

**BioVeL – Biodiversity Virtual e-Laboratory**

## **Workflow Documentation**

# **Simulate stochastic growth from a sequence of matrices 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



## Simulate stochastic growth from a sequence of matrices Workflow

## 1 Description

The Simulate stochastic growth from a sequence of matrices Workflow provides an environment to simulate stochastic growth by projection using whole matrix selection techniques in an independently and identically distributed (iid) environment from a set of 2 or more projection matrices.

This workflow is based on the popbio package (script: stoch.projection - Simulate stochastic growth from a sequence of matrices, Stubben, Milligan and Nantel, 2013) based on the The MATLAB code in Box 7.3 (Morris and Doak 2002). For more details of the analysis see: Exploring population growth in a variable environment with computer simulations (Morris and Doak 2002, page: 229-233 and Caswell 2001, Chapter 14).

Outputs:

- Equal selection data
- Stochastic projection graph
- Unequal selection data

We use *Gentiana pneumonanthe* collected in Terschelling, The Netherlands as example how to run this workflow, besides that this data was not taken in a stochastic environment.

## 2 General

### 2.1 Name of the workflow in myExperiment.

Name: Simulate stochastic growth from a sequence of matrices.

Download: The workflow pack can be downloaded at myExperiment under the following address:  
<http://www.myexperiment.org/packs/645.html> or only the workflow:  
<http://www.myexperiment.org/workflows/4349.html>

### 2.2 Date, version and licensing

Last updated: 30<sup>th</sup> July 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:

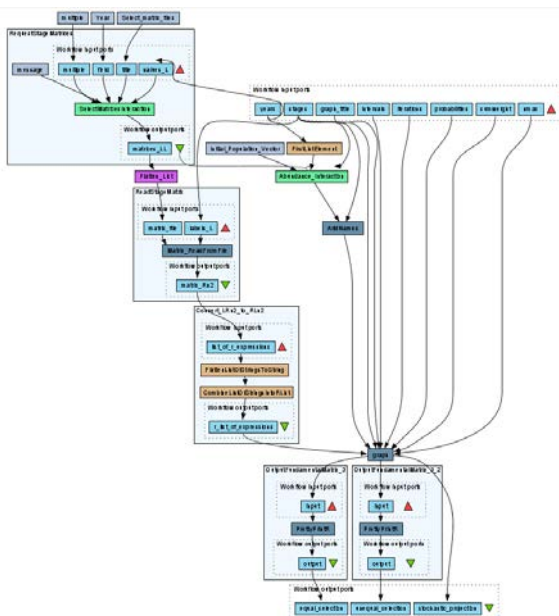
The input data and results reported in this publication (tutorial) come from data (Dr. Gerard Oostermeijer unpublished results and publication: Oostermeijer, J.G.B. M.L. Brugman, E.R. de Boer; H.C.M. Den Nijs. 1996. Temporal and Spatial Variation in the Demography of *Gentiana pneumonanthe*, a Rare Perennial Herb. *The Journal of Ecology*, 84: 153-166.) using BioVeL workflows and services ([www.biovel.eu](http://www.biovel.eu)). Simulate stochastic growth from a sequence of matrices 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, demography, *Gentiana pneumonanthe*, independently and identically distributed (iid), matrix population models, matrix, model, package 'popbio' in r, population, stage matrix, stochastic environment, stochastic growth, stochastic projection, Morris and Doak 2002.

### 3.2 Scientific workflow description



The aim of the Simulate stochastic growth from a sequence of matrices workflow provides an environment to simulate stochastic growth by projection using whole matrix selection techniques in an independently and identically distributed (iid) environment from a set of 2 or more projection matrices. The workflow accepts input data (matrices) in a .txt format (decimal numbers indicated by dots e.g.: 0.578). The output is provided as a set of R results and graphic plots.

**Figure 1. A)** Graph of the Simulate stochastic growth from a sequence of matrices 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

This tutorial requires 6 input files or matrices of 6 consecutive years (from 87-88 to 92-93) of the same place (Terschelling, The Netherlands, Oostermeijer *et al.*, 1996).

### 5.2 Input data.

#### 5.2.1 Data preparation/format

The workflow accepts input data (matrices) in a .txt format, all decimal numbers in each matrix must be indicated by dots e.g.: 0.578. All examples matrices for the tutorial are available in the PACK: <http://www.myexperiment.org/packs/645.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 matrices

- [MTers87\\_88.txt](#)
- [MTers88\\_89.txt](#)
- [MTers89\\_90.txt](#)
- [MTers90\\_91.txt](#)
- [MTers91\\_92.txt](#)
- [MTers92\\_93.txt](#)

### 5.3 Select input data dialogue boxes.

The first step is to fill out the input ports:

#### 5.3.1 INPUTPORTS

1) **Graph title:** Is the main title to be display in the stochastic projection graph. Clicks in add value and then write the text in the right space. Please do not use a title longer than the suggest it, 11 word or 83 characters with spaces (Fig. 2)

*e.g.: Projection of stochastic growth for Gentiana using equal and unequal probabilities.*

## Simulate stochastic growth from a sequence of matrices Workflow

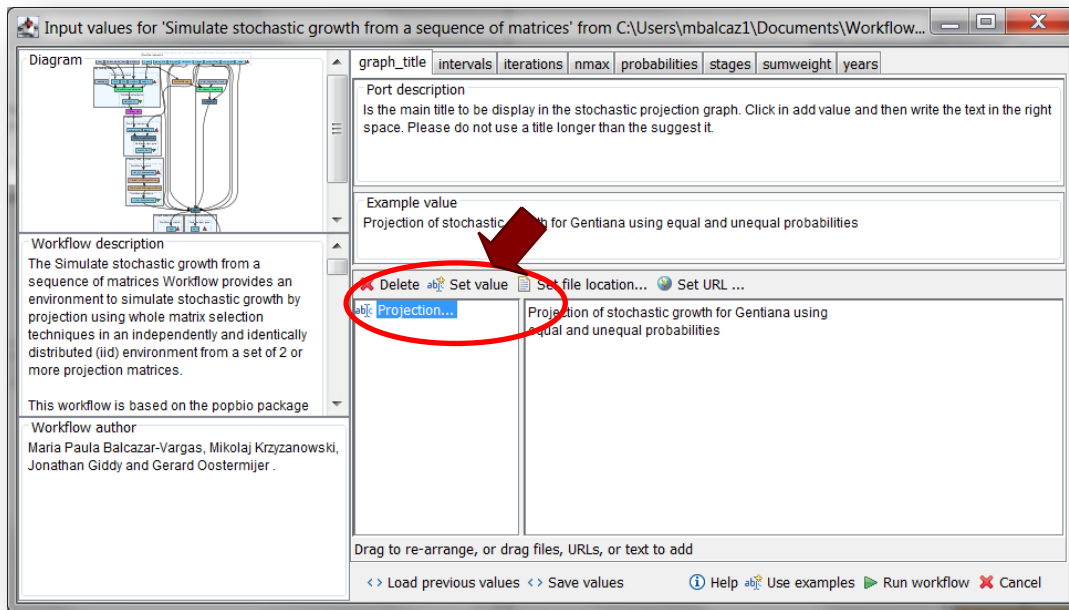


Figure 2. Stochastic projection graph title to be filled in (example).

2) **Intervals**: This value is the number of time steps or projection intervals to predict future population size or in other words the time to predict future population size. Click in add value and write the number in the right space (Fig. 3).

e.g.: 25

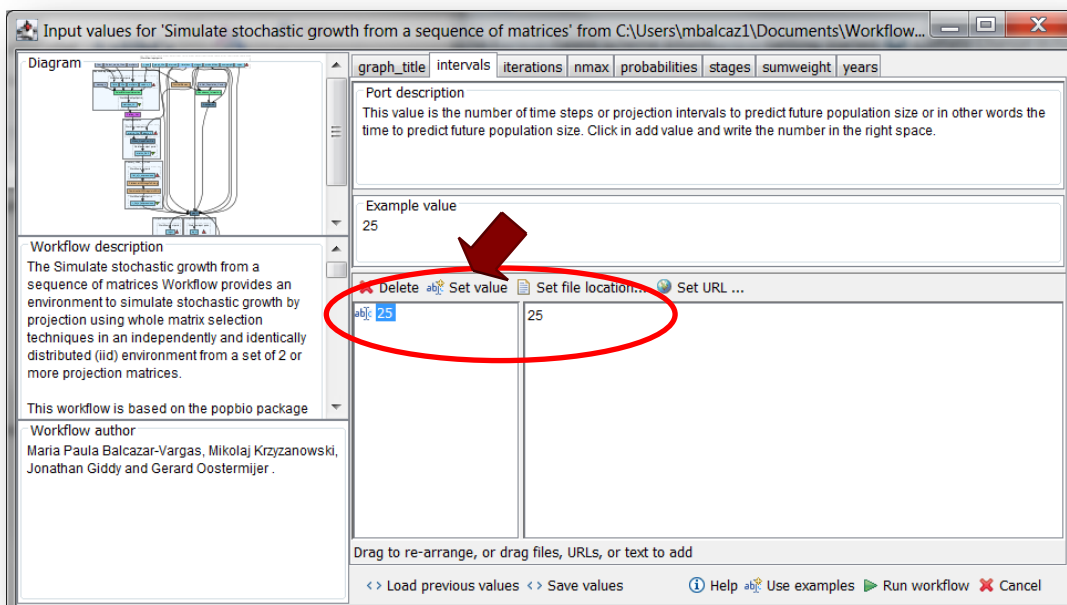


Figure 3. The number of intervals to be filled in.

## Simulate stochastic growth from a sequence of matrices Workflow

3) **Iterations:** It is the number of iterations or the number of realizations of population growth to simulate. Click in add value and write the number in the right space (Fig. 4).

e.g.: 1000

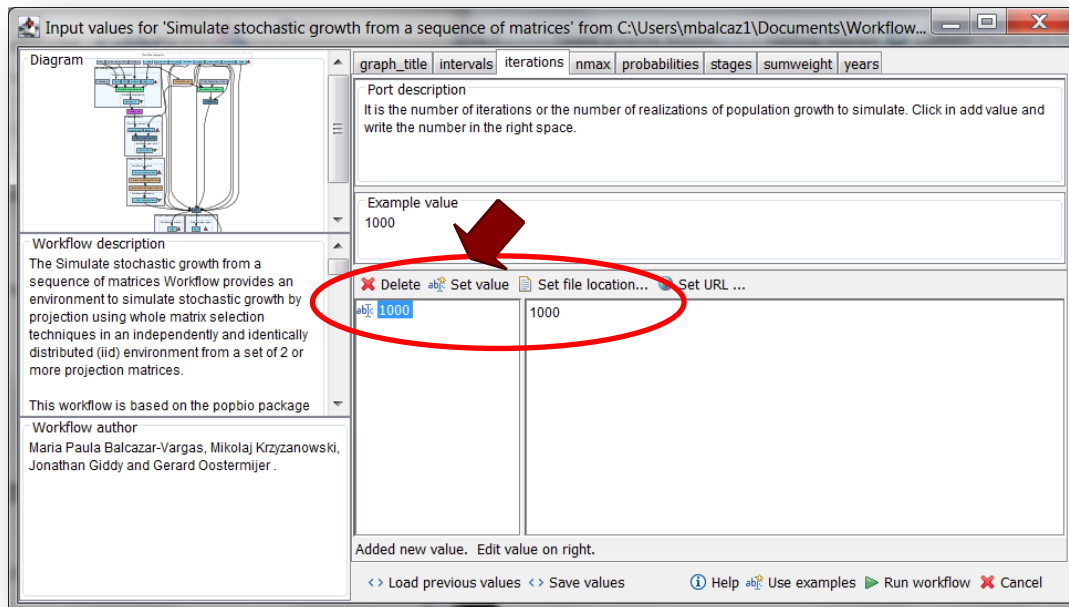


Figure 4. Number of iterations to be filled in.

4) **nmax:** it is a maximum number of individuals beyond which population projections cannot exceed.

### No density dependence:

To use *no density dependence*: first, click in add value, then in the left window a List is added with a [] element in the right window. Delete the []. After deletion, the "List" has no elements in the right window. When running the workflow, Rshell script will assign NULL to the input. (Fig 5 - 7).

This tutorial uses No density dependence (see results).



## Simulate stochastic growth from a sequence of matrices Workflow

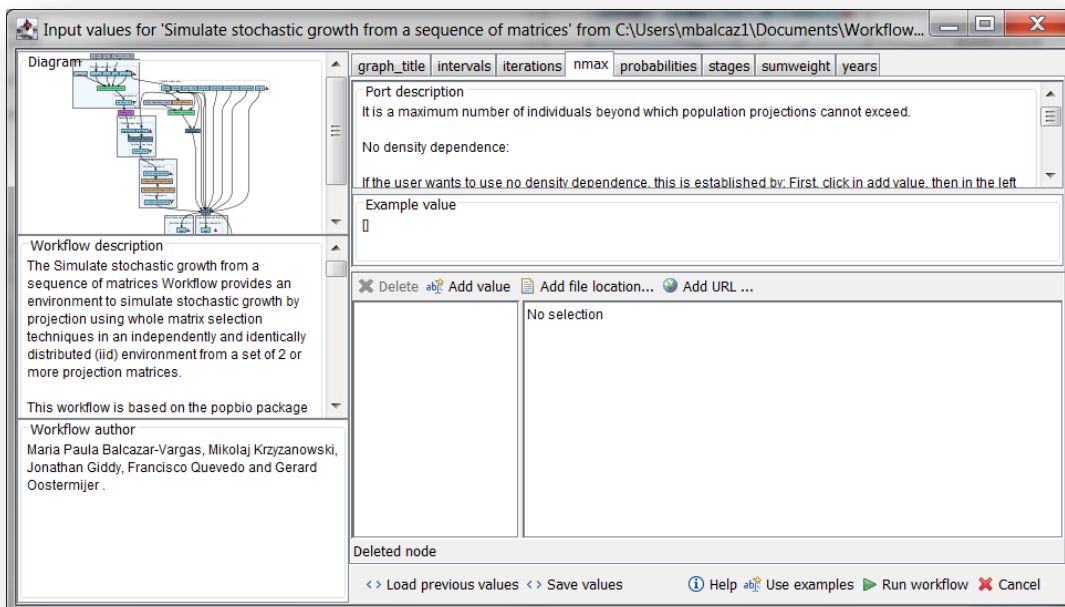


Figure 5. No density dependence: First, click in add value.

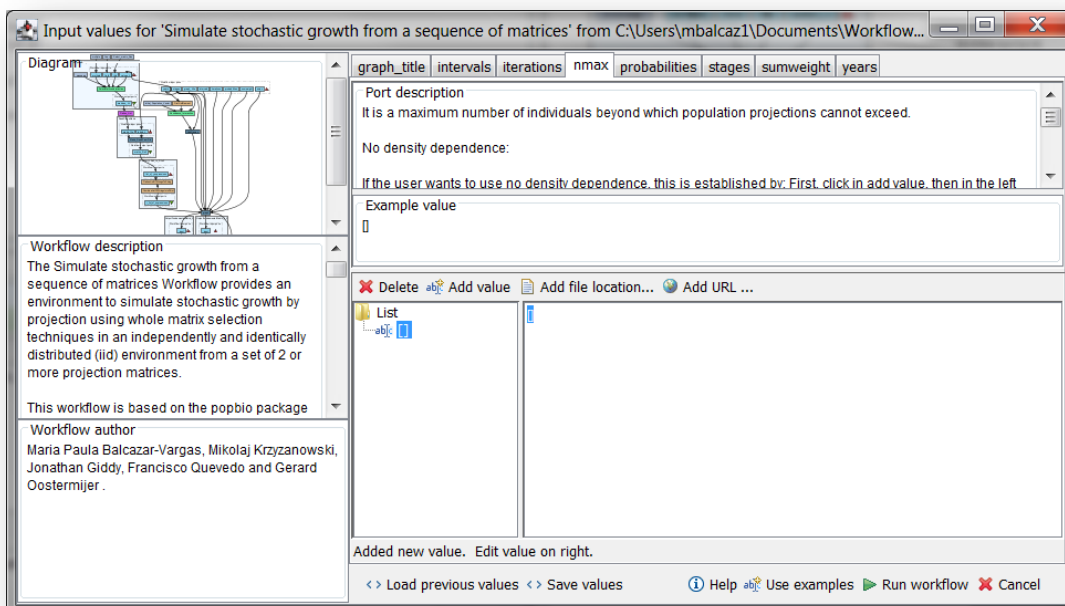
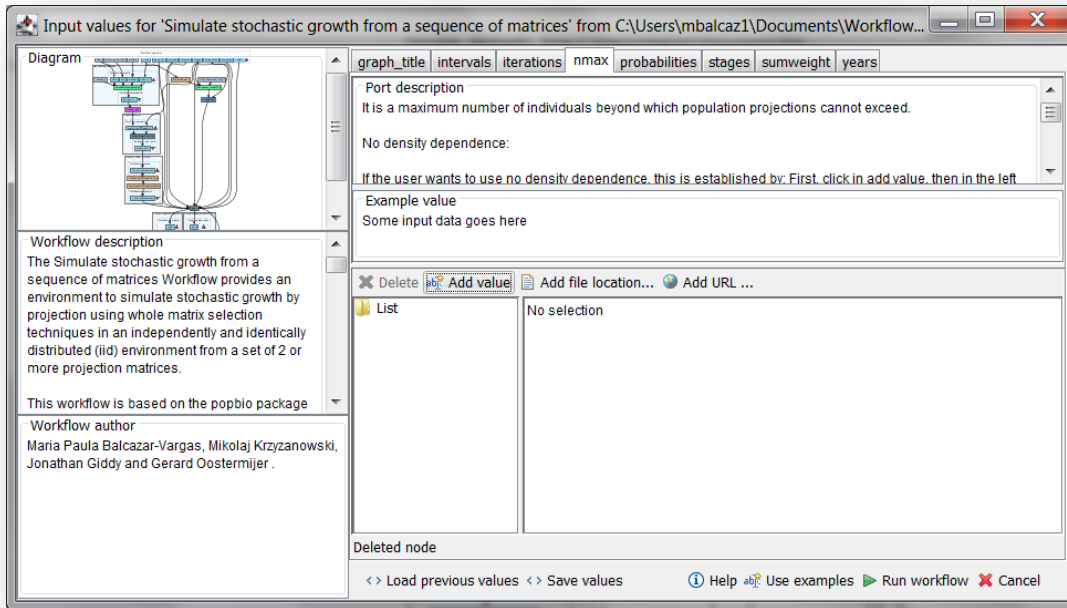


Figure 6. No density dependence: then in the left window a List is added with a [] element in the right window.

## Simulate stochastic growth from a sequence of matrices Workflow

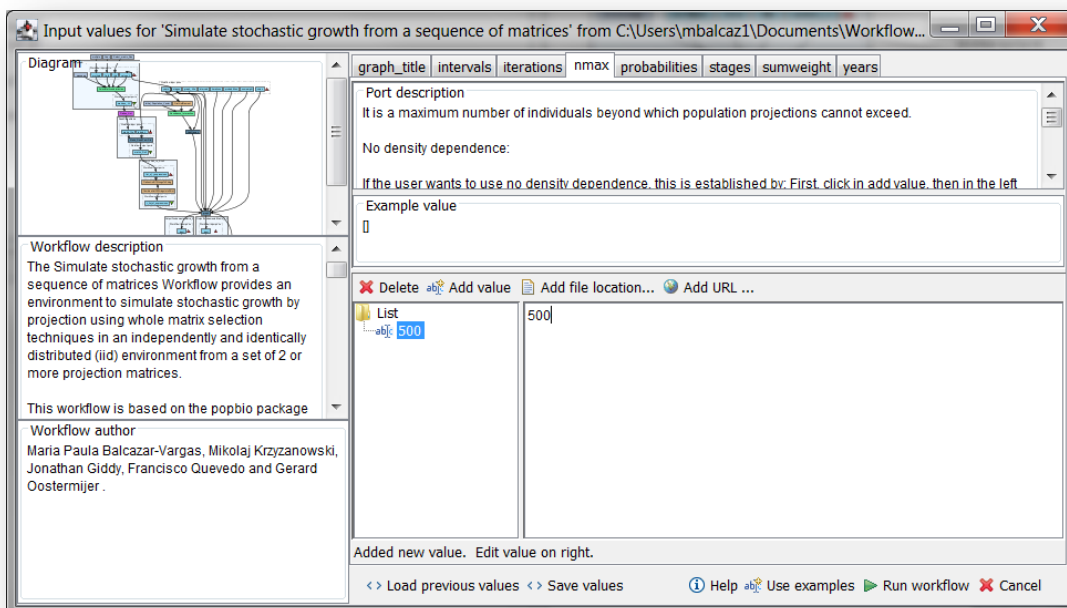


**Figure 7.** No density dependence: Delete the []. After deletion of [] the "List" has no elements in the right window.

### Density dependence:

If the user wants to introduce density dependence: Click in add value and write the number in the right space (Fig. 8).

e.g. 500



**Figure 8.** Density dependence: Click in add value and write the number in the right window (e.g. 500).

## Simulate stochastic growth from a sequence of matrices Workflow

5) **Probabilities:** it is a vector of probability weights used by sample for selecting the projection matrices. In the following example, 6 matrices of 6 consecutive years are used, therefore will be 6 probabilities weights, all sum up 1.

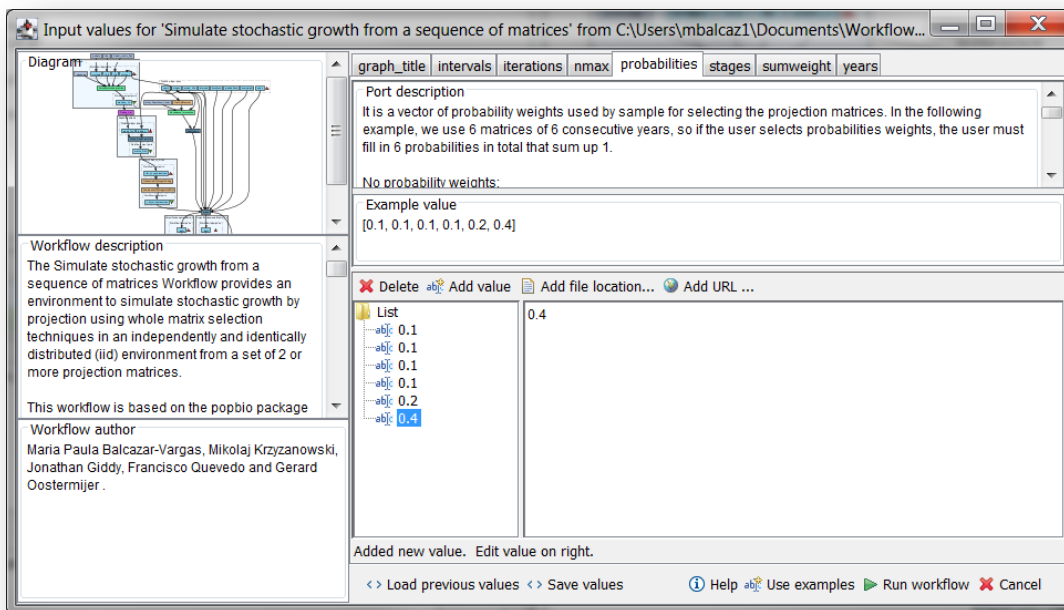
### No probability weights:

To use *no density dependence*: first, click in add value, then in the left window a List is added with a [] element in the right window. Delete the []. After deletion, the "List" has no elements in the right window. When running the workflow, Rshell script will assign NULL to the input (Fig 5 - 7).

### Probability weights:

The respective probabilities must be filled one by one and they must sum up 1. First, press add value (see arrow in Fig. 9), fill a probability number and press enter; then press add value and fill once again the next probability, repeat the action until you have filled the probabilities for all matrices. The order of the probabilities must follow the order of the years input. Each year represent a matrix. The users can change the frequencies with which the different matrices are chosen. *In this example not all the 6 matrices is equally likely to occur.*

e.g. [0.1, 0.1, 0.1, 0.1, 0.2, 0.4] (all the probabilities must sum up 1)



**Figure 9.** Probability weights: Click in add value and write the number in the right window.

6) **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 10), 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.

## Simulate stochastic growth from a sequence of matrices Workflow

In the following example, the matrix has 5 stages or categories:

The stages of this matrix are called:

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

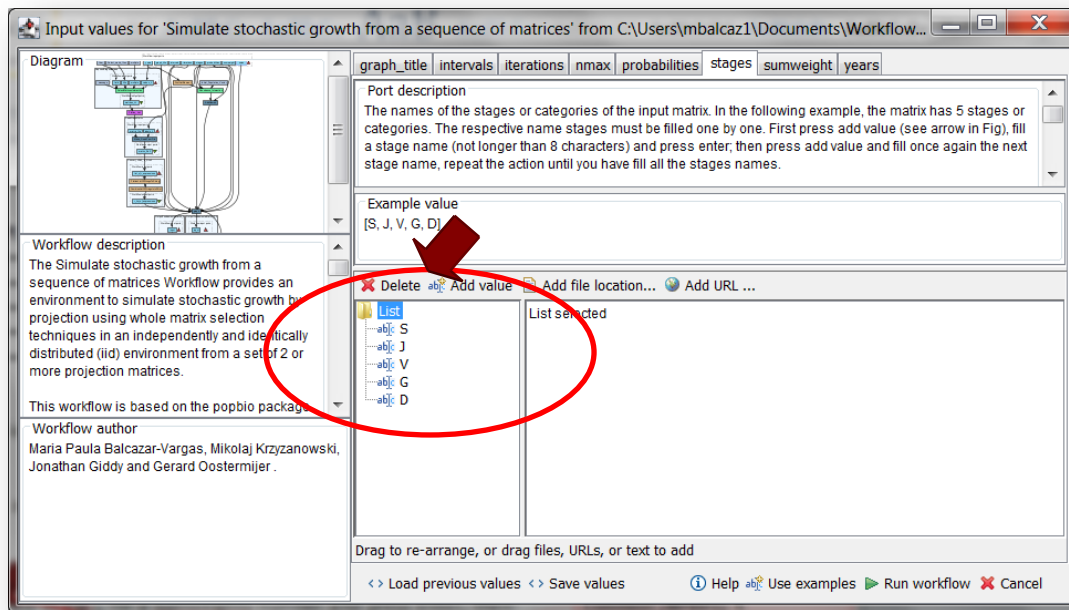


Figure 10. Inputs stages names to be filled in (example).

7) **Sumweight:** A vector of ones and zeroes used to omit stage classes when checking density threshold.

The respective sumweights must be filled one by one, the values must be 1 if the stage is included or 0 if is absent. First, press add value (see arrow in Fig. 11), fill a sumweights number and press enter; then press add value and fill once again the next sumweight, repeat the action until you have filled the sumweights for all stages. The order of the sumweights must follow the order of the stage input.

e.g. [1, 1, 1, 1, 1] (all the stages are included), this is the values used in this tutorial.

e.g. [0, 1, 1, 0, 1] (stage S and G are omit)

## Simulate stochastic growth from a sequence of matrices Workflow

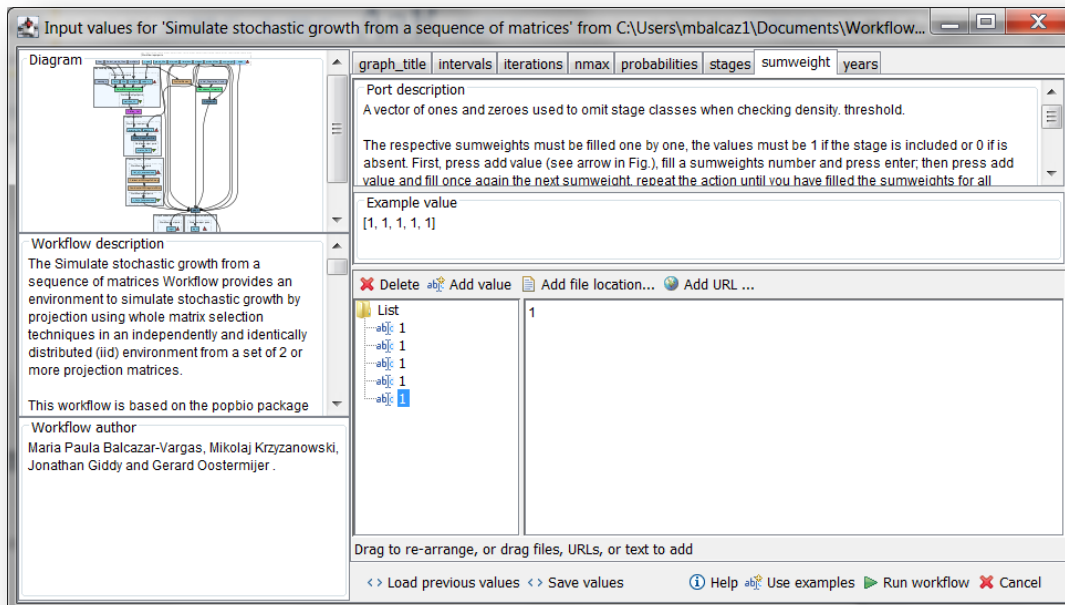


Figure 11. Sumweights to be filled in.

8) **Years:** Years in which the data of the matrices were collected. The years must be added one by one as list. First click in add value, fill the first year and press enter, then press add value and fill once again the next year, repeat the action until you have fill all the years (Fig 12).

The years can be display as follow:

e.g.

- 1987 or 1987-1988

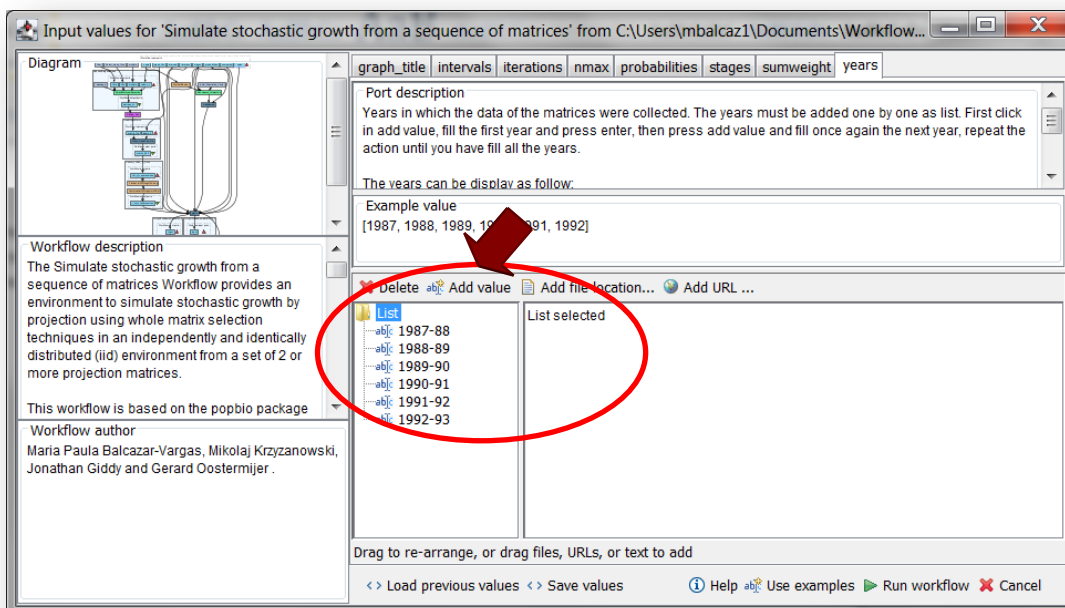


Figure 12. Inputs years to be filled in (example).

## Simulate stochastic growth from a sequence of matrices Workflow

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) **Years:** Please be shore that all the files are saved in the same directory. This interaction page will show the submitted years. When the dialogue appears (Fig 13), click Browse for the first year (1987-88). A file dialog appears. Select the file (a stage matrix) and click Open (Fig 14). The selected matrix appears in the browser (Fig 15). Repeat the procedure for each year (Fig 16). Once all the matrices are submitted, click the Confirm button (Fig 16). After clicking Confirm, the confirmation "Matrices submitted" appears (Fig 17). When the analyses are completed, they appear on different windows under results in Taverna, the user can save each output separately.

In the following example, there are 6 years or 6 matrices:

- 1) 1987-88
- 2) 1988-89
- 3) 1989-90
- 4) 1990-91
- 5) 1991-92
- 6) 1992-93

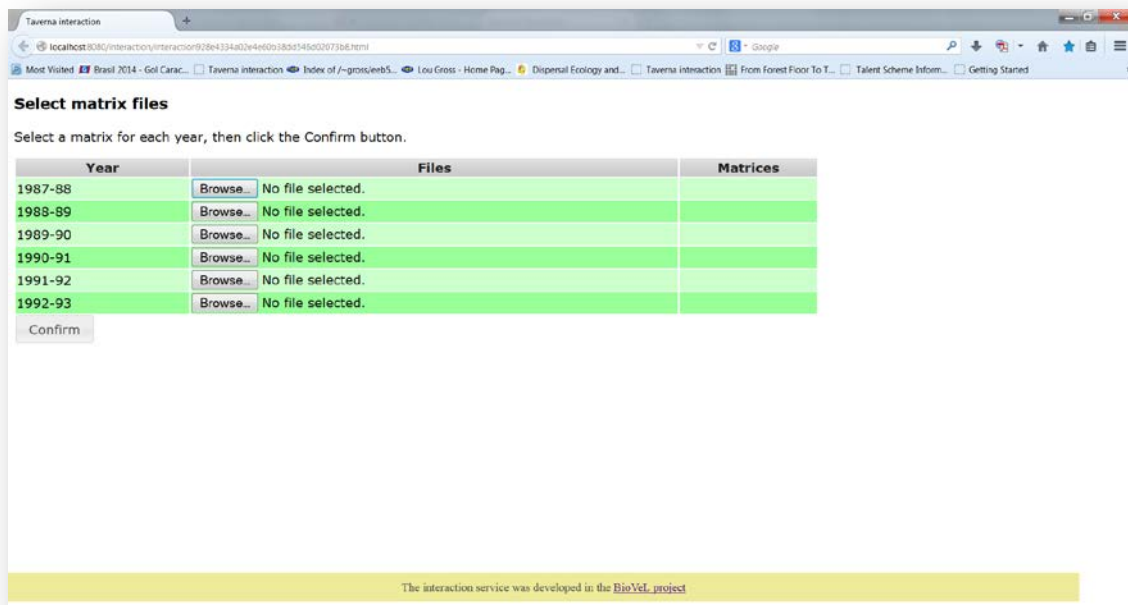
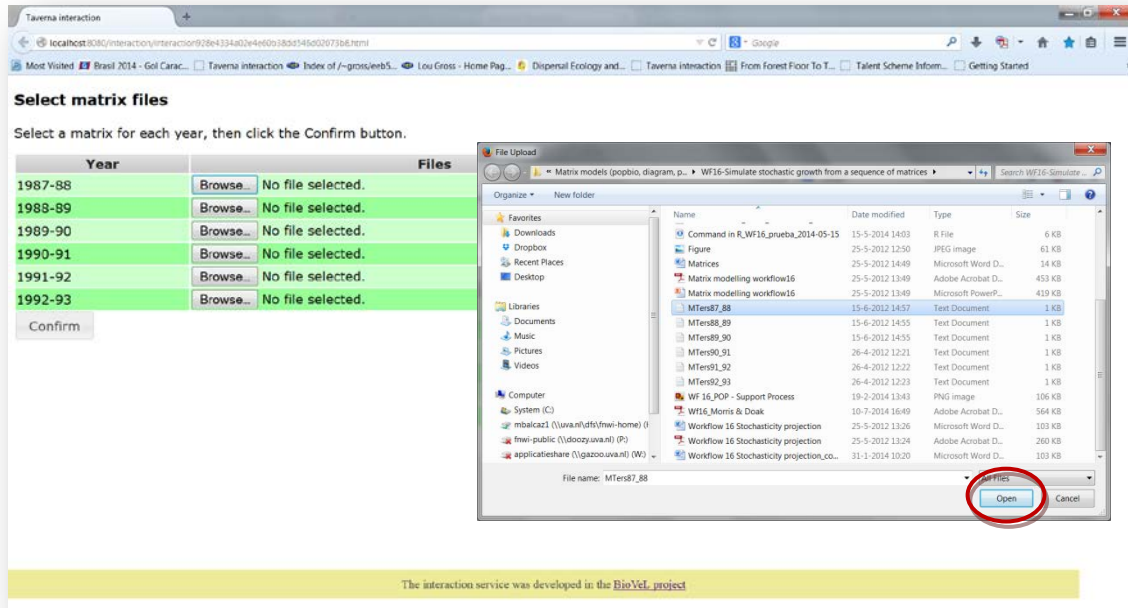
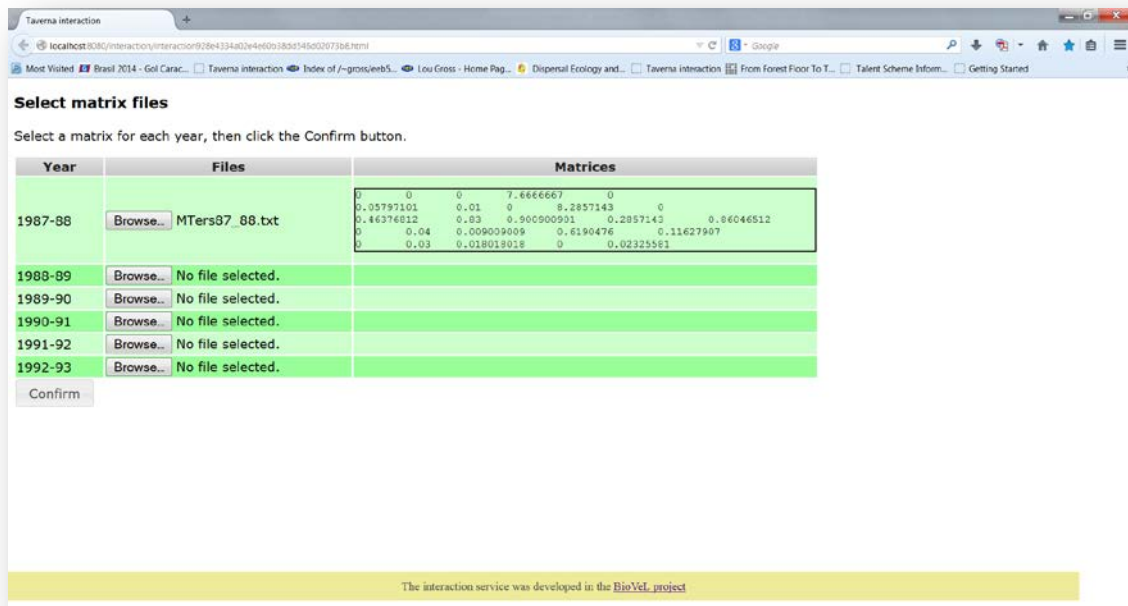


Figure 13. Places dialogue appears.

## Simulate stochastic growth from a sequence of matrices Workflow

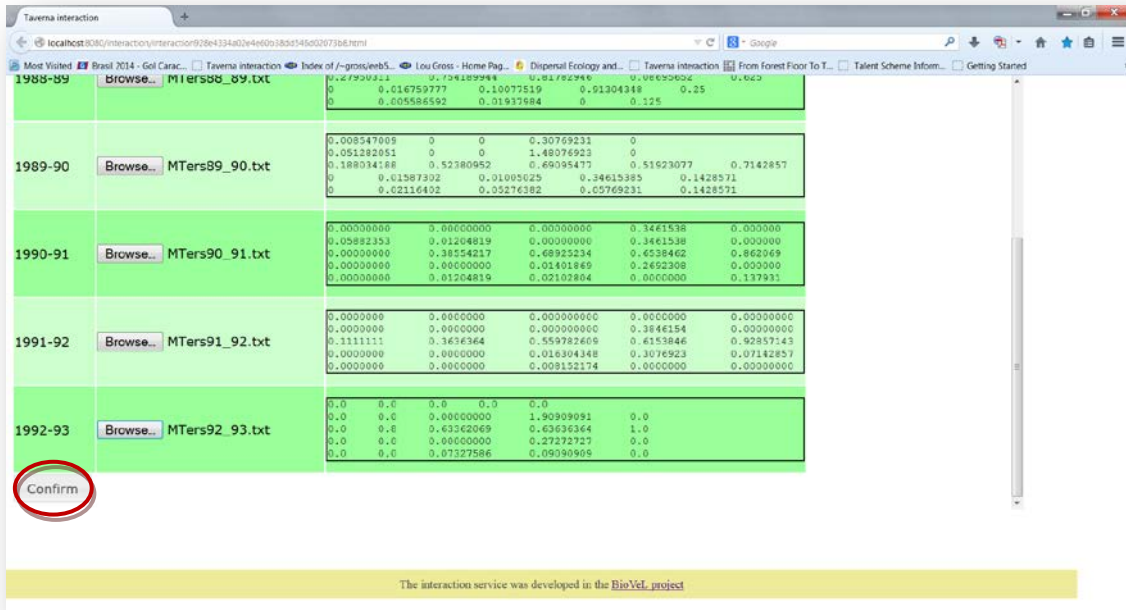


**Figure 14.** Years dialogue browse. Click in browse (1987-88), and select the file and then click the Open button.

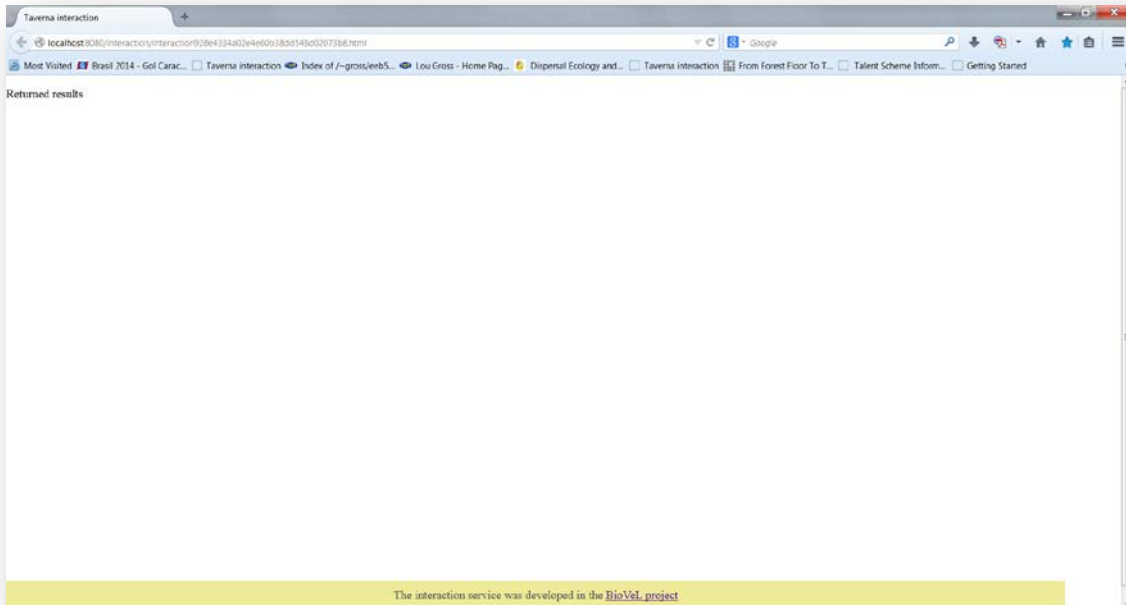


**Figure 15.** Year's dialogue. The selected matrix appears. Repeat the procedure for each year.

## Simulate stochastic growth from a sequence of matrices Workflow



**Figure 16.** Year's dialogue. Once all the matrices are submitted, click in Confirm.



**Figure 17.** Year's dialogue. After the user clicks Confirm, the confirmation "Returned results" appears.



## Simulate stochastic growth from a sequence of matrices Workflow

2) **Initial population vector:** In this dialogue appear the fields with the initial abundance per stage observed in the field in the first year (Fig 18 - 20). As an example *Gentiana pneumonanthe* has 5 stages with its respective abundance. Abundances of the first year, 1987:

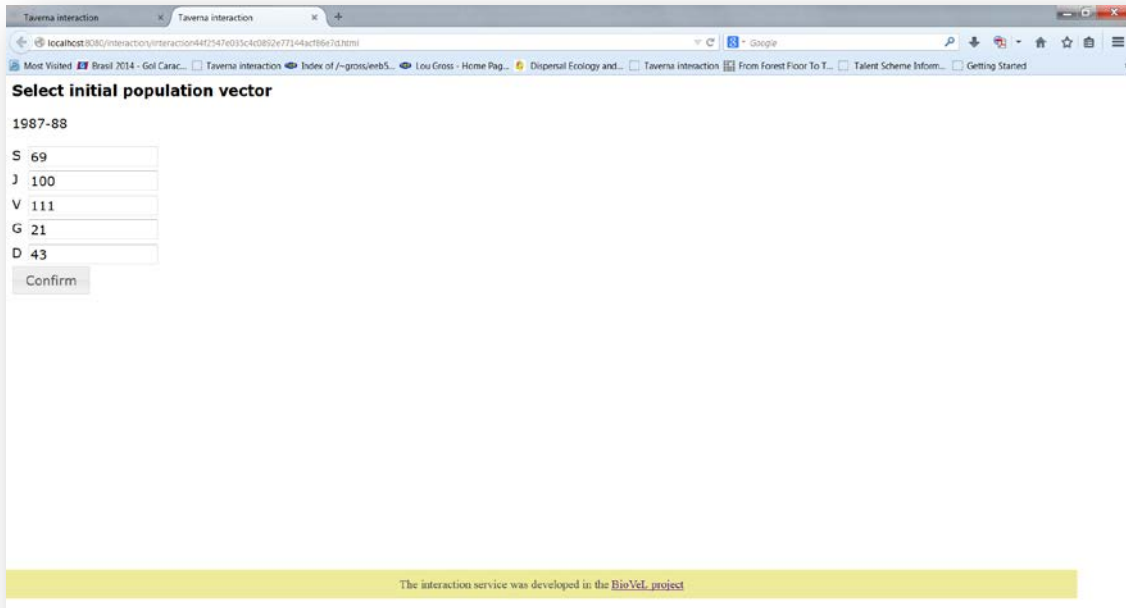
Stage abundance

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

The screenshot shows a web browser window with the title 'Taverna interaction'. The address bar contains the URL 'localhost:8080/interaction/interaction0442547e0135c4e0852e77144act86e7d.html'. The main content area is titled 'Select initial population vector' and has a sub-header '1987-88'. Below this, there are five input fields labeled S, J, V, G, and D, each containing the number '0'. A 'Confirm' button is positioned below the input fields. At the bottom of the page, a yellow banner contains the text 'The interaction service was developed in the [BioVeL project](#)'.

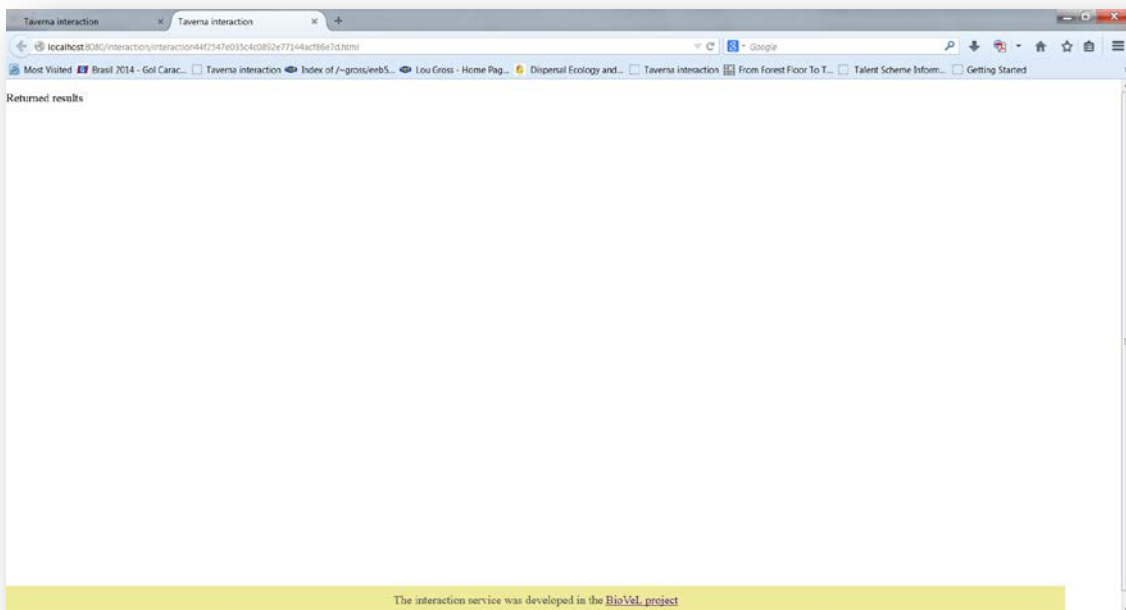
**Figure 18.** Initial population vector dialogue. Fill in the abundance per stage appears.

## Simulate stochastic growth from a sequence of matrices Workflow



The screenshot shows a web browser window with the title 'Taverna interaction'. The address bar shows a localhost URL. The main content area is titled 'Select initial population vector' and displays the year '1987-88'. Below this, there are five input fields for population counts: S (69), J (100), V (111), G (21), and D (43). A 'Confirm' button is located below the input fields. At the bottom of the page, a yellow footer contains the text: 'The interaction service was developed in the [BioVeL project](#)'.

**Figure 19.** Initial population vector dialogue. Once the abundances values were submitted. The user confirms the field.



The screenshot shows the same web browser window as Figure 19, but the main content area is now empty except for the text 'Returned results' at the top. The 'Confirm' button is no longer visible. The yellow footer at the bottom remains the same, with the text: 'The interaction service was developed in the [BioVeL project](#)'.

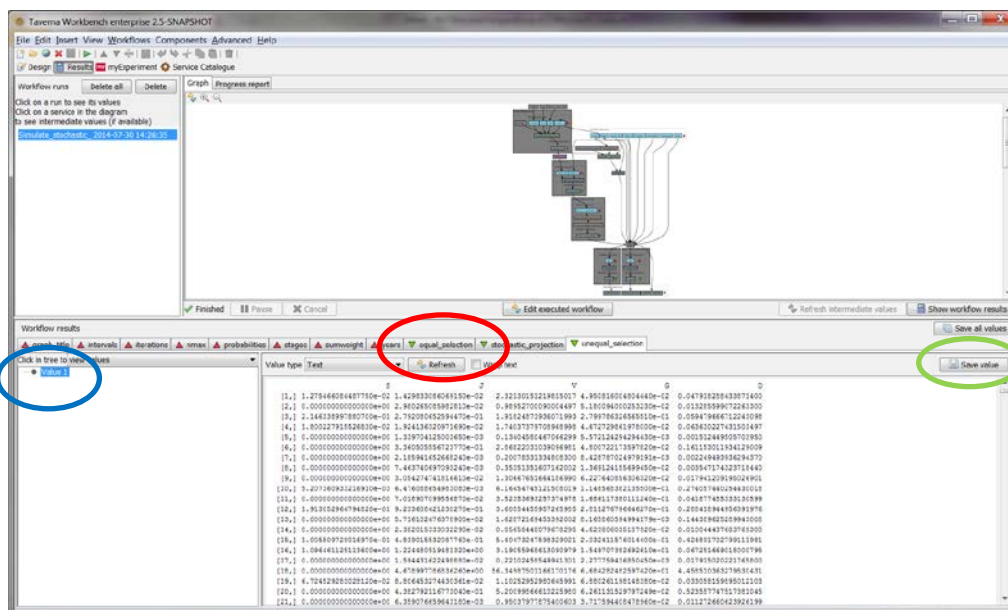
**Figure 20.** After the user clicks Confirm, the confirmation “Returned results” appears. Refer back to Taverna workbench.

After the user confirms the fields, the workflow performs the analysis. To complete all the analysis may take several minutes, depend on the iterations and number of matrices.

## 5.4 Save data/results

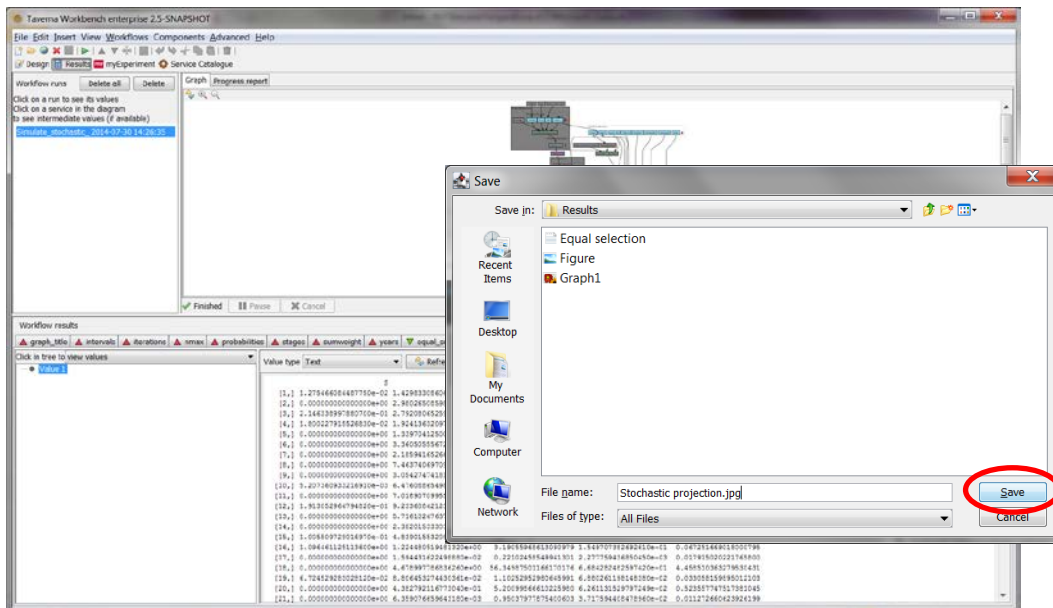
### 5.4.1 OUTPUTS

When the analyses are completed, they appear on different windows under results in Taverna. First, click in the selected result window e.g. **Equal selection** (Fig 21, red oval). Second, click in the left window on value 1 (Fig 21, blue oval) and you will see the **Equal selection** results in the right window (Fig 21). Third, on the right window click on save value (Fig 21, green 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 22). Finally, save the file in the chosen map (Fig 22).



**Figure 21.** Taverna workbench results. When the analyses are completed, they appear on different windows under results in Taverna. Click on **Equal selection** window (red oval). Do click on Value 1 on the left window of the **Equal selection** output results (blue oval). Save value on the right window of the **Equal selection** output results (green oval).

## Simulate stochastic growth from a sequence of matrices Workflow



**Figure 22.** Name the file and add the extension, for a text file: e.g. .csv or .txt. For images: e.g.: .jpg or png. Click on Save.

Each simulated realization of population growth begins with an ***Initial population vector*** with entries equal to the numbers of marked plants in each size class over all study plots in the first year of Oostermijer's demographic study, and runs for 25 years (***Intervals***). The initial total population density (the sum of the initial population vector) is 344 (see sum of ***Initial population vector***). The distribution of total population density at year 25 for 1000 independent realizations is shown in ***Stochastic projection*** (Fig 25). Please refer to Morris and Doak 2002, pages:229-233.

1) ***Equal selection:*** (Fig 23).

The screenshot shows a Microsoft Excel spreadsheet titled 'Equal selection'. The data is organized in columns A through F and rows 1 through 29. The values are in scientific notation, representing population density over time.

	A	B	C	D	E	F
1	5					
2	[1]	1,465+01	1,595+01	1,59,3450	4,55E+00	3,43E+00
3	[2]	5,06E-02	5,52E-02	0,492583	1,54E-02	5,07E-02
4	[3]	0,00E+00	9,60E-01	35,80814	1,70E+00	4,81E-01
5	[4]	1,11E+00	1,57E+00	7,305803	1,05E+00	1,79E-01
6	[5]	9,52E-01	4,60E+00	48,61667	2,38E+00	3,95E+00
7	[6]	1,27E+01	1,39E+01	46,9311	2,23E+00	1,11E+00
8	[7]	0,00E+00	1,34E+00	40,5191	1,73E+00	2,99E-01
9	[8]	4,24E+00	4,58E+00	12,68925	4,89E-01	2,54E-01
10	[9]	8,91E+00	4,50E+01	357,6048	1,52E+01	2,05E+01
11	[10]	2,12E+01	2,31E+01	82,37666	3,38E+00	1,74E+00
12	[11]	1,63E+00	2,89E+00	6,35847	6,89E-01	9,00E-02
13	[12]	1,39E+00	1,49E+00	3,058001	1,53E-01	7,34E-02
14	[13]	2,75E-01	2,99E-01	2,867121	8,61E-02	6,20E-02
15	[14]	1,06E+00	1,08E+00	48,12512	1,74E+00	1,48E+00
16	[15]	2,31E-01	1,11E+00	6,991744	3,74E-01	5,42E-01
17	[16]	3,11E+00	3,38E+00	6,691423	3,52E-01	1,36E-01
18	[17]	1,25E-02	1,45E-02	0,706407	2,27E-02	2,48E-02
19	[18]	0,00E+00	5,63E-02	9,267335	3,61E-01	1,10E-01
20	[19]	5,13E-01	5,36E-01	8,606358	5,45E-01	2,73E-01
21	[20]	1,04E-01	1,29E-01	3,57415	1,37E-01	1,36E-01
22	[21]	1,11E+00	1,20E+00	4,420792	1,76E-01	9,03E-02
23	[22]	0,00E+00	2,89E-02	1,481056	7,06E-02	3,76E-02
24	[23]	0,00E+00	4,30E-01	3,168744	6,26E-02	3,53E-02
25	[24]	1,19E-02	5,74E-02	2,35336	8,79E-02	2,01E-01
26	[25]	0,00E+00	6,58E-01	64,25824	2,68E+00	7,24E-01
27	[26]	2,74E+00	3,12E+00	13,13746	4,48E-01	2,75E-01
28	[27]	0,00E+00	1,35E-02	0,64098	3,11E-02	7,56E-03
29	[28]	0,00E+00	6,12E-01	86,82922	3,32E+00	1,06E+00
30	[29]	8,88E-01	1,18E+00	14,29234	9,01E-01	4,28E-01

**Figure 23.** Equal selection data(.txt or .csv).

## Simulate stochastic growth from a sequence of matrices Workflow

### 2) Unequal selection: (Fig 24).

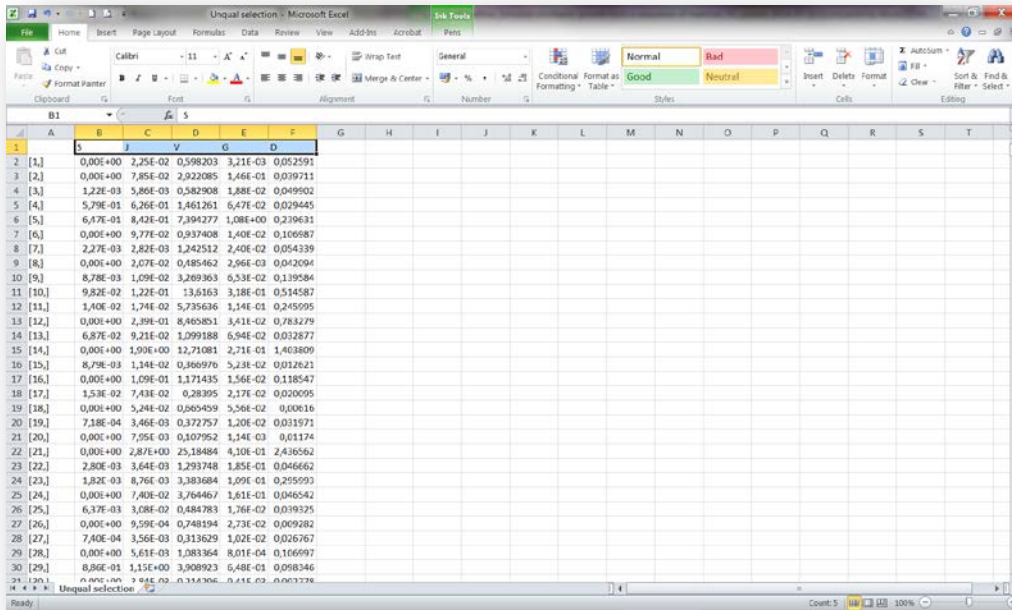


Figure 24. Unequal selection (.csv or text files).

### 3) Stochastic projection (Fig 25). Results of simulating a population of Gentiana in an iid stochastic environment. The initial population size was 344 (see Initial population vector).

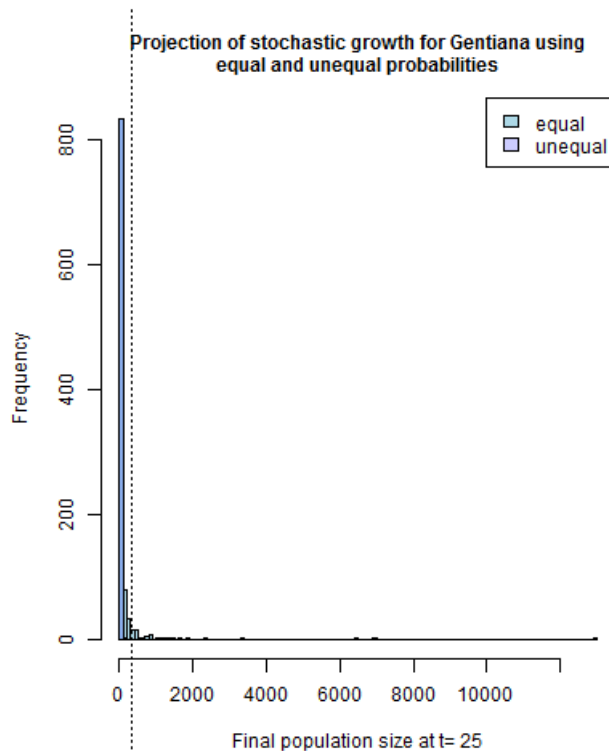


Figure 25. Stochastic projection (.jpg or .png files).

## 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 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, 2<sup>nd</sup> Edition. Sinauer Associates, Sunderland, Massachusetts.

**Morris, W. F., and D. F. Doak.** 2002. Quantitative conservation biology: Theory and practice of population viability analysis. Sinauer, Sunderland, Massachusetts, USA. 480 pages

**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. **Mikolaj Krzyzanowski**, 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. **Francisco Quevedo**, Cardiff School of Computer Science and Informatics, Cardiff University, Cardiff CF24 3AA, United Kingdom.

## Simulate stochastic growth from a sequence of matrices Workflow

5. *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