HAL open science

A tropical Atlantic dynamics analysis by combining machine learning and satellite data

Sabine Arnault, Sylvie Thiria, Michel Crépon, François Kaly

▶ **To cite this version:**

Sabine Arnault, Sylvie Thiria, Michel Crépon, François Kaly. A tropical Atlantic dynamics analysis by combining machine learning and satellite data. Advances in Space Research, 2021, 68 (2), pp.467-486. 10.1016/j.asr.2020.09.044 . hal-03336066

HAL Id: hal-03336066

https://hal.science/hal-03336066v1

Submitted on 13 Jun 2023

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.

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Version of Record: https://www.sciencedirect.com/science/article/pii/S0273117720307092
Manuscript_2203362de781cc611509b8b87227cd81

A Tropical Atlantic Dynamics Analysis by Combining Machine Learning and Satellite Data

Sabine Arnault (*), Sylvie Thiria (*), Michel Crépon (*) and François Kaly (**)

(*) Sorbonne Université, CNRS, IRD, MNHN
Laboratoire d'Océanographie et du Climat: Expérimentations et Approches Numériques (LOCEAN)
75005 Paris
France

(**) Laboratoire de Traitement de l'Information (LTI)
Ecole Supérieure Polytechnique,
B.P 5085 Dakar-Fann
Sénégal

September 2020

Corresponding author email: sa@locean-ipsl.upmc.fr

Co-authors emails: fak_01@hotmail.com

thiria@locean-ipsl.upmc.fr

© 2020 published by Elsevier. This manuscript is made available under the CC BY NC user license
https://creativecommons.org/licenses/by-nc/4.0/

Abstract

The western tropical Atlantic Ocean is a very energetic and highly variable region. It is one of the main contributors to the inter-hemispheric mass and heat transports. This study aim is to give a new picture of the space and time variability of this region using statistical tools applied to five different satellite measurements (Sea Surface Temperature, Sea Surface Salinity, ocean topography, wind stress vectors). We first processed each data set by using a Self-Organizing Maps (SOM), which is an efficient clustering methodology based on non-linear artificial neural networks to compress the information embedded in the data. The SOM was then combined with a Hierarchical Ascendant Classification (HAC) to cluster the different phenomena in a small number of classes whose physical characteristics are easy to identify. Three classes were identified which allowed us to analyse the dynamics of the North Brazil Current, and the North Equatorial Countercurrent, respectively, and their links with the Inter-Tropical Convergence Zone and the Amazon and Orinoco river runoffs. The SOM + HAC analysis gave a coherent picture of the concomitant seasonal variability of the variables. Furthermore, we were able to point out the correlations existing between salinity features recently discovered and wind, temperature, and dynamic topography structures. Applying our method to the interannual signals, we showed a year to year variability which deserves further analysis.

Keywords: Atlantic, North Brazil Current, Machine Learning Self-Organizing Map, Satellite observations.

1. Introduction

The western tropical Atlantic Ocean (Fig. 1) is a very energetic and highly variable region. It houses the North Brazil Current (NBC), which is a powerful western boundary current that carries warm and salty waters of South Atlantic origin into the northern hemisphere, flowing northwest along the coast of Brazil, by crossing the equator. As a result, the NBC closes wind-driven circulation within the equatorial gyre (Gordon, 1986). The NBC feeds a system of zonal countercurrents, such as the surface North Equatorial Countercurrent (NECC), although the portion of the NBC transport retroflected is debated (e.g. Bourlès et al., 1999; Csanady, 1985; Philander and Pacanowski, 1986; Richardson et al., 1994; Schott and Böning, 1991). Furthermore, it contributes to the cross-equatorial transport of upper-ocean waters as part of the Atlantic meridional overturning cell (Johns et al, 1998). Large anticyclonic rings shed by the current swirl northwestwards along the South American coast, often reaching the eastern edges of the Lesser Antilles, where they eventually become absorbed into the Caribbean and Florida Currents (Arnault et. al., 1999; Fratantoni et. al., 1995; Paris et al., 2002; Schott et. al., 1998). The Sea Surface Salinity (SSS) in this region reflects the influence of both high precipitation under the Inter-Tropical Convergence Zone (ITCZ) and the discharge of major rivers along the South American coast (Mignot et al., 2007). When the NBC retroflects in the boreal summer-fall season, the Amazon plume water is deflected eastward in the NECC producing a freshwater signature which, with accompanying ITCZ rainfall, dramatically reduces salinities between 5° and 10°N (Lentz, 1995; Muller-Karger et al., 1988). Recently, however, Grodsky et al. (2014a) identified with the Aquarius/SAC-D satellite SSSs, a local salinity maximum centered at 8°N in boreal spring-early winter. Its existence seems to be the result of the different phases of the seasonal variations of Amazon discharge and ocean currents.

In this paper, we revisit the western tropical North Atlantic dynamics, using machine learning techniques. Machine learning is a field of statistical research for training computational algorithms that split, sort, and transform a set of data to maximize the ability to classify, predict, cluster, or discover patterns in a target data set. Indeed, a deluge of Earth system observation data has become

available, coming from a plethora of sensors measuring ocean states, air-sea fluxes, and intensive or time/space-integrated variables. They include satellite remote sensing observations done by different sensors such as SST (Sea Surface Temperature) and SSS as well as in situ observations (increasingly from autonomous sensors) at and below the surface and in the atmosphere. One key challenge is to extract interpretable information and knowledge from these new data sets and integrating between variables (Reichstein et al., 2019).

Thus, we decided to apply machine learning analysis to different satellite-derived oceanic data sets (surface salinity and temperature, dynamic topography, surface wind stress) in the tropical western Atlantic region. The aim (and the challenge) of this study is to investigate how this promising technology is able to identify oceanic phenomena and concomitant variabilities between parameters in an area where so many space and time scales are implied.

The paper is organized as follows. Section 2 presents the data sets and the methods (the Self-Organizing Map -SOM-, Hierarchical Ascendant Classification -HAC- approaches) we used. Results are given in section 3 and discussed in section 4. The conclusion follows in section 5.

2. Data and methods

2.1. Data

In this study, we used four oceanographic parameters: the SSS, the SST, the Absolute Dynamic Topography (ADT), and the surface wind stress (τx and τy), most of these data being issued from satellite data.

SMOS SSS

The SSS data were derived from the SMOS (Soil Moisture and Ocean Salinity) mission, which is part of the ESA (European Space Agency)'s Living Planet Programme. A new methodology for correcting systematic SSS biases (Boutin et al., 2018) has been used in these SMOS SSS. The

SMOS SSS are affected by biases coming from various nonphysical contaminations such as the so-called land contamination and latitudinal biases likely due to the thermal drift of the instrument. These biases are relatively weak and have almost no impact on soil moisture retrieval. On the contrary, for salinity estimation, the impact is non-negligible and can reach more than one salinity unit in some regions close to the coasts. These biases are not easy to characterize because they exhibit very strong spatial gradients and they depend on the coast orientation in the Field Of View (FOV) and on the position on the swath. LOCEAN/IPSL (UMR CNRS/Sorbonne Université/IRD/MNHN) and ACRI-st have derived a methodology for correcting systematic SSS biases. This version uses an improved debiasing technique that is also implemented in the CATDS-CPDC processing; only maps generation (filtering and smoothing) and geographical coverage differ from the CATDS-CPDC version. Boutin et al. (2018) found that the Root Mean Square (RMS) differences between the resulting 18-day SMOS SSS and 100-km averaged ship SSS are of the order of 0.20 in the open ocean. These new SMOS SSSs are mapped with a resolution of 0.25°x0.25° and produced every 4 days from 2010 till 2016. To check the consistency of these new SSS data in the studied region, we compared them with salinity data from the tropical Atlantic PIRATA (for now Prediction and Research Moored Array in the Tropical Atlantic, Bourlès et al. (2008)) network. We selected the buoys at 8°N, 38°W and 12°N, 38°W (Fig. 1). The comparison refers to the PIRATA mooring salinity at 1-meter depth (hereafter named SSS) and the SMOS SSS from the 0.25°x0.25° pixel located around the moorings. The processing of satellite measuring the SSS has led to study the effects of the near-surface stratification and subfootprint variability on the salinity, both effects being required when comparing in situ and satellite SSS values (e.g. Boutin et al., 2013. For instance, several studies have estimated that, on average, rain-induced surface freshening occurs ~16% of the time when considering the tropics (Boutin et al., 2013; Anderson and Riser, 2014; Drucker and Riser, 2014; Meissner et al., 2014). Anderson and Riser (2014) found that salinity in the upper 4 m is, in most cases, well mixed and that the difference between salinity at a few centimeters and 4 m below the sea surface is less than 0.1 for 97% of the observations.

Consequently, the near-surface fresh anomalies produced by rainfall are eliminated quickly, typically within a few hours. Freshwater plumes, causing vertical and horizontal gradients, can also complicate the comparison of satellite and in situ salinity measurements. Guiffard et al. (2019) compared 2 simulations of the tropical Atlantic Ocean with and without the Amazon runoff. They found that most of the impacts of the Amazon river on SSS occurred west of 45°W. The PIRATA moorings we selected along 38°W are thus far enough from the Amazon run off so that we can reasonably think the SSS gradients will be considerably attenuated by mixing and advection. The impact of the SMOS gridding on a 0.25°x0.25° mesh size is more difficult to assess. Vinagradova and Ponte (2013) quantified SSS variability within 1°x1° bins to be as high as 0.2 near western boundary currents and in river outflow regions. We will assume a value of 0.2 as a maximum error for the SMOS SSS. For the PIRATA SSS, Freitag et al. (1999) assumed that errors in the moored salinity time series are generally reduced to about 0.02.

The correlations between these two different SSS signals vary from 0.78 at 12°N to 0.91 at 8°N (Fig. 2). The RMS differences range between 0.24 (12°N) to 0.33 (8°N) which is close to the SMOS error. In particular, the important SSS decrease (sometimes below to 33) occurring every year in late boreal fall, is well captured by the SMOS observations. As mentioned previously, the SSS in this region reflects the influence of both high precipitation under the ITCZ and the discharge of major rivers along the American coast deflected eastward in the NECC. For instance, Arnault et al. (1992) observed low SSS (<33) around 28°W and between 6 and 8°N, during the IRD/CNES ARAMIS1 experiment in September-October 1988. These results confirm that these space SSS data are good representatives of the surface salinity variability in the studied region.

DUACS ADT

The altimeter products were processed by Ssalto/Duacs and Aviso+, with support from CNES (https://www.aviso.altimetry.fr/en/data/products/sea-surface-height-products/global/ssaltoduacs-experimental-products.html). They are presently distributed by CMEMS (Copernicus Marine

Environment Monitoring Service). We used the multi-mission altimeter data set reprocessing DUACS DT2014 (Pujol et al., 2016) whose main priority was to improve the monitoring of the mesoscales in the global ocean. The data from all altimeter missions (Jason1 and 2, TOPEX/Poséïdon, ENVISAT, GEOSAT Follow On, ERS1 and 2, and GEOSAT) are processed by the Ssalto/Duacs system to provide a consistent catalogue of products for varied applications, both for near real time applications and offline studies. We used the "twosat" daily series to keep a homogeneous data set over the period.

OSTIA SST

We also used the SST obtained from the Group for High Resolution Sea Surface Temperature (GHRSST). It consists of level 4 SST analysis produced daily on an operational basis at the UK Met Office using optimal interpolation (OI) on a global 0.054-degree grid. The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) analysis uses satellite data from sensors that include the Advanced Very High Resolution Radiometer (AVHRR), the Advanced Along Track Scanning Radiometer (AATSR), the Spinning Enhanced Visible and Infrared Imager (SEVIRI), the Advanced Microwave Scanning Radiometer-EOS (AMSRE), the Tropical Rainfall Measuring Mission Microwave Imager (TMI), and in situ data from drifting and moored buoys (UK Met Office. 2005. GHRSST Level 4 OSTIA Global Foundation Sea Surface Temperature Analysis. Ver. 1.0. PO.DAAC, CA, USA. Dataset accessed [YYYY-MM-DD] at http://dx.doi.org/10.5067/GHOST-4FK01; Martin et al., 2016; Stark et al., 2007).

Wind stresses

The wind stresses, τx (Eastward), τy (Northward), were obtained from the European Center for Medium Range Weather Forecasts (ECMWF) ERA-interim reanalysis data (http://www.ecmwf.int/research/era/do/get/era-interim).

All these data sets have been limited over the tropical Atlantic in the region defined by 0°N to 15°N

and 30°W to 80°W. The series extend from January 2010 to December 2016. We used daily products re-interpolated every 4 days on the SMOS time resolution and a 0.75°X 0.75° gridded map similar to the wind space resolution.

2.2. Methods

In this study, our methodology consists in applying an efficient clustering method performed in two steps (Fig. 3). The above oceanographic and atmospheric data corresponding to the 2010-2016 period have been clustered into a large number of prototypal oceanic situations by using a SOM. This number was then reduced by using an HAC in order to facilitate the analysis in terms of physical processes.

2.2.1 The SOM map

The SOM is an unsupervised classification method made of a competitive neural network structured in two layers (Kohonen, 2001) suitable for clustering large data sets. The first layer (or input layer) receives the data. The second one is a neuron grid. Each neuron is associated with a referent vector that has the dimension of the data and is statistically representative of a cluster of the data set. The SOM projects data to a lower-dimensional space (2D or 3D) constituted of the neuron grid and their associated referent vectors. This space is called a map. The map consists of neuronal sets interconnected by a non-oriented graph structure. The map grid summarizes the information contained in the multivariate data set. The neuron grid presents a topological ordering, which means that close neighbour regions on the SOM map represent similar situations, while dissimilar patterns are mapped further apart. The reference vectors are determined through a learning process (Kohonen, 2001), by minimizing a non-linear cost function. Initially, the SOMs were used for statistical research, but they can be used for any kind of purpose, especially in climate research (Richardson et al., 2007). The SOM is well adapted to the problem we solve. It has a higher discriminant power to separate the different classes than the Empirical Orthogonal Functions (EOF)

(Lorenz, 1956) or the K-Means algorithm (Badran et al., 2005) due the topological constrain among the different classes. Besides the SOM is able to deal with non-linear aspects of the phenomenon that we analyse, contrary to EOF, which mainly deals with linear aspects of the phenomena (Liu et al., 2006). More details about the SOM and its advantages in regards with other classification methods like the k-means can be found in the literature (e.g. Baçao et al., 2005; Badran et al., 2005; Kiang and Kumra, 2001; Liu and Weisberg, 2005; Murtagh and Hernandez-Pajares, 1995). For instance, Liu et al. (2006) compared the feature extraction performance of both SOM and EOF techniques by using artificial data representative of known patterns. The SOM was shown to extract the patterns of a linear progressive sine wave as the EOF did, even with noise added. However, in an experiment with multiple sets of more complex patterns, the EOF method failed in choosing all these patterns while the SOM succeeded. Thus, SOM applications are becoming increasingly useful in geosciences because it has been demonstrated to be an effective feature extraction technique that has many advantages over conventional data analysis methods (Liu et al., 2006; Liu and Weisberg, 2011). In oceanography, the SOMs have been mainly used in satellite algorithms (Brajard et al., 2007; Gueye et al, 2015; Niang et al., 2003; 2006; Tanguy, 2011), for biological and ocean color studies (e.g. El Hourany et al., 2019; Hardman-Mountford et al., 2003; Richardson et al., 2003, Saraceno et al., 2006) and to analyse oceanic phenomena such as current variability (e.g. Jouini et al., 2016; Leloup et al., 2007; Liu and Weisberg, 2005; 2007; Liu et al., 2006). An extensive review of SOM applications in oceanography is provided in Liu and Weisberg (2011), where the studies are referenced from the data sets, oceanic variables, and oceanic areas they consider. More recently, Liu et al. (2016) propose a dual-SOM technique to explore data in both spatial and temporal domains.

The SOM has many tunable parameters. Liu et al. (2006) analyzed the impact of the choice of these parameters through a sensitivity study over artificial data representative of known patterns. In the present study, we used the well-known SOFT Toolbox developed in the Laboratory of Information and Computer Science in the Helsinki University of Technologies (Copyright (C) 2000-2005 by Esa

Alhoniemi, Johan Himberg, Juha Parhankangas, and Juha Vesanto). The architectures of the SOM maps have been determined using the default settings of the Kohonen package that means batch training (which is the fastest training algorithm, Vesanto et al., 2000); linear initialization; hexagonal map lattice structure (which is the most popular neighborhood and results in a smoother map); sheet map shape and Gaussian neighborhood function; default sizes for the maps. Vesanto et al. (2000) and Liu et al. (2006) state that increasing the map size leads to more accurate results, but also more patterns. Therefore, our trade-off will be between accuracy and compressing information into manageable few patterns. We found that these default sizes that provide a large number of neurons are preferred to smaller ones that could mix up data of different meanings. In the present study, we used rectangular SOMs composed of 24*12 = 288 neurons.

2.2.2 The HAC clustering

The large number of neurons (288 in the present case) provided by the SOM map permits to take into account the complexity of the input data set but could remain difficult to handle both from a computer time-consuming point of view and from a visualization and interpretation one (Meza-Padilla et al., 2019). To counteract this difficulty, this large number of neurons is aggregated into a smaller number of categories based on the similarities of the subsets. We extracted a few pertinent classes from the subsets by clustering subsets having similar statistical properties, expecting that the classes can be associated with common oceanic characteristics. For that, a HAC is applied to the SOM maps using the Ward distance in order to aggregate the neurons into a small number of classes facilitating their interpretation as physical processes. The HAC algorithm is a bottom-up hierarchical classification (Jain and Dubes, 1988; Jain et al., 1999). This method iteratively computes a partition hierarchy (Badran et al., 2005). From the initial partition containing the neuron groups of the SOM map, two neurons of the same neighbourhood are aggregated at each iteration. These two subsets are selected by measuring their similarity according to the Ward criteria. The iterations continue until all the subsets of the partition are regrouped together. We aggregated the

neurons into 3 statistically significant classes. This number of classes has been selected by choosing the most significant discriminative partition with respect to the full dendrogram of the HAC according to the Davies-Bouldin and to the gap indexes (Davies and Bouldin, 1979; Tibshirani et al., 2001). The first one is an index that takes into account not only the intra-class homogeneity by comparing the distance between each situation vector and its reference vector, but also the heterogeneity between the different classes. The second one compares the within-class dispersion with that expected under an appropriate null dispersion. These indexes have been used with success by Jouani et al. (2016) in their analysis of the Sicilian channel circulation. We also performed several sensitivity tests on this number of classes and found that 3 was a good compromise between a value that minimizes the Davies-Bouldin index and a value that maximizes the gap index and that these 3 classes were able to detect the most dominant patterns of the large scale and low-frequency variability of our tropical Atlantic domain. In a recent study considering a finer scale circulation in the western Gulf of Mexico, Meza-Padilla et al. (2019) chose an optimal value of 7.

2.2.3 The 'strong forms'

An important step is to identify groups of ocean situation (or image) constituted of the five variables taken together. To avoid the curse of dimensionality due to the large number of variables with respect to the number of data, we trained 5 different SOM maps, one for each variable (SST, ADT, SSS, τx and τy). For each variable, the SOM algorithm has been applied as a vector quantization in order to compress the information contained in the dataset. Thus for each SOM map, a reduced number of “similar” cases was estimated using the HAC onto each set of the SOM neurons: each variable is represented by a number “generic cases”. Since each variable belongs to a class obtained at the end of the HAC classification, the inter-dependence of the five variables (the correlations) was taken into account by determining a common partition considering the class of each variable: the so-called “strong form” partition. This partition was done as follows: for instance, if at time N, the SST image belongs to class 1, the ADTs to class 3, the SSSs to class 2, the zonal wind stresses

to class 2 and the meridional to class 2, the associated "strong form" can be labelled (1,3,2,2,2). The strong forms present an effective structuration of the data set constituted of the five variables. This new clustering is representative of characteristic situations of the ocean: i.e. the number associated with a "strong form" permits to characterize it: it is a good compromise between the representativeness of a cluster and its general character, which will be verified in section 4. As said in section 2.2.2, we chose to reduce each initial map into 3 classes and determine the "strong forms" from these partitions. This leads to $3^5 = 243$ potential "strong forms", but due to the structuration of the data imposed by the ocean physics, we encountered a small number of "strong forms" only, that are interpreted in section 3. In the following we identified a "strong form" by a number; we heuristically tried to design similar "strong forms", i. e. which are contiguous in time, by close numbers.

Another method would have been to cluster the different variables with a single SOM map. This might give weak results because each variable does not have the same behavior and characteristic depending on space and time. We deal with ocean variables (SST, SSS, ADT) on the one hand and with atmospheric variables (τx, τy on the other hand) on the other hand which may have different spatial (and also to a lesser degree temporal) structures. Consequently, clustering a five variable set with a unique SOM might give confusing results, which were not easily physically interpretable.

3. Results

3.1 Analysis of the SOM according to ocean geography

We first made a spatial analysis of the SOM in order to check the data series and the performance of the SOM algorithm. We decomposed the studied period (January 2010 - December 2016) into 635-time steps of 4 days. The SST, SSS, ADT, τx, and τy variables are organized as follows: at each time step, each variable is represented by a latitude X longitude geographical map of 771 pixels (the continental values have been removed). The full dataset is therefore represented by a matrix of

dimensions (771, 3175) which contains the merged SST, ADT, SSS, and wind information for 5X635 (3175) time dependant data. The aim of our analysis is to summarize the information contained in the learning set by producing a small number of referent vectors that are statistically representative of the learning set. The SOM reduced the information to 24*12 = 288 referent vectors, each referent vector being associated with a dedicated neuron. After the SOM computation, each (771 X 3175) data is assigned a referent vector (neuron). We then monthly average the neuron values to reconstruct a monthly climatology from the SOM map. These reconstructions are presented for February and August, two opposite months of the tropical Atlantic cycle (e.g. Arnault, 1987) in Figures 4&5. If the SOM is pertinent over our data sets, these maps should reveal structures equivalent to the monthly climatology. This is clearly the case (Fig. 4&5). For both months, the mean differences are less than 0.02°C for the SST, 0.3 cm for the ADT, 0.02 for the SSS, and 0.02 N/m2 for the wind. In February (Fig. 4), both SOM reconstructions and climatology maps show southward easterlies over most of the area, weakening southwards due to the southern location of the ITCZ during that period. The SSTs present temperature gradients with SST decreasing polewards. The dynamic topography is rather flat with only a poorly shaped high along 5°N. The SSSs show two pools of freshwaters along the American coast due to the presence of the Amazon and Orinoco river mouths. In August (Fig. 5), the ITCZ migrates northward around 7°N-10°N so that the region shows northward easterlies South of 7°N, and southward ones North of that latitude. The SSTs indicate a warm tongue beneath the ITCZ, starting from the American coast and curving southwards, and colder waters north and south of it. The dynamic topography now presents well-marked zonal highs and lows so that the resulting zonal geostrophic NECC and North Equatorial Current (NEC) strengthen. The SSSs show a large area of freshwater extending eastward at ~5-10°N.

As stated earlier, this spatial study was mostly devoted to check the coherence of our data sets and the relevance of our SOM computation over them. This encouraging result allowed us to continue with the investigation on the temporal variability.

3.2 Time analysis

A time analysis will be far more interesting. Indeed, a global view of the time variability of the different phenomena in the western tropical Atlantic Ocean will be useful. We analysed the oceanographic situations by performing a clustering analysis data set from 2010 to 2016 using the SOM algorithm. To do this, we ordered the data in order to have the time steps as individuals and the spatial points as variables.

The study can be decomposed in two steps:

- First, as previously we trained 5 different SOM maps, one for each variable (SST, ADT, SSS, τx and τy).
- Second, the existing correlations between the 5 variables through the "generic cases" have been used to exhibit a limited number of coherent behaviours from an oceanographic point of view.

Before proceeding to the six years of analysis, we conducted the first experiment using the climatology estimated from these six years. This was done to show the efficiency of the "strong form" approach to enlighten the link between the different variables.

3.2.1. Climatological analysis

In this analysis we focused our attention on the second step of our methodology and estimated the seasonal climatological signal from the 6-year data set. For each variable, a database of 12 monthly climatological data and 771 pixels was formed (12X771) and we apply the HAC using the Ward distance for the intra-classes similarity. As mentioned above, the number of classes has been selected by choosing the most significant discriminative partition with respect to the full dendrogram of the HAC according to the Davies-Bouldin and to the gap indexes. Figure 6 shows, as an example, the SST dendrogram and histogram of the data after the HAC clustering. This SST example is representative of all the other oceanic parameters. These figures help to determine the

most significant discriminative partition. Following these analyses, 3 classes have been retained for deepening our study.

Figure 7 presents these three classes for the zonal τx and meridional τy wind stress components. Class 1 signal presents a North-South gradient for τx with trade zonal component (τx always westwards) increasing northwards. τy shows a negative extremum around 10°N in the East. Class 2 shows again τx values increasing from the Equator to 15°N which means a northward intensification of the trades. Compared to class 1, however, the increase is weaker. The pattern for τy shows the zero line (ITCZ) located near the Equator. Class 3 shows an extremum around 5°-10°N for τx implying that τx is reduced in that area. τy presents the ITCZ zero line in the same area between 5 and 10°N. Figure 8 presents the three classes for the SST. SST class 1 shows warm waters (>28°C) along the American coast, extending over the basin between 0 and 5°N approximately. SST class 2 presents a warm tongue (> 29°C) extending south-eastwards from the north-western area of the region. SST class 3 shows a northward gradient of temperatures with SST lower than 22°C in the north-eastern area of our domain. Figure 9 presents the SSS climatological classes. SSS class 3 shows a zonal band of very fresh SSS (<34) along the American coast, extending eastwards across the basin at ~7°N and northward between 50 and 60°W. SSS class 2 still offers these patterns but attenuated. The northern minimum is separated from the coastal and zonal one. They are more weakened on the SSS class 1 with the zonal minimum displaced northward around 11°N and the northern one nearly vanished. The results concerning the ADT classes (Fig. 10) show a highly contrasted relief with a succession of zonal highs and lows for the ADT class 2, especially for the NECC high along 3-4°N, and the NEC low between 10 and 15°N. It contrasts with the amorphous structures presented by the ADT class 3. The ADT class 1 shows an in-between situation. This first step of our analysis offers a coherent picture for any of our oceanic parameters.

Then, as explained in section 2.2, we obtain 10 strong forms for the climatological series. The time

evolution of these different strong forms presents a clear seasonal signal over the mean climatological year (Fig. 11). For instance, strong form Number 1 (Fig. 12) occurs in late boreal spring. The associated winds present an ITCZ located around 2-4°N. The trades are thus more intense in the north-Eastern part of the region. The SST is minimum while beneath the ITCZ where the winds weaken, a large area of maximum SST appears. Due to the evaporation linked to the intense winds, SSS increases north-eastwards. However, along the American coast, the Amazon runoff, which is maximum at that time, freshens the sea waters. Carried away by the NBC and the Ekman transport, these waters are mostly advected northwards. In response to these SST and SSS signals, the ADT presents rather similar patterns. Indeed, due to thermosteric and halosteric contributions, where the SST decreases and SSS increases (north-western area), the ADT decreases. In places where the SST increases and SSS decreases (along the coast), the ADT increases. Looking at the ADT gradients, thus at the geostrophic surface velocity, we can conclude that the NECC is only moderate which explains the relatively weak advection of the Amazon and Orinoco fresh waters eastwards. In October (strong form Number 4, boreal fall, Fig. 13), the ITCZ has migrated northwards. It settles around 7°N, resulting in rather intense winds North and South of this ITCZ area, and weak ones beneath. In response to the Ekman pumping resulting from wind curl, the ADT topography increases with a pronounced trough between 10 and 15°N, and a crest between the Equator and 5°N. The resulting geostrophic zonal NECC and coastal NBC intensify, advecting eastwards (up to 35°W) warm SSTs and relatively fresh waters which now extend across the region and northwards until the limit of our domain.

These preliminary results are encouraging. The picture they give of the common variability of winds, SSTs, SSSs, and ADTs is coherent and in agreement with the climatological cycle of the tropical Atlantic variability. A classification was therefore done on the whole 2010-2016 series.

3.2.2. 2010-2016 seasonal signal

We applied the same method to the whole 2010-2016 data of the oceanic parameters, but now we

used the SOM quantization and the HAC onto the neurons of the maps to determine the 3 “generic cases” of each variable. The “strong form” clustering is applied during the climatological analysis. The time-space dimensions of the starting matrices are (635, 771). We obtained about 70 strong forms, which are numbered according to the description given in section 2.2.3. Figure 14 gives the time evolution of these different ‘strong forms’ characterized by their associated number over the studied period. The signal is seasonal. From an oceanographic point of view, this is not a surprise due to the seasonal character of the tropical Atlantic Ocean, contrary to the Pacific Ocean for instance, where inter-annual variability associated with the El Niño-Southern Oscillation phenomenon prevails (e.g. Merle et al., 1979). However, due to the innovation of our approach in the geophysical fluid dynamics domain, it is worth noting. Figures 15 and 16 present 2 ‘strong forms’ occurring during the same seasons as those selected for climatology: boreal spring and fall. Strong form Number 11 occurs in spring every year (Fig. 15). It associates the same wind, SST, SSS and ADT patterns as those described previously for the climatology: the winds present an ITCZ located near the Equator: the SST is minimum in the north-eastern part of the zone where the winds are stronger, and warmer in areas where the winds weaken; the SSS increases in the strong wind areas, whilst the Amazon and Orinoco runoffs freshen the coastal sea waters. The ADT is linked to thermosteric and halosteric contributions, in such a manner that in areas where SST decreases and SSS increases (north-western area), ADT decreases. In areas where SST increases and SSS decreases (along the coast), ADT increases. In boreal fall (Fig. 16) the picture given by the strong form Number 53 is again very similar to the one described for the same climatological season: the ITCZ is further North, resulting in moderate (beneath it) to intense winds (North and South of it). Ekman pumping increases the ADT relief. The geostrophic NECC and NBC intensify, advecting eastwards and northwards warm SSTs and relatively fresh water.

3.2.3. 2010-2016 interannual signal

Despite the shortness of our time series, a first attempt was made to investigate the interannual

variability, or more precisely of the year to year variability of the oceanic parameters between 2010 and 2016. As the seasonal cycle prevails in the tropical Atlantic variability, the anomalies of these parameters (referring to their respective seasonal cycle) have been obtained, and the monthly means of these anomalies computed. The time-space dimensions of the starting matrices are (84, 771).

Figure 17 gives the time evolution of the 'strong forms' we determined. Interestingly, a kind of "contrast" can be observed between the years 2012-2014 on the one hand, and the year 2010 and 2016 on the other hand. Figure 18 presents the wind, SST, ADT, and SSS characteristics obtained for the strong form Number 49, representative of the boreal winter 2012, and the strong form Number 2, representative of the same season but in 2016. Clearly, both winters show contrasted situations. In 2012, the southward trades intensify north of 5°N. The SST show negative anomalies then. Two SSS striking features occur: first, a strong minimum anomaly between 50° and 58°W, extending northwards from the American coast. SSS freshens there. Then, along ~7°N, a local maximum (positive SSS anomalies) is embedded in mostly negative but weaker SSS anomalies. The ADT signal reveals a kind of frontal zone along 7-9°N with negative eddy-shaped anomalies North of it, and positive ones southward. In 2016, the southward trades decelerate. The SSTs increase. The SSSs now show positive anomalies in areas where the 2012 minimum was previously observed. The ADT anomalies are weak.

4. Summary and Discussion

After checking the consistency and the ability of our methods to extract information from a spatial analysis which gives coherent results, we analysed how the seasonal cycle of SST, ADT, SSS, and wind stress data sets are linked over the western tropical Atlantic Ocean, first from their climatological averages, then from the whole 2010-2016 period. The dominance of the seasonal cycle with respect to other time periods is clearly depicted. Every year during the boreal spring, our results show that the ITCZ is located around 2-4°N so that the trades are more intense in the north-

eastern part of the region. The SSTs are minimum while beneath the ITCZ, a large area of a maximum SST appears. Due to the evaporation, the SSS increases north-eastwards while river runoffs freshen the signal along the American coast. The ADT presents a combined response through halosteric and thermosteric contributions with only poor relief, thus moderate geostrophic circulation, during that season. In fall, the ITCZ migrates northwards. The ADT shows a pronounced trough between 10 and 15°N, and a crest between the Equator and 5°N in response to wind curl and Ekman pumping contribution. The resulting geostrophic zonal NECC and coastal NBC intensify, advecting eastwards (up to 35°W) warm SSTs and relatively freshwaters. This is in agreement with previous studies. Indeed, the tropical Atlantic is known to be in close equilibrium with the seasonally varying wind (e.g. Garzoli and Katz, 1983; Katz et al., 1977; Katz, 1981; Philander, 1979; Philander and Pacanowski, 1981). For instance, Merle and Arnault (1985) showed that the seasonal dynamic topography distribution is dominated by a series of zonally oriented ridges and troughs strongly contrasted in the summer-fall season and only poorly outlined in spring. Philander and Pacanowski (1986) indicate that the deepening of the thermocline in the western Atlantic is caused by the intensification of the zonal component of the equatorial trade winds piling up water in the western basin coast during the boreal summer-fall. This deepening is maximum at about 3°N in the vicinity of the NECC ridge. The asymmetry is caused by additional effects on the thermocline displacement of the meridional wind stress component and wind stress curl: the northward winds induce a downwelling at 3°N which is enhanced by the wind stress curl. Carton and Zhou (1997) regarding the annual cycle of the tropical Atlantic SST from a model simulation, observed strong SST variations off northwest Africa with minimum SSTs which result from local upwelling during the winter-spring season when the trades intensify. The way our analysis connects the winds, SST, and ADT signals at a seasonal time scale is thus in perfect agreement with the knowledge previously acquired on each of these variables separately. Adding the SSS information is a new step made enabled thanks to the satellite contribution. We observe that the SSS seasonal signal in this Atlantic region is mainly controlled by the high precipitations associated with the

ITCZ and the discharge of the South American rivers for the freshwater presence, and by local evaporation due to intense winds for the high salinity maximum. The role of the horizontal advection by surface currents is also shown from their geostrophic components linking ADT and SSS patterns, or from the Ekman transport linking wind and SSS structures. Recently, Grodsky et al. (2014a) point out a local SSS maximum using the Aquarius/SAC-D remote sensed data. This maximum is centred at 8°N. Despite the Aquarius data set only extends over 2 years, these authors hypothesize that this SSS maximum is seasonal and occurs during boreal winter. It could be due to salty South Atlantic waters advected by the NBC and retroflected into the western part of the NECC without any dilution from the local ITCZ rainfalls which have migrated southwards at that time. Similar SSS maxima can be noticed in the SSS climatology of Dessier and Donguy (1994) but none in the hydrography-based World Ocean Atlas of Boyer et al. (2012). Therefore, the challenge was to look at the potential existence of such a SSS maximum in our results and how, if any, it is connected to the other variables. Looking at figure 9 and the 3 mean SSS classes extracted from the climatological analysis, we found that class 1 is the only one which presents such a SSS maximum located at ~7-8°N. The second step was then to identify when this class 1 is present in the 'strong form' quintet composed of (SST, ADT, SSS, τx, τy). This occurred only twice, for the strong forms Number 6 and 9. Strong form Number 6 occurs in November (Fig. 19). It associates a wind pattern with an ITCZ around 2-3°N, and the remnant of an ADT relief in the NECC area. The hypothesis of salty waters being advected eastwards by the NECC without rainfall effect is coherent with this picture. Form 9 occurs one month later, in December (not shown). Then, the ITCZ is at its southernmost location, the ADT relief implies the disappearance of the NECC so that the SSS maximum, although not being diluted again, is no more advected eastwards. These conclusions are in perfect agreement with Grodsky et al. (2014a)'s hypothesis.

Once the demonstration of the ability of our method to point out coherent geophysical information on the seasonal variability of surface oceanic parameters has been established, a natural extent was to study the interannual ("year to year") variability. This constitutes a real challenge again due to

the weakness of the interannual signal in the tropical Atlantic. We observed that the interannual strong form evolutions present an opposition between the years 2012-2014 and the other ones. For instance, in the boreal winter 2012, the southward trades intensify north of 5°N. Not surprisingly, the SST shows negative anomalies then. Furthermore, this southward trade intensification accentuates the northward Ekman transport. In spring 2012, the Amazon discharge was particularly important. Consequently, Ekman transport advected a pool of freshwaters to the North up to the Caribbean which can explain the strong SSS negative anomaly extending northwards from the American coast. The ADT signal revealed a frontal zone along 7-9°N with eddy-shaped anomalies. This signal can be the remaining of Tropical Instability Waves along 5°N (Mélice and Arnault, 2017), implying a strong and unstable NECC during the year 2012. In early fall 2012, Grodsky et al. (2014b) mentioned higher salinity than in 2011 off the Amazon river. The SSS positive anomaly we observe offshore in our results along 7°N could be the advection of these high salinities by the strong NECC. In winter 2016, the situation reversed. Again, our method points out interesting and coherent features on the inter-annual/"year-to-year" variabilities of the wind, SST, SSS, and ADT variables in the north-western tropical Atlantic.

5. Conclusion

In this paper, we investigated how machine learning algorithms could help in understanding oceanic dynamics from satellite observations. Indeed, with the arrival during the past decade, of a new generation of satellite missions dedicated to the Earth observing system (e.g. Topex/Poseidon/Jason altimeters, SMOS, Aquarius radiometers...) oceanographers are facing the crucial problem of exploiting voluminous data sets to extract information and knowledge. For this, we applied machine learning algorithms (SOM and HAC) to SST, ADT, SSS, and wind data sets over the north-western tropical Atlantic Ocean to evidence coherent structures and the potential relations among the different variables. We demonstrated that the computations over the seasonal cycle of our data sets give a coherent picture of the concomitant variability of the variables. Furthermore, we pointed out

the correlation between salinity features recently evidenced and wind, temperature, and dynamic topography structures. Applying our method to the interannual signals, we showed a year to year variability which is promising although it needs further investigations.

Thus, the method makes it possible to identify coherent patterns from huge data sets, to classify them and to extract information and knowledge from the variables. These very encouraging results bring new perspectives in the oceanographic domain. The next step will be to investigate more pending questions about the tropical Atlantic Ocean such as the interannual variability which is still poorly described. In the future, the same SOM can also be applied to the time domain of the same data set to extract characteristic time series for different regions (e.g., Liu et al., 2006; Meza-Padilla et al., 2019).

Acknowledgement

This work has been funded by the french CNES (Centre National d'Etudes Spatiales) and the IRD (Institut de Recherche pour le Développement) organisations. The Group for High Resolution Sea Surface Temperature (GHRSST) Multi-scale Ultra-high Resolution (MUR) SST data were obtained from the NASA EOSDIS Physical Oceanography Distributed Active Archive Center (PO.DAAC) at the Jet Propulsion Laboratory, Pasadena, CA (http://dx.doi.org/10.5067/GHGMR-4FJ01). The Ssalto/Duacs altimeter products were processed by Aviso+, with support from CNES. They are now distributed by CMEMS (Copernicus Marine environment monitoring service). The L3_DEBIAS_LOCEAN_v2 Sea Surface Salinity maps have been produced by LOCEAN/IPSL (UMR CNRS/UPMC/IRD/MNHN) laboratory and ACRI-st company that participate to the Ocean Salinity Expertise Center (CECOS) of Centre Aval de Traitement des Donnees SMOS (CATDS). This product is distributed by the Ocean Salinity Expertise Center (CECOS) of the CNES-IFREMER Centre Aval de Traitement des Donnees SMOS (CATDS), at IFREMER, Plouzane (France). The wind stresses are obtained from the European Center for Medium Range Weather Forecasts (ECMWF) ERA-interim reanalysis data (http://www.ecmwf.int/research/era/do/get/era-interim). S. Arnault is funded by the french Institut de Recherche pour le Développement (IRD).

REFERENCES

Andersen, J.E., Riser, S.C., 2014. Near-surface variability of temperature and salinity in the near-tropical ocean: Observations from profiling floats. J. Geophys. Res., 119 (C11), 7433-7448, doi.org/10.1002/2014JC010112

Arnault, S., 1987. Tropical Atlantic geostrophic currents and ship drifts. J. Geophys. Res., 92, (C5), 5076-5088, doi:10.1029/JC092iC05p05076.

Arnault S., Gourdeau, L., Ménard, Y., 1992. Comparison of the altimetric signal with in-situ measurements in the Tropical Atlantic Ocean. Deep Sea Res., 39, 3/4, 481-499. doi: 10.1016/0198-0149(92)90084-7.

Arnault, S., Bourles, B., Gouriou, Y., Chuchla, R., 1999. Intercomparison of upper layer circulation of the western equatorial Atlantic Ocean: In situ and satellite data. J. Geophys. Res. 104 (C9): 21171–21194, doi:10.1029/1999JC900124.

Bação, F., Lobo, V., Painho, M., 2005. Self-organizing maps as substitutes for K-means clustering. In: Sunderam VS et al (eds) Computational science. Springer, Berlin, 476–483. doi: 10.1007/b136575.

Badran, F., Yacoub, M., Thiria, S., 2005. chap. Self-organizing maps and unsupervised classification. in edited by G. Dreyfus, 379–442, Springer, Neural Networks, Berlin Heidelberg, doi:10.1007/3-540-28847-3.

Bourles, B., Molinari, R.L., Johns, W.E., Wilson, W.D., 1999. Upper layer currents in the western tropical North Atlantic (1989-1991)". J. Geophys. Res. 104 (C1), 1361–1375. doi:10.1029/1998JC900025.

Bourles, B., Lumpkin R., McPhaden, M.J., Hernandez, F., Nobre, P., Campos, E., Yu, L., Planton, S., Busalacchi, A., Moura, A.D., Servain, J., Trotte, J., 2008. The PIRATA Program: History, Accomplishments, and Future Directions. Bull. Amer. Meteor. Soc., 89, 1111-1125.

Boutin J., Reverdin, G., Yin, X., Gaillard, F., 2013. Sea surface freshening inferred from SMOS and ARGO salinity: Impact of rain. Ocean Science, 9(1), 183-192, doi: 10.5194/os-9-183-2013.

Boutin J., Chao, Y., Asher, W.E., Delcroix, T., Drucker, R., Drushka, K., Kolodziejczyk, N., Lee, T., Reul, N., Reverdin, G., Schanze, J., Soloviev, A., Yu, L., Anderson, J., Bruckner, L., Dinnat, E., Santos-Garcia, A., Jones, W.L., Maes, C., Meissner, T., Tang, W., Vinogradova, N., Ward, B., 2016. satellite and in situ salinity, understanding near surface stratification and subfootprint variability. Bull. Atm. Meteorol. Soc., 97 (8): 1391–1407. https://doi.org/10.1175/BAMS-D-15-00032.1

Boutin J., Vergely, J.L., Marchand, S., d'Amico, F., Hasson, A., Kolodziejczyk, N., Reul, N., Reverdin, G., Vialard, J., 2018. New SMOS Sea Surface Salinity with reduced systematic errors and improved variability. Remote Sensing of Environment, 214, 115-134, doi.org/10.1016/j.rse.2018.05.022.

Boyer, T.P., Levitus, S., Antonov, J.I., Reagan, J.R., Schmid, C., Locarnini, R., 2012. [Subsurface salinity] Global oceans [in State of the Climate in 2011]. Bull. Atm. Meteorol. Soc., 93(7), S72-S75.

Brajard, J., Niang, A., Sawadogo, S., Fell, F., Santer R., Thiria S., 2007. Estimating aerosol parameters above ocean from MERIS observation using topological maps. Int. J. Remote Sensing, Vol 28(3-4), 781-795.

Carton, J.A., Zhou, Z., 1997. Annual cycle of sea surface temperature in the tropical Atlantic Ocean. J. Geophys. Res., 102, C13, 27813-27824.

Csanady, G.T., 1985. A zero potential vorticity model of the North Brazilian Current. J. Mar. Res., 43, 553-579.

Davies, D., Bouldin, D., 1979. A cluster separation measure. IEEE Trasactions on Pattern Anaysis and Machine Intelligence, 2, 224-227, doi: 10.1109/TPAMI.1979.4766909.

Dessier, A., Donguy, J.R., 1994. The sea surface salinity in the tropical Atlantic between 10°S and 30°N – Seasonal and interannual variations (1997-1989). Deep Sea Res., Pat 1, 41, 81-100.

Drucker, R., Riser, S.C., 2014.Validation of Aquarius sea surface salinity with Argo: Analysis of error due to depth of measurement and vertical salinity stratification. J. Geophys. Res., 119 (C7), 4626-4637, doi: 10.1002/2014JC010045.

El Hourany, R., Abboud-Abi Saab, M., Faour, G., Meija, C., Crépon, M., Thiria, S., 2019. Phytoplankton diversity in the Mediterranean Sea from satellite data usinf Self-Organizing Maps. J. Geophys. Res., 124, 5827-5843, https://doi.org/10.1029/2019JC015131.

Fratantoni, D.M., Johns, W.E., Townsend, T.L., 1995. Rings of the North Brazil Current: Their structure and behavior inferred from observations and a numerical simulation. J. Geophys. Res., 100, 10633-10654, doi:10.1029/95JC00925.

Freitag, H.P., McCarty, M.E., Nosse, C., Lukas, R., McPhaden, M.J., Cronin, M.F., 1999. COARE Seacat data: calibrations and quality control procedures. NOAA Technical Memorandum ERL PMEL-115, 89pp

Garzoli, S., Katz, E., 1983. The forced annual reversal of the Atlantic North Equatorial Countercurrent. J. Phys. Oceanogr., 13, 2082–2090.

Gordon, A.L., 1986. Interocean exchange of thermocline water. J. Geophys. Res., 91 (C4), 5037–5046. Doi:10.1029/JC091iC04p05037.

Grodsky, S.A., Carton, J.A., Bryan, F.O., 2014a. A curious local surface salinity maximum in the northwestern tropical Atlantic. J. Geophys. Res., vol. 119, 484-495, doi:10.1002/2013JC009450.

Grodsky, S.A., Reverdin, G., Carton, J.A., Coles,V.J., 2014b. Year to year salinity changes in the Amazon plume: contrasting 2011 and 2012 Aquarius/SACD and SMOS satellite data.Rem. Sens. Of Environ., 140, 14-22, doi:10.1016/j.rse.2013.08.033.

Gueye, M.B., Niang, A., Arnault, S., Thiria, S., 2015. A neural approach for salinity profile completion using a recursive algorithm. Res. J. of Applied Sciences, engineering and technology, 10(2), 221-229, 10.19026/rjaset.10.2575.

Guiffard, P., Llovel, W., Jouanno, J., Morvan, G., Decharme, B., 2019. Contribution of the Amazon river discharge to regional sea level in the topical Atlantic Ocean. Water, 11, 2348, doi:10.3390/w11112348.

Hardman-Mountford,N., Richardson, A.J., Boyer, D.C., Kreiner, A., Boyer, H.J., 2003. Relating sardine recruitment in the Northern Benguela to satellite-derived sea surface height using a neural network pattern recognition approach. Prog. Oceanogr., 59 (2-3), 241-255, doi:10.1016/j.pocean.2003.07.005.

Jain, A.K., Dubes, R.C., 1988. Algorithms for Clustering Data. Prentice-Hall, Englewood Cliffs.

Jain, A.K., Murty, M., Flynn, P., 1999. Data clustering, a review. ACM Comput. Surv., 31, 264-323.

Johns, W.E., Lee, T.N., Beardsley, R.C., Candela, J., Limeburner, R., Castro, B., 1998. Annual cycle and variability of the North Brazil Current. J. Phys. Oceanogr., 28 (1), 108–128, doi:10.1175/1520-0485(1998)028<0103:ACAVOT>2.0.CO;2.

Jouini, M., Béranger, K., Arsouze, T., Beuvier, J., Thiria, S., Crépon, M., Taupier-Letage, I., 2016. The Sicily Channel surface circulation revisited using a neural clustering analysis of a high-resolution simulation. J. Geophys. Res. Oceans, 121, 4545–4567, doi: 10.1002/2015JC011472.

Katz, E.J., 1981. Dynamic topography of the sea surface in the Equatorial Atlantic. J. Mar. Res., 39, 53-63.

Katz, E.J., Belevitsch, R., Bruce, J., Bubnov, V., Cochrane, J., Duing, W., Hisard, P., Lass, H., Meincke, J., DeMesquita, A., Miller, L., Rybnikov, A., 1977. Zonal pressure gradient along the equatorial Atlantic, J. Mar. Res., 35(2), 293-307.

Kiang, M. Y., Kumar, A., 2001. An Evaluation of Self-Organizing Map Networks as a Robust Alternative to Factor Analysis in Data Mining Applications. Information Systems Res., 12(2), 177-194.

Kohonen, T., 2001. Self-Organizing Maps. Third, Extended Edition. Springer Series in Information Sciences vol. 30, Berlin, Germany: Springer-Verlag, ISBN 978-3-540-67921-9

Leloup, J., Lachkar, Z., Boulanger, J., Thiria, S., 2007. Detecting decadal changes in ENSO using neural networks. Climate dynamics, 28(2), 147-162.

Lentz, S.J., 1995. Seasonal variations in the horizontal structure of the Amazon plume inferred from historical hydrographic data. J. Geophys. Res., 100(C2), 2391-2400, doi: 10.1029/94JC01847.

Liu, Y., Weisberg, R.H., 2005. Patterns of ocean current variability on the West Florida Shelf using the self-organizing map. J. Geophys. Res., 110, C06003, doi:10.1029/2004JC002786.

Liu, Y., Weisberg, R.H., 2007. Ocean currents and sea surface heights estimated across the West Florida Shelf. J. Phys. Oceanogr., 37, 1697-1713, doi: 10.1175/JPO3083.1.

Liu, Y. and Weisberg, R.H., 2011. A Review of Self-Organizing Map Applications in Meteorology and Oceanography, Self Organizing Maps—Applications and Novel Algorithm Design. In: Mwasiagi Josphat Igadwa Ed., IntechOpen, 22 pp., doi: 10.5772/13146

Liu, Y., Weisberg, R.H., Mooers, N.K., 2006. Performance evaluation of the self-organizing map for future extraction. J. Geophys. Res., 111, C05018, doi:10.1029/2005JC003117.

Liu, Y., Weisberg, R.H., Vignudelli S., Mitchum G., 2016. Patterns of the loop current system and regions of sea surfaceheight variability in the eastern Gulf of Mexico revealed by theself-organizing maps. J. Geophys. Res., 121, 2347–2366, doi:10.1002/2015JC011493.

Lorenz, E.N., 1956. Empirical orthogonal functions and statistical weather prediction, Stat. Forecasting Proj. Rep., Tech. Rep. 1, 1–49, Dep. of Meteorol., MIT, Cambridge.

Martin, M., Fiedler, E., Roberts-Jones, J., Blockley, E., McLaren, A., Good, S., 2016. PRODUCT USER MANUAL For OSTIA Near Real Time Level 4 SST products over the global ocean SST-GLO-SST-L4-NRT-OBSERVATIONS-010-001, CMEMS version scope: version 2.3, Approval Date : 24 October 2016, 34 pp.

Meissner, T., Wentz, F., Scott, J., Hilburn. K., 2014. "Upper ocean salinity stratification and rain freshening in the tropics observed from Aquarius." 2014 IEEE Geoscience and Remote Sensing Symposium, 5111-5114 [10.1109/igarss.2014.6947647].

Mélice, J.L., Arnault, S., 2017. Investigation of the intra-annual variability of the North Equatorial Countercurrent/North Brazil Current eddies and of the instability waves of the North tropical Atlantic Ocean using satellite altimetry and Empirical Mode Decomposition. J. Atm. Ocean. Techn., 34, 2295-2310, doi:10.1175/JTECH-D-17-0032.1.

Merle, J., Fieux M., Hisard, P., 1979. Annual signal and interannual anomalies of sea surface temperatures in the Eastern Equatorial Atlantic Ocean. Deep-Sea Res.. GATE supplement II to V, 26, 77-101. https://doi.org/10.1016/B978-1-4832-8366-1.50023-6.

Merle, J., Arnault, S., 1985. Seasonal variability of the surface dynamic topography in the tropical Atlantic 0cean. J. Mar. Res., 43, 267-288.

Meza-Padilla, R., Enriquez, C., Liu, Y., Appendini, C.M., 2019. Ocean Circulation in the Western Gulf of Mexico Using Self-Organizing Maps. J. Geophys. Res., vol. 124, 6, 4152-4167, doi:/10.1029/2018JC014377.

Mignot, J., de Boyer Montégut, C., Lazar, A., Cravatte, S., 2007. Control of salinity on the mixed layer depth in the world ocean: 2. Tropical areas. J. Geophys. Res., 112, C100010, doi:10.1029/2006JC003954.

Muller-Krager, F.E., McClain, C.R., Richardson, P.L., 1988. The dispersal of Amazon's water. Nature, 333, 56-58.

Murtagh, F., Hernández-Pajares, F., 1995. The Kohonen self-organizing feature map method: an assessment. J. of Classification, 12, 165–190.

Niang, A., Gross, L., Thiria, S., Badran, F., Moulin, C., 2003. Automatic Neural Classification of Ocean Colour Reflectance Spectra at the Top of the Atmosphere with Introduction of Expert Knowledge. Remote Sensing of Environment, vol. 86(2), 257-271.

Niang, A., Badran, F., Moulin, C., Crepon, M., Thiria, S., 2006. Decoding aerosol typology and optical thickness over the Mediterranean from SeaWifs images using neural methodology. Remote

Sensing of Environment, vol. 100, 82-94.

Paris, C.B., Cowen, R.K., Lwiza, K.M.M., Wang, D.P., Olson, D.B., 2002. Multivariate objective analysis of the coastal circulation of Barbados, West Indies: implications for larval transport. Deep-Sea Res. Part I, 48(8), 1363-1386.

Philander, S.G.H., 1979. Variability of the tropical ocean. Dyn. Atmos. Oceans., 3, 191-208.

Philander, S.G.H., Pacanowski, R.C., 1981. Response of equatorial oceans to periodic forcing. J. Geophys. Res., 86, 1903-1916.

Philander, S.G.H., Pacanowski, R.C., 1986. A Model of the Seasonal Cycle in the Tropical Atlantic Ocean. J. Geophys. Res., 91, 14192-14206, doi:10.1029/JC091iC12p14192.

Pujol, M.-I., Faugère, Y., Taburet, G., Dupuy, S., Pelloquin, C., Ablain, M., Picot, N., 2016. DUACS DT2014: the new multi-mission altimeter data set reprocessed over 20 years, Ocean Sci., 12, 1067-1090, doi:10.5194/os-12-1067-2016.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, M., Carvalhais, N., Prabhat, 2019. Deep learning and process understanding for data-driven Earth system science. Nature, 566, 195-204, doi.org/10.1038/s41586-019-0912-1

Richardson, A.J., Risien, C., Shillington, F.A., 2003. Using self-organizing maps to identifiy patterns in satellite imagery. Prog. Oceanogr., 59, 223-239.

Richardson, P.L., Hufford, G.E., Limeburner, R., Brown, W.S., 1994. North Brazil Current retroflection eddies. J. Geophys. Res., 99, C3, 5081-5093, doi:10.1029/93JC03486.

Richardson, A., Risien, C., Shillington, F., 2007. Using self organizong maps to identifiy patterns in satellite imagery. Prog. In Oceanogr., 59(2-3), 223-239.

Saraceno, M., Provost, C., Lebbah, M., 2006. Biophysical regions identification using an artificial neuronal network: A case study in the South Western Atlantic. Advances in Space Research 37 (2006) 793–805, doi:10.1016/j.asr.2005.11.005.

Schott, F. A., Boning, C. W., 1991. The WOCE model in the western equatorial Atlantic: Upper layer circulation, J. Geophys. Res., 96, 6993-7004.

Schott, F. A., Fischer, J., Stramma, L., 1998. Transports and pathways of the upper-layer circulation in the western tropical Atlantic. J. Phys. Oceanogr., 28(10), 1904-1928.

Stark, J.D., Donlon, C.J., Martin, M.J., McCulloch, M.E., 2007. OSTIA : An operational, high resolution, real time, global sea surface temperature analysis system., Oceans 07 IEEE Aberdeen, conference proceedings. Marine challenges: coastline to deep sea. Aberdeen, Scotland.IEEE.

Tanguy, Y., 2011. Variabilité de la dynamique et de la thermodynamique dans l'Atlantique tropical : projet ARAMIS. Thèse de l'université Pierre et Marie Curie. Paris. France. 230 pp.

Tibshirani, R., Walther, G., Hastle, T., 2001. Estimating the number of cluster in data sets via the gap statistic. J. R. Stat. Soc., Ser. B., 63(2), 411-423, doi:10.1111/1467-9868.00293.

Vesanto, J., Himberg, J., Alhoniemi, E., Parhankangas, J., 2000. SOM toolbox for Matlab 5, report,

Helsinki University of Technology, Helsinki, Finland.

Vinogradova, N.T., Ponte, R.M., 2013. Small-Scale Variability in Sea Surface Salinity and Implicationsfor Satellite-Derived Measurements. J. Atm. Oceanogr. Technology, 30, 2689-2694, doi: 10.1175/JTECH-D-13-00110.1.

LIST OF THE MAIN ACRONYMS USED IN THIS PAPER

AATSR : Advanced Along Track Scanning Radiometer
ACRI-ST : ACRI-Sciences de la Terre
ADT : Absolute Dynamic Topography
AMSRE : Advanced Microwave Scanning Radiometer-EOS
AVHRR : Advanced Very High Resolution Radiometer
Aquarius/SAC-D : Aquarius/Satélite de Aplicaciones Científicas-D
ARAMIS : Altimétrie sur un Rail Atlantique et Mesures InSitu
AVISO : Archiving, Validation and Interpretation of Satellite Oceanographic Data
CATDS-CPDC : Centre Aval de Traitement des Données SMOS-Data Production Center
CMEMS : Copernicus Marine Environment Monitoring Service
CNES : Centre National d'Etudes Spatiales
CNRS : Centre National de la Recherche Scientifique
ECMWF : European Center for Medium Range Weather Forecasts
ENVISAT : Environment Satellite
EOF : Empirical Orthogonal Functions
ERS1-2 : European Remote Sensing satellite 1-2
ESA : European Space Agency
FOV : Field Of View
GEOSAT : GEOdetic SATellite
GHRSST : Group for High Resolution Sea Surface Temperature
HAC : Hierarchical Ascendant Classification
IRD : Institut de Recherche pour le Développement
ITCZ : InterTropical Convergence Zone
LOCEAN/IPSL : Laboratoire d'Océanographie et du Climat: Expérimentations et Approches Numériques / Institut Pierre Simon Laplace
LTI : Laboratoire de Traitement de l'Information
MNHN : Muséum National d'Histoire Naturelle
NBC : North Brazil Current
NECC : North Equatorial CounterCurrent
NEC : North Equatorial Current
OSTIA : Operational Sea Surface Temperature and Sea Ice Analysis
PIRATA : now Prediction and Research Moored Array in the Tropical Atlantic
RMS : Root Mean Square
SMOS : Soil Moisture and Ocean Salinity
SOM : Self-Organizing Map
SEVIRI : Spinning Enhanced Visible and Infrared Imager
SSALTO/DUACS : Segment Sol Multimission Altimetry and Orbitography/Developing Use of Altimetry for Climate Studies-Data Unification and Altimeter Combination System
SSS : Sea Surface Salinity
SST : Sea Surface Temperature
TMI : Tropical Rainfall Measuring Mission Microwave Imager
τx τy : Zonal and meridional wind stress component
TOPEX/Poséidon : Ocean Topography Experiment/Poséidon
UMR : Unité Mixte de Recherche

FIGURE CAPTION

Figure 1: Main characteristics of the investigated region with the Amazon and Orinoco mouths; the Intertropical Convergence Zone (ITCZ), the North Brazil Current (NBC) and North Equatorial Countercurrent (NECC) mean locations; the red dots indicate the 2 PIRATA (now Prediction and Research Moored Array in the Tropical Atlantic) moorings used in the study.

Figure 2: Comparison of SMOS SSS with salinity data from the tropical Atlantic PIRATA network, for the buoy at 12°N, 38°W (up), and at 8°N, 38°W (bottom). The correlation between the 2 signals is higher than 0.7 and the RMS difference around 0.3. Error on SMOS SSS is assumed to be 0.2, error on PIRATA SSS is assumed to be 0.02.

Figure 3: Flow diagram of our methodology.

Figure 4: The SOM reconstructed maps (up) and climatological maps (bottom) for February and for the different data sets (SST, ADT, SSS, wind stress). Units are (°C) for SST, (m) for ADT and (N/m^2s) for wind stress.

Figure 5: Same as Figure 4 but for August.

Figure 6: Dendogram and data repartition for the first 20 clusters for the SST associated with the climatological SOM + HAC analysis.

Figure 7: First 3 mean classes retained for the wind stress τx τy classification in the climatological analys*is*. The zero line for τy is in bold to underline the ITCZ location. Units are N/m^2s.

Figure 8: Same as Figure 7 but for the SST classification. Units are °C.

Figure 9: Same as Figure 7 but for the SSS classification.

Figure 10: Same as Figure 7 but for the ADT classification. Units are m.

Figure 11: Time evolution over the climatological months of the different strong forms obtained after the SOM + HAC classification on the variable quintet (SST, SSS, ADT, τx, τy).

Figure 12: Mean characteristics of the different oceanic variables: SST, ADT, SSS and the wind stresses climatological components for the strong form Number 1 in May-June. Units are (°C) for SST, (m) for ADT and (N/m^2s) for wind stress.

Figure 13: Same as figure 12 but for strong form Number 4 in October.

Figure 14: Time evolution over the 2010-2016 period of the different strong forms obtained after the SOM + HAC classification on the variable quintet (SST, SSS, ADT, τx, τy).

Figure 15: Mean characteristics of the different oceanic variables: SST, ADT, SSS and the wind stresses for strong form Number 11 occurring during boreal spring every year. Units are (°C) for SST, (m) for ADT and (N/m^2s) for wind stress.

Figure 16: Same as Figure 15 but for strong form Number 53 occurring in boreal fall every year.

Figure 17: Time evolution over the 2010-2016 period of different strong forms obtained after the SOM+HAC classification on the variable quintet (SST, SSS, ADT, τx, τy). The seasonal cycle has been removed from all time-series. This can be considered as a "year-to-year" variability.

Figure 18: Mean characteristics of the different oceanic variables SST, ADT, SSS and wind stresses "interannual" components referred to their climatological averages and for the strong forms occurring during boreal winter 2012 (up) and boreal winter 2016 (bottom). The zero lines are bold. Units are (°C) for SST, (m) for ADT and (N/m^2s) for wind stress.

Figure 19: Mean characteristics of the different oceanic variables: SST, ADT, SSS and wind stresses climatological components for the strong form Number 6 occurring in November. Units are (°C) for SST, (m) for ADT and (N/m^2s) for wind stress.

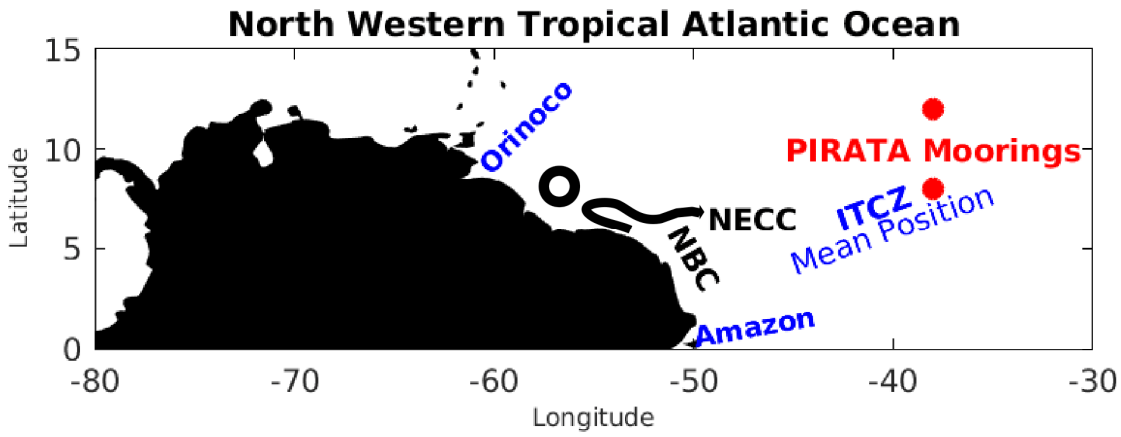


Figure 1: Main characteristics of the investigated region with the Amazon and Orinoco mouths; the Intertropical Convergence Zone (ITCZ), the North Brazil Current (NBC) and North Equatorial Countercurrent (NECC) mean locations; the red dots indicate the 2 PIRATA (now Prediction and Research Moored Array in the Tropical Atlantic) moorings used in the study.

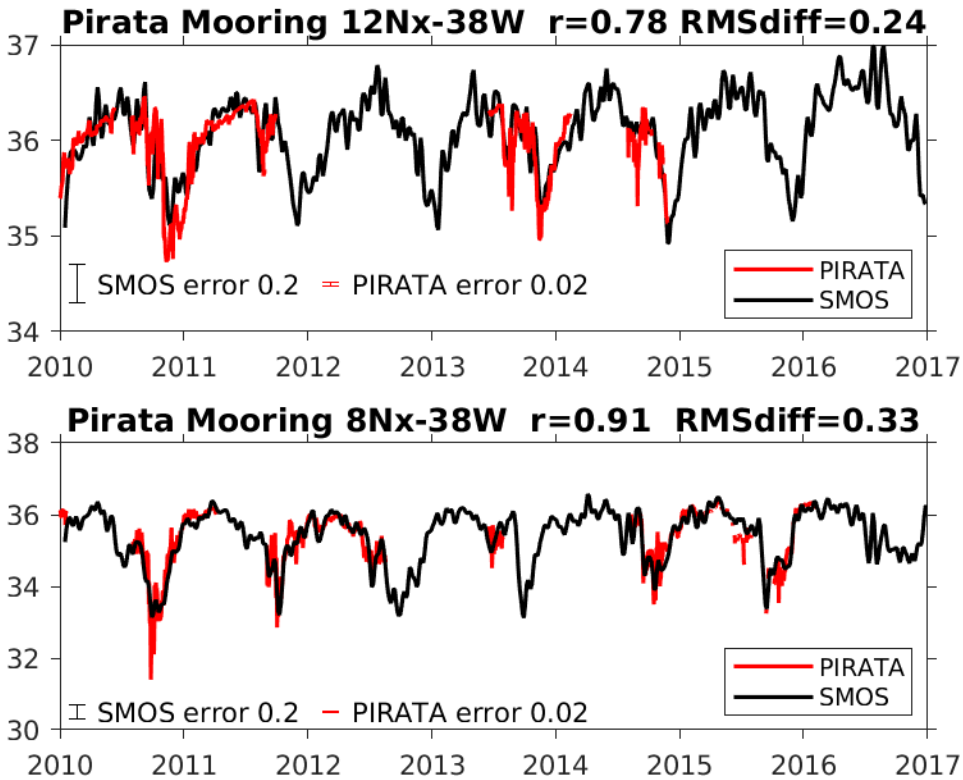


Figure 2: Comparison of SMOS SSS with salinity data from the tropical Atlantic PIRATA network, for the buoy at 12°N, 38°W (up), and at 8°N, 38°W (bottom). The correlation between the 2 signals is higher than 0.7 and the RMS difference around 0.3. Error on SMOS SSS is assumed to be 0.2, error on PIRATA SSS is assumed to be 0.02.

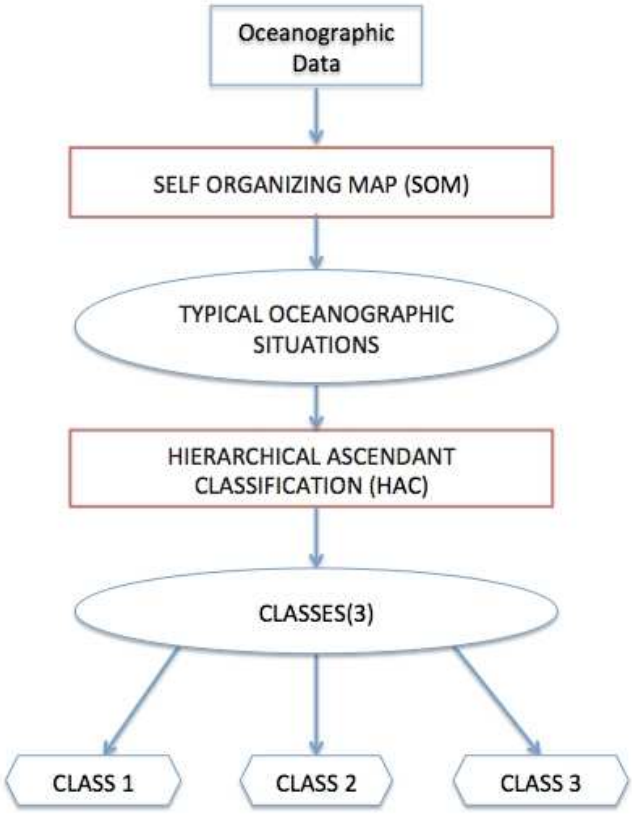


Figure 3: Flow diagram of our methodology

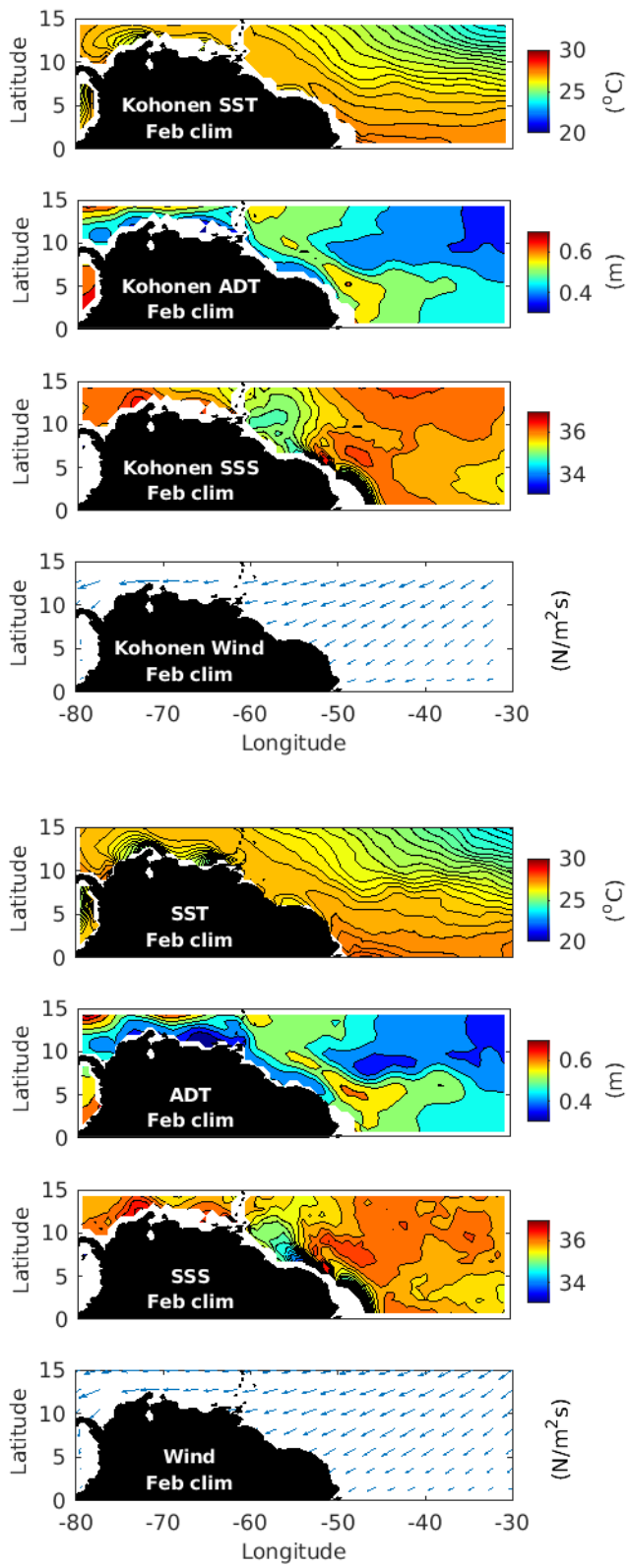


Figure 4: The SOM reconstructed maps (up) and climatological maps (bottom) for February and for the different data sets (SST, ADT, SSS, wind stress). Units are (°C) for SST, (m) for ADT and (N/m^{2}s) for wind stress.

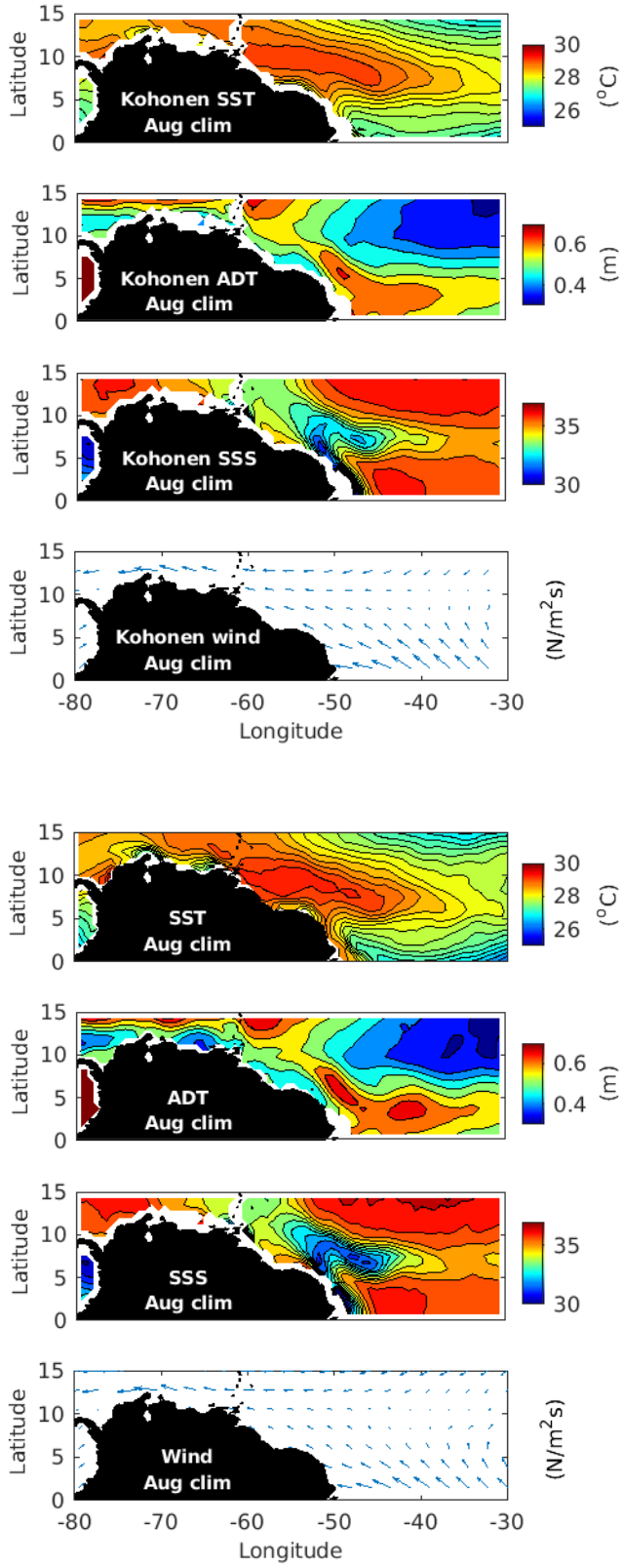


Figure 5: Same as Figure 4 but for August.

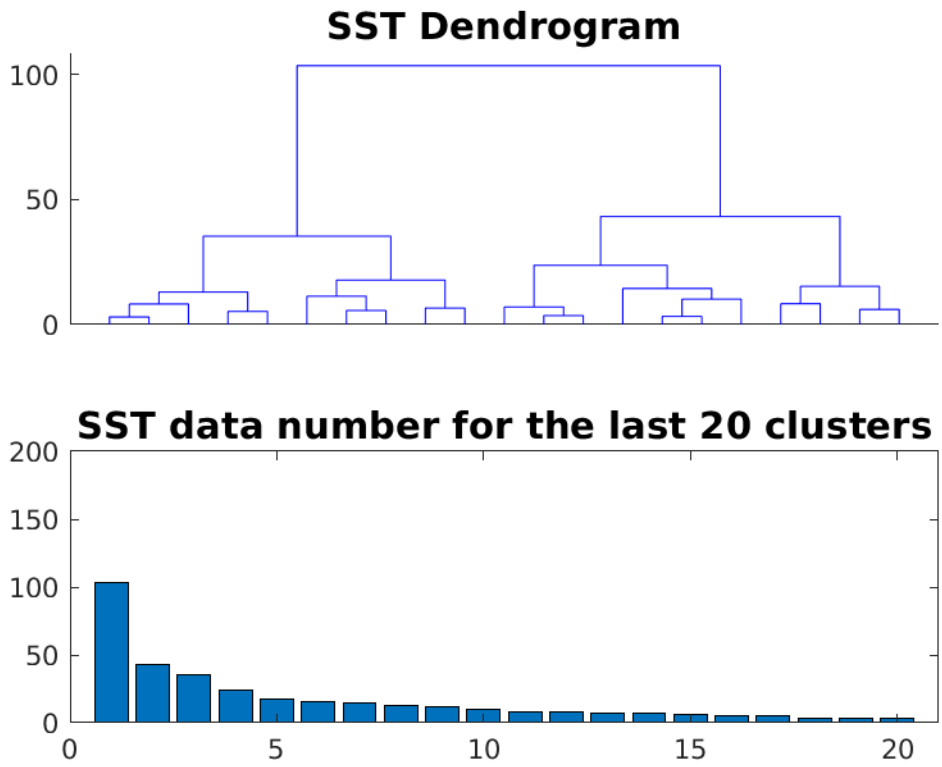


Figure 6: Dendogram and data repartition for the first 20 clusters for the SST associated with the climatological SOM + HAC analysis.

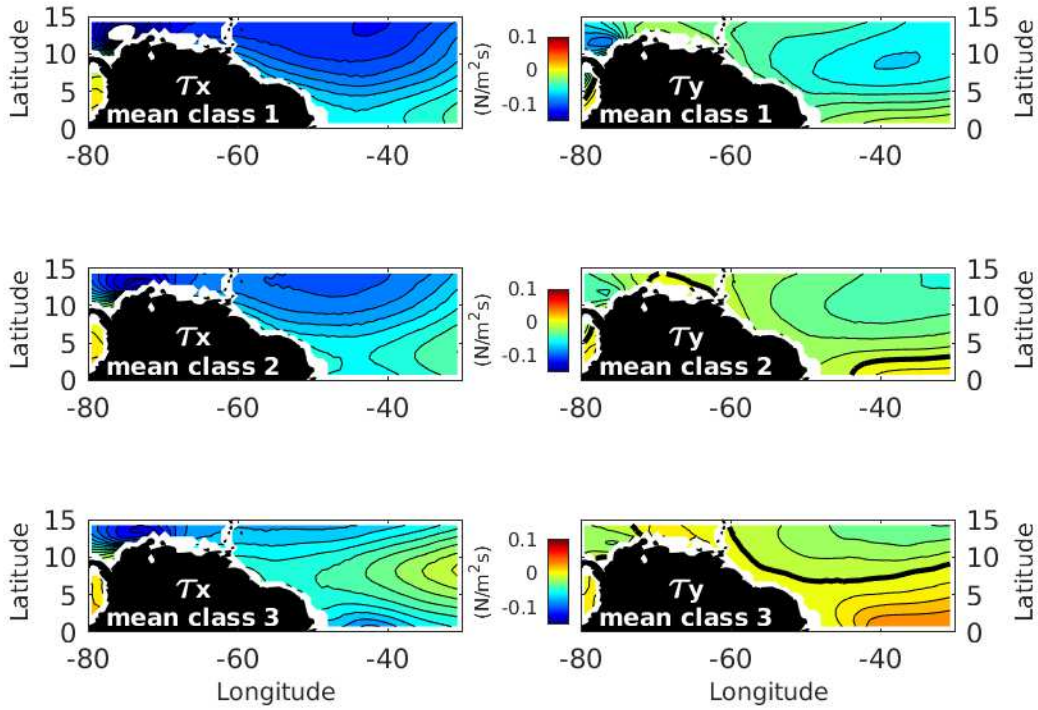


Figure 7: First 3 mean classes retained for the wind stress τx τy classification in the climatological analysis. The zero line for τy is in bold to underline the ITCZ location. Units are N/m^{2}s.

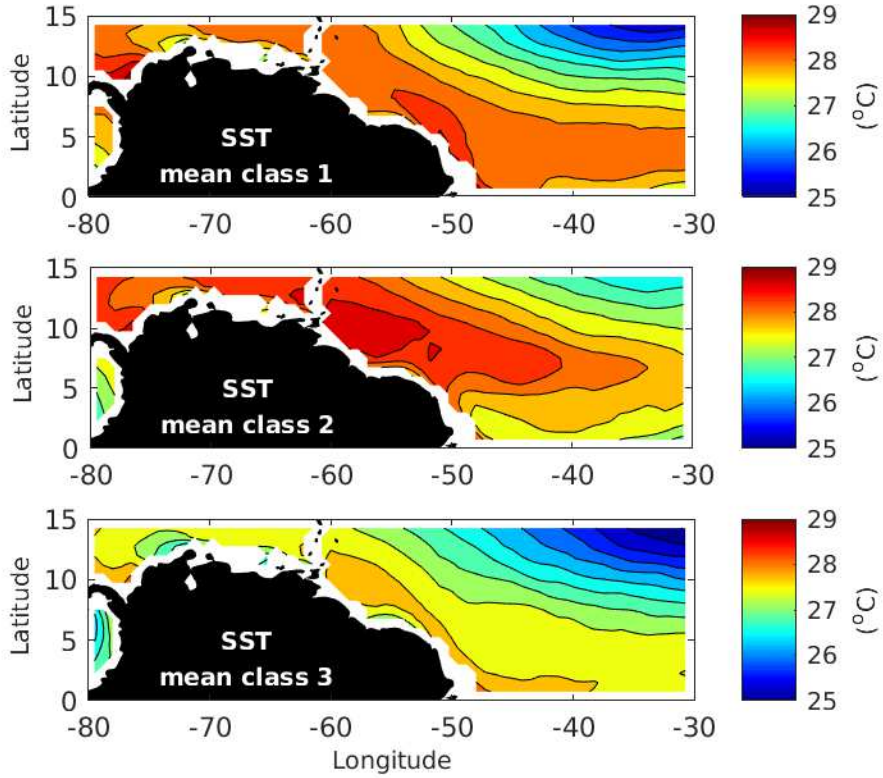


Figure 8: Same as Figure 7 but for the SST classification. Units are °C.

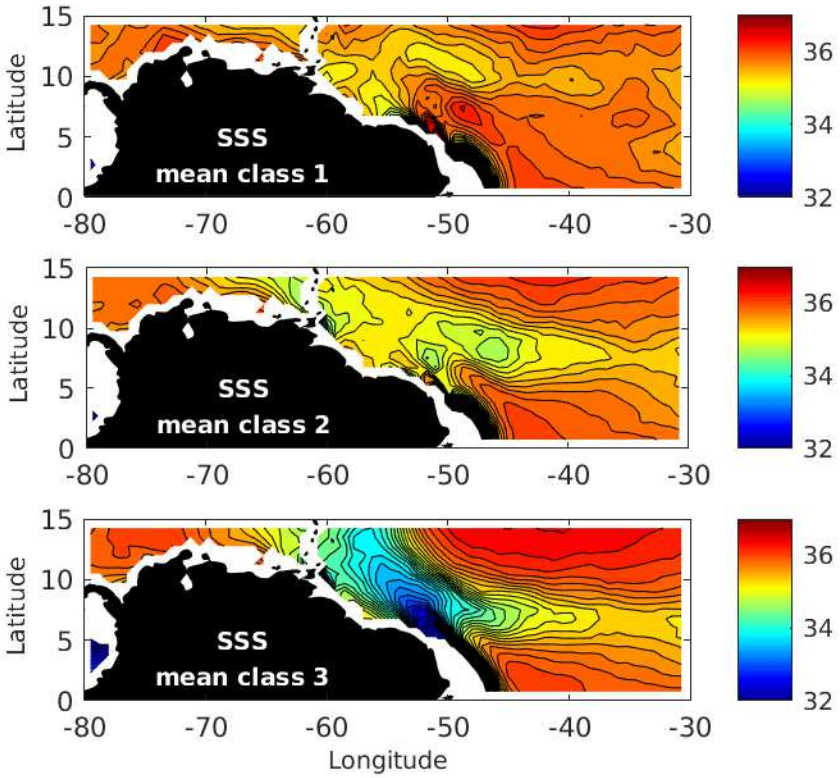


Figure 9: Same as Figure 7 but for the SSS classification.

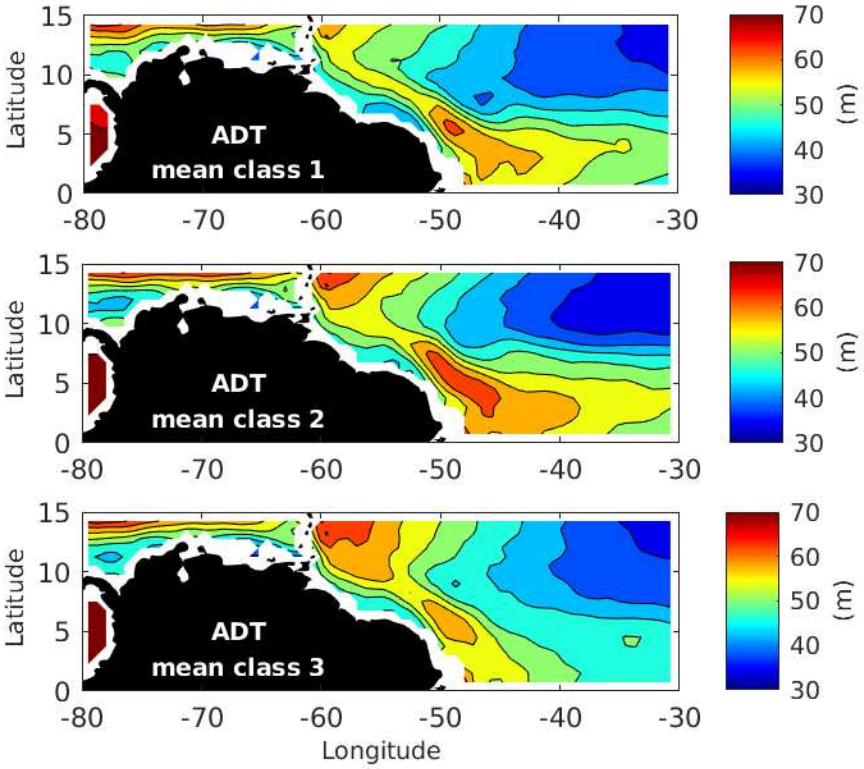


Figure 10: Same as Figure 7 but for the ADT classification. Units are m.

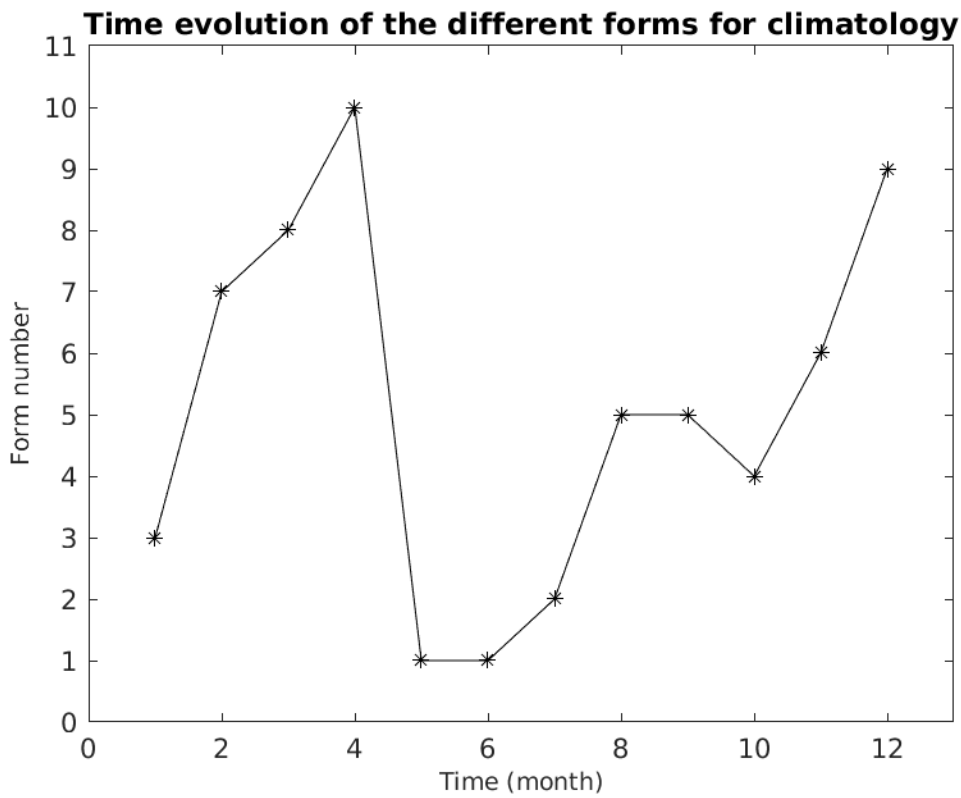


Figure 11: Time evolution over the climatological months of the different strong forms obtained after the SOM+ HAC classification on the variable quintet (SST, SSS, ADT, τx, τy).

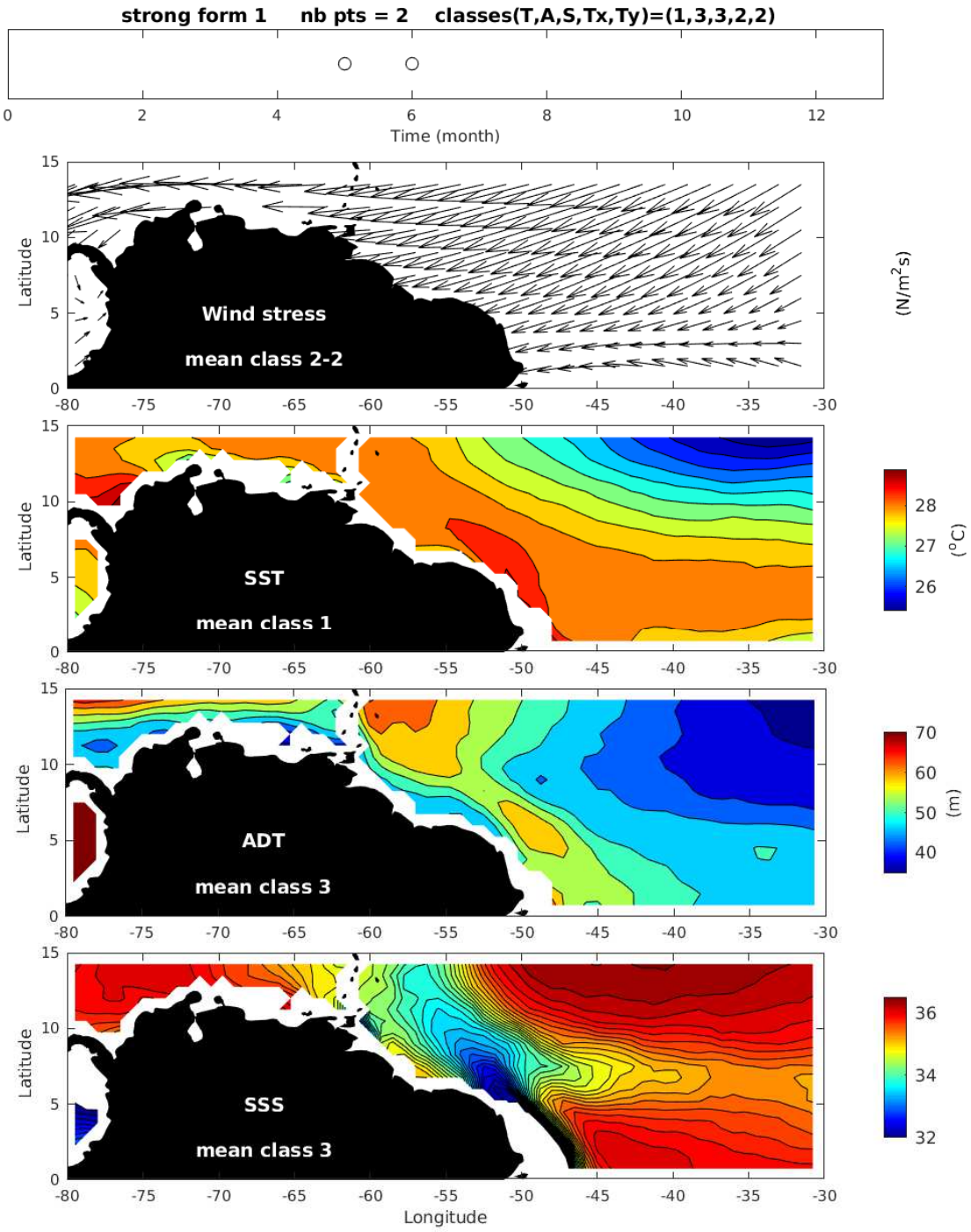


Figure 12: Mean characteristics of the different oceanic variables: SST, ADT, SSS and wind stress climatological components for strong form Number 1 in May-June. Units are (°C) for SST, (m) for ADT and (N/m²s) for wind stress.

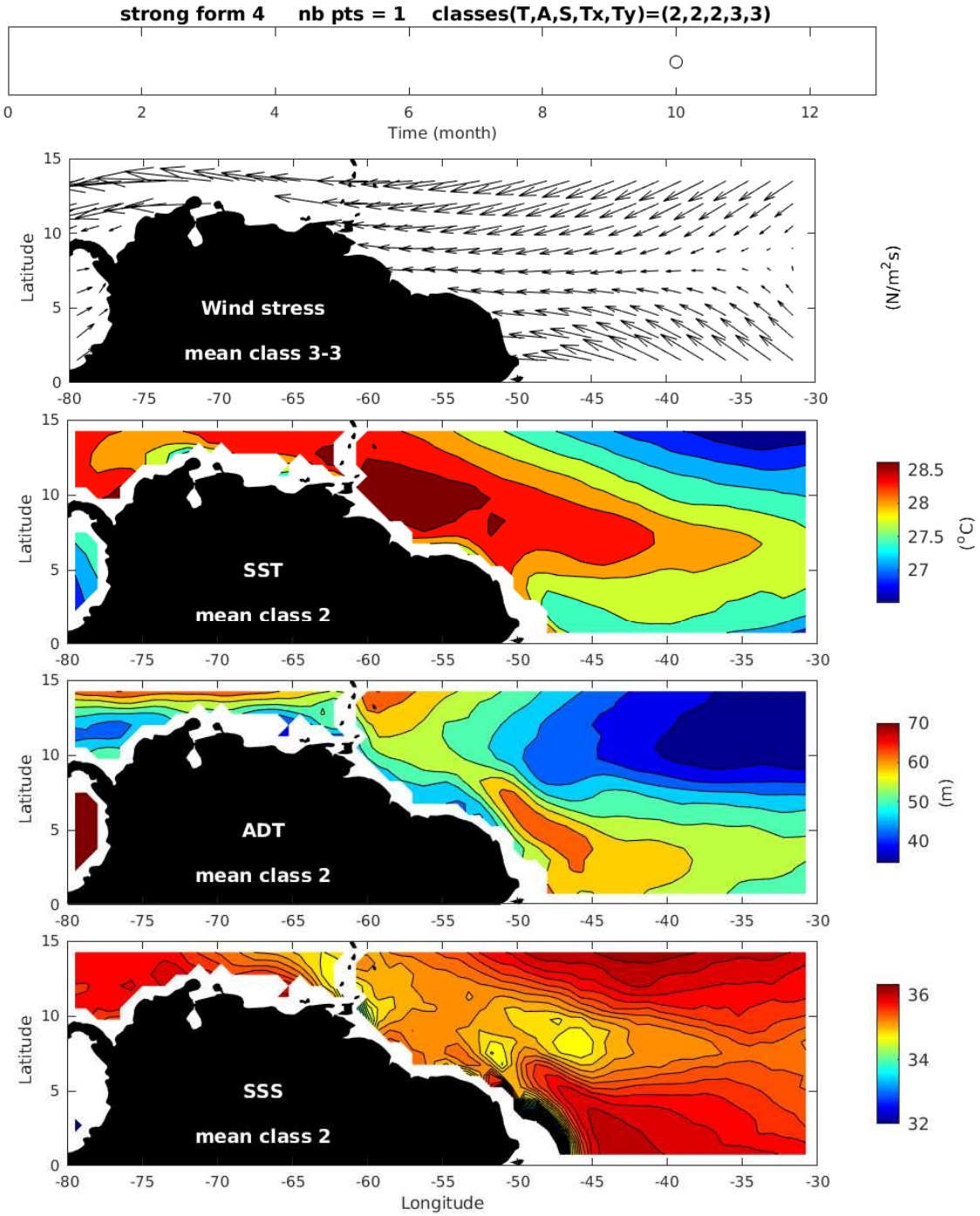


Figure 13: same as Figure 12 but for form Number 4.

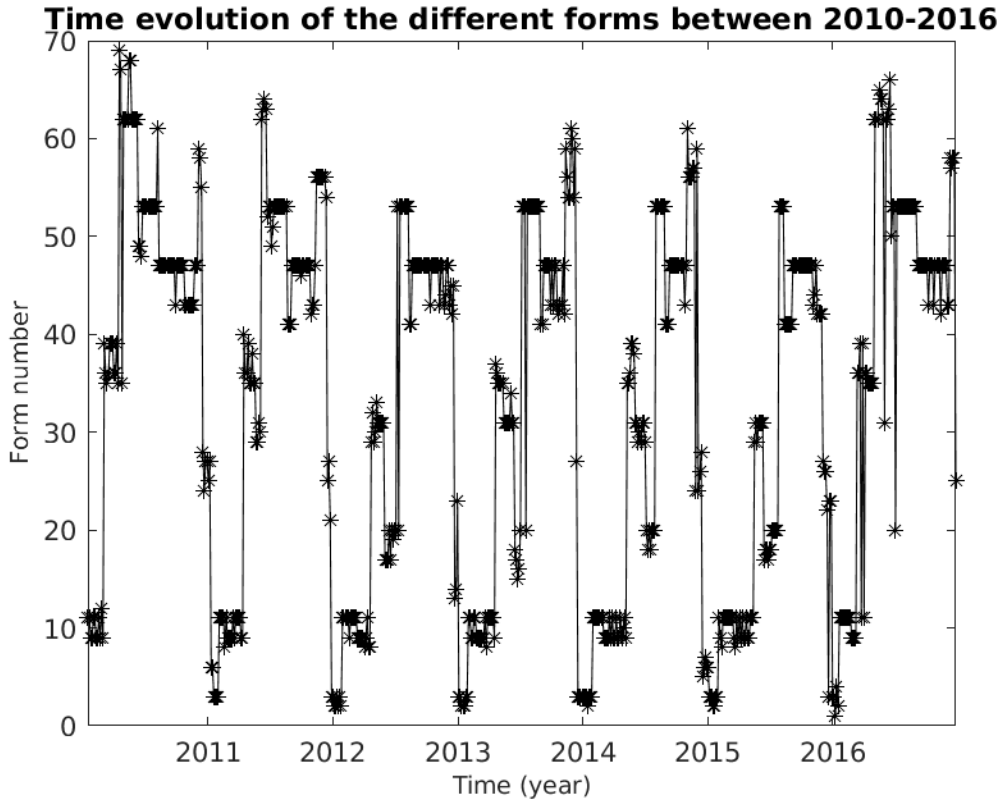


Figure 14: Time evolution over the 2010-2016 period of the different strong forms obtained after the SOM+HAC classification on the variable quintet (SST, SSS, ADT, τx, τy).

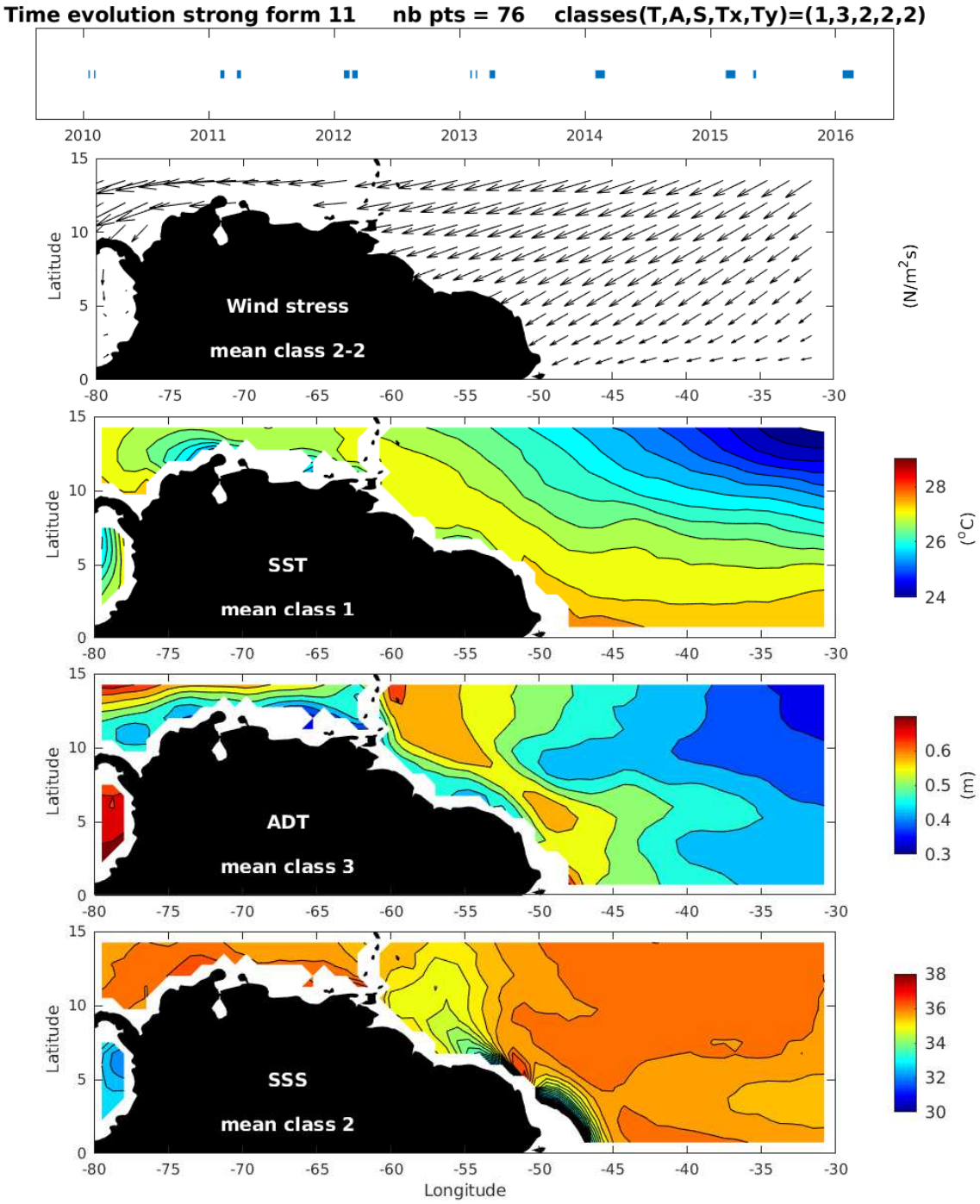


Figure 15: Mean characteristics of the different oceanic variables: SST, ADT, SSS and wind stress components for strong form Number 11 occurring during boreal spring every year. Units are (°C) for SST, (m) for ADT and (N/m^{2}s) for wind stress.

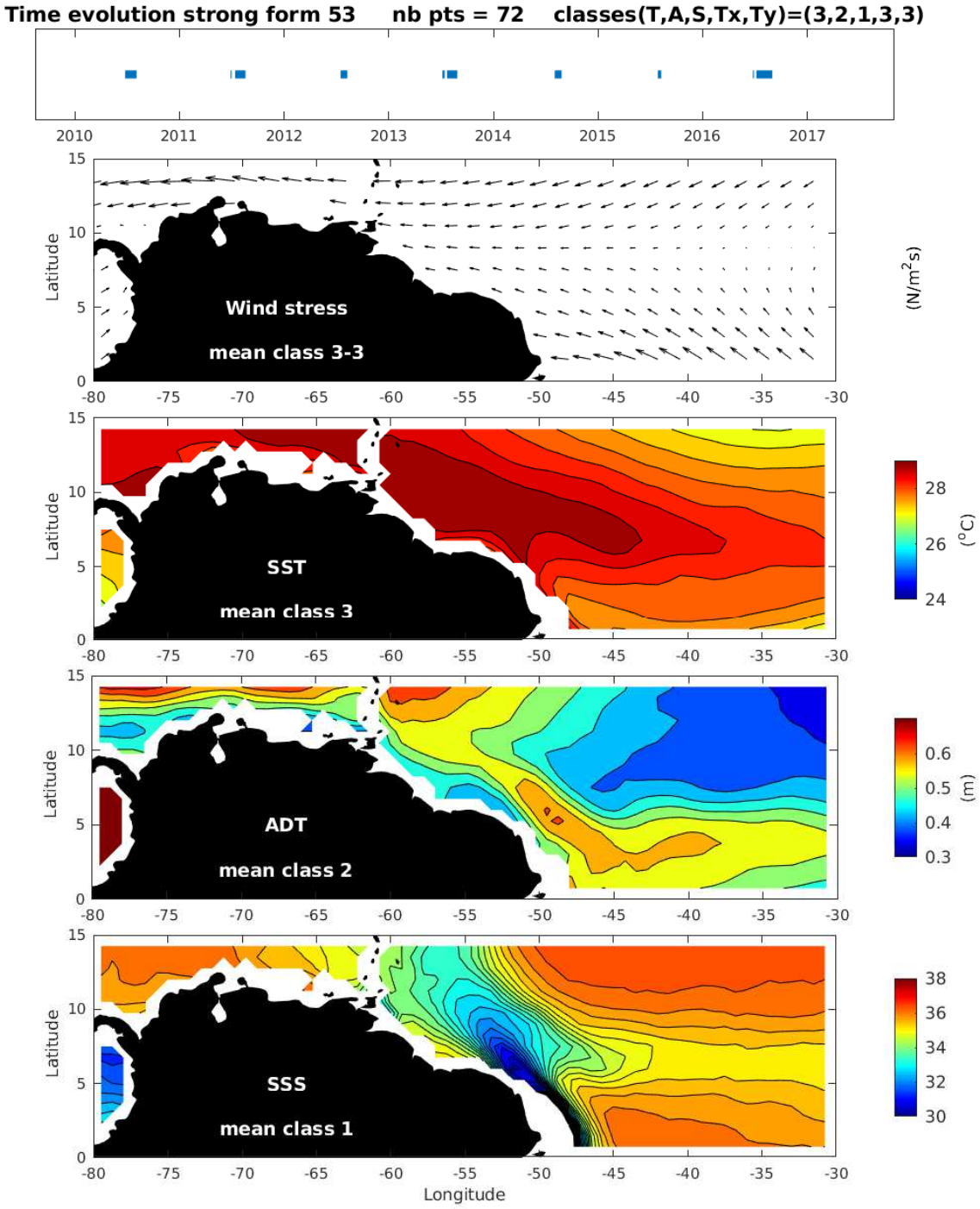


Figure 16: Same as Figure 15 but for strong form Number 53 occurring during boreal fall every year.

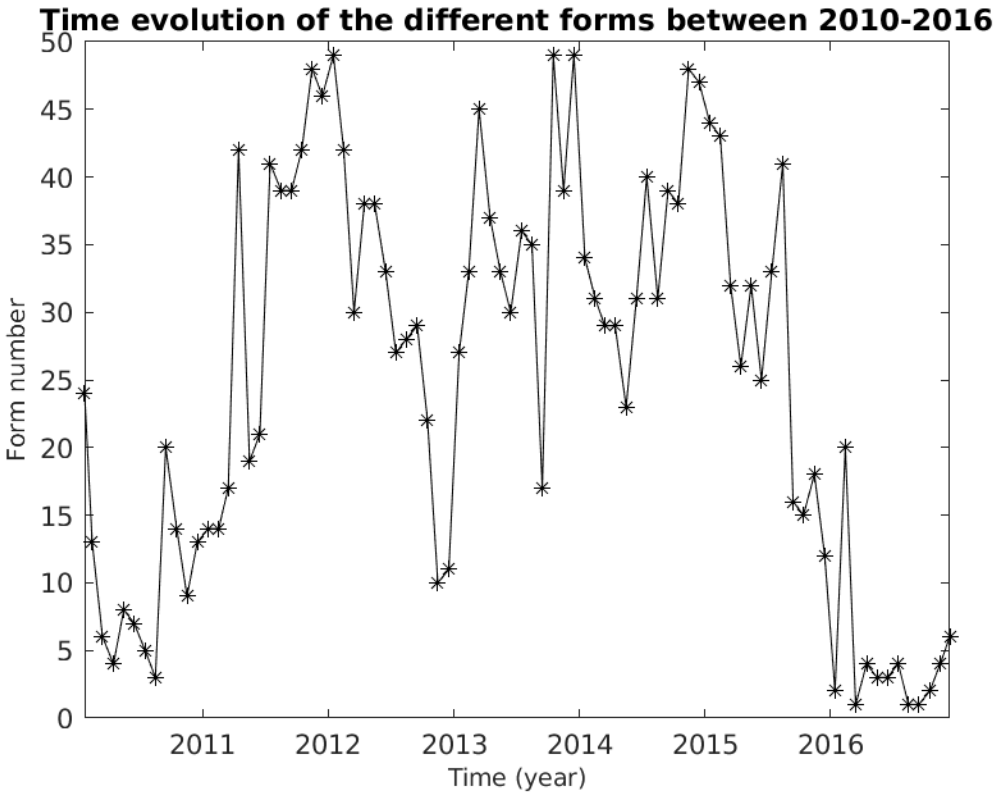
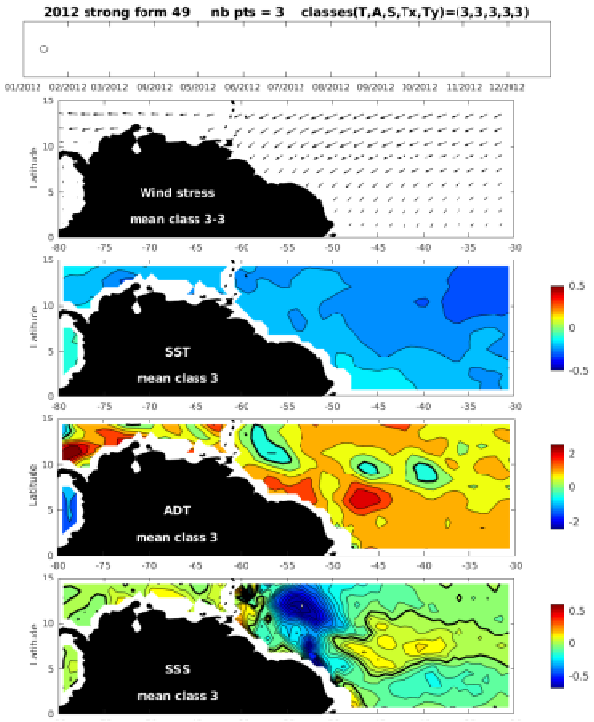


Figure 17: Time evolution over the 2010-2016 period of the different strong forms obtained after the SOM+HAC classification on the variable quintet (SST, SSS, ADT, τx, τy). The seasonal cycle has been removed from all time series. This can be considered as a "year-to-year" variability.

Boreal winter 2012



Boreal winter 2016

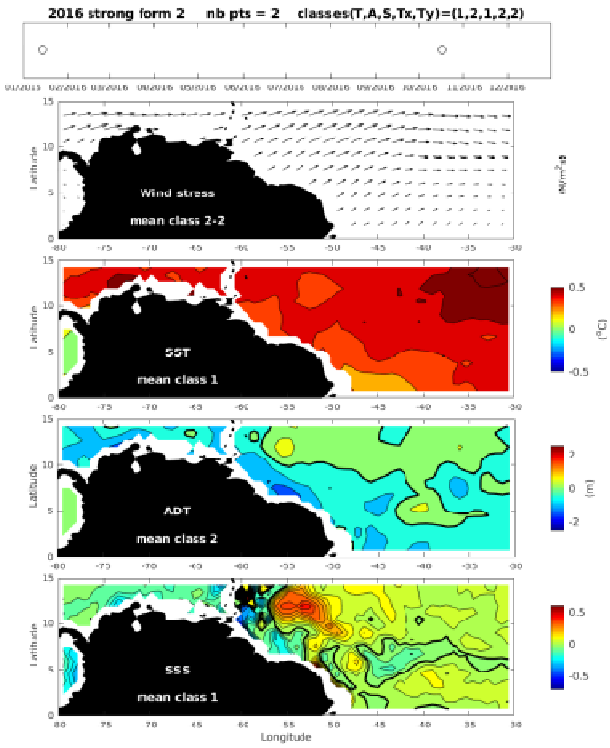


Figure 18: Mean characteristics of the different oceanic variables SST, ADT, SSS and wind stress "interannual" components referred to their climatological averages and for the strong forms occurring during boreal winter 2012 (up) and boreal winter 2016 (bottom). The zero lines are in bold. Units are (°C) for SST, (m) for ADT and (N/m^{2}s) for wind stress.

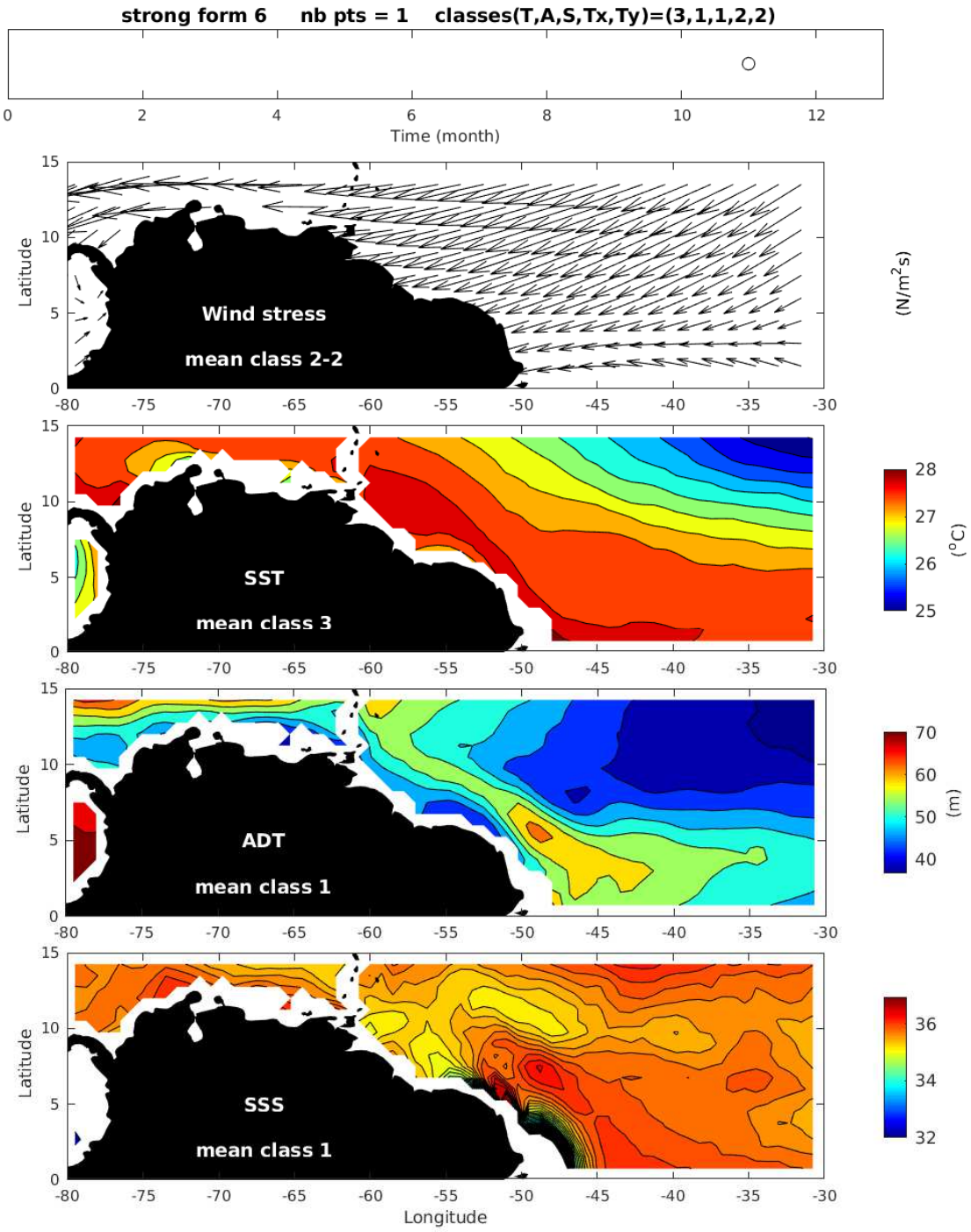


Figure 19: Mean characteristics of the different oceanic variables SST, ADT, SSS, and wind stresses climatological components for strong form Number 6 occurring in November. Units are (°C) for SST, (m) for ADT and (N/m^{2}s) for wind stress.