

Optimisation de variétés de tournesol sous incertitude climatique

Ronan Trépos (master 2 de Bastien Poublan)

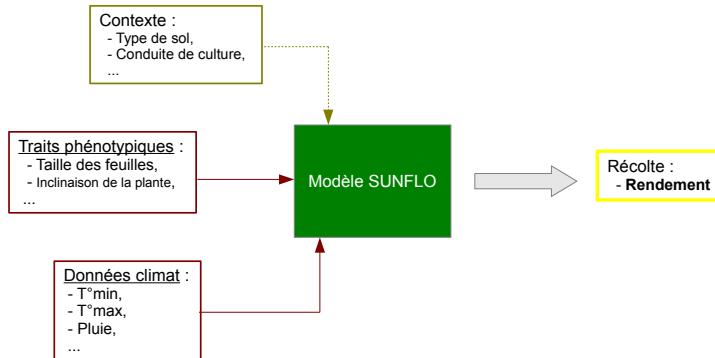
31 octobre 2014

Collaborations : B. Poublan, V.Picheny et R.Trépos



- 1 Présentation du stage
- 2 Méthodes développées
 - Sélection d'un sous-ensemble de climat
 - Reconstruction des distributions de rendements
- 3 Résultats obtenus
- 4 Conclusion

- Modèle de simulation dynamique :



- Objectif : trouver le phénotype permettant d'obtenir les meilleurs rendements

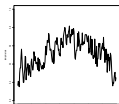
Tournesol

- Différentes caractéristiques :
 - 8 traits phénotypiques
 - Minimum et maximum pour chaque trait
 - Variations continues

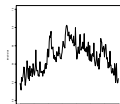
Gestion de l'incertitude

- Séries temporelles pour la description du climat :
 - Données journalières
 - Données de 1975 à 2012 sur 5 emplacements en France
⇒ 190 climats différents

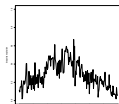
Température minimale



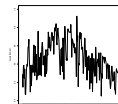
Température maximale



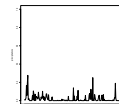
Évapotranspiration



Ensoleillement



Précipitations



- Prise en compte de l'incertitude climatique \Rightarrow Temps de calcul important
- Comment réduire le nombre de simulations ?
- Quels critères statistiques de performance ?
 - ◇ Climat variable aléatoire \rightarrow Rendement variable aléatoire

Objectifs du stage

- \Rightarrow Formalisation du problème d'optimisation
- \Rightarrow Solution algorithmique

Optimisation : quelle formulation ?

Critères statistiques

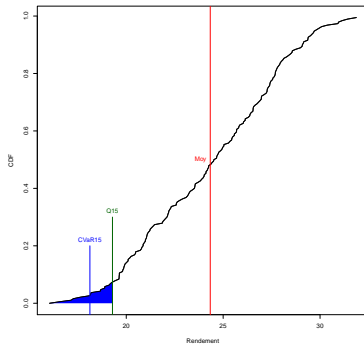
- Moyenne, variance, quantile ...
- Utilité espérée, "CVaR" ...

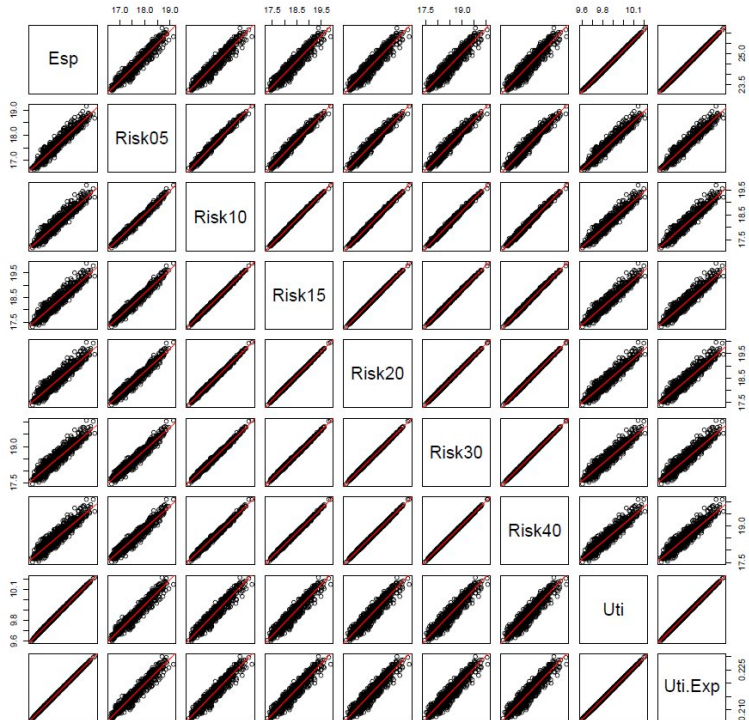
Données

- 1 phénotype : 190 rendements

Étude statique : plan d'expérience à 1000 points

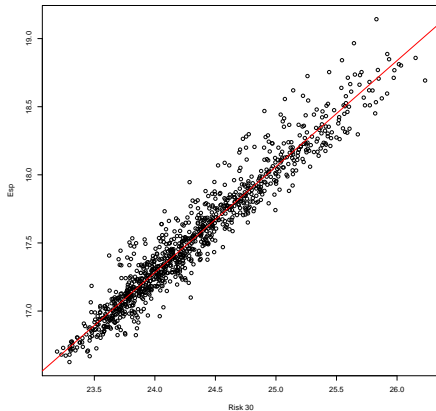
- 1000 phénotypes avec 190 climats
⇒ Indicateurs de performance





Formulation retenue

$$\text{Maximiser : } \begin{cases} E[R] \\ \text{CVaR}_{30\%}(R) \end{cases}, R \in \mathbb{R}^{190}$$

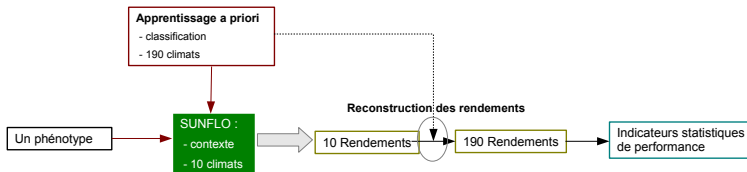


Différentes approches

- Approche par échantillonnage exhaustif :



- Approche à l'aide d'un apprentissage à priori :



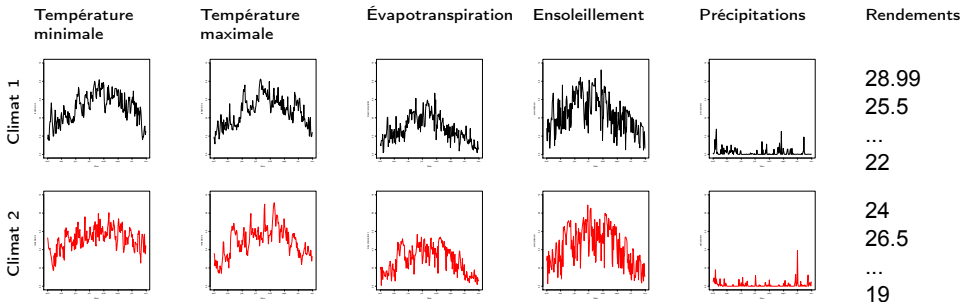
- Séparation apprentissage a priori et optimisation
- Adaptable à un autre simulateur
- Approche sur la classification
 - Séries climatiques différentes
 - Rendements différents
- Limites :
 - Résolution d'un problème approché
 - Pas de correction *en ligne* sur l'apprentissage

Sélection d'un sous-ensemble de climat

Recherche de 10 climats différents donnant 10 rendements différents

Étapes de la classification

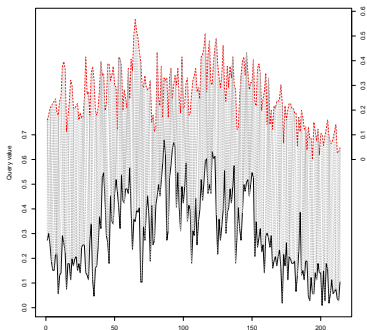
- Calcul de distance entre les séries temporelles
- Agrégation des distances (Normalisation + somme pondérée)
- Algorithme de classification non supervisée (*K-means*, *PAM*, *classification hiérarchique*)
⇒ Sous ensemble de climats



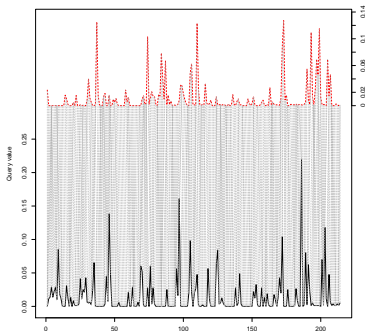
Distance entre deux séries (même variable)

- Distance DTW (Dynamic Time Warping)
- Prise en compte de “décalage” entre deux séries
- Température minimale, maximale, ensoleillement, évapotranspiration : ± 7 jours
- Précipitation : ± 3 jours

Évapotranspiration



Précipitations



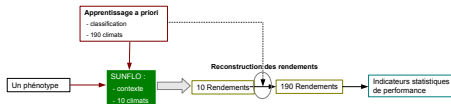
Distances

- Distance DTW inter-climats
- Distance euclidienne inter-rendements
- Normalisation des distances

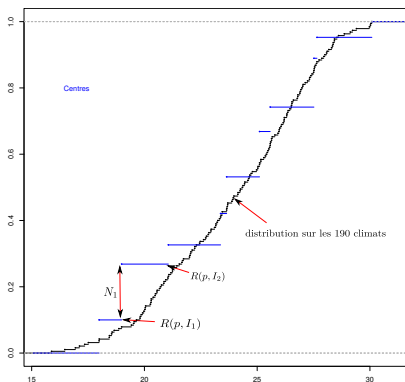
Utilisation dans la classification de la distance suivante :

$$\begin{aligned} \text{Distance}_{\text{Totale}} &= \frac{1}{2}(\alpha_1 \bar{D}_{Tmin} + \alpha_2 \bar{D}_{Tmax} + \alpha_3 \bar{D}_{Evapo} \\ &+ \alpha_4 \bar{D}_{Ensol} + \alpha_5 \bar{D}_{Pluie}) \\ &+ \frac{1}{2}(\mu_1 \bar{D}_{Pheno1} + \mu_2 \bar{D}_{Pheno2} + \dots + \mu_{10} \bar{D}_{Pheno10}) \end{aligned}$$

Reconstruction de la distribution (10 rendements \Rightarrow 190 rendements)



• Approches non paramétriques



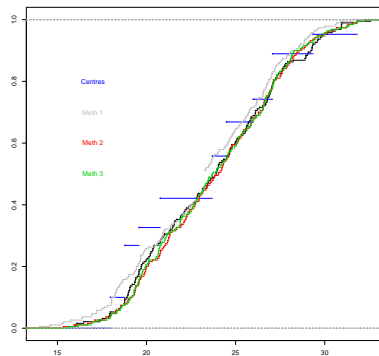
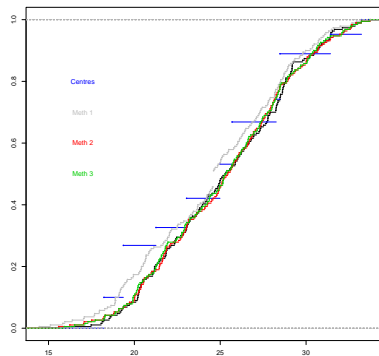
Pour un nouveau phénotype p et à partir des rendements obtenus sur les 10 climats représentatifs des classes I_1, \dots, I_{10} : $R(p, I_1), \dots, R(p, I_{10})$, reconstruire une estimation des rendements sur les 190 climats : $\hat{R}(p, c_1), \dots, \hat{R}(p, c_{190})$.

Reconstruction des 190 rendements (méthode 3)

- Soit $\tilde{R} = \{\overbrace{R(p, l_1), \dots, R(p, l_1)}^{N_1}, \overbrace{R(p, l_2), \dots, R(p, l_2), \dots, R(p, l_{10})}^{N_2}\}$
- Soit une base B de phenotypes évalués sur les 190 climats :
 $B \in \mathbb{R}^{\|B\|, N}$
- on pose : $\tilde{B}_i \in \{\overbrace{B(i, l_1), \dots, B(i, l_1)}^{N_1}, \overbrace{B(i, l_2), \dots, B(i, l_2), \dots, B(i, l_{10})}^{N_2}\}$
- calcul de résidus : $\epsilon_{i,j} = B_{i,j} - \tilde{B}_{i,j}$
- alors $\hat{R}_j = \tilde{R}_j + \zeta_j$
- où ζ_j est une perturbation du rendement pour le climat j dont le signe est donné par les $\epsilon_{i,j}$ et l'ampleur est fonction de la variance intra classe du climat j , la variance de \tilde{R} et la variance de \tilde{B} .

remarque : comme base B , on peut prendre les 10 pheno utilisés pour la classification pour lesquels on a simulé le rendement sur les 190 climats.

Rendements obtenus pour deux phénotypes :

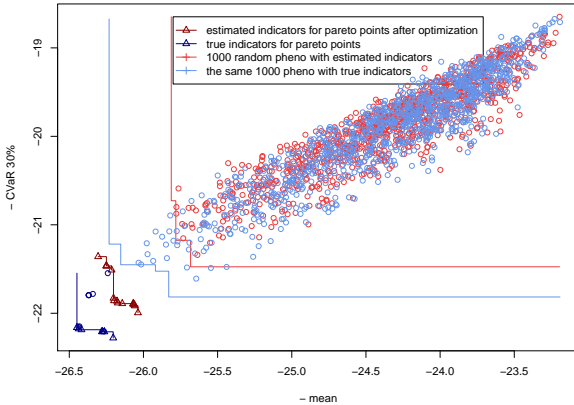


Noir : distribution exacte

Méthode d'optimisation :

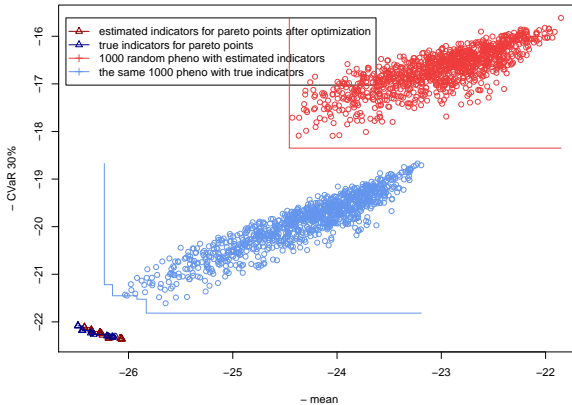
- Utilisation de deux critères de performance
- Optimisation multicritère NSGA2
 - 80 individus
 - 80 générations
 - 10 climats
- Deux lancements de la méthode d'optimisation
- Analyse post traitement avec 190 climats
- Comparaison avec 1000 phénotypes aléatoires

Pareto sets using 10 random pheno for reconstruction



- Globalement bonne reconstruction de l'espérance et la CVar30 sur les 1000 phénos du LHS
- Mauvaise reconstruction des phéno du front de Pareto

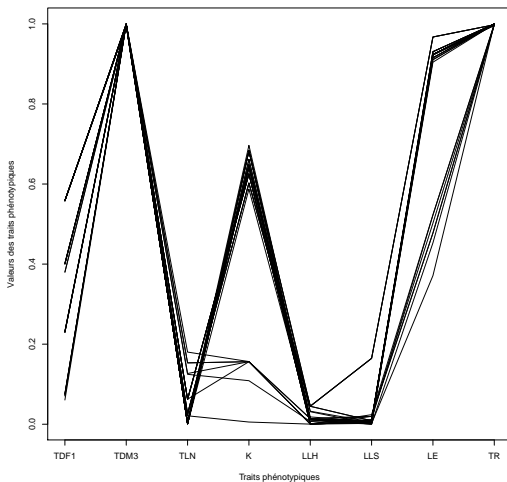
Pareto sets using 10 pheno from previous pareto front



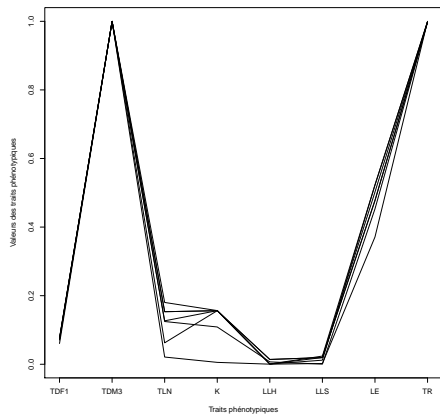
Difference : on utilise le front de Pareto issu du 1er lancement de NSGA-II comme base B pour la reconstruction au lieu de celle des 10 phenos utilisés pour la classification. → requiert la simulation exhaustive (sur les 190 climats) des phéno du front de pareto.

Résultats obtenus

Représentation des caractéristiques des phénotypes présents dans le front de Pareto



Stratégie prudente



- Période de floraison courte
 - Faible capacité d'interception de la lumière (Inclinaison de la plante)
 - Eviter les risques hydriques et solaires
- CVaR grande

- Sélection d'un sous-ensemble de climats par classification
- Distance "DTW" pour la classification, mise en place d'un algorithme (k-means)
- Différents critères de comparaison retenus :
 - Rendement moyen
 - "Conditional Value at Risk" à $\alpha\%$ du rendement pour différentes valeurs de α
- Résultats satisfaisants dans l'optimisation
- Perspectives :
 - Poids pour l'agrégation des distances
 - Apprentissage dans l'optimisation (modifications de l'algorithme d'optimisation MO-PSO)

Merci de votre attention