

Plug N Check module 1 : Etat et perspectives

AELOS/ATLANMOD LS2N - UMR CNRS 6004
{Firstname.Lastname}@univ-nantes.fr

3 octobre 2019

Plateforme PlugNCheck

C'est le contexte, pas le contenu ! Mise à disposition d'outils
Ici, généralisation de l'outil de Benoît qui parcourt un espace de paramètres d'un modèle biologique spécifique (croissance des méduses).

Description rapide

- Le modèle dépend de paramètres réels bornés.
- L'exécution d'une simulation pour de multiples valeurs des paramètres
- Comparaison à des données expérimentales (in situ et labo) pour évaluer l'adéquation des paramètres.

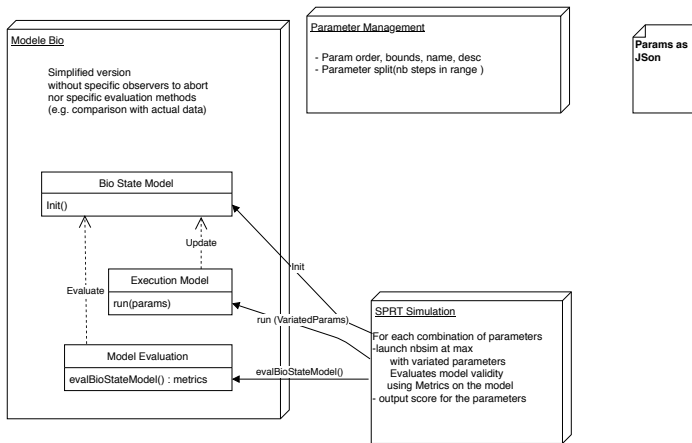
Espace de paramètres

Paralléloptope : N paramètres réels, ça fait beaucoup de valeurs possibles

- découpage en plages de valeurs, on passe d'une infinité de valeurs à un nombre fini de plages
- du coup on a découpé l'espace en $\prod_{i=1}^N nbplages_i$ "cases"
- méthode de monte carlo et SPRT : tirage au hasard dans ces plages de valeurs, en fixant un risque alpha, beta et un nombre de simulations par case pour obtenir une précision "suffisante"

Abstraction

Si on généralise , on isole le modèle, la gestion des parametres et celle de la méthode de simulation :



Parcours

On veut en plus :

- interrompre et reprendre, varier les parcours.

Préalable : discrétisation puis numérotation des cases.

- Un ordre sur les params, puis une construction en base variable.
- On peut traiter les cases comme des entiers ou des points, voire des hypervoxels

Inconvénients du parcours de base

Principe : on parcourt toutes les cases.

On fait un produit cartésien

problème :

- on s'ennuie vite
- les valeurs aux extrêmes de range sont pas forcément les plus intéressantes
on prévoit souvent un peu large pour pas louper des trucs
- ça revient à fixer tous les params sauf 1 au minimum, le faire varier, puis augmenter un 2eme param, recommencer ...

Parcours 1 : Modifier l'ordre

On peut changer l'ordre pour le parcours pour prioriser les dimensions intéressantes, ou les paramètres qu'on juge les plus sensibles.

Version simple :

- on décide d'un nouvel ordre
- on crée une liste de mapping du nouvel ordre vers l'ordre initial (du programme de simulation),
- on l'applique au produit cartésien généré à partir de nouvel ordre.

Parcours 2 : Random

Aléatoire, mais sans répétition ni oubli, c'est un parcours, pas un échantillonnage !
là, ça devient couteux en espace :

- on construit un tableau des nbcases premiers entiers et on le shuffle $O(n)$
- sur le cas bio centaines de milliers de cases
- en python c'est infame niveau mémoire, on est sur du *6 ou *8 par rapport à du C pour cet usage
- mais bon on est censé distribuer

Construction vs Génération

Parcours 3 : Radial

On part d'un point et on étend au voisinage de plus en plus lointain.

- On doit utiliser une distance appropriée (Tchebychev/ chessboard - voisinage de Moore) et respecter les bornes
(une distance euclidienne laisse des trous et même si on les comble on n'a pas de belle partitions)
- pour permettre un découpage en parallélotopes, qui correspondent bien à des sous ensembles de plages
d'ou assemblage, voire cohabitation avec des granularités distinctes.

Actuel : construction de liste et tri , mais passer à une génération de cases, moins gourmand en mémoire

Parcours 4 : Précision progressive ?

L'objectif : Construire des tendances et affiner progressivement

- partir des extremes, le milieu, puis les quarts ...
- Pour chaque range de valeurs, construire un mapping vers ce parcours
- Faire un parcours radial partant de 0.0 et transformer les coordonnées grace au mapping

Distribution

- split de range , de regions ? quelles stratégies raisonnables de partitionnement ?
- Comment distribuer/interrompre/reprendre/changer de parcours sans redites ni oublis.
- Les coûts d'une blacklist distribuée/vs indices avec des générateurs de liste (lookahead pour prendre en compte la progression)

Et remontée/centralisation des infos

Visualisation

- progression (% complete par case), distribution
- résultats, partiels (tendances), requetes sur régions ?

Diagnostics ? Une case trop grosse ? Variance maximale comme indicateur, avec dump des valeurs pour tenter de déterminer la ou les dimensions à découper ?

Plan

