



Original software publication

mStatGraph: Exploration and statistical treatment software to process, compute and validate oceanographic data

H.L. Varona^{a,b,*}, C. Noriega^{a,b}, J. Araujo^c, S.M.A. Lira^{d,e}, M. Araujo^{a,b,c}, F. Hernandez^f^a Laboratory of Physical, Coastal and Estuarine Oceanography (LOFEC), Oceanography Department (DOCEAN), Federal University of Pernambuco, Recife, PE, Brazil^b Center for Risk Analysis and Environmental Modeling (CEERMA), Federal University of Pernambuco, Recife, PE, Brazil^c Brazilian Research network on Global Climate Change (Rede CLIMA), São José dos Campos, SP, Brazil^d Laboratory of Zooplankton, Department of Oceanography, DOCEAN, Federal University of Pernambuco, Recife, PE, Brazil^e Laboratory of Plankton Ecology, Department of Biology, Federal Rural University of Pernambuco, Recife, PE, Brazil^f Laboratoire d'Etudes en Géophysique et Océanographie Spatiales (LEGOS), 18 avenue Edouard Belin, 31401, Toulouse Cedex 9, France

ARTICLE INFO

Keywords:

Statistical tests
Physical oceanography
Biogeochemistry
Data comparison
River carbonate system

ABSTRACT

mStatGraph is a software designed to facilitate geoscientists to study the environment in a fast and efficient way. This software integrates statistical tests, parameter trend studies, comparison of numerically simulated data with observed data, computation and plotting of oceanographic and meteorological parameters, and computation of the carbonate system in rivers. It also includes plotting of maps of the spatial distribution of collected data, plotting of current and wind roses, and Hovmöller plotting. All these functions integrated in a single tool allows to streamline data processing and in turn improve the knowledge of the environment.

Code metadata

Current code version	v1.0
Permanent link to code/repository used for this code version	https://github.com/SoftwareImpacts/SIMPAC-2023-321
Permanent link to reproducible capsule	https://codeocean.com/capsule/5551220/tree/v1
Legal code license	MIT License
Code versioning system used	Zenodo
Software code languages, tools and services used	MATLAB
Compilation requirements, operating environments and dependencies	MATLAB 2021
If available, link to developer documentation/manual	https://zenodo.org/record/8152683/files/mStatGraph%20UserManual_v1.7.pdf?download=1
Support email for questions	humberto.varona@ufpe.br

1. Introduction

At the core of any scientific investigation lies basic statistics, which provide a foundation for understanding data distributions, central tendencies, and variability. Measures such as mean, median, mode, and standard deviation are essential in describing and summarizing oceanographic data. They offer insights into the central values and spread

of observations, facilitating the identification of outliers or anomalies. Comparing data sets is a critical step in oceanographic research, allowing scientists to identify significant differences between samples or study areas. Statistical tests like t-tests and Mann–Whitney U-tests are commonly employed for this purpose [1–4]. The t-test assesses if the means of two populations significantly differ, whereas the Mann–Whitney U-test compares distributions when assumptions for the t-test

DOI of original article: <https://doi.org/10.1016/j.rsma.2023.103071>.

The code (and data) in this article has been certified as Reproducible by Code Ocean: (<https://codeocean.com/>). More information on the Reproducibility Badge Initiative is available at <https://www.elsevier.com/physical-sciences-and-engineering/computer-science/journals>.

* Corresponding author at: Laboratory of Physical, Coastal and Estuarine Oceanography (LOFEC), Oceanography Department (DOCEAN), Federal University of Pernambuco, Recife, PE, Brazil.

E-mail addresses: humberto.varona@ufpe.br (H.L. Varona), carlos.delnor@gmail.com (C. Noriega), juliamdearaujo@gmail.com (J. Araujo), simonealira@gmail.com (S.M.A. Lira), moa.ufpe@gmail.com (M. Araujo), fabrice.hernandez@ird.fr (F. Hernandez).

<https://doi.org/10.1016/j.simpa.2023.100571>

Received 30 June 2023; Received in revised form 18 August 2023; Accepted 19 August 2023

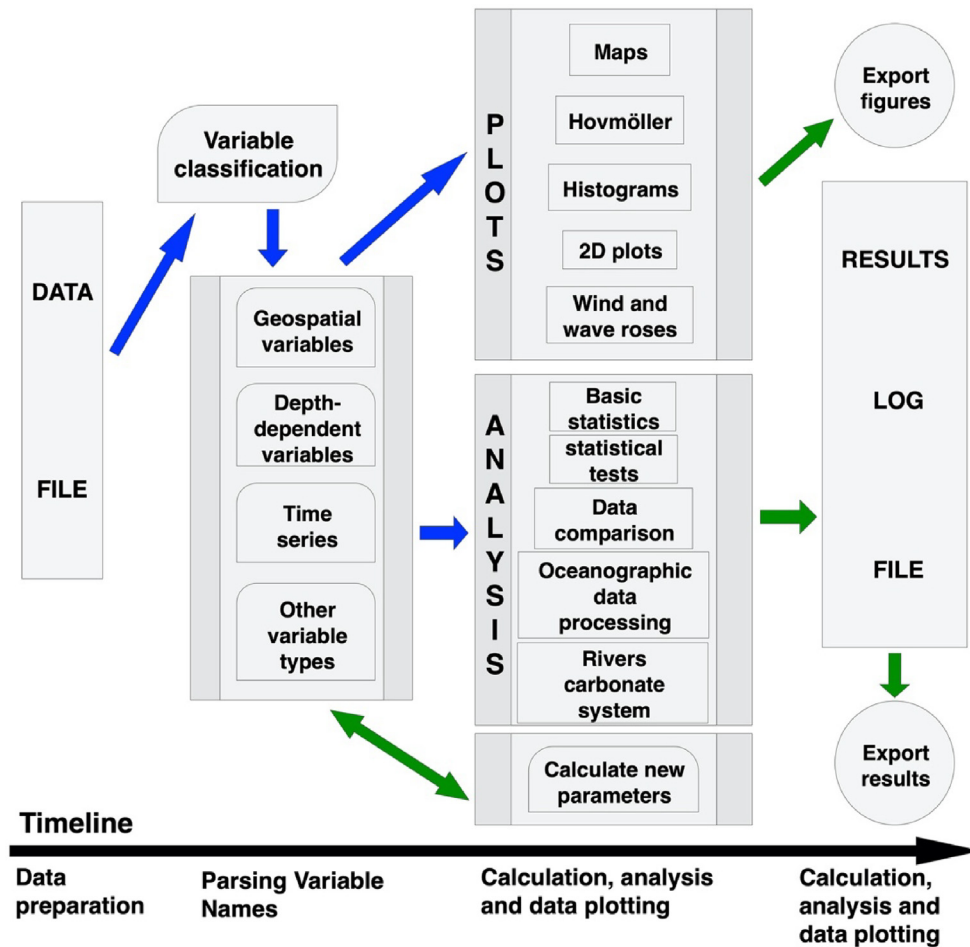


Fig. 1. Logical steps to perform the calculation and statistical graphs and ocean parameters with mStatGraph.

are not met [5]. These techniques enable researchers to evaluate the significance of differences and make informed comparisons.

Assessing the quality and accuracy of collected data is crucial in oceanography. Bias, a measure of systematic error, quantifies the deviation between measured values and a reference standard [6–8]. Root Mean Square Error (RMSE) provides a measure of the differences between observed and predicted values [7,8], offering insights into the overall accuracy of models or measurements. The scatter index represents the spread of data points around the regression line, helping evaluate the precision and reliability of measurements [9]. Oceanographic research often involves analyzing time series data to uncover long-term trends or variability [10,11]. Statistical techniques, such as linear regression, allow scientists to identify and quantify trends, providing valuable information on processes like climate change or oceanic circulation patterns. Trend analysis provides a basis for predicting future conditions and understanding the dynamics of the marine environment.

Multivariate analysis plays a crucial role in exploring complex relationships among multiple variables in oceanographic studies. The Kruskal–Wallis test [5], a non-parametric alternative to ANOVA, examines whether there are significant differences between multiple groups. It is particularly useful when the assumptions of normality and equal variances are not met. ANOVA (Analysis of Variance) assesses if the means of multiple groups are significantly different, providing insights into factors influencing oceanographic phenomena. Dunn and Tukey tests are post-hoc tests frequently used after significant differences are found in ANOVA [12]. These tests enable scientists to pinpoint

which group means differ significantly from others, assisting in the identification of specific sources of variation.

2. Software details

The mStatGraph software is a MATLAB 2021b package designed to assist physical oceanographers, biogeochemists and meteorologists in their research. It provides an integrated set of functions, including calculations of oceanographic parameters such as mixed layer depth (MLD), isothermal layer depth (ILD), and barrier layer thickness (BLT). The software also provides graphics commonly used by oceanographers, such as Hovmöller diagrams, CTD profiles (conductivity, temperature and depth), spatial distribution maps, wind roses, currents and waves. It can handle input data from various sources, including observed data (in situ, radar, satellite) or numerical model output, and automatically sorts variables to trigger relevant options. The software can perform statistical analyses, data comparisons, and generate visualizations in the form of graphs and 2D maps, which can be exported in various formats and saved for later review (Fig. 1). In addition, mStatGraph includes an equation solver, which allows users to work with custom equations involving loaded variables and computed data.

The software functionalities are adapted to the input data; if longitude and latitude are present, map options are activated and, if time variables are detected, time series plots are available. Similarly, if the data contain depth, potential temperature and salinity, calculation options for MLD, ILD and BLT are enabled, along with the

CTD plot. Other options, such as ocean current, wind and wave roses, are activated depending on the presence of relevant components in the input data. Statistical tests and data comparison options are automatically activated according to the columns selected in the data table. The software efficiently processes oceanographic data, performs river carbonate system computations if appropriate data are present, and records all results in a “Summary” tab. Users can also customize plotting and mapping options based on variable classifications, adding features such as bathymetry, boundaries, and customizing collection point representations.

3. Software algorithms

The mStatGraph software automatically computes basic statistics for all variables after loading the data file. Understanding these basic statistical measures is essential for describing and analyzing data in various fields of oceanography. They provide information about the central tendency, variability, dispersion, shape, and distribution of data, allowing researchers to make informed interpretations and draw meaningful conclusions from their analyses. A complete list of comparative statistics, dispersal indices, and physical oceanographic and biogeochemical parameters is available in the Supplementary Material (MS).

4. Related works

Through mStatGraph, data analysis has been performed in the articles [11,13], which has facilitated the processing of the data used. Environmental problems caused by the accumulation of nitrogen and phosphorus in river water as it flows into the ocean were analyzed using mStatGraph software [11]. It was also used in the validation of the simulated data with the WAVEWATCH 3 model, comparing these with data observed from in situ instruments. Being able to reproduce the extreme waves that occurred in March 2016 on the coasts of Rio Grande do Sul, Brazil [13].

5. Impact overview

The generation of maps associated with the numerical results using the mStatGraph software in [11] facilitated the interpretation of the results of the hydro-chemical parameters observed in 60 tropical rivers. The data series of chemical parameters such as nitrogen and phosphorus were analyzed temporally and spatially; mainly trying to find statistical patterns through internal analysis, such as the relationship between natural and anthropogenic factors. Correlations between these factors were analyzed using Pearson and Spearman tests. Trends associated with population growth and temporal variations in the concentrations of chemical elements such as nitrogen and phosphorus could be tested using the Mann–Kendall test. On the other hand, mStatGraph allowed validations of three numerical simulations of wind, wind wave and swell behavior during the presence of a strong low-pressure gradient in Rio Grande do Sul, Brazil [13]; these validations were performed using t-test, bias, root mean square error and scatter index in time series comparing the simulated data with data observed in situ.

The mStatGraph software is designed to facilitate the work of geoscientists. An example of this is represented in Fig. S1, which shows maps made with mStatGraph of the spatial distribution of physical and biogeochemical parameters collected. The pH values oscillated between 5 and 8.5 units; with lower values in the southern region of the study area (Fig. S1a). Water temperature variations (Fig. S1b) recorded during the period (23–29 °C) showed typical values for tropical waters (mean = 27 °C). According to the environmental resolution (CONAMA 305; 2005) dissolved oxygen (DO) values should not be less than 5 mg l⁻¹. The values obtained here showed DO concentrations above this limit except in some points (~10 points) with values lower than 5 mg l⁻¹ (Fig. S1c). According to the observations of the period, the TOC

varied between 2.7–70.0 mg l⁻¹, with an accumulation of values in the range of 10–20 mg l⁻¹ and an average value of 13.0 ± 10 mg l⁻¹ (Fig. S1d). This value is typical for tropical regions. High TOC values (20–50 mg l⁻¹) can be found in urban areas where the amount of organic matter included in urban sewage is greater than in natural waters.

Another example would be a trend analysis (Fig. S2), in the heavy metals group, Fe did not show significant trend (Kendall tau test; $p = 0.85$; for $\alpha = 0.05$) (Fig. S2a), while, Mn showed a significant positive trend during the study period (Kendall tau test; $p = 0.0028$; $\alpha = 0.05$). Two peaks in the concentration of this element were recorded; however, most values were below 0.05 mg Mn l⁻¹ (Fig. S2b). Fe is not considered in the list of heavy metals in Brazilian brackish water legislation. According to the Environmental Protection Agency (EPA), the maximum limit in natural waters is 300 mg Fe m⁻³ (0.3 mg Fe l⁻¹). The Brazilian Ministry of Health (Ordinance 518/2004) considers a value of 0.3 mg Fe l⁻¹ for water intended for human consumption, while Resolution 430/2011 of the Brazilian Ministry of Environment for fluent releases considers a limit value of 15 mg Fe l⁻¹. The results obtained for Fe indicated an average value of 0.37 mg Fe l⁻¹. This value is higher than that recommended by the EPA and the Brazilian Ministry of Health. The positive trend observed for these two parameters indicates an increase in the slope of the trend line from 2002 onwards. Although Fe and Mn have quite similar chemical behavior, the extent to which Fe and Mn dissolve in water depends on the amount of DO in the water and, to a lesser extent, on the acidity, that is, pH. The results of Fe and Mn showed positive trends (Fig. S2a and Fig. S2b); however, the average values were not above the limits required by environmental legislation.

6. Future development of the software

In the future, two types of multivariate data analysis will be implemented in mStatGraph: the Principal Components Analysis (PCA) and Canonical Correlation Analysis (CCA). Normalized Bias, normalized root mean square error, Percentage Error (Equation S16), and Model Performance Index (Equation S17) will also be implemented to compare numerically simulated data with observed data from different sources (these parameters are detailed in the MS).

7. Final remarks

mStatGraph has versatile applications both within and beyond the academic realm, making it a practical and efficient tool for conducting statistical tests and computations of oceanographic and meteorological parameters. Its wide usability enables geoscientists to easily access and analyze crucial data for studying and monitoring the environment.

CRediT authorship contribution statement

H.L. Varona: Conceptualization, Methodology, Software, Validation, Formal analysis, Visualization, Writing – original draft, Writing – review & editing. **C. Noriega:** Visualization, Writing – original draft, Writing – review & editing. **J. Araujo:** Visualization, Writing – original draft, Writing – review & editing. **S.M.A. Lira:** Visualization, Writing – original draft, Writing – review & editing. **M. Araujo:** Conceptualization, Methodology, Writing – review & editing, Supervision, Funding acquisition. **F. Hernandez:** Methodology, Visualization, Software.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Humberto L. Varona and S. M. A. Lira report financial support was provided by European Union.

Acknowledgments

HLV and SMAL acknowledge the TRIATLAS project, which has received funding from the European Union's Horizon 2020 Research and Innovation Program under grant agreement no. 817578. CN acknowledges the Fade/UFPE agreement n° 04/2018 (5850.0108524.18.9) through grant postdoctoral stage (UFPE). JA and MA acknowledge the support of the Brazilian Research Network on Global Climate Change-Rede CLIMA, Brazil (FINEP grants 01.13.0353-00), and to the International Joint Laboratory TAPIOCA (IRD-UFPE-UFRPE).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.simpa.2023.100571>.

References

- [1] P.A. Umesh, M.R. Behera, Performance evaluation of input-dissipation parameterizations in WAVEWATCH III and comparison of wave hindcast with nested WAVEWATCH III-SWAN in the Indian Seas, *Ocean Eng.* 202 (2020) 106959, <http://dx.doi.org/10.1016/j.oceaneng.2020.106959>, Elsevier BV.
- [2] T.A. Capuano, M. Araujo, M. Silva, H.L. Varona, G. Cambon, A. Koch-Larrouy, T-s and hydrodynamical structures within the deltaic regions and continental platforms adjacent to two northeastern Brazilian rivers, *Reg. Stud. Mar. Sci.* 51 (2022) 102219, <http://dx.doi.org/10.1016/j.rsma.2022.102219>, Elsevier BV.
- [3] P.D. Bromirski, D.R. Cayan, R.E. Flick, Wave spectral energy variability in the northeast Pacific, *J. Geophys. Res. Oceans* 110 (C3) (2005) <http://dx.doi.org/10.1029/2004jc002398>, American Geophysical Union (AGU).
- [4] E. Charles, D. Idier, P. Delecluse, M. Déqué, G. Le Cozannet, Climate change impact on waves in the Bay of Biscay, France, *Ocean Dyn.* 62 (6) (2012) 831–848, <http://dx.doi.org/10.1007/s10236-012-0534-8>, Springer Science and Business Media LLC.
- [5] K. Silva, H.L. Varona, M. Araujo, C. Medeiros, L. Santos, Characterization of dissolved inorganic carbon and dissolved oxygen in two impacted Amazonian lakes, *Pan-Am. J. Aquat. Sci.* 17 (2) (2022) <http://dx.doi.org/10.54451/panamjas.17.2.97>, Instituto de Ecologia y Ciencias Ambientales (IECA).
- [6] M.V.B. Silva, B. Ferreira, M. Maida, S. Queiroz, M. Silva, H.L. Varona, T.C.M. Araújo, M. Araújo, Flow-topography interactions in the western tropical Atlantic boundary off Northeast Brazil, *J. Mar. Syst.* 227 (2022) 103690, <http://dx.doi.org/10.1016/j.jmarsys.2021.103690>, Elsevier BV.
- [7] L. Zou, L. Liu, Z. Wang, Y. Chen, Evaluation Wave Condition using Wavewatch III Around Island Area in the South China Sea, Elsevier BV, 2023, <http://dx.doi.org/10.2139/ssrn.4361713>.
- [8] Y. Sheng, W. Shao, S. Li, Y. Zhang, H. Yang, J. Zuo, Evaluation of typhoon waves simulated by WaveWatch-III model in shallow waters around Zhoushan Islands, *J. Ocean Univ. China* 18 (2) (2019) 365–375, <http://dx.doi.org/10.1007/s11802-019-3829-2>, Springer Science and Business Media LLC.
- [9] W. Qiao, J. Song, H. He, F. Li, Application of different wind field models and wave boundary layer model to typhoon waves numerical simulation in WAVEWATCH III model, *Tellus A* 71 (1) (2019) 1657552, <http://dx.doi.org/10.1080/16000870.2019.1657552>, Stockholm University Press.
- [10] C. Noriega, C. Medeiros, H.L. Varona, L. Rodrigues, M. Araujo, S. Monteiro, A.X. da Silva, N.A. Pereira, E.E.S. de Lima, D.S.T. da Silva, S. de Campos Pereira, J. Araujo, M. Rollnic, Water quality in a tropical estuarine channel: current conditions, trends, and trophic status (1990–2016), *Water Air Soil Pollut.* 233 (9) (2022) <http://dx.doi.org/10.1007/s11270-022-05852-x>, Springer Science and Business Media LLC.
- [11] C. Noriega, M. Araujo, H.L. Varona, M. Costa, A.E. Calzada, C. Medeiros, A.X. Silva, L.P. Portela, D.S. Bezerra, R. Jeronimo, P.P. Barros, J. Araujo, M. Rollnic, Long-term trend of nitrogen and phosphorus transport in 12 tropical coastal watersheds in Northeast Brazil, *Química Nova* 46 (6) (2023) 616–626, <http://dx.doi.org/10.21577/0100-4042.20230058>, Sociedade Brasileira de Química (SBQ).
- [12] C.S. de Santana, S.M. de A. Lira, H.L. Varona, S. Neumann-Leitão, M. Araujo, R. Schwaborn, Amazon river plume influence on planktonic decapods in the tropical Atlantic, *J. Mar. Syst.* 212 (2020) 103428, <http://dx.doi.org/10.1016/j.jmarsys.2020.103428>, Elsevier BV.
- [13] D.M. Borges, H.L. Varona, M.F. Alonso, M. Araujo, F. Hernandez, Investigating an extreme meteo-oceanographic event in the southern Brazil from in situ observations and modeling results, *Reg. Stud. Mar. Sci.* 65 (2023) 103071, <http://dx.doi.org/10.1016/j.rsma.2023.103071>.