Is Observed Data Adequate to Automate the Construction of Causal Models?

Wang Zhao[1], Susan Howick[2], John Quigley[2]

Abstract – System dynamics models are built on a set of cause-and-effect assumptions representing the understanding of the causal relationships in a real system. Can machines automatically build system dynamics models by inferring causal assumptions directly from observed data? In this paper, we provide a review of existing works on automated model building and critically discuss their strength and limits. Relating the question to recent studies of causal inference, we conclude that state-of-the-art techniques are still inadequate to automatically infer causal relations from observed data, that the inadequacy has to do with different understandings of causality, and that the relation between causal model and empirical data deserves a thorough consideration as we introduce more data science and artificial intelligence techniques into the system dynamics field.

Keywords – Causal inference, System dynamics, Observed data, World model

Introduction

This paper is centred on automated inference of System Dynamics models (SDMs) from observed data. Since entering the new millennium, there have been pioneering studies and practices trying to achieve this goal using techniques from statistics and machine learning. In order to carry out a comprehensive analysis of the benefits and shortcomings of these pieces of work, their underlying assumptions need to be investigated. This will include how each piece of work views system dynamics models and how it understands the modelling process that is needed. We first briefly revisit classic system dynamics literature to make clear what 'building a model' means for experienced modellers, then examine work on automated model building using two questions: how does the work understand the model building process and how does it replicate this process with machine-based algorithms.

System dynamics modelling and causal assumptions

"System dynamics is a computer-aided approach to policy analysis and design. With origins in servomechanisms engineering and management, the approach uses a perspective based on information feedback and circular causality to understand the dynamics of complex social system." (Richardson, 1991, p. 144) After initially created by Forrester (1961, 1958), the approach has been widely applied to business analysis and the study of complex social-ecological systems.

Of the utmost importance in system dynamics is the strong link between model structure and its behaviour (Sterman, 2000). System dynamics models (SDMs) are causal models, which means that they represent causal relationships in an individual system or population (Hitchcock, 2019). In a system dynamics modelling process, modellers retrieve these causal relationships from their observation and interaction with the real system either by themselves or through stakeholders who supposedly know the situation better (Sterman, 2000). A causal

[1] Ph.D. student, Department of Management Science, University of Strathclyde, Glasgow.
Email: wang.zhao@strath.ac.uk
[2] Professor, Department of Management Science, University of Strathclyde, Glasgow

relationship could be as simple as ‘more equipment defects will lead to a higher breakdown rate’(Sterman, 2000, p. 69). A sufficient set of such relationships would be able to characterise the system of interest within a certain boundary. The characterisation not only clarifies the elements which the system consists of, but also demonstrates how these elements influence each other. The set of relationships therefore explain the mechanisms underlying a system through ‘telling a story’ or ‘explaining a theory’ (Forrester, 1994). SDMs are therefore called ‘theory-like’ models (Barlas and Carpenter, 1990).

In system dynamics literature, readers often encounter terms like ‘assumption’, ‘hypothesis’, and ‘theory’. These terms relate to the causal relationships extracted from the real system. For example, Forrester (1994) used ‘hypothesis (a theory)’ to indicate what is generated from describing a system: “…the relevant system must be described and a hypothesis (theory) generated for how the system is creating the troubling behaviour” (p.246). In the same work, by writing “[n]o one in the company differed with the assumptions in the model…” (p.254) Forrester referred to ‘assumption’ as something underlying a system dynamics model. It is arguable that although there are slight differences in the meaning of ‘assumption’, ‘hypothesis’, and ‘theory’, they are interchangeable in a system dynamics context. In order to avoid confusing the reader with too many almost synonymous terms, in this paper we use ‘causal assumption’ to refer to our understanding of the causal relationships in the system, with which we start to create visible model structures such as stocks and flows.

Causal assumptions are part of a ‘mental model’. “In system dynamics, the term ‘mental model’ includes our beliefs about the networks of causes and effects that describe how a system operates, along with the boundary of the model (which variables are included and which are excluded) and the time horizon we consider relevant – our framing or articulation of a problem.” (Sterman, 2000, p. 16) A mental model is one’s inner perception of a real system. A mental model is subjective and could vary from person to person. As shown in Figure 1, while the process ‘perceiving a real system’ internalises the social reality, the process ‘translating mental models into formal models’ externalises mental models into a concrete and visible form that enables interpersonal communication and comparison.

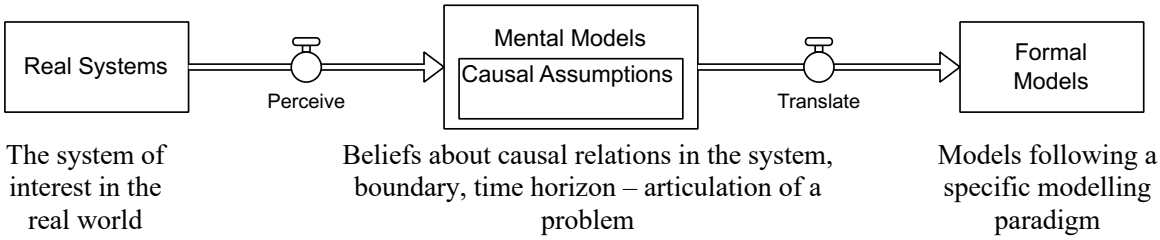

Figure 1. From mental model to formal model

Information sources for system dynamics modelling

Forrester (1980) considered mental data as the most important source of information used in system dynamics modelling process, which includes “observations about structure and policies, expectations about system behaviour, and actual observed system behaviour” (Forrester, 1980, p. 556). Although slightly different from Sterman (2000)’s definition of mental model, it is arguable that they both include individual understanding of the causal relationships in a system. Beside mental data, the other two sources are written data and numerical data.

Written data includes both 'recording of information from the mental store' and 'concepts and abstractions that interpret other information sources' (Forrester, 1980, p. 557). Compared with mental data and numerical data, this source of information is less relevant to our discussion for the time being as practices in automated modelling have not used information in text format so far. However, there have been studies on inferring causal relations from raw text (Tirunagari et al., 2012; Kim and Jun, 2015), so we may expect to see studies trying to build system dynamics models by text mining in the future.

Numerical data is key to our discussion as it has been the main source of information used in automated model building. Numerical data includes 'specific information on some parameter values', 'numerical summary of typical characteristics of economic behaviour', and 'time series data' (Forrester, 1980, p. 558). However, Forrester holds a conservative attitude toward numerical data as it tells us something about the system but misses "direct evidence of structure and policies that created the data", and "do not reveal the cause-to-effect direction between variables" (Forrester, 1980, p. 558). Regarding how numerical data should be used, Forrester stressed that in system dynamics modelling, "time series data are used much less as the basis for parameter values than in econometric models", but "the model itself generates synthetic output time series data that can be compared in a variety of ways with the real time series data" (Forrester, 1980, p. 558). In other words, use of numerical data might add to the validity of a model, but the mental data base is still the main source of information for modelling.

The comparatively little importance of numerical data is also due to the purpose of studying a model's behaviour in this field. As Meadows (1980, p. 31) wrote, "system dynamicists are generally unconcerned with precise numerical values of system variables in specific years. They are more interested in general dynamic tendencies; whether the system as a whole is stable or unstable, oscillating, growing, declining, or in equilibrium". Obviously, system dynamicists believe that a model should be more correct in replicating the system's structure than precise in reproducing the system's behaviour, and the study of model behaviour should focus more on the behaviour's pattern instead of its precision (Barlas, 1989, 1996). In doing so, it is more important to extract causal assumptions from stakeholders' mental models than focusing on observed numerical data.

Automated model building as a part of SD × Data

Over recent decades there has been a closer connection between system dynamics and data. This is mainly due to three factors. First, there is a constant search for better tools in the system dynamics field. Innovations in implementing computer-based algorithms to automate tasks that were previously only doable (or not even doable) by humans has taken place in different stages of the modelling process. For example, pattern recognition algorithms are used to categorise dynamic patterns observed from time-series data to facilitate model validation (Barlas and Kanar, 1999), calibration and behaviour analysis (Yücel and Barlas, 2015). Techniques from data science such as sampling and clustering are applied to facilitate data pre-processing before modelling, and help with analysis of model-generated big data (Pruyt et al., 2014). Using an algorithm that dynamically calculates impact scores for causal links, Schoenberg et al. (2019) proposed a method to analyse feedback loop dominance over a model's simulation. These innovative tools facilitate the modelling process and encourage researchers and practitioners to improve the tools further.

Second, in the era of big data, modelling projects often include a data set much bigger than was previously possible. Traditional modelling paradigms such as system dynamics, agent-based modelling, and econometrics are all challenged to make use of big data. Different from the aforementioned classic literature where model-based consultancy is centred on mental data, consulting projects now require system dynamics models to be more (numerical) data-driven and able to interpret a big, heterogeneous, and often ever-changing (dynamic) data set. Different modelling paradigms compete over capabilities of inferring from data, and some of them have innate advantage due to a close connection to statistical methods, for instance econometrics. An early comparison between system dynamics and econometrics by Meadows (1980) shows that compared with system dynamics, econometrics modelling is strong in estimating parameters from observed data. This advantage is supposedly more obvious when it comes to big data sets.

Third, the fast development in artificial intelligence in recent years encourages researchers to apply methods in machine learning and data science to traditional tasks in various disciplines outside the machine learning field. They do this mainly by translating tasks from different fields into tasks of machine learning, a process called 'problem casting', then solve the tasks using methods such as searching or optimisation. These attempts lead to cross-disciplinary research and often benefit both the 'method field' and 'problem field' through a dialogue between them.

Pruyt et al. (2014) offered a detailed analysis of incorporating system dynamics with big data-related methods. They observed three ways "in which big data and data science may play a role in SD: (1) to obtain useful inputs and information from (big) data, (2) to infer plausible theories and model structures from (big) data, and (3) to analyse and interpret model-generated "brute force data" (Pruyt et al., 2014, p. 1). Their analysis focused mainly on (1) and (3), but less on "inferring plausible theories and model structures from (big) data", which this paper focusses on. Zhao (2019) analysed works in this area spanning early 2000s to 2018 and found that the state-of-the-art approaches to "inferring theories and model structures from data" all take time-series data as the main (if not only) source of information. This means that the methods are almost end-to-end: with time-series data fed into the algorithm as an input and a model structure expected as the output.

Different methods used for inferring structures lead to different representations of outcome (Zhao, 2019). Drobek et al. (2014) used Pearson product-moment correlation coefficient to estimate the likelihood of causation between any pair of variables and represented the inferred structure as a graph network whose nodes are variables in a system and edges are assigned with correlation coefficients. Drobek et al. (2015) used artificial neural networks to approximate the deterministic structural equations between variables. The result was a set of neural networks that each takes input from other networks and feeds its own output to other networks. Abdelbari and Shafi (2016, 2017, 2018) used echo-state neural networks to identify cyclic causal relationships (i.e. feedback loops) from time-series data from variables in a system, and represent the outcome in causal loop diagrams. Chen et al. (2011) used a genetic algorithm to construct and test a large number of structural equations to find the ones that best reproduce the historical dynamic behaviour of the system. This is a method similar to symbolic regression which makes Chen et al.'s (2011) outcome the only analytic one among works in the same area.

Although the above works are innovative in methods, their performance is not as promising. Most of the works are not able to produce useful model structures. When it comes to a complex

situation, for instance a system with many variables and multiple feedback loops, the algorithms often generate model structures we cannot interpret and "redundant causal relations still exist in great amount" (Zhao, 2019, p. 22).

A further concern is whether these works follow the basic assumptions underlying the system dynamics field. These assumptions include 'structure drives behaviour', the feedback theory, and an endogenous perspective (Meadows, 1980). They are important for a system dynamics work to be interpretable to, and subsequently useful for, stakeholders.

While lack of performance could be resolved through refinement of algorithms, questions about assumptions cannot. Based on learning from the classic literature and observation of the state-of-the-art studies for automated modelling, we will discuss whether the methods driven by observed data are adequate for building a causal model.

Data-based inference of causal model

A basic assumption in system dynamics is that a system's behaviour is driven by its causal structure (Sterman, 2000). Causal assumptions formulated from observing the real system are foundational to such a causal structure. To build causal models, the first challenge for machines will be to automatically formulate causal assumptions that lay foundations for causal models, in other words, to perform causal inference. This is not a new topic: pioneers in artificial intelligence have been pursuing the ability of machines to perform causal inference since the time of McCarthy (1963). For a system dynamics model built by machine to be useful it has to include causal assumptions of the system, otherwise the model would be a correlative analysis of a set of variables, as clearly noted by Drobek et al. (2014). However, for the time being, time-series data still constitutes the majority of – if not all – data used in these studies. This leads to a question that has been asked for centuries: can we infer causation from only observed data?

The short answer is no. Pearl (2019) reviewed the history of causality in academia and found that even the existence of 'causality' has been in debate for centuries. This scepticism about causality is rooted in the positivist thinking of traditional statisticians like Karl Pearson. They believed science is only a description of the existing and observable, and for description of observed patterns, correlation was a better descriptor than causation (Pearl, 2019). Practically, even if we believe the existence of causality, statistics only proves to be effective in discovering regularities (covariations, correlations) but not causation, and findings from statistical analysis are not cause-and-effect relationships (Pearl, 2009, p. 42).

Therefore, the previously mentioned works on automated model building perform only statistical analysis of time-series data instead of causal inference. They look for model structures represented in a system dynamics format that generate behaviour close enough to reference modes characterised by time-series data, without carrying causal assumptions of the real system. As stated by the authors of those works: their outcomes are only potential or possible causal relationships and demand further validation by humans (Abdelbari and Shafi, 2017; Drobek et al., 2014). In our previous discussion we mentioned that causal assumptions are revealed in mental models rather than numerical data. However, the mental data, which according to Forrester (1980) is the most important information source for building models, has not been used in existing studies. As a comparison, humans extract model structure from their mental models that include causal assumptions, while machines approximate model structure from observed historical behaviour, which does not include casual assumptions.

These approximations can therefore only guess the causal structures. We therefore believe that the state-of-the-art techniques are still inadequate to automatically build system dynamics models from observed data.

This reflects a somewhat common negligence when applying techniques from one field to another. Although it may seem straightforward to simplify a model building process to the generation of model structures and therefore see it as a task of search and optimisation, important assumptions in system dynamics are missed. Model structures should be relevant to causal processes in the real world, and fitness in behaviour pattern is not the only criterion for model validity - patterns in historical behaviour should be reproduced for the right reason. In contrast, work that integrates data science techniques with data pre-process or model analysis (as analysed in Pruyt et al. (2014)) is less likely to face this missing assumption, because their tasks are originally statistical and do not have to do with the model's causal assumptions.

Model-based causal inference of observed data

Our discussion so far has considered why contemporary work cannot infer causal models from observed data but has not addressed what the relationship between a model and data should be.

By using the concept 'causality', system dynamicists have been silently but critically walking away from the positivist philosophical paradigm. They believe that causation comes from peoples' perceptions and therefore is socially constructed and not objective. Accepting the existence of a single social reality and different individual interpretations of such a reality at the same time seems to fall into the trap of dualism. The former indicates an objective perspective while the latter indicates a subjective one and these are based on almost contradicting ontological and epistemological assumptions. However, instead of declaring that system dynamics has a shaky foundation in social theories, Lane (1999) accepted the possibility that subjective and objective assumptions could co-exist and proposed that system dynamics could potentially be such a method to make this co-existence practically possible. In other words, while we may not understand the mechanisms underlying a social system perfectly, we can each have our own mental model about the system. By putting our mental models into formal models such as SDMs, we can compare and contrast with each other and use new observations to update our models, which is a progressive learning process.[3]

The same shift in philosophical stance has also occurred in understanding causality. Different from the traditional refutation to causality, Pearl (2019) believed that we each can have our own causal model describing the relations between a set of variables. These causal models can be used to answer questions about interventions and counterfactuals. Intervention questions are such as 'what will happen if we do this', and counterfactual questions are such as 'what would have happened if we had behaved in a different way', both of which are difficult (if not impossible) to answer using statistical methods and observed data. Pearl (2019) also believed that, over time, we constantly update our own causal models with new observations of the real world, which is also a progressive learning process.

[3] Mingers (2006) resonated this thinking from the philosophical side and advocated a new philosophical paradigm – critical realism – to reconcile the divide between the subjective and objective ends on the philosophical spectrum.

We can easily find similarities between how Pearl sees causal models and how system dynamicists see system dynamics models. People first formulate untested causal assumptions from observing and interacting with the world. Once obtained, these assumptions help us to interpret what we see and imagine what may happen but are still subject to further modifications and updates.

This critical shift from objective causalities (directly observable from the objective world) to subjective causalities (opinion or belief about the object world by an individual) changes the task of 'causal inference' from extracting causal relationships from observed data into interpreting observed data with one's (causal) mental model. Similarly, Forrester believes that building a system dynamics model involves filtering observed data through one's internal knowledge to recall a model structure that could explain the input (Forrester, 1980, p. 560). Forrester and Pearl therefore converge on the relation between causal models and data, suggesting that ideas about how the world works should come before interpreting the observed data, instead of vice versa.

World model: a way to the future

One question still remains unanswered: if causal models cannot be inferred from observed data, but should be used to infer observed data, then where should causal models come from?

Until now, the most reliable way of gaining confidence in a causal relationship from observed data is through randomised controlled trials (RCT) (Rubin, 1974)[4]. However, this is not to suggest that a machine should autonomously design RCTs – it is too costly, and more importantly it is intervening in nature and therefore different from how a human perceives day-to-day causal relationships. It is easy to say that perceiving the world through causality is human nature, but when it comes to machines, it is still in debate whether they can replicate human nature.

Hume (1711-1776) wrote in his *An Enquiry concerning Human Understanding* that "knowledge about causes is never acquired through *a priori* reasoning, and always comes from our experience of finding that particular objects[5] are constantly associated with one another (Hume, 2010, p. 12)." Hume's words indicate that our sense of cause and effect comes from observing regularities – or covariances – and such 'experience' accumulates over time as knowledge *a priori*. In other words, Hume believes that humans do perceive causal relationships from empirical observations, which is the same kind of data as time-series.

However, there are two critical differences between how humans use empirical observation and how machines use observed data. First, humans have a long period of time to formulate their basic understanding of the world through exposure to various regularities and covariances in day-to-day life. Second, humans use their causal knowledge acquired *a priori* in some previous situations to perform causal inference in a new situation, which is a process of 'transfer learning', while machines do not have knowledge *a priori* and only face the new situation.

[4] However, in randomized controlled trials one would still have type 1 and type 2 errors. So it would never really conclude with certainty that there was a causal relationship.
[5] Editor's remark: "When Hume is discussing cause and effect, his word 'object' often covers events as well as things."

These two differences suggest that we may need to think about equipping machines with domain knowledge to enable them to perform causal inference. In Pearl (2018)'s new framework for causal inference, causal knowledge is built into machines as a 'world model' to help them answer questions about causality with the help of empirical data. The same discussion on a world model has occurred in the field of machine learning. Researchers have realised that a lack of common sense and causal knowledge of the world has hindered artificial intelligence from mastering higher level tasks – tasks that requires planning and reasoning rather than recognising speeches and shapes. It has therefore been proposed that a world model should be built into artificial intelligence to provide these types of knowledge (LeCun, 2018).

In this paper we have reviewed literature on the basic assumptions in system dynamics, literature on automated construction of system dynamics models, and literature on causal inference. We believe that observed data is not sufficient for extracting causal relationships, and therefore contemporary approaches relying only on statistical analysis of observed data are inadequate to build system dynamics models by themselves. We believe a shift from objective understanding of causality to subjective understanding of causality is critical to understand the relation between causal model and empirical data. Mental models should not be seen as something extracted from observed data, but as a part of accumulated domain knowledge that helps with the understanding of observed data.

For future research, we recommend that we focus less on extracting model structures from observed data and more on machines' formulation of mental models, probably through equipping them with a world model, mimicking the way in which a human learns by observing and interacting with the real world over years. This will be difficult, and existing assumptions under automated model building need to be fundamentally changed.

References

Abdelbari, H., Shafi, K., 2018. Learning structures of conceptual models from observed dynamics using evolutionary echo state networks. J. Artif. Intell. Soft Comput. Res. 8, 133–154.

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.

Abdelbari, H., Shafi, K., 2016. Optimising a constrained echo state network using evolutionary algorithms for learning mental models of complex dynamical systems. Presented at the 2016 International Joint Conference on Neural Networks (IJCNN), IEEE, pp. 4735–4742.

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., 1989. Multiple tests for validation of system dynamics type of simulation models. Eur. J. Oper. Res. 42, 59–87. https://doi.org/10.1016/0377-2217(89)90059-3

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.

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

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.

Forrester, J.W., 1994. System dynamics, systems thinking, and soft OR. Syst. Dyn. Rev. 10, 245–256.
Forrester, J.W., 1980. Information sources for modeling the national economy. J. Am. Stat. Assoc. 75, 555–566.
Forrester, J.W., 1961. Industrial dynamics. Cambridge, Mass. M.I.T. Press and Wiley, Cambridge, Mass.
Forrester, J.W., 1958. Industrial Dynamics. A major breakthrough for decision makers. Harv. Bus. Rev. 36, 37–66.
Hitchcock, C., 2019. Causal Models, in: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.
Hume, D., 2010. An Enquiry concerning Human Understanding. Jonathan Bennett.
Kim, J.-M., Jun, S., 2015. Graphical causal inference and copula regression model for apple keywords by text mining. Adv. Eng. Inform. 29, 918–929.
LeCun, Y., 2018. Learning world models: The next step towards AI. IJCAI KeyNote Speech.
McCarthy, J., 1963. Situations, Actions, and Causal Laws (No. AI-MEMO-2). STANFORD UNIV CA DEPT OF COMPUTER SCIENCE.
Meadows, D.H., 1980. The unavoidable a priori. Elem. Syst. Dyn. Method 23–57.
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.
Pearl, J., 2019. The book of why : the new science of cause and effect. Penguin, London.
Pearl, J., 2009. Causality : models, reasoning, and inference, 2nd ed.. ed. Cambridge, Cambridge.
Pruyt, E., Cunningham, S., Kwakkel, J., De Bruijn, J., 2014. From data-poor to data-rich: system dynamics in the era of big data. Presented at the 32nd International Conference of the System Dynamics Society, Delft, The Netherlands, 20-24 July 2014; Authors version, The System Dynamics Society.
Richardson, G.P., 1991. System dynamics: Simulation for policy analysis from a feedback perspective, in: Qualitative Simulation Modeling and Analysis. Springer, pp. 144–169.
Rubin, D.B., 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. J. Educ. Psychol. 66, 688.
Schoenberg, W., Davidsen, P., Eberlein, R., 2019. Understanding model behavior using loops that matter. ArXiv190811434 Phys.
Sterman, J., 2000. Business dynamics : systems thinking and modeling for a complex world. Boston : Irwin/McGraw-Hill, Boston.
Tirunagari, S., Hanninen, M., Stanhlberg, K., Kujala, P., 2012. Mining causal relations and concepts in maritime accidents investigation reports. Int. J. Innov. Res. Dev. 1, 548–566.
Yücel, G., Barlas, Y., 2015. Pattern recognition for model testing, calibration, and behavior analysis, in: Analytical Methods for Dynamic Modelers. pp. 173–206.
Zhao, W., 2019. Automated Model Conceptualization and Interactive Modeling Environment: A Software Prototype (Master's Thesis). The University of Bergen.