

Simple and efficient tools to explore complex models in agroecology with an eye on interactions

Robert Faivre, Victor Picheny, Hervé Monod

INRA - Mathématiques et Informatique Appliquées
MIA Toulouse and Jouy-en-Josas

Nice, July 3rd, 2013



Genesis

A lot of models are developed in agroecology (from the simplest to very complex models)

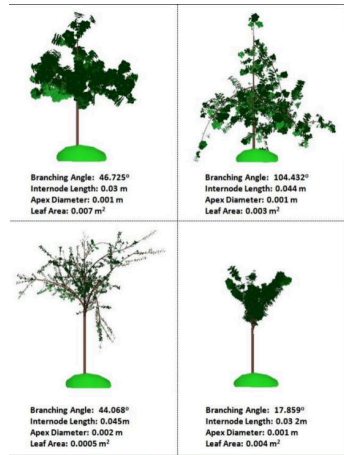
MappleT (Markov Apple Tree)
(Costes *et al.*, 2008)

Manipulation of tree architecture (genetical parameters)

Optimizing light interception

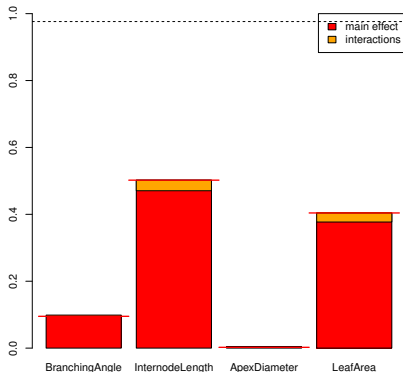
Four parameters : leaf area, internode length, top shoot diameter, branching angle

Computing time ($\sim 1h$ for a 5 years-old tree)



Sensitivity analysis

Multiple polynomial metamodel of degree 3 based on a LHS of size 300



97.6 % of variance of light interception explained.

Polynomial Linear Metamodel

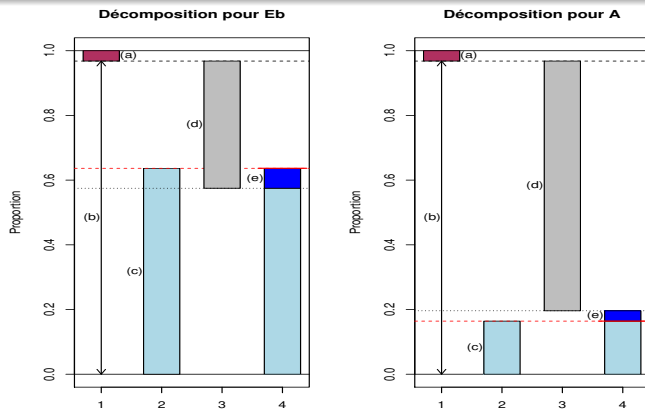
$$Y = \sum_{a=1}^A \beta_a \left(\prod_{k=1}^K X_k^{d_{a,k}} \right) + \eta$$

- K is the number of input parameters ;
- $A = C_{K+D}^D$ is the number of cross product terms ($0 \leq \sum_k d_{a,k} \leq D$) ;
- D the maximal degree of the polynomial ;
- η is a centred random term independent of the X_k variables.

Decomposition of the sources for X_1 (% of explained variance, R^2)

- Main effect : $X_1 + X_1^2 + X_1^3$
- Total : $X_1 + X_1^2 + X_1^3 + X_1 X_2 + X_1 X_2^2 + X_1^2 X_2 + X_1 X_3 + X_1 X_3^2 + X_1^2 X_3 + X_1 X_2 X_3 + \dots$

Sources of variation



- (b) = R_C^2 coefficient of determination of the complete model ; (a) = $1 - R_C^2$ the residual
- (c) = R_k^2 corresponds to the main effect of the parameter X_k negligible
- (d) = R_{-k}^2 is the coefficient of determination of the plmm when X_k is not taken into account
- (b) - (d) = $R_C^2 - R_{-k}^2$ is the specific contribution of X_k
- (e) = interaction (left) and/or confusion (right) ?

Design

Most typical sampling designs

MC Monte Carlo

lhs Latin hypercube

Common property (= common problem)

- independent sampling across dimensions
- \Rightarrow no control of d -dimensional subspaces, for $d > 1$

Higher-dimensional-subspace-filling designs : many methods, including

OA-lhs OA-based Latin hypercube

tms-net digital net (with scrambling)

OA-lhs and tms-nets : basic principles

For a design of size $N = p^s$ and maximum strength

- equally replicate p^s hyper-rectangles
- methods differ by which rectangle shapes are considered

Construction

- huge literature, websites (see Lemieux book, 2009)
- R package PLANOR (Monod, Bouvier, Kobilinsky, 2012)
 - define s nested pseudofactors at p levels for each factor
 - construct a suitable regular factorial fractional design
 - recompose quantitative levels and scramble

lhs property ?

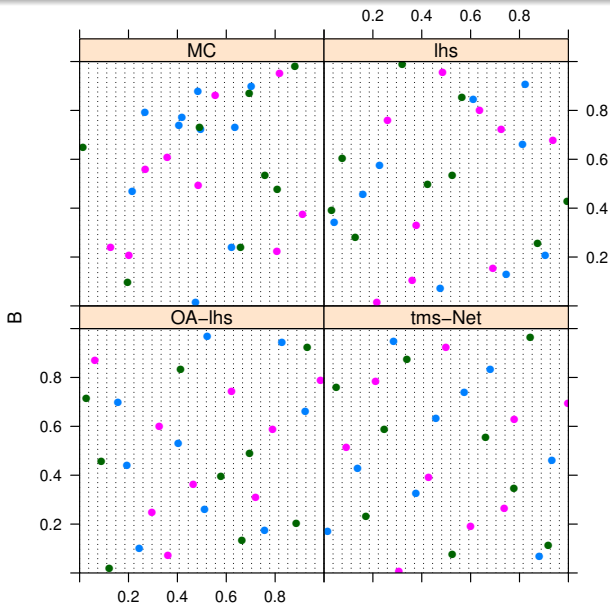
Example :

3 factors

$N = 27$

basis $p = 3$

Design	?
MC	No
lhs	Yes
OA-lhs	Yes
tms-net	Yes



OA property (strength 3)?

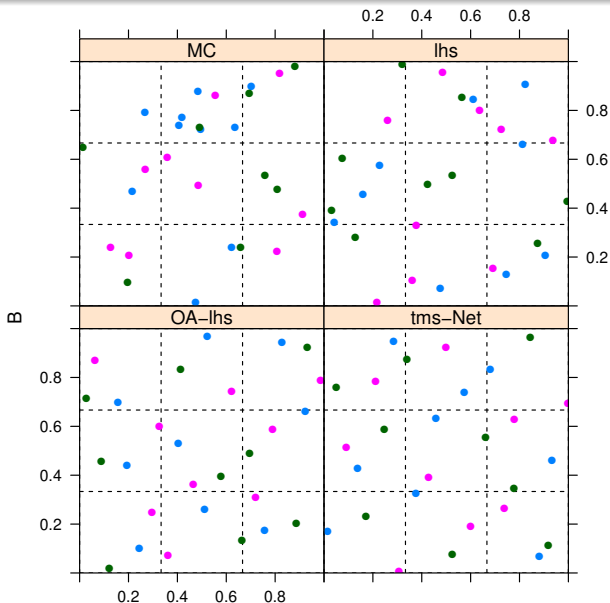
Example :

3 factors

$N = 27$

basis $p = 3$

Design	?
MC	No
lhs	No
OA-lhs	Yes
tms-net	Yes



equidistribution properties ?

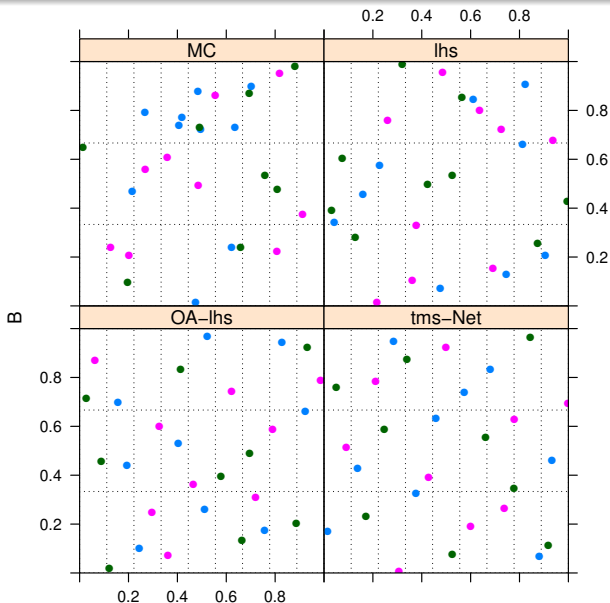
Example :

3 factors

$N = 27$

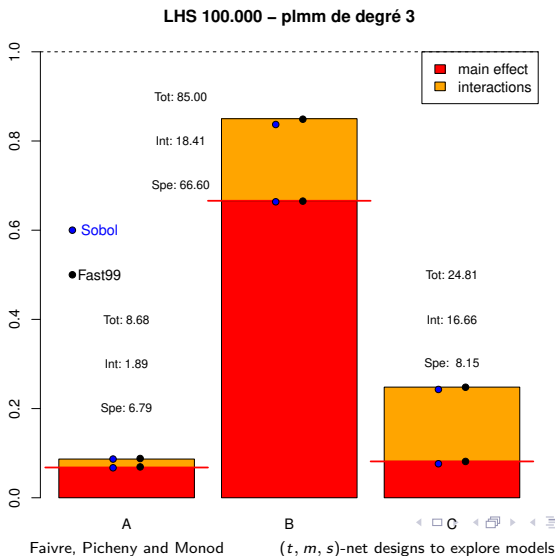
basis $p = 3$

Design	?
MC	No
lhs	No
OA-lhs	No
tms-net	Yes



Case study

A simple model : $1 + x_1 - 0.1x_1^2 + 0.2x_1^3 + x_2 + 0.3x_2^2 - x_1x_2 - 2x_3 + 3x_2x_3^2$

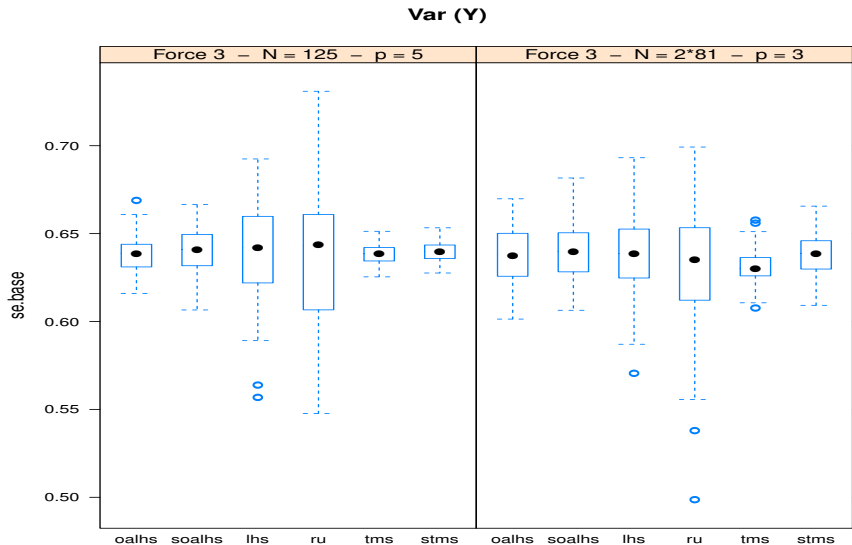


Simulations

- sampling designs : oalhs (*fixed* and scrambled), lhs, Monte Carlo, (t, m, s) -net (*fixed* and scrambled)
- depending on force and number of levels
- with size depending on force and the number of levels

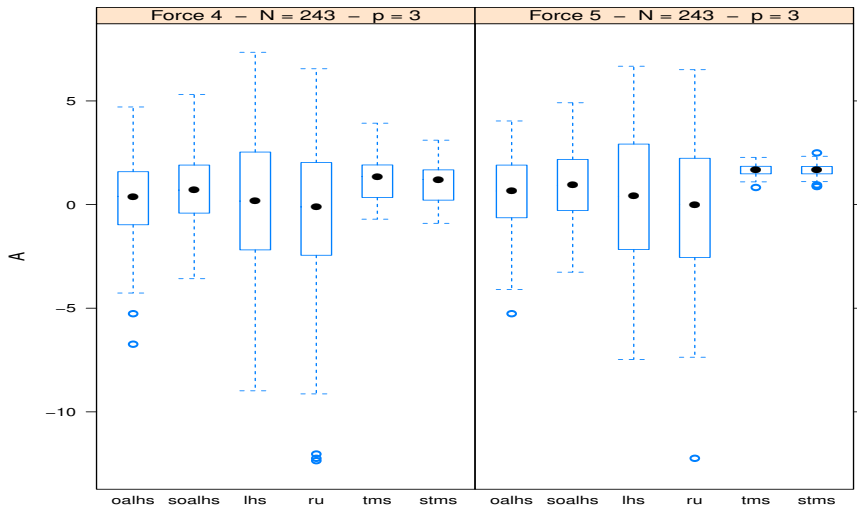
	Strength		
	2	3	4
3 levels	(9, 27, 81)	(27, 81, 243)	(81, 243)
5 levels	(25, 125)	(125)	(625)

- twice sized : paired oalhs (*fixed* and scrambled)
- analysis with plmm
- 99 replicates

Results : Estimation of $\text{var}(Y)$ 

Results

Interaction/Confusion for A





Conclusion

- **preliminary results**
- sampling designs : lower interval of variation
 (t, m, s) -net $>$ oalhs $>$ lhs $>$ Monte Carlo
- number of levels : $5 > 3$
- the strength is with us (conditionned to the size N)

to be continued

References

- E. Costes, et al., 2008. "MAppleT : simulation of apple tree development using mixed stochastic and biomechanical models," *Functional Plant Biology*, vol. 35, 2008, pp. 936-950.
-  Faivre R., looss B., Mahévas S., Makowski D., Monod H., editors, 2013. *Analyse de sensibilité et exploration de modèles. Applications aux modèles environnementaux. Collection « Savoir Faire »*, Quae, 2013, 352p. 
- Han L, Da Silva D, Boudon F, Pradal C, Faivre R, Costes E., 2012. Investigating influence of geometrical traits on light interception efficiency of apple trees : a modelling study with MAppleT. In *PMA12 : Fourth International Symposium on Plant Growth Modeling, Simulation, Visualization and Applications*, pp. 152-159. Edited by Guo Y, Kang MZ, Dumont Y. Beijing, China : IEEE Press.
- Lemieux C., 2009. *Monte Carlo and Quasi-Monte Carlo Sampling*. Springer Series in Statistics.
- Monod H., Bouvier A., Kobilinsky A., 2012. Generation of regular factorial designs. R package available on the CRAN.

Afin d'aborder les grands enjeux liés au changement climatique et à la gestion durable des ressources naturelles ou exploitées, des modèles sont développés par les chercheurs en agronomie, écologie, environnement, halieutique, gestion de l'eau, océanographie, etc. Ces modèles intègrent de plus en plus les dynamiques et les processus liés à des systèmes complexes. Pour explorer leurs propriétés et juger de leur pertinence pour assister la décision, il est nécessaire de faire appel à des méthodes d'analyse et d'exploration adaptées. Il est alors souvent fait référence à une grande classe de méthodes, les analyses de sensibilité globales.

Les auteurs de cet ouvrage sont membres pour la plupart du réseau de recherche interinstitutionnel Mexico (Méthodes pour l'exploration informatique de modèles complexes). Forts de leur expérience dans l'organisation d'écoles-chercheurs, ils ont souhaité transférer leur vision globale des différentes méthodes d'analyse de sensibilité et d'exploration ainsi que certaines règles pratiques d'analyse des modèles.

Ce livre s'adresse aux modélisateurs et utilisateurs de modèles qui souhaitent acquérir ou consolider leur maîtrise des méthodes d'analyse et d'exploration de modèles par simulation.

Robert Faivre, docteur en modélisation, calcul scientifique et statistique de l'Université Paris-Sud, Orsay, est directeur de recherche à l'Inra.

Bertrand looss, docteur en géostatistique de l'École des Mines de Paris, habilité à diriger des recherches, est chercheur senior au sein d'EDF R&D.

David Makowski, statisticien, agronome, habilité à diriger les recherches en sciences de la vie (Université Paris-Sud), est directeur de recherche à l'Inra.

Stéphanie Mahévas, docteur en mathématiques appliquées de l'Université Rennes 1, habilitée à diriger des recherches, est chercheur à l'Ifremer.

Hervé Monod, ingénieur agronome, docteur en statistique, est directeur de recherche à l'Inra. Il dirige l'unité MIA du centre de Jouy-en-Josas.

Illustrations de couverture :
Chaos polynomial exploré par un échantillon © Robert Faivre et Claude Bruchou
Surface de réponse, logo du réseau Mexico © Nicolas Dumoulin
L'écosystème marin pour la recherche halieutique © Bénédicte Voile
Résumé graphique d'une analyse de sensibilité © Robert Faivre

éditions
Quæ
Éditions Crad, Premer, Inra, Inrae
www.quae.com

55 €

ISBN : 978-2-7592-1906-3



9 782759 219063

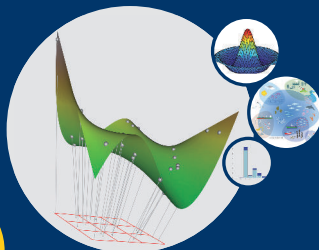
ISSN : 1952-1251
Ref. : 02370Savoir
faireSavoir
faire

Analyse de sensibilité et exploration de modèles

Application aux sciences de la nature
et de l'environnement

R. Faivre, B. looss, S. Mahévas, D. Makowski,
H. Monod, éd.

Analyse de sensibilité et exploration de modèles

éditions
Quæ