The COMPADRE and COMADRE User Guide and Data Entry Protocol

The Compadre Team

Last updated on 2020-11-20

Contents

Chapter 1

Welcome

The COMPADRE and COMADRE plant and animal matrix databases contain demographic data in the form of matrix population models for hundreds of plant and animal species, as well as algae, fungi, bacteria, and viruses!

COMPADRE is a long-term enterprise initiated in 1989 by Jonathan Silvertown and Miguel Franco. COMADRE was instituted by Salguero-Gómez in 2011. Currently these databases are supported and hosted at the Max Planck Institute for Demographic Research (Germany).

To access data and to learn more about the project, who's involved, and how it is funded, please visit the main web site for COMPADRE and COMADRE databases at https://www.compadre-db.org.

1.1   This site

This web site provides reference material about the variables contained in the databases (**User Guide**), and about how data are entered into the databases (**Data Entry Protocol**). Use the navigation facility to find the area you are interested in.

> **Note:** The book is a "work in progress" and will change during the course. The latest version can always be found at the website.
> It is currently maintained by Owen Jones (jones@biology.sdu.dk). We welcome reports of errors and suggestions for improvement.

Chapter 2

User Guide

Architecture
GrowthType

Taxonomy
SpeciesAuthor
SpeciesAccepted
Authority
TaxonomicStatus
TPLVersion
InfraspeciesAccepted
GenusAccepted
Genus
Family
Order
Class
DicotMonoc
AngioGymno
Phylum
Kingdom

Location
MatrixPopulation
LatDeg
LatMin
LatSec
LatNS
LonDeg
LonMin
LonSec
LonWE
Altitutide
Country
Continent
Ecoregion

Matrices
A
U
F
C

Matrix Class Information
MatrixClassAuthor
MatrixClassOrganized
MatrixClassNumber

Source of Information
Authors
Journal
YearPublication
DOI/ISBN
AdditionalSource

Matrix Details
StudiedSex
MatrixComposite
Captivity
MatrixStartYear
MatrixStartSeason
MatrixStartMonth
MatrixEndYear
MatrixEndSeason
MatrixEndMonth
MatrixSplit
MatrixFec
Observation
MatrixDimension
SurvivalIssue

Study Details
StudyDuration
StudyStart
StudyEnd
AnnualPeriodicity
NumberPopulations
MatrixCriteriaSize
MatrixCriteriaOntogeny
MatrixCriteriaAge

Version information
Version
DateCreated
NumberSpeciesAccepted
NumberStudies
NumberMatrices
Agreement

metadata   mat   matrixClass   version

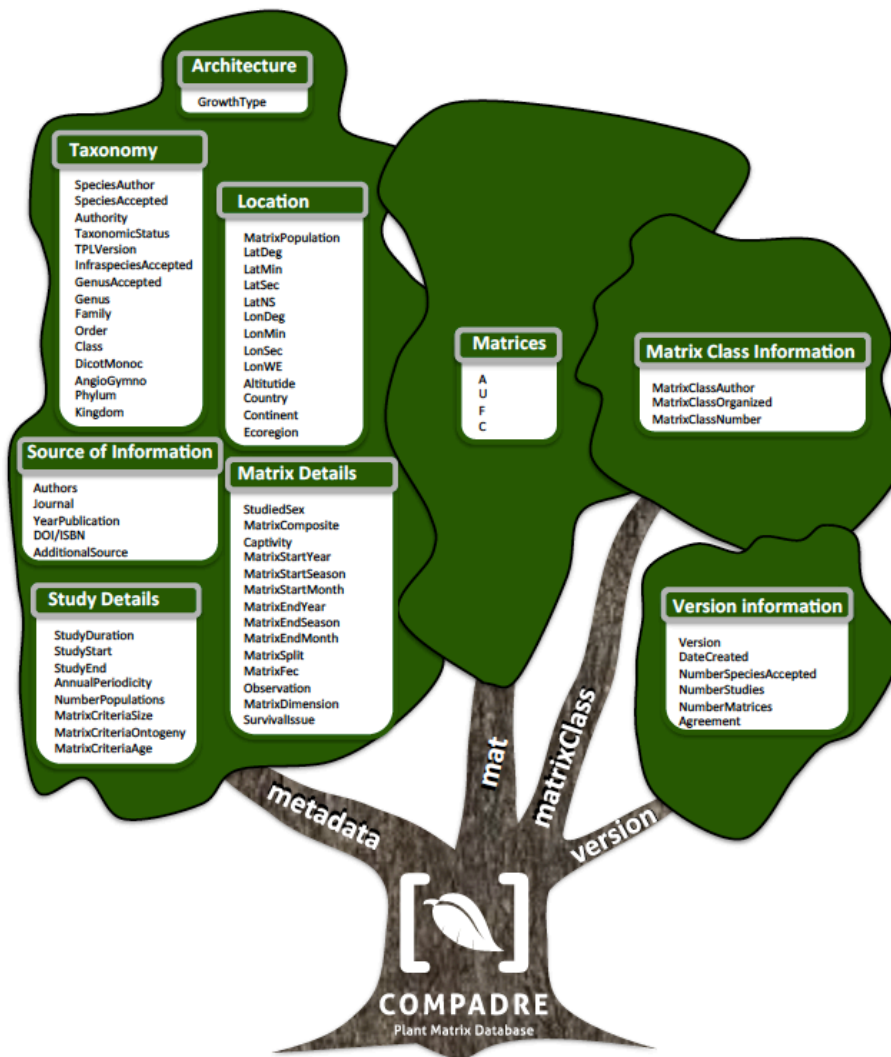**COMPADRE**
Plant Matrix Database

Figure 1:   *Variables archived in COMPADRE, organized according to the general category to which they are associated. The R data object compadre 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.*

2.1   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.

2.1.1   `MatrixID`

Definition Unique identification number given to each MPM and its associated metadata

Possible values:

- `MatrixID` - e.g. 238263

2.1.2   `SpeciesAuthor` (Population form and the Species form)

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. "`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`"

2.1.3   `SpeciesAccepted` (Population form)

Definition: Currently accepted latin name. This information is obtained from GBIF

Possible values:

- Genus species - e.g. "`Taraxacum officinale`", "`Icaricia icarioides fenderi`", "`Iris pallida subsp. cengialti`", etc.

Note on nomenclature in COMPADRE and COMADRE: Species in COMPADRE are displayed following guidelines in the International Code of Nomenclature for algae, fungi, and plants (ICN) (i.e. abbreviated infraspecific type **is** displayed between the specific and subspecific epithets), and species in COMADRE are displayed following guidelines in the International Code of Zoological Nomenclature (ICZN) (i.e. abbreviated infraspecific type **is not** displayed between the specific and subspecific epithets).

Examples:

- "`Anthyllis vulneraria subsp. alpicola`" and "`Hosta undulata var. undulata`" in COMPADRE
- "`Panthera tigris sumatrae`" and "`Somateria mollissima borealis`" in COMADRE

2.1.4 `CommonName` (Population form)

Definition: Common name as used in the publication

Possible values:

- Common_name - e.g. `American black bear`

2.1.5 `Kingdom` (Species form)

Definition: Kingdom to which species belongs

Possible values:

- kingdom - e.g. Plantae, Fungi, Rhodophyta, Chromista (COMPADRE includes fungi and algae as well as plants, and COMADRE includes viruses and bacteria as well as animals)

2.1.6 `Phylum` (Species form)

Definition: Phylum to which species belongs

Possible values:

- phylum - e.g. `Magnoliophyta`

2.1.7 `Class` (Species form)

Definition: Class to which species belongs

Possible values:

- class - e.g. `Magnoliopsida`

2.1.8 `Order` (Species form)

Definition: Order to which species belongs

Possible values:

- order - e.g. `Caryophyllales`

2.1.9 `Family` (Species form)

Definition: Family to which species belongs

Possible values:

- family - e.g. `Polygonaceae`

2.1.10 `Genus` (Species form)

Definition: Taxonomic genus name of study *SpeciesAccepted*

Possible values:

- genus - e.g. `Chorizanthe`

2.1.11 `Species` (Species form)

Definition: Taxonomic species name of study *SpeciesAccepted*

Possible values:

- species - e.g. sapiens

2.1.12 `Infraspecies` (Species form)

Definition: Taxonomic infraspecies name of study *SpeciesAccepted*

Possible values:

- `infraspecies` - e.g. fenderi

2.1.13 `InfraspeciesType` (Species form)

Definition: Infraspecific type of study *SpeciesAccepted*

Possible values:

- `infraspecies type` - e.g. subspecies, variety, color form

2.1.14 `OrganismType` (Species form)

Definition: General organism type, based mainly on architectural organization. The species was assigned to one of these possible values using the description of plant growth type provided by the author for COMPADRE; animal taxonomic class, virus, or bacteria for COMADRE; and other external sources (e.g. other publications)

Possible values for **COMPADRE**:

- `Algae` - A brown, green or red alga. Note that green algae are in kingdom Plantae, but should still be recorded as OrganismType = Algae
- `Fungi` - A fungus species, including yeasts and molds as well as multicellular fungi.
- `Annual` - 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 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
- `Bryophyte` - a bryophyte
- `Epiphyte` - an epiphyte
- `Fern` - a fern
- `Herbaceous perennial` - a herbaceous perennial
- `Liana` - a liana
- `Palm` - a palm
- `Shrub` - a shrub
- `Succulent` - a succulent
- `Tree` - a tree

Possible values for **COMADRE**:

- taxonomic class - e.g. "`Mammalia`", "`Aves`", "`Gastropoda`"

2.1.15 `DicotMonoc` (Species form)

Definition: Whether species is a dicot or monocot

Possible values:

- `Eudicot` - species is a dicot
- `Monocot` - species is a monocot
- `NA` - Species is neither a dicot nor a monocot (not an angiosperm) or species in COMADRE

2.1.16 `AngioGymno` (Species form)

Definition: Whether species is an angiosperm or a gymnosperm

Possible values:

- `Angiosperm` - Species is an angiosperm
- `Gymnosperm` - Species is a gymnosperm
- `NA` - Neither angiosperm nor gymnosperm or species in COMADRE

2.1.17 `Authors` (Publication Form)

Definition: Last (family) name of all authors

Possible values:

- name(s) - Separated with ";" e.g. "`Smith; Jones`"

2.1.18 `Journal` (Publication form)

Definition: The name of the journal from which data were collected

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
- `NA` - Data sourced from other material

2.1.19 `SourceType` (Publication form)

Definition: The type of source from which the data were collected

Possible values:

- `Journal Article` - Matrices are from a journal article
- `Book` - Matrices are from a book
- `Book Chapter` - Matrices are from a book chapter
- `Thesis` - Matrices are from a doctoral dissertation or masters thesis
- `Report` - Matrices are from a report
- `Conference` - Matrices are from a conference talk or poster
- `Other` - Matrices are from another type of source, e.g., website, preprint

2.1.20 `OtherType` (Publication form)

Definition: Type of source from which the data were collected when `SourceType` is 'Other'

Possible values:

- `Internet` - Matrices are from an internet source, e.g. website
- `Unpublished` - Matrices are unpublished and provided by the data collector
- `Preprint` - Matrices are from an article preprint, e.g. bioRxiv

2.1.21 `YearPublication` (Publication form)

Definition: Year of publication

Possible values:

- yyyy - e.g. `2012`

2.1.22 `DOI_ISBN` (Publication form)

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

Possible values:

- DOI or ISBN - e.g. "`10.1111/1365-2745.12334`"

2.1.23 `AdditionalSource` (Publication form)

Definition: Additional source(s) used to reconstruct the matrix, or to obtain additional information on ancillary data

Possible values:

- Lastname Journal YearPublication - e.g. `Godinez-Alvarez Bot Rev 2003`

2.1.24 `StudyDuration` (This is calculated using the fields Study End and Study Start; Publication form)

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`
- `Simulation` - The matrices in the study are from simulated data only and not from real-world or experimental data

2.1.25 `StudyStart` (Publication form)

Definition: First year of study

Possible values:

- yyyy - e.g. `2012`
- `Simulation` - The matrices in the study are from simulated data only and not from real-world or experimental data

2.1.26 `StudyEnd` (Publication form)

Definition: Final year of study

Possible values:

- yyyy - e.g. `2014`
- `Simulation` - The matrices in the study are from simulated data only and not from real-world or experimental data

2.1.27 `ProjectionInterval` (Formerly `AnnualPeriodicity`; Matrix form)

Definition: Indicates the time step (periodicity) for which the seasonal or annual transition matrix was constructed, e.g., '1' indicates that the MPM iteration period is 1 year; 0.5 indicates that the MPM iterates once every 0.5 years or 6 months; 2 indicates that the MPM iteration occurs every 2 years

Possible values:

- numeric value - e.g. `1`, `2`, `0.2` etc.

2.1.28 `MatrixCriteriaSize` (Population form)

Definition: Indicates whether the matrix population model contains stages based on size

Possible values:

- `Yes` - The matrix population model contains stages based on size (e.g. DBH, stem length, stem height, stem number, etc.)
- `No` - The matrix population model is not based on size

2.1.29 `MatrixCriteriaOntogeny` (Population form)

Definition: Indicates whether the matrix population model contains stages based on development

Possible values:

- `Yes` - The matrix population model contains at least one class that is based on development/ontogeny (e.g. seedbank, reproductive, vegetative, dormant, adult, etc.)
- `No` - Matrix population model is not based on ontogeny

2.1.30 `MatrixCriteriaAge` (Population form)

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

Possible values:

- `Yes` - The matrix population model contains at least one class that is based on age (e.g. 1, 2, 3, etc.)
- `No` - The matrix population model is not based on age

2.1.31 `MatrixPopulation` (Formerly PopulationName; Population form)

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. "`Brazeau Creek, Florida`".

2.1.32 `NumberPopulations` (This has to be calculated, it is not a field, at least not one that can be observed on our end)

Definition: Number of populations of the study species (`SpeciesAccepted`) within a publication. 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)

Possible values:

- integer - e.g. `3` (for three populations)

2.1.33 `Lat` (From the coordinates field. If it was recorded in DMS then it is transformed to decimal degrees; Population form)

Definition: Degree value of latitude in decimal degree format. Negative values indicate that the location is in the southern hemisphere

Possible values:

- -90-90 - e.g. `64.87`, `-43.7`

2.1.34 `Lon` (From the coordinates field. If it was recorded in DMS then it is transformed to decimal degrees; Population form)

Definition: Degree value of longitude in decimal degree format. Negative values indicate that the location is to the west of the Greenwich meridian (0 degrees)

Possible values:

- -180-180 - e.g. `137.78`, `-78.32`

2.1.35 `Altitude` (Population form)

Definition: Altitude of studied population defined as height above or below sea level in meters of the population. Negative numbers indicate populations below sea level

Possible values:

- numeric value - e.g. `653`

2.1.36 `Country` (Population form)

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.

2.1.37 `Continent` (Population form)

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

2.1.38 `Ecoregion` (Population form)

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 exceptions for this are the codes `LAB` for studies conducted in laboratory or greenhouse conditions or `NA` for populations in which an assigning an ecoregion is not practical (e.g. human data pooled for an entire country)

Possible values:

- `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
- `SRE` - Freshwater - small river 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
- `NA` - Populations in which an assigning an ecoregion is not practical (e.g. human data pooled for an entire country)

2.1.39 `StudiedSex` (Matrix form)

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

2.1.40 `MatrixComposite` (Formerly Composition; Matrix form)

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

Possible values:

- `Individual` - A matrix population model constructed for a single study $\times$ species $\times$ population $\times$ treatment $\times$ 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

2.1.41 `MatrixSeasonal` (Matrix form)

Definition: Indicates whether a matrix population model is seasonal, e.g., a matrix population model that does not describe a full annual transition, but rather a seasonal (< 1 yr) transition

Possible values:

- `Yes` - The matrix population model is seasonal
- `No` - The matrix population model is not seasonal

2.1.42 `MatrixTreatment` (Formerly Treatment; Matrix form)

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)

2.1.43 `MatrixCaptivity` (Formerly Captivity; Matrix form)

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 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.

2.1.44 `MatrixStartYear` (Matrix form)

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

2.1.45 `MatrixStartSeason` (Now MatrixSeasonal PLUS 2 more fields: MatrixStartSeason and MatrixEnd-Seaon; Matrix form)

Definition: First season of study

Possible values:

- `Spring`
- `Summer`
- `Autumn`
- `Winter`
- `NA`

2.1.46 `MatrixStartMonth` (Matrix form)

Definition: First month of study

Possible values:

- *1-12* - e.g. `1` = January, `6`= June, etc.

2.1.47 `MatrixEndYear` (Matrix form)

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`

2.1.48 `MatrixEndSeason` (Matrix form)

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)

Possible values:

- `Spring`
- `Summer`
- `Autumn/Fall`
- `Winter`
- `NA`

2.1.49 `MatrixEndMonth` (Matrix form)

Definition: Last month of study

Possible values:

- *1-12* - e.g. `1` = January, `6` = June, etc.

CensusType

Definition: Sampling census used by the authors

Possible values:

- `Prebreeding` - Population was censused before a birth pulse
- `Postbreeding` - Population was censused after a birth pulse
- `Intermediate` - Population was censused between successive birth pulses (e.g. bears give birth while hidden in winter dens, which makes pre- or post-breeding censuses impractical
- `Birth-flow` - Population for which reproduction occurs continually over an interval (e.g. humans)

2.1.50 `MatrixSplit` (Matrix form)

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:

- `Divisible` - 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 sub-matrices **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

2.1.51 `MatrixFec` (Matrix form)

Definition: This records whether fecundity was measured at all for this matrix model

The rationale for this is that a fecundity (in the **F** matrix) may be recorded as 0 either because it was not measured, or because reproduction was measured and estimated to be zero. As a general rule, if any values for fecundity (e.g. in the **F** matrix) are non-zero, then `MatrixFec` should be recorded as 'Yes'. If all values are zero, then you we check the paper carefully to check whther this is because no fecundity was detected, or whether no attempt was made to measure fecundity

Possible values:

- `Yes` - Reproduction was included in the MPM, and thus values 0 should appear in the submatrix **F** if sexual reproduction occurred in these stages/ages. Note that it is possible that reproduction was measured and modelled but none occurred. In that case, values of 0 are possible in the **F** matrix

- `No` - Reproduction was not modelled by the matrix population model, either because it was not the goal of the study or due to logistical impossibility. In some **F** submatrices the elements where sexual reproduction would have appeared contain `NA` values, and in others they contain 0 values
- `NA` - it is not known whether fecundity was measured for the matrix

2.1.52 `Observations` (Matrix form)

Definition: This variable provides information that is not provided elsewhere but may nonetheless be important (e.g. plant canopy, burning intervals, etc.)

Possible values:

- any text - Any text string giving useful additional information

2.1.53 `MatrixDimension` (Matrix form)

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

Possible values:

- *integer* - e.g. 5

2.1.54 `SurvivalIssue` (This is not a field, it needs to be calculated from the survival matrix)

Definition: This is a calculated field used for checking matrix data. It is the maximum stage-specific survival value in the survival **U** submatrix. To calculate `SurvivalIssue`, we sum each column in the **U** matrix individually. The maximum value among those sums **is** the `SurvivalIssue`.

Note: This value is only calculated with a number if `MatrixSplit` = "Divided", that is, if the matrix is able to be split into submatrices. If `MatrixSplit` = "Indivisible", the matrix is not divisible, so there is no survival **U** submatrix. In the case of indivisible matrices, the `SurvivalIssue` field will be empty (`NA`) and "Not Calculated" (`NC`) in the SQL database

Possible values:

- *numeric* - e.g. `0.9`, `1.2`, `1.45`
- `NA` - Matrix population model is not divisible into submatrices

2.1.55 `_Database`

Description: An internal check used for database creation and displays the database name (COMPADRE or COMADRE)

2.1.56 _PopulationStatus

Description: An internal check used for database creation. All values should display as `Released`

2.1.57 _PublicationStatus

Description: An internal check used for database creation. All values should display as `Released`

2.2 Variables in `matrixClass`

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

2.2.1 `MatrixClassAuthor` (Class Author; Matrix form)

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`, `cm^2`)

2.2.2 `MatrixClassOrganized` (Class Organized; Matrix form)

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 (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 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

2.2.3 `MatrixClassNumber` (The number of classes in a matrix. This has to be calculated from `MatrixClassOrganized`)

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)

2.3   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

2.3.1   `matA` (Matrix)

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

2.3.2   `matU` (Survival matrix)

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`)

2.3.3   `matF` (Fecundity matrix)

Description: 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 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`)

2.3.4   `matC` (Clonal matrix)

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`)

2.4 Variables in `version`

`version` is a list that provides some summary information.

2.4.1 Version

Description: The version number of the database (Major_Year_Month_Patch). Past versions used (Major_Minor_Patch) format. Numbering follows semantic versioning (www.semver.org)

2.4.2 Type

Description: The type of database release

Possible values:

- `Release` - The standard COMPADRE or COMADRE release
- `Patch` - A release version that has been patched since the previous standard release. Patches are released if there are significant additions or error corrections to the database between standard releases
- `Test` - A version used strictly for internal testing of the databases. These versions do not follow semantic versioning, are not preserved indefinitely, and are not intended for public release

2.4.3 DateCreated

Description: Date of creation of the `Rdata` database object in the format `MMM_DD_YYYY`. E.g. Aug_06_2015

2.4.4 TimeCreated

Description: Time of creation of the `Rdata` database object in the format `HH:MM`. E.g. 19:08. Times are in GMT

2.4.5 Agreement

Description: The URL link for the data user agreement containing the open-access conditions to the use of these data (https://www.compadre-db.org/UserAgreement)

2.4.6 How to Cite

Description: The URL link for instructions on how to cite the data (https://compadre-db.org/Help/HowToCite)

2.4.7 `GeneratorScriptVersion`

Description: Version number for the code used to generate the COMPADRE and COMADRE databases. This version number is for internal tracking and troubleshooting

2.5   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).
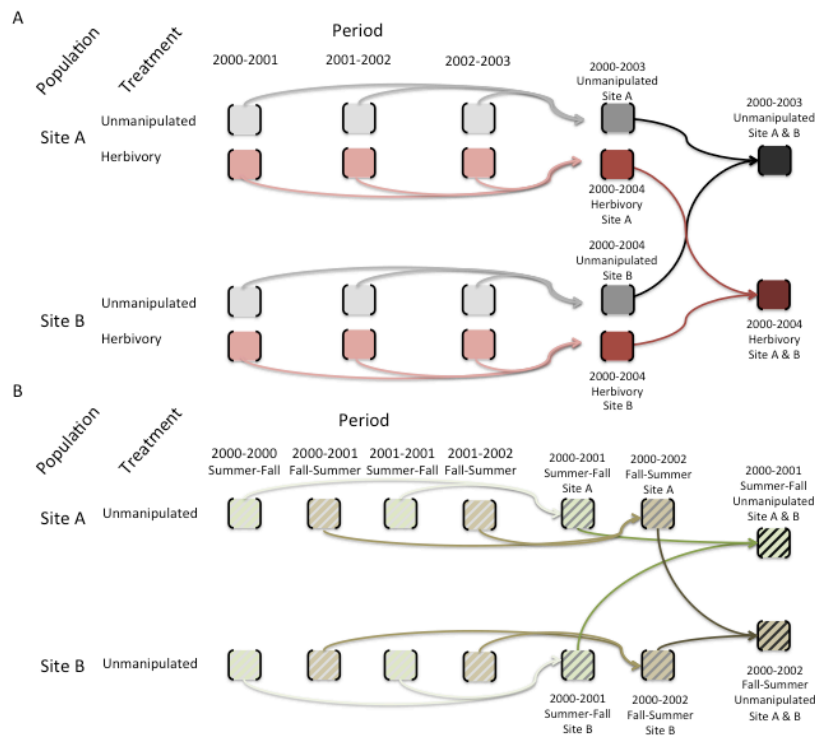
Figure 2. *Schematic showing how mean population matrix models are derived from all available population matrix models in a study.* **A.** *For annual population matrix models (those that follow population dynamics from one year to the next, or sometimes more – some tree species are studied with a 5-year 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).*
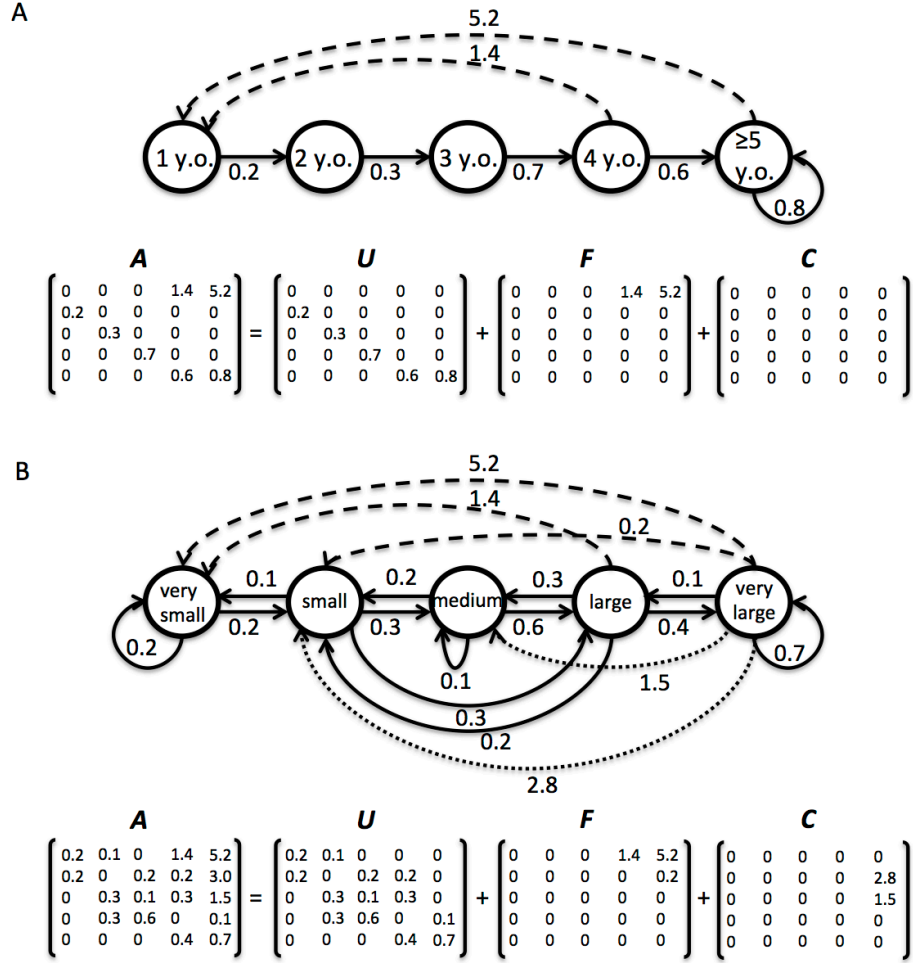
A

$$
\begin{bmatrix}
0 & 0 & 0 & 1.4 & 5.2 \\
0.2 & 0 & 0 & 0 & 0 \\
0 & 0.3 & 0 & 0 & 0 \\
0 & 0 & 0.7 & 0 & 0 \\
0 & 0 & 0 & 0.6 & 0.8
\end{bmatrix}
=
\begin{bmatrix}
0 & 0 & 0 & 0 & 0 \\
0.2 & 0 & 0 & 0 & 0 \\
0 & 0.3 & 0 & 0 & 0 \\
0 & 0 & 0.7 & 0 & 0 \\
0 & 0 & 0 & 0.6 & 0.8
\end{bmatrix}
+
\begin{bmatrix}
0 & 0 & 0 & 1.4 & 5.2 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix}
+
\begin{bmatrix}
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix}
$$

B

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

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).*

Chapter 3

Data Entry Protocol

3.1 Preamble

The COMPADRE Plant Matrix Database (COMPADRE hereafter) and the COMADRE Animal Matrix Database (COMADRE hereafter) are compiled from (mostly) published literature. The majority of this literature was published as peer-reviewed articles in academic journals, with a smaller number of data obtained from books, conference presentations, etc.

The data digitisation team (The COMPADRinos) carefully examine the literature and extract information in order to reconstruct the matrix population models analysed by the original researchers. The extracted data are entered into our web-based digitization forms, which are then stored as structured SQL tables. Wherever there are doubts about aspects of the data source, the team attempts to contact the researchers to resolve them. There are three types of digitization forms in COMPADRE/COMADRE: (1) The PUBLICATION form which contains metadata on the publication, such as bibliographic and study level information; (2) The POPULATION form with metadata that allows for the interpretation of the matrix data, and it pertains to information at the species/population level; and (3) the MATRIX form, which contain the matrix population models themselves, as well as information that is matrix-specific.

These data are compiled using a set of R scripts and released to the public as open access data via the website: www.compadre-db.org.

This document outlines the way that the digitisation team enters data into our website. It is primarily intended for internal use and, as such, it includes information on some variables which are not released into the public domain. Nevertheless, following our philosophy on open access, replicability and transparency, we provide open access to the document here. Users external to our team with questions can direct them at compadrecontact@gmail.com. Specific questions can also be directed to Rob Salguero-Gómez (rob.salguero@zoo.ox.ac.uk) or Owen Jones (jones@biology.sdu.dk).

3.2   General information

- If you find additional sources *always* attach them and note them in "Additional Source". An additional source is a paper that references matrices from another source already in the database.
- Download the PDF attached to the publication form assigned to you, all pieces of information extracted from PDFs must be highlighted directly on the PDF in yellow and re-upload the PDF to the publication form.
- If a PDF contains matrices/data from other papers and/or studies, this must be noted as "Not primary source" in the notes variable of the publication form.
- If additional publications are found with respect to a given species, these must be inspected for any additional matrix population models with different model assumptions (e.g. different stages) and/or different study years, populations and/or treatments. If the paper provides new information, it must be included in the database as a new publication.
- Authors must only be contacted from the official email: compadrecontact@ gmail.com.
- Where matrices can be split, please always include an "A" matrix as well as "U", "F" & "C" matrices (for explanation see here in the protocol for matrix form). Present matrices from left-to-right in the order A, U, F, C. **Always** split matrices when possible.
- There are also species that do not fall into the Kingdoms Plantae or Animalia. Fungi and Algae should be included in COMPADRE, with the plants. Any other species (e.g. bacteria) should be included in COMADRE.

3.3   Values used throughout the database

The following values are commonly used throughout the database.

- `NA` - *Not available* - The information is not available either because the author does not have it or because the information is not recognized as "priority" (see below) for the digitisation.
- `AFI` - *Asked for it* - The information is not yet in the database, but the COMPADRino has already contacted the author requesting it (i.e. the author has not yet replied).
- `NDY` - *Not digitised yet* - The information has not yet been digitised. This is likely to occur when some priority information has been digitised for a given species, and the COMPADRino has moved onto another species without contacting an author for the missing information.
- `NC` - *Not calculated* - Information was not calculated because its calculation is not sensible. For instance, latitudes and longitudes of various populations of the same species are not averaged and assigned to the grand-mean population matrix of the species. This is the same for ranges, where author provides a range of latitudes and longitudes for a given population. Ranges are instead noted in the observation column.

- `LAB` - *Laboratory* - The author collected demographic data for the studied species in a laboratory or under greenhouse conditions. Controlled, typically indoor, conditions that mean the study population is not affected by the environmental conditions typical of the **actual** geographic location of the study.
- `FRANK` - *Frankenstein matrix* - Frankenstein matrices are MPMs where the author parameterised their model using demographic data from 2+ other papers/sources; in other words, when the matrix values come from a source that was not affected by the same conditions as the rest of the matrix. These MPMs should be clearly indicated with "`FRANK`" in the observation column. Furthermore, geographic location data, time information (start/end year, season etc.) should be recorded only if the author reports ONE main location for the matrix; if multiple locations are reported, then it should be recorded as `FRANK` and not `NA.`
- `Simulation` - *A simulated matrix* - The matrix model is a simulation, and perhaps only loosely based on a specific real species. Typically, these simulated matrices are used to explore the behaviour of a population in a theoretical way. In those cases, there is no geographic location, ecoregion etc. For these data we do not calculate averages or ranges for locations and time periods but instead enter `Simulation` (not `NA`).

3.4 Important 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.
- Salguero-Gómez, R., Jones, O. R., Archer, C. R., Buckley, Y. M., Che-Castaldo, J., Caswell, H., Hodgson, D., Scheuerlein, A., Conde, D. A., Brinks, E., de Buhr, H., Farack, C., Gottschalk, F., Hartmann, A., Henning, A., Hoppe, G., Römer, G., Runge, J., Ruoff, T., Wille, J., Zeh, S., Davison, R., Vieregg, D., Baudisch, A., Altwegg, R., Colchero, F., Dong, M., de Kroon, H., Lebreton, J.-D., Metcalf, C. J. E., Neel, M. M., Parker, I. M., Takada, T., Valverde, T., Vélez-Espino, L. A., Wardle, G. M., Franco, M., Vaupel, J. W. (2015), The COMPADRE Plant Matrix Database: an open online repository for plant demography. Journal of Ecology, 103: 202–218. https://doi.org/10.1111/1365-2745.12334
- Kendall, B.E., Fujiwara, M., Diaz-Lopez, J., Schneider, S., Voigt, J., Wiesner, S. 2019. Persistent problems in the construction of matrix population models. Ecological Modelling 406, 33–43 https://doi.org/10.1016/j.ecolmodel.2019.03.011

3.5   Protocol for the publication form

3.5.1   Status

This is the status for the publication. Please READ the website-status protocol to determine the status that should be assigned to the publication.

- `Assigned` – the publication has been assigned to a node. The compadrino will get publications in this status. Once the compadrino has started working on the publication, it should be changed to "Digitizing"
- `Digitizing` – compadrino is working/editing the population
- `Rejected/Deleted`
- `Awaiting info` – compadrino has determined s/he requires more information to complete digitization
- `Review-Ready` – the population is ready to be reviewed by a database curator
- `Released` – population can be now seen by public users
- `Error` – an error has been found in the population and needs to be corrected

3.5.2   Node Assignment

The node the publication has been assigned for. Only coordinators and Admins edit this field.

3.5.3   Assigned to

The name of the compadrino that has claimed or has been assigned the publication.

3.5.4   Database

The database to which the publication belongs to, either Compadre (plants) or Comadre (animals). Usually determined by the database coordinator. When you are assigned a publication, always make sure the assigned database is correct.

- `Compadre` (plants)
- `Comadre` (animals)

3.5.5   Attach a document

Attach all files used to digitize this publication. Include original journal article, additional sources, appendices, files from author communications, etc. Usually all proposed/assigned publication will have attached the original journal article.

3.5.6 DOI/ISBN

The DOI of the journal article. If you have the DOI, you could automatically fill the bibliographic information fields. Check to make sure those fields were populated correctly.

3.5.7 Authors

Publication authors. This is usually done by the database coordinator. Make sure all authors have been added and are in the right order. If an author is missing search for the last name in the search box; make sure to select the correct author. If the author is not in the list, use the button "Create author" to add a new one:

- *Author name*: The last name of the author.
- *Author initials*: The initials of the first, middle and other names of the author, do not include the last name initials.
- *Author email*: if it is the corresponding author, include the contact email.

3.5.8 Reference type

Record the type of reference from which data were sourced.

- *Book*: Matrices are from a book
- *Book Chapter*: Matrices are from a single chapter in a book
- *Journal Article*: Matrices come from a Journal paper.
    - *Journal name*: Record the standard abbreviation of the journal using the BIOSIS format. For journal abbreviations, click "Full list of journal abbreviation" (https://images.webofknowledge.com/images/help/WOS/A__abrvjt.html). If no abbreviation is available in the list, search past records of the journal from the website or look for it in the left corner of the paper.
    - *Year*: the year the paper was published.
- *Thesis*: Matrices are from a thesis such as PhD thesis, MSc thesis, etc.
    - *Thesis type*: Select the correct option.
- *Report*: Matrices are from a report.
- *Conference*: Matrices are from a conference in forms such as a talk or a poster.
- *Others*: Matrices are from other sources including, but not limited to archives.

3.5.9 Purposes

Select all that apply. Describes the purpose(s) of the study detailed in the manuscript. This information is usually obvious from the abstract and the last paragraph of the introduction. Several purposes are possible per study.

We have categorised the purposes as follows:

- *Comparative demography* - When the authors studied two or more species
- *Spatial demography* - When the authors studied the demography of two or more populations within the same species
- *Abiotic impacts* - When the authors studied the effects of abiotic factors (e.g fire, hurricanes, rain, fertilisers, droughts, etc)
- *PVA* - When the authors report a Population Viability Analysis of an endangered species
- *Species dynamics description* - When there is no actual treatment in the study (Unmanipulated). Instead, the goal was to describe the natural dynamics of a species
- *Interspecific interactions* - When the authors studied the interactions between different species (e.g herbivory, pollinators, inter-specific competition, predation, parasitism, etc)
- *Management evaluation* - When the authors explored the effects of a specific management plan (e.g. to eradicate an invasive species, to preserve an endangered species or assess effects of management on other species).
- *Methodological advancement* - When the novelty of the paper is not an ecological question, but mostly to introduce a new mathematical tool often using already published matrices.

3.5.10   Embargo

If the author does not wish to make personally communicated data available, the data will be under embargo until this date. On that date, an automatic reminder will go out to assigned Compadrino, Admin and Coordinator. The data will not become public until explicit permission has been given by the author. Update the embargo date if the author sets a new date. Leave blank when there is no embargo.

3.5.11   Additional source

Record all additional sources used to construct matrices or to obtain additional ancillary data. Additional sources should be formatted as LastName Journal YearPublished additional information (e.g. Godinez-Alvarez Bot Rev 2003, Schmidt J Ecol 2007 Appendix A1) both here and in the attached file names. Be sure to update both whenever changes are made to the data associated with the publication. Attach the necessary PDF/excel files/documents to the "Attach a document" field.

3.5.12   Additional Notes

This field is intended to record any important issues or general notes about the data for the use of the COMPADRINO team. These entries should only include sensible informative key points that others can understand. Use English only.

3.5.13 Missing data

Data that is not yet available and thus not yet digitized for this publication. Update this information whenever changes are made to populations and matrices associated with this publication.

3.5.14 Treatment

Record all treatment categories used in the study. Include both environmental and experimental treatments. Write "No" if nothing was manipulated.

3.5.15 Within Site Replication

Number of plots or transects sampled at each population and the size description of the plots. For example, 6(4x4m^2) means that in each population, there were 6 permanent plots of 4x4m2. If different populations have different plot sizes, record all of them separated by semicolon. (e.g. A: 2(15ha); B: 3(20ha))

3.5.16 Study start (year)/ Study End (year)

An alert will appear if the start year is greater than the end year, or if the year recorded is greater than 2020

3.5.17 Release version

The Version number of the database where the data were released (no longer applies). Do not enter anything here.

3.5.18 Database source

The institute that collated the data together in a mini database before COMPADRE and COMADRE were developed.

3.5.19 Under "Author contacts" tab

Add *one or more* authors. Use "`Add author contact`" button to add all the authors that falls into the following categories:

- The corresponding author of the publication
- Other authors contacted when the corresponding author of the study was unreachable
- The authors were referred to by the corresponding author
- Authors of additional sources contacted for information

3.5.20   Corresponding author

Select the corresponding author from the list of the authors in the bibliographic information. If the author is not added into the database yet, click "Create New Author" to add a new author.

`Create New Author` and `Add New Author buttons`

- Author email: If the email of the author is provided, add the email address

`Edit author` button in `Manage Authors` tab

- Update the contact information of the author. Put year or "dead" in parentheses to keep track of multiple emails. (e.g. aaa@compadre.org (dead); aaa@compadre.org (2016); xxx@compadre.org (2019))

3.5.21   Date contacted

Record the date(s) when the publication author was contacted asking for information in US date format (mm/dd/yyyy). Authors should always be contacted from the compadrecontact@gmail.com email address rather than a personal email address.

Procedure for lack of response depends on the Year of Publication of the source:

- For papers older than 1990: When the corresponding author does not reply to your third email, or you cannot find their email address (e.g. in old publication) or reach their coauthors, then change the publication status from "Digitizing" to "Review-ready."
- For more recent papers (published after 1990): If the authors (and coauthors) do not reply after the third time you have contacted them, please pass the email address that you have sent them, together with a brief summary of the information you are after and the attached PDF to Owen and Rob. If they do not make successful contact with the authors, they will change the publication status from "Digitizing" to "Review-ready."

If you were not able to find an email address of any of the authors of a paper (and also Rob and Owen don't know one), but that paper has a matrix population model that can be digitised – enter the matrix and all information that are available from the paper. All missing information get an NA and we list the species as "Review-Ready", still enter the missing information into Missing Data field.

3.5.22   Date contacted again

Date(s) of the other attempts to contact this corresponding author. (mm/dd/yyyy; mm/dd/yyyy).   Make another attempt to contact the author again if there has not been a response in 3 months.

3.5.23   Sent email content

This records the information that was requested from the author. Use the same format as for "Missing Data" above. If multiple pieces of information are requested, separate with ";". If the author has been asked for something other than the standard missing information, a short note should be left, e.g. "is the species clonal?"

3.5.24   Author's reply email content

Did the author reply to a request for further information. This should record a brief answer given by the author, or "No" (no response yet). Note: Always change Author's Reply from "`NA`"" / "`Nothing`" to "No" when author has been contacted and update the entry as soon as a reply is received. Always reply to the author with a short thank-you note when a reply is received. Change it to the codes from "`Missing Data`" (M; I; L; S; O) when the information was received from the author. If the author has been asked for information other than the standard information, make a short note of what it is. Also, record the date author replied next to the information received. (e.g. `M; S (11/11/2018)`)

3.6   Protocol for the Population form

3.6.1   Status

This is the status for the population. Please **READ** the website-status protocol to determine the status that should be assigned to the population. * `Digitizing` – compadrino is working/editing the population * `Rejected/Deleted` * `Awaiting info` – compadrino has determined s/he requires more information to complete digitization * `Review-Ready` – the population is ready to be reviewed by a database curator * `Released` – population can be now seen by public users * `Error` – an error has been found in the population and needs to be corrected

3.6.2   Species Accepted

Currently accepted Latin binomial name for the species.

- Search for the correct species name. This might be different from the Latin name the author used (e.g. author used an old Latin name for the species). If you cannot find the correct species name in this field contact a website Admin.

3.6.3   Species Author

The Latin binomial name of the species as reported in the publication. Note that this may not be the same as the current accepted name. If an author uses

a Latin name with a typo, please leave it as is, as this variable represents the name as used by the authors. This applies, as well, if the author(s) used an outdated taxonomic name

- Record as Genus_species.
- When intra-specific names are given by authors, record them as `Genus_species_subsp._subspecies`, `Genus_species_var._variety`, or `Genus_species_cf._colour` form. (e.g. `Sarracenia_alata`, `Sambucus_racemosa_subsp._sieboldiana`)
- Always use an underscore "_" to substitute spaces (e.g. write "`Cryptantha_flava`" instead of "`Cryptantha flava`"). This only applies to the *Species Author* field

3.6.4 Population name

The name of the population where the data for the matrix was recorded. Usually given by the author. When no name provided, use other characteristic information, such as a brief geographic location. 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.

3.6.5 Species common name

The common name for the species. This is pulled automatically from the selected species but can be overwritten if desired/needed.

3.6.6 Ecoregion

Description of the Ecoregion for the study. Follow the link to obtain the ecoregion for the set of coordinates the population is located in: https://patrickbarks.shinyapps.io/biomes/

The options are:

Terrestrial ecoregions

- `TMB` tropical and subtropical moist broadleaf forests
- `TDB` tropical and subtropical dry broadleaf forests
- `TSC` tropical and subtropical coniferous forests
- `TBM` temperate broadleaf and mixed forests
- `TCF` temperate coniferous forests
- `BOR` boreal forests/ taiga
- `TGV` tropical and subtropical grasslands, savannas and shrublands
- `TGS` temperate grasslands, savannas, and shrublands
- `FGS` flooded grasslands and savannas
- `MON` montane grasslands and shrublands
- `TUN` tundra
- `MED` Mediterranean forests, woodlands and scrubs

- `DES` deserts and xeric shrublands
- `MAN` mangroves

Freshwater ecoregions

- `LRE` large river ecosystems
- `LRH` large river headwater ecosystems
- `LRD` large river delta ecosystems
- `SRE` small river ecosystems
- `SLE` small lake ecosystems
- `LLE` large lake ecosystems
- `XBE` xeric basin ecosystems

Marine ecoregions

- `POE` polar ecosystems
- `TSS` temperate shelf and seas ecosystems
- `TEU` temperate upwellings
- `TRU` tropical upwellings
- `TRC` tropical coral

For a more inclusive description of water ecoregions, see http://worldwildlife. org/biomes

3.6.7 Country

Choose the ISO 3 Country Code(s) of where the study took place. You can search by the country's name or the country's code.

3.6.8 Population nautical miles

For oceanic studies, this is the distance in nautical miles from the closest land mass. If the population was studied both in French and Spanish waters 22 miles from land the country code field would be FRA; SPA and the nautical miles field would be NODC-22. If it was only studied in international waters (normally defined as 12 nautical miles) then you would record it as NODC-22.

3.6.9 Continent

Continent where study took place. Select the appropriate one:

- `Africa`
- `Asia`
- `Europe`
- `N America` (includes Canada, USA and Mexico)

- `S America` (incluces the rest of the countries in the Americas not outlined in "`N America`")
- `Antarctica`
- `Oceania` (various definitions for Oceania exist, but here we opted for this one: https://en.wikipedia.org/wiki/List_of_Oceanian_countries_by_population)

3.6.10   Matrix criteria size

Records whether the matrix model is based on size (e.g. stem length/number, body length/mass, etc).

- `Yes` - The matrix population model contains stages based on size.
- `No` - The matrix population model is not based on size.

3.6.11   Matrix Criteria Ontogeny

Records whether the matrix model is based on ontogenetic stage (e.g. seedbank, reproductive, vegetative, dormant, etc).

- `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.

3.6.12   Matrix Criteria Age

Records whether there are any classes that are 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

3.6.13   Coordinates and Altitude

Get coordinates from https://www.gps-coordinates.net/

Estimate the geographic coordinates to the nearest location possible. Check if they reasonably fit the location of the study by looking them up on Google maps or Google Earth. If the terrestrial plant or animal has aquatic coordinates, obtain the coordinates by estimating them from the study site provided by the authors.

When geographic coordinates are not attainable, select an option that corresponds with the study situation. The possible options are `LAB`, `Simulation`, `NC`, `FRANK`, and `AFI.`

- `NC` (Not Calculated): When the matrices are built from mean or pooled data of multiple populations with different study locations. Record each geographic coordinate in the observation of matrix metadata.

- `LAB` (Laboratory): When the study was carried out in a laboratory or under greenhouse conditions where study populations were not affected by the environmental conditions of the actual geographic location.

- `FRANK` (Frankenstein matrix): When the matrices are constructed from two or more sources that were in different conditions. Enter the main geographic coordinates when available. For FRANK studies, specify that it is a FRANK study in the observation with a brief summary of the methods. (e.g. FRANK; Survival rates estimated from the field study, but fecundity values obtained from a greenhouse study).

- `Simulation`: theoretic studies that are not or are loosely based on a real population.

- `AFI` (Asked for it): contacted the author for the information but the author hasn't replied yet.

3.6.14 Coordinates Format

Select the adequate format for the coordinates of the population. Options include:

- `Degrees, minutes and seconds (DMS`
- `Decimal degrees`
- `LAB`
- `Simulation`
- `NC`
- `Frank`
- `AFI`

3.6.15 Latitude Degrees / Latitude Minutes / Latitude Seconds

Record the degree (0-90), minute (0-60) and second (0-60) values of the locations of the matrix.

Note: if there are no seconds (or minutes) given, you should record as NA rather than 0.

3.6.16 Latitude N or S

Cardinal direction * `N` - North * `S` - South

Note: Cardinal direction might be given with a "+" or "-" whereat "+" means North and "-" means South.

3.6.17 Longitude Degrees / Longitude Minutes / Longitude Seconds

Records the degree (0-90), minute (0-60) and second (0-60) values of the locations of the matrix.

3.6.18 Longitude W or E

Cardinal direction * `W` - West * `E` - East Note: Cardinal direction might be given with a "+" or "-" wherby "+" means East and "-" means West.

The following link is useful for converting between different ways of presenting lat/long information: http://www.earthpoint.us/Convert.aspx

3.6.19 Altitude

Altitude of studied population in m = Height above sea level of specific population in meters e.g. `208` (without "m").

Note: Altitude is either given in the paper, by author communication, or we look it up with *Google Maps*. If you enter the coordinates into *Google Maps* you'll see the altitude at the bottom right corner of the window. If you scroll around in *Google Maps* with your mouse the altitude will change, so pay attention that you read the altitude directly after entering the coordinates without scrolling around!

3.6.20 General notes on location information

If you don't have a detailed description of the geolocation-coordinates enter the following sentence into the Observation-Column: The GPS coordinates were approximated to the closest geographic location described in the reference. Sometimes the Lat and Long given in the paper do NOT match the study area. In this case enter the following in the `Observation`-Column: GPS co-ordinates cited exactly as reported in original paper. These do not match written study site description.

Pay attention to the means: We don't calculate ranges for mean matrices. In that case, choose the option `NC` for "not calculated" in the "`Coordinates format`" field.

3.7 Protocol for the matrix metadata form

### 3.7.1 Composition

Indicates the type of matrix population model

- `Individual` - single population studied over a single time period, for example, over one year
- `Mean` - a mean matrix was calculated by averaging across more than one population studied over a single time period, for example, over one year
- `Pooled` - a matrix was calculated by pooling data across more than one population studied over a single time period, for example, over one year
- `NDY` - not done yet / unset

3.7.2 Seasonal

A matrix population that does not describe a full annual transition, but rather a seasonal transition.

- Specify the seasons in the "Start Season" and "End Season" fields of the matrix form.

A matrix can be seasonal even if it is constructed over several years e.g. spring in years 2000-2012. Make sure you detail that there are multiple years data in the observation column.
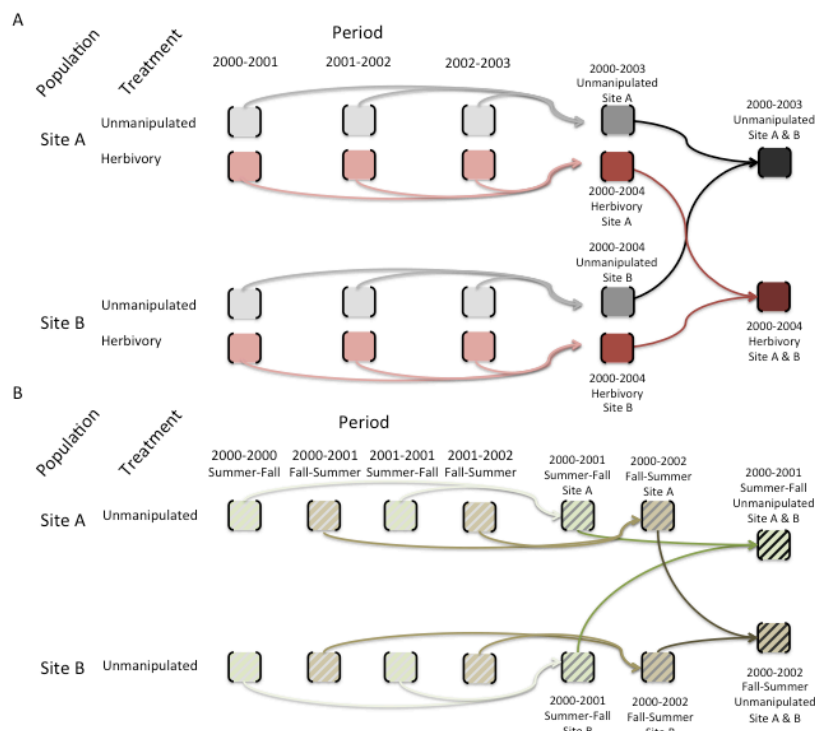
This figure is a schematic showing how mean population matrix models are derived from all available population matrix models in a study. A. For annual population matrix models (those that follow population dynamics from one year to the next, or sometimes more – some tree species are studied with a 5-year

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).

3.7.3   Dimensions

Select the dimension of the matrix.

- Expand the drop-down menu and type in the number to navigate faster.

3.7.4   Class Author

The stage description as indicated in the source by the publication author.

- If author has used abbreviations, include the explanations of it.
- Use periods (.) for decimals instead of commas (,).

- Always include stage names, stage categories and units. (e.g. S (Seedlings): 1.5 cm   stem height < 3.5 cm, Adult: age 5+ yrs, etc.).
- Keep the authors' descriptions even if the paper is written in another language.
- Always add units (when possible) (e.g. 0-9 cm DBH, not 0-9 DBH)
- Be clear in what is described (e.g. 0-9 cm DBH/ height, not only 0-9 cm)
- Make sure that there is space between the number and the unit (e.g. 1.45 cm, not 1.45cm)
- Also write down abbreviations - if the author used them frequently (e.g. if the author explains his stages in the text and after that only uses "S1, S2, J1, J2, J3, A, R" write them down, too (= s1 = dormant seeds)

3.7.5   Class Organized

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

- 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 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.

3.7.6 Start Year

First year of study. Year t in annual matrix population model that describes population dynamics from time t to t+1. e.g. `1995`.

3.7.7 End Year

Last year of study. Year t+1 in matrix population model that describes population dynamics from time t to t+1. e.g. `2001`.

3.7.8 Start Month/End Month

Record the start and end months if provided

- For monthly transition matrices, record the start month and end month (m to m+1).
- If demographic information was collected during a period of time, please state the period separated by a hyphen (e.g. `6-8` for June through August) in observations.

3.7.9 Start Season/End Season

Record the start season and end season for the seasonal matrices.

3.7.10 Census type

Choose the type of sampling census the authors use (see Kendall et al. 2019 for additional details). The authors may have specified if they sampled before or after reproductive events. If this is not provided in the publication, please contact the corresponding author.

- `Prebreeding` – censused the population just before breeding occurred.
- `Postbreeding` – censused the population just after breeding occurred.
- `NDY`
- `AFI`
- `NA`

3.7.11   Projection interval

The time step (periodicity) of the transition matrix. Be aware this is NOT equal to sampling frequency. If a matrix has 6-month projection interval, it would be `0.5` since it is 0.5 years. In the same way, if a matrix has 5-year projection interval, it would be `5`.

3.7.12   Split

Divisible if enough information is given to divide the survival and reproductive portions of the matrix. The matrix is divisible when there is no reproductive portion in the A matrix. Usually, indivisible matrices contain mixed cells in A matrix that cannot be divided into U, F, and C matrices due to insufficient information.

- An **A** matrix should ALWAYS be recorded, even if no reproduction data was available.
- `Divided` - The matrix population model A has successfully been divided into the process-based sub matrices 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 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

This is how a divided matrix could look like (species: *Troglodytes troglodytes*):

CELL COLOR   No Highlighting

Matrix

Age 0-1: active	Age 1-2: active	Age > 2: active
1.01	0.88	0.78
0.43	0	0
0	0.52	0.46

Survival Matrix

Age 0-1: active	Age 1-2: active	Age > 2: active
0	0	0
0.43	0	0
0	0.52	0.46

Fecundity Matrix

Age 0-1: active	Age 1-2: active	Age > 2: active
1.01	0.88	0.78
0	0	0
0	0	0

Clonal Matrix

Age 0-1: active	Age 1-2: active	Age > 2: active
0	0	0
0	0	0
0	0	0

This is how an indivisible matrix could look like (species: *Syngonanthus nitens*):

Edit Matrices   ×
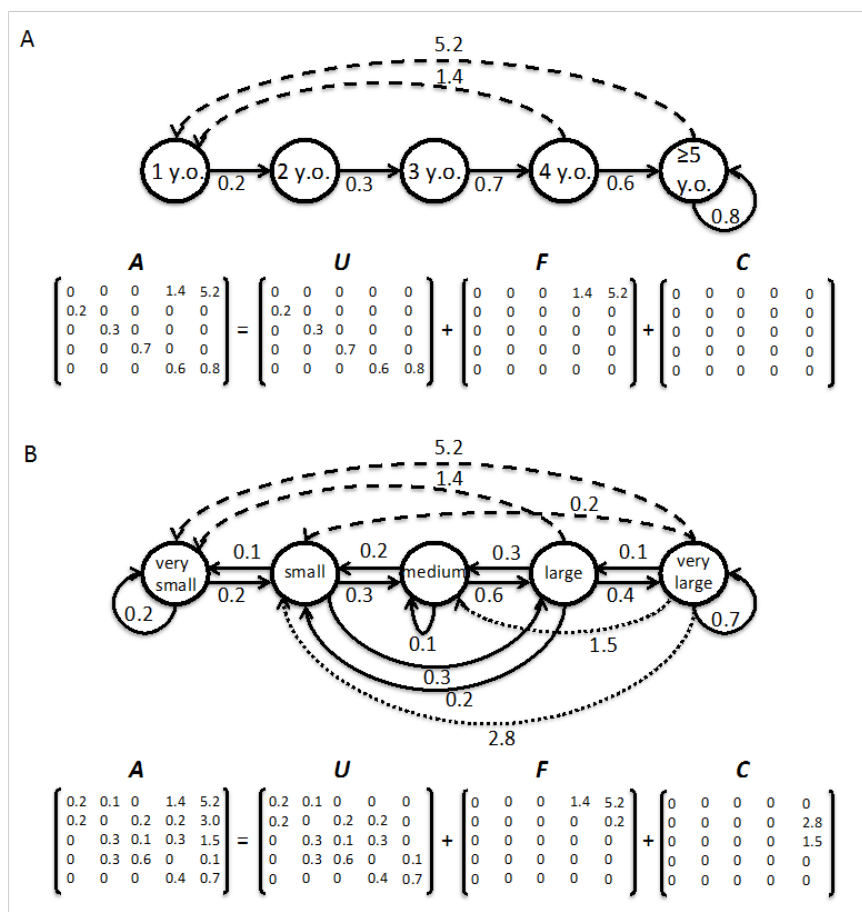
CELL COLOR | No Highlighting   ▾

Matrix

Seedling: single rosettes diameter < 2cm: active	Small ramets: rosettes diameter <2cm, that are part of larger genet: active	Adults: single rosettes: active	Adult ramets 1: adults in genet with 2-4 ramets: active	Adult ramets 2: adults in genets with more than 4 ramets: active
0.3684	0.0667	0.2875	0.3099	0.0271
0.0526	0.5333	0.1386	0.1816	0.1034
0.5263	0.0667	0.6544	0.0581	0.0051
0.0526	0.2	0.6386	0.753	0.0034
0	0.1333	0	0.5714	0.7

The diagram below shows the 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). This model includes "retrogression" or shrinkage in addition to "progression" or growth. 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 Split).

A

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

B

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

3.7.13 Treatment

A brief description of experimental treatments of the population intentionally imposed by humans.

- Write "Unmanipulated" if no experimental, human-imposed treatment was applied. Natural events of non-intentional occurrence should be

recorded in the treatment in the publication form and in observations column here, especially if they are used as an explanatory factor in the study.

- For example, population dynamics could be compared in disturbed vs. undisturbed areas. In these cases MatrixTreatment should be recorded as "Unmanipulated", a note should be added in Observation and the name of the population should include a clear indication of the conditions. e.g. "Population 1 - disturbed", "Population 2 - undisturbed".

3.7.14   Captivity

Select the following options:

- `W` – Wild: Studied in natural conditions
- `C` – Captive: studied for most part of the life cycle of the species in a botanical garden, greenhouse, laboratory, etc.
- `CW` – Captured from wild: if the species was taken from a wild population but studied in labs or garden.
- `C;W` – Mixed studies that incorporate vital rates from both captive and wild populations

3.7.15   Studied Sex

Indication of the sex(es) considered when constructing the matrix population model.

- `M` - Studied only males
- `F` - Studied only females
- `H` - Hermaphrodites
- `M/F` - Males and females separately in the same population matrix model
- `A` - All sexes together
- `NA` - There are no distinct sexes (e.g. a virus)

3.7.16   Sources

Where the data was obtained to construct the matrix. If the data is from an appendix that is attached at the end of the original publication, record it as "Paper" and NOT as "Online Appendix". "Online Appendix" is only for appendices that are in separate documents. Reselect the option to unselect it. If sources other than those below are used, a short but useful description should be given. The aim of this field is that everybody can find the related information ASAP e.g. if something needs to be fixed.

- Author communication - Matrices derived from personal communication with author.
- Paper - Matrices derived from a published manuscript

- Old matrices - Matrices derived from an old matrices file (no more detailed description is available)
- Database - Matrices derived from one of the various databases that have been merged into COMPADRE/COMADRE see Database above.
- Book: Matrices come from a book or book chapter
- Online appendix: The matrices come from an online appendix that is not attached at the end of the original publication. Most be a separate document.
- Calculated: compadrino self-calculated the matrix from a life table or other data sources. Include what the original data type is (e.g. life table, integral projection model, individual-based model, etc.) and information on how you calculated the matrix in the observation field.
- Altered: the original published matrix was altered for accuracy. Include specific reason(s) for why the changes were made (e.g., seed bank issue, typo corrected through author communications, recalculating from pre-breeding to post-breeding, etc.) and information on how you calculated the matrix in the observation field.

3.7.17   Fec (Matrix fecundity)

This records whether fecundity was measured at all for this matrix model. This will help distinguish whether a 0 in the fecundity transition(s) indicates no fecundity or fecundity was not measured. We include fecundity, recruitment, or sexual reproduction here ignoring the technical differences.

The rationale for this is that a fecundity (in the F matrix) may be recorded as 0 either because it was not measured, or because reproduction was estimated to be zero. As a general rule, if any values for fecundity (e.g. in the F matrix) are non-zero, then Fec should be recorded as 'Yes'. If all values are zero, then you should check the paper carefully to check whether this is because no fecundity was detected, or whether no attempt was made to measure fecundity. There are some technical differences between fecundity, sexual reproduction, and recruitment. For the purposes of Fec these differences do not matter - if estimates fecundity, recruitment or sexual reproduction are collected, Fec = "`Yes`".

- `Yes` - fecundity was measured for the matrix
- `No` - fecundity was not measured for the matrix
- `NA` - it is not known whether fecundity was measured for the matrix

3.7.18   Observations

Add any useful information about the population matrix. (e.g. geographic co-ordinates of pooled populations, unusual estimations or treatments, etc.)

Some standardised sentences:

- The GPS coordinates were approximated to the closest geographic location described in the reference. (see Geolocation)

- GPS co-ordinates cited exactly as reported in original paper. These do not match written study site description. (see Geolocation)
- Geolocation / Altitude from descriptor: (if no matrices are entered yet but a geolocation information are available)
- Matrix is calculated with stochastic model
- X years since last fire
- Calculated with demographic data from Author JournalName YearPublished
- Data is from a greenhouse experiment
- Matrix Class Author is not clarified

3.7.19   Create/Edit Matrix

The population matrix models. Including the overall matrix (A) and, where possible, the sub-matrices U (survival), F (fecundity), C (clonality). A range of numeric values is possible in each element of the matrix. Sometimes only a U matrix is available because data on reproduction were not collected. In these cases the U matrix should be used as the A matrix, and a note taken in Observations (see the rest of this section for more details). i.e. an A matrix.

- The A matrix - The overall matrix including all processes. The A matrix can usually be "split" into consituent processes represented by the U, F and C sub-matrices such that A = U + F + C. Transition probabilities are constrained between 0 and 1. Reproductive values are always positive but have no upper bound.
- The U matrix - The population sub-matrix model U (a sub-matrix of A), describes the survival-dependent dynamics of a population under the conditions described above. Values are constrained to be between 0 and 1. These include only survival-dependent vital rates (no sexual or clonal reproduction). If only the U matrix is available, it should also be used as the A matrix (see "The F matrix" below).
- The F matrix - 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). Sometimes sexual reproduction is not recorded (e.g. because it was not the focus of the study). This should be noted in observations and recorded using Fec (see elsewhere in this guide). Where possible, the "known unknowns" should be recorded as NA. For example, if you know that the "adult class" reproduces sexually, but simply was not estimated you should record NA for that element's fecundity, and "No" for Fec.
- The C matrix - 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. Where there is no clonality, this matrix should be included, but will be all 0 values.

The values for the U, F and C matrices are only digitised for "Divided" matrices (indicated in the field Split)

3.7.20   Notes for Matrix construction

- Each cell represents transition from the column stage to the row stage.
- You may enter a number or NA (Not Available).
- For survival rates, the values are restricted between 0 and 1. The sum of the survival rates from a single stage should not exceed one.
- Reproductive values do not have upper bounds but should not be a negative number. This also applies to compound cells where survival rates and reproductive values are combined.
- "0" should not be colored
- Values across the U, F, and C matrices should add up to the A matrix. If there is not enough information to divide compound cells in A matrix, the matrix should be marked "Indivisible" and only A matrix should be recorded.

3.8   Protocol for reporting an error

3.8.1   Report Error button

Populations in "Review-ready" and "Released" status may be found to have an error. When the checker (Admins or coordinators) changes the status from "Review-ready" to an "error", they must report the error through the `Report Error` button on the population page. Any user can also report the error through the `Report Error` button on the population page. The Admins, Coordinator, or the person assigned to fix database errors ("Error Handler"; a role selected through a tick box in Manage Users table) will be notified by email. If it is determined to be an actual error, the Admin/Coordinator/Error Handler will change the status to "Error." If it is not a simple error and more information is needed, the publication can then be changed to "Awaiting info" status and assigned to a Compadrino, at which point the assigned Compadrino may continue to work on it. If the error is fixed by the Admin/Coordinator/Error Handler, the publication can go to "Released" status (since this will serve as the check for the publication if it was previously "Review-Ready," or if it was previously "Released" it can return to the "Released" status after the error is fixed).