Learning about Functions in a Data-Rich Environment[1]

Tim Erickson, Senior Scientist
Epistemological Engineering

For some time now, progressive mathematics educators have been trying, with varying success, to get more data into the mathematics curriculum. Why use data? Working with data answers the "when are we ever going to use this?" question (Witmer 2002); as citizens, we need to deal with data in our everyday lives; and data gives a dose of tangible reality to a subject that, left to its own, often becomes too dry and abstract. In the USA, this effort has produced published materials (e.g., Burrill et al. 1997; Murdock et al. 2002; Carlson and Winter 1997), presentations, and workshops for secondary-school teachers—especially as sponsored by the purveyors of high-powered calculators. Despite these resources, however, many mathematics courses are still data- (and context-) poor.

What can help math teachers use data in meaningful ways? As part of a *physics* curriculum development project (Erickson 2002), our students have been using dynamic data analysis software to work with elementary functions as models for their data. Not surprisingly, many of their challenges are fundamentally mathematical; and solving the mathematical conundrum leads (we believe) to better conceptual understanding in physics (we are not alone, see, e.g., Wells et al. 1995). It is a relatively small step to view these activities as mathematics curriculum with a physics context.[2] Elsewhere (Erickson 2004) we discuss the process of constructing mathematical models and how mathematical elements such as function parameters correspond to physical meaning. There, we also discuss some particular challenges students seem to face coordinating their understanding of data and functions. In this paper, we will see a few examples of the mathematical issues students face when they work with data. We will see how technology lets us focus on ideas that are not particularly accessible without it. How important are these ideas? How essential is the technology? What insights from technology-rich learning apply in general? How does working with data in this way affect students' overall mathematical understanding? Our ongoing research is only beginning to address these questions.

Fitting Curves to Data: Making Parameters and Using Them

In a typical modeling activity, students fit a function to data. Such activities have been around for years, but technology may open the door to new approaches. Let's look at an example. Figure 1 shows data for how far a ball rolls on a pool table as a function of its initial speed.

How do the students make and alter the function? In our project, students use Fathom (KCP Technologies, 2001). They use a formula editor to enter the function they want to plot. In this

---

[1] This material is based upon work supported by the National Science Foundation under Award Number DMI-0216656. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

[2] Must it be physics? No, but physics is particularly easy to connect to functions.

example, they set the parameter **A** using a *slider*, which is essentially a pointer on a number line. They drag the pointer to change the value. The graph of the function updates as they do this, so students see the effect of changing the parameter in real time.

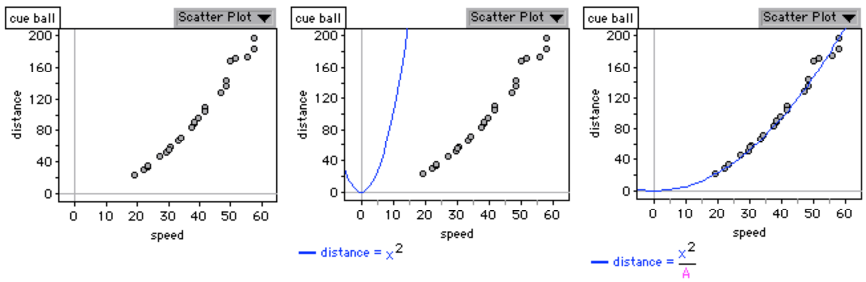


Figure 1. If students recognize the original shape (in the left-hand picture) as being more or less quadratic, they begin by plotting $y = x^2$ on the graph (middle picture). That does not fit, so they must transform the equation; here they have introduced a parameter for the denominator and plotted $y = x^2/\mathbf{A}$ (right-hand picture). They have set **A** = 17.

Already, we can see (or rather *imagine*, in this static medium) a benefit from technology: since the function moves in synchrony with the slider, students get a visceral "feel" for the effect of the parameter on the shape of the function. But the flexibility of software has additional benefits. The traditional task might give away the functional form ("find the value of A such that $y = x^2/\mathbf{A}$ fits the data as well as possible") but that constraint may not be necessary. Let the student find the functional form. If Bob thinks, "the points are below the curve, I should subtract," he could enter $y = x^2 - \mathbf{A}$. If Maria thinks it looks exponential, she could enter $y = \mathbf{A}^x$. But when each varies the **A** slider, they will see immediately that their functions will not fit the points for any value of **A**.

So finding functions is not just a matter of sliding values. A student can choose the function itself, how to express the function, what parameters to use, what to slide and how far, and if there is more than one parameter, which parameter and in what order. These more open-ended possibilities become more practical when the student has technological help. To reflect these possibilities, we must also upgrade the questions we ask students. In addition to "what is the best value of **A**?" we can ask the deeper, "what happened to the function as you changed **A**, and why?"

Let's look at a 3-parameter situation. Suppose we have data for a projectile in *x* and *time*. We might take the data from a video (shown) where we measure *time* in frames and *y* in bricks. For this parabola, we need three parameters, but we have a choice whether to write it as $\mathbf{A}t^2 + \mathbf{B}t + \mathbf{C}$ or $\mathbf{A}(t - \mathbf{H})^2 + \mathbf{K}$. Our students tried it both ways and decided that the second version was easier to make fit—because each parameter did something understandable: **A** changed the curvature, and **H** and **K** translated the function horizontally and vertically.[3]

[3] The physics students also had to make the correspondence with the "physics" formula, $h = h_0 + v_1 t + (1/2)gt^2$. These physics students said, "hey, we saw that in math class—I never knew we would actually *use* it for anything."

But getting the parabola close to *all* the points is difficult; as students make fine adjustments to one parameter—to fit one part of the data better—another part gets worse. We need a new tool.

Importance of Residuals

Students can use *residual plots* to show the vertical distances from data points to the curve. If the fit is good, these residuals will be randomly scattered about zero. A residual plot acts as a "magnifying glass," expanding the vertical scale so that students see fine deviations that are hard to detect, especially where the function is steep. When students see that there is still a pattern in the residuals, they know that there are still adjustments to be made. But which parameter to adjust? In this situation, students discover that when they look at the residuals, **A** changes the curvature and **K** changes the vertical position (as before), but that now, **H** changes the residuals' *slope*. So students develop a system: use **A** to make the residuals *straight*, then use **H** to make them *flat*, and finally use **K** to make then *zero*.

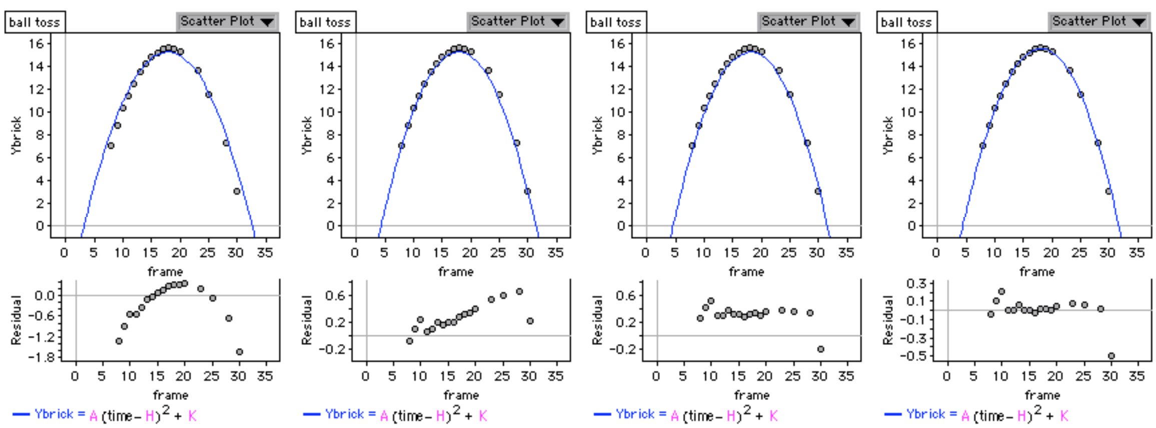


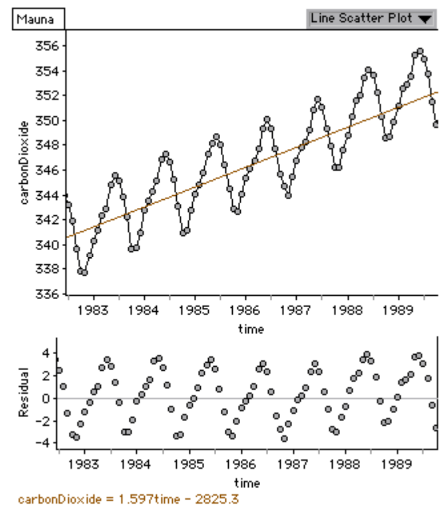
Figure 2: The left picture shows our first try, where the top points fit pretty well. The residual plot (below each main plot) shows how the left-hand points are lower than the curve. Residuals are as large as 1.5 bricks. To get the second figure, we adjusted **A**; the residuals are straight, but tilted. In the third, we adjusted **H**, and now the residuals are flat, but low. In the final picture, we adjusted **K** so that the residuals cluster around zero. Now the residuals are more or less random, and mostly less than 0.1 bricks.

Deeper understanding comes when students try to explain *why* the parameters (especially **H**) have the effect that they do. More experienced students can do so algebraically (seeing how the x^2 term cancels, leaving only a linear function), but other students can still offer reasonable, qualitative explanations.

Some computer programs and calculators will, of course, perform a least-squares fit to a wide variety of functions and give you the "best" parameter values. But the point of this activity is not only to fit the data, but also to learn about the functions: and that happens when you build the functions—and set the parameters—yourself.

Discussion

These examples serve to give a taste of the kinds of things we can now do in teaching about functions in a data-rich environment. It's worth musing a little about sources of data and mathematical possibilities: temperatures falling towards equilibrium form beautiful exponentials, as do the times between bounces of a bouncing ball. The (signed) distances of Jupiter's moons from the planet give us convincing sinusoids, as does the light intensity transmitted through crossed polarizing filters. The CO_2 concentrations as measured from the summit of Mauna Loa (shown at right; students manipulate the line directly by dragging its ends) show a steady increase with a periodic fluctuation; the *residuals* from the line are roughly sinusoidal—teaching us about addition of functions. And (in our lab, at least) the temperature gain in a microwave oven is a linear function of the time spent heating, but a power law function of the mass of the material being heated. Other techniques for fitting, especially transforming data in order to make it linear, make sense in this context as well.



Could students do these things by hand, or using a graphing calculator? Of course. But the speed and the dynamic nature of the software give students a more vivid experience of how functions work. We believe this results in correspondingly more understanding.

How is what we describe different from what is already going on? How do those differences translate into mathematical learning opportunities? We'll look at three things: dynamism, redoability, and authenticity.

Dynamism. Our first principle is that we use what we might call "dynamic" technologies. A data tool that lets students dynamically manipulate a function's parameters will help them see how functions live in families, and help them characterize the effect each parameter has on a function's shape. That the function changes "during the drag" is vitally important. We might say that with dynamic technologies, students can observe and explain mathematical *phenomena*.

What do we mean by a mathematical phenomenon? An example: Suppose we have data well-fit to a sinusoid, controlled by up to four parameters (i.e., $y = \mathbf{A} \sin(\mathbf{B}(x - \mathbf{C})) + \mathbf{D}$). Now we change the phase slider, **C**. What will we see in the residuals? They are near zero when **C** is correct and show a sinusoid with an amplitude of 2**A** when **C** is 180° out of phase. So as we change **C**, we see a phenomenon in the residuals: they vanish, then return at double strength. Explaining that phenomenon helps us understand about sinusoids, and about interference. A subtler issue is, what is the pattern in the residuals when **C** has some intermediate value? This is not a question about data, but about the behavior of functions, in this case, the difference between two sinusoids. Yet the question arises naturally—you see it happen and wonder what's going on—when you're exploring actual data (and you have enough technological power to drive the display).

We have often been surprised by what happens in a data display. And we tend to notice a *happening* rather than something static. On reflection, we often find that this phenomenon, this dynamic occurrence, illuminates the data, but also—most important in this context—illuminates a mathematical principle.

Redoability. Speed and ease of use also make it easier for students to redo an analysis. This “redoability” is important as we try to develop more open-ended, constructivist activities. In an example above, Maria can even choose the wrong function (exponential instead of quadratic) and her assignment is not ruined; the technology helps her redo the work quickly. This makes exploration less risky and more attractive, and naturally gives students a wider range of experiences with more functions.

Authenticity. We could just as well have used simulated, cookbook data. What is the value in using real data, such as data from physics—especially if the students aren’t learning physics?

First, data being real can be a point of entry for a student weary of abstraction. The more engaging the context the better, but the “wow” factor is not as important as reality. Second, real data have variation, and students need to see that the curve does not always go right through all of the points. But finally—and related to redoability—using real data helps students learn to be critical about which function works. Without real data, there’s really no choice, no chance for students to see (for example) that not all upwardly-curving functions are the same, nor the chance to find out that distinguishing between them can sometimes be dicey.

References

Burrill, Gail F. et al. 1997, 1999. *Data Driven Mathematics: Mathematics in a World of Data* (series). Dale Seymour Publications (Pearson Learning).

Carlson, R. J. and Winter, M. J. 1997. *Transforming Functions to Fit Data: Mathematical Explorations Using Probes, Electronic Data Collection Devices, and Graphing Calculators*. Emeryville, CA: Key Curriculum Press.

Erickson, T. 2002–04. *Integrating Science and Mathematics though Data.* Award DMI-0216656. National Science Foundation, USA.

Erickson, T. 2004. Stealing from Physics: Modeling with Mathematical Functions in Data-Rich Contexts. ICME-2004, Copenhagen.

KCP Technologies. 2001. *Fathom Dynamic Statistics® Software*. Emeryville, CA; Key Curriculum Press.

Murdock, J., Kamischke, E., and Kamischke, E. 2002. *Advanced Algebra through Data Exploration*. Emeryville, CA: Key Curriculum Press.

Wells, M. Hestenes, D., and Swackhamer, G. 1995. A Modeling Method for high school physics instruction. *Am. J. Phys*. 63 (7), July 1995, 606-619.

Witmer, J. 2002. *Data Driven Activities: Teaching Statistics Through Mathematics*. Paper presented at the 6th International Conference on Teaching Statistics (ICOTS 6), Cape Town, South Africa.