Combining Predictive Schemes in Long-Range Forecasting

KLAUS FRAEDRICH* AND NEVILLE R. SMITH

Bureau of Meteorology Research Centre, Melbourne, Australia

5 November 1987 and 24 October 1988

ABSTRACT

The linear combination of two statistical forecast schemes of a single observable provides, in the average, a more accurate prediction than the individual forecasts alone. This method is applied to long-range forecasting of the monthly mean tropical Pacific sea surface temperatures.

1. Introduction

Long-range weather forecasting is still considered as "an empirical art . . . with a mix of techniques that are statistical to some extent" (Gilman 1985). There are, however, promising results from both deterministic and statistical prediction models of the El Niño/Southern Oscillation process (Cane et al. 1986; Barnett 1984). To enhance the performance of any objective long-range prediction method, it is suggested that, in some objective manner, consensus forecasts should be made to improve the accuracy. Only recently has it been shown in practice that, when applied to short-term probability of rainfall prediction, the combination of a stochastic forecast scheme and a deterministic NWP model can improve the average skill of the individual methods considerably (Fraedrich and Leslie 1987). It appears to be the combination of two independent forecasts (compared with the combination of one prediction, i.e., an NWP model, and data), which provides an increase of the accuracy. It is the purpose of this note to show that the mix of prediction models, i.e., the error minimizing linear combination of forecast schemes, may achieve some gains for the long-range forecast skill.

To illustrate this claim we extend Thompson's (1977) analysis, who advocated combination forecasts in meteorology more than a decade ago, and apply the results to the forecast of the monthly mean sea surface temperature (SST) anomalies in the east equatorial Pacific. In section 2 the method and the prediction schemes are introduced; in section 3 two applications are presented.

* Permanent affiliation: Institut für Meteorologie, Freie Universität Berlin, D-1000 Berlin 41, FRG.

Corresponding author address: Dr. Klaus Fraedrich, Institut fur Meteorologie, Freie Universität Berlin, Dietrich Schäfer Weg 6–8, D-1000 Berlin 41, Federal Republic of Germany.

2. Combining forecast schemes

Consider two normalized anomaly forecasts $a_i = (x_i(t) - \langle x_i \rangle)/s_i$, $i = 1, 2$, predicting the observable x or its corresponding anomaly a, where $\langle x_i \rangle$ is the ensemble mean and s_i the respective standard deviation. These forecasts can be linearly combined,

$$a_*(t) = \alpha a_1(t) + \beta a_2(t), \tag{1}$$

minimizing the ensemble mean square error of the combination forecasts with respect to α and β:

$$E_* = \langle (a - a_*)^2 \rangle. \tag{2}$$

The weights $\alpha = (R_1 - rR_2)/(1 - r^2)$ and $\beta = (R_2 - rR_1)/(1 - r^2)$ lead to the hindcast skill

$$S_* = 1 - E_* = \frac{R_1^2 + R_2^2 - 2rR_1R_2}{1 - r^2}. \tag{3}$$

Here R_i are the correlation coefficients between the forecasts (x_i or a_i) and the observable (x or a), r is the correlation between the two forecasts (x_1, x_2 or a_1, a_2).

The ensemble mean square error, E_*, is regarded as amplitude sensitive, whereas the ensemble mean correlation P_* between combination prediction a_* and verification a is phase sensitive. As it can be shown that the weights α, β maximize P_*, one can conclude that they provide, in the average, minimum errors in both amplitude and phase.

The following individual prediction schemes are introduced (Table 1) before combining them in optimal fashion:

The *climate* mean prediction, $a_C(t) = 0$ serves as the reference forecast from which the hindcast skills, S, of all other schemes can be evaluated; that is, the climate hindcast provides the zero skill base. The most simple forecast model is *persistence*, $a_P(t) = a(t - n)$; here it is applied as a normalized anomaly prediction. Other schemes are the *autoregression*, a_A, or the *regression*, a_R, which predict the observable by linearly extrapo-

TABLE 1. List of predictive schemes and their hindcast skill. Here R, R_R and r are correlation coefficients: the lag-n auto-correlation of the observable, the lag-m cross-correlation (regression) between predictand a and predictor b, and the cross correlation between the forecasts to be combined (see section 2).

Forecast schemes	Normalized anomaly predictions	Skill
Climate	$a_C(t) = 0$	$S_C = 0$
Persistence	$a_P(t) = a(t-n)$	$S_P = 2R(t-n, t) - 1$
Autoregression	$a_A(t) = a(t-n)R(t-n, t)$	$S_A = R^2(t-n, t)$
Regression	$a_R(t) = b(t-m)R_R(t-m, t)$	$S_R = R_R^2(t-m, t)$
Persistence-climate	$a_{PC}(t) = a_A(t)$	$S_{PC} = S_A = R^2(t-n, t)$
Persistence-regression	$a_{PR}(t) = ((R - rR_R)a_P + (R_R - rR)a_R)/(1-r^2)$	$S_{PR} = (R^2 + R_R^2 - 2RR_Rr)/(1-r^2)$

lating it from the same variable $a(t-n)$ or from another one, say $b(t-m)$.

To show how the combination technique improves the average forecast skill of any one of the individual components, the persistence model is linearly combined with climate and the regression scheme (Table 1):

(a) The persistence-climate combination is identical with the individual autoregression prediction, $a_A = a_{PC} = \alpha a_P$. The weight $\alpha = R(t-n, t)$ is the lag n autocorrelation. It is easily realized that the persistence–climate skill is greater than (or equal to) the skill of persistence and of climate alone:

$$S_{PC} - S_P = [R(t-n, t) - 1]^2 > 0. \quad (4)$$

(b) The persistence-regression combination ($a_1 = a_P$, $a_2 = a_R$): The weights [Eq. (1)] are $\alpha = (R(t-n, t) - rR_R(t-m, t))/(1-r^2)$ and $\beta = (R_R(t-m, t) - rR(t-n, t))/(1-r^2)$; R is the cross correlation between the observable and the prediction $a_1 = a_P$, and here it is identical with the lag n autocorrelation; R_R is the cross correlation between the observable a and the prediction $a_2 = a_R$, which can be shown to be identical with the respective cross-correlation coefficient; r is the cross-correlation coefficient between the predictions $a_1 = a_P$ and $a_2 = a_R$. Here it reduces to the correlation between the predictor, $a(t-n)$, for the persistence forecast and the predictor, $b(t-m)$, for the regression forecast; that is, $r = \langle a_1 a_2 \rangle / \sigma_1 \sigma_2$ with the covariance $\langle a_1, a_2 \rangle = \langle a(t-n)b(t-m) \rangle R(t-n, t)R_R(t-m, t)$ and the standard deviation $\sigma_1 = R(t-n, t)$ and $\sigma_2 = R_R(t-m, t)$. As expected, the gain in skill is largest when the forecast schemes are independent ($r = 0$ in Table 1). Again it can be shown that, on the average, the persistence-regression combination gains skill over all other individual predictions: regression, persistence, autoregression, and climate:

$$\left.\begin{aligned} S_{PR} - S_R &= (R - rR_R)^2/(1-r^2) > 0 \\ S_{PR} - S_P &= (R_R - rR)^2/(1-r^2) + (1-R)^2 > 0 \\ S_{PR} - S_A &= (R_R - rR)^2/(1-r^2) > 0 \\ S_{PR} - S_C &= S_{PR} = (R_R - R)^2/(1-r^2) + R^2 > 0 \end{aligned}\right\}. \quad (5)$$

It should be mentioned that the hindcast skill of the persistence-autoregression combination is identical with the hindcast skill of the autoregression–regression combination, $S_{PR} = S_{AR}$. For a first-order autoregressive AR(1)-process, $R(t-n, t) = R^n(t-1, t)$.

In section 3 the combination technique is applied to the statistical forecasts of the monthly mean sea surface temperatures of the eastern equatorial Pacific.

3. Applications: Forecasting monthly mean SST of the equatorial Pacific

The variability of the tropical Pacific sea surface temperature (SST) and the related atmospheric circulation pattern—El Niño/Southern Oscillation (ENSO) phenomenon—has received considerable interest in modeling and forecasting large-scale and long-range dynamics. Due to its impact, various strategies for its prediction have been developed. In this section we shall briefly demonstrate the improvement gained by (a) objectively combining two readily available statistical techniques for the prediction of the monthly mean eastern equatorial Pacific SST anomaly, and (b) objectively combining a statistical technique and a deterministic model forecast for the prediction of the mean July SST at Talara, Peru.

(a) The first scheme is the persistence of the SST anomaly. The second one is the prediction by a regression between the SST anomaly and one of its predictors; here it is the date of the onset of the Australian summer monsoon (Holland 1986). It is used as a measure of the zonal wind in the equatorial region of the western and central Pacific, which has been identified as the key predictor in ENSO studies (e.g. Barnett 1984).

A comment on the physical interpretation of the individual forecast schemes is in order. The persistence and/or autoregression SST forecasts provide a zero-order dynamic of the surface layer of the tropical Pacific with its relatively large inertia contributing to the ENSO process. The zonal wind/SST regression relates the atmospheric forcing to the subsequent oceanic response in SST, which is phase-locked to the annual cycle and depends on the lead time of the forecast. The linear combination, finally, provides the minimum error forecast of the SST signal using both informations.

The published lagged autocorrelations of the equatorial Pacific SST anomalies for July (Wright 1985; Fig. 6a therein) and the lagged cross correlation between the date of the onset of the Australian summer monsoon at Darwin and east equatorial Pacific SST in July (Holland 1986; Fig. 10 therein) suffice as the first estimates for a demonstration (Table 2). Note that the monsoon onset/July SST correlation $R_R(t - m, t)$ is obtained in February (at the latest) and remains unchanged when the actual time proceeds to July. However, the correlations between monsoon onset and the SSTs of the other months between February and July change. The hindcast skill (variance explained) using the monsoon onset/SST regression is comparable with Barnett's (1984, Fig. 2 therein) results from predictions for the SST off South America, which are based on the zonal and meridional wind fields in the near equatorial regions. From these estimates the hindcast skills of the various predictions for the July SST anomaly are easily deduced: persistence, autoregression, regression and persistence–regression combination. The hindcast skill of the prediction schemes needs to be reduced by a zero order estimate of (twice) the artificial skill, $S_{art} \sim m(1 - S)/(N - m)$, when their performance using independent data (Davis 1978) is considered; m is the number of predictors and N the effective number of cases ($m = 2$, $n \approx 30$ for the combination).

The July SST anomaly hindcast skills of the different prediction schemes are shown in Fig. 1. The following features should be noted:

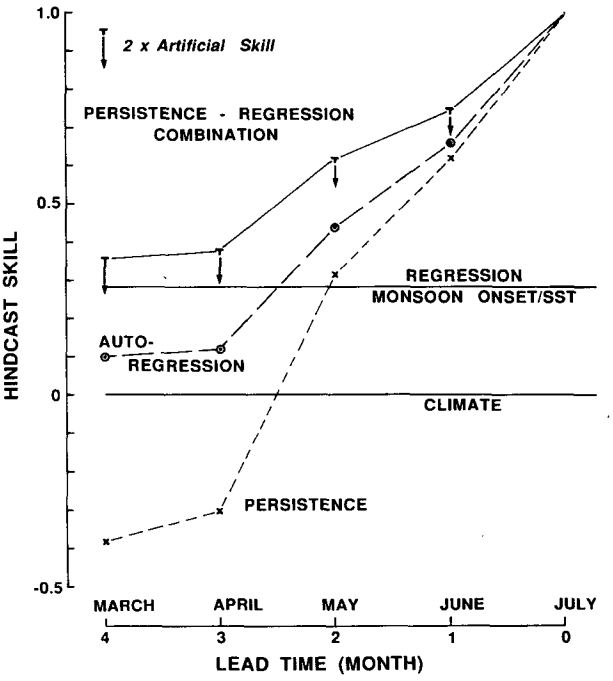

FIG. 1. Hindcast skill for the prediction of the monthly mean sea surface temperature of the eastern equatorial Pacific (July) using the individual forecast schemes (climate, regression, autoregression, persistence) and the persistence–regression combination. A zero-order measure of the artificial skill of the combination forecast is included to estimate the skill reduction when using independent data.

1) The hindcast skill rises after the lead time has passed the months of largest instability (March and April) of the ENSO process.

2) Maximum increased skill is gained from a linear combination of two forecast schemes when they are independent. For the persistence– (and autocorrelation–) regression combination, this is the case in March, April and May, when $r \sim 0$ (Table 2).

3) As the persistence–climate (or zero skill) combination is equivalent to the autoregression forecast, it is not surprising that persistence–regression and autoregression–regression combination are equivalent. This indicates that the two models to be combined need to have independent skill. Furthermore, the amount of skill gained from the persistence-regression combination over the regression alone is provided by the skill of the autoregression prediction, whenever the forecast schemes are approximately uncorrelated ($r \sim 0$). This is observed in March, April and, to a certain extent in May.

4) The (twofold) artificial skill reduces the combination hindcast skill to its anticipated independent forecast skill, which appears to be comparable with the hindcast levels of the individual schemes (i.e., the regression and/or the autoregression); these levels, however, need also to be reduced by their corresponding artificial skills, when the individual schemes are applied to independent data.

TABLE 2. Auto- and cross-correlation coefficients used in evaluating the hindcast skill of example (a): The lag auto-correlation (R) of the equatorial Pacific SST in July (Wright 1985); the cross correlation between the date of the Australian monsoon onset and the east equatorial Pacific SST in July (R_R, i.e., constant with decreasing lead time), and the same for other monthly SSTs (r), (Holland 1986).

Parameter	March	April	May	June	July
July SST autocorrelation R	.31	.35	.66	.81	1.00
Monsoon onset/July SST R_R	−.53	−.53	−.53	−.53	−.53
Monsoon onset/monthly SST r	−0.5	−.07	−.17	−.30	−.53

(b) Here we use the July mean SST at a particular site—Talara, Peru—as the predictand, and use persistence (or equivalently autoregression) of Talara SST at 1-, 2- and 3-month lead times as the first scheme. As in example (a), this scheme can be interpreted as a zero-order thermal model of SST change in the eastern Pacific taking into account local effects and the thermal inertia of tropical waters, elements which are absent in the deterministic model. The second scheme consists of 1-, 2- and 3-month forecasts for the upper-layer thickness anomaly (a proxy for thermocline change) in the eastern Pacific averaged over 10°S–10°N, 80°–100°W for 1963–1987. They are determined from the Florida State University tropical Pacific model (Inoue

and O'Brien 1984, 1986; O'Brien, personal communication) and regressed to Talara July SST. This scheme can be regarded as a diagnostic tool for interpretation of the dynamic effect of surface wind forcing in the central and western Pacific at a remote site in the eastern Pacific. Table 3 sets out the correlation coefficients for the predictors and predictands.

TABLE 3. Auto- and cross-correlation coefficients used in evaluating the hindcast skill of example (b): R is the lag autocorrelation of Talara SST in July; R_R is the cross-correlation between the 1-, 2- and 3-month forecasts for upper-layer thickness anomaly of the FSU dynamic model (Inoue and O'Brien 1984, 1986) and Talara SST in July; r is the same but for SSTs in the 3 preceding months.

Parameter	April	May	June	July
July Talara SST autocorrelation R	.77	.87	.93	1.00
FSU model/July SST R_R	.54	.52	.57	.46
FSU model/monthly SST r	.27	.42	.53	.46

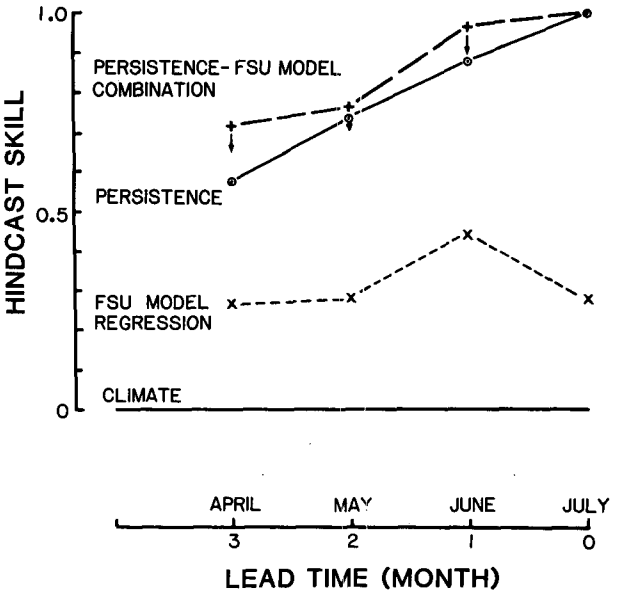

FIG. 2. Hindcast skill for the prediction of Talara SST (July) using individual forecast schemes (persistence and regression) and a combination scheme (SST persistence − FSU model regression). The data for the FSU June 1-month forecasts were available only for the period 1979–85. Continuous wind (perfect boundary) conditions were used for FSU July simulation. A zero-order measure of the artificial skill level of the combination forecast is included to estimate the skill reduction when using independent data.

The period for the hindcast experiment is dictated by the common availability of data for the predictors and the results are shown in Fig. 2. The skill of the FSU model alone is generally less than half that of the statistical model. The choice for the predictand (July SST at a *site* rather than an areal *average*) and the lack of model thermodynamics contribute to this difference. (Note that the skill here is relative to observed climate whereas Inoue and O'Brien (1984) use an evaluation based upon the model climate.) The combination of persistence and the FSU deterministic model regression (or equivalently autoregression–regression) demonstrates improvement in the hindcast skill (allowing for the artificial skill level), most notably at longer lead times when the two schemes are more nearly independent (Table 3). We must be wary, however, of overinterpreting this result, since the correlation coefficients and the skill evaluation both derive from the same data. A true evaluation must await longer data records with the skill being based on comparisons between the combined forecasts and independent data. This method is also a demonstration of the optimal use of model diagnostics in cases where the model is unable to capture key features of the thermodynamic and thus contains an inherent bias.

4. Conclusion

It has been shown that considerable hindcast skill can be gained when combining two predictive schemes in an optimal (i.e. error-minimizing) fashion. Long-range forecasting is probably one area of fruitful application, where such a combination technique need not be confined to statistical schemes. For example, the subjective consensus between monthly predictions as presently applied to long-range forecasting (Kalnay and Livezey 1985). The objective combination of both stochastic and deterministic long-range forecast models of, for example, the ENSO phenomenon (as a continuous process) appears to be as promising as the combination of Markov chains and a numerical weather prediction model in probabilistic short-term weather forecasting (Fraedrich and Leslie 1987).

Acknowledgments. We wish to thank Masamichi Inoue, Jim O'Brien and David Legler for kindly making the FSU model data available to us.

REFERENCES

Barnett, T. P., 1984: Prediction of the El Niño of 1982–83. *Mon. Wea. Rev.*, **112**, 1403–1407.

Cane, M. A., S. E. Zebiak, and S. C. Dolan, 1986: Experimental forecasts of El Niño. *Nature*, **321**, 827–832.

Davis, R. E., 1978: Predictability of sea level pressure anomalies over the North Pacific Ocean. *J. Phys. Oceanogr.*, **8**, 233–246.

Fraedrich, K., and L. M. Leslie, 1987: Combining predictive schemes in short-term forecasting. *Mon. Wea. Rev.*, **116**, 1640–1644.

Gilman, D. L., 1985: Long-range forecasting: The present and the future. *Bull. Amer. Meteor. Soc.*, **66**, 159–164.

Holland, G. J., 1986: Interannual variability of the Australian summer monsoon at Darwin: 1952–82. *Mon. Wea. Rev.*, **114**, 594–604.

Inoue, M., and J. J. O'Brien, 1984: A forecasting model for the onset of a major El Niño. *Mon. Wea. Rev.*, **112**, 2326–2337.

——, and ——, 1986: Predictability of the decay of the 1982/83 El Niño. *Mon. Wea. Rev.*, **114**, 967–972.

Kalnay, E., and R. Livezey, 1985: Weather predictability beyond a week: An introductory review. *Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics*. M. Ghil, R. Benzi and G. Parisi, Eds., North Holland, 311–346.

Thompson, P. D., 1977: How to improve accuracy by combining independent forecasts. *Mon. Wea. Rev.*, **114**, 228–229.

Wright, P. B., 1985: The Southern Oscillation: An ocean-atmospheric feedback system? *Bull. Amer. Meteor. Soc.*, **66**, 398–412.