J. Dairy Sci. 99:1–21
http://dx.doi.org/10.3168/jds.2015-10032

Comparison of model predictions with measurements: A novel model-assessment method

N. R. St-Pierre[1]
Department of Animal Sciences, The Ohio State University, Columbus 43210

ABSTRACT

Frequently, scientific findings are aggregated using mathematical models. Because models are simplifications of the complex reality, it is necessary to assess whether they capture the relevant features of reality for a given application. An ideal assessment method should (1) account for the stochastic nature of observations and model predictions, (2) set a correct null hypothesis, (3) treat model predictions and observations interchangeably, and (4) provide quantitatively interpretable statistics relative to precision and accuracy. Current assessment methods show deficiencies in regards to at least one of these characteristics. The method being proposed is based on linear structural relationships. Unlike ordinary least-squares, where the projections from the observations to the regression line are parallel to the y-axis and inverse regression where they are parallel to the x-axis, the generalized projection regression method (GePReM) projects the observations on a regression line in a direction determined by the ratio of the precision of the observations to that of the mathematical model predictions. Estimation and testing issues arise when the model is expressed in the common slope-intercept format. A polar transformation circumvents these issues. The parameter for the angle between the regression line and the horizontal axis has symmetrical confidence intervals and is equivariant to the exchange of X and Y. The null hypothesis for the equivalence test is that the model predictions are not equivalent to the observations. Information size is calculated as the simple ratio of the variance of the true values of the observations and of the computer model predictions divided by their respective precision. This information size plays a critical role in determining the number of observations required and the size of the zone of practical tolerance for the equivalence tests. The terminology used in the comparison of measurement methods is adapted to that of model assessment based on the equivalence tests on the relative precision, regression slope, and mean bias. Two examples are presented, with complete details of the calculations required for parameter estimation, equivalence tests, and confidence intervals. The assessment method proposed is an alternative to other assessment methods available. Further research is required to establish the relative benefits and performance of this proposed method compared with others available in the literature.
Key words: model assessment, model validation, generalized projection regression method (GePReM)

INTRODUCTION

Frequently, science leads to hypotheses and theories that are best expressed using the language of mathematics. The mathematics involved can be as simple as a single function or much more complex, resulting in what are known as mathematical models. Such models can take many forms and be classified as dynamic or static, mechanistic or empirical, deterministic or stochastic (Thornley and France, 2007). Because models are abstractions and simplifications of the much more complex reality, they cannot fully characterize reality in its most intricate details. This leads to the inevitable need to assess the adequacy of a given model in representing sufficiently well the features of the real world relevant to a defined task or set of objectives. This is the essence of model assessment.

In mathematical modeling work, the model is often constructed and parameterized using domain expertise and small data sets. Eventually, external research data (i.e., data not used in model identification and parameterization) become available. These data are then used to assess the model's properties. This situation is quite different from the traditional statistical one, where the same data are used for model identification, parameterization, and model assessment.

Many methods of model assessment have been proposed and most were recently reviewed by Tedeschi (2006). In general, methods fall into one of the following categories: linear regression (Mayer et al., 1994), including orthogonal regression (Warton et al., 2006)

Received June 30, 2015.
Accepted February 21, 2016.
[1]Corresponding author: st-pierre.8@osu.edu

and modified regression (St-Pierre, 2003); analyses of deviations (Mitchell, 1997); analyses of residuals (Draper and Smith, 1988); concordance correlation coefficient (Lin, 1989); mean square error of prediction (MSEP; Bibby and Toutenburg, 1977), partitioning of MSEP into error in central tendency (i.e., mean bias), errors due to regression (i.e., linear bias), and errors due to disturbances (or random errors; Theil, 1961). All of these methods of model assessment suffer from one or more deficiencies in that they either set an incorrect model, test an incorrect hypothesis, provide metrics that are not easily interpretable, or fail to answer the right question. In addition, a useful model assessment method should provide, a priori, the characteristics of the data necessary to a useful model assessment, something akin to an a priori power determination before conducting an experiment.

The objectives of this paper are (1) to identify the most important characteristics of an ideal model assessment method, (2) to present a novel method of model assessment that meets all these characteristics, and (3) to show its application using 2 examples, the first consisting of DMI predictions in growing dairy goats (NRC, 2007), and the second dealing with predictions of microbial N flow to the duodenum of dairy cows (NRC, 2001). The new method, the generalized projection regression method (**GePReM**), will be presented without any mathematical proofs. The GePReM sets a statistical model, whereas the assessment process is for a mathematical model. The statistical model yields predictions and so does the mathematical model. To avoid confusion between the 2 models, we will refer to the mathematical model, the one being assessed, as “the computer model” in the balance of this paper, and its predictions as “the computer model predictions,” although it should be clear that a mathematical model does not necessarily require a computer to yield predictions.

DESIRABLE CHARACTERISTICS OF AN IDEAL ASSESSMENT METHOD

An ideal computer model-assessment method should exhibit many desirable features (Tedeschi, 2006). Among all the desirable characteristics, arguably the most important ones can be stated as follows.

Accounting for the Stochastic Nature of Observations and Predictions

All measurements and computer models have inherent uncertainty (i.e., errors). Often the uncertainty in the predictions is not explicitly acknowledged by the model developers and is not incorporated in the computerized form of the model, but overlooking uncertainty and errors does not negate their existence.

Simple computer models can mathematically be represented by the following set of undefined functions:

$$\mathbf{Y} = \mathbf{f}(\mathbf{X}, \mathbf{B}) + \mathbf{e}, \qquad [1]$$

where $\mathbf{Y}$ is a vector of n observations, $\mathbf{f}$ is a set of undefined functions, $\mathbf{X}$ is an $n \times p$ matrix of input variables, $\mathbf{B}$ is a vector of parameter estimates, and $\mathbf{e}$ is an n vector of residual errors. In this notation, stochasticity enters the computer model in many ways. First, the values of the input variables $\mathbf{X}$ are seldom known with certainty. For example, the weight of an animal when used to estimate DMI is not perfectly known. Second, the vector $\mathbf{B}$ refers to estimates of the true parameters $\boldsymbol{\beta}$, which themselves are seldom (if ever) known. By definition, the statistical estimation of parameters implies uncertainty represented by a matrix of variances and covariances of the estimated values. Third, the functional forms in $\mathbf{f}$ are rarely known with certainty. Sometimes they can be based on prevailing theories (e.g., Michaelis-Menten kinetics); many times, they are chosen among a set of candidate functions based on best-fit statistics. Hence, there is generally uncertainty regarding the specific functional forms that were chosen. Last, the residual errors cannot be ignored post-estimation. This error (uncertainty) would remain even in a perfect world, where $\mathbf{f}$, $\mathbf{X}$, and $\mathbf{B}$ would be errorless. In short, all computer model predictions are truly stochastic. Estimating prediction errors from computer models is not trivial (Marino et al., 2008). Analytical solutions are seldom available, but numerical methods such as Monte Carlo methods can generally be used quite successfully (e.g., St-Pierre and Thraen, 1999).

As for observations, their errors are generally intuitive and have been recognized in most, if not all, computer model-assessment methods.

Setting a Correct Null Hypothesis

If the comparison involves a set of parameters θ, the significance test should not be based on the conventional set of hypotheses:

$$\mathrm{H}_0\text{: } \theta = \theta_0 \text{ versus } \mathrm{H}_1\text{: } \theta \neq \theta_0. \qquad [2]$$

That is, the computer model predictions should not be deemed equal to the observations unless there is enough evidence to the contrary. Instead, hypothesis tests should be set as in equivalence studies: