

## ***Interactive comment on “Par@Graph – a parallel toolbox for the construction and analysis of large complex climate networks” by H. Ihshaish et al.***

**H. Ihshaish et al.**

hisham.ihshaish@uwe.ac.uk

Received and published: 15 April 2015

Dear J.F. Donges (Referee):

Thank you for the invaluable input, remarks and suggested corrections. With regards to the first part of your review, please find our answers inlined below :

### **- In what programming language is the toolbox implemented? Is there an easy to use interface in a scripting language such as Python?**

The toolbox's Constructor module is implemented in C++ and the parallelism is provided using MPI and MPI/IO standards, to enable the processing of networks' construction on distributed memory parallel computers. On the other hand, based on *igraph's* source code in C, OpenMP directives are used to provide multithreading for the Analy-

C460

sis Engine of Par@Graph so that the analysis of the constructed networks are run on multi-core machines (a multi-core master computing element). An enhanced version of some of the provided algorithms have been implemented in C++11 and its multithreading library.

The choice to use a low-level programming language over Python, particularly in the Constructor module, is basically to facilitate low-level memory manipulation, where it is crucial when processing large correlation matrices. In Python, memory management is ensured internally by the Python memory manager interacting with the memory manager of the operating system. This limits the possibility to optimise memory usage during the construction of large-scale networks.

With regards to an easy-to-use user interface, Par@Graph is equipped with a user configuration file in the format of raw text. Wide range of experimental settings are provided in this file and could be specified to match user particular requirements. Additionally, users are not required to have much experience in C++ since the toolbox functions are provided as a library. Only a simple “main” program (along with the user configuration file) is required to call the provided functions. However, a better approach might be what you mention – providing an interfacing scripting language, e.g., a Python module which certainly would deliver more flexibility to potential users with scripting experience. This will be provided in future versions of the software.

### **- It would be important to know where the software can be obtained for interested researchers and whether it will be available as an open source code.**

The toolbox is not licensed and therefore is not yet available as open source code. Since the development of Par@Graph has been carried out through a collaborative project in which a private company has been a partner, there will be a decision to make shortly about the type of license to provide the software with. For now, however, interested researchers are recommended to contact the authors for the source code, and a remark will be included accordingly in a “Code availability” section in the revised

C461

version of our manuscript.

**- Also, how flexible is the toolbox with regards to including new measures of time series similarity (such as event synchronization [1]) to the Network Constructor as well as additional measures for network quantification (e.g., node-weighted measures for taking into account inhomogeneous grids [2] or motif densities in directed networks [3,4])?**

**- What are the steps required towards including recently proposed methodologies for reconstructing statistically causal climate networks [5,6] in a parallel computing way using the toolbox? The challenge here is that links do not simply arise from a pairwise computation involving two time series, but involve a larger set of time series each because essentially, partial correlations and conditional probabilities are taken into account to exclude the potentially misleading effects of indirect couplings or common drivers.**

The toolbox is equipped with two main functions (Pearson Correlation Coefficient and Mutual Information) to quantify the similarity between pair-wise time series. Additional functions are provided to find maximum correlation value between two time series within a prescribed maximum delay. Other pair-wise measures, including e.g., event synchronisation to quantify the synchronicity between events in time series by counting the events that can be uniquely associated within a prescribed maximum delay are technically possible to include in the toolbox with no much effort since the processing of the whole dataset will follow the same provided computational paradigm. Similarly, functions to weight nodes based on, for example, their densities in climate networks to make the network more representative of the studied domain are certainly possible to include. Of course this will need the inclusion of further algorithms to consider these weights, like for instance, the area-weighted connectivity as in [a].

On the other hand, discovering potential causal links (pathways) in a network is also possible to include in the presented toolbox. As such, although more extensive pro-

C462

cessing will be involved at each ring process (as in the paper) in order to find conditional dependencies between more than two variables, but this is possible since each process is indeed assigned a chunk of time series rather than two. We are actually keen to further discussions on this subject to study the inclusion of additional techniques for the reconstruction of climate networks and analysis, and therefore we welcome any collaborative developments and relevant suggestions.

**- How is the toolbox equipped to deal with coupled climate networks [7] including links between time series in different climatological fields? In this case, it is important to allow for multiple thresholds for network construction, applying different thresholds within each field as well as between fields.**

Although not yet tested with Par@Graph, however, in order to reconstruct a network of networks or such coupled networks, not much of additional work is needed once a subnetwork “flag” is given to each node, then one applies the chosen threshold suitable within each “flag” nodes (time series), and another one when nodes have different “flags” (to define cross-links). Indeed this is one of the motivations behind the development of Par@Graph – not only to reconstruct networks from high-resolution models representing one climatological field, but also to facilitate the reconstruction and analysis of large-scale networks resulted from coupling two (or multiple) single-variable correlation networks. Additional developments will have to be added in order to provide algorithms to calculate cross-network metrics like the cross-degree, cross-clustering coefficient and son on.

**More detailed comments:**

**- p. 322, footnote 3: Please use the following URL for referring to pyunicorn: <http://tocsy.pik-potsdam.de/pyunicorn.php>. The suggested descriptive reference is: Donges, J.F., J. Heitzig, J. Runge, H.C. Schultz, M. Wiedermann, A. Zech, J.H. Feldhoff, A. Rheinwalt, H. Kutza, A. Radebach, N. Marwan and J. Kurths (2013, April).**

C463

Both the recommended URL and reference will be added to the final version of the revised manuscript.

- p. 322, line 24: change to -> "... developed at the Potsdam Institute ...".

Suggested change will be added added.

- p. 323, line 4: in earlier works, also climate networks involving  $\mathcal{O}(10^4)$  have been studied. For example, the frequently analyzed NCEP/NCAR reanalysis surface air temperature data set involves more than 10,000 grid points or nodes.

Suggested change will be considered - and a reference to a climate network of  $\sim 10^4$  nodes will be included.

- Section 2 and later Sections: there are some minor inconsistencies in the notation: a) Please use  $C_{\{ij\}}$  consistently (sometimes  $C_{\{i,j\}}$  is also used). b)  $N(i)$  is introduced on p. 326 to denote the neighborhood of node  $i$ , but later, sometimes  $k(i)$  is used for the same purpose, even though  $k(i)$  also denotes the degree of node  $i$ . c) Use subscript or brackets consistently for denoting the node  $i$  for a local network measure. In the paper, e.g.,  $C_i$  is used for the clustering coefficient, but  $H(i)$  is used for the entropy measure.

Suggested changes will be added to the final revised manuscript.

- p. 326, Equation 5: a factor of 2 is missing in the nominator. This is because in the definition of the clustering coefficient, the number of links  $\rho_i$  between neighbors of  $i$  is divided by the maximum possible number of such links which is  $\binom{k_i}{2} = k_i(k_i-1) / 2$ .

Correct. Mainly for the undirected networks which we reconstruct. Correction will be added accordingly.

- p. 327, line 18: for consistency, use  $BC_i$  instead of  $BC(v)$ .

- p. 329, line 5: change to -> "..., then the pair  $(i,j)$  is copied to a ..."

C464

- p. 329, line 7: explain what  $p_0$  is here (the master compute node).

Suggested changes will be added to the final revised manuscript.

- There appear to be some capitalization issues on the front page p. 319 (e.g., look at the names of the institutions authors are affiliated to).

"scientific software engineers" will be changed to "Scientific Software Engineers".

- p. 334, line 5/6: change to -> "... repeated annual cycle ..."

- p.334, line 24: capitalization: change to -> "Intel".

Suggested changes will be considered in the final revised manuscript.

## References

[a] A. A. Tsonis, K. L. Swanson, and P. J. Roebber. What Do Networks Have to Do with Climate? BAMS, 87(5):585–595, 2006.

Interactive comment on Geosci. Model Dev. Discuss., 8, 319, 2015.