage (2021), the Ukraine war (2022), and the Houthi attacks on container ships in the Red Sea (2024) led to the shutdown of factories, delays in maritime transport, shortage of essential products, and extremely high prices (Zhao et al., 2023; Berger, 2024). In 2023, disruptions caused an average of $82 million in annual losses per company in key industries (Reuters, 2023). These disruptions affect the legal supply chain as well as illicit supply chains, causing shortages and losses in the former and opportunities and reduced risk in the latter.

In this dissertation, we use supply chains for counterfeit Personal Protective Equipment (PPE) as the example case. The COVID-19 pandemic led to a steep rise in the worldwide demand for PPE such as face masks, respirators, gloves, and goggles (Omar et al., 2022). PPE can be divided into two categories: medical and non-medical. Medical PPE is certified and typically comes with a higher price and profit margin, making it an attractive target for fraud-involved organizations (i.e., legitimate companies engaged in fraud) (Ippolito et al., 2020). As a result, a significant number of organizations engaged in fraud entered the market after the initial stages of COVID-19, trying to sell non-medical PPE as medical (Hashemi et al., 2022). However, detecting counterfeit PPE has been challenging for law enforcement as (1) criminals take advantage of legitimate supply chains to mask their counterfeits, also known as piggybacking, (2) legitimate supply chains are impacted by COVID-19, on which there is little historical data, and (3) fraud-involved organizations obfuscate their data as much as possible. Thus, detecting and effectively intervening in this largely invisible supply chain is difficult for law enforcement.

The counterfeit PPE case is just one example where supply chain visibility is of the utmost importance (Zhao et al., 2023). Supply chain visibility means the ability to track parts, components, or products in transit from supplier to customer, addressing the actors' capability to monitor and trace the movement of goods with accurate and timely information (Saqib et al., 2019; Kalaiarasan et al., 2022). When supply

chain visibility increases, logistical processes within the supply chain can be more effectively aligned (Srinivasan & Swink, 2018; Kalaiarasan et al., 2022). In recent years, supply chain visibility has become key for improving supply chain management and design (Busse et al., 2017; Roy, 2021). Successful supply chain management is heavily dependent on the availability of information shared by multiple actors within the supply chain (Brun et al., 2020). More specifically, a supply chain is a network of actors that produce and distribute a specific product or service from supplier to end-user, i.e., from raw materials to end-product (Fisher, 1997; Christopher, 2016). Three main flows can be distinguished for supply chains: (1) goods flows, (2) information flows, and (3) financial flows (Min & Zhou, 2002; Stadtler & Kilger, 2002).

Even in this digital era, the data required to improve supply chain visibility, such as data on demand, inventory levels, processing times of a manufacturer, and transportation times, is often sparse (Guida et al., 2023; Spreitzenbarth et al., 2024). Only 6% of the companies claim to have complete supply chain visibility, according to the GEODIS Supply Chain Worldwide Survey, even though over 50% of the supply chain companies use or are planning to use digital technologies (Macri, 2018; GEODIS, 2020; The Business Continuity Institute, 2022). One of the causes for data sparseness is reluctance among actors within a supply chain to share (high-quality) data for various reasons such as competition and high costs (Boone et al., 2019), or because of illegal behavior of supply chain partners involved in fraud (Ficara et al., 2021).

Data sparseness leads to uncertainties about the operations of actors within the supply chain (e.g., inventory levels, transportation times) as well as about the overall structural supply chain composition and geographical locations (e.g., how many actors are involved, where the actors are located). Especially in the case of illicit supply chains, there is little information and there are many uncertainties around the operations and the supply chain's structural composition, making it challenging to effectively disrupt this illegal supply chain. Criminals use various modi operandi, routes, transportation modes, multiple actors, communication channels, and business models, impacting the flow of goods and the structure and geographical aspects of the supply chain (Duijn et al., 2014; Anzoom et al., 2021). Such a supply chain is characterized by deep uncertainty due to the wide range of possible structural compositions within the supply chain. Deep uncertainty is defined as a situation "where analysts do not know, or the parties to a decision cannot agree on, (1) the appropriate conceptual models that describe the relationships among the key driving forces that will shape the long-term future, (2) the probability distributions used to represent uncertainty about key variables and parameters in the mathematical representations of these conceptual models, and/or (3) how to value the desirability of alternative outcomes" (Lempert et al., 2003, p. xii).

Improving supply chain visibility is very challenging for a supply chain characterized by complexity, sparse data, and (deep) uncertainty. Gaining more insight into the effect of data sparseness on supply chain visibility is essential for making improvements. The first step is to clearly define the concept of data sparseness in the context of supply chains. Although a large variety of types of poor data is presented in the literature, a clear and concise formalization of data sparseness is still lacking, especially in the field of supply chain management (Laranjeiro et al., 2015; Kalaiarasan et al., 2022). Studies show that sparse data resulting from data errors impacts sup-

ply chain visibility and decision-making (Oliveira & Handfield, 2019; Agrawal et al., 2022). However, the exact effect of different dimensions of data sparseness on supply chain visibility is still poorly understood in the literature.

In recent years, many studies in the supply chain field have focused on a new generation of information and communication technology (ICT) systems to collect data for improving supply chain visibility (Topsector Logistiek, 2019; Kalaiarasan et al., 2023). Examples of these systems are Internet-of-Things, Radio-frequency identification transponders, and Blockchain (Pero & Rossi, 2014; Calatayud et al., 2019; Kumar et al., 2022). One of the key aspects of many of these systems is the collaboration between actors within the supply chain (Pero & Rossi, 2014; Kalaiarasan et al., 2023). In our example case of counterfeit PPE supply chains, collaboration and sharing of data among fraud-involved organizations and between fraud-involved organizations and law enforcement are out of the question, since data can potentially reveal illegal activities. Collaboration between supply chain partners is, therefore, not always a given. So, the central question in this research is: how to improve supply chain visibility, given that the currently available data is sparse?

1.1.1. MODELING SUPPLY CHAINS

For improving supply chain visibility, it is important to gain insight into the supply chain itself (descriptive analytics), explore possible future situations (predictive analytics), and examine how to transform the future supply chain into a desired state (prescriptive analytics) (Wang et al., 2016; Tiwari et al., 2018). A powerful approach for analyzing a supply chain is simulation (Tiwari et al., 2018). Simulation is a way of getting insight into the behavior of complex systems, recognizing relations over time, and exploring future ("what-if") scenarios (Shannon, 1998; Law et al., 2000). A simulation model is conceptualized as consisting of variables and relations. Many variables in the model need an initial value to capture the initial model state and behavioral characteristics that are consistent with the behavior of the system. These initial values are called parameters; more specifically, parameters of components of the model. Some of the parameter values might be observed directly, while others are unobservable and thus have to be tuned to match the state and behavior of the simulation model with its real-world counterpart.

Model calibration is the process of tuning and estimating the model parameters, using observed system data, to improve the similarity between the model and the real-world system (Wigan, 1972; Ören, 1981). The goal of model calibration is to find those parameter values for which the behavior of the simulation model is as close as possible to the observed behavior of the real system using real data (Liu et al., 2017). Two types of uncertainty in models for which calibration is needed can be distinguished: parametric uncertainty, i.e., uncertainty in the initial values of the model's parameters, and structural uncertainty, i.e., uncertainty in the structure of the model like the modeling equations and structural composition of the model (Webster & Sokolov, 1998; Park & Schneeberger, 2003; Park & Qi, 2005). Model calibration primarily involves adjusting model parameters rather than altering the model structure (Wigan, 1972; Ören, 1981; Coenen et al., 2018). Similarly, most studies in the field of logistics have explored parametric uncertainty rather than addressing structural uncertainty (Halim et al., 2016; Coenen et al., 2018; Moallemi & Köhler, 2019). Especially

in a supply chain (for an illicit product) characterized by sparse data, the models are characterized by both parametric and structural uncertainty.

Additionally, model calibration should be able to handle the available sparse data. A subset of model calibration techniques that seems to be able to handle sparse data in other fields can be identified. For example, Evolutionary Algorithms optimize high-dimensional problems with sparse data (Ren & Wu, 2013), Bayesian Inference handles sparse datasets in machine learning (Jalali et al., 2017), and Data Assimilation predicts simulation models in real-time with sparse data (Xie, 2018). A number of studies have investigated the calibration of simulation models in the context of data sparseness (Liu et al., 2017; de Groot & Hübl, 2021; De Santis et al., 2022). These studies focus, however, on a case study, e.g., an emergency department, with one sparse dataset, and they apply only one model calibration technique, making the effectiveness of these techniques unclear for supply chains or broader levels of data sparseness. Hence, it is still unknown how various model calibration techniques perform for modeling supply chains given a varying degree of data sparseness.

1.1.2. MODELING ROBUST INTERVENTIONS

When modeling a complex system, like a supply chain, characterized by sparse data and deep uncertainty, there is a level of equifinality among the simulation models. In terms of model calibration, this means that multiple versions of the supply chain simulation model are coherent with the sparse real-world data. Hence, a complex system cannot be captured by a single theory or model (Page, 2021). Only focusing on one model for analysis could lead to a "poor" understanding of the system, and hence ineffective interventions in the real world (Thompson & Smith, 2019). Thus, an ensemble of models is needed to analyze the effectiveness of interventions in such cases instead of a single model (Veit, 2020).

In this case, the effectiveness of an intervention refers to the robustness of the intervention. An intervention is robust when it performs in a satisfactory way across a large majority of an ensemble of models (Walker et al., 2013). An analysis of robustness includes evaluating the performance of an intervention in many plausible models varying over a large set of assumptions, instead of describing the best-estimate model and evaluating the performance of the intervention only within this model (Lempert et al., 2013). The models within the ensemble needed for robustness analysis, therefore, have to be similar but distinct (Weisberg, 2012).

However, model calibration techniques typically lead to a single simulation model that fits the data best, instead of multiple simulation models. An ensemble of models can be created by choosing a user-defined number of near-optimal solutions, e.g., the top five. Yet, it is likely that this ensemble only contains very similar models, as configurations with slightly different parameters from the optimal solution typically outperform those with vastly different parameters. For robustness, a diverse group of explainable models is more desirable, yet harder to find (Durán & Formanek, 2018)[1]. Therefore, this research focuses on how to generate a diverse set of plausible supply chain models, using a simulation calibration approach that can deal with sparse data.

[1] For example, when using Google Maps to navigate between city A and city B, it provides three distinct routes using different highways. However, the three most optimal routes probably involve a slight adjustment in direction at the beginning or the end within the city, rather than using different highways.

1.2. RESEARCH GOAL AND QUESTIONS

This research aims to generate a diverse set of plausible supply chain models that can be used to identify robust interventions, where the model calibration techniques have to deal with sparse data. Hence, the main research question is:

How to generate a diverse ensemble of reconstructions of a supply chain, in cases where the available data is sparse?

To answer this research question, the following sub-questions need to be answered:

1. **How to classify data sparseness for supply chain visibility?**
   Gaining more insights into the effect of data sparseness on supply chain visibility is essential for making improvements. As a first step, a clear and concise classification of dimensions of data sparseness in the field of supply chain management is needed. A systematic review of the current state-of-the-art literature on supply chain visibility and data quality will be performed. The exact effect of different dimensions of data sparseness on supply chain visibility is still poorly understood. To gain more insight, a quantitative analysis will be carried out for a stylized simulation model of a moderately complex illicit supply chain. This sub-question (1) provides a classification of data sparseness in the context of supply chain visibility, and (2) assesses the impact of data sparseness on supply chain visibility.

2. **To what extent can various model calibration techniques identify the parameters of a supply chain simulation model when varying the degree of data sparseness?**

   Complex supply chains, for example, those involving counterfeit PPE, are characterized by partly unobservable behavior and sparse data, making it challenging to construct a reliable simulation model. Model calibration can help with this, as it is the process of tuning and estimating the model parameters with observed data of the system. A subset of model calibration techniques, Genetic Algorithms and Bayesian Inference, seems to be able to deal with sparse data in other fields (Vrugt & Beven, 2018; Mirjalili, 2019). However, it is unknown how these techniques perform when calibrating simulation models with sparse data. This sub-question analyzes the quality-of-fit of two model calibration techniques for a counterfeit PPE supply chain simulation model given an increasing degree of data sparseness.

3. **To what extent can various model calibration techniques reconstruct the underlying structure of a supply chain when varying the degree of data sparseness?**

   Data sparseness in complex supply chains, such as a counterfeit PPE supply chain, makes the supply chain largely invisible, resulting in uncertainty about its structural composition. This, in turn, makes it challenging to intervene and stop crime in a complex system like a supply chain. Simulation is a way to get

insight into the behavior of complex systems with computer models, using calibration to tune the parameters of the model to match its real-world counterpart. However, the extent to which model calibration techniques can accurately reconstruct the structure of a supply chain characterized by sparse data has not yet been investigated. To answer this question, we have to model structural uncertainty rather than only parametric uncertainty. This sub-question assesses the quality-of-fit of various model calibration techniques given structural uncertainty with a varying degree of data sparseness.

4. **How feasible is the quality diversity algorithm for generating a diverse ensemble of reconstructions of a supply chain when varying the degree of data sparseness?**
   Data on supply chains is often sparse due to reluctance among actors to share their data, making simulation modeling of supply chains difficult. Particularly, supply chain simulation models suffer from parametric and structural uncertainties as a result of this data sparseness. When calibrating a simulation model, there is a large variety of plausible simulation models that could explain the sparse observations about the real-world supply chain. A relatively unknown approach to generate this diverse set of plausible models is the Quality Diversity algorithm (Mouret & Clune, 2015; Fontaine et al., 2020). This study evaluates the feasibility of using the Quality Diversity algorithm to generate a diverse ensemble of supply chain simulation models for a varying degree of data sparseness.

Although generating a diverse ensemble of plausible supply chain models can potentially help to design robust policies and interventions in cases where only sparse data is available, this dissertation does not explicitly evaluate whether the plausible supply chain models actually enable more effective interventions. This would be a follow-up step after the development and evaluation of the model calibration techniques from this dissertation.

1.3. RESEARCH METHODS

The main research goal of this dissertation is to more accurately model supply chains in the presence of sparse data and theoretically evaluate various model calibration techniques to improve the fit between the model and observations. This section describes the research method for each sub-question. First, the systematic literature review for sub-question 1 is described. Second, the quantitative analysis for all subquestions using a ground truth set-up is outlined. Third, the method for developing the ground truth simulation model is discussed. Last, the set-up of the model calibration analysis for sub-questions 2 to 4 is presented.

1.3.1. LITERATURE REVIEW

The first sub-question is addressed by conducting a systematic literature review on two bodies of literature: (i) supply chain visibility and (ii) data quality. We follow the systematic search method described by van Wee and Banister (2016). Database engines such as Scopus and Google Scholar are used to identify the relevant literature.

The scope of this literature review is restricted to academic papers and books in English. Papers are selected based on the number of citations, while taking into account how recently the papers are published, not to miss recent contributions. For both streams of literature, the first step is to search the current state-of-the-art literature on specific keywords within a date range of 2000 to 2023. The second step focuses on finding additional papers using snowballing. Results of the literature review are used for identifying and operationalizing a classification for data sparseness in the field of supply chain visibility. This classification and its operationalization are the basis for the quantitative analysis for each of the remaining sub-questions.

1.3.2. QUANTITATIVE ANALYSIS USING GROUND TRUTH SET-UP

Each sub-question (1-4) involves a quantitative analysis using (a part of) a ground truth set-up. Figure 1.1 presents an overview of the ground truth set-up used in this project. Sub-question 1 is answered using only the upper part of the figure. Sub-questions 2, 3, and 4 are answered using the entire set-up. First, a ground truth simulation model of a stylized supply chain is developed. From this model, we extract ground truth data representing the true maximum or 0% of data sparseness. Next, we degrade the ground truth data by adding sparseness. For example, we randomly delete 10% of the data to account for missing values. When extracting data from the real world, data is inherently sparse and is therefore comparable to the sparse data in this figure. With the sparse data from the ground truth set-up, we calculate supply chain visibility to answer sub-question 1.

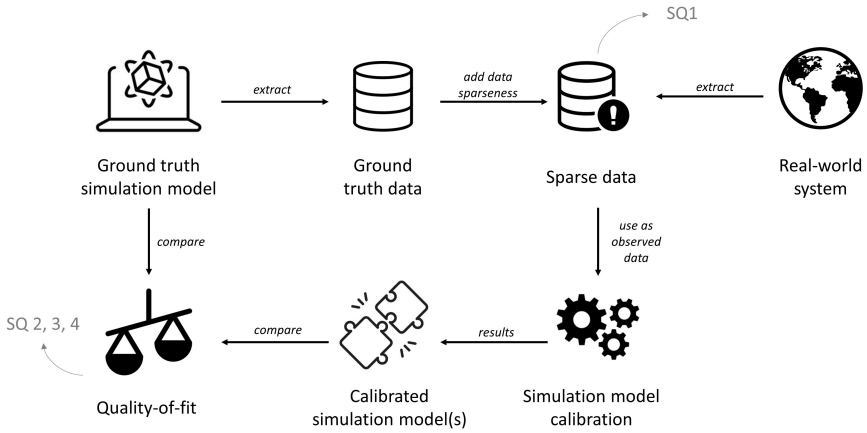


Figure 1.1: Overview of the Ground Truth Set-Up.

For the remaining sub-questions, we use the sparse data from the ground truth set-up as “observed” data of the system for the simulation model calibration. The simulation model calibration process involves creating many supply chain configurations and optimizing the most plausible configuration based on the smallest distance between the simulated data from the configurations and the sparse observation data. This results in optimal solutions(s) for the configuration of the supply chain simulation model, i.e., calibrated simulation model(s). While the model calibration

techniques aim to minimize the distance between the simulated data and sparse observation data, it does not guarantee that the calibrated simulation model(s) will be close to that of the ground truth model. The reasons are that the optimization is using sparse data, that different models can have a similar outcome, i.e., equifinality, or that the calibration technique gets stuck in a local optimum. Hence, we evaluate the quality-of-fit of the model calibration technique by comparing (the parameters and structure of) the calibrated simulation model(s) with the ground truth simulation model. Sub-questions 2 to 4 use this set-up to theoretically assess to what extent various model calibration techniques can reconstruct the ground truth supply chain when varying the degree of data sparseness.

1.3.3. GROUND TRUTH SIMULATION MODEL

The case study used for the ground truth simulation model is a stylized counterfeit PPE supply chain. This supply chain is interesting as it is inherently characterized by sparse data because (1) the production of counterfeit PPE during COVID-19 presented a new and unexpected phenomenon with little historical data, and (2) counterfeit PPE supply chains are operated by organizations selling fraudulent products that obfuscate as much data as possible (Hashemi et al., 2023). Corruption enables counterfeit PPE by facilitating fake certifications, bypassing inspections, and allowing illicit goods through customs, undermining law enforcement. In the supply chain connecting suppliers in the source country to customers in the destination country, different types of actors are involved, and a variety of transport modalities are used. Data on the operations of the supply chain is gathered using openly available information and expert interviews with multiple law enforcement agencies. This ensures that the ground truth model is realistic, allowing the research findings to be applicable in real-world settings. Joint research with the Terrorism, Transnational Crime & Corruption Center at George Mason University in the USA allows for gathering real-world information on the counterfeit PPE supply chain from a United States perspective. The gathered set of data is used for the development of a conceptual model of the supply chain and its constraints. The conceptual model is validated with experts. Figure 1.2 shows an example of a high-level conceptual model of a stylized counterfeit PPE supply chain.

For simulating the PPE supply chain, we use a discrete event simulation model since it is a widely used approach for simulating supply chains (Law et al., 2000; Robinson, 2005; Schmitt & Singh, 2009). Discrete event modeling simulates the operations of a system as a sequence of events at discrete time points, where each event changes the state within the system (Robinson, 2004). Discrete event simulation typically uses queues and resources to describe the system (Law et al., 2000). The models are stochastic in nature and the parameters are often represented by the use of statistical distributions (Tako & Robinson, 2008). This makes it suitable for representing the stochastic dynamics of supply chain operations. A limitation of discrete event simulation is that a lot of data is required to develop a model of the detailed operation of a supply chain.

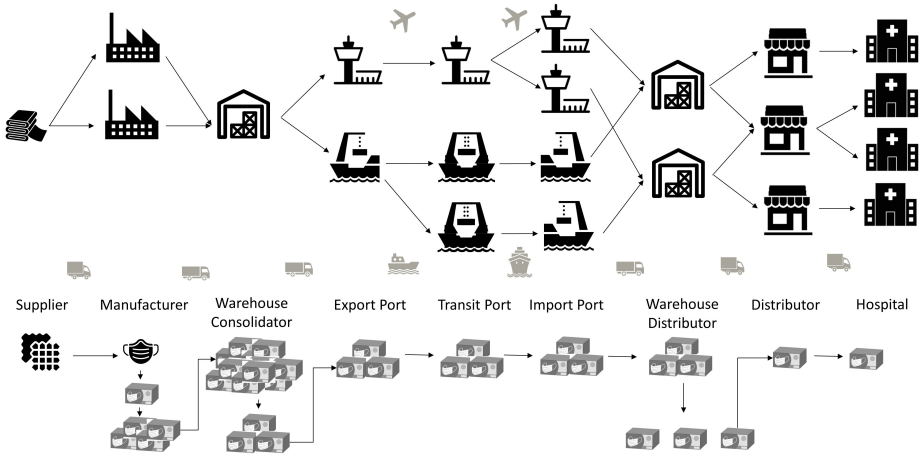


Figure 1.2: Example of Counterfeit PPE Supply Chain.

The simulation models in this project are created using the library *pydsol-core* [2] and *pydsol-model* [3] in Python. The library *pydsol-core* is a Python implementation of the Distributed Simulation Object Library (DSOL), originally implemented in Java by Jacobs (2005), and based on the formal definition of simulation models of Zeigler et al. (2018). The library *pydsol-model* is an additional layer on *pydsol-core* and includes standard model objects suitable for developing discrete event simulation models. The libraries are developed during this Ph.D. project by Alexander Verbraeck and Isabelle M. van Schilt. The code for the ground truth simulation model of this dissertation is available at GitHub[4].

[2] https://github.com/averbraeck/pydsol-core
[3] https://github.com/imvs95/pydsol-model
[4] The code for the ground truth simulation model is available at https://github.com/imvs95/complex_stylized_supply_chain_model_generator, and for the graph generator of open source shipping data at https://github.com/imvs95/port_data_graphs. The large datasets used in this research are available at https://doi.org/10.4121/adf4373c-7a9a-4d9c-a1ff-0f893d8d0b06.v1.

1.3.4. CONFIGURATION OF THE MODEL CALIBRATION ANALYSIS

The main goal of sub-questions 2 to 4 is to evaluate the quality-of-fit of various model calibration techniques when varying the degree of data sparseness. Table 1.1 gives an overview of the model calibration analysis per sub-question.

	Uncertainty	Dimensions of Sparseness	Model Calibration Techniques	Outcome
Sub-question 2	Parametric	Missing values	Powell, ABC, GA	One optimal solution
Sub-question 3	Structural	Noise, Bias, Missing values	Powell, ABC, GA, BO	One optimal solution
Sub-question 4	Parametric, Structural	Noise, Bias, Missing values	QD	Multiple optimal solutions

Table 1.1: Overview of Model Calibration Analysis per Sub-Question. (Powell = Powell's Method, ABC = Approximate Bayesian Computing, GA = Genetic Algorithms, BO = Bayesian Optimization, QD = Quality Diversity algorithm)

The first step is to assess to what extent model calibration techniques that seem suitable for handling sparse data perform in the case of parametric uncertainty. Parametric uncertainty means uncertainty in the values of the parameters of the simulation model (Webster & Sokolov, 1998; Parker, 2014). For sub-question 2, we focus on calibrating parameters on one dimension of data sparseness. We choose missing values as this impacts the quantity of the data, making it a more straightforward way of degrading data (Oliveira et al., 2005; Ehrlinger & Wöß, 2022). Regarding the model calibration techniques, we select one reference technique and two techniques that seem suitable for handling sparse data. We use Powell's Method as the reference technique since it is commonly used for calibrating simulation models (Liu et al., 2017). The reference technique is used to evaluate whether the other techniques work better. The other two model calibration techniques are Approximate Bayesian Computing (ABC) and Genetic Algorithms (GA). ABC is a technique for estimating the posterior distribution of model parameters using Bayesian statistics (Sadegh & Vrugt, 2014; Vrugt, 2016). This technique seems to be one of the most suitable techniques using Bayes' theorem for calibrating with sparse data as it is likelihood-free (Vrugt & Beven, 2018). GA is one of the oldest and most well-known evolutionary algorithms, i.e., a population-based optimization technique. It is widely used for calibration, especially in high-dimensional optimization problems where data is often sparse (Park & Qi, 2005; Ren & Wu, 2013; Slowik & Kwasnicka, 2020). The outcome of the model calibration techniques for sub-question 2 is one single optimal solution for the parameters of the supply chain simulation model.

On top of parametric uncertainty, the structure of a supply chain characterized by sparse data is often unknown, especially in the case of criminal activities. In the case of a supply chain, many interdependencies exist among various actors, making it difficult to view actors as independent components in a simulation model, unlike parameters (Baldissera Pacchetti, 2021). Since model calibration mostly focuses on tuning the model parameters and not the model structure, tuning both simultaneously is far more challenging than just tuning the parameters (Moore & Doherty, 2005; Co-

enen et al., 2018). Thus, sub-question 3 focuses on model calibration given structural uncertainty while keeping the parameters the same. This means that the model calibration techniques aim to find the best fitting model structure to match the underlying supply chain given a varying degree of data sparseness. For this sub-question, we use three dimensions of data sparseness, noise, bias and missing values, for a holistic analysis of their impact. We examine the impact of data sparseness dimensions both individually and in combination. In terms of the model calibration techniques, we again analyze Powell's Method, ABC, and GA. We include Bayesian Optimization (BO) as an additional prospective technique. BO is a technique that uses Bayes' theorem to search for the optimum by constructing the posterior distribution (van Hoof & Vanschoren, 2021). It is among the few techniques in the field of machine learning that are able to handle small data sets (Jalali et al., 2017). Similar to sub-question 2, the outcome of the selected model calibration techniques in sub-question 3 gives one single optimal solution for the structure of the supply chain.

In sub-questions 2 and 3, the model calibration techniques result in one single optimal outcome. However, the overall goal of this research is to generate an ensemble of reconstructions of a supply chain, not a single reconstruction. Thus, sub-question 4 focuses on generating a diverse set of plausible supply chain configurations. For this, we use a quality diversity (QD) algorithm. QD algorithms use evolutionary concepts to find optimal solutions at each point of the user-defined search space (Mouret & Clune, 2015; Chatzilygeroudis et al., 2021). QD is mostly used in the field of robotics and reinforcement learning (Pugh et al., 2015; Lim et al., 2022; Tjanaka et al., 2023). Since QD is a relatively new approach, it is still unexplored to what extent the technique would work for calibrating simulation models and for dealing with data sparseness (Schneider et al., 2022). For this sub-question, we combine the parametric and structural uncertainty, meaning that the model is calibrated on both the parameters and the structure. We design scenarios that incorporate all three chosen dimensions of data sparseness, reflecting the type of sparseness we would expect in real-world data of supply chains. The outcome of the QD algorithm is a diverse set of plausible configurations of the parameters and the structure of a supply chain simulation model.

1.4. OUTLINE OF THESIS

The dissertation is structured as shown in Figure 1.3. Chapter 2, answering sub-question 1, proposes a classification for data sparseness and assesses its impact on supply chain visibility. Using this classification, Chapter 3 focuses on parametric uncertainty and presents the quality-of-fit of various model calibration techniques on parameters when systematically increasing the degree of data sparseness. This chapter answers sub-question 2. Chapter 4 answers sub-question 3 by showing the quality-of-fit of various model calibration techniques for identifying the underlying structure when varying the degree of data sparseness. For generating more than one optimal solution, Chapter 5 explores the use of the quality-diversity algorithm for calibrating simulation models in case of sparse data, answering sub-question 4. Chapter 6 presents an overarching discussion of this research. Chapter 7 concludes the research by answering the research questions and presenting recommendations for further research.

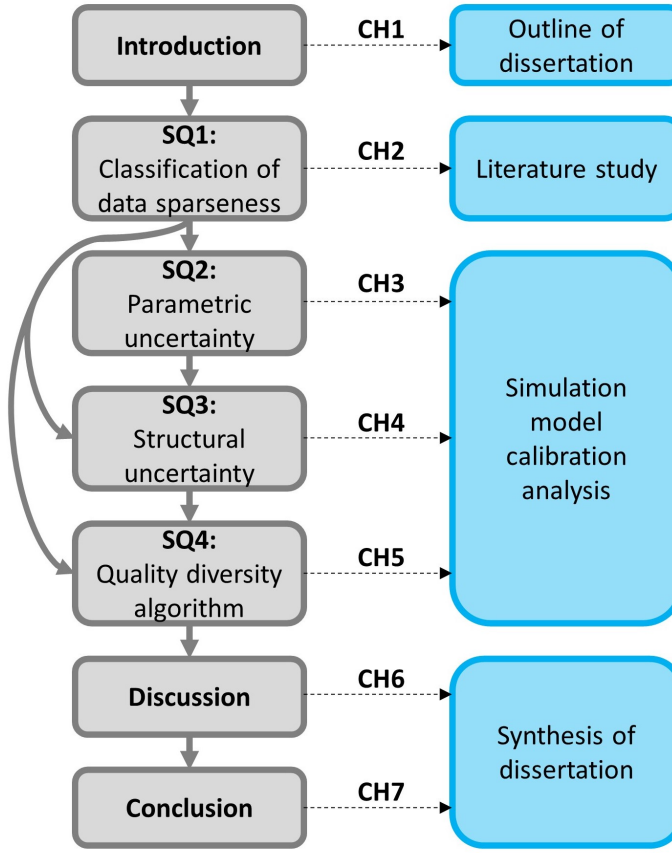


Figure 1.3: Outline of the dissertation.

ROTTERDAM

2

DIMENSIONS OF DATA SPARSENESS AND THEIR EFFECT ON SUPPLY CHAIN VISIBILITY

Supply chain visibility concerns the ability to track parts, components, or products in transit from supplier to customer. The data that organizations can obtain to establish or improve supply chain visibility is often sparse. This chapter presents a classification of the dimensions of data sparseness and quantitatively explores the impact of these dimensions on supply chain visibility. Based on a review of supply chain visibility and data quality literature, this study proposes to characterize data sparseness as a lack of data quality across the entire supply chain, where data sparseness can be classified into three dimensions: noise, bias, and missing values. The quantitative analysis relies on a stylized simulation model of a moderately complex illicit supply chain. Scenarios are used to evaluate the combined effect of the individual dimensions from actors with different perspectives in the supply chain, either supply or demand-oriented. Results show that when a data sparseness of 90% is applied, supply chain visibility reduces to 52% for noise, to 65% for bias, and to 32% for missing values. The scenarios also show that companies with a supply-oriented view typically have a higher supply chain visibility than those with a demand-oriented view. The classification and assessment offer valuable insights for improving data quality and for enhancing supply chain visibility.

This chapter has been published as: van Schilt, I. M., Kwakkel, J. H., Mense, J. P., & Verbraeck, A. (2024) Dimensions of data sparseness and their effect on supply chain visibility. *Computers & Industrial Engineering, 191*, pp. 110108. https://doi.org/10.1016/j.cie.2024.110108.
The code is available at https://github.com/imvs95/scv_sparse_data.

2.1. INTRODUCTION

The COVID-19 pandemic caused a steep rise in the worldwide demand for Personal Protective Equipment (PPE) such as face masks, gloves, goggles, and glasses (Omar et al., 2022). To enable proper planning for purchasing and producing PPE in such a high-demand situation, a good insight into the overall supply chain is required. There is a range of PPE products available, which can generally be classified as medical PPE or non-medical PPE. Medical PPE is certified and has a higher price and profit margin. This made it attractive for fraudulent organizations to enter the market, and sell non-medical PPE as medical (Ippolito et al., 2020). Hashemi et al. (2022) found that during the initial stages of the COVID-19 pandemic, the majority of PPE manufacturers producing fraudulent products emerged in Asia. However, counterfeit PPE activities and related logistics operations remained largely invisible due to little historical data on COVID-19 and on organizations selling fraudulent products trying to obfuscate their data (van Schilt et al., 2023). This counterfeit PPE case exemplifies a scenario where supply chain visibility is of the utmost importance, but it is hampered by sparse data (Zhao et al., 2023).

Supply chain visibility focuses on the ability to track parts, components, or products in transit from supplier to customer, addressing the actors' capability to monitor and trace the movement of goods with accurate and timely information (Saqib et al., 2019; Kalaiarasan et al., 2022). When supply chain visibility increases, logistical processes within the supply chain can be more effectively aligned (Srinivasan & Swink, 2018; Kalaiarasan et al., 2022). For example, hospitals can more effectively prepare for stock-outs of medical PPE, or align with trustworthy organizations from whom they can buy legitimate medical PPE.

Even in this digital era, many supply chain organizations still face challenges in processing and retrieving visibility data. (Wang et al., 2016; Tiwari et al., 2018; Wang & Zhuo, 2020). Additionally, the data required to improve supply chain visibility, such as data on demand, inventory levels, processing times of a manufacturer, and transportation times, is often sparse (Somapa et al., 2018; Kuipers, 2021). One of the causes for data sparseness is reluctance among actors within a supply chain to share (high-quality) data for various reasons such as competition and high costs (Boone et al., 2019), or because of illegal behavior of supply chain partners engaged in fraud (Ficara et al., 2021). Other potential problems in data collection and sharing are malfunctioning sensors leading to biased values or missing data points, inconsistency in data formats between different systems, or simply typos (Oliveira & Handfield, 2019).

Gaining more insight into the effect of data sparseness on supply chain visibility is essential for making improvements. A first step is to define data sparseness for supply chains. Unfortunately, there is no clear and agreed definition of data sparseness in the context of supply chain management. Various data quality issues can be seen as data sparseness, such as noise, bias, missing values, out-of-date information, different representations of the same data, or data that is not relevant for its use (Laranjeiro et al., 2015; van Schilt et al., 2023). Laranjeiro et al. (2015) presents a large variety of poor data instances and how they impact data quality. Although a large variety of poor data instances is presented in the literature, a clear and concise formalization of data sparseness is still lacking, especially in the field of supply

chain management. Oliveira and Handfield (2019) found that information quality plays a key role in supply chain visibility, and poor data resulting from data errors impacts decision-making. For example, when supply chain partners act on incomplete, inaccurate, and outdated data, this can lead to forecasting errors and supply chain disruptions (Agrawal et al., 2022). Certain errors may have a more significant impact on supply chain visibility than others. For example, missing data values can result in a complete lack of knowledge of the supply chain, while noisy observations provide some indication of the value's magnitude in the supply chain (Laranjeiro et al., 2015). The exact effect of these different types of data errors (i.e., data sparseness) on supply chain visibility is still poorly understood.

This chapter, therefore, focuses on the conceptualization of data sparseness in the context of supply chain management and the impact of data sparseness on supply chain visibility. Based on a review of supply chain visibility and data quality literature, a 3-dimensional classification of data sparseness is derived. Next, the effects of these dimensions of data sparseness on supply chain visibility are quantitatively assessed through a case study. A simulation model of a stylized supply chain of counterfeit PPE is used as ground truth. Complete data is extracted from this model, and then this data is systematically modified to increase sparseness along each of the three dimensions. Next, we assess how supply chain visibility changes. To evaluate the role of interaction effects between the three dimensions, we use scenarios to investigate the combined effect of the three dimensions of data sparseness. These scenarios describe data sparseness situations that could occur in real-life supply chains from the perspective of different actors, such as those positioned at the beginning of the supply chain (supply-oriented) or at the end (demand-oriented).

The contribution of this research is two-fold: (i) to provide a classification of data sparseness, and (ii) to assess its impact on supply chain visibility. By explicitly including data sparseness, our study is novel compared to the most recent systematic literature reviews on supply chain visibility of Kalaiarasan et al. (2022) and Somapa et al. (2018). Although both studies discuss data quality, they do not specifically focus on the dimensions of data sparseness and their impact on supply chain visibility. As for managerial implications, it is important for companies in a supply chain to be aware of the different dimensions of data sparseness and the differences in their impact on supply chain visibility. This might help companies to prioritize how to improve their data and, thereby, their visibility. Supply chain visibility is key for making the supply chain operations more efficient (Srinivasan & Swink, 2018; Sodhi & Tang, 2019).

The chapter is structured as follows. Section 2.2 presents the method for performing the literature review. Section 2.3 discusses the current state-of-the-art for supply chain visibility. Section 2.4 reviews the literature on data quality. Section 2.5 combines these two bodies of literature and presents a classification of data sparseness. Section 2.6 formalizes data sparseness and supply chain visibility, explains the design of the simulation experiment, and introduces the case study. Section 2.7 presents the effects of an increasing degree of sparseness for each of the identified dimensions of data sparseness on supply chain visibility, and evaluates the effect of data sparseness on supply chain visibility for plausible real-life scenarios. Section 2.8 discusses the results. Section 2.9 concludes this study and provides directions for further research.

2.2. LITERATURE REVIEW METHOD

A literature review was conducted for papers in the fields of supply chain visibility and data quality. For a comprehensive overview, the authors have executed a systematic search for relevant literature following the method described by van Wee and Banister (2016). Database engines such as Scopus and Google Scholar were used to identify the relevant literature. The scope of this literature review was restricted to academic papers and books in English. Papers were selected based on the number of citations, while taking into account how recently the papers were published, to not miss recent contributions. Papers had to meet a minimum threshold of 50 citations, subject to their year of publication and relevance to the topic, with the exception of a few papers that provided a key insight into the literature but had fewer citations. The specified publication date range is from 2000 to 2023, permitting a few exceptions for older literature that is still heavily cited. This research examines two bodies of literature: supply chain visibility, and data quality. In searching for the papers, we explicitly looked for different viewpoints and approaches for supply chain visibility and data quality over the years.

For supply chain visibility, the first step was to search for current state-of-the-art literature defining supply chain visibility using the search keywords: “supply chain visibility”, “supply chain transparency”, “supply chain visibility definition”, and “supply chain management and visibility”. The date range for filtering the literature is from 2000 to 2023. Papers were selected based on the number of citations. In the second step, additional papers were found using snowballing. In the third step, the search was focused on the literature for measuring supply chain visibility with the date range of 2000 to 2023 using the search keywords: “measure”, “calculate supply chain visibility”, “assessing supply chain visibility”, and “operationalize”. We limited the papers to those that include the calculation of supply chain visibility, and rejected papers that only mention the characterization of supply chain visibility. For all papers, the title, keywords, introduction, conclusion, and approach section were scanned. Papers were selected based on the number of citations, taking into account the publication date of the paper. In the fourth step, related papers were searched using snowballing.

For data quality, the first step was to search for current state-of-the-art literature on data quality with a date range from 2000 to 2023 using the specific search keywords: “data quality”, “data quality dimensions”, “characterize data quality”, and “sparse data quality”. In the second step, related papers were searched using snowballing. Some papers from before the year 2000 were also included in the literature search as there is a relatively older body of literature about data quality. In the third step, the search was targeted toward literature on data quality issues from year 2000 onwards using the keywords: “data issues”, “degraded data”, “data completeness”, and “poor data”. More in-depth papers on the definition of data quality issues were searched in the fourth step using snowballing and the specific keywords “measuring data quality issues” and “calculate noise/bias/missing values”. In addition to recent research from the years 2000 to 2023, literature from before 2000 has also been included as a basis of reference.

Through this systematic literature review of the two bodies of literature, the overlap between the topics was examined, facilitating the identification and classification of dimensions of data sparseness in relation to supply chain visibility. The evaluation of the obtained publications involved assessing their quality and comprehensiveness through the application of a quality filter at the beginning of the search and during snowballing. The quality filter checked the relevance of the literature based on the publication's keywords, title, and abstract, as well as the impact factor of the journal of publication. The filter has been applied for the initial list of literature and for the literature resulting from snowballing. To obtain extra feedback, the results were presented and discussed by the researchers during a conference in the field of transport and logistics.

2.3. SUPPLY CHAIN VISIBILITY

In recent years, supply chain visibility has become key for improving supply chain management and design (Busse et al., 2017; Roy, 2021). Successful supply chain management is heavily dependent on the availability of information shared by multiple actors within the supply chain (Brun et al., 2020). Research shows that to improve competitiveness by reducing costs, fulfilling demand, enhancing operational efficiency, or increasing customer service, it helps to have a more visible supply chain (Lavastre et al., 2014; Swift et al., 2019). Supply chain visibility creates a valuable opportunity to gain insights and exchange knowledge with other stakeholders in the network, which in turn is beneficial for designing an efficient supply chain system (Wei & Wang, 2010; Somapa et al., 2018). Moreover, it facilitates action and reduces (decision) risk, making the supply chain more resilient (Saqib et al., 2019; Rogerson & Parry, 2020).

The outbreak of COVID-19 showed the vulnerabilities of supply chains with low visibility, leading to a vast array of distribution issues and shortages (Junaid et al., 2023; Zhao et al., 2023). Both a lack of upstream visibility to the suppliers and downstream visibility to the customers existed (Busse et al., 2017; Kalaiarasan et al., 2022).

Our literature overview focuses on the definition of supply chain visibility, and the methods for assessing and measuring it. The current state-of-the-art papers on these topics are used for operationalizing supply chain visibility for this research.

2.3.1. DEFINITION

Supply chain visibility is a commonly and broadly used term in supply chain and logistics with a variety of meanings. Francis (2008, p. 182) proposes a general definition based on a literature review: *"Supply chain visibility is the identity, location and status of entities transiting the supply chain, captured in timely messages about events, along with the planned and actual dates/times for these events."* Similar to Saqib et al. (2019), this definition assumes that a detailed picture of the entities, i.e., any object moving through the supply chain, is needed. Providing complete information about all objects in the supply chain presents a challenge for the stakeholders in the supply chain, who might need to provide confidential and competitive information, and as a result, they are often reluctant to share such information (Pero & Rossi, 2014; Wang

& Zhuo, 2020). Second, not all stakeholders benefit from improved supply chain visibility: having too much information without a clear use case can be a distraction. Barratt and Oke (2007) includes the extent to which data is key or useful for supply chain visibility according to their definition. This definition is often referred to by other authors (Kalaiarasan et al., 2022). Concluding, a weakness in the general definition offered by Francis (2008) is the absence of the relevance of the information for the stakeholders. Schoenthaler (2003), McCrea (2005), and Barratt and Oke (2007) do include this relevance in their definitions of supply chain visibility.

Later, Williams et al. (2013) adds the quality of supply and demand information on accuracy, timeliness, completeness, and usability in their definition of supply chain visibility. Kalaiarasan et al. (2022, p. 4) takes this a step further by defining supply chain visibility as *"the extent to which actors within a supply chain have visual access to the timely and accurate demand and supply information that they consider to be key or useful to their operations and supply chains."*

Most literature indicates that supply chain visibility is dependent on good data, either stating usefulness or data quality dimensions. Some definitions require a detailed picture of the entire supply chain (Francis, 2008), while other definitions are more aggregated on either the supply or the demand side (Barratt & Oke, 2007; Williams et al., 2013; Kalaiarasan et al., 2022). Combining the major insights from the literature, supply chain visibility for this research is defined as:

Supply chain visibility refers to the ability of tracking parts, components or products in transit from supplier to customer through relevant data of stakeholders.

Next to the dependence on good quality data, supply chain visibility also depends on the willingness of organizations to share this data. Bartlett et al. (2007) uses transparency as a measure of visibility, and combines it with a degree of obscurity. Sodhi and Tang (2019) refers to supply chain visibility as the company's effort to gather information and data, and supply chain transparency as the company's willingness to share information with the public. Brun et al. (2020) notes that collaboration amongst supply chain partners and the level of trust should increase to achieve supply chain visibility. Since our study does not focus on the general public but on supply chain partners, transparency in the context of supply chain visibility is defined as the willingness to share relevant data with stakeholders.

2.3.2. METHODS FOR ASSESSING SUPPLY CHAIN VISIBILITY

Somapa et al. (2018) is the most recent literature review that discusses the characterization and the quantification of supply chain visibility in a network. They define three characteristics to capture supply chain visibility: (1) accessibility of information, (2) quality of information, and (3) usefulness of information. The first characteristic focuses on the capability of information and communication technology (ICT) systems to collect data, whereas the other two characteristics focus on the quality of information for obtaining the organization's goal. In recent years, a new generation of ICT systems has arisen to collect data for improving supply chain visibility. One of the most interesting recent concepts is the Internet-of-Things (IoT), consisting of Internet-embedded sensors and ICT components to provide data on supply

chain and logistics activities (Calatayud et al., 2019). IoT can make more data accessible such that the supply chain is more visible for all actors in real-time (Kumar et al., 2022). Another useful concept is the Radio-frequency identification transponder (RFID), an auto-identification system for detecting objects and elements while they move along the supply chain (Pero & Rossi, 2014). Kalaiarasan et al. (2023) notes that many studies show how these concepts can be used for improving supply chain visibility. The authors show the potential of this new generation of ICT systems, including IoT, RFID and blockchain, for collecting data in real-time from all stakeholders. One of the key aspects for these systems is the collaboration between actors within the supply chain (Pero & Rossi, 2014; Kalaiarasan et al., 2023). As mentioned before, competition can limit the necessary collaboration between actors. In our example case of counterfeit PPE supply chains, the necessary collaboration and sharing of data are out of the question, since data can give away information about illegal activities. Collaboration between supply chain partners is therefore not always a given.

Somapa et al. (2018) gives an overview of quantitative and qualitative approaches for measuring supply chain visibility. Common quantitative methods are regression analysis, visibility scorecards, utilization ratios, and mathematical models rooted in, e.g., set theory. Only a few methods consider the global supply chain level instead of the firm level (Somapa et al., 2018). One of these methods is presented by Zhang et al. (2011) who measure supply chain inventory visibility by using set theory. They define visibility as the capability to access and provide information among several companies. Lee and Rim (2016) uses the Six Sigma method to evaluate the end-to-end supply chain visibility with a focus on operational capabilities. In contrast to studies that focus on the information perspective of visibility, they focus on the visibility of processes to asses whether the supply chain has the capability to execute the supply chain plan (Somapa et al., 2018). Lee and Rim (2016) calculate the mean and standard deviation of individual processes for lead time, yield, quality, and utilization.

Another method to determine supply chain visibility that includes the end-to-end supply chain is the calculation of geometric means of information quantity and quality shared between the other actors and the focal company, as designed by Caridi et al. (2010, 2013). A strength of this paper is that the authors focus on measuring supply chain visibility in complex networks, which is particularly challenging. In contrast, most of the literature focuses on relatively simple two-tier or linear supply chains. Caridi et al. (2010, 2013) is a notable exception by giving a quantitative approach to assess the degree of supply chain visibility in complex systems for inbound and outbound logistics. They distinguish four types of information flows for supply chain visibility: (1) transactions/events, (2) status information, (3) master data, and (4) operational plans. They measure visibility as the amount and the quality of information the focal company possesses, compared to the total information that could be obtained. First, the visibility that the focal company has of each individual actor in the supply chain is measured by supply chain managers who judge the quality and the quantity of information available for providing visibility. These judgments are collected for each type of information flow and for each supply chain actor on a relative scale from 1 (lowest) to 4 (best). An argument against this technique is that it is subjective. After obtaining the judgments, the individual visibility measures are com-

bined to calculate the global visibility. The global measure is the weighted average of visibility for each actor. The weight for each actor is based on how much the focal company purchases from an actor, and how much an actor buys from the focal company, and the distance between the companies in terms of the number of tiers and vertical integration. So, the more an actor sells to or buys from the focal company or the closer it is to the focal company in the supply chain, the higher the weight.

2.3.3. OPERATIONALIZATION

Combining the insights of Caridi et al. (2010, 2013), Somapa et al. (2018), Calatayud et al. (2019) and Kalaiarasan et al. (2022), this research measures supply chain visibility as *the weighted average of the available information quantity and quality divided by their theoretical maximum for all actors in the supply chain given the goods, information, and financial flows.* The characteristics for measurement can be captured by the quality and the quantity of information (Caridi et al., 2010, 2013; Somapa et al., 2018). This means that accessibility of information (e.g., the capability of IT systems, IoT, RFID) will be out of scope. Three types of flows can be distinguished for measuring supply chain visibility: the goods flow, the information flow, and the financial flow (Min & Zhou, 2002; Stadtler & Kilger, 2002). For each of these flows, data can be extracted to assess visibility. This study primarily focuses on the goods flow.

To measure supply chain visibility, the available quantity and quality of the information are compared to their theoretical maximum (Caridi et al., 2010). Instead of using expert judgments, quantitative measures are used to calculate the quantity and the quality. Quantity is measured as the percentage of the number of data points that are available to the actor in comparison to the full data set. Quality is measured as the mean absolute percentage error of the data set of the actor compared to the full data set. Along the lines of Caridi et al. (2010), these percentages are combined into a geometric mean to determine the supply chain visibility of an individual actor.

Similar to Caridi et al. (2010), supply chain visibility is first measured for each actor, but without the presence of a focal company. Next, the supply chain visibility scores of individual actors are aggregated into a global measure using a weighted average. The weight of an actor is determined by the number of orders and the costs they represent. The weight is assigned to each corresponding actor to determine the weighted visibility of the actor. The sum of the visibility scores of all actors in the supply chain results in a percentage value for the global supply chain visibility.

2.4. DATA QUALITY

Data quality is a topic that has been researched for many years and in various disciplines (Ehrlinger & Wöß, 2022). Data quality management involves data collection (data profiling), the characterization of data quality, the measurement of data quality, and data quality monitoring (Bronselaer, 2021). Our research focuses on sparse data with a low volume, whereas big data literature focuses on high volumes of data (see e.g., Günther et al. (2017), Jeble et al. (2018)). Therefore, literature on big data, e.g., the 5 V's for the quality of data: Volume, Variety, Velocity, Veracity, and Value (Wamba et al., 2015) is kept out of scope.

2.4.1. DATA QUALITY DIMENSIONS

Several papers have provided a categorization of data quality. Wang and Strong (1996) defines data quality as "the fitness of use" and presents a framework of data quality aspects that are important to data customers. They identify four main categories: (1) accuracy, (2) relevance, (3) representation, and (4) accessibility. Pipino et al. (2002) defines a detailed list of sixteen data quality dimensions based on a survey of healthcare, finance, and consumer product companies. Most of them fall into the categories of Wang and Strong (1996). One new dimension has been added: ease of modification, i.e., the level to which data is easy to modify. Fan and Geerts (2012) states five central issues for data quality: (1) data consistency, i.e., the validity of data to the real-world, (2) data deduplication, i.e., multiple points referring to the same real-world entity (3) data accuracy, i.e, closeness of a value to its true value, (4) information completeness, i.e., complete data to answer the question, and (5) data currency, i.e., timeliness.

Huang (2013) aggregates data quality into three main categories: (1) syntactic quality, the level to which data follows the rules of a data model, with subcriteria including accuracy and consistency, (2) semantic quality, the level to which data is relevant and required for the purpose, with subcriteria including accuracy, completeness, and mapping consistency, and (3) pragmatic quality, the level to which data is suitable for a given application, with subcriteria including completeness, timelineness, and presentation suitability. Hazen et al. (2014) defines four dimensions of data quality in the context of supply chain management: (1) accuracy, the degree to which data has errors, i.e., the degree to which it is similar to the "real" value, (2) timeliness, the degree to which data is up-to-date, (3) consistency, the degree to which similar data is presented in the same format, and (4) completeness, the degree to which necessary data is available. These dimensions are similar to the subcriteria of Huang (2013) and the taxonomy presented in Gao et al. (2016). In a comparison study on data quality frameworks, Cichy and Rass (2019) shows that accessibility, accuracy, completeness, consistency, and timeliness have the highest number of occurrences as data quality criteria. Although there is an ongoing discussion on the dimensions of data quality in literature, the criteria identified by the above authors (accuracy, timeliness, consistency, completeness) are the most frequently used ones to describe data quality (Ehrlinger & Wöß, 2022).

2.4.2. DATA QUALITY ISSUES

Data quality issues such as data sparseness or errors in the data for one or more of the dimensions of data quality lead to poor decision quality (Heinrich et al., 2018; Bronselaer, 2021). Accuracy decreases when data deviates from the "real" value; timeliness decreases when the data is outdated; consistency decreases when different data points are not presented in the same format; completeness decreases when there is missing data (Souibgui et al., 2019). Additionally, in the case of (partly) illicit supply chains, data can be manipulated or masked to avoid detection (van Schilt et al., 2023). In terms of the data quality dimensions, this study identifies and addresses three main data quality issues that are relevant for decision-making: noise, bias, and missing values (Oliveira et al., 2005; Janssen et al., 2017).

Noise in data results in corrupted or distorted data, potentially rendering it meaningless (Sáez et al., 2014). Noise in a data point is generally defined as a deviation of that particular data point, where the distribution of the deviation has a mean and a noise width (Gaussian noise) (Teng, 1999; Zhu & Wu, 2004; Zhu et al., 2004). So, a data point with noise results in the original value plus or minus a deviation. There is a difference between noise that is inherent (natural), and injected (artificial). When analyzing noise, it is important to take this distinction into account (Seiffert et al., 2014).

Bias in data means that the data is not representative of the population or the phenomenon of study (Tripepi et al., 2010). Bias means that some members are more likely to be included than others, thus the probability of a member being included is unequally distributed. Data produced by humans may contain bias as a result of human preferences or human observational capabilities. The most common types of bias are (i) selection bias, i.e., group representation, (ii) reporting bias, i.e., some observations are more likely to be reported than others, and (iii) detection bias, i.e., a phenomenon is more likely to be observed than others (Ntoutsi et al., 2020).

Missing values relate to the completeness of a data set. Peng et al. (2023) presents a review and notes that missing values are a widespread data quality problem. They categorize missing values into three categories, building on research by Rubin (1976). This first category addresses data that is missing completely at random, meaning the absence of a data value is based on a random sample of the complete data set. The second category of missing values is missing at random, meaning the absence of a data value is related to some properties of the observed data (the data set without the missing values) but not to the missing data. The third category is missing not at random. meaning the absence of a data value is systematically correlated to properties of the missing data itself (Fox, 2015). As an example of the second category, people with a higher age are more likely to withhold information on their income, meaning that the probability of missing data depends on the age (a property of the observed data). As an example of the third category, people with a higher income are more likely to withhold information on their income, meaning the probability of missing data depends on the income level itself (a property of the missing data).

2.5. CLASSIFICATION OF DATA SPARSENESS

In this section, the literature on supply chain visibility is combined with the literature on data quality for classifying data sparseness in the field of supply chain management. First, the overlap between the two bodies of literature is discussed. Next, the classification of data sparseness based on the literature review is presented.

Supply chain visibility is primarily determined by the quality and the quantity of the data (Caridi et al., 2010; Kalaiarasan et al., 2022). For quality, the data quality criteria of Huang (2013), Gao et al. (2016), and Ehrlinger and Wöß (2022) are used as these are the most frequently used ones to describe data quality. Quality and quantity of data are specified by the syntactic and semantic criteria, more specifically by their accuracy, consistency, and completeness. In the field of supply chain management, these data quality criteria are of relevance for enhancing supply chain visibility, and

for informed decision-making (Munir et al., 2020; Kalaiarasan et al., 2022). Accuracy ensures precise information on important supply chain variables such as inventory, order status, and lead times. This helps, for example, to make accurate demand forecasts to prevent excess inventory, which is important for the predictive real-time nature of the supply chain. Consistency ensures reliable data of supply chain operations that is shared between stakeholders. For example, a consistent data format between stakeholders helps to track the movement of goods. Considering the multi-sourced and geospatial characteristics of a supply chain, a consistent data set shared among the many stakeholders is of high importance to enhance supply chain visibility by enabling the tracking of the movement of goods and inventory levels. Completeness ensures comprehensive data of the supply chain operations. For example, this helps to anticipate demand and avoid stockouts, or to enable efficient planning when looking at the temporal characteristics of the supply chain.

For data *sparseness*, three main issues for data quality are distinguished: noise, bias, and missing values (Oliveira et al., 2005; Janssen et al., 2017; van Schilt et al., 2023). These issues are classified as the dimensions of data sparseness. The logical relationships between the dimensions of data sparseness and data quality criteria including their impact on supply chain visibility are illustrated as follows: Noise impacts the accuracy and the consistency of data quality. For example, in a case where the inventory of medical PPE is monitored manually, a typo in the data leads to noise. Inaccurate data on the inventory levels affects the accuracy and reliability of the supply chain visibility. Bias impacts the consistency and completeness of the data. For example, using the PPE case again, large hospitals could be overrepresented in the supply chain data, making small hospitals invisible. This would make the supply chain data skewed and incomplete as there is less information on small hospitals. As a result, fewer resources could be allocated to smaller hospitals, leading to stockouts. Missing values impact the completeness criteria. As an example in the PPE case, there could be no data on the lead times from the supplier to the hospitals, meaning that the hospitals have no visibility on how to manage their stock.

Other criteria, such as the pragmatic criteria of Huang (2013) and the timeliness of Ehrlinger and Wöß (2022), are not included in our classification. These criteria describe the relevance of the data, and indicate whether it is suitable and up-to-date for a given application. However, relevance is a very different kind of criterion than noise, bias, and missing values. Relevance concerns the applicability of the data set as a whole given a specific type of analysis or decision, whereas the other dimensions concern the modification of values within the data set for any analysis or decision purpose (Bronselaer, 2021). Especially considering the temporal and dynamic nature of a supply chain, the relevance of the data is subject to time-sensitive and up-to-date information. For example, accurate demand forecasting needs a relevant and up-to-date data set but still faces challenges when some data values with the data set are inaccurate, inconsistent, and incomplete.

Data in a supply chain can either be sparse by itself (i.e., unintentional sparseness) or sparse by manipulation (i.e., intentional sparseness) (Bartlett et al., 2007). Intentional manipulation of data is also a data quality issue (Janssen et al., 2017). The willingness of stakeholders to share this sparse data is the primary factor that deter-

mines transparency in the context of supply chain visibility (Wang & Zhuo, 2020; Baah et al., 2022). Combining (un)intentional sparseness and (non-)transparency leads to four cases, where stakeholders are either (1) willing to share unintentionally sparse data to *improve* supply chain management, (2) unwilling to share unintentionally sparse data to *hide* data, (3) willing to share intentionally sparse data to *mislead* other stakeholders, or (4) unwilling to share intentionally sparse data to *prevent* poor data availability. The fraction of intentional sparseness of the data has an impact on how to cope with data in supply chain management and how to use it in decision-making (Oliveira & Handfield, 2019; Bronselaer, 2021). For example, if a supplier intentionally withholds key production data about the fabric of medical PPE to gain a competitive advantage, the manufacturer may make sub-optimal decisions on inventory levels, leading to potential disruptions in the supply chain and increased costs.

This literature study led to the following definition of sparse data in relation to supply chain visibility:

Sparse data in supply chain management refers to the lack of data quality across the entire supply chain for the quality dimensions: noise, bias, and missing values, where a certain fraction of data sparseness is intentional.

Table 2.1 presents the classification of sparse data in the context of supply chain management. In summary, there are three dimensions of data sparseness: (1) **noise**, i.e., values in the data set are distorted; (2) **bias**, i.e., values in the data set are not representative of the population or the phenomenon of study; (3) **missing values**, i.e., values in the data set are missing. Each dimension has a certain fraction of intentional sparseness. Thus, each dimension of data sparseness consists of (i) the level of data quality, and (ii) the fraction of intentional sparseness.

	Description	**Level of data quality**	**Fraction of intentional sparseness**
Noise	Distortedness.	Value is modified by adding a deviation following a distribution in x% of original data elements.	Noise is for y% intentionally sparse in the data.
Bias	Representativeness.	Value is structurally more likely to be present in x% of the original data elements.	Bias is for y% intentionally sparse in the data.
Missing values	Completeness.	Value is missing in x% of the original data elements.	Missing values is for y% intentionally sparse in the data.

Table 2.1: Classification of data sparseness in three dimensions.

2.6. METHODS

In this research, the effect of the identified dimensions of data sparseness on supply chain visibility is assessed by systematically increasing the degree of sparseness in the data. First, the quantification of the dimensions of data sparseness is described. Second, the formalization of global supply chain visibility is discussed. Next, the design of experiments using a ground truth simulation is explained. Last, the case study used in this research for performing experiments is presented.

a_n	quantity percentage of data for node n
u_{tn}	bias value of time t and node n for the ground truth data
v_{tn}	value of time t and node n for the ground truth data
v'_{tn}	value of time t and node n for the sparse data
scv_n	supply chain visibility for node n
scv	global supply chain visibility
q_n	quality percentage of data for node n
w_n	weight for node n based on average inventory
A_n	set of values that are not *NaN* for each node n, $\forall\, t \in T$
N	set of nodes in the data of the supply chain model, $n \in N$, where each node represents an actor in the supply chain network
T	set of elements in the time domain in the ground truth data of the supply chain model, $t \in T$
T^*	set of elements in the time domain in the ground truth data indicating bias

Table 2.2: Table of Notation.

2.6.1. FORMALIZATION OF DIMENSIONS

Let $t \in T$ be an index t in the set of elements of the time domain T in the data. Let $n \in N$ be a node n (in this case, an actor) in the set of nodes N in the data. Let v_{tn} be a value of time t and node n for the ground truth data, and let v'_{tn} be a value of time t and node n for the sparse data. The degree of data sparseness is systematically increased on the three identified dimensions of data sparseness as follows:

Noise level of $x\%$ is defined as $x\%$ of original data elements are modified by adding a deviation following a distribution. This means that, over the entire data set, $x\%$ of the data has noise. It is randomly determined, using a discrete Uniform distribution, which elements of the data set have noise. The deviation of the noise follows a Gaussian distribution with a standard deviation of 1. A value with noise can be defined as:

$$v'_{tn} \sim v_{tn} + \mathcal{N}(\mu = 0, \sigma = 1) \tag{2.1}$$

Bias level of $x\%$ is defined as values that are structurally more likely to be present in $x\%$ of the original data elements. A sample of $x\%$ of the rows is randomly drawn to represent bias. Every row is allocated a weight through a log-normal distribution with $\mu = 0$ and $\sigma = 1$, and a sample is selected based on these weights. For example, there are 100 data rows with 25% bias. This means that on average 25 rows are sampled using the weights resulting from the log-normal distribution, and replace a randomly selected row from the ground truth data set. The higher the weight given the log-normal distribution, the more likely the row will be sampled and will be more often present in the data set. The other 75 rows remain the same as the ground truth data set. Let $T^* \subset T$ be the set of elements in the time domain indicating bias. Let u_{tn} be a historical value that is already present in the data set, and used to create bias:

$$u_{tn} \in \{v_{t'n'} : t' \in T, n' \in N\} \tag{2.2}$$

A value with bias can be defined as:

$$v'_{tn} = \begin{cases} v_{tn}, & \forall t \notin T^*, \\ u_{tn}, & \forall t \in T^* \end{cases} \tag{2.3}$$

Missing values level of $x\%$ means that a value is missing in $x\%$ of the original data elements. Similar to the noise level, it is randomly determined which $x\%$ of data points are missing over the entire data set by following a discrete Uniform distribution. A missing value can be defined by a non-value (*NaN*, indicating *Not a Number*) as follows:

$$v'_{tn} = NaN \tag{2.4}$$

Important to note is that a non-value differs from a zero value. In the context of supply chain management, many true values can be zero, such as zero inventory of a product, so a missing value is encoded as *NaN* rather than as zero (Heinrich et al., 2018).

2.6.2. FORMALISATION OF SUPPLY CHAIN VISIBILITY

Supply chain visibility is measured by comparing the available quantity and quality of the information to its theoretical maximum, as described in Section 2.3.3. The calculation of supply chain visibility in our research is as follows: first, the quantity and the quality of the information at each node are measured. For each node $n \in N$, the quality as a percentage is defined as follows:

$$q_n = 100 - MAPE(v_n, v'_n) \tag{2.5}$$

where MAPE is the mean absolute percentage error relative to the average of the data elements of the node. Hereby, the magnitude of the mean absolute percentage error is taken into account. MAPE is defined as,

$$MAPE(v_n, v'_n) = \frac{100}{\#T} \sum_{t \in T} \left| \frac{v_{tn} - v'_{tn}}{\overline{v_n}} \right| \tag{2.6}$$

For each node $n \in N$, the quantity as a percentage is defined as follows:

$$a_n = 100 \times \frac{\#A_n}{\#T}, A_n = \{v'_{tn} \neq NaN, \forall t \in T\} \tag{2.7}$$

The supply chain visibility for each $n \in N$ is calculated as:

$$scv_n = \sqrt{q_n \times a_n} \tag{2.8}$$

Second, the weight of each node in the supply chain is determined. The weight is based on the average number of orders w_n of each node n. The average number of orders is normalized over all nodes. This gives,

$$w_n = \frac{\overline{v_n}}{\sum_{n \in N} \overline{v_n}} \tag{2.9}$$

The global supply chain visibility as a percentage can be calculated as follows:

$$scv = \sum_{n \in N} w_n \times scv_n \tag{2.10}$$

2.6.3. DESIGN OF EXPERIMENTS

This research uses a ground truth simulation model to evaluate and compare the supply chain visibility for varying degrees of data sparseness in each of the dimensions. The simulation model calculates the ground truth values to obtain the theoretical maximum quality and quantity of information. This set-up allows for correctly assessing how supply chain visibility changes as the true maximum is known which is often not the case in real life (Khondoker et al., 2016).

Figure 2.1 presents the method used for calculating supply chain visibility for various degrees of data sparseness using the ground truth. First, the ground truth data for each time element $t \in T$ and each node $n \in N$, v_{tn}, is extracted from the simulation model. This ground truth data does not include any sparseness. Then, a certain percentage of data sparseness is added: noise, bias, and missing values. Next, the ground truth data values (v_{tn}) and the sparse data values (v'_{tn}) are used to calculate the supply chain visibility. First, the supply chain visibility is calculated for each node, $n \in N$, using the ground truth data and the sparse data. The quantity and the quality of the sparse data is compared to the ground truth. Next, the weights of each node are determined based on the average number of orders. Then, these measures are combined to a global supply chain visibility (indicated by *SCV* in Figure 2.1) as a percent value.

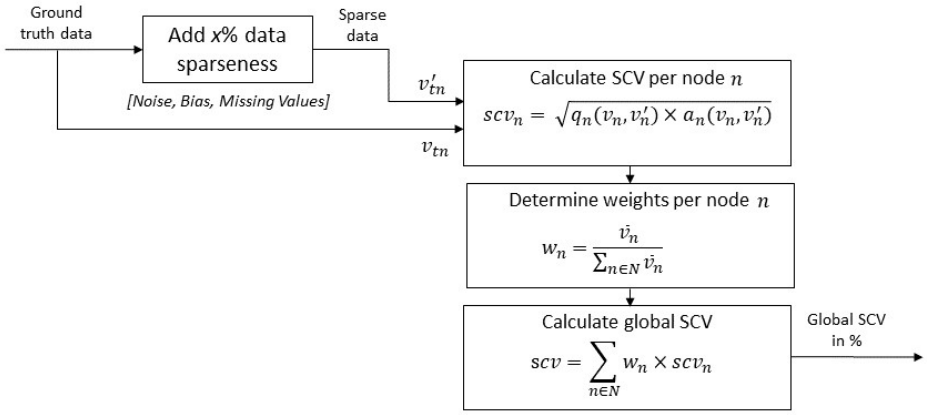


Figure 2.1: Method for calculating global Supply Chain Visibility (SCV).

Two experiments are performed in this study: (1) systematically increase the degree of data sparseness for each individual dimension, and (2) design and evaluate plausible real-life scenarios with regard to data sparseness. First, the degree of data sparseness is systematically increased by 10% for each dimension. More specifically, the experiments are 0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90%. For the

ground truth data, i.e., the base case, there is no data sparseness in any dimension so the supply chain visibility is always 100%.

Second, the effect of the three dimensions of data sparseness on stylized scenarios that are theoretically plausible in real-life supply chains is evaluated. In these stylized scenarios, all three dimensions of data sparseness are used as most real-world data sets include all these dimensions of data sparseness. The dimensions are added in the following order to the data set: (1) add bias, so bias only exists for true values of the data, (2) add noise to this biased data set, (3) delete values to create missing values. The configuration of these stylized scenarios is presented in Section 2.7.2.

Each experiment is performed with 200 unique seeds to account for the effect of stochasticity on supply chain visibility. By transforming the ground truth data using the same seeds for each experiment, it is ensured that the exact same observations are modified for each dimension of data sparseness.

2.6.4. CASE STUDY

In this research, a stylized counterfeit PPE supply chain is used as a case study for performing experiments. This supply chain is characterized by sparse data since (1) the production of counterfeit PPE during COVID-19 presented a new and unexpected phenomenon with little historical data, and (2) counterfeit PPE supply chains are operated by organizations selling fraudulent products that obfuscate as much data as possible (Hashemi et al., 2023).

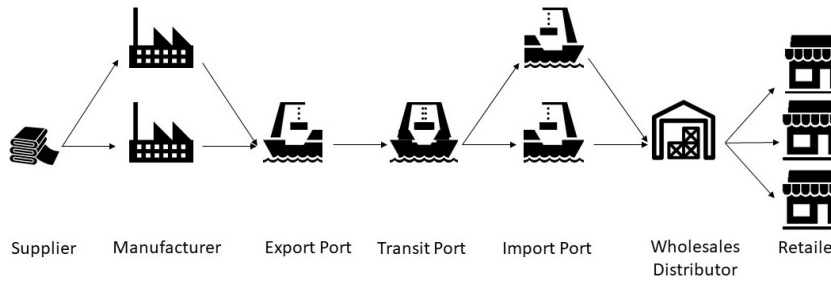


Figure 2.2: Stylized Supply Chain of Counterfeit PPE.

Figure 2.2 visualizes the stylized counterfeit PPE supply chain simulation model. The symbols in the figure represent the main actors in the supply chain, and the arrows represent the transportation flows. The supply chain starts with the raw materials supplier, placed in this stylized case in Vietnam, who supplies products for PPE such as fabrics. Next to China (source of the majority of the medical counterfeits) and India, many PPE come from Vietnam including the general productive mask production (Nikkei Asia, 2020). These products are transported over land to one of the two manufacturers in the same country, Vietnam. These manufacturers produce protective masks (mislabeling them as medical) in the factory and pack them in batches for transport. Each batch has a certain quantity of counterfeit PPE. For example, a batch consists of 2000 boxes of 200 PPE which equals a quantity of 200,000 PPE in total. Next, a batch of finished counterfeit PPE is transported from the manufacturers' lo-

cation via a truck to the export port in Hai Phong, Vietnam. The batch is loaded into a 40 ft container and transported by a small container ship to the transit port, Tanjung Pelepas, Malaysia. The small container ship unloads the container with counterfeit PPE at the transit port. At the same port, the container is loaded onto a larger container ship for overseas transport. The destination of this ship, also the import port, is either the Port of Rotterdam, The Netherlands, or the Port of Antwerp, Belgium. The container is unloaded at one of these ports and waits for inland transport to the (illegal) wholesales distributor in Eindhoven, The Netherlands. This means when the container arrives in Antwerp, the truck crosses a land border to arrive at the wholesales distributor. At the wholesales distributor, the batch of counterfeit PPE in the container is equally divided into three smaller batches for the three retailers. These smaller batches are transported by small trucks to the retailers in Amsterdam, Utrecht, and Venlo in The Netherlands. When the counterfeit PPE arrives at the retailer, the products are being sold with or without knowing that they are counterfeit.

A discrete event simulation model of this stylized configuration of a counterfeit PPE supply chain from Vietnam to stores in the Netherlands is used to gather the ground truth data. Table 2.3 shows the input parameters for the actors and the links used in the stylized simulation model.

In the simulation model, most uncertainties such as delays of transport modalities and speed of transport modalities follow triangular distributions inspired by real-world data of a fashion retailer (Kuipers, 2021). Table 2.4 shows the input parameters and the distributions of the speed and the delays of the transport modalities for the simulation model of this study. This case study represents a complex network suitable for our study due to the many uncertainties in the supply chain simulation model (e.g., delay in transport modalities, loading and unloading times). For example, the retailer's inventory can fluctuate very much, depending on whether a vessel has a 1-day delay or a 7-day delay.

Actors				**Links**		
Input Parameter	**Distribution**	**Value**	**Unit**	**Name**	**Value**	**Unit**
Interarrival time of product at supplier	Exponential	1.5	days	Supplier to manufacturer 1	50	km
Time at manufacturer	None	2.5	days	Supplier to manufacturer 2	45	km
Time at ports	Triangular	1, 2, 2	days	Manufacturer 1 to export port	125	km
Time at wholesales distributor	Triangular	0.5, 1, 2	days	Manufacturer 2 to export port	100	km
Time at retailers	Exponential	0.2	days	Export port to transit port	1656	nautical miles
				Transit port to import port Rotterdam	9286	nautical miles
				Transit port to import port Antwerp	9195	nautical miles
				Import port Rotterdam to wholesales distributor	135	km
				Import port Antwerp to wholesales distributor	100	km
				Wholesales distributor to retailer Amsterdam	125	km
				Wholesales distributor to retailer Utrecht	92	km
				Wholesales distributor to retailer Venlo	60	km

Table 2.3: Input parameters of actors and links for the simulation model of the stylized counterfeit PPE supply chain.

Transport modalities							
Input Parameter	**Distribution**	**Value**	**Unit**	**Input Parameter**	**Distribution**	**Value**	**Unit**
Speed of small truck	Triangular	0, 100, 120	km/h	Delay of small truck	Triangular	0, 0.2, 0.5	days
Speed of large truck	Triangular	0, 80, 120	km/h	Delay of large truck	Triangular	0, 0.5, 1	days
Speed of feeder	Triangular	10, 18, 25	knots	Delay of feeder	Triangular	0, 4, 16	days
Speed of vessel	Triangular	10, 18, 25	knots	Delay of vessel	Triangular	0, 7, 16	days

Table 2.4: Input parameters of speed and delay of the transport modalities for the simulation model of the stylized counterfeit PPE supply chain.

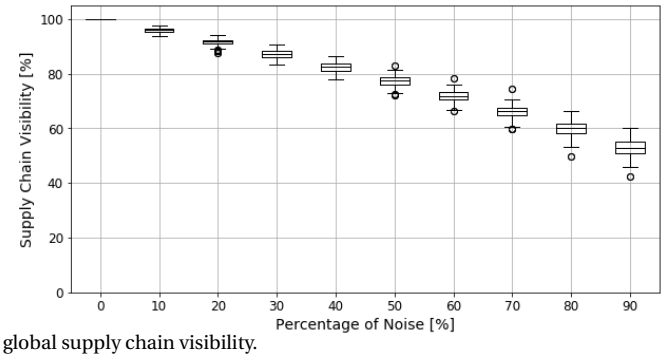
From this simulation model, time series data on the stylized supply chain is extracted as ground truth data. The time series data entails data on the inventory that is located at each actor (e.g., manufacturer, export port, import port) in the supply chain per day. Each data element in the time series data is thus the inventory of an actor at a specific time. The mean inventory value per day is calculated for the multiple replications of the simulation model. A simulation time of 52 weeks with 20 unique replications is used. The simulation model has been developed with the library *pydsol* in Python. This library is a Python implementation of the Distributed Simulation Object Library (DSOL), originally implemented in Java (Jacobs, 2005).

2.7. RESULTS

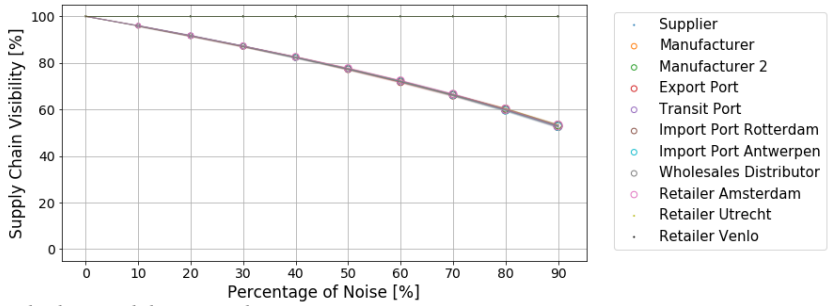
This section presents the results of variations in supply chain visibility given an increasing degree of sparseness for each of the identified dimensions of data sparseness using the case study. Next, the plausible scenarios that could theoretically occur in real life are described. These scenarios are evaluated for the impact of data sparseness on supply chain visibility.

2.7.1. EFFECT OF THE INDIVIDUAL DIMENSIONS

Figures 2.3, 2.4, and 2.5 show, for each individual dimension of data sparseness, a boxplot of global supply chain visibility for various degrees of data sparseness. The boxplot displays the minimum, the 1st quartile (i.e., 25th percentile), the median, the 3rd quartile (i.e., 75th percentile), and the maximum of the percentage of supply chain visibility for each degree of data sparseness. Also, the average supply chain visibility of each actor in the supply chain over various degrees of data sparseness including a 95% confidence interval is shown. The size of the markers in the plot is equal to the size of the 95% confidence interval.



(a) Boxplot of global supply chain visibility.

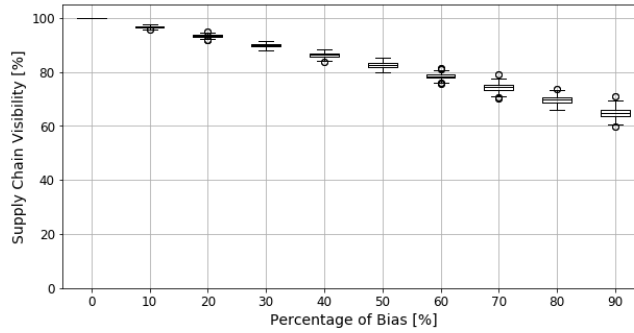


(b) Supply chain visibility per node.

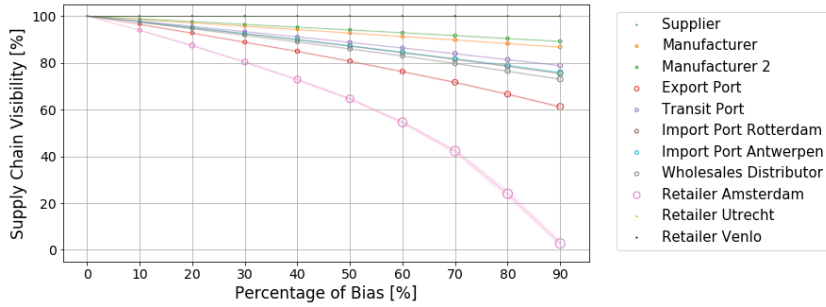
Figure 2.3: Results for dimension noise for 200 seeds for various degrees of data sparseness.

Figure 2.3a presents the boxplot of supply chain visibility when adding noise to the ground truth data. It shows that the global supply chain visibility gradually decreases when more noise is present in the data with average steps of 4% to 7% per 10% of extra noise. The highest median value of supply chain visibility, excluding the base case, is 95.9% at 10% noise. The lowest median value of supply chain visibility is 52.9% at 90% noise. The spread of the supply chain visibility over the 200 seeds becomes wider with a higher degree of noise, meaning that the interquartile distance (i.e., the distance between the 1st and 3rd quartiles) becomes wider. However, this distance stays limited to at most 4.3%. The distance between the minimum and the maximum value of supply chain visibility becomes even wider over the various degrees of noise with the largest distance of 14% at 90% data sparseness.

When looking more closely at which actors contribute to this spread, Figure 2.3b shows that most actors follow the same decreasing trend over the various degrees of noise regarding their supply chain visibility. Represented by the size of the marker in this figure, the retailer in Amsterdam has the widest confidence interval of 1.2% when increasing the degree of noise in the data. Other actors have a confidence interval between 0.8% to 1.0% at the highest degree of data sparseness (90%).



(a) Boxplot of global supply chain visibility.



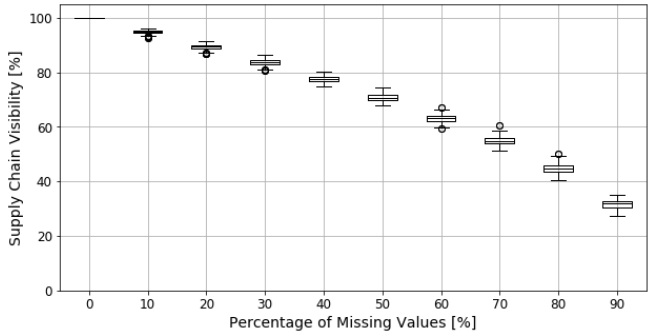
(b) Supply chain visibility per node.

Figure 2.4: Results for dimension bias for 200 seeds for various degrees of data sparseness.

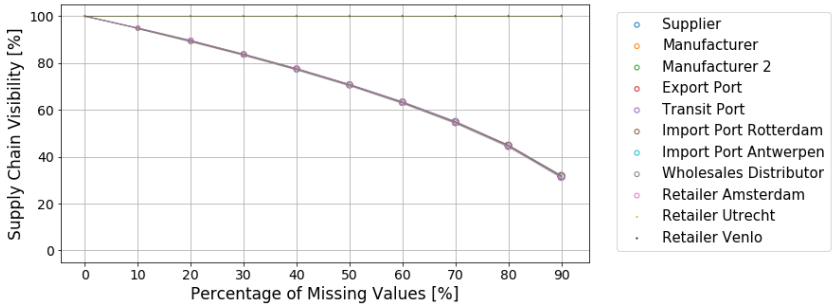
Figure 2.4a shows the boxplot of global supply chain visibility when adding bias to the ground truth data. The boxplot shows that supply chain visibility decreases when more bias is present in the data with on average steps of 3% to 5% per 10% of extra bias. The highest median value of supply chain visibility, excluding the base case, is 96.8% at 10% bias. The lowest median value of supply chain visibility is 64.9% at 90% bias. The spread of supply chain visibility becomes wider up to 50% bias with an interquartile distance from 0.5% to 1.4%, and the distance between the minimum and the maximum values from 1.8% to 5.5%. At 60% bias, the spread becomes smaller (4.5%) and afterwards, it increases by 2% for 70% data sparseness. After 70%, the spread becomes wider with the widest spread at 90% bias with an interquartile distance of 4.9% and a distance between the minimum and the maximum of 8.8%.

When looking at the supply chain visibility per actor including the 95% confidence interval in Figure 2.4b, it shows that the average supply chain visibility percentage of the actors retailer in Amsterdam converges to 2.7% at 90% bias. From 60% onwards, the average supply chain visibility of retailer in Amsterdam is decreasing steeply with steps of 10% to 20%, and with a confidence interval higher than 1.2%. This could explain why the spread of the global supply chain visibility is smaller at 60% bias, and becomes considerably wider afterwards. Also, for this actor, the average supply chain visibility percentage decreases relatively steeply compared to the other actors. The percentage of supply chain visibility of the manufacturers gradually

decreases with on average steps of 1% to 2% per 10% bias increase over the various degrees of bias as data sparseness. For the transit port, import ports, and wholesales distributor, supply chain visibility decreases with average steps of 2% to 3%. The percentage of supply chain visibility of the actor export port decreases slightly more steep with average steps of 4% to 5% when increasing bias in the data.



(a) Boxplot of global supply chain visibility.



(b) Supply chain visibility per node.

Figure 2.5: Results for dimension missing values for 200 seeds for various degrees of data sparseness.

Figure 2.5a presents the boxplot of global supply chain visibility when adding missing values to the ground truth data. It shows that the supply chain visibility decreases when more missing values are present in the data. The decrease starts with steps of 5% to 6% per 10% increase in missing values. From 50% missing values onwards, the median value of supply chain visibility decreases with 7% to 13% per 10% step. The highest median value of supply chain visibility, excluding the base case, is 94.8% at 10% missing values. The lowest median value of supply chain visibility is 31.9% at 90% missing values in the data. The spread of the supply chain visibility is relatively small but increases over the various degrees of missing values. The interquartile distance is 0.8% at 10% missing values and is gradually increasing to 2.4% at 90% missing values. The distance between the minimum and the maximum value of supply chain visibility is increasing from 2.8% to 8.6%.

When looking at the supply chain visibility for each actor in Figure 2.5b, it shows that most actors in the supply chain follow the same trend regarding the average per-

centage of supply chain visibility over the various degrees of missing values. The 95% confidence interval of all the actors, except for the retailer in Utrecht and the retailer in Venlo, becomes slightly wider when the percentage of data sparseness increases. However, this is still not more than 0.7%.

	Noise		Bias		Missing	
Percentage of Data Sparseness	*Mean*	*Std*	*Mean*	*Std*	*Mean*	*Std*
0%	100.0	0.0	100.0	0.0	100.0	0.0
10%	95.9	0.8	96.8	0.4	94.8	0.6
20%	91.7	1.2	93.4	0.5	89.5	0.9
30%	87.2	1.4	89.9	0.7	83.7	1.1
40%	82.4	1.6	86.3	0.8	77.5	1.2
50%	77.3	1.8	82.5	1.0	70.7	1.4
60%	71.8	2.0	78.5	1.0	63.2	1.5
70%	66.1	2.3	74.3	1.4	54.8	1.5
80%	59.9	2.7	69.8	1.5	44.9	1.7
90%	52.8	3.0	65.0	1.9	31.7	1.7

Table 2.5: Supply chain visibility (%) mean and standard deviation for each dimension of data sparseness and for various degrees of data sparseness.

Table 2.5 shows the mean and the standard deviation, i.e., the spread, of the supply chain visibility as a percentage for each dimension in more detail. It can be observed that the standard deviation of the supply chain visibility increases when more noise is added to the data. The increase of the standard deviation from 0.8% at 10% sparseness to 3.0% at 90% sparseness is the highest of all dimensions. The table also makes clear that missing values assert the most influence on supply chain visibility. For missing values, the average percentage of supply chain visibility decreases all the way down to 31.7% for 90% data sparseness. Missing values has the lowest standard deviation over most degrees of data sparseness compared to noise and bias.

2.7.2. SCENARIO ANALYSIS

To compare the effect of data sparseness on real-life supply chain cases, plausible scenarios that theoretically could occur in a supply chain for assessing supply chain visibility are developed. Table 2.6 presents the configuration of the percentages of noise and missing values of four stylized scenarios: (i) competitor, (ii) key actor, (iii) supply-oriented, and (iv) demand-oriented. For each scenario, a bias of 25% over the entire data set is added as real-life data often includes values that are structurally more present than others. For example, companies have structurally more information on their own inventory than on the inventory of other actors.

	Scenarios							
	Competitor		**Key Actor**		**Supply-Oriented**		**Demand-Oriented**	
	Noise	*Missing*	*Noise*	*Missing*	*Noise*	*Missing*	*Noise*	*Missing*
Supplier	10	25	10	25	10	25	80	95
Manufacturer 1	10	25	10	25	20	35	70	85
Manufacturer 2	95	95	10	25	20	35	70	85
Export port	10	25	10	25	30	45	50	65
Transit port	10	25	95	95	40	55	40	55
Import port 1 & 2	10	25	10	25	50	65	30	45
Wholesales	10	25	10	25	70	85	20	35
Retailer 1, 2 & 3	10	25	10	25	80	95	10	25

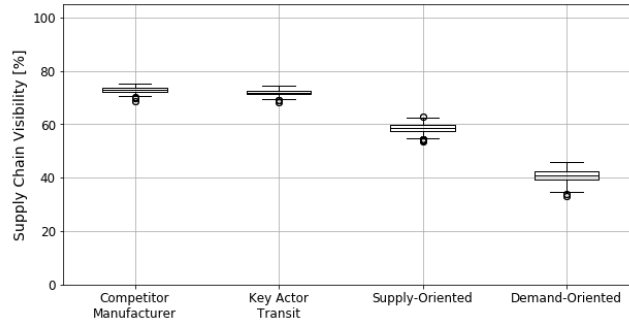
Table 2.6: Configuration of percentage of noise and missing value for each actor in % for four scenarios.

The first scenario, competitor, reconstructs the case where only one of two actors in a competitive position in a supply chain is willing to share data. A possible reason is that an actor is reluctant to share good data for competitive reasons. In our case, this is a manufacturer (referred to as manufacturer 2) with a noise of 95% and missing values of 95%. In real life, it is unlikely that the data of the other actors is perfect. To account for this, the other actors have a noise of 10% and 25% missing values.

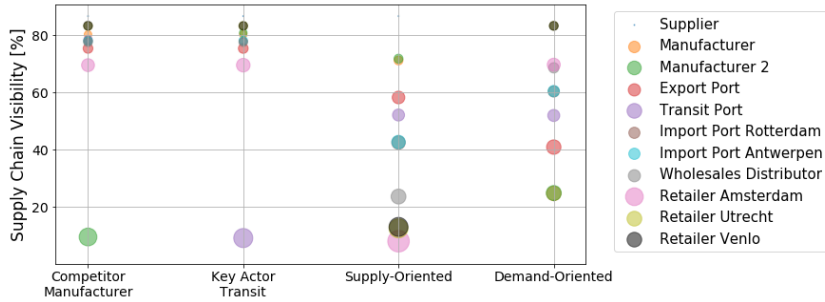
The second scenario, key actor, shows the case where an actor at a key position, i.e., in the middle of the supply chain, only provides sparse data to the rest of the supply chain with noise and missing values of 95%. Similar to the competitor scenario, the other actors have a noise of 10% and 25% missing values.

The third scenario, supply-oriented, represents the case where much is known on the supply side (starting with only 10% noise and 25% missing values for the supplier), and less is known on the demand side (ending with 80% noise and 95% missing values for the retailers). This often holds for suppliers as they have more high-quality information on actors upstream than downstream, represented by gradually degrading data over the actors in the supply chain.

The fourth scenario, demand-oriented, represents the case where much is known on the demand side (starting with only 10% noise and 25% missing values for the retailers), and less is known on the supply side (ending with 80% noise and 95% missing values for the supplier). The retailers have a higher quality and quantity of information on the actors close to them, i.e., downstream. Similar to the supply-oriented scenario, this is represented by a gradual increase in the percentage of noise and missing values following the sequential ordering of the upstream actors in the supply chain.



(a) Boxplot of global supply chain visibility.



(b) Supply chain visibility per node per scenarios. The size of the markers represents the 95% confidence interval of the supply chain visibility for each actor.

Figure 2.6: Results of the scenarios (1) Competitor, (2) Key Actor, (3) Supply-Oriented, (4) Demand-Oriented over 200 seeds.

Figure 2.6a shows the boxplot of the average global supply chain visibility percentage for each scenario. In the scenarios where only one actor provides sparse data, the competitor and the key actor, the global supply chain visibility is 72.9% and 71.9% respectively. The spread of these two scenarios over the 200 seeds is small as the interquartile distance for both scenarios is only 1.3%, and the distance between the minimum and the maximum values is at most 5.1%. Figure 2.6b presents the average supply chain visibility percentage per actor with the size of the marker representing the 95% confidence interval. From this, it can be observed that the decrease in global supply chain visibility is directly correlated with a low average supply chain visibility of the particular actor that has a high noise and a high number of missing values in each of the two scenarios. The average supply chain visibility of Manufacturer 2 and the transit port is around 9.2% with a 95% confidence interval of 0.6%. Other actors have an average supply chain visibility between 69.5% to 86.6%.

In the scenarios where noise and missing values are gradually added to the actors in the supply chain, either supply-oriented or demand-oriented, the global supply chain visibility is 58.5% and 40.6% respectively (see Figure 2.6a). For the supply-oriented scenario, the spread is small with an interquartile distance of 2.0% and a distance between the minimum and maximum values of 7.8%. For the demand-oriented scenario, the spread is wider with an interquartile distance of 3.0% and a distance between the minimum and the maximum of 11.1%.

When looking at the supply-oriented scenario, the average supply chain visibility per actor in Figure 2.6b decreases over the supply chain. Given the sequential ordering of the actors in the supply chain, the supplier and the manufacturers have the highest average supply chain visibility between 86.6% and 89.7%. The actors with the lowest average supply chain visibility in the supply-oriented scenario are the retailers; between 7.9% to 12.9% with a relatively wide confidence interval.

For the demand-oriented scenario, a similar pattern of sequentially decreasing average supply chain visibility over the actors in the supply chain is present but then reversed in comparison to the supply-oriented scenario. The actors with the highest average supply chain visibility are the retailers in Utrecht and Venlo with 83.5%, and the supply chain visibility of the retailer in Amsterdam is 69.6%. Actors with the lowest average supply chain visibility are the manufacturers with 24.7% and a relatively small confidence interval. Figure 2.6b also shows that the global supply chain visibility of this scenario is the lowest and it has the widest spread.

2.8. DISCUSSION

Six main elements are addressed that are essential for properly understanding and interpreting the results of this study: (1) impact of artifacts of the simulation model, (2) use of sampling method, (3) way of calculation of supply chain visibility, (4) specificity to a sequential supply chain, (5) lack of including intentionality, and (6) limitation on the incorporation of the data collection process.

First, the results show that in all three individual dimensions of data sparseness, the actors at the outer end of the supply chain, i.e., the supplier, the retailer in Utrecht, and the retailer in Venlo, have zero to little spread in their supply chain visibility or are not influenced (i.e., the visibility remains at 100%) when adding data sparseness. A reason is that the inventory of these actors is often zero as this is the starting or the ending node of the supply chain. This is an artifact of the simulation model as the product does not stay at the supplier for long (e.g., not longer than 1 day), and products are assumed to be sold or used quite quickly after arriving at the retailer. Since the average inventory of these actors is low, the weights for calculating the global supply chain visibility are also low (Caridi et al., 2010, 2013). Therefore, these outliers have little impact on the resulting global supply chain visibility. Interestingly, when adding all three dimensions of data sparseness to each actor in the scenarios, the supply chain visibility of the supplier, the retailer in Utrecht, and the retailer in Venlo, are somewhat affected by data sparseness, but the effects are very limited.

Second, the sampling method for the dimensions of data sparseness affects the results depending on the data quality criterion for which data sparseness is introduced (Laranjeiro et al., 2015). The missing values dimension results in a small 95% confidence interval and the lowest standard deviation (not more than 1.7%) for the supply chain visibility. An explanation is that the missing values dimension only impacts the quantity of the data. It is more straightforward in which way the data is transformed, so the spread is low. For noise and bias, dimensions that affect the quality of the data, the ranges on how the data can be transformed are wider and, therefore, the spread in supply chain visibility outcomes is larger. Also, as bias is

sampled using a log-normal distribution, there is a higher probability that the correct information of some actors is more often present in the data than the correct information of others. This leads to a higher quality of the data of those actors and, therefore, a higher supply chain visibility (Kalaiarasan et al., 2022). It explains that, in the bias dimension, the average supply chain visibility of most actors is relatively high in comparison to the fast-decreasing supply chain visibility of the retailer in Amsterdam. The data is sampled using a Uniform distribution for the dimensions noise and missing values. As the data of all actors are equally likely to be modified following this Uniform distribution, the actors logically follow the same trend regarding the impact on visibility in these dimensions.

Third, the way of calculating supply chain visibility is of importance when interpreting the results. For the scenarios where only one actor is impacted, the competitor scenario and the key actor scenario, the average supply chain visibility percentage is approximately the same. However, in supply chain theory, a "bull-whip" effect of information would be expected in the key actor scenario, i.e., every actor upstream of the key actor would also be less visible due to the sparse data of the key actor (Lee et al., 1997). This means that, theoretically, degrading data in the key actor scenario leads to a lower global supply chain visibility than in the competitor scenario. However, this effect is not represented in the formulas of global supply chain visibility, and therefore, the results of these two scenarios are the same. This lack of including the "bull-whip" effect is a limitation for the calculation.

When comparing demand-orientation and supply-orientation, the results show that the demand-oriented scenario has a lower global supply chain visibility than the supply-oriented scenario. Additionally, the supply-oriented scenario includes more actors with low supply chain visibility. A cause for this phenomenon is that the weights assigned to each actor for calculating global supply chain visibility are based on average order quantity in units (i.e., inventory levels), following Caridi et al. (2010). Actors upstream in the supply chain generally have more average inventory than those downstream as they use a make-to-stock approach. More specifically, the PPE supply chain is a push supply chain where the supplier and manufacturer create inventory for the long-term demand instead of a pull supply chain where they respond to real-time demand (Nag et al., 2014). This entails that the supplier and the manufacturer have a high weight, contributing more to the global supply chain visibility according to the formula used in our study. Thus, the results hold for cases where the average inventory is a key indicator for determining global supply chain visibility. In other words, the supply chain characteristics are important for calculating the average inventory and, therefore, for the validity of our results. Next to the push and pull characteristic, the structure of the supply chain plays a crucial role in determining the average inventory of actors (Li et al., 2020). For example, if an assembly supply chain of a car were studied with many suppliers of small products like windows and steering wheels, the inventory load might be differently distributed than in the case of PPE. It would be interesting to examine whether these results hold for different types of complex supply chains where inventory is distributed differently.

Fourth, the results are specific to the linear counterfeit PPE supply chain model used in our study. A supply chain is often represented as a sequential network, mean-

ing that, for example, there is a one-directional flow between supplier and manufacturer. On the one hand, this direct and linear dependency between the actors could lead to a more straightforward calculation of supply chain visibility, being a limitation to the generalizability of the results. On the other hand, many supply chains are characterized by a sequential network, even when there are more actors involved. Thus, the effect of the dimensions of data sparseness on supply chain visibility is generalizable to other supply chains with similar complexity.

Fifth, the quality and the quantity of the data, hence the supply chain visibility, are not directly affected by the intentionality of data sparseness as it does not matter whether the actor intentionally transformed the data for calculating supply chain visibility in this study. Therefore, the intentionality aspect of data sparseness is not included in our analysis. However, coping with sparse data and using it for decision-making is different when data is intentionally transformed (Janssen et al., 2017; Oliveira & Handfield, 2019). For example, when bias is intentionally added to the data of counterfeit PPE, it is most likely that fraud-involved organizations try to mask their real activities, and planning effective interventions on this biased data is difficult. Whereas, if data is unintentionally sparse, masking of data for one specific actor in the supply chain does not take place, and effective interventions can still be planned on the biased data. The studied scenarios for a key actor hiding information and a competitor hiding information could be seen as first experiments with intentional data sparseness. As the fraction of intentional sparseness impacts how to cope with data and how to use it in decision-making, it would be interesting to examine the impact of intentionality on data sparseness for decision-making (Bronselaer, 2021).

Last, a limitation of the systematic literature review on supply chain visibility and data quality is that the data collection phase was kept out of scope. For the purpose of this research, only the impact of data sparseness on supply chain visibility has been studied. The literature study provided some possibilities on decreasing data sparseness during the data collection phase, such as the use of IoT, RFID, and blockchain (Pero & Rossi, 2014; Kumar et al., 2022). Extending this research by analyzing how to improve data quality for all phases of the data management process and how to rank these solutions would be interesting for academics and practitioners.

2.9. CONCLUSION

Improving data quality is crucial for enhancing supply chain visibility, because accurate and comprehensive data allows for informed decision-making, monitoring operations, enhancing resilience, and mitigating potential inefficiencies (Munir et al., 2020; Bronselaer, 2021). Poorly informed supply chain management decisions may result from data sparseness, creating challenges for stakeholders to coordinate effectively, and potentially resulting in shortages of products (Janssen et al., 2017; Kalaiarasan et al., 2022). Therefore, it is important to make supply chain practitioners aware of the different dimensions of data sparseness and how these dimensions impact supply chain visibility. However, no clear and concise formalization of data sparseness exists in the current state-of-the-art literature on supply chain management. Additionally, a knowledge gap exists in understanding the extent of the impact caused by different dimensions of data sparseness. Addressing these knowledge gaps

is essential for enhancing the ability of supply chain practitioners to deal with data sparseness, and for contributing to further developments in the supply chain field by explicitly including the notion of data sparseness and its impact.

This research addresses the gaps in the existing literature by providing a classification of data sparseness in the context of supply chains and assessing its impact on supply chain visibility. First, using a systematic literature review, data sparseness is classified into three dimensions: (1) noise, i.e., values in the data set are distorted, (2) bias, i.e., data is not representative of the population or the phenomenon of study, (3) missing values, i.e., values are missing in the data. Each dimension has a certain fraction of intentional sparseness. Thus, sparse data in relation to supply chain visibility is referred to as: *"lack of data quality across the entire supply chain for the quality dimensions: noise, bias, and missing values, where a certain fraction of data sparseness is intentional"*.

Next, the impact of these dimensions on the supply chain visibility is evaluated for an increasing degree of data sparseness. A stylized counterfeit PPE supply chain simulation model is used as ground truth. Data is extracted from this model, and then data sparseness for the three dimensions is systematically added to this data. Hereby, the magnitude of change in supply chain visibility for an increasing degree of data sparseness on each individual dimension is assessed. Four stylized scenarios that could occur in real life regarding data sparseness and their effect on supply chain visibility are also examined.

The main research findings demonstrate that data sparseness greatly affects the visibility of the counterfeit PPE global supply chain. More specifically, data sparseness impacts supply chain visibility, leading to a reduction of up to 52.8% for noise, 65.0% for bias, and 31.7% for missing values. For all three individual dimensions, the average percentage of global supply chain visibility decreases when more sparseness is added to the data, and the visibility values have a small 95% confidence interval. The missing values dimension has the largest impact on the decrease in supply chain visibility, whereas bias has the least impact. The results show the relative importance of the dimensions of data sparseness for actors in the supply chain. The scenario analysis shows that the location of an actor who is unwilling to share data (either a competitor or a key actor) makes no difference for the global supply chain visibility percentage when using the current formulas. The scenario analysis also shows that the demand-oriented scenario has the lowest average global supply chain visibility at 40.6%. A reason is that the global supply chain visibility percentage decreases more when actors with a high average inventory provide sparse data. It also shows that companies with a supply-oriented view will have a better insight into the supply chain visibility than those with a demand-oriented view.

To provide practical advice, this study helps supply chain practitioners by providing information on the relationship between dimensions of data sparseness and supply chain visibility. The primary impact on supply chain visibility appears to be missing data, suggesting that supply chain practitioners should prioritize addressing missing values to improve supply chain visibility. Additionally, companies with a demand-oriented view should prioritize collecting data from upstream as much as possible. This would enhance their decision-making capabilities.

Future research should focus on evaluating the impact of data sparseness on different supply chain configurations in the context of supply chain visibility, e.g., non-sequential supply chain networks. Following on this, expanding the complexity of the simulation model (e.g., including more actors), and therefore, the complexity of the data set is also a direction for future research. Another research direction is to investigate the inclusion of the "bull-whip" effect in the calculation of supply chain visibility, and to include intentionality for evaluating decision-making with data sparseness. A final research direction is to research methods to enhance the data quality management process.

SINGAPORE

3

CALIBRATING SIMULATION MODELS WITH SPARSE DATA

Counterfeit supply chains during Covid-19

COVID-19 related crimes like counterfeit Personal Protective Equipment (PPE) involve complex supply chains with partly unobservable behavior and sparse data, making it challenging to construct a reliable simulation model. Model calibration can help with this, as it is the process of tuning and estimating the model parameters with observed data of the system. A subset of model calibration techniques seems to be able to deal with sparse data in other fields: Genetic Algorithms and Bayesian Inference. However, it is unknown how these techniques perform when accurately calibrating simulation models with sparse data. This research analyzes the quality-of-fit of these two model calibration techniques for a counterfeit PPE simulation model given an increasing degree of data sparseness. The results demonstrate that these techniques are suitable for calibrating a linear supply chain model with randomly missing values. Further research should focus on other techniques, larger set of models, and structural uncertainty.

This chapter has been published as: van Schilt, I. M., Kwakkel, J. H., Mense, J. P., & Verbraeck, A. (2023) Calibrating simulation models with sparse data: Counterfeit supply chains during COVID-19. In B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C. Corlu, L. Lee, E. Chew, T. Roeder, & P. Lendermann (Eds.), *Proceedings of the 2022 Winter Simulation Conference* (pp. 496–507). https://doi.org/10.1109/WSC57314.2022.10015241.
The code is available at https://doi.org/10.4121/a772fd6f-ec0b-4038-8e54-5b9901f060ad.v1.

3.1. INTRODUCTION

During COVID-19, a rise in counterfeit Personal Protective Equipment (PPE) and related criminal activities was detected. Suddenly, there was a high worldwide demand for PPE such as face masks, particulate filter respirators, gloves, goggles, and glasses (Omar et al., 2022). Medical PPE for hospitals have stricter requirements, such as certification, than non-medical PPE. Certified PPE are more valuable than non-certified PPE, making it attractive for criminals to try and sell non-certified PPE as certified PPE. Detecting counterfeit PPE has been challenging since (1) COVID-19 is a new and unexpected phenomenon so there is little historical data, and (2) criminals generally try to share as little data as possible. Together, this makes it hard to get insight into criminal activities pertaining counterfeit PPE, making it a complex system.

Simulation is a way to get insight into complex systems, recognizing relations, and exploring future scenarios (Shannon, 1998). In particular, the focus of this chapter is on discrete event simulation for representing complex socio-technical systems (Schmitt & Singh, 2009). A model can be conceptualized as consisting of variables and relations, and many variables need an initial value in the model to capture an initial state and behavior that is consistent with the state and behavior of the system. These initial values are called parameters; more specifically parameters of components of the model. Some of the parameter values might be observed directly, while others are unobservable and thus have to be tuned to match the behavior of the simulation model with its real world counterpart.

Model calibration can help with constructing a model close to the real world. It is the process of tuning and estimating the model parameters with observed data of the system to improve the similarity between the model and the system. The goal of model calibration is to find those parameter values for which the behavior of the simulation model is as close as possible to the observed behavior of the real system by using real data.

In case of criminal activities in general, and in particular for counterfeit PPE, data is sparse. Criminals want to stay off the grid and generally do not voluntarily share information about their criminal activities. In case of COVID-19 related crimes, data sparseness is even more pronounced due to its novelty. This makes it even more challenging to calibrate models. In cases like this, model calibration should be able to handle sparse observed data. Data sparseness can be classified in three dimensions: (1) noise, (2) bias, and (3) missing values (Huang, 2013; Hazen et al., 2014). This research focuses on one of the three dimensions of data sparseness, missing values. The goal of model calibration with sparse data is to find the most likely model configuration that matches the underlying system.

A subset of model calibration techniques seems to be able to handle sparse data in other fields. For example, Evolutionary Algorithms are widely applied for high-dimensional optimization problems where data often becomes sparse (Ren & Wu, 2013). Bayesian Inference is often used for uncertainty analysis, and is one of the few techniques in machine learning that is able to handle sparse data sets (Jalali et al., 2017; Vrugt & Beven, 2018). Data Assimilation is a promising technique for predicting simulation models in real-time with sparse data (Xie, 2018; Kuipers, 2021). However, it it yet unknown how these techniques perform for the calibration of simulation models given sparse data.

Therefore, this study analyzes two model calibration techniques that are likely suitable for calibration in the case of sparse data. To test these techniques, a case study of a counterfeit PPE supply chain with a focus on intentionally mislabeled PPE is used. We use a stylized discrete event simulation model of a counterfeit PPE supply chain as ground truth. We extract data from this model, systematically increase the degree of sparseness of the data, and assess the extent to which the selected model calibration techniques can still identify the underlying supply chain. We also test a commonly used model calibration technique as reference. This chapter is the first step towards analyzing and comparing various model calibration techniques on simulation models of complex systems in the case of sparse data.

The chapter is structured as follows. In Section 3.2, we discuss the current state-of-the-art literature on model calibration with sparse data, and select the model calibration techniques for this study. In Section 3.3, we explain the design of experiments used to test the selected model calibration techniques. In Section 3.4, we outline the simulation model of the case study, and present the results of the quality-of-fit of the selected model calibration techniques on the case study given an increasing degree of data sparseness. In Section 3.5, we discuss our results. In Section 3.6, we conclude our study, and provide some directions of further research.

3.2. MODEL CALIBRATION TECHNIQUES

Calibration of simulation models is defined as finding values for parameters of the model by using real data until there is a “good" agreement, i.e., as close as possible, between the model data and the observed data over a given time interval (Wigan, 1972; Ören, 1981; Hofmann, 2005). Optimization techniques are commonly used for model calibration as the objective is to parenting the difference between the model data and the observed data (Liu et al., 2017).

3.2.1. RELATED WORK

Malleson (2014) discusses the calibration of simulation models in the field of criminology. The author focuses on the goodness-of-fit in spatial structures. He presents three computer algorithms that help with exploring the parameter space: (1) Hill Climbing, (2) Simulated Annealing, and (3) Genetic Algorithms. Malleson (2014) emphasizes the need for gathering reliable observed data from the criminal system as this is not present yet. He notes that the calibrated model would not represent the real system when data is sparse. In our study, we do not focus on gathering this data but we focus on how to present the real system using model calibration given sparse observed data.

Liu et al. (2017) are one of the first to explicitly addresses calibration of a simulation model under data sparseness. They propose a simulation-optimization approach to automatically calibrate a simulation model with sparse data. They formulate the problem as a series of local minimum search problems. An agent-based model of an emergency department is used as case study. Following from this, De Santis et al. (2022) focus on calibration of a discrete event simulation model under data sparseness. They use the observable values from the target system for finding

values of the simulation model on the level of model parameters, e.g., the time difference between known time stamps. de Groot and Hübl (2021) use calibration as a form of validation. In their case, validation of the simulation model is difficult due to the sparseness of data. They manually adjust parameters and behavior of the model to increase validity.

The main differences between the related work and our research are that (a) we compare various optimization techniques in the case of data sparseness instead of selecting one, and (b) we do not assume that one calibration technique works best for all types of sparse data.

3.2.2. SELECTED TECHNIQUES FOR MODEL CALIBRATION WITH SPARSE DATA

We select a commonly used model calibration technique as reference technique: an exact solver using Powell's Method. As a first attempt to analyze the performance of techniques that seem to be able to deal with sparse data for calibrating simulation models, we select two model calibration techniques: Genetic Algorithms and a Markov Chain Monte Carlo sampling approximate Bayesian computation. The following sections describe these model calibration techniques in more detail.

POWELL'S METHOD

Exact solvers calibrate a model through exact mathematical optimization that guarantees to find (local or global) optimal solutions during model calibration (Puchinger & Raidl, 2005). A commonly used exact algorithm for calibrating simulation models is Powell's Method (Liu et al., 2017). In a rugged high-dimensional fitness landscape typical for discrete event simulations, Powell's Method might be one of the best techniques for calibrating due to its search speed (Zhong & Cai, 2015). Powell's Method is a gradient-free minimization algorithm using a repeated line search introduced by Powell (1964). In more detail, the algorithm selects a starting point and draws two different lines as search directions. On one of these lines, the algorithm performs an one-dimensional optimization to find a new optimal point. From this point on, an one-dimensional optimization is performed on the other line representing the different search direction. With these optimal points, a conjugate search direction is drawn where also an one-dimensional optimization is performed. These steps are repeated until the algorithm finds the optimal solution or when stopping criteria are reached (Vassiliadis & Conejeros, 2009). In this research, the number of iterations and functions evaluations are used as stopping criteria.

GENETIC ALGORITHM

Evolutionary algorithms calibrate a model through population-based, i.e., "survival-of-the-fittest", techniques. One of the oldest and well-known evolutionary algorithms are Genetic Algorithms (GA) (Slowik & Kwasnicka, 2020). GA are widely applied as optimization algorithm in the field of model calibration (Park & Qi, 2005; Malleson, 2014). Classic GA are based on Darwin's theory of natural selection. The idea is that fittest individuals have a higher change to survive, and thus their genes contribute more to the reproduction of the next generation (Whitley, 1994). Each parameter of

the optimization represents a gene. Each solution of the optimization corresponds to a combination of genes, also known as a chromosome of an individual.

GA follow four steps: (1) initialization, (2) selection, (3) recombination, and (4) mutation (Mirjalili, 2019). At the initialization, a random population to ensure diversity in the solution space is spawned. Next, a selection of the best solutions based on their fitness value is created. The fitness value is calculated by the user defined fitness function, i.e., the objective function of the optimization. After this, the chromosomes are combined to produce new chromosomes, also called recombination. This means that two solutions (parents solutions) are selected to produce new solutions (children solutions). Cross-over operators are used to combine and swap the genes of the parent solutions to produce children solutions. In the last step, the genes of some children solutions are altered, also called mutation. In this way, the algorithm maintains the diversity of the population since a certain level of randomness is included in the population. This avoids the probability that GA stay in the local optimum (Mirjalili, 2019).

GA are iterative processes, meaning that it keeps on creating new populations using selection, recombination, and mutation until some user defined stopping criterion is reached. In this research, we use the number of function evaluations as a stopping criterion.

APPROXIMATE BAYESIAN COMPUTATION

Model calibration is a core application of Bayesian data analysis using Bayes' theorem (Csilléry et al., 2010). In the case of sparse data and uncertainties, approximate Bayesian computing (ABC) is one of most suitable techniques for calibrating as it is likelihood-free (Vrugt & Beven, 2018). ABC is a technique for estimating the posterior distribution of model parameters using Bayesian statistics.

One of the most efficient sampling algorithms for ABC is Differential Evolution Adaptive Metropolis (DREAM), a multi-chain Markov Chain Monte Carlo Sampling algorithm (Sadegh & Vrugt, 2014). DREAM combines a multi-chain Markov Chain with differential evolution, as also found in some GA, for population evolution with a Metropolis selection rule. More specifically in the case of calibration, DREAM draws samples using the Markov Chain Monte Carlo Sampling method. These samples are used to run the simulation model, and to collect data. The distance between the simulated data and the observed data is used to either accept or reject a sample using an adaptive selection rule. DREAM uses multiple parallel chains to explore the solutions space adequately, and cross-over of solutions between the chains exists (Vrugt, 2016).

The above steps in each chain are repeated until a stopping criterion, i.e., the number of draws, is reached. When this happens, the accepted samples are used to approximate the posterior parameter distribution.

3.2.3. DISTANCE METRIC

In order to minimize the difference between the simulation model data and the observed data, a so-called distance metric needs to be defined. The distance metric represent the distance between the simulation model data and the observed data given a certain function. Generally, standard statistical functions such as the mean square

error, Kolmogorov-Smirnov metrics, or Euclidean (L2) distance are used as distance metric. However, most of these standard statistical functions do not properly adapt to the data of a specific problem (Suárez et al., 2021). In our case, the metric has to incorporate data of stochastic models in combination with sparse observed data of the system. Moreover, in simulating complex and large systems we typically deal with a high-dimensional space as a result of the many components with their parameters, which makes it challenging to find an appropriate and meaningful distance metric (Aggarwal et al., 2001).

Mirkes et al. (2020) note that the classic distance metric, such as L1 and L2, are highly efficient for complex and high-dimensional data applications. Thus, for the purpose of this study, we use a classic distance metric: the Manhattan (L1) distance. The Manhattan distance is the distance between data points as the sum of the absolute differences normalized for all dimensions.

3.3. DESIGN OF EXPERIMENTS USING GROUND TRUTH

We perform experiments to analyze the quality-of-fit of the selected model calibration techniques for different degrees of data sparseness. First, we explain the set-up for evaluating the quality-of-fit for the three selected model calibration techniques by using the ground truth. Next, we discuss the configuration of each technique.

3.3.1. GROUND TRUTH SET-UP FOR EVALUATING THE QUALITY-OF-FIT

This research uses a ground truth set-up to evaluate the quality-of-fit of the model calibration techniques over various degrees of data sparseness. For replicating the observed data of the system, we use a simulation model that serves as a ground truth and extract data from this model. By using this set-up, we can assess how close the estimation of the calibration is to the true values as these are known. This is nearly impossible with real data (Khondoker et al., 2016).

Figure 3.1 presents the method used for evaluating the model calibration techniques. First, we define a ground truth simulation model with as input decision variable X with ground truth $X = x$. The output of the ground truth simulation model is the *ground truth data*, which does not include any sparseness. Next, we add data sparseness to the ground truth data. For example, 10% of the ground truth data elements are transformed into missing values. This leads to *sparse observed data*. The simulation model is calibrated to the *sparse observed data*. For the calibration, each iterative model calibration technique in essence selects a candidate value for the decision variable, $X = v$ (Frank et al., 2013). Five replications of the simulation model are ran based on the candidate values, leading to the *simulation model data* as output. The replications are combined using the mean value, standard deviation, 5th and 95th percentile, and the average time interval of quantity per actor type. Then, the distance between the *simulated model data* and the *sparse observed data* is calculated using the distance metric. This distance is minimized by the model calibration technique. Based on the distance, the model calibration technique selects new candidate values for the decision variable. This process stops when a stopping criterion is reached. The result is a value for the decision variable, $X = v^*$, that best describes

the ground truth model, according to the model calibration technique.

Although the model calibration techniques minimizes the distance between the simulated and observed *output* data, the decision variable of the calibrated simulation model is not necessarily close to the decision variable of the ground truth model. So, we introduce the quality-of-fit of the decision variables which are defined as the normalized distance between the ground truth decision variable, $X = x$, and the optimal decision variable resulting from the simulation model calibration, $X = v^*$. This quality-of-fit is calculated by normalizing the difference between the ground truth input, $X = x$, and the solution, $X = v^*$, given the upper and lower bounds of the decision variable X. A quality-of-fit of 0 means that the optimal solution resulting from the calibration is not close to the ground truth; a quality-of-fit of 1 means that the optimal solution resulting from the calibration is the same as the ground truth.

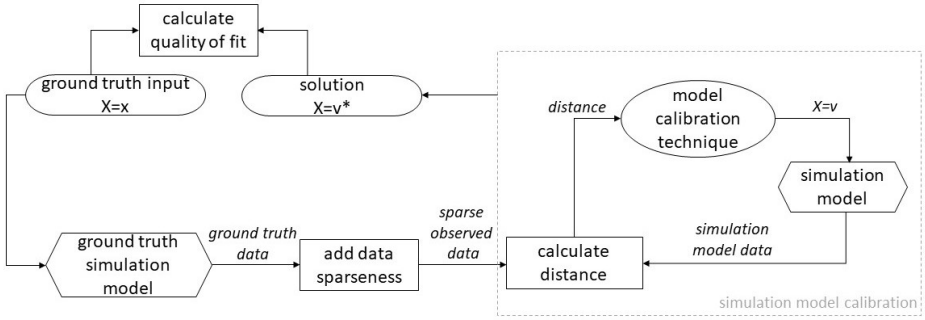


Figure 3.1: Method for evaluating calibration of a simulation model with sparse data.

The above steps represent one experiment for evaluating the quality-of-fit of a model calibration technique, given a certain degree of data sparseness. We systematically increase the degree of data sparseness added to the *ground truth data*. We evaluate for 10%, 25%, 50%, 75%, and 90% data sparseness. A degree of *x%* means that *x%* of the original data elements are missing values. It is randomly determined which *x%* of data elements are missing over the entire data set. Additionally, the model calibration techniques are examined for 0% of data sparseness, i.e., *ground truth data*, as a base case. Each experiment is performed with 8 seeds to account for the effect of stochasticity on the quality-of-fit. For each seed, we first transform *x%* of the data set to missing values, and then we use this as input for all three model calibration techniques. This means that the exact same observations were left out of the data set that is presented to the different techniques for simulation model calibration.

3.3.2. CONFIGURATION OF MODEL CALIBRATION TECHNIQUES

To calculate the quality-of-fit for the selected model calibration techniques, the ground truth decision variable, $X = x$, is compared to the optimal solution, $X = v^*$, for each of the model calibration techniques. The result of Powell's Method and GA is a single optimal solution of the decision variable, so $X = v^*$. However, the result of ABC is an approximate posterior distribution of the decision variable. To extract

one optimal value of decision variable X from this resulting posterior distribution, we select the value with the highest frequency, i.e., the mode, for that specific distribution. In this way, the most often accepted value of the decision variable represents the optimal solution for ABC as $X = v^*$.

For pragmatic reasons, we define a stopping criterion for finding the optimal solution for each technique. The stopping criteria for these experiments are based on an empirical analysis on the convergence of the model calibration techniques over 5 seeds. For the reference technique, Powell's Method, we limit the number of function evaluations to 1500 and the number of iterations to 100. For GA, we use 15.000 function evaluations as a stopping criterion. The analysis shows that with 15.000 function evaluations, the number of improvements stays constant for every seed. For ABC, we use 20.000 draws as the stopping criterion. The analysis shows that there is convergence of ABC determined by the Gelman-Rubin statistics at 20.000 draws for 3 of the 5 seeds (Gelman & Rubin, 1992).

3.4. CASE STUDY: COUNTERFEIT PPE SUPPLY CHAIN

To evaluate the model calibration techniques, we use a case study of a counterfeit PPE supply chain. First, we introduce the stylized simulation model based on this case study. Next, we discuss the analysis and comparison of the various model calibration techniques given the simulation model of this case study.

3.4.1. INTRODUCTION OF THE SIMULATION MODEL

A discrete event simulation model of a stylized configuration of a counterfeit PPE supply chain from Vietnam to stores in the Netherlands is used. We assume that the counterfeit PPE are produced in Vietnam; one of the countries where most PPE come from, next to China and India. Most of these products are transported over sea to Europe following the legitimate transport flows. After arrival in Europe, they are distributed over various stores.

Figure 3.2 visualizes the stylized counterfeit PPE supply chain in more detail. The symbols represent the main actors in the supply chain, and the arrows represent the transportation flows. Starting from the supplier, supplies for PPE such as fabrics are delivered to the manufacturer over land in the production country, Vietnam. The manufacturer produces the protective mask (mislabeling them as medical) PPE in the factory and packs them in batches for transport. Each batch has a certain quantity of counterfeit PPE. For example, a batch consists of 1000 boxes of 100 PPE that equals a quantity of 100,000 PPE in total. Next, a truck transports a batch of finished counterfeit PPE to the export port in Hai Phong, Vietnam. The batch is loaded into a container and transported by a feeder to the transit port, Tanjung Pelepas, Malaysia. Once the batch is loaded on the feeder, it becomes part of the legitimate transport flow. At the transit port, the feeder unloads the container with counterfeit PPE. At the same port, the container is loaded onto a vessel, i.e., a larger container ship, for international transport. After a certain amount of days on international waters, the vessel arrives at the import port in Rotterdam, The Netherlands. The container is unloaded here, and waits for inland transport to the wholesales distributor in Eindhoven, The

Netherlands. The wholesales distributor can also be seen as the stash location for the counterfeit PPE. At the wholesales distributor, the batch of counterfeit PPE in the container is equally divided into three smaller batches for the retailers. These smaller batches are transported by small trucks to the retailer. When the counterfeit PPE arrive at the retailer, customers (either businesses or individual customers) can purchase the products with or without being informed that they are counterfeit.

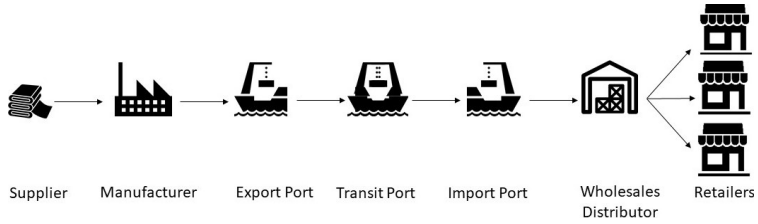


Figure 3.2: Visualization of the stylized counterfeit PPE supply chain.

The structure of the supply chain is linear. Due to the many uncertainties in the supply chain (e.g., delay in transport modalities, loading and unloading times), the supply chain becomes complex. For example, the retailer's inventory can fluctuate very much, depending on whether a vessel has a 1-day delay or a 7-day delay. In the simulation model, most uncertainties such as delays of transport modalities and speed of transport modalities follow triangular distributions inspired by real world data of a fashion retailer (Kuipers, 2021).

In this research, the manufacturing duration, also referred to as manufacturing time, is the system parameter to be calibrated. More specifically, we use the manufacturing time as the decision variable in the simulation model calibration, meaning that we seek for the most likely value for the system parameter of manufacturing time. Table 3.1 shows the configuration of the manufacturing time as a decision variable. Manufacturing time has been chosen as an uncertain system parameter in this study for three reasons: (1) manufacturing time in another country is typically unobservable from the client's location, (2) there were many orders due to COVID-19 that could lead to extreme delays, and (3) delays at the beginning of the supply chain often have an unpredictably high impact on the rest of the supply chain due to the snowball effect.

Decision Variable	**Ground Truth**	**Lower Bound**	**Upper Bound**	**Unit**
Manufacturing Time (X)	2.5	1	10	Days

Table 3.1: Configuration of manufacturing time.

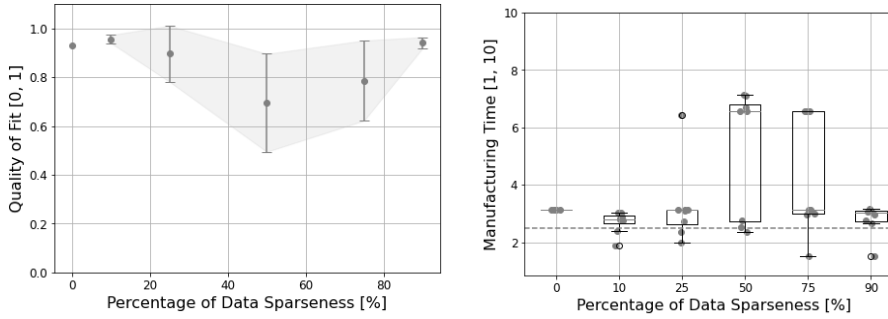
Given the value of the decision variable, in this case the manufacturing time, the simulation model is evaluated using a time series of the inventory levels of PPE for each actor in the supply chain (e.g., manufacturer, export port, import port) per day. The time series over multiple replications are combined using the mean values per day. Aggregated statistics of these combined time series are created, serving as the *simulation model data*. The statistics to represent the time series of each actor are

the mean, standard deviation, 5th percentile, 95th percentile, and the average interval time (i.e., interval between the arrival of batches at actors). The data used for calibration includes the aggregated statistics of all actors.

The discrete event simulation model is developed with the library pydsol in Python. This library is a Python implementation of the Distributed Simulation Object Library (DSOL), originally implemented in Java (Jacobs, 2005).

3.4.2. ANALYSIS OF POWELL'S METHOD, GA & ABC

We analyze the quality-of-fit for the reference technique, Powell's Method, and the selected techniques, GA and ABC, given certain degrees of data sparseness and using 8 replications with unique seeds. For each technique, we show a graph of the average quality-of-fit with a 95% confidence interval to visualize the spread of the solutions over various replications. Besides, we show a boxplot of the calculated optimal values of the decision variable *manufacturing time* resulting from the various replications. In addition, a table is presented to compare the reference technique and the selected model calibration techniques by the average quality-of-fit and the standard deviation.

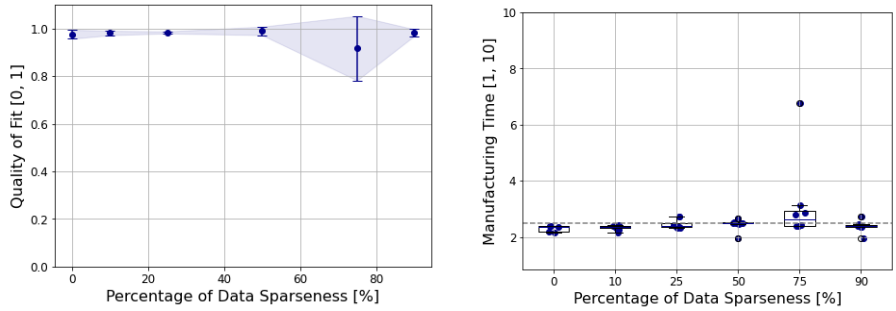


(a) Average quality-of-fit with the 95% confidence interval.

(b) Boxplot of the optimal values of manufacturing time. The dashed gray line is the ground truth value of manufacturing time: 2.5 days.

Figure 3.3: Results for Powell's Method for 8 seeds for various degrees of data sparseness.

Figure 3.3a shows that Powell's Method has an average quality-of-fit between 0.70 to 0.96. When data sparseness is more than 10%, the average quality-of-fit decreases and the 95% confidence interval becomes wider. Figure 3.3b shows that from 10% data sparseness onward, the algorithm finds optimal values of more than 6 days for the manufacturing time. Interestingly, there are no optimal values found between 3 and 6 days. Surprisingly, Powell's method has a high quality-of-fit with a small confidence interval with 90% data sparseness.

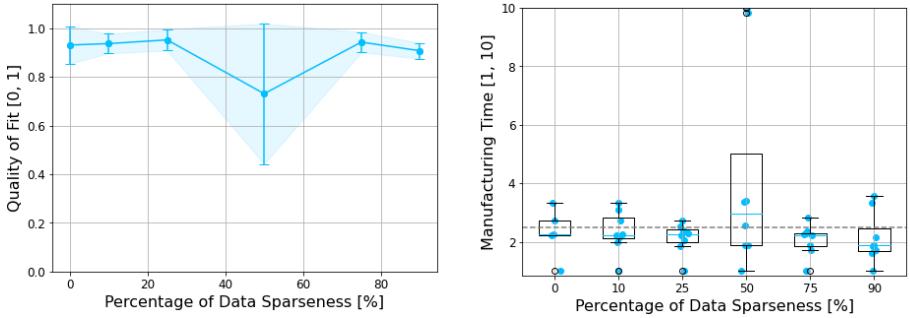


(a) Average quality-of-fit with the 95% confidence interval.

(b) Boxplot of the optimal values of manufacturing time. The dashed gray line is the ground truth value of manufacturing time: 2.5 days.

Figure 3.4: Results for Genetic Algorithm for 8 seeds for various degrees of data sparseness.

Figure 3.4a shows that GA has an average quality-of-fit between 0.92 and 0.99. The quality-of-fit and the correlated spread stays constant for most of the chosen values for data sparseness. The 95% confidence interval is narrow for the different degrees of data sparseness. Only with 75% data sparseness, there are more solutions that have a lower quality-of-fit and the 95% confidence interval is wider. In Figure 3.4b, we see that for 75% data sparseness, most optimal solutions for the decision variable are slightly above the ground truth value. There is one outlier where the optimal manufacturing time is calculated to be more than 6 days.



(a) Average quality-of-fit with the 95% confidence interval.

(b) Boxplot of the optimal values of manufacturing time. The dashed gray line is the ground truth value of manufacturing time: 2.5 days.

Figure 3.5: Results for approximate Bayesian computation for 8 seeds for various degrees of data sparseness.

Figure 3.5a shows that ABC has an average quality-of-fit between 0.78 and 0.98. For most of the degrees of data sparseness, the average quality-of-fit is around 0.95 and the 95% confidence interval is narrow. Only at 50% data sparseness, the average quality-of-fit is the lowest, i.e., around 0.78, and the confidence interval is relatively wide. Figure 3.5b shows that at 50% data sparseness, the algorithm has a wide spread of optimal solutions for the value of manufacturing time. Some solutions are close to

the lower bound and the upper bound of this decision variable. Other solutions are closer to the ground truth value, i.e., between 3 and 4 days, but are still relatively far from the ground truth compared to experiments with other degrees of data sparseness. The resulting posterior distribution of the algorithm for 50% data sparseness follows a bimodal distribution.

	Powell's Method		**Genetic Algorithm**		**Approximate Bayesian Computation**	
Percentage of Data Sparseness	*Mean*	*Std*	*Mean*	*Std*	*Mean*	*Std*
0%	0.93	0.00	0.98	0.01	0.93	0.05
10%	0.96	0.02	0.98	0.01	0.94	0.04
25%	0.90	0.13	0.98	0.00	0.95	0.04
50%	0.70	0.22	0.99	0.02	0.73	0.32
75%	0.79	0.18	0.92	0.15	0.94	0.05
90%	0.94	0.03	0.98	0.02	0.91	0.04

Table 3.2: Quality-of-fit in mean and standard deviation for each model calibration techniques for various degrees of data sparseness.

Table 3.2 presents the results of the average quality-of-fit and the corresponding standard deviation of the reference technique and the two selected model calibration techniques for various degrees of data sparseness. It shows that GA outperforms Powell's Method and ABC for all percentages of data sparseness in terms of a higher average quality-of-fit and a lower standard deviation. ABC performs slightly better on the average quality-of-fit than Powell's Method. However, Powell's Method and ABC both have a relatively high standard deviation compared to GA, meaning that there is more variation in the distance of the optimal solution to the ground truth value. Over the various degrees of data sparseness, Powell's Method has the highest standard deviation. It is quite remarkable that Powell's Method and ABC have the lowest average quality-of-fit and the highest standard deviation for 50% data sparseness. For both techniques, the average quality-of-fit increases again for 75% and 90% data sparseness.

Overall, GA and ABC outperform the reference technique for calibrating the counterfeit PPE supply chain simulation model over various values for data sparseness. From this analysis, GA shows to be the most promising for calibrating a simulation model with sparse data due to the high average quality-of-fit, the narrow 95% confidence interval, and a small standard deviation over all degrees of data sparseness.

3.5. DISCUSSION

Overall, the results show that the selected model calibration techniques seem to have a high quality-of-fit for calibrating the counterfeit PPE simulation model with sparse data. There are three limitations for generalizing the results: (1) local vs. global optimum, (2) specific to supply chains, and (3) lack of including structural uncertainty and of including other dimensions of data sparseness.

Regarding the local vs. global optimum, it is remarkable that Powell's Method and ABC both have the lowest quality-of-fit and the highest standard deviation at 50% data sparseness. A possible explanation for this result for Powell's Method is that the algorithm sometimes gets stuck in a local optimum, instead of reaching the global optimum (Powell, 1964), possibly caused by two input spaces of interest. A possible explanation for ABC is that the algorithm results in a bimodal distribution, with more than one region in the input space that results in optimal solutions. Calibration with these algorithms can yield multiple counterfeit PPE supply chains that could represent the real world supply chain to a certain extent. We should therefore be careful in choosing the configuration to gain insights from. Not doing so could lead to a "wrong" view on criminal activities in the real world counterfeit PPE supply chain. In addition, a wider set of optimization algorithms could be explored for their effectiveness in model calibration.

The second limitation is that the results could be specific to the linear counterfeit PPE supply chain model. In general, a supply chain is often presented as a sequential network. This means, for example, that there is an one-directional flow between the supplier and the manufacturer. On the one hand, this direct and linear dependency between the actors could lead to more straightforward calibration of the simulation model with sparse data. This challenges the generalizability of the results to other systems. The linear supply chain also has a single parameter that needed to be calibrated, where in real situations, data of multiple parameters might be sparse. On the other hand, the results of this chapter give a proof of concept on how data sparseness effects the ability to calibrate a linear supply chain using sparse data.

The third limitation is that the lack of including structural uncertainty and of including other dimensions of data sparseness. Keeping the structure of the simulation model the same for the ground truth and the calibrated model could be a crucial element for being able to find the optimal value for the parameter(s). When structure is included as a parameter, this could mean that it is more difficult for the model calibration techniques to converge to a solution with a high degree of data sparseness. In our example case, data sparseness in the form of missing data values were random, where in reality there could be patterns, such as missing data only during the night. Finally, data sparseness consists of more dimensions than missing values: examples are noise and bias. The effect of these other types of data sparseness on calibration quality is still unknown, making it difficult to generalize the results to all types of data sets. Nonetheless, this study gives insight in the quality-of-fit for one parameter when increasing the percentage of missing values, a type of uncertainty that often occurs in criminal cases, specifically during COVID-19.

3.6. CONCLUSION AND FUTURE WORK

This research is a first attempt to analyze the quality-of-fit of model calibration techniques that are likely to be suitable for calibrating simulation models in the case of sparse data. Due to the high data sparseness in counterfeit PPE supply chains, we used a PPE supply chain as our case study. We selected a reference technique that is often used for calibration of simulation models: Powell's Method. We selected GA and ABC as model calibration techniques that are likely to be suitable in case of sparse data. By using a ground truth set-up for evaluating the quality-of-fit, we assessed how accurately the three model calibration techniques find the optimal system parameter value for the simulation model with an increasing degree of data sparseness. The results demonstrate that the selected model calibration techniques are suitable for calibrating simulation models when faced with sparse data, at least for a linear supply chain with randomly missing values. This shows that with sparse data due to COVID-19 and criminals masking their data, the selected model calibration techniques can help to gain insight in underlying counterfeit PPE supply chains.

The main directions for future research are including more model calibration techniques, evaluating for a larger set of simulation models, introducing structural uncertainty and other dimensions of data sparseness.

WASHINGTON D.C.

4

IDENTIFYING THE STRUCTURE OF ILLICIT SUPPLY CHAINS WITH SPARSE DATA

A simulation model calibration approach

Illicit supply chains for products like counterfeit PPE are characterized by sparse data and great uncertainty about the operational and logistical structure, making criminal activities largely invisible to law enforcement and challenging to intervene in. Simulation is a way to get insight into the behavior of complex systems, using calibration to tune model parameters to match its real-world counterpart. Calibration methods for simulation models of illicit supply chains should work with sparse data, while also tuning the structure of the simulation model. Thus, this study addresses the question: "To what extent can various model calibration techniques reconstruct the underlying structure of an illicit supply chain when varying the degree of data sparseness?" We evaluate the quality-of-fit of a reference technique, Powell's Method, and three model calibration techniques for sparse data: Approximate Bayesian Computing, Bayesian Optimization, and Genetic Algorithms. For this, we use a simulation model of a stylized counterfeit PPE supply chain as ground truth. We parameterize structural uncertainty using System Entity Structure. The results demonstrate that Bayesian Optimization and Genetic Algorithms are suitable for reconstructing the underlying structure of an illicit supply chain for a varying degree of data sparseness. Both techniques identify a diverse set of optimal solutions that fit with the sparse data. For a comprehensive understanding of illicit supply chain structures, we propose to combine the results of the two techniques. Future research should focus on developing a combined algorithm and incorporating solution diversity.

This chapter has been published as: van Schilt, I. M., Kwakkel, J. H., Mense, J. P., & Verbraeck, A. (2024) Identifying the structure of illicit supply chains with sparse data: A simulation model calibration approach. *Advanced Engineering Informatics, 62*, pp. 102926. https://doi.org/10.1016/j.aei.2024.102926.
The code is available at https://github.com/imvs95/structure_calibration_sparse_data.

4.1. INTRODUCTION

During the COVID-19 pandemic, there has been a major increase in demand for Personal Protective Equipment (PPE) like face masks, gloves, and glasses (Omar et al., 2022). PPE can be divided into two categories: medical and non-medical. Medical PPE is certified and typically comes with a higher price and profit margin, making it an interesting target for organizations engaged in fraud(Ippolito et al., 2020). A significant number of fraud-involved PPE manufacturers entered the market during the initial stages of COVID-19, trying to sell non-certified PPE as certified PPE (Hashemi et al., 2023). Law enforcement detected and seized over 58 million counterfeit 3M respirators since the pandemic's beginning (as of May 2022), yet this only represents a fraction of the total (Hashemi et al., 2022). Detecting counterfeit PPE has been challenging as little historical data on COVID-19 is available, and fraud-involved organizations obfuscate their data as much as possible. Consequently, criminal activities and the related logistics operations remain largely invisible (van Schilt et al., 2023). Therefore, identifying counterfeit PPE and effectively intervening in this largely invisible supply chain is difficult for law enforcement.

The counterfeit PPE supply chain is just one example of an illicit supply chain in which law enforcement faces challenges for intervening and stopping criminal activities (Nellemann et al., 2018). Often, only sparse information and data is available of any illicit supply chain. This results in uncertainties regarding the operational and logistical working of the illicit supply chain (e.g., processing times, travel times), as well as the overall structural composition of the supply chain (e.g., how many actors are involved, which sequence of supply chain activities is used, where the actors are located) (Eser et al., 2015; Ficara et al., 2021). More information on the supply chain can be gathered using the experiences of law enforcement, asking for information from criminals, open-source data, and theories on legal supply chains (Magliocca et al., 2019). Information collection is difficult in the context of illicit supply chains; for example, data on police operations is often incident-based, criminals either withhold information, or data is still insufficient for understanding the complete logistics operations of the criminals (Anzoom et al., 2021).

Especially in the case of illicit supply chains, the operational and logistical structure, including geographical boundaries, is often not known to law enforcement (Ficara et al., 2021). This structure is crucial for identifying opportunities to disrupt such a supply chain. Criminals use various modi operandi, routes, communication channels, and business models, impacting the flow of goods and, hence, the structure and geographical context of the supply chain (Duijn et al., 2014; Anzoom et al., 2021). Also, criminals often take advantage of legal supply chains to mask their illicit activities, i.e., piggybacking, which could make the illicit supply chain even more invisible (Grossman & Shapiro, 1986; Shelley, 2018). The structure and geographical scope of supply chains for illicit products are also based on factors like corruption that enable bypassing inspections, clearing customs, and weakening law enforcement. This wide variety of possibilities for carrying out criminal activities and their invisibility complicate the efforts of law enforcement to uncover details about illicit supply chains, including the identities and details of particular persons, the actual operational and logistical structure, methods, and modes. Complex supply

chains characterized by sparse data and structural uncertainty make it challenging to stop crime.

Simulation can help to get insight into complex systems, understand behavior and relations, and explore future scenarios using computers (Banks, 1998; Zeigler et al., 2018). This chapter focuses on the use of discrete event simulation models for understanding illicit supply chains (Schmitt & Singh, 2009; Magliocca et al., 2022). Simulation models require data to mimic the behavior of the real world, either for the parameters of components of the model, such as processing times, or for defining the structure of the components in the model, such as the network of the supply chain. For this, model calibration is used as it is the process of tuning and estimating the model parameters with observed data of the system to improve the similarity between the model and the system (Wigan, 1972; Ören, 1981; Hofmann, 2005).

In the case of simulating illicit supply chains, model calibration should be able to handle sparse observed data (Anzoom et al., 2021; Ficara et al., 2021; Lian et al., 2024). Three dimensions of data sparseness are defined: (1) noise, (2) bias, and (3) missing values (van Schilt et al., 2024). A number of studies have investigated the calibration of simulation models in the context of data sparseness while assuming that the structure of the model is known and fixed (Liu et al., 2017; de Groot & Hübl, 2021; van Schilt et al., 2023). However, it has not yet been investigated how simulation model calibration techniques perform in the case of sparse data combined with structural uncertainty. Assessing the performance of model calibration techniques in the case of sparse data for studying illicit supply chains is further complicated by how structural uncertainty is modeled. Many interdependencies exist among various actors in a supply chain, making it difficult to view actors as independent components in a simulation model, unlike parameters (Baldissera Pacchetti, 2021). Since model calibration mostly focuses on tuning the model parameters and not the model structure, tuning both simultaneously is far more challenging than just tuning the parameters (Moore & Doherty, 2005; Coenen et al., 2018).

This study assesses the extent to which model calibration techniques can accurately reconstruct the underlying structure of the supply chain with a varying degree of data sparseness. First, we review related work on the modeling and simulation of illicit supply chains, and model calibration and its challenges when data is sparse. Next, we evaluate the quality-of-fit of a set of model calibration techniques for accurately reconstructing the structure of the illicit supply chain. For this, we use a stylized ground truth simulation model of a counterfeit PPE supply chain based on real-world data. We extract data from this simulation model, systematically vary the degree of data sparseness, and assess to which extent the selected model calibration techniques can reconstruct the structure of the supply chain.

More explicitly, our study aims to address the question: *"To what extent can various model calibration techniques reconstruct the underlying structure of an illicit supply chain when varying the degree of data sparseness?"*. Accordingly, this chapter lays the foundation for modeling (illicit) supply chains characterized by structural uncertainty and sparse data, to get insights into their operations and hence, allow law enforcement agencies to effectively intervene to stop crime.

This chapter is structured as follows. Section 4.2 presents the related work on modeling and simulation for illicit supply chains. Section 4.3 discusses the current state-of-the-art literature on model calibration and its challenges when dealing with sparse data. Section 4.4 describes the design of experiments, the simulation model of the case study, and the configuration of the selected model calibration techniques. Section 4.5 shows the quality-of-fit for reconstructing the structure of the illicit supply chain using the simulation model calibration approach. Section 4.6 discusses our results. Section 4.7 concludes our work and provides directions for further research.

4.2. MODELING AND SIMULATION FOR ILLICIT SUPPLY CHAINS

This section describes the current state-of-the-art literature on modeling and simulation of illicit supply chains. First, related work on illicit supply chains using simulation is examined. Second, structural uncertainty in illicit supply chain simulation models is described.

4.2.1. RELATED WORK

Simulation is a vital computational approach for understanding the behavior of illicit supply chains, and for exploring further scenarios like the effect of interventions (Anzoom et al., 2021; Magliocca et al., 2022; van Schilt et al., 2023). Anzoom et al. (2021) present a literature review of illicit supply chain network research focusing on operational research, management science, and industrial engineering. Most studies focus on network design, optimization, or social science theories. Their review reveals that only a few simulation studies of illicit supply chains have been conducted.

Some of these studies simulate the criminal network by focusing on the business model, the roles, and how the network evolves over time, drawing on social science theories. Duijn et al. (2014) simulates a criminal cannabis cultivation network to understand the dynamics of resilience in this network as a consequence of disruptions. The authors primarily focus on the dynamics between different roles of actors within the network, and not on the logistical operations. van der Zwet et al. (2019) design an agent-based model for emergent opponent behavior, which is present in organized crime groups that, for example, traffic illicit products.

Other simulation studies focus on replicating the supply and demand in the illicit supply chain to evaluate the effect of disruption strategies in the drug market (Caulkins, 1993; Rydell et al., 1996). More recent studies focus on developing more detailed simulation models to understand disruption strategies in a specific supply chain. For instance, Dray et al. (2008) develop an agent-based model for interaction between individuals and the supply in the heroin market. Kovari and Pruyt (2012) create a system dynamic simulation model of human trafficking for evaluating the effect of policy interventions in the Netherlands. Kretschmann and Münsterberg (2017) present a discrete event simulation model for testing one specific detection method at the border.

Specifically on trafficking, Magliocca et al. (2019) develop a spatial agent-based simulation model of cocaine traffickers to the United States via Central America

based on qualitative data such as theoretical perspectives, media reports, empirical studies, and field research. Their model produces realistic patterns of cocaine trafficking in space and time in response to interventions. Jensen and Dignum (2019) model the illegal cocaine trafficking supply chain based on legal supply chain theories. The authors investigate the difference between the legal and illegal supply chain with a focus on trust. They indicate that more work on the simulation model itself has to be done to enhance its accuracy when representing illegal supply chains. González Ordiano et al. (2020) identify potential geographical hotspots in the illicit supply chain using a variable state resolution Markov Chain, assuming three scales of connectivity (e.g., countries, regions, continents). Their approach consists of two steps: (1) to create a series of Markov Chain models that describe the network in different state spaces, and (2) to select the model that describes the network best. Benatia et al. (2022) evaluates frequent pattern mining for tracing counterfeit products in a supply chain, specifically cosmetics, using a multi-agent simulation model.

In the most recent studies, simulation and optimization models are coupled to analyze interventions in illicit supply chains. Magliocca et al. (2022) introduce coupled agent-based and spatial optimization models for examining the deployment of interventions and the correlated adaptive response of the drug network over time. Their results show that increasing interventions lead to diversifying of the routes and dispersing of the illicit shipment volumes, making it more difficult to seize illicit products. Hashemi et al. (2023) use a simulation-optimization framework to model counterfeiters' behavior and analyze different disruption strategies. They use a scenario tree structure to model the uncertainties in the simulation and optimize the supply chain operations of the criminals on maximizing profit and minimizing risk. van Schilt et al. (2023) test the performance of various optimization techniques for accurately calibrating the parameters of a discrete event simulation model of an illicit supply chain when increasing the degree of data sparseness. Their results show that the simulation model calibration of parameters successfully works in situations with sparse data. They note that an interesting further direction of research is to investigate the performance for finding the underlying structure of the supply chain.

Unlike most previous research that typically uses a single simulation model structure with uncertain parameters, we address structural uncertainty. Our study uses a similar simulation model calibration approach as van Schilt et al. (2023), but it focuses on finding the most representative structure of the real-world supply chain instead of just finding the most likely parameters' values for an assumed structure. Compared to previous studies using a simulation-optimization approach, we focus on calibrating the underlying structure of the supply chain rather than optimizing interventions.

4.2.2. STRUCTURAL UNCERTAINTY IN ILLICIT SUPPLY CHAIN SIMULATIONS

Building a simulation model for an illicit supply chain requires knowledge to ensure it aligns with the system, e.g., the real-world illicit supply chain under study. Certain aspects of such an illicit supply chain remain uncertain, while others are observ-

able. For an illicit supply chain, the knowledge that is required to design a simulation model is often deeply uncertain, meaning that there is no clear consensus on the conceptual model of the system, the probability distributions, or the desirability of outcomes of the model (Lempert et al., 2003; Marchau et al., 2019; Ficara et al., 2021).

We can distinguish two types of uncertainty in illicit supply chain simulation models that have to match the system's counterpart: (1) parametric uncertainty, i.e., uncertainty in (initial) values of the model's parameters or conditions, and (2) structural uncertainty, i.e., uncertainty in the structure of the model (Webster & Sokolov, 1998; Parker, 2013; Parker, 2014). Parametric uncertainty describes the uncertainty in initial values of the model for capturing an initial state and behavior that matches its real-world counterpart. For example, uncertainty about the parameters used to choose a route based on maximizing profit of a fraudulent actor like the cost of transport, or on minimizing risk like the parameter of the risk of getting caught. Structural uncertainty describes uncertainty in the modeling equations, structure, or behavior of the model (Baldissera Pacchetti, 2021). For example, uncertainty about the number of fraudulent actors and their relation in a supply chain, and how these actors choose a route.

In the field of logistics, research has been performed on exploring parametric uncertainty but not on structural uncertainty (Halim et al., 2016; Coenen et al., 2018; Moallemi & Köhler, 2019). Especially in the case of illicit supply chains, the structure is often uncertain (van Schilt et al., 2023). Therefore, the innovative contribution of this research is addressing structural uncertainty for simulation models related to illicit supply chains.

4.3. MODEL CALIBRATION AND THE CHALLENGES WITH SPARSE DATA

This section describes the current state-of-the-art literature on model calibration and its challenges in the case of sparse data. First, related work on model calibration with sparse data is discussed. Second, an overview of model calibration techniques that seem suitable for dealing with sparseness is presented. Third, a modeling approach for structural uncertainty regarding calibration is described.

4.3.1. RELATED WORK

Few studies have investigated the calibration of simulation models in the context of data sparseness. Liu et al. (2017) are one of the first to explicitly address the calibration of a simulation model under data sparseness. They propose a simulation-optimization approach to calibrate an agent-based simulation model with sparse data automatically using an emergency department as a case study. The problem is formulated as a series of local minimum search problems. Subsequently, De Santis et al. (2022) focus on the calibration of a discrete event simulation model under data sparseness. Observable values from a real-world system are used to determine the parameter values of the simulation model, for example, the time interval between known time stamps. de Groot and Hübl (2021) use calibration as a form of validation, and in their case, the sparseness of data makes validating the simulation model

challenging. Consequently, they manually fine-tune the parameters and dynamics of the model to enhance validity. Hao et al. (2021) uses evolutionary neural networks to build more accurate surrogate simulation models with limited data. van Schilt et al. (2023) compare various calibration techniques for simulation models when increasing the degree of data sparseness. They calibrate the parameters of a discrete event simulation model on a counterfeit PPE supply chain.

In line with this, our study compares the performance of various calibration techniques for data sparseness rather than selecting one. We assume that a single calibration technique is most probably not able to deal with all types of sparse data. The novelty of our work is that we apply the simulation model calibration approach to identify the underlying structure of the simulation model, as opposed to only focusing on the parameters.

4.3.2. MODEL CALIBRATION TECHNIQUES FOR SPARSE DATA

Calibration of simulation models involves finding parameter values by comparing the model's output with real data until a "good" match is achieved, meaning that the model data closely matches the observed data over a given time interval (Wigan, 1972; Ören, 1981; Hofmann, 2005). As model calibration aims to minimize the difference between the model data and the observed data, optimization techniques are commonly used for this purpose (Liu et al., 2017; van Droffelaar et al., 2024). We distinguish four families of calibration techniques that are interesting when dealing with sparse data: (1) Deterministic mathematical solvers, (2) Evolutionary algorithms, (3) Bayesian inference, and (4) Data assimilation.

Figure 4.1 shows an overview of the families, the techniques, and the algorithms that can be applied for model calibration in the case of sparse data. Note that this is a non-exhaustive overview.

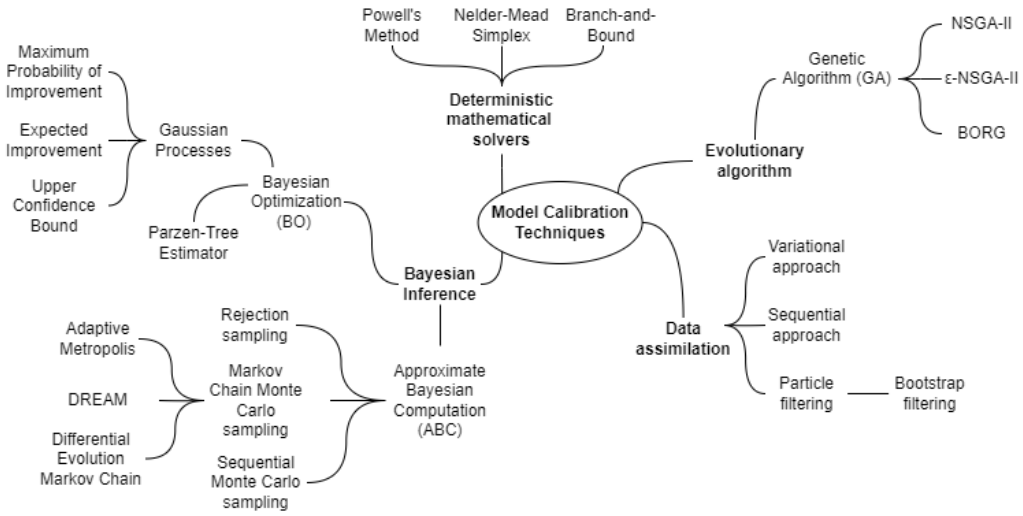


Figure 4.1: Overview of model calibration families and techniques for sparse data.

Deterministic mathematical solvers calibrate models through deterministic mathematical optimization that guarantees to discover (local or global) optimal solutions (Puchinger & Raidl, 2005). A commonly used deterministic algorithm for calibrating simulation models is Powell's Method (Liu et al., 2017). Powell's Method is a gradient-free minimization algorithm using a repeated line search introduced by Powell (1964). Due to its fast search speed, this method is preferred for calibrating discrete event simulation models that are typically characterized by a rugged high-dimensional fitness landscape (Zhong & Cai, 2015). Another example of a deterministic mathematical solver for model calibration is the Nelder-Mead Simplex algorithm (Olsson & Nelson, 1975). Moreover, the Branch-and-Bound algorithm is also commonly used for optimizing linear or mixed-integer programs (Lawler & Wood, 1966; Morrison et al., 2016).

Evolutionary algorithms calibrate a model through population-based, also called, "survival-of-the-fittest", techniques. One of the oldest and well-known evolutionary algorithms are Genetic Algorithms (GA) (Slowik & Kwasnicka, 2020). GA are widely applied in the field of model calibration, especially in high-dimensional problems where data is often sparse (Park & Qi, 2005; Ren & Wu, 2013; Malleson, 2014). Classic GA are based on Darwin's theory of natural selection. The main idea is that the fittest individuals are more likely to survive, and thus contribute more to the next generation (Whitley, 1994). A classic and popular algorithm is Non-dominated Sorting Genetic Algorithm II (NSGA-II) (Deb et al., 2002; Reed et al., 2013). Based on NSGA-II, ε-NSGA-II was introduced that merges NSGA-II with a ε-search algorithm to define the search precision for more efficiency, reliability, and more ease of use (Kollat & Reed, 2007). For more complex and multi-objective problems, BORG is a suitable algorithm (Salazar et al., 2016). BORG is an extension of ε-NSGA-II with adaptive operator selection, meaning that it adapts to the most appropriate operator based on the performance (Hadka & Reed, 2013).

Bayesian inference uses Bayes' theorem to calibrate models. Model calibration is a core application of Bayesian data analysis (Csilléry et al., 2010). Approximate Bayesian Computing (ABC) is one of most suitable techniques for handling sparse data and uncertainties due to its likelihood-free nature (Vrugt & Beven, 2018). ABC is a technique for estimating the posterior distribution of model parameters using Bayesian statistics. There are three sampling methods for ABC: (1) rejection sampling, (2) Markov Chain Monte Carlo sampling, and (3) sequential Monte Carlo sampling (Csilléry et al., 2010). An algorithm for ABC is the Differential Evolution Markov Chain algorithm that combines an evolutionary algorithm with Markov Chain Monte Carlo sampling (Braak, 2006). Another algorithm for ABC with Markov Chain Monte Carlo sampling is Adaptive Metropolis (Wöhling & Vrugt, 2011). This algorithm updates the Gaussian distribution for sampling using the information gathered so far in the process. Sadegh and Vrugt (2014) introduce a multi-chain approximate Bayesian computation with Markov Chain Monte Carlo Sampling algorithm, also called Differential Evolution Adaptive Metropolis (DREAM). This sampling method is based on a multi-chain Markov Chain method that uses differential evolution for population evolution with a Metropolis selection rule. Additionally, subspace sampling is applied to enhance search efficiency. It is shown that DREAM is one of the most efficient

sampling algorithms for ABC (Sadegh & Vrugt, 2014).

Another technique in the family of Bayesian inference is Bayesian Optimization (BO). Bayesian optimization techniques are among the few techniques in the field of machine learning that are able to handle small data sets (Jalali et al., 2017). BO is a technique that uses Bayes' theorem to search for the optimum by constructing the posterior distribution. This can either be defined by Gaussian Processes, also referred to as Kriging, or by using the Parzen-Tree Estimator (van Hoof & Vanschoren, 2021). It balances between exploration and exploitation of the solution space based on a Maximum Probability of Improvement, an Expected Improvement, or an Upper Confidence Bound function. The most acquisition function for exploration is Expected Improvement (Jones, 2001; Bischl et al., 2023).

The last family of methods is data assimilation that calibrates models by dynamically incorporating observed data into the model. This is a promising technique for calibrating with sparse data when estimating unobservable states in a running simulation model (Hu & Wu, 2019; Kuipers, 2021). There are three approaches for data assimilation for discrete event simulation models: (1) variational approach, (2) sequential approach, and (3) particle filtering (Xie, 2018). The variational approach chooses a time interval and treats the data within that interval in the same manner to produce estimates of the state variables of the system. The sequential approach assimilates data sequentially over time with the goal to correct the estimated state when a new observation becomes available. It only updates the specific state for the specific time that an observation becomes available. Particle filtering follows the steps of the sequential approach but aims to estimate the conditional distribution of all states up to a user-defined time given all available measurements. A commonly used algorithm for particle filtering is the bootstrap filter algorithm (Xie, 2018).

In this research, we use a deterministic mathematical solver using Powell's Method algorithm as a reference, given it is one of the most commonly used model calibration techniques. Moreover, we compare three model calibration techniques that are most promising in the case of sparse data: (1) a Genetic Algorithm (GA) using the ε-NSGA-II algorithm, (2) Approximate Bayesian Computation (ABC) using the DREAM algorithm, and (3) Bayesian Optimization (BO) using Gaussian Processes with the Expected Improvement function. Our study excludes the family of data assimilation techniques since the focus is not on calibrating real-time (running) simulation models.

4.3.3. MODELING STRUCTURAL UNCERTAINTY FOR CALIBRATION

Evaluating model calibration techniques' performance in the case of sparse data is further complicated by how structural uncertainty is modeled. Uncertainty in the structure of a discrete event simulation model is often implemented by parametrization (Baldissera Pacchetti, 2021). A fully comprehensive model is built, incorporating all potential components and links within specific search ranges. Binary parameters are then utilized to determine the inclusion or exclusion of each component and/or link in the model. Researchers can randomize the values of the binary parameters to include uncertainty in their model runs, or can calibrate these binary parameters to find a structure close to the real world (Baldissera Pacchetti, 2021). However, three

primary drawbacks are encountered in our study when using this implementation: (1) designing a fully comprehensive simulation model is time-consuming and memory heavy, (2) performing model runs or calibrating the model is computationally heavy because of the many decision variables, and (3) designing for interdependencies between the components and links (e.g., no links between suppliers and manufacturers is not realistic in a supply chain) causes many additional modeling rules or optimization constraints (Folkerts et al., 2020). These three drawbacks make it difficult to capture the structural uncertainty in a supply chain simulation model easily.

Another way to include uncertainty in the structure of the simulation model for experiments is model composability (Yilmaz, 2019; Folkerts et al., 2020). This means that multiple distinct system configurations are created by coupling components of the system (e.g., different actors in a supply chain and transport modes in various ways). A system configuration defines the structure of the components in a system and the associated parameters (Folkerts et al., 2020). To describe these components of the system in a simulation model, a referential ontology is used. Referential ontologies, such as Extensible Markup Language, Unified Modeling Language, and System Entity Structure (SES), support the development of models by describing real-world entities (Zeigler & Hammonds, 2007; Hofmann, 2013). In this study, we focus on the ontology framework of SES as it is a powerful framework specifically designed for modeling and (discrete event) simulation (Zeigler & Hammonds, 2007; Tolk et al., 2023).

Zeigler (1984) introduces SES for composing multiple system configurations (e.g., various supply chain structures) for simulation. SES defines a set of system configurations, helpful for generating a set of simulation models for a family of systems. It is represented by a tree structure including entity nodes, descriptive nodes, and attributes (Folkerts et al., 2020). Entity nodes describe an object of the system, e.g., an actor in the supply chain. Descriptive nodes describe the composition among at least two entities using aspect nodes. An aspect node describes the composition of an entity, either physical or non-physical. For example, a PPE manufacturer consists of a production facility, supply inventory, and manufactured product inventory like respirators. A multi-aspect node describes the composition of an entity consisting of many entities of the same type. For example, the set of respirators consists of (many) identical respirators. A specialization node describes the entity's categorization. For example, the respirator can either be certified or not.

The process of deriving a single configuration (e.g., a specific supply chain) of the SES is called pruning. For each single configuration, a specific structure and parametrization is defined. Given the increasing complexity of systems such as a supply chain and many possible system configurations, it is preferred to conduct pruning automatically (Zeigler & Sarjoughian, 2013). Automated pruning for a specific system requires knowledge on the degrees of freedom to ensure valid system configurations and thus, valid simulation models (Folkerts et al., 2020). Each entity and descriptive node has specific rules for composing a valid system. For example, in the case of a supply chain, at least one type of each actor in the SES has to be present in a system configuration. All knowledge and rules necessary for automatic pruning have to be known at the beginning of the pruning process, using scripts or a set of constraints

(Pawletta et al., 2016; Deatcu et al., 2018; Hermans, 2022).

The novelty of this study is that we focus on structural uncertainty for calibration simulation models, instead of most research that only focuses on parametric uncertainty. This study uses SES to examine structural uncertainty in simulation models. This allows us to calibrate a supply chain simulation model using a set of system configurations efficiently based on a theoretical ontology. More explicitly, the contribution of this study is to evaluate the quality-of-fit of various model calibration techniques for identifying the structure of a supply chain with sparse data.

4.4. METHODS

In this research, we examine to which extent a set of model calibration techniques can correctly match the structure of a simulation model for a varying degree of data sparseness. First, the design of experiments using a ground truth simulation is explained. Second, the configuration of the selected model calibration techniques is presented. Third, the formalization and parametrization of the simulation model as a case study is described. Last, the formalization of the stylized system entity structure is presented.

4.4.1. DESIGN OF EXPERIMENTS USING THE GROUND TRUTH

This section presents the design of experiments for our study. First, the ground truth set-up is presented. Second, the quality-of-fit is discussed. Last, the experiments are described.

GROUND TRUTH SET-UP

A ground truth set-up is used to evaluate the performance of the selected model calibration techniques over various degrees of data sparseness. One stylized simulation model acts as a ground truth to produce the observed data of the system, and data is extracted from this model. This set-up allows us to measure the calibration's closeness to the "true" values, which is challenging with real data that inherently has some degree of sparseness (Khondoker et al., 2016).

Figure 4.2 shows the method used for evaluating the model calibration techniques. In this research, we calibrate using the graphs representing the supply chain model to identify the underlying structure. More specifically, we focus on a directed acyclic graph that consists of vertices and edges, i.e., $g = (V, E)$. Vertices represent the actors in the supply chain, meaning the type and number of actors. Edges represent the connectivity between these actors in the supply chain.

First, we define a ground truth simulation model based on the directed ground truth graph, g^o, with vertices, V^o, and edges, E^o. The output of the ground truth simulation model is the *ground truth data*, not including any sparseness. Next, we add data sparseness with a degree of x% to the ground truth data. A degree of x% means that x% of the original data elements have noise, are biased, or are missing values. For example, 10% of the ground truth data elements are transformed into missing values. It is randomly determined which x% of data elements are sparse over the entire data set. We adopt the detailed implementation of randomly assigning sparseness to data on noise, bias, and missing values from van Schilt et al. (2024). This results in *sparse observed data*.

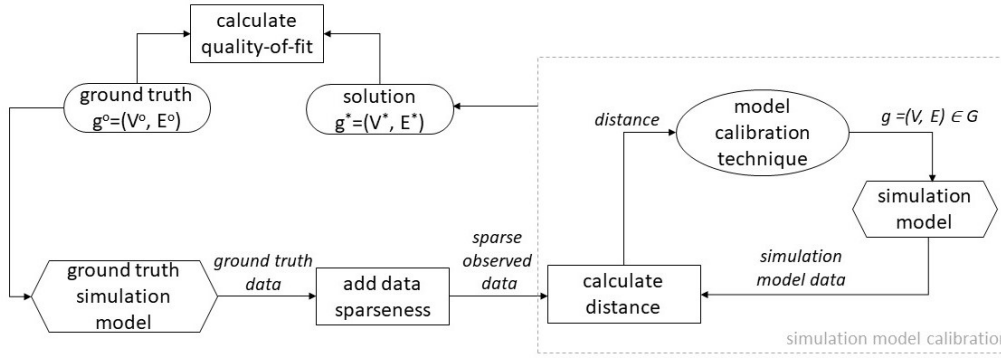


Figure 4.2: Method for evaluating calibration of a simulation model on the graph with sparse data.

When the sparse data is defined, the simulation model calibration process starts. We have a large set of plausible graph configurations of the supply chain under study, with many dependencies between the vertices and edges in such a graph. Thus, we use SES to define a set of plausible graph configurations of the supply chain, G. Each graph in this set, $g \in G$, is a randomly generated directed acyclic graph with vertices and edges, (V, E). A large set of graphs of plausible supply chains is created using SES as input for the model calibration techniques to select candidate solutions. In our study, we use a set of 40.000 randomly generated graphs, balancing between an adequate size for exploration and computational efficiency.

The model calibration technique essentially selects a candidate graph, $g = (V, E) \in G$. The simulation model is run for 5 unique replications based on this candidate graph, resulting in *simulation model data* as output. Next, the distance between the *simulation model data* and the *sparse observed data* is calculated using a distance metric. Based on the resulting distance, the model calibration technique selects a new candidate graph. The process repeats and stops when a stopping criterion is reached, e.g., the number of iterations or a certain number of solutions close to the ground truth. The solution is the graph, $g^* = (V^*, E^*)$, that best describes the structure of the ground truth model according to the calibration technique.

QUALITY-OF-FIT

While model calibration aims to minimize the distance between the simulated and sparse observed *output* data, it does not guarantee that the graph of the calibrated simulation model will be close to that of the ground truth. Therefore, we assess the quality-of-fit of the solution graph and the ground truth graph. Assessing the similarity of graphs is complex, making it challenging and computationally expensive to determine a single metric for evaluating the quality-of-fit (Wills & Meyer, 2020). Thus, we compare the graphs using various feature-based distances of (1) the number of vertices, (2) the number of edges, and (3) average betweenness centrality, i.e., the average fraction of all shortest paths that pass through a vertex. Additionally, a commonly used similarity measure is the graph edit distance (Wang et al., 2021). The graph edit distance defines the cheapest set of graph edit operations (e.g., node inser-

tion, edge deletion) needed to transform one graph to the other graph (Abu-Aisheh et al., 2015). For computational reasons, we use an approximated greedy graph edit distance of Riesen et al. (2015) by transposing this problem to an assignment problem. The python library *GMatch4py*[1] is used.

EXPERIMENTS

The steps in Figure 4.2 outline a single experiment for evaluating the quality-of-fit of a model calibration technique, given a certain degree of data sparseness. We systematically increase the degree of data sparseness added to the *ground truth data* with steps of 10%. Thus, we evaluate for 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%. Additionally, the model calibration techniques are examined for 0% of data sparseness, i.e., *ground truth data*, as a base case. Following the results of the individual dimensions, we analyze a set of experiments in which we combine the dimensions of data sparseness to study the interaction effects.

Each experiment is conducted with 6 seeds to account for the impact of stochasticity on the simulation model calibration outcome. Each seed produces a set of results that are presented individually. For each seed, we first transform *x%* of the data set with sparseness, and then we use this as input for all the model calibration techniques. This means that the exact same observations were transformed in the data set and are provided to the different techniques for simulation model calibration.

4.4.2. CONFIGURATION OF THE MODEL CALIBRATION TECHNIQUES

Recalling the selected model calibration techniques for this research in Section 4.3.2, Powell's Method is considered as a reference technique, while ABC, BO, and GA are identified as suitable options for dealing with sparse data.

A distance metric for the model calibration techniques needs to be defined to minimize the difference between the simulation model data and observed data. Common metrics like mean square error, Kolmogorov-Smirnov, or Euclidean distance are often used, but they may not adapt well to specific problems (Aggarwal et al., 2001; Suárez et al., 2021). Our study requires a metric that considers stochastic models and sparse observed data in complex, high-dimensional systems. According to Mirkes et al. (2020), classic metrics like L1 and L2 are effective for complex, high-dimensional data tasks. Therefore, we use the Manhattan (L1) distance, measuring the sum of absolute differences between data points across all dimensions after normalization.

For calculating the quality-of-fit for the selected model calibration techniques, we compare the ground truth graph, $g^o = (V^o, E^o)$, with the graph of the optimal solution, $g^* = (V^*, E^*)$. The result of Powell's Method, GA, and BO is a single optimal solution of the decision variable; in this case, the value of the graph's index. In contrast, ABC produces an approximate posterior distribution of the indexes of graphs. To obtain one optimal solution of a graph from this resulting posterior distribution, we select the graph's index with the highest frequency, i.e., the mode, for that specific distribution. Thus, the most often accepted graph, obtained by this index value, serves as the optimal solution for ABC.

[1] https://github.com/jacquesfize/GMatch4py

For each technique, a stopping criterion for finding the optimal solution is defined. The stopping criteria for these experiments are based on an empirical analysis on the convergence of the model calibration techniques over 6 seeds. For the reference technique, Powell's Method, we limit the number of function evaluations to 1500 and the number of iterations to 100. For ABC, we use 15.000 draws as the stopping criterion. The analysis shows that there is convergence of ABC determined by the Gelman-Rubin statistics at 15.000 draws for most of the 6 seeds (Gelman & Rubin, 1992). For BO, we use 3750 iterations as a stopping criterion. We use 100 initial points. With this number of iterations, the number of improvements remained constant for every seed. For GA, we use 10.000 function evaluations as a stopping criterion. The analysis shows that with 10.000 function evaluations, the number of improvements is stable across all seeds.

4.4.3. FORMALIZATION OF GROUND TRUTH SIMULATION MODEL

The case study used for the ground truth simulation model is a stylized counterfeit PPE supply chain. Figure 4.3 visualizes the structure of the ground truth counterfeit PPE supply chain simulation model from China to the northeast USA as a graph. The symbols in the figure represent the main actors (vertices of the graph) in the supply chain, and arrows represent the transportation flows (edges of the graph).

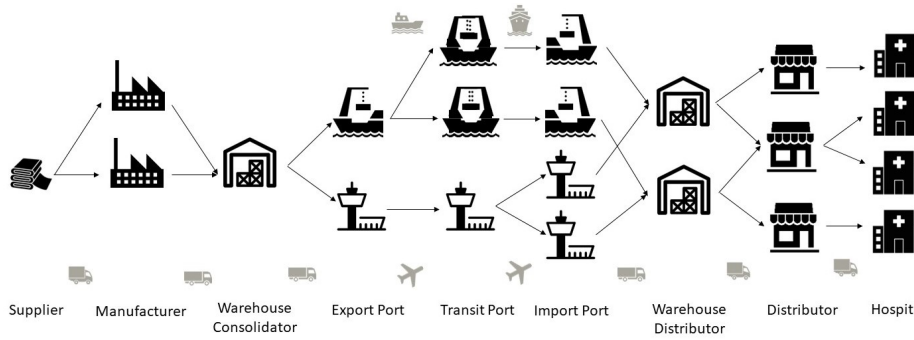


Figure 4.3: Stylized Supply Chain of Counterfeit PPE.

The supply chain starts at the supplier of raw materials, placed in Guangdong, China, who supplies products for PPE such as fabrics. These products are transported overland to one of the two manufacturers in the same area. These manufacturers produce counterfeit PPE in the factory, pack them in boxes, and consolidate them into batches for transportation. Each batch contains a specific quantity of counterfeit PPE, such as 2000 boxes with 20 PPE units per box, resulting in a total of 20,000 PPE units per batch. Next, a batch of finished counterfeit PPE is transported from the manufacturers' location via a truck to the consolidation warehouse close to the border of Hong Kong. Batches from several manufacturers are stored here. We identified two strategies for handling these batches: (1) wait until an order arrives; then the specific order is picked and shipped to the customer, or (2) wait until the stock reaches the level required to directly fill one container. In the ground truth model, we assume that batches are handled based on random order arrivals with an average interarrival

time of 1.2 days. When the orders are picked in the warehouse, they are transported overland by truck to the export seaport or airport in Hong Kong, depending on the mode of transport. For transportation overseas, the batch is loaded into a 40 ft container and transported by a small container ship to the transit port. Upon arrival at the transit port, the small container ship unloads the container carrying counterfeit PPE. At the same port, the container is loaded onto a larger container ship for overseas transport. Depending on the destination port, the route that the container follows is either (1) from Hong Kong to New York, USA via Singapore, or (2) from Hong Kong to Boston, USA via Shanghai, China. For transportation via air, the batch is loaded into the cargo hold of an international airplane using pallets. The destination of this batch is either New York, USA (airport JFK) or Boston, USA (airport BOS). In both cases, there is a transit at Amsterdam Schiphol Airport (AMS), where the batch is moved from one airplane to another. Arriving at the import port, the batch in a container or pallet is unloaded at one of these ports, and waits for inland transport to one of the two (illegal) wholesales distributor in the area of New York, USA or Boston, USA. Here, the batch of counterfeit PPE is equally divided into smaller batches for the two distributors they serve. Small trucks directly transport these smaller batches to distributors in New Hampshire, Connecticut, and New Jersey. Next, the distributors transport the batch to hospitals in Portsmouth, Providence, New Haven, and Philadelphia. When the counterfeit PPE arrive at the hospital, the products are used for medical reasons without knowing that they are counterfeit.

Actors			
Input Parameter	**Distribution**	**Value**	**Unit**
Interarrival time of product at supplier	Exponential	10	days
Time at manufacturer	Gamma	1.5, 0.8	days
Time at warehouse consolidator	Triangular	0.5, 1, 1	days
Time to pickup at warehouse consolidator	Triangular	0.5, 1, 2	days
Probability of counterfeit PPE in shipping container at warehouse consolidator		0.5	
Time at sea ports	Triangular	0.5, 1, 2	days
Time at air ports	Triangular	0.5, 1, 1	days
Waiting time at yard for transport at import sea port	Uniform	0.5, 3	days
Probability of counterfeit PPE extracted at import sea port		0.5	
Waiting time at yard for transport at import air port	Uniform	0.5, 1	days
Probability of counterfeit PPE extracted at import air port		0.5	
Time at warehouse distributor	Triangular	1, 2, 2	days
Time at distributor	Exponential	0.2	days
Time at hospital	Exponential	0.1	days

Table 4.1: Input parameters of actors for the simulation model of the stylized counterfeit PPE supply chain.

Table 4.1 and Table 4.2 show the input parameters for the actors and the links used in the ground truth simulation model. In the simulation model, most uncertainties such as delays of transport modalities and speed of transport modalities follow triangular distributions inspired by real-world data of a fashion retailer and expert interviews (Kuipers, 2021; Hashemi et al., 2022). Table 4.3 shows parametrization of the speed and the delays of the transport modalities for the simulation model of this study.

Links		
Name	**Value**	**Unit**
Supplier to manufacturer 1	50	km
Supplier to manufacturer 2	80	km
Manufacturer 1 to warehouse consolidator	140	km
Manufacturer 2 to warehouse consolidator	75	km
Warehouse consolidator to export sea port	45	km
Warehouse consolidator to export air port	60	km
Export sea port to transit sea port Shanghai	2.8	days
Export sea port to transit sea port Singapore	9	days
Transit sea port Shanghai to import sea port Boston	42.5	days
Transit sea port Singapore to import sea port New York	26	days
Export air port to transit air port Amsterdam	9274	km
Transit air port Amsterdam to import air port Boston and New York	5547, 5847	km
Import sea and air port Boston to warehouse distributor Boston	15, 20	km
Import sea and air port New York to warehouse distributor New York	80, 72	km
Warehouse distributor Boston to distributor New Hampshire, Connecticut	105, 150	km
Warehouse distributor New York to distributor Connecticut, New Jersey	175, 150	km
Distributor New Hampshire to hospital Portsmouth	15	km
Distributor Connecticut to hospital Providence, New Haven	140, 60	km
Distributor New Jersey to hospital Philadelphia	50	km

Table 4.2: Input parameters of links for the simulation model of the stylized counterfeit PPE supply chain.

Time series data is extracted from the simulation model of this specific system configuration as ground truth data. The time series data entails data on when a quantity of PPE arrives at an actor, including the location and the type of actor. For example, a batch with a quantity of 20.000 PPE arrives at the export airport in Hong Kong on day 3. Data of the time series is summed per day, and is aggregated over the actor types that are represented in the SES (see Figure 4.4). Multiple replications are combined using the mean value per day per actor type. A simulation time of 52 weeks with 5 unique replications is used. The simulation model has been developed with the library *pydsol-core* and *pydsol-model* in Python in combination with *networkx*. The library *pydsol* is a Python implementation of the Distributed Simulation Object Library (DSOL), originally implemented in Java (Jacobs, 2005).

Transport modalities							
Input Parameter	**Distribution**	**Value**	**Unit**	**Input Parameter**	**Distribution**	**Value**	**Unit**
Speed of small truck	Triangular	0, 100, 120	km/h	Delay of small truck	Triangular	0, 0.2, 0.5	days
Speed of large truck	Triangular	0, 80, 120	km/h	Delay of large truck	Triangular	0, 0.5, 1	days
Speed of train	Triangular	25, 40, 75	km/h	Delay of train	Triangular	0, 0.3, 0.5	days
Speed of feeder	Triangular	10, 18, 25	knots	Delay of feeder	Triangular	0, 4, 16	days
Speed of vessel	Triangular	10, 18, 25	knots	Delay of vessel	Triangular	0, 7, 16	days
Speed of airplane	Uniform	740, 930	km/h	Delay of airplane	Triangular	0, 1, 4	hours

Table 4.3: Input parameters of speed and delay of the transport modalities for the simulation model of the stylized counterfeit PPE supply chain.

4.4.4. FORMALIZATION OF SYNTHETIC STRUCTURAL UNCERTAINTY

In our research, we use a stylized SES to incorporate structural uncertainty for designing the set of plausible simulation models. All configurations of the simulation models result from the SES, including the ground truth model which is one specific configuration. Figure 4.4 presents the SES of the counterfeit PPE supply chain simulation model. The tree starts with a supply chain with multiple actors, shown by the physical multi-aspect node. A discrete event simulation model of a supply chain has the following model elements: source (i.e., creating entities), server (i.e., processing entities), sink (i.e., destroying entities), and links to connect these elements (Banks, 1998). A supplier acts as a source in the simulation model, as the supply chain starts here. The sink describes the end of the supply chain with the type (export) customer. There are multiple types of servers, as indicated by the specialization node. Any actor that processes entities, in this case PPE products, is a server. There are three types of ports described in the SES of the stylized supply chain case: import port, transit port, and export port. Moreover, a supply chain has links to connect the actors. This SES includes two links: a sea link based on the travel time overseas, and a link for land and air transport based on the distance.

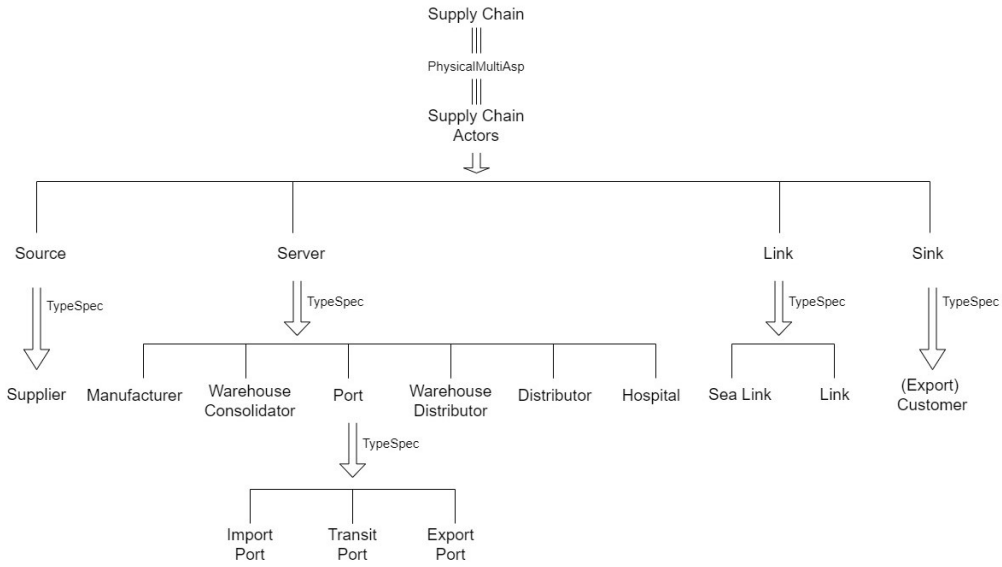


Figure 4.4: System Entity Structure of the Counterfeit PPE Supply Chain Simulation Model.

For composing system configurations from the SES, specifying rules and constraints have to be set. One important rule for our case is to have at least one representative of each actor type in the supply chain that is arranged in a specific sequence and interconnected through links. For example, a supplier has to be connected to a manufacturer, who in his turn has to be connected with a warehouse consolidator. This determines the incoming and outgoing degrees of each actor. Also, counterfeit PPE is commonly shipped across borders using routes that align with the legal sup-

ply chains. Hence, the travel time and transit time of international transport overseas (i.e., sea links) is based on open-source data of the shipping schedules given by MSC, Maersk, HMM, and Evergreen. Through the distinct shipping alliances that these four companies are part of, we gain a comprehensive understanding of the schedules of the leading shipping firms. A distribution is fitted for the travel time of each leg for the seaports (port-to-port) and the scheduled processing times at the transit ports based on four months of schedules in 2023 and 2024. The airport network is based on an open-source flight route database from 2017. The travel distance between the airports is calculated using the Haversine distance between two airports (Robusto, 1957). The distance for links over land relies on expert interviews. For this, we use the information from open-source data to identify the real-world locations of ports, and from there, we determine the positions of other actors based on expert information. For example, the warehouse locations are often driving distance from the ports. See Appendix A.1 for more details on the specifying rules and distances.

For our study, we randomly generate a set of graphs using the synthetic SES. First, the number of vertices is randomly determined using the SES, relying on the actor type and the corresponding constraint on the number of vertices per type. Second, the edges are randomly defined based on the connectivity between the vertices, i.e., the amount of incoming and outgoing degrees. Third, we use the open-source data on ports to determine the international routes and their travel time. The generated export and import ports are randomly assigned to real-world locations. Given the edges between the export and import ports, plausible real-world routes between these ports are identified. Fourth, the travel distance between the other vertices are defined, e.g., between suppliers and manufacturers. The distance per edge is chosen using a Uniform distribution of the minimum and maximum distance between actors. Last, the graphs are sorted on their average betweenness centrality, i.e., the fraction to which a vertex lies on the shortest path between other vertices averaged over all vertices. In terms of illicit networks, the betweenness centrality of an actor determines the centrality of an actor in the network, e.g., an actor with a high betweenness centrality often has a broker position (Morselli, 2010; Diviák et al., 2019). We use the average betweenness centrality of the graph as a descriptor since it is shown to be the most effective topology for planning interventions (Cavallaro et al., 2020; Ficara et al., 2021).

4.5. RESULTS

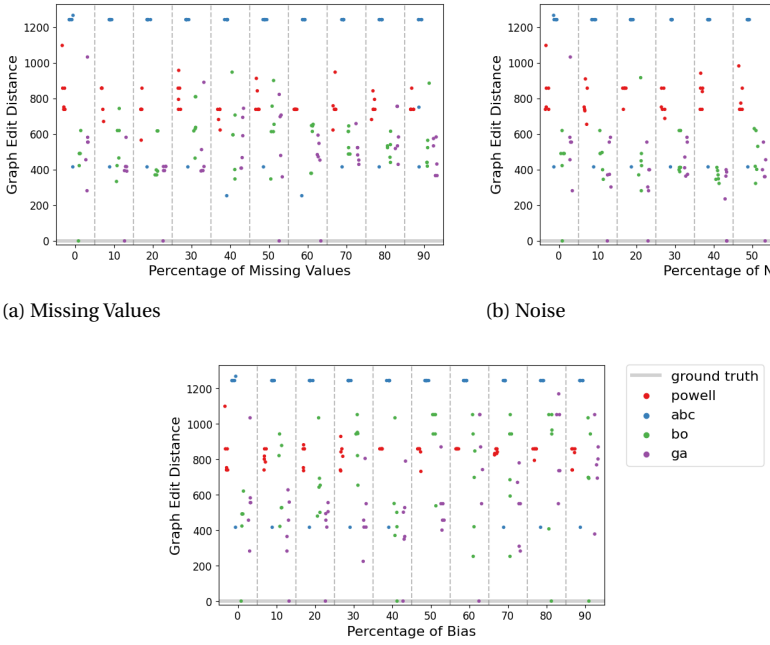
We discuss the results of the model calibration techniques when varying the degree of data sparseness, both individually per dimension and in combination. The results are presented in a scatter plot where each point represents the optimal solution arising from the model calibration technique for one unique seed.

4.5.1. ANALYSIS OF THE INDIVIDUAL DATA SPARSENESS DIMENSIONS

This section presents the analysis of the various metrics for quality-of-fit for the four selected model calibration techniques when increasing the degree of the data sparseness dimensions individually. We discuss the metric of graph edit distance, the ranking of the average betweenness centrality, and the number of vertices and edges. For more results on the average betweenness centrality and the Manhattan distance, see Appendix A.2.

GRAPH EDIT DISTANCE

The graph edit distance measures the difference between the ground truth graph and the optimal graph chosen by the model calibration technique based on graph edit operations (Abu-Aisheh et al., 2015). A graph edit distance of zero indicates that the optimal graph and the ground truth are identical. Hence, the lower the graph edit distance, the higher the quality-of-fit. Figure 4.5 shows the graph edit distance of the optimal solutions of the model calibration techniques for six unique seeds per percentage of missing values, noise, and bias.



(a) Missing Values

(b) Noise

(c) Bias

Figure 4.5: Graph Edit Distance per Dimension of Data Sparseness for Powell's Method, ABC, BO, GA

Powell's Method demonstrates consistency in finding optimal solutions for missing values, noise, and bias. The graph edit distance for each solution is between 623 and 1098 across all percentages of data sparseness. Hence, the found solutions need a relatively high number of graph edit operations to transform into the ground truth. The spread of the solutions in terms of graph edit distance is relatively small compared to other techniques. However, Powell's Method fails to identify the ground truth successfully.

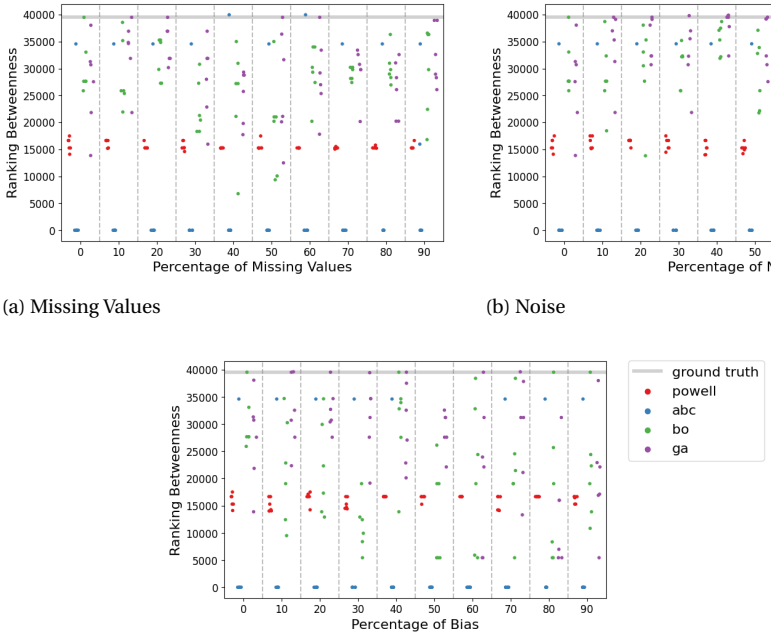
Similar to Powell's Method, ABC shows consistency in terms of graph edit distance for missing values, noise, and bias. The graph edit distance for most solutions is either 416 or 1243 graph edit operations across most percentages of missing values and all percentages of noise and bias. At 40% and 60% missing values, Figure 4.5a shows an outlier with a graph edit distance of only 254 node operations from the ground truth. Nevertheless, ABC fails to identify the ground truth.

BO has a more diverse graph edit distance of the solutions across the various percentages of data sparseness, but the majority of the graph edit distances still lies between 400 and 707 operations. For a data sparseness of 0%, BO identifies the ground truth represented by a graph edit distance of 0. With 10%, 20%, 40% and 60% missing values, BO finds optimal solutions that have a relatively low graph edit distance between 334 to 400 edit operations. However, the ground truth is not identified for any percentage of missing values. Comparatively, BO does identify the ground truth at 80% noise. Additionally, the differences in graph edit distance between the solutions for each percentage of noise are less spread out. Especially at 40%, 80%, and 90% noise, all solutions have a relatively low graph edit distance between 323 to 413 graph edit operations. For bias, we see in Figure 4.5c that BO finds optimal solutions with a relatively high graph edit distance compared to missing values and noise for each percentage of data sparseness, meaning more solutions with 942 to 1051 graph edit operations from the ground truth. In contrast, the ground truth is identified most frequently for bias at 40%, 80%, and 90%.

GA shows a diverse graph edit distance of the solutions across the various percentages of data sparseness, where we observe an identification of the ground truth as well as relatively high graph edit distances. Other solutions have graph edit distances in the range of 236 to 801 graph edit operations. For 0% data sparseness, GA finds an optimal solution with a relatively high graph edit distance of 1033 graph edit operations. The ground truth is not found for 0% of data sparseness. GA identifies the ground truth for 10%, 20%, 50%, and 60% of missing values. For noise, GA is able to identify the ground truth most often as visible in Figure 4.5b. The ground truth is identified for the majority of noise percentages, excluding 30% and 60%. Especially for 40% and 90% of noise, the other solutions that are identified are in close proximity to each other with a graph edit distance between 236 and 456 graph edit operations. Overall, the graph edit distance for noise stays limited to 582 edit operations. Next, GA identifies the ground truth at 10%, 20%, 40%, and 60% of bias. Solutions with a relatively high graph edit distance, i.e., of 869 and 1051 operations, are found after 60% bias. This means that, for bias, GA found more complex graph structures that explain the sparse data compared to missing values and noise.

RANKING OF AVERAGE BETWEENNESS CENTRALITY

The graphs in the set for calibration are ranked based on their average betweenness centrality (as described in subsection 4.4.1), where higher rankings correspond to higher average betweenness centrality scores. The ground truth graph ranking is 39520, with a ranking from 0 to 40000, and the average betweenness centrality is 0.046, with a range of 0.003 to 0.110. Figure 4.6 presents rankings for all dimensions of data sparseness across the four techniques. Results on the exact values of the average betweenness centrality are provided in Appendix A.2.



(a) Missing Values

(b) Noise

(c) Bias

Figure 4.6: Ranking of Average Betweenness Centrality per Dimension of Data Sparseness for Powell's Method, ABC, BO, GA

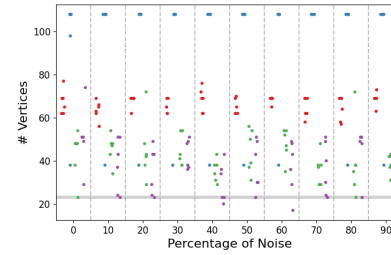
Figure 4.6 shows that Powell's Method consistently identifies a few specific graphs as optimal solutions over the increasing percentage of missing values, noise, and bias. All optimal graphs of Powell's Method are found in a specific area of the set of graphs with a ranking of around 15000 and an average betweenness centrality of 0.01. However, the solutions are not close to the ground truth, based on their ranking nor on their average betweenness centrality. The solutions of Powell's Method typically have minimal overlap with the solutions of the other techniques, especially for the dimension noise (see Figure 4.6b).

For ABC, we see that this technique reaches two optimal graphs consistently over the increasing percentage of missing values, noise, and bias. One optimal graph with a ranking of 0 is quite distant from the ground truth, while the other is in closer proximity with a ranking of 34589. For missing values, Figure 4.6a displays that the outliers of 40% and 60% of missing values have a ranking of betweenness of 39999, seemingly in close proximity to the ground truth. Their graph edit distance of 254 graph edit operations indicates that the graph is indeed close to the ground truth. Except for 70% noise and for 50% and 60% bias, where only one optimal solution is discovered, the two graphs consistently emerge as optimal solutions across increasing percentages of data sparseness.

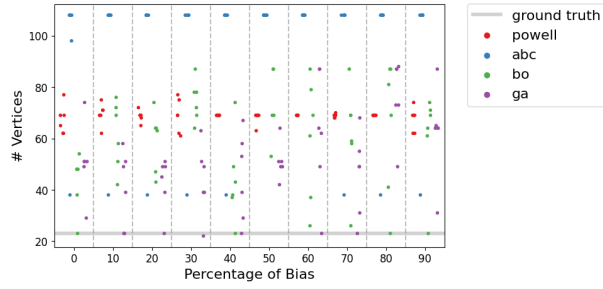
Comparatively, BO finds a variety of graphs for missing values, noise, and bias when looking at the betweenness ranking. For missing values, the diversity of solutions increases when the percentage of missing values increases up to 50% of missing values. Especially at 70% and 80% of missing values, the optimal graphs are relatively close to each other in terms of average betweenness centrality. For noise, the optimal graphs align closely with the ground truth in terms of betweenness ranking but they display a wide spread in average betweenness centrality. For bias in Figure 4.6c, BO identifies a diverse set of graphs based on betweenness ranking as optimal solutions, yet most frequently finds the ground truth.

GA also identifies a variety of graphs as solutions for missing values, noise, and bias based on the ranking of average betweenness centrality. As missing values increase to 50%, the diversity of solutions increases, but decreases beyond that point. This indicates a closer proximity in solutions, especially at 70% missing values. For noise, Figure 4.6b shows that the optimal solutions are close to the ground truth for betweenness ranking. Yet, their average betweenness centrality is relatively high and wide spread. For bias, GA identifies diverse optimal graph solutions in terms of ranking of betweenness and average betweenness centrality. As bias increases, the diversity of solutions increases, meaning a higher percentage of bias leads to a more diverse set of optimal graph structures.

(a) Missing Values



(b) Noise



(c) Bias

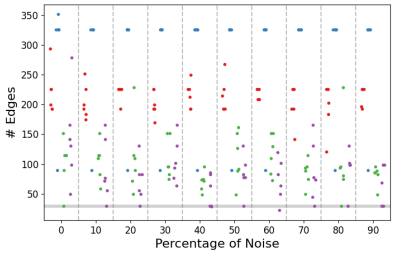
Figure 4.7: Number of Vertices per Dimension of Data Sparseness for Powell's Method, ABC, BO, GA

VERTICES AND EDGES

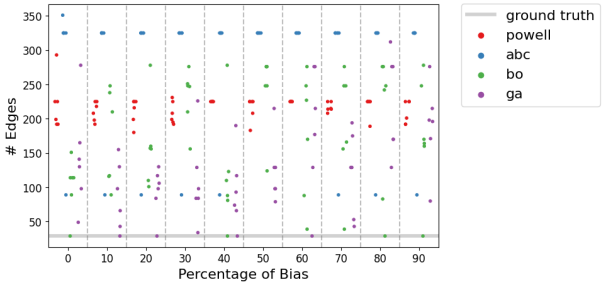
Figure 4.7 and Figure 4.8 show the number of vertices and edges of each solution graph per dimension of data sparseness for the four techniques. The ground truth graph contains 23 vertices and 29 edges.

For all three dimensions of data sparseness, Powell's Method identifies optimal graphs within a distinct range of vertices and edges, specifically between 58 to 78 vertices and between 130 to 293 edges. Particularly for noise, the range of vertices and edges has a minimal overlap with other techniques. Next, ABC identifies two optimal solutions for most percentages of data sparseness: one graph with 108 vertices and 325 edges, and the other graph with 38 vertices and 89 edges. For the outliers at 40% and 60% missing values, an optimal graph is discovered with 21 vertices and 31 edges, with a close proximity to the ground truth. The optimal solutions of ABC have limited overlap with the other techniques in terms of vertices and edges.

(a) Missing Values



(b) Noise



(c) Bias

Figure 4.8: Number of Edges per Dimension of Data Sparseness for Powell's Method, ABC, BO, GA

BO identifies optimal solutions with a varying number of vertices and edges for missing values, noise, and bias. These solutions generally have higher vertex and edge counts compared to the ground truth, meaning a richer structure that is coherent with the sparse data. Figure 4.7a shows that the number of vertices increases up to 30% of missing values. For noise, BO identifies graphs with vertex and edge counts that are closer to, yet still higher than, the ground truth. For bias, Figure 4.7c and 4.8c show that BO discovers a diverse range of graphs in terms of vertices and edges. The vertices and edges of the optimal solutions of BO overlap mostly with those of GA.

GA identifies optimal solutions with vertex and edge counts both lower and higher than the ground truth across different percentages of noise and bias. For missing values, GA only finds graphs that have higher vertex and edge counts than the ground truth. For noise, GA discovers graphs closer to the ground truth in terms of vertices and edges, also identifying optimal solutions with fewer vertices and edges. Figure 4.7b shows that GA typically finds an optimal graph with around 50 vertices for each noise percentage except 40%. For bias, GA identifies a wide variety of graphs with a varying number of vertexes and edges, with diversity notably increasing after 50% bias.

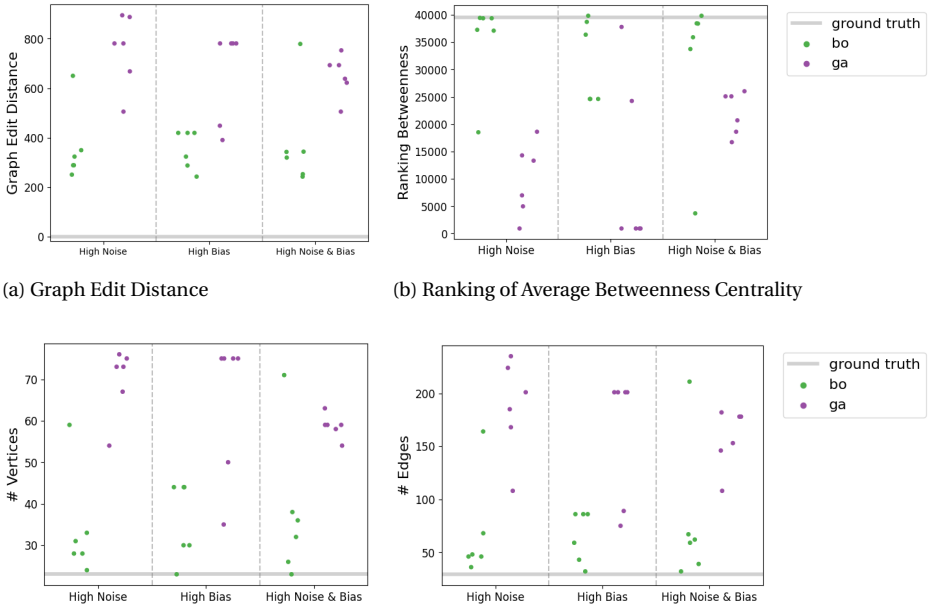
4.5.2. ANALYSIS OF COMBINATIONS OF DATA SPARSENESS DIMENSIONS

Having analyzed the dimensions of data sparseness separately, we evaluate to what extent the model calibration techniques can still reconstruct the supply chain structure when combining the different dimensions of data sparseness. For this, we use the two best-performing techniques of individual analysis: BO and GA. The model calibration techniques are evaluated using scenarios where the dimensions of data sparseness are combined. Tabel 4.4 presents the percentage of data sparseness per scenario. In the scenarios, we use percentages of noise and bias of 20% and 80%, since BO most frequently identifies the ground truth with bias, and GA with noise. The goal is to see whether these techniques can still find the ground truth when a combination of the various dimensions of data sparseness is added.

Scenario	**Noise**	**Bias**	**Missing Values**
High Noise	80%	20%	25%
High Bias	20%	80%	25%
High Noise & Bias	80%	80%	25%

Table 4.4: Configuration of Scenarios

Figure 4.9 shows that BO and GA fail to identify the ground truth for each scenario. We see in Figure 4.9a that BO has a relatively low graph edit distance between 242 and 419 edit operations in all three scenarios, with two outliers with 650 and 779 operations, respectively, for the scenario High Noise and for the scenario High Noise & High Bias. The graph edit distance of GA is relatively high with a range of 390 to 895 operations. The scenario High Noise has the highest graph edit distance, meaning GA finds solutions that have denser graph structures. In line with this, Figure 4.9b shows that the solutions of BO are in close proximity to the ground truth in terms of the ranking of average betweenness centrality. Yet, GA identifies a diverse set of solutions relatively distant from the ground truth, and it has a lower average centrality betweenness. Also, Figure 4.9c and Figure 4.9d illustrate that BO identifies optimal graphs in terms of the number of vertices and edges, closer to the ground truth, whereas GA tends to be more distant from the ground truth. Especially for the scenario of High Noise and the scenario of High Bias, the solutions of BO and GA are distant from each other. For all three scenarios, BO and GA identify optimal solutions within distinct subsets of the graph features. A pair plot is presented in Appendix A.2 for a more detailed picture of the difference between BO and GA.



(a) Graph Edit Distance

(b) Ranking of Average Betweenness Centrality

(c) Number of Vertices

(d) Number of Edges

Figure 4.9: Results of the Experiments for BO and GA

4.6. DISCUSSION

This section reflects on the results of the model calibration techniques and discusses the limitations of our study.

4.6.1. REFLECTION ON MODEL CALIBRATION TECHNIQUES

The results show that Powell's Method and ABC are not suitable techniques for calibrating the structure of a supply chain simulation model with sparse data. Similar to van Schilt et al., 2023, Powell's Method gets stuck in a local optimum instead of reaching the global optimum. We see this in our results as Powell's Method is consistent in identifying a specific range of graphs for all features. Also, ABC consistently identifies two optimal solutions that deviate from the ground truth across most percentages of each data sparseness dimension. A possible explanation is ABC gets stuck in local optima for two reasons: (1) the algorithm results in a bimodal distribution, meaning multiple regions of the input space lead to optimality, and (2) the algorithm does not mutate fast enough, hindering its ability to reach a global optimum. Moreover, despite the different impact of data sparseness dimensions on Manhattan distance (Appendix A.2), Powell's Method and ABC consistently generate similar outcomes across varying percentages of missing values, noise, and bias.

BO and GA are suitable techniques for calibrating the structure of a supply chain simulation model with sparse data. Both techniques result in a diverse set of optimal solutions with various graph edit distances when increasing the degree of sparseness individually. The diverse set of optimal solutions for BO and GA largely overlap. BO can identify the ground truth, especially for a high percentage of bias. The main property of BO is that it uses a Gaussian process to model the distribution of the unknown objective function while balancing exploration and exploitation, making it efficient and relatively fast for sparse data situations (Jalali et al., 2017). However, especially in a non-linear decision space such as calibrating discrete event simulation models and having a combination of data sparseness, distinguishing between identifying promising regions and exploring uncertain regions may not be straightforward. Next, GA is the only technique that successfully identifies the ground truth for all dimensions of data sparseness, especially for noise. Through the population-based nature of GA and the use of evolutionary operators such as crossover and mutation, GA is able to cope with noise, and other dimensions of data sparseness relatively well since the algorithm relies on the properties of the population rather than the individual evaluations (Slowik & Kwasnicka, 2020).

Additionally, a reason for GA outperforming BO could be the implementation of the decision variables in the algorithms. BO tries to fit a continuous distribution that only works with floating point numbers, whereas the decision variable requires integers to rank the graphs. In BO, we rounded the floating point number to integers, leading to possible smaller steps taken by the algorithm and fewer evaluations of unique solutions. In contrast, GA allows for the direct use of the integers, allowing for taking larger steps and more exploration.

Although BO and GA are suitable when varying individual dimensions of data sparseness, they both fail to identify the ground truth when combining these dimensions. In contrast to the individual analysis, the results of the combined scenario show a minimal overlap between the optimal solutions of BO and GA. Further research is needed to investigate the extent to which model calibration techniques can cope with the combination of data sparseness dimensions for accurately identifying the ground truth.

To obtain a comprehensive overview of the various graphs approximating the ground truth with sparse data, we advise using a combination of the results of BO and GA. For further work, it would be interesting to develop an algorithm that combines the exploration and exploitation using the Gaussian process of BO with the population-based approach of GA, and evaluate the suitability for calibrating the structure with sparse data.

In both individual analyses and scenarios, BO and GA identify graphs with high graph edit distances, which often indicates significantly more vertices and edges than the ground truth. Having more vertices and edges suggests a more dense structure of a graph and more complexity. This complexity carries a risk of overfitting as it allows the simulation models of the dense graphs to reproduce the sparse data better than those that slightly differ from the ground truth. Denser graphs, then, seem optimal for the model calibration techniques. For example, with noise and bias, the simulation models of the dense graphs fill in the gaps created by the data sparseness, and

for missing values, "anything goes". Especially in a highly rugged fitness landscape, typical for discrete event simulations, BO and GA favor these dense structures over those closer to the ground truth.

To address this, we shift towards a different direction in terms of the metric of calibration, i.e., match the simulated data to the ground truth data. One possible approach is to limit the likelihood of dense graphs being considered optimal by penalizing denser structures in the objective of the calibration or by restricting the degrees of freedom during the model calibration. In the case of illicit supply chains, the maximum number of vertices and edges of the graph can be limited to ensure less dense graphs can be identified as optimal. However, the dense graphs found by BO and GA do not necessarily lead to "wrong" results since they could explain the sparse data. Restricting the degrees of freedom for model calibration could lead to an unrealistic and (too) narrow view of the potential structures of the supply chain.

Another approach is to embrace and control the diversity of simulation models that explain the sparse data. Instead of a model calibration technique leading to a single optimal solution distant from the ground truth, we want a diverse set of optimal solutions ranging from those relatively close to the ground truth to those further away. For real-world illicit supply chains, this diverse set of plausible structures that could explain the sparse data, including dense structures, is realistic. For example, a dense structure with more actors and routes spreads risk effectively, but it also increases vulnerability to detection since more actors are involved (Morselli, 2010; Anzoom et al., 2021). Further research should focus on finding a diverse set of optimal solutions in terms of simulation model calibration with sparse data.

4.6.2. LIMITATIONS

Designing a large set of graphs using System Entity Structures (SES) as input for the model calibration techniques comes with limitations. First, the set of graphs is merely a sample representation of potential structures, and the set is not exhaustive. This could lead to an overrepresentation of certain configurations of supply chains based on constraints. In this research, it results in many generated graphs in the set having more vertices and edges than the ground truth. Second, the constraints of the SES are chosen by the user. In our study, we use a known ground truth to inform constraint selection. However, in real-world situations where the ground truth is unknown, setting constraints can be difficult. Different perspectives result in varying constraints, affecting the outcomes and optimal solutions of model calibration techniques (Hermans, 2022). Hence, incorporating diverse perspectives is crucial for modeling structural uncertainty. Despite the limitations, SES remains a powerful method for describing structural uncertainty within complex models, like an illicit supply chain model, to approximate the ground truth.

Another limitation of this research is the assumption that the exact percentage of the dimensions of data sparseness is known, whereas these are often unknown in real life. Irrespective, the type and degree of dimensions of data sparseness can be determined based on the characteristics of a real-world supply chain. For example, criminals try to hide data and overrepresent outdated information on their operations as much as possible, resulting in a high degree of missing values and bias.

4.7. CONCLUSION

This research addresses the question: *"To what extent can various model calibration techniques reconstruct the underlying structure of an illicit supply chain when varying the degree of data sparseness?"* We evaluate the quality-of-fit of a reference technique, Powell's Method, and three model calibration techniques that promise to be able to handle sparse data: Approximate Bayesian Computing (ABC), Bayesian Optimization (BO), and Genetic Algorithms (GA). For this, we use a case study of a counterfeit PPE supply chain as ground truth, and formalize structural uncertainty with System Entity Structures (SES). Our analysis shows that:

- SES is a powerful approach for defining structural uncertainty in a supply chain simulation model to approximate the ground truth using calibration. Incorporating diverse perspectives of users on the system is crucial for modeling structural uncertainty.
- Powell's Method and ABC fail to reconstruct the underlying structure of an illicit supply chain for any dimension of data sparseness. These algorithms result in local optima instead of global.
- GA and BO are suitable for reconstructing the underlying structure of an illicit supply chain for a varying degree of data sparseness individually. For a comprehensive understanding of the various graphs approximating the ground truth, we recommend combining the results of BO and GA.
- Denser graph structures, i.e., more vertices and edges, tend to describe and reproduce the sparse data the best. Many optimal solutions from the model calibration techniques are, therefore, distant from the ground truth but are not necessarily incorrect. We highlight the need for identifying a diversity of solutions that are optimal with sparse data, instead of only one.

Reconstructing the underlying structure of an illicit supply chain helps to get insight into the operations of criminals, and it allows law enforcement agencies to effectively plan their interventions.

Further work is needed to investigate the extent to which model calibration techniques can cope with a combination of the dimensions of data sparseness. Future studies should focus on developing an algorithm that combines BO and GA, and evaluating generalizability to various types of supply chains. Additionally, further research should focus on incorporating the diversity of graphs that are coherent with sparse data for analysis and measuring the quality-of-fit of model calibration techniques.

MEXICO

5

A SIMULATION-BASED APPROACH FOR RECONSTRUCTING A DIVERSE SET OF SUPPLY CHAIN MODELS WITH SPARSE DATA USING A QUALITY DIVERSITY ALGORITHM

Data on supply chains is often sparse due to reluctance among actors to share their data, making supply chain simulation modeling difficult. Particularly, supply chain simulation models suffer from parametric and structural uncertainties as a result of this data sparseness. When calibrating a simulation model, there is a large variety of plausible simulation models that could explain the sparse observations about the real-world supply chain. Constructing this diverse set is not an easy task. A relatively unknown approach to generating this diverse set of plausible models is the Quality Diversity (QD) algorithm. This study evaluates the feasibility of using QD to generate a diverse ensemble of supply chain simulation models for a varying degree of data sparseness. The results show that QD is able to generate a diverse ensemble of supply chain models, including the ground truth. As expected, QD successfully identifies the structure of the ground truth most frequently at 0% of data sparseness. When more data sparseness is present, QD is prone to overfitting on more complex structures. Additionally, this study highlights the importance of gathering information on the upstream supply chain. Further research should focus on reviewing the calibration metric for sparse data, and evaluating the effectiveness of the diverse ensemble of plausible supply chain configurations for identifying robust interventions.

This chapter is submitted as: van Schilt, I. M., Kwakkel, J. H., Mense, J. P., & Verbraeck, A. (2024) A simulation-based approach for reconstructing a diverse set of supply chain models with sparse data using a quality diversity algorithm.
The code is available at https://github.com/imvs95/quality_diversity_simulation.

5.1. INTRODUCTION

During COVID-19, there was a steep rise in demand for Personal Protective Equipment (PPE), such as goggles, gloves, face masks, and respirators (Omar et al., 2022). Many countries suddenly needed a large supply of medical PPE to protect caretakers in hospitals. This high demand, unfortunately, also led to an opportunity for the illicit market to produce and distribute counterfeit PPE (Hashemi et al., 2022). Fraud-involved organizations were, for instance, selling non-medical PPE as medical PPE with a high profit margin, leading to unacceptable health risks (Ippolito et al., 2020; Hashemi et al., 2023). Due to the lack of historical data on COVID-19 and criminals trying to obfuscate their data as much as possible, the illicit activities and related logistics of these organizations selling fraudulent products were mostly invisible (van Schilt et al., 2023). This made it difficult for law enforcement to intervene effectively, seize counterfeit PPE, and stop crime.

Simulation is a way to understand a system and to measure the effectiveness of interventions by modeling the system's behavior over time (Banks, 1998; Zeigler et al., 2018). This chapter focuses on discrete event simulation models to represent illicit supply chains (Schmitt & Singh, 2009; Magliocca et al., 2022). A discrete event simulation model consists of components, their parameters, and their behavior over time, as well as the relations between the components. For example, a simulation model of a supply chain consists of actors with parameters such as the time to process a product, inventory levels, and transportation times. The relations define the connections between the actors to represent the network, i.e., the structure, of the supply chain. Calibration uses data to tune the parameters of a simulation model in such a way that the model behavior sufficiently matches the system behavior in the real world (Wigan, 1972; Ören, 1981; Hofmann, 2005).

However, data on supply chains, such as demand, inventory levels, processing times, or transportation times, is often sparse due to reluctance among supply chain actors to share their (correct) data (Somapa et al., 2018; van Schilt et al., 2024). This reluctance has various causes, such as competition, high data cost, or illicit activities in the case of fraudulent supply chains (Ficara et al., 2021). Given that data is sparse, there is a high probability of equifinality during the calibration process, i.e., different input values can result in the same outcome (van Schilt et al., 2023). More specifically, there are multiple versions of the supply chain simulation model that explain the sparsely observed data. Especially for illicit supply chains, there are many possible involved actors, many steps in the process, a large variety of modi operandi, and many possible transport routes (Duijn et al., 2014; Anzoom et al., 2021). Combining this wide variety of possibilities to carry out criminal activities with data sparseness, calibration can result in a large variety of plausible simulation models to describe a real-world illicit supply chain.

Thus, a system with sparse data, such as an illicit supply chain, cannot be fully explained by a single theory or model but needs a variety of models (Mitchell, 2002; Veit, 2020). This is in line with the many-model thinking approach of Page (2021) that emphasizes the need for an ensemble of models to understand and analyze a complex system. More specifically for illicit supply chains, van Schilt et al. (2023) note that choosing one supply chain configuration could lead to a "wrong" view on the crim-

inal supply chain, and hence, making "wrong" decisions on interventions. Accordingly, identifying effective interventions in a system with sparse data, such as a counterfeit PPE supply chain, means evaluating the robustness of interventions over an ensemble of models, i.e., multiple plausible explanations of the real world (Marchau et al., 2019). Many calibration algorithms that can generate multiple model configurations often converge to similar solutions, as configurations with parameters that differ slightly from the best-matching solution typically outperform those with vastly different parameters (Page, 2021). Finding a *diverse* set of, say, 20 solutions during calibration is, therefore, harder than finding the top-20 best solutions, yet a diverse set is more desirable in terms of robustness (Durán & Formanek, 2018).

An approach for generating a diverse set of plausible models is Quality Diversity (QD) algorithms. QD algorithms use evolutionary concepts to find optimal solutions at multiple points of the user-defined search space (Mouret & Clune, 2015; Chatzilygeroudis et al., 2021). It is mostly used in the field of robotics and reinforcement learning (Pugh et al., 2015; Lim et al., 2022; Tjanaka et al., 2023). Recent work applied QD for multi-objective optimization (Pierrot et al., 2022), hyperparameter tuning of a machine learning model (Schneider et al., 2022), and for identifying the most preferred solutions to decision-makers (Kent & Branke, 2023). However, QD remains unexplored in many other application areas, since it is a relatively new approach for evolutionary computation (Pugh et al., 2015; Schneider et al., 2022). More specifically to our study, it has not been researched yet whether QD algorithms can generate a diverse set of plausible configurations of supply chain simulation models characterized by sparse data. This raises the question whether the simulation model configurations proposed by the QD algorithms for such data-sparse supply chains align with plausible real-world supply chain configurations.

Therefore, we evaluate the feasibility of a QD algorithm for generating a diverse ensemble of supply chain simulation models when varying the degree of data sparseness. First, we review related work on generating a diverse set of simulation models that can be calibrated with sparse data, and on QD algorithms. Next, we assess the feasibility of a QD algorithm for generating a diverse set of supply chain configurations that can explain the observed (sparse) data. To test the approach, we use a ground truth simulation model of a synthetic counterfeit PPE supply chain. For the analysis, we extract data from the ground truth model and vary the degree of data sparseness. Next, we assess whether the QD algorithm can generate the ground truth as a feasible solution among its diverse set of solutions. Hence, our study offers a first insight into the potential of using QD algorithms to generate an ensemble of diverse and plausible configurations of simulation models, particularly for supply chains with sparse data. Such an ensemble of plausible configurations, in turn, enables decision-makers to make more robust decisions on, for example, interventions that are effective for the *ensemble* of supply chain models, rather than for a single explanation of the observed data.

The chapter is structured as follows. Section 5.2 presents the relevant state-of-the-art literature. Section 5.3 describes the method for evaluating the results of the QD algorithm, and the used case study. Section 5.4 shows the results of the QD algorithm when applying it to simulation models for sparse data. Section 5.5 discusses our results. Section 5.6 concludes our work and presents further research.

5.2. LITERATURE REVIEW

This section presents the current state-of-the-art for our research. First, we show the related work on simulation with sparse data. Second, we position our chapter in relation to the literature on a pluralist view on simulation modeling. Third, we examine the related work on QD and provide more insight into the QD algorithm itself.

5.2.1. SIMULATION WITH SPARSE DATA

Simulating a real-world system becomes challenging when data is sparse (Srikrishnan & Keller, 2021). Data sparseness makes it more difficult to accurately mimic the behavior of the real world in a simulation model (Vanbrabant et al., 2019; Anzoom et al., 2021; van Schilt et al., 2023). Few studies have examined simulation models calibration to the real world in the context of data sparseness.

One of the first authors to explicitly address the calibration of a simulation model under data sparseness is Liu et al. (2017). They propose a simulation-optimization approach to automatically tune the parameters of an agent-based simulation model with sparse data using an emergency department as a case study. The problem is formulated as a series of local minimum search problems. Vanbrabant et al. (2019) present a framework for assessing real-world input data quality problems for emergency department simulation models. Next, De Santis et al. (2022) focus on the calibration of a discrete event simulation model under data sparseness. Observable values from the real-world system are used to determine the parameter values of the simulation model, for example, values for time intervals. de Groot and Hübl (2021) use calibration as a validation method; in their case, the sparseness of data makes validating the simulation model challenging. They manually fine-tune the parameters and dynamics of the model to enhance validity. Srikrishnan and Keller (2021) calibrate an agent-based simulation model on housing abandonment under flood risk, and show that limited data can be insufficient for correctly identifying the model structure. van Schilt et al. (2023) compare various calibration techniques for simulation models when increasing the degree of data sparseness. They calibrate the parameters of a discrete event simulation model of a counterfeit PPE supply chain. van Schilt et al. (2024) evaluate the effect of three dimensions of data sparseness (noise, bias, and missing values) on supply chain visibility using simulation. They use a discrete event simulation model of a counterfeit PPE supply chain. In line with van Schilt et al. (2023, 2024), our study systematically varies the degree of data sparseness for noise, bias, and missing values as well. This enables us to identify the extent to which the behavior of the real world can be represented in a simulation model.

5.2.2. A PLURALIST VIEW ON SIMULATION MODELING

When modeling a complex phenomenon characterized by sparse data and uncertainty, a level of equifinality among the simulation models can exist. This means that many plausible simulation models are coherent with the available sparse data of the real world. Only focusing on one model to analyze the system could lead to a “wrong” view on the phenomenon, and hence ineffective interventions (van Schilt et al., 2023). Thus, a complex phenomenon, such as a supply chain, cannot be captured by a sin-

gle theory or model when data is sparse (Page, 2021). Illicit supply chains are a good example of supply chains where data is intentionally sparse, but legal supply chains also suffer from incomplete and erroneous data. Identifying the structure and parameter values of a supply chain for simulation purposes using sparse data is typically a case where the research philosophy of pluralism applies.

Pluralism as a research philosophy refers to a diversity of views, theories, or models that are required to explain a complex phenomenon, rather than using just a single view, theory, or model. From a research philosophy standpoint, Mitchell (2002) notes that pluralism in science reflects complexity. Building on this, Lenhard and Winsberg (2010) note that having a plurality of models for making forecasts is essential for future science, especially in the field of global climate change models. With respect to analysis with models, Weisberg (2012) refers to robustness analysis for a similar but distinct group of models. The author argues that the more models are available, the more likely robust properties amongst the models can be found that can be related to the real world. Similarly, Durán and Formanek (2018) note that a heterogeneous ensemble of models is needed for robustness analysis. Veit (2020) takes these statements even further and argues that (1) any successful analysis should be focused on a target set of models, and (2) for almost any aspect of a phenomenon, scientists require multiple models to achieve a goal. As one simulation model is a limited representation of the world, it only gives one system view with a very precise formulation (Tolk et al., 2023). This is especially undesirable for systems analysis when there is so little data available about a model property, that it is not even possible to use probability distributions in the model, a phenomenon also known as deep uncertainty (Marchau et al., 2019).

From a modeling perspective, Thompson and Smith (2019) state that having an ensemble of models reduces the possibility of errors. Simulation models require user-defined initial conditions given by scientists based on real-world data. The idea is to generate an ensemble of models with different initial conditions to account for these errors in the initial conditions. Similarly, Page (2021) notes that scientists often cannot use one single simulation model, and introduces the many-model approach. This approach refers to the need for multiple models to understand a complex system.

Both from a research philosophy and from a modeling perspective, the current state-of-the-art literature argues that pluralism in the context of model diversity is essential to an analysis of a complex phenomenon. This holds especially for cases where a phenomenon is characterized by uncertainty and data sparseness.

5.2.3. QUALITY DIVERSITY ALGORITHMS

Quality Diversity (QD) algorithms, or illumination algorithms, aim to find the most diverse set of close-to-optimal solutions using evolutionary concepts (Mouret & Clune, 2015). Traditional optimization algorithms aim to find the best solution within a specific search space, while illumination algorithms focus on providing the highest-performing solution at every user-defined point within that search space (Mouret & Clune, 2015; Chatzilygeroudis et al., 2021). QD algorithms are a relatively novel approach for evolutionary computation, and they are not yet heavily used in many application areas (Pugh et al., 2015; Schneider et al., 2022).

RELATED WORK

QD algorithms are widely applied in the field of reinforcement learning and robotics (Lim et al., 2022; Tjanaka et al., 2023). A classic example is maze navigation where QD is used to generate a set of diverse behaviors for the robot movement to solve the maze (Pugh et al., 2015; Gravina et al., 2018; Fontaine & Nikolaidis, 2021; Grillotti & Cully, 2022). The diverse behaviors are specified in the form of input parameters to the robot. No trivial mapping of the input parameters to the robot movement exists, making quality diversity interesting for these types of problems.

The application of QD algorithms for model calibration is a relatively unexplored area. Schneider et al. (2022) compare various quality diversity algorithms for hyper-parameter optimization of a machine learning model. To our knowledge, no research has been performed on calibrating simulation models using QD.

QD algorithms can be used to present a diverse set of solutions to decision-makers. Recent work of Kent and Branke (2023) combines a quality diversity search with Bayesian optimization. Their approach efficiently identifies the most preferred solutions by including the decision-makers in the process. This interactive illumination process helps decision-makers to understand the problem and to find the most preferred solution(s). Although Bayesian optimization is known for creating high-quality models with only a few observations, this chapter does not focus on using an interactive approach to calibrate models with sparse data (Jalali et al., 2017; Kent & Branke, 2023). Rather, we use QD for the automatic generation of a diverse set of models that can explain the sparse data that is available for a supply chain.

Compared to the other studies, we examine the use of QD algorithms for calibrating simulation models as an emerging field of application. Additionally, we assess the feasibility of QD algorithms in sparse data situations where it is desirable to have model diversity (Page, 2021).

EXPLANATION OF THE QD ALGORITHM

QD algorithms are based on evolutionary concepts of "survival-of-the-fittest" (Pugh et al., 2015; Chatzilygeroudis et al., 2021). The algorithms use three parameter spaces: (1) input space, (2) behavior space, and (3) output space. The input space includes the so-called genotypes x, also known as the input parameters. The behavior space includes the so-called phenotypes $b(x)$, also known as the dimensions describing the behavior of the input parameters x. The output space is the solution space, $f(x)$.

The general idea is that the QD algorithms create diversity by discretizing the behavior space, and fill each container in this space with an optimal solution (Chatzilygeroudis et al., 2021). For this, a mapping between the input space and the behavior space is needed. In a sense, a dimension reduction happens when mapping the genotypes x to the phenotypes $b(x)$. For example, the phenotypes when designing a robot can be the height, weight and energy consumption of that robot (Mouret & Clune, 2015). A straightforward mapping from x to $b(x)$, also direct encoding, occurs when genotypes affect an independent component of the phenotypes. A complex mapping from x to $b(x)$, also known as indirect encoding, occurs when a genotype affects multiple components of the phenotypes, meaning that it is not independent (Mouret & Clune, 2015). Choosing the mapping highly impacts the quality of the solutions, but

unfortunately, it is quite challenging to define a good quality behavior space (Cully & Demiris, 2017). One way to define the dimensions of the behavior space is to use techniques such as Principal Component Analysis. However, in many cases, defining the dimensions of the behavior space cannot be carried out automatically, and expert knowledge is required (Chatzilygeroudis et al., 2021).

The behavior space needs to be discretized to identify subspaces or "containers" in which optimality can be found. Discretization of the behavior space is often done by a grid or a Centroidal Voronoi Tessellation (CVT) (Chatzilygeroudis et al., 2021). A grid-based approach is easy to understand and implement for QD algorithms. A user-defined number of discrete intervals per dimension of the behavior space is needed to create the multi-dimensional hypercuboid containers using the grid. For high-dimensional behavior spaces, Vassiliades et al. (2017) introduces CVT as a geometry tool. CVT explicitly controls the number of containers, and the resulting containers have a convex polygonal shape with corresponding centroids.

Finding the optimal solution for each container in the behavior space is done with a QD algorithm. One of the first and most widely applied QD algorithms is MAP-Elites (Mouret & Clune, 2015). MAP-Elites starts with generating a set of candidate solutions with randomly chosen genotypes x following a Gaussian distribution. The behavior $b(x)$ and the output $f(x)$ of these candidate solutions are calculated. Next, the candidate solutions are placed into the containers to which they belong in the behavior space. When multiple candidate solutions are placed in the same container, the highest-performing one (i.e., the best output) is kept. After initialization, the search algorithm starts. It randomly selects a container in the discretized behavior space. Mutation and crossover is used to generate offspring from the candidate solution in the container. If the offspring has the highest-performing output, then it replaces the current candidate solution. This process repeats until a stopping criterion, e.g., the number of function evaluations, is reached.

Another commonly used QD algorithm is Novelty Search with Local Competition, introduced by Lehman and Stanley (2011). This algorithm compares the quality and the diversity of a candidate solution relative to its neighbor. It optimizes the candidate solutions on (1) quality: maximizing the output relative to its neighbors, and (2) diversity: maximizing the novelty objective on how far the solution in the behavior space is distant from its neighbors. The main limitation of this algorithm is that this algorithm creates two individual sets of solutions for quality and diversity. This is less effective than having one whole set of solutions generated by MAP-Elites (Chatzilygeroudis et al., 2021).

More recently, Fontaine et al. (2020) proposed the Covariance Matrix Adaptation MAP-Elites (CMA-ME). This algorithm combines the popular MAP-Elites with the single-objective optimization algorithm called Covariance Matrix Adaptation Evolution Strategy (CMA-ES). This hybrid algorithm efficiently explores new areas in the search space using MAP-Elites, while using the selection and adaptation rules of CMA-ES to find high-quality solutions. Fontaine and Nikolaidis (2021) shows that CMA-ME outperforms MAP-Elites for finding a diverse set of optimal solutions, and could work in the case of ill-conditioned objectives and measure functions.

Another recent extension of MAP-Elites is the multi-objective MAP-Elites of Pierrot et al. (2022). In addition to MAP-Elites, it uses multi-objective optimization to create a Pareto front for each container of the behavior space. This approach is interesting for most real-life problems where multiple objectives are conflicting, providing insights into trade-offs for this diverse set of Pareto optimal solutions.

The main limitation of QD algorithms is that it is not guaranteed that all the containers in the discretized behavior space are filled (Lehman & Stanley, 2011; Mouret & Clune, 2015; Chatzilygeroudis et al., 2021). Since the discretization of the behavior space is a user-defined process, it can be that the genotypes do not map to some phenotypes. This especially plays a role in the case of a complex mapping process, i.e., indirect coding. Another limitation is that the algorithm cannot directly search in the behavior space due to the mapping (Mouret & Clune, 2015). This means that it is possible that many candidate solutions with different genotypes could be present in the same container.

5.3. METHOD

This section outlines the method for generating and evaluating a diverse set of optimal supply chain configurations using a QD algorithm. First, the formalization of the case study is discussed. Second, the configuration of the QD algorithm is presented. Last, the design of experiments using a ground truth set-up is explained.

5.3.1. FORMALIZATION OF CASE STUDY

The case study used in this research is a stylized counterfeit PPE supply chain. Based on open-source data and expert interviews, we use one specific configuration of a stylized counterfeit PPE supply chain as ground truth.

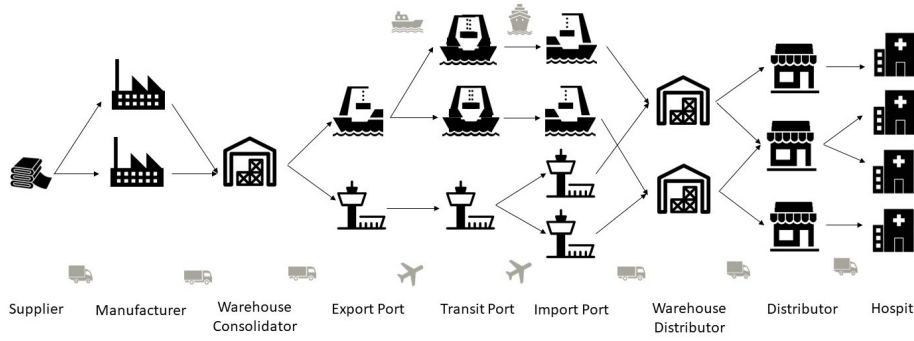


Figure 5.1: Synthetic Ground Truth Supply Chain of Counterfeit PPE.

Figure 5.1 visualizes the structure of the ground truth counterfeit PPE supply chain simulation model from China to the northeast United States of America (USA). The symbols in the figure represent the main actors in the supply chain, and the arrows represent the transportation flows. The supply chain starts at the supplier of raw materials, located in Guangdong, China, who supplies products for PPE such as

fabrics. These products are transported overland by a small truck to one of the two manufacturers in the same area. The manufacturers produce counterfeit PPE in the factory, pack them in boxes, and consolidate them into batches for transportation. Each batch contains a specific quantity of counterfeit PPE, such as 2000 boxes with 20 PPE units per box, resulting in a total of 20,000 PPE units per batch.

Next, a batch of finished counterfeit PPE is transported from the manufacturers' location via a truck to the consolidation warehouse close to the border of Hong Kong. Batches from several manufacturers are stored here until an order arrives and then, this specific order gets picked. On average every 1.2 days, an order is picked up from the consolidation warehouse. The order of counterfeit PPE can either be directly put in a shipping container for international transport at pickup, or it can be picked up and put in a shipping container later at the port of destination. The probability that the counterfeit PPE is in a shipping container at the warehouse consolidator is 0.5. In both cases, the order is transported overland to the export seaport or airport in Hong Kong.

For transportation over sea, the batch is loaded into a 40 ft shipping container and transported by a small container ship to the transit port. Upon arrival at the transit port, the small container ship unloads the shipping container carrying counterfeit PPE. At the same port, the shipping container is loaded onto a larger container ship for overseas transport. Depending on the destination port, the route that the shipping container follows is either (1) from Hong Kong to New York, USA via Singapore, or (2) from Hong Kong to Boston, USA via Shanghai, China. For transportation via air, the batch is loaded into the cargo hold of an international airplane using pallets. The destination of this batch is either New York, USA (airport JFK) or Boston, USA (airport BOS). In both cases, there is a transit at Amsterdam Schiphol Airport (AMS), where the batch is moved from one airplane to another.

Arriving at the import port, the batch in a shipping container or pallet is unloaded at one of these ports, and waits at the yard for inland transport to one of the two (illegal) warehouse distributor in the area of New York, USA or Boston, USA. The counterfeit PPE can already be extracted from the shipping container at the import port, or it can happen at the warehouse distributor. The probability of the counterfeit PPE being extracted at the import port is 0.5. At the warehouse distributor, the batch of counterfeit PPE is equally divided into smaller batches for the two distributors they serve. Small trucks directly transport these smaller batches to distributors in New Hampshire, Connecticut, and New Jersey. Lastly, the distributors transport the batch to hospitals in Portsmouth, Providence, New Haven, and Philadelphia where the counterfeit PPE are used as non-counterfeit.

We develop a discrete event simulation model of this stylized configuration of a counterfeit PPE supply chain from China to hospitals in the northeast USA. In the simulation model, most uncertainties such as processing times at actors, delays of transport modalities, and speed of transport modalities follow triangular distributions inspired by real-world data of a fashion retailer and expert interviews (Kuipers, 2021; Hashemi et al., 2022). Table 5.1 and 5.2 show the input parameters for the actors and the links used in the ground truth simulation model. Table 5.3 shows parametrization of the speed and the delays of the transport modalities for the sim-

ulation model of this study. The simulation model has been developed with the library *pydsol-core* and *pydsol-model* in Python in combination with *networkx*. The library *pydsol* is a Python implementation of the Distributed Simulation Object Library (DSOL), originally implemented in Java (Jacobs, 2005).

Actors			
Input Parameter	**Distribution**	**Value**	**Unit**
Interarrival time of product at supplier	Exponential	10	days
Time at manufacturer	Gamma	1.5, 0.8	days
Time at warehouse consolidator	Triangular	0.5, 1, 1	days
Time to pickup at warehouse consolidator	Triangular	0.5, 1, 2	days
Probability of counterfeit PPE in shipping container at warehouse consolidator		0.5	
Time at sea ports	Triangular	0.5, 1, 2	days
Time at air ports	Triangular	0.5, 1, 1	days
Waiting time at yard for transport at import sea port	Uniform	0.5, 3	days
Probability of counterfeit PPE extracted at import sea port		0.5	
Waiting time at yard for transport at import air port	Uniform	0.5, 1	days
Probability of counterfeit PPE extracted at import air port		0.5	
Time at warehouse distributor	Triangular	1, 2, 2	days
Time at distributor	Exponential	0.2	days
Time at hospital	Exponential	0.1	days

Table 5.1: Input parameters of actors for the simulation model of the synthetic counterfeit PPE supply chain.

Links		
Name	**Value**	**Unit**
Supplier to manufacturer 1	50	km
Supplier to manufacturer 2	80	km
Manufacturer 1 to warehouse consolidator	140	km
Manufacturer 2 to warehouse consolidator	75	km
Warehouse consolidator to export sea port	45	km
Warehouse consolidator to export air port	60	km
Export sea port to transit sea port Shanghai	2.8	days
Export sea port to transit sea port Singapore	9	days
Transit sea port Shanghai to import sea port Boston	42.5	days
Transit sea port Singapore to import sea port New York	26	days
Export air port to transit air port Amsterdam	9274	km
Transit air port Amsterdam to import air port Boston and New York	5547, 5847	km
Import sea and air port Boston to warehouse distributor Boston	15, 20	km
Import sea and air port New York to warehouse distributor New York	80, 72	km
Warehouse distributor Boston to distributor New Hampshire, Connecticut	105, 150	km
Warehouse distributor New York to distributor Connecticut, New Jersey	175, 150	km
Distributor New Hampshire to hospital Portsmouth	15	km
Distributor Connecticut to hospital Providence, New Haven	140, 60	km
Distributor New Jersey to hospital Philadelphia	50	km

Table 5.2: Input parameters of links for the simulation model of the synthetic counterfeit PPE supply chain.

Transport modalities							
Input Parameter	**Distribution**	**Value**	**Unit**	**Input Parameter**	**Distribution**	**Value**	**Unit**
Speed of small truck	Triangular	0, 100, 120	km/h	Delay of small truck	Triangular	0, 0.2, 0.5	days
Speed of large truck	Triangular	0, 80, 120	km/h	Delay of large truck	Triangular	0, 0.5, 1	days
Speed of train	Triangular	25, 40, 75	km/h	Delay of train	Triangular	0, 0.3, 0.5	days
Speed of feeder	Triangular	10, 18, 25	knots	Delay of feeder	Triangular	0, 4, 16	days
Speed of vessel	Triangular	10, 18, 25	knots	Delay of vessel	Triangular	0, 7, 16	days
Speed of airplane	Uniform	740, 930	km/h	Delay of airplane	Triangular	0, 1, 4	hours

Table 5.3: Input parameters of speed and delay of the transport modalities for the simulation model of the synthetic counterfeit PPE supply chain.

The ground truth discrete event simulation model is developed to produce the observed data of the system. We extract time series data from the simulation model that describes when a quantity of PPE arrives at an actor, including the location and the type of actor. For example, a batch with a quantity of 20.000 PPE arrives at the export airport in Hong Kong on day 3. Data of the time series is summed per day, and is aggregated over the actor types. Multiple replications are combined using the mean value per day per actor type. In this research, the model runs for a simulation time of 52 weeks with 10 replications with unique seeds.

5.3.2. CONFIGURATION OF QUALITY DIVERSITY ALGORITHM

Our study uses a QD algorithm to generate a diverse ensemble of optimal supply chain simulation models. In essence, this algorithm is used to calibrate the simulation model to find a diverse set of plausible supply chain simulation models. The configuration of the QD algorithm includes the description of the three spaces: the input space, the behavior space, and output space. Next, we outline the configuration of the algorithm for this study.

Input Space The input space of the QD algorithm defines the parameters to calibrate. These are uncertain parameters in the simulation model that need to be tuned such that the model's behavior matches the real-world behavior. In the case of illicit supply chains, the structure and the actor's parameters of the supply chain simulation model are uncertain. We design various profiles of the input space that combine structural and parametric uncertainty for specific parts of the supply chain. Table 5.4 gives an overview of the various profiles of the input space and the corresponding parameters. Each parameter in the input space has bounds to ensure that the algorithm chooses feasible candidate solutions. For example, a simulation model cannot schedule an event in the past, so a parameter related to time cannot be lower than zero. Note the parameters excluded from a composition's input space are considered known and correspond to the ground truth.

The first profile of the input space is defining the structure of the supply chain. For the structural uncertainty, a large set of supply chain configurations with different numbers of actors and connectivity is created using the System Entity Structure approach (Zeigler, 1984; Zeigler & Hammonds, 2007). The set of graphs is ranked on the density in each graph, and the graph index is part of the input space. Second, we add parametric uncertainty to the existing structural uncertainty for the other pro-

files of the input space. This means that both the structure of the supply chain (the profile *structure*) and the parameters of a specific part within the supply chain are uncertain. We divide the supply chain into five parts for defining the profiles of the input space: modus operandi, source, legal, import, and destination.

Profiles	Distribution	Bounds	Unit
Structure			
Graph structure (index)		(0, 40.000)	int
Modus Operandi			
Probability of counterfeit PPE in shipping container at warehouse consolidator		(0, 1)	
Probability of counterfeit PPE extracted at import sea port		(0, 1)	
Probability of counterfeit PPE extracted at import air port		(0, 1)	
Source			
Interarrival time of product at supplier	Exponential	(1, 15)	days
Time at manufacturer	Gamma	[(0.1, 10), 0.5]	days
Time at warehouse consolidator	Triangular	[(0.1, 9.9), (0.1, 10), (0.2, 10)]	days
Time to pickup at warehouse consolidator	Triangular	[(0.1, 10), (0.1, 20), (0.5, 20)]	days
Legal			
Time at sea ports	Triangular	[(0.5, 2), (0.5, 5), (1, 5)]	days
Time at air ports	Triangular	[(0.1, 1), (0.1, 2.5), (0.3, 2.5)]	days
Import			
Waiting time at yard for transport at import sea port	Uniform	[(0.1, 1), (0.5, 5)]	days
Waiting time at yard for transport at import air port	Uniform	[(0.1, 1), (0.3, 2)]	days
Destination			
Time at warehouse distributor	Triangular	[(0.1, 9.9), (0.1, 10), (0.2, 10)]	days
Time at distributor	Exponential	(0.1, 5)	days
Time at hospital	Exponential	(0.1, 5)	days

Table 5.4: Overview of the profiles of the input space and the corresponding parameters. The actor's parameters following a distribution have more parameters in the input space (min, mode, max) and hence, more bounds.

Behavior Space The behavior space defines the dimensions on which the diverse set of solutions is positioned. The input space is mapped to the behavior space to ensure that the dimensions describe the behavior of the input parameters. For illicit supply chains, it is interesting to find diverse and optimal supply chain simulation models with different transport costs and various degrees of network vulnerability (Anzoom et al., 2021). Transport costs are an essential part of profit-driven crime; lower costs mean more profit (Snaphaan & van Ruitenburg, 2024). Network resilience shows the extent to which the network is vulnerable to interventions of law enforcement; more resilient is more interesting for fraudulent organizations (Ficara et al., 2021). Both dimensions are important for fraudulent organizations when designing a supply chain. For example, a supply chain with low transport costs and high network resilience increases the profit and reduces the probability that their illicit business will be inactive. Additionally, it could help law enforcement to gain an understanding of the supply chain configurations that fraudulent organizations most likely choose depending on the organizations' perspective on transport cost and network resilience, and hence, business model.

In more detail, the 33 input parameters are mapped to the two dimensions in the behavior space. First, the transport cost in the behavior space is defined as the average transport cost of a product. Let the graph of interest be $g = (V, E)$, let $p \in P$ be the set of finished products at the end customers, and let $m \in M$ be the set of possible

transport modes with $M = \{smalltruck, largetruck, feeder, vessel, train, airplane\}$. Let t_p^{total} be the total time of product p in the supply chain, starting from the supplier to the hospital. Let $t_{p,e,m}$ be the travel time of product p on edge e with transport mode m. Let c_m be the cost per time unit per mode of transport. Let v be a linear function defining the time discount of the market value of PPE as $v(t) = \frac{-0.34}{730}t + 0.59$ [1]. The starting market value of any PPE product p is $v(0) = 0.59$. The average transport cost C over all products P is:

$$C = \frac{\sum_{p \in P} \sum_{e \in E} \sum_{m \in M} c_m t_{p,e,m} \left(1 - \frac{v(t_p^{total}) - v(0)}{v(0)}\right)}{\#P} \tag{5.1}$$

Second, network resilience is the resistance of the network to disruption, being an intervention from law enforcement, and the adaptation following this disruption (Anzoom et al., 2021). From a criminal perspective, three factors influence the resilience of the network: (1) a diversity of links in a complex network, (2) the nodal position and their criticality, and (3) human capital (Cavallaro et al., 2020). On a network level, a criminal network containing a diversity of links makes it tolerant to random disruptions, and hence, resilient. On the node level, the position of the actor on centrality and visibility determines the vulnerability or resilience of that actor (Morselli, 2010; Diviák et al., 2019). In this research, the behavior space entails resilience on a network level, and hence, the diversity of links. In line with this, Gao et al. (2016) state that density, i.e., the ratio of the number of edges to the possible number of edges in a network, is one of the key factors influencing a network's resilience. Therefore, we use density as a measure of network resilience in the behavior space. Equation 5.2 describes the formula of the density with $g = (V, E)$ as the directed graph of interest.

$$R = \frac{\#E}{\#V(\#V - 1)} \tag{5.2}$$

The behavior space is discretized in a grid of 10 x 10 containers. We refer to these as QD containers. The range of transport costs in this behavior space is between \$250 to \$1250, and, for density, between 0.02 and 0.07. The behavior space only has two dimensions, making a grid suitable for enhancing the understandability and interpretability of the results (Chatzilygeroudis et al., 2021).

Output Space The output space defines the objective to minimize, in this case, the distance between the ground truth output and the output of the candidate solution. A distance metric is used to describe the distance between the simulation model data and the observed data given a certain function. In this research, we use a classic distance metric: the Manhattan (L1) distance. The Manhattan distance is the sum of the

[1] Discount of the market value of PPE is extracted from the statistics on face masks from the beginning of COVID-19 (2020) to the end (2022). See https://www.statista.com/outlook/cmo/tissue-hygiene-paper/face-masks/worldwide#volume.

absolute differences for each dimension of the data points. This distance metric is highly efficient for complex and high-dimensional data applications such as discrete event simulation models (Aggarwal et al., 2001; Mirkes et al., 2020). In our research, we compare the aggregated time series data of each actor resulting from the ground truth simulation model and the candidate simulation model. We normalize the Manhattan distance of each actor using the 5th percentile and 95th percentile of the actor's ground truth data. Next, we sum the normalized Manhattan distance of each actor to get the overall Manhattan distance between the ground truth and candidate solution.

Configuration of Algorithm In this research, the QD algorithm Covariance Matrix Adaptation MAP-Elites (CMA-ME) of Fontaine et al. (2020) is used for finding a diverse set of optimal solutions due to its high performance. Emitters are instances of the CMA-ME algorithm that generate new candidate solutions, adapt, and save the population of solutions. The algorithm is initialized with an emitter across ten unique seeds. Each emitter generates 96 candidate solutions in each iteration. A convergence analysis is performed on the number of quality improvements and diversity to determine the number of iterations required. For the profile structure, the QD algorithm is run for 110 iterations, meaning a total of (96 x 110 =) 10560 function evaluations per seed. For the profiles that include parametric and structural uncertainty, we use (96 x 156 =) 14976 function evaluations for convergence. The quality diversity algorithm is implemented using the python library *pyribs* (Tjanaka et al., 2023).

5.3.3. DESIGN OF EXPERIMENTS

For our experiments, a ground truth set-up is used to assess the feasibility of applying QD under a varying degree of data sparseness. This set-up allows us to measure how closely QD calibrates the "true" values, which is challenging when dealing with real-world data (Khondoker et al., 2016; van Schilt et al., 2024). A stylized simulation model serves as the ground truth, from which we extract data representing the observed data of the system. Figure 5.2 visualizes the ground truth set-up using the following steps:

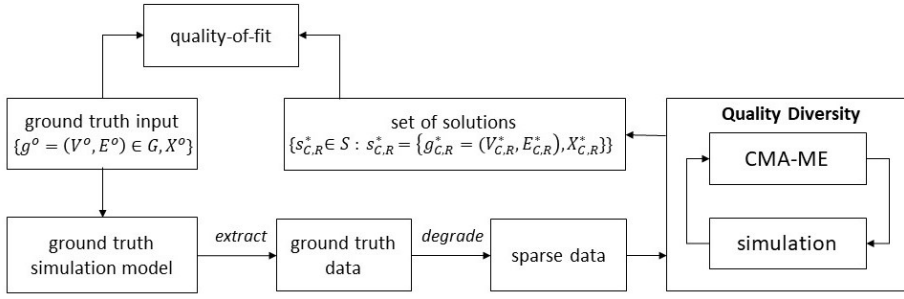


Figure 5.2: Ground Truth Set-Up.

- Start with the input data for the ground truth, being a graph with vertices and edges, $g^o = (V^o, E^o)$, and input parameters, X^o. The ground truth is presented in Section 5.3.1.
- Use the input data to design and run a ground truth simulation model. From this model, we extract the ground truth data.
- Degrade the ground truth data on three dimensions of data sparseness (missing values, noise, and bias) following van Schilt et al. (2024). For the experiments, we design scenarios that incorporate all dimensions of data sparseness, reflecting real-world data of supply chains. We categorize each dimension with a low sparseness of 20%, a medium sparseness of 50%, and a high sparseness of 80%. Table 5.5 presents an overview of the scenarios used in this study.

Scenarios	Bias	Noise	Missing Values
All Low	20%	20%	20%
All Medium	50%	50%	50%
All High	80%	80%	80%
Bias Low	20%	50%	50%
Bias High	80%	50%	50%
Noise Low	50%	20%	50%
Noise High	50%	80%	50%
Missing Low	50%	50%	20%
Missing High	50%	50%	80%

Table 5.5: Scenarios for the Dimensions of Data Sparseness.

- Use the sparse data as input for the QD process. This research uses the CMA-ES algorithm for calibration. From the input space, we select candidate graphs and candidate input parameters as candidate solutions. Then, the simulation runs 10 replications and compares the simulated output with the sparse data to determine the objective Manhattan distance. Additionally, the transport cost and the density of the network are outputs of the simulation model needed for the behavior space. The QD process continues until a stopping criterion is reached.
- Collect the set of optimal solutions resulting from the QD process. This set of optimal solutions, S, contains one optimal solution for each QD container in the behavior space, i.e., $s^*_{C,R} \in S$. Each solution entails a combination of a graph and input parameters that assign the solution to a part of the behavior grid, i.e., $s^*_{C,R} = \{g^*_{C,R} = (V^*_{C,R}, E^*_{C,R}), X^*_{C,R}\}$.
- Analyze the quality-of-fit by comparing the ground truth input and the set of solutions resulting from QD. While QD minimizes the gap between the simulated data and the sparse data, this does not necessarily mean that the set of solutions captures the ground truth. Hence, we determine the proximity of the set of solutions to the ground truth by assessing how often the ground truth

is identified by QD across various unique seeds. In addition, we compare the solutions on various features such as transport cost, density, objective value (Manhattan distance), the number of vertices, the number of edges, and the graph edit distance, i.e., the cheapest set of graph edit operations needed to transform one graph to the other graph (Abu-Aisheh et al., 2015). We use an approximated greedy graph edit distance of Riesen et al. (2015) for computational reasons.

5.4. RESULTS

We discuss the results of calibrating the supply chain simulation model for each profile using the QD algorithm when varying the degree of data sparseness. First, we present the results of the convergence of the QD algorithm. Second, we analyze the extent to which QD can identify the ground truth across the seeds. Third, we combine the results of the seeds into a single QD front and examine the QD container where the ground truth is expected. Last, we evaluate the overall QD front.

In this section, we refer to the Quality Diversity results mapped into the behavior space as the QD front. Figure 5.3 presents an example of the QD front of the profile structure with 0% of data sparseness. In this figure, we display the container on the QD front where the ground truth is expected. The behavior values of the ground truth model are $709.9 transport cost and 0.0573 density. In the discretized behavior space, this means the ground truth fits in the QD container between $650 to $750 transport cost and a density of 0.055 to 0.060. We refer to this as the ground truth container.

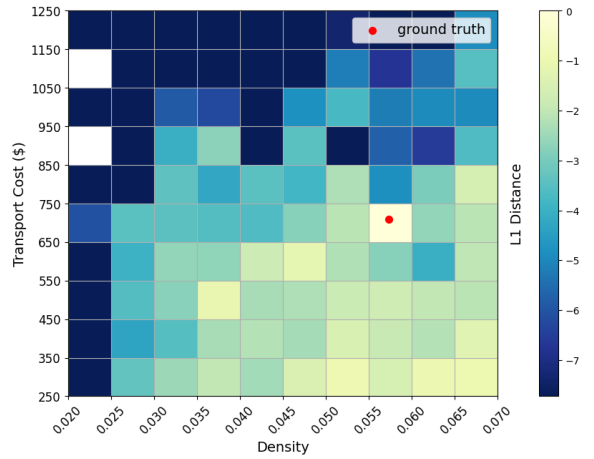
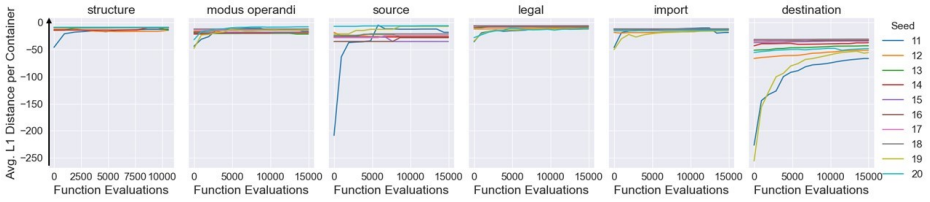


Figure 5.3: Quality Diversity Results of Profile Structure with 0% of Data Sparseness.

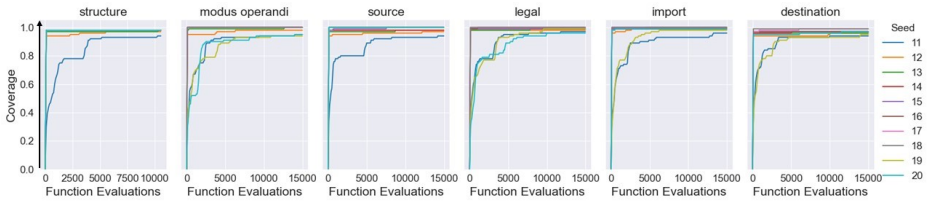
To properly compare the different data sparseness scenarios, the output space, in terms of its Manhattan (L1) distance, is normalized using the minimum and the maximum objective values of the QD front. This means the objective value closest to the ground truth is 0, and the objective value the most far from the ground truth is 1. The direction of desirability is towards 0. We refer to the normalized objective value as normalized L1 distance. More information on the normalization can be found in Appendix B.1.

5.4.1. CONVERGENCE OF QD ALGORITHM

We evaluate the convergence of the QD algorithm for calibrating the structure and the parameters of the counterfeit PPE simulation model. For this, we focus on the scenario with 0% of data sparseness. A measure for convergence is the quality of the QD front per seed by the average L1 distance between the QD containers and the ground truth. This measure indicates the magnitude of improvements from each function evaluation. A constant value means no substantial improvements are made to the front. Another measure for convergence is the diversity of the QD front per seed by the coverage, i.e., the percent of QD containers in the behavior space that contains a solution. A higher coverage means that more QD containers in the QD front contain a solution. We divide the behavior space into 100 QD containers, so a coverage of 0.94 means that 94 of the 100 QD containers hold a solution.



(a) Quality Measured as Average L1 Distance Between the QD Containers and the Ground Truth per Profile



(b) Density Measured as Coverage Percent

Figure 5.4: Convergence on Quality and Diversity per Seed at 0% of Data Sparseness.

Figure 5.4 shows that the average L1 distance and the coverage are constant for most of the seeds for all the profiles. In general, the average L1 distance across the QD containers and the coverage stays relatively constant starting from the initial sample, except for seed 11 and seed 19. There is still much room for improvement for these seeds after initializing, but the average L1 distance and the coverage become closer in proximity to the other seeds over the function evaluations. Overall, Figure 5.4a shows that calibrating for the profile source and destination leads to the variation between the seeds on the average L1 distance across the QD containers. Figure 5.4b shows that most seeds have a high coverage percent and, therefore, lead to a high degree of diversity.

In some experiments, the figures show that the average L1 distance over the QD containers decreases, whereas the aim is to increase towards 0. This can be explained by the diversity of the QD front in that experiment. For example, for the profile source at seed 11, we see a peak of the average L1 distance at 5760 function evaluations of -4.9 with a coverage of 0.74. For the next iteration at 6720 function evaluations, the

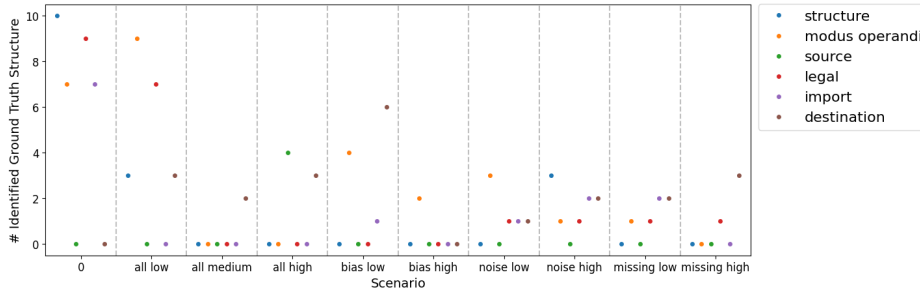


Figure 5.5: Number of Times the Ground Truth Structure is Identified per Profile and per Scenario across Ten Seeds.

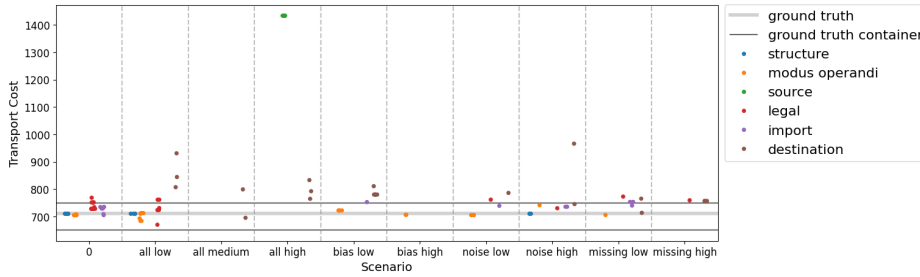


Figure 5.6: Transport Cost of the Identified Ground Truth Structures per Profile and per Scenario across Ten Seeds.

average L1 distance across the QD containers decreases to -11.8 with a coverage of 0.76. This means that the average L1 distance decreases when the coverage increases.

5.4.2. IDENTIFYING THE GROUND TRUTH ACROSS SEEDS

Figure 5.5 displays how many times the ground truth structure has been identified across the ten seeds for each profile and the various data sparseness scenarios. Overall, the ground truth structure is most often identified at 0% of data sparseness, especially for the profile where only the structure is calibrated. Remarkable is that for the profile source and the profile destination, QD fails to identify the ground truth structure at 0% of data sparseness. For the profile source, the ground truth structure is successfully identified only once for the scenario all high. For the profile destination, the ground truth structure has been identified in scenarios with more data sparseness and, most frequently, in the case of the scenario with a low bias percentage. For the other profiles, the results generally show that more sparseness leads to less or similar identification of the ground truth structure.

Examining the solutions that contain the ground truth structure, Figure 5.6 shows that many of these solutions have a higher transport cost compared to the ground truth. This means that these solutions are placed in QD containers other than the ground truth container. Especially for the profile legal and destination, the identified

ground truth transport cost is often higher than the ground truth. This could mean that the QD calibrates the legal and destination parameters often too high compared to the ground truth such that the configuration results in a higher transport cost. For the source profile, we see that, in the only scenario where the ground truth structure has been identified, the configuration of the source parameters leads to extremely high transport costs. This is caused by relatively low interarrival time and relatively high warehouse consolidator times. Thus, the behavior of this specific solution does not align closely with the ground truth in terms of transport cost. See Appendix B.2 for the graphs on the parameter values.

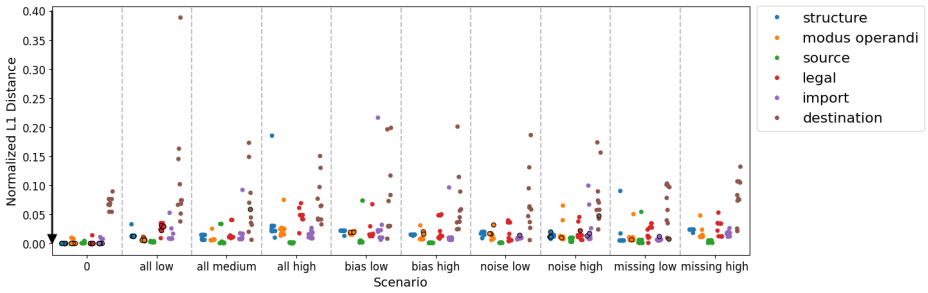


Figure 5.7: Normalized L1 Distance for the Ground Truth Container per Profile and per Scenario across Ten Seeds. The solutions containing the ground truth structure have a black outer edge. The arrow represents the direction of desirability.

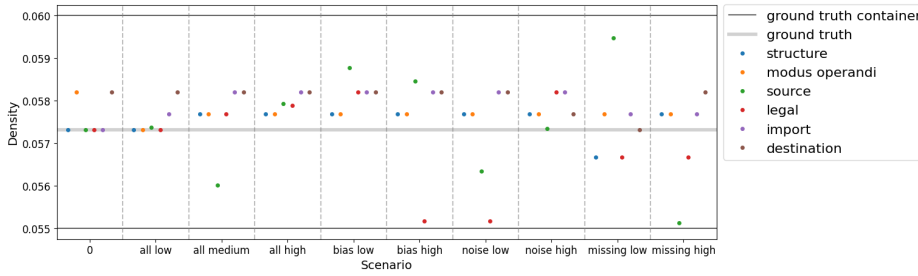
Zooming in on the QD container in the behavior space where the ground truth is expected, Figure 5.7 displays the most optimal supply chain configurations for each seed over the various scenarios and profiles in this specific QD container. When combining the results of the seeds to a single QD front, the solution with the lowest L1 distance is chosen. In the figure, we see that the solution containing the ground truth structure is the most optimal for the majority of the profiles at 0% of data sparseness, and the scenario all low. However, for the other scenarios of data sparseness, the solution containing the ground truth frequently fails to be the most optimal across the seeds and, consequently, does not appear in the single QD front.

5.4.3. ANALYZING THE GROUND TRUTH CONTAINER OF QD FRONT

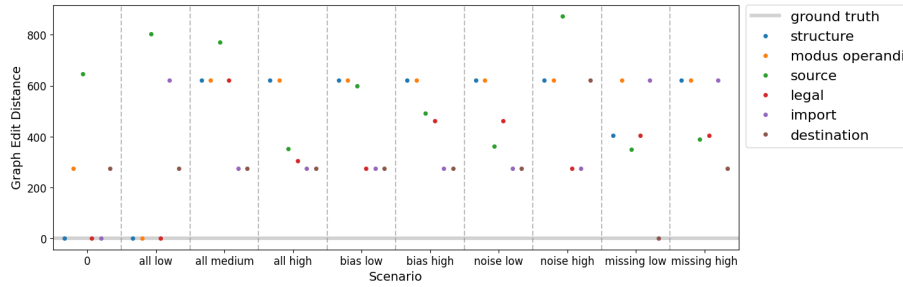
The solutions of the seeds are combined to create a single QD front for each profile in each scenario. Similar to the QD algorithm, the most optimal solution for each QD container across the seeds is included in the QD front. We analyze the QD container where the ground truth model fits to assess the quality-of-fit of the QD front for each profile in each scenario.

Figure 5.8 shows the density, i.e., the ratio of the number of edges to the possible number of edges in the graph, and the graph edit distance, i.e., the cheapest set of graph edit operations to transform the graph to the ground truth. The figure shows that the ground truth is most often identified for the scenario of 0% data sparseness and the scenario of all dimensions having a low data sparseness. For the other scenar-

ios, the optimal solutions found in the ground truth container have a higher density (Figure 5.8a). The graph edit distance of solutions that did not result in the ground truth structure is mostly between 276 and 621 edit operations (Figure 5.8b).



(a) Density.



(b) Graph Edit Distance.

Figure 5.8: Characteristics of the Solutions in the Ground Truth Container in the Quality Diversity Front per Scenario and per Profile.

Figure 5.8 shows that two graph structures are most often identified as optimal across most scenarios. The graph structure with a density of 0.0577 and a graph edit distance of 621 is often identified as optimal for the profile structure and modus operandi. The graph structure with a density of 0.0582 and a graph edit distance of 276 is often identified as optimal for the profile import and destination. An exception is the profile source that did not identify either of the two optimal graph structures for any scenario. This profile has identified solutions with a relatively high or relatively low graph edit distance compared to the ground truth. Moreover, the solutions identified for this profile have the highest graph edit distance, with a range between 362 to 873.

In more detail, Table 5.6 presents the ground truth structure and the graph structures of the two most often identified optimal solutions. The graph with a density closer to the ground truth (1) has a higher graph edit distance and a higher number of vertices and edges than the graph with a higher density (2). The figures of the graph structures also show that graph (1) has many more transit ports, import ports, distributors, and hospitals than graph (2). Interestingly, all three structures have one supplier, one or two manufacturers, and one warehouse consolidator. Additionally,

Ground Truth	(1)	(2)
Density: 0.0573	Density: 0.0577	Density: 0.0582
Graph Edit Distance: 0	Graph Edit Distance: 621	Graph Edit Distance: 276
Vertices: 23	Vertices: 52	Vertices: 28
Edges: 29	Edges: 153	Edges: 44

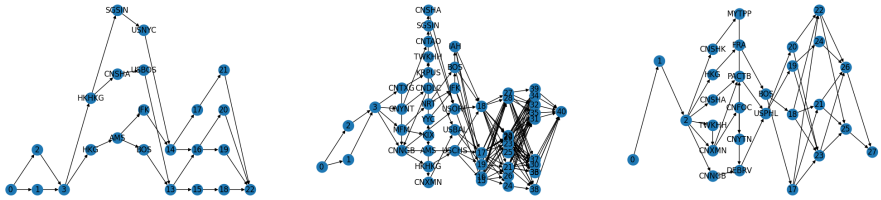


Table 5.6: Characteristics of the graph structure of the ground truth and the two most often identified solutions (1)-(2) in the ground truth container of the Quality Diversity front. Each step in the horizontal line of the structure plot represents a set of actors, going from left to right: supplier, manufacturer, warehouse consolidator, export port, transit port, import port, warehouse distributor, distributor, and hospital.

all graphs have a combination of sea and air transport, with an overlap of ports in the graphs, such as Hong Kong, Amsterdam, Shanghai, and Singapore. Notably, the three graphs have the airport in Boston as an import port. Additional figures on the transport costs, the number of vertices, and the number of edges for each solution in the ground truth container in the QD front can be found in Appendix B.3.

5.4.4. ANALYZING OVERALL QD FRONT

We analyze the overall QD front per profile and per scenario on diversity and quality. Regarding diversity, the QD front for each scenario and each profile has a relatively high coverage when merging the ten unique seeds (see Figure 5.4). For the single QD front, the coverage is between 0.94 and 1.0.

Figure 5.9 shows the distribution of quality of the solutions in the QD front using the number of occurrences (count) of the normalized L1 distance. A histogram with a binwidth of 0.1 is used to determine the counts, meaning there are 10 bins in total. We refer to each bin by using the minimum value of that particular bin, e.g., the bin of 0.1 to 0.2 is referred to as the value 0.1. A high count suggests that this value is more frequently observed in the QD front. In general, we see that each profile has a high count around a normalized L1 distance of 0.0, meaning that most solutions in the QD front are relatively close to the ground truth. The distribution is skewed towards the left. Most profiles have a normalized L1 distance of 0.4 to 0.8, with another small peak around the bin of 0.9.

When looking at the modus operandi and legal profiles, the figure shows two different directions of counts across the data sparseness scenarios. One has a higher number of occurrences around a normalized L1 distance of 0.2, whereas the other has no occurrences around 0.2. The results show that the scenarios that have a high number of occurrences in the modus operandi profile tend to have a low number of occurrences in the legal profile. For example, the scenario with 0% of data sparseness and all high have no occurrences at 0.2 of normalized L1 distance for the modus

operandi profile. In contrast, the scenario of 0% of data sparseness and all high have a count of 10 at a normalized L1 distance of 0.2.

For the other profiles (structure, source, import, and destination), a difference in the number of occurrences between the scenarios is shown around the value 0. For structure, source, and import, we see that all scenarios have the highest number of occurrences around the value 0.1 of the normalized L1 distance. Generally, the scenario bias low generally has the lowest count. For destination, we see a high dispersion in the number of occurrences between the scenarios for the normalized L1 distance of 0, where the scenario bias high has the highest count and the scenario all medium has the lowest.

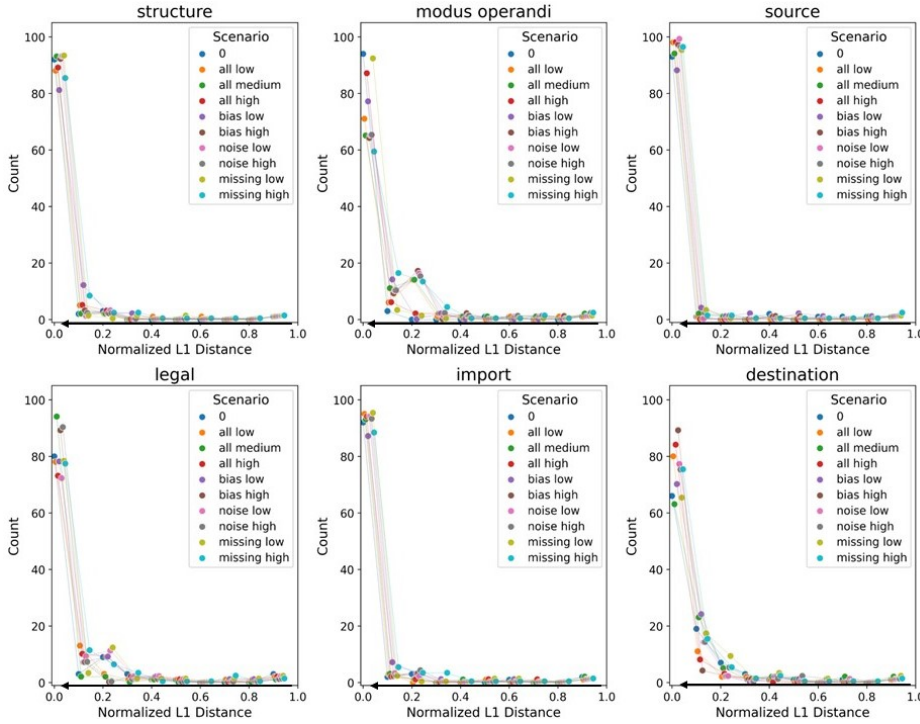


Figure 5.9: Number of Occurrences (Count) of the Normalized L1 Distance of the Solutions in the Quality Diversity Front. A histogram with a binwidth of 0.1 is used for the counts, meaning a total of 10 bins between 0 and 1. The points are the count of one particular bin, e.g., 80 occurrences between 0.0 and 0.1, and are plotted at the starting value of the bin. For visualization purposes, we added some jitter to limit the overlap in the data points.

For the source, legal, import, and destination profiles, the scenario with 0% of data sparseness is on the relatively low segment for the number of occurrences at the value 0, with several occurrences for a higher normalized L1 distance. This could suggest that the scenarios with more sparseness lead to more solutions with a lower normalized L1 distance. Comparatively, the scenario with 0% of data sparseness has a relatively high number of occurrences for the value 0, and little counts for a higher

normalized distance for the profiles structure and modus operandi. This suggests that the scenarios with more data sparseness lead to more solutions with a higher normalized L1 distance.

5.5. DISCUSSION

This section reflects on the results of the Quality Diversity (QD) algorithm and discusses the implications of the results for real-world applications.

5.5.1. REFLECTION ON QUALITY DIVERSITY ALGORITHM

The QD algorithm successfully shows its feasibility for calibrating a supply chain simulation model. A key notion for using QD in this field is its sensitivity to the initialization of the seeds of the algorithm, which determines the initial sample and the randomness for selecting candidate solutions. The convergence results show that for some seeds, there is still much room for improvement in terms of the average Manhattan distance of the solutions and the coverage, whereas other seeds instantly reach a satisfactory level of quality and diversity. An explanation for this is that, in a highly rugged fitness landscape typical for discrete event simulation (Azadivar, 1999), QD needs to perform additional iterations to reach convergence when the initial sample is chosen poorly. Thus, when using QD to calibrate a discrete event simulation model, it is crucial to use various seeds.

In terms of diversity, the results demonstrate that QD fills at least 96% of the QD containers for calibrating this counterfeit PPE supply chain model across all profiles and scenarios of data sparseness. Although the main limitation of QD is that it does not guarantee to fill every QD container in the discretized behavior space (Lehman & Stanley, 2011; Mouret & Clune, 2015; Chatzilygeroudis et al., 2021), QD successfully reaches a high coverage for calibrating this simulation model. Nevertheless, it is necessary to consider the trade-off between the diversity enforced by the algorithm, e.g., solutions are "all over the place", and the quality of the QD front, e.g., limited solutions that are highly optimal.

In terms of the quality-of-fit for calibrating the structure, the results show that QD is able to identify the ground truth for most profiles and various scenarios of data sparseness across the seeds. The ground truth structure is most frequently identified at 0% of data sparseness. In the case of more data sparseness, the solution containing the ground truth structure often has a higher normalized L1 distance and is not the most optimal solution across the seeds. Thus, solutions with another graph structure better fit the sparse data than the ground truth. These solutions have high graph edit distances, and more vertices and edges than the ground truth, thus more complexity. Simulation models with more complex structures often reproduce the sparse data better than those that slightly differ from the ground truth, indicating a risk of overfitting. However, the simulation models with more complex structures do not necessarily lead to "wrong" results since they could explain the sparse data. In addition, the results of this study show much overlap between the ground truth structure and the overfitted structures (Table 5.6), which is potentially interesting for decision-making. For example, all structures include Boston Logan Airport as an import port, and this could be marked as a potential intervention hotspot for counterfeit PPE.

Regarding the quality-of-fit for calibrating the actors' parameters, the results demonstrate that the parameters of the source and destination actors seem to have the most impact on the simulation outcomes and the resulting QD front. For the source profile, QD has difficulty fitting the parameters and the structure, even with 0% of data sparseness. For the destination profile, the ground truth structure is more often identified with sparse data, but the identified actors' parameters result in high transport costs, leading to a different container on the QD front. For the other profiles, the actors' parameters are not necessarily similar to the ground truth, but the QD front does not substantially change. Even though some profiles have a higher number of parameters to calibrate, the parameters have less impact on the simulation model outcomes. In summary, the impact of the actors' parameters on the simulation model outcomes, rather than the quantity, influences the quality-of-fit for QD.

However, the quantity of the parameters does appear to have an impact on the QD algorithm. Specifically, for the profiles with the fewest parameters to calibrate — 1 for structure and 4 for modus operandi — the scenario with 0% data sparseness shows a high number of occurrences of the normalized L1 distance of 0 on the overall QD front. This can be explained by the nature of the QD algorithm CMA-ME, which uses a covariance matrix (Fontaine et al., 2020). With fewer parameters, the covariance structure is simpler, enhancing the exploration of solutions with 0% of data sparseness close to the ground truth than with more parameters. For calibrating a simulation model with more than 5 parameters, the results demonstrate that more data sparseness could lead to more solutions that are coherent with the sparse data. In line, we see that the solutions with different structures and different actors' parameters than the ground truth could explain the sparse data better. This makes it increasingly difficult to detect the ground truth when using the classical way of calibrating a simulation model that minimizes the distance between the models' data and the observed data (Wigan, 1972; Ören, 1981; Hofmann, 2005). Therefore, it would be interesting for future research to reconsider whether the classical way of calibrating and its' metric fits for sparse data situations.

5.5.2. REFLECTION ON REAL-WORLD APPLICATION

This study highlights the importance of gathering information on the upstream supply chain. The findings indicate that calibrating the source profile is the most challenging for QD. Additionally, the parameters of the source actors have a high impact on the simulation model outcomes. Also, the detailed graph structures (Table 5.6) show similarities with the ground truth, particularly on the source part of the supply chain. More specifically, there is a correlation between the number of suppliers and the objective value of QD; the more suppliers a structure has, the higher the normalized L1 distance, and the more distant from the ground truth (Appendix B.4). In the case of a push-pull supply chain, like counterfeit PPE, the upstream supply chain's effectiveness directly impacts the cost efficiency and the lead times of the remainder of the supply chain. Thus, gaining more information on the upstream supply chain helps to identify a diverse ensemble of plausible supply chains and, hence, potentially helps to design robust interventions in real life.

Next, the choice of the behavior space is crucial for applying the QD results in real life. In this study, we only focused on the transport cost and the network vulnerability, whereas other factors, such as the cost of bribing officials, the market value of the counterfeit goods, the trust between actors, and the detectability of certain modi operandi, also play a role in illicit supply chains. For further research, it would be interesting to include these factors in the simulation models.

Last, the aim of generating a diverse ensemble of plausible supply chain configurations is to identify robust interventions for real-world applications. In this study, we investigate the feasibility of the QD algorithm to generate such a diverse ensemble, and highlight the potential for these plausible supply chain configurations to support robust interventions for the counterfeit PPE supply chain. However, further research should evaluate the actual effectiveness of this ensemble for identifying robust interventions, both theoretical and in practice.

5.6. CONCLUSION

This research examines the feasibility of the Quality Diversity (QD) algorithm for generating a diverse ensemble of supply chain simulation models when the available data is sparse. For this, we use a case study of a counterfeit PPE supply chain as the ground truth, extract data from the ground truth, and vary the degree of data sparseness. We assess whether QD can identify the ground truth among the diverse set of solutions, in the case of structural and parametric uncertainty.

Our analysis demonstrates that QD is able to generate a diverse ensemble of supply chain simulation models. Due to the algorithms' sensitivity to seed initialization, it is crucial to use various seeds. QD identifies the structure of the ground truth most frequently for 0% of data sparseness. In case of more data sparseness, simulation models of more complex structures, i.e., more vertices and edges, than the ground truth, tend to describe and reproduce the sparse data better. These complex structures are not necessarily "wrong", as they show overlap with the ground truth in parts of the supply chain. For parametric uncertainty, the impact of the parameters on the simulation model outcomes, rather than the quantity, influences the quality-of-fit of the solutions of the QD algorithm for identifying the ground truth. Additionally, the results show that, in the case of structural and parametric uncertainty, more data sparseness could lead to more solutions that are coherent with the sparse data. Simulation models with different structures and parameters can reproduce the sparse data better than the ground truth, making it difficult to identify the ground truth when only minimizing the models' data with the observed data – the classical way of calibrating.

For practical implications, our study offers a first insight into the potential of using QD algorithms to generate a diverse ensemble of reconstructions of a supply chain, in particular for supply chains with sparse data. The results emphasize the importance of gathering information on the upstream supply chain to identify such an ensemble. This could help decision-makers to make more robust decisions on, for example, interventions that are effective for the ensemble of supply chain models rather than for a single model.

Further research should focus on reviewing the way of calibrating simulation models and their metrics in sparse data situations, and extending the simulation model to other types of supply chains and adding additional (upstream) information. It would also be interesting to systematically increase the degree of the dimensions of data sparseness to assess the feasibility of QD instead of scenario-based. Last, further research should evaluate the actual effectiveness of the diverse ensemble of plausible supply chain configurations for identifying robust interventions, theoretically and in practice.

BOSTON

6

DISCUSSION

This chapter presents a methodological reflection and a practical reflection regarding this dissertation, showcasing directions for further research.

6.1. METHODOLOGICAL REFLECTION

This section reflects on the following methodological limitations and implications of this dissertation: (1) the impact of the dimensions of data sparseness, (2) the properties of the model calibration techniques, (3) the overfitting of the dense graphs in the calibration, and (4) the need for diversity.

Impact of Dimensions of Data Sparseness

Throughout this dissertation, we use three dimensions of data sparseness – noise, bias, and missing values – as defined in Chapter 2. For all three dimensions of data sparseness, the average percentage of global supply chain visibility decreases when more sparseness is added to the data. The research results show that the "missing values" dimension has more impact on the decrease of supply chain visibility than noise and bias. For model calibration purposes (Chapter 3, 4, and 5), we see a similar impact of the different dimensions of data sparseness (as in Chapter 2) on the objective value of the model calibration, the Manhattan (L1) distance. Also here, the Manhattan distance decreases most when missing values are added, and decreases less for noise and bias.

For the results of the model calibration techniques, the impact of the different dimensions is less visible. For example, Powell's Method and Approximate Bayesian Computing generate the same solutions across a varying percentage of noise, bias, and missing values. This implies that for certain algorithms, such as Powell's Method and Approximate Bayesian Computing, the dimension of data sparseness becomes irrelevant since it does not impact the identified solutions. For the other model calibration techniques – Bayesian Optimization, Genetic Algorithm, and the Quality Diversity algorithm – we see that the three dimensions of data sparseness affect the identified solutions differently, though not as distinctly as for supply chain visibility as observed in Chapter 2.

Properties of Model Calibration Techniques

In this dissertation, we evaluate five model calibration techniques: Powell's Method, Approximate Bayesian Computing, Bayesian Optimization, Genetic Algorithm, and the Quality Diversity algorithm. We examine to what extent these algorithms can reconstruct the underlying parameters and structure of the ground truth supply chain when varying the degree of data sparseness.

In Chapter 3 and Chapter 4, we show that Powell's Method and Approximate Bayesian Computing have the lowest quality-of-fit for both parametric and structural uncertainty, when we aim to identify a single optimal solution. The results show that there are multiple input spaces of interest, meaning that different models explain the same outcome, i.e., we observe the principle of equifinality. This causes Powell's Method to get stuck in a local optimum instead of converging to the global optimum (Powell, 1964). For Approximate Bayesian Computing, the technique results in a bimodal distribution if multiple regions of the input space lead to optimality, implying it could get stuck in a local optimum as well (Vrugt & Beven, 2018). Additionally, the highly rugged fitness landscape -– typical for discrete event simulation – makes it increasingly difficult to escape from this local optimum (Azadivar, 1999). We see that Approximate Bayesian Computing seems to perform slightly better for reconstructing the ground truth parameters (Chapter 3) than for reconstructing the underlying structure (Chapter 4).

Bayesian Optimization and Genetic Algorithms both successfully reconstruct the underlying structure of the counterfeit PPE supply chain when increasing the degree of data sparseness for each dimension separately. Genetic Algorithms outperform Bayesian Optimization as it identifies the ground truth most frequently for all dimensions of data sparseness (Chapter 4). Due to the population-based nature of Genetic Algorithms and the use of evolutionary operators such as crossover and mutation, this algorithm is able to cope with all three dimensions of data sparseness (noise, bias, and missing values) relatively well since the algorithm relies on the properties of the population rather than on the individual population members (Slowik & Kwasnicka, 2020). This is in line with Chapter 3 in which Genetic Algorithms has the highest quality-of-fit for identifying the parameters of the supply chain simulation model across an increasing percentage of missing values. In comparison, the main property of Bayesian Optimization is that it uses a Gaussian process to model the distribution of the unknown objective function while balancing exploration and exploitation, making it efficient and relatively fast for sparse data situations (Jalali et al., 2017). However, distinguishing between identifying promising regions and exploring uncertain regions may not be straightforward in a highly rugged fitness landscape in combination with data sparseness.

Although the Genetic Algorithms technique identifies the ground truth more frequently than Bayesian Optimization, both techniques yield a diverse set of optimal graph structures when multiple regions of the input space are of interest as shown in Chapter 3. We recommend using both techniques when calibrating the underlying structure of a supply chain simulation model in the case of sparse data to obtain a comprehensive overview of the various graphs approximating the ground truth. For further research, it would be interesting to develop an algorithm that combines the

exploration and exploitation using the Gaussian process of Bayesian Optimization with the population-based approach of Genetic Algorithm, and evaluate the suitability of this combined method for calibrating the structure of a supply chain simulation model with sparse data.

Chapter 5 shows that the Quality Diversity algorithm is able to reconstruct the underlying supply chain simulation model successfully. The Quality Diversity algorithm used in this chapter (Covariance Matrix Adaptation Evolution Strategy), draws new candidate solutions from a multivariant Gaussian distribution using a covariance matrix between all the input parameters (Fontaine et al., 2020). For the input spaces with fewer parameters (including structural uncertainty), the covariance matrix is simpler, enhancing the exploration of the solutions closer to the ground truth compared to input spaces with more parameters. Chapter 5 shows that the ground truth is most frequently identified for 0% of data sparseness compared to the other scenarios where all dimensions of data sparseness are present. The Quality Diversity algorithm has not been specifically selected for its ability to cope with data sparseness, but rather for creating diversity. For this algorithm, it is necessary to consider the trade-off between the diversity enforced by the algorithm, e.g., solutions that are "all over the place", and the quality of the Quality Diversity front, e.g., limited solutions that are highly optimal. Further research should focus on combining the Quality Diversity algorithm with the properties of Bayesian Optimization and Genetic Algorithms to create an algorithm specifically designed to generate a diverse set of plausible simulation models when only sparse data is available. Additional research could investigate different ways of creating diversity when calibrating simulation models.

Next, in Chapter 5, we only present scenarios of data sparseness when combining the dimensions of data sparseness instead of analyzing the dimensions individually. In Chapter 4, we show that Genetic Algorithms and Bayesian Optimization fail to identify the underlying supply chain structure, when combining the dimensions of data sparseness. For further research, it would be interesting to systematically investigate to which extent Genetic Algorithms, Bayesian Optimization, and Quality Diversity algorithm can cope with the combination of data sparseness dimensions for accurately identifying the ground truth.

Overfitting of Dense Graphs in the Calibration

Chapter 4 concludes that denser graph structures, i.e., more vertices and edges, tend to reproduce the sparseness in the data. Similarly, Chapter 5 shows that graphs with more vertices and edges have a better fit to the sparse data than the ground truth, even when controlling for diversity in the graph density. In both chapters, the model calibration techniques are overfitting in more complex graph structures as they often explain and reproduce the sparse data better than those that slightly differ from the ground truth. With noise and bias, the simulation models of the dense graphs fill in the gaps created by the data sparseness, and for missing values, "anything goes". From a decision-making perspective, this does not necessarily lead to "wrong" results since the graphs are consistent with the sparse data. The results of Chapter 5 still show much overlap between the ground truth and the dense graphs in parts of the supply chain, which is potentially interesting for decision-making (e.g., planning interventions).

Having structures that are much more complex than the ground truth is not desirable from a classical model calibration perspective. The goal of model calibration is to tune the model parameter such that it represents the real system by minimizing the difference between the model data and the observed data. Hence, the results from the model calibration techniques should ideally be close to the ground truth. In essence, a high-quality calibration should describe the sparse data using the smallest set of assumptions, similar to Occam's razor (Hamilton, 1861). This is in line with one of the measures used for the quality-of-fit for the model calibration techniques in Chapter 4 and Chapter 5: graph edit distance, i.e., the smallest set of graph edit operations (e.g., node insertion, edge deletion) needed to transform one graph to the other graph (Abu-Aisheh et al., 2015). Although the graph edit distance is used for analyzing the results of the model calibration techniques in this dissertation, it has not been incorporated in the model calibration process itself to, for example, penalize for complexity. Further research should focus on reviewing the classical way of calibration and its metric of quality in the case of sparse data while considering Occam's razor.

Need for Diversity

This dissertation presents three main reasons that highlight the need for diversity when calibrating a supply chain simulation model when the available data is sparse. First, the highly rugged fitness landscape, which is typical for discrete event simulation models (Azadivar, 1999), makes model calibration techniques sensitive to initialization. The initialization determines the initial sample and the operators for selecting candidate solutions in this highly rugged fitness landscape. Especially for the Quality Diversity algorithm in Chapter 5, the initial sample can differ a lot across the seeds and, consequently, influences the number of function evaluations needed for convergence. Additionally, Chapter 4 shows that Bayesian Optimization and Genetic Algorithms result in a diverse set of solutions across the seeds. Thus, we need to use multiple seeds to capture the diversity of the solutions.

Second, finding a single optimal solution in a highly rugged fitness landscape is challenging, as described in Chapter 3 and Chapter 4. This dissertation shows that there are multiple configurations of the counterfeit PPE supply chain that could represent the real-world supply chain given the sparse data that is available. Relying on a single configuration for gaining insight and choosing the "wrong" one could lead to a "wrong" view of the supply chain activities and, potentially, poor decision-making. Thus, we emphasize the importance of a diverse ensemble of plausible calibrated supply chain simulation models in the case of sparse data instead of a single model.

Third, this dissertation studies the effects of diversity in Chapter 5 by applying the Quality Diversity algorithm. The results of this chapter show that, generally, more data sparseness could lead to more supply chain configurations that are coherent with the sparse data available. Additionally the algorithm identifies many supply chain simulation models that reproduce the sparse data better than the ground truth, where they still have an overlap in characteristics of the supply chain with the ground truth. These results highlight the need for diversity in supply chain simulation modeling, when the available data is sparse.

6.2. PRACTICAL REFLECTION

This section reflects on the following practical implications concerning this dissertation: (1) the importance of the upstream supply chain, (2) the sparseness in real-world data, (3) the generalizability of the research to other supply chains, and (4) the effectiveness of identifying robust interventions.

Importance of the Upstream Supply Chain

This dissertation highlights the importance of gathering upstream information on the supply chain for real-world applications. In Chapter 2, we show that supply-oriented companies have more visibility on the supply chain than demand-oriented companies. This result holds for a push-pull supply chain where upstream actors generally have more inventory than those downstream. In Chapter 5, we demonstrate that the parameters of the source actors in the supply chain, e.g., processing time at manufacturer and warehouse consolidator, have a high impact on the outcomes of the supply chain simulation model. Additionally, the chapter illustrates that most similarities between the ground truth and the supply chains that reproduce the sparse data the best, are in the sourcing part of the supply chain. This is in line with the feature scoring analysis, i.e., the relationship between model inputs and outputs, in Appendix C. The analysis shows that the number of suppliers has the highest impact on the Manhattan (L1) distance, and the number of warehouse consolidators has the highest impact on the transport cost.

For the real-world applications of illicit supply chain going through in the Netherlands, a destination country for international transport, disrupting a smaller number of large shipments from the import (upstream) is supposedly more efficient than disrupting a lot of small shipments at distribution (downstream). Additionally, most information on illicit supply chains is likely focused on the Netherlands, meaning it is demand-oriented instead of supply-oriented. Therefore, it would be valuable for law enforcement to gather more information on the upstream illicit supply chain.

Sparseness in Real-World Data

One of the key novelties of this dissertation is that the degree of data sparseness is systematically varied for comparing its impact on supply chain visibility (Chapter 2), and for evaluating the various model calibration techniques (Chapter 3 to Chapter 5). This allows us to theoretically assess the impact of the dimensions of data sparseness on supply chain analysis with the use of the ground truth set-up. However, the exact percentage of data sparseness per dimension is often unknown in real life. For a real-world application of this dissertation's results, the type and the degree of data sparseness have to be measured or estimated based on the characteristics of the real-world supply chain. For example, criminals try to hide data on their operations as much as possible, resulting in a high degree of missing values (van der Plas, 2022; Mendes, 2023). Also, data can either be sparse by itself or sparse by manipulation, i.e., intentional sparseness (Chapter 2). The intentionality level can be of importance for understanding the sparseness based on the characteristics of the real-world supply chain.

In this dissertation, we focus on the data quality issues for the modification of the values within a dataset, i.e., on numerical sparseness. In real life, datasets often face non-numerical issues, such as timeliness or the relevance of the data set for a certain analysis (Chapter 2). Timeliness is crucial for illicit supply chains, as criminals are opportunistic, which causes information to get rapidly outdated (Anzoom et al., 2021; Ficara et al., 2021). For further research, it would be interesting to assess the impact of these non-numerical dimensions of data sparseness on supply chain visibility, and on model calibration techniques.

Generalizability of the Research to Other Supply Chains

Throughout this dissertation, a stylized counterfeit PPE supply chain is used as a case study. This supply chain is a sequential network, meaning that, for example, there is a one-directional flow between the supplier and the manufacturer. This leads to a direct and linear dependency between the actors within the supply chain, where complexity arises due to the many actors and numerous steps involved in the supply chain. Moreover, the case study is a relatively linear push supply chain up to the warehouse consolidator and a diverged pull supply chain from the warehouse consolidator onwards. The results are generalizable to other supply chains with a similar push-pull structure, and a similar complexity in terms of actors, dependencies between actors, and modalities. It would be interesting to examine whether the results still hold for supply chains with other characteristics, such as a pull-push supply chain, an assembly supply chain, or a circular supply chain.

The product moving through the supply chain is counterfeit PPE, a relatively small and non-perishable product. Although the product is counterfeit, the steps in the supply chain are similar to those of legitimate supply chains. In many cases, the counterfeits even exploit the legitimate supply chain by piggybacking. Hence, the results of this dissertation can be generalized to supply chains of other non-perishable products, such as clothes, tools, chocolate, coffee, or other illicit goods. For real-world applications, the simulation calibration approach of this dissertation can remain the same when using it for other supply chains, but the underlying simulation model needs to be adjusted, and there is no guarantee that the algorithms give the same quality results.

Last, this dissertation primarily focuses on the goods flow in the illicit supply chain. To get a holistic understanding of the supply chain for real-world applications, the communication and the financial flow should also be analyzed. The main challenge of adding these flows to the current counterfeit PPE supply chain simulation model is that it would increase complexity in terms of the number of actors and the number of geographical locations. For example, money can be transferred from locations entirely different from where the goods are moving through, and intermediary companies can be used for information exchange. Moreover, the communication and the financial flow of the illicit supply chain do not always follow similar steps as the legitimate supply chains, making data most likely even sparser.

In summary, this dissertation offers the first insight into generating a diverse ensemble of plausible supply chain simulation models where the available data is sparse. The results of this dissertation are generalizable to a sequential push-pull supply chain for non-perishable products. For practical implications, the simulation calibration approach of this dissertation can potentially be used for other supply chains if the underlying simulation model is modified. A possible extension of the current counterfeit PPE simulation model is adding the communication flows and the financial flows.

Effectiveness for Identifying Robust Interventions

This dissertation offers a first insight into generating a diverse ensemble of reconstructions of a supply chain, in cases where the available data is sparse. The goal of such an ensemble is to support robust decision-making. For the example case of the counterfeit PPE supply chain, this means disrupting the supply chain by identifying robust interventions. This dissertation shows that, in the case of sparse data, it is possible to reconstruct a supply chain simulation model close to or overlapping with the ground truth. In particular, Chapter 5 illustrates that supply chains that reproduce the sparse data the best, have structural similarities with the ground truth. These similarities are potentially interesting for identifying robust interventions and decision-making. However, this dissertation did not explicitly assess the effectiveness of interventions using the ensemble of reconstructions. Hence, the primary follow-up research of this dissertation should focus on explicitly evaluating the effectiveness of a diverse ensemble of reconstructions (this dissertation) for identifying robust interventions. It would be valuable to evaluate this theoretically and in real life.

Due to the dynamic nature of a supply chain, the effectiveness of the diverse ensemble of reconstructions for decision-making highly depends on the moment it is generated. For example, the counterfeit PPE supply chain before, during, and after COVID-19 has constantly changed -– e.g., there was a higher supply during COVID-19, resulting in an increased use of air transport than before, where deep-sea shipping was the default transportation mode (Hashemi et al., 2022). For effective decision-making, the reconstructions of this supply chain have to change accordingly. Additionally, an intervention changes the counterfeit PPE supply chain system as well. For example, when customs control seizes many counterfeit PPE at Boston Logan Airport, the counterfeiters most likely do not send their counterfeits through this airport anymore, and change their import port. This is called the “waterbed” effect, meaning disrupting in one location will not lead to a decrease in the flow of goods, but they will pop up at another location (Klaassen, 2021). For measuring the effectiveness, future research should include the timeliness of the reconstructions and the “waterbed” effect by techniques such as game theory (Aerden, 2023) and adversarial learning.

DELFT

7

CONCLUSION

Even in this digital era, the data required to improve supply chain visibility and to identify robust interventions is often sparse due to the supply chain actors' reluctance to share information. This data sparseness leads to uncertainties about the operations within a supply chain (e.g., inventory, transportation times) as well as about the overall structural composition and geographical locations (e.g., number and location of the actors). Illicit supply chains, in particular, suffer from limited information and a high level of uncertainty, making it challenging to effectively disrupt these supply chains.

Simulation is a way of getting insight into the behavior of complex systems, recognizing relations over time, and exploring future ("what-if") scenarios. Model calibration, i.e., the process of tuning and estimating the simulation model parameters using observed data to match the real system, is essential. However, research on calibrating supply chain simulation models, given a varying degree of data sparseness, is still lacking. Additionally, most research on model calibration in logistics primarily involves adjusting the parameters (i.e., parametric uncertainty) rather than altering the model structure (i.e., structural uncertainty).

When calibrating a simulation model, there is a large variety of plausible simulation models that could explain the sparse observations from the real-world supply chain. Research on how to generate a diverse ensemble of plausible supply chain models that could be used for identifying robust intervention is lacking.

In this dissertation, we investigate how to generate a diverse ensemble of reconstructions of a supply chain that can be used to identify robust interventions, where the model calibration techniques have to deal with sparse data. Throughout this dissertation, we use a simulation calibration approach in combination with a ground truth set-up of a stylized counterfeit Personal Protective Equipment (PPE) supply chain as case study. We extract data from this model, systematically vary the degree of data sparseness, and asses the extent to which various model calibration techniques can still reconstruct the underlying supply chain.

This research is carried out in four steps. First, a classification of data sparseness for supply chain visibility is needed. A literature review is conducted on data sparseness and supply chain visibility, and a quantitative analysis is performed to assess the impact of data sparseness on supply chain visibility. This classification is used in the remaining research steps. Second, we analyze the extent to which various model calibration techniques can identify a parameter of a supply chain model when varying the degree of data sparseness. This step offers a first insight into the quality-of-fit of model calibration techniques in the case of sparse data with parametric uncertainty. Third, we evaluate the quality-of-fit for model calibration techniques for reconstructing a supply chain given structural uncertainty when only sparse data is available. Fourth, we assess the feasibility of the Quality Diversity algorithm for calibrating supply chain simulation models in the case of sparse data, for both parametric and structural uncertainty. The aim of this step is to offer initial insights into the potential of using the Quality Diversity algorithm for generating an ensemble of diverse and plausible configurations of a supply chain simulation model with sparse data.

This chapter answers the research questions introduced in Chapter 1, and provides a general conclusion of this dissertation. Next, an outlook for future research is given, and policy recommendations are listed.

7.1. ANSWERING THE RESEARCH QUESTIONS

The main research question of this dissertation is:

How to generate a diverse ensemble of reconstructions of a supply chain, in cases where the available data is sparse?

The main research question is divided into four sub-questions. We present a conclusion for each of these sub-questions, and end with a general conclusion.

1. **How to classify data sparseness for supply chain visibility?**

   The research for this sub-question provides a classification of data sparseness in the context of supply chains and assesses its impact on supply chain visibility. First, using a systematic literature review, data sparseness can be classified along three dimensions: (1) noise, i.e., values in the data set are distorted, (2) bias, i.e., data is not representative of the population or the phenomenon of study, (3) missing values, i.e., values are missing in the data. Thus, sparse data in relation to supply chain visibility is referred to as: *"lack of data quality across the entire supply chain for the quality dimensions: noise, bias, and missing values, where a certain fraction of data sparseness is intentional"*.

   Next, the impact of sparseness for each of these dimensions on supply chain visibility is evaluated. The main research findings demonstrate that data sparseness greatly affects the visibility of the counterfeit PPE global supply chain, leading to a reduction of visibility up to 52.8% for noise, 65.0% for bias, and 31.7% for missing values. For all three individual dimensions, the average percentage of global supply chain visibility decreases when more sparseness

is added to the data. The missing values dimension has the largest impact on the decrease in supply chain visibility, whereas bias has the least impact. The results show the relative importance of the dimensions of data sparseness for supply chain visibility. In addition, our analysis shows that the location of an actor in the chain who is unwilling to share data (either a competitor or a key actor) makes no difference for the global supply chain visibility percentage when using the calculations in this dissertation. We also show that the demand-oriented scenario has the lowest average global supply chain visibility at 40.6%. A reason is that the global supply chain visibility percentage decreases more when actors with a high average inventory provide sparse data. It also shows that companies with a supply-oriented view will have a better insight into the supply chain visibility than those with a demand-oriented view.

To provide practical advice, the primary impact on supply chain visibility seems to be missing data, suggesting that supply chain practitioners should prioritize addressing missing values to improve supply chain visibility. Additionally, companies with a demand-oriented view should prioritize collecting upstream data as much as possible, to enhance their decision-making capabilities.

2. **To what extent can various model calibration techniques identify the parameters of a supply chain simulation model when varying the degree of data sparseness?**

   The research in this sub-question is a first attempt to analyze the quality-of-fit of model calibration techniques that are likely to be suitable for calibrating simulation models in the case of sparse data. We select a reference technique that is often used for the calibration of simulation models: Powell's Method. We select Genetic Algorithm (GA) and Approximate Bayesian Computing (ABC) as model calibration techniques that seem to be able to handle sparse data. By using a ground truth set-up for evaluating the quality-of-fit, we assess how accurately the three model calibration techniques find the ground truth system parameter values for the simulation model when we apply an increasing degree of data sparseness. The results demonstrate that the selected model calibration techniques are suitable for calibrating the parameter values of the simulation models when faced with sparse data, at least for a linear supply chain with randomly missing values. For our case study, this shows that with sparse data due to COVID-19 and criminals masking their data, the selected model calibration techniques can help to gain insight in underlying counterfeit PPE supply chains.

3. **To what extent can various model calibration techniques reconstruct the underlying structure of a supply chain when varying the degree of data sparseness?**

   The research in this sub-question evaluates the quality-of-fit of various model calibration techniques to reconstruct a supply chain characterized by structural uncertainty and sparse data. We analyze a reference technique, Powell's Method, and three model calibration techniques that promise to be able to handle sparse data: Approximate Bayesian Computing (ABC), Bayesian Optimization (BO), and Genetic Algorithms (GA). For this, we formalize structural uncertainty with System Entity Structures (SES). Our analysis shows that:

   - SES is a powerful approach for defining structural uncertainty in a supply chain simulation model to approximate the ground truth using calibration.
   - Powell's Method and ABC fail to reconstruct the underlying structure of an illicit supply chain for all three dimensions of data sparseness. These algorithms often converge to local optima instead of global ones.
   - GA and BO are suitable for reconstructing the underlying structure of an illicit supply chain for a varying degree of data sparseness individually. For a comprehensive understanding of the various graphs approximating the ground truth, we recommend combining the results of BO and GA.
   - Denser graph structures, i.e., more vertices and edges, tend to describe and reproduce the sparse data the best. Many optimal solutions from the model calibration techniques are, therefore, distant from the ground truth but are not necessarily incorrect. We highlight the need to identify a diverse set of solutions that have a good fit with the sparse data instead of only one solution.

   For the case of the illicit PPE supply chain, reconstructing the underlying structure of this supply chain helps to get insight into the operations of criminals, and it potentially allows law enforcement agencies to effectively plan their interventions.

4. **How feasible is the quality-diversity algorithm for generating a diverse ensemble of reconstructions of a supply chain when varying the degree of data sparseness?**

   The research of this sub-question examines the feasibility of using the Quality Diversity (QD) algorithm for generating a diverse ensemble of supply chain simulation models when the available data is sparse. We assess whether QD can identify the ground truth (the stylized counterfeit PPE supply chain) among the diverse set of solutions, in the case of structural and parametric uncertainty.

   Our analysis demonstrates that QD is able to generate a diverse ensemble of supply chain simulation models. Due to the algorithm's sensitivity to seed initialization, it is crucial to use various seeds. QD identifies the structure of the

ground truth most frequently for 0% of data sparseness. In case of more data sparseness, simulation models of more complex structures, i.e., more vertices and edges than the ground truth, tend to describe and reproduce the sparse data better. These complex structures are not necessarily "wrong", as they show overlap with the ground truth in parts of the supply chain. For parametric uncertainty, the impact of the parameters on the simulation model outcomes, rather than the quantity, influences the quality-of-fit of the solutions of the QD algorithm for identifying the ground truth. Additionally, the results show that, in the case of structural and parametric uncertainty, more data sparseness could lead to more solutions that are coherent with the sparse data. Simulation models with different structures and parameters can reproduce the sparse data better than the ground truth, making it difficult to identify the ground truth when only minimizing the models' data with the observed data – the classical way of calibrating.

The research of this sub-question offers a first insight into the potential of using QD algorithms to generate an ensemble of diverse and plausible configurations of simulation models, particularly for supply chains with sparse data. The results emphasize the importance of gathering information on the upstream supply chain to identify such an ensemble. This could help decision-makers to make more robust decisions on, for example, interventions that are effective for the ensemble of supply chain models rather than for a single explanation of the observed data.

7.2. GENERAL CONCLUSION

This dissertation offers a first insight into generating a diverse ensemble of reconstructions of a supply chain in the case of sparse data using a simulation approach. We highlight three main scientific contributions. First, this dissertation is the first study that systematically varies the degree of data sparseness to evaluate the effectiveness of various model calibration techniques. Although the exact degree of data sparseness is often unknown in real life, this research gives a scientific insight into the impact of the degree of data sparseness on supply chain visibility and supply chain modeling. This set-up allows us to first theoretically assess the quality-of-fit of model calibration techniques before applying them in real life.

Second, this dissertation fills a research gap concerning the calibration of the structure of a supply chain simulation model in addition to fitting its parameters. Especially in the case of (illicit) supply chains, the structural composition and geographical locations are of importance for decision-making. Using the QD algorithm, this research provides a method for creating various structures of a supply chain simulation model by using model calibration, and presents metrics to compare these structures.

Third, the results identify model calibration techniques that are suitable for accurately reconstructing a supply chain characterized by sparse data, for both parametric and structural uncertainty. However, the techniques often overfit to more complex graphs than the true underlying supply chain. Additionally, when calibrating a sim-

ulation model with sparse data, diversity should be included in terms of the use of multiple seeds, the combination of multiple techniques, and the generation of a variety of solutions.

For supply chain practitioners and decision-makers, this dissertation presents three main contributions to practice. First, this dissertation offers insight into data sparseness for supply chain visibility and modeling. It is valuable for supply chain management to have an understanding of how to cope with data sparseness and how this impacts supply chain visibility. Second, this research highlights the importance of gathering information on the upstream supply chain. Third, a contribution of this dissertation to practice is that supply chain practitioners should recognize that there is not a single model for a supply chain when the available data is sparse, but there are multiple feasible models. This could help in making more robust decisions on, for example, effective interventions.

7.3. FUTURE RESEARCH

Based on the chapters in this dissertation and the reflection in Chapter 6, we present suggestions for future research on the model calibration techniques, the possible extensions of this research, and the further application of this research. This dissertation is the first study to evaluate model calibration techniques given a systematically varying degree of data sparseness. Our study has revealed limitations in the applicability of several common model calibration techniques, it has shown the risk of overfitting by the model calibration techniques when using sparse data, and it has highlighted the need for diversity in several dimensions. We recommend the following steps for future research on model calibration techniques that are dependent on sparse data:

- Investigate the extent to which Bayesian Optimization, Genetic Algorithms, and Quality Diversity can cope with the combination of data sparseness dimensions for accurately identifying the ground truth.
- Develop an algorithm that combines the exploration and exploitation phases using the Gaussian process of Bayesian Optimization together with the population-based approach of Genetic Algorithms, and evaluate its suitability for calibrating the underlying structure of a supply chain simulation model with sparse data.
- Combine the properties of Bayesian Optimization, Genetic Algorithms, and the Quality Diversity algorithm to develop an algorithm specifically suitable for generating a diverse ensemble of plausible simulation models when the available data is sparse.
- Investigate different algorithms for creating diversity, next to the Quality Diversity algorithm used in this dissertation.
- Review the metric that defines the quality of the calibration process for a simulation model in cases of data sparseness.

Second, it would be valuable to include additional characteristics of illicit supply chains into the supply chain simulation models and data, and extend the research of this dissertation in the following ways:

- Assess the impact of non-numerical dimensions of data sparseness, e.g., timeliness, on supply chain visibility, and on the model calibration techniques.
- Examine the "waterbed" effect of an illicit supply chain using techniques such as game theory and adversarial learning.

Third, it would be interesting to examine whether the results of this dissertation hold for the following application areas:

- Supply chains with other characteristics, like a pull-push supply chain, an assembly supply chain, or a circular supply chain.
- Different types of simulation models, such as agent-based models or System Dynamics models.

Finally, this dissertation presents a simulation approach for generating a diverse ensemble of reconstructions of an illicit supply chain given sparse data, which is potentially useful for identifying robust interventions. The primary follow-up research should focus on explicitly evaluating the effectiveness of the diverse ensemble of plausible supply chain configurations for identifying these robust interventions.

7.4. POLICY RECOMMENDATIONS

We present policy recommendations that emerge from the practical implications of this dissertation and from the lessons learned when executing this research. First, this dissertation shows the importance of data sparseness for improving supply chain visibility, and it highlights the importance of the upstream supply chain for disrupting illicit supply chains. This leads to the following recommendations:

- Raise awareness among decision-makers about the dimensions of data sparseness and its impact on supply chain visibility.
- Prioritize the reduction of missing values in supply chain data to improve supply chain visibility, and prioritize the collection of data from upstream actors.

Second, this dissertation presents a simulation calibration approach for gaining more insight into illicit supply chains. The following recommendations relate to the application of this approach in real life:

- Examine how the simulation calibration approach from this dissertation can be embedded in the workflow of decision-makers, e.g., law enforcement agencies, to gain more understanding about the illicit supply chain and to make better informed decisions.

- Investigate whether and how to use the simulation calibration approach in the security and law enforcement domain from a legal perspective.
- Collaborate with partners in the security domain, e.g., law enforcement agencies, and with transport alliances, to get more insight into the illicit supply chain in a systematic manner. In particular, it is recommended to collaborate with upstream stakeholders for information collection.

Third, it would be interesting to extend the research of this dissertation to make it better applicable for various applications:

- Integrate communication and financial flows of the supply chain into the simulation model. This adds complexity regarding geographical interactions, such as meetings and financial transactions that occur independently of logistical flows.
- Extend this research to other supply chains with sparse data in different domains, such as illicit supply chains for drugs or human trafficking. Another example could be to study sustainability issues and violations of supply chains for products like chocolate or coffee.
- Study supply chains that merge different domains; for example, persons involved in selling counterfeit PPE likely engage in the sale of other counterfeit items, such as bags.
- Expand this research to consider the possible impact of interventions on livability. While disrupting illicit supply chains reduces the availability of illicit goods in society, it could negatively affect the overall livability of a region or a country. In the case of drug supply chains, increased drug detection might result in increased violence.

Finally, this dissertation successfully demonstrates how to generate a diverse ensemble of reconstructions of a supply chain with sparse data using a simulation approach. The aim of this ensemble is to help identify effective interventions in the case of legal and illicit supply chains. The next crucial step for this research is to evaluate the effectiveness of this diverse ensemble of reconstructions of a supply chain for identifying these robust interventions, both theoretically and in practice.

BIBLIOGRAPHY

Abu-Aisheh, Z., Raveaux, R., Ramel, J.-Y., & Martineau, P. (2015). An exact graph edit distance algorithm for solving pattern recognition problems. In A. Fred, M. De Marsico, & M. Figueiredo (Eds.), *4th International Conference on Pattern Recognition Applications and Methods* (pp. 271–278). https://doi.org/10.5220/0005209202710278

Aerden, V. (2023). *Crime in equilibrium: A study on the criminal supply chain in the port of rotterdam, using simulation and game theory* [M.Sc. Thesis]. Faculty of Technology, Policy and Management, Delft University of Technology. https://resolver.tudelft.nl/uuid:f83b7feb-5d6f-4613-9155-dbc8f9fe4601

Aggarwal, C. C., Hinneburg, A., & Keim, D. A. (2001). On the surprising behavior of distance metrics in high dimensional space. In J. Van den Bussche & V. Vianu (Eds.), *International Conference on Database Theory* (pp. 420–434). https://doi.org/10.1007/3-540-44503-X_27

Agrawal, T. K., Kalaiarasan, R., Olhager, J., & Wiktorsson, M. (2022). Supply chain visibility: A Delphi study on managerial perspectives and priorities. *International Journal of Production Research*, 1–16. https://doi.org/10.1080/00207543.2022.2098873

Anzoom, R., Nagi, R., & Vogiatzis, C. (2021). A review of research in illicit supply-chain networks and new directions to thwart them. *Institute of Industrial and Systems Engineers Transactions*, *54*(2), 134–158. https://doi.org/10.1080/24725854.2021.1939466

Azadivar, F. (1999). Simulation optimization methodologies. In P. Farrington, H. B. Nembhard, D. T. Sturrock, & G. W. Evans (Eds.), *Proceedings of 1999 Winter Simulation Conference* (pp. 93–100). Association for Computing Machinery. https://doi.org/10.1145/324138.324168

Baah, C., Opoku Agyeman, D., Acquah, I. S. K., Agyabeng-Mensah, Y., Afum, E., Issau, K., Ofori, D., & Faibil, D. (2022). Effect of information sharing in supply chains: Understanding the roles of supply chain visibility, agility, collaboration on supply chain performance. *Benchmarking: An International Journal*, *29*(2), 434–455. https://doi.org/10.1108/BIJ-08-2020-0453

Baldissera Pacchetti, M. (2021). Structural uncertainty through the lens of model building. *Synthese*, *198*(11), 10377–10393. https://doi.org/10.1007/s11229-020-02727-8

Banks, J. (1998). *Handbook of simulation: Principles, methodology, advances, applications, and practice*. John Wiley & Sons.

Barratt, M., & Oke, A. (2007). Antecedents of supply chain visibility in retail supply chains: A resource-based theory perspective. *Journal of Operations Management*, *25*(6), 1217–1233. https://doi.org/10.1016/j.jom.2007.01.003

Bartlett, P. A., Julien, D. M., & Baines, T. S. (2007). Improving supply chain performance through improved visibility. *The International Journal of Logistics Management*, *18*(2), 294–313. https://doi.org/10.1108/09574090710816986

Benatia, M. A., Baudry, D., & Louis, A. (2022). Detecting counterfeit products by means of frequent pattern mining. *Journal of Ambient Intelligence and Humanized Computing*, *13*(7), 3683–3692. https://doi.org/10.1007/s12652-020-02237-y

Berger, P. (2024). *New disruptions, geopolitics hang over 2024 supply chains*. Retrieved May 7, 2024, from https://www.wsj.com/articles/new-disruptions-geopolitics-hang-over-2024-supply-chains-6fca8f1e

Bischl, B., Binder, M., Lang, M., Pielok, T., Richter, J., Coors, S., Thomas, J., Ullmann, T., Becker, M., Boulesteix, A.-L., Deng, D., & Lindauer, M. (2023). Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *13*(2), e1484. https://doi.org/10.1002/widm.1484

Boone, T., Ganeshan, R., Jain, A., & Sanders, N. R. (2019). Forecasting sales in the supply chain: Consumer analytics in the big data era. *International Journal of Forecasting*, *35*(1), 170–180. https://doi.org/10.1016/j.ijforecast.2018.09.003

Braak, C. J. T. (2006). A markov chain monte carlo version of the genetic algorithm differential evolution: Easy bayesian computing for real parameter spaces. *Statistics and Computing*, *16*, 239–249. https://doi.org/10.1007/s11222-006-8769-1

Bronselaer, A. (2021). Data quality management: An overview of methods and challenges. In T. Andreasen, G. De Tré, J. Kacprzyk, H. Legind Larsen, G. Bordogna, & S. Zadrożny (Eds.), *Flexible Query Answering Systems* (pp. 127–141). Springer International Publishing. https://doi.org/10.1007/978-3-030-86967-0_10

Brun, A., Karaosman, H., & Barresi, T. (2020). Supply chain collaboration for transparency. *Sustainability*, *12*(11), 4429. https://doi.org/10.3390/su12114429

Busse, C., Schleper, M. C., Weilenmann, J., & Wagner, S. M. (2017). Extending the supply chain visibility boundary: Utilizing stakeholders for identifying supply chain sustainability risks. *International Journal of Physical Distribution & Logistics Management*, *47*(1), 18–40. https://doi.org/10.1108/IJPDLM-02-2015-0043

Calatayud, A., Mangan, J., & Christopher, M. (2019). The self-thinking supply chain. *Supply Chain Management: An International Journal*, *24*(1), 22–38. https://doi.org/10.1108/SCM-03-2018-0136

Caridi, M., Crippa, L., Perego, A., Sianesi, A., & Tumino, A. (2010). Measuring visibility to improve supply chain performance: A quantitative approach. *Benchmarking: An International Journal*, *17*(4), 593–615. https://doi.org/10.1108/14635771011060602

Caridi, M., Perego, A., & Tumino, A. (2013). Measuring supply chain visibility in the apparel industry. *Benchmarking: An International Journal*, *20*(1), 25–44. https://doi.org/10.1108/14635771311299470

Caulkins, J. P. (1993). Local drug markets' response to focused police enforcement. *Operations Research*, *41*(5), 848–863. https://doi.org/10.1287/opre.41.5.848

Cavallaro, L., Ficara, A., De Meo, P., Fiumara, G., Catanese, S., Bagdasar, O., Song, W., & Liotta, A. (2020). Disrupting resilient criminal networks through data analysis: The case of sicilian mafia. *PLOS ONE*, *15*(8), e0236476. https://doi.org/10.1371/journal.pone.0236476

Chatzilygeroudis, K., Cully, A., Vassiliades, V., & Mouret, J.-B. (2021). Quality-diversity optimization: A novel branch of stochastic optimization. In P. Pardalos, V. Rasskazova, & M. Vrahatis (Eds.), *Black Box Optimization, Machine Learning, and No-Free Lunch Theorems* (pp. 109–135). Springer Optimization; Its Applications. https://doi.org/10.1007/978-3-030-66515-9_4

Christopher, M. (2016). *Logistics & Supply Chain Management*. Pearson UK.

Cichy, C., & Rass, S. (2019). An overview of data quality frameworks. *IEEE Access*, *7*, 24634–24648. https://doi.org/10.1109/ACCESS.2019.2899751

Coenen, J., van Der Heijden, R. E., & van Riel, A. C. (2018). Understanding approaches to complexity and uncertainty in closed-loop supply chain management: Past findings and future directions. *Journal of Cleaner Production*, *201*, 1–13. https://doi.org/10.1016/j.jclepro.2018.07.216

Csilléry, K., Blum, M. G., Gaggiotti, O. E., & François, O. (2010). Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution*, *25*(7), 410–418. https://doi.org/10.1016/j.tree.2010.04.001

Cully, A., & Demiris, Y. (2017). Quality and diversity optimization: A unifying modular framework. *Transactions on Evolutionary Computation*, *22*(2), 245–259. https://doi.org/10.1109/TEVC.2017.2704781

De Santis, A., Giovannelli, T., Lucidi, S., Messedaglia, M., & Roma, M. (2022). A simulation-based optimization approach for the calibration of a discrete event simulation model of an emergency department. *Annals of Operations Research*, 1–30. https://doi.org/10.1007/s10479-021-04382-9

Deatcu, C., Folkerts, H., Pawletta, T., & Durak, U. (2018). Design patterns for variability modeling using ses ontology. In A. D' Ambrogio & U. Durak (Eds.), *Proceedings of the Model-driven Approaches for Simulation Engineering Symposium* (pp. 23–43).

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, *6*(2), 182–197. https://doi.org/10.1109/4235.996017

de Groot, L., & Hübl, A. (2021). Developing a calibrated discrete event simulation model of shops of a dutch phone and subscription retailer during covid-19 to evaluate shift plans to reduce waiting times. In S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo, & M. Loper (Eds.), *Proceedings of the 2021 Winter Simulation Conference* (pp. 1–12). Institute of Electrical; Electronics Engineers, Inc.

Diviák, T., Dijkstra, J. K., & Snijders, T. A. (2019). Structure, multiplexity, and centrality in a corruption network: The czech rath affair. *Trends in Organized Crime*, *22*, 274–297. https://doi.org/10.1007/s12117-018-9334-y

Dray, A., Mazerolle, L., Perez, P., & Ritter, A. (2008). Policing australia's 'heroin drought': Using an agent-based model to simulate alternative outcomes. *Journal of Experimental Criminology*, *4*(3), 267–287. https://doi.org/10.1007/s11292-008-9057-1

Duijn, P. A., Kashirin, V., & Sloot, P. M. (2014). The relative ineffectiveness of criminal network disruption. *Scientific Reports*, *4*, 4238. https://doi.org/10.1038/srep04238

Durán, J. M., & Formanek, N. (2018). Grounds for trust: Essential epistemic opacity and computational reliabilism. *Minds and Machines*, *28*, 645–666. https://doi.org/10.1007/s11023-018-9481-6

Ehrlinger, L., & Wöß, W. (2022). A survey of data quality measurement and monitoring tools. *Frontiers in Big Data*, *5*, 850611. https://doi.org/10.3389/fdata.2022.850611

Eser, Z., Kurtulmusoglu, B., Bicaksiz, A., & Sumer, S. I. (2015). Counterfeit supply chains. *Procedia Economics and Finance*, *23*, 412–421. https://doi.org/10.1016/S2212-5671(15)00344-5

Fan, W., & Geerts, F. (2012). *Foundations of Data Quality Management* (3st). Springer Nature. https://doi.org/10.1007/978-3-031-01892-3

Ficara, A., Cavallaro, L., Curreri, F., Fiumara, G., De Meo, P., Bagdasar, O., Song, W., & Liotta, A. (2021). Criminal networks analysis in missing data scenarios through graph distances. *PLOS ONE*, *16*(8), e0255067. https://doi.org/10.1371/journal.pone.0255067

Fisher, M. L. (1997). What is the right supply chain for your product? *Harvard Business Review*, *75*(2), 105–117.

Folkerts, H., Pawletta, T., Deatcu, C., & Zeigler, B. P. (2020). Automated, reactive pruning of system entity structures for simulation engineering. In F. J. Barros, X. Hu, H. Kavak, & A. A. D. Barrio. (Eds.), *Proceedings of the 2020 Spring Simulation Conference* (pp. 1–12). https://doi.org/10.22360/SpringSim.2020.Mod4Sim.001

Fontaine, M., & Nikolaidis, S. (2021). Differentiable quality diversity. *Advances in Neural Information Processing Systems*, *34*, 10040–10052.

Fontaine, M. C., Togelius, J., Nikolaidis, S., & Hoover, A. K. (2020). Covariance matrix adaptation for the rapid illumination of behavior space. *Proceedings of the Genetic and Evolutionary Computation Conference*, 94–102. https://doi.org/10.1145/3377930.3390232

Fox, J. (2015). *Applied Regression Analysis and Generalized Linear Models* (3rd). Sage Publications.

Francis, V. (2008). Supply chain visibility: Lost in translation? *Supply Chain Management: An International Journal*, *13*(3), 180–184. https://doi.org/10.1108/13598540810871226

Frank, M., Laroque, C., & Uhlig, T. (2013). Reducing computation time in simulation-based optimization of manufacturing systems. *Proceedings of the 2013 Winter Simulations Conference*, 2710–2721. https://doi.org/10.1109/WSC.2013.6721642

Gao, J., Xie, C., & Tao, C. (2016). Big data validation and quality assurance–Issuses, challenges, and needs. *Symposium on Service-Oriented System Engineering*, 433–441. https://doi.org/10.1109/SOSE.2016.63

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

GEODIS. (2020). *New research by geodis and accenture interactive reveals ecommerce challenges facing global brands*. Retrieved October 8, 2020, from https://geodis.com/newsroom/new-research-geodis-and-accenture-interactive-reveals-ecommerce-challenges-facing-global

González Ordiano, J. Á., Finn, L., Winterlich, A., Moloney, G., & Simske, S. (2020). On the analysis of illicit supply networks using variable state resolution-markov chains. In M.-J. Lesot, S. Vieira, M. Z. Reformat, J. P. Carvalho, A. Wilbik, B. Bouchon-Meunier, & R. R. Yager (Eds.), *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems* (pp. 513–527). https://doi.org/10.1007/978-3-030-50146-4_38

Gravina, D., Liapis, A., & Yannakakis, G. N. (2018). Quality diversity through surprise. *Transactions on Evolutionary Computation*, *23*(4), 603–616. https://doi.org/10.1109/TEVC.2018.2877215

Grillotti, L., & Cully, A. (2022). Unsupervised behavior discovery with quality-diversity optimization. *Transactions on Evolutionary Computation*, *26*(6), 1539–1552. https://doi.org/10.1109/TEVC.2022.3159855

Grossman, G. M., & Shapiro, C. (1986). Counterfeit-product trade. *The American Economic Review*, *78*(1), 59–75.

Guida, M., Caniato, F., Moretto, A., & Ronchi, S. (2023). The role of artificial intelligence in the procurement process: State of the art and research agenda. *Journal of Purchasing and Supply Management*, 100823. https://doi.org/10.1016/j.pursup.2023.100823

Günther, W. A., Mehrizi, M. H. R., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems*, *26*(3), 191–209. https://doi.org/10.1016/j.jsis.2017.07.003

Hadka, D., & Reed, P. (2013). Borg: An auto-adaptive many-objective evolutionary computing framework. *Evolutionary Computation*, *21*(2), 231–259. https://doi.org/10.1162/EVCO_a_00075

Halim, R. A., Kwakkel, J. H., & Tavasszy, L. A. (2016). A scenario discovery study of the impact of uncertainties in the global container transport system on european ports. *Futures*, *81*, 148–160. https://doi.org/10.1016/j.futures.2015.09.004

Hamilton, W. (1861). *Discussions on Philosophy and Literature, Education and University Reform*. Harper & Brothers.

Hao, J., Ye, W., Jia, L., Wang, G., & Allen, J. (2021). Building surrogate models for engineering problems by integrating limited simulation data and monotonic engineering knowledge. *Advanced Engineering Informatics*, *49*, 101342. https://doi.org/10.1016/j.aei.2021.101342

Hashemi, L., Huang, E., & Shelley, L. (2022). Counterfeit ppe: Substandard respirators and their entry into supply chains in major cities. *Urban Crime. An International Journal*, *3*(2), 74–109. https://doi.org/10.26250/heal.panteion.uc.v3i2.290

Hashemi, L., Jeng, C. C., Mohiuddin, A., Huang, E., & Shelley, L. (2023). Simulating counterfeit personal protective equipment (ppe) supply chains during covid-19. In B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C. Corlu, L. Lee, E. Chew, T. Roeder, & P. Lendermann (Eds.), *Proceedings of the 2022 Winter Simulation Conference* (pp. 522–532). https://doi.org/10.1109/WSC57314.2022.10015398

Hazen, B. T., Boone, C. A., Ezell, J. D., & Jones-Farmer, L. A. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics*, *154*, 72–80. https://doi.org/10.1016/j.ijpe.2014.04.018

Heinrich, B., Hristova, D., Klier, M., Schiller, A., & Szubartowicz, M. (2018). Requirements for data quality metrics. *Journal of Data and Information Quality*, *9*(2), 1–32. https://doi.org/10.1145/3148238

Hermans, B. (2022). *Structural uncertainty in supply chain simulation models* [M.Sc. Thesis]. Faculty of Technology, Policy and Management, Delft University of Technology. https://resolver.tudelft.nl/uuid:e19d2957-eb33-4171-8dc1-8053de3d9e1c

Hofmann, M. (2005). On the complexity of parameter calibration in simulation models. *The Journal of Defense Modeling and Simulation*, *2*(4), 217–226. https://doi.org/10.1177/154851290500200405

Hofmann, M. (2013). Ontologies in modeling and simulation: An epistemological perspective. In A. Tolk (Ed.), *Ontology, Epistemology, and Teleology for Modeling and Simulation: Philosophical Foundations for Intelligent M&S Applications* (pp. 59–87). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-31140-6_3

Hu, X., & Wu, P. (2019). A data assimilation framework for discrete event simulations. *ACM Transactions on Modeling and Computer Simulation*, *29*(3), 1–26. https://doi.org/10.1145/3301502

Huang, Y. (2013). *Automated simulation model generation* [Ph.D. Thesis]. Faculty of Technology, Policy and Management, Delft University of Technology. https://doi.org/10.4233/uuid:dab2b000-eba3-42ee-8eab-b4840f711e37

Ippolito, M., Gregoretti, C., Cortegiani, A., & Iozzo, P. (2020). Counterfeit filtering facepiece respirators are posing an additional risk to health care workers during covid-19 pandemic. *American Journal of Infection Control*, *48*(7), 853. https://doi.org/10.1016/j.ajic.2020.04.020

Jacobs, P. H. M. (2005). *The DSOL Simulation Suite* [Ph.D. Thesis]. Faculty of Technology, Policy and Management, Delft University of Technology. https://doi.org/10.4233/uuid:4c5586e2-85a8-4e02-9b50-7c6311ed1278

Jalali, H., Van Nieuwenhuyse, I., & Picheny, V. (2017). Comparison of kriging-based algorithms for simulation optimization with heterogeneous noise. *European*

Journal of Operational Research, *261*(1), 279–301. https://doi.org/10.1016/j.ejor.2017.01.035

Janssen, M., van der Voort, H., & Wahyudi, A. (2017). Factors influencing big data decision-making quality. *Journal of Business Research*, *70*, 338–345. https://doi.org/10.1016/j.jbusres.2016.08.007

Jeble, S., Dubey, R., Childe, S. J., Papadopoulos, T., Roubaud, D., & Prakash, A. (2018). Impact of big data and predictive analytics capability on supply chain sustainability. *The International Journal of Logistics Management*, *29*(2), 513–538. https://doi.org/10.1108/IJLM-05-2017-0134

Jensen, M., & Dignum, F. (2019). Drug trafficking as illegal supply chain — a social simulation. In P. Ahrweiler & M. Neumann (Eds.), *Advances in Social Simulation: Proceedings of the 15th Social Simulation Conference* (pp. 9–22). https://doi.org/10.1007/978-3-030-61503-1_2

Jones, D. R. (2001). A taxonomy of global optimization methods based on response surfaces. *Journal of Global Optimization*, *21*(4), 345–383. https://doi.org/10.1023/A:1012771025575

Junaid, M., Zhang, Q., Cao, M., & Luqman, A. (2023). Nexus between technology enabled supply chain dynamic capabilities, integration, resilience, and sustainable performance: An empirical examination of healthcare organizations. *Technological Forecasting and Social Change*, *196*, 122828. https://doi.org/10.1016/j.techfore.2023.122828

Kalaiarasan, R., Agrawal, T. K., Olhager, J., Wiktorsson, M., & Hauge, J. B. (2023). Supply chain visibility for improving inbound logistics: A design science approach. *International Journal of Production Research*, *61*(15), 5228–5243. https://doi.org/10.1080/00207543.2022.2099321

Kalaiarasan, R., Olhager, J., Agrawal, T. K., & Wiktorsson, M. (2022). The abcde of supply chain visibility: A systematic literature review and framework. *International Journal of Production Economics*, *248*, 108464. https://doi.org/10.1016/j.ijpe.2022.108464

Kent, P., & Branke, J. (2023). Bayesian quality diversity search with interactive illumination. In L. Paquete (Ed.), *Proceedings of the Genetic and Evolutionary Computation Conference* (pp. 1019–1026). https://doi.org/10.1145/3583131.3590486

Khondoker, M., Dobson, R., Skirrow, C., Simmons, A., & Stahl, D. (2016). A comparison of machine learning methods for classification using simulation with multiple real data examples from mental health studies. *Statistical Methods in Medical Research*, *25*(5), 1804–1823. https://doi.org/10.1177/0962280213502437

Klaassen, R. (2021). *The route of crime: Analysing the impact of risk vs gain tradeoffs on international criminal supply chains* [M.Sc. Thesis]. Faculty of Technology, Policy and Management, Delft University of Technology. https://resolver.tudelft.nl/uuid:c1bce36d-1c92-4657-b410-83b29336fac6

Kollat, J. B., & Reed, P. M. (2007). A computational scaling analysis of multiobjective evolutionary algorithms in long-term groundwater monitoring applications. *Advances in Water Resources*, *30*(3), 408–419.

Kovari, A., & Pruyt, E. (2012). Prostitution and human trafficking: A model-based exploration and policy analysis. In E. Husemann & D. Lane (Eds.), *Proceedings of the 30th International Conference of the System Dynamics Society* (pp. 1–24). https://systemdynamics.org/wp-content/uploads/assets/proceedings/2012/2012proceed.html

Kretschmann, L., & Münsterberg, T. (2017). Simulation-framework for illicit-goods detection in large volume freight. In W. Kersten, T. Blecker, & C. Ringle (Eds.), *Proceedings of the 23th Hamburg International Conference of Logistics* (pp. 427–448). https://hdl.handle.net/10419/209320

Kuipers, L. (2021). *Increasing supply chain visibility with limited data availability: Data assimilation in discrete event simulation* [M.Sc. Thesis]. Faculty of Technology, Policy and Management, Delft University of Technology. https://resolver.tudelft.nl/uuid:5f68b82f-205e-4509-9a64-22082c46065f

Kumar, D., Singh, R. K., Mishra, R., & Wamba, S. F. (2022). Applications of the internet of things for optimizing warehousing and logistics operations: A systematic literature review and future research directions. *Computers & Industrial Engineering*, *171*, 108455. https://doi.org/10.1016/j.cie.2022.108455

Laranjeiro, N., Soydemir, S. N., & Bernardino, J. (2015). A survey on data quality: Classifying poor data. In G. Wang, T. Tsuchiya, & D. Xiang (Eds.), *2015 IEEE 21st Pacific Rim International Symposium on Dependable Computing* (pp. 179–188). IEEE. https://doi.org/10.1109/PRDC.2015.41

Lavastre, O., Gunasekaran, A., & Spalanzani, A. (2014). Effect of firm characteristics, supplier relationships and techniques used on supply chain risk management (SCRM): an empirical investigation on French industrial firms. *International Journal of Production Research*, *52*(11), 3381–3403. https://doi.org/10.1080/00207543.2013.878057

Law, A. M., Kelton, W. D., & Kelton, W. D. (2000). *Simulation Modeling and Analysis* (Vol. 3). McGraw-Hill.

Lawler, E. L., & Wood, D. E. (1966). Branch-and-bound methods: A survey. *Operations Research*, *14*(4), 699–719. https://doi.org/10.1287/opre.14.4.699

Lee, H. L., Padmanabhan, V., & Whang, S. (1997). Information distortion in a supply chain: The bullwhip effect. *Management Science*, *43*(4), 546–558. https://doi.org/10.1287/mnsc.43.4.546

Lee, Y., & Rim, S.-C. (2016). Quantitative model for supply chain visibility: Process capability perspective. *Mathematical Problems in Engineering*, *2016*. https://doi.org/10.1155/2016/4049174

Lehman, J., & Stanley, K. O. (2011). Evolving a diversity of virtual creatures through novelty search and local competition. In N. Krasnogor (Ed.), *Proceedings of the Genetic and Evolutionary Computation Conference* (pp. 211–218). Association for Computing Machinery. https://doi.org/10.1145/2001576.2001606

Lempert, R., Kalra, N., Peyraud, S., Mao, Z., Tan, S. B., Cira, D., & Lotsch, A. (2013). Ensuring robust flood risk management in Ho Chi Minh City. *World Bank Policy Research Working Paper*, (6465), 1–63. https://ssrn.com/abstract=2271955

Lempert, R. J., Popper, S. W., & Bankes, S. C. (2003). *Shaping the next one hundred years: New methods for quantitative, long-term policy analysis*. RAND Corporation.

Lenhard, J., & Winsberg, E. (2010). Holism, entrenchment, and the future of climate model pluralism. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, *41*(3), 253–262. https://doi.org/10.1016/j.shpsb.2010.07.001

Li, Y., Zobel, C. W., Seref, O., & Chatfield, D. (2020). Network characteristics and supply chain resilience under conditions of risk propagation. *International Journal of Production Economics*, *223*, 107529. https://doi.org/10.1016/j.ijpe.2019.107529

Lian, Z., Zhou, Z., Hu, C., Feng, Z., Ning, P., & Ming, Z. (2024). Interpretable large-scale belief rule base for complex industrial systems modeling with expert knowledge and limited data. *Advanced Engineering Informatics*, *62*, 102852. https://doi.org/10.1016/j.aei.2024.102852

Lim, B., Allard, M., Grillotti, L., & Cully, A. (2022). Accelerated quality-diversity for robotics through massive parallelism. *Workshop on Agent Learning in Open-Endedness*. https://openreview.net/forum?id=SUqU4mPbUZ9

Liu, Z., Rexachs, D., Epelde, F., & Luque, E. (2017). A simulation and optimization based method for calibrating agent-based emergency department models under data scarcity. *Computers & Industrial Engineering*, *103*, 300–309. https://doi.org/j.cie.2016.11.036

Macri, K. P. (2018). *Only 6% of companies believe they've achieved full supply chain visibility*. Retrieved February 10, 2024, from https://www.supplychaindive.com/news/supply-chain-visibility-failure-survey-geodis/517751/

Magliocca, N. R., McSweeney, K., Sesnie, S. E., Tellman, E., Devine, J. A., Nielsen, E. A., Pearson, Z., & Wrathall, D. J. (2019). Modeling cocaine traffickers and counterdrug interdiction forces as a complex adaptive system. *Proceedings of the National Academy of Sciences*, *116*(16), 7784–7792. https://doi.org/10.1073/pnas.1812459116

Magliocca, N. R., Price, A. N., Mitchell, P. C., Curtin, K. M., Hudnall, M., & McSweeney, K. (2022). Coupling agent-based simulation and spatial optimization models to understand spatially complex and co-evolutionary behavior of cocaine trafficking networks and counterdrug interdiction. *Institute of Industrial and Systems Engineers Transactions*, 1–14. https://doi.org/10.1080/24725854.2022.2123998

Malleson, N. (2014). Calibration of simulation models. *Encyclopedia of Criminology & Criminal Justice*, *40*, 115–118. https://doi.org/10.1007/978-1-4614-5690-2_688

Marchau, V. A. W. J., Walker, W. E., Bloemen, P. J. T. M., & Popper, S. W. (2019). Introduction. In V. A. W. J. Marchau, W. E. Walker, P. J. T. M. Bloemen, & S. W. Popper (Eds.), *Decision Making under Deep Uncertainty: From Theory to Practice* (pp. 1–20). Springer. https://doi.org/10.1007/978-3-030-05252-2_1

McCrea, B. (2005). EMS completes the visibility picture. *Logistics Management*, *44*(6), 57–61.

Mendes, H. L. (2023). *Guns & roses: Scenario-based analysis of the robustness of police interventions in the flower-oriented sector: Strategies for criminal apprehension* [M.Sc. Thesis]. Faculty of Technology, Policy and Management, Delft University of Technology. https://resolver.tudelft.nl/uuid:04bd0498-8414-43e8-af15-e67b02b3b508

Min, H., & Zhou, G. (2002). Supply chain modeling: Past, present and future. *Computers & Industrial Engineering*, *43*(1-2), 231–249. https://doi.org/10.1016/S0360-8352(02)00066-9

Mirjalili, S. (2019). Genetic algorithm. In *Evolutionary Algorithms and Neural Networks. Studies in Computational Intelligence* (pp. 43–55, Vol. 780). Springer. https://doi.org/10.1007/978-3-319-93025-1_4

Mirkes, E. M., Allohibi, J., & Gorban, A. (2020). Fractional norms and quasinorms do not help to overcome the curse of dimensionality. *Entropy*, *22*(10), 1–31. https://doi.org/10.1145/358598.358601

Mitchell, S. D. (2002). Integrative pluralism. *Biology and Philosophy*, *17*, 55–70. https://doi.org/10.1023/A:1012990030867

Moallemi, E. A., & Köhler, J. (2019). Coping with uncertainties of sustainability transitions using exploratory modelling: The case of the matisse model and the uk's mobility sector. *Environmental Innovation and Societal Transitions*, *33*, 61–83. https://doi.org/10.1016/j.eist.2019.03.005

Moore, C., & Doherty, J. (2005). Role of the calibration process in reducing model predictive error. *Water Resources Research*, *41*(5), 1–14. https://doi.org/10.1029/2004WR003501

Morrison, D. R., Jacobson, S. H., Sauppe, J. J., & Sewell, E. C. (2016). Branch-and-bound algorithms: A survey of recent advances in searching, branching, and pruning. *Discrete Optimization*, *19*, 79–102. https://doi.org/10.1016/j.disopt.2016.01.005

Morselli, C. (2010). Assessing vulnerable and strategic positions in a criminal network. *Journal of Contemporary Criminal Justice*, *26*(4), 382–392. https://doi.org/10.1177/1043986210377105

Mouret, J.-B., & Clune, J. (2015). Illuminating search spaces by mapping elites. *arXiv preprint arXiv:1504.04909*. https://doi.org/10.48550/arXiv.1504.04909

Munir, M., Jajja, M. S. S., Chatha, K. A., & Farooq, S. (2020). Supply chain risk management and operational performance: The enabling role of supply chain integration. *International Journal of Production Economics*, *227*, 107667. https://doi.org/10.1016/j.ijpe.2020.107667

Nag, B., Han, C., & Yao, D.-q. (2014). Mapping supply chain strategy: An industry analysis. *Journal of Manufacturing Technology Management*, *25*(3), 351–370. https://doi.org/10.1108/JMTM-06-2012-0062

Nellemann, C., Stock, J., & Shaw, M. (2018). *World Atlas of Illicit Flows. Global Initiative Against Transnational Organized Crime*. %5Ctextrm % 7Bhttps://globalinitiative.net/wp-content/uploads/2018/09/Atlas-Illicit-Flows-FINAL-WEB-VERSION.pdf%7D

Nikkei Asia. (2020, September 16). Vietnam revamps as 'world's mask factory' to offset COVID hit. Retrieved March 15, 2022, from https://asia.nikkei.com/

Economy/Trade/Vietnam-revamps-as-world-s-mask-factory-to-offset-COVID-hit

Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M.-E., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., et al. (2020). Bias in data-driven artificial intelligence systems — An introductory survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(3), e1356. https://doi.org/10.1002/widm.1356

Oliveira, M. P. V. d., & Handfield, R. (2019). Analytical foundations for development of real-time supply chain capabilities. *International Journal of Production Research*, *57*(5), 1571–1589. https://doi.org/10.1080/00207543.2018.1493240

Oliveira, P., Rodrigues, F., & Henriques, P. (2005). A formal definition of data quality problems. In F. Naumann, M. Gertz, & S. Madnick (Eds.), *Proceedings of the MIT Information Quality Conference* (pp. 1–14). MIT.

Olsson, D. M., & Nelson, L. S. (1975). The nelder-mead simplex procedure for function minimization. *Technometrics*, *17*(1), 45–51. https://doi.org/10.1080/00401706.1975.10489269

Omar, I. A., Debe, M., Jayaraman, R., Salah, K., Omar, M., & Arshad, J. (2022). Blockchain-based supply chain traceability for covid-19 personal protective equipment. *Computers & Industrial Engineering*, *167*, 107995. https://doi.org/10.1016/j.cie.2022.107995

Ören, T. I. (1981). Concepts and criteria to assess acceptability of simulation studies: A frame of reference. *Communications of the Association for Computing Machinery*, *24*(4), 180–189. https://doi.org/10.1145/358598.358605

Page, S. E. (2021). *The model thinker: What you need to know to make data work for you* (2nd ed.). Hachette Book Group.

Park, B., & Qi, H. (2005). Development and Evaluation of a Procedure for the Calibration of Simulation Models. *Transportation Research Record*, *1934*(1), 208–217. https://doi.org/10.1177/0361198105193400122

Park, B., & Schneeberger, J. (2003). Microscopic simulation model calibration and validation: Case study of vissim simulation model for a coordinated actuated signal system. *Transportation Research Record*, *1856*(1), 185–192. https://doi.org/10.3141/1856-20

Parker, W. (2014). Values and uncertainties in climate prediction, revisited. *Studies in History and Philosophy of Science Part A*, *46*, 24–30. https://doi.org/10.1016/j.shpsa.2013.11.003

Parker, W. S. (2013). Ensemble modeling, uncertainty and robust predictions. *Wiley Interdisciplinary Reviews: Climate Change*, *4*(3), 213–223. https://doi.org/10.1002/wcc.220

Pawletta, T., Schmidt, A., Zeigler, B. P., & Durak, U. (2016). Extended variability modeling using system entity structure ontology within matlab/simulink. *Proceedings of the 49th Annual Simulation Symposium*, 22:1–22:8.

Peng, J., Hahn, J., & Huang, K.-W. (2023). Handling missing values in information systems research: A review of methods and assumptions. *Information Systems Research*, *34*(1), 5–26. https://doi.org/10.1287/isre.2022.1104

Pero, M., & Rossi, T. (2014). RFID technology for increasing visibility in ETO supply chains: A case study. *Production Planning & Control*, *25*(11), 892–901. https://doi.org/10.1080/09537287.2013.774257

Pierrot, T., Richard, G., Beguir, K., & Cully, A. (2022). Multi-objective quality diversity optimization. *Proceedings of the Genetic and Evolutionary Computation Conference*, 139–147. https://doi.org/10.1145/3512290.3528823

Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, *45*(4), 211–218. https://doi.org/10.1145/505248.506010

Powell, M. J. (1964). An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The Computer Journal*, *7*(2), 155–162. https://doi.org/10.1093/comjnl/7.2.155

Puchinger, J., & Raidl, G. R. (2005). Combining metaheuristics and exact algorithms in combinatorial optimization: A survey and classification. *First International Work-Conference on the Interplay Between Natural and Artificial Computation*, 41–53. https://doi.org/10.1007/11499305_5

Pugh, J. K., Soros, L. B., Szerlip, P. A., & Stanley, K. O. (2015). Confronting the challenge of quality diversity. In S. Silva (Ed.), *Proceedings of the Genetic and Evolutionary Computation Conference* (pp. 967–974). Association for Computing Machinery. https://doi.org/10.1145/2739480.2754664

Reed, P. M., Hadka, D., Herman, J. D., Kasprzyk, J. R., & Kollat, J. B. (2013). Evolutionary multiobjective optimization in water resources: The past, present, and future. *Advances in Water Resources*, *51*, 438–456. https://doi.org/10.1016/j.advwatres.2012.01.005

Ren, Y., & Wu, Y. (2013). An efficient algorithm for high-dimensional function optimization. *Soft Computing*, *17*(6), 995–1004. https://doi.org/10.1007/s00500-013-0984-z

Reuters. (2023). *Costs from supply chain disruptions drop by over 50% but headwinds remain -survey*. Retrieved February 10, 2024, from https://www.reuters.com/markets/costs-supply-chain-disruptions-drop-by-over-50-headwinds-remain-survey-2023-08-09/

Riesen, K., Ferrer, M., Dornberger, R., & Bunke, H. (2015). Greedy graph edit distance. In P. Perner (Ed.), *International Conference of Machine Learning and Data Mining in Pattern Recognition* (pp. 3–16). https://doi.org/10.1007/978-3-319-21024-7_1

Robinson, S. (2004). Simulation: The practice of model development and use.

Robinson, S. (2005). Discrete-event simulation: From the pioneers to the present, what next? *Journal of the Operational Research Society*, *56*(6), 619–629. https://doi.org/10.1057/palgrave.jors.2601864

Robusto, C. C. (1957). The cosine-haversine formula. *The American Mathematical Monthly*, *64*(1), 38–40. https://doi.org/10.2307/2309088

Rogerson, M., & Parry, G. C. (2020). Blockchain: Case studies in food supply chain visibility. *Supply Chain Management: An International Journal*, *25*(5), 601–614. https://doi.org/10.1108/SCM-08-2019-0300

Roy, V. (2021). Contrasting supply chain traceability and supply chain visibility: Are they interchangeable? *The International Journal of Logistics Management*, *32*(3), 942–972. https://doi.org/10.1108/IJLM-05-2020-0214

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592. https://doi.org/10.1093/biomet/63.3.581

Rydell, C. P., Caulkins, J. P., & Everingham, S. S. (1996). Enforcement or treatment? modeling the relative efficacy of alternatives for controlling cocaine. *Operations Research*, *44*(5), 687–695. https://doi.org/10.1287/opre.44.5.687

Sadegh, M., & Vrugt, J. A. (2014). Approximate bayesian computation using markov chain monte carlo simulation: DREAM (ABC). *Water Resources Research*, *50*(8), 6767–6787. https://doi.org/10.1002/2014WR015386

Sáez, J. A., Galar, M., Luengo, J., & Herrera, F. (2014). Analyzing the presence of noise in multi-class problems: Alleviating its influence with the one-vs-one decomposition. *Knowledge and Information Systems*, *38*(1), 179–206. https://doi.org/10.1007/s10115-012-0570-1

Salazar, J. Z., Reed, P. M., Herman, J. D., Giuliani, M., & Castelletti, A. (2016). A diagnostic assessment of evolutionary algorithms for multi-objective surface water reservoir control. *Advances in Water Resources*, *92*, 172–185. https://doi.org/10.1016/j.advwatres.2016.04.006

Saqib, Z., Saqib, K., & Ou, J. (2019). Role of visibility in supply chain management. In *Modern Perspectives in Business Applications* (pp. 1–14). IntechOpen. https://doi.org/10.5772/intechopen.87202

Schmitt, A., & Singh, M. (2009). Quantifying supply chain disruption risk using monte carlo and discrete-event simulation. In M. Rossetti, R. R. Hill, & B. Johansson (Eds.), *Proceedings of the 2009 Winter Simulation Conference* (pp. 1237–1248). Institute of Electrical; Electronics Engineers, Inc. https://doi.org/10.1109/WSC.2009.5429561

Schneider, L., Pfisterer, F., Thomas, J., & Bischl, B. (2022). A collection of quality diversity optimization problems derived from hyperparameter optimization of machine learning models. In J. E. Fieldsend (Ed.), *Proceedings of the Genetic and Evolutionary Computation Conference Companion* (pp. 2136–2142). https://doi.org/10.1145/3520304.3534003

Schoenthaler, R. (2003). Creating real-time supply chain visibility. *Electronic Business*, *29*(8), 12.

Seiffert, C., Khoshgoftaar, T. M., Van Hulse, J., & Folleco, A. (2014). An empirical study of the classification performance of learners on imbalanced and noisy software quality data. *Information Sciences*, *259*, 571–595. https://doi.org/10.1016/j.ins.2010.12.016

Shannon, R. E. (1998). Introduction to the art and science of simulation. In D. Medeiros, E. F. Watson, J. S. Carson, & M. S. Manivannan (Eds.), *Proceedings of the 1998 Winter Simulation Conference* (pp. 7–14). Institute of Electrical; Electronics Engineers, Inc. https://doi.org/10.1109/WSC.1998.744892

Shelley, L. I. (2018). *Dark commerce: How a new illicit economy is threatening our future*. Princeton University Press.

Slowik, A., & Kwasnicka, H. (2020). Evolutionary algorithms and their applications to engineering problems. *Neural Computing and Applications*, *32*(16), 12363–12379. https://doi.org/10.1007/s00521-020-04832-8

Snaphaan, T., & van Ruitenburg, T. (2024). Financial crime scripting: An analytical method to generate, organise and systematise knowledge on the financial aspects of profit-driven crime. *European Journal on Criminal Policy and Research*, 1–21. https://doi.org/10.1007/s10610-023-09571-9

Sodhi, M. S., & Tang, C. S. (2019). Research opportunities in supply chain transparency. *Production and Operations Management*, *28*(12), 2946–2959. https://doi.org/10.1111/poms.13115

Somapa, S., Cools, M., & Dullaert, W. (2018). Characterizing supply chain visibility - a literature review. *The International Journal of Logistics Management*, *29*(1), 308–339. https://doi.org/10.1108/IJLM-06-2016-0150

Souibgui, M., Atigui, F., Zammali, S., Cherfi, S., & Yahia, S. B. (2019). Data quality in ETL process: A preliminary study. *Procedia Computer Science*, *159*, 676–687. https://doi.org/10.1016/j.procs.2019.09.223

Spreitzenbarth, J. M., Bode, C., & Stuckenschmidt, H. (2024). Artificial intelligence and machine learning in purchasing and supply management: A mixed-methods review of the state-of-the-art in literature and practice. *Journal of Purchasing and Supply Management*, 100896. https://doi.org/10.1016/j.pursup.2024.100896

Srikrishnan, V., & Keller, K. (2021). Small increases in agent-based model complexity can result in large increases in required calibration data. *Environmental Modelling & Software*, *138*, 104978. https://doi.org/10.1016/j.envsoft.2021.104978

Srinivasan, R., & Swink, M. (2018). An investigation of visibility and flexibility as complements to supply chain analytics: An organizational information processing theory perspective. *Production and Operations Management*, *27*(10), 1849–1867. https://doi.org/10.1111/poms.12746

Stadtler, H., & Kilger, C. (2002). *Supply Chain Management and Advanced Planning: Concepts, Models, Software, and Case Studies* (Vol. 4). Springer Berlin Heidelberg.

Suárez, J. L., García, S., & Herrera, F. (2021). A tutorial on distance metric learning: Mathematical foundations, algorithms, experimental analysis, prospects and challenges. *Neurocomputing*, *425*, 300–322. https://doi.org/10.1016/j.neucom.2020.08.017

Swift, C., Guide Jr, V. D. R., & Muthulingam, S. (2019). Does supply chain visibility affect operating performance? Evidence from conflict minerals disclosures. *Journal of Operations Management*, *65*(5), 406–429. https://doi.org/10.1002/joom.1021

Tako, A. A., & Robinson, S. (2008). Model building in system dynamics and discrete-event simulation: A quantitative comparison. *Proceedings of the 2008 International Conference of the System Dynamics Society*, 20–24.

Teng, C.-M. (1999). Correcting noisy data. In I. Bratko & S. Dzeroski (Eds.), *Proceedings of the 16th International Conference on Machine Learning* (pp. 239–248). Morgan Kaufmann Publishers Inc.

The Business Continuity Institute. (2022). *Supply chain resilience report 2021* (Report). Retrieved February 10, 2024, from https://dhlinsights.dhlsupplychain.dhl.com/ao_thought-leadership_social-recommend/report_2021-supply-chain-resilience

Thompson, E. L., & Smith, L. A. (2019). Escape from model-land. *Economics: The Open-Access, Open-Assessment E-Journal*, *13*(40), 1–15. https://doi.org/10.5018/economics-ejournal.ja.2019-40

Tiwari, S., Wee, H.-M., & Daryanto, Y. (2018). Big data analytics in supply chain management between 2010 and 2016: Insights to industries. *Computers & Industrial Engineering*, *115*, 319–330. https://doi.org/10.1016/j.cie.2017.11.017

Tjanaka, B., Fontaine, M. C., Lee, D. H., Zhang, Y., Balam, N. R., Dennler, N., Garlanka, S. S., Klapsis, N. D., & Nikolaidis, S. (2023). Pyribs: A bare-bones python library for quality diversity optimization. *arXiv preprint arXiv:2303.00191*. https://doi.org/10.48550/arXiv.2303.00191

Tolk, A., Page, E. H., Graciano Neto, V. V., Weirich, P., Formanek, N., Durán, J. M., Santucci, J. F., & Mittal, S. (2023). Philosophy and modeling and simulation. In T. Ören, B. P. Zeigler, & A. Tolk (Eds.), *Body of knowledge for modeling and simulation: A handbook by the society for modeling and simulation international* (pp. 383–412). Springer. https://doi.org/10.1007/978-3-031-11085-6_16

Topsector Logistiek. (2019). *Actieagenda Topsector Logistiek 2020-2023. Op weg naar een concurrerende en emissieloze logistiek in Nederland*. Topsector Logistiek. https://topsectorlogistiek.nl/wptop/wp-content/uploads/2020/02/Actieagenda-2020-2023.pdf

Tripepi, G., Jager, K. J., Dekker, F. W., & Zoccali, C. (2010). Selection bias and information bias in clinical research. *Nephron Clinical Practice*, *115*(2), c94–c99. https://doi.org/10.1159/000312871

van der Plas, R. (2022). *Trafficking and trust: Understanding the role of trust in a criminal supply chains* [M.Sc. Thesis]. Faculty of Technology, Policy and Management, Delft University of Technology. https://resolver.tudelft.nl/uuid:a581592f-b006-4b2c-a50a-3655d6bfa28a

van der Zwet, K., Barros, A. I., van Engers, T. M., & van der Vecht, B. (2019). An Agent-Based Model for Emergent Opponent Behavior. In J. M. F. Rodrigues, P. J. S. Cardoso, J. Monteiro, V. V. K. Roberto Lam, M. H. Lees, J. J. Dongarra, & P. M. Sloot (Eds.), *Proceedings of 19th international conference of computational science* (pp. 290–303). https://doi.org/10.1007/978-3-030-22741-8_21

Vanbrabant, L., Martin, N., Ramaekers, K., & Braekers, K. (2019). Quality of input data in emergency department simulations: Framework and assessment techniques. *Simulation Modelling Practice and Theory*, *91*, 83–101. https://doi.org/10.1016/j.simpat.2018.12.002

van Droffelaar, I. S., Kwakkel, J. H., Mense, J. P., & Verbraeck, A. (2024). Simulation-optimization configurations for real-time decision-making in fugitive inter-

ception. *Simulation Modelling Practice and Theory*, *133*, 102923. https://doi.org/10.1016/j.simpat.2024.102923

van Hoof, J., & Vanschoren, J. (2021). Hyperboost: Hyperparameter optimization by gradient boosting surrogate models. *arXiv:2101.02289*.

van Schilt, I. M., Kwakkel, J. H., Mense, J. P., & Verbraeck, A. (2024). Dimensions of data sparseness and their effect on supply chain visibility. *Computers & Industrial Engineering*, *191*, 110108. https://doi.org/10.1016/j.cie.2024.110108

van Schilt, I. M., Kwakkel, J., Mense, J. P., & Verbraeck, A. (2023). Calibrating simulation models with sparse data: Counterfeit supply chains during covid-19. In B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C. Corlu, L. Lee, E. Chew, T. Roeder, & P. Lendermann (Eds.), *Proceedings of the 2022 Winter Simulation Conference* (pp. 496–507). https://doi.org/10.1109/WSC57314.2022.10015241

van Wee, B., & Banister, D. (2016). How to write a literature review paper? *Transport reviews*, *36*(2), 278–288. https://doi.org/10.1080/01441647.2015.1065456

Vassiliades, V., Chatzilygeroudis, K., & Mouret, J.-B. (2017). Using centroidal voronoi tessellations to scale up the multidimensional archive of phenotypic elites algorithm. *Transactions on Evolutionary Computation*, *22*(4), 623–630. https://doi.org/10.1109/TEVC.2017.2735550

Vassiliadis, V. S., & Conejeros, R. (2009). Powell method. In C. A. Floudas & P. M. Pardalos (Eds.), *Encyclopedia of optimization* (pp. 3012–3013). Springer. https://doi.org/10.1007/978-0-387-74759-0_516

Veit, W. (2020). Model pluralism. *Philosophy of the Social Sciences*, *50*(2), 91–114. https://doi.org/10.1177/00483931198948

Vrugt, J. A. (2016). Markov chain monte carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling & Software*, *75*, 273–316. https://doi.org/10.1016/j.envsoft.2015.08.013

Vrugt, J. A., & Beven, K. J. (2018). Embracing equifinality with efficiency: Limits of acceptability sampling using the DREAM (LOA) algorithm. *Journal of Hydrology*, *559*, 954–971. https://doi.org/10.1016/j.jhydrol.2018.02.026

Walker, W. E., Marchau, V. A., & Kwakkel, J. H. (2013). Uncertainty in the framework of policy analysis. In W. A. Thissen & W. E. Walker (Eds.), *Public Policy Analysis* (pp. 215–261). Springer. https://doi.org/10.1007/978-1-4614-4602-6_9

Wamba, S. F., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How 'big data'can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, *165*, 234–246. https://doi.org/10.1016/j.ijpe.2014.12.031

Wang, G., Gunasekaran, A., Ngai, E. W., & Papadopoulos, T. (2016). Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics*, *176*, 98–110. https://doi.org/10.1016/j.ijpe.2016.03.014

Wang, J., & Zhuo, W. (2020). Strategic information sharing in a supply chain under potential supplier encroachment. *Computers & Industrial Engineering*, *150*, 106880. https://doi.org/10.1016/j.cie.2020.106880

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, *12*(4), 5–33. https://doi.org/10.1080/07421222.1996.11518099

Wang, R., Zhang, T., Yu, T., Yan, J., & Yang, X. (2021). Combinatorial learning of graph edit distance via dynamic embedding. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5241–5250. https://doi.org/10.1109/CVPR46437.2021.00520

Webster, M. D., & Sokolov, A. P. (1998). *Quantifying the uncertainty in climate predictions. Program report on the Science and Policy of Global Change*. Massachusetts Institute of Technology.

Wei, H.-L., & Wang, E. T. (2010). The strategic value of supply chain visibility: Increasing the ability to reconfigure. *European Journal of Information Systems*, *19*(2), 238–249. https://doi.org/10.1057/ejis.2010.10

Weisberg, M. (2012). *Simulation and similarity: Using models to understand the world*. Oxford University Press.

Whitley, D. (1994). A genetic algorithm tutorial. *Statistics and Computing*, *4*(2), 65–85. https://doi.org/10.1007/BF00175354

Wigan, M. R. (1972). The fitting, calibration, and validation of simulation models. *Simulation*, *18*(5), 188–192. https://doi.org/10.1177/003754977201800506

Williams, B. D., Roh, J., Tokar, T., & Swink, M. (2013). Leveraging supply chain visibility for responsiveness: The moderating role of internal integration. *Journal of Operations Management*, *31*(7-8), 543–554. https://doi.org/10.1016/j.jom.2013.09.003

Wills, P., & Meyer, F. G. (2020). Metrics for graph comparison: A practitioner's guide. *PLOS ONE*, *15*(2), e0228728. https://doi.org/10.1371/journal.pone.0228728

Wöhling, T., & Vrugt, J. A. (2011). Multiresponse multilayer vadose zone model calibration using markov chain monte carlo simulation and field water retention data. *Water Resources Research*, *47*(4), W04510. https://doi.org/10.1029/2010WR009265

Xie, X. (2018). *Data assimilation in discrete event simulations* [Ph.D. Thesis]. Faculty of Technology, Policy and Management, Delft University of Technology. https://doi.org/10.4233/uuid:d0c47163-3845-430b-a8ce-013c41faa2ea

Yilmaz, L. (2019). Toward self-aware models as cognitive adaptive instruments for social and behavioral modeling. In *Social-Behavioral Modeling for Complex Systems* (pp. 569–586). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119485001.ch24

Zeigler, B. P. (1984). *Multifacetted Modelling and Discrete Event Simulation*. Academic Press Professional, Inc.

Zeigler, B. P., & Hammonds, P. E. (2007). *Modeling and Simulation-Based Data Engineering: Introducing Pragmatics into Ontologies for Net-Centric Information Exchange*. Elsevier.

Zeigler, B. P., Muzy, A., & Kofman, E. (2018). *Theory of Modeling and Simulation: Discrete Event & Iterative System Computational Foundations* (3rd). Academic Press. https://doi.org/10.1016/C2016-0-03987-6

Zeigler, B. P., & Sarjoughian, H. S. (2013). *Guide to Modeling and Simulation of Systems of Systems. Simulation Foundations, Methods and Applications*. Springer.

Zhang, A. N., Goh, M., & Meng, F. (2011). Conceptual modelling for supply chain inventory visibility. *International Journal of Production Economics*, *133*(2), 578–585. https://doi.org/10.1016/j.ijpe.2011.03.003

Zhao, N., Hong, J., & Lau, K. H. (2023). Impact of supply chain digitalization on supply chain resilience and performance: A multi-mediation model. *International Journal of Production Economics*, *259*, 108817. https://doi.org/10.1016/j.ijpe.2023.108817

Zhong, J., & Cai, W. (2015). Differential evolution with sensitivity analysis and the Powell's Method for crowd model calibration. *Journal of Computational Science*, *9*, 26–32. https://doi.org/10.1016/j.jocs.2015.04.013

Zhu, X., & Wu, X. (2004). Class noise vs. attribute noise: A quantitative study. *Artificial Intelligence Review*, *22*(3), 177–210. https://doi.org/10.1007/s10462-004-0751-8

Zhu, X., Wu, X., & Yang, Y. (2004). Error detection and impact-sensitive instance ranking in noisy datasets. *Proceedings of the 19th National Conference on Artificial Intelligence*, 378–384. https://www.aaai.org/Papers/AAAI/2004/AAAI04-061.pdf

A

APPENDIX FOR CHAPTER 4: SUPPLEMENTARY RESULTS

A.1. DETAILS OF SYSTEM ENTITY STRUCTURE

This appendix presents a detailed overview of the specifying rules and distances used for the SES of our study.

Type	**Number**		**Incoming Type**	**Outgoing Type**	**Incoming Degree**		**Outgoing Degree**	
	min	***max***			***min***	***max***	***min***	***max***
Supplier	1	5		Manufacturer	0	0	1	5
Manufacturer	1	5	Supplier	Warehouse Consolidator	1	5	1	6
Warehouse Consolidator	1	10	Manufacturer	Export Port	1	6	1	inf
Export Port	1	21	Warehouse Consolidator	Transit Port	1	10	1	inf
Transit Port	1	10	Export Port	Import Port	1	inf	1	inf
Import Port	1	26	Transit Port	Warehouse Distributor	1	inf	1	inf
Warehouse Distributor	1	5	Import Port	Distributor	1	26	1	inf
Distributor	1	10	Warehouse Distributor	Hospital	1	inf	1	inf
Hospital	1	15	Warehouse Distributor	(Export) Customer	1	inf	1	1
(Export) Customer	1	1	Hospital		1	inf	0	0

Table A.1: Specifying rules for the actor types in the SES.

Table A.1 shows the specifying rules for each type of actor, including the incoming and outgoing actors, to ensure a feasible sequence. Each actor must appear at least once and must be linked to other actors, requiring a minimum count of one for each actor. Additionally, every actor must have incoming and outgoing connections with a minimum degree of one, except for the supplier and export customer, since they represent the start and end of the supply chain. To limit the complexity of the problem, each actor is assigned a maximum value determined through expert interviews. The maximum number of export ports and import ports is determined given the real-world number of ports of interest for the case study, i.e., the origin countries China and Hong Kong, and the destination country, the northeast United States of America (USA). Each actor has an incoming and outgoing type of actor that determines the sequence of the randomly generated supply chain. The maximum value for both incoming and outgoing degrees is typically constrained by the maximum allowable number for the incoming or outgoing actor's type, or it may be unbounded (infinite). Moreover, we make the assumption that there is a single (export) customer serving as the final destination of the supply chain, implying that hospitals have only one outgoing connection.

From	To	Bounds (km)	From	To	Bounds (km)
Export Port	Warehouse Consolidator	[5, 150]	Import Port	Warehouse Distributor	[5, 150]
Warehouse Consolidator	Manufacturer	[50, 300]	Warehouse Distributor	Distributor	[25, 200]
Manufacturer	Supplier	[25, 200]	Distributor	Hospital	[5, 100]

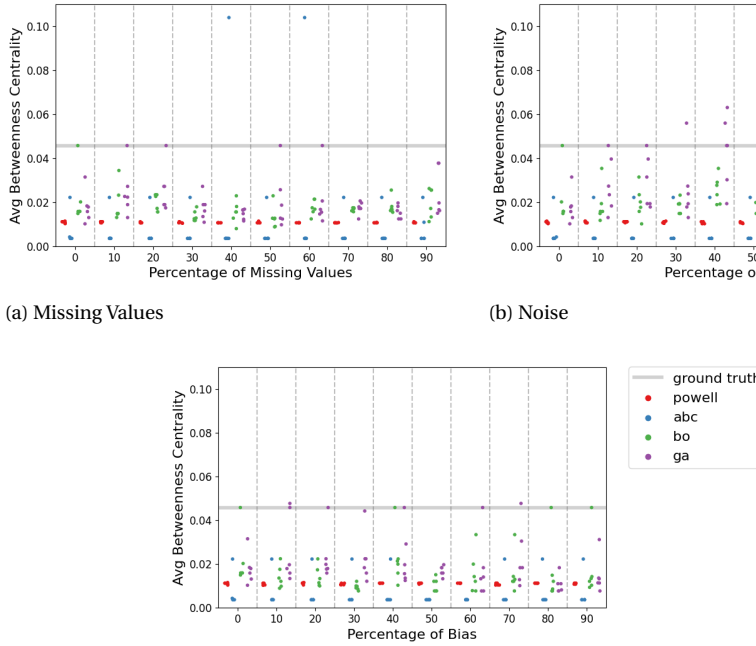
Table A.2: Parameters for determining the travel distance of the land link per actor type pair with the minimum and maximum bounds in kilometers.

The distance for links over land relies on expert interviews, and follows a Uniform distribution with a minimum and maximum distance between actors (see Table A.2). We use the information from open-source data to identify the real-world locations of ports. From there, we determine the positions of other actors based on expert information in ascertaining direction. First, the warehouse consolidator and the warehouse distributor can be next to the port (in this case, 5 kilometers distance) or at driving distance from the port with a maximum of 150 kilometers. On the origin country side (China and Hong Kong), the manufacturer is the most vulnerable location as the PPE products are made here. Therefore, this location is often kept secret and can be far from the warehouse consolidator, e.g., even in another country. For this, the minimum and maximum values of 50 to 300 kilometers are chosen as distance. The supplier can be close to the manufacturer to ensure quick delivery to the manufacturer (e.g., 25 kilometers) or further away with a maximum of 200 kilometers. On the destination country side (northeast USA), the distance between the warehouse distributor and the distributor location can be 25 kilometers to 200 kilometers as the distributor is often located close to a city. The hospital can be next to the distributor (5 kilometers) or in the surroundings of that particular city (up to 100 kilometers).

A.2. RESULTS

This appendix presents the results of the average betweenness centrality and the Manhattan distance for the individual analysis. Also, we show the results of the scenario analysis in a pairplot.

AVERAGE BETWEENNESS CENTRALITY



(a) Missing Values

(b) Noise

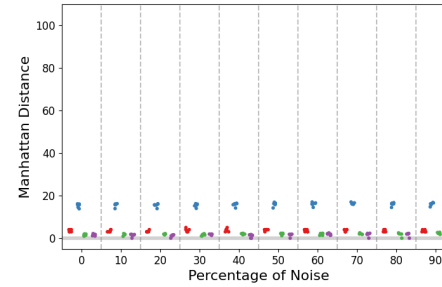
(c) Bias

Figure A.1: Average Betweenness Centrality per Dimension of Data Sparseness for Powell's Method, ABC, BO, GA

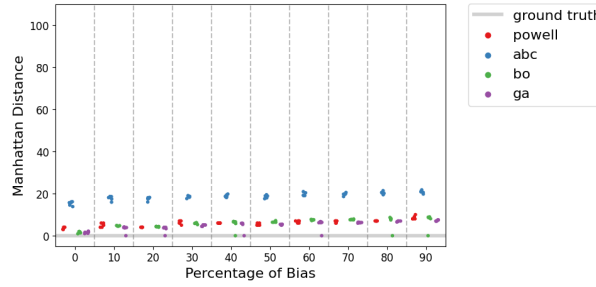
Figure A.1 demonstrates that, generally, the average betweenness centrality of solutions across each techniques tends to be lower than the ground truth. Solutions of ABC result in the lowest average betweenness centrality of 0.005. In Figure A.1a, ABC identifies two outliers that have a high average betweenness centrality of 0.11. Powell's Method has an average centrality betweenness of around 0.01 for all three dimensions of data sparseness. For BO and GA, a diversity of average betweenness centrality of each solution exists between 0.01 and 0.06. The diversity is the highest for the dimension of noise (see Figure A.1b), especially for GA. In this case, many optimal solutions from GA exist that have a higher average betweenness centrality than the ground truth. For noise, BO also identifies solutions with a higher variety of average betweenness centrality.

MANHATTAN DISTANCE

(a) Missing Values



(b) Noise



(c) Bias

Figure A.2: Manhattan Distance per Dimension of Data Sparseness for Powell's Method, ABC, BO, GA

Figure A.2 shows that the Manhattan distance, i.e., the objective value, follows the same trend for each model calibration technique on the three dimensions of data sparseness. For missing values in Figure A.2a, the Manhattan distance increases exponentially when more data is randomly removed. Powell's Method and ABC result in a higher Manhattan distance than BO and GA, meaning a larger gap between the solution and the ground truth with more missing values. For noise and bias, the Manhattan distance of the solutions stays relatively constant over the various percentages of data sparseness. In both cases, ABC has a higher Manhattan distance. Figure A.2c shows a slight increase in Manhattan distance when increasing the percentage of bias.

PAIRPLOT FOR SCENARIOS

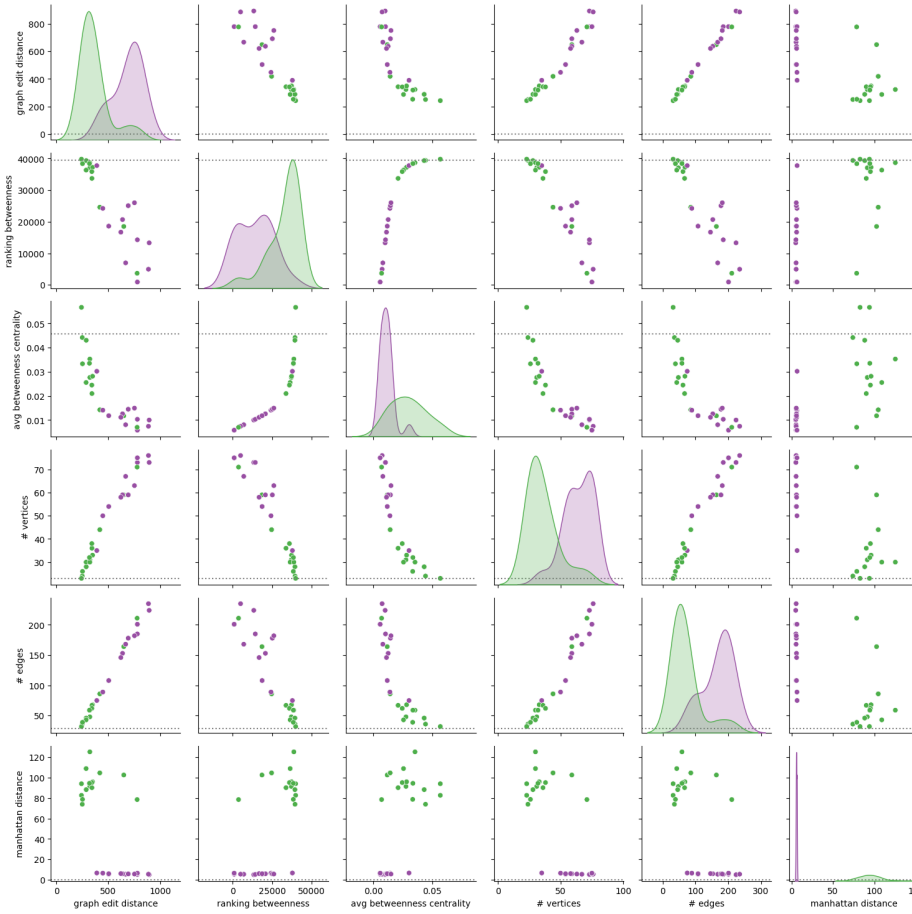


Figure A.3: Pair plots of the scenarios for BO and GA. Green represents the solutions of BO, and purple represents the solutions of GA. Gray dotted line is the ground truth.

Figure A.3 shows that BO and GA identify optimal solutions within distinct subsets of all features. BO identifies optimal graphs with a (ranking of) average betweenness centrality, the number of vertices and edges, closer to the ground truth. Comparatively, GA tends to be more distant from the ground truth, but shows more proximity in terms of Manhattan distance. Moreover, the pair plot shows that a graph with fewer vertices tends to have fewer edges, a lower average betweenness centrality and graph edit distance.

B

APPENDIX FOR CHAPTER 5: SUPPLEMENTARY RESULTS

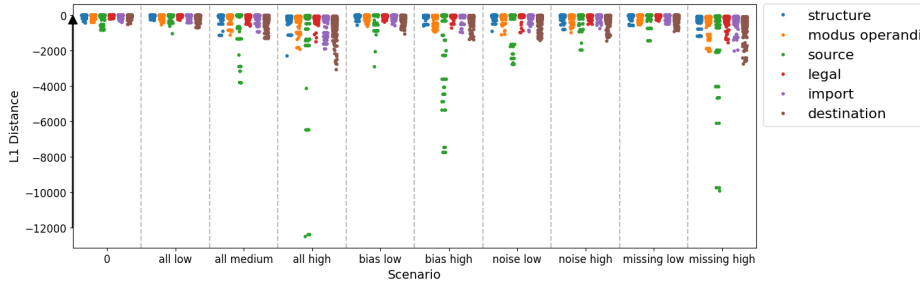
B.1. NORMALIZED L1 DISTANCE

Figure B.1 presents the objective value (Manhattan distance) and the normalized L1 distance over the Quality Diversity (QD) front of all solutions across the ten seeds. We see that the minimum and the maximum Manhattan distance differ for the various scenarios. For example, the source profile shows that with more data sparseness, the Manhattan distance becomes lower, but the solution is not necessarily less optimal for that particular data sparseness scenario. This makes it difficult to compare the QD fronts of the various scenarios on the objective value for the quality-of-fit (van Schilt et al., 2024). Therefore, we normalize the objective value using the minimum and the maximum of the QD front.

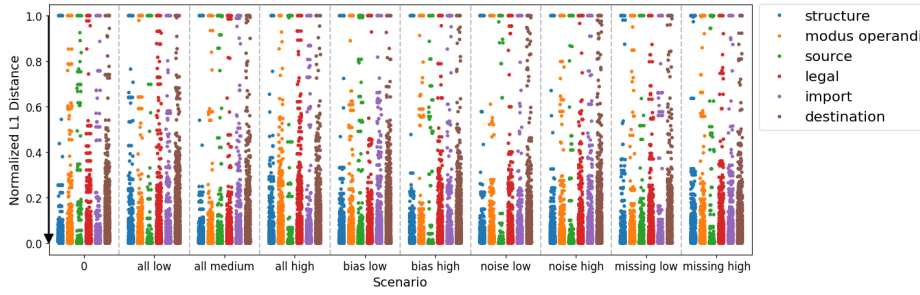
In more detail, Table B.1 shows the minimum and the maximum Manhattan distance for the solutions in the QD front where the ten seeds are merged. We also see that the maximum Manhattan distance the QD front achieved is lower with more data sparseness, especially for the scenario all high and missing high.

Profile	Structure		Modus Operandi		Source		Legal		Import		Destination	
Scenario	*min*	*max*	*min*	*max*	*min*	*max*	*min*	*max*	*min*	*max*	*min*	*max*
0%	-194	0	-795	-12	-96	-1	-50	0	-185	0	-275	-1
All Low	-258	-7	-76	-6	-454	-5	-102	-6	-414	-6	-635	-9
All Medium	-498	-12	-123	-12	-64	-7	-510	-12	-475	-12	-572	-14
All High	-1124	-67	-1035	-66	-1713	-38	-414	-64	-1623	-65	-2566	-68
Bias Low	-341	-11	-306	-10	-41	-6	-155	-11	-319	-11	-741	-13
Bias High	-531	-17	-134	-17	-4073	-9	-317	-16	-482	-17	-1246	-18
Noise Low	-498	-12	-123	-12	-1649	-7	-120	-12	-492	-12	-958	-13
Noise High	-448	-13	-121	-12	-216	-7	-285	-12	-425	-12	-881	-14
Missing Low	-575	-7	-256	-7	-235	-5	-72	-7	-493	-7	-375	-9
Missing High	-1186	-83	-432	-83	-2091	-51	-545	-81	-1194	-78	-2030	-86

Table B.1: Minimum and Maximum Value of the Manhattan (L1) Distance for the Solutions in the QD Front per Profile and per Scenario. We refer to the QD front where the results of the seeds are combined.



(a) Manhattan (L1) Distance Between QD Container and the Ground Truth.



(b) Normalized L1 Distance Between QD Container and the Ground Truth.

Figure B.1: Objective Value of the Solutions of the Ten Seeds.

B.2. PARAMETER VALUES FOR THE IDENTIFIED GROUND TRUTH STRUCTURES

Figure B.2 shows the parameter values of the identified ground truth structures versus the normalized L1 distance per profile across seeds. For each profile, we display the actor's parameters that had to be calibrated. The modus operandi, legal, and import

profile show no correlation between the parameter values, the normalized L1 distance, and whether it fits in the ground truth container or not (so transport cost). For the source profile, there is only one solution that identifies the ground truth structures. This solution has a low normalized L1 distance, so it is relatively close to the ground truth. The interarrival time is relatively low compared to the ground truth, whereas the warehouse consolidator parameters are relatively high compared to the ground truth. This explains the high transport cost. For the destination profile, we see that the higher the warehouse distributor and retailer time, the higher the normalized L1 distance. The figure also shows that even with a marginal difference in the actor's parameter value of the destination, the solution does not fit within the ground truth container.

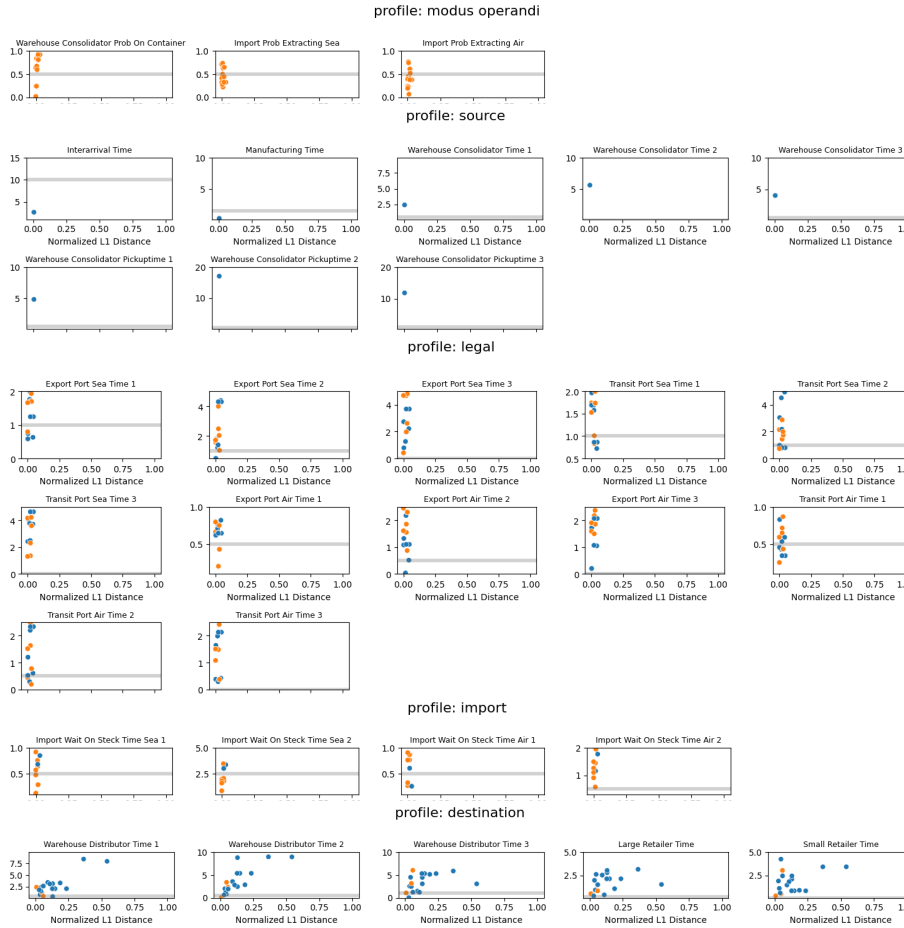
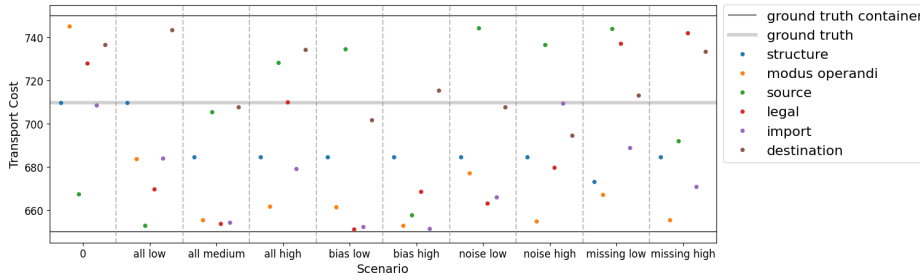


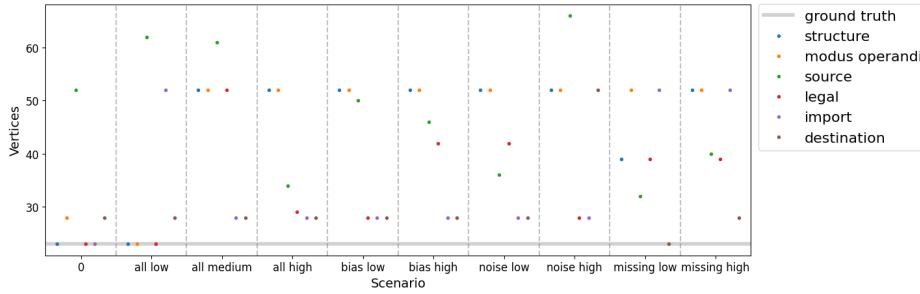
Figure B.2: Parameter Values for the Solutions that Identified the Ground Truth Structure. Orange indicates that the solution fits in the ground truth container, and blue indicates that the solution does not fit in the ground truth container. The grey line indicates the ground truth value.

B.3. ADDITIONAL FEATURES OF SOLUTIONS IN GROUND TRUTH CONTAINER

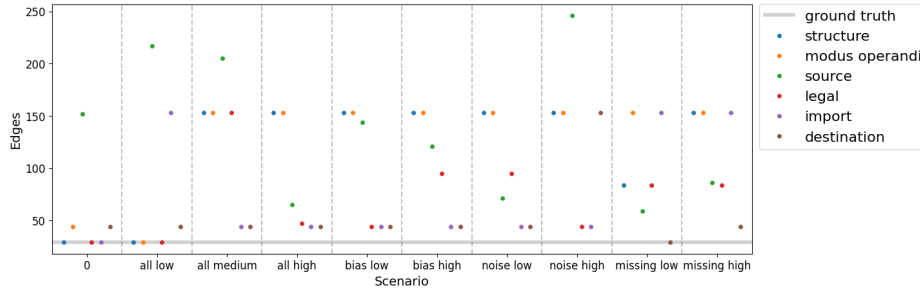
Figure B.3 visualizes the solutions in the ground truth container of the single QD front per scenario and per profile. Figure B.3a shows that most solutions have a lower transport cost than the ground truth. For the profile where only the structure is calibrated, it shows that the transport cost is constant for each graph structure. For all the other profiles, the transport cost also depends on the calibrated parameters and not only on the graph structure. Figure B.3b and Figure B.3c show that all graph structures that are found (not being the ground truth) have a higher number of vertices and edges than the ground truth. The source profile identifies the graph structures with the highest number of vertices and edges.



(a) Transport Cost.



(b) Number of Vertices.



(c) Number of Edges.

Figure B.3: Additional Features of Solutions in Ground Truth Container.

B.4. IMPACT OF THE NUMBER OF SUPPLIERS, MANUFACTURERS, AND WHOLESALES CONSOLIDATORS

Figure B.4 shows the normalized L1 distance for all solutions of the ten seeds for each scenario per the number of suppliers in the graph. It displays that, for each scenario, the more suppliers a graph has, the higher the normalized L1 distance. Also, most solutions seem to have a graph with only one or two suppliers.

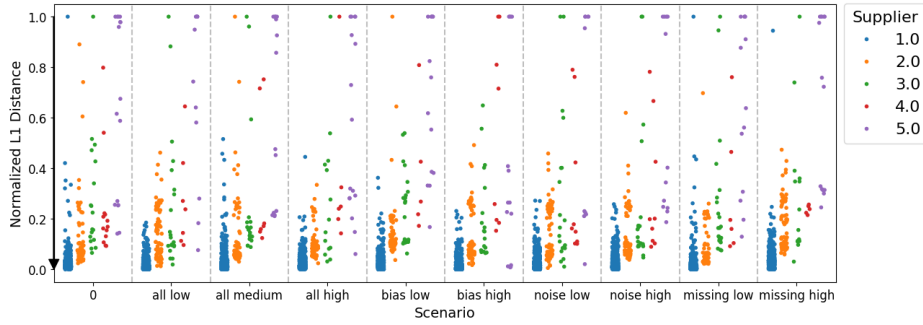


Figure B.4: Normalized L1 Distance per Scenario per Number of Suppliers.

Figure B.5 shows the normalized L1 distance for all solutions of the ten seeds for each scenario per the number of manufacturers in the graph. It displays that most solutions have one to three manufacturers. There is no clear correlation between the number of manufacturers and the normalized L1 distance.

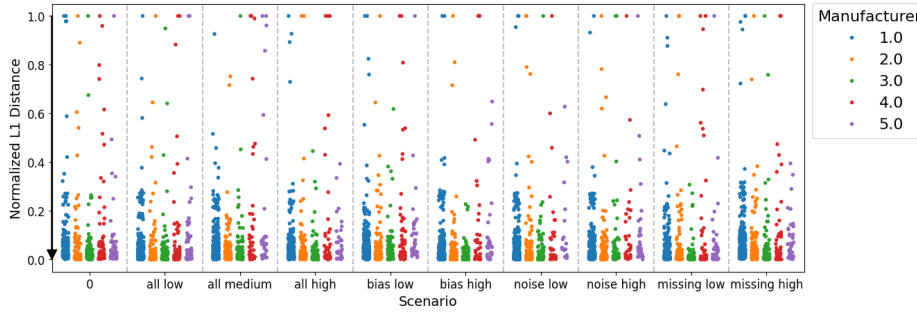


Figure B.5: Normalized L1 Distance per Scenario per Number of Manufacturers.

Figure B.6 shows the normalized L1 distance for all solutions of the ten seeds for each scenario per the number of warehouse consolidators in the graph. It displays that most solutions have one warehouse consolidator with a normalized L1 distance between 0 and 0.3. Also, graphs with more warehouse consolidators lead to a lower spread of the solutions around a normalized distance of 0.

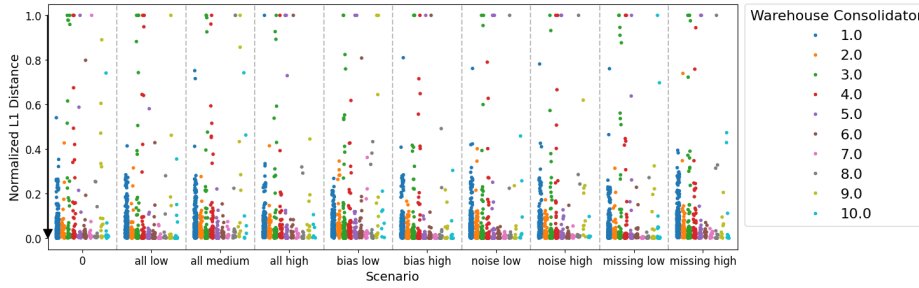


Figure B.6: Normalized L1 Distance per Scenario per Number of Warehouse Consolidators.

C

APPENDIX FOR CHAPTER 6: FEATURE SCORING

The feature scoring analysis identifies the relationship between the model inputs and the model outputs. A high feature score indicates that the model input has a great impact on the model output. For each model output, the model input scores sum up to 1. The synthetic counterfeit PPE supply chain simulation model is used for the feature scoring analysis. As model inputs, we use the actors' input parameters, and graph characteristics such as the number of vertices, the number of edges, and the number of actors (based on the System Entity Structure). As model outputs, we use the Manhattan (L1) distance, the transport cost, and the graph edit distance. More information on the model inputs and outputs can be found in Chapter 4 & 5.

First, the ground truth simulation model is analyzed using feature scoring with 10.000 different variations of the actors' parameters. The underlying supply chain graph does not change, and therefore, the graph edit distance is not included as model output. Figure C.1 shows that all actors' parameters have a similar impact on the Manhattan (L1) distance and on the Key Performance Indicator (KPI) transport cost. The color map displays some slight differences between the parameters in terms of impact (maximum difference around 0.003), and this is seen as negligible.

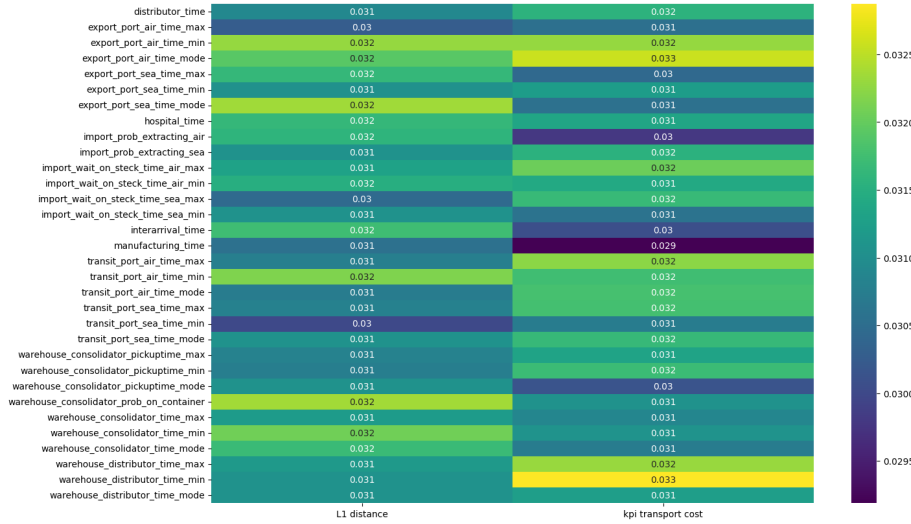


Figure C.1: Feature Scoring Analysis for the Ground Truth.

Second, we analyze the impact of the different supply chain structures on the model outputs while keeping the actors' parameters constant. We run 10.000 different configurations of the structure of the supply chain. Figure C.2 shows that the number of edges and vertices have the highest impact on the graph edit distance. Next, the number of suppliers has the highest impact on the Manhattan (L1) distance, followed by the warehouse distributor. For the transport cost, the number of warehouse consolidators has a very high impact with 0.53. This is followed by the number of suppliers.

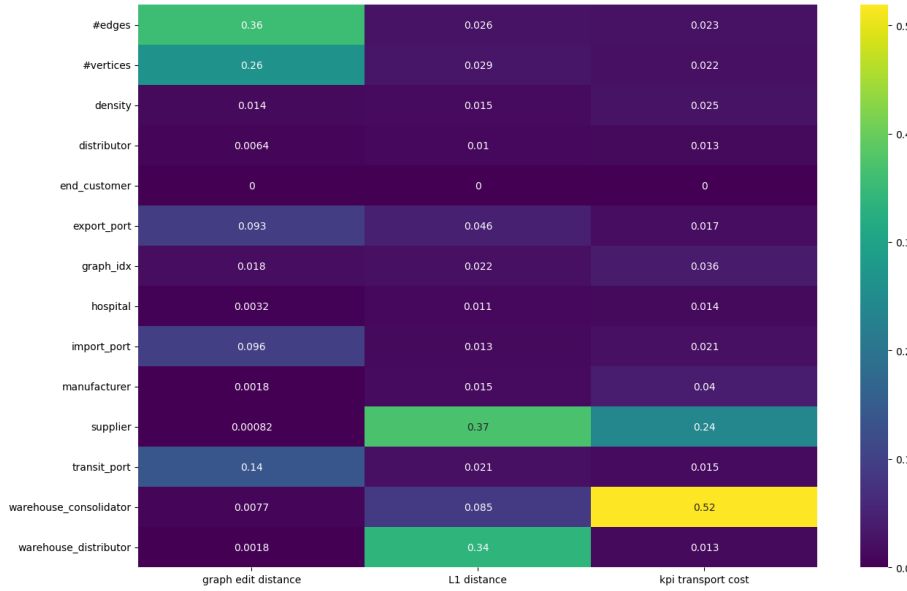


Figure C.2: Feature Scoring Analysis for the Supply Chain Structures.

Third, we analyze the impact of the actors' parameters combined with various supply chain structures on the model outputs. We run 25 different supply chain structures for 5.000 different variations of the actors' parameters, so in total, 125.000 model runs. Similar to the ground truth analysis, Figure C.3 shows that all actors' parameters present a similar relationship for each model output. These parameters have the most impact on the transport cost, with a maximum difference of 0.002. Figure C.4 presents the impact of the graph characteristics on the model outputs. We see that the number of suppliers has a very high impact on the Manhattan (L1) distance. The impact of the number of warehouse distributors has become almost negligible compared to varying only the structure. Moreover, we see that the number of warehouse consolidators has a high impact on transport costs. However, it is much less than when only varying the structure of the supply chain.

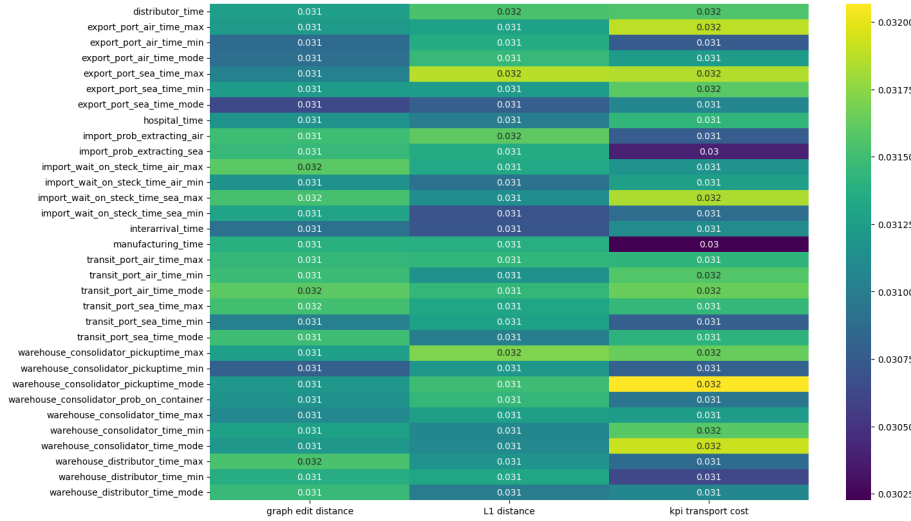


Figure C.3: Feature Scoring Analysis for Actors' Parameters and Supply Chain Structures: Parameters

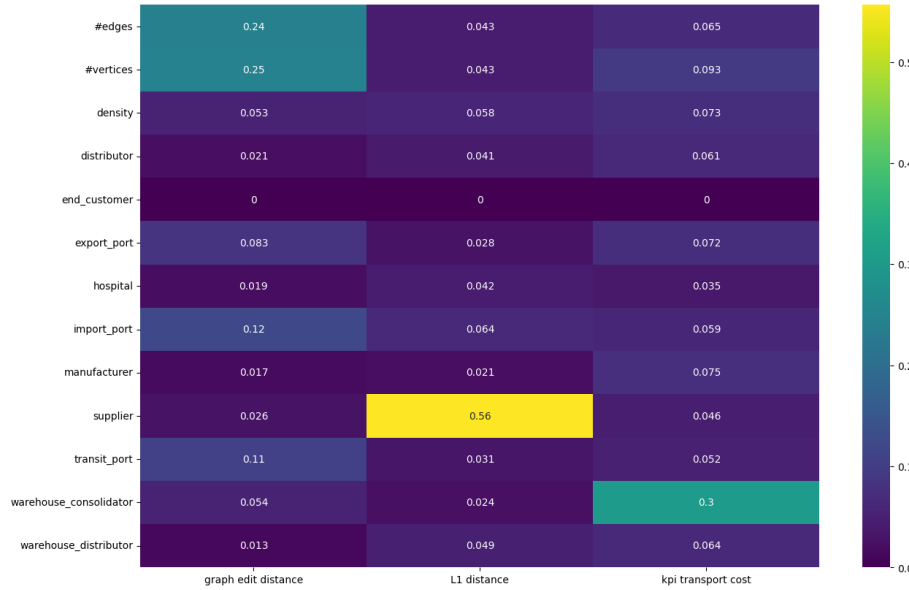


Figure C.4: Feature Scoring Analysis for Actors' Parameters and Supply Chain Structures: Graph Characteristics

AUTHOR BIOGRAPHY

Isabelle was born in Nieuwerkerk aan den IJssel, the Netherlands, on August 16, 1995. During high school, she developed an interest in the combination of technology and management with the ultimate dream of being a true Swiss army knife. She studied her bachelor at the Faculty of Technology, Policy and Management at Delft University of Technology (TU Delft) with a specialization in Transportation and Logistics and a minor in Applied Mathematics. Isabelle continued with the master programme Engineering and Policy Analysis at TU Delft. She represented her fellow master students during a board year at Study Association Curius. During her master, she conducted research on last-mile delivery at Massachusetts Institute of Technology in the Center of Transportation and Logistics. For her thesis, she worked on creating an optimization model for the flight schedule at KLM Royal Dutch Airlines.

Since May 2020, Isabelle is a PhD candidate at the Policy Analysis section of the Faculty of Technology, Policy and Management. Her PhD research focused on a simulation-based approach for improving supply chain visibility with sparse data. Together with her promoter, she developed an efficient and easy-to-use discrete event simulation library in Python. This project is performed in collaboration with the Dutch National Police Artificial Intelligence Lab (NPAI). During the course of the project, Isabelle has worked together with various departments of the Dutch Police and other law enforcement institutions to translate research findings into real-world applications. On the academic collaboration side, she visited the Terrorism, Transnational Crime and Corruption Center (TraCCC) at George Mason University in Washington D.C. in the United States of America. Also, she established long-lasting relationships throughout her studies and visited the Center of Transportation and Logistics at Massachusetts Institute of Technology in Cambridge in the United States of America, and EADA Business School in Barcelona, Spain.

Isabelle is a positive-minded and energetic person who is passionate about data-driven solutions in the field of global supply chains and logistics. She is driven by a systems approach for complex problems using multidisciplinary perspectives and is enthusiastic about inspiring stakeholders. She looks forward to apply innovations to the industry as a connector between stakeholders and engineers, and strives for a smart, sustainable and efficient world.

PUBLICATIONS AND PRESENTATIONS

PUBLICATIONS

RELATED TO THIS THESIS

1. **van Schilt, I. M.**, Kwakkel, J. H., Mense, J. P., & Verbraeck, A. (2024) Identifying the structure of illicit supply chains with sparse data: A simulation model calibration approach. *Advanced Engineering Informatics, 62*, pp. 102926. https://doi.org/10.1016/j.aei.2024.102926

2. **van Schilt, I. M.**, Kwakkel, J. H., Mense, J. P., & Verbraeck, A. (2024) Dimensions of data sparseness and their effect on supply chain visibility. *Computers & Industrial Engineering, 191*, pp. 110108. https://doi.org/10.1016/j.cie.2024.110108

3. **van Schilt, I. M.**, Kwakkel, J. H., Mense, J. P., & Verbraeck, A. (2023) Calibrating simulation models with sparse data: Counterfeit supply chains during COVID-19. In B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C. Corlu, L. Lee, E. Chew, T. Roeder, & P. Lendermann (Eds.), *Proceedings of the 2022 Winter Simulation Conference* (pp. 496–507). https://doi.org/10.1109/WSC57314.2022.10015241

OTHER

1. **van Schilt, I. M.**, van Kalker, J., Lefter, I., Kwakkel, J. H., & Verbraeck, A. (2024) Buffer scheduling for improving on-time performance and connectivity with a multi-objective simulation–optimization model: A proof of concept for the airline industry. *Journal of Air Transport Management, 115*, pp. 102547. https://doi.org/10.1016/j.jairtraman.2024.102547

2. Difrancesco, R. M., **van Schilt, I. M.**, & Winkenbach, M. (2021). Optimal in-store fulfillment policies for online orders in an omni-channel retail environment. *European Journal of Operational Research, 293*(3), 1058-1076. https://doi.org/10.1016/j.ejor.2021.01.007

CONFERENCE PRESENTATIONS

1. **van Schilt, I.M.**, Difrancesco, R. M., & Winkenbach, M. (2024, July). Online orders fulfillment with lateral transshipment in an omni-channel environment: Trading-off economic and environmental sustainability. *European Operations Management Association (EurOMA) Conference*. Barcelona, Spain.

2. **van Schilt, I.M.**, Kwakkel, J.H., Mense, J. P., & Verbraeck, A. (2024, January). Simulation for criminal supply chains. *TechGym*. Delft, The Netherlands. (online)

3. **van Schilt, I.M.**, Kwakkel, J.H., Mense, J. P., & Verbraeck, A. (2023, December). Simulation for criminal supply chains. *Police AI Lab meeting*. Delft, The Netherlands. (poster)

4. **van Schilt, I.M.**, Kwakkel, J.H., Mense, J. P., & Verbraeck, A. (2023, October). Simulation for criminal supply chains. *Disrupting Illicit Supply Chains of Counterfeits Series*. Virginia, USA and Delft, The Netherlands. (online)

5. **van Schilt, I.M.**, & Verbraeck, A. (2023, May). Workshop on Python Distributed Discrete Event Simulation Object Library: pydsol-core & pydsol-model. *Massachusetts Institute of Technology (MIT) Megacity Logistics Lab*. Massachusetts, USA.

6. **van Schilt, I.M.**, Kwakkel, J.H., Mense, J. P., & Verbraeck, A. (2023, May). Supply chain visibility with sparse data using simulation: Counterfeit supply chains during COVID-19. *George Mason University (GMU) Terrorism, Transnational Crime and Corruption Center*. Virginia, USA.

7. **van Schilt, I.M.**, Difrancesco, R. M., & Winkenbach, M. (2023, April). Online orders fulfillment with lateral transshipment in an omni-channel environment: Trading-off economic and environmental sustainability. *International Purchasing & Supply Education & Research Association (IPSERA) Conference*. Barcelona, Spain.

8. **van Schilt, I.M.**, Kwakkel, J.H., Mense, J. P., & Verbraeck, A. (2022, December). Calibrating simulation models with sparse data: Counterfeit supply chains during COVID-19. *Winter Simulation Conference*. Singapore, Singapore.

9. **van Schilt, I.M.**, Kwakkel, J.H., Mense, J. P., & Verbraeck, A. (2022, November). Finding a diverse set of representative criminal supply chain models to identify robust interventions. *Annual Meeting of the Society for Decision Making Under Deep Uncertainty*. Mexico City, Mexico.

10. **van Schilt, I.M.**, Kwakkel, J.H., Mense, J. P., & Verbraeck, A. (2022, September). Structural uncertainty in supply chain simulation models. *Conference of the Netherlands Research School on Transport, Infrastructure and Logistics*. Utrecht, The Netherlands.

11. **van Schilt, I.M.**, Difrancesco, R. M., & Winkenbach, M. (2024, June). Online orders fulfillment with lateral transshipment in an omni-channel environment. *Research Seminar EADA Business School*. Barcelona, Spain.

12. **van Schilt, I.M.**, van Droffelaar, I.S., Kwakkel, J.H., Mense, J. P., & Verbraeck, A. (2022, April). Simulation. *Police AI Lab meeting*. Utrecht, The Netherlands.

13. **van Schilt, I.M.**, Kwakkel, J.H., Mense, J. P., & Verbraeck, A. (2021, November). Simulations for criminal supply chains. *Conference Directorate General of Police and Security Regions, Police, and Police Academy on Research at, for and by the police*. The Hague, The Netherlands. (online)

14. **van Schilt, I.M.**, Kwakkel, J.H., Mense, J. P., & Verbraeck, A. (2021, September). The effect of sparse data on the performance of machine learning techniques for supply chain visibility. *International Conference on Computational Logistics*. Enschede, The Netherlands. (online)

15. **van Schilt, I.M.**, Kwakkel, J.H., Mense, J. P., & Verbraeck, A. (2021, April). Supply chain visibility with sparse data: A conceptualization. *Conference of the Netherlands Research School on Transport, Infrastructure and Logistics*. (online)

SUPERVISED MSC THESES

1. Denekamp, H., **van Schilt, I.M.**, Lefter, I., & Verbraeck, A. (2024) Geospatial Intelligence for Identifying Illicit Supply Chains with Satellite Data and Machine Learning. M.Sc. Thesis. Delft University of Technology. *In Progress.*

2. Lopes Mendes, H., **van Schilt, I.M.**, Torensma, T., Annema, J.A., & Verbraeck, A. (2024) Guns & Roses: Scenario-Based Analysis of the Robustness of Police Interventions in the Flower-Oriented Sector: Strategies for Criminal Apprehension. M.Sc. Thesis. Delft University of Technology. URL: http://resolver.tudelft.nl/uuid:04bd0498-8414-43e8-af15-e67b02b3b508

3. Aerden, V., **van Schilt, I.M.**, van Halem, G.H., Staňková, K., & Verbraeck, A. (2023) Crime in Equilibrium: A study on the criminal supply chain in the Port of Rotterdam, using simulation and game theory. M.Sc. Thesis. Delft University of Technology. URL: http://resolver.tudelft.nl/uuid:f83b7feb-5d6f-4613-9155-dbc8f9fe4601

4. van der Plas, R., **van Schilt, I.M.**, van Halem, G.H., van Rijswijk, E., Stoppelenburg, P., van der Wal, C.N., & Kwakkel, J.H. (2022) Trafficking and Trust: Understanding the role of trust in a criminal supply chain. M.Sc. Thesis. Delft University of Technology. URL: http://resolver.tudelft.nl/uuid:a581592f-b006-4b2c-a50a-3655d6bfa28a

5. Hermans, B., **van Schilt, I.M.**, Huang, Y., & Kwakkel, J.H. (2022) Structural uncertainty in supply chain simulation models: An approach to account for structural uncertainty in supply chain simulation models. M.Sc. Thesis. Delft University of Technology. URL: http://resolver.tudelft.nl/uuid:e19d2957-eb33-4171-8dc1-8053de3d9e1c

6. Klaassen, R., **van Schilt, I.M.**, van den Bosch, R., Baas, H., van der Voort, H.G., Kwakkel, J.H., & Verbraeck, A. (2021) The Route of Crime: Analysing the impact of risk vs gain trade-offs on international criminal supply chains. M.Sc. Thesis. Delft University of Technology. URL: http://resolver.tudelft.nl/uuid:c1bce36d-1c92-4657-b410-83b29336fac6

7. Kuipers, L., **van Schilt, I.M.**, Zandanel, F., Huang, Y., Kwakkel, J.H., & Verbraeck, A. (2021) Increasing supply chain visibility with limited data availability: Data assimilation in discrete event simulation. M.Sc. Thesis. Delft University of Technology. URL: http://resolver.tudelft.nl/uuid:5f68b82f-205e-4509-9a64-22082c46065f

TRAIL THESIS SERIES

The following list contains the most recent dissertations in the TRAIL Thesis Series. For a complete overview of more than 400 titles, see the TRAIL website: www.rsTRAIL.nl.

The TRAIL Thesis Series is a series of the Netherlands TRAIL Research School on transport, infrastructure and logistics.

Schilt, I.M. van, *Reconstructing illicit supply chains with sparse data: A simulation approach*, T2025/2, January 2025, TRAIL Thesis Series, the Netherlands

Ruijter, A. de, *Two-Sided Dynamics in Ridesourcing Markets*, T2025/1, January 2025, TRAIL Thesis Series, the Netherlands

Fang, P., *Development of an Effective Modelling Method for the Local Mechanical Analysis of Submarine Power Cables*, T2024/17, December 2024, TRAIL Thesis Series, the Netherlands

Zattoni Scroccaro, P., *Inverse Optimization Theory and Applications to Routing Problems*, T2024/16, October 2024, TRAIL Thesis Series, the Netherlands

Kapousizis, G., *Smart Connected Bicycles: User acceptance and experience, willingness to pay and road safety implications*, T2024/15, November 2024, TRAIL Thesis Series, the Netherlands

Lyu, X., *Collaboration for Resilient and Decarbonized Maritime and Port Operations*, T2024/14, November 2024, TRAIL Thesis Series, the Netherlands

Nicolet, A., *Choice-Driven Methods for Decision-Making in Intermodal Transport: Behavioral heterogeneity and supply-demand interactions*, T2024/13, November 2024, TRAIL Thesis Series, the Netherlands

Kougiatsos, N., *Safe and Resilient Control for Marine Power and Propulsion Plants*, T2024/12, November 2024, TRAIL Thesis Series, the Netherlands

Uijtdewilligen, T., *Road Safey of Cyclists in Dutch Cities*, T2024/11, November 2024, TRAIL Thesis Series, the Netherlands

Liu, X.,*Distributed and Learning-based Model Predictive Control for Urban Rail Transit Networks*, T2024/10, October 2024, TRAIL Thesis Series, the Netherlands

Clercq, G. K. de, *On the Mobility Effects of Future Transport Modes*, T2024/9, October 2024, TRAIL Thesis Series, the Netherlands

Dreischerf, A.J., *From Caveats to Catalyst: Accelerating urban freight transport sustainability through public initiatives*, T2024/8, September 2024, TRAIL Thesis Series, the Netherlands

Zohoori, B., *Model-based Risk Analysis of Supply Chains for Supporting Resilience*, T2024/7, October 2024, TRAIL Thesis Series, the Netherlands

Poelman, M.C., *Predictive Traffic Signal Control under Uncertainty: Analyzing and Reducing the Impact of Prediction Errors*, T2024/6, October 2024, TRAIL Thesis Series, the Netherlands

Berge, S.H., *Cycling in the age of automation : Enhancing cyclist interaction with automated vehicles through human-machine interfaces*, T2024/5, September 2024, TRAIL Thesis Series, the Netherlands

Wu, K., *Decision-Making and Coordination in Green Supply Chains with Asymmetric Information*, T2024/4, July 2024, TRAIL Thesis Series, the Netherlands

Wijnen, W., *Road Safety and Welfare*, T2024/3, May 2024, TRAIL Thesis Series, the Netherlands

Caiati, V., *Understanding and Modelling Individual Preferences for Mobility as a Service*, T2024/2, March 2024, TRAIL Thesis Series, the Netherlands

Vos, J., *Drivers' Behaviour on Freeway Curve Approach*, T2024/1, February 2024, TRAIL Thesis Series, the Netherlands

Geržinič, N., *The Impact of Public Transport Disruptors on Travel Behaviour*, T2023/20, December 2023, TRAIL Thesis Series, the Netherlands

Dubey, S., *A Flexible Behavioral Framework to Model Mobilityon-Demand Service Choice Preference*, T2023/19, November 2023, TRAIL Thesis Series, the Netherlands

Sharma, S., *On-trip Behavior of Truck Drivers on Freeways: New mathematical models and control methods*, T2023/18, October 2023, TRAIL Thesis Series, the Netherlands

Ashkrof, P., *Supply-side Behavioural Dynamics and Operations of Ride-sourcing Platforms*, T2023/17, October 2023, TRAIL Thesis Series, the Netherlands

Sun, D., *Multi-level and Learning-based Model Predictive Control for Traffic Management*, T2023/16, October 2023, TRAIL Thesis Series, the Netherlands

Brederode, L.J.N., *Incorporating Congestion Phenomena into Large Scale Strategic Transport Model Systems*, T2023/15, October 2023, TRAIL Thesis Series, the Netherlands

Hernandez, J.I., D*ata-driven Methods to study Individual Choice Behaviour: with applications to discrete choice experiments and Participatory Value Evaluation experiments*, T2023/14, October 2023, TRAIL Thesis Series, the Netherlands

Aoun, J., *Impact Assessment of Train-Centric Rail Signaling Technologies*, T2023/13, October 2023, TRAIL Thesis Series, the Netherlands

Pot, F.J., *The Extra Mile: Perceived accessibility in rural areas*, T2023/12, September 2023, TRAIL Thesis Series, the Netherlands

TRAIL

Summary

Even in this digital era, the data for improving supply chain visibility is often sparse due to the actors' reluctance to share information or because of illicit activities. This dissertation demonstrates a simulation approach for reconstructing illicit supply chains with sparse data, that could help robust decision-making on effective interventions. We use a simulation calibration approach in combination with a ground truth simulation model of a stylized counterfeit Personal Protective Equipment (PPE) supply chain as case study.

About the Author

Isabelle M. van Schilt holds an MSc degree in Engineering and Policy Analysis from Delft University of Technology. She conducted her PhD at the faculty of Technology, Policy and Management, Delft University of Technology, as part of the National Police Artificial Intelligence Lab (NPAI).

TRAIL Research School ISBN 978-90-5584-358-9

Erasmus University Rotterdam · Radboud University · rijksuniversiteit groningen · TU Delft Delft University of Technology · University of Twente. · TU/e Technische Universiteit Eindhoven University of Technology