

CATS

Manual

Version 1.0.2

Andreas Gattringer

Contents

1	Glossary	5
2	Introduction	6
2.1	About this document	7
2.2	High-level overview	7
2.3	Quick start guide	7
2.4	Running a CATS simulation	7
3	The structure of a CATS simulation	8
3.1	Local populations	8
3.2	A plant's life cycle	8
3.3	A single year in CATS	9
3.4	Simulation duration and phases	10
3.4.1	Simulation phase	11
3.4.2	(Optional) Burn-in and warm-up phases	11
4	Demographic model	11
4.1	Vital rates	12
4.2	Processes and their associated vital rates	12
4.2.1	Carrying capacity application	12
4.2.2	Clonal growth (optional)	13
4.2.3	Germination	13
4.2.4	Maturation	14
4.2.5	Seed production	14
4.2.6	Dispersal	14
4.2.7	Year-to-year survival and age transition	15
4.3	Population density effects	16
4.4	Other vital parameters	17
4.4.1	Seed persistence	17
4.4.2	Minimum and maximum age of maturity	17
4.5	Presets	18
4.6	Poisson processes	18
4.6.1	Poisson processes options	18

5	Dispersal model	19
5.1	Kernel based dispersal	19
5.2	Local dispersal	21
5.3	Long-range dispersal	21
6	Vital rate parametrisation	21
6.1	Parametrisation modes	21
6.2	Hybrid parametrisation mode	21
6.2.1	Model calibration	22
6.2.2	Vital rate to suitability link functions	22
6.3	Direct parametrisation mode	24
6.4	Vital rate specification	26
6.4.1	Options available in all parametrisation modes	26
6.4.2	Options available only in hybrid parametrisation mode	27
6.4.3	Options available only in direct parametrisation mode	27
6.5	Vital rate defaults	28
6.5.1	Hybrid parametrisation mode defaults	28
6.5.2	Direct parametrisation mode defaults	28
7	Initial population distribution	28
8	Environments	28
8.1	Temporally changing environments	30
8.1.1	Temporally constant environments	30
8.1.2	Discretely changing environments	30
8.1.3	Linearly interpolated environments	31
8.2	Hybrid parametrisation mode environments	31
8.2.1	Suitability environments	31
8.3	Direct parametrisation mode environments	32
8.3.1	Pre-calculated vital rate environments	32
8.3.2	GLM environments	32
8.3.3	Constant value environments	33
9	Overlays	33
9.1	Exclusion overlays	33
9.2	Habitat layer (carrying capacity) overlay	34
9.2.1	CSV file format	34
10	CATS Output	34
10.1	Population grids	34
10.2	Grid statistics	34
10.3	Population growth rates grids	34
11	Input data	34
11.1	Filename patterns	35
11.2	Raster files	35
11.2.1	Dispersal kernels	35
11.3	Configuration files	35
11.3.1	Configuration file format	35
11.3.2	Configuration sections	36
11.3.3	Key-value pairs	36

11.3.4	Comments	37
11.3.5	Configuration parameter value data types	37
12	Configuration parameters	38
12.1	general (section)	38
12.1.1	output interval (required)	38
12.1.2	run name (required)	38
12.1.3	starting year (required)	38
12.1.4	time steps (required)	38
12.1.5	always include sign in year (optional)	39
12.1.6	burn-in phase year (optional)	39
12.1.7	burn-in phase length (conditional)	39
12.1.8	compression (optional)	39
12.1.9	debug (optional)	39
12.1.10	file format (optional)	39
12.1.11	ignore raster projection (optional)	39
12.1.12	output directory (optional)	39
12.1.13	poisson dampening factor (optional)	39
12.1.14	poisson dampening minimum (optional)	40
12.1.15	poisson maximum draw difference (optional)	40
12.1.16	poisson maximum (optional)	40
12.1.17	stats populated threshold (optional)	40
12.1.18	threads (optional)	40
12.1.19	warm-up phase year (optional)	40
12.1.20	write all (optional)	41
12.1.21	zero pad year digits (optional)	41
12.2	species (section)	41
12.2.1	dispersal (required)	41
12.2.2	initial population filename (required)	41
12.2.3	OT (required, hybrid mode only)	41
12.2.4	maximum age of maturity (required)	41
12.2.5	minimum age of maturity (required)	42
12.2.6	seed persistence (required)	42
12.2.7	suitability (required, hybrid mode only)	42
12.2.8	<vital rate> environment (required, direct mode only)	42
12.2.9	<vital rate> maximum (required)	42
12.2.10	ZT (required, hybrid mode only)	43
12.2.11	demographic module (optional)	43
12.2.12	demographic slope (optional, hybrid mode only)	43
12.2.13	hapaxanthly (optional)	43
12.2.14	increase initial population to carrying capacity (optional)	43
12.2.15	initial population suitability threshold (optional, hybrid mode only)	43
12.2.16	maximum adult fraction of carrying capacity (optional)	43
12.2.17	preset (optional)	43
12.2.18	save lambda grid (optional)	43
12.2.19	save lambda grid density (optional)	44
12.2.20	<vital rate> density (optional)	44
12.2.21	<vital rate> density threshold (optional)	44

12.2.22	<vital rate> environment multiplier (optional, direct mode only)	44
12.2.23	<vital rate> function (optional, hybrid mode only)	44
12.2.24	<vital rate> minimum (optional)	44
12.2.25	<vital rate> minimum scale factor (optional, hybrid mode only)	45
12.2.26	<vital rate> scale factor (optional, hybrid mode only)	45
12.2.27	<vital rate> suitability cutoff (optional, hybrid mode only)	45
12.2.28	<vital rate> value at OT (optional, hybrid mode only)	45
12.3	overlay (section, multiple allowed)	45
12.3.1	filename pattern (required)	46
12.3.2	type (required)	46
12.3.3	default value (optional)	46
12.3.4	enabled (optional)	46
12.3.5	reload interval (optional)	46
12.3.6	table filename (optional)	46
12.4	environment (section, multiple allowed)	46
12.4.1	type (required)	47
12.4.2	filename pattern (required)	47
12.4.3	interpolation interval (optional)	47
12.4.4	replacement interval (optional)	47
12.4.5	save environment (optional)	48
12.4.6	suitability divisor (optional, hybrid mode only)	48
12.4.7	value (optional)	48
12.5	dispersal (section)	48
12.5.1	kernel filenames (required, list)	48
12.5.2	minimum weights (required, list)	48
12.5.3	maximum weights (required, list)	48
12.5.4	local dispersal proportion (optional)	48
12.5.5	long range radius (optional)	49
12.5.6	long range target count (optional)	49
12.5.7	long range event probability (optional)	49
12.6	module (section)	49
12.6.1	filename (required)	49
12.7	glm (section)	49
12.7.1	family (required)	49
12.7.2	intercept (required)	49
12.7.3	predictors (required, list)	49
12.7.4	<predictor> linear (required)	50
12.7.5	type (required)	50
12.7.6	<predictor> quadratic (optional)	50
12.7.7	save environment (optional)	50
13	Command line arguments	50
13.1	Configuration option precedence	50
13.2	List of command line options	50
13.2.1	General command line options	51
13.2.2	Hybrid mode only options	51
13.2.3	Random number options	51
13.2.4	Debugging options	52

14 Running a CATS simulation	52
14.1 MPI simulations	52
15 Configuration examples	53
16 Modules	53

1 Glossary

raster layer spatially explicit numerical data. For input data, **CATS** only supports raster data with square cells in GeoTIFF format. See [Input data](#) for more information about input data.

vital rates Local population dynamics (within a cell) are controlled by species-specific vital rates representing biological processes. Examples are flowering frequency, seed yield and carrying capacity. For a full list, see [Vital rates](#). Each **vital rate** is determined by an environment, which specifies local conditions that can vary from cell to cell.

simulation grid the spatial simulation extent. It is a rectangular grid comprised of cells each potentially representing a population of the study species. Local population dynamics are modelled separately in each cell, but populations interact via dispersal.

parametrisation mode The parametrisation modes specifies, how the vital rates are specified and linked to the **environment**. CATS supports two parametrisation modes:

hybrid parametrisation mode In this mode, **vital rates** depend on the species' occurrence probability (suitability) and a specified value for each rate representing the maximum achievable under the most beneficial in situ conditions. The functional link (sigmoidal, linear or constant) between suitability and vital rate value can be selected for each vital rate, otherwise defaults are used. The same suitability value is used for all vital rates. See [Hybrid parametrisation mode](#).

direct parametrisation mode In this mode, the values of each vital rate are directly specified by a provided raster layer, capped at the specified vital rate maximum. CATS can also apply simple regression models (GLMs), in which one or more environmental raster layers are combined to calculate a vital rate's values. See [Direct parametrisation mode](#).

environment one or more **raster layers** associated with a vital rate. Depending on the parametrisation mode an environment can either be

- a raster layer, which contains occurrence probabilities (i.e., suitability ranging from 0-1) derived from a species distribution model (SDM). A link function (see parametrisation mode) is applied to generate the

value of the vital rate (hybrid parametrisation mode).

- a raster layer of pre-calculated values for a vital rate. (direct parametrisation mode)
- a set of raster layers, like temperature or precipitation, which are combined using a generalized linear model (GLM) to obtain the values of the vital rate (direct parametrisation mode). See [Environments](#).

simulation year/year CATS simulations proceed in annual time steps. In each year, the biological processes (clonal growth, germination, maturation, seed production, and dispersal) defining the demographic model are modelled for each cell based on their associated local vital rates values.

life stages and age classes CATS distinguishes three **life stages** for plants: seeds, juveniles and adults. Juveniles are defined as non-reproductive plants independent of size (from the seedling stage up to plants in the year before their first flowering). Adults are mature, reproductive individuals. Seeds and juveniles are structured in annual **age classes**. At the end of each year, (surviving) seeds and juveniles advance to the next higher age class (or are discarded if they reach their maximum age defined by the parameters [seed persistence \(required\)](#) and [maximum age of maturity \(required\)](#)).

configuration parameter a specification in the configuration file consisting of a **configuration parameter name** (i.e., the identifier) and the associated **configuration parameter value**. An example:
`maximum germination rate = 0.3`

overlay layers additional, optional, spatially resolved information which modify the simulation locally based on their value, can be specified, e.g. by excluding non-habitable locations like water, or modifying the carrying capacity based on habitat types. In CATS, these are called overlay layers, see [Overlays](#).

2 Introduction

CATS is a program to model changes in species distribution, spatially and temporally explicit under changing environmental conditions. In the current version of **CATS** only plants can be modelled. **CATS** is raster-based, i.e. **CATS** simulations run on a grid. The smallest spatial unit is a square cell. Each time step corresponds to one year. By changing the environmental conditions over time, the spread and/or decline of a species can be modelled, e.g. to study the effect of climatic changes, not only on a species potential distribution, e.g. like species distribution models (SDM), but also whether the species can keep up with this change via migration.

2.1 About this document

This document should give an overview of all possible configuration options and parameters in **CATS**. It is a work in progress. While every option is at least mentioned and shortly described, additional comments and design decisions will be added as needed.

2.2 High-level overview

A **CATS** simulation starts from an initial population distribution, i.e. spatially resolved abundances or presence/absence data. The simulated area is divided into a grid with square cells. The population dynamics within the cell are driven by vital rates, like germination rate and survival rate. **CATS** supports two different methods to determine spatial and temporal variation in these vital rates, the so called parametrisation modes, in which the vital rates can be either linked to a single suitability raster (hybrid parametrisation mode), or specified directly by providing input rasters for all vital rates (direct parametrisation mode). See [Vital rate parametrisation](#) for more information.

The input raster(s) for vital rates are called environments and have to be provided at least for some years. **CATS** can either linearly interpolate between years for which the input rasters are provided, or periodically change input rasters. It is also possible to leave the input rasters constant over time. See [Environments](#).

A simulation spans many years (time steps), in which population dynamics, based on the local population abundance and vital rates are simulated. The demographic model is run for each year and cell individually. At the end of each year, interaction between cells occurs via dispersal of seeds. Certain grid statistics are output at end of each year.

2.3 Quick start guide

A quick-start guide for CATS can be found <https://cats.univie.ac.at/>.

2.4 Running a CATS simulation

CATS is a command line program. The main input of a CATS simulation is the configuration file. It is a text file containing the parameters of the simulation, including species parametrisation. It also specifies the name and location of all other input files:

- environment rasters
- (optional) overlay rasters and data
- initial adult population distribution raster
- dispersal kernels

For more information about the configuration file format, see [Configuration](#)

[files](#). To specify a command line argument for CATS, call the executable with the path of the configuration file as the last argument. Additionally you can also specify other information directly on the command line. See [Command line arguments](#). For a detailed description, see [Running a CATS simulation](#).

3 The structure of a CATS simulation

CATS is spatially explicit, i.e. a simulation runs on a grid, consisting of square cells of the same size. It is also temporally explicit. A simulation spans one or multiple years. Each year is sub-divided into a series of biological processes, like germination and seed production, and book-keeping actions like loading data and calculating statistics. The list of processes and actions remains the same over the whole simulation duration. Each process and action is run in the same order for the whole grid. The simulation advances to the next process or action, only after these have run for each cell of the grid. For a list of processes and actions see [A single year in CATS](#).

3.1 Local populations

In CATS, the population of each cell (simulation site) is modelled individually. Cells interact only via dispersal, i.e. the process of transporting seeds from their cell of origin to other cells. Each seed is only dispersed (transported) once. Seeds that are not dispersed beyond their cell of origin are added to the cell's seed bank. The concept of local population in CATS describes the abundance (count) of adult and juvenile plants at a given size (cell). Juveniles are age-structured, that is the abundance of juveniles of different ages (in years) are stored separately, while adults are aggregated into a single class which is not resolved further.

Note

In CATS an 'individual' is defined as a sexually reproductive unit, i.e. in case of clonal plants as a ramet, independent of whether it is physically separated from other ramets or not (such as, e.g., the shoots of a tussock grass).

3.2 A plant's life cycle

Adult in the context of CATS means a individual that is able to reproduce sexually, i.e. produce (viable) seeds. CATS represents three life stages of plants: seeds, juveniles and adults. The first two life stages are age-structured, i.e. structured into one class per year of age. Adults are aggregated into a single class; only their abundance is stored. The first process simulated within each year is clonal growth of adults, if exhibited by the simulated species followed by shiftings seeds and juveniles to the next life stage; i.e. from seed to juvenile or from juvenile to adult. The next process simulated is seed germination. Seeds that do not germinate can survive a number of years in the ground (configuration parameter seed persistence) and have a chance to germinate at the start of a later year, if they survive the transition to the next year. Juveniles require a number of years (configuration parameters minimum age of maturity) before

they can become adult. If this value is zero, a plant can become adult in the same year it germinated (e.g. annual plants). Any juveniles that have not matured before reaching a maximum age (configuration parameter maximum age of maturity), will be discarded. After maturation, , i.e. transition from the juvenile to the adult stage, flowering and seed production can occur. The seeds produced by all individuals of a cell will be dispersed, both in their own cell, and cells further away, depending on dispersal kernel and long range dispersal options.

Each annual simulation step ends with simulating the survival of seeds, juveniles and adults from the previous year. Any surviving seeds and juveniles are then advanced to the next age class, or discarded, if they reach their maximum age.

3.3 A single year in CATS

CATS years are subdivided into actions and processes, which occur every year in the same order and are listed below.

- Statistics reset
At the start of the year, all statistics are reset.
- Environment update
Environments are updated. Suitability, direct rate and predictor raster layers are loaded from disk, if required. For a detailed description see [Environments](#).
- Overlay update
Overlay layers are updated. For a detailed description of overlays, see [Overlays](#).
- Carrying capacity application
After updating environments and overlays, the (potentially affected) carrying capacity is updated. The adult and juvenile populations of each cell will be capped, if they exceed their carrying capacity limits, see [Carrying capacity application](#).
- Clonal growth (optional)
For species which produce clonally, i.e. their clonal growth rate maximum is greater than 0, clonal growth is calculated, see [Clonal growth \(optional\)](#).
- Carrying capacity application
The carrying capacity is applied again. While every biological process in CATS should ensure that the carrying capacity is not exceeded, the carrying capacity will be applied after each task that can either modify the carrying capacity, or increase the population abundance.
- Germination
Seeds that were dispersed in previous years have a chance to germinate, i.e. to advance to the juvenile stage. For details see [Germination](#).

- **Maturation**
Juveniles have a chance to become adult, if the number of years since germination has reached or exceeded the minimum age of maturity. If the age of maturity is set to zero, a plant can germinate and mature in the same year. For details see [Maturation](#).
- **Carrying capacity application**
- **Seed production**
Mature plants (adults) have a chance to produce seeds, depending, amongst others, on flowering frequency and seed yield. For details see [Seed production](#).
- **Dispersal**
Seeds that were produced in the current year are dispersed using dispersal kernels, i.e. rasters that specify the probability to travel to a cell in relation to the source cell. Optionally, a fraction of seeds can be directly deposited into the source cell or transported via long-range events. This happens before the dispersal using kernels, which then transports the remaining seeds. See [Dispersal model](#).
- **Grid output**
Depending on the configuration parameter output interval (required), the adult population abundance grids (and optionally others) are saved to disk. See [Population grids](#).
- **Statistics output**
Grid statistics, i.e. the number of populated cells are calculated and saved to a CSV file. See [Grid statistics](#).
- **Year-to-year survival and age transition**
In the last process of the year, survival rates are applied. Any seeds and juveniles that survive the year-to-year transition and have not yet reached an age equal to seed persistence or maximum age of maturity are advanced to their respective new age class, any others are discarded. See [Year-to-year survival and age transition](#).

3.4 Simulation duration and phases

At the start of a CATS simulation, the adult initial population distributed is loaded from the specified initial population raster. Depending on the configuration, the initial population abundance is either read directly from the input raster (but capped at the local carrying capacity), or set to the local carrying capacity (e.g. when only presence/absence data is available). See [Initial population distribution](#).

3.4.1 Simulation phase

The simulation phase is the main phase of a CATS simulation. Starting from an adult initial population, the demographic model and dispersal will be run for the specified number of years. The simulation phase is always enabled.

3.4.2 (Optional) Burn-in and warm-up phases

The initial population only consists of adults – no seeds or juvenile plants are present – and is therefore not in an equilibrium state. To rectify this, CATS supports optional burn-in and warm-up phases which are run before the simulation phase. Instead of directly starting the simulation, the burn-in phase starts with the climatic situation at a pre-defined year and will run for a specified number of years with static environmental conditions. After the specified number of years have been reached, the warm-up phase starts. In the warm-up phase, the environmental conditions are shifted over multiple years (from the burn-in year to the simulation start year) to the environmental conditions of the actual simulation, which then starts with an initial population of seeds, juveniles and adults based on the previous phases. To simulate this shift, the environmental conditions are linearly interpolated over that time span. During the burn-in and warm-up phases, no population abundance grids are output. To enable the burn-in and warm-up phase, the following configuration parameters have to be set: [burn-in phase year \(optional\)](#), [burn-in phase length \(conditional\)](#), [warm-up phase year \(optional\)](#).

Example:

The simulation of a species' spatial spread over a 100 year period starting in the year 2020 requires an initial population distribution. This data corresponds to absence/presence data recorded between 2000 and 2020. The burn-in year is set to 2010, which is repeated for 100 years. After the burn-in phase, the environmental conditions are linearly interpolated between 2010 and 2020 over 10 (2020 – 2010) years and continues with the actual simulation.

Note

During the warm-up phase the species can colonize new areas, so at the start of the simulation phase the new spatial distribution can be potentially much larger. This can be rectified by restricting the population to their initial distribution via the suitability raster.

4 Demographic model

Vital rates define the local population dynamics. The term vital rate is also used for the carrying capacity, which follows a similar mechanic. The local population dynamics model consists of a series of processes, that are applied to the local population for each year in the same order. Each process (with the exception of dispersal, see below) is quantified by one or more vital rates.

4.1 Vital rates

Vital rates are the quantifiable representation of the driving processes of the population dynamics model. Each vital rate can be either constant or dependent on an environment. CATS makes no assumptions about how the local environment affects vital rates. Vital rates are primarily parametrised by their maximum value, i.e. the rate they can achieve under optimal conditions, their actual value (either specified spatially and temporally explicit, or as function on the suitability) and their dependence on the adult population density. For parametrisation, see [Vital rate parametrisation](#). For each vital rate, a maximum value has to be specified, which serves as an upper limit for that rate. This maximum value should represent the value that a rate can reach under optimum conditions. Vital rates are always positively formulated, i.e. a larger value corresponds with higher abundance or performance.

4.2 Processes and their associated vital rates

Processes are listed in the order they appear in the annual cycle. They are quantified by vital rates. In the following text, 'local' vital rate values are the calculated or specified vital rate values for environmental conditions at that specific site in that specific year, modified by population density effects, if any.

4.2.1 Carrying capacity application

In the CATS model, each cell (geographic location) has an upper limit of individuals (adults and juveniles) it can support. CATS makes no assumptions about the underlying limitation. The carrying capacity is split between adults and juveniles. The associated vital rate is carrying capacity.

Carrying capacity

The carrying capacity specifies the maximum number of individuals co-occurring in a single cell. It is split between adults and juveniles. The maximum number of adults that can fit into a cell is determined by the local carrying capacity value (the vital rate value depending on the local conditions) times the adult carrying capacity fraction (default 0.5, i.e. 50%, see configuration parameter [maximum adult fraction of carrying capacity \(optional\)](#) and [Presets](#)). Adults take precedence over juveniles. Juveniles can use any remaining carrying capacity. Adults count fully towards the carrying capacity, while juveniles are weighted and count, depending on age, less than an adult. The weight for an juvenile in the n th year after germination (where $n = 0$ is the year of germination) is calculated as

$$w_n = \left(\frac{n+1}{m+1} \right)^2 \quad (1)$$

where w_n is the weight, and m the maximum age of maturity. Carrying capacity constraints are applied whenever the carrying capacity itself is changed (e.g. by changes to the environment on which the carrying capacity depends, or changes to overlay layers modifying the carrying capacity), or the number of individuals

(either adults or juveniles) in a cell increases. This guarantees that after each process these constraints are not violated.

Once this limit has been reached, no additional adults can be added. Any excess adults will be discarded. The specified carrying capacity maximum value must be a number ≥ 1 . The carrying capacity is a special case, as it cannot depend on the population density, which is defined as the fraction of the used carrying capacity.

Note

Adult individuals in CATS are reproductive units, i.e. ramets in case of clonal growth.

Note

The carrying capacity is a special case when it comes to density effects: it can never depend on population density.

4.2.2 Clonal growth (optional)

The clonal growth process represents asexual reproduction, which increases the number of adults (ramets). The clonal growth rate specifies how many additional ramets are formed from an already existing ramet, by multiplying the latter with $1 +$ the local clonal growth rate (a number ≥ 0). This process is only enabled if a clonal growth rate maximum > 0 is specified in the species parametrisation. Each clone is immediately added to the local adult population. The associated vital rate is the clonal growth rate.

Note

A possible delay in the time between asexual production of ramets and their maturity (their ability to reproduce sexually) is currently not implemented.

Clonal growth rate

The local clonal growth value is capped at zero as lower bound, e.g. under non-ideal conditions at worst no clonal growth occurs. If not specified, clonal growth is disabled, which is equivalent to specifying a maximum clonal growth rate of 0. Clonal growth is a Poisson process, see [Poisson processes](#).

Note

The clonal growth rate can never fall below 0, no matter the parametrisation, i.e. the number of adults can never be reduced by clonal growth.

4.2.3 Germination

Germination is the process that transforms seeds into juvenile individuals. The local germination rate is multiplied with the number of seeds of each age-class of the seed bank. The number of germinated seeds is subtracted from the corresponding age-class of the seed bank and added to the juvenile population

abundance of age 0 (germinated in the current year). The associated vital rate is the germination rate.

Germination rate

The germination rate specifies the probability that a seed germinates (advances to the juvenile stage). Currently CATS does not allow the germination rate to depend on the age of the seeds. See assumptions and limitations. The specified germination rate maximum value has to be a number in range (0, 1].

Germination is a Poisson process, see [Poisson processes](#).

4.2.4 Maturation

Maturation is the process that transforms juveniles into adult individuals, i.e. the transformation into an individual that is able to reproduce. Juveniles are structured into age classes. Only juveniles that have reached the specified minimum age of maturity (see [Minimum and maximum age of maturity](#)) are considered for maturation. The maturation rate is calculated from the associated vital rate [Germination to adult survival rate](#). The maturation rate is age-specific, i.e. it scales linearly with the juvenile age class. For age classes below the minimum age of maturity, the maturation rate is zero, and reaches the juvenile transition rate (see [Germination to adult survival rate](#)) at the maximum age of maturity. For each juvenile age class, the abundance is multiplied with the age-specific maturation rate. The result is the number of new adults, which are subtracted from the corresponding juvenile class. Maturation is a Poisson process, see [Poisson processes](#).

4.2.5 Seed production

Seed production is a two-step process, consisting of flowering and actual seed production. The number of adults in a cell is multiplied with the local flowering frequency rate, the result is then multiplied with the local seed yield rate to calculate the number of seeds that are produced (and available for dispersal) locally. The associated vital rates are flowering frequency and seed yield. Seed production is a Poisson process, see [Poisson processes](#).

Flowering frequency

The flowering frequency is one of the two vital rates associated with reproduction and specifies the number of adults that will produce seeds. The specified flowering frequency maximum value has to be a number in range (0, 1].

Seed yield

The seed yield specifies the number of seeds that an adult will produce (if it flowers). The specified seed yield maximum value has to be a number greater or equal to one.

4.2.6 Dispersal

Dispersal is described in [Dispersal model](#) and is not dependent on the local environment and therefore has no associated vital rate.

4.2.7 Year-to-year survival and age transition

At the end of each simulation year, survival (mortality) and age transitions for all age classes (of both seeds and juveniles) are applied.

For seeds, the survival is defined by the seed survival rate. For each seed age class, the local mortality is given by one minus the local seed survival rate. The result is then subtracted from the corresponding seed age class. All surviving seeds are then advanced to the next age class or discarded, if their age exceeds the seed persistence. The associated vital rate for seeds is the [Seed survival rate](#) and the vital parameter [Seed persistence](#).

Juveniles are treated similar to seeds. The number of discarded juveniles for each age class is the number of juveniles of that class multiplied by the local mortality rate of one minus the local juvenile transition rate, which is calculated from the germination-to-adult survival rate. The juvenile transition rate does not depend on the age of the juvenile individuals. Any surviving juveniles are advanced to the next age class, or discarded if their age exceeds the maximum age of maturity.

The associated vital rate is [Germination to adult survival rate](#) and associated vital parameters are [Minimum and maximum age of maturity](#).

For adults, the adult population abundance is multiplied with the local adult survival rate to calculate the number of individuals that survive from the current simulation year to the next. The associated vital rate is the [Adult survival rate](#). Survival processes are Poisson processes, see [Poisson processes](#).

Adult survival rate

The adult survival rate specifies the probability that an adult survives from the end of one vegetation period to the start of the next. The specified adult survival rate maximum value has to be a number in range [0, 1].

Note

For annual and biannual plants the [adult survival rate maximum](#) should be set 0.

Seed survival rate

The seed survival rate specifies the probability that a seed survives the transition between years and advances to the next age class. Seeds need to survive at least once to be able to germinate. Any seeds that have not germinated within a number of years equal to the value of the seed persistence will be discarded. The specified seed survival rate maximum value has to be a number in range (0, 1].

Note

The current version of CATS operates under the assumption that the seed survival rate is independent of seed age.

Germination to adult survival rate

The germination to adult survival rate specifies the probability that a germinated seedling reaches adulthood, i.e. the reproductive state. It is not used directly, but from this rate, the juvenile transition rate and the maturation rate are derived.

The juvenile transition rate specifies the probability that a juvenile survives the year-to-year transition and gets advanced to the next age class. The current version of CATS operates under the following assumptions, which are used to calculate the juvenile transition rate from the germination to adult survival rate.

- juvenile survival between vegetation periods is independent of juvenile age
- the maturation rate scales linearly between the minimum age and maximum age of maturity.
- the maturation rate at the maximum age of maturity is equal to the juvenile transition rate (survival rate)
- the maturation rate is zero before the minimum age maturity

The specified germination-to adult survival rate maximum value has to be a number in range (0, 1]. See also [Poisson processes](#).

4.3 Population density effects

CATS takes population density effects into account by modifying affected vital rates. Currently only negative density effects, i.e. larger abundances lead to lower vital rates, are considered. Each vital rate can be made density dependent, see the individual vital rates for which this is enabled as default. For the population density effects only the adult population abundance is considered, and compared to the (adult) carrying capacity. Under default settings in hybrid parameterisation mode, density effects

- are only enabled if the population abundance is greater than the carrying capacity at the occurrence threshold, i.e. there are not density effects for suitability values below OT.
- will only reduce the vital rate value to a minimum of the vital rate value at OT.

For hybrid mode parametrisation, the density dependence is a multiplier to the vital rate value is

- for linear vital rate link function

$$r(N) = r(0) \times \frac{K - N}{K} \text{ for } \frac{N}{K} > t, r(0) \text{ otherwise} \quad (2)$$

- for the constant vital rate link function the population density is ignored

$$r(N) = r(0) \quad (3)$$

- for the sigmoid vital rate link function

$$r(N) = \min \left(r(0), r(0) \times \frac{K - N}{K} + \frac{N}{K} \times r_{max} \times s \right) \quad (4)$$

where $r(x)$ is the vital rate value at an adult population abundance x , N is the population abundance, K is the maximum carrying capacity, r_{max} is the vital rates maximum value, s is the scale factor, and t is the population density threshold above which density dependence effects are enabled. The default threshold is the occurrence threshold OT.

For direct mode parametrisation, the density dependence is a multiplier to the pre-calculated vital rate value

$$r(N) = r(0) \times \frac{K - N}{K} \text{ for } \frac{N}{K} > t, r(0) \text{ otherwise.} \quad (5)$$

where $r(x)$ is the vital rate value at an adult population abundance x , N is the population abundance, K is the maximum carrying capacity, and t is the population density threshold above which density dependence effects are enabled. The default threshold is 0, i.e. density effects will have an effect at any population size, if enabled.

4.4 Other vital parameters

4.4.1 Seed persistence

The seed persistence specifies the maximum number of year-to-year transitions a seed can survive. At the end of each year, all seeds older than the seed persistence, and which have not yet germinated, will be discarded. Seed persistence is independent of the environment. The seed persistence is a number of years ≥ 1 .

4.4.2 Minimum and maximum age of maturity

The minimum and maximum age of maturity specify the lower and upper age bounds needed for a juvenile to become adult. Juveniles who have not matured after a number of years equal to the maximum age of maturity will be discarded.

Note

For annual plants the minimum and maximum age of maturity needs to be set to zero (juveniles can only mature in the year of germination), for biannual plants both values need to be set to one (juveniles can only mature in the year after germination). See also [Presets](#).

4.5 Presets

Presets are an optional feature that sets selected vital rates and parameters to pre-selected values. Values that are defined via a preset are not required to be and should not be specified in the configuration file, even if they would normally be required parameters. The following presets currently exist in CATS:

- **tree** which is intended for long living species like trees. Currently it uses the same presets as other, i.e. no special presets
- **annuals** for which the **minimum age of maturity** and **maximum age of maturity** is set to 0, the **adult survival rate maximum** is set to 0 and the **maximum adult fraction of carrying capacity** is set to 1.
- **biennials** for which the **minimum age of maturity** and **maximum age of maturity** is set to 1, the **adult survival rate maximum** is set to 0 and the **maximum adult fraction of carrying capacity** is set to 0.5.
- **other**, which is the default preset and does not change any values.

4.6 Poisson processes

To add stochastic variation to the local population dynamics, Poisson processes are used. Whenever a quantity is derived by multiplying a vital rate with an population abundance (seeds, juveniles, or adults), a Poisson process is used. The product of vital rate and abundance is used as the expectation value of a Poisson distribution from which the result is drawn. This does not apply to the carrying capacity, which is used as a limit and is not multiplied with an abundance. When applicable, processes are mathematically reformulated that only the change is drawn from a Poisson distribution, e.g. the adult survival is not calculated by multiplying the adult abundance with the adult survival rate, but the number of plants that will not survive is calculated by multiplying the abundance with one minus the survival rate, i.e. the change of abundance is stochastically varied, not the resulting quantity, which would otherwise lead to exaggerated numerical fluctuations. The resulting changing quantity is also capped to the total quantity where applicable, e.g. the number of dying adults cannot exceed the total number of adults, and the number of germinated seeds is capped at the number of available seeds.

4.6.1 Poisson processes options

CATS offers the option to modify these Poisson processes. You can specify the maximum quantity that is drawn from the Poisson distribution. This is enabled per default for performance reasons. CATS uses the GSL (GNU Scientific library, <https://www.gnu.org/software/gsl/>) library for drawing from a Poisson distribution. The corresponding configuration parameter is **poisson maximum (optional)**, which defaults to 5000. If the quantity exceeds this maximum, either the whole quantity will be rounded, or split in two parts - the maximum will be drawn from the Poisson distribution, which is then added to the rounded remainder. This will change the shape of the distribution, but

not the expectation value. This behaviour is controlled by the configuration parameter [poisson maximum draw difference \(optional\)](#) which defaults to true.

Additionally, dampening can be applied to reduce the magnitude of stochastic processes. The configuration parameter [poisson dampening factor \(optional\)](#) specifies the complement of the fraction of the quantity that gets drawn from the Poisson distribution (with the above mentioned maximum limit), which is then added to the remainder. This dampening will only be applied if the original quantity is above the minimum threshold as specified by the configuration parameter [poisson dampening minimum \(optional\)](#) which defaults to 1.

5 Dispersal model

While the demographic model is restricted to cell-wise processes, the dispersal model connects these otherwise unconnected cells. The dispersal model is used to transport seeds, allowing interaction between cells. Every produced seed will undergo dispersal, but in practice most seeds will end up in their cell of origin. Seeds will only disperse once, in the year they are produced. CATS currently supports three different dispersal modes:

- kernel (raster) based dispersal, the default and only required mode, in which a raster (the dispersal kernel) is used to specify the fraction of seeds that are dispersed into cells relative to the source cell. Multiple kernels can be used in conjunction to model different transport vectors.
- (optional) long-range dispersal, where seed packets are transported to a random cell within a given range
- (optional) local dispersal, where a fraction of the locally produced seeds are directly deposited in the originating cell, bypassing other dispersal mechanisms.

The order is 1) local dispersal, 2) long-range dispersal, 3) kernel based dispersal(s).

Note that cell-local and kernel-based dispersal will disperse fractions of a cell's seed yield. For each cell of the simulation grid the number of seeds that were dispersed into it will be summed up after all dispersal processes have finished and will be used as expectation value in a Poisson process. If for example the sum of seeds dispersed into a cell far away from any populations is 1/10000, there is a 1/10000 chance that a seed will actually be assigned to that cell. See [Poisson processes](#).

5.1 Kernel based dispersal

This is the default dispersal mode. Multiple kernels can be used, e.g. to simulate different dispersal vectors, like endo-zoochoric, exo-zoochoric and wind dispersal. To fully parametrise this mode, a dispersal kernel (a raster layer) and a minimum and maximum weight need to be provided. For each cell where seeds

have been produced (and non-dispersed seeds are left after applying other, optional dispersal modes), the dispersal raster will be overlaid over the simulation raster, so that the middle cell of the dispersal raster is at the location of the originating cell. The values of the cells in the dispersal raster give the fraction of seeds transported using this kernel, that are transported to the corresponding simulation grid cell.

The raster layer has to fulfil the following requirements

- the raster layer must consist of square cells with the same resolution as the simulation grid
- the raster layer must be square, i.e. the number of rows must match the number of columns
- the number of cells per row and column has to be odd, so that there is a middle cell
- all cell values have to be in range $[0, 1]$
- the sum of all cells must be equal or less than one. If the sum is less than one, seeds are lost. This can be, for example, used to simulate endo-zoochoric dispersal, where some seeds are destroyed in the digestive system.

Multiple dispersal kernels of different sizes can be used. For each dispersal kernel, a minimum and maximum weight has to be specified. All seeds available to kernel based dispersal are assigned to the different kernels according to the weights. The weights are numbers in range $[0, 1]$ and have to sum up to less or equal one. A special value of -1 can be used to indicate that that kernel should take the remaining weight. The weights are processed in order of specification. For each cell, a random number between the minimum and maximum number is chosen from an uniform distribution for each dispersal kernel, and the corresponding amount of seeds assigned to that kernel.

Note

The kernel weights are processed in the order they are specified. If the minimum weights are 0.2, -1, 0.2 and the maximum weights are 0.4, -1, 0.4, then the first kernel will receive between 20% and 40% of all seeds, the second kernel the rest (60% to 80%), and no seeds will be dispersed using the third kernel. Kernel-based dispersal is the last dispersal mode to be simulated.

Note

The dispersal rasters' extent will have an effect on the simulation CPU time.

5.2 Local dispersal

Local dispersal is used to bypass any other dispersal mode. It is fully parametrised by a single value, a number in range $[0, 1]$, that specifies what fraction of seeds should immediately be dispersed into the originating cell and not available to other dispersal modes.

5.3 Long-range dispersal

Long range dispersal is used for rare long-range dispersal events and is parametrised using three values: a dispersal event probability, a dispersal radius (in cells) and a long range target count. For each cells in which seeds are available for dispersal, the probability for a long-range dispersal event to occur is given by the long range dispersal event probability, and is not influenced by other or previous long range dispersal events. If a long range dispersal event occurs, the target count specifies into how many target cells seeds transported by this long range dispersal event should be dispersed to. The corresponding number of cells are drawn from a square area centred on the source cell, extending to the source cell row $\pm$ dispersal radius cells rows and the source cell column $\pm$ dispersal radius columns. A cell can be selected multiple times, and excluded cells can also be selected, in which case the seeds are lost. The number of seeds that are dispersed into each target cell is equal to the seed yield of a single adult in the source cell, as long as there are remaining seeds to be dispersed in the source cell.

6 Vital rate parametrisation

6.1 Parametrisation modes

In CATS, vital rates depend on the local environment. CATS supports two parametrisation modes by which vital rate values are specified. Only a single mode can be chosen for a single simulation. That mode applies to all vital rates.

6.2 Hybrid parametrisation mode

In this mode, only a spatially explicit occurrence probability (e.g. from a species distribution model, SDM) is provided as input. This occurrence probability is used as a proxy of how suitable the environment is for the plant. The suitability is a value between 0 and 1. All vital rate values are either constant, or depend on this suitability. In general, CATS makes no assumption on the type of suitability dependence or shape of functional link between vital rate and the suitability. For non-constant vital rates, the actual vital rate will take a value between 0 and the specified maximum value for that vital rate. A larger suitability value will result in a larger vital value. Vital rates are formulated so that larger values correspond to larger local population growth rates.

Two thresholds have to be defined for the hybrid parametrisation mode:

- **occurrence threshold (OT)**, which is the threshold for population growth. The demographic model is calibrated, such that an isolated pop-

ulation (i.e. without seed immigration) stays constant over time (while allowing stochastic fluctuations) at OT. A good choice for the OT is the SDM's probability threshold for binary absence/presence prediction.

- **zero threshold (ZT)**, is the threshold for sink populations. Seeds will germinate, but not be produced. Under this threshold some vital rates values are set to zero. 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.

Note

CATS defines that at a suitability value equal to OT, the population size of an isolated cell (no incoming dispersal from other cells) stays constant over time. At suitability values less than OT the population abundance will shrink for an isolated cell, at suitability values greater than OT the population abundance will grow for an isolated cell.

6.2.1 Model calibration

CATS defines that at the occurrence threshold (OT) a (isolated, i.e. in a cell which does not receive seeds from the outside) population abundance remains stable over time. This is achieved by adding a correction term to the functions linking suitability to vital rate value. See [Vital rate to suitability link functions](#) for a list of available functions.

This correction term is a function of a hidden variable, the scale factor, which is the value linking the suitability to vital rates takes on at OT. To calculate the scale factor, a stage structured matrix model formulation of CATS is used. See [CATSpaper](#). The scale factor is varied until the dominant eigenvalue of this matrix model reaches 1 within a certain accuracy. This enables population abundance long-time (100000+ simulation cycles) stability, while allowing stochastic abundance fluctuations.

6.2.2 Vital rate to suitability link functions

CATS currently supports three different functions to link a vital rate value to suitability. The corresponding configuration parameter value is in brackets:

Constant link (constant)

The value of rate is independent of the suitability and equal to the vital rate maximum value above the suitability cut-off (if specified), and zero below.

Note

The constant link cannot depend on the population density.

Linear link (linear)

The link between suitability and vital rate is linear. The vital rate's value is equal to the maximum vital rate value at a suitability of 1.0 and is fit through the scale factor's value at a suitability equal to the occurrence threshold (OT).

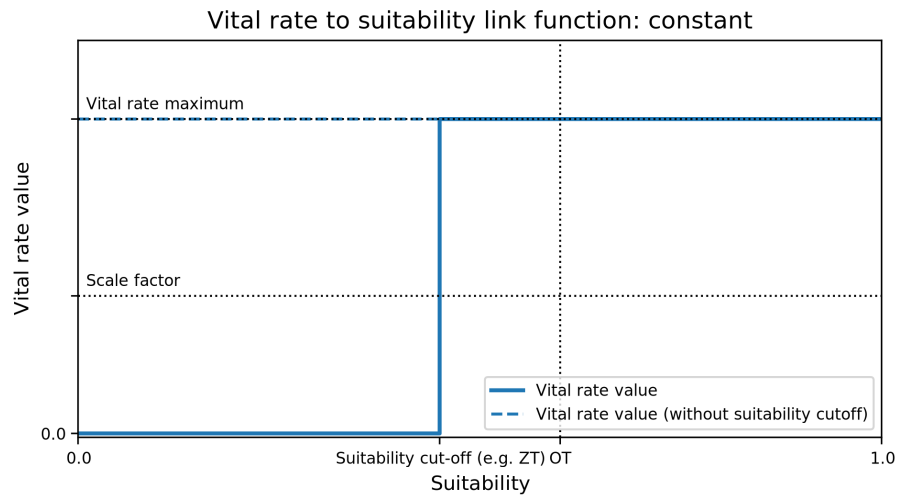


Figure 1: **Constant vital rate to suitability link.** Shown are the vital rate values (blue), with a suitability cut-off (solid) and without (dashed). The vital rate value is always equal to the vital rate maximum, except when a suitability cutoff is set. In this case, for suitability values less than this cutoff, the vital rate value is zero. The vital rate minimum (configuration parameter `<vital rate> minimum (optional)` is ignored.)

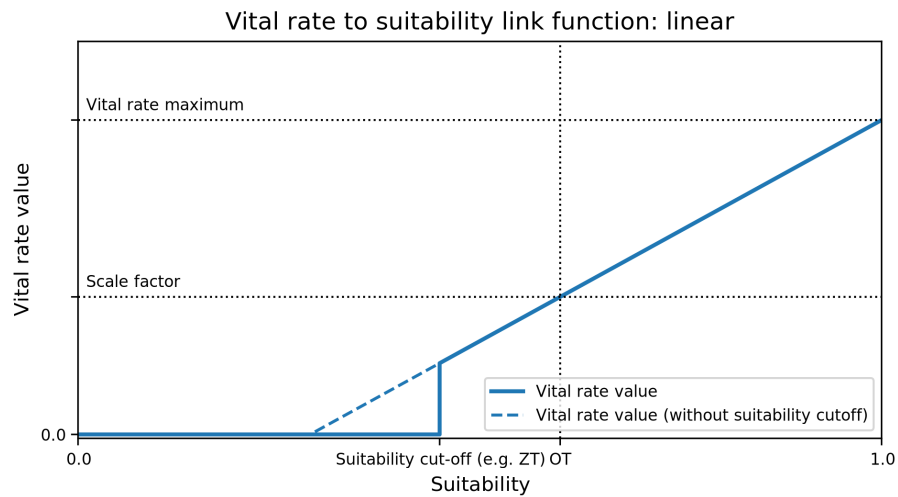


Figure 2: **Linear vital rate to suitability link.** Shown are the vital rate values (blue), with a suitability cut-off (solid) and without (dashed). The line is fitted so that the vital rate value is equal to the vital rate maximum value at suitability 1, and equal to the scale factor at OT. If a suitability cutoff is set, the vital rate value is zero for suitability less than this cutoff. Not shown are the optional population density effects.

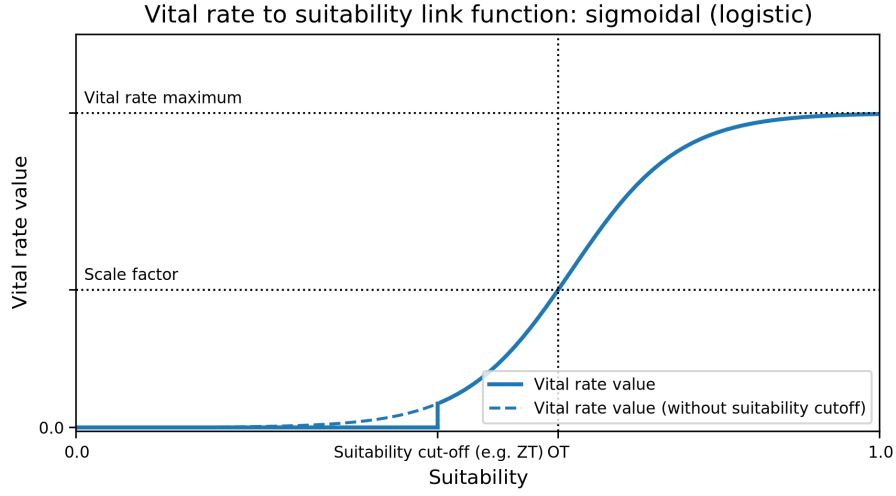


Figure 3: **Sigmoid (logistic) vital rate to suitability link.** Shown are the vital rate values (blue), with a suitability cutoff (solid) and without (dashed). The line is fitted so that the vital rate value is equal to the vital rate maximum value at suitability 1, equal to the scale factor at OT, and zero at suitability zero. If a suitability cutoff is set, the vital rate value is zero for suitability less than this cutoff. Not shown are the optional population density effects.

Logistic (sigmoid)

The function is logistic, spanning from zero at a suitability value of zero to the vital rate's maximum value at a suitability equal to one. The curve is fitted, so that its value, at a suitability equal to the occurrence threshold OT, is equal to the scale factor:

$$f(s) = \frac{\text{vital rate maximum}}{1 + \exp(ds \times (OT - s)) \times \left(\frac{\text{vital rate maximum}}{\text{rate at OT}} - 1 \right)} \quad (6)$$

where s is the local suitability, *rate at OT* is a scale term, which defaults to the vital rate's maximum value times the scale factor, and ds is the demographic slope term, which defaults to 15 and can be manually configured with the configuration parameter [demographic slope \(optional, hybrid mode only\)](#).

Note

The rate at OT can be directly specified, in which case the scale factor is not used. See [<vital rate> value at OT \(optional, hybrid mode only\)](#).

6.3 Direct parametrisation mode

In direct parametrisation mode, the concept of occurrence probability is not used. The values for each vital rate have to be specified directly for each cell of the

Density-dependent demographic rates for OT 0.5 and scale factor 0.5

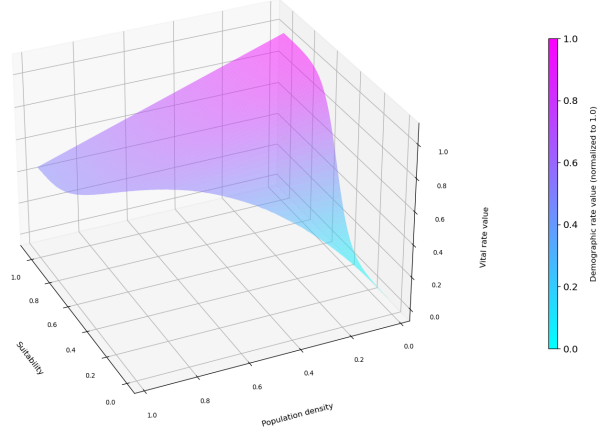


Figure 4: Vital rate value for the sigmoid al (logistic) vital rate to suitability link depending on suitability and population density (as fraction of the carrying capacity maximum). For this example, OT and scale factor were set to 0.5, and not suitability cut-off was set. See <https://cats.univie.ac.at/videos/tbd> for an animated version.

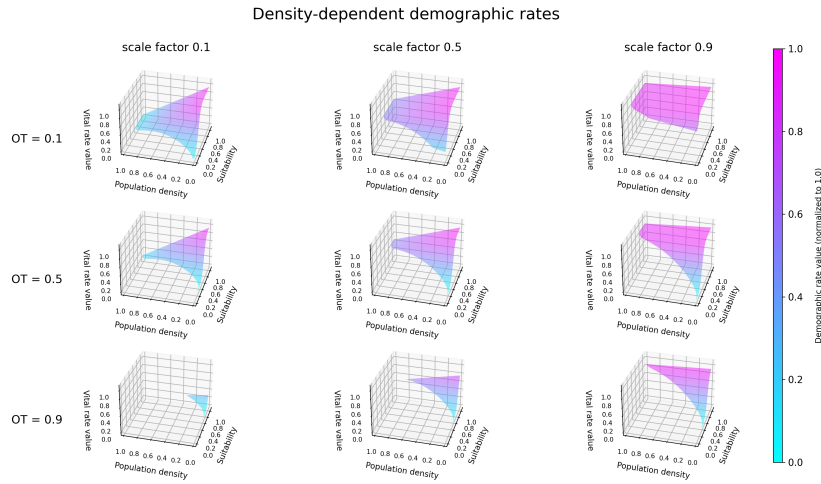


Figure 5: Effect of the scale factor and OT on the value for the sigmoid al (logistic) vital rate to suitability link. This figure shows the same information as Figure 4 for nine combinations of different scale factors and OT values.

simulation grid. Additionally, for each vital rate it can be specified whether it is affected by the adult population density. For each vital rate this can be done by either by

- providing input raster layer.
- calculating the vital rate values on the fly using generalized linear models (GLMs) of one or more predictor variables. The GLMs have to be fitted outside of CATS. The predictor coefficients are specified in the configuration file and are applied to spatially resolved predictor raster layers.
- providing a constant value for that vital rate. This value will be used for all years and raster cells.

If raster layers are used, either for directly specifying the vital rate values, or for use as predictor values for GLMs, they can also change in time, either by replacing the value after a specified number of simulation years, or by linear interpolation over time. See [Linearly interpolated environments](#).

Note

In direct parametrisation mode the maximum values for each vital rate has to be specified like in the hybrid parametrisation mode. The calculated or specified values will be capped at 0 and the maximum value, respectively.

6.4 Vital rate specification

Depending on parametrisation mode, the behaviour of vital rates can be specified. At a minimum, two values have to be specified for each vital rate:

- the vital rate's maximum value, corresponding to ideal conditions
- the environment to which the vital rate is linked or read from.

In hybrid parametrisation mode, all vital rates are linked to the suitability and the environment does not have to be explicitly specified individually. For direct parametrisation mode, the associated environment has to be specified for each vital rate individually.

Additionally, the behaviour of vital rates can be specified individually for each vital rate. If not explicitly specified, the behaviour of vital rates will use default values.

Below is a list of the most commonly used vital rate options. For a full list of options see [species \(section\)](#).

6.4.1 Options available in all parametrisation modes

Maximum value

The maximum numeric value a vital rate can reach under optimal conditions.

Vital rates can never exceed their maximum value, and will be capped at this value. See [<vital rate> maximum \(required\)](#).

Minimum value

The minimum numeric value for vital rates. Defaults to 0. The vital rate's calculated value, after applying all other options, will be clamped to the range between minimum and maximum value.

Density

whether the vital rate exhibits a (negative) dependence on adult population density, see [<vital rate> density \(optional\)](#).

Density threshold

The minimum adult population size (as fraction of the carrying capacity maximum) for population density effects. Below this threshold, no population density effects will occur. See [<vital rate> density threshold \(optional\)](#). If not otherwise specified, the default threshold is set to 0.0 for direct parametrisation mode and equal to scale factor for hybrid parametrisation mode. This ensures that at OT an isolated population stays stable in size, independent of the number of adults.

6.4.2 Options available only in hybrid parametrisation mode

Function

the shape of the functional link between suitability and rate value, see [Vital rate to suitability link functions](#) and [<vital rate> function \(optional, hybrid mode only\)](#).

Suitability cut-off

the minimum suitability threshold under which the rate is zero, see [<vital rate> suitability cutoff \(optional, hybrid mode only\)](#).

Scale factor

Allows to override the scale factor for a specific vital rate. See [<vital rate> scale factor \(optional, hybrid mode only\)](#). The scale factor for a specific vital rate is taken into account when calculating the scale factor for all other vital rates.

Minimum scale factor

Allows to specify a minimum value for a vital rate's scale factor. See [<vital rate> minimum scale factor \(optional, hybrid mode only\)](#). The minimum scale factor for a specific vital rate is taken into account when calculating the scale factor for all other vital rates.

6.4.3 Options available only in direct parametrisation mode

Environment

The name of the linked environment, see [<vital rate> environment \(required, direct mode only\)](#).

Environment multiplier

A numeric value that will be multiplied with the environment's value for this specific vital rate. Used e.g. when multiple vital rates share the same environment with different scales. See [<vital rate> environment multiplier \(optional, direct mode only\)](#).

6.5 Vital rate defaults

6.5.1 Hybrid parametrisation mode defaults

In [Hybrid parametrisation mode defaults](#) the default parametrisation values for all vital rates are given. The column `suit. cut-off` specifies the minimum suitability under which the vital rate's value is equal to zero. Note that for the hybrid parametrisation modes density effects are only enabled if the local population abundance exceeds the carrying capacity at the occurrence threshold, or the maximum carrying capacity times the scale factor, if the carrying capacity is constant.

6.5.2 Direct parametrisation mode defaults

In direct parametrisation mode, all optional configuration parameter are disabled or zero, except density dependence, which is enabled for the same vital rates as in hybrid parametrisation mode, however there is no minimum suitability for density effects to occur, as the concept of suitability does not exist in direct parametrisation mode.

7 Initial population distribution

CATS requires an initial, starting population distribution provided as raster file, see configuration parameter [initial population filename \(required\)](#). The initial population grid will be used as the reference grid for size, resolution, and optionally extent and projection information. The initial distribution can be either provided as absence/presence or abundance data. The default behaviour of CATS is to increase the initial population sizes to the local carrying capacity. This can be controlled via the configuration parameter [increase initial population to carrying capacity \(optional\)](#). In hybrid parametrisation mode the initial population can be restricted to cells with a minimum suitability, see configuration parameter [initial population suitability threshold \(optional, hybrid mode only\)](#).

8 Environments

Environments are spatially resolved data, i.e. raster layers describing the simulation area. They are used to either calculate or directly specify the values of vital rates. Each environment (except the constant value environment) consists of one or more input raster of values. There are three different types of environment (and additionally a special case, the constant value layer, see [Constant value environments](#))

Table 1: Hybrid parametrisation vital rate defaults

Vital rate	link to suitability	density dependence	suitability cut-off
carrying capacity	sigmoid	no (not allowed)	0
clonal growth rate	sigmoid	yes (negative)	ZT
germination rate	sigmoid	yes (negative)	ZT
flowering frequency	sigmoid	no	OT
seed yield	sigmoid	no	OT
germination to adult survival	sigmoid	yes (negative)	ZT
seed survival rate	constant	no	0
adult survival rate	sigmoid	no	0

- suitability layers (hybrid parametrisation mode only)
- pre-calculated vital rate layers (direct parametrisation mode only)
- GLM vital rate layers, consisting of one or more predictor rasters (direct parametrisation mode only)

For each environment the environment type, and a file name or file name pattern (except for constant environments) has to be specified. Depending on how the environment changes over time, some elements of this file name pattern will be substituted.

8.1 Temporally changing environments

Environments usually change over the course of the simulation, e.g. due to the effect of changing climatic conditions on vital rates. How environments change over time in CATS can be specified on a per-environment basis. At the start of the simulation, each environment is loaded. There are three options for temporal development of environments.

Note

For GLM environments, the temporal changes have to be specified per predictor, not per environment. Different predictors within the same GLM can have different temporal behaviour, e.g. a soil raster could be static, a precipitation layer discretely replaced every year, and a temperature raster linearly interpolated over 5 year intervals.

8.1.1 Temporally constant environments

Static environment stay constant over the whole simulation duration, but value varies in space. The file name pattern is used as is, no substitution will be done.

Example:

```
[environment: soil]
type = predictor
filename pattern = input-layers/soil-layer.tiff
```

8.1.2 Discretely changing environments

Discretely changing environments stay constant over a specified time, and are replaced after a specified number of years. For these environments, a replacement interval (in years) and a file name pattern have to be specified.

In the first year of the simulation and after a number of simulation years equal to the replacement interval, the file name of the input raster filename is assembled using the filename pattern and the current simulation year. Any occurrence of '%Y' in the filename pattern is replaced with the current simulation year. See [Filename patterns](#) for details. In between the replacement intervals the environment stays constant.

Example:

```
[environment: pinus_mugo_adult_survival]
type = direct rate
replacement interval = 10
filename pattern = input-layers/pinus_mugo_%Y.tiff
```

8.1.3 Linearly interpolated environments

Similar to the discretely changing environments. Instead of the a reload interval a interpolation interval is specified. In the first simulation year, the raster layer for that year, and for the year plus the interpolation is loaded. For years in between these years, the values will be interpolated linearly. After each number of interpolation interval simulation years, these input raster values will be updated for the corresponding start and end year.

Example:

```
[general]
start year = 2030
simulation length = 32

[environment:pinus_mugo]
type = suitability
interpolation interval = 10
filename pattern = input-layers/pinus_mugo_%Y.tiff
```

For this example, in the first simulation year (2030) the input rasters 'input-layers/pinus_mugo_2030.tiff' and 'input-layers/pinus_mugo_2040.tiff' will be loaded. For the simulation year 2030 and 2040 the corresponding values will be used, in between the values will be linearly interpolated. In the simulation year 2040, the values will be updated to 'input-layers/pinus_mugo_2040.tiff' and 'input-layers/pinus_mugo_2050.tiff' and so on. This means, that the input raster layers for 2030, 2040, 2050 and 2060 2070 have to be available. If the simulation length was only 30 years, only the input rasters for 2030 to 2060 have to be available.

8.2 Hybrid parametrisation mode environments

In hybrid parametrisation mode only one type of environment is supported: [Suitability environments](#).

8.2.1 Suitability environments

Suitability environments are only used for the hybrid mode of CATS, where all vital rates for species depend on a single suitability environment. A suitability environment consist of a raster describing the (climatic or otherwise) suitability

of each cell in the simulation environment, usually derived from a SDM's occurrence probability, which is calculated outside of CATS. Each cell is assigned a value between 0 (totally unsuitable) to 1 (perfectly suitable).

Note

If the input data is not in range $[0, 1]$ but in $[0, N]$ it is possible to specify a divisor in the configuration file, see [suitability divisor \(optional, hybrid mode only\)](#).

Note

"No data" values are treated like suitability 0. If you wish to exclude cells, use an exclusion overlay.

8.3 Direct parametrisation mode environments

8.3.1 Pre-calculated vital rate environments

Pre-calculated vital rate environments are rasters that directly specify the spatially resolved values for a vital rate. The values for the vital rates are taken directly from this raster and clamped between zero and the specified maximum value for the corresponding vital rate.

A multiplier value for the corresponding range, if specified, is applied pre-clamping. This allows the same vital rate environment to be used for multiple vital rates with different ranges, e.g. seed yield and flowering frequency, if these rates are multiples of each other.

8.3.2 GLM environments

Note

GLM environments were rewritten and are not yet fully tested.

GLM environments are a convenience feature for the direct parametrisation mode of CATS. GLM environments are calculated from one or more input (predictor) raster, e.g. precipitation or temperature information, to which a generalized linear model is applied. In addition to the predictor rasters, a list of coefficients for the model is needed. Multiple environments can share the same predictors with different coefficients. CATS only supports a small subset of GLMs: predictor terms up to second (quadratic) order, without interaction terms. These GLMs have to be parametrised outside of CATS. To use them in CATS, the raster layer for the predictors and their model coefficients have to be provided.

Note

Currently only polynomial GLMs up to 2nd degree without interaction terms between predictors are supported in CATS. If the GLMs are calculated in R, the option `"raw = TRUE"` has to be specified for the `poly()` function.

Supported GLM families

Currently, the following GLM families are supported in CATS:

- binomial
- poisson
- bernoulli
- gaussian

GLM environment example

8.3.3 Constant value environments

Constant value environments are a special case. They are constant in space and time, and have no external raster layers associated with them. They are used for vital rates that stay constant over the whole simulation duration. The same specified value is used for all cells and each year. No input raster layer is needed, no filename pattern is required. The only required configuration parameter is its value.

Example:

```
[environment:pinus_mugo_seed_survival]
type = constant
value = 0.75
```

9 Overlays

Overlays are rasters that assign each cell a value, modifying the cell's properties. Overlays are an optional feature. A single simulation can only have one of each type of overlay. Overlays usually describe discrete properties of the underlying cell and therefore are never interpolated over time, but can either be static (valid for the whole duration of the simulation), or replaced with another input raster at regular intervals. CATS currently supports two types of overlays: exclusion overlay and habitat layer (carrying capacity) overlays.

9.1 Exclusion overlays

Exclusion overlays allow you to exclude cells from the simulation. Any cells marked excluded are ignored, and any existing population individuals will be destroyed, so will seeds that are dispersed into excluded cells. If a change of the exclusion overlay results in an excluded cell no longer being excluded, the cell is treated like a regular simulation cell again, although it starts empty. Excluded cells do not inhibit or change dispersal across it. This can be used to mark cells as ocean, or other inhospitable areas, in which no plants can occur. Raster used for the exclusion overlay should be integer-valued. Any non-zero value in the

corresponding cell will mark it as excluded. No data values in the input raster will be treated as excluded cells.

9.2 Habitat layer (carrying capacity) overlay

This overlay allows to link (abstract) habitat types to a modifier carrying capacity. It consists of two parts: an input raster, which defines the habitat type (an integer number) for each cell, and an associated CSV file, that lists a carrying capacity multiplier (a number ≥ 0) for each habitat type. For no data raster cell values, the multiplier is 0. The multiplier is applied to the calculated carrying capacity, not its maximum. Additionally, a default multiplier can be defined for values in the input raster, that have no corresponding entry in the CSV file.

9.2.1 CSV file format

The CSV file for this overlay is a text file with comma separated values and a header line, which defines the column names. The habitat type is found in column 'habitat_type' and the carrying capacity multiplier in column 'cc_multiplier'. The minimal CSV implementation in CATS does not support quoting, so any comma, even if escaped or quoted, will be used as a field separator.

Example:

```
habitat_type,cc_multiplier
1,0.1
7,0.3
3,0
9,1.5
```

10 CATS Output

CATS will print informative data to the screen. In addition, grid statistics for each year will be saved to a CSV file, and the adult population abundance distribution as GeoTIFF file. Output files will include the run name (see [run name \(required\)](#)) and the replicate number (see [List of command line options](#)), in addition to the species name (see [species \(section\)](#), where applicable). All output data will be saved in sub-directories of the output directory, see [output directory \(optional\)](#).

10.1 Population grids

10.2 Grid statistics

10.3 Population growth rates grids

11 Input data

CATS uses three different file formats for input. The configuration uses a text-based file format based on INI files, which consists of key-value pairs structured

into sections. For spatially resolved data rasters in GeoTIFF format are used. Only the data in the first band is currently used. With the exception of dispersal kernels, all rasters have to have same resolution, projection information and extent. Dispersal kernels only need to match in resolution. The last format is a (subset of) CSV, which is currently only used for the habitat (carrying capacity) overlay, see [CSV file format](#). The simulation resolution and projection is read from the initial population file.

11.1 Filename patterns

11.2 Raster files

CATS uses GDAL ([GDAL](#)) for loading and writing spatially resolved raster data. Currently only one raster band, the first, is used and only [GeoTIFF](#) files and [Arc/Info ASCII Grid](#) for dispersal kernels) have been tested. Only rasters with square cells, i.e. the same horizontal and vertical resolution are supported. An equal-area projection is recommended if using larger simulation extents.

Note

CATS uses values as-is, without applying band scaling.

11.2.1 Dispersal kernels

Dispersal kernels are rasters that specify the dispersal probability for each cell relative to the middle cell, which is overlaid over the dispersal source cell. This means that dispersal kernels are required to have an odd number of rows and columns. Additionally, the current version of CATS expects the the number of rows and columns to be the same. Only the resolution of the dispersal kernel is matched against the simulation resolution. See [Kernel based dispersal](#).

11.3 Configuration files

CATS configuration files are based on the INI file format, i.e. they are structured text-files. A configuration file is required for each CATS simulation. The configuration file includes information about the simulation duration, species and dispersal parametrisation, and specifies input and output file names. The name of the configuration file is passed to CATS as a command line argument, see [Command line arguments](#).

Note

For annotated configuration file examples see the [Quick start guide](#), or [Configuration examples](#).

11.3.1 Configuration file format

The configuration file is a text based on the INI format. The data in CATS configuration files is contained in key-value pairs, which are grouped into sections. Each configuration value can uniquely identified by the section and key

name. This means that each section name may only appear once per configuration file, and each key name only once per section. The order of sections within the configuration file, and the order of key-value pairs within each section does not matter. Only one section declaration one or key-value pair or one may be specified per line.

11.3.2 Configuration sections

A section is specified by a section name in square brackets. The section name consists of a type identifier, and a user chosen identifier, separated by a colon. This allows multiple sections of the same type to be specified. The exception is the `[general]` section, which does not have a user chosen identifier. Section type identifiers are pre-defined in CATS, while the user chosen identifier can be chosen freely. Some configuration values might reference other sections, in this case only the user chosen identifier has to be specified.

Example for section names:

- `[general]`
section type 'general'
- `[species:pinus_mugo]`
section type 'species', identifier 'pinus__mugo'
- `[environment:germination_rate]`
section type 'environment', identifier 'germination__rate'

11.3.3 Key-value pairs

All key-value pairs following the section declaration are associated with the most recently declared section. If the same key is specified more than once per section, the resulting configuration behaviour is undefined.

A key-value pair consists of a key name and a value, separated by an equal sign (`=`). Key names are case-sensitive. White-space before and after key name and value are ignored, but not within the key name or value. Depending on the key, valid values can be numbers, strings, or list of either numbers or strings. List of numbers or strings are separated by a comma `,`.

Disallowed characters in keys and values:

- comma `,` is used for separating values in a list
- equal sign `=` is used for separating key and value
- comment character `#` are used to indicate comments. All characters following the comment character until the end of the line are ignored. See [Comments](#) below.

```
some_key_name=some_value
some_key_name=some_value
some_key_name=some_value
```

are equivalent ways to assign `some_value` to `some_key_name`, but

```
some_key_name=some_value
```

or

```
some_key_name=some_value
```

are not, because the spaces within keys or values refer to different keys and values.

11.3.4 Comments

Comments are indicated by the comment character `#`. Lines starting with `#` are considered comment lines and are ignored. Comments may also be placed after section names or key-value pairs. Any characters following (and including) the comment character are also ignored. This means that `#` is not allowed as part of section names, key names or values.

CATS optionally outputs all (optional) configuration parameters it tried to read but that were not specified, and values that were specified in the configuration file but not read, see [Debugging options](#). If a required configuration parameter is not found, CATS will terminate with an error message.

Note

Additionally to the configuration file, some data can only be provided by command line arguments to CATS, e.g. the replicate number. See [Command line arguments](#).

11.3.5 Configuration parameter value data types

CATS configuration parameter values can be one of the following data types.

- **Strings values:** String values are specified as is, without quotation marks. Space characters before the first and after the last non-space character are ignored, see above.
- **Floating point values:** Can be either specified as numerical value, or in exponential notation, e.g. `1.2e1` and `12.0` are equivalent.
- **Integer values:** Same as floating point values, only that the number is cast to an integer, and the part after the decimal separator is discarded, e.g. `12.7` will result in an integer value of `12`.
- **Boolean values:** Either `true` or `false`. Values of `1`, `t`, `y`, `yes`, or `true`

will be interpreted as true, while values of 0, f, n, no, or false will be interpreted as false. Boolean values are case insensitive.

12 Configuration parameters

This section lists all required and optional configuration sections, and their required and optional keys. For each section, first all required configuration parameters are listed in alphabetical order, followed by the optional configuration parameters in alphabetical order. Some required configuration parameters are only used depending on another optional parameter, and therefore are marked as optional.

The section and parameter names are listed in their respective headings. Additional information is stated in brackets, which are not part of the name.

Some configuration parameter names contain a term angular brackets, e.g. `<vital rate>` or `<predictor>`. This term (including the angular brackets) is a placeholder, which needs to be substituted with either the name of a vital rate, or the name of a user-specified predictor (see there).

12.1 general (section)

The general section is required. Unlike other sections, it can only appear once per configuration file, and as such, the section name only consists of the section type identifier; a user-specified identifier is not allowed.

12.1.1 output interval (required)

Integer value ≥ 0 . Specifies how often (in years) adult population and other distribution grids should be saved to disk. A value of zero disables the output. If a value greater than zero is specified, the grids will be saved in the first simulation year, and every `output interval` years after that.

12.1.2 run name (required)

String value. This parameter is used, in conjunction with the replicate number, to identify a single simulation run. It is used as part of the file name of output files like grid statistics, adult population grids, etc.

12.1.3 starting year (required)

Integer value. Specifies the calendar year of the first simulation year. This is used to assign each simulation time step a calendar year. This information is used to load the correct environment data for the corresponding simulation year, and for output file names.

12.1.4 time steps (required)

Integer value > 0 . Specifies the number of simulation time steps (years) a simulation should run. This value is only used for the simulation phase, not including optional burn-in or warm-up phase years.

12.1.5 always include sign in year (optional)

Boolean value, default false. If enabled, the date format in file names for loading and saving data will always include the sign (either – for years less than, and + for years greater than zero). Otherwise, only negative years will be prefixed with a minus sign.

12.1.6 burn-in phase year (optional)

Integer value. Specifies the calendar year for the burn-in phase, which is used for filename. If specified, enables the burn-in phase.

12.1.7 burn-in phase length (conditional)

Integer > 0 . Required if burn-in phase is enabled. Specifies how many years the burn-in phase should last.

12.1.8 compression (optional)

Legacy, maybe remove.

12.1.9 debug (optional)

Still undocumented.

12.1.10 file format (optional)

Legacy, maybe remove.

12.1.11 ignore raster projection (optional)

Boolean value, default false. If enabled, input raster projection information will not be verified, only the resolution and number of rows and columns of all input rasters are required to match. The exceptions are dispersal kernel raster layers, which can be of any extent, but have to match in resolution.

12.1.12 output directory (optional)

String value. Path to the base output directory, which needs to exist, either absolute or relative to the current working directory. Defaults to cats-output. The output directory will be created if needed.

Note

This configuration parameter can be overridden by the comment line argument if specified, see [General command line options](#).

12.1.13 poisson dampening factor (optional)

Floating point value in range $[0, 1]$. Default 0. For a description see [Poisson processes options](#).

Note

This configuration parameter can be overridden by the comment line argument if specified, see [Random number options](#).

12.1.14 poisson dampening minimum (optional)

Floating point value in range ≥ 0 . Default 1.0. For a description see [Poisson processes options](#).

Note

This configuration parameter can be overridden by the comment line argument if specified, see [Random number options](#).

12.1.15 poisson maximum draw difference (optional)

Boolean value. Default true. For a description see [Poisson processes options](#).

Note

This configuration parameter can be overridden by the comment line argument if specified, see [Random number options](#).

12.1.16 poisson maximum (optional)

Floating point value ≥ 0 . Default value 5000. For a description see [Poisson processes options](#).

Note

This configuration parameter can be overridden by the comment line argument if specified, see [Random number options](#).

12.1.17 stats populated threshold (optional)

Integer > 0 . Default 2. All grid statistic summaries will be calculated twice, first a cell will count as populated if the adult population size is at least 1, and an additional time, where a cell only counts as populated if the adult population exceeds this threshold.

12.1.18 threads (optional)

Integer value ≥ 1 , default 1. Specifies the maximum number of threads that are spawned for computations, i.e. the maximum number of CPU cores that will be used at the same time, if possible.

12.1.19 warm-up phase year (optional)

Integer value. Needs to be less than starting year. Specifies the start year of the warm-up phase. If this value is set, the warm-up phase will be enabled. See [Simulation phase](#).

12.1.20 write all (optional)

Boolean value, default false. If this option is enabled, in addition to the adult population grids, the juvenile and seed count grids (one per age class each) will be saved to disk at the same time.

12.1.21 zero pad year digits (optional)

Integer value ≥ 0 , default 0. Specifies the format of the year string in input and output file names. The year will be zero-padded to a minimum of the specified number of digits (including the sign), corresponding to the `%0xd` format of `printf`, where `x` is the specified value.

Example:

`zero pad year digits = 4` will result in the year string 0023 for year 23, +023 for the same year, if `always include sign in year = true`, and -32552 for year -32552.

12.2 species (section)

The species section contains the parametrisation of a single species. The user-specified identifier is used to define the name of that species and is used for output file names and information.

Example:

`[species:pinus_mugo]` contains the parametrisation for species with the identifier `pinus_mugo`.

12.2.1 dispersal (required)

String. Specifies the dispersal configuration section associated with this species. Only the user-specified identifier has to be given.

Example:

`dispersal = some dispersal` specifies that the dispersal information is contained in section `[dispersal:some_dispersal]`

12.2.2 initial population filename (required)

String (filename). Specifies the location of the initial population distribution for this species.

12.2.3 OT (required, hybrid mode only)

Numeric value in range $(0, 1]$. Specifies the occurrence threshold for this species.

12.2.4 maximum age of maturity (required)

Integer ≥ 0 . Must be greater than or equal to `minimum age of maturity`. Specifies what age (in years) a juvenile plant can reach (starting at 0 in the year of germination), before it either becomes adult or is discarded. Juvenile

plants have a chance to become mature in the years between minimum and maximum age of maturity and will be discarded afterwards. For annual plants, both minimum and maximum age of maturity is 0. For biannual plants, both ages are 1.

12.2.5 minimum age of maturity (required)

Integer ≥ 0 . Specifies what age (in years) a juvenile plant must reach (starting at 0 in the year of germination), before it either can become adult. Juvenile plants have a chance to become mature in the years between minimum and maximum age of maturity and will be discarded afterwards. For annual plants, both minimum and maximum age of maturity is 0. For biannual plants, both ages are 1.

12.2.6 seed persistence (required)

Integer ≥ 1 . Number of years a seed can survive in the ground and has a chance to germinate, before it is discarded. A number of 1 means that seeds have only one chance at germination, in the year after dispersal. If they seeds fail to germinate in this time span, they are discarded.

12.2.7 suitability (required, hybrid mode only)

String, required for hybrid mode. Specifies the user-defined identifier for the link environment. That environment must be of type 'suitability'.

Example

: `suitability = pinus_mugo_rcp45` specifies that the linked suitability is specified in section `[environment:pinus_mugo_rcp45]`

12.2.8 <vital rate> environment (required, direct mode only)

String value, user identifier of a environment section (except of type predictor) or a GLM section. This value specifies the environment to use for this corresponding vital rate.

12.2.9 <vital rate> maximum (required)

Numeric value. Specifies the maximum value for the corresponding rate. Vital rate values will be capped at this value. The <vital rate> maximum must be in range (0, 1] with the following exceptions:

- except seed yield and carrying capacity, where the value must be equal or greater than 1.
- except clonal growth rate, where the value must be equal or greater than 0. The clonal growth rate maximum can be omitted (i.e. is not required), in which case it is set to 0 (clonal growth disabled).

12.2.10 ZT (required, hybrid mode only)

Numeric, in range [0, OT). Specifies the zero threshold for this species. This value is used as a default value for the [<vital rate> suitability cutoff \(optional, hybrid mode only\)](#) for vital rates related to reproduction, see [Hybrid parametrisation mode defaults](#).

12.2.11 demographic module (optional)

Undocumented, because the implementation has not yet been finalized.

12.2.12 demographic slope (optional, hybrid mode only)

Numeric value > 0 . Default value 15.0. Specifies the slope of the sigmoid function used in hybrid mode. See [Logistic \(sigmoid\)](#).

12.2.13 hapaxanthly (optional)

Numeric value in range [0, 1]. Default value 0. If enabled (> 0), specifies the fraction of adults that will die after having produced seeds.

12.2.14 increase initial population to carrying capacity (optional)

Boolean value, default true. If enabled, the initial population will be set to local carrying capacity, independent of the abundance specified by the initial population raster. If disabled, the initial population will be capped at the local carrying capacity, but never increased beyond the specified value.

12.2.15 initial population suitability threshold (optional, hybrid mode only)

Numeric value in range [0, 1] or the strings **ZT** or **OT**. Default value 0.0. Specifies the minimum required suitability for the initial population to be set.

12.2.16 maximum adult fraction of carrying capacity (optional)

Numeric value in range (0, 1]. Specifies the maximum fraction of the carrying capacity that adult individuals can occupy. Default 0.5. [Carrying capacity](#) The default value can be overridden with presets, see [Presets](#).

12.2.17 preset (optional)

String value, one of **tree**, **annual**, **biennial**, or **other**. Selects some default vital rate and parameter default values. See [Presets](#).

12.2.18 save lambda grid (optional)

Boolean value, default false. If enabled, the calculated population growth (lambda) value for each cell of the grid (in isolation) will be saved to disk. Density effects will not be considered, a population abundance of 1 will be assumed.

The resulting calculation is very slow! In hybrid mode, values will be cached, but the calculation will still be slow.

12.2.19 `save lambda grid density` (optional)

Boolean value, default false. As `save lambda grid`, but the cell's actual population abundance will be considered.

Warning:

Very slow calculation!

12.2.20 `<vital rate> density` (optional)

Boolean value. Default value depends on vital rate, see [Hybrid parametrisation mode defaults](#). If enabled, the vital rate is negatively population density-dependent, see [Population density effects](#).

12.2.21 `<vital rate> density threshold` (optional)

Numeric value in range $[0, 1]$. Specifies what the minimum population (as fraction of the maximum carrying capacity) is to enable population density effects. For hybrid parametrisation mode, the default is equal to the scale factor, if the vital rate is density dependent. For direct parametrisation mode, the default value is 0.0

12.2.22 `<vital rate> environment multiplier` (optional, direct mode only)

Numeric value. Default value 1.0. Specifies an optional multiplier which is applied to the environment's values.

12.2.23 `<vital rate> function` (optional, hybrid mode only)

String value. Specifies the name of the function that is applied to the suitability to calculate the vital rate's value. See [Vital rate to suitability link functions](#) for a list of supported values.

12.2.24 `<vital rate> minimum` (optional)

Numeric value. Default value 0.0. Specifies the minimum value for vital rate.

Note

In hybrid mode, a vital rate's value can fall below the minimum if a suitability cut-off is specified. In this case, the vital rate's value will be zero if the suitability falls below the specified threshold.

Note

In hybrid mode the minimum is ignored if the link function to suitability is constant. Only the (mandatory) maximum value and the suitability cut-off will be used.

12.2.25 <vital rate> minimum scale factor (optional, hybrid mode only)

Numeric value in range (0, 1). Default 0. If set, the scale factor for will not be less than the specified value, or the calculated scale factor, if larger. The minimum scale factor for a specific vital rate is taken into account for calculating the scale factor for all other vital rates.

Note

The default minimum scale factor for the `adult survival rate` is not 0, see [Presets](#).

12.2.26 <vital rate> scale factor (optional, hybrid mode only)

Numeric value in range (0, 1). If set, the scale factor for the corresponding vital rate will be exactly this value, independent of the calculated scale factor's value. The scale factor for a specific vital rate is taken into account for calculating the scale factor for all other vital rates.

12.2.27 <vital rate> suitability cutoff (optional, hybrid mode only)

Numeric value in range [0, 1] or one of ZT or OT. Default value depending on vital rate, see [Vital rate defaults](#). Specifies the minimum required suitability for this vital rate. For suitability values less than this cut-off, the vital rate's value will be zero, even if a vital rate minimum value (configuration parameter `<vital rate> minimum (optional)`) is set.

12.2.28 <vital rate> value at OT (optional, hybrid mode only)

Numeric value in range [0, 1]. If specified, this option sets the vital rate's value at a suitability equal to the occurrence threshold OT. Otherwise, the vital rate's value at OT will be equal to the scale factor times the vital rate's maximum value. This option is primarily used for ensuring that adult survival does not fall off to quickly with falling suitability.

Note

This value is ignored if the vital rate's functional link to suitability is constant.

Note

This option will effectively disable the scale factor for this vital rate. Model calibration via calculation of the scale factor will be affected and might not be possible, if too few vital rates use the scale factor.

12.3 overlay (section, multiple allowed)

Overlays are raster layers that assign each cell of the simulation area a value, that modifies the local behaviour. See [Overlays](#). Currently supported overlay types are:

- **exclusion**
- **habitat carrying capacity**

12.3.1 filename pattern (required)

String. Specifies the pattern of the overlay raster filename, i.e. the path (relative to the current working directory) and filename of the input raster. If the string %Y appears in the pattern, and a reload interval is specified, it will be replaced with the current simulation year (according to the year formatting options specified in section general).

12.3.2 type (required)

String. Must be one of the supported overlay types. See above. Only one overlay per overlay type can be used in a simulation.

12.3.3 default value (optional)

Numeric value in range [0, 1]. Required (and only used) for overlay type **habitat carrying capacity**. Specifies the carrying capacity multiplier in the input raster not found in the translation table.

12.3.4 enabled (optional)

Boolean value, default true. If set to false, the overlay will not be used (but has to be correctly specified).

12.3.5 reload interval (optional)

Integer value ≥ 0 , default 0. If set to a value larger than zero, the input raster for this overlay will be replaced at this interval in simulation years, with %Y in the filename pattern replaced with the current simulation year.

12.3.6 table filename (optional)

String. Required (and only used) for overlay type **habitat carrying capacity**. Specifies the name of the translation table, which assign carrying capacity multipliers to overlay raster values. See [Habitat layer \(carrying capacity\) overlay](#).

12.4 environment (section, multiple allowed)

Environments are used as input raster layers for demographic processes, and quantify the vital rates' local values, either directly (suitability, direct rate, constant), or indirectly (GLMs). Currently supported environment types are:

- **suitability** Hybrid mode only.
- **predictor** For use with GLMs only. Cannot be used on their own, but only as part of a GLM
- **direct rate** Direct mode only.

- **constant** Direct mode only. Special case, does not require a raster file.

12.4.1 type (required)

String value. One of the supported environment types, see [environment](#) (section, multiple allowed).

12.4.2 filename pattern (required)

String. Required except for environment type **constant**.

12.4.3 interpolation interval (optional)

Integer value ≥ 0 , default 0. Not used for environment type 'constant'. Mutually exclusive with replacement interval (see below). If specified, values in this environment will be cell-wise linearly interpolated over time with endpoints the specified number of years apart.

Example:

The following (partial) configuration example, the simulation starts in the year 2021 and runs for 15 years, until and including the year 2046. The calculated suitability values would be calculated following this scheme:

```
year 2021: 1.0 × suit/pmugo_2021.tiff + 0.0 × suit/pmugo_2031.tiff
year 2022: 0.9 × suit/pmugo_2021.tiff + 0.1 × suit/pmugo_2031.tiff
...
year 2030: 0.1 × suit/pmugo_2021.tiff + 0.9 × suit/pmugo_2031.tiff
year 2031: 1.0 × suit/pmugo_2031.tiff + 0.0 × suit/pmugo_2041.tiff
...
year 2046: 0.5 × suit/pmugo_2041.tiff + 0.5 × suit/pmugo_2051.tiff
```

All input layers, i.e. for the years 2031, 2041, and 2051 have to exist.

[general]

```
starting year = 2021
time steps = 25
```

[environment:pinus_mugo]

```
type = suitability
interpolation interval = 10
filename pattern = suitability/pmugo_%Y.tiff
```

12.4.4 replacement interval (optional)

Integer value ≥ 0 , default 0. Not used for environment type 'constant'. Mutually exclusive with interpolation interval (see above). If specified, the input rasters will be replaced after every specified years, with

12.4.5 save environment (optional)

Boolean value, default false. Not used for environment type **constant**. If enabled, this environment's current values (after interpolation or predicting a GLM) are saved to disk in the directory 'environments'. Used for debugging.

12.4.6 suitability divisor (optional, hybrid mode only)

Numeric value $\neq 0$. Only used for environments of type **suitability**. If specified, the values of the input raster will be divided by this value. Used to scale raster layers from range $[0, N]$ to valid range $[0, 1]$.

12.4.7 value (optional)

Numeric value. Required (and only used) for environments of type **constant**. Instead of reading the values for the environment from a raster file, all cells of this environment will be set to the specified value.

12.5 dispersal (section)

The dispersal configuration section describes the parametrisation of all dispersal processes for a single species. Required are only dispersal kernels (a minimum of one), and its weights. Additionally, long-range and local dispersal can be specified, see .

12.5.1 kernel filenames (required, list)

String values, separated by commas. List of file names for the dispersal kernel rasters. The rasters have to match the simulation rasters in resolution, but not extent or projection. For the format, see [Kernel based dispersal](#) and [Dispersal kernels](#).

12.5.2 minimum weights (required, list)

Numeric values in range $[0, 1]$ or -1. A list of values that specify the minimum weight for each kernel (as listed in kernel filename). The list has to have the same number of entries as the number of kernel filenames. The special value of -1 means that this kernel should use the remaining weight. For description of the kernel weights see [Kernel based dispersal](#).

12.5.3 maximum weights (required, list)

As minimum weights. The maximum weight of each kernel have to be equal or greater than its minimum weight.

12.5.4 local dispersal proportion (optional)

Numeric value in range $[0, 1]$, default value 0. The fraction of seeds that should be directly dispersed into the source cell, before any other dispersal mechanism.

12.5.5 long range radius (optional)

Integer value ≥ 0 , default 0. If this value is greater than zero, long range dispersal will be enabled. This parameter specifies the maximum radius (in cells) that a seed can travel. See [Long-range dispersal](#).

12.5.6 long range target count (optional)

Integer value ≥ 0 , default 0. Required > 0 if long range dispersal is enabled. Specifies the number of target cells that should be randomly drawn and dispersed into. See [Long-range dispersal](#).

12.5.7 long range event probability (optional)

Numeric value in range $[0, 1]$. Required > 0 if long range dispersal is enabled. Specifies the probability, that a long range dispersal even should happen per cell that has produced seeds and year. [Long-range dispersal](#).

12.6 module (section)

12.6.1 filename (required)

12.7 glm (section)

Currently these GLM types are supported in CATS:

- `quadratic glm`
- `linear glm`

Currently these GLM families are supported in CATS:

- `binomial`
- `gaussian`
- `bernoulli`
- `poisson`

12.7.1 family (required)

String. One of the supported GLM families, see above and [GLM environments](#).

12.7.2 intercept (required)

Numeric value. Specifies the GLM intercept. See [GLM environments](#).

12.7.3 predictors (required, list)

List of strings, separated by a comma. Each string is required to be the user-defined name of an environment section of type `predictor`.

12.7.4 <predictor> linear (required)

Numeric value. The linear coefficient of the predictor.

12.7.5 type (required)

String. One of the supported GLM types, see above and [GLM environments](#).

12.7.6 <predictor> quadratic (optional)

Numeric value. Required for quadratic GLMs. The quadratic coefficient of the predictor.

12.7.7 save environment (optional)

Boolean value. See [save environment \(optional\)](#).

13 Command line arguments

CATS is a command line program and expects at least one command line argument: the configuration file, which contains all necessary information for a simulation:

```
./cats path/to/configuration.conf
```

By specifying a replicate number (an integer equal or greater than zero), it is possible to distinguish multiple simulations with the same configuration file. The replicate number will be part of all output file names. If no replicate number is specified, it will default to zero. To specify the replicate number, run CATS with the `-r` or `--replicate` command line option:

```
./cats -r 2 path/to/configuration.conf
```

```
./cats --replicate=2 path/to/configuration.conf
```

For a full list of command line options, you can run cats with the `--help` argument:

```
./cats --help
```

13.1 Configuration option precedence

Some configuration options can be both specified in the configuration file and on the command line. Command line options take precedence over configuration parameters specified in the configuration file, which take precedence over default values (which are used when neither are specified).

13.2 List of command line options

For a list of command line options, you can run CATS with the `--help` command line options

```
./cats --help
```

There is only one required command line argument: the path of the configuration file. All other arguments are optional. Some only work in conjunction with hybrid parametrisation mode, and others are mutually exclusive.

13.2.1 General command line options

- `--replicate=<replicate>` or `-r <replicate>`
Set the replicate number (Integer ≥ 0)
- `--output-directory=<path>` or `-o <path>`
Equivalent to the configuration parameter [output directory \(optional\)](#)
- `--log-level=<log level>` or `-l <log level>`
set the log level, one of the following strings: `debug`, `info`, `important`, `warning`, or `error`
- `--version` or `-v`
Show CATS version and quit.
- `--json`
Output additional information in JSON format to stderr, prefixed with `JSON::` (e.g. for use with a GUI)

13.2.2 Hybrid mode only options

- `--scale-factor=<scale factor>` or `-s <scale factor>`
Directly set the scale factor for testing purposes in range (0, 1)
- `--scale-only`
Run only the scale factor calculation, output information and quit without running the simulation.
- `--scale-test`
Mutually exclusive with `--scale-gradient`. Calculate the scale factor and run a simulation on 1x1 grid for 10000 years at a suitability of OT.
- `--scale-test-cc=<carrying capacity>`
Only in conjunction with `--scale-test` or `--scale-gradient`. Override the carrying capacity specified in the configuration file.
- `--scale-test-years=<years>`
Only in conjunction with `--scale-test` or `--scale-gradient`. Override the number of years of the test simulation. Default 10000.
- `--scale-gradient`
Mutually exclusive with `--scale-test`. Run a simulation on a virtual suitability grid with default simulation area, which is a horizontal suitability gradient.

13.2.3 Random number options

- `--poisson-dampening`
Which fraction [0.0, 1.0] of random quantities is rounded instead of drawn from a poisson distribution. Default 0. Equivalent to configuration parameter [poisson dampening factor \(optional\)](#).

- **--poisson-dampening-min**
When using dampening > 0, only round if the rounded quantity is at least this value. Default 1. Minimum 1. Equivalent to configuration parameter [poisson dampening minimum \(optional\)](#)
- **--poisson-maximum**
Maximum parameter when drawing from poisson distribution instead of rounding for performance reasons. Default 5000. Equivalent to configuration parameter [poisson maximum \(optional\)](#).
- **--poisson-maximum-draw-difference**
Draw the difference to the maximum from a poisson distribution and add to the rounded remainder, instead of rounding the entire quantity. Default: yes Equivalent to configuration parameter [poisson maximum draw difference \(optional\)](#).

13.2.4 Debugging options

- **--debug-lambda**
Show additional output for problems with the calculation of the population growth rate lambda.
- **--debug-config-unread**
Show entries in the configuration files that were not read and quit.
- **--debug-config-unspecified**
Show optional configuration parameters that were not specified and quit.
- **--debug-config-parsing**
Same as --debug-config-unread and --debug-config-unspecified
- **--debug-vital-rates**
Undocumented. Outputs several GB of debug data of vital rate values for all link types over the whole range of suitability and population density values. In development.

14 Running a CATS simulation

14.1 MPI simulations

CATS also comes as MPI (Message Passing Interface, <https://www.mpi-forum.org/>) version (although there is no pre-compiled Windows available). The MPI version runs a single simulation distributed across multiple compute nodes. This allows to e.g. use larger extents which would not fit in the memory of a single node, and use more CPU cores in parallel. The number of threads specified in the configuration file is the number of threads that will be used on each node. The executable file is cats-mpi, which is run in conjunction with mpirun. Otherwise, it is used exactly as the standard version of CATS, including using the same configuration file.

Note

The only feature that is currently not supported in the MPI version is the long range dispersal.

In the MPI version, the simulation extent is split into multiple, equally sized (or as close as equally sized as possible) horizontal strips. The number of strips is equal to number of nodes specified with `mpirun`, and each strip is handled on an individual node. CATS demographic module considers the individual grid cells in isolation, so only the dispersal module needs to cross node boundaries. Seeds that would be dispersed into strips in neighbouring strips will get copied to their respective nodes. This way a simulation should not differ between the MPI and standard version of CATS. Grid statistics and population grids are calculated and collected on the individual nodes, and sent to a single node where data is aggregated and saved to disk.

15 Configuration examples

For example configuration files, see the quick start guide [example data](#).

16 Modules

Modules are loadable libraries that extend and modify the behaviour of CATS. There is currently no documentation, as the implementation is still under development and will change.: