Putting the Data in MoDa: Integrating Agent-Based Modeling, Quantitative Data Analysis, and Teacher Responsivity to Investigate Complex Phenomena

Aditi Wagh, Concord Consortium, MIT Scheller Teacher Education Program, awagh@concord.org
Michelle Wilkerson, University of California, Berkeley, mwilkers@berkeley.edu
Adelmo Eloy, Teachers College, Columbia University, adelmo@fablearn.net
Tamar Fuhrmann, Teachers College, Columbia University, tf2464@tc.columbia.edu
Paulo Blikstein, Teachers College, Columbia University, paulob@tc.columbia.edu

Abstract: This tech demo introduces key enhancements to MoDa: A free, open-source web based modeling and data analysis system designed to support students in making sense of complex systems. MoDa is a domain-specific, block-based computational modeling tool that allows students to build models side-by-side to real-world data. Our latest enhancements include integrating quantitative data in the system using the Common Online Data Analysis Platform and Activity Player to support structured data-rich investigations, allowing students to use a wider variety of real-world data sources to refine their model toward reproducing important features of system behavior both *qualitatively*, by reproducing important visual and dynamic behaviors, and *quantitatively* by allowing closer comparison of patterns, relationships, and variability.

Introduction & Motivation

A central goal of science education is for students to interpret patterns in complex phenomena in the world and reason about potential causes underlying them. Sense-making of scientific phenomena and proposing solutions to problems often begins with finding patterns in data and attending to variability in those patterns. Students' lives are filled with such patterns (e.g., The water seemed warmer at my beach this summer - is it getting warmer everywhere?) and sources of variability (e.g., My cousins, who live across the country, aren't noticing higher water levels - why might that be?). Once initial trends emerge, students can propose causal mechanisms to explain the observed patterns.

Typically, modeling activities at the K-12 level engage students in making observations of a real-world phenomenon, constructing models that seek to explain or predict system behavior, and then validating and refining those models based on comparisons to the real-world phenomenon. Most often, such comparisons are *qualitative*—the simulation is compared to the visual and behavioral characteristics of the real-world phenomenon (Schwarz et al., 2009), with less attention to its mathematical and probabilistic aspects. However, there is growing evidence that attending to the *quantitative* characteristics of systems and their corresponding models can deepen student engagement. For example, deeper analysis of statistical features of real-world data can introduce students to important ideas about variability and error that can then be incorporated into models (Lehrer & Romberg, 1996; Lehrer & Schauble, 2004). Conversely, juxtaposing data produced by simulations with data generated in sensor-equipped laboratory settings can encourage students to use patterns to validate and revise their models (e.g., Blikstein, 2014; Bumbacher, 2019; Fuhrmann, 2018). Similarly, prompting students to reason about quantitative details of simulation output, such as random noise, supports them in connecting the micro and macro level behaviors of a system (Wilkerson & Wilensky, 2015). We extend this work to explore comparing simulated data to real-world data collected *in situ*, particularly large-scale public data streams.

MoDa is a tool to support integrating computational modeling and multiple forms of data analysis, including large-scale public quantitative data, in ways that are adaptive and responsive to teachers' and students' needs. As a domain-specific, block-based modeling toolkit with embedded support to compare students' constructed models to real-world observations side-by-side (Wagh et al., 2022; Fuhrmann et al., under review) our latest design of MoDa adds support for quantitative and statistical analysis, through integrating students' model output and real-world datasets. In particular, it integrates with the Common Online Data Analysis Platform (CODAP), a web-based data analysis tool designed for middle and high school students, and widely used in schools (W. F. Finzer & Damelin, 2014).

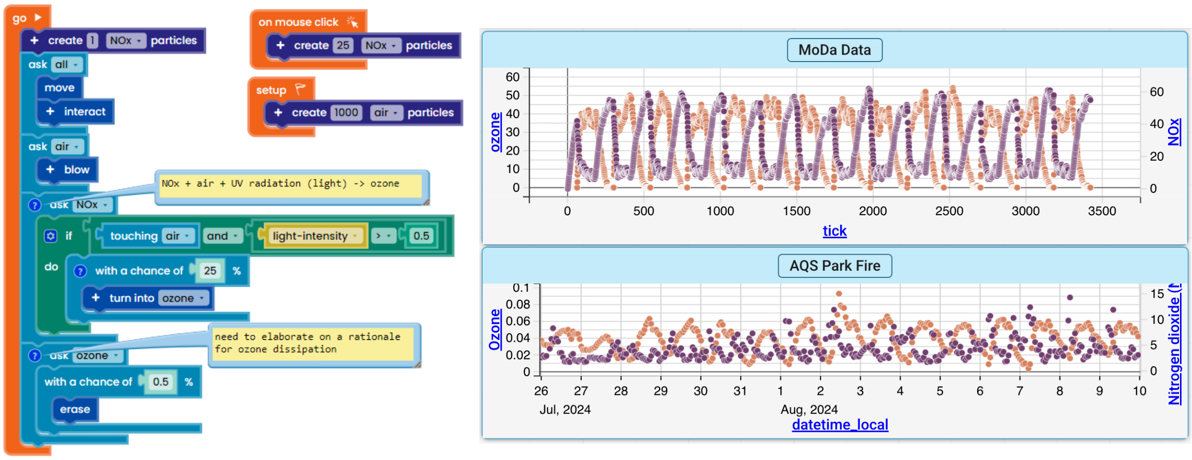
Because integrating this variety of modalities and evidence sources requires careful coordination, we are designing MoDa with a variety of highly adaptable features so that units can be adjusted in response to students' and teachers' immediate classroom needs. Below, we describe three aspects of this new design iteration: (1) A user interface for creating domain-specific agent-based blocks; (2) Exporting data from student models into CODAP for comparison with quantitative real-world data; and, (3) Leveraging Activity Player to support modular technology development in order to respond to teacher and student needs.

Quantitative Analysis of Models

Data explorations support student reasoning about scientific relationships, patterns, and variability (Rivet & Ingber, 2017). Students engage in such explorations when they construct, analyze and use data. Because "data" can take many forms including simulation output, student-generated measurements, and public data (Lee & Wilkerson, 2018), it is crucial for students to understand how different "data-texts" can work together to explore scientific phenomena and the different functions they serve toward building models (Duschl et al., 2021). It is also critical for students to attend to variability in the data so they can reason about patterns and associated levels of certainty in a phenomenon (Lehrer & Romberg, 1996; Lehrer & Schauble, 2004). Figure 1 features a block library focused on modeling air quality, along with the resulting simulated data and real air quality data.

MoDa is inspired by work on agent-based modeling (ABM) in education, and NetLogo in particular. ABM supports reasoning about causal mechanisms and probabilistic factors at the individual level, and simulates the behavior of elements in a complex system at different levels of analysis using simple probabilistic rules (de Jong et al., 2013; Wilensky & Resnick, 1999). By creating and studying these models, students can examine how population-level patterns emerge from individual-level processes (Wilensky & Reisman, 2006), supporting reasoning about mechanisms (Dickes et al., 2016; Hsiao et al., 2019) and probability (Abrahamson et al., 2006).

Figure 1. The new design iteration of MoDa integrates domain-specific programming (left) with CODAP to allow model output (top right) to be analyzed and compared with real-world data (bottom right).



Data explorations and ABM are complementary practices. Both data explorations and ABM are rich scientific activities that involve multiple practices highlighted in science education reforms (e.g., NGSS Lead States, 2013). Uniting these practices through a coherent technology and curriculum has the potential to support students in reasoning about patterns in complex phenomena and their underlying causes in new ways. Emphasizing the probabilistic nature of systems, and the inexact science of extracting meaningful patterns from "messy" data, also aligns with important emerging ideas in computing education (what Tedre et al., 2021 call "CT 2.0"). Though our focus is on developing scientific models from data rather than machine learning, the emphasis on deriving meaning from data, parallel processing, and assessing the "correctness" of a model as evaluating its "accuracy at modeling some selected features of a phenomenon" are foregrounded in an approach that brings together quantitative data analysis and agent-based modeling. However, integrating these perspectives and levels of analysis is not straightforward for students and requires careful design and scaffolding (Wilkerson & Wilensky, 2015). That is what we aim to do through this work.

Responsivity to Teachers' Pedagogies and Student Ideas

Introducing multiple sources of data across multiple scientific practices can be pedagogically challenging. To support teachers and students to meaningfully engage in this complex work, we aim to be attentive and adaptive to teachers' needs and students' ideas as they engage in this ambitious work. To do this, we draw on the notion of *responsivity in technology design*: Being responsive (Robertson et al., 2015) to students' ideas through the design of a computational environment (Walkoe et al., 2017).

In prior work, we have considered how responsivity could be built into computational environments by designing for relatively rapid changes to the environment to reflect students' ideas (Wilkerson et al., 2015). We identified multiple forms of responsivity ranging from renaming, adding, or removing existing blocks to better align with student ideas to extending out an existing blocks to enable students to pursue new investigations beyond the intended design to modifying the definition of a particular block, allowing for "finer tuning" of students' models. We believe that attending to responsivity in this current project is especially important. When working with real-world data, these types of responsivity can allow students to tap into the everyday language they may use to describe patterns, explore the quantitative dynamics of systems, and model more subtle differences between otherwise qualitatively similar patterns. Moreover, because integrating quantitative data analysis and ABM is pedagogically challenging work, enabling teachers to modify the computational environment to tailor to their students' ideas can help them better support their students.

Goals & Technology Design

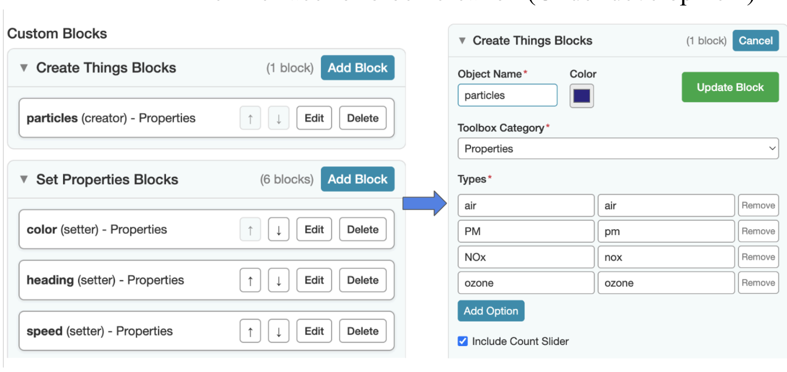
To combine the scientific practices of quantitative data exploration and ABM into a flexible platform, we are integrating data explorations in CODAP (Common Online Data Analysis Platform) with ABM in MoDa into a curriculum wrapper called Activity Player (The Concord Consortium, N.D.). The latest design of MoDa runs on a lightweight agent-based simulation engine called Atomic Agents (McNeill, 2022/2025). It can export simulated data directly into CODAP using Activity Player provides a platform to house MoDa and CODAP components, enable data transfer between the two so that simulated data can be directly compared with real-world sources, and provides curricular support infrastructure surrounding these technology components.

User Interface for educators to create domain-specific agent-based blocks

An underlying vision for the latest iteration of MoDa is to democratize access to creating and modifying block libraries to support agent-based modeling investigations in classrooms. Our conjecture is that this can enable educators and educational designers (including teachers) to support modeling investigations that are responsive to students' ideas in ways that go beyond the research team's intent.

To support this, we have been developing a visual interface to create and adapt domain-specific blocks (Figure 2) to design a block library for a particular unit. The interface is intended to support multiple levels of responsivity so that a user without much prior coding experience can modify names of entities, variables, and behaviors in the library, or remove blocks as needed, while users with more coding expertise could also modify the code for an existing block or generate a new block.

Figure 2
The interface for block creation (Under development)



Besides supporting block creation, we are envisioning additional ways to make MoDa responsive to teacher and student needs. For instance, earlier implementations of MoDa enacted *block unpacking*: Students being able to inspect, and modify the definition of a particular domain-specific block (Eloy et al., 2025). We envision making it possible for educators to select which blocks to make "unpackable" based on which mechanisms they would like their students to explore. For instance, depending on whether a teacher wants to focus on mechanisms for ozone production and/or dissipation, they can select which specific blocks to unpack so students can do a deeper dive into those specific scientific ideas (See Figure 1).

Simulating and Contextualizing Quantitative Data using ABMs and CODAP

To enable students to compare quantitative data from their ABMs with real world data, we are building an interface to export simulated quantitative data generated by students' ABM directly to CODAP for further exploration. This function builds on work that has embedded and exported data from simulations directly into CODAP (W. Finzer, 2014; H.-S. Lee et al., 2014), further coupling such exports with real-world comparative datasets. To allow for flexible use of these functions across contexts, students will be able to export (as relevant) one or more of the following kinds of data from their ABM: (1) Population data such as the number of a type of entity (e.g., "particles" in Figure 1); (2) Data about the environment in the form of "patch" variables; and (3) Sensors embedded in the model (e.g., a probe collecting data) or global reporter variables relevant to an activity.

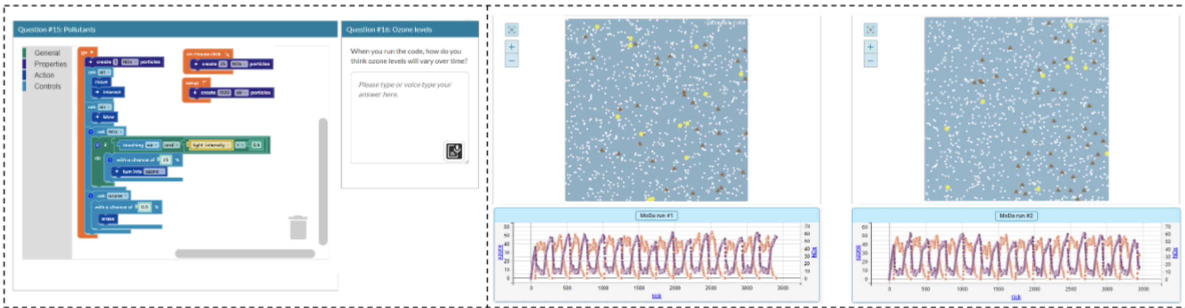
A central goal of the project is to support students to compare patterns in quantitative data from the real world and their ABMs for sense-making of mechanisms resulting in those patterns. In contrast to simpler systems, doing this in complex systems often involves connecting aggregate quantitative patterns to microscopic mechanisms that are described at a different unit of analysis - that of the agent. Hence, it is vital that each aggregate dataset exported from an ABM to CODAP be associated with the corresponding agent-level mechanisms that produced it. Our hope is that this can help students validate which agent-level mechanisms resulted in relevant aggregate trends that best match the real-world data. We are working towards providing this agent-level context to a dataset about the aggregate system in CODAP in two ways: (1) Enabling students to record and export visual data in the form of a video-recording of an agent-based simulation run; and, (2) Enabling students to "tag" their model code to the dataset in some way. Collectively, these can provide a context for students to compare quantitative trends in data to examine underlying agent-level mechanisms.

Modular Design/Integration of Technology Components

To enable responsivity to teachers' pedagogical needs and students' ideas, we will leverage the built-in modularity of Activity Player through both the design of the technology and curriculum. The technology described here around integrating ABM and quantitative data consists of four modular components: (1) Blocks-based ABM code; (2) The simulation; (3) Quantitative data from an ABM exported to CODAP; and (4) Real-world quantitative data represented in CODAP. Using Activity Player, we are currently exploring how these components can be combined in different ways to support a range of pedagogical and curricular goals. For instance, to support comparison of trends and variability in data from ABM and the real world, (3) and (4) can be adjacent (See Figure 1). On the other hand, to support code reflections (Wagh et al., In review), (1) can be used independently to ask students to predict what trends a model would result in (Figure 3a), or multiple model outputs can be combined to have students predict what output a program would result in (Figure 3b).

Figure 3a & 3b

(a) A code reflection for students to make a code prediction; (b) Comparing outputs from two model runs



(a) (b)

Conclusion

This tech demo presents the most recent design iteration of MoDa, a tool to support integrating computational modeling and quantitative data analysis in ways that are adaptive and responsive to teachers' and students' needs. At this tech demo, participants will be able to interact with the Activity Player-based MoDa system. They will have a choice of exploring data and building models related to key air quality or water quality dynamics. They will also be able to preview accompanying curricular documents we are developing for the environment.

Authorship Statement

AW initiated and conceptualized the paper. AW and MW wrote the initial draft of the paper. AE contributed to initial brainstorming about the paper, provided cleaned datasets to generate data visualizations, and created screenshots. All authors contributed to editing the paper. All authors approve the paper.

References

Abrahamson, D., Janusz, R. M., & Wilensky, U. (2006). There Once Was a 9-Block …- A Middle-School Design for Probability and Statistics. *Journal of Statistics Education*, *14*(1). https://doi.org/10.1080/10691898.2006.11910573

Blikstein, P. (2014). Bifocal Modeling: Promoting Authentic Scientific Inquiry Through Exploring and Comparing Real and Ideal Systems Linked in Real-Time. In A. Nijholt (Ed.), *Playful User Interfaces: Interfaces that Invite Social and Physical Interaction* (pp. 317–352). Springer Singapore. https://doi.org/10.1007/978-981-4560-96-2_15

Bumbacher, E. W. (2019). *Supporting evidence‑based reasoning in science education: An examination of pedagogical and technological approaches* [Doctoral dissertation, Stanford University]. https://purl.stanford.edu/cj780bz0155

Concord Consortium. (n.d.). Activity Player [Software]. https://activity-player.concord.org

de Jong, T., Linn, M. C., & Zacharia, Z. C. (2013). Physical and Virtual Laboratories in Science and Engineering Education. *Science*, *340*(6130), 305–308. https://doi.org/10.1126/science.1230579

Dickes, A. C., Sengupta, P., Farris, A. V., & Basu, S. (2016). Development of Mechanistic Reasoning and Multilevel Explanations of Ecology in Third Grade Using Agent-Based Models. *Science Education*, *100*(4), 734–776. https://doi.org/10.1002/sce.21217

Duschl, R., Avraamidou, L., & Azevedo, N. H. (2021). Data-Texts in the Sciences. *Science & Education*, *30*(5), 1159–1181. https://doi.org/10.1007/s11191-021-00225-y

Eloy, A., Fuhrmann, T., de Deus Lopes, R., Blikstein, P., & Wagh, A. (2025). *Opening the Black Box: Student Motivations in an Unpackable-Block-Based Computational Modeling Environment*. https://repository.isls.org//handle/1/11698

Finzer, W. (2014). Games, Data, and Habits of Mind. In T. Wassong, D. Frischemeier, P. R. Fischer, R. Hochmuth, & P. Bender (Eds.), *Mit Werkzeugen Mathematik und Stochastik lernen – Using Tools for Learning Mathematics and Statistics* (pp. 71–84). Springer Fachmedien. https://doi.org/10.1007/978-3-658-03104-6_6

Finzer, W. F., & Damelin, D. (2014). Common Online Data Analysis Platform (CODAP). *NSF Award*, *14*, 35470.

Fuhrmann, T., Schneider, B., & Blikstein, P. (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), 867–893. https://doi.org/10.1080/09500693.2018.1453175

Hsiao, L., Lee, I., & Klopfer, E. (2019). Making sense of models: How teachers use agent‑based modeling to advance mechanistic reasoning. *Prof. Klopfer via Mark Szarko*. https://dspace.mit.edu/handle/1721.1/124342

Lee, H.-S., Pallant, A., Tinker, R., & Horwitz, P. (2014). *High School Students' Parameter Space Navigation and Reasoning during Simulation-Based Experimentation*. https://repository.isls.org//handle/1/1180

Lee, V. R., & Wilkerson, M. H. (2018). *Data use by middle and secondary students in the digital age: A status report and future prospects* [Commissioned Paper for the National Academies of Sciences, Engineering, and Medicine, Board on Science Education, Committee on Science Investigations and Engineering Design for Grades 6-12].

Lehrer, R., & Romberg, T. (1996). Exploring Children's Data Modeling. *Cognition and Instruction*, *14*(1), 69–108. https://doi.org/10.1207/s1532690xci1401_3

Lehrer, R., & Schauble, L. (2004). Modeling Natural Variation through Distribution. *American Educational Research Journal*, *41*(3), 635–679.

McNeill, G. (2025). *Gjmcn/atomic-agents* [JavaScript]. https://github.com/gjmcn/atomic-agents (Original work published 2022)

NGSS Lead States. (2013). *Next Generation Science Standards: For States, By States*. National Academies Press. https://doi.org/10.17226/18290

Rivet, A. E., & Ingber, J. (2017). Analyzing and interpreting data. In *Helping students make sense of the world using next generation science and engineering practices* (pp. 159–180). NSTA Press. https://scholar.google.com/scholar?cluster=11238032367674721865&hl=en&oi=scholarr

Robertson, A. D., Scherr, R., & Hammer, D. (2015). *Responsive Teaching in Science and Mathematics*. Routledge.

Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., Shwartz, Y., Hug, B., & 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. https://doi.org/10.1002/tea.20311

Tedre, M., Denning, P., & Toivonen, T. (2021). CT 2.0. *Proceedings of the 21st Koli Calling International Conference on Computing Education Research*, 1–8. https://doi.org/10.1145/3488042.3488053

Wagh, A., Fuhrmann, T., Eloy, A. A. da S., Wolf, J., Bumbacher, E., Blikstein, P., & Wilkerson, M. (2022). MoDa: Designing a Tool to Interweave Computational Modeling with Real-world Data Analysis for Science Learning in Middle School. *Proceedings of the 21st Annual ACM Interaction Design and Children Conference*, 206–211. https://doi.org/10.1145/3501712.3529723

Walkoe, J., Wilkerson, M., & Elby, A. (2017). *Technology-Mediated Teacher Noticing: A Goal for Classroom Practice, Tool Design, and Professional Development*. https://repository.isls.org//handle/1/215

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. https://doi.org/10.1207/s1532690xci2402_1

Wilensky, U., & Resnick, M. (1999). Thinking in Levels: A Dynamic Systems Approach to Making Sense of the World. *Journal of Science Education and Technology*, *8*(1), 3–19. https://doi.org/10.1023/A:1009421303064

Wilkerson, M. H., & Wilensky, U. J. (2015). Patterns, probabilities, and people: Making sense of quantitative change in complex systems. *Journal of the Learning Sciences*, *24*(2), 204–251. https://doi.org/10.1080/10508406.2014.976647

Wilkerson, M., Wagh, A., & Wilensky, U. (2015). Balancing curricular and pedagogical needs in computational construction kits: Lessons from the DeltaTick project. *Science Education*, *99*(3), 465–499. https://doi.org/10.1002/sce.21157

Acknowledgments

We would like to thank Leslie Bondaryk and her development team at Concord Consortium for leading technical development on this work. We would also like to thank Lucas Coletti for his work with testing design prototypes and supporting data cleaning. This material is based upon work supported by the National Science Foundation under Grant No. 2445609.