User guide to the COMADRE Animal Matrix Database

Contents

User guide version information	3
General Instructions	3
Database Organization	3
The meanings of NA in COMADRE	4
Disclaimer	4
What's new in this version	4
Version 1.0.0	4
Variables in metadata	6
SpeciesAuthor	6
SpeciesAccepted	6
CommonName	6
CoLCheckOK	7
CoLCheckDate	7
Infraspecies	7
SpeciesEpithetAccepted	7
GenusAccepted	7
GenusAuthor	8
Family	8
Order	8
Class	8
Authors	9
Journal	9
YearPublication	9
DOI/ISBN	9
AdditionalSource	10
StudyDuration	10
StudyStart	10
StudyEnd	10
AnnualPeriodicity	11
NumberPopulations	
MatrixCriteriaSize	11

	MatrixCriteriaOntogeny	11			
	MatrixCriteriaAge	12			
	MatrixPopulation	12			
	LatDeg	12			
	LatMin	12			
	LatSec	13			
	LatNS	13			
	LonDeg	13			
	LonMin	13			
	LonSec	13			
	LonWE	14			
	Altitude	14			
	Country	14			
	Continent	14			
	Ecoregion	15			
	StudiedSex	16			
	MatrixComposite	16			
	MatrixTreatment	17			
	Captivity	17			
	MatrixStartYear	17			
	MatrixStartSeason	17			
	MatrixStartMonth	18			
	MatrixEndYear	18			
	MatrixEndSeason	18			
	MatrixEndMonth	19			
	MatrixSplit	19			
	MatrixFec	19			
	Observation	19			
	MatrixDimension	20			
	SurvivalIssue	20			
Variables in matrixClass 20					
	MatrixClassAuthor	20			
	MatrixClassOrganized	20			
	MatrixClassNumber	21			

Variables in mat	21		
matA	21		
matU	21		
matF	22		
matC	22		
Variables in version	22		
Version	22		
DateCreated	22		
NumberAcceptedSpecies	22		
NumberStudies	22		
NumberMatrices	23		
Agreement	23		
Appendix - The calculation of mean matrix population models			
Appendix - The calculation of mean matrix population models	23		
References	26		

User guide version information

Version 1.0.0

Date modified: Oct 5th 2015

Contact: comadre-contact@demogr.mpg.de

General Instructions

Database Organization

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

• COMADRE_v_MAJOR_MINOR_PATCH.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.

The meanings of NA in COMADRE

NA in the COMADRE 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 COMADRE. 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)

Disclaimer

The COMADRE 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.comadre-db.org. However, we claim no responsibility for any damage that may arise from using COMADRE. 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.

What's new in this version

Version 1.0.0

 \bullet 402 studies with 346 taxonomically accepted animal species and a total of 1,1603 population matrix models

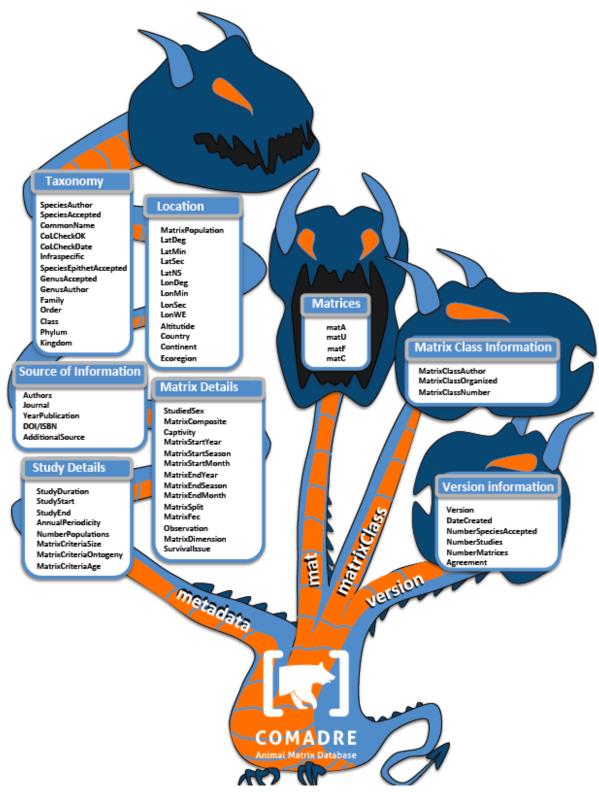


Figure 1. Variables archived in COMADRE, organized according to the general category to which they are associated. The R data object COMADRE is a list containing further list and data.frame objects (named metadata, mat, matrixClass and version), here represented by each branch of the tree. The canopy contains the variables in each data.frame/list, organized by general categories.

Variables in metadata

The object metadata is a data.frame detailing the ancillary metadata that give context to each matrix population model (MPM). These are detailed below.

SpeciesAuthor

Definition: Latin name of the species as indicated in the publication; this may not be the same as the currently accepted name (See SpeciesAccepted below)

Possible values:

- 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 numeric suffix after Genus_species starting with "2" (e.g. "Ursus_americanus", "Ursus_americanus_2", "Ursus_americanus_3").
- <...>var.<...> Where the publication refers to an intraspecific variety, this is indicated with "var." followed by the variety name
- <...>subsp.<...> Where the study refers to an intraspecific subspecies, this is indicated with "subsp." followed by the subspecies name, e.g. "Ursus_americanus_subsp._floridanus"

SpeciesAccepted

Definition: Currently accepted taxonomic name according to the Catalogue of Life (www.catalogueoflife.org)

Possible values:

• <Genus_species> - e.g. Ursus americanus

CommonName

Definition: Common name as used in the publication

Possible values:

• <Common_name> - e.g. American black bear

CoLCheckOK

Definition: Whether the taxonomy detailed here has been verified at the Catalogue of Life (http://www.catalogueoflife.org)

Possible values:

- TRUE Taxonomy has been verified by Catalogue of Life
- FALSE Taxonomy has not been verified by Catalogue of Life

CoLCheckDate

Definition: The date that the taxonomy was checked at the Catalogue of Life

Possible values:

• DDMMYYYY - E.g. 28052015 stands for 28th of May 2015

Infraspecies

Definition: Taxonomic infraspecific name of study SpeciesAuthor

Possible values:

 \bullet <infraspecies> - e.g. "floridanus", for "Ursus americanus subsp. floridanus"

SpeciesEpithetAccepted

Definition: Taxonomic species epithet name of study SpeciesAccepted, as per CoL

Possible values:

• <epithet> - e.g. "americanus", for "Ursus americanus subsp. floridanus". In some cases, the author may have used a different species epithet than the currently accepted genus.

GenusAccepted

Definition: Taxonomic genus name of study SpeciesAccepted, as per CoL

• <genus> - e.g. "Ursus", for "Ursus americanus subsp. floridanus may have used a different genus than the currently accepted genus.</genus>	'. In some cases,	the autho
GenusAuthor		
Definition: Genus used in SpeciesAuthor		
Possible values:		
• <genus> - e.g. "Ursus"</genus>		
Family		
Definition: Family to which species belongs		
Possible values:		
• <family> - e.g. Ursidae</family>		
Order		
Definition: Order to which species belongs		
Possible values:		
• order - e.g. Carnivora		
Class		
Definition: Class to which species belongs		
Possible values:		
• <class> - e.g. Mammalia</class>		

Authors

Definition: Last (family) name of all authors

Possible values:

• <name(s)> - Separated with ";" e.g. "Smith; Jones"

Journal

Definition: The document from which data were sourced.

Possible values:

- <abbreviated journal name> 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 are from a conference talk
- Conference poster Matrices are from a conference poster

YearPublication

Definition: Year of publication

Possible values:

• <yyyy> - e.g. 2012

DOI/ISBN

Definition: Digital Object Identification or International Standard Book Number codes that identify the publication.

P	ossi	ble	val	lues:
	וססט	DIC	va	lucs.

• <DOI> or <ISBN> - e.g. "10.1101/027821"

AdditionalSource

 $\textbf{Definition:} \ \, \text{Additional source}(s) \ \, \text{used to reconstruct the matrix, or to obtain additional information on ancillary data}$

Possible values:

• <Lastname Journal YearPublication> - e.g. Naujokaitis-Lewis Cons Biol 2009

 ${\tt StudyDuration}$

Definition: The number of years of observation. 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

Possible values:

• <integer> - e.g. 14

StudyStart

Definition: First year of study

Possible values:

• <yyyy> - e.g. 2012

${\tt StudyEnd}$

Definition: Final year of study

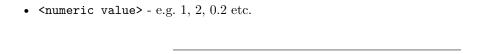
Possible values:

• <yyyy> - e.g. 2014

AnnualPeriodicity

Definition: Indicates the time frame for which the seasonal or annual the MPM was constructed. e.g. 1 indicates that the MPM iteration period is 1yr; 2 indicates that the MPM iterates twice in a year; 0.2 indicates that the MPM iteration occurs every 5yrs.

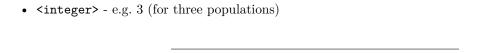
Possible values:



NumberPopulations

Definition: Number of populations of the study species. This variable refers to the number of populations as defined by the author. Within-site replication of permanent plots or similar are not considered to be different populations (see below)

Possible values:



MatrixCriteriaSize

Definition: Indicates whether the matrix population model contains stages based on size, and if so, what variable of size was measured.

Possible values:

- Yes The matrix population model contains stages based on size (e.g. wing span, weight, height, etc)
- No The matrix population model is not based on size

MatrixCriteriaOntogeny

Definition: 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. juvenile, reproductive, dormant, etc)
- No Matrix population model is not based on ontogeny

MatrixCriteriaAge

Definition: Indicates whether the matrix population model contains any stage based on age

Possible values:

- 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

${\tt MatrixPopulation}$

Definition: Name of population where matrix was recorded, usually as given by the publication author. In some cases, where the author provides no name, we give the closest geographic location instead. 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)

Possible values:

• <any text string> - e.g. "Mitta Mitta River".

LatDeg

Definition: Degree value of latitude.

Possible values:

• <0-90> - e.g. 64

LatMin

Definition: Minute value of latitude

Possible values:

• <0-60> - e.g. 34

LatSec

Definition: Seconds value of latitude (decimates are possible). Note that a value of 0 may mean that precise values are not available.

Possible values:

• <0-60> - e.g. 23.14 or 34

LatNS

Definition: Cardinal direction

Possible values:

- N North
- S South

LonDeg

Definition: Degree value of longitude

Possible values:

• <0-180> - e.g. 137

LonMin

Definition: Minute value of longitude

Possible values:

• <0-60> - e.g. 45

LonSec

Definition: Seconds value of longitude (decimates are possible). Note that a value of 0 may mean that precise values are not available.

Possible	values:
----------	---------

• <0-60> - e.g. 34.335

LonWE

Definition: Cardinal direction

Possible values:

- W West
- E East

Altitude

Definition: Altitude of studied population defined as height above sea level of the population in meters.

Possible values:

 \bullet <numeric value> - e.g. 653

Country

Definition: Country/ies where study took place. We use ISO 3 three letter country codes (https://en. wikipedia.org/wiki/ISO_3166-1_alpha-3). If the study involves multiple countries, these are separated by ":".

Possible values:

• <ISO 3 country code(s)> - e.g. DEU; DNK for a study carried out in Germany and Denmark.

Continent

Definition: Continent where study took place

Possible values:

- 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

Definition: Indication 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 The one exception for this is the code LAB used for studies conducted in laboratory or greenhouse conditions.

- 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

Definition:

Possible values:

- M Studied only males
- F Studied only females
- H Studied hermaphrodites
- M/F Males and females separately in the same population matrix model
- A All sexes modeled together

MatrixComposite

Definition: Indicates the type of matrix population model (also see Figure 2 below for more information).

- Individual A matrix population model constructed for a single study × species × population × treatment × period combination.
- 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

Definition: Describes if a treatment was applied or not and the nature of the treatment.

Possible values:

- <treatment> 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

Definition: Whether the study species was studied in the wild or under controlled conditions for most of its life cycle

Possible values:

- W Wild: study in natural conditions
- C Captive: studied for most part of the life cycle of the species in a laboratory, zoo, etc.
- CW Captured from Wild: If the species was taken from a wild population but studied in labs or zoos etc.

MatrixStartYear

Definition: First year of study. Year t in annual matrix population model that describes population dynamics from time t to t+1.

Possible values:

• <yyyy> - e.g. 1994

MatrixStartSeason

Definition: 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).

Possible values:

- 1 Spring
- 2 Summer
- 3 Autumn/Fall
- 4 Winter

MatrixStartMonth

Definition: First month of study.

Possible values:

• <1–12> - e.g. 1 = January, 6= June etc.

${\tt MatrixEndYear}$

Definition: Last year of study. Year t+1 in matrix population model that describes population dynamics from time t to t+1.

Possible values:

• <yyyy> - e.g. 1999

MatrixEndSeason

Definition: 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/Fall
- 4 Winter

MatrixEndMonth

Definition: Last month of study.

Possible values:

• <1-12> - e.g. 1 = January, 6 = June etc.

MatrixSplit

Definition: This indicates whether the MPM **A** matrix has been divided into its constituent **U**, **F** and **C** submatrices. See Figure 3 below for further details.

Possible values:

- Divided The matrix population model **A** has been 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 submatrices U, F and C because insufficient information is available to classify the various demographic processes for each sub-matrix. In indivisible matrices, only A has values given the other matrices are filled with NA values.

MatrixFec

Definition: Whether reproduction was modeled by the MPM.

Possible values:

- Yes Reproduction was included in the MPM, and thus if it occurred in the specific matrix interval of interest values >0 should appear in the submatrix **F**.
- No Reproduction was not included in the matrix population model, either because it was not the goal of the study or due to logistical impossibility. In some ${\bf F}$ submatrices the elements where sexual reproduction would have appeared contain NA values.

Observation

Definition: This variable provides information that is not provided elsewhere but may nonetheless be important (e.g. "This is a matrix constructed from a long term dataset in the UK: size was estimated in a single season but female fecundity from a long term dataset")

D	•1	1	1	
Poss	าก	ıe	val	mes:

• <any text> - any text string giving useful additional information.

MatrixDimension

Definition: Dimension of the matrix population model **A**.

Possible values:

• <integer> - e.g. 5

SurvivalIssue

Definition: Minimum stage-specific survival value in the matrix population model submatrix \mathbf{U} when > 1

Possible values:

• <numeric> - e.g. 0.9, 1.2, 1.45

Variables in matrixClass

The object matrixClass is a list of data.frames detailing the class names used in each MPM.

MatrixClassAuthor

Definition: 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, cm²).

${\tt MatrixClassOrganized}$

Definition: We standardize all stages in a given population matrix model to one of three stages (prop, active, dorm) to facilitate analyses.

Possible values:

- prop Propagule. This applies to every stage defined by the author as a spore-like stage. 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 additional gap-years exists in the study species, as 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 dormant after having been born and established.

MatrixClassNumber

Definition: A numerical representation of classes in the population matrix model.

Possible values:

• <integer> - e.g. 1,2,3, ... n, where n is the dimension of the MPM (see MatrixDimension below)

Variables in mat

The object mat is a list of lists giving the actual MPMs in the form of the A matrix, and (where possible), the U, F and C matrices.

\mathtt{matA}

Description: The matrix population model **A** which describes the population dynamics of a population under conditions described by MatrixComposite, MatrixTreatment, MatrixStartYear, MatrixEndYear, Population, and Observation in the metadata object.

matU

Description: The population sub-matrix model U (a sub-matrix of A), which describes the survival-dependent dynamics of a population under the conditions described above. A range of numeric values are possible, but are constrained to be between 0 and 1. These include only survival-dependent vital rates (no sexual or clonal reproduction). The values are only digitized for Divided matrices (indicated by variable MatrixSplit)

${\tt matF}$

Description: The population sub-matrix model \mathbf{F} (a sub-matrix of \mathbf{A}), which describes the sexual-reproduction dynamics of a population under the conditions described above. A range of numeric values are 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
Description: 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)
Variables in version
version is a list that provides some summary information.
Version
Description: The version number of the database (Major_Minor_Patch). Past versions include: 3.0.0. Current version is 3.1.0. Numbering follows semantic versioning (www.semver.org).
DateCreated
Description: Date of creation of the Rdata database object in the format MMM_DD_YYYY. E.g. Oct_05_2015.
NumberAcceptedSpecies
Description: Total number of taxonomically accepted species. The total number of accepted species is based on variable SpeciesAccepted (above). E.g. 346

NumberStudies

Description: Total number of species \times studies in the database. e.g. 402

NumberMatrices

Description: Total number of	f matrices (not including s	plit matrices \mathbf{U} , \mathbf{F} a	(\mathbf{C}) in the databa	se. e.g. 1603
_				
Agreement				
Description: Link to the UR use of these data: http://www		9	*	itions to the

Appendix - The calculation of mean matrix population models

- In COMADRE 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 COMADRE 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).

Appendix - The calculation of mean matrix population models

- In COMADRE 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 COMADRE 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).

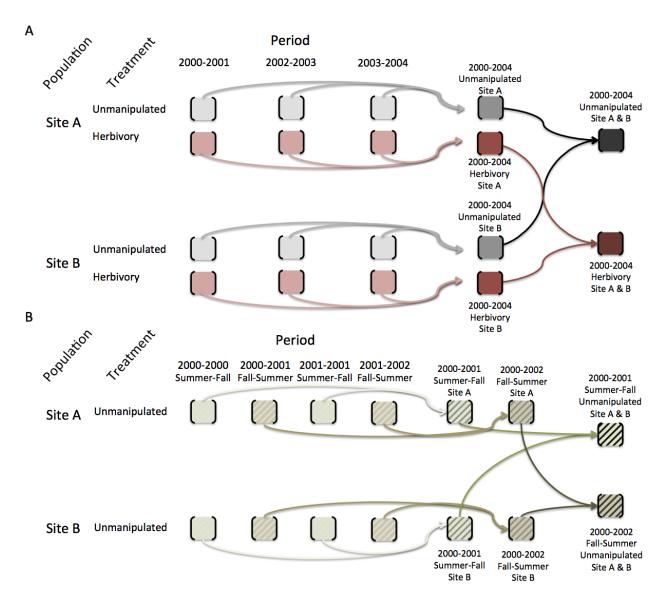


Figure 2. Schematic showing how mean population matrix models are derived from all available population matrix models in a study. A. For annual population matrix models (those that follow population dynamics from one year to the next, or sometimes more – some tree species are studied with a 5-year interval), 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).

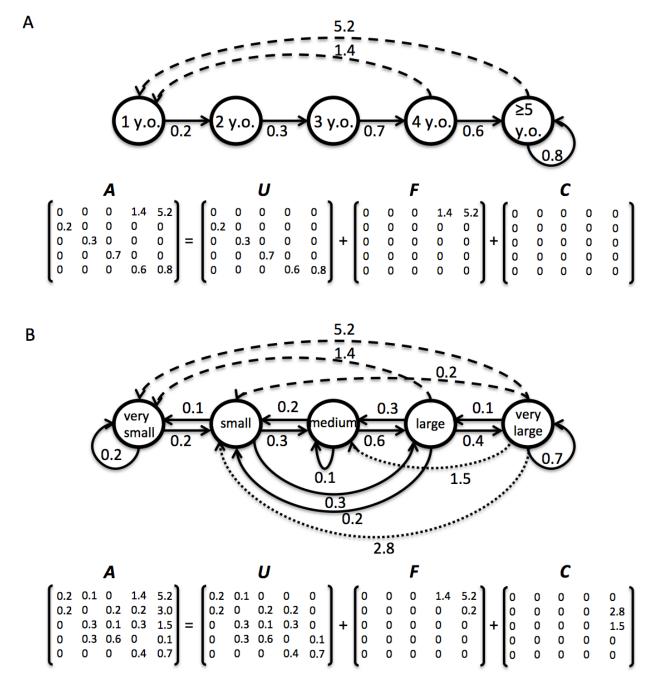


Figure 3. Life cycle of two hypothetical animal populations based on age (\mathbf{A}) and size (\mathbf{B}) , with their corresponding matrix population models A, and underlying basic demographic processes of survival (U sub-matrix; solid arrows), sexual reproduction (F sub-matrix; dashed arrows) and clonal reproduction (C sub-matrix; dotted arrows). In the Leslie matrix model example (\mathbf{A}) , the division of sub-matrices is more straightforward than in the Lefkovitch matrix model example (\mathbf{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).

References

Caswell, H. (2001) Matrix population models. Second edition. 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.