Identification of the North Brazil Current through spatial motifs in fixed time slices

Ricardo Buçard de Castro[1], Vinicius Correia Monteiro[1], Rafaelli Coutinho[1], Heraldo Borges[1], Eduardo Ogasawara[1]

[1]Federal Center for Technological Education of Rio de Janeiro - CEFET/RJ

ricardo.castro@aluno.cefet-rj.br,vinicius.monteiro@aluno.cefet-rj.br

rafaelli.coutinho@cefet-rj.br,heraldo.borges@cefet-rj.br,eogasawara@ieee.org

Abstract. *Spatial-temporal motif analyzes can provide insights into the data. It can also be particularly interesting to analyze spatial patterns by fixing a time slice. One variable in which the space and time relationship is present is the Sea Surface Temperature (SST). SST is related to several natural phenomena that severely impact the lives of millions of people. The SST can also be analyzed in a time slice, seeking to discover the spatial relationships that may reflect sea currents' behavior at a given moment. This work evaluates a method for identifying motifs in spatial-temporal series by comparing the occurrences of these motifs with previously mapped sea currents. As a proof of concept, the objective is to identify the North Brazil Current. This current is affected by the Intertropical Convergence Zone, identified in the extensive bibliography as responsible for the droughts in northeast Brazil. For this, after the discovery of spatial-temporal motifs, a linear regression of each found motif is performed. The angle of this regression is then compared to the current presented at the Windy platform. The proposed method achieved 70% of hits within the ten best-ranked motifs.*

Resumo. *As análises de motifs espaço-temporais podem fornecer intuições sobre os dados. Também pode ser particularmente interessante analisar os padrões espaciais fixando-se uma fatia temporal. Uma variável na qual a relação espaço e tempo é presente é a da Temperatura da Superfície do Mar (TSM). A TSM está relacionada aos diversos fenômenos naturais que impactam severamente a vida de milhões de pessoas. A TSM também pode ser analisada em uma fatia de tempo, procurando-se descobrir as relações espaciais, que podem refletir o comportamento das correntes marítimas em um determinado momento. Este trabalho avalia um método de identificação de motifs em séries espaço-temporais por meio da comparação das ocorrências destes motifs com correntes marítimas previamente mapeadas. Como prova de conceito, o objetivo é ser capaz de identificar a Corrente Norte Brasil, corrente que é afetada pela Zona de Convergência Intertropical, identificada em ampla bibliografia como responsável pelas secas no nordeste do Brasil. Para isso, após a descoberta de motifs espaço-temporais, é feita uma regressão linear de cada motif encontrado. O ângulo dessa regressão é então comparado com o ângulo no qual a corrente está se movimentando de acordo com a plataforma Windy. O método proposto foi capaz de atingir resultados de 70% de acerto dentro dos dez motifs mais bem ranqueados.*

1. Introduction

Several areas search for patterns to understand specific behaviors of the phenomena they study, such as weather forecasting, wind generation, and image recognition. It is important to emphasize that each phenomenon studied has its particular character. It is necessary to analyze their relationships in time and space in some cases. An example would be the prediction of the Sea Surface Temperature (SST) through data obtained by several geolocated sensors [Salles et al., 2016].

Spatial-temporal motif analysis provides some insights into the data. From them, we observe the same pattern repeating itself in space and time. For example, the discovery of motifs in seismic *datasets* brings relationships with horizons present in the subsoil [Borges et al., 2020b]. It can also be particularly interesting to analyze spatial patterns by fixing a time slice. An example would be the SST, which can bring information about the sea currents when analyzed in a time slice. In this article, the North Brazil Current (NBC) patterns are analyzed using the TSM *dataset* obtained from the *National Oceanic and Atmospheric Administration* (NOAA) [NOAA, 2022].

Therefore, the *Combined Series Approach* (CSA) [Borges et al., 2020a] method was used to discover spatial motifs in a time slice. The discovered motifs were confronted with the angulation of sea currents movement of the *Wind map & weather forecast* (Windy) [Windy, 2022]. From the exact latitude and longitude positions of the occurrence of the motif, a linear regression of the occurrences was performed to obtain the angulation of these patterns. Such angulation was compared with the angles reported in Windy. It was possible to verify which motifs are associated with the real behavior of the NBC current. The proposed method could associate 70% of the top-10 discovered motifs to the currents.

This work is divided into six sections. In the 2 Section, the spatial-temporal motifs are defined. In the 3 section, the related works are presented. Section 4 presents the proposed method and the *dataset* used. Section 5 addresses the results achieved. Finally, Section 6 makes the final remarks.

2. Spatial-temporal motifs

A spatial-temporal series can be described as a pair (t, p), where a time series t is associated with a spatial position p [Shekhar et al., 2015]. This position in space can be of different types, such as geographic coordinates or any other reference representing the location where the data were observed. If the position varies with time, it is a trajectory spatial-temporal series. Otherwise, it is a permanent spatial-temporal series [Borges et al., 2020b].

Given a sequence q and a time series t, q is a motif of t with support σ, if and only if q is included in t at least σ times. The length of a q motif ($|q|$) is called the word size. Formally, given a sequence q and a time series t, where $W = sw_{|q|}(t)$, *motif(q,t,σ)* $\iff \exists R \subseteq W$, $(|R| \geq \sigma)$, such that $\forall w_i \in R$, $w_i = q$ [Mueen, 2014].

When analyzing a set of spatial-temporal series D, it can be seen that there are also patterns that are frequently not only in time but also in space. These patterns are known as *spatial-temporal motifs*. One can formalize a spatial-temporal motif as a subsequence q that occurs at least σ times in D and occurs in at least κ different close spatial-temporal series, where σ and κ are two support values such that $\sigma \leq \kappa$ [Borges et al., 2020b]. The process for discovering *spatial-temporal motifs* is composed of five steps: (1) normalization and indexing; (2) partition of the spatial-temporal series; (3) combination of blocks and discovery of motifs; (4) aggregation and evaluation of constraints; (5) ranking of found motifs.

In Step 1, the *dataset* is normalized to z-score and indexed via *Symbolic Aggregation Approximation* (SAX) [Lin et al., 2007] of size a. In Step 2 the spatial-temporal series are separated into blocks (B) space-time: temporal block size (tt) and spatial block size (te). In Step 3, all the k sequences within a block are combined into a single time series (CS). Thus, CS is the concatenation of the sequences within the block $b_{i,j}$. Formally, $cs = q_1||\cdots||q_k$ and $|cs| = te \cdot tt$. Then, a traditional time series motif discovery algorithm is applied.

In Step 4, the motifs found in each block are evaluated by two constraints: (i) whether the number of occurrences of the pattern is equal to or greater than σ; (ii) whether the number of spatial-temporal series is equal to or greater than κ. Finally, in Step 5, we seek to distinguish the motifs through a ranking. Such ranking considers the entropy of the motif, the number of motifs occurrences, and the distance between motifs.

3. Related Works

Both works involving spatial-temporal motifs and those based on pattern discovery using NOAA *datasets* were analyzed. Regarding the discovery of motifs, there are several traditional approaches to discovering motifs in time series [Torkamani and Lohweg, 2017]. They can be divided into exact methods [Jiang et al., 2008; Mueen et al., 2009] and approximate methods [Chiu et al., 2003; Lin et al., 2007].

Du et al. [2009] analyze the trajectory of financial data in a state-space model of the companies and corporations to which these data belong, analyzing motifs in the trajectory of company data. Likewise, Oates et al. [2013] analyzes trajectory data from moving objects. In both cases, the approaches are significantly different from the CSA [Borges et al., 2020b]. The CSA assumes that the analyzed data are collected in fixed positions, unlike the others, which focus on studying trajectories.

Regarding works using NOAA *datasets*, one can find mainly those focused on attesting to the accuracy of the data provided by the agency [Sakaida and Kawamura, 1992; Huang et al., 2021] and validating the data [Li et al., 2001]. Sakaida and Kawamura [1992] review linear (MCSST) and nonlinear (CPSST) regression equation algorithms that estimate SSTs in the oceans around Japan. Huang et al. [2021] analyzes several NOAA SST products, comparing the data with readings from buoys and floating sensors. Finally, Li et al. [2001] validates SST data from two NOAA satellites, using two different SST algorithms, one nonlinear (NLSST) and one multichannel (MCLSST) in the Gulf of Mexico, Northeast and Southeast USA, and Great Lakes.

In addition to these, Xylogiannopoulos et al. [2019] propose the use of *big data* analysis to find multivariate motifs in diverse climatic data for a location in a given period. This process differs from the one presented in this work, both in terms of the type of motif sought and the type of data analyzed. However, it points out the importance of studying and analyzing motifs in climatic data.

4. Method

This work aims to detect NBC behavior by comparing Windy's angulation data with the motifs found in NOAA's *dataset* using CSA. Figure 1 summarizes the proposed method. The process begins with the SST data extract, transform, and load (ETL) steps. The chosen NOAA *dataset* was *OI SST V2 High Resolution Dataset*[1].

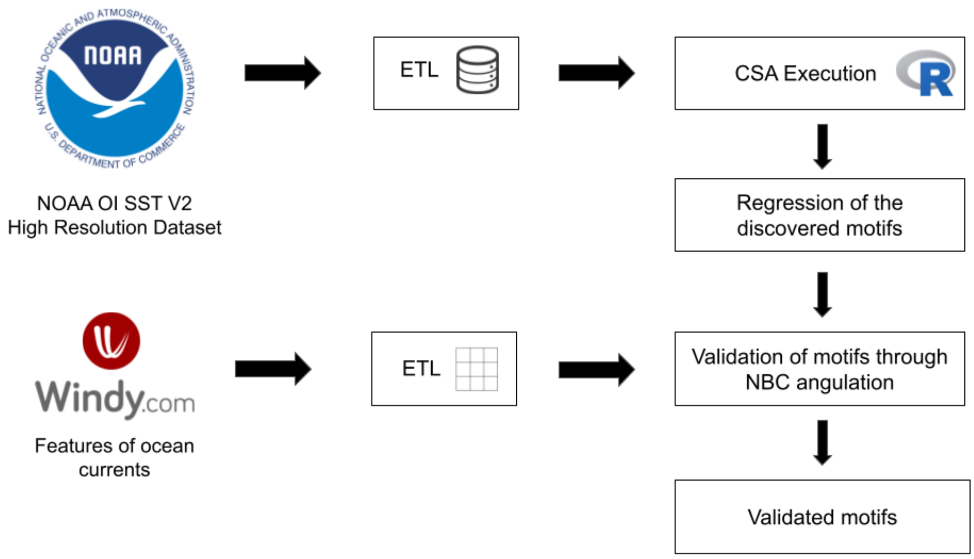

Figure 1. Proposed process

The *dataset* presents a 1440 x 720 dimension grid covering latitudes between 89.875S to 89.875N and longitudes between 0.125L to 359.875L. Both latitudes and longitudes vary by 0.25 at each position on the grid. This *dataset* has daily SST data from 1981 to 2021. With this *dataset* you can discover patterns and compare them to Windy's real-time data.

Figure 2 contains the entire geographic space that the NOAA *dataset* covers and the temperature readings for October 4, 2021, using as a base a color palette ranging from increasing purple-pink with transitory colors such as blue, green, orange and yellow. Temperature readings are in degrees Celsius and range from -1.8°C to 33.32°C.

On the SST set was normalization by z-score and SAX indexing ($a = 11$). Next, CSA is applied to discover and rank motifs using the R package STMotif [2]. Since the objective is to analyze a time slice, the latitude dimension was associated with time, and the longitude dimension was associated with space. In this context, the following parameterization was used for the CSA: $w = 4$, $te = 20$, $tt = 10$, $\sigma = 3$ and $\kappa = 2$. The occurrences of each motif underwent a linear regression to characterize a direction. The angle of the discovered top-k motifs was compared to the currents lifted on the Windy on a given day.

In Windy's ETL process, the data of the analyzed stream is enriched. An average angulation of sectors in the region of interest is produced. These data were adjusted

[1] https://www.psl.noaa.gov/data/gridded/data.noaa.oisst.v2.highres.html
[2] https://cran.r-project.org/web/packages/STMotif/index.html

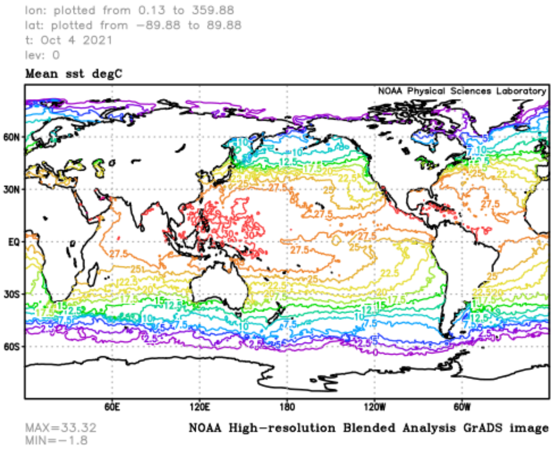

Figure 2. NOAA SST Preview for October 4, 2021

to characterize the direction of the currents. It enables the comparison with the motifs discovered within the latitude and longitude ranges in which the current is found. Finally, tolerance is adopted as the individual current measurements are discretized in ten-degree intervals. It is based on a confidence interval between the difference in the angulation of the discovered patterns and the average value of the current in a certain interval.

5. Experimental Evaluation

We only have real-time Windy angulation data available. We collected the angulation data for October 4, 2021, as a Gold-standard to evaluate our proof of concept experiment. The objective was to evaluate whether the arrangement of the motifs occurs in agreement with the currents observed. One of the challenges faced is the perception that some currents are underwater; therefore, this type of stream may not be able to be identified by the SST *dataset*.

In Figure 3, it is possible to observe the currents that are present between the north of Brazil and Africa. The colors represent the speed at which the water is being moved. Its intensity decreases through a palette (white, blue, pink, orange, green, and gray). The NBC corresponds to the pink, blue and white areas that start above Fortaleza and continue to a little above Amapá. In this region, retroflection occurs, and we observe a blue band in the opposite direction, which corresponds to the North Equatorial Countercurrent.

Notably, the NBC is the current that stands out the most for the intensity of the movement compared to others in the same region. Furthermore, the NBC is currently affected by the Intertropical Convergence Zone (ITCZ) and, therefore, relevant to the rainfall distributions in the north of the Brazilian Northeast, of great value for study.

A *dataset* clipping was made with fifteen degrees of longitude by seven degrees of latitude to analyze the desired data. It corresponds to the latitudes between 0.125N and 7.375N and the longitudes between 48.875W and 34.125W. In Figure 4, it is possible to visualize the stretch of the Atlantic Ocean that represents the area that was used for the experiments.

The direction of current movements was disregarded to compare the Windy *dataset* angle with the linear regression angle. We took in taking into account only the angulation.

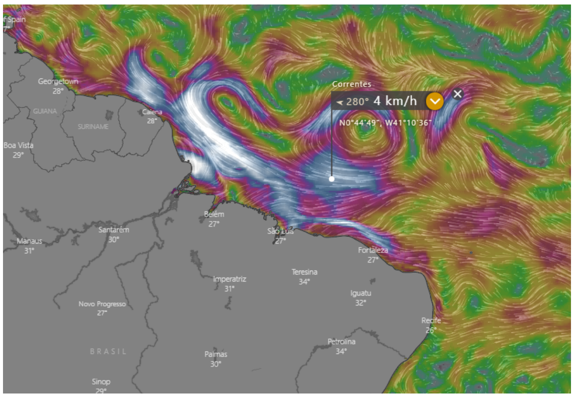

Figure 3. Movement of sea currents affected by the ITCZ in Brazil

Figure 4. Region analyzed: 0.125N to 7.375N (lat) and 48.875O to 34.125O (long)

Finally, these angles are compared with the angles of the discovered top-k motifs.

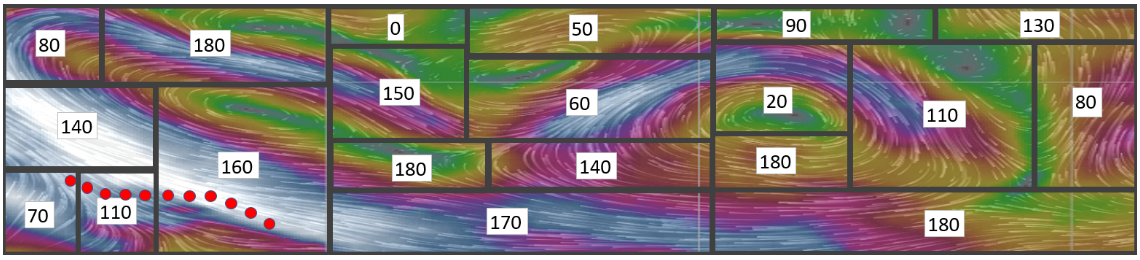

Figure 5. Overlay of NBC angulation with Windy sea current movement data on October 4, 2021. The red marking indicates occurrences of the $aaab$ motif in the region

From the lists of motifs discovered in the study region, each mapped angle is compared with the motif angle obtained through linear regression for the top-15 discovered motifs. To exemplify the execution of this step, we calculate the linear regression of the $aaab$ motif. In Figure 5, it is possible to observe the occurrences of the motif $aaab$ in space through the red dotted markings. The markings bring a line represented by the equation $y = -0.31x + 9.7$. Furthermore, from the arctangent of the angular coefficient,

-0.31, an angle of 162 was obtained. As a result, the motif $aaab$ was validated since the value felt within the confidence interval of 160 degrees. The procedure was done for the remaining top-15 motifs. Therefore, a list of nine validated motifs was obtained, namely: $kjih$, $cbaa$, $aaaa$, $kkkk$, $baaa$, $kkji$, $cbba$, $gfee$ and $aaab$. We got a 70% validation rate for the top ten (top-10) and 60% validation for the top fifteen (top-15).

6. Conclusion

In this work, a methodology is proposed for the identification of sea currents from the discovery of spatial motifs from a time slice of the SST *dataset*. SST data were analyzed using the CSA algorithm to identify spatial motifs in a time slice. Due to the fixed time setting, the longitude was associated with the space dimension. Additionally, the latitude was associated with the time dimension in the CSA. For each motif, a linear regression of its occurrences and the angle of this regression were calculated. The angle of occurrences of each motif was compared with the average angle of movement of sea currents obtained in Windy.

From the proof of concept study, the performance of the proposed method indicated that 70% of the ten best-ranked motifs presented a direction compatible with the analyzed currents, reaching the objectives of this work. The results achieved open space to study the discovery of spatial-temporal motifs formed from compositions of spatial patterns that can be repeated in time.

Acknowledgment

The authors would like to thank CAPES (finance code 001), CNPq, and FAPERJ for partially funding this paper.

References

Heraldo Borges, Amin Bazaz, Esther Pacitti, and Eduardo Ogasawara. STMotif: Discovery of Motifs in Spatial-Time Series. Technical report, https://cran.r-project.org/web/packages/STMotif/index.html, 2020a.

Heraldo Borges, Murillo Dutra, Amin Bazaz, Rafaelli Coutinho, Fábio Perosi, Fábio Porto, Florent Masseglia, Esther Pacitti, and Eduardo Ogasawara. Spatial-time motifs discovery. *Intelligent Data Analysis*, 24(5):1121–1140, 2020b.

Bill Chiu, Eamonn Keogh, and Stefano Lonardi. Probabilistic discovery of time series motifs. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 493–498, 2003.

Xiaoxi Du, Ruoming Jin, Liang Ding, Victor E Lee, and John H Thornton Jr. Migration motif: a spatial-temporal pattern mining approach for financial markets. In *Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144, 2009.

Boyin Huang, Chunying Liu, Eric Freeman, Garrett Graham, Tom Smith, and Huai-Min Zhang. Assessment and Intercomparison of NOAA Daily Optimum Interpolation Sea Surface Temperature (DOISST) Version 2.1. *Journal of Climate*, 34(18):7421–7441, 2021.

Tao Jiang, Yucai Feng, Bin Zhang, Jie Shi, and Yuanzhen Wang. Finding motifs of financial data streams in real time. In *International Symposium on Intelligence Computation and Applications*, pages 546–555. Springer, 2008.

Xiaofeng Li, W Pichel, P Clemente-Colon, V Krasnopolsky, and J Sapper. Validation of coastal sea and lake surface temperature measurements derived from NOAA/AVHRR data. *International Journal of Remote Sensing*, 22(7):1285–1303, 2001.

Jessica Lin, Eamonn Keogh, Li Wei, and Stefano Lonardi. Experiencing SAX: a novel symbolic representation of time series. *Data Mining and knowledge discovery*, 15(2): 107–144, 2007.

Abdullah Mueen. Time series motif discovery: dimensions and applications. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 4(2):152–159, 2014.

Abdullah Mueen, Eamonn Keogh, Qiang Zhu, Sydney Cash, and Brandon Westover. Exact discovery of time series motifs. In *Proceedings of the 2009 SIAM international conference on data mining*, pages 473–484. SIAM, 2009.

NOAA. National Oceanic and Atmospheric Administration. Technical report, https://www.noaa.gov/, 2022.

Tim Oates, Arnold P Boedihardjo, Jessica Lin, Crystal Chen, Susan Frankenstein, and Sunil Gandhi. Motif discovery in spatial trajectories using grammar inference. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, pages 1465–1468, 2013.

Futoki Sakaida and Hiroshi Kawamura. Accuracies of NOAA/NESDIS sea surface temperature estimation technique in the oceans around Japan. *Journal of Oceanography*, 48(3):345–351, 1992.

R. Salles, P. Mattos, A.-M.D. Iorgulescu, E. Bezerra, L. Lima, and E. Ogasawara. Evaluating temporal aggregation for predicting the sea surface temperature of the Atlantic Ocean. *Ecological Informatics*, 36:94–105, 2016.

Shashi Shekhar, Steven K Feiner, and Walid G Aref. Spatial computing. *Communications of the ACM*, 59(1):72–81, 2015.

S. Torkamani and V. Lohweg. Survey on time series motif discovery. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 7(2), 2017.

Windy. Windy: Wind map & weather forecast. Technical report, https://www.windy.com/, 2022.

Konstantinos Xylogiannopoulos, Panagiotis Karampelas, and Reda Alhajj. Multivariate motif detection in local weather big data. In *2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 749–756. IEEE, 2019.