# *AsympPDC Package v. 2b User Guide*

Koichi Sameshima & Luiz A. Baccalá

# Contents

Abstract

This manual describes AsympPDC Package version 2.0b user level functionality aimed at inferring multivariate time series connectivity structure via Partial Directed Coherence (PDC) (and its variants gPDC and iPDC) via frequency domain null hypothesis tests that employ PDC's asymptotic statistics. The package also provides PDC confidence limits when connectivity cannot be rejected. Eleven examples from the literature illustrate its functionalities and capabilities.

# Introduction

## 1.1 Preliminaries

This is a AsympPDC Package version 2.0b release, prepared for accompanying the book "Methods in Brain Connectivity Inference through Multivariate Time Series Analysis", of an implementation of asymptotic statistics for partial directed coherence (PDC) that allows frequency domain Granger causality estimation via multivariate time-series vector autoregressive modeling by testing for the null hypothesis of significant PDC via its computed asymptotic statistical properties [Takahashi et al., 2007, 2010, de Brito et al., 2010]. Present release only MATLAB code is provided. A future release including a Python implementation is planned.

This documentation describes the basic package organization and aims at allowing data analysts to quickly access the present connectivity inference methods for application on their own data sets. Didactic examples are included in the package to illustrate its use and functionality together with the nature of possible results.

This user manual is organized as follows: after the present brief introduction, Section 1.2 discusses installation and legal aspects. This is followed by a description of the package's organization (Section 2.1), its main routines and folder structure. Note that Section 2.1 presents a brief list of the illustrations provided in the package. Section 2.1 describes a template file whose comments point to details of how to call each function while Section 3.1 portrays the results of running one of the provided examples. Finally Section 3.2 lists known issues.

To provide expedited access to the present computational methods, this release does not provide a user friendly interface environment of the type often provided by real world professional data analysis packages. This is made up by providing the full integrated code for PDC estimation and null hypothesis testing. In doing so, our goal is to provide tools to explain/clarify the key concepts, PDC's advantages and limitations and (2) to give data analysts the chance to experiment the methods on their own data in a way that they can incorporate the code in their own analysis procedures.

## 1.2 Installation and Legal Aspects

### Installation and Requirements

The compressed package distribution file is provided in the book accompanying CD-ROM, asymppdc.zip, or can be downloaded from PDC website.[1] After uncompressing it, copy/transfer the folder structure to your working folder. Then set MATLAB path access to that folder and subfolders using File Set Path... command from MATLAB Desktop window. The **asymppdc** main folder contains three subfolders, **routines**, **examples** and **supporting**, and a pair of files, a

[1] http://www.lcs.poli.usp.br/~baccala/pdc/

comprehensive template m-file for PDC connectivity analysis, **pdc_analysis_template.m** (see Appendix A), and a **readme.txt** file.

The AsympPDC package has been tested under Windows, Mac OS X and Linux-Ubuntu platforms running MATLAB version 7.1 and higher. The package uses some routines from Signal Processing, Statistics, and Control System toolboxes.

## License and Distribution

This release of the AsympPDC package is a MATLAB version of PDC connectivity tools for time series analysis and its content (not including **supporting folder**'s codes) is released as open source code under the GNU general public license version 3.

MATLAB© is a product from Mathworks Inc™.

## Disclaimer

The software package is supplied "as is" and all use is at your own risk. The authors disclaim all warranties of any kind, either express or implied, as to the software package, including, but not limited to, implied warranties of fitness for a particular purpose, merchantability or non-infringement of proprietary rights. Neither this agreement nor any documentation furnished under it is intended to express or imply any warranty that the operation of the software package will be error-free.

Under no circumstances shall the authors of the software be liable to any user for direct, indirect, incidental, consequential, special, or exemplary damages, arising from the software, or user's use or misuse of the software. Such limitation of liability shall apply whether the damages arise from the use or misuse of the data provided or errors of the softwares.

# Implementation

## 2.1 General Organization

A short description of folder (module) content follows.

### Routines Module

There are four main routine types: (1) the MVAR module itself that estimates the VAR model via four algorithm options and the choice of different model order selection criteria; (2) a time domain Granger causality test (GCT) implementation that includes an instantaneous Granger causality test (IGCT); (3) the PDC and asymptotic statistics calculation routine (**asymp_pdc.m**) implementing all three PDC formulations and finally (4) a basic plotting routine, **xplot**, that provides graphic representation of the asymp_pdc routine results.

These modules and auxiliary m-files are briefly described:

**Module and routine descriptions**

- **A_to_f.m**
  Computes $\mathbf{A}(f)$ matrix in the frequency domain.

- **arfitcaps.m**
  If available, the **ARfit package** can be used as an alternative for the natively implemented VAR estimation algorithms. The ARfit package was developed by Tapio Schneider and Arnold Neumaier. Please visit Tapio Schneider's site. For further information about ARfit see [Scheneider and Neumaier, 2001].

  **arfitcaps.m** is capsule routine for calling the arfit.m, which is part of the "ARfit: Multivariate Autoregressive Model Fitting" package. If you would like use ARfit algorithm for VAR model estimation, you can get it at http://www.clidyn.ethz.ch/arfit/index.htmll
  Please also read the allied license term before using it.

- **asymp_pdc.m**
  Computes the PDC connectivity measure and its asymptotic statistics taking as input arguments the time series.

- **cmlsm.m**
  VAR least squares estimator.

- **coh_alg.m**
  Calculates the cross coherence functions from spectral density matrix.

- **gct_alg.m**
  Performs the Granger causality test, including instantaneous causality.

- **getCij.m**
  Extracts the (i,j) index variable structure from the c structure that results from asymp_pdc and employs the following syntax

  ```
  Cij(f) = getCij(c,i,j,nFreq)
  ```

  c is a structured variable that stands for either c.pdc, c.th, c.ic1, c.ic2, c.SS, c.coh, or p.pdc_th, as returned by **asymp_pdc.m** and **pdc_alg.m**.

- **mcarns.m**
  Calculates the coefficients of vector auto-regressive matrix using the Nuttall-Strand algorithm (a generalization of single channel harmonic method).

- **mcarvm.m**
  Calculates the coefficients of vector auto-regressive matrix using the Vieira-Morf algorithm, a generalization of single-channel geometric method.

- **mvar.m**
  Estimates the VAR matrix based on algorithm choice and model order selection criteria

- **mvarresidue.m**
  Residues test for whiteness.

- **pdc_alg.m**
  Computes partial directed coherence measure from the time series given by "options". If you want just to calculate PDC measure, and the asymptotic statistic this is probably the most useful routine

- **ss_alg.m**
  Calculates the spectral density function (SS) given the auto-regressive matrix, $A$, and covariance residue.

- **standardize.m**
  Data transformation by standardization of time series imposing zero mean and unit standard-deviation.

- **xplot.m**
  Connectivity plot in matrix layout with power spectra or PDC along the main diagonal.

- **xplot_title.m**
  This is an auxiliary routine that can be used with **xplot.m** to put a text title above the matrix layout plot.

## Examples Module

The folder **examples** contains 11 examples borrowed from the literature, four auxiliary m-files used by the examples (**example_analysis_parameters.m**, **example_pre_processing.m**, **example_mvar_estimation.m**, and **example_pdc_analysis.m**), and a batch file that runs all the examples in batch-mode (**run_all_examples.m**) for operating system compatibility testing purposes.

Those curious about the performance and characteristics of PDC, gPDC and iPDC, play along with the examples in the folder, read the comments and try alternatives.

The examples presented in this package have been borrowed from Andrews and Herzberg [1985]), Sunspot-Melanoma 1936-1972 series, and from the literature, Baccalá and Sameshima

[2001a], Baccalá and Sameshima [2001b], Schelter et al. [2005], Schelter et al. [2006], Schelter et al. [2009], Guo et al. [2008], Gourévitch et al. [2006], and an extended variant of Winterhalder et al. [2005].

All examples can be run in batch to verify if all routines and necessary toolboxes are available and MATLAB path has been set by running the example batch command

>>run_all_examples

## Supporting Module

Mostly comprised of user-contributed MATLAB code:

- **boundedline.m**
- **shadedErrorBar.m**
- **shadedplot.m**
- **subplot2.m**
  One can control the subplots spacing by editing the **kspacescale** parameter in **subplot2.m** routine. Try varying it in [0 1] range.
- **suplabel.m**
- **suptitle.m**
- **tilefig.m** and **tilefigs.m**
  Figure tiling routines for organizing the visualization of several open figures. See for instance **andrews_herzberg.m**, the Sunspot-Melanoma example.

## PDC Analysis Getting Started Template File

The **pdc_analysis_template.m** file is meant to be self-explanatory. Please read and play with it. The file contains a description of how to interconnect the various routines to make the analysis of one's data. The **pdc_analysis_template.m** file is listed in Appendix 5 for convenience.

# Examples

## 3.1 Examples from literature

### Sunspot-Melanoma 1936-1972 Series: andrews_herzberg.m

This is an interesting very short data set, 37 data points in all (see Figures 3.1 and 3.2), which can be used to investigate the interdependence between the cycles of solar activity given by the annual sunspot number and the epidemiological record of the number annual melanoma cases in the state of Connecticut from 1936 to 1972. The data are provided in the **sunmeladat.m** file contained in the **extras** directory.

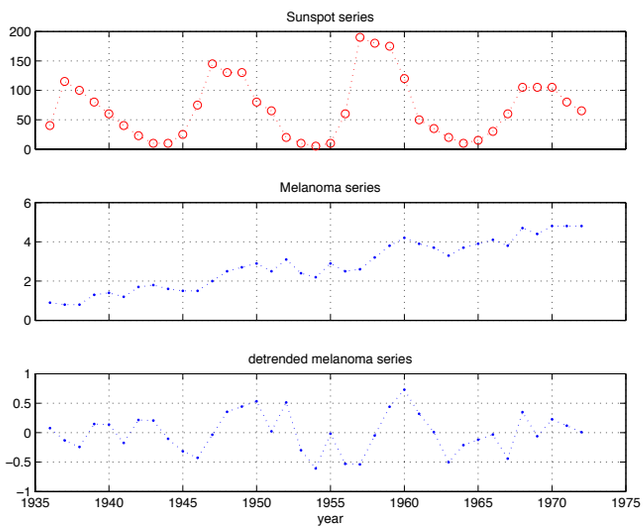


Figure 3.1: Detrended Sunspot-Melanoma 1936-1972 Series .

Original PDC estimation of detrended but not standardized times series yields the result depicted in Figure 3.3.

Using generalized PDC for $\alpha = 1\%$ leads to the results in Figure 3.4 where the apparent contradictory results from Figure 3.3 become resolved.

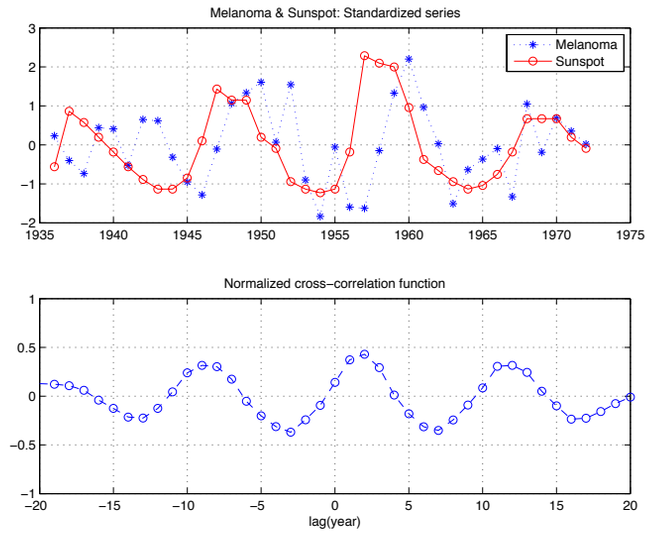


Figure 3.2: Standardized Sunspot-Melanoma series plotting and corresponding normalized cross-correlation function. Note that the peak of cross-correlation of 2 years lag led by the sun's activity.

**Checking asympPDC package log output on Command Window in more detail:**

```
==========================================================================
     Andrews and Herzberg's Sunspot and Melanoma 1936-1972 Data
                Sunspot --> Melanoma or other way?
==========================================================================

Setting up default analysis parameters.
Running simple data pre-processing routines.
Time series were detrended.
Running MVAR estimation and GCT analysis routines.
maxOrder limited to 30
IP=1  vaic=418.156614
IP=2  vaic=415.350982
IP=3  vaic=409.454496
IP=4  vaic=411.220477

Number of channels = 2 with 37 data points; MAR model order = 3.
--------------------------------------------------------------------------
                  MVAR Residues test for whiteness
==========================================================================
Good MAR model fitting! Residues white noise hypothesis NOT rejected.
Pass =
    0.0250
st =
   85.8928
--------------------------------------------------------------------------
                          GRANGER CAUSALITY TEST
==========================================================================
Connectivity matrix:
Tr_gct =
    -1     0
     1    -1
Granger causality test p-values:
```

```
pValue_gct =
   -1.0000    0.0796
    0.0000   -1.0000
------------------------------------------------------------------------
                    INSTANTANEOUS GRANGER CAUSALITY TEST
========================================================================
Instantaneous connectivity matrix:
Tr_igct =
    -1     0
     0    -1
Instantaneous Granger causality test p-values:
pValue_igct =
   -1.0000    0.0762
    0.0762   -1.0000
>>>> Instantaneous Granger causality NOT detected.
(...)
%=======================================================================
            End of Sunspot-Melanoma series analysis example.
%=======================================================================
```

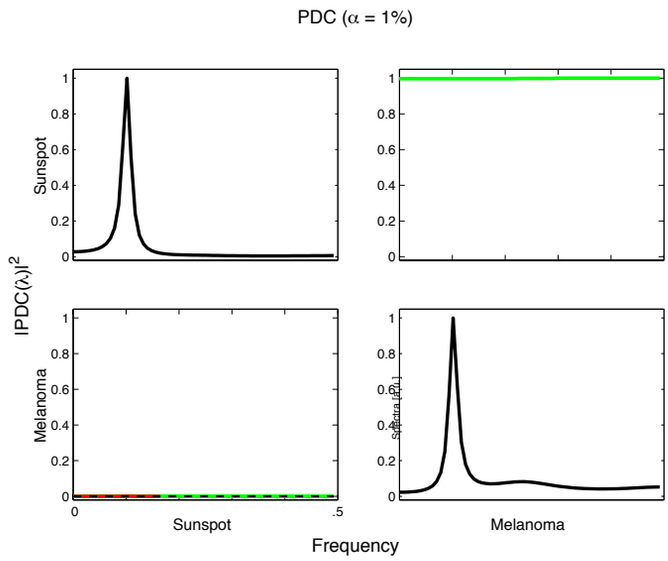


Figure 3.3: Matrix layout plot of squared original PDC ([Baccalá and Sameshima, 2001a]) calculated from detrended but nonstandardized sunspot and melanoma series. The direction of influence is from column to row variables. Observe that in this case the magnitude of PDC from $Melanoma \rightarrow Sunspot$ is high, close to 1, yet as the green line indicates, PDC is not significant. In the reverse direction, i. e. $Sunspot \rightarrow Melanoma$ even though PDC's magnitude seems almost zero, one can see a red line at the lower frequencies indicating significant connectivity in this direction. The gray lines depict coherence function, a symmetric measure.

## baccala2001a_ex3.m

Example 3 of Baccalá and Sameshima [2001a] (pag.468), a five-dimensional VAR[3] with feedback between $x_4$ and $x_5$. Note that squared original PDC is plotted on the main diagonal instead of usual power spectra (see Figure 3.5).

## baccala2001a_ex4.m

Example 4 of Baccalá and Sameshima [2001a], a five-dimensional VAR[2] also with feedback between $x_4$ and $x_5$. The squared original PDC is also plotted along the main diagonal (Figure 3.6).

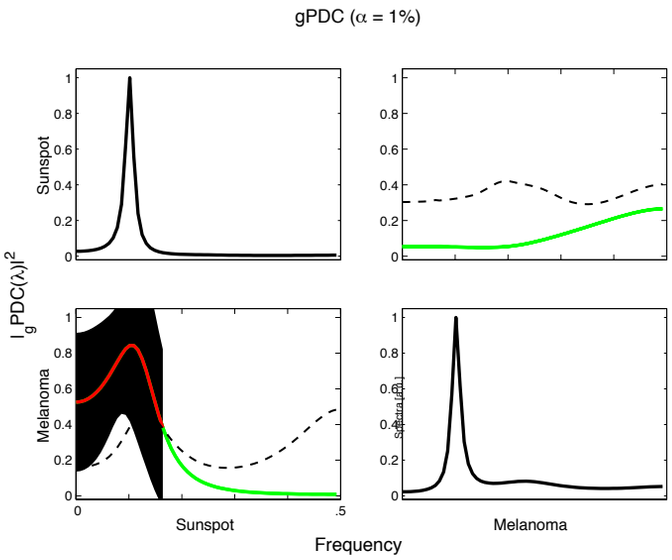


Figure 3.4: Matrix layout plot of squared generalized PDC ([Baccalá and Sameshima, 2007]) estimates of the same sunspot and melanoma series of Figure 3.3. Now one observes that $|gPDC(f)|^2 < 0.3$ for the Melanoma to Sunspot activity, but more importantly it is not statistically significant. In the reverse direction, $Sunspot \rightarrow Melanoma$ now reveals a clear picture of Sun's activity onto the number of melanoma cases, with a peak at the lower frequency range, which correponds to approximately 11-year cycle indicating significant (red line) connectivity in this direction. In the gPDC plots, the black dashed-line indicates the 1% significance level, and the pair of blue dashed-lines the 99% confidence interval of the significant range of gPDC.

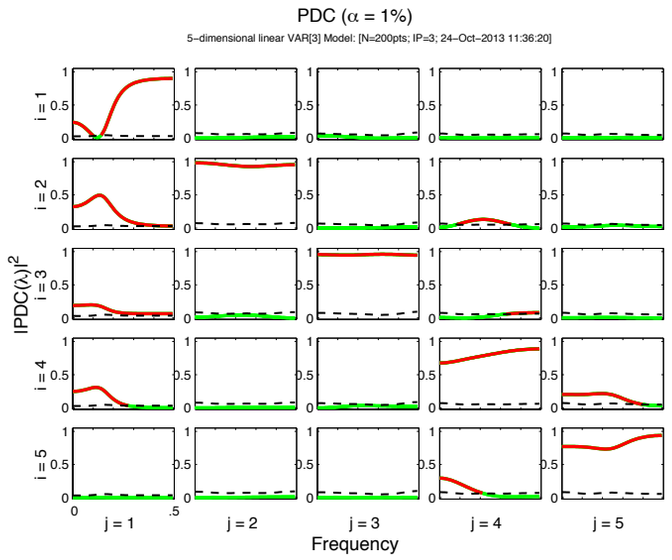


Figure 3.5: Squared PDC estimates of **baccala2001a_ex3.m** simulation.

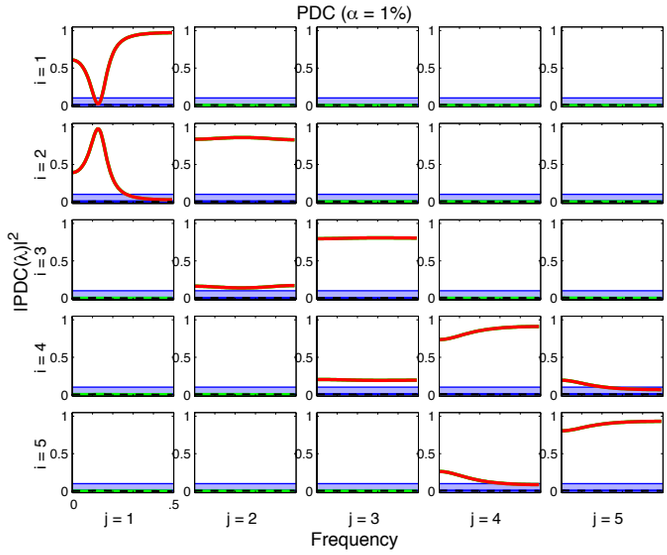


Figure 3.6: Squared PDC estimates of **baccala2001a_ex4.m** simulation.

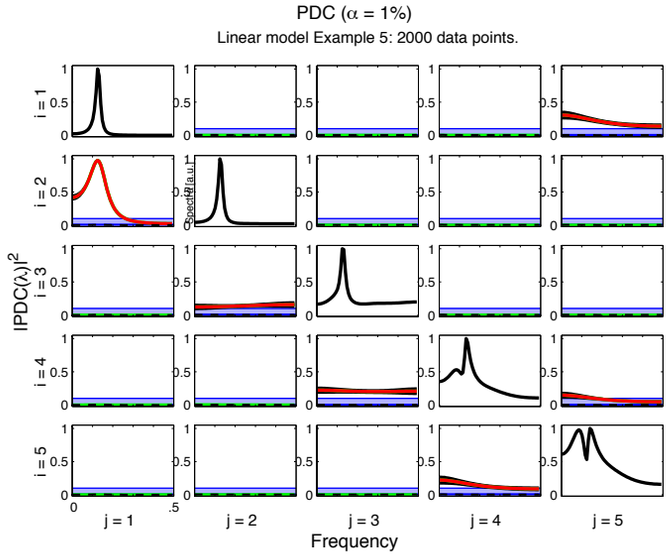


Figure 3.7: Squared PDC estimates of **baccala2001a_ex5.m** simulation.

## baccala2001a_ex5.m

Example 5 of Baccalá and Sameshima [2001a] is a slight variation of previous example, now with loop close from $x_5$ and $x_1$. A pair of figures are generated, first, Figure 3.7, with standard y-axis scale [0 1] and second, Figure 3.8, using colored rescaling plots as explained in **pdc_analysis_template.m**.

The range of PDC and coherence is from [01], but maximum peak amplitude may be small ($< .1$), or even smaller, $< .01$, so that in these cases it might be interesting to plot PDC at finer y-axis scale. If flgColor not equal zero is chosen, the off-diagonal plottings use white-background for full-scale [0 1] y-axis, while light-blue-background fills [0 .1] scale, or more correctly [.01 .1] and light-purple-background shows very fine detail of small, [0 .01] range y-axis,usually not significant PDCs. Edit example m-file and try flgColor = 0 or 1, or both [0 1]. The y-axis scale is controled by flgScale parameters. For detailed options, issue help xplot command.

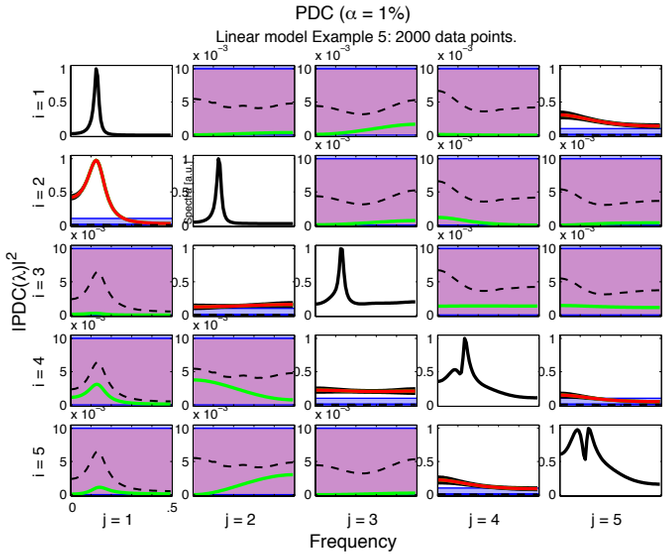


Figure 3.8: Rescaled plotting of Figure 3.7.

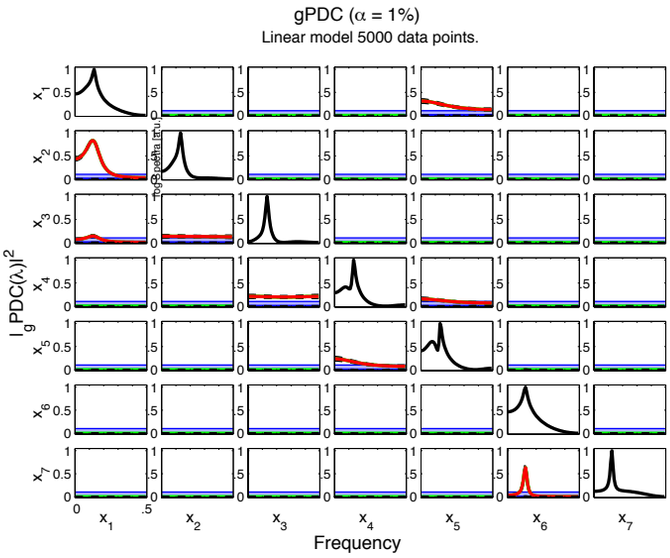


Figure 3.9: Outcome of a **baccala2001b.m** simulation.

## baccala2001b.m

A 7-dimensional VAR[2] model with loop and feedback from Baccalá and Sameshima [2001b], $|gPDC|^2$ calculation. Compare the results of Granger Causality Test listed on MATLAB Command Window with gPDC matrix layout plotted results. As this is an analysis of linear model with quite large amount of data, 5000 data points, the outcomes from GCT (in time domain) and gPDC (in frequency domain) should agree fairly well. Note that the confidence intervals of significant $|gPDC|^2$ are very tight. Observe that the processes $x_6$ and $x_7$ are isolated from other variables, however if the simulation is repeated many times you might see some false-positive connectivity occurrences (with approximately at the rate of 1%, which is the significance level chosen for this example).

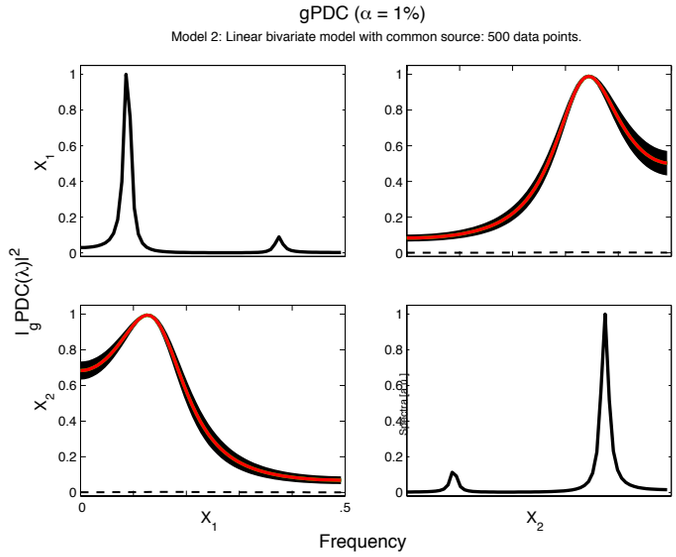


Figure 3.10: Simulation output of **gourevitch2006_model2.m**

## gourevitch2006_model2.m

This example is from Gourévitch et al. [2006], a linear bivariate model with bidirectional influence and common source. Each process has peak of power spectrum at different frequencies. Due to common source influence, significant instantaneous Granger causality is detected between them. Check the Command Window log output.

## guo2008_linear.m

This is an example from Guo et al. [2008], a five-dimensional VAR[3] process modified from a version presented in Baccalá and Sameshima [2001a], in which was added a large common exogenous inputs to all variables. The exogenous inputs introduce large common variance that overpower the magnitude of directed interactions. As you may notice from Command Window output, there are significant instantaneous Granger causality between all pair of variables. Due to the large common exogenous white noise often one may see false-positive and false-negative connectivity in some simulations, which will also be dependent on your choice of significance level.

One immediately sees that the PDC, Figure 3.11, or gPDC, Figure 3.12, do not resemble PDC plot in Fig.1 of Guo et al. [2008]. Our best guess is that Guo and colleagues used an incorrect PDC estimator. Actually you may also notice that our PDC and gPDC estimates are very similar to PGC shown by Guo et al. [2008].

In all three PDC formulations, the significant PDC frequency range is similar. It is important fo stress that iPDC gives a measure of size effect, which is very small in this examples due to the large common exogenous inputs. Even so iPDC is significant in the same frequency range as PDC and gPDC. See Figure 3.13. The same figure can be seen rescaled in Figure 3.14.

So we must conclude that the statement by Guo and collaborators that PDC can not uncover the connectivity pattern in large common noise does hold.

## schelter2005.m

Five-dimensional VAR[4]-process Eq. (5) example from Schelter et al. [2005].

The $x_4$ power spectrum and $|g\pi_{2,4}|^2$ of our simulation, Figure 3.15, differ significantly from Schelter et al. [2005] 's results. Our guess is that Schelter et al. may have used slightly different parameters from what they stated in Eq. 5 in Schelter et al. [2005].

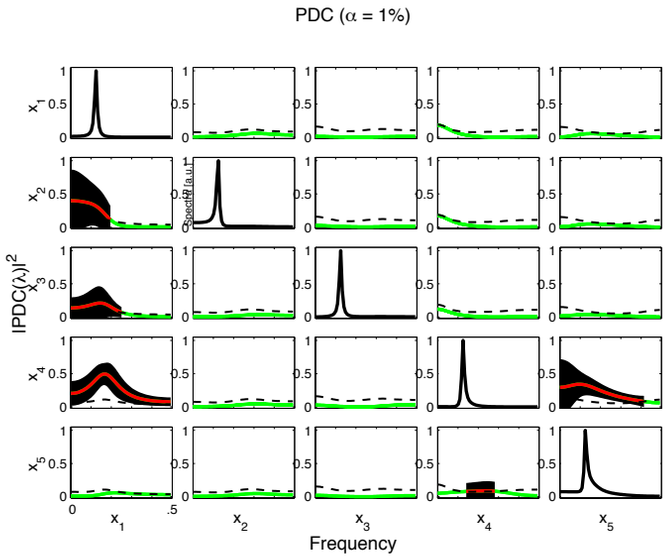


Figure 3.11: Simulation $|PDC|^2$ output of **guo2008_linear.m**

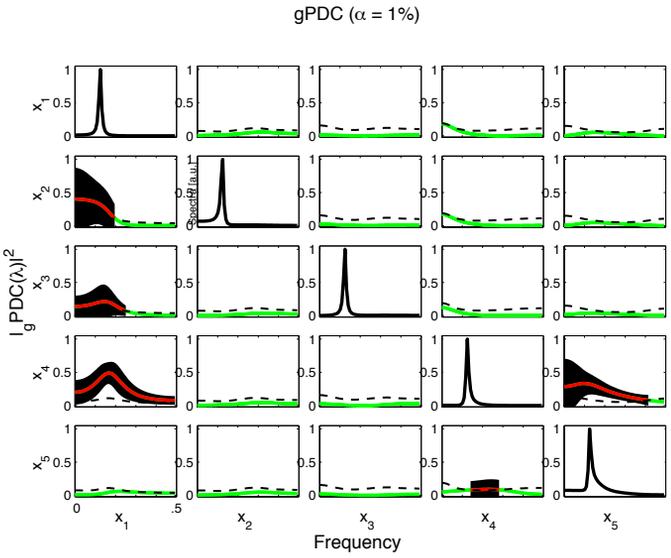


Figure 3.12: $|gPDC|^2$ estimates, without confidence interval, using the same data set of Fig. 3.11 given by **guo2008_linear.m**. Note that for this example PDC and gPDC estimates are very similar.

Note that, for linear model with balanced innovation, the maximum of gPDC estimates is roughly proportional to the autoregressive model coefficients. As we are plotting squared gPDC, the amplitude is proportional to the square of coefficient.

## schelter2006.m

A 5-dimensional VAR[4] model example presented by Schelter et al. [2006]. Comparing this plot with Fig.3 in Schelter et al. [2006], one sees that their is PDC amplitude plots, while this example shows squared PDC.

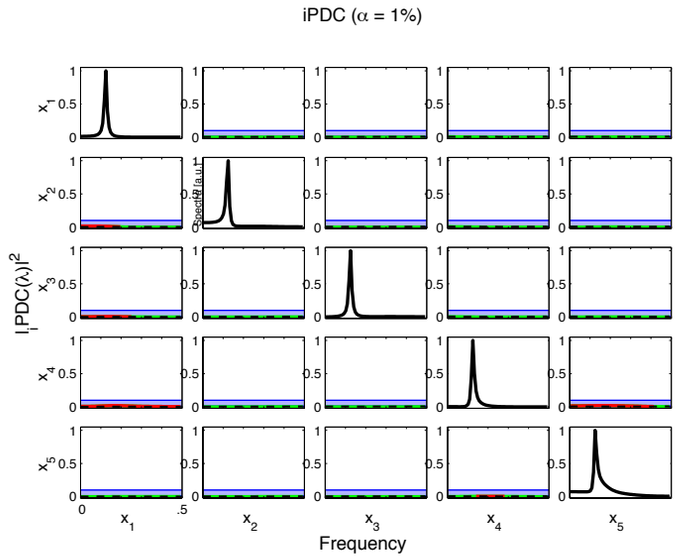


Figure 3.13: $|iPDC|^2$ estimates using the same data set of Fig. 3.11, which depicts very small amplitude influences and reflects actual size effect of interactions in the frequency domain.

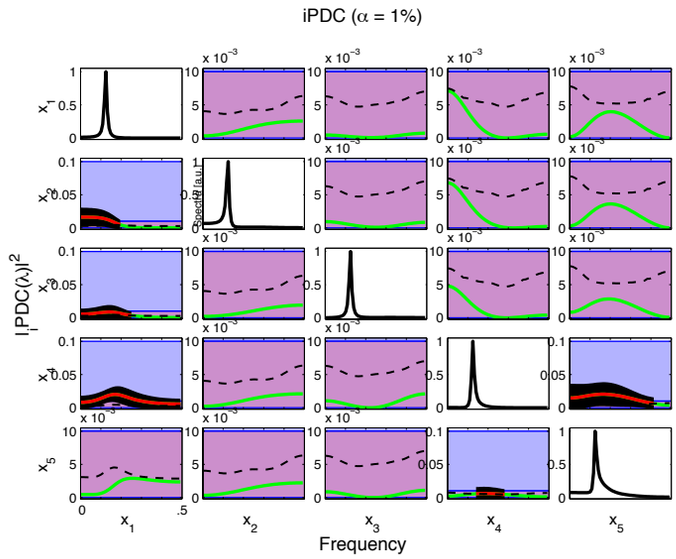


Figure 3.14: Rescaled plots of Figure 3.13.

## schelter2009_vap1.m

A 5-dimensional VAR[3] process example presented by Schelter et al. [2009]. The example shows squared gPDC plots, without power spectra on main diagonal. Observe in Figure that there are two false-positive connections detection from $x_1 \rightarrow x_5$ and $x_3 \rightarrow x_5$ with very small amplitude. Using color-rescaling, Figure 3.17, one sees that the peaks of false-positive squared gPDC are smaller than 0.01.

## winterhalder2005_variant.m

Expanded version of three random independent processes presented by Winterhalder et al. [2005], here with seven variables using unbalanced innovation noises with following values: $\sigma_1 = 500; \sigma_2 = 1; \sigma_3 = 500; \sigma_4 = 1; \sigma_5 = 1; \sigma_6 = 1; \sigma_7 = 1$. Figure 3.19 shows PDC estimates, with some non-significant large values. The squared gPDC outcome is shown in Figure 3.20.

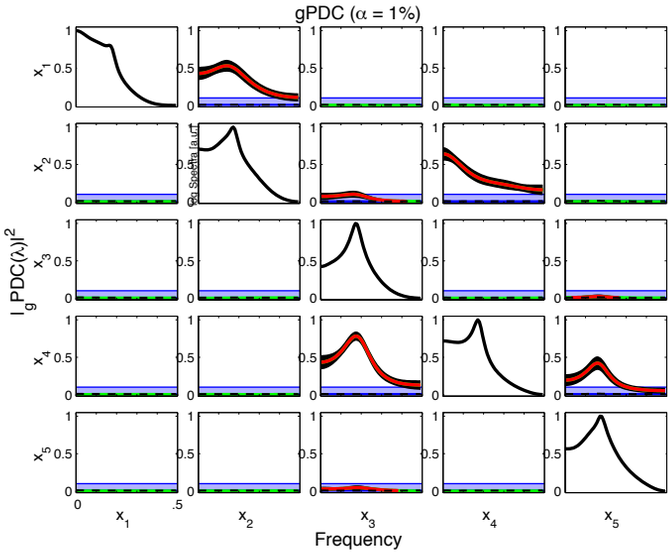


Figure 3.15: $|PDC|^2$ estimate of a schelter2005.m simulation. Note log-scale power spectra plots on main diagonal.

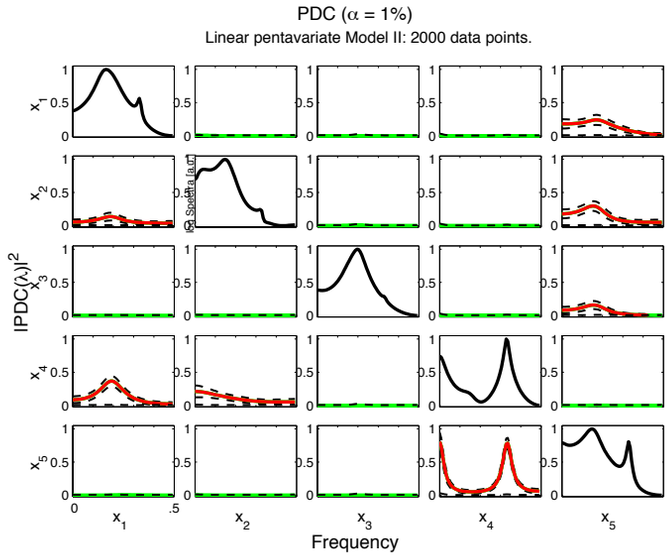


Figure 3.16: $|PDC|^2$ plot of schelter2006.m simulation.

## 3.2 Known issues

1. The x-axis scaling plot label does not work for any other than normalized unit frequency, i.e. $fs = 1$.

2. The **asymp_pdc**.m routine as provided is not optimized. Setting $alpha = 0$, provides PDC without its asymptotic statistics, and is thus much faster.

3. A known issue among MATLAB subplot users is that any figure reformatting with subplot requires replotting of everyone of its components which accounts for the slow speed of the xplot routine.

Please help us by referencing your use of this package and by reporting any bugs you find. You may do so by email to ksameshi@usp.br or baccala@lcs.poli.usp.br.

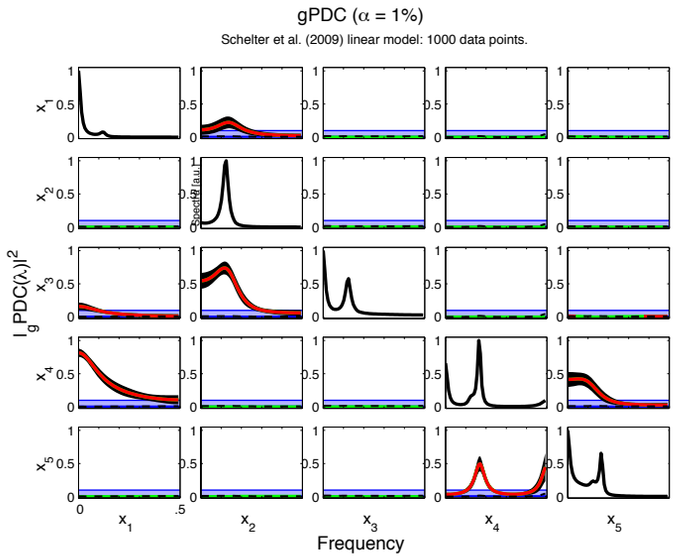


Figure 3.17: $|gPDC|^2$ plot of schelter2009_vap1_gPDC.m simulation with 99% confidence interval plots. Ocasionally one might see some false-positive connectivities inference in eventual simulations, which could be seen in more detail using a rescaled figure, such as Figure 3.18 (not observed in this simulation).

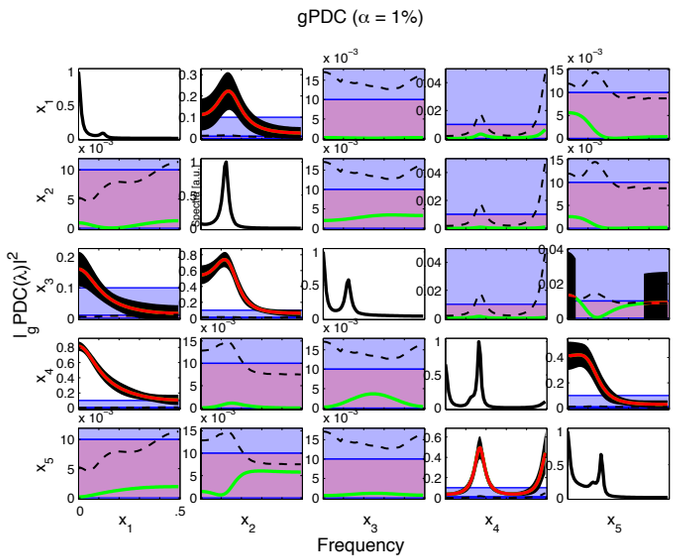


Figure 3.18: Rescaled $|gPDC|^2$ plot of Figure 3.17.

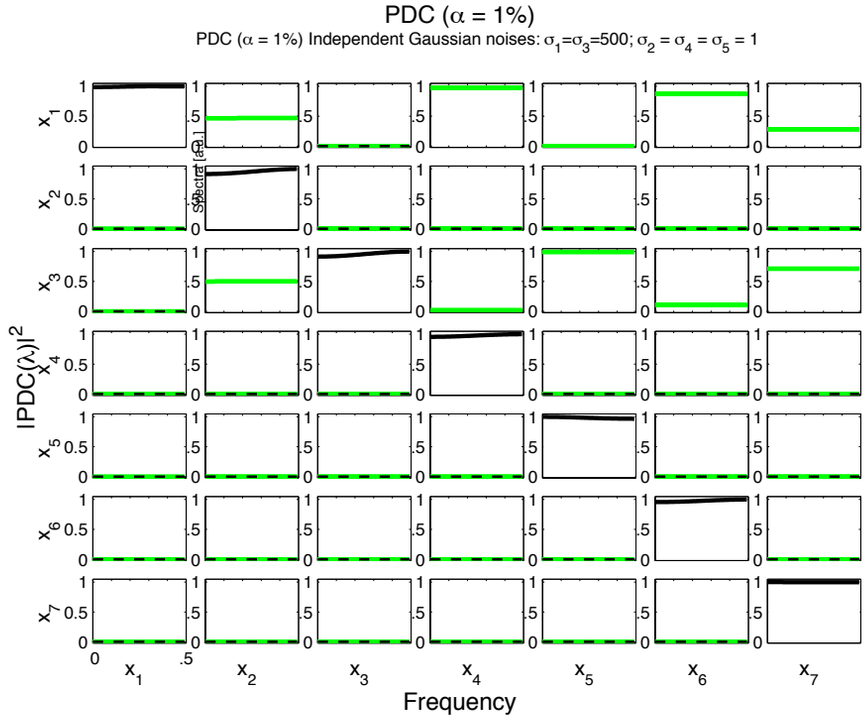


Figure 3.19: Squared PDC estimates for seven independent random processes with unbalanced innovation noises. One immediately sees that PDC estimates from processes with small innovation noises toward large one may show very high but nonsignificant values. The normalization introduced with gPDC definition resolves this issue (See Figure 3.20.)

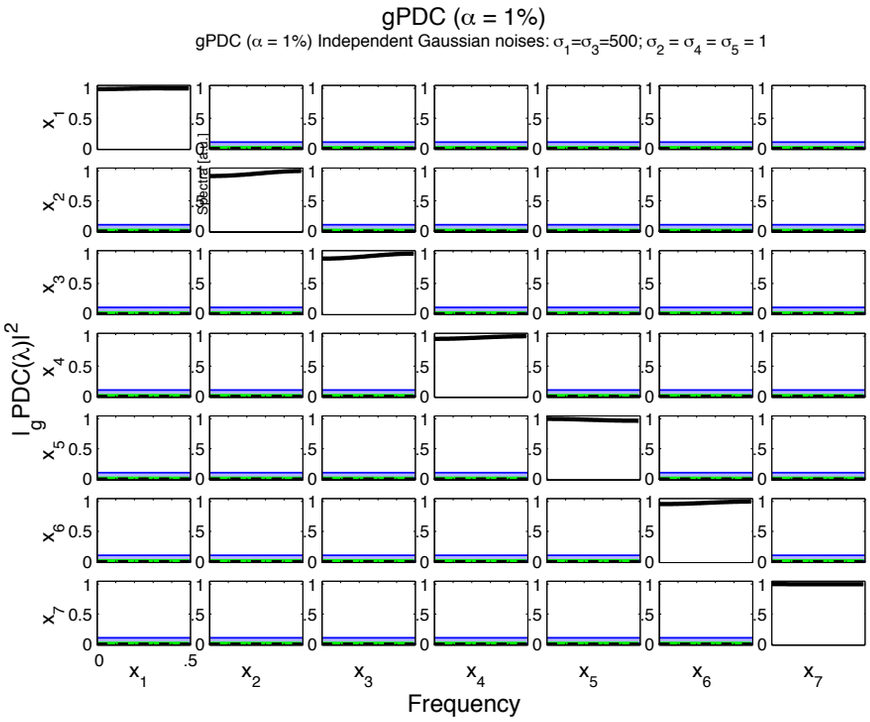


Figure 3.20: Squared gPDC estimates for seven independent random processes with unbalanced innovation noises.

# Acknowledgements

## Acknowledgements:

K.S. and L.A.B. gratefully acknowledge support from the Fundação de Apoio à Pesquisa do Estado de São Paulo (FAPESP - São Paulo Research Foundation) Grant 2005/56464-9 (CInAPCe Program). Daniel Y. Takahashi received fellowships from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and FAPESP Grant 2008/08171-0 during the AsympPDC package development. L.A.B. was also supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) Grant 304404/2009-8. K.S. was supported by Fundação Faculdade de Medicina and CNPq Grant 309381/2012-6. Carlos Stein Naves de Brito was supported by CAPES fellowship. K.S. is grateful to the warmth and unconditional support from the Department of Radiology and Oncology, Faculdade de Medicina - Universidade de São Paulo.

# Bibliography


D. F. Andrews and A. M. Herzberg. *Data: A Collection of Problems from Many Fields for the Student and Research Worker*. Springer, New York, 1985.

L. A. Baccalá and K. Sameshima. Partial directed coherence: a new concept in neural structure determination. *Biological Cybernetics*, 84:463–74, 2001a.

L. A. Baccalá and K. Sameshima. Overcoming the limitations of correlation analysis for many simultaneously processed neural structures. *Progress in Brain Research*, 130:33–47, 2001b.

L. A. Baccalá and K. Sameshima. Generalized partial directed coherence. *Proceeding of the 15th International Conference on Digital Signal Processing (DSP 2007)*, pages 163–166, 2007.

C. S. N. de Brito, L. A. Baccalá, D. Y. Takahashi, and K. Sameshima. Asymptotic behavior of generalized partial directed coherence. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 1718–1721, 2010.

B. Gourévitch, R. L. Bouquin-Jeannès, and G. Faucon. Linear and nonlinear causality between signals: methods, examples and neurophysiological applications. *Biological Cybernetics*, 95: 349–369, 2006.

S. Guo, J. Wu, M. Ding, and J. Feng. Uncovering interactions in the frequency domain. *PLoS Computational Biology*, 4:1–10, 2008.

B. Schelter, M. Winterhalder, M. Eichler, M. Peifer, B. Hellwig, B. Guschlbauer, C. H. Lücking, R. Dahlhaus, and J. Timmer. Testing for directed influences among neural signals using partial directed coherence. *Journal of Neuroscience Methods*, 152:210–219, 2005.

B. Schelter, M. Winterhalder, B. Hellwig, B. Guschlbauer, C. H. Lücking, and J. Timmer. Direct or indirect? Graphical models for neural oscillators. *Journal of Physiology, Paris*, 99:37–46, 2006.

B. Schelter, J. Timmer, and M. Eichler. Assessing the strength of directed influences among neural signals using renormalized partial directed coherence. *Journal of Neuroscience Methods*, 179:121–130, 2009.

T. Scheneider and A. Neumaier. Algorithm 808: ARfit - A Matlab package for the estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Transactions on Mathematical Software*, 27:58–65, 2001.

D. Y. Takahashi, L. A. Baccalá, and K. Sameshima. Connectivity inference between neural structures via partial directed coherence. *Journal of Applied Statistics*, 34:1259–1273, 2007.

D. Y. Takahashi, L. A. Baccalá, and K. Sameshima. Information theoretic interpretation of frequency domain connectivity measures. *Biological Cybernetics*, 103:463–469, 2010.

M. Winterhalder, B. Schelter, W. Hesse, K. Schwab, L. Leistritz, D. Klan, R. Bauer, J. Timmer, and H. Witte. Comparison of linear signal processing techniques to infer directed interactions in multivariate neural systems. *Signal Processing*, 85:2137–2160, 2005.

# Appendix

## PDC_analysis_template.m listing

```
%PDC analysis getting started template file
%
% Edit this file to analyze your data. You might want to choose analysis
% parameters followed by comment containing "<***>". Check bellow.
%
% Some important input and output parameters and variables:
% input:
%        u     - data in columns
%        fs    - Sampling frequency
%        maxIP - externally defined maximum IP
%        alg   - for algorithm (1: Nutall-Strand),(2: mlsm) ,
%                              (3: Vieira Morf),  (4: ARfit)
%        criterion - for AR order selection =>
%                                    1: AIC; 2: Hanna-Quinn; 3: Schwartz;
%                                    4: FPE, 5: fixed order in MaxIP
%        alpha  -  PDC test significance level
%
% output:
%         c.pdc - original/generalized/informational PDC
%         c.th -  threshold level by Patnaik approximation
%         c.pdc_th - above threshold pdc values otherwise equal NaN
%         c.ic1,c.ic2 - superior and inferior confidence interval
%         c.p - VAR model order
%         c.SS - Power spectrum
%         c.coh - coherece function


%===========================#
% Times series for analysis /
%===========================#
% u     - data in columns.
%         The variable u must contain the time series
%         If flgExample=1 the template file will analyze a
%         5 variables Gaussian independent noises.

format compact
clear all; clc

```

```
% Choose Example 1 == Five independent random variables model
%                 2 == Sunspot-melanoma time series
%                 3 == Baccala & Sameshima (2001) 5 variables linear model
%                 4 == Takahashi(2009) Thesis' example model
flgExample=2;

disp('================================================================================');
disp('============= PDC analysis getting started template ==================')
disp('================================================================================');

switch flgExample
   case 1,
      u=randn(2000,5);    %<***> Example (1)
      disp('            Random Independent Process with 5 variables')
      disp('================================================================================');

   case 2
      u=sunmeladat([4 3]); %<***> Example (2)
      disp('     Andrews and Herzberg''s Sunspot and Melanoma Example');
      disp('                  Sunspot --> Melanoma or other way?');
      disp('================================================================================');
   case 3
      u=baccala2001a_ex5data(200);
   case 4,
      u=takahashi_thesis_dat(200);
   otherwise
      error('Wrong example selection.')
end;

fs = 1; %<***> Sampling frequency

[nSegLength,nChannels]=size(u);
if nSegLength < nChannels, error('The data might be transposed.'); end;

%============================#
%  Channel identification   /
%============================#

switch flgExample
  case 1,
    chLabels  = {'x_1';'x_2';'x_3';'x_4';'x_5'}; %<***> Example (1)
    strTitle2 = 'Five independent Gaussian noises '; %Title info
  case 2
    chLabels  = {'Sunspot';'Melanoma'}; %<***> Example (2)
    strTitle2 = 'Sunspot & Melanoma 1936-1972 ';
  case 3,
    chLabels  = [];                     %<***> Example (3)
    strTitle2 = 'Five variables Baccal?+Sameshima(2001) examples ';
  case 4,
    chLabels  = {'X'; 'Y'; 'Z'};        % Takahashi thesis example
    strTitle2 = 'Takahashi 2008 (Thesis) example';
end;

flgLabels = ¬isempty(chLabels);
```

```
93  if flgLabels,
94    if nChannels ≠ max(size(chLabels))
95      error('Numbers of labels and channels do not match.')
96    end;
97  end;
98
99  %============================#
100 %       Action flags         /
101 %============================#
102 flgDetrend = 1;    %<***> Usually it's recommended to detrend the time series.
103
104 flgStandardize = 1; %<***> For PDCn estimation standardization has no effect.
105
106 if flgStandardize,
107   disp('Be aware that the data standardization does not affect the generalized')
108   disp('   PDC estimates nor its statistics results, so that data standardization')
109   disp('   is not necessary.')
110 end;
111
112 %============================#
113 %    Analysis parameters     /
114 %============================#
115 nFreqs = 128; %<***> number of points on frequency scale;
116               %      use either 64 or 128.
117
118 metric = 'euc';
119 %        metric   'euc'  - Euclidean     -> original PDC;
120 %                 'diag' - diagonal      -> gPDC;
121 %                 'info' - informational -> iPDC;
122
123 maxIP = 30; % maxIP - externally defined maximum IP %<***>
124
125 %============================#
126 %    MAR algorithm           /
127 %============================#
128 % Choose one of algorithm for MAR estimation
129 % alg   - for algorithm  (1: Nutall-Strand),(2: mlsm) ,
130 %                        (3: Vieira Morf),  (4: QR artfit)
131 alg = 1; %<***> Nuttall-Strand (alg=1) algorithm, it seems to be a good
132        %      and robust method.
133
134 %============================#
135 %MAR order selection criteria/
136 %============================#
137 % criterion - for AR order choice
138 %  1: AIC; 2: Hanna-Quinn; 3: Schwartz;
139 %  4: FPE, 5: fixed order in MaxIP
140 criterion = 1; %<***> AIC, Akaike information criterion (Our preferred one)
141
142 %=================
143 alpha = 0.01;         %<***> Significance level for PDC null hypothesis
144                       % testing, it is usually 1% or 5%
145                       %
146                       % IMPORTANT: if alpha == 0, no asymptotic statistics
```

```
                    % calculation is performed and ASYMP_PDC (see bellow)
                    % will only returns PDC. This option is interesting
                    % if you want faster PDC calculation.
                    %
gct_signif = alpha;  % Granger causality test significance. Choose other
                    % value if you have good reason for using different
                    % one from PDC statistical testing.
igct_signif = alpha; % Instantaneous Granger causality test significance level.
VARadequacy_signif = 0.05; % VAR model adequacy significance level

%=========================================================================
%                    Plotting options
%=========================================================================

% The following parameters are used by xplot routine

%             [1 2 3 4 5 6 7]
flgPrinting = [1 1 1 3 3 0 2]; % See xplot.m for more detail.
%              | | | | | | 7 Power Spectra (0: w/o SS; 1: Linear; 2: Log-scale)
%              | | | | | 6 Coherence in gray line
%              | | | | 5 Confidence interval
%              | | | 4 Confidence interval(1:dashed;2:shaded;3:error barr
%              | | 3 Plot PDC in red line
%              | 2 Patnaik threshold level in black dashed-line
%              1 PDC in green line

flgColor = 1; % Choosing background color for exposing y-axis scale
%        flgColor - 0: white background;
%                   1: white [.1 1], light-blue [.01 .1], purple [0 .01]
%                   2: white [.1 1], light-gray [.01 .1], gray [0 .01]
flgScale = [1 2 3];
%        flgScale - 1: [0 1] / {if max(PDC/DTF) > 1}:[0 max(PDC/DTF)]
%                   2: [0 {if max(PDC/DTF) > 1}:max(PDC/DTF)]/[0 1]/[0 .1]/[0 .01]
%                                  based on flgMax(PDC/DTF/Threshold/CI/all)
%                   3: [0 max(PDC/DTF/Thr/CI/all)]
%                                  based on flgMax(PDC/DTF/Threshold/CI/all)
%        Plotting option for scaling y-axis according to the max
%        amplitude either of:
%                  PDC/DTF - max value of these measures,
%                  Thr - max Patnaik threshold,
%                  CI - max upper confidence interval value, max of
%                  TCI - max Threshold or confidence value, or
%                  all - max of all of them.
%                chosen by flgMax for eventually exposing small values
%                detail of PDC/DTF.
flgMax = 'TCI';
%        flgMax - {'PDC'; 'DTF'; 'Thr'; 'CI'; 'TCI'; 'all'} measure used as
%                   upper limit for scale. See flgScale.
flgSignifColor = 3;
%        flgSignifColor -  0: black lines
%                          1: black / gray -> significant /not signif PDC/DTF
%                          2: red  /  gray ->      "
%                          3: red  / green ->      "
%                          4: red  / black ->      "
```

```
%                           5: black / green ->     "

axis_scale = [0 0.50 -0.02 1.05];
w = fs*(0:(nFreqs-1))/2/nFreqs;
w_max = fs/2; %<***> Usually half of sampling frequency = Nyquist frequency


%=========================================================================
%=========================================================================
%        ATTENTION: BELOW THIS LINE PROBABLY YOU MIGHT NOT WANT TO EDIT,
%            UNLESS YOU WANT TO CUSTOMIZE YOUR ANALYSIS ROUTINE.
%=========================================================================

%=========================================================================
%                     Detrend and standardization options
%=========================================================================

[nChannels,nSegLength]=size(u);
if nChannels > nSegLength, u=u.';
   [nChannels,nSegLength]=size(u);
end;

if flgDetrend,
   for i=1:nChannels, u(i,:)=detrend(u(i,:)); end;
   disp('Time series were detrended.');
end;

if flgStandardize,
   for i=1:nChannels, u(i,:)=u(i,:)/std(u(i,:)); end;
   disp('Time series were scale-standardized.');
end;



%=========================================================================
% Additional info for title (optional)

strTitle1 = ['PDC(' '{\alpha = ' int2str(100*alpha) '%}' ') '];
switch metric
   case 'euc'
      %NOP
   case 'diag'
      strTitle1 = ['g' strTitle1];
   case 'info'
      strTitle1 = ['i' strTitle1];
   otherwise
      error('Unknown metric.')
end;
% or set strTitle1 = [];

%==================
switch alg
  case 1
    disp('VAR estimation using Nutall-Strand algorithm.')
```

```
  case 2
    disp('VAR estimation using least-squares estimator.')
  case 3
    disp('VAR estimation using Vieira-Morf algorithm.')
  case 4
    disp('VAR estimation using QR-Arfit algorithm.')
end;

%============================#
%MAR order selection criteria/
%============================#
switch criterion
   case 1
      disp('Model order selection criteria: AIC.')
   case 2
      disp('Model order selection criteria: Hanna-Quinn.')
   case 3
      disp('Model order selection criteria: Schwartz (BIC).')
   case 4
      disp('Model order selection criteria: FPE.')
   case 5
      disp('Model order selection criteria: fixed order in maxIP.')
   otherwise
      error('Model order selection criteria: NOT IMPLEMENTED YET.')
end;

%=========================================================================
%                            VAR model estimation
%=========================================================================
[IP,pf,A,pb,B,ef,eb,vaic,Vaicv] = mvar(u,maxIP,alg,criterion);


disp(['Number of channels = ' int2str(nChannels) ' with ' ...
  int2str(nSegLength) ' data points; MAR model order = ' int2str(IP) '.']);

%=========================================================================
%    Testing for adequacy of MAR model fitting through Portmanteau test
%=========================================================================
   h = 20; % testing lag
   aValueVAR = 1 - VARadequacy_signif;
   flgPrintResults = 1;
[Pass,Portmanteau,st,ths]=mvarresidue(ef,nSegLength,IP,aValueVAR,h,...
                                                        flgPrintResults);

%=========================================================================
%         Granger causality test (GCT) and instantaneous GCT
%=========================================================================
   flgPrintResults = 1;
[Tr_gct, pValue_gct, Tr_igct, pValue_igct] = gct_alg(u,A,pf,gct_signif, ...
                                                    flgPrintResults);

%=========================================================================
%            PDC, threshold and confidence interval calculation.
%=========================================================================
```

```
309
310 % if alpha == 0, no asymptotic statistics is performed. ASYMP_PDC returns
311 % only the PDC. This option is much faster!!
312  c=asymp_pdc(u,A,pf,nFreqs,metric,alpha);
313
314 % Power spectra and coherence calculation
315 c.SS = ss_alg(A, pf, nFreqs);
316 c.coh = coh_alg(c.SS);
317
318 % Statistically significant PDC on frequency scale
319 if alpha ≠ 0,
320    pdc_temp = ((abs(c.pdc)-c.th) > 0).*c.pdc + ((abs(c.pdc)-c.th) ≤ 0)*(-1);
321    pdc_temp(ind2sub(size(pdc_temp),find(pdc_temp == -1))) = NaN;
322    c.pdc_th = pdc_temp;
323 end;
324
325 %Adding further analysis details in the figure title.
326 %strTitle3 = ['[N=' int2str(nSegLength) '; IP=' int2str(c.p) ']'];
327 % or
328
329 strTitle3 = ['[N=' int2str(nSegLength) 'pts; IP=' int2str(c.p) '; ' ...
330    datestr(now) ']'];
331
332 % or leave emptied: strTitle3=[];
333
334 %=========================================================================
335 %              Matrix Layout Plotting of the Analysis Results
336 %=========================================================================
337
338 w_max = fs/2;
339 strTitle = [strTitle1 strTitle2 strTitle3];
340 strWindowName = 'pdc Analysis Template Example';
341
342
343 % The following "for loop" through flgColor values, 0 and 1, and yields a
344 % pair of plots, one without and other with color scale rearrangement option.
345 % Value range of PDC and Coherence is from [0 1], but sometimes the maximum
346 % peak value is small (<0.1), or even smaller, (<.01), so in these cases it
347 % might be interesting to have a plot with finer smaller y-axis scale. The
348 % white-background plot indicates full-scale [0 1] y-axis, while
349 % light-blue-background stands for intermediate [0 .1] scaling and
350 % light-purple-background shows very fine detail of small, usually not
351 % significant PDCs. Try flgColor = 0 or 1, or both [0 1].
352
353 for kflgScale = flgScale,
354    h=figure;
355    set(h,'NumberTitle','off','MenuBar','none', ...
356       'Name', strWindowName )
357
358    [hxlabel,hylabel] = xplot(c,flgPrinting,fs,w_max,chLabels, ...
359                              flgColor,kflgScale,flgMax,flgSignifColor)
360
361 % The title suplabel command should (not sure) follow the xplot routine
362 % In MacOS X, for flgPrinting(7) = 4 or 5, the main diagonal plotting
```

```
% gets misaligned if suplabel with 't' option is used more than once.

   [ax,hT]=suplabel( strTitle, 't' );
   set(hT,'FontSize',8)
end;

tilefigs

%========================== xplot ========================================
%Plot legend:  Blue lines on the main diagonal = Power spectra;
%              Black dashed lines are Patnaik threshold for pdcn;
%              Green lines = non significant pdcn;
%              Red lines = significant pdcn;
%              Light-gray lines = coherence function.
%
% Notes:       a.The main diagonal of matrix layout contains power spectra.
%              b.Coherences are symmetric, e.g.,
%                   Coh_{Sunspot,Melanoma}(f) = Coh_{Melanoma,Sunspot}(f).
%              c.pdcn is asymmetric relation, and the pdcn graphics should
%              be read as if the flow of information is been from the
%              x-axis variable toward y-axis variable.
%
%              For sunspot and melanoma example, one only sees significant
%              pdcn from Sunspot to Melanoma, which could eventually be
%              interpreted that "Sunspot", or the Sun's activity
%              modulates the incidence of melanoma.
%========================== xplot ========================================
disp('======================================================================');
disp('===========pdc_ANALYSIS_TEMPLATE SUCCESSFULLY FINISHED ==============')
disp('======================================================================');
```