NOTES AND CORRESPONDENCE

Anomaly Correlation and an Alternative: Partial Correlation

UWE RADOK*

Geophysical Institute, University of Alaska, Fairbanks, Alaska

TIMOTHY J. BROWN

Cooperative Institute for Research in Environmental Studies, University of Colorado, Boulder, Colorado

2 December 1991 and 31 August 1992

ABSTRACT

The "anomaly correlation" between a predicted time series and a verifying series, both of which have had the same climatic history subtracted, combines their "direct" correlation with their two climate correlations and with the variance of the historical series. The direct correlation can be freed of climatic effects to become the "partial" correlation, which therefore seems a better parameter for judging predictive skill.

The anomaly correlation coefficient has been widely used to verify model predictions (e.g., Miyakoda et al. 1986; Tracton et al. 1989; Palmer et al. 1990). In terms of the relevant spatial covariance (cov) and variances (var), it is defined as

$$R = \frac{\operatorname{cov}[(O - C)(M - C)]}{[\operatorname{var}(O - C)\operatorname{var}(M - C)]^{1/2}}, \qquad (1)$$

where O are the observed values, M the modeled values, and C the climatological values for the points of a grid at a given time. Murphy and Epstein (1989) have shown that R is just one component in a "decomposited" mean-square-error–type skill score and that under very special conditions, R contains an in-built bias of 0.5. Yet in view of its wide use, R itself deserves a fuller decomposition.

It will be assumed that the observed field O and the modeled field M have spatial variances that are multiples of the climatic field variance C, so that $\operatorname{var}O = b_1 \operatorname{var}C$ and $\operatorname{var}M = b_2 \operatorname{var}C$; when time means are considered both the b could approach 1. The correlations r_{OC} and r_{MC} of the observed and modeled fields with the climate field are also needed; they reflect primarily the phase relationships of the three large-scale patterns (L. Gandin, personal communication).

It follows then that $\operatorname{var}(O - C) = \operatorname{var}O + \operatorname{var}C - 2\operatorname{cov}OC$ can be put into the form

$$\operatorname{var}(O - C) = (1 + b_1 - 2b_1^{1/2} r_{OC}) \operatorname{var}C. \qquad (2a)$$

Similarly,

$$\operatorname{var}(M - C) = (1 + b_2 - 2b_2^{1/2} r_{MC}) \operatorname{var}C, \qquad (2b)$$

and (1) becomes

$$R = \frac{(\operatorname{cov}OM - \operatorname{cov}OC - \operatorname{cov}MC + \operatorname{var}C)}{\gamma \operatorname{var}C}, \qquad (3)$$

where $\gamma = [(1 + b_1 - 2b_1^{1/2} r_{OC})(1 + b_2 - 2b_2^{1/2} \times r_{MC})]^{1/2}$

To simplify the algebra both b_1 and b_2 are approximated by their mean b. The first covariance in (3) defines the correlation r_{OM} between O and M themselves; this is initially close to unity and tends to decline as the modeling progresses and the large-scale patterns O and M develop phase differences due to the growth of baroclinic disturbances. The other two covariances, $\operatorname{cov}OC$ and $\operatorname{cov}MC$, jointly define the average, $\bar{r}_c$, of the two climate correlations, r_{OC} and r_{OM}. The arithmetic mean can be approximated using the geometric mean $\rho_C = (r_{OC} r_{MC})^{1/2}$ in the linear term resulting from the expansion of the product under the root in γ, $[-4b^{1/2}(1 + b)\bar{r}_c]$. Writing this as $[-4b^{1/2}(1 + b)\rho_C]$ simplifies the factor γ to $\gamma' = 1 + b - 2b^{1/2}\rho_C$. Using ρ_C in place of $\bar{r}_c$ will introduce no appreciable error in R until the observed and modeled large-scale patterns develop substantial phase differences.

* Also affiliated with CIRES, Boulder, Colorado.

Corresponding author address: Dr. Uwe Radok, University of Colorado, Campus Box 449, Boulder, CO 80309-0449.

With these changes, (3) finally takes the approximate form

$$R \approx \frac{b}{\gamma\gamma'}\left(r_{OM} - \frac{2}{b^{1/2}}\rho_C\right) + \frac{1}{\gamma'}. \quad (4a)$$

For increasing b (diminishing climate var$C$) the value of R approaches that of r_{OM}. When all the variances are approximately equal, for example, for forecasts of time averages, $b \approx 1$ and $\gamma' \approx 2(1 - \rho_C)$, (4a) then takes the simpler form

$$R \approx \frac{1}{2} + \frac{r_{OM} - \rho_C}{2(1 - \rho_C)}. \quad (4b)$$

Now R has the value 0.5 whenever the basic correlation r_{OM} equals the average climate correlation ρ_C. This agrees with the bias previously found by Murphy and Epstein (1989) to affect R in more special conditions. In the general case, a smaller bias amounting to $1/\gamma'$ is indicated by (4a).

To illustrate the implications of this decomposition, (4a) and (4b) have been evaluated for realistic ranges of r_{OM} and ρ_C and three variance factors b. The resulting values of R are given in Table 1. For small values of the average climate correlations ρ_C, the anomaly correlation R, due to its in-built positive bias, comes out larger than r_{OM} instead of falling below r_{OM} when the climate connection is discounted. The anomaly correlations become indeed smaller than r_{OM} for larger climate correlations, but the differences are largest for intermediate values of the variance ratio b. Such a complicated dependence of the anomaly correlation coefficient R on the average climate correlation ρ_C, and on the relative magnitudes of the different variances involved, puts into doubt the usefulness of R as a skill diagnostic, at least from a statistical point of view.

A simple alternative verification parameter is the "partial correlation," $r_{OM.C}$, between the deviations of O and M from their regression lines on C (e.g., Yule and Kendall 1950, chapter 12); this is given by

$$r_{OM.C} = \frac{r_{OM} - r_{OC}r_{MC}}{[(1 - r_{OC}^2)(1 - r_{MC}^2)]^{1/2}}, \quad (5a)$$

or with the assumptions made here,

$$r_{OM.C} \approx \frac{(r_{OM} - \rho_C^2)}{(1 - \rho_C^2)}. \quad (5b)$$

This is free of (linear) climatic effects on both O and M and is also given in Table 1; it closely approximates the correlation r_{OM} for negligible climate effects and registers a decrease in skill as those effects become more substantial. It is intriguing that for some values of ρ_C and b, the anomaly correlation closely approximates $r_{OM.C}$.

The advantages of the partial correlation coefficient include that its frequency distribution can be rendered

TABLE 1. Values of the anomaly correlation coefficient R and partial correlation $r_{OM\cdot C}$ for different variance ratios b and climate correlations ρ_C.

	$\rho_C = 0.1$				$\rho_C = 0.3$			
r_{OM}	$R\ (b = 1)$	$R\ (b = 2)$	$R\ (b = 5)$	$r_{OM.C}$	$R\ (b = 1)$	$R\ (b = 2)$	$R\ (b = 5)$	$r_{OM.C}$
0.90	0.94	0.93	0.91	0.90	0.93	0.91	0.89	0.89
0.80	0.90	0.85	0.82	0.80	0.86	0.81	0.79	0.78
0.70	0.83	0.77	0.72	0.70	0.79	0.72	0.68	0.67
0.60	0.78	0.71	0.64	0.60	0.71	0.63	0.57	0.56
0.50	0.72	0.63	0.54	0.49	0.64	0.54	0.46	0.45
0.40	0.67	0.56	0.46	0.39	0.57	0.44	0.36	0.34
0.30	0.61	0.48	0.37	0.29	0.50	0.35	0.25	0.23
0.20	0.56	0.41	0.28	0.19	0.43	0.26	0.14	0.12
0.10	0.50	0.34	0.19	0.09	0.36	0.16	0.03	0.01
0.00	0.44	0.26	0.10	−0.01	0.29	0.07	−0.07	−0.10
	$\rho_C = 0.5$				$\rho_C = 0.7$			
r_{OM}	$R\ (b = 1)$	$R\ (b = 2)$	$R\ (b = 5)$	$r_{OM.C}$	$R\ (b = 1)$	$R\ (b = 2)$	$R\ (b = 5)$	$r_{OM.C}$
0.90	0.90	0.87	0.87	0.87	0.83	0.80	0.83	0.80
0.80	0.80	0.75	0.73	0.73	0.67	0.61	0.65	0.61
0.70	0.70	0.62	0.60	0.60	0.50	0.41	0.48	0.41
0.60	0.60	0.50	0.47	0.47	0.33	0.22	0.30	0.22
0.50	0.50	0.37	0.34	0.33	0.17	0.02	0.12	0.01
0.40	0.40	0.25	0.20	0.20	0.00	−0.18	−0.05	−0.18
0.30	0.30	0.12	0.07	−0.07	−0.17	−0.37	−0.21	−0.37
0.20	0.20	−0.01	−0.05	−0.07	−0.33	−0.57	−0.39	−0.57
0.10	0.10	−0.14	−0.20	−0.20	−0.50	−0.76	−0.57	−0.76
0.00	0.00	−0.26	−0.33	−0.33	−0.67	−0.96	−0.74	−0.96

TABLE 2. Correlations and related parameters for six 24-h forecasts of the 500-hPa profile along 45°N (Charney and Eliassen 1949; Figs. 4 and 8–13): C — climate, O — observed, and M — forecast (model).

Starting date	8 January 1946	9 January 1946	10 January 1946	11 January 1946	12 January 1946	13 January 1946
(a) Correlations computed from data						
r_{OM}	0.88	0.87	0.91	0.96	0.93	0.76
$r_{OM.C}$	0.80	0.78	0.85	0.77	0.81	0.29
$R_{(O\text{-}C,M\text{-}C)}$	0.80	0.78	0.84	0.88	0.86	0.32
(b) Anomaly correlation R computed from (4a) and auxiliary parameters used						
$R_{(O\text{-}C,M\text{-}C)}$	0.77	0.79	0.83	0.87	0.85	0.41
r_{OC}	0.60	0.63	0.74	0.90	0.85	0.75
r_{MC}	0.75	0.62	0.63	0.93	0.81	0.90
$\rho_C = \sqrt{r_{OC}r_{MC}}$	0.67	0.68	0.69	0.92	0.83	0.82
$b_1 = \text{var}O/\text{var}C$	2.79	3.74	2.66	3.72	4.87	2.34
$b_2 = \text{var}M/\text{var}C$	1.92	3.72	3.44	3.60	4.59	4.74
$b = (b_1 + b_2)/2$	2.36	3.73	3.05	3.66	4.73	3.54

almost Gaussian with Fisher's z transformation (e.g., Fisher 1941, section 35);

$$z = \frac{1}{2}\log_e\left(\frac{1+r}{1-r}\right). \tag{6}$$

If the observed distribution of the meteorological correlations can be normalized in this way, (6) provides a chance model for assessing deviations of individual skill values from the average skill of many model runs, as was attempted for anomaly correlations in Fig. 7 of Tracton et al. (1989).

Partial correlation was suggested for model verification by Grant (1955) and used in a barotropic NWP model experiment by Voice et al. (1972). Grant's original argument used the results of the historical first numerical weather prediction experiment of Charney and Eliassen (1949), which can serve to illustrate the different parameters here discussed. The experiment provided six 24-h forecasts of the 500-hPa height profile along 45°N, starting from 0400 UTC on six successive days in January 1946. The different correlations have been calculated from the climate (C), observed (O), and forecast (M) heights at 12 longitudes 30° apart, starting at 150°W, and are shown in Table 2a. The anomaly correlations as computed with (4a) are shown in Table 2b together with the other parameters used in that equation. The arithmetic and geometric means of the two climate correlations closely agree in all cases, so only ρ_C is listed in Table 2b.

The anomaly correlations R and partial correlations $r_{OM.C}$ generally are similar to and smaller than the direct correlation between observed and predicted heights; larger differences appear for the fourth and fifth forecasts. A marked drop in both climate-adjusted correlations is seen for the last forecast, the only one with a direct correlation r_{OM} smaller than 0.8. Table 2b shows that only in this case were the variance ratios b_1 and b_2 sufficiently different to call for the use of (3) for R instead of (4a) with their mean value as b.

A much larger set of similar data, including the smaller correlations r_{OM} that tend to occur in longer-range forecasts, would be needed to describe the full behavior of the anomaly and partial correlations; we believe that the analysis here presented provides at least a useful overview.

Acknowledgments. We thank an anonymous reviewer of an earlier version of this note for pointing out the need to consider variance differences.

REFERENCES

Charney, J. G., and A. Eliassen, 1949: A numerical method for predicting the perturbations of the middle-latitude westerlies. *Tellus*, **1**, 38–54.

Fisher, R. A., 1941: *Statistical Methods for Research Workers*, 8th ed. Oliver and Boyd, London, 344 pp.

Grant, A. M., 1955: On results of forecasting with the barotropic model. *Tellus*, **7**, 275–276.

Miyakoda, K., J. Sirutis, and J. Ploshay, 1986: One-month forecast experiments—Without anomaly boundary forcings. *Mon. Wea. Rev.*, **114**, 2363–2401.

Murphy, A. H., and E. S. Epstein, 1989: Skill scores and correlation coefficients in model verifications. *Mon. Wea. Rev.*, **117**, 572–581.

Palmer, T. N., C. Brancovic, F. Molteni, S. Tibaldi, L. Ferranti, A. Hollingsworth, U. Cubasch, and E. Klinker, 1990: The European Center for Medium-Range Weather Forecasts (ECMWF) program on extended-range prediction. *Bull. Amer. Meteor. Soc.*, **71**, 1317–1330.

Tracton, M. S., K. Mo, and W. Chen, 1989: Dynamical extended range forecasting (DERF) at the National Meteorological Center. *Mon. Wea. Rev.*, **117**, 1604–1635.

Voice, M. E., U. Radok, D. Jenssen, and J. Bayliss, 1972: A barotropic numerical weather prediction experiment for the Australian region. *Aust. Meteor. Mag.*, **20**, 129–173.

Yule, U. G., and M. G. Kendall, 1950: *An Introduction to the Theory of Statistics*. Hafner, 701 pp.