OFFPRINT

On the effects of lag-times in networks constructed from similarities of monthly fluctuations of climate fields

G. Tirabassi and C. Masoller

Please visit the new website
www.epljournal.org

epl

A Letters Journal Exploring the Frontiers of Physics

EPL Compilation Index

www.epljournal.org

Visit the EPL website to read the latest articles published in cutting-edge fields of research from across the whole of physics.

Each compilation is led by its own Co-Editor, who is a leading scientist in that field, and who is responsible for overseeing the review process, selecting referees and making publication decisions for every manuscript.

- Graphene
- Liquid Crystals
- High Transition Temperature Superconductors
- Quantum Information Processing & Communication
- Biological & Soft Matter Physics
- Atomic, Molecular & Optical Physics
- Bose–Einstein Condensates & Ultracold Gases
- Metamaterials, Nanostructures & Magnetic Materials
- Mathematical Methods
- Physics of Gases, Plasmas & Electric Fields
- High Energy Nuclear Physics

If you are working on research in any of these areas, the Co-Editors would be delighted to receive your submission. Articles should be submitted via the automated manuscript system at **www.epletters.net**

If you would like further information about our author service or EPL in general, please visit **www.epljournal.org** or e-mail us at **info@epljournal.org**

Biaxial strain on lens-shaped quantum rings of different inner radii, adapted from **Zhang et al** 2008 EPL **83** 67004.

Artistic impression of electrostatic particle–particle interactions in dielectrophoresis, adapted from **N Aubry and P Singh** 2006 EPL **74** 623.

Artistic impression of velocity and normal stress profiles around a sphere that moves through a polymer solution, adapted from **R Tuinier, J K G Dhont and T-H Fan** 2006 EPL **75** 929.

IOP Publishing

Image: Ornamental multiplication of space-time figures of temperature transformation rules (adapted from T. S. Bíró and P. Ván 2010 EPL **89** 30001; artistic impression by Frédérique Swist).

On the effects of lag-times in networks constructed from similarities of monthly fluctuations of climate fields

G. TIRABASSI and C. MASOLLER

Departament de Fisica i Enginyeria Nuclear, Universitat Politecnica de Catalunya - Colom 11, Terrassa 08222, Barcelona, Spain, EU

Abstract – The complex network framework has been successfully applied to the analysis of climatological data, providing, for example, a better understanding of the mechanisms underlying reduced predictability during *El Niño* or *La Niña* years. Despite the large interest that climate networks have attracted, several issues remain to be investigated. Here we focus on the influence of the periodic solar forcing in climate networks constructed via similarities of monthly averaged Surface Air Temperature (SAT) anomalies. We shift the time series in each pair of nodes such as to superpose their seasonal cycles. In this way, when two nodes are located in different hemispheres we are able to quantify the similarity of SAT anomalies during the winters and during the summers. We find that data time-shifting does not significantly modify the network Area Weighted Connectivity (AWC), which is the fraction of the total area of the Earth to which each node is connected. This unexpected network property can be understood in terms of how data time-shifting modifies the strength of the links connecting geographical regions in different hemispheres, and how these modifications are washed out by averaging the AWC.

Introduction. – In recent years the application of the complex networks to climatological data lead to the development of the field of climate networks [1–9], where the nodes represent geographical coordinates and the links quantify the degree of statistical similarity of the climate: if the similarity in two nodes is above a certain threshold, then these nodes are interconnected with a link [1]. Since the network is defined over a regular grid of nodes, taking into account that two nodes that are geographically close tend to have similar climate, strong local links can be expected [10]. Distant nodes will have weaker similarity values and thus, teleconnections will, in general, be represented by weak links.

Climate networks have been found consistent with well-known climate phenomena, such as the teleconnection between El Niño and Asian Monsoon basin [11], and have also provided new insight into our climate. As teleconnections are not static, but respond to global climate changes, it has been shown that during periods of global warming, the number of long-range links increases due to an increase of the strength of teleconnections [12]. It has also been shown that is possible to predict the occurrence of an El Niño Southern Oscillation (ENSO) event by analysing the evolution of the network topology in El Niño basin [3,13].

Despite the successful application of complex networks to climate data analysis, to the best of our knowledge the influence of the solar forcing has not yet been investigated. The network approach to the study complex systems, such as the brain or our climate [1,14], is based on identifying similarities via time series analysis, usually employing cross-correlation or mutual information measures. In analogy with functional brain networks, where links between different brain regions can be clearly identified, which might or might not have a known underlying anatomical connection in the brain, in climate networks the links reflect climate similarities between pairs of nodes, which are not necessarily connected via ocean and/or atmospheric processes. However, a difference with brain functional networks is that in climate networks the common solar forcing might induce correlations which could result in a distortion of the true network backbone. An underlying assumption of the climate network approach is that, by building the network from similarities of climate anomalies, the influence of the common solar forcing is removed. Since the climate system is a complex system, with a wide

range of time-scales, the interplay of nonlinearities and noise with a small external periodic forcing is highly non-trivial [15,16] and a key issue is to analyse the role of the common solar forcing in the connectivity of the network.

To analyse the influence of the solar annual cycle we consider monthly averaged Surface Air Temperature (SAT) Anomalies (SATA). Before constructing the network we first calculate the lag-times between any pair of nodes by shifting their time series of SAT data (including the annual cycle) such as to superpose their seasonal cycle. This allows comparing the anomalies during the same season in nodes located in different hemispheres.

Since the network connectivity is based in the statistical similarity of the time series, an increase of connectivity in the mid-latitudes (where the seasonal cycle is strong and well defined) and a decrease of connectivity in low-latitudes (where the seasonal cycle is weak or non-existent) could be expected if SATA data have a residual effect of the annual solar forcing. To track the changes in the network induced by time-shifting we used the usual graphical representation of climate networks, the Area Weighted Connectivity (AWC), which is the fraction of the Earth to which each network node is connected. When comparing the AWC of the networks obtained with and without time-shifting we found only minor differences.

This unexpected network property can be understood from the variation of the similarity values (the Cross-Correlation, CC, or the Mutual Information, MI) when these are computed with lag-times. The short-range links connecting neighbouring nodes, and the long-range links connecting nodes in the same latitude tend to have zero lag; thus, their similarity value is not modified and thus, they do not modify the AWC. The links representing the well-known teleconnection between El Niño and Indian Monsoon regions have also zero lag and do not change the AWC. With respect to the rest of long-range links that connect nodes in different hemispheres, most of them have non-zero lags and the time-shifting indeed changes their values; however, these changes are in general random and are mainly washed out when computing the average connectivity.

Dataset. – We consider monthly averaged SAT, re-analysis data from the Center for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) [17]. The data covers the period from January 1948 to May 2012 ($N = 773$ months) and have a spacial grid resolution of $2.5°$ ($M = 10226$ nodes). Removing the annual cycle in each node results in zero-mean anomalies (SATA), which are then normalized to have unitary variance.

Identification of lag-times. – To compare the climate in any two nodes, i and j, under the same stage of the annual solar cycle, we shift the time series in one node, and find the time shift that gives maximum similarity, which is referred to as the *lag-time*, τ_{ij}. Specifically, we calculated the Cross-Correlation coefficient ($CC_{ij}(\tau)$) between the
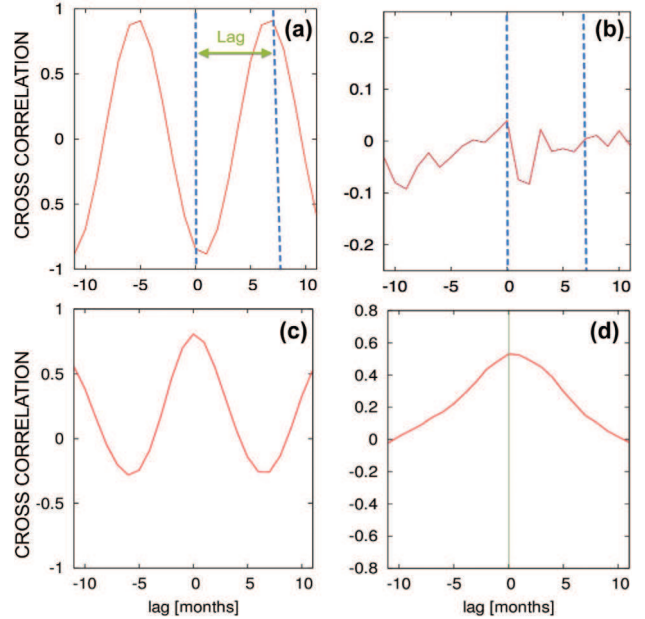

Fig. 1: (Color online) Representation of the method used to identify the lag-time, τ_{ij}, between nodes i and j. As an example we plot in panel (a) the cross-correlation of the SAT time series of nodes close to Rome and Buenos Aires; there the maximum in the interval [0, 11] occurs at $\tau_{ij} = 7$ months. (b) CC between the same nodes but now computed from SATA data (notice that there is no significant maximum). (c), (d): as (a), (b) but for the teleconnection between a node in the equatorial Indian Ocean (7 S, 65 E) and a node in the *El Niño* basin (12 S, 145 W). In this case there is a pronounced maximum at zero lag in (d) (vertical line).

SAT time series (including the annual cycle) in i and j, shifting one series τ months with respect to the other. Then, τ_{ij} is chosen as the value of τ in the interval [0, 11] where $CC_{ij}(\tau)$ is maximum (see fig. 1). There are pairs of nodes for which $CC_{ij}(\tau)$ has two maxima in the interval [0, 11]; this can be frequent when one of the nodes is in the tropical regions, were there is a well-known semi-annual periodicity. In these cases the lag-time was chosen as the smallest of the two possible lag-times.

To demonstrate that this procedure indeed gives meaningful lag-times, in fig. 2 we display in color code the lag-times of three nodes, one in the North Hemisphere (NH), one in the South Hemisphere (SH), and one in the tropics. The maps reveal clear characteristic patterns, signatures of climatic regions. We can also note the memory effect induced by the oceans, and the almost perfect 6 months symmetry between the NH and SH.

For the sake of clarity we remark that τ_{ij} if determined from the similarity of the SAT field, while the network is constructed from the similarity of anomalies (SATA field).

Construction of climate networks. – We measure the degree of *statistical similarity* of the time series in two nodes by the absolute value of the Cross-Correlation (CC) coefficient and by the Mutual Information (MI), using the
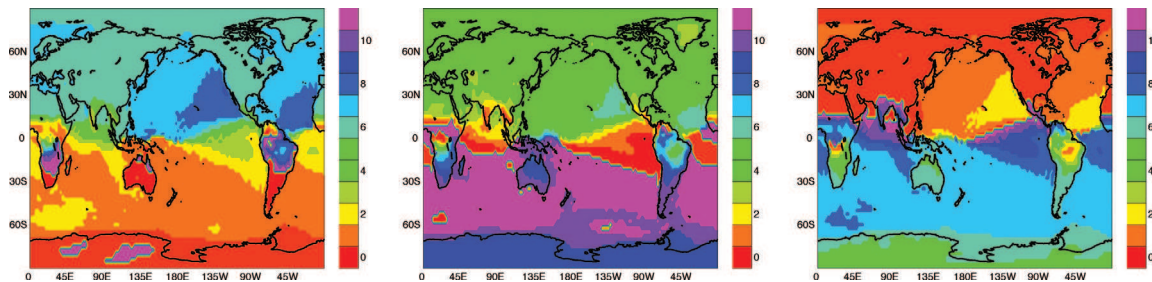
Fig. 2: (Color online) Lag-times of a node in Mongolia (left), in Australia (center) and in the El Niño basin (right).

identified lag-times:

$$CC_{ij} = \left| \frac{1}{N} \sum_{t=0}^{N} a_i(t) \ a_j(t + \tau_{ij}) \right|, \qquad (1)$$

where the a_i and a_j are the SATA time series in nodes i and j, τ_{ij} is their lag-time, and $N = 773$ months;

$$MI_{ij} = \sum_{m,n} p_{ij}(m,n) \log_2 \left(\frac{p_{ij}(m,n)}{p_i(m)p_j(n)} \right), \qquad (2)$$

where the p_i, p_j are the probability distributions associated to $a_i(t)$ and $a_j(t)$, and p_{ij} is the joint probability of $a_i(t)$, $a_j(t + \tau_{ij})$. These probabilities were estimated by 8-bin frequency histograms [2].

We also used the *Mutual Information Ordinal Pattern* (MIOP), for which the probabilities p_i, p_j and p_{ij} are computed from the ordinal representation of the time series $a_i(t)$ and $a_j(t)$, *i.e.*, from the sequence of Ordinal Patterns (OPs), $n_i(t)$, $n_j(t)$, that keep information about the order in which the values appear in the time series [5,6,18]. This symbolic transformation allows to compute the MI —and from that to obtain a network— that takes into account memory effects at different time-scales, depending on the way the OPs are constructed.

We consider OPs of length $D = 4$, and thus the number of possible different OPs is 24. In this way, the SATA data in each node i, $a_i(t)$, is transformed into a sequence of integer numbers, $n_i(t)$ with $n_i \in [1, 24]$. From the symbolic sequences in two nodes we compute the probabilities of the OPs, and then, the MIOP from eq. (2).

The advantage of the symbolic method is that the OPs can be constructed with either consecutive or non-consecutive months. We constructed the OPs within four consecutive months as well as within four equally spaced months covering a one-year period. In the following sections we will present the results only in the former case because the main conclusions are the same for both ways of constructing the OPs.

Distribution of link strengths and lag-times. – A first test of the influence of the lag-times in the climate network is provided by the analysis of the degree of similarity, by comparing the distributions of the values of CC and MI computed with and without lag-times. This
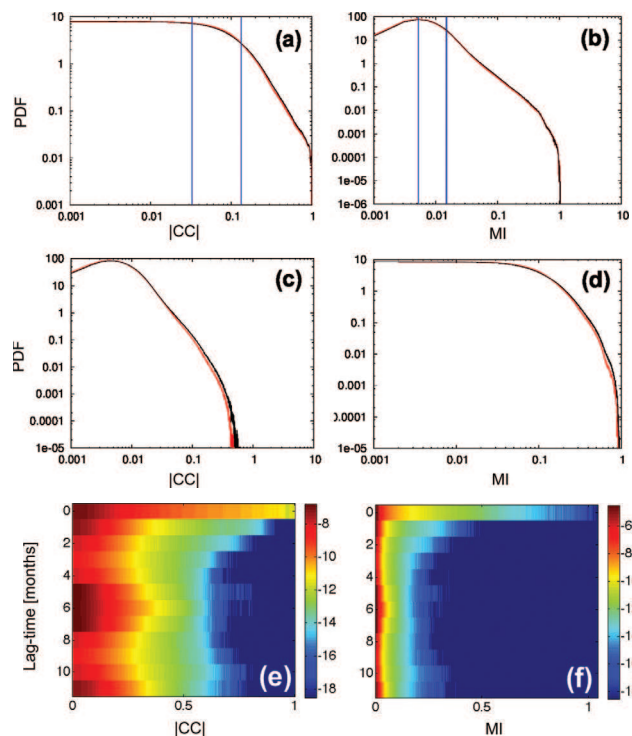

Fig. 3: (Color online) (a), (b): histograms (in log-log scales) of the absolute value of the Cross-Correlation ($|CC|$) and of the Mutual Information (MI), computed with (red) and without (black) lag-time shifting of the time series as can be seen the histograms are nearly identical. The blue lines indicate the two thresholds with which 50% of the links are extracted (see text for details). (c), (d): histograms of CC_{ij} and MI_{ij} such that $\tau_{ij} \neq 0$. The difference with the histograms in (a) and (b) is a decrease of the maximum value due to the removal of local links, which are at $\tau_{ij} = 0$. (e), (f): two-dimensional plots of the probability (in color code in logarithmic scale) *vs.* the value of $|CC|$ or MI and the lag-time. It can be seen that most of the links with high $|CC|$ or MI values have $\tau_{ij} = 0$, which can be expected as the strongest links are short-range connections.

is presented in figs. 3(a), (c) (CC) and figs. 3(b), (d) (MI), where one can observe that the distributions are almost identical, with no significant influence of the lag-times in the mean values or in the shape of the distributions.

Further insight can be obtained by analysing the relation between the strength of a link and its associated lag-
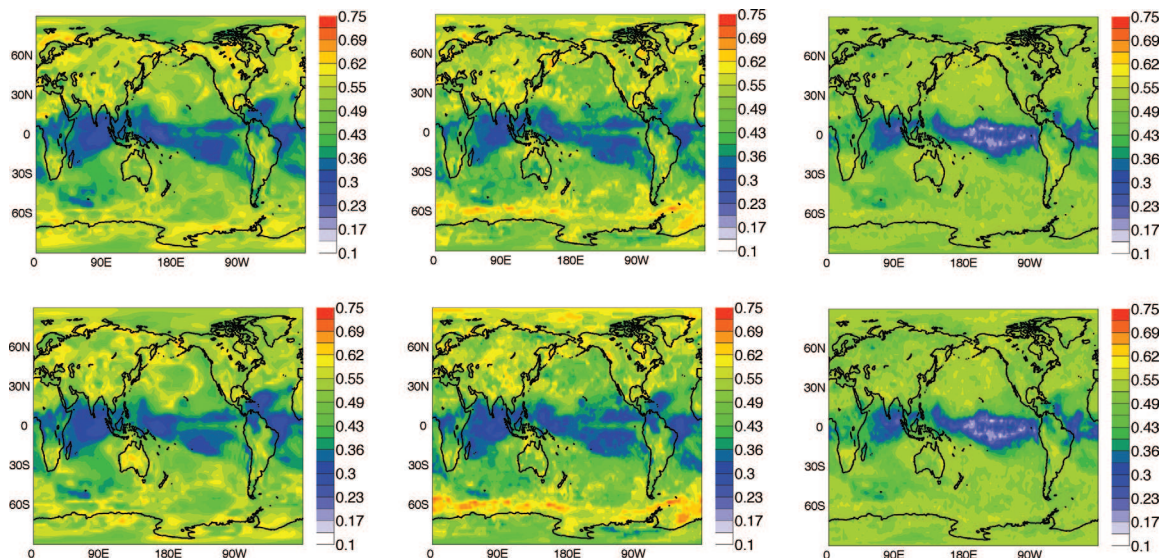
Fig. 4: (Color online) AWC plots: the degree of statistical similarity is quantified with the absolute value of the CC (left), the MI (center) and the MIOP (right), with and without lag-times (lower and upper row, respectively). The thresholds to construct the network are such that the link density is 50%, with the strongest and weakest links removed.

time. As discussed in the introduction, climate networks are defined over a regular grid, in which some long-range teleconnections occur. This fact is reflected in the two-dimensional probability distributions of similarity values in the plane (lag-time, CC/MI value) displayed in fig. 3(e) (CC) and in fig. 3(f) (MI); in both cases the probability is represented in logarithmic scale. As was expected, the higher occurrence of high correlations/mutual information values is at zero lag, confirming the local character of the vast majority of the connections.

Extracting relevant links with non-zero lag-times. – In order to build climate networks we need to define a significance criterion, such that, if the similarity measure of nodes i and j exceeds the threshold they are linked, otherwise they are not.

However, since we are interested in observing the effects of the lag-times in the network topology, we cannot keep only the strongest links, as these are mainly local and thus, have zero lag. The influence of the lag-times in the network connectivity can be observed only if the network contains weak links. However, when links that have very low CC or MI values are included, they might not be relevant as these low correlations might occur just by chance.

As a compromise solution we decided to use two thresholds, in the following way: we chose a high threshold to remove the 25% strongest links (which are at zero lag and will obscure the influence of the lag-times), and a low threshold to remove the 25% weakest links (which are considered noise). In this way the network extracted preserves 50% of the total links, those that are in the second and third "quartiles" (see figs. 3(a) and (b)). To justify the use of a low threshold, we point out that, as the CC value in eq. (1) is the sum of about 770 terms, it can,

for simplicity be assumed to be the sum of independent identically distributed random variables with zero mean and standard deviation equal to 1. Such sum can then be expected to follow the central limit theorem, thus the cross-correlation —without the absolute value as in eq. (1)— can be expected to be a zero-mean Gaussian distributed with standard deviation around 0.04. Then, correlation values above 0.03 are likely to occur by chance with a probability of about 0.4.

Results. – The networks obtained after computing the three similarity matrices, CC_{ij}, MI_{ij} and $MIOP_{ij}$ (with the ordinal patterns constructed with four consecutive months), and after filtering with the thresholding technique described previously, are then graphically represented with the AWC in each node i [12],

$$AWC_i = \frac{\sum_j^M A_{ij} \cos(\lambda_j)}{\sum_j^M \cos(\lambda_j)}. \tag{3}$$

where A_{ij} is the adjacency matrix and λ_j is the latitude of the node j.

The results are presented in fig. 4: in the bottom row the lag-times were included when we computed the similarity matrices; in the top row, instead, the similarity matrices were computed at zero lag.

As can be noticed, the introduction of the lag-times results only in tiny changes in the network, although the network constructed via CC values seems to be more influenced. This is the main and surprising result of our analysis, since it means that the synchronization of the time series according to the annual seasonal cycle (that is, comparing winters with winters, summers with summers and so on) does not increase the degree of connectivity.
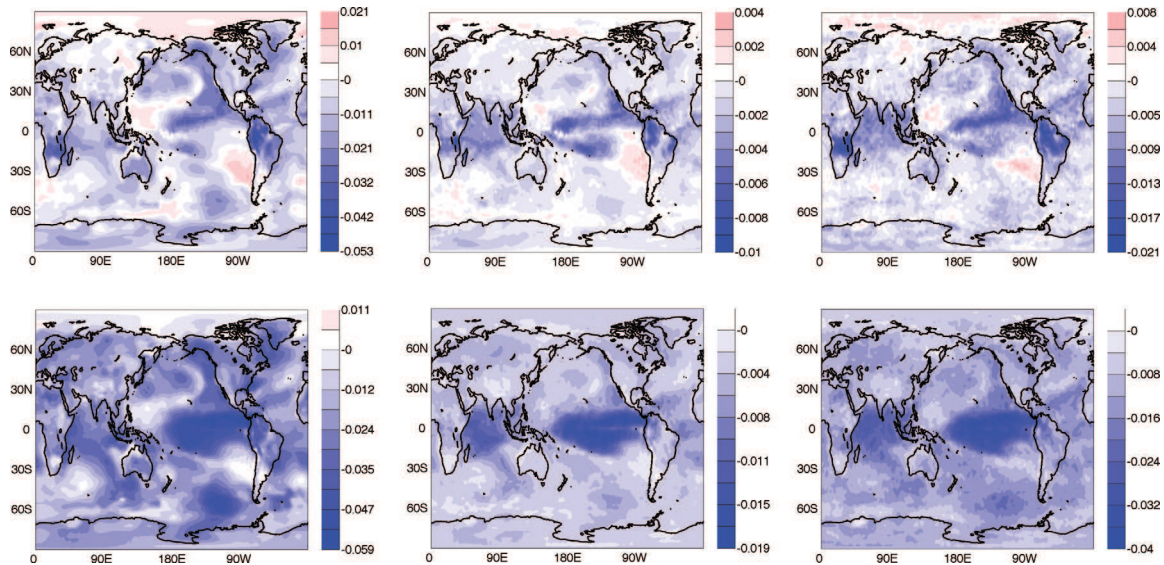
Fig. 5: (Color online) Variation of the average degree of similarity (WAWC, see text). The top row displays the variation when the lag-times used are those computed from SAT data, the bottom row displays the variation when the lag-times are chosen randomly in the interval [0, 11]. The degree of similarity is quantified with the absolute value of the CC (left), the MI (center) and the MIOP (right).

To investigate if the influence of the lag-times could be hidden by noise, *i.e.*, by the presence of weak links that are not significant, or by strong links (those in the third quartile, as the strongest quartile has already been removed), we applied an extra filtering technique. To each link two labels have been assigned: a CC-label and a MI-label, each of them ranging from 1 to 4, representing the quartile the CC or MI value belongs to. Then, we filtered out the links that had one of the two labels equal to 4 (*i.e.*, they were in the strongest quartile, either in terms of CC or MI value). In this way most of the local connections were eliminated. Then, we also eliminated the links that were in the first quartile of CC and in the first two quartiles of MI. In other words, we retained links with small CC values but with intermediate MI values. In this way we obtained networks with density of the order of 50%, but also their AWC maps (not shown) did not reveal any significant influence of lag times.

We also considered the influence of a random choice of the lag-times in the interval [0, 11], and to our surprise, again the AWC maps did not reveal any clear influence of the random data shifts.

It is possible that the introduction of lag-times results in small variations of the strength of the links, that cannot be observed due to the thresholding process, by which the similarity matrices, CC_{ij} and MI_{ij}, are transformed into adjacency matrices, A_{ij}, of 0's and 1's. To investigate this issue, in each node we computed a *Weighted* AWC (WAWC) defined as in eq. (3), but replacing the adjacency matrix, A_{ij}, with the similarity measure (CC or MI). Then, we plotted the difference between the WAWC calculated with lag-times and the WAWC calculated with zero lags. The results are presented in the first row of

fig. 5, where a positive difference indicates an increase of the average correlation.

As we can see the changes are indeed very small, and this observation is consistent in the three similarity measures, with a localized region of weak correlation enhancement, which does not occur in the mid-latitudes. On the contrary, it seems that an average loss of correlation influences these regions and, more in general, the global AWC map. Therefore, not only the introduction of lag-times produces only very small effects, but these occur in unexpected regions, such as western equatorial and southeastern Pacific.

We tested these results against a random choice of lag-times in the interval [0, 11] (fig. 5 second row). As can be seen in this case there is only a loss of correlation, especially in the ENSO basin. This latter effect is due to the fact that El Niño region is connected with almost all the tropical belt, that is a region at zero lag; shifting randomly the series partially destroys the correlations (see the correlation plot in fig. 1(d)), resulting in a decrease in WAWC.

We performed further checks of the analysis. We removed the local links by a spatial threshold of 3000 km, searching for lag effects in the strongest fourth quartile. In this way only the strong teleconnections have been considered. However the inclusion of lag-times did not results in increased in mid-latitude connectivity. We also considered longer lag-times (all lags equal to 5 years) which resulted in the loss of the network architecture and the emergence of noisy structures, as expected.

We point out that, since the analysis was done with SATA data of 773 months, the network obtained is an average of the climate network over more than 64 years. If we use shorter windows the results might be different

and the network topology might show a different sensitivity to time shifts. However, by dividing the 773 months in shorter data sets we can compromise the robustness of the analysis as it would be performed over short data sets, and the variability of the results can then be attributed to either i) insufficient statistics, or ii) the evolution of the network topology. Nevertheless, the possible influence of time shifts in networks defined over shorter time intervals can be an interesting study when performed over data set consisting, for example, of daily or weekly averaged SAT values (instead of averaged monthly as here).

We also carried out an explorative analysis of the geopotential height field at 1000 and 500 hPa, and also these fields seem to show the same behaviour as the SATA field here described.

Conclusions. – We have shown that the introduction of time shifts when computing statistical similarity measures does not produce significant changes in the AWC of climate networks built from monthly averaged SAT anomalies. This is an unexpected finding since the shifts were aimed at comparing the time series in two nodes at the same phase of the annual solar cycle (*i.e.*, comparing winters with winters, summers with summers). This observation is robust with respect to the choice of the statistical similarity measure, and is robust with respect to a random choice of time shifts. Our results can be understood in terms of the lag-time distribution of the links, and the fact that most of the links with non-zero lags are weak links (see the 2D histograms in figs. 3(e), (f)). For these links the time-shifting indeed changes their similarity values; however, these changes appear to be random (see fig. 1(b)) and tend to be washed out when computing the AWC.

$$* * *$$

REFERENCES

[1] Tsonis A. and Roebber P., *Physica A*, **333** (2004) 497.
[2] Donges J., Zou Y., Marwan N. and Kurths J., *EPL*, **87** (2009) 48007.
[3] Gozolchiani A., Havlin S. and Yamasaki K., *Phys. Rev. Lett.*, **107** (2011) 148501.
[4] Berezin Y., Gozolchiani A., Guez O. and Havlin S., *Sci. Rep.*, **2** (2012) 666.
[5] Barreiro M., Marti A. and Masoller C., *Chaos*, **21** (2011) 013101.
[6] Deza J., Barreiro M. and Masoller C., to be published in *Eur. Phys. J. ST* (2013).
[7] Steinhaeuser K., Chawla N. and Ganguly A., *ACM SIGKDD Explor. Newslett.*, **12** (2010) 25.
[8] Hendrix W., Tetteh I., Agrawal A., Semazzi F., Liao W. and Choudhary A., *Community dynamics and analysis of decadal trends in climate data*, in *Proceedings of the Data Mining Workshops (ICDMW), 2011, IEEE 11th International Conference on Data Mining* (IEEE) 2011, pp. 9–14.
[9] Feng A., Gong Z., Wang Q. and Feng G., *Theor. Appl. Climatol.*, **109** (2012) 635.
[10] Paluš M., Hartman D., Hlinka J. and Vejmelka M., *Nonlinear Process. Geophys.*, **18** (2011) 751.
[11] Malik N., Bookhagen B., Marwan N. and Kurths J., *Clim. Dyn.*, **39** (2012) 971.
[12] Tsonis A., Swanson K. and Roebber P., *Bull. Am. Meteorol. Soc.*, **87** (2006) 585.
[13] Tsonis A. A. and Swanson K. L., *Phys. Rev. Lett.*, **100** (2008) 228502.
[14] Bialonski S., Horstmann M.-T. and Lehnertz K., *Chaos*, **20** (2010) 013134.
[15] Nicolis C. and Nicolis G., *Tellus*, **33** (1981) 225.
[16] Benzi R., Parisi G., Sutera A. and Vulpiani A., *Tellus*, **34** (1982) 10.
[17] Kistler R. *et al.*, *Bull. Am. Meteorol. Soc.*, **82** (2001) 247.
[18] Bandt C. and Pompe B., *Phys. Rev. Lett.*, **88** (2002) 174102.