

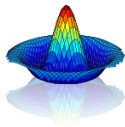
Le fameux PLMM

Polynomial Linear MetaModel

Robert Faivre

INRA - Mathématiques et Informatique Appliquées
MIA Toulouse

Toulouse, le 23 mai 2014



MEXICO
WEXICO

Référence : Exploration par construction de métamodèles

Afin d'aborder les grands enjeux liés au changement climatique et à la gestion durable de 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 la prise en compte de dynamiques et de 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. Forts de leur expérience dans l'organisation d'écoles-chercheurs, les auteurs de cet ouvrage, membres pour la plupart du réseau de recherche interinstitutionnel Mexico (Méthodes pour l'exploration informatique de modèles complexes), ont souhaité transférer par cet ouvrage leur vision globale des différentes méthodes d'analyse de sensibilité et d'exploration et certaines règles d'analyse des modèles développés.

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 Iooss, 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.

éditions
Quæ
Éditions Crad, Ifremer, Inra, Inraea
www.quæ.com

55 €
ISBN : 978-2-7592-1906-3



9 782759 219063
ISSN : 1952-1251
Ra1 : 02370

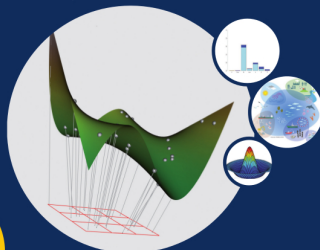
Savoir
faireSavoir
faire

Analyse de sensibilité et exploration de modèles

Application aux sciences de la nature
et de l'environnement

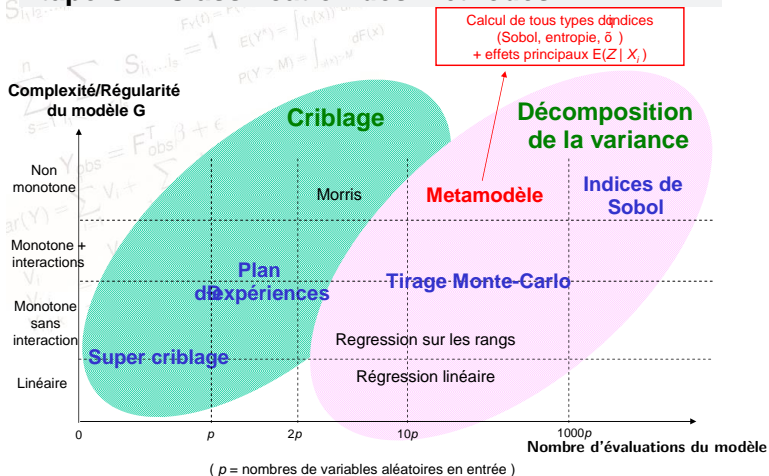
R. Faivre, B. Iooss, S. Mahévas, D. Makowski,
H. Monod, éd.

Analyse de sensibilité et exploration de modèles

éditions
Quæ

Contexte grille

Étape CDE Classification des méthodes



Écriture du modèle

- Métamodèle polynomial de degré 2

$$y = \beta_0 + \sum_{j=1}^K (\beta_j x_j + \beta_{j_2} x_j^2) + \sum_{j=1}^{K-1} \sum_{k=j+1}^K \beta_{jk} x_j x_k + \varepsilon$$

lm (Biomasse ~ polym(Eb,Lmax,B,degree=2))

- Métamodèle polynomial de degré D

$$y = \beta_0 + \sum_{j=1}^K (\beta_{j_1} x_j + \dots + \beta_{j_D} x_j^D) + \sum_{l=1}^{C_{K+D}^D} \beta_l x_j^{l_1} \dots x_j^{l_j} \dots x_j^{l_K} + \varepsilon$$

avec $\sum l_j \leq D$ aux termes diagonaux près

lm (Biomasse ~ polym(Eb,Lmax,B,degree=D))

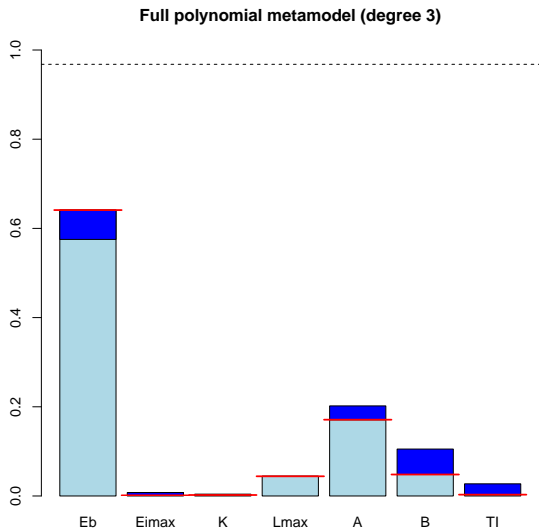
- $\hat{\Theta} = (X'X)^{-1}X'Y$

Calculs pour l'Analyse de Sensibilité

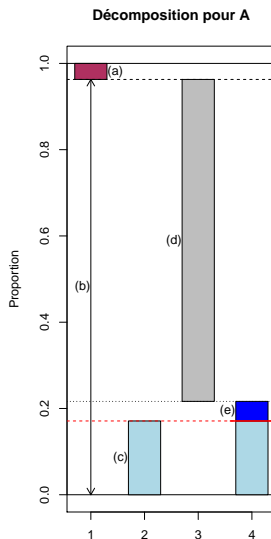
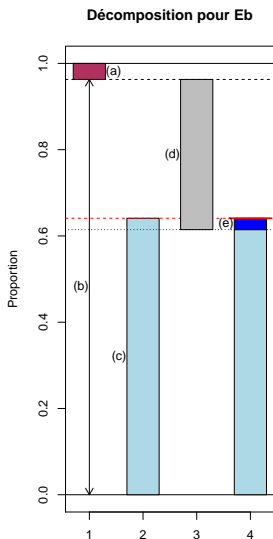
Calculer les contributions des facteurs

- part de variance expliquée par le modèle complet
 $lm(\text{Biomasse} \sim \text{polym}(\text{Eb}, \text{Eimax}, \text{K}, \text{Lmax}, \text{A}, \text{B}, \text{TI}, \text{degree} = 3))$
Multiple R-squared : 0.9679, Adjusted R-squared : 0.9636
- part de variance expliquée par une seule variable
 $lm(\text{Biomasse} \sim \text{poly}(\text{Eb}, \text{degree} = 3))$
Multiple R-squared : 0.6411, Adjusted R-squared : 0.6400
- part de variance expliquée par tous sauf la variable
 $lm(\text{Biomasse} \sim \text{polym}(\text{Eimax}, \text{K}, \text{Lmax}, \text{A}, \text{B}, \text{TI}, \text{degree} = 3))$
Multiple R-squared : 0.3927, Adjusted R-squared : 0.3376

Modèle wwdm et contributions des facteurs



Décomposition des contributions des facteurs



Ajout : Procédure stepwise de choix de modèle

Percentage of factor contributions (multiple R-squared): 96.28389

Polynomial Linear MetaModel of degree 3 (stepwise selection)

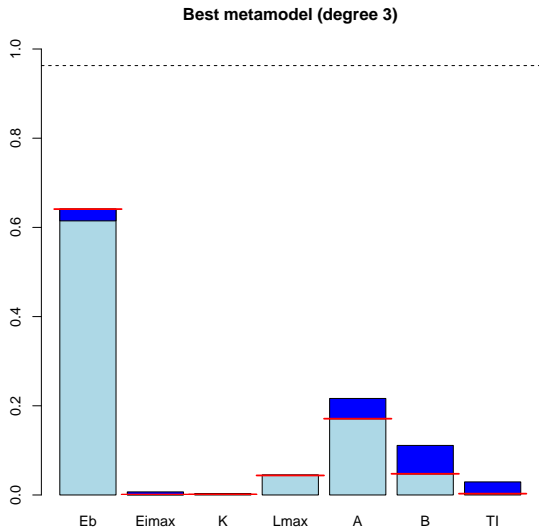
```
Biomasse ~ '1000000' + '0000100' + '0000110' + '0000010' +
'0001000' + '0000200' + '0000101' + '1000100' + '0000210' +
'0000011' + '0100000' + '0002000' + '1000010' + '1000110' +
'0000300' + '1000200' + '1001000' + '0000020' + '0001100' +
'0000111' + '1000101' + '0010000' + '0000201' + '0000002' +
'0011000' + '1100000' + '1002000' + '1000011' + '1110000' +
'0010010' + '0010100', data = expand X
```

Initial standard error of 710.5708 with 999 degrees of freedom

Residual standard error of 139.1544 with 968 degrees of freedom

	First Order SI	Total SI	Interaction
Eb	61.4624564	64.1021987	-2.6397424
Eimax	0.0836550	0.6881091	0.6044541
K	0.1341365	0.3125939	0.1784574
Lmax	4.3793347	4.5039537	0.1246190
A	17.1052521	21.6438003	4.5385482
B	4.7512409	11.1110778	6.3598369
TI	0.2941507	2.9329036	2.6387529

Modèle wwdm et contributions des facteurs



Implémentation

- Polynomes orthogonaux (**polym**) lourds. Utilisation du produit des polynomes orthogonaux individuels (**poly(X_1)*...*poly(X_K)**)
- Procédure stepwise **stepAIC** du package **MASS** en partant du modèle constant.

```
stepAIC(modele, scope = list(lower = lower, upper = upper),
direction = "both", trace = 0)
```

Récupération du meilleur modèle dans l'élément sortie `$best$call`

- Calculs avec les R^2 ajustés
- Quatre structures de sortie : full, best, full.R2adjusted, best.R2adjusted
- Une fonction summary, avec en choix, la structure considérée
- Une fonction plot avec les mêmes options

Intégration dans MTK

```
mtk.analyserAddons(where="newplmm.R", authors="R. Faivre, INRA-MIA Toulouse",
name="PLMM", main="plmm", summary="summary.plmm", plot="plot.plmm")
source("mtkPLMM2Analyser.R")
```

Le fichier newplmm.R contient différentes fonctions (plmm, plmm.mtk, summary.plmm et plot.plmm)

```
plmm <- function(X, Y, degree.pol = 1, rawX = FALSE, numY = 1, listeX = NULL, all = FALSE, which = "best",
lang = "en", digits = options()\$digits, colors = c("red", "orange", "blue"), legend.loc = NULL, ...)
{
if(ncol(Y)>1) { nomY <- names(Y)[numY] ; Y <- data.frame(Y[,numY]) ; names(Y) <- nomY }
if(!is.null(listeX)) X <- X[ , listeX]
DATA <- cbind(X,Y)
resultats <- plmm.mtk(DATA,degpol = degree.pol, raw = rawX
information <- list(AnalysisMethod ="PLMM", nomsX = names(X),nomY = resultats$best$nomY,
degree.pol = degree.pol, rawX = rawX, numY = numY, listeX = listeX, all = all, which = which,
lang = lang, digits = digits, colors = colors, legend.loc = legend.loc)
resultat <- list(main = resultats, information = information)
class(resultat) <- c(class(resultat), "plmm")
resultat
}
```

Exemple avec MTK

```

# Generates the factors
data(WWDM.factors)

# 1) to create a sampler with the Monte-Carlo method
sampler <- mtkNativeDesigner("BasicMonteCarlo", information = list(size=100) )

# 2) to create a simulator with the WWDM model
model <- mtkNativeEvaluator("WWDM", information = list(year=3))

# 3) to create a partial workflow (design and evaluation)
experience <- mtkExpWorkflow(expFactors=WWDM.factors, processesVector=c(design=sampler, evaluate=model) )

run(experience)

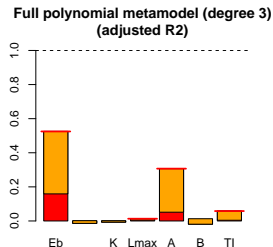
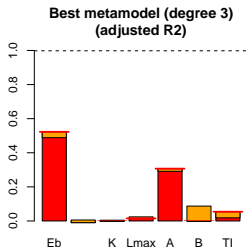
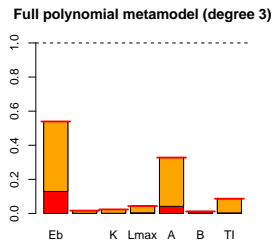
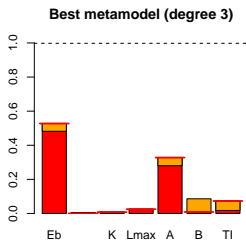
# 4) to create and to add to the workflow the analyser PLMM
analyser <- mtkNativeAnalyser("PLMM", information = list(degree.pol = 3))
addProcess(experience, p = analyser, name = "analyze")
run(experience)

summary(getProcess(experience,name="analyze"))
summary(getProcess(experience,name="analyze"),which="full")
summary(getProcess(experience,name="analyze"), lang="fr", all = TRUE, digits= 3)
extractData(experience,name="analyze")$best$call

plot(getProcess(experience,name="analyze"))
plot(getProcess(experience,name="analyze"),all=TRUE)

```

Modèle wwdm et contributions des facteurs



plmm

R Documentation

The `plmm` method for sensitivity analysis

Description

A `mtk` compliant implementation of the `plmm` method for sensitivity analysis using polynomial linear metamodeling.

Usage

- `mtkplmmAnalyser(listParameters = NULL)`
- `mtkNativeAnalyser(analyze="plmm", information=NULL)`

Parameters

`degree.pol`:

the maximum degree of polynomials (the sum of the degrees of cross products of polynomials is lower or equal to `degree.pol`). See details.

`rawX`:

orthogonal polynomials (default value `FALSE`) or raw polynomials (`TRUE`). See `poly`, `polym`.

`numY`:

the column number of the dependent variable (default is the first column of the dataframe of outputs).

`listeX`:

the column numbers of the dependent variables (default is all the dependent variables).

Parameters for auxiliary functions

`all`:

all the specific summaries and plots are displayed if `TRUE` (default is `FALSE`). Else, see the `which` option.

`which`:

when `all=FALSE`, the name of the specific summary or plot. Options are "best" (default), "full", "best.adjustedR2", "full.adjustedR2". See details.

`lang`:

language of the summary and plot ("en" (default) for english, "fr" for french).

`digits`:

number of digits in the summary (default = `options()$digits`).

`colors`:

colors used in plot (default = `c("red", "orange", "blue")`).

`legend.loc`:

location of the legend in plot (default no legend (`NULL`), options are "topleft", "topright", ... see (`help(legend)`)).

Details

1. The `plmm` metamodeling approach consists in estimating 3 models and comparing the percentage of variance (coefficient of determination) explained by these 3 models. The 3 models are `polym(A,B,C)`, `poly(A)`, `polym(B,C)` where `polym` computes orthogonal polynomials. `polym(A,B,C)` gives the total variance explained by the full metamodel, `poly(A)` gives the variance that can be explained by factor A only (in the sense of polynomials of A) and `polym(B,C)` gives the variance not explained by factor A. Total sensitivity index of factor A is computed as $\frac{\max(R^2(\text{poly}(A)), 1 - R^2(\text{polym}(A,B,C)) - R^2(\text{polym}(B,C)))}{R^2(\text{polym}(A,B,C))}$ where $R^2(M)$ is the coefficient of determination of model M, and first order sensitivity index as $\min(R^2(\text{poly}(A)), 1 - R^2(\text{polym}(A,B,C)) - R^2(\text{polym}(B,C)))$. The `plmm` function computes a best and full model on the basis of stepwise model selection starting with the polynomial model, with different degrees both (see `stepAIC` for more details). Total sensitivity and first order indices are computed in the same way. Additional results are given when using adjusted R^2 for both best and full models.