The salinity data in the ocean are non-uniform and irregular, so methods for salinity estimation using available predictors (i.e. temperature data or others) are urgently needed. For the purpose of calculating salinity profiles in the top 500 meters of the Red Sea, a collection of regression models based on surface salinity, temperature profiles, and other variables are provided. The curvature observed in temperature is captured using both surface salinity and temperature. Julian day is utilized to capture seasonal variability in the region, while salinity plots, latitude, and longitude are employed to capture regular spatial changes over the fitting regions. Thus, for the studied region, the lowest errors of the salinity estimations and the best-fit regression curve have been determined. In terms of temperature, this regression model is quadratic for the whole region at all depths, and linear for surface salinity, longitude, latitude, and day of the year. Even without the surface salinity measurement we can estimate the salinity with good reduction of RMS errors for all depths below 150m.

INTRODUCTION

The Red Sea is a narrow, semi-enclosed basin extending between 12.5°N and 30°N latitude. Its average width and depth are 220 km and 524m, with maximum depth of about 3000m (Fig. 1). The Bab el-Mandeb Strait, which connects to the Gulf of Aden at its southernmost point and has a minimum width of around 25 km and a maximum depth of 160 m, is the only outlet to the Indian Ocean (Radwan, 2008).

The World Ocean Database is available for measurements of temperature (T) and salinity (S), but unfortunately its coverage of the Red Sea is almost non-uniform. There are two methods to use these data in various applications because of this problem. First, a one-degree grid is interpolated using the climatic mean values of T and S, which are conventionally used and taken from this database. The second approach is to create techniques that, using temperature readings and the statistical link between S and T, enable salinity to be estimated at any location in the ocean. For the purpose of initializing numerical models and assimilation of data, the second approach appears very promising as it can result in a more uniform and accurate distribution of T-S data over the Global Ocean research region (Thacker et al., 2007 and Korotenko, 2007).
It has long been known that salinity can be estimated in the ocean by statistically correlating salinity with temperature. The formula for determining salinity is based on the connection $\hat{S}(z) = \hat{S}[T(z)]$ that was proposed by Stommel (1947), in which $\hat{S}(z)$ is the average salinity value at a given temperature and $\hat{S}(z)$ is an estimate of the salinity fluctuation with depth ($z$). The so-called salinity mean approach $\hat{S}(z) = \langle S(z) \rangle$, where $\langle S(z) \rangle$ is the climatological mean salinity value, can yield a more accurate estimate of salinity in some places (Emery and Brien, 1978). An additional adjustment to Stommel's (1947) approach of estimating salinity using temperature profiles and salinity readings at the ocean's surface was put forth by Dungey et al. (1986).

The primary limitation of the Stommel method (1947) and all of its modifications is that while all methods are completely worthless in strata where $T$ and $S$ have different depth-dependence functions, they can replicate (recover) salinity anomalies that are associated with temperature anomalies (Korotenko, 2007). The so-called barrier layers are these.

Hansen and Thacker (1999) and Hussein (2016) proposed and implemented a regression method for salinity estimates from temperature measurements, which included the introduction of salinity's dependence on temperature, season, and geographic location, as well as its dependence on latitude, longitude, and day of the year. General regression equation (relationship) is written as:
\[ S(z) = \langle S(z) \rangle + \sum_i a_i(z) (P_i - \langle P_i \rangle) \quad (1) \]

Whereas variables utilized as predictors are indicated by \( P_i \), climatological averages are represented by angle brackets, and parameter values \( a_i \) are obtained via regressions at each depth \( z \). Adjustments to the climatic salinity profile caused by observed predictor deviations from their climatological averages are known as estimates.

**Hussein et al. (2011)** examined salinity estimations for the upper 500m layer in the southeast Mediterranean Sea for several predictor versions in (Eq. 1). Only the upper 50 m layer's estimation error could be reduced by the authors by adding surface salinity; the 50–500 m layer's estimation error could be reduced by computing longitude.

The use of regression relationships with high-order polynomials to predict \( P_i \) was the next stage in the development of regression approaches for predicting ocean salinity. Consequently, the regression that follows was used in place of (1):

\[ \hat{S}(z) = \sum a_{i,l}(z) P_{l,i} + \varepsilon \quad (2) \]

For instance, the following is how this connection is expressed when dealing with \( n \)-degree polynomials for temperature:

\[ \hat{S}(z) = a_0 + a_{T,1} T + a_{T,2} T^2 + a_{T,n} T^n \quad (3) \]

Attempts were made by Thacker (2007) and Thacker and Sindlinger (2007) to find appropriate regression equations. To estimate the Gulf of Mexico salinity and in the northwest Atlantic area \((25°-45°N \times 65°-35°W)\) based on NODC WOD-2001 data. In the Gulf of Mexico region, the second-order polynomial in Eq. (3) was sufficient to approximate the salinity picture.

**Hussein (2016)** applied different forms of predictors in Eq. (3) to salinity estimates for the upper 500m layer in the southeastern Mediterranean region. In this region, the surface salinity added to the fourth-order polynomial in Eq. (3) was better for the upper 130 m while when adding longitude \( X \), latitude \( Y \) and day \( D \) of the year to the third-order polynomial in Eq. (4). It was the best relative to the rest of the depths.

\[ \hat{S}(z) = a_0 + a_{T,1} T + a_{T,2} T^2 + a_{T,3} T^3 + a_{lon} X + a_{lat} Y + a_{day} d \quad (4) \]

The use of date \( d \) of measurements, reflecting the dependence of observations on the season, along with temperature and surface salinity \( S(0) \), longitude \( X \), and latitude \( Y \), gave an opportunity for improvement. Salinity estimates.

This work aim to estimate the salinity profile of the Red Sea (in the upper 500 m) from two models: the first one from the measurements of temperature profiles only and second from the temperature and sea surface salinity profiles. These profiles can be utilized for applications where salinity profiles might be required or for estimating salinity profiles that can be included into numerical circulation models.

### MATERIALS AND METHODS

The locations of the 111 CTD profiles that were selected to be employed in this study to establish empirical relationships between salinity and temperature for the region spanning \( 5° \times 6° \) (33.5°E to 38.5°E and 22°N to 28°N) (Fig. 2). For this study, the NODC
The WOD-2001 database was used and represents the period between 1962 and 2011. The data covers each profile at the following standard depths 0, 10, 20, 30, 50, 75, 100, 150, 200, 300, 400, and 500 meters. Only the data for the upper 500m were utilized because that region had the most profiles accessible.

The 111 temperature and salinity profiles were split into two sets: the training data for the model fitting process consisted of 74 profiles, and the remaining 37 profiles were placed aside for independent verification (Fig. 2). The scatter plots of T-S for the research region at different depths (Fig. 3).

![Fig. 2. Selected data profiles locations of the 111 CTD stations in study area as training and verification sampling stations](image)

The mean of salinity profiles \(\langle S(z) \rangle\) and temperature \(\langle T(z) \rangle\) are shown in Fig. (4) together with their standard deviations. Correlation coefficients between \(S(z)\) and \(T(z)\), and surface salinity \(S(0)\) and \(S(z)\) at different depth interval are shown in Fig. (5).

Correlation between salinity and temperature is small and negative in the upper 150m, nearly zero in 300m, also small and positive between 300m and 500m depth. Correlation with surface salinity is strong in the above 75m, moderate between 100 and 150m and negligible elsewhere. These correlations complimentary nature implies that using them together should be beneficial.
Fig. 3. Scatter plot of CTD data of Salinity and Temperature at different standard depths of the study area.
Fig. 4. The blue curve displays the average temperature (a) and salinity (b) for all 111 data sets in this research. Other curves display the standard deviations from the mean.

Fig. 5. Blue curve represents correlation coefficient between salinity $S(z)$ and temperature $T(z)$ and red curve between $S(0)$ and $S(Z)$ at different depth interval

The salinity profiles indicate that there is some variation of 1.7 in the surface mixed layer. Since the haline mixed layer frequently reaches depths of thirty meters or more, surface salinity in this area can be used as a reliable predictor of upper salinity. The dispersion between the profiles gets smaller below 200 meters below the surface. The temperature profiles show that there is comparatively little scatter below the thermocline and a great deal of variability in the surface mixed layer. Under 23°C, the T-S relationship is clearly defined; but, as one approaches the surface, the relationship gets less and less defined (Fig. 6).
The analysis methods

Finding regression models for each pressure level that can explain the data in (Fig. 3) is the first step in the salinity estimation process. The independent verification data for the respective levels will be used to evaluate these models' skill. The scatter plot in Figure 3 indicates that a first-degree (linear) polynomial of temperature or a higher degree might be used to describe salinity, and this turned out to be the case. The training data at each pressure level was fitted with a polynomial of temperature with varying degrees. At each depth level, the following equations (around 12 different types of regression models, from Eq. (5) to Eq. (16)) were applied:

\[ S = P_1(T) = a_0 + a_1T \]  
\[ S = P_2(T) = a_0 + a_1T + a_2T^2 \]  
\[ S = P_3(T) = a_0 + a_1T + a_2T^2 + a_3T^3 \]  
\[ S = P_4(T) = a_0 + a_1T + a_2T^2 + a_3T^3 + a_4T^4 \]  
\[ S = P_3(T) + \text{day} + \text{lat} + \text{long} = a_0 + a_1T + a_2T^2 + a_3T^3 + a_4d + a_5x + a_6y \]  

The next regression models were used at each depth level corresponding to the combinations between \( P_2(T), P_3(T), P_4(T) \) and surface salinity in addition to day, latitude and longitude.

\[ S = P_2(T) + S(0) = a_0 + a_1T + a_2T^2 + a_3S(0) \]  
\[ S = P_3(T) + S(0) = a_0 + a_1T + a_2T^2 + a_3T^3 + a_4S(0) \]  
\[ S = P_4(T) + S(0) = a_0 + a_1T + a_2T^2 + a_3T^3 + a_4T^4 + a_5S(0) \]  
\[ S = P_3(T) + \text{day} + \text{long} = a_0 + a_1T + a_2T^2 + a_3T^3 + a_4S(0) + a_5 \text{long} \]  
\[ S = P_3(T) + S(0) + \text{lat} = a_0 + a_1T + a_2T^2 + a_3T^3 + a_4S(0) + a_5 \text{lat} \]  
\[ S = P_3(T) + S(0) + \text{day} = a_0 + a_1T + a_2T^2 + a_3T^3 + a_4S(0) + a_5 \text{day} \]  
\[ S = a_0 + a_1T + a_2T^2 + a_3S(0) + a_4d + a_5x + a_6y \]  

Fitting to the local training data yielded the coefficients for each model: \( a_0, a_1, a_2, a_3, a_4, a_5, a_6, \) and \( a_7 \). Where \( T, S(0), d, x, \) and \( y \) stand for measured temperature, surface salinity, Julian day of the year, longitude, and latitude, respectively, and \( S \) represents the

---

**Fig. 6.** Profiles of temperature and salinity and temperature-salinity scatter plot for the 111 data sets used in this study.
estimated salinity. The root mean square (RMS) errors has been used to represent discrepancies between the measured and predicted salinities for the verification.

RESULTS AND DISCUSSION

I. A regression methods without mean salinity (<S(z)>)

For the upper 500 m, salinity profiles were estimated using various polynomial degrees of temperature profile. For each of the 37 profiles in the verification data set, there are 12 different regression technique iterations from Eqs (5–16). The RMS error illustrated in Figs. (7, 8).

I.1. The temperature polynomial (TP) models

For the first five types of models for Eqs. (5-9) the RMS error decreased with depth. The first four models used Eqs. (5-8) are nearly indistinguishable. The third which comes from Eq. (7) and fourth from Eq. (8) models enhanced RMS error than first of Eq. (5) and second of Eq. (6) models as shown in Fig. (7a). For depths between 200m and 400m, i.e., the third and fourth degree polynomials of temperature may be used to predict salinity below the thermocline with RMS errors less than 0.06 due to the positive tight association between temperature and salinity (Fig. 7a). By adding day, latitude and longitude to third degree of polynomial (Eq. 9) RMS errors enhanced than the previous Eqs. (5-8) between surface and depths less than 200m. Below 200m until 400m depth it is coincide with the Eqs. (5-8) models. Below 400m depth, RMS reached to 0.045 (Fig. 7a).

I.2. TP and surface salinity (TPS) models

For the top 75 m depth (0.69 to 1), the next three model for Eqs. (10, 11 & 12) proposed benefiting from a high correlation between S(0) and the S(Z). Below 200 meters, the association between S(0) and S(Z) is significantly reduced. One problem with using surface salinity readings is that they don't fully account for the fluctuation that exists in the top tens of meters. The upper 250 m depth's RMS values, which ranged from 0.05 to 0.08, were more accurate than the results from TP models. It rose to 0.12 between 50 and 100 meters, altered from 0.06 to 0.08 between 100 and 200 meters, and varied between 0.04 and 0.07 below 200 meters (Fig. 7b). The RMS range values of the three models (Eqs. 10 to 12) at all depths between surface and 100m were represented nearly indistinguishable. The RMS values of the regression model of Eq. (11) were lower than the other two models of Eqs. (10 & 12) for the depth between 100m and 400m, and coincide with model of Eq. (12) at 500m depth. In general, the RMS range values of these three models were better than the TP regression models which have been described in the previous section.

Data must be taken from a large region (such as 5°x6° region) in order to provide statistically meaningful results. Over this region, horizontal gradients of water characteristics may considerably contribute to the variations about the mean profiles. Because the climatological structure in the Red Sea may be largely zonal, the Latitude (y) and Longitude (x) would be added to the other predictors in order to assess the feasibility of capturing some of this variability as those outlined below. As indicated in Eqs. (13, 14 and 15), the longitude, latitude, and day of the year were added separately to P3(T), and the result was provided by Eq. (16).
I.3. TPS, longitude, latitude and day of the year

There is no RMS error valuable difference between models of Eq. (13, 14, 15 & 16) as shown in Fig. (8a); RMS values of these models nearly coincide with each other.

In comparison between the RMS values of Eqs. (12 and 16) (Fig. 8b), the output of the model of Eq. (16) illustrates that RMS errors enhanced salinity estimation until 150m depth than model of Eq. (12) and coincide with each other downward 500m depth.

II. A regression method with mean salinity <S(z)>

Using the mean salinity profile and six different regression procedures for the 37 profiles in the verification data set, salinity profiles were computed for the top 500 m. For the verification profiles, the RMS discrepancies between the estimated and measured salinities were calculated and displayed in Fig. (9a, b).

II.1. The mean salinity method

This method examines the estimation of salinity by its climatologically mean:

$$\hat{S}(z) = \langle S(z) \rangle$$  \hspace{1cm} (17)

This method's RMS errors were labelled “mean salinity” and displayed in Fig. (9a). Because of the discrepancy between the training and verification data, errors somewhat
surpass the variability near the surface. This technique records a modest amount of fluctuation at 500m and higher variability at 200m deep.

II.2. The temperature (T) method
Combining these approaches, utilizing observed temperature to improve upon the climatological mean salinity in a regression model, is suggested by the complementarity of depths at which the T-S and climatologically mean methods work best (Hansen & Thacker 1999 and Hussein, 2011):

\[ \hat{S}(z) = \langle S(z) \rangle + a_T(z) [T(z) - \langle T(z) \rangle] \]  

(18)

The training set of data was used to fit this approach as well as the ones that follow. The coefficients varied with depth even though the model was fitted independently at every level of the chosen interval. The RMS errors, shown in Fig. (9a) as “Temperature”; show that this approach realizes better than the mean salinity method in the upper 30m; in the depth intervals 50 to 100 m and 200m to 500m the two methods are equivalents; in the depth interval between 100m and 150m RMS value decreased. These results are corroborated by the correlation theme in Fig. (5), which shows that salinity and temperature have a moderately negative correlation for the top 30m layer, that correlation increases between the depth intervals of 100 and 150 m, and that it changes to a positive, low correlation value depth interval below 250 to 500m.

II.3. The surface salinity (SS) method
Addressing to the issue of using surface salinity data to partially capture the variability found in the top several tens of meters. Initially, we examine their application when a temperature profile is not detected. Regression equations show how variations in the measured surface salinity from its mean affect the estimates derived from the mean salinity profile at each depth \( z \):

\[ \hat{S}(z) = \langle S(z) \rangle + a_S(z) [S(0) - \langle S(0) \rangle] \]  

(19)

The RMS estimation error, labelled "Surface salinity" in Fig. (9a), is decreased to 0.15 in the upper 50m and to 0.1 in the range between 75m and 150m depth due to the substantial correlation of \( S(z) \) with \( S(0) \) in the upper 100m of the water column. Beyond 200 meters below the surface, surface salinity becomes useless, the regression coefficient

Fig. 9. RMS errors for different approaches to the estimation of the salinity profiles in the verified data. RMS errors for surface salinity, temperature, and mean salinity profile regressions (a). RMS errors for regression on latitude, longitude, and both surface salinity and temperature (b)
a_s(z) approaches zero, and the "surface salinity" curve in Fig. (9a) aligns with the "mean salinity" and "temperature" curves.

II.4. The SS and T (Both) method

To estimate salinity deviations, it is simple to incorporate temperature and surface salinity profile variations from their means.

\[ \tilde{S}(z) = \langle S(z) \rangle + a_T(z) [T(z) - \langle T(z) \rangle] + a_s(z) [S(0) - \langle S(0) \rangle] \]  

(20)

The coefficients values \( a_S \) and \( a_T \) must be found by fitting Eq. (20) to the verification data; they are not the same as those for Eqs. (18) and (19). As \( z \) decreases with depth to virtually zero and \( a_T(z) \) is roughly equal to that determined for Eq. (18), it turns out that at depths where surface salinity provides no information about the subsurface salinity. The curve in Fig. (9b) labelled "Both" shows that this extension does not improve the use of surface salinity at the surface, but it does marginally reduce errors at all depths deeper than 75 m.

II.5. The SS and T (Both) plus latitude

The Latitude was added in this method to the set of predictor as shown in Eq. (21). The results were indistinguishable from those obtained using both Surface salinity and temperature method as shown in Fig. (9b).

\[ \tilde{S}(z) = \langle S(z) \rangle + a_T(z) [T(z) - \langle T(z) \rangle] + a_s(z) [S(0) - \langle S(0) \rangle] + a_y [y - \langle y \rangle] \]  

(21)

II.6. The SS and T (Both) plus longitude

The Longitude was added in this method to the set of predictor as shown in Eq. (22). The results also like previous method were indistinguishable from those obtained using both, surface salinity and temperature method or surface salinity and temperature plus latitude as shown in Fig. (9b).

\[ \tilde{S}(z) = \langle S(z) \rangle + a_T(z) [T(z) - \langle T(z) \rangle] + a_s(z) [S(0) - \langle S(0) \rangle] + a_x [x - \langle x \rangle] \]  

(22)

II.7. The SS and T (Both) plus day, latitude and longitude

According to Eq. (23), the day \( d \) was added to the set of predictors in this manner. Similar to the previous technique, the findings could be distinguished in certain depths from the ones produced by combining the methods of surface salinity and temperature and surface salinity and temperature plus latitude (Fig. 10).

\[ \tilde{S}(z) = \langle S(z) \rangle + a_T(z) [T(z) - \langle T(z) \rangle] + a_s(z) [S(0) - \langle S(0) \rangle] + a_y [y - \langle y \rangle] + a_x [x - \langle x \rangle] \]  

(23)

The current work presents the findings from the seven prior methods of Eqs. (17 to 23) for calculating salinity profiles, including one traditional method that used mean salinity. The curves depicted in Figs. (9 and 10) self-organize into three near-surface classes and three distinct classes in deepwater environments. At a depth of about 300 meters, the seven techniques are almost identical. In the vicinity of the surface, the mean salinity approach exhibits the greatest inaccuracies. With the addition of temperature data, the mean salinity is marginally improved. The RMS estimation error is significantly decreased when salinity alone is used as a predictor. When surface salinity data is added to climatological profiles, the estimation errors in the upper 50 m are reduced to 0.15.

Regression analysis shows that temperature data is more beneficial to include at depths between 50 and 200 m than mean salinity. All temperature-based approaches yield comparable findings at depths more than 200 m, with errors reduced to less than 0.08. At a depth of 100 meters, surface salinity offers little improvement. Temperature and surface salinity are still the best options when it comes to latitude and longitude. Day, latitude, and longitude are added for further improvement in depths less than 50 meters, producing
almost the fewest errors over almost all depths. In all depth ranges, the model of Eq. 23 performs almost perfectly.

Finally, from the above results (Fig. 11a), model of Eq. (16) enhances the salinity estimate value than model of Eq. (23) between 50m and 75m depth and also below 200m depth. Model of Eq. (23) gives some enhancement in salinity estimate from 150m to 250m depth than model of Eq. (16).

![Fig. 10. RMS errors for models Eq. (20) and Eq. (23) (a), RMS errors for models Eq. (21) and Eq. (23) (b)](image)

In case of the surface salinity is available, the regression model represents by Eq. (16) is best one to estimate salinity in the study area. While in case of the surface salinity is not available, one can use the regression model represents by Eq. (9) which is nearly coincide with results of Eq. (16) for the depth below 75m (Fig. 11b).

The optimum model when surface salinity is available is the regression model denoted by Eq. (16). All 37 observed and predicted salinity profiles at each chosen interval of the verification data set are shown in (Fig. 12) to demonstrate the capacity of the regression model of Eq. (16) to reproduce individual salinity profiles. Furthermore, these samples' corresponding temperature profiles are displayed in (Fig. 13).

![Fig. 11. Comparison between RMS errors for models of Eq. (16) & Eq. (23) (a) and RMS errors for models of Eq. (9) & Eq. (16) (b)](image)
Case Study

Another experiment were conducted for the same study area (Red Sea) on more recent data than we used before. These data were taken from WOD for period 2000 to 2018. So, the number of experimental data was 73 profiles and the number of verification data was 34 profiles, as shown in Fig. (12). Models (9, 12 and 16) were applied to the profiles data to obtain the regression coefficients, and then we used the verification profiles data to estimate the salinity profiles based on the used model Eqs. (9, 12 and 16). The RMS errors results are as shown in the Fig. (13a, b).

Through the results obtained from the two mentioned cases study, we can confirm that the estimation of the salinity values given by Eq. (16) is the best of all, in case the surface salinity data are available. If the surface salinity data are not available, we can use Eq. (9) as it corresponds to the results obtained from Eq. 16 at a depth of 75m downward to 500m depth as shown in Fig. (13a, b).

![Fig. 12. Experimental and verification data profiles (2000 to 2018)](image)

![Fig. (13). RMS errors (a) for models of Eqs. (9, 12 and 16), RMS errors (b) for models of Eqs. (9 and 16)](image)
**CONCLUSION**

Historical data on salinity are highly erratic and non-uniform due to the unique challenges associated with measuring and observing salinity in the world ocean. In recent times, it has been feasible to identify the primary statistical relationships between salinity and various predictors, including temperature, season, and day of the year. These geographic features also contribute to the statistical relationships, making them useful for estimating ocean salinity. Using polynomial regression analysis of the relationships between salinity (S), temperature (T), latitude (x), longitude (y), and day of the year (d), the method proposed here develops regression coefficients of the oceanographic data on salinity and temperature. This makes it possible to estimate the salinity at any location of the Red Sea that provided the temperature values. In actuality, the method's application is limited to 500 meters below the surface of the Red Sea. The regression model shown by Eq. (16) is the most accurate one to estimate salinity in the research area if surface salinity is provided. If surface salinity is unavailable, the regression model represented by Eq. (9) can be utilized, as it almost perfectly aligns with the outcomes of Eq. (16) for depths less than 75 meters.

**REFERENCES**


