

Validating mathematical models of biological systems: a concordance correlation coefficient that incorporate bias, scale shift and accuracy of measurements versus model predictions. N. R. St-Pierre. The Ohio State University, Columbus, OH-43210.

Mathematical models are now frequently used to quantify complex biological systems. The validation of such models is done by comparing model predictions to observed data. Various statistical methods have been used to assess a model's validity: the Pearson correlation coefficient, the paired t-test, the least-square analysis of slope ($=1$) and intercept ($=0$), and the coefficient of variation or the intraclass correlation coefficient. None of these can completely assess the desired reproducibility characteristics. The Pearson correlation coefficient only measures precision of a linear relationship, not accuracy. Both the paired t-test and least squares analysis can falsely reject (accept) the hypothesis of high agreement when the residual error is small (large). The coefficient of variation and the intraclass correlation coefficient assume a dependent and an independent variable. More importantly, they fail to recognize the duality (interchangeability) of predictions with observations. Both are mathematical transforms of measurements. Both have random errors from measurements and parameter estimates. And both have structural errors due to the simplification of the complex real world. The relevant question is not whether a model predicts observed data but whether the model and the observation method measure the same thing, whether the methods agree and how good is the agreement. This requires a joint assessment of precision and accuracy.

The Committee on Animal Nutrition of the National Research Council (NRC) is charged with producing tables of nutrient requirements of various classes of animals. Nutrient requirements are expressed in the form of computerized mathematical models. In a recent publication, the NRC (2001) produced a new model for estimating the nutritional requirements of dairy cattle. A key step in the calculation of protein and amino acid requirements is the estimation of the amount of bacterial protein synthesized in the rumen. In ruminants, the net supply of protein and amino acids is derived from two separate sources: a variable portion of the feed protein not broken down by the ruminal micro-flora passes to the duodenum (small intestine) where it can be digested and absorbed by the animal. The second portion consists of microbial protein synthesized by the ruminal micro-flora using carbon skeletons, ATP, ammonia, amino acids, and short peptides. The quantification of the net supply from each process is very important to the optimal feeding of ruminant animals and their environment impact (N excretion). The measurements of microbial and undegraded feed protein to the duodenum must rely on surgically altered animals and inert markers. Thus, the measurements of microbial protein (MiN) and non-ammonia-non-microbial protein flows (NANMN) to the duodenum are subject to substantial errors of measurements, plus structural errors (i.e., the non-digestible markers are not perfect markers) and possibly errors in parameter estimates. The prediction of MiN is based on total digestible nutrient intake (TDN) which is a function of the (uncertain) chemical composition of the feedstuffs and their (uncertain) bio-availabilities. Thus, both observed and predicted MiN and NANMN have errors from measurements, parameter estimates, and structural forms. This situation, where predictions and observations are interchangeable is very frequent in biology. The

question is whether we can use predictions of MiN and NANMN to replace measured values when estimating nutrient requirements.

A single scaled statistic called concordance correlation coefficient (CCC) has been suggested as an omnibus statistic to jointly assess precision and accuracy. Let Y_1 be the observed values and Y_2 be the predictions. The concordance correlation coefficient $\rho^c = 1 - \{E(Y_1 - Y_2)^2 / E[(Y_1 - Y_2) | Y_1, Y_2 \text{ are uncorrelated}]\} = 2 \sigma_{12} / [\sigma_1^2 + \sigma_2^2 + (\mu_1 - \mu_2)^2] = \rho_{12} \chi_{12}$, where $\mu_1 = E(Y_1)$, $\mu_2 = E(Y_2)$, $\sigma_1^2 = \text{Var}(Y_1)$, $\sigma_2^2 = \text{Var}(Y_2)$, and $\sigma_{12} = \text{Cov}(Y_1, Y_2) = \sigma_1 \sigma_2 \rho_{12}$. The CCC is a product of two components: precision (ρ_{12}) and accuracy (χ_{12}), where $\chi_{12} = 2 \sigma_1 \sigma_2 / [\sigma_1^2 + \sigma_2^2 + (\mu_1 - \mu_2)^2] = [(v_{12} + 1/v_{12} + u_{12}^2) / 2]^{-1}$, with $v_{12} = \sigma_1 / \sigma_2$ representing scale shift, and $u_{12} = (\mu_1 - \mu_2) / (\sigma_1 \sigma_2)^{1/2}$ representing location shift relative to the scale.

Application to a dataset of 256 measured and predicted values of MiN from 56 published studies shows that predictions and measurements are concordant ($\rho^c = 0.476$), have small scale shift (1.54) and location shift (-0.02), and are accurate (0.913) but that they lack precision (0.522). Expressed differently, the deviance (0.573) is composed of a very small bias (0.0003; or 0.05% of the deviance), a small scale shift (0.095; or 16.5% of the deviance), and a large imprecision (0.479; or 83.5% of the deviance). Thus, little gain in model precision can be expected until superior methods of measurements, with much greater precision are found.