

BioVeL – Biodiversity Virtual e-Laboratory

Workflow Documentation

Calculate quasi-extinction threshold 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:
Project Co-ordinator:
Project Homepage:
Duration of Project:

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

Calculate quasi-extinction threshold Workflow

1 Description

This workflow estimates by simulation the quasi-extinction probability time cumulative distribution function for a structured population in an independently and identically distributed (iid) stochastic environment. This workflow is based on the *popbio* package (stoch.quasi.ext - Calculate quasi-extinction threshold, Stubben, Milligan and Nantel, 2013) based on the MATLAB code in Box 7.5 (Morris and Doak 2002). For more details of the analysis see: Calculating the probability of hitting a quasi-extinction threshold by time t , method: simulating extinction probabilities (; Morris and Doak 2002, pag: 241-244 and Caswell 2001, pag: 443-444).

This model do not permit extinction, however we can study *quasi-extinction* (Caswell 2001). A population is quasi-extinct if it shrinks to a specific fraction of its current size (Ginsbur et al 1982 in Caswell 2001). The quasi-extinction threshold can be chosen in the belief that it would pose a significant threat of the real exaction (Caswell 2001).

METHOD: SIMULATING EXTINCTION PROBABILITIES Keep track of whether the total population density (or the density summed across a subset of the classes with which we are particularly concerned, such as the reproductive classes) has fallen below the quasiextinction threshold each year. The fraction of realizations that first hit the threshold during or before year t gives the cumulative probability of extinction (Morris and Doak 2002). This workflow performs such a calculation.

2 General

2.1 Name of the workflow in myExperiment.

Name: Calculate quasi-extinction threshold workflow

Download: The workflow pack can be downloaded at myExperiment under the following address: <http://www.myexperiment.org/packs/653.html> or only the workflow: <http://www.myexperiment.org/workflows/4435.html>

2.2 Date, version and licensing

Last updated: 14th 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:

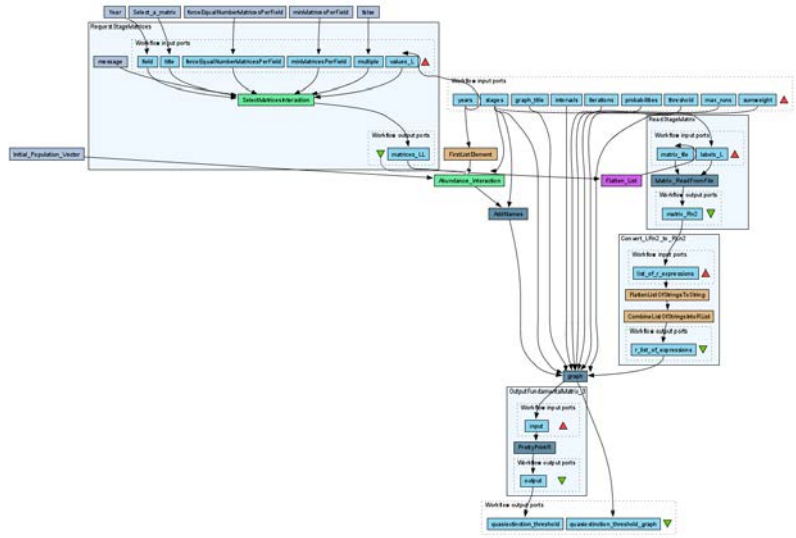
The input data and results reported in this publication (tutorial) come from data (Dr. Gerard Oostermeijer unpublished results and publication: Oostermeijer, J.G.B. M.L. Brugman, E.R. de Boer; H.C.M. Den Nijs. 1996. Temporal and Spatial Variation in the Demography of *Gentiana pneumonanthe*, a Rare Perennial Herb. *The Journal of Ecology*, 84: 153-166.) using BioVeL workflows and services (www.biovel.eu). Calculate quasi-extinction threshold 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, cumulative distribution function (CDF), demography, *Gentiana pneumonanthe*, matrix, matrix population models, Morris and Doak 2002, package 'popbio' in r, population viability analysis (PVA), probability of quasi-extinction threshold by time *t*, simulating extinction probabilities, stage matrix, stochastic environment.

3.2 Scientific workflow description



The aim of the Calculate quasi-extinction threshold workflow is to provide a connected environment to estimates by simulation the quasi-extinction probability time cumulative distribution function for a structured population in an independently and identically distributed (iid) stochastic environment. 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 a graphic plot.

Figure 1. A) Graph of the Calculate quasi-extinction threshold 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).

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/653.html> or here below. **All the files must be saved in the same directory.**

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](#)

We should mention that this kind of analysis should not be apply to successional species as *Gentiana pneumonanthe*; but, we only do it as example how to run this workflow.

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 quasiextinction threshold graph. The number 100, used in the title is the quasiextinction threshold value used in this tutorial run. Please do not use a title longer than the suggest it. Click in set value and then write the text in the right space.

Calculate quasi-extinction threshold Workflow

e.g.: *Time to reach a quasi-extinction threshold of 100 individuals.*

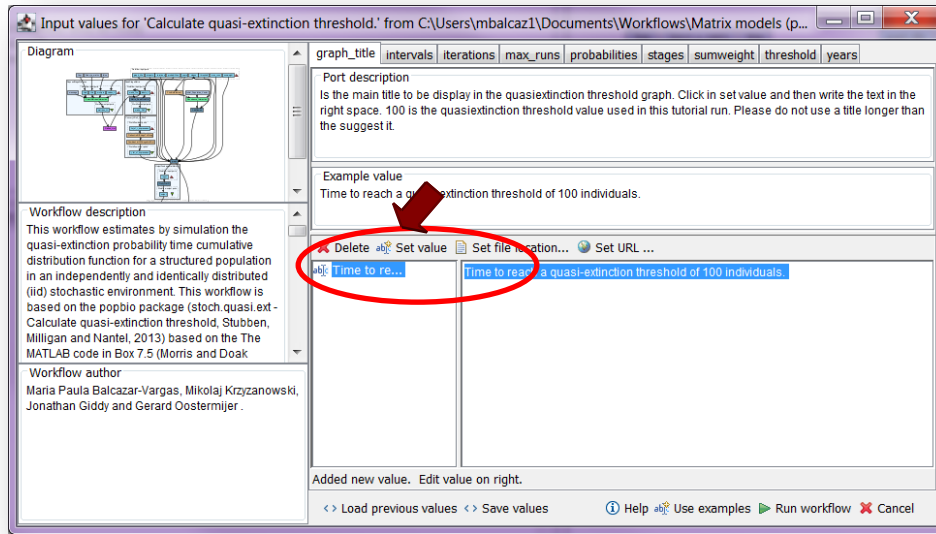


Figure 2. Graph title to be filled in (example).

2) **Intervals**: This value is the farthest future time horizon (e.g. 25) and therefore the years of axis X of the output graph. Click in set value and then write the number in the right space (Fig 3).

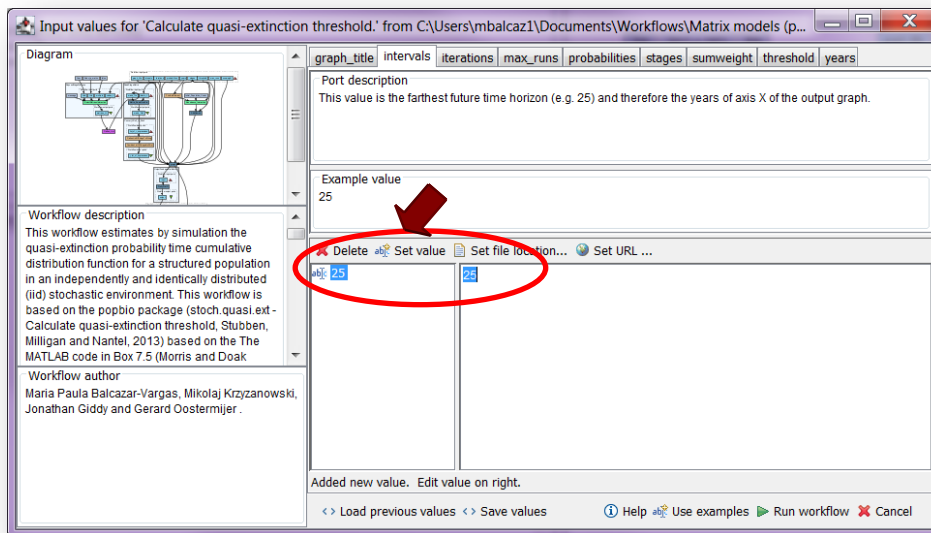


Figure 3. The number of intervals to be filled in.

3) **Iterations**: the number of realization of population growth to simulate each run or number of iterations. Click in set value and then write the number in the right space (Fig. 4).

Calculate quasi-extinction threshold Workflow

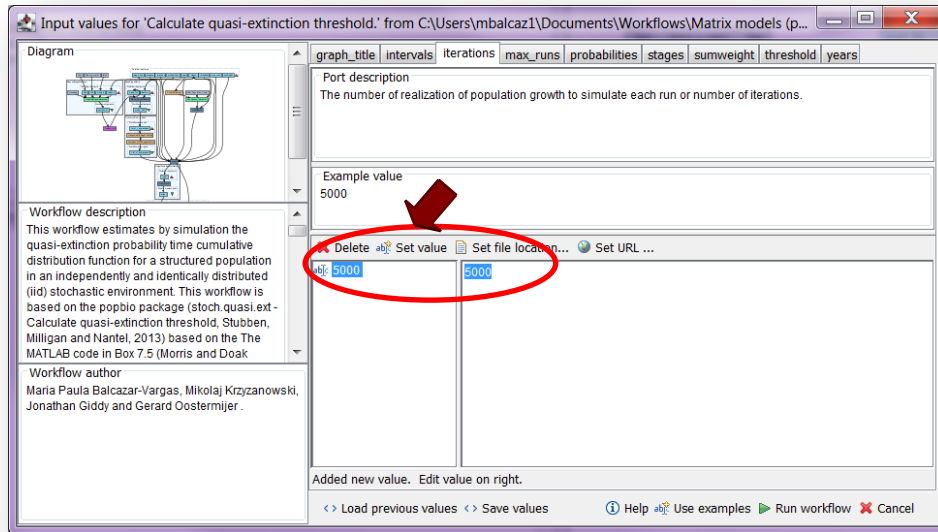


Figure 4. Iterations to be filled in (e.g. 5000)

4) **max_runs**: the number of times to simulate CDF (cumulative distribution function). Click in set value and then write the number in the right space (Fig. 5).

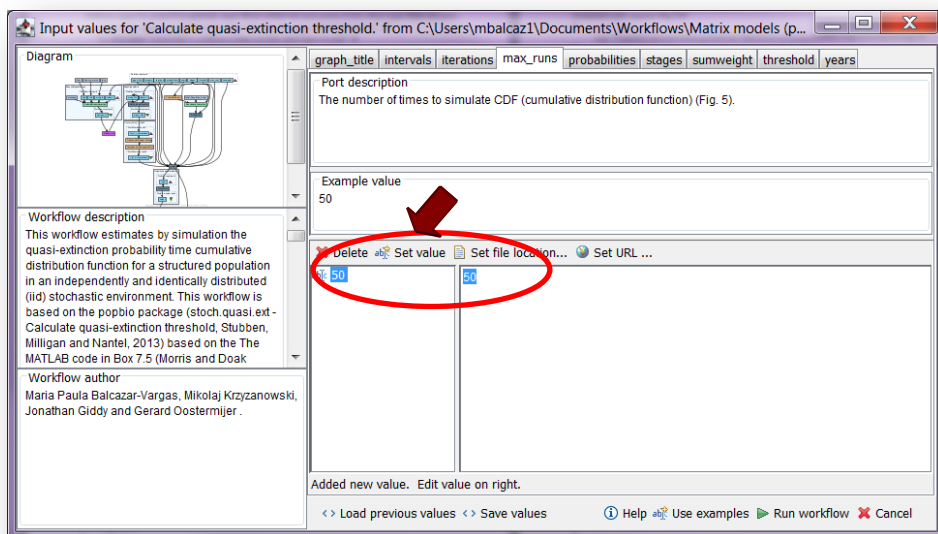


Figure 5. max_runs to be filled in (e.g. 50)

Calculate quasi-extinction threshold Workflow

5) **probabilities:** probabilities of choosing each submitted matrix (Fig. 6). Note that The value of probabilities must equal to 1. This value can be ignored if the user does want use probabilities [].

Use probabilities:

Each probability corresponding to each year and each year corresponds to a matrix. Probabilities and years should follow the same order. First press add value (see arrow in Fig 6), fill a probability value and press enter; then press add value and fill once again the next value, repeat the action until you have fill all the probabilities.

e.g. In this tutorial, 6 matrices are used, therefore 6 probabilities must be used, the probabilities must follow the same order as the submitted matrices (years).

[0.1, 0.2, 0.2, 0.2, 0.1, 0.2]. The probabilities must be settled to 1.

The first matrix (year = 1987) has a 0.1% probability to be chosen over 1.

The second matrix (year = 1988) has a 0.2% probability to be chosen over 1. etc.

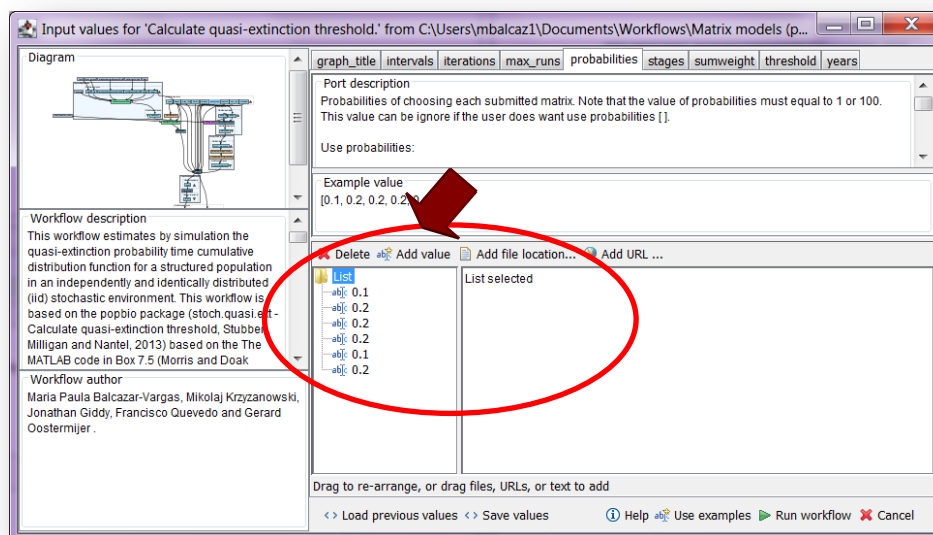


Figure 6. Probabilities to be filled in (0.1, 0.2, 0.2, 0.2, 0.1, 0.2)

Ignore probabilities:

When the user wants to ignore the use of probabilities []: Under probabilities, click in Add value (Fig. 7A), by default a list of selected probabilities will appear (Fig. 7B). Select and delete them. Once they have been deleted the user sees (Fig. 7C). Click in Delete (red arrow in 7C), in the right window will appears: No selection (Fig. 7D).

Calculate quasi-extinction threshold Workflow

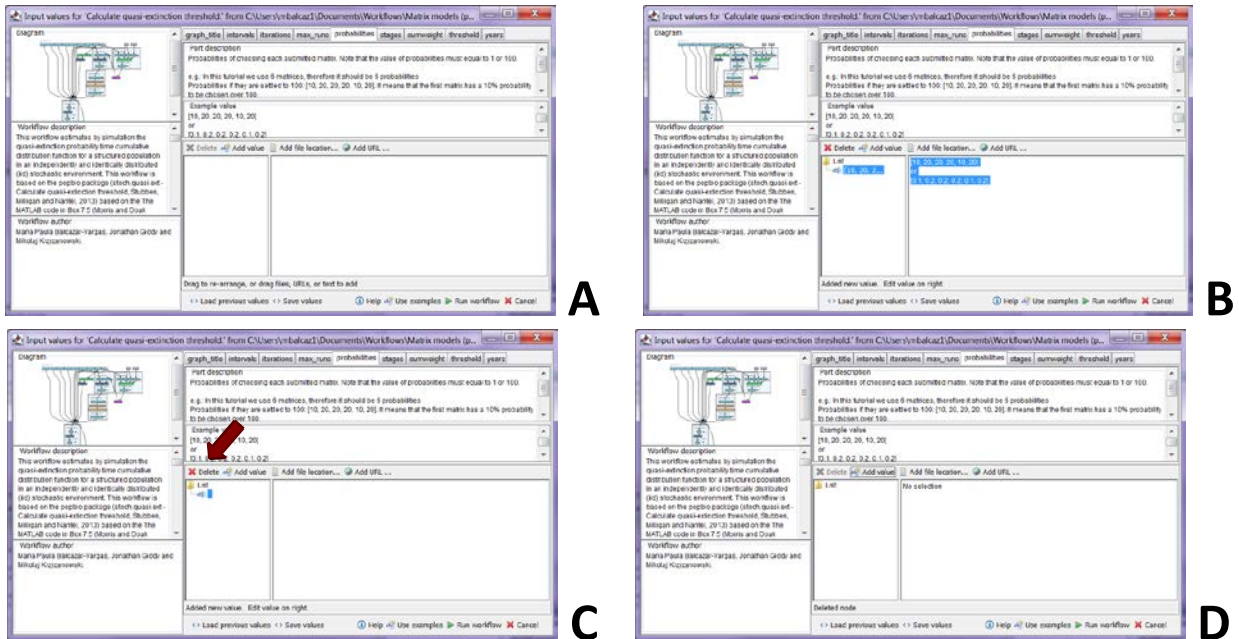


Figure 7. Ignore probabilities.

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 8), fill a stage name 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 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 |

e.g. [S, J, V, G, D]

Calculate quasi-extinction threshold Workflow

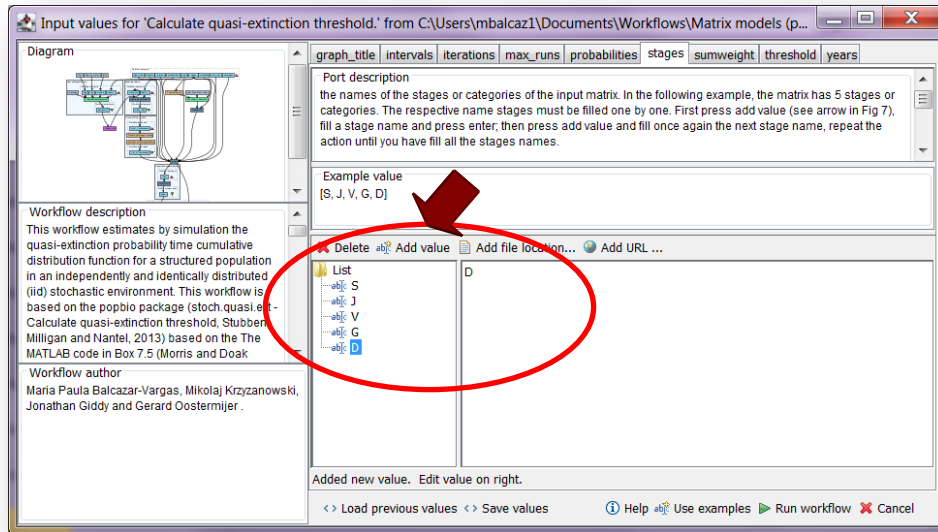


Figure 8. Inputs stages names to be filled in (example).

7) **Sumweight:** summed density to compare the quasi-extinction threshold. A vector of ones and zeroes used to omit stage classes when checking quasiextinction threshold (0 = omit a stage, 1= include). The respective sumweights must be filled one by one. First press add value (see arrow in Fig 9), fill a 0 or 1 and press enter; then press add value and fill once again the next value, repeat the action until you have fill all the sumweights.

Use sumweight, ignore stages:

In this tutorial, matrices have 5 stages, therefore 5 sumweight data must be used, the sumweight values must follow the same order as the submitted stages.

In this example, the zero instruct the workflow to ignore seedlings (S is the first stage, see **Stages**) when summing the densities across classes to compare to the quasi-extinction threshold (Figure 9).

e.g.: [0, 1, 1, 1, 1]

Calculate quasi-extinction threshold Workflow

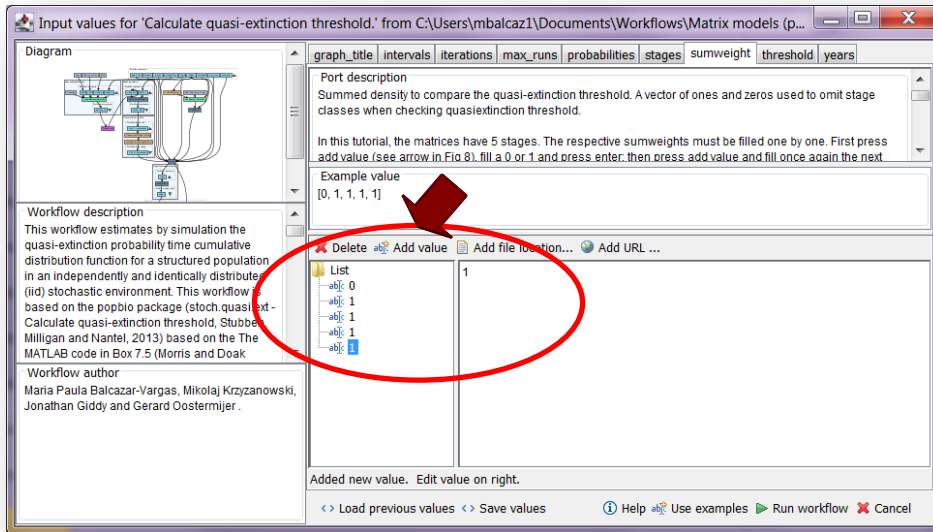


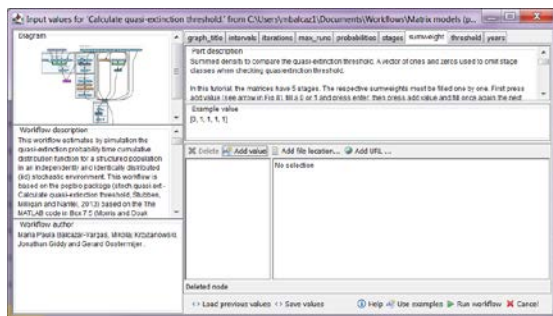
Figure 9. Sumweights to be filled in (example).

If we wanted to set the threshold in terms of only the reproductive plants, we would set all but the penultimate entry in sumweight to zero. (G is the penultimate stage, see [Stages](#))

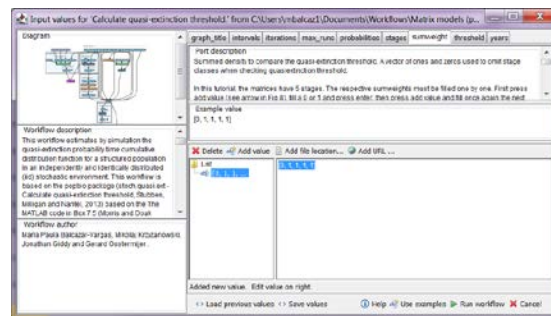
e.g.: [0, 0, 0, 1, 0]

Ignore sumweight:

When the user wants to ignore the use of sumweight [:]: Under sumweight, click in Add value (Fig. 10A see this and following figures under [probabilities](#)) by default a list of selected probabilities will appear (Fig. 10B). Select and delete them. Once they have been deleted the user sees (Fig. 10C). Click in Delete (red arrow in 10C), in the right window will appears: No selection (Fig. 10D).



A



B

Calculate quasi-extinction threshold Workflow

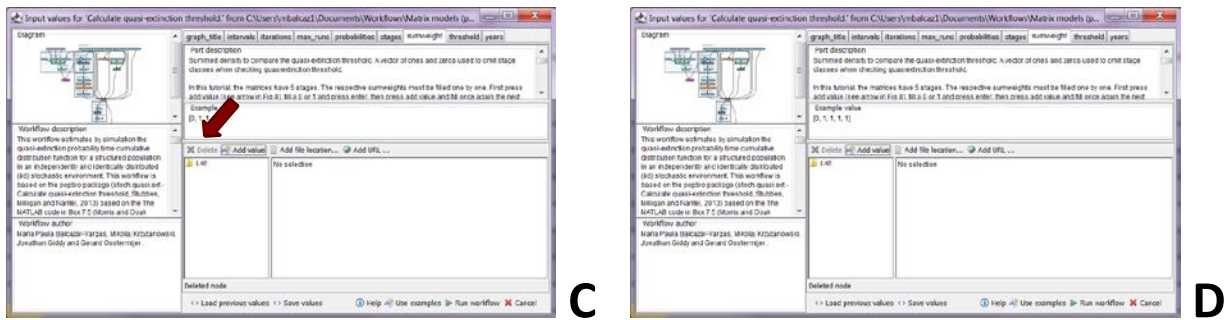


Figure 10. Ignore Sumweights.

8) **Threshold:** The quasi-extinction threshold, express as a density. The entries in the **initial population vector** represent densities, but quasi-extinction thresholds are typically expressed in terms of numbers of individuals (Morris and Doak 2002). Thus when we ask whether the summed densities across all or a subset of the classes has fallen below the quasi-extinction threshold, we first need to be sure that a numerical threshold has been converted to a density. For example, let's assume that the densities in the population vector represent mean numbers of individuals per hectare, that the population occupies 1,000 hectares, and that we have decided to set the quasiextinction threshold at a total of 100 individuals in the reproductive classes. One hundred reproductive individuals in the entire population represents an average density of 0.1 individuals per hectare summed across the reproductive classes, so we would consider the population to have hit the threshold if this sum hits 0.1, not 100. Alternatively, we could multiply the starting densities by 1,000 to arrive at an estimate of total numbers and set the quasiextinction threshold to 100 (Morris and Doak 2002) (Fig 11).

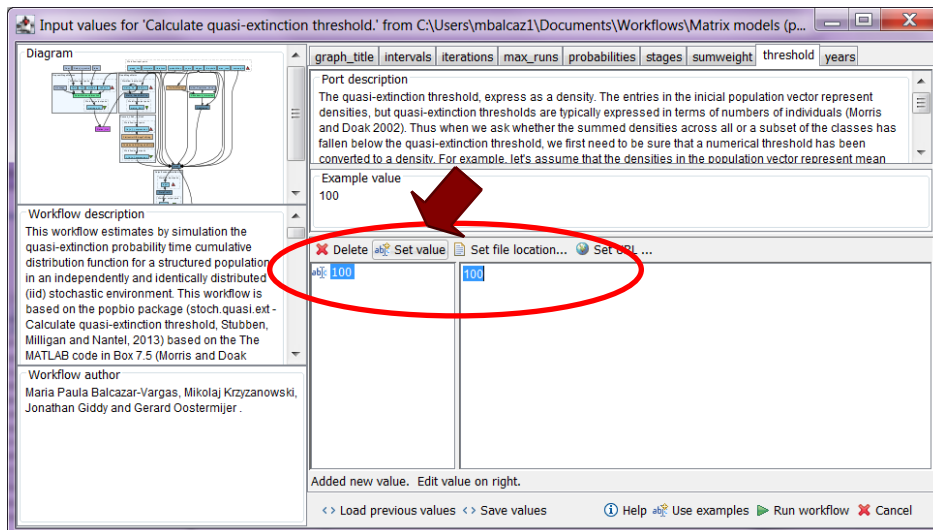


Figure 11. The quasi-extinction threshold to be filled in (example).

9) **Years:** each year represents a matrix and therefore characterizes a period of transition (Fig. 12). The respective years must be filled one by one. First press add value (see arrow in Fig 12), fill a year (e.g. 1987) and press enter; then press add value and fill once again the next year (1988), repeat the action until you have fill all the years.

In this tutorial, we have 6 matrices (see 5.2.2 **Input data**) that represent 6 years (e.g.: **1987** =data interval between 1987 and 1988; **1988**= data interval between 1988 and 1989).

Calculate quasi-extinction threshold Workflow

Years:

- 1) 1987
- 2) 1988

- 3) 1989
- 4) 1990

- 5) 1991
- 6) 1992

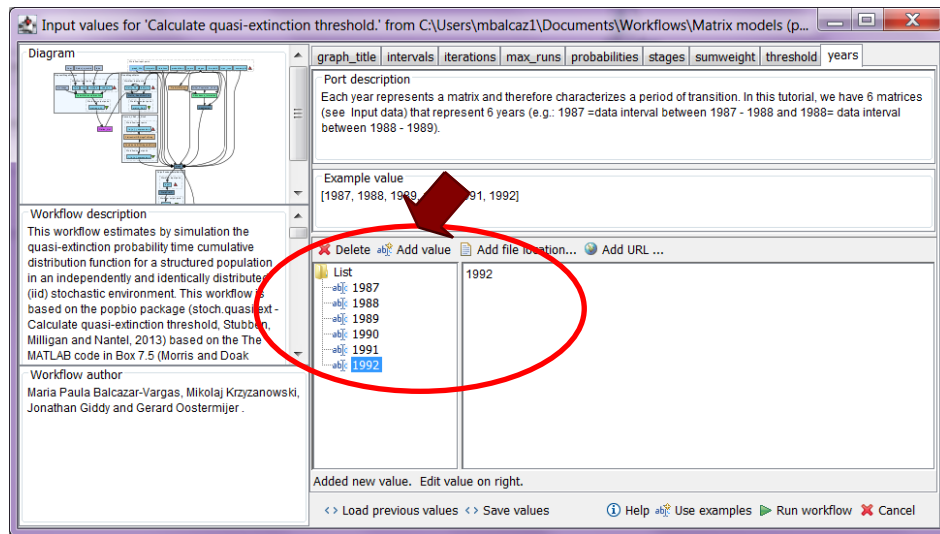


Figure 12. Inputs years to be filled in (example).

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) **Select a matrix for each year:** all the matrices must be in the same folder. The interaction page will show the submitted years. When the dialogue appears (Fig 13), click Browse for the first year (e.g. 1987). A file dialog appears. Select the file (one matrix that belong to that year) and clicks Open (Fig 14). Then 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 "Returned Results" appears (Fig 17). The workflow controls whether each year have submitted a matrix.

In the following example, there are 6 years:

- 1) 1987
- 2) 1988
- 3) 1989
- 4) 1990
- 5) 1991
- 6) 1992

Calculate quasi-extinction threshold Workflow

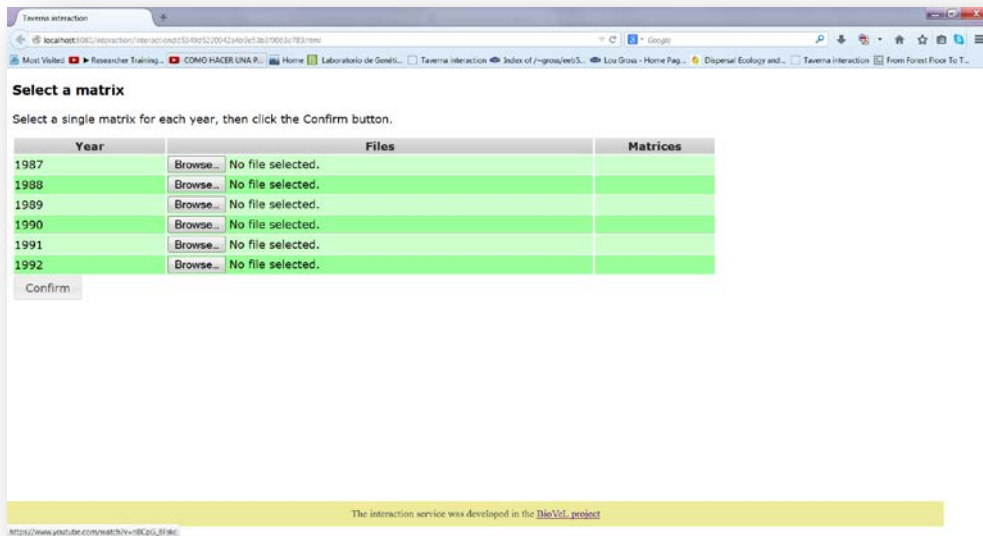


Figure 13. Select a matrix for each year dialogue appears.

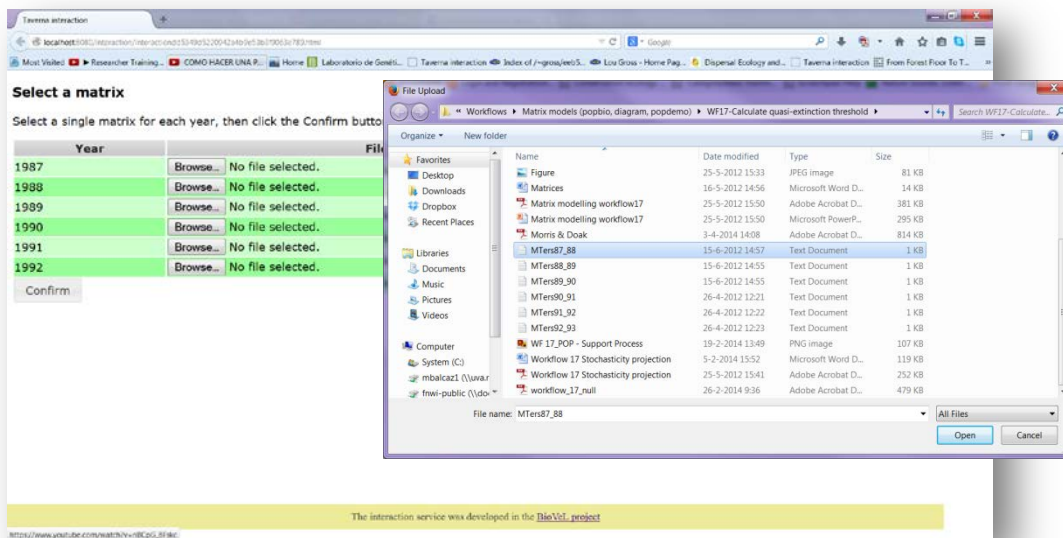


Figure 14. Select a matrix for each year dialogue browse. Click in browse (1987), and select the file (one file e.g. M_Ters87-88.txt) and then click in Open button.

Calculate quasi-extinction threshold Workflow

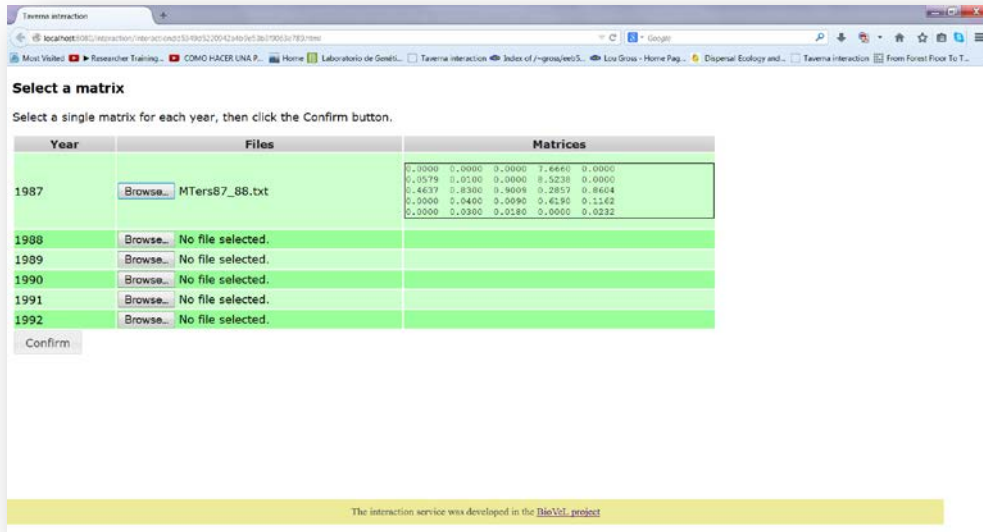


Figure 15. Select a matrix for each year dialogue. The selected matrices appear. Repeat the procedure for each year.

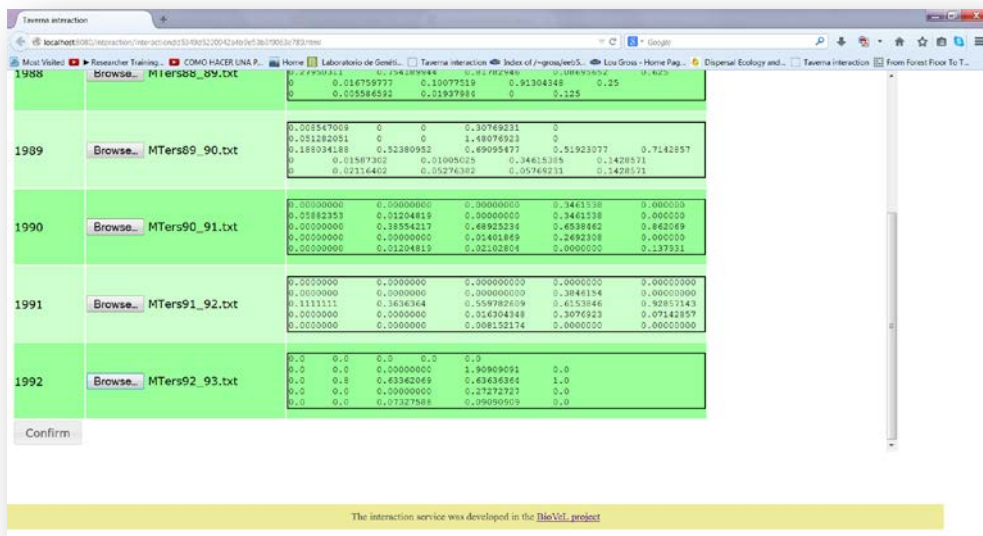


Figure 16. Select a matrix for each year dialogue. Once all the matrices are submitted, click in Confirm.

Calculate quasi-extinction threshold Workflow

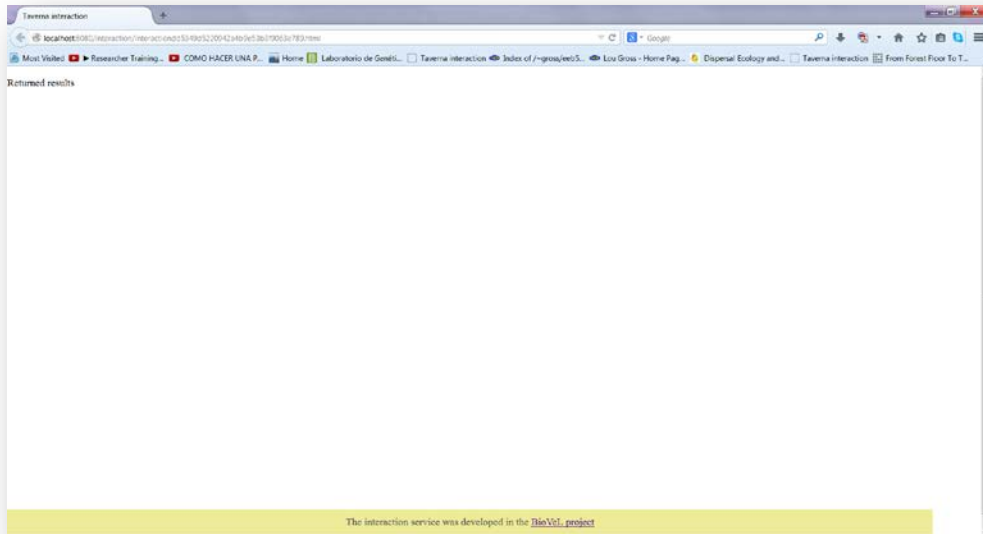
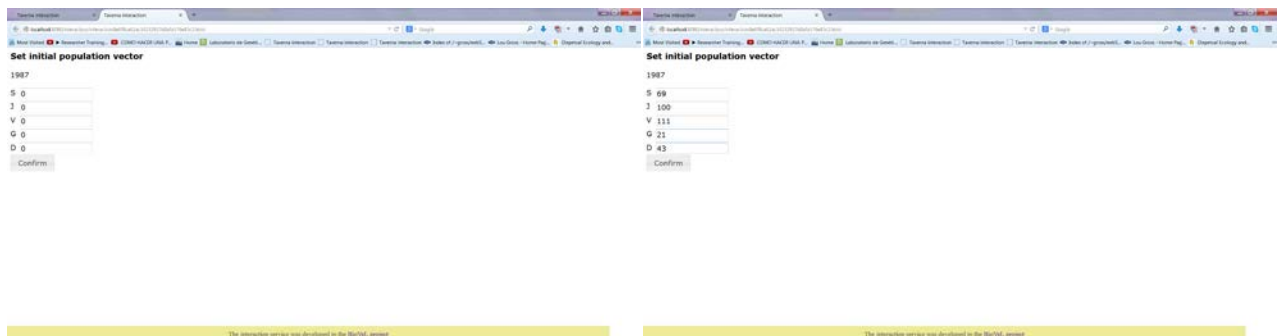


Figure 17. Select a matrix for each year dialogue. After the user clicks Confirm, the confirmation “Returned Results” appears. Refer back to Taverna workbench.

2) **Initial population vector:** In dialogue appear the fields with the initial population vectors per stage observed in the field (for more details see Morris and Doak). As an example *Gentiana pneumonanthe* has 5 stages with its respective densities. Densities of the year 1987 (Fig 18):

Stage densities

1) S (seedlings)	69
2) J (Juveniles)	100
3) V (vegetative)	111
4) G (reproductive individuals)	21
5) D (dormant plants)	435



Calculate quasi-extinction threshold Workflow



Figure 18. Initial population vectors dialogue. The density per stage dialogue appears, fill in the observed densities. Click in confirm, the confirmation “Returned Results” appears.

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

Calculate quasi-extinction threshold Workflow

5.4 Save data/results

5.4.1 OUTPUTS

First, click in the selected result window e.g. **quasi-extinction threshold** (Fig 19, red oval). Second, click in the left window on value 1 (Fig 19, blue oval) and you will see the numerical results. Third, on the right window click on save value (Fig 19, 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. Finally, save the file in the chosen map (Fig 20). Repeat the procedure for other outputs.

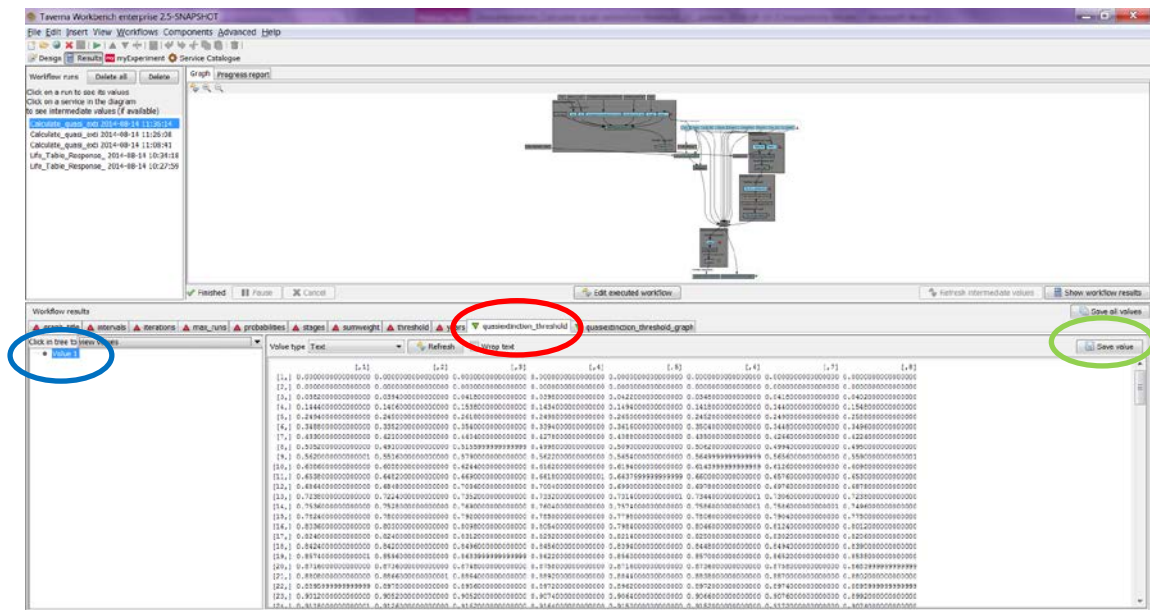


Figure 19. Taverna workbench results. When the analysis is completed, they appear on different windows under results in Taverna.

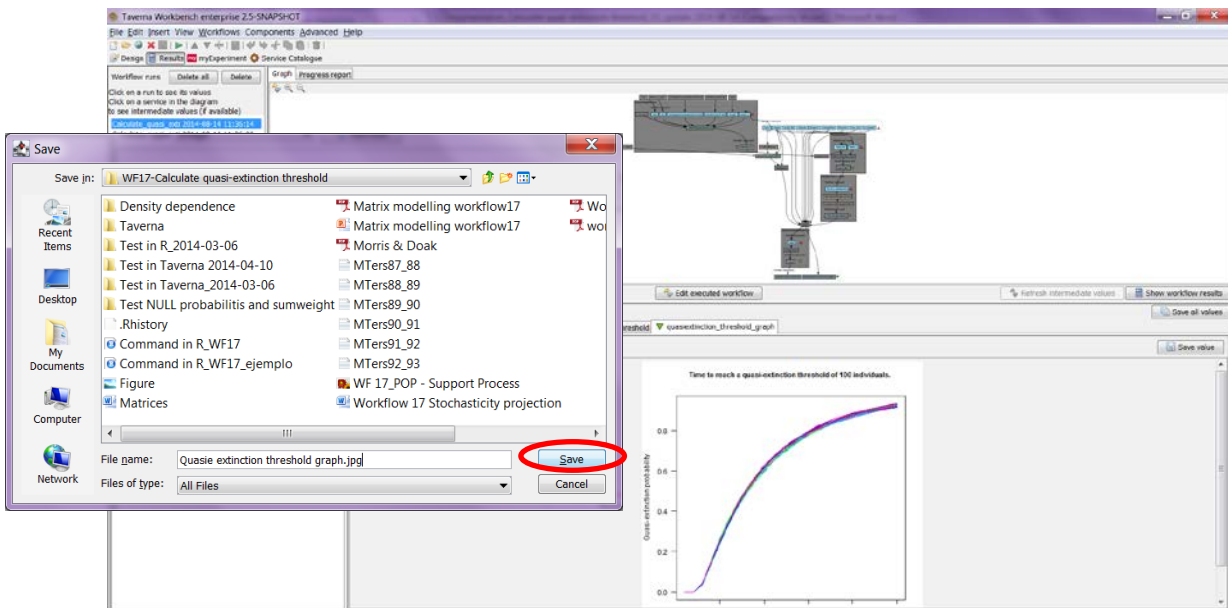


Figure 20. Name the file, click on Save value, name the file and add the extension for the file. For a text file: e.g. .csv or .txt. For images: e.g.: .jpg or png.

Calculate quasi-extinction threshold Workflow

1) Quasi-extinction threshold numeric results

This text has been adapted to the *Gentiana* example following Morris and Doak (2002). The Figure 19 and 20 are the numeric and graphic results of 50 separate runs of this workflow, each with 5,000 separate realizations of population growth and a quasi-extinction threshold of 100. Figure 20 shows little variation among the separate runs, indicating that in this example, 5,000 realizations provide a reasonably good estimate of the extinction time CDF (Morris and Doak 2002, advocate increasing the number of realizations per run until separate runs yield very similar estimates for the CDF). Starting with a population density of 344 individual plants, the probability of hitting a threshold density of 50 individual plants reaches a value of 0.1 after only about 7.5 years, and exceeds 0.8 by year 25 (see Morris and Doak 2002).

In this tutorial, it is computed the probability that the total density over all classes except seedlings (S) falls below 50. In this example the quasi-extinction threshold is applied to a subset of the stages (J, V, G and D). For this purpose, the sumweights were [0 1 1 1 1 1] (see input ports example **Sumweight** ignoring seedlings (S) when summing the densities across classes to compare to the quasi-extinction threshold (Morris and Doak 2002).

The entries in the **Initial population vector** (see interactions) represent densities, but quasi-extinction thresholds are typically expressed in terms of numbers of individuals. Thus when it is asked whether the summed densities across all or a subset of the classes has fallen below the quasi-extinction threshold, first it is necessary to be sure that a numerical threshold has been converted to a density. For example, let's assume that the densities in the population vector represent mean numbers of individuals per hectare, that the population occupies 100 hectares, and that we have decided to set the quasiextinction threshold at a total of 50 individuals. One hundred reproductive individuals in the entire population represents an average density of 0.5 individuals per hectare summed across the J, V, G and D stages, so it would be considered the population to have hit the threshold if this sum hits 0.5, not 50. Alternatively, it could be multiplied the starting densities by 100 to arrive at an estimate of total numbers and set the quasiextinction threshold to 50 (Morris and Doak 2002) (Fig 21).

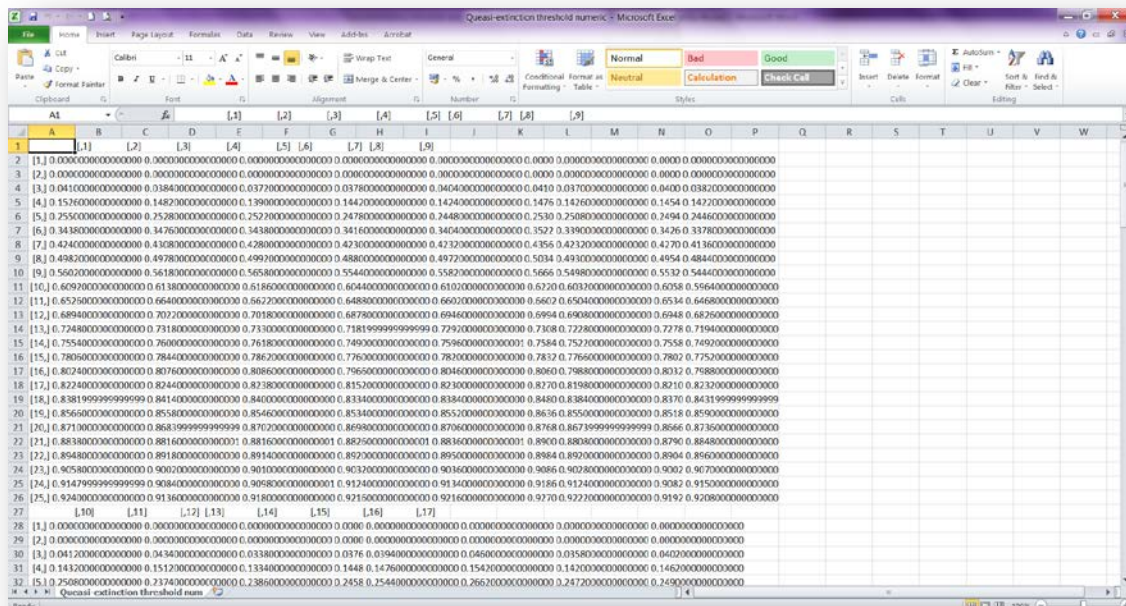


Figure 21. Quasi-extinction threshold numeric results (.csv or text files).

Calculate quasi-extinction threshold Workflow

2) **Quasi-extinction threshold graph:** Plots the numeric results of the Quasi-extinction threshold numeric results (Fig 22).

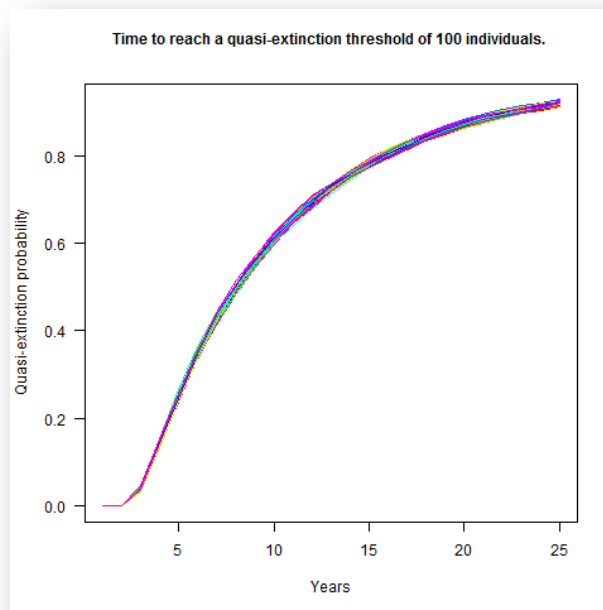


Figure 22. Quasi-extinction threshold graph (.png or .jpg).

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.

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

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

Calculate quasi-extinction threshold 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