Projets Atelier « *SSA* » (Analyse des données)

All data referred to below as well as related files can be found on the following website: http://www.environnement.ens.fr/perso/claessen/ateliers/SSA/

General remarks

In general, the idea is to formulate a **research question** about the system in question (see description of projects below), and to use the SSA-MTM Toolkit to answer that question. This sounds obvious, but it may not be so easy to find a good (and answerable) question, nor to find the answer that particular question. Obviously, what needs to be avoided is collecting many results in different directions without much coherence. So it may be worthwhile to take some time in the beginning to think carefully about the project and what you would like to do with it. Some questions are presented below that can be used as a starting point for your project.

What can you do with SSA, and the other methods offered by the SSA-MTM toolkit?

- Analyse the spectrum of a time series (BT, MEM)
- Filter a time series (SSA)
- Analyse the filtered time series (SSA+BT or MEM)
- Filter a time series of multiple variables at the same time (MSSA)
- Analyse the spectrum of the resulting filtered time series one by one (MSSA+BT or MEM)

The analysis of time series of multiple variables is particularly interesting for detecting the influence of one process on another one, for example, the dynamics of a population may be forced by El Nino or other fluctuating climate factors. This may be detected, for example, by finding the same peak in the power spectrum of either time series.

Filtering Different options are possible. (1) You can look for the major oscillatory component(s) of a time series, and leave out the rest (supposedly noise). This way you filter out the noise and keep the supposedly essential oscillations of the time series. (2) You can try to filter our a periodic but “uninteresting” oscillatory component: for example your data may contain an annual cycle that is due to seasonality but your are more interested in the existence of multi-annual cycles (perhaps caused by climate oscillations such as El Nino), that may be obscured by the presence of the annual cycle. In that case you need to identify which SSA components correspond to this uninteresting annual cycle and reconstruct the time series *without* these components (but possibly leaving noise components in the reconstructed time series; and of course leaving the interesting oscillations in the reconstructed time series). Often, it is interesting or necessary to first do (2) and then do (1) to find the interesting oscillations.

Data handling From the above it becomes clear that in doing such analyses, careful data handling is very important, because during the process you will create new data sets (the results of filtering), which subsequently need to be analysed further. For example, if we start with a time series in a vector called “data” that was read from a file, doing an SSA will give us a filtered time series called “ssarcvec”. In order to proceed with a spectral method (BT or MEM) we can either write this new time series to a file and then open it, or run a spectral analysis directly on the vector “ssarcvec”. Then in a next step, based on the found power spectrum, we might want to re-filter the time series, thus creating again a new data set. Giving adequate names to data vectors and matrices (and files) becomes very useful.

Of course if we are working on multivariate time series, this could become even more complicated, since we may have to proceed with a subsequent analysis using only one particular column of a matrix ("mssarcmat") that was produced by doing an MSSA reconstruction (and then repeat the same analysis on another column, etc). For the latter operation, please read carefully the part "**3.a. Getting started**" (File/Data etc) of the SSA-MTM toolkit online User guide (see atelier website for link).

Data pre-treatment Even before starting an SSA or other analysis with the toolkit, you may have to treat your data to put them in a format that can be used in the software, and that is appropriate for your particular research question. **Reading data file:** First, make sure there are no empty lines in the data file, nor text lines (i.e., remove the header row). Second, the toolkit can read either a vector, or a matrix. If you ask to read a vector, and if the file contains only a single column, then of course that column will be read. If the file contains two columns, however, the toolkit will read only the second column. If you ask to read a matrix, the toolkit will read all columns in the file. In a next step, you can of course select one column of the data matrix to do a mono-variate SSA or something like that (see again User guide, Getting started, Matrix/Vector). **Normalising data**: this seems to be unnecessary for SSA of single variables, but appears very important for Multichannel SSA (MSSA). So you have to make sure that each column of data has mean=0 and SD=1. This can be done in any spreadsheet program such as Open Office Calc.

Common sense Finally, before starting any analysis and data treatment, it is highly recommended to plot your data in different ways to do a visual inspection of the data you have in your hands. This might give you some ideas that can guide your analysis: for example you may recognise fluctuations (and have a rough idea of a their period); you may suspect correlations between different time series; etc. For example, the choice of M (the window size) may be influenced by such kind of first explorations. Also, you may want to compute new, derived time series based on the "raw" data that may be more interesting to analyse than the raw data (think, for example, of Catch per Unit Effort, CPUE, rather than total catch or total effort).

Project « Lake Windermere »

The data set contains data on two fish populations, perch and pike. See Winfield et al 2008 for a general introduction of the lake system and the environmental variables. The columns in the data set contain the following variables:

```
year
perch3_N        perch at least 3 years old, North Basin
perch2_N        perch 2 years old, North Basin
perch3_S        perch at least 3 years old, South Basin
perch2_S        perch 2 years old, South Basin
pike3_N         idem for pike
pike2_N
pike3_S
pike2_S
effortperch_N   fishery effort on perch in North Basin
effortperch_S   fishery effort on perch in South Basin
```

```
effortpike_N     idem for pike
effortpike_S
temperature
wind.N           wind speed North Basin
wind.S           wind speed South Basin
phosph_N         P concentration North Basin
phosph_S         P concentration South Basin
```

The fish populations abundances are catch-mark-recapture estimates of population size; they are hence not similar to “catch” data in fisheries data (see below). Fish younger than 2 years are invulnerable to the fishing method; hence there is no data on this age class. The ecosystem is influenced by several environmental factors that change over time: harvesting, temperature, wind and nutrient load (phosphorus).
NB some time series (temperature, wind, P) are **incomplete**. So to read the data as a matrix into the toolkit, you have to choose to either (i) remove these columns; or (ii) remove the rows that contain missing values.

Different questions can be asked about this data set:

- Do young and old perch/pike fluctuate in similar fashion?
- Do pike and perch fluctuate in similar fashion? Can you detect evidence of coupling between the fish species?
- In 1976 an epidemic decimated the perch population (affecting old individuals mostly). Can you detect a change in the dynamics corresponding to this event?
- Is there evidence for exogenous forcing of the fish dynamics?

References (see the PDF on the atelier website)
Edeline, E., S. M. Carlson, L. C. Stige, I. J. Winfield, J. M. Fletcher, J. B. James, T. O. Haugen, L. A. Vøllestad, N. C. Stenseth. Trait changes in a harvested population are driven by a dynamic tug-of-war between natural and harvest selection. *PNAS* 140:15799-15804
Winfield, Ian J.; James, J. Ben; Fletcher, Janice M.. 2008 Northern pike (Esox lucius) in a warming lake: changes in population size and individual condition in relation to prey abundance. *Hydrobiologia*, 601 (1). 29-40. doi:10.1007/s10750-007-9264-1

Project « Collembola »

The data is the result of an experiment done by Thomas Tully, during his PhD at the ENS in the Ecology & Evolution lab (UMR 7625). In the experiment, the population dynamics of 5 different clones (genotypes) of the clonally reproducing collembolan Folsomia candida are monitored under controlled (identical) conditions. For each clone, 4 replicates are followed, each in a separate box. For each box, the total number of individuals, as well as the numbers subdivided in six different size classes are counted. The life stages are defines as size classes of 0.5 mm intervals. The idea is that the different clones may have different population dynamical behaviour. The main data file contains the following data columns:

```
clone       clone indentity (AP, DK, GB, TO, or US)
box         clone & replicate number
num         replicate number
day         the date of measurement
total       total population in the box
st0         number in stage 0 (0.5    1 mm)
st1         number in stage 1 (1    1.5 mm)
st2         etc
```

```
st3
st4
st5
st6         number in stage 6 (>3 mm)
```

A problem with this data is that the points of the time series are not equidistant in time (see the data). SSA requires that all time intervals are equal (question for you: why?). For this atelier we will ignore this complication and treat the data as if they were equidistant. However, it may be necessary to remove for instance the last point of each series, as the last interval is relatively large.

Possible questions:

- What are the characteristics of the dynamics of total population size? Are there differences between clones? Are the dynamics consistent between the replicates of a single clone?
- Do different size classes fluctuate in a correlated way?

References (PDFs on website)

Tully T & Ferriere R (2008) Reproductive flexibility: genetic variation, genetic costs, and long-term evolution in a collembola. PLoS One 3(9): e3207 doi:10.1371 /journal.pone.0003207.

Tully T (2004) Facteurs maternels, génétiques et environnementaux de l'expression des traits d'histoire de vie chez le collembole Folsomia candida, Willem: Université Pierre et Marie Curie, Paris 6.

Project « Tuna fishery and Indian Ocean climate »

The data of this project have been analysed and published by Meynard et al 2007, using wavelet analysis. (A short introduction to wavelet analysis is found in their article, as well as in the review by Ghil et al 2002). The study concerns fishery data of two tuna populations (bigeye tuna and yellowfin tuna) in the Indian Ocean. The data includes total fishing effort and the catch data for each species during 1955-2002. A separate time series for yellowfin tuna in another region is given for a shorter period (yellowfin tuna, 1984-2001). A time series of the IOI (Indian Ocean Index) is also given. The data file contains the following data:

```
year
month
IOI         The climatic index (Indian Ocean Index)
#hooks            Fishing effort
Cbet        Catch of BET = Bigeye tuna
Cyft        Catch of YFT = Yellowfin tuna
Ctot        Total catch
cpue bet    Catch per unit effort BET
cpue yft    Catch per unit effort YFT
cpue tot    Catch per unit effort total
```

Possible questions:

- What is the influence of climatic oscillations [based on the Indian Oscillation Index (IOI)] on monthly catch rates of two tropical tuna species in the equatorial Indian Ocean?

- Are the tuna (two species) population dynamics correlated?

- Is the fishery activity influenced by climatic conditions?
- Are catch and effort indeed correlated, as expected?

References (PDFs on website)

Ghil M., R. M. Allen, M. D. Dettinger, K. Ide, D. Kondrashov, M. E. Mann, A. Robertson, A. Saunders, Y. Tian, F. Varadi, and P. Yiou, 2002: "Advanced spectral methods for climatic time series," Rev. Geophys.,40(1), pp. 3.1-3.41, 10.1029/2000RG000092

Ménard, F. F., Marsac, F., Bellier, B. & Cazelles, B., 2007. Climatic oscillations and tuna catch rates in the Indian Ocean: A wavelet approach to time series analysis. Fisheries Oceanography, 16, 95-104.