

BioVeL – Biodiversity Virtual e-Laboratory

Workflow Documentation

Age specific survival workflow for local execution

June 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



Age specific survival workflow for local execution

1 Description

The age-specific survival analysis workflow gives the basic information on the mean, variance and coefficient of variation (cv) of the time spent in each stage class and the mean and variance of the time to death.

Analyses:

- Fundamental matrix (N)
- Variance (var)
- Coefficient of variation (CV)
- Meaneta
- Vareta

2 General

2.1 Name of the workflow in myExperiment.

Name: Age specific survival workflow for local execution.

2.2 Date, version and licensing

Last updated: 2th November 2012

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 Gerard Oostermeijer unpublished results and publication: Oostermeijer, J.G.B. M.L. Brugman, E.R. de Boer; H.C.M. Den Nijs. 1996.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 (<u>www.biovel.eu</u>). Age specific survival 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

Matrix Population Models, Age specific survival, Fundamental matrix (N), Variance (var), Coefficient of variation (cv), Meaneta, Vareta.

3.2 Scientific workflow description

The age-specific survival analysis workflow gives the basic information on age-specific survival; this includes the mean, variance and coefficient of variation (CV) of the time spent in each stage class and the mean and variance of the time to death.

The aim of the age-specific survival analysis workflow is to provide a connected environment to calculate the mean, variance and coefficient of variation (CV) of the time spent in each stage class and the mean and variance of the time to death. The workflow accepts input data in a .txt format. The output is provided as a set of R results.

- a) Fundamental matrix (N): is the mean of the time spent in each stage class.
- b) Variance (var): is the variance in the amount of time spent in each stage class.
- c) Coefficient of variation (CV): is the coefficient of variation of the time spent in each class (SD/mean- the ratio of the standard deviation to the mean).
- d) Meaneta: is the mean of time to death, of life expectancy of each stage.
- e) Vareta: is the variance of time to death.



Figure 1. Age-specific survival analysis 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 popdemo 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...

"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 will use an input file MTers87_88.txt, called in myexperiment.org: Stage Matrix of *Gentiana pneumonanthe* 1987-88 File. You will use this file to perform the age specific survival on it. 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 each matrix must be indicated by dots e.g.: 0.578. The example matrix for the tutorial is available in: http://www.myexperiment.org/files/1135.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

• MTers87 88.txt

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.

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) *columns fecundity*: the column(s) in which the fecundity values are found, should be selected.

In the example of the *Gentiana* species (Oostermeijer et al. 1996. *Journal of Ecology*): The selected column is G (reproductive individuals), the number 4 will be used to

	S	J	V	G	D
S	0	0	0	7,666	0
J	0,0579	0,01	0	8,5238	0
V	0,4637	0,83	0,9009	0,2857	0,8604
G	0	0,04	0,009	0,619	0,1162
D	0	0,03	0,018	0	0,0232

identify the fecundity column (G) (see matrix next to text).

The numbers of the reproductive columns must be added one by one (in the case that there is more than one reproductive column. In this example there is only one). First press add value, fill the number of the column and press enter, then press add value and fill once again the next reproductive column, repeat the action until you have fill all the numbers of all reproductive columns (Fig 2).

e.g.: 4

Diagram .	columns_fecundity rows_fecundity stage_matrix_file stages
	Port description To perform the age-specific survival analysis, the column(s) in which the fecundity values are found, should be selected. In the example of the Gentiana species (Oostermeijer et al. 1996. The Journal of Ecology): The selected column is G (reproductive individuals): therefore number 4 will be used to identify the fecundity
Workflow description The age-specific survival analysis workflow gives the basic information on age-specific survival, this includes the mean, variance and coefficient of variation (cv) of the time spent in each stage class and the mean and variance of the time to death. Fundamental matrix (N): is the mean of the time spent in each stage class. Workflow author Maria Paula Balcázar-Vargas, Jonathan Giddy and Gerard Oostermeijer	Example value 4 C Delete Add value Add URL List able C
	Added new value. Edit value on right.
	(3) Help aik Use examples ▶ Run workflow ¥ Cancel

Figure 2. The number(s) of the column(s) to be filled in.

2) *rows fecundity:* the row(s) in which the recruitment values are found, should be selected.

In the example of the *Gentiana* species (Oostermeijer et al. 1996. *Journal of Ecology*): Selected rows are S and J (seedlings and juveniles; these stages receive recruits each year from the stage G): therefore, the numbers 1 and 2 are used to identify these rows (see matrix next to the text).

	S	J	V	G	D
S	0	0	0	7,666	0
J	0,0579	0,01	0	8,5238	0
V	0,4637	0,83	0,9009	0,2857	0,8604
G	0	0,04	0,009	0,619	0,1162
D	0	0,03	0,018	0	0,0232

The numbers of the stages that receive recruits each year from the stage G must be added one by one (in the case that there are more than one recruit stage row). First press add value, fill the number of the row and press enter, then press add value and fill once again the next recruit stage row, repeat the action until you have fill all the numbers of all recruits rows (Fig. 3).

e.g.: 1 and 2



Figure 3. The number(s) of the recruit(s) row(s) to be filled in.

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

Input values for 'Age-specific survival an Diagram	columns_fecundity rows_fecundity stage_matrix_file stages									
	columns_recunality rows_recunality stage 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 provid bt-file.									
	Example value 0.0000 0.0000 0.0000 0.0000									
Workflow description The age-specific survival analysis workflow gives the basic information on age-specific survival, this includes the mean, variance and coefficient of variation (cv) of the time spent in each stage class and the mean and variance of the time to death. Fundamental matrix (N): is the mean of the time spent in each stage class. Workflow author Maria Paula Balcázar-Vargas, Jonathan Giddy and Gerard Oostermeijer	0.0579 0.0100 0.0000 0.55238 0.0000 0.4637 0.8300 0.0000 0.2857 0.8604	•								
	Deleted node									
	↔ Load previous values ↔ Save values (1) Help at Use examples > Run workflow X Can	cel								

Figure 4. Set the Location of the matrix file.

Diagram	Port description The stage matrix file input		u see in the example). It should be provide as a	-
Look jn: WF4-fundamental matrix or A Recent Test Deskop Deskop Wy Occuments Computer	v4 v4 natrix	7.6660 8.5238 0.2857 ration	0.0000 0.0000 0.8604	•
File name: MTers87_88.bdt Network Files of type: All Files		<u>Open</u>		
	<> Load previous values	<>> Save values (i) Hel	p 🞼 Use examples 🕨 Run workflow 💢 Can	cel

Figure 5. Choose the matrix file and click in Open.

4) **<u>stages</u>**: Here come the names of the stages or categories of the input matrix. No longer than 8 characters.

	S	J	V	G	D
S	0	0	0	7,666	0
J	0,0579	0,01	0	8,5238	0
V	0,4637	0,83	0,9009	0,2857	0,8604
G	0	0,04	0,009	0,619	0,1162
D	0	0,03	0,018	0	0,0232

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

e.g. the stages of this matrix are called:

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

The name of stages or categories must be added one by one. First press add value, fill the name of the stage no longer than 8 characters and press enter; then press add value and fill once again the next name of the stage; repeat the action until you have fill all the names of all stages (Fig 6).

Age specific survival workflow for local execution

Input values for 'Age-specific survival anal	ysis.' from C:\Users\mbalcaz1\Documents\Workflows\Matrix models (popbio, d🗖 💷 🗪	٢
Diagram	columns_fecundity rows_fecundity stage_matrix_file stages	
	Stage: Stage: Here come the names of the stages or categories of the input matrix. The name of the stages must be added one by one. First press add value, fill a stage name (not longer than 8 characters) and press enter, then press add value and fill once adain the next stage name. repeat the action until you have fill all the stages names. Example value S	•
Workflow description The age-specific survival analysis workflow gives the basic information on age-specific survival, this includes the mean, variance and	J V Velete av Add value Add file location @ Add URL	Ŧ
coefficient of variation (cv) of the time spent in each stage class and the mean and variance of the time to death.	List D 	
Fundamental matrix (N): is the mean of the time spent in each stage class.		
Workflow author Maria Paula Balcázar-Vargas, Jonathan Giddy and Gerard Oostermeijer		
	Added new value. Edit value on right.	
	Audeo new value: - cut value of right. ↔ Load previous values <> Save values ③ Help ab [®] Use examples ▶ Run workflow X Cancel	1

Figure 6. The name of the stages to be filled in.

After the user has filled out the input ports and has clicked the **Run Workflow button**, the workflow performs the analysis. To complete all the analysis may take few seconds.

When the analyses are completed, they appear on a different window 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 **Age_specific_survival** (Fig 7). 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. (Fig 8), Finale, save the file in the chosen map.

Taver a Workberch enterprise 2.5-SN De Edit Joseft View Workflows Comp De Station (Comp Design (Comp) (Comp) (Comp) People (Comp) (Comp)	anenti Advanced Brip 41 % E I B	
Non Tillow nov. Before all Online Color or not too too belook Data or a nortoo or the chapter and faith reaction of the chapter and faith reaction of the chapter of the ch	Doph International Violed ow nour ports Violed ow nour ports (columnal_fecundar) Integra (columnal_f	
	✓ Protect III more X tand	🐴 natioat internediate values 🛛 📓 shan variation results
Ock in tree to size values	date nath & A sage V As specific annual	See sale
4 11 for 2000-0000	Frag S S 6 C.2000000 C.20000000 C.20000000 C.2000000 C.20000000 C.20000000 C.20000000 C.20000000 C.20000000 C.200000000000000 C.2000000000000000000000000000000000000	

Figure 7. Age_specific_survival output window. Save value on the right window of the Age_specific_survival output results.

Taverna Workbench enterpris	# 2.5-SNAPSHOT	Management of the Owner, State of Street, or other		
ie Edit Insert View Workflow		d Belp		
Posign 😫 Result: 🖬 wyExperim	art O Sorara Catalogue		-	
and the second se	Delete Craph Progress	report	🛃 Save	
de on a nun to see its values de on a service in the diagram	444		Bare Sure	
ox on a centro in the chaptern see intermediate velves (if available		World ow input ports	Save in	: 🚺 WF4-fundamental matrix or Age specific survival 🔹 🍺 📂 🛄-
		columna_tecundty n		📜 Taverna
		-	Recent	I. Test
			Items	Rhistory
				Command in R_WF4
				Matrix modelling workflow4
			Desktop	Matrix modelling workflow4
	Firschoo II	Lincol X Cancel		MTers87_88
Neriflow results				
lick in tree to view volves	Canoda 🛡 opple-unitarilitie	dages ▼ Apr.peofc_trivial Volue type Test volue type Test	My	MTers87_88
• (78100 T			Documents	Workflow 4 fundamental.matrix
		Proc 0 V 6 0 S 0.10001000 0.00000000 0.00000000 0.00000000 0.000000000000 0.0000000000000000 0.0000	Computer	r Workflow 4 fundamental.matrix
		ACV C V B B 5 5.050200 5.00000 Field Bodi Set 4 4.055120 5.100000 Field Bodi Bodi Set 9 4.055120 5.100000 Field Bodi Bodi Set 9 4.055120 5.0402000 5.041241 1.2431435 6.0402324 5.0402345	Network	File game: Age_specific_survival.cs/
		E 3.253333 2.24744E 2.2797535 2.4852479 0.4559435 Aumaintia		Files of type: All Files
		8.202203 25.72022 12.99900 23.124054 14.300234	<u> </u>	
		Praneta 3 V G E 149,803 191.0125 101.4011 197,850 101.0129		-

Figure 8. Name the file and determinate the extension file. For a text file: e.g. .csv or .txt.

Age specific survival

The age specific survival workflow gives the basic information on age-specific survival (Fig 9):

11) **Fundamental matrix (N):** is the mean of the time spent in each stage class. e.g.: For our *Gentiana* example means a J plants will spends, on average, about 1 year as a Juvenile, 11 as vegetative plant, less than a year a reproductive and dormant plant.

12) Variance (var): is the variance in the amount of time spent in each stage class.

13) *Coefficient of variation (CV):* is the coefficient of variation of the time spent in each class (SD/mean-the ratio of the standard deviation to the mean).

14) *Meaneta:* is the mean of time to death, of life expectancy of each stage. e.g. The mean age at death is the life expectancy; the life expectancy of a new individual seedling is 8 years.

15) *Vareta:* is the variance of time to death.

	Insert									maunacsy	 Microsoft E 	xcei	-			-				
											3	100	•na •na	.000 - 9	Show Detail					
				Prope	erties 24	ZA	C Rea	oply	5 10	-ŏ	.	1	*1	-						
n From	From From	Other Exis	ting Refr	esh se Edit L	inks 🕌	Sort F	ilter y Adv	anced Text	to Remove	Data	Consolidate	What-If	Group Ungrou	p Subtotal						
	Get External D	ces * Conni lata	ections Al	Connection		Sort	& Filter	Colu	mns Duplicate	Data Tool	A A	unalysis *		Outline	6					
J25		- (°		Contraction.	з Д.	5011	Section 2	*		0010 100										
	В			F	F	G	н	1	1	К	L	М	N	0	P	0	R	s	Т	11
SN C	0		U	-		0				IN.				U		ų		3		0
	S	J	v	G	D															
	1				0															
		1,010101		0																
1	6,890559	11,99681	13,35817	10,01862	12,9606															
5	0,208448	0,466788	0,391118	2,918338	0,691977															
C				0,184814																
var																				
			V	G	D															
				0 0	0															
		0,010203			0															
1				157,2694																
5				5,598359																
)	0,18007	0,321332	0,31526	0,247831	0,332007															
cv	-			-	-															
		J			D															
		NaN			NaN															
		0,1		NaN 1,25174	NaN															
				0,810765																
)				2,693662																
,	5,251050	2,240300	2,270505	2,033002	0,450254															
meanet	a									1										
	s	J	v	G	D					•										
	8,286472																			
vareta																				
	S	J	v	G	D															
	143,4207	181,1844	181,6537	177,2333	181,2319															
		natrix ⁄ 🕽											4							

Figure 9. Age specific survival opens in excel.

For further details see: References

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) and popdemo (Stott, Hodgson and Townley, 2013)

- **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.
- **Stott, I., D.J. Hodgson and S. Townley**. 2013. popdemo: Provides Tools For Demographic Modelling Using Projection Matrices. Version 0.1-3.
- **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