University of Strathclyde

Department of Management Science

Modelling Parameter Interdependence in System Dynamics: A Data-Driven Bayesian Network Approach to Assessing Uncertainties in Simulations

Wang Zhao

A thesis presented in fulfilment of the requirements for the degree of
Master of Philosophy

2026

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed: Wang Zhao

Date: 21 August 2025

Contents

Chapter 1 Introduction 1
1.1 Background and Motivation 1
1.1.1 System Dynamics and Parameter Estimates 1
1.1.2 Parameter Interdependence 1
1.1.3 Bayesian Networks as an Approach to Modelling Parameter Dependencies 2
1.1.4 Relevance to Mixing Methods 3
1.2 Problem statement 3
1.3 Research Objective and Question 4
1.4 Thesis Outline 4
Chapter 2 Literature Review 6
2.1 System Dynamics 6
2.1.1 Origin and Philosophical Foundations 6
2.1.2 Formal Representation of Systems 7
2.1.3 Dynamic Behaviour 9
2.1.4 Data Sources and Utilisation 10
2.1.5 Modelling Process 11
2.1.6 Debates around System Dynamics 11
2.2 Parameter Estimation of System Dynamics Models 13
2.2.1 Judgemental Methods for Parameter Estimation 14
2.2.2 Statistical Methods for Parameter Estimation 15
2.2.3 Special Treatment for Large Models 16
2.2.4 Estimating Interdependent Parameters 17
2.3 Sensitivity Analysis of System Dynamics Models 18
2.3.1 Sensitivity Analysis in System Dynamics 18
2.3.2 Parameter Interdependence in Sensitivity Analysis 19
2.3.3 Discussion on Parameter Dependencies 20
2.4 Modelling Multivariate Dependence Structure with Copula 21
2.4.1 Copulas: An Approach to Correlated Sampling 21
2.4.2 Archimedean and Elliptical Copulas 21
2.4.3 Limitations in Capturing Complex Conditional Structures Using Copulas 27
2.5 Bayesian Networks 28
2.5.1 Origin and Philosophical Foundations 28
2.5.2 Formal Representation of System 30

2.5.3 Data Sources and Utilisation 32
2.5.4 Modelling Process 33
2.6 Learning Bayesian Network from Data 33
2.6.1 Structure Learning Algorithms 34
2.6.2 Parameter Learning Algorithms 36
2.7 Combining SD and BN 38
2.7.1 The Mixing Methods Approach 38
2.7.2 Strategy for Examining Existing Combinations of SD and BN 39
2.7.3 Searching for Existing Studies 40
2.7.4 Taxonomy of SD-BN Design Patterns 40
2.7.5 Discussion of Existing SD-BN Combination 48
2.8 Summary of Gaps 50
Chapter 3 Methodology and Conceptual Approach 52
3.1 Research Philosophy and Methods 52
3.1.1 Research Philosophy 52
3.1.2 Research Methods 53
3.2 Overall Approach 54
3.3 Illustrative Example: Outline 56
3.3.1 System Dynamics SIR Model 57
3.3.2 Selected Parameters and Uncertainty Metrics 61
3.3.3 Exploring Dependence Structures Using Copulas 61
3.3.4 Modelling Parameter Dependencies Using Bayesian Networks 62
3.4 Chapter Summary 63
Chapter 4 Exploring Dependence Structures with Copula 64
4.1 Specification of Copulas 64
4.2 Monte Carlo Experiment 65
4.2.1 Sampling from Copulas 66
4.2.2 Monte Carlo Simulation 68
4.2.3 Colleting KPIs and Measuring Uncertainties 68
4.3 Influence of Parameter Dependence Structure 68
4.3.1 Monte Carlo Outcomes Across Correlation Levels 69
4.3.2 Observations and Interpretation 70
4.4 Chapter Summary 71
Chapter 5 Modelling Parameter Dependence with Bayesian Network 73
5.1 Data Acquisition and Preparation for Learning BN 73

5.1.1 Adjustment Start Time and Contact Adjustment Level 74
5.1.2 Disease Duration and Infectivity 76
5.1.3 Contact Rate 77
5.1.4 Vaccination Rate 78
5.1.5 Therapeutic Interventions 78
5.1.6 Summary of Data Preparation 78
5.2 Learning BN Structure and Parameter 79
5.2.1 Algorithm Choice: PC 79
5.2.2 Variable Discretisation 80
5.2.3 Structure and Parameter Learning 81
5.3 Monte Carlo Experiment 84
5.3.1 Generating Parameter Samples 84
5.3.2 Monte Carlo Simulation and Outcome 87
5.4 Validation of the Bayesian Network Approach 88
5.5 Influence of Parameter Dependence Structure: Revisit 89
5.5.1 Motivation for Comparing BN and Copulas Results 89
5.5.2 Comparison Setup, Execution, and Result 89
5.6 Chapter Summary 93
Chapter 6 Conclusion and Future Work 94
6.1 Overall Conclusion 94
6.2 Contributions 95
6.3 Relevance beyond Epidemiological Modelling 95
6.4 Practical Implications 96
6.5 Limitations and Future Work 97
6.5.1 Technical Limitations 97
6.5.2 Methodological Limitations 99
6.5.3 Closing Reflection 99
References 101
Appendix A Search Documentation 116
A1 Search Sessions 116
A2 Screening Criteria and Coding 116
A3 Flow Diagram 117
A4 Included Studies 118

Abstract

System Dynamics (SD) modelling supports decision-making by simulating and projecting key performance indicators (KPIs) of a system. To assess uncertainty in these KPIs, modellers typically vary model parameters and run multiple simulations. This process illustrates potential scenarios and generates KPI distributions that provide quantified measures of uncertainty through statistical analysis.

However, parameters often do not vary independently. Studies across sectors have reported correlations and dependencies among parameters (Krefeld-Schwalb et al., 2022; Li and Vu, 2013). When varying parameters in SD models, whether these dependencies are accounted for can shape the combinations of parameter values, influence the distributions of projected KPIs and the derived insights. This issue has not been thoroughly addressed in the SD literature.

To highlight the importance of parameter dependence, we present a copula-based experiment, modelling dependencies between SD model parameters in several different ways and comparing the resulting KPI distributions. The experiment demonstrates that both the strength of correlations and the structure of dependencies can affect KPI uncertainty. These findings motivate the adoption of more flexible approaches to adequately model such dependencies.

The main contribution of this thesis is a method that models dependencies among SD model parameters using Bayesian Networks (BNs) to improve KPI uncertainty assessment. BNs provide a flexible framework and algorithms for uncovering complex conditional relationships from data and integrating them with expert knowledge. We apply the approach to an epidemic SD model, where dependencies among epidemiological parameters are estimated from a cross-country COVID-19 dataset using a BN and validated against domain knowledge. The learned BN produces input for analysing KPIs of the epidemic SD model and yields uncertainty envelopes that differ from those generated by independent or single-copula priors.

This study offers SD practitioners a practical way to incorporate empirically grounded multi-parameter dependencies into their models, enhancing the defensibility of uncertainty assessments while keeping additional data-collection effort manageable. The proposed method contributes to both the SD and mixed-methods research literatures.

Chapter 1 Introduction

1.1 Background and Motivation

1.1.1 System Dynamics and Parameter Estimates

System Dynamics (SD) is a method developed by Jay Forrester in the 1950s, designed to support managerial decision-making through modelling and computer simulation (Forrester, 1961). At its core, SD uses feedback loops to represent the nonlinear and dynamic nature of complex systems. By creating conceptual or computational models that replicate real-world systems, SD enables analysts and decision-makers to explore how different policies or interventions might unfold over time (Sterman, 2000).

In SD models, equations are used to characterise the causal relationships between variables and parameters are used to quantify the strength of each relationship. Accurate parameter estimates play a crucial role in SD modelling because they enable a simulation to provide more credible model outputs, often presented as key performance indicators (KPIs) projections, on which decisions are based. Methods for estimating parameters have been a focus of SD research. Sterman (2000) identified two categories of parameter estimating methods: judgemental methods that elicit expert judgement on parameter values via workshops, interviews, or structured elicitation processes (Ford and Sterman, 1998; Vennix, 1996), and statistical methods that infer parameter values via fitting model behaviour to historical patterns (Graham, 1980; Osgood and Liu, 2015; Struben et al., 2015).

Parameter estimates are subject to uncertainty. In judgemental methods, uncertainty arises from the varying levels of confidence, knowledge, and experience among experts. In statistical methods, uncertainty stems from variability in data sources, measurement errors, and inherent system complexities. To capture these uncertainties, modellers typically represent parameters as probability distributions rather than fixed point estimates (Osgood and Liu, 2015; Struben et al., 2015).

1.1.2 Parameter Interdependence

In SD models, parameters are treated as constant, independent values. However, the uncertainties – or errors – associated with these parameters often exhibit interdependencies. Empirical studies across various modelling contexts, including biological models (Li and Vu,

2013) and cognitive systems (Krefeld-Schwalb et al., 2022), consistently report correlations among parameter errors. While these correlations have no impact on model KPIs when parameters are fixed, they do influence how parameters co-vary and, consequently, affect the uncertainty assessment of KPIs. Despite its potential significance, this issue has received limited attention in the SD literature, with only a few studies – most notably McNaught (2003), Eker et al. (2014), and Ford and Flynn (2005) – explicitly acknowledging its relevance. In this thesis, we demonstrate this point using copulas, which are mathematical functions that describe the dependence between two or more random variables (Trivedi and Zimmer, 2007), to create joint distributions of parameters with the same association measure but different association structure. The use of copulas illustrates how dependence structure can influence KPI distributions and distort uncertainty assessment and thus highlight the need to adequately model dependence between parameters.

1.1.3 Bayesian Networks as an Approach to Modelling Parameter Dependencies

Within in the SD realm, it is recommended that when parameters are interdependent, the model boundary be extended to incorporate such interdependence as part of the model's causal structure (Ford, 1990; Ford and Flynn, 2005). Dependencies in SD models are typically represented using analytical equations or lookup tables (Eker et al., 2014; Rios-Ocampo and Gary, 2022; Sterman, 2000). However, these methods can become cumbersome or insufficient when capturing complex dependency structures or multivariate correlations, such as those found among SD model parameters, which are often highly non-linear and multidimensional. This challenge calls for a more flexible approach.

Bayesian Networks (BNs) offer a well-established statistical framework capable of representing complex interdependencies among variables (Koller and Friedman, 2009; Pearl, 1985). A BN is a probabilistic graphical model that represents variables and their conditional dependencies using a directed acyclic graph (DAG). These dependencies can be interactively elicited from experts (Bertone et al., 2018; Quigley and Walls, 2021; Quigley et al., 2008) or algorithmically learned by pooling data from similar events or instances, allowing associations between variables to be inferred from observed patterns (Daly et al., 2011; Scanagatta et al., 2019).

BNs thus provide a flexible means of modelling the dependencies among SD model parameters, drawing on expert knowledge, empirical data, or a combination of both. Bertone et al. (2018), for example, used expert-elicited BN to represent interdependent parameters in SD models and support scenario generation. However, the data-driven aspect of BNs, which is particularly valuable in situations where expert knowledge is scarce, but data is abundant, has not been fully explored. In this thesis, we explore using BNs learned from empirical data to model parameter dependencies in System Dynamics models. The method is primarily data-driven, with BN's structure validated against domain knowledge, as literature from a relevant field suggests the combination of data and domain knowledge can outperform pure data-driven methods in learning a BN (Abdulkareem et al., 2019).

1.1.4 Relevance to Mixing Methods

This thesis adopts a mixing methods perspective (Howick and Ackermann, 2011; Mingers and Brocklesby, 1997), integrating SD and BN in a way that bridges different methods of quantitative modelling. In management science and operational research (MS/OR), mixing methods typically involves selecting and combining complementary methods or techniques to address the complexity of real-world problems more robustly. In this thesis, we borrow SD's strengths in feedback-based dynamic simulation and combine them with BN's data-driven, conditional probability modelling to capture nuanced, empirical dependencies among parameters. This synergy reflects the idea that no single approach fully addresses all aspects of uncertainty analysis in SD, and that weaving together SD and BN can yield deeper insights into how data-informed multi-parameter dependencies drive model outcomes. In the subsequent chapters, we demonstrate how this mixing methods approach introduces an empirical approach that operationalise data learning of BN to model dependencies between SD model parameters.

1.2 Problem statement

While the previous section outlined the critical role of informed parameter estimates and potential effects of not modelling multi-parameter dependencies adequately, an important gap remains unaddressed in the SD literature:

- Lack of a data-driven method to model multi-parameter dependencies in System Dynamics (SD)

  In System Dynamics (SD), dependencies between parameters can affect the assessment of uncertainty in model KPIs. When such dependencies are ignored or oversimplified, the resulting combinations of parameter values can lead to distorted KPI projections and inaccuracies in uncertainty assessment. Despite this risk, most existing SD studies rely on theoretical or uniform parameter distributions (Eker et al., 2014; Ford and Flynn, 2005; Jadun et al., 2017), with limited attention given to adequately capturing multi-parameter dependencies. Some efforts – such as Bertone et al. (Bertone et al., 2018) – have used expert-elicited Bayesian Networks (BNs) to represent dependencies. However, such approaches depend heavily on expert knowledge, and a data-driven method for learning parameter dependencies directly from empirical data remains largely absent in the SD literature. This gap is especially critical in contexts where expert knowledge is scarce, but data is available, underscoring the need for a data-driven approach to modelling multi-parameter dependencies in SD.

Addressing this gap forms the central motivation of this thesis.

1.3 Research Objective and Question

In response to the gap outlined above, this thesis pursues the following objective:

- Demonstrate a data-driven method for modelling SD model parameter interdependence

  We introduce and implement Bayesian Networks (BN) as a data-driven tool to build correlated distributions for parameters in System Dynamics (SD) models. By learning from empirical datasets, BNs can adequately capture dependencies among parameters in the real world and facilitate the uncertainty assessment of SD model KPIs.

The Research Question arising from this objective is:

- *RQ*: How can a data-driven Bayesian Network be employed to capture the dependencies among SD model parameters?

1.4 Thesis Outline

To address the objective and answer the research question, the thesis is organised into six chapters:

- Chapter 1: Introduction

This chapter establishes the background and motivation for modelling parameter interdependence in SD models, highlights the research gap, and defines the thesis objective.

- Chapter 2: Literature Review

  This chapter reviews existing literature on System Dynamics, Bayesian Networks, and the intersection of them. Emphasis is placed on how parameters are estimated for SD models, how parameter dependency structures can be explored, and how BN fits the role of modelling the joint distribution of parameters.

- Chapter 3: Methodology and Conceptual Approach

  This chapter presents the conceptual workflow for assessing the influence of parameter correlations on the uncertainty of model output and introduces the idea of using BN to empirically model parameter correlations. It also outlines an illustrative SD model to demonstrate the approach.

- Chapter 4: Exploring Dependence Structures with Copula

  This chapter describes how a copula-based procedure is used to explore various parameter dependency structures of the SD model and illustrate their influence on KPI uncertainty.

- Chapter 5: Modelling Parameter Dependence with Bayesian Network

  This chapter demonstrates how a BN-based approach captures nuanced, real-world relationships among parameters and illustrates it through the case study using the SIR SD model.

- Chapter 6: Conclusion and Future Work

  This chapter presents the conclusion, summarises the main contribution, discusses the limitations, and suggests avenues for future research.

Chapter 2 Literature Review

This chapter begins by surveying the current state of research on System Dynamics (SD), focusing on sensitivity analysis methods and their treatment of parameter distributions. It then reviews Bayesian Networks (BN), examining the role of Bayesian Networks in capturing dependencies in data and how copulas have traditionally been used to model correlations. It also reviews Mixing Methods, a theoretical framework that underpins combinations of different methods such as SD and BN to solve complex problems.

2.1 System Dynamics

2.1.1 Origin and Philosophical Foundations

SD was invented by applying ideas from servomechanism, including feedback control and delays, to the study of socio-economic systems (Forrester, 1961). The studied systems were seen as counterpart of mechanical systems in the socio-economic sphere. A strongly positivist, deterministic, and realistic worldview was adopted. Models were used to describe what exists in the social world and validated using statistical methods to ensure the simulation results they generated comply with historical records (Sterman, 1984).

In the 1990s the diagrammatic representation of SD was combined with participatory methods to support group decision-making, creating the field of Group Model Building (GMB) (Vennix, 1996, 1999). GMB uses SD under a different worldview that people's perception of their situation influences their behaviour and in turn affects the broader environment. This worldview assumes that social reality is of shared and intersubjective meaning. Under this worldview, SD was used in GMB as a tool to create such shared meaning through eliciting subjective knowledge from stakeholders (Vennix, 1996).

Lane (Lane, 1999) offered an analysis of the philosophical basis of different SD schools, based on which he suggested that SD has the potential to bridge the gap of 'dualism' in social science studies, which is a philosophical divide between the objective, structure-centred worldview and the subjective, agent-centred worldview. This suggestion is partly resonated by Mingers (2006) who advocates the social theory of Critical Realism (CR). Critical Realism adopts a mixed worldview, assuming the existence of social reality and mechanism but also recognises individuals' subjective understanding of the reality; it is through a continuous, dynamic process of hypothesis-test that people contrast their understanding with what they observe and therefore advance their knowledge. Mingers suggests that SD could be a suitable tool to practice the CR ideology, because firstly, SD can represent individuals' understanding of social existence and secondly, SD practises the hypothesis-test process through building and testing a model, which makes it an ideal bridge between the mentioned two worldviews (Mingers, 2006).

2.1.2 Formal Representation of Systems

2.1.2.1 Diagrammatic Representation

SD has two diagrammatic representations, stock-and-flow diagram (SFD) and causal loop diagram (CLD). SFDs distinguish between stocks, flows, and the rest of variables, which are represented using different symbols, while in CLDs all variables are represented in the same way. All SD models that can be simulated are capable of representation as SFDs. SFDs are originally used by Forrester in *Industrial Dynamics* to represent SD models (Forrester, 1961). CLDs emerged later, over time they have been identified as useful tools for system description and model conceptualisation (Coyle, 1977; Goodman, 1997; Randers, 1980; Sterman, 2000).

Morecroft (1982) reviewed the two representations. He found that SFDs has the strength of being explicit about the types of variables in the system but requires more details about the system to construct, while on the other hand, CLDs have the strength of providing an overview of loop structure and are useful for behaviour analysis but lack discrimination of the different types of variables. He critiqued that both representations have weaknesses in facilitating model conceptualisation as they lack representation of (1) the groups of components of the system and (2) the decision-making process.

SD models a system as variables and causal relations between them. SD distinguishes between two types of causal relations: instantaneous causal relations and cumulative causal relations. The key difference between them is that when the cause presents, in an instantaneous causal relation the effect takes place immediately, while in a cumulative causal relation the effect takes place either gradually over a period or immediately but after a period. This period is called a "delay", which is a key concept in SD.

If the effect of a cause eventually traces a series of influences back to the cause itself through multiple connected causal relations, this circular causal relation is termed "feedback", and the pathway formed by the connected causal relations is termed a "feedback loop". It can be deduced that a feedback loop must include at least one cumulative causal relation, because otherwise the effect of the cause will come back to the cause immediately. The number of cumulative causal relations (i.e., delays) within a feedback loop is termed as the "order" of the feedback loop.

SD distinguishes between two types of feedback loops: reinforcing feedback loops and balancing feedback loops. If the change of a variable in a certain direction eventually causes the variable to change further in the same direction (self-enhancing), the feedback loop is a reinforcing loop. Otherwise, if the change of a variable in a certain direction eventually causes the variable to change in the opposite direction (self-correcting), the feedback loop is a balancing loop. Figure 2.1 shows an example of an SFD and a CLD of the same system consisting of a negative feedback loop (balancing loop).

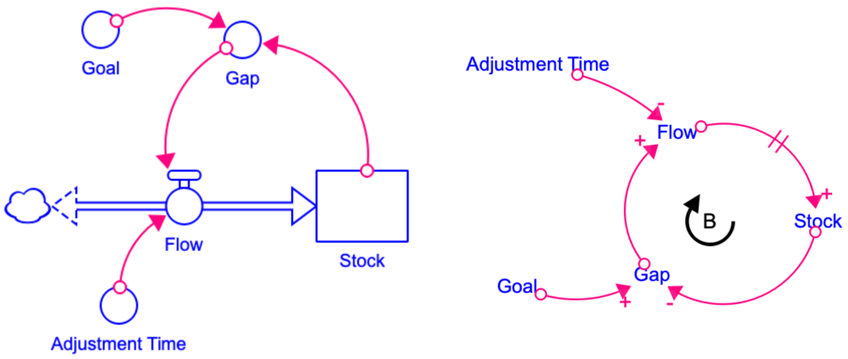


Figure 2.1 Stock-and-flow diagram and causal loop diagram of a negative feedback loop

2.1.2.2 Types of Variables

There are three basic types of variables in SD models: stocks (also known as levels), flows, and auxiliaries (convertors). Stocks are used to model real-world entities of which the level takes time to adjust. Flows are used to model the adjustments in their levels. Commonly used examples include (1) arrival (inflow) and shipment (outflow) of goods adjust the level of storage (stock) and (2) perception of the market (flow) adjusts the knowledge of a customer (stock). These examples show that stocks can be used to model both tangible levels (materials, such as goods) and intangible levels (information, such as knowledge) in the system. A principle of SD is that stocks can only be adjusted by flows. The rest of the elements in the system are modelled as auxiliaries, including intermediate variables, external inputs that vary over time, and constant external inputs (i.e. parameters).

Recent developments in SD software such as Stella have introduced more variants of stocks, for example "conveyors" on which materials are first-in-first-out (FIFO) and "ovens" which release all their content at a certain time (iseesystems, 2020). These variants enhanced SD's suitability to discrete problems without compromising the fundamental feedback principles.

2.1.3 Dynamic Behaviour

In SD, the dynamics of a system refers to the change of the system's state over time, also known as the system's dynamic behaviour, which is technically the change of the value of all variables over time. SD employs an endogenous perspective, which seeks to explain the most characteristic dynamics of the system's behaviour through the endogenous feedback structure of the system rather than exogeneous drivers (such as a varying input from outside the system). This leads to another principle of SD "structure drives behaviour". This principle suggests that (1) when characterising the system, changes in state of the system are attributed to both the state of the system itself and inputs from outside the system; and (2) when building the model, flows are determined (i.e., calculated from) by both the levels of the stocks and external inputs.

The origin of the dynamic behaviour can therefore be outlined as follows: at any certain point in time, the current state of the system and the external input to the system jointly determine the change that is "about to happen"; this change is then added to the state to determine the state of the system at the next point in time. In the real world, we normally see time as continuous[1]; however, in SD models, we see the dynamics of a system as a series of states at multiple discrete time points.

The interval between two such time points is referred to as a "time step", also known as a "dt" (delta time) in SD simulation packages such as Vensim or Stella. A simulation run of an SD model normally involves multiple dts. Within each dt, two steps of calculation are carried out: the software first calculates the values of the flows using the values of the stocks and external inputs, then adds the flow values to (or subtract them from) the stocks that the flows are connected to. Recalling the dichotomy between instantaneous causal relations and cumulative

[1] Modern physics sees time as discrete with an interval of 5.39×10^{-44}s, which is referred to as Planck time (Wendel et al., 2020). This does not affect that conventionally, people in their daily life still see time as continuous.

causal relations, we can see the first step as a realisation of all instantaneous causal relations, and the second step as a realisation of all cumulative causal relations.

In SD models, instantaneous causal relations are represented in the form of analytic equations. These equations denote how one variable is arithmetically calculated using others. These equations reflect the modeller's belief and knowledge about the underlying causal mechanism in the real world (Forrester, 1961; Sterman, 2000).

2.1.4 Data Sources and Utilisation

Forrester (Forrester, 1980) classifies information sources for SD modelling into three types: mental data, written data, and numerical data. Mental data refers to knowledge possessed by experts and stakeholders, also known as mental model (Sterman, 2000). Written data refers to the recorded mental data and "concepts and abstractions that interpret other information sources" (Forrester, 1980, p. 557). Numerical data refers to time-series data of interested quantities and some parameter values. Forrester believes that mental data are the richest source of information for building an SD model, while numerical data are the least informative as they do not themselves provide causal knowledge but are only useful to validate a model by its behaviour.

SD modellers utilise mental data through working with clients on their problems and eliciting their causal knowledge (Forrester, 1980). Techniques have been developed for modellers to work with group clients, for example group model building (GMB) (Andersen and Richardson, 1997; Scott, 2018; Vennix, 1996) and participatory system dynamics modelling (PSDM) (Antunes et al., 2006; Videira et al., 2017).

Sterman agrees that mental data is a critical source of information for SD models (Sterman, 2000), but argues that numerical data are more informative in the contemporary time than in Forrester's time, due to the advance in analytical methods over time (Sterman, 2018). These methods include using summary statistics to evaluate the fit of the SD model to historical data (Sterman, 1984), validating the model by calibration (Oliva, 2003), using Bayesian statistics to reduce uncertainties in value of parameters and formulation of equations (Osgood and Liu, 2015; Rahmandad et al., 2021; Wakeland and Homer, 2020; Zhou et al., 2020), and using machine learning to identify possible model structure from empirical data (Abdelbari and Shafi, 2017; Chen et al., 2011; Drobek et al., 2015, 2014; Duggan, 2008; Jeng et al., 2006; Schoenberg, 2020). Pruyt (2016) reviewed the data science methods and techniques used in SD modelling for processing both empirical data and simulated data, and called for more investments from the SD community into the integration of big data and SD.

2.1.5 Modelling Process

SD modelling follows a multi-stage process, normally including (1) describing a real-world problem and the system it is embedded in, (2) developing a conceptual model and a simulation model, (3) calibrating the model and running simulations, and (4) drawing insights and policy suggestions (Forrester, 1994; Richardson and Pugh, 1981; Sterman, 2000). This process often involves significant iterative work where a modeller returns to a previous stage from a later stage to further clarify the problem, update causal assumptions, modify model structure, or collect additional data to characterise key quantities (Homer, 1996).

The SD modelling process can result in a range of outcomes or combinations of them, including (1) a better organised, clarified, and unified body of knowledge of the system where the problematic behaviour is observed (Forrester, 1987); (2) a better understanding of the cause for the failure of a previous policy (Forrester, 2007, 1969); (3) a shared learning of a group if multiple stakeholders are involved (Rouwette et al., 2002; Vennix, 1996); (4) a better understanding of the dynamics between different drivers and potential policy levers (Ford and Flynn, 2005; Ford, 1999; Guneralp, 2006; Schoenberg et al., 2020), (5) a forecast of the system's future behaviour under reasonable scenarios (Lyneis, 2000), and finally (6) policy suggestions made through model-based testing and optimisation (Fiddaman, 2002; Hamarat et al., 2014). The exact outcome would depend on the purpose of the work. The impact of the work can be negatively affected by various adverse factors, such as low degree of involvement of the organisation's executives (Grossler, 2007).

2.1.6 Debates around System Dynamics

As a management science / operational research method, SD has an engineering origin but is also used to model and simulate social systems. SD models are built with both expert knowledge and empirical data. It is used for both creating shared knowledge of systems and for bringing changes to them.

The SD field has seen philosophical and methodological debates since its early days. Such debates include (1) model validity versus usefulness, (2) numerical precision versus plausibility of trend of model behaviour, and (3) whether a quantitative simulation model is always better than a qualitative one. Review of the three debates helps elaborate the historical and contemporary concerns of the method and highlights the value of this work.

The first debate is about SD models' validity. SD has been criticised as not being sufficiently scientific as it does not employ "formal, objective, quantitative model validation procedures" (Barlas and Carpenter, 1990, p. 148; Nordhaus, 1973; Zellner, 1980). The counterargument is made based on an acknowledgement of different philosophical schools: unlike econometrics which is based on empiricist philosophy of science, SD is more consistent with relativist philosophy; justification of SD models is a social process which establishes usefulness to stakeholders rather than objective validity (Barlas, 1996; Barlas and Carpenter, 1990; Meadows, 1980). However, not all SD works position themselves within the interpretivist philosophy (Lane, 1999). Recent developments in philosophy have brought forward Critical Realism, which seeks to bridge the gap between the empiricist and relativist schools (Bhaskar, 2013). In some's view, SD has been advocated for as a method consistent with Critical Realism, for its capability of representing beliefs and critically selecting them in light of social realities (Lane, 1999; Mingers, 2006).

The second debate is about the interpretation of SD models' behaviour. SD heavily uses simulation but there have been different opinions on how the outcomes (i.e., the simulated data) should be interpreted. Criticisms have been made about the lack of numerical validation of the simulations (Nordhaus, 1973; Zellner, 1980). Others believe the simulation reveals more of trends, and therefore its analysis should focus more on patterns rather than numerical precision (Barlas and Kanar, 1999; Hekimoglu and Barlas, 2016; Yücel and Barlas, 2007). Some argue if rigorous summary statistics is used, numerical validation is possible and can add to SD models' credibility (Sterman, 1984). Recent developments use sensitivity analysis and Bayesian statistical inference to quantify the uncertainty in simulations, using a calculated degree of uncertainty to reconcile the different opinions on how model behaviour should be interpreted (Ford and Flynn, 2005; Jadun et al., 2017, p. 201; Kwakkel and Pruyt, 2013; Osgood and Liu, 2015; Rahmandad et al., 2021; Wakeland and Homer, 2020).

The third debate is about the quantification of SD models. As reviewed in section 2.1.2, there are two mainstream representations of SD models: causal loop diagrams (CLDs) and stock-and-flow diagrams (SFDs). CLDs only involve the causal relations between variables and are more suitable for qualitative SD models. However, if one wishes to build a simulation model, they will need to turn the CLD into an SFD and specify analytic equations for each of the variables. This leads to a problem: some variables in a CLD are not straightforward enough to be represented by a quantity, neither are their causal relations straightforward enough to be represented by an equation. These variables are referred to as "soft" variables and their quantification are made difficult by (1) uncertainty in their causal relations and (2) lack of data, sometimes resulting in misleading values and equations, making simulations unreliably (Coyle, 2000). Coyle (2000) reviewed several cases where quantification of models caused problems and suggested (1) more careful considerations are needed when deciding whether to pursue a simulation model (2) more research is needed on quantifying soft variables, handling non-linear effects, and measuring the impact of uncertainties introduced in quantification on model reliability. Homer and Oliva (2001) counterargued that model quantification, if following a rigorous manner, can almost always add value beyond causal mapping and suggested the use of analytical methods such as sensitivity analysis to address uncertainties with soft variables and incomplete data. A similar point is made by Warren (2004), where he discussed the downside of employing only the feedback perspective when tackling dynamic problems and called for more attention on the stock-flow accumulation. The field has later seen attempts to improve model quantification, for example using multiple criteria decision analysis (MCDA) to facilitate the quantification of soft variables (McLucas, 2003), and using Bayesian inference to address uncertainties in equation formulation in light of empirical data (Zhou et al., 2020).

These three debates originate from the fundamental characteristics of SD and have profound impact in the field till today. They also help to reveal some weaknesses of SD, for example the difficulty of learning SD models from data, as pointed out by Pruyt (2016) and Sterman (2018).

2.2 Parameter Estimation of System Dynamics Models

Parameters in SD models are exogenous variables that are constant and not dependent on other variables in the model. A formal quantitative SD model can only be simulated when it is fully parameterised, i.e., each parameter is assigned with a value. Parameters values may be determined via formal statistical estimation from numerical data or judgemental estimation from expert opinion (Sterman, 2000). This section reviews the literature on parameter estimation for SD models.

2.2.1 Judgemental Methods for Parameter Estimation

Judgemental methods estimate parameter values using expert opinion gleaned from interviews, workshops, archival materials, direct experience, and other methods (Sterman, 2000). Judgemental methods are effective ways to utilise mental data – or in other words, mental models, which Forrester (1980) believed to be holding the majority of information relevant to the system among three types of data (mental, written, and numerical).

Methods to elicit information from experts for SD modelling have been developed (Antunes et al., 2006; Scott, 2018; Vennix, 1996). Ford and Sterman (1998) observed that most of such methods have been used in the early phases of modelling to articulate problem, select model boundary, identify variables and map the causal relationships. Meanwhile, less literature exists that addresses methods for eliciting the specific information necessary to accurately estimate parameters, establish initial conditions, and define behavioural relationships required in the later phases. An examination of recent literature indicates that this observation remains valid today.

One notable exception is Lee et al. (1998), who outlined a detailed process to elicit key parameters from a group of experts. The process includes three steps, namely "clarifying parameter definition and location", "estimating parameters by individual judgement", and "turning numerical value of parameters by group judgement". Similarly, Ibrahim et al. (2021) described a comparable individual-then-group judgement elicitation process, albeit with less specificity. Although these processes are operable and valuable, they do not explicitly address the determination of parameter *bounds*, which reflect expert confidence in parameter values or indicate the plausible ranges of parameters. Consequently, this omission results in an absence of quantified parameter uncertainty and, subsequently, a lack of measurable uncertainty in model behaviour.

Beyond the field of System Dynamics, the broader literature has extensively examined the elicitation of expert judgement on uncertainties as probability distributions. Quigley and Walls (2021) and Quigley et al. (2008) provide comprehensive reviews of these methods. Among the approaches identified, the extended Stanford Research Institute (SRI) Elicitation Process (Merkhofer, 1987; Spetzler and Stael von Holstein, 1975) is highlighted as a robust framework. This method comprises seven stages of expert engagement: motivating, structuring, conditioning, encoding, verifying, aggregation, and discretisation.

However, a significant challenge associated with these processes is the cognitive burden placed on experts. Implementing the full procedure can demand substantial time and intensive engagement, a burden that can hinder practical application (Barons et al., 2022; Werner et al., 2017). Additionally, expert elicitation is susceptible to various biases at different stages, most notably motivational, cognitive, and methodological biases, for which techniques have been developed for mitigation (Cooke, 1991; Ferrell, 1985; Gosling, 2018; Moore, 1987; O'Hagan et al., 2006; Spetzler and Stael von Holstein, 1975).

2.2.2 Statistical Methods for Parameter Estimation

Statistical methods estimate parameter values using numerical data and serve as a complement to judgemental methods. Graham (1980) identified two categories of data applicable to parameter estimation. The first category comprises data *below* the level of aggregation of model variables, commonly referred to as "unaggregated data". These data describe individual elements within a stock or discrete events within a flow, effectively characterising the components of a model at a granular level (p. 144). For instance, estimating the average lifespan of houses in a region by calculating the mean of individual houses' lifespans exemplifies the use of unaggregated data. Graham argued that, in most System Dynamics applications, parameter values are predominantly derived from such descriptive, disaggregated data, noting that all factual knowledge about a system – except for aggregated statistics aligned with model variables – belongs to this category.

The second category encompasses data *at* the level of aggregation of model variables. Such data typically appear as system-level statistics and can be used either to directly assign values to parameters and initial stock conditions or to indirectly derive them through model equations. In the latter approach, equations that define relationships among variables are calibrated using available data to approximate the real-world relations. Continuing the earlier example, the average lifespan of houses can also be estimated using aggregate data through the model equation: *rate of house demolition = number of house / average house lifespan*, assuming data on housing stock and demolition rates are available. This process resembles linear regression, where coefficients are inferred to ensure that model behaviour aligns with observed relationships (Graham, 1980).

Beyond estimating parameters using a single equation, it is also common to use the entire set of model equations – that is, the full System Dynamics model – to infer one or more parameter values (Graham, 1980; Sterman, 2000). This approach involves running repeated simulations to identify parameter values that allow the model to reproduce historical time series data of key stocks or flows. Unlike the analytical single-equation method, this whole-model approach is numerical and simulation-based. As Sterman (2000) notes, traditional regression techniques are generally inadequate for estimating SD model parameters due to the presence of feedback, nonlinearity, and dynamic complexity. To estimate parameters numerically, methods such as maximum likelihood (Struben et al., 2015) and maximum a posteriori estimation (Osgood and Liu, 2015; Wakeland and Homer, 2020) are employed, often in conjunction with adaptive sampling techniques like Markov Chain Monte Carlo (MCMC) (Osgood and Liu, 2015) or Hamiltonian Monte Carlo (HMC) (Andrade and Duggan, 2020). However, a key limitation of these approaches is that a good fit to historical data does not ensure the correctness of parameter values, as different parameter sets can produce similar system behaviour (Graham, 1980; Sterman, 2000).

To address this ambiguity, Graham (1980) suggests that greater confidence in model validity is achieved when parameter values derived from disaggregated data (rather than from curve-fitting) also enable the model to reproduce historical data. Indeed, if a model fits observed patterns without being explicitly tuned to do so, that replication strengthens confidence in the model's structure and parameter assumptions. Sterman (2000) also emphasises the value of integrating statistical and judgemental methods. Expert knowledge helps constrain the plausible ranges for parameters, while statistical methods provide a means to test and refine these estimates empirically.

2.2.3 Special Treatment for Large Models

For large or complex models, partial model estimation (Homer, 2012) offers a practical solution to the difficulty of estimating all the critical parameters at one time. This method involves isolating a key structure or decision rule from the full model – essentially cutting feedback loops – and using historical data to drive its inputs. Parameters are then estimated (either judgementally or statistically) so that the output of this subsystem aligns closely with observed data. This technique has been applied in several studies (Fiddaman, 1997; Homer, 2012; Oliva, 1996; Taylor, 1999) to improve parameter credibility while maintaining model tractability.

Another useful approach for managing parameter estimation in large System Dynamics models is sensitivity analysis. As Sterman (2000) explains, sensitivity analysis plays a crucial role in guiding data collection efforts. Since all parameters inherently involve uncertainty, modelers must prioritise where to focus their resources. While most parameters can, in theory, be estimated more accurately with sufficient time and efforts, it is rarely feasible to refine all of them to the same degree. Sensitivity analysis helps prioritise this process: parameters that significantly influence model behaviour warrant further investigation and more precise estimation, whereas those with minimal impact can often be approximated without compromising the robustness of the results. This targeted approach not only enhances model credibility but also optimises time and effort. For example, Ford and Flynn (2005) introduce a statistical screening method that effectively identifies the most influential parameters, which could be used to guide parameter estimation. Section 2.3 will review more techniques and applications of sensitivity analysis in System Dynamics modelling.

2.2.4 Estimating Interdependent Parameters

Another important challenge in estimating System Dynamics model parameters is the potential interdependence between the parameters. While parameters in SD models are technically constant values rather than variables – and thus, by definition, not interdependent – this distinction becomes blurred during the estimation process. When parameters are treated as unknown variables with uncertainty bounds, correlations can emerge between their estimation errors, reflecting underlying interdependencies (Bertone et al., 2018; McNaught, 2003).

This phenomenon is not unique to SD modelling but is common across system modelling and system identification more broadly. For instance, Krefeld-Schwalb et al. (2022) examined parameter correlations in computational models of cognition and found that strong intercorrelations can impair estimation accuracy. Similarly, Li and Vu (2013) studied this issue in nonlinear dynamic biological systems, identifying data requirements that help mitigate such effects.

In the SD literature, the impact of parameter dependencies on parameter estimation remains largely unexplored. Osgood and Liu (2015) examined how the choice of prior distribution of parameters – which implicitly encode assumptions about dependencies – affects parameter estimation from empirical data. They found that these dependencies exerted only a limited influence compared with the effect of the empirical data itself. However, the generalisability of this finding remains uncertain, particularly given the relatively small scale of the experimental model. Wakeland and Homer (2020) also briefly addressed this issue, noting informal comments from peer modellers who argued that “ranges [of uncertain parameters] should not include implausible values” – a statement that, when interpreted broadly, implies the exclusion of implausible combinations of values from different parameters. Wakeland and Homer concluded that a more systematic investigation of this question is warranted.

2.3 Sensitivity Analysis of System Dynamics Models

2.3.1 Sensitivity Analysis in System Dynamics

Sensitivity analysis is “the assessment of how uncertainty in a model’s output can be apportioned to uncertainty in the model’s input factors” (Jadun et al., 2017, p. 312; Saltelli et al., 2010). It allows for “the exploration of model behaviour across a broad range of conditions” (Jadun et al., 2017, p. 312). Sensitivity analysis is important in decision support using simulation models in that it helps to understand the response of the model’s behaviour to different policy interventions.

Sterman (2000) identified two types of sensitivity analysis in System Dynamics (SD): univariate sensitivity analysis and multivariate sensitivity analysis. Univariate sensitivity analysis involves varying a single input variable while observing its effect on the output variable, providing insights into the isolated impact of that variable. In contrast, multivariate sensitivity analysis examines the joint effects of multiple input variables on the output variable. This is achieved by systematically and simultaneously varying all relevant input variables to assess their combined influence on the model's behaviour. The focus of this case study is on multivariate sensitivity analysis, given its relevance to understanding the interplay of multiple factors in complex systems.

The use of sensitivity analysis in the field of System Dynamics has a long history. Eberlein (1984) described sensitivity analysis as a tool for simplifying SD models. Sterman (2000) in his textbook differentiated between local sensitivity analysis and global sensitivity and suggested modellers to do both. Kleijnen (1995) employed linear regression to calculate coefficients of the input variables and their interactions with the output variable. Ekker (2014) applied sensitivity analysis to graphic functions to study their impact to model behaviour. Hekimoglu and Barlas (2016) used behaviour patterns as the indicator for simulation output and analysed the pattern sensitivity, instead of numerical sensitivity, to inputs. Ford (1990) and Ford and Flynn (2005) used partial correlation coefficients between input and output variables to identify important inputs, a method known as ‘screening’ that is similar to Sobol’s indices in that both use the fractions of variable variances to total variance. Jadun et al. (2017) used a variance decomposition-based method for sensitivity analysis, applying both Sobol’s indices and total effect indices in their sensitivity analysis of an SD model and calculated them using a Monte Carlo method, and concluded that variance-decomposition based methods can quantitatively identify interaction effects distinguishes itself from other methods previously adopted in the field.

2.3.2 Parameter Interdependence in Sensitivity Analysis

Current sensitivity analysis practices in SD modelling is the common assumption that parameters are independent (e.g., Jadun et al., 2017). This assumption often fails to reflect real-world conditions. McNaught (2003) Eker et al. (2014) have highlighted that certain combinations of parameter values can lead to unrealistic or internally inconsistent inputs, producing implausible simulation outcomes. While parameters are typically treated as exogenous boundaries of an SD model (Sterman, 2000), there is no inherent guarantee that they are independent. In practice, parameters may be interdependent or influenced by shared underlying factors – a phenomenon known as confounding (as discussed in Section 2.2.4).

As a result, sensitivity analyses that treat parameters as independent overlook critical interactions and ignore the constraints these dependencies impose on the joint combinations of parameter values. This can significantly distort the distribution of simulation outputs, undermining the reliability of the analysis (McNaught, 2003).

One direct strategy to handle parameter dependencies is to embed them within the model structure itself, by specifying one parameter as a function of another (Ford, 1990; Ford and Flynn, 2005), using either analytical equations or graphical functions (Eker et al., 2014; Rios-Ocampo and Gary, 2022; Sterman, 2000). These approaches are adequate for low-dimensional or well-understood relationships but become cumbersome and insufficient when the dependencies are complex, high-dimensional, or poorly specified.

This challenge has motivated the exploration of more flexible modelling approaches. McNaught (2003) proposed using Bayesian Networks (BNs) to represent interdependencies among SD model parameters, including latent variables that exert simultaneous influence on multiple parameters. BNs offer a structured way to encode parameter relationships while maintaining probabilistic coherence. Similarly, Bertone et al. (2018) employed an expert-elicited BN to reflect how broader policy and contextual factors shape parameter collectively, enabling coordinated variation of parameters across different simulation scenarios.

While expert-elicited BNs offer a powerful means of incorporating domain expertise, they also present cognitive and practical burdens (Barons et al., 2022), particularly as the number of parameters and their interconnections increase. An appealing alternative is to learn parameter dependencies directly from empirical data, such as cross-sectional or panel datasets representing multiple instances of similar systems. Although such data-driven BN learning methods are well established (and reviewed in Section 2.6), their application to sensitivity analysis in SD remains largely unexplored. The closest example is provided by Rahmandad et al. (2021), who used a hierarchical Bayesian model to estimate pandemic model parameters from a cross-country dataset. However, their focus was on coupling the same parameter across different countries, rather than uncovering dependencies between different parameters. This therefore represents a notable gap in the current literature and an opportunity for methodological advancement.

2.3.3 Discussion on Parameter Dependencies

The preceding sections indicate that parameter dependencies in SD models are relevant to parameter estimation, sensitivity analysis, and scenario exploration – contexts where uncertainties across multiple parameters are present. However, the impact of such dependencies remains insufficiently explored in all these situations, with only sporadic mentions by a few authors. A comprehensive investigation of this issue would require a systemic assessment of its effects and a comparative evaluation of available tools for modelling parameter dependencies. While a full-scale treatment is beyond the scope of this thesis, we design a focused experiment to illustrate how parameter dependencies can affect KPI uncertainty assessment for a specific SD model, thereby motivating the development of a method for modelling such dependencies when necessary.

In Section 2.4, we review the literature on copula, a class of statistical tools capable of modelling a range of dependence structures with varying strengths of correlation, which can support the experimental design. Furthermore, the literature suggests that Bayesian Networks (BNs) are a promising method for capturing complex and multidimensional parameter dependencies in SD models (Bertone et al., 2018; McNaught, 2003). Section 2.5 reviews BNs and their relevant features for this purpose.

2.4 Modelling Multivariate Dependence Structure with Copula

2.4.1 Copulas: An Approach to Correlated Sampling

Trivedi and Zimmer (2007) defined copulas as "functions that connect multivariate distributions to their one-dimensional margins". A copula exclusively captures the dependence structure among variables in a multivariate distribution and can be integrated with any set of univariate distributions to define the marginal distributions (Ruppert, 2011). According to Sklar's (1973, p. 449) theorem: "If G is an n-dimensional joint distribution function with 1-dimensional margins $F_1, \cdots, F_n$, then there exists a function C (called an "n-copula") from the unit n-cube to the unit interval such that

$$G(x_1, \dots, x_n) = C\big(F_1(x_1), \dots, F_n(x_n)\big)$$

for all real n-tuples $(x_1, \dots, x_n)$."

Copulas are valuable in modelling dependence structures across various fields, including finance, actuarial science, and micro-economics, capturing tail dependence, modelling joint survival probabilities, and establishing the joint distribution of variables from different parametric families (Trivedi and Zimmer, 2007).

2.4.2 Archimedean and Elliptical Copulas

2.4.2.1 Archimedean Copulas

Archimedean copulas constitute a widely used family of copulas due to their mathematical simplicity and flexibility in capturing various dependence structures. "Archimedean copulas are popular because they are easily derived and are capable of capturing wide ranges of dependence." (Trivedi and Zimmer, 2007, p. 41). Archimedean copulas are defined through an inverse Laplace transform applied to a generator function. According to Ruppert (Ruppert, 2011, p. 178), a copula is an Archimedean copula if its generator function is in the form:

$$C(u_1, \dots, u_d) = \Phi^{-1}\{\Phi(u_1) + \cdots + \Phi(u_d)\},$$

where the function φ if the generator of the copula and satisfies

1. Φ is a continuous, strictly decreasing, and convex function mapping [0, 1] onto [0, ∞].
2. $\Phi(0) = \infty$, and
3. $\Phi(1) = 0$.

Three of the most frequently applied Archimedean copulas are the Clayton, Gumbel, and Frank copulas. These copulas differ in how they capture dependence, particularly in the tails of the distribution. The Clayton copula, for instance, is well-suited for modelling strong lower tail dependence, making it applicable in scenarios where extreme negative outcomes tend to co-occur. Conversely, the Gumbel copula is effective in capturing upper tail dependence, making it useful in cases where extreme positive values exhibit strong dependence. The Frank copula, unlike the previous two, is more symmetric and does not emphasise tail dependence as strongly. Due to their simple analytical forms, Archimedean copulas are particularly advantageous in bivariate modelling or cases where dependencies across variables are expected to be similar.

Frank Copula

The Frank copula is:

$$C^{Frank}(u_1, \ldots, u_d) = -\frac{1}{\theta}\log\left\{1 + \frac{\prod_i \left(e^{-\theta u_1} - 1\right)}{e^{-\theta} - 1}\right\}, -\infty < \theta < \infty.$$

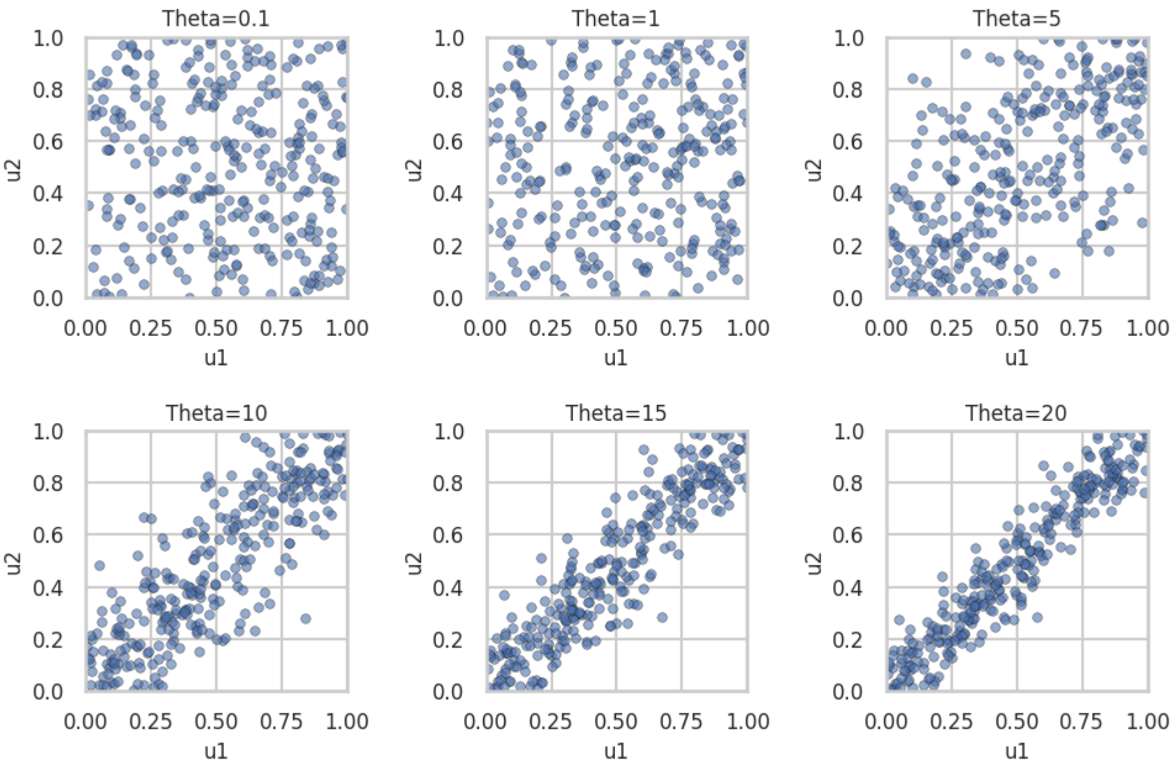


Figure 2.2 Samples from Frank copula with 6 different theta values

Figure 2.2 shows 300 samples drawn from Frank copula with 6 different theta values. The samples are symmetrically distributed with no increased dependence on both tails.

Clayton Copula

The Clayton copula is:

$$C^{Clayton}(u_1, \dots, u_d) = \left(u_1^{-\theta} + \cdots + u_d^{-\theta} - d + 1\right)^{-\frac{1}{\theta}},\ \theta > 0.$$

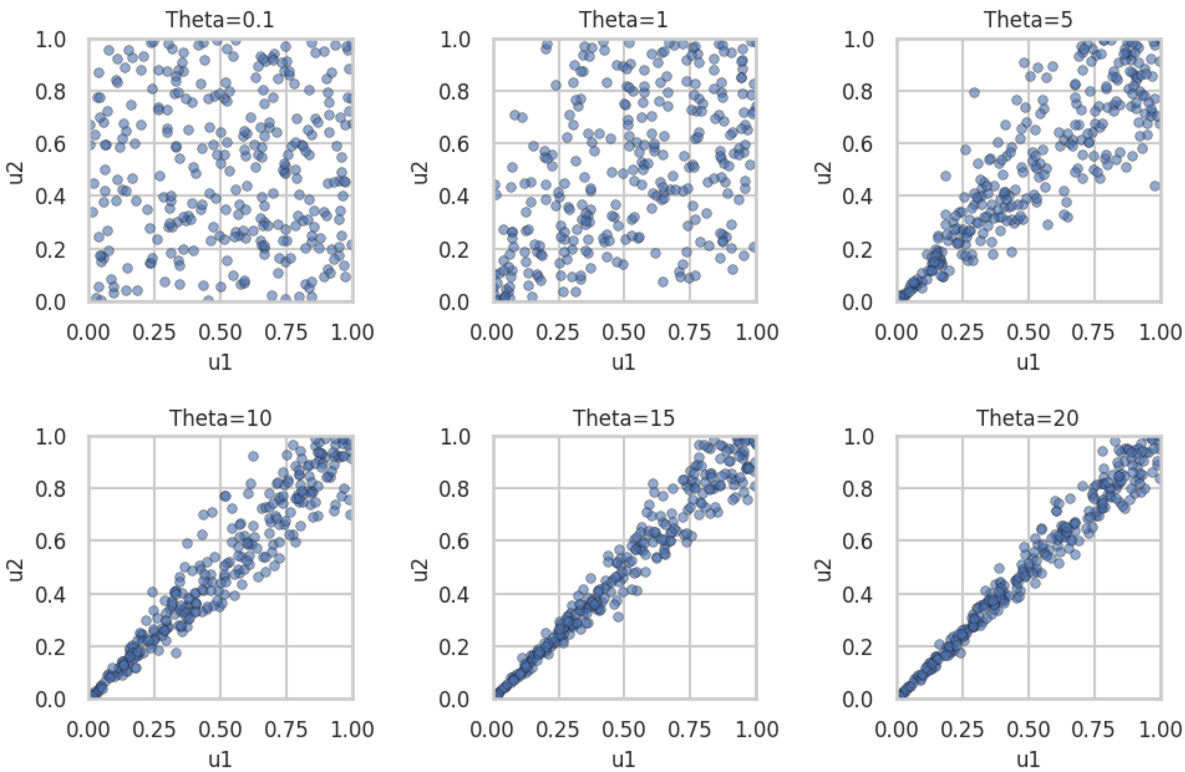


Figure 2.3 Samples from Clayton copula with 6 different theta values

Figure 2.3 shows 300 samples drawn from Clayton copula with 6 different theta values. The samples show stronger lower tail dependence as theta increases.

Gumbel Copula

The Gumble Copula is:

$$C^{Gumbel}(u_1,\ldots,u_d) = exp\left[-\{(\log u_1)^\theta + \cdots + (\log u_d)^\theta\}^{\frac{1}{\theta}}\right], \theta \geq 1.$$

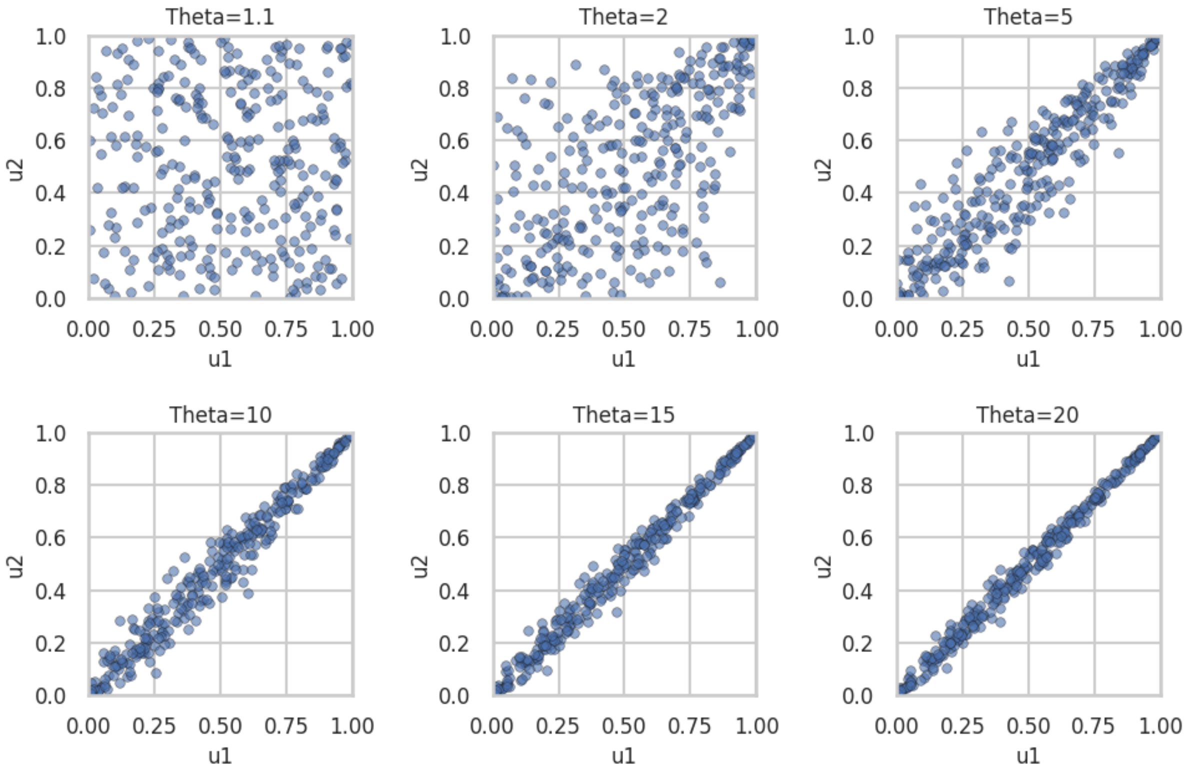


Figure 2.4 Samples from Gumbel copula with 6 different theta values

Figure 2.4 shows 300 samples drawn from Gumbel copula with 6 different theta values. The samples show stronger upper tail dependence as theta increases.

2.4.2.2 Elliptical Copulas

Elliptical copulas, including the Gaussian and t-copulas, originate from multivariate elliptical distributions, such as the multivariate normal or t-distributions. These copulas are characterised by their dependence solely on the correlation structure of the underlying variables rather than their marginal distributions (Ruppert, 2011, p. 177). In the case of the Gaussian copula, dependence is fully determined by the correlation matrix, denoted as R:

$$C^{Gauss}(u_1, \ldots, u_d | \Omega) =$$

$$C_R^{\text{Gauss}}(u) = \Phi_R\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\right)$$

where Φ^{-1} is the inverse cumulative distribution function (cdf) of a standard normal distribution and Φ_R is the joint cdf of a multivariate normal distribution with mean vector 0 and covariance matrix equal to the correlation matrix R.

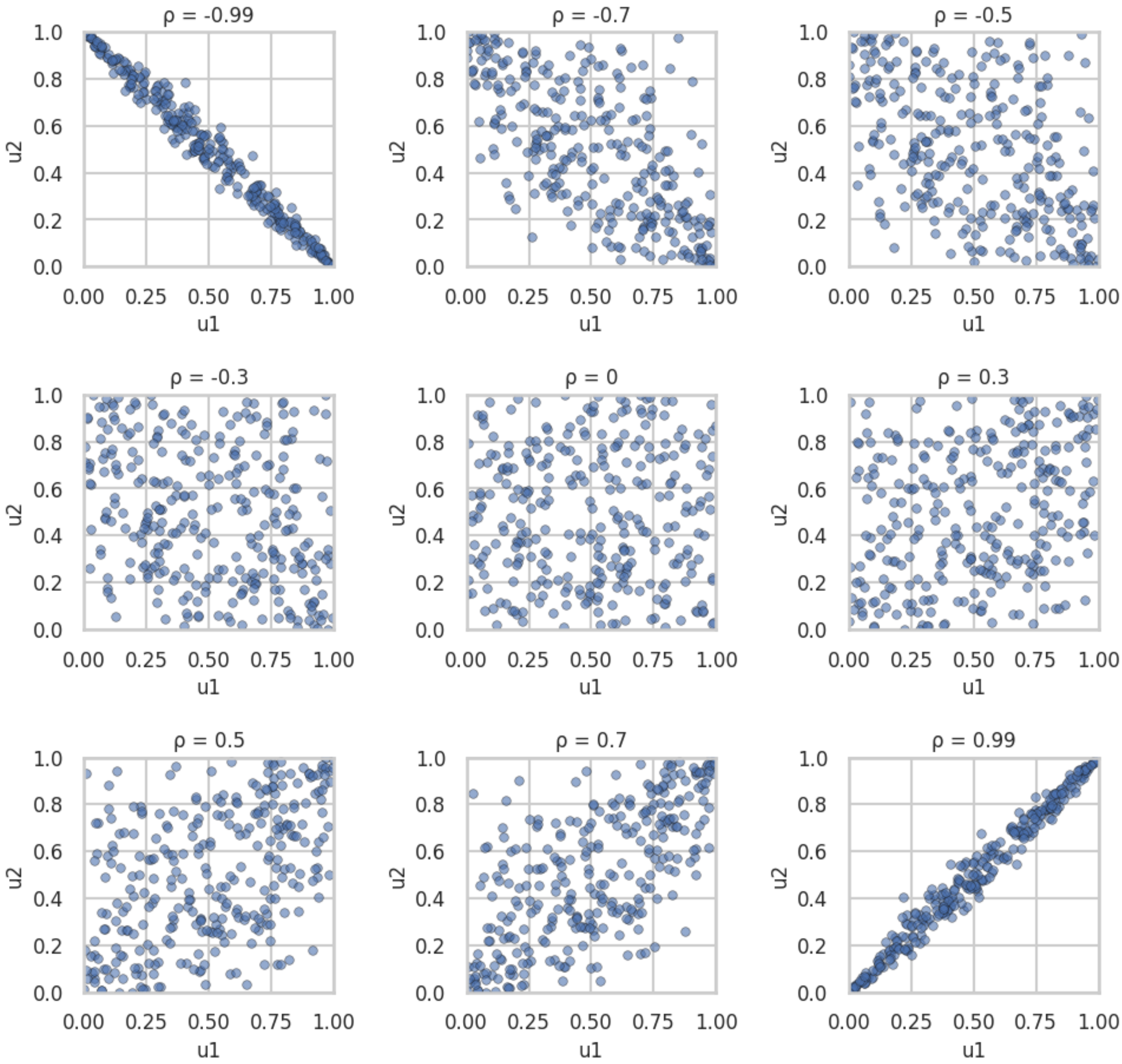


Figure 2.5 Samples Gaussian copula with 9 different levels of correlation

Figure 2.5 shows 300 samples drawn from Gaussian copula at 9 different levels of correlation. The samples show stronger dependence on both tails as correlation between the two dimensions increases.

The t-copula extends the Gaussian copula by introducing an additional shape parameter, ν, which controls the heaviness of the tails. This feature enables the t-copula to model tail dependence. The degree of tail dependence increases as ν decreases, allowing for a more flexible representation tail dependence (Ruppert, 2011, p. 178).

While both Archimedean and elliptical copulas provide valuable tools for dependence modelling, their suitability depends on the nature of the application. Archimedean copulas are preferred when computational simplicity and a flexible dependence structure are required, particularly in bivariate settings. In contrast, elliptical copulas offer advantages in multivariate contexts where correlation-based dependence or flexible tail dependence is crucial.

2.4.3 Limitations in Capturing Complex Conditional Structures Using Copulas

Despite their utility for modelling pairwise or simple multivariate dependence, copulas often struggle to capture complex conditional structures that arise in real-world systems. Many widely used copula families, such as Gaussian and the one-parameter Archimedean copulas, assume specific symmetrical or monotonic forms of dependence across variables. In cases where certain variables exhibit context-specific or regime-dependent correlations – such as changes in dependence strength when a particular variable becomes high or low – these families may fail to reflect the true underlying relationships. This is because copulas can encode joint relationships only through one or a limited number of parameters which limits their flexibility. Moreover, in the case of multivariate Archimedean copulas, the correlations between all dimensions are assumed to be identical, which can lead to underfitting or misrepresenting the real multi-lateral dependence (Nelsen, 2006, p. 154). The fact that Archimedean copulas (if used strictly) only model a limited range of positive dependence is another limiting factor for their usefulness. This will be less a problem for modelling assets or risks, but in broader contexts, modelling negative dependence might be necessary (Embrechts et al., 2003, p. 39).

When assessing uncertainties in SD models, such misrepresentation of underlying parameter dependencies can potentially pose risks. For example, parameter draws that systematically omit stronger – or weaker – dependencies in certain operating conditions may yield overly optimistic or pessimistic outcomes in simulation. Practitioners might inadvertently dismiss policies that are robust under realistic conditional dynamics, or they may incorrectly accept flawed strategies due to understated tail dependence. Through nonlinearities introduced by the feedback loops, mismatches in joint parameter sampling may further distort the distribution of model outputs, misleading decision makers, as cautioned by Eker et al. (2014). Consequently, while copulas help unify marginal distributions into a coherent multivariate framework, their limited flexibility for high-dimensional, conditional, and domain-specific dependencies remains a shortcoming in many real-world MS/OR applications. It is for this reason that copulas may not be a suitable method to model interdependent parameters in SD models in real-world contexts. Nevertheless, their ability to explore diverse dependence structures makes them a valuable tool for evaluating the impact of such dependence on model outcomes.

2.5 Bayesian Networks

2.5.1 Origin and Philosophical Foundations

From the early days of its invention, Bayesian Networks (BNs) were closely associated with beliefs as they are designed to model human's inferential reasoning and interpretation of data (Pearl, 1985), for this reason it was also called Bayesian Belief Networks (BBN). Variables are modelled as conditional probability distributions, capturing both their uncertainties and their causalities. The causalities are operationalised through belief propagation, a process of using observed data on some variables to infer the posterior probability distributions of other variables. In this section, we review the philosophical assumptions of BNs, in particular three concepts: probability, causality, and Bayesian statistics.

Probability

Literature has offered multiple distinctions and interpretations of *probability*. A distinction is a pair of contradicting views on a certain characteristic of probability. Distinctions can be used to examine interpretations in terms of which view an interpretation takes regarding a certain characteristic of probability. Williamson (2005) reviewed three distinctions and four interpretations. The three distinctions are:

(1) Single-Case / Repeatable, drawing distinction on whether a random variable can be assigned values once or more than once.

(2) Mental / Physical, drawing distinction on whether a probability is interpreted as a characteristic of an agent's mental state or part of the physical world.

(3) Subjective / Objective, drawing distinction on whether two agents with the same background knowledge can disagree on a probability value without any of them being wrong.

The four interpretations of probability are:

(1) Frequency theory, where probability is defined through repeated observations of the same random trial (Mises, 2014; Reichenbach, 1971; Venn, 1888).

(2) Propensity theory, where probability is attached to the repeatable experiment (i.e., the random trial) instead of to the generated collection of observations (Popper, 1959).

(3) Chance, where the probability of a single-case random variable is determined by 'the whole physical situation' of the universe at that time (Popper, 1990, p. 17).

(4) Bayesianism, where probabilities are interpreted as an agent's rational degrees of belief determined by her background knowledge (De Finetti, 1937; Ramsey, 1926; Williamson, 2005).

This research does not intend to contribute further to the above discussion of philosophical ground of probability but instead will propose to align BN with the philosophical school of critical realism (CR). Further discussion will be offered in the Methodology Chapter.

Causality

The concept of 'causality' has also been under volumes of debates. David Hume is known for his scepticism about causality by arguing that our understanding of cause and effect is not based on a logical or empirical connection between events, but rather on the human habit of associating two events that regularly occur in sequence, and that what humans can perceive is not a connection between cause and effect, but only a constant conjunction of events (Hume, 1896). Bertrand Russell recognised the limitation of the traditional use of 'causality', and advocated for 'functional relationships' or 'laws of change' instead of causal laws, which are more aligned with modern science (Russell, 1912). Similar efforts to Russell's have been made to reduce causal relations to other non-causal concepts (Williamson, 2005, p. 110), including physical process (Dowe, 2000; Salmon, 1998, 1984), physical probabilistic relations, counterfactual laws (Lewis, 1973), and agents' ability to manipulate the cause for a different effect (Menzies and Price, 1993; Price, 1992a, 1992b, 1991).

Williamson (2005, p. 111) claimed that the same distinctions discussed above for probability are also applicable to the different interpretations of causality. Especially, he highlighted the Bayesianism interpretation of causality, which interprets causality as "the causal beliefs that an agent ought to adopt are determined by her background knowledge" (Williamson, 2005, p. 138).

Bayesian Statistics

From an epistemic perspective, Bayesian statistics is used to create causal knowledge. It has been identified as having a goal to learn general laws from observations, and thus follows an inductive way of reasoning (Bernardo and Smith, 1994; Earman, 1992; Howson and Urbach, 2006). Gelman and Shalizi (Gelman and Shalizi, 2013) disagree with this opinion, arguing that the inductive perspective omits the important role prior probabilities play in Bayesian inference. They argue that prior probabilities – probabilities with which a BN is initialised – are substantially elicited from experts' knowledge and can be seen as hypotheses. The later process of updating the prior probabilities with new evidence is virtually to test those hypotheses, during which a prior probability undergoing a major change is deemed as a hypothesis being rejected. This account is aligned with Popper's hypothetico-deductive view of causal discovery (Popper, 2005).

Williamson (2005, p. 148) combined the inductive view and the hypothetico-deductive view of causal discovery and proposed a synthesis. In the synthesis, the hypotheses represented as prior probabilities in Bayesian statistics are seen as induced from the agent's background knowledge. Therefore, the process of causal knowledge discovery is summarised as a four-step scheme, including Hypothesising (inductive reasoning of hypothesis from background knowledge), Predicting (deductive reasoning of expected observation from hypothesis), Testing (contrasting hypothesis against real observation), and Updating (tested hypothesis becoming a new part of background knowledge).

2.5.2 Formal Representation of System

2.5.2.1 Diagrammatic Representation

A BN is a multivariate joint probability distribution (1) factorised using the conditional dependencies between its variables and (2) represented as a directed acyclic graph (DAG) that encompasses the conditional dependencies. In the DAG, a node (or vertex) represents a variable, whereas an edge (or arc) represents an immediate probabilistic dependence or influence relationship, pointing from the influencing variable to the influenced variable. A conditional probability distribution is specified for each node (except for the root node, of which the distribution is not conditional) to represent the dependence of its value on its parent nodes (Pearl, 1985; Sigurdsson et al., 2001). The example DAG shown in **Figure 2.6** represents a joint distribution $P(A, B, C, D)$, with factorisation $P(A, B, C, D) = P(D|B, C)P(B|A)P(C|A)P(A)$.

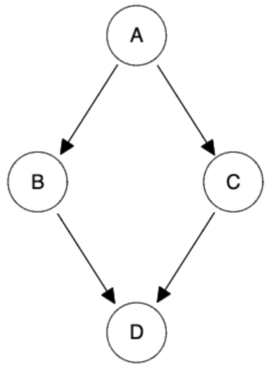


Figure 2.6 A directed acyclic graph

Factorising a joint probability distribution offers strength in studying complex multivariate uncertainties. In a complex system consisting of random variables $X_1, X_2, \ldots, X_n$, a state of the system is defined as the combination of the state of each variable and noted as $(X = x_1, X_2 = x_2, \ldots, X_n = x_n)$, where $x_1, x_2, \ldots, x_n$ are the states of $X_1, X_2, \ldots, X_n$. Here a state of X means one of the multiple possible values X can take. $P(X_1, X_2, \ldots, X_n)$ is the joint probability distribution across all random variables. The number of parameters to store this distribution grows exponentially as the number of random variables grows[2]. However, if knowledge about the interdependencies between variables are leveraged, the joint distribution can be factorised as a product of conditional distributions, following the chain rule of probability:

$$
\begin{aligned}
&P(X_1, X_2, \ldots, X_n) \\
&= P(X_1|X_2, X_3, \ldots, X_n)P(X_2, X_3, \ldots, X_n) \\
&= P(X_1|X_2, X_3, \ldots, X_n)P(X_2|X_3, X_4, \ldots, X_n) \ldots P(X_n)
\end{aligned}
$$

It is common in complex systems that a variable is usually not directly influencing or influenced by many the other variables, but only a few (Watts and Strogatz, 1998). For example,

[2] Assume each random variable has 2 states. To store the joint probability distribution of N variable, which has in total 2^N possible states, $2^n - 1$ parameters are needed, whereas the last probability can be calculated by subtracting the rest of the probabilities from 1.

if X_1 is independent of the rest of the variables given X_2, term $P(X_1|X_2, X_3, \dots, X_n)$ in the above product form can be simplified as $P(X_1|X_2)$, which requires much fewer parameters to store[3] and is easier to specify by experts or to learn from data. The key to this simplification is to study the real system and identify the conditional dependencies for every variable in the system and use them to construct conditional probability distributions. Once the conditional probability distribution for every variable is specified, they can be used to build a BN.

2.5.2.2 Types of Variables

In BN, most variable are chance nodes that models a probability distribution, with additional classes including deterministic nodes, equation nodes, decision nodes, and value nodes. Deterministic nodes are chance nodes with a zeros-and-one bivariate conditional probability distribution, working similarly to the 'if-then-else' function used in SD. Equation nodes are nodes characterised by analytic equations. Decision nodes are nodes that can be manually set to one of its states to simulate decision makers' decisions as input to the model. Lastly, value nodes are nodes that yield different values given different combinations of their parent nodes.

2.5.3 Data Sources and Utilisation

The building blocks of BN – conditional probability distributions – can be specified from both expert knowledge and empirical data. Expert judgements reflect experts' knowledge and experience in their domains. Expert judgements can be elicited and turned into probability distributions following certain procedure. Quigley and Walls (2021) and Quigley et al. (2008) review the relevant literature and suggest that the extended Stanford Research Institute (SRI) Elicitation Process (Merkhofer, 1987; Spetzler and Stael von Holstein, 1975) is a useful procedure for eliciting expert judgements. It involves 7 stages of working with experts: motivating, structuring, conditioning, encoding, verifying, aggregation, and discretisation. However, due to the subjectivity of expert judgements, various biases could be introduced to the elicitation at each of the stages, such as motivational bias, cognitive bias, and

[3] Assume each random variable has 2 states. To store $P(X_1|X_2)$ we need $(2-1) \times 2^1 = 2$ parameters, whereas to store $P(X_1|X_2, X_3, \dots, X_n)$ we need $(2-1) \times 2^{N-1} = 2^{N-1}$ parameters. For a conditional probability distribution, fewer parent nodes will drastically reduce the number of parameters needed to store it.

methodological bias. Methods to minimise them have been discussed in the literature (Cooke, 1991; Ferrell, 1985; Gosling, 2018; Moore, 1987; O'Hagan et al., 2006; Spetzler and Stael von Holstein, 1975).

Empirical data is another rich source of information for conditional probability distributions. Daly et al. (2011) and Scanagatta et al. (2019) reviewed machine learning / data mining algorithms for inferring BN from data, such as the hill-climber Monte Carlo (HCMC) algorithm and the expectation maximum (EM) algorithm. The inference deals with two objectives: (1) inferring the BN structure, which is to identify the edges in the DAG, and (2) inferring the BN parameters, which is to assign values to the distributions. The complexity of inference grows as the number of variables in the BN increases. A solution to this problem is to use heuristic algorithms to make the inference solvable in reasonable time, but accuracy of the inference becomes lower and how to balance accuracy and efficiency has been a key issue (Daly et al., 2011).

2.5.4 Modelling Process

BNs have been applied to problem solving in MS/OR contexts. The literature has involved several proposed procedures of using the BN to solve operational problems. Sigurdsson et al. (2001) outline a procedure of three stages and seven steps, including (1) problem structuring (identifying variables, identifying network structure, and expressing as statistical variables), (2) instantiation (specifying conditional probabilities), and (3) inference (entering evidence, propagating beliefs, and interpret results). Chen and Pollino (2012) proposed a similar procedure, including examining model purpose and context, reviewing of existing knowledge, specifying model structure and parameters, and model evaluation. This procedure does not involve the decision support stage as in Sigurdsson et al. (2001) because it is more focused on the development of BN. Daly et al. (2011) and Scanagatta et al. (2019) reviewed algorithms to learn BN entirely or partially from observed data. This data-driven method can play an important role in the modelling process where expert knowledge is limited but empirical data is available, which will be reviewed in the next section.

2.6 Learning Bayesian Network from Data

Learning a BN from data is a critical task in fields such as medicine, biology, epidemiology, economics, and social sciences, enabling reasoning under uncertainty. This section synthesises methods for learning BNs from data, focusing on both historically significant algorithms and the latest developments, covering both structure learning and parameter learning.

2.6.1 Structure Learning Algorithms

Structure learning involves determining the DAG that best represents the dependencies in the data, a problem known to be algorithmically NP-hard due to the super-exponential growth in the number of possible DAGs as the number of variables increases. The literature identifies three primary approaches: constraint-based, score-based, and hybrid methods.

2.6.1.1 Constraint-Based Methods

Constraint-based algorithms rely on conditional independence (CI) tests to construct the graph, often returning Markov Equivalence Classes (MECs) as Partially Directed Acyclic Graphs (PDAGs) or Complete PDAGs (CPDAGs). These methods are grounded in the causal Markov and faithfulness assumptions, making them suitable for large datasets but potentially less accurate with small samples or complex dependencies.

The PC algorithm, introduced by Spirtes, Glymour, and Scheines in 1993 in their book "Causation, Prediction, and Search" (Spirtes et al., 2001) is a cornerstone of constraint-based methods. PC algorithm is a global-discovery algorithm, it starts with a complete graph and iteratively removes edges based on CI tests, with a complexity of $O(n^2 2^n)$, where n is the number of vertices in the DAG.

PC algorithm's variants include PC-Stable (Colombo and Maathuis, 2014), which is order-independent and improves accuracy and stability in learning the true graph in high-dimensional settings at the cost of being slightly slower than PC (Kitson et al., 2023). Conservative PC (or CPC) (Ramsey et al., 2006) uses weaker adjacency-faithfulness, offering better arrowhead precision but worse recall. For latent variables, FCI (Spirtes et al., 1995) and its stable version (Colombo and Maathuis, 2014) produce Pattern Graphs (PAGs), with RFCI (Colombo et al., 2012) being around 250 times faster for sparse graphs (Kitson et al., 2023).

Unlike the PC algorithm and its variants, which adopt a global discovery strategy by testing conditional independencies among all variables, local discovery algorithms focus on identifying the dependencies of each target variable and then merging these local structures to produce a global network. These algorithms often rely on discovering the Markov Blanket (MB) or the parents and children (PC) set for each variable. For instance, the Incremental Association Markov Blanket (IAMB) family of algorithms (Tsamardinos et al., 2003) starts by tentatively adding parents and children of a target variable based on statistical tests (e.g., conditional independence tests) and refines the candidate sets until a stable Markov Blanket is identified. Local discovery methods can be especially efficient in high-dimensional settings where a global approach like PC may become prohibitively expensive, as they limit conditional independence testing to subsets of variables relevant to a specific target node. Moreover, local structures can be more interpretable for exploratory analyses in which understanding the immediate influences around a particular variable is crucial (Aliferis et al., 2003; Kitson et al., 2023).

2.6.1.2 Score-Based Methods

Score-based algorithms frame BN structure learning as an optimisation problem. A scoring function quantifies how well a candidate structure explains the observed data, and search techniques seek to maximise (or minimise) that score (Daly et al., 2011; Scanagatta et al., 2019). Historically, some of the earliest score-based approaches include the K2 algorithm (Cooper and Herskovits, 1992) and Hill-Climbing , which used a Bayesian score under simplifying assumptions to greedily add parents to each node given a fixed ordering. Despite its practicality, K2 can be sensitive to variable orderings and tends to get stuck in local optima when the data are noisy, or the ordering is suboptimal.

Later explorations in this direction introduced more robust scoring functions, notably the Bayesian Dirichlet (BD) family – such as the BDe (Heckerman et al., 1995) and BDeu (BDe with a uniform prior) – and frequentist-inspired criteria like the Bayesian Information Criterion (BIC) or Minimum Description Length (MDL). These scores leverage decomposability, meaning that the overall network score sums contributions from each variable and its parent set (Daly et al., 2011). This property allows algorithms to compare and update each node's parents independently, improving computational efficiency.

Score-based structure learning can be carried out by exact or approximate search strategies. Exact methods, typically based on dynamic programming (Koivisto and Sood, 2004; Silander and Myllymaki, 2006), guarantee finding the global optimum but often become infeasible even for moderate numbers of variables due to super-exponential growth of the DAG space. Approximate search strategies – such as greedy hill-climbing, best-first search, or genetic algorithms – often scale better but do not guarantee a global optimum. A popular family of approximate algorithms iterates between making small local changes (adding, removing, or reversing edges) and accepting moves that improve the global score. Greedy Equivalence Search (GES) (Chickering, 2002) refines this principle by orienting edges in a way that moves through Markov Equivalence Classes rather than individual DAGs, often converging to high-scoring networks.

2.6.1.3 Hybrid Methods

Hybrid algorithms combine the advantages of both constraint-based and score-based approaches (Kitson et al., 2023). Their workflow often starts by using conditional independence tests to constrain the structure – typically finding a skeleton or partially directed acyclic graph (PDAG) in a manner akin to PC – and then applies a scoring criterion to refine or orient edges within those constraints (Tsamardinos et al., 2006). By pruning clearly absent edges early on, hybrid approaches can reduce the size of the search space for the subsequent scoring step, leading to efficiency gains in high-dimensional problems.

One prominent example is the Max–Min Hill-Climbing (MMHC) algorithm (Tsamardinos et al., 2006), which proceeds in two phases. First, a constraint-based procedure identifies a skeleton using a max–min heuristic for conditional independence tests; second, a local hill-climbing score-based search orients edges among the remaining candidates. This blend often results in structures that are more accurate than purely constraint-based methods in the presence of limited sample sizes, while still preserving many of the computational benefits that constraint-based filtering provides. Other hybrid techniques employ alternative independence tests or scoring functions, but the core principle remains consistent: use constraints to reduce the search space and then refine with a global scoring step.

2.6.2 Parameter Learning Algorithms

Once the BN structure is determined, the next step is to learn the parameters – namely, the conditional probability distributions associated with each variable and its parents. Parameter learning methods can be broadly divided into two categories: Frequentist (Maximum Likelihood) approaches and Bayesian approaches.

2.6.2.1 Maximum Likelihood Estimation (MLE)

Under the frequentist paradigm, the parameters for discrete BNs are typically learned by counting how often each variable and parent configuration appears in the data (Daly et al., 2011). For instance, if X has parents Π_X, the conditional probability table $\theta_{x|\pi_x}$ is computed by:

$$\theta_{x|\pi_x} = \frac{count(X = x, \Pi_X = \pi_x)}{\sum_{x'} count(X = x', \Pi_X = \pi_x)}$$

where $count(\cdot)$ is the frequency of observing a particular configuration in the dataset. MLE is straightforward and often works well for sufficiently large samples but can be susceptible to overfitting when data are sparse. Smoothing techniques, such as Laplace correction (or more generally, Dirichlet priors), help mitigate zero-frequency issues.

2.6.2.2 Bayesian Parameter Learning

Bayesian methods treat the network parameters themselves as random variables endowed with prior distributions (Koller and Friedman, 2009). A common choice is the Dirichlet prior for discrete variables, leading to closed-form updates of the posterior parameter distributions. This approach naturally incorporates regularisation, as prior distributions can be tuned to reflect domain knowledge or to impose smoothing, thus avoiding overfitting in small-sample regimes (Steck and Jaakkola, 2002). Posterior inference over parameters can be done analytically (e.g., with a conjugate prior) or via Markov Chain Monte Carlo (MCMC) when no closed-form solution exists (Daly et al., 2011).

2.6.2.3 Handling Missing Data

In many real-world applications, not all variables are observed for every data instance. Expectation-Maximisation (EM) (Lauritzen, 1995) is the canonical algorithm used to handle missing data for parameter estimation. EM iterates between computing expected sufficient statistics for the missing entries (E-step) given a current estimate of the parameters and then updating parameters (M-step) using these expected counts as if they were observed data. Convergence is guaranteed to a local maximum of the likelihood. For Bayesian approaches, a variant of EM known as the Expectation-Maximisation algorithm with a Dirichlet prior (a.k.a. Bayesian EM) (Beal and Ghahramani, 2003) incorporates prior information during the M-step.

2.6.2.4 Continuous Variables

Parameter learning for continuous variables typically assumes a parametric family for the local conditional distributions. A common assumption is that continuous nodes follow a Gaussian distribution, leading to Linear Gaussian BNs where each continuous child is modelled via a linear combination of its (continuous) parents plus Gaussian noise (Koller and Friedman, 2009). Parameter updates then involve standard linear regression techniques or maximum likelihood estimates under normal assumptions. Other distributions (e.g., Conditional Linear Gaussian, nonparametric) can be employed when linearity or normality are not appropriate (Daly et al., 2011).

Overall, parameter learning is comparatively more straightforward than structure learning, as it typically involves well-established statistical estimation techniques once the BN structure is fixed. Nevertheless, the choice of parameter learning method – frequentist vs. Bayesian, "best-guess" vs. EM for missing data, Gaussian vs. nonparametric distribution – can substantially impact the learning outcome of a BN's parameters.

2.7 Combining SD and BN

2.7.1 The Mixing Methods Approach

Before delving into the specific ways of combining System Dynamics (SD) and Bayesian Networks (BN), it is useful to acknowledge the broader literature on mixing methods in management science and operational research (MS/OR). Scholars have observed that single methods (e.g., a single simulation approach) may fail to capture the complexity of organisational or social problems, thus motivating a multimethodology perspective that draws from multiple paradigms, methodologies, and techniques (Mingers and Brocklesby, 1997).

In the MS/OR context, multimethodology is often described as a framework in which different "methodologies" (e.g., SD, BN, or discrete-event simulation) can be partitioned and recombined to form a tailored intervention strategy (Mingers and Brocklesby, 1997). This may involve taking one part of a method (e.g., the causal loop diagrams from SD) alongside elements of another (e.g., the probabilistic inference from BN).

Meanwhile, scholars such as Howick and Ackermann (2011) refer to "mixing methods" or "mixed methods" to capture the idea of combining tools, techniques, methods, methodologies, and/or paradigms within a single research or modelling study. In practice, mixing methods can include everything from straightforward "parallel" usage – where two methods run side by side to compare or triangulate their insights – to a more "interactive" or "enrichment" design, where one method's outcomes feed into another. The difference between these designs can be subtle, but it generally centres on how integrated or sequential the methods are (Bennett, 1985; Morgan et al., 2017).

2.7.2 Strategy for Examining Existing Combinations of SD and BN

In the next section, we examine how SD and BN have been combined in the literature. Identifying which combination design is appropriate may hinge on the problem context, data availability, and philosophical considerations about what type of knowledge the study aims to produce (Mingers and Brocklesby, 1997). By aligning with these mixing-methods principles, we can more systematically explore how and why SD and BN can enhance one another in an intervention or modelling study. The following lens and aspects are derived from the mixing methods literature such as Howick and Ackermann (2011) and mixing methods frameworks such as Morgan (2013), and will be applied to study each combination of SD and BN.

1. **Problem context** refers to the characteristics of the problems where this mixing SD and BN design is suitable.
2. **System view** refers to the problem boundary and level of details for each methodology. This aspect often hugely influences the initial selection of methods to mix.
3. **Method dominance** pertains to the level of focus or priority given to each method in the context of the project.
4. **Mixed method design** concerns the collaborative use of the methodologies, specifically the sequence in which they are applied. The existing designs seem to conform well to the categories set out by Morgan et al. (2017), including parallel design, sequential design, interaction design, and enrichment and integration design. While Morgan et al. (2017) examined the integration of SD with Discrete-Event Simulation (DES) and Nguyen (2022) focused on SD with Agent-Based Modelling (ABM), their categorisation frameworks are applicable here due to structural similarities in mixing-methods modelling. Here enrichment and integration are merged following Nguyen (2022) to avoid being stuck in nuance, as the difference between them is often marginal.

We therefore adopt the following four categories: **Parallel, Sequential, Interaction, and Enrichment**.

5. **Technical justification** concerns the suitability of the design to the problem and its technical feasibility.
6. **Benefit** is based on the authors' reasoning for integrating different methods and how this integration yields superior results compared to using a single method.

2.7.3 Searching for Existing Studies

A focused narrative review was undertaken to identify studies combining SD and BN. Searches were last executed on 8 August 2025 in Scopus and Google Scholar using queries to capture the relevant studies (exact queries and search process are documented in Appendix A), no year limits were applied, and results were limited to English. Records were exported and deduplicated. In Scopus, titles and abstract were screened for the presence of both SD and BN; in Google Scholar, screening extended to full-text matches. Potentially relevant records underwent full-text assessment, and studies in which both SD and BN were used were retained. For each retained paper, backward citation checking (reference lists) and forward citation checking (Google Scholar/Scopus "Cited by") were performed; newly discovered records entered the same screening loop until no new eligible items were found. As of 8 August 2025, this process identified 41 relevant studies. A PRISMA-style flow count and the screening results are provided in Appendix A. The approach aims for transparency, comprehensiveness, and reproducibility without claiming exhaustiveness.

2.7.4 Taxonomy of SD-BN Design Patterns

Whilst modest in size, the literature combining SD and BN has grown steadily over the past two decades. To avoid a study-by-study narrative and align with the multimethodology lens in Sections 2.7.1 and 2.7.2, the synthesis below is organised by design patterns – recurrent ways SD and BN are combined in practice. Table 2.1 summarises each pattern in terms of typical purpose, data flow, relationship of system views, and methodological dominance, with representative examples. The sections below provide concise descriptions and pointers to illustrative cases; the full list of included studies by pattern is provided in Appendix A .

Table 2.1 Summary of SD–BN design patterns

Design	Typical purpose	Data flow	System view	Dominance	Examples
Parallel	Compare/triangulate conclusions on the same problem	None (SD ‖ BN)	Shared boundary; similar abstraction	Balanced	Sušnik et al. (2013);
	SD and BN each models a different input to a later phase	BN & SD → third sector	Different aspects of the same problem	Balanced	Chikkagoudar et al. (2022); Yan and Wood (2017)
Sequential	BN quantifies parameters/inputs for SD scenarios	BN → SD	Partially overlapping views	Often balanced	McNaught (2003); Bertone et al. (2018);
	SD generates synthetic data to learn /parameterise/update a BN	SD → BN	Shared boundary; different emphasis	BN is often the practical focus	Wang et al. (2016); Crookes (2017); Niu et al. (2020);
	Use SD artefacts to structure/parameterise a BN (e.g., CLD→BN nodes/links)	SD → BN	BN is the system view; SD informs BN structure	BN-dominant	Rodriguez-Ulloa (2018); Farmani and Savic (2008);
Interactive	Exchange state/parameter info during simulation	SD ↔ BN (time-stepped)	Complementary, connected sub-systems	Balanced	Punyamurthula and Badurdeen (2018), Mohaghegh et al, (2009);
Enrichment	Embed BN to represent probabilistic/nonlinear relations inside an SD structure	internal (within SD)	SD is the system view; BN refines sub-processes	SD-dominant	Hafezi et al. (2017, 2021); Tan et al. (2021); Bertone et al. (2016); Liu et al. (2024)
	Estimate SD parameters/equations with data	(data → SD via Bayesian update)	SD is the system view; Bayesian machinery supports inference	SD-dominant	Osgood and Liu (2015): parameters; Zhou et al. (2020): equations

2.7.4.1 Parallel Comparison

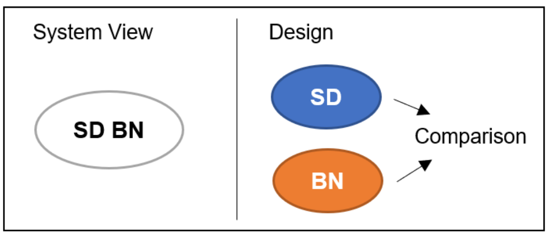


Figure 2.7 Parallel comparison

Problem context: Two independent models (SD and BN) are built for the same problem using comparable inputs to compare conclusions (e.g., Sušnik et al. (2013)).

System view: Models share the same boundary and a similar level of abstraction, and they address the same research question.

Dominance: The two methods have balanced dominance – neither method is privileged in modelling effort or interpretation.

Design: The two models are used parallelly with no data exchange; the results are compared after both methods are implemented.

Technical justification: This design provides a structured way to assess robustness and method sensitivity when different epistemic lenses are applied to identical evidence.

Benefit: Triangulation reveals convergence and divergence, clarifies each method's strengths and limitations, and strengthens confidence in policy conclusions.

2.7.4.2 Parallel Inputs to a Subsequent Stage

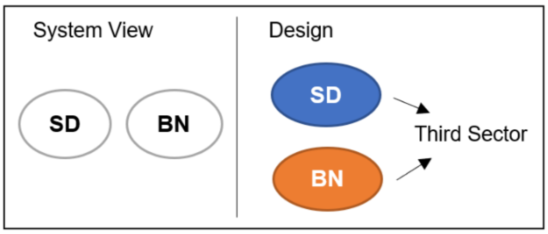


Figure 2.8 Parallel inputs to a later stage

Problem context: Here SD and BN are developed independently to generate complementary inputs that are combined in a subsequent stage such as scenario generation (e.g., Chikkagoudar et al. (2022); Yan & Wood (2017)).

System view: Each model captures a different aspect of the same problem, so the views are complementary rather than overlapping. However, their views may not be collectively exhaustive, as the third sector could still have an additional view.

Dominance: The methods are treated even-handedly because each provides a distinct input required downstream.

Design: The design is parallel followed by integration in a third sector where the separate outputs are synthesised for decision making.

Technical justification: The separation allows each method to operate where it is strongest and avoids forcing premature integration.

Benefit: The subsequent synthesis benefits from richer and more transparent evidence, with the provenance of each input remaining clear.

2.7.4.3 Sequential BN to SD

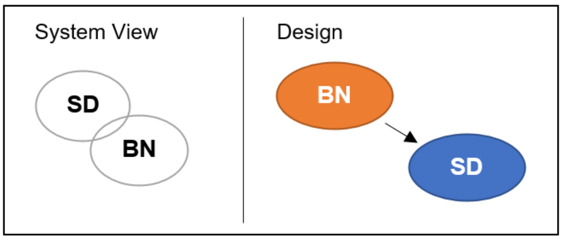


Figure 2.9 Sequential BN to SD

Problem context: A BN is elicited or learned to quantify uncertain parameters or scenario inputs that then drive an SD model (e.g., McNaught (2003); Bertone et al. (2018)).

System view: The views are partially overlapping: the BN concentrates on uncertainty in determinants while the SD model captures the dynamic consequences of those determinants.

Dominance: Dominance is typically balanced because the BN resolves parameter uncertainty, and the SD model conducts the dynamic exploration.

Design: The design is sequential and one-way, with information flowing from BN to SD and no feedback in the opposite direction.

Technical justification: The BN integrates expert knowledge and sparse data into probabilistic inputs that the SD model can propagate through feedback-rich dynamics.

Benefit: The resulting scenario analysis is explicitly uncertainty-aware and provides policy testing that is grounded in structured probabilistic inputs.

2.7.4.4 Sequential SD to BN via Synthetic Data

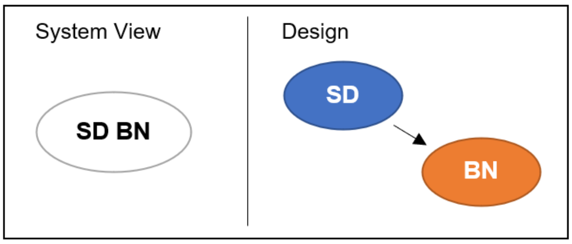


Figure 2.10 Sequential SD to BN via synthetic data

Problem context: When empirical data are insufficient to fit a BN, an SD model is used to simulate the system and produce synthetic datasets for learning or parameterising the BN (e.g., Wang et al. (2016); see also Crookes (2017); Niu et al. (2020)).

System view: The two models share the same boundary but emphasise different aspects: SD represents accumulations and flows, whereas BN represents probabilistic dependencies and diagnosis.

Dominance: In practice BN is often the focal product, while SD serves as the generator of the necessary training data.

Design: The design is sequential and one-way, with information flowing from SD to BN; once the BN is trained, the SD model may no longer be required.

Technical justification: Structural validation of the SD model provides credibility for the simulated data, allowing the BN to be learned despite limited observations.

Benefit: The approach enables probabilistic inference and diagnosis while preserving dynamic realism in the synthetic data used for training.

2.7.4.5 Sequential SD to BN via SD artefacts

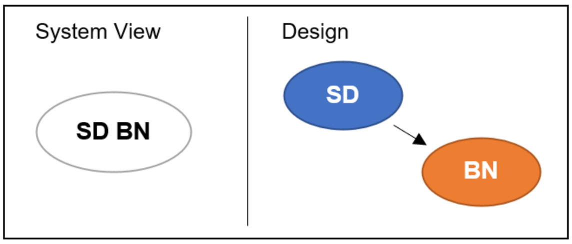


Figure 2.11 Sequential SD to BN via SD artefacts

Problem context: SD artefacts – most commonly causal loop diagrams – are developed and translated into BN structure to support probabilistic reasoning (e.g., Rodriguez-Ulloa (2018); Farmani & Savić, 2008).

System view: The two models share the same boundary but emphasise different aspects: BN provides the overall system view, while SD contributes causal hypotheses that shape BN nodes, arcs, and prior distributions.

Dominance: The BN is dominant because the SD artefacts function primarily as structuring devices rather than as an executable simulation.

Design: The design is sequential knowledge transformation in which SD representations are mapped to a BN that is then used for learning or inference.

Technical justification: CLDs encode directional causal beliefs that can be formalised as BN structure, enabling uncertainty propagation and evidence updating. During the transformation, feedback loops in CLDs are removed by adding nodes with temporal meaning to the BN, so that the BN remains acyclic. No reviewed study transforms CLDs into Dynamic Bayesian Networks (DBNs).

Benefit: The method retains the convenience of modelling and narrative causal insight of SD while gaining BN's rigorous treatment of uncertainty and data.

2.7.4.6 Interactive Coupling

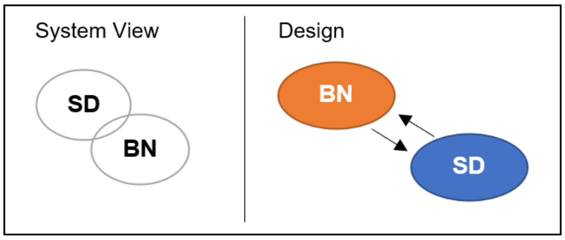


Figure 2.12 Interactive coupling

Problem context: Some applications require time-evolving feedback dynamics and probabilistic risk reasoning simultaneously, motivating two-way coupling between SD and BN (e.g., Punyamurthula and Badurdeen (2018); Mohaghegh et al. (2009)).

System view: The models represent complementary but connected sub-systems, such as process flows in SD and risk states in BN, with a shared interface for exchange.

Dominance: Dominance is generally balanced because both models are essential to model the problem and demand substantial effort.

Design: The design is interactive and bidirectional: at each time step SD state variables supply inputs to the BN as evidence, and BN posteriors update SD parameters for the next step. Figure 2.13 (Fig. 3 in Punyamurthula and Badurdeen (2018)) provides a visualisation of one such design.

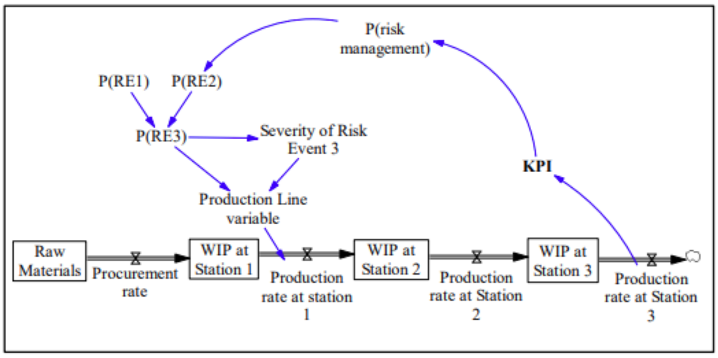


Figure 2.13 Interaction between BN (upper) and SD (lower) models (Punyamurthula and Badurdeen, 2018 fig. 3)

Technical justification: SD provides feedback and accumulation, whereas BN offers belief updating and uncertainty propagation; together they yield time-consistent probabilistic dynamics.

Benefit: The coupled system captures the co-evolution of processes and risks with uncertainty carried through time.

2.7.4.7 Enrichment: BN embedded within SD

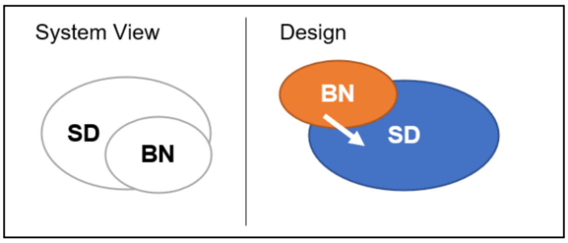


Figure 2.14 BN embedded within SD as Enrichment

Problem context: Certain non-linear relationships within an SD model are best represented probabilistically to summarise the general trend before translated into algebraic or Boolean equations. A BN is built to model the dependencies between some variables, then transformed and embedded as a part of the SD model's structure (e.g., Hafezi et al., (2017, 2021); Tan et al., (2021); Bertone et al., (2016); Liu et al., (2024)).

System view: The SD model retains the overall system view, whereas the BN adds a probabilistic and non-linear lens to parts of the system not well represented within SD.

Dominance: The SD model is dominant because it governs the system's dynamics while the BN augments local structure of the SD.

Design: The design is tight integration within SD, where the BN is transformed into algebraic or Boolean equations. This distinguishes it from the BN to SD sequential design, where BN retains its CPT-based form and only sends parameter values to the SD model.

Technical justification: BNs conveniently encode expert knowledge and uncertainty that are inconvenient to express directly as deterministic equations.

Benefit: The enrichment improves local realism and allows uncertainty propagation without sacrificing SD's feedback structure.

2.7.4.8 Enrichment: Bayesian update within SD

This design is not visualised using a diagram as Bayesian update is used as a model calibration tool instead of system modelling tool.

Problem context: Parameter values or equation forms in an SD model are uncertain and can be constrained by data, so Bayesian update is used for parameter estimation or equation-form selection (e.g., Osgood & Liu (2015); Zhou et al. (2020)).

System view: The SD model provides the comprehensive system view, while the Bayesian machinery supplies an inference layer over parameters or structures.

Dominance: The SD model remains dominant because the goal is a better-calibrated SD model rather than a stand-alone BN.

Design: Priors over parameters or equation forms are updated with observed data to yield posterior estimates that inform subsequent modelling and simulations.

Technical justification: Bayesian update offers a coherent probabilistic framework for combining empirical evidence with prior knowledge and for quantifying residual uncertainty.

Benefit: The approach reduces parameter and equation-form uncertainty, provides credible intervals, and improves the explanatory and validity of the SD model.

2.7.5 Discussion of Existing SD-BN Combination

Across the reviewed studies, SD-BN combination clusters into eight recurrent design patterns:

1. Parallel comparison (SD and BN on the same problem),
2. Parallel inputs to a subsequent stage (SD and BN to a third sector),
3. Sequential BN to SD (BN generates inputs for SD),
4. Sequential SD to BN via synthetic data,
5. Sequential SD to BN via SD artefacts,
6. Interactive coupling of SD and BN,
7. Enrichment with BN embedded within SD, and
8. Enrichment via Bayesian update within SD to estimate parameters or equations.

In these combinations, more evidence is observed on sequential and enrichment designs, whereas explicit parallel and interactive designs are less common. Collectively, these patterns show that SD and BN have been combined to (a) triangulate insights or provide complementary inputs, (b) move information one-way to address parameter or data limitations, (c) exchange information during simulation, (d) embed Bayesian inference within SD to represent uncertainty or calibrate models, or (e) ease the development of one model by using the other model as an intermediate.

Although the enrichment designs documented here predominantly embed BN components within an SD model, the converse – embedding SD-type deterministic relations within a BN – is also feasible and used in practice. Contemporary BN software (e.g., GeNIe) supports equation (or deterministic) nodes (see Section 2.5.2.2), in which a node's value is computed from analytic/algebraic expressions rather than a conditional probability table, analogous to auxiliary or flow equations in SD. While no study in our review explicitly reported such SD-into-BN embedding under the inclusion criteria, the capability is supported by common tooling. Conceptually, this corresponds to a BN-dominant enrichment: the BN provides the system view, and selected substructures are specified deterministically using SD-style formulations.

Table 2.2 summarises the circumstances in which each SD-BN design pattern is most appropriate. It serves as a quick, literature-based reference that complements the detailed accounts in Sections 2.7.4.1 – 2.7.4.8. The table is not exhaustive, and further work is required to develop more comprehensive guidance on selecting and combining SD and BN designs.

Table 2.2 Mapping of Designs to Their Advantages and Risks

Pattern	When to use
Parallel comparison	User needs triangulation or a method comparison on the same problem with comparable inputs.
Parallel inputs to a later stage	User needs complementary outputs from different aspects of the problem that will be combined in a downstream analysis or decision process.
Sequential BN to SD	User needs to quantify uncertain parameters/scenarios for SD using expert knowledge or sparse data.
Sequential SD to BN via synthetic data	BN learning is data-limited but an SD structure can credibly simulate the needed observations.
Sequential SD to BN via SD artefacts	CLDs/sector maps can structure a BN when running an SD simulation is unnecessary.
Interactive coupling	User requires real-time exchange between dynamic states and probabilistic inferences during simulation.
Enrichment: BN embedded within SD	Some local relationships are best expressed probabilistically inside an SD model.
Enrichment: Bayesian update within SD	Parameters or equation forms in SD need calibration or selection using data within a coherent probabilistic framework.

Among the existing studies, two – McNaught (2003) and Bertone et al. (2018) – are closely aligned with the use explored in this thesis because they employ expert-elicited BNs to model dependencies among SD parameters and to generate scenario inputs. However, as highlighted in Sections 2.5 and 2.6, one of the key strengths of BNs – their ability to be learned directly from empirical data – has received little attention within the SD literature. Building on the existing use of BNs for modelling parameter dependencies, this thesis takes a different direction by investigating how such BNs can be data-driven rather than relying exclusively on expert knowledge, thereby enabling the capture of empirically grounded parameter relationships.

The following section summarises the gaps identified in the literature and outlines the methodological contribution proposed by this thesis.

2.8 Summary of Gaps

The previous sections have demonstrated that System Dynamics (SD), despite its success in modelling feedback-rich problems, often relies on oversimplified assumptions about parameter dependencies for uncertainty analysis.

Copulas have been identified as a parametric approach that can be used to explore dependence structures between variables and therefore may be useful for assessing the impact of parameter dependence on uncertainty assessment of SD models.

Although Bayesian Networks (BNs) have been used to model SD models parameters through expert elicitation, their potential as a data-driven tool for modelling a wide range of dependence structures remains largely unexplored in the SD literature.

These observations echo the gap outlined in Section 1.2:

- Lack of a data-driven method to model multi-parameter dependencies in System Dynamics (SD)

  In System Dynamics (SD), dependencies between parameters can affect the assessment of uncertainty in model KPIs. When such dependencies are ignored or oversimplified, the resulting combinations of parameter values can lead to distorted KPI projections and inaccuracies in uncertainty assessment. Despite this risk, most existing SD studies rely on theoretical or uniform parameter distributions (Eker et al., 2014; Ford and Flynn, 2005; Jadun et al., 2017), with limited attention given to adequately capturing multi-parameter dependencies. Some efforts – such as Bertone et al. (Bertone et al., 2018) – have used expert-elicited Bayesian Networks (BNs) to represent dependencies. However, such approaches depend heavily on expert knowledge, and a data-driven method for learning parameter dependencies directly from empirical data remains largely absent in the SD literature. This gap is especially critical in contexts where expert knowledge is scarce, but data is available, underscoring the need for a data-driven approach to modelling multi-parameter dependencies in SD.

Bringing together these themes, the methodological need becomes clear: a method to model parameter interdependence in System Dynamics models. The literature suggests that Bayesian Networks can potentially fill this gap.

To demonstrate the need to consider the influence of parameter dependence structures on the outputs of a SD model, in Chapter 4 , we use a copula-based process. This motivates the use of BN to model the dependence more adequately, thereby reducing the risk of distorting KPI uncertainty assessment and improving the overall credibility of model analysis.

Before proceeding to Chapter 4 , the next chapter will lay the methodological foundation for this thesis. It first establishes the philosophical stance and research methods underpinning the study, then introduces the overall conceptual workflow. The chapter also presents an illustrative Susceptible-Infected-Recovered (SIR) System Dynamics model, along with the copula- and BN-based techniques used to examine and model parameter dependencies, which together form the basis for the experiments and analyses described in the subsequent chapters.

Chapter 3 Methodology and Conceptual Approach

This chapter outlines the methodology guiding this thesis. It introduces the research philosophy, methods, and conceptual workflow, along with the illustrative compartmental System Dynamics model and the copula- and BN-based techniques used to analyse parameter dependencies.

3.1 Research Philosophy and Methods

3.1.1 Research Philosophy

In this thesis, we adopt Critical Realism (CR) as our overarching philosophical stance, recognising both the objective dimension of real-world phenomena and the subjective dimension of stakeholder perceptions and modeller judgments. Originally developed by Roy Bhaskar (1997), CR posits that an external reality exists independently of our conceptions, yet our knowledge of it is inevitably theory-laden and fallible. These ideas have been further advanced within management science and operational research contexts by scholars such as John Mingers (2006), who argues that CR usefully bridges the gap between purely positivist approaches (where social phenomena are treated as objective facts akin to physical laws) and purely interpretivist approaches (where reality is seen as wholly constructed by human beliefs).

Central to Critical Realism are (1) the distinction between the empirical (what is observed), the actual (events occurring whether or not observed), and the real (the deeper causal mechanisms underlying observed phenomena), and (2) the dynamic process by which agents continuously update their beliefs about the real through perceiving the empirical. In an SD modelling context, this translates into a process wherein we test competing model structures and parameterisations against observational data, acknowledging that people's opinions or domain knowledge are important sources of insight but must be evaluated for plausibility through empirical scrutiny. For example, different hypothesised feedback loops or parameter values may initially derive from stakeholder beliefs or prior studies; once implemented in simulation, however, they can be checked against real-world data to distinguish more plausible from less plausible system representations.

Our choice of CR aligns well with System Dynamics in that SD itself frequently emphasises an iterative process of building, testing, and refining models based on both empirical evidence and stakeholder understanding. In addition, our proposed BN-based approach for capturing parameter interdependence is fundamentally a statistical treatment of data and, as such, is less explicitly tied to any single philosophical paradigm. If a researcher wishes to embed BN-driven parameter dependencies within a different philosophical framework – e.g., a conventional positivist or interpretivist stance – the data-driven process of BN learning would still apply. Thus, while we position this work within a Critical Realist context, the methodology outlined here does not, in principle, exclude other philosophical choice made by researchers who seek a more robust approach to modelling multi-parameter uncertainty.

3.1.2 Research Methods

To meet our objective and address the gap identified in Chapter 1 and Chapter 2 , we employ four main research methods:

1. **Literature Review**

   We began by systematically reviewing existing scholarly work on System Dynamics, Bayesian Networks, mixing methods (including mixing SD with BN), and copulas. This review situates our work, surfaces the need for a data-driven BN approach, and reveals specific shortcomings in simpler correlation modelling. The review clarifies the gap in the literature around how to incorporate parameter dependence into SD modelling.

2. **Secondary Data Analysis**

   While our illustrative example (to be outlined later in this chapter) is not heavily reliant on large external datasets, the general BN approach presupposes the ability to glean parameter relationships from either domain knowledge or real-world observations. Where possible, we collate relevant data from, published data sources, documented studies, or prior analyses to learn Bayesian Network. In a real application (e.g., epidemiological data on infection rates), one would rely more heavily on primary data sets and possibly expert elicitation as well to specify conditional dependencies.

3. **Simulation**

   The core of our System Dynamics demonstration involves simulating an SIR model, constructed around stocks of susceptible, infected, and immune populations. Simulation experiments enable us to observe how different parameter sets (uniform vs. copula-based vs. BN-driven) affect key outcome metrics (i.e., KPIs) such as peak infection or total infections. By comparing results across these metrics' distributions, we illustrate how ignoring or simplifying dependencies can mislead decision makers about the outbreak's dynamics.

4. **Stochastic Analysis**

   Monte Carlo experiments are employed to conduct stochastic analysis, enabling the quantification, visualisation, and comparison of uncertainty in SD model KPI projections. This includes drawing samples from parameter distributions modelled using copulas and Bayesian Networks. Each random-draw strategy yields a distribution of simulation outputs, which can be visualised and compared to highlight the differences in KPI uncertainties under different assumptions of parameter correlation and motivate modelling parameter dependencies adequately.

Taken together, these research methods align with an overarching Critical Realist mindset: we combine domain knowledge, empirical data, and iterative model simulation to refine our representation of complex phenomena. The next sections detail how these methods unify into the proposed methodological approach.

3.2 Overall Approach

Figure 3.1 (below) illustrates the workflow adopted in this study to assess the importance of parameter dependencies and reduce uncertainties through BN-informed SD parameter priors. The intermediate step – Motivational Copula Experiments – is not strictly required when the sole objective is to model parameter dependencies using a BN. In such cases, the workflow can proceed directly along path "A", bypassing this step. However, when it is important to demonstrate the impact of parameter dependencies – for example, to persuade stakeholders of the need to collect additional parameter data – the copula experiment can be incorporated by following path "B".

Figure 3.1 Flowchart of the proposed approach

The flow begins with Problem Articulation, where the SD model's purpose, scope, and intended outputs (e.g., key performance indicators or KPIs) are clarified. Next, Parameter Identification and Screening ensure we focus on those parameters whose uncertainty may significantly affect the model's behavioural modes or KPIs. Following path "B", we then carry out a set of Experiments, using copula-based parameter distributions to show that for our chosen SD model, different parameter dependence structures can be consistent with the same correlation measure but yield different KPI uncertainty envelopes, motivating that parameter dependence needs to be more adequately modelled to avoid distorting uncertainty assessment of KPIs. Data-Driven

BN Construction therefore follows, in which a Bayesian Network that captures observed correlations between the SD model parameters is learned from an empirical dataset. This BN then Generates Correlated Priors, feeding into the SD model's Monte Carlo Simulation, enabling us to Assess the Results, which are the distributions of model KPIs. We close this workflow with a Refinement & Feedback loop to revisit any stage if the analysis suggests a new piece of data or a refined BN definition is needed.

This chart serves as an overall roadmap to follow in this thesis, but can also be potentially adapted to similar contexts, for example, other modelling and simulation methods.

3.3 Illustrative Example: Outline

To demonstrate the proposed approach in a concrete setting, we employ a simplified SIR (Susceptible – Infected – Recovered) SD model commonly used in epidemic studies as an illustrative example. This model serves strictly as **proof of concept** to demonstrate the methodology. It is not intended to capture the full complexity of real-world disease dynamics, nor should the results or insights derived from its analysis be interpreted as having direct policy relevance.

The choice of this model is motivated by three considerations. First, the SIR model is canonical in SD education, which lowers the barrier to understanding for readers. Second, the model exhibits sufficient feedback structure and involves multiple interacting parameters, making it suitable for illustrating how assumptions about parameter distributions and their joint dependence influence simulation outcomes. Third, the SIR model benefits from the availability of rich cross-country data collected during the COVID-19 pandemic, reducing the practical effort required for assembling empirical datasets.

Although the empirical dataset is derived from COVID-19 observations, whose transmission dynamics are more accurately represented by SEIR (Susceptible – Exposed – Infected – Recovered) type models with an incubation stage, the present study applies these data to a simplified SIR structure. This introduces a structural mismatch at the compartment level. In this work, however, the empirical data primarily describe contextual and policy-related parameters rather than compartment-level disease transition statistics and are used to learn cross-context parameter dependencies rather than to calibrate mechanistic epidemic dynamics. These higher-level parameter relationships are not tied to the presence or absence of an explicit exposed compartment and are therefore expected to transfer qualitatively across closely related epidemic model structures. The SIR model is thus used here as a transparent proof-of-concept vehicle for demonstrating the BN-based dependency-learning workflow, rather than as a structurally faithful representation of COVID-19 transmission dynamics. In applications where structural epidemiological fidelity is required, the same BN procedure could be coupled with an SEIR or more detailed compartmental model instead.

The subsections below provide an overview of the SIR model (Section 3.3.1), define the key uncertainty metrics and parameters of interest (Section 3.3.1.2), explain how copulas are used to evaluate the sensitivity of SD model outcomes to variations in dependence (Section 3.3.3), and how BN is used to empirically model dependencies (Section 3.3.4).

3.3.1 System Dynamics SIR Model

3.3.1.1 Model Structure

The SIR model tracks three main stocks: *Susceptible, Infected,* and *Immune*. Individuals begin in the susceptible population, can become infected after sufficient risky contacts, and eventually move into immune status through recovery or vaccination. The model is implemented with daily time steps and includes a simple feedback mechanism: the more infected individuals there are, the higher total contacts become, thus elevating risk to susceptible individuals. Figure 3.2 shows the stock-and-flow diagram of the model. Equations and parameter values are documented following the figure.

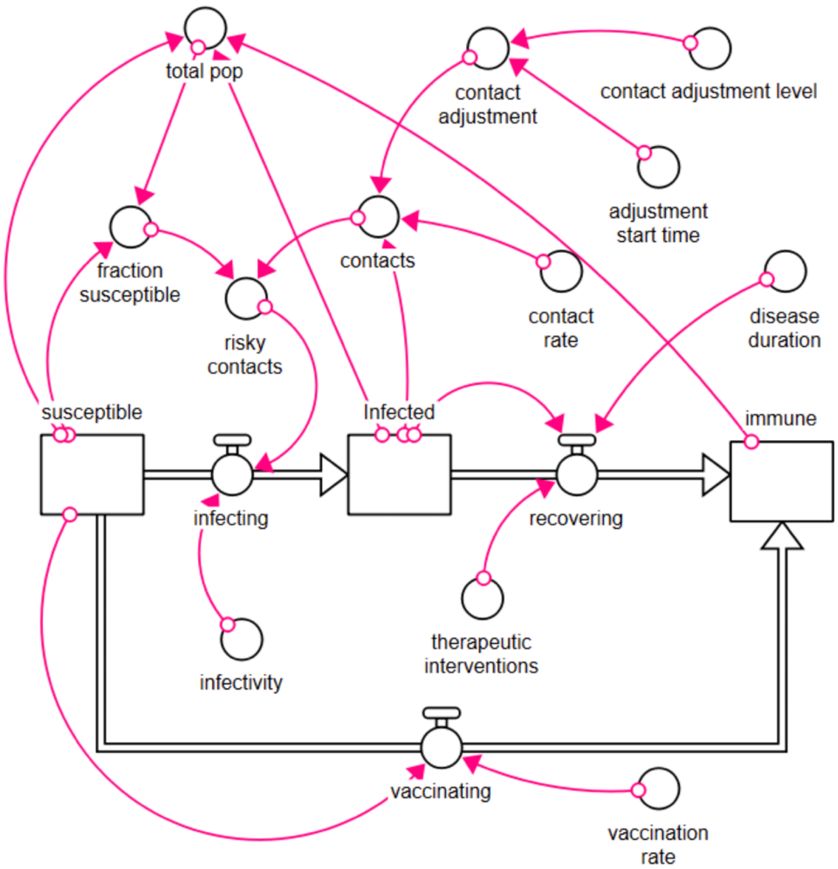


Figure 3.2 Stock-and-flow diagram of the SIR SD model

Equations in the model are defined as follows:

Stocks:

$$\text{infected}(t) = \text{infected}(t-\Delta t) + [\text{infecting} - \text{recovering}] \times \Delta t,$$

$$\text{immune}(t) = \text{immune}(t-\Delta t) + [\text{recovering} + \text{vaccinating}] \times \Delta t,$$

$$\text{susceptible}(t) = \text{susceptible}(t-\Delta t) - [\text{infecting} + \text{vaccinating}] \times \Delta t.$$

Initial values:

$$\text{infected} = 1,$$

$$\text{immune} = 0,$$

$$\text{susceptible} = 1{,}000{,}000.$$

Flows:

$$\text{infecting} = \text{risky\_contacts} \times \text{infectivity},$$

$$\text{recovering} = \frac{\text{Infected}}{\text{disease\_duration}} \times \text{therapeutic\_interventions},$$

$$\text{vaccinating} = \text{susceptible} \times \text{vaccination\_rate}.$$

Converters/Auxiliaries:

$$\text{contacts} = \text{Infected} \times \text{contact\_rate} \times \text{contact\_adjustment},$$

$$\text{risky\_contacts} = \text{contacts} \times \frac{\text{susceptible}}{\text{total\_pop}},$$

$$\text{contact\_adjustment} = \begin{cases} \text{contact\_adjustment\_level}, & \text{if TIME} > \text{adjustment\_start\_time}, \\ 1, & \text{otherwise} \end{cases}$$

$$\text{total\_pop} = \text{susceptible} + \text{infected} + \text{immune}$$

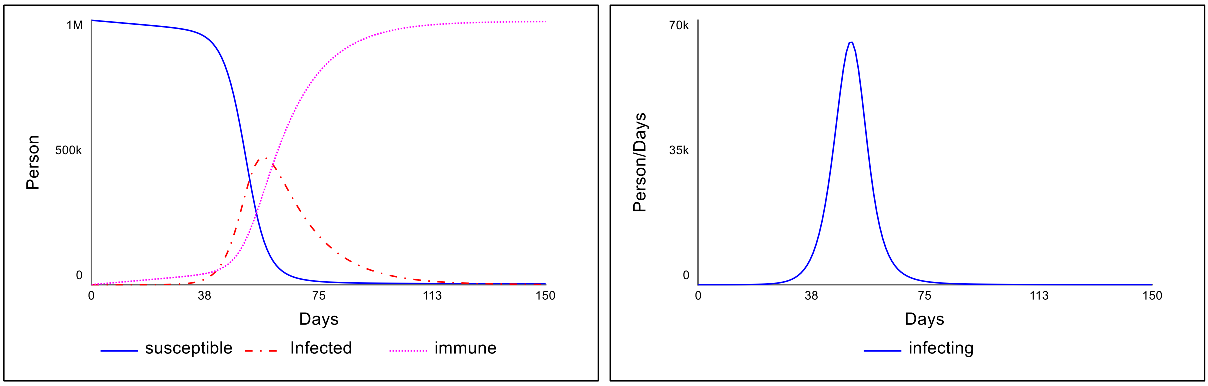
Parameters:

Table 3.1 List of parameters in the SIR model

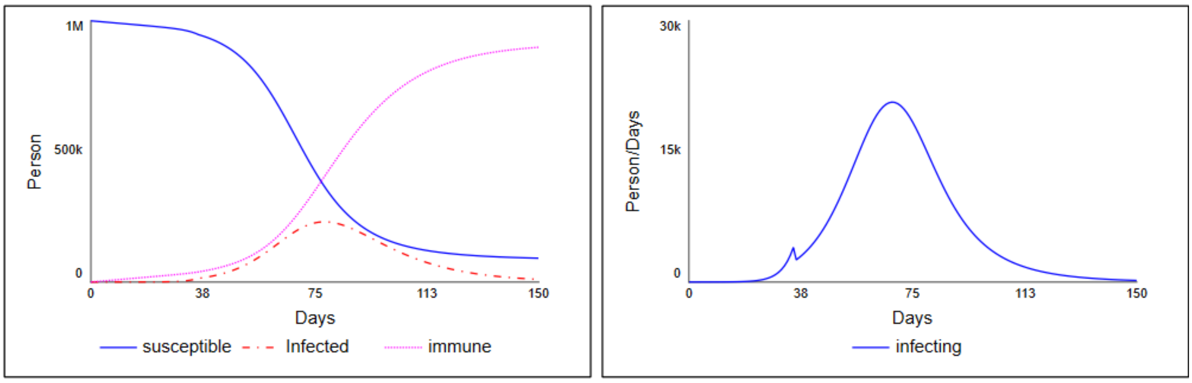
Parameter	Definition	Default Value	Unit
Contact rate	Average daily contacts made by an infected individual	85	Touch/Person/Day
Disease duration	Mean infectious period	14	Days
Infectivity	Infection probability per contact	0.0045	Person/Touch
Contact adjustment level	Strength of intervention on social contacts – 0 is no intervention; 1 is no contact	0	Dimensionless
Therapeutic interventions	Factor that modifies the recovery rate	1	Dimensionless
Vaccination rate	Fraction of susceptible population vaccinated per day	0.001	1/Day
Adjustment start time	Day at which contact reduction (*contact adjustment level*) is triggered	35	Days

3.3.1.2 Model Behaviour

Figure 3.3 presents the simulated behaviour of the model using in two example scenarios. In scenario (a), all parameters take their default values, whereas in scenario (b) an intervention is introduced on day 35 that halves the contact rate. The left panels illustrate the trajectories of the three key stock variables: susceptible, infected, and immune individuals; the right panels display the infection flow rate over time.



(a)



(b)

Figure 3.3 Behaviour of key stocks and flows under two scenarios: (a) without intervention (b) with an intervention that halves contact rate on day 35

In scenario (a), as expected in an SIR model, the number of susceptible individuals decreases over time in an S-shaped curve, while the immune population increases with a similarly shaped rise. The infected population and the infecting rate both follow a rise-and-fall pattern, peaking mid-simulation before declining as immunity builds in the population. In scenario (b), a small dip in infecting rate is noticeable around day 35, where it temporarily declines before rising again. This pattern corresponds to the contact adjustment intervention introduced on day 35, which temporarily reduces the infection rate. A visual comparison of the growth gradients of the infected in the two scenarios after day 35 suggests that the intervention helped slow the spread and flatten the peak of infections.

It is important to note that all model parameters, including total population size, are assumed and not calibrated to any specific real-world population. These two simulation runs are intended to illustrate model dynamics for demonstration purposes.

3.3.2 Selected Parameters and Uncertainty Metrics

For illustrative purposes, we do not exclude any parameter at this stage. All seven parameters of the SIR model are assumed to potentially influence the model's behavioural modes or KPI outcomes and are therefore included in the subsequent copula experiments and BN modelling steps. In real-world applications, however, it is often advisable to apply a parameter screening technique – such as that proposed by Ford and Flynn (2005) – to identify the most influential parameters. The selection criteria for this process should be determined collaboratively between the modeller and relevant stakeholders, allowing later steps to focus on the parameters that matter most and thereby reducing modelling effort without sacrificing insight.

Because our ultimate interest is in assessment of the uncertainty in the model's key performance indicators (KPIs) and how parameter distributions shape their distributions, we define a few high-level metrics to quantify the uncertainty in the model's behaviour:

1. **Max Infecting**: The maximum new infections/day across the simulation run.
2. **Time of Max Infecting**: The day on which max new infection occurs.
3. **Max Infected**: The maximum existing number of infections across the simulation run.
4. **Time of Max Infected**: The day on which max infected occurs.

These metrics reflect common epidemiological concerns in SIR models – i.e., how large the outbreak might become, how quickly it peaks, and how many people are currently infected. In a real application, one could also track final immune or mortality rates.

3.3.3 Exploring Dependence Structures Using Copulas

Given the selection of influential parameters – all seven in our SIR SD model – we now proceed to the next step: Motivational Copula Experiments (see Figure 3.1). The purpose of this step is to demonstrate how the structure of dependencies among parameters can influence the uncertainty envelope of KPIs. This, in turn, motivates the adoption a more adequate and flexible method to model parameter dependencies, such as using a Bayesian Network (BN).

We use four different copulas – Gaussian and selected Archimedean families (e.g., Frank, Clayton, Gumbel) – to impose various correlation structures between parameters, while preserving their marginal distributions (*uniform* in our case). To control but also explore the effect of correlation strength, we repeat this process at different levels of correlation (measured using the *Spearman rank correlation coefficient*). The resulting joint distributions are used to sample parameter sets that reflect both the imposed dependency structure and the specified correlation level. These samples are then used in Monte Carlo simulations, generating model behaviours from which KPIs are extracted. The associated uncertainties are assessed using *normalised standard deviation* and compared across different copulas and correlation levels. The detailed procedure of carrying out the copula experiments will be presented in Chapter 4 .

While copulas allow us to examine how different parameter dependency structures affect KPI uncertainties, we note that these families typically involve only one or two tuning parameters, resulting in relatively simple dependence structures. Additionally, standard Archimedean copulas are not well-suited for capturing negative correlations in high dimensions. As such, copulas alone are insufficient for fully capturing the complexity of real-world parameter dependencies. However, their ability to represent a range of correlation structures makes them valuable for showcasing the influence of parameter dependence structure on SD model KPI uncertainties, motivating the need of a more flexible and empirically grounded methods for modelling dependencies more adequately.

3.3.4 Modelling Parameter Dependencies Using Bayesian Networks

Following the previous step, we assume – for demonstration purposes – that the influence of parameter dependencies on KPI uncertainty of the chosen SIR SD model is significant enough to warrant further investigation. This assumption is based on the differences observed in KPI uncertainty envelopes across copulas at the same correlation level, which we assume to be sufficient to prompt stakeholders to conclude that the dependencies between parameters should be modelled more adequately.

Within this context, we present a data-driven approach in which a BN is learned from empirical data gathered across multiple observations – in our case, outbreaks of different virus strains in various global locations. The learned BN structure is validated against domain knowledge, as literature from a relevant field suggests the combination of data and domain knowledge can outperform pure data-driven methods in learning a BN (Abdulkareem et al., 2019). The detailed procedure of data gathering and BN learning will be presented in Chapter 5 . The learned BN captures conditional dependencies between model parameters and is then used to generate parameter samples through forward sampling. These samples reflect the joint distribution of parameters as inferred from observed data.

The sampled parameter sets are input into Monte Carlo simulations of the SD model, producing output behaviours from which KPIs are extracted. The resulting KPI uncertainty distributions are visualised and analysed to support interpretation and decision-making.

3.4 Chapter Summary

This chapter stated the philosophical foundation of this thesis, and the research methods employed to achieve its objective. It then introduced the overall approach as a step-by-step workflow and outlined an illustrative demonstration, using an SIR SD model as an example. It also introduced the copula-based method to explore parameter dependence structures and the BN-based method to empirically model parameter dependence. The subsequent chapters will detail how the approach is implemented in practice.

Chapter 4 Exploring Dependence Structures with Copula

Building upon the outlined conceptual approach, this chapter and the next describe the implementation in detail and present the analyses. They show how structure of dependence of SD model parameters influences KPI uncertainty (Chapter 4), and how an empirical BN can be used to model the dependence more adequately (Chapter 5).

4.1 Specification of Copulas

To illustrate the influence of the structure of parameter dependence on the SD model's KPI uncertainty, we first need to define a variety of dependence structures. Copulas offer a parametric way to couple each parameter's marginal distribution into a joint distribution with controllable correlation. Here, we use four copulas, i.e., Gaussian, Frank, Clayton, and Gumbel, and systematically sample across a set of Spearman rank correlation levels to produce a range from near-zero (as in the case of uniform distribution or independent variables) to moderately and then highly correlated samples. Table 4.1 summaries the 44 copulas definitions (4 families × 11 correlation levels per family) we populate in this section.

Table 4.1 Definitions of copulas representing a variety of dependence structures

Copula Family	Levels of Spearman Rank Correlation (ρ)				
Gaussian	0.0	0.1	…	0.9	1.0
Frank	0.0	0.1	…	0.9	1.0
Clayton	0.0	0.1	…	0.9	1.0
Gumbel	0.0	0.1	…	0.9	1.0

We are not separating uniform/independent sampling from copula sampling as a standalone strategy, because a copula can be configured to yield uncorrelated samples that effectively mimic independent uniform distributions (via correlation matrix = 0 for Gaussian copula, or θ close to 0 or 1 for Archimedean copulas). Thus, we do not run a separate "uniform" scenario because our lowest correlation setting for each copula family approximates that.

Directly defining an Archimedean copula at a specific correlation level can be challenging, as the copula parameter θ does not correspond directly to the Spearman correlation coefficient. For the three Archimedean copula families (Frank, Gumbel, and Clayton), we adopt an iterative empirical approach using binary search to identify appropriate θ values. For each copula family and for each target correlation level ρ (ranging from 0.0 to 1.0 in increments of 0.1), we search for θ within a bounded range $[\theta_{min}, \theta_{max}]$, respecting family-specific constraints (e.g., $\theta \geq 1$ for Gumbel, $\theta > 0$ for Clayton). At each iteration, we sample 3,000 observations from the candidate copula and compute the empirical Spearman correlation. The search terminates early if the absolute difference between empirical and target correlation falls below 0.02 (ideal threshold), or after a maximum of 30 iterations. We accept θ values where the final difference remains within 0.2 (acceptable threshold); otherwise, the candidate copula is rejected. For the Gaussian copula, the correlation parameter is set directly to the target value, as its correlation matrix structure allows explicit specification of dependence strength.

Applying this search strategy to the 33 Archimedean copula combinations (3 families × 11 correlation levels), we successfully identified acceptable θ values for all cases with no rejections. Among these, 16 combinations achieved ideal precision (|difference| $\leq$ 0.02), while the remaining 17 met the acceptable threshold.

We omit extensive negative correlation exploration for the high-dimensional scenario (7 dimensions in our case), as standard Archimedean families (Frank, Gumbel, Clayton) do not naturally handle partial negative correlation across multiple dimensions. A preliminary 2D test confirmed that negative correlation can indeed alter variance in single-pair settings, but extending that approach (e.g., reverse one of the two columns for a negative θ) grows complicated or infeasible with more than two parameters. Since the goal of copula experiments is to show that dependence structure can influence KPI uncertainty even at the same level of correlation measure, we are satisfied with only positively correlated samples.

4.2 Monte Carlo Experiment

This section describes the Monte Carlo simulation procedure used in the copulas-based experiments, including the sampling of parameters for the SIR SD model, the execution of multiple simulation runs, and the subsequent calculation of KPI uncertainties. By comparing the resulting KPI distributions under different parameter-sampling schemes, particularly those with varying dependence structures at the same correlation level, we can assess how the structure of parameter dependencies influences uncertainty in model outputs.

4.2.1 Sampling from Copulas

Since we are interested in all seven parameters of the SIR SD model (as discussed in Section 3.3.2), we use each of the 44 copulas definitions specified in Section 4.1 (see Table 4.1 for summary) to generate random samples $(u_1, \ldots, u_d) \in [0,1]^d$, where $d = 7$. Here, a *sample* refers to a 7-dimensional vector, where each element u_i corresponds to one model parameter. Because these values are still normalised within the $[0,1]$ interval, they must be mapped to the actual range $[a_i, b_i], i \in [1,7]$ of each parameter in order to be used as inputs in the SIR SD model.

Although each parameter in the SIR model may have a different plausible range, we adopt a uniform scaling approach for demonstration purposes. Acknowledging that parameters in the SIR SD model can vary substantially across regions – due to differences in virus strain, geography, culture, and policy responses – we set the range for each parameter to [20%, 180%] of its original value (i.e., $\pm 80\%$,). This simplification keeps the process transparent and tractable for illustrative analysis. In real-world applications, parameter ranges could be defined more precisely based on expert judgement or stakeholder input, without requiring substantial additional effort, since the primary goal of the copula experiments is to raise awareness of parameter dependencies.

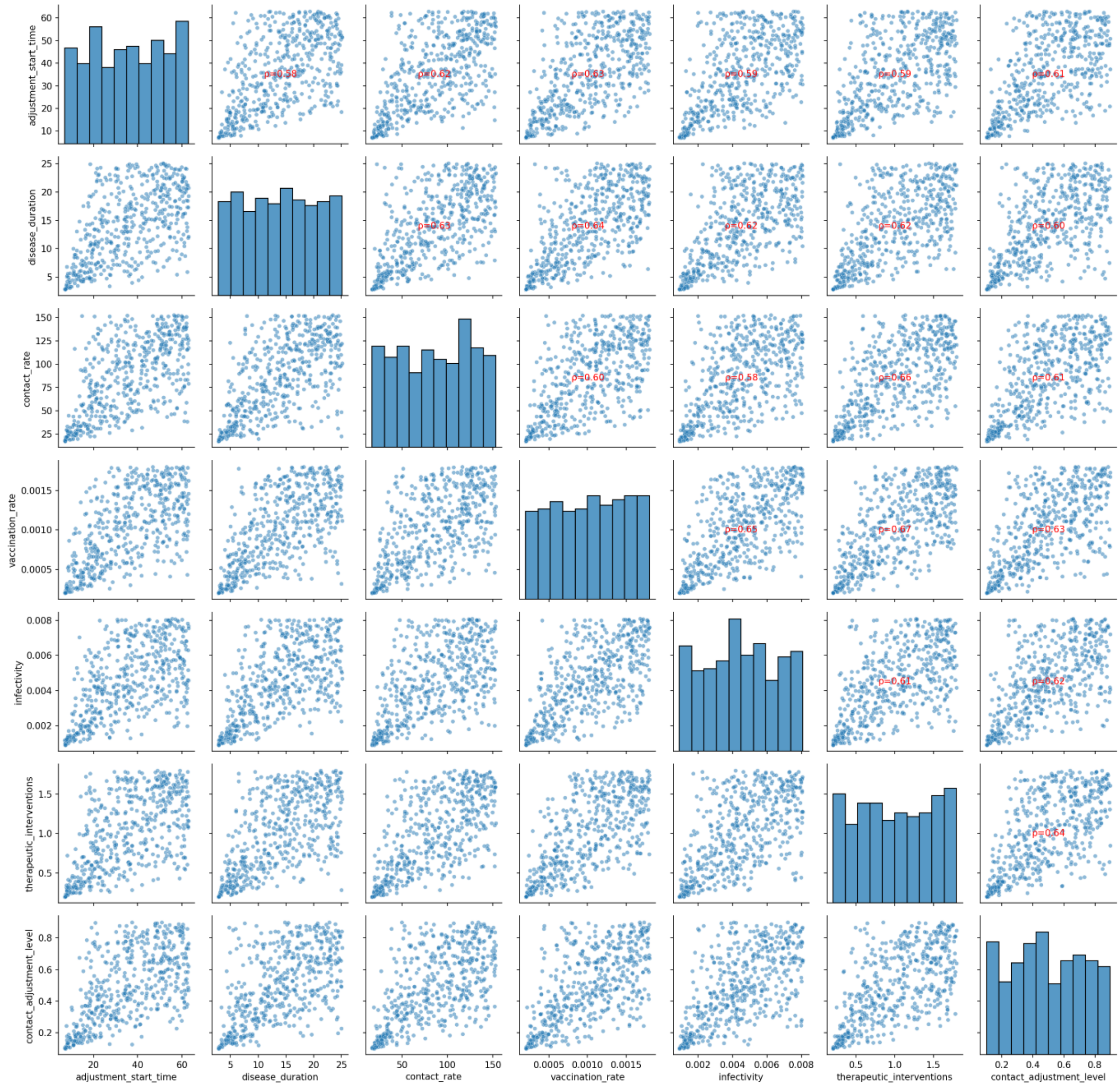


Figure 4.1 Samples drawn using Clayton copula with pairwise correlation at 0.6

For each of the 44 copula definitions, we draw 500 samples, resulting in a total of 22,000 simulations. This sample size is sufficient to visualise the imposed dependence structures – e.g., via pairwise parameter plots (see Figure 4.1) – while keeping the computational burden of Monte Carlo simulation manageable, as it scales linearly with the number of samples. Hence, by covering a range of Spearman's rank correlation coefficient from 0.0 to 1.0 (in increments of 0.1) for four copula families, we obtain a rich set of parameter samples to run Monte Carlo simulations of the SIR SD model.

4.2.2 Monte Carlo Simulation

In each iteration of the Monte Carlo simulation, we use a sample parameter vector to instantiate the SIR SD model and perform a time-stepped simulation (e.g., daily steps until a specified end date). This yields a complete time series of variable trajectories, including, for example, the daily new infections, the total infected, and the daily vaccination rate.

4.2.3 Colleting KPIs and Measuring Uncertainties

After each run, we retrieve the time-series for variables of interest (i.e., Infected and Infecting) and determine the KPIs, as specified in Section 3.3.2, including:

1. **Peak Infected**

   The maximum infected value, useful for estimating number of bed needed in the hospital for the peak of infection wave.

2. **Time to Peak in Infected**

   Index in the time series at which infected is at its max, useful for estimating how quick the additional beds should be added.

3. **Peak Infecting**

   The maximum infecting value, useful for estimating service capacity needed for outpatient and emergency departments in hospitals.

4. **Time to Peak in Infecting**

   Index in the time series at which infecting is at its max, useful for estimating how quickly the corresponding services should be scaled up.

We record the four KPI values from each simulation run, resulting in a total of 22,000 KPI sets. For each of the 44 combinations of correlation level and copula family, we calculate the standard deviation of each KPI across the 500 runs as a measure of uncertainty.

4.3 Influence of Parameter Dependence Structure

This section presents the results of the copula experiments, showing how different structures of parameter dependence affect the KPI uncertainties.

4.3.1 Monte Carlo Outcomes Across Correlation Levels

The uncertainties associated with each of the four KPIs are visualised and compared across different correlation levels and copula families (Figure 4.2, Figure 4.3, Figure 4.4, and Figure 4.5). Instead of presenting the raw uncertainty values (i.e., standard deviations), we normalise the results by using the standard deviation at the zero-correlation level as baseline. The differences from this baseline are then expressed as percentages to facilitate easier comparison across settings. Additionally, boxplots are included for each correlation level and copula family to examine the skewness and distributional shape of the KPI values.

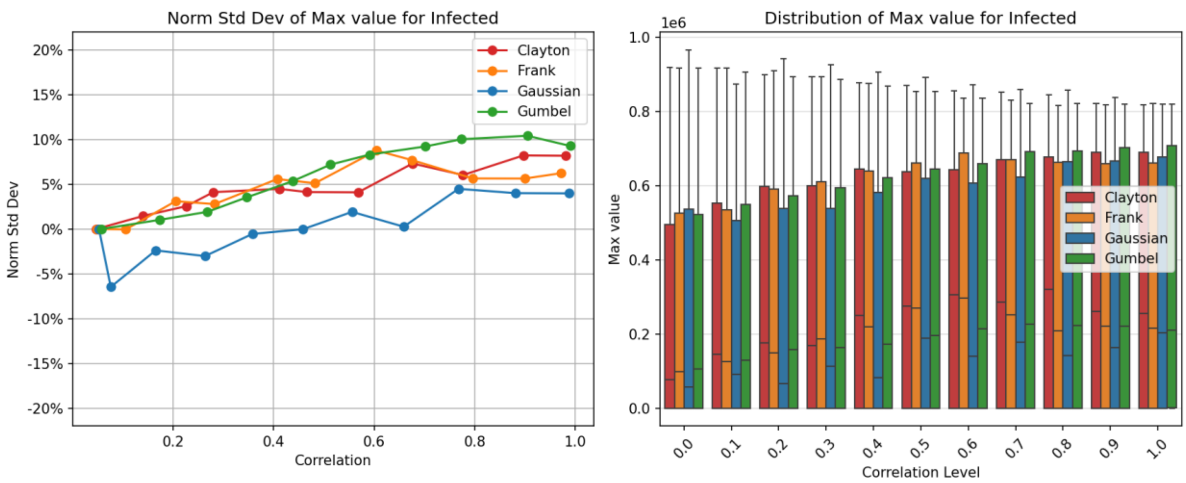


Figure 4.2 KPI uncertainty at different correlation levels by copula, max infected

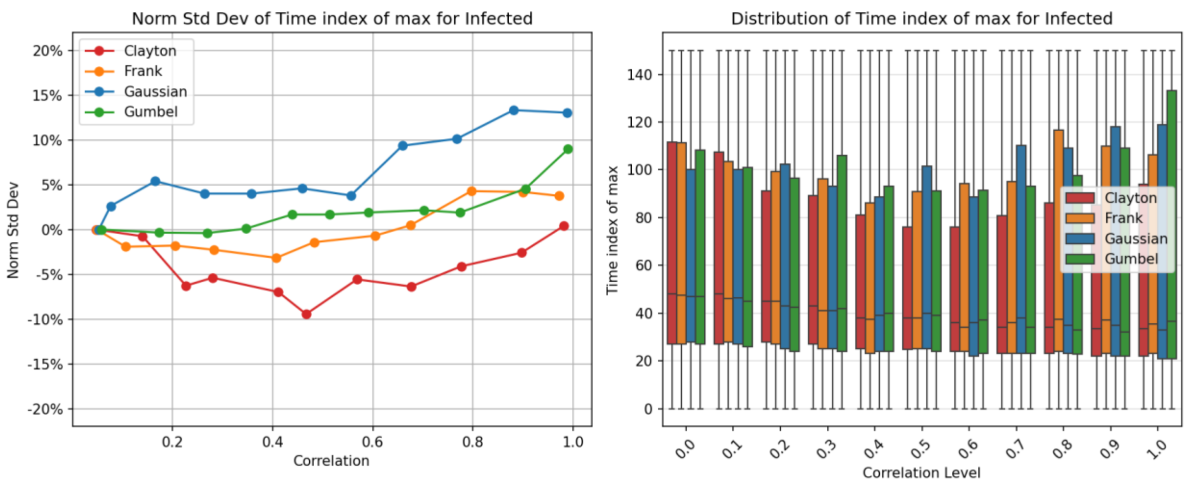


Figure 4.3 KPI uncertainty at different correlation levels by copula, time of max infected

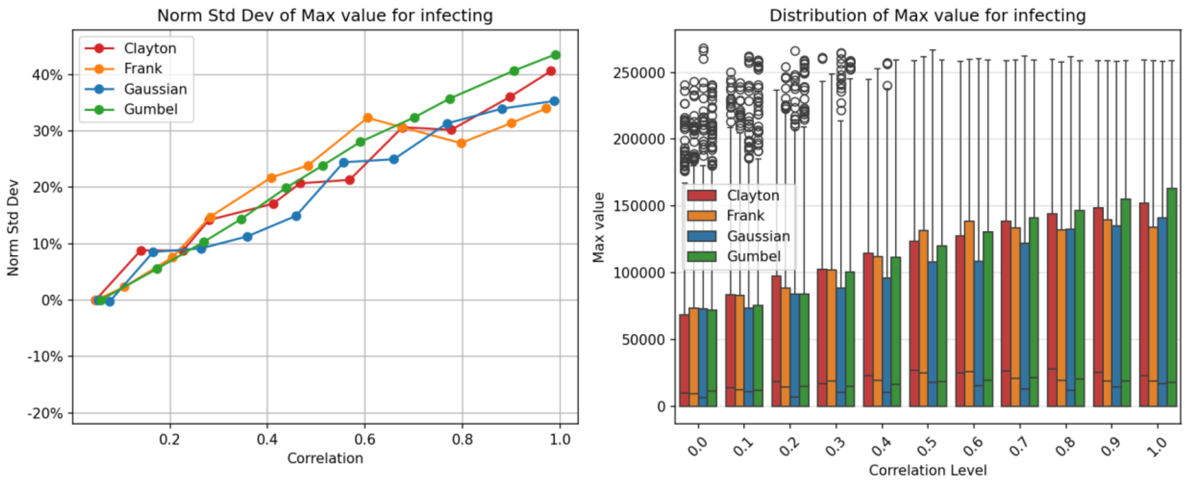


Figure 4.4 KPI uncertainty at different correlation levels by copula, max infecting

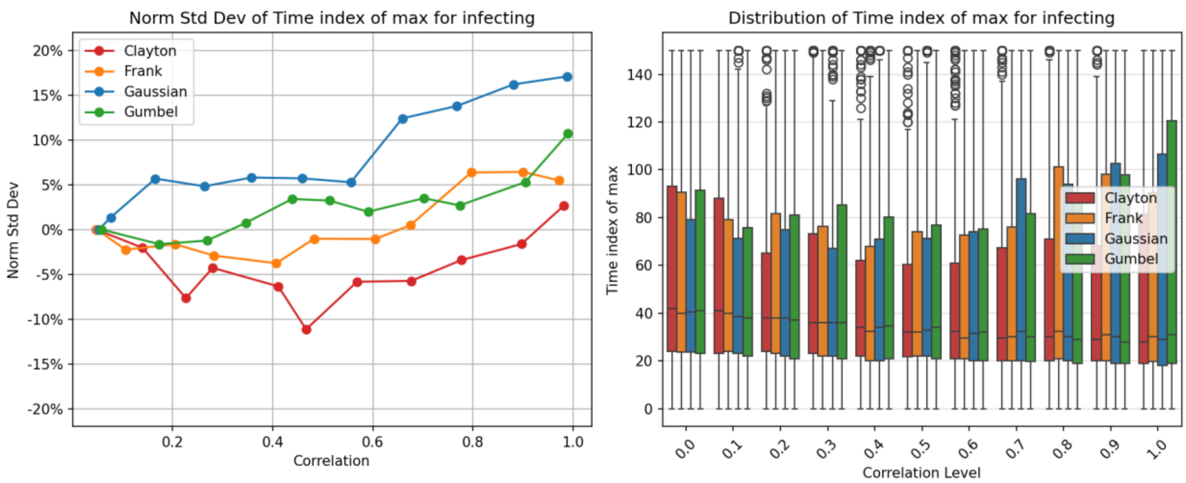


Figure 4.5 KPI uncertainty at different correlation levels, time of max infecting

4.3.2 Observations and Interpretation

The difference in output variance across correlation levels observed in Figure 4.2 to Figure 4.5 shows the influence of parameter dependence on KPI uncertainties of the SIR model. The four figures above show that distributions for all KPIs are affected by correlation level and copula type to different extents, mostly between 5% and 15% with extreme cases up to around 40%. For reference, when comparing two samples of size 500 using a two-tailed F-test at the 5% level, the critical ratio of standard deviations required for a statistically significant difference is approximately 1.09, which is about 9% above/below the baseline. In light of this, differences in Figure 4.2 to Figure 4.5 smaller than 9% would therefore not be statistically distinguishable, while some of the larger differences observed (above 10%) lie above this detection threshold.

The practical significant of such differences, however, depends critically on the decision context. In exploratory modelling or long-range scenario analysis, a 10-15% change in the width of an uncertainty envelope may have little impact on qualitative insights or policy assessment. By contrast, in threshold-sensitive settings such as hospital bed planning, vaccine stockpiling, or triggering non-pharmaceutical interventions during a pandemic, a 10-15% shift in peak magnitude or peak timing uncertainty could materially alter capacity buffers or decision thresholds. The same numerical effect size can therefore be negligible or consequential depending on the operational stakes and the tolerance for risk. In the present illustrative model, the bounds of practical importance remain unclear: while some differences are statistically detectable, this study does not establish whether they would systematically alter policy conclusions or resource decisions in applied SD settings.

Among the four KPIs, *Max infecting* (Figure 4.4) is more sensitive to correlation level and less to copula type, while the other three show more sensitivity to copula type, which means even at a similar level of correlation, different copulas (i.e., different dependence structures) can make the KPI distribution different. For instance, Figure 4.5 shows that at the correlation level of 0.8, Gaussian copula yields a wider spread *time of max infecting* distribution than Clayton (+15% baseline vs. -3% baseline).

These observation supports the conclusion that the structure of parameter dependencies *can* influence the distribution of certain KPIs in this SIR SD model. Even when parameters are correlated to the same degree, different dependency structures result in noticeably different uncertainty envelopes. This suggests that for this specific model, relying on uninformed or oversimplified assumptions about parameter dependence *can* distort the assessment of KPI uncertainty. Consequently, it motivates the need to incorporate additional information to model dependencies more adequately.

4.4 Chapter Summary

In this chapter, we presented copula-based experiments to demonstrate that for correlated SD model parameters for the example model, different dependence structures can yield the same correlation measure but result in different KPI distributions, thus having an impact on uncertainty assessment. This sets the stage for pursuing a data-driven approach to modelling parameter dependence – namely, a Bayesian Network – which we introduced in Chapter 2 . In the next chapter, we demonstrate the procedure for learning a BN of parameters from empirical data and using it to assess KPI uncertainties.

Chapter 5 Modelling Parameter Dependence with Bayesian Network

The previous chapter established the role of parameter dependency, motivating the adoption of a more flexible method to model such dependency. In this chapter, we demonstrate how a BN-based approach captures nuanced, real-world relationships among parameters from data and illustrate it through the case study using the SIR SD model.

5.1 Data Acquisition and Preparation for Learning BN

This section describes the empirical data collection and pre-processing for the seven parameters in our SD model: *adjustment start time, contact adjustment level, disease duration, infectivity, contact rate, vaccination rate, and therapeutic interventions*.

The dataset comprises COVID-19 infection waves and corresponding government interventions across 42 countries, spanning from early 2020 to the end of 2021. For each country, we identified three distinct data points associated with the dominant variants during that period – namely, the original (wild-type), Alpha, and Delta variants – resulting in a total of 126 observations or data points. By leveraging cross-country and cross-variant data, we aim to uncover generalisable correlations among system dynamics (SD) model parameters. The following subsections detail the rationale, data sources, and specific transformations applied to each parameter.

Because several parameters are not directly observable at the country level, this study relies on proxies and simplified transformations. These choices are made to ensure cross-country comparability and tractability for a proof-of-concept demonstration, but they inevitably introduce modelling assumptions that may influence the learned dependencies. Throughout this section, we therefore highlight the rationale for each proxy, the potential direction of bias it may introduce, and the implications for interpreting the resulting BN.

The public data sources we used to create our dataset include:

- Oxford COVID-19 Government Response Tracker (OxCGRT) ("COVID-19 Government Response Tracker | Blavatnik School of Government," 2020; Hale et al., 2021) for tracking government response across countries.
- China National Centre for Bioinformation (CNCB) (Zhao and Song, 2020) for identifying dominant variant at different points in time.

- Urbanization by Sovereign State (“Urbanization by sovereign state,” 2025) for estimating base contact rate for each country.
- Our World in Data (Mathieu et al., 2020) for estimating vaccination rate for each country.
- OECD Hospital Beds (Organisation for Economic Co-operation and Development, 2024) for number of beds per 1000 inhabitants as a proxy for therapeutic intervention.

For each country, we extracted key dates (first reported COVID-19 case, date of tightening policies, etc.), policy indices (OxCGRT’s composite stringency index), and epidemiological and medical indicators (vaccine uptake rates, therapeutic capabilities, etc.). For simplicity, we used a variant’s known global dominance time window for all countries and averaged indices within this time window to remove time factor and reflect typical national characteristics for individual countries.

This aggregation suppresses within-wave temporal dynamics and treats each country-variant combination as a representative static context. As a result, parameters that may in reality vary over the course of a wave are approximated by their average levels. This simplification risks masking systematic within-wave changes in policy intensity or epidemiological conditions and may bias the learned dependencies toward long-run national characteristics rather than short-term adaptive responses. Future applications intended for policy analysis should consider more granular temporal resolution or explicitly model time-varying parameters.

5.1.1 Adjustment Start Time and Contact Adjustment Level

These two parameters reflect how quickly and how intensively a country implements non-pharmaceutical interventions (NPIs), such as school closures or travel restrictions, after the first confirmed COVID-19 case. We approximate this via a composite policy index from the Oxford COVID-19 Government Response Tracker (OxCGRT) (“COVID-19 Government Response Tracker | Blavatnik School of Government,” 2020; Hale et al., 2021) that sums eight key measures:

1. School closing (0–3)
2. Workplace closing (0–3)
3. Cancel public event (0–2)

4. Restrictions on gatherings (0–4)
5. Close public transport (0–2)
6. Stay at home requirements (0–3)
7. Restrictions on internal movement (0–2)
8. International travel controls (0–4)

The *adjustment start time* is defined as the number of days from the first confirmed case in a country until the composite index first reaches or exceeds 5. If the index was already ≥ 5 before the first confirmed case (e.g., due to pre-emptive measures), we set *adjustment start time* to 1. The threshold of 5 was selected as a pragmatic indicator of substantive intervention activity, corresponding to the activation of at least two major NPIs, and chosen to provide a consistent rule across countries to determine an *adjustment start time*. This assumption may bias the adjustment start time toward earlier detection in countries that adopt multiple mild measures rather than a single stringent one. If this threshold correlates poorly with actual behaviour change, the BN may learn attenuated or misaligned timing effects. Future applications should assess sensitivity to alternative thresholds or employ mobility-based indicators that more directly reflect contact reduction.

It is noted that for waves of infections other than the initial one, it is difficult to identify the time from which the adjustment starts because the adjustment (NPIs) often does not stop but remains in place with variation in tightness across the entire COVID-19 period, neither does the variation show very clear correlation to the infection data. We therefore duplicate the value of *adjustment start time* of the same country in the initial wave for the later waves, to best reflect the characteristics of that country in response to an outbreak. This assumption treats responsiveness as a stable national characteristic and enforces persistence of policy timing across successive waves. While appropriate for capturing broad response styles in a proof-of-concept setting, it risks overstating cross-wave similarity and suppressing learning effects or adaptive behaviour of countries. If responsiveness changed substantially over time, the BN may learn weaker correlations between responsiveness and, for example, infectivity.

We map the OxCGRT-based stringency score to a bounded *contact adjustment level* in [0.1, 0.9] to express intervention effectiveness as a percentage reduction factor that can be applied directly in the SD equations. This improves interpretability and keeps the parameterisation mathematically straightforward. However, the mapping introduces two sources of uncertainty: first, the choice of bounds (e.g., whether the upper bound should be 0.9 rather than 0.8 or 0.7) lacks strong empirical support in this study; second, applying a common mapping across countries implicitly assumes that a given policy score translates into a similar level of behavioural change everywhere, which may impose forced homogeneity across heterogeneous social and institutional contexts. If these assumptions are violated, the BN may learn distorted or spurious dependencies involving the intervention parameter. Future applications should calibrate the mapping using mobility, contact-survey, or compliance indicators, potentially allowing country- or region-specific scaling.

5.1.2 Disease Duration and Infectivity

Because it is challenging to gather real-time, accurate nation-level data for average infection length (*disease duration*) or *infectivity*, we rely on dominant variant analyses. Following an approach similar to Alhamlan and al-Qahtani (2025) and Stepanova et al. (2022), we identify periods in which a variant (original/wild-type, Alpha, Delta) is globally prevalent using sources including China National Centre for Bioinformation (CNCB) (Zhao and Song, 2020) and assign typical epidemiological traits:

1. Original (wild-type): *disease duration*≈14 days, infectivity≈3.
2. Alpha: *disease duration*≈11 days, infectivity≈4.
3. Delta: *disease duration*≈8 days, infectivity≈6.

These are representative numbers reflecting shorter disease duration but higher infectivity as the virus evolves. The number assigned to *infectivity* is our indicated level of infectivity and will later be translated into an array of probabilities [0.001, 0.004, 0.009] by which a close contact between a susceptible person and an infected person turns into a new instance of infection. To translate the ordinal infectivity levels into a numeric transmission probability usable by the SIR SD equations, we apply a min–max scaling that maps the lowest and highest infectivity categories in the dataset to a bounded interval $[0.001, 0.009]$. This interval is chosen for technical consistency with the parameter-range design used in the copula experiments (Chapter 4): the default per-contact transmission probability in the illustrative SIR model is 0.0045 (Table 3.1 in Section 3.3.1.1, rounded to 0.005 for implementation), and applying the same ±80% scaling yields $[0.005 \times 0.2, 0.005 \times 1.8] = [0.001, 0.009]$. The resulting mapping preserves the ordinal ranking across variants while keeping parameter magnitudes within a numerically stable range for simulation. Values are rounded to three decimal places for consistency with the discretised BN categories. This construction is intended as a pragmatic parameterisation for proof-of-concept; future applications should replace it with a calibrated mapping based on empirical transmission estimates or sensitivity analysis over alternative probability ranges.

The time window of each global dominant variants is as based on the date range in which the variant dominated weekly global new cases. The original type is taken as dominant from the beginning to December 22 of 2020, Alpha is taken as dominant from December 23 of 2020 to May 31 of 2021, and Delta from June 1 of 2021 to December 19 of 2021.

5.1.3 Contact Rate

Contact rate represents the average daily interactions each infected individual has. In practice, a purely epidemiological measure is difficult to glean from real-world data. We selected urbanisation rate ("Urbanization by sovereign state," 2025) as a coarse, globally available proxy, positing that a higher proportion of urban population correlates with denser and more frequent interactions, providing a single scalar compatible with the SD parametrisation. Although detailed country-specific contact matrices are available for most countries (e.g., Prem et al., 2021), these matrices are themselves largely synthetic projections derived from a small number of empirical surveys and reweighted using national demographic structures. Incorporating such matrices would therefore introduce an additional modelling layer based on demographic assumptions rather than directly observed contact behaviour.

Moreover, the Prem et al. (2021) matrices are high-dimensional and stratified by age and setting, whereas the present SD model requires a single effective contact parameter aligned with time-averaged intervention conditions. Collapsing matrices into a scalar would require further discretionary choices regarding age weighting, location mixing, and averaging of such mixing over the course of intervention, each of which could influence the learned dependencies. The present proxy therefore favours transparency and cross-country comparability over epidemiological preciseness.

However, this choice biases contact rate toward demographic density rather than structured mixing and may introduce confounding with development-related parameters, such as *therapeutic interventions*. Future applications should consider deriving scalar contact indicators from synthetic matrices or validating simpler proxies against them, particularly in models intended to inform policy.

5.1.4 Vaccination Rate

Vaccine uptake was zero or negligible during the earliest (wild-type) stage, so for data points involving the original virus, we set *vaccination rate*=0. By contrast, for Alpha (Dec 23, 2020–May 31, 2021) and Delta (Jun 1–Dec 19, 2021), we use daily average vaccinations per million from Our World in Data (Mathieu et al., 2020) over that variant's time window.

5.1.5 Therapeutic Interventions

This parameter indicates healthcare capacity. Ferrara et al. (2022) who founded that per-capita number of hospital beds significantly predicted the COVID-19 death rate, indicating it is closely related to the capability of delivering therapeutic interventions. We used number of hospital beds per thousand inhabitants, from OECD Hospital Beds (Organisation for Economic Co-operation and Development, 2024), as a proxy for how effectively a country can treat severe cases, then scaled it to a ratio between itself and the average level of all countries. A higher ratio indicates more capable medical infrastructure, hence a larger factor for accelerating recovery. However, hospital bed density captures baseline capacity but does not reflect ICU availability, staffing levels, treatment quality, or surge capability. If treatment effectiveness varies independently of bed density, the learned dependencies may misrepresent clinical effects. Future applications should incorporate richer healthcare capacity indicators where available.

5.1.6 Summary of Data Preparation

Piecing together the collected and pre-processed data, we created a dataset including 126 data points for the seven parameters; the first few are shown in Table 5.1. Taken together, these preprocessing choices reflect deliberate trade-offs between epidemiological fidelity, cross-country comparability, and modelling tractability. The BN therefore learns dependencies among *constructed contextual descriptors* defined through discretised categories and proxy variables rather than among directly observed biological parameters. While the categorical structure of BN captures relative ordering and co-occurrence patterns robustly, the subsequent numeric encoding used for SD simulation can influence the resulting KPI distributions.

Accordingly, the learned structures are intended to demonstrate methodological feasibility rather than to deliver calibrated epidemiological inference and should not be interpreted as causal or policy-valid without further domain-specific validation and sensitivity analysis. Applications intended to inform decision making would require stronger justification of proxies, exploration of alternative constructions, and explicit assessment of robustness.

Table 5.1 First 5 out of 126 rows of the parameter dataset

Country name	Adjustment start time	Contact adjustment level	Disease duration	Contact rate	Vaccination rate	Infectivity	Therapeutic interventions
Austria	18	0.741	14	59.5	0	0.001	1.58
Belgium	38	0.763	14	98.2	0	0.001	1.222
Brazil	9	0.649	14	87.8	0	0.001	0.496
Bulgaria	6	0.627	14	76.7	0	0.001	1.646
Canada	51	0.718	14	81.9	0	0.001	0.554
…	…	…	…	…	…	…	…

5.2 Learning BN Structure and Parameter

This section explains how we build and learn a Bayesian Network (BN) from the data of 7 SD model parameters. We adopt the PC algorithm for structure learning, run in the GeNIe Academic 5.0 software, and discretise parameters to produce a purely discrete BN.

5.2.1 Algorithm Choice: PC

A variety of approaches can learn a BN structure from data, broadly classified into constraint-based or score-based algorithms (see Section 2.6.1 for an overview). For this study, we use the PC algorithm, for three reasons:

1. **Literature Support**: PC is one of the classical constraint-based algorithms, widely recognised for recovering causal graphs under certain assumptions (e.g., no latent confounders). Although our scenario is not strictly "causal", PC remains effective at identifying potential edges from conditional independence tests (Spirtes et al., 2001).
2. **Interpretability**: By iteratively removing edges inconsistent with the data's conditional independences, PC produces a partially directed acyclic graph (DAG), facilitating a structured, stepwise view of how the BN emerges from the dataset.

3. **Software Compatibility**: GeNIe implements PC with user-friendly parameter settings. This keeps the workflow consistent and reproducible.

However, the choice of the PC algorithm does not preclude the use of other algorithms on this dataset, nor does it imply that PC is the most suitable option. Comparing the performance of different algorithms on this dataset as well as the resulted BNs falls beyond the scope of this study, which aims to illustrate the operationalisation of BN learning for System Dynamics. Nevertheless, this comparison could be a valuable direction for future research.

5.2.2 Variable Discretisation

While two parameters – *disease duration* and *infectivity* – take only a few discrete values (strictly linked to virus variants), the other five (*adjustment start time, contact adjustment level, contact rate, vaccination rate, and therapeutic interventions*) are fundamentally continuous. To keep the entire BN in a discrete format, we carried out the following:

1. **Examine Histograms**: For each of the five continuous parameters, we plotted a histogram across all 126 data points, visually identifying potential cut points.
2. **Create Bins**: We grouped each variable into categories labelled from 1 to N with equal width of interval. We manually determined N for each variable to best preserve characteristics of its distribution.
3. **Record Category Means**: For every parameters, we recorded the mean of the data points within each interval. This allows us to map the BN-generated samples back to their original ranges in a later step.

This binning strategy yields a final dataset in which all variables have discrete states, as shown in Figure 5.1.

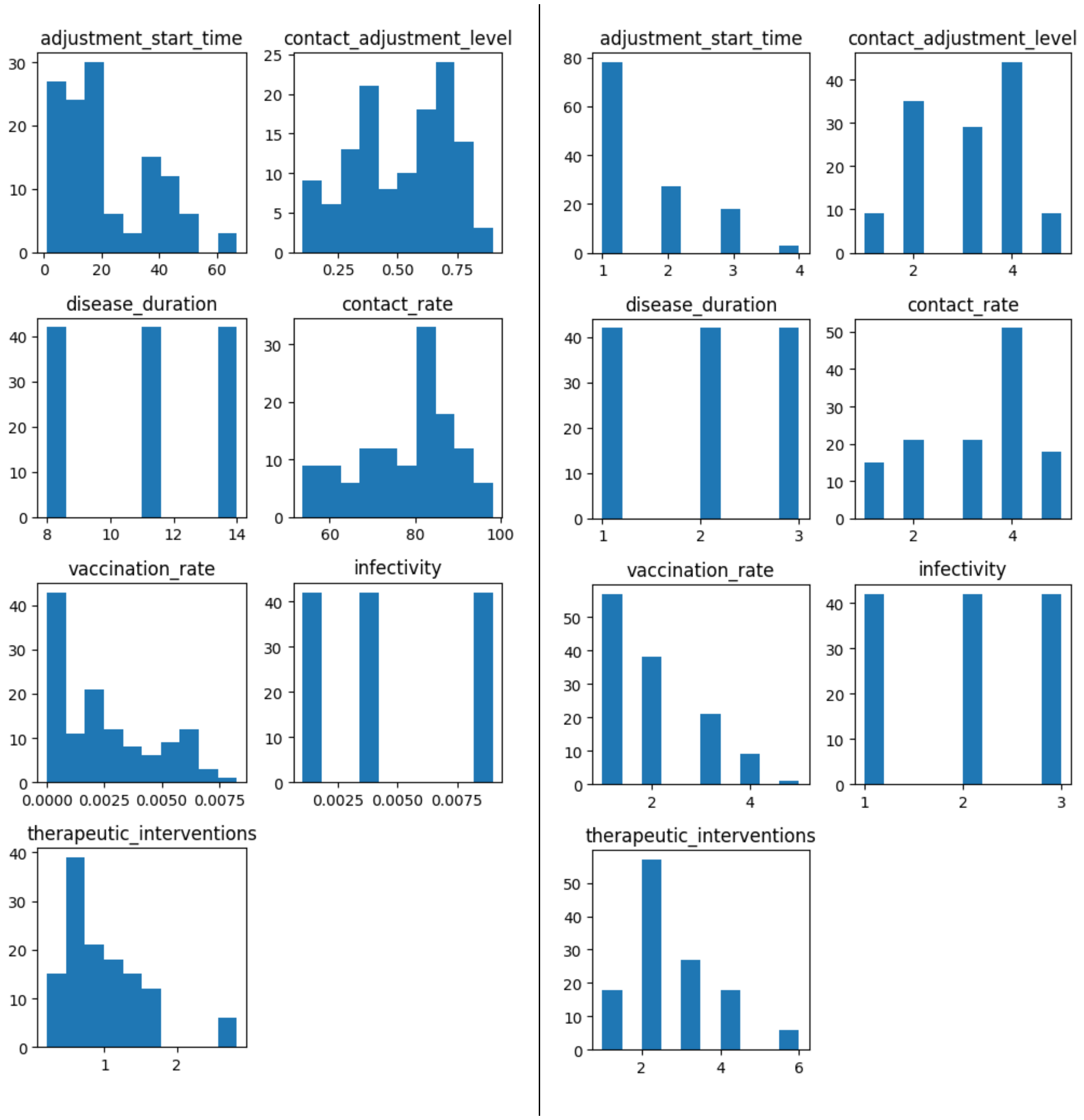


Figure 5.1 Histogram of parameters before (left, x axis is value) and after binning (right, x axis is index)

5.2.3 Structure and Parameter Learning

To start the BN learning procedure, we let PC explore all potential links among the seven parameters, subject to the default algorithm settings:

- Max adjacency size: 8
- Significance threshold: 0.05
- Max time: unlimited

Hence, PC performed a series of conditional independence tests among the discrete variables, progressively removing edges that contradicted these tests until reaching a partially oriented DAG. Edges in that DAG are categorised by the arrow "head" configuration:

- 0 arrowhead: The algorithm is uncertain about the orientation.
- 1 arrowhead: A fully oriented arrow from node A to node B.
- 2 arrowheads: A hidden common cause is suspected but not present in the dataset.

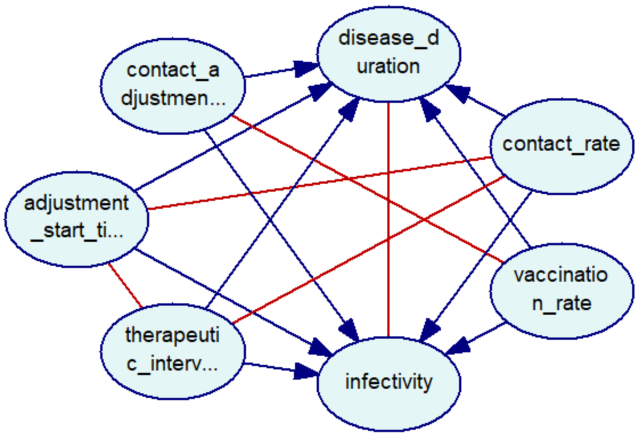


Figure 5.2 Structure learning outcome of the PC algorithm (final run)

We performed a manual review of the suggested edges (Figure 5.2). We specified directions for the undirected or bi-directional edges and examined those directional edges using our domain knowledge (at a level of an average person rather than an expert). For instance, we reversed arrows from *disease duration* and *infectivity* to *contact rate*, drawing on our knowledge that if people perceive a longer disease duration and higher infectivity, they may develop a fear to the disease and thus reduce contacts. Table 5.2 provides interpretation for each learned edge using average-person reasoning.

Table 5.2 Interpretation of the learned parameter dependencies

No.	Learned edge	Interpretation using average-person reasoning
1	Adjustment start time → Contact rate	Countries that respond earlier may end up with lower average contact rates during the wave because restrictions and caution begin sooner.
2	Therapeutic interventions → Contact rate	Better healthcare capacity could correlate with stricter public health management.

3	Therapeutic interventions → Infectivity	Better healthcare systems may correlate with lower effective transmission through better testing/isolation, treatment access, or public health infrastructure.
4	Therapeutic interventions → Disease duration	Higher treatment capacity may shorten illness duration or reduce time spent infectious through earlier care and better management.
5	Infectivity → Contact rate	If the disease is perceived as more infectious or outcomes look worse, people and governments may reduce contacts more.
6	Infectivity → Disease duration	More infectious variants may also be associated with shorter clinical courses.
7	Infectivity → Contact adjustment level	Higher perceived infectivity may trigger stronger policy restrictions or compliance, reflected as higher contact adjustment.
8	Disease duration → Contact rate	If illness lasts longer or people think it does, individuals may reduce contacts to avoid infection or prolonged disruption to their life.
9	Disease duration → Contact adjustment level	Longer disease duration may motivate stronger restrictions to avoid prolonged burden on society and healthcare.
10	Vaccination rate → Infectivity	Higher vaccination reduces effective transmission, lowering "effective infectivity" at the population level.
11	Vaccination rate → Disease duration	Higher vaccination could correlate with shorter average illness duration through milder cases at population level.
12	Vaccination rate → Contact adjustment level	Higher vaccination may coincide with looser restrictions (lower adjustment level) because perceived risk declines.

After deciding orientations for uncertain edges and dropping spurious ones, we let the PC algorithm continue with parameter learning. Within GeNIe, this is straightforward for discrete data: each node's CPT is derived from the observed frequency distribution over itself and its parent nodes. Because our dataset is small but complete (i.e., no missing values), the maximum likelihood estimate (MLE) method was sufficient; more advanced methods could be used if missing or noisy data were a concern.

Overall, this approach yields a purely discrete BN that preserves the observed dependence patterns in our dataset of 126 data points, subject to minimal (but sensible) modifications based on domain knowledge. The structure and the marginal distribution of each node of the BN is shown in Figure 5.3.

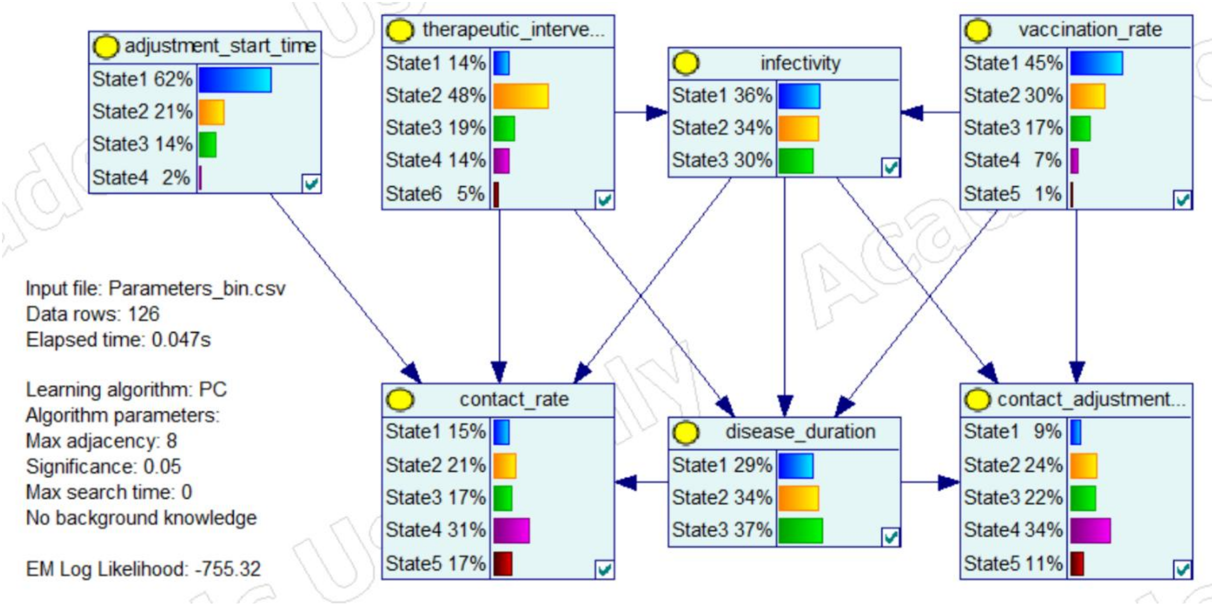


Figure 5.3 Bayesian Network of parameter dependence

While many learned edges are at least qualitatively interpretable at the level of common-sense epidemiological and behavioural reasoning, we also acknowledge that some dependencies are more ambiguous than others. Moreover, limitations of the learning algorithm may lead it to represent latent confounding structures as direct dependencies: when two observed variables are both driven by an unobserved factor (A←C→B), the absence of that latent variable (C) forces the algorithm to choose an arbitrary direction between them (A→B or B→A), even when neither direction represents a meaningful relationship. The table therefore serves as a face-validity check rather than a confirmation of causal structure. In applied settings where the BN is intended to inform policy, these learned dependencies should be further validated through expert review, robustness checks to alternative discretisation and mapping choices, and comparison against independent epidemiological evidence before being treated as substantive relationships.

5.3 Monte Carlo Experiment

5.3.1 Generating Parameter Samples

The Bayesian Network described in Section 5.2 yields a joint discrete distribution of the seven parameters. At this stage, we produce BN-based samples using forward sampling: from root nodes to leaves, following each node's conditional probability table. All variables remain in the discrete states used for BN learning, so each sample is simply a 7-dimensional vector, with each of the seven elements being a category index for the corresponding BN variable. To map BN categories back into numeric values for the SD model, we assign each category its mean value that was calculated beforehand (e.g., "*contact rate*: 3" → 75.2). This yields an input for the SIR simulation (as in the copula experiments). We considered that the sample size N must be large enough to reproduce the dependence between parameters but in the meantime not so large that it hampers the computational feasibility of the Monte Carlo simulation. We chose $N = 500$ after a few trials, which produces a set of 500 discrete parameter vectors, each representing a plausible configuration of the seven model parameters based on empirical relationships rather than assumed correlations.

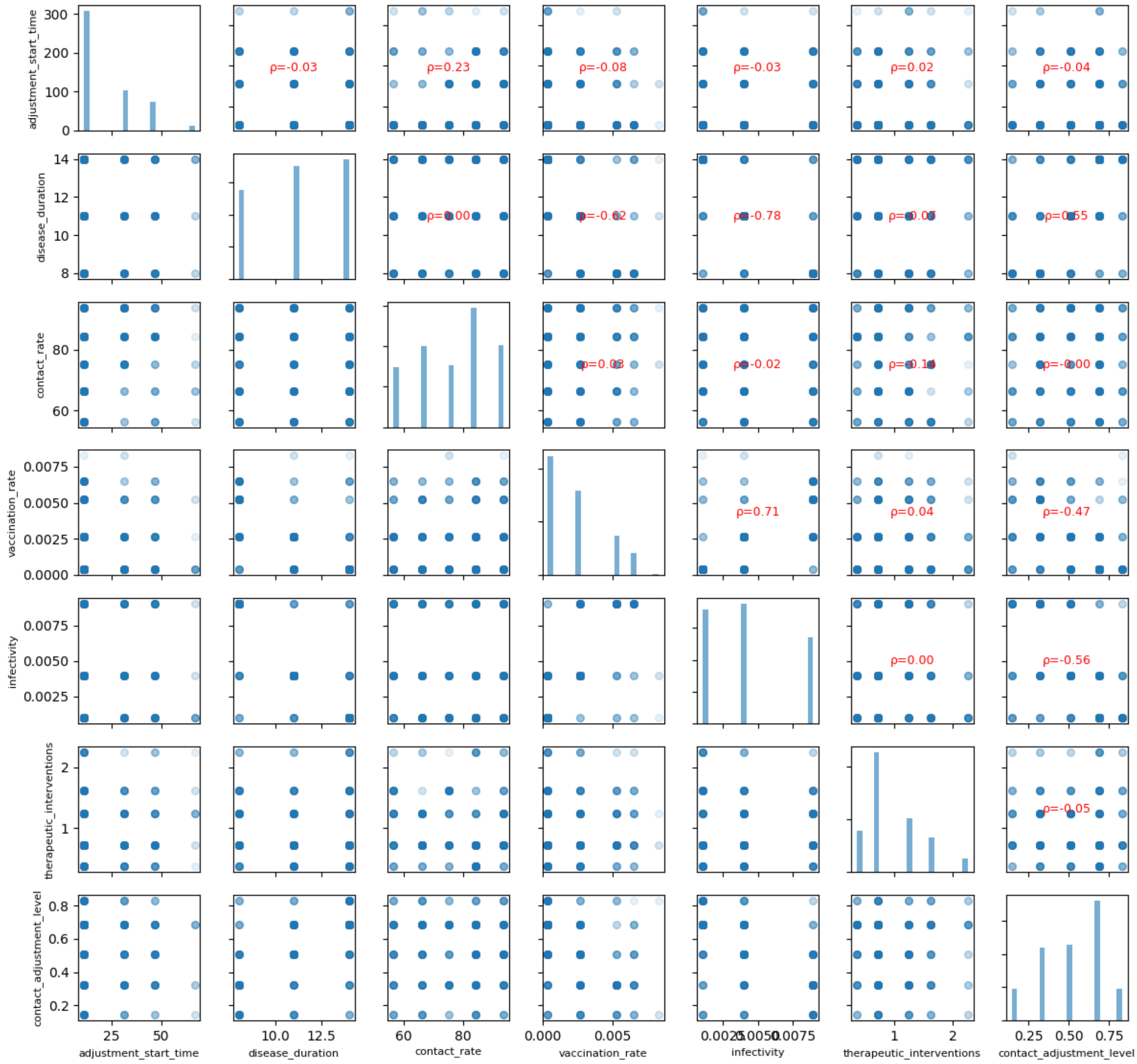


Figure 5.4 Pair plots of samples drawn from forward sampling of BN, where darker colour means higher probability density

Figure 5.4 illustrates the 7×7 pairwise scatter plots for these BN-sampled parameters, highlighting two key observations:

1. **Variation in correlation**: Certain parameter pairs, such as *vaccination rate* vs. *infectivity* and *disease duration* vs. *contact adjustment level*, show moderate or strong correlation, while more than half of the parameter pairs (13 out of 21) show weaker correlation ($\rho < 0.1$).

2. **Discrete clusters**: Because the BN learning process bins continuous variables and assigns them the mean value of their intervals, each parameter's distribution appears in a series of discrete values, forming "grid-like" patterns in the pairwise plots.

5.3.2 Monte Carlo Simulation and Outcome

In each iteration of the Monte Carlo simulation, a parameter vector is taken from the 500 samples to parameterise the SIR SD model. The SD model then performs a time-stepped simulation (e.g., daily steps until a specified end date), which yields a time series projection for each variable, such as daily new infections and vaccination flows. A total number of 500 simulations were run and their results were collected. Distributions of the KPIs are visualised and measured in the same way as for the copulas, as explained in Section 4.2.3.

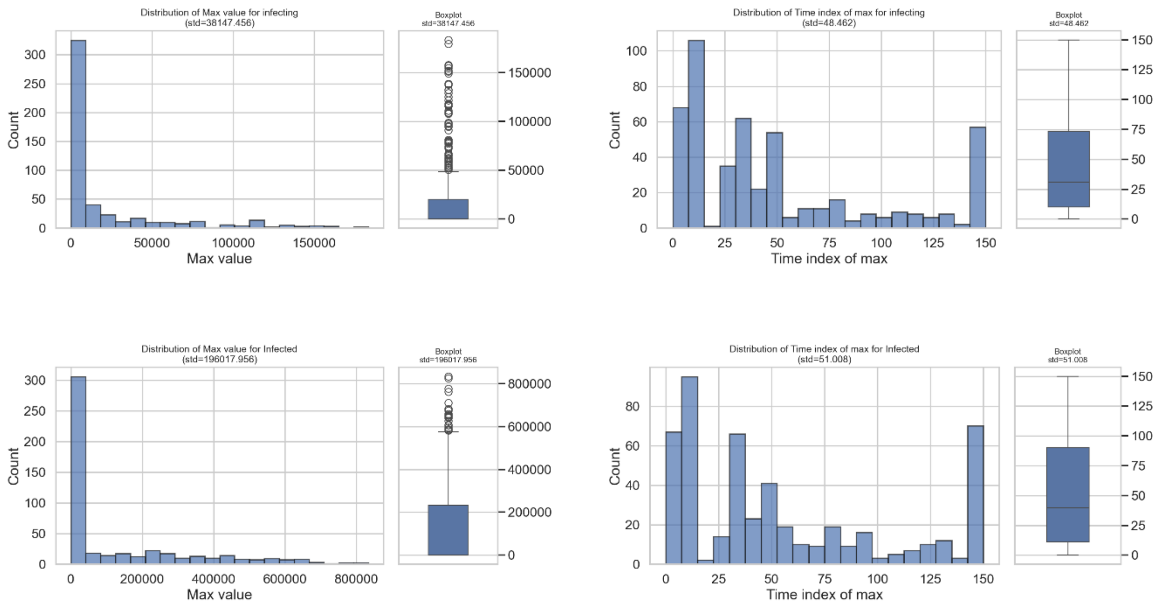


Figure 5.5 Distributions of four KPIs from Monte Carlo simulation using BN-sampled parameters

Figure 5.5 presents the distributions of the four KPIs obtained from Monte Carlo simulations of the SIR SD model, using parameter values sampled from the empirical BN. The peak-magnitude KPIs (Max Infecting and Max Infected) are both highly right-skewed. In the majority of runs, peak daily new infections remain below approximately 10,000 and peak concurrent infections below approximately 40,000, indicating that most BN-consistent parameter configurations generate relatively contained outbreaks. At the same time, the long upper tails and numerous outliers indicate the presence of a smaller subset of parameter combinations that produce very large peaks, reflecting rare but severe epidemic trajectories.

The timing KPIs (Time of Max Infecting and Time of Max Infected) exhibit substantial dispersion and clear multi-modality. A large share of runs peak early or in the mid-course of the simulation, while a pronounced mass near the end of the time horizon (around day 150) indicates that, for a non-negligible subset of parameter draws, the epidemic peak has not arrived within the simulated period. In these cases, the reported "maximum" corresponds to the end-of-horizon value rather than to an interior turning point, suggesting slow- but still-growing outbreaks under certain parameter configurations.

Taken together, the four KPI distributions indicate that most BN-consistent parameter configurations generate relatively contained outbreaks, with moderate peak magnitudes and peaks occurring early or in the middle of the simulation horizon, while a smaller subset produces very large and sometimes late peaks. This pattern is qualitatively epidemiologically plausible: high and early peaks are consistent with combinations of high effective contact rates and slower recovery, whereas more moderate or delayed peaks arise from lower transmission intensity or stronger intervention effects. The coexistence of mild and severe regimes mirrors the diversity of epidemic trajectories observed across countries during the COVID-19 pandemic.

At the same time, these distributions should be interpreted with caution. They are conditional on the specific BN structure, discretisation scheme, structure-learning algorithm, and the quality and coverage of the underlying dataset. Alternative modelling choices could yield different dependence structures and, consequently, different uncertainty envelopes. The purpose of this exercise is therefore not to claim substantive epidemiological insight or policy relevance, but to demonstrate the feasibility and mechanics of coupling a data-driven BN with an SD model to propagate empirically learned parameter dependence into KPI uncertainty assessment. Systematic comparison across alternative BN learning strategies and validation against external benchmarks remain important directions for future work.

5.4 Validation of the Bayesian Network Approach

A question remains as to how the KPI distributions may be validated. Since many – if not most – of the simulated scenarios are hypothesised without a real-world counterpart, it is impractical to validate each individual scenario against an observation (in our case, a COVID-19 outbreak). Alternatively, could the KPI distributions be validated as a whole instead? One might propose comparing the simulated KPI distributions with empirical KPI data collected from the real world (in this case, the real-world KPIs associated with the 126 observations in our dataset) and derive numerical measure of the difference between two distributions – the empirical and the simulated. However, this approach is also flawed, as the empirical KPI distribution is almost certainly less comprehensive than the simulated one (126 data points compared to 500 or more), meaning that many scenarios generated by the simulations may not present in the observed data. An alternative option is to validate the learned BN directly, rather than the KPIs. This validation can be conducted either through expert review (i.e., asking domain experts to assess the learned BN) or statistically, by comparing the parameter dataset to the empirical BN and deriving measures such as Kullback-Leibler divergence (Kullback and Leibler, 1951; Moral et al., 2021). While expert review requires additional effort from domain experts, it is arguably less demanding than eliciting the entire BN from scratch.

5.5 Influence of Parameter Dependence Structure: Revisit

5.5.1 Motivation for Comparing BN and Copulas Results

In section 4.3, we compared KPI distributions from Monte Carlo simulations using parameter values sampled from different copulas at different levels of correlation. The result suggests that the structure of the dependence between parameters influences KPI uncertainty, motivating the use of BN to model the dependence more adequately.

With the BN-based approach detailed in Section 5.1, 5.2 and its outcome presented in Section 5.3, we believe it is relevant to contrast the BN-based KPI distributions with the copula-based KPI distributions. Since BN defines the dependency structure differently from copulas, the difference in KPI distributions may provide additional evidence, further confirming that parameter dependence structure influences KPI uncertainty.

5.5.2 Comparison Setup, Execution, and Result

To ensure fair comparison between BN and copula results, all factors other than the dependence structure must be held constant across the two methods. This means copula samples should be at the same correlation level and drawn from the same range as the BN samples. In Section

4.2.1, copula samples were drawn from a range of $\pm 80\%$ of the parameter's original value; but for the current comparison, their ranges need to be adapted to those of the BN samples, which are drawn from ranges empirically informed by the collected data (although *contact adjustment level* did adopt a range of $\pm 80\%$ when mapped from its proxy variable).

Table 5.3 Arbitrarily chosen vs. empirical data-informed parameter ranges

Parameter	**+/-80% of original value**	**Min/Max from BN samples**
Adjustment start time	(7.0, 63.0)	(10.846, 67.0)
Contact adjustment level	(0.1, 0.9)	(0.144, 0.832)
Disease duration	(2.8, 25.2)	(8.0, 14.0)
Contact rate	(17.0, 153.0)	(56.52, 93.733)
Vaccination rate	(0.0, 0.002)	(0.0, 0.008)
Infectivity	(0.001, 0.008)	(0.001, 0.009)
Therapeutic interventions	(0.2, 1.8)	(0.372, 2.25)

To set up the comparison, we first measured an overall Spearman rank correlation across all parameter pairs in the BN-sampled dataset (similar to the approach in our copula sampling), which is $\rho_{BN} \approx 0.213$. We then replicated this correlation with each of the four copula types (Gaussian, Frank, Gumbel, Clayton) using the numerical method introduced in Section 4.1[4]. Next, we drew 500 samples from each copula and mapped them to the empirically informed range of each parameter, as summarised in Table 5.3, to perform Monte Carlo simulation of the SIR model. Finally, we compared the BN-based KPI uncertainties to the results from the matched-correlation copulas. Results are shown in Figure 5.6 to Figure 5.9.

[4] Estimated θ: $\theta_{Frank} = 1.251; \theta_{Gumbel} = 1.111; \theta_{Clayton} = 0.313$.

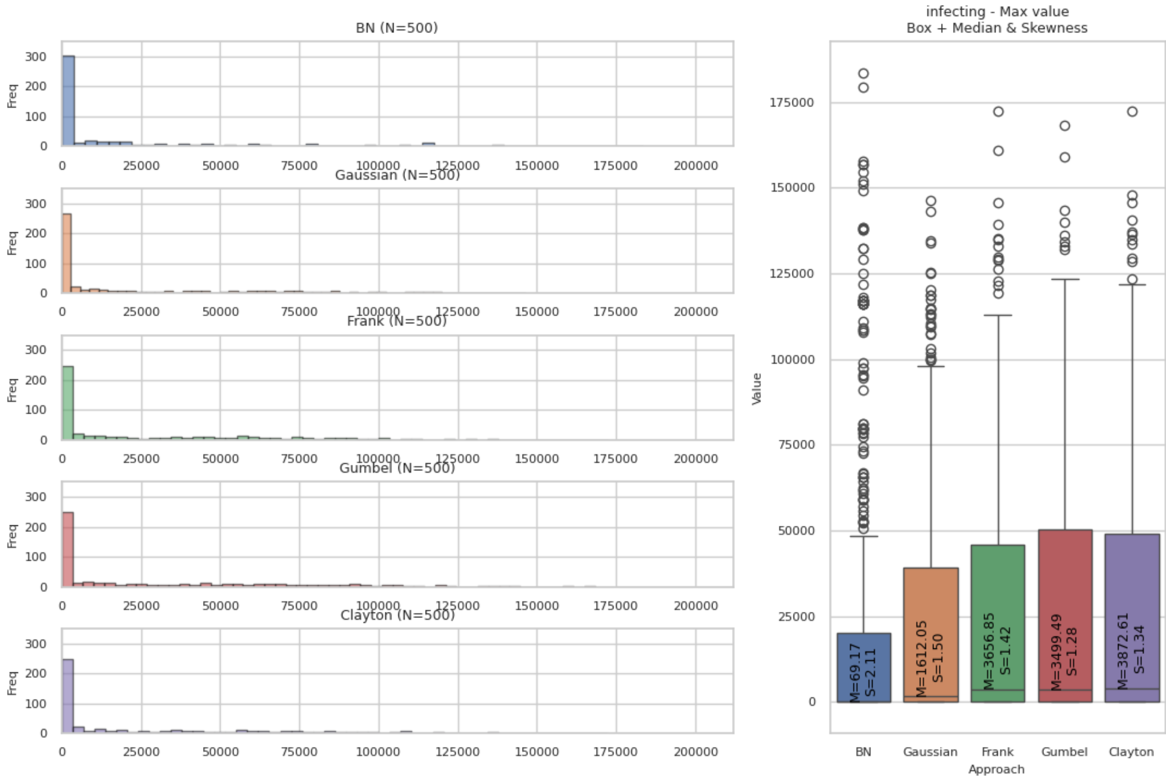


Figure 5.6 BN vs. Copula distribution of KPI (same range): max infecting

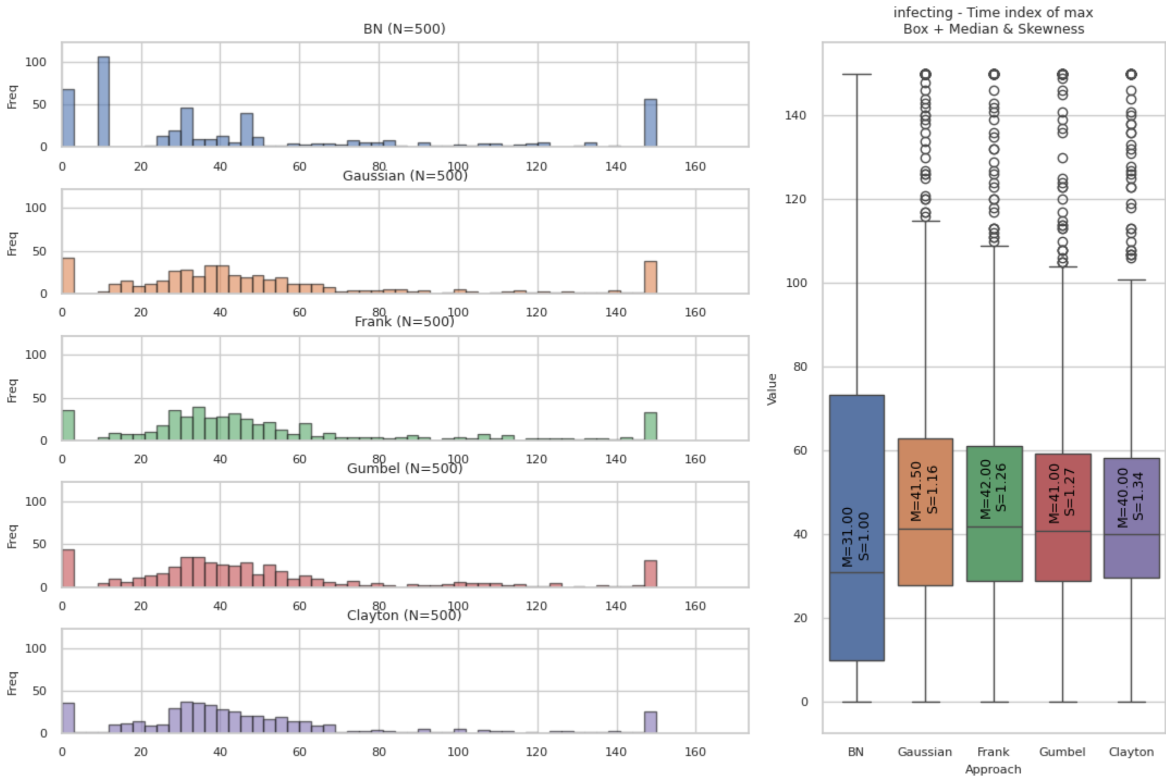


Figure 5.7 BN vs. Copula distribution of KPI (same range): time of max infecting

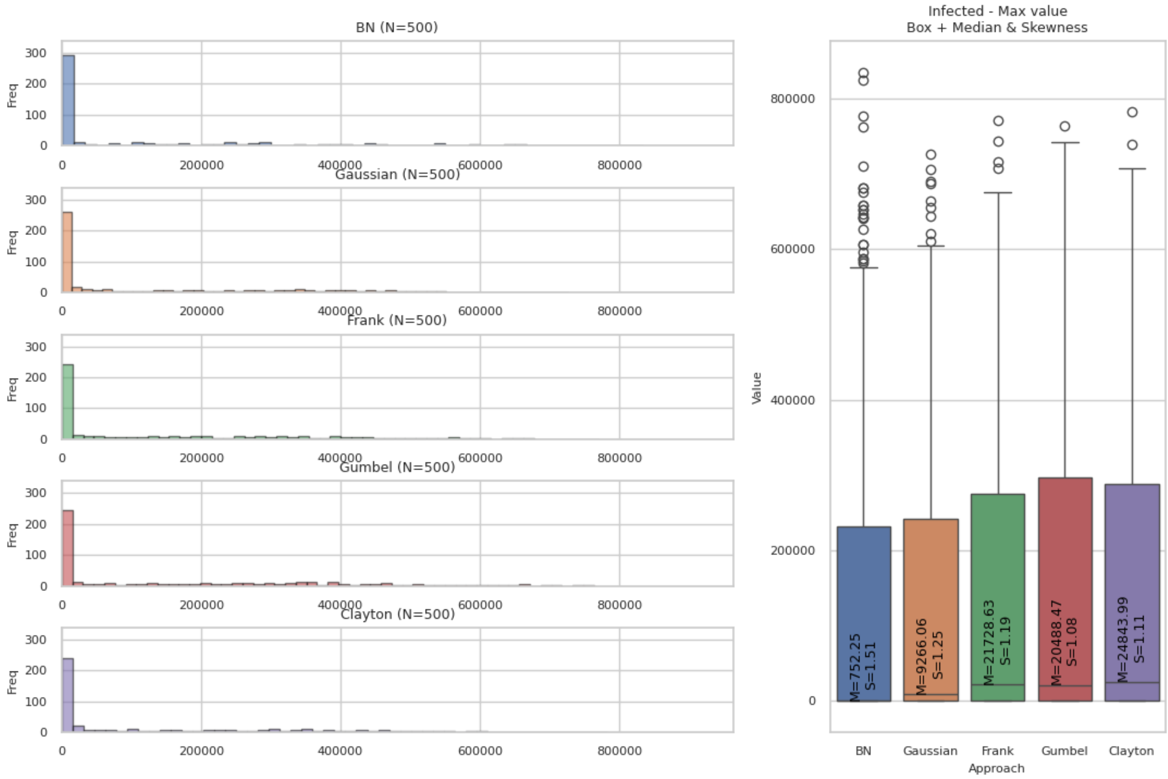


Figure 5.8 BN vs. Copula distribution of KPI (same range): max infected

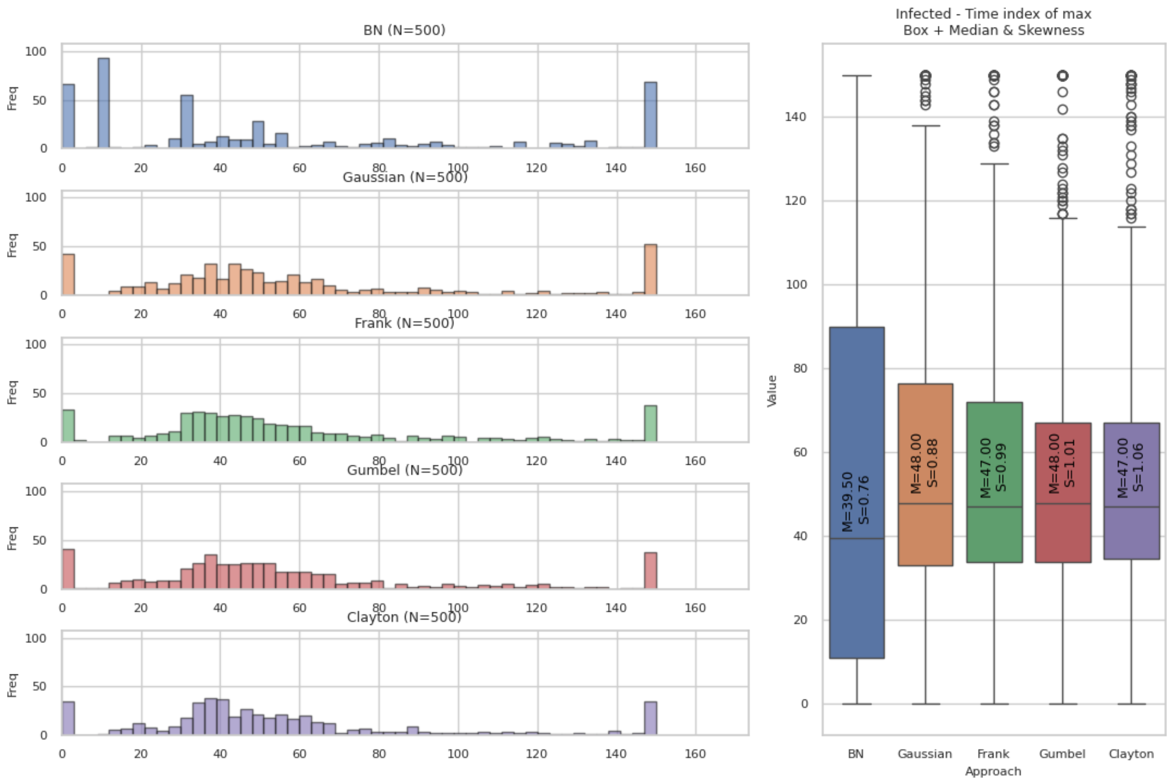


Figure 5.9 BN vs. Copula distribution of KPI (same range): time of max infected

The results show how the BN-based KPI distributions compare with copula-based KPI distributions, when BN and copulas samples are correlated at the same level and drawn from the same ranges (so that *dependence structure* is the only difference). The BN-based KPI distributions in Figure 5.6 to Figure 5.9 (identical to those in Figure 5.5) differ obviously from copula-based KPI distributions in all 4 KPIs. For instance, for KPIs *time of max infecting* and *time of max infected*, BN-based KPI distributions show much higher variance than the copula-based ones, especially on the lower end, suggesting that many potential combinations of parameters – especially those leading to earlier arrival of the peak of infection – are not adequately explored with copulas, and the corresponding scenarios (arguably also the more adverse ones) are *missing*. This may be explained by the fact that we only drew positively correlated samples from the copulas due to the limitation in modelling negatively correlated variables in a high-dimensional space.

These observations provide additional support for our previous conclusion that parameter dependence structure can influence KPI uncertainties and further motivate the use of a more flexible approach to model the multivariate distribution of parameters in SD models.

5.6 Chapter Summary

In this chapter, we detailed the data-driven approach for modelling parameter dependencies. We began by detailing how we acquired and pre-processed cross-country COVID-19 data to form a dataset that covers seven parameters across multiple countries and waves of infections. Next, we constructed a Bayesian Network (BN) from those parameters by discretising continuous variables, applying the PC algorithm to learn the structure in GeNIe, and then refining the partially oriented graph for better plausibility. This yielded a BN that embeds conditional dependencies from actual policy, epidemiological, and healthcare capacity measures. The BN's fine-grained relationships lead to distinct behaviours compared to the simpler, one-parameter copula families. This fulfils the objective of this study, namely, demonstrating BN as a viable empirical method to model interdependence of parameters in SD models.

Chapter 6 Conclusion and Future Work

This chapter closes the thesis by summarising the conclusions, contributions, as well as limitations of this research. It also highlights the broader implications for practitioners and researchers who aim to integrate informed parameter dependence into System Dynamics (SD). Finally, it discusses potential avenues to extend and refine the proposed data-driven Bayesian Network (BN) approach.

6.1 Overall Conclusion

This thesis demonstrates that Bayesian Networks offer a practical, data-driven means of embedding empirically observed parameter interdependencies into System Dynamics models. Using a cross-country COVID-19 dataset, a BN was learned and linked to a seven-parameter SIR SD model, enabling parameter samples that respect observed correlations to drive Monte-Carlo simulations and produce more defensible uncertainty envelopes for KPIs.

While the empirical demonstration is confined to a single epidemiological case, the study establishes a general workflow for learning parameter dependence from data and propagating it into SD uncertainty analysis. The results show that even when marginal parameter ranges are identical, assumptions about dependence structure can alter KPI uncertainty. This finding suggests that independence assumptions – common in SD sensitivity analysis – may be consequential in a broader class of models where parameters are influenced by shared contextual or latent drivers.

The empirical results of this thesis should be interpreted with appropriate caution. The single proof-of-concept case study demonstrates that parameter dependence *can* materially affect uncertainty assessment and that a BN *can* be feasibly learned and coupled with an SD model, but it does not establish that such effects will arise in *all* SD applications, nor that the proposed workflow will *always* improve decision support. The study illustrates methodological feasibility rather than empirical prevalence. In particular, it does not quantify how often parameter dependence is consequential, how sensitive results are to data quality, or how the approach compares systematically with alternative uncertainty-modelling strategies across domains. These questions remain open and constitute important directions for future empirical validation.

6.2 Contributions

Firstly, the thesis proposes a data-driven workflow for modelling dependence among SD model parameters using BN. The workflow comprises assembling panel data, discretising variables, learning BN structure and parameters, and forward-sampling coherent parameter sets to assess of KPI uncertainty. Together, these steps provide SD practitioners with a scalable alternative to resource-intensive expert elicitation, ad-hoc assumptions, or purely parametric dependence models, enabling more defensible uncertainty assessments while keeping additional data-collection effort within manageable bounds. The thesis also introduces a copula-based procedure as an optional motivational step to illustrate when dependence may matter, as summarised in Figure 3.1.

Secondly, the thesis advances the SD literature on the role of parameter dependence in uncertainty analysis. The copula experiments demonstrate that dependence structure – not merely the strength of correlation – can distort KPI uncertainty envelopes, lending empirical support to concerns raised by McNaught (2003) and Eker et al. (2014) regarding implausible parameter combinations. The BN-based approach offers a practical alternative to expanding the model boundary (Ford and Flynn, 2005), providing greater flexibility than structural equations and more scalable handling of multivariate dependence than lookup tables.

Finally, the thesis contributes to the growing body of work integrating System Dynamics and Bayesian Networks. Closely related designs include Bertone et al. (2018), who used an expert-elicited BN to support scenario generation. The present work extends this line by showing that scenario BN can also be learned from empirical data.

6.3 Relevance beyond Epidemiological Modelling

Although the proof-of-concept application is epidemiological, the relevance of parameter dependencies extends well beyond infectious-disease modelling. In many SD applications, the same structural model is applied across heterogeneous contexts – countries, regions, firms, or policy regimes – where parameters are shaped by shared background conditions that are external to the model boundary. For example, when adapting a market-diffusion model across countries, income levels may correlate with price sensitivity; in public-health models, compliance rates may co-vary with reporting delays.

In such cases, the correlations arise not from direct causal relations within the SD structure, but from contextual drivers – institutional quality, cultural norms, or regulatory regimes – that influence several parameters simultaneously. In such settings, assuming independent variation among parameters risks generating implausible combinations and distorted uncertainty envelopes. As McNaught (2003, p. 8) observed, recognising parameter interdependence "adds another dimension to the elicitation process," shifting attention from isolated ranges to coherent configurations of background conditions.

At the same time, parameter dependencies will not be equally consequential in all SD applications. In models where parameters are well identified from a single, homogeneous context, or where KPI behaviour is dominated by a small number of structural feedbacks, independence assumptions may be largely acceptable. Similarly, when parameter uncertainty is narrow or when KPIs depend weakly on parameter interactions, modelling dependence may offer little practical gain. The relevance of parameter dependence is therefore inherently model- and context-specific, and should be assessed pragmatically through preliminary sensitivity analysis before investing in more elaborate dependence modelling.

6.4 Practical Implications

From a practical perspective, the thesis underscores that following implications:

1. Ensure that scenario exploration is based on parameter assumptions that are coherent with observed data.
2. Derive parameter correlations from pooled data of similar events, as an alternative to judgemental or statistical methods.
3. Use partial domain knowledge to refine the network (e.g., exclude nonsensical edges) to avoid overfitting to the data.

Nevertheless, adopting a BN approach requires sufficient data coverage. Where data of similar events is scarce or highly noisy, this approach can be challenging or even impractical. The practitioners may still need to adopt a more effortful approach by involving experts or sourcing data on individual parameters.

6.5 Limitations and Future Work

6.5.1 Technical Limitations

6.5.1.1 The Continuity-Flexibility-Interpretability Trade-off

Statistical dependence among multiple variables is challenging to model because three desirable properties – continuity, flexibility, and interpretability – cannot be maximised simultaneously. Discrete BNs give clarity but lose information during discretising; parametric or hierarchical Bayes models preserve continuity but constrain shape, resulting in limited flexibility in capturing dependence structures; neural networks for probability density estimation (i.e., neural density estimators) (Bishop, 1994) deliver continuity and almost unlimited flexibility but with minimal transparency. However, there is a rich trade-off space between these three objectives. This subsection discusses potential extensions to our work within this broader picture.

In this thesis, we used equal-width binning for our discrete BN, a choice that is quick but not necessarily optimal. Studies have found that alternative binning scheme such as k-means discretisation (Sari et al., 2021) and supervised discretisation (Dougherty et al., 1995) can retain more information. Metrics such as model accuracy, F-measure, Kullback-Leibler divergence can quantify how much predictive power is lost when continuous variables are discretised (Beuzen et al., 2018; De Campos, 2006; Marcot, 2012). Future work should benchmark several binning strategies – equal-width, equal-frequency, machine learning-based, and expert-elicited – against these information loss metrics before selecting one for SD practice.

Moreover, discretisation is not the only path. Conditional Linear Gaussian BNs keep some nodes continuous while maintaining exact inference for the discrete portion of the network (Madsen, 2008; McGeachie et al., 2014). Other work shows that the Gaussian density can be swapped out for more flexible piece-wise continuous densities that still allow exact probability propagation within the BN, including mixtures of truncated exponentials (MTEs) (Moral et al., 2001), mixtures of polynomials (MoPs) (Shenoy and West, 2009), and their tailed extension tMoPs (Luengo et al., 2025). This flexibility in representing capability of modelling continuous variables is particularly relevant to SD parameter modelling, as most of them are continuous, thus worth exploring in future work.

Moving toward greater flexibility, neural density estimators (NDEs) – such as mixture-density networks (Bishop, 1994) and normalising flows (Kobyzev et al., 2021; Papamakarios et al., 2021) – can approximate virtually any continuous joint distributions, thus providing both continuity and flexibility. Their hidden layers, however, often obscure the causal narratives valued by policy stakeholders. A promising compromise is to hybridise the approaches: retain an explicit high-level causal skeleton (e.g., a BN-style graph) but model each complex conditional distribution with a compact neural density estimator. Such "graph-informed neural networks" have already appeared in the causal-inference literature (e.g., Goudet et al., 2018) and could be adapted to the SD context for modelling more complex parameter dependencies.

6.5.1.2 Modelling Time-Varying Parameters and Dependencies

The present study treats all parameters as static. In real systems, however, parameters such as contact rate can change over time – and so can the dependencies among parameters. Dynamic Bayesian Networks (DBNs) extend standard BNs by replicating the graph over successive time slices and adding edges from each variable's previous-time instance to its current instance, so that every variable can be conditioned on its own past state (as well as on other variables) (Mihajlovic and Petkovic, 2001; Murphy, 2002; Scutari, 2020). DBNs have been applied to model longitudinal datasets and are robust to missing observations (Begashaw et al., 2024; Liu et al., 2019). Future work could integrate a DBN layer into an SD model to capture the influence of time-varying exogeneous drivers on a dynamic system.

6.5.1.3 Validation, Robustness, and Bayesian Workflow

Whichever modelling path is adopted, validation of the BN should become routine parts of the BN-SD uncertainty workflow. The quickest option would be to forward-simulate new data from the learned BN and compare key summaries of the simulated data to the real observations, as discussed in Section 5.4. It is computationally light and often adequate when the dataset is large and BN parameter uncertainty is small.

For a more comprehensive assessment, one can perform a posterior predictive check (PPC) (Rubin, 1984; Gelman et al., 1996; Berkhof et al., 2000). In a BN setting, the idea is to repeatedly draw a new set of CPTs from their posterior distributions (generated by the BN-learning algorithm), forward-simulate a synthetic data set from each draw and then compare summary features of these replicated data sets with the same features in the real data on which the BN is learned. The collection of replicated summaries forms an empirical “reference distribution”; if the observed summary falls in the extreme tails of this reference distribution, the BN is flagged as mis-fitting that aspect of the data. PPCs can reveal mis-fit that a point-estimate likelihood overlooks – for example, tails that are too light or multimodality not captured by the BN. Crawford (2014) provides a comprehensive guide to applying PPC on BN.

In addition to predictive checks, cross-validation across discretisation schemes, BN-learning algorithms, or alternative networks produced by different random seeds will reveal how sensitive the BN is to each modelling choice. Without such checks, the workflow risks unwarranted confidence in the learned BN, and subsequently in the seemingly empirically grounded KPI envelopes.

6.5.2 Methodological Limitations

Beyond technical considerations, the study has several methodological limitations. First, the approach presumes that repeated observations across comparable contexts are sufficiently available for data-driven BN modelling to be meaningful; where such data are sparse, dependence learning may be unreliable. Second, the workflow focuses on dependence among parameters rather than uncertainty in the SD structure itself, leaving open questions about interactions between parameter dependence and structural misspecification. Third, discretisation and structure-learning choices for learning a BN introduce modelling degrees of freedom that can influence inferred dependencies, underscoring the importance of predictive checks and sensitivity analysis. Finally, the study does not provide a principled criterion for when modelling parameter dependence is warranted, suggesting the need for preliminary screening procedures to guide its application in practice.

6.5.3 Closing Reflection

Taken together, these limitations highlight that advances in dependence modelling inevitably come with new demands on data, judgement, and interpretation. While new methods create new opportunities, they also require careful judgement about whether and when to adopt them and raise the bar for stakeholder engagement. As model complexity increases, transparency and ease of explanation tend to diminish. Practitioners therefore face a multi-objective trade-off – continuity, flexibility, interpretability, and dynamic realism – that rarely allows one goal to be maximised without compromise. The optimal balance is inherently context-dependent and should be chosen to serve the decision at hand, the available evidence, and the audience's capacity – their technical fluency, time constraints, and willingness to engage. As Box (1976) famously observed, "all models are wrong, but some are useful"; the task is not to eliminate imperfection, but to ensure that added sophistication yields commensurate decision value. By clarifying these tread-offs and outlining next steps, this section aims to chart a pragmatic route for advancing uncertainty analysis in System Dynamics.

References

Abdelbari, H., Shafi, K., 2017. A computational Intelligence-based Method to 'Learn' Causal Loop Diagram-like Structures from Observed Data. Syst. Dyn. Rev. 33, 3–33.

Abdulkareem, S.A., Mustafa, Y.T., Augustijn, E.-W., Filatova, T., 2019. Bayesian networks for spatial learning: a workflow on using limited survey data for intelligent learning in spatial agent-based models. GeoInformatica 23, 243–268. https://doi.org/10.1007/s10707-019-00347-0

Alhamlan, F.S., Al-Qahtani, A.A., 2025. SARS-CoV-2 Variants: Genetic Insights, Epidemiological Tracking, and Implications for Vaccine Strategies. Int. J. Mol. Sci. 26, 1263. https://doi.org/10.3390/ijms26031263

Aliferis, C.F., Tsamardinos, I., Statnikov, A., 2003. HITON: A Novel Markov Blanket Algorithm for Optimal Variable Selection. AMIA. Annu. Symp. Proc. 2003, 21–25.

Andersen, D.F., Richardson, G.P., 1997. Scripts for group model building. Syst. Dyn. Rev. J. Syst. Dyn. Soc. 13, 107–129.

Andrade, J., Duggan, J., 2020. An evaluation of Hamiltonian Monte Carlo performance to calibrate age-structured compartmental SEIR models to incidence data. Epidemics 33, 100415.

Antunes, P., Santos, R., Videira, N., 2006. Participatory decision making for sustainable development—the use of mediated modelling techniques. Land Use Policy, Resolving Environmental Conflicts:Combining Participation and Muli-Criteria Analysis 23, 44–52. https://doi.org/10.1016/j.landusepol.2004.08.014

Barlas, Y., 1996. Formal aspects of model validity and validation in system dynamics. Syst. Dyn. Rev. J. Syst. Dyn. Soc. 12, 183–210.

Barlas, Y., Carpenter, S., 1990. Philosophical roots of model validation: two paradigms. Syst. Dyn. Rev. 6, 148–166.

Barlas, Y., Kanar, K., 1999. A dynamic pattern-oriented test for model validation. Presented at the Proceedings of 4th systems science European congress, pp. 269–286.

Barons, M.J., Mascaro, S., Hanea, A.M., 2022. Balancing the Elicitation Burden and the Richness of Expert Input When Quantifying Discrete Bayesian Networks. Risk Anal. 42, 1196–1234. https://doi.org/10.1111/risa.13772

Beal, M.J., Ghahramani, Z., 2003. The Variational Bayesian EM Algorithm for Incomplete Data: With Application to Scoring Graphical Model Structures, in: Lindley, V., Bernardo, J.M., Bayarri, M.J., Berger, J.O., Dawid, A.P., Heckerman, D., Smith, A.F., West, M. (Eds.), Bayesian Statistics 7. Oxford University PressOxford, pp. 453–463. https://doi.org/10.1093/oso/9780198526155.003.0025

Begashaw, G.B., Zewotir, T., Fenta, H.M., 2024. Dynamic Bayesian network modeling for longitudinal data on child undernutrition in Ethiopia (2002–2016). Front. Public Health 12, 1399094. https://doi.org/10.3389/fpubh.2024.1399094

Bennett, P.G., 1985. On linking approaches to decision-aiding: issues and prospects. J. Oper. Res. Soc. 36, 659–669.

Berkhof, J., van Mechelen, I., Hoijtink, H., 2000. Posterior predictive checks: Principles and discussion. Comput. Stat. 15, 337–354. https://doi.org/10.1007/s001800000038

Bernardo, J.M., Smith, A.F.M., 1994. Bayesian theory, 1st ed, Wiley series in probability and mathematical statistics. Wiley.

Bertone, E., Sahin, O., Richards, R., Roiko, A., 2016. Modelling with stakeholders: a systems approach for improved environmental decision making under great uncertainty, in: Proceedings - 8th International Congress on Environmental Modelling and Software, iEMSs 2016. Presented at the 8th International Congress on Environmental Modelling and Software, Toulouse, France.

Bertone, E., Sahin, O., Stewart, R.A., Zou, P.X.W., Alam, M., Hampson, K., Blair, E., 2018. Role of financial mechanisms for accelerating the rate of water and energy efficiency retrofits in Australian public buildings: Hybrid Bayesian Network and System Dynamics modelling approach. Appl. Energy 210, 409–419. https://doi.org/10.1016/j.apenergy.2017.08.054

Beuzen, T., Marshall, L., Splinter, K.D., 2018. A comparison of methods for discretizing continuous variables in Bayesian Networks. Environ. Model. Softw. 108, 61–66. https://doi.org/10.1016/j.envsoft.2018.07.007

Bhaskar, R., 2013. A realist theory of science. Routledge.

Bhaskar, R., 1997. On the Ontological Status of Ideas. J. Theory Soc. Behav. 27, 139–147. https://doi.org/10.1111/1468-5914.00031

Bishop, C.M., 1994. Mixture density networks. Aston University.

Box, G.E.P., 1976. Science and Statistics. J. Am. Stat. Assoc. 71, 791–799. https://doi.org/10.1080/01621459.1976.10480949

Chen, S.H., Pollino, C.A., 2012. Good practice in Bayesian network modelling. Environ. Model. Softw. 37, 134–145. https://doi.org/10.1016/j.envsoft.2012.03.012

Chen, Y., Tu, Y., Jeng, B., 2011. A Machine Learning Approach to Policy Optimization in System Dynamics Models. Syst. Res. Behav. Sci. 28, 369–390. https://doi.org/10.1002/sres.1089

Chickering, D.M., 2002. Learning equivalence classes of Bayesian-network structures. J. Mach. Learn. Res. 2, 445–498.

Chikkagoudar, S., Chatterjee, S., Bharadwaj, R., Ganguly, A., Kompella, S., Thorsen, D., 2022. Assurance by Design for Cyber-physical Data-driven Systems, in: Douglass, R., Gremban, K., Swami, A., Gerali, S. (Eds.), IoT for Defense and National Security. Wiley, pp. 191–212. https://doi.org/10.1002/9781119892199.ch11

Colombo, D., Maathuis, M.H., 2014. Order-independent constraint-based causal structure learning. J Mach Learn Res 15, 3741–3782.

Colombo, D., Maathuis, M.H., Kalisch, M., Richardson, T.S., 2012. Learning high-dimensional directed acyclic graphs with latent and selection variables. Ann. Stat. 294–321.

Cooke, R., 1991. Experts in uncertainty: opinion and subjective probability in science. Oxford University Press on Demand.

Cooper, G.F., Herskovits, E., 1992. A Bayesian method for the induction of probabilistic networks from data. Mach. Learn. 9, 309–347. https://doi.org/10.1007/BF00994110

COVID-19 Government Response Tracker | Blavatnik School of Government [WWW Document], 2020. URL http://www.bsg.ox.ac.uk/research/covid-19-government-response-tracker (accessed 2.28.25).

Coyle, G., 2000. Qualitative and quantitative modelling in system dynamics: some research questions. Syst. Dyn. Rev. 16, 225–244. https://doi.org/10.1002/1099-1727(200023)16:3%253C225::AID-SDR195%253E3.0.CO;2-D

Coyle, R.G., 1977. Management System Dynamics. Wiley.

Crawford, A., 2014. Posterior Predictive Model Checking in Bayesian Networks. ProQuest LLC. Arizona State University.

Crookes, D.J., 2017. Does a reduction in the price of rhino horn prevent poaching? J. Nat. Conserv. 39, 73–82. https://doi.org/10.1016/j.jnc.2017.07.008

Daly, R., Shen, Q., Aitken, S., 2011. Learning Bayesian networks: approaches and issues. Knowl. Eng. Rev. 26, 99–157. https://doi.org/10.1017/S0269888910000251

De Campos, L.M., 2006. A scoring function for learning Bayesian networks based on mutual information and conditional independence tests. J. Mach. Learn. Res. 7.

De Finetti, B., 1937. Foresight: Its logical laws, its subjective sources, in: Breakthroughs in Statistics: Foundations and Basic Theory. Springer, pp. 134–174.

Dougherty, J., Kohavi, R., Sahami, M., 1995. Supervised and Unsupervised Discretization of Continuous Features, in: Machine Learning Proceedings 1995. Elsevier, pp. 194–202. https://doi.org/10.1016/B978-1-55860-377-6.50032-3

Dowe, P., 2000. Physical Causation. Cambridge University Press.

Drobek, M., Gilani, W., Molka, T., Soban, D., 2015. Automated equation formulation for causal loop diagrams, in: International Conference on Business Information Systems. Springer, pp. 38–49.

Drobek, M., Gilani, W., Soban, D., 2014. A data driven and tool supported CLD creation approach, in: The 32nd International Conference of the System Dynamics Society, Delft. pp. 1–20.

Duggan, J., 2008. Equation-based policy optimization for agent-oriented system dynamics models. Syst. Dyn. Rev. 24, 97–118. https://doi.org/10.1002/sdr.393

Earman, J., 1992. Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory. Cambridge, ma: MIT Press.

Eberlein, R.L., 1984. Simplifying dynamic models by retaining selected behavior modes (PhD Thesis). Massachusetts Institute of Technology.

Eker, S., Slinger, J., van Daalen, E., Yucel, G., 2014. Sensitivity analysis of graphical functions. Syst. Dyn. Rev. https://doi.org/10.1002/sdr.1518

Embrechts, P., Lindskog, F., McNeil, A., 2003. Modelling dependence with copulas and applications to risk management. Handb. Heavy Tailed Distrib. Finance 8, 329–384.

Ferrara, N., Campobasso, C.P., Cocozza, S., Conti, V., Davinelli, S., Costantino, M., Cannavo, A., Rengo, G., Filippelli, A., Corbi, G., 2022. Relationship between COVID-19 Mortality, Hospital Beds, and Primary Care by Italian Regions: A Lesson for the Future. J. Clin. Med. 11, 4196. https://doi.org/10.3390/jcm11144196

Ferrell, W.R., 1985. Combining individual judgments, in: Behavioral Decision Making. Springer, pp. 111–145.

Fiddaman, T.S., 2002. Exploring policy options with a behavioral climate–economy model. Syst. Dyn. Rev. J. Syst. Dyn. Soc. 18, 243–267.

Fiddaman, T.S., 1997. Feedback complexity in integrated climate-economy models (PhD Thesis). Massachusetts Institute of Technology.

Ford, A., 1990. Estimating the impact of efficiency standards on the uncertainty of the Northwest electric system. Oper. Res. 38, 580–597.

Ford, A., Flynn, H., 2005. Statistical screening of system dynamics models. Syst. Dyn. Rev. J. Syst. Dyn. Soc. 21, 273–303.

Ford, D.N., 1999. A behavioral approach to feedback loop dominance analysis. Syst. Dyn. Rev. 15, 3–36. https://doi.org/10.1002/(SICI)1099-1727(199921)15:1%253C3::AID-SDR159%253E3.0.CO;2-P

Ford, D.N., Sterman, J.D., 1998. Expert knowledge elicitation to improve formal and mental models. Syst. Dyn. Rev. 14, 309–340. https://doi.org/10.1002/(SICI)1099-1727(199824)14:4%253C309::AID-SDR154%253E3.0.CO;2-5

Forrester, J.W., 2007. System dynamics—a personal view of the first fifty years. Syst. Dyn. Rev. 23, 345–358. https://doi.org/10.1002/sdr.382

Forrester, J.W., 1994. System dynamics, systems thinking, and soft OR. Syst. Dyn. Rev. 10, 245–256.

Forrester, J.W., 1987. Lessons from system dynamics modeling. Syst. Dyn. Rev. 3, 136–149.

Forrester, J.W., 1980. Information sources for modeling the national economy. J. Am. Stat. Assoc. 75, 555–566.

Forrester, J.W., 1969. Urban Dynamics. Pegasus Communications, Waltham, MA.

Forrester, J.W., 1961. Industrial dynamics. Cambridge, Mass. M.I.T. Press and Wiley, Cambridge, Mass.

Gelman, A., Meng, X.-L., Stern, H., 1996. Posterior Predictive Assessment of Model Fitness Via Realized Discrepancies. Stat. Sin. 6, 733–760.

Gelman, A., Shalizi, C.R., 2013. Philosophy and the practice of Bayesian statistics. Br. J. Math. Stat. Psychol. 66, 8–38. https://doi.org/10.1111/j.2044-8317.2011.02037.x

Goodman, M.R., 1997. Study Notes in System Dynamics. J. Oper. Res. Soc. 48, 1147–1147. https://doi.org/10.1057/palgrave.jors.2600963

Gosling, J.P., 2018. SHELF: the Sheffield elicitation framework, in: Elicitation. Springer, pp. 61–93.

Goudet, O., Kalainathan, D., Caillou, P., Guyon, I., Lopez-Paz, D., Sebag, M., 2018. Learning Functional Causal Models with Generative Neural Networks, in: Escalante, H.J., Escalera, S., Guyon, I., Baró, X., Güçlütürk, Y., Güçlü, U., van Gerven, M. (Eds.), Explainable and Interpretable Models in Computer Vision and Machine Learning. Springer International Publishing, Cham, pp. 39–80. https://doi.org/10.1007/978-3-319-98131-4_3

Graham, A.K., 1980. Parameter estimation in system dynamics modeling. Elem. Syst. Dyn. Method 143–161.

Grossler, A., 2007. System dynamics projects that failed to make an impact. Syst. Dyn. Rev. https://doi.org/10.1002/sdr.365

Guneralp, B., 2006. Towards coherent loop dominance analysis: progress in eigenvalue elasticity analysis. Syst. Dyn. Rev. https://doi.org/10.1002/sdr.343

Hafezi, M., Sahin, O., Stewart, R., Ware, D., Mackey, B., 2017. Participatory dynamic modelling approach for adaptation planning needs focusing on Small Islands Developing States.

Hafezi, M., Stewart, R.A., Sahin, O., Giffin, A.L., Mackey, B., 2021. Evaluating coral reef ecosystem services outcomes from climate change adaptation strategies using integrative system dynamics. J. Environ. Manage. 285, 112082. https://doi.org/10.1016/j.jenvman.2021.112082

Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S., Tatlow, H., 2021. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). Nat. Hum. Behav. 5, 529–538. https://doi.org/10.1038/s41562-021-01079-8

Hamarat, C., Kwakkel, J.H., Pruyt, E., Loonen, E.T., 2014. An exploratory approach for adaptive policymaking by using multi-objective robust optimization. Simul. Model. Pract. Theory, Simulation-Optimization of Complex Systems: Methods and Applications 46, 25–39. https://doi.org/10.1016/j.simpat.2014.02.008

Heckerman, D., Geiger, D., Chickering, D.M., 1995. Learning Bayesian networks: The combination of knowledge and statistical data. Mach. Learn. 20, 197–243. https://doi.org/10.1007/BF00994016

Hekimoglu, M., Barlas, Y., 2016. Sensitivity analysis for models with multiple behavior modes: a method based on behavior pattern measures. Syst. Dyn. Rev. https://doi.org/10.1002/sdr.1568

Homer, J., Oliva, R., 2001. Maps and models in system dynamics: a response to Coyle. Syst. Dyn. Rev. 17, 347–355. https://doi.org/10.1002/sdr.224

Homer, J.B., 2012. Partial-model testing as a validation tool for system dynamics (1983). Syst. Dyn. Rev. 28, 281–294. https://doi.org/10.1002/sdr.1478

Homer, J.B., 1996. Why we iterate: scientific modeling in theory and practice. Syst. Dyn. Rev. J. Syst. Dyn. Soc. 12, 1–19.

Howick, S., Ackermann, F., 2011. Mixing OR methods in practice: Past, present and future directions. Eur. J. Oper. Res. 215, 503–511. https://doi.org/10.1016/j.ejor.2011.03.013

Howson, C., Urbach, P., 2006. Scientific reasoning: the Bayesian approach. Open Court Publishing.

Hume, D., 1896. A treatise of human nature. Clarendon Press.

Ibrahim, A., Daniyal, H., Asmawaty, T., Kamaludin, A., 2021. Potential Data Collections Methods for System Dynamics Modelling: A Brief Overview. Int. J. Adv. Comput. Sci. Appl. 12. https://doi.org/10.14569/IJACSA.2021.0120332

iseesystems, 2020. Stocks [WWW Document]. URL https://www.iseesystems.com/resources/help/v2/#08-Reference/01-ObjectsAndProperties/01-BuildingBlocks/Stocks.htm (accessed 2.1.24).

Jadun, P., Vimmerstedt, L.J., Bush, B.W., Inman, D., Peterson, S., 2017. Application of a variance-based sensitivity analysis method to the Biomass Scenario Learning Model. Syst. Dyn. Rev. 33, 311–335. https://doi.org/10.1002/sdr.1594

Jeng, B., Chen, J., Liang, T., 2006. Applying data mining to learn system dynamics in a biological model. Expert Syst. Appl. 30, 50–58.

Kitson, N.K., Constantinou, A.C., Guo, Z., Liu, Y., Chobtham, K., 2023. A survey of Bayesian Network structure learning. Artif. Intell. Rev. 56, 8721–8814. https://doi.org/10.1007/s10462-022-10351-w

Kleijnen, J.P.C., 1995. Sensitivity analysis and optimization of system dynamics models: Regression analysis and statistical design of experiments. Syst. Dyn. Rev. 11, 275–288. https://doi.org/10.1002/sdr.4260110403

Kobyzev, I., Prince, S.J.D., Brubaker, M.A., 2021. Normalizing Flows: An Introduction and Review of Current Methods. IEEE Trans. Pattern Anal. Mach. Intell. 43, 3964–3979. https://doi.org/10.1109/TPAMI.2020.2992934

Koivisto, M., Sood, K., 2004. Exact Bayesian structure discovery in Bayesian networks. J. Mach. Learn. Res. 5, 549–573.

Koller, D., Friedman, N., 2009. Probabilistic Graphical Models: Principles and Techniques, 1st ed, Adaptive Computation and Machine Learning series. The MIT Press.

Krefeld-Schwalb, A., Pachur, T., Scheibehenne, B., 2022. Structural parameter interdependencies in computational models of cognition. Psychol. Rev. 129, 313–339. https://doi.org/10.1037/rev0000285

Kullback, S., Leibler, R.A., 1951. On information and sufficiency. Ann. Math. Stat. 22, 79–86.

Kwakkel, J.H., Pruyt, E., 2013. Exploratory Modeling and Analysis, an approach for model-based foresight under deep uncertainty. Technol. Forecast. Soc. Change, Future-Oriented Technology Analysis 80, 419–431. https://doi.org/10.1016/j.techfore.2012.10.005

Lane, D.C., 1999. Social theory and system dynamics practice. Eur. J. Oper. Res. 113, 501–527.

Lauritzen, S.L., 1995. The EM algorithm for graphical association models with missing data. Comput. Stat. Data Anal. 19, 191–201.

Lee, T.-P., Zagonel, A.A., Andersen, D.F., Rohrbaugh, J., Richardson, G.P., 1998. A Judgment Approach to Estimating Parameters in Group Model Building: A Case Study of Social Welfare Reform at Dutchess County, in: Proceedings of the 16th International Conference of the System Dynamics Society.

Lewis, D., 1973. Causation. J. Philos. 70, 556–567. https://doi.org/10.2307/2025310

Li, P., Vu, Q.D., 2013. Identification of parameter correlations for parameter estimation in dynamic biological models. BMC Syst. Biol. 7, 91. https://doi.org/10.1186/1752-0509-7-91

Liu, M., Stella, F., Hommersom, A., Lucas, P.J.F., Boer, L., Bischoff, E., 2019. A comparison between discrete and continuous time Bayesian networks in learning from clinical time series data with irregularity. Artif. Intell. Med. 95, 104–117. https://doi.org/10.1016/j.artmed.2018.10.002

Liu, Y., Ma, X., Qiao, W., Han, B., 2024. A methodology to model the evolution of system resilience for Arctic shipping from the perspective of complexity. Marit. Policy Manag. 51, 867–882. https://doi.org/10.1080/03088839.2023.2187472

Luengo, J.C., Ramos-López, D., Rumí, R., 2025. Modeling continuous distributions in hybrid Bayesian networks using mixtures of polynomials with tails. Comput. Stat. Data Anal. 212, 108246. https://doi.org/10.1016/j.csda.2025.108246

Lyneis, J.M., 2000. System dynamics for market forecasting and structural analysis. Syst. Dyn. Rev. 16, 3–25.

Madsen, A.L., 2008. Belief update in CLG Bayesian networks with lazy propagation. Int. J. Approx. Reason., Special Section on Probabilistic Rough Sets and Special Section on PGM'06 49, 503–521. https://doi.org/10.1016/j.ijar.2008.05.001

Marcot, B.G., 2012. Metrics for evaluating performance and uncertainty of Bayesian network models. Ecol. Model. 230, 50–62. https://doi.org/10.1016/j.ecolmodel.2012.01.013

Mathieu, E., Ritchie, H., Rodés-Guirao, L., Appel, C., Gavrilov, D., Giattino, C., Hasell, J., Macdonald, B., Dattani, S., Beltekian, D., 2020. Coronavirus (COVID-19) vaccinations. Our World Data.

McGeachie, M.J., Chang, H.-H., Weiss, S.T., 2014. CGBayesNets: Conditional Gaussian Bayesian Network Learning and Inference with Mixed Discrete and Continuous Data. PLoS Comput. Biol. 10, e1003676. https://doi.org/10.1371/journal.pcbi.1003676

McLucas, A.C., 2003. Incorporating soft variables into system dynamics models: a suggested method and basis for ongoing research, in: Proceedings of the 21st International Conference of the System Dynamics Society. pp. 20–24.

McNaught, K., 2003. Influences and connections between system dynamics and decision analysis, in: Proceedings of the 21st International Conference of the System Dynamics Society. Citeseer.

Meadows, D.H., 1980. The unavoidable a priori. Elem. Syst. Dyn. Method 23–57.

Menzies, P., Price, H., 1993. Causation as a Secondary Quality. Br. J. Philos. Sci. 44, 187–203. https://doi.org/10.1093/bjps/44.2.187

Merkhofer, M.W., 1987. Quantifying judgmental uncertainty: Methodology, experiences, and insights. IEEE Trans. Syst. Man Cybern. 17, 741–752. https://doi.org/10.1109/TSMC.1987.6499281

Mihajlovic, V., Petkovic, M., 2001. Dynamic bayesian networks: A state of the art.

Mingers, J., 2006. A critique of statistical modelling in management science from a critical realist perspective: its role within multimethodology. J. Oper. Res. Soc. 57, 202–219.

Mingers, J., Brocklesby, J., 1997. Multimethodology: Towards a framework for mixing methodologies. Omega 25, 489–509. https://doi.org/10.1016/S0305-0483(97)00018-2

Mises, R. von, 2014. Mathematical Theory of Probability and Statistics. Academic Press.

Mohaghegh, Z., Kazemi, R., Mosleh, A., 2009. Incorporating organizational factors into Probabilistic Risk Assessment (PRA) of complex socio-technical systems: A hybrid technique formalization. Reliab. Eng. Syst. Saf. 94, 1000–1018. https://doi.org/10.1016/j.ress.2008.11.006

Moore, C.M., 1987. Group techniques for idea building. Sage Publications, Inc.

Moral, S., Cano, A., Gómez-Olmedo, M., 2021. Computation of Kullback–Leibler Divergence in Bayesian Networks. Entropy 23, 1122. https://doi.org/10.3390/e23091122

Moral, S., Rumi, R., Salmerón, A., 2001. Mixtures of Truncated Exponentials in Hybrid Bayesian Networks, in: Benferhat, S., Besnard, P. (Eds.), Symbolic and Quantitative Approaches to Reasoning with Uncertainty. Springer, Berlin, Heidelberg, pp. 156–167. https://doi.org/10.1007/3-540-44652-4_15

Morecroft, J.D., 1982. A critical review of diagramming tools for conceptualizing feedback system models. Dynamica 8, 20–29.

Morgan, J.S., 2013. Exploring frameworks for mixing Discrete Event Simulation and System Dynamics methods in theory and in practice (Ph.D.). University of Strathclyde.

Morgan, J.S., Howick, S., Belton, V., 2017. A toolkit of designs for mixing Discrete Event Simulation and System Dynamics. Eur. J. Oper. Res. 257, 907–918. https://doi.org/10.1016/j.ejor.2016.08.016

Murphy, K.P., 2002. Dynamic bayesian networks. Probabilistic Graph. Models M Jordan 7, 431.

Nelsen, R.B., 2006. An Introduction to Copulas, Second Edition. ed, Springer Series in Statistics. Springer New York, New York, NY. https://doi.org/10.1007/0-387-28678-0

Nguyen, L.K.N., 2022. Hybrid health systems simulation modelling: controlling Covid-19 infections in care homes (http://purl.org/coar/resource_type/c_db06). University of Strathclyde.

Niu, H., Ma, C., Han, P., Li, S., Ma, Q., Zhang, S., 2020. Research on Safety Modeling and Analysis to the Task Process of Airborne Weather System, in: 2020 IEEE 20th International Conference on Software Quality, Reliability and Security Companion (QRS-C). Presented at the 2020 IEEE 20th International Conference on Software Quality, Reliability and Security Companion (QRS-C), pp. 317–323. https://doi.org/10.1109/QRS-C51114.2020.00062

Nordhaus, W.D., 1973. World Dynamics: Measurement Without Data. Econ. J. 83, 1156–1183. https://doi.org/10.2307/2230846

O'Hagan, A., Buck, C.E., Daneshkhah, A., Eiser, J.R., Garthwaite, P.H., Jenkinson, D.J., Oakley, J.E., Rakow, T., 2006. Uncertain judgements: eliciting experts' probabilities.

Oliva, R., 2003. Model calibration as a testing strategy for system dynamics models. Eur. J. Oper. Res. 151, 552–568. https://doi.org/10.1016/S0377-2217(02)00622-7

Oliva, R., 1996. A dynamic theory of service delivery: Implications for managing service quality (PhD Thesis). Massachusetts Institute of Technology.

Organisation for Economic Co-operation and Development, 2024. Hospital Beds.

Osgood, N., Liu, J., 2015. Combining Markov chain Monte Carlo approaches and dynamic modeling, in: Analytical Methods for Dynamic Modelers. The MIT Press, pp. 125–169.

Papamakarios, G., Nalisnick, E., Rezende, D.J., Mohamed, S., Lakshminarayanan, B., 2021. Normalizing flows for probabilistic modeling and inference. J Mach Learn Res 22, 57:2617-57:2680.

Pearl, J., 1985. Bayesian networks: A model of self-activated memory for evidential reasoning, in: Proceedings of the 7th Conference of the Cognitive Science Society, University of California, Irvine, CA, USA. pp. 15–17.

Popper, K., 2005. The logic of scientific discovery. Routledge.

Popper, K.R., 1990. A world of propensities. Thoemmes, Bristol.

Popper, K.R., 1959. The propensity interpretation of probability. Br. J. Philos. Sci. 10, 25–42. https://doi.org/10.1093/bjps/X.37.25

Prem, K., Zandvoort, K. van, Klepac, P., Eggo, R.M., Davies, N.G., Group, C. for the M.M. of I.D.C.-19 W., Cook, A.R., Jit, M., 2021. Projecting contact matrices in 177 geographical regions: An update and comparison with empirical data for the COVID-19 era. PLOS Comput. Biol. 17, e1009098. https://doi.org/10.1371/journal.pcbi.1009098

Price, H., 1992a. Agency and Causal Asymmetry. Mind 101, 501–520. https://doi.org/10.1093/mind/101.403.501

Price, H., 1992b. The Direction of Causation: Ramsey's Ultimate Contingency. PSA Proc. Bienn. Meet. Philos. Sci. Assoc. 1992, 253–267.

Price, H., 1991. Agency and Probabilistic Causality. Br. J. Philos. Sci. 42, 157–176. https://doi.org/10.1093/bjps/42.2.157

Pruyt, E., 2016. Integrating Systems Modelling and Data Science: The Joint Future of Simulation and 'Big Data'Science. Int. J. Syst. Dyn. Appl. IJSDA 5, 1–16.

Punyamurthula, S., Badurdeen, F., 2018. Assessing Production Line Risk using Bayesian Belief Networks and System Dynamics. Procedia Manuf. 26, 76–86. https://doi.org/10.1016/j.promfg.2018.07.010

Quigley, J., Walls, L., 2021. Characteristics of a Process for Subjective Probability Elicitation. Expert Judgement Risk Decis. Anal. 287–318.

Quigley, J.L., Bedford, T.J., Walls, L.A., 2008. Prior distribution elicitation, in: Ruggeri, F., Kenett, R., Faltin, F. (Eds.), Encyclopaedia of Statistics in Quality and Reliability. John Wiley & Sons Inc.

Rahmandad, H., Lim, T.Y., Sterman, J., 2021. Behavioral dynamics of COVID-19: estimating underreporting, multiple waves, and adherence fatigue across 92 nations. Syst. Dyn. Rev. 37, 5–31. https://doi.org/10.1002/sdr.1673

Ramsey, F.P., 1926. Truth and probability, in: Readings in Formal Epistemology: Sourcebook. Springer, pp. 21–45.

Ramsey, J., Zhang, J., Spirtes, P., 2006. Adjacency-Faithfulness and Conservative Causal Inference, in: Dechter, R., Richardson, T. (Eds.), Proceedings of the Twenty-Second Conference Conference on Uncertainty in Artificial Intelligence (2006). AUAI Press, pp. 401–408.

Randers, J., 1980. Guidelines for model conceptualization. Elem. Syst. Dyn. Method 117, 139.

Reichenbach, H., 1971. The Theory of Probability. University of California Press.

Richardson, G.P., Pugh, A.L., 1981. Introduction to system dynamics modeling with DYNAMO. Productivity Press Inc.

Rios-Ocampo, J., Gary, M.S., 2022. Using analytical equations to represent nonlinear relationships. Syst. Dyn. Rev. 38, 354–370. https://doi.org/10.1002/sdr.1718

Rodriguez-Ulloa, R., 2018. Systemic methodology for risks evaluation and management in the energy and mining sectors (SYSMEREM-EMS) using bayesian networks. J. Decis. Syst. 27, 191–200. https://doi.org/10.1080/12460125.2018.1468157

Rouwette, E., Vennix, J., van Mullekom, T., 2002. Group model building effectiveness: a review of assessment studies. Syst. Dyn. Rev. https://doi.org/10.1002/sdr.229

Rubin, D.B., 1984. Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician. Ann. Stat. 12, 1151–1172. https://doi.org/10.1214/aos/1176346785

Ruppert, D., 2011. Copulas, in: Statistics and Data Analysis for Financial Engineering, Springer Texts in Statistics. Springer New York, New York, NY, pp. 175–200. https://doi.org/10.1007/978-1-4419-7787-8_8

Russell, B., 1912. On the Notion of Cause. Proc. Aristot. Soc. 13, 1–26.

Salmon, W.C., 1998. Causality and explanation. Oxford University Press.

Salmon, W.C., 1984. Scientific explanation and the causal structure of the world. Princeton University Press.

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., Tarantola, S., 2010. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Comput. Phys. Commun. 181, 259–270. https://doi.org/10.1016/j.cpc.2009.09.018

Sari, D.P., Rosadi, D., Effendie, A.R., Danardono, D., 2021. Discretization methods for Bayesian networks in the case of the earthquake. Bull. Electr. Eng. Inform. 10, 299–307. https://doi.org/10.11591/eei.v10i1.2007

Scanagatta, M., Salmerón, A., Stella, F., 2019. A survey on Bayesian network structure learning from data. Prog. Artif. Intell. 8, 425–439. https://doi.org/10.1007/s13748-019-00194-y

Schoenberg, W., 2020. Inferring Causality in Feedback Systems by Combining Neural Nets and Ordinary Differential Equations. Presented at the 38th International Conference of the System Dynamics Society.

Schoenberg, W., Davidsen, P., Eberlein, R., 2020. Understanding model behavior using the Loops that Matter method. Syst. Dyn. Rev. 36, 158–190. https://doi.org/10.1002/sdr.1658

Scott, R., 2018. Group model building: using systems dynamics to achieve enduring agreement, 1st ed. 2018.. ed, SpringerBriefs in operations research. Springer, Singapore.

Scutari, M., 2020. Bayesian network models for incomplete and dynamic data. Stat. Neerlandica 74, 397–419. https://doi.org/10.1111/stan.12197

Shenoy, P.P., West, J.C., 2009. Mixtures of polynomials in hybrid Bayesian networks with deterministic variables, in: Proceedings of the Eighth Workshop on Uncertainty Processing. pp. 202–212.

Sigurdsson, J.H., Walls, L.A., Quigley, J.L., 2001. Bayesian belief nets for managing expert judgement and modelling reliability. Qual. Reliab. Eng. Int. 17, 181–190. https://doi.org/10.1002/qre.410

Silander, T., Myllymaki, P., 2006. A simple approach for finding the globally optimal Bayesian network structure. arXiv. https://doi.org/10.48550/arXiv.1206.6875

Sklar, A., 1973. Random variables, joint distribution functions, and copulas. Kybernetika 9, 449–460.

Spetzler, C.S., Stael von Holstein, C.-A.S., 1975. Exceptional paper—Probability encoding in decision analysis. Manag. Sci. 22, 340–358.

Spirtes, P., Glymour, C., Scheines, R., 2001. Causation, Prediction, and Search. The MIT Press. https://doi.org/10.7551/mitpress/1754.001.0001

Spirtes, P., Meek, C., Richardson, T., 1995. Causal inference in the presence of latent variables and selection bias, in: Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence, UAI'95. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 499–506.

Steck, H., Jaakkola, T., 2002. On the Dirichlet Prior and Bayesian Regularization, in: Advances in Neural Information Processing Systems. MIT Press.

Stepanova, M., Lam, B., Younossi, E., Felix, S., Ziayee, M., Price, J., Pham, H., de Avila, L., Terra, K., Austin, P., Jeffers, T., Escheik, C., Golabi, P., Cable, R., Srishord, M., Venkatesan, C., Henry, L., Gerber, L., Younossi, Z.M., 2022. The impact of variants and vaccination on the mortality and resource utilization of hospitalized patients with COVID-19. BMC Infect. Dis. 22, 702. https://doi.org/10.1186/s12879-022-07657-z

Sterman, J., 2018. System dynamics at sixty: the path forward. Syst. Dyn. Rev. 34, 5–47.

Sterman, J., 2000. Business dynamics : systems thinking and modeling for a complex world. Boston : Irwin/McGraw-Hill, Boston.

Sterman, J.D., 1984. Appropriate summary statistics for evaluating the historical fit of system dynamics models. Dynamica 10, 51–66.

Struben, J., Sterman, J., Keith, D., 2015. Parameter estimation through maximum likelihood and bootstrapping methods, in: Analytical Methods for Dynamic Modelers. pp. 3–38.

Susnik, J., Molina, J., Vamvakeridou-lyroudia, L.S., Savic, D.A., Kapelan, Z., 2013. Comparative Analysis of System Dynamics and Object-Oriented Bayesian Networks Modelling for Water Systems Management. Water Resour. Manag. Dordr. 27, 819–841. http://dx.doi.org/10.1007/s11269-012-0217-8

Tan, S., Weinert, D., Joseph, P., Moinuddin, K.A.M., 2021. Incorporation of technical, human and organizational risks in a dynamic probabilistic fire risk model for high-rise residential buildings. Fire Mater. 45, 779–810. https://doi.org/10.1002/fam.2872

Taylor, H.F. (Henry F., 1999. Modeling paper material flows and recycling in the US macroeconomy (Thesis). Massachusetts Institute of Technology.

Trivedi, P.K., Zimmer, D.M., 2007. Copula modeling: an introduction for practitioners. Found. Trends® Econom. 1, 1–111.

Tsamardinos, I., Aliferis, C.F., Statnikov, A., 2003. Time and sample efficient discovery of Markov blankets and direct causal relations, in: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Presented at the KDD03: The Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, Washington, D.C., pp. 673–678. https://doi.org/10.1145/956750.956838

Tsamardinos, I., Brown, L.E., Aliferis, C.F., 2006. The max-min hill-climbing Bayesian network structure learning algorithm. Mach. Learn. 65, 31–78. https://doi.org/10.1007/s10994-006-6889-7

Urbanization by sovereign state, 2025. . Wikipedia.

Venn, J., 1888. The logic of chance, 3rd ed. MacMillan and Co., London.

Vennix, J., 1996. Group Model Building: Facilitating Team Learning Using System Dynamics, 1st edition. ed. Wiley, Chichester ; New York.

Vennix, J.A.M., 1999. Group model-building: tackling messy problems. Syst. Dyn. Rev. 15, 379–401. https://doi.org/10.1002/(SICI)1099-1727(199924)15:4%3C379::AID-SDR179%3E3.0.CO;2-E

Videira, N., Antunes, P., Santos, R., 2017. Participatory Modelling in Ecological Economics: Lessons from practice, in: Routledge Handbook of Ecological Economics. Routledge, pp. 362–371. https://doi.org/10.4324/9781315679747-43

Wakeland, W., Homer, J., 2020. Addressing Parameter Uncertainty in SD Models with Fit-to-history and Monte-Carlo Sensitivity Methods.

Wang, G., Wang, S., Kang, Q., Duan, H., Wang, X., 2016. An integrated model for simulating and diagnosing the water quality based on the system dynamics and Bayesian network. Water Sci. Technol. 74, 2639–2655. https://doi.org/10.2166/wst.2016.442

Warren, K., 2004. Why has feedback systems thinking struggled to influence strategy and policy formulation? Suggestive evidence, explanations and solutions. Syst. Res. Behav. Sci. Off. J. Int. Fed. Syst. Res. 21, 331–347.

Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks. Nature 393, 440–442. https://doi.org/10.1038/30918

Wendel, G., Martínez, L., Bojowald, M., 2020. Physical Implications of a Fundamental Period of Time. Phys. Rev. Lett. 124, 241301. https://doi.org/10.1103/PhysRevLett.124.241301

Werner, C., Bedford, T., Cooke, R.M., Hanea, A.M., Morales-Nápoles, O., 2017. Expert judgement for dependence in probabilistic modelling: A systematic literature review and future research directions. Eur. J. Oper. Res. 258, 801–819. https://doi.org/10.1016/j.ejor.2016.10.018

Williamson, J., 2005. Bayesian nets and causality: philosophical and computational foundations. Oxford University Press, New York.

Yan, H., Wood, R., 2017. A structural model for estimating losses associated with the mis-selling of retail banking products. J. Oper. Risk 12, 1–19.

Yücel, G., Barlas, Y., 2007. Pattern-based system design/optimization. Presented at the Proceedings of the 25th international conference of the system dynamics society, Boston.

Zellner, A., 1980. Information Sources for Modeling the National Economy: Comment. J. Am. Stat. Assoc. 75, 567–569. https://doi.org/10.2307/2287645

Zhao, W., Song, S., 2020. Covid-19 Database. Hereditas 42, 212–221. https://doi.org/10.16288/j.yczz.20-030

Zhou, W., O'Neill, E., Moncaster, A., Reiner, D.M., Guthrie, P., 2020. Forecasting urban residential stock turnover dynamics using system dynamics and Bayesian model averaging. Appl. Energy 275, 115388. https://doi.org/10.1016/j.apenergy.2020.115388

Appendix A Search Documentation

A1 Search Sessions

A focused narrative search was undertaken to identify studies combining System Dynamics (SD) and Bayesian networks (BN). Search was last executed on 8 August 2025 in Scopus and Google Scholar. Language was restricted to English, and no year limits were applied. In Scopus, title/abstract/keywords were searched using a query; in Google Scholar, results were sorted by relevance (using the default option) and top 200 records were screened due to diminishing yield. Scopus served as the primary source and Google Scholar as a supplementary source; when a record appeared in both, it was attributed to Scopus as the source of record. Exact queries, dates, filers, and stop rules are summarise in Table A.1.

Table A.1 Databases, queries, dates, filters, and stop rules

Platform	Exact query	Date run	Filters/limits	Stop rule	Note
Scopus	TITLE-ABS-KEY ("system dynamics") AND TITLE-ABS-KEY (Bayesian W/3 network*) AND (LIMIT-TO (LANGUAGE, "English"))	8 August 2025	English; no year limits	Exported all hits	"Bayesian W/3 network*" captures terms such as "Bayesian network", "Bayesian networks", and "Bayesian belief networks"
Google Scholar	System Dynamics Bayesian Networks	8 August 2025	English; sort=relevance	Screened top 200 results	Searches span full text where indexed; not restricted to title/abstract/keywords

A2 Screening Criteria and Coding

Two sets of codes were used to record primary reasons for exclusion at each stage (Table A.2 and Table A.3). T-codes apply at title/abstract (TA) screening; F-codes apply at full-text (FT) assessment. Reasons are not mutually exclusive, but for reporting (e.g., PRISMA counts), one primary reason was assigned per record. T1A and T2A are specific sub-categories of T1 and T2, respectively. Specifically, T2A is used for cases where the phrase "system dynamics" was used but not following the usage consistent with the discipline of System Dynamics (Forrester, 1961; Sterman, 2000).

Table A.2 Title and abstract screening codes (T-codes)

Code	Definition	Example/Notes
T1	Terms do not present in title or abstract.	Neither "system dynamics" nor "Bayesian network(s)" appears in title/abstract.
T1A	"Bayesian Neural Network(s)" matched instead of "Bayesian (Belief) Network(s)".	Sub-case of T1 used to separate false positives due to "neural".
T2	Out of scope/domain.	Topic unrelated after abstract review (wrong field/problem) – e.g., deep learning or physics.
T2A	"System dynamics" used in a sense not consistent with Forrester (1961) (e.g., generic "dynamics of systems").	Sub-case of T2 to flag terminology drift.
T3	Not a research paper.	Editorial or non-scholarly content.

Table A.3 Full-text assessment codes (F-codes)

Code	Definition	Example/Notes
F1	Only SD is substantively present; BN absent.	Paper uses SD modelling without BN.
F2	Only BN is substantively present; SD absent.	Paper builds BN but no SD modelling.
F3	Both terms appear, but only as a mention or with no substantial combination.	No integration of SD and BN in methods / results; literature reviews.
F4	Outside scope after full text.	Context/problem falls outside predefined scope – e.g., improving a Bayesian method.
F5	Full text unavailable despite reasonable attempts.	Unreachable through university's library and no payable options for access can be found.
F6	Duplicating or summarising another record already included.	Slides/review/same study restated without new evidence or further improvement.

A3 Flow Diagram

Figure A.1 summarises identification, deduplication, title/abstract screening, full-text assessment, snowballing, and final inclusion.

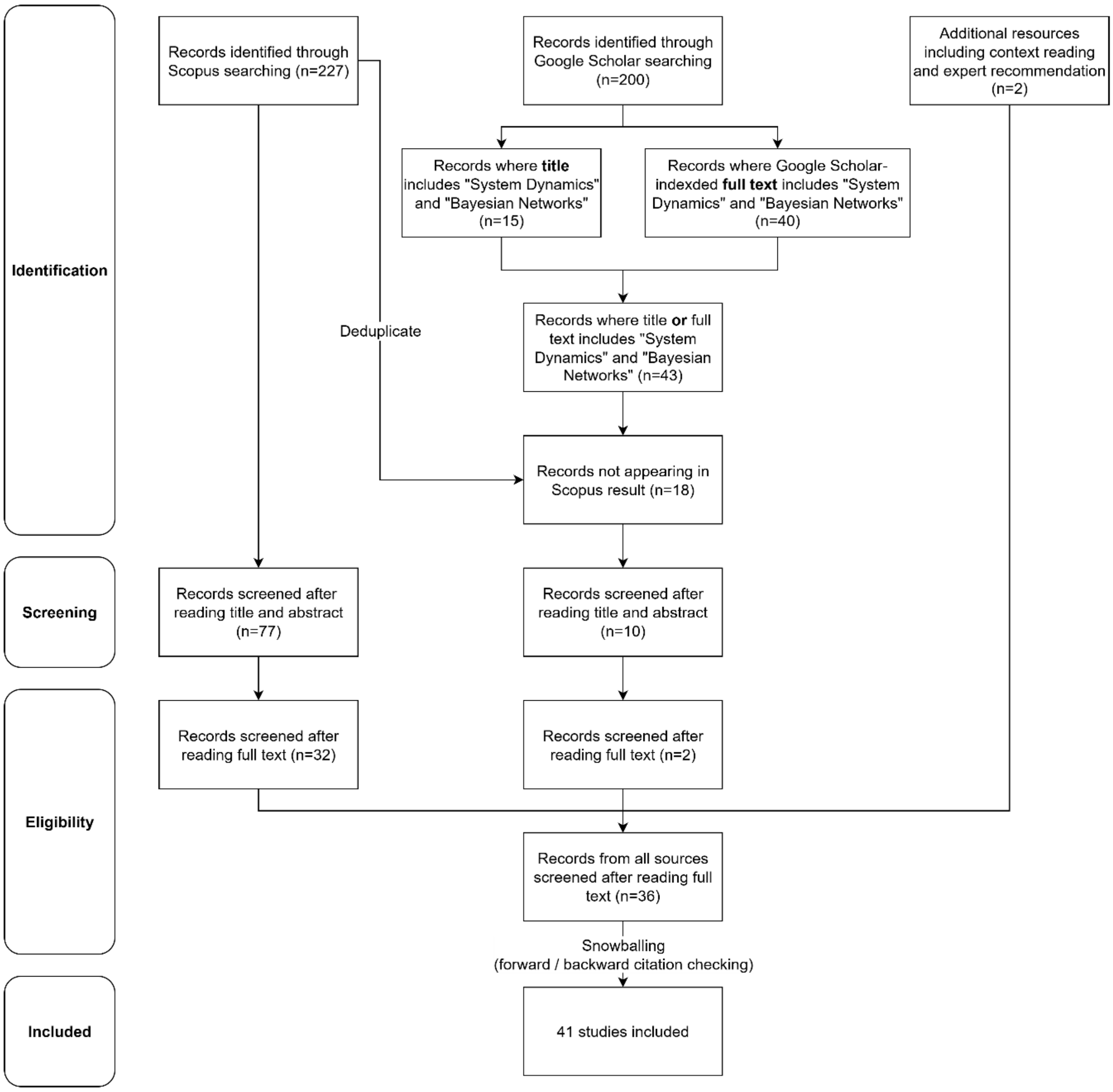


Figure A.1 Identification, deduplication, screening, full-text assessment, snowballing, and final inclusion

A4 Included Studies

As of 8 August 2025, the process identified 41 relevant studies meeting the inclusion criteria (combining SD and BN). Table A.4 lists the included records and their categorisation. Complete screening log is available on reasonable request from the author.

Table A.4 Included studies

Year	Authors	Title	How SD and BN Combine	Design	Identification Source
2025	Bai J.; Wang J.; Li X.	Interruption Risk Propagation and Resilience Evaluation of Supply Chain of Emergency Medical Supplies Under Information Sharing Mechanism	BN generates parameters values for SD simulation	Sequential	Scopus full result
2025	Zhang J.; Wu L.; Zhang T.; Wu D.	Systemic and dynamic risk analysis of drilling construction based on bayesian network and system dynamics model	BN outputs are used to initialise SD model	Sequential	Google Scholar top 200
2024	Dehghani S.; Massah Bavani A.; Roozbahani A.; Sahin O.	Assessment of Climate Change-Induced Water Scarcity Risk by Using a Coupled System Dynamics and Bayesian Network Modeling Approaches	SD generates input for BN update	Sequential	Scopus full result
2022	Baroroh I.; Ariana I.M.; Dinariyana A.A.B.	Risk Evaluation Concept of Engine Room Module Installation on Modular Construction Shipbuilding	BN is used to determine inputs to SD simulation	Sequential	Scopus full result
2022	Wu J.; Zhang L.; Bai Y.; Reniers G.	A safety investment optimization model for power grid enterprises based on System Dynamics and Bayesian network theory	Sensitivity analysis applied on BN to identify factors to inform SD model boundary	Sequential	Scopus full result
2021	Ginters E.; Revathy J.C.	Hidden and latent factors' influence on digital technology sustainability development	BN is used to analyse SD simulation outcome	Sequential	Scopus full result
2020	Niu H.; Ma C.; Han P.; Li S.; Ma Q.; Zhang S.	Research on Safety Modeling and Analysis to the Task Process of Airborne Weather System	SD generates input for BN update	Sequential	Papers citing Mohaghegh (2010)
2019	Bertone E.; Sahin O.; Richards R.; Roiko A.	Assessing the impacts of extreme weather events on potable water quality: The value to managers of a highly participatory, integrated modelling approach	BN generates parameters values for SD simulation	Sequential	Scopus full result
2019	Sun K.; Luxhøj J.T.	A hybrid approach for modelling dynamic flows and systemic risks in supply chains	DBN generates input to SD simulation	Sequential	Scopus full result
2018	Bertone E.; Sahin O.; Stewart R.A.; Zou P.X.W.; Alam M.;	Role of financial mechanisms for accelerating the rate of water and energy efficiency retrofits in Australian public	BN generates parameters values for SD simulation	Sequential	Scopus full result

Year	Authors	Title	How SD and BN Combine	Design	Identification Source
	Hampson K.; Blair E.	buildings: Hybrid Bayesian Network and System Dynamics modelling approach			
2018	Rodriguez-Ulloa R.	Systemic methodology for risks evaluation and management in the energy and mining sectors (SYSMEREM-EMS) using bayesian networks	CLD maps out critical sectors of the system for further sensitivity and scenario analysis using BN	Sequential	Scopus full result
2017	Crookes D.J.	Does a reduction in the price of rhino horn prevent poaching?	SD generates prior distributions that characterise BN nodes	Sequential	Scopus full result
2016	Wang G.; Wang S.; Kang Q.; Duan H.; Wang X.	An integrated model for simulating and diagnosing the water quality based on the system dynamics and Bayesian network	SD generates synthetic simulation data for learning a BN	Sequential	Scopus full result
2016	Sticha P.J.; Axelrad E.T.	Using dynamic models to support inferences of insider threat risk	SD generates synthetic simulation data for learning a BN	Sequential	Scopus full result
2016	Akinyemi O.O.; Adebiyi K.A.	Modelling uncertainty in runway safety intervention performance evaluation	BN is used to model an accident probability variable in an SD model	Sequential	Scopus full result
2016	Phan T.D.; Sahin O.; Smart J.C.R.	System dynamics and Bayesian network models for vulnerability and adaptation assessment of a coastal water supply and demand system	SD generates synthetic data for learning a BN	Sequential	Scopus full result
2015	Wang Y.F.; Li Y.L.; Zhang B.; Yan P.N.; Zhang L.	Quantitative Risk Analysis of Offshore Fire and Explosion Based on the Analysis of Human and Organizational Factors	BN models a probability, which is projected to the long-term using SD	Sequential	Scopus full result
2010	Oh E.H.; Deshmukh A.; Hastak M.	Vulnerability assessment of critical infrastructure, associated industries, and communities during extreme events	BN models probability of failure to use as input for SD simulation of time-varying vulnerability	Sequential	Scopus full result
2008	Farmani R.; Savic D.	An evolutionary Bayesian belief network-based methodology for adaptive water management	CLD is used to surface drivers, feedback, and delays which are later translated into a BN	Sequential	Scopus full result
2003	McNaught, K.	Influences and connections between system dynamics and decision analysis	BN generates parameters values for SD simulation	Sequential	Expert recommendation
2023	Brereton, C.	Children's environmental health in least developed countries: a modelling	SD and BN are independently developed for the same problem and	Parallel	Papers citing Mohaghegh (2010)

Year	Authors	Title	How SD and BN Combine	Design	Identification Source
		approach to support policy decisions (PhD Thesis)	comparisons drawn at fixed points		
2022	Chikkagoudar S.; Chatterjee S.; Bharadwaj R.; Ganguly A.; Kompella S.; Thorsen D.	Assurance by Design for Cyber-physical Data-driven Systems	SD and BN each models a different input for analysis in a later phase	Parallel	Papers citing Wang et al. 2015
2017	Yan H.; Wood R.M.	A structural model for estimating losses associated with the mis-selling of retail banking products	SD and BN each models a different input for analysis in a later phase	Parallel	Scopus full result
2013	Sušnik J.; Molina J.-L.; Vamvakeridou-Lyroudia L.S.; Savić D.A.; Kapelan Z.	Comparative Analysis of System Dynamics and Object-Oriented Bayesian Networks Modelling for Water Systems Management	SD and BN are independently developed for the same problem and comparisons drawn at fixed points	Parallel	Scopus full result
2012	Balaban M.A.; Banks C.M.; Sokolowski J.A.	Engaging m&s to characterize cause and effect patterns of us withdrawal	BN is built independently to validate part of causal structure of a CLD	Parallel	Scopus full result
2008	Gregoriades, A.	Human Error assessment in complex Socio-Technical systems—System Dynamics versus Bayesian Belief Network	SD and BN are independently developed for the same problem and comparisons drawn at fixed points	Parallel	Google Scholar top 200
2018	Punyamurthula S.; Badurdeen F.	Assessing Production Line Risk using Bayesian Belief Networks and System Dynamics	SD and BN interact as two parts of one simulation model	Interaction	Scopus full result
2017	Kazemi R.; Mosleh A.; Dierks M.	A Hybrid Methodology for Modeling Risk of Adverse Events in Complex Health-Care Settings	SD and BN interact as two parts of one simulation model	Interaction	Scopus full result
2017	Bui H.; Pence J.; Mohaghegh Z.; Reihani S.; Kee E.	Spatio-temporal socio-technical risk analysis methodology: An application in emergency response	SD and BN interact as two parts of one simulation model	Interaction	Papers citing Mohaghegh (2010)
2016	Wang F.; Ding L.; Love P.E.D.; Edwards D.J.	Modeling tunnel construction risk dynamics: Addressing the production versus protection problem	SD and BN interact as two parts of one simulation model	Interaction	Scopus full result
2010	Mohaghegh Z.	Combining system dynamics and Bayesian belief networks for socio-technical risk analysis	SD and BN interact as two parts of one simulation model	Interaction	Scopus full result
2009	Mohaghegh Z.; Kazemi R.; Mosleh A.	Incorporating organizational factors into Probabilistic Risk Assessment (PRA) of	SD and BN interact as two parts of one simulation model	Interaction	Scopus full result

Year	Authors	Title	How SD and BN Combine	Design	Identification Source
		complex socio-technical systems: A hybrid technique formalization			
2024	Liu Y.; Ma X.; Qiao W.; Han B.	A methodology to model the evolution of system resilience for Arctic shipping from the perspective of complexity	BN is mapped into SD structure	Enrichment	Scopus full result
2021	Hafezi M.; Stewart R.A.; Sahin O.; Giffin A.L.; Mackey B.	Evaluating coral reef ecosystem services outcomes from climate change adaptation strategies using integrative system dynamics	BN models nonlinear relationships within SD model structure	Enrichment	Scopus full result
2021	Tan S.; Weinert D.; Joseph P.; Moinuddin K.A.M.	Incorporation of technical, human and organizational risks in a dynamic probabilistic fire risk model for high-rise residential buildings	BN is embedded within SD model structure	Enrichment	Scopus full result
2021	Tan S.; Weinert D.; Joseph P.; Moinuddin K.	Sensitivity and uncertainty analyses of human and organizational risks in fire safety systems for high-rise residential buildings with probabilistic t-h-o-risk methodology	BN is embedded within SD model structure	Enrichment	Scopus full result
2020	Tan S.; Weinert D.; Joseph P.; Moinuddin K.	Impact of technical, human, and organizational risks on reliability of fire safety systems in high-rise residential buildings—applications of an integrated probabilistic risk assessment model	BN is embedded within SD model structure	Enrichment	Scopus full result
2020	Zhou W.; O'Neill E.; Moncaster A.; Reiner D.M.; Guthrie P.	Forecasting urban residential stock turnover dynamics using system dynamics and Bayesian model averaging	Bayesian update used to estimate SD model parameter values from empirical data	Enrichment	Scopus full result
2017	Hafezi, M.; Sahin, O.; Stewart, R.; Ware, D.; Mackey, B.	Participatory dynamic modelling approach for adaptation planning needs focusing on Small Islands Developing States	BN models nonlinear relationships within SD model structure	Enrichment	Papers citing Susnik et al. (2013)
2016	Bertone E.; Sahin O.; Richards R.; Roiko A.	Modelling with stakeholders: A systems approach for improved environmental decision making under great uncertainty	BN models nonlinear relationships within SD model structure	Enrichment	Scopus full result
2015	Osgood, N.; Liu, J.	Combining Markov chain Monte Carlo approaches and dynamic modeling	Bayesian update used to estimate SD model parameter values from empirical data	Enrichment	Contextual reading