

Modeling data with linear, quadratic, exponential, and other functions

Mary Parker, Austin Community College
Hunter Ellinger, Exemplar Technologies, Inc.

<http://www.austincc.edu/mparker/modeling/>

Includes links to the materials themselves, which are available for others to use and modify.

WHY INCLUDE MODELING IN A COURSE?

[Instructor's edition] Why is modeling an appropriate topic for general education?

Modeling provides an accessible way of connecting the regularities in natural situations to mathematical formulas.

Modeling makes use of people's natural pattern-recognition skills in choosing formulas (including combinations of basic models), detecting outliers, and distinguishing between noise and structural deviations.

By delegating the numerical computations and iterative search to the "transparent box" of a spreadsheet, modeling emphasizes the need for strategic thinking by students in setting up the process and in assessing its results.

Showing the extent to which underlying patterns inherent in natural datasets can be extracted by modeling techniques gives students a concrete example of abstraction. This is particularly effective when the modeling formulas used are expressed in "natural" parameters that match the way students think about the situation that produced the data.

[Student's edition] What is modeling good for?

Modeling lets you quickly find the numerical values that measurement data imply for the parameters reflecting how you think about a situation.

Modeling also provides information about the amount of random noise in a set of data, and about how well the predictions of any specified kind of mathematical formula can match that data.

EDUCATIONAL GOALS FOR MODELING IN *MATHEMATICS FOR MEASUREMENT*

Teaching about modeling as well as about mathematics

MFM modeling is presented as a "transparent box", in which the mechanisms of parameterized formulas, goodness-of-fit indicators, and iterative fitting are all openly reflected in spreadsheets that are first used, and then constructed, by students.

The basic mechanism for transparency is the use of a spreadsheet to contain the data and the model. Cells in the spreadsheet are designated for each parameter (e.g., intercept and slope for a linear model), and references to these cells are used in making model predictions for each data row. Changes to the parameter cells thus change all the predictions, and consequently the deviations computed for each data row as the difference between the data and the prediction.

The openness extends to discussion of what indicator is appropriate to choose a best-fit model. The standard approach of minimizing the sum of squared deviations is used in most cases, but students are shown how alternative strategies (e.g., minimizing the maximum deviation or basing the indicator on the economic costs of deviations) can be implemented when such goals for the fitting process would better address the needs of a situation.

Using natural parameters for mathematical modeling formulas

The format in which standard mathematical formulas are presented is usually designed for compactness and generality, not for accessibility to people with limited mathematical backgrounds.

Alternate representations (such as the vertex form of a quadratic equation) are often preferable in instructional and practical situations because they make the connection between number and meaning obvious. *Mathematics for Measurement* uses such “natural” parameters wherever it can.

This strategy is particularly fruitful in modeling because it is then often the case that the answers to the questions posed by the problem are simply parameter values (e.g., “When was the ball at its highest point? How high?” when data is fit with a vertex-form quadratic). This encourages students to state their models in terms of what they want to fit, often making post-fitting algebraic manipulation unneeded.

INSTRUCTIONAL SEQUENCE OF M.F.M. MODELING TOPICS

The course includes some review, modeling, error propagation from calculations with approximate numbers, and applied trigonometry using both right triangles and general triangles. Each of the three topics is covered in about one-third of the semester. The list below only includes the modeling topics.

Preparatory skills

- Algebra: linear and proportional equations; evaluating a formula using parentheses appropriately in the calculator/spreadsheet; graphing a formula on a particular domain by point-plotting, including choosing appropriate scales.
- Spreadsheet use
 - Introduction to spreadsheet formulas
 - creating number patterns (to use as x variables)
 - formula evaluation for a variety of formula types
 - graphs of formulas
 - Parametric formulas
 - Standard spreadsheet functions (SUM, AVERAGE, etc.)
- Linear formulas: writing the equation of a line through two points; solving word problems leading to exact linear formulas, which includes defining variables, deciding which is output variable, interpreting parameter values, redefining the input variable (e.g. “years since 1990”), making predictions using the formula, and backward calculations of ‘find the x that gives $y = k$ ’ using the formula.
- Dealing with measurement data sets
 - Basic statistics for repeated measurements: mean and standard deviation
 - Bias and calibration
 - Graphing measurement data
 - awareness of the effects of automatic scaling
 - choosing the orientation appropriate to what you intend to predict

Modeling with spreadsheets

- Linear and quadratic models, fit by hand with modeling templates
- These include redefining the input value as needed, using the formulas to make predictions, the graphs and spreadsheet values to do backwards calculations of ‘find the x that gives $y = k$ ’, and interpreting the parameters.
- Natural parameters for models are used to facilitate estimation and interpretations.
 - convenient starting/ending points (e.g., linear intercept at initial data year)
 - vertex form for quadratic models, rather than polynomial form
- Discuss that the linear formula is a constant amount of change; mention that quadratic formula is a constant acceleration
- Method for systematic improvement of parameter estimates

- Exponential models, fit by hand with modeling templates
 - includes all the same ideas as used for the linear and quadratic models. The natural parameters are from the growth-rate form rather than the exponential-coefficient form.
- Discuss that the exponential formula is derived from a constant percentage change
- Revise the model instead of re-defining the variables. (Use “x-1780” in the model instead of revising the x-variable to be “years since 1780.”)

- Automated fitting for any kind of model formula
 - adding a goodness-of-fit indicator, which is primarily the sum of squared deviations
 - using the spreadsheet’s “Solver” capability to find the parameter values that minimize the sum of the squared deviations. (Students are expected to get a somewhat reasonable set of initial values for the parameters before using Solver.)
 - discuss how it is also optimizing on standard deviation (using the correct degrees of freedom to allow for the models having different numbers of parameters.)
- Comparison of quality of fits from best-fit models of different kinds (which needs standard deviation rather than sum of squares.) Compare by looking at the graphs, by comparing standard deviations, and by investigating residual deviations.
- Discuss how two different models (such as linear and exponential) may both do well in interpolation, but give very different results in extrapolation.
- **Students now make modeling spreadsheets for new formulas from blank worksheets**, using Solver on the sum of the squared deviations. Some formulas are $y = \frac{1}{ax+b}$ and $y = a \cdot b^x$.
Initial parameter values are given because Solver may be sensitive to initial values.
- Extensions are briefly investigated
 - recognizing outliers, and removing them from the fitting process by simply zeroing the sum of squared deviations value.
 - alternative criteria for “best fit” – maximum deviation, relative standard deviation

The following is material we have prepared but haven’t used yet because of time constraints.

Advanced formulas

- Logistic model. Parameters: baseline, height, transition, slope at transition
- Normal density curve. Parameters: center (average), width (standard deviation), area
- Sinusoidal model. Parameters: wavelength, amplitude, phase, average
- Discussion of applications of each formula, which includes the role of each of the natural parameters in the descriptions of the situations.

Semi-log graphs and log-log graphs (as a graphing topic, not really a modeling topic.)

Advanced modeling techniques

- Combination of models by addition (with warnings about parameter confounding)
 - exponential plus baseline (e.g., cooling to unknown room temperature)
 - linear plus sinusoidal modulation (e.g., to assess daily temperature effects)
 - sum of two normal curves (e.g., to extract parameters for unresolved populations)
- Combination of models by composition
 - explicit range definition with IF function (e.g., to find step-function transition)
 - implicit range definition with MAX or MIN (e.g., for multiple-constraint process)
- Redefinition of goodness-of-fit indicator to reflect situation-specific economic costs