BioVeL – Biodiversity Virtual e-Laboratory

Workflow Documentation

Calculate and Plot abundance over the years Workflow for local execution

Augustus 2014

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

Grant Agreement No: 283359
Project Co-ordinator: Mr Alex Hardisty
Project Homepage: http://www.biovel.eu
Duration of Project: 36 months

1 Description

This workflow calculates and plots the abundance per stage per year based on two or more years census data that contains transitions probabilities from each stage to the next, of tagged or mapped individuals of an animal or plant population.

2 General

2.1 Name of the workflow in myExperiment.

Name: Calculate and Plot the abundance over the years workflow

Download: The workflow pack can be downloaded at myExperiment under the following address: http://www.myexperiment.org/packs/664.html   or   only   the   workflow: http://www.myexperiment.org/workflows/4442.html

2.2 Date, version and licensing

Last updated: 28Th Augustus 2014

Version: 1

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 Gerard Oostermeijer unpublished results and publication: Oostermeijer, J.G.B. M.L. Brugman, E.R. de Boer; H.C.M. Den Nijs. 1996.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 Calculate and Plot the abundance over the years 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

Abundance, BioVeL, demography, *Gentiana pneumonanthe*, matrix population models, package '*popbio*' in r, survival curve.

3.2 Scientific workflow description
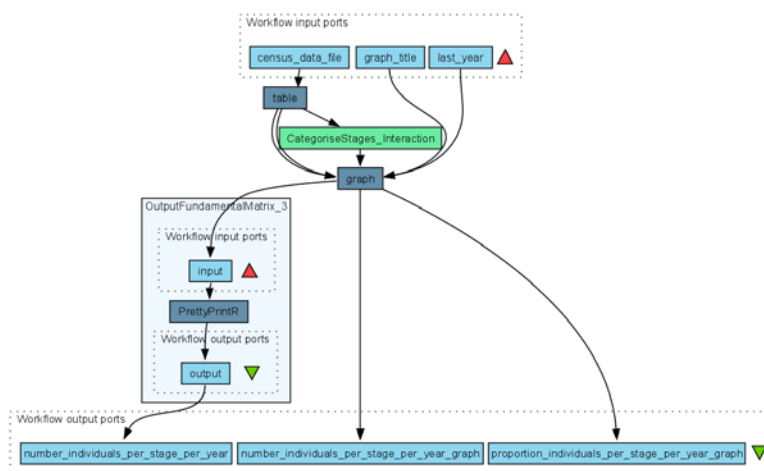
The aim of this workflow is to provide a connected environment to calculate and plot the abundance over the years. The workflow accepts input data in a .csv format. The output is provided as a set of R results.

Figure 1. A) Graph of the 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:
    - o install.packages("Rserve")
- Install package popbio
    - o install.packages("popbio")
- Local R Server: (Rserve) running at port 6311. See https://wiki.biovel.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

This tutorial requires a .txt file e.g. Data_Gentiana.csv. (Oostermeijer *et al.*, 1996).

5.2 Input data.

5.2.1 Data preparation/format

The workflow accepts input data (table) in a .csv format. The input data (a .csv-file) has to have the format of a table containing the demographic data on a series of individuals in two or more 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 and cannot be longer than 8 characters. When an individual died after a year or later the stage will be: dead (Table 1).

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

No	year	stage
2	1987	D
4	1987	J
5	1987	D
31	1987	V
36	1987	D
37	1987	J
41	1987	J
43	1987	S
46	1987	V
47	1987	S
54	1987	J
55	1987	V
57	1987	J
58	1987	V
59	1987	V
61	1987	V
62	1987	V
63	1987	V
66	1987	S
67	1987	V
68	1987	D
71	1987	G
72	1987	J

5.2.2 Input data

The table for this tutorial is available in http://www.myexperiment.org/files/1266.html or here below.

Terschelling matrices
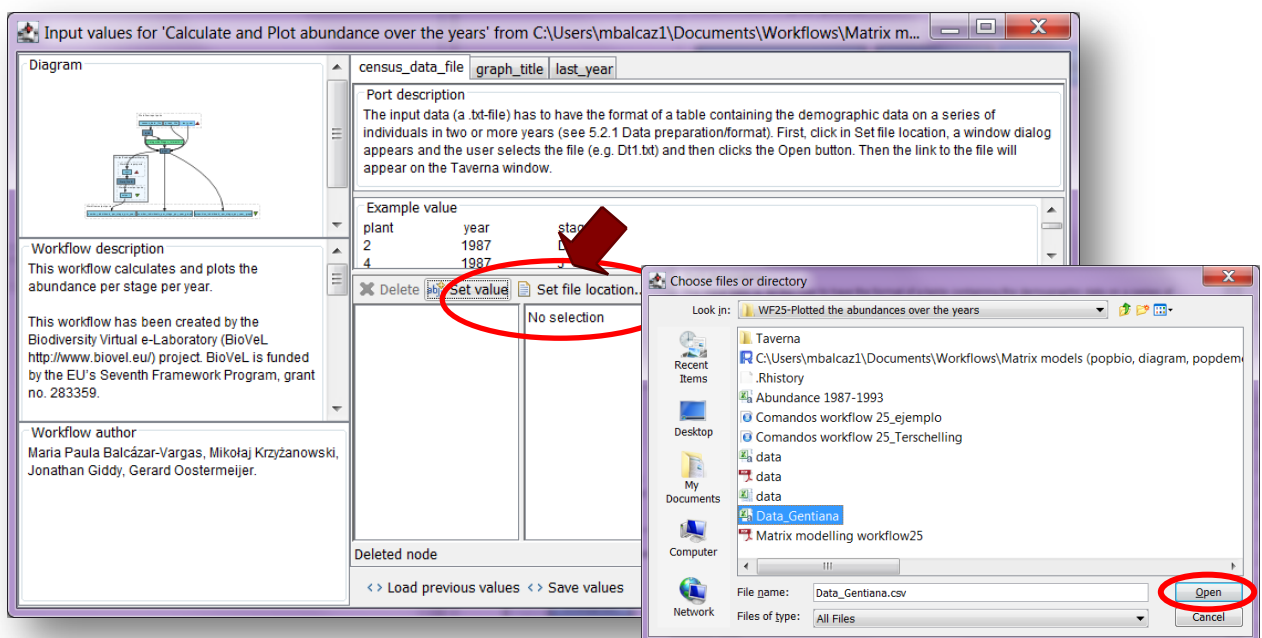- Data_Gentiana.csv

5.3 Select input data dialogue boxes.

The first step is to fill out the input ports:

5.3.1 INPUTPORTS

1) ***Census_data_file*:** The input data file must have the format of a table (a .csv-file) containing the demographic data on a series of individuals over several years. (see 5.2.1 Data preparation/format). First, click in Set file location (Fig 2), a window dialog appears and the user selects the file (e.g. Data_Gentiana.csv) and then clicks the Open button. Then the link to the file will appear on the Taverna window.

Figure 2. Input file to be filled in. Click in Set file location (red arrow), a window dialog appears and the user selects the file (e.g. Data_Gentiana.csv) and then clicks the Open button.

2) ***graph_title*:** The title of the plots. As example, it can be the name of the species or the name of the place where the research has been conducted. First press add value (Fig 3), fill the plot title and press enter.
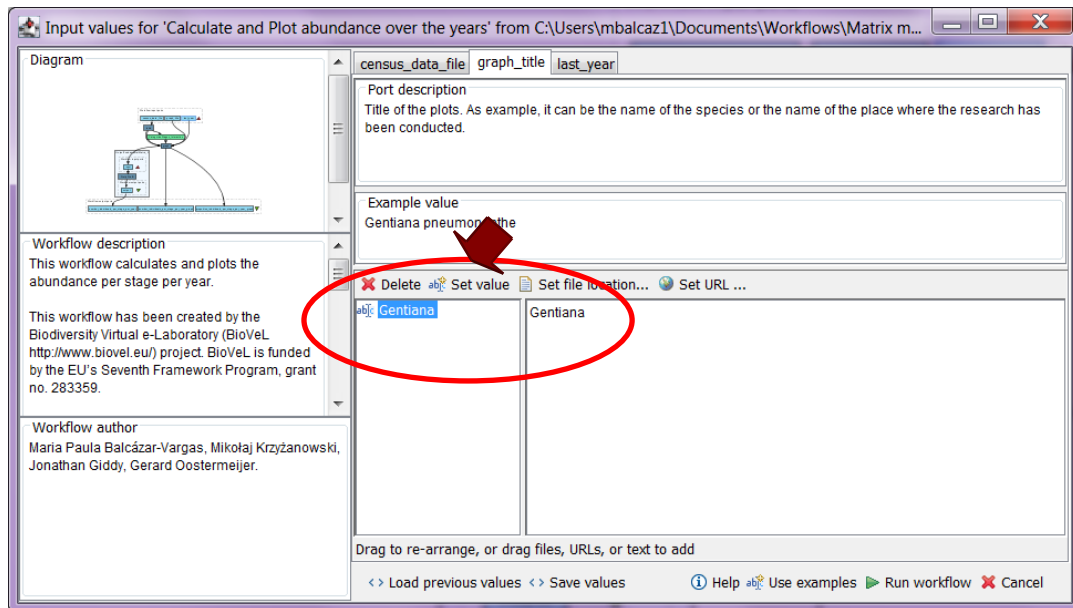
e.g. *Gentiana pneumonanthe*

Figure 3. Inputs years to be filled in (example).

2) **_last_year:_** this value is used to calculate and plot the abundances of the submitted data. For example, data from 1987 to 1993 (Data_Gentiana.csv). If the year 1993 is submitted, the abundances will be calculated and graphed from 1987-1993.
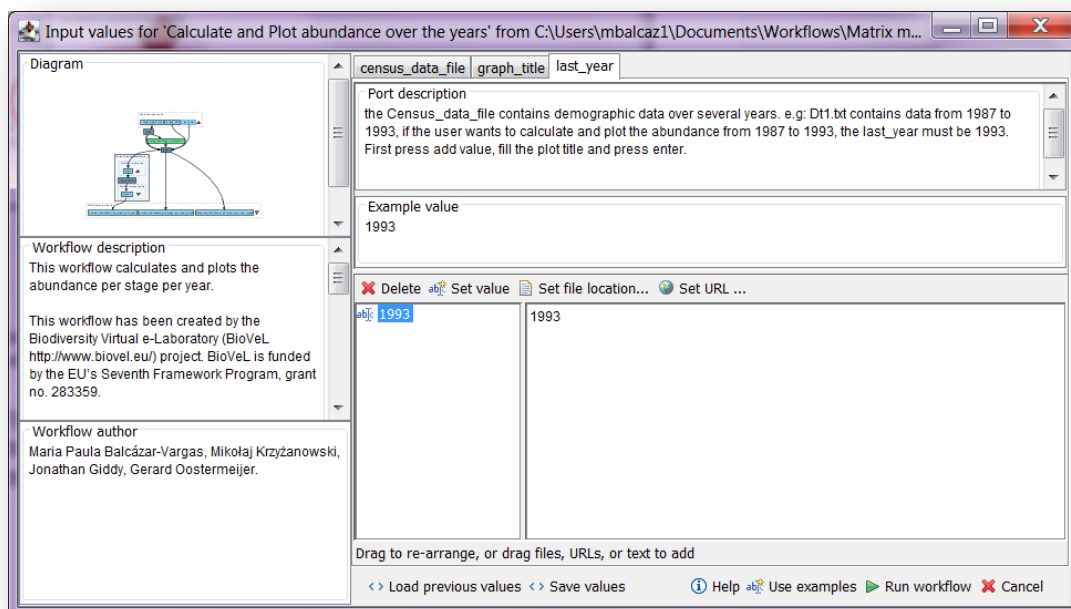
  e.g. 1993

Figure 4. Inputs last_year to be filled in (example).

After the user has filled out the input ports, click **_Run Workflow_** button, a dialogue pages appear in the user's internet browser.

5.3.2 INTERACTION

1) _**Stages**_: With this dialogue automatically appear the names of the stages or categories of the census data file. When the dialogue appears, the stages are in disorder (Fig 5), 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 6 and 7). 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 Data_Gentiana.txt, dead is use to denote when a plant has died in the second, third etc. year, the user can select dead in the excluded column or not. Then the user clicks in confirm and you will read stages submitted (Fig 8).

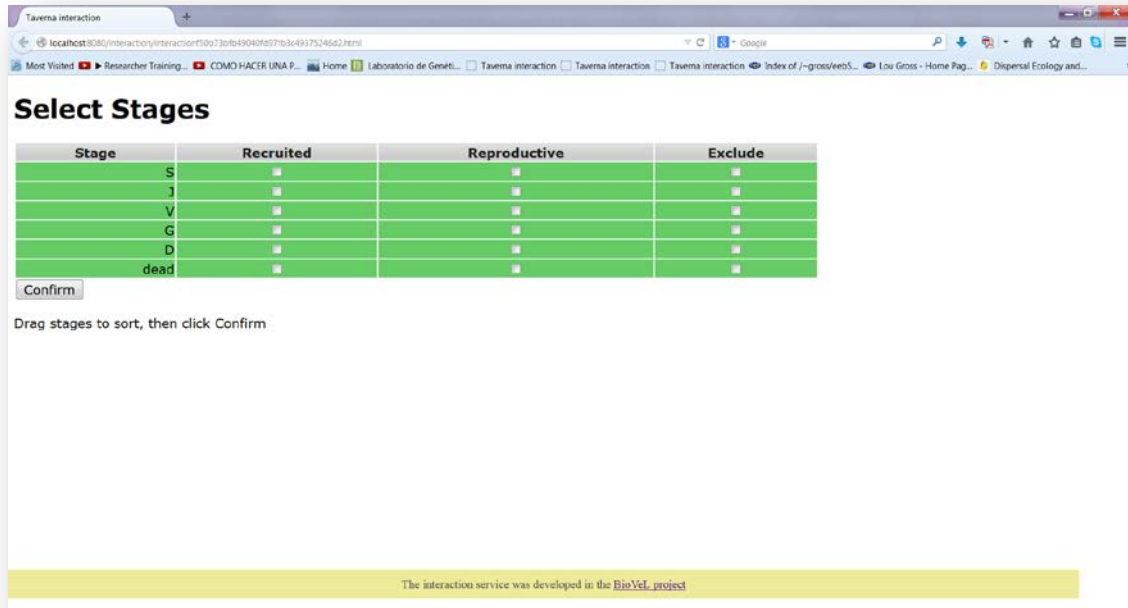In the following example, the life cycle of _Gentiana pneumonanthe_ has 5 stages or categories:
1) Seedlings	S
2) Juveniles	J
3) Vegetative	V
4) Reproductive individuals	G
5) Dormant plants	D
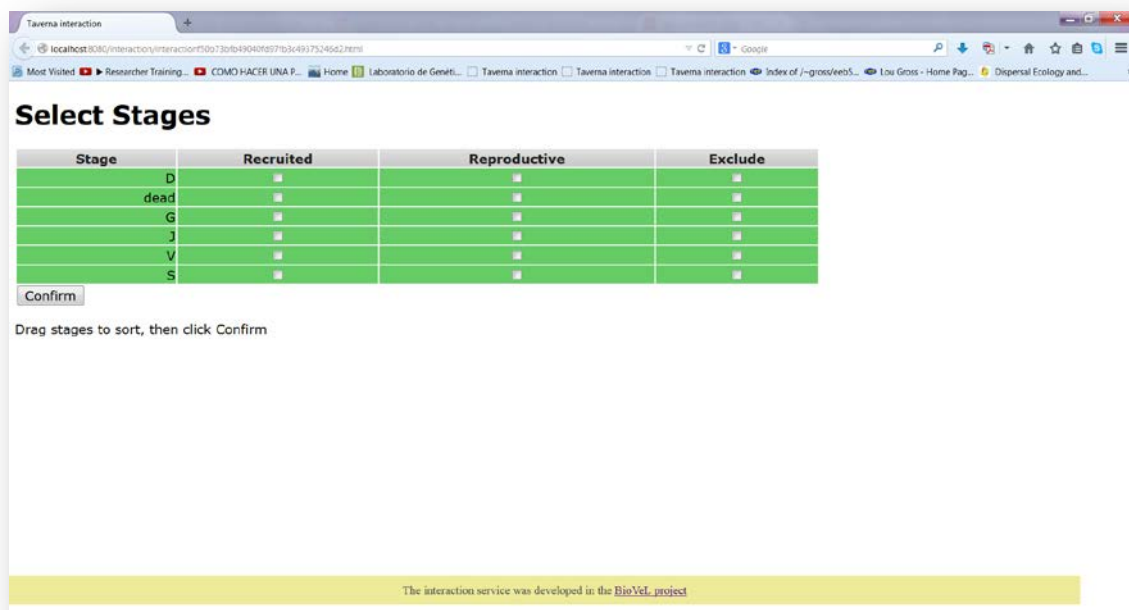
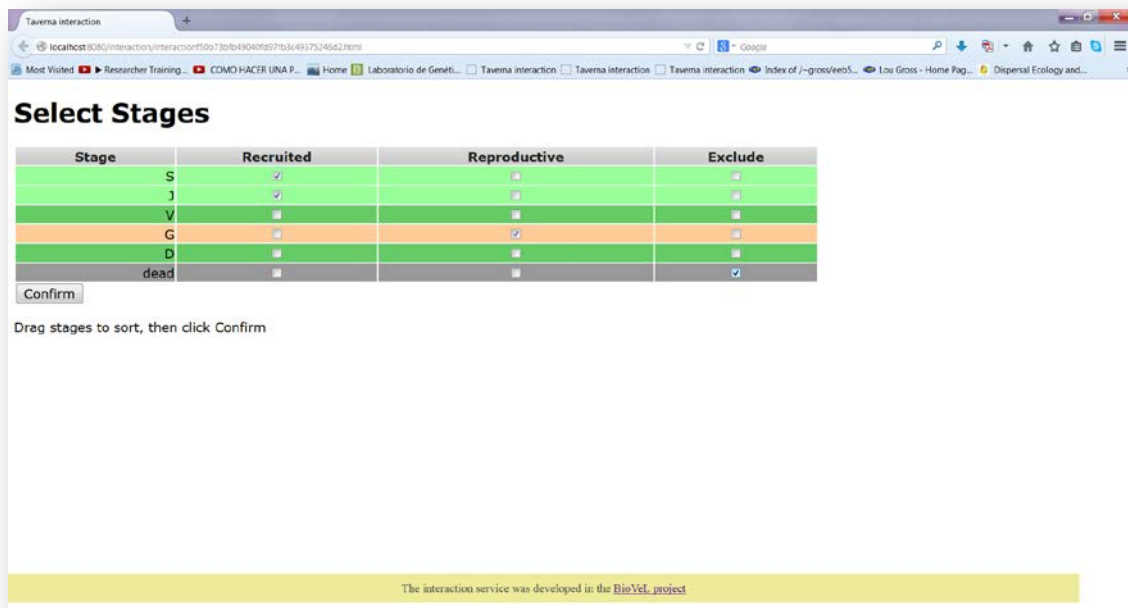**Figure 5**. Stages dialogue. When the dialogue appears, the stages are in disorder.
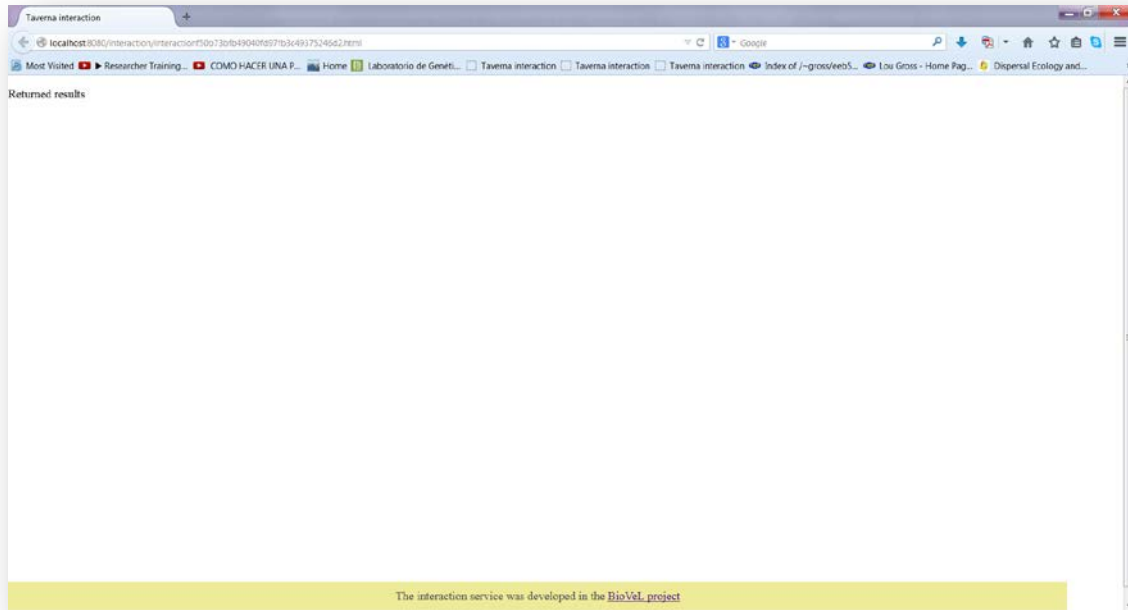
Figure 6. Stages dialogue. Drag each stage to the proper order. The stages should be organized as in the life cycle by the user (e.g. S, J, V, G, D, dead).

Figure 7. Stages dialogue. Choose if the stage belongs to the recruited or reproductive category, or should be excluded. In this example, dead is excluded.

Figure 8. Stage dialogue. After the user clicks Confirm, the confirmation "Returned Results" appears. Refer back to Taverna workbench.

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. *number_individuals_per_stage_per_year* (Fig 9, red oval). Second, click in the left window on value 1 (Fig 9, blue oval) and you will see the numerical results. Third, on the right window click on save value (Fig 9, green oval). Fourth, name the file and determinate the extension file, for a text files: e.g. .csv or .txt, for images: .jpg or .png. Finally, save the file in the chosen map (Fig 10).
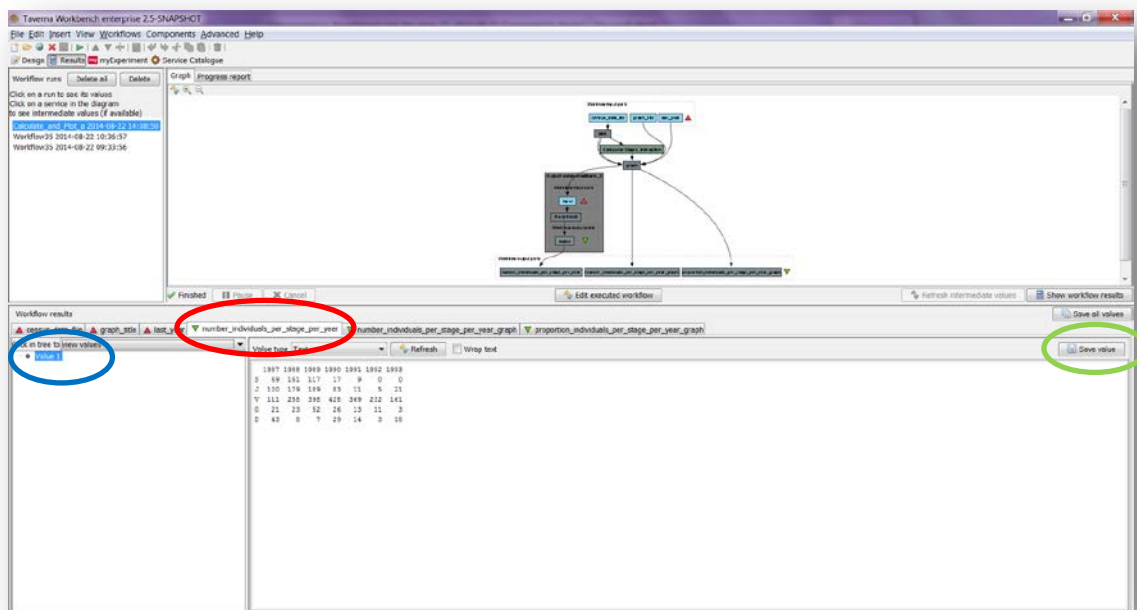
Figure 9. Taverna workbench results. When the analysis is completed, they appear on different windows under results in Taverna.

Calculate and Plot abundance over the years Workflow
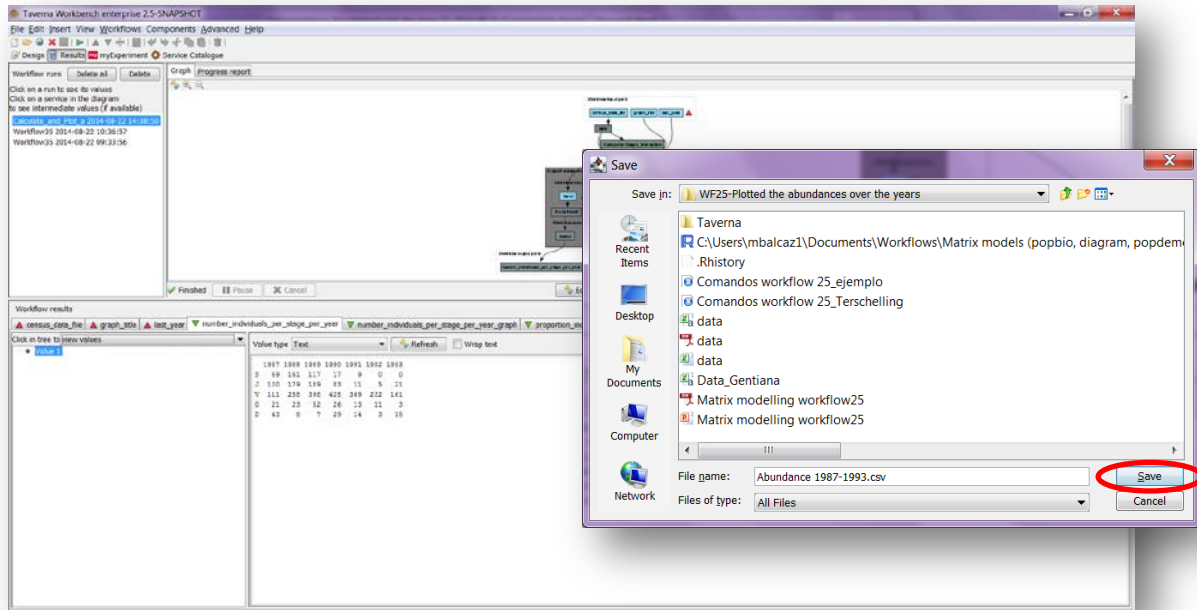
Figure 10. Name the file, click on Save value, name the file and add the extension for the file. For a text file: e.g. .csv or .txt.

1) number_individuals_per_stage_per_year

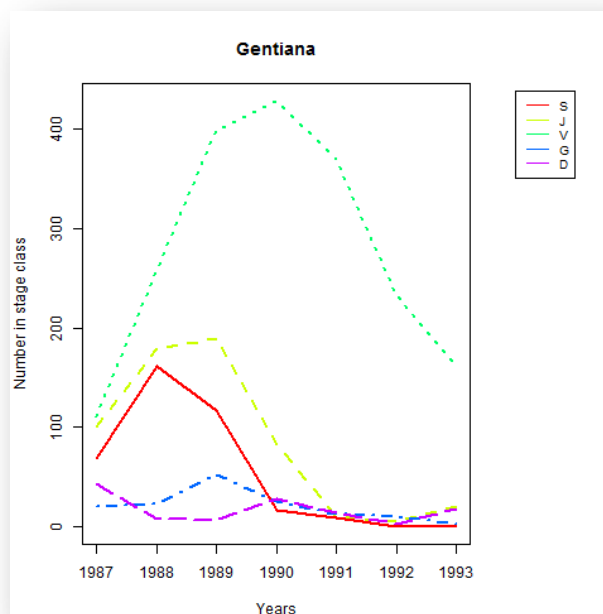The abundance per stage per year of Gentiana (1987-1993). The csv file after apply text to columns (Fig 11).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		1987	1988	1989	1990	1991	1992	1993							
2	S	69	161	117	17	9	0	0							
3	J	100	179	189	83	11	5	21							
4	V	111	258	398	428	369	232	161							
5	G	21	23	52	26	13	11	3							
6	D	43	8	7	29	14	3	18							

Figure 11. Abundance per stage per year, numeric results (.csv or text files).

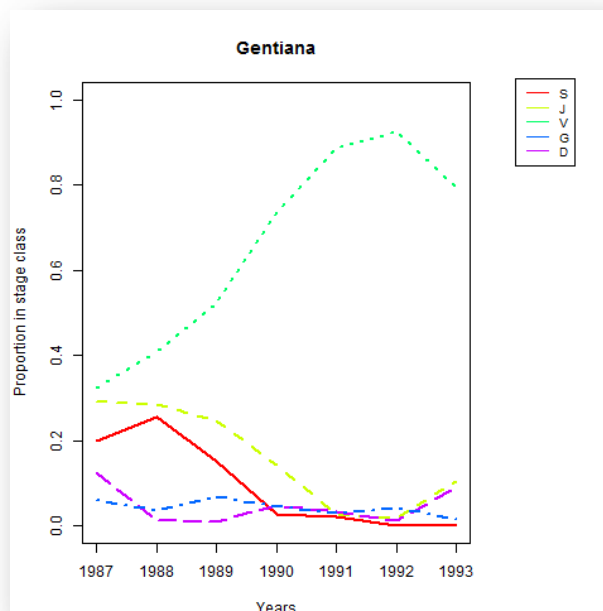*2) number_individuals_per_stage_per_year_graph*

The abundance per stage per year of Gentiana (1987-1993). The png file after (Fig 12).

Figure 12. Number of individuals per stage per year graph, Graph results (.png or jpg files).

3) proportion_individuals_per_stage_per_year_graph

The proportion of individuals per stage per year of Gentiana (1987-1993). The png file after (Fig 13).

Figure 13. Proportion of individuals per stage per year graph, Graph results (.png or jpg files).

6 Support

For questions with using the workflow, please write ***support @biovel.eu***

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

7 Bibliography

This workflow was created using and based on Package '*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.

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

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