

Moussa, H.; Benallal, M. A.; Goyet, C.; Lefèvre, N.; El Jai, M. C.; Guglielmi, V.; Touratier, F.

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. (English. French summary) [Zbl 1338.86015](#)
ESAIM, Proc. Surv. 49, 65-77 (2015).

Summary: 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*.

MSC:

- 86A32 Geostatistics
- 86A05 Hydrology, hydrography, oceanography

Software:

[cftool](#); [ITE](#); [Neural Network Toolbox](#)

Full Text: [DOI](#)

References:

[1] : Residual $r_i = y_i - \hat{y}_i$ observed value \hat{y}_i fitted response $r_i^2 = (y_i - \hat{y}_i)^2$ $S = \sum r_i^2$ 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]. ESAIM: PROCEEDINGS AND SURVEYS 71 Second experiment: $\beta_0 = -3.925$ $\beta_1 = +0.5998$ $\beta_2 = +2.828$ $\beta_3 = +0.001887$ $\beta_4 = -0.02251$ $\beta_5 = -0.05022$ 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 MATLABtm, 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.

 $net_j = w_{ij} \cdot I_i(7)$ 72 ESAIM: PROCEEDINGS AND SURVEYS 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]. $RMSE = \sqrt{\frac{1}{n} \sum (x_i - y_i)^2}$
- MAE: This is a linear score which means that all individual differences are weighted equally in the average [5]. $MAE = \frac{1}{n} \sum |x_i - y_i|$

 ESAIM: PROCEEDINGS AND SURVEYS 73 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. 74 ESAIM: PROCEEDINGS AND SURVEYS Experiments Models MBERMSE MAE IaMNR 4.2598 x 10⁻¹⁵ 0.30510.2306bMLP-9.3441 x 10⁻⁴⁰ 0.12060.0900 IIaMNR-4.9171 x 10⁻¹³⁰ 0.28440.2144bMLP-0.02900.18940.1177 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. ESAIM: PROCEEDINGS AND SURVEYS 75 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). 76ESAIM: PROCEEDINGS AND SURVEYS Figure 10. Interpolation results with PIRATA data cruise in experiment II; (a): MNR, (b): MLP. Experiments Datasets Models MBERMSE MAE IbMLP-0.17850.58100.2700aMNR-0.18710.43360.3316 1Satellite data (MODISAqua)aMNR0.09010.25220.1965 2In situ data (Santa Maria cruises)bMLP0.22471.96081.5621 IIbMLP0.33892.31120.6521aMNR-0.15880.36140.2473 3Satellite data (MODISAqua)aMNR0.09720.26740.2032 4In situ data (PIRATA cruise)bMLP-0.17810.37750.3081 ESAIM: PROCEEDINGS AND SURVEYS 77 Figure 11. Error histograms in interpolation, (a): MNR, (b): MLP. Or for interpolation with RMSE = 0.2, within 23oS-55oW to 23oN-20oW: $\{\beta\}_0 = -3.925$ $\{\beta\}_1 = +0.5998$ $\{\beta\}_2 = +2.828$ $\{\beta\}_3 = +0.001887$ $\{\beta\}_4 = -0.02251$ $\{\beta\}_5 = -0.05022$ At present, additional datasets are required to better assess the model. We propose to perform a model for 78ESAIM: PROCEEDINGS AND SURVEYS

- [2] S.H. Brown, Multiple Linear Regression Analysis: A Matrix Approach with MATLAB, ALABAMA JOURNAL OF MATHEMATICS, 2009.
- [3] S.L. Baker, Simple Regression Theory I, Lecture notes, 2010.
- [4] S. Seung, Multilayer perceptrons and backpropagation learning, Lecture 4, September 2002.
- [5] 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.
- [6] The MathWorks, Curve Fitting Toolbox User's Guide, 2014.
- [7] The MathWorks, Neural Network Toolbox User's Guide, 2014.

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.