An analogue prediction method for global sea surface temperature

N. Agarwal*, C. M. Kishtawal and P. K. Pal
Atmospheric Sciences Division, Meteorology and Oceanography Group, Space Applications Centre, Ahmedabad 380 053, India

A new analogue method for predicting the global sea surface temperature (SST) is presented in this paper. In order to predict the SST for the future week, the present method searches for a similar time evolution of SST in the historical data set. Weekly analysed SST fields of Reynold and Smith¹ have been used as the main database for the present study. The present algorithm shows reasonable skill in predicting global SST patterns for a lead-time up to 25 weeks.

SEA surface temperature (SST) is the most important time-varying boundary condition for extended-range dynamic predictions, and one requires predicted SST fields at each time step of model integration, for up to two months. One solution to this problem is to use coupled models, in which the SST prediction is done by the ocean part of the coupled model. The other alternative is to use some empirical technique for SST prediction. There are various methods of empirical prediction of SST. The simplest method may be to extrapolate the values with the assumption of persistence, but in this method the prediction error grows very rapidly with time. The same thing may happen with the persistence gradient method. Statistical methods have emerged as a useful tool for predicting the SST fields during the past one decade. A statistical prediction method based on principal oscillation pattern (POP) analysis method is used to forecast SSTs in the equatorial Pacific Ocean, known as Nino-3 (140–180 W; 5 S–5 N)²,³. Penland and Magorian⁴ used the method of inverse linear modelling for SST prediction. Their method involved a linear transformation of a predictor field to produce a forecast of a predictand field, where the predictand field is the predictor field at some later time. This approach was similar to the concept of Canonical Correlation Analysis (CCA)⁵,⁶. CCA derives a transformation that maximizes the correlation between the predictand and the transformed predictor fields. In contrast, POP analysis is based on an eigenvalue problem resulting from minimizing the error between the predictand and the transformed predictor field.

The technique of using the historical analogues to formulate a forecast, known as ‘analogue’ prediction method has been used to varying degrees of success by different researchers⁷,⁸. This method is based on the premise that interseasonal changes in the climate system occur similarly from one instance to another, such that when the system is in the same state it was for the same season in some past year, a sequence of events similar to those which occurred in the past instance may be expected now also. Such analogue methods are in vogue in seasonal climate prediction and also in cyclone track prediction. In case of SST prediction, the above argument may hold good if (a) the ocean dynamics responsible for changes in SST fields are linear or weakly nonlinear, (b) the SST data reflect most of the relevant dynamics in the evolution of SST, and (c) the driving effect of noise on the system is not so large as to preclude forecasting with some meaningful lead time⁹. Penland and Sardeshmukh¹⁰ have reasoned that SST alone contains all of the relevant dynamics to a large extent, and thus there is sufficient ground to use past SST observations for making predictions for up to 9 months. Besides conceptual simplicity, the main advantage of empirical prediction methods like analogue technique is that predictions are made using observed data alone, whereas in case of coupled models the predicted SST fields may be affected by the way different physical processes (e.g. air–sea coupling) have been parameterized in the model.

Here we present the results from a new analogue prediction method for SST prediction. In this method, the immediate past tendency of SST for one year is matched with the earlier similar cycles and the best-matched year is picked up for each grid point of the global domain. The matching is done by defining a norm for distance between the two cycles with the help of differences in absolute values, tendencies and spatial gradients at each grid point. The motivation for the present research came from the fact that the dynamic extended range prediction models need predicted SST fields as an important boundary condition, for prediction of monthly to seasonal mean monsoon rainfall. However, SST prediction in itself may have various other applications, such as the prediction of El-Nino and non El-Nino events.

Data and methodology

The basic data used for the present study is the weekly SST analysis prepared by Reynold and Smith¹. These
fields were computed weekly on a 1° grid using objective analysis of in situ and satellite measurements of SST, by means of optimum interpolation technique. The overall global mean accuracy of the final fields was estimated to be about 0.7 K. At the beginning of this study, the weekly SST analysis existed for the period January 1982 through April 2000 (~ 18 years).

At present, various operational centres use the climatological mean SST for extended range predictions. However, at different regions of the globe, the actual SST fields may deviate significantly from their climatological mean values. Figure 1 shows the mean deviation of observed weekly SST from the climatological mean values. For about 75% of the oceanic regions, the deviation is larger than 0.5 K. This discrepancy may be enough to cause considerable departures of predicted climatic states from the actual ones.

Our approach of SST prediction is based on the search of appropriate analogues of present SST time series (past 52 weeks from the current week) in the historical SST time series, at each 1° grid point. For this, a 52-week time series of SST is searched from the past data set that has the maximum likelihood with the current SST time series of 52 weeks. The total length of the past data set is \( N \) of 52 week length, where, \( N \) is the total number of weeks in the data set, and also, \( N \) is the index of the current week. Maximum likelihood criterion is matched if the following matching function \( F \) is minimized:

\[
F^P = \sum_{i=1}^{52} \alpha |\text{SST}^C_i - \text{SST}^P_i| + \beta \left| \frac{\partial \text{SST}^C_i}{\partial t} - \frac{\partial \text{SST}^P_i}{\partial t} \right|
\]

The superscripts \( C \) and \( P \) represent current and past cycles, respectively, while \( i \) is the index for week, which varies from 1 to 52 for a cycle. The year is counted backwards from the current year and refers to the 52-week period. This is different from the normal calendar year. \( \alpha \) and \( \beta \) are arbitrary scaling parameters. If the above function minimizes at cycle \( J = P \), then the predicted SST at a given grid point is

\[
\text{SST}^{C+1}_i = \text{SST}^{J+1}_i + \Delta \text{SST}.
\]

Figure 1. Global distribution of standard deviation (\( \sigma \)) of SST. Areas with \( \sigma > 0.5 \) K are shaded.
After the prediction is made, the predicted value replaces the current value, and the procedure given by eqs (1) and (2) is repeated iteratively. Some initial treatment of the analysed SST fields would be necessary to compute the matching function $F$ in eq. (1). This includes the removal of annual cycle and then filtering of time series of SST anomalies using 3-point averaging of residual SST fields. The latter step was found to be helpful in minimizing the effects of high frequency nonlinear dynamics in SST time series. In eq. (2), $\Delta$SST indicates the mean difference between present and past best match SST time series. A schematic representation of the method is depicted in Fig. 2.

**Results and discussion**

In order to access the skill of the present prediction algorithm, a comparison of predicted and actual SSTs has been carried out for six zones (Table 1).

The location of these regions is shown in Fig. 3. Out of these, the regions situated in equatorial Pacific are known to exhibit large inter-annual variations of SST, in the form of El-Nino and La-Nina. The global impact of El-Nino–La-Nina cycle on climate and its influence on

<table>
<thead>
<tr>
<th>Region</th>
<th>Lat.-Long. range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nino 3</td>
<td>10 N–10 S and 130 W–80 W</td>
</tr>
<tr>
<td>Nino 4</td>
<td>10 N–10 S and 180 W–130 W</td>
</tr>
<tr>
<td>Indian Ocean (Region-A)</td>
<td>15 S–25 N and 55 E–95 E</td>
</tr>
<tr>
<td>North West Pacific (Region-B)</td>
<td>30 N–60 N and 180 W–120 W</td>
</tr>
<tr>
<td>North East Pacific (Region-C)</td>
<td>30 N–60 N and 120 E–180 E</td>
</tr>
<tr>
<td>North Atlantic (Region-D)</td>
<td>30 N–60 N and 70 W–00 W</td>
</tr>
</tbody>
</table>

**Figure 4.** Comparison of rms errors of prediction (solid line) with rms errors of climatological guess (dashed line) for 30 weeks of prediction, starting from the first week of May. Each curve is an average of six different predictions between 1994 and 1999. Different panels show the comparison for different geographic regions.
Figure 5 a–d. Comparison of predicted (solid line) and actual (dashed line) SST anomalies for 30 weeks of prediction, starting from the first week of May 1996, for a, Region-A, b, Region-B, c, Region-C, and d, Region-D.

Figure 5 e–h. e, Comparison of actual (dashed line) and predicted SST anomalies over Nino-3 region for 30 weeks of prediction, starting from the first week of May 1996. f, Same as (e), but for Nino-4 region; Difference between actual and predicted SST anomalies for g, Nino-3 region; and h, Nino-4 region.
human life has motivated many researchers to find methods to make accurate prediction of SST over this region.

Figure 4 shows weekly values of the rms errors of predicted SST anomalies (solid line) along with the rms errors of weekly climatological values (dashed line), averaged for the period 1994–1999, for different regions mentioned earlier. For all the curves, the starting week for prediction is the first week of May. These curves indicate that the predicted fields provide improved estimates of SST fields for at least the first 10 weeks (2 months) of time. The error reduction is particularly noticeable over the tropical Pacific Ocean, which comprises Nino-3 and Nino-4 regions, where the present method improves the prediction by as much as 20% over the climatological guess.

Figure 5 a–d shows the time series of predicted and actual SST anomalies for 30 weeks, starting from the first week of May 1996, for the regions of study. The time variation of predicted and actual SST anomalies shows an overall matching, and the maximum deviation of predicted SST anomalies from actual values is within 0.5 K. Over Nino-3 and Nino-4 regions, there is a very good match between predicted and observed SST anomalies (Figure 5 e–h). Deviations are somewhat larger over regions B and C (both over Pacific Ocean) during 1997 (Figure 6 a–d), which was a very strong El-Nino year. Relatively poor performance of the method for 1997 can be attributed to the smaller size of the data set. Since this method relies on the search of an appropriate analogue of current SST variations in the past data records, SST variations similar to 1997 were absent from the data set, which starts from the first week of January 1982. Even though the search method minimizes at the most similar event during the past time, the predicted values may deviate significantly from the actual ones, during the extreme events as in case of 1997. In other words, time evolution pattern similar to 1997, never existed in past records of our data set and it was difficult to predict such unusual events which had no analogues in the past. This also explains relatively larger differences between the actual and predicted SST anomalies over the Nino-3 region (Figure 6 e–h). However, over Nino-4 region, where the magnitude of SST anomalies was mild, our method predicted reasonably better. Despite larger error over Nino-3 region, the present method was able to predict some positive anomaly over the Nino-3 region, which was still better than the climatological mean field, which would have predicted a zero anomaly. In a modified scheme, we used a variable size of analogue cycle in such a way, that for initial weeks of prediction, the analogue cycle size is small (~3 weeks) and gradually increases to 52 weeks, after the twentieth week of prediction. This scheme slightly improves the prediction of SST anomalies at...

Figure 6 a–d. Same as Figure 5 a–d but for 1997. Additional dotted curve shows the prediction, with variable analogue cycle size.

CURRENT SCIENCE, VOL. 80, NO. 1, 10 JANUARY 2001 53
Nino-3 and Nino-4 regions (Figure 6, dotted curves). This observation suggests that further work is needed on the selection of the size of analogues and the impact of the recent past observations on the predictability of SST patterns. Figure 7 shows a comparison of predicted and actual SST anomaly patterns for December 1997. During a non El-Nino year, the predicted SST anomalies compare quite well with the actual values, which is evident from Figure 4. Figure 8a and b shows the global mean RMS errors of climatological mean and predicted anomalies using the present method, for the first fifteen weeks of prediction for the years 1997 and 1998. Figure 8 clearly indicates that at least for the first 3 months, the time over which the dynamic models need predicted SST fields for making extended range prediction, is superior to the climatological guess fields.

The analogue prediction technique described above has been compared with the existing NCEP coupled model. The results for the NCEP coupled model were obtained from Climate Diagnostic Bulletin (January 2000). In this model the forecasts were made from four consecutive weekly oceanic initial conditions (2, 9, 16 and 23 January 1997).

Figure 6 e–h. Same as Figure 5 e–h but for 1997. Additional dotted curve shows the prediction, with variable analogue cycle size.

Figure 7. Comparison of predicted (upper panel) and actual (lower panel) SST anomalies for December 1997. This is the 30th week prediction starting from May 1997.
For each oceanic initial condition, four forecasts were made by using different atmospheric initial conditions. Figure 9 shows the difference between the ensemble mean of all the sixteen forecasts by the NCEP coupled model and the observed SST anomaly (dashed line). It also shows the difference between the SST anomaly as predicted from the analogue method starting from the end of January and the observed SST anomaly (solid line). It can been seen that for most of the points the analogue method gives significantly more accurate estimates of SST anomalies than the ensemble mean prediction by NCEP coupled model.

**Summary**

A new type of analogue prediction technique is proposed for forecasting global SST patterns on weekly time scales. Using SST analysis of Reynolds and Smith as the main database, the present algorithm shows some skill in predicting the El-Nino and non-El-Nino events, with a lead period of 30-40 weeks. One of the limitations of our method is the small size of historical data from where the appropriate analogues are picked up. It can be expected that the accuracy of the present method will improve as the data size increases with time.


Received 7 July 2000; revised accepted 9 October 2000