Contextualizing Large-Scale Domain Knowledge for Conceptual Modeling and Simulation

Sungeun An[1] , **Spencer Rugaber**[1] , **Jennifer Hammock**[2] and **Ashok K. Goel**[1]

[1]School of Interactive Computing, Georgia Institute of Technology
[2]National Museum of Natural History, Smithsonian Institution
sungeun.an@gatech.edu

Abstract

We present an interactive modeling tool, VERA, that scaffolds the acquisition of domain knowledge involved in conceptual modeling and agent-based simulations. We describe the knowledge engineering process of contextualizing large-scale domain knowledge. Specifically, we use the ontology of biotic interactions in Global Biotic Interactions, and the trait data of species in Encyclopedia of Life to facilitate the model construction. Learners can use VERA to construct qualitative conceptual models of ecological phenomena, run them as quantitative simulations, and review their predictions.

1 Introduction

Modeling is central to human cognition and scientific reasoning [Schwarz and White, 2005]. Qualitative representations in the form of conceptual models capture the components and mechanisms that explain the observed phenomenon [Forbus, 1984][Schwarz and White, 2005], which helps learners externalize, share, and simulate the system the model represents. Simulation models are executable with specific values for the system's input variables, enabling determination of the temporal evolution of the values of the system's output variables [White and Frederiksen, 1990][De Jong and Van Joolingen, 1998].

The importance of learning how to construct, use, evaluate, and revise models has been advocated by many researchers [Frederiksen and White, 2002]. In particular, with the importance of addressing environmental problems, there is growing interest in empowering novice learners with little scientific training or expertise to engage in modeling activities to increase their understanding of the local environments and to learn through various experiments. They are also interested in capturing and simulating conceptual knowledge [Huang *et al.*, 2017].

However, many studies have revealed that modeling is a difficult process for novice learners and hampered by lack of domain knowledge [Hogan and Thomas, 2001][Sins *et al.*, 2005][VanLehn, 2013]. Ecological knowledge is inherently heterogenous, including both qualitative and quantitative aspects [Salles and Bredeweg, 2003]. The qualitative aspect of ecological knowledge includes definitions of components,

a set of variables associated with them, and causal relations between the variables associated with the different components [Liem and others, 2013]. For example, learners should be able to answer some questions, such as "How much energy will an alligator acquire from eating a muskrat?" and "What are the relationships between body mass and respiratory rate?" Without the domain's necessary knowledge as a pre-requisite, the learner will not be able to benefit from using and creating models because there is no knowledge to differentiate and integrate [Tennyson and Breuer, 2002].

Of course, large amounts of knowledge about many domains are now readily accessible on the internet. However, much of this general-purpose knowledge is not particular to any specific task and thus difficult to comprehend by many learners. The research question then becomes can we provide access to large-scale domain knowledge in a comprehensible manner? Our research hypothesis is that access to large-scale domain knowledge should be contextualized and that contextualized acquisition of this knowledge may help learners achieve deeper understanding about the domain and generate richer models.

In this paper, we describe the knowledge engineering process of contextualizing large-scale domain knowledge in an online modeling called VERA. The Virtual Experimentation Research Assistant (VERA; vera.cc.gatech.edu; [An *et al.*, 2020]) supports ecological modeling using large-scale domain knowledge through Smithsonian's Encyclopedia of Life [EOL; eol.org; [Parr *et al.*, 2016]] and Global Biotic Interaction [GloBI; globalbioticinteractions.org; [Poelen *et al.*, 2014]]. Specifically, we used the ontology of biotic interactions in GloBI and the trait data of species in EOL to facilitate the model construction. Learners can use VERA to construct qualitative conceptual models of ecological phenomena, run them as quantitative simulations, and review their predictions.

2 Related Work

Much cognitive systems research has explored interactive tools for qualitative modeling and qualitative simulation and their use for promoting science education and research [Bredeweg and Forbus, 2003][Forbus *et al.*, 2005][Leelawong and Biswas, 2008][Bredeweg *et al.*, 2013]. VERA has a close similarity to Forbus et al.'s Vmodel [2005], Bredeweg et al.'s Garp-3 system [2009], and DynaLearn [2013] in that it allows the user to first create qualitative models of ecologi-

cal phenomena and then qualitatively simulate them [Forbus *et al.*, 2005][Bredeweg *et al.*, 2009][Bredeweg *et al.*, 2013]. In contrast to these qualitative modeling tools, VERA uses Component-Mechanism-Phenomena language to assist the process of authoring conceptual models. VERA also uses a translator between the conceptual models and the off-the-shelf NetLogo engine for directly spawning the agent-based simulations from the conceptual models.

VERA is built on the previous work that used Component-Mechanism-Phenomenon (or CMP) causal models [Joyner *et al.*, 2011] that arose from Structure-Behavior-Function models [Goel *et al.*, 2009]. In CMP, Components represent the Structure of the system, emphasizing the physical pieces of an ecological system and their properties. Mechanism represents the Behavior of the system, a sequence of developments linked together in a causal chain that results in some observable Phenomenon. Phenomenon represents what was formerly described as the Function of the system; the Phenomenon is the initial observable event that the investigator aims to explain. VERA is also built on the previous work that used an off-the-shelf agent-based simulation system called NetLogo [Joyner *et al.*, 2014] because agent-based simulations are especially well suited for ecological modeling [Wilensky and Reisman, 2006].

Although the focus of this paper is on the knowledge engineering in VERA, we should note that earlier work has demonstrated VERA's usefulness for enabling learning of domain knowledge [An *et al.*, 2018]. In particular, we found that learners who used the full cycle of model construction, evaluation, and revision created models of higher quality than learners who only constructed conceptual models or only observed the simulations [An *et al.*, 2022].

3 VERA

VERA (Virtual Experimentation Research Assistant) is an online learning environment that enables learners to construct conceptual models of ecological systems and run agent-based simulations of these models. This allows learners to explore ecological systems and perform "what-if" experiments to either explain an existing ecological system or attempt to predict the outcome of future changes to one.

From a human-learning perspective, the key steps in the overall procedure for learning in VERA are as follows. A learner often starts with the identification of an atypical or abnormal phenomenon (e.g., overpopulation). The learner may then develop multiple hypotheses for explaining the phenomenon. The VERA's interface for CMP conceptual modeling coupled with the contextualized domain knowledge helps the learner to express and elaborate his or her hypotheses. The learner may then evaluate his or her hypotheses through the agent-based simulations, which may lead to a revision of his or her hypothesis.

This paper focuses on describe the knowledge engineering process of contextualizing large-scale domain knowledge in VERA. As shown in Figure 1, VERA uses domain knowledge in two ways. First, VERA's taxonomy of interactions among biotic components in a conceptual model is based on the ontology of the interactions used by Global Biotic Interactions

(GloBI). Second, VERA uses trait data from Encyclopedia of Life to suggest the initial parameter values and define the relationships between variables.

This section introduces the design of our system including conceptual models using CMP modeling language (Section 3.1) and an AI compiler that translates the conceptual models into agent-based simulation (Section 3.3). Section 3.2 provides how VERA retrieves and contextualize domain knowledge.
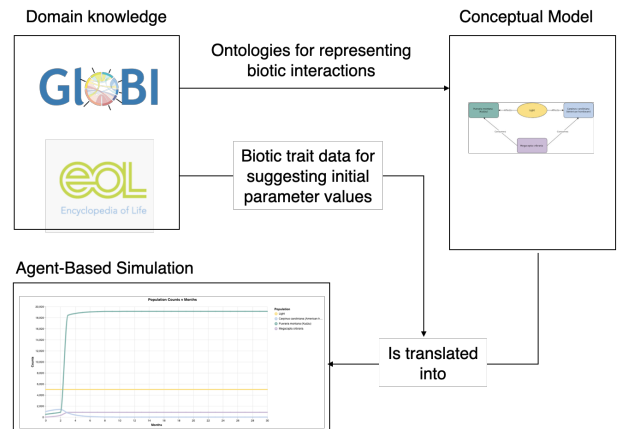

Figure 1: Schematic Overview of Using Domain Knowledge for Conceptual Modeling and Simulation.

3.1 Conceptual Models

The declarative conceptual models of an ecological phenomenon in VERA are based on the Component-Mechanism-Phenomenon (CMP) models. Components in VERA can be either biotic or abiotic, and each component has a set of variables associated with it, thirteen for biotic and three for abiotic components to represent the structure of the system. For example, biotic components are defined by their *lifespan, reproductive maturity, reproductive interval, offspring count, starting population, minimum population, body mass, carbon biomass, respiratory rate, photosynthesis rate, assimilation efficiency, move direction, and move velocity*. Abiotic substances are defined by their *amount, minimum amount, and growth rate*. These properties were selected and adapted to represent ecologically relevant attributes and primitives for ecological modeling (see Table 1).

To describe casual relationships among components in an ecological system, we used the ontology of the interactions from a digital library called Global Biotic Interactions (GloBI) [Poelen *et al.*, 2014]. GloBI provides open access to finding species interaction data (e.g., predator-prey, pollinator-plant, pathogen-host, parasite-host) by combining existing open datasets using open-source software. GloBI provides 22 possible interaction types between species. Among 22 interaction types, we reduced to five interaction types by integrating redundant interactions and extracting "primitive" interactions between components that give rise to the behavior of the system as a whole (see Table 2). Examples of such primitive component–component interactions

Component	Property	Description	EOL Trait Data
Biotic	Lifespan	Average lifespan of organisms in this population in months.	"life span" or "total life span"
	Reproductive Maturity	Age when organisms in this population are able to begin reproduction in month.	"age at first birth","age at first reproduc-tion","age at maturity","onset of fertility","egg laying begins"
	Reproductive Interval	Frequency with which organisms in this population are able to reproduce in months.	"inter-birth interval"
	Offspring Count	Average number of offspring per spawning individual for a reproduction cycle.	"offspring" "litters per year"
	Body Mass	Average body mass per organism.	"body mass" If it is not available, attempt to estimate it based on taxonomic ancestry traits ("body length")
	Carbon Biomass	Average carbon biomass in an individual organism.	"carbon biomass" Attempt to estimate it based on taxonomic ancestry traits ("plant height", "body mass")
	Respiratory Rate	Average basal metabolic rate, measured as respiration (loss) of carbon biomass.	"respiratory rate" If it is not available, attempt to estimate it based on taxonomic ancestry traits ("basal metabolic rate", "body mass")
	Photosynthesis Rate	Average addition of carbon biomass from photosynthesis for a square meter of density-based populations (kg/month).	"photosynthetic rate" If it is not available, attempt to estimate it based on taxonomic ancestry traits ("net carbon fixation rate")
	Assimilation Efficiency	Efficiency of assimilating carbon biomass via consumption (0.0 - 1.0).	Attempt to estimate it based on taxonomic ancestry traits.

Table 1: The Selected Properties of Biotic Components and their Trait Data for Conceptual Modeling.

include *consumes* (one biotic organism consuming another), *produces* (a biotic organism producing an abiotic substance), and *destroys* (an abiotic substance harming a biotic organism).

To instantiate the principles of CMP and better facilitate modeling of ecological phenomena, VERA provides a visual interface for Component-Mechanism-Phenomenon language with a well-defined semantics to represent conceptual models clearly. Figure 2(A) shows a screenshot of the model canvas in VERA where a learner can build a conceptual model by adding biotic (rectangular), abiotic (ellipse), and relationships among them. The learner can draw a directed relationship between two components and choose a pre-dfined relationship type from the drop-down menu. The right-side panel in Figure 2(B) is the modeling and simulation parameter editor. The simulation parameters of each component can affect its simulation behavior.

Figure 2(A) illustrates a causal conceptual model of kudzu (*Pueraria Montana*), kudzu bug (*Megacopta Cribraria*), and American hornbeam (*Carpinus Caroliniana*) in the Southern United State. In this model, there are four components: light (abiotic), kudzu (biotic), American hornbeam (biotic), and kudzu bug (biotic). Regarding mechanisms (relationships), kudzu bug consumes both kudzu and American hornbeam, and light affects both kudzu and American hornbeam. The di-

rection of the arrow between the variables of two components indicates the direction of causal influence. For example, the arrow from kudzu bug to kudzu indicates that the population of kudzu bug consumes the population of kudzu.

A Phenomenon is an observation of the system of interest. For example, the mechanism illustrated in Figure 2(A) can exhibit three different phenomena depending on the manipulation of the kudzu bug population. First, when kudzu bug is low, kudzu grows fast and outcompetes American hornbeam for the shared resource of light, and American hornbeam does not survive the competition with kudzu. Second, when the population of the kudzu bug is adequate (medium), the kudzu population is controlled while American hornbeam also survives (as shown in Figure 2(C)). Lastly, when the population of the kudzu bug is high, the Kudzu and the American hornbeam population both die off due to the large kudzu bug population.

3.2 Contextualize Domain Knowledge

We use EOL's the trait data to suggest initial parameter values for species of interest (see Table 1). EOL is the world's largest aggregated and curated database of species data with almost two million species and eleven million attribute records in the biological domain [Parr *et al.*, 2016]. Over 10 million trait data records are available for 1.7 million taxa. Figure 3 shows
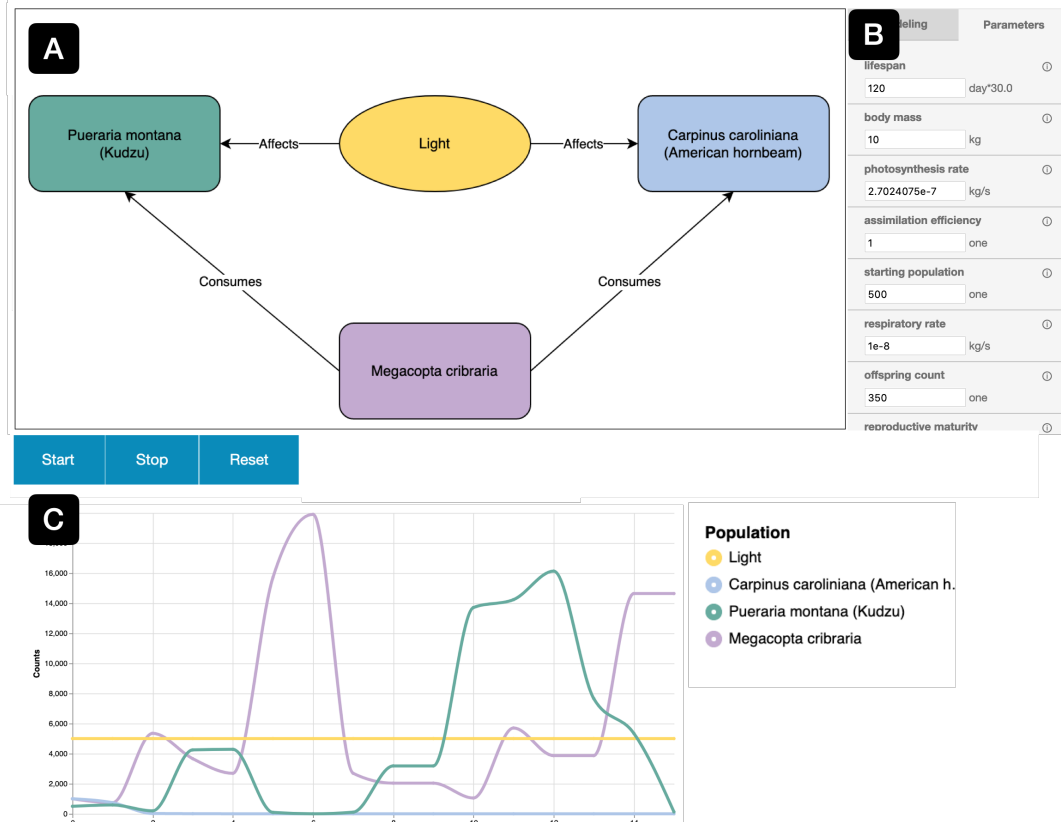
Figure 2: The VERA system. (A) Conceptual Model. (B) Simulation parameters. (C) The Simulation Output Graph – x axis: Time (months); y axis: Population.

a screenshot of trait data of Red Tailed Hawk in EOL. Trait-Bank aggregates trait records from many sources, and for a given trait, multiple records may exist from different studies under a variety of conditions. To support large-scale queries, EOL provides API services that allow queries for on-demand JSON output about EOL taxa, ecological interactions, and organism attributes in the Cypher language.

VERA derives simulation parameters from EOL traits if possible, otherwise attempts to estimate them based on trophical ancestor and its body mass. If the trait data is available in EOL, values are directly derived from the EOL trait data. For example, in Table 1, *lifespan* is retrieved from existing EOL traits such as *life span* or *total life span* (In case of redundancy, the average value is used). *Reproductive maturity* is retrieved from traits such as *age at first birth*, *age at first reproduction*, *age at maturity*, *onset of fertility*, *egg laying begins*. However, if the necessary trait data is not available in EOL, VERA attempts to estimate the value with the information of the species' ancestor. For example, if *carbon biomass* is not available in EOL, it is calculated using its *body mass*: $C_{reptilia}(.122) \times bodymass$ (in case its ancestor is Reptilia in freshwater); $C_{mammalia}(.16) \times bodymass$ (in case of Mammalia). If its taxonomic ancestry traits are not available, VERA provides reasonable default values for ecological plausibility (e.g., $C_{default}(.1) \times bodymass$).

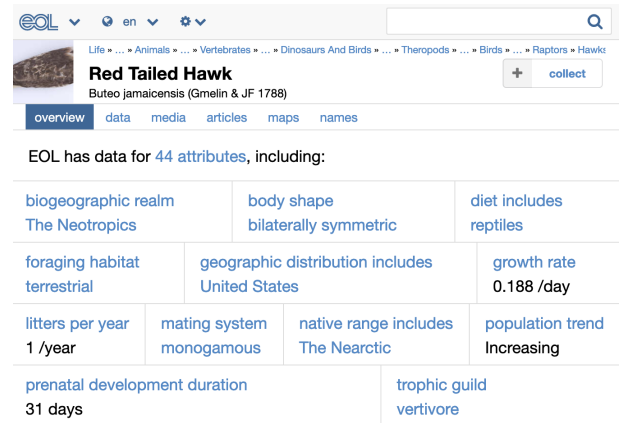VERA provides the contextualized domain knowledge via

Figure 3: Screenshot of EOL trait data of Red Tailed Hawk

"Lookup EOL" feature. Figure 4 shows the process of adding a biotic component via Lookup EOL. (1) The learner clicks on "LookUp EOL" button in VERA and queries species name (either scientific or common names). (2) Then the system returns a list of species names that matches the input via EOL Search API. (3) Then, the learner selects one species from the list, and the system calls EOL TraitBank API for retrieving specific traits of the species. Our inference engine uses the

Relationship	Property	Description	GloBI Interactions
X Consumes Y	Consumption rate, interaction probability	When X interacts with Y, it will partially or wholly consume Y, with carbon transfer to X from Y.	"eat", "get eaten by", "preys on", "get preyed on by"
X Destroys Y	Destruction rate, interaction probability	When X interacts with Y, it partially or wholly destroys a simulation entity of type Y with no carbon transfer to X.	"kill", "is killed by", "parasi-tize", "get parasitized by", "get infected by"
X Produces Y	Production rate	X will produce Y with some stochastic timing and amount.	"visits flowers of", "flowers visited by", "pollinate", "get pollinated by", "spread", "get spread by."
X Affects Y	Growth rate, interaction probability	This is a generic growth modifier that allows for growth rates (negative or positive) to modify Y when X interacts with it, where none of the above relationships apply.	"interacts with" (+, -), "related to" (+, -), "parasitize" (-), "get parasitized by" (-), "hosts", "get hosted by."
X Becomes Y on Death	Percent body mass	When X expires, it produces Y.	-

Table 2: The Taxonomy of Interactions among Components.

retrieved traits to preset the simulation parameters.

3.3 Translating Conceptual Models into Simulations

Following our earlier work on the MILA-S system [Joyner *et al.*, 2011][Joyner *et al.*, 2014][Goel *et al.*, 2016], VERA uses an artificial intelligence compiler to automatically translate the patterns in the conceptual models into the primitives of agent-based simulation of NetLogo. In other words, VERA automatically generates the simulations directly from learner's casual model. After constructing a CMP conceptual model, the learner uses the suggested initial parameter values by EOL for simulation generation (as described in Section 3.2).

Figure 5 illustrates mechanisms for translating the semantics of CMP conceptual models into the semantics of the Netlogo agent-based simulations, which is composed of four steps. First, the compiler extracts conceptual model from visual layout via mxGraph java library. Then it is represented as VeraWeb Domain Model to handle high level simulation concepts. Next, Domain Simulation Builder decomposes high-level domain-specific behaviors and logic into domain independent simulation operations. Finally, the resulting simulation abstract syntax tree (AST) and abstract semantic graph (ASG) compiles abstract simulation into target simulation native constructs such as NetLogo.

We developed detailed knowledge representations for each component (agent)-component (agent) interaction that will specify how to set up the corresponding simulation. This includes knowledge about the mechanism by which an agent–agent interaction occurs as described in Table 2; for example, in the case of consumption, the consuming organism receives energy and nutrients while the consumed organism perishes. It will also include numerical information relating the various parameters of the mechanism. Some of this information will be specific to the simulation engine so that the AI

compiler can in fact automatically generate the corresponding simulation in the given simulation platform.

The components and properties in VERA (in Figure 2) are translated to NetLogo models [Wilensky and Reisman, 2006]. Biotic components in VERA are translated to Turtles–the primary agents in NetLogo, and their attributes are also translated back to turtles' attributes. For example, carbon parameters, *carbon biomass* (the amount of carbon), *respiratory rate* (loss of carbon biomass), *photosynthesis rate* (gain of carbon biomass), and *assimilation efficiency* (percentage of carbon biomass retained by the consumer) in VERA are calculated to measure *energy* in NetLogo. Spatial parameters, *move velocity* and *move direction*, in VERA are translated to *xcor*, *ycor*, and *heading* in NetLogo (The initial positions are randomly initialized in VERA).

Figure 2(C) illustrates the time-series graph of the NetLogo simulation results for the conceptual model in Figure 2(A). Note that all four components of the causal model are represented in the simulation: the Light in yellow, Kudzu in green, American Hornbeam in light blue, and Kudzu Bug in purple. Before running a simulation, the user clicks the Reset button to apply input parameters to a new simulation. The user next clicks the Start button to start the time steps of the simulation and clicks the Stop button to pause the simulation.

We should note that VERA is a conceptual modeling tool specifically designed for non-scientists with little scientific training or expertise. Therefore, the core aim of VERA's agent-based simulation is not to provide accurate scientific predictions, but to give learners the ability to manipulate various variables and to see for themselves how the changes may affect the system. By translating conceptual models into simulations, VERA facilitates the higher-level rapid model revision process. The running of the simulation enables the learner to observe the evolution of the system variables over time, and iterate through the model-simulate-refine loops. In this way, VERA integrates both qualitative reasoning in the
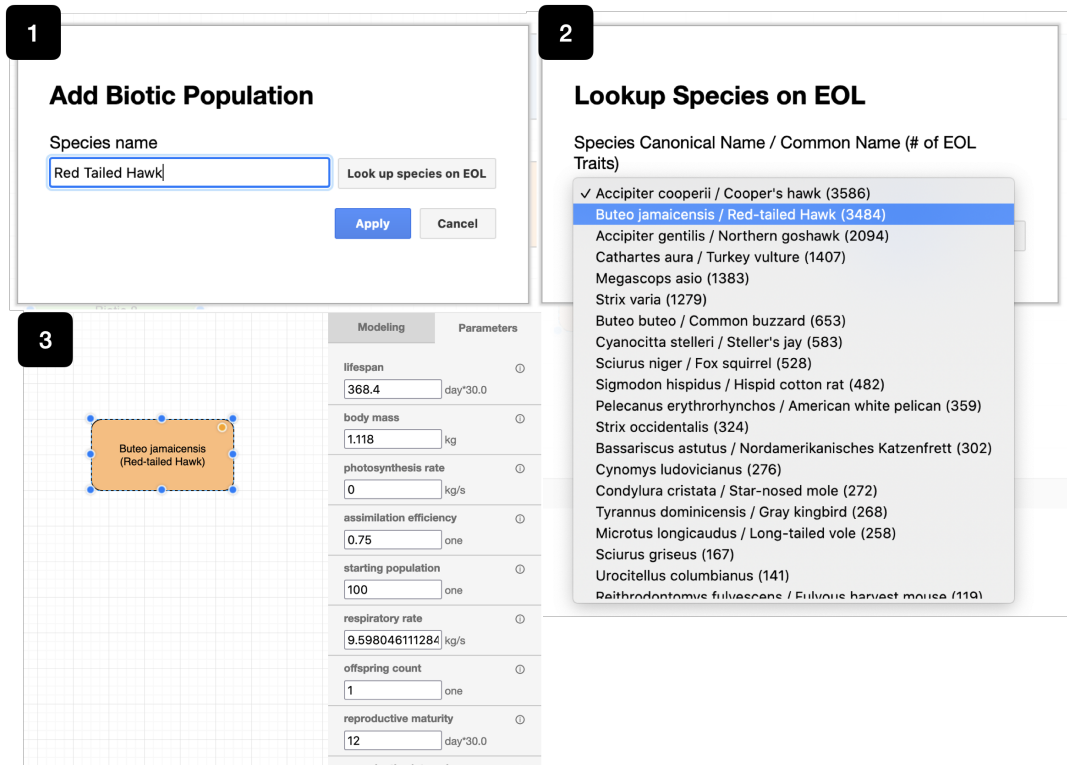
Figure 4: Using EOL TraitBank Data to Set Up Simulation Values.

conceptual model and quantitative reasoning in the simulation reasoning on one hand, and explanatory reasoning and predictive reasoning.

4 Conclusions

Modeling requires domain knowledge, e.g., relationships between variables describing the system being modeled. The research question in this work is how might we provide access to large-scale domain knowledge for learners engaged in constructing qualitative models of ecological systems in a comprehensible manner? The research hypothesis is that contextualizing large-scale domain knowledge can scaffold the task of authoring conceptual models. VERA contextualizes EOL's large scale domain knowledge to support modeling of ecological systems. It also uses Globi's ontology of ecological interactions to ground VERA's CMP models.

The VERA system has been introduced and evaluated in various contexts including college courses [An *et al.*, 2018] and self-directed learning on the web [An *et al.*, 2022]. Additionally, our data analysis indicates promising results on the usefulness of contextualized domain knowledge in building conceptual models. Based on theses results, the contextualized domain knowledge is expected to help novice learners to use both qualitative and quantitative aspects of ecological knowledge to model, analyze, explain, and predict problems in ecological systems.

References

[An *et al.*, 2018] Sungeun An, Robert Bates, Jennifer Hammock, Spencer Rugaber, and Ashok Goel. Vera: popularizing science through ai. In *International Conference on Artificial Intelligence in Education*, pages 31–35. Springer, 2018.

[An *et al.*, 2020] Sungeun An, Robert Bates, Jen Hammock, Spencer Rugaber, Emily Weigel, and Ashok Goel. Scientific modeling using large scale knowledge. In *International Conference on Artificial Intelligence in Education*, pages 20–24. Springer, 2020.

[An *et al.*, 2022] Sungeun An, Spencer Rugaber, Jennifer Hammock, and Ashok Goel. Understanding self-directed learning with sequential pattern mining. In *Interna-*
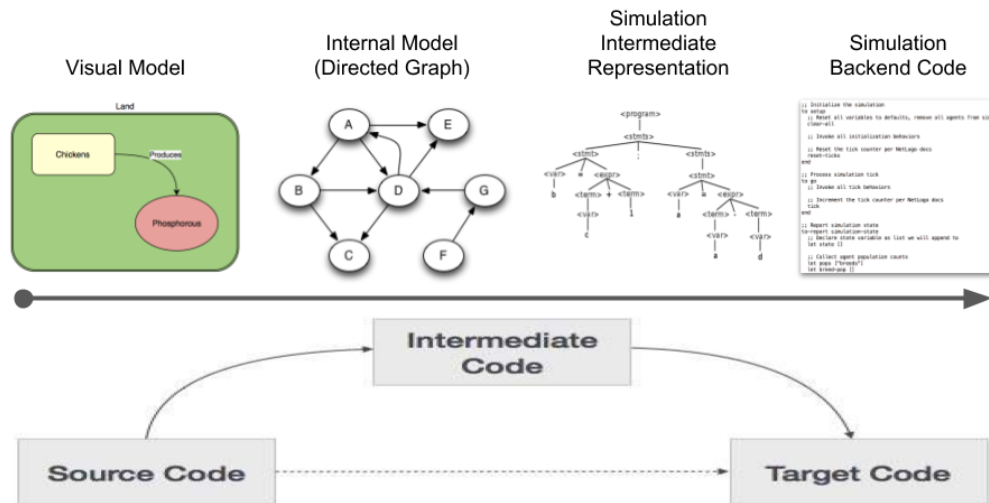
Figure 5: Mechanisms of the Visual Conceptual Model to a Simulation.

tional Conference on Artificial Intelligence in Education. Springer, 2022.

[Bredeweg and Forbus, 2003] Bert Bredeweg and Kenneth D Forbus. Qualitative modeling in education. *AI magazine*, 24(4):35–35, 2003.

[Bredeweg et al., 2009] Bert Bredeweg, Floris Linnebank, Anders Bouwer, and Jochem Liem. Garp3—workbench for qualitative modelling and simulation. *Ecological informatics*, 4(5-6):263–281, 2009.

[Bredeweg et al., 2013] Bert Bredeweg, Jochem Liem, Wouter Beek, Floris Linnebank, Jorge Gracia, Esther Lozano, Michael Wißner, René Bühling, Paulo Salles, Richard Noble, et al. Dynalearn–an intelligent learning environment for learning conceptual knowledge. *AI Magazine*, 34(4):46–65, 2013.

[De Jong and Van Joolingen, 1998] Ton De Jong and Wouter R Van Joolingen. Scientific discovery learning with computer simulations of conceptual domains. *Review of educational research*, 68(2):179–201, 1998.

[Forbus et al., 2005] Kenneth D Forbus, Karen Carney, Bruce L Sherin, and Leo C Ureel II. Vmodel: A visual qualitative modeling environment for middle-school students. *Ai Magazine*, 26(3):63–63, 2005.

[Forbus, 1984] Kenneth D Forbus. Qualitative process theory. *Artificial intelligence*, 24(1-3):85–168, 1984.

[Frederiksen and White, 2002] John R Frederiksen and Barbara Y White. Conceptualizing and constructing linked models: Creating coherence in complex knowledge systems. *The role of communication in learning to model*, pages 69–96, 2002.

[Goel et al., 2009] Ashok K Goel, Spencer Rugaber, and Swaroop Vattam. Structure, behavior, and function of complex systems: The structure, behavior, and function modeling language. *Ai Edam*, 23(1):23–35, 2009.

[Goel et al., 2016] Ashok K Goel, David A Joyner, and Taylor J Hartman. Preparing mila-s for college. In *IJCAI 2016 Workshop on Qualitative Reasoning*, 2016.

[Hogan and Thomas, 2001] Kathleen Hogan and David Thomas. Cognitive comparisons of students' systems modeling in ecology. *Journal of Science Education and Technology*, 10(4):319–345, 2001.

[Huang et al., 2017] Joey Huang, Cindy Hmelo-Silver, Rebecca Jordan, Troy Frensley, Steven Gray, and Greg Newman. Scientific discourse of citizen scientists: A collaborative modeling as a boundary object. Philadelphia, PA: International Society of the Learning Sciences., 2017.

[Joyner et al., 2011] David A Joyner, Ashok K Goel, Spencer Rugaber, Cindy Hmelo-Silver, and Rebecca Jordan. Evolution of an integrated technology for supporting learning about complex systems. In *2011 IEEE 11th International Conference on Advanced Learning Technologies*, pages 257–259. IEEE, 2011.

[Joyner et al., 2014] David A Joyner, Ashok K Goel, and Nicolas M Papin. Mila–s: generation of agent-based simulations from conceptual models of complex systems. In *Proceedings of the 19th international conference on intelligent user interfaces*, pages 289–298, 2014.

[Leelawong and Biswas, 2008] Krittaya Leelawong and Gautam Biswas. Designing learning by teaching agents: The betty's brain system. *International Journal of Artificial Intelligence in Education*, 18(3):181–208, 2008.

[Liem and others, 2013] Joachim Liem et al. *Supporting conceptual modelling of dynamic systems: A knowledge engineering perspective on qualitative reasoning*. Universiteit van Amsterdam [Host], 2013.

[Parr *et al.*, 2016] Cynthia S Parr, Katja S Schulz, Jennifer Hammock, Nathan Wilson, Patrick Leary, Jeremy Rice, and Robert J Corrigan Jr. Traitbank: Practical semantics for organism attribute data. *Semantic Web*, 7(6):577–588, 2016.

[Poelen *et al.*, 2014] Jorrit H Poelen, James D Simons, and Chris J Mungall. Global biotic interactions: An open infrastructure to share and analyze species-interaction datasets. *Ecological Informatics*, 24:148–159, 2014.

[Salles and Bredeweg, 2003] Paulo Salles and Bert Bredeweg. Qualitative reasoning about population and community ecology. *AI Magazine*, 24(4):77–77, 2003.

[Schwarz and White, 2005] Christina V Schwarz and Barbara Y White. Metamodeling knowledge: Developing students' understanding of scientific modeling. *Cognition and instruction*, 23(2):165–205, 2005.

[Sins *et al.*, 2005] Patrick HM Sins, Elwin R Savelsbergh, and Wouter R van Joolingen. The difficult process of scientific modelling: An analysis of novices' reasoning during computer-based modelling. *International Journal of Science Education*, 27(14):1695–1721, 2005.

[Tennyson and Breuer, 2002] Robert D Tennyson and Klaus Breuer. Improving problem solving and creativity through use of complex-dynamic simulations. *Computers in Human Behavior*, 18(6):650–668, 2002.

[VanLehn, 2013] Kurt VanLehn. Model construction as a learning activity: A design space and review. *Interactive Learning Environments*, 21(4):371–413, 2013.

[White and Frederiksen, 1990] Barbara Y White and John R Frederiksen. Causal model progressions as a foundation for intelligent learning environments. *Artificial intelligence*, 42(1):99–157, 1990.

[Wilensky and Reisman, 2006] Uri Wilensky and Kenneth Reisman. Thinking like a wolf, a sheep, or a firefly: Learning biology through constructing and testing computational theories—an embodied modeling approach. *Cognition and instruction*, 24(2):171–209, 2006.