

Analyses de sensibilité: application à des modèles agronomiques

Hervé Monod

Unité MIA

INRA-Centre de Jouy-en-Josas



- ❑ Modélisation et besoins en outils d'exploration numérique à l'INRA
- ❑ Analyse de sensibilité sur un modèle dynamique
- ❑ Analyse de sensibilité sur un modèle spatial

1. *La modélisation à l'INRA*

Caractéristiques :

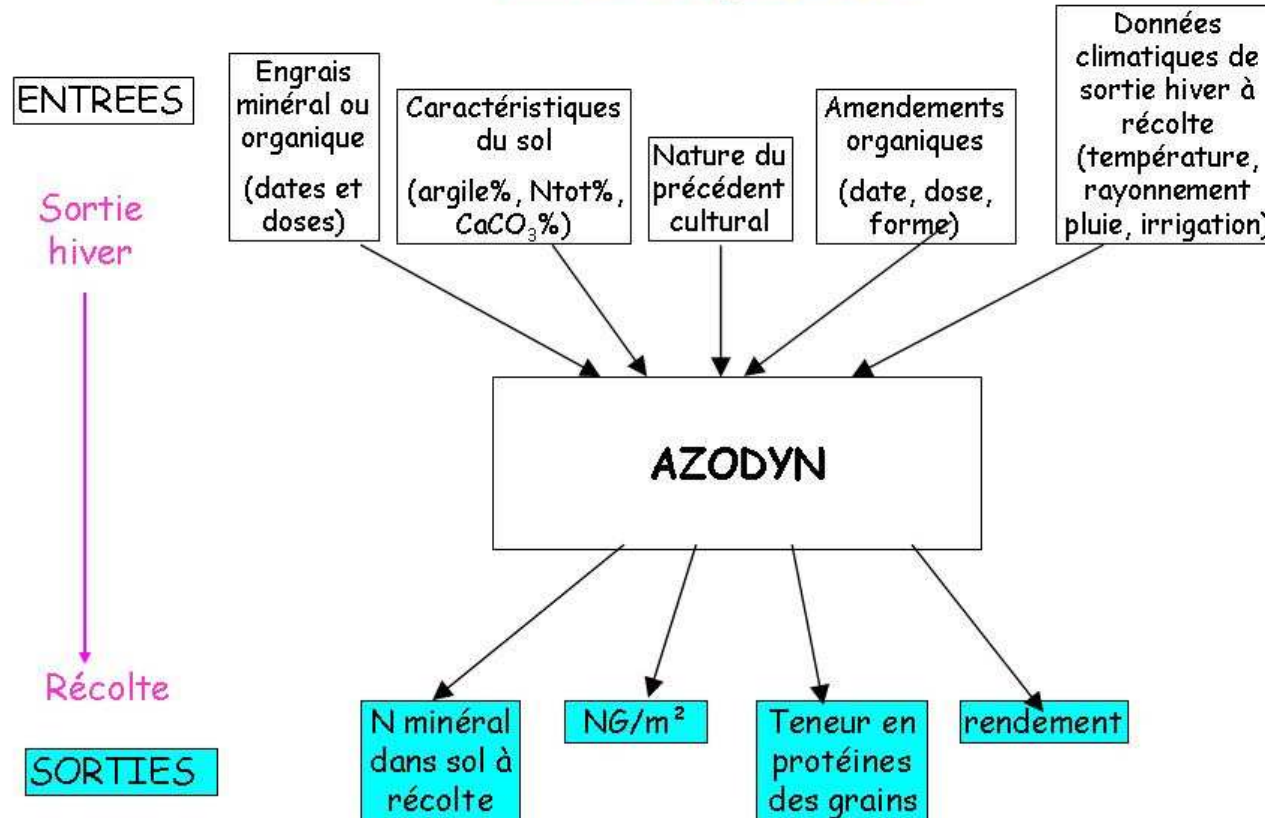
- *Diversité* :
 - biologie moléculaire → systèmes socio-économiques
 - approches et cultures de la modélisation très variées
- *Complexité croissante* :
 - étude de systèmes sur plusieurs niveaux d'organisation
 - intégration de données hétérogènes
 - *multiples sources de variabilité*
- *Rôle de + en + stratégique* :
 - inter-disciplinarité
 - quantification de phénomènes non/mal observables
 - outil de compréhension et de préconisation

Quelques exemples

A. Modèle de culture (AZODYN, M.-H. Jeuffroy, INRA-Grignon)

Ecole-Chercheur « Systèmes de culture », Nouan le Fuzelier, 13-17/05/02

Les entrées du modèle doivent être facilement accessibles par les utilisateurs potentiels



Exemple d'équations

$$\text{Biomass}(t + 1) = \text{Biomass}(t) + E [1 - \exp(-K \text{Surf.Fol}(t))] \text{Radiation}(t)$$

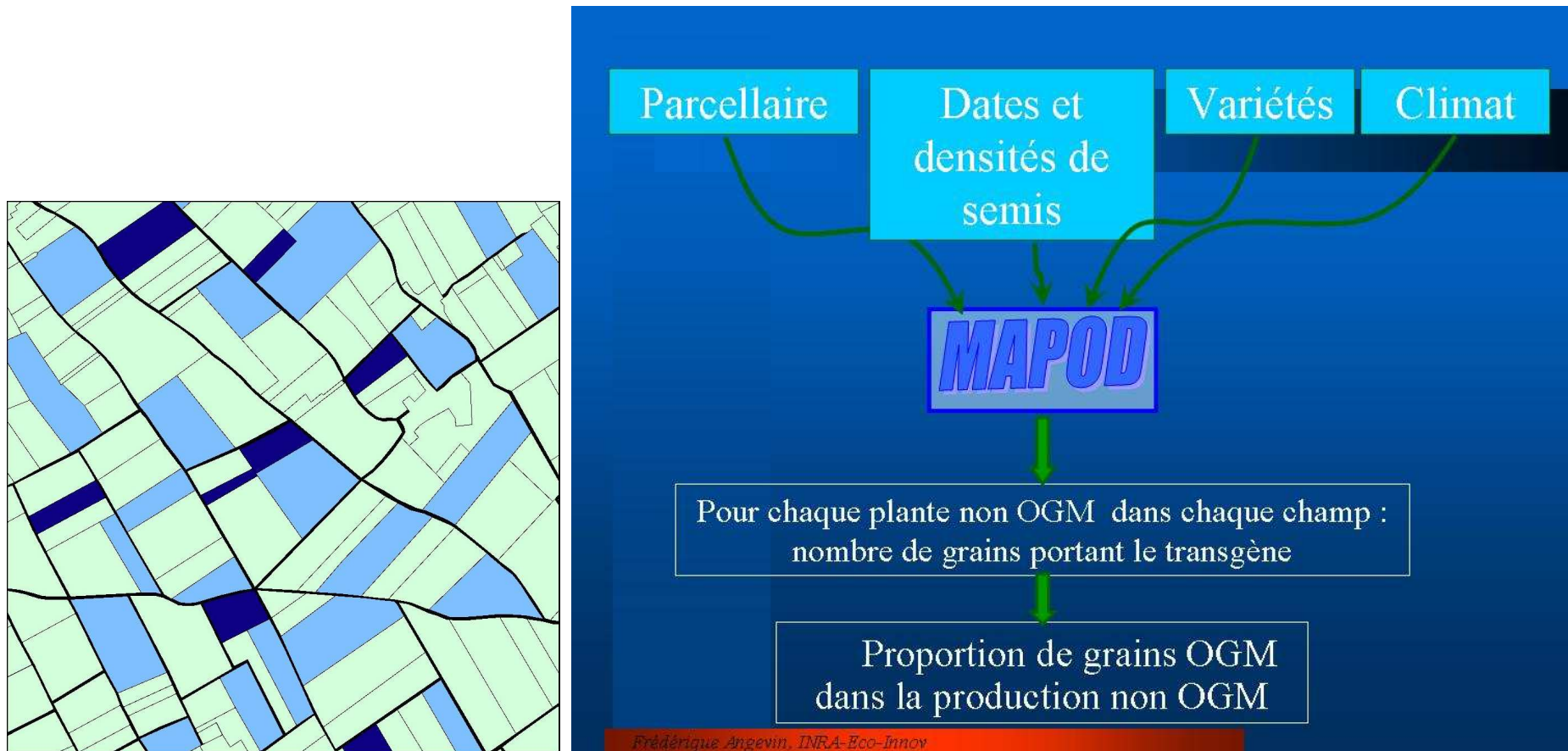
$$\text{Surf.Fol}(t) = L_{\text{MAX}} [1 - A \exp(-B \text{IndicedeNutrition}(t))]$$

Équations et paramètres définis à partir d'expérimentations plein-champ sur plusieurs sites-années

Applications :

- comparer l'influence des conduites de culture sur rdt, qualité, environnement
- optimiser les conduites
- prédire le rdt, la qualité, le risque de pollution

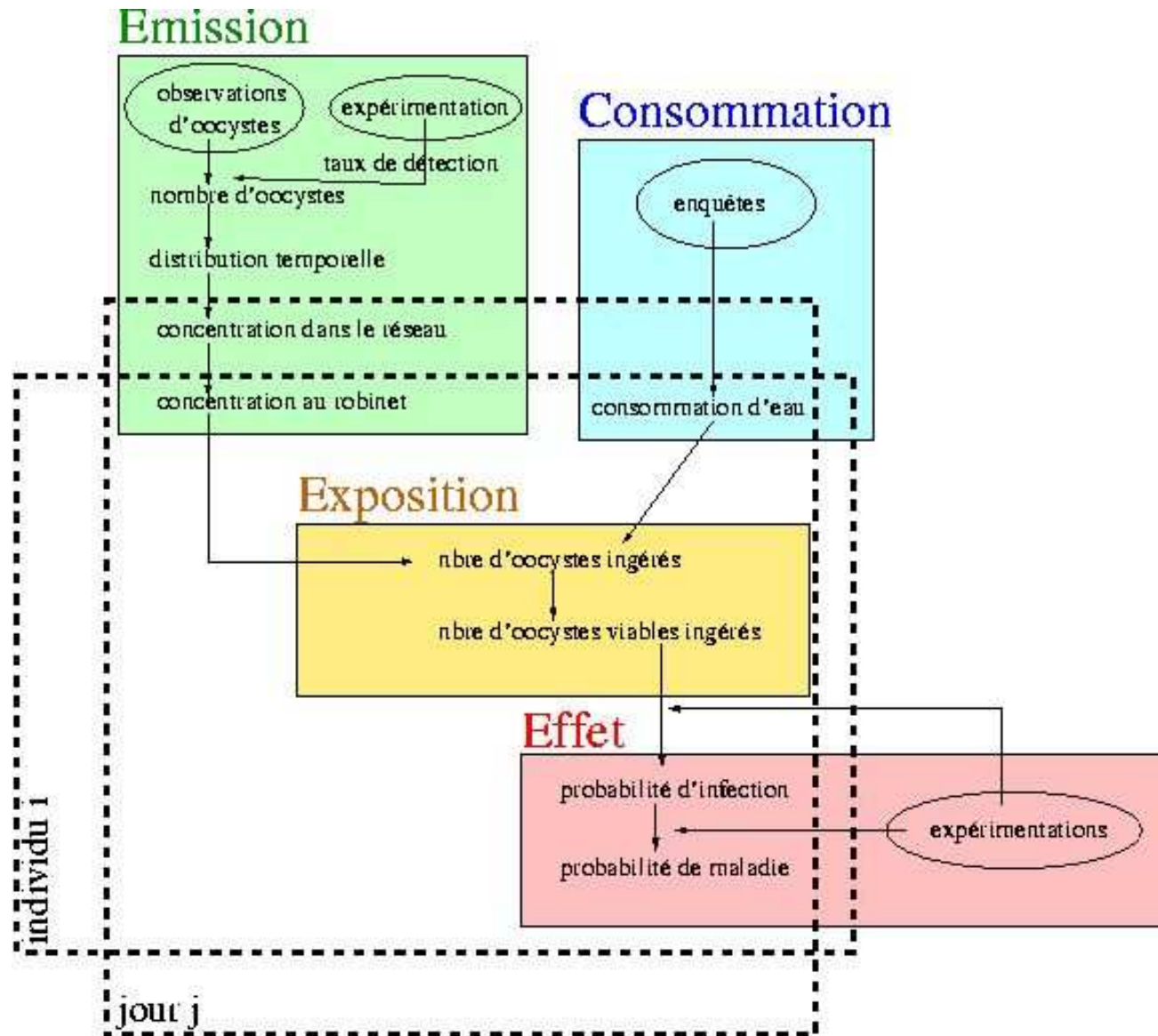
B. Modèle de flux de gènes (Mapod, F.Angevin-E.Klein; Genesys, N. Colbach; SIGMEA)



Echelle	Recherche	Gestion
	<p>Modèles agronomiques</p> <p>Mapod (Angevin, Klein, Meynard, 2001)</p> <p>Genesys (Colbach, Clermont Dauphin, Meynard, 2001)</p> <p>Simulations</p>	<p>Étude des systèmes de cultures</p>
	<p>Modèles de fonctionnement des exploitations (Angevin <i>et al.</i>, 2002)</p> <p>simulations</p>	<p>Étude des pratiques mises en œuvre par les agriculteurs</p>
	<p>Modèles de bassins d'approvisionnement (Le Bail, 2003)</p> <p>simulations</p>	<p>Étude des pratiques de collecte stockage</p>

Frédérique Angevin, INRA-Eco-Innov

C. Modèle d'évaluation de risque en alimentation (cryptosporidies; J.-B. Denis, Jouy et coll. AFSSA)



Besoins en exploration numérique de modèles

- ❑ comprendre le comportement du modèle (analyses de sensibilité)
- ❑ identifier les paramètres et variables d'entrée clés
 - *Exemples*
 - Agronomie : modèles de culture (dynamiques journaliers)
 - Eco-toxicologie : croissance de populations de gambusies
 - Epidémiologie animale : développement d'une épidémie dans un troupeau
- ❑ prédire en intégrant l'incertitude
 - *Exemples*
 - Risque alimentaire : évaluation quantitative
 - Risque environnemental : dispersion de transgènes
- ❑ comparer/optimiser *in silico*
 - *Exemples*
 - Distances d'isolement entre champs
 - Recherche de stratégies optimales

⇒ **Projet d'animation interne : formation, échanges, programmes**

2. Analyse de sensibilité d'un modèle dynamique

Campbell, McKay, Williams, SAMO 2004

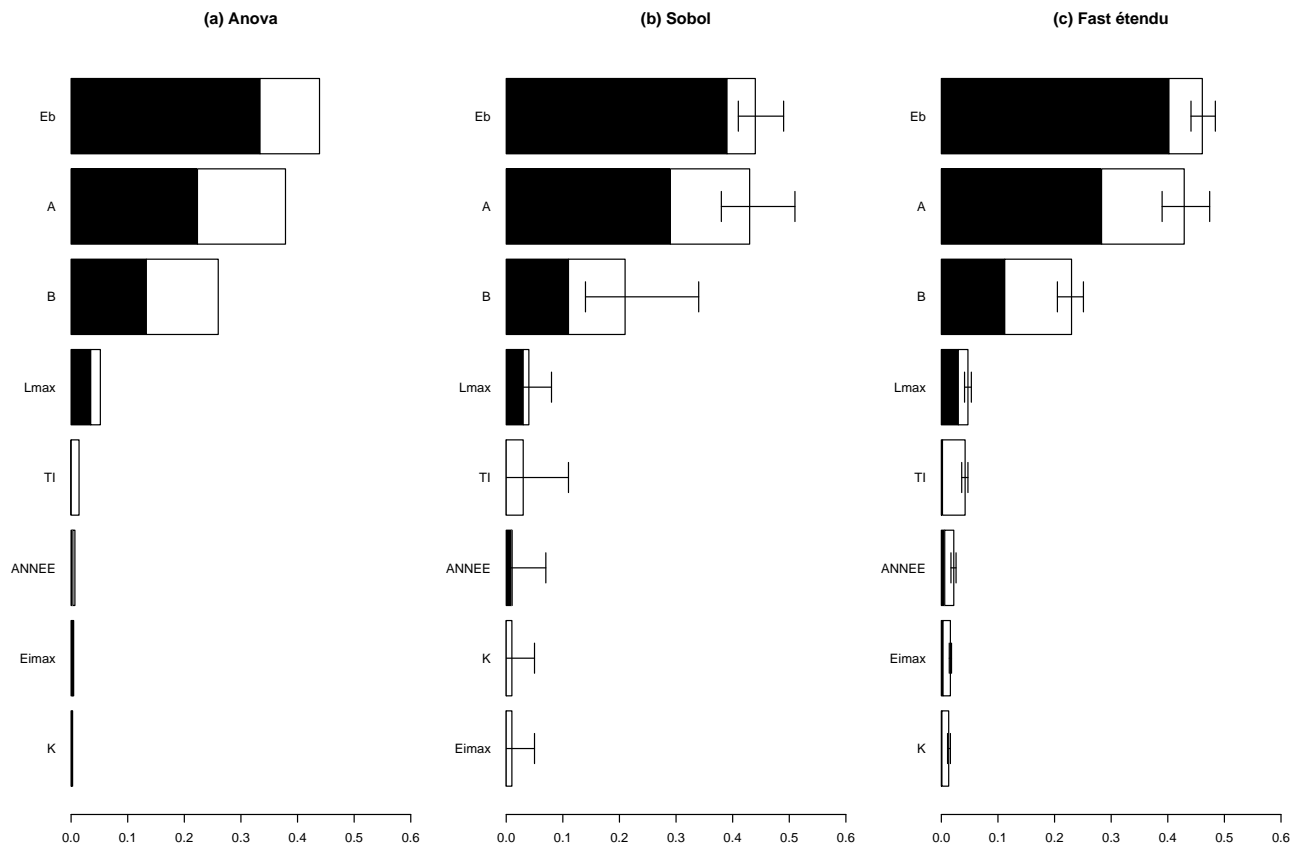
Le modèle AZODYN :

- ❑ Modèle dynamique simulant la croissance d'une culture de blé.
- ❑ Réponses :
 - rendement, teneur en protéines, indice de nutrition azotée journalier (INN)
- ❑ Variables d'entrée :
 - climat journalier, composition du sol, dates de fertilisation
- ❑ 69 paramètres
- ❑ 13 paramètres génétiques

- Intervalles d'incertitude sur les paramètres génétiques :

Paramètre	Définition	Intervalle d'incertitude	Unité
RDTMAXVAR	rdt maximal	10.0 - 13.7	t.ha ⁻¹
Tep.flo	Durée entre épiaison et floraison	100 - 200	°C.jour
E_{bmax}	efficacité d'utilisation du rayonnement	2,7 - 3,3	g.MJ ⁻¹
E_{imax}	ratio rayonn ^t intercepté/incident	0.0 - 0.99	-
D	Rapport surf.fol./azote critique	0.02 - 0.045	-
K	Coefficient d'extinction	0,6 - 0,8	-
R	Ratio azote total/aérien	1.0 - 1.5	-
Lambda	Efficacité d'utilisation N	25 - 45	-
Mu	Efficacité d'utilisation N	0.6 - 0.9	-
NGM2MAXVAR	Capacité à produire un nb de grains/m ²	107.95 - 146.05	-
P1GMAXVAR	Poids maxi d'un grain	47 - 65	mg
REM2	Fraction d'azote remobilisé	0.5 - 0.9	-
DJPF	Seuil de température	150 - 250	°C.jour

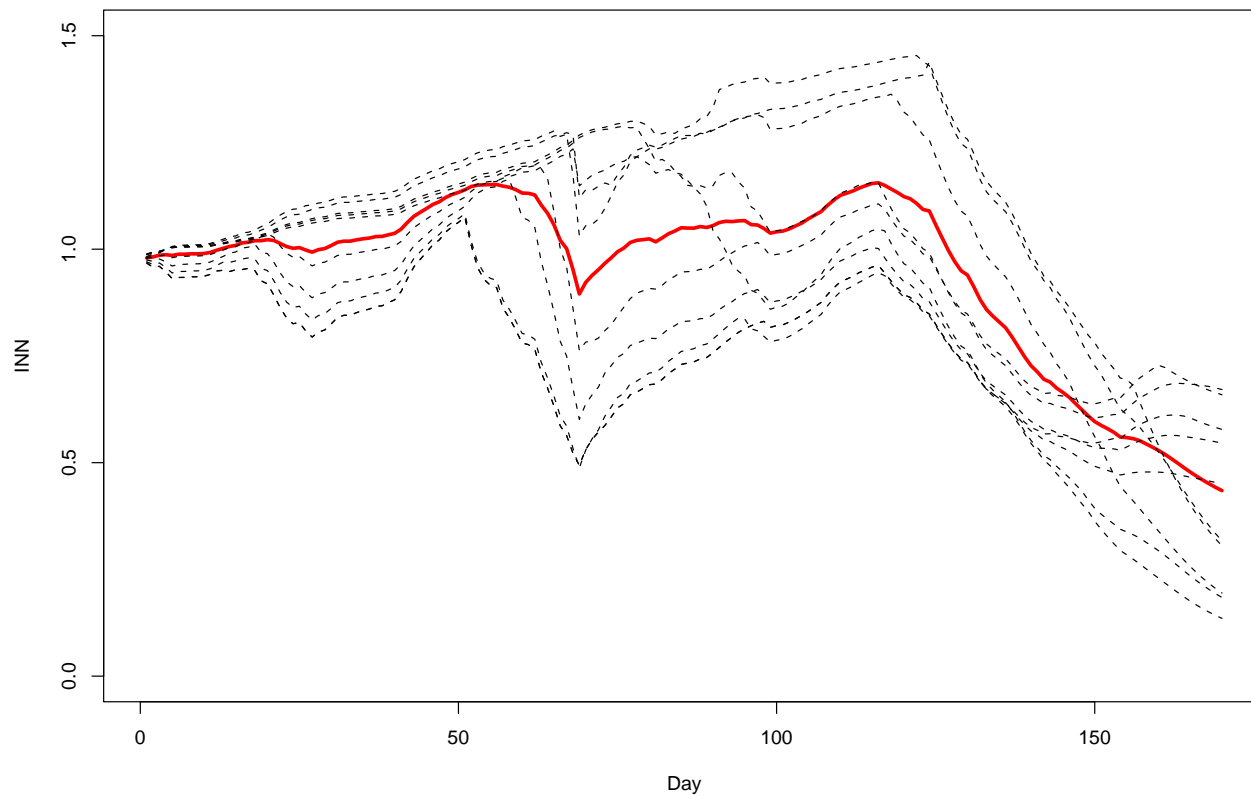
Rendement, taux de protéine



INN :

La sortie du modèle est une courbe : comment caractériser sa sensibilité aux paramètres ?

- Réponse moyenne du modèle et qqs variations :

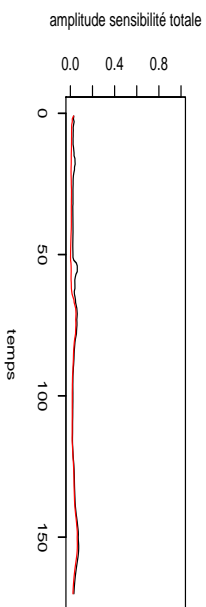


Possibilités: Analyses de sensibilité ...

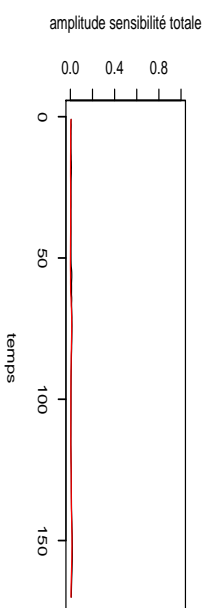
- ❑ ... dynamiques au jour le jour
- ❑ ... sur sorties “synthétiques” d'intérêt
- ❑ ... sur coefficients d'une décomposition dans une base fixée
- ❑ ... sur coefficients d'une décomposition dans une base données-dépendante

Dynamiques au jour le jour

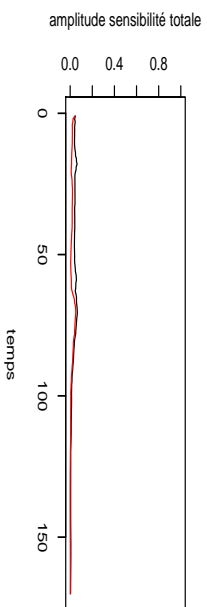
Ebmax



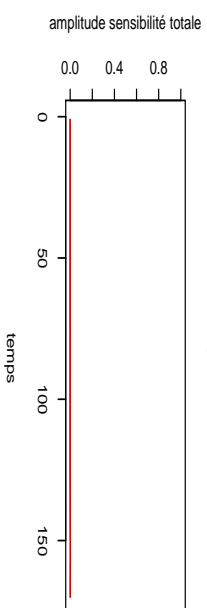
Eimax



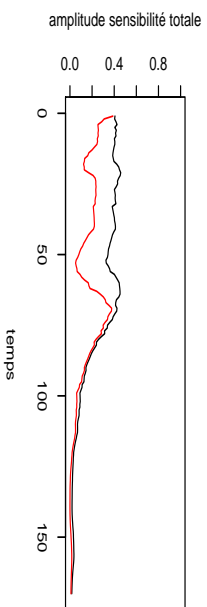
K



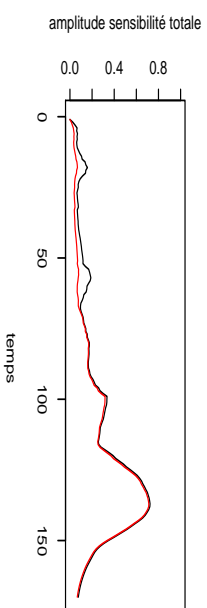
Tep.flo



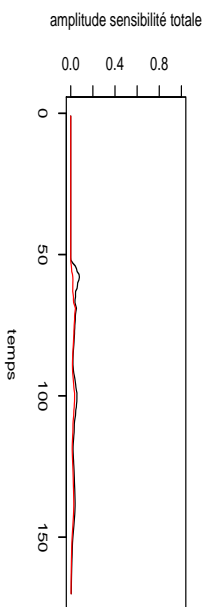
D



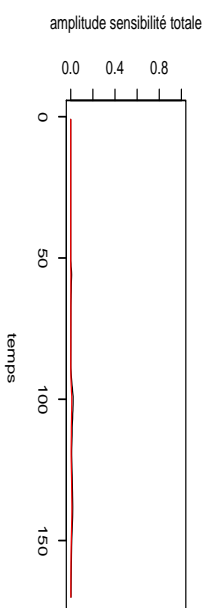
R



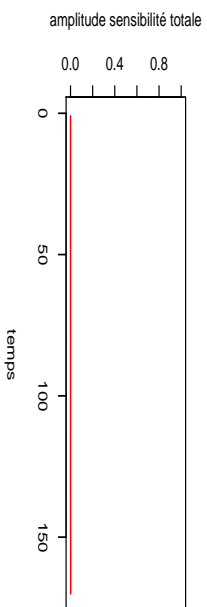
Lambda



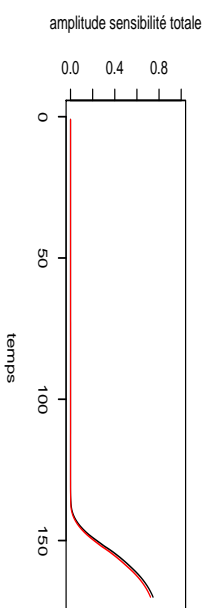
Mu



DUPF



REM2



Décomposition données-dépendantes : Principales étapes

1. Plan d'expériences → simulations

$$\mathbf{X} = \begin{pmatrix} x_{1,1} & \dots & x_{1,n} \\ \vdots & & \vdots \\ x_{u,1} & \dots & x_{u,n} \\ \vdots & & \vdots \\ x_{N,1} & \dots & x_{N,n} \end{pmatrix} \implies \mathbf{Y} = \begin{pmatrix} y_{1,1} & \dots & y_{1,T} \\ \vdots & & \vdots \\ y_{u,1} & \dots & y_{u,T} \\ \vdots & & \vdots \\ y_{N,1} & \dots & y_{N,T} \end{pmatrix}$$

1. **Plan d'expériences** → simulations

2. **Analyse multivariée** → ex: ACP

$\mathbf{v}_1, \dots, \mathbf{v}_K$ base orthonormée de \mathbf{R}^T t.q. \mathbf{v}_1 maximise la variance des $(\langle \mathbf{v}_1, \mathbf{y}_u \rangle)_u$, etc.

→ décomposition: $\mathbf{y}_u - \mathbf{y}_\bullet = \sum_{k=1}^K h_{u,k} \cdot \mathbf{v}_k$

$$\mathbf{X} = \begin{pmatrix} x_{1,1} & \dots & x_{1,n} \\ \vdots & & \vdots \\ x_{u,1} & \dots & x_{u,n} \\ \vdots & & \vdots \\ x_{N,1} & \dots & x_{N,n} \end{pmatrix} \implies \mathbf{Y} = \begin{pmatrix} y_{1,1} & \dots & y_{1,T} \\ \vdots & & \vdots \\ y_{u,1} & \dots & y_{u,T} \\ \vdots & & \vdots \\ y_{N,1} & \dots & y_{N,T} \end{pmatrix} \implies \mathbf{H} = \begin{pmatrix} h_{1,1} & \dots & h_{1,K} \\ \vdots & & \vdots \\ h_{u,1} & \dots & h_{u,K} \\ \vdots & & \vdots \\ h_{N,1} & \dots & h_{N,K} \end{pmatrix}$$

3. **Analyses de sensibilité** sur les scores $(h_{u,k})_u$

- Sensibilités de premier ordre :

$$\hat{S}_i = \frac{SC_i}{SC_T}$$

- Sensibilités totales :

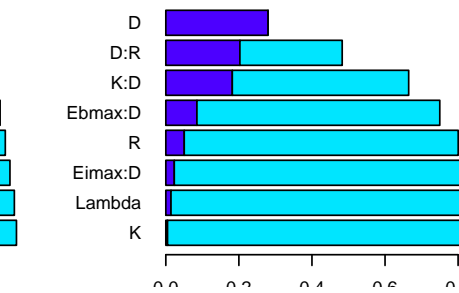
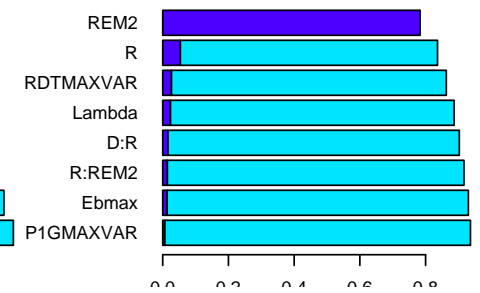
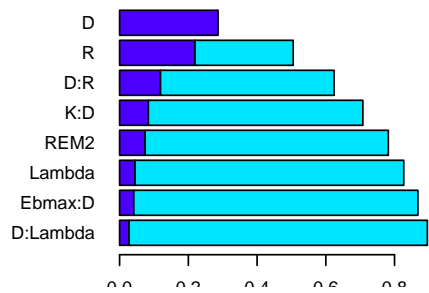
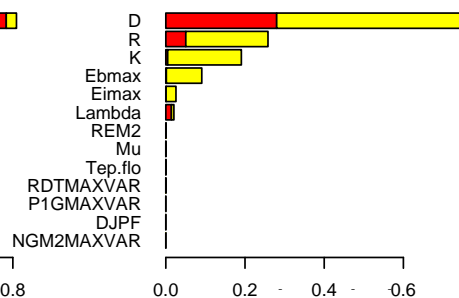
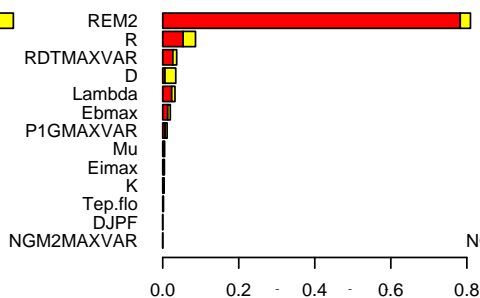
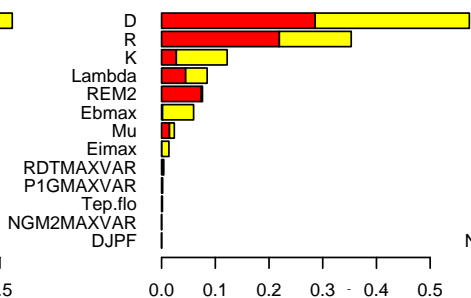
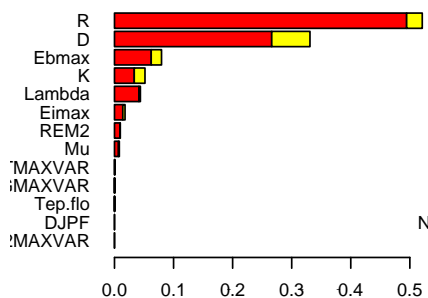
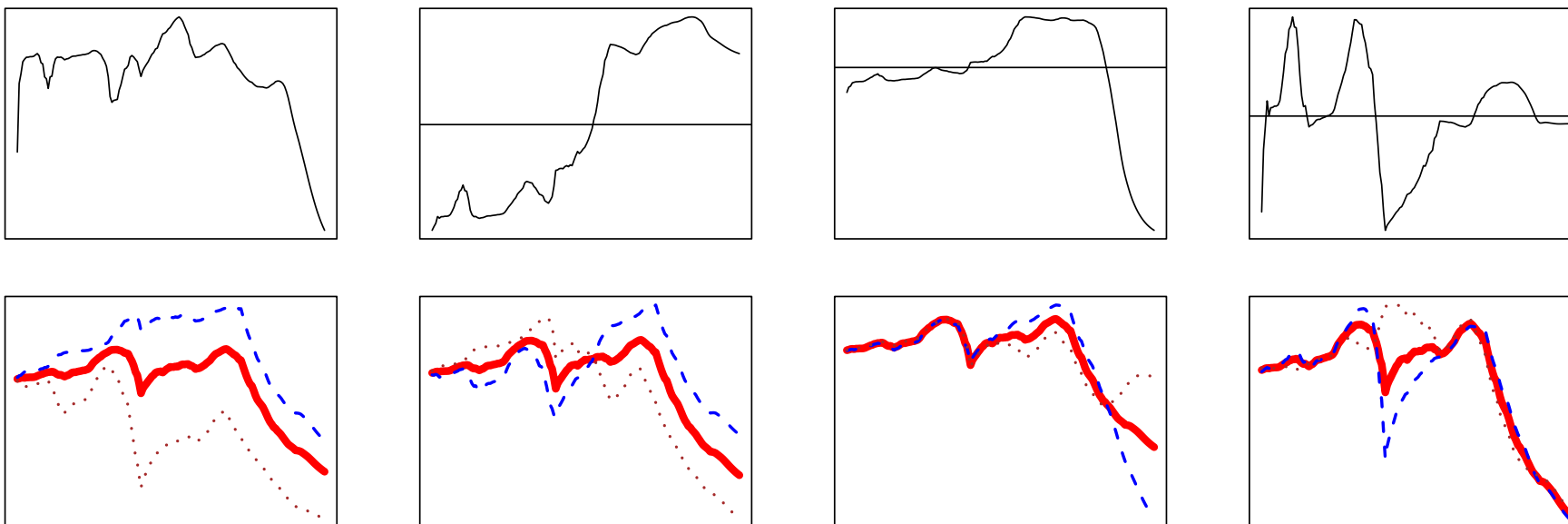
$$\hat{S}_i^{tot} = \frac{SC_i + \sum_j SC_{ij} + \sum_{j,k} SC_{ijk} + \dots + SC_{1, \dots, i, \dots, n}}{SC_T}$$

Application à AZODYN

1. Plan factoriel à deux niveaux : le *min* et le *max* de chaque intervalle d'incertitude.
2. ACP normée : les dates sont les variables.
3. Analyse de la variance sur les premières composantes principales.
4. Calcul des indices de sensibilité.

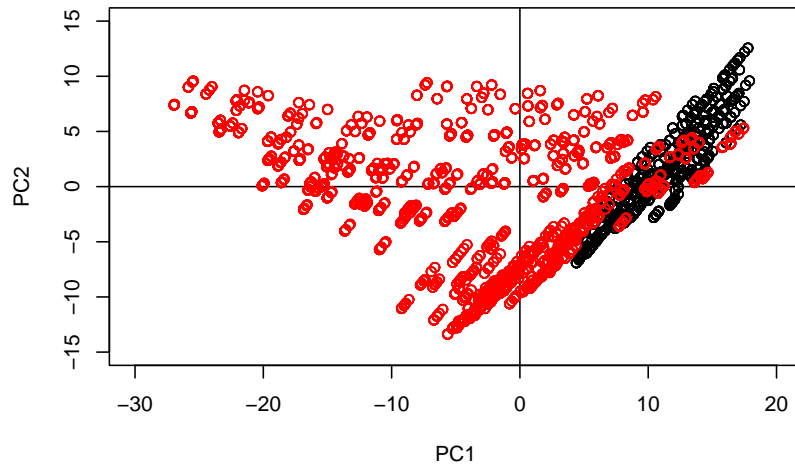
Résultats

- 85,08% d'inertie sur le premier plan factoriel
- Inertie de la troisième composante principale : 7,15%

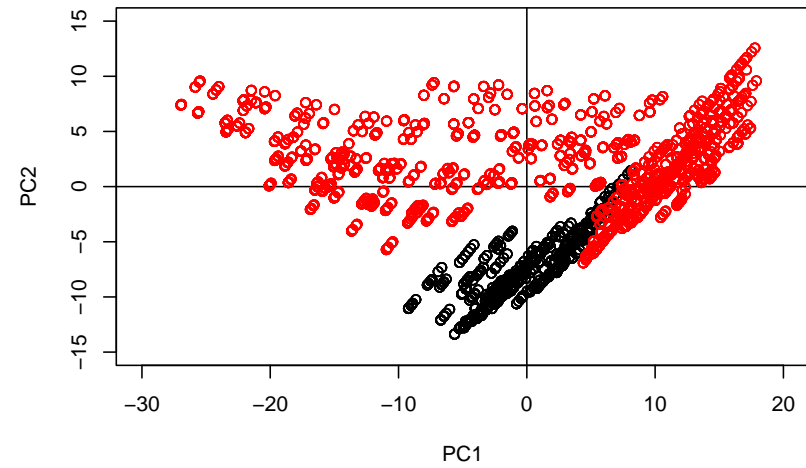


Projection sur le premier plan factoriel :

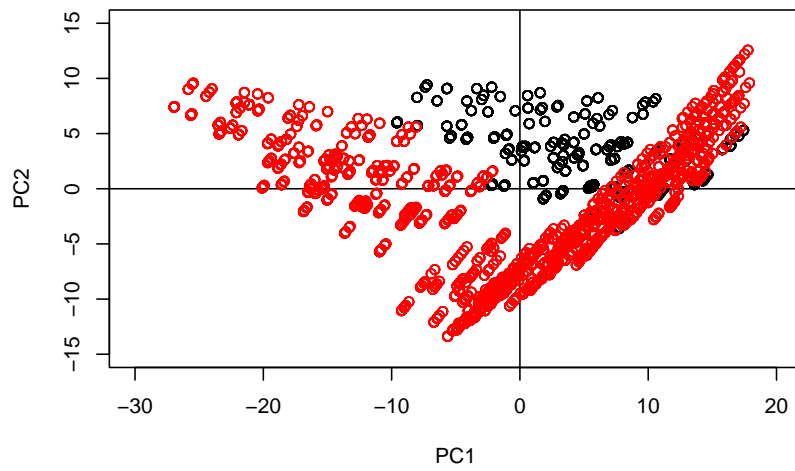
D=0.02 et R=1



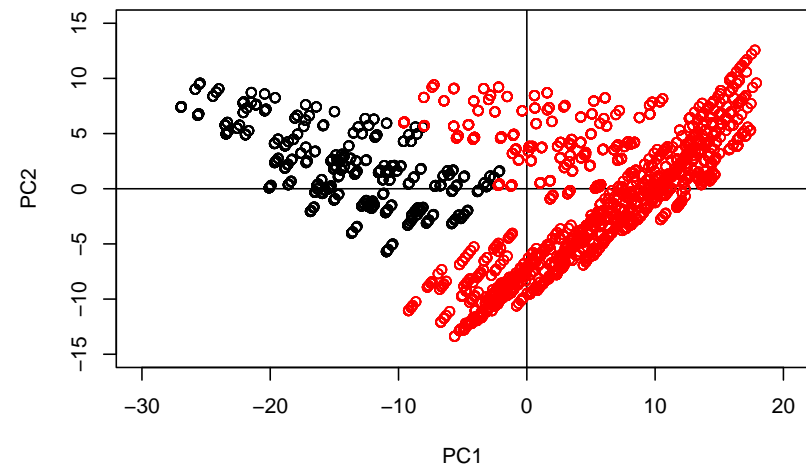
D=0.02 et R=1.5



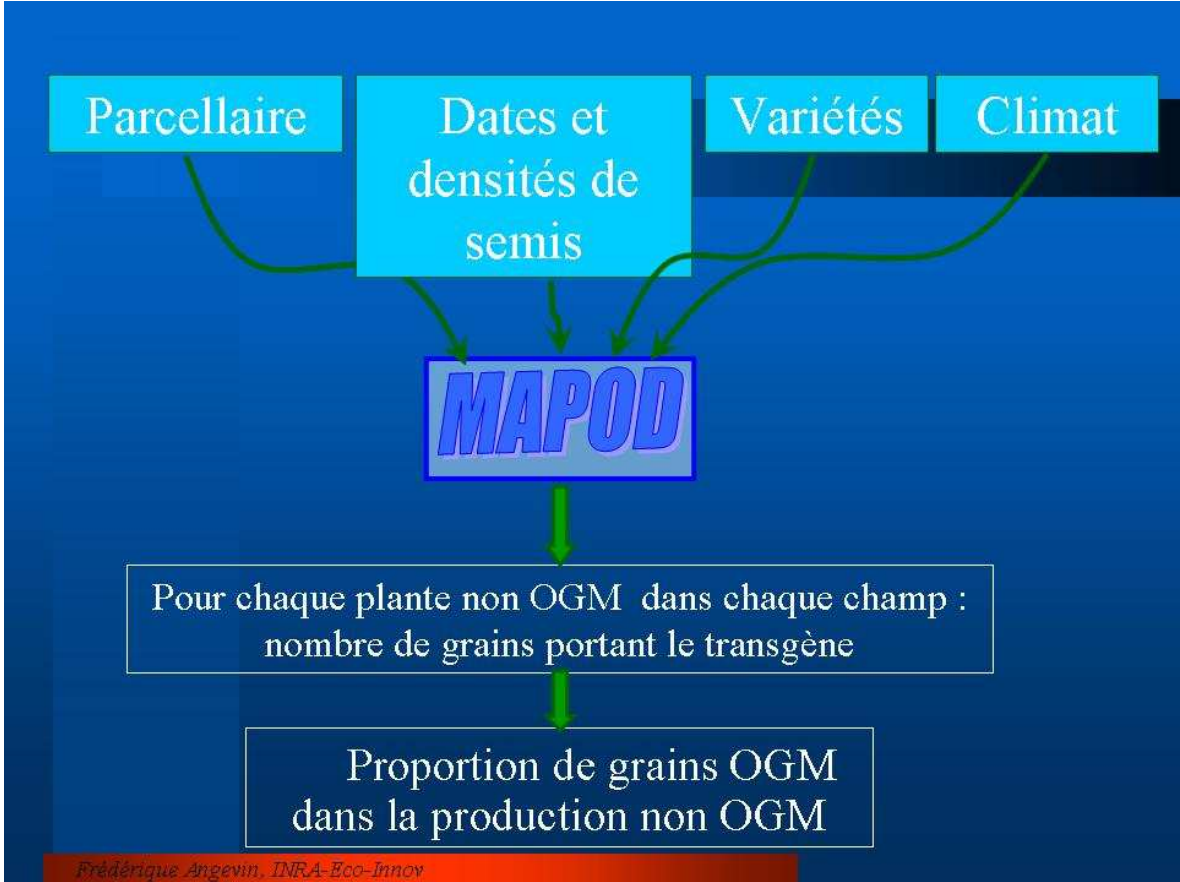
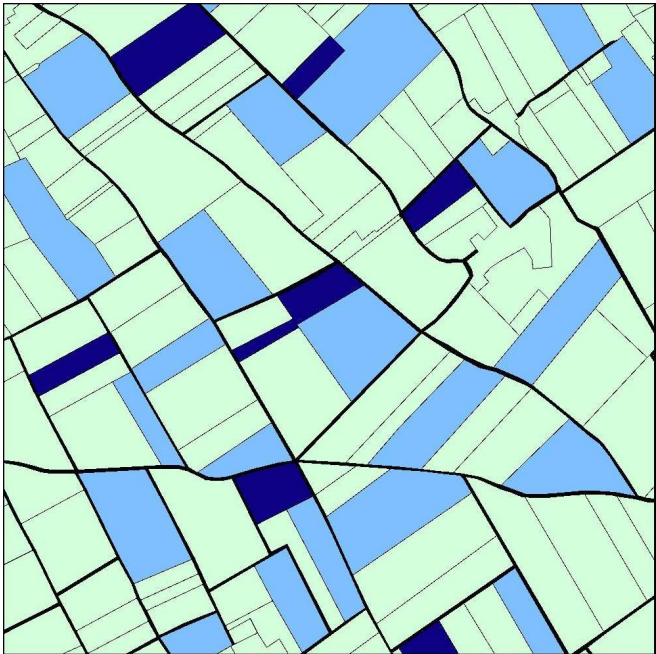
D=0.045 et R=1



D=0.045 et R=1.5



3. Analyse de sensibilité d'un modèle de flux de gènes



Facteurs d'entrée :

Classe du paramètre	Paramètre	Gamme de variation globale
paysage	descripteurs spatiaux ...	-
	pourcentage de maïs	-
	pourcentage d'OGM	-
météo	températures	-
	précipitations	-
	vent	-
	...	-
Courbe de dispersion	paramètres	-
	-	-
+ caract. variétés	fbraison	-
	...	-
+ conduite de culture	...	-
	...	-

Chaîne de calcul :

Spécification des gammes de variations des facteurs d'intérêt



Spécification du plan d'expériences ou d'échantillonnage



Construction des fichiers paysages et météos désirés



Calcul de variables ponctuelles locales et parcellaires



Simulations du modèle (*Matlab*)



Calcul des sorties ponctuelles : taux de pollution



Calcul des sorties agrégées par parcelles



Calcul des sorties agrégées par simulation



Analyses des résultats

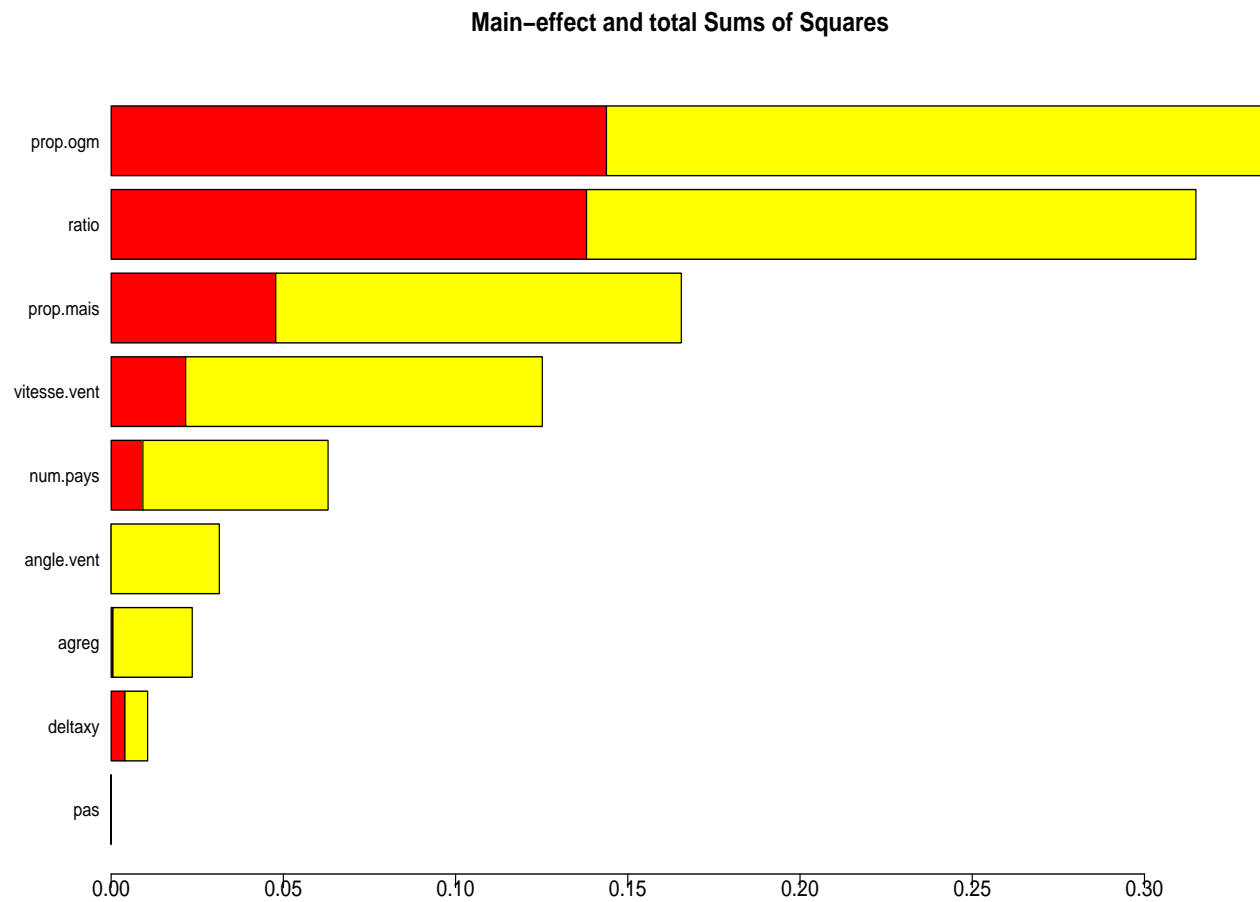


Figure 1: exemple de sortie recherchée

Conclusion

- ❑ besoins d'outils diversifiés
- ❑ intérêt d'un environnement complet en outils math-stat
- ❑ nécessité de prendre en compte différents niveaux de variabilité
- ❑ adapter le plan de simulations à des questions précises
- ❑ corrélations en entrée; multivarié en sortie; stochastique entre les deux
- ❑ complémentarité avec l'expérimentation: estimation/validation

Référence suppl.: Dreesbeke, Fine, Saporta (Eds) (1997). Plans d'Expériences: Applications à l'Industrie. Technip.

Contributions-collaborations: Frédérique Angevin, Arnaud Gauffreteau, Marie-Hélène Jeuffroy, Claire Lavigne, David Makowski ; Katarzyna Adamczyk, David Makowski, Cédric Naud, Kevin Perraud, Flavien Chevauchet, Thomas Romary, Valérie Viaud
projet MRT "Impact des OGM"; projet européen sigma