UDC 551.590.2
IRSTI 37.23

GRAPH THEORY IN ANALYSIS OF PALEOCLIMATIC TIME SERIES

I.S. Knyazeva[1], N.G. Makarenko[2], A.A. Pak[3]
[1,2]The Central Astronomical Observatory of
the Russian Academy of Sciences at Pulkovo, Saint-Petersburg, Russia
[3]Institute of Information and Computational Technologies, Almaty, Kazakhstan
[1]iknyazeva@gmail.com, [2]ng-makar@mail.ru, [3]aa.pak83@gmail.com
[1]ORCID ID: https://orcid.org/0000-0003-0026-6047
[2]ORCID ID: https://orcid.org/0000-0001-9917-1421
[3]ORCID ID: https://orcid.org/0000-0002-8685-9355

Abstract. At the current work, we present the results of diagnostic of similarities of various paleo reconstructions of temperatures with the help of network approaches. The first step of data processing is transformations of correlational patterns of time series into the geometry of the corresponding network, which is an input for further processing steps by methods of algebraic topology. To detect the possible non-linear connections between climatic series and solar activity, we propose the approach based on network generation with the help of embedding time series into the feature space of the corresponding dimension. In conclusion, we give markovian chains for climatic reconstructions and Wolf numbers.
Keywords: climatic series, Wolf numbers, solar activity, algebraic topology

Introduction

The information about the climate of Earth at previous periods of time before the epoch with instrumental recordings is about the extraction data from auxiliary sources such as annual tree rings, glacial kerns, marine deposits, glacial varved clays, thermospeleological series. Various thermal reconstructions can be calculated from correlations of these proxy indicators with temperature of instrumental period. So, the result of such reconstructions of averaged temperature in North hemisphere is an inference about "hockey stick", namely the steep increase of averaged temperature in 20s century, which was after relatively stable state of temperature during 1000 years. These point of view is not common. It is not unlikely that the unique of current increase of temperature can be erroneous, due to the fact of understaged low-frequency trends in data [1].

There is an open question about the quality of the reconstruction. It is worth noting, that regression models of Sluckogo-Yula, which are the basis of the majority of reconstructions, actually, they are applicable only for stationary (or weakly stationary) random series, where randomness is endogenic property. That's poorly agreed with commonly approved climatic scenarios [2]. As the result of stated above, there are several main problems of inferences about the future scenarios of climate:

How to check the quality of reconstructions?
How to choose the best reconstruction?
What's the meaning of the "best" one of reconstructions?
How correctly compare reconstructions?
How correctly average reconstructions?

Recently we tried to address some of these questions. The main idea was about the checking, is it correct to use a differentiable model to generate reconstruction, e.g., Takens models [3]. The answer was negative: the Hölder regularity of paleoreconstructions turned out to be worse than that of the gauge instrumental series [4].

In this article, we use the methods of graph (network) theory to compare different reconstructions with each other and detect the sun-conditioned component in them. Networks are built directly from time series. The presence of such a component would make it possible to construct a criterion for choosing the best reconstruction within the framework in a correct way. Networked approaches have been used previously to visualize climate and meteorological data [5], [6].

We use quantitative network descriptors from time series, (1) to compare paleoreconstructions with each other and (2) to highlight a possible solar component in temperature reconstructions. For this, the annual series of Wolf numbers from 1090 to 2002 was used as a reconstruction of solar activity [7].

Considered 6 temperature reconstructions of the previous millennium, described in the work [8]. Let's list them: reconstructions by Briffa[9] and Esper [10], Jones [11] and Mann's reconstruction [12].

The last mentioned is considered the most comprehensive because it sourced paleoclimatic records from more than 100 different locations. It took into account data with different time resolutions and scales. The Crowley reconstruction [13] is based on a sample that uses only the data that is available at all times. Finally, Moberg's latest reconstruction [14] is based on high frequency (tree rings) and low frequency data, which are considered separately using wavelet decomposition. All of these reconstructions were investigated using network approaches.

2 Construction of complex networks from series

The main idea of constructing networks from series is to transform a time series into a complex network or graph [15], [16]. A graph is an ordered set of vertices (nodes) and edges (arcs). We describe below three ways to build networks from rows.

2.1 Loop networks

This method, proposed in [17], refers to the construction of complex networks from pseudo-periodic time series. The approach assumes that the time series is cyclical, and a separate cycle is taken as the base unit or node for building the network. A network of such rows is constructed as follows. The time series is divided into cycles, which can be of different lengths; each cycle is considered a network node. So, for the time series of Wolf numbers, the nodes are separate cycles. To construct edges between nodes, they are compared with each other in a suitable metric, for example, in Euclidean [16].

Let $d(C_i, C_j)$ is a distance between two loops: $C_i = (x_1, x_2, ..., x_{n_i}), C_j = (y_1, y_2, ..., y_{n_j}),$ and Δ is corresponding cutoff. Then the adjacency matrix $a(i, j)$ is defined as follows: if $d(C_i, C_j) < \Delta$, then $a(i, j) = 1$ and loops C_i, C_j considered to be connected. Contra versa they are disconnected, $a(i, j) = 0$. With the help of an adjacency matrix and appropriate algorithm one can produce a network [18].

2.2 Networks of embedded data

Another method [16], allows to build networks for arbitrary, not necessarily cyclic, time series using the technique of topological embeddings of scalar data into R^m. The choice of dimension and lag is a stand alone problem of chaotic dynamics and depends on the data [3]. After the data is embedded, each point, i.e. - vector of lagging coordinates, is considered as a separate node. The network is obtained by connecting a node with its k nearest neighbors, for some fixed k. In the case when two points are mutually close to each other, the connection is added only once, so that the network is built as a simple undirected graph. The average complexity of the network will be 2k. After that, by evaluating the correlation function, links caused by the close arrangement of nodes are excluded. The resulting network is an NxN connectivity matrix filled with zeros and ones. Various approaches can be used to visualize this matrix; general requirements for the construction of a graph are reduced to minimizing the number of intersections of the edges of the graph and the distance between the vertices [18]. Figure 1 shows an example of a time series representing one of the coordinates of the solution to the Ressler equation [19] and the corresponding network, for the following nesting parameters: dimension m =3, lag τ = 10.

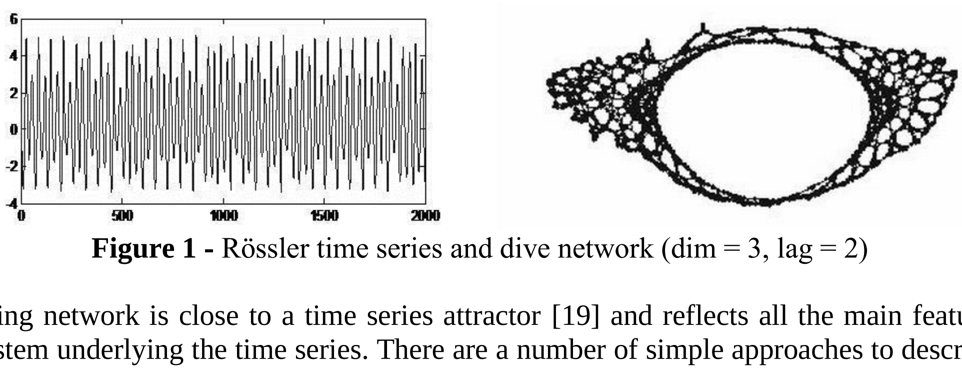

Figure 1 - Rössler time series and dive network (dim = 3, lag = 2)

The resulting network is close to a time series attractor [19] and reflects all the main features of the dynamical system underlying the time series. There are a number of simple approaches to describe graphs formally. One of them is based on the so-called "motives" that arise in the network and the frequency of their occurrence [20] . A motive of size n is called a subnet consisting of n nodes. In fact, each motif represents one of the possible ways of connecting n points. For simplicity of calculations, motives of size 4 are usually considered, which contains 6 variants of motives (see Fig. 2)

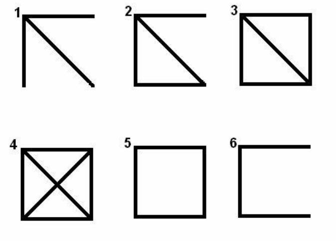

Figure 2- Variants of motives for a subnet with a number of nodes equal to 4.

It turns out that for time series generated by different dynamical systems, the distribution of motives is significantly different. Table 1 shows the frequencies of occurrence of motives from Fig. 2. for networks built by nesting three time series.

Table 1. Distribution of the occurrence of motives for three model series

Motif №	Lorenz	Noise	Resler
1	0,0736	0,02904	0,04229
2	0,25606	0,33112	0,29853
3	0,10145	0,09766	0,12309
4	0,01038	0,02226	0,01938
5	0,06893	0,01171	0,04808
6	0,4896	0,50822	0,46863

.

Note that more informative descriptors of the graph complexity are the spectrum of eigenvalues of the Laplace-Kirchhoff matrix [21].

2.3 Markovian networks

Consider a dynamic system (M,T) like mapping $T: M \to M$ defined on a compact subset $M \in R^d$. The random trajectory of this system is a sequence of points: $\{x, Tx, T^2x, ...\}$. To get a coarse-grained version of the dynamics, let's split M into a finite number of connected non-empty subsets: $\bigcup_{i-1}^{n} A_i \subseteq M$, $A_i \cap A_j = \varnothing$. Let's identify $i-$ state of a dynamic system with a cell A_i. We do not discuss the choice of a suitable partition here. It is known that the Markov model of dynamics is given by the matrix of transition probabilities: $\hat{P}_{ij} = \text{Prob}_{\mu}\{Tx \in A_j \,|x \in A_i\}$ [Froyland, 2001]. Here the probability is understood with respect to some natural ergodic measure μ, assuming that for almost all starting points $x \in M$ the trajectories have the same distribution. However, this measure is usually unknown. Therefore, instead of μ use the normalized Lebesgue measure:

$$P_{ij} = \frac{m\left(A_i \cap T^{-1} A_j\right)}{m\left(A_i\right)}$$

In other words, the probability of transition from the state i (or cell A_i) into state j (cell A_i) is determined by the relative proportion of points A_i who moved to A_j by affect of action T. The described ideas underlie the Markov networks proposed in the work [22].

Let's construct a histogram of samples for a time series so that each bin contains approximately the same number of points. In this case, the width of the bins will be of course different, and the bins themselves will play the role of partitioning the phase space, i.e. $\{A_i\}_{i=1}^{n}$. Let us assign a serial number to each bin, and then encode each value of the time series in accordance with the belonging to a certain bin. As a result of such a procedure, instead of the original row, a new row consisting of bin numbers will be obtained. Then we calculate the frequency of all consecutive transitions from one number to another, we get a transition matrix of size $N \times N$, where N is number of bins.

In the event that the row is smooth, adjacent transitions will prevail and the matrix will be close to a diagonal form, but if, on the contrary, the structure is very chaotic, then the matrix will have many filled elements far from the main diagonal.

3. Numerical results

Below are the results of a network analysis of the monthly average Wolf numbers, the reconstruction of the annual values of a series of Wolf numbers (from 1090 to 2002) obtained by Yu.A. Nagovitsin in [7], and temperature reconstructions of the northern hemisphere over the last thousand years described in the introduction.

3.1 Comparison of solar activity cycles

The Wolf number series is one of the main indices of solar activity, and therefore we used it to identify possible solar-terrestrial connections. Wolf's Zurich series of monthly averages has been available since 1749 and contains 23 complete cycles. We have built cyclical networks for this series to identify similarities between different cycles. The division into nodes of the network was carried out in accordance with the known dates of the beginning and end of the cycles according to the list of solar activity cycles.

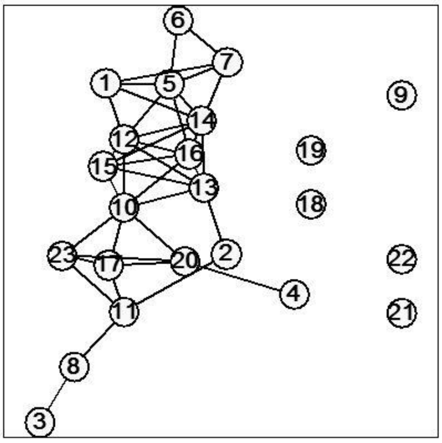

Figure 3- Relationships between cycles at the threshold $\Delta = 300$. Five cycles form isolated vertices.

Further, between each pair of cycles, the distance in the metric L_2 was calculated and a threshold $\Delta = 350$ was selected for which the adjacency matrix was constructed. The choice of this value was due to the fact that this is the minimum threshold at which each node has at least one connection with others. Figure 3 shows the minimum energy graph [18] for a threshold $\Delta = 300$ as an example. This analysis did not lead to any clusters associated with the shape of the cycles.

3.2 Analysis of the dynamics of solar activity and temperature reconstructions

To diagnose possible connections between temperature reconstructions and solar activity, we used embedded nets (see Section 2.2). For different lags, the networks constructed by the series of Wolf numbers demonstrated a stable topological structure, similar to the networks obtained for systems of deterministic chaos (see Figs. 1 and 4). Stability is understood here in the following sense of the similarity of the network built over the entire sample and its part. Figure 4 (top panel) clearly shows the similarity of the networks for the complete Wolf series and the network for the first ten cycles. In the bottom panel of Fig. 4, on the right, the net for the Ressler attractor for $m = 3, \tau = 2$, and on the left, the net for the Gaussian noise. The similarity of the networks for the Wolf numbers and the Ressler attractor confirms the conclusions of [19]. An analysis of the frequency of occurrence of motives also confirmed the connection that a number of monthly Wolf numbers belong to chaotic processes.

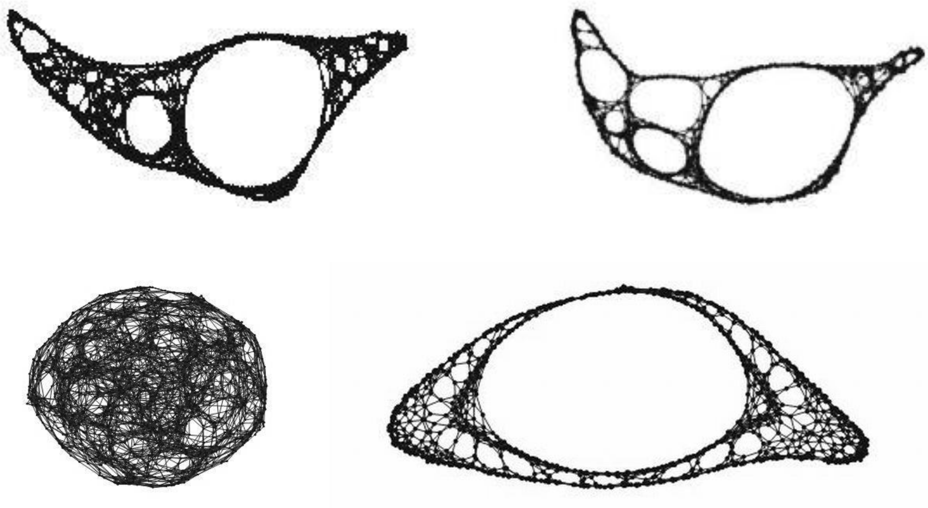

Figure 4- Networks for different tiers. Above - Wolf numbers, complete series (left), right first 10 cycles of Wolf series ($m=3, \tau=5$). Below, on the right - the network for the Rossler attractor ($m=3, \tau=2$), on the left - Gaussian noise ($m=3, \tau=2$)

To analyze climatic regimes, we also used dive nets for annual temperature reconstructions from 1000 to 1995 by various authors and for various diving parameters. In all cases, it turned out that networks are structures that are unstable in time and differ from each other. Here, as above, instability means the difference in the form of networks for different time intervals. Qualitatively, these networks do not resemble either a random series or networks of deterministic chaos.

Figure 5 shows examples of such networks for temperature reconstructions at immersion parameters $m=24, \tau=1$. We chose the dimension $m=24$ in connection with the indication of the possible presence of the Hale cycle in the climatic series [2]. It turned out that the networks of all temperature reconstructions differ to varying degrees from each other. Table 2 and Table 3 show the results of the numerical analysis of the graphs. It is interesting to note that the reconstructions of Briffa and Esper have a rather visual structure and similar parameters of the distribution of motifs and curvature. Perhaps this is due to the fact that they rely only on dendrodata. The rest of the reconstructions, despite the fact that they use similar multi-proxy data, differ significantly from each other. The Crowley and Moberg reconstructions stand out in particular. Thus, it can be assumed that the internal dynamics of the reconstructions is not sufficiently invariant with respect to the use of the initial data.

Figure 5- Types of networks for various climatic reconstructions. Left, net for Mann row, right, for Moberg row for $m=24, \tau=1$.

To analyze possible solar-terrestrial connections, networks for reconstruction of the annual Wolf numbers [7] were constructed with the same parameters as for temperature reconstructions. The results were quite unexpected. It turned out that year-on-year data is already poorly matched to networks for processes with a dynamic chaos regime. Moreover, visually, they are more similar to climatic reconstructions. It turned out that with an increase in the dimension of the phase space, the structure of the network changes and becomes closer to the structure of temperature reconstructions

Markov networks for temperature reconstructions also correspond to the picture obtained above: the correlation structure and internal dynamics of the variants are different. The degree of dominance of the main diagonal corresponds to the smoothness of the row; the more pronounced it is, the smoother the row.

4. Conclusions

The results obtained can be summarized as follows:

1. The structure of cyclic networks for the monthly average Wolf numbers does not contain any distinguished clusters, which could be expected in the case of the existence of one or several “typical” cycles with analogs close to them, in the sense of the metric used.

2. Immersion networks for monthly average Wolf numbers are qualitatively and in terms of the distribution of motives similar to dynamic chaos (like the Ressler attractor). The nets for the full row and its half are similar. Perhaps this indicates the existence of a certain scale-invariant measure.

3. Networks for paleoclimatic series are qualitatively different from both random networks and dynamic chaos networks. They differ significantly from each other, indicating a lack of coordinated dynamics of reconstruction. Mutual differences decrease with an increase in the dimension of investment, which is possibly due to the presence of long correlations (24 years, 60 years or more). Markov networks support these findings.

4. The structure of the networks for the reconstruction of the annual Wolf numbers substantially depends on the dimension of the immersion. At low dimensions, networks are close to dynamic chaos; at large, they are similar to the networks of climatic series. Perhaps this is due to the absence of significant long-term relationships between cycles in the analyzed time interval.

References

[1] Christiansen B. “Reconstructing the NH mean temperature: Can underestimation of trends and variability be avoided?” *J. Clim.*, (2011). 24(3). 674–692, doi: 10.1175/2010JCLI3646.1.

[2] Monin A. S., D. M. Sonechkin, *Observational Climate Fluctuations*. Moscow: Nauka, 2005.

[3] Ott E., Sauer T., Yorke J. A., Coping with chaos. Analysis of chaotic data and the exploitation of chaotic systems. (1994).

[4] Makarenko N. G., Karimova L. M., Helama S., Eronen M. “Evaluating direct and indirect evidence of climatic change by Hölder regularity and order pattern in time series,” *Geomagn. Aeron.*, (2013). 53. 8. 922–926, Dec., doi: 10.1134/S0016793213080148.

[5] Tsonis A. A., Swanson K. L., Roebber P. J. What do networks have to do with climate? Bull. Am. Meteorol. Soc. 2006. 87(5). 585–595, doi: 10.1175/BAMS-87-5-585.

[6] Paluš M., Hartman D., Hlinka J., Vejmelka M. Discerning connectivity from dynamics in climate networks. Nonlinear Process. Geophys. 2011. 18(5). 751–763, doi: 10.5194/npg-18-751-2011.

[7] Nagovitsin Y. A. Nonlinear mathematical model of the solar cycle process and possibilities for reconstruction of past activity. Pisma v AZh. 1997. 23(11).

[8] Ogurtsov M.G., Junger H., Lindholm M., Helama S., Dergachev V. A. Quasisecular cyclicity in the climate of the earth’s nothern hemisphere and its possible correlation with solar activity variations. Solar-Terrestrial Phys. 2007. 10(123). 10–15.

[9] Briffa K. R. Annual variability in the Holocene: Interpreting the message of ancient trees. Quat. Sci. Rev. 2000. 19. 87–105.

[10] Esper J., Cook E. Schweingruber F. H. Low-frequency signals in long tree-ring chronologies and the reconstruction of past temperature variability. *Science (80-.).* 2002. 295. 2250–2253.

[11] Jones P. D., Briffa K. R., Barnett T. P. Tett F. B. High-resolution palaeoclimatic records for the last millennium: interpretation, integration and comparison with General Circulation Model control-run temperatures. *The Holocene*. 1998. 8(4). 455–471, doi: 10.1191/095968398667194956.

[12] Mann M. E., Bradley M. K., Raymond S. Hughes. Northern hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations. *Geophys. Res. Lett.* 1999. 26. 759–762.

[13] Crowley T. J. Lowery T. S. How warm was the Medieval warm period?. *Ambio. 2000*. 29. 51–54.

[14] Moberg A., Sonechkin D. M., Holmgren K., Datsenko M. H., Karlén W. Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data *Nature*. 2005. 433(7026). 613–617, doi: 10.1038/nature03265.

[15] Gao Z. K., Small M., Kurths J. Complex network analysis of time series. *EPL. 2016.* 116(5). 50001, doi: 10.1209/0295-5075/116/50001.

[16] Small M., Zhang, J. Xu X. Transforming time series into complex networks. in *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering.* 2009. 5(2). 2078–2089, doi: 10.1007/978-3-642-02469-6_84.

[17] Zhang J., Small M. Complex network from pseudoperiodic time series: Topology versus dynamics. *Phys. Rev. Lett. 2006.* 96 (23). 238701, doi: 10.1103/PhysRevLett.96.238701.

[18] Battista D., Eades G., Tamassia R., Tollis I.G., Graph Drawing: Algorithms for the Visualization

of Graphs. Prentice Hall. 1999.

[19] Letellier C., Aguirre L. A., Maquet J., and R. Gilmore, Evidence for low dimensional chaos in sunspot cycles. *Astron. Astrophys. 2006.* 449(1). 379–387. doi: 10.1051/0004-6361:20053947.

[20] Xu X. K., Zhang J., Li P., Small M. Changing motif distributions in complex networks by manipulating rich-club connections. *Phys. A Stat. Mech. its Appl. 2011*. 390(23–24). 4621–4626, doi: 10.1016/j.physa.2011.06.069.

[21] Alekseev V. V., Makarenko N. G., Knyazeva I. S. Graph Dynamics of Solar Active Regions: Morse–Smale Complexes and Multiscale Graphs of Magnetograms. *Astron. Lett. 2020.* 46(7). 488–500, doi: 10.1134/S1063773720070014.

[22] Shirazi A. H., Jafari G. R., Davoudi J., Peinke J., Rahimi M. R. Tabar, Sahimi M. Mapping stochastic processes onto complex networks *J. Stat. Mech. Theory Exp.* 2009. 9(7). P07046, doi: 10.1088/1742-5468/2009/07/P07046.