Noname manuscript No.
(will be inserted by the editor)

Warping

Victor Picheny · Coralie Picard · Gael Thebaud

Received: date / Accepted: date

[Coralie's comment: les figures ne sont pas toutes de la meme taille, mais je n'ai pas passe trop de temps dessus avant qu'on se mette d'accord sur celles qui seraient dans l'article ou pas]

Abstract On peut [Victor's comment: faire un commentaire] [Coralie's comment: chacun avec sa couleur], on peut aussi ~~enlever des trucs~~ ou bien ajouter d'autres trucs, et Gael aussi.

Keywords to do

1 Introduction

[Victor's comment: toi ou moi] **Mathematical modelling for design, in particular in epidemiology**

Mathematical models are increasingly used in many research fields to understand and optimize a process. More particularly, they are useful in epidemiology to predict epidemics and to propose efficient control options. [Coralie's comment: (Cunniffe, Koskella, et al. 2015 Cunniffe et al. 2016 Mushayabasa and Tapedzesa 2015 Tildesley et al. 2006 Bajardi et al. 2012 Kompas et al. 2017 VanderWaal et al. 2017 Grechi et al. 2012).] However, these epidemiological studies are moslty focused on improving one control option which generally depends on only one or two parameters in their model, although various control actions are usually applied simultaneously to

V. Picheny
MIAT, Université de Toulouse, INRA, Castanet-Tolosan, France
Tel.: +33561285551
E-mail: victor.picheny@inra.fr

C. Picard
BGPI, Montpellier SupAgro, INRA, Univ. Montpellier, Cirad, TA A-54/K, 34398, Montpellier

G. Thebaud
to do

manage an epidemic. All these actions could be jointly optimized but taking into account numerous management parameters in an optimization problem can be difficult, especially when the management efficiency depends on the interaction between these parameters.

[Victor's comment: pour toi... mais peut-Łtre plus facile faire une fois que le reste aura avanc]

The sharka model and objectives

In this study, we analyse a simulation model of sharka disease spread and management. This disease, caused by a virus transmitted by aphids through *Prunus* orchard, is one of the most damaging diseases of stone fruit trees belonging to the genus Prunus (e.g. peach, apricot and plum) (Cambra et al. 2006; Rimbaud et al. 2015). Our model includes epidemiological parameters which vary between simulations, and various landscapes on which the virus can spread, which means that this model is stochastic. In addition, management parameters allow to simulate orchard surveillance. Here, we aim to optimize these management parameters using a efficient optimization algorithm.

[Victor's comment: le reste de l'intro pour moi]

Generalities on optimization

Bayesian optimization

Problem at hand: dealing with local invariances

Outline

2 Model description and problem set-up

[Victor's comment: Section remplir par toi ! Suggestion de plan dtaill.]

What does it model

How the model works

What problem do we want to solve

Inputs description

Invariances descriptions

Table of inputs with range of variation

Table of invariance relations

The simulation model that we analyze in this work is a stochastic, spatially explicit, SEIR (susceptible-exposed-infectious-removed) model that simulates sharka spread and management actions (including surveillance, removals and replantations, Pleydell et al. 2018; Rimbaud et al. 2018a, 2018b). This model is orchard-based, with a discrete time step of one week. It allows to perform simulations on landscapes composed of uncultivated areas and patches on which peach trees are grown. The patches can be more or less aggregated in the landscape however, we only use in this work the 30 landscapes with a high level of patch aggregation as described by Picard et al. 2018. During the simulation, the trees in the patches are characterized by different states. When the simulation begins, they are not infected: they are in the "susceptible" state. Then, the virus is introduced the first year of the simulation in one of the patches and spreads through orchards (new introductions can also occur during the entire simulation on all patches). The virus causes changes in tree status: from ”susceptible”, they become ”exposed” (infected but not yet infectious or symptomatic), ”infectious hidden” (after the end of the latent period), ”infectious detected” (when specific symptoms are detected on the tree during a survey), and ”removed” (when the tree is removed from the patch). The model output is an economic criterion, the net present value (NPV), which accounts for the benefit generated by the cultivation of productive trees and the costs induced by fruit production and disease management (Rimbaud et al. 2018b).

In order to simulate wide range of epidemic and management scenarios, the model includes 6 epidemiological and 23 management parameters (Rimbaud et al. 2018b, Picard et al. 2018). In this work, we will use the 6 epidemiological parameters and only 10 management parameters (related to the surveillance of the orchards). They include distances of 3 zones for which the surveys are more or less frequent as well as their duration, the probability of the infected tree detection, and a contamination threshold which can request to increase the surveillance frequency in the focal zone. Details of epidemiological and management parameters used in this study are presented in Fig.1 and Table 1 (this table also includes the variation ranges of the parameters in the model).

Here, we aim to optimize the management strategy of the disease (i.e. to find the combination of management parameters allowing to obtain the best NPV), taking into account the epidemic stochasticity. However, we note that some combinations of management parameters can represent the same management, which may cause problems in the optimization process. Indeed, we observe that some management parameters are not useful when other parameters have a value of 0, which means that they can take any values without modifying the simulation. For example, when a zone radius is 0, the associated surveillance frequency have no impact on the NPV (regardless its value). The methodological developments that are proposed in this work address this issue by removing the parameter combinations which lead to the same management. The parameter invariances removed from the model are listed in Table 2.

Table 1 Epidemiological and management parameters implemented in the previously developed model with minimum and maximum values corresponding to the variation range of each parameter.

		Min	Max
Epidemiological parameters			
q_K	Quantile of the connectivity of the patch of first introduction	0	1
ϕ	Probability of introduction at plantation (before management)	0,02	0,02
	Probability of introduction at plantation (during management)	0,0046	0,0107
p_{MI}	Relative probability of massive introduction (before management)	0,4	0,4
	Relative probability of massive introduction (during management)	0	0,1
W_{exp}	Expected value of the dispersal weighting variable	0,469	0,504
β	Transmission coefficient	1,25	1,39
θ_{exp}	Expected duration of the latent period duration (years)	1,71	2,14
Management parameters			
ρ	Probability of detection of a symptomatic tree	0	0,66
γ_O	Duration of observation zones (years)	0	10
ζ_s	Radius-distance of security zones (m)	0	5800
ζ_f	Radius-distance of focal zones (m)	0	1
ζ_{eO}	Radius-distance of observation epicenter (m)	0	1
$1/\eta_0$	Maximal period between 2 observations (year)	1	15
η_s	Observation frequency in security zones (year-1)	0	8
η_f	Observation frequency in focal zones (year-1)	0	8
η_{f*}	Modified observation frequency in focal zones (year-1)	0	8
χ_o	Contamination threshold in the observation epicenter, above which the observation frequency in focal zone is modified	0	1

Table 2 Invariances of management parameters. For instance, when γ_O = 0 or when ρ = 0, χ_o does not influence the model output.

Warping	Management parameters	OR	OR	OR
No warping	ρ			
	$1/\eta_0$			
	γ_O			
Warping based on warped variables	χ_o	$\gamma_O = 0$	$\rho = 0$	
	ζ_{eO}	$\gamma_O = 0$	$\zeta_s = 0$	$\rho = 0$
	ζ_f	$\gamma_O = 0$	$\zeta_s = 0$	
Circular conditions	η_{f*}	$\gamma_O = 0$	$\rho = 0$	
	ζ_s	$\gamma_O = 0$	$\eta_s = 0$	
	η_s	$\gamma_O = 0$		
	η_f	$\gamma_O = 0$		

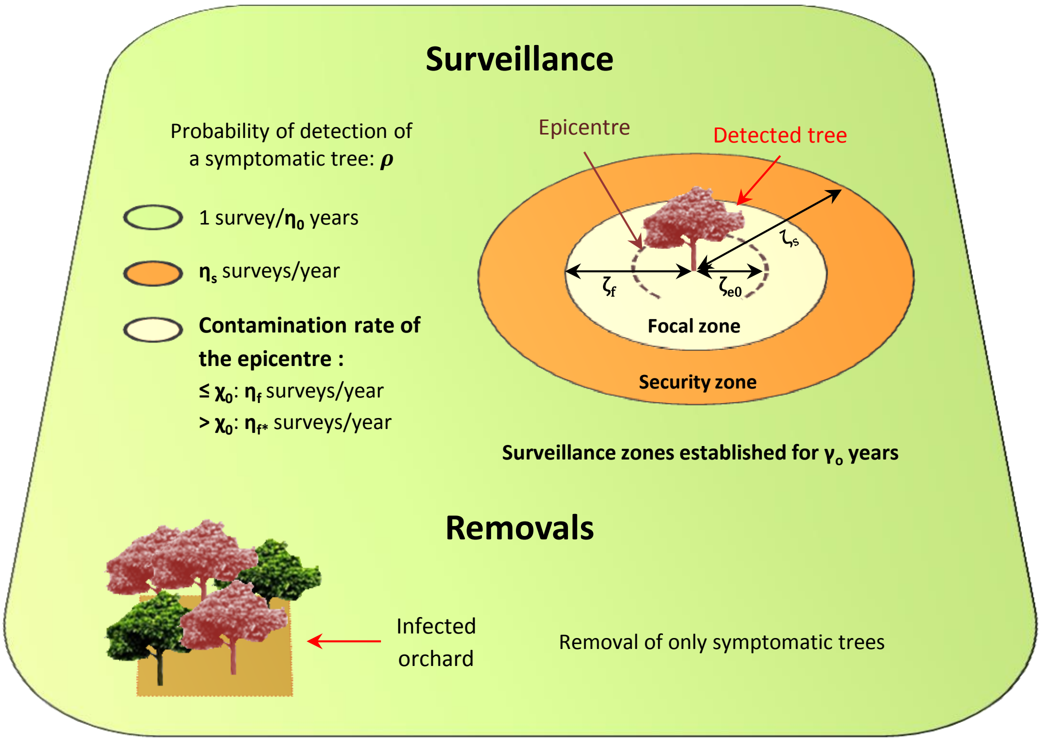

Fig. 1 Management actions implemented in the model

3 Methods Bayesian optimization

3.1 Overview

3.2 Bayesian optimization of stochastic simulators

3.3 Bayesian optimization with invariances

3.3.1 Definitions

3.3.2 Simple warping

3.3.3 Warping based on linear relations

3.3.4 Combining warpings

4 Experiments on toy problems

4.1 Problem descriptions

4.2 Comparison metrics

4.3 Results

5 A warping-based Bayesian optimization of the Sharka model

5.1 Numerical setup

5.1.1 Experiments description

[Victor's comment: Premier jet par toi ?] [Coralie's comment: j'ai l'impression que les parties krigeage et warping devraient se trouer dans cette partie et non pas dans le 5.1.2]

To evaluate the benefits of including the warping step in the optimization process (i.e. reducing the parameter space removing the combinations which lead to the same management), we conducted 50 independent optimizations of sharka management parameters with and without the warping step. The criterion to optimize was the mean of the NPV ($\overline{NPV}$). For this to happen, we randomly selected 50 times 200 management strategies using a maximin Latin hypercube sampling design (Fang, Li, and Sudjianto 2005). Then, for each sampling design of 200 strategies, we performed 2 optimizations in parallel: with and without the warping step. For one optimization, we performed sequentially 200 iterations allowing to choose 200 new strategies, resulting in a total of 400 evaluated strategies.These 200 new strategies were selected each time among 100,000 randomly generated candidate points over the parameter space and 10,000 more locally around the best point found. In addition, for each evaluated strategy, 1000 simulations were carried out (with different random seed) to take into account the variability due to the epidemic and landscape characteristics.

5.1.2 Comparison with standard BO

Description of comparison metrics

[Victor's comment: Idem juste pour les mthodes de comparaison, je me charge du paragraphe pour dire quoi on se compare et je m'occupe de la partie krigeage et warping.]

[Coralie's comment: on compare ici les resultats obtenus sans le probleme d identification : avec ton script denoise.results.v6.R. Mais je ne sais pas trop comment l expliquer ici]

We firstly compared the optimization results by subtracting the $\overline{NPV}$ achieved using the optimization with the warping step and the optimization without the warping step (obtained from the same sampling design).

In addition, we compared the optimization speed between the optimizations with or without warping. To this end, we used two different ways. Firstly, we performed a nonlinear regression of $\overline{NPV}$ obtained for all the selected strategies during the optimization process with and without the warping step, and we compared the growth parameter c of the following regression: $a + b \times exp^{-c \times x}$. Secondly, we used a specific algorithm developed by [Coralie's comment: reference???]. Briefly, we uniformly defined 100 α values between a minimum and a maximum values. Then, for each iteration performed in the optimization process (i.e. for each of the 200 evaluated strategies), we add: the number of optimizations (under 50) which exceed α 1, the number of optimizations which exceed α 2, ..., the number of optimizations which exceed α 100. We used $\alpha \in$ [0;18,012.12], and then $\alpha \in$ [10,000;18,012.12]. The value 18,012.12 corresponds to the maximal value of $\overline{NPV}$ identified in all the optimizations.

5.2 Results and insights into the Sharka model

We firstly subtracting the $\overline{NPV}$ obtained with optimizations with and without the warping step. In 24 out of the 50 optimization cases, we obtained better $\overline{NPV}$ with the warping step than without (Fig.2). This result means that with 200 iterations in the optimization, the final optimization result is not impacted by the use of a warping step.

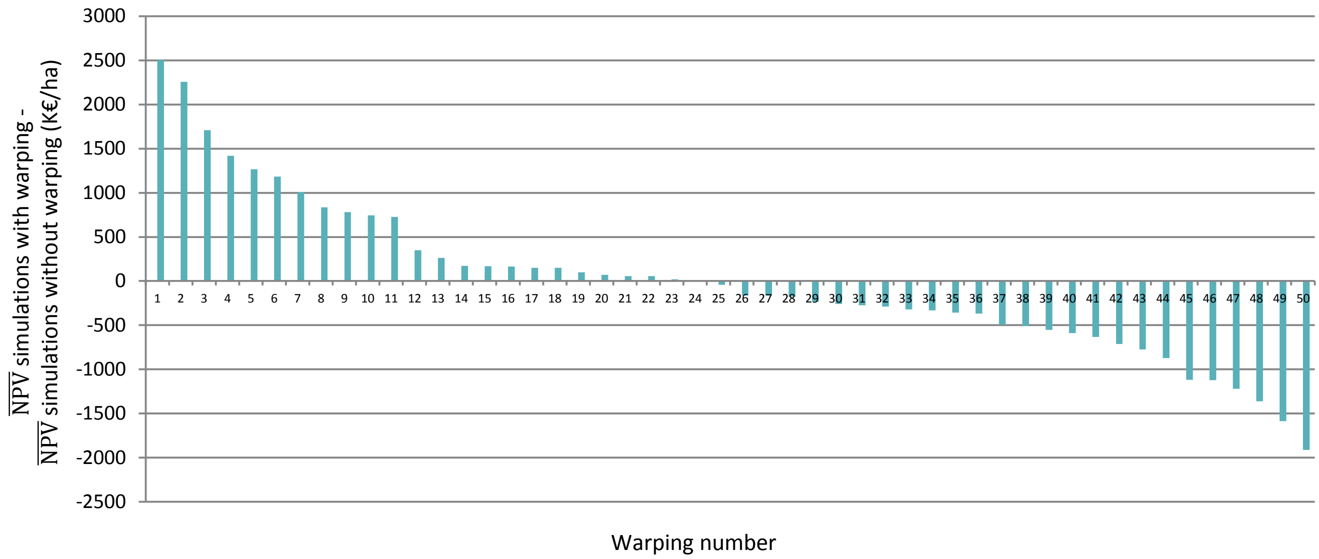

Fig. 2 Comparison of $\overline{NPV}$ obtained at the end of the optimization with and without warping.

However, we showed that the warping can impact the optimization speed (Fig.3). Indeed, the parameter c corresponding to the growth parameter of a nonlinear regression was higher with (0.26) than without (0.18) warping (Fig.4). In addition, we can visually observe that the warping step allow to improve the optimization speed on the Fig.5 and 6 which present the results of the algorithm developped by [Coralie's comment: reference???].

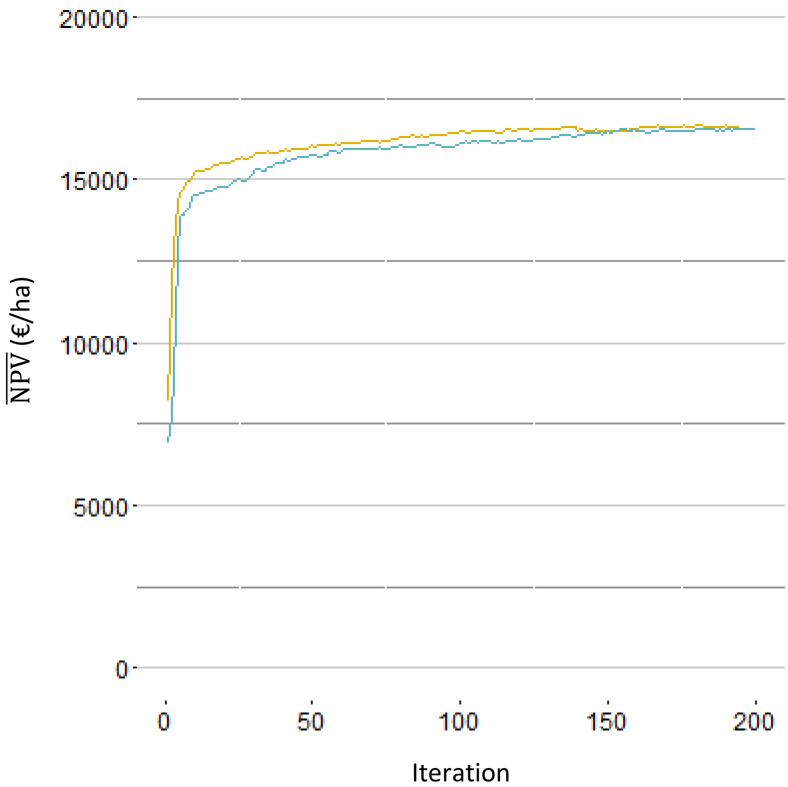

Fig. 3 Comparison of $\overline{NPV}$ obtained during optimizations with and without warping. Yellow and blue lines represent the mean of the $\overline{NPV}$ selected at each iteration for the 50 optimizations respectively perfomed with and without the warping step.

6 Conclusion

What we did (the problem we solved)

What we proposed: warping to tackle invariances. Proof of concept

Possible extensions

References

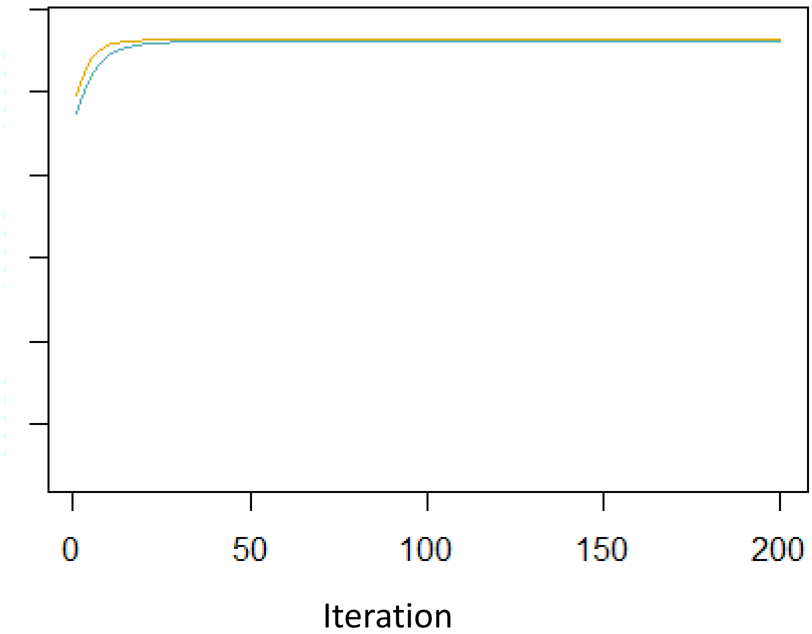

Fig. 4 Non linear regression on $\overline{NPV}$ obtained at each iteration of the optimizations with (yellow) and without (blue) warping.

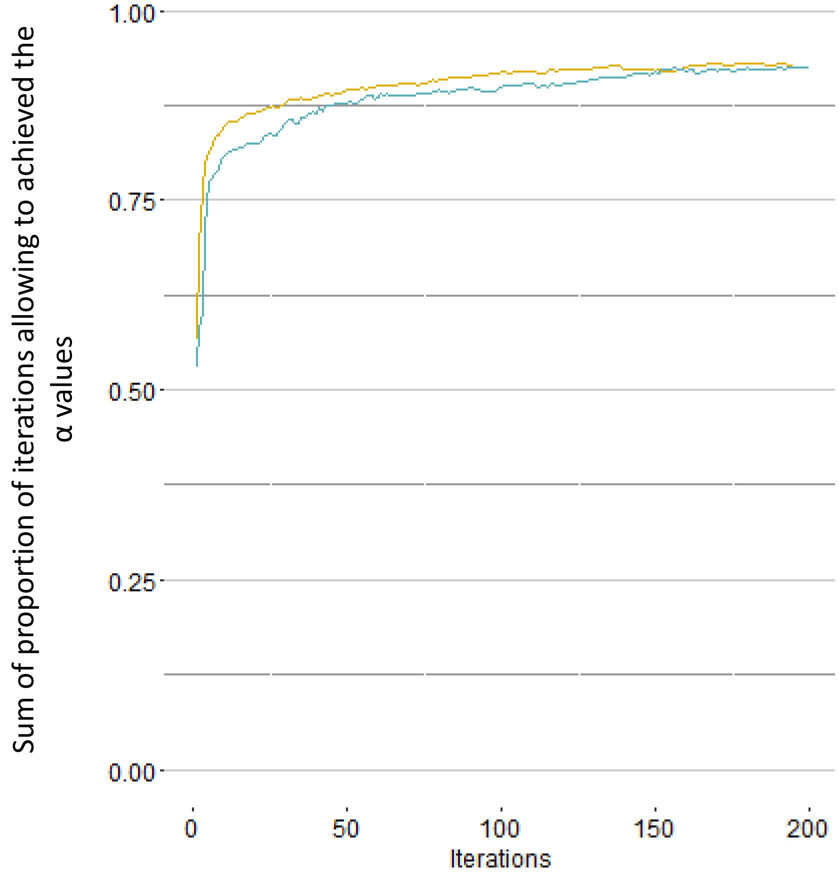

Fig. 5 Results of the Coco algorithm with (yellow) and without (blue) warping ($\alpha \in [0;18{,}012.12]$).

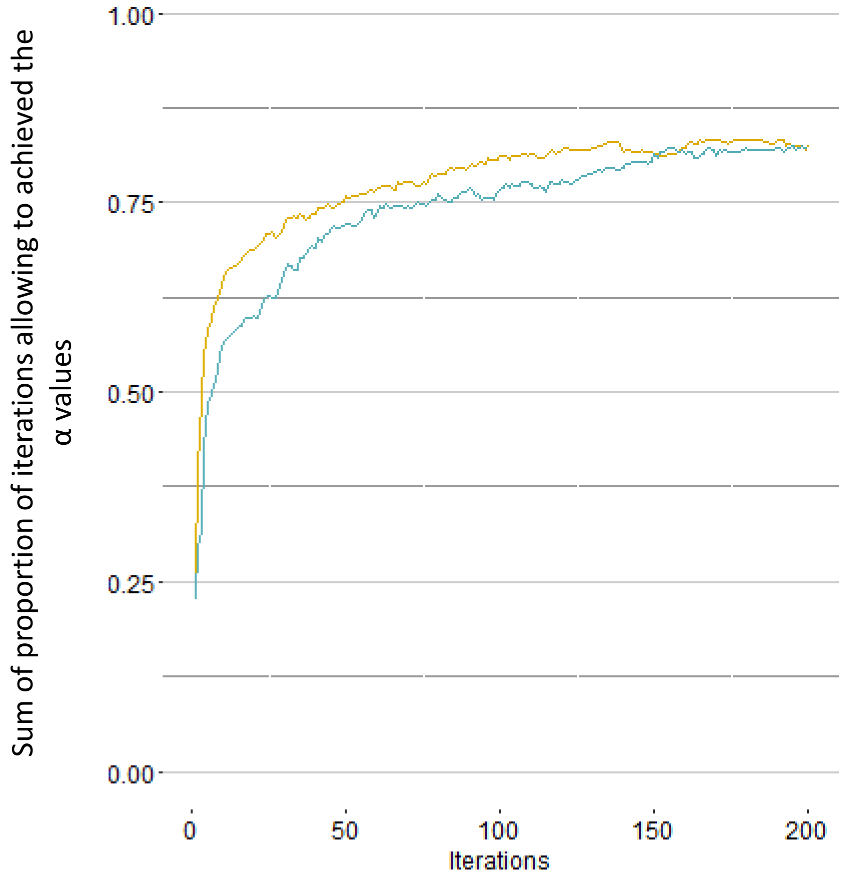

Fig. 6 Results of the Coco algorithm with (yellow) and without (blue) warping ($\alpha \in [10{,}000;18{,}012.12]$).