



UNIVERSITY
OF WOLLONGONG
AUSTRALIA

University of Wollongong
Research Online

Faculty of Engineering and Information Sciences -
Papers: Part A

Faculty of Engineering and Information Sciences

2002

MTD method for better prediction of sea surface temperature

Velappa Ganapathy
University of Malaya

K Meena
Shrimati Indira Gandhi College

Kashem M. Muttaqi
Queensland University of Technology, kashem@uow.edu.au

Publication Details

V. Ganapathy, K. Meena & M. A. Kashem, "MTD method for better prediction of sea surface temperature," *International Journal of Remote Sensing*, vol. 18, (23) pp. 3725-3743, 2002.

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library:
research-pubs@uow.edu.au

MTD method for better prediction of sea surface temperature

Abstract

A class of incremental learning procedures known as the Modified Temporal Difference (MTD) method is introduced in this paper for fixed-step prediction problems which uses the functional features of Multilayer Perceptron. The method is applied for weekly prediction of the Sea Surface Temperature (SST) from oceanographic data. Temporal Difference (TD) methods suggest how each output of a temporal sequence must be changed, whereas a back-propagation algorithm decides which part(s) of a network to change in order to influence its output and reduce the overall error. In other words, TD methods and back-propagation address temporal credit and structural credit assignment issues, respectively. While the two methods address different sides of the same issues they are quite compatible and easily combined. A new scheme is formed by combining the advantage of back-propagation and TD methods catering to fixed-step problems and is named as the MTD method. The back-propagation algorithm is modified to propagate the temporal error. For prediction problems, the exponential recency has not been found suitable due to its large negative slope. In this paper a weighing scheme is introduced in which alterations are made to past predictions according to a newly proposed recency factor. The stochastic method, back-propagation algorithm, TD and MTD methods are applied to predict the SST values in the Arabian Sea, the Bay of Bengal and Central Indian Ocean and a comparative study is made. From the study it is observed that the proposed alternative recency factor in the MTD method leads to better prediction than the exponential recency.

Keywords

mtd, method, better, surface, prediction, temperature, sea

Disciplines

Engineering | Science and Technology Studies

Publication Details

V. Ganapathy, K. Meena & M. A. Kashem, "MTD method for better prediction of sea surface temperature," *International Journal of Remote Sensing*, vol. 18, (23) pp. 3725-3743, 2002.

“This article was originally published as”

MTD method for better prediction of sea surface temperature

Modified Temporal Difference Method for Better Prediction of Sea Surface Temperature

V. Ganapathy*, K. Meena** M. A. Kashem***,
* Multimedia University, Cyberjaya, Selangor, Malaysia
** Shrimati Indira Gandhi College, Tiruchirapalli, India
*** Electrical & Electronic Systems Engineering, Queensland University of Technology,
Brisbane, Australia

Abstract

A class of incremental learning procedures known as Modified Temporal Difference (MTD) method is introduced in this paper for fixed step prediction problems which uses the functional features of Multilayer Perceptron. The method is applied for the weekly prediction of the Sea Surface Temperature (SST) from the oceanographic data. Temporal Difference (TD) methods suggest how each output of a temporal sequence must be changed, whereas back-propagation algorithm decides which part(s) of a network to change in order to influence its output and reduce the overall error. In other words, TD methods and back-propagation address temporal credit and structural credit assignment issues respectively. While the two methods address different sides of the same issue, they are quite compatible and easily combined. A new scheme is formed by combining the advantages of back-propagation and TD methods catering to fixed step problems and is named as the Modified Temporal Difference (MTD) method. The back-propagation algorithm is modified to propagate the temporal error. For prediction problems, the exponential recency has not been found suitable due to its large negative slope. In this paper a weighing scheme is introduced in which alterations are made to past predictions according to a newly proposed recency factor. The stochastic method, back-propagation algorithm, TD and MTD methods are applied to predict the Sea Surface Temperature

(SST) values in the Arabian Sea, the Bay of Bengal and the Central Indian Ocean and a comparative study is made. From the study it is observed that the proposed alternative recency factor in the MTD method leads to better prediction than the exponential recency.

Keywords: Sea Surface Temperature, Temporal Difference Method, Back-Propagation, Multilayer Perceptron, and Recency Factor.

I. Introduction

Conventional prediction has been approached by the techniques of stochastic modelling and design. Research in this area has forked into two broad categories of analysis - the Kendall [1] and the Wiener [2] approaches. Cox [3] has reviewed processes with long range dependence with emphasis on their connections with second order time series analysis. These models are dependent on the time span in which they are operating and concentrate mainly on structure and development rather than on predictions. Though the Wiener model gives comparatively good results, it has not been widely applied owing to its difficult mathematics.

Temporal Difference (TD) learning is a technique developed by Sutton [4] and provides a class of incremental learning procedures specialized for prediction problems. It is proven that for most real world predictions, TD methods require less memory and less computation than conventional methods and they produce more accurate predictions. The hallmark of this technique is its sensitivity to changes in successive prediction rather than

to overall error between prediction and the actual outcome. Any prediction problem can be cast in the supervised learning paradigm by taking the first item to be the data based on which the prediction must be made and the second item to be the actual outcome, namely what the prediction should have been. This type of predictions cannot be directly employed for making prediction for a fixed time later. Although this involves a sequence of predictions, TD methods cannot be used because each prediction is of a different event and there is no desired relationship between them. Hence, they need to use functional features of Multilayer Perceptron and as a result, extra layers, nodes and non-linearities of the latter have to be accommodated. Though Multilayer Perceptron has the capability to perform mapping of any nonlinear input-output relation and satisfy some of the criteria like adaptability and fault tolerant characteristics of its architecture, it by itself, might not yield solutions for the problem of prediction.

The exponential decay function that determines the weightage to successive predictions in TD methods induces a forgetting of learning laws and does not lead to optimal predictions. Further, they decide which part of the network to change so as to influence the network's output to reduce the overall error. They are referred to as the temporal and structural credit assignment issues respectively. Since back-propagation and TD methods address different sides of the same issue, they are perfectly compatible and easily combined. For example, Anderson [5] has implemented such a combination of back-propagation and TD methods, successfully applying it to both broom-stick balancing task and the "Towers of Hanoi" problem. However, both these domains provide only variable step problems. Narendran et. al. [6] have designed a scheme wherein the advantages of

back-propagation and TD methods have been combined in a single model for monthly rainfall prediction. They have assumed the value to be predicted as a real valued scalar and each prediction was assumed to be a function of only its preceding observation but these suppositions could be done away with. In [10] Narendran et. al. have obtained new levels of accuracy in daily and weekly rainfall predictions. They used a combination of back-propagation and TD methods to predict the quantum of yearly rainfall. Separate mechanisms for both structural and temporal credit assignment issues have been adopted. Sutton [8] has shown clearly that learning to predict is very important since it determines the outcome of the future event based on the current observations of the state of the environment. For making prediction for a fixed time later, such as the weekly prediction of Sea Surface Temperature (SST) from oceanographic data, both the temporal credit assignment and structural credit assignment issues have to be addressed. That is, it should be decided how each output of a temporal sequence of outputs needs to be changed and also which part of a network should be changed so as to influence the output in order to reduce the overall error.

In this paper, a particular TD procedure is introduced by relating it to a classical supervised learning procedure, namely, the Widrow-Hoff rule [9]. However, TD procedures cannot be used in making prediction for a fixed amount of time later, since each prediction is of a different event and there is no clear desired relationship between them. Back-propagation cannot be directly applied to the problem of prediction, since it needs to know the outcome for the back-propagation of error. But it is not available till a certain amount of time elapses resulting in large storage requirements. A new scheme is

designed in this paper wherein the advantage of back-propagation and TD methods have been combined in a single model which caters to fixed step problems. Conventionally the weightage to successive predictions in TD methods is achieved by an exponential decay function [6]. However for a fixed step prediction, such as, for example, prediction of SST, the exponential recency is not appropriate, since it has a large negative slope and hence an alternative recency which leads to better prediction is proposed. One of the most important parameters in the climatic changes and oceanographic studies is the SST [10]. The stochastic, back-propagation, TD and MTD methods are used to predict SST values in the Arabian Sea, the Bay of Bengal and the Central Indian Ocean. It is observed from the experimental results that the statistical methods suffer from their lack of adaptability and unsuitability while back-propagation gets entangled into local minima problems. TD methods are inappropriate for prediction for fixed time later because of large negative slope in the weighing scheme, and hence MTD methods are shown to yield better results for the weekly prediction of SST.

II. Temporal Difference (TD) and Modified Temporal Difference (MTD) Methods

There are two kinds of prediction learning problems namely the single-step and multi-step. In single-step problems, all information about the correctness of each prediction is revealed at once. In multi-step problems, only partial information relevant to its correctness is revealed at each step. In single-step problems, data naturally comes in as "observation-outcome pairs". These problems are ideally suited to the pair-wise supervised learning approach. In the above case, TD methods cannot be distinguished

from supervised learning methods. However, TD methods show much improvement over conventional methods in multi-step problems. It is shown that the two procedures produce exactly the same weight changes but that the TD procedure can be implemented incrementally and therefore requires far less computational power. This is more widely used as a conceptual bridge to a larger family of TD procedures, which produce different weight changes than any supervised learning method.

II.A Learning Procedures for Temporal Difference Methods

In multi-step problems, the experience comes in as observation-outcome sequences of the form $X_1, X_2, X_3, \dots, X_m, z$ where each X is a vector of observations available at time t in the sequence, and z is the outcome of the sequence. The components of each X are assumed to be real-valued measurements (or) features and z is assumed to be a real-valued scalar. For each observation-outcome sequence, the TD method produces a corresponding sequence of predictions $p_1, p_2, p_3, \dots, p_m$ each of which is an estimate of z . Here each p_t is assumed to be a function of all preceding observation vectors up through time t .

In general, all learning procedures will be expressed as rules for updating W . Let us assume that W is updated only once for each complete observation-outcome sequence and thus does not change during a sequence. For each observation, an increment to W , denoted by ΔW_t is determined. After a complete sequence has been processed, W is changed by the sum of all increments as,

$$W = W + \sum_{t=1}^m \Delta W_t \tag{1}$$

The Prototypical supervised learning procedure is

$$\Delta W_t = \langle (z - p_t) \rangle_{\Delta W_t} p_t \quad (2)$$

where \langle is a positive parameter affecting the rate of learning, and the gradient ΔW_t is the vector of partial derivatives of p_t with respect to each component of W .

Another instance of the prototypical supervised learning procedure is the "generalized delta rule" or back-propagation algorithm [11]. It is noted that all ΔW_t in equation (2) depends on z and thus cannot be determined until the end of the sequence when z becomes known. Thus all observations and predictions made during a sequence must be remembered until all the W_t 's are computed. The key is to present the error $z - p_t$ as a sum of changes in predictions; that is,

$$z - p_t = \sum_{k=t}^m \langle (p_{k+1} - p_k) \rangle \quad (3)$$

Substituting the value of ΔW_t from Equation (2) in (1) and simplifying, we get

$$W = W + \Delta W_t \quad (4)$$

$$\text{where, } \Delta W_t = \langle (p_{t+1} - p_t) \rangle \sum_{k=1}^t \Delta W_k \quad (5)$$

This equation can be computed incrementally, since each prediction depends only on a pair of successive predictions and the sum of all past values for ΔW_t . The procedure given by equation (5) is referred to as the TD(1) procedure [4]. It is also shown that TD(1) produces the same weight changes per sequence, as the Widrow-Hoff rule on multi-step problems.

Instead of sensitivity to overall error between predictions and the final outcome, sensitivity to changes in successive predictions is the speciality of TD methods. As stated in the previous case, in response to an increase (decrease) in prediction from p_t to p_{t+1} , an increment $\Delta W(t)$ is calculated that increases (decreases) the predictions for some (or all) of the preceding observation vectors X_1, X_2, \dots, X_t . The procedure given by equation (5) is a special case in which all those predictions are altered to an equal extent. Conventionally, the weightage to successive predictions in TD methods is achieved by an exponential function. This induces a forgetting of learning laws and does not lead to optimal predictions. An exponential weighing scheme with recency has been used in which alterations to past predictions are made in k steps and are weighed according to β^k for $0 \leq \beta \leq 1$. That is,

$$\Delta W(t) = \langle [p_{t+1} - p_t] \sum_{k=1}^t \beta^{t-k} \nabla_w p_k \rangle \quad (6)$$

The gradient $\nabla_w p_k$ is the vector of partial derivatives of p_k with respect to each component of W . Equation (6) represents the TD(β) family of learning procedures. If we now make $\beta = 1$, then equation (6) becomes

$$\Delta W(t) = \langle [p_{t+1} - p_t] \sum_{k=1}^t \nabla_w p_k \rangle \quad (7)$$

An important advantage of equation (7) compared to the exponential form is that it can be computed incrementally.

For $\beta < 1$, TD(β) produces weight changes in the same way as the supervised learning method. The difference is the greatest in the case of TD(0) (where $\beta = 0$), in which the

weight increment is determined only by its effect on the prediction associated with the most recent observation:

$$\Delta W_t = \langle (p_{t+1} - p_t) \rangle_w p_t \quad (8)$$

This is similar to the well known supervised learning procedure,

$$\Delta W_t = \langle (z - p_t) \rangle_w p_t \quad (9)$$

which is a gradient descendant procedure.

III. Shortcomings of TD Methods and BP Algorithm for Fixed Step Prediction

Though TD methods have substantial improvement over conventional learning paradigms, they still fall short of the requirement necessary for real world prediction problems. They are restricted to variable step problems. They cannot be used for making direct prediction for a later time because each prediction relates to a different event with no desired relationship between them. The above formulation of TD methods is applicable to linear or simple nonlinearities with only one processing element. Obviously, for real - world prediction, they need to use the functional features of Multilayer Perceptron and hence the extra layers, nodes and nonlinearities of the latter need to be accommodated. Thus TD methods have to be slightly modified if they are to be utilized for real world prediction problems. Also backpropagation cannot be used for direct prediction because it is essential to know the outcome for the backpropagation of error. But it is not available till a certain amount of time elapses resulting in large storage requirements.

IV. Modified Temporal Difference (MTD) Methods

The object of combining backpropagation and TD methods is to get accurate credit assignment. By backpropagation we learn which part(s) of a network to change in order to influence its output and reduce the overall error. TD methods suggest how each output of a temporal sequence should be changed. In other words, backpropagation and TD methods address structural credit assignment and temporal credit assignment [12] issues respectively. While the two methods address different sides of the same issue, it is significant to note that they are quite compatible and easily combined. The key requirement is that the gradient $\partial_w p_k$ should be compatible.

IV.A Prediction by a Fixed Interval

For making prediction for a fixed time later, three key steps have to be followed. This is demonstrated for weekly prediction of SST in the three oceans mentioned earlier.

- i) The problem is embedded in a larger class of problems to produce an approximate sequence of predictions.

In each month t , not only $p_4(t)$ is formed, but also $p_3(t)$, $p_2(t)$ and $p_1(t)$ are formed, where each $p_{TM}(t)$ is an estimate of SST TM weeks later.

- ii) Recursive equations that express the desired relationship between predictions at different times in the sequence are written down.

The above formulation provides overlapping sequences of interrelated predictions, for example, $p_4(t)$, $p_3(t+1)$, $p_2(t+2)$ and $p_1(t+3)$ all of the same event in this case, of the SST in week $(t+4)$.

If the predictions are accurate, then

$$p_{TM}(t) = p_{TM}(t+1), \quad 1 < TM < 5 \quad (10)$$

where, $p_0(t)$ is defined as the actual outcome at time t .

iii) An update rule is constructed that uses a mismatch in the recursive equations to derive weight changes towards a better match. The update rule for weight vector

$\square W_{TM}$ used in [4] to compute $p_{TM}(t)$ is rewritten as,

$$\square W_{TM} = \langle [p_{TM}(t+1) - p_{TM}(t)] \square_{k=1}^t p_{TM}(k) \rangle \quad (11)$$

IV.B Backpropagating Temporal Errors - Vector Representation

In a Multilayer Perceptron, $\square_w p_{TM}(k)$ is calculated by a backpropagation algorithm as described below.

The backpropagation algorithm has been modified so that it backpropagates the temporal error [13] rather than the conventional error. Let us assume that the predictions take vector values, i.e., $p_{TM}(k) = U_{1,i}$ where $U_{1,i}$ is the output of the i th node in layer 1 and values of TM can be varied. The value of $U_{1,i}$ is given by

$$U_{1,i} = f \left\{ \square_{i=0}^{N_1-1} W_{1,i,j} U_{1,j} \right\} \quad (12)$$

where $f(\cdot)$ is the sigmoid nonlinearity. As a result, the partial derivatives of the error measure with respect to each connection weight are computed in a backward propagation sweep through the network.

Hence, for the weights in the final layer

$$\frac{\partial p_{\text{TM}}(k)}{\partial W_{l,i,j}} = \frac{\partial U_{l,i}}{\partial W_{l,i,j}} = U_{l,i} (1 - U_{l,i}) U_{l\Box,j} \quad (13)$$

because of sigmoid nonlinearity. For weights in the preceding layer, we have

$$\frac{\partial p_{\text{TM}}(k)}{\partial W_{l\Box,i,j}} = \frac{\partial p_{\text{TM}}(k)}{\partial U_{l\Box,i}} \frac{\partial U_{l\Box,i}}{\partial W_{l\Box,i,j}} \quad (14)$$

Then the partial derivative of $p_{\text{TM}}(k)$ with respect to each weight in the network is obtained

by using the chain rule as follows:

$$\frac{\partial p_{\text{TM}}(k)}{\partial W_{l\Box,i,j}} = W_{l,i,j} U_{l\Box,i} (1 - U_{l\Box,i}) U_{l\Box2,i} \quad (15)$$

Similarly for the earlier layer,

$$\frac{\partial p_{\text{TM}}(k)}{\partial W_{l\Box2,i,j}} = \prod_{k=0}^{N_{l\Box1}} \frac{\partial p_{\text{TM}}(k)}{\partial U_{l\Box1,k}} \frac{\partial U_{l\Box1,k}}{\partial W_{l\Box2,i}} \frac{\partial U_{l\Box2,i}}{\partial W_{l\Box2,i,j}} \quad (16)$$

This process is continued and the partial derivatives in the update rule are obtained.

V. Studies on Recency Factors

In conventional procedure, the weightage to successive predictions in TD methods is achieved by an exponential decay function. But exponential recency is not suited for prediction for a fixed time later, since it has a large negative slope. So an alternative

recency factor which leads to better predictions and still retains some functional advantages of the original function is proposed.

A weighing scheme is a scheme in which alterations to past predictions are made according to λ^k . In this paper a new weighing scheme $1/(1 + \lambda k)$ is used in the place of λ^k . The use of $1/(1 + \lambda k)$ instead of λ^k has several advantages.

- i) The value of the function is by definition $\lambda < 1$. Hence the range of possible λ values is not restricted to the range (0 - 1).
- ii) It provides lesser slope than the conventional λ^k . This proves to be advantageous for the real world problems, because the weightage of successive predictions does not decrease in an exponential manner.
- iii) With the introduction of an intermediate function, the new weighing scheme can be expressed as follows.

$$\square W(t) = \langle [p_{t+1} \square p_t] \square_{k=1}^t \frac{1}{1 + \lambda(t \square k)} * \square_w p_k \rangle \quad (17)$$

which is obtained from Equation (6).

The λ in λ^k and that in $1/(1 + \lambda k)$ have different meanings. For example, the λ in λ^k represents the value of the function at $k = 1$ whereas the λ in $1/(1 + \lambda k)$ represents the value of the function at $k = (1 - \lambda)/\lambda^2$. Thus these two instances of λ should be interpreted differently. The advantage of this is that both TD and MTD represent a decaying function whose values at $k = 0$ are 1, 1 and λ , $1/(\lambda+1)$ at $k = 1$ respectively. This helps for a direct comparison of these two methods. Therefore, a common parameter can be defined as

$\beta = \gamma$ in the TD methods and, $\beta = 1/(1 + \gamma)$ in the MTD methods.

Figures 1 and 2 illustrate the different weighing schemes for γ^k and $1/(1 + \gamma^k)$ respectively. The k values and β values are taken along the x and y axes respectively and the graph is plotted for specific values of γ namely 0.9, 0.7, 0.6, 0.3 and 0.1. It is observed from the above Figures that MTD method provides lesser slope than the TD methods for the corresponding values of k . The differences between γ^k and $1/(1 + \gamma^k)$ are more significant for low values of β , and hence MTD methods are better for prediction of SST in such domains. This method is proved to be a natural choice for the above problem since a gradual reduction in weightage is desired rather than exponential decay.

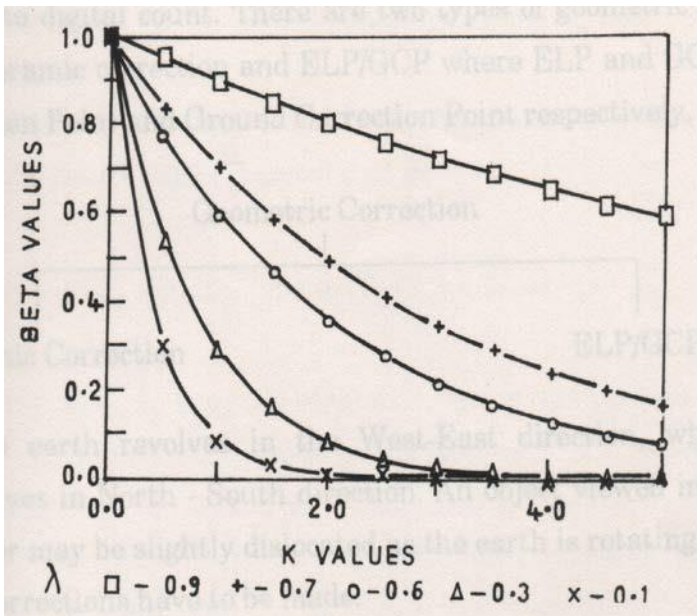


Fig.1: Weighing scheme for Conventional TD Learning

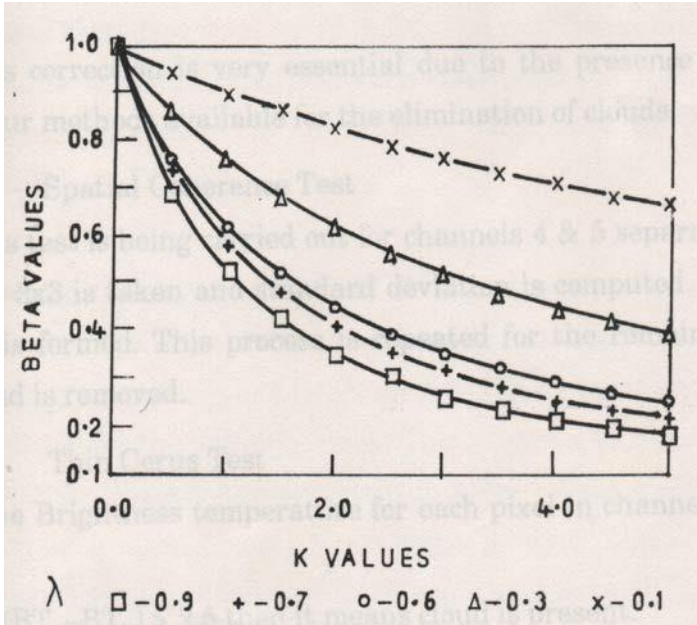


Fig.2: Weighing scheme for Modified TD Learning

VI. Experimental Results

The proposed MTD method is applied to predict the SST of the Arabian Sea, the Bay of Bengal and the Central Indian Ocean. Also, Stochastic, TD and BP methods are used to predict the same. The results are compared with the actual temperatures.

VI.A Calculation of SST

The Sea Surface Temperature (SST) is one of the most important parameters in the climatic changes and oceanographic studies. This section presents a brief description of the retrieval of SST from NOAA-AVHRR Data. The Advanced Very High Resolution Radiometer (AVHRR) is a scanning radiometer flown onboard NOAA series of satellites and operates primarily at five spectral regions (channels 4 and 5 are of same band).

Processing of AVHRR Data for the retrieval of SST involves several steps namely Radiometric calibration, SST conversion, geometric correction, etc. [10]. Oceanographic data [14] contains five channels 1, 2, 3, 4 & 5 and only channels 4 and 5 are used to find SST [15]. The various steps involved in the preparation of SST are (i) Calibration Correction, (ii) Geometric Correction, and (iii) Atmospheric Correction.

In Calibration correction, the SST data measured in radians is converted into digital count. There are two types of geometric correction namely panoramic correction and ELP/GCP where ELP and GCP are the Earth Location Point and Ground Correction Point respectively. The earth revolves in the West-East direction, whereas the sensor revolves in North-South direction. An object viewed in the earth by the sensor may be slightly dislocated as the earth is rotating and hence necessary corrections have to be made.

The Atmospheric Correction is very essential due to the presence of clouds. There are four methods available for the elimination of clouds.

(i) Spatial Coherence Test

This test is being carried out for channels 4 & 5 separately. First pixel of size 3 x 3 is taken and standard deviation is computed. If it is 0.2 then cloud is formed. This process is repeated for the remaining pixels and the cloud is removed.

(ii) Thin Cerus Test

The Brightness temperature for each pixel in channels 4 & 5 is computed.

If $(BT_4 - BT_5) > 2.5$ then it means cloud is present.

(iii) Cross Cloud Check

Here only channel 5 data is considered and if the brightness temperature $BT_5 < 280$ then it indicates the formation of cloud.

(iv) Reflectance Test

For finding the reflectance, the channel 2 data is taken. A parameter called 'Absorb' is defined as,

Absorb = (Energy reflected from the surface)/(Energy incidence on the surface).

If 'Absorb' of channel2 $> 12\%$ then it indicates the presence of cloud. The Sea Surface Temperature (SST) is calculated as in [13].

$$SST = a_1T_4 + a_2(T_4 - T_5) + a_3(T_4 + T_5) + (\text{Sec} \setminus - 1) + a_4$$

where a_1, a_2, a_3 & a_4 are constants.

The errors present in the data are deleted and the refined data is in the form of $1^0 \times 1^0$ weekly average data. Normally land is marked as 1 and cloud is marked as 0. Generally SST values vary from 20^0C to 35^0C . The processing of NOAA - AVHRR Data for the retrieval of SSTs has been undertaken at the National Remote Sensing Agency (NRSA), Hyderabad, India since 1986, and was made operational in early 1981 with a support of Department of Ocean Development (DOD), Government of India.

The backpropagation, TD and MTD methods are used to predict the SST values in the three oceans under study. The $1^0 \times 1^0$ weekly average SST data for a given latitude and longitude is fed as an input and the output value is predicted for the following week. This is shown in Fig.3. Four Physical variables namely surface stress in kg/m^2 , daily wind

speed in km per hour average for a week, density of sea water in kg/m^3 and SST in $^{\circ}\text{C}$. In order to account for the fixed step problem, an additional input namely week is introduced into the 2 hidden-layer Multilayer Perceptron.

The training set consists of SST data for three years (1990-1992). The algorithms are repeatedly presented in the training set until the convergence of the weight vector. The performance of the algorithms is evaluated by using a test set which consists of 2 years data following the training set.

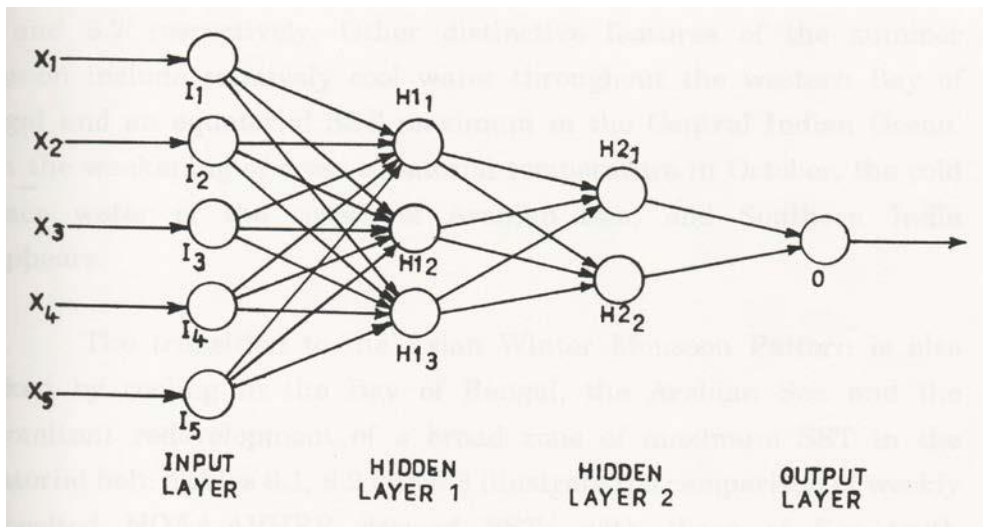


Fig.3: Network showing the prediction of SST for the oceanographic data.

- At the input layer, $1^{\circ} \times 1^{\circ}$ weekly average SST data for a particular latitude and longitude is fed as an input to the network. The output is the predicted temperature.

The surface waters in the Indian Ocean experience little annual temperature variation and stay relatively cool throughout the year. During the winter monsoon months of November to February, a broad zone of maximum SST occurs in the equatorial belt from which temperature decreases Northwestward into the Bay of Bengal and the Arabian Sea [16].

The SST pattern characteristics of the Asian summer monsoon evolves gradually from March to June, being partially dependent on the establishment of the South-West monsoon flow itself. Other distinctive features of the summer monsoon include relatively cool water throughout the western Bay of Bengal and an equatorial SST maximum in the Central Indian Ocean. With the weakening of cross-equatorial temperature in October, the cold surface water of the coasts of Arabian Sea, and Southern India disappears. The changes in SST in all the three oceans during the winter and summer monsoon by considering the SST values for the months of December and May are shown in Fig. 4, 5, 6 and 7 respectively. For comparison, the SST calculated using the conventional stochastic method is also shown in the Figs. 4 and 6. In the Figures, a few test samples of the SSTs are shown for different region of the Arabian Sea, Indian Ocean and Bay of Bengal. Actual SST values obtained from AVHRR Satellite are given for various region in the Figures. In the test samples shown in the Figures, the SST values are in the order of Actual, Stochastic and BP (Figures 4 and 6) and in the order of Actual, TD and MTD (Figures 5 and 7).

The transition to the Asian Winter Monsoon Pattern is also marked by cooling in the Bay of Bengal, the Arabian Sea and the concomitant redevelopment of a broad zone of maximum SST in the equatorial belt. Tables 1, 2 and 3 illustrate the comparison of weekly composited NOAA-AVHRR derived SSTs with those of Sea truth observations for the three different oceans namely the Arabian Sea, the Bay of Bengal and the Central Indian Ocean. It is observed that the MTD values more nearly coincide with the AVHRR values than the BP and TD values.

VI.B Performance Evaluation of Stochastic, BP, TD and MTD Methods

The stochastic techniques are really two different procedures namely the Wiener approach and the auto-regressive scheme. Auto-regressive approach advocates a class of linear stochastic models of auto-regressive nature not necessarily stationary. A distribution item is included that designates a class of independent identically distributed random variables with zero expectation and positive finite variance. It deals with the simultaneous solution of large sets of normal equations to extract the relevant parameters which are determined on the least squares minimization principle.

In backpropagation algorithm, the training inputs have to be compiled beforehand i.e., the observations are reconstructed into the form (X_1, Y_1) (X_2, Y_2) (X_3, Y_3) etc. This is because the algorithm needs to know the actual outcome for backpropagation of the error and it is not available till a certain amount of time elapses. Thus extra preprocessing is needed and as a result it repeatedly gets entangled into local minima problems [17]. To overcome this, a momentum term and several reinforcing variables that supplement the original factors are added to the input layer of the Multilayer Perceptron. Second order and third order terms are used to form these variables.

TD methods demonstrated faster learning than the backpropagation algorithm described above. The γ value is varied from 0 to 0.9 in steps of 0.1 and the learning rate is selected to yield the lowest error for a specific value of β of 0.3. The exponential recency is not

appropriate for prediction for a fixed time later since it has a large negative slope as illustrated in the weighing scheme.

MTD method uses an alternative recency factor $1/(1 + \lambda^k)$ instead of the conventional exponential term. The weighing schemes for the conventional TD and MTD methods provide a direct comparison of these two methods. Both $TD(\lambda)$ and $MTD(\lambda)$ represent a decaying function whose value at $k = 0$ is 1 and α at $k = 1$. It is clear that the MTD method provides a lesser slope than TD methods for corresponding values of α . The performance of Stochastic, Backpropagation, TD and MTD methods and their optimum predictions are also studied and compared.

VII. Conclusion

The structural and temporal credit assignment issues in making prediction for a specified time later have been addressed. TD methods or backpropagation algorithm cannot by itself address both the issues. The relative shortcomings of both are discussed in detail. It is thus found that a new technique, designed by combining backpropagation and TD methods with proposed recency factor, produces better prediction for a fixed time later. For the weekly prediction of SST, recursive equations consisting of overlapping sequences of interrelated predictions at different times are formed. The above factors describe the SST pattern characteristics of the summer and winter monsoon in the three oceans namely the Indian Ocean, the Bay of Bengal and the Arabian Sea.

By combining back-propagation and TD methods, a new scheme is designed for making prediction for a specified time later. Two such networks are developed - one trained with the conventional back-propagation algorithm and the other with Modified Temporal Difference (MTD) methods. The back-propagation algorithm has been modified so that it back-propagates the temporal error rather than the conventional error. The changes in the SST throughout the year in the Arabian Sea, the Central Indian Ocean and the Bay of Bengal are studied. A weighing scheme is introduced in which alterations to past predictions are made according to both exponential decay function and the proposed recency factor. The root mean squared error values are arrived at by considering actual values and the values obtained by other means like stochastic method, back-propagation algorithm, TD and MTD methods.

The Modified Temporal Difference method has been successfully used to predict the Sea Surface Temperature in the three oceans. The choice of an alternative recency factor in this method leads to better prediction than the exponential decay. It is found that statistical methods lack adaptability and backpropagation methods suffer from local minima problems. And TD methods also suffer from inappropriate prediction for a fixed time later since it has a large negative slope in the weighing scheme. It is proven that the proposed MTD method yields better prediction for a fixed time later.

Acknowledgement

The authors gratefully acknowledge the data provided for this research by the National Remote Sensing Agency (NRSA), Hyderabad, India.

References:

1. M. G. Kendall, 'Time Series', Charles Griffin and Company Ltd., London and High Wycombe, 1976.
2. N. Wiener, 'Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications', MIT Press, 1949.
3. D. R. Cox, 'Long-Range Dependence: a Review', In David H. A. and David H. T. (Eds.), 'Statistics: An Appraisal', Proceedings of the 50th Annual Conference, Iowa State Statistical Laboratory, 1984.
4. R. S. Sutton, 'Learning to Predict by the Methods of Temporal Differences', Machine Learning Journal, Vol. 3, 1988, pp.6-44.
5. C. W. Anderson, 'Learning and Problem Solving with Multilayer Connectionist Systems', Doctoral Dissertation, Dept. of Computer Information Science, University of Massachusetts, Amherst, 1986.
6. R. Narendran, V. Ganapathy, and M. V. Somasundaram, 'Faster Learning with a Combination of Backpropagation and Temporal Difference Methods for Prediction Problems', ANNSI-93, BARC, Trombay, Bombay, 1993.
7. R. Narendran, V. Ganapathy, and M. V. Somasundaram, 'A Study on the Effects of Recency Factors on Prediction in Real World Domains', Proc. Of the IEEE Int. Conf. On Neural Networks, Orlando, Florida, 1994.
8. R. S. Sutton, A. G. Barto, 'An Adaptive Network that Constructs and Uses an Internal Models of its Environment', Cognition and Brain Theory, 4, 1981, pp.217-246.

9. B. Widrow, and M. E. Hoff, 'Adaptive Switching Circuits', IRE Western Electric Show and Convention Record, Part 4, New York, 1960, pp.96-104.
10. A. Narendranath, M. V. Rao, and K. H. Rao, 'Technical Note: Observed High Temperatures in the Sunlight Area Over the North Indian Ocean', Int. Journal on Remote Sensing, Vol.14, No.5, 1993, pp.849-853.
11. M. A. Sartori, and P. J. Antsaklis, 'A Simple Method to Derive Bounds on the Size and to Train Multilayer Neural Networks', IEEE Trans. on Neural Networks, Vol.2, 1991, pp.467-471.
12. R. S. Sutton, 'Temporal Credit Assignment in Reinforcement Learning', Doctoral Dissertation, Dept. of Computer and Information Science, University of Massachusetts, Amherst, 1984.
13. R. S. Sutton, and A. G. Barto, 'A Temporal-Difference Model of Classical Conditioning', Proc. Of the 9th Annual Conf. Of the Cognitive Science Society, Seattle, WA: Lawrence Erlbaum, 1987, pp. 355-378.
14. K. L. Wyrki, 'Oceanographic Atlas of the International Ocean Expedition', National Science Foundation, US Government Printing Office, Washington DC, 1971.
15. J. C. Price, 'Analysis of Some Methods for Obtaining Sea Surface Temperature from Satellite Observations', NRSA Rep., 1973, pp.20, 108, 651-73.
16. S. R. Shetye, 'A Model Study of the Seasonal Cycle of the Arabian Sea Surface Temperature', Journal of Marine Research, 44, 1986, pp.521-542.
17. Xiao-Hu Yu, 'Can Backpropagation Error Surface Not Have Local Minima', IEEE Trans. on Neural Networks, Vol.3, No.6, 1992.

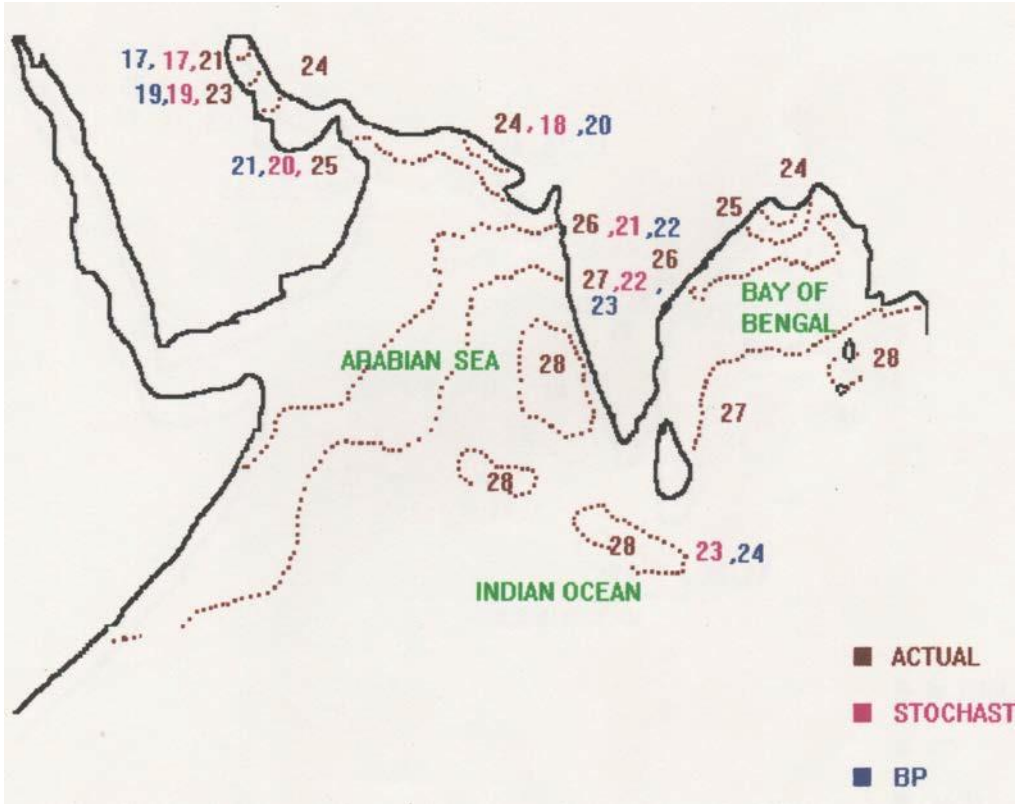


Fig.4: Changes in SST in all the three Oceans during the winter monsoon (December) (Actual, Stochastic and Backpropagation values).

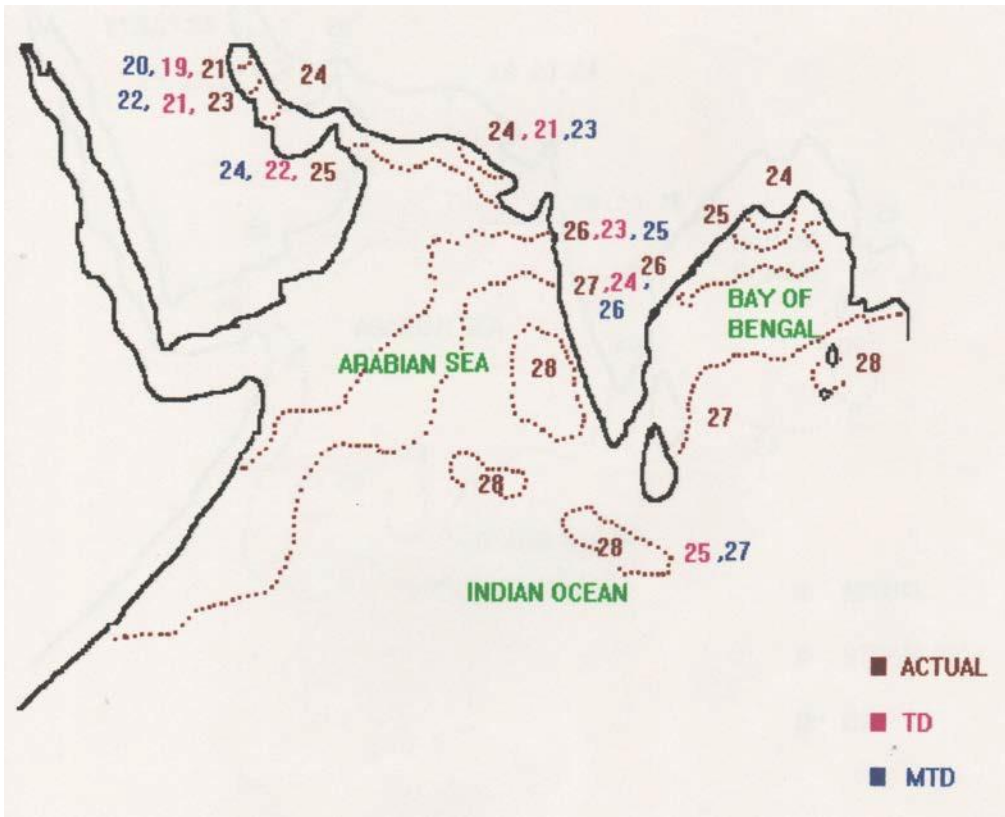


Fig.5: Changes in SST in all the three Oceans during the winter monsoon (December) (Actual, TD and MTD values).

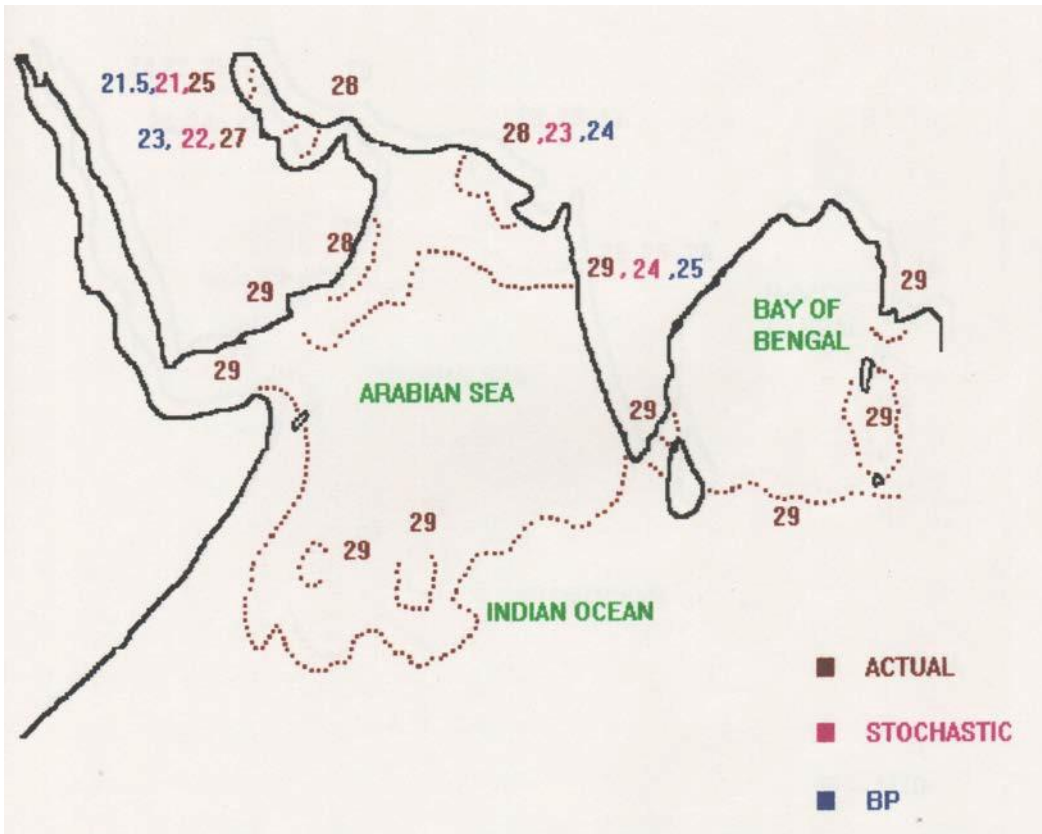


Fig.6: Changes in SST in all the three Oceans during the summer monsoon (May) (Actual, Stochastic and Backpropagation values).

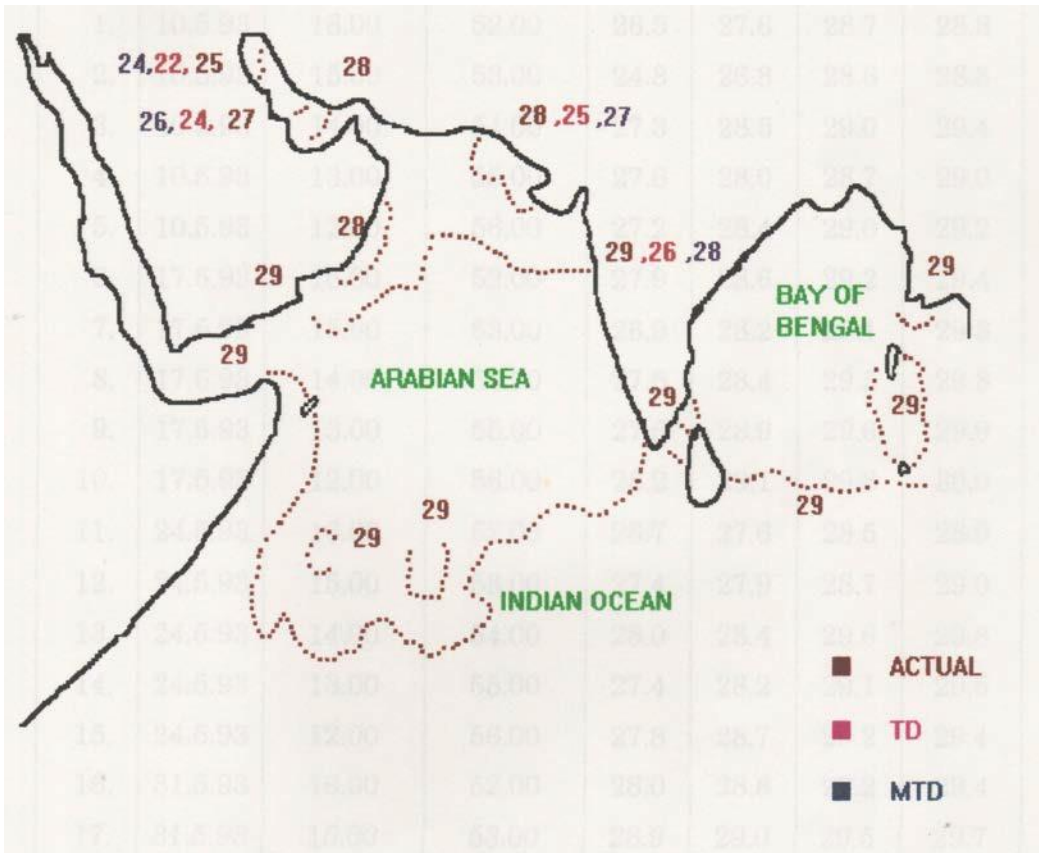


Fig.7: Changes in SST in all the three Oceans during the summer monsoon (May) (Actual, TD and MTD values).

Table 1: Comparison of Weekly Compositated NOAA-AVHRR Derived SSTs with those of Observations Obtained by Using Different Methods for the Arabian Sea

Sl. No.	Date	Latitude	Longitude	BP	TD	MTD	AVHRR
1	10.5.93	16.00	52.00	26.5	27.6	28.7	28.8
2	10.5.93	15.00	53.00	24.8	26.8	28.6	28.8
3	10.5.93	14.00	54.00	27.3	28.5	29.0	29.4
4	10.5.93	13.00	55.00	27.6	28.0	28.7	29.0
5	10.5.93	12.00	56.00	27.2	28.4	29.0	29.2
6	17.5.93	16.00	52.00	27.9	28.6	29.2	29.4
7	17.5.93	15.00	53.00	26.9	28.2	29.1	29.3
8	17.5.93	14.00	54.00	27.5	28.4	29.5	29.8
9	17.5.93	13.00	55.00	27.6	28.9	29.6	29.9
10	17.5.93	12.00	56.00	28.2	29.1	29.8	30.0
11	24.5.93	16.00	52.00	26.7	27.6	28.5	28.9
12	24.5.93	15.00	53.00	27.4	27.9	28.7	29.0
13	24.5.93	14.00	54.00	28.0	28.4	29.6	29.8
14	24.5.93	13.00	55.00	27.4	28.2	29.1	29.5
15	24.5.93	12.00	56.00	27.8	28.7	29.2	29.4
16	31.5.93	16.00	52.00	28.0	28.8	29.2	29.4
17	31.5.93	15.00	53.00	28.9	29.0	29.5	29.7
18	31.5.93	14.00	54.00	28.7	29.1	29.5	29.8
19	31.5.93	13.00	55.00	28.8	29.2	29.7	29.8
20	31.5.93	12.00	56.00	28.9	29.4	29.9	30.2
Average				27.62	28.44	29.2	29.46
Standard Deviation				1.60	0.63	0.412	0.404

Table 2: Comparison of Weekly Compositated NOAA-AVHRR Derived SSTs with those of Observations Obtained by Using Different Methods for the Bay of Bengal

Sl. No.	Date	Latitude	Longitude	BP	TD	MTD	AVHRR
1	10.5.93	16.00	70.00	29.2	29.6	29.9	30.1
2	10.5.93	15.00	71.00	29.5	29.8	30.0	30.1
3	10.5.93	14.00	72.00	29.6	29.8	30.0	30.2
4	10.5.93	13.00	73.00	29.7	29.8	30.0	30.1
5	10.5.93	12.00	74.00	29.8	30.0	30.2	30.3
6	17.5.93	16.00	70.00	29.5	29.8	30.0	30.1
7	17.5.93	15.00	71.00	29.0	29.4	29.9	30.1
8	17.5.93	14.00	72.00	29.4	29.6	30.0	30.3
9	17.5.93	13.00	73.00	29.2	29.5	29.8	30.0
10	17.5.93	12.00	74.00	29.5	29.7	30.0	30.3
11	24.5.93	16.00	70.00	29.4	29.8	30.0	30.2
12	24.5.93	15.00	71.00	29.4	29.7	30.0	30.1
13	24.5.93	14.00	72.00	29.7	29.8	30.1	30.2
14	24.5.93	13.00	73.00	29.4	29.8	30.0	30.1
15	24.5.93	12.00	74.00	29.6	29.9	30.1	30.3
16	31.5.93	16.00	70.00	28.8	29.2	29.4	29.8
17	31.5.93	15.00	71.00	28.9	29.3	29.6	29.9
18	31.5.93	14.00	72.00	28.8	29.2	29.7	30.0
19	31.5.93	13.00	73.00	29.0	29.6	29.8	30.1
20	31.5.93	12.00	74.00	29.4	29.7	30.0	30.3
Average				29.34	29.65	29.93	30.13
Standard Deviation				0.30	0.22	0.18	0.13

Table 3: Comparison of Weekly Composited NOAA-AVHRR Derived SSTs with those of Observations Obtained by Using Different Methods for the Central Indian Ocean

Sl. No.	Date	Latitude	Longitude	BP	TD	MTD	AVHRR
1	10.5.93	21.00	89.00	26.0	26.7	27.0	27.3
2	10.5.93	20.00	90.00	26.1	26.8	27.1	27.4
3	10.5.93	19.00	91.00	26.8	27.0	27.5	27.9
4	10.5.93	18.00	92.00	27.1	27.9	28.4	28.7
5	10.5.93	17.00	93.00	27.0	27.8	28.6	28.9
6	17.5.93	21.00	89.00	26.5	27.0	27.2	27.7
7	17.5.93	20.00	90.00	26.3	26.9	27.2	27.9
8	17.5.93	19.00	91.00	27.0	27.8	28.0	28.2
9	17.5.93	18.00	92.00	27.3	27.9	28.4	28.7
10	17.5.93	17.00	93.00	27.8	28.0	28.4	28.7
11	24.5.93	21.00	89.00	26.7	27.0	27.4	27.7
12	24.5.93	20.00	90.00	26.9	27.1	27.6	27.9
13	24.5.93	19.00	91.00	26.8	27.5	27.9	28.2
14	24.5.93	18.00	92.00	27.5	28.0	28.5	28.7
15	24.5.93	17.00	93.00	27.3	27.9	28.5	28.7
16	31.5.93	21.00	89.00	26.5	26.9	27.5	27.7
17	31.5.93	20.00	90.00	26.6	27.0	27.6	27.9
18	31.5.93	19.00	91.00	26.8	27.0	27.5	27.9
19	31.5.93	18.00	92.00	27.6	28.0	28.4	28.6
20	31.5.93	17.00	93.00	27.8	28.0	28.5	28.7
Average				26.92	27.41	27.86	28.17
Standard Deviation				0.50	0.49	0.54	0.48