

Process-oriented models of autumn leaf phenology: ways to sound calibration and implications of uncertain projections

Michael Meier, Christof Bigler

<https://doi.org/10.5194/gmd-16-1-2023-supplement>

Correspondence to: Michael Meier (michael.meier@cefe.cnrs.fr)

S1 Optimisation algorithms

The two Bayesian algorithms that we evaluated are Efficient Global Optimisation algorithms based on kriging (Krige, 1951; Picheny and Ginsbourger, 2014). One of these algorithms (EGO) is purely Bayesian, whereas the other (TREGO) combines Bayesian optimisation with a deterministic derivative-free trust-region formation. In the former, the expected improvement criterion is maximised over the entire search space, whereas in the latter, the expected improvement criterion is maximised over the entire search space or over a smaller trust region.

The three non-Bayesian algorithms that we evaluated are the Generalised Simulated Annealing (GenSA), the Particle Swarm Optimisation (PSO), and the Covariance Matrix Adaptation with Evolutionary Strategies (CMA-ES) algorithm. The GenSA algorithm follows an iterative search for the global optimum, where points with better solutions become new starting points for the evaluation of the neighbours (Xiang et al., 1997; Xiang et al., 2013). Points with worse solutions may also be accepted as new starting points, but with a decreasing probability as iterations increase and differences to the best solution become larger. The neighbours to be sampled for the next iteration are determined with a Cauchy-Lorentz distribution. The PSO algorithm samples the search space at a constant number of points (named particles) at each iteration (Clerc, 2011, 2012; Marini and Walczak, 2015). For the initial iteration, a random position, search direction, and velocity is assigned to each particle, under the assumption of a uniform distribution. After the initial iteration as well as after unsuccessful iterations, each particle shares information regarding position, search direction, velocity, and personal best solution with a given number of neighbouring particles. Thus, every particle derives the next sampling point from its own position, search direction, velocity, and best solution as well as from its corresponding knowledge of its neighbours. The CMA-ES algorithm iteratively samples the search space, deriving the sampling points anew from a multivariate normal distribution at each iteration (Hansen, 2006, 2016). While the shape of the distribution ellipsoid is determined by a covariance matrix based on principal component analysis (adaptation), the standard deviation of the distribution (step length) depends on the history of the search (evolution). The mean of the ellipsoid represents the best solution in the current iteration.

All algorithms were set to find the global minimum of the root mean square error (RMSE) between the observed and modelled autumn phenology. Phenology models may return NA values if the accumulated senescence rate does not meet the corresponding threshold value before the end of the current year, while certain algorithms cannot handle iterations that lead to NA values. Therefore, we set the modelled day of year to one for NA values and hence calculated

the RMSE accordingly. All R functions for the applied algorithms include different settings and control parameters, depending on if they were executed in a normal or extended mode and on the number of free parameters in the corresponding phenology model (Table S1).

Table S1. The optimisation algorithms and the R-packages and R-functions with which they were applied.

Algorithm	R-package::R-function()	Settings and controls	normal	extended	Source
EGO	DiceOptim::easyEGO()	budget = n.init = inneroptim = "BFGS" TR = F nugget = 1e-5 algo = "EGO" multistart = 1	$10 \times n$ $5 \times n$	$30 \times n$ $10 \times n$	Pi21
TREGO	DiceOptim::easyEGO()	budget = n.init = inneroptim = "BFGS" TR = T nugget = 1e-5 algo = "TREGO" multistart = 1	$10 \times n$ $5 \times n$	$30 \times n$ $10 \times n$	Pi21
GenSA	phenor::pr_fit_parameters()	max.call =	$20000 \times n$	$80000 \times n$	Hu18
PSO	pso::psoptim()	maxit =	$500 \times n$	$5000 \times n$	Be12
CMA-ES	cmaes::cma_es()	maxit =	$100 \times n^2$	$1000 \times n^2$	Tr11

Note: For every function, the settings and controls in general as well as in dependence of the normal vs. extended mode and of the number of free model parameters (n) are listed, together with the source of the respective R-package (Pi21: Picheny et al. (2021); Hu18: Hufkens et al. (2018); Be12: Bendtsen (2012); Tr11: Trautmann et al. (2011)).

S2 Calibration and validation samples

We followed random, systematic, or stratified sampling procedures to select observations and sites for the various calibration and validation samples, which differed by calibration mode. The different sampling and selection procedures are explained and illustrated below.

S2.1 Site-specific calibration

We performed 5-fold cross-validation using random, systematic continuous, and systematic balanced sampling procedures. The systematic sampling procedures were based on year, mean annual temperature, or the timing of autumn phenology. All applied procedures are explained and illustrated in Table S2 and Figure S1, respectively.

Table S2. Sampling procedures and corresponding selection of observations, together with the respective abbreviations used in this study.

Sampling procedure	Selection of observations	Abbreviation	N:q min(N:7) – max(N:2)
Random	Random selection	Rdm.	2.3 – 26.0 (beech) 2.3 – 26.6 (oak) 2.3 – 12.0 (larch)
Systematic	Continuous selection according to year	Sys. (cts.; YR)	
	Continuous selection according to mean annual temperature	Sys. (cts.; MAT)	
	Continuous selection according to autumn phenology	Sys. (cts.; AP)	
	Balanced selection according to year	Sys. (bal.; YR)	
	Balanced selection according to mean annual temperature	Sys. (bal.; MAT)	
	Balanced selection according to autumn phenology	Sys. (bal.; AP)	

Note: Further, the range of the sample size proxy N:q (i.e. 80% of the number of observations per site divided by the number of free model parameters) is given per species. The smallest values for N:q were obtained with $q = 7$, while $q = 2$ for the largest values.

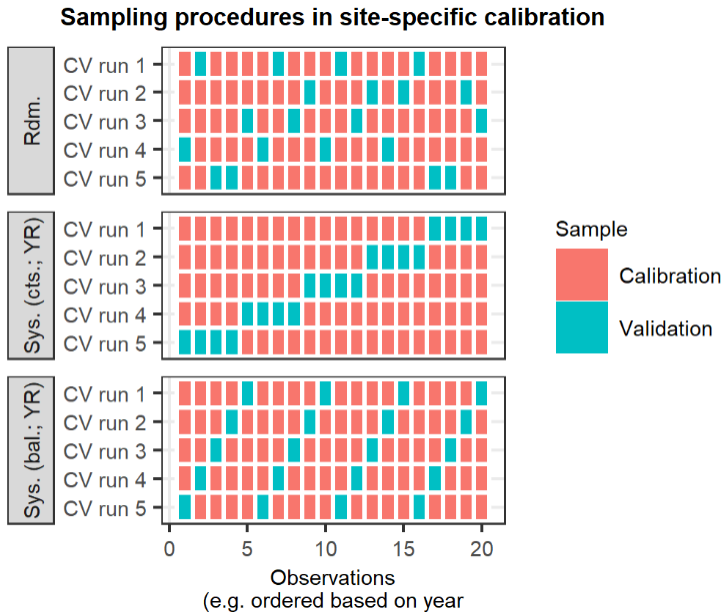


Figure S1. Illustration of random, systematic continuous, and systematic balanced sampling procedures. The allocation to the calibration and validation sample (red and turquoise, respectively) for each of the five cross-validation (CV) runs is shown as an example for a site with 20 observations. The abbreviations for the sampling procedures are explained in Table S2.

S2.2 Species-specific calibration

After allocating observations per site to calibration or validation, we formed different calibration samples with the sites from our 500-site populations. We validated the models with 25% randomly selected observations per site (rounded up to the nearest integer) within the calibration sample or within the whole population. The calibration samples consisted of the remaining 75 % of observations from 2 to 500 sites, selected using random, systematic, or stratified sampling procedures. These procedures were based on the annual mean temperature or the timing of autumn phenology. All procedures used are explained and illustrated in Table S3 and Figure S2, respectively. Note that in the special case of the calibration sample with the entire 500 sites, no sampling procedure was applied and the validations within the sample and within the population were identical to each other.

Table S3. Sampling procedures and corresponding selection of sites, together with the respective abbreviations used in this study.

Sampling procedure	Selection of sites	Abbreviation	$\bar{N}:q$ [range, ($\bar{N}:7$) – ($\bar{N}:2$)]	$s:S$ (no of sites)	Number of samples per combination and species
Full sample	Entire population	Full sample	3.6 – 12.7 (beech) 3.6 – 12.5 (oak) 2.4 – 8.5 (larch)	1.0 (s = 500)	1
Random	Random selection from entire population	Rdm. (entire p.)	3.0 – 17.5 (beech) 2.6 – 16.0 (oak) 2.2 – 9.0 (larch)	0.004 (s = 2) 0.01 (s = 5) 0.02 (s = 10) 0.04 (s = 20) 0.1 (s = 50) 0.2 (s = 100) 0.4 (s = 200)	5 5 5 5 5 5 5
	Random selection from lower 50% according to average mean annual temperature	Rdm. (lwr.; MAT)	2.1 – 17.8 (beech) 2.9 – 17.8 (oak) 2.3 – 9.0 (larch)		
	Random selection from lower 50% according to average autumn phenology	Rdm. (lwr.; AP)	2.6 – 17.2 (beech) 3.1 – 19.0 (oak) 2.1 – 9.0 (larch)		
	Random selection from upper 50% according to average autumn phenology	Rdm. (upr.; AP)	2.5 – 15.5 (beech) 3.2 – 15.0 (oak) 2.3 – 9.0 (larch)		
	Balanced selection according to average mean annual temperature	Sys. (bal.; MAT)	2.9 – 20.2 (beech) 2.8 – 15.2 (oak) 2.4 – 9.8 (larch)		
Systematic	Balanced selection according to average autumn phenology	Sys. (bal.; AP)	2.2 – 16.1 (beech) 2.4 – 15.2 (oak) 2.1 – 9.0 (larch)		
Stratified	Random selection from equal sized bins according to average mean annual temperature	Str. (MAT)	2.1 – 10.0 (beech) 6.4 – 25.1 (oak) 2.4 – 8.9 (larch)	0.024 (12)	5
	Random selection from equal sized bins according to average autumn phenology	Str. (AP)	4.4 – 17.0 (beech) 4.0 – 16.0 (oak) 2.4 – 8.6 (larch)	0.034 (17)	5

Note: Further, the ranges of the sample size proxies $\bar{N}:q$ and $s:S$ (i.e. the average number of observations per site divided by the number of free model parameters and the number of sites relative to the 500 sites of the entire population, respectively) are given per species. The smallest values for $\bar{N}:q$ were obtained with $q = 7$, while $q = 2$ for the largest values.

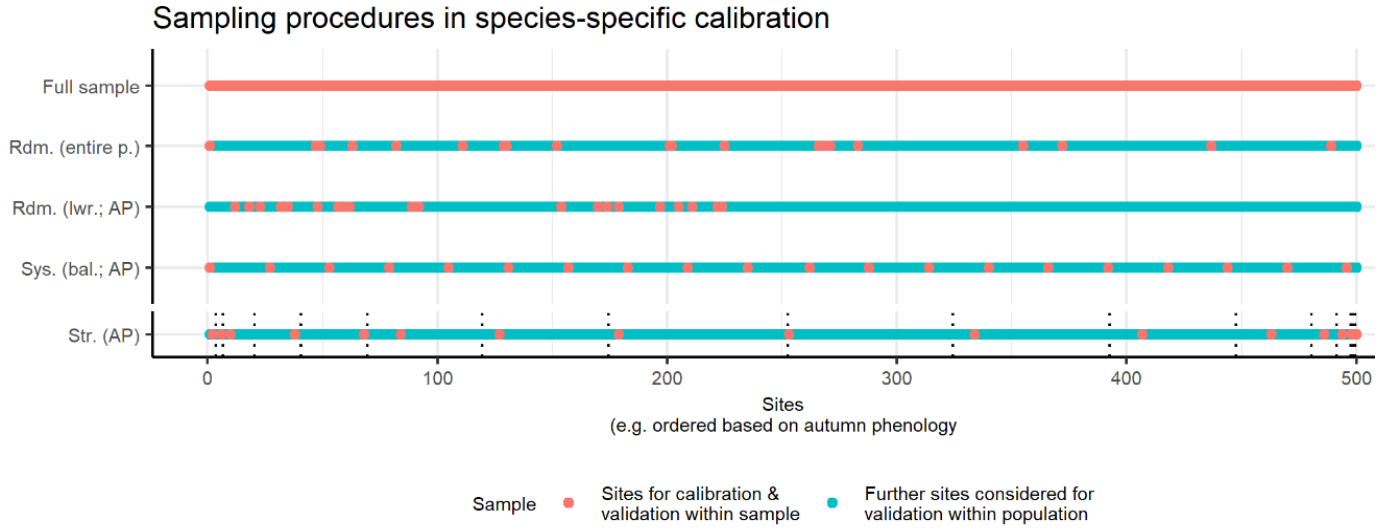


Figure S2. Illustration of the full sample and the calibration samples from random, systematic, and stratified sampling procedures. An example of an allocation of 20 or 17 sites to the calibration samples (red) is shown for the respective random and systematic or stratified sampling procedures. The remaining 480 or 483 sites in the population (turquoise) were included in the validation within the population. The 17 equally sized bins in the stratified sample have identical widths according to average autumn phenology (i.e. identical difference in days between lower and upper bound of bin) and are separated by dashed black lines. The abbreviations for the sampling procedures are explained in Table S3.

S3 Performance statistics

We quantified model performance according to the root mean square error (RMSE; Eq. S1):

$$RMSE = 1/n \sqrt{\sum_{i=1}^n (O_i - \hat{O}_i)^2}$$

(Eq. S1)

Here, O_i and \hat{O}_i are the respective observed and modelled day of year of the i^{th} element of a dataset of n data.

References

- Bendtsen, C.: pso: Particle Swarm Optimization. R package version 1.0.4 [Code], 2012.
- Clerc, M.: From Theory to Practice in Particle Swarm Optimization, in: Handbook of Swarm Intelligence: Concepts, Principles and Applications, edited by: Panigrahi, B. K., Shi, Y., and Lim, M.-H., Springer Berlin Heidelberg, Berlin, Heidelberg, 3-36, 10.1007/978-3-642-17390-5_1, 2011.
- Clerc, M.: Standard Particle Swarm Optimisation. hal-00764996. <https://hal.archives-ouvertes.fr/hal-00764996>, 2012.
- Hansen, N.: The CMA evolution strategy: A comparing review, in: Towards a new evolutionary computation: Advances in the estimation of distribution algorithms, edited by: Lozano, J. A., Larrañaga, P., Inza, I., and Bengoetxea, E., Springer Berlin Heidelberg, Berlin, Heidelberg, 75-102, 10.1007/3-540-32494-1_4, 2006.
- Hansen, N.: The CMA evolution strategy: A tutorial, arXiv preprint arXiv:1604.00772, 2016.

- Hufkens, K., Basler, D., Milliman, T., Melaas, E. K., and Richardson, A. D.: An integrated phenology modelling framework in R, *Methods Ecol. Evol.*, 9, 1276-1285, 10.1111/2041-210x.12970, 2018.
- Krige, D.: A statistical approach to some mine valuations and allied problems at the Witwatersrand, Unpublished Master's Thesis, University of the Witwatersrand, Witwatersrand, South Africa, 1951.
- Marini, F. and Walczak, B.: Particle swarm optimization (PSO). A tutorial, *Chemometrics and Intelligent Laboratory Systems*, 149, 153-165, <https://doi.org/10.1016/j.chemolab.2015.08.020>, 2015.
- Picheny, V. and Ginsbourger, D.: Noisy kriging-based optimization methods: A unified implementation within the DiceOptim package, *Computational Statistics & Data Analysis*, 71, 1035-1053, 10.1016/j.csda.2013.03.018, 2014.
- Picheny, V., Ginsbourger Green, D., and Roustant, O.: DiceOptim: Kriging-Based Optimization for Computer Experiments. R package version 2.1.1 [Code]. <https://CRAN.R-project.org/package=DiceOptim>, 2021.
- Trautmann, H., Mersmann, O., and Arnu, D.: cmaes: Covariance Matrix Adapting Evolutionary Strategy. R package version 1.0-12 [Code], 2011.
- Xiang, Y., Sun, D. Y., Fan, W., and Gong, X. G.: Generalized simulated annealing algorithm and its application to the Thomson model, *Phys. Lett. A*, 233, 216-220, 10.1016/s0375-9601(97)00474-x, 1997.
- Xiang, Y., Gubian, S., Martin, F., Suomela, B., and Hoeng, J.: Generalized Simulated Annealing for Global Optimization: The GenSA Package, *R Journal*, 5, 13-28, 10.32614/RJ-2013-002, 2013.