

Visualization of Geospatial Time Series from Environmental Modeling Output

Patrick Köthür¹, Mike Sips¹, Julian Kuhlmann¹, and Doris Dransch¹

¹ Helmholtz Centre Potsdam – GFZ German Research Centre for Geosciences, Germany

Abstract

Environmental models produce geospatial time series containing many spatio-temporal patterns. Scientists need to understand these patterns to analyze the behavior of the simulated environmental systems. We combine clustering and visualization to generate an intuitive visual summary of geospatial time series that captures the data's prominent spatio-temporal information. As a first step, we evaluated our approach with well-understood observational data. Our visualization depicted all prominent features of these data suggesting that our method is readily applicable to environmental model output.

Categories and Subject Descriptors (according to ACM CCS): I.3.8 [Computer Graphics]: Applications—
I.6.6 [Simulation and Modeling]: Simulation Output Analysis—

1. Introduction

The aim of environmental simulation modeling is to study or predict the behavior of System Earth, e.g., ocean circulation, landslides, flood inundation, or earthquake induced ground motion. Simulations of such real-world systems produce geospatial time series.

To study the behavior of environmental systems, scientists need to understand the spatio-temporal patterns hidden in the geospatial time series. Visualization has proven to be an effective approach to gain insight into time series [AMST11]. In this paper, we focus on geospatial time series where each time step describes a spatial configuration represented by a 2D grid of scalar values. Two prominent techniques for visualizing geospatial time series are small multiples and map animation [Tuf90, Tuf01, AAG03]. However, these techniques are only appropriate for rather small geospatial time series. Small multiples show only a very limited number of time steps because of screen space limitations. For map animations, the user may have difficulties to perceive important geospatial patterns in a large stream of images due to change blindness [TMB02, FGB11].

In this paper, we introduce an intuitive visual summary of geospatial time series that depicts the data's prominent spatio-temporal patterns in a compact visualization. This visual summary is based on clustering that reduces the numerous spatial configurations of a time series to a small number

of representative clusters. We use the cluster labels to segment the geospatial time series into blocks of similar spatial configurations. These analytical results are visually encoded in two components: a spatial configuration view that depicts extracted spatial patterns, and a sequence view that displays their occurrence over time. A first evaluation with well-understood observational data shows that our approach captures the data's prominent spatio-temporal information.

2. Related work

As related work we briefly discuss domain specific spatio-temporal clustering in the geosciences and the combination of clustering and visualization for time series analysis.

The aim of clustering is to divide data into groups of similar objects. The identified clusters provide a condensed description of the original data (see [JMF99] or [HK06] for further readings). Within the geosciences, meteorologists apply clustering to geospatial time series. The clustering serves as an automated analysis step to extract prominent spatial configurations of the atmospheric circulation [Hut96, Hut00, Hor10, RVLS10]. These approaches do not combine their spatio-temporal clustering results with interactive visualization for an in-depth exploration of spatio-temporal patterns.

A combination of clustering and interactive visualization to facilitate the exploration of time series data is utilized in many visualization approaches [vWvS99, LKL05, HMJ*12].

However, these techniques do not cover spatial data. Approaches specifically addressing (geo)spatial time series focus on the comparison of spatial regions with regard to their temporal behavior [DJMK06, WS09]. In contrast to these methods, we want to capture the global spatio-temporal patterns in the data. Bruckner and Möller [BM10] also focus on global patterns, but their visualization is tailored to visual effects design; a different application problem.

3. Background

Our application scenario is ocean modeling. We collaborated with several geoscientists and adopted a user- and task-centered approach [DKS*10] to derive a thorough understanding of the domain problem.

3.1. Ocean modeling

Ocean modeling serves two purposes. First, it is a way of evaluating existing theories about different processes in the ocean by comparing the model output to measured data. Second, it is a way of performing experiments that scientists cannot conduct in the real world [Ste05].

Ocean models produce different kinds of time series data (1D, 2D gridded, volumetric). We focus on time series of regularly structured 2D grids.

3.2. Task analysis

We distinguish two main objectives requiring visualization in ocean modeling: first, the debugging and refinement of the model, and second, to gain scientific insight about the system under study. Model debugging involves, among others, the identification of outlying spatio-temporal patterns. If scientists cannot explain a specific pattern with geophysical laws or other expert knowledge, the spatio-temporal pattern probably results from erroneous model code and demands debugging and refinement.

Gaining scientific insight from ocean models requires the detection of prominent spatio-temporal patterns. Scientists often describe these patterns as temporary, possibly recurring, regional outliers. A prior specification of what constitutes a prominent spatio-temporal pattern is generally not possible because their definition is highly dependent on the spatio-temporal context. Scientists need to study many characteristics, such as geographic location, spatial extent, patterns in neighboring geographic regions, nature of emergence and disappearance of patterns, duration, reoccurrence, etc., to decide whether a specific spatio-temporal pattern is prominent.

Based on our task analysis, we identified the following design requirements.

DR1 Extract prominent spatio-temporal patterns.

DR2 Present the prominent spatio-temporal patterns to the user.

DR3 Preserve the spatio-temporal context.

DR4 Point to potential outlying patterns.

DR5 Point to recurring patterns.

DR6 Allow for interactive exploration of spatio-temporal patterns.

The approach introduced in this paper supports design requirements 1 through 4. Design requirements 5 and 6 will be future work.

4. Our approach

The spatio-temporal context of environmental simulation model output plays an important role (see Section 3.2). Therefore, only very limited aggregation of the spatial and temporal dimensions is feasible. For example, computing moment statistics such as mean or variance for each time step (= spatial configuration) would cause a complete loss of the spatial information. Likewise, computing these statistics for each grid point over all time steps would result in a loss of the temporal information.

In our approach, we combine clustering and visualization to focus on prominent patterns while preserving spatio-temporal context information. In the first step, we cluster the time steps of an ocean model output. This reduces the time series to a small number of clusters with each cluster representing a certain type of spatial pattern occurring in the data. Hence, time steps containing similar spatial patterns are grouped together in the same cluster. In the second step, we partition the time series into blocks containing the same type of spatial pattern by utilizing the cluster labels. Finally, we present these analytical results in a combination of two different visual components to the user.

4.1. Clustering and segmentation

We apply bottom-up hierarchical clustering to all time steps of the time series. Initially, every time step forms its own cluster. A similarity measure successively merges the time steps into bigger clusters, starting with the two most similar clusters. The result is a binary tree called dendrogram. The dendrogram depicts the cluster hierarchy with all clusters being merged into one at its root. The resulting cluster hierarchy provides a full description of the data. It also facilitates top-down exploration which is important in our application scenario. Scientists do not need to specify the number of clusters in advance. They may start with a small number of clusters, i.e., a coarse description of the data, and gradually move down the hierarchy to increase the number of clusters and, hence, the level of detail. Inspired by its frequent use in image retrieval and image sequence segmentation [VRB00, KCB03, DG03, LYJ05], we apply the sum of squared errors to measure the similarity between time steps. The agglomeration method used was average linkage.

In a second step, we segment the original time series by labeling each time step with its associated cluster number. The result is a list denoting cluster affiliation of the time steps. This list successfully reduces the data to its representative spatio-temporal patterns, satisfying design requirement 1 (DR1).

4.2. Visualization

We propose a visualization that combines two views to capture the data's prominent spatio-temporal patterns (Figure 1).

The *spatial configuration view* depicts representative types of spatial configurations in the time series. We derive a representative spatial pattern for each cluster by computing an average grid of its members. We show the representative grids in separate maps and arrange them in a small multiple layout. A colored frame around each representative grid denotes its cluster affiliation. We use one of ColorBrewer's [HB03] qualitative color schemes for color coding.

The *sequence view* presents the temporal occurrence of spatial patterns in the time series. We visually encode the time series as a horizontal bar that maps the time steps to their associated clusters. Cluster affiliation is mapped to color, as shown in Figure 1 (the solid black line serves evaluation purposes, see Section 5.2). Due to spatial and temporal autocorrelation in environmental modeling data, subsequent time steps often have the same cluster affiliation. This has the effect of presenting the time series as a sequence of coherent blocks, which leads to a concise representation of the temporal information. This approach even scales to large time series. Note, scalability and the visuals have their limits when the mapping of time steps to clusters does not yield visually coherent blocks. However, these extreme data are very unlikely in environmental modeling.

We visually link the spatial configuration view and the sequence view using a consistent color scheme. Vertical arrangement of the views results in a compact display and emphasizes the visual linking through color. The combination of the two views presents an overview of the spatio-temporal patterns in the model output, meeting DR2 and DR3.

The spatial configuration view and the sequence view may be combined with further visualizations to meet other design requirements. For example, we suggest a histogram of cluster occurrences to further facilitate the identification of outliers and to satisfy DR4 (Figure 1, bottom). Small cluster sizes imply rare occurrences of the associated spatial patterns. Therefore, users may be pointed to outlying patterns by studying this histogram. Please note that it can only hint at potential outliers. Users will still have to consult the spatial configuration view and the sequence view to further analyze the spatio-temporal context.

5. Evaluation

Our application example stems from ocean modeling. We evaluated our method using well-understood observational data. In the following, we introduce the test data and present our findings.

5.1. Test data set

We use sea-level anomalies data obtained from a combination of several satellite altimeters. This altimeter product was produced by Ssalto/Duacs and distributed by Aviso, with support from Cnes (<http://www.aviso.oceanobs.com/duacs/>). The geospatial time series consists of weekly global sea-level data ranging from October 1992 to July 2009. In a preprocessing step, we subtracted a global trend and spatially rescaled the data obtaining 876 grids with a spatial resolution of 194×96 .

The test data set has well-defined characteristics that should become apparent in our visualization:

- a pronounced seasonal cycle,
- interannual variations in the Tropics (El Niño/La Niña),
- a very strong El Niño/La Niña event in 1997/98.

5.2. Results

We evaluated our visualization in collaboration with a geoscientist. We generated visual summaries of the test data for different numbers of clusters; gradually increasing the number of clusters from two to twelve. The geoscientist considered eight clusters an adequate description of the test data's prominent characteristics (Figure 1).

The periodic appearance of the clusters A, B, and C in the sequence view describes a pronounced seasonal cycle. Cluster B clearly shows a Northern Hemisphere winter/spring pattern. Negative sea-level anomalies in the Northern Hemisphere can be attributed to a low volume of the ocean induced by low temperature. Likewise, clusters A and C resemble a Northern Hemisphere summer/autumn state. For further validation, we overlaid the sequence view with an El Niño/La Niña index [NOA11] (solid black line). The interpretation of this index is straightforward. If the index exceeds the upper dotted line, we should observe an El Niño pattern (with some small delay). A La Niña pattern should follow shortly after the index undershoots the lower dotted line. Our visualization clearly correlates with the index. Clusters E, F, and G describe the outstanding 1997/98 El Niño/La Niña, and Clusters D and H represent El Niño or La Niña patterns of lower intensity. In addition, the histogram of cluster occurrences also points to the El Niño/La Niña patterns as outstanding features, since the associated clusters are rather small.

Applying our approach to well-understood observational data yielded promising results. The correlation between our

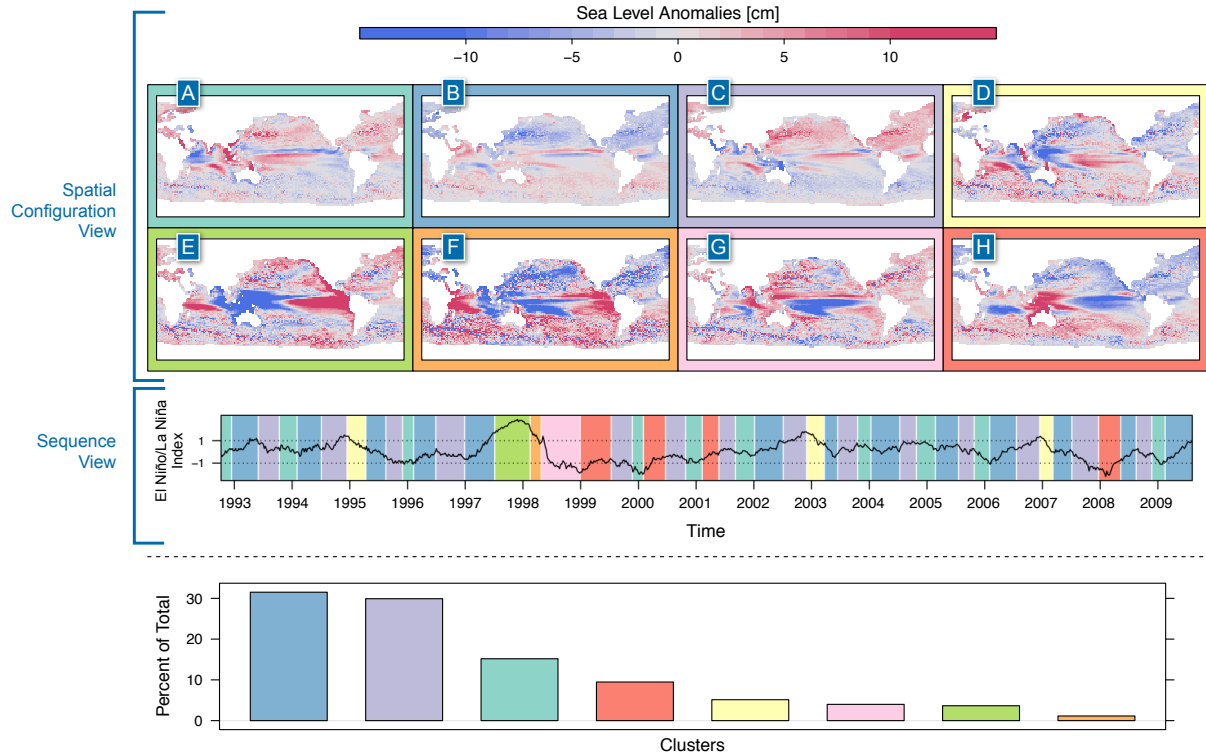


Figure 1: The result of our approach: An intuitive visual summary of geospatial time series that captures the data's prominent spatio-temporal patterns. Shown here are weekly sea-level anomalies from October 1992 to July 2009. The spatial configuration view shows a representative spatial pattern for each cluster (A to H). The sequence view depicts the occurrence of these patterns over time. A consistent color scheme visually links all views. Summer and winter states are shown in clusters A, B, and C. Clusters D, and H depict El Niño and La Niña patterns. A very strong El Niño/La Niña event in 1997/98 is captured by clusters E, F, and G. Our visualization correlates with an El Niño/La Niña index that is depicted as a solid black line in the sequence view. In addition, a histogram of cluster occurrences points to potentially outlying patterns (here: El Niño/La Niña).

visualization and an El Niño/La Niña index shows that our approach captures the prominent spatio-temporal patterns in geospatial time series (see Section 5.1) and successfully meets DR1 through DR4. This encourages the application of our method to actual environmental model output.

6. Discussion and future work

We demonstrated that our novel visualization provides a concise visual summary of prominent spatio-temporal features in geospatial time series. This is a first step towards a comprehensive visual analytics approach that meets all design requirements. Future work will facilitate interactive exploration of the spatio-temporal patterns shown in the visual summary (DR6). Interaction should enable the geoscientist to determine clustering parameters, explore the cluster hierarchy, zoom, filter, and query for detailed information.

A second focus is on further evaluation. Although recurring patterns (DR5) become apparent in our exemplary visu-

alization, there are established techniques for depicting periodicity in data (e.g., [CK98, Chu98]). We will evaluate them regarding our application. In addition, our segmentation of geospatial time series by means of clustering can be regarded as a symbolic representation of the time series. This allows the application of well-known data mining algorithms to automatically extract well-defined patterns, including periodicities.

At last, we plan to extend our approach to multi-run simulations of environmental systems. The aim will be to provide an overview of spatio-temporal patterns in multi-run data and to facilitate the exploration of input-output relations in environmental simulation models.

Acknowledgments

This work was partially supported by the German Federal Ministry for Education and Research (BMBF) via the Potsdam Research Cluster for Georisk Analysis, Environmental Change and Sustainability (PROGRESS).

References

- [AAG03] ANDRIENKO N., ANDRIENKO G., GATalsky P.: Exploratory spatio-temporal visualization: an analytical review. *Journal of Visual Languages & Computing* 14, 6 (2003), 503–541. doi:10.1016/S1045-926X(03)00046-6. 1
- [AMST11] AIGNER W., MIKSCH S., SCHUMANN H., TOMINSKI C.: *Visualization of time-oriented data*. Springer, London, 2011.
- [BM10] BRUCKNER S., MÖLLER T.: Result-Driven Exploration of Simulation Parameter Spaces for Visual Effects Design. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (2010), 1468–1476. doi:10.1109/TVCG.2010.190. 2
- [Chu98] CHUAH M.: Dynamic aggregation with circular visual designs. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis)* (1998), Wills G., Dill J., (Eds.), IEEE Computer Society, pp. 35–43. doi:10.1109/INFVIS.1998.729557. 4
- [CK98] CARLIS J. V., KONSTAN J. A.: Interactive visualization of serial periodic data. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)* (1998), ACM Press, pp. 29–38. 4
- [DG03] DENGSHENG ZHANG, GUOJUN LU: Evaluation of similarity measurement for image retrieval. In *Proceedings of the 2003 International Conference on Neural Networks and Signal Processing* (2003), vol. 2, IEEE, pp. 928–931. doi:10.1109/ICNNSP.2003.1280752. 2
- [DJMK06] DIANSHENG GUO, JIN CHEN, MACEachREN A. M., KE LIAO: A Visualization System for Space-Time and Multivariate Patterns (VIS-STAMP). *IEEE Transactions on Visualization and Computer Graphics* 12, 6 (2006), 1461–1474. doi:10.1109/TVCG.2006.84. 2
- [DKS*10] DRANSCH D., KÖTHÜR P., SCHULTE S., KLEMAN V., DOBSLAW H.: Assessing the quality of geoscientific simulation models with visual analytics methods – a design study. *International Journal of Geographical Information Science* 24, 10 (2010), 1459–1479. doi:10.1080/13658816.2010.510800. 2
- [FGB11] FISH C., GOLDSBERRY K. P., BATTERSBY S.: Change Blindness in Animated Choropleth Maps: An Empirical Study. *Cartography and Geographical Information Science* 38, 4 (2011), 350–362. doi:10.1559/15230406384350. 1
- [HB03] HARROWER M., BREWER C. A.: ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps. *The Cartographic Journal* 40, 1 (2003), 27–37. doi:10.1179/000870403235002042. 3
- [HK06] HAN J., KAMBER M.: *Data mining: Concepts and techniques*, 2 ed. Morgan Kaufmann, San Francisco, 2006. 1
- [HMJ*12] HAO M. C., MARWAH M., JANETZKO H., DAYAL U., KEIM D. A., PATNAIK D., RAMAKRISHNAN N., SHARMA R. K.: Visual exploration of frequent patterns in multivariate time series. *Information Visualization* 11, 1 (2012), 71–83. doi:10.1177/1473871611430769. 1
- [Hor10] HORENKO I.: On clustering of non-stationary meteorological time series. *Dynamics of Atmospheres and Oceans* 49, 2–3 (2010), 164–187. doi:10.1016/j.dynatmoce.2009.04.003. 1
- [Hut96] HUTH R.: An intercomparison of computer-assisted circulation classification methods. *International Journal of Climatology* 16, 8 (1996), 893–922. doi:10.1002/(SICI)1097-0088(199608)16:8<893::AID-JOC51>3.0.CO;2-Q. 1
- [Hut00] HUTH R.: A circulation classification scheme applicable in GCM studies. *Theoretical and Applied Climatology* 67, 1–2 (2000), 1–18. doi:10.1007/s007040070012. 1
- [JMF99] JAIN A. K., MURTY M. N., FLYNN P. J.: Data clustering: a review. *ACM Computing Surveys* 31, 3 (1999), 264–323. doi:10.1145/331499.331504. 1
- [KCB03] KOKARE M., CHATTERJI B., BISWAS P.: Comparison of similarity metrics for texture image retrieval. In *IEEE TENCON 2003* (2003), Allied Publishers Private Ltd., pp. 571–575. doi:10.1109/TENCON.2003.1273228. 2
- [LKL05] LIN J., KEOGH E., LONARDI S.: Visualizing and discovering non-trivial patterns in large time series databases. *Information Visualization* 4, 2 (2005), 61–82. doi:10.1057/palgrave.ivs.9500089. 1
- [LYJ05] LIWEI WANG, YAN ZHANG, JUFU FENG: On the Euclidean distance of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27, 8 (2005), 1334–1339. doi:10.1109/TPAMI.2005.165. 2
- [NOA11] NOAA NATIONAL WEATHER SERVICE CLIMATE PREDICTION CENTER: Climate Prediction Center - Monitoring and Data: Current Monthly Atmospheric and Sea Surface Temperatures Index Values, 2011. URL: <http://www.cpc.ncep.noaa.gov/data/indices/>. 3
- [RVLS10] RUST H. W., VRAC M., LENGAINNE M., SULTAN B.: Quantifying Differences in Circulation Patterns Based on Probabilistic Models: IPCC AR4 Multimodel Comparison for the North Atlantic. *Journal of Climate* 23, 24 (2010), 6573–6589. doi:10.1175/2010JCLI3432.1. 1
- [Ste05] STEWART R. H.: Introduction to Physical Oceanography, 2005. URL: http://oceanworld.tamu.edu/resources/ocng_textbook/contents.html.
- [TMB02] TVERSKY B., MORRISON J. B., BETRANCOURT M.: Animation: can it facilitate? *International Journal of Human-Computer Studies* 57, 4 (2002), 247–262. doi:10.1006/ijhc.2002.1017. 1
- [Tuf90] TUFTE E. R.: *Envisioning information*. Graphics Press, Cheshire and CT, 1990. 1
- [Tuf01] TUFTE E. R.: *The visual display of quantitative information*, 2 ed. Graphics Press, Cheshire and CT, 2001. 1
- [VRB00] VENEAU E., RONFARD R., BOUTHEMY P.: From video shot clustering to sequence segmentation. In *Proceedings of the 15th International Conference on Pattern Recognition* (2000), Sanfeliu A., (Ed.), IEEE Computer Society Press, pp. 254–257. doi:10.1109/ICPR.2000.902907. 2
- [vWvS99] VAN WIJK J. J., VAN SELOW E. R.: Cluster and calendar based visualization of time series data. In *Proceedings of the 1999 IEEE Symposium on Information Visualization (InfoVis'99)* (1999), Wills G., Keim D., (Eds.), IEEE, pp. 4–9. doi:10.1109/INFVIS.1999.801851. 1
- [WS09] WOODRING J., SHEN H.-W.: Multiscale Time Activity Data Exploration via Temporal Clustering Visualization Spreadsheet. *IEEE Transactions on Visualization and Computer Graphics* 15, 1 (2009), 123–137. doi:10.1109/TVCG.2008.69. 2