Analyse d'incertitude du modèle SUNFLO sous RECORD : Approche bayésienne et R-VLE

Francois Brun, Arnaud Bensadoun, Daniel D. Wallach, Philippe P. Debaeke, Luc Champolivier, Jean-Pierre Palleau, Emmanuelle Mestries

Analyse d'incertitude du modèle SUNFLO sous RECORD : approche bayésienne et R-VLE.

François Brun (ACTA), Arnaud Bensadoun (CETIOM/INRA), Daniel Wallach (INRA), Philippe Debaeke (INRA), Luc Champolivier (CETIOM), Jean-Pierre Palleau (CETIOM), Emmanuelle Mestries (CETIOM)

2ème Journée d'animation de la plate-forme RECORD
Toulouse, 27 janvier 2012

Objectif de la présentation

- **Illustrer sur un cas concret l'utilisation de l'interfaçage de RECORD-VLE avec le logiciel statistique R**

Analyse d'incertitude de SUNFLO

- Projet CASDAR 2010-2012 « associer un niveau d'erreur aux modèles »
  (RMT modélisation)
- Sur SUNFLOv1 sous RECORD-VLE
  - Utilisation pour l'accompagnement de l'évaluation variétale par les ingénieurs CETIOM
    - Aide à l'interprétation des données des réseaux d'évaluation
    - Comparaison des performances des variétés dans d'autres conditions (sol-climat-conduite) (expérimentation virtuelle)
  - Démarche d'analyse d'incertitude

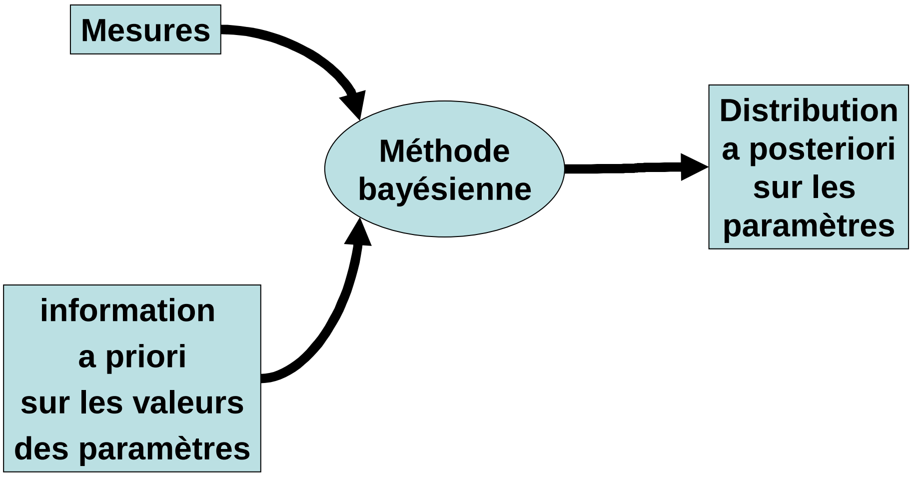
Démarche de l'analyse d'incertitude

- Comment avoir des prédictions de rendement et de teneur en huile assorties d'un indice de fiabilité ?

Etapes	Tâches
Définition des besoins et des contraintes	1) explicitation des variables d'intérêt (par rapport à utilisation)
	2) choix d'indicateurs d'incertitude pour les variables d'intérêt
Analyse d'incertitude	3) identification des sources d'incertitude
	4) caractérisation des informations disponibles
	$\boxed{\text{5) quantification des sources d'incertitudes}}$
	6) propagation de l'incertitude (Distribution de chaque variable d'intérêt)
	7) « meilleure réponse » (valeur moyenne de la variable d'intérêt)
	8) valeur des indicateurs d'incertitude
Analyse des résultats	9) analyse des contributions des différentes sources d'incertitude
Vérification des hypothèses	10) vérification avec des données
	11) explicitation et analyse des hypothèses

Tableau 1. Proposition d'une démarche opérationnelle et générique pour associer un niveau d'incertitude aux sorties d'un modèle.

Méthode bayésienne : dans le cadre de l'estimation des paramètres



Le modèle statistique

- $y = f(X, \theta) + \varepsilon$ Avec $\varepsilon \sim N(0, {\sigma_\varepsilon}^2)$
- On s'intéresse ici à la distribution de θ et ${\sigma_\varepsilon}^2$ sachant y
  - θ sachant y et ${\sigma_\varepsilon}^2$
  - ${\sigma_\varepsilon}^2$ sachant y et θ
- En statistique Bayésienne, on l'appelle la distribution *a posteriori*

$$P(\theta, \sigma_\varepsilon^2 | Y) \propto P(Y | \theta, \sigma_\varepsilon^2) \times \pi(\theta, \sigma_\varepsilon^2)$$

➔ **Exploration numérique de la distribution par chaîne de Markov**

Metropolis-Hastings within Gibbs

- Echantillonnage de distributions conditionnelles par étape

  - Etape 1 $P(\theta|Y, \sigma_\varepsilon^2) = P(Y, \sigma_\varepsilon^2|\theta) \times \pi(\theta)$

  - Etape 2 $P(\sigma_\varepsilon^2|Y, \theta) = P(Y, \theta|\sigma_\varepsilon^2) \times \pi(\sigma_\varepsilon^2)$

- Nécessite la définition

  - ✓ Des distributions *a priori* pour θ et σ_ε^2
  - ✓ D'une fonction de vraisemblance
  - ✓ D'une loi de proposition

$$g\left(\theta^{(n+1)}, \theta^{(n)}\right) = N(\theta^{(n)}, \Sigma)$$

Metropolis-Hastings within Gibbs

La vraisemblance

- Fonction qui relie les données aux paramètres

- Probabilité que les données aient été observé pour des valeur de paramètres données

$$P(Y|\theta, \sigma_\varepsilon^2\,) \propto \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}^{N/2}} \times e^{-\sum_{k=1}^{N}[Y_k - \boxed{f(X_k,\theta)}]^2/2\sigma_\varepsilon^2}$$

Step 1: Metropolis-Hastings

- On tire une valeur $\theta^{*(n+1)}$ dans $N(\theta^{(n)},\Sigma)$

($\theta^{(0)}$ et $\sigma_\varepsilon^{2\,(0)}$ sont fixés arbitrairement)

- On calcul le rapport

$$\alpha = \frac{P(Y|\theta^{*(n+1)},\sigma_\varepsilon^{2\,(n)}) \times \pi(\theta^{*(n+1)})}{P(Y|\theta^{(n)},\sigma_\varepsilon^{2\,(n)}) \times \pi(\theta^{(n)})}$$

- Si $\alpha > 1$: $\theta^{(n+1)} = \theta^{*(n+1)}$

- Sinon : $\theta^{(n+1)} = \theta^{*(n+1)}$ avec probabilité α
  $\theta^{(n+1)} = \theta^{(n)}$ avec probabilité $1\text{-}\alpha$

➔ **Grand nombre de simulations dépendantes entre elles**

Step 2 : Gibbs

- $P(\sigma_\varepsilon^2|Y,\theta) \propto \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}^{\,N/2}} \times e^{-\sum_{k=1}^{N}[Y_k - f(X_k,\theta)]^2/2\sigma_\varepsilon^2} \times \frac{1}{\sigma_\varepsilon^2}$

On pose $\tau = \frac{1}{\sigma_\varepsilon^2}$ pour obtenir

- $P(\tau|Y,\theta) \sim \Gamma\left(shape = \frac{N}{2} + 2,\ scale = 2/\sum_{k=1}^{N}[Y_k - f(X_k,\theta)]^2\right)$

➔ Tirage direct dans la distribution *a posteriori* (inverse gamma)

$$\sigma_\varepsilon^{2\,(n+1)} \sim 1/\Gamma(shape, scale)$$

➔ Recalcule de la vraisemblance de $\theta^{(n+1)}$ sachant $\sigma_\varepsilon^{2\,(n+1)}$

$$P(Y|\theta^{(n+1)},\sigma_\varepsilon^{2\,(n+1)})$$

R-VLE

- **Fonctions:**
  -lire des fichiers vpz (modèle)
  -modifier les valeurs d'entrées et les paramètres
  -lancer des simulations
  -récupérer les résultats sous forme d'objet R : *matrices* ou *dataframes*

R-VLE

- **Fonctions:**
  -lire des fichiers vpz (modèle)
  -modifier les valeurs d'entrées et les paramètres
  -lancer des simulations
  -récupérer les résultats sous forme d'objet R : *matrices* ou *dataframes*

➔ Suffisant pour estimer les paramètres d'un modèle VLE en profitant des outils statistiques (RNG, test, coda,…) de R

Démarche

- La commande de base : `run(monModèle, cond1=v1, cond2=v2,...)`

- Dans notre cas : 75 conditions (variables d'entrées et paramètres confondus)

➔ Il faut pouvoir changer certains paramètres (ceux qu'on estime) à chaque appel du modèle mais conserver exactement les mêmes données (celles sur lesquelles on veut estimer les paramètres)

Démarche

- Création d'une fonction pour formaliser le plan d'expérience : `A.plandexp()`

`plan_exp<-A.plandexp(tabUSM,tabLAI,tabTDM,var.out,type.out)`

Entrées :

- Définition des USM
- Observations sur les USM
- Valeurs des paramètres fixes
- Variable(s) de sortie désirée(s) (Rdt, teneur huile, LAI, Biomasse)
- Type de sortie désirée (Valeurs simulées, Valeur simulées et observées, SEQ)

Sorties :

- Une liste contenant le plan d'expérience

Démarche

- **Le modèle peut tourner via la fonction** `lancement.SUNFLO`

```
x<-lancement.SUNFLO(plan_exp,param.optim,thread)
```


Démarche

- **Le modèle peut tourner via la fonction** `lancement.SUNFLO`

```
x<-lancement.SUNFLO(plan_exp,param.optim,thread)
```


➔ **Reste à estimer ses paramètres**

Démarche

Définition des a priori à partir d'une table: `plan_RIBS<-B.planprior(tab)`

Paramètre	Valeur	Distribution	B inf	B sup	Moyenne	Var	Optim
Eb_fin	0.015	Normale	0		0.1	0.001	1
Eb_max	1	Uniforme	2	4			1
TBase	4.8						0

`B.planprior()` fournit également une matrice de variance-covariance pour la loi de proposition

Démarche

- **A notre disposition :**
  - **Des données**
  - **Un plan d'expérience**
  - **Un modèle**
  - **Des a priori sur les paramètres à estimer**

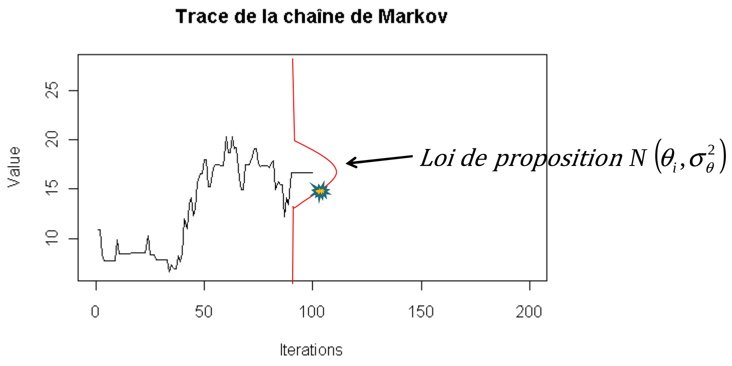
Démarche

- **A notre disposition :**
  - **Des données**
  - **Un plan d'expérience**
  - **Un modèle**
  - **Des a priori sur les paramètres à estimer**

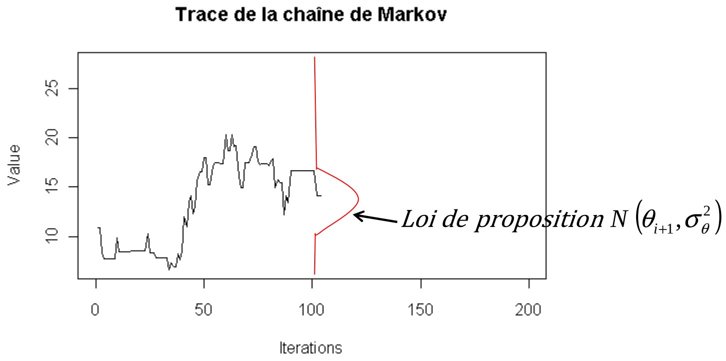
➔ ***Suffisant pour appeler RIBS***

RIBS is an Implementation of a Bayesian Sampler

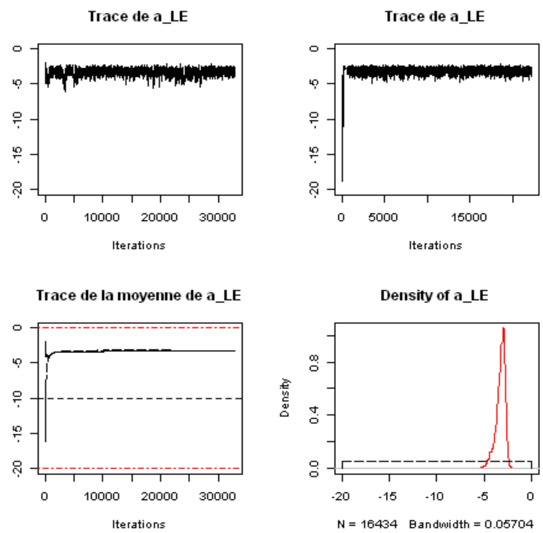
La méthode



La méthode



Résultats



Conclusion

- Interactions avec le modèle avec R-VLE
  - définition des paramètres, des unités de simulation, lancement du modèle, récupération des sorties
  - pas de passage par fichier (rapide)
  - éventuellement quelques modifications mineures dans le modèle VLE (« map » non pris en charge par RVLE)

⇒Faciliter d'interfacer les méthodes d'estimation de paramètres, d'analyse d'incertitude,… avec les modèles sous Record-VLE

Perspectives

- Faciliter l'accès aux méthodes disponibles
  - travail en cours dans le cadre du RMT modélisation (www.modelia.org)
  - création de ressources pédagogiques sous R
  - différents modèles exemples pour l'agronomie et l'environnement
  - un exemple encodé sous RECORD-VLE (Azodyn Colza à confirmer)

⇒Édition d'un livre d'exercice

⇒Sessions de Formations

Merci

La vie c'est comme une chaîne de Markov, on ne sait jamais sur quoi on va tomber!