Paper 2:

Student Ideas about Computational Thinking Concepts When Learning About Modeling Hydrologic Systems

Kristin L. Gunckel, University of Arizona
Judith A. Cooper, University of Arizona
Daniel L. Moreno, University of Arizona
Beth A. Covitt, University of Montana
Garrett Love, North Carolina School of Science & Mathematics
Randall Boone, Colorado State University
Alan Berkowitz, Cary Institute of Ecosystem Studies
John C. Moore, Colorado State University

Presented at the Annual International Meeting of NARST
March, 2018
Atlanta, Georgia

This material is based upon work supported by the National Science Foundation DRL – 1543228 Comp Hydro: Integrating data computation and visualization to build model-based water literacy. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Student Ideas about Computational Thinking Concepts When Learning About Modeling Hydrologic Systems

Integrating computational thinking into science instruction is a relatively new focus in science education. Computational thinking is listed in the *Framework for K-12 Science Education* (National Research Council, 2012) and the Next Generation Science Standards (NGSS Lead States, 2013) as one of the eight scientific practices that students should participate in while learning science. Yet, defining what computational thinking is, identifying what students should learn about it, providing examples of what it looks like in a science curriculum, and understanding how students think about and engage in computational thinking practices are all new territory. In this paper we lay out our framework for integrating computational thinking into instruction about water in environmental systems and present some data on student ideas about computational thinking concepts.

Computational Thinking and Modeling Groundwater Systems

At its core, computational thinking is an analytical approach useful for understanding and solving problems (Grover & Pea, 2018; National Research Council, 2010; Sengupta, Kinnebrew, Basu, Biswas, & Clark, 2013). It involves dealing with abstractions on multiple levels to be able to define problems, see essential relationships, and search for solutions (Wing, 2006, 2011). Computational thinking is deeper than just knowing how to use computers and broader than just knowing how to code. While using computers and coding may fall under the umbrella of computational thinking, the umbrella itself is more of way of thinking that provides an avenue for using computers and coding, among other things, to understand and solve complex problems (Grover & Pea, 2018).

Computational thinking is deeply intertwined with the practice of modeling complex systems (National Research Council, 2012; Sengupta et al., 2013). Modeling is a process of abstracting, decomposing, or simplifying a problem in order to develop explanations and predictions of phenomena (Schwarz et al., 2009). All phenomena are embedded in connected systems that operate at multiple scales according to scientific principles (Goldstone & Wilensky, 2008; Kali, Orion, & Eylon, 2003). Modeling systems relies on computational thinking to produce generalizable abstract representations, taking into consideration the foundational scientific principles relevant to the system. Producing and using the model requires understanding and using computational thinking concepts to engage in computational thinking practices, such as developing clear boundaries, assigning relevant parameters, recognizing a hierarchical organization that crosses spatial and temporal scales, identifying patterns, and testing and validating models (Ben-Zvi Assaraf & Orion, 2005; National Research Council, 2012). Thus, computational models can help explain phenomena, but understanding models also requires understanding how computation is part of the model itself (Wilensky & Reisman, 2006; Wofford, 2009).

Comp Hydro focuses on modeling the flow of water through environmental systems. The basic scientific principle is that water flows from high potential energy to low potential energy along pathways of highest hydraulic conductivity. In watersheds, this principle translates to water flowing in the direction of slope of the topographic surface. Similarly, in unconfined aquifers, water roughly flows in the direction of the slope of the water table. In aquifers, hydraulic conductivity is also related to the permeability of the substrate through which the water moves, influencing the rate at which the water flows. Modeling the flow of water through systems requires using two important computational thinking practices: discretization and parameterization. Discretization is the process of dividing a problem space (e.g., topographic surface or stratigraphy of an aquifer) into discrete, equal-sized chunks that can be represented and analyzed in a computer model. For example, contour lines of elevation or contaminant concentration are a way of discretizing a three-dimensional surface. Continuous space, such as an area underground, can also be divided into equal-area cells. Each cell is then assigned relevant parameters based on the scientific principles involved. When assembled, the chunks produce a pixelated, or rasterized image. In modeling the flow of groundwater, for example, two important parameters for each discretized cell are potential energy and permeability. Understanding discretization requires understanding the purpose of dividing an area into distinct units and the advantages and disadvantages of using different size units for the process. For example, larger sized cells require fewer data points, resulting in a less detailed, or more pixelated representation. Determining the average or modal values for parameters for larger cells may be more difficult than for smaller cells.

We were interested in designing instructional sequences that support students in learning and using computational thinking concepts while modeling water flow through watersheds and aquifers. We also wanted to use these instructional sequences to gain some insights into how students make sense of these computational thinking concepts. Our research questions were:

1. How can instructional sequences be designed to support students in learning and using computational thinking such as discretization and parameterization to model water flowing through environmental systems?
2. How do students make sense of discretization and parameterization when modeling water in environmental systems?

Computational Thinking in Instruction about Water in Environmental Systems

The common approach to integrating computational thinking into the curriculum is to engage students in coding models of systems, using programs such as Scratch or Net Logo (Lye & Koh, 2014). In Comp Hydro, we took a different approach. Our goal was to engage students in learning to understand how computer models trace water as a way to make the link between a computational concept, such as discretization or parameterization, and how that concept actually plays out when using a computer to model a complex system. To this end, we wanted to create experiences that would help students think about what happens within a computer model to solve a problem, such as modeling the pathway of water through a system. Rather than asking students to think like a computer, we tried to create experiences that would put students inside a computer model, figuratively speaking, to understand how computer models discretize space and use values for assigned parameters in an algorithm to model a pathway. We reasoned that understanding how computer models use hydrologic principles to trace water flow, such as water flows from high potential energy to low potential energy, could help students understand the hydrologic principle. Furthermore, we reasoned that being able to understand, use, evaluate, and potentially create models requires that students understand how computer programs use computational concepts to generate models. For example, when interpreting a computer-generated map of a groundwater contamination plume, students might ask where the data points used to produce the map are located in order to assign some confidence in the plume boundaries depicted by the model output.

To achieve these goals, we wrote four curriculum units, each focusing on either a groundwater or surface water issue of importance and relevance to the students in the schools using the Comp Hydro instructional materials. In the Arizona version of Comp Hydro, students studied a trichloroethelyne (TCE) and 1,4 dioxane groundwater contamination plume caused by historic use of these chemicals to clean airplane parts at the nearby airport in the 1950s (Tillman, 2009). Similarly, in the Montana version, students studied a groundwater heavy metal contamination plume emanating from a lead and zinc smelter (Burns & Marcussen, 2016). The Maryland version focused on surface water flooding issues related to urban runoff (Smith & Smith, 2015) and the Colorado version examined how surface water resources are partitioned among various users to analyze the impacts of a proposed water supply reservoir (Duggan, 2016). We hoped that students studying Comp Hydro in each of these four sites would find the local water problem engaging and motivate them to learn more about water, how it moves through their communities, and how computers and computer models could help them understand and evaluate locally relevant water problems (Bennett, Lubben, & Hogarth, 2007).

Each of the four curriculum units used a variety of models and types of models to support students in making connections among physical phenomena, two-dimensional representations, and computational models, and for making connections between hydrologic phenomena and computational thinking concepts. For example, the water table, defined as the elevation at which pore spaces underground become saturated, is an abstract construct of a physical phenomenon that is hidden from view. Visualizing an abstract, hidden, spatial construct is especially challenging to students (Kali & Orion, 1996; Piburn et al., 2005). Furthermore, maps of water tables are two-dimensional representations of three-dimensional surfaces. Moving back and forth between the three-dimensional conceptualization of the water table and a two-dimensional contour map representation of the water table is a challenge for many students and limits their progress towards developing a model-based understanding of groundwater (Gunckel, Covitt, Salinas, & Anderson, 2012). Interpreting, evaluating, and eventually producing contoured models of the water table requires understanding of computational thinking concepts such as interpolation and discretization of data. We designed Comp Hydro to support students in making all of these connections.

To help students understand and visualize the water table and learn interpolation and discretization concepts, the groundwater versions of Comp Hydro (i.e., Montana and Arizona), involved students using string or electric tape to measure the water level in an above-ground array of vertically-oriented PVC pipes that represented wells. Students marked the water level on the outside of the pipes with blue tape and connected the pipes with blue yarn to visualize the water table as a surface (Covitt, Podrasky, Fassnacht, Paquette, & Woessner, 2018). In the next lesson, students translated their data of the water levels measured in the PVC pipes into elevation data on a map and then contoured the water table. In this process, students had to grapple with interpolation between data points to draw the contours and discretization to decide the contour interval. In a subsequent lesson, students used a Net Logo application (Wilensky & Reisman, 2006) to contour the same data. This process was designed to help the students understand how the computer had to interpolate and discretize the data, just like they had done when contouring their maps, to produce computer-generated representations that looked similar to the maps they had drawn. This instructional sequence was also intended to help the students learn to interpret contour maps as representations of three-dimensional surfaces and visualize the water table as a three-dimensional surface underground. The watershed versions of Comp Hydro had a similar sequence for visualizing and contouring spatial rainfall data.

All four versions of Comp Hydro had a similar instructional sequence for modeling the flow of water through either aquifers or watersheds. In the groundwater sequences, students first observed the flow of dye through a three-dimensional groundwater modeling tank filled with layers of sand and gravel. Students also used permeameters to test the permeability of sand, gravel, and a mixture of sand and gravel. These models gave students physical experiences from which to visualize groundwater flow and to make sense of the hydrologic principles that govern water flow (i.e., potential energy and permeability). In the next lesson, students experienced how a computer program traced the flow of water through the same groundwater tank. A two-dimensional grid representing the problem space of an aquifer was laid out on a large piece of paper. Each cell in the grid was assigned values for potential energy and permeability parameters. Students took turns moving markers representing water from cell to cell according to the hydrologic principles they had learned in the previous lesson. For example, students had to use the potential energy values assigned the cell they were in to determine the next adjacent cell to move to each time the teacher called out the next iterative turn of the model. Similarly, students had to use the permeability values assigned to the cell they were in to determine how many cells they could move during each turn. Cells representing sand told students to move only two cells per turn while cells with higher permeability values told students to move three cells per turn. In this way, students applied hydrologic principles using the assigned parameters to follow the algorithmic rules of the computer model to trace water through the aquifer. The follow-up lesson engaged students in using and manipulating a Net Logo model of the same groundwater scenario to try to clean up groundwater contamination. A similar sequence was used in the watershed units. This sequence was designed to engage students in making connections across physical and computational models, help students actualize the principles of groundwater flow, and support students in making sense of parameterization as an important computational thinking concept in modeling.

Methods for Exploring Student Computational Thinking

Context

Comp Hydro took place at four sites across the United States (Arizona, Colorado, Maryland, Montana). We developed four instructional units contextualized in local water issues as described above: two focused on groundwater flow and two focused on watersheds. Teams writing the instructional units included science education researchers, hydrologists, computer scientists, modelers, and teachers. Each unit included Net Logo models to illustrate the computational thinking concepts and hydrologic principles incorporated into the units. Units included teacher guides, student materials, necessary powerpoint decks, Net Logo models, and the physical lab materials necessary to conduct all lessons.

Lead researchers at each site conducted professional development with four to eight high school teachers to introduce them to the Comp Hydro instructional units. Teachers involved taught Earth science, integrated science, environmental science, honors biology, computer science, and engineering. Professional developments activities at each site ranged from four days to two weeks, depending on the school district constraints at each site, and introduced teachers to the hydrologic and computational thinking concepts, the instructional activities, the Net Logo models, and the assessments.

Teachers at each site enacted the Comp Hydro lessons in their classrooms. The units were designed to be about three weeks of instructional time. Comp Hydro research staff were usually available, either in person or via email, to support teachers in teaching the lessons.

Participants

Comp Hydro sites spanned a diverse range of contexts, from predominantly rural/small town school with majority white populations (Montana) to suburuban schools with majority white populations (Colorado), to urban schools with minority-majority populations (Arizona and Maryland). The data for this paper were from high school students in Arizona.

Data Collection

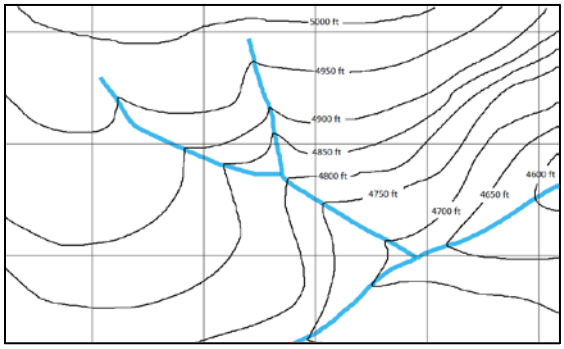
To assess student thinking we used both written assessments and interviews.

Written Assessments. Computer-based written assessments were designed to assess student understanding and use of both computational thinking concepts and hydrologic principles. Most items were situated in two water scenarios similar to the scenarios students studied during the Comp Hydro lessons. One scenario involved reading maps to trace water through a watershed. The other scenario involved interpreting maps and cross-sections of a groundwater contamination situation. Both scenarios were based on actual water issues.

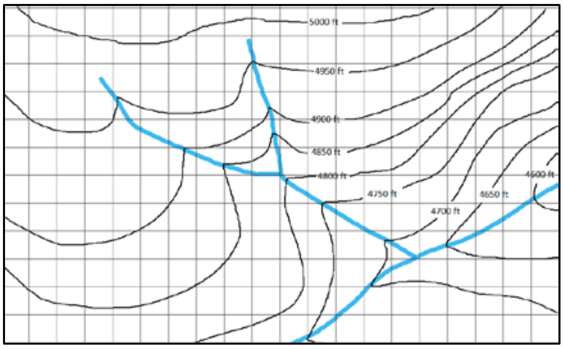
For this paper we analyzed student responses to three items designed to elicit student responses related to discretization of watersheds and two items designed to elicit student responses related to parameterization of a model of aquifer flow.

Discretization Items

The diagram below shows two different grids to divide the map into cells to develop a computer model of water flow. Use this diagram for the questions below.



Grid A



Grid B

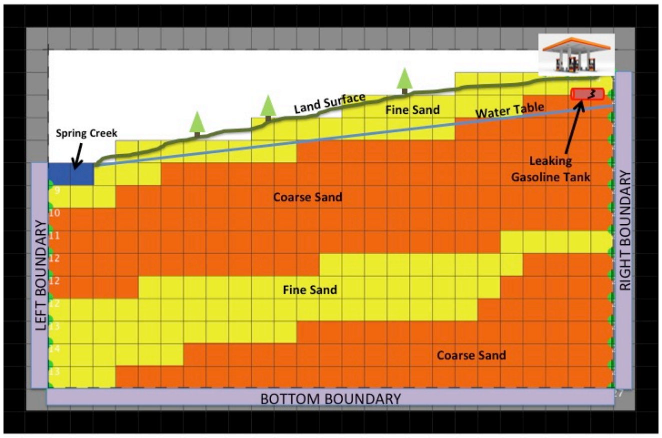
Discretization Prompt 1. What is the purpose of dividing the area into cells?

Discretization Prompt 2. Give at least one advantage of using Grid B (smaller cells) for your computer model.

Discretization Prompt 3. Give at least one disadvantage of using Grid B (smaller cells) for your computer model.

Parameterization Items

The image below shows a cross-section of the area where the underground gasoline tank is leaking. A grid has been applied over the cross-section to begin making a computer model of the gasoline spill.



Parameterization Prompt 1: What information about each cell in the grid would be needed to compute and predict the flow of water and MTBE through the system?

Parameterization Prompt 2: Please explain why each type of information (parameter) you listed is important.

Interviews. We conducted semi-structured interviews with a small set of students at each site (2 students per teacher) to better understand how they were thinking about the assessment prompts. Students were asked the same questions that they answered on the written assessment. In some cases, we were able to provide students with copies of the answers they wrote on the assessments. We then asked students to elaborate on their answers. We used follow-up questions to probe their thinking in order to better understand how they were answering the assessment items. Interviews were audio and video recorded. Artifacts from the interviews were scanned and saved. All interviews were transcribed.

Analysis. To analyze the data for this paper, we began with the interviews from 16 students at one urban Comp Hydro site (Arizona). These students completed the Comp Hydro instructional activities in either 9th-grade integrated science (9 students), 9th-grade honors biology (2 students), or upper-division (10th-12th grade) Earth Science (5 students). We focused on the interview questions that aligned with the discretization and parameterization assessment prompts.

We used a grounded theory approach (Strauss & Corbin, 1994) to identify categories of students thinking. Two researchers went through an iterative process of reading each student's answers and comparing them to other students' answers, looking for similarities in how the students were responding to the questions. Through this process we constructed a set of initial categories and identified indicators for each category.

We then moved to the assessment data from the same site. We randomly sampled 50 responses to the discretization and parameterization items from a total of 588 student tests. We looked for statements in these responses that were similar to the categories identified from the initial interview analysis. In this process we identified many more types of responses than were evident in the interviews. Through an iterative process of moving back and forth between the assessment responses and interview responses, we grouped multiple indicators into four categories.

At this point we returned to the assessment data and randomly sampled another 50 student responses. Two researchers independently coded the responses to the discretization and parameterization items, discussed their differences, and refined the indicators. They repeated this process two more times until they had greater than 80% interrater reliability.

Findings

From the student responses to the interview and assessment prompts, we identified four broad categories representing qualitatively distinct ways of making sense of and using discretization and parameterization concepts. The categories represent increasingly more sophisticated responses from answers that merely repeat the prompt to responses that connect computational thinking concepts to hydrologic principles. Tables 1 and 2 shows examples of assessment responses in each category. Below, we refer to this table and add examples from interviews to describe the categories.

Literal Construers

Responses in this category restated phrases in the prompt, referred to concepts unrelated to the hydrologic context of the prompt, or were ambiguous to the point of not being able to interpret deeper meanings.

Discretization. The discretization prompt showed students two different grids on a map and stated that the grids were used to divide the map into cells in order to develop a computer model of water flow. The prompt then asked students the purpose of dividing the map into cells. In designing the prompt we felt it was important to provide some context for the grids but wanted to know more about how students were thinking about how one would use the grids to translate the maps into a computer model. Literal construer responses often stated that the cells are used to develop a computer model, essentially restating the prompt and providing no further elaboration (Table 1). When describing advantages and disadvantages of using the grid with the smaller cells, responses in this category often talked about how the smaller cells made it easier or possible to model water flow, without explaining why, again essentially restating the prompt. Other responses made ambiguous references to the size or number of cells in the smaller grids, essentially restating what they saw in the grid. Literal construers also sometimes responded to the prompt by listing the biological functions of cells, such as reproduction, growth, or maintenance. These responses suggest that the students were bringing a biological, rather than computational lens to interpreting the prompts.

Parameterization. The parameterization prompt asked what information about each cell the computer would need in order to model the flow of water and gasoline underground and why that information is important. Similar to the examples from the discretization prompt, typical responses in this category restated portions of the prompt without providing elaboration or additional explanation (Table 2). Often these responses seemed to be circular, stating that to model the flow of water, the computer needs to know the flow of the water.

Another type of response that we lumped into this category were answers that focused on the effects of the gasoline spill and saw no connection to modeling. For example, one response stated, “It really wouldn’t help why because the gasoline spill will leak everywhere.” The response suggests that modeling the spill is not going to be useful because spills are unpredictable. We grouped these responses in the literal construer category because they seemed to be focused on the literal concern of the gasoline spill and not on modeling or using computers to define or solve problems.

Overall, the literal construers category included the least sophisticated responses. The responses suggested that the students were making sense of the prompts at the most literal level and sometimes did not recognize a connection between the phenomenon of the spill and the process of modeling the flow of water in an aquifer. We note that the responses in the literal construer category came only from the written assessments and were not present in the interviews. It could be that the short-answer nature of the assessment prompts on the assessment did not provoke students to think more deeply about what the items were asking.

Model Describers

Many of the responses to the prompts described what the models would show. Unlike literal construers, these responses suggested a connection between the model and the phenomenon, often describing what one would see in the model.

Discretization. On the discretization prompt, students often talked about how the grids were helpful for seeing something on the map. An interview with Collette provided an exemplar. The interviewer began by reading the assessment prompt.

143. INTERVIEWER: A scientist wants to make a computer model of the flow of water on this land. And, the first thing they have to do is they have to divide the map into cells. So they put this grid on top. And, what would be the purpose of putting a grid on top of the map to make a computer model?
144. COLLETTE: Um—a grid?
145. INTERVIEWER: Mm hmm. The grid just means this—um—dividing it up into these squares.
146. COLLETTE: Um—I don’t know, I guess to see each section?
147. INTERVIEWER: So say that again?
148. COLLETTE: To see each section of the map—
149. INTERVIEWER: Okay.
150. COLLETTE: —in little parts.
151. INTERVIEWER: And why would they want to do that?
152. COLLETTE: Mm—(pause) for better information?

In line 148, Collette said that the purpose of dividing the map into cells was “to see each section of the map in little parts. She further elaborated in line 152 that the cells would provide information, although she does not say what information one would get from each cell. Her statement suggests that to her the grid highlighted information, possibly the elevations or maybe just the location of the rivers, provided on the map.

The interviewer then went on to ask Collette about the advantages and disadvantages of the two cells sizes for producing a three-dimensional models of the map topography.

173. INTERVIEWER: Okay. So, thinking about that, how does it help you think about why, um—what would be an advantage of choosing one with the smaller squares—Grid B—if you were going to put that into the computer model?

174. COLLETTE: Um—I guess with the smaller squares, you could see more, and it gives you more points.
175. INTERVIEWER: Okay. And then, what would be the disadvantage of it, though?
176. COLLETTE: Mm—just like maybe it's a little too much?

In line 174, Collette again focused on what one could see when looking at the model. She described that smaller cells provide more points, possibly for drawing the picture, but also explained that having more points might provide too much information. Across these excerpts, Collette's responses focused on describing the map with the overlaid cells and describing the sizes of the cells.

Themes related to seeing, showing, for focusing on information on the map also emerged from the assessment data (Table 1). These responses also state that the grids, especially the smaller grids, provide more detail in what can be seen. Similarly, responses to the advantages and disadvantages prompts talked about smaller grids allowing one to see in greater detail or seeing smaller or larger areas.

Parameterization. On the parameterization assessment and interview prompts, responses in the model describers category described the pathway and direction of the water flow through the cross-section in the prompt. For example, Abigail responded this way:

153. INTERVIEWER: Look at the grid here, all of these squares marked on this diagram. What information would you need to tell each cell in this grid so that way it would know which way the water would flow?
154. ABIGAIL: Probably the finished information we'd need is where the contamination first started. [00:25:54]
155. INTERVIEWER: Okay.
156. ABIGAIL: So – because it started right here we need to know what direction it would go, so probably test if it's going down or to the side.

Abigail was thinking about the pathway that the water and/or gasoline would take through the aquifer. She was thinking that in order to trace contamination, one would need to know where the contamination started (line 154), which was shown on the cross-section provided in the prompt. In line 156 she was imagining different pathways the water and contamination could possibly take. Rather than thinking about how a model could use information such as potential energy or permeability to create the pathway, she was describing the overall pathway that the water would take. In other words, she was not thinking about the individual cells or how a computer would model the flow; she was tracing the flow of water herself.

Table 2 shows example responses to the parameterization assessment prompts. Like Abigail, these responses described the flow of water through the scenario shown in the diagram that accompanied the prompt. They often gave detailed descriptions of the pathway the water and/or gasoline would take. These responses seemed to be interpreting the prompt as asking how the water would flow through the aquifer, rather than how a computer would trace the pathway.

Responses in the model describers category are about the "what" of computer models. They essentially describe what the models show. These responses suggest that students understand some principles of hydrologic flow in order to be able to interpret the cross-section and trace the water flow themselves. However, because the responses in this category were about the water flow only and not the parameters used in a computer model or the cells shown on the diagram, we viewed the responses in this category as less sophisticated than the categories that follow.

Model Users

Some students described how they would use the cells on the maps or cross-sections to interpret the maps or trace the flow of water.

Discretization. Below is an excerpt of an interview transcript with Mario. To understand this excerpt, it is important to know that while the interviewer asked Mario about the grids on the topo map, Mario answered in the context of an activity that he did in class that involved moving a marker representing TCE contamination in groundwater through a cross-section representation of the groundwater system marked with a grid. Each cell contained information about the relative potential energy and the hydraulic conductivity of the cell. He referred to this representation as the "table model" because the class was modeling the flow of groundwater on the table rather than in a computer model.

86. INTERVIEWER: Good, you answered my next question. (chuckle) (pause) Let's go on to this one. This one, we've got that same map, right?
87. MARIO: Mm-mm.
88. INTERVIEWER: But we put two grids on it, and we did that because we're wanting to divide the area up into these different cells, so that we can a computer model, kind of like we did in Comp Hydro. What was the purpose of putting a grid on top of the map?
89. MARIO: Making it easier to make the cells for the model.
90. INTERVIEWER: Why do we have cells in the model?
91. MARIO: To show the way the water would flow easier; that's the way I looked at it when we looked at the table model, because there was less cells at the top than at the bottom. At the bottom, it was perfectly like a diagonal –
92. INTERVIEWER: Here, you can draw on this (paper noise).
[Mario draws a rectangle with a grid]
93. MARIO: Since it was a rectangle, and there was a bunch of boxes. Pretty much the best way the TCE flowed in this chart was this way. They came down, went all the way to the left, and then it popped back up and it ended right there. But it still had a different ways to flow, but mainly the most cells were at the bottom of those three cells. That's pretty much how it flowed. It shows the way the water flows easier.
94. INTERVIEWER: What was inside each of those cells?
95. MARIO: The energy; was it potential energy of the cells?
96. INTERVIEWER: Mm-mm.
97. MARIO: The way it flowed. (whispering) What is it? There was another one that was in the box; I just can't remember (overlapping voices)

98. INTERVIEWER: Why were those things in the box?
99. MARIO: The potential energy?
100. INTERVIEWER: Yeah.
101. MARIO: I don’t know, actually. I didn’t know much about the potential energy, but I understood the cells part. And the reason why, because when he said move down to a different one, but it has to be less. Or the potential energy has to be less, if that makes sense. I don’t know how to explain it.

In lines 91 and 93, Mario described the flow of water through the cross-section. In this respect, because there was no flow of water marked on the representation he was describing, his response was similar to the responses in the model describer category. In line 93 he also refers specifically to the cells. He talks about how there are fewer cells on the top and the bottom, and by this we think he actually means there are fewer pathways through the cells at the top of the diagram. As such he is still describing the pathways of the water. But in line 95 he began to refer to the cells as containers of potential energy. Although there is no evidence that he understood what potential energy is or why it is important for being able to trace groundwater and contaminants through an aquifer system, he did recognize that each cell divided the field into discrete sections that were then assigned rules for how to trace the groundwater and TCE. Mario’s response suggest a procedural understanding of models as useful for tracing water flow and cells as useful components of the models. Granting that Mario did not answer the question about the purpose of the cells for digitizing the topographic watershed map, his response did reflect the notion that cells are useful for using models to solve a problem, in the case that he described, the flow of water through a system.

Analysis of assessment responses provided additional examples of responses that described cells as being useful (Table 1). Often, responses indicated that the cells would be useful for finding locations on a map or cross-section, or making comparisons between points on the cross-section. When discussing advantages and disadvantages of different size cells, responses in this category sometimes referred to the how more cells could be useful for solving the problem more easily or in a better way but could make the process of using the cells more difficult or confusing. Unlike the model describer category, all of these answers in the model users category specifically referenced the cells and indicated that the cells were somehow useful for solving a problem.

Parameterization. Mario’s transcript above also provides an example of parameterization responses in this category. Mario did identify a relevant parameter, potential energy (line 95 and line 99). In line 101 he also referred to how he used the potential energy, stating that he moved his marker to a cell with lower potential energy. Because there is no evidence that Mario understood why this is a useful rule for modeling water flow, we cannot assign this response to a more sophisticated category. However, he did recognize that computer modeling involves assigning values to parameters that are used according to specific rules to trace water flow.

Assessment responses showed similar characteristics to Mario’s comments (Table 2). Responses identified information that the students thought was important in order to model the flow of water. Sometimes this information was relevant to modeling groundwater flow, such as potential energy or permeability, and sometimes it was information relevant to the problem, such as the location of wells, but not necessary to trace water flow. The reasoning about the importance or relevance of these parameters often referenced the idea that the computer needed this information without identifying the relevant hydrologic principle or how the computer model would use the information to trace water flow.

Responses in the model users category are more sophisticated that model describers because they were able to identify that certain, specific information is necessary to model the flow of water, they recognized the general function of discretizing the problem space, and they saw modeling as useful for solving problems. Essentially, the model users explain how to use computer models to trace water flow. The responses often indicate that students' understanding of the parameters they identify may be weak and they do not always identify the most relevant parameters. However, unlike model describers, they recognize the function for discretization and parameterization for helping to solve a problem.

Model Interpreters

Responses in the model interpreters category connected hydrologic principles governing the flow of water through watersheds and aquifers to computational thinking concepts. We called this category model interpreters because many of the answers indicated an ability to evaluate model input and/or output based on hydrologic and computational thinking concepts.

Discretization. Brisa's interview provided an exemplar of the modelers for the discretization category. Somewhat like Mario, Brisa answered this question in the context of an activity she had done in class. She described how she used cells in a Net Logo activity to make a three-dimensional model of a two-dimensional map. In the activity the topo map the students used was color coded by elevation. Students had to assign an elevation to each cell superimposed on the map based on the overall color of the cell. They then entered these elevations into the model to produce the three-dimensional view of the map surface.

139. INTERVIEWER: The first thing the scientist does is divide the map into cells. What's the purpose for dividing the map into cells?
140. BRISA: Well, when they divide it into cells it's kind of grouping that cell under one specific set. So like when we did this we had to group it into colors based on the elevation, so what the majority of the color was inside. So like if this one… like let's say one of these grids had like mostly red, then that would resemble something, and we'd put that whole grid under red.
141. INTERVIEWER: Uh huh. [0:14:00]
142. BRISA: So that's kind of like an easier way for the scientists to go through and get a good estimate of what it was in that area.

Brisa began by describing how she would use the grid over the map to enter data from the map into a model (line 140). In line 142 she went beyond just describing the procedure to explain that this procedure was a way to "get a good estimate of what it was in that area." This statement suggests that she recognized that the cells were useful for dividing the area of the map into discrete chunks and assigning a value of to the cells to represent an average elevation for the area covered by that cell. This understanding was by far the most sophisticated understanding of discretization and most in line with what we had hoped that students would learn from the instructional activities.

Additional examples from the assessment data show understanding of using the cells to chunk continuous data (Table 1). In this category, responses to the prompt referred to elevation as information that would be used in the computer model to trace direction of water flow. Responses to the advantages and disadvantages prompts usually gave some form of the general principle that more data provides greater accuracy and detail but also requires more work on either the part of the computer or the person entering the data into the computer. There were a few responses that also recognized that computers interpolate data, which has advantages and disadvantages.

Parameterization. For the parameterization prompt, responses in the model interpreters category identified a relevant hydrologic parameter and were able to explain the hydrologic principle that would be used to trace water. Again, Brisa provides the best example.

267. INTERVIEWER: So what do you have to do in order to tell a computer how to trace the flow of contamination?

268. BRISA: What can you do is you would need to know the energy level of all of the grids. Because as I mentioned earlier, the water, it won't be able to go from lower… a lower energy level to a higher energy level. So it would need to know what the energy level of each section is to determine the possible ways that the water could have gone.

Here, Brisa identified one of the relevant parameters (potential energy) and explained the hydrologic principle for how this parameter is used to trace water flow. In line 268 we interpret that when she referred to "it," she was referring to the computer model using this information to trace the water flow. Assessment responses were similar (Table 2). They identified at least one relevant parameter and provided the hydrologic principle used to model the flow of water. Due to the shortness of the responses, however, we sometimes had to infer that students were describing how the computer would use the information, since that aspect was given in the prompt.

Model interpreters are not necessarily model makers. The goal of our instructional sequence was not to have students be able to make a computer model. However, we did want them to have some understanding of how hydrologic principles and computational thinking concepts are used to model water flow. We see the model interpreters category as including responses that provide some indication of why computer models provide the pathway that they do in the model output. Responses in this category were the most sophisticated of the responses and indicated that students were making connections to be able to think about why models work and to potentially be able to evaluate both the input and output of computational models.

Table 1: Example Student Responses to the Discretization Prompts
Prompt Stem: The diagram below shows two different grids to divide the map into cells to develop a computer model of water flow.

Category	**Discretization Prompt 1: What is the purpose of dividing the area into cells?**	**Interpretation of Prompt 1**	**Discretization Prompt 2: Give at least one advantage of using the smaller grid for your computer model.**	**Discretization Prompt 3: Give at least one disadvantage of using the smaller grid for your computer model**	**Interpretation of Prompts 2 & 3**
Literal Construers	To develop a computer model of water flow	Restates that cells are used to make a computer model.	makes it easier to develop in a computer model of water flow	You would have to have that many cells to make a model	Restates that cells are used to make a computer model.
	To make a computer model.	Restates that cells are used to make a computer model.	Mapping it	Model the graph	Restates that cells are used to make a computer model.
	cell division is a step in reproduction and is necessary for growth and maintenance.	Refers to cell biology.	there smaller	more and bigger	Ambiguous reference to size of cells
Model Describers	The cells marks off certain areas	Grids make areas visible for study.	The smaller cells show a much better scale of certain areas.	The cells might show too small amount of an area.	Grids show information
	To see the over all look and to see much closer.	Grids make areas visible	We are able to see more detail into the grid.	The grid doesn't cover as much data.	Grids show information; focus on detail.
	The purpose of dividing the areas into cells is to get a better look, or get better information from them	Grids make areas visible; help show information.	You can see a lot of the map in greater detail in the smaller squares.	It would probably take way too long to look at every single piece of the grid, and personally, i would rather look at the whole map rather than tiny little squares in a grid.	Focus on detail.
Model Users	make it easier to locate using grids	Cells are useful for locating information	Being able to find exact locations.	Having to use the middle of the cells if a location isn't exact.	Cells are useful for locating information
	each section might have a different water levels, contamination and in each place the water might be very different	Cells are useful for making comparisons; notes the problem that cells are useful for solving.	each section might [information] have different that can be very useful	that's alot of data you have to go through in grid b	Cells as containers of information; more cells makes the work more difficult.
	Dividing an area into cells makes it easier to speak about, it allows you to section off different areas and identify locations with coordinates.	Cells are useful for making comparisons and for locating points	Grid B allows for the location of different things to be more pinpoint as the grid squares are smaller.	Smaller grid squares could make the data very busy and possibly hard to read.	More cells produce more accuracy but can also be more difficult to work with.

Category	**Discretization Prompt 1: What is the purpose of dividing the area into cells?**	**Interpretation of Prompt 1**	**Discretization Prompt 2: Give at least one advantage of using the smaller grid for your computer model.**	**Discretization Prompt 3: Give at least one disadvantage of using the smaller grid for your computer model**	**Interpretation of Prompts 2 & 3**
Model Interpreters	The purpose of dividing the area into cells is to get a general idea of what the elevation must be in that cell	Shows understanding chunking continuous data.	You get a more accurate description and therefore better data.	The computer might fill in missing information in the blank areas.	Recognizes that data may not be available for all cells and computers often interpolate.
	The purpose is to what area is in which square and you will be able to enter it into the computer easier	Indicates understanding of chunking continuous data and entering it into a computer model.	Grid B would give the computer more input for a model that may be more reliable and or show more information on how water flows	takes the person and the computer longer to generate the model	More data provides greater accuracy but requires more work.
	This could help scientists see where the water would go by using grids to determine the elevation of the land.	Recognizes that the grid cells are useful for working with elevation data on the map.	The elevation would be more accurate instead of being a general broad area.	This could take a considerable more amount of time to create	More data provides greater accuracy but requires more work.

Table 2: Example Student Responses to the Parameterization Prompts
Prompt stem: The image below shows a stratigraphic cross-section of the area where the underground gasoline tank is leaking. A grid has been applied over the cross-section to begin making a computer model of the gasoline spill.

Category	**Parameterization Prompt 1: What information about each cell in the grid would the computer need to know in order to model the flow of water and MTBE through the groundwater?**	**Parameterization Prompt 2: Please explain why each type of information (parameter) you listed is important.**	**Interpretation**
Literal Construers	water flow	it shows where the gasoline leaked.	Restated that in order to model the flow of water, one would need to know the flow of water. Restated that the gasoline flows with the water.
	the flow of the water	they wont know which way the water will spread	Restated that in order to model the flow of water and gasoline, one would need to know the flow of water.
	It really wouldn't help why because the gasoline spill will leak everywhere.	Its important because it will be in separate areas	Suggests that spills are unpredictable and modeling may not be helpful.
Model Describers	the flow	water flow leads to the spring creek so with the gas leak it will go into the groundwater and makes its way to the grid because theirs the coarse sand that makes it hard to go through but helps slide off to the side of the flow	Describes the flow of water to Spring Creek through the course sand labeled in the cross-section but does not refer to the cells or the parameters.
	the leak is flowing through fine sand to get into the river	because it shows how the leak is flowing	Refers to what the model will show of the water flowing through the fine sand.
	the contamination is definitely going into the creek	due to the way the water table is slanted the contamination will surly lead into the creek	Interprets the direction of the water through the aquifer based on the cross-section, but does not connect to computer modeling.
Model Users	it would need to know which area the permeability and it needs to go below the permeabilty , how many cells each one would need to move, so we need more information	So you know how many spaces to move to get the the destination and you would	Identifies relevant parameters and indicates that they are useful to solve a problem but does not connect to the relevant hydrologic principle or explain how the computer model would use the information to trace water flow.
	The computer would need to know the amount of energy in each cell.	Because you won't be able to really model anything if you don't have enough data to do that.	Identifies relevant parameters and indicates that they are useful to solve a problem but does not connect to the relevant hydrologic principle or explain how the computer model would use the information to trace water flow.
	The computer would need to know where the contamination started and where it is spreading and the elevation	It would need to know these things order to know what way the contamination flows. The elevation and shows what way it flows towards.	Identifies a parameter, but not the relevant parameter for modeling. States that computer would use this information but doesn't explain how.

Category	Parameterization Prompt 1: What information about each cell in the grid would the computer need to know in order to model the flow of water and MTBE through the groundwater?	Parameterization Prompt 2: Please explain why each type of information (parameter) you listed is important.	Interpretation
Model Interpreters	The computer needs to know each cells permeability, energy, and maybe even elevation because each of these are needed to understand where the contamination will flow.	Permeability is important because the computer would need to know if water flows through that cell fast or slow and it also depends on what material each cell is. Energy is important because the computer needs to know how much energy it takes foe water to go through it. Elevation might help as well to know where the contamination should flow.	Identifies the relevant parameters and explains the hydrologic principles used.
	The computer would need to know the elevation of each cell grid n order to do that. It would also need to know what type of sediments it is going to have to go through.	First elevation is going to choose the direction of the flow of the MTBE. And also the sediment is going to see if the MTBE could actually go through it.	Identifies the relevant parameters and explains the hydrologic principles used.
	The computer would need to know the potential energy of each cell.	Knowing the potential energy of each cell is important because the contamination has to travel from high to low potential energy.	Identifies one relevant parameters and explains the hydrologic principles used.

Discussion

Comp Hydro provides an example of what integrating computational thinking into science instruction can look like. While there is a growing literature on computational thinking in the computer science journals (Lye & Koh, 2014), Comp Hydro explores how students use computational thinking when learning science.

Much of the work in defining computational thinking has been focused on high level of concepts like abstraction, logic, and pattern recognition (Grover & Pea, 2018; Wing, 2011). In our work we have focused on smaller grain-size concepts like discretization and parameterization. These concepts are the link to understanding and creating system models (Sengupta et al., 2013). Furthermore, they are necessary to be able to evaluate the output of computer models and assign confidence to the results that are produced. Comp Hydro then connects these concepts to the hydrologic concepts so that students not only understand how water moves through systems, but how computer models use these concepts to produce the visualizations that they see.

Our work also provides an approach to integrating computational thinking concepts to scientific principles, in our case, hydrologic principles. Other approaches so far have often engaged students in coding, which is definitely important (Lye & Koh, 2014). However, we argue that in order to code models of complex systems, one must understand the relevant hydrologic principles, the important computational thinking concepts, and how they are connected. Our approach situates student learning in actual water issues of importance and relevance to the students studying Comp Hydro (Bennett et al., 2007). We use a range of models, from physical models to maps and cross-sections, to Net Logo computer models, to help students not only visualize the phenomenon of water flow through watersheds and aquifers and the hidden aspects of these systems such as water tables, but also break open the black box of how computer models produce these visualizations.

A third contribution to this work is our focus on student thinking about computational concepts and modeling of water in environmental systems. Much of the work in computational thinking has been top down, focusing on defining what expert computer scientists' know and therefore what students should learn. Our work starts from the bottom up, and begins to look closely at how students think when working with computational models of water systems. We have identified four categories of types of responses that are qualitatively distinct. Our categories are not about how correct these answers are. In fact, even the most sophisticated responses that we received were not perfect answers and might be downgraded if one were using them to assign a grade to student work. Our focus was on how students were thinking about the questions asked. The least sophisticated answers, literal construers, focused only on was given in the prompt. The other three categories, however, showed increasingly more sophisticated ways of making sense of the prompts given using computational concepts and hydrologic principles. Model describers responses used hydrologic principles to trace water, but do not yet indicate how computational concepts like discretization and parameterization are useful for solving problems. Model user responses, however, include indications that students were using the concepts to think about problems. Responses in this category may be somewhat proceduralized and there were missteps that indicate that students did not yet fully understand all aspects of the computational or hydrologic principles they used. Nevertheless, they recognized the computational models can be useful and they attempted to use them to solve problems. Finally, model interpreters were able to connect hydrologic principles and computational thinking concepts to translate hydrologic data into computational models of hydrologic systems. Like the model users, there were some places where students' understandings of either hydrologic or computational concepts were incomplete; nevertheless, their responses indicated that they viewed the connection between computational concepts and hydrologic concepts in a way that could be useful for creating and evaluating models. While these categories are not yet a learning progression, they do outline increasing levels of sophistication in the ways that students think about computational concepts and scientific principles.

There has been little work so far on assessing and measuring student computational thinking in science instruction. Some work has looked at different approaches of assessing student computational thinking during programming activities, such as analyzing student projects, engaging students in design scenarios, or conducting artifact-based interviews (Brennan & Resnick, 2012). Our assessment approach, using science scenario-based assessment items, focuses on students' use of computational thinking concepts while reasoning about scientific phenomena. Comp Hydro assessment items were accessible to students who have a range of experiences and ways of thinking about models, water systems, and computers and our framework for interpreting student responses provides criteria for interpreting student thinking.

A limitation of this work is that we have used only data from one site. We plan to analyze data from other sites soon. Also, so far, we have only been able to describe student thinking. Our next steps are to use our categories of student responses to code a large sample of student responses. The eventual goal is to be able to describe the distribution of student responses across these categories and use item response theory to test for a learning progression for computational thinking about water in environmental systems.

References

Ben-Zvi Assaraf, O., & Orion, N. (2005). Development of system thinking skills in the context of earth system education. *Journal of Research in Science Teaching, 42*(5), 518-560.

Bennett, J., Lubben, F., & Hogarth, S. (2007). Bringing science to life: A synthesis of the research evidence on the effects of context-based and STS approaches to science teaching. *Science Education, 91*(3), 347-370.

Brennan, K., & Resnick, M. (2012). *New frameworks for studying and assessing the development of computational thinking.* Paper presented at the Proceedings of the 2012 annual meeting of the American Educational Research Association, Vancouver, Canada.

Burns, B., & Marcussen, C. (2016). *Fourth five-year review report for East Helena Superfund site*. Denver, CO: United States Environmental Protection Agency.

Covitt, B., Podrasky, A., Fassnacht, D., Paquette, R., & Woessner, W. (2018). Delving below the sruface with the Montana Groundwater Academy. *Connected Science Learning, 1*(5). http://csl.nsta.org/2018/02/delving-below-the-surface/

Duggan, K. (2016). New NISP plan: Add to Poudre Flows in Fort Collins: Coloradoan.

Goldstone, R. L., & Wilensky, U. (2008). Promoting transfer by grounding complex systems principles. *Journal of the Learning Sciences, 17*(4), 465–516.

Grover, S., & Pea, R. (2018). Computational Thinking: A competency whose time has come. *Computer Science Education: Perspectives on Teaching and Learning in School*, 19.

Gunckel, K. L., Covitt, B. A., Salinas, I., & Anderson, C. W. (2012). A learning progression for water in socio-ecological systems. *Journal of Research in Science Teaching, 49*(7), 843-868. doi: 10.1002/tea.21024

Kali, Y., & Orion, N. (1996). Spatial abilities of high-school students in the perception of geologic structures. *Journal of Research in Science Teaching, 33*(4), 369-391.

Kali, Y., Orion, N., & Eylon, B.-S. (2003). Effect of knowledge integration activities on students' perception of the Earth's crust as a cyclic system. *Journal of Research in Science Teaching, 40*(6), 545-565.

Lye, S. Y., & Koh, J. H. L. (2014). Review on teaching and learning of computational thinking through programming: What is next for K-12? *Computers in Human Behavior, 41*, 51-61.

National Research Council. (2010). *Report of a workshop on the scope and nature of computational thinking*: National Academies Press.

National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas.* Washington, D. C.: National Academies Press.

NGSS Lead States. (2013). *Next Generation Science Standards: For States, By States*: Achieve, Inc. on behalf of the twenty-six states and partners that collaborated on the NGSS.

Piburn, M. D., Reynolds, S. J., McAuliffe, C. M., Leedy, D. E., Birk, J. P., & Johnson, J. K. (2005). The role of visualization in learning from computer-based images. *International Journal of Science Education, 27*(5), 513-527.

Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., . . . Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching, 46*(6), 632-654.

Sengupta, P., Kinnebrew, J. S., Basu, S., Biswas, G., & Clark, D. (2013). Integrating computational thinking with K-12 science education using agent-based computation: A theoretical framework. *Education and Information Technologies, 18*(2), 351-380.

Smith, B. K., & Smith, J. A. (2015). The flashiest watersheds in the contiguous United States. *Journal of Hydrometeorology, 16*(6), 2365-2381.

Strauss, A., & Corbin, J. (1994). Grounded theory methodology: An overview. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 273-285). Thousand Oaks, CA: Sage.

Tillman, F. D. (2009). *Results of the analyses for 1,4-dioxane of groundwater samples collected in the Tucson Airport Remediation Project area, south-central Arizona, 2006-2009*. U.S. Geological Survey Open-File Report 2009-1196.

Wilensky, U., & Reisman, K. (2006). Thinking like a wolf, a sheep, or a firefly: Learning biology through constructing and testing computational theories—an embodied modeling approach. *Cognition and Instruction, 24*(2), 171-209.

Wing, J. M. (2006). Computational thinking. *Communications of the ACM, 49*(3), 33-35.

Wing, J. M. (2011). *Research notebook: Computational thinking—What and why?* The Link Magazine, Carnegie Mellon University, (Spring). Pittsburgh.

Wofford, J. (2009). K-16 Computationally rich science education: A ten-year review of the Journal of Science Education and Technology (1998–2008). *Journal of Science Education and Technology, 18*(1), 29-36.