Reply

ALAN J. FALLER

Institute for Physical Science and Technology, University of Maryland, College Park 20742

2 March 1982

Glahn's comments are indeed appropriate and point the way to more general possibilities. The limited intent of F(7) [Eq. (7) in Faller (1981)] was a combined measure of the reduction of variance accomplished by a certain set of multiple regression equations or equivalent prediction schemes. As an afterthought, F(17) was proposed as a measure of predictability using the same set of equations in cases where J_n, the number of cases to be predicted with the nth equation, was not the same for all of the N equations.

Glahn points out that in certain common cases J_n may be proportional to I_n, the number of data sets originally used in each of the N classes. He goes further to note that if, in addition, the classification into N sets is part of the overall prediction scheme, the variance explained by the classification should be included in the measure of predictability. Then R_1^2, originally discarded by this author as irrelevant, is in fact the more appropriate measure of the overall reduction of variance. An example, illustrated by Fig. 1 and the values of Table 1, may help remove this exchange from the realm of pure symbolism.

We are concerned with 24 h predictions of a variate y, perhaps temperature, using first a numerical model and then a set of regression equations with input based upon the 24 h output of the numerical model. We would like a consistent measure of the expected overall predictability of this procedure for, say, one year, and one suitable measure is determined to be an overall reduction of variance of y.

The independent data is first classified into the four seasons, $n = 1 - N$, $N = 4$. By consensus, meteorologists do not include climatological trends in their measure of forecasting skill. It is generally the departures from climatology that count and so, in this example, the variance among classes will be omitted from our measure of predictability.

The number of data sets in each class, I_n, is different, perhaps because this particular station was a summer resort and had several years when winter data were missing. For predictions, however, we anticipate an equal number for each season (class).

Within each class there is a different number of subclasses, $m = 1 - M$, $M = M(n)$, the number M being determined, perhaps, by air mass type. The subclassification of the original data sets, for later multiple regression, is to be based upon the results of historical 24 h predictions using the numerical model, rather than observations, because the same numerical model will be used to identify the subclasses for future applications of the regression equations.

The number of data sets in each subclass, I_{mn}, is different because of the different predicted frequencies of the air masses. Lacking information to the contrary it must be assumed *a priori* that the frequency of prediction of each subclass will remain proportional to the corresponding I_{mn}. Moreover, since the subclass events themselves are to be predicted, it will be appropriate, following Glahn's suggestion, to include the variance among subclasses when the measure of predictability is formulated.

Fig. 1 also illustrates the nomenclature. $\bar{y}$ is a grand mean, to be defined; $\bar{y}^n$ is a class mean with sample size I_n; $\bar{y}^{mn}$ is a subclass mean with sample

0021-8952/82/081203-03$04.75

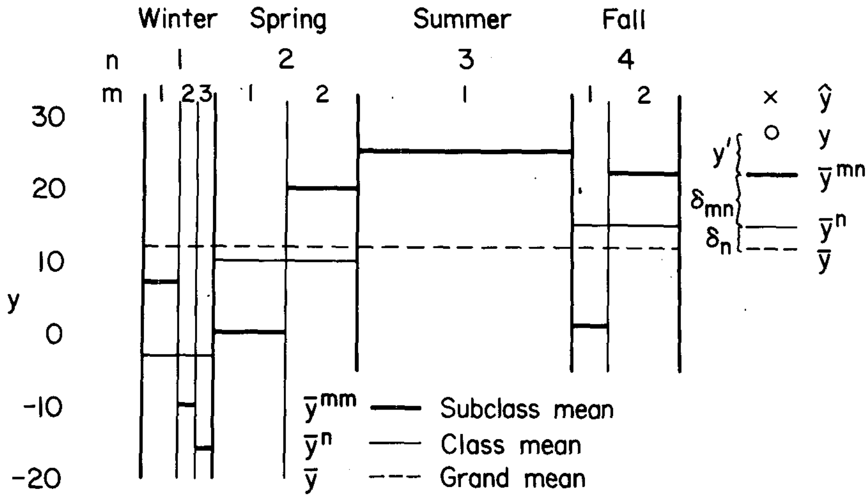

FIG. 1. A graphical representation of the hypothetical problem discussed in the text for the values given in Table 1. The horizontal dimensions are proportional to the number of data sets in the subclasses.

size I_{mn}; the deviation from the subclass mean is $y'_{imn} = y_{imn} - \bar{y}^{mn}$; and $\hat{y}_{imn}$ is an estimate of y_{imn}. Then setting

$$\delta_n = \bar{y}^n - \bar{y} \quad \text{and} \quad \delta_{mn} = \bar{y}^{mn} - \bar{y}^n \text{ (see Fig. 1)},$$

it follows that

$$y = \bar{y} + \delta_n + \delta_{mn} + y', \qquad (1)$$

where the subscripts on y and y' have been dropped.

The problem of weighting is easily handled in this example by assigning the weight $1/I_n$ to each data set in the sample. Then the weighted grand mean is defined by

$$\bar{y} = \frac{\sum\limits_{imn} y/I_n}{\sum\limits_{imn} 1/I_n} = \frac{1}{N}\sum_n \bar{y}^n. \qquad (2)$$

In a similar way the weighted standard error of estimate is

$$s_e^2 = \overline{(\hat{y} - y)^2} = \frac{1}{N}\sum_n \overline{(\hat{y} - y)^2}^n. \qquad (3)$$

These formulas weight each class equally, but each subclass is weighted within its class, according to the number of data sets I_{mn}.

The weighted variance of the entire data set is now

$$s_y^2 = \overline{(y - \bar{y})^2} = \frac{1}{N}\sum_n \overline{(y - \bar{y})^2}^n. \qquad (4)$$

Using (1), Eq. (4) reduces to

$$s_y^2 = \frac{1}{N}\sum_n (\delta_n^2 + \overline{\sigma_{mn}^2}^n + \overline{y'^2}^n), \qquad (5)$$

TABLE 1. Hypothetical numerical values for the example of Fig. 1. Values below the center line are derived from "data" given above the line.

Class n	1			2		3	4	
Subclass m	1	2	3	1	2	1	1	2
I_{mn}	1000	500	500	2000	2000	5000	1000	2000
$\overline{y'^2}^{mn}$	100	36	9	36	64	64	100	25
$\overline{(\hat{y} - y)^2}^{mn}$	75.0	18.0	0.9	7.2	32.0	19.2	90.0	15.0
$\bar{y}^{mn}$	7	−10	−16	0	20	25	1	22
r_{mn}^2	0.25	0.50	0.90	0.80	0.50	0.70	0.10	0.40
I_n		2000		4000		5000	3000	
$\bar{y}^n$		−3		10		25	15	
$\bar{y}$				11.75				
δ_{mn}	10	−7	−13	−10	10	0	−14	7
δ_n		−14.75		−1.75		13.25	3.25	

which is simply the sum of the variances involving δ_n, δ_{mn} and y'.

The total reduction of variance including the climatological effect is clearly

$$R_a^2 = 1 - \frac{s_e^2}{s_y^2}, \tag{6}$$

where s_e^2 and s_y^2 are given by (3) and (5). To eliminate the effect of the seasonal (class) variance it is only necessary to drop δ_n^2 from (5) leaving, instead of (6),

$$R_b^2 = 1 - \frac{\sum_n \overline{(\hat{y} - y)^2}^n}{\sum_n (\overline{\delta_{mn}^2}^n + \overline{y'^2}^n)}. \tag{7}$$

Eq. (7) thus measures the combined predictability of the subclasses (by the numerical model) and the departures y' (by the regression equations) *after* the seasonal classification.

Finally, if the intent is to measure the predictability of the regression equations alone, *after* first forming classes and subclasses, the term in δ_{mn}^2 is dropped from (7) leaving

$$R_c^2 = 1 - \frac{\sum_n \overline{(\hat{y} - y)^2}^n}{\sum_n \overline{y'^2}^n}. \tag{8}$$

The sample values listed in Table 1 illustrate the calculation of the three measures listed in (6), (7) and (8). The sums entering (6) are

$$\sum_n \overline{(\hat{y} - y)^2}^n = 121.025, \quad \sum_n \delta_n^2 = 406.750,$$

$$\sum_n \overline{\delta_{mn}^2}^n = 302.500, \quad \sum_n \overline{y'^2}^n = 225.250.$$

It follows that

$$R_a^2 = 0.870, \quad R_b^2 = 0.771, \quad R_c^2 = 0.463.$$

Thus there are many possible definitions of a combined reduction of variance (or an average correlation coefficient), but in each case the measure should be clearly defined on a rational basis.

REFERENCE

Faller, A. J., 1981: An average correlation coefficient. *J. Appl. Meteor.*, **20**, 203–205.