Title

MoDa: Designing a Tool to Interweave Computational Modeling with Real-world Data Analysis for Science Learning in Middle School

ISBN
9781450391979

Authors
Wagh, Aditi
Fuhrmann, Tamar
da Silva Eloy, Adelmo Antonio
et al.

Publication Date
2022-06-27

DOI
10.1145/3501712.3529723

Peer reviewed

MoDa: Designing a Tool to Interweave Computational Modeling with Real-world Data Analysis for Science Learning in Middle School

ADITI WAGH, Massachusetts Institute of Technology, USA

TAMAR FUHRMANN, Teachers College, Columbia University, USA

ENGIN BUMBACHER, Teacher University Lausanne, Switzerland

ADELMO ELOY, Teachers College, Columbia University, USA

JACOB WOLF, Teachers College, Columbia University, USA

PAULO BLIKSTEIN, Teachers College, Columbia University, USA

MICHELLE HODA WILKERSON, University of California, Berkeley, USA

Coordinating modeling and real-world data is central to building scientific theories. This paper examines how a complementary focus on modeling and data contributed to 8th grade students' learning of mechanisms underlying wildfire smoke spread in MoDa, a web-based environment that integrates computational modeling side-by-side with real-world data for comparison and validation. Epistemic network analysis of student responses in pre-post tests revealed a shift from primarily macro-level explanations to explanations that integrated macro and micro-level explanations of the phenomenon. Video data analysis revealed three design elements that contributed to student learning: Naming of the blocks, match between data and model visualization, and collective reflections on models. We reflect on implications for the design of environments that integrate computational modeling with real-world data analysis.

CCS Concepts: • **Applied computing → Education**; • **Human-centered computing → Human computer interaction (HCI)**; **Interaction design**.

Additional Key Words and Phrases: computational modeling, agent-based modeling, design, data practices, science education

ACM Reference Format:

Aditi Wagh, Tamar Fuhrmann, Engin Bumbacher, Adelmo Eloy, Jacob Wolf, Paulo Blikstein, and Michelle Hoda Wilkerson. 2022. MoDa: Designing a Tool to Interweave Computational Modeling with Real-world Data Analysis for Science Learning in Middle School. In *Interaction Design and Children (IDC '22), June 27–30, 2022, Braga, Portugal*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3501712.3529723

1 INTRODUCTION AND THEORETICAL BACKGROUND

In K-12 science education, modeling and data practices are seen as central to student learning [1]. There has been research on developing learning technologies to support these practices in classrooms. For instance, domain-specific modeling tools—in which the building blocks for a model are named and defined in the context of a specific phenomenon or domain—have been found to be readily interpretable for students [2, 9, 12, 18]. Research with these modeling tools have shown that they support mechanistic reasoning about phenomena [5, 16]. Similarly, students benefit from meaningful engagement with data in scientific inquiry activities, both in terms of science learning, and understanding of the nature of data [10]. General educational tools such as CODAP (Common Data Analysis Platform) help students manipulate, transform and visualize different types of data [3]. Other tools enable students to analyze fixed curated datasets (e.g., BGuILE [11]), or help collect and analyze real-world data from sensor-based measurement tools [15].

Few technologies facilitate an integrative approach by enabling students to coordinate between modeling and real-world data [4]. This is a missed opportunity, as integrating modeling and real-world data can be productive for science learning [6, 7]. Without tools that interweave real-world data analysis with model simulations, it is difficult to engage students in computational models and real-world data for theory building. This paper presents a pilot study of MoDa, a web-based environment for interweaving computational modeling and real-world data analysis for comparison and validation. This study was conducted as part of a larger project to investigate ways to make computational modeling a sustained practice in middle school classrooms. This paper examines how a unit enacted with MoDa supported student learning of mechanisms underlying the scientific phenomenon of wildfire smoke spread.

2 DESIGN OF THE *HOW WILDFIRE SMOKE SPREADS* MODULE WITH MODA

MoDa is a web-based environment that juxtaposes computational modeling side by side with real-world data for comparison and validation. The environment includes a modeling area and a real-world data area. The modeling area consists of (1) A *Coding workspace* where learners program their models using a library of domain-specific blocks. This is a custom block-based coding tool built on Google's Blockly library; (2) A *Simulation workspace* where students can run and manipulate the model they built in the coding workspace. Models are generated using the NetLogo engine [17]; and, (3) A *Model Data Visualization area* that displays plots for selected data generated by the model in real time.

A real-world data area is also being developed to support different data types and data sources.These can be qualitative, like video recordings or images, or quantitative, like data loaded from csv files or streamed live from digital sensors. Data sources can be physical experiments, quantitative measurements with sensor-equipped microcontrollers taken by students, or publicly-available quantitative datasets like (NASA's open-data platform). Students can examine the real-world data in comparison with the simulated data from their models, in real-time or asynchronously, to inform and evaluate their models. A prototype real-world data area was used for this study in which students were provided with a two relevant video clips, though they could not yet upload their own real-world data.

In the *How Wildfire Smoke Spreads* module, students compare computational agent-based models with satellite video data of wildfires to explore how interactions between micro-level entities of air and smoke particles impact the macro-level phenomenon of smoke spread. Two key ideas drive the module: (1) Diffusion and wind effects interact to create different patterns of smoke spread; and (2) Temperature gradients around a wildfire differently impact PM2.5 and PM10 particles, causing the lighter PM2.5 particles to travel further. This pilot study focuses on the first key idea.

2.1 Modeling: Building computational agent-based models

In the *How Wildfire Smoke Spreads* module (figure 1), students are presented with a bird's eye-view of a landscape. The block library includes three categories of blocks. Property blocks allow then to create air particles and smoke particles and set their properties such as speed or color. Action blocks model the behavior of particles: "move" which causes particles to move along a trajectory, "bounce off" which causes particles to bounce off each other when they collide, and "apply wind" which causes the particles to update their trajectories to follow direction and speed specified by corresponding wind direction and speed sliders. Control blocks can be used to specify behaviors for when a particle touches another particle. Some blocks are "unpackable" (differentiated by a plus sign) and as such serve as functions that aggregating other blocks into ready-made routines. One example is the "interact" block, which wraps together control structures with the "bounce off" block to create a behavior where particles bounce off one another when touching.

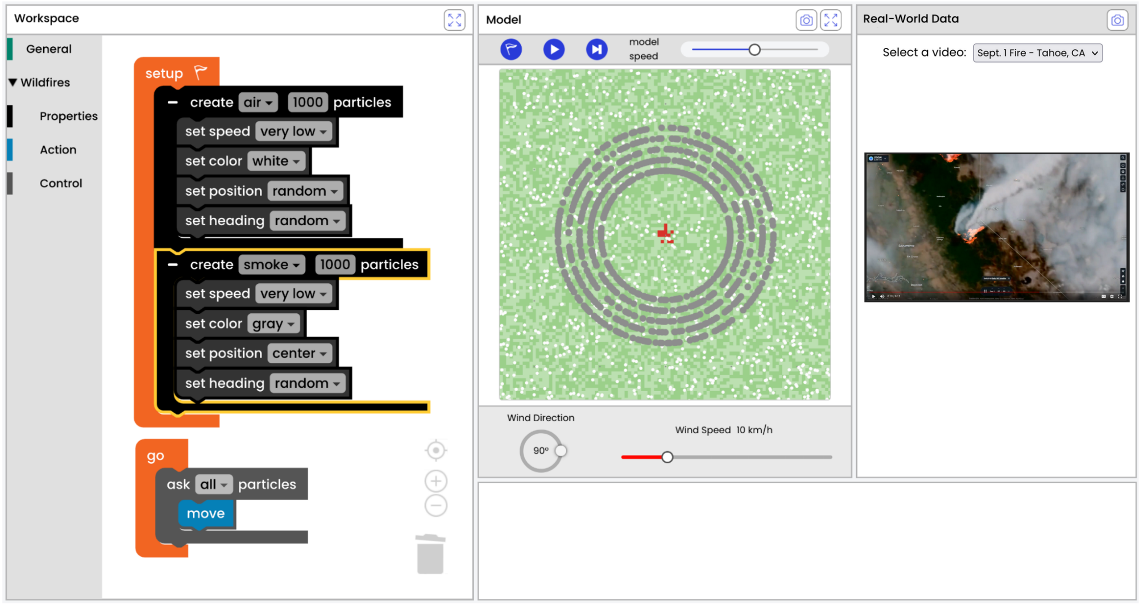

Fig. 1. MoDa showing a wildfire smoke model during a run alongside sattelite video. *Note that the speed setting includes a degree of randomness, creating multiple circles of smoke particles.*

2.2 Real-world data: Analyzing video data of wildfire smoke for comparison & model validation

Students have access to two clips of wildfire smoke spread selected from publicly available satellite video data. The videos are meant to draw students' attention to contrasting patterns in the shape of smoke: In the presence of strong winds, smoke spreads in the shape of a cone while in milder wind conditions, diffusion plays a more important role, causing the smoke to spread all around. The video clips are intended to encourage students to compare and validate their models to simulate the shape of smoke spread under different conditions. The video clips for this study were selected from a local wildfire that took place two months before the lesson.

3 THE STUDY

We ask: *How did the complementary focus on data and modeling contribute to student learning of the mechanisms underlying smoke spread?* Thus we examine how student responses shifted from the pre- to post-test, and how the design contributed to those shifts. The study was conducted in four 8th grade science classes taught by the same teacher in a public school in the Bay Area, California, and took place over 2 science class periods and lasted for 2 hours 25 minutes. The science teacher and a researcher co-taught the classes.

3.1 Unit Sequence

The unit sequence is presented in figure 2. Due to space constraints, we only describe the second part of the unit (items 3-8 in the figure). After an introduction (1) and student drawing (2), the research facilitator introduced the MoDa environment and asked students to add smoke and air particles to a model, and make them move (3). After this, the facilitator projected two or three student models and led a discussion around the model code and resulting simulation

(4). In the data analysis activity (5), the teacher presented the two videos described in section 2.2 and asked students to analyze the "path" and "shape" of smoke spread. In the following whole-class discussion (6), the teacher asked students to share their interpretations and propose explanations for the differences in patterns. In the second modeling activity (7), students revisited their initial models to recreate the patterns they noticed in the videos. Finally, the facilitator led a whole-class discussion (8) using and two or three student models an occasionally modifying them to better fit the data.

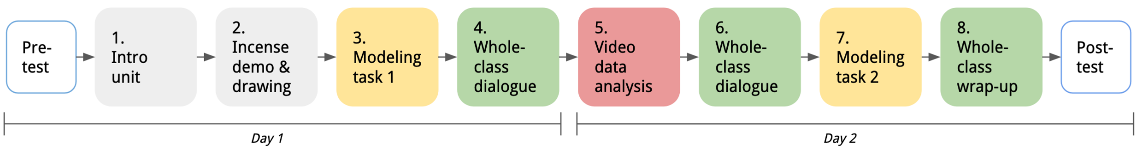

Fig. 2. Unit sequence

3.2 Data Collection and Analysis

A total of 87 students consented to be in the study. Of these students, we focus our analysis on 33 students from two classes in which we implemented a pre- and post-test. The test was designed to assess their understanding of how diffusion and wind impact visual patterns in the spread of wildfire smoke. The test included one open-ended question and two multiple-choice questions. Because of space constraints, we only present our analysis of the open-ended question which asked: "*On September 9, 2020, the sky around the Bay Area appeared red as smoke from fires around Northern California blotted out the sun. However, the closest fire was more than 50 miles away. How do the smoke particles make their way through the air?*". In addition, we video-recorded two to three focal pairs of students in each class (18 students) and all whole-class discussions of student models and data analysis.

To examine shifts in student responses from pre to post, we used epistemic network analysis (ENA) [14]. ENA is a technique to track expressed knowledge about key ideas and how those ideas are connected with each other. This analysis generates a graph in a two-dimensional Cartesian space (figure 3). In this graph, the codes applied to a unit of analysis (here, individual student responses) are represented as nodes positioned in the space. The positioning of nodes is done using optimization routines that maximize the differences between two groups (here, pre- and post-test responses). The resulting locations of nodes in an ENA graph can be used to interpret the dimensions of the projected space. For our research purposes, employing ENA showed us the frequency of co-occurrences or the strength of association between our codes in student responses, and whether or not they shifted from pre to post.

Our coding scheme draws on an existing framework for mechanistic reasoning [13] and was iteratively developed to represent the entities, properties and behaviors in this system. Two researchers independently coded student responses and used this framework to identify codes relevant to the phenomenon based on student responses. Through subsequent rounds of coding, 13 codes were finalized (See coding scheme in Appendix A.1). ENA was applied to the coded responses to graphically represent the shift in the frequency of codes and the co-occurrence of codes from pre- to post-test.

In addition, we analyzed video data of the three whole-class discussions (marked in green in the instructional sequence) from one class section. We chose to analyze these discussions because they included multiple student models and played a critical role in how students came to integrate data and modeling to build explanations about this phenomenon. We used collaborative viewing [8], in which multiple researchers collectively interpreted the videos and corresponding student models. This involved noting moments of shift towards identifying relevant micro-level entities and behaviors and linking them to the macro-level phenomenon of smoke spread. These moments were transcribed and analyzed to identify specific design elements that seemed to support the shift.

4 FINDINGS

4.1 How did student responses shift from pre to post?

The ENA graph in figure 3.a shows individual student responses as colored circles: The red circles represent student responses at pre, the blue circles represent student responses at post, and the black circles are specific codes that were used to analyze the data. As can be seen in the figure, from pre to post, student responses shifted from the left side of the graph to the right side. A closer look at the placement of the nodes reveals that the quadrants on the left and right side represent qualitatively different nodes and student responses. Broadly, the left side represents macro-level entities (e.g., smoke, wind) and their behaviors (e.g., blow, spread). In contrast, the right side represents micro-level entities (e.g, particles) and their behaviors (e.g., "push" and "bounce"). Note that the nodes towards the middle (e.g., "strength" and "direction") are ones that cannot be strictly categorized as either macro or micro. Finally, the weighted links in Figure 3.a represent a comparison of frequency of word associations between average pre and post answers.

The shift from pre to post was found to be statistically significant. Along the X axis, a two sample t test assuming unequal variance showed student pre-test reponses (mean=-0.36, SD=0.45, N=33) were different from post-test responses (mean=0.36, SD=0.409, N=33) at the alpha=0.05 level (t(63.34)= -6.85, p<0.001, Cohen's d=1.69). This shift is also reflected in the comparison between average networks for pre- and post-test responses, represented by the weighted connections between nodes. The thick red connections between "smoke," "wind," and "blow" indicate that associating those macro-level codes was more common at pre than post. The thick blue connection between "smoke," "particles," and "bounce" indicates an increase in students describing smoke and its behavior from an agent-level perspective. For example, one student's response shifted from "[...] *the more things burn, more* ***smoke*** *is produced and the* ***air*** *current* ***spreads*** *out the* ***smoke***." to "*The* ***smoke particles*** *interact with the* ***air particles****. The* ***smoke particles bounce off*** *the air particles so the* ***smoke moves*** *in the* ***direction*** *the* ***air*** *is moving since* ***air*** *applies* ***wind****. [...]*", after the unit (Figure 3.b).

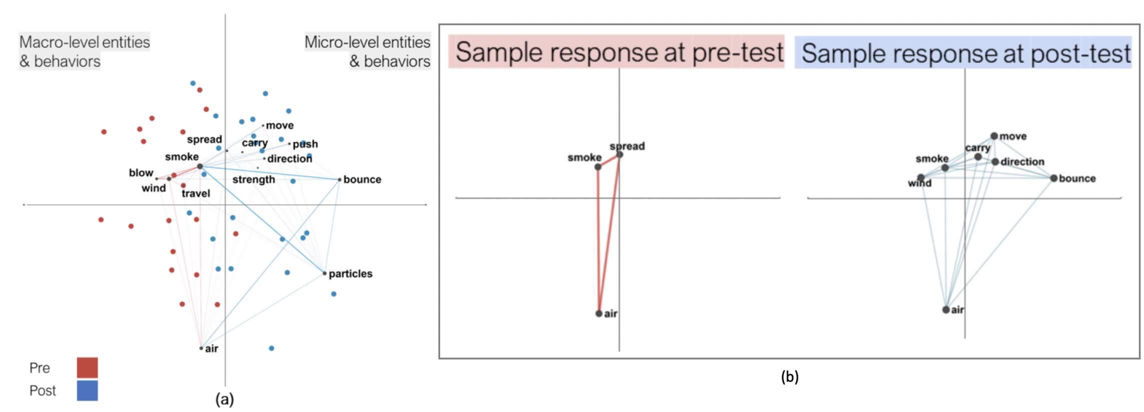

Fig. 3. Plots generated from ENA: (a) networks for student responses; (b) example of an individual student's response at pre and post.

4.2 How did the design support shifts in student responses from pre to post?

We found that three design elements played an important role in supporting student learning. The **naming of blocks** using simple, domain-specific language helped them reason about the function of a block. The **match between visualizations** of real and simulated data enabled them to use the data to inform and evaluate their models. And, **collective reflection on models** through whole-class discussions engaged students in a scaffolded analysis and comparison. In what follows, we describe three episodes from our data, drawn from whole-class discussion and listed chronologically, that illustrate how these design elements played a role in furthering student thinking.

4.2.1 How naming of the blocks (#1) and collective reflection (#3) supported initial observations and explanations of smoke spread. The first collective reflection established this activity as one in which students were to identify connections between code and patterns (or lack of) in the movement of air and smoke particles. For instance, when comparing the difference in the code of two student models, students noted that there was a smaller number of particles in the initialization of one model. Upon running the model, the facilitator asked if they could tell there were fewer particles visually. Jason said, "[...] *it's like they start as one circle of particles that are all covered, but once they spread out you don't really see many.*" This student noticed that, in the beginning, all smoke particles are clumped together in a circle, but when they start dispersing, it was clear that there weren't many particles. (Figure 4a)

Linking the movement of particles in the simulation with the code was further aided by the design of the block names. In this first collective reflection, many students had only used some of the blocks in the library. Yet the names of the blocks supported students in generating explanations of how the code worked. When the class was making predictions about another student's model code (Figure 4b), one student said, "*Well when it says **all particles**, and they're all **moving**, then when you **apply wind**, they're probably going to **interact** with the wind and **bounce off***". This student relied on the block names (apply wind, interact, and bounce off) to interpret what the code would do. Others used the block names to provide a more mechanistic explanation; for instance, "*Well it says they're going to **bounce off** of each other, so I think **once they bounce off each other, they're going to kind of spread out** and make the green a little bit more visible*". This student used the block "bounce off" to link the macro level outcome of "spreading out" of the smoke with the micro-level behaviors followed by individual smoke particles. In this way, collective reflection anchored the visual phenomenon of smoke spread and explanations using code as focal points. It also set students up for the video analysis task that followed.

4.2.2 How the match between visualizations (#2) supported the comparison between models and data. The video analysis activity was designed for students to attend to the "shape" of the smoke spread (a "cone" in strong wind v/s a "circle" in weak wind) in the two videos (See video screenshots in Figure 5b). In the whole-class discussion, students quickly came to consensus that the wind strength and direction explained the difference between the shape of smoke spread in the two videos. For instance, one student asserted that in the first video, "*the wind was blowing the smoke more towards the **northeast side***" while "*in the second video there **wasn't really much wind***".

In the final discussion, students shared that they included strength and direction of the wind to match their model to the video data. For instance, in Figure 5a, the student tried to match not only the cone shape, but the specific direction of smoke in the video by setting the wind speed to "medium" and the heading to "right" under the "apply wind" block. Similarly, one student observed during a class discussion that the shape of the smoke in the simulation was like a "circle", others agreed that it matched the second video. When the facilitator asked the class to predict what the simulation would look like if the wind speed was higher, students responded that it would look like the first video in which "*it* *started off as a big line and then it* ***separated into two*** *and then eventually it turned into a big cloud."* The match between the data and the simulation visualization gave them a visual reference of what their model simulation should look like.

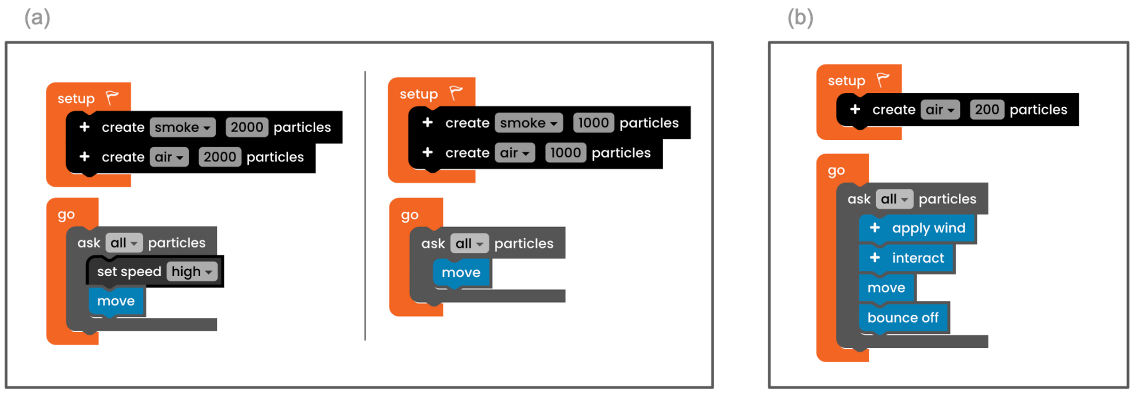

Fig. 4. (a) Two student models, one with more smoke particles (left) and one with fewer smoke particles (right). (b) Another student's model which used more particle actions.

4.2.3 How block names (#1), visualization match (#2) and collective reflection (#3) contributed to mechanistic reasoning. The final collective reflection also revealed that all three design elements contributed to students being able to explain code and reason about mechanism. Student model 5a, initially resulting from an effort to match visual features of the video and model, also offers an example of this next step in reasoning. When asked to describe the model code, the student said "*I added* ***more smoke particles than air particles****, and I* ***asked the air particles to apply wind*** *and make it* ***move right at medium speed****, and then I* ***asked all particles to interact by touching any particle it would bounce off*** *and* ***move****.*" This student's description included many of the necessary elements to model this phenomenon.

The facilitator then told students he would make three changes to this code to have it more closely match the data. He set the speed of air particles to "wind speed", the heading to "wind direction," and increased the number of air particles (Figure 5a). Manipulating the slider for wind speed in this modified model could make the simulation flexibly resemble one video or the other (Figure 5b). Though no students we observed generalized the model in this way, they were able to explain the resulting patterns in terms of causal mechanism. For instance, on running the modified model with a high wind speed, the facilitator asked students to explain how the code resulted in a cone shape in the simulation. One student explained that "*there's so much air particles that it's kind of push the smoke particles towards the same direction*". In this explanation, the student recognized that the air particles represented the wind ("*towards the same direction*") and that collisions between the air and smoke particles led to the smoke blowing in the same direction as the wind.

5 DISCUSSION: DESIGNING TO SUPPORT COORDINATING MODELING AND REAL-WORLD DATA

Here, we reflect on the role of each design element named above and how it relates to our broader design goal.

The naming of the blocks helped students link the visual movement of the particles in the simulation with the blocks code, and connect the macro-level outcome with the micro-level behavior of particles. One challenge we face is how closely to align the language within the blocks and environment with the language students use to interpret the data. For instance, when interpreting the data, students used cardinal direction to describe the wind. However, in

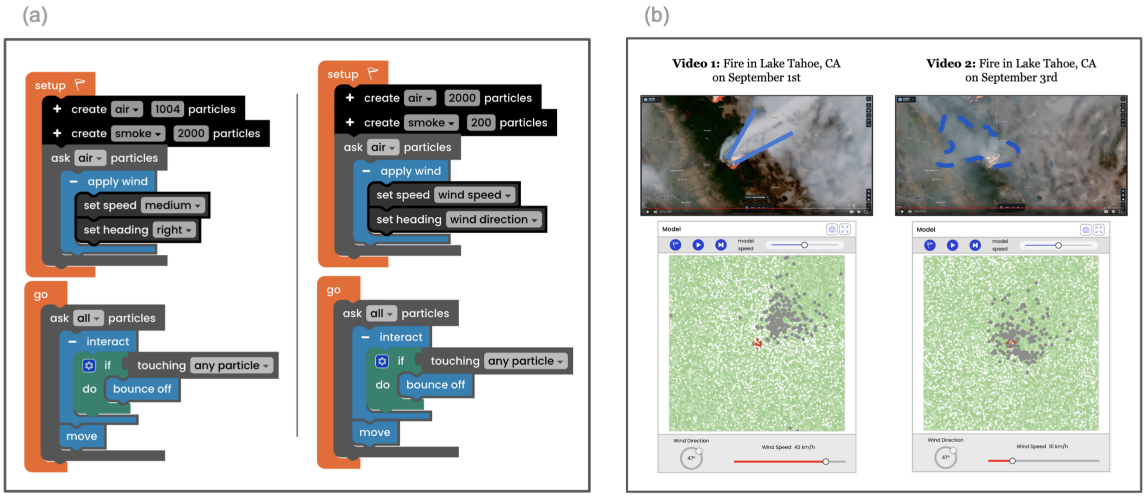

Fig. 5. (a) Student's model before and after the facilitator edits and (b) the resulting model output at different wind speeds compared with satellite video data of smoke spread shown to students (blue shape highlighting added for this paper).

Logo-derived languages such as NetLogo on which our environment is built, the "heading" is defined as 0-360 degrees. We are now considering using cardinal directions to align with how students readily describe the data.

The match between the data and the simulation visualizations made it possible for students to visually transition from the model to the data as they made comparisons to inform their modeling process. However, this visual matching has raised tensions around maintaining the same scale and time in the data and model while making the necessary abstractions visible for students to attend to. For instance, we had to strike a balance between making the visualization look like smoke without losing the agent perspective of being able to see *individual* smoke particles.

The whole-class collective reflection provided a way to make student ideas explicit, and then developed and refined through discussion of models. However, it was not straightforward for the facilitator to project multiple student models for discussion, as he had to quickly load student models on his computer in the middle of the discussion. To support seamless collective reflections with student models, we are designing a feature within the MoDa environment for teachers to view and tag student models and make modifications to their code without actually changing a student's saved model. Moreover, in these collective reflections, the teacher and research facilitator made several moves to help students consider the more generalized potential function of models to accommodate multiple presentations of data. We are analyzing these moves to inform the design of professional development for our partner teachers.

Looking ahead, we are working towards extending design elements identified in this study that will translate to other data sources beyond video. Expanding the design to include a wider range of data types will broaden the field's understanding of how to juxtapose computational modeling and real-world data for comparison and model validation.

6 SELECTION AND PARTICIPATION OF CHILDREN

87 students consented to participate in the study. The science teacher used curriculum that included computer simulations and thought students would be excited to get to build their own. They described the study as an opportunity for students to meet with researchers who "do work in science" and to build computer models. We prepared a letter for teachers to

share with families that described study procedures, what data that would be collected and how it would be shared in language understandable to adolescents. The teacher emailed this letter and the consent form to parents and guardians, inviting all students to participate. Students were told that they would do the lesson even if they chose not to be a part of the study. Three authors of this paper joined the class on Zoom to briefly describe their professional careers. The researcher who was in class (also a co-author on this paper), also answered student questions about his major in college and professional trajectory. This seemed to be of interest to many students.

ACKNOWLEDGMENTS

Thanks to our science consultants Beth Covitt, Daniel Banuti, Stephen Guerin, and Torsten Hadrich. This work is supported by National Science Foundation grant DRL-2010413. ENA analysis was done using WebENA, funded in part by NSF grants DRL-1661036 and DRL-1713110, the Wisconsin Alumni Research Foundation, and the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison. The opinions, findings, and conclusions do not reflect the views of the funding agencies, cooperating institutions, or other individuals.

REFERENCES

[1] 2013. *Next Generation Science Standards: For States, By States.* The National Academies Press, Washington, DC. https://doi.org/10.17226/18290
[2] Umit Aslan, Nicholas LaGrassa, MS Horn, and Uri Wilensky. 2020. Code-first learning environments for science education: a design experiment on kinetic molecular theory. *Constructionism 2020* (2020).
[3] Tom Bielik, Lynn Stephens, Dan Damelin, and Joseph S Krajcik. 2019. Designing Technology Environments to Support System Modeling Competence. In *Towards a Competence-Based View on Models and Modeling in Science Education*. Springer, 275–290.
[4] Engin Bumbacher, Zahid Hossain, Ingmar Riedel-Kruse, and Paulo Blikstein. 2018. Design Matters: The Impact of Technology Design on Students' Inquiry Behaviors. International Society of the Learning Sciences, Inc.[ISLS].
[5] Amanda Catherine Dickes, Pratim Sengupta, Amy Voss Farris, and Satabdi Basu. 2016. Development of mechanistic reasoning and multilevel explanations of ecology in third grade using agent-based models. *Science Education* 100, 4 (2016), 734–776.
[6] Tamar Fuhrmann, Bertrand Schneider, and Paulo Blikstein. 2018. Should students design or interact with models? Using the Bifocal Modelling Framework to investigate model construction in high school science. *International Journal of Science Education* 40, 8 (2018), 867–893.
[7] Julia Svoboda Gouvea and Aditi Wagh. 2018. Exploring the unknown: Supporting students' navigation of scientific uncertainty with coupled methodologies. International Society of the Learning Sciences, Inc.[ISLS].
[8] Brigitte Jordan and Austin Henderson. 1995. Interaction analysis: Foundations and practice. *The journal of the learning sciences* 4, 1 (1995), 39–103.
[9] Ken Kahn. 2007. Building computer models from small pieces.. In *SCSC*. 931–936.
[10] Victor Lee and Michelle Wilkerson. 2018. Data use by middle and secondary students in the digital age: A status report and future prospects. (2018).
[11] Brian J Reiser, Iris Tabak, William A Sandoval, Brian K Smith, Franci Steinmuller, and Anthony J Leone. 2001. BGuILE: Strategic and conceptual scaffolds for scientific inquiry in biology classrooms. *Cognition and instruction: Twenty-five years of progress* (2001), 263–305.
[12] Alexander Repenning and James Ambach. 1997. The agentsheets behavior exchange: Supporting social behavior processing. In *CHI'97 Extended Abstracts on Human Factors in Computing Systems*. 26–27.
[13] Rosemary S Russ, Rachel E Scherr, David Hammer, and Jamie Mikeska. 2008. Recognizing mechanistic reasoning in student scientific inquiry: A framework for discourse analysis developed from philosophy of science. *Science education* 92, 3 (2008), 499–525.
[14] David Williamson Shaffer. 2018. Epistemic network analysis: Understanding learning by using big data for thick description. *International handbook of the learning sciences* (2018), 520–531.
[15] Robert Tinker and Joseph Krajcik. 2012. *Portable technologies: Science learning in context*. Vol. 13. Springer Science & Business Media.
[16] Aditi Wagh and Uri Wilensky. 2018. EvoBuild: A quickstart toolkit for programming agent-based models of evolutionary processes. *Journal of Science Education and Technology* 27, 2 (2018), 131–146.
[17] Uri Wilensky. 1999. NetLogo. Evanston, IL: Center for connected learning and computer-based modeling, Northwestern University.
[18] Michelle Wilkerson-Jerde, Aditi Wagh, and Uri Wilensky. 2015. Balancing curricular and pedagogical needs in computational construction kits: Lessons from the DeltaTick project. *Science Education* 99, 3 (2015), 465–499.

A CODING SCHEME

Entities: Smoke/air/wind/particles; Properties: Direction/strength; Behaviors: Blow/bounce/carry/move/push/spread/travel