CONCORDANCE CORRELATION FOR MODEL PERFORMANCE ASSESSMENT: AN EXAMPLE WITH REFERENCE EVAPOTRANSPIRATION OBSERVATIONS

by D. W. Meek

USDA-ARS National Soil Tilth Laboratory, 2150 Pammel Dr., Ames, IA 50011-4420

TEL: (515) 294-2246; FAX: (515) 294-8125; Email: Dave.Meek@ars.usda.gov

Introduction

Model performance assessment in agronomy and agroecology has employed many procedures in the past and continues to do so in the present. With models becoming modular and increasingly complex, Fila et al. (2003) developed the public domain software library called IRENE_DLL (Integrated Resources for Evaluating Numerical Estimates - Dynamic Link Library). This library has most of the commonly used analyses including procedures called *difference based methods* and procedures called *association based methods*. Difference based measures include methods like estimation of the root mean square error (RMSE), of the mean absolute error (MAD), and of the mean bias error (mbe) or *bias* for brevity with a t-test (Fox, 1981). Association based methods include Pearson or Spearman correlation coefficients and regression analysis.

Using either a t-test alone or a correlation coefficient alone can result in a misleading or inadequate assessment (Lin, 1989). An omnibus procedure combining the essential assessments of these two analyses exists, is well-known, and is often used in other scientific literature. Lin (1989) introduced the concordance correlation coefficient as a bivariate analysis to assess agreement between paired measurements. The analysis can be used for other similar assessments including comparing modeling predictions with observations (Lin et al., 2002). In order to promote the use of this procedure in the agronomic sciences, this work presents a brief review of the procedure and an example comparing the performance of two reference evapotranspiration models to lysimeter based measurements.

The Concordance Correlation Coefficient

To assess the relationship y=1x, Lin (1989) developed an insightful test statistic called the concordance correlation coefficient, denoted r_c. The r_c statistic is an adjusted version of the well-known Pearson product-moment correlation coefficient, r, and so can be formally evaluated in the same way. Let the y variable, here the observations, have a mean, μ_y, and standard deviation, δ_y. Let the x variable, here the model predictions, have a mean, μ_x, and standard deviation, δ_x. Then $r_c = rC_b$ where $C_b = [(v + 1/v + u^2)/2]^{-1}$. Here $v = \delta_x/\delta_y$ and is called a *scale shift* while $u = (\mu_x - \mu_y)/(\delta_x\delta_y)^{1/2}$ and is called a *location shift relative to scale* with $\mu_x - \mu_y$ is an expression for the mbe. A pure location shift could have the data scatter parallel to the 45° line (a.k.a. "1 to 1 line") through the origin. If u < 0 then the scatter is above the 45° line. In a pure scale shift the data scatter would cross the 45° line. In general both shifts are present to some degree.

The Data Set and Models

Daily totals for hourly weighing lysimeter observations and hourly model estimates from a nearby weather station are taken from the Phene et al. (1986) calibration study conducted during 1985 at Five Points, CA. Daily reference evapotranspiration data, ET0 in mm, from every other date for a total of 50 d are taken from the longer data set in the study. This selection reduces first-order serial correlation effects on the statistical analysis. The reason is because the individual series as well as the difference series have notable autocorrelations and so violate the independence criteria. The weighing lysimeter is described in Howell et al. (1985). The weather station is described in Howell et al. (1984). For the purpose of this study, the daily lysimeter observations are considered valid having no bias or any other matters of concern.

The first model considered, Model-1, is a modified Penman Equation with the empirical wind function of Doorenbos and Pruitt (1977). The second model considered, Model-2, uses the same radiative term but uses an iteratively estimated, atmospheric-stability-based, wind function (p. 218, Brutsaert, 1982).

Results

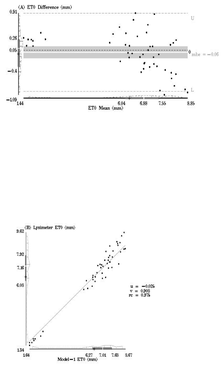
Figures 1 and 2 display results for Models 1 and 2 respectively. Table 1 allows for the comparison of selected performance measures. Both models are biased low with respect to the measures (u, mbe < 0). In fact u or mbe for Model-2 are about an order of magnitude larger than the corresponding values for Model-1. In addition, both models have systematically lower variability when compared to the lysimeter's variability (v < 1) with the v value for Model-2 concordance much lower than v for the Model-1 concordance. The probability of a Model-1 versus Model-2 r_c difference is p<0.004. Model-1 estimates are likely acceptable for most uses being in reasonable statistical agreement with observations, while Model-2 estimates are probably unacceptable.

Table 1. Evapotranspiration model performance statistics†.

	-- Correlation Measures--			---------- Difference Measures ----------			
Model‡	r	r_c	p_{diff}	u	v	mbe	RMSE
						mm	mm
Model-1	0.980	0.975	0.227	-0.025	0.908	-.055±0.069 (p≤0.428)	0.488
Model-2	0.982	0.946	0.000	0.215	0.841	.452±0.077 (p≤0.001)	0.701

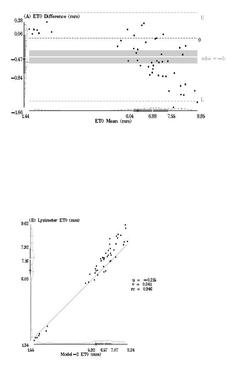
†r Pearson product-moment correlation coefficient.
r_c Concordance correlation coefficient.
p_{diff} Probability of r, r_c difference (based on associated normal scores).
u Location shift relative to the scale.
v Scale shift.
mbe Mean Bias Error ± standard error of the mbe (α probability for mbe=0).
RMSE Root Mean Square Error.
‡There are 50 observations in the data set. Model-1 has the wind function based on Doorenbos and Pruitt (1977). Model-2 has the wind function based on Brutsaert (1982).

Fig. 1



Model-1 performance results. The top panel (A) is an mean-difference (m-d) plot with box-plot axes. Both of the thin axis lines span the range of the data. The thick gray bars span the inter-quartile range with the gap at the median; the "+" is the mean. An empirical histogram for each axis appears to the right of the vertical axis for the paired differences and above the horizontal axis for the paired means. The values in the scatterplot are black dots, the gray band is the 95% confidence interval for the mean bias error (mbe), the gap in the band is the mbe, and the black dotted line is the zero reference line. The dashed gray lines labeled "U" and "L" are mbe ± $1.96\sigma_{mbe}$. Notice that the mbe is not different from 0 (p<0.05). The bottom panel (B) is a bivariate (b-v) plot with the same kind of box-plot axes as in (A) with black dots for the data, a gray 45° line, and concordance correlation statistics indicating reasonable agreement between observations and the predictions. Here both |u| and |v| < 0.1.

Fig. 2



Model-2 performance results. Although r for Model-2 is slightly larger than r for Model-1, notice in the m-d plot, panel (A), the confidence band excludes 0. In addition, in the b-v plot, panel (B), both |u| and |v| > 0.1; consequently r_c is much lower than r.

Discussion and Conclusions

Sound model development and assessment includes more than setting goals for performance measures. It should include a careful conceptual analysis tailored to the job at hand (Oreskes et al., 1994). In addition, it is not good practice to rely on either a single performance measure, let alone on the evaluation of a single data set. None-the-less, given the appropriate careful considerations, this example points out the merits of concordance correlation. While any of the difference measures alone reveal the serious bias in Model-2 estimates, they do not reveal or deal with the scale shift. An r value alone does not adjust for either location or scale shift (u and v) but the r_c does. The modeler's analytical toolbox should therefore include concordance correlation analysis for model performance assessment because it provides a simple and sound statistically based omnibus test that can also add analytical insight.

Acknowledgements

This work was supported by the USDA-ARS National Soil Tilth Laboratory, Ames, IA, Dr. J.L. Hatfield, Director. Thanks to J.W. Singer, USDA-ARS, Ames, IA and R.P. Ewing, ISU, Ames, IA for comments.

References

Brutsaert, W. 1982. Evaporation into the atmosphere: Theory, history, and applications. D. Reidel Publishing Co., Boston, MA. pp. 229.

Doorenbos, J. and W.O. Pruitt. 1977. Crop water requirements. FAO Irrigation and drainage paper No. 24. pp. 124.

Fila, G., G. Bellocchi, M. Donatelli, and M. Acutis. 2003. IRENE_DLL: A class library for evaluating numerical estimates. Agron. J.95: 1330-1333.

Fox, D.G., workshop chair, 1981. Judging air quality model performance: A summary of the AMS workshop on dispersion model performance. Bull. Am. Meteorol. Soc., 62(5): 599-609.

Howell, T.A., D.W. Meek, C.J. Phene, K.R. Davis, and R.L. McCormick. 1984. Automated weather data collection for research on irrigation scheduling. Trans. ASAE 27(2): 386-391, and 396.

Howell, T.A., R.L. McCormick, and C.J. Phene. 1985. Design and installation of large weighing lysimeters. Trans. ASAE 28(1): 106-112 and 117.

Lin, L.I. 1989. A Concordance Correlation Coefficient to Evaluate Reproducibility. Biometrics 45: 255-268.

Lin, L.I., A.S. Hedayat, B. Sinha, M. Yang. 2002. Statistical Methods in Assessing Agreement: Models, Issues, and Tools. J. Am. Stat. Assoc. 97(457): 257-270.

Oreskes, N., K. Shrader-Frechette, and K. Belitz. 1994. Verification, validation, and confirmation of numerical models in the earth sciences. Science 263: 641-646.

Phene C.J., D.W. Meek, R.L. McCormick, and K.R. Davis. 1986. Calibration and use of Penman's equation for irrigation scheduling. ASAE Paper No. 86-2594, ASAE, St. Joseph, MI. 25 p.