

LINC

Learning about Interacting Networks in Climate

Marie Curie Initial Training Networks (ITN)

FP7- PEOPLE - 2011- ITN

Grant Agreement No. 289447

WP1: Network Construction and Analysis

Deliverable D1.4

Toolbox for the identification of most important teleconnections and their stability

Jürgen Kurths, PIK, Potsdam

Release date: 1 November 2014

Status: Public



Executive Summary

This document represents the deliverable **D1.4 - Toolbox for identification of most important teleconnections and their stability** and involves a software package which was used for the results in Deliverable 1.5. The software is available on the LINC website (<http://www.climatelinc.eu/results/software/>).

The aim of this deliverable is to provide a toolbox for climate network reconstruction and analysis, which allows to identify the most important teleconnections in climate networks and estimate their stability with time. In particular, first, the deliverable includes software for the Climate Network Reconstruction using different methods (*Pearson correlation, Mutual information, Event synchronization, Weighted networks*), which allows to calculate networks measures (*degree, area weighted connectivity, betweenness, clustering coefficient*), in order to use them for identification of the teleconnections. Second, it includes *Common Component Correlation Function* method, which can be used for estimating stability of the teleconnections with time. Third, the deliverable includes one of the most important tools for climate network analysis and detection of teleconnections in climate – that of the Visualization tool for Climate Networks. Such visualization tool was developed in the LINC partner PIK. This software can improve the presentation of the results of climate network analysis, as well as can be useful for detection of the most important teleconnections in climate. The software can be used by LINC fellows (ERs and ESRs), and the link to the software can be found on the LINC website <http://www.climatelinc.eu/results/software/>.

Impact on other WPs: This toolbox was created as a part of the research undertaken by LINC fellows (ERs and ESRs) of the WP1 – Network reconstruction and analysis and WP2 – Interacting networks. Also, it can be beneficial for LINC fellows from other workpackages: WP3 – Natural Climate variability, WP4 - Future Climate Change and WP5 – Tipping points in the Climate system, since the methods and techniques provided in the toolbox can be applied to a variety of teleconnections in Climate system on different time scales, and therefore, can help to better understand natural climate and transitions between different climate regimes.

It is planned that this toolbox will be updated and added to throughout the course of the project. The web pages will be maintained so as the latest version of the toolbox is publicly accessible at all times.

Target audience of the deliverable: all partners, ESRs and ERs of the LINC project.

Deliverable Identification Sheet

Grant Agreement No.	PITN-GA-2011-289447
Acronym	LINC
Full title	Learning about Interacting Networks in Climate
Project URL	http://climatelinc.eu/
EU Project Officer	Lucia PACILLO

Deliverable	D1.4 - Toolbox for identification of most important teleconnections and their stability.
Work package	WP1 – Network construction and analysis

Date of delivery	Contractual	M 30	Actual	31-May-2014
Status	version. 1.00		final <input checked="" type="checkbox"/>	draft <input type="checkbox"/>
Nature	Prototype <input type="checkbox"/> Report <input type="checkbox"/> Database <input checked="" type="checkbox"/> <input type="checkbox"/>			
Dissemination Level	Public <input checked="" type="checkbox"/> Consortium <input type="checkbox"/>			

Authors (Partner)	Juergen Kurths		
Responsible Author	Juergen Kurths	Email	kurths@pik-potsdam.de
	Partner	PIK	Phone <optional>

Abstract (for dissemination)	This deliverable is a cover note for deliverable D1.4 - Toolbox for identification of most important teleconnections and their stability.	
Keywords	Climate, networks construction and analysis	

Version Log			
Issue Date	Rev No.	Author	Change(s)
31-05-2014	001	J. Kurths	Final report

TABLE OF CONTENTS

EXECUTIVE SUMMARY.....	i
1 INTRODUCTION.....	1
2 TOOLBOX CONTENT.....	1
2.1 Software for Climate network construction and analysis.....	1
2.2 Software for identification of the stability of the teleconnections.....	7
2.3 Software for visualization of the Climate Networks....	8
REFERENCES.....	10

1. INTRODUCTION

In this document we summarize some of the tools from the complex network theory, climate network theory, statistics and information theory which can be used for the climate network reconstruction, analysis and identification of the most important teleconnection in climate and their stability. A toolbox is written in Python with implementation of C, in order to reduce the calculation time, and requires Python to conduct such analysis. Most of the toolbox is included in PyUnicorn package which is available on the LINC website <http://www.climatelinc.eu/results/software/>. Other part of the toolbox, which includes such tools as *Network Construction using Event synchronization method*, *Network Construction using Weighted networks* and *Common Component Correlation Function*, is provided in separate software packages, as they require a separate procedure to run the codes. This part of the deliverable is provided in the attachment and also can be found on the LINC website. Finally, the last part of the deliverable consists of the Visualization tool for Climate Networks.

2. TOOLBOX CONTENT

2.1 Software for Climate network construction and analysis

Finding similarities between the dynamics of the nodes of the climate network is an interesting and important problem. Depending on what kind of connection between the nodes one look at, in case when one can not defined a physical connection (e.g. through model), the only possible way to establish connections between the nodes is to establish statistically defined connections. There are several commonly used methods to construct a climate network, depending on the correlation that measures similarity of time series of nodes. In this deliverable, a list of the main methodologies is presented for construction of the climate networks from statistical dependencies, which is accompanied with a software tools to use them. Presented in the software methods for climate network construction are:

1. Method for network construction from correlation matrix of time series based on Pearson correlation
2. Method for network construction using mutual information between the pairs of time series

3. Event synchronization approaches
4. Weighted climate networks

Method for network construction from correlation matrix of time series based on Pearson correlation

Correlation matrix as a base for climate network can be constructed using Pearson correlation coefficient. Let us consider grid point i and j - nodes of the climate network. Let us define $x_i(t)$ and $y_j(t)$ as time series of the grid points i and j . Then, Pearson correlation coefficient for these times series is determined in a following way:

$$R_{ij} = \frac{Cov(x_i(t), y_j(t))}{S_{x_i(t)} S_{y_j(t)}} \quad (2)$$

where R_{ij} - is a Pearson correlation, $Cov(x_i(t), y_j(t))$ - is the covariance of time series $x_i(t)$ and $y_j(t)$, and defined as:

$$Cov(x_i(t), y_j(t)) = \frac{\sum_{t=1}^{t=N} (x_i(t) - \langle x_i(t) \rangle)(y_j(t) - \langle y_j(t) \rangle)}{N - 1} \quad (3)$$

where $\langle x_i(t) \rangle$ and $\langle y_j(t) \rangle$ are the arithmetic mean for the time series $x_i(t)$ and $y_j(t)$ respectively, and N - number of values in the time series (times series $x_i(t)$ and $y_j(t)$ must have the same length), $S_{x_i(t)}$ and $S_{y_j(t)}$ - are standard deviations of the $x_i(t)$ and $y_j(t)$ respectively, and defined as:

$$S_{x_i(t)} = \sqrt{\frac{\sum_{t=1}^{t=N} (x_i(t) - \langle x_i(t) \rangle)^2}{N - 1}} \quad (5)$$

$$S_{y_j(t)} = \sqrt{\frac{\sum_{t=1}^{t=N} (y_j(t) - \langle y_j(t) \rangle)^2}{N - 1}} \quad (6)$$

Repetition of the described procedure for all pairs of nodes of the climate network, yields a correlation matrix which can be used for the construction of the undirected cli

mate network. Formula (6) represents so-called zero-lag Pearson correlation coefficient, because two time series which are compared are taken at the same time (e.g. dynamics in the nodes of the climate network is compared without time delay between the nodes). Examples of zero-lag Pearson correlation coefficient as a measure of similarity between the nodes of the climate networks can be found in Donges et al. (2009a); Stolbova et al. (2014); Tupikina et al. (2014); Molkenhain et al. (2014).

The code for the climate network construction using Pearson correlation is included in PyUnicorn package which is available on the LINC website.

Method for network construction using mutual information between the pairs of time series

For measuring linear relationship between the times series of the climate data one can use Pearson correlation coefficient or Spearman rank order correlation coefficient. However, climate phenomena show strongly nonlinear relationship between each other, as well as climate variables (Maraun and Kurths, 2005). To uncover these nonlinear relationships a number of nonlinear measures of statistical interdependence have been used in climate science. The most well-known nonlinear method for the climate network construction is Mutual information:

$$M_{ij} = \int_{m,n} p_{ij}(m,n) \log \frac{p_{ij}(m,n)}{p_i(m)p_j(n)}, \quad (8)$$

where $p_i(m)$ - is a probability density function (PDF) of the time series $x_i(t)$ in the grid point i , $p_j(n)$ - is a PDF of the time series $y_j(t)$ in the grid point j , and $p_{ij}(m; n)$ - is the joint PDF of a pair $(x_i(t), y_j(t))$ (Kraskov et al., 2004; Cellucci et al., 2005; Khan et al., 2007). By definition, M_{ij} is symmetric, so that $M_{ij} = M_{ji}$. The standard unit of measurement of the Mutual information is a bit, if logarithms to base 2 are used.

Probability densities of the time series $x_i(t)$ and $y_j(t)$ and joint probability density can be estimated using a simple histogram approach with equally sized bins for all pairs i, j of grid points (Donges et al., 2009a).

M_{ij} is a measure of the degree of statistical interdependency: if the two time series are independent, $p_{ij}(m; n) = p_i(m)p_j(n)$ and $M_{ij} = 0$.

Correlation matrix based on the Mutual information can be obtained by calculating Mutual information for each pair of the grid points. It is a symmetric matrix, so it can be used for the construction of the undirected climate network. Examples of climate networks obtained using Mutual information nonlinear measure of statistical interdependencies of the time series of the grid points can be found in Donges et al. (2009b,a); Deza et al. (2013b,a).

The code for the climate network construction using Mutual information is included in PyUnicorn package which is available on the LINC website and detailed description of the package can be found there as well.

Event synchronization approaches

Not all climate data are smooth and continuous as temperature or pressure fields. Sometimes, in climate science there is a need to study event - like time series, such as rainfall (which may happen or not, and has various intensities), fires, different extreme events (precipitation, fires), and etc. Considering event - like time series, it is important to choose an appropriate method to infer similarity of dynamics between different nodes. One possible way is to analyze event - like time series, is called event synchronization (ES). ES is a method to measure synchronization and time - delay patterns between signals. It was introduced by Quiroga et al. (2002) for Electrocardiograms of rat and human brain, and modified and used as a similarity measure for the construction of a climate network of extreme rainfall events by Malik et al. (2010, 2011); Boers et.al. (2013), and Stolbova et al. (2014).

This method has advantages over other time-delayed correlation techniques (e.g., Pearson lag-correlation), specifically for studying precipitation data, as it allows us to define extreme event series of rainfall, depending on the kind of extreme, and uses a dynamic time delay. The latter refers to a time delay that is adjusted according to the two time series being compared, which allows for better adaptability to the region of interest. Another advantage of this method is that it also can be applied to a non-gaussian and event-like data set.

Detailed description of the method can be found in Deliverable 1.5 and in the works of Quiroga 2009, Malik et.al 2010, Boers et. al. 2013 and Stolbova et. al. 2014. This method requires pre-processing of the climate data in an event-like way, so, that occurrence of the event is counted as 1, in respect with homogenous time discretization, and 0 - otherwise.

The code for the Event synchronization *ES.py*, can be found of the link website, under Software. The code is written in Python with partial implementation of C, and requires cluster for the parallel calculation of the results.

Weighted climate networks

Climate networks do not necessarily have to be unweighed. Thresholding procedure of the correlation matrix to obtain an adjacency matrix inevitably leads to the loss of so-called "weak" links. However, in case when one is interested in all links of the climate network, omitting thresholding procedure leads to weighted climate network.

A way to construct a weighted climate network which is based on the lagged optimal correlations, was introduced by Yamasaki et al. (2008), and modified by Gozolchiani et al. (2011). In the software the weighted climate network is constructed using modified version described in Gozolchiani et al. (2011) and Wang et. al. (2013).

This method for the weighted climate network construction implies measuring lagged Pearson correlation coefficient between two grid points i and j : $R_{ij}(x_i(t), y_j(t + t))$, for the range of $[-t_{\max} \leq t \leq t_{\max}]$. Value of t_{\max} is determined using characteristic time scales of the climate phenomenon which is aimed to be described by the weighted climate network. Then, the optimal lag can be determined, which corresponds to the maximum Pearson correlation coefficient. Therefore, for each pair of the grid points i and j , there are two values: maximum correlation value and optimal lag.

Gozolchiani et al. (2011) proposed the modification of the Yamasaki et al., (2008) method, which includes the following normalisation of the weight of the link in the weighted climate network:

$$W_{ij} = \frac{\max(R_{ij}(x_i(t), y_j(t + t_{\text{optimal}}))) - \text{mean}(R_{ij}(x_i(t), y_j(t + t_{\text{optimal}})))}{\text{std}(R_{ij}(x_i(t), y_j(t + t_{\text{optimal}})))}, \quad (14)$$

where W_{ij} - is a weight of the link of the weighted climate network, $[-t_{\max} \leq t \leq t_{\max}]$.

This modification of the link weight normalization allows to separate "real" optimal correlation by giving them higher weight than "background" correlation.

The code for the Weighted climate networks *Weighted_CN.py*, can be found of the link website, under Software. The code is written in Python with partial implementation of C, and requires cluster for the parallel calculation of the results.

Network measures

Climate networks constructed using one of the previously described methods can be analysed using various network measures. Filtering of via network measures can help to identify the most important teleconnections in climate (e.g. Donges et. al. 2009, Stolbova et. al., 2014). PyUnicorn package allows to calculate network measure as perform a filtering of the most important links. It allows calculation of such network measures as degree, betweenness, the maximal geographical link length, area weighted connectivity, clustering and etc.

For example, the degree of node j (geographical site) in the network gives the number of all nodes in which extreme events are synchronized with extreme events in node j (e.g. how many links each individual node has). Nodes with high degree are usually referred to as supernodes or hubs. In this study, we extend this term to the regions of spatially adjacent nodes with high degree.

In order to define the betweenness of a node in the network, let us assume that moisture travels through rainfall network via the shortest path. The shortest path – is the way to go from one node in the network to another using a minimal number of links. For given nodes i and j in the network, S_{ij} shortest paths exist. If a large fraction of the shortest paths from a node i to a node j pass through node v , then node v is an important mediator for the moisture transport in the network. The physical interpretation of betweenness is that it indicates the moisture pathways through the network of extreme rainfall. As a result, patterns with high betweenness can play a crucial role in understanding the mechanism of information transport through the climate network.

The maximal (average) geographical link length (L_j) of a node is the geographically longest (average) link that this node has.

Table 2: Examples of network measures

Examples of network measures		
	Formula	Definition
Degree, D_j	$D_j = \frac{\sum_{i=1}^N A_{ij}}{N-1}$	N – is the total number of nodes
Betweenness, B_v	$B_v = \sum_{i \neq j \neq v \in \{V\}} \frac{\sigma_v(i, j)}{\sigma(i, j)}$	$\sigma_v(i, j)$ – is the number of the shortest paths between i and j passing through v , $\sigma(i, j)$ – is the total number of shortest paths between i and j .
Maximal geographical link length, L_{ij}	$L_{ij} = \alpha_{ij} A_{ij} R$	α_{ij} – is angular geographical distance R – is radius of the Earth, 6371.009

2.2 Software for identification of the stability of the teleconnections

Software for identification of the stability of the teleconnections includes newly developed technique, which allows to track evolution of the climate networks on different time scales - the common component evolution function (CCEF). This tool characterizes network development over time, and is developed to track changes in the links of the evolving in time networks and allows to determine episodes of the breaking down of the links. These episode of breaking down of the links in the climate networks allow to estimate the stability of all links in the Climate Network, and, most importantly, the stability of teleconnections. An example of the application of this technique to investigate the teleconnection between such climate phenomena as the

Indian Summer Monsoon and El - Niño -Southern Oscillation, and formulation of the method can be found in Deliverable D - 1.5 and Tupikina et.al. 2014.

The link to the software can be found on <http://www.climatelinc.eu/results/software/>.

2.3 Software for visualization of the Climate Networks

Visualization is an established flexible tool to analyze climate related data and to communicate climate research findings to scientific community and public. However, there are multiple challenges: the very large and heterogeneous data sets, the multiple tasks to be performed and the underlying complexity of the scientific knowledge. However, these problems can be solved by appropriate visualization tools and techniques which allow to analyze large climate data and visualize Climate networks, in order to identify the most important teleconnections in climate. LINC partner PIK provides such visualization tool that can be widely used by researcher for the Climate data and Climate Network analysis. The link to the software is available on the link website <http://www.climatelinc.eu/results/software/>. An example of climate network visualization is presented on Figures 1 and 2.

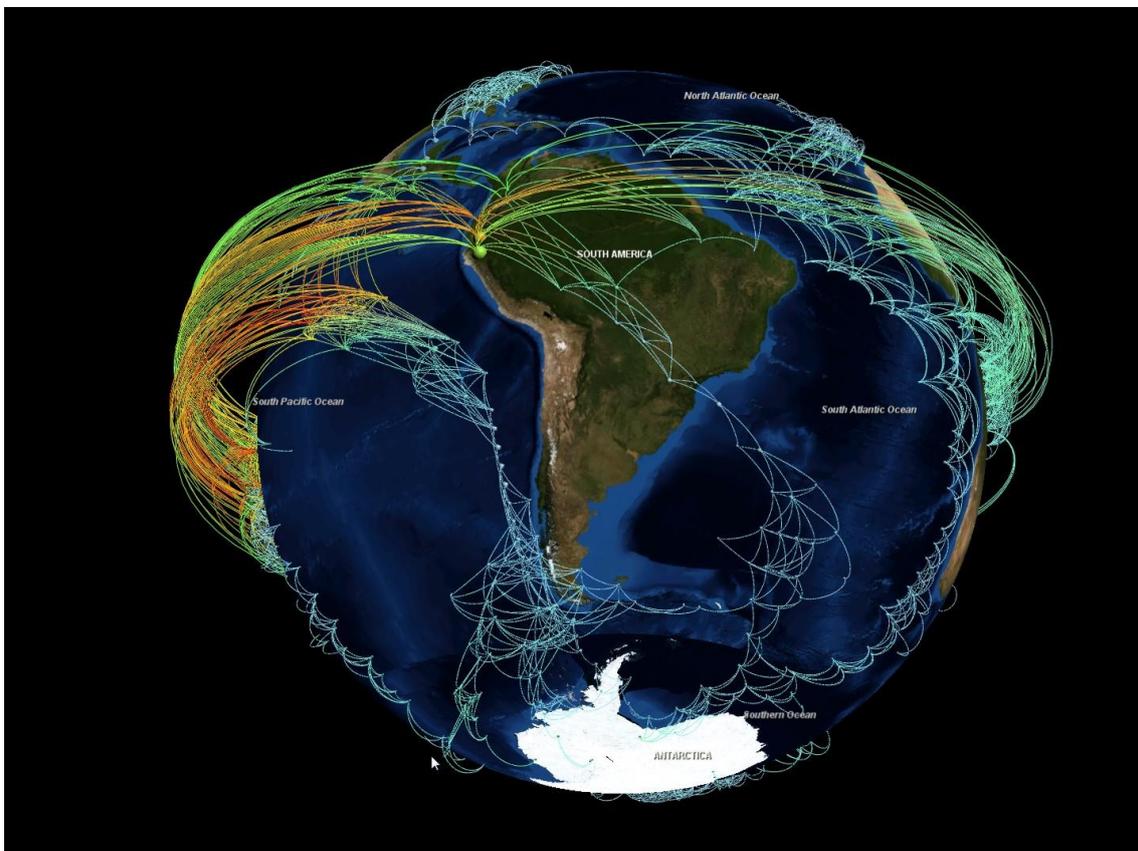


Figure 1. Spherical node link representation of a global climate temperature network

with 3D arcs, depicting both long distant teleconnections and regional structures.



Figure 2. Spherical node link representation of a regional climate extreme rainfall synchronization (Indian Summer monsoon region) network with 3D arcs, depicting both long distant teleconnections and regional structures.

REFERENCES

- [1] PyUnicorn Package: <http://www.pik-potsdam.de/~donges/pyunicorn/>
- [2] Data Visualization, <https://www.pik-potsdam.de/institute/archive/scientific-departments/data-computation/sdm/data%20visualization>
- [3] V. Stolbova, P. Martin, B. Bookhagen, N. Marwan, J. Kurths, Topology and seasonal evolution of the network of extreme precipitation over the Indian subcontinent and Sri Lanka, *Nonlin. Processes Geophys.*, 21, 901–917, 2014, [doi:10.5194/npg-21-901-2014](https://doi.org/10.5194/npg-21-901-2014)
- [4] L. Tupikina, K. Rehfeld, N. Molkenhain, V. Stolbova, N. Marwan, and J. Kurths, Characterizing the evolution of climate networks, *Nonlin. Processes Geophys.*, 21, 705–711, 2014, [doi:10.5194/npg-21-705-2014](https://doi.org/10.5194/npg-21-705-2014)
- [5] N. Molkenhain, K. Rehfeld, V. Stolbova, L. Tupikina, and J. Kurths, On the influence of spatial sampling on climate networks, *Nonlin. Processes Geophys.*, 21, 651–657, 2014, [doi:10.5194/npg-21-651-2014](https://doi.org/10.5194/npg-21-651-2014)
- [6] Donges, J. F., Zou, Y., Marwan, N., and Kurths, J.: The backbone of the climate network, *Europhys. Lett.*, 87, 48007, [doi:10.1140/epjst/e2009-01098-2](https://doi.org/10.1140/epjst/e2009-01098-2), 2009a.
- [7] Donges, J. F., Zou, Y., Marwan, N., and Kurths, J.: Complex networks in climate dynamics, *Eur. Phys. J. Spec. Top.*, 174, 157–179, 2009b.
- [8] Gozolchiani, A., Yamasaki, K., Gazit, O., and Havlin, S.: Pattern of climate network blinking links follows El Niño events, *Europhys. Lett.*, 83, 28005, [doi:10.1209/02955075/83/28005](https://doi.org/10.1209/02955075/83/28005), 2008.
- [9] Gozolchiani, A., Havlin, S., and Yamasaki, K.: The Emergence of El Niño as an Autonomous Component in the Climate Network, *Phys. Rev. Lett.*, 107, 14, [doi:10.1103/PhysRevLett.107.148501](https://doi.org/10.1103/PhysRevLett.107.148501), 2011.
- [10] Malik, N., Bookhagen, B., Marwan, N., and Kurths, J.: Analysis of spatial and temporal extreme monsoonal rainfall over South Asia using complex networks, *Clim. Dynam.*, 39, 971–987, 2011.
- [11] Tsonis, A. and Roebber, P.: The architecture of the climate network, *Physica A*, 333, 497–504, 2004.
- [12] Tsonis, A. A. and Swanson, K.: Topology and Predictability of El Niño and La Niña Networks, *Phys. Rev. Lett.*, 100, 228502, [doi:10.1103/PhysRevLett.100.228502](https://doi.org/10.1103/PhysRevLett.100.228502), 2008.
- [13] Tsonis, A. A., Swanson, K. L., and Roebber, P. J.: What Do Networks Have to Do with Climate?, *B. Am. Meteorol. Soc.*, 87, 585–595, 2006.

- [14] Tsonis, A. A., Swanson, K. L., and Wang, G.: On the Role of Atmospheric Teleconnections in Climate, *J. Climate*, 21, 2990–3001, 2008.
- [15] Wang, Y., Gozolchiani, A., Ashkenazy, Y., Berezin, Y., Guez, O., and Havlin, S.: Dominant Imprint of Rossby Waves in the Climate Network, *Phys. Rev. Lett.*, 111, 138501, [doi:10.1103/PhysRevLett.111.138501](https://doi.org/10.1103/PhysRevLett.111.138501), 2013.
- [16] Yamasaki, K., Gozolchiani, A., and Havlin, S.: Climate Networks around the Globe are Significantly Affected by El Niño, *Phys. Rev. Lett.*, 100, 228501, [doi:10.1103/PhysRevLett.100.228501](https://doi.org/10.1103/PhysRevLett.100.228501), 2008.
- [17] Yamasaki, K., Gozolchiani, A., Havlin, S., and Gan, R.: Climate Networks Based on Phase Synchronization Analysis Track El- Niño, in: *Progress of Theoretical Physics Supplement*, 178–188, [doi:10.1143/PTPS.179.178](https://doi.org/10.1143/PTPS.179.178), 2009.