



Introduction à la calibration de modèles complexes

Stéphanie Mahévas

Unité Ecologie et Modèles pour
l'halieutique, Nantes



- Calibration : que diable est la calibration ?
- Pourquoi calibrer un modèle ?
- Comment calibrer un modèle ?

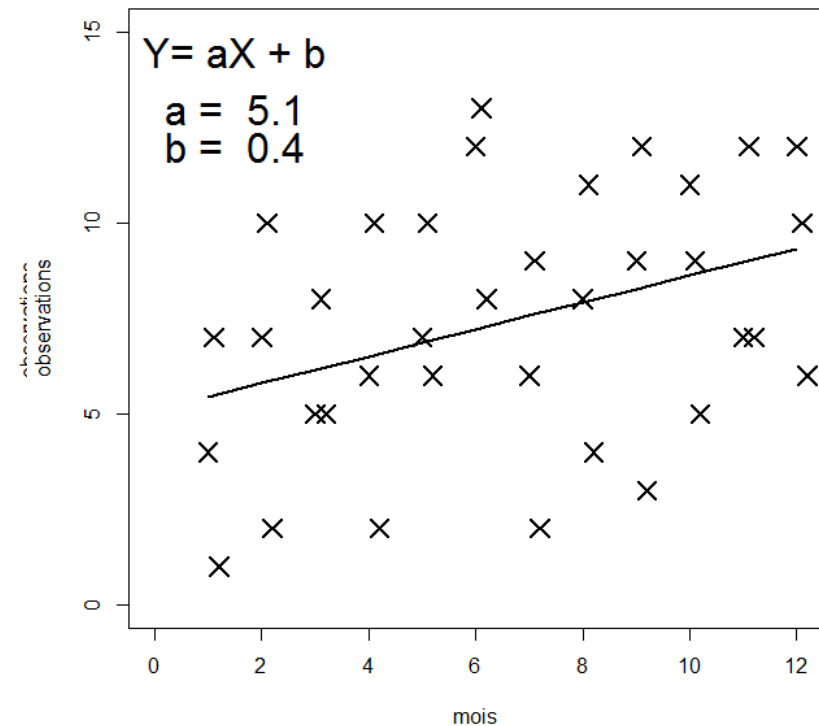
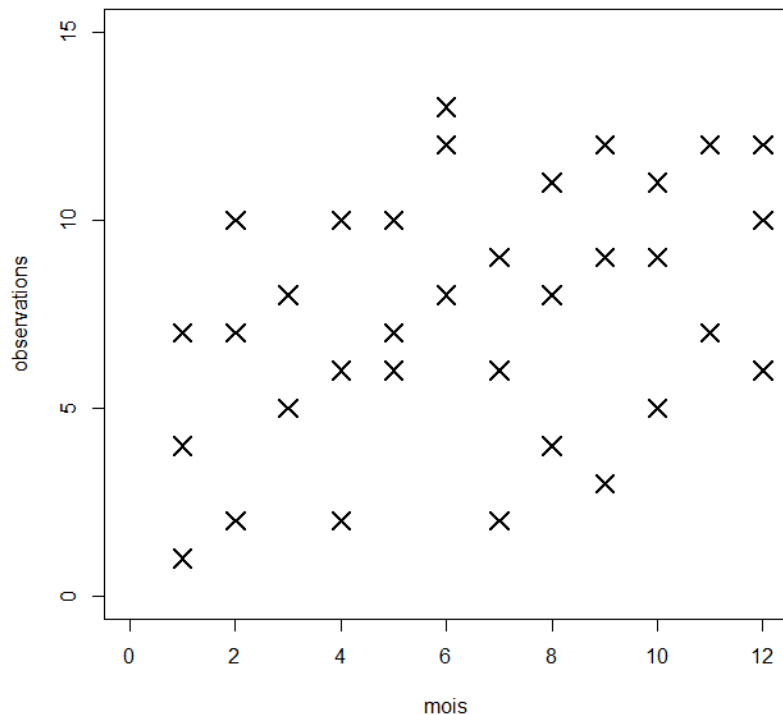
Calibration : que diable est la calibration ?

- En métrologie : étalonnage (calibrage) : comparaison avec un dispositif test qui rend compte avec précision de la mesure standard (étalon)
- En statistiques : méthode d'estimation (méthode inverse) des paramètres - étant donné x et $y=f(x)$, connaissant y , on cherche x ($=f^{-1}(y)$)
- Calibration de modèle : processus d'ajustement des paramètres d'un modèle en intégrant l'incertitude des paramètres et/ou du modèle pour obtenir une représentation du système modélisé qui satisfasse un critère prédéfini

Calibration d'un modèle linéaire (2 paramètres)

Jeu d'observations mensuelles

Quel est le meilleur modèle linéaire à 2 paramètres permettant de reproduire ce jeu d'observations ?



$$Y = 5.1 + 0.35 \text{ mois}$$

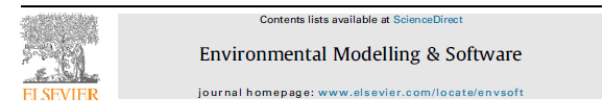
Modèle complexe

- Modèle linéaire à 2 paramètres : résolution analytique – estimation par les moindres carrés
- Modèles complexes
 - Nombreux paramètres
 - Nombreuses sorties
 - Des processus mal connus

Le problème
devient plus
difficile

Pourquoi calibrer un modèle ?

- Estimer des paramètres qui ne peuvent pas ou difficilement être mesurés
- Comprendre le fonctionnement du système étudié
- Crédibiliser/améliorer un modèle pour l'utiliser en décision, prédiction...



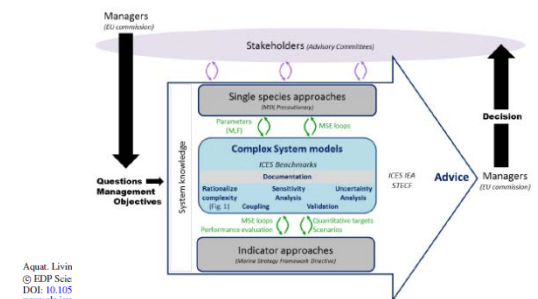
Optimisation as a process for understanding and managing river ecosystems

E.J. Barbour ^{a, h, *}, L. Holz ^c, G. Kuczera ^d, C.A. Pollino ^e, A.J. Jakeman ^a, D.P. Loucks ^f



Selection and validation of a complex fishery model using an uncertainty hierarchy

Sigrid Lehuta ^{a, *}, Pierre Petitgas ^a, Stéphanie Mahévas ^a, Martin Huret ^b, Youen Vermard ^c, Andrés Uriarte ^d, Nicholas R. Record ^e



Aquat. Livin.
© EDP Scie
DOI: 10.105
www.alb-jou

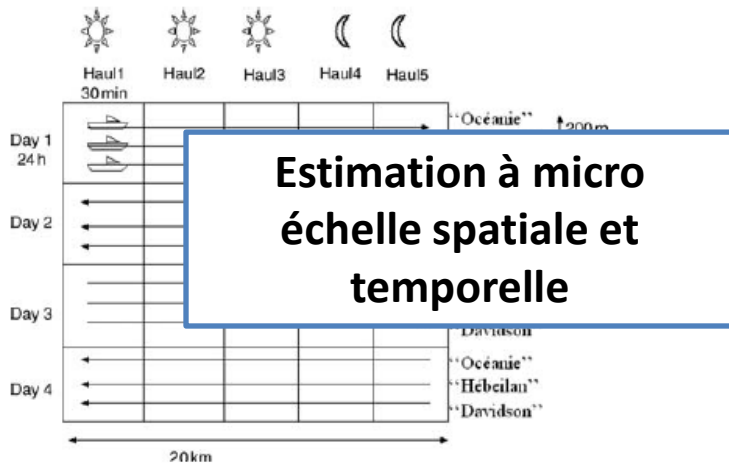
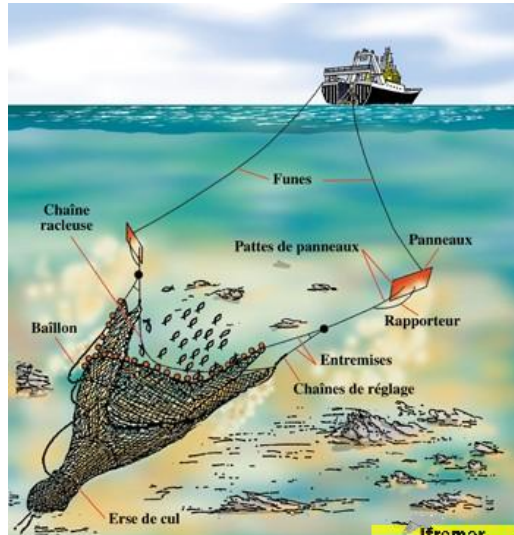
Reconciling complex system models and fisheries advice:
Practical examples and leads

Sigrid LEHUTA ^{1, *}, Raphaël GIRARDIN ², Stéphanie MAHÉVAS ¹, Morgane TRAVERS-TROLET ²
and Youen VERMARD ¹

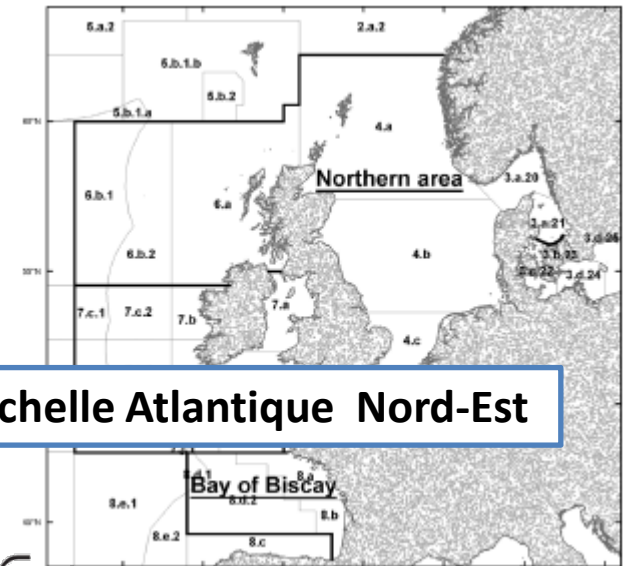
Paramètres difficiles à mesurer

Capturabilité du merlu (*merluccius merluccius*) **q** ?

Captures \sim **q** x Effort x Abondance



Echelle Atlantique Nord-Est



Saison	1	2	3	4
Valeur $q_s = q_s^0$	$1.69 * 10^{-6}$	$8.20 * 10^{-7}$	$7.24 * 10^{-7}$	$7.63 * 10^{-7}$

ICES Journal of Marine Science (2011), 68(1), 107–118. doi:10.1093/icesjms/fsq140

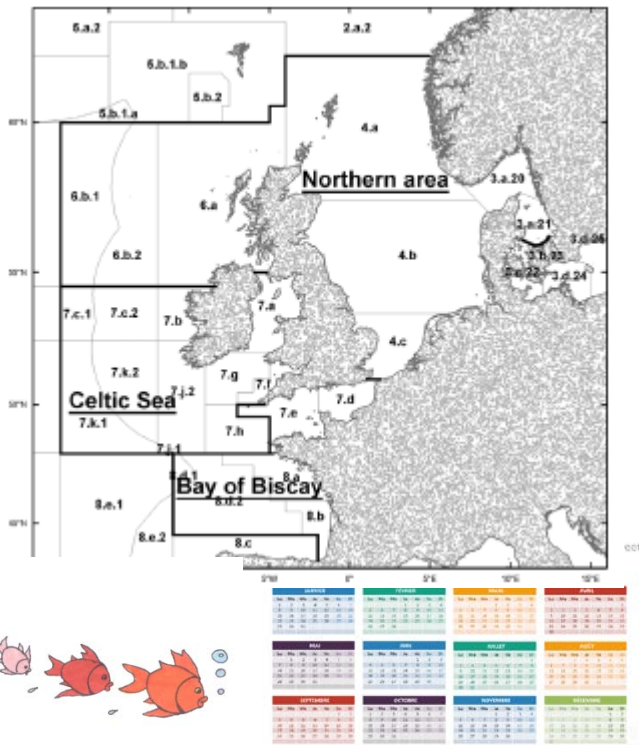
Hake catchability by the French trawler fleet in the Bay of Biscay: estimating technical and biological components

Stéphanie Mahévas*, Verena M. Trenkel, Mathieu Doray, and Arnaud Peyronnet

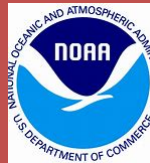
Vigier, A., Bertignac M., Mahévas S. In Prep

Mieux comprendre le système

La dynamique spatio-temporelle de la population de merlu (merluccius merluccius) en Atlantique Nord-Est



Stock Synthesis



Améliorer la qualité des sorties du modèles



AVANT LA
CALIBRATION

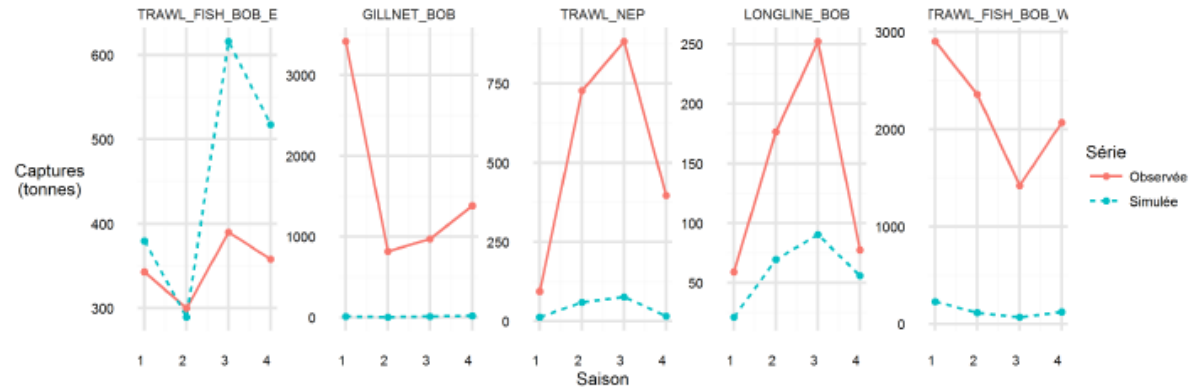


FIGURE 3.5 – Ajustement des captures simulées par ISIS-Fish (trait pointillé) aux captures observées (trait plein) pour 2010 par super-métier (cadran) et saison (abscisse) à l'étape 1 pour $q^p = 0.85$.

APRES LA
CALIBRATION

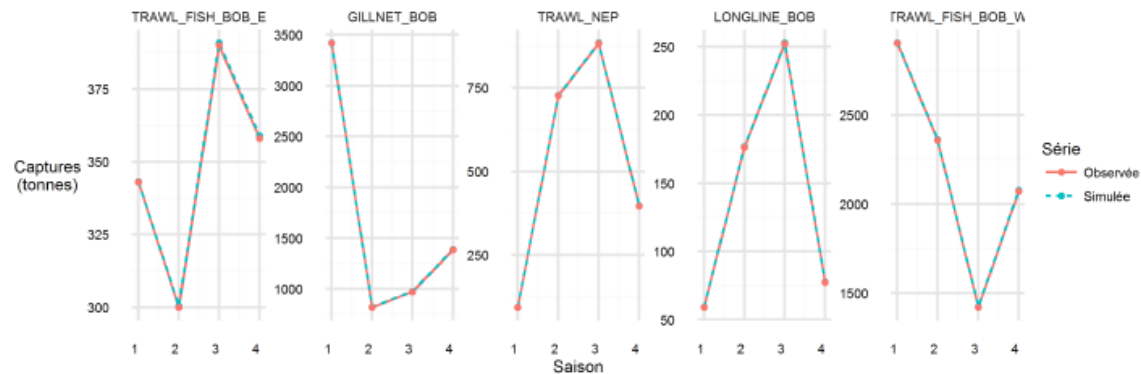
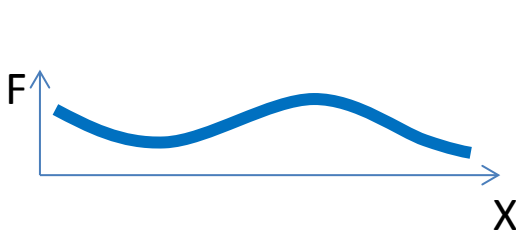


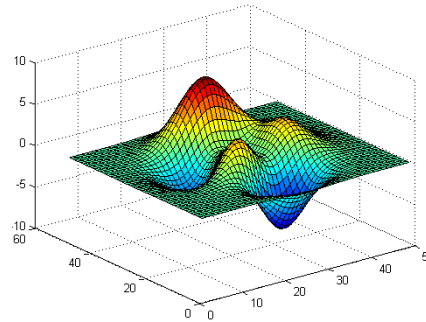
FIGURE 3.7 – Ajustement des captures simulées par ISIS-Fish (trait pointillé) aux captures observées (trait plein) pour 2010 par super-métier (cadran) et saison (abscisse) à l'itération 3 de l'étape 2 pour $Tarf^p = 1$.

Pourquoi la calibration de modèles complexes peut être difficile ?

- **Nombre de paramètres : dimension de l'espace d'exploration très grande**



1 paramètre



2 paramètres

...

Beaucoup de paramètres

- **Temps de simulation : évaluation coûteuse pour une valeur des paramètres**

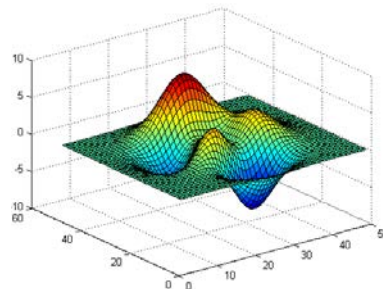
plusieurs minutes

à

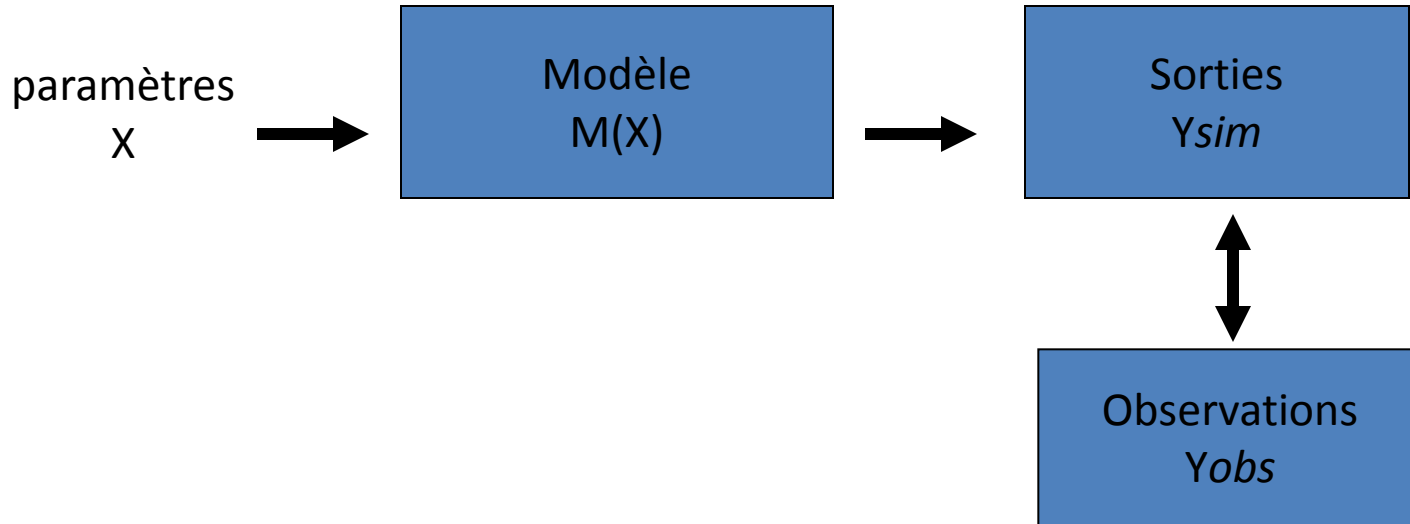
plusieurs heures

- **Multi-modalités**

- **Stochasticité**



Calibration



$$F(X) = \text{dist}(Y_{sim}, Y_{obs}) = \text{dist}(M(X), Y_{obs})$$

X?

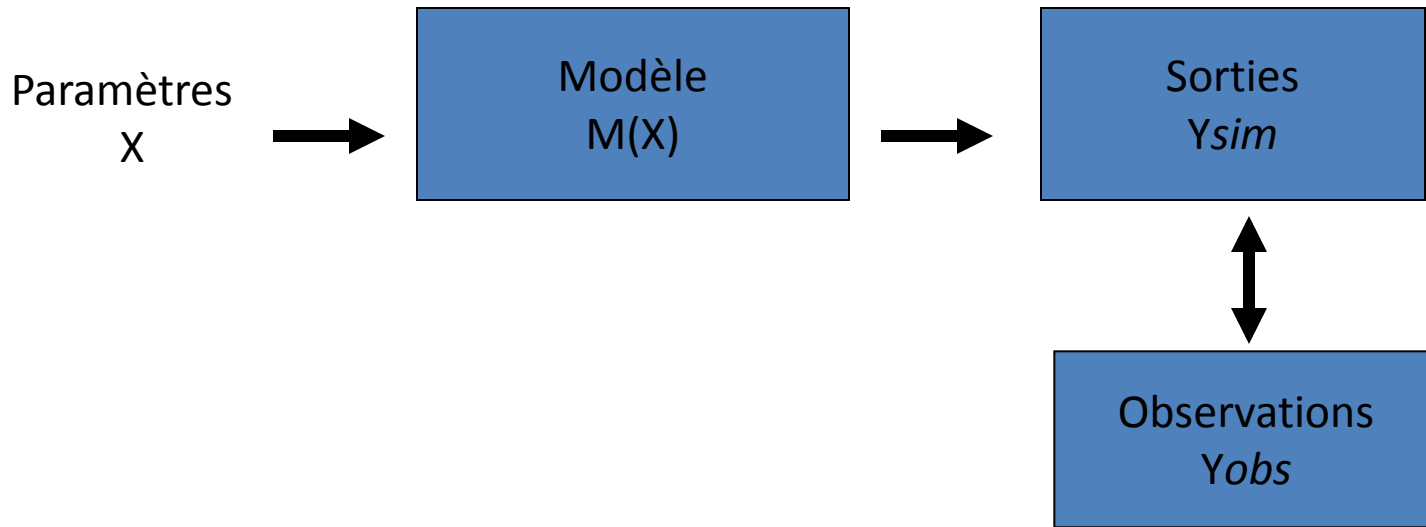
$$X_{opt} = \text{Arg}(\min(\text{dist}(Y_{sim}, Y_{obs})))$$

F non analytique

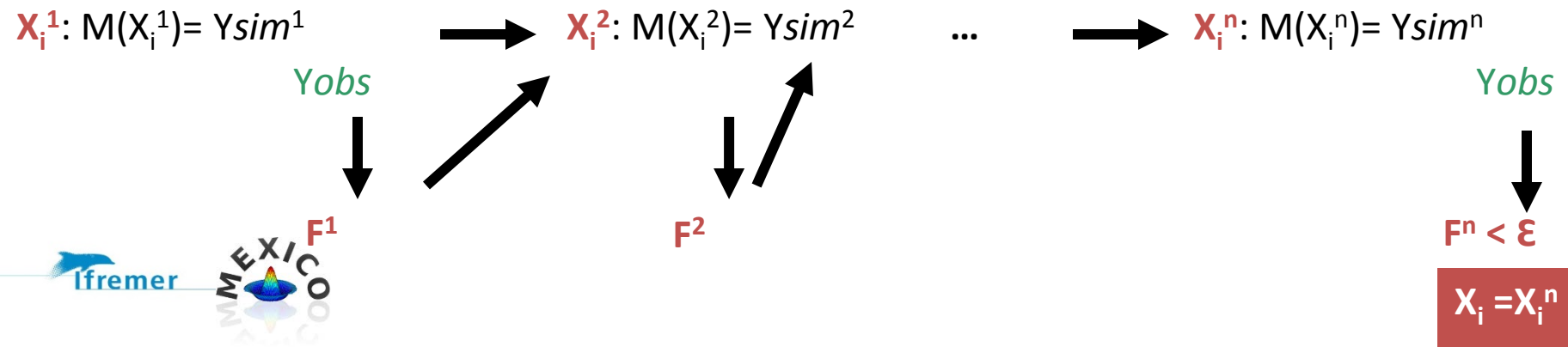
**Non linéaire, multimodale, ...
Coûteuse à évaluer**

**Optimisation – approche numérique
Optimiseur = algorithme itératif**

Calibration



$$X?$$
$$X_{opt} = \text{Arg}(\min(\text{dist}(Y_{sim}, Y_{obs})))$$



CALIBRATION
=
Optimisation d'une fonction mathématique

Pas d'hypothèse de
distribution sur **Y**
Pas d'hypothèse de
distribution sur **X**

Analyse
numérique

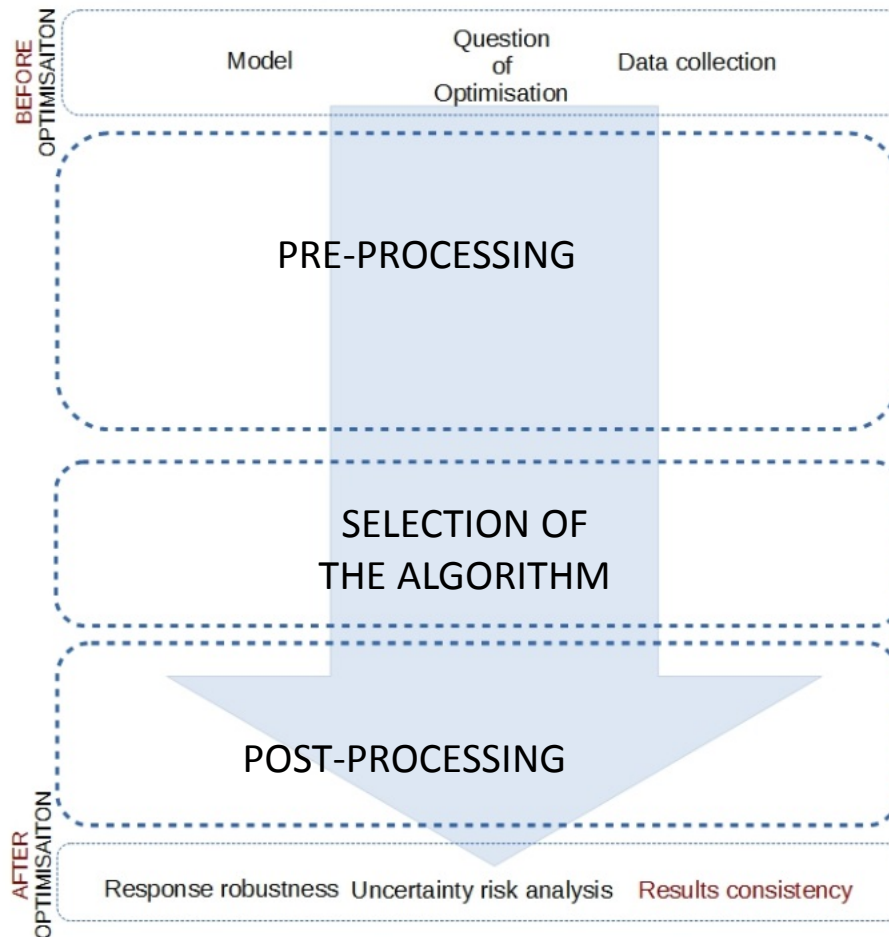
Hypothèse de
distribution sur **Y**
Pas d'hypothèse de
distribution sur **X**

Statistique
Inférentielle

Hypothèse de
distribution sur **Y**
Hypothèse de
distribution sur **X**

Statistique
Bayésienne

Mise en œuvre de la calibration



Pre-processing

Rarement... voire jamais explicité

Les classiques indispensables:

- Bien expliciter la question d'optimisation
capacité prédictive du modèle ? Valeurs des paramètres? compréhension du système?
- Liste des données disponibles: observations, connaissance experte
- Liste de tous les paramètres à estimer : bornes, contraintes, discrètes/continues
- Liste des incertitudes : données, processus (modèle)
- Construire une première fonction d'objectif : point le plus critique?

Les plus communes : moindres carrés, vraisemblance

Les plus spécifiques : statistiques pour l'Approximate Bayesian Computation (ABC) (Fearnhead and Prangle, 2012)

Multi-objectifs : pondérations (Francis, 2011), fronts Pareto (dominance, >4 difficile) (Deb and Sundar, 2006)

Pre-processing

Rarement... voire jamais explicité

Les classiques indispensables:

- Bien expliciter la question d'optimisation
capacité prédictive du modèle ? Valeurs des paramètres? compréhension du système?
- Liste des données disponibles: observations, connaissance experte
- Liste de tous les paramètres à estimer : bornes, contraintes, discrètes/continues
- Liste des incertitudes : données, processus (modèle)
- Construire une première fonction d'objectif : point le plus critique?

$$F(X) = \text{dist}(Y_{\text{sim}}, Y_{\text{obs}}) = \text{dist}(M(X), Y_{\text{obs}})$$

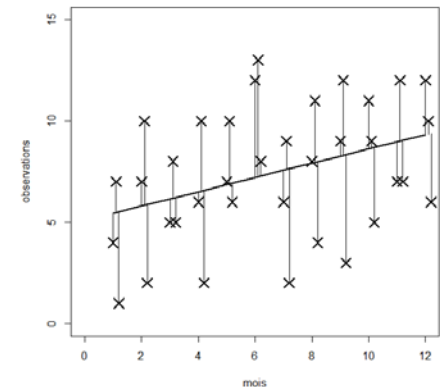
Les plus communes : moindres carrés, vraisemblance

Les plus spécifiques : statistiques pour l'ABC

Multi-objectifs : pondérations, fronts Pareto (dominance, >4 difficult)

$$F(X) = \sum_{k=0}^n (M(X) - Y_{\text{obs}})^2$$

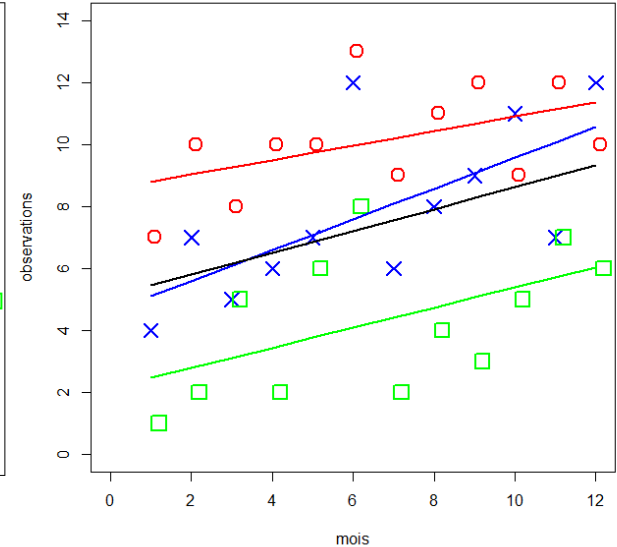
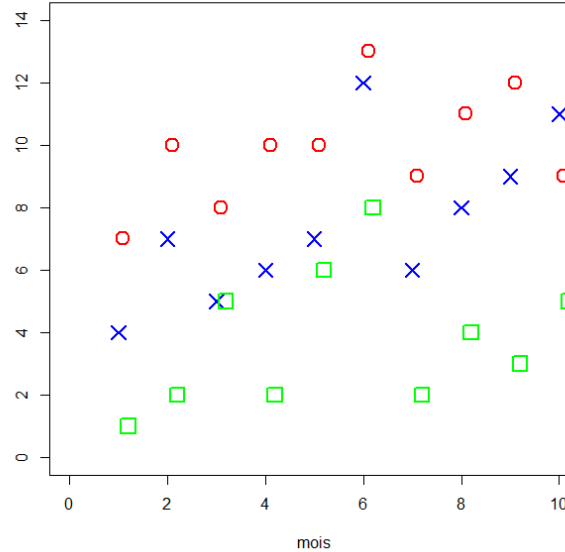
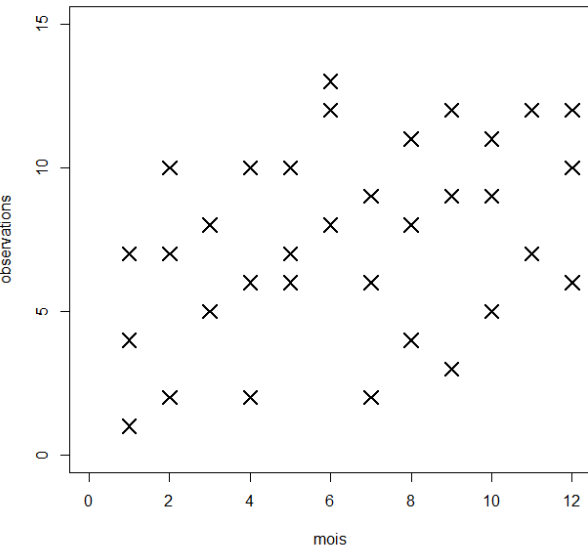
$$F(X) = L(Y = Y_{\text{obs}}|X)$$



Hétérogénéité spatiale

Variabilité temporelle

Jeu d'observations mensuelles



O Zone 1- Année 1

X Zone 2 - Année 1

□ Zone 1- Année 2

$$M(X) \rightarrow M'(X')$$

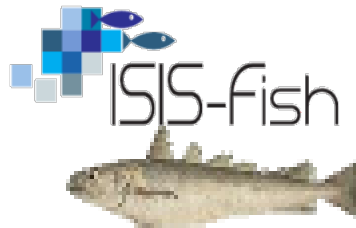
$$F(X') = \alpha F1(X') + \beta F2(X') + \gamma F3(X')$$

$$FO = \sum_{\substack{p \in P \\ s \in S \\ Smet \in SMET}} (\omega_{Smet,p,s}^{LFD} * \alpha * FO_1 + \omega_{Smet,p,s}^{weight} * (\beta * FO_2 + \gamma * FO_3))$$

$$FO_1 = \left(\frac{C_{Smet,l,s,p}^{obs}}{\sum_{l \in L} C_{Smet,l,s,p}^{obs}} - \frac{C_{Smet,l,s,p}^{sim}}{\sum_{l \in L} C_{Smet,l,s,p}^{sim}} \right)^2$$

$$FO_2 = \left(\frac{\sum_{l \in L} C_{Smet,l,s,p}^{obs} - \sum_{l \in L} C_{Smet,l,s,p}^{sim}}{\sum_{l \in L} C_{Smet,l,s,p}^{obs}} \right)^2$$

$$FO_3 = \left(\frac{\sum_{\substack{l \in L \\ s \in year}} C_{Smet,l,s,p}^{obs} - \sum_{\substack{l \in L \\ s \in year}} C_{Smet,l,s,p}^{sim}}{\sum_{\substack{l \in L \\ s \in year}} C_{Smet,l,s,p}^{obs}} \right)^2$$



Pre-processing

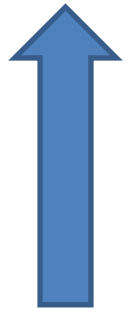
Rarement... voire jamais explicité

Les classiques indispensables:

- Bien expliciter la question d'optimisation
- Liste des données disponibles
- Liste de tous les paramètres à estimer
- Liste des incertitudes : données, processus (modèle)
- Construire une première fonction d'objectif : point le plus critique?

Les indispensables moins classiques:

- Exploration des données et réduction de la dimension
outliers, surdispersion, correlations, etc. et analyse de sensibilité, ACP,...
- Exploration et adaptation de la fonction d'objectif



Quel algorithme ?

Une abondante littérature (très technique surtout pour les numériciens)

Struct Multidisc Optim (2010) 41:219–241
DOI 10.1007/s00158-009-0420-2

RESEARCH PAPER

**Survey of modeling and optimization strategies
to solve high-dimensional design problems
with computationally-expensive black-box functions**

Songqing Shan · G. Gary Wang

HYDROLOGICAL PROCESSES
Hydrol. Process. (2008)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/hyp.7152

**Evaluation of global optimization algorithms for parameter
calibration of a computationally intensive hydrologic
model**

Xuesong Zhang,^{1,2*} Raghavan Srinivasan,¹ Kaiguang Zhao¹ and Mike Van Liew³

J Glob Optim (2013) 56:1247–1293
DOI 10.1007/s10898-012-9951-y

**Derivative-free optimization: a review of algorithms
and comparison of software implementations**

Luis Miguel Rios · Nikolaos V. Sahinidis

J Glob Optim (2009) 45:3–38
DOI 10.1007/s10898-008-9332-8

A review of recent advances in global optimization

C. A. Floudas · C. E. Gounaris

Environmental Modelling & Software 34 (2012) 67–86



ELSEVIER

Contents lists available at SciVerse ScienceDirect

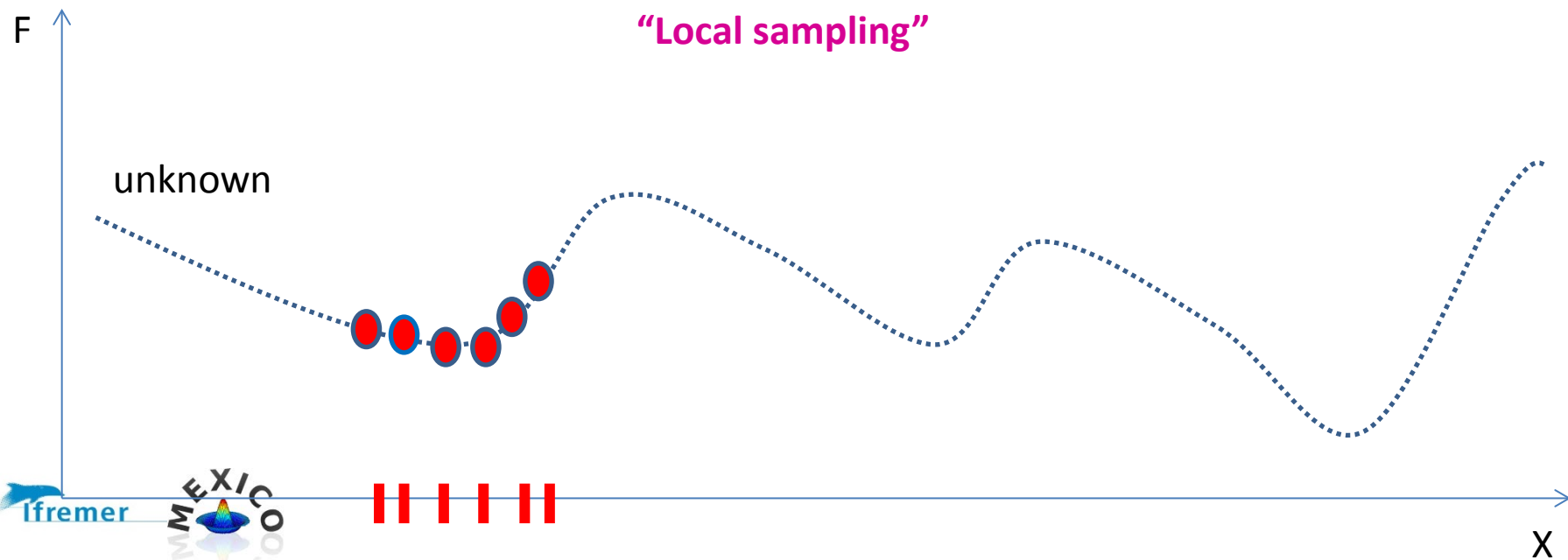
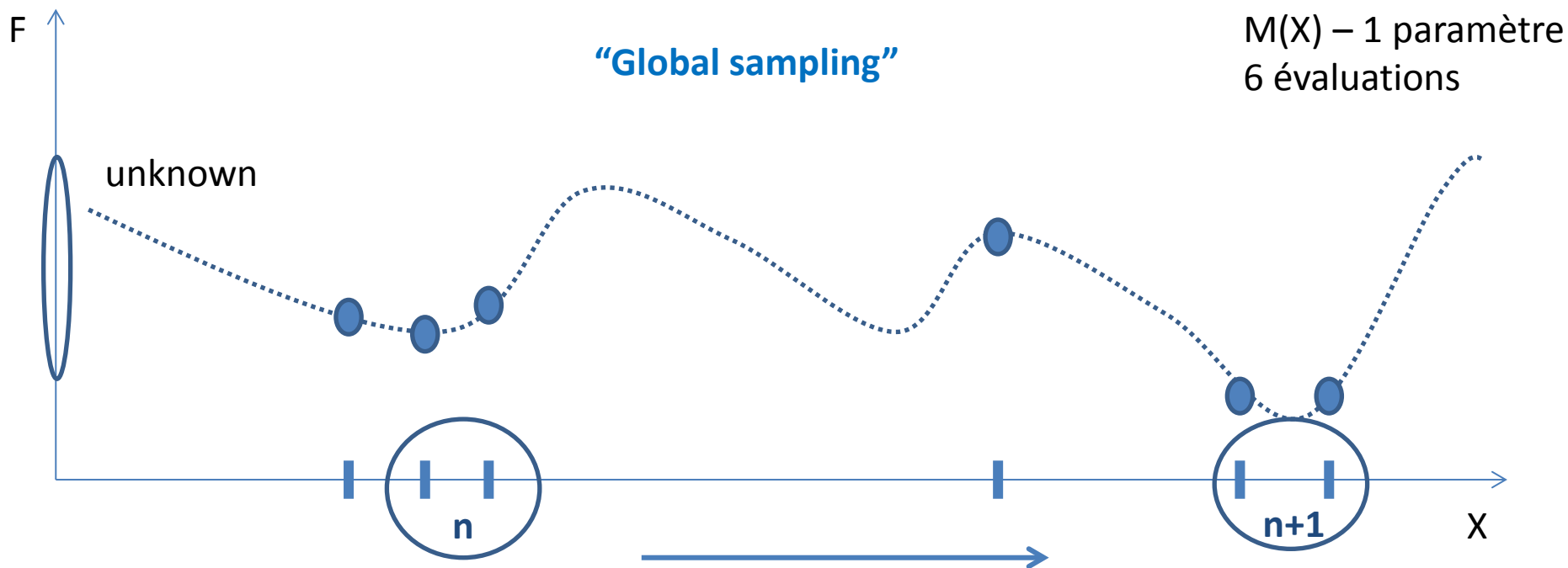
Environmental Modelling & Software

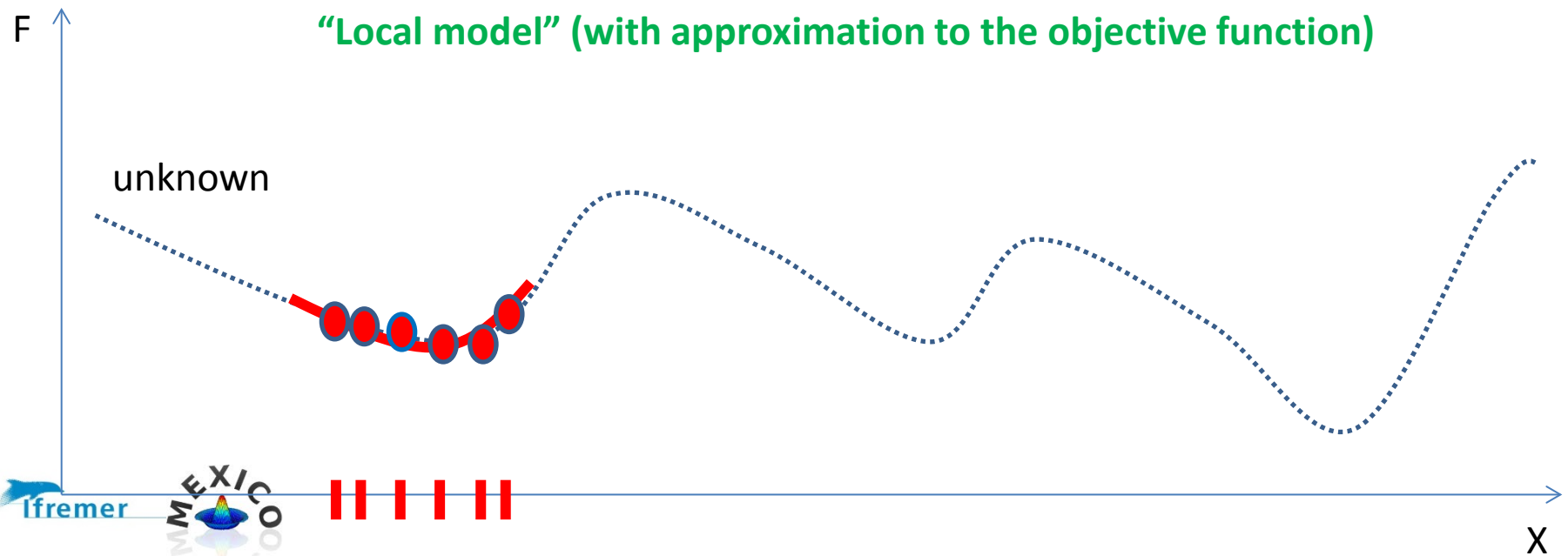
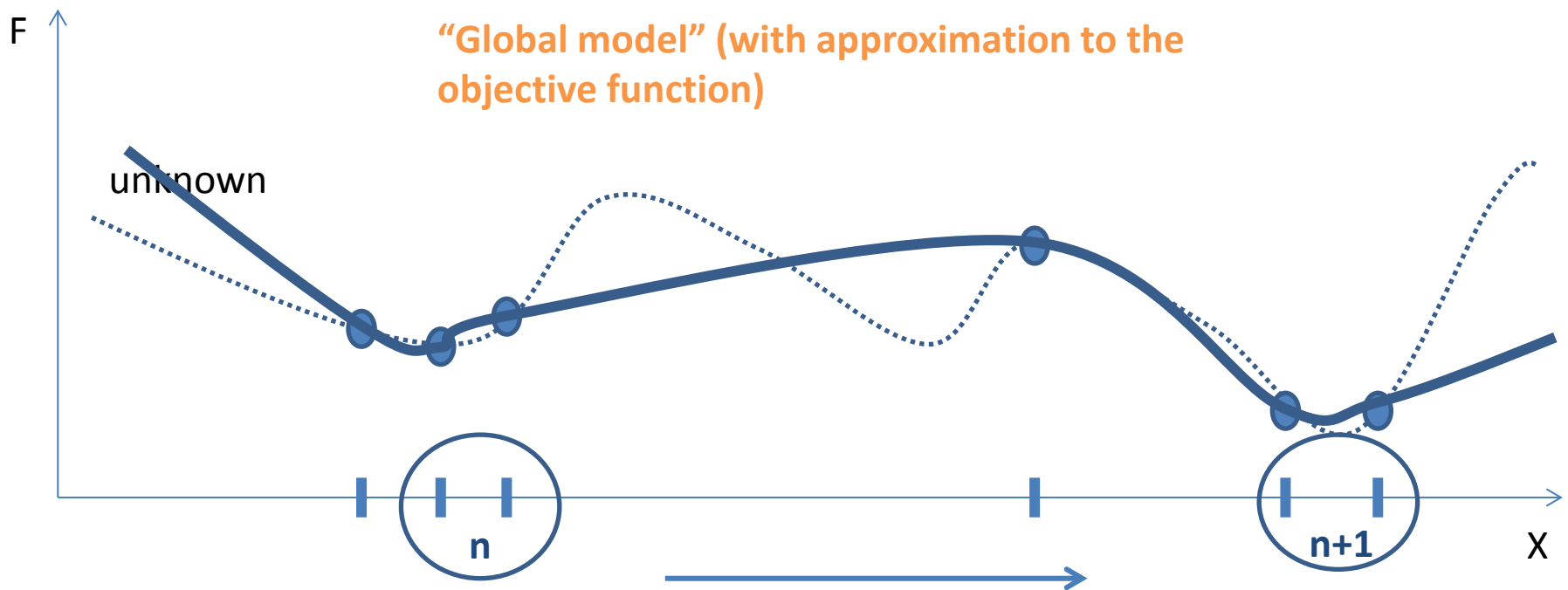
journal homepage: www.elsevier.com/locate/envsoft



**Numerical assessment of metamodeling strategies in computationally intensive
optimization**

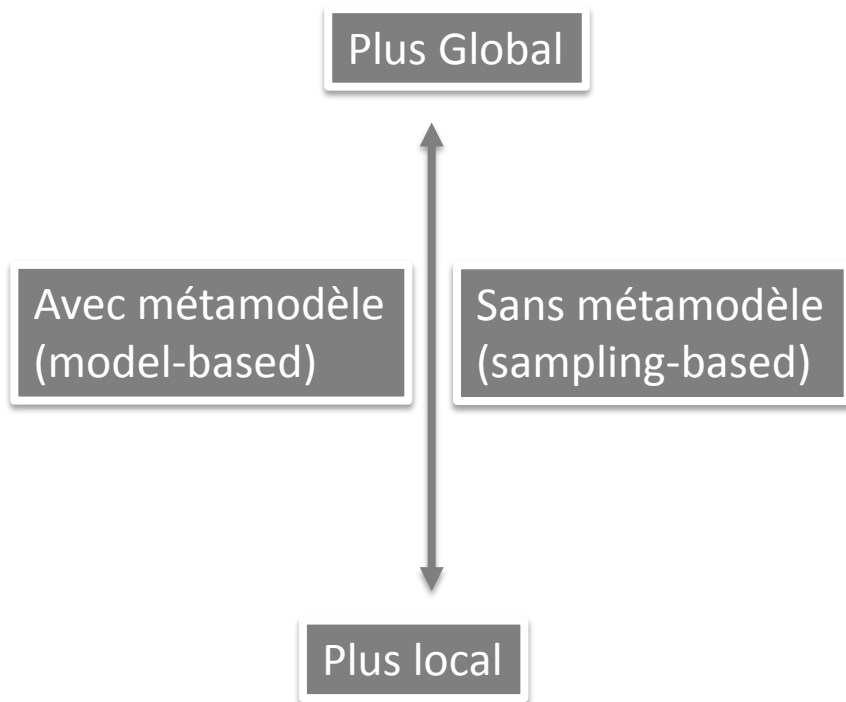
Saman Razavi*, Bryan A. Tolson, Donald H. Burn



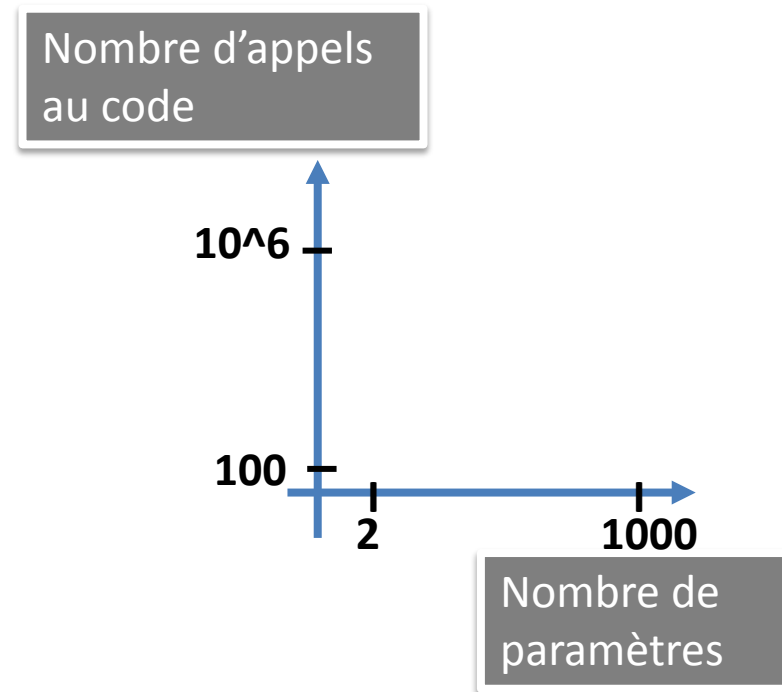


Deux grilles pour guider la sélection

Grille 1 : Espace de projection



Grille 2: Aide à la sélection



- Grille 1 : positionner les différentes familles dans l'espace des deux critères
- Grille 2 : aider au choix de la famille d'optimisation

Deux grilles pour guider la sélection

Grille 1 : Espace de projection

Grille 2: Aide à la sélection

Plus Global

Plus local

number of parameters

- Grille 1 : positionner les différentes familles dans l'espace des deux critères
- Grille 2 : aider au choix de la famille d'optimisation

Post-processing

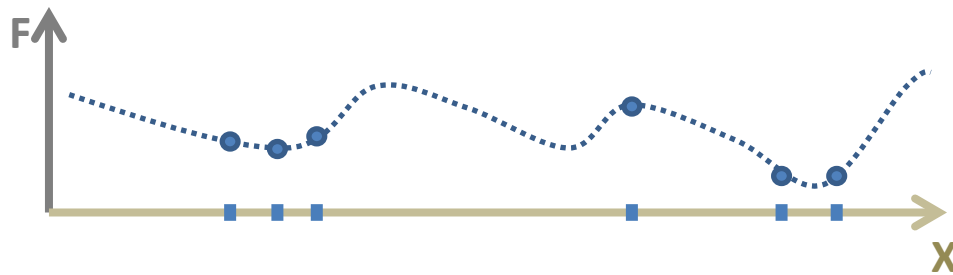
Des tentatives mais peu voire pas d'outils disponibles clés en main

- Évaluer la qualité de l'optimisation
 - convergence : “est-ce que l'on est loin du minimum ?”
 - global/local : “est-ce que l'algorithme a plongé dans creux local?”
 - l'identifiabilité des paramètres : “est-ce que plusieurs solutions donnent le même minimum ?”
- Résoudre le multi-critères
- Stop ou encore ?

Pour tous les algorithmes

A chaque itération, l'algorithme calcule un ensemble de solutions et les valeurs de la fonction d'objectif associées :

la trace de l'algorithme dans l'espace de X et dans l'espace de F



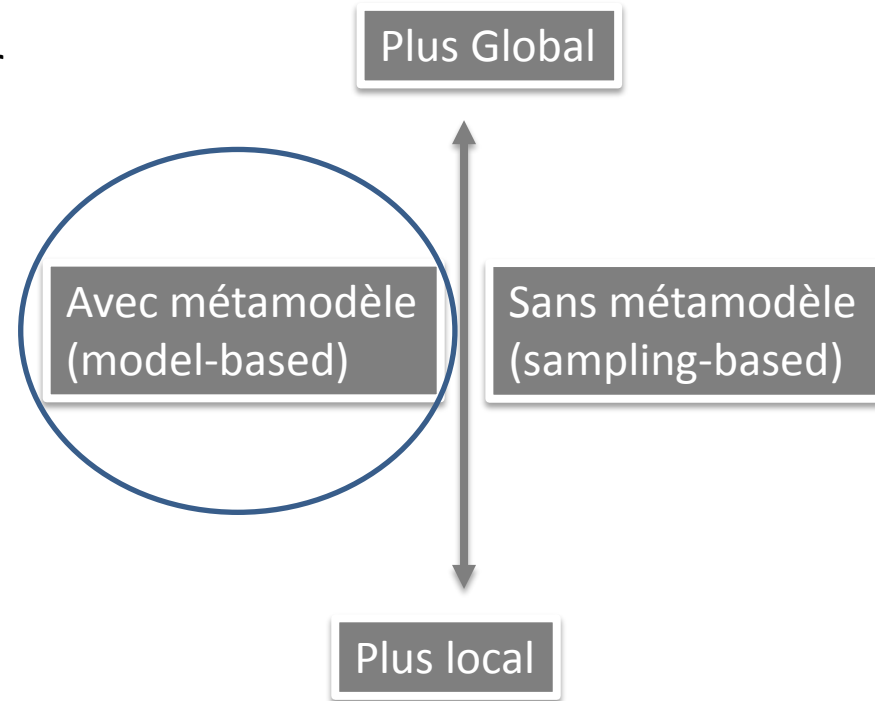
- Sur X : Oscillations , distances entre solution, directions dominantes, fréquences, Sur F : Série des meilleures solutions, (e.g. Maier et al 2014)
- Sensibilité aux points initiaux
- Analyse de sensibilité autour de la solution (e.g. Kleijnen et Sargent 2000)
- Comprendre les propriétés de F : trous, barrières, plateaux, corrélations (e.g. Wright 1932)



Reformulation de la fonction d'objectif
Reparamétrisation
Changement d'algorithmes ou paramètres de contrôle

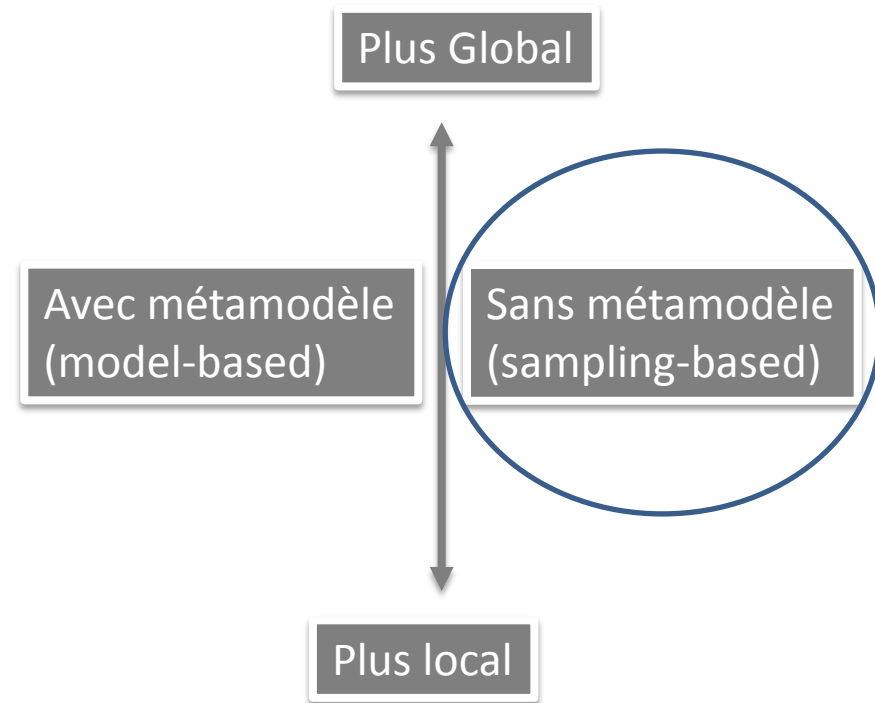
Selon la famille d'algorithmes

- **Algorithmes avec métamodèle:** une approximation de la fonction, des dérivées de premier et second ordre de la fonction autour de l'optimum (optimum ?, identifiabilité ?, intervalles de confiance ?) – Hessian (e.g. Gill et al., 1981)

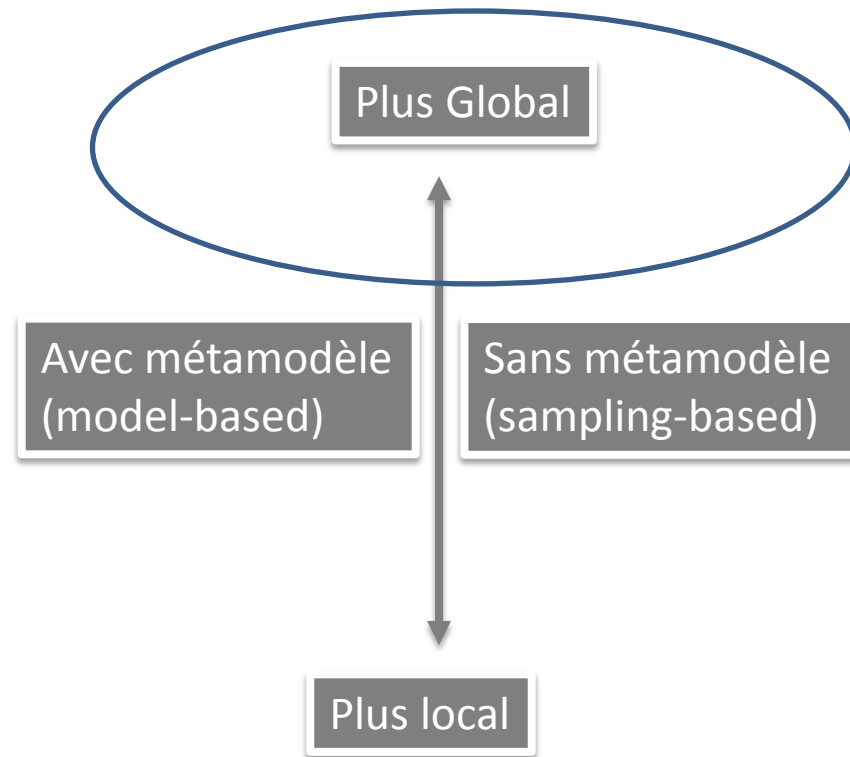


Selon la famille d'algorithmes

- **Algorithmes sans métamodèle :**
donnent un optimum mais aussi une famille de solutions autour de l'optimum (approcher la forme de la fonction d'objectif et des covariances des paramètres, distribution ...) (e.g. Kendall and Nichols 2002)



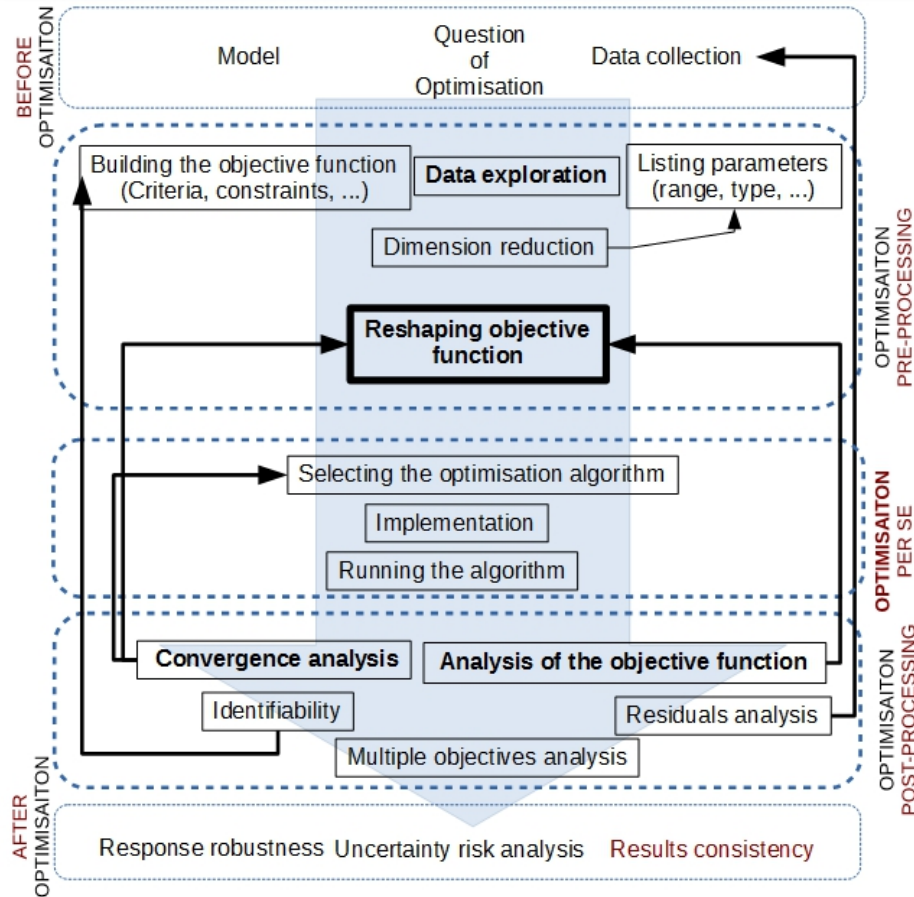
Selon la famille d'algorithmes



- **Algorithmes de recherche globale :**
capturent une forme approchée de la fonction d'objectif sur l'espace des variables (la précision dépendra de l'équilibre entre la phase d'exploration et celle d'intensification)

En résumé

L'optimisation : une démarche pas si linéaire



ODDO : Overview, Design, Details of Optimisation

(ODD Grimm 2010)

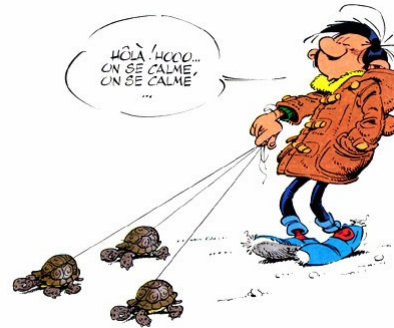
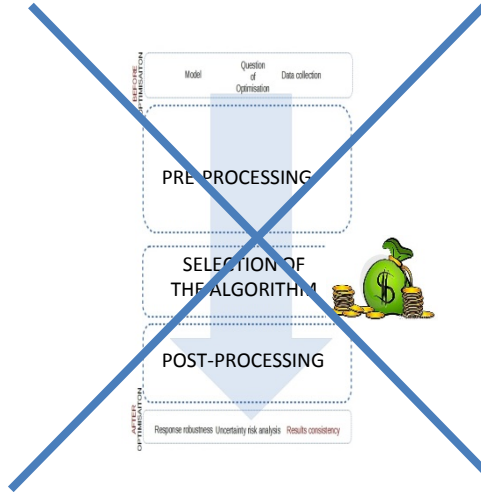
Pre-processing	Problem Formulation	Model	
		Question	
		Data	
		Parameters/variables – Bounds&constraints	
		uncertainty	
	Objective Function	Initial objective function	
		building	
		reshaping	
	Exploratory Analysis	final	
		data	
		Reduction dimension	

Algorithm	Family	
	Description-Justification	
	Adaptation	
	Settings	

Pos-p processing	Convergence	
	Optimum properties	
	Identifiability	
	Residual analysis	
	Multicriteria	

Comments & failures	
---------------------	--

Recommandations



(slow science academy 2010)



Numériciens



Probleme	Model	
	Question	
	Data	
	Collection	
Pre-processing	Pre-processing	
	Selection of the algorithm	
	Post-processing	
	Results	
Algorithm	Building	
	Validation	
	Testing	
	Deployment	
Post-processing	Building	
	Validation	
	Testing	
	Deployment	
Conclusions & Results	Building	
	Validation	
	Testing	
	Deployment	

ODDO en annexe

Remerciements

Victor Picheny, Patrick Lambert, Nicolas Dumoulin, Lauriane Rouan, Jean-Christophe Soulié, Hilaire Drouineau, Rodolphe Leriche, Robert Faivre, Sidrid Lehuta, Dimo Brockoff

Références

- Bolker t al. « Strategies for Fitting Nonlinear Ecological Models in R, AD Model Builder, and BUGS ». *Methods in Ecology and Evolution* 4, no 6 (2013): 501–512. doi:10.1111/2041-210X.12044.
- Fearnhead, Paul, et Dennis Prangle. « Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation ». *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 74, no 3 (2012): 419–474.
- Gill, Philip E., Walter Murray, et Margaret H. Wright. *Practical optimization*. Academic press, 1981.
- Gimenez, O, A Viallefont, E A Catchpole, R Choquet, et B J T Morgan. « Methods for investigating parameter redundancy ». *Animal Biodiversity and Conservation* 27, no 1 (2004): 561–572.
- Grimm V, Berger U, DeAngelis D L, Polhill J G, Giske J and Railsback S F (2010). The ODD protocol: A review and first update. *Ecological Modelling* 221 (23), 2760-2768. [doi:10.1016/j.ecolmodel.2010.08.019]
- Kendall, William L., et James D. Nichols. « Estimating State-Transition Probabilities for Unobservable States Using Capture-Recapture/Resighting Data ». *Ecology* 83, no 12 (2002): 3276–3284. doi:10.2307/3072078.
- Kleijnen, Jack P. C., et Robert G. Sargent. « A methodology for fitting and validating metamodels in simulation1 ». *European Journal of Operational Research* 120, no 1 (janvier 2000): 14–29. doi:10.1016/S0377-2217(98)00392-0.
- Maier, H. R., Zoran Kapelan, J. Kasprzyk, J. Kollat, L. S. Matott, M. C. Cunha, G. C. Dandy, et al. « Evolutionary algorithms and other metaheuristics in water resources: current status, research challenges and future directions ». *Environmental Modelling & Software* 62 (2014): 271–299.