Short and Long Range Air Temperature Forecasts near an Ocean

H. M. VAN DEN DOOL

Department of Meteorology, University of Maryland, College Park, MD 20742

J. L. NAP

Royal Netherlands Meteorological Institute, DeBilt, The Netherlands

(Manuscript received 6 September 1984, in final form 31 January 1985)

ABSTRACT

Using a predetermined statistical scheme, forecasts are made of the daily air temperature (AT) at San Diego, starting from local antecedent information concerning AT and sea surface temperature (ST) only. These forecasts are verified by calculating skill scores (S) over 1948–79. In this maritime area such simple schemes turn out to have high S for lead times up to a month and small but positive S out to a year. Differences in S of the various one predictor schemes ($ST \rightarrow AT$, $ST \rightarrow ST$, $AT \rightarrow AT$, $AT \rightarrow ST$) are discussed; $ST \rightarrow ST$ is far superior to any of the other three. For most schemes S is low in late summer; this is attributed to the shallowness of the ocean's mixed layer in that season. The effects of time averaging the predictor and/or predictand are discussed for the $ST \rightarrow AT$ scheme. For long enough lead times averaging appears to improve forecast skill. The localness of the prognostic information carried by ST is investigated by comparing S for San Diego and an inland station (Escondido). At a forecast lead time of three days S decreases by 50% over a distance of 25 km. Further analysis shows that this decay is primarily caused by a decrease in skill of daily maximum temperature forecasts.

In view of the similarity of the present results to those obtained at the Dutch coast, we conclude that local information about the state of the surface has probably enough prognostic potential to be incorporated in existing operational schemes of short and long range air temperature forecasts near oceans and lakes.

1. Introduction

Suppose we want to make a weather forecast for a given location A. What data to describe the present do we need in order to make such a forecast? It is almost imperative to refer to Fig. 18.9 in Smagorinsky (1974) in this context. The message of that figure is that for short-range prediction one needs data for a small area (of which A is roughly in the center) and that with increasing lead time data for larger and larger areas are required. Ultimately for long enough lead times one needs to specify initial data over the whole globe including the upper oceans and the earth's surface.

In view of this, the present paper is unusual. We will try to make forecasts of the air temperature for lead times ranging from one day to one year *starting from antecedent local information only*. This will be done for two sites at the West coast of a continent, namely San Diego (California) and Den Helder (The Netherlands). These sites are chosen because the nearby ocean is thought to provide the slowly changing lower boundary conditions needed to make a skillful forecast of subsequent air temperatures. One may describe our attempts as single station forecasting although (in contrast to the habits in this old profession) we will use surface observations only.

The basic experiment will be to predict daily mean air temperatures (AT) from antecedent instantaneous sea surface temperature (ST), notation $ST \rightarrow AT$. This is done by using an *a priori* defined statistical forecast method which is outlined in Section 2. Most of the results discussed here are for San Diego's AT starting from ST at the nearby Scripps Pier in La Jolla. It has been known for a long time (Hubbs, 1948; Roden, 1960) that ST and AT are related in this area. Hubbs noted that sea temperature at La Jolla corresponded closely to air temperature records for San Diego; his evidence was based on monthly mean data for 1915–48. Roden (1960) using updated datasets showed that anomalies in monthly mean AT and ST were correlated, the coherence increasing from 0.2 at 1 cycle per 2 months to 0.8 at the multiyear time-scales. At none of these time-scales did the ocean appear to lead the atmosphere or reverse; the two media were roughly in phase.

Restricting ourselves to using just local initial data means that the results of this study may not be directly applicable in practice where one would like to use a Numerical Weather Prediction (NWP) model as a primary forecasting tool, at least for the short-range prediction. However, we will encounter here an example of amazingly high skill for both short and long lead times using simple information and simple methods only. This naturally leads to the question whether a combination of NWP and local information (incorporated via Model Output Statistics say) would lead to better forecasts.

The simplicity of our approach allows us to address several interesting questions.

1) What is the skill of air temperature forecasts made by local methods at places where external memory (the ocean) is so close? We will investigate lead times of up to a year.

2) How do the following schemes compare? $ST \rightarrow AT$ (basic one), $AT \rightarrow ST$, $ST \rightarrow ST$ and $AT \rightarrow AT$.

3) Is there seasonality in the skill?

4) Does it help to use two predictors, say $(AT, ST) \rightarrow AT$ rather than ST or AT alone?

5) Does averaging (in time) of predictor and/or predictand improve the skill of forecasting AT?

6) What is the difference in skill in forecasting maximum and minimum temperature from antecedent ST?

7) How does the skill of the $ST \rightarrow AT$ scheme decay away from the area where ST is measured.

2. Data, forecast method and verification

In this study we use pairs of datasets. One such pair consists of i) daily mean (≡ average max/min) air temperature (AT) at San Diego's airport (Lindbergh Field) and ii) daily sea surface temperature (ST) at the Scripps Pier in La Jolla (one observation, in the morning). These two sites are about 20 km apart which is close enough to call the procedure to be discussed local forecasting. We will use data over 1948–79. The second pair consists of daily mean (≡ average over 6 or 8 observations) ST at lightvessel *Texel* (North Sea) and daily mean (based on 24 hourly observations) AT in Den Helder (at the Dutch coast) for the period 1890–1977. These two sites are 25 km apart. The latter pair of datasets has been discussed earlier in Van den Dool and Nap (1981), and will be used here only occasionally for a comparison with the results of San Diego. In Section 3g, we will use AT at Escondido, which is 25 km inland rather than right at the edge of the ocean. Finally in Sections 3f and 3g daily maximum and minimum temperatures at San Diego and Escondido are used.

As a first step the data for San Diego, Escondido and Scripps Pier were converted from °F (whole degrees) to °C (tenths of a degree). We then detrended the AT and ST series by subtracting a 10 day/10 year running mean from each of the daily values, i.e.,

$$T(k, j) = T^*(k, j) - \sum_{j-4}^{j+5} \sum_{k-4}^{k+5} T^*(k, j)/100 \qquad (1)$$

where T^* refers to raw data, k is the day ($k = 1, 365$; 366 is the same as 1 etc.) and j is the year ($j = 1, 32$). For $j < 5$ or $j > 28$ the average was taken over the first and last ten years respectively. The aim of applying (1) is to take out the annual cycle and multiyear trends in a smooth way. The resulting anomaly series is used throughout the remainder of this paper and referred to as AT and ST.

No attempts were made to explicitly correct the data for known changes in observational site, instruments and changes in exposure. Such changes may have introduced some spurious low-frequency variability in the time series. One of the objectives of taking out a ten-year running mean (rather than a grand mean) is to exclude such artificial low-frequency variability. However, it is realized that taking out trends (whichever way it is done) is always arbitrary. Another check on the influence of data problems is provided below by considering the results for the first and second half of the record separately.

Both for background information and later reference, Fig. 1 shows the frequency distribution of (detrended) ST and AT in the month of October. The climatological mean temperatures (over 32 years) for October (vertical arrows) have been added to ST and AT in plotting this histogram. As can be seen in Fig. 1, AT tends to be both higher and more variable than ST. With the exception of one or two winter months ST tends to be lower than AT (Tont, 1981), thereby contributing to the static stability of the lower atmosphere. Except for the summer months AT has a much larger variance than ST.

Frequency distributions such as the one shown in Fig. 1 have been made for each calendar day both for ST and AT. Given the distribution, each day during 1948–79 has been categorized in terciles (predictand) and quintiles (predictor), that is (1) B(elow), N(ear normal) and A(bove), the chances being as close as possible to ⅓ and (2) M(uch) B(elow), N(ormal) and M(uch) A(bove) the chances being as close as possible to 20, 60 and 20%. Our forecasting scheme then reads (for the basic experiment)

$$ST(0) \in \binom{MA}{MB} \rightarrow AT(\tau) \in \binom{A}{B} \qquad (2)$$

where $\tau \geq 0$. In words: if ST at day 0 falls in the class $MA(MB)$ then we forecast that AT, at all lead times τ, will be in $A(B)$. There is no forecast if $ST(0) \in N$ (we will briefly discuss the effect of this in Section 3), i.e. there is a forecast during 40% of the time. Expression (2) is a predetermined form of damped cross-persistence. This is likely to give some results for San Diego since persistence at long time scales in this area is higher than anywhere else in the United States (van den Dool, 1984). Expression (2) is for the $ST \rightarrow AT$ scheme but it is straightforward to make the $ST \rightarrow ST$, $AT \rightarrow AT$ and $AT \rightarrow ST$ schemes in a similar fashion. In all of these cases quintiles apply to the predictor and terciles to the predictand.

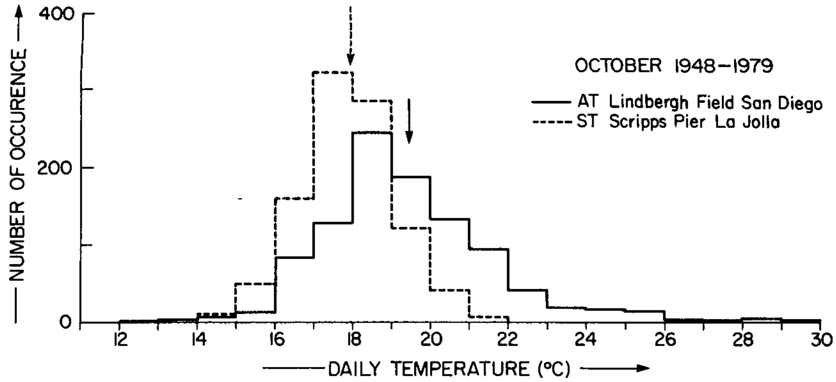

FIG. 1. Frequency distribution of daily values of anomalies in air temperature at Lindbergh Field (San Diego) and anomalies in sea surface temperature at the Scripps Pier at La Jolla in the month of October. The period is 1948–79. The 32-year mean October temperature has been added as a reference (the vertical arrows). The anomalies, however, are relative to a mean, defined by Eq. (1), that changes smoothly with calendar date and year. Total cases: 992. The number of occurrence is given for 1°C intervals, i.e., $12.0 \leq AT \leq 12.9$°C etc.

The forecasts are verified in the following way. A forecast is a hit if the class is predicted correctly; there are no partial hits. From a large number (M) of forecasts the skill score (S) can be calculated by

$$S = \frac{H - C}{M - C} \tag{3}$$

where H is the number of hits and C is the number of hits expected by chance which in our case is close to $M/3$.[1] So S is 1 for perfect forecasts and 0 for forecasts made without (detectable) insight in the problem. We will study S for lead times τ ranging from 0 (specification rather) to a year, as a function of season, for daily values, for maximum and minimum temperature separately, for time averaged temperatures (both predictor and predictand) and for the two geographical locations.

3. Results

a. *Year-round results*

In Fig. 2 some year-round results of the $ST \rightarrow AT$ scheme are shown for lead times up to two months. The number of forecasts that went into the skill score calculation (for one τ) is roughly 40% of 365×32. The curve based on detrended data (lower part of Fig. 2) is the basic experiment. As can be seen ST specifies AT with a skill score of 38%. With increasing lead time S drops below 20% at 6 days but remains 10% or more up to 60 days. This is truly unusual because it means that in San Diego there is skill in daily air temperature forecast beyond the range of what is usually called deterministic predictability (2–3 weeks), skill that can be effectuated with extremely simple means. The fact that the skill levels off at 10–15% (rather than zero) indicates that there is coherent low-frequency variability in the AT and ST series. This also means that we are dealing here with a two time-scale problem, which is to be expected for air–sea interaction.

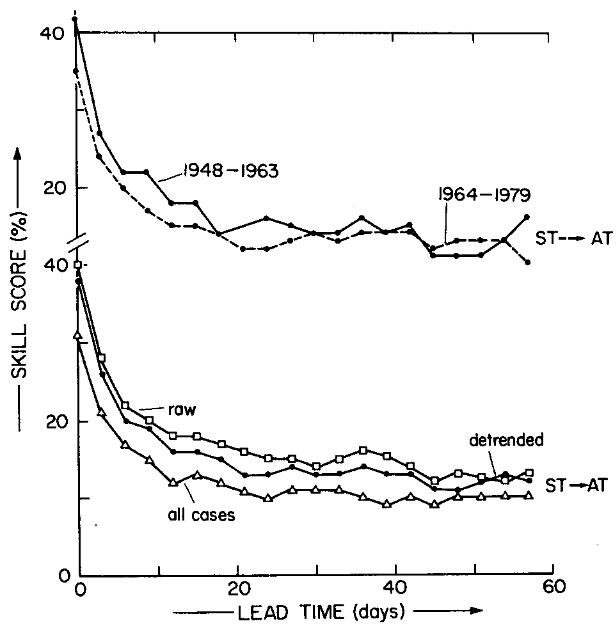

FIG. 2. Year-round skill score (%) of forecasts of the air temperature in San Diego from antecedent sea surface temperature in La Jolla ($ST \rightarrow AT$), for lead times from 0 to 60 days. Forecast scheme for the raw data as well as the detrended data is based on Exp. (2). The upper part of the figure shows verification for 1948–63 and 1964–79 separately, the lower part is based on all data. For convenience the symbols are plotted every three days.

[1] It is not always possible to distribute 32 temperature values over 3 (5 or any) equally probable classes. Therefore we kept track of the exact distributions over 1948–79 and used the occurrences in each class during 1948–79 in the calculation of C. This rules out artificial skill.

The effect of taking out the trend is shown in Fig. 2 by comparing the raw and detrended curves. Detrending takes away some of the skill that, most likely, has to be appreciated as artificial. Our way (or any) of detrending [Eq. (1)] is arbitrary of course. We rerun the skill score calculation with 1 day/10 year and 30 day/10 year climatologies and found essentially the same results. The curve labelled "all cases," refers to an experiment where both *ST* and *AT* were categorized in a two-class system, above and below median. The forecast then reads [in the spirit of (2)]

$$ST(0) \in \begin{pmatrix}\text{Above Median}\\ \text{Below Median}\end{pmatrix} \rightarrow$$

$$AT(\tau) \in \begin{pmatrix}\text{Above Median}\\ \text{Below Median}\end{pmatrix}$$

With this procedure a forecast can be made in all cases, rather than 40% of the time [with (2)]. Including near-normal predictors lowers the skill somewhat but the difference is not dramatic. In the upper part of Fig. 2 the forecasts of the basic experiment [detrended, Exp. (2)] are verified separately for 1948–63 and 1964–79. This gives some indication of the sampling error in the year-round skill score. The rms difference between the two sets of 16 years is 3.1% which can be interpreted as $\sigma_{16}\sqrt{2}$ where σ_{16} is the sampling error in 16 year mean skill. So σ_{32} should be about 1.6%.

In Fig. 3 a comparison is made between the $ST \rightarrow AT$, $ST \rightarrow ST$, $AT \rightarrow AT$ and $AT \rightarrow ST$ schemes, out to a lead time of 60 days. In general the *ST* predicting itself is far superior to any of the other three combinations. If one's target is to predict the air temperature it does not seem to make very much difference whether one starts from antecedent *ST* or *AT*; they both contain a similar amount of information about future *AT*. For $\tau < 25$ days $AT \rightarrow AT$ performs somewhat better than $ST \rightarrow AT$ but the reverse becomes true for longer lead times. One wonders whether we observe here the familiar finding that for short range forecasting the initial state of the atmosphere is most important while for longer lead times the state of the surface becomes the predominant source of information. Figure 3 also indicates that, for lead times up to a month, it is easier to forecast *ST* from *AT* than *AT* from *ST*. In other words, on short time-scales it is mainly the atmosphere that drives the ocean, rather than the other way around. Because he used monthly mean data Roden (1960) could not detect this lead-lag relationship.

Results of the $ST \rightarrow AT$ and $AT \rightarrow AT$ schemes are repeated in Fig. 4 along with skill scores out to a year and a comparison with *Texel*–Den Helder for the first month. At San Diego the skill remains positive (although very small) out to about a year for both schemes, a truly remarkable case of persistence of anomalies. The encircled values are 10-day averages of the skill of daily forecasts made in exactly the same way for Den Helder. During the first 10 days the forecast skill is equally high in these very different areas but beyond 20 days knowledge of *ST* (or *AT*) does not help *AT* forecasting very much any more at Den Helder. In both San Diego and Den Helder the $AT \rightarrow AT$ scheme is superior (inferior) to the $ST \rightarrow AT$ scheme for short (long) lead times.

b. Seasonality of daily forecasts

Figure 5 shows the skill score *S* as a function of calendar month of the predictor and lead time for the $ST \rightarrow AT$ scheme. The number of forecasts verified per month is about 40% of 30 × 32 (for each lead time). Skills larger than 20 are hatched, lower than 10 are stippled. The *S* at 3 days is given specifically by a number, ranging from 31 in December and February to only 11 in September. The skill scores are low in late summer for all lead times. Figure 6 is the same as Fig. 5 but now for $AT \rightarrow AT$. At a lead time of 3 days $AT \rightarrow AT$ outperforms $ST \rightarrow AT$ in all months of the year but the reverse becomes true for nearly all months at longer lead

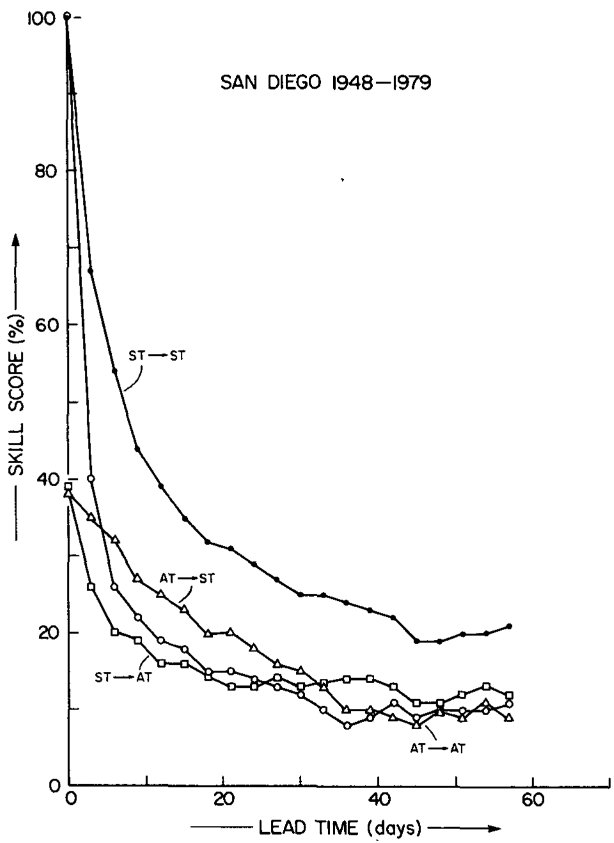

FIG. 3. Year round skill score (%) of the $ST \rightarrow ST$, $ST \rightarrow AT$, $AT \rightarrow AT$ and $AT \rightarrow ST$ schemes. All data are detrended, over 1948–79 and for the San Diego setting. At $\tau = 0$, *ST* and *AT* specify themselves (trivially) with $S = 100\%$.

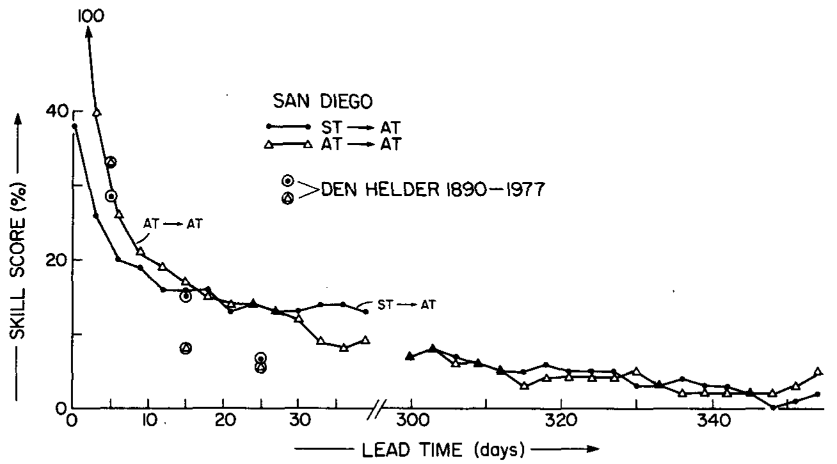

FIG. 4. Year-found skill score (%) of the $ST \rightarrow AT$ and $AT \rightarrow AT$ schemes for lead times out a year. The curves are for San Diego, 1948–79. The encircled dots and asterisks are 10-day averages of the skill score of daily forecasts made in exactly the same way for the pair light-vessel *Texel* and Den Helder over the period 1890–1977.

times. Seasonality in the skill of the $AT \rightarrow AT$ scheme is virtually absent. The clearest seasonality is visable in the $ST \rightarrow ST$ scheme (see Fig. 7). ST is highly predictable in all months of the year except when the mixed layer is shallow (in late summer). Figure 7 is essentially similar to the plot of autocorrelation of ST versus month and lag, compare to Fig. 7 in Van den Dool and Horel (1984). Combining the information of Figs. 5, 6 and 7, it would seem that the lack of skill of the $ST \rightarrow AT$ scheme in late summer is caused by the short time scale of the ocean's anomalies.

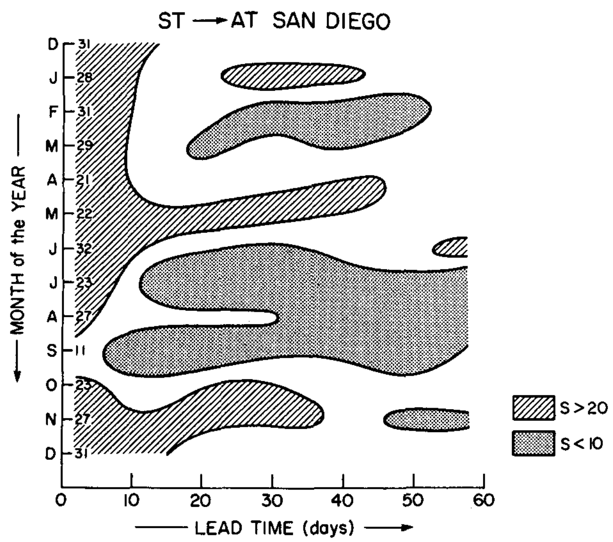

FIG. 5. The skill score (%) of the $ST \rightarrow AT$ scheme as a function of month (i.e. month of the predictor) and lead time. Skill scores larger than 20 are hatched, less than 10 stipled. The vertical row of numbers at $\tau = 3$ is the skill score of 3 day forecasts for all months.

c. Averaging the predictand

In this subsection we present year round results for the $ST \rightarrow AT$ scheme for various lengths of time averages working on AT. We start on day 0 from $ST(0)$ and ask the question whether the skill of predicting time averaged AT, notation $AT^i(\tau)$, will increase with the length of the averaging operator i. To that end we replaced daily values $AT(k, j)$ by

$$\sum_{k'=k-(i-1)/2}^{k+(i-1)/2} AT(k', j)/i$$

and repeated the procedure of calculating frequency distributions, making forecasts, etc. In Fig. 8a one can see that, indeed, at a lead time of 20 days (as an example) there is more skill in predicting a 5-day

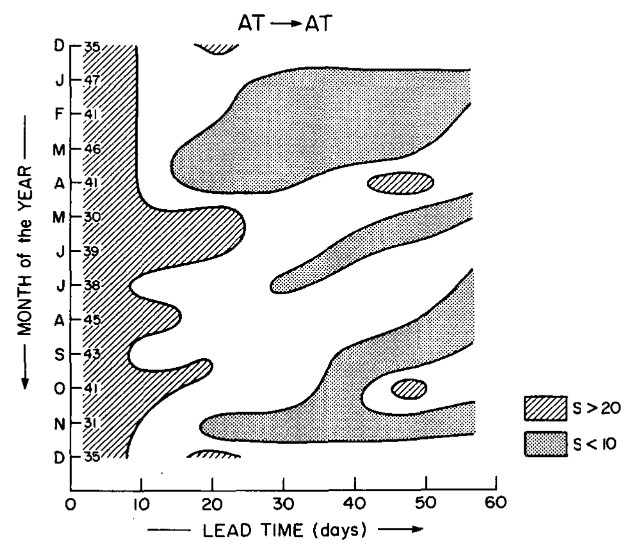

FIG. 6. As in Fig. 5 but now for $AT \rightarrow AT$.

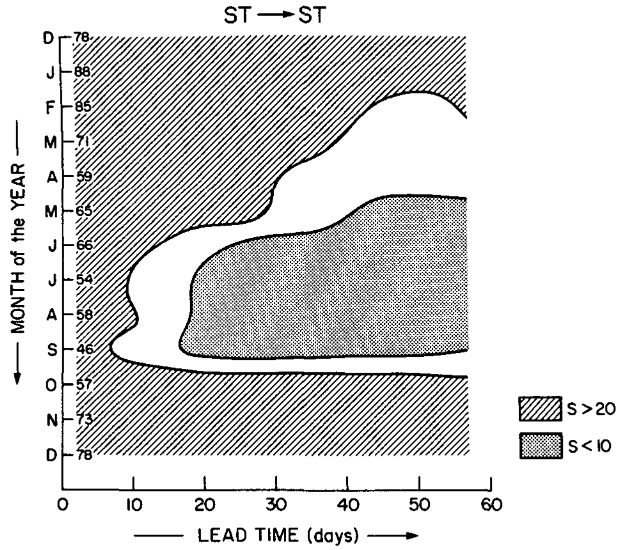

FIG. 7. As in Fig. 5 but now for $ST \to ST$.

mean AT than an instantaneous AT. And so on for 15 and 29 day mean AT. One explanation of this increase in skill could be that by averaging in time the noise in the AT series is suppressed, so as to bring out more clearly the signal associated with ST. A less favorable explanation would be that by increasing the averaging period i, the first values included in $AT^i(\tau)$ move backward in time towards the moment at which $ST(0)$ was observed. Since the skill of instantaneous forecasts is larger for smaller lead time this may explain the increase with i in Fig. 8a. Actually there is a problem in defining the notion lead time τ. At first sight, it seems natural to define the lead time τ by the distance from the day at which ST is measured to the center of the average that operates on AT. The results in Fig. 8a are plotted in that fashion (see inset of Fig. 8a). However, for small τ and/or large i the $AT^i(\tau)$ value is composed of individual AT values stretching back prior to the day at which ST is measured. Since this is clearly undesirable, those parts of the curves in Fig. 8a suffering from that problem are dashed. Another way to look upon the notion of lead time would be to define it as the distance between the day at which ST is measured and the first daily value of AT included in $AT^i(\tau)$. This defines the second notion of lead time as

$$\tau' = \tau - (i - 1)/2 \qquad (4)$$

See the inset in Fig. 8b for a clarification.

In Fig. 8b the results are plotted as a function of τ'. This changes the picture quite a bit. For small lead times (<5 days) there is more skill in forecasting an instantaneous AT value than a time-averaged AT. More specifically, yesterday's sea surface temperature contains more information about today's air temperature than about any time-averaged AT starting today. However, for longer lead times time-averaging does improve the skill. For instance, today's $ST(0)$ has more to say about AT^{29} starting 20 days from now than about an instantaneous AT 20 days ahead. This unambiguously shows the positive impact of reducing the noise in the AT series on the skill of the $ST \to AT$ forecasting scheme.

d. Averaging the predictor

In the previous section we fixed the predictor and considered various averages applied on the predictand. In the present section we will fix the predictand to be AT^{29} and ask the question whether averaging the predictor (ST) is of any help in increasing skill of predicting AT^{29}. Again we have to deal with the ambiguity in the notion lead time. In Fig. 9 there are two displays of the same results. In Fig. 9a τ is defined as the distance (in days) between the centers of gravity of the time averages over ST and AT respectively, and in Fig. 9b τ' is the distance separating the end of the average over ST and the beginning of the average over AT. For a prediction of $AT^i(\tau)$ starting from $ST^j(0)$ τ relates to τ' by

$$\tau' = \tau - (i - 1)/2 - (j - 1)/2. \qquad (5)$$

Insets in Figs. 9a and 9b clarify the meaning of τ and τ'.

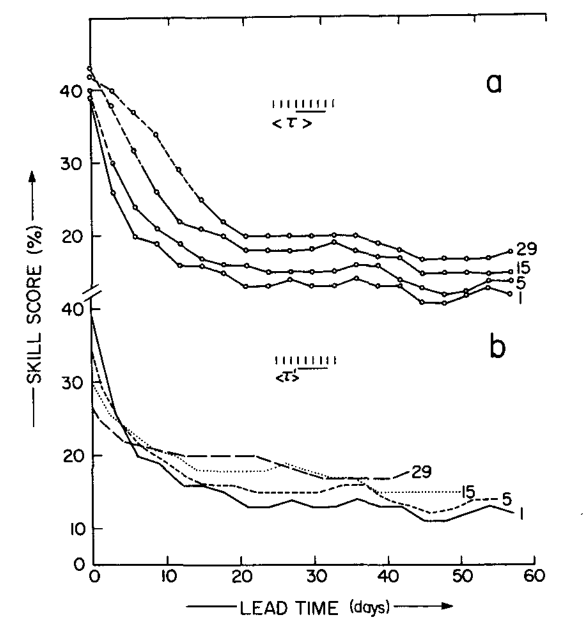

FIG. 8. Year-round skill score (%) of the $ST \to AT^i$ scheme where AT^i is averaged over i = 1, 5, 15 or 29 days. The curves are for San Diego, 1948–79. In the upper panel (a) the results are plotted versus a lead time (τ) which is the distance between $ST(0)$ and the center of the average working on AT (see inset for an example). In the lower panel the same results are plotted against the lead time τ' which is the distance between $ST(0)$ and the first daily AT value included in the average AT^i [See Eq. (4) and inset]. The dashed part of the curves in (a) suffer from definitional problems described in the text.

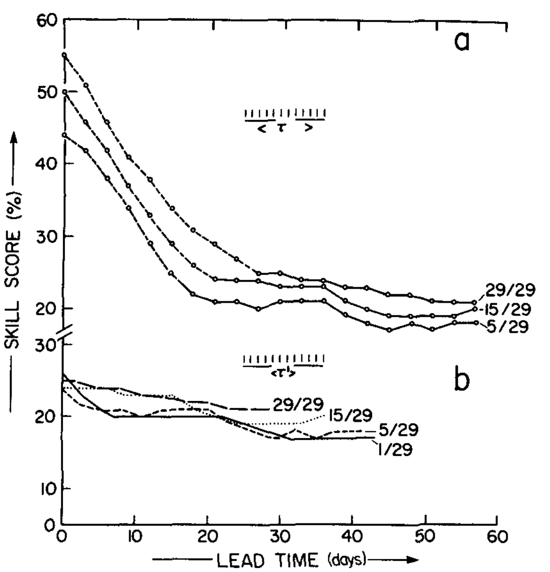

FIG. 9. Year-round skill score (%) of the $ST^j \rightarrow AT^i$ scheme where ST^j is averaged over j = 1, 5, 16 and 29 days and i = 29 days. The curves are for San Diego, 1948–79. In the upper panel the results are plotted versus a lead time (τ) which is the distance between the center of the two averages (see inset for example) while in the lower panel the lead time is defined as the distance from the last value included in ST^j to the first included in AT^i [see inset and Eq. (5)]. In (a) the $ST^1 \rightarrow AT^{29}$ curve has been omitted since it virtually coincides with 5/29. The dashed part of the curves in (a) suffer from definitional problems described in the text.

Figure 9a shows how averaging the predictor increases our skill score. Figure 9b indicates that even with a very conservative view on the notion lead time the skill score generally increases with increasing averaging of the predictor (at least away from τ' = 0). Therefore we conclude that suppressing noise in one way or another in the predictor time series is advantageous to making a forecast—at least for the San Diego case.

e. More than one predictor

We now return to forecasting daily AT from antecedent daily information. In Figs. 2, 3, 5 and 6 it was shown that although $AT \rightarrow AT$ performs somewhat better than $ST \rightarrow AT$ (at least for $\tau < 25$ days) the level of skill of these two schemes is not all that different. One may suspect actually that the high instantaneous correlation between AT and ST forces these schemes to behave more or less the same. Therefore it is of interest to find out whether a two-predictor scheme based on antecedent AT and ST [$(AT, ST) \rightarrow AT$] can add anything to the information contained already in ST or AT. In Fig. 10 we compare year round results of the $ST \rightarrow AT$ and $AT \rightarrow AT$ scheme (the same as in Fig. 3) with the $(ST, AT) \rightarrow AT$ scheme defined by

$$ST(0) \quad \text{and} \quad AT(0) \in \begin{pmatrix} MA \\ MB \end{pmatrix} \rightarrow AT(\tau) \in \begin{pmatrix} A \\ B \end{pmatrix} \qquad (6)$$

(6) is a natural extension of (2). As it turns out (6) has considerably more skill at all lead times than either of the single predictor schemes. However, this in itself does not necessarily mean that a two-predictor scheme is better. One may argue that by requiring $AT(0)$ to be extreme one in fact selects cases where $ST(0)$ is much more extreme than being a member of MA or MB. And, indeed, choosing a more extreme $ST(0)$ as a predictor would increase the skill score. However, this possibility is ruled out by considering the mean anomaly. For the $ST \rightarrow AT$ scheme the mean absolute anomaly in $ST(0)$ is 1.58°C and for the $(ST, AT) \rightarrow AT$ scheme this value is 1.68°C, not much more extreme. Likewise, for the $AT \rightarrow AT$ scheme the mean absolute anomaly in $AT(0)$ is 2.42°C and for the $(ST, AT) \rightarrow AT$ scheme this value is 2.47°C. Hence the two predictor scheme is not a super extreme one predictor scheme in disguise, and therefore we conclude that two predictors (ST and AT) contain more information about future AT than just one (ST or AT). This conclusion is independent of lead time. A likely physical explanation is that requiring AT to be extreme guarantees the ST to become (or stay) more extreme for the next few days than it would have done otherwise. This is supported by the fact that the $AT \rightarrow ST$ has a high skill at short lead times (see Fig. 3). So using AT as a predictor works out, to some extent, as a look ahead ST, or in other words as reducing the effective lead time of the $(ST, AT) \rightarrow AT$ scheme.

f. Predicting max/min temperature

Up to this point we have been concerned with predicting the daily averaged temperature. In this section we will consider the prediction of maximum and minimum temperature (T_x and T_n) separately. The predictor will be ST, so we have (1) $ST \rightarrow T_x$, (2) $ST \rightarrow T_n$ and for comparison (3) $ST \rightarrow AT$ ($=(T_x + T_n)/2$). Each of these schemes has a decay of skill with increasing lead time in the fashion shown in

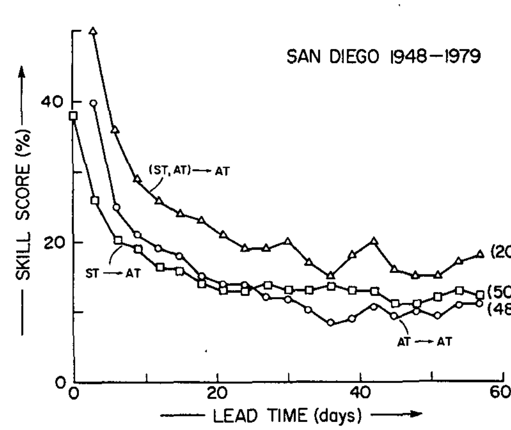

FIG. 10. Year-round skill score (%) of the $ST \rightarrow AT$, $AT \rightarrow AT$ and $(ST, AT) \rightarrow AT$ schemes. The latter is defined by Eq. (6). The numbers in parenthesis are the number of forecasts made with a given scheme.

Fig. 2. Figure 11 shows the results for the 3-day forecast [$ST(0) \rightarrow T_x$ (3) etc.] as a function of the month of the year. The seasonality in the three curves is similar, each having minimum skill in September. With the exception of December the $ST(0)$ has more to say about future T_n than about future T_x by a wide margin. Apparently minimum air temperature is controlled much more by the ocean's temperature than maximum air temperature is. This is quite evident also from Fig. 1 which indicates that AT is rarely lower than ST while the opposite is a common feature. It is only in the absence of incoming solar radiation that ST would determine T_x, but such days are rare in San Diego. It is also interesting to see that AT is about as predictable as T_n. Apparently averaging T_n with T_x suppresses noise and even though T_x is much less related to ST than T_n is, the resulting $(T_x + T_n)/2$ is as predictable as T_n.

g. Outside the coastal strip

In Van den Dool and Nap (1981) the central issue was the very rapid decrease of skill of forecasting when moving landward from the North Sea coast of the Netherlands. While employing ST as a predictor, they found the skill of forecasts of monthly mean AT to be reduced to ½ its coastal value over a distance of only 20–50 km. To test this behavior in Southern California we repeated the calculations for air temperature in Escondido, situated about 25 km away from the coast and 40 km north-northeast of San Diego. Figure 12 is as Fig. 11 but now for Escondido's air temperature predicted from ST at the Scripps Pier. The results are somewhat amazing. There is a radical decrease in skill compared to Fig. 11, but mostly in the T_x prediction which has dropped from 15 to only 4% (yearly-averaged skill). In fact the T_n predictions for Escondido are almost as skillful as those for San Diego. The AT skill has dropped from 26 to 16% which fits in very well with the results for the Netherlands where the skill score of the $ST \rightarrow AT$ scheme, S being averaged over the first 10 days,

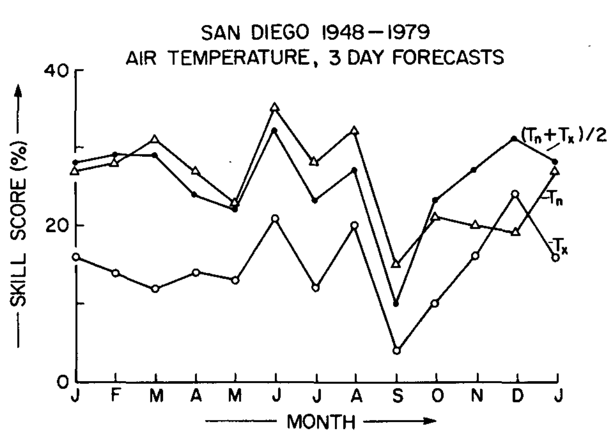

FIG. 11. Skill score (%) of 3-day forecasts of maximum (T_x), minimum (T_n) and daily mean temperature at San Diego starting from antecedent ST as a function of the month of the year.

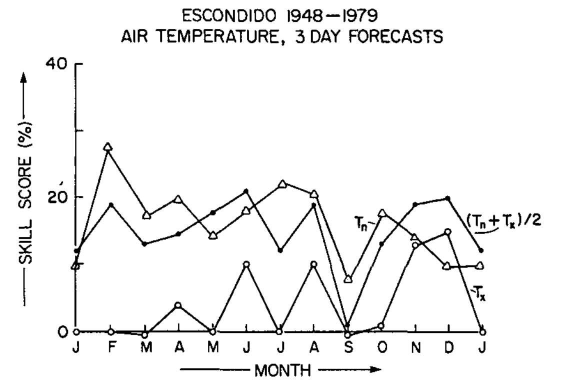

FIG. 12. As in Fig. 11 but for Escondido's air temperature.

drops from 29% at Den Helder to 18% for AT at De Bilt (about 50 km away from the coast).

At a lead time of 15 days the skill of the $ST \rightarrow T_x$ and $ST \rightarrow T_n$ schemes has leveled off to small values (0.05–0.10 at San Diego and 0.05 at Escondido). So the decrease of skill when going inland is most clearly visible for time scales less than 15 days.

4. Conclusions and discussion

The conclusions (valid for San Diego unless stated otherwise) of this study are:

1) Using simple forecast techniques and readily available local initial data a significant amount of skill is achieved in air temperature forecasts for stations near the ocean, San Diego in particular. Sampling just today's sea surface temperature (ST) or air temperature (AT) turns out to have prognostic value for the air temperature out to a month (Dutch coast) and out to a year (!) (San Diego).

2) The skill score of the $ST \rightarrow ST$ scheme is superior (at all lead times) to $ST \rightarrow AT$, $AT \rightarrow AT$ and $AT \rightarrow ST$.

3) It is easier, for lead time less than a month, to forecast $AT \rightarrow ST$ than to forecast $ST \rightarrow AT$.

4) For lead times less than a month $AT \rightarrow AT$ is generally superior to $ST \rightarrow AT$; for longer lead times the reverse is true. This phenomenon is observed throughout most of the year in San Diego and seems to hold up at the Dutch coast as well (except that the turn around point is 10 rather than 25 days).

5) The skill score of the $ST \rightarrow AT$, $ST \rightarrow ST$ and $AT \rightarrow AT$ schemes is generally high throughout the year with the exception of late summer. This is attributed to the shallowness of the mixed layer of the ocean by the end of the summer.

6) Averaging the predictand in the $ST \rightarrow AT$ scheme has a positive (negative) impact on the skill score for long (short) lead times. (A lot of caution is needed in defining lead time in schemes where either the predictor or the predictand is averaged in time.) Averaging the predictor in the $ST \rightarrow AT$ scheme seems to improve the skill of predicting monthly means for nearly all lead times. The latter result

differs from Roads and Barnett's (1984) results concerning statistical prediction of 500 mb height. They found that the most recent piece of information is the best predictor of all time averages up to a month. Such a result can be expected for processes close enough to red noise; San Diego's AT and ST do not obey the red noise process. But San Diego may very well be an exception. In the *Texel*-Den Helder case the skill score of the $ST^j \rightarrow AT^{30}$ ($\tau' = 0$) scheme decreases monotonically from 0.28 for $j = 1$ to 0.19 for $j = 30$, in agreement with Roads and Barnett (1984).

7) Using two predictors (ST and AT) seems to improve AT prediction. The reason for this could be that AT has a lot of prognostic value concerning ST.

8) The skill score rapidly decays if one tries to predict AT at Escondido (25 km inland) from antecedent ST. This is in agreement with earlier findings by Van den Dool and Nap (1981) for the Dutch area. The decay in skill is mostly in the decay of the skill in predicting the maximum temperature for lead times less than 15 days.

9) For both San Diego and Escondido, the ST has much more prognostic information about upcoming minimum temperature than maximum temperature.

The results reported here have obvious restrictions. The weather and climate of San Diego could very well be a rare case where local forecasting happens to be successful out to incredible lead times. So the relevance to other areas may be small. One may even argue that we have not studied local forecasting in its most restricted sense. In the San Diego area one single observation seems to be enough to inform us about climatic anomalies of large spatial and temporal extension. The sea surface temperature at the Scripps Pier can be anomalous for 1 or 2 years and these anomalies correlate highly (especially in winter) with sea surface temperature anomalies all over the North Pacific (Cayan, personal communication, 1984).

Conditions are completely different in the North Sea area. A single observation in that area does not seem to contain prognostic information for lead times longer than a month. As discussed in Van den Dool and Nap (1981) there is a much stronger seasonality in the skill probably owing to the seasonal dependence of the stability of the lower atmosphere. However, combination of the results from these two cities does suggest that predictability of AT from antecedent local information could play a significant role along boundaries of large water bodies. In practice ST could be incorporated into the post processing [Model Output Statistics (MOS)] of a large-scale numerical weather prediction model. Often antecedent AT has been used in MOS and in a physical sense AT may be a proxy to describe ST (or the state of the surface more in general). However it is desirable to include ST directly both for short and long range prediction.

A very old question in the context of long-range prediction is whether it is advisable or inevitable to average the predictand and/or predictor in time. This question seems to have been answered in the affirmative since it has become almost tradition to forecast monthly and seasonal means. In this paper (as well as Roads and Barnett, 1984) this question was approached in a very formal way, that is by investigating whether skill scores improve upon averaging. But there is also the practical consideration. From a practical point of view it does not make any difference (to a user) whether we forecast that there is an above normal chance that the monthly mean will be cold or that there is an above normal chance that any given day of the coming month will be cold. So the scientific finding that skill improves upon averaging the predictand could be quite meaningless to the user of the forecast. The tradition to forecast time means has another drawback: it obscures the fact that most of the skill is really in the first few days. It is better to look upon long-range prediction as information concerning the slowly varying part of the weather and to verify these predictions also against daily data (either instantaneous or filtered daily data) for various lead times rather than against one single monthly or seasonal mean alone.

Acknowledgments. The sea surface temperature data concerning the Scripps Pier were put at our disposal by Betsy Stewart from Scripps Institution for Oceanography (SIO). Most of the work was done while the first author visited the Climate Research Group at SIO. The final product benefitted a great deal from discussions with John Horel, John Roads, Daniel Cayan and Tim Barnett. This research was partially supported by the Climate Dynamics Program, Division of Atmospheric Sciences, National Science Foundation under Grants ATM-8314431 and ATM-8123279.

REFERENCES

Hubbs, C. L., 1948: Changes in the fish fauna of western North America correlated with changes in ocean temperature. *J. Mar. Res.*, **7**, 459–482.

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

Roden, G. I., 1960: On nonseasonal temperature and salinity variations along the west coast of the United States and Canada. *California Cooperative Oceanic Fisheries Investigations*, **8**, 95–119.

Smagorinsky, J., 1974: Global atmospheric modeling and the numerical simulation of climate. *Weather and Climate Modification*, W. N. Hess, Ed., Wiley and Sons, 633–686.

Tont, S. A., 1981: Upwelling: effects on air temperature and solar radiance. *Coastal and Estuarine Sci.*, **1**, 57–62.

van den Dool, H. M., 1984: Long-lived air temperature anomalies in the midlatitudes forced by the surface. *Mon. Wea. Rev.*, **112**, 555–562.

——, and J. L. Nap, 1981: An explanation of persistence of monthly mean temperatures in the Netherlands. *Tellus*, **33**, 123–131.

——, and J. Horel, 1984: An attempt to estimate the thermal resistance of the upper ocean to climatic change. *J. Atmos. Sci.*, **41**, 1601–1612.