

# COMPADRE/COMADRE Data Digitization Protocol

*March 27 2018*

## Contents

Preamble . . . . .	2
General information . . . . .	3
<b>Values used throughout the database</b>	<b>3</b>
<b>Priority variables</b>	<b>4</b>
<b>References</b>	<b>5</b>
<b>DESCRIPTOR FILE</b>	<b>5</b>
Student . . . . .	5
Notes . . . . .	6
VersionRelease . . . . .	6
Transferred . . . . .	6
MissingData . . . . .	7
DateDigitalization . . . . .	7
Embargo . . . . .	7
StudyChecked . . . . .	7
StudyDifficulty . . . . .	8
OrganismType . . . . .	9
SpeciesAuthor . . . . .	10
CommonName . . . . .	10
Taxonomy columns . . . . .	11
Authors . . . . .	11
Journal . . . . .	11
YearPublication . . . . .	13
CorrespondingAuthor . . . . .	13
Email . . . . .	13
Contacted . . . . .	13
ContentEmail . . . . .	14
Reply . . . . .	14
DOLISBN . . . . .	14
AdditionalSource . . . . .	14
Database . . . . .	14
IUCNStatus, InvasiveStatus, ConservationStatusUS . . . . .	15
Ecoregion . . . . .	15
GrowthFormRaunkiaer (P) . . . . .	17
Purposes . . . . .	18
Treatment . . . . .	19
StudyStart . . . . .	19
StudyEnd . . . . .	19
StudyDuration . . . . .	19
AnnualPeriodicity . . . . .	19
NumberPopulations . . . . .	20
WithinSiteReplication . . . . .	20
MatrixCriteriaSize . . . . .	20
MatrixCriteriaOntogeny . . . . .	20

MatrixCriteriaAge . . . . .	20
<b>STACKED MATRIX FILES</b>	<b>21</b>
EnteredBy . . . . .	21
EnteredDate . . . . .	21
Source . . . . .	21
SpeciesAuthor . . . . .	21
StudiedSex . . . . .	21
MatrixComposite . . . . .	22
MatrixTreatment . . . . .	23
MatrixCaptivity . . . . .	23
YEAR/SEASON/MONTH . . . . .	23
MatrixStartSeason . . . . .	24
MatrixStartMonth . . . . .	24
MatrixEndYear . . . . .	24
MatrixEndSeason . . . . .	24
MatrixEndMonth . . . . .	25
MatrixPopulation . . . . .	25
Geolocation columns . . . . .	25
MatrixSplit . . . . .	28
MatrixFec . . . . .	31
Observation . . . . .	31
MatrixClassOrganized . . . . .	31
MatrixClassAuthor . . . . .	32
MatrixClassNumber . . . . .	32
matU/matF/matC/ . . . . .	32
Color coding . . . . .	33

## 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 Excel spreadsheets in a structured way. Wherever there are doubts about aspects of the data source, the team attempts to contact the researchers to resolve them. There are two types of Excel file in COMPADRE/COMADRE: (1) The DESCRIPTOR FILE(s) which contain metadata allowing the interpretation of the matrix data, and it pertains to information at the species/study level; and (2) the STACKED MATRIX FILE(s) 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](http://www.compadre-db.org).

This document outlines the way that the digitisation team enters data into the Excel sheets. 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 [compadre-contact@demogr.mpg.de](mailto:compadre-contact@demogr.mpg.de). Specific questions can also be directed to Rob Salguero-Gómez ([salguero@demogr.mpg.de](mailto:salguero@demogr.mpg.de)) or Owen Jones ([jones@biology.sdu.dk](mailto:jones@biology.sdu.dk)).

## General information

- This protocol is intended for both COMPADRE and COMADRE. Some variables pertain to *only* one or the other database and are indicated as such in the text. Specifically “P” pertains only to plants (COMPADRE), whereas “A” pertains only to animals (COMADRE).
- PDFs of new publications emailed to the COMPADRinos are to be saved under the “Literature” sub-folder of COMPADRE or COMADRE in the N://COMPADRE & COMADRE directory of the Max Planck Institute for Demographic Research (MPIDR) server. Immediately after saving the PDF, a new entry must be opened with the “SpeciesAuthor”, “Authors”, “Journal”, and “YearPublication” in both the descriptor and stacked file.
- If you find additional sources *always* save them and note them in “AdditionalSource”. All pieces of information extracted from PDFs must be highlighted directly on the PDF in yellow.
- 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 Species Descriptors file.
- 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 an additional species entry, followed with a numeric subindex following “SpeciesAuthor” (e.g. “Pinus\_nigra” and “Pinus\_nigra\_2”).
- Authors must only be contacted from the official email: compadrecontact@gmail.com.
- Latin names of species for variable “SpeciesAccepted” must be checked using reliable sources (e.g. Catalogue of Life <http://www.catalogueoflife.org/>, Wikipedia <https://en.wikipedia.org/>) for spelling mistakes. If an author uses a Latin name with a typo, please leave it as is, as the variable “SpeciesAuthor” (below) represents the name as used by the authors.
- The element-by-element mean values of unmanipulated (control) and treatment matrix population models are to be kept separate (i.e. never calculate mean matrices from a mixture of unmanipulated and treated matrices). Furthermore, mean matrices should never mix different treatments (e.g. do not calculate mean matrices for matrices with different grazing regimes). However, mean matrices *should* be calculated for unmanipulated matrices, and for matrices that have received the same treatments..
- When digitising matrices, only colors from the color palette, which is given inside the Excel file under the tab “Color palette”, can be used in the Stacked Matrix file. Where matrices can be split, please always include an “A” matrix as well as “U”, “F” & “C” matrices (for explanation see here in the stacked matrix file. Present matrices from left-to-right in the order A, U, F, C. **Always** split matrices when possible.
- The COMPADRINOs have writtenguides to plant growth types and animal classes which we categorise species to. They are here (plant growth type) and here (animal classes).

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.

---

## 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. For these data we do not calculate averages or ranges but instead enter **LAB** (not **NA**).

**FRANK** - Frankenstein matrix - Frankenstein matrices are MPMs where the author parameterised their model using demographic data from 2+ other papers/sources. These MPMs should be clearly indicated with “FRANK” in the observation column. Furthermore, for these matrices, temporal and geographic information is not useful for comparative analyses and therefore geographic location data, time information (start/end year, season etc.) should be recorded as “FRANK” (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**).

---

## Priority variables

The following is a list of variables that the COMPADRino team digitises to the best of its effort, contacting the author when the information is not available in the manuscript when necessary.

- SpeciesAuthor
- Authors
- Email
- Journal
- YearPublication
- DOI.ISBN
- Ecoregion
- GrowthFormRaunkiaer (only relevant for COMPADRE)
- Purposes
- StudyDuration
- AnnualPeriodicity
- MatrixStartYear
- MatrixEndYear
- MatrixStartMonth
- MatrixEndMonth
- NumberPopulations
- StudiedSex
- Geolocation and Altitude
- Continent
- Country
- The full taxonomy (Kingdom, Phylum, etc.)

## 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., Viereg, 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. doi: 10.1111/1365-2745.12334
- 

## DESCRIPTOR FILE

### Student

First name of COMPADReino

Possible values

- Alex
- AlexNH
- Anne
- Christoph
- Claudia
- Elisa
- Erik
- France
- Gabriel
- Gesa
- Hendrik
- Jakob
- Janek
- Julia
- Rob
- Sandra
- Stefan
- Tara
- Veronika

\_\_\_\_Information for the Stacked File:\_\_\_\_

The color coding in the Student column of the Stacked File means the following:

- RED There is some problem with this entry, the student who entered the species is informed and works on the species/species again
- LIGHT ORANGE The *Transferred* column has been checked. The entry in that column is correct. If something was missing (see *MissingData*) author has been contacted again. Everything else is fine.

- **DARK ORANGE** The *Transferred* column has been checked. The entry in that column is correct but a problem occurred which is explained in the *Notes* column (e.g. no mail address of author found).
- **GREEN** The entry has been checked twice. No faults have been found. Species/study is ready to go online.

Note: Excel might color the cell (in green or orange etc.) automatically when entering a new species. Please always change the fill color for all new entries into “no fill”. The students who do the error-checking will change into orange/red/green when necessary.

---

## Notes

This column is intended to record any important issues or general notes about the data for the use of the COMPADRE/COMADRE team. It should be ensured that these entries include only sensible informative key points/sentences that also others can understand. As with all variables in the database, use **English** only.

---

## VersionRelease

The Version number of the database where the data is/will be released. e.g. 3.0.0, 3.1.0. Where the data are not yet to be released (for example, because the data are incomplete) an “X” is used in place of a version number. For all future digitization, an “X” should be placed in the version record. These will be replaced with an appropriate version number by Owen or Rob only at a later date when a final check is done to release more data periodically.

---

## Transferred

Record the status of transference of matrices into the stacked file.

**Yes** – All variables (geolocation, individual matrices) for which the author had data have been transferred. It’s impossible to get missing data. All available matrices have been transferred. No more matrices are available.

If there is no email address for an author of a paper, but that paper has a matrix population model that can be digitized – enter the matrix population model to the stacked file and enter all information that is available from the paper into the descriptor, all missing information get an NA and we list the species as “transferred-Yes”

**Partly** - Only sections have been transferred. Some information is missing, which is reported in the column “Missing data” below. At least one matrix is transferred into the stacked file. Check out E-Mail - Reply under “Contacted” to find out what to do when author does not answer.

**No** - Nothing has yet been transferred, but the COMPADRino has started working on this species.

If the matrix population model(s) cannot be digitized, that paper reference should be deleted from both the species descriptor file *and* the stacked matrix file. Make a note of this deleted paper, using the normal format, in a NEW PAGE of the species descriptor file, called “deleted papers”, make a short note why it was deleted (for example no actual matrix population model was used).

---

## MissingData

What data are still missing?

- **M** - All matrices missing
- **I** - Only individual matrices are missing
- **L** - Labels are missing (e.g. information about the populations and years)
- **S** - Information about study site missing (e.g. country, geolocation, WithinSite etc)
- **O** - Other priority information missing
- **No** - Everything is digitised

If more than one piece of information is missing, these are separated with “;” (without white-space in between), e.g. “I; S; O”.

---

## DateDigitalization

This column records the date when the COMPADRINO started working on this species. Specifically when the first matrix population model of that study was digitised in stacked file. Dates should be recorded in the following (US) date format: mm.dd.yyyy, e.g. 12.25.2015 for 25th Dec 2015.

---

## Embargo

If the author does not wish to make personally communicated data available, the data will be under embargo until the following date. At that point, an automatic reminder will go out to the author. The data will not become public until explicit permission has been given by the author. The date should be recorded in US date format (mm.dd.yyyy), or as NA when there is no embargo.

---

## StudyChecked

This variable records how much the MPM has been checked by us. It is a numeric variable that gets incremented by +1 each time the data in a set of matrices for a study are checked. The first time the data are entered, StudyChecked is 0 (it has been *entered* not checked). Then, if for whatever reason we revisit the paper/data to check details, we add 1 to the number. e.g. an study that has been entered, then checked twice should have a value of 2.

- 0 – Digitized only
- 1 – Error checked once since digitization
- 2 – Error checked twice since digitization
- 3 – Error checked three times since digitization
- 4 – Etc.

## NOTE -

In the files we are already using there is some colour coding that we can carry on using in the *Student* column. In this column we use ORANGE color to give some indication of check status:

We will now modify this to two shades of orange:

- LIGHT ORANGE The *Transferred* column has been checked. The entry in that column is correct. If something was missing (see MissingData) author has been contacted again. Everything else is fine.
  - DARK ORANGE The *Transferred* column has been checked. The entry in that column is correct but a problem occurred which is explained in the *Notes* column (e.g. no mail address of author found).
- 

## StudyDifficulty

This indicates how difficult the MPM information was to extract from the paper. It is recorded on a 3 point scale (values of 1, 2, 3), with 1 being easy to 3 being hard. How do we decide on difficulty?

The points below give the characteristics of “easy”, “medium” and “hard” studies to digitise. These are not exhaustive, but should allow you to roughly categorise the study you are working on to one of the three levels.

### 1) EASY

- MPMs have a single vital rate per matrix element
- Not a “Frankenstein matrix”
- The MPMs freely available to digitize and it was 100% clear what population, year and/or treatment it was referring too.
- No clonality present.
- Easy and clearly described life cycle.
- Matrix information present in the paper in the form of a matrix/table.
- It was NOT necessary to contact the publication authors for clarification or more data.

### 2) MEDIUM

- Multiple vital rates per matrix element but calculations of each element from vital rates is clearly described
- Frankenstein matrix from clearly distinguishable sources
- The matrix was freely available to digitize and it was 100% clear, but the population, year and/or treatment each matrix relates too was unclear but manageable from the literature
- Clonality was present and the clonal components were clearly described within the literature and the matrix could therefore be split accurately into component matrices (U, C, F)
- Complex life cycle but clear definition of life cycle provided
- Matrix not present in the paper and therefore matrix elements needed to be calculated from vital rates and life cycle diagrams

### 3) HARD

- Multiple vital rates per matrix element but calculations of each element from vital rates is not described.
- Frankenstein matrix with undistinguishable sources.
- The matrix was freely available to digitize and it was 100% clear, but the population, year and/or treatment each matrix relates too was undeterminable.
- Clonality was present within the matrix but the clonal components were not clearly described within the literature and therefore the matrix could not be split into component matrices (U, C, F).
- No information on life cycle available / poor description of life cycle.
- Matrix elements had to be calculated from tables.
- Vital rates were modeled as a function of other factors, for example, density dependence, weather.
- It was necessary to contact the publication authors for clarification or more data.



---

## OrganismType

### Plants (COMPADRE)

In **plants** this entry records the plant/algae type, based mainly on architectural organisation. This information is intended mostly for internal use and reflects how our PDF library is organized on our servers, as well as the tab in the Matrix Stacked file where the matrices are stored.

Possible values are:

- Algae
- Fungi
- Annual
- Bryophyte
- Epiphyte
- Fern
- Herbaceous perennial
- Liana
- Palm
- Shrub
- Succulent
- Tree

The species should be assigned to one of these possible values using the description of plant growth type provided by the author and by reference to other sources (e.g. other publications).

Note about the Annual category: Typically plant ecologists also refer to pseudo-annual species as being “biennials”. The main difference between the annuals and biennials is that annual species complete their life cycle (are born, grow, reproduce and die) within a single year, whereas biennials have the possibility to forget of reproduction in the first year and expanding their life cycle to their second year, when reproduction is fatal.

### Animals (COMADRE)

For animals, with the exception of humans, taxonomic class is recorded in both the Descriptors file and the Stacked Matrix file. For humans (*Homo sapiens*) we record “Homo\_sapiens”. Again, this information is intended for internal use and reflects how our PDF library is organised on our servers, as well as the tab in the Matrix Stacked file where the matrices are stored.

Possible values include:

- Actinopterygii
- Amphibia
- Anthozoa
- Arachnida
- Ascidiacea
- Aves
- Bivalvia
- Branchiopoda
- Cephalaspidomorphi
- Chondrichthyes
- Chromadorea
- Clitellata
- Demospongiae

- Diplopoda
- Echinoidea
- Eurotatoria
- Gastropoda
- Gymnolaemata
- Malacostraca
- Mammalia
- Maxillopoda
- Merostomata
- Onychophorida
- Ostracoda
- Polychaeta
- Reptilia
- Secernentea
- Spirochaetes
- Homo\_sapiens
- Virus
- Bacteria

Contact Owen or Rob if you find something that doesn't fit into these groups.

---

## SpeciesAuthor

The Latin binomial name of the species as reported in the publication. Note that this may not be the same as the currently accepted name. This should be recorded as "Genus\_species". It should *always* include the genus and the species names. Where the paper authors have given infra-specific names information on variety and subspecies should also be added (see below).

**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"

The word following "var.", "subsp." or "cf." must not be capitalised. Excel will try to capitalise it via auto-correct because it follows a period punctuation.

Always use an underscore "\_" to substitute spaces (e.g. write "Cryptantha\_flava" instead of "Cryptantha flava").

If two or more different demographic studies exist for a particular species, these should be digitised with concatenated numbers after the genus\_species starting with "2" onwards (e.g. "Rosa\_sempervirens", "Rosa\_sempervirens\_2", "Rosa\_sempervirens\_3"). Note that the first hit does not get a "1" (e.g. "Rosa\_sempervirens" instead of "Rosa\_sempervirens\_1")

Exactly the same naming convention **must** be used in both the Species Descriptor and Stacked Matrix files.

---

## CommonName

Records the common name as used in the publication. Separate multiple names with a semi-colon (";"). If there is no common name used in the publication, record "NA".

## Taxonomy columns

This series of columns includes taxonomic information. Please obtain (and check) this information using The Plant List (<http://www.theplantlist.org>) for plants and the Catalogue of Life (<http://www.catalogueoflife.org/col/>) for animals. For algae, you can use <http://www.algaebase.org>.

- Kingdom - Kingdom to which species belongs.
- Phylum - Phylum to which species belongs e.g. Magnoliophyta.
- Class - Class to which species belongs e.g. Magnoliopsida.
- Family - Family to which species belongs e.g. Polygonaceae.
- Order - Order to which species belongs e.g. Caryophyllales.
- Genus - Genus used in SpeciesAuthor e.g. Chorizanthe.
- SpeciesAccepted - Currently accepted Latin name.
- DateOfCheck - the date when the taxonomy was checked.
- AngioGymno Whether species is an angiosperm or a gymnosperm (record Angiosperm or Gymnosperm). If it is neither angiosperm nor gymnosperm fill in NA. Use NA for **all** animals.
- DicotMonoc - Whether species is a dicot or monocot (record Eudicot or Monocot). If species is neither dicot nor monocot (i.e. if it is not an angiosperm) fill in NA. Use NA for **all** animals.

For COMPADRinos to get an idea of what kingdom, class etc means:

---

## Authors

Last name of all authors on the publication. Separate multiple names with “;”. Note that **all** authors named in the publication should be recorded in the same order as in the publication. E.g. for the Compadre-paper: Salguero-Gómez; Jones; Archer; Buckley; Che-Castaldo; Caswell; Scheuerlein; Conde; Baudisch; Brinks; de Buhr; Farack; Gottschalk; Hartmann; Henning; Hoppe; Römer; Runge; Zeh; Vieregge; Altwegg; Colchero; Dong; Hodgson; de Kroon; Lebreton; Metcalf; Neel; Parker; Takada; Valverde; Vélez-Espino; Wardle; Franco; Vaupel.

---

## Journal

Record the name of the journal from which data were sourced. If the data were sourced from a book, thesis, report, or conference talk/poster, record that as appropriate. For journal abbreviations see list here: [https://images.webofknowledge.com/images/help/WOS/A\\_abrvjt.html](https://images.webofknowledge.com/images/help/WOS/A_abrvjt.html).

If no abbreviation is available in both of those list, take a look into the paper. Often there are abbreviations in the left corner of the paper or at the first page.

- Journal Abbreviation (e.g. J Ecol for Journal of Ecology) - 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. Note: If our source is a book name the author of the related chapter in Authors, not the author of the book.
  - PhD thesis - Matrices are from a doctoral thesis.
  - MSc thesis - Matrices are from a masters thesis.
  - Report - Matrices are from a report.
  - Conference talk - Matrices reported in a conference talk.
  - Conference poster - Matrices reported in a conference poster.
-

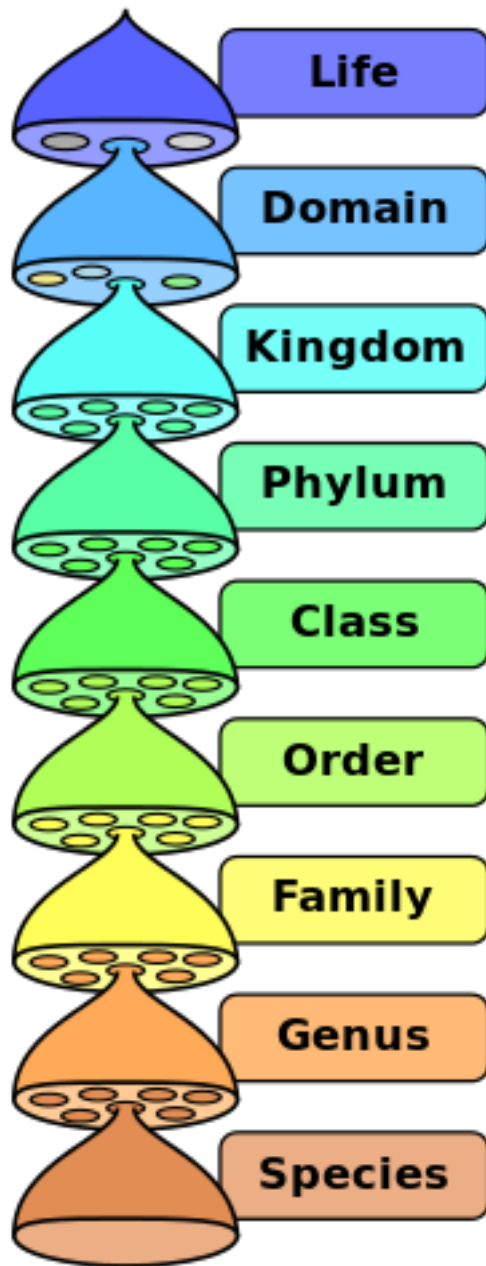


Figure 1: Taxonomy

## YearPublication

Year of publication of the source publication. This should be recorded as YYYY (e.g. 2002).

---

## CorrespondingAuthor

Last name of corresponding author(s). Note that the corresponding author may not necessarily be the first author. When no author is named as the corresponding one in the paper we use the first author. Separate multiple names with “;” when pertinent.

---

## Email

For communication with the corresponding author record the email address, followed by the year in which the email address last worked or when the paper has been published to distinguish between old and new addresses. If email address doesn’t work anymore enter “dead” instead of the year (note that this does not mean that the author is dead!). The author’s email address can be obtained from the publication, from related papers, from internet searches, and personal communication etc.

e.g. jones@biology.sdu.dk (2015)

---

## 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 YearPublication 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 “Partly” to “Yes” in Transferred column.
- 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 “Partly” to “Yes” in Transferred column.

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 to the Matrix Stacked file and enter all information that are available from the paper into the Species Descriptor file, all missing information get an NA and we list the species as “transferred-Yes”, still enter the missing information into MissingData

---

## ContentEmail

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?”.

---

## Reply

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 **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.

---

## DOI.ISBN

This entry records the Digital Object Identifier (DOI) or International Standard Book Number (ISBN) codes that identify the source material. In more recent papers you may find the DOI somewhere at the beginning of the paper. You may also search for it in the Web of Knowledge.

Use the DOI resolution service at <http://dx.doi.org/> to confirm the DOI. Do this by appending the DOI after this web page. For example, for the DOI “10.1111/1365-2745.12334” –

<http://dx.doi.org/10.1111/1365-2745.12334>

This should take you to the journal page for the article.

Only enter the DOI itself into this field do not add additional text like the full web-link, or “DOI:” or other text.

Note that older references may not actually have a DOI. In these cases, use “NA”.

---

## AdditionalSource

Record additional source(s) used to reconstruct the matrix, or to obtain additional ancillary data. This should be formatted as Lastname Journal YearPublication additional information, e.g. Godinez-Alvarez Bot Rev 2003, or e.g. Schmidt J Ecol 2007 Appendix A1

---

## Database

Over the years COMPADRE database has obtained data from similar matrix databases. This column records whenever this is the case. The possible values are:

- Exeter - David Hodgson
- COMPADRE I - Miguel Franco & Jonathan Silvertown
- Ramula - Satu Ramula
- NCEAS - Jennifer Williams et al.

- UQ - Yvonne Buckley and collaborators
- 

## **IUCNStatus, InvasiveStatus, ConservationStatusUS**

We no longer digitise this information. Instead we point out to end users of the databases how to get this information from the relevant databases.

---

## **Ecoregion**

Description of the Ecoregion for the study, using the categories described in Figure 1 of Olson et al. (2001).

### **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**

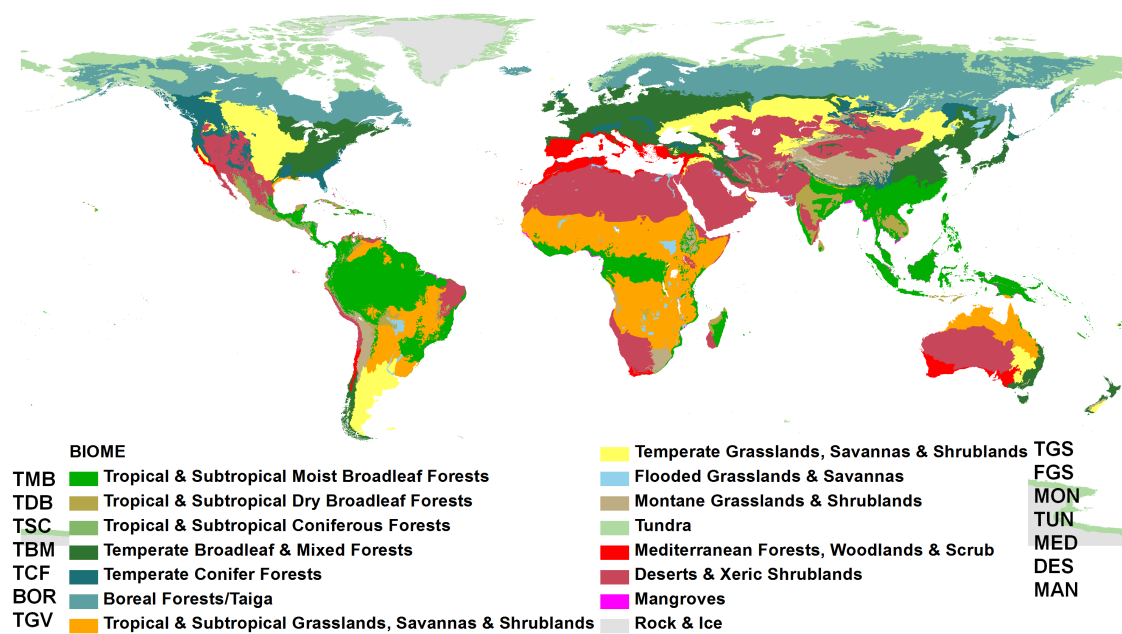
- POE polar ecosystems
- TSS temperate shelf and seas ecosystems
- TEU temperate upwellings
- TRU tropical upwellings
- TRC tropical coral

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

This is a drop-down variable. Separate with “;” if more than one

To find out within which terrestrial ecoregion your study has been made, search for the place on the following map:

---



From: Olson et al. (2001) Terrestrial ecoregions of the world: New map of life on earth. Bioscience 51:933-938

Figure 2: Ecoregions



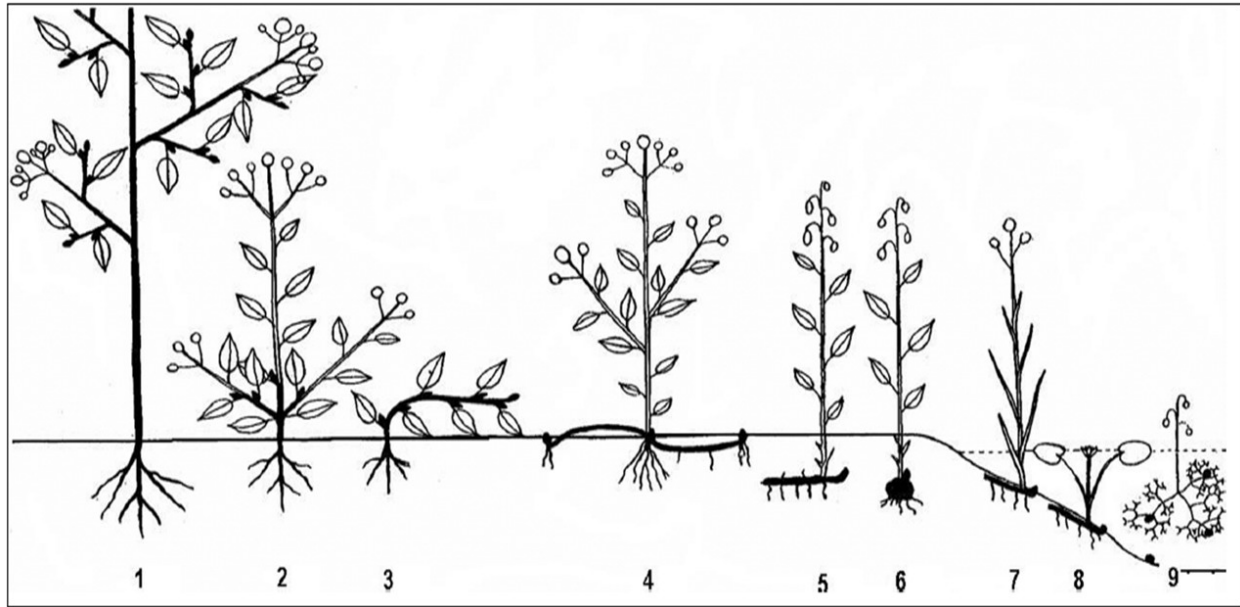


Figure 3: Raunkiaer

### GrowthFormRaunkiaer (P)

Plant classification according to Raunkiaer (classified after the position of the overwintering buds). Enter “NA” for **all** animals in COMADRE.

Possible values

- Megaphanerophyte Overwintering buds positioned at >30 m height (e.g. large trees)
- Mesophanerophyte Overwintering buds positioned at 8-30 m height (e.g. medium trees)
- Nanophanerophyte Overwintering buds positioned at 0.25-8 m max height (e.g. small trees and all shrubs)
- Chamaephyte Buds on persistent shoots near the ground
- Hemicryptophyte Buds at or near the soil surface (e.g. daisy, dandelion)
- Geophyte Buds resting under dry ground (e.g. crocus, tulip)
- Helophyte Buds resting in marshy ground (e.g. reedmace, marsh-marigold)
- Hydrophyte Resting by being submerged under water (e.g. water-lily, frogbit)
- Therophyte Mostly annual and biennial plants
- Epiphyte Plants living on to of other plants

For further information see [https://en.wikipedia.org/wiki/Raunkiaer\\_plant\\_life-form](https://en.wikipedia.org/wiki/Raunkiaer_plant_life-form).

1. Phanerophytes
2. Chamaephyte
3. Chamaephyte
4. Hemicryptophyte
5. Cryptophytes, more specifically: Geophyte
6. Cryptophytes, more specifically: Geophyte
7. Cryptophytes, more specifically: Helophyte
8. Cryptophytes, more specifically: Hydrophyte
9. Cryptophytes, more specifically: Hydrophyte
10. Therophyte, aerophyte and epiphyte not shown

The description offered here is at the subsection level (e.g. “Cryptophytes” are classified as “Hydrophyte”,

“Helophyte” or “Geophyte”; “Phanerophytes” are classified as “Nanophanerophytes”, “Mesophanerophytes” or “Megaphanerophytes”). When in doubt, information is to be retrieved from the internet (e.g. Wikipedia, Google photos) or the author is to be contacted or ask Ro. However, the variable “Plant type” typically relates to the variable “Growth form Raunkiaer” as it follows

- “Tree”, “Palm” and “Liana” are “Nanophanerophyte” (0.25-8 m height), “Mesophanerophyte” (8-30 m height) or “Megaphanerophyte” (>30 m height), depending on their maximum attainable heights as indicated in parentheses. \* If you don’t get this info from the paper, a quick Google photo search can give you an idea. “Shrub” and “Succulent” are either “Chamaephyte” or (most often) “Nanophanerophyte”, but in some species they be “Mesophanerophyte”.
- “Herbaceous perennials” may be “Hydrophyte”, “Helophyte”, “Epiphyte”, “Geophyte”, “Hemicryptophyte” or “Chamaephyte”.
  - Hydrophyte: herbaceous perennial that lives in or really close to a lake, river or sea
  - Helophyte: herbaceous perennial that lives in a marsh or wetland
  - Geophyte: herbaceous perennial that overwinters below ground. These species have a storage below ground of the following types: rhizome, tuber, bulb or similar
  - Hemicryptophyte: herbaceous perennial without storage organ below ground that will overwinter right at ground level
  - Chamaephyte: herbaceous perennial that will overwinter above ground but no higher than 25 cm.
- “Annual” (and biennials or pseudo-annuals) are “Therophyte”.
- “Algae” is “Hydrophyte”.
- “Bryophyte” can be “Epiphyte” or “Hemicryptophyte”.
- “Epiphyte” is “Epiphyte”. Note that the Raunkiaer growth form “Aerophyte” is not used in this database. All “Aerophyte” (e.g. genus *Tillandsia*) are classified as “Epiphyte”.

---

## Purposes

The purpose columns describe the purpose(s) of the study detailed in the manuscript. This information is usually obvious from the abstract and the last paragraph of the intro. Several purposes are possible per study and for each purpose a 0 or 1 must be entered, representing “was **not** a purpose” and “was a purpose” of the study respectively.

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

The names of the columns are: PurposeComparativeDemography, PurposeSpatialDemography, PurposeAbioticImpacts, PurposePVA, PurposeSpeciesDynamicsDescription, PurposeInterspecificInteractions, PurposeManagementEvaluation and PurposeMethodologicalAdvancement.

---

## Treatment

Records any treatments of any kind other than control/unmanipulated. Be concise with the treatment description. Multiple treatments should be separated with a semi-colon “;”. Enter “No” if the matrices are only obtained from unmanipulated conditions (for definition of “unmanipulated” see “MatrixTreatment” below).

---

## StudyStart

Records the first year of the study, e.g. 2008

---

## StudyEnd

Records the last year of the study, e.g. 2012

---

## StudyDuration

Records the number of years of observation, e.g. 5. This is calculated as: “StudyEnd” year - “StudyStart” year + 1. The calculation thus overlooks any missing years in the middle of the study period. Note that Excel should calculate the StudyDuration automatically since the formula is already set as default for that column. If it is not calculated by Excel please enter formula (not a calculated number) in case we get better information later.

---

## AnnualPeriodicity

The periodicity of the transition matrix presented in the paper. For example:

- 1 - A yearly time-step
- 2 - A 6-month time step where the model increments twice per year (e.g. in January and in June)
- 0.2 - A 5 year time step where the model increments once every 5 years (often seen in models for trees).
- ? - Other annual periodicities

Annual periodicities often, *but not always*, correspond to the frequency of fieldwork visits. If that is the case, the AnnualPeriodicity can be calculated as follows: If the study site has been visited more than once a year calculate  $1 * N$ , where N is the number of visits to the site per year, e.g.  $1 * 2 = 2$ . If the study site has been revisited less than once per year calculate  $1/Y$ , where Y is the number of years, e.g.  $1/5 = 0.2$ .

---

## NumberPopulations

Number of populations of the study species. This could be any integer value and refers to the number of populations as defined by the author. Within-site replication of permanent plots are not considered to be different populations.

---

## WithinSiteReplication

Reports the number of plots/transects sampled at each population and should give an accurate size description of the plots. For example, “6 (4x4 m<sup>2</sup>)” means that in each population, the authors established and followed six permanent plots of 4x4 m<sup>2</sup>. Be as specific but as brief as possible.

---

## MatrixCriteriaSize

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

- criteria - The matrix population model contains stages based on size and this briefly describes the aspect of size that was measured to classify individuals (e.g. DBH, stem length, stem height, stem number, etc).
  - No - The matrix population model is not based on size.
- 

## MatrixCriteriaOntogeny

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.
  - Based on size - Where the authors report ontogenetic stage class(es) but the stage class(es) themselves are based on size criteria (e.g. vegetative individuals defined as those <1 m height)
- 

## MatrixCriteriaAge

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
  - Based on size - If the stage classes are in reality based on size criteria (e.g. old individuals defined as those with >10 stems)
-

## STACKED MATRIX FILES

---

### EnteredBy

See **Student** above.

---

### EnteredDate

See **DateDigitalization** above.

---

### Source

Origin of data. If sources other than those below are used, a short but useful description should be given. The aim of this column 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)
  - *any database* - Matrices derived from one of the various databases that have been merged into COMPADRE/COMADRE see **Database** above.
  - Online appendix - Matrices derived from the online appendix of publication
  - Book - Matrices derived from a book
  - Other sources - Matrices derived from any other source.
- 

### SpeciesAuthor

See **SpeciesAuthor** above.

---

### StudiedSex

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)
-

## MatrixComposite

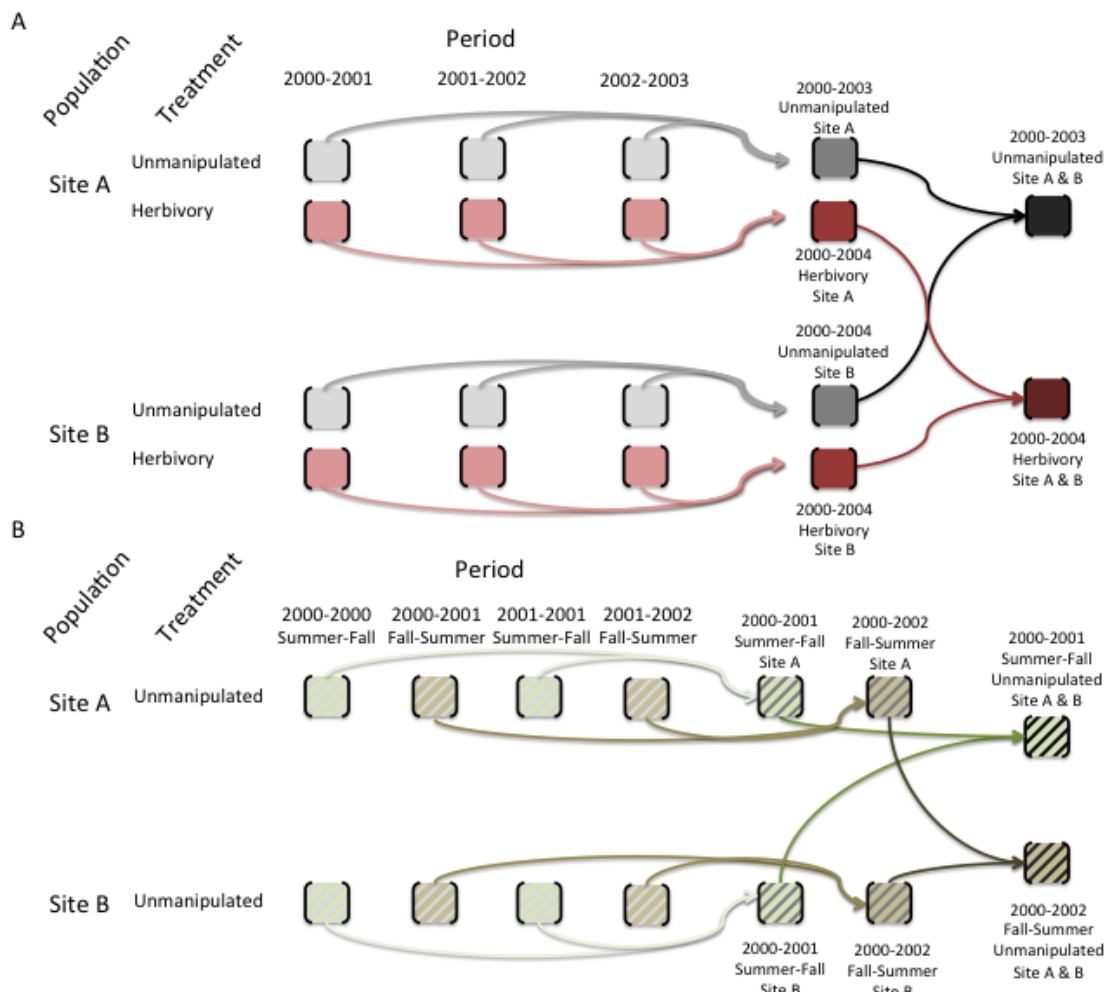
Indicates the type of matrix population model (See the figure below for more information).

- Individual - A matrix population model constructed for a single study  $\times$  species  $\times$  population  $\times$  treatment  $\times$  period combination
- Mean - An average (arithmetic mean) of other matrix population models (e.g. element-by-element arithmetic mean of a population's matrices across several time periods available)

**Note:** Generally, the mean values of unmanipulated and treatment matrices are to be kept independent of each other (i.e. never calculate mean matrices across both unmanipulated and treatment matrices), and a mean matrix for different treatments should not be calculated/digitised in the database.

- 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

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).

---

## MatrixTreatment

Describes if a treatment was applied by researchers or not, and the nature of the treatment.

- *treatment* - Provides a brief description of the treatment applied to the population described by the matrix population model. We define treatment as an action intentionally imposed by humans. If more than one applies to a matrix, multiple 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.

Note: Sometimes a treatment is not *applied* by researchers but is still 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”.

---

## MatrixCaptive

Indicates whether the study species was studied in the wild or under controlled conditions for most of its life cycle.

- **W** - Wild: the study was carried out in natural conditions.
- **C** - Captive: the study was carried out for most part of the life cycle of the species in a botanical garden, green house, laboratory, etc.
- **CW** - Captured from wild: the study was carried out on a population that was captured from wild.
- **C;W** - Mixed studies that incorporate vital rates from both captive and wild populations.

---

## YEAR/SEASON/MONTH

These 6 columns describe the temporal aspects of the study for each matrix.

### MatrixStartYear

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

---

### MatrixStartSeason

Numeric indication of the 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 (i.e. summer in the southern hemisphere corresponds to winter in the northern hemisphere)

1. Spring
  2. Summer
  3. Autumn or Fall
  4. Winter
- 

### MatrixStartMonth

Numeric indication of the first month of study. Month  $m$  in periodic matrix population model that describes population dynamics from month  $m$  to  $m + 1$ . If the 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).

1. January
  2. February
  3. March
  4. April
  5. May
  6. June
  7. July
  8. August
  9. September
  10. October
  11. November
  12. December
- 

### MatrixEndYear

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

---

### MatrixEndSeason

Numeric indication of the end season of the matrix.

1. Spring



2. Summer
  3. Autumn or Fall
  4. Winter
- 

## MatrixEndMonth

Numeric indication of the last month  $m$  of study. Month  $m + 1$  in periodic matrix population model that describes population dynamics from month  $m$  to  $m + 1$ . If the 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).

1. January
2. February
3. March
4. April
5. May
6. June
7. July
8. August
9. September
10. October
11. November
12. December

**Note:** All study time information always belong to the related matrix. So for a mean matrix the whole study duration must be recorded (e.g. *MatrixStartYear 1998 together with MatrixEndYear 2005*) while for a mean matrix the specific year or seasonal information must be given (e.g. *Matrix StartYear 2001 together with MatrixEndYear 2002*, or “*MatrixStartYear 2001, MatrixStartSeason 1*” together with “*MatrixEndYear 2002, MatrixEndSeason 2*”

This is how it possibly looks like (this is a 6x6 mean (the upper one) and individual (the lower one) matrix for the algae *Alaria\_nana*)

---

## MatrixPopulation

The name of population where the matrix was recorded, usually as given by the publication author. For example “Brazeau Creek, Florida”. When the author provides no name, the closest geographic location is assigned as name. If there are multiple populations in the study and their names are not pertinent/available, sequential names in alphabetical order are assigned for each population in the study, separated by “;” (e.g. “A”; “B”; “C”; etc).

---

## Geolocation columns

These 11 columns describes the location of each matrix.

**Always** check geolocation with Google Earth! If the geolocation from Google Earth does not fit the study site description ask author for further information and use one of the comments for the **Observation**.

I	J	K	L	M	N
MatrixStartYear	MatrixStartSeason	MatrixStartMonth	MatrixEndYear	MatrixEndSeason	MatrixEndMonth
1997	NA	5	2001	NA	9
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
1997	NA	5	1997	NA	9
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA

Figure 4: Alaria

### LatDeg / LatMin / LatSec

Record the degree (0-90), minute (0-60) and second (0-60) values of the locations of the matrix. If no LatSec is available type in a zero. This means: the value “0” may mean precise values are not available! If the geolocation is given as e.g. 58° 06’ N 17° 26’ E we do not enter “06” into the LatMin / LatSec column but only “6”. **Note:** if there are no seconds (or minutes) given, you should record as NA rather than 0.

### LatNS

Cardinal direction

- N - North
- S - South

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

### LonDeg / LonMin / LonSec

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

### LonWE

Cardinal direction

- W - West
- E - East

**Note:** Cardinal direction might be given with a “+” or “-“ whereat “+” 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>



Figure 5: Altitude

### Altitude

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

**Note:** Height 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!

### Country

The ISO 3 Country Code(s) of where the study took place. If the study involves multiple countries, these are separated by “;”. [https://en.wikipedia.org/wiki/ISO\\_3166-1\\_alpha-3](https://en.wikipedia.org/wiki/ISO_3166-1_alpha-3).

Where studies are conducted in seas or oceans this becomes a bit more challenging. We will provide country information where pertinent (i.e. in waters under the jurisdiction of a country) but where this is not possible, we will use a code to describe that sea or ocean: NODC For example, a fish species studied in French territorial waters in the Bay of Biscay would be recorded as “Country” = “FRA; NODC-22” (for France, and the NODC code for the Bay of Biscay). If the population was studied both in French and Spanish waters the code would be “FRA; SPA; NODC-22”. If it was only studied in international waters (normally defined as 12 nautical miles (for explanation see here) then you would record it as “NODC-22”.

### Continent

Continent where study took place

- Africa
- Asia
- Europe
- N America (includes Canada, USA and Mexico)
- S America (includes the rest of the countries in the Americas not outlined in “N America”)
- Antarctica

P	Q	R	S	T	U	V	W	X	Y	Z
Lat	Lat	Lat						Altit		
Deg	Min	Sec	LatN	LonE	LonW	LonE	LonW	ude	Country	Continent
NC	NC	NC	NC	NC	NC	NC	NC	NC	USA	N America
NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
48	24	0	N	124	44	0	W	1	USA	N America
NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Figure 6: Geolocation

- Oceania (various definitions for Oceania exist, but here we opted for this one: [https://en.wikipedia.org/wiki/List\\_of\\_Oceanian\\_countries\\_by\\_population](https://en.wikipedia.org/wiki/List_of_Oceanian_countries_by_population))

### 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 e.g. for mean matrices. In that case we enter NC for "not calculated" into the columns LatDeg, LatMin, LatSec, LatNS, LonDeg, LonMin, LonSec, LonWE and Altitude. In columns Country and Continent we separate with ";" if the study was located on more than one country / continent.

This is how it possibly looks like (this is a 6x6 mean (the upper one) and individual (the lower one) matrix for the algae *Alaria\_nana*)

### MatrixSplit

Indicates the type of matrix population model, i.e. can A it be divided into U, F and C sub-matrices? If a matrix is divisible, we divide it! Then always insert all sub-matrices – even if e.g. C only has zeros! 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 it could look (these are two divisible 4x4 matrices for the herb *Dactylorhiza lapponica*):

	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW	AX
MatrixCI assNum ber	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19	
	1	0	0	227	0	NA	0	0	0	0	NA	0	0	227	0	NA	0	0	0	
	2	0,0033	0,6155	0,639	0,861	NA	0,0033	0,6155	0,639	0,861	NA	0	0	0	0	NA	0	0	0	
	3	0	0,296	0,2025	0,0575	NA	0	0,296	0,2025	0,0575	NA	0	0	0	0	NA	0	0	0	
	4	0	0,0445	0,0965	0,078	NA	0	0,0445	0,0965	0,078	NA	0	0	0	0	NA	0	0	0	
	1	0	0	170	0	NA	0	0	0	0	NA	0	0	170	0	NA	0	0	0	
	2	0,0046	0,637	0,571	0,892	NA	0,0046	0,637	0,571	0,892	NA	0	0	0	0	NA	0	0	0	
	3	0	0,273	0,221	0,049	NA	0	0,273	0,221	0,049	NA	0	0	0	0	NA	0	0	0	
	4	0	0,042	0,115	0,059	NA	0	0,042	0,115	0,059	NA	0	0	0	0	NA	0	0	0	

Figure 7: Splitting

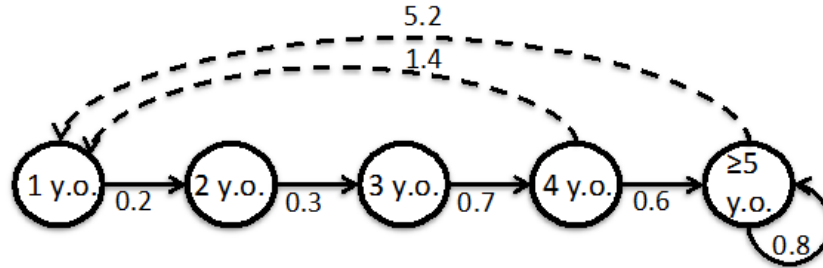
assNum	A1	A2	A3	A4	A5
1	0,3684	0,0667	0,2875	0,3099	0,0271
2	0,0526	0,5333	0,1386	0,1816	0,1034
3	0,5263	0,0667	0,6544	0,0581	0,0051
4	0,0526	0,2	0,6386	0,753	0,0034
5	0	0,1333	0	0,5714	0,7

Figure 8: Splitting

This is could look (this is a indivisible 5x5 matrices for the herb *Syngonanthus nitens*):

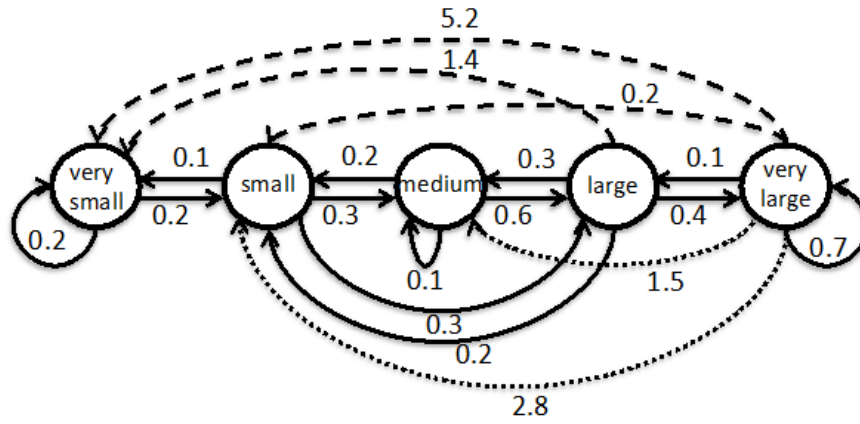
The diagram below shows the life cycle of two hypothetical plant populations based on age (A) and size (B), with their corresponding matrix population models. A, and underlying basic demographic processes of survival (U sub-matrix; solid arrows), sexual reproduction (F sub-matrix; dashed arrows) and 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 MatrixSplit in above).

A



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

B



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

\*\*\*\*\*

## MatrixFec

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 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 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 the **MatrixFec** these differences do not matter - if estimates fecundity, recruitment or sexual reproduction are collected, MatrixFec = “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
- 

## Observation

Any useful additional information about the population matrix model that is not provided elsewhere but may nonetheless be important (e.g. plant canopy, burning intervals, 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
- MatrixClassAuthor is not clarified

**Note:** *Information in the observation-column refer only to the matrix next to the column! If an information is important for all matrices of that species it must be copy & pasted for every matrix.*

---

## MatrixClassOrganized

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.

---

## MatrixClassAuthor

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<sup>2</sup>)

Some information on how to correctly write things in that column:

- First letter must be a capital letter (e.g. Adult: 100-150 cm DBH)
- If the Author uses abbreviations first write them down, followed by a “:” and the explanation of that abbreviations
- Use the most detailed description
- 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 you use “.” Instead of “,” when typing numbers (e.g. 1.45 cm, not 1,45 cm)
- Make sure that there is space between the number and the unit (e.g. 1.45 cm, not 1.45cm)
- We use exact the same description as the author did. If the paper is in another language (e.g. all the Spanish stuff from Iriondo Book 2009) we also write down the stages in that foreign language (e.g. in Spanish).
- 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)

---

## MatrixClassNumber

A numerical representation of classes in the population matrix model. Numbers starting with 1 (**not 0**). e.g. 1,2,3, ...  $n$ , where  $n$  is the dimension of the population matrix.

---

## matU/matF/matC/

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 constituent 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



	= Gray:	stasis
	= Green:	progression / growth
	= Orange:	retrogression / shrinkage
	= Yellow:	seed bank
	= Dark brown:	vegetative dormancy
	= Dark green:	awakening from vegetative dormancy
	= Dark orange:	fragmentation
	= Pink:	sexual reproduction
	= Blue:	asexual reproduction / clonal growth
	= Purple:	colonization

Figure 9: Unique

	= Asexual reproduction & Any Kind of Transition (U + C)
	= Sexual Reproduction & Any Kind of Transition (U + F)
	= Sexual & Asexual Reproduction & Any Kind of Transition (U + F + C)

Figure 10: Mixed

of the study). This should be noted in observations and recorded using **MatrixFec** (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 **MatrixFec**.

- 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 by variable **MatrixSplit**)

**Note:** *Leave one column filled with NAs between two separated matrices.*

## Color coding

Within the Excel sheets we color code all matrices in the same standardised way. The “color palette” can easily be found in (and copy pasted from) all stacked files (last tab).

**unique cells** (only one reproduction- or transition-form exists)

**mixed cells** (more than one reproduction- or transition-form exists)

**other colored cells**

NA	=	Not available	The information is not available
NDY	=	Not digitized yet	The information has not yet been digitized.
NC	=	not calculated	Information was not calculated because its calculation is not sensible
AFI	=	Asked for it	Author has been contacted (but no information in database yet)
End	=	End	All matrices for one species are entered, new one starts below

Figure 11: Other

		Seedbank						Sexual + Transition			
prop	Seed	0.400	0	0	0	0	0	1.050	12.639	0	<sexual
active	<8.cm.(Seedling)	0.134	0.463	0.013	0.010	0.003	0	0	9.123	0	<sexual & clonal
active	8-25.cm.(Small.juvenile)	0	0.338	0.877	0.051	0.031	0.008	0.008	2.500	0	<clonal
	25-45.cm.(Medium.juvenile)	0	0	0.056	0.748	0.025	0.015	0	0	0	
active	>.45.cm.(Large.juvenile)	0	0	0	0.130	0.802	0.086	0.019	0	0	<retrogression
active	>.45.cm.(Prereproductive)	0	0	0	0.005	0.111	0.823	0	0.001	0	<awakening
active	>.45.cm.(Small.adult)	0	0	0	0	0	0.069	0.912	0.017	0	
dorm	Vegetatively dormant	0	0	0	0	0	0	0.059	0.978	0	
								^	^	^	
								growth	dormancy	dormancy	

Figure 12: Example

All matrices are to be digitised using color-codes for the matrix elements  $a_{ij} > 0$ . If more than one process co-occur in the same cell, the author is to be contacted (or the database curator to be consulted). Used only the colors indicated in the tab “color palette” (and above). Leave one column filled with NAs between two separated matrices.

**Note:** Use only this colors for each of the different demographic processes. Our automatised matrix splitting procedure will not recognise the process otherwise and you will have to end up doing it by hand yourself otherwise.

**Example:** Matrix example for color-coding. Note that the coloring scheme is for internal usage only. The color names are the ones defined by Excel, and the user will not see them, as only text-based files will be made available to end users. We use these colors to separate divisible matrices automatically through a machine learning algorithm into U (survival), F (sexual reproduction) and C (clonal reproduction) matrices.