MODELING AND INQUIRY LEARNING APPLICATION: AN EXPLORATION OF SPATIAL SIMULATIONS IN AN INQUIRY LEARNING APPLICATION

A Thesis
Presented to
The Academic Faculty

by

Taylor J. Hartman

In Partial Fulfillment
of the Requirements for the Degree
Bachelor of Science with the Research Option in the
School of Computer Science

Georgia Institute of Technology
May 2017

MODELING AND INQUIRY LEARNING APPLICATION: AN EXPLORATION OF SPATIAL SIMULATIONS IN AN INQUIRY LEARNING APPLICATION

Approved by:

Professor Ashok Goel, Advisor
Department of Computer Science
Georgia Institute of Technology

Date Approved: May 2017

I dedicate this thesis to my family for all of the love and support they have given me throughout the years, to Dr. Ashok Goel for being a wonderful mentor and always pushing me to dig deeper and explore more, and to my team, GT Wreck, for providing the much needed outlet and support away from home.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

Modeling and simulation has been a common technique to explore phenomena in ecology without having to spend an enormous amount of time with data collection and waiting on an ecosystem to develop. The practice of simulating an ecological phenomenon requires a good amount of practice and ability in scripting and programming. A lack of knowledge in the domain of computer modeling and simulation leads to creating a working model and simulation of an ecological phenomenon to be difficult. Additionally, simulations are created on a per-project basis. Each simulation has dozens of specific parameters that are implemented and adjusted by the creator of the simulation in order to get the simulation to perform well and produce results. The result of this practice is many variations of one simulation whose changes are specific to the given project the simulation was created for. A lot of simulations are built from scratch or from code snippets from other sources and the communication of how these simulators work tend to be either difficult to understand and extremely unclear or altogether nonexistent which was pointed out by Grimm, et. al [7].

The current Modeling and Inquiry Learning Application (MILA) system solves a lot of these problems when attempting to model food chain ecology. A user of MILA uses the interface to create a conceptual model that is compiled down into a simulation to be run in the NetLogo simulator. MILA allows a user to conceptually model a food chain without being bogged down by the overhead of also setting up a simulation of the food chain itself. MILA takes the conceptual model that has been created and interprets it in such a way so that it can be simulated in NetLogo. The MILA system has already been found useful and produces excellent results in the domain of education and teaching ecological phenomena in a classroom setting[3], but its applicability to real world biological and ecological exploration and simulation is limited by its current implementation. In order to add functionality to the system and expand MILAs applicability to the scientific community, this project integrates a spatially explicit simulator into MILA and creates a new interface to allow conceptual modeling of spatial relationships. This integration is a huge step in the direction of being able to model and simulate interactions of entire ecosystems of biological species and abiotic materials. This integration is the first attempt to model a meta-ecosystem- a set of ecosystems connected by spatial flows of energy, materials, and organisms.

1.1 Approach

The Modeling and Inquiry Learning Application-Simulation (MILA-S) platform was originally designed as a classroom tool that allowed students to model an ecological food chain system and then run a simulation based off of the model. The strength of the current system is that it compiles for the user a simulation from a conceptual model, allowing users with very little simulation-writing experience to run simulations based off of the model they created. Further, MILA-S encourages scientific inquiry learning in users of the application in a middle school setting.

In an experiment performed with middle school students, MILA-S was shown to significantly increase students' scientific inquiry ability. However, when a similar experiment was run in a college setting, there was no increase in the students' inquiry abilities. While there are a few possible reasons that this experiment did not produce the expected results, one main reason that was focused on was that the conceptual model and resultant simulation was not complex enough to have a significant effect on the scientific abilities of a student in higher education. In order to make a more complicated simulation and conceptual model, two things needed to be accomplished:

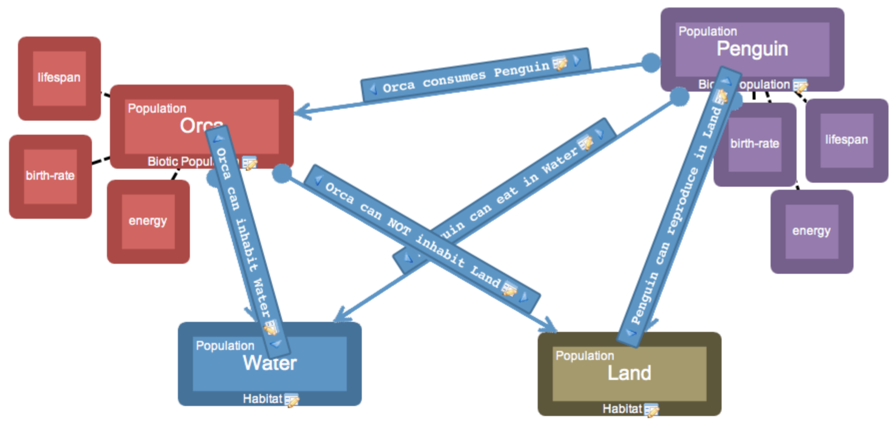
1. create stable simulations in the current MILA-S system

2. integrate a spatial component into the current MILA-S system

At the time of these early experimentations it was difficult to produce stable simulations that would result in the expected predator-prey consumer cyclic behavior. The simulations resulted in either a single organism's population exploding and causing the simulation to crash or all organisms quickly dying off. It was important to be able to set up a stable simulation in MILA so that the user could have a base simulation to work off of when tweaking or adding things to the environment and seeing how it affected the ecosystem. This was mediated in early Spring 2016.

Introducing a spatial component into the current MILA-S system would allow for more complicated simulations and models to be created. Spatial components like habitats and defining where an organism is allowed to move would allow users to start seeing even more ecological phenomena be produced by the system such as boundary effects. The next step of introducing a spatial component into MILA-S was accomplished in mid-Spring 2016. This system will be referred to as MILA-Spatial. In order to achieve this integration the same model for the rest of the conceptual model compiler was used in order to fit a spatial simulator onto the system. This approach worked and allowed users to define habitats and boundaries using the same approach as they were able to use in the original MILA-S.

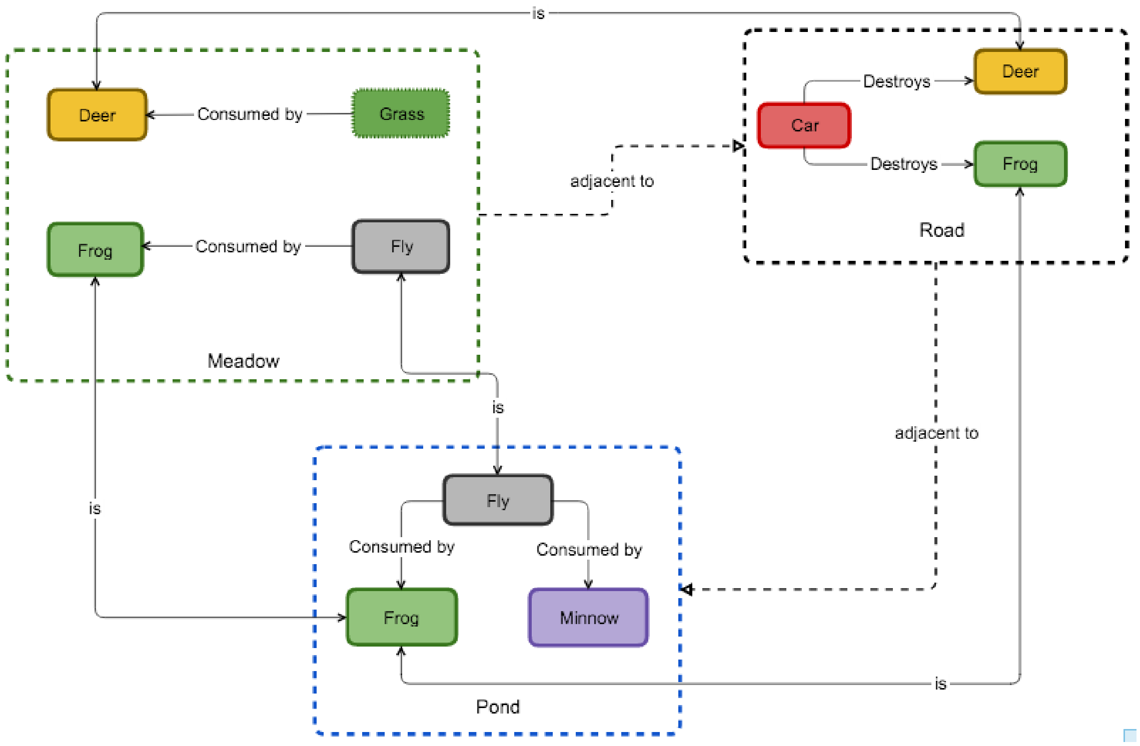
However, the introduction of spatial components revealed some fundamental shortcomings of the causal modeling language MILA-S uses. As can be seen in the image below, just two organisms and two habitats result in a messy conceptual model:



One can imagine an overwhelming model that contains five organisms and three habitats. With the current conceptual modeling language being used, it is impractical to create larger and more complex ecosystems that contain spatially explicit components purely due to the conceptual model quickly getting out of hand. Because of the shortcomings in the causal modeling language mentioned above,it has come to my attention that the Structure-Behavior-Function (SBF) modeling language seems to be missing a definition and representation of spatially explicit components of an environment or system and further how to relate these new components to the existing components in the model. It is not enough to just categorize a spatially explicit concept as a structure in SBF. The behaviors ascribed to structures in SBF do not cover all of the important interactions that a spatially explicit node can have with other spatial nodes as well as nodes representing agents in the environment. To elaborate, in SBF, the structures are ascribed functions that result in the behaviors of the entire system. However, the language does not support situations in which many structures' ascribed behaviors or functions are dependent on other structures. For example, a deer's behavior may be dependent on its spatial positioning (restricting deer to only reproduce on land). This type of relationship that relies on structures' properties or behaviors being dependent on another structures' existence or location relative to it is not present in current causal modeling languages. Therefore, in order to be able to model an ecosystem with spatial dimensions, a new causal modeling language is needed to fully express all of the relationships structures can have with each other.

MILA currently makes use of a flavor of SBF called Component-Mechanism-Phenomenon (CMP). In order to achieve the previously described changes to the conceptual modeling language, CMP needed to be edited and restructured into a new causal modeling language and visual representation that includes a notion of "interaction" between components. These interactions define relative behavior based on the spatial relationships of components. Additionally this would allow components to have spatial properties as well. One example of a spatial property a non-spatial node could have would be a property that results in flocking behavior. The incorporation of interactions also allow for ambient properties to exist in the model. Ambient properties like temperature or season do not fit into the current modeling scheme but are important for modeling a complex ecosystem. Ambient components of a system would be able to define behaviors and properties of other components through this new concept of interaction.

It is necessary to change the visual representation of the model in order to accommodate the planned changes in the language. Below I have provided an example of what a system might look like composed of three habitats, five biotic nodes, and one abiotic node.



The big changes between the current visual representation of the model present in MILA and this new model are

- Habitat nodes are expanded to encompass the biotic or abiotic nodes that can live/exist in the habitat. This allows for an easy way to represent where biotic nodes can travel or be in without having a separate link for each habitat.

- Habitat nodes now have spatially explicit links that allow relationships between the habitats themselves. These links describe spatial layouts in the above example but could also be used to further describe shared boundaries or other relationships between spatial landscapes.

- Currently there is a new link that allows two nodes to be connected and described as being the same node. This link is present for clarity.

- Biotic nodes that are associated with a particular habitat and are immobile have a border style that matches the habitat's color. This is an implementation that allows for the inclusion of plants in an ecosystem model.

These additions to the system are only the beginning. As the system is iteratively developed and more feedback is received both in reference to the model itself and possible ecological concepts that can be integrated, this new system has the potential to be able to structure and simulate fairly complicated models. After the system has reached a beta version in development, an experiment on college-level students will be run again and the results compared to those of the previous experimental runs.

In conclusion, the current conceptual model being used needs to be changed or updated in order to accommodate spatially explicit features of a system. These changes will be integrated into a new version of MILA called MILA-Hierarchical Spatial Simulator or MILA-HSS. This work would have implications reaching farther than just the realm of ecology or education, however. Creating a conceptual modeling language that is able to represent spatially explicit components in a system has uses in other fields of science and technology, including but not limited to the problem of modeling relationships of magnets in the context of the physical sciences. With further work on this new causal modeling language and conceptual model representation, we believe that there is high potential for this research to have a resounding impact in the context of conceptual models and scientific exploration in education and field work.

CHAPTER II

LITERATURE REVIEW

Inquiry-driven modeling of scientific phenomenon is an important process in how scientists make sense of natural phenomena [5]. Modeling and simulation on its own has a storied importance in the scientific community. Scientists use models to make sense of phenomena then run simulations to test hypotheses that create these models. In the current state of the science, there are plenty of unique simulators that can be used for a myriad of different purposes. In general, simulations are created on a per-research question basis, resulting in a large basis of simulations that are created and recreated repeatedly and also require a high level of expertise to both reproduce and even run in the first place [7]. For citizen scientists, this use of high level modeling and simulation may not be easily accessible. There is always a need for an increase in the number of tools that can be used to facilitate scientific inquiry and discovery [3]. Inquiry based modeling is a cornerstone type of modeling that some think accurately describe how scientists go about forming and revising hypotheses. These models allow scientists to abstract a reality to a model so that they can more easily reason over it and generate explanations of observations and allow for clearer organization, evaluation, and expansion of current understanding [1]. The Modeling and Inquiry Learning Application (MILA) is a tool being developed to allow for this process of scientific discovery to be easier by removing the middle layer of the process of creating the simulation parameters [3].

The next step in the development of the MILA system is to integrate spatial simulation to allow for the habilitation of a new classification of scientific discoveries. Crook, Castle, and Batty describe how GISs can be used more widely by integrating them into agent-based models [1]. The problem with working with GIS data and trying to create models from GIS data is that it requires a high level of expertise in both simulation creation and data modeling. Creating a front-facing application that removes the requirement of this level of expertise makes way for citizen scientists to be able to contribute to their local scientific communities. This would also greatly improve teaching these types of ideas in a classroom setting [3], but for this research project we will be focusing on a contribution to the scientific community.

However, modeling spatial representations of ecosystems to present and analyze data is no easy task. For example, how do we connect the ideas of the ecosystem and environment to the predator/prey lifecycle of individual species interacting with each other and their environment? The two leading theories of spatial ecology models were food web meta-communities and landscape ecosystem ecology. Food web meta-communities focused on movement in the model as movement via traits. They focused mainly on predation and ignored abiotic processes and materials. Landscape ecosystem ecology focused on geographical structure of ecosystems, movements of materials and energy among ecosystems, and how it affects the functioning of an ecosystem. Landscape ecosystem ecology traditionally explained patterns as opposed to predicted outcomes. Massol et. al. introduced the idea of a meta-ecosystem to simplify these relationships. A meta-ecosystem is a set of ecosystems connected by spatial flows of energy, materials, and organisms[4]. This provides a brand new framework that links the two main spatial models of ecology into one, providing the modeling strategy that will be adopted and implemented in this research.

In addition to questions concerning precise implementation, there are other salient problems in creating an Agent-Based Model for Geo-Spatial Simulation[1]. These problems are addressed as follows. In regards to the first challenge described, the modeling that will occur will be based off of conceptual models created by a user to explore a certain state space and different combinations of conceptual models.

The goal of the models will not be to predict possible outcomes of tweaking different parameters. This project is focused on scientific discovery and less about accurate predictions. Second, domain knowledge for the application will be provided by Encyclopedia of Life and other sources, such as geographical databases, which will result in a modeling system that is specific to ecology. This system will be tailored to suit the needs of current ecologists and educators and would not be concerned with being generalized to other domains' needs. Concerning agent representation, the agents in each model will be defined as moving agents and represent biotic and abiotic substances present in an ecosystem. For the fourth challenge of validity, the subsequent MILA system will be implemented around answering use-case questions. After the system is developed to answer a few specific instances and research questions, the system will be generalized and abstracted to a level where it can be used more generally to answer a wide variety of research questions.

CHAPTER III

EXPERIMENTAL DESIGN

The purpose of this experiment is to determine the validity of the integration of a spatially explicit modeling language and system in the current MILA system and to ensure that the system is still useable. In this experiment the system that was used was MILA-Spatial, the integration of spatial components with the original MILA-S system. This system was used as opposed to the MILA-HSS system due to the MILA-Spatial system being much farther along in development and usability than MILA-HSS. It was determined that the difference in development stages between MILA-S and MILA-HSS would negatively impact users' experiences with MILA-HSS this skewing the data.

The hypothesis is as follows: the introduction of a spatially explicit modeling language to use in exploring spatial relationships among organisms allows for a greater number and complexity of model hypotheses therefore allowing for greater scientific exploration. Using a spatially explicit modeling language will allow more hypotheses about a problem to be explored as opposed to not having a spatially explicit modeling language available.

The hypothesis will be measured through the creation of hypotheses about the specific experiments the subjects are running. They will be allowed to log their hypotheses and the results of their experiments either on paper or through MILAs interface.

3.1 Materials

Participants were evaluated based on an evaluation created by our lab to asses the effects of the system. The full evaluation questionnaire is presented in the appendix.

3.2 Recruitment of Participants

Participants were current students attending the Georgia Institute of Technology. They were recruited on a volunteer basis from various sources including different departments and schools in order to gain a wide variety of participants and not just students familiar with biology. Participants were not compensated for their time and were entirely volunteers. All of the participants were recruited as a convenience sample.

All participants were female undergraduate students from the Georgia Institute of Technology aged 19-23 years old. All participants were familiar with computers and adapting to new technologies and were all currently studying in STEM fields. Therefore this sample was fairly homogeneous.

3.3 Procedure

The experimental procedure is as follows in the list below. All participants filled out a pre- and post-questionnaire as well which can be found in the appendix.

1. Give the problem statement.
2. Have them set up and run experiments using the original MILA-S system. As they perform these experiments they record their hypotheses.
3. Have them set up and run experiments using the MILA-Spatial system that is the subject of this paper. As they perform these experiments they record their hypotheses.

Half of the participants in the study followed the above procedure and half of the participants followed the procedure but used MILA-Spatial system before the original MILA-S system. This was done so as to avoid any ordering effects in the results.

The participants are asked questions about the system and about ecology. These questions and their answers are discussed in the results section.

The problem statement that was presented to participants was that of the phenomenon of Adelie penguin breeding colonies disappearing in the Antarctic. The full problem statement can be found in the appendix.

This problem statement was chosen due its inherent spatial components.

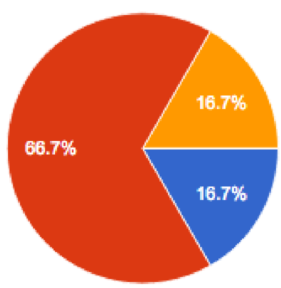
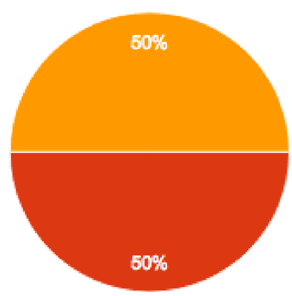
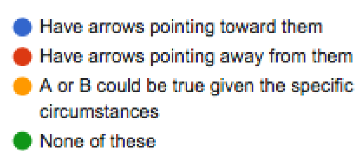
CHAPTER IV

RESULTS AND DISCUSSION

4.1 Ecological Modeling and Simulation Knowledge Results

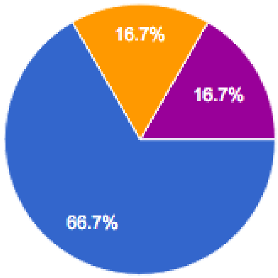
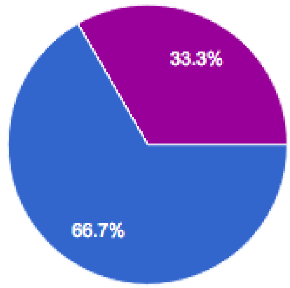
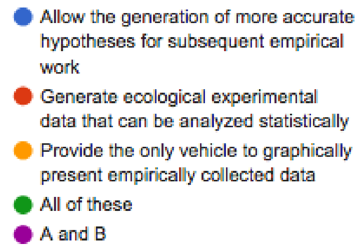
There were questions in the pre and post questionnaires intended to measure the participants' knowledge of ecological ideas and modeling and siulating these ideas. Descriptions of the questions and their results follow.

- The first question measured the participants' understandings of how an ecological diagram of a food chain works.

pre-questionnaire	post-questionnaire	answers
		

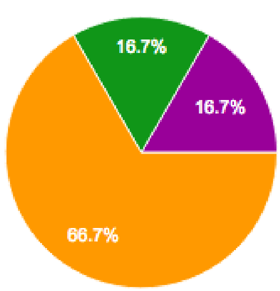
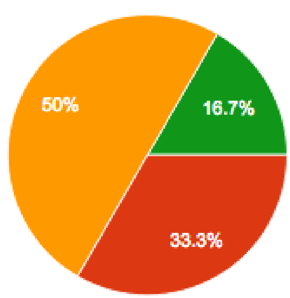
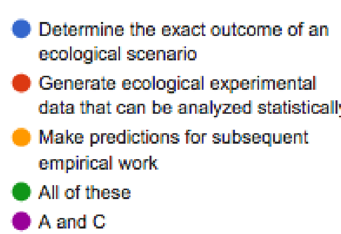
  The correct answer is shown with the color red, or selection B from the questionnaire. This does not show anything statistically significant but there seems to be an indicator that the system does not necessarily help the users learn the correct answer to this question.

- The second problem inquired about the users' understandings of how a concpetual model can be incorporates into the scientific method.

pre-questionnaire	post-questionnaire	answers
		

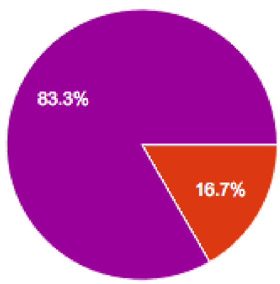
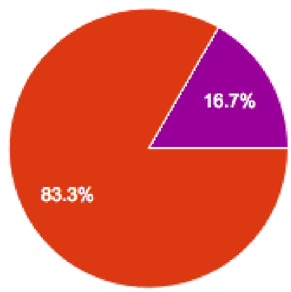
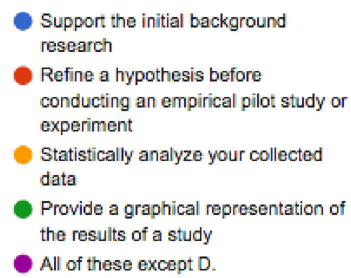
The correct answer is A, the blue selection. The percentage that chose this selection remained the same from pre-experiment to post-experiment. However, the participant that chose C instead chose the last selection which included the correct answer, A. This is interesting and shows that there may be a misunderstanding somewhere. For exmple it could be the case that the participants are not familiar with statistical analysis of experimental data. Alternatively it could be the case that the wording of this question is not clear to the participants. The selection of considering B to be just as correct as choice A may be contributed to by the provided graphings that the system does for the user. The user may interpret this as empirical data.

- Question three tests the users' knowledge of how a simulation model can be incorporated into the scientific method.

pre-questionnaire	post-questionnaire	answers
		

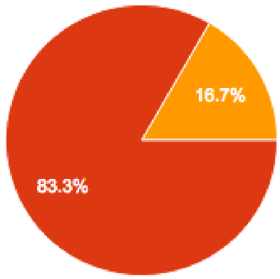
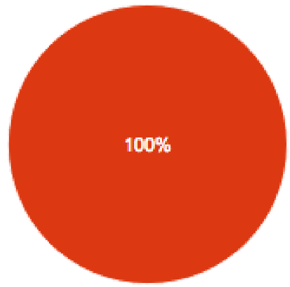
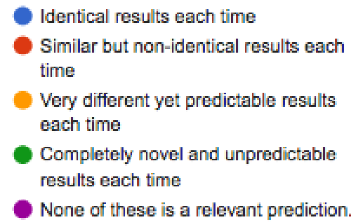
Here we see a negative result between the pre- and post-questionnaires. The correct selection is the selection shown in yellow - part C. There is a reduction in the selection of the correct answer as well as a decrease in selection of an option that inludes the correct answer. However, these changes are not statistically significant.

- Question four tests the users' understanding of how a conceptual model paired with a simulation model can be incorporated into the scientific method.

pre-questionnaire	post-questionnaire	answers
		

The correct answer is represented by the red color, option B. The difference on this question from the pre- to post-questionnaire is more drastic than any of the other results analyzed so far. This result is statistically significant with a comparative error rate of 42.21. This shows that through the use of the systems the participants were able to gain an intuition concerning the scientific value of models and simulations of processes.

- This question tested users' understandings of how simulations work and the impications of running the same simulation multiple times.

pre-questionnaire	post-questionnaire	answers
		

The correct selection is shown in orange. For reference to the questionnaire it is selection B.

These results show that while most participants were able to answer the question correctly on the pre-questionnaire, the participant that did not correctly answer this question was able to after using the systems. This result is not statistically significant but good to note nonetheless and would be interesting to keep in mind when a larger study is performed.

4.1.1 Results

The results from the pre- and post-questionnaires concerning ecological knowledge and understanding of models and simulations overall were not statistically significant. However, there were some interesting trends displayed that can be looked at and explored further in future experimentation.

4.2 Model Complexity Analysis

The average number of hypotheses explored in the spatially explicit version of MILA was 2.5. The average number of hypotheses explored in the original MILA system was 2.8333. Therefore I cannot accept the hypothesis that the spatially explicit system would foster greater number of hypotheses explored. However this might not necessarily be indicative of a problem with the system. For example, many of the hypotheses explored in the spatially explicit version of MILA were centered around habitat loss as well as population dynamics. This indicates that it was able to help model the environment in a way that the user was desiring. For a future study it would be beneficial to provide a problem statement that is less obviously focused on habitat loss and actually a little more ambiguous possibly with more concrete results. This might allow the user to truly explore more hypotheses as oppose to settling on a possible answer after a few iterations of model to simulation experimentation.

4.3 Preference of System

66.66% of the participants preferred using the simulator with the spatial component added to it. The participants that preferred the original MILA system gave reasons such as "It seemed like the fluctuations in populations among competing species had a greater influence in the population of the Adelie penguins" and "It was simpler, so it was easier to keep track of which variables I was changing." The first quote indicates that the participant felt that the population dynamics were more important in finding the final hypothesis than the spatial components. Overall these two participants felt that the spatial component did not contribute to their experimentation. This could be for many reasons- an ill defined problem statement, a problem that does not require the spatial relationships to be modeled in order to be effective, etc. Additionally this could indicate that in order to be useful the spatial implementations need to be more robust. On the other hand the participants that chose the spatially-infused version of MILA cited an ability to add more detail and the ability to conform the model better to the problem statement as reasons for preferring it. This would be an interesting avenue to explore in a larger study since the results are fairly close.

CHAPTER V

CONCLUSION

This pilot study was performed with a very small sample size of 6. Because of this none of the results are statistically significant except one, and even this one result cannot be taken at face value due to the sample size. However, the overall results of the study show trends that are worthwhile to explore in a larger study.

First there is a preference for the spatial simulator in MILA-Spatial over the original simulator in MILA-S. This preference was marked with comments about how it fit better with the provided problem statement. This shows MILA-Spatial's ability to model real-world environments and phenomenon not only on a surface level, but these interactions worked well enough in the simulation that they were believable to the users and not a hinderence to their ability to experiment with the system.

Next there was the only statistically significant result between the pre and post-test. This indicates that using these systems overall allows users to learn something about how conceptual models and simulations fit into the scientific method. This does not necessarily indicate anything specific about the spatial simulator.

Finally, the level of the number of hypotheses explored between the original MILA-S and MILA-Spatial did not seem to have that large of a difference. However it would be worthwhile to explore this further with a different problem statement that encouraged more exploration and experimentation. This problem statement concerned global warming and its effects on the Antarctic. This is a problem that is fairly mainstream at the moment and the idea of the ice caps melting was probably one of the first thoughts for the participants of this study. Giving a problem statement that encourages more exploartion and does not necessarily have a seeminlgy obvious explanation would be beneficial in further examining the validity of this system.

CHAPTER VI

DISCUSSION

In addition to the study previously described, the implementation of the MILA-HSS system was a big part of this project. The work done in this aspect is described below. This work entails taking the work done on the conceptual modeling language discussed in the introduction and implementing it into an actual working system. This was achieved in the Fall of 2016 with collaboration and help from a colleague, Marc Marone.

There is a close resemblance of the HSS conceptual model compiler to the original compiler in terms of the types of functions it is creating. The main differences are highlighted below.

1. The compiler and the model are completely separate

    - In the implementation of MILA-S and MILA-Spatial the compiler and the model were intrinsically related. The NetLogo code needed for the individual agent was present in the actual model of the agent in code. This required any changes to revolve around editing these components in the code and did not leave any flexibility for the user of the application.
    - What I have implemented with the help of my colleague now is a completely separate model and compiler. The model is built and any additions to the model are recorded by an "omnipresent" representation of the model. I will refer to this model from now on as the Overview Model. The overview model is what changes directly with each addition/deletion/edit of the model in the conceptual model interface.

- Once all editing is done and the user is ready to run their simulation, once the run button is clicked the compiler becomes engaged and takes over.

2. The compiler now has the ability to become more intelligent

   - A future iteration of this compiler has a lot of potential with making knowledge-based decisions about how the NetLogo code should be written. For example, if the compiler knows this is a simple predator-prey model it can decide to not waste any time with checking for patterns of migration.
   - The compiler will now be able to be optimized much more. Before the code that was being written for the simulation had the potential to be very involved for no reason. Now all movement for all agents can occur in one method instead of n different methods were n is the total number of moving agents

3. Developers have an easier time extending the application.

   - Instead of going in and having to write or override multiple classes in order to add a new implementation for something, the workflow will be to add anything needed on the conceptual side. Hook it up to the compiler side. Tell the compiler how to treat this new item.
   - The compiler and model are overall much more flexible and ready for additions than previous implementation.

The changes made to the Component-Mechanism-Phenomenon conceptual model used in previous versions of MILA allow the conceptual modeler to be used to explore more complex biological phenomena. By introducing the idea of spatial components to the conceptual model we have made it possible to model and simulate ecological phenomena spatially.

By adding this ability to the conceptual modeler, steps had to be taken to further develop the SBF-Netlogo compiler described by Vattam, et. al. [6]. The major changes to the CMP-Netlogo compiler are changes that can be replicated by future model-to-simulation compilers and the principles described at coming to the conclusion of each change is a process that can be replicated for future modeling languages. The ability to simulate more complex phenomena has multiple implications. First, for the widespread use of MILA in higher level education. MILA has been found to show statistically significant increases in scientific inquiry in middle schools [3], but the same experiments run at the college level show no differences in pre-tests and post-tests. One reason for this could be that the level of complexity older versions of MILA could achieve were not complex enough to result in inquiry learning gains at the college level. With more complex phenomena able to be modeled, MILA could be used in college-level introductory biology courses. Second, this change is the first of its kind in conceptual modeling. The Structure-Behavior-Function (SBF) modeling language was originally intended to model physical devices [2]. The Component-Mechanism-Phenomenon (CMP) flavor of SBF was able to convert SBF into a causal modeling language, but was still limited spatially. As far as the author is concerned, this version of CMP is the first conceptual modeling language that enables a user to explicitly conceptually model components' interactions not only with each other but also their environment.

CHAPTER VII

FUTURE WORK

The immediate next steps for this research would be to perform the study performed in this paper on a larger group for further and more robust analysis. The same procedure can be used as was used in the pilot study. One thing to add would be to run the study on just the spatial simulator as well in order to see the results on the pre- and post-questionnaires without the effects of the original MILA-S system.

In addition to a larger study it would also be beneficial to run an experiment to determine if this improved conceptual modeler shows a statistically significant increase in students' scientific inquiry literacy and ability. An experiment run on multiple levels of education such as middle school, high school, and college level would be beneficial. It would be beneficial to run these experiments in order to determine if the new CMP language still hold true to the deep learning experienced by users of the SBF language. In addition to experiments in education, it would be worthwhile to thoroughly test and deploy the tool to citizen scientists for use in the field. This type of conceptual modeler is certainly a breakthrough in modeling and simulation and is perfectly fit for those scientists without a lot of simulation writing experience. MILA was originally designed as a tool for education and still has a lot of potential in that realm. It would be interesting to explore how the ability to increase model complexity would affect scientific inquiry and curiosity.

In addition to these explorations, the MILA system has the potential to be used as a tool for resource managers, ecological scientists, citizen scientists, and all levels of education. There are a lot of directions to take the system in the ecological world. Aside from just ecology, the use of MILA in other domains can also be explored.

With this addition to the conceptual modeling language there might be interesting results in applying and re-defining this addition to the language in order to explore other domains. One such domain may be physical processes. As a learning tool being able to model the spatial dimension would open the door to modeling spatial relationships conceptually. The MILA-S system was originally designed for a very specific educational purpose. However with the ability to modify the conceptual modeling language and underlying compiler shown in this project, the potential for the MILA system to be used in other domains is present.

APPENDIX A

PRE AND POST QUESTIONNAIRE

1. First Name:

2. Last Name:

3. In a food web diagram, producers:

   (a) Have arrows pointing toward them

   (b) Have arrows pointing away from them

   (c) A or B could be true given the specific circumstances

   (d) None of these

4. A conceptual model is a diagrammatic or text explanation of how a process occurs, including key elements and how they affect each other (for example, a concept map is a type of conceptual model). Based on your current knowledge, conceptual models can be incorporated into the scientific method because they

   (a) Allow the generation of more accurate hypotheses for subsequent empirical work

   (b) Generate ecological experimental data that can be analyzed statistically

   (c) Provide the only vehicle to graphically present empirically collected data

   (d) All of these

   (e) A and B

5. A simulation model builds and generates a digital prediction of a conceptual model, using simplifying assumptions, to predict the dynamics and outcome of

the modeled scenario. Based on your current knowledge, simulation models can be incorporated into the scientific method because they

(a) Determine the exact outcome of an ecological scenario

(b) Generate ecological experimental data that can be analyzed statistically

(c) Make predictions for subsequent empirical work

(d) All of these

(e) A and C

6. Consider an extended version of the scientific method that incorporates conceptual and simulation models. A conceptual model paired with a simulation model of the same concepts fit into the scientific method to

(a) Support the initial background research

(b) Refine a hypothesis before conducting an empirical pilot study or experiment

(c) Statistically analyze your collected data

(d) Provide a graphical representation of the results of a study

(e) All of these except D

7. Imagine an ecological model where individuals in one finite, dynamic population interact with individuals in another finite, dynamic population using a given set of assumptions and parameters. If you ran the same simulation multiple times, the simulator would generate

(a) Identical results each time

(b) Similar but non-identical results each time

(c) Very different yet predictable results each time

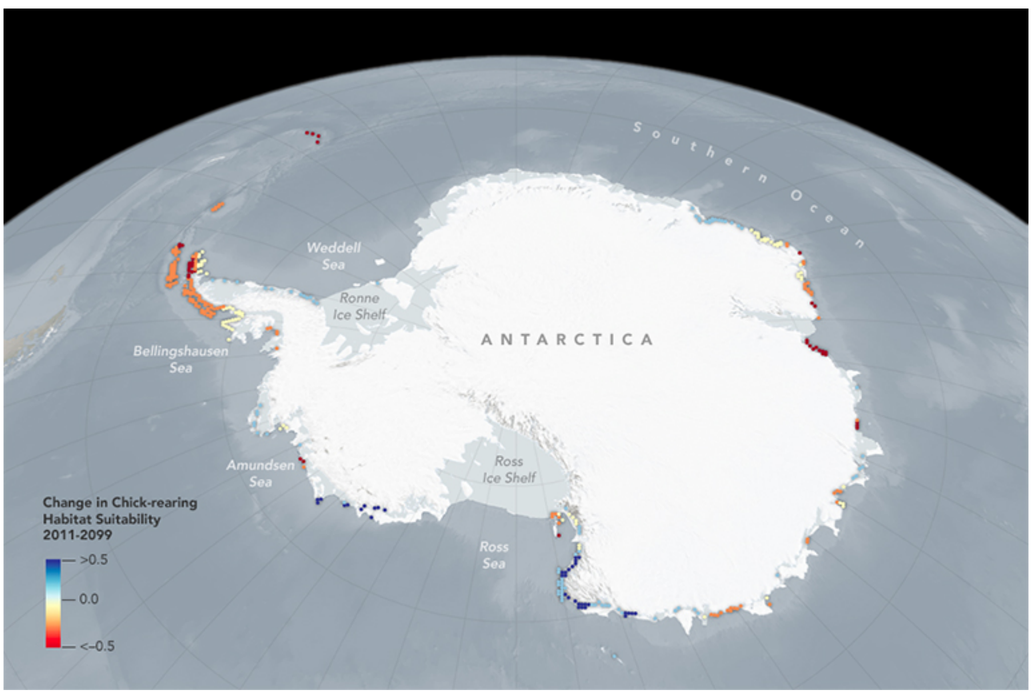
(d) Completely novel and unpredictable results each time

(e) None of these is a relevant prediction.

APPENDIX B

PROBLEM STATEMENT

There are two types of penguins that live in Antarctica: Emperor and Adelie. The Adelie penguin species breeds all throughout the Antarctic continent. The population of this species are experiencing a decline along the West Antarctic Peninsula. There has been an 80% decrease in penguin colonies on the West Antarctic Peninsula recorded since the 1970s. At the same time the Adelie populations are stable or increasing in other areas of Antarctica. The below graphic shows the locations of these colonies and their statuses as either increasing, stable, or decreasing.



The problem presented to you as a citizen scientist is to explore the possible reasons for this population decline in this area of the Antarctic continent. You are free to use the MILA system to its full abilities as well as the following information about Adelie Penguins and the Antarctic Continent.

Why are Adelie penguin populations decreasing along the West Antarctic Peninsula?

REFERENCES

[1] A. Crooks, C. C. and Batty, M., "Key challenges in agent-based modelling for geo-spatial simulation," *Computer Environmental Urban Systems*, vol. 32, no. 6, pp. 417–430, 2008.

[2] A. Goel, S. R. and Vattam, S., "Structure, behavior, and function of complex systems: The structure, behavior, and function modeling language," vol. 23, 2008.

[3] D. Joyner, A. G. and Papin, N., "Mila-s : Generation of agent-based simulations from conceptual models of complex systems," 2014.

[4] F. Massol, D. Gravel, N. M. M. W. C. T. F. and Leibold, M. A., "Linking community and ecosystem dynamics through spatial ecology," *Ecology Letters*, vol. 14, no. 3, pp. 313–323, 2011.

[5] Joyner, D. A., *Metacognitive Tutoring for Inquiry-Driven Modeling*. PhD thesis, Georgia Institute of Technology, 2015.

[6] S. Vattam, A. K. G. and Rugaber, S., "Behavior patterns: Bridging conceptual models and agent-based simulations in interactive learning environments," pp. 139–141, 2011.

[7] V. Grimm, U. Berger, F. B. S. E. V. G. J. G. J. G.-C. T. G. S. K. H. G. H. A. H. J. U. J. C. J. W. M. M. B. M. G. P. C. P. S. F. R. A. M. R. M. M. R. E. R. N. R. E. S. S. S. R. a. S. R. V. U. V. and DeAngelis, D. L., "A standard protocol for describing individual-based and agent-based models," *Ecological Modeling*, vol. 198, no. 1-2, pp. 115–126, 2006.

[8] WENNING, C. J., "Assessing inquiry skills as a component of scientific literacy," *Journal of Physics Teacher Education Online*, vol. 4, no. 2, pp. 21–24, 2007.