

BioVeL – Biodiversity Virtual e-Laboratory

Workflow Documentation

Parametric Bootstrap or Resample a projection matrix Workflow for local execution

July 2014

Capacities Programme of Framework 7: EC e-Infrastructure Programme –
e-Science Environments - INFRA-2011-1.2.1

Grant Agreement No:
Project Co-ordinator:
Project Homepage:
Duration of Project:

283359
Mr Alex Hardisty
<http://www.biovel.eu>
36 months



Parametric Bootstrap or Resample a projection matrix Workflow for local execution

1 Description

The Parametric Bootstrap or Resample a projection matrix Workflow provides an environment to resample a projection matrix using a multinomial distribution for transitions and a log normal distribution for fertilities (Stubben, Milligan, and Nantel. 2011). The resample is based on number of plants surveyed. The projection matrix A is first split into separate transition and fertility matrices. Dead fates are added to the transition matrix and the columns are then sampled from a Multinomial distribution based on the size in each corresponding stage class in n. The fertility rates are sample from a Log Normal distribution using the Inorms function. The same variance is applied to all rates by default. (Stubben, Milligan and Nantel 2013, Caswell 2001 see section 12.1.5.2).

The goal of a demographic analysis is very often to estimate lambda, because lambda is estimated from imperfect data, such estimation are uncertain. Therefore, when the results have policy implications it is important to quantify that uncertainty. Confidence interval is one of the traditional tools to doing so (see outputs: **Confidence interval of Lambda**).

Analyses:

- Lambda (λ)
- Mean matrix
- Variance matrix
- Histogram
- Confidence interval of Lambda
- X= List of resampled matrices

2 General

2.1 Name of the workflow in myExperiment.

Name: Parametric Bootstrap Workflow for local execution

Download: The pack can be downloaded at myExperiment under the following address:
<http://www.myexperiment.org/packs/643.html> or the workflow:
<http://www.myexperiment.org/workflows/4309.html>

2.2 Date, version and licensing

Last updated: 25th July 2014

Version: 3

Licensing: CC-BY-SA

2.3 How to cite this workflow

To report work that has made use of this workflow, please add the following credit acknowledgement to your research publication:

The input data and results reported in this publication (tutorial) come from data (Dr. Gerard Oostermeijer unpublished results and publication: Oostermeijer, J.G.B. M.L. Brugman, E.R. de Boer; H.C.M. Den Nijs. 1996. Temporal and Spatial Variation in the Demography of *Gentiana pneumonanthe*, a Rare Perennial Herb. *The Journal of Ecology*, 84: 153-166.) using BioVeL workflows and services (www.biovel.eu). Parametric Bootstrap workflow was run on *<date of the workflow run>*. BioVeL is funded by the EU's Seventh Framework Program, grant no. 283359.

3. Scientific specifications

3.1 Keywords

BioVeL, Matrix Population Models, stage matrix with not density dependence, Lambda (λ), Confidence intervals of λ , Parametric Bootstrap or Resample a projection matrix, mean matrix, variance matrix.

3.2 Scientific workflow description

The aim of the Parametric Bootstrap or Resample a projection matrix Workflow is to resample a projection stage matrix using a multinomial distribution for transitions and a log normal distribution for fertilities (Stubben, Milligan, and Nantel. 2011). The workflow accepts input data in a .txt format. The output is provided as a set of R results and a graphic plot.

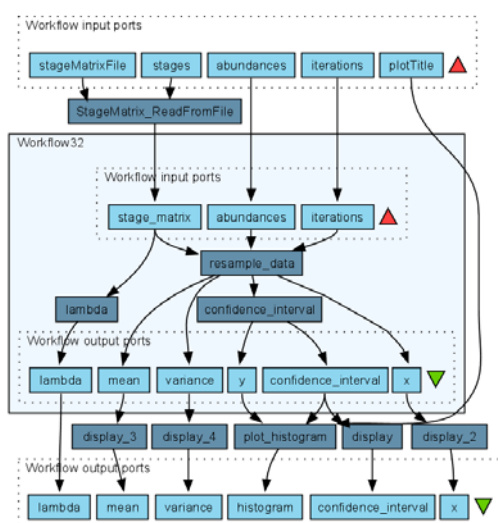


Figure 1. Parametric Bootstrap workflow in Taverna workbench

For more detailed description of the functions, please visit the Tutorial section.

4 Technical specifications

4.1 Execution environment and installation requirements

The Workflow requires a Taverna Engine. The simplest way to install a Taverna Engine is to install Taverna Workbench. The workflow also requires an Rserve installation with popbio package installed. It is possible to setup the workflow to use a remote Rserve. However, instructions for installing a local Rserve are provided below.

4.2 Taverna installation, including updates and plugins

4.2.1 Taverna installations

- Taverna Workbench: Version 2.4 or 2.5. For installation files and instructions, please go to <http://www.taverna.org.uk/download/workbench>

4.2.2 Taverna Dependencies

- Install R software in your computer. See: <http://www.r-project.org/>
- Start R, and install package Rserve:
 - `install.packages("Rserve")`
- Install package popbio
 - `install.packages("popbio")`
- Local R Server: (Rserve) running at port 6311. See <https://wiki.biodel.eu/x/3ICD> for additional information.

4.2.3 How it works

- First, open R, once R is opened, type `library(Rserve)` and press enter; then type `Rserve()` and press enter again. You will see then the following message: Starting Rserve...

```
"C:\PROGRA~1\R\R-30~1.1\library\Rserve\libs\x64\Rserve.exe"
```

After this operation you can open Taverna and run the workflow.

5 Tutorial

5.1 Introduction

In this tutorial, you will use an input file MTers87_88.txt, called in myexperiment.org: Stage Matrix of *Gentiana pneumonanthe* 1987-88 File. This file will be used to run this workflow. Finally you will be able to export your results to different formats.

5.2 Input data.

5.2.1 Data preparation/format

The workflow accepts input data (matrices) in a .txt format, all decimal numbers in the matrix must be indicated by dots e.g.: 0.578. The example matrix for the tutorial is available in: <http://www.myexperiment.org/packs/643.html> or here below.

5.2.2 Input data

The input files are in a .txt format: to download click here in each file:

Terschelling

- [MTers87_88.txt](#)

Example from:

J. Gerard B. Oostermeijer; M.L. Brugman; E.R. de Boer; H.C.M. Den Nijs. 1996. Temporal and Spatial Variation in the Demography of *Gentiana pneumonanthe*, a Rare Perennial Herb. *Journal of Ecology*, Vol. 84(2): 153-166.

5.3 Select input data dialogue boxes.

The first step is to fill out the input ports:

5.3.1 INPUTPORTS

1) **Initial abundance:** In this dialogue appear the fields with the initial abundance per stage observed in the field (Fig 2). As an example *Gentiana pneumonanthe* has 5 stages with its respective abundance.

The abundance of the stages or categories must be added one by one. First press Add value, fill the abundance of the first stage and press enter; then press Add value and fill once again the next stage abundance; repeat the action until you have fill all the abundance of all stages (Fig 2).

Parametric Bootstrap or Resample a projection matrix Workflow for local execution

Stage abundance of the year 1987:

S (seedlings)	69
J (Juveniles)	100
V (vegetative)	111
G (reproductive individuals)	21
D (dormant plants)	43

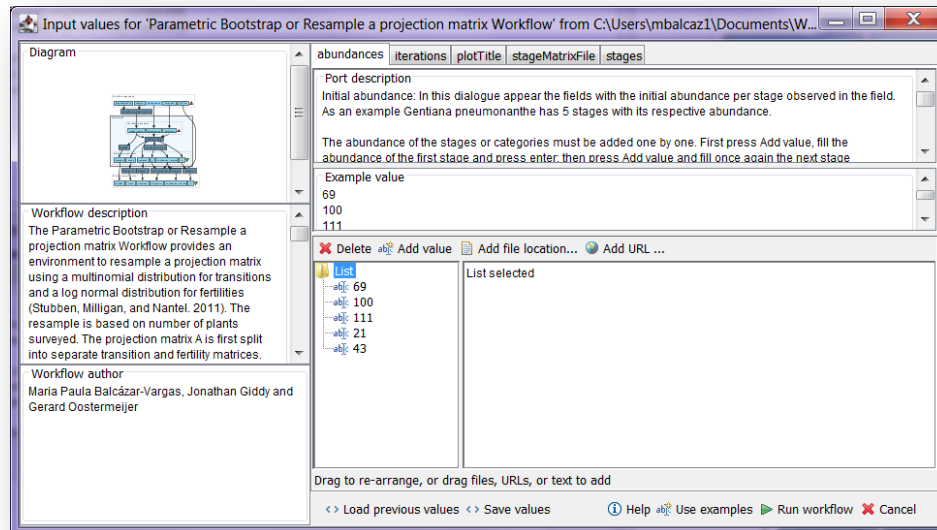


Figure 2. Abundance input. Fill in the abundance per stage appears.

2) ***Iterations***: Number of iterations for calculation of the resample analysis of the parametric Bootstrap. Click in Set Value (red arrow in Fig 3), tip the number in the right window.

e.g.: 1000

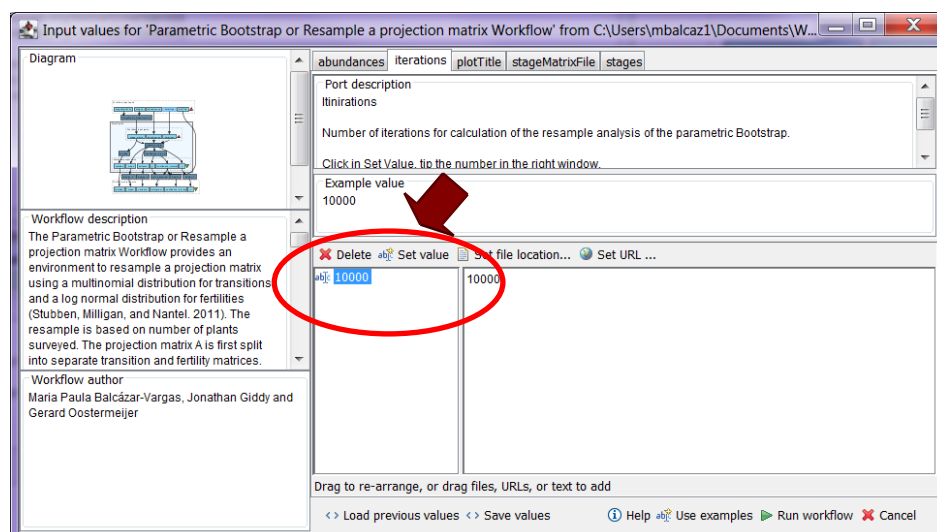


Figure 3. A number of iteration's to be filled in.

Parametric Bootstrap or Resample a projection matrix Workflow for local execution

3) **plotTitle:** Descriptive title for labelling generated output graph. Click in Set Value (red arrow in Fig 4), tip the text in the right window.

e.g.: *Gentiana pneumonanthe*, Terschelling (species taxonomic name, place where the research was conducted)

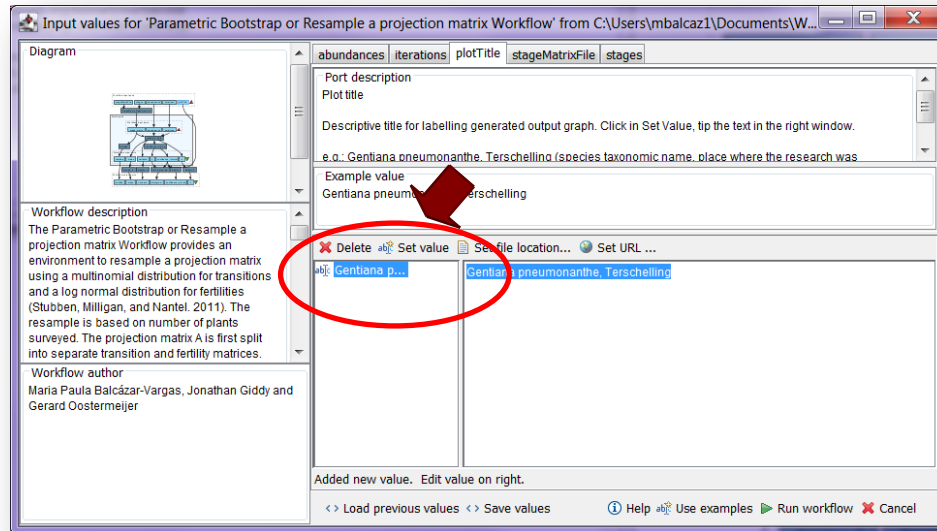


Figure 4. Text to be filled in.

4) **StageMatrixFile:** Matrix in a .txt file format. Click in Set file location (red arrow in Fig 5), a window dialog appears and the user selects the file (e.g. [MTers87_88.txt](#), Fig 6) and then clicks the Open button. Then the link to the file will appear on the Taverna window.

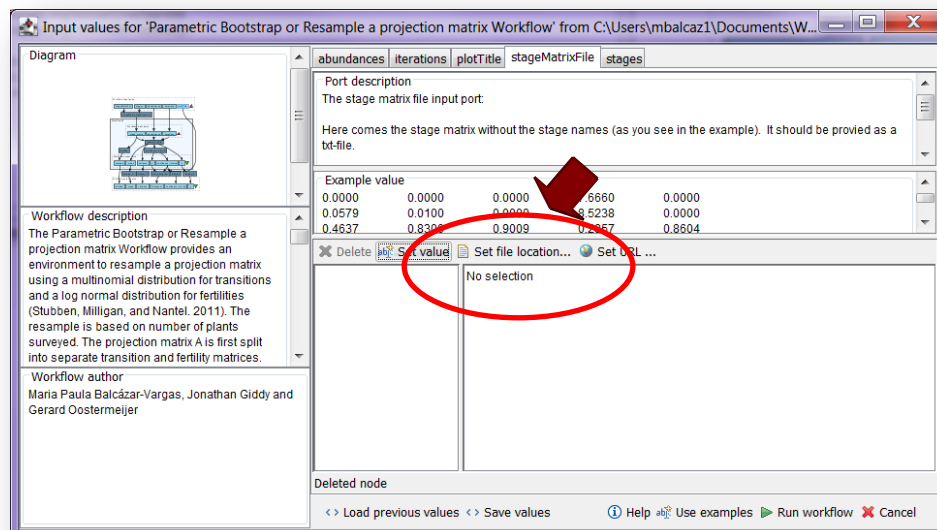


Figure 5. Set the Location of the matrix file.

Parametric Bootstrap or Resample a projection matrix Workflow for local execution

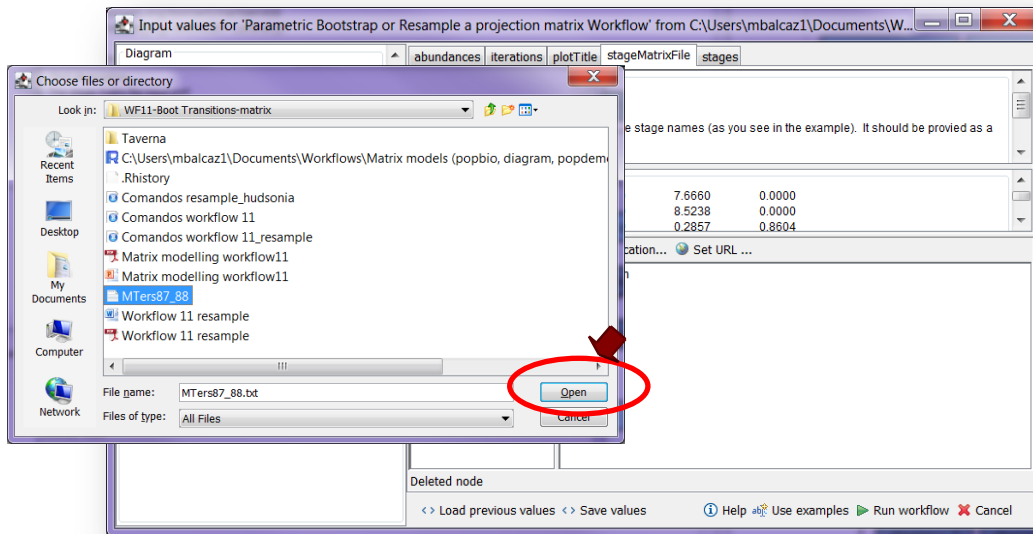


Figure 6. Choose the matrix file and click in Open.

5) **Stages:** the names of the stages or categories of the input matrix. In the following example, the matrix has 5 stages or categories. The respective name stages must be filled one by one. First press add value (see arrow in Fig 7), fill a stage name (not longer than 8 characters) and press enter; then press add value and fill once again the next stage name, repeat the action until you have fill all the stages names.

In the following example, the life cycle of *Gentiana pneumonanthe* has 5 stages or categories:

- | | |
|-----------------------------|---|
| 1) Seedlings | S |
| 2) Juveniles | J |
| 3) Vegetative | V |
| 4) Reproductive individuals | G |
| 5) Dormant plants | D |

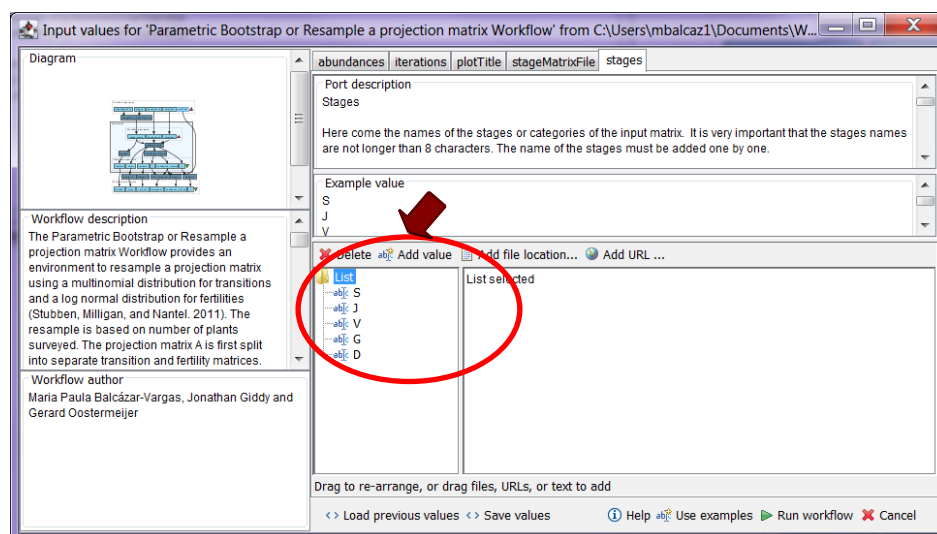


Figure 7. Stages to be filled in.

After the user has filled out the input ports and has clicked on the **Run Workflow button**, to complete

Parametric Bootstrap or Resample a projection matrix Workflow for local execution

all the analysis may take several minutes depend on the number of iterations.

When the analyses are completed, they appear on different windows under results in Taverna, the user have to save each output separately.

5.4 Save data/results

5.4.1 OUTPUTS

First, click in the selected result window e.g. **confidence_interval** (green oval in Fig 8). Second, click in the left window on value 1 (blue oval), Third, on the right window click on the save value (red oval). Fourth, name the file and determinate the extension file. For a text file: e.g. .csv or .txt. For images: .jpg or png (Fig 9), Finale, save the file in the chosen map.

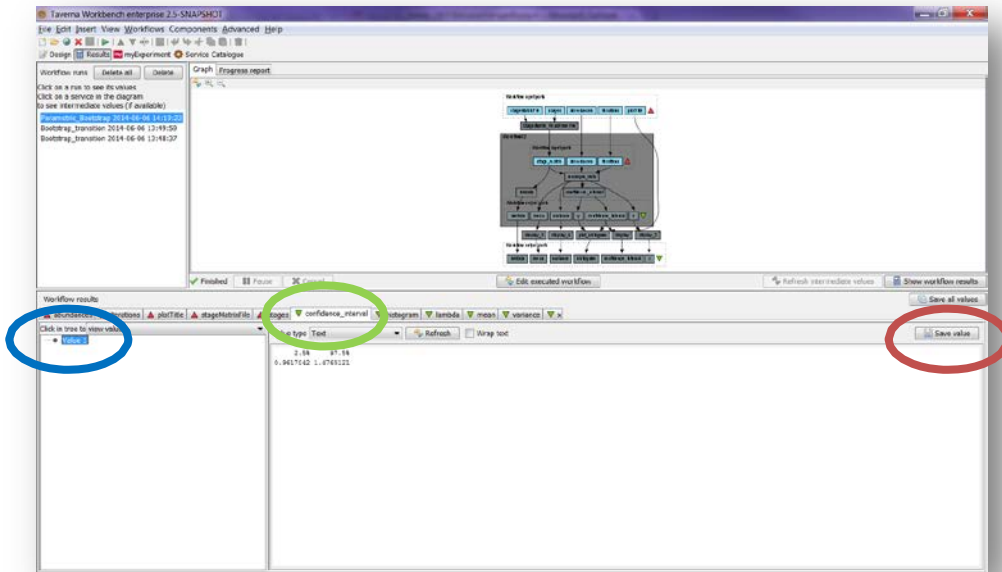


Figure 8. Confidence interval output window.

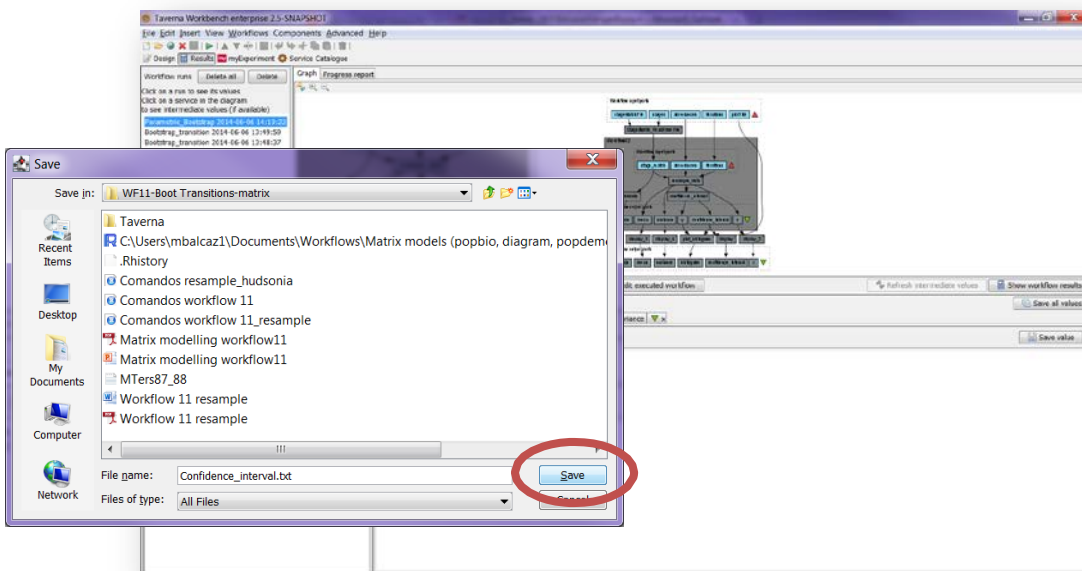


Figure 9. Name the file and determinate the extension file. For a text file: e.g. .csv or .txt. For images: e.g.: .jpg or png.

Parametric Bootstrap or Resample a projection matrix Workflow for local execution

Bootstrap analysis

1) **Confidence_interval**: 95% Confidence interval of Lambda. In statistics, a confidence interval (CI) is a type of interval estimate of a population parameter and is used to indicate the reliability of an estimate. It is an observed interval (i.e. it is calculated from the observations), in principle different from sample to sample, that frequently includes the parameter of interest if the experiment is repeated. How frequently the observed interval contains the parameter is determined by the confidence level or confidence coefficient

e.g.: 2.5% 97.5%
 0.9617 1.4765

2) **CI_histogram**: Histogram plotting the frequencies of the lambda values and the 95% confidence intervals resulting from the bootstrap analysis (Fig 10).

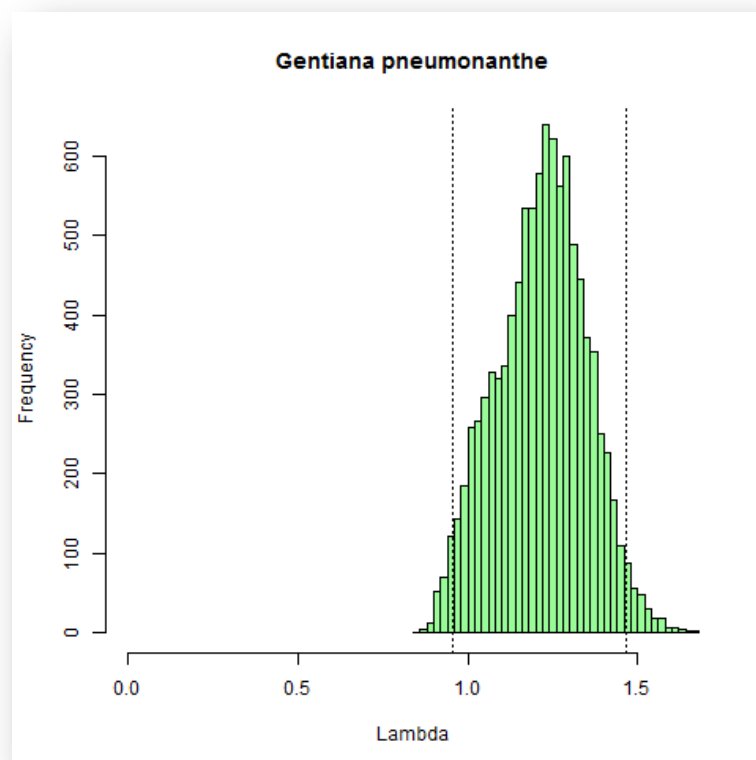


Figure 10. The confidence interval histogram opens as a graph.

3) **Lambda (λ)**: Lambda or dominant eigenvalue, this value describes the population growth rate of a stage matrix. The population will be stable, grow or decrease at a rate given by lambda: e.g.: $\lambda = 1$ (population is stable), $\lambda > 1$ (population is growing) and finally $\lambda < 1$ (population is decreasing).

e.g.: 1.2375

4) **Mean matrix**: the mean matrix is calculated based on the matrices generated by bootstrap analysis and the number of the matrices at the same time is based on the number of iterations (Fig 11).

Parametric Bootstrap or Resample a projection matrix Workflow for local execution

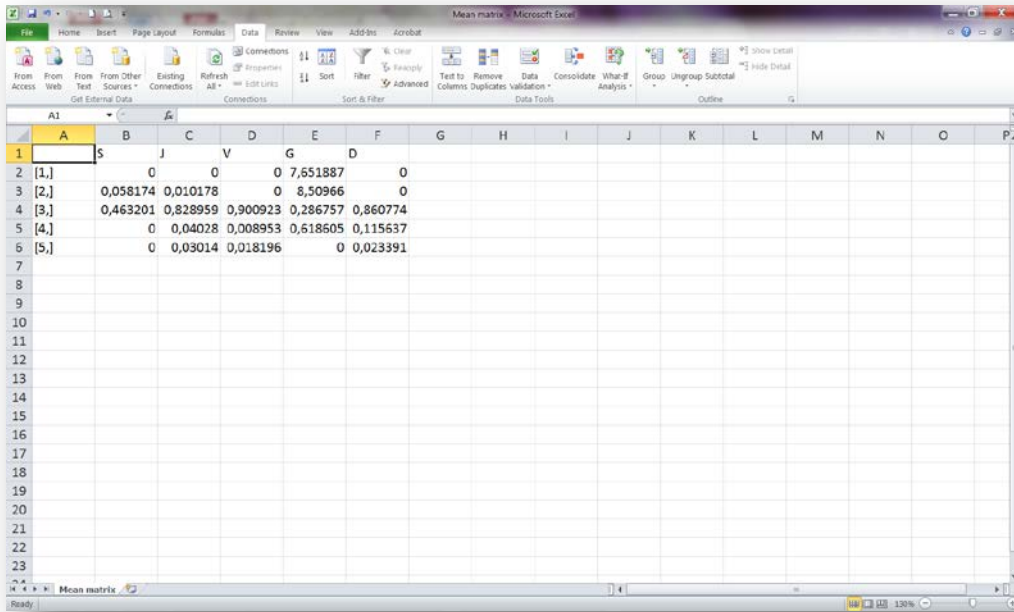


Figure 11. *The mean matrix* opens in excel.

- 5) **Variance matrix**: the variance matrix is calculated based on the matrices generated by bootstrap analysis and the number of the matrices at the same time is based on the number of iterations (Fig 11).

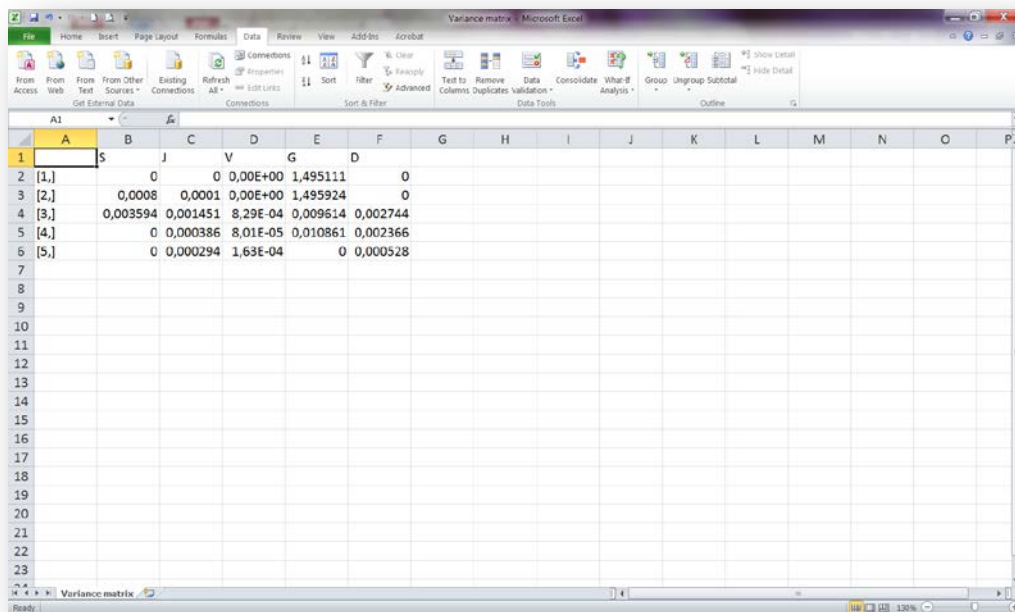


Figure 11. *The variance matrix* opens in excel.

- 6) **X= List of resampled matrices**: it is the list of matrices generated by bootstrap analysis and the number of the matrices at the same time is based on the number of iterations..

6 Support

For questions with using the workflow, please write [support @biovel.eu](mailto:support@biovel.eu)

For definitions of technical and biological terms, please visit the BioVeL glossary page:
<https://wiki.biovel.eu/display/BioVeL/Glossary>

7 References

This workflow was created using and based on Packages 'popbio' in R. (Stubben & Milligan 2007; Stubben, Milligan & Nantel 2011).

Caswell, H. 2001. Matrix population models: Construction, analysis and interpretation, 2nd Edition. Sinauer Associates, Sunderland, Massachusetts.

Jongejans E. & H. de Kroon. 2012. Matrix models. Chapter in Encyclopaedia of Theoretical Ecology (eds. Hastings A. & Gross L) University of California, p415-423

Oostermeijer J.G.B., M.L. Brugman; E.R. de Boer; H.C.M. Den Nijs. 1996. Temporal and Spatial Variation in the Demography of *Gentiana pneumonanthe*, a Rare Perennial Herb. The Journal of Ecology, Vol. 84(2): 153-166.

Stubben, C & B. Milligan. 2007. Estimating and Analysing Demographic Models Using the popbio Package in R. Journal of Statistical Software 22 (11): 1-23

Stubben, C., B. Milligan, P. Nantel. 2011. Package 'popbio'. Construction and analysis of matrix population models. Version 2.3.1

7.1 Acknowledgements

7.1.1 Authors

1. **Maria Paula Balcazar-Vargas** Instituut voor Biodiversiteit en Ecosysteem Dynamica (IBED), Universiteit van Amsterdam.
2. **Jonathan Giddy** Cardiff School of Computer Science and Informatics, Cardiff University, Cardiff CF24 3AA, United Kingdom.
3. **J. Gerard B. Oostermeijer** Instituut voor Biodiversiteit en Ecosysteem Dynamica (IBED), Universiteit van Amsterdam.

Parametric Bootstrap or Resample a projection matrix Workflow for local execution

7.1.2 Project funding

The workflow described in this documentation has been designed and implemented as part of the BioVeL project.

BioVeL is funded by the European Commission 7th Framework Programme (FP7) as part of its e-Infrastructures activity. Under FP7, the e-Infrastructures activity is part of the Research Infrastructures programme, funded under the FP7 'Capacities' Specific Programme. It focuses on the further development and evolution of the high-capacity and high-performance communication network (GÉANT), distributed computing infrastructures (grids and clouds), supercomputer infrastructures, simulation software, scientific data infrastructures, e-Science services as well as on the adoption of e-Infrastructures by user communities.

7.2 Publications