178 Progress of Theoretical Physics Supplement No. 179, 2009

Climate Networks Based on Phase Synchronization Analysis Track El-Niño

Kazuko YAMASAKI,[1,*] Avi GOZOLCHIANI[2] and Shlomo HAVLIN[2]

[1] *Tokyo University of Information Sciences, Chiba 265-8501, Japan*
[2] *Minerva Center and Department of Physics, Bar Ilan University, Ramat Gan, Israel*

Eight networks, based on temperature records from four different geographical regions in two pressure levels, are created in resolute snapshots of time for the last 28 years. The links represent the level of robust phase synchronization between places in the interior of each regime. The number of links appears to be a sensitive measure of the El-Niño influence even on regimes far away from the El-Niño basin. A comparison between the statistical information of the phase synchronization network and the similar information obtained previously using cross correlation technique is provided.

§1. Introduction

Climate can be represented as a network of interacting heterogeneous special units.[1)] Information flows between these units (represented by nodes) in the form of heat transfer, wind, flow of water and transfer of other materials (represented by links). There are also global external changes which influence all the nodes in the climate network (for example — variation of the radiation from the sun). In the current view, the external influences as well as internal processes that occur within each of the nodes (such as precipitation) are considered to be the natural autonomic behavior of the individual node. The former form of interaction which relates the dynamics of two nodes is the direct coupling. A flow of heat or material between one node and another one through a third intermediate node is an indirect coupling. In general there can be more than just one intermediate node on the way. The influence usually has a time scale (e.g. a few days) and a preferred direction (e.g. from the western node to the southern one). These set a time delay.

Between most of the pairs of nodes in the climate network there exists an indirect coupling. This means that a few localized severe events (such as massive heat transfer between the ocean and the atmosphere in a restricted zone of the pacific) can, in principle be felt as a change of the coupling between two nodes outside this zone. These changes may be tracked by a rapid change in the correlation pattern between the two nodes.

On the basis of such insight our group has recently constructed a sensitive measure (the number of links in the network) that tracks the El-Niño influence around the world.[2),3)] This work, however, integrated only a small part of the available information (temperature records within two pressure levels, inside four representative geographical regimes, with a grid of measurements taken coarser than the one currently available). A further work that integrates the amount of information avail-

*) E-mail: yamasaki@rsch.tuis.ac.jp

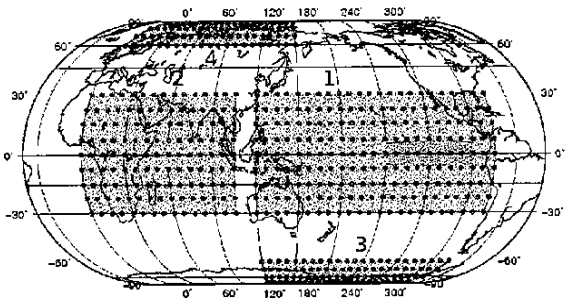

Fig. 1. The four geographical zones used for building the four climate networks studied. The dots represent the nodes of the network. The rectangular geographical area inside zone 1 shows the standard basin for which El-Niño/La-niña effects on temperature and pressure is significantly observed. (See Fig. 3(a).)

able is still needed, if one aims at a tool that provides better time resolution and detection power.

In the current paper we integrate our method into information about the phase synchronization between pairs of nodes. Techniques that track phase synchronization are known to be complementary to correlation based techniques in the information they provide about the interaction between systems. We therefore summarize here a comparison between the information about El-Niño that we collected using both methods.

§2. Network constructed by phase synchronization

Synchronization is a tendency of two interacting systems, a and b, to modify their pace towards some static functional relation.[4] This relation might be complete identity, and then the systems are said to be in complete synchrony. It is an established fact that even chaotic systems can achieve complete synchrony whilst preserving chaos.[5]

Phase synchronization is a more loose relation, very common in nature.[6)–10)] If a and b have an internal quantity, a phase $(\phi_a(t), \phi_b(t))$, which advances monotonically in time (when the systems are not attached), and of which small perturbations do not grow or shrink in time, a phase synchronized state is defined — a state in which $P(\phi_a - \phi_b)$, the distribution of the phase differences, is significantly different from a flat distribution.[11] We will follow the definitions adopted in Refs. 6) and 7) for the precise meaning of this criterion.

A phase synchronization measure yields information about the coupling between a and b which is complementary to correlations based methods. The range of applicability of methods based on the two phenomena is not a clear cut, but there are a few cases which help to draw the line between the two. First, if a and b are not coupled at all, but the measuring apparatus measures in one detector mostly signal

from a but also some small fraction of the signal from b and vice versa, cross correlation is expected to yield a clear indication of relation between the measurements, while phase synchronization measures will not indicate dependence at all.[11] A second case is one in which the signals of both a and b are of the form $Ae^{i\omega(t)t}$ where A is a constant (or very slowly varying) amplitude, and $\omega(t)$ is a frequency which changes slowly (slower than $1/\omega$) in time. In such a case, if a and b significantly interact, a phase synchronization will clearly show relation between the two, while cross correlation is expected to fail. Both cases are relevant to climate records which are based on models interpolating between measuring stations.

There are many time series analysis methods that quantify phase synchrony.[11] We chose a method that takes into account the statistical background as the method of preference.[6] In this method, phase is defined as the angle of the Gabor analytic signal[12] in the complex plane. Let us sketch the main steps.

Given the l-th record of time dependent temperature anomaly $u_l(t)$, we can compute its Hilbert transform $v_l(t)$:

$$v_l(t) = \frac{1}{\pi}\mathrm{PV}\int_{-\infty}^{\infty}\frac{u_l(\tau)}{t-\tau}d\tau. \tag{2·1}$$

The complex analytic signal is defined:

$$z_l(t) = u_l(t) + iv_l(t) = A_l e^{i\phi_l(t)}. \tag{2·2}$$

$\phi_l(t)$ is the phase time series of the l-th record. The phase difference of two records indexed by l and r is:

$$\varphi_{l,r}(\tau,t) = (\phi_l(t) - \phi_r(t+\tau)) \bmod 2\pi, \tag{2·3}$$

where we include a time delay τ between the two records. When τ is negative we will take the symmetric definition: $\varphi_{l,r}(-\tau,t) = -\varphi_{r,l}(\tau,t)$.

The next step is to compute the probability density function of $\varphi_{l,r}(\tau,t)$ by choosing a fine division of $[0,2\pi]$ to M bins of size $2\pi/M$, and then to evaluate the probability $p^k_{l,r}(\tau)$ that the phase difference is in the k-th bin. A synchronization index $\chi_{l,r}(\tau)$[11] is defined, using the Shannon entropy[13] $S_{l,r}(\tau)$:

$$\chi_{l,r}(\tau) = \frac{S_{\max} - S_{l,r}(\tau)}{S_{\max}}, \quad S_{l,r}(\tau) = -\sum_{k=1}^{M} p^k_{l,r}(\tau)\ln\left(p^k_{l,r}(\tau)\right). \tag{2·4}$$

By definition, the maximum entropy is $S_{\max} = \ln(M)$ and the range of the synchronization index is $0 \le \chi_{l,r}(\tau) \le 1$. $\chi_{l,r}(\tau) = 1$ means complete phase synchronization (also termed *phase lock*), in this case the distribution of the phase differences $p^k_{l,r}(\tau)$ is a Dirac-delta function.

Phase synchronization is finally quantified by a single number — the synchronization strength, $W_{l,r}$, which is the relation between the highest value of $\chi_{l,r}(\tau)$ and the background value met because of statistical fluctuations, but also because of inherent biases in the synchronization index:[6]

$$W_{l,r} = max\left(\chi_{l,r}(\tau) - mean\left(\chi_{l,r}(\tau)\right)\right)/std\left(\chi_{l,r}(\tau)\right). \tag{2·5}$$

where *max*, *mean*, and *std* are the maximal, the mean, and the standard deviation value of $\chi_{l,r}(\tau)$ within the range of τ, respectively.

The second quantity which characterizes a pair of records l, r is the time delay. The time delay $D_{l,r}$ between the two records is estimated by the τ value that yields the maximal $\chi_{l,r}(\tau)$.

Using a threshold value Q, we can make a network with a network matrix $\rho_{l,r}$.

$$\rho_{l,r} = \Theta(W_{l,r} - Q). \tag{2·6}$$

Next we propose some methods which are useful when we track the feature of the successive networks indexed by time stump y (a network at a time interval y, the network at the next interval $y+1$ ...). In some cases, the strength of the link $W^y_{l,r}$ goes above or below Q depending on time (it is now customary to call this behavior "blinking links"[3]).

In the first method we count the conjugate group of links — namely that group of links that do not blink. $n_k(y)$ is the number of links which appeared more than k times in a row before the time interval y (including its current appearance).

$$M^y_{l,r} = \sum_{n=0}^{y-1} \prod_{m=y-n}^{y} \rho^m_{l,r}, \quad n_k(y) = \sum_{l=0}^{N} \sum_{r=l+1}^{N} \Theta\left(M^y_{l,r} - k + 1\right). \tag{2·7}$$

Here k is the number of intervals a link has to survive in order to be included in the network, and N is the total number of links. The summand on the rhs of Eq. (2·7) represents the network adjacency matrix.

The second method counts the number of blinking links. This is more subtle since there are a lot of blinking patterns. We chose links that either appear, disappear and appear again (1,0,1) during three subsequent values of y, or have the opposite pattern (0,1,0). Let us denote the number of blinking links that follow these patterns $b(y)$.

§3. Results

We analyze daily temperature records taken from a grid of 7.5° resolution (available at *) in various geographical zones (shown in Fig. 1) measured on sea level and on a 500 mb pressure level. The measurements are taken for the years 1979–2007, for which 8 known El-Niño events occur (events A-E in Fig. 3, when C is actually composed of three adjacent small events and the last event is at the end of the period, thus not enabling analysis based on 365 days periods at this event). We have also included in the figure for completeness the main La-Niña events (pointed in the figure with small letters).

To avoid the trivial effect of seasonal trends we subtract from each day's temperature, the yearly mean temperature of that day. Specifically, if we take the temperature signal of a given site in the grid to be $\widetilde{T}^y(d)$, where y is the year and d is the day

*) NCEP Reanalysis data, NOAA/OAR/ESRL PSD, Boulder, Colorado, USA. http://www.cdc.noaa.gov

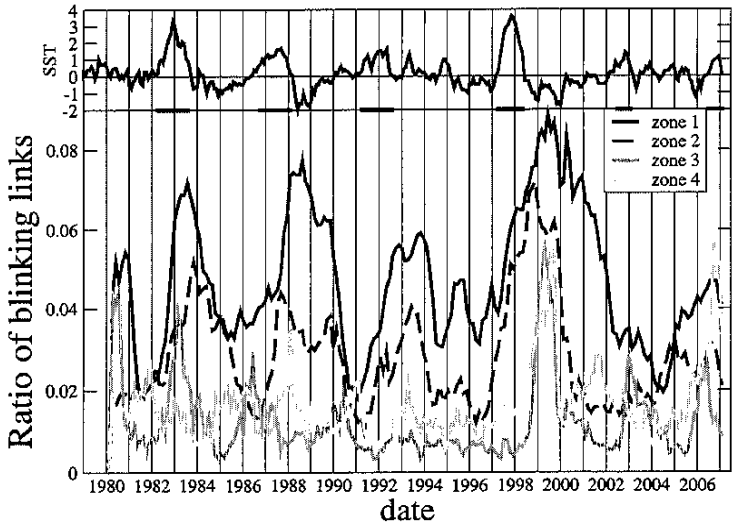

Fig. 2. Number of blinking links $b(y)$ as a function of time. The top panel shows the temperature based El-Niño index NINO3. The bottom panel shows the number of blinking links as a function of time that exhibit the patterns 1,0,1 or 0,1,0 in the four zones shown on Fig. 1.

(ranging from 1 to 365), the new filtered signal will be $T^y(d) = \widetilde{T}^y(d) - \frac{1}{Y}\sum_y \widetilde{T}^y(d)$ (where $Y = 29$ is the number of years available in the record).*)

As in previous works that employed cross correlation based techniques, the nodes of our network are the records of$T^y(d)$ from the different grid points.[2),3),15)] However, differently from these works, we use the synchronization function $\chi_{l,r}(\tau)$ in order to decide whether two nodes are linked together. We advance the beginning date y of the temperature time series by 50 days (and the finishing day is after 1 year), so we have snapshots of our networks in a 50 days resolution. The maximal time delay considered is $\tau_{\max} = 200$ days.

The variation of $W^y_{l,r}$ in time y is such that during El-Niño events its distribution is lowered Fig. 4. Since we seek for dynamical changes rather than static structure we pick a threshold on the synchronization measure, which is on the left end of the scale for normal times, but not so during El-Niño. So we chose $Q = 3.0$.

The important part of the story is the fact that the synchronization values that go below Q in El-Niño times, and which seem to be the hallmark of the event's influence, blink (appear and disappear) erratically during the event. To search exactly for such massive amount of fluctuations we developed two complementary methods. In the first method we count the number of survival links $n_k(y)$. In the current work, the y resolution is 50 days and $k = 5$, so $n_k(y)$ is the number of links that survive more than 250 days. The results shown below are not sensitive to the choice of k. The second method counts the number of blinking links.

In Fig. 3 we show the time dependent number of surviving links $n_k(y)$. For reference we show in the top panel (A) two standard indexes of El-Niño, based on

*) After filtering is finished, we can use overlapping year periods, s represents the index of a year period (from 1 to 200).

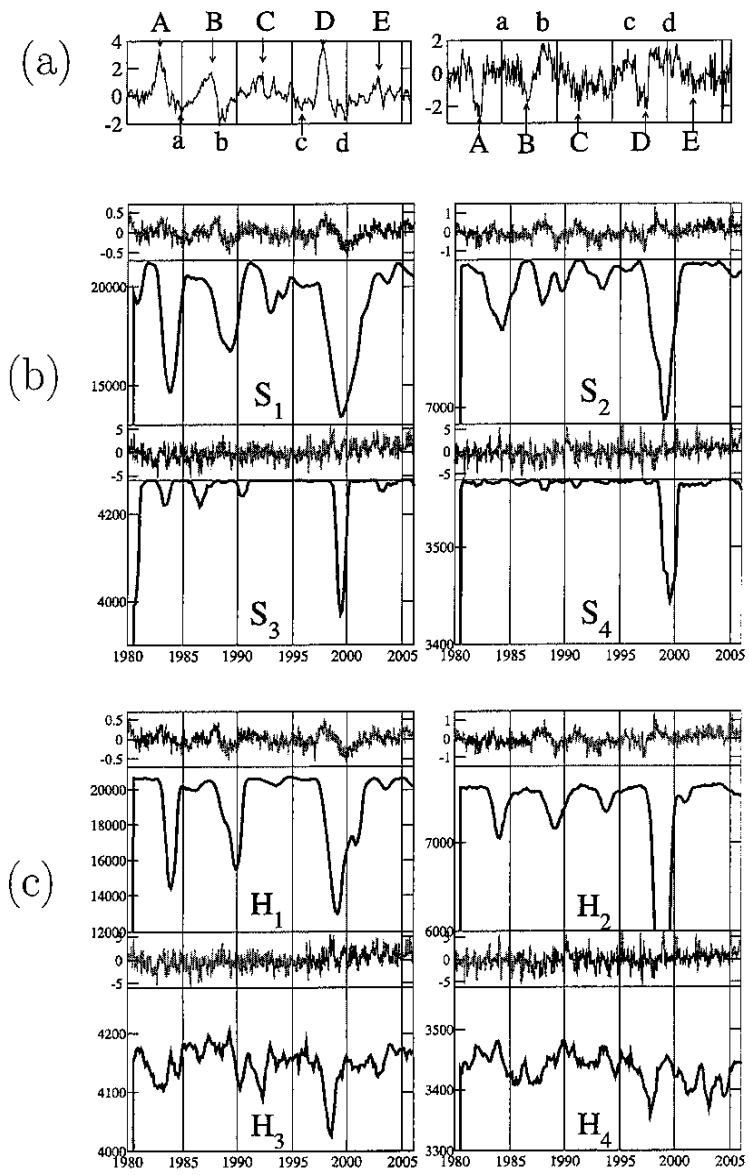

Fig. 3. (a) Mean sea surface temperature (left) in the standard basin shown in the rectangle (NINO3) inside zone 1 in Fig. 1 , and the difference in sea level pressure (right) between Tahiti and Darwin, both are standard indexes for El-Niño/La-niña (see e.g. 14)). (b) and (c) The upper curves represent the temperature anomaly series in zones (b) S_1, S_2, S_3, S_4 (temperature measured at surface) and (c) H_1, H_2, H_3, H_4 (temperature measured at 500 mb pressure level). The lower curves present by phase synchronization the number of links $n_k(s)$ that survive in the network as a function of time, for these same zones.*)

*) The date related to each point in the graph is the end of the k (which equals 5 in our case) subsequent periods that we take into account. The time resolution (of s) for which we compute the network is 50 days, which means that each point uses 250+365 days of information from its past. Because phase-difference shifts the data, we also use 200 points from the future of that point (which is our maximal shift).

temperature measurements (left) and pressure difference measurements (right) in the pacific. The El-Niño events appear in the form of strong peaks in the left panel, and strong values in the right one, also pointed by capital letters (A-E), measured in the standard basin region inside zone 1 of Fig. 1.*) On the bottom of each of the eight panels the function $n_k(y)$ is shown, for both sea level $(S_1 - S_4)$ and 500 mb pressure level $(H_1 - H_4)$. Each minimum in these plots corresponds to an El-Niño event. In general, there is an excellent agreement between the results obtained here, and previous results obtained with a cross correlation method,[2] but the details are not completely similar.

The most significant response is seen on zone S_1, which includes the El-Niño basin. In this zone the cross correlation based technique and the phase synchronization one yield an almost exactly identical picture, apart from the 2002 event (E) which in the cross correlation work was completely masked by the effect of the 1997 (D) event, while in Fig. 3 it is nicely distinguishable.

In zone S_3 the 1987 (B) event is also seen much more significantly on the current phase synchronization based method. However, the 2002 (E) event has a stronger effect on the cross correlation based network.

The rest of the zones, including all the measurement at 500 mb level show similar behavior. In zones H_3, H_4 the phase synchronization $n_k(y)$ is more noisy, but the information that is supplied by the two methods is similar.

Above all the panels of Fig. 3 we also show the mean temperature anomaly of the respective zone and height level. One can see that the information collected using a network approach is very different from a direct measurement of the temperatures, which is more noisy and in general cannot directly indicate the influence of El-Niño.

The count of blinking links $b(y)$ in Fig. 2 shows peaks that correspond to El-Niño events. The response observed here is, in general, of better quality than the response we had in cross correlation based method.[3] This is most evident in zones 3 and 4.

A somewhat peculiar fact is that the 1987 (B) event (which was a relatively long one) is seen as a close double peak in zones 2 and 3.

Another insight is met by observation of the distribution of the synchronization strength $W^y_{l,r}$. The left panel of Fig. 4 shows how this distribution changes dramatically between normal times (dashed) and the A-E El-Niño events. On the right panel we show that the links that later on weaken as a result of the El-Niño event have a-priori a slightly weaker distribution. This is also with an excellent agreement with the results known from the cross correlation method.

A link that survives a 250 days period (like these links that we count on the first method) has an inherent special property. These kinds of link have a much shorter time delays, and they were termed SC (strongly connected) links in a previous paper.[3] On the right panel of Fig. 5 this difference in the distribution of SC links' time delays is explicitly shown. In this figure we show the difference between the two probability distribution functions in normal times, but it persisted also when we consider El-Niño and La-Niña times. The distribution of all time delays, however,

*) This partial region is the standard region from which the El-Niño effect is determined and quantified.

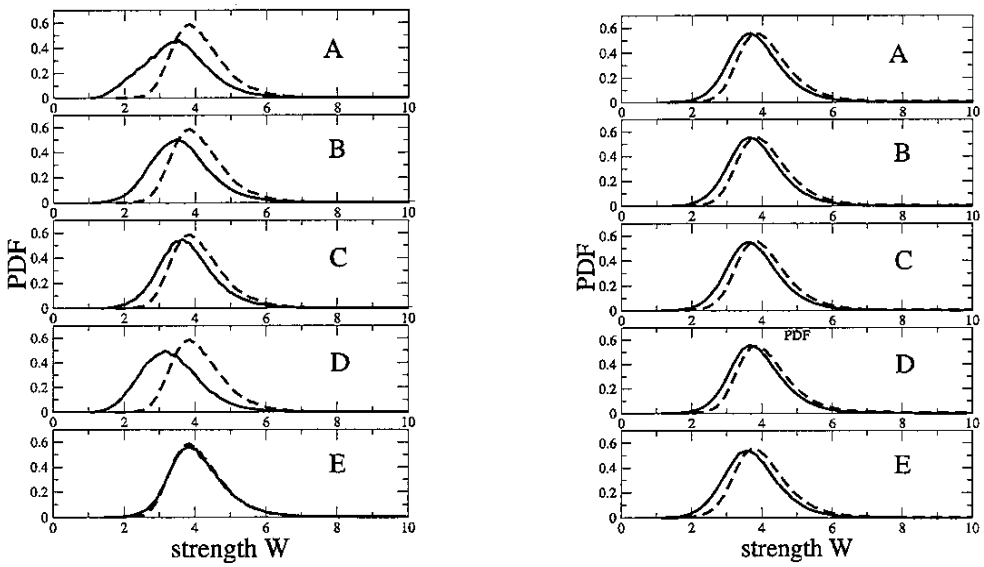

Fig. 4. The distribution of correlation strength $W^s_{l,r}$ between pairs of records in zone 1 (see Fig. 1). The five rows correspond to the five El-Niño events, which are the most noticeable in zone 1 (Fig. 3). Left (A-E): The dashed curves describe the distribution of $W^s_{l,r}$ in reference periods, where the influence of El-Niño on the network does not show. The solid curves describe the distribution of $W^s_{l,r}$ for five El-Niño events. Right (A-E): Two distributions of $W^s_{l,r}$ in the reference periods. The solid curves describe the distributions of the links that do not survive during 250 days in one of the El-Niño events , and the dashed curves are the distribution of the ones that survive during 250 days in the considered event.*)

resides in all periods, disregarding the existence of either an El-Niño or a La-Niña event.

As a final step in the comparison we show in Fig. 6 (left) the overall agreement of correlation strength and synchronization strength. This indicates that the two methods are generally in their applicability region, but they do supply a bit different information. As explained in §2, this kind of agreement is not a trivial outcome, and there are systems in which it completely fails to exist.

We also show in Fig. 6 (right) the agreement between delay times. One should have in mind that when $W^y_{l,r}$ is small, there are a lot of peaks, representing a lot of competing time scales that control the coupling between the l-th and the r-th grid

*) The curves (left) are the distributions of the strength (or delays) of all pairs of the nodes in 5 periods. (The number of the elements for a distribution is $N(N-1)/2 \times 5$ of $W^s_{l,r}$ (or $D^s_{l,r}$).) The dotted curves(left) in different rows are same and calculated by the data in normal periods. The solid curves (left) are calculated by the data in one of the El-Niño/normal periods. For calculation of the distribution in right panel, we separate the links of normal times into two groups, a group that will break in one of the events and another group that will reside in this event. The links that are said to reside in the event 5 continuous periods of 50 days (250 days) in one of El-Niño/normal periods. For the reference, we used 30 reference periods where the influence of El-Niño on the network does not show. The dotted curves (right) are the distributions of the strength (or delay) of the links from the 30 periods that will reside in an event. The solid curves (right) are the distribution of the strength (or delay) of the links in the same 30 periods that will break in the event.

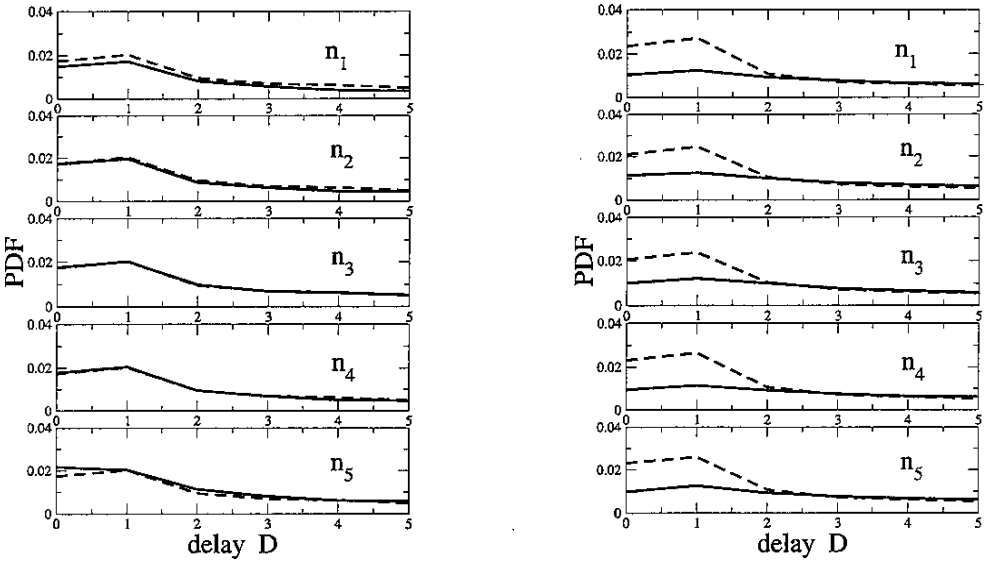

Fig. 5. Distribution of time delays $D^s_{l,r}$ in zone 1 (see Fig. 1). The five rows correspond to five chosen periods where the influence of El-Niño on the network does not show (Fig. 3). Left (n_1-n_5): The dashed curves describe the distribution in 30 reference periods. *) The solid curves describe the distribution for the five periods which we picked. Right (n_1-n_5): Two distributions of $D^s_{l,r}$ in the reference periods. The solid curves are the distributions of the links that do not survive during 250 days in one of the five chosen periods, and the dashed curves are the distribution of the ones that survive during 250 days.*)

points (see Ref. 3) for a discussion about the intermediately connected, IC, links), as well as peaks that come from noise. Such peaks are not expected to fall in the same place using different methods. However the diagonal density is quite impressive, especially in the long term delays (hundreds of days). It reminds us that although the climate system is chaotic, some parts of the patterns that exist in the system may reside for a very long period.

§4. Conclusion

We have tested how phase synchronization based information may improve a recently suggested technique for tracking the El-Niño influence around the world. We followed the dynamical changing of a network based on phase synchronization between temperature measurements. This was done for 4 zones around the globe, and in two pressure levels.

The structural changes obtained for the phase synchronization based networks are in general in good agreement with the structural changes obtained from the cross correlation based network. These include a general dynamical decrease of the number of links during El-Niño (which is also indicated in the work by Tsonis and Swanson.[15]) However, the dynamical process of breaking was shown by us in a previous, cross correlation based work.[2]). Individual breaking links are usually fluctuating back

*) See the footnote *) on page 185.

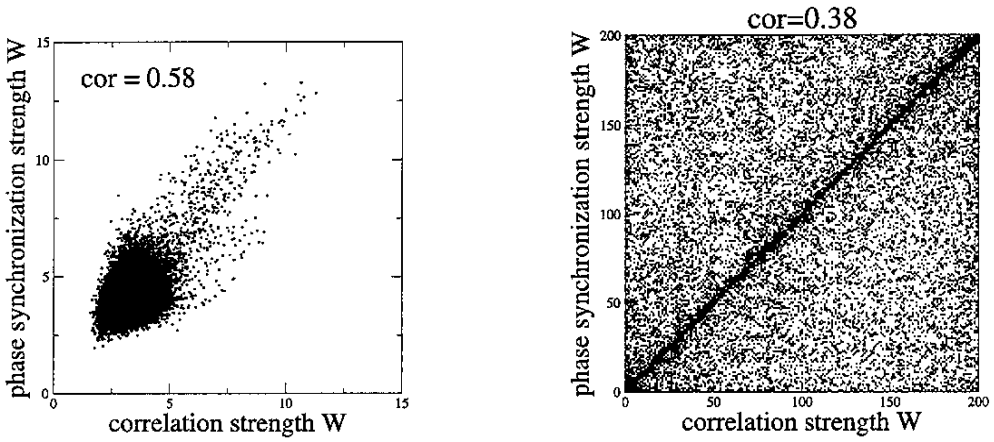

Fig. 6. Left: A scatter plot of phase synchronization strength values $W^y_{l,r}$ vs the cross correlation strength obtained in a previous work, for non-El-Niño period. Right: A scatter plot of the time delay τ obtained using a phase synchronization method Vs. the same time delay obtained using a cross correlation method in a previous work.

and forth between strong and weak synchronization, a phenomenon that has been previously termed "blinking links".[6] The blinking links have generally longer time delays, and are distributed differently even in normal times.

The two methods are almost identical with regard to the stable strongly connected links (see Fig. 6), while they show a complementary information for the intermediately connected links (see Fig. 6, as well as the comparative discussion in §3). Specifically this means that for weaker events, or a sequence of events that occur shortly one after the other, sometimes the phase synchronization based method indicates the details of the event better, and sometimes does the cross correlation based. This is seen mainly for the sequence of the small events that began on 1991, as well as for the 2002 event that occurred right after the influences of the famous giant 1997 event decayed. This work was supported by the Academic Frontier Joint Research Promoting Center of Tokyo University of Information Sciences.

References

1) A. A. Tsonis and K. L. Swanson and P. J. Roebber, Bull. Am. Meteorol. Soc. **87** (2006), 585.
2) K. Yamasaki, A. Gozolchiani and S. Havlin, Phys. Rev. Lett. **100** (2008), 228501.
3) A. Gozolchiani, K. Yamasaki, O. Gazit and S. Havlin, Europhys. Lett. **83** (2008), 28005.
4) L. Kocarev and U. Parlitz, Phys. Rev. Lett. **76** (1996), 1816.
5) S. Boccaletti, J. Kurths, G. Osipov, D. L. Valladares and C. S. Zhou, Phys. Rep. **366** (2002), 1.
6) A. Gozolchiani, S. Moshel, J. Hausdorff, E. Simon, J. Kurths and S. Havlin, Physica A **366** (2006), 552.
7) D. Rybsky, S. Havlin and A. Bunde, Physica A **320** (2003), 601.
8) T. J. Walker, Science **166** (1969), 891.
9) J. Buck, Q. Rev. Biol. **63** (1988), 265.
10) R. Bartsch, J. W. Kantelhardt, T. Penzel and S. Havlin, Phys. Rev. Lett. **98** (2007), 054102.
11) M. Rosenblum, A. Pikovsky, J. Kurths, C. Schäfer and P. A. Tass, *Handbook of Biological Physics* (Elsevier Science, 2001).
12) D. Gabor, J. IEE London **93** (1946), 429.

13) C. E. Shannon, The Bell System Technical Journal **27** (1948), 379.
14) H. A. Dijkstra, *Nonlinear Physical Oceanography* (Kluwer Academic Publishers, 2000).
15) A. A. Tsonis and K. L. Swanson, Phys. Rev. Lett. **100** (2008), 228502.