Sunflower phenotype optimization under climatic uncertainties using crop models

Victor Picheny, Ronan Trépos , Bastien Poublan (INRA MIAT),
Pierre Casadebaig (INRA AGIR)

MOABI 2015/11/27

Institut National de la Recherche Agronomique

Model assisted crop improvement and ideotype design

Ideotype : combination of morphological and/or physiological traits optimizing crop performances to a particular biophysical environment and crop management (see Martre et al. 2015 for a review).

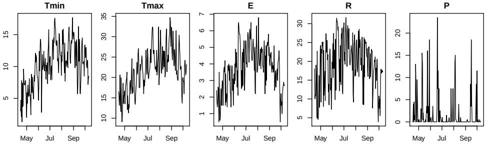
In silico approaches : finding an ideotype using ecophysiological dynamic models to predict crop performances :

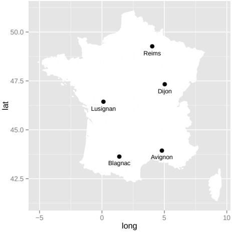
- eg. Semenov et al. 2014, Grechi et al. 2012, Wu et al. 2012, QuilotTurion et al. 2012, Qi et al. 2010, Letort et al. 2008.

SUNFLO : ecophysiological model of sunflower crop

$$\begin{aligned} y : \mathbb{X} \times \Omega &\rightarrow \mathbb{R}^+ \\ \mathbf{x}, c &\longmapsto y(\mathbf{x}, c). \end{aligned}$$

- $y(\mathbf{x}, c)$: annual yield on a sunflower crop (tons per hectare)
- $\mathbb{X} \subset \mathbb{R}^8$: space of phenotype traits combination
- Ω : space of climatic series ($|\Omega| = 190$)



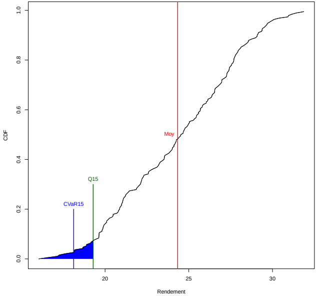


Robust optimization of yield : problem formulation

Multiobjective optimization

$$\begin{cases} \max & \mathbb{E}\left[Y(\mathbf{x})\right] \\ \max & \mathrm{CVaR}_\alpha\left[Y(\mathbf{x})\right] \\ \text{s.t.} & \mathbf{x} \in \mathbb{X}. \end{cases}$$

- $Y(\mathbf{x})$: random variable of $y(\mathbf{x}, c)$
- $\mathrm{CVaR}_\alpha\left[Y(\mathbf{x})\right] =$ $\mathbb{E}\left[Y(\mathbf{x}) | Y(\mathbf{x}) \leq Q_\alpha\left[Y(\mathbf{x})\right]\right]$



Solving the problem

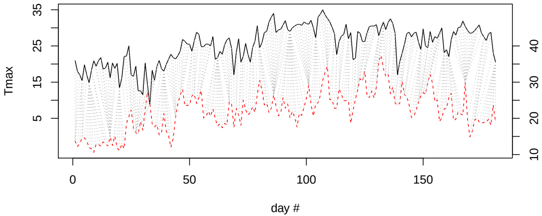
- 'black box' optimization : complex dynamic system
- population based algorithms : state of the art for multiobjective (nsga2, mopso-cd), parallelization of simulations

→ pb : to reduce the number of simulations for evaluation.

Method Two-step MOPSO-CD

1. Build a representative set Ω_K of Ω ($K = 10$)

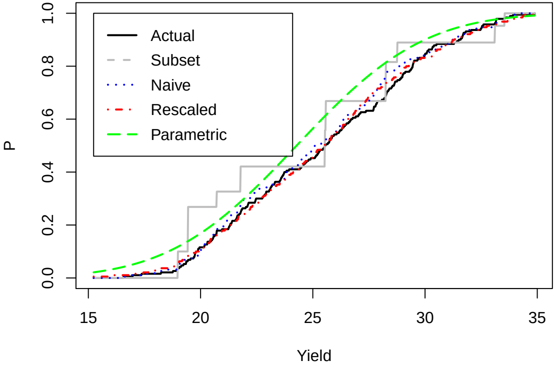
   Use of a K-medoids method with a distance which is a combination of time series distance and few model outputs.



Method Two-step MOPSO-CD

1. Build a representative set Ω_K of Ω ($K = 10$)
2. Use a mixture model : $f_{Y(\mathbf{x})}(y) = \sum_{k=1}^{K} \frac{N^k}{N} f_{Y^k(\mathbf{x})}(y)$

Decomposition of the component $Y^k(\mathbf{x}) = y(\mathbf{x}, \omega^k) + \varepsilon^k(\mathbf{x})$.



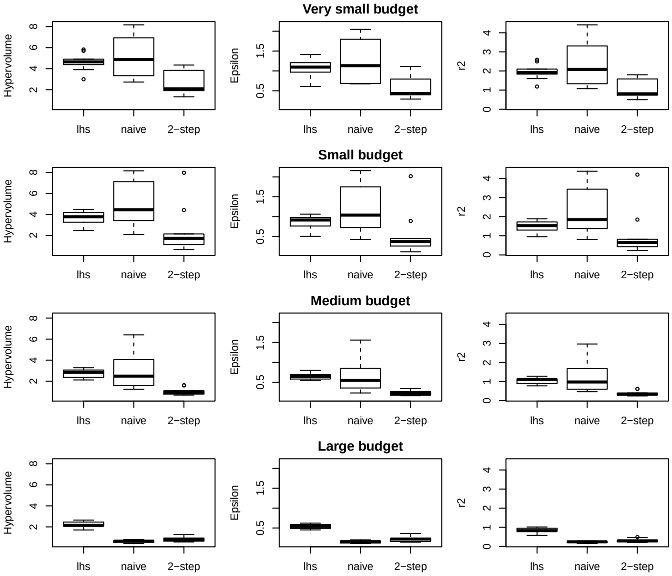
Method Two-step MOPSO-CD

1. Build a representative set Ω_K of Ω ($K = 10$)
2. Use a mixture model : $f_{Y(\mathbf{x})}(y) = \sum_{k=1}^{K} \frac{N^k}{N} f_{Y^k(\mathbf{x})}(y)$
3. Use a population based optimization algorithm (mopso-cd)

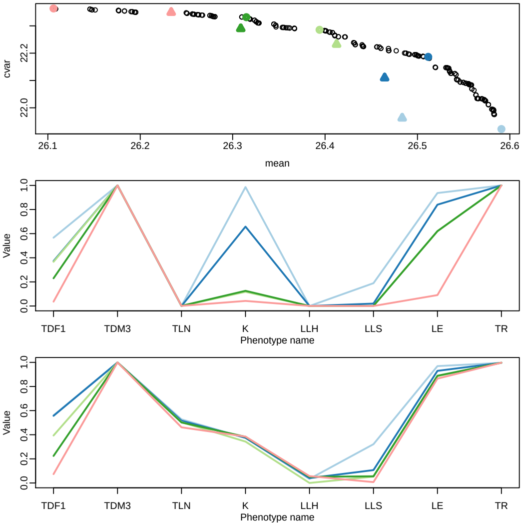
Experiments performed

Optimization experiment	Budget	Nb of iterations	Pop size	Real nb of simulations
Intensive	Very large	$300(\times 2)$	200	$\simeq 2 \times 10^7$
Random (or LHS)	very small	-	60	11,400
	small	-	125	23,750
	medium	-	500	95,000
	large	-	2,000	380,000
Naive MOPSO-CD	very small	12	5	12,350
	small	25	5	24,700
	medium	50	10	96,900
	large	100	20	383,000
Two-step MOPSO-CD	very small	$42(\times 2)$	9	11,540
	small	$71(\times 2)$	14	23,960
	medium	$152(\times 2)$	30	95,600
	large	$308(\times 2)$	61	380,780

Performance results (hypervolume, epsilon, r2)



Results in input space



'red' group :

- early flowering (TDF1)
- low light extinction efficiency (K)
- low plant leaf area (LLS)

→ conservative resource management strategy

Thank you !

- preprint : http ://arxiv.org/abs/1509.05697
- this work has been implemented on the RECORD platform
- computing resources have been provided by the genotoul bioinformatics platform Toulouse Midi-Pyrenees