

BioVeL – Biodiversity Virtual e-Laboratory

Workflow Documentation

Parametric Bootstrap or Resample a projection matrix 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 <u>http://www.biovel.eu</u> 36 months



1 Description

The Parametric Bootstrap or Resample a projection matrix Workflow provides an environment to resample a projection matrix using a multinomial distribution for transitions and a log normal distribution for fertilities (Stubben, Milligan, and Nantel. 2011). The resample is based on number of plants surveyed. The projection matrix A is first split into separate transition and fertility matrices. Dead fates are added to the transition matrix and the columns are then sampled from a Multinomial distribution based on the size in each corresponding stage class in n. The fertility rates are sample from a Log Normal distribution using the lnorms function. The same variance is applied to all rates by default. (Stubben, Milligan and Nantel 2013, Caswell 2001 see section 12.1.5.2).

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: *Confidence interval of Lambda*).

Analyses:

- Lambda (λ)
- Mean matrix
- Variance matrix
- Histogram
- Confidence interval of Lambda
- X= List of resampled matrices

2 General

2.1 Name of the workflow in myExperiment.

Name: Parametric Bootstrap Workflow for local execution

Download: The pack can be downloaded at myExperiment under the following address:http://www.myexperiment.org/packs/643.htmlortheworkflow:http://www.myexperiment.org/workflows/4309.htmlortheworkflow:

2.2 Date, version and licensing

Last updated: 25th July 2014

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:

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 Gerard Oostermeijer unpublished results and publication: Oostermeijer, J.G.B. M.L. Brugman, E.R. de Boer; H.C.M. Den Nijs. 1996.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). Parametric Bootstrap 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, Lambda (λ), Confidence intervals of λ , Parametric Bootstrap or Resample a projection matrix, mean matrix, variance matrix.

3.2 Scientific workflow description

The aim of the Parametric Bootstrap or Resample a projection matrix Workflow is to resample a projection stage matrix using a multinomial distribution for transitions and a log normal distribution for fertilities (Stubben, Milligan, and Nantel. 2011). The workflow accepts input data in a .txt format. The output is provided as a set of R results and a graphic plot.



Figure 1. Parametric Bootstrap 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. The simplest way to install a Taverna Engine is to install Taverna Workbench. The workflow also requires an Rserve installation with 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 Dependencies

- Install R software in your computer. See: <u>http://www.r-project.org/</u>
- Start R, and install package Rserve:
 - install.packages("Rserve")
- Install package popbio
 - install.packages("popbio")
- Local R Server: (Rserve) running at port 6311. See <u>https://wiki.biovel.eu/x/3ICD</u> for additional information.

4.2.3 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...

 $\label{eq:c:PROGRA~1\R\-30~1.1\library\Rserve\libs\x64\Rserve.exe} "C:\PROGRA~1\R\-30~1.1\library\Rserve\libs\x64\Rserve.exe"$

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 MTers87_88.txt, called in myexperiment.org: Stage Matrix of *Gentiana pneumonanthe* 1987-88 File. This file will be used to run this workflow. Finally you will be able to export your results to different formats.

5.2 Input data.

5.2.1 Data preparation/format

The workflow accepts input data (matrices) in a .txt format, all decimal numbers in the matrix must be indicated by dots e.g.: 0.578. The example matrix for the tutorial is available in: http://www.myexperiment.org/packs/643.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
 - <u>MTers87_88.txt</u>

Example from:

J. Gerard B. Oostermeijer; 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. *Journal of Ecology*, Vol. 84(2): 153-166.

5.3 Select input data dialogue boxes.

The first step is to fill out the input ports:

5.3.1 INPUTPORTS

1) *Initial abundance*: In this dialogue appear the fields with the initial abundance per stage observed in the field (Fig 2). As an example *Gentiana pneumonanthe* has 5 stages with its respective abundance.

The abundance of the stages or categories must be added one by one. First press Add value, fill the abundance of the first stage and press enter; then press Add value and fill once again the next stage abundance; repeat the action until you have fill all the abundance of all stages (Fig 2).

Stage abundance of the year 1987:

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



Figure 2. Abundance input. Fill in the abundance per stage appears.

2) <u>Itinirations</u>: Number of iterations for calculation of the resample analysis of the parametric Bootstrap. Click in Set Value (red arrow in Fig 3), tip the number in the right window.

e.g.: 1000

Input values for 'Parametric Bootstrap or F	lesample a projection matrix Workflow' from C:\Users\mbalcaz1\Documents\W 😑 💷 🗾
Input values for 'Parametric Bootstrap or F Diagram Workflow description Workflow description The Parametric Bootstrap or Resample a projection matrix low first and a log normal distribution for transitions and a log normal distribution for transitic a log normal distributions and a log normal distribution for	tesample a projection matrix Workflow' from C:\Users\mbalcaz1\Documents\W abundances iterations plotTitle stageMatrixFile stages Port description tininizations Number of iterations for calculation of the resample analysis of the parametric Bootstrap. Click in Set Value. to the number in the right window. Example value 10000 10000
	Drag to re-arrange, or drag files, URLs, or text to add
	Caller all use examples with worknow a caller

Figure 3. A number of itineration's to be filled in.

3) *plotTitle*: Descriptive title for labelling generated output graph. Click in Set Value (red arrow in Fig 4), tip the text in the right window.

e.g.: Gentiana pneumonanthe, Terschelling (species taxonomic name, place where the research was conducted)

Diagraffi	•	abundances iterations plotTitle stageMatrixFile stages
	• III	Port description Plot title Descriptive title for labelling generated output graph. Click in Set Value, tip the text in the right window. e.a.: Gentiana pneumonanthe. Terschelling (species taxonomic name. place where the research was Example value Gentiana pneumonanthe. Set Value, the text in the right window.
Workflow description he Parametric Bootstrap or Resample a rojection matrix Workflow provides an nvironment to resample a projection matrix sing a multinomial distribution for transitions nd a log normal distribution for fertilities Stubben, Milligan, and Nantel. 2011). The sample is based on number of plants urveyed. The projection matrix A is first split to separate transition and fertility matrices. Workflow author Iaria Paula Balcázar-Vargas, Jonathan Giddy a ierard Oostermeijer		Delete all ^o Set value Set URL Set gention a p Gention a p
		Added new value. Edit value on right.

Figure 4. Text to be filled in.

4) **<u>StageMatrixFile</u>**: Matrix in a .txt file format. Click in Set file location (red arrow in Fig 5), a window dialog appears and the user selects the file (e.g. <u>MTers87_88.txt</u>, Fig 6) and then clicks the Open button. Then the link to the file will appear on the Taverna window.

Diagram	abundances iterations plotTitle stageMatrixFile stages							
	Port description The stage matrix file input port: Here comes the stage matrix without the stage names (as you see in the example). It should be provied as a bt-file.							
Workflow description The Parametric Bootstrap or Resample a projection matrix Workflow provides an environment to resample a projection matrix using a multinomial distribution for transitions and a log normal distribution for fertilities (Stubben, Milligan, and Nantel. 2011). The resample is based on number of plants surveyed. The projection matrix A is first split into separate transition and fertility matrices. Workflow author Maria Paula Balcázar-Vargas, Jonathan Giddy and Gerard Oostermeijer	Example value 0.0000							
	Deleted node							
	↔ Load previous values ↔ Save values (1) Help 🐗 Use examples 🕨 Run workflow 💢 Cance	el						

Figure 5. Set the Location of the matrix file.

	Diagram Diagram Diagram	stageMatrixFile stages
Choose files	or directory	
Look in:	📜 WF11-Boot Transitions-matrix 🔹 🦻 🖽 -	
e.	L Taverna	e stage names (as you see in the example). It should be provied as a
Recent	R C:\Users\mbalcaz1\Documents\Workflows\Matrix models (popbio, diagram, popdem)	
Items	Rhistory	
	Comandos resample_hudsonia	7.6660 0.0000
	Comandos workflow 11	0.2857 0.8604
Desktop	Comandos workflow 11_resample	astion (A) Cot UDI
	Matrix modelling workflow11	Lation Set OKL
My	Matrix modelling workflow11	n a construction of the co
Documents	E MTers87_88	
	Workflow 11 resample	
	TWorkflow 11 resample	
Computer	4	
	File name: MTers87_88.bd Qpen	
Network	Files of type: All Files Cancer	
_		
	Deleted node	
		a b b b b b b b b b b

Figure 6. Choose the matrix file and click in Open.

5) **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 7), 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.

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 7. Stages to be filled in.

After the user has filled out the input ports and has clicked on the *Run Workflow button*, to complete

all the analysis may take several minutes depend on the number of itinerations.

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. **confidence_interval** (green oval in Fig 8). Second, click in the left window on value 1 (blue oval), Third, on the right window click on the save value (red 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 9), Finale, save the file in the chosen map.

Norkfox runs Deleta all Delete	Graph Progress report		
Sict os a revis to servise. Sict os a servicos de cagara o ser terrenciados volves (f.analóbe) mentode a terre y 2014 (f.a. 600) Boltetray, travisten 2014 (f.e. 64) (3):45:30 Boltetray, travisten 2014 (f.e. 64) (3):45:30			
Workflow results	v Freehed III Found Stream	Entropy Technology (State of the State of th	Ty Raheak sterrinsders volves Stow workflow results
Ad in the to vige value	2.08 1764 2.08 29.16 0.0617621.070122	 Spikefeeb Wrap sock 	in the obs

Figure 8. Confidence interval output window.

Save Save Save Save Computer Recent Recent recents workflows/lares	- 6 ×
www.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww	
Save Save Save Save I Tavena C.Users/nable/attal/Documents/Workflows/Matrix models (popbio, diagram, popdem, Ristory) Desktop Desktop Desktop Comandos resample_hudsonia Comandos resample_hudsonia Comandos resample_hudsonia Comandos resample_hudsonia Comandos resample_hudsonia Comandos resample_hudsonia Comandos workflow 11 Matrix modelling workflow11 Matrix modelling workflow11 Matrix modelling workflow11 Mitres7,88 Workflow 11 resample Mitres7,88 Workflow 11 resample Workflow 11 resample Workflow 11 resample Workflow 11 resample Mitres7,88 Workflow 11 resample Mitres7,88 Workflow 11 resample Workflow 11 resample Mitres7,88 Workflow 11 resample	
Save Save	
Save Save In: WF11-Boot Transitions-matrix Recent Recent Recent Desktop Documents Workflow 11 C Comandos workflow 11 Matrix modelling workflow 11 File game: Confidence_interval.bt Files of type: All Files	
Save jr: WF11-8oot Transitions-matrix Image: Computer Structure S	
Intakema Recent Recent Recent Restor O Comandos resample_hudsonia O Comandos workflow 11 Matrix modelling workflow11 File game: Confidence_interval.bd File game: Confidence_interval.bd Files Contact	
Recent Items R C\Users\mbalca21\Documents\Workflows\Matrix models (popbio, diagram, popdem Rhistory Beskop Comandos resample_hudsonia Comandos workflow 11 Comandos workflow 11 Comandos workflow 11 Matrix modelling workflow11 Matrix modelling workflow11 Matrix modelling workflow11 Matrix modelling workflow11 Matrix modelling workflow11 Workflow 11 resample Workflow 11 resample Workflow 11 resample Workflow 11 resample Workflow 11 resample Image: Confidence_interval.txt File game: Confidence_interval.txt Files of type: All Files	
Lems LKNStory © Comandos resample_hudsonia © Comandos workflow 11 © Comandos workflow 11 © Comandos workflow 11. © Comandos workflow 11. © Matrix modelling workflow11 Matrix modelling workflow11 Matrix modelling workflow11 Matrix modelling workflow11 Matrix modelling workflow11 Workflow 11 resample © Workflow 11 resample Workflow 11 resample © Workflow 11 resample Workflow 11 resample © File game: Computer III File game: Confidence_interval.bd Files of type: All Files	
^O Comandos resample_hudsonia Desktop ^O Comandos workflow 11 ^M Matrix modelling workflow 11 ^M Matrix modelling workflow 11 ^M Matrix modelling workflow 11 ^M Merss8_28 ^W Workflow 11 resample ^W Workflow 11 resample ^M Workflow 11 resample ^M File game: Confidence_interval.bd File of type: All Files	
Deskop © Comandos workflow 11 Deskop © Comandos workflow 11_resample Matrix modelling workflow 11 Matrix modelling workflow 11 My Matrix modelling workflow 11 Motification Matrix modelling workflow 11 Workflow 11 resample Workflow 11 resample Workflow 11 resample Workflow 11 resample Workflow 11 resample Ele game: Computer III File game: Confidence_interval.btd Files of type: All Files	Show workflow results
Desktop © Comandos workflow 11_resample Matrix modelling workflow 11 Matrix modelling workflow 11 Matrix modelling workflow 11 Matrix modelling workflow 11 Workflow 11 resample Workflow 11 resample Workflow 11 resample Workflow 11 resample Image: State 1	Save al values
Image: Second	
My Matrix modelling workflow11 Documents Workflow 11 resample Workflow 11 resample Workflow 11 resample Workflow 11 resample III Image: Confidence_interval.bt Save Files of type: All Files	Save value
Motor MTers87_88 Workflow 11 resample Workflow 11 resample Workflow 11 resample Workflow 11 resample Image: Second seco	
Workflow 11 resample Workflow 11 resample Workflow 11 resample Computer File game: Confidence_interval.bd Files of type: All Files	
Workflow 11 resample Computer Network File game: Confidence_interval.bt Save Files of type: All Files	
Computer Ketwork Network Files of type: All Files	
Image: Confidence_interval.bd Network File game: Confidence_interval.bd Save	
File name: Confidence_interval.bt Network Files of type: All Files	
Network Files of type: All Files	

Figure 9. Name the file and determinate the extention file. For a text file: e.g. .csv or .txt. For images: e.g.: .jpg or png.

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%
	0.9617	1.4765

2) *Cl_histogram:* Histogram plotting the frequencies of the lambda values and the 95% confidence intervals resulting from the bootstrap analysis (Fig 10).



Figure 10. The confidence interval histogram opens as a graph.

3) Lambda (λ): Lambda or dominant eigenvalue, this value describes the population growth rate of a stage matrix. The population will be stable, grow or decrease at a rate given by lambda: e.g.: $\lambda = 1$ (population is stable), $\lambda > 1$ (population is growing) and finally $\lambda < 1$ (population is decreasing).

e.g.: 1.2375

4) *Mean matrix*: the mean matrix is calculated based on the matrices generated by bootstrap analysis and the number of the matrices at the same time is based on the number of itinerations (Fig 11).

a1	From From Web Test	From Other Sources * Ci sternal Data	Existing Refr	Convections	ns 61 112 11 Sort	W Gear Filter S Factor S Adva	need Column	Remove E s Duplicates Valio Da	lata Consoldat Lation - ta Tools	e What-If Analysis -	Group Ungroup Subs	*] show Detail *] Hide Detail tail				
	A1	• (*	fa:													4
	A	В	С	D	E	F	G	Н	1	J	K	L	M	N	0	P.
l		S	1	V	G	D										
	[1,]	0	C	0 0	7,651887	0										
	[2,]	0,058174	0,010178	0	8,50966	0										
	[3,]	0,463201	0,828959	0,900923	0,286757	0,860774										
	[4,]	0	0,04028	0,008953	0,618605	0,115637										
	[5,]	0	0,03014	0,018196	0	0,023391										
1	H Mean m	natrix 🖉								14			él		Contract of the	× 1

Figure 11. The mean matrix opens in excel.

5) *Variance matrix*: the variance matrix is calculated based on the matrices generated by bootstrap analysis and the number of the matrices at the same time is based on the number of itinerations (Fig 11).

1.5	4	- 3	1.		-				-	Var	ance matri	- Micros	soft Excel		-	-			_	-	6 X
file	H	те	Insert Page	Layout F	Formula	s Data Re	view View	Add-In	s Acrobat					-				-		0	0 - 9 -
A	From s Web	From Test Get Ed	From Other Sources * C ternal Data	Existing onnections	Refres All •	Connection Properties h Edit Links Connections	1 1 1 1	Filter Sort & Fi	% Cear S Feacpi S Advano ter	Text to Column	Remove s Duplicate	Data Validation Data To	Consolidate 1*	What-If Analysis	Group	Chigroup Su Outli	*] show bet *] Hide Deta stotal	ii L			
	A1		• (*	fa .																	
	A		В	C		D	E		F	G	- F	1	1	J		К	L	м	N	0	P
			s	J		V	G	D													
	[1,]	1	C	1	0	0,00E+00	1,495111		0												
	[2,]		0,0008	0,0	001	0,00E+00	1,495924	l.	0												
	[3,]		0,003594	0,001	451	8,29E-04	0,009614	0,00	02744												
	[4,]		0	0,000	386	8,01E-05	0,010861	0,00	2366												
	[5,]		0	0,000	294	1,63E-04	0	0,00	00528												
		mance	e matrix A											14							
ŝ	0		CINCUIN CO											1.4	_				HINT	95 (·)	0 0

Figure 11. The variance matrix opens in excel.

6) **X= List of resampled matrices**: it is the list of matrices generated by bootstrap analysis and the number of the matrices at the same time is based on the number of itinerations..

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: <u>https://wiki.biovel.eu/display/BioVeL/Glossary</u>

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, 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.

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