Extended Forecasting

WINTER SEASON TEMPERATURE FORECASTS USING DISCRIMINANT ANALYSIS

Robert P. Harnack (1) and Michael W. Cammarata (1a)
Department of Meteorology and Physical Oceanography
Cook College—New Jersey Agricultural Experiment Station
Rutgers—The State University of New Jersey
New Brunswick, NJ 08903

Paper of the Journal Series
New Jersey Agricultural Experiment Station
Cook College
Rutgers—The State University of New Jersey
New Brunswick, NJ 08903

ABSTRACT

A set of winter temperature forecast experiments were conducted using multiple discriminant analysis with a "jackknife" procedure. Nine separate experiments were conducted using various predictor fields including North Pacific sea surface temperature (SST), tropical Pacific SST, North Atlantic SST, and Northern Hemisphere 700 mb heights defined for various periods (November, October, fall, and summer). For most trials the data sample consisted of the years 1950–78. Principal conclusions were:

1. Overall skill for most winter forecast trials were statistically significant at the 95% confidence level using the binomial distribution. Overall percent correct skill ranged from 59% (summer heights) to 71% (November All SST) for two-class models and 39% (summer heights) to 54% (November All SST) for three-class models.

2. A direct comparison between discriminant and regression analyses, using November Pacific SST predictors, winter sub area temperatures as predictands and a 1950–1978 data sample, showed that discriminant analysis gave superior results. Overall, the two-class skill was 67% versus 59% and the three-class skill was 48% versus 39%.

3. Overall skill decreased only slightly as forecast lead time increased. The skill for October and summer SST predictor models (68% and 67% respectively for two-class) were as good as operational short lead-time winter forecasts issued now.

1. INTRODUCTION

This is a brief update for interested readers on the latest in a series of long-range forecast experiments emphasizing United States winter season temperature forecasts. The history of past work, from which the present work is an extension, starts with the use of discriminant analysis applied to winter temperature at 19 locations in the eastern United States (Harnack and Landsberg, 2) followed by the use of linear regression applied to winter temperature at 9 then later at 15 subareas of the U.S. (Harnack, 3, 4). In these experiments most predictors were in the form of principal component amplitude time series for November fields of data (i.e., sea surface temperature and 700 mb heights mainly) and independent testing was conducted on a relatively small sample (i.e., less than ten cases). Significant skill was found for the best models, those using sea surface temperatures (SST), though skill levels were only slightly better than those obtained by using simple persistence.

This paper will summarize the results from a number of new winter season temperature forecast experiments using statistical methods and applied to the contiguous United States. The verification results for each method are compared to those of random chance and to simple persistence. The methods employed include linear regression and discriminant analysis. New results using the first method in a "jackknife" procedure have been reported recently (Harnack et. al., 5), and are reported only for comparison purposes. Results using discriminant analysis are reported for the first time in a comprehensive way. In a direct comparison, it will be shown that discriminant analysis produces superior results to those of linear regression.

Some of the other published papers describing operational and experimental U.S. seasonal forecast methods and results, several of which involve somewhat similar methodology, are: analog forecast experiments by Barnett and Preisendorfer (6) and Bergen and Harnack (7); linear regression experiments by Barnett (8); the statistical-physical methodology by Namias (9); skill levels of some experimental climate forecasts in Preisendorfer and Mobley (10); and descriptions of National Weather Service operational procedures and forecast skill by Gilman (11) and Harnack (12). All of these references indicate that winter season temperature forecasts are skillful relative to random chance overall, though skill levels are modest. As always, there is considerable geographical variation in skill shown.

2. PROCEDURE

Separate descriptions of the predictors, predictands, statistical methods, and verification procedures employed in this study follow.

a. Predictors

The experiments used raw SST data for the North Pacific, tropical Pacific and North Atlantic as well as Northern Hemisphere 700 mb heights. In separate trials, predictors were defined for summer and fall seasons, and the months of October and November. The spatial domains for each field consisted of the following:

1. The North Pacific SST consisted of 50 grid points in the region bounded by 25–55°N, 125°W–160°E with staggered 5 degree latitude by 10 degree longitude spacing. The data source was Scripps Institution of Oceanography.

2. The Tropical Pacific SST consisted of 34 grid points in the region bounded by 5°S–10°N, 85°W–165°W with staggered 5 degree latitude by 10 degree longitude resolution. The data sources and procedure are described in Harnack (4).

3. The North Atlantic SST consisted of 50 grid points in the region from 25–65°N, spanning the entire basin using a staggered 5 degree latitude by 10 degree longitude resolution. The data were also provided by Scripps.

4. Northern Hemisphere 700 mb heights consisted of 127 grid points over the latitude band 25–65°N covering the entire hemisphere with a staggered five degree latitude by ten degree longitude spacing. The data source was the National Meteorological Center.

Table 1. Listing of discriminant forecast trials showing model name and predictors available in predictor pool for screening. PC's denote principal component amplitudes. Numbers in parentheses are the number of PC's in predictor pool.

MODEL NAME	PREDICTOR POOL
Nov. Pacific SST	No. Pac. SST PC's (5), Trop. Pac. SST PC's (1)
Nov. All SST	No. Pac. SST PC's (5), No. Atl. SST PC's (4), Trop. Pac. SST PC's (1)
Nov. Heights	NH 700 mb Ht. PC's (6)
Fall Pacific SST	No. Pac. SST PC's (5), Trop. Pac. SST PC's (1)
Fall Heights	NH 700 mb Ht. PC's (6)
Oct. All SST	No. Pac. SST PC's (4), No. ATl. SST PC's (4) Trop. Pac. SST PC's (1)
Oct. Heights	NH 700 mb Ht. PC's (8)
Summer All SST	No. Pac. SST PC's (5), No. Atl. SST PC's (5), Trop. Pac. SST PC's (2)
Summer Heights	NH 700 mb Ht. PC's (5)

All predictor fields were subjected to principal component (PC) analysis and the number of retained PC's was determined using Monte Carlo simulations as previously described by Harnack et. al. (5). Table 1 describes the degree of data reduction accomplished. The total explained variance for each field in winter trials by the retained PC's were: 72% for North Pacific SST, 49% for tropical Pacific SST, 62% for North Atlantic SST, and 71% for 700 mb heights. For each forecast trial, predictors were in the form of time-varying amplitudes of the PC's.

b. Predictands

As in several earlier studies, area-averaged winter temperatures for 15 regions (subareas) of the contiguous United States were employed as predictands to be matched with various predictor combinations. The period of record included winters from 1950–78, labeled according to the year of January. Monthly climatic division (CD) temperatures were used as raw data. The subareas are defined in Figs. 1–3. Temperature classes were assigned by first ranking winter temperatures for the period 1941–70 for establishing tercile limits, and then tercile designation (three class); or finding subarea median values, then assigning above or below to each winter (two class).

c. Statistical methods

1. Discriminant analysis

Discriminant analysis was used in this study to derive classification equations for winter temperature class separately for each subarea. Generally, both two and three temperature classes were employed. This statistical method was tried on the winter

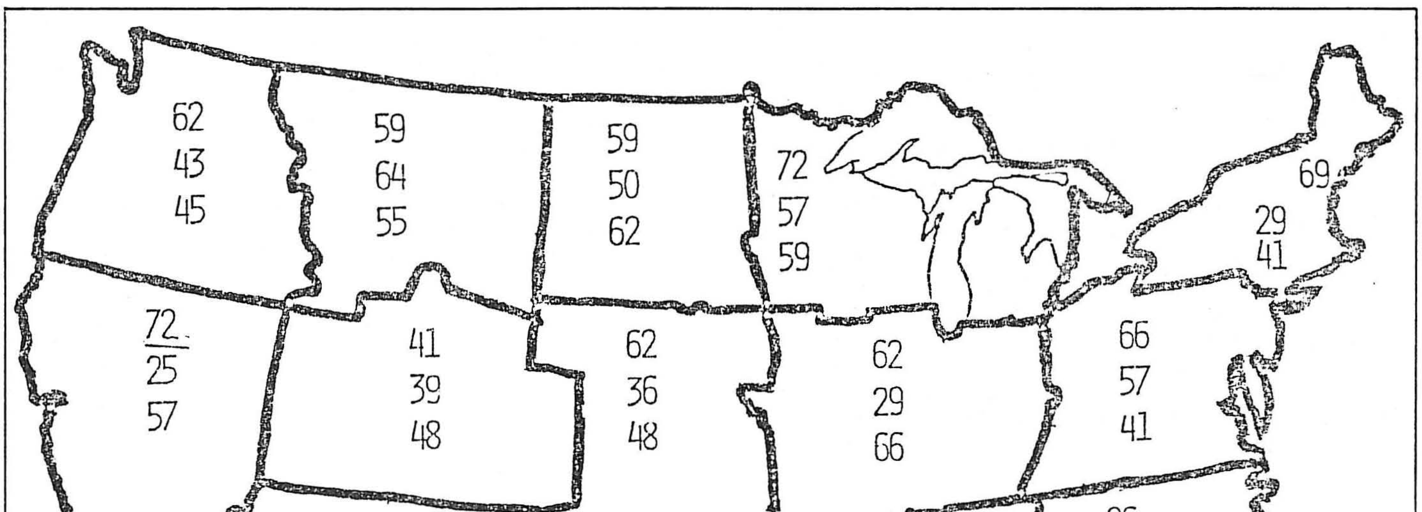

Fig. 1. Mean percent correct for two-class winter temperature forecasts made for 1950–78 with the November Pacific SST model by discriminant method, regression method, and simple persistence (top to bottom). Underline for the first two methods, denotes statistical significance at the 95% level of confidence.

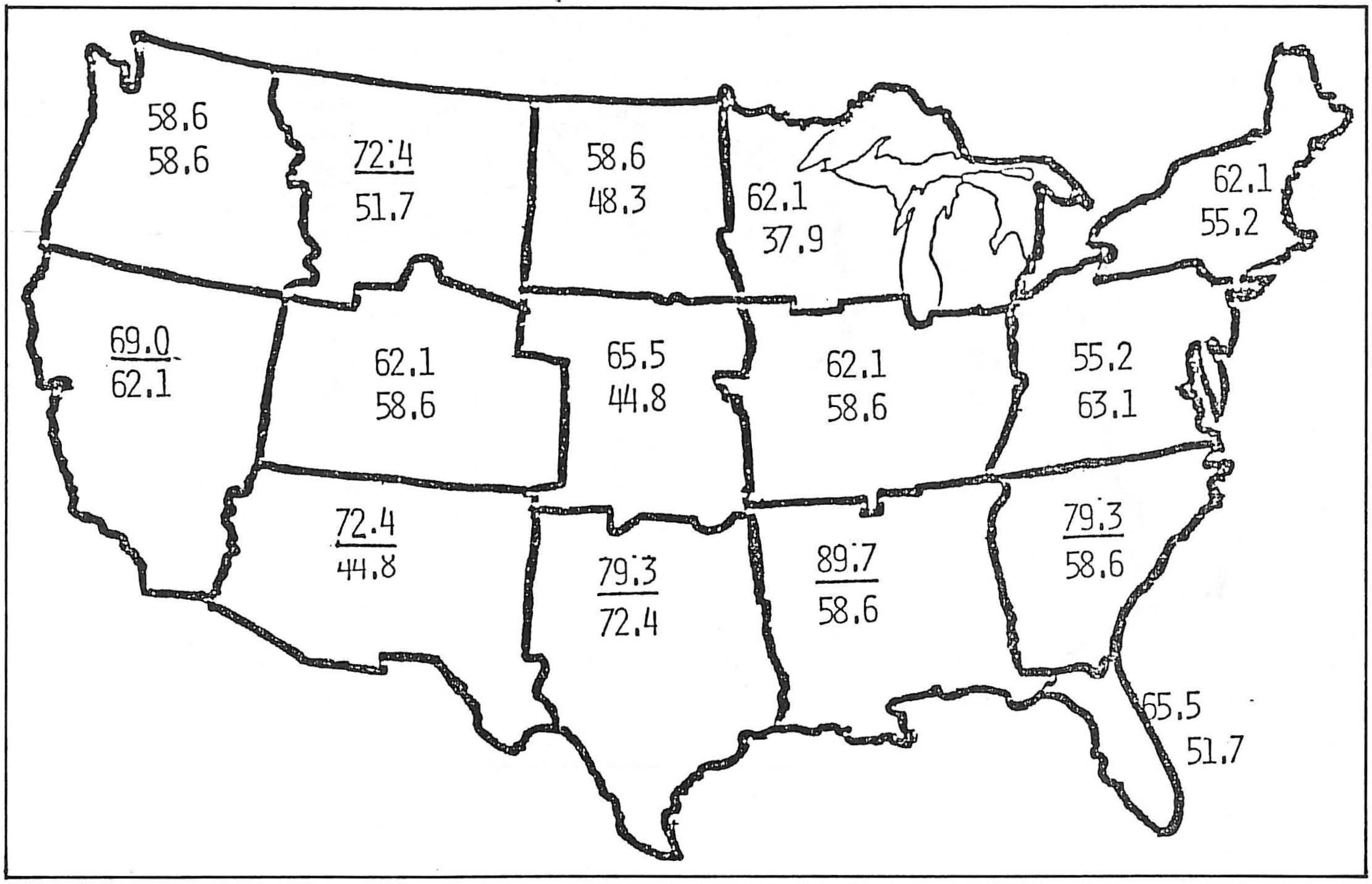

Fig. 2. Mean percent for two-class winter temeprature forecasts made for 1950–78 by the October All SST model (top number) and simple persistence (bottom number). Underline denotes statistical significance at the 95% level.

temperature forecast problem in an attempt to benefit from the use of predictor transformation to maximize group distinction. Though the classification equations are linear, the overall effect of using discriminant analysis is to allow for some non-linearity between data groups. This was a statistical experiment in which no physical reasoning for possible non-linearity is offered. Classification functions, one for each temperature class, are weighted sums of the original predictor variables that provide the maximum separation between classes. Given the temperature classes and predictors involved, the computation develops the best set of weights (coefficients) possible from the observations and produces a sum of squares (discriminant score) that best distinguishes between the temperature classes. In sum, linear combinations of original predictor variables are found which exhibit large differences in group means for the discriminant scores. Multiple discriminant analysis (MDA) is then a technique for finding those combinations that will separate the group means of the predictor variables to the maximum degree allowed by the predictors chosen for use. An excellent description of the technique, including the mathematics, is contained in Neubert (13). The BMDP software was used in this study for performing stepwise discriminant analysis.

2. Verification and statistical significance

The results for all methods employed are presented in terms of percent correct using two or three temperature classes which is conventional for many reported long-range forecast experiments. Most of the scores presented here were obtained by aggregating the number of correct forecasts made using the entire original sample employed in a "jackknife" procedure. Given 29 cases in the sample, 29 sets of discriminant classification equations are formulated for each subarea using data obtained from 28 cases. Each statistical relationship is tested and verified on the one case left out in its formulation. By selecting the 28 cases that are taken from the original 29, all available cases in the sample are used to make forecasts and are therefore included in the verification statistics. In addition, for one of the models (i.e., the one using November North Pacific SST predictors) equations were tested on a separate eight year independent sample.

Local skill (percent correct at individual subareas over all forecasts) and global skill (percent correct for individual years over all subareas) were obtained for each trial. Significance testing was performed on local skill and overall skill (all subareas and years combined) using the binomial distribution applied in a cumulative fashion (see Bergen and Harnack, 7). For the overall skill tests, it was assumed the temporal degrees of freedom (df) was equal to 19 based on the average lag 1 autocorrelation of .22 and the method of estimating effective sample size presented by Laurmann and Gates (14). They suggested:

$$N^* = N\left[\frac{1-\varphi_1}{1+\varphi_1}\right]$$

where N^* = effective sample size, N = actual sample size, and φ_1 = lag 1 autocorrelation. The spatial df was set at 6 based on Monte Carlo testing described in Harnack et al. (5, page 1952). Therefore, for overall significance 114 (i.e., 19 times 6) was used as the effective sample size. For local significance testing, the effective sample size was determined locally by application of

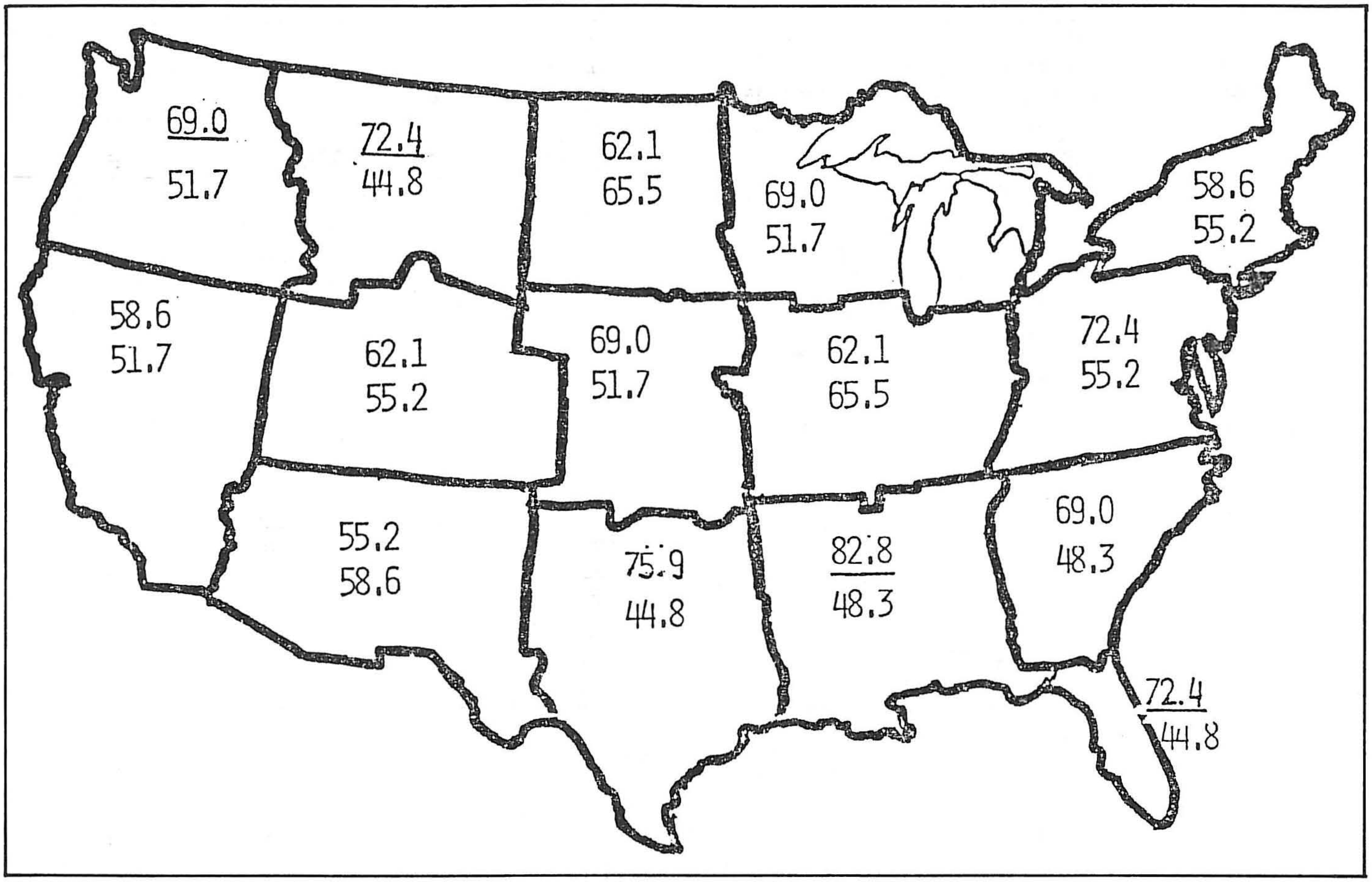

Fig. 3. Mean percent correct for two-class winter temperature forecasts made for 1950–78 by the summer All SST model (top number) and simple persistence (bottom number). Underline denotes statistical significance at the 95% level.

the Laurmann and Gates (14) formula using individual subarea lag 1 autocorrelations. These ranged from −.20 to .58, thereby yielding effective sample sizes ranging from 8 to 29. The binomial test was applied using the appropriate sample size, 95 percent confidence level, and probability of success by chance of 0.33 (three-class verification) or .50 (two-class verification). For all experiments, simple persistence was applied (i.e., assuming that the antecedent period observed class was the forecast). which is shown with most of the results for comparison purposes. To determine overall statistical significance relative to simple persistence, the probability of success is set equal to the overall percent correct for simple persistence, then the binomial test is applied as before.

d. Experimental trials performed

Table 1 shows the nine discriminant model trials conducted including the number of PC's available in the predictor pool for screening. The winters of 1950–78 constituted the sample for which MDA was employed in the "jackknife" procedure described earlier. In the case of the model using November North Pacific SST only as predictors, the winters of 1979–86 were also used for further independent testing. SST predictors were used in this study because of the encouraging results found in earlier work by the author and by others, as well as physical justification which Namias (9), in particular, has articulated over the past decades. Height predictors were included for comparison purposes and because antecedent circulation is an important component of the National Weather Service (NWS) long-range forecast indicators. The longer lead-time forecast trials, those using October and summer predictors, were conducted after encouraging results were obtained using fall predictors.

3. RESULTS

Overall forecast trial performance is summarized in Table 2. All trials except for two attained statistical significance relative to random chance, with both exceptions being from the three class set of trials, and most trials also attained statistical significance relative to simple persistence. The lower skill and nonsignificant models tended to be from the three class trials and models which used 700 mb heights as predictor variables. Clearly, models using SST predictors performed better than those using heights overall, though in some comparisons, it is a standoff.

The one regression model result (taken from Harnack et. al., 5) given in Table 2, line two, shows inferior skill to that of the discriminant model having the same predictor pool and sample (i.e., 59% vs. 67% for two class trials). This, plus the fact that the fall and October predictor model skill shown here also exceeds those found in somewhat similar earlier regression trials (Harnack, 4), suggests that discriminant analysis is superior to regression for long-range forecasts of winter temperature class.

Subarea comparisons between discriminant and regression analysis capabilities are shown in Fig. 1. This map also compares model performance with simple persistence and gives geographical variation of skill. Examination of the data, shows that the discriminant model was at least nominally supeior to regression in all but one subarea (Northern Rockies). The differences are particularly striking in three subareas: New England, the Ohio

Table 2. Overall verification statistics (percent correct) using all subareas and cases combined. Independent cases are used here in a "jackknife" framework. D denotes discriminant analysis and R denotes regression analysis. Persistence forecast verification shown for comparison. Underlined indicates models deemed significant at the 95% confidence level while * indicates models deemed significantly better than persistence at the 95% level.

	STAT.	2 CLASS (% CORRECT)		3 CLASS (% CORRECT)	
MODEL	METHOD	MODEL	PERST	MODEL	PERST
Nov. Pac. SST	D	67*	52	48*	36
Nov. Pac. SST	R	59	52	39	36
Nov. All SST	D	71*	52	54*	36
Nov. Hts.	D	60	52	42	36
Fall Pac. SST	D	65*	56	52*	40
Fall Hts.	D	67*	56	48*	40
Oct. All SST	D	68*	56	52*	38
Oct. Hts.	D	69*	56	47*	38
Sum. All SST	D	67*	53	50*	39
Sum. Hts.	D	59	53	39	39

Valley, and Far West. The greatest skill for the November Pacific SST discriminant model is seen in three subareas in the southeast where the two-class percent correct exceeds 80 percent, far better than the skill shown by simple persistence. Local significance at the 95 percent confidence level was achieved only in the southeast for this model, but still was also impressive in the Great Lakes, New England, southern Plains and the Far West subareas. A similarly prepared map for three-class verification (not shown), supports most of the foregoing conclusions except for the significant skill relative to chance found in the New England, Florida, and Northern Plains subareas. Percent correct values in these subareas were 59 to 62%. The lowest skill tended to occur in interior subareas of the West for both the two- and three-class verifications. Table 3 shows the two-class verification contingency table for the November Pacific SST model.

Table 3. Verification contingency table for November Pacific SST model. All subareas and years (1950–79) combined. A = ABOVE, B = BELOW.

PREDICTED	A	B	Totals
OBSERVED			
A	144	76	220
B	70	145	215
TOTALS	214	221	

The encouraging skill found for the short lead-time models lead to trials using SST and height predictors for October and summer. Surprisingly, the overall results shown in Table 2 indicate very little diminution of skill using the All SST models for comparison as the lead time increases. Two-class skill maps for the October and summer All SST models are shown in Figs. 2 and 3, respectively. These indicates that, like the November models, the best skill is noted in the southeast but high skill also tends to extend to the southwest for the October model. The two-class verification contingency table for October and summer All SST models are shown respectively in Tables 4 and 5. It should be noted that each of the contingency tables has a large degree of symmetry, indicating little preferential skill for above versus below forecasts.

Table 4. Verification contingency table for summer All SST model. All subareas and years (1950–79) combined. A = ABOVE, B = BELOW.

PREDICTED	A	B	Totals
OBSERVED			
A	150	70	220
B	72	143	215
TOTALS	222	213	

Table 5. Verification contingency table for October All SST model. All subareas and years (1950–79) combined. A = ABOVE, B = BELOW.

PREDICTED	A	B	Totals
OBSERVED			
A	146	74	220
B	67	148	215
TOTALS	213	222	

North Pacific SST data were obtained for the Novembers of 1978–1985 in order to test the November Pacific SST discriminant model for the additional winters of 1979–1986. These winter cases were used as completely independent cases (i.e., not in a "jackknife" mode). The other models were not similarly tested since the needed data were not available (i.e., Atlantic SST) at the time when these calculations were being made. Verification results for the two-class model included an overall percent correct of 46%, which increased to 49% when the model classification equations were updated each year, then employed to classify the next year. These near-random chance skill scores cast some doubt on the encouraging skill found in the "jackknife" mode. A comparison of skill for these winters for "jackknife" versus non-"jackknife" trials indicated that the major contributors to lower skill for this period were the winters of 1979–80 and 1984–85. Discussions with personnel at the National Weather Service's Climate Analysis Center indicate that their long-range forecasts performed with lower skill in recent years as well, suggesting that an unusual set of factors may have been operating in the atmosphere- ocean systems. Whether statistically based forecast methods using SST will return to higher levels of skill in the future is uncertain, but the results reported here suggest that users should be skeptical until the stability of the statistical relationships is verified.

4. CONCLUSIONS

A set of long-range forecast experiments were conducted along the lines of previously reported studies with more use made of North Atlantic SST. More importantly, multiple discriminant analysis was employed as an alternative to multiple linear regression to account for some non-linearity in relationships. As in the most recent published study (Harnack et al., 5), a "jackknife" approach was employed so that the amount of "independent" testing of model equations could be maximized. Nine separate experiments were conducted using various predictor fields including North Pacific SST, tropical Pacific SST, North Atlantic SST, and Northern Hemisphere 700 mb heights defined for various periods (November, October, fall, and summer). For most trials the data sample consisted of the winters 1950–78. Comparisons were made between skill achieved using regression analysis and discriminant analysis; between skill achieved with models having different lead times and predictors; and between skill achieved from one subarea to another. Principal conclusions were:

(1) Overall skill for most winter forecast trials were statistically significant at the 95% confidence level using the binomial distribution. Exceptions were the three-class November Pacific SST and summer heights models. Overall percent correct skill ranged from 59% (summer heights) to 71% (November All SST) for two-class models and 39% (summer heights) to 54% (November All SST) for three-class models.

(2) A direct comparison in which predictors (November Pacific SST), predictands (winter subarea temperatures), and data sample (1950–78) were the same but the statistical method was varied (discriminant versus regression analysis), showed that discriminant analysis gave superior results. Overall the two-class skill was 67% versus 59% and the three-class skill was 48% versus 39%.

(3) SST predictor models generally performed better than 700 mb height predictor models at all lead times, but the differences were most pronounced for the three-class models.

(4) Overall skill decreased only slightly as forecast lead times increased. The skill for October and summer SST predictor models (68% and 67% respectively for two-class) are as good as operational short lead-time winter forecasts now issued.

(5) A separate model validation on independent winters of 1979–86 for the November Pacific SST model (two-class) show much lower skill (49% correct) than the "jackknife" results seen for 1950–78, suggesting extra caution in interpreting results and using the derived models.

ACKNOWLEDGEMENTS

New Jersey Agricultural Experiment Station Publication No. D-13507-6-86 supported by State funds; by the Climate Dynamics Program, Division of Atmospheric Sciences, National Science Foundation (Grant ATM-8217215); and by the National Oceanic and Atmospheric Administration (Grant NA83-AA-D-00023).

The authors wish to thank Mr. Daniel Cayan of the Climate Research Group of Scripps Institute of Oceanography for supplying the sea surface temperature data set, Val Bowers for typing assistance, and the anonymous reviewer(s) for many helpful suggestions.

NOTES AND REFERENCES

1. Dr. Robert P. Harnack received his B.S. in Meteorology from Rutgers University and his M.S. and Ph.D from the University of Maryland. He is presently Professor of Meteorology at the Department of Meteorology and Physical Oceanography, Cook College, Rutgers University, New Brunswick, New Jersey. Dr. Harnack's speciality is extended forecasting and he has written many articles on this subject. He is the author of NWA Monograph 1-86 entitled "Principles and Methods of Extended Period Forecasting in the United States, 1986."

1a. Michael Cammarata is a research meteorologist at the NWS Technical Development Laboratory (TDL) in Silver Spring, MD. He is currently working on the Local AFOS MOS Program (LAMP). Mr. Cammarata received his M.S. in Meteorology from Rutgers University.

2. Harnack, R.P. and H.E. Landsberg, 1978: Winter season temperature outlooks by objective methods. J. Geophys. Res., 83, *3601–3616.*

3. Harnack, R.P., 1979: A further assessment of winter temperature predictions using objective methods. Mon. Wea. Rev., 107, *250–267.*

4. Harnack, R.P., 1982: Objective winter temperature forecasts: an update and extension to the West. Mon. Wea. Rev., 110, *287–295.*

5. Harnack, R.P., J. Harnack and J.R. Lanzante, 1986: Seasonal temperature prediction using a jackknife approach with an intraseasonal variability index. Mon. Wea. Rev., 114, *1950–1954.*

6. Barnett, T.P. and R.W. Preisendorfer, 1978: Multifield analog prediction of short-term climate fluctuations using a climate state vector. J. Atmos. Sci., 35, *1771–1787.*

7. Bergen, R. and R.P. Harnack, 1982: Short-term climatic anomaly predictions using a simple analog approach. Mon. Wea. Rev., 110, *1083–1099.*

8. Barnett, T.P., 1981: Statistical prediction of North American air temperatures from Pacific predictors. Mon. Wea. Rev., 109, *1021–1041.*

9. Namias, J., 1984: Short period climatic variations, Collected works of J. Namias, Vol. III, 1975 through 1983, University of California, San Diego, 393 pp.

10. Preisendorfer, R.W. and C.D. Mobley, 1984: Climate forecast verifications, United States mainland, 1974–1983. Mon. Wea. Rev., 112, *810–825.*

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

12. Harnack, R.P., 1986: Principles and methods of extended period forecasting in the United States, 1986. Monograph No. 1-86 of the National Weather Association, 36 pp.

13. Neubert, W.W., 1977: Multiple discriminant analysis. Selected topics in statistical meteorology. Air Weather Service (AWS-TR-77-273), *4-1 to 4-19.*

14. Laurmann, J. and L. Gates, 1977: Statistical considerations in the evaluation of climate experiments with atmospheric general circulation models. J. Atmos. Sci., 34, *1187–1199.*