



**HAL**  
open science

## A comparison of Multiple Non-linear regression and neural network techniques for sea surface salinity estimation in the tropical Atlantic ocean based on satellite data

Hadger Moussa, Mohamed Benallal, Catherine Goyet, Nathalie Lefèvre, M. C. El Jai, Véronique Guglielmi, Franck Touratier

### ► To cite this version:

Hadger Moussa, Mohamed Benallal, Catherine Goyet, Nathalie Lefèvre, M. C. El Jai, et al.. A comparison of Multiple Non-linear regression and neural network techniques for sea surface salinity estimation in the tropical Atlantic ocean based on satellite data. CMACS 2014 - Colloque de Modélisation, Analyse et Contrôle des Systèmes, May 2014, Ifrane, Morocco. pp.65-78, 10.1051/proc/201549006 . hal-01140426

**HAL Id: hal-01140426**

**<https://hal.science/hal-01140426>**

Submitted on 8 Apr 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## A COMPARISON OF MULTIPLE NON-LINEAR REGRESSION AND NEURAL NETWORK TECHNIQUES FOR SEA SURFACE SALINITY ESTIMATION IN THE TROPICAL ATLANTIC OCEAN BASED ON SATELLITE DATA

H. MOUSSA<sup>1,2</sup>, M.A. BENALLAL<sup>1,2</sup>, C. GOYET<sup>1,2</sup>, N. LEFEVRE<sup>3,4</sup>, M. C. EL JAI<sup>1,2</sup>, V. GUGLIELMI<sup>1,2</sup> AND F. TOURATIER<sup>1,2</sup>

**Abstract.** Using measurements of Sea Surface Salinity and Sea Surface Temperature in the Western Tropical Atlantic Ocean, from 2003 to 2007 and 2009, we compare two approaches for estimating Sea Surface Salinity : Multiple Non-linear Regression and Multi Layer Perceptron. In the first experiment, we use 18,300 *in situ* data points to establish the two models, and 503 points for testing their *extrapolation*. In the second experiment, we use 15,668 *in situ* measurements for establishing the models, and 3,232 data points to test their *interpolation*. The results show that the Multiple Non-linear Regression is an admissible solution whether it be *interpolation* or *extrapolation*. Yet, the Multi Layer Perceptron can be used only for *interpolation*.

**Résumé.** En utilisant des mesures de Salinité et de Température à la surface de la mer, dans l'ouest de l'océan Atlantique tropical, de 2003 à 2007 puis 2009, on compare deux approches pour la prédiction de la Salinité dans l'eau de mer de surface : la Régression Non-linéaire Multiple et le Perceptron Multi Couches. Dans la première expérience, 18 300 mesures *in situ* sont utilisées dans la construction des deux modèles et 503 points pour tester leur *extrapolation*. Dans la deuxième expérience, 15 668 mesures *in situ* sont utilisées pour établir les deux modèles et 3 232 points pour tester leur *interpolation*. Les résultats montrent que la Régression Non-linéaire Multiple peut être appliquée à la fois pour l'*extrapolation* et l'*interpolation*. Cependant, le Perceptron Multi Couches ne peut être utilisé que pour l'*interpolation*.

### CONTENTS

Introduction	66
1. Datasets and methods	66
1.1. Datasets	66
1.2. MNR (Multiple Non-linear Regression) Model	69
1.3. MLP (Multi-Layer Perceptron) Model	71
1.4. Models evaluation	71
2. Results and discussion	72
2.1. Estimation of SSS using MNR	75

<sup>1</sup> IMAGES.ESPACE-DEV, Université de Perpignan Via Domitia, 52 Avenue Paul Alduy, 66860 Perpignan Cedex 9, France.

<sup>2</sup> ESPACE-DEV, UG UA UR UM IRD, Maison de la télédétection, 500 Rue Jean-Francois Breton, 34093 Montpellier Cedex 5, France.

<sup>3</sup> IRD LOCEAN, Université Pierre et Marie Curie, UMR 7159, Paris, France.

<sup>4</sup> Lab. Oceanografia Física Estuarina e Costeira, Universidade Federal de Pernambuco, Av. Arquitetura, Brasil.

2.2. Estimation of SSS using MLP	75
Acknowledgements	77
References	77

## INTRODUCTION

The tropical Atlantic Ocean (23°S - 23°N) is a highly dynamic region. It is generally characterized by high carbon dioxide fugacity ( $fCO_2$ ) in surface sea-water. Consequently, this region is considered as a source of  $CO_2$  for the atmosphere [1]. In order to quantify the variation of  $fCO_2$  in this area, we will determine the variation of the  $CO_2$  absorption capacity of the ocean [1]. Data used in such studies, are usually collected during oceanographic cruises which are expensive in terms of time and money. Hence, here we also use another data source like satellite data, to cover much wider area than oceanographic cruises.

In order to compute  $fCO_2$ , we need several parameters, such as  $SSS$  (Sea Surface Salinity) and  $SST$  (Sea Surface Temperature). Unlike other satellite parameters such as  $SST$  which are available for several years, the  $SSS$  data provided from  $SMOS$  (Soil Moisture and Ocean Salinity) mission are available only since 2010. Yet the available *in situ* data are only from 2003 to 2007 and 2009. Thus it is necessary to determine  $SSS$  from other available parameters. Figure 1 illustrates the temporal variations of  $SSS$  and  $SST$ . Figure 2.a illustrates that there is an *inverse* relationship between these two parameters (when  $SST$  increases,  $SSS$  decreases and vice versa), so we chose to use the widely available  $SST$  to estimate  $SSS$ . However, due to the wide range of latitudes (Figure 1), we have to consider the dynamically complex system of currents flows in the western tropical Atlantic Ocean [4]. Each cruise covers a limited area of this region. Figure 1 and Figure 2 show the difference of temperature and salinity between these areas. In order to consider this criterion, we include *Latitude* in the models. Figure 2.b illustrates the relation between  $SSS$  and *Latitude*. In both techniques the function can be written as follows :

$$SSS = f(Latitude, SST). \quad (1)$$

In this paper, we compare two mathematical techniques to estimate  $SSS$  from  $SST$  measured by satellite : Multiple Non-linear Regression and Multi-Layer Perceptron. Two experiments are implemented. In both experiments, we start by establishing the models from a dataset which contain these three parameters, namely *Latitude*,  $SST$  and  $SSS$ . Thus, in the first experiment, models are tested with another dataset that is outside of the original observation area, this is an "*Extrapolation*". In the second experiment, the dataset used in the test is also different from the original dataset, but it is inside the observation area, this is an "*Interpolation*". We used MATLAB™ software for the implementations of these programs.

## 1. DATASETS AND METHODS

### 1.1. Datasets

This study concentrates on one period of the year from February to April. Five datasets are used. Four of them contain data collected during the following cruises:

- Santa Maria (in 2003/2004/2005/2006/2007),
- COLIBRI (in 2006/2007/2009),
- PIRATA BRXI (in 2009),
- Monte Olivia (in 2009),

which have been performed as part of the EC-funded project CARBOOCEAN IP program (<http://dataportal.carbo-ocean.org/>). These datasets cover a large part of the Western Tropical Atlantic Ocean, with 18803 data points. The Underway  $fCO_2$ ,  $SST$ , and  $SSS$  were recorded on the research vessels. In this work, we use only  $SST$ ,

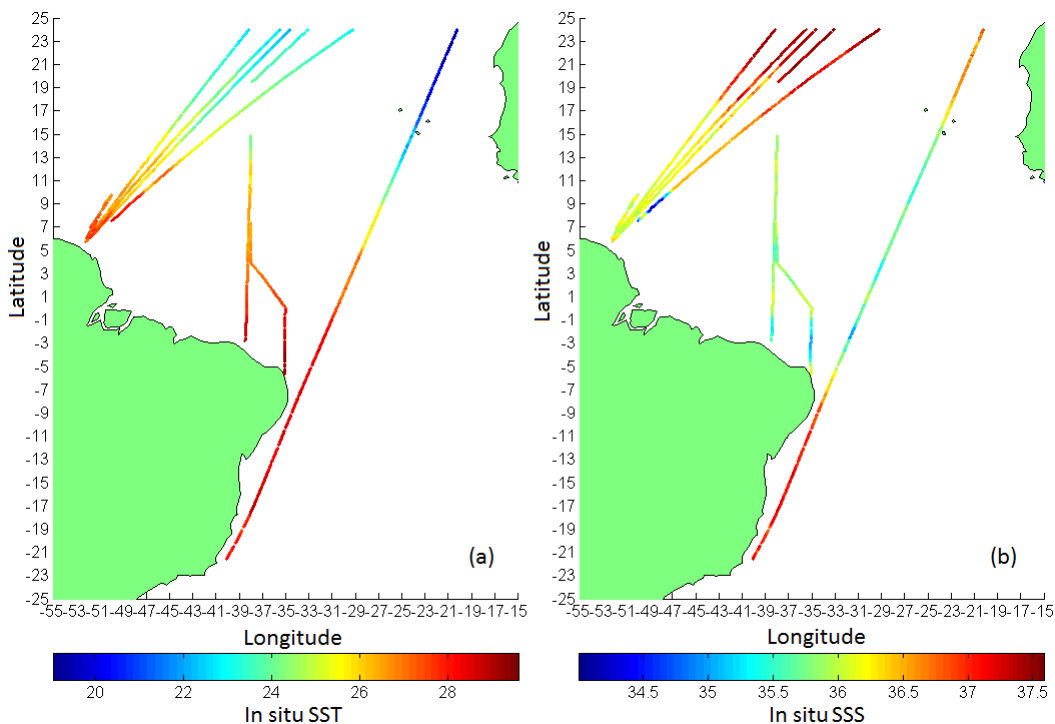


FIGURE 1. (a): *in situ* SST, (b): *in situ* SSS, collected during cruises : COLIBRI (western track, in 2006/2007/2009), PIRATA BRXI (central track, in 2009), Monte Olivia (eastern track, in 2009).

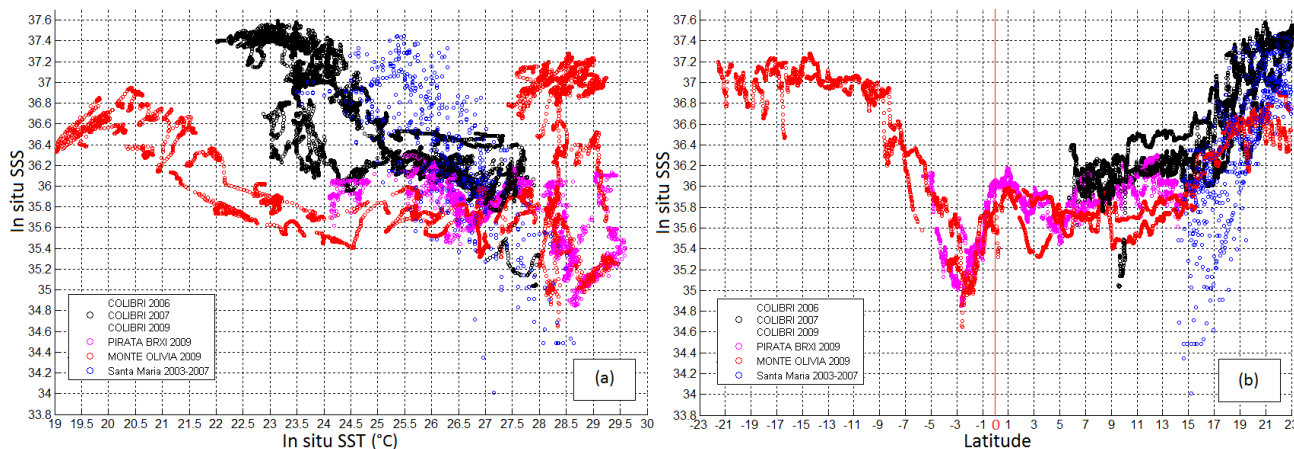


FIGURE 2. (a) : *in situ* SSS in terms of *in situ* SST. (b) : *in situ* SSS in terms of Latitude

and SSS. The fifth dataset contains SST satellite data from MODISAqua (Moderate Resolution Imaging Spectroradiometer Aqua) mission (<http://oceandata.sci.gsfc.nasa.gov/>) at 4 Km x 4 Km resolution. This satellite was launched on May 4, 2002. It passes south to north over the equator in the afternoon, and it is viewing the entire earth's surface every 1 to 2 days. Data are produced in HDF (Hierarchical Data Format) format which is the standard data format for all NASA Earth Observing System (EOS) data products. Cruises dates, tracks,

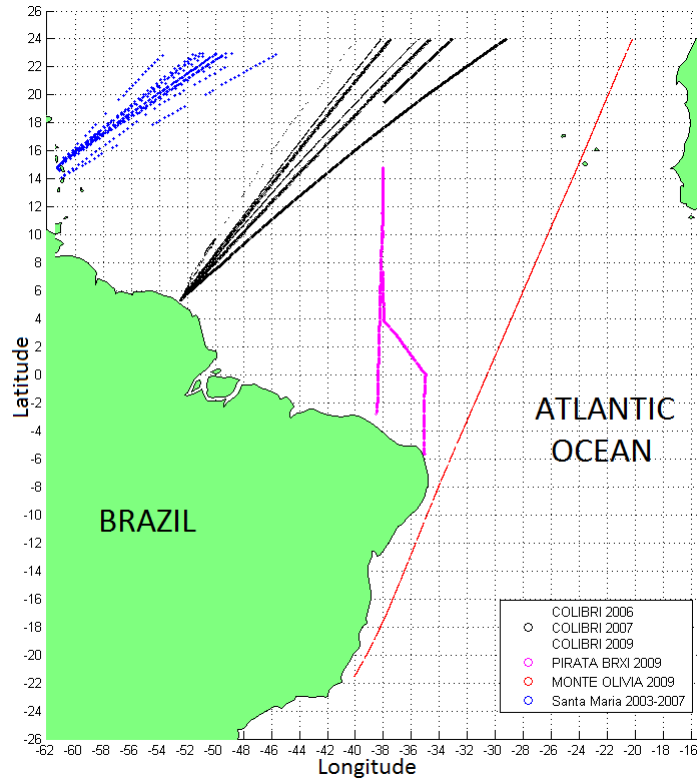


FIGURE 3. Cruises tracks : COLIBRI (2006/2007/2009), PIRATA BRXI (2009), Monte Olivia (2009), Santa Maria (2003/2004/2005/2006/2007).

Cruises	Dates	Latitude Min-Max of cruises tracks	Research vessels
Santa Maria	January - 2003	14°N - 23°N	MV Santa Maria
	Jan, Feb, Mar, April - 2004		
	Feb, Mar, April - 2005		
	Jan, Feb, April - 2006		
	Feb, Mar, April - 2007		
Colibri	March - 2006	5°N - 23°N	MN Colibri
	March, April - 2007		
	Feb, Mar, April - 2009		
Monte Olivia	March, April - 2009	23°S - 23°N	Monte Olivia
PIRATA	March, April - 2009	5°S - 15°N	N.Oc. Antares

TABLE 1. Datasets informations : cruises dates, latitudes, and research vessels.

and vessel names are presented in Table 1. Figure 3 shows the cruises tracks.

In the first experiment, Colibri, PIRATA, and Monte Olivia data are used for building the models. MODISAqua data is used only for validation. Santa Maria data cruise is used for testing the *extrapolation*. In the second experiment, models are established using data from Santa Maria, Colibri, and Monte Olivia cruises. Validation is made using data from MODISAqua. Data from the PIRATA cruise is used for testing the *interpolation*. The origin and number of data used for each model is reported in Table 2.

	Building models	Data points number	Test	Data points number
Experiment I	Colibri PIRATA Monte Olivia	18300	Santa Maria	503
			MODISAqua	9139
Experiment II	Santa Maria Colibri Monte Olivia	15668	PIRATA	3232
			MODISAqua	8247

TABLE 2. The origin and number of data used in the two experiments.

## 1.2. MNR (Multiple Non-linear Regression) Model

A simple linear regression illustrates the relation between the dependent variable  $y$  and the independent variable  $x$  based on the regression equation [5]:

$$y_i = \beta_0 + \beta_1 x_i, \quad i = 1, 2, \dots, n \quad (2)$$

The regression coefficients  $\beta_0$  and  $\beta_1$ , represent the *intercept* and the *slope*, respectively.

In our case, we have to find a relation between three variables  $SSS$ ,  $SST$ , and  $Latitude$ . The dependent variable ( $SSS$ ) is related to two independent variables ( $SST$  and  $Latitude$ ). The linear model for  $k$  variables is:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}, \quad i = 1, 2, \dots, n \quad (3)$$

Figure 2 shows that there is a non-linear relationship between  $(SSS, Latitude)$  and  $(SSS, SST)$ . Fitting a regression plane (4) will not be representative.

$$y_i = \beta_0 + \beta_1 * Lat + \beta_2 * SST \quad (4)$$

We consider a quadratic surface to fit the data sets. The regression equation is:

$$y_i = \beta_0 + \beta_1 * Lat + \beta_2 * SST + \beta_3 * Lat^2 + \beta_4 * Lat * SST + \beta_5 * SST^2 \quad (5)$$

The Curve Fitting Toolbox in MATLAB™ provides the most widely used techniques for fitting curves and surfaces to data, using linear and non-linear regression, interpolation, and smoothing. To find the *surface* that best fits the data, we use the most popular method to estimate the regression coefficients in (5), namely the *Least Squares Method*.

### 1.2.1. Least Squares Method

This method is one of the oldest techniques of modern statistics. It was developed in the late 18<sup>th</sup> century and early 19<sup>th</sup>. It allows to compare experimental data, generally flawed, to a mathematical model supposed to describe this data. This method allows to minimize the experimental errors impact, by adding information in the measuring process [3]. In matrix form, the model is given by the formula:

$$y = f(X, \beta) + \epsilon \quad (6)$$

Where:

$y$  is an  $n$ -by-1 vector of responses.  
 $f$  is a function of  $\beta$  and  $X$ .  
 $\beta$  is a  $m$ -by-1 vector of coefficients.  
 $X$  is the  $n$ -by- $m$  design matrix for the model.  
 $\epsilon$  is an  $n$ -by-1 vector of errors.  
 $n$  is the number of data points included in fit.

The  $m$  regression coefficients are held in the vector  $\beta$ . To obtain this coefficient estimates, this method minimizes the summed square of residuals. The residual for the  $i^{th}$  data point  $r_i$  is defined as the deviation of the points from the regression line, in another term, it represent the difference between the observed response value  $y_i$  and the fitted response value  $\hat{y}_i$  [9]:

$$\begin{aligned}
 \text{Residual } r_i &= y_i - \hat{y}_i \\
 y_i &\text{ observed value} \\
 \hat{y}_i &\text{ fitted response} \\
 S &= \sum_i^n r_i^2 = \sum_i^n (y_i - \hat{y}_i)^2
 \end{aligned}$$

In Figure 4, we show some data points and a regression line, the plotted squares represent the squares of residuals. If we change the slope and/or the intercept of this line, the sizes of squares would be changed. The *Least Squares Method* find the line that minimizes the total area of these squares [6].

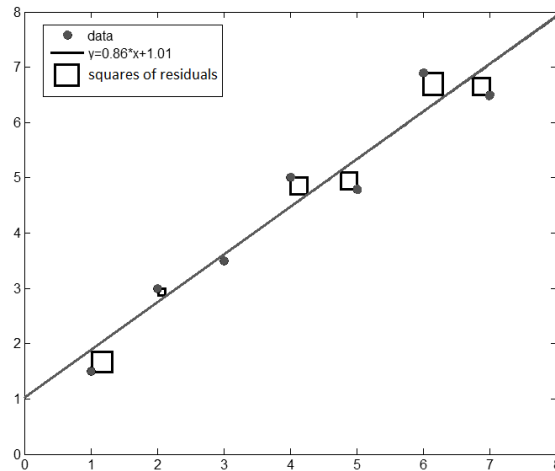


FIGURE 4. Representation of linear regression with the residuals squares.

We sought to compute a non-linear relationship expressing  $SSS$  as a function of Latitude ( $Lat$ ) and  $SST$ . We performed the multi-variable non-linear regression function which gives less errors :

$$SSS = f(Lat, SST) = \beta_0 + \beta_1 * Lat + \beta_2 * SST + \beta_3 * Lat^2 + \beta_4 * Lat * SST + \beta_5 * SST^2$$

with

First experiment:

$$\begin{aligned}
 \beta_0 &= -5.855 \\
 \beta_1 &= +0.5229 \\
 \beta_2 &= +3.061 \\
 \beta_3 &= +0.002243 \\
 \beta_4 &= -0.01997 \\
 \beta_5 &= -0.05623
 \end{aligned}$$

Second experiment:

$$\begin{aligned}\beta_0 &= -3.925 \\ \beta_1 &= +0.5998 \\ \beta_2 &= +2.828 \\ \beta_3 &= +0.001887 \\ \beta_4 &= -0.02251 \\ \beta_5 &= -0.05022\end{aligned}$$

### 1.3. MLP (Multi-Layer Perceptron) Model

The Artificial Neural Networks (ANNs) introduced in the 1960s, are based on the human nervous system functioning to design processing information machines [7]. It is composed of two or more layers. Each one contains a set of neurons (e.g layers in Figure 5). The connections between the layers are associated with weights. There are many types of Neural Network (NN). They differ in structure, and the learning algorithm. In this work we proposed to apply the MLP (Multi-Layer Perceptron) that is a supervised method [7]. It requires a desired output in order to learn using a back-propagation algorithm, for more details see [7]. Model that maps the input to the output is then created. The goal is to produce an output when the desired output is unknown. Figure 5 shows the structure of MLP.

The standard MLP network that is used for function fitting in the Neural Network Toolbox of MATLAB™, is a two-layer feedforward network, with a sigmoid transfer function (Equation (8)) in the hidden layer and a linear transfer function (Equation (9)) in the output layer [10]. Before learning starts, the default dataset division is made as follow:

- 70% for training,
- 15% to validate that the network is generalizing,
- 15% used as a completely independent test of network generalization.

$$net_j = \sum w_{ij} * I_i \quad (7)$$

$$f(net_j) = \frac{1}{1 - e^{-net_j}} \quad (8)$$

$$O_j = f(net_j) \quad (9)$$

$I$  is the input vector which contains *Latitude* and *SST*.  $O$  is the output vector that represents *SSS*.  $w_{ij}$  is the weight from input  $i$  to neuron  $j$ .

In the algorithm, we propose, initially the hidden layer contained one neuron, at each new iteration a new NN is created, on which one hidden neuron is added. Finally there are as many networks as iterations. In order to find the NN that gives a best performance, we compute MSE (Equation (12)) for each one. In the first experiment, the NN with 48 hidden neurons was the best one, while it is the NN with 83 hidden neurons in the second experiment.

### 1.4. Models evaluation

Willmot and Matsura (2005) have suggested that RMSE (Root Mean Square Error) is an inappropriate and misinterpreted measure of average error, and that MAE (Mean Absolute Error) is a more natural measure [2]. Chai and Draxler (2014) have demonstrated that the RMSE is not ambiguous in its meaning, and is more appropriate to use than the MAE when model errors follow a normal distribution [8].

In model evaluation studies, there are three rules which are regularly employed to measure the average magnitude of the error.



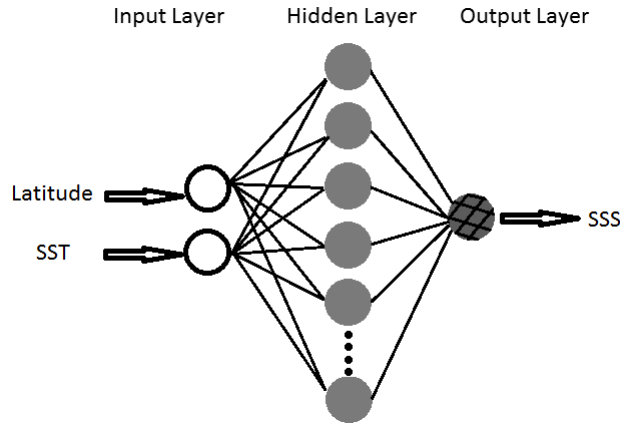


FIGURE 5. MLP structures obtained after data training: 2 input neurons correspond to *Latitude* and *SST*, 1 hidden layer (with 48 neurons in first experiment and 83 neurons in the second one), and 1 output neuron corresponds to estimated *SSS*.

- RMSE: As shown in Equation (10), using this quadratic scoring rule, since the errors are squared before they are averaged, the RMSE gives a relatively high weight to large errors. Thus, this rule is most useful when large errors are particularly undesirable [8].

$$RMSE = \sqrt{\frac{\sum_i (x_i - y_i)^2}{n}} \quad (10)$$

- MAE: This is a linear score which means that all individual differences are weighted equally in the average [5].

$$MAE = \frac{\sum_i |x_i - y_i|}{n} \quad (11)$$

- MSE: We look that MSE (Mean Square Error) is analogue to RMSE [8].

$$MSE = \frac{\sum_i (x_i - y_i)^2}{n} \quad (12)$$

- MBE: (Mean Bias Error) Here, the signs of the errors are not removed.

$$MBE = \frac{\sum_i (x_i - y_i)}{n} \quad (13)$$

Where  $n$  is the number of data,  $y$  is the output of the model and  $x$  is the target.

## 2. RESULTS AND DISCUSSION

To evaluate the goodness of fit test, in both experiments, for each technique, *SS* was calculated at the *in situ SST* and compared with the *in situ SSS* data. In the first experiment, Figure 6.a and Figure 6.b show fitted *SSS* by the MNR and the MLP, respectively, as function of *in situ SSS*. Figure 7.a and Figure 7.b show results of the second experiment. We note that in all figures which represent estimated *SSS* as function of *in situ SSS*, two lines are represented, the theoretical line " $y = x$ ", and the regression line " $y = a.x + b$ " that represents the relationship between this two parameters. The MBE, the RMSE, and the MAE of models are reported in Table 3. In building models, for both experiments, MLP gives smaller errors than MNR.

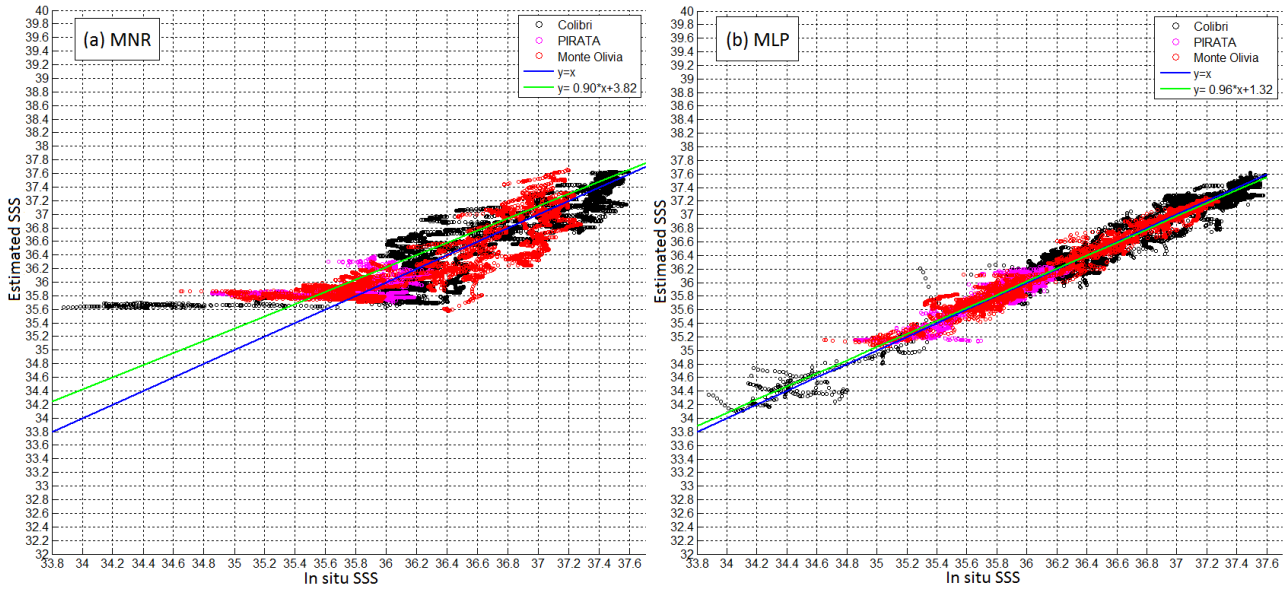


FIGURE 6. Experiment I (for *extrapolation*): evaluation of MNR (a) and MLP (b) using the *in situ* data; *SSS* estimated as function of *in situ* *SST*.

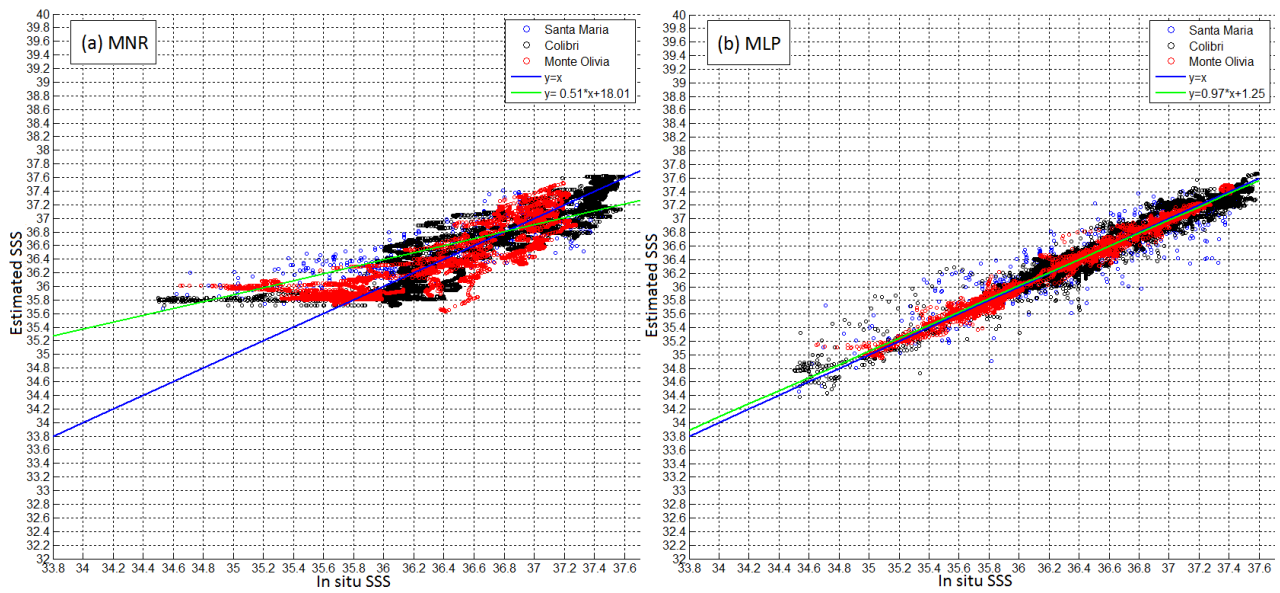


FIGURE 7. Experiment II (for *interpolation*): evaluation of MNR (a) and MLP (b) using the *in situ* data; *SSS* estimated as function of *in situ* *SST*.

In order to assess which approach would best estimate *SSS*, we used the remaining data subsets (Table 2). To validate the two models, we used daily averaged satellite data (from MODISAqua mission) at the dates of the *in situ* data used to build the models. Figure 8 shows an example of one day of satellite *SST* as well as the measured *in situ* data of *SST*. To extract data we need, we sought satellite points corresponding to *in situ*

Experiments	Models	MBE	RMSE	MAE
I	a MNR	$4.2598 \times 10^{-15}$	0.3051	0.2306
	b MLP	$-9.3441 \times 10^{-4}$	0.1206	0.0900
II	a MNR	$-4.9171 \times 10^{-13}$	0.2844	0.2144
	b MLP	-0.0290	0.1894	0.1177

TABLE 3. Goodness of fit for each model using *in situ* data : (I) first experiment, and (II) second experiment.

points and containing *SST* information. Then, we use points that include both *in situ SST* and satellite *SST* (Figure 8 in red). Several parameters, such as cloud cover (Figure 8 in white) prevent satellite observations, hence decrease of data point number that we can use from 18300 (*in situ* data) to 9139 (with satellite data) in the first experiment, and from 15668 to 8247 in the second one (Figure 8 in blue represents points with only *in situ SST* data). We used both polynomial functions and neural networks to estimate *SSS* from these satellite data. Estimated *SSS* is then compared with *in situ SSS* used in building the models.

After this first step of validation, the models are tested with data points which are situated in a different geographical area. In the first experiment, we use data from the Santa Maria cruise. In the second experiment, we use data from the PIRATA cruise. Because of the use of a dataset which is in another geographical area, errors have increased compared to the validation step. RMSE of satellite *SST* (Table 4), show that in datasets validation, there are some satellite points with a large error in *SST*, which leads to large errors with the MLP method.

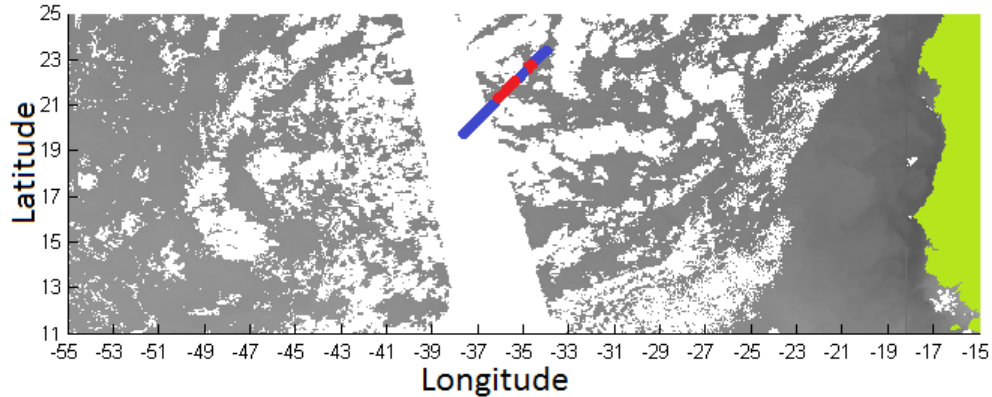


FIGURE 8. One day averaged Satellite SST data (gray scale); *in situ* SST data (blue); *in situ* SST data and satellite SST data (red).

Experiment	MBE	RMSE	MAE
I	0.4464	0.7477	0.5271
II	0.4289	0.7533	0.5150

TABLE 4. Satellite SST residuals (compared to *in situ* SST) in both experiments.

### 2.1. Estimation of SSS using MNR

Results of the first step of validation are reported in Table 5.I.1.a for the first experiment, and Table 5.II.3.a for the second one. Comparing these results with those of Table 3, we note that in the two experiments, errors have improved. This is because the validation dataset is in the same geographical area than the original data, and the number of data points is smaller than that of construction model. Table 5.I.2.a contains results of the *extrapolation*. In Figure 9.a, we see that the MNR agree relatively well on estimating salinity. Table 5.II.4.a shows that *interpolation* results are better than those of *extrapolation*. However, Figure 10.a shows a large margin between the theoretical and the regression lines. If we compare this fit with the MLP one (Figure 10.b), we see that the MLP fit is better, which is contradictory with the values of errors (Table 5.II.4.a and Table 5.II.4.b). For this reason, we computed the histogram of errors. As shown in Figure 11, in *interpolation* with MNR we have 67.98% of errors between  $[0 - 0.2]$ , and 32.02% between  $]0.2 - 1.2]$ . Unlike *interpolation* with MLP, we have 35.18% in  $[0 - 0.2]$ , and 64.82% in  $]0.2 - 1.2]$ . We note that in all MNR regression (construction of models and tests), the model tends towards to overestimate *SSS* at the lowest *in situ SSS*.

### 2.2. Estimation of SSS using MLP

Table 5.I.1.b and Table 5.II.3.b show results of the first validation step. Unlike MNR, MLP is sensitive to large errors. For both experiments, we note that due to the large errors in satellite *SST*, validation errors are large compared to errors in building model. The large errors in *extrapolation* results (Table 5.I.2.b, Figure 9.b), show that in spite of small errors in building model (Table 3.II.b), and in spite of an excellent regression (Figure 6.b), it is not adapted at all for *extrapolation*. However, for an *interpolation*, it can provide reasonable results (see Table 5.II.4.b and Figure 10.b).

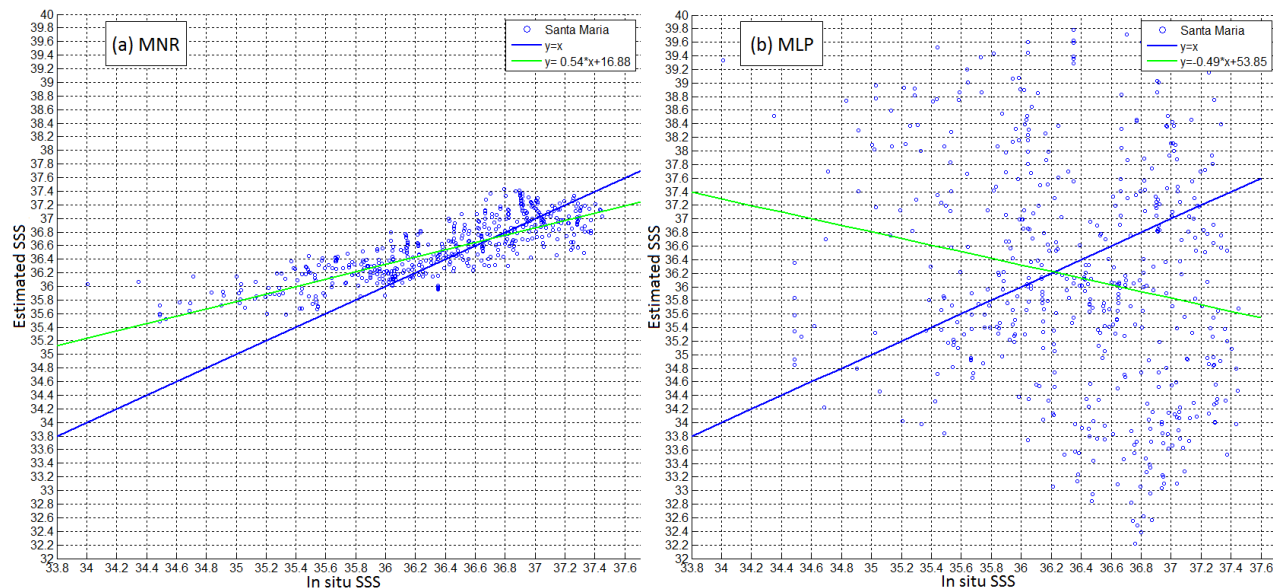


FIGURE 9. *Extrapolation* results with Santa Maria data cruises in experiment I; (a): MNR, (b): MLP.

This work shows that the MLP method needs to be significantly improved to be used to interpolate and extrapolate *SSS* from *SST*. A neural network is more complicated than a polynomial function. Thus for obtaining a robust neural network, many parameters can be changed, especially the structure, for example change the number of hidden layers, the number of hidden neurones, connections between neurones... The

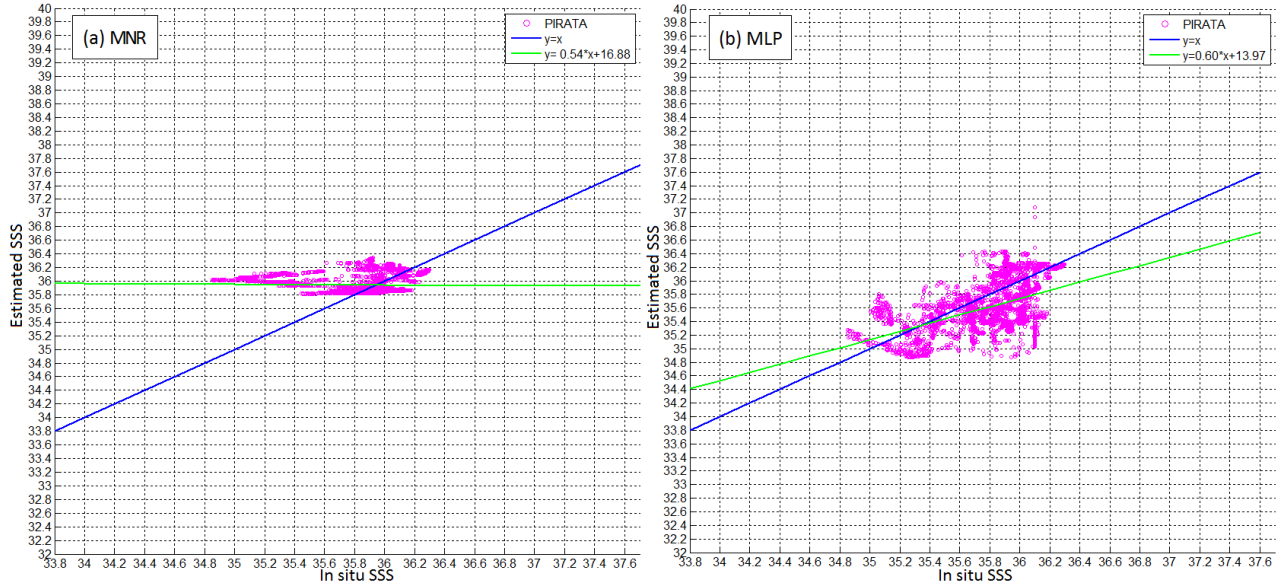


FIGURE 10. *Interpolation* results with PIRATA data cruise in experiment II; (a): MNR, (b): MLP.

Experiments	Datasets		Models	MBE	RMSE	MAE	
I	1	Satellite data (MODISAqua)	a	MNR	0.0901	0.2522	0.1965
			b	MLP	-0.1785	0.5810	0.2700
	2	<i>In situ</i> data (Santa Maria cruises)	a	MNR	-0.1871	0.4336	0.3316
			b	MLP	0.2247	1.9608	1.5621
II	3	Satellite data (MODISAqua)	a	MNR	0.0972	0.2674	0.2032
			b	MLP	0.3389	2.3112	0.6521
	4	<i>In situ</i> data (PIRATA cruise)	a	MNR	-0.1588	0.3614	0.2473
			b	MLP	-0.1781	0.3775	0.3081

TABLE 5. Estimated residuals *SSS* with the four test datasets (1,2,3,4), and the two techniques (a: MNR, b: MLP).

training algorithm could also be changed.

This study further indicates that the popular Non-linear Regression with Least Squares method provides better results than MLP in this area and for this parameters (*SST*, *SSS*). The MNR equation:

$$SSS = f(Lat, SST) = \beta_0 + \beta_1 * Lat + \beta_2 * SST + \beta_3 * Lat^2 + \beta_4 * Lat * SST + \beta_5 * SST^2$$

is an admissible solution whether it is for *extrapolation* with  $RMSE = 0.3$ , within  $23^\circ\text{S}$ - $61^\circ\text{W}$  to  $23^\circ\text{N}$ - $20^\circ\text{W}$ :

$$\begin{aligned} \beta_0 &= -5.855 \\ \beta_1 &= +0.5229 \\ \beta_2 &= +3.061 \\ \beta_3 &= +0.002243 \\ \beta_4 &= -0.01997 \\ \beta_5 &= -0.05623 \end{aligned}$$

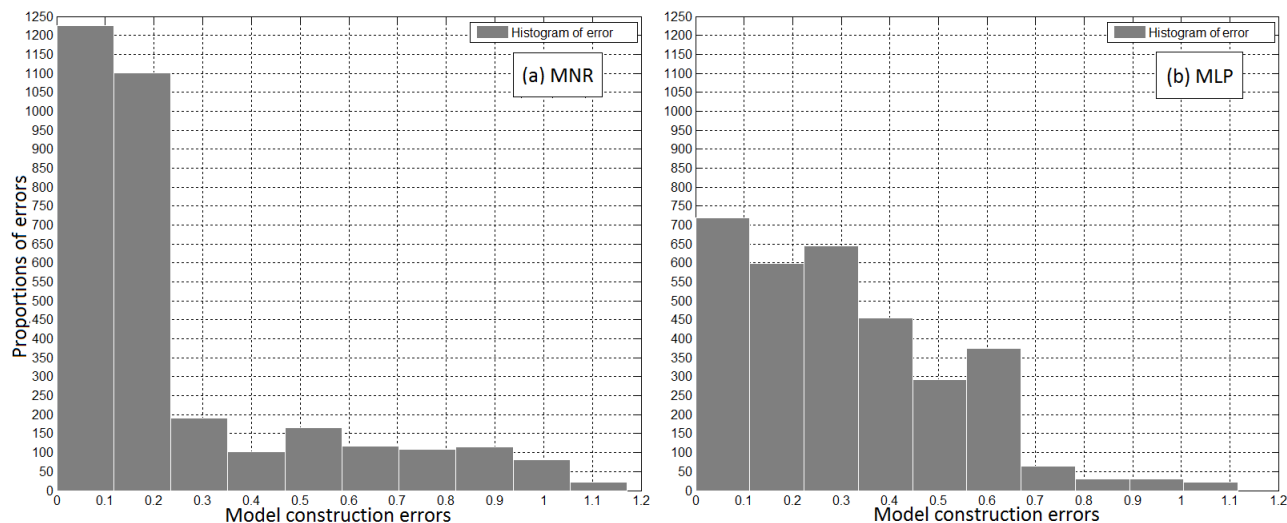


FIGURE 11. Error histograms in *interpolation*, (a): MNR, (b): MLP.

Or for *interpolation* with  $RMSE = 0.2$ , within  $23^{\circ}\text{S}$ - $55^{\circ}\text{W}$  to  $23^{\circ}\text{N}$ - $20^{\circ}\text{W}$ :

$$\begin{aligned}\beta_0 &= -3.925 \\ \beta_1 &= +0.5998 \\ \beta_2 &= +2.828 \\ \beta_3 &= +0.001887 \\ \beta_4 &= -0.02251 \\ \beta_5 &= -0.05022\end{aligned}$$

At present, additional datasets are required to better assess the model. We propose to perform a model for each month, or for each season of the year (Fall: September-October-November, Winter: December-January-February, Spring: March-April-May, Summer: June-July-August) depending upon the availability of *in situ* data. A threshold could be used to exclude the wrong satellite *SST*. Overall, this study opens the route of mapping *SSS* at the same temporal and spatial resolution as *SST* from satellites.

#### ACKNOWLEDGEMENTS

We are grateful to the MESRS (Ministère Algérien de l'Enseignement Supérieur et de la Recherche Scientifique) for funding H.MOUSSA's PhD.

#### REFERENCES

- [1] C. Goyet, R. Adams, and G. Eiseheid, *Observations of the  $CO_2$  system properties in the tropical Atlantic Ocean*. MARINE CHEMISTRY, Vol. 60, P. 49-61, February 1998.
- [2] C.J. Willmott and K. Matsuura, *Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance*, CLIMATE RESEARCH, Vol. 30, P. 79-82, 2005.
- [3] H. Abdi, *The Method of Least Squares*, In N.J. Salkind, D.M. Dougherty, & B. Frey (Eds.): *Encyclopedia of Research Design*, P. 705-708, 2010.
- [4] N. Lefèvre, D.F. Urbano, F. Gallois, and D. Diverrès, *Impact of physical processes on the seasonal distribution of the fugacity of  $CO_2$  in the western tropical Atlantic*, JOURNAL OF GEOPHYSICAL RESEARCH : OCEANS, Vol. 119, P. 646-663, February 2014.

- [5] S.H. Brown, *Multiple Linear Regression Analysis: A Matrix Approach with MATLAB*, ALABAMA JOURNAL OF MATHEMATICS, 2009.
- [6] S.L. Baker, *Simple Regression Theory I*, Lecture notes, 2010.
- [7] S. Seung, *Multilayer perceptrons and backpropagation learning*, Lecture 4, September 2002.
- [8] T. Chai and R.R. Draxler, *Root mean square error (RMSE) or mean absolute error (MSE)?*, GEOSCIENTIFIC MODEL DEVELOPMENT, Vol. 7, P. 1525-1534, February 2014.
- [9] The MathWorks, *Curve Fitting Toolbox User's Guide*, 2014.
- [10] The MathWorks, *Neural Network Toolbox User's Guide*, 2014.