

**BioVeL – Biodiversity Virtual e-Laboratory**

## **Workflow Documentation**

# **Creation of a stage matrix model 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



## Matrix population models construction Workflow for local execution

# 1 Description

The Creation of a stage matrix model Workflow provides an environment to construct a stage matrix model. This workflow create a stage matrix based on two years census data that contains transitions probabilities from each stage to the next, of tagged or mapped individuals of an animal or plant population. The resulting is a stage matrix model. This workflow can construct one o several matrices of different years in the same analysis.

## 2 General

### 2.1 Name of the workflow in myExperiment.

Name: Matrix Population Models construction Workflow for local execution

Download: The workflow pack can be downloaded at myExperiment under the following address:  
<http://www.myexperiment.org/packs/626.html> or only the workflow:  
<http://www.myexperiment.org/workflows/3276.html>

### 2.2 Date, version and licensing

Last updated: 23th July 2014

Version: 2

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 Gerard Oostermeijer unpublished results and publication: Oostermeijer, J.G.B. M.L. Brugman, E.R. de Boer; H.C.M. Den Nijs. 1996.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](http://www.biovel.eu)). Matrix Population Models construction workflow was run on *<date of the workflow run>*. BioVeL is funded by the EU's Seventh Framework Program, grant no. 283359.

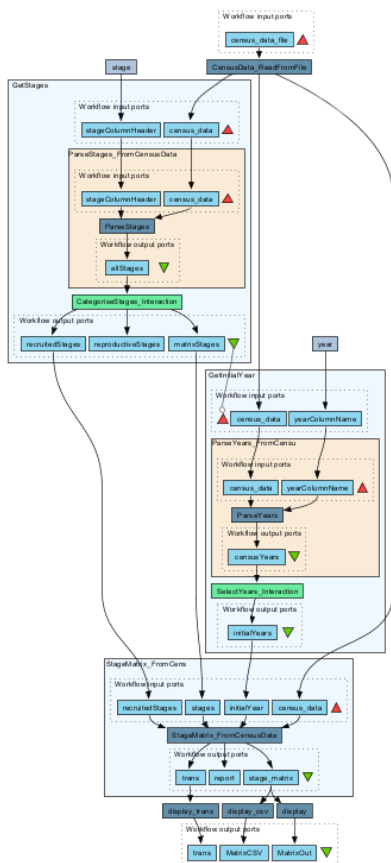
## 3. Scientific specifications

### 3.1 Keywords

BioVeL, Matrix Population Models, stage matrix with not density dependence.

### 3.2 Scientific workflow description

The aim of the workflow is to provide a connected environment for construct a stage matrix o several at the same time. The workflow accepts input data in a .txt format. The output is provided as a set of R results (Fig 1):



**Figure 1.** Stage matrix construction 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 may use input files in a .txt format: e.g. Dt1.txt. 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. When an individual died after a year the stage will x (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	

## Matrix population models construction Workflow for local execution

### 5.2.2 Input data

The example data [Dt1.txt](http://www.myexperiment.org/files/853.html) for the tutorial is available at: <http://www.myexperiment.org/files/853.html> or here below.

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

#### Terschelling matrices

- [Dt1.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.

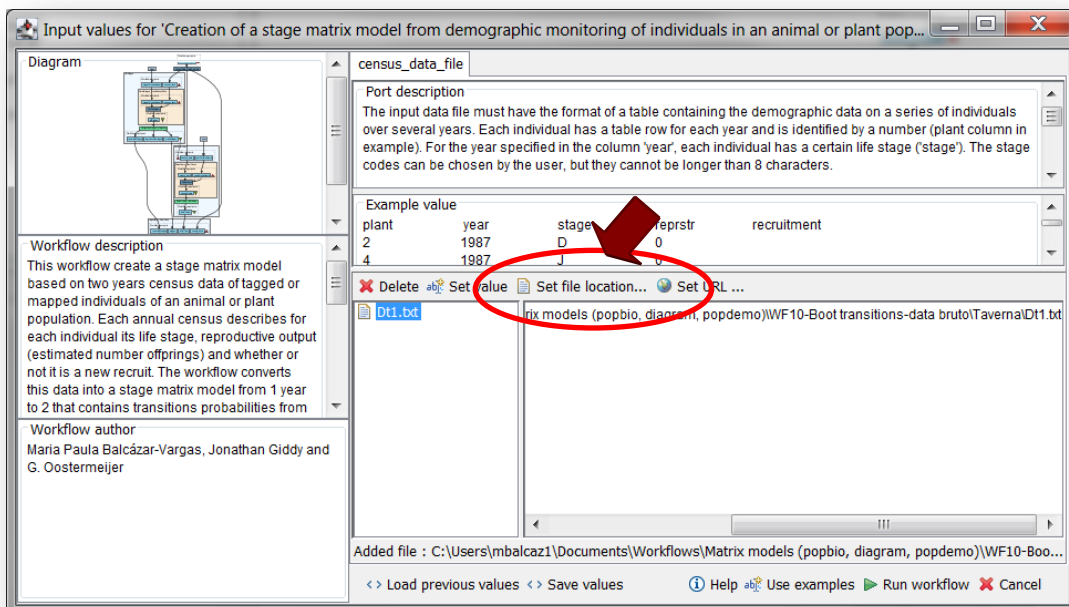
## 5.3 Select input data dialogue boxes.

The first step is to fill out the input ports:

### 5.3.1 INPUTPORTS

1) **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 (see 5.2.1 Data preparation/format)

Click in Set file location (Fig 2), a window dialog appears and the user selects the file (e.g. Dt1.txt) 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. Dt1.txt) and then clicks the Open button.

## Matrix population models construction Workflow for local execution

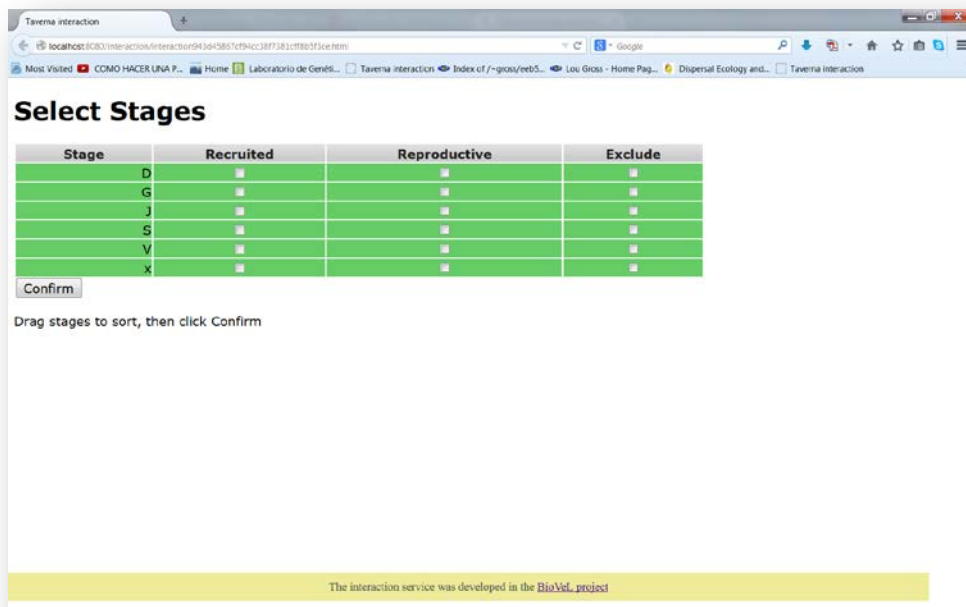
After the user has filled out the input ports and has clicked the **Run Workflow button**, a dialogue pages appear in the user's internet browser. See the next section:

### 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 3), 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 4 and 5). 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 Dt1.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 3, 4 and 5).

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

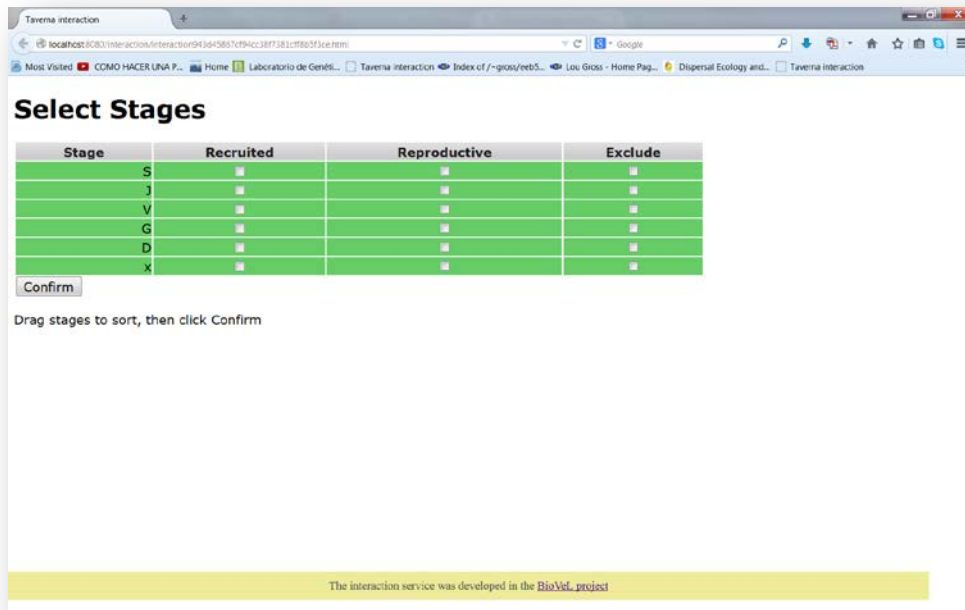
- |                             |   |
|-----------------------------|---|
| 1) Seedlings                | S |
| 2) Juveniles                | J |
| 3) Vegetative               | V |
| 4) Reproductive individuals | G |
| 5) Dormant plants           | D |



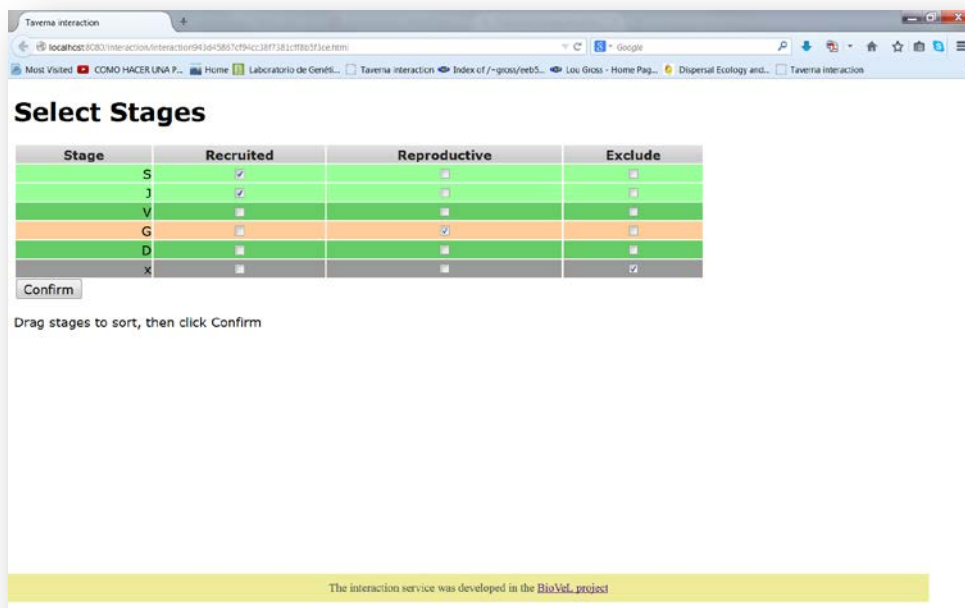
**Figure 3.** Stages dialogue. When the dialogue appears, the stages are in disorder.



## Matrix population models construction Workflow for local execution



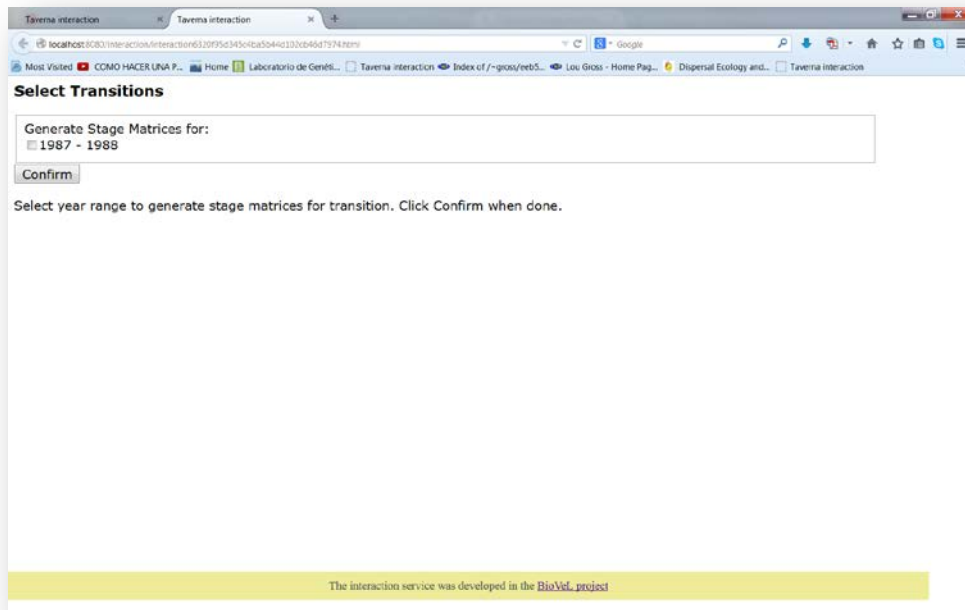
**Figure 4.** Stages dialogue. Drag each stage to the proper order. The stages should be organized as in the life cycle by the user (e.g. S, J, V, G, D, x).



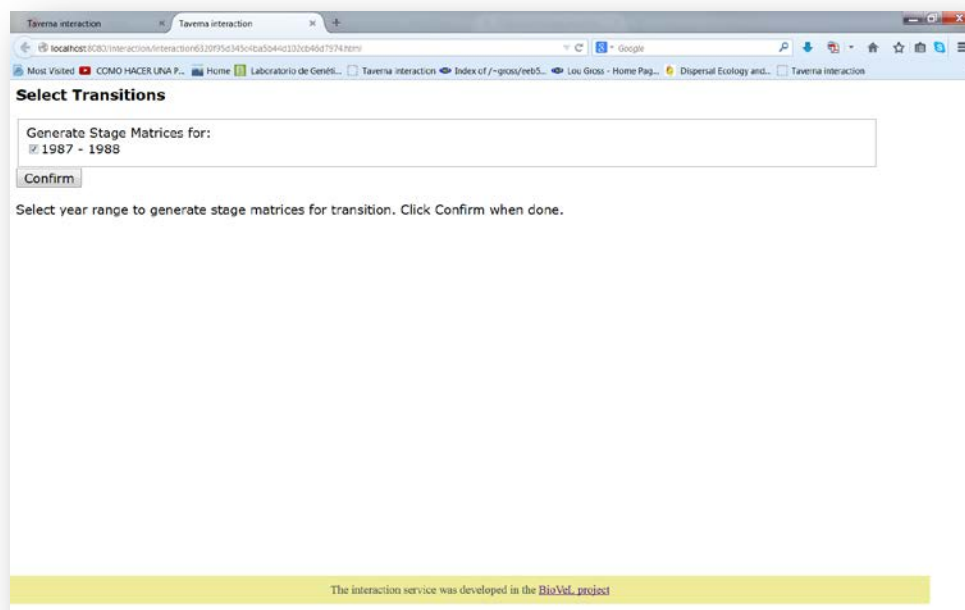
**Figure 5.** Stages dialogue. Choose if the stage belongs to the recruited or reproductive category, or should be excluded.

2) **Year:** In this dialogue view the years of the census data file (Fig 6, example file from 1987 to 1988). Select one or multiple census years. In this example the years 1987 - 1988 has been chosen (Fig 7 and 8). That means that the analysis will create one matrices, for the interval 1987 -1988. After this step, the user should confirm the selection (Fig 8).

## Matrix population models construction Workflow for local execution



**Figure 6.** Years dialogue. When the dialogue appears, the user selects the year interval of one or more year matrices.



**Figure 7.** Year dialogue. The user selects the year interval, in this case 1987-1988, for a year matrix 1987-1988.

## 5.4 Save data/results

### 5.4.1 OUTPUTS

First, click in the selected result window e.g. **MatrixCSV** (Fig 6, red oval). Second, click in the left window on value 1 (Fig 6, blue oval, because we calculated one matrix, there is only one value. When the user calculates e.g. 6 matrices in the same analysis, it will appear value 1 to value 6 below each other). Third, on the right window click on the save value (Fig 6, green oval). Fourth, name the file and determinate the extension file. For a text file: e.g. .csv or .txt., Finale, save the file in the chosen map (Fig 7). Save the results of the **MatrixCSV** as .csv and the results of the **MatrixOut** as .txt.

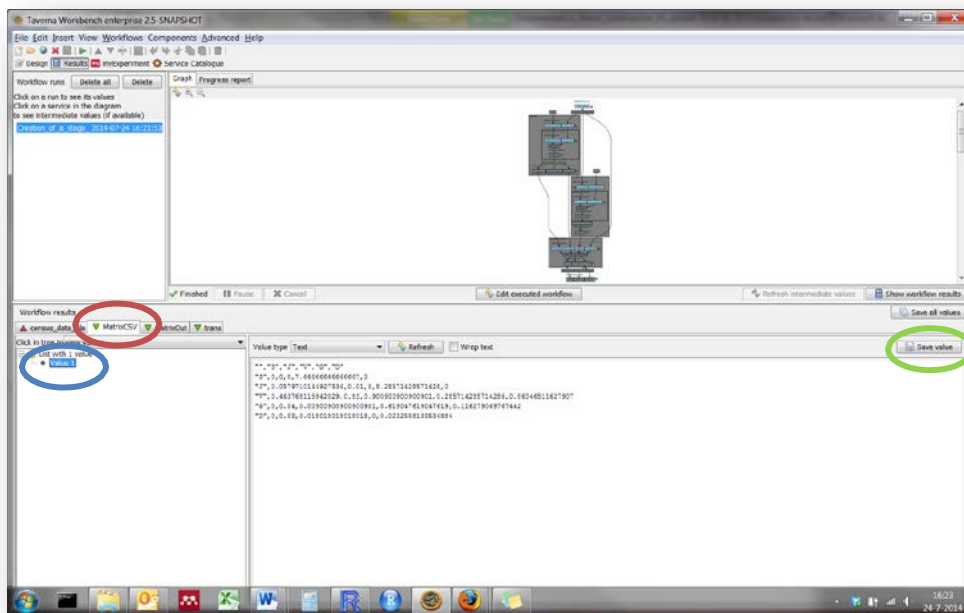


Figure 6. Stage-matrix output window.

## Matrix population models construction Workflow for local execution

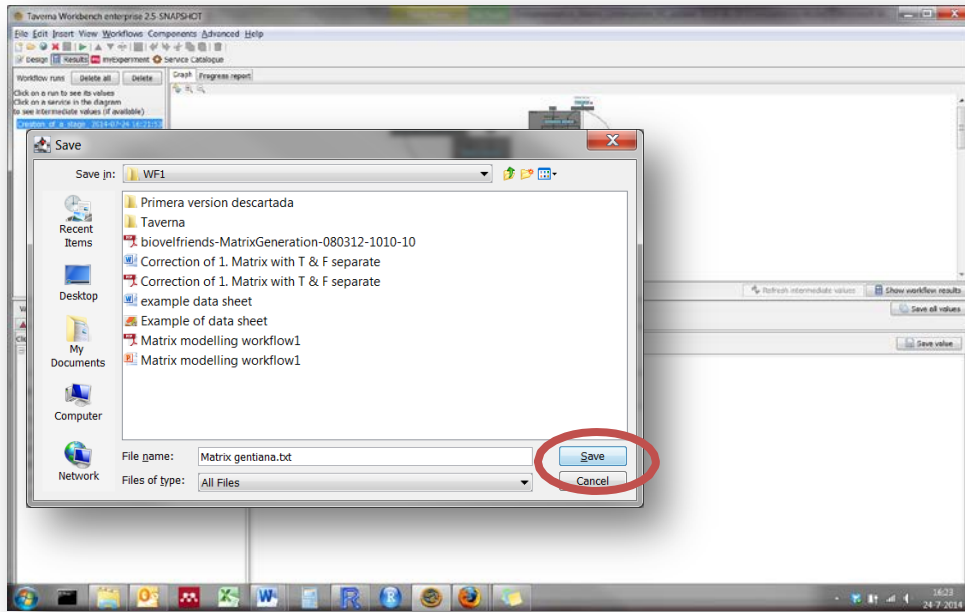
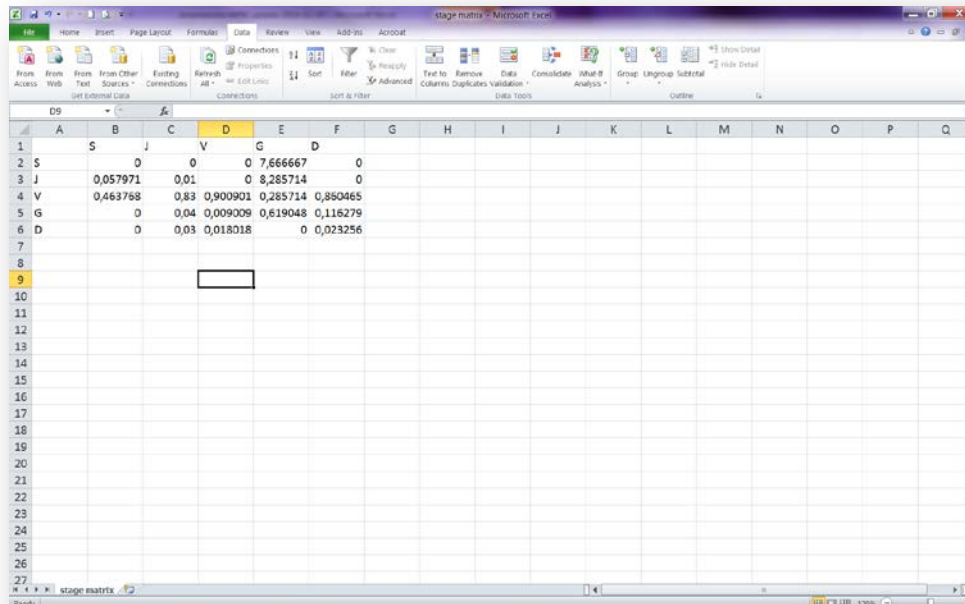


Figure 7. Name the file and determine the extension file. For a text file: e.g. .csv or .txt.

### Construction matrix

1) **MatrixCSV and MatrixOut**: are stage matrices containing transitions probabilities from each stage to the next. In this example the selected species is *Gentiana pneumonanthe* has a matrix of 5 x 5 stages (Fig. 8; Oostermeijer et al., 1996).



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	S	J	V	G	D												
2	S		0	0	0	7,666667	0										
3	J		0,057971	0,01	0	8,285714	0										
4	V		0,463768	0,83	0,900901	0,285714	0,850465										
5	G		0	0,04	0,009009	0,619048	0,115279										
6	D		0	0,03	0,018018	0	0,023256										
7																	
8																	
9																	
10																	
11																	
12																	
13																	
14																	
15																	
16																	
17																	
18																	
19																	
20																	
21																	
22																	
23																	
24																	
25																	
26																	
27																	

Figure 14. MatrixCSV output file open in excel.

## 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, 2<sup>nd</sup> 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.

## **Matrix population models construction 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**