

CATS

a quick start guide

About this document

This document is meant to help you get your first simulation running, focussing on the practical and technical aspects of doing so. For more, detailed information, please refer to [link-to-paper](#), or the manual and other documentation available at <https://cats.univie.ac.at>.

About CATS

CATS¹ is a software tool to simulate the spatio-temporal dynamics of plant species under potentially changing environmental conditions, on either real or virtual landscapes. It is discrete in space (grid-based) and time (using an annual cycle). The spatial extent and resolution, as well the simulation length is constrained by the availability of computation resources².

Starting from an initial distribution of adult populations, the local population dynamics, driven by environmental conditions, are modelled on a cell-based raster. The population dynamics are controlled via vital rates (e.g. germination rate, adult survival rate, and so on), which can either be linked to the output of a species distribution model (SDM), or directly defined. The individual cells of the simulation interact via dispersal.

CATS is currently limited to single-species simulation. The spatial resolution of a CATS simulation is defined by the resolution of the input data. CATS is written in C, with a focus on performance and allows the modelling of large spatial extents (depending on available RAM up hundreds of million to billions of cells). It has also been used for simulations spanning thousands of years (Willner et al, in review).

This quick start guide showcases a simple, predefined application of CATS. In addition it is meant to illustrate how you can quickly get a CATS simulation running with your own data. CATS together with the exemplary data sets and the configuration file illustrated below in this document as well as a comprehensive manual are available at <https://cats.univie.ac.at/>.

CATS consists of two core elements. The first is the demographic module, which simulates how the environment drives the population dynamics. CATS has two different ways how the vital rates are parametrized. In this quick start guide, only the hybrid parametrisation mode, in which the vital rates depend on the occurrence probability derived from an SDM is used. Vital rates can also be specified directly, see section 6 of the manual. CATS offers a lot of flexibility to configure how the vital rates react to the environment, but will use defaults if not otherwise specified. The second core element is the dispersal module, which links the individual cells by modelling propagule (e.g. seed) transport.

¹ The name "CATS" is an acronym of "Cellular Automaton-like Tool to Simulate plant spread", which described the software when it was first written. While this is no technically longer quite accurate, the name has been retained.

² available RAM is the limit for the grid size, but an MPI version to use the shared resources of a computer cluster for a single simulation is also available

Running CATS

Technical information

CATS simulations are run on the command line. While it is generally tested on run on Linux, a pre-compiled version for Windows is available online. It should be possible to compile and run CATS under other operating systems, but this has not been tested.

Required Information

Each simulation requires a configuration file and input data provided as raster layers. The minimum required information to run a simulation in hybrid parametrisation mode is

- maximum values for vital rates: adult survival rate, seed survival rate, germination to adult survival rate, germination rate, flowering frequency, seed yield, and optionally clonal growth rate
- carrying capacity maximum
- age information for minimum and maximum age of maturity, and seed persistence
- one or more raster files with occurrence probability derived from an SDM. The values should be in range [0, 1].
 - **zero threshold, ZT**: the minimum occurrence probability needed for sink populations (under default settings, no seed production but germination). A good choice of ZT is the lowest projected occurrence probability value at which the species was present in the data used to fit the SDM.
 - **occurrence threshold, OT**: the minimum occurrence probability needed for stable populations. A good choice for the OT is the SDM's probability threshold for binary absence/presence prediction.
- a raster file of the initial adult population distribution (either as absence/presence information or population abundances)
- one or more dispersal kernel raster files and their minimum and maximum weights

A short description of each of these points can be found below. For more information see the CATS paper or the manual.

The configuration file contains information describing the simulation parameters, the demographic information of the species to be modelled, and the location of the input raster files.

For an example see [Configuration file example](#). Dependent on these specifications, a specific set of input data layers, comprising the environmental suitability, the initial distribution of adult populations of the study species and its dispersal kernels are required. All layers need to be in GeoTIFF raster format³ (only the first band is used), must have the same cell resolution, extent and origin⁴. The exception are the input rasters for the dispersal kernels, which only need to have the same resolution – see [Dispersal kernel\(s\)](#).

Simulation parametrisation

As an important first step the temporal extent of the simulation need to be set, by defining a starting year and the simulation duration in years. In contrast, the spatial extent is not directly specified, but is taken from the input data layers (more specifically from the initial population

³ Other raster file formats supported by the GDAL library (<https://gdal.org/>) might work, but are not tested.

⁴ This is not always possible for technical reasons, so CATS can be told to ignore projection information, in which case only the number of rows and columns of the input grid have to be the same. See configuration parameter `ignore projection information` in the manual.

layer, see below).

Species parametrisation

For the simulated species, a few mandatory characteristics have to be specified. CATS uses a list of vital rates (see below) to model the local demography of species. In the **hybrid parametrisation mode**, which is used for this quick-start guide, only the **vital rates' maximum values**, which are the values they can reach under optimal conditions, have to be specified. The actual value of these vital rates depends on the local environmental conditions, i.e. the suitability (occurrence probability derived from an SDM). The function link between the suitability and the vital rate values can be configured, but if not otherwise specified, defaults will be used⁵.

Individuals are structured into different life stages: seeds, juveniles and adults. Juveniles are defined as non-reproductive plants from the seedling stage up to plants in the year before their first reproduction. Adults are mature, reproductive individuals. Stages prior to reproductive ability are structured into annual age classes, introducing propagation delays for newly colonised habitats.

Note

For clonally reproducing species, an adult individual in CATS is defined to be a single ramet.

CATS uses the following vital rates (in bold) to describe the local demography of species, for which the potential maximum values have to be specified: **flowering frequency** (probability that an adult individual reproduces sexually within a year), **seed yield** (number of seeds produced per adult), **germination rate** (probability that a seed germinates and survives its first growing season), **adult survival rate** (probability to be alive in the following year), **germination-to-adult survival rate** (probability that a germinated seedling survives until maturation; i.e., its first sexual reproduction), **seed survival rate** (probability that a seed survives to the next year) and **carrying capacity** (maximum number of individuals per cell, see the manual). Moreover, the ability of a species for **clonal growth** can be considered as an annual multiplier for the number of additional adult individuals in a cell (i.e. the proportion of additional daughter rametes per mother ramet; Accordingly, a value of 0 reflects no clonal growth).

The specified maximum values for these vital rates should reflect the maximum performance of plants under favourable natural conditions (results from lab-/garden experiments need to be considered with caution).

Additionally, the time from germination to maturation needs to be specified (as **minimum and maximum age of maturity**, in number of years) and the **seed persistence** (i.e., the maximum number of years that seeds can survive in the soil) have to be specified.

For detailed information on these vital rates and their configuration possibilities see sections 4 and 6 of the manual.

⁵ the default settings depend on the vital rate and are either a sigmoid or constant link to suitability

Note

For annuals and biennials, set the minimum and maximum age of maturity to 0 years for annuals, and 1 for biennials, and the adult survival rate maximum and clonal growth rate maximum to 0. The flowering frequency should be set to 1. Alternatively, presets can be used – for more information see the manual.

Environmental data: suitability file(s) and thresholds

In the hybrid parametrisation mode⁶, CATS usually relies on the output of SDMs. The occurrence probability is used as a proxy for the species environmental suitability. A higher suitability is linked to a higher population growth rate. The suitability is a value in range [0, 1], which corresponds to the predicted occurrence probability. It is likely to be derived from a species distribution model (SDM) but can also be of other origin). Additionally two thresholds have to be specified: the **zero threshold (ZT)**, the lowest suitability for sink populations, where plants can survive and germinate, but not reproduce, and the **occurrence threshold (OT)**, which is the suitability at which the CATS model will be fitted to have a population growth rate of 1. See the paper and section 6.2 of the manual for more details. As a rule of thumb you can use the lowest projected occurrence probability where the species was observed in the data to fit the SDM as ZT, and the SDM's threshold probability for binary projections as OT.

These suitability values have to be specified as raster files. Temporal changes in environmental conditions are represented by providing suitability raster files for different time steps. For longer simulation periods (e.g. postglacial range dynamics up to present) it is neither necessary from a biological point of view nor computationally feasible to derive suitability rasters for each year. Therefore, suitability layers can be provided to CATS at specific intervals (specified in the configuration file). The annual suitability within an interval is either computed by linear interpolation between consecutive suitability rasters (e.g. between 2020 and 2030 for a 10-years interval) or remains constant for the whole interval (e.g. at the level of 2020 until 2029 and changes in 2030). However, it is also possible to run CATS assuming stable environmental conditions (i.e. using only one suitability raster).

Initial distribution of the study species For each CATS simulation, it is necessary to specify the spatial distribution pattern of the study species in the first year of the simulation. Thereby, for each cell the population size (i.e., either the number of adult individuals or absence/presence data) of the species need to be given as raster files. Be aware, that for CATS an adult individual refers to a ramet in case of clonally propagating species. These data can be obtained from direct measurements, derived from SDMs (e.g. by defining all cells above an suitability threshold as occupied), or of other origin.

Dispersal kernel(s)

In addition to the demographic model, CATS also uses a dispersal model, which is used to link the otherwise independent cells. The dispersal module is used to transport seeds produced in cells to areas in their vicinity. The dispersal model is parametrised by one or more dispersal kernels, i.e. raster grids that specify the probability of seeds transported relative to the cell of origin, and how these kernels are weighted. CATS make no assumptions about the underlying transport mechanism, but in the past has been used to simulate anemochorous, exo- as well as endozoochorous dispersal. Each vector is represented by an own data layer. Dispersal kernels reflect the dispersal ability of the study species and are used as moving window covering all cells. Its dimensions are, hence, independent to those of the simulation extent, but are generally

⁶ See [Options for more advanced applications](#) for a more direct method of the parametrisation of vital rates

much smaller, squared and have an odd number of rows/columns. Furthermore, the map origin is not considered. Each cell of the dispersal kernel represents the proportion of seeds dispersed via the respective vector that fall into the respective cell. Thereby, the central cell representing the source population (i.e. the population whose seeds should be spread). This means that the individual values in the dispersal kernel raster have to range $[0, 1]$, and the sum over all cells in the kernel has to be at most 1 (sums lower than 1 indicate that seeds were lost in the dispersal process, e.g. by digestion when simulating endozoochoric dispersal).

To allocate the overall seed yield of a population to specific vectors, multiple kernels of different sizes can be specified and a minimum and maximum weight can be assigned. For each dispersal event a random number between the minimum and maximum weight for each kernel is drawn from a uniform distribution. A weight of -1 indicates that all the remaining seeds (those not dispersed via previously defined kernels) should be allocated to that kernel. Thus, this value can be assigned only to the last kernel.

Model calibration

In hybrid parametrisation mode, the model will be calibrated, that population will stay stable (population growth rate = 1) for an isolated cell at a suitability equal to the occurrence threshold. This is achieved by shifting the vital rate values at this suitability, so that the dominant eigenvector of CATS Leslie-matrix formulation is equal to 1.

Output data

Simulation results are written to the output directory (cats-output, unless otherwise specified), with **adults** (containing raster files with the number of adults per cell – grids with juveniles and seeds can optionally also be stored), **stats** for general statistics of the simulation (see section 10.2 of the manual for more information) and **lambda** for population growth rates (optional, see section 10.3 of the manual). All output file names start with the name of the simulation run (configuration parameter **run name** in section **[general]**).

Configuration files

Configuration files are text files (INI file format) and can be generated and edited using any text editor. A configuration file consists of sections (e.g. species parametrisation and dispersal information), which logically group together configuration parameters.

The section identifier, which is enclosed in square brackets, consists of a section type (from a list of allowed type), which indicates what kind of values are contained in the section, and a user specified identifier, separated by a colon. This way multiple sections of the same type can be specified, which only differ in the identifier, e.g. when specifying multiple overlay layers (chapter 9 of the manual), each in their own section. The exception is the 'general' section, which does not need a user specified identifier, as it can only appear once per configuration file. The combination of section type and user-specified identifier is required to be unique in the file. Duplicate configuration parameter names are now allowed in the same section.

Each section contains one or more key-value pairs. The key is also called the configuration parameter name, the value the configuration parameter value. They are separated by an equal sign.

Some configuration parameter values consist of multiple values, which are separated by commas. White space around key, equal sign and value(s) are ignored. Comments are indicated by the hash sign (#). Any character after and including this comment character up to the end of the

line, is ignored.

Here an example:

```
# a free standing comment
[section type: identifier]
configuration parameter name 1 = value # another comment
configuration parameter name 2 = value1, value 2
```

Sometimes configuration parameter names reference other sections. Only the user-specified identifier has to be given, the section type is automatically inferred.

For the minimum example, four section are required. The contained configuration parameters are the minimum needed to run a CATS simulation in hybrid mode. The sections are:

[general] for information regarding the simulation itself

[environment:Astmons-rcp45] for the environment, i.e. the informations about the suitability rasters

[species:Astmons] for the species parametrisation of species Astmons (=Astragalus Monspessalanus)

[dispersal:Astmons-dispersal] for the dispersal information

Some configuration parameter values reference source data locations. These paths can either be specified using absolute paths or paths relative to the working directory. '%Y' (without the single quotes) can be used as replacement character for the current simulation year in file names. For additional formatting options when replacing '%Y' by the year, see section Filename patterns in the manual. CATS uses default options for unspecified parameters, e.g. the density dependence of each vital rate. The defaults are listed in the manual, but are also printed as CATS output.

Running a CATS simulation

CATS simulations are run on the command line. All file paths in the configuration file, unless they are absolute paths, are relative to the working directory, i.e. the directory in which CATS is started.

CATS on Windows

To run the above example, download the CATS quick start zip file (<https://cats.univie.ac.at/data/cats-quick-start-window.zip>) and extract it.

To start the simulation, open a terminal window and change to the directory where you extracted the ZIP file (it should contain the executable **cats.exe**).

Note

You can open the Windows command line shell by navigating to the directory which contains **cats.exe**. Click in the location bar, type **cmd** and then press Enter.

Then run the following command:

```
cats example\Astmons.conf
```


Note

For technical reasons, the performance of the pre-compiled Windows version is likely to be worse than a self-compiled version or the Linux version.

To see which other command line options are supported by CATS, start the program with the `--help` option:

```
cats --help
```

You can download the CATS executable and the example data separately from <https://cats.univie.ac.at>.

CATS on other operating systems

Download the source code from <https://cats.univie.ac.at> and follow the build instructions. Download the example data from <https://cats.univie.ac.at/cats/data/cats-quick-start-data.zip> and extract it to the directory with the built executable and run

```
./cats example/Astmons.conf
```

You'll find CATS output in the directory `cats-output`.

Options for more advanced applications

CATS has much more to offer than demonstrated in this quick start guide. For example, overlay layers, which provide information like habitat type or exclusion status, can be used to modify the local demography of species. Besides, additional dispersal vectors can be enabled.

Note

It is possible to repeatedly run a simulation using the same configuration file but generating different output files. This is feasible as CATS uses stochastic elements, generating variable results despite identical input data and specifications. Therefore, a number (the replicate number, 0, if not otherwise specified) is included at the end of output file names (prefixed with `_r`). To specify the replicate number, use the `-r` option when running CATS

In addition, the link between suitability and vital rates can be configured individually for each vital rate. This e.g. includes the functional link, suitability thresholds and whether a vital rate is affected by the local population density. However, in the direct parametrisation mode, CATS simulations also allows to parametrise the vital rates directly, based on empirical data; i.e., simulations are decoupled from SDM-derived suitabilities/occurrence probabilities at all. A comprehensive description of all options and details can be found in the manual.

Configuration file example

Here is an example configuration file (with some optional configuration parameters, which are labelled as such). The configuration parameters are explained in the comments. This file is part of the example data set <https://cats.univie.ac.at/data/cats-quick-start-data.zip> and can also be downloaded from <https://cats.univie.ac.at/data/Astmons.conf>.

```
# Example configuration file.
# Everything after '#' is a comment and ignored by CATS
# This file is a simplified version of the Astmons rcp45 simulation, on a smaller extent
# see <link to paper>
```

```

#
# Configuration options are grouped in [sections]
# For a detailed explanation see the quick start guide and the
# manual available at https://cats.univie.ac.at
# All configuration options not marked (optional) are mandatory!

[general]
# this parameter is used to identify this simulation and will be used for output file names
run name = Astmons_rcp45

starting year = 2020

# how many years the simulation should run for
time steps = 80

# how often (in years) should population grids be saved to disk
# note: statistics are saved every year, independent of the output interval
# This is independent of the interval input data is provided, see below
# (optional, default 1)
output interval = 10

# how many CPU cores should be used in parallel
# use -1 to use all available cores
# (optional, default 1)
threads = -1

[overlay: landuse]
# (optional) note: you can remove this whole section to see how the result would look like
# if we do not take habitats into account, i.e. if the plant could occur everywhere where it is
# climatically suitable. See the manual for an explanation of these parameters
#
# we use land use information to adapt constrain the
# the available carrying capacity for each cell
# note: for this simple example we assume that the land use does not change over time
# See section Habitat layer (carrying capacity) overlay in the manual for details.
type = habitat carrying capacity
filename pattern = example/landuse/landuse-rcp45-2020.tiff
table filename = example/landuse/Astmons_habitats.csv
default value = 0

# species parametrisation for species 'Astmons' = Astragalus monspessalanus
[species: Astmons]
# where to look for the suitability information for this species:
# in the section [environment: Astmons-rcp45]
suitability = Astmons-rcp45

# where to look for the the dispersal information for this species: [dispersal: Astmons-dispersal]
dispersal=Astmons-dispersal

# Initial population distribution information
# file name of the distribution pattern in the first year of the simulation, here as a relative path
initial population filename = example/Astmons-initial.tiff
# (optional) only set the initial population where the suitability is above OT
initial population suitability threshold = OT
# (optional) instead of using the presence/absence value specified in the input file, set the
# initial population size to the local carrying capacity
increase initial population to carrying capacity = yes

# Hybrid mode parameters: suitability thresholds
# CATS uses a range of [0, 1] for the cell's suitabilities.
# If your input data is in another range, you can use the option
# suitability divisor (see below) to transform the data read from the raster.
# occurrence threshold (OT), ideally the lowest projects occurrence probability
# of a presence used to parametrise the SDM. Used to calibrate the model.

```



```

# At OT, the demographic model will result in a stable population in isolation.
# A good choice for the OT is the SDM's probability threshold for binary absence/presence prediction
OT = 0.566
# Zero threshold, the minimum occurrence probability needed for sink population, where plants can
# germinate and survive, but not reproduce.
# a good choice is the lowest projected occurrence probability at which the species was present
# in the data used to fit the SDM
ZT = 0.081

# The maximum number of individuals per cell.
carrying capacity maximum = 1730

# To parametrise vital rates, the minimum required information is their maximum
# value, which is achieved under optimal environmental conditions.
# For further information see the manual.
adult survival rate maximum = 0.9
flowering frequency maximum = 0.4
germination rate maximum = 0.3
germination to adult survival rate maximum = 0.33
seed yield maximum = 100
seed survival rate maximum = 1.0
# (optional, default 0)
clonal growth rate maximum = 1

# additional age information, used for the stage structured model

# number of years a seed can survive in the ground, after which it is discarded.
# Minimum 1, which means it only gets one chance at germination, in the year after dispersal
seed persistence = 5

# the age of a plant (years after germination) reproduces sexually for the first time under highly
# favourable environmental conditions. If this value is zero, the plant can become adult in the year
# of germination.
minimum age of maturity = 2

# the age of a plant (years after germination) reproduces sexually for the first time under highly
# favourable environmental conditions. If this value is zero, the plant can become adult in the year
# of germination.
maximum age of maturity = 5

[environment: Astmons-rcp45]
# environment (suitability) information referenced above
# this environment is of type suitability (required for hybrid mode)
type = suitability

# (optional, default 1)
# how often (in years) does the suitability change (i.e. only for these years the suitability files
# will be loaded). For years in between the values will be interpolated linearly.
# If not specified, a suitability grid for each year has to be specified.
interpolation interval = 10

# where to find the suitability rasters.
# '%Y' is a placeholder that will be replaced with the current simulation
# year, i.e. 2020, 2030 to assemble the correct filename
filename pattern = example/suitability/Astmons_rcp45_%Y.tiff

# (optional, default 1.0)
# the values in the file are in range [0, 1000], while CATS expects
# suitability values in range [0, 1] so we divide the data read from
# the raster by 1000
suitability divisor = 1000

# Information describing the dispersal behaviour
[dispersal: Astmons-dispersal]

```

```
# Comma separated list of dispersal kernel raster files.  
# In this example the species is dispersed by endo-, exozoochoric, and wind dispersal  
  
# This should all be on one line.  
# The format of the dispersal kernel is described in the manual.  
kernel filenames =  
↪ example/Astmons_2dim_endo.asc,example/Astmons_2dim_exo.asc,example/Astmons_2dim_anemo.asc  
# Associated minimum and maximum weights (in order)  
# Here: 1% to 5% of all seeds will be dispersed using the endo- and exozoochoric kernels,  
# the rest (-1) using the wind kernel.  
minimum weights = 0.01,0.01,-1  
maximum weights = 0.05,0.05,-1
```