Identifying the structure of illicit supply chains with sparse data
A simulation model calibration approach

van Schilt, Isabelle M.; Kwakkel, Jan H.; Mense, Jelte P.; Verbraeck, Alexander

DOI
10.1016/j.aei.2024.102926

Publication date
2024

Document Version
Final published version

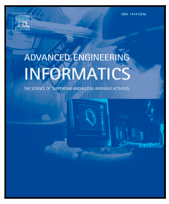
Published in
Advanced Engineering Informatics

Advanced Engineering Informatics 62 (2024) 102926

Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei



Identifying the structure of illicit supply chains with sparse data: A simulation model calibration approach

Isabelle M. van Schilt [a], Jan H. Kwakkel [a,*], Jelte P. Mense [b], Alexander Verbraeck [a]

[a] Faculty of Technology, Policy and Management, Delft University of Technology, Jaffanlaan 5, Delft, 2628 BX, The Netherlands
[b] National Policelab AI, Utrecht University, Princetonplein 5, Utrecht, 3584 CC, The Netherlands

ARTICLE INFO

Keywords:
Illicit
Supply chain
Structural uncertainty
Sparse data
Simulation

ABSTRACT

Illicit supply chains for products like counterfeit Personal Protective Equipment (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 that have shown promise 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 extract data from this ground truth and systematically vary its sparseness. 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.

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 [1]. 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 fraudulent organizations [2]. A significant number of fraudulent PPE manufacturers entered the market during the initial stages of COVID-19, trying to sell non-certified PPE as certified PPE [3]. 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 [4]. Detecting counterfeit PPE has been challenging as little historical data on COVID-19 is available, and fraudulent organizations obfuscate their data as much as possible. Consequently, criminal activities and the related logistics operations remain largely invisible [5]. 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 [6]. 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) [7,8]. 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 [9]. 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 [10].

Especially in the case of illicit supply chains, the operational and logistical structure, including geographical boundaries, is often not

* Corresponding author.
E-mail address: j.h.kwakkel@tudelft.nl (J.H. Kwakkel).

https://doi.org/10.1016/j.aei.2024.102926
Received 12 June 2024; Received in revised form 21 October 2024; Accepted 30 October 2024
Available online 15 November 2024

known to law enforcement [8]. 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 [10,11]. 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 [12,13]. 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 [14,15]. This paper focuses on the use of discrete event simulation models for understanding illicit supply chains [16,17]. 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 [18–20].

In the case of simulating illicit supply chains, model calibration should be able to handle sparse observed data [8,10,21]. Three dimensions of data sparseness are defined: (1) noise, (2) bias, and (3) missing values [22]. 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 [5,23,24]. 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 [25]. 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 [26,27].

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 paper 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.

Our paper is structured as follows. Section 2 presents the related work on modeling and simulation for illicit supply chains. Section 3 discusses the current state-of-the-art literature on model calibration and its challenges when dealing with sparse data. Section 4 describes the design of experiments, the simulation model of the case study, and the configuration of the selected model calibration techniques. Section 5 shows the quality-of-fit for reconstructing the structure of the illicit supply chain using the simulation model calibration approach. Section 6 discusses our results. Section 7 concludes our work and provides directions for further research.

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.

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 [5,10,17]. Anzoom et al. [10] 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. [11] 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. [28] 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 [29,30]. More recent studies focus on developing more detailed simulation models to understand disruption strategies in a specific supply chain. For instance, Dray et al. [31] develop an agent-based model for interaction between individuals and the supply in the heroin market. Kovari and Pruyt [32] create a system dynamic simulation model of human trafficking for evaluating the effect of policy interventions in the Netherlands. Kretschmann and Münsterberg [33] present a discrete event simulation model for testing one specific detection method at the border.

Specifically on trafficking, Magliocca et al. [9] 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 [34] 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. [35] 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. [36] 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. [17] 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. [3] 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. [5] 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. [5], 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.

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 observable. 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 [8,37,38].

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 [39–41]. 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 [25]. 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 [27,42,43]. Especially in the case of illicit supply chains, the structure is often uncertain [5]. Therefore, the innovative contribution of this research is addressing structural uncertainty for simulation models related to illicit supply chains.

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.

3.1. Related work

Few studies have investigated the calibration of simulation models in the context of data sparseness. Liu et al. [23] 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. [44] 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 [24] 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. [45] uses evolutionary neural networks to build more accurate surrogate simulation models with limited data. van Schilt et al. [5] 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.

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 [18–20]. As model calibration aims to minimize the difference between the model data and the observed data, optimization techniques are commonly used for this purpose [23,46]. 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.

Fig. 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.

Deterministic mathematical solvers calibrate models through deterministic mathematical optimization that guarantees to discover (local or global) optimal solutions [47]. A commonly used deterministic algorithm for calibrating simulation models is Powell's Method [23]. Powell's Method is a gradient-free minimization algorithm using a repeated line search introduced by Powell [48]. 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 [49]. Another example of a deterministic mathematical solver for model calibration is the Nelder–Mead Simplex algorithm [50]. Moreover, the Branch-and-Bound algorithm is also commonly used for optimizing linear or mixed-integer programs [51,52].

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) [53]. GA are widely applied in the field of model calibration, especially in high-dimensional problems where data is often sparse [54–56]. 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 [57]. A classic

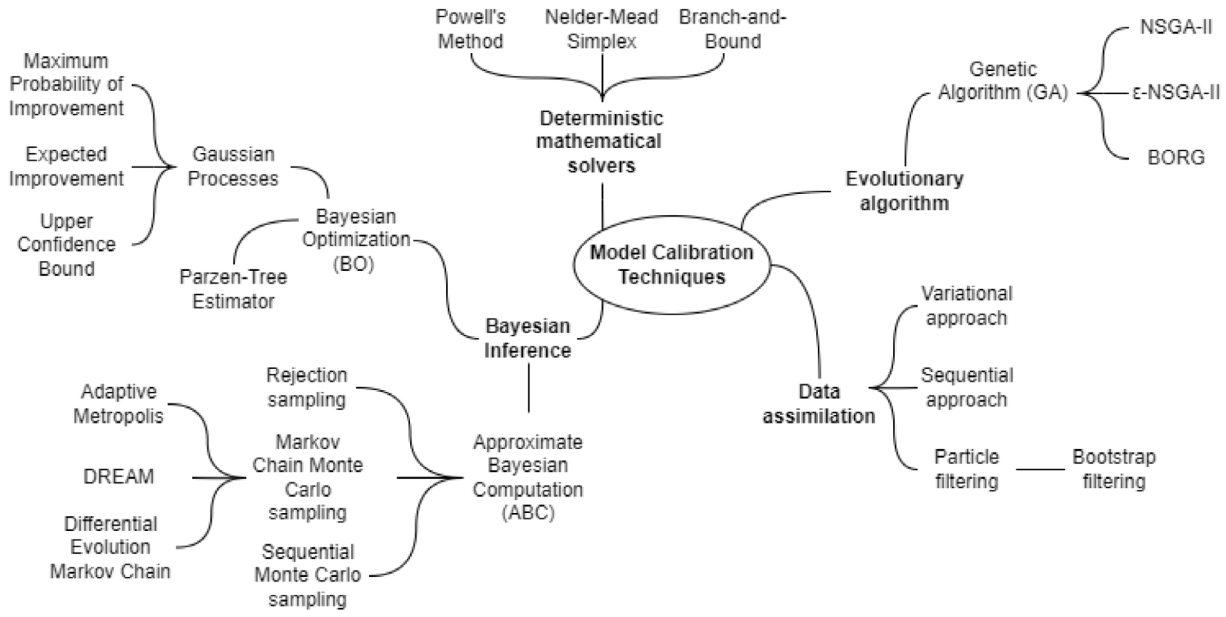


Fig. 1. Overview of model calibration families and techniques for sparse data.

and popular algorithm is Non-dominated Sorting Genetic Algorithm II (NSGA-II) [58,59]. 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 [60]. For more complex and multi-objective problems, BORG is a suitable algorithm [61]. 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 [62].

Bayesian inference uses Bayes' theorem to calibrate models. Model calibration is a core application of Bayesian data analysis [63]. Approximate Bayesian Computing (ABC) is one of most suitable techniques for handling sparse data and uncertainties due to its likelihood-free nature [64]. 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 [63]. An algorithm for ABC is the Differential Evolution Markov Chain algorithm that combines an evolutionary algorithm with Markov Chain Monte Carlo sampling [65]. Another algorithm for ABC with Markov Chain Monte Carlo sampling is Adaptive Metropolis [66]. This algorithm updates the Gaussian distribution for sampling using the information gathered so far in the process. Sadegh and Vrugt [67] 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 [67].

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 [68]. 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 [69]. 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 [70,71].

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 [72,73]. There are three approaches for data assimilation for discrete event simulation models: (1) variational approach, (2) sequential approach, and (3) particle filtering [74]. 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 [74].

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.

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 [25]. 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 [25]. However, three primary drawbacks are encountered in our study when using this implementation: (1) designing a fully comprehensive simulation

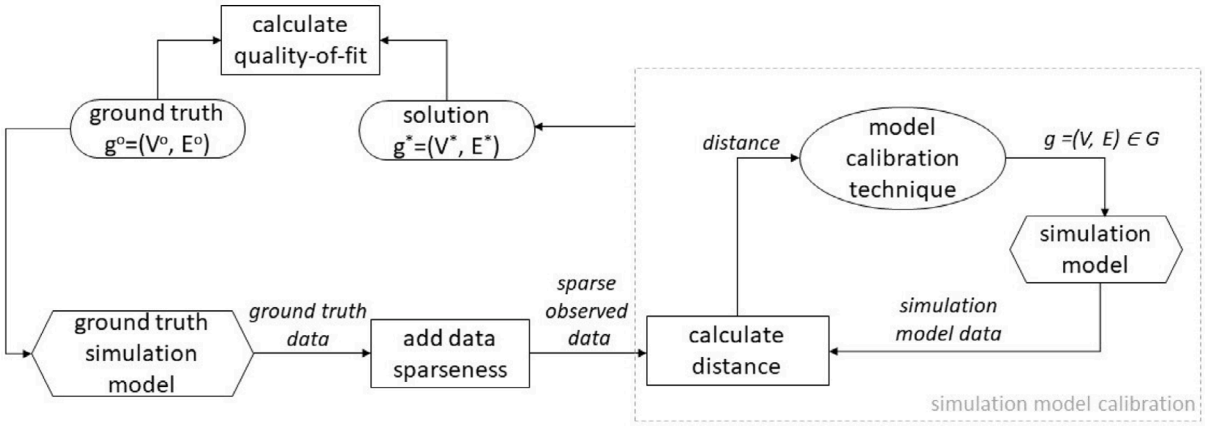


Fig. 2. Method for evaluating calibration of a simulation model on the graph with sparse data.

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 [75]. 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 [75,76]. 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 [75]. 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 [77,78]. 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 [77,79].

Zeigler [80] 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 [75]. 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 [81]. Automated pruning for a specific system requires knowledge on the degrees of freedom to ensure valid system configurations and thus, valid simulation models [75]. 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 [82–84].

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. 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.

4.1. 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.

4.1.1. 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 [85].

Fig. 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. [22]. This results in *sparse observed 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.

4.1.2. 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 [86]. 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 [87]. The graph edit distance defines the cheapest set of graph edit operations (e.g., node insertion, edge deletion) needed to transform one graph to the other graph [88]. For computational reasons, we use an approximated greedy graph edit distance of Riesen et al. [89] by transposing this problem to an assignment problem. The python library *GMatch4py*[1] is used.

4.1.3. Experiments

The steps in Fig. 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.

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

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.

4.2. Configuration of the model calibration techniques

Recalling the selected model calibration techniques for this research in Section 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 [90,91]. Our study requires a metric that considers stochastic models and sparse observed data in complex, high-dimensional systems. According to Mirkes et al. [92], 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.

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 [93]. 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.

4.3. Formalization of ground truth simulation model

The case study used for the ground truth simulation model is a stylized counterfeit PPE supply chain. Fig. 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).

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

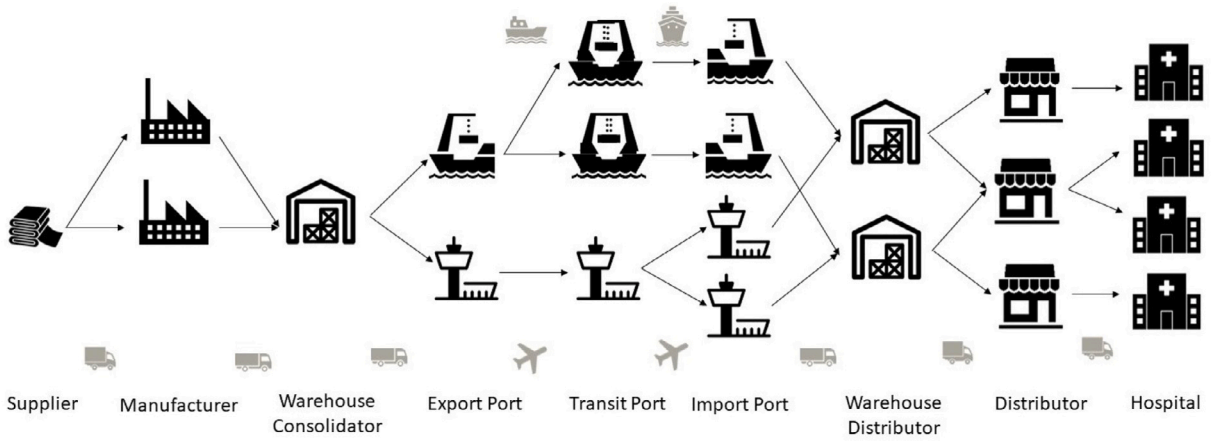


Fig. 3. Stylized supply chain of counterfeit PPE.

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.

Table 1 shows 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 [4,72]. Table 2 shows parametrization of the speed and the delays of the transport modalities for the simulation model of this study.

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 Fig. 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 [94].

4.4. Formalization of stylized 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. Fig. 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 [14]. 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.

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 supply 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

Table 1
Input parameters of actors and links for the simulation model of the stylized counterfeit PPE supply chain.

Actors				Links		
Input Parameter	Distribution	Value	Unit	Name	Value	Unit
Interarrival time of product at supplier	Exponential	10	days	Supplier to manufacturer 1	50	km
Time at manufacturer	Gamma	1.5, 0.8	days	Supplier to manufacturer 2	80	km
Time at warehouse consolidator	Triangular	0.5, 1, 1	days	Manufacturer 1 to warehouse consolidator	140	km
Time to pickup at warehouse consolidator	Triangular	0.5, 1, 2	days	Manufacturer 2 to warehouse consolidator	75	km
Probability of counterfeit PPE in container at warehouse consolidator		0.5		Warehouse consolidator to export sea port	45	km
Time at sea ports	Triangular	0.5, 1, 2	days	Warehouse consolidator to export air port	60	km
Time at air ports	Triangular	0.5, 1, 1	days	Export sea port to transit sea port Shanghai	2.8	days
Waiting time at yard for transport at import sea port	Uniform	0.5, 3	days	Export sea port to transit sea port Singapore	9	days
Probability of counterfeit PPE extracted at import sea port		0.5		Transit sea port Shanghai to import sea port Boston	42.5	days
Waiting time at yard for transport at import air port	Uniform	0.5, 1	days	Transit sea port Singapore to import sea port New York	26	days
Probability of counterfeit PPE extracted at import air port		0.5		Export air port to transit air port Amsterdam	9274	km
Time at warehouse distributor	Triangular	1, 2, 2	days	Transit air port Amsterdam to import air port Boston and New York	5547, 5847	km
Time at distributor	Exponential	0.2	days	Import sea and air port Boston to warehouse distributor Boston	15, 20	km
Time at hospital	Exponential	0.1	days	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 2
Input parameters of speed and delay of the transport modalities 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 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

an open-source flight route database from 2017. The travel distance between the airports is calculated using the Haversine distance between two airports [95]. 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 for more details on the specifying rules and distances.

For our study, we randomly generate a set of graphs using the stylized 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

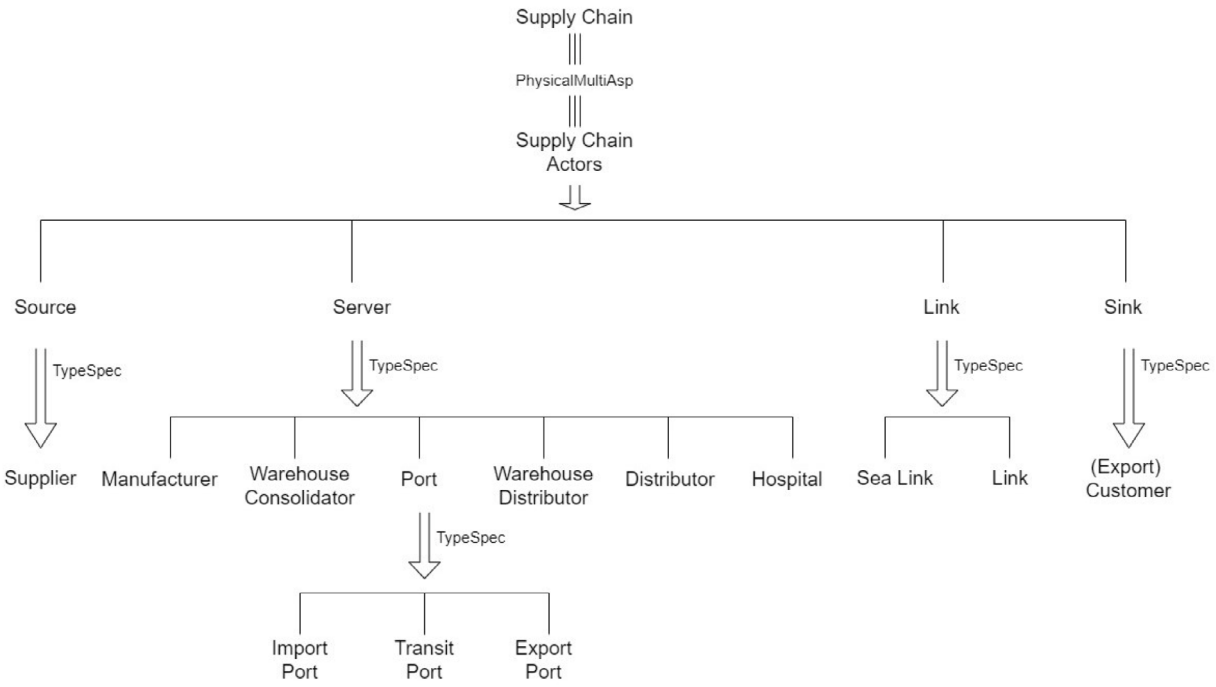


Fig. 4. System entity structure of the counterfeit PPE supply chain simulation model.

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 [96,97]. 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 [8,98].

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.

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 B.

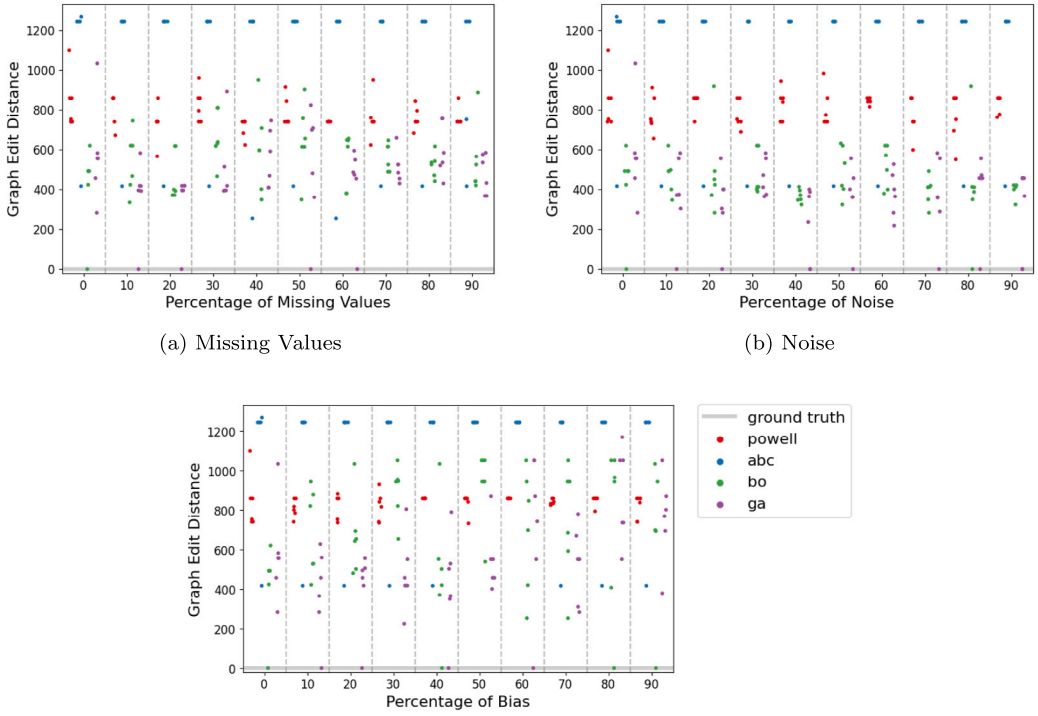
5.1.1. Graph edit distance

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 [88]. 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. Fig. 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.

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, Fig. 5(a) 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 Fig. 5(c) that BO finds optimal solutions with a relatively high graph edit distance compared to missing



(a) Missing Values (b) Noise

(c) Bias

Fig. 5. Graph edit distance per dimension of data sparseness for Powell's method, ABC, BO, GA.

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 Fig. 5(b). 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.

5.1.2. Ranking of average betweenness centrality

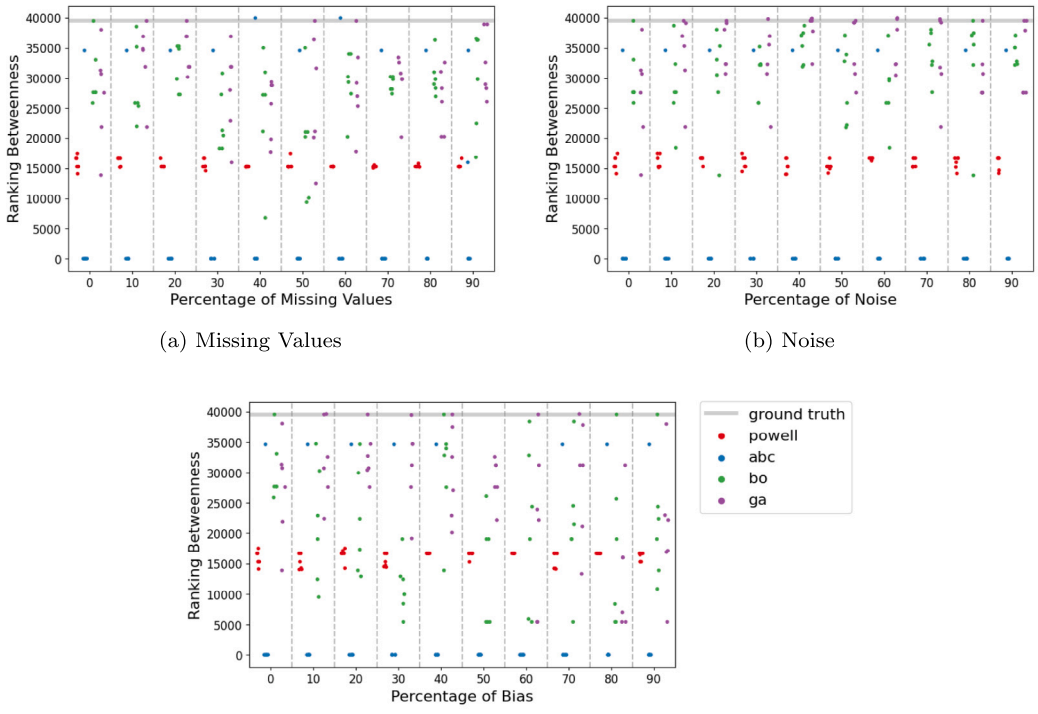
The graphs in the set for calibration are ranked based on their average betweenness centrality (as described in Section 4.1), where higher rankings correspond to higher average betweenness centrality scores. The ground truth graph ranking is 39520, with a ranking from 0 to 40000, and the average betweenness centrality is 0.046, with a range of 0.003 to 0.110. Fig. 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 B.

Fig. 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 Fig. 6(b)).

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, Fig. 6(a) 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 Fig. 6(c), 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, Fig. 6(b) shows that the optimal solutions are close to the ground truth for betweenness



(a) Missing Values (b) Noise

(c) Bias

Fig. 6. Ranking of average betweenness centrality per dimension of data sparseness for Powell's method, ABC, BO, GA.

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.

5.1.3. Vertices and edges

Fig. 7 and Fig. 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.

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 resonates with the sparse data. Fig. 7(a) 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, Figs. 7(c) and 8(c) 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. Fig. 7(b) 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.

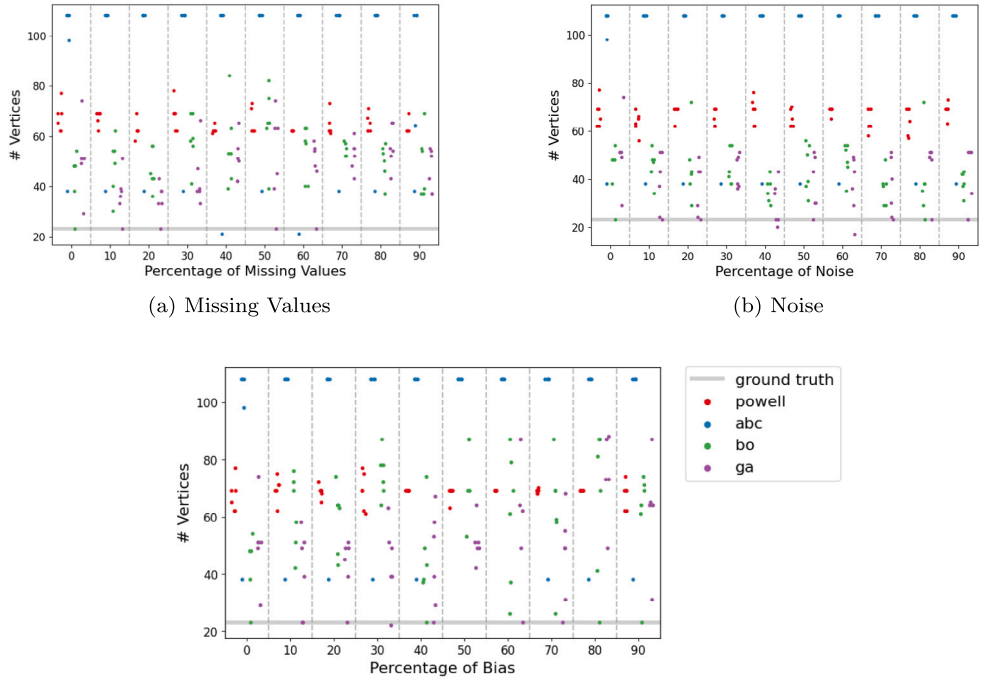
Table 3
Configuration of scenarios.

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

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. Table 3 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.

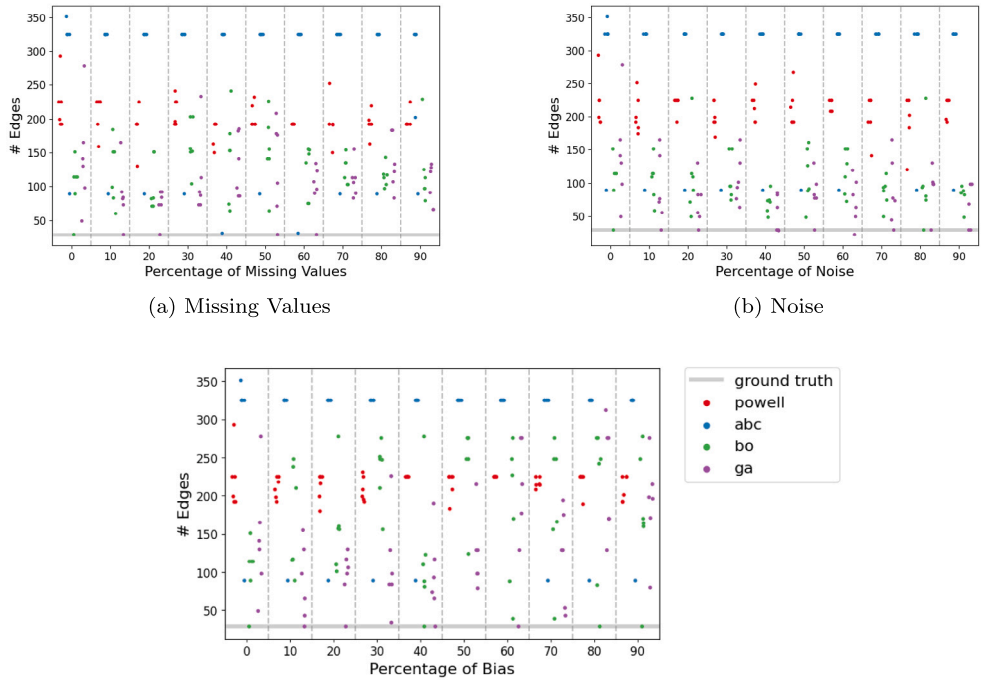
Fig. 9 shows that BO and GA fail to identify the ground truth for each scenario. We see in Fig. 9(a) that BO has a relatively low graph edit distance between 242 and 419 edit operations in all three scenarios, with two outliers with 650 and 779 operations, respectively, for the scenario High Noise and for the scenario High Noise & High Bias. The graph edit distance of GA is relatively high with a range



(a) Missing Values (b) Noise

(c) Bias

Fig. 7. Number of vertices per dimension of data sparseness for Powell's method, ABC, BO, GA.

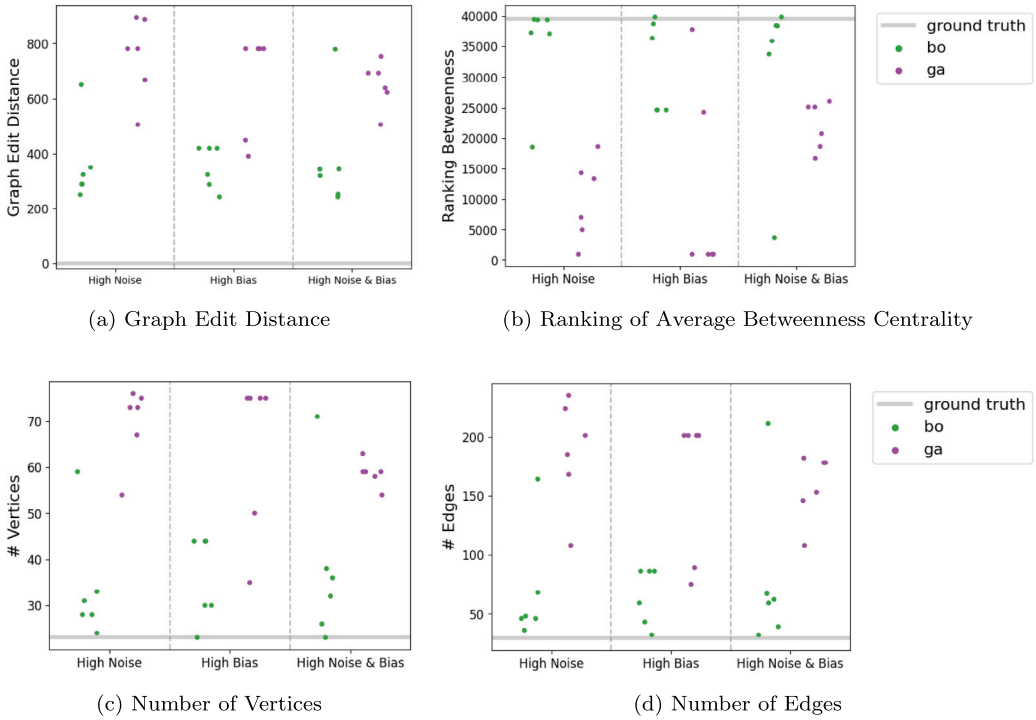


(a) Missing Values (b) Noise

(c) Bias

Fig. 8. Number of edges per dimension of data sparseness for Powell's method, ABC, BO, GA.

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, Fig. 9(b) 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, Fig. 9(c) and Fig. 9(d) 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



(a) Graph Edit Distance (b) Ranking of Average Betweenness Centrality

(c) Number of Vertices (d) Number of Edges

Fig. 9. Results of the experiments for BO and GA.

solutions within distinct subsets of the graph features. A pair plot is presented in Appendix B for a more detailed picture of the difference between BO and GA.

6. Discussion

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

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 [5], 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 B), 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 [68]. 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 [53].

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 resonate and 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 [10,96]. Further research should focus on finding a diverse set of optimal solutions in terms of simulation model calibration with sparse data.

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 [84]. 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.

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 resonate with sparse data. 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.

CRediT authorship contribution statement

Isabelle M. van Schilt: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Jan H. Kwakkel:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Software, Methodology, Conceptualization. **Jelte P. Mense:** Writing – original draft, Visualization, Validation, Supervision, Resources, Conceptualization. **Alexander Verbraeck:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Methodology, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Isabelle M. van Schilt reports financial support was provided by Nederlandse Politie. One of the co-author is editor of the special issue. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.aei.2024.102926.

Data availability

Data will be made available on request.

References

[1] I.A. Omar, M. Debe, R. Jayaraman, K. Salah, M. Omar, J. Arshad, Blockchain-based supply chain traceability for COVID-19 personal protective equipment, Comput. Ind. Eng. 167 (2022) 107995, http://dx.doi.org/10.1016/j.cie.2022.107995.

[2] M. Ippolito, C. Gregoretti, A. Cortegiani, P. Iozzo, Counterfeit filtering facepiece respirators are posing an additional risk to health care workers during COVID-19 pandemic, Am. J. Infect. Control 48 (7) (2020) 853, http://dx.doi.org/10.1016/j.ajic.2020.04.020.

[3] L. Hashemi, C.C. Jeng, A. Mohiuddin, E. Huang, L. Shelley, 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, Institute of Electrical and Electronics Engineers, Inc., Singapore, 2023, pp. 522–532, http://dx.doi.org/10.1109/WSC57314.2022.10015398.

[4] L. Hashemi, E. Huang, L. Shelley, Counterfeit PPE: Substandard respirators and their entry into supply chains in major cities, Urban Crime. An Int. J. 3 (2) (2022) 74–109, http://dx.doi.org/10.26250/heal.panteion.uc.v3i2.290.

[5] I.M. van Schilt, J. Kwakkel, J.P. Mense, A. Verbraeck, 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, Institute of Electrical and Electronics Engineers, Inc, Singapore, 2023, pp. 496–507, http://dx.doi.org/10.1109/WSC57314.2022.10015241.

[6] C. Nellemann, J. Stock, M. Shaw, World Atlas of illicit flows. Global initiative against transnational organized crime, 2018, URL: https://globalinitiative.net/wp-content/uploads/2018/09/Atlas-Illicit-Flows-FINAL-WEB-VERSION.pdf.

[7] Z. Eser, B. Kurtulmusoglu, A. Bicaksiz, S.I. Sumer, Counterfeit supply chains, Procedia Econ. Finance 23 (2015) 412–421, http://dx.doi.org/10.1016/S2212-5671(15)00344-5.

[8] A. Ficara, L. Cavallaro, F. Curreri, G. Fiumara, P. De Meo, O. Bagdasar, W. Song, A. Liotta, Criminal networks analysis in missing data scenarios through graph distances, Plos One 16 (8) (2021) e0255067, http://dx.doi.org/10.1371/journal.pone.0255067.

[9] N.R. Magliocca, K. McSweeney, S.E. Sesnie, E. Tellman, J.A. Devine, E.A. Nielsen, Z. Pearson, D.J. Wrathall, Modeling cocaine traffickers and counterdrug interdiction forces as a complex adaptive system, Proc. Natl. Acad. Sci. 116 (16) (2019) 7784–7792, http://dx.doi.org/10.1073/pnas.1812459116.

[10] R. Anzoom, R. Nagi, C. Vogiatzis, A review of research in illicit supply-chain networks and new directions to thwart them, Inst. Ind. Syst. Eng. Trans. 54 (2) (2021) 134–158, http://dx.doi.org/10.1080/24725854.2021.1939466.

[11] P.A. Duijn, V. Kashirin, P. Sloot, The relative ineffectiveness of criminal network disruption, Sci. Rep. 4 (1) (2014) 1–15, http://dx.doi.org/10.1038/srep04238.

[12] G.M. Grossman, C. Shapiro, Counterfeit-product trade, Am. Econ. Rev. 78 (1) (1986) 59–75.

[13] L.I. Shelley, Dark Commerce: How a New Illicit Economy Is Threatening Our Future, Princeton University Press, Princeton, NJ, USA, 2018.

[14] J. Banks, Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice, John Wiley & Sons, New York, NY, USA, 1998.

[15] B.P. Zeigler, A. Muzy, E. Kofman, Theory of Modeling and Simulation: Discrete Event & Iterative System Computational Foundations, third ed., Academic Press, FL, USA, 2018, http://dx.doi.org/10.1016/C2016-0-03987-6.

[16] A. Schmitt, M. Singh, 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, Institute of Electrical and Electronics Engineers, Inc., Austin, Texas, 2009, pp. 1237–1248, http://dx.doi.org/10.1109/WSC.2009.5429561.

[17] N.R. Magliocca, A.N. Price, P.C. Mitchell, K.M. Curtin, M. Hudnall, K. McSweeney, Coupling agent-based simulation and spatial optimization models to understand spatially complex and co-evolutionary behavior of cocaine trafficking networks and counterdrug interdiction, Inst. Ind. Syst. Eng. Trans. (2022) 1–14, http://dx.doi.org/10.1080/24725854.2022.2123998.

[18] M.R. Wigan, The fitting, calibration, and validation of simulation models, Simulation 18 (5) (1972) 188–192, http://dx.doi.org/10.1177/003754977201800506.

[19] T.I. Ören, Concepts and criteria to assess acceptability of simulation studies: A frame of reference, Commun. Assoc. Comput. Machinery 24 (4) (1981) 180–189, http://dx.doi.org/10.1145/358598.358605.

[20] M. Hofmann, On the complexity of parameter calibration in simulation models, J. Defense Model. Simul. 2 (4) (2005) 217–226, http://dx.doi.org/10.1177/154851290500200405.

[21] Z. Lian, Z. Zhou, C. Hu, Z. Feng, P. Ning, Z. Ming, Interpretable large-scale belief rule base for complex industrial systems modeling with expert knowledge and limited data, Adv. Eng. Inform. 62 (2024) 102852, http://dx.doi.org/10.1016/j.aei.2024.102852.

[22] I.M. van Schilt, J.H. Kwakkel, J.P. Mense, A. Verbraeck, Dimensions of data sparseness and their effect on supply chain visibility, Comput. Ind. Eng. 191 (2024) 110108, http://dx.doi.org/10.1016/j.cie.2024.110108.

[23] Z. Liu, D. Rexachs, F. Epelde, E. Luque, A simulation and optimization based method for calibrating agent-based emergency department models under data scarcity, Comput. Ind. Eng. 103 (2017) 300–309, http://dx.doi.org/10.1016/j.cie.2016.11.036.

[24] L. de Groot, A. Hübl, 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, Institute of Electrical and Electronics Engineers, Inc., Phoenix, Arizona, 2021, pp. 1–12, http://dx.doi.org/10.1109/WSC52266.2021.9715306.

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

[26] C. Moore, J. Doherty, Role of the calibration process in reducing model predictive error, Water Resour. Res. 41 (5) (2005) 1–14, http://dx.doi.org/10.1029/2004WR003501.

[27] J. Coenen, R.E. van Der Heijden, A.C. van Riel, Understanding approaches to complexity and uncertainty in closed-loop supply chain management: Past findings and future directions, J. Clean. Prod. 201 (2018) 1–13, http://dx.doi.org/10.1016/j.jclepro.2018.07.216.

[28] K. van der Zwet, A.I. Barros, T.M. van Engers, B. van der Vecht, An agent-based model for emergent opponent behavior, in: J. ao 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, Springer, Faro, Portugal, 2019, pp. 290–303, http://dx.doi.org/10.1007/978-3-030-22741-8_21.

[29] J.P. Caulkins, Local drug markets' response to focused police enforcement, Oper. Res. 41 (5) (1993) 848–863, http://dx.doi.org/10.1287/opre.41.5.848.

[30] C.P. Rydell, J.P. Caulkins, S.S. Everingham, Enforcement or treatment? Modeling the relative efficacy of alternatives for controlling cocaine, Oper. Res. 44 (5) (1996) 687–695, http://dx.doi.org/10.1287/opre.44.5.687.

[31] A. Dray, L. Mazerolle, P. Perez, A. Ritter, Policing Australia's 'heroin drought': Using an agent-based model to simulate alternative outcomes, J. Exp. Criminol. 4 (3) (2008) 267–287, http://dx.doi.org/10.1007/s11292-008-9057-1.

[32] A. Kovari, E. Pruyt, 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, System Dynamics Society, St. Gallen, Switzerland, 2012, pp. 1–24, URL: https://systemdynamics.org/wp-content/uploads/assets/proceedings/2012/2012proceed.html.

[33] L. Kretschmann, T. Münsterberg, 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, Institute of Business Logistics and General Management, Hamburg, Germany, 2017, pp. 427–448, URL: https://hdl.handle.net/10419/209320.

[34] M. Jensen, F. Dignum, 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, Springer, Mainz, Germany, 2019, pp. 9–22, http://dx.doi.org/10.1007/978-3-030-61503-1_2.

[35] J.Á. González Ordiano, L. Finn, A. Winterlich, G. Moloney, S. Simske, On the analysis of illicit supply networks using variable state resolution-Markov chains, in: M.-J. Lesot, S. Vieira, M.Z. Reformat, J. ao Paulo Carvalho, A. Wilbik, B. Bouchon-Meunier, R.R. Yager (Eds.), International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Communications in Computer and Information Science, Lisbon, Portugal, 2020, pp. 513–527, http://dx.doi.org/10.1007/978-3-030-50146-4_38.

[36] M.A. Benatia, D. Baudry, A. Louis, Detecting counterfeit products by means of frequent pattern mining, J. Ambient Intell. Humaniz. Comput. 13 (7) (2022) 3683–3692, http://dx.doi.org/10.1007/s12652-020-02237-y.

[37] R.J. Lempert, S.W. Popper, S.C. Bankes, Shaping the Next One Hundred Years: New Methods for Quantitative, Long-Term Policy Analysis, RAND Corporation, Santa Monica, CA, 2003, http://dx.doi.org/10.7249/MR1626.

[38] V.A.W.J. Marchau, W.E. Walker, P.J.T.M. Bloemen, S.W. Popper, Introduction, in: Decision Making under Deep Uncertainty: From Theory To Practice, Springer International Publishing, Cham, ISBN: 978-3-030-05252-2, 2019, pp. 1–20, http://dx.doi.org/10.1007/978-3-030-05252-2_1.

[39] M.D. Webster, A.P. Sokolov, Joint Program Report Series Report, Quantifying the Uncertainty in Climate Predictions, vol. 37, MIT Joint Program on the Science and Policy of Global Change, 1998, URL: https://globalchange.mit.edu/publication/14359.

[40] W.S. Parker, Ensemble modeling, uncertainty and robust predictions, Wiley Interdiscip. Rev. Clim. Change 4 (3) (2013) 213–223, http://dx.doi.org/10.1002/wcc.220.

[41] W. Parker, Values and uncertainties in climate prediction, revisited, Stud. Hist. Philos. Sci. A 46 (2014) 24–30, http://dx.doi.org/10.1016/j.shpsa.2013.11.003.

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

[43] E.A. Moallemi, J. Köhler, Coping with uncertainties of sustainability transitions using exploratory modelling: The case of the MATISSE model and the UK's mobility sector, Environ. Innov. Soc. Trans. 33 (2019) 61–83, http://dx.doi.org/10.1016/j.eist.2019.03.005.

[44] A. De Santis, T. Giovannelli, S. Lucidi, M. Messedaglia, M. Roma, A simulation-based optimization approach for the calibration of a discrete event simulation model of an emergency department, Ann. Oper. Res. (2022) 1–30, http://dx.doi.org/10.1007/s10479-021-04382-9.

[45] J. Hao, W. Ye, L. Jia, G. Wang, J. Allen, Building surrogate models for engineering problems by integrating limited simulation data and monotonic engineering knowledge, Adv. Eng. Inform. 49 (2021) 101342, http://dx.doi.org/10.1016/j.aei.2021.101342.

[46] I.S. van Droffelaar, J.H. Kwakkel, J.P. Mense, A. Verbraeck, Simulation-optimization configurations for real-time decision-making in fugitive interception, Simul. Model. Pract. Theory 133 (2024) 102923, http://dx.doi.org/10.1016/j.simpat.2024.102923.

[47] J. Puchinger, G.R. Raidl, Combining metaheuristics and exact algorithms in combinatorial optimization: A survey and classification, in: First International Work-Conference on the Interplay Between Natural and Artificial Computation, Springer, Las Palmas, Spain, 2005, pp. 41–53, http://dx.doi.org/10.1007/11499305_5.

[48] M.J. Powell, An efficient method for finding the minimum of a function of several variables without calculating derivatives, Comput. J. 7 (2) (1964) 155–162, http://dx.doi.org/10.1093/comjnl/7.2.155.

[49] J. Zhong, W. Cai, Differential evolution with sensitivity analysis and the Powell's method for crowd model calibration, J. Comput. Sci. 9 (2015) 26–32, http://dx.doi.org/10.1016/j.jocs.2015.04.013.

[50] D.M. Olsson, L.S. Nelson, The Nelder-Mead simplex procedure for function minimization, Technometrics 17 (1) (1975) 45–51, http://dx.doi.org/10.1080/00401706.1975.10489269.

[51] E.L. Lawler, D.E. Wood, Branch-and-bound methods: A survey, Oper. Res. 14 (4) (1966) 699–719, http://dx.doi.org/10.1287/opre.14.4.699.

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

[53] A. Slowik, H. Kwasnicka, Evolutionary algorithms and their applications to engineering problems, Neural Comput. Appl. 32 (16) (2020) 12363–12379, http://dx.doi.org/10.1007/s00521-020-04832-8.

[54] N. Malleson, Calibration of simulation models, Encycl. Criminol. Crim. Justice 40 (2014) 115–118, http://dx.doi.org/10.1007/978-1-4614-5690-2_688.

[55] B. Park, H. Qi, Development and evaluation of a procedure for the calibration of simulation models, Transp. Res. Rec. 1934 (1) (2005) 208–217, http://dx.doi.org/10.1177/0361198105193400122.

[56] Y. Ren, Y. Wu, An efficient algorithm for high-dimensional function optimization, Soft Comput. 17 (6) (2013) 995–1004, http://dx.doi.org/10.1007/s00500-013-0984-z.

[57] D. Whitley, A genetic algorithm tutorial, Stat. Comput. 4 (2) (1994) 65–85, http://dx.doi.org/10.1007/BF00175354.

[58] P.M. Reed, D. Hadka, J.D. Herman, J.R. Kasprzyk, J.B. Kollat, Evolutionary multiobjective optimization in water resources: The past, present, and future, Adv. Water Resour. 51 (2013) 438–456, http://dx.doi.org/10.1016/j.advwatres.2012.01.005.

[59] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE Trans. Evol. Comput. 6 (2) (2002) 182–197, http://dx.doi.org/10.1109/4235.996017.

[60] J.B. Kollat, P.M. Reed, A computational scaling analysis of multiobjective evolutionary algorithms in long-term groundwater monitoring applications, Adv. Water Resour. 30 (3) (2007) 408–419.

[61] J.Z. Salazar, P.M. Reed, J.D. Herman, M. Giuliani, A. Castelletti, A diagnostic assessment of evolutionary algorithms for multi-objective surface water reservoir control, Adv. Water Resour. 92 (2016) 172–185, http://dx.doi.org/10.1016/j.advwatres.2016.04.006.

[62] D. Hadka, P. Reed, Borg: An auto-adaptive many-objective evolutionary computing framework, Evol. Comput. 21 (2) (2013) 231–259, http://dx.doi.org/10.1162/EVCO_a_00075.

[63] K. Csilléry, M.G. Blum, O.E. Gaggiotti, O. François, Approximate Bayesian computation (ABC) in practice, Trends Ecol. Evolut. 25 (7) (2010) 410–418, http://dx.doi.org/10.1016/j.tree.2010.04.001.

[64] J.A. Vrugt, K.J. Beven, Embracing equifinality with efficiency: Limits of acceptability sampling using the DREAM (LOA) algorithm, J. Hydrol. 559 (2018) 954–971, http://dx.doi.org/10.1016/j.jhydrol.2018.02.026.

[65] C.J.T. Braak, A Markov chain Monte Carlo version of the genetic algorithm differential evolution: easy Bayesian computing for real parameter spaces, Stat. Comput. 16 (2006) 239–249, http://dx.doi.org/10.1007/s11222-006-8769-1.

[66] T. Wöhling, J.A. Vrugt, Multiresponse multilayer vadose zone model calibration using Markov chain Monte Carlo simulation and field water retention data, Water Resour. Res. 47 (4) (2011) W04510, http://dx.doi.org/10.1029/2010WR009265.

[67] M. Sadegh, J.A. Vrugt, Approximate Bayesian computation using Markov chain Monte Carlo simulation: DREAM (ABC), Water Resour. Res. 50 (8) (2014) 6767–6787, http://dx.doi.org/10.1002/2014WR015386.

[68] H. Jalali, I. Van Nieuwenhuyse, V. Picheny, Comparison of kriging-based algorithms for simulation optimization with heterogeneous noise, European J. Oper. Res. 261 (1) (2017) 279–301, http://dx.doi.org/10.1016/j.ejor.2017.01.035.

[69] J. van Hoof, J. Vanschoren, Hyperboost: Hyperparameter optimization by gradient boosting surrogate models, 2021, http://dx.doi.org/10.48550/arXiv.2101.02289, arXiv:2101.02289.

[70] D.R. Jones, A taxonomy of global optimization methods based on response surfaces, J. Global Optim. 21 (4) (2001) 345–383, http://dx.doi.org/10.1023/A:1012771025575.

[71] B. Bischl, M. Binder, M. Lang, T. Pielok, J. Richter, S. Coors, J. Thomas, T. Ullmann, M. Becker, A.-L. Boulesteix, D. Deng, M. Lindauer, Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges, Wiley Interdiscip. Rev.: Data Min. Knowl. Discov. 13 (2) (2023) e1484, http://dx.doi.org/10.1002/widm.1484.

[72] L. Kuipers, 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, 2021, URL: https://resolver.tudelft.nl/uuid:5f68b82f-205e-4509-9a64-22082c46065f.

[73] X. Hu, P. Wu, A data assimilation framework for discrete event simulations, ACM Trans. Model. Comput. Simul. 29 (3) (2019) 1–26, http://dx.doi.org/10.1145/3301502.

[74] X. Xie, Data Assimilation in Discrete Event Simulations (Ph.D. thesis), Faculty of Technology, Policy and Management, Delft University of Technology, Delft, Netherlands, 2018, http://dx.doi.org/10.4233/uuid:d0c47163-3845-430b-a8ce-013c41faa2ea.

[75] H. Folkerts, T. Pawletta, C. Deatcu, B.P. Zeigler, 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, Institute of Electrical and Electronics Engineers, Inc., Fairfax, VA, USA, 2020, pp. 1–12, http://dx.doi.org/10.22360/SpringSim.2020.Mod4Sim.001.

[76] L. Yilmaz, Toward self-aware models as cognitive adaptive instruments for social and behavioral modeling, in: Social-Behavioral Modeling for Complex Systems, John Wiley & Sons, Ltd, ISBN: 9781119485001, 2019, pp. 569–586, http://dx.doi.org/10.1002/9781119485001.ch24.

[77] B.P. Zeigler, P.E. Hammonds, Modeling and Simulation-Based Data Engineering: Introducing Pragmatics into Ontologies for Net-Centric Information Exchange, Elsevier, USA, 2007.

[78] M. Hofmann, 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, Springer, Heidelberg, 2013, pp. 59–87, http://dx.doi.org/10.1007/978-3-642-31140-6_3.

[79] A. Tolk, E.H. Page, V.V. Graciano Neto, P. Weirich, N. Formanek, J.M. Durán, J.F. Santucci, S. Mittal, 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, Springer, 2023, pp. 383–412, http://dx.doi.org/10.1007/978-3-031-11085-6_16.

[80] B.P. Zeigler, Multifacetted Modelling and Discrete Event Simulation, Academic Press Professional, Inc., San Diego, CA, USA, 1984.

[81] B.P. Zeigler, H.S. Sarjoughian, Guide to Modeling and Simulation of Systems of Systems. Simulation Foundations, Methods and Applications, Springer, Londen, UK, 2013.

[82] T. Pawletta, A. Schmidt, B.P. Zeigler, U. Durak, Extended variability modeling using system entity structure ontology within MATLAB/Simulink, in: Proceedings of the 49th Annual Simulation Symposium, Society for Computer Simulation International, San Diego, CA, USA, 2016, pp. 22:1–22:8.

[83] C. Deatcu, H. Folkerts, T. Pawletta, U. Durak, 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, Society for Computer Simulation International, Braunschweig, Germany, 2018, pp. 23–43.

[84] B. Hermans, Structural Uncertainty in Supply Chain Simulation Models (M.Sc. thesis), Faculty of Technology, Policy and Management, Delft University of Technology, 2022, URL: https://resolver.tudelft.nl/uuid:e19d2957-eb33-4171-8dc1-8053de3d9e1c.

[85] M. Khondoker, R. Dobson, C. Skirrow, A. Simmons, D. Stahl, A comparison of machine learning methods for classification using simulation with multiple real data examples from mental health studies, Stat. Methods Med. Res. 25 (5) (2016) 1804–1823, http://dx.doi.org/10.1177/0962280213502437.

[86] P. Wills, F.G. Meyer, Metrics for graph comparison: A practitioner's guide, Plos One 15 (2) (2020) e0228728, http://dx.doi.org/10.1371/journal.pone.0228728.

[87] R. Wang, T. Zhang, T. Yu, J. Yan, X. Yang, Combinatorial learning of graph edit distance via dynamic embedding, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 2021, pp. 5241–5250, http://dx.doi.org/10.1109/CVPR46437.2021.00520.

[88] Z. Abu-Aisheh, R. Raveaux, J.-Y. Ramel, P. Martineau, 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, Lisbon, Portugal, 2015, pp. 271–278, http://dx.doi.org/10.5220/0005209202710278.

[89] K. Riesen, M. Ferrer, R. Dornberger, H. Bunke, Greedy graph edit distance, in: P. Perner (Ed.), International Conference of Machine Learning and Data Mining in Pattern Recognition, Springer, Hamburg, Germany, 2015, pp. 3–16, http://dx.doi.org/10.1007/978-3-319-21024-7_1.

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

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

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

[93] A. Gelman, D.B. Rubin, Inference from iterative simulation using multiple sequences, Statist. Sci. 7 (4) (1992) 457–472, http://dx.doi.org/10.1214/ss/1177011136.

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

[95] C.C. Robusto, The cosine-haversine formula, Amer. Math. Monthly 64 (1) (1957) 38–40, http://dx.doi.org/10.2307/2309088.

[96] C. Morselli, Assessing vulnerable and strategic positions in a criminal network, J. Contemp. Crim. Justice 26 (4) (2010) 382–392, http://dx.doi.org/10.1177/1043986210377105.

[97] T. Diviák, J.K. Dijkstra, T.A. Snijders, Structure, multiplexity, and centrality in a corruption network: the Czech Rath affair, Trends Organ. Crime 22 (2019) 274–297, http://dx.doi.org/10.1007/s12117-018-9334-y.

[98] L. Cavallaro, A. Ficara, P. De Meo, G. Fiumara, S. Catanese, O. Bagdasar, W. Song, A. Liotta, Disrupting resilient criminal networks through data analysis: The case of sicilian mafia, Plos One 15 (8) (2020) e0236476, http://dx.doi.org/10.1371/journal.pone.0236476.