# Developing Computational Methods to Measure and Track Learner's Spatial Reasoning

BY

# ADITI KRISHNA MALLAVARAPU

B.E. Computer Engineering, University of Pune, Maharashtra, India, 2011

# THESIS

Submitted as partial fulfillment of the requirements for the degree of Master of Science in Computer Science in the Graduate College of the University of Illinois at Chicago, 2014

Chicago, Illinois

Defense Committee:

Dr. Leilah Lyons, Chair and Advisor, Computer Science Dr. Barbara Di Eugenio, Computer Science Dr. Brian Ziebart, Computer Science To my family and friends

#### ACKNOWLEDGEMENT

I am thankful to family especially my parents and my new family here in US, for standing by me through the process encouraging me at every step of my personal and professional life. To my friends for being around to help around every time I needed help.

I am also grateful to advisor, Dr. Leilah Lyons, for not only advising me but also for constantly motivating me to do better. She has always been a role model for me, to provide me the strength and inspiration to excel in my academics and also in my life struggles. I would like to mention the support and encouragement I received from my fellow researchers. It wouldn't have been possible without you all.

I would like to thank my thesis committee Dr. Barbara Di Eugenio and Dr. Brian Ziebert for their unwavering support and assistance. They provided guidance in all areas that helped me accomplish my research goals and enjoy myself in the process.

Lastly, I would like to take a moment and thank everyone who have been a part of my life during this process and for tolerating my tantrums. Thank you for keeping me focused.

AKM

iii

<u>C</u>	<u>CHAPTER</u>		
1	IN	VTRODUCTION	1
	1.1	Open-Ended Problems in EDM	
	1.2	Problem Space: Green Infrastructure in Urban Planning	5
	1.3	Approach & Contribution	6
2	BA	ACKGROUND & PRIOR WORK	
	2.1	The Current State of Educational Data Mining	
	2.2	Educational Data Mining for Open-Ended Problems	
	2.3	EcoCollage: a Platform for Promoting and Studying Spatial Reasoning	around Urban
	Planr	ning Problems	11
	2.4	Metrics for Spatial Phenomena	15
	2.4	4.1 Spatial Pattern Characterization Methods	16
	2.4	4.2 Details on the Ripley's K(t) function	
3	Al	NALYSIS OF ECOCOLLAGE	
	3.1	Available Data	
	3.2	Initial Attempt to Understand EcoCollage strategies: Creating a Visualiza	tion Tool 25
	3.3	Second Revision to Visualization Tool	
4	DA	ATA MINING APPROACH	
	4.1	Reducing the State Space of the Solutions	30
	4.2	Initial State Space Reduction Approach: Condensing State Space to "Cha	nge" Space 30

# TABLE OF CONTENTS

# TABLE OF CONTENTS (continued)

<u>CHAPTER</u> <u>PAGE</u>		
4.3	Revised State Space Reduction Approach: Simple Spatially-Sensitive Density Metrics	
	33	
4.4	Final State Space Reduction Approach: Radii-Preservation	
4.5	Strategy Detection Approach	
5 RI	ESULTS	
5.1	Regression Results: Identifying Differences in Effective Spatial Strategies across	
Inter	faces	
5.2	Solution Discovery: Comparing and Tracking Learner's Use of Spatial Strategies 51	
6 C(	ONCLUSION & FUTURE WORK 60	
BIBLIC	DGRAPHY	
VITA.		

EQUATION	PAGE
I: Theoretical K(t) function	
II:Ripley's K Univariate	19
III: Ripley's K for Bivariate	
IV: Ripley's L Equation	
V: Placement Dissimilarity	
VI: Abundance Dissimilarity	
VII: Spatial Dispersion Dissimilarity	
VIII: Multiple Linear Regression	41

# LIST OF EQUATIONS

# LIST OF FIGURES

FIGURE	PAGE
I: Plot showing the relationship between infiltration and density	
II: A Depiction of Paper Condition (left), Multi-mouse (top right), Single- Mouse	(bottom right)
	13
III: The simulation Interface:	14
IV: Illustration showing how Ripley's univariate calculation	19
V: Illustration of how Ripley's K Bivariate	21
VI: Study area and a search circle	
VII: The format of the data file (input) to the Java application	27
VIII: The second revision to the visualization tool	
IX: Visualization of learners' trials	
X: Plot of the Univariate L-Values	
XI: Plot of normalized L values	39

# LIST OF TABLES

TABLE	<b>PAGE</b>
I: Sum and Average of Trials attempted per condition by groups	
II: Correlation matrix for the density metrics	
III: The coefficients of the significant variables for the paper trials	
IV: The coefficients of the significant variables for the multi-mouse trials	47
V: the coefficients of the significant variables for the single mouse trials	
VI:Coefficients of the three regression models	
VII: Goodness of fit for the models	50
VIII: The coefficients of the significant variables for all the trials	51
IX: An alternate representation of the coefficients in Table VII above	
X: The total number of good strategies, by condition	53
XI: The average delta in number of good strategies used, trial-to-trial	
XII: The number of unique good strategies tested by groups	55
XIII: Illustration of how for Group 9's exploration of the solution space	
XIV: Summary of the appearance of good strategies (as identified by the multi-lin	ear regression
model) in trials generated in the three conditions	
XV: ANOVA results for order of discovery of good strategies	58
XVI: Post-hoc tests for the order discovery of good strategies	59

#### SUMMARY

Interactive learning environments can provide learners with opportunities to explore rich, real-world problem spaces, but it can be hard for educators and educational designers to (a) understand what students are "up to", and to subsequently (b) provide guidance or feedback to help learners make progress. Educational Data Mining (EDM) offers the potential to help diagnose student activities within an interactive learning environment, but it has historically been applied to constrained and fairly well-understood problem spaces. This work represents part of a growing body of research that is applying EDM techniques to more open-ended problem spaces.

The open-ended problem space under study here was an environmental science simulation, where learners were confronted with the real-world challenge of figuring out where it is effective to place green infrastructure in an urban neighborhood so as to reduce surface flooding. Learners could try out many different arrangements of green infrastructure and use the simulation to test each solution's impact on flooding. The solutions proposed by the learners were logged during a series of experimental trials with different user interface designs for the simulation. Analyzing this data was difficult due to the large possible solution state space, and because there are many possible good solutions and may possible paths to discover good solutions.

This work proposes a procedure for reducing the state space while maintaining critical spatial properties of the solutions. Spatial reasoning problems are a class of problems not yet examined by EDM, so this work will set the stage for further research in this area. This work also demonstrates a procedure for discovering effective spatial strategies and solution paths, and demonstrates how this information can be used to give formative feedback to the designers of the interactive learning environment.

ix

## 1 Introduction

Educational data mining is aimed at developing methods to analyze the data that emerges from educational settings to further understand the behavior of students [1]. The increasing prevalence of technology in classrooms and other learning environments has the potential to both affect the way students learn things, as well as to help educators and educational designers get a better window onto the processes by which students learn things. This latter capacity, being able to track how students act within technological learning environments, will become increasingly important as our ability to create rich interactive learning experiences outstrips our ability to design assessments. Teachers most often informally assess learners' progress via observation or via strategies like pop quizzes, and formally assess learners' performance via written paper tests. These formats don't easily cover the wide range of learning possible within an interactive learning environment. For example, learners can exhibit a range of skills and epistemic knowledge while engaged in a task that they could seldom learn from reading a textbook passage or express on a written test. This can hamper the ability of teachers to give meaningful feedback and guidance to students, and can have larger negative implications in the current high stakes testing atmosphere of education.

"Stealth assessment" is one approach to automated, embedded assessment: usually applied to interactive simulations or games, in this approach student performance is monitored and tracked while they play, often without them being aware of it [2]. Compared to paper-based assessments, one can gain a much richer picture of learner capabilities, but with added complications: how the "stealth assessment" data is analyzed is highly dependent on who will use this data, and for what purpose. Any assessment designer can attest that the nature of the learner performance evidence that must be gathered rests on a number of factors, including whether the assessment is intended to assist future learning (a formative assessment), to assess individual achievement (a summative assessment), or to assess the quality and effectiveness of an educational intervention (in other words, the assessment isn't really about the individual learner, but about the learning experience itself) [3]. Once the purpose of the assessment has been decided upon, assessment designers must then specify the three aspects of the "assessment triangle": the model of student *cognition* and/or a model of how learning takes place in the domain in an ideal case (these can be different), the nature of the *observations* that will provide evidence of learner performance, and the *interpretation* process that will convert observations into evidence [3].

Educational data mining researchers, then, are not always trying to solve the same kinds of problems. In particular, the design of the technological learning environment can greatly affect the types of *observations* of learner performances that are available for analysis. For example, while there have been many examples of using data mining to track students' progress through interactive learning environments like games [2, 4, 5] and simulations [6, 7, 8, 9, 10, 11, 12, 13] using log files, most of these learning experiences are intentionally designed to be highly constrained so as to maximize the informational value of the logged observations. For example, learners may be given a well-defined, fixed goal where there are known fixed number of steps to reach this goal, and a known, fixed number of choices that can be made by the learner - in such circumstances, any user action can be constructed as taking them closer to or farther away from the goal. These well-constrained problem spaces can be used by data miners to do things like provide formative feedback to the students [7, 13] or to their teachers [6] so that student learning can be better managed, to provide formative feedback to the environment designers so that flaws in the design can be detected and corrected [4, 6], or to provide evaluative feedback on the nature and scope of mistakes made by learners in the environment [5, 8]. However, these constrained

problem spaces often do not reflect problems found in the real world. Real world problems often have many different solutions, which can be reached via many different paths. While presenting learners with simplified and constrained problems can be a good way to help them come to understand the core properties of a domain (i.e., conceptual learning), it does not prepare learners for grappling with the messiness of problem-solving in real scenarios (what can sometimes be called process learning, or the acquisition of a disciplinary disposition). Giving learners less constrained, more open-ended problems can help them get experience with disciplinary processes and dispositions, but one barrier to doing so is that a lot more work is required to assess student progress within open-ended problem spaces. Educational data mining offers the possibility of making assessment of open-ended problem-solving more tractable, and thus allowing for the increased educational use of open-ended problems.

#### 1.1 Open-Ended Problems in EDM

Increasing the open-endedness of the problem makes it harder to use the *observation* of a given action as direct evidence of progress towards or away from a solution. When designing a highly-constrained learning environment, the designers usually already have a well-defined model of *cognition* underpinning the domain. In open-ended problem spaces, this cognitive model may as yet be poorly understood, or there may be multiple competing models of cognition at play. While these challenges might stymie traditional assessment designers, one of the strengths of data mining is its ability to discover patterns in data, and to use those patterns to build models. The earlier uses of EDM were just to find correlational relationships in data (e.g., demonstrating that a certain type of action tends to be associated with a successful outcome), but EDM is increasingly being used to build models of student behavior [1]. This allows researchers (and eventually,

educators) to do more than just compare student performance against a single (assumed) model of desirable performance, rather, it allows for the presence of multiple models of performance. This capacity for model discovery is especially useful for understanding how learners grapple with rich, open-ended problem spaces.

The work presented here addresses a learning environment that is decidedly open-ended, with many degrees of freedom, and with no clear prescriptions for good and bad solutions: an environmental simulation where the relative, not absolute, spatial placements of elements matter. Others have begun to apply data mining approaches to open-ended learning problems as in [4], but in this case the problem was still fairly constrained in the following ways: (1) there were not many degrees of freedom in learner actions, and at each step the learner's action was highly contingent on earlier learner actions, reducing the size of the problem space; (2) even though there were multiple solution paths, there was a single well-defined model for what constituted a "good" solution. The problem space we are confronted in our work with has a very large state space – there are 324! different possible solutions (2.28899746 E+674) - and a large number of those solutions are likely to be fairly equivalent in terms of their performance, even though they may not be at all structurally similar. These aspects make a straightforward application of existing EDM approaches difficult. My work, then, extends the current work on EDM for open-ended problems by 1) devising an approach for reducing the state space of a spatial problem with a genuinely large and non-contingent set of possible learner actions so that it would be tractable for analysis, (2) using my state-space reduction approach to discover the spatial strategies that may undergird student performance in the open-ended problem space, and (3) using the results of this strategy discovery to compare how different user interfaces can impact learners' use of those spatial strategies.

## 1.2 Problem Space: Green Infrastructure in Urban Planning

Our problem space is one drawn from urban planning and environmental science: the challenge of integrating green infrastructure into existing urban infrastructure in order to reduce surface flooding in urban areas. The urban planners' problem task involves deciding how to arrange green infrastructure elements (like swales, a special type of water-retention garden) across a landscape to maximize the capture of rainwater (while minimizing costs). This challenge entails reasoning spatially about 2-dimensional patterns of swales (which can exist at different levels of scale to either reinforce one another, or counteract one another). A given spatial pattern might be very effective at one level of scale but ineffective at another level of scale, or in combination with other patterns at different levels of scale. So a planner can't just recommend to say, clump all the swales near each other, as this approach may work at some levels of scale but not at others.

The learning environment used in this work was developed for an NSF grant to further environmental science education, and is an agent-based simulation implemented in *NetLogo<sup>1</sup>* and adapted from a green infrastructure planning simulation, *L-GrID*, developed for the Illinois EPAfunded *Green Infrastructure Plan for Illinois* project. The goal was to provide learners with software that would allow them to practice making decisions about where to place green infrastructure elements in an urban setting, and to give them dynamic simulated feedback on their choices. To support collaborative placement decisions, a paper-based Tangible User Interface (TUI) front-end was constructed for use with the simulation, as TUIs are theorized to provide benefits for spatial reasoning tasks. For comparison, a custom drag-and-drop front-end was also

<sup>&</sup>lt;sup>1</sup> Available at: http://ccl.northwestern.edu/netlogo/

created to support two control conditions — single mouse or multiple mice — that would link to the back-end ABM simulation just as the TUI does.

In all three conditions, the swale coordinates are saved to a file that the simulation uses on set-up. The maps used in the experiment consist of different elements – (a) two types of land cover: impermeable (e.g., roads, buildings), and highly permeable (swales), (b) sewers (which drain water from the surface but prevent groundwater infiltration), (c) an elevation gradient, and (d) an outflow point at the lowest point of the map (from which water leaves the map, emulating how water will flow from one terrain to another – this is not a closed system).

## 1.3 Approach & Contribution

This work explored a method to detect spatial characteristics of the patterns created by learners so that different solutions could be compared against one another. A multivariate linear regression approach was then used to determine which patterns at which spatial scales were associated with improvements in rainwater capture. Essentially, we used the data generated by learners interacting with the problem space to bootstrap the development of a model of how these novice learners engage with the problem space (i.e., the combinations of spatial strategies they found to be effective at capturing rainwater). (It should be noted that while we are, in a sense, modeling the strategies used by learners, our findings in no way approach a full-fledged cognitive model, as might be used within an Intelligent Tutoring System – much more future work is needed to reach that stage). We then used these results to examine if the user interface design affected the way in which learners approached exploring the problem space: did they use different spatial strategies, or discover them more quickly or more slowly, when using different user interfaces? We found that this was indeed the case - certain spatial strategies were more often present in some user

interface conditions than others. We also used the results to examine if the patterns of spatial strategy exploration differed across user interface conditions, and found that certain interface designs did seem to promote earlier discovery of spatial strategies.

This work is among the first in educational data mining to tackle an open-ended problem space with both multiple solutions and multiple solution paths, and is the first to our knowledge to approach a spatial reasoning problem in educational data mining domain. It demonstrates the potential for educational data mining approaches to help educators develop models of effective solution strategies in rich problem spaces. We also demonstrate that such "discovered" models can be used to evaluate and compare different learning environment designs. Additionally, these results suggest that we may be able to summatively examine the meta-strategies (e.g., the ordering of strategy exploration) to determine which patterns of exploration may be more or less effective in this complex spatial problem space. This is a finding that could be used in future work to develop dynamic formative feedback to help learners engage with complex spatial problem spaces.

# 2 Background & Prior Work

### 2.1 The Current State of Educational Data Mining

There are many potential uses of educational data mining of learner log data. Student log data has been extensively used to analyze various components of student behavior and track student progress, but for different purposes. The use, and the nature of the log data itself, both shape which EDM analytic techniques should be applied. The past decade has seen an increase in the efforts of the researchers to utilize the hidden information in these log files. A review history and current trends of EDM is indicative to a shift of trends in the research from relationship mining to discovery with models and prediction in these recent years [1]. The logs have been extensively used in the past years to evaluate learning material [14, 15, 16], to study how different types of student behavior overtime [19, 20, 21, 22]. Improvement of these student models have also attracted a lot of attention [1] mostly to know students current knowledge, motivation, and attitudes.

As concerns classifying the purpose of analysis in terms of the use of log data we have two major trends. One purpose can be to use the log data in a "formative" fashion to improve future learning. For example, Harpstead et al. use the log data to improve the design of the simulation by clustering features in the solutions through context free 2D grammar, aiming to improve learning [4]. Martinez-Maldonado et al., uses a process discovery tool on the log data which generates a meaningful abstraction of general process by distinguishing actions that are important to detect when and who needs scaffolding [6], and Corbett et al. emphasized how modelling student individual differences could be used to increase student learning [23]. Another purpose can be to use the log data in a "summative" fashion to judge or evaluate past learning, as shown by Gobert

and Sao Pedro through a machine learning algorithm to detect models of student inquiry skills [7]. Rafferty et al. uses L1 regularization to evaluate if paired student interactions are predictive of both students' post test scores average and their individual scores [8]. There are also some cases where the log data is used both formatively and summatively - in computer adaptive testing, for example, student responses are used formatively to shape the next summative question to be given to them.

One of the more common analytic approaches seen, especially with stealth assessment, is to devise approaches that can compare learner performance against an ideal model of how an expert or other maximally efficient person would accomplish the same task, as a way of summative evaluation of their performance [8, 10], to share with teachers, parents, administrators, etc. This approach works well with content areas that are well-understood (where it is possible to have uncontroversial models of what expert performance should look like), and well-bounded (the solution paths are relatively linear, so that at any point the "distance" between the learner's current position in the problem space and the solution can be determined). For open-ended problems, learners may have many "varied" number of steps and solution paths [13].

There has also been considerable work in exploring the effectiveness of pedagogical support (in learning softwares and collaborative learning environments). [14, 24] investigated which types of supports were more effective overall, among different groups of students and in different situations. Beck et al. used learning decomposition to find the best fit model to infer the relative effectiveness of each support type for promoting learning by understanding the key factors impacting learning and to design better learning systems [14].

## 2.2 Educational Data Mining for Open-Ended Problems

As discussed in the preceding section, compared to other problems open ended problems lack structure, as in they are not well-bounded (the distance between the current state and the goal is difficult to be determined), and the problem's solution set may comprise large number of candidates which makes it impossible to evaluate each of the possible solutions [13]. This property makes us incapable of using complex methodologies like Bayesian knowledge tracing and Markov models, which when applied to better-defined problems spaces give us a detailed insight of the of the learners' behavior, as done by Pelanek et al. [10], Falakmasir et al. [25]. Also these methodologies consider the problem space to be divided into snippets on time. Each snippet refers to a move and each move would incur upon them a reward that would motivate to go ahead and explore in the next move, so the next move would be influenced by the previous move and the reward obtained in the previous move.

There have been very few researchers who have explored open-ended problem spaces in educational data mining. A successful attempt to extract learning behavior of students from paired student interaction data with an open-ended chemistry lab to understand and estimate the individual knowledge is depicted in Rafferty et al. [8]. Their goal was to investigate whether paired interactions with an open-ended chemistry tutor can be used to predict individual student post-test performances. The interactive lab required the students to apply their chemistry knowledge in authentic, real-world contexts (like whether factories are reporting accurate pollution levels). The students were placed in pairs and could approach the problem however they felt comfortable by analyzing and testing various chemical reactions in the virtual lab, giving the open-endedness to the learning experience. They identified 12 features to characterize each of the activities the students performed, categorizing the activities as help seeking, holistic, and practical. The outcome

of the study was to predict an individual's post-test scores through paired performance and pretest scores.

One other research attempt at detecting student strategies from open ended games was in the paper by Harpstead et al. [4], where they used a combination of data mining and automaton theory to extract features of the learners' solutions and compare them to designers' solutions. The paper describes a game called RumbleBlocks [4] where kids in the age group of 5- 8 years of age have to place blocks to build a stable structure. Their understanding of center of mass and stability are being assessed through the game. For each student solution a decision tree was constructed; once they have these trees they tried to extract features to create a vector that would describe how and what students were doing. The features of those structures were then matched to the features from the solutions of the designer of the game. The study focuses on helping the designers and researchers redesign aspects of the learning experience that seemed to produce discrepancies between how the players used it and how the designers had envisioned its use

# 2.3 <u>EcoCollage: a Platform for Promoting and Studying Spatial Reasoning around Urban</u> Planning Problems

The Next Generation Science Standards (NGSS) recommends that learners engage with simulated models both to deepen their content knowledge of systems within the content domain and to acquire practice skills [26]. For domains like environmental science, which spans the disciplines of ecology and urban planning, system functions are dependent on the relative spatial positions of elements (buildings, permeable surfaces, habitats) [27]. For example, much of "green infrastructure" planning for storm water management involves making decisions about where

green infrastructure elements such as green roofs and drainage swales need to be placed to yield maximum benefit.

We know that relative spatial placement of swales can impact outcomes important to the urban planning problem space, like the amount of infiltration of rainwater into the groundwater system. This has important consequences for urban planning: while, in general, installing more swales results in more infiltration (and thus reductions in flooding), there is not a perfectly linear relationship between the number of swales and the amount of infiltration - certain kinds of placements can have greater or lesser impacts. See *Figure I*.

We know that, owing to the dynamics of how water flows across paved versus unpaved surfaces, that certain types of patterns (like clustering swales, or deliberately spacing them out) can help the swales function as more than just sponges: they can also slow or direct the flow of water so that it can be better absorbed by other swales (or, less desirably, but still preferable to flooding on the ground, to be captured by sewers).



Figure I: Plot showing the relationship between infiltration and density

From a pedagogical perspective, it's critical to give learners practice with spatial problem solving. Exploring and discussing the effects of different spatial configurations of green infrastructure on flooding and groundwater infiltration, for example, is central to building an understanding of environmental science. The fields of ecology and urban planning increasingly make use of Agent-Based Modeling (ABM) software to model and test hypotheses about complex human-natural systems [28]. The most common method for specifying spatial arrangements in ABMs is via a programming interface, but it seemed to our team that a direct-manipulation user interface (where users could directly place and move representations of the simulation elements on a representation of the environment) would be more appropriate for novice learners, for both usability and pedagogical reasons.



Figure II: A Depiction of Paper Condition (left), Multi-mouse (top right), Single- Mouse (bottom right)



Figure III: The simulation Interface: The Green squares represent swales, and the blue coloration shows the depth of water at the moment when this screen shot was taken.

In a small pilot study, we confirmed that using a paper-based tangible user interface (TUI) to specify a (preset) configuration of 16 elements was over 7 times faster (1m 11s) than an expert user manually programming the same configuration (8m 18s) [29]. Theories of embodied reasoning claim that abstract visual and spatial concepts are acquired from embodied sensorimotor experiences [30], such as those gained by interacting with TUIs. If TUIs better align with users' schema for perceiving and reasoning about spatial relationships, then this would allow for users' spatial problem solving to be more streamlined and less cognitively taxing.

Surprisingly, we could find little experimental work that verified these embodied benefits of TUIs. For instance, while our pilot demonstrated an efficiency benefit *vis à vis* programming [29], it didn't allow us to claim that the benefit was afforded by the embodied nature of the TUI alone

(as programming is not a direct input method). For a more fair comparison, we conducted a second pilot that contrasted a TUI to a multiple mouse direct input interface, and found surprisingly few usability and collaboration differences [31]. This motivated us to investigate the interaction of access, tangibility, and collaboration by studying the same task across three conditions: a TUI, a multiple mouse interface, and a single mouse interface. In contrast to the second study [31], adding a single-mouse condition enabled us to isolate equality of access as a factor. We were interested in investigating how users' problem solving was impacted, both in terms of their exploration of the problem space (i.e. breadth of problem solving) and in terms of their optimization of proposed solutions (i.e. depth of problem solving). To begin to compare problem solving across these conditions, though, we first needed to establish how to measure users' spatial strategies.

#### 2.4 <u>Metrics for Spatial Phenomena</u>

Everybody uses spatial skills knowingly or unknowingly in day to day routine like arranging the closet, placing jars in the kitchen etc. Spatial reasoning, which is the ability to mentally visualize and manipulate two- and three-dimensional objects, also is a great predictor of talent in science, technology, engineering and math, collectively known as STEM [32]. However, if we want to go beyond just noting this correlation to improving the success children and adults have in science and technology, more must be known about spatial learning [33]. Though research has explored how different pedagogical strategies in classrooms can improve spatial skills on posttests of spatial abilities [34], there is as yet little evidence for (or methods for studying) how spatial reasoning evidences itself during the learning process.

#### 2.4.1 Spatial Pattern Characterization Methods

Before we can hope to study how people reason about spatial problems, we need some way of measuring the spatial properties of their proposed solutions. The literature supporting statistics to measure spatial patterns is extensive. Spatial statistics caters to several fields like plant ecology, animal ecology, geography, mining, engineering and many more. These fields made use of spatial statistics either for explorative or inference purposes, and had different approaches for the same which included mathematical approaches like counting methods, covariance, variance, etc. [35]. The spatial statistic to be adopted by the research is influenced by research objective, measurement types and sample data [35]. For this problem space, we are concerned with the relative placement of items: are they near one another, or spread apart? The relative distances between swales and other swales, and between swales and other water-capturing elements (like sewers) meaningfully affect the patterns of flooding that emerge in urban settings.

Though there are many statistical methods to measure spatial patterns there are only a few that fit the purpose of research and the sampling data concerned with our problem domain. Our problem domain needs a method to characterize arrangements in a 2-dimensional space. All the spatial metrics assume the points corresponding to the location of objects or events of interest are surveyed and mapped [35].

The Nearest Neighbor (NN) method is perhaps the simplest relative location metric to apply. The procedure iterates through all points, tallies the distance of each point to its nearest neighbor, and averages together those distances. This allows one to measure the degree of clumping found across the map. By comparing the nearest neighbor metric against what one would expect from a random distribution of the same number of points across the same area, one can discover if the points are more likely to be spatially close than expected by chance alone. If one wished to compare two patterns on the basis of NN, however, only limited conclusions could be reached – any spatial patterning that exists at radii beyond the NN radii for the two patterns goes unrecognized. It could be the case that when a NN comparison suggests that two patterns are different because one has a larger NN metric, in fact, the pattern with the smaller NN metric also has a very similar pattern at the same scale as the larger NN pattern. So these patterns would be marked as being dissimilar even though they shared spatial characteristics.

Another metric called the Cumulative Distribution function (CDF) of distances, given the probability distribution of distance k, calculates for all events, the probability of event-to-nearestevent distances is less than or equal to the distance k. This probability distribution can then be used to compare two patterns at a given scale of distance, which would resolve the issue with NN comparisons pointed out above, but the comparison is still limited by the window size, and cannot recognize certain spatial properties (like whether the patterns exhibit spatial regularities like overdispersion).

The Ripley's K metric is considered as a refinement of the nearest-K neighbors metric. The Ripley's K function can be used to summarize a point pattern, test hypotheses about the pattern, estimate parameters and fit models [32]. It quantifies the spatial pattern intensity of points for various sizes of circular windows, i.e. it computes the mean number of points lying within the circular search window. Alternative metrics like the nearest neighbor distance, cumulative distribution function (CDF) of distance (of points from random points to the nearest neighbors) do not calculate the metrics on varying scales of distances [32]. Ripley's K metric can thus successfully detect combinations of effects like clustering at large scales while simultaneously being sensitive to regularity at smaller scales. It can also compute these properties for both univariate (where the arrangement of an event is observed with respect the other events of the same

type) and multivariate (the arrangement of an event with respect to one or more other event types) patterns. By using circular window Ripley's K gives an isotropic (i.e., non-direction-sensitive) cumulative count of all points at distance 0 to some distance t, specified by the analyst [35]. For this reason, we chose to employ the Ripley's K metric (technically, we used a normalized version, the Ripley's L measure) to convert the state space for our problem from one of absolute Cartesian placements of swales to the much smaller (and more meaningful) space of the relative spatial arrangements of swales.

### 2.4.2 Details on the Ripley's K(t) function

Here we provide extra details on the Ripley's metrics because they ended up being very important to our analyses. The function K(t) is defined to give the probability of finding the elements of interest in the specified window size given the overall density of elements in that area. The general definition of the Ripley's K-function for a certain distance t is

 $K(t) = 1/\lambda[E(t)]$ 

# Equation I: Theoretical K(t) function

Where

 $\lambda$  is the density in the study plot obtained from n (number of points) and

A (area of the study plot)

E(t) = expected number of points within the distance t of an arbitrary point



Figure IV: Illustration showing how Ripley's univariate calculation

The figure above shows an illustration of how Ripley's univariate calculation works for the radius r. It sweeps a region of size r around each item of interest, i. The approach repeats this calculation for radii of size 1 to t, producing t different K statistics.

For Univariate computations:

$$K(t) = \frac{1}{\lambda} \sum_{i=1}^{n} \sum_{j=1}^{m} w(i, j) I(i, j)$$

Equation II: Ripley's K Univariate

Where  $i \neq j$ 

 $\lambda(density) = n/A$ 

n = number of events or points

A = Area of the map under consideration

 $w(i,j) = edge \ correction \ factor$ 

= 1 if the search circle is centered at i and passing through j is completely inside the study area otherwise it is the proportion of the search circle in the study area.

I(i, j) = Indicator function 1 if the point j is within distance t of point i else 0

(Determined by the Euclidian distance between point i and j)

With 
$$d(i,j) = \sqrt{(xi - xj)^2 + (yi - yj)^2}$$

being the Euclidian distance between the points pi=(xi, yi) and pj=(xj, yj) within the study region.

For Bivariate computations:

$$K(t) = \frac{1}{\lambda_1 * \lambda_2} \sum_{i=1}^n \sum_{j=1}^m w(i,j) I(i,j)$$

# Equation III: Ripley's K for Bivariate

Where  $i \neq j$  and  $\lambda_1$  and  $\lambda_2$  are the densities of the two elements of interest.

and m and n are the number of elements of interest in each category.



Figure V: Illustration of how Ripley's K Bivariate

The figure illustrates how Ripley's K Bivariate sweeps the study space, tallying the number of items of a second type found within radius r of each item i of the first type. The map is divided into a grid. For Bivariate calculation the metric calculates the number of green elements (swales) in the radius r around the yellow elements (in our case would be sewers).

Because of its hyperbolic behavior, the interpretation of K-function is not straightforward, especially if one wishes to compare the spatial characteristics of one map against another. For this reason, a modification called L-function has been proposed to normalize it

$$L(t) = \sqrt{K(t)/\Pi} - t$$

Equation IV: Ripley's L Equation

The expected value of the univariate L-function under CSR (complete spatial randomness) is 0 for all *t*. Complete spatial randomness (CSR) describes a point process whereby point events occur within a given study area in a completely random fashion. Such a process is modeled using only one parameter  $\lambda$ , i.e. the density of points within the defined area [36]. Poisson distribution is used to express the probability of given number of events occurring in a fixed interval of space and/or time independently of the last event. Thus, when the L value is positive, indicating that the pattern is more tightly-packed than one would expect to see by chance, we know that the pattern tends to be clustered, and when the L-value is negative the pattern is tending towards being overdispersed or regular [32].

The accuracy of the K value highly depends on the size and shape of the study area and the edge effects. The edge effect, if uncorrected, would overestimate how much "empty space" surrounds points of interest at the boundaries of the study plot as compared to those in the center of the study plot. For this reason, we curtail the size of the area used to compute the density for edge-adjacent points. Common practice while considering edge corrections is to keep the maximum search circle radius about one half of the shortest dimension of the study area [37], as this reduces the number of assumptions being made about the pattern.

The edge effects should be considered when the search circle intersects the edge of the study plot. As shown in the figure the search circle consists of two distinctive parts one inside the study plot and other outside the study plot. If the points under the portion of the search circle denoted by  $A(r)^-$  are considered the area A(r) would have fewer points than expected [37]. Edge effects calculate the proportion of the search circle inside the study plot and utilize the area of this proportion in the calculations. The mechanism becomes complicated when the shape of the study plot is irregular, but in our case, we are using rectangular maps and this method suffices.



Figure VI: Study area and a search circle

The figure illustrates the Study area and a search circle ci(r) with radius r centered on a point i within this region  $A(r)^-$  and  $A(r)^+$  are the area of the search circle outside and inside respectively

# **3** Analysis of EcoCollage

#### 3.1 Available Data

The data used in this work was collected from a controlled within-subject with-rotation experiment with three conditions (each testing a different user interface for the simulation, one that used a paper-based TUI, one that used a single mouse input with a graphical user interface, and one which used a multi-mouse input with a graphical user interface).

This data covered 30 different experiments, each of which involved 3 undergraduate students working together across the three conditions. In each of the three conditions (the order of which was varied to counterbalance any practice effects), the students tested out different possible placements of swales on a map of an urban setting. The participants judged the success of their placements (which we term "trials" in this documents) based on the amount of water captured by the swales, called the infiltration, and the "cost" to place that number of swales on the map – a fixed amount of \$10,000/swale. The software converted these in-game metrics into real-world monetary rewards that the participants could earn (they would leave the experiment with a payout tallied from each of the three condition). Participants were motivated to improve the infiltration while keeping cost low by dint of an economic reward – they would receive more or less money for their participation in the experiment depending on their improvements to these two in-game scores. The table below shows the total number of trials attempted by the participants, as well as the average number of trials each participant group produced per condition.

	paper	multi- mouse	single mouse
sum	182	205	211
AVERAGE	6.07	6.83	7.03

*Table I: Sum and Average of Trials attempted per condition by groups* 

### 3.2 Initial Attempt to Understand EcoCollage strategies: Creating a Visualization Tool

Our efforts to inspect the strategies used by participants initially started with a visualization tool that was able to "play back" the trials attempted by the learners while tackling the EcoCollage problem, originally created by PhD student Tia Shelley. The visualization tool was created using Processing.

The visualization tool has two modes: one is default mode in which you can compare two trials that occur next to each other and the other is the overview mode where you can see the different garden placements across all the trials in that particular study. My first attempt at computing meaningful spatial metrics for this space involved implementing metrics proposed in [38]: Placement Differential (PD), Abundance Differential (AD), and Spatial Dispersion Dissimilarity (SDD) for both univariate and bivariate calculations. These metrics are described in a bit more detail in the next section, for now suffice to say that they were intended to capture trial-to-trial differences in how the participants placed the swales, as we were interested in studying their solution paths. (Although ultimately, we did not find these metrics as useful for this purpose as we had hoped).

After implementing the PD, AD, SDDu, and SDDb metrics, I augmented the visualization tool to show bar graphs of PD, AD, SDDu and SDDb along with the trials. We discovered that while

viewing these metrics shed some light on how dramatically players would vary their arrangements, because we distilled the two spatial arrangement metrics down to a single number, we missed a great deal of the nuance in how the participants were varying the placement of their swales.

#### 3.3 Second Revision to Visualization Tool

Thus, we felt a need of further ammunition to aid our analysis. Many mathematical platforms like R [39], Programita [40] and many others have a functional implementations of the Ripley's K function. These implementations though very elaborate lacked the transparency and help required to be adopted by our problem domain. We initially started out calculating the spatial autocorrelations with the tool called Programita, but we wanted to streamline our data mining process of analysis. We found it more convenient to develop our own analytic tool than trying to assimilate the third party tool for our analysis. I developed a Java application that would calculate the Ripley's K function and provide us with the classification of the point patterns under each radii as either clustered, overdispersed and random. We implemented the Monte-Carlo simulations to calculate the confidence intervals, running 500 iterations of random placements of points of the same number used by the players for each of the player-generated maps.

Each trial produced by the participants was represented by the set of (x-y) coordinates representing the swale positions as placed by the learners in a text file. The format for one such text file is shown.



Figure VII: The format of the data file (input) to the Java application

The figure shows the format of the data file (input) to the Java application. The first line indicates the date and time of the trail when it was attempted. The next few lines indicate the experiment number, condition, conditions type trial number, etc. The series of numbers following the meta data, are the x and y coordinates for the swales separated by space. Also each trial was associated with a particular map, which was represented as a text file as well containing the coordinates of the sewers on that map. When calculating the Ripley's L value for a trial, the application would read the coordinates from the respective text file, and give the respective univariate L values. For bivariate L values, the application also refers to the text file containing the coordinates of the sewers for that particular map. The application stores all the L values in different tables in the MySQL database, with one entry for univariate and bivariate for each trial. So the application would give as output L values which are stored in the MySQL database tables with the trial names as the unique keys. Each table would have a maximum of 598 rows (1 for each trial) and 12 columns representing the trail name, and the 11 L values at the radii from 1 through
11. While calculating the Ripley's L value, the edge effects were the most important consideration to be accounted for.

We incorporated these metrics through the Ripley's L function into a third version of our visualization for easy interpretation. The visualization tool enabled us to make sense of the Ripley's L values by observing the visual placements on the study maps and the plots for the Ripley's L values (univariate and bivariate for the whole map only) simultaneously. The screenshot of the visualization tool is presented below.



Figure VIII: The second revision to the visualization tool

The figure shows the visualization tool developed to aid analysis showing two trials (played one after the other) with the Univariate and the Bivariate values (line graphs) for each of the maps below the particular trial. The left panel indicates the legend, the right panel indicates the line graphs for the outcome metrics, and the bar graphs representing Abundance Differential,

Placement Differential, and Spatial Dispersion Dissimilarity for univariate and bivariate. The dots above represent the various studies and the dots below represent the various trials in each study. The color coding is unique for each mode of technology used to attempt the trials.

This visualization was helpful for manually inspecting the solutions produced by the participants, which suggested differences in how participants were exploring the problem space, but we quickly realized that if we wanted more conclusive evidence a data mining approach would be warranted.

### **4 Data Mining Approach**

#### 4.1 <u>Reducing the State Space of the Solutions</u>

As with many open-ended learning problems, in our problem too there are many different ways that learners could attempt to solve the problem. Some may be better than others, but there is no single "right answer" we can judge learners efforts by or guide learners towards. In this problem space, learners can place swales in any of 324 different locations on the map. This means that there are 324! different possible solutions (2.28899746 E+674). To even begin to attempt to apply educational data mining techniques to this problem space, we need to explore ways to "bin" solutions into classes or categories of strategies. These so called strategies were the different spatial arrangements that the learners employed to strike a balance between infiltration and the cost invested. Our goal was to use data mining to track changes in the patterns to infer the spatial strategy or strategies the learners were using to improve their scores.

# 4.2 Initial State Space Reduction Approach: Condensing State Space to "Change" Space

Our first attempt in this direction was in [38] where we designed metrics to reflect changes learners could make to the maps. We designed the 4 differential metrics. These metrics were designed to track trial-to-trial changes in placements, so we could understand how learners were exploring the problem space. We initially thought that the specific strategies used might be less important than the nature of their exploration pattern: for example, slow-and-steady changes (analogous to "hill climbing") might be more effective at producing good outcomes than sharp changes in strategy (analogous to "random restarting"). We thus chose two metrics that could indicate how similar or dissimilar each solution seemed to one another, in terms of the "actions" taken by the users to produce the solutions: Placement Dissimilarity: (PD) was designed to note the changes in the placements of swales. We designed this metric using the Hamming distance called the edit distance. Assuming the map being divided into grid of 1 \* b blocks (1 being the length of the map and b being the breadth) so the edit distance would consider each point on the map to be within the coordinate system of this map. This metric counts the number of points on the map that do not match if the maps are aligned on each other, normalized by the number of swales in each map.

Mathematically, PD was denoted as

$$PD = \frac{\sum_{i=0}^{n} M_a^i - M_b^i}{(N_a + N_b)}$$

### Equation V: Placement Dissimilarity

Where a and b are the subscripts for the map a and map b

 $M_a - M_b$  represents the difference in the sum of squares of the first map that do not match with the second map

 $N_a$ ,  $N_b$  represents the number of swales in map a and number of swales in map b respectively

• Abundance Dissimilarity: (AD) this measure was designed to track changes in the number of swales the learners used. We normalized this metric on the maximum number of swales of the two maps that were being considered. AD was calculated as:

$$AD = \frac{(N_a - N_b)}{MAX(N_a - N_b)}$$

Equation VI: Abundance Dissimilarity

We also wanted to be sure to capture nuance in how (perhaps even slight) changes in placements might cause large impacts to spatial patterns, and so devised two spatially-dependent metrics:

• Spatial Dispersion Dissimilarity: (SDDU, SDDB) To compute the Spatial Dispersion Dissimilarity (SDD) across two maps, a and b, we need only to compute a variant of the Hamming distance across the tuples, where we normalize the value across the number of radii, r, used in the spatial dispersion computations:

$$SDD = \frac{1}{r} \sum_{i=1}^{n} |L_r^a - L_r^b|$$

Equation VII: Spatial Dispersion Dissimilarity

Where r is the radius

 $L_r^a, L_r^b$  are the normalized Ripley's L-values of map a and map b

Though these metrics were able to tell us about the changes the learners made, it appeared our assumptions about the importance of change trajectories in producing effective strategies for this problem space were wrong. We were not able to significantly and reliably correlate these changes in patterns in the trials to changes in the outcome metrics like the infiltration, so using these metrics as the basis for data mining was a non-starter. We needed to re-examine our approach - perhaps a different model of what constituted a strategy was needed.

#### 4.3 <u>Revised State Space Reduction Approach: Simple Spatially-Sensitive Density Metrics</u>

We used the visualization tool I had augmented to revisit the problem space and manually inspect the learner solutions, hoping that browsing through them would yield alternative ideas about how to condense the problem state space. We noticed that the solutions proposed by the learners did seem to vary in interesting ways in terms of density. Rather than examining change, then, we thought we would attempt to characterize each solution in terms of different (seemingly meaningful) types of density. When changing the arrangement pattern of the swales students make changes in the densities of the swales in different regions of the map. We identified a few of them as hot spots like:

- around the sink (i.e., the lowest elevation point on the map)
- around the apogee (i.e., the highest elevation point on the map)
- in the upstream area (i.e., the half of the map closest to the apogee)
- in the downstream area (i.e., the half of the map closest to the sink)
- overall density

While placing the swales on the map the learners would pay more attention to the sink, which motivated us to divide the map into upstream and downstream portions. Here the sink and the apogee are recognized to be the points of lowest elevation and the point of highest elevation respectively. By dividing the map we could observe the patterns that affect infiltration and where those patterns are located -in upstream or downstream. We calculated the density of the swales in and around these hot spots to devise these metrics.

We suspected the changes in the density to be closely related to the infiltration, but this approach lost much of the spatial nuance of the problem – like the prior state space reduction

techniques, the resulting metric were too simplified. Also, to understand how these density values were effecting the infiltration we would have to regress them against the outcome metrics. We realized after an initial inspection that this approach would not work, as these metrics were highly correlated with each other, making it difficult to interpret the results if regression was to be performed.

correlation	sink	Apogee	Number of swales	up swales	down swales
Sink	Х	0.83	0.84	0.63	0.82
apogee	Х	Х	0.79	0.83	0.65
Number of swales	Х	Х	X	0.80	0.95
up swales	Х	Х	Х	Х	0.57
down swales	Х	Х	Х	Х	X

Table II: Correlation matrix for the density metrics

Multiple linear regression requires the variables to be strictly independent. When we computed the correlation matrix for these variables we found them to be highly correlated which was a result of the data being closely linked. This was true as changes in density in one of these metrics would have a ripple effect in some other densities too. For example, if we changed the density near the sink, the density in the downstream area and also the overall density metrics would change simultaneously. Such multi-colinearity in the data can have negative effects like exclusion of the most significant variable from the model [41]. Therefore, we decided to investigate yet

another method of state space reduction, once again returning to the Ripley's metric, but this time preserving its radius-dependent information.

#### 4.4 Final State Space Reduction Approach: Radii-Preservation

Part of the problem with these earlier approaches was that we were losing important information related to *spatial scale*. For example, with the SDDu and SDDB metrics, we were collapsing the spatial metrics from a format that retained information about the radii at which the spatial distribution patterns emerged to a single-value difference metric. Similarly, although density metrics contain a certain inherent scale (i.e., the area used to compute the density), it would conflate all scales that were smaller than that of the area. For our problem space, though, the scale at which a pattern exists matters (in terms of how well it supports water capture), so we needed to revisit our approach so as to be better able to link outcomes (like water capture) to the spatial patterns that exist or coexist across spatial scales.

The Ripley's L calculation, developed for studying the spatial patterns (extensively used to study spatial patterns of plants in ecology), is designed to identify Complete Spatial Random (CSR) patterns in the spatial arrangements at different windows of areas (defined radially). Deviations from CSR can be used to detect two things: whether the elements comprising a given pattern are placed closer together (i.e., are clumped) or are spaced farther apart from one another (i.e., overdispersed) than one would expect by chance, given the radius. This metric allows us to characterize the arrangement patterns the student make with the swales as clumped, overdispersed and random statistically, and allows us to do so at multiple radii. (It can be very possible that a spatial pattern that exhibits overdispersion at one radius of examination can exhibit clumping at a different radius). So at each radii we characterize the swale arrangement made by students to be clumped, random, overdispersed. This is done in a univariate fashion, to discover how students are placing swales relative to each other, and in a bivariate fashion, to discover how the students are placing swales relative to other map elements, like sewers.

Thus, in this new state-space reduction approach, the Ripley's L values are calculated for each trial as univariate and bivariate, and at each radii varying from 1 to 11 (half of the length of the study area) we observe the pattern to either to be clumped, random or overdispersed. We calculate the Ripley's L values on all the trails:

- Univariate pattern: This observes the pattern that the swales form with respect to the other swales placed in the study map.
- Bivariate pattern: This observes the pattern that the swales form with respect to the sewers on the study map.

Each of these Ripley's L metrics are a vector of length 11, describing the L-value at each radii 1 through 11. In order to test the deviation from randomness (dispersion or clustering) of the point patterns using the univariate or the bivariate functions, we computed a 99% confidence interval of L(t) using the Monte Carlo method from 500 simulated CSR patterns with the same number of points contained inside a region with the same geometry [32]. The points above the confidence interval displayed clumped patterns whereas the points below the lower confidence interval displayed an over-dispersed pattern (regular pattern). I will now provide an illustrative example:



Figure IX: Visualization of learners' trials

The figure shows a Visualization of the learners map in one of the trials. The green patches represent the swales they placed on the map.

 $L_u = \{1.211, 10.184, 21.339, 26.323, 29.863, 33.919, 37, 264, 42.014, 45.811, 49.486, 51.841\}$ 

These are Ripley's L values for the map shown above. These values are also plotted in the graph below with the lower and the upper bounds of the confidence intervals.



Figure X: Plot of the Univariate L-Values

The plot above shows the Univariate L-values (blue line) within the confidence intervals (red line and grey line). The points above the red line exhibit a clumped pattern and below the grey line would be overdispersed, the points between the confidence interval exhibit a random pattern. The plot is shown for the arrangement shown above, at radii 3 through 11 the arrangement is clumped and at radii 1 and 2 it random.

Based on the confidence intervals we denote the radii which have clumped arrangements by 1, overdispersed arrangements by -1 and random arrangements by 0. We call this notation to be normalized L value notation. We hypothesized that we could find variations in the strategies at different radii by comparing these normalized values. For the visualization map shown in *figure IX*, then, the normalized L-values can be written as:

$$L_{norm} = [0,0,1,1,1,1,1,1,1,1,1]$$



Figure XI: Plot of normalized L values

Radius

The figure above shows a plot of the normalized L values at various radii in three consecutive trials in Study 1. The plot exhibits how variation in strategies can be seen with the normalized L-values. The first plot shows clumping at all radii except for radii 1 and 2, the second plot shows clumping at 3 through 11 and random pattern at the other radii, whereas the third plot shows clumping only through 5 to 10 and random pattern at the other radii. What this progression shows is that the participants changed the "granularity" of the clumping of the swales as they explored the problem space, moving from fairly tightly-packed clumps to more loosely-packed clumps.

For ease of later analysis, we split each 11-tuple into two: one which captured the presence of any significant clumping, and one which captured the presence of any significant overdispersion. Because we were computing these metrics for both univariate and bivariate distributions of swales, we ended up with a 44-tuple to represent each solution, giving us a state space with a theoretical max of 44! (or 2.7 e54), although because one can never have a distribution at a radius which is both clumped and overdispersed at the same time, the maximum is actually 22! (or 1.1 e21). While large, this is still a great deal smaller than our original 324! state space.

## 4.5 <u>Strategy Detection Approach</u>

We used linear regression to help us discover which of the strategies the learners adopted more prominently helped them to improve infiltration (because no matter where they place the swales some infiltration is bound to happen, although some arrangements are superior to others). In order to get more detailed picture of which patterns at which radii would increase the infiltration, we used linear multiple regression to find the relationship between the patterns at different radii and the infiltration. By doing so we identified the strategies or explicit type of arrangements that would support increase in infiltration with respect to the data generated by participants during the experimental trials.

Regression analysis is a statistical process of estimating the relationship among the variables. It helps understand how the typical value of the dependent variable changes when any one of the independent variables are varied while others are held fixed assuming that the independent variables do not possess any multi-collinearity. Multi-collinearity is present among the independent variables when the independent variables are correlated with each other. As we noted above, in explaining why we abandoned our density-based state-space reduction approach, multi-collinearity is not desirable as it reduces the precision of parameter estimates and thus makes it difficult to draw conclusions from the model [41].

The result of regression is an estimation function of independent variables called the regression function. These regression functions can be used to understand which among the independent variables are related to the dependent variables and to explore the forms of these relationships. Among many forms of linear regressions are parametric, in that the regression function defined in terms of unknown parameters that are estimated from the data  $(b_i)$ . Multiple regression is a type of linear regression that has multiple independent variables and one outcome variable [42].

Multiple linear regression is basically a model represented by:

 $y = b_0 + x_1 b_{1+} x_2 b_{2+\dots+} x_k b_k$ 

Equation VIII: Multiple Linear Regression

The equation represents the model specified by the data,

Where y= outcome variable or the dependent variable

 $x_1, x_2, \dots, x_k$  = independent variables

 $b_1, b_2, \dots, b_k$  = coefficients of the independent variables

We used multiple linear regression to validate the value of the spatial strategies we identified. Linear regression is extensively used to provide descriptions of the relationship between variables, for screening or selecting variables which explain a significant amount of variation in the outcome variables and also for prediction [41]. We wanted to know if these identified strategies had a linear relationship with the outcome metrics of infiltration and the cost rewards and if so which arrangements better explain the variation in infiltration. It is important for us to be able to

consider strategies combinatorially because we know that in this spatial problem space, certain strategies often necessarily co-exist (for example, attempts to over-disperse swales at smaller radii are also likely to increase the clumpiness of swale placements when they are considered at larger radii).

We utilized multiple linear regression in two different ways: first to confirm whether or not the strategies we pre-identified as likely to have an impact on the outcome variables which would be evident through the linear relationship, and secondly to allow us to identify the strategies which had a positive impact on the outcome metrics and which had a negative impact, as this could be determined by the signs on the parameters of the model.

We performed regressions on 4 sets of data. First, each subset of solutions generated during each user interface condition (182 solutions in the paper condition, 205 in multimouse, and 211 in single mouse) then on the entire corpus of solutions generated during the experiments (598 solutions). We first modeled the strategies used to good effect within each condition because we.expected the interface design to affect which strategies the participants explored. Differences in the regression parameters would indicate that the participants' exploration of the strategy space was mediated by the interface design.

Each of these regressions gave us a model describing the variables which affected the outcome variables and they could explain the variation in the outcome variables. Every term in the linear relationship would have a coefficient  $(b_i)$  associated with it which indicates the slope, it estimates the average value of y changes by  $b_i$  units for each 1 unit change in  $x_i$  holding all other variables constant. These regression models are evaluated for accuracy by a metric called the coefficient of determination and described by the term  $R^2$ . The  $R^2$  value reports the proportionality of total variation in y explained by all x variables taken together. The value is

between 0 and 1. A best fit would be described by the coefficient of determination to be 1. Another value called the adjusted coefficient of determination (Adjusted  $R^2$ ) reports the proportionality of variation in y explained by all x variables adjusted for the number x variables used. When dealing with multiple regression the Adj  $R^2$  is a better metric to measure the goodness of fit to the model. Besides these measures which tell us the strength and efficiency of the model, the significance of individual variable can be explained by the t-tests or the P-values of the individual slopes  $b_i$ .

In order to find the model that fits the data well we used stepwise regression. The stepwise regression starts from one model and proceeds in a greedy fashion, only including those independent variables in the model that are significant. The significance of each variable is calculated at each step deciding either to include or exclude the variable from the model. We used the Akaike information criterion that gives the measure of the statistical model for a given set of data for selecting the variables for the model [43].

Now, one might argue that a better approach would have been to run regression on the full problem space (of size 324!), which would have entailed generating all possible maps, running the simulation on them to obtain the infiltration values, and performing the Ripley's analysis to generate their 44-tuple strategy description for the regression. Technically possible, but time-consuming, as each run of the simulation takes about 1 minute on a normal computer. Because the effective strategies are highly dependent on the underlying geography of the maps, this approach would also not be very scalable, as the vision for the methods we devise here would be to allow educators to construct their own maps to better match a neighborhood familiar to the learners. A simple change in block size would completely alter the efficacy of different spatial strategies.

Alternatively, one could also propose to reverse-engineer the strategy space (recall this is smaller, of size 22!), to construct Cartesian arrangements of swales that would produce a given strategy tuple, so that the simulation could be run on the swale arrangement to produce the infiltration value. While technically feasible, the fact is that there are many possible ways to create a map to match a given strategy tuple, and so great care would be needed in reverse-engineering the strategies to produce swale coordinates so that this process does not inadvertently produce systematic nature in the candidate set of maps that would skew the results of the regression. Perhaps the best argument for using the experimental data, though, is the fact that because there are so many possible maps that are fairly equivalent to one another in terms of their influence on infiltration, a multiple linear regression on the entire strategy space would have produced an equation with an unwieldy number of influential terms, many of which would likely reflect strategies never even attempted by users. We are more interested in examining the portion of the strategy space actually explored by learners, not in exploring the strategy space itself. Put in educational terms, we are interested in examining the strategies within the learners' "Zone of Proximal Development" [44] – meaning, the space of strategies within reach of their current (novice-level) understanding of the problem. For all of these reasons, we decided to use learnergenerated solutions to help us flag potentially beneficial spatial strategies.

# 5 Results

### 5.1 Regression Results: Identifying Differences in Effective Spatial Strategies across Interfaces

Regression identified the patterns that efficiently explained the relationships between arrangements and the outcome metrics of infiltration. We regressed both the univariate and bivariate spatial metrics for the whole map against the infiltration measure. For the regression results to make more sense we held the random column under each radius as a comparative "strategy." So the coefficients we would get from the regression would express how the clumped and overdispersed strategies at a particular radii compared against instances where the swales were effectively placed randomly. We regressed this separately for all conditions, and separately for paper, multi mouse, and single mouse against infiltration, in each case we got different models indicating differences in the significant spatial strategies at play.

The coefficients can be interpreted as comparisons against a random placement at a given radius; to a certain extent, this random placement may be regarded as an absence of a spatially-sensitive strategy at that level of scale. So a positive coefficient for the clumped pattern at radius X, would mean that it had more of a positive influence than the random placement at radius X and a negative coefficient at radius Y would mean it had a negative influence than the random placement at radius Y on the infiltration. The same logic applies to coefficients for the overdispersion strategies. We identified the strategies with positive coefficients as "good strategies" and strategies with negative coefficients as "bad strategies," because the positive coefficients terms add up to give a better infiltration whereas the negative coefficient terms tend to reduce infiltration.

The tables below indicate the coefficients for the paper trials, followed by the multi-mouse, and the single mouse trials. The table only shows the coefficients that are significant for the model. Each variables' significance is indicated by the pValues column shown.

Significant Variables	Estimate	tStat	pValue
clumped Univariate 1	1090.3	2.16	3.19E-02
overdispersed Univariate 1	2740.8	5.34	3.03E-07
clumped Univariate 2	-1044.8	-1.58	1.15E-01
clumped Univariate 4	3632.7	3.71	2.82E-04
clumped Univariate 6	-1817.2	-2.06	4.10E-02
clumped Univariate 8	2985.7	5.12	8.43E-07
clumped Bivariate 1	3520.7	6.65	3.95E-10
clumped Bivariate 2	3671.1	6.21	3.94E-09
overdispersed Bivariate 4	1403.8	2.79	5.79E-03
clumped Bivariate 7	4231.4	9.72	5.60E-18
overdispersed Bivariate 10	-1632.8	-1.76	7.88E-02
overdispersed Bivariate 11	3555.8	3.97	1.05E-04

Table III: The coefficients of the significant variables for the paper trials

Significant Variables	Estimate	tStat	pValue
overdispersed Univariate 1	2907.5	5.07	9.39E-07
overdispersed Univariate 2	1827.2	3.21	1.55E-03
clumped Univariate 3	1968.4	3.69	2.95E-04
clumped Univariate 8	961.47	1.44	1.50E-01
clumped Univariate 11	1511.7	2.24	2.59E-02
Overdispersed Univariate 11	-2471.2	-3.51	5.54E-04
overdispersed Bivariate 1	2267.5	3.49	5.93E-04
clumped Bivariate 2	2498.4	2.96	3.44E-03
overdispersed Bivariate 3	1127.7	1.90	5.85E-02
overdispersed Bivariate 5	2235	4.38	1.90E-05
clumped Bivariate 6	4951	4.02	8.14E-05
clumped Bivariate 7	2753.2	3.14	1.96E-03
clumped Bivariate 10	-1929.4	-3.38	8.87E-04

Table IV: The coefficients of the significant variables for the multi-mouse trials

Significant Variables	Estimate	tStat	pValue
clumped Univariate 1	1200.5	3.92	1.22E-04
overdispersed Univariate 1	1264.6	2.52	1.26E-02
overdispersed Univariate 2	1880.6	3.00	3.05E-03
overdispersed Univariate 6	7406.9	5.46	1.46E-07
clumped Univariate 7	1248	2.92	3.90E-03
overdispersed Univariate 7	-5193.4	-2.79	5.79E-03
clumped Univariate 10	-2365	-2.79	5.76E-03
clumped Univariate 11	3691.4	4.61	7.25E-06
overdispersed Bivariate 1	2694.5	5.49	1.28E-07
clumped Bivariate 2	2334.3	3.17	1.78E-03
overdispersed Bivariate 2	2450	4.60	7.84E-06
clumped Bivariate 4	1978.6	2.00	4.69E-02
overdispersed Bivariate 4	-1157.5	-2.52	1.25E-02
clumped Bivariate 6	2770.2	4.41	1.72E-05
overdispersed Bivariate 6	1402.5	2.51	1.28E-02
clumped Bivariate 7	1455.1	2.39	1.76E-02
overdispersed Bivariate 7	2435	3.55	4.94E-04
clumped Bivariate 8	-1015.2	-1.64	1.03E-01
overdispersed Bivariate 8	-1685.5	-2.37	1.88E-02
clumped Bivariate 9	2262.2	5.01	1.25E-06
overdispersed Bivariate 9	1367.1	1.99	4.79E-02
overdispersed Bivariate 10	2670.2	4.64	6.29E-06

Table V: the coefficients of the significant variables for the single mouse trials

It can be hard to compare these 3 tables, so we also created a table that just visualizes the polarity and magnitude of the coefficients for the three conditions:

radius	paper - i	nfiltration			multi mouse - infiltration			ation single mouse - infiltration				
	o-u	o-b	c-u	c-b	o-u	o-b	c-u	c-b	o-u	o-b	c-u	c-b
1	2740.8		1090.3	3520.7	2907.5	2267.5			1264.6	2694.5	1200.5	
2			- 1044.8	3671.1	1827.2			2498.4	1880.6	2450		2334.3
3						1127.7	1968.4					
4		1403.8	3632.7							- 1157.5		1978.6
5						2235						
6			- 1817.2					4951	7406.9	1402.5		2770.2
7				4231.4				2753.2	- 5193.4	2435	1248	1455.1
8			2985.7				961.47			۔ 1685.5		- 1015.2
9										1367.1		2262.2
10		- 1632.8						۔ 1929.4		2670.2	-2365	
11		3555.8			- 2471.2		1511.7				3691.4	

Table VI: Coefficients of the three regression models. Each cell indicates a different strategy at a different radius: so, for example, we can see that the Overdispersed Univariate (o-u) strategy at radius 1 was a "good" strategy regardless of condition

*Table VI* shows the coefficients for each of the three regression models, using coloration to indicate degree of positive (green) or negative (red) of each strategy's impact on infiltration. Here c-u indicates clumped univariate, c-b indicates clumped bivariate, o-u indicates overdispersed univariate, o-b indicates overdispersed bivariate Note that while there are some commonalities (for example, overdispersed-univariate was found to be effective at radius 1 for all three models) there are still a fair number of disagreements across the models. Note also that in the single mouse case, users discovered an interesting boundary where univariate overdispersion went from being effective at radius 6 to being negative at radius 7 (this contrast likely arises because blocks are of size 6x6 - apparently, participants in the single-mouse case discovered that overdispersed patterns which spanned blocks were not as effective as overdispersed patterns within blocks).

	The	models	obtained	from	regression	had	a	fairly	good	fit	to	the	data,	their
R² val	ues a	and adju	sted R <sup>2</sup> v	alues a	re indicated	in th	e ta	ble bel	ow.					

Conditions	Paper	Multi-mouse	Single mouse
Rsq	0.86	0.81	0.90
adjusted Rsq	0.85	0.80	0.89
F-statistic	87.3	62.6	77.3
pValue	2.09E-65	1.44E-61	1.81E-81

Table VII: Goodness of fit for the models

The models were able to mine the solutions for each interface with a fairly good coefficient of determination for each model. We were able to identify that the spatial strategies that bubbled up as being effective were different for the different interfaces. This variation in significant variables across the different interfaces suggested that the interface did have an effect in shaping the strategies used by the learners. Mind you, this does not necessarily speak to the overall effectiveness of one interface over another at producing effective strategies – on average, participants in this within-subject trial obtained an infiltration of 6115.0 gallons when using the paper condition, 6229.8 gallons when using the multimouse condition, and 5226.7 when using the single mouse condition. None of the differences in infiltration were statistically significant. What this model regression is telling us, though, is that while the participants may have attained similar levels of performance when using each interface, the *way* they went about obtaining those outcomes was possibly affected by the interface condition. In other words, their exploration of the strategy space was likely affected. The next section will investigate this claim further.

# 5.2 Solution Discovery: Comparing and Tracking Learner's Use of Spatial Strategies

While the models constructed by individually regressing the paper, multi-mouse, and single mouse trials could suggest whether there were differences in which strategies were more or less prominent in producing results, this by itself does not count as evidence – it could be the case that the differences in the models were artifacts of the data and not indicative of differences in the influence of the interface on spatial reasoning. We needed a mechanism to compare the strategies across interface designs more transparently.

Significant Variables	Estimate	tStat	pValue
clumped Univariate 1	1256.3	4.54	6.65E-06
overdispersed Univariate 1	2866.2	8.88	8.33E-18
overdispersed Univariate 2	977.01	2.60	9.62E-03
clumped Univariate 3	1050.5	3.29	1.07E-03
clumped Univariate 8	1332.5	3.12	1.87E-03
overdispersion Univariate 10	-1921.2	-3.19	1.50E-03
clumped Univariate 11	1211	3.00	2.77E-03
overdispersed Bivariate 1	2635.2	8.22	1.28E-15
clumped Bivariate 2	2809.5	6.28	6.49E-10
overdispersed Bivariate 3	541.62	1.67	9.53E-02
clumped Bivariate 4	1119.7	1.89	5.92E-02
overdispersed 5 Bivariate	726.19	2.21	2.76E-02
clumped Bivariate 6	1691.7	2.95	3.31E-03
overdispersed Bivariate 6	1565.1	4.50	8.36E-06
clumped Bivariate 7	3488.8	8.86	9.66E-18
clumped Bivariate 9	1002.1	2.22	2.68E-02
clumped Bivariate 10	-1390.3	-3.22	1.37E-03
overdispersed Bivariate 10	891.09	2.96	3.24E-03

Table VIII: The coefficients of the significant variables for all the trials

For this purpose, we regressed all the trials, irrespective of the interface that they were attempted in. The regression model would give us the significant variables which are found to be more generally effective at producing positive infiltration, and we could then use these parameters to compare across the conditions. The parameters of the model obtained from the regression are tabulated below. The model had a coefficient of determination ( $R^2$ ) value to be 0.802 (p< 0.05). This model fairly good fit for the data at hand.

	o-u	o-b	c-u	c-b
1	2866.2	2635.2	1256.3	
2	977.01			2809.5
3		541.62		
4				1119.7
5		726.19	1050.5	
6		1565.1		1691.7
7				3488.8
8			1332.5	
9				1002.1
10	-1921.2	891.09		-1390.3
11			1211	

Table IX: An alternate representation of the coefficients in Table VII above

*Table IX.* is an alternate representation of the coefficients in *Table VIII* above highlighting their magnitude and polarity. False coloring indicates the degree of positive (green) or negative (red) impact each strategy had on infiltration.

We used this model to examine the learners' exploration mechanism in the problem space. This model allowed us to ask if the different interface designs might affect things like:

- The total number of good spatial strategies used in the different conditions (i.e., are some of the interfaces more likely to encourage the use of good strategies than others?),
- How the participants went about exploring the space of strategies identified as good strategies, as measured by trial-to-trial changes in the applications of good strategies
- If participants were more likely to *explore* good strategies or to *exploit* good strategies (measured by the number of unique strategies participants tested within a condition i.e., strategies used only once. Exploitation would show fewer unique strategies)
- How long it took participants to discover good strategies (in terms of the number of trials it took before a strategy was employed)

Let's examine each of these in turn.

	paper	multi-mouse	single mouse
TOTAL	572	676	530
AVERAGE PER GROUP	19.07	22.53	17.67
(STDEV)	(19.63)	(20.58)	(19.16)
AVERAGE PER TRIAL (STDEV)	2.95 (2.63)	3.25 (2.89)	2.40 (2.35)

Table X: The total number of good strategies, by condition

We compared the total number of good strategies across the interface trials, and it seemed that, in the multi-mouse condition, participants employed slightly more total good strategies, followed by the paper condition, with the single-mouse condition showing the smallest total number of good strategies used, although we found that none of these differences were significant (see *Table X*).

	paper	multi-mouse	single mouse
AVERAGE	1.52	1.17	0.75
STDEV	1.56	1.10	0.68

Table XI: The average delta in number of good strategies used, trial-to-trial.

We compared the average of trial to trial (delta) changes in specific good strategies used. We found that in the paper condition, participants showed more change in specific "good" strategies they employed from one trial to the next, followed by the multi-mouse condition, and with the singlemouse condition showing the smallest amount of change in the strategies used. This difference was significant according to a within-subject ANOVA (F=4.43, p =0.0162). A post-hoc Bonferroni-Holm correction revealed that the only significant pairwise difference was between the paper and the single mouse conditions. A higher trial-to-trial delta in good strategies employed indicates that a given trial is less similar to the trial that preceded it. Because we are only tracking the change in the application of strategies known to positively impact the outcome, the presence of a higher delta indicates that the participants are more active in exploring the space of good solutions, an activity that is more likely to yield a meaningfully different outcomes. It can be seen as a marker of productive exploration of the *strategy* space (which is different from exploring the problem space: with an open-ended problem space, wide exploration can all too easily be nonproductive). Thus, participants explored the strategy space significantly more effectively in the paper condition than in the single mouse condition. The fact that the strategy exploration of the participants was middling for the multi-mouse condition suggests that distributed control,

regardless of whether it is accomplished with a TUI or with mice, does seem to promote more strategy space exploration.

	paper	multi-mouse	single mouse
AVERAGE UNIQUE STRATEGIES, PER GROUP	5.30	5.83	4.03
STDEV	4.40	4.56	3.39
AVERAGE UNIQUE STRATEGIES, PER TRIAL	1.01	0.94	0.61
STDEV	0.93	0.88	0.48

Table XII: The number of unique good strategies tested by groups

The interpretation that the interface condition seems to affect strategy exploration also seems to be supported when we examine the number of unique strategies (i.e., good strategies tested only once), using this as a proxy for exploration (as opposed to exploitation). In the multi-mouse condition, when summed across all of their trials participants employed more unique good strategies (5.83), which is the count of strategies used once and only once across all trials in the condition. This is followed by the paper condition (5.30) with the single-mouse condition showing the smallest total number of unique strategies used (4.03). This difference was significant according to a within-subject ANOVA (F=3.81, p =0.0278). A post-hoc Bonferroni-Holm correction revealed that the only significant pairwise difference was between the multi and the single mouse conditions. This suggests that participants were most likely to return to previously-used strategies in the single-mouse condition, which again reinforces the idea that participants were not exploring the strategy space as much in the single mouse condition as they were in the multi-mouse condition. One possible explanation for this marker of a lack of exploration is the lack of multi-user control - it can be hard to make meaningful change to a proposed solution when

only one person is working on doing so, as individuals may be more prone to prematurely committing to a strategy, and thus more likely to explore variations around that strategy rather than exploring the strategy space more broadly. When one breaks down the unique use of good strategies on a per-trial basis, however, the differences become even more apparent. (This actually might be a more fair comparison, since the speed of use of the interfaces differed, as seen in the differences in the number of trials participants could complete in each condition). Although in the multi-mouse condition, participants employed more unique good strategies in total, when the number of unique strategies is averaged by trial, we see that the paper condition averages the largest number of unique good strategies per trial (1.01), with multi-mouse not far behind (0.94), and the single mouse condition showing once again the smallest average of unique strategies used (0.61). This difference was significant according to a within-subject ANOVA (F=4.67, p =0.01316). A post-hoc Bonferroni-Holm correction revealed that there were two significant pairwise differences, between the paper and the single mouse conditions, and between the multi and the single mouse conditions. This suggests that participants explored the solution space more in the paper and multi-mouse control conditions, and did so more efficiently, trying out an average of 1.04 and 0.94 new unique strategies in each trial, respectively. The much lower ratio of 0.61 new unique strategies per trial in the single mouse case further suggests that participants did not explore as broadly in that condition.

We also analyzed the data to determine the order of discovery for the good strategies. We wanted to observe how the interface designs influenced this discovery process. By "discovery" we refer to how many trials were needed before a group employed a given strategy – a strategy used during trial 1 would have a discovery order of 1, and a strategy never used until the third trial would have a discovery order of 3. To obtain these numbers, we first converted each trial into a binary 16-

tuple indicating the presence or absence of each of the 16 significantly "good" strategies highlighted by the regression model. Then we multiplied each tuple by the order of that trial within it's condition – so if a given trial was the fourth attempted, any 1s in the 16-tuple would be converted to 4s. Then, for each of the 3 conditions within the 30 experiments, we created another 16-tuple that recorded the earliest occurrence of each of the 16 strategies.

Group #	Trial #	c- u r1	c- u r3	c- u r8	c-u r11	od- u r1	od- u r2	c- b r2	c- b r4	c- b r6	c- b r7	c- b r9	0- b r1	o- b r3	0- b r5	0- b r6	o-b r10
9	1																
9	2		2	2													
9	3	3	3	3	3												
9	4	4	4	4	4												
9	5	5	5	5	5									5	5		
9	6	6	6	6	6							6		6	6		
9	7	7	7	7	7							7		7	7		
9	8	8	8	8	8							8		8	8		
9	9	9	9	9	9							9		9	9		
First Appearance:		3	2	2	3							6		5	5		

*Table XIII: Illustration of Group 9's exploration of the solution space* 

*Table XIII* is an illustration of how for Group 9's exploration of the solution space proceeded in the multimouse condition, and how we distilled the order of their strategy. They uncovered the clumped univariate strategies fairly early during their 9-trial exploration, and took a bit longer to uncover the clumped bivariate strategy at radius 9 and the overdispersed bivariate strategies at radii 3 and 5.

The following table was constructed with the average of the sums of these trials.

		1	
Conditions	paper	multimouse	single mouse
Strategy appearance count	159	175	128
Average first appearance of strategies	2.77	2.82	3.70
STDEV first appearance of strategies	2.16	2.21	2.59

*Table XIV: Summary of the appearance of good strategies (as identified by the multi-linear regression model) in trials generated in the three conditions* 

The table suggests that the interface designs had some impact on the order of discovery of the good strategies. Moreover, in the paper and multi-mouse interface the learners were found to be discovering the good strategies faster than single mouse trials.

Order Discovery	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	74.931	2	37.465	7.05	0.001
Within Groups	2439.115	459	5.314		
Total	2514.045	461			

Table XV: ANOVA results for order of discovery of good strategies

Order Discovery	(I) condition	(J) condition	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	paper	multi	-0.044	0.253	0.984	-0.64	0.55
		single	922*	0.274	0.002*	-1.57	-0.28
	multi	paper	0.044	0.253	0.984	-0.55	0.64
		single	878*	0.268	0.003*	-1.51	-0.25
	single	paper	.922*	0.274	0.002*	0.28	1.57
		multi	.878*	0.268	0.003*	0.25	1.51

Table XVI: Post-hoc tests for the order discovery of good strategies

An analysis of variance (ANOVA) on these scores again yielded significant variation among conditions, F (2, 259) = 7.05, p < 0.05. A post hoc Tukey test showed that the paper, and multimouse conditions differed significantly from the single mouse condition at p < 0.05; indicating that the interface designs were influencing the strategy discovery patterns.

### 6 Conclusion & Future Work

This paper showcases one of the first attempts at devising metrics to track spatial reasoning and sets the stage for studying how that skill is acquired. There are whole classes of problems involving complex systems and spatial reasoning that are appearing in national educational standards, but which are not currently instructionally supported, owing to a lack of teaching tools and a lack of assessment approaches. Until these problems can be addressed, these standards are nothing more than empty mandates. Our approach has contributed by devising a mechanism to analyze such problems and give feedback to the designers about the effectiveness of the interfaces. These analytic procedures could be adapted to problems involving spatial reasoning even outside urban planning and environmental science, like in analyzing road accidents, disease outbursts, etc. This paper exemplifies how known statistical tools from different domains can be used together to address an ill-structured problem, the innovation here was to put them together to analyze the situation.

The results we uncovered here provide evidence that user interface design can in fact impact how learners explore spatial problem spaces. The recommendation to designers seems to be that if one wants to promote more and earlier exploration (as opposed to exploitation) of spatial strategies, providing an interface that allows multiple users to all contribute to the spatial pattern is a good approach. This work also enables a number of other analyses to be performed (which we have not yet attempted): examining if learners "get stuck" exploring patterns at certain radii (which would indicate that they might need guidance to help them consider incorporating new spatial scales), if learners struggle to perceive certain types of spatial patterns (an initial examination of the data, not reported on here, suggests that learners might take longer to realize that overdispersion is a strategy that can be employed), if certain types of explorations of the strategy space (for example, more systematically combining good strategies) is more effective at discovering good outcomes, and, ultimately, if any of these observable patterns of behavior result in a greater understanding of spatial phenomena (for which we would need to conduct another experiment where we interviewed participants before and after their use of the software).

Through this work we were able to reduce the solution space of a complex spatial problem, while maintaining the spatial features of the data intact. This was crucial for our task as we wanted to study which spatial patterns were effective in this problem space, and how learners went about discovering and applying those spatial patterns. In future work we wish to follow up on this nascent exploration of the model space underpinning spatial reasoning in this domain, and to use our findings to help guide learners to more productively and methodically explore spatial problem spaces. We see great potential in this, as it is too often the case that learners can get "lost" exploring open-ended problems, which could result in them not being able to get adequate exposure to comparing and contrasting effective strategies. As we further refine our model of the spatial reasoning learners may exhibit in this domain, we would be able to devise software-based interventions (termed "scaffolding" in the education literature) to guide the learner towards better explorations even in a complex solution space environment.

### BIBLIOGRAPHY

- R. S. Baker and K. Yacef, "The State of Educational Data Mining in 2009: A Review and Future Visions," *Journal of Educational Data Mining*, vol. 1, no. 1, pp. 3-17, 2009.
- [2] V. J. Shute, "Stealth Assessment in computer based games to support learning," in *Computer Games and Instruction*, Charlotte, NC, Information Age Publishers, 2011, pp. 503-524.
- [3] R. Glaser, J. W. Pellegrino and N. Chudowsky, Knowing What Students Know: The Science and Design of Educational Assessment, Washington D.C.: The National Academic Press, 2001.
- [4] E. Harpstead, C. J. MacLellan, K. R. Koedinger, V. Aleven, S. P. Dow and B. A. Myers, "Investigating the Solution Space of an Open-Ended Educational Game Using Conceptual Feature Extraction," in *6th International Conference on Educational Data Mining*, Memphis, Tennessee, 2013.
- [5] E. Andersen, Yun-En Liu, E. Apter, F. Boucher-Genesse and Z. Popovic´, "Gameplay Analysis through State Projection," in *International Conference on the Foundations of Digital Games*, Monterey, California, 2010.
- [6] R. Martinez-Maldonado, K. Yacef and J. Kay, "Data Mining in the Classroom: Discovering Groups Strategies at a Multi-tabletop Environment," in 6th International Conference on Educational Data Mining, Memphis, Tennessee, 2013.

- [7] J. Gobert, M. Sao Pedro, J. Raziuddin and R. Baker, "From Log Files to Assessment Metrics for Science Inquiry Using Educational Data Mining," *Journal of the Learning Sciences*, vol. 22, no. 4, pp. 521-563, 2013.
- [8] A. N. Rafferty, J. Davenport and E. Brunskill, "Estimating Student Knowledge from Paired Interaction Data," in 6th International Conference on Educational Data Mining, Memphis, Tennessee, 2013.
- [9] M. Desmarais and F. Lemieux, "Clustering and Visualizing Study State Sequences," in 6th International Conference on Educational Data Mining, Memphis, Tennessee, 2013.
- [10] P. Jarusek, M. Klusacek and R. Pelanek, "Modeling Students' Learning and Variability of Performance in Problem Solving," in 6th International Conference on Educational Data Mining, Memphis, Tennessee, 2013.
- [11] K. E. DiCerbo and K. Kidwai, "Detecting Player Goals from Game Log Files," in 6th International Conference on Educational Data Mining, Memphis, Tennessee, 2013.
- [12] J. Muller, A. Kretzschmar and S. Greiff, "Exploring Exploration: Inquiries into Exploration Behavior in Complex Problem Solving," in 6th International Conference on Educational Data Mining, Memphis, Tennessee, 2013.
- [13] G. Biswas, K. M. Loretz and J. R. Segedy, "Model-Driven Assessment of Learners in an Open-Ended Learning Environment," in *Third International Conference on Learning Analytics and Knowledge*, New York, NY, 2013.
- [14] J. Beck and J. Mostow, "How who should practice: Using learning decomposition to evaluate the efficacy of different types of practice for different types of students," in 9th International Conference on Intelligent Tutoring Systems, Montreal, 2008.
- [15] T. Hsieh and T. Wang, "A Mining based approach on discovering courses pattern for constructing suitable learning path," *Expert System and Applications*, vol. 37, no. 6, pp. 4156-4167, 2010.
- [16] R. Baker, S. M. Gowda and A. T. Corbett, "Automatically detecting a student's preparation for future learning: Help use is key," in 4th international conference on educational data mining, Eindhoven, Netherlands, 2011.
- [17] M. Cocea, A. Hershkovitz and R. Baker, "The Impact of Off-task and Gaming Behaviors on Learning: Immediate or Aggregate?," in 14th International Conference on Artificial Intelligence in Education, Brighton, UK, 2009.
- [18] C. Forsyth, H. Butler, A. C. Graesser, D. Halpern, K. Millis, Z. Cai and J. Wood, "Higher contributions correlate with higher learning gains," in *3rd International Conference on Educational Data Mining*, Pittsburgh, PA, 2010.
- [19] H. Jeong and G. Biswas, "Mining Student Behavior Models in Learning-by-Teaching Environments," in *1st International Conference on Educational Data Mining*, Montreal, Canada, 2008.
- [20] I. Arroyo, H. Mehranian and B. Woolf, "Effort based tutoring: An empirical approach to intelligent tutoring," in *3rd International Conference of Educational Data Mining*, Pittsburgh, PA, 2010.

- [21] F. Bouchet, R. Azevedo, J. Kinnebrew and G. Biswas, "Identifying students' characteristic learning behaviors in an intelligent tutoring system foresting self regulated learning," in 5th International Conference of Educational Data Mining, Chania, Greece, 2012.
- [22] K. Muldner, W. Burleson, B. Van de Sande and K. Van Lehn, "An analysis of students' gaming behavior in an intelligent tutoring system predictors and impacts," *User Modelling and User Adapted Interaction*, vol. 21, no. 1-2, pp. 99-135, 2011.
- [23] A. Corbett, "Cognitive Computer Tutors: Solving the Two-Sigma Problem," in *International Conference on User Modeling*, Sonthofen, Germany, 2001.
- [24] M. Pechenizkiy, T. Calders, E. Vasilyeva and P. De Bra, "Mining the Student Assessment Data: Lessons Drawn from a Small Scale Case Study," in *1st International Conference on Educational Data Mining*, Montreal, Canada, 2008.
- [25] M. H. Falakmasir, Z. A. Pardos, G. J. Gordon and P. Brusilovsky, "A Spectral Learning Approach to Knowledge Tracing," in 6th International Conference on Educational Data Mining, Memphis, Tennessee, 2013.
- [26] N. R. Council, A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas. Committee on a Conceptual Framework for New K-12 Science Education Standards., Washington D.C.: The National Academic Press, 2011.
- [27] E. S. Minor and D. L. Urban, "A Graph-Theory Framework for Evaluating Landscape Connectivity and Conservation Planning," *Conservation Biology*, vol. 22, no. 2, pp. 297-307, 2008.

- [28] M. L. Zellner, "Embracing Complexity and Uncertainty: The Potential of Agent-Based Modeling for Environmental Planning and Policy," *Planning Theory and Practice*, vol. 9, no. 4, pp. 237-457, 2008.
- [29] T. Shelley, L. Lyons, J. Shi, E. Minor and M. Zellner, "Paper to parameters: designing tangible simulation input," in *12th ACM International Conference adjunct papers on Ubiquitous computing*, New York, NY, 2010.
- [30] T. Falcão and S. Price, "What have you done! The role of 'interference' in tangible environments for supporting collaborative learning," in 8th International Conference on Computer Supported Collaborative Learning, Rhodes, Greece, 2009.
- [31] T. Shelley, L. Lyons, E. Minor and M. Zellner, "Evaluating the embodiment benefits of a paper- based TUI for educational simulations," in 29th International Conference on Human Factors in Computing Systems, New York, NY, 2011.
- [32] P. M. Dixon, "Ripley's K function," in *Encyclopedia Environmetrics*, NJ, Wiley, 1995, pp. 1796-1803.
- [33] M. Cimons, "Science of Spatial Learning," U.S.News, 17 May 2012.
- [34] M. Stieff, B. Dixon, B. Kumi and M. Hegarty, "Strategy Training Eliminates Sex Differences in Spatial Problem Solving in a STEM Domain.," *Journal of Educational Psychology*, vol. 106, no. 2, pp. 390-402, 2014.
- [35] M. J. Fortin, M. R. Dale and J. V. Hoef, "Spatial Analysis in Ecology," in *Encyclopedia of Environmetrics*, NJ, Wiley, pp. 2051-2058.

- [36] O. Maimon and L. Rokach, Data Mining and Knowledge Discovery Handbook, Springer, 2010.
- [37] J. M. B. Protázio, W. C. Pereira and F. Elayne Jesus de Castro, "Explicit Formulas for an Area Based Edge Effect Correction Method and their Application to Ripley's K-Function," *Journal of Vegetation Science*, vol. 10, no. 3, pp. 433-438, 1999.
- [38] L. Lyons, E. Minor, M. Zellner, B. Slattery and T. Shelley, "Parsing Patterns: Developing Metrics to Characterize Spatial Problem Solving Strategies within an Environmental Science Simulation," in 11th International Conference of the Learning Sciences, Boulder, 2014.
- [39] "spatstat: An R Package for Analyzing Spatial Point Patterns," Journal of Statistical Software, vol. 12, no. 6, pp. 1-42, 2005.
- [40] T. Wiegand, F. He and S. P. Hubbell, "A systematic comparison of summary characteristics for quantifying point patterns in ecology," *Ecography*, vol. 36, no. 1, pp. 92-103, 2013.
- [41] M. Kutner, C. Nachtsheim and J. Neter, Applied Linear Regression Models, McGraw Hill, 2004.
- [42] M. Hazewinkel, Encyclopedia of Mathemetics, Springer, 1994.
- [43] H. Akaike, "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716-723, 1974.
- [44] L. Vygotsky, Mind in Society: The development of higher mental processes, Cambridge, MA: Harvard University Press, 1978.

## VITA

NAME:	Aditi Krishna Mallavarapu
EDUCATION:	B.E., Computer Engineering, University of Pune, India,2011
	M.S., Computer Science, University of Illinois at Chicago, 2014
EXPERIENCE:	Programmer Analyst Trainee, Cognizant Technology Solutions, India
	2012
	Network Software Intern, Tarana Wireless Inc., Santa Clara, CA, 2013