Proceedings of the 2020 Winter Simulation Conference
K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, eds.

CONCEPTUAL MODELS IN SIMULATION STUDIES: MAKING IT EXPLICIT

Pia Wilsdorf
Fiete Haack
Adelinde M. Uhrmacher

Institute for Visual and Analytic Computing
University of Rostock
Albert-Einstein-Str. 22
Rostock, 18059, GERMANY

ABSTRACT

Conceptual models play an important role in conducting simulation studies. A formal or at least explicit specification of conceptual models is key for effectively exploiting them during simulation studies and thereafter, for interpreting and reusing the simulation results. However, the perception of conceptual models varies strongly and with it possible means for specification. A broad definition of the conceptual model, i.e., as a loose collection of early-stage products of the simulation study, holds the potential to unify existing definitions, but also poses specific challenges for specification. To approach these challenges, without claiming to be exhaustive, we identify a set of products, which includes research question, data, and requirements, and define relations and properties of these products. Based on a cell biological case study and a prototypical implementation, we show how the formal structuring of the conceptual model assists in building a simulation model.

1 INTRODUCTION

The conceptual model plays an important role in conducting simulation studies, which is documented in various life cycle models (Balci 2012; Sargent 2013). However, no unanimous definition of conceptual model does exist for modeling and simulation (Fujimoto et al. 2017).

In other areas of computer science, and in particular in software engineering, widespread agreement about the definition and specification of conceptual models has been established. There the conceptual model refers to identifying the software structure, important operations and interfaces, before the technical realization and development of algorithms (Schewe and Thalheim 2005). Common methods for conceptualization are ER diagrams, Petri nets, state charts, flow charts, or algebras. Some of these approaches are sophisticated, formal modeling languages such as Petri nets, whereas others such as UML have a rather ambiguous semantics that is open to interpretation (Thalheim 2010). The decision on which method to use largely depends on the task at hand as well as experience and common practices.

In the modeling and simulation community, the perception of conceptual models varies strongly. Be it a rather narrow view that defines the conceptual model as an abstract model description (Nance 1994), or a wider interpretation as loosely-coupled construct that integrates a variety of different artifacts (Balci 2012), or even broader including the model's context (Robinson 2008a). Another discussion is how to specify the conceptual model and its parts. There the ideas range from informal specification using verbal narratives and sketches (Grimm et al. 2020) to formal conceptual modeling languages (Guizzardi and Wagner 2012).

As a result of this dispute, so far little support for conceptual modeling does exist. In contrast, in software engineering there is a long-standing research community for conceptual modeling (Embley and Thalheim 2012). The efforts of that community are paying off, e.g., in automatic code generation, or automatic regression testing and consistency checking. For simulation studies, thorough conceptual modeling

978-1-7281-9499-8/20/$31.00

can also enable automatic support of certain steps such as reuse and composition of models, or verification and validation (Balci et al. 2011). However, "developing an engineering discipline of conceptual modeling [in modeling and simulation] will require much better understanding of: 1. how to make conceptual models explicit and unambiguous [...], 2. the processes of conceptual modeling [..., as well as] 3. architectures and services for building conceptual models" (Fujimoto et al. 2017).

In this paper, we will focus on the first question, that is, how to make conceptual models, their parts, and their relations explicit. Our approach aims at unifying the heterogeneous definitions of "conceptual model" in modeling and simulation and the various ways of specification. The resulting, rich definition of the conceptual model allows for automatic exploitation in supporting the building and analysis of the corresponding simulation model due to incorporation of formal approaches, while maintaining flexibility and ease of use. We also focus on a specific type of simulation study, namely those where the building of a valid simulation model is in the focus of interest. Consequently, some parts of the conceptual model, such as input data or behavioral requirements, will play a key role in the conceptual model as we define it. Based on a cell biological case study and a prototypical wiki implementation, we will show how the formal and explicit definition of the conceptual model assists in building a simulation model, and later discuss how the collected information could be exploited.

2 PERSPECTIVES ON CONCEPTUAL MODELS IN MODELING AND SIMULATION

During a simulation study, the conceptual model assists the diverse problem analysis, model building and experimentation activities. This is reflected in life cycle models such as Balci (2012) and Sargent (2013), see Figure 1. If the ultimate goal of the simulation study is to obtain a valid simulation model from which meaningful conclusions can be drawn, these activities are closely intertwined and the conceptual model as well as the simulation model are repeatedly revised, changed, and validated.

However, despite its centrality for the modeling and simulation life cycle, no unanimous definition of the conceptual model has been agreed upon (Robinson et al. 2015). Figure 2 summarizes the standpoints of selected authors regarding the content of the conceptual model and its specification: whether it is interpreted having a more narrow or wider scope, and whether it is seen as a formal or informal construct. In the following, we will describe the work on conceptual models in more detail.

In early definitions, authors distinguished between the notions of conceptual model and communicative model (Nance 1994; Fishwick 1995). The conceptual model was seen as vague and ambiguous, and typically referred to concepts in the modelers mind, whereas the communicative model referred to an informal representation of these concepts.

Other authors concentrate on formal languages for specifying conceptual models, which may then be automatically transformed into a computerized model, e.g., based on Petri nets or DEVS, thereby obscuring the line between conceptual and computerized model (Guizzardi and Wagner 2012; Cetinkaya et al. 2012). Typically, these types of conceptual models are applicable only to specific types of problems, and the expressiveness is constrained by the underlying formalism.

Balci (2012) describes the conceptual model as a "repository of highlevel conceptual constructs and knowledge specified in a variety of communicative forms (e.g. animation, audio, chart, diagram, drawing, equation, graph, image, text, and video) intended to assist in the design of any type of large-scale complex M&S application." In Balci's definition these conceptual constructs all refer to the simulation model itself, and are kept separate from artifacts that refer to the context of the simulation model, i.e., the problem formulation, objectives, and requirements. Sargent (2013) provides a similar, however shorter, definition.

In contrast, Robinson (2008a), Robinson (2008b) subsume all these different artifacts under the term conceptual model, as all of them, including research questions, requirements, etc., are useful for conducting the simulation study. Robinson also identifies further contents of the conceptual model that should be made explicit: general project objectives regarding, e.g., visualization or simulation speed; model inputs, outputs, as well as used data; scope, level of detail, assumptions, and simplifications; entities, activities, and what modeling approaches to use; and finally justifications for each design choice. Pace (2000) identifies a

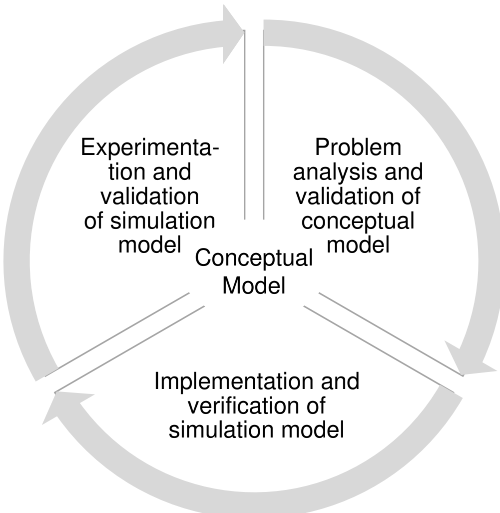


Figure 1: Roles of the conceptual model in the modeling and simulation life cycle, adapted from (Sargent 2013).

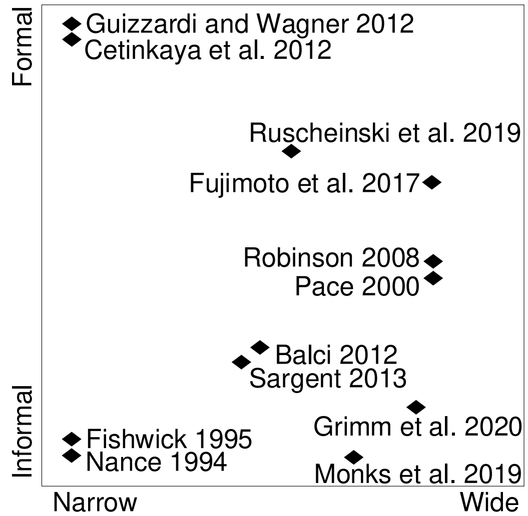


Figure 2: The spectrum of conceptual model definitions from narrow to wide and informal to formal.

similar list of important conceptual model parts. Both Robinson and Pace say that a partly formal approach will be needed, however, neither give concrete suggestions for the notation of conceptual models.

Fujimoto et al. 2017 defines the conceptual model as a collection of early-stage products that integrate and provide information and requirements for a variety of simulation study aspects. They advocate the development of more explicit and formal conceptual models based on domain-specific languages, ontologies, and other suitable knowledge representations. The problem, however, is that although many such formalisms exist, they are not accessible to domain experts, and integration into the usual workflows is difficult.

The work on reporting guidelines for simulation models faces similar challenges (Monks et al. 2019; Grimm et al. 2020). The ODD (Overview, Design concepts, and Details) protocol for example has recently been extended to the context of simulation studies and now includes purpose and requirements (called "patterns") of the model, state variables, processes, design concepts, initialization, input data, and submodels. However, reporting documents are typically created retrospectively to summarize the work done in a mainly verbal format, whereas the activity of conceptual modeling accompanies the whole modeling process.

Looking at the published definitions of the conceptual model in Figure 2, narrow definitions of the conceptual model use either formal or informal approaches exclusively. Wider definitions, on the other hand, use a wider variety of specification formats, and thus are often semi-formal. However, there is a lack of *wide and formal* approaches. In the artifact-based workflow by Ruscheinski et al. 2019 some of the discussed conceptual aspects have been integrated with special focus on formally defined requirements, thereby making this part of the conceptual model formally explicit and exploitable. However, it is a heavyweight approach which also states something about the process of designing a conceptual model and how to store it. We are aiming at a more lightweight approach for integrating the diverse parts of a conceptual model using formal methods where applicable.

3 MAKING EARLY-STAGE PRODUCTS EXPLICIT

The aim of our research is to integrate all the different conceptual aspects and to make them explicit. At the same time, we want our framework to explicitly support both formal and informal approaches, as both are crucial for the conceptual modeling phase. Our definition encompasses the main artifacts research objective *obj*, requirement *req*, approach *appr*, assumption *assum*, qualitative model *qmo*, and data or information source *src*. In this section these artifacts and the relations between them are further identified and formalized.

We define the conceptual model *cmo* as a tuple that comprises lists (denoted by *) of the different artifacts.

$$cmo = \langle obj^*, req^*, appr^*, assum^*, qmo^*, src^* \rangle$$

3.1 Research Objective

A simulation model is built for a system and an experiment to answer specific questions about the system (Cellier and Greifeneder 1991). These questions, i.e., the research objectives, determine when a suitable abstraction of the system of interest has been achieved. The objective *obj* is a tuple that integrates a number of descriptions of the overall objective and its context. This, typically informal, specification of the objective may be further substantiated by specific subgoals, called requirements. Thus, the objective contains a list of links (denoted by @) to requirement artifacts *req* that express what observed phenomena shall be reproduced by the simulation model. This approach is prominent in areas where the simulation models are built for explanation rather than prediction.

$$obj = \langle descr^*, @req^* \rangle$$

A description *descr* is defined as multimedia object consisting of a type (such as text, images, or video), a format and a tool for handling a description of that type, and finally the actual specification. Moreover, links to *src* artifacts allow to set the objective into context, e.g., if the current study builds on previous work. The description will be part of all the different ingredients of the conceptual model, and thus add flexibility by allowing the use of different communicative forms as proposed in Balci's definition of conceptual model (Balci 2012).

$$descr = \langle descrType, descrFormat, descrTool, descrSpec, @src^* \rangle$$
$$descrType = \text{'text'} \mid \text{'image'} \mid \text{'video'} \mid \ldots \mid \text{'other'}$$

3.2 Requirement

Requirements usually refer to the output and thus the behavior of the simulation model. If the expected behavior is expressed formally as a logic formula, e.g., the temporal logic MITL (Maler and Nickovic 2004), it can be checked automatically via statistical model checking (Agha and Palmskog 2018) in a tool like SESSL (Ewald and Uhrmacher 2014). The definition therefore contains, in addition to a general description, the *logicType*, a tool for interpreting the logic specification (*logicTool*), and the actual (*logicFormula*). Note, that we do not use a format here, since the syntax used to express logic formulas heavily relies on the implementation at hand. However, often format and tool will be redundant information, and only the file format would suffice. We, nonetheless, include both as sometimes the tool is crucial for exact reproduction of simulation results.

Another type of behavioral requirement is to reproduce data, i.e., the requirement is directly linked to a data source *src* that describes the intended behavior. The check, whether the output of the simulation model is sufficiently close to the expectation, is often done by applying face validation.

However, requirements might also be more fundamental, and refer, e.g., to the performance of the simulations, or to the choice of the modeling and simulation approach (Robinson 2008b). For example, if spatial resolution plays a role, the modeling and simulation approach should take space into account (Bittig and Uhrmacher 2010). If small numbers need to be considered, a stochastic approach based on the Gillespie algorithm (Gillespie 1977) might be more suitable than a deterministic one. Also, for some simulation studies the approach *appr* is determined from the project start due to other constraints.

Since links are bidirectional in our definition to allow tracing the origin as well as the usages of an artifact, the requirement is linked to the objective it refers to. In many simulation studies there will be only one objective, however, for larger projects multiple independent objectives could be formulated.

$$req = \langle descr^*, reqType, @obj^* \rangle$$
$$reqType = behavioralReq \mid @appr \mid \ldots$$
$$behavioralReq = \langle logicType, logicTool, logicFormula \rangle \mid @src \mid \ldots$$

3.3 Approach

The approach artifacts may refer to the modeling metaphor such as agent-based, population-based, reaction-based, or event-driven systems (Page 1994). They may refer to the system dynamics that are deemed most adequate such as discrete-event-based, discrete-stepwise, continuous or hybrid simulation (Zeigler et al. 1995). If the system at hand exhibits spatial behaviors, a suitable spatial modeling approach has to be selected. Sometimes, decisions about the modeling tools, like NetLogo for agent-based modeling and simulation (Tisue and Wilensky 2004), have to be made at the conceptual level. Similarly, analysis approaches such as experiment designs (Sanchez et al. 2018) have to be carefully selected beforehand. This is important also for simulation studies that use models for prediction.

In a hybrid modeling and simulation setting it is especially important to make explicit which parts of the simulation model each approach applies to. Therefore, the definition includes links to the different parts of the qualitative model, i.e., the submodels, parameters, entities, attributes, and interactions. Again, a description can be added to explicate and justify the choice of approach as well as the rationale behind it.

Since the reproducibility crisis of science, thinking about how to make the own research reproducible and well-documented has become an important task in the modeling and simulation community (Dalle 2012). Therefore, approaches used for traceability should be selected at the beginning of a simulation study to allow a documentation of the entire model building and experimentation process. Approaches for traceability include reporting guidelines like STRESS (Monks et al. 2019) and ODD (Grimm et al. 2020), scientific- (Oinn et al. 2004) and artifact-based workflows (Ruscheinski et al. 2019), and provenance (Ruscheinski and Uhrmacher 2017).

$$\begin{aligned} appr &= \langle descr^{*}, apprType, apprConcept, apprFormat, apprTool, @req, \\ &\quad @qmo^{*}, @ent^{*}, @att^{*}, @act^{*}, @param^{*} \rangle \\ apprType &= \text{'metaphor'} \mid \text{'dynamics'} \mid \text{'spatial'} \mid \text{'format'} \mid \text{'tool'} \mid \text{'analysis'} \mid \text{'traceability'} \mid \ldots \mid \text{'other'} \end{aligned}$$

Some of the specified approaches are related to requirements, i.e., they are mandatory with respect to a research objective or other project requirements. Most of these related requirements refer to the semantics of the simulation approach, e.g., if a continuous-time Markov chain is required, we still could use a stochastic process algebra, or a stochastic Petri net for modeling. In return, a specific format or tool may cover a multitude of these concepts. Thus, in the definition of an approach, we always combine the specification of a concept (here *apprConcept*) with a format or tool realizing it. To support the choice of approaches, formats, and tools, an ontological classification will be required that structures the general concepts of modeling and simulation (Guizzardi and Wagner 2012). Furthermore, uniquely identifying each concept with an ontology tag would facilitate an automatic exploitation of the conceptual model.

3.4 Assumption

While requirements look at the expectations we have referring to the simulation study, and thus typically focus on the output, the assumptions describe the starting point of a simulation study, its scope, and simplifications that are made. Assumption therefore determine how the simulation results might be interpreted. Assumptions *assum* can be stated more generally in the form of text, or specifically using mathematical formulas, which could be interpreted, e.g., by a specific tool. The description of an assumption artifact can provide further meaning and could include mathematical derivations. Assumptions may be derived from data, and thus links to data and information source artifacts can be included.

The formulated assumptions may be related directly to individual entities and causal relations of the system of interest. Therefore, the definition integrates links to the various parts of the qualitative model specification. For example, in a chemical model one could assume that the overall number of phosphorylated entities remains constant during the simulation, or one could make an assumption about the distribution of a parameter or attribute.

$$\begin{aligned} assum &= \langle descr^*, assumType, assumFormat, assumTool, assumSpec, @src^*, \\ &\quad @qmo^*, @ent^*, @att^*, @act^*, @param^* \rangle \\ assumType &= \text{'text'} \mid \text{'formula'} \mid \ldots \mid \text{'other'} \end{aligned}$$

3.5 Qualitative model

The qualitative model is what nearly all literature about conceptual models expect from it, i.e., a list of variables and the causal relations between those. The qualitative model *qmo* could be provided simply as a sketch, in a simple rule-based formalism, or as a Boolean model using suitable tools or file formats. In addition, we provide a means to make the important parts of the qualitative model such as entities, attributes, and interactions explicit. This allows us to reference them in other parts of the conceptual model, e.g., as part of assumptions and approaches, or map them to a specific data source. Our definition of the *qmo* follows the epistemological categories of systems (Klir and Rozehnal 1990).

The *qmo* may be structured into a set of submodels, e.g., for multilevel (Maus et al. 2011) or multiscale modeling (Hoekstra et al. 2019). Each submodel is also a *qmo*, i.e., a recursive definition is allowed. Entities *ent* are specified such that they may be either simple variables (i.e., able to hold a single value), or structured entities that may contain subentities and have attributes *att* (to be used in a compartment- or agent-based modeling metaphor). Thus, an entity is characterized by a name, a description, a potential list of subentities, and a list of attributes. Interactions *act* describe the dynamics of a system. An interaction refers to one or more entities, and model parameters *param* such as rate constants. For a more complex formalism that , e.g., supports the specification of the interaction type, stoichiometric matrices, or to make initial states explicit, a suitable approach, format, or tool can be chosen.

$$\begin{aligned} qmo &= \langle descr^*, qmoType, qmoFormat, qmoTool, qmoSpec, qmo^*, ent^*, act^*, param^*, @appr^*, @assum^*, @src^* \rangle \\ qmoType &= \text{'text'} \mid \text{'sketch'} \mid \text{'rule-based model'} \mid \text{'boolean model'} \mid \ldots \mid \text{'other'} \\ ent &= \langle name, descr^*, att^*, ent^*, @act^*, @appr^*, @assum^*, @src^* \rangle \\ att &= \langle name, descr^*, attType, @appr^*, @assum^*, @src^* \rangle \\ act &= \langle name, descr^*, @ent^*, @param^*, @appr^*, @assum^*, @src^* \rangle \\ param &= \langle name, descr^*, value, unit, @act^*, @appr^*, @assum^*, @src^* \rangle \end{aligned}$$

3.6 Data and information source

Data is central for simulation studies. Data can comprise input data, i.e., the input parameters of the model such as rate constants, initial concentrations, etc. Data can be used for validation or calibration of the simulation model, to establish a theory, or for illustration of a problem or process. Data may be experimental data, i.e., obtained from real measurements, or data generated by other simulations, so to support calibration or (cross-)validation.

But data that are helpful in conducting the simulation study may also be of a more general nature, and thus called information *src*. These can reference, e.g., scientific literature, or scientific notebooks (Kluyver et al. 2016). The description may provide additional information about the source, e.g., why a publication was selected, or how a data set was created.

Data or information sources can be linked to the other parts of the conceptual model, e.g., an approach, an assumption, or specific entities or parameters of the *qmo*. The sources thereby provide meaning to these other artifacts as outlined in the previous subsections. In particular, each source artifact can be resolved to a unique *srcID* such as a DOI or URL. Or, a local file can be uploaded, e.g., if data have just been produced in own wetlab experiments, and therefore have not been published yet.

To support the analysis or display of the data and information sources, we also include the specification of an appropriate file format *srcFormat* or tool *srcTool*. The choice of an appropriate tool for a specific file format may again be further supported by ontologies and other knowledge bases. However, the formatting of data or information is a topic of its own, e.g., how to express observational data and their semantics using standard formats.

$$\begin{aligned} src &= \langle descr^*, srcType, srcRole, srcIdType, srcIdSpec, srcFormat, srcTool, @obj^*, @req^*, @appr^*, @assum^*, \\ &\quad @qmo^*, @ent^*, @att^*, @act^*, @param^* \rangle \\ srcType &= \text{'experimental data'} \mid \text{'simulation data'} \mid \text{'literature'} \mid \text{'notebook'} \mid \ldots \mid \text{'other'} \\ srcRole &= \text{'input'} \mid \text{'validation'} \mid \text{'calibration'} \mid \text{'theory'} \mid \text{'illustration'} \mid \ldots \mid \text{'other'} \\ srcIdType &= \text{'DOI'} \mid \text{'URL'} \mid \text{'ISBN'} \mid \ldots \mid \text{'other'} \end{aligned}$$

4 CASE STUDY: MODELING THE WNT SIGNALING PATHWAY

We exemplarily realize the proposed formal definition of the conceptual model in a prototypical implementation based on *MediaWiki*. Using this wiki we build a conceptual model for a simulation study of WNT signaling in human neuronal progenitor cells during early differentiation, following the storyline of (Haack et al. 2015). We show how the conceptual model can be developed alongside the simulation model, and therefore is constantly revised and updated. Our wiki is publicly available for illustration and allows the interested reader to export its pages to set up their own wiki (CMoWiki 2020).

4.1 Implementation as a Wiki

Wikis are a popular means for collecting, structuring, and sharing knowledge of all kinds. *MediaWiki* is a free and open-source wiki engine which allows for simple web-based editing of information with a well-known markup language (https://www.mediawiki.org). Extensions provide additional features, e.g., *Semantic MediaWiki* enables semantic annotations (Krötzsch et al. 2006), and *Page Forms* allows for easy page creation and data insertion via forms.

First, after setting up a wiki named *CMoWiki*, the data structure has to be established. Therefore, we create categories for each artifact type and other substructures such as the *cmo*, *obj*, *req*, *descr*, etc. We then create properties for each individual input, e.g., *logicFormula* and *logicTool*, and define the input types, e.g., *Text*, *Image*, or *Page* for links to other artifacts. We thereafter create templates which relate inputs to the categories, and define the page layout. Based on the templates, we create forms, allowing users to edit pages in a graphical dialogue. After these preparatory steps, users can start creating pages and adding data. For example, the user can add a new page of the category *Conceptual Model* using the corresponding form. Figure 3 shows the main page of the *Wnt Simulation Study*, which contains individual sections for the different artifacts of the conceptual model. In the first section, an objective called *Main Objective* has been created, and further specified on the corresponding page, i.e., by adding a description. Also, a link to a requirement has been added to refine the objective. Since in this snapshot the requirement page of *Req1* has not been specified yet, the link is presented in red. By clicking on the red link, the modeler is directed to the *Create Requirement: Req1* form to enter information about the requirement.

4.2 Starting out - Development of Objective and Requirements

When first starting the simulation study, a particular focus is laid on membrane dynamics, in particular lipid rafts. Lipid rafts are small domains in the membrane (microdomains) with high local concentration of cholesterol, sphingolipids, and protein receptors. Therefore, the research objective is defined as: *What is the role of membrane lipid rafts on canonical Wnt signaling in human neural progenitor cells?*

Based on this rather general objective, first behavioral requirements of the membrane processes can be defined. For instance, lipid rafts typically cover $25-30\%$ of the membrane, and therefore the concentration of homogeneously distributed LRP6 within lipid rafts should stay between 25 and 30% of total LRP6 within any given time interval, whereas the concentration of the membrane-associated protein CK1y should

Figure 3: The Conceptual Modeling Wiki for creating, editing, and linking artifacts.

stay between 75 and 80%. We specify this requirement in the time interval of 60 to 720 minutes using Metrical Interval Temporal Logic (MITL) (Maler and Nickovic 2004) as supported by the statistical model checking module of SESSL (Ewald and Uhrmacher 2014):

$$\begin{aligned}
\phi_1 = \;& \mathtt{G}(60,720)((\mathtt{OutVar}(\mathtt{d_lrp6}) >= \mathtt{OutVar}(\mathtt{lrp6}) * \mathtt{Constant}(0.25))\mathtt{and} \\
& (\mathtt{OutVar}(\mathtt{d_lrp6}) <= \mathtt{OutVar}(\mathtt{lrp6}) * \mathtt{Constant}(0.3))) \\
\phi_2 = \;& \mathtt{G}(60,720)((\mathtt{OutVar}(\mathtt{d_ck1y}) >= \mathtt{OutVar}(\mathtt{ck1y}) * \mathtt{Constant}(0.75))\mathtt{and} \\
& (\mathtt{OutVar}(\mathtt{d_ck1y}) <= \mathtt{OutVar}(\mathtt{ck1y}) * \mathtt{Constant}(0.85)))
\end{aligned}$$

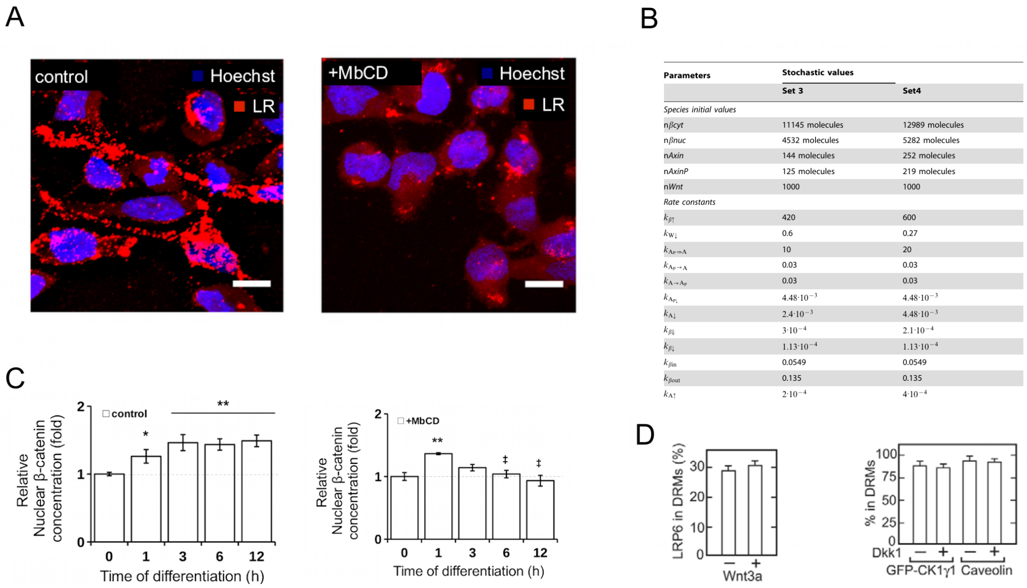
To illustrate how these information are matched to our definition of the requirement artifact *req*, consider the following filled tuple. However, note that actually all the formal definitions presented in Section 3 are encapsulated in the data structure of the wiki implementation.

$$\begin{aligned}
\text{Req1} = \;& \langle descr = \langle descrType = \text{'text'}, descrFormat = \text{' '}, descrTool = \text{' '}, descrSpec = \text{'The distribution of...'}, \ldots \rangle, \\
& behavioralReq = \langle logicType = \text{'MITL'}, logicTool = \text{'SESSL'}, logicFormula = \phi_1 \wedge \phi_2 \rangle, obj = \text{@Main Objective} \rangle
\end{aligned}$$

4.3 First Data and Information Sources

A variety of data have to be collected as prerequisites to actually pursue the initially defined research question in-vitro and in-silico (see Figure 4). Fluorescence microscopy data of in-vitro neural progenitor cells, e.g., confirms the existence of lipid rafts, and a partial localization of the receptors within these membrane structures. A second prerequisite and important data set for model calibration is obtained by analyzing the protein concentration of β-catenin through immunohistochemistry. These in-vitro data confirm that disruption of membrane lipid rafts in neural progenitor cells attenuates canonical Wnt signaling. Apart from conducting these two experiments in the lab, a variety of literature is consulted and various conceptual material is collected, such as parameter values from (Mazemondet et al. 2012), and the distribution of LRP6 and CK1y from (Sakane et al. 2010). In the conceptual model, for each of the data sets a new *src* artifact is created, e.g., the information about the β-catenin data looks like this:

$$\begin{aligned}
BetaCateninData = \;& \langle descr = \langle descrType = \text{'text'}, \ldots, descrSpec = \text{'The analysis of...'}, src = \text{' '} \rangle, \\
& srcType = \text{'experimental data'}, srcRole = \text{'calibration'}, srcIdType = \text{'DOI'}, \\
& srcIdSpec = \text{'https://doi.org/10.1371/journal.pcbi.1004106.s010'}, \\
& srcFormat = \text{'csv'}, srcTool = \text{'MS Excel'}, \ldots \rangle
\end{aligned}$$



Parameters	Stochastic values	
	Set 3	Set4
Species initial values		
n*βcyt*	11145 molecules	12989 molecules
n*βnuc*	4532 molecules	5282 molecules
n*Axin*	144 molecules	252 molecules
n*AxinP*	125 molecules	219 molecules
n*Wnt*	1000	1000
Rate constants		
$k_{\beta\uparrow}$	420	600
$k_{W\downarrow}$	0.6	0.27
$k_{A_P \Rightarrow A}$	10	20
$k_{A_P \rightarrow A}$	0.03	0.03
$k_{A \rightarrow A_P}$	0.03	0.03
$k_{A_{P\downarrow}}$	$4.48 \cdot 10^{-3}$	$4.48 \cdot 10^{-3}$
$k_{A\downarrow}$	$2.4 \cdot 10^{-3}$	$4.48 \cdot 10^{-3}$
$k_{\beta\Downarrow}$	$3 \cdot 10^{-4}$	$2.1 \cdot 10^{-4}$
$k_{\beta\downarrow}$	$1.13 \cdot 10^{-4}$	$1.13 \cdot 10^{-4}$
$k_{\beta in}$	0.0549	0.0549
$k_{\beta out}$	0.135	0.135
$k_{A\uparrow}$	$2 \cdot 10^{-4}$	$4 \cdot 10^{-4}$

Figure 4: Microscopy data (A), parameter table of the intra-cellular Wnt model (B) adopted from (Mazemondet et al. 2012), relative protein concentration of beta-catenin (C), and distribution of LRP6 and CK1y in membrane lipid rafts (D), as depicted in Figure 1 and 2 in (Sakane et al. 2010).

4.4 Qualitative Model and Choice of Modeling Approach

In this study, the choice of the approach, i.e., the modeling formalism and the corresponding modeling tool, is closely intertwined with the development of the qualitative model. At first, based on the available data, a diagram identifying the main structures and reactions is built (see Figure 5). It becomes clear that due to the previously defined objective and requirements a hierarchical, attributed modeling approach should be used. In particular, nested entities are needed to represent the structure of the cell and the separation of the membrane into individual lipid rafts; attributes are needed to account for the various binding and activation (phosphorylation) states of proteins, such as the receptor LRP6. The rule-based modeling language ML-Rules (Maus et al. 2011) supports all the desired features. Below, we show an excerpt of the more detailed qualitative model specification, i.e., the specification of the *entity* LRP6, thereby making explicit the related reactions, attributes, and information sources.

$$\langle name = \text{'LRP6'}, \ldots, att = \langle name = \text{'phos'}, \ldots \rangle, \ldots, act = \langle name = R2, \ldots \rangle, \ldots, src = \text{@Sakane et al. 2010} \rangle$$

4.5 Assumptions

This simulation study relies on two existing simulation models (Lee et al. 2003; Mazemondet et al. 2012), from which most assumptions are inherited. For example, according to Lee et al. the autocrine Wnt signal W can be modeled as a negative exponential process which in the corresponding assumption artifact is expressed using the formula: $W = \begin{cases} 0 & , \text{ for } t < t_0 \\ e^{-\lambda(t-t_0)} & , \text{ for } t \geq t_0 \end{cases}$.

4.6 Validating against Wet Lab Data

While developing and testing the first implemented version of the simulation model, we observe that the given qualitative model and the assumptions do not suffice to reproduce the experimental results. In detail, β-catenin concentration is increased at certain time points despite the disruption of the lipid rafts, which means that the requirement associated with the objective cannot be met by the simulation model. This indicates necessary changes in the conceptual model, e.g., by further experimental measurements, additional

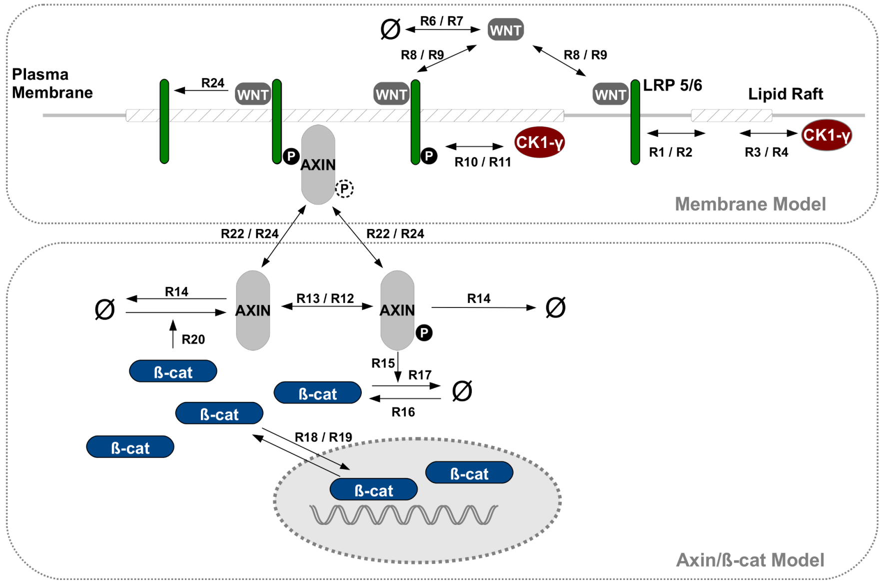


Figure 5: The qualitative model is provided as a diagram describing the submodels, entities, and reactions.

literature research, and possibly the definition of additional requirements. Indeed, in the follow-up study by (Peng et al. 2017) it is shown how the simulation model was extended to resolve the problem. The explicit representation of requirements and, based on this, the automatic generation of simulation experiments for the revised model showed to be of high value.

5 DISCUSSION

Our approach advocates a broad interpretation of the conceptual model in modeling and simulation. Diverse early-stage products of a simulation study can be made explicit and related to each other within the conceptual model. This supports managing the documentation during the simulation study in a structured manner, and should also facilitate later assessments of a simulation study and its results. Each early-stage product, be this an objective, a requirement, an assumption, data, or the qualitative model, is captured by some form of multimedia description, the “product” itself, and further means that help interpreting the product. Thus, in the ideal case, products ship with a formal semantics, that enables an automatic interpretation by respective tools for acting upon this semantics. E.g., if specified unambiguously with a clear semantics, requirements (Ruscheinski et al. 2019) or hypotheses (Lorig 2019) can automatically be checked during the simulation study. An unambiguous documentation of parameters can be used, e.g., for automatically setting up a sensitivity analysis (Ruscheinski et al. 2018). Recording, maintaining and exploiting the different aspects of the conceptual model promises to be particularly useful if various simulation models shall be developed. It reveals implicit relations, as well as similarities and differences between the early-stage products of the individual studies. Therefore, sharing conceptual models, e.g., through a wiki, can improve collaborative development and community reuse of artifacts. While the current implementation and conceptualization provide a good starting point for this, further case studies are required to develop and probe additional means for formalization and evaluate their implications for conducting a simulation study more systematically and effectively. Moreover, to be most valuable to modelers, our approach for conceptual modeling needs to be integrated with other approaches for supporting simulation studies, e.g., for tracing provenance (Ruscheinski and Uhrmacher 2017) or automatic experiment generation.

ACKNOWLEDGMENTS

This work was funded by the research grants DFG UH 66/18 ’GrEASE’ and DFG CRC 1270 ’Elaine’.

REFERENCES

Agha, G., and K. Palmskog. 2018. “A Survey of Statistical Model Checking”. *ACM Transactions on Modeling and Computer Simulation* 28(1):1–39.

Balci, O. 2012. “A Life Cycle for Modeling and Simulation”. *SIMULATION* 88(7):870–883.

Balci, O., J. D. Arthur, and W. F. Ormsby. 2011. “Achieving reusability and composability with a simulation conceptual model”. *Journal of Simulation* 5(3):157–165.

Bittig, A. T., and A. M. Uhrmacher. 2010. “Spatial Modeling in Cell Biology at Multiple Levels”. In *Proceedings of the 2010 Winter Simulation Conference*, edited by B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan, 608–619. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Cellier, F. E., and J. Greifeneder. 1991. *Continuous System Modeling*. Springer Science & Business Media.

Cetinkaya, D., A. Verbraeck, and M. D. Seck. 2012. “Model Transformation from BPMN to DEVS in the MDD4MS Framework”. In *Proceedings of the 2012 Symposium on Theory of Modeling and Simulation - DEVS Integrative M&S Symposium*. San Diego, CA: Society for Computer Simulation International.

CMoWiki 2020. Conceptual Modeling Wiki. https://cmowiki.informatik.uni-rostock.de, accessed 1*st* December.

Dalle, O. 2012. “On Reproducibility and Traceability of Simulations”. In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. M. Uhrmacher, 1–12. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Embley, D. W., and B. Thalheim. 2012. *Handbook of Conceptual Modeling: Theory, Practice, and Research Challenges*. Springer.

Ewald, R., and A. M. Uhrmacher. 2014. “SESSL: A Domain-specific Language for Simulation Experiments”. *ACM Transactions on Modeling and Computer Simulation* 24(2):11:1–11:25.

Fishwick, P. A. 1995. “Simulation Model Design”. In *Proceedings of the 1995 Winter Simulation Conference*, edited by C. Alexopoulos, K. Kang, W. R. Lilegdon, and D. Goldsman, 209–211. Piscataway, New Jersey: IEEE Computer Society.

Fujimoto, R., C. Bock, W. Chen, E. Page, and J. Panchal. (Eds.) 2017. *Research Challenges in Modeling and Simulation for Engineering Complex Systems*. Simulation Foundations, Methods and Applications Series. Springer International Publishing.

Gillespie, D. T. 1977. “Exact Stochastic Simulation of Coupled Chemical Reactions”. *The Journal of Physical Chemistry* 81(25):2340–2361.

Grimm, V., S. F. Railsback, C. E. Vincenot, U. Berger, C. Gallagher, D. L. DeAngelis, B. Edmonds, J. Ge, J. Giske, J. Groeneveld, A. S. Johnston, A. Milles, J. Nabe-Nielsen, J. G. Polhill, V. Radchuk, M.-S. Rohwäder, R. A. Stillman, J. C. Thiele, and D. Ayllón. 2020. “The ODD Protocol for Describing Agent-Based and Other Simulation Models: A Second Update to Improve Clarity, Replication, and Structural Realism”. *Journal of Artificial Societies and Social Simulation* 23(2):7.

Guizzardi, G., and G. Wagner. 2012. “Conceptual Simulation Modeling with Onto-UML Advanced Tutorial”. In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. M. Uhrmacher, 1–15. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Haack, F., H. Lemcke, R. Ewald, T. Rharass, and A. M. Uhrmacher. 2015. “Spatio-Temporal Model of Endogenous ROS and Raft-Dependent WNT/Beta-Catenin Signaling Driving Cell Fate Commitment in Human Neural Progenitor Cells”. *PLoS Computational Biology* 11(3).

Hoekstra, A. G., B. Chopard, D. Coster, S. Portegies Zwart, and P. V. Coveney. 2019. “Multiscale Computing for Science and Engineering in the Era of Exascale Performance”. *Philosophical Transactions of the Royal Society A* 377(2142):20180144.

Klir, G. J., and I. Rozehnal. 1990. “Epistemological Categories of Systems: An Overview and Mathematical Formulation”. In *Computer Aided Systems Theory — EUROCAST ’89*, edited by F. Pichler and R. Moreno-Diaz, 7–32. Berlin, Heidelberg: Springer Berlin Heidelberg.

Kluyver, T., B. Ragan-Kelley, F. Pérez, B. E. Granger, M. Bussonnier, J. Frederic, K. Kelley, J. B. Hamrick, J. Grout, S. Corlay, P. Ivanov, D. Avila, S. Abdalla, C. Willing, and Jupyter Development Team. 2016. “Jupyter Notebooks-a Publishing Format for Reproducible Computational Workflows.”. In *20th International Conference on Electronic Publishing*, 87–90.

Krötzsch, M., D. Vrandečić, and M. Völkel. 2006. “Semantic MediaWiki”. In *The Semantic Web - ISWC 2006*, edited by I. Cruz, S. Decker, D. Allemang, C. Preist, D. Schwabe, P. Mika, M. Uschold, and L. M. Aroyo, 935–942. Berlin, Heidelberg: Springer Berlin Heidelberg.

Lee, E., A. Salic, R. Krüger, R. Heinrich, and M. W. Kirschner. 2003. “The Roles of APC and Axin derived from Experimental and Theoretical Analysis of the Wnt Pathway”. *PLoS Biology* 1(1).

Lorig, F. 2019. *Hypothesis-Driven Simulation Studies - Assistance for the Systematic Design and Conducting of Computer Simulation Experiments*. Springer.

Maler, O., and D. Nickovic. 2004. “Monitoring Temporal Properties of Continuous Signals”. In *Formal Techniques, Modelling and Analysis of Timed and Fault-Tolerant Systems*, 152–166. Berlin, Heidelberg: Springer Berlin Heidelberg.

Maus, C., S. Rybacki, and A. M. Uhrmacher. 2011. “Rule-based Multi-level Modeling of Cell Biological Systems”. *BMC Systems Biology* 5(1):166.

Mazemondet, O., M. John, S. Leye, A. Rolfs, and A. M. Uhrmacher. 2012, 08. "Elucidating the Sources of β-Catenin Dynamics in Human Neural Progenitor Cells". *PLOS ONE* 7(8):1–12.

Monks, T., C. S. Currie, B. S. Onggo, S. Robinson, M. Kunc, and S. J. Taylor. 2019. "Strengthening the Reporting of Empirical Simulation Studies: Introducing the STRESS Guidelines". *Journal of Simulation* 13(1):55–67.

Nance, R. E. 1994. "The Conical Methodology and the Evolution of Simulation Model Development". *Annals of Operations Research* 53(1):1–45.

Ruscheinski, A., and A. M. Uhrmacher. 2017. "Provenance in Modeling and Simulation Studies - Bridging Gaps". In *Proceedings of the 2017 Winter Simulation Conference*, edited by W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, 872–883. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Ruscheinski, A., T. Warnke, and A. M. Uhrmacher. 2019. "Artifact-based Workflows for Supporting Simulation Studies". *IEEE Transactions on Knowledge and Data Engineering* 32(6):1064–1078.

Ruscheinski, A., K. Budde, T. Warnke, P. Wilsdorf, B. C. Hiller, M. Dombrowsky, and A. M. Uhrmacher. 2018. "Generating Simulation Experiments based on Model Documentations and Templates". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 715–726. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Oinn, T., M. Addis, J. Ferris, D. Marvin, M. Senger, M. Greenwood, T. Carver, K. Glover, M. R. Pocock, A. Wipat, and P. Li. 2004. "Taverna: a Tool for the Composition and Enactment of Bioinformatics Workflows". *Bioinformatics* 20(17):3045–3054.

Pace, D. K. 2000. "Ideas about Simulation Conceptual Model Development". *Johns Hopkins APL Technical Digest* 21(3):327–336.

Page, E. H. 1994. *Simulation Modeling Methodology: Principles and Etiology of Decision Support*. Ph. D. thesis, Virginia Polytechnic Institute and State University.

Peng, D., T. Warnke, F. Haack, and A. M. Uhrmacher. 2017. "Reusing Simulation Experiment Specifications in Developing Models by Successive Composition — a Case Study of the Wnt/β-Catenin Signaling Pathway". *SIMULATION* 93(8):659–677.

Robinson, S. 2008a. "Conceptual Modelling for Simulation Part I: Definition and Requirements". *Journal of the Operational Research Society* 59(3):278–290.

Robinson, S. 2008b. "Conceptual Modelling for Simulation Part II: a Framework for Conceptual Modelling". *Journal of the Operational Research Society* 59(3):291–304.

Robinson, S., G. Arbez, L. G. Birta, A. Tolk, and G. Wagner. 2015. "Conceptual Modeling: Definition, Purpose and Benefits". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, 2812–2826. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Sakane, H., H. Yamamoto, and A. Kikuchi. 2010, January. "LRP6 is Internalized by Dkk1 to Suppress its Phosphorylation in the Lipid Raft and is Recycled for Reuse". *Journal of Cell Science* 123(3):360–368.

Sanchez, S. M., P. J. Sánchez, and H. Wan. 2018. "Work Smarter, not Harder: a Tutorial on Designing and Conducting Simulation Experiments". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 237–251. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Sargent, R. G. 2013. "Verification and Validation of Simulation Models". *Journal of Simulation* 7(1):12–24.

Schewe, K.-D., and B. Thalheim. 2005. "Conceptual Modelling of Web Information Systems". *Data & Knowledge Engineering* 54(2):147–188.

Thalheim, B. 2010. "Towards a Theory of Conceptual Modelling". *Journal of Universal Computer Science* 16(20):3102–3137.

Tisue, S., and U. Wilensky. 2004. "Netlogo: A Simple Environment for Modeling Complexity". In *Fifth International Conference on Complex Systems*, Volume 21, 16–21.

Zeigler, B. P., H. S. Song, T. G. Kim, and H. Praehofer. 1995. "DEVS Framework for Modelling, Simulation, Analysis, and Design of Hybrid Systems". In *Hybrid Systems II*, 529–551: Springer Berlin Heidelberg.

AUTHOR BIOGRAPHIES

PIA WILSDORF is a Ph.D. candidate in the Modeling and Simulation group at the University of Rostock. Her e-mail address is pia.wilsdorf@uni-rostock.de.

FIETE HAACK is a Postdoctoral researcher in the Modeling and Simulation group at the University of Rostock. His e-mail address is fiete.haack@uni-rostock.de.

ADELINDE M. UHRMACHER is Professor at the Institute for Visual and Analytic Computing, University of Rostock, and head of the Modeling and Simulation group. Her e-mail address is adelinde.uhrmacher@uni-rostock.de.