		 Département de mathématiques et Informatique appliquées 	
<b>Proposition d'une feuille de route pour la mise en oeuvre d'une Boite A Outil Mexico (BAOM)</b>					
Réf.	propositionDeRealisationBAOM				
Projet	<b>Méthodes pour l'EXploration Informatique des modèles COMplexes</b>				
Auteur(s)					
Claude	Bruchou	INRA – MIA Avignon			
Robert	Faivre	INRA – BIA toulouse			
Hervé	Monod	INRA -MIA Jouy			
Hervé	Richard	INRA – MIA Avignon			
Juhui	Wang	INRA – MIA -Jouy			
Date de création : 16 janvier 09			Date de la dernière Modification : 21 Avril 2009		
Statut du document					
Version	Date	Etat	Description de la modification		
0.00	16 janv. 09	En cours	Premier jet		
0.02	02 Fév. 09	En cours	Seconde révision		
0.03	13Fév. 09	Fixe	Version diffusée dans MEXICO pour proposition		
0.04	21 Avr. 09	MaJ	Intégration de l'étude de cas ISIS (IFREMER)		
0.05	22 mai 09	MaJ	Intégration de l'étude de cas Ecomeristem (CIRAD BIOS)		

## 1 Contexte et objectifs

Après trois ans d'existence du réseau MEXICO, la réunion du mois de décembre 2008 à Montpellier a été l'occasion de faire le point des projets informatiques en cours de développement. Après la définition collective des schémas XML et d'une architecture générale du système informatique, la dynamique s'est poursuivie et enrichie de succès d'une part, de nouveaux questionnements d'autre part. A titre d'exemple, on peut citer les travaux réalisés autour de ISIS-Fish qui montrent l'intérêt d'avoir un outil générique indépendant du modèle de simulation, mais aussi les difficultés que cela oblige à résoudre; les bibliothèques R réalisées autour de la plateforme VLE qui démontrent la possibilité d'utiliser R comme environnement de développement dans le cadre du projet MEXICO; les différents projets de thèse ou post-doc engagés à MIA ou au Cemagref qui offrent une vision plus large des méthodes pouvant être exploitées dans le cadre du projet MEXICO; la réflexion sur le « reporting » engagée dans le cadre de la plateforme « SimExplorer » qui permet d'aller au-delà de l'analyse statistique proprement dite.

Ces avancées s'appuient de façon significative sur les échanges du réseau Mexico en général et de son groupe informatique en particulier. Cependant une difficulté majeure du fonctionnement de ce réseau réside dans la nécessité de conserver un degré de généricité important afin que chaque

équipe puisse participer facilement à la réflexion et à la réalisation, assurer l'intégration de ces avancées dans son propre projet sans avoir besoin de produire du code « projet dépendant ». Pour redynamiser la lancée initiale, nous pensons que le projet devrait reprendre et afficher résolument l'objectif de développement et de mise en œuvre d'outils informatiques, qui soient d'une part indépendants et accessibles par des personnes externes au groupe, d'autre part utiles pour les projets propres aux différents membres du réseau qu'il s'agisse de ISIS-Fish, de Sim-Explorer, de VLE ou de Record.

Notre démarche est motivée en particulier par le souhait de

- 1) valoriser le travail collectif déjà accompli par le groupe informatique de Mexico (schéma XML et code R);
- 2) réactiver le groupe, homogénéiser et généraliser ses acquis ;
- 3) passer à une phase plus active d'intégration de méthodes.

Nous proposons trois objectifs opérationnels principaux :

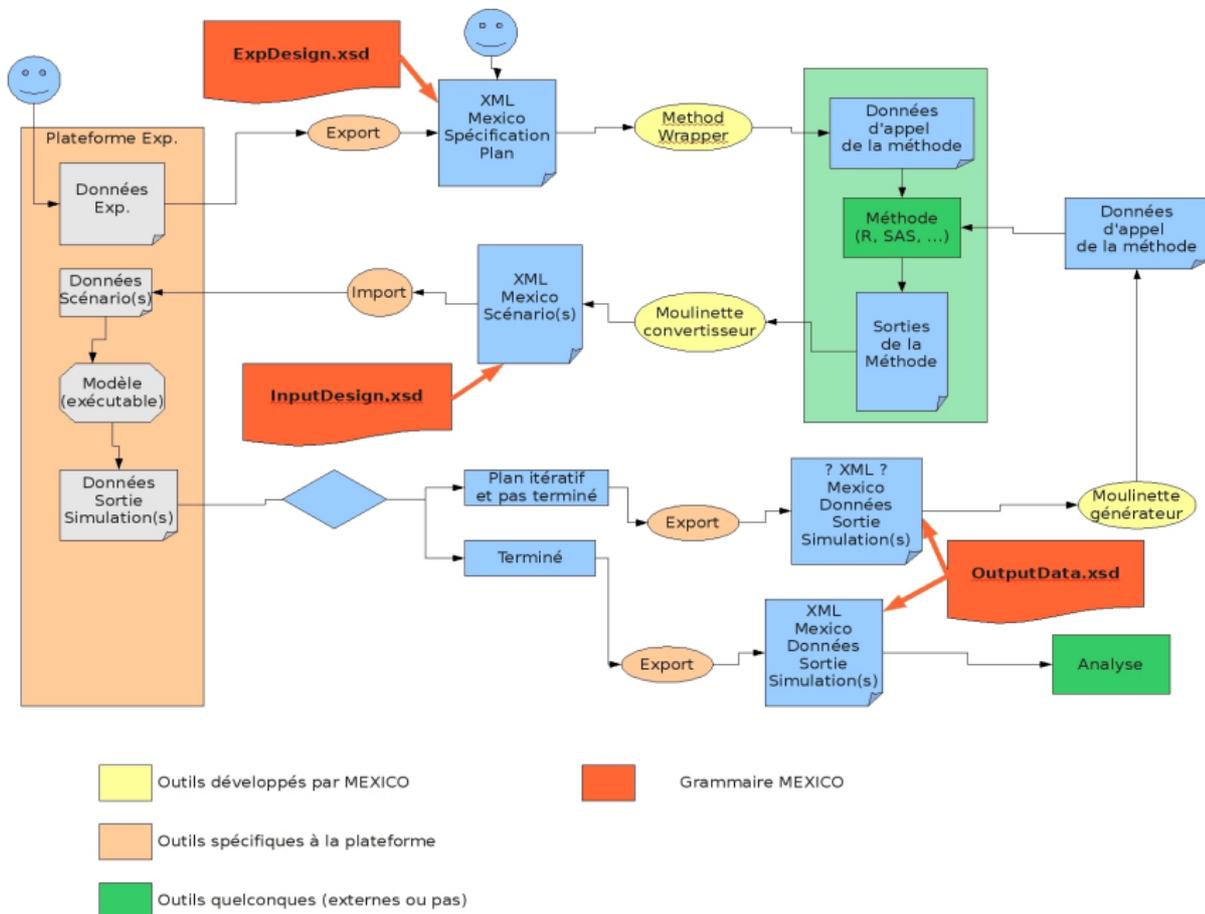
- valider le formalisme mexico à l'aide d'études de cas réels ;
- développer une librairie R mettant en œuvre une large gamme de méthodes ;
- développer une interface informatique rendant conviviale et ergonomique l'utilisation de la boîte à outil développée.

## **2 Méthode de travail**

Nous privilégions une approche par l'utilisateur. Ceci consiste à identifier quelques exemples types d'études de cas « réels », et à les exécuter étape par étape afin de mieux spécifier nos besoins et afin par la suite d'illustrer nos réalisations. Nous proposons d'articuler dans un premier temps notre travail autour de l'analyse de sensibilité et d'inclure au fur et à mesure d'autres méthodes liées à l'exploration de modèle de simulation numérique qui peuvent être des méthodes existantes ou des méthodes développées dans le cadre d'activités de recherche du réseau. Nous nous appuyerons sur l'environnement R et les schémas XML développés par le groupe de travail informatique.

## 2.1 Architecture générale du package informatique « MEXICO »

Le schéma ci-dessous rappelle l'enchaînement des flux de données entre les différents modules de la boîte à outil Mexico. Un des objectifs réalisés par le groupe informatique était de spécifier le formalisme qui décrit les données échangées (*schémas XML en orange*).



## 2.2 Mode de fonctionnement

Il s'agit de constituer, au sein des différentes équipes partenaires, un binôme ou trinôme modélisateur-statisticien /informaticien autour d'une étude spécifique basée sur un modèle existant. L'informaticien sera chargé d'exprimer les différentes étapes de cette étude dans le formalisme XML Mexico. Le groupe organisera une confrontation des ces différentes études pour statuer sur la robustesse du formalisme et documentera celui-ci au fur et à mesure de sa validation.

Il faudra définir un mode de fonctionnement pour tracer les éléments XML qui posent problème, une approche possible est la suivante :

Pour un cas donné, proposer une solution, diffuser cette solution pour que chacun puisse la tester dans ses propres études de cas et décider collectivement (et unanimement) de la valider ou de la rejeter définitivement. Cependant il faudra bien s'assurer de proposer une solution qui remette au minimum en cause le reste de la construction des schémas. Cela nécessite également que chaque

équipe réagisse assez rapidement à chaque proposition pour ne pas bloquer le processus. D'autre part nous nous interrogeons sur l'aide que pourrait apporter des outils de gestion de bug (style « bug trackers ») dans le suivi et la traçabilité des schémas. Il faudra probablement murir cette réflexion pour une coordination optimale des équipes).

### 2.3 *Choix d'étude de cas*

Des modèles de simulation existants et réellement exploités dans le cadre de nos collaborations et activités de recherche seront utilisés comme projets pilotes pour tester et valider les résultats du développement. Nous proposons de nous appuyer sur la réalisation de scénarii avec des modèles proposés par des modélisateurs de notre environnement proche, c'est-à-dire dans nos unités respectives, afin de pouvoir les impliquer dans la démarche pour vérifier l'intérêt de celle-ci. A titre d'exemple, nous décrivons ci-dessous les applications envisagées à MIA-Jouy et BioSP Avignon. Nous invitons chacun à réfléchir sur cette question et à proposer des modèles et des cas d'étude qui soient intéressants. La généralité de nos travaux exige la représentativité et la diversité des modèles. Néanmoins, il nous paraît nécessaire que le modèle proposé soit déjà fonctionnel et que ces modèles n'aient pas besoin des développements supplémentaires dans le cadre du réseau MEXICO.

Les études de cas sélectionnés sont présentés en annexe du document.

En ce qui concerne les autres participants, BIA-Toulouse pourrait par exemple proposer un modèle lié au projet RECORD, ISIS-Fish est déjà un cas pilote, et le LISC, le LIL de ULCO, ou le CIRAD pourraient sans doute élargir la diversité des modèles abordés.

### 2.4 *Réalisation du scénario*

Les études d'exploration numérique seront menées avec les chercheurs porteurs des modèles en respectant et en analysant les étapes suivantes :

- 0) poser le problème en termes d'*Analyse de Sensibilité*. Cette étape impose de réfléchir au travail nécessaire pour passer de la formalisation courante du modèle à une formalisation adaptée à des objectifs précis d'exploration numérique. Par exemple on peut penser à des paramètres liés spatialement ou temporellement et que l'on souhaite faire varier de façon conjointe, ou à la distinction entre variabilité et incertitude: dans quelle mesure la distinction doit-elle rester explicite pendant l'exploration ? Dans quelle mesure ce travail relève-t-il du modélisateur ou bien des outils que l'on souhaite développer ?
- 1) formaliser le problème d'analyse de sensibilité selon le schéma *expDesign.xsd* ;
- 2) construire, en s'appuyant sur le R-package « Sensitivity » dans un premier temps, le plan de simulations et le représenter à l'aide d'un fichier XML selon le schéma *inputDesign.xsd* ;
- 3) exécuter les simulations ;
- 4) récupérer les résultats de sorties dans un fichier XML selon *outputData.xsd* ;
- 5) appliquer, en s'appuyant sur le R-package « Sensitivity » dans un premier temps, une méthode d'analyse de sensibilité ;
- 6) explorer les résultats sous R
  1. réaliser des graphiques, des résumés, ...
  2. en déduire le « reporting » pertinent.

Ce travail s'appuiera sur celui entamé par Danièle Vanbaelinghem au cours de son stage de master I « Ingénierie en Logiciel Libre » de l'Université du Littoral côte d'Opale (mars – juillet2008).

[ Vanbaelinghem Danièle , 2008. « Réalisation d'un processus de workflow pour l'exploration automatisée des simulations sous R » ]

## 2.5 Bilan

Ces opérations successives sur plusieurs modèles doivent servir à fixer de manière consensuelle les schémas XML que nous pourrions ensuite rendre disponibles via une URL à partir du serveur mexico. Par ailleurs, les méthodes à intégrer dans la librairie de la boîte à outil Mexico s'appuieront sur ces spécifications, que ce soit pour la construction du dispositif ou les représentations graphiques.

## 3 Développement informatique

Il s'agira ensuite de réaliser les développements informatiques proprement dits qui permettront l'automatisation des différentes tâches identifiées. Nous pensons que ce travail est réalisable sous l'environnement R. Néanmoins, certaines réflexions doivent être engagées :

- **Pour commencer** : s'assurer d'une vision partagée de l'articulation entre le travail du groupe Mexico et les plateformes existantes telles que ISIS-Fish, VLE, RECORD, ou SimExplorer ;
- **Choix de l'environnement de développement** : doit-il se faire sous R ou dans un langage plus neutre capable d'être appelé par différents outils de calculs comme R, Matlab, SciLab ? Quelles conclusions tirons-nous de l'expérience acquise dans le cadre de VLE-MEXICO et de SimExplorer ?
- **Une étape, un module** : comment organiser la modularité et l'enchaînement des modules entre planification et analyse, entre séquences planification-analyse si l'on aborde des plans séquentiels ?
- **Élargir à d'autres types de méthodes et d'analyse**: conformément aux objectifs du réseau, le package que nous proposons doit permettre d'aller au-delà du R-package « Sensitivity » disponible actuellement et de rassembler les méthodes utilisées par le réseau MEXICO et homogénéiser leurs utilisations. Quelles sont les méthodes qui nous intéressent et la même démarche utilisée dans l'intégration du R-package « Sensitivity » est-elle transposable pour ces nouvelles méthodes ? Quelles réflexions nouvelles et développements informatiques seront nécessaires ?
- **Construire des outils de reporting** : les reporting peuvent apporter une réelle plus value dans la démarche d'analyse pour peu que nous sachions identifier les informations pertinentes à collecter. L'expérience issue de SimExplorer peut-elle nous servir de point de départ ?
- **Construire une interface** simplifiant l'enchaînement des modules. Un outil de ce type peut être un moyen efficace d'attirer des chercheurs plus réticents à une utilisation des outils du domaine.

## 4 Conclusion

Le projet présenté ci-dessus se veut en continuité de ce qui a déjà été accompli dans le cadre de Mexico. Il laisse de nombreux points à préciser avec le reste de la communauté « structure des données » du réseau MEXICO.

Pour nous donner des repères dans le temps, nous proposons le calendrier suivant :

- 1) **mi-mars** pour une version de ce document décrivant les choix techniques et organisationnels

après retour d'information de la communauté;

2) **juin** pour une réunion sur le bilan d'expérience de la mise en place des scénarii définis dans le section 2.3 et la validation des formalismes Mexico (des échanges pourront bien sûr avoir eu lieu au cours de l'Ecole-Chercheur);

3) **fin octobre** pour une réunion sur l'avancement du développement informatique et la publication du package R ;

4) **fin décembre** pour une réunion sur l'avancement du développement informatique pour l'interface graphique et les perspectives du réseau.

# Validation du formalisme XML : Études de cas

## Étude de cas proposée par MIA-Jouy

Il s'agit de modèles issus de l'épidémiologie animale. Développés par Suzanne TOUZEAU dans le cadre de la modélisation d'épidémies dans des élevages, ce sont des modèles de dynamique de populations, de type compartimentaux à temps discret ou temps continu (EDO, EDP) et modèles SIR (susceptibles, infectés, retirés/immunisés).

Ces modèles se distinguent par le fait que :

- les populations d'animaux sont gérées activement par l'homme : la gestion de décisions dans les modèles (démographie, introduction éventuelle de contraintes pratiques et de coûts) fait donc partie intégrante du modèle ;
- l'espace d'états du modèle peut être de grande dimension.

Les paramètres peuvent être :

- quantitatifs (ex.: taux de croissance), ou qualitatifs (ex.: type de renouvellement) ;
- en assez grand nombre (10-20 paramètres) ;
- avec des intervalles d'incertitude plus ou moins larges selon les connaissances accessibles.

Les sorties représentent l'évolution de la maladie dans le troupeau (ex. : prévalence, i.e. proportion d'infectés au cours du temps) et sont donc dynamiques et multi-variées.

Quant à l'analyse de sensibilité, nous cherchons surtout à

- évaluer l'impact de paramètres incertains ou mal connus sur les sorties, afin de mieux orienter les expérimentations ou enquêtes à mener (en particulier sur les paramètres épidémiologiques) ;
- estimer l'impact de caractéristiques de l'élevage sur les sorties, afin de généraliser les résultats à d'autres types d'élevage.

Les applications concrètes du modèle portent sur l'évolution de la tremblante dans un troupeau ovin et le portage de salmonelles en filière porcine [Lurette et al. 2009].

Référence: A. Lurette, S. Touzeau, M. Lamboni, H. Monod (2009). Sensitivity analysis to identify key parameters influencing Salmonella infection dynamics in a pig batch. *Journal of Theoretical Biology*, 2009 (to appear).

## Étude de cas proposée par MIA-Avignon

Il s'agit d'un modèle d'épidémiologie végétale développé par Frédéric Fabre (INRA Unité Pathologie Végétale, Avignon).

Dans un contexte d'utilisation de gènes de résistance aux maladies virales dans les plantes cultivées on cherche des critères prédictifs de la durabilité des gènes de résistances. Il est nécessaire de hiérarchiser l'effet des nombreux facteurs pouvant potentiellement concourir à l'émergence de variants virulents. Nous proposons un modèle aléatoire construit à partir d'un système d'équations différentielles, permettant de quantifier l'effet des facteurs génétiques et épidémiologiques sur la durabilité des résistances. De plus, le modèle simule à deux échelles, hôte et parcelle, la dynamique démo-génétique au cours d'une épidémie.

Au total, 10 facteurs quantitatifs ou qualitatifs sont inclus dans le modèle :

- (1) le nombre de mutations nécessaires au contournement,
- (2) la nature de ces mutations (transitions ou transversions),
- (3) le taux de mutation,
- (4) le ratio taux de transitions sur taux de transversions,
- (5) la vitesse de multiplication virale dans l'hôte,
- (6) le coût de fitness associé aux mutations,
- (7) le degré de ségrégation spatiale,
- (8) la vitesse de développement de l'épidémie,
- (9) le temps de latence de l'infection,
- (10) le nombre de particules virales transmises à une plante saine.

Des analyses de sensibilité globale (méthodes Sobol et ANOVA) ont déjà été réalisées [Fabre et al. 2009]. Référence : Fabre, F., Bruchou, C., Palloix, A., & Moury, B. 2009. Key determinants of resistance durability to plant viruses: insights from a model linking within- and between-host dynamics. *Virus Res.*, In press.

## Étude de cas proposée par IFREMER EMH/ Code Lutin

**Objectif :** Explorer numériquement le modèle ISIS (Mahevas et Pelletier 2004) (sélectionner les paramètres à faire varier, choisir une méthode d'exploration des domaines de définition des paramètres sélectionnés, construire le plan d'expériences, lancer un plan d'expériences, analyser le plan d'expériences)

**Exemple :** pêche virtuelle ecoMod2 (Pelletier et al sous presse)

**Contrainte :** les développements nécessaires doivent être cohérents avec les développements à venir pour utiliser l'outil Mexico – formats d'échange d'information entre ISIS et logiciel Mexico\_R (format Mexico XML)

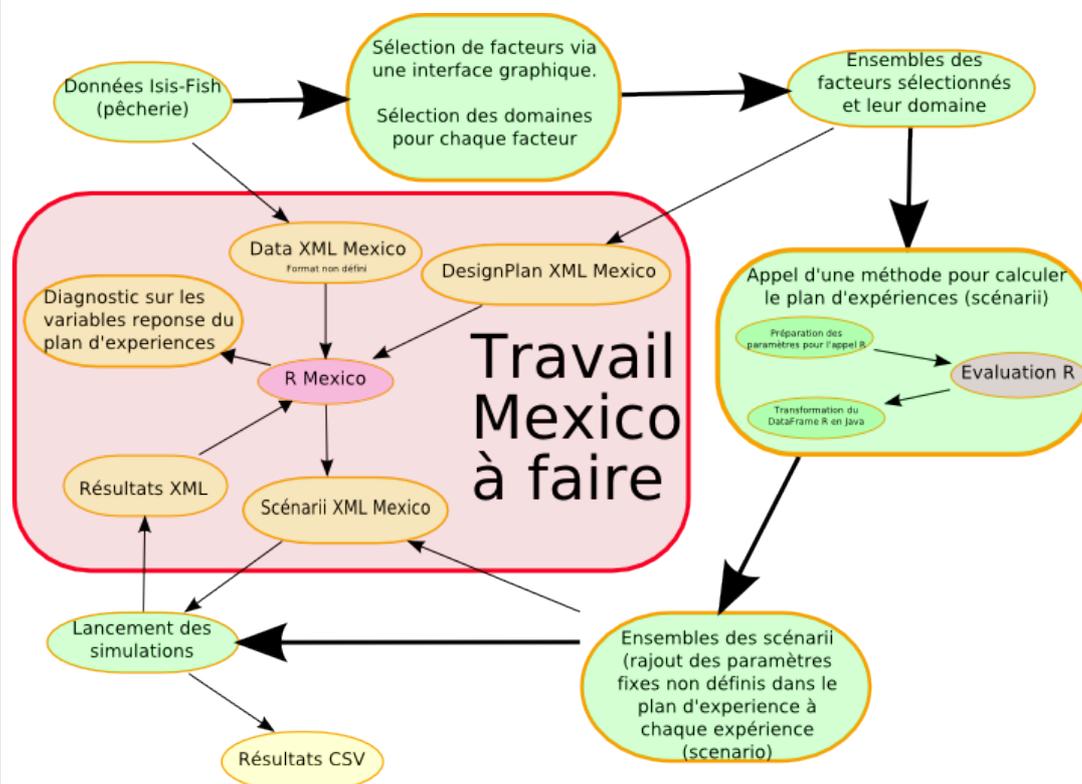


Figure 1 : Étapes de mise en œuvre d'une exploration numérique d'ISIS-Fish (domaine vert) par analyse de sensibilité en utilisant R (domaine Rose).

**Étape 1 :** sélectionner les paramètres d'Isis à faire varier

2 choix : développer nos propres interfaces, ou utiliser les interfaces Sim Explorer dans les deux cas : il faut extraire la BD d'ISIS et l'exporter dans un format XML qui puisse être interpréter pour remplir les interfaces – format Mexico XML

Les deux choix seront développés. La priorité est mise sur nos propres interfaces (pour ne pas être trop dépendant de SimExplorer avant l'école chercheur).

**Étape 2 :** sélection des paramètres et de leur domaine de définition – choix de la méthode d'exploration des domaines définition (dans une liste correspondant dans un premier temps à la librairie Sensitivity puis élargie dans un second temps à la librairie

R-Mexico – Au moyen de l'interface

**Étape 3 :** Lancement depuis l'interface (ISIS, ou SimExplorer, ou Mexico) du plan d'expérience sous R

**Étape 4 :** A partir des sorties de R – export au format Mexico\_XML des paramétrages de lancement des simulations

**Étape 5 :** lancement des simulations (mode batch ou autre)  
relecture des sorties R au format Mexico\_XML et visualisation dans l'interface (ISIS, ou SimExplorer, ou Mexico) – choix du mode de lancement directement sur la machine, mode batch pour lancer le plan sur un super-calculateur (ex Caparmor ou machine Cemagref)

**Étape 6 :** récupérations des exports ISIS (sorties des simulations) et mise au format Mexico\_XML pour en permettre l'analyse sous R –

**Étape 7 :** sous R analyse du plan d'expériences

Exemples d'analyses de sensibilité d'ISIS-Fish

1 pêche Merlu-langoustine golfe de Gascogne (Drouineau et al 2005)  
plan d'expériences : D-optimal design et plan factoriel fractionnaire (Min,Max)  
Analyse des sorties : Coefficient de sensibilité pour chaque variable réponse et ACP pour caractériser la sensibilité pour l'ensemble des variables réponse

2 pêche Anchois golfe de Gascogne (Lehuta et al soumis et Mahévas et al 2008)  
plan d'expériences : plan factoriel fractionnaire, plan complet (Min,Max), plans hiérarchiques  
Analyse des sorties : PLS pour des coefficients de sensibilité multivariés

3 pêche de Morue de mer Baltique (Kraus et al 2009)  
plan d'expériences : simulation de 2 scenarii d'environnement – pas de plan  
Analyse des sorties : indicateurs

Références

Lehuta S., Mahévas, S., Petitgas, P, Pelletier D. (soumis) Evaluating Marine Protected Areas impact using sensitivity analysis and simulation : the case of the Bay of Biscay Anchovy (*Engraulis encrasicolus*) fishery. ICES Journal of Marine Science  
Pelletier, D., Mahévas S., Drouineau H., Vermard Y., Thebaud O., Guyader O., Poussin B. (2009). Evaluation of the bioeconomic sustainability of complex fisheries under a wide range of policy options using ISIS-Fish. Ecological Modelling. Sous presse.

## Étude de cas proposée par le CIRAD - BIOS «BIological Systems»

*Proposition de*

*Jean-Christophe Soulie CIRAD - BIOS «BIological Systems» Department*

### **Nature du modèle**

*Ecomeristem* est un modèle mécaniste simulant la morphogénèse de la plante (notamment le riz, modèle pour les céréales) et sa plasticité en réponse à son environnement abiotique (eau, lumière, température). La topologie de la plante est simulée au pas de temps journalier à l'échelle de l'organe, la feuille au niveau aérien (les racines n'étant qu'un seul compartiment accumulant de la biomasse). Les processus composant la morphogénèse sont contrôlés par des paramètres génotypiques (variétaux):

(1) de type constitutif: exprimant un potentiel génétique (eg. dimensionnement et rythme d'initiation des feuilles et ramifications (talles).

(2) de type 'seuil' de réponse à des variables d'état journalières de la plante (eg. paramètre *Ict*, fixant le seuil d'*Ic* à partir duquel la plante émettra des talles ou ramifications, avec *Ic*: ratio entre offre et demande en assimilats carbonés dans la plante (offre: résultat de la photosynthèse; demande: somme des croissances d'organes). De façon similaire, des paramètres seuils contrôlent la réponse de l'expansion et transpiration foliaire au stress hydrique.

### **Problématique**

De par sa nature, le modèle *Ecomeristem* simule une réponse non linéaire de la morphogénèse de la plante soumise aux contraintes du milieu extérieur. Les processus de régulation mis en œuvre au sein de ce mécanisme étant fortement interdépendants (e.g. taille vs. vitesse de croissance et d'apparition des feuilles vs. nombre de feuilles ou ramification), les paramètres qui y sont attachés se trouvent de la même manière liés que cela soit d'un point de vue physiologique, génétique mais aussi mathématique.

Par ses concepts fondateurs, les paramètres du modèle communément considérés par ses utilisateurs pour discriminer des génotypes (eg. par des méthodes heuristiques), sur leur productivité et vigueur en réponse à un environnement donné, sont ceux contrôlant de façon directe les mécanismes composites cités ci-dessus (par exemple, dimensionnement potentiel (MGR) et rythme d'émission (phyllochron) des feuilles, sensibilité du tallage à la disponibilité en photosynthétats (*Ict*), cf. Fig. 1).

Cependant, une analyse de sensibilité systématique du modèle, relative à l'ensemble de ses paramètres biologiques, n'a jamais été réalisée, au risque de manquer des hypothèses de recherche fortes. En particulier, une meilleure appréhension des facteurs contrôlant dans le modèle la vigueur de croissance de la plante dans un environnement donné pourrait permettre d'améliorer une exploration de la diversité génétique au niveau des ses processus à l'aide du modèle.

### **Objectif :**

L'objectif de cette étude sera atteint en deux phases :

- 4 mener une analyse de sensibilité du modèle quasi-exhaustive au travers de ses paramètres 'plante', en considérant une seule variable de sortie, la plus intégrative possible : la biomasse totale de la plante
- 5 a- confirmer ou non le rôle-clé des paramètres jusqu'à maintenant utilisés pour

discriminer des génotypes ;

b- mener une seconde analyse de sensibilité du modèle vis-à-vis des paramètres ayant montré les indices de sensibilité les plus forts sur la biomasse de la plante ;

c- quantifier leur impacts sur des variables de sorties constitutives de la production de biomasse (notamment, le nombre de feuilles sur la plante, taille des feuilles à une position donnée, le nombre de talles, les réserves en amidon dans la plante).

### Approche choisie :

### Expériences de simulation :

on se place dans des conditions environnementales non contraignantes (pas de stress), avec un jeu de données climatiques issues de.... Sur 30j.

Dans le premier cas, 10 paramètres seront adressés

Dans le deuxième, les 5 paramètres les plus importants selon l'analyse précédente.

### Analyse de sensibilité globale du modèle :

Compte tenu de la non-linéarité, du nombre relativement faible de paramètres d'entrée du modèle *Ecomeristem* mais aussi de son faible coût en temps de calcul (de l'ordre 300ms pour une simulation) une approche basée sur la décomposition de la variance de la réponse (type Sobol, FAST) sera privilégiée.

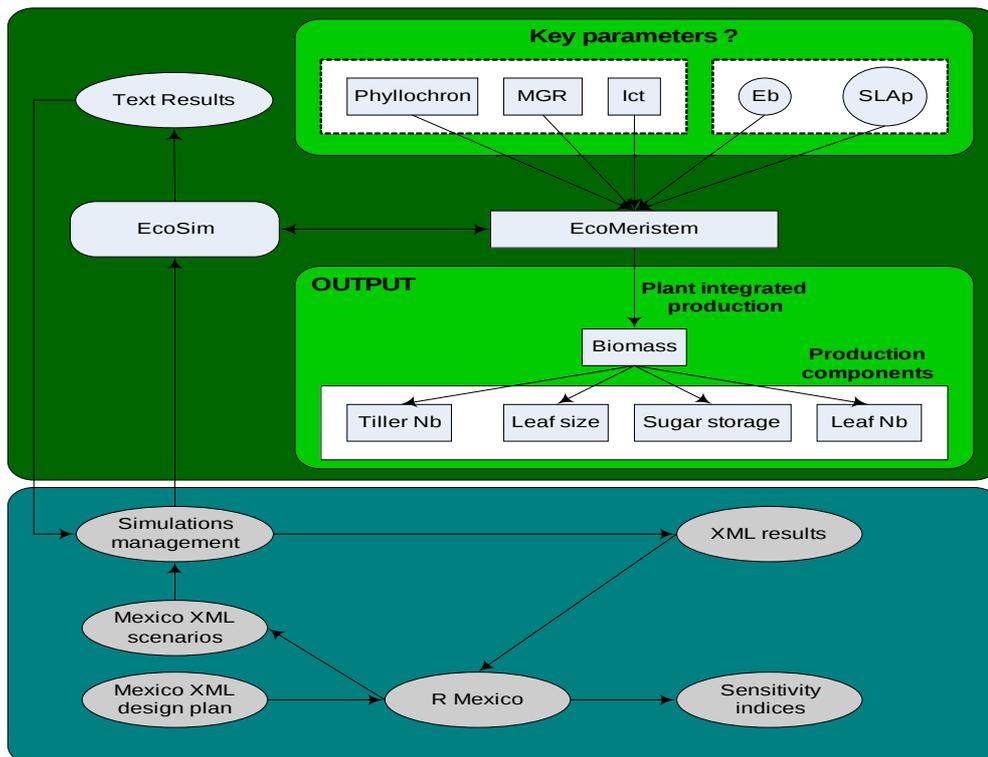


Fig. 1 : Intégration de la problématique adressée dans le BAO Mexico