Calage de modèles, point de vue math-info

Arthur, Peter, Vincent

MOISE, STEEP, SET

Introduction

- L'instantiation d'un modèle numérique requiert en général la détermination de certains paramètres du modèle (calage).
  Cela peut être effectué :
  - "manuellement"
  - en adoptant les valeurs des paramètres d'une autre instance du modèle
  - par estimation "automatique"
  - par une combinaison de ces méthodes
- Souvent, on applique une procédure itérative :
  - évaluation de la qualité des paramètres courants :
    - évaluation "à l'œil"
    - vérification si les paramètres ont des valeurs expectées, respectent des contraintes d'ordre, etc.
    - on fait tourner le modèle sur des données tests et on évalue sa sortie
    - combinaison de ces méthodes
  - selon le résultat de cette évaluation, on met à jour les paramètres
  - on recommence

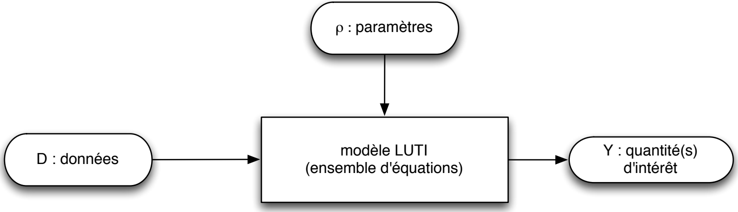
Introduction

- Point de vue math-info :
  - Idéalement, on formalise au maximum possible l'évaluation de la qualité des paramètres, *via* :
    - une fonction objectif qui "mesure" leur qualité
    - des fonctions exprimant des contraintes sur les paramètres : bornes sur leurs valeurs, ordonnancement de différents paramètres, etc.
  - Possibilités pour la mise en place de la fonction objectif :
    - mesure du décalage entre sorties du modèle et des observations
    - cette mesure peut être une simple différence, mais idéalement elle se base sur une connaissance de l'incertitude sur les observations
    - toute autre fonction qui décrit la qualité ou la "plausibilité" des paramètres (si pas possible de formuler cela sous forme de contraintes)

  - Une fois qu'on a formalisé comment évaluer la qualité des paramètres, on les détermine par optimisation numérique

Optimisation

Soit le modèle $Y = f(D, \rho)$



Optimisation

Cadre général

En supposant qu'on ait défini une fonction objectif

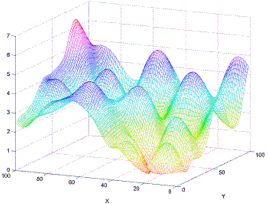
$$g(\rho) : \mathbb{R}^N \mapsto \mathbb{R}$$

Le but est de trouver les valeurs de ρ qui minimisent cette fonction, sous les contraintes suivantes :

- Contraintes d'égalité : $c_i(\rho) = 0$
- Contraintes d'inégalité : $c_j(\rho) < 0$

Les points difficiles

- La non linéarité (plusieurs minima)
- La grande dimension
- Etc.



Optimisation

- Caractéristiques de problèmes ou méthodes d'optimisation :
  - problèmes contraints ou non contraints ;
  - nombre de paramètres et de contraintes ;
  - problèmes avec fonction objectif linéaire, quadratique, non-linéaire ;
  - contraintes linéaires, quadratiques, non-linéaires ;
  - contraintes d'égalité ou d'inégalité (ou les deux) ;
  - type des paramètres : réels, entiers, permutations d'entiers, ...
  - méthodes nécessitant le calcul des dérivées de la fonction de coût ou pas ;
  - méthodes globales vs. méthodes locales ;
  - méthodes déterministes vs. méthodes stochastiques.

Métholodogie

- Ingrédients d'une méthodologie d'estimation de paramètres :
  - Fonction objectif
  - Contraintes
  - Données : observations, valeurs par défaut de certains paramètres, etc.
  - Une ou des méthodes d'optimisation
  - Une ou des stratégies, par exemple :
    - découpage du problème, optimisation de certains paramètres puis des autres, etc.
    - détermination des paramètres d'un modèle simplifé (e.g. moindre résolution spatiale, temporelle, descriptive), puis du modèle complet
    - agencement d'étapes automatiques et manuelles
    - etc.

Optimisation

- Problèmes potentiels :
  - Même un calage “parfait” peut donner lieu à des prévisions fausses :
    - modèle inadéquat (ne correspond pas à la réalité du phénomène)
    - données incertaines, trop peu de données
  - $\rightarrow$ Nécessité de valider
  - La grande dimension ... en entrée
  - ... mais aussi en sortie $\rightarrow$ difficulté d'analyse des résultats
  - Le manque de données
  - Pas de forme analytique pour g (e.g. si l'évaluation de g requiert qu'un processus dynamique sous-jacent ait convergé)
  - Modèle f stochastique
  - Contrainte pratique : temps de calcul

Schéma global

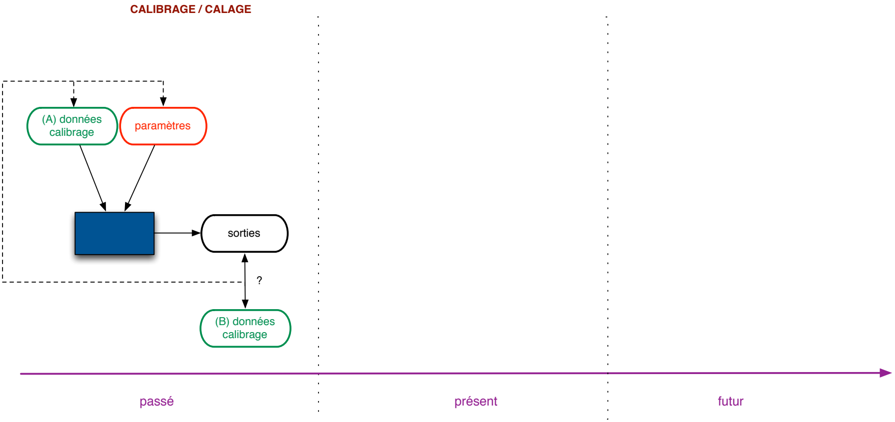


Schéma global

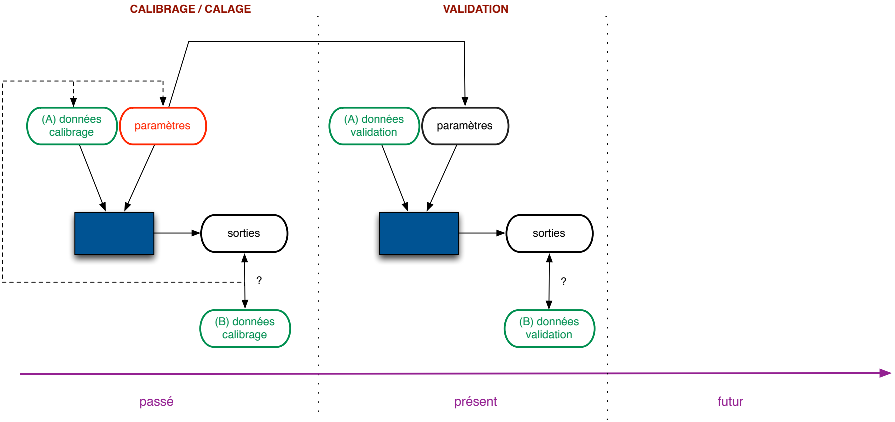


Schéma global