The Effect of Intraseasonal Circulation Variability on Winter Temperature Forecast Skill*

KEITH W. DIXON

Geophysical Fluid Dynamics Laboratory/NOAA, Princeton University, Princeton, NJ 08540

ROBERT P. HARNACK

Department of Meteorology and Physical Oceanography, Cook College—New Jersey Agricultural Experiment Station, and Rutgers—The State University of New Jersey, New Brunswick, NJ 08903

(Manuscript received 12 July 1984, in final form 26 August 1985)

ABSTRACT

The prediction of winter temperatures in the United States from Pacific sea surface temperatures was examined using a jackknifed regression scheme and a measure of intraseasonal atmospheric circulation variability. Employing a jackknifed regression methodology when deriving objective prediction equations allowed forecast skill to be better quantified than in past studies by greatly increasing the effective independent sample size. The procedures were repeated on three datasets: 1) all winters in the period 1950–79 (30 winters), 2) the 15 winters having the highest Variability Index (VI), and 3) the 15 winters having the lowest VI. The Variability Index was constructed to measure the intraseasonal variability of five-day period mean 700 mb heights for a portion of the Northern Hemisphere. Verification results showed that statistically significant skill was achieved in the complete sample (overall mean percent correct of 39 and 59 for three- and two-category forecasts respectively), but improved somewhat for the low VI sample. In that case, corresponding scores were 44 and 64 percent correct. In contrast, the high VI sample scores were lower (34 and 58 percent correct) than for the complete sample, indicating that skill is likely dependent on the degree of intraseasonal circulation variability.

1. Introduction

This study was motivated by the idea that seasonal predictability of the atmosphere may be a function of the consistency and strength of some forcing mechanism external to the atmosphere, and that *intra*seasonal circulation variability is at least an indirect measure of this consistency. Specifically, the purpose was to determine if winter season temperature forecast skill for United States subareas, determined by verifying objective forecasts made using empirical methods, is related to winter 700 mb height variability. Previous studies (Harnack, 1979, 1982) indicated that November mean sea surface temperatures for the North Pacific and tropical Pacific could be used to skillfully predict winter temperature categories for United States subareas, based on independent testing of multiple regression equations for nine cases. Overall skill was modest (mean percent correct was 47% for three category forecasts), and was quite spatially dependent. A similar study by Barnett (1981) tends to confirm these assessments.

For the previously cited reasons, and because skill levels were both modest and still somewhat uncertain due to the smallness of the independent sample, the earlier studies were repeated in part, but with two major differences. The first was that statistical models were separately derived and verified for three samples using: 1) all available years, 2) low intraseasonal variability years, and 3) high intraseasonal variability years. The second major difference was the use of the so-called "jackknife" regression technique. The use of this method allows one to maximize the number of independent test cases. The procedures used are discussed in section 2.

There has not been a great deal of past research conducted concerning intraseasonal variability of tropospheric circulation. This lack of investigation is in spite of the belief held by some that the success of long range predictions may be significantly affected by intraseasonal variability. This point was mentioned by Madden (1977) in analyzing why long range predictability is greater for summer (due to less intraseasonal variability) than winter. In addition, Blackmon (1976) and Blackmon et al. (1977) found that when Northern Hemisphere 500 mb heights were subjected to different filters, most of the unfiltered time series' variance was contained in the low-pass filter, which was related to perturbations with periods of 10 to 90 days in length.

In an attempt to quantify intraseasonal variability and its relationship to seasonal-mean circulation, Harnack and Crane (1984) formulated an intraseasonal

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

variability index (VI) using individual season standard deviations of five-day mean 700 mb heights, which were normalized by long term standard deviations on a grid point by grid point basis. Principal components of five-day 700 mb data were correlated with concurrent intraseasonal variability, as measured by the VI. Among other things, the expansion of the midlatitude westerlies in winter was related to increased intraseasonal variability, while decreased variability was observed concurrently with contracted westerlies. Second, low intraseasonal variability for the spring and winter seasons tended to occur when the Aleutian low was well developed in the eastern and central Pacific area. The study reported here used VI to stratify winter cases so as to quantify the dependency of forecast skill on intraseasonal variability.

2. Procedure

a. Determination of intraseasonal variability

The variability index (VI) derived for use in the Harnack and Crane (1984) study was employed here. The dataset, from which variability indices were derived, consisted of daily 700 mb heights for the Northern Hemisphere as obtained on magnetic tapes from the National Climatic Data Center. A 148 point grid was selected to center on North America, and to include much of the North Pacific and Atlantic Oceans. This 148 point grid has 10 degree latitude by 10 degree longitude spacing, and extends from 0 degrees longitude westward to 160°E, and from the North Pole southward to 20°N.

Daily 700 mb heights were used to compute five-day averages at each gridpoint. Five-day averages were chosen so that much of the day-to-day variability in the data would be eliminated, leaving information on variability at a longer time scale, yet still allow for a reasonably long sample size for each session. Each of the four seasons were defined to contain 18 five-day periods. The winter season was defined as commencing on 1 December, spring on 1 March, summer on 1 June, and autumn on 1 September.

The data included the period from winter 1947/48 through winter 1979/80. After reducing the dataset size via the 148 point grid and the use of five-day averages, a variability index was derived. The VI was designed to represent the intraseasonal variability of five-day averages of 700 mb heights, over the 148 point domain, for each of the individual seasons of interest. The following procedure is repeated, with only minor clarification, from Harnack and Crane (1984) describing the derivation of the variability index:

1) Using the five-day 700 mb height data, standard deviations were computed at each grid-point for each *individual* season, by combining the 18 five-day period values.

2) Using these computed values, a long-term average standard deviation was computed at each grid-point for each season type (combining standard deviation values for all winters, springs, etc. separately).

3) Ratio values were then computed at each grid-point, for each *individual* season, by dividing individual season standard deviations by the long-term average standard deviation. This procedure normalized for the latitudinal and seasonal dependence of the standard deviations. When the ratio value is greater than one, the individual season is more variable than normal, while values less than one indicate below normal variability at the grid-point.

4) For each individual season, the overall grid variability was represented by one value (referred to as the VI), by spatially averaging the grid point values, after appropriate weighting by latitude.

The winter VI time series is shown in Fig. 1. Fluctuations and trends of VI are discussed in Harnack and Crane (1984).

b. Investigation of relationships between winter temperature forecasts and intraseasonal variability

1) DATA USED IN DERIVING TEMPERATURE FORECAST EQUATIONS

The data used as the predictands in the screening regression procedure were averaged temperatures for the months of December, January, and February. Seasonal temperatures were determined for each subarea shown in Fig. 2, by using data from all climatic divisions (CDs) within each of the fifteen subareas, calculating an average temperature for the entire subarea for each month, and then taking the simple arithmetic average of the three winter months.

The predictors used were those that were included previously in the most successful prediction models, namely North Pacific and tropical Pacific sea surface temperature (SST) fields. The specific dataset used included mean November North Pacific SST data on a 50 point grid, extending from 25 to 55°N latitude,

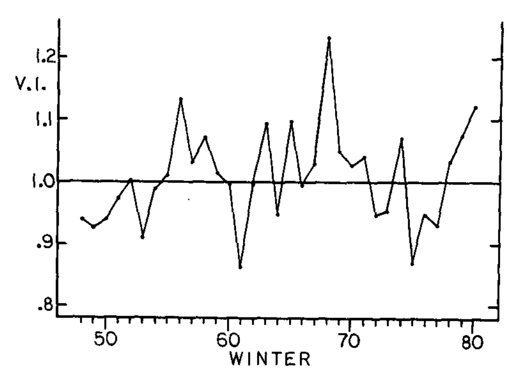

FIG. 1. Plot of winter Variability Index (VI) for 1947/48 through 1979/80.

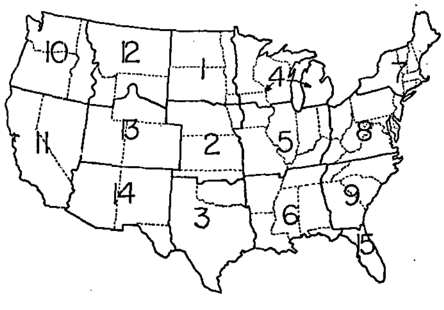

FIG. 2. Location of mean winter temperature subareas with identifying numbers.

125°W to 160°E longitude. The grid points were spaced 5 degrees latitude by 10 degrees longitude. The data were provided by the Scripps Institution of Oceanography (Climate Research Group).

Mean November SST for the tropical Pacific were used to define the second set of predictors. The raw data were on a staggered 34 point grid (using spacing of 5 degrees latitude by 10 degrees longitude), which covered the region from 10°N to 5°S latitude, 85°W to 165°W longitude. This dataset was first assembled for a study by Harnack et al. (1984).

2) SCREENING REGRESSION: FORMULATION, SKILL, AND SIGNIFICANCE

A total of six predictors were included in the predictor pool for use in the multiple linear regression models, with subarea winter temperatures as the predictands. The six consisted of the time-varying amplitudes of the first five principal components (PC) derived from November North Pacific SST data, and the time-varying amplitude of the first tropical Pacific SST PC. These were the only PC found to be significant (not random noise) at the 95 percent confidence level, as determined by Monte Carlo procedures (Overland and Preisendorfer, 1982). The five North Pacific PCs explain 72% of the variance and the first tropical Pacific PC explains 49% of the variance in their respective datasets. PCs were extracted from the correlation matrix.

Each of the thirty winter seasons from 1949/50 through 1978/79 inclusive, were categorized as having either low or high intraseasonal 700 mb height variability, by ranking the VI values and designating the lower half as "low" and the upper half as "high" intraseasonal variability. The 15 low variability winters had VIs ranging from 0.864 to 1.003, while VIs for the 15 high variability winters fell between 1.009 and 1.239. Thus, there were three datasets employed: 1) all years, consisting of the aforementioned thirty winters; 2) the low VI winters only, consisting of 15 winters; 3) the high VI winters only, also consisting of 15 winters.

In an effort to determine if winter season temperature forecast equations, empirically derived using sea surface temperature data as predictors, performed differently for winters with high intraseasonal variability versus winters characterized by low intraseasonal variability, the following procedure was used for each of the three datasets individually:

1) First, a screening regression procedure was performed, in which temperatures for each subarea were the predictands and the time-varying amplitudes of the six PCs discussed above were the potential predictors. For each subarea, the one, two, three and four-variable models which explained the largest portion of the variance in the dependent temperature data were noted (i.e., those having the highest R-squared statistic). Information concerning the best five- and six-variable models was not retained, since preliminary work showed that the addition of the fifth and sixth predictors to the equations resulted in quite minimal and insignificant increases in the R-squared statistic.

2) The screening regression procedure was modified to execute Monte Carlo tests. The methodology followed "shuffled" the predictands and randomly assigned them to predictor groups in that dataset, as suggested by Lund (1970). The screening regression procedure was then performed on each subarea's 15 or 30 sets of shuffled predictors and predictands, storing the R-squared values for the best one through four variable models. This shuffling of predictands was repeated 300 times. After the 300th trial, the mean and standard deviation of the R-squared statistic were computed for each of the 15 subareas and one through four variable models. These statistics were used to perform t-tests for each subarea. Of the four models computed in step 1, the one which had the highest t-score was declared to be the "best" model for predicting winter temperatures from November Pacific SST data, for the dataset of concern (all winters, high variability or low variability winters). This determined the number of predictors to be used in the prediction model for each subarea, as described in step three.

3) The regression models constructed in step 1 used all 30 or 15 years in the derivation process, and therefore produced R-squared values which give the variance of winter temperatures explained for a *dependent* sample. Less skill would be expected if these same equations were used to predict temperatures from an *independent* sample of winters. It is this "artificial predictability" (the amount by which the skill of forecasts diminish when tested on an independent sample) that must be addressed, if reliable skill levels are to be determined.

A procedure known as "jackknifing" or "cross-validation" allows for independent testing and therefore a direct assessment of skill, despite a relatively small sample size (Mosteller and Tukey, 1977). As applied in this study, the number of independent tests using the jackknife procedure equaled the number of winters in the dataset for which it was being applied (30 or 15). For the full dataset of 30 winters, a series of 30 regression equations were derived for each subarea (the number of variables in each subarea's models was set to match that of the best equation derived from the dependent data in steps 1 and 2. Each equation was derived from 29 of the winters in the dataset and tested on the one winter omitted. Given the 30 possible combinations of 29-leave-out-one, 30 separate forecasts were made and tested. In each instance, an independent test was conducted. The verification results presented represent aggregate skill determined in this manner. The same procedure was carried out for the two datasets consisting of 15 sets of predictors and predictands, with there being 15 combinations of 14-leave-out-one.

4) Though the regression equations produced predictions of winter temperature anomalies, forecasts are more commonly issued in terms of temperature categories. Therefore, skill levels of both the two and three category forecast types were evaluated. For each subarea, the median and upper and lower terciles of the thirty years of winter temperatures were determined. The tercile boundaries of the observed data were used to establish the limits of the below, near normal, and above categories for the three category forecast verifications. The median served to discriminate between the below and above categories for the two category forecasts. All forecast and observed temperatures were converted from numeric to categorical form. Categorical skill is given as the percentage of cases for which the predicted and observed categories were identical (percent correct).

The binomial distribution was applied in performing one-tailed tests using the normally distributed large sample z test statistic,

$$z = \frac{\hat{p} - p_0}{(p_0 q_0 / n)^{1/2}} \tag{1}$$

where $\hat{p}$ is the portion of forecasts verified as correct, p_0 is the portion expected to be correct by chance (0.33 or 0.50), $q_0 = 1 - p_0$, and n is the number of degrees of freedom.

The percent correct skill levels for the three datasets were each individually tested using the null hypothesis that the calculated skill equalled that which would be achieved had the forecasts been drawn from a population of randomly generated categorical forecasts (i.e., 33% for the three-category verifications and 50% for the two-category type forecasts). The alternative hypothesis that the skill level of the regression equations was greater than that expected by random chance was accepted if the z score calculated was greater than the critical value of the test statistic (1.282 for the 90% confidence level, 1.645 for the 95% confidence level).

Similarly, to test whether the low variability verification scores were significantly greater than those of the high variability dataset, one-tailed z-tests were performed where

$$z = \frac{\hat{p}_L - \hat{p}_H}{[\hat{p}\hat{q}(1/n_2 + 1/n_H)]^{1/2}} \tag{2}$$

where the L subscript denotes the low variability dataset and H the high, $\hat{p} = (\hat{p}_L + \hat{p}_H)/2$ and $\hat{q} = 1 - \hat{p}$.

H_0: $\hat{p} = p_0$; prediction score verifications are not significantly better than random chance.

H_A: $\hat{p} > p_0$; prediction scores are better than random chance would dictate,

where $\hat{p}$ is the portion of forecasts verified as correct and p_0 is the portion of forecasts expected by random chance to be correct (0.33 for 3-category verification, 0.50 for 2-category verification).

The number of degrees of freedom (DF) used in calculating the critical values was set at 180 for the "all" years sample (6 DF times 30 winters) and 90 for the low and high VI samples (6 DF times 15 winters). This assumes temporal independence between winters; the same assumption which was applied when using the binomial distribution to determine local significance for the skill levels of individual subareas (DF equal to the sample size). This is a reasonable assumption considering the work of Madden (1977), who found that the spectra of winter season temperatures for most locations in North America resembled "white noise".

The value of six spatial DF was found by using a Monte Carlo approach. A total of 1000 randomly generated series of 30 numbers were each correlated with the winter temperature time series for the 15 subareas. For each of the trials, the percentage of the area of the continental United States represented by temperature subareas which achieved statistically significant correlations at the 95 percent confidence level according to F-tests, was noted. Combining the 1000 trials produced a frequency distribution of the area expected to achieve local significance at the 95 percent confidence level by random chance. From this distribution, the critical percent area was designated as that value above which only five percent of the generated distribution fell. The number of DF was found by reversing the procedure described by Livezey and Chen (1983), in which they used the binomial distribution to find the critical percent area, given the spatial DF and confidence level. Here the area and confidence level were known, so the spatial DF was determined.

3. Results

Table 1 and Figs. 3–5 summarize the main results derived by using the jackknife procedure previously described for each of the three datasets, then verifying the forecasts in both the three- and two-category formats. Overall mean percent correct scores are given in Table 1, and the spatial distribution of mean percent correct for each of the three datasets are shown in Figs. 3–5.

Table 1 shows that the jackknifed regression models for all three datasets performed significantly better than random chance when verified in the two-category for-

TABLE 1. Mean percent correct obtained by applying jackknife regression technique to each dataset. Asterisks denote statistical significance at the 90 percent confidence level; double asterisks, the 95 percent confidence level.

Dataset	Verification: Three-category	Verification: Two-category
All years	39.3**	59.3**
Low VI	44.0**	64.0**
High VI	33.8	58.2*

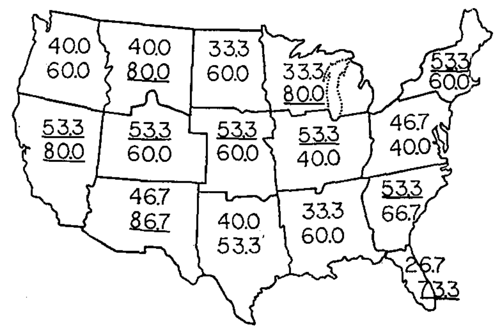

FIG. 4. As in Fig. 3 except dataset includes low VI years only.

mat across the United States, as did the three-category predictions for the all years and low VI datasets. Comparing the high and low VI results reveals that the 44 percent correct achieved by the low VI prediction equations in the three-category verification is significantly better than the high VI results, at the 95 percent confidence level. However, one can be only about 90 percent certain that the two-category, low VI results represent a skill level greater than that of the high VI data.

The "best" regression equations derived from all the data in the three datasets, as described in steps 1 and 2 of the procedure, appear in Table 2, for each temperature subarea as labeled in Fig. 2. The PC amplitudes used as predictors were not normalized prior to use in regression equations. When the jackknifing procedure was used, the regression coefficients produced were generally quite stable. This was especially true for the equations representing the eastern and southern portions of the United States. The coefficients most affected by which year was left out in the course of the jackknife procedure were typically those derived for subareas in the Northern Plains, and West of the Rocky Mountains. Regression coefficients calculated for the high variability dataset tended to be less stable than the others.

Those regression equations with just one variable and a relatively small regression coefficient, tended to produce forecasts of near normal temperatures more than one-third of the time. In the all-years dataset this was most evident in the three southwestern subareas, where about two-thirds of the forecasts were for near normal temperatures. For regions such as these, the two-category verification scores might be examined to assess if winter temperatures tend to be related to the November SST data, since the three-category verification scores are not likely to stray far from thirty-three percent correct.

The effectiveness of using the jackknifing procedure to account for artificial predictability in this study is evident in that the prediction equations derived from all 30 years of data had 45.6% correct versus 39.3% for the jackknifed results. Similarly, the two-category verifications would have been correct in 66.4% of the 450 tests, instead of the 59.3% correct score achieved when the 29-leave-out-one method was used.

These results indicate that: 1) November SST can be used to objectively forecast area-averaged winter temperature categories with significant, though modest skill, and 2) forecast skill appears to be a function of the intraseasonal circulation variability, with lower skill in the higher variability winters. Perhaps the high VI winters tend to be those in which the strength of Pacific air–sea interactions is weak, so that forecast skill based on the use of SST predictors is less. The first conclusion is not new, since it was made in the earlier papers as

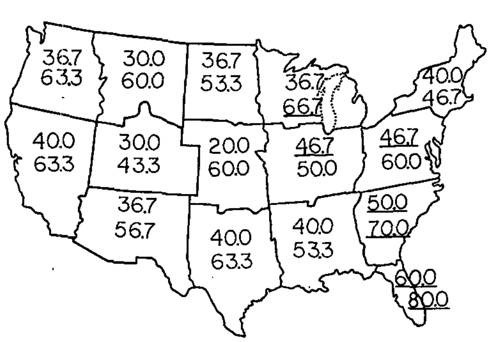

FIG. 3. Mean percent correct scores for categorical forecasts produced using the jackknife regression technique on all years (1949–79). Top number applies to three-category forecasts and bottom number to two-category forecasts. Underlined denotes statistical significance at the 90 percent confidence level.

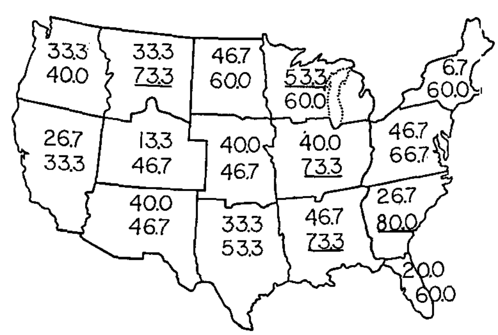

FIG. 5. As in Fig. 3 except dataset includes high VI years only.

TABLE 2. Winter temperature forecast equations derived using all observations in the dependent sample. Predictors PAC1 through PAC5 refer to the first five principal components calculated from the North Pacific SST data, and TROP1 refers to the first tropical Pacific principal component.

	Dataset					
	All-years		Low VI		High VI	
Subarea number	Predictors in best model	Regression coefficients	Predictors in best model	Regression coefficients	Predictors in best model	Regression coefficients
1	PAC1	0.28	PAC2	−0.46	PAC1	0.77
					PAC5	−1.23
					PAC4	−0.88
					TROP1	−0.53
2	PAC3	−0.37	PAC3	−0.43	PAC5	−1.04
3	TROP1	−0.19	TROP1	−0.23	TROP1	−0.35
	PAC3	−0.38	PAC3	−0.51	PAC5	−0.84
	PAC2	0.25	PAC2	0.30	PAC4	−0.46
	PAC4	0.26	PAC4	0.32		
4	PAC3	−0.41	PAC3	−0.49	PAC1	0.45
	PAC1	0.19	PAC1	0.30	TROP1	−0.48
	PAC2	0.20			PAC4	−0.54
	PAC4	0.25				
5	PAC3	−0.62	PAC3	−0.78	TROP1	−0.35
	PAC2	0.46	PAC2	0.55		
	PAC4	0.52				
6	PAC3	−0.60	PAC3	−0.58	TROP1	−0.45
	PAC2	0.56	PAC2	0.61		
	PAC4	0.55	TROP1	−0.26		
	PAC1	−0.25	PAC4	0.55		
7	PAC2	0.29	PAC3	−0.70	PAC2	0.26
	PAC3	−0.33	PAC2	0.42	TROP1	−0.24
	PAC4	0.34			PAC1	0.22
	PAC5	0.25			PAC5	0.44
8	PAC2	0.51	PAC2	0.69	TROP1	−0.34
	PAC4	0.53	PAC3	−0.75		
	PAC3	−0.45				
9	PAC2	0.56	PAC2	0.65	TROP1	−0.45
	PAC4	0.55	TROP1	−0.23		
	PAC3	−0.41	PAC4	0.54		
	PAC1	−0.28	PAC3	−0.43		
10	PAC5	−0.50	PAC2	−0.46	PAC5	−0.86
	PAC2	−0.26			TROP1	0.27
11	PAC2	−0.15	PAC2	−0.44	PAC5	−0.53
			TROP1	−0.19	TROP1	0.33
					PAC1	−0.23
12	PAC5	−0.49	PAC2	−0.63	PAC5	−1.04
	PAC2	−0.26				
13	TROP1	−0.11	TROP1	−0.32	PAC5	−0.55
			PAC2	−0.47	TROP1	0.40
					PAC2	0.25
					PAC1	−0.22
14	TROP1	−0.12	TROP1	−0.24	PAC2	0.27
					TROP1	0.27
					PAC1	−0.21
15	PAC2	0.43	PAC2	0.53	PAC1	−0.49
	PAC1	−0.31	PAC4	0.53		
	PAC4	0.48	TROP1	−0.18		
	PAC3	−0.33	PAC1	−0.18		

well, however in this work the number of forecast verifications has been lengthened considerably (from 9 to 30), making the significance assessment and level of skill more reliable.

Examination of Fig. 3 reveals that most of the skill is contributed from the eastern subareas, especially the extreme southeast, Florida, Ohio Valley, and Great Lakes regions. However, using just the low VI winters, significant skill was found in most sections (Fig. 4), including the Far West. Figure 5 indicates that for high VI winters skill is again confined mainly to the East.

As an additional measure of skill, the correlation between predicted and observed temperature anomalies was computed for each of the data samples, for each subarea. The variation in space of R-squared (i.e., explained variance) tended to be similar to that depicted in Figs. 3–5 for the mean percent correct statistic. For the complete sample (30 years), R-squared values were between 14 and 28% for the previously identified "best" subareas (mainly in the East), and less than 10% for the remainder of the country. The correlation coefficients themselves were positive in all but a few cases. For the low VI sample, the "best" subareas had R-squared ranging from 15 to 45% (for six of the 15 subareas), with the remainder at less than 10%. In contrast, the high VI sample results showed no subarea with R-squared greater than 8%.

4. Summary and conclusions

This study attempted to better quantify forecast skill for winter temperatures from that obtained in previous studies by using a jackknife regression approach, so that each case in the sample could be used as an independent forecast case. The predictors were principal components of the November North Pacific and tropical Pacific sea surface temperature fields, and the predictands were area-averaged December–February temperature anomalies for fifteen United States subareas. The procedures were repeated on three datasets: 1) all winters in the period 1950–79 (30 winters), 2) the 15 winters having the highest Variability Index (VI), and 3) the 15 winters having the lowest VI. The VI was constructed to measure the intraseasonal variability of five-day period mean 700 mb heights for a portion of the Northern Hemisphere. The purpose of stratifying the regression analyses was to determine if forecast skill was a function of intraseasonal variability.

The verification results, given as mean percent correct skill scores for forecasts formated in two and three categories, showed that statistically significant skill was achieved in the all years sample (overall mean percent correct of 39 and 59 for three- and two-category forecasts respectively), but improved somewhat for the low VI sample. In that case the corresponding scores were 44 and 64 percent correct. In contrast, the high VI sample scores were lower (34 and 58 percent correct) than for the complete sample, indicating that skill is likely dependent on the degree of intraseasonal circulation variability. Other measures of prediction skill, such as the correlation between predicted and observed temperature anomalies, also confirmed the difference in skill when prediction models were stratified by high versus low intraseasonal variability. This result was expected based on physical reasoning but still needed to be objectively confirmed. Of course, this information cannot be used as a forecast aid at this time since intraseasonal variability must be predicted first, and there is no current evidence that this can be done skillfully. The geographical distribution of skill had the largest contribution from the East for the complete and high VI samples, but showed less regional preference for the low VI sample.

Acknowledgments. New Jersey Agricultural Experiment Station, Publication No. D-13507-6-84, supported by State funds and by the Climate Dynamics Program, Division of Atmospheric Sciences, National Science Foundation, under Grant ATM-8217215.

The authors wish to thank the Climate Research Group of Scripps Institution of Oceanography for supplying the sea surface temperature dataset, Jeremi Harnack for miscellaneous tasks, John Lanzante for advice, and Ruth Smith for typing assistance.

REFERENCES

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

Blackmon, M. L., 1976: A climatological spectral study of the geopotential height of the Northern Hemisphere. *J. Atmos. Sci.*, **33**, 1607–1623.

——, J. Wallace, N. C. Lau and S. Mullen, 1977: An observational study of the Northern Hemisphere wintertime circulation. *J. Atmos. Sci.*, **34**, 1040–1053.

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

——, 1982: Objective winter temperature forecasts: An update and extension to the western United States. *Mon. Wea. Rev.*, **110**, 287–295.

——, and M. W. Crane, 1984: Intraseasonal tropospheric circulation variability and its relation with concurrent atmospheric and oceanic fields. *Mon. Wea. Rev.*, **112**, 1768–1781.

——, J. R. Lanzante and J. Harnack, 1984: Associations among the tropical Pacific wind and sea surface temperature fields and higher latitude circulation. *J. Climatol.*, **2**, 267–290.

Livezey, R., and W. Chen, 1983: Statistical field significance and its determination by Monte Carlo Techniques. *Mon. Wea. Rev.*, **111**, 46–59.

Lund, I. A., 1970: A Monte Carlo method for testing the statistical significance of a regression equation. *J. Appl. Meteor.*, **9**, 330–332.

Madden, R. A., 1977: Estimates of the autocorrelations and spectra of seasonal mean temperatures over North America. *Mon. Wea. Rev.*, **105**, 9–18.

Mosteller, F., and J. Tukey, 1977: *Data Analysis and Regression.* Addison Wesley, p. 586.

Overland, J. E., and R. W. Preisendorfer, 1982: A significance test for principal components applied to a cyclone climatology. *Mon. Wea. Rev.*, **110**, 1–4.