Proceedings of the 2017 Winter Simulation Conference
W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, eds.

USING STRUCTURAL EQUATION-BASED METAMODELING FOR AGENT-BASED MODELS

Kai G. Mertens
Iris Lorscheid
Matthias Meyer

Institute of Management Accounting and Simulation
Hamburg University of Technology
Am Schwarzenberg-Campus 4
Hamburg, 21073, GERMANY

ABSTRACT

Trustworthy statistical modeling is an emerging challenge in agent-based modeling (ABM). However, typical characteristics of ABM such as the potential for high numbers of entities and parameters, interdependent relations between entities, several layers of effects, and emergent social phenomena challenge this process. In particular, aggregated outcomes emerging from individual agent interactions are, at least partly, difficult to measure. This might impede the statistical modeling process and thus the formulation of trustable conclusions. For this reason, we introduce structural equation modeling (SEM) as a promising statistical modeling method to analyze the behavior of ABMs. SEM allows for the estimation and evaluation of highly networked systems by explicating interactions between types of agents, measuring emergent phenomena, and identifying output patterns of simulation models. Overall, these contributions foster the credibility and trustworthiness of ABMs, and also support the communication and understanding of simulation models' behavior and their output.

1 INTRODUCTION

There is an increasing interest in trustworthy conclusions from complex dynamic agent-based models. Recently, coupled systems or interrelated agents have complicated the analysis of agent-based modeling (ABM) (Polhill et al. 2016). Traditional concepts of scenario and sensitivity analysis by means of metamodels are central approaches to understanding simulation model behavior, but might fail to address the increasing complexity in computational social and behavioral science (Schulze et al. 2017). It might be the case that standard functional forms of metamodels for ABMs are difficult to present and understand, especially to stakeholders (Grimm and Berger 2016a), and therefore uncertainty concerning simulation model behavior remains. Thus, the potential usefulness of these models might suffer from a lack of understanding (Grimm and Berger 2016b). In addition, the analysis of ABMs consumes many resources.

ABM has a strong emphasis on explanatory questions in the social and behavioral sciences and contributes to generative explanations with respect to underlying patterns and dynamics, as well as aggregated behavior of heterogeneous agents (Epstein 2007; Squazzoni 2012). However, this emphasis can result in several layers of effects, reciprocal interactions, latent emergent phenomena, or macroscopic effects (Schulze et al. 2017). For instance, interactions, depending on a minimum set of rules at the individual level, possibly lead to emergent phenomena such as oscillation. Consequently, interesting outcomes might be hard to capture by means of standard statistical models, and relevant effects potentially remain hidden in the system. This provides challenges for simulation model analysis. In addition, very little is known about ways to deal with them adequately. In particular, far too little attention has been paid to understanding interacting agents, coupled systems, and their resulting emergence in accessible ways, which is a notable benefit in agent-based modeling. This limitation becomes especially visible when it comes to the complex social interactions in nonlinear simulation models with such phenomena.

This paper suggests a structured functional form for simulation metamodels in ABM, providing an accessible methodology for understanding complex model behavior. Our approach allows for the incorporation of prior conceptual model information such as individual behaviors or rules and the underlying simulation model data, aiming at analyzing the simulation model behavior. Structural equation modeling (SEM) is a common method for analyzing social and behavioral effects in empirical studies (Hair et al. 2014) and has been recently introduced for the analysis of simulation model behavior by means of simulation metamodeling (Mertens et al. 2015). Consequently, this metamodeling technique is able to address complex models, in particular interacting agents as well as high order phenomena, by means of convenient software such as smartPLS (Ringle et al. 2015). Overall, we see a possible method to support modelers to meet the challenges of analyzing and presenting ABMs, which might clarify entangled individual behavior and their resulting emergent phenomena.

We exemplify the capabilities of SEM using a socio-ecological simulation model. The chosen predator-prey model has coupled non-linear dynamics and macroscopic phenomena. First, a statistical regression model from Law (2014) is used as a baseline. We show that this can lead to counterintuitive results that may be further analyzed by SEM. Additionally, regression provides limited information about macroscopic system behavior due to missing endogenous and reciprocal relations. Here, SEM has the capability to address such relations in simple expressions. The ability of SEM to estimate interrelated relationships and to capture latent emergent social behavior by means of constructs and paths helps to assess simulation model behavior from a new but accessible perspective. Consequently, this approach may reveal the underlying explanatory processes and crucial drivers for an emergent phenomenon. This is particularly relevant because modelers typically need to design specific measures or analyses for the examination of emergent behavior (see Gore et al. 2017). In summary, the methodology might increase the rigor of simulation modeling referred to as generative explanations.

The paper proceeds as follows. First, we introduce in Section 2 the standard regression-based process of using metamodels in ABM. Next, in Section 3, we introduce structured functional metamodels by means of SEM and show disparities to the benchmark metamodel concerning resulting inferences. Here, we generate two types of metamodels of a complex nonlinear simulation model based on Law (2014). In Section 4, we propose the structured functional model as a new treatment of specific challenges in agent-based simulation. Finally, we provide some concluding remarks in Section 5.

2 REGRESSION-BASED METAMODELS FOR AGENT-BASED MODELS

Simulation as a research method provides ways to generate a particular research question, and statistical modeling is an established tool that enables comprehensible analyses of results and consumable communication of conclusions (Kleijnen 2015). In this line, metamodeling is a popular technique and has the merit of being accessible, which could result in a trustworthy analysis (Goldspink 2000). Metamodeling also fulfills a vital role in understanding and presenting simulation model behavior and its results.

Not surprisingly, simulation model analysis is also an emerging concern in ABM, particularly because emergent outputs entrench themselves in rather indirect observations. Each outcome of an ABM is grounded in individual model activities, which are able to emerge as aggregated patterns. Likely aggravating circumstances of complexity in ABM are dynamics, interrelated relationships, and macroscopic phenomena, which possibly pose a threat to a clear understanding of model behavior and output. In addition, emergent behavior appears to be rather latent and difficult to operationalize.

To investigate and present SEM for ABM as a metamodeling approach further, we consider a predator-prey model from the NetLogo Library (Wilensky 1999), which represents a complex nonlinear simulated socio-ecological system (Wilensky 1997). Besides these characteristics, the model was chosen because it is used in Law (2014) as an example of a regression-based metamodel, which will be used in the following as a benchmark.

The simulation model embodies a complex socio-ecological system, including interactions of three entities (wolves, sheep, and grass). The main model assumptions are intuitive. First, if grass is eaten, then it will regrow after a determined amount of time. Second, the sheep eat the grass to gain energy, and finally the wolves eat the sheep to gain energy as well. These assumptions and their variations generate fluctuations in the persistence of the ecosystem, which is the emergent phenomenon. For a more detailed description of the model, see Law (2014), Lorscheid and Meyer (2016), and Wilensky (1997).

The generation of metamodels requires simulation data that are ideally generated using designed experiments (Lorscheid et al. 2012). We adopt the Latin hypercube design with the identical factor ranges as presented in Law (2014). The design is generated with m = 50 levels and k = 5 factors by means of MATLAB R2015b to achieve similar parameters. We also use the same terms for the factors: x_1 = sheep-gain-food; x_2 = sheep-reproduce; x_3 = wolves-gain-food; x_4 = wolves-reproduce; and x_5 = grass-regrowth. This dataset is used for all following analyses. Based on these data, we were able to replicate the metamodel from Law (2014) of the wolves (y_{wolves}= number of wolves), leading to the nearly identical regression-based metamodel with $R^2 = 0.933$:

$$\begin{aligned} y_{wolves} = {} & 68.65 + 37.39\,x_1 + 12.09x_2 + 29.57x_3 - 16.96x_5 - 8.12x_1x_2 - 17.26x_1x_3 + 5.90x_2x_3 \\ & -2.54x_3x_5 - 8.29x_3x_5 - 16.81x_1^2 - 10.52x_3^2 \end{aligned} \tag{1}$$

Trying to understand the simulation model behavior can be tedious, and metamodels of ABMs can be particularly challenging to interpret. Initially, the high order effects get attention because of the eight different terms. So the information concerning the simulation model behavior is aggravated, but it might even be misleading because if we neglect 2 – 4 % of the design points then we achieve completely different high order terms. Thus, caution is needed because aspects of overfitting might appear. In addition, we do not get insights about the other agents, and thus there is the necessity to generate a second metamodel for the sheep (y_{sheep}). Again, this statistical model also has a predictive power and a high fit of $R^2 = 0.995$ and can be expressed by:

$$y_{sheep} = 165.4 + 7.41x_1 - 0.79x_2 - 38.01x_3 - 5.55x_5 - 7.08x_2x_3 - 1.53x_1^2 - 12.60x_3^2 \tag{2}$$

According to this model, grass-regrowth (x_5) has a minor negative effect on sheep. This is counterintuitive, as it is expected that the sheep population should strongly increase if they are able to eat more grass. Therefore, we would expect a positive high effect of grass on sheep. In general, the final interpretation of the simulation model behavior remains on a general and aggregated level of the regression function.

In contrast, the socio-ecological system comprises different entities that possibly interact in a complex, multi-stage process, which might result in relevant findings that have not yet been addressed by the metamodel. The persistence and oscillation of the ecological system represents an output of interest and there is no defined way to address these emergent behaviors, referring to the proposed baseline metamodels. As mentioned, baseline metamodeling techniques do not typically emphasize all endogenous relations such as simultaneous causalities. Nevertheless, this might neglect relevant information concerning the behavior of the system.

3 STRUCTURAL EQUATION MODELING FOR AGENT-BASED MODELING

This paper uses partial least squares (PLS)-SEM, which is a statistical modeling technique widely known across several disciplines, as it helps to connect information from theoretical models and the data obtained. The application of this methodology benefits from numerous characteristics such as measuring latent phenomena, being nonparametric regarding data, and convenient software tools (Ringle et al. 2015). In defining relationships and rules *a priori* by means of expert knowledge and theory, it consequently yields to a causal network, which needs a model-based specification by measureable factors. In general, the method has linkages with the semantic of parallel distributed processing and is closely related to regression algorithms (Rumelhart et al. 1987). Thus, SEM consists of inputs, outputs, and intermediate instances with their own output equations. Concerning parallel distributed processing, this approach has neither hidden units nor learning and activation rules, which contrasts with neural networks. Regarding the analysis and understanding of simulation model behavior, however, the combination of a conceptual semantic presentation of the simulation model and the corresponding measures from the simulation output constitute a promising new way to relate agents' behaviors to their observable attributes.

This part demonstrates the application of PLS-SEM by means of a socio-ecological system, which tries to reproduce the findings from Law (2014). The first step is to operationalize the simulation entities or agents, respectively. In this model, there exist three characteristic entities, namely grass, sheep, and wolves. They will be represented by five factors: grass-regrowth (x_5) is the only factor of grass and, thus, it determines the entity 'grass regrowth time'; sheep-gain-food (x_1) and sheep-reproduce (x_2) are the characteristics that drive sheep, and therefore, they are aggregated to the 'sheep characteristic'; and finally, the wolves comprise wolves-gain-food (x_3) and wolves-reproduce (x_4), and in the same way they present the entity 'wolves characteristic.' Next, we define all populations (i.e. grass y_{grass}, sheep y_{sheep}, and wolves y_{wolves}) as separate entities, because they interact with each other. This is a second layer that allows us to represent intuitive conceptual information or behavior between the agents, respectively (Mertens et al. 2015).

Figure 1 (next page) shows the structural metamodel, meaning the factor relationships, and its estimated unstandardized parameter values. The rectangles represent the specified factors that are related to the entities, which are symbolized by circles. The relationships are the inherent structural rules. The path coefficients result from a partial least square algorithm, which is performed by smartPLS (Ringle et al. 2015). Subsequently, the parameter values represent the underlying behavior of the simulation data concerning its relationships. Due to these characteristics, this metamodel increases the transparency and consumability of the model behavior and thus supports the communication by explicating several sources of information.

Furthermore, this metamodel unveils a detailed effect decomposition that might support the simulation model analysis. To exemplify our argument, we analyze the effects of grass regrowth time. If we investigate solely the direct path coefficient of grass regrowth time, negative effects on the wolves, sheep, and grass are apparent again. Thus, if the regrowth time is high (i.e., the grass needs more time to regrow after being eaten by the sheep), it will result in a lower population, especially of sheep (-28.730). This reflects the observation from the baseline regression model that the higher the time required for grass regrowth, the smaller the sheep population.

Thus, we further consider the effects of grass regrowth time on grass and then on sheep. Grass regrowth time has a negative effect on the grass quantity (-23.012), and so the longer the regrowth time, the less grass there is to be eaten by the sheep. Grass again has a negative effect on sheep (-0.997). The explanation behind this effect is the "paradox of enrichment" (Rosenzweig 1971), meaning that an unrestricted resource supply may destabilize ecosystems. Here, too much grass may lead to sharp increases in sheep populations and thus also lead to an excess of wolves, which again may lead to the reduction or even extinction of sheep and, in the end, the whole population. This explains the negative relation of grass quantities on sheep population. In total, by considering the decomposition effect of the indirect and direct effect and by combining them, we get a small negative parameter value by means of offsetting: $((-23.012 \times -0.997) - 28.730 = -5.78)$. This value is nearly the same as the value in the baseline metamodel (see equation (2)), yet this effect decomposition makes explicit the causal network and might, therefore, offer a more trustworthy explanation of the complexity in the simulation model behavior. Thus, we can explain the counterintuitive effect by means of explicating direct and indirect relations based on prior assumptions from the conceptual model. This should illustrate our point that SEM can reveal explanations in depth within ABM.

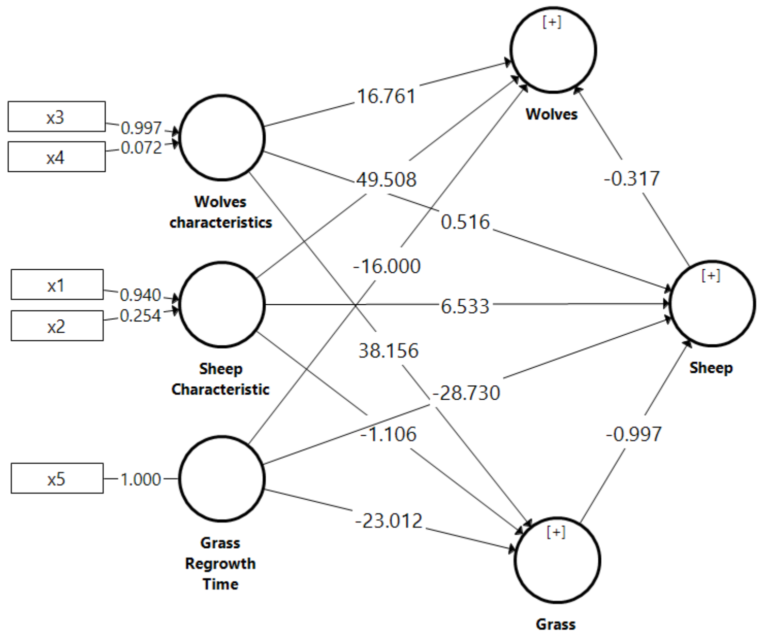

Figure 1: Structured functional metamodel by means of structural equation-based metamodeling.

In this section, we showed that an overly simple metamodel can provide misleading or blurred inferences. Both the regression-based baseline models discussed offer rather inscrutable functional forms and contain too many fitted terms. In addition, the baseline models hide the underlying patterns concerning agents and emergent behavior. Using SEM with the same simulation model data, we are able to give more details about the complex effects and consequently the simulation model behavior. Nevertheless, both the SEM and regression-based metamodel are still working at the surface in examining the underlying simulation model behavior. In order to address this issue in more detail, we examine in the next section the interdependent relations between entities and measure a higher order phenomenon of the ecosystem.

4 SIMULATION MODEL ANALYSIS

4.1 Understanding Interdependent Relationships between Agent Types

The nature of ABM comprises interdependent relationships between agents or systems, which is one of its main advantages concerning modeling capabilities. Reciprocal effects between agents occur and yield to dynamic model behavior that may result in emergent phenomena such as oscillation. Estimating such effects plays a decisive role in understanding the simulation model behavior (Schulze et al. 2017). For example, more grass will increase the number of sheep, but as the sheep eat the grass this will reduce the amount of grass. In this line, a key issue is to estimate such effects consistently in order to shed light on the underlying dynamic patterns. However, a well-established approach for analyzing such simultaneous model behavior in agent-based models is missing. Thus, we claim that methods for the analysis of those reciprocal relationships might be useful for understanding the interactions between agent types and might support a trustworthy simulation model analysis.

Using the same experimental data, we perform an additional analysis by means of two-stage least square (2SLS) regression with our structural metamodel in order to estimate the consistent parameters of reciprocal relationships. A major problem regarding the baseline metamodel is that the estimation of reciprocal interactions results in inconsistent parameter values. Agents in dynamic environments effect each other reciprocally, which is reflected in a kind of endogeneity in ABM data. This results in a potential inconsistency for statistical modeling, which is due to correlated error terms (i.e., cov(agent $type_X$; agent $type_Y$) $\neq 0$). In other words, if one wants to analyze the effect between agent $type_X$ and agent $type_Y$ and neglects their mutual dependency, their consistent individual effects cannot be captured.

To address this issue, 2SLS splits the estimation process into two stages. First, we denote an identified reduced structural equation for each agent type, which dedicates the corresponding variables to the agent, to *ex-ante* cover the correlated error. This actually is a decisive problem in empirical studies, yet modelers should know their simulation models and profit from orthogonal input factors. Next, in using predicted values from the reduced equations (stage one), the fitted values should no longer contain correlated errors. Subsequently, we may estimate the interactions suggested using the fitted values.

In tangible terms, the first stage addresses the correlated errors by specifying the agent types individually by their input factors and reciprocal interactions, whereas the second stage profits from an adjusted variance, which allows for the estimation of agent interactions consistently. Endogeneity hampers the ability to determine consistent individual effects between agent types, yet it can be treated by means of SEM. Further details and a comprehensive mathematical description of the methodology can be found in Angrist and Pischke (2009).

Considering our simulation model, wolves eat sheep and sheep eat grass, but – concerning a longer time in the system – they also decline their potential nutrition. Conclusively, we argue that these entities have reciprocal effects and require special treatment. Therefore, we propose the described approach that allows for the estimation of consistent parameters of simultaneous relationships in simple terms.

For performing a 2SLS analysis, we first assign the factors of the simulation model (x_1 to x_5) to their agents (y_{wolves}, y_{sheep}, y_{grass}), including their coupled effects (i.e. y_{grass} affects y_{sheep}; y_{sheep} affects y_{grass}), in order to identify three structural equations for each agent (i.e. grass quantity: $x_5 + y_{sheep} = y_{grass}$). As mentioned above, it is important to identify each exogenous and endogenous association correctly to address the correlated variance (i.e. cov(y_{sheep}, y_{wolves}); cov(y_{sheep}, y_{grass}) = 0). Afterwards, we replace the endogenous factor (e.g. y_{grass}) with its fitted values ($\hat{y}_{grass}$) from the reduced structural equation. Finally, we are able to estimate the consistent unstandardized parameter values for the reciprocal interactions as shown in Figure 2 (next page).

Unstandardized [standardized] parameter values describe the effects between agent types and help to explain their basic relationships. For instance, let the sheep population be bigger than the wolves at a certain time step. Concerning loop (2), wolves eat sheep and so the sheep population declines (-0.233 [-0.301]), which results in a positive development of their own (-0.788 [-0.507]). Subsequently, loop (1) points out that less sheep supports grass growth and vice versa (-1.057 [-0.672]), and if the grass quantity increases, the sheep will need to keep declining otherwise it will not grow steadily (-0.581 [-0.815]). Summarized, both loops are constantly proceeding until the sheep population achieves a low level, which impedes predation from the wolves. Then, the sheep can enjoy a fruitful environment for themselves, while the grass quantity is large and the wolves are declining. Here, the structural equation-based metamodel possibly provides conclusions about the agents' interactions, and so it likely helps us to understand the aggregated picture of the socio-ecological system.

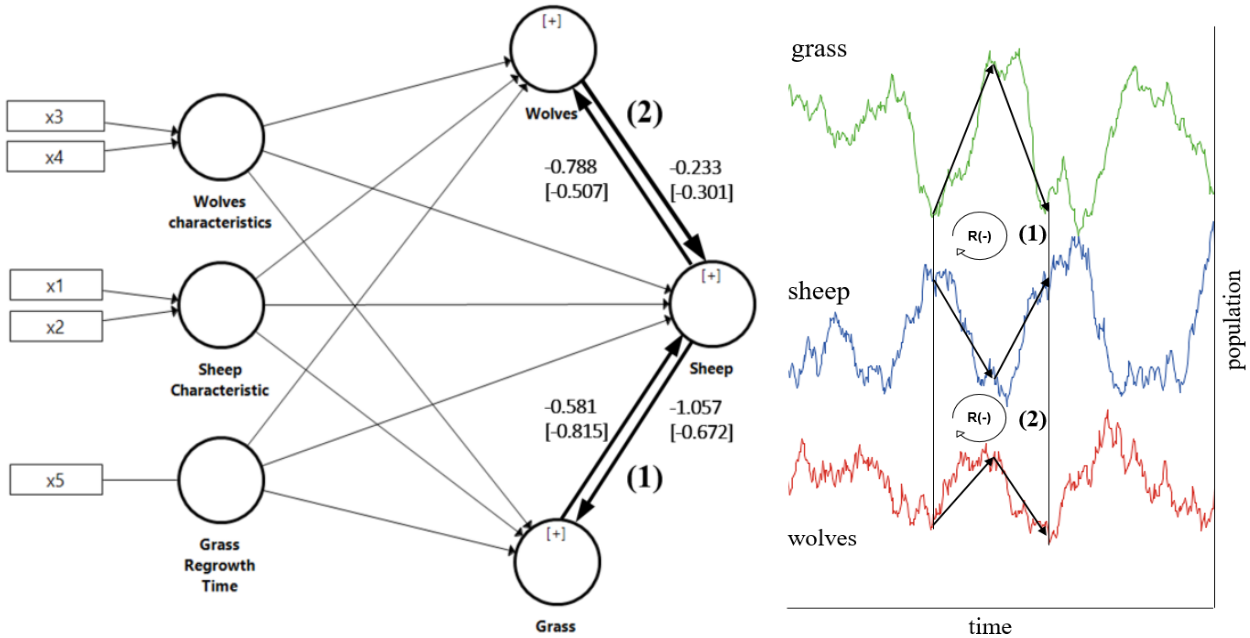

Figure 2: A demonstration of the usability and relevance of interdependent relationships referred to in the simulation model output. We first can identify two negative reinforcing loops (1) and (2) concerning their negative unstandardized [standardized] parameter values. This indicates a cause of oscillation. In addition, the parameter values provide an explanation for dissimilar peak amplitudes. When the information is taken together, the metamodel may explain the emergent oscillation, while it identifies the individual pattern in context to each other.

Closer inspection of the results in Figure 2 also reveals information about reference loop pattern (1) and potential statements about the amplitudes (2), which allows for the explanation of the emergent oscillation pattern (macroscopic) from an agent's perspective (microscopic). Even though we see the dynamics in the model output, smaller structures or detailed findings might be hidden and hamper a full understanding.

First (1), both agent interactions indicate negative parameter values, which can be further identified as self-reinforcing loops. Loops, in particular reinforcing ones, are potential drivers for destabilization, presumably because they cannot be directly controlled. Examining the metamodel from our example, we can denote a reference pattern with two negative reinforcing loops, meaning that the agents (i.e. wolves and sheep) are promoting their own self-reinforcing feedback in decreasing the other.

Second (2), feedback mechanisms with delays are a prominent phenomenon for oscillation that is usually based in agent interactions. Grounded in the estimation process by means of PLS-SEM with 2SLS, we are able to identify parameter values for each path. Here, a higher standardized strength possibly indicates a higher volatility of the oscillation, for instance in the case of loop (2), which has a stronger feedback than direct effect (-0.788 to -0.233). An initial increase of wolves does not immediately influence the sheep population, while one wolf is not able to reduce a *complete sheep* (i.e. one does not find a sheep). As a result, the standard deviation of the wolves, depending on the steady state, is about three times higher than the sheep. It might sound intuitive, but it could be a reason why wolves are more vulnerable to parameter changes.

To conclude this section, the disentanglement of the macroscopic oscillation pattern by individual agents' interactions by means of SEM likely provides explanations for emergent phenomena. ABMs are not implementing oscillation patterns in advance, as they particularly result from the underlying individual agent activities. Therefore, understanding and identifying such aggregated behavior starts to link heterogeneous agents in context to their assumed behavior. Recently, studies have shown that this as an important challenge (Polhill et al. 2016; Schulze et al. 2017), and they have highlighted the necessity for further research. In our study, we are able to address this issue by means of analyzing referential patterns and parameter values in order to identify decisive agents' interactions in the model behavior. The proposed approach provides the possibility of understanding dynamic models via specifying causal loops. We consider them to be accessible maps to examining the system behavior, while it relates simulation model output dynamics to reciprocal interactions of agents.

4.2 Understanding Macroscopic Phenomena by means of Construct Building

Capturing and analyzing aggregated system behavior quantitatively can be resource intensive for agent-based modelers. This paper suggests a process to use already implemented factors for this objective. In ABM, the macro-level phenomenon is typically a core subject of interest. Nonetheless, this process takes time and might hamper condensed communication, especially to stakeholders. For example, modelers need to implement and define specific measures for analyzing the resulting system behavior quantitatively.

In order to support and simplify such processes, we show a way that approximates aggregated behavior of social systems directly in referring to existing factors such as agents' attributes or outcomes. According to the simulated socio-ecological system, we define an entity that represents the persistence of the system. Here, persistence means that all species are alive and the ecosystem is resilient in the long run. We propose to capture the multifaceted nature of persistence by means of the individual populations, and to achieve this, we compute a new measure by means of a formative construct. The underlying measures determine a latent effect by means of a formative construct or – to put it another way – the computed factors cause the corresponding construct (see Hair et al. 2014). In terms of this example, persistence encompasses the population levels of wolves and sheep. Therefore, this is a new measure that allows for analyzing shrinks in the persistence of populations.

Concerning all apparent and significant layers of effects in Figure 3 (i.e. agents' factors, agents' interactions, and a final collective entity), we are able to explore the underlying model structures, which allows for the explication and understanding of emerging patterns in model behavior. First, following the paths from agents' attributes to respective agents helps to understand their effect on the population size, which further allows for suggestions about emerging loop patterns. In detail, if grass regrowth time is diminishing, then it leads to a high quantity level of grass (-23.012). Again, grass regrowth time regulates the time it takes for grass to grow back, and a lower value leads to more grass. In addition, the wolves' characteristic affects the grass, sheep, and wolf population. There, it reduces the sheep (-37.533), increases the grass (38.156), and intuitively supports the wolves (28.667), and it can shift the system to rather volatile situations, because two of the three direct paths end in a negative influence on persistence (i.e. grass: $38.156 \times -0.488 \times 0.387$; sheep -37.533×0.587).

Connecting this behavior with the documented behavior above, we can depict the following: If the grass quantity is getting large, it will endanger persistence ($0.580 \times -0.587 = -0.341$), while the feedback loop (2) might have enough potential to destabilize the system. Here, we can confirm that the simulation model behaves as it is theoretically described by Rosenzweig (1971), who found that ecological systems suffer from a 'paradox of enrichment.' Overall, the social system model behavior remains complex and entangled, but with SEM it is possible to track relevant effects as well as to discover decisive simulation model patterns in simple and understandable terms.

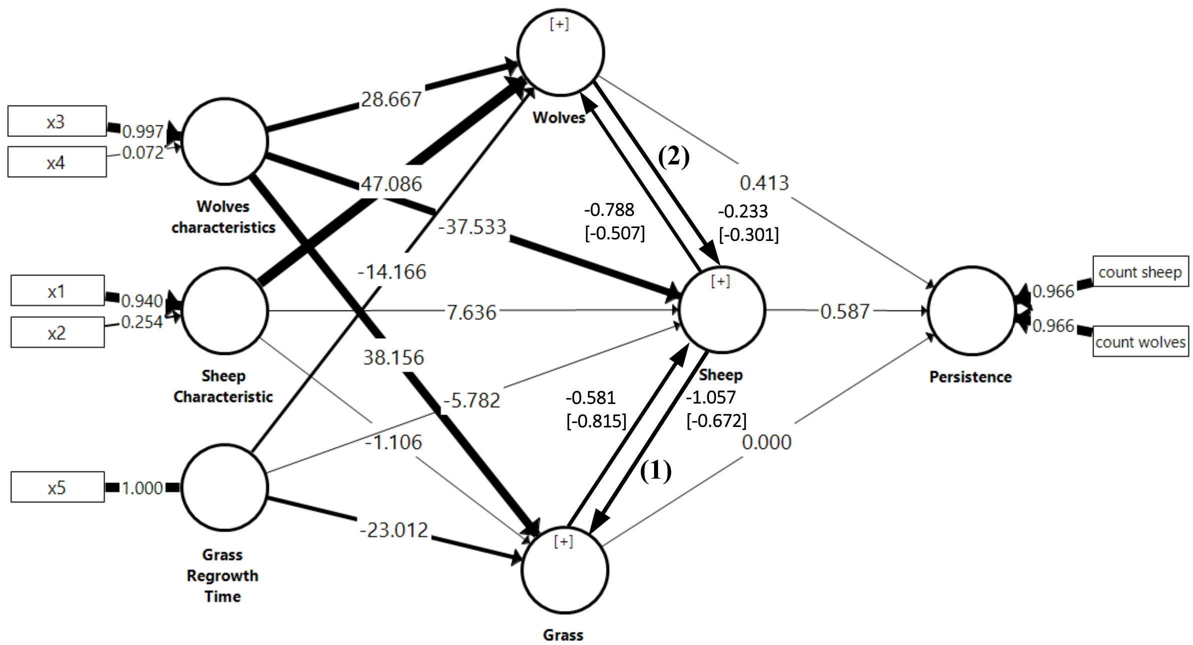

Figure 3: Complete structural equation-based metamodel of a socio-ecological system. The comprehensive metamodel allows for detailed qualitative as well as quantitative analyses and helps us to understand the social-ecological system. To understand an increase of a certain agent population, one needs to consider the driving characteristics. In order to explain the emergent oscillation of the simulation model output, one could look at agent interactions and use the identified parameter values. Finally, the effect on the persistence is likely to be visible in the last layer, where – for example – the quantities of grass as well as the wolf population have the possibility of a negative effect on persistence (i.e. -0.581 × 0.587 = -0.341;-0.233 × 0.587 = -0.136).

5 CONCLUDING REMARKS

The present study aims to introduce a SEM-based approach for the metamodeling of ABMs to facilitate simulation model analysis. The second aim was to investigate and provide ways to handle specific challenges that researches have to face in ABM, such as emergent phenomena.

One finding from this study is that the application of regression-based metamodels to explore underlying behaviors in a complex socio-ecological system can lead to potentially misleading results or can result in not all emerging patterns being explained. We therefore introduce SEM as a promising technique for metamodeling in the context of simulation model analysis in ABM. The proposed metamodeling approach offers a consumable functional form that allows for decomposing effects, which increases transparency and traceability. The comparison of the conceptual model (i.e. implemented behaviors) with the simulation model behavior reflected in the simulated data enhances the understanding of the simulation model behavior. This project can contribute to a framework for the explanation of complex social systems and support researchers in their generative conclusions.

A key contribution of the paper is to document the challenge of emergence in simulation model analyses in ABM and how it might be addressed. A number of studies have postulated ways to measure aggregated effects, yet ours is the first work to report a rather simple solution and presentation for such phenomena. Emergent behaviors of social systems still need specific treatment, which can be both complex and time-consuming, and baseline metamodels possibly provide limited support, as mentioned. This study proposed a way to utilize implemented relationships and agents' attributes from the individual level to explain the macroscopic phenomena. This possibility should prove to be particularly valuable for presenting patterns of social systems, and ideally it will enhance the trustworthiness of complex models. In particular, one crucial contribution of SEM in empirical studies is to measure unobservable social phenomena; however, this advantage could also be used in ABM. Therefore, depicting new structure and measurement models might help to measure macroscopic social behaviors, in linking already existing measures and relations without high efforts. Our introduced metamodel assesses conceptual (e.g. top-down information) and data information (e.g. bottom-up information) combined, and allows for the explanation of patterns that are emerging from the simulation model behavior. This rather new possibility of metamodeling might also have a bearing on further analyses of agent-based simulation, because it allows for the identification of generative structures. Thus far, our findings may support researchers in finding a plausible and acceptable depiction of the simulation model behavior, including emergent patterns, without additional effort. Eventually, the proposed metamodel fosters trustworthy and reliable explanations as well as ensuring communication to stakeholders.

Further research is needed to test the suggested approach with other ABM models, because we expect this new approach to be a fruitful area for further work and applications. Also, a natural progression of this work is to analyze the metamodel's effects in order to transform an ABM into a system dynamics model, which allows for the generalizability of our findings. A limitation of this specific study is that we currently cannot incorporate spatial aspects, which might be relevant for the problem studied in our case study and in other ABMs.

REFERENCES

Angrist, J. D., and J.-S. Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, New Jersey: Princeton University Press.

Epstein, J. M. 2007. *Generative Social Science: Studies in Agent-Based Computational Modeling*. Edited by S. A. Levin, and S. H. Strogatz. London, UK: University Press Group Ltd.

Goldspink, C. 2000. "Modelling Social Systems as Complex: Towards a Social Simulation Meta-Model." *Journal of Artificial Societies and Social Simulation* 3 (2).

Gore, R., S. Diallo, C. Lynch, and J. J. Padilla. 2017. "Augmenting Bottom-up Metamodels with Predicates." *Journal of Artificial Societies and Social Simulation* 20 (1).

Grimm, V., and U. Berger. 2016a. "Robustness Analysis: Deconstructing Computational Models for Ecological Theory and Applications." *Ecological Modelling* 326:162-167.

Grimm, V., and U. Berger. 2016b. "Structural Realism, Emergence, and Predictions in Next-Generation Ecological Modelling: Synthesis from a Special Issue." *Ecological Modelling* 326:177-187.

Hair, J. F., G. T. M. Hult, C. M. Ringle, and M. Sarstedt. 2014. *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. Thousand Oaks and Los Angeles, California: Sage.

Kleijnen, J. P. C. 2015. *Design and Analysis of Simulation Experiments*. Second Edition, *International Series in Operations Research & Management Science*. Cham, CH: Springer International Publishing.

Law, A. M. 2014. *Simulation Modeling and Analysis*. 5th International Edition. New York, New York: McGraw-Hill, Inc.

Lorscheid, I., B.-O. Heine, and M. Meyer. 2012. "Opening the 'Black Box' of Simulations: Increased Transparency and Effective Communication through the Systematic Design of Experiments." *Computational & Mathematical Organization Theory* 18 (1):22-62.

Lorscheid, I., and M. Meyer. 2016. "Divide and Conquer: Configuring Submodels for Valid and Efficient Analyses of Complex Simulation Models." *Ecological Modelling* 326:152-161.

Mertens, K. G., I. Lorscheid, and M. Meyer. 2015. "Structural Equation Modeling for Simulation Metamodeling." In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. M. Macal, and M. Rosetti, 701-712. Piscatawa, New Jersey: Institute of Electrical and Electronics Engineers.

olhill, J. G., T. Filatova, M. Schlüter, and A. Voinov. 2016. "Modelling Systemic Change in Coupled Socio-Environmental Systems." *Environmental Modelling & Software* 75:318-332.

Ringle, C. M., S. Wende, and J. Becker. 2015. "Smartpls 3." *Bönningstedt: SmartPLS: Retrieved from http://www.smartpls.com* [Accessed on 2017-06-30].

Rosenzweig, M. L. 1971. "Paradox of Enrichment: Destabilization of Exploitation Ecosystems in Ecological Time." *Science* 171 (3969):385-387.

Rumelhart, D. E., J. L. McClelland, and P. R. Group. 1987. "Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations." Edited by J. A. Feldmann, P. J. Hayes, and D. E. Rumelhart, *Computational Models of Cognition and Perception*. Cambridge, Massachusetts: MIT Press.

Schulze, J., B. Müller, J. Groeneveld, and V. Grimm. 2017. "Agent-Based Modelling of Social-Ecological Systems: Achievements, Challenges, and a Way Forward." *Journal of Artificial Societies and Social Simulation* 20 (2) (8).

Squazzoni, F. 2012. "What Is Agent-Based Computational Sociology All About?". In *Agent-Based Computational Sociology*, 1-32. John Wiley & Sons.

Wilensky, U. (1997). NetLogo Wolf Sheep Predation model. http://ccl.northwestern.edu/netlogo/models/. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, Illinois.

Wilensky, U. (1999). NetLogo. http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, Illinois.

AUTHOR BIOGRAPHIES

KAI G. MERTENS is an associate researcher at the Institute of Management Accounting and Simulation at the Hamburg University of Technology, under the mentorship of Prof. Dr. habil. Matthias Meyer. He graduated with a Master's degree in industrial engineering and his research focuses on modeling and simulation of systems. His email address is kai.mertens@tu-hamburg.de.

IRIS LORSCHEID is a post-doctoral researcher and lecturer at the Institute of Management Control and Accounting at the Hamburg University of Technology. Her research interests include agent-based modeling and simulation model analysis. She received her diploma in computer science from the University of Koblenz and her PhD from the Hamburg University of Technology. Her email address is iris.lorscheid@tu-hamburg.de.

MATTHIAS MEYER is a Professor of Management Control and Accounting and Head of the Institute of Management Accounting and Simulation at the Hamburg University of Technology. His research interests include simulation data analysis, simulation modeling, the design of cost accounting systems, and modeling of accounting routines. He has a habilitation degree from the WHU – Otto Beisheim School of Management and he received a PhD from Ludwig-Maximilian University in Munich. His email address is matthias.meyer@tu-hamburg.de.