Rowan University

Rowan Digital Works

Faculty Scholarship for the College of Science & Mathematics

College of Science & Mathematics

9-19-2016

Representing uncertainty on model analysis plots

Trevor I. Smith
Rowan University, smithtr@rowan.edu

Follow this and additional works at: https://rdw.rowan.edu/csm_facpub

Part of the Physics Commons

Let us know how access to this document benefits you - share your thoughts on our feedback form.

Representing uncertainty on model analysis plots

Trevor I. Smith[*]

Department of Physics and Astronomy and Department of STEAM Education, Rowan University,
Glassboro, New Jersey 08028, USA

Model analysis provides a mechanism for representing student learning as measured by standard multiple-choice surveys. The model plot contains information regarding both how likely students in a particular class are to choose the correct answer and how likely they are to choose an answer consistent with a well-documented conceptual model. Unfortunately, Bao's original presentation of the model plot did not include a way to represent uncertainty in these measurements. I present details of a method to add error bars to model plots by expanding the work of Sommer and Lindell. I also provide a template for generating model plots with error bars.

I. INTRODUCTION

Model analysis is a powerful tool for representing student learning in terms of both increases in the use of correct models and decreases in the use of incorrect models. Bao and Redish introduced model analysis as a complement to typical representations of learning gains that focus on student correctness [1]. The model plot simultaneously shows how much a class's use of the correct model increases and how much their use of a well-defined incorrect model decreases (or vice versa). Student use of these models are often measured by a mutiple-choice survey, such as the Force Concept Inventory (FCI) [2] or the Force and Motion Conceptual Evaluation (FMCE) [3]. Smith, Wittmann, and Carter have used model analysis in conjunction with statistical analyses of students' normalized gains to compare the effects of various instructional strategies on student learning at several colleges and universities [4]. Rakkapao, Pengpan, Srikeaw, and Prasitpong also report on the benefits of using model analysis to represent the rich variety of data that come from comparing instructional methods, including cases in which student use of both the correct and common incorrect model increase [5].

Smith, Wittmann, and Carter introduced a method for adding error bars to a model plot as a representation for experimental uncertainty [4]. In this paper I refine this process and provide additional details about the methods and assumptions for generating errors bars. I also provide templates for generating model plots that include error bars using either MATHEMATICA or the R software environment.

[*]smithtr@rowan.edu

II. DENSITY MATRICES AND THE MODEL PLOT

The main goal of model analysis is to use response frequencies to determine the probabilities of students in a particular class using each well-defined model. One step is to create a density matrix to represent a class's knowledge state at a given time [1]. The class density matrix D is the sum of students' individual density matrices, each of which is determined by the measured frequencies of each student using each of the models.

$$D = \sum_{i=1}^{N} \begin{pmatrix} p_{1,i} & \sqrt{p_{1,i}p_{2,i}} & \cdots & \sqrt{p_{1,i}p_{n,i}} \\ \sqrt{p_{2,i}p_{1,i}} & p_{2,i} & \cdots & \sqrt{p_{2,i}p_{n,i}} \\ \vdots & & \ddots & \vdots \\ \sqrt{p_{n,i}p_{1,i}} & \sqrt{p_{n,i}p_{2,i}} & \cdots & p_{n,i} \end{pmatrix}, \quad (1)$$

where $p_{j,i}$ is the probability that the ith student uses the jth model to answer a particular question. Typically model 1 is the correct Newtonian model (e.g., net force is proportional
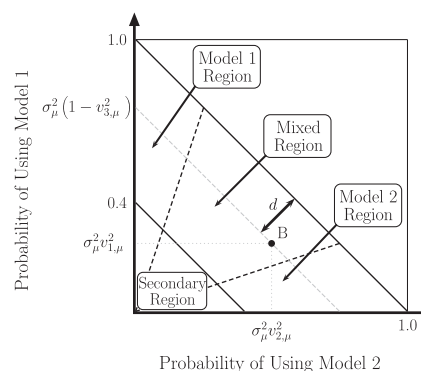

FIG. 1. Various regions of a model plot with two well-documented mental models. Figure recreated from Ref. [1] and originally published in Ref. [4], where σ_μ^2 is the μth eigenvalue of the class density matrix, and $v_{k,\mu}$ is the kth component of the μth eigenvector.

to acceleration), models 2 through $(n-1)$ are associated with well-documented incorrect models (e.g., net force is proportional to velocity), and model n is the catch all for any "other incorrect" responses.

The eigenvalues and eigenvectors of D are used to characterize the class's knowledge state [1]. The primary eigenvalue and the components of its associated eigenvector are used to create a single point on a model plot representing a class's probability of using each model at a given time. Figure 1 shows a sample model plot for a physical situation with two well-defined models [1,4].

III. UNCERTAINTY AND THE NEED FOR ERROR BARS

One shortcoming of model analysis and the model plot is that statistical uncertainty is not represented in the results. Is data point "B" in Fig. 1 really in the mixed region, or could it be in the model 2 region? It is impossible to have confidence in the interpretation of a class's "state" without error bars on the model plot.

A. Uncertainty in eigenvalues

Sommer and Lindell recognized the omission of measures of statistical power and proposed a method for determining the uncertainty in the eigenvalues of the class density matrix [6]. Their method considers that the measured probability that a student uses a particular model has an associated uncertainty $\epsilon_{j,i}$ that may be positive or negative ($-1 \leq \epsilon_{j,i} \leq 1$). The real probability may be as high (or low) as $p_{j,i} + \epsilon_{j,i}$. This results in a single-student error matrix, e_i,

$$e_i = \begin{pmatrix} \epsilon_{1,i} & e_{12,i} & \cdots & e_{1n,i} \\ e_{21,i} & \epsilon_{2,i} & \cdots & e_{2n,i} \\ \vdots & & \ddots & \vdots \\ e_{n1,i} & e_{n2,i} & \cdots & \epsilon_{n,i} \end{pmatrix}, \quad (2)$$

where

$$e_{k\ell,i} \equiv \sqrt{(p_{k,i} + \epsilon_{k,i})(p_{\ell,i} + \epsilon_{\ell,i})} - \sqrt{p_{k,i}p_{\ell,i}}. \quad (3)$$

Unfortunately, the uncertainty of a single student choosing a particular model is typically not knowable from data sets of pre- and post-test surveys. Therefore, Sommer and Lindell assume that the error matrix for the class will have the same form as that of Eq. (2):

$$E = \begin{pmatrix} \epsilon_1 & E_{12} & \cdots & E_{1n} \\ E_{21} & \epsilon_2 & \cdots & E_{2n} \\ \vdots & & \ddots & \vdots \\ E_{n1} & E_{n2} & \cdots & \epsilon_n \end{pmatrix}, \quad (4)$$

$$E_{k\ell} \equiv \sqrt{(D_{kk} + \epsilon_k)(D_{\ell\ell} + \epsilon_\ell)} - \sqrt{D_{kk}D_{\ell\ell}}, \quad (5)$$

where D_{kk} is one of the diagonal elements of the class density matrix.

Given that the error in the measured probability could be positive or negative, each term in E could also be either positive or negative. This information is used to generate a set of specific error matrices. Because the $(n \times n)$ density matrix is symmetric, the general error matrix is also symmetric, yielding $2^{n(n+1)/2}$ specific matrices with different combinations of positive and negative terms. By adding each of these specific error matrices to the class density matrix D and computing the eigenvalues of each of the resulting adjusted density matrices, one can determine the upper and lower bounds for each of the eigenvalues [6]. One may now be confident that the actual eigenvalue falls within the range $[\sigma_{\mu,\text{min}}^2, \sigma_{\mu,\text{max}}^2]$.

While this is a step in the right direction, it falls short of providing a mechanism for representing statistical uncertainty within the model plot (the points on which depend on both eigenvalues and the associated eigenvectors). Moreover, this method requires an initial assumption of the values of the uncertainties, ϵ_i, that are used to create the general error matrix.

B. Creating error bars on the model plot

To create error bars on the model plot one must translate the uncertainty associated with the eigenvalue of the density matrix to an uncertainty in each dimension of the model plot. As shown in Fig. 1, the horizontal coordinate (x) corresponds with the probability of choosing model 2 and is defined as the product of the primary eigenvalue σ_μ^2 with the square of the second component of the associated eigenvector $v_{2,\mu}^2$. Similarly, the vertical coordinate (y) is associated with model 1 and the first component of the eigenvector:

$$x = \sigma_\mu^2 v_{2,\mu}^2, \quad (6)$$

$$y = \sigma_\mu^2 v_{1,\mu}^2. \quad (7)$$

The uncertainty in each of these coordinates is determined by the uncertainty in the primary eigenvalue and the components of its associated eigenvector, but this relationship is not necessarily straightforward.

Smith, Wittmann, and Carter assumed that the fractional uncertainty in the primary eigenvalue will be the same as the fractional uncertainty of each coordinate [4],

$$\frac{\Delta_\mu}{\sigma_\mu^2} = \frac{\Delta_x}{x} = \frac{\Delta_y}{y}, \quad (8)$$

where Δ_μ is defined by the upper and lower bounds:

$$\Delta_\mu = \frac{(\sigma_{\mu,\max}^2 - \sigma_\mu^2) + (\sigma_\mu^2 - \sigma_{\mu,\min}^2)}{2}. \qquad (9)$$

The actual coordinates will fall within the ranges $x \pm \Delta_x$ and $y \pm \Delta_y$. However, this assumption causes the error bars to be proportional to the value of the coordinate, which may not accurately reflect the uncertainty in the class's model state, e.g., $\Delta_x = x(\Delta_\mu/\sigma_\mu^2)$.

A less restrictive method is to determine the coordinates (x_k, y_k) for each of the adjusted density matrices by calculating the eigenvalues and eigenvectors of each. The uncertainty represented by the error bars would then be $[x_{\min}, x_{\max}]$ and $[y_{\min}, y_{\max}]$. I provide an example in Sec. IV that shows the results of each assumption.

C. Choosing an initial estimate of uncertainty

Sommer and Lindell propose using a single uncertainty for simplicity ($\epsilon = \max\{\epsilon_1, \epsilon_2, ..., \epsilon_n\}$), but provide no straightforward method for determining an initial estimate [6]. There are several options for choosing an initial estimate of the uncertainty based on the pre- and post-test data. The choices I present are based on the standard error of a particular data set.

One of the simplest choices is to assume that the uncertainty will be the same for all models and will be the same for both pre- and post-test data. To accomplish this I use the pooled standard error in terms of the standard deviations of both the pre- and post-test data sets:

$$\epsilon = \sqrt{\left(\frac{(N_{pre} - 1)s_{pre}^2 + (N_{post} - 1)s_{post}^2}{N_{pre} + N_{post} - 2}\right)\left(\frac{1}{N_{pre}} + \frac{1}{N_{post}}\right)}, \qquad (10)$$

where N is the number of students and s is the standard deviation of the number of correct answers for each data set [7].

For cases in which one uncertainty may not fit the data one may choose to calculate the standard error for the pretest and post-test separately:

$$\epsilon_{pre} = \frac{s_{pre}}{\sqrt{N_{pre}}}, \qquad (11)$$

$$\epsilon_{post} = \frac{s_{post}}{\sqrt{N_{post}}}. \qquad (12)$$

Additionally, one may choose not to accept the assumption proposed by Sommer and Lindell that the uncertainty is the same for all models. Equations (10)–(12) may all be applied to the data sets of students using each of the models. In the following section I provide examples for a single value of uncertainty for all models pre- and postinstruction (the most restrictive assumption)

and different uncertainties for all models (the least restrictive assumption).

IV. EXAMPLE OF GENERATING ERROR BARS

I present an example to illustrate the process of creating error bars on the model plot and to examine the implications of each of the assumptions for choosing an initial error estimate. For brevity I only present pretest and post-test data for a single class in one year, answering questions within a single question cluster [8]. In this question cluster there are two well-defined models (correct and common incorrect) and one "other" model. The pre- and post-test class density matrices are

$$D_{pre} = \begin{pmatrix} 0.222 & 0.162 & 0.123 \\ 0.162 & 0.500 & 0.270 \\ 0.123 & 0.270 & 0.278 \end{pmatrix}, \qquad (13)$$

$$D_{post} = \begin{pmatrix} 0.559 & 0.089 & 0.140 \\ 0.089 & 0.218 & 0.090 \\ 0.140 & 0.090 & 0.224 \end{pmatrix}. \qquad (14)$$

These density matrices have eigenvalues and associated eigenvectors that yield coordinates on the model plot:

$$\sigma_{pre}^2 = 0.758, \qquad \mathbf{v}_{pre} = \begin{pmatrix} 0.355 \\ 0.773 \\ 0.525 \end{pmatrix}, \qquad (15)$$

$$x_{pre} = 0.453, \qquad y_{pre} = 0.095, \qquad (16)$$

$$\sigma_{post}^2 = 0.641, \qquad \mathbf{v}_{post} = \begin{pmatrix} 0.896 \\ 0.265 \\ 0.357 \end{pmatrix}, \qquad (17)$$

$$x_{post} = 0.045, \qquad y_{post} = 0.514. \qquad (18)$$

As seen in Fig. 2(a) this class starts in the model 2 region and moves to the model 1 region. In the following sections I examine the implications of each assumption: equal fractional uncertainties vs coordinate-specific uncertainties, and equal initial error estimates vs model-specific errors.

A. Assuming equal fractional uncertainty and equal error estimates

As a starting point I assume that the fractional uncertainties for both coordinates are the same as that of the primary eigenvalue and that a single estimate of uncertainty will suffice for all models for both the pre- and post-test data [9]. The statistics for each data set are contained in Table I. The pooled standard error for the correct responses

is $\epsilon = 0.0470$, which is higher than either of the individual standard errors for pre- or postdata. This provides general error matrices of

$$E_{\text{pre}} = \begin{pmatrix} \pm 0.0470 & \pm 0.0504 & \pm 0.0472 \\ \pm 0.0504 & \pm 0.0470 & \pm 0.0488 \\ \pm 0.0472 & \pm 0.0488 & \pm 0.0470 \end{pmatrix}, \quad (19)$$

$$E_{\text{post}} = \begin{pmatrix} \pm 0.0470 & \pm 0.0517 & \pm 0.0514 \\ \pm 0.0517 & \pm 0.0470 & \pm 0.0470 \\ \pm 0.0514 & \pm 0.0470 & \pm 0.0470 \end{pmatrix}, \quad (20)$$

where the "$\pm$" indicate that each cell could be positive or negative.

Using every combination of positive and negative cells yields 64 adjusted density matrices. For the pretest the eigenvalues span the range [0.629, 0.892], and on the posttest they span the range [0.540, 0.762]. This gives general uncertainties of $\Delta_{\mu,\text{pre}} = 0.131$ and $\Delta_{\mu,\text{post}} = 0.111$. Assuming equal fractional uncertainties provides an uncertainty for each coordinate:

$$\Delta_{x,\text{pre}} = 0.0786, \qquad \Delta_{y,\text{pre}} = 0.0166, \qquad (21)$$

$$\Delta_{x,\text{post}} = 0.0078, \qquad \Delta_{y,\text{post}} = 0.0893. \qquad (22)$$

As can be seen in Fig. 2(b) the point in the model 2 region has a much larger uncertainty in the horizontal coordinate and vice versa. This is a direct result of the assumption of equal fractional uncertainties for which the uncertainty in a coordinate is proportional to the value of that coordinate.

B. Assuming coordinate-specific uncertainty and model-specific error estimates

I now present the results of rejecting both the assumption of equal fractional uncertainties and the assumption of equal initial error estimates. Using the standard error for each model in each data set (see Table I) provides new general error matrices:

$$E_{\text{pre}} = \begin{pmatrix} \pm 0.0253 & \pm 0.0276 & \pm 0.0231 \\ \pm 0.0276 & \pm 0.0263 & \pm 0.0233 \\ \pm 0.0231 & \pm 0.0233 & \pm 0.0202 \end{pmatrix}, \quad (23)$$

$$E_{\text{post}} = \begin{pmatrix} \pm 0.0422 & \pm 0.0399 & \pm 0.0368 \\ \pm 0.0399 & \pm 0.0337 & \pm 0.0318 \\ \pm 0.0368 & \pm 0.0318 & \pm 0.0299 \end{pmatrix}. \quad (24)$$

One may use these to calculate x and y coordinates for each of the $2^{n(n+1)/2}$ adjusted density matrices.

$$x_{\text{pre}} \in [0.399, 0.502], \qquad (25)$$

$$y_{\text{pre}} \in [0.060, 0.139], \qquad (26)$$

$$x_{\text{post}} \in [0.012, 0.104], \qquad (27)$$

$$y_{\text{post}} \in [0.437, 0.585]. \qquad (28)$$

As shown in Fig. 2(c) this gives a much different representation of the uncertainty in the coordinates on the model plot. This provides a more accurate representation of the span of the model space each point could occupy given that these uncertainties do not depend directly on the values of the coordinates themselves. These error bars also
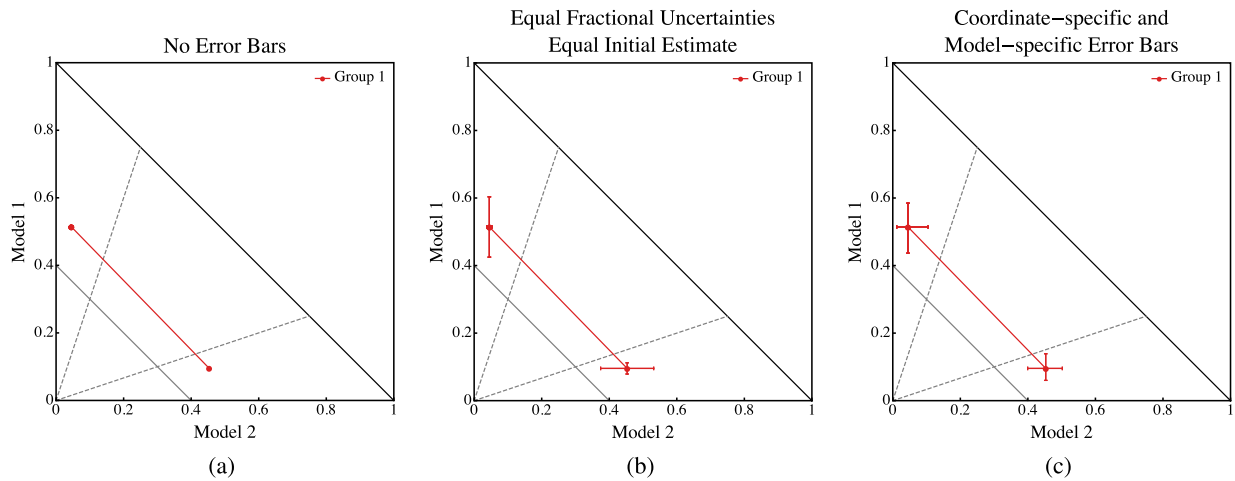

FIG. 2. Model plots for the example data: (a) no error bars, (b) assuming equal fractional uncertainties for both coordinates and equal initial error estimates for all models (pooled standard error), and (c) assuming coordinate-specific uncertainties and model-specific initial error estimates (standard error of the distribution for each model).

TABLE I. Statistics regarding the number of students (N), average values, standard deviations (s), and standard errors (ϵ) of Model 1 and Model 2 scores for both the pre- and post-test data.

	N	Model 1: Correct			Model 2: Common		
		Average	s_{M1}	ϵ_{M1}	Average	s_{M2}	ϵ_{M2}
pre	109	0.222	0.264	0.0253	0.500	0.274	0.0263
post	85	0.559	0.389	0.0422	0.218	0.311	0.0337
pooled				0.0470			0.0421

show that the uncertainty in the post-test data is greater than in the pretest data (see Table I).

V. SUMMARY

The assumptions involved in calculating error bars can have a dramatic effect on the interpretations reflected in the model plot. The sample data above show that using the standard error of the data yields error bars with a non-negligible extent on the model plot. This is most visible in Fig. 2(c) where the model-specific error estimates and coordinate-specific uncertainties provide macroscopic error bars in both coordinates for both points. These results support the notion that it is imperative to represent the uncertainty in the coordinates on the model plot in some fashion. The class model state is not precisely known as would be implied otherwise.

In an effort to facilitate the use of this method I have included several template files in the Supplemental Material that may be used to generate density matrices and model plots with error bars [10]. The EXCEL file (`MA_FMCE_template.xlsx`) generates class density matrices from student responses to the FMCE [11]. This file may be modified in order to create density matrices for any other multiple choice data with well-defined models. The text files include the commands for importing data from EXCEL, performing the necessary matrix calculations, and generating model plots with error bars using the open-source R software environment [12]. The file `MA_3Model_templateR.txt` assumes a question cluster with three models (two well-defined models and one "other" incorrect, as is the case with the above example), and the file `MA_4Model_templateR.txt` assumes a question cluster with four models [13]. I have also included MATHEMATICA template files (`MA_3Model_template.nb` and `MA_4Model_template.nb`) that perform the same functions as the R files but are not as flexible. All files include instructions for creating model plots with error bars starting from a class set of multiple-choice survey data.

[1] L. Bao and E. F. Redish, Model analysis: Representing and assessing the dynamics of student learning, Phys. Rev. ST Phys. Educ. Res. **2**, 010103 (2006).

[2] D. Hestenes, M. Wells, and G. Swackhamer, Force concept inventory, Phys. Teach. **30**, 141 (1992).

[3] R. K. Thornton and D. R. Sokoloff, Assessing student learning of Newton's laws: The force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula, Am. J. Phys. **66**, 338 (1998).

[4] T. I. Smith, M. C. Wittmann, and T. Carter, Applying model analysis to a resource-based analysis of the force and motion conceptual evaluation, Phys. Rev. ST Phys. Educ. Res. **10**, 020102 (2014).

[5] S. Rakkapao, T. Pengpan, S. Srikeaw, and S. Prasitpong, Evaluation of POE and instructor-led problem-solving approaches integrated into force and motion lecture classes using a model analysis technique, Eur. J. Phys. **35**, 015016 (2014).

[6] S. R. Sommer and R. S. Lindell, Accounting for errors in model analysis theory: A numerical approach, AIP Conf. Proc. **720**, 49 (2004).

[7] Equation (10) assumes that student scores are recorded as a percentage correct. If instead scores are recorded directly, then the entire expression should be divided by the number of questions in the particular cluster.

[8] These data were originally published in Ref. [4]: School 1, Year 2, Energy cluster.

[9] These are the same assumptions that were used to generate the model plots in Ref. [4].

[10] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevPhysEducRes.12.023102 for template files for performing model analysis and generating model plots with coordinate-specific and model-specific error bars.

[11] The EXCEL template also works in OPENOFFICE.

[12] EXCEL files must have the `.xlsx` extension to be imported into R.

[13] The two additional files `MA_3Model_templateR-ods.txt` and `MA_4Model_templateR-ods.txt` are designed to import data from OPENOFFICE (`.ods`) files rather than EXCEL files.