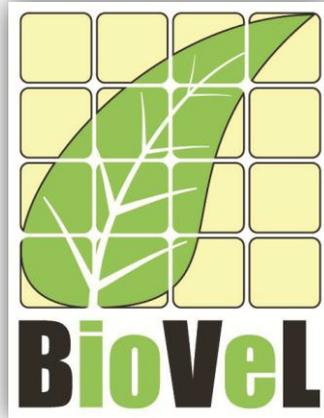


Bootstrap of observed census transitions (Confidence intervals of λ) workflow for local execution



BioVeL – Biodiversity Virtual e-Laboratory

Workflow Documentation

**Bootstrap of observed census transitions
(Confidence intervals of λ) 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



Bootstrap of observed census transitions (Confidence intervals of λ) workflow for local execution

1 Description

The Bootstrap of observed census transitions (confidence intervals of λ) workflow for local execution calculates bootstrap distributions of population growth rates (λ), stage vectors, and projection matrix elements by randomly sampling with replacement from a stage-fate data frame of observed transitions. 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: Bootstrap analysis).

A detailed description of resampling methods to estimate confidence intervals for demographic estimates is described Caswell (2001, Chapter 12)

2 General

2.1 Name of the workflow in myExperiment.

Name: Bootstrap of observed census transitions (Confidence intervals of λ) workflow for local execution.

Download: The pack can be downloaded at myExperiment under the following address:
<http://www.myexperiment.org/packs/642.html> or the workflow:
<http://www.myexperiment.org/workflows/3295.html>

2.2 Date, version and licensing

Last updated: 24th July 2012

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:

Bootstrap of observed census transitions (Confidence intervals of λ) workflow for local execution

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). Bootstrap of observed census transitions (Confidence intervals of λ) 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

Matrix Population Models, Lambda (λ), Bootstrap of observed census transitions (confidence intervals of λ).

3.2 Scientific workflow description

The aim of the Bootstrap of observed census transitions workflow calculates bootstrap distributions of population growth rates (λ), stage vectors, and projection matrix elements by randomly sampling with replacement from a stage-fate data frame of observed transitions. 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.

The workflow accepts input data in a .txt format. The output is provided as a set of R results and a graphic plot (Fig 1):

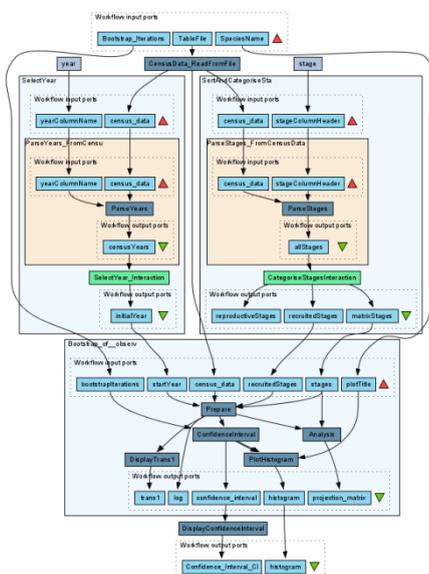


Figure 1. Bootstrap of observed census transitions workflow in Taverna workbench.

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 including the Interaction Service plugin. The simplest way to install a Taverna Engine is to install Taverna Workbench, and then install the Interaction Service plugin.

The workflow also requires an Rserve installation with the 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 Plug-ins

- Please install the [Interaction](#) plugin: version 1.0.2+

4.2.3 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.

4.2.4 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 *Gentiana87-91.txt*, called in *myexperiment.org*: Demographic data *Gentiana pneumonanthe* 1987-1991 from Terschelling File. You will use this file to perform the bootstrap analysis. Finally you will be able to export your results to different formats.

5.2 Input data.

5.2.1 Data preparation/format

The input data (a .txt-file) has to have the format of a table containing the demographic data on a series of individuals in two years. Each individual has a table row for each year and is identified by a number (plant column in example). For the year specified in the column 'year', each individual has a certain life stage ('stage'). The stage codes can be chosen by the user, but they have to be specified with a string. For all individuals, the number of offspring is specified in a chosen way (in the example, 'reprstr' gives the number of flowers/fruits for each plant. In the example, it can be seen that only generative adults (stage=G) had flowers. Individuals without offspring have to be indicated by filling in '0' in this column, which cannot have empty cells. The column 'recruitment' specifies those individuals that are new in the second year census to the population by means of a code. In the example, 'RS' is a new seedling, and 'RJ' a new juvenile plant. As can be seen in the example, no code is needed in this column for individuals that were already present (Table 1).

Table 1. Input file in a .txt format to create a stage matrix.

plant	year	stage	reprstr	recruitment
2	1987	D	0	
4	1987	J	0	
5	1987	D	0	
31	1987	V	0	
36	1987	D	0	
37	1987	J	0	
41	1987	J	0	
43	1987	S	0	
46	1987	V	0	
47	1987	S	0	
2	1988	V	0	
4	1988	D	0	
5	1988	V	0	
9	1988	J	0	RJ
14	1988	J	0	RJ
15	1988	J	0	RJ
16	1988	J	0	RJ
17	1988	J	0	RJ
18	1988	J	0	RJ
24	1988	S	0	RS
25	1988	S	0	RS
31	1988	V	0	
32	1988	S	0	RS
33	1988	J	0	RJ
35	1988	S	0	RS
36	1988	V	0	
37	1988	V	0	

The example data for this tutorial is available at: <http://www.myexperiment.org/files/850.html> or here

Bootstrap of observed census transitions (Confidence intervals of λ) workflow for local execution

below.

5.2.2 Input data

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

Terschelling matrices

- [Gentiana87-91.txt](#)

Please note that the workflow as well as the tutorial is a beta version, and may contain errors. We hope you will still find it interesting to work with this workflow, and look forward to your feedback.

5.3 Select input data dialogue boxes.

The first step is to fill out the input ports:

5.3.1 INPUTPORTS

1) **Bootstrap Iterations**: Number of iterations for calculation of bootstrap distributions.

To submit the number of iterations. First press add value, fill the number of iterations and press enter (Fig 2).

e.g.: 1000

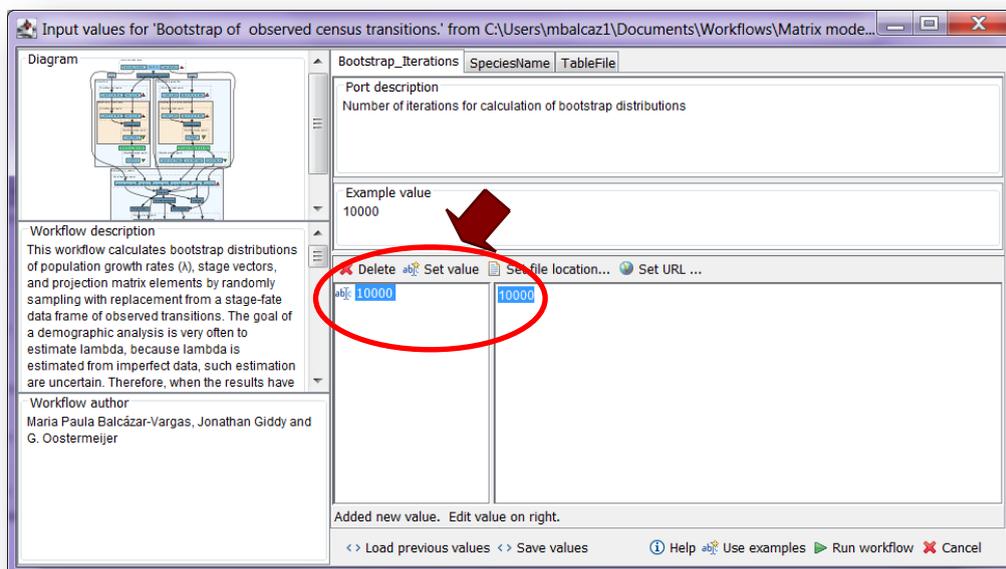


Figure 2. Set the number of iterations.

Bootstrap of observed census transitions (Confidence intervals of λ) workflow for local execution

2) **Species Name:** Descriptive title for labelling generated output (graph).

To submit the title of graph. First press add value, fill the desire title and press enter (Fig 3).

e.g.: *Gentiana pneumonanthe* (species taxonomic name)

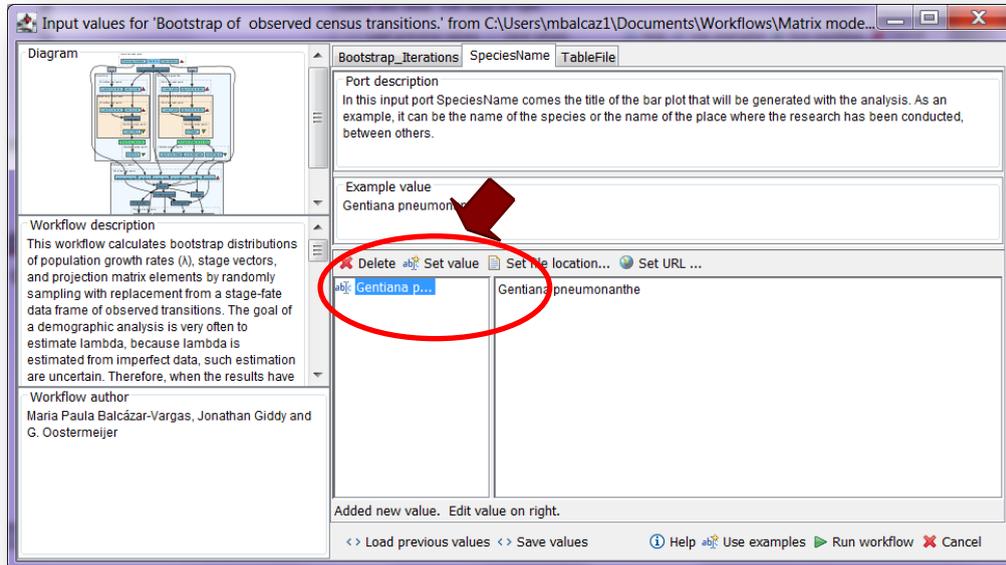


Figure 3. Set the graph title.

2) **TableFile:** The input data (a .txt-file) has to have the format of a table containing the demographic data on a series of individuals in two years. Each individual has a table row for each year and is identified by a number (plant column in example). For the year specified in the column 'year', each individual has a certain life stage ('stage'). The stage codes can be chosen by the user, but they have to be specified with a string. For all individuals, the number of offspring is specified in a chosen way (in the example, 'repstr' gives the number of flowers/fruits for each plant. In the example, it can be seen that only generative adults (stage=G) had flowers. Individuals without offspring have to be indicated by filling in '0' in this column, which cannot have empty cells. The column 'recruitment' specifies those individuals that are new in the second year census to the population by means of a code. In the example, 'RS' is a new seedling, and 'RJ' a new juvenile plant. As can be seen in the example, no code is needed in this column for individuals that were already present (See Table 1 in section 5.2.1 Data preparation/format).

To open the file. Click in Set file location (red arrow in Fig 4), a window dialog appears and the user selects the file (e.g. [Gentiana87-91.txt](#), Fig 5) and then clicks the Open button. Then the link to the file will appear on the Taverna window.

Bootstrap of observed census transitions (Confidence intervals of λ) workflow for local execution

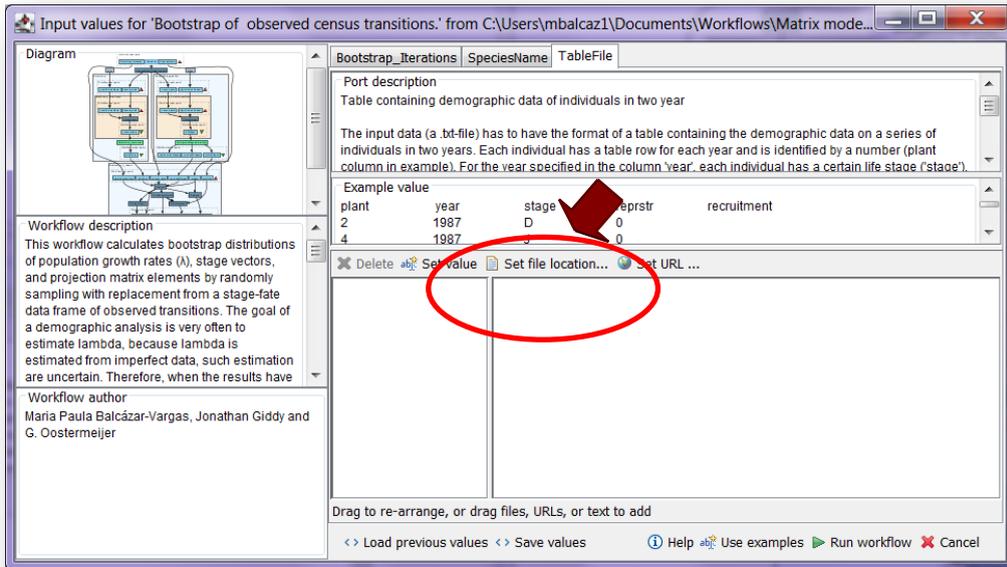


Figure 4. Set the Location of the matrix file.

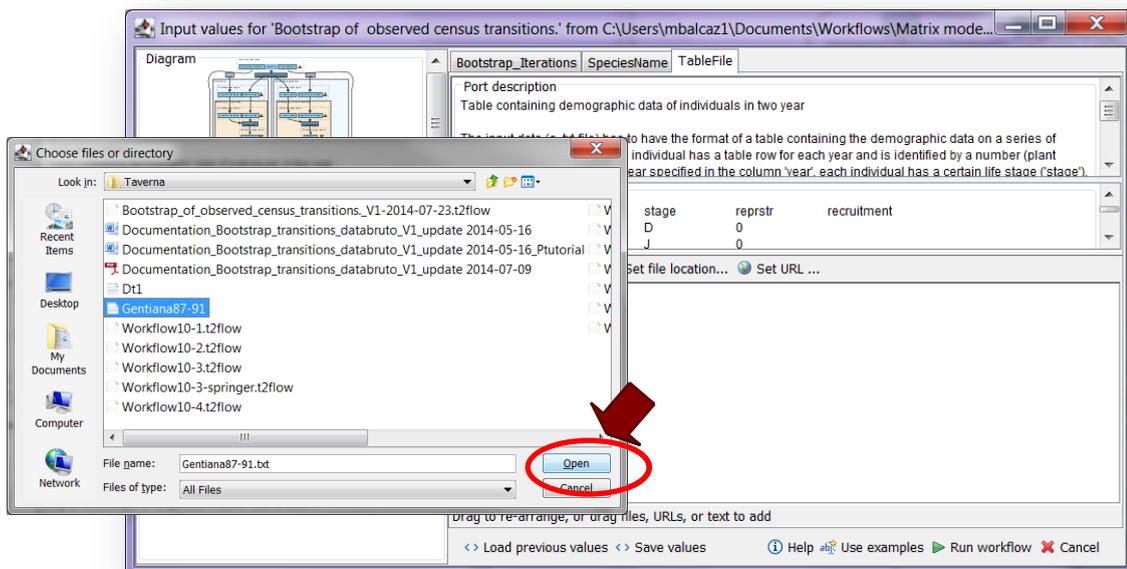


Figure 5. Choose the file and click in Open.

After the user has filled out the input ports and has clicked the **Run Workflow button**, a dialogue pages appear in the user's internet browser. See the next section:

Bootstrap of observed census transitions (Confidence intervals of λ) workflow for local execution

5.3.2 INTERACTION

1) **Stages:** With this dialogue, automatically appears the names of the stages or categories of the census data file. When the dialogue appears, the stages are in disorder (Fig 6), so the user drags and organizes the stages according to the order in the life cycle. Then, the author chooses if the stage belongs to the recruited, reproductive category or it should be excluded (Fig 7 and 8). Recruited means that new individuals can be recruited to this (these) stage(s). Reproductive stages are those that reproduce (produce offspring) (in this example the stage G). In the census data file [Gentiana87-91.txt](#), x is use to denote when a plant has died in the second year, so the user must selected in the excluded column. Then the user clicks in confirm and you will read stages submitted (Fig 7 and 8).

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 |

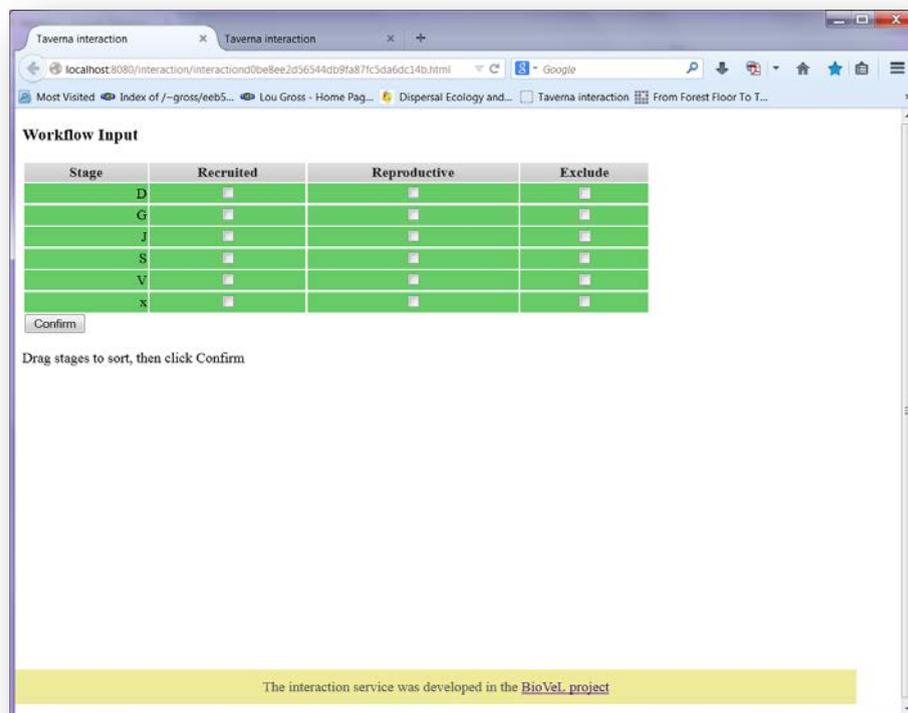


Figure 6. Stages dialogue. When the dialogue appears, the stages are in disorder.

Bootstrap of observed census transitions (Confidence intervals of λ) workflow for local execution



Figure 7. Stages dialogue. Drag each stage to the proper order. The stages should be organized as in the life cycle by the user (e.g. S, J, V, G, D, x).

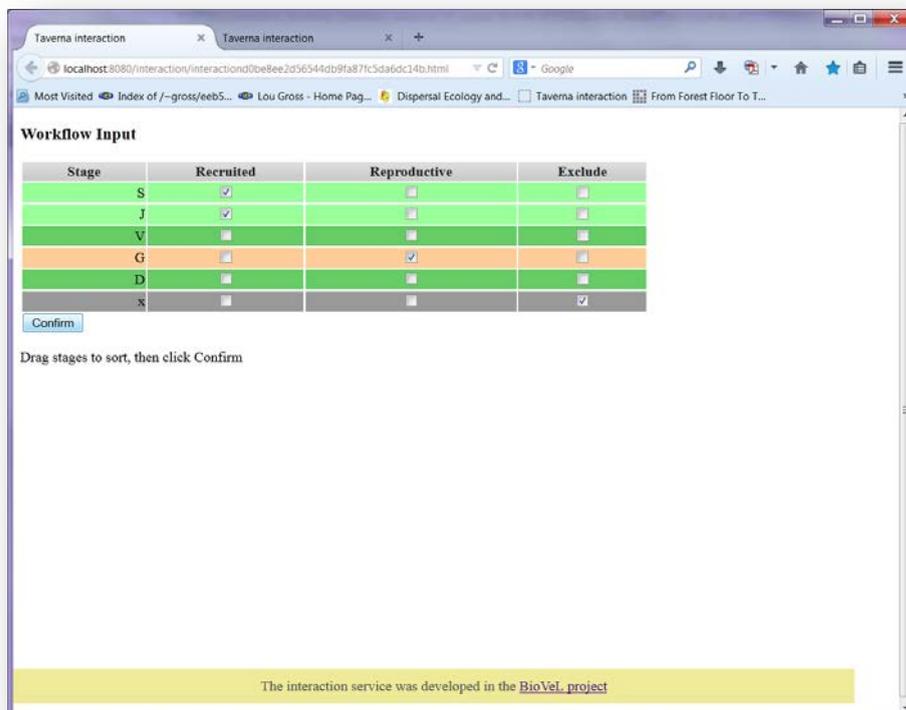


Figure 8. Stages dialogue. Choose if the stage belongs to the recruited or reproductive category, or should be excluded. Then click Confirm.

Bootstrap of observed census transitions (Confidence intervals of λ) workflow for local execution

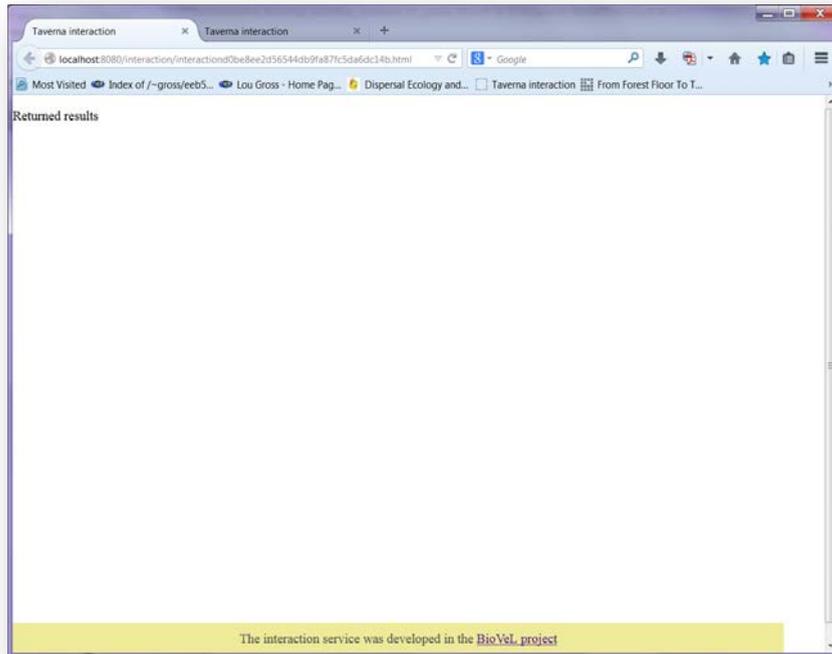


Figure 9. Stages dialogue. After click Confirm, the dialog Return results appears. Go to the second interaction page Year.

2) **Year:** In this dialogue, select the initial year of the census data file (Fig 10). In this example the year 1987 has been chosen (Fig 11). That means that the bootstrap analysis the interval 1987 -1988 of the data. After this step, the user should confirm the selection (Fig 11).

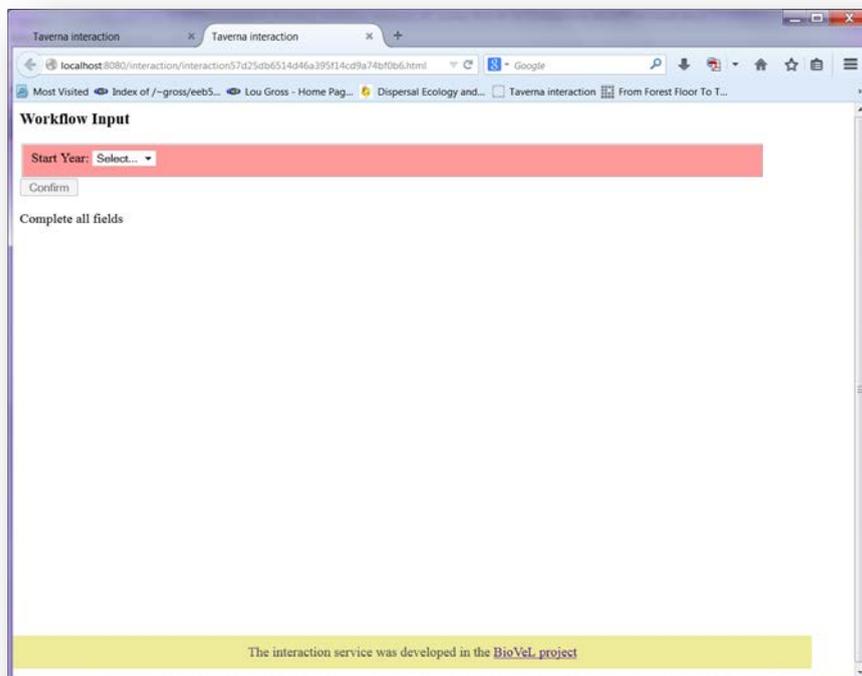


Figure 10. Years dialogue. When the dialogue appears, the user selects the initial year of the interval.

Bootstrap of observed census transitions (Confidence intervals of λ) workflow for local execution

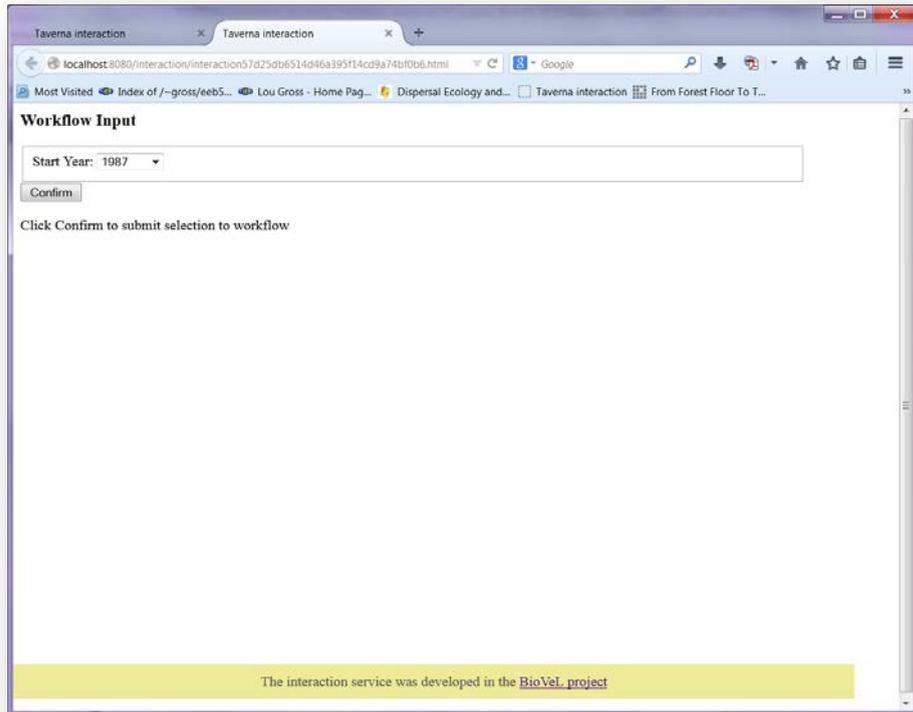


Figure 11. Year dialogue. The user selects the year interval, in this case 1987 and click Confirm.

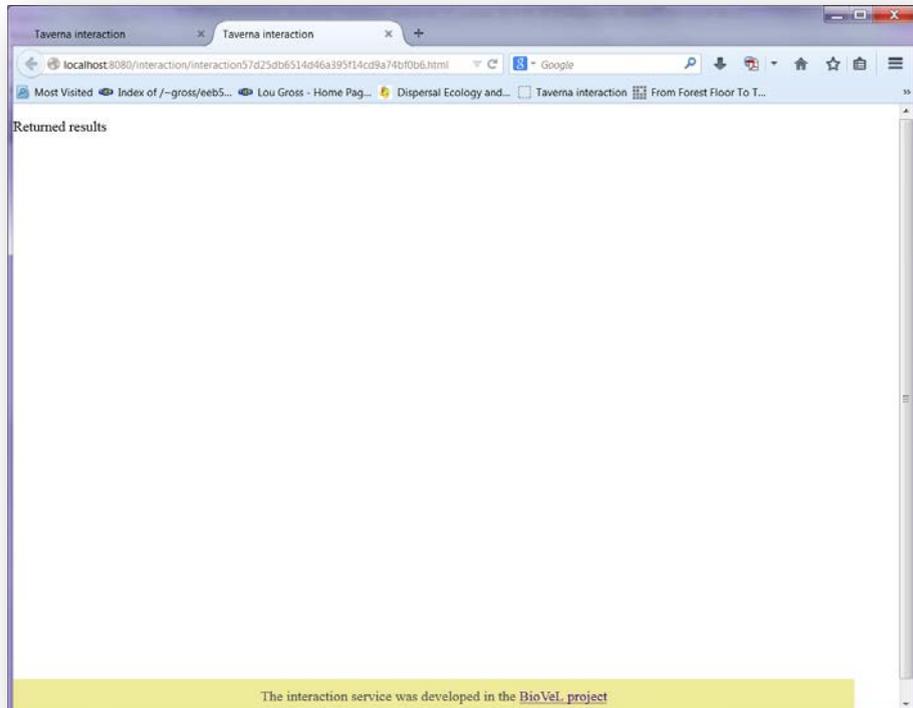


Figure 12. Year dialogue. After click Confirm, the dialog Return results appears. Return to Taverna workbench.

Bootstrap of observed census transitions (Confidence intervals of λ) workflow for local execution

After the user confirms the field, the workflow performs the analysis. To complete all the analysis may take several minutes.

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

Bootstrap of observed census transitions (Confidence intervals of λ) workflow for local execution

5.4 Save data/results

5.4.1 OUTPUTS

First, click in the selected result window e.g. **Confidence_Interval_CI** (Fig 13). Second, click in the left window on value 1 (blue oval), Third, on the right window click on the save value (green oval). Fourth, name the file and determine the extension file. For a text file: e.g. .csv or .txt. For images: .jpg or png (Fig 14, red oval), Finale, save the file in the chosen map.

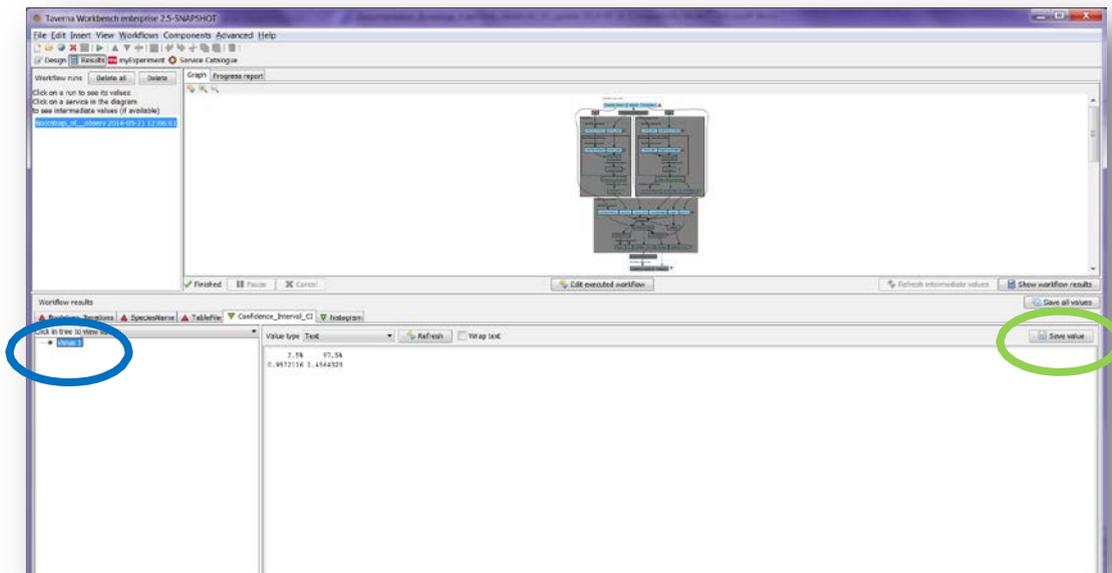


Figure 10. Confidence_Interval_CI output window. Value 1 (blue oval) on the left window.

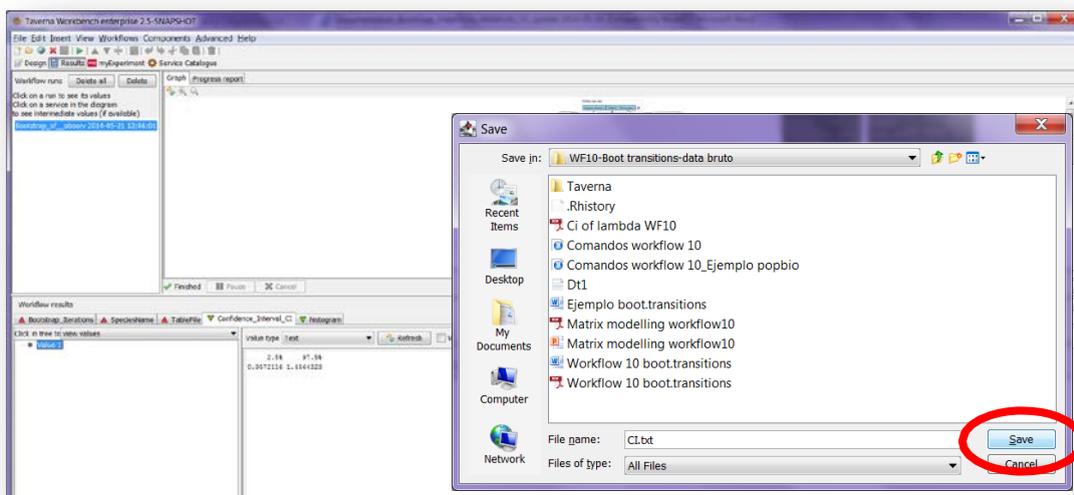


Figure 13. Name the file and determine the extension file. For a text file: e.g. .csv or .txt. For images: e.g.: .jpg or png.

Bootstrap of observed census transitions (Confidence intervals of λ) workflow for local execution

Repet the steps with the second result window **Histogram** (Fig 14).

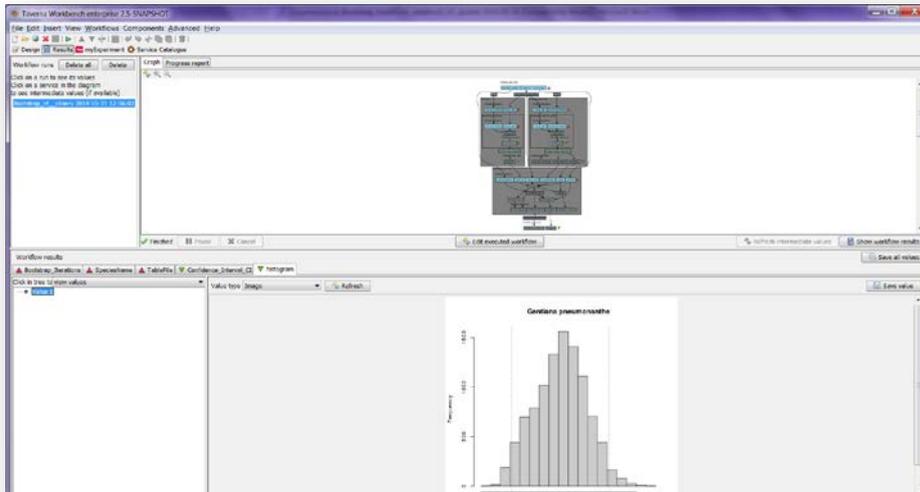


Figure 14. Histogram output window.

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%
0.9566	1.4648

2) **CI_histogram**: Histogram plotting the frequencies of the lambda values and the 95% confidence intervals resulting from the bootstrap analysis (Fig 26).

Bootstrap of observed census transitions (Confidence intervals of λ) workflow for local execution

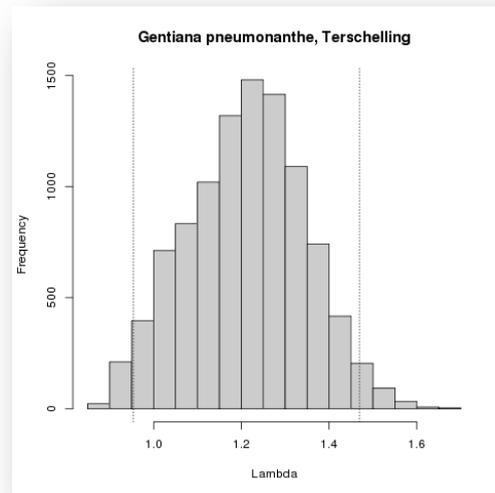


Figure 15. *The confidence interval histogram* opens as a graph.

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) and popdemo (Stott, Hodgson and Townley, 2013)

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.

Bootstrap of observed census transitions (Confidence intervals of λ) 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