COMPADRE
Plant Matrix Database

MAX-PLANCK-INSTITUT FÜR DEMOGRAFISCHE FORSCHUNG
MAX PLANCK INSTITUTE FOR DEMOGRAPHIC RESEARCH

COMPADRE
Plant Matrix Database

A user's guide to

COMPADRE

THE PLANT MATRIX DATABASE

Date modified: November 10th 2014

Contact: compadre-contact@demogr.mpg.de

Contents

General Instructions

Database Organization

The data associated with COMPADRE are provided in a single *R* data object (extension *.Rdata*) file and a *Nexus* file (extension *.nex*). In addition, these files are accompanied by *R* scripts available in the Supplementary Information of the manuscript introducing COMPADRE, and in our github repository (https://github.com/jonesor/compadreDB).

- ***COMPADRE_Data_MONTH_DAY_YEAR_version_XX.RData***: Contains basic information regarding the source of publication, as well as ecological, biogeographic, and taxonomic details of the demographic study for each study species, the demographic information (i.e. the matrix population model) and metadata.

- ***COMPADRE_Phylogeny_MONTH_DAY_YEAR.nex***: A standard Nexus format file containing a phylogenetic tree for the species in COMPADRE.

Disclaimer

- The COMPADRE digitization team does its best to ensure data accuracy, and every piece of information goes through multiple error-checks prior to its release in www.compadre-db.org. However, we claim no responsibility for any damage that may arise from using COMPADRE. A list of error checks and potential issues in the use and interpretation of the database are described in the main manuscript. The end user is ultimately responsible for his/her interpretations of the data.

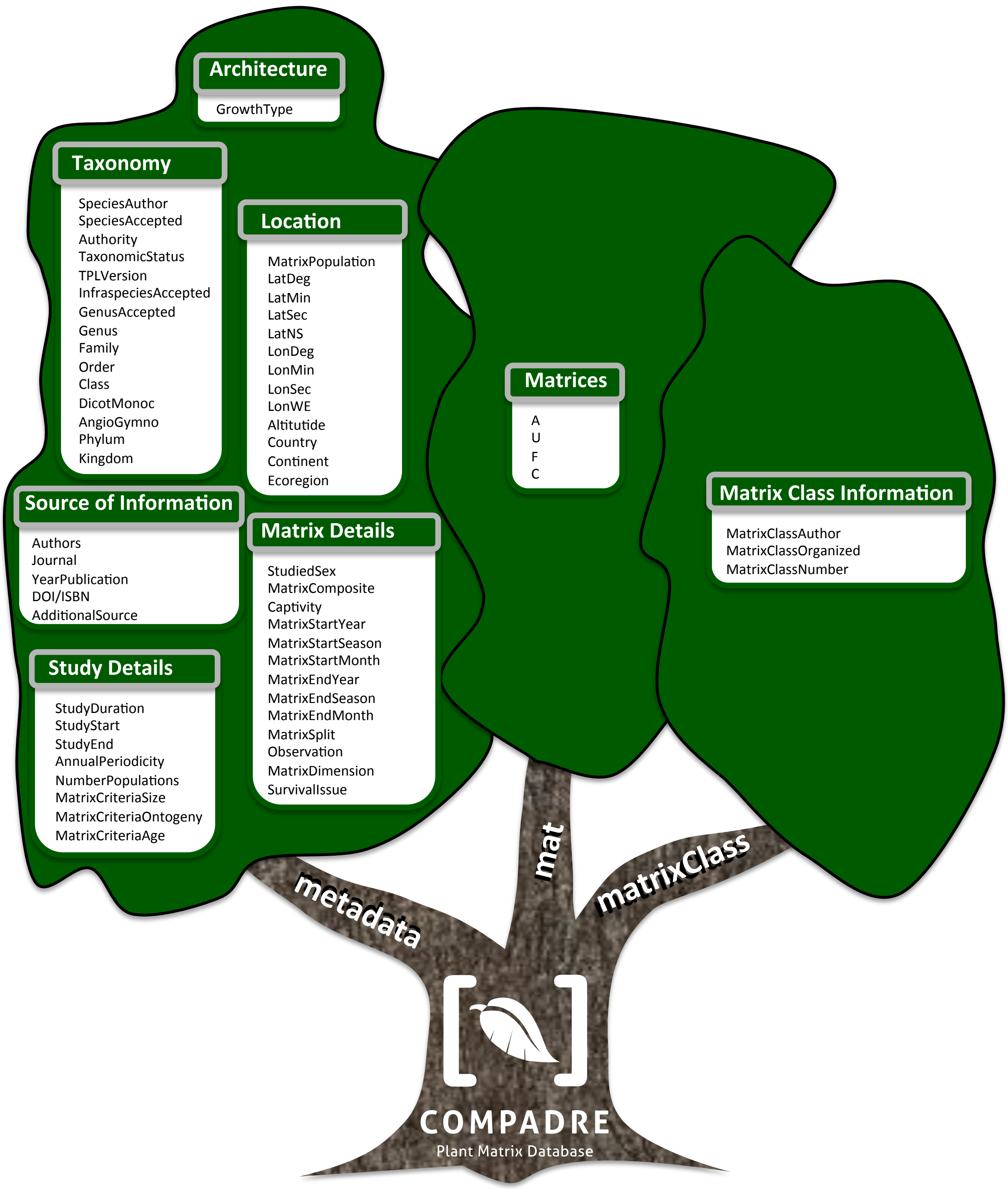
The meanings of NA in COMPADRE

NA in the COMPADRE data generally means that the data are not available/applicable. This could be because the data were not reported by the author(s), or it could mean that the data have not yet been digitized for this version of COMPADRE. It could also mean that the data are not applicable in this case (e.g. where data are derived from several geographic locations we do not report a single average latitude/longitude value).

COMPADRE_Data_MONTH_DAY_YEAR_version_X.X.RData

The following table below contains definitions of the variables digitized in the *COMPADRE_Data_MONTH_DAY_YEAR_version_X.X.*RData and, where appropriate, information about the associated values. Each variable is associated with one of seven categories, as described in figure 1.

Figure 1. Name of variables archived in COMPADRE, organized according to the general category to which they are associated. The R data object "compadre" is a list with three data.frames ("metadata", "mat" and "matrixClass"), here represented by each branch of the tree. The canopy contains the variables in each data.frame, organized by general categories.



Variable	Variable Description	Possible values	Description of values and relevant notes
SpeciesAuthor	Latin name of the species as indicated in the publication; this may not be the same as the currently accepted name (See *SpeciesAccepted* below)	*Genus_species*	This variable always includes the genus and the species names and, where appropriate, information on variety and subspecies (see below). If two or more demographic studies exist for a single species, these are digitized with a numerical suffix after *Genus_species* starting with “2” (e.g. “Rosa_sempervirens”, “Rosa_sempervirens_2”, “Rosa_sempervirens_3”).
		<…>var.<…>	Where the publication refers to an intraspecific variety, this is indicated with "var." followed by the variety name, e.g. "*Pityopsis_aspera_var._aspera*"
		<…>subsp.<…>	Where the study refers to an intraspecific subspecies, this is indicated with "subsp." followed by the subspecies name, e.g. "*Anthyllis_vulneraria_subsp._alpicola*"
SpeciesAccepted	Currently accepted latin name. This information is is obtained from [The Plant List]	*<Genus_species>*	e.g. Chorizanthe
Authority	Taxonomic authority of *SpeciesAccepted*, as per [The Plant List]	<Authority>	e.g. Benth.
TaxonomicStatus	Taxonomic correspondence of	NoMatch	Species is a plant, but it does not match with the data repository of The Plant List.

	SpeciesAuthor with *SpeciesAccepted*	Unresolved	No match because species is not a plant (this is always the case with brown algae, Kingdom Chromalveolata)
		Accepted	*SpeciesAuthor* is identical to *SpeciesAccepted*
		Synonym	*SpeciesAuthor* is synonymous of *SpeciesAccepted*
TPLVersion	Version of The Plant List used for taxonomic validation	1.0	Version 1.0
InfraspeciesAccepted	Taxonomic infraspecific name of study *SpeciesAccepted*, as per The Plant List	<infraspecies>	e.g. pungens
SpeciesEpithetAccepted	Taxonomic species epithet name of study *SpeciesAccepted*, as per The Plant List	<epithet>	e.g. pungens
GenusAccepted	Taxonomic genus name of study *SpeciesAccepted*, as per The Plant List	<genus>	e.g. Chorizanthe
Genus	Genus used in *SpeciesAuthor*	<genus>	e.g. Chorizanthe
Family	Family to which species belongs	<Family>	e.g. Polygonaceae
Order	Order to which species belongs	<Order>	e.g. Caryophyllales

Class	Class to which species belongs	<Class>	e.g. Magnoliopsida
DicotMonoc	Whether species is a dicot or monocot	Eudicot	Species is a dicot
		Monocot	Species is a monocot
		NA	Species is neither a dicot nor a monocot (not an angiosperm)
AngioGymno	Whether species is an angiosperm or a gymnosperm	Angiosperm	Species is an angiosperm
		Gymnosperm	Species is a gymnosperm
		NA	Neither angiosperm nor gymnosperm
Phylum	Phylum to which species belongs	<Phylum>	e.g. Magnoliophyta
Kingdom	Kingdom to which species belongs	<Kingdom>	e.g. Plantae
GrowthType	General plant/algae type, based mainly on architectural organization	Algae	The species was assigned to one of these possible values using the description of plant growth type provided by the author and other external sources (e.g. other publications) *Note about the Annual category:* Typically plant ecologists also refer to pseudoannual species as being "biennials". The main difference between the annuals and biennials is that annual species complete their lifecycle (are born, grow, reproduce and die) within a single year, whereas biennials have the
		Annual	
		Bryophyte	
		Epiphyte	
		Fern	
		Herbaceous perennial	
		Liana	
		Palm	

		Shrub	potential to stretch that time-window up to two years. Both annuals and biennials are often associated with periodic (or seasonal) matrices. Biennial (and pseudoannual) species are categorized as “Annual” in COMPADRE
		Succulent	
		Tree	
Authors	Last (family) name of all authors	<name>	Separated with “; ” e.g. "Smith; Jones"
Journal	The document from which data were sourced	<code>	Where the data come from a scientific journal article, the abbreviated journal name is given. We use the standard abbreviation of the journal using the BIOSIS Format
		Book	Matrices are from a book, or book chapter
		PhD thesis	Matrices are from a doctoral thesis
		MSc thesis	Matrices are from a masters thesis
		Report	Matrices are from a report
		Conference talk	Matrices reported in a conference talk
		Conference poster	Matrices reported in a conference poster
YearPublication	Year of publication	<yyyy>	e.g. 2012
DOI/ISBN	Codes that identifies the reference	<DOI> or <ISBN>	***DOI:*** “Digital object identifier” – a unique code that identifies an electronic document. ***ISBN***: “International Standard Book Number” – a unique code that identifies books
AdditionalSource	Additional source used to reconstruct the matrix, or to obtain additional information on ancillary data	<Lastname Journal YearPublication>	e.g. Godinez-Alvarez Bot Rev 2003
StudyDuration	The number of years of observation	<xx>	This is calculated as (“StudyEnd” year - “StudyStart” year) + 1 (see below). The calculation thus overlooks any missing years in the middle of the study period

StudyStart	First year of study	<yyyy>	e.g. 2012
StudyEnd	Last year of study	<yyyy>	e.g. 2013
AnnualPeriodicity	Frequency with which seasonal or annual the matrix population model was constructed	1	Once a year
		2	Twice per year
		0.2	Once every five years
		etc	
NumberPopulations	Number of populations of the study species.	<xx>	This variable refers to the number of populations as defined by the author. Within-site replication of permanent plots are not considered to be different populations (see below)
MatrixCriteriaSize	Indicates whether the matrix population model contains stages based on size, and if so, what variable of size was measured	<criteria>	The matrix population model contains stages based on size and this briefly describes the aspect of size that was measured to classify individuals (e.g. DBH, stem length, stem height, stem number, etc)
		No	The matrix population model is not based on size
MatrixCriteriaOntogeny	Indicates whether the matrix population model contains stages based on development	Yes	The matrix population model contains at least one class that is based on development/ontogeny (e.g. seedbank, reproductive, vegetative, dormant, etc)
		No	Matrix population model is not based on ontogeny
		Based on size	Where the authors report ontogenetic stage class(es) but the stage class(es) themselves are based on size criteria (e.g. vegetative individuals defined as those <1 m height)
MatrixCriteriaAge	Indicates whether the matrix population model contains any stage based on age	Yes	The matrix population model contains at least one class that is based on age (e.g. individuals of 0 year old)
		No	The matrix population model is not based on age

		Based on size	If the stage classes are in reality based on size criteria (e.g. old individuals defined as those with >10 stems)
MatrixPopulation	Name of population where matrix was recorded, usually as given by the publication author	<…..>	For example “Brazeau Creek, Florida”. When the author provides no name, the closest geographic location is assigned as name. If there are multiple populations in the study and their names are not pertinent/available, sequential names in alphabetical order are assigned for each population in the study (e.g. “A”, “B”, “C”, etc)
LatDeg	Degree value of latitude	<0 to 90>	
LatMin	Minute value of latitude	<0 to 60>	
LatSec	Seconds value of latitude (decimates are possible).	<0 to 60>	Note: The value “0” may mean precise values are not available
LatNS	Cardinal direction	N	North
		S	South
LonDeg	Degree value of longitude	<0 to 180>	
LonMin	Minute value of longitude	<0 to 60>	
LonSec	Seconds value of longitude (decimates are possible)	<0 to 60>	Note: The value “0” may mean precise values are not available
LonWE	Cardinal direction	W	West
		E	East
Altitude	Altitude of studied population	<xxxx>	Height above sea level of specific population in meters
Country	Country/ies where study took place	ISO 3 Country Code	If the study involves multiple countries, these are separated by "; ". We use standard codes

			from the ISO 3 list
Continent	Continent where study took place	Africa	
		Asia	
		Europe	
		N America	Includes Canada, USA & Mexico
		S America	Countries in the Americas except Canada, USA and Mexico
		Oceania	Various definitions for Oceania exist, but here we opted for this one: http://en.wikipedia.org/wiki/List_of_Oceanian_countries_by_population
Ecoregion	Description of the Ecoregion for the study, using the categories described in Figure 1 of *Olson et al.* (2001). For a more inclusive description of water ecoregions, see http://worldwildlife.org/biomes	TMB	Terrestrial - tropical and subtropical moist broadleaf forests
		TDB	Terrestrial - tropical and subtropical dry broadleaf forests
		TSC	Terrestrial - tropical and subtropical coniferous forests
		TBM	Terrestrial - temperate broadleaf and mixed forests
		TCF	Terrestrial - temperate coniferous forests
		BOR	Terrestrial - boreal forests/ taiga
		TGV	Terrestrial - tropical and subtropical grasslands, savannas and shrublands
		TGS	Terrestrial - temperate grasslands, savannas, and shrublands
		FGS	Terrestrial - flooded grasslands and savannas
		MON	Terrestrial - montane grasslands and shrublands
		TUN	Terrestrial – tundra
		MED	Terrestrial - Mediterranean forests, woodlands and scrubs

		DES	Terrestrial - deserts and xeric shrublands
		MAN	Terrestrial – mangroves
		LRE	Freshwater - large river ecosystems
		LRH	Freshwater - large river headwater ecosystems
		LRD	Freshwater - large river delta ecosystems
		SRD	Freshwater - small river delta ecosystems
		SLE	Freshwater - small lake ecosystems
		LLE	Freshwater - large lake ecosystems
		XBE	Freshwater - xeric basin ecosystems
		POE	Marine - polar ecosystems
		TSS	Marine - temperate shelf and seas ecosystems
		TEU	Marine - temperate upwellings
		TRU	Marine -tropical upwellings
		TRC	Marine - tropical coral
		LAB	Laboratory or greenhouse conditions – controlled, usually indoor, conditions that mean the study species is not affected by the environment conditions typical of the actual geographic location of the study
StudiedSex	Sex(es) considered to construct the matrix population model	M	Studied only males
		F	Studied only females
		H	Hermaphrodite
		M/F	Males and females separately in the same population matrix model
		A	All sexes together
MatrixComposite	Indicates the type of matrix population model (See **Figure 2** below	Individual	A matrix population model constructed for a single study × species × population × treatment × period combination

	for more information)	Mean	An average (mean) of other matrix population models (e.g. element-by-element arithmetic mean of a population's matrices across several time periods available)
		Pooled	A matrix population model that has been constructed by pooling individual-level demographic information across populations and/or periods. This type of matrix, when available, has always been provided by the author either in the publication or through personal communications
		Seasonal	A matrix population model that does not describe a full annual transition, but rather a seasonal (< 1 yr) transition
MatrixTreatment	Describes if a treatment was applied or not and the nature of the treatment	<treatment>	Provides a brief description of the treatment applied to the population described by the matrix population model. We define treatment as an action intentionally imposed by humans. If more than one applies to a matrix, treatments are separated with a "; "
		Unmanipulated	No experimental, human-imposed treatment was applied. Natural events of non-intentional occurrence (e.g. fire, hurricanes) are recorded as "Unmanipulated", but these incidences are described in the "Observations" variable (below)
Captivity	Whether the study species was studied in the wild or under controlled conditions for most of its life cycle	W	Wild: study in natural conditions
		C	Captive: studied for most part of the life cycle of the species in a botanical garden, green house, laboratory, etc
		CW	Captured from Wild: If the species was taken from a wild population but studied in labs or gardens etc.
MatrixStartYear	First year of study. Year *t* in annual	<xxxx>	e.g. 1995

	matrix population model that describes population dynamics from time *t* to *t*+1		
MatrixStartSeason	First season of study. Season *s* in periodic matrix population model that describes population dynamics from season *s* to *s*+1. Seasons are naturally hemisphere-specific, and this information can be deduced from variables *Country* and *LatDeg*, below. Here season is used as described in manuscript by the authors (Summer in the Southern Hemisphere corresponds to Winter in the Northern Hemisphere)	1	Spring
		2	Summer
		3	Autumn or Fall
		4	Winter
MatrixStartMonth	First month of study. Month *m* in periodic matrix population model that describes population dynamics from	1	January
		2	February
		3	March
		4	April
		5	May

	month *m* to *m*+1	6	June
		7	July
		8	August
		9	September
		10	October
		11	November
		12	December
MatrixEndYear	Last year of study. Year *t*+1 in matrix population model that describes population dynamics from time *t* to *t*+1	<xxxx>	e.g. 2001
MatrixEndSeason	Last season of study. Season *s*+1 in periodic matrix population model that describes population dynamics from season *s* to *s*+1. Here season is used as described in manuscript by the authors (Summer in the Southern Hemisphere corresponds to Winter in the Northern Hemisphere)	1	Spring
		2	Summer
		3	Autumn or Fall
		4	Winter
MatrixEndMonth	Last month of study. Month *m+1* in periodic matrix population model	1	January
		2	February
		3	March

	that describes population dynamics from month *m* to *m*+1	4	April
		5	May
		6	June
		7	July
		8	August
		9	September
		10	October
		11	November
		12	December
MatrixSplit	Type of matrix population model, i.e. can ***A*** it be divided into ***U***, ***F*** and ***C*** submatrices?	Divided	The matrix population model ***A*** has successfully been divided into the process-based submatrices ***U***, ***F*** and ***C***
		Indivisible	The matrix population model ***A*** has not been divided into the process-based sub-matrices ***U***, ***F*** and ***C*** (see Figure 3 below for further details) because insufficient information is available to classify the various demographic processes for each sub-matrix. In indivisible matrices, only ***A*** (see below) is presented
Observation	Other information about the population matrix model	<Information>	This is for the provision of information that is not provided elsewhere but may nonetheless be important (e.g. plant canopy, burning intervals, etc.)
MatrixClassAuthor	Stage description as given by the publication author	<class description>	The stage description as indicated in the source by the publication author of the matrix population model. When possible, we add units (e.g. cm, m, cm^2)
MatrixClassOrganized	We standardize all stages in a given population matrix model to one of three stages (prop, active, dorm) to	prop	Propagule (seed). This applies to every stage defined by the author as seed bank or seed. Users are encouraged to carefully examine matrices with these stages and to implement the appropriate calculations to avoid a spurious additional year being added when no seedbank exists in the study species, as

	facilitate analyses.		explained by Caswell (2001, p. 60)
		active	This includes stages that can neither be placed in the "prop" nor "dorm" (see below) stages
		dorm	A stage that is vegetatively dormant after having germinated and becoming established
MatrixClassNumber	A numerical representation of classes in the population matrix model.	<Integer>	e.g. 1,2,3, … n, where n is the dimension of the population matrix model (see *MatrixDimension* below)
MatrixDimension	Dimension of the matrix population model ***A***	<Integer>	e.g. 5, 8, 15
SurvivalIssue	Minimum stage-specific survival value in the matrix population model submatrix U when > 1	<Integer>	e.g. 1.2, 1.45
matA	The matrix population model ***A*** which describes the population dynamics of a population under conditions described by *MatrixComposite*, *MatrixTreatment*, *MatrixStartYear*, *MatrixEndYear*, *Population*, and *Observation*	A range of numeric values is possible	Transition probabilities are constrained between 0 and 1. Reproductive values are always positive but have no upper bound
matU	The population sub-matrix model ***U*** (a sub-matrix of **A**), which describes the	A range of numeric values possible, but are constrained to be	These include only survival-dependent vital rates (no sexual or clonal reproduction). The values are only digitized for "Divided" matrices

	survival-dependent dynamics of a population under the conditions described above	between 0 and 1	(indicated by variable *MatrixSplit*)
matF	The population sub-matrix model ***F*** (a sub-matrix of **A**), which describes the sexual-reproduction dynamics of a population under the conditions described above	A range of numeric values possible, but are constrained to be positive	These include only estimates of sexual reproduction (no clonal reproduction). The values are only digitized for “Divided” matrices (indicated by variable *MatrixSplit*)
matC	The population sub-matrix model ***C*** (a sub-matrix of **A**), which describes the clonal-reproduction dynamics of a population under the conditions described above	A range of numeric values are possible, but these are constrained to be positive	These include only estimates of clonal reproduction The values are only digitized for “Divided” matrices (indicated by variable *MatrixSplit*)

Appendix

The calculation of mean matrix population models

- In COMPADRE we present the individual matrix population models for each season, year, study population and treatment that can be calculated from a paper, as well as those personally communicated by the publication authors to the COMPADRE digitization team. When pertinent, mean matrix population models that are element-by-element averages (arithmetic mean) across all *Unmanipulated* matrix population models (defined above) and across each treatment group are also given. Mean matrix population models that combine estimates from different treatments, or that combine estimates from treatments with unmanipulated conditions, are not calculated because of their lack of biological interpretation (Figure 2).

- Seasonal matrix population models, where more than one matrix population model exists for a single year, are marked as "Seasonal". For these, we calculate average matrix population models for each season but not for each year (Figure 3). Deriving an annual model from seasonal data is straightforward and the calculations are described by Caswell (2001, p. 349).

Figure 2. Schematic showing how mean population matrix models are derived from all available population matrix models in a study. **A**. For annual population matrix models (those that follow population dynamics from one year to the next, or sometimes more – some tree species are studied with a 5-year interval), element-by-element means are first calculated across the population matrix models describing the population dynamics for time periods within each population. Then the grand element-by-element mean is calculated across the matrix population models for the different populations. This procedure is repeated for models describing dynamics under each imposed treatment (pink matrices below) and for those describing the dynamics of unmanipulated populations (gray). Note that mean population matrix models are only calculated from population matrix models that share the same treatment (or absence of treatment). **B**. For seasonal or periodic population matrix models (e.g. for annual species), the mean seasonal population matrix model is calculated within populations and then across populations, both for matrices describing treated populations and for those describing populations with no imposed treatment (i.e. unmanipulated).

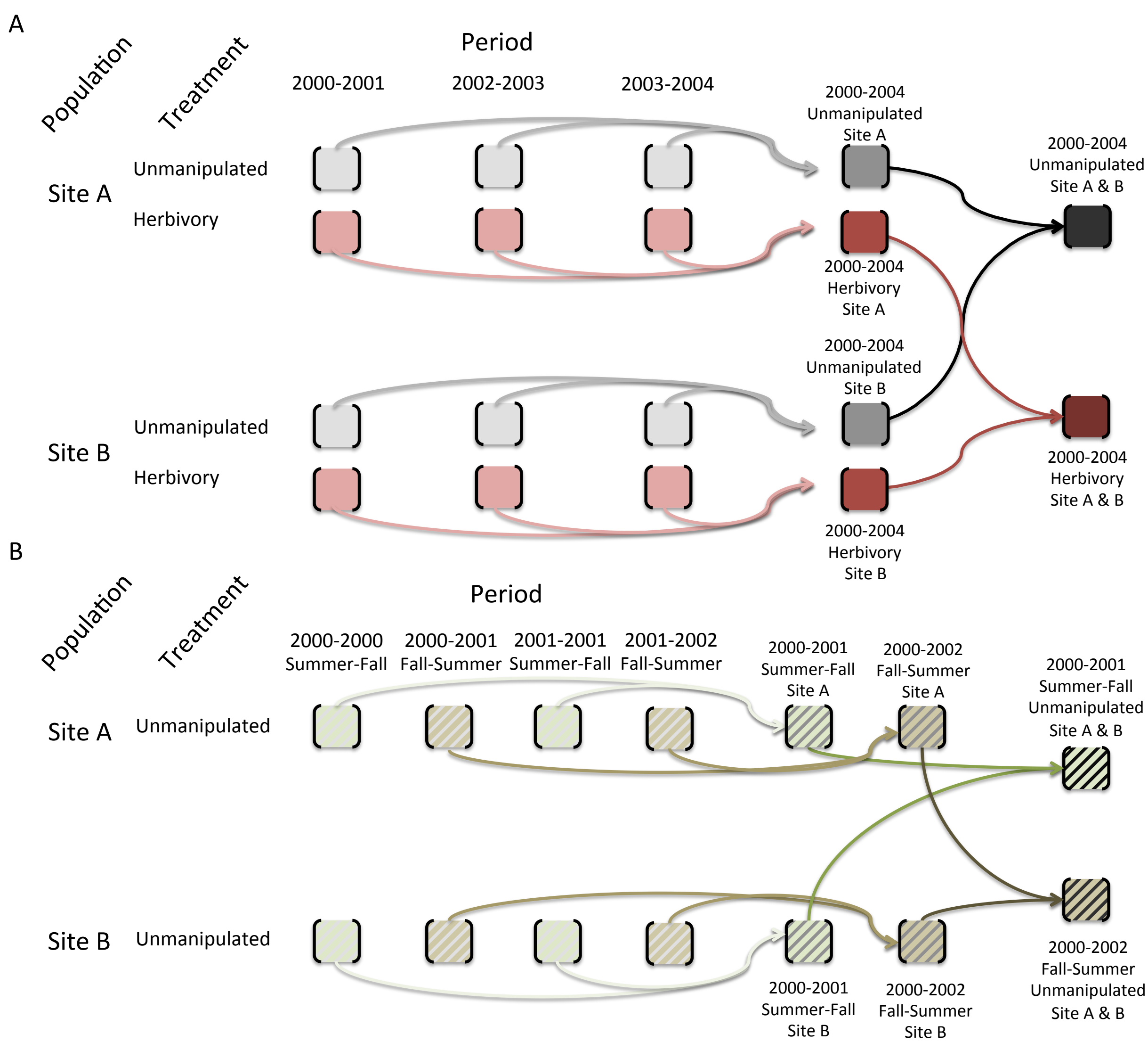
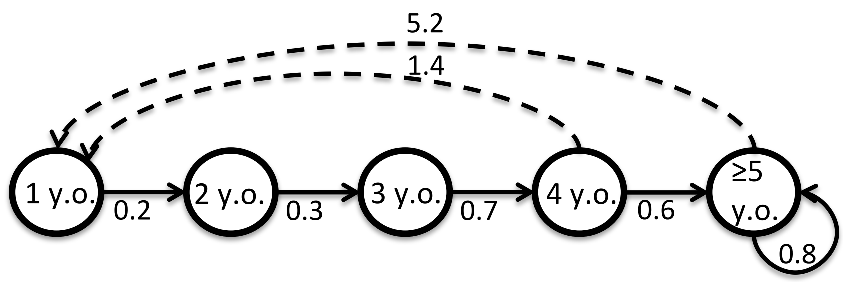


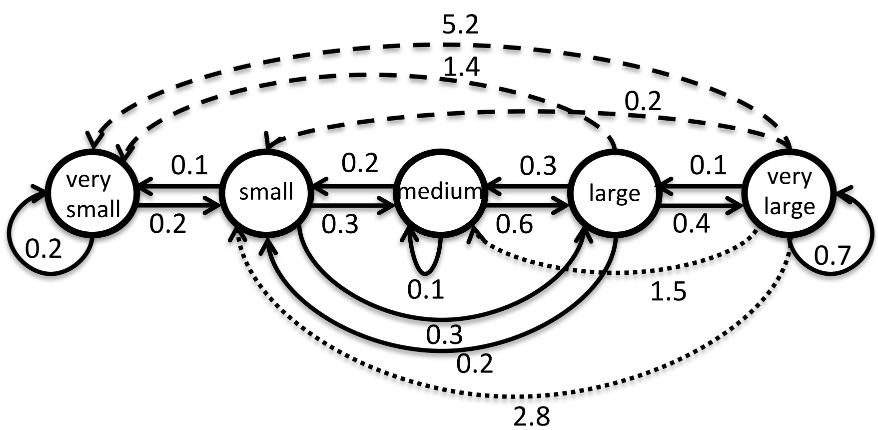
Figure 3. Life cycle of two hypothetical plant populations based on age (A) and size (B), with their corresponding matrix population models ***A***, and underlying basic demographic processes of survival (***U*** sub-matrix; solid arrows), sexual reproduction (***F*** sub-matrix; dashed arrows) and clonal reproduction (***C*** sub-matrix; dotted arrows). In the Leslie matrix model example (A), the division of sub-matrices is more straightforward than in the Lefkovitch matrix model example (B). In the latter imaginary example, individuals can transition into the same stage as they can contribute with sexual and/or clonal offspring (e.g. small stage). In these cases, splitting ***A*** into sub-matrices ***U***, ***F*** and ***C*** is only feasible when sufficient information is provided by the authors (see variable *MatrixSplit* in above).

A



$$\boldsymbol{A} = \begin{bmatrix} 0 & 0 & 0 & 1.4 & 5.2 \\ 0.2 & 0 & 0 & 0 & 0 \\ 0 & 0.3 & 0 & 0 & 0 \\ 0 & 0 & 0.7 & 0 & 0 \\ 0 & 0 & 0 & 0.6 & 0.8 \end{bmatrix} = \boldsymbol{U} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0.2 & 0 & 0 & 0 & 0 \\ 0 & 0.3 & 0 & 0 & 0 \\ 0 & 0 & 0.7 & 0 & 0 \\ 0 & 0 & 0 & 0.6 & 0.8 \end{bmatrix} + \boldsymbol{F} \begin{bmatrix} 0 & 0 & 0 & 1.4 & 5.2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \boldsymbol{C} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

B



$$\boldsymbol{A} = \begin{bmatrix} 0.2 & 0.1 & 0 & 1.4 & 5.2 \\ 0.2 & 0 & 0.2 & 0.2 & 3.0 \\ 0 & 0.3 & 0.1 & 0.3 & 1.5 \\ 0 & 0.3 & 0.6 & 0 & 0.1 \\ 0 & 0 & 0 & 0.4 & 0.7 \end{bmatrix} = \boldsymbol{U} \begin{bmatrix} 0.2 & 0.1 & 0 & 0 & 0 \\ 0.2 & 0 & 0.2 & 0.2 & 0 \\ 0 & 0.3 & 0.1 & 0.3 & 0 \\ 0 & 0.3 & 0.6 & 0 & 0.1 \\ 0 & 0 & 0 & 0.4 & 0.7 \end{bmatrix} + \boldsymbol{F} \begin{bmatrix} 0 & 0 & 0 & 1.4 & 5.2 \\ 0 & 0 & 0 & 0 & 0.2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \boldsymbol{C} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2.8 \\ 0 & 0 & 0 & 0 & 1.5 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

References

Caswell, H. (2001) Matrix population models. Secondedition. *Sunderland, MA: Sinauer Associates*.

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V., Underwood, E.C., D'amico, J.A., Itoua, I., Strand, H.E. & Morrison, J.C. (2001) Terrestrial Ecoregions of the World: A New Map of Life on Earth A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience*, **51**, 933–938.