Long-Lead Seasonal Temperature Prediction Using Optimal Climate Normals

JIN HUANG, HUUG M. VAN DEN DOOL, AND ANTHONY G. BARNSTON

Climate Prediction Center, National Centers for Environmental Prediction, Washington, D.C.

(Manuscript received 23 June 1993, in final form 7 July 1994)

ABSTRACT

This study is intended to determine the spatially varying optimal time periods for calculating seasonal climate normals over the entire United States based on temperature data at 344 United States climate divisions during the period of 1931–1993. This is done by verifying the seasonal climate normals as a forecast for the same season next year. The forecast skill is measured by the correlation between the predicted and observed anomalies relative to the 30-yr normal. The optimal time periods are chosen to produce the highest correlation between the forecasts and the observation.

The results indicate that generally (all seasons and all locations) *annually updated* climate normals averaged over *shorter* than 30-yr periods are better than the WMO specified *30-yr* normal (updated only every 10 years), in terms of the skill in predicting the upcoming year. The spatial pattern of the optimal averaging time periods changes with season. The skill of optimal normals comes from both the annual updating and the shorter averaging time periods of these normals. Using optimal climate normals turns out to be a reasonably successful forecast method. Utility is further enhanced by realizing that the lead time of this forecast is almost one year. Forecasts at leads beyond one year (skipping a year) are also reasonably skillful.

The skill obtained from the dependent verification is lowered to take account of the degradation expected on independent data.

In practice the optimal climate normals with a variable averaging period were found to be somewhat problematic. The problems had to do primarily with the temporal continuity and spatial consistency of the forecasts. For the time being, a constant time period of 10 years is used in the operational seasonal temperature forecasts for all seasons and locations.

1. Introduction

In meteorology, climatology and their applications to many other fields, the term ''climate normals'' has become standard. Nevertheless, the notion of a climate normal is somewhat controversial (Kunkel and Court 1990). To many users, a climate normal may imply that the climate is stationary and has no trends, in which case it would only be a matter of sampling a long enough (preferably infinite) period to determine the normals as accurately as possible. However, it has been known (or suspected) that the climate may not exactly be constant, in which case normals should be calculated from recent data only. Notwithstanding their ambiguity, climate normals have been found useful in many long-term planning applications. For example, crop choices, agricultural practices, utility rates, etc., are based in part on assumptions about climatic conditions in the near future. For such purposes, climate normals are often used as the best possible baseline ''prediction'' of the future climate. We place ''prediction'' in quotation marks because usually (in meteorology) prediction implies an ability to forecast departures from normal with skill.

Then the question is: what are the optimal climate normals in terms of their predictive skill? Here *optimal climate normals* (OCN) are defined as the average over the most recent K years, where K is selected such that the K-year average gives the best prediction for the next year. In other words, what is the best climate persistence forecast for next winter (for example)? Is it last winter's value, or the average over the K most recent winters? An OCN acts as a low-pass filter, finding a characteristic timescale of variability and identifying the associated signal. Because the climate is not stationary, a large K does not necessarily produce the lowest errors in forecasts based on climate normals.

The traditional climate normals are based on a specified 30-yr period, which is updated at the start of each decade (such as 1951–1980, 1961–1990, etc.), according to WMO's recommendation.[1] That is, they are ''fixed'' 30-yr normals that are shifted forward once

Corresponding author address: Dr. Jin Huang, Climate Analysis Center, National Centers for Environmental Prediction, W/NMC51, Washington, D.C. 20233.

[1] WMO has recommended only 1901–1930, 1931–1960, and 1961–1990 as normals, but the in-between 30-yr averages are so common and in widespread use that they are treated here also as WMO normals.

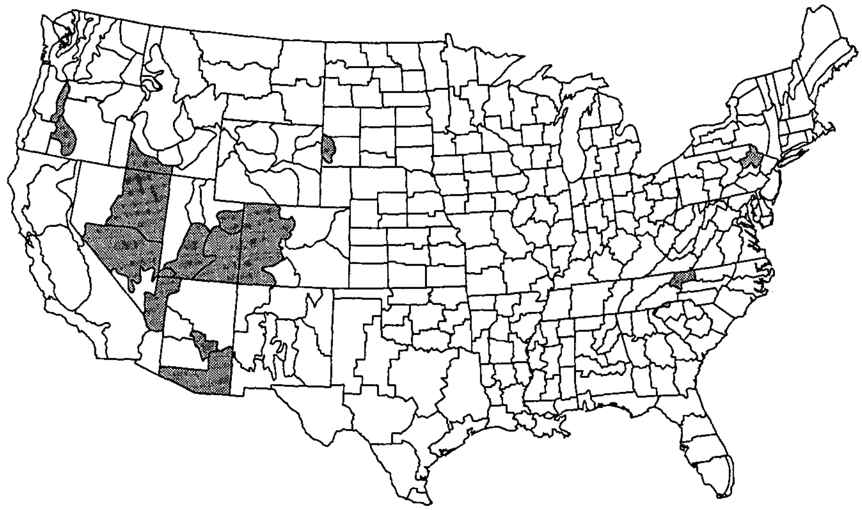

FIG. 1. Maps of the 344 U.S. climate divisions. Here 17 climate divisions (shaded) are excluded in the analyses due to data quality.

every 10 years. The advantage of following the WMO recommendation is that it allows for universal usage of an agreed upon standard of reference. Anomalies (departures from the normal) can thus be used unambiguously to describe abnormally cold–warm years or seasons. However, there is no obvious reason why the 30-yr average would be the best climate prediction at all places, for all weather elements. Indeed, many studies have found that 30-yr normals have less predictive skill than averages over the most recent K years, where $K < 30$ (e.g., Court 1967–1968; Lamb and Changnon 1981; Dixon and Shulman 1984; Kunkel 1987).

The idea of OCN is not at all new. In fact, many ideas about the use of optimal climate averages are as old as climatology itself and were summarized in a series of scientific reports by Court (1967–1968). However, most previous studies were based on data at only a few locations in the United States and rarely considered tests on independent data. In this study, we examine the entire continental United States to determine the skill of the seasonal temperature forecasts based on OCN, that is, based on persistence of the mean anomaly over the most recent K years for the same season that we are predicting. We also discuss the seasonal and spatial distribution of K and test results on independent data. Recently, we have started to apply OCN as one of the prediction approaches in operational seasonal temperature forecasts at the Climate Prediction Center (CPC, formerly Climate Analysis Center) of the National Center for Environmental Prediction (NCEP, formerly National Meteorological Center) (Van den Dool 1994). Problems encountered in operational practice are also considered and reported in section 8.

OCN provides a forecast with a lead time of almost one year (more precisely the lead is nine months, e.g., we can predict the spring of 1997 at the end of spring 1996 or whenever the data for 1996's spring are available). Seasonal forecasts at such a long lead are a rarity so far. Because WMO normals are not updated every year, OCN can yield higher predictive skill than WMO normals both because $K < 30$ and because we average over the *most recent* K years rather than over a period of K years that ended up to 13 years ago. The problem of the aging of the WMO normals has only recently been considered (Angel et al. 1993).

What distinguishes this article from most previous work on OCN is also an attempt to test the results on independent data or, more generally, to determine an a priori skill estimate. This turns out to be an unusually difficult problem.

2. Data and definition

This study uses seasonally averaged daily temperature during 1931–1993 for 344 United States' climate divisions. The geographic distribution of the United States' climate divisions is shown in Fig. 1. In the analyses, 17 climate divisions (shaded in the map) are excluded due to data-quality problems (Cayan et al. 1986), resulting in 327 climate divisions.

Suppose T_i, $i = 1, \cdots, n$, is a time series of data for a given climate division for a particular season at yearly intervals for n years. A second series of backward-looking averages $\overline{T}_{i,k}$ is constructed as follows:

$$\overline{T}_{i,k} = \frac{1}{k}\sum_{j=1}^{k} T_{i-j}, \quad (1)$$

for $k = 1, 2, \cdots, 30$ and for $i = 31, 32, \cdots, n$. For data during 1931–1993, $n = 63$. The k-index represents the number of years over which the climate average is calculated. Climate means for each value of k are then used to make forecasts of the upcoming year, and this is done for each year from year number $i = 31$ (i.e., 1961) onward.

The skill of OCN to predict the upcoming year is measured here by the correlation between the forecast anomaly and observed anomaly during the verification period of 1961–1993. The correlation is defined as

$$\mathrm{COR_{dep}}(k) = \frac{\sum_{i=31}^{n} \hat{T}_i^f \hat{T}_i^{ob}}{[\sum_{i=31}^{n} (\hat{T}_i^f)^2 \sum_{i=31}^{n} (\hat{T}_i^{ob})^2]^{1/2}}, \quad (2)$$

where

$$\hat{T}_i^f = \bar{T}_{i,k} - C_{\mathrm{WMO}}, k = 1, 30$$
$$\hat{T}_i^{ob} = T_i^{ob} - C_{\mathrm{WMO}}, \quad (2a)$$

where the referenced 30-yr mean (C_{WMO}) is defined in Table 1, considering the situation in practice where the official normals are updated every ten years with several years delay (e.g., the 1961–1990 normal has been used only after 1993). Later in section 5, we will discuss the difference in skill when using an annually updated 30-yr normal as the reference.

The correlation defined in Eq. (2) is for dependent data (hence $\mathrm{COR_{dep}}$) because we searched for K, that is, fitted the forecast model to the data to estimate one parameter. Nowadays, a cross-validation (CV) method is often used to establish expected skill on independent future data. The CV method implies that each of the years to be verified (which are 1961–1993) is held out in turn when calculating the optimal K and then used as the forecast target. However, in this case, because of the configuration of the long set of predictor years being immediately followed by the predictand year, a CV test is not straightforward. In this paper, the dependent correlation ($\mathrm{COR_{dep}}$) is lowered to account for the shrinkage expected on independent data [see Eq. (7) in Barnston and Van den Dool 1993]:

$$\mathrm{COR_{indep}} = \frac{N\mathrm{COR_{dep}}}{N-1} - \frac{1}{(N-1)\mathrm{COR_{dep}}} \quad (3)$$

for $\mathrm{COR_{dep}} > 1/(N)^{1/2}$ where $N = 20$ is the effective sample size of the dependent verification during 1961–1993. (The reason for not using the original $N = 33$ will be given in section 7.) Unless stated otherwise, we show $\mathrm{COR_{indep}}$ in the next sections.

Note that the anomalies are defined as the departures from the WMO specified 30-yr average (C_{WMO}), and thus skill is defined as the improvement over the skill obtained when using the aging 30-yr mean as the forecast. The optimal averaging time K is that k for which the correlation is maximum in the dependent period.

TABLE 1. Definition of WMO 30-yr normals (C_{WMO}).

Year of verification	WMO specified 30 years
1961–1963	1931–1960 (1921–1950 unavailable)
1964–1973	1931–1960
1974–1983	1941–1970
1984–1993	1951–1980
>1993	1961–1990

There are other criteria for choosing an optimal K; they will be discussed in section 4.

The method employed here to determine K is mathematically identical to that used by Harnack et al. (1984), Roads and Barnett (1984), and Van den Dool (1985), who, in the context of monthly forecasting, sought to optimize the skill of the forecast for the next month by persisting an average of the last K *days* of the previous month. An important technical difference between the climate normals approach and that used for monthly forecasts in Van den Dool (1985) is that in the former case the magnitude of the climate anomalies goes to zero as k goes to 30. This does not happen in the monthly application.

This manuscript deals with only one special case: using a K-year average to predict a single upcoming year. The more general situation would be to use a K-year average to predict an average over n years with a lead of m years. Court (1967–1968, part V) found that the optimal period for predicting a single value m years beyond the end of the averaging period is m years shorter than for predicting the next year's value, that is, K-m. This is opposite to the experience in short-term forecasting (Van den Dool 1985). In the latter case, K tends to become larger as the future average (n) increases and also as the lead (m) increases.

3. Spatial distribution of optimal averaging time

We have examined the spatial distributions of the optimal averaging period K for 12 seasons shifted by one month, but only show the standard four seasons (MAM, JJA, SON, and DJF) in Fig. 2. The optimal averaging period K is substantially less than 30 for all seasons in almost all areas. In other words, the WMO's 30-yr climatic normals are usually not the best predictor for the upcoming year's temperature for the given season. This result supports and further extends results from the previous studies (e.g., Lamb and Changnon 1981; Dixon and Shulman 1984; Kunkel 1987).

The spatial distributions of K are different for different seasons, as found earlier by Court (1967–1968). However, the pattern changes more or less smoothly month by month, which can be seen in maps for all 12 seasons (only four are shown). In fall low K ($K \leqslant 5$) values are found mainly in the western half of the country, while in winter K is lowest ($K \leqslant 5$; down to $K = 1$) in the east. In southern Illinois in summer, me-

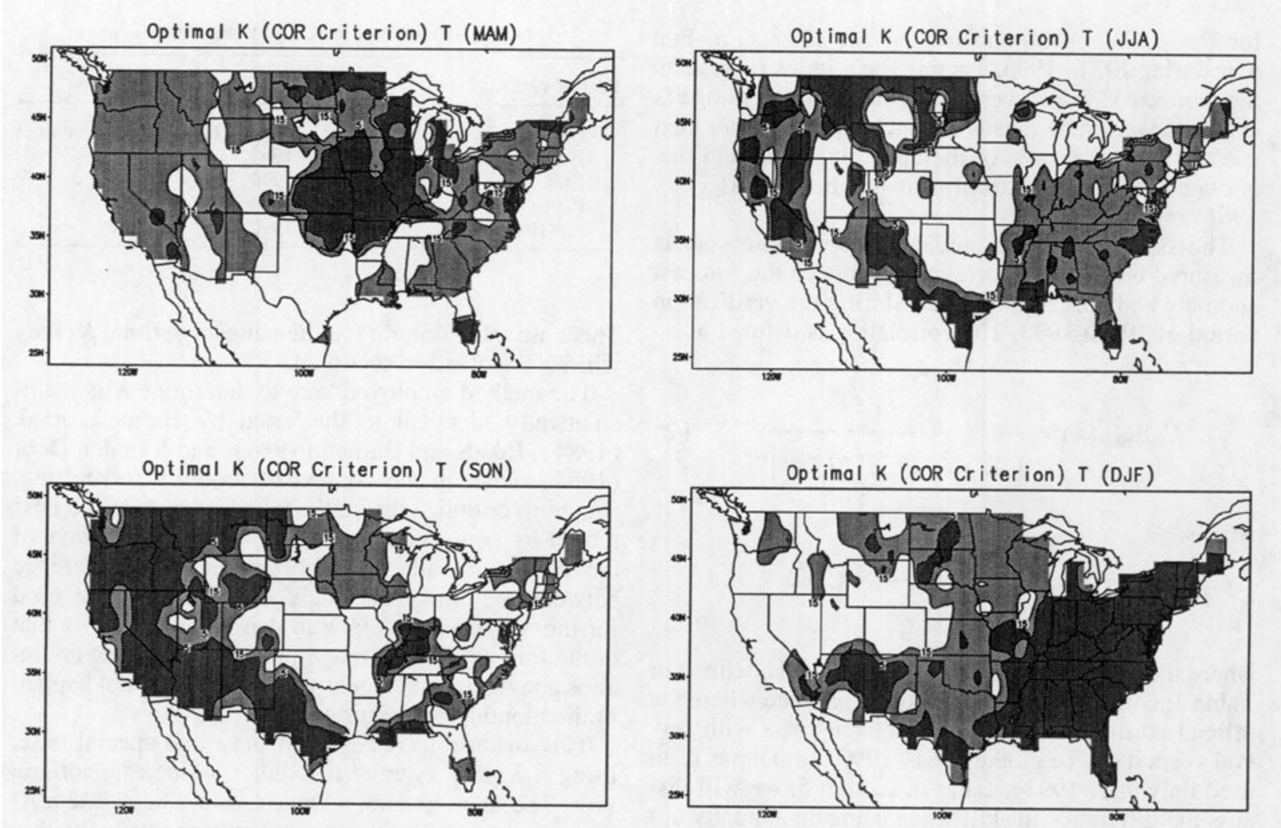

FIG. 2. Spatial distribution of optimal averaging time K (in years) for climate normals for (a) MAM; (b) JJA; (c) SON; and (d) DJF. Heavily shaded areas are for $K \leq 5$ and lightly shaded areas are for $5 < K \leq 15$. Blank areas are for $15 < K < 30$.

dium-term ($5 < K < 15$) averages have the most predictive information, which agrees with the results of Lamb and Changnon (1981). The K parameter is a crude measure of the spectrum of interannual variation in temperature; that is, the timescale of variation. When K is small there is marked year to year persistence. A longer K might imply significant interdecadal variations in the climate. Interestingly, some similarity can be seen with the results of the studies in long-range potential predictability (Madden and Shea 1978; Madden 1981; Shea and Madden 1990).

Figure 3 shows the correlation between the OCN forecasts and the observations. The correlations are derived using Eq. (2) and appropriately lowered by Eq. (3). Only correlations greater than 0.3 are plotted, the rest of the field is "masked out" to be blank. Because skill is defined to show improvement over skill using the WMO normal, generally higher correlation is most likely found where the optimal K is much different from 30, that is, much lower than 30 in our case.

4. Different criteria

To measure the predictive accuracy of the climate normals, several other reasonable criteria have been proposed (e.g., Lamb and Changnon 1981; Dixon and Shulman 1984; Kunkel 1987), besides the correlation (COR). The optimal averaging period that produces the minimum root-mean-square (rms) error has been considered (e.g., Kunkel 1987). A mean absolute difference (ABS) between forecast and observation has also been used (Lamb and Changnon 1981). Yet another proposed method is, for each k, to count the number of times for which the k-year average makes the best prediction, that is, has the smallest error for each of the years to be verified. The optimal K is the one with the highest frequency of being the best (Kunkel 1987).

To make a comparison, we have plotted in Fig. 4 the three quantities, COR, rms, and ABS as a function of k at several chosen climate divisions for JJA. The rms error of C_{WMO} is also plotted, which will be further discussed in section 5. Here the COR shown is calculated from Eq. (2), that is, dependent correlation. The result of the "frequency" method is not shown but will be mentioned later. For each calendar season, we picked three locations with $K \leq 5$, $5 < K \leq 15$, and $15 < K \leq 30$, respectively, with the guidance of Fig. 2 and Fig. 3, the latter figure showing the locations having skill. Because K is regionally consistent (Court 1967–1968), individual climate divisions are representative. Figure 4 only shows the JJA season as an

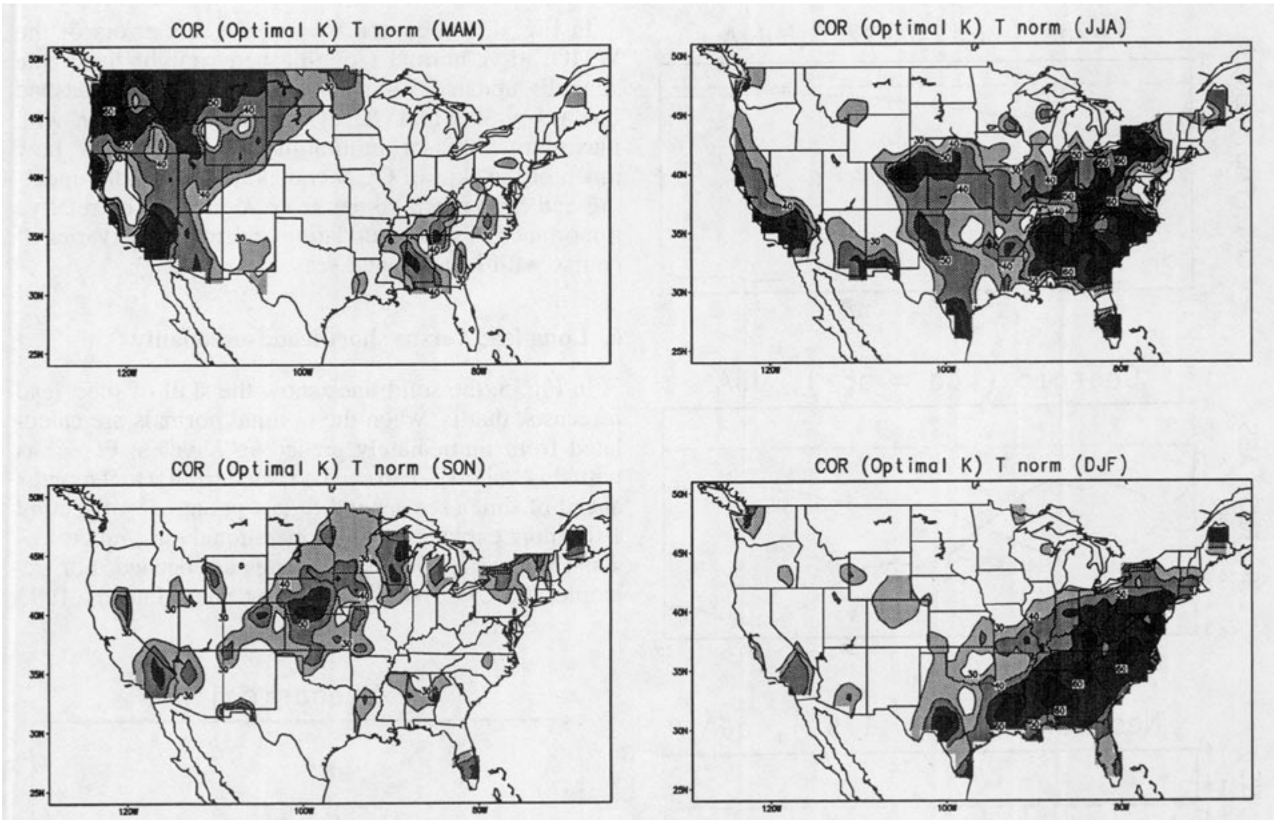

FIG. 3. Spatial distribution of the cross correlation (×100) between the forecast (optimal normal) and the actual values for (a) MAM; (b) JJA; (c) SON; and (d) DJF. Correlations smaller than 0.3 are not plotted. The contour intervals are 0.1.

example. Generally, COR is consistent with rms and ABS; that is, small rms or ABS tends to correspond to large COR, and vice versa. Near the optimum the criterion often depends weakly on k (as found also by Court). The optimal K values shown in Fig. 2 (particularly for Georgia and Nebraska) could thus easily be altered by a few years by 1) changes in dataset (i.e., sampling variability) or 2) using a different criterion. By the same token, though, changing K by a few years for a given criterion does not change the skill of the forecast very much. We conclude that COR, rms, and ABS all give essentially the same answer.

We found that the "frequency" method leads to very short K (1 or 2 years) in most of the United States' climate divisions in each season. This is presumably due to the last 1- or 2-yr persistence of climate anomalies, which is seen quite often (Kunkel 1987). However, the other three methods rarely pick the shortest periods to be optimal. In summer, for example, K is nine in Georgia based on all three criteria. The difference between the "frequency" method and the other three methods has also been noted before (Lamb and Changnon 1981; Kunkel 1987). Although persistence of short-lived anomalies is important and a noteworthy feature of the United States climate, a sudden change in pattern can occur as well, which would result in a large rms error for small K. Being conservative, we prefer the COR, rms, or ABS criteria.

It has become conventional to use correlations to describe forecast skill at NCEP. The rms, ABS, and COR are interrelated measures, as exemplified in Fig. 4. The relationship between COR and rms is discussed for standardized data in Barnston (1992) and in more general terms in Murphy and Epstein (1989).

5. Difference between annually updated and aging 30-yr mean

The positive correlations shown in Fig. 3 indicate that the annually updated climate normals averaged over shorter than 30-yr periods have higher predictive skill than the aging 30-yr mean. However, it is not clear at this point whether the skill comes mainly from the shorter averaging periods or from the annual updating. In order to identify the source of the improvement in skill, we compare the skill of OCN using two different 30-yr means as the reference: One is an annually updated 30-yr mean, that is, $\overline{T}_{i,k=30}$ [see Eq. (1) for the definition] (referred to as scenario I, which, we believe, is used in most of the literature on OCN), and the other is the aging 30-yr mean ($C_{\rm WMO}$) as defined in Table 1 (referred to as scenario II, which was used above and also in practice) (O'Lenic 1994; Van den Dool 1994).

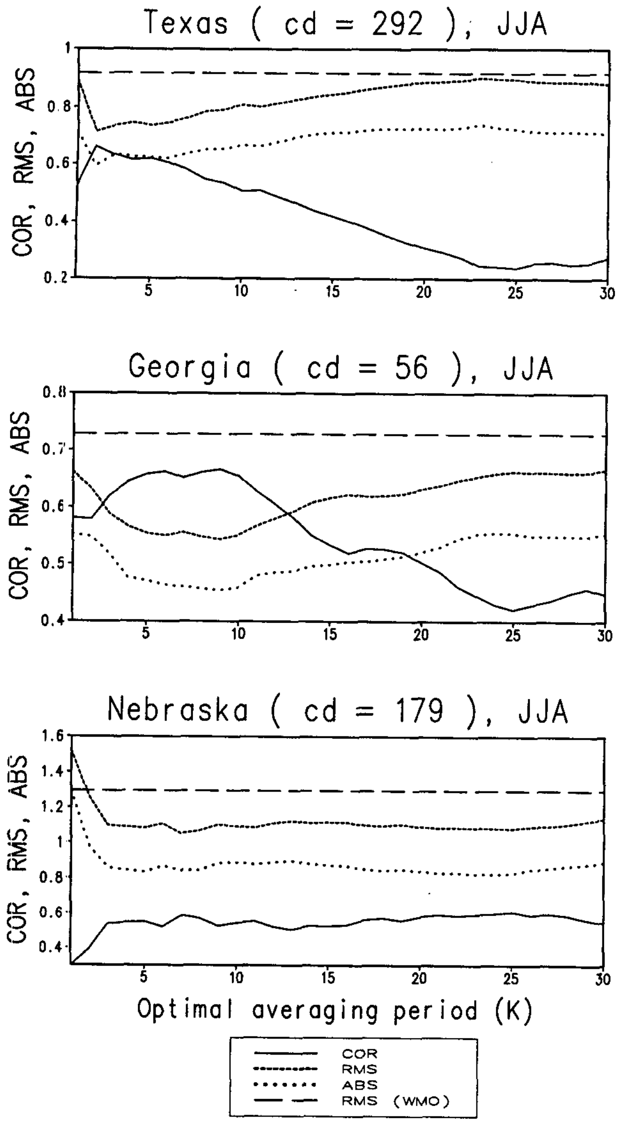

FIG. 4. Three different criteria, rms error (short dashed line, unit: degrees Celsius), absolute error (ABS, dotted line), correlation (COR, solid line; unit: degrees Celsius), and rms error of WMO 30-yr normal (long dashed line) as a function of averaging period k in (a) Texas (JJA) with $K \leq 5$; (b) Georgia (JJA) with $5 < K \leq 15$; and (c) Nebraska (JJA) with $15 < K < 30$.

Figure 5 shows the spatially averaged correlation over the United States as a function of season from NDJ (season 1) to OND (season 12) for scenario I and scenario II (see solid lines in both panels). (The spatial average also includes all climate divisions with correlations less than 0.3). The higher skill in scenario II than in scenario I indicates that the skill of OCN is partly due to the annual updating of the normals. The positive sign of the correlation in scenario I suggests that the skill also comes from using the shorter averaging periods to calculate the optimal normals.

In Fig. 4, the comparison of the rms errors of the WMO 30-yr normal (long-dashed straight line), the annually updated 30-yr normal (value of short-dashed line at $k = 30$), and OCN (short-dashed line at K with maximum COR or minimum rms) also show how much of the skill of OCN comes from annually updating and how much comes from $K < 30$. The relative importance of annual updating and smaller K varies of course with location and season.

6. Long lead versus short lead; seasonality

In Fig. 5, the solid lines show the skill of short lead forecasts, that is, when the optimal normals are calculated from immediately preceding K years. Forecasts with this "short" lead can be issued from 0 to 9 months ahead of time (see second to last paragraph of section 1 for more explanation). In operational long-range seasonal forecasts, even longer leads are needed. For example, the CPC/NCEP is issuing since January 1995

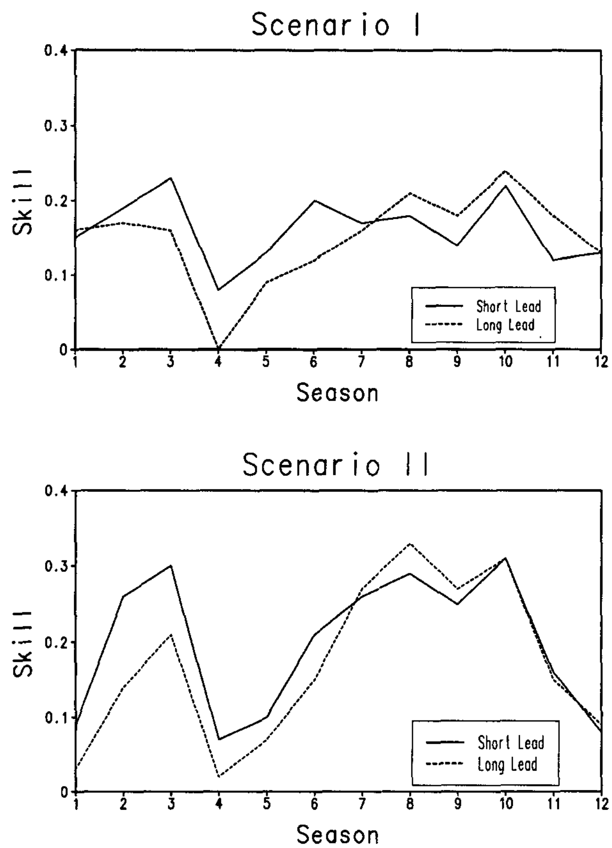

FIG. 5. The correlation spatially averaged over United States as a function of season from NDJ (season 1) to OND (season 12). The top panel is for the case with the annually updated 30-yr mean as the reference; the lower panel is for the case with the aging WMO 30-yr mean as the reference. The solid line is for short lead (0 to 9 months) and the dashed line is for the longer leads (skipping one year, that is, 10 to 21 months).

long-lead forecast with a lead time of up to 13 months (O'Lenic 1994; Van den Dool 1994). Moreover, we have to take into account the practicalities related to the unavailability of recent climate division data in real time.

Forecasts with the longer lead (10 to 21 months) are made by skipping one year between the predictors and the predictand. We found that the difference in skill between the short lead and long lead is not very large (dashed line in Fig. 5). This certainly implies that the skill of OCN is not very sensitive to the lead time. In other words, forecasts beyond one year are about as skillful as those with shorter leads (Note that for long leads, K may differ from the K for short leads).

The skill of OCN changes with season, as seen from Fig. 5 and also from Fig. 3. Basically, the skill is higher in summer and winter during the verification period (1961–1993), especially under scenario II. The skill is lower in transition seasons, particularly in fall and early spring.

Scenario II is relevant to the operational forecasts. It is interesting to note the skill maxima in late winter and summer. This appears to be a feature of many forecast methods (e.g., Barnston 1994).

7. Test on independent data

The correlations shown in Fig. 3 and Fig. 5 are a proxy for cross-validated correlations obtained here from the dependent correlation [Eq. (2)], lowered according to estimated shrinkage expected on independent data [Eq. (3)]. We follow this theoretical procedure because it is not clear how to perform cross validation for OCN forecasts in the way it is done on a regression-type forecast (Barnston and Van den Dool 1993). The shrinkage of the skill depends on the sample size and the correlation itself. The higher the dependent correlation and the larger the sample size N, the smaller the shrinkage. Given $N = 20$ in Eq. (3), for example, $COR_{indep} = 0.14$ for $COR_{dep} = 0.3$, while $COR_{indep} = 0.42$ for $COR_{dep} = 0.5$.

Note that the N in Eq. (3) is the effective sample size N_{eff}, estimated to be 20 for 1961–1993. We now explain how we estimated N_{eff}. To determine N_{eff}, we first carried out an old-fashioned independent verification for the 8-yr (1986–1993) period. The truly independent skill can be used in conjunction with the dependent skill to estimate N_{eff}, using Eq. (3). For each verified season during 1986–1993, the optimal K and the dependent skill are determined based on the data ending one year before (e.g., to forecast a season in 1986, we use 1961–1985 as forecast targets to find the K and to determine the dependent skill) for each climate division and season. Because there are only eight years, we do the comparison by pooling data in space and time. The *dependent skill* averaged over the eight years and all seasons and all climate divisions is *0.35.* The *actual skill* (i.e., the correlation between the actual forecasts and the observed anomalies) is *0.19.* In calculating those correlations, forecast and verification anomalies are divided by local standard deviations to prevent the pattern correlation from being dominated by climate divisions with large standard deviations.

We now estimate the shrinkage using Eq. (3). The 8-yr averaged dependent sample size N ($N = 25$ for 1961–1985, $N = 26$ for 1961–1986, etc) is 28.5, and thus the estimated independent skill, using $N_{eff} = 28.5$, would be $COR_{indep} = 0.26$ (rather than 0.19) given $COR_{dep} = 0.35$. Apparently, Eq. (3) underestimates the shrinkage when using the original sample size (28.5). The likely reason is that the forecasts (i.e., K-year-averaged seasonal climate normals) will be similar from year to year especially when K is large. When forecast skill is relatively high, the observations must also tend to have some year to year persistence. We thus found that the effective N should be about 17 in order to get $COR_{indep} = 0.19$. Thus, the ratio of effective N to original N is about 0.60. Applying this ratio to $N = 33$ (for 1961–1993), we get the effective sample size $N = 20$, which is used in Eq. (3) to produce Figs. 3 and 5. But we have to remember that this correction is an approximation based on 8-yr independent verification only. Note that the same feature that makes OCN a success (interannual persistence of anomalies) hinders in showing OCN's statistical significance.

In practice (Van den Dool 1994), forecasts will be made only at locations that are known in advance to have usable skill. When including only climate divisions with dependent correlations [Eq. (2)] of at least 0.4, the annually averaged skill for the eight independent years becomes 0.30 (0.58, 0.10, 0.43, and 0.38 for MAM, JJA, SON, and DJF, respectively). The shrinkage of 0.4 down to 0.3 is consistent with $N_{eff} = 20$.

An alternative test on independent data is to use the K-values that are optimal for long leads for short-lead forecasts. The similarity in skill for short and long leads (see Fig. 5) and particularly the similarity of the K-values for short and long leads points to robust results. In this case we can use the entire 1961–1993 dataset. For the short-lead forecasts the dependent COR, annually averaged, equals 0.34. When using the K optimal for the long lead, this short-lead skill drops to 0.24. This drop is quite consistent with the shrinkage approach above and suggests that $N_{eff} = 20$ for 1961–1993 is a very reasonable estimate.

8. Practical considerations in operational forecasts

When applying OCN in practice we encountered several new problems that led us to reconsider some of its design aspects for operational forecasts. The problems had to do primarily with the temporal continuity and spatial consistency of OCN forecasts. Above we made the point that skill of OCN does not depend too much on the precise value of K, that is, particularly when optimally $K = 4$, $k = 3$, or $k = 5$ will still give nearly the same skill. That may be true, but the forecast itself *does* change a lot. We were thus faced with abrupt changes in forecast at the same location when going, for example, from DJF to JFM. A similar problem exists in space. Even though K is regionally consistent and skill has regional scales, the forecast itself is sensitive to K when K is less than about five. This leads to hard to accept positive and negative forecast anomalies side by side on a map for the same target season.

Variable K is obviously not always a blessing. To be conservative and simple in practice, we recalculated a single K value for all 12 seasons and all 327 CDs combined. The COR for this grand sample is shown as a function of k (Fig. 6a). The results are very telling. Using $K = 30$ always and everywhere explains most of the gain over using WMO normal (COR = 0.24). Since we are fairly confident that the optimal K is shorter than 30, we pick $K = 10$ (although $K = 23$ would do nearly as well) and so raised the COR to 0.28. Because of the large sample used, there should be negligible difference here between dependent and independent sample results. Figure 6b shows the skill of OCN (COR_{dep}) during 1961–1993 for variable K (i.e., optimal K) and $K = 10$. The skill of OCN (optimal K) is of course higher, but this may be mostly sampling. For now we sacrifice the option of *variable* K, because (a) it helps relatively little, (b) it is difficult to cross validate the results (see section 7), and (c) it leads to problems in real-time forecasts.

Operational methods should preferably be reliable in the sense that their skill should not vary wildly. While the skill in Fig. 6b is far from constant, there are only five (out of 33) years where COR < 0. Generally 0.1–0.3 is to be expected. We note with interest in Fig. 6b is that the skill of OCN is considerably higher during the 1960s featuring COR = 0.5 averaged over all seasons and all locations. Also, 1991–1992 featured high skill, but 1993 had COR = 0. It would in interesting to investigate why.

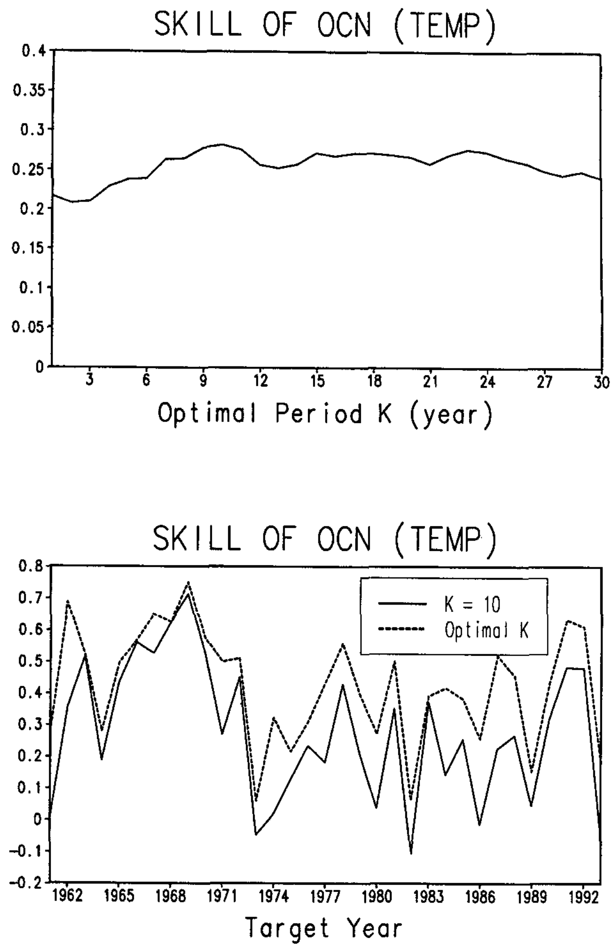

FIG. 6. (a) The correlation averaged for all seasons and all climate divisions as a function of K for short lead. (b) The skills of OCN during 1961–1993 for $K = 10$ (solid line) and optimal K (dashed line).

9. Summary and discussion

This study has examined seasonal temperature data at 344 U.S. climate divisions during the period of 1931–1993. The purpose is to determine the spatially varying optimal time periods for climate normals over the United States when the normals are used as forecasts for the same season in the following year and the year after. The skill is measured by the correlation between the predicted and observed anomalies. The skill is defined as the improvement over the WMO specified 30-yr normal (updated only every 10 years). While this method appears merely to "beat the system," it is fair because it formulates forecasts precisely in the way done by other methods and expected by the operational community. The optimal time periods are chosen to maximize the skill in predicting the target season in the upcoming year. The use of OCN to predict the next year's temperature for a given season provides a forecast with a lead time of almost one year.

The results indicate that for almost all seasons and locations annually updated climate normals averaged over shorter periods are better than the WMO specified 30-yr normal, in terms of the skill in seasonal temperature prediction in the upcoming year. The improvement in skill is contributed by both using shorter averaging periods and annual updating. The spatial patterns of the optimal averaging time periods (K) change smoothly with season. The details of the spatial and seasonal variations in K certainly contain sampling error. On the other hand, the forecast skill is not terribly sensitive to changes in K by a few years.

The correlation obtained from the dependent verification is lowered to take account of the degradation expected on independent data. This correction incorporates an estimation of the effective sample size and the dependent correlation. Two independent tests are conducted in this study. The first one is to use the period of 1961–1985 to estimate K and to use the period of 1986–1993 as the verification period, which is used to lower the dependent correlation. The second method is to use the K optimal for the long lead when making short-lead forecasts. It allows to use the entire 1961–1993 for verification and the shrinkage is consistent with the other independent test.

Forecasts beyond one year have been found to be about as skillful as those with shorter leads. It is found that the optimal K by pooling all seasons and locations is 10 for the short lead and 9 for the long lead, that is, K is *one* year smaller when predicting *one* year beyond. This confirms Court's finding (1967–1968) mentioned at the end of section 2.

The skill of OCN changes with season. The skill is higher in summer and winter, based on the verification during 1961–1993. The seasonality of the skill could be different for some other periods. As discussed in section 7, for example, the skill was the highest in spring for 1986–1993.

For precipitation we also found climate normals over shorter periods to have higher predictive skill than the 30-yr average (not shown). However, the optimal averaging periods have spatial patterns different from those for temperature (especially in summer), and the skill is lower. We plan to try using the median instead of the mean for the precipitation study, as suggested by Court (1967–1968).

The optimal normals could be determined by different criteria proposed in literature. Methods based on COR, rms, or ABS show similar characteristics, while the "count" method (Kunkel 1987) tends to choose the shortest period to be optimal.

Included in our OCN forecasts are also unknown nonmeteorological features, such as artificial changes in measurements. For example, if a thermometer suddenly acquired a 10°C positive bias, K would soon become very small.

In practice we found a variable K to be problematic and, for the time being, settled for a constant K for all seasons and locations (K = 10 for temperature and K = 15 for precipitation).

Acknowledgments. We are grateful to Edward S. Epstein, Kenneth E. Kunkel, Robert E. Livezey, A. James Wagner, the editor (P. J. Lamb), and the two anonymous reviewers for their helpful comments and suggestions.

REFERENCES

Angel, J. R., W. E. Easterling, and S. W. Kirtsch, 1993: Towards defining appropriate averaging periods for climate normals. *Climatol. Bull.*, **27**, 29–44.

Barnston, A. G., 1992: Correspondence among the correlation, rmse, and Heidke forecast verification measures: Refinement of the Heidke score. *Wea. Forecasting,* **7,** 699–709.

——, 1994: New long-lead forecasts: Seasonal and geographic variations in skill. *Proc. 19th Annual Climate Diagnostics Workshop,* College Park, MD, 412–415.

——, and H. M. van den Dool, 1993: A degeneracy in cross-validated skill in regression-based forecasts. *J. Climate,* **5,** 963–977.

Cayan, D. R., C. F. Ropelewski, and T. R. Karl, 1986: *An Atlas of United States Monthly and Seasonal Temperature Anomalies December 1930–November 1984.* NOAA-U.S. Climate Program Office, 244 pp.

Court, A., 1967–68: Climatic normals as predictors: Parts I–V. Sci. Rep. Air Force Cambridge Res. Lab., Bedford, MA, Contract AF19(628)-5176. [NTIS AD-657 358, AD-686 163, AD-672 103, AD-687 137, AD-687 138].

Dixon, K. W., and M. D. Shulman, 1984: A statistical evaluation of the predictive abilities of climatic averages. *J. Appl. Meteor.*, **23,** 1542–1552.

Harnack, R., C. Kleupfel, and R. Livezey, 1984: Prediction of monthly 700 mb height using antecedent and simulated height predictors. *Proc. Ninth Annual Climate Diagnostics Workshop,* Corvallis, OR, 305–316.

Kunkel, K. E., and A. Court, 1990: Climatic means and normals—a statement of the American Association of State Climatologists (AASC). *Bull. Amer. Meteor. Soc.,* **71,** 201–203.

Lamb, P. J., and S. A. Changnon, 1981: On the "best" temperature and precipitation normals: The Illinois situation. *J. Appl. Meteor.,* **20,** 1383–1390.

Madden, R. A., 1981: A quantitative approach to long-range prediction. *J. Geophys. Res.,* **86,** 9817–9825.

——, and D. J. Shea, 1978: Estimates of the natural variability of time-averaged temperatures over the United States. *Mon. Wea. Rev.,* **111,** 1696–1703.

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

O'Lenic, E., 1994: A new paradigm for production and dissemination of the NWS's long lead-time seasonal climate outlooks. *Proc. 19th Annual Climate Diagnostics Workshop,* College Park, MD, 408–410.

Roads, J. O., and T. P. Barnett, 1984: Forecasts of the 500-mb height using a dynamically oriented statistical method. *Mon. Wea. Rev.,* **112,** 1354–1369.

Shea, D. J., and R. A. Madden, 1990: Potential for long-range prediction of monthly mean surface temperatures over North America. *J. Climate,* **3,** 1444–1451.

Van den Dool, H. M., 1985: Prediction of daily and time-averaged temperature for lead times of 1–30 days. *Reprint, Ninth Conf. on Probability and Statistics in Atmospheric Sciences.* Virginia Beach, VA, Amer. Meteor. Soc., 144–155.

——, 1994: New operational long-lead seasonal climate outlooks out to one year: Rationale. *Proc. 19th Annual Climate Diagnostics Workshop,* College Park, MD, 405–407.