

Modèles Dynamiques de la Croissance des Plantes : Succès et Défis

Paul-Henry P.-H. Cournède

▶ **To cite this version:**

Paul-Henry P.-H. Cournède. Modèles Dynamiques de la Croissance des Plantes : Succès et Défis. Potentiels de la science pour une agriculture durable, Académie d'Agriculture de France. Edition: Jean-Claude Pernollet., , 2018. hal-02285935

HAL Id: hal-02285935

https://centralesupelec.hal.science/hal-02285935

Submitted on 13 Sep 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Modèles Dynamiques de la Croissance des Plantes : Succès et Défis *

Paul-Henry Cournède

Laboratoire de Mathématiques et Informatique pour la Complexité et les Systèmes, CentraleSupélec, Université Paris-Saclay

Résumé

Depuis quelques décennies, le travail conjoint des biologistes, agronomes, mathématiciens et informaticiens a permis le développement des modèles de croissance de plantes. S'il existe une typologie assez variée, dépendant en particulier des échelles de description et des processus étudiés, ces modèles partagent l'ambition d'offrir un cadre pour l'analyse de l'interaction génotype-environnement.
Nous présentons dans cet article les grandes étapes du développement des modèles de croissance de plantes et les concepts clés sur lesquels ils reposent, notamment pour les modèles dits structure-fonction, qui combinent la description du fonctionnement de la plante et son développement architectural. Nous détaillons en particulier le cadre mathématique adapté à l'intégration de ces concepts et soulignons l'importance des problèmes statistiques posés par l'analyse des données pour la paramétrisation et la validation des modèles.
Une fois la démarche de modélisation présentée, nous illustrons par quelques exemples d'applications comment l'amélioration de la prévision des phénotypes par ces modèles ouvre des perspectives importantes pour l'agriculture : cela concerne l'aide à la décision et l'optimisation des pratiques culturales ou encore le processus d'amélioration variétale. Finalement, nous soulignons un défi à venir pour les modèles : comment peuvent-ils bénéficier de l'information génomique désormais facilement accessible ?

Abstract

For a few decades, the joint work of biologists, agronomists, mathematicians and computer scientists has allowed the spectacular development of plant growth models. If their typology is very diverse, mostly according to the scales at which the different growth processes are described, these models share the ambition to propose a framework to analyze the genotype×environment interaction.
In this paper, we present the main historical steps in plant growth model development and the key concepts underlying their construction, with a particular focus on functional-structural plant models which combine the description of plant ecophysiological functioning and architectural development. We also detail the mathematical formalism adapted to the integration of biological concepts and underline the statistical issues raised by data analysis for model parameterization and validation.

*Collection : Potentiels de la science pour une agriculture durable, Académie d'Agriculture de France. Editeur : Jean-Claude Pernollet. 2018. https://www.academie-agriculture.fr/academie/groupes-de-travail/potentiels-de-la-science-pour-une-agriculture-durable

Once the modelling methodology exposed, we illustrate by a few applicative examples the important perspectives in agriculture : by an improved prediction of phenotypes, models can support new decision aid tools for farmers or help optimizing the breeding process. Finally, we will conclude by underlining a new challenge for plant models : how can they take advantage of the now easily available genomics data ?

1 Histoire récente de la modélisation de la croissance des plantes

L'histoire de la modélisation de la croissance des plantes a connu un essor spectaculaire dans les années 1990, lorsque deux approches jusqu'alors relativement séparées se sont rencontrées : l'approche géométrique et descriptive développée en foresterie et l'approche écophysiologique et énergétique de l'agronomie.
Cette dichotomie historique peut sans doute s'expliquer par la différence de l'objet d'étude dans les deux disciplines. Si l'arbre cache la forêt, le champ cache sans doute l'épi de blé... Aussi, en foresterie, les scientifiques ont longtemps développé des études dendrométriques, c'est à dire des études basées sur la mesure de caractéristiques géométriques de l'arbre, bien sûr le fameux diamètre "à hauteur de poitrine", mais également la hauteur, la forme du houppier, et pour les plus détaillées les structures des branches principales. Une approche plus analytique de la botanique développée notamment par Hallé and Oldeman (1970) a permis de coupler ces mesures avec une description plus fine des structures végétales. De fins modélisateurs et des infographistes de talent ont ainsi tiré parti de cette analyse pour créer les premiers modèles graphiques de plantes, décrivant de façon fine leur géométrie (de Reffye et al., 1988; Prusinkiewicz et al., 1988) avec des résultats graphiques spectaculaires pour l'époque (Figure 1). Des procédures d'interpolation entre des mesures à différents âges permettent alors de donner une impression de croissance à ces plantes. Le succès fut alors au rendez-vous, pour l'utilisation dans les projets d'aménagement architecturaux ou encore dans les jeux vidéos. Malgré tout, le caractère purement descriptif reste trop limité pour des applications plus larges.



FIGURE 1 – *Simulation et visualisation de la croissance d'un épicéa à 10 ans et à 30 ans par le logiciel AMAP (de Reffye et al. (1988)). Images par courtoisie de Marc Jaeger, Cirad.*

D'un autre côté, l'agronomie s'est historiquement intéressée de façon prioritaire à l'objectif premier de l'agriculture, la production végétale, et plus précisément le rendement des cultures. La compréhension des mécanismes permettant d'expliquer la constitution du rendement doivent permettre d'améliorer ce dernier. A l'approche statistique de Fisher a succédé les premiers modèles d'écophysiologie, en particulier sous l'impulsion de Monteith

and Moss (1977) et De Wit et al. (1979). La culture est considérée comme un continuum (la plante individuelle n'est pas considérée), on évalue l'interception lumineuse, l'évaporation de la culture et le bilan hydrique qui permettent de déterminer un stress hydrique, le besoin en azote, la conversion en composantes du rendement...La description de l'interaction entre la culture et l'environnement est au cœur de l'approche. Plusieurs modèles se développent, citons en particulier l'école américaine, autour de CERES et de ses successeurs (Jones et al., 1986), l'école hollandaise à la suite des travaux de De Wit (Bouman et al., 1996), l'école australienne avec APSIM (McCown et al., 1996), et l'école française autour de STICS à l'INRA, projet porté par la regrettée Nadine Brisson (Brisson et al., 1998). Ces modèles sont souvent génériques, très riches du point de vue de l'interaction plante-environnement, et sont les premiers à être concrètement utilisés pour des applications en agriculture. Malgré tout, leur description trop sommaire de l'architecture de la plante les rend difficilement capables de prendre en compte certains aspects de la plasticité de la culture (dont l'architecture est un élément fondamental) en particulier en situation de stress, et ils souffrent d'un manque de robustesse dans ces situations. Une autre limitation importante de ces modèles est la difficulté de leur utilisation pour l'amélioration génétique car une part importante des traits de la plante sur lesquels les sélectionneurs travaillent sont architecturaux.

Dans les années 1990, ces deux approches se sont enrichies l'une l'autre, pour donner naissance aux modèles structure-fonction (Functional-Structural Plant Models, FSPM en anglais), qui combinent la description du développement structural des plantes et leur fonctionnement écophysiologique, ainsi que l'interaction entre les deux processus. Le premier FSPM est sans doute celui proposé par E. D. Ford et ses collègues (Sorrensen-Cothern et al., 1993) pour le sapin, qui décrit de façon fine l'interception de la lumière par la plante grâce à sa structure 3D. L'effet de l'auto-ombrage et de la compétition avec les voisins est pris en compte. L'adaptation de la plante à l'environnement est décrite par le développement différencié de chaque partie de l'arbre, par l'effet de l'acquisition locale de biomasse.

La première conférence FSPM s'en est ensuivie en Finlande en 1996, avec les premières contributions majeures de quelques groupes qui allaient marquer les décennies à venir par leurs apports, notamment les Finlandais et le modèle Lignum (Sievänen et al., 1997), et les Français du CIRAD et de l'INRA de l'unité AMAP à Montpellier (de Reffye et al., 1997). Les premiers résultats concernent principalement les arbres, mais assez rapidement la même démarche est appliquée pour les cultures, en insistant sur l'écophysiologie (voir par exemple Fournier and Andrieu (1998) de l'INRA Grignon sur le ï s).

Nous allons dans la suite donner quelques concepts clés pour la formulation de ces modèles.

2 Concepts clés

Comme l'explique l'historien des sciences Franck Varenne (Varenne, 2003), un des enjeux importants qui a suivi la création des premiers modèles structure-fonction a été leur re-mathématisation : ces modèles, basés sur des programmes informatiques de plus en plus complexes, ont rapidement fini par s'identifier aux logiciels qui les simulent et non aux formalismes mathématiques sous-jacents, avec le cloisonnement qui en découle : chaque équipe développe alors son modèle, identifié à son logiciel, et ce modèle est souvent réalisé dans l'optique d'étudier une question scientifique particulière, importante pour l'équipe de recherche mais difficile à généraliser.

Le travail sur la formalisation mathématique des modèles et le développement de méthodes mathématiques pour leur analyse est l'objectif majeur de l'équipe Digiplante que nous

avons créée en 2004 avec Philippe de Reffye et Marc Jaeger, entre l'Ecole Centrale Paris, l'INRIA et le CIRAD, et qui est aujourd'hui une équipe du laboratoire de Mathématiques et Informatique de CentraleSupélec, à l'Université Paris-Saclay.
La démarche mise en place est véritablement celle de la modélisation mathématique, en s'inspirant de la méthodologie dite des bonnes pratiques de modélisation ('good modelling practice', Scholten et al. (2007)). Avant toute chose, il est important de bien préciser l'objectif du modèle : pourquoi le construit-on ? Il est intéressant de constater que bien souvent, en modélisation de systèmes biologiques, le modèle est avant tout vu comme un support à l'intégration de toute la connaissance existante, qui est souvent gigantesque en biologie. Malgré tout, une telle approche est confrontée à des écueils mathématiques majeurs : gestion d'échelles spatiales et temporelles très différentes, paramétrisation du modèle et propagation d'incertitudes, interactions entre processus mal connues... Il convient donc de préciser avant tout dans quelle optique le modèle est construit : si on veut prédire le rendement d'un champ de blé, il y a fort à parier que le modèle le plus efficace ne sera pas le même que si l'on veut prédire l'effet physiologique d'une mutation génétique sur une variété. En ce qui concerne la formalisation mathématique, la théorie des systèmes dynamiques à espaces d'états offre un cadre formel adapté pour décrire la croissance de la plante. Le système est la plante individuelle, et l'on peut écrire de façon très générique :

$$\begin{cases} X(n+1) = F(X(n), U(n), \theta) \\ X(0) = X_0 \end{cases} \tag{1}$$

où $X(n)$ représente les variables d'états du système au temps n, par exemple les masses et dimensions des différents organes de la plante, θ représente les paramètres du système, que l'on souhaite caractéristique d'une espèce, et même d'un génotype, $U(n)$ représente les variables externes au temps n, typiquement les conditions pédoclimatiques et les pratiques culturales, et F, la fonction d'état, décrit les processus biophysiques de la croissance pris en compte dans le modèle. X_0 donne la condition initiale, de façon simplifiée, on peut l'assimiler à la masse de la graine par exemple. Notons que le système ainsi formalisé est déterministe. Bien souvent, il est également nécessaire de se placer dans un cadre stochastique, pour décrire la variabilité des réalisations du système. $X(n)$ est alors un vecteur aléatoire, et on peut généralement écrire

$$X(n+1) = F(X(n), U(n), \theta, \xi(n))$$

où $\xi(n)$ est un vecteur aléatoire, représentant les différents aléas sous-jacents, que l'on modélise par des distributions de probabilité particulières. De façon équivalente, il s'agit de modéliser la loi conditionnelle $X(n+1)|X(n), \theta, U(n)$, (Cournède et al., 2013).
Une des particularités de la modélisation structure-fonction est qu'elle doit décrire à la fois le développement structural et le fonctionnement écophysiologique de la plante.

2.1 Organogenèse

Un type de grammaires formelles, les systèmes de Lindenmayer, se sont révélés un outil extrêmement puissant pour modéliser la dynamique de mise en place des organes et de la structure des plantes (Smith, 1984; Prusinkiewicz and Lindenmayer, 1990). Le modèle GreenLab utilise en plus de cette grammaire, quelques éléments de botanique analytique introduits par Barthélémy and Caraglio (2007). En particulier, GreenLab repose sur la conception hiérarchique de la structure de la plante : les organes sont rassemblés par phytomère, qui est l'unité de base, généralement constituée d'un entrenoeud, portant une feuille, potentiellement des fruits et des bourgeons axillaires, ces derniers pourront donner

des ramifications. Les métamères sont eux-mêmes rassemblés par unité de croissance qui définit l'ensemble des phytomères qui sur le même axe apparaissent en même temps. Une horloge d'organogenèse rythme l'apparition des nouvelles unités de croissance. Pour les arbres des régions tempérées, généralement une nouvelle unité de croissance apparaît au printemps. Pour les plantes de cycle plus court, comme le maïs, l'arabidopsis et la betterave, c'est le temps thermique, c'est à dire la somme des degrés.jours accumulés, qui détermine la vitesse d'apparition des nouvelles unités de croissance. La succession des unités de croissance donne les axes, et les branchements entre différents axes, donne la structure complète. Le concept d'âge physiologique permet de différencier les différents types d'axes ou d'organes, cf. Figure 2.

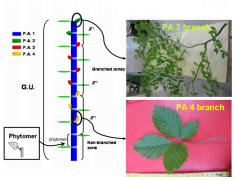


FIGURE 2 – *Description analytique d'une unité de croissance (G.U. pour growth unit en anglais) de l'axe principal (age physiologique 1) pour le hêtre : l'unité de base est le phytomère. Ils se distinguent en fonction de l'âge physiologique (P.A.) du bourgeon axillaire qu'ils portent, ici de l'âge physiologique 2 pour les branches les plus vigoureuses à l'âge physiologique 4 pour les petits rameaux (Letort et al., 2008a).*

Une fois cette analyse botanique réalisée, la formalisation en L-Système, permet de gérer la production des bourgeons et donc la dynamique de mise en place de la structure de la plante, comme illustré sur la Figure 3.
Les règles de production sont généralement stochastiques (Kang et al., 2008; Loi and Cournède, 2008) ou déterminées par des règles de disponibilité des ressources pour la plante (Mathieu et al., 2009). Ce modèle d'organogenèse permet une grande souplesse de représentation et s'adapte aussi bien aux plantes annuelles, comme le maïs, ou le colza, qu'aux arbres.

2.2 Fonctionnement

L'architecture de la plante est en quelque sorte son squelette. Pour prendre chair, elle va se reposer sur deux phénomènes : la production de biomasse, grâce à la photosynthèse, et l'allocation, qui est la distribution de la biomasse nette produite entre tous les organes existants et les nouveaux organes à paraître.
La Figure 4 schématise les principaux processus de la croissance.
Il est à noter que deux horloges (pas de temps) différentes peuvent potentiellement être utilisées dans le modèle, le fonctionnement est plus simplement évalué en se calant sur le temps calendaire, alors que comme nous l'avons évoqué, l'organogenèse est calée sur le temps thermique.
De façon simplifiée, nous pouvons écrire la production $\Delta Q(n)$ au pas de temps n de la façon suivante :

$$\Delta Q(n) = E(n)\mu S_p \left(1 - e^{-kS(n)/S_p}\right)$$

où :

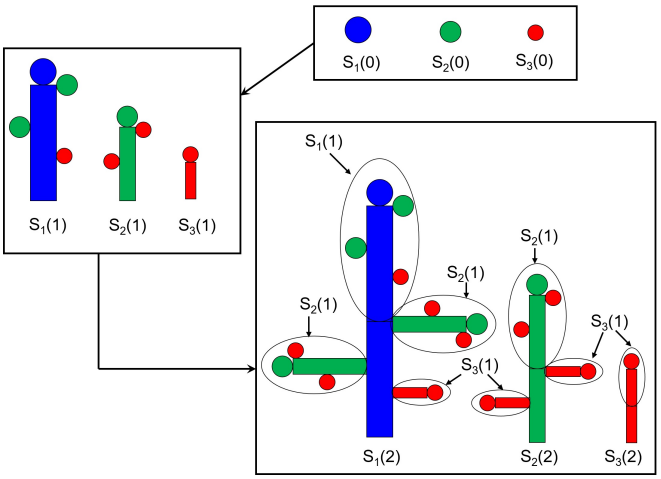


FIGURE 3 – *Exemple de L-Système avec 3 âges physiologiques. Chacun des trois types de bourgeon a sa règle de production d'une unité de croissance, qui contient elle-même des bourgeons (généralement un terminal et potentiellement plusieurs axillaires) et par composition, les structures sont créées : $S_k(j)$ représente la structure d'âge physiologique k et d'âge chronologique j. $S_k(0)$ correspond donc au bourgeon d'âge physiologique k, $S_k(1)$ à une unité de croissance, et $S_k(2)$ à des structures après deux cycles de croissance. Un des points intéressants, est que les structures sont construites par composition, et ainsi se répètent un grand nombre de fois dans la plante. On imagine ce qu'une prise en compte de cette propriété permet en termes de gain pour la simulation des arbres, cf. (Cournède et al., 2006).*

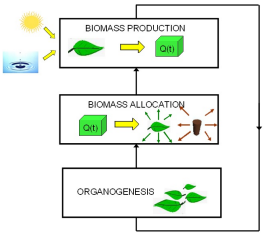


FIGURE 4 – *La production de biomasse résulte de l'interception lumineuse qui permet la photosynthèse. Cette biomasse est ensuite répartie dans l'architecture de la plante et permet la mise en place des nouveaux organes. A noter qu'à l'initiation la biomasse est fournie par la graine.*

- $E(n)$ est une fonction des variables environnementales au temps n, principalement les radiations incidentes, mais potentiellement des fonctions de stress en particulier hydrique ou azoté,
- μ est une efficience de conversion énergétique,
- S_p est une surface caractéristique de la plante (plus il y a de compétition, plus cette surface est petite)
- $S(n)$ est la surface foliaire de la plante au temps n
- k est le facteur d'interception de la loi de Beer.

$S_p\left(1-e^{-kS(n)/S_p}\right)$ peut être vu comme la fraction de radiation interceptée. Dans certains modèles, au lieu d'être évaluée de façon empirique, cette fraction peut être calculée de façon plus fine en regardant ce qui se passe au niveau de chaque feuille par lancer de

rayons ou par des calculs de radiosité (Fournier and Andrieu, 1998).
Cet incrément de biomasse est alors distribué entre les différents organes, en fonction de leurs puits d'attraction de biomasse, qui régissent ainsi la compétition. Ce puits dépend du type de l'organe (feuille, fruit, tige...) ainsi que de son âge (généralement calculé en temps thermique), les organes ne demandant pas la même chose à tous les moments de leur vie. Si on considère $p_i(n)$ le puits au pas de temps n de l'organe i, nous avons alors l'incrément de biomasse $\Delta m_i(n)$ pour l'organe i qui est simplement calculé de la façon suivante :

$$\Delta m_i(n) = p_i(n)\frac{\Delta Q(n)}{D(n)}, \text{ avec } D(n) = \sum_i p_i(n) \ .$$

Dans le modèle GreenLab original (De Reffye and Hu, 2003), la fonction puits est une fonction empirique de type bêta, dont les différents paramètres sont déterminés par calibration (Cournède et al., 2011). La somme de tous ces puits donnent la demande totale de la plante $D(n)$. Les incréments de biomasse concernent bien sûr également les nouveaux organes, ceux dont le module d'organogenèse a prédit l'apparition.
Notons que comme montré dans la Figure 3, on peut trouver au même moment un grand nombre d'organes similaires dans la plante : en conséquence, il n'est pas nécessaire de calculer chacun individuellement. On factorise (mathématiquement et informatiquement) les objets identiques. On a alors :

$$D(n) = \sum_i N_i(n)p_i(n) \ .$$

où cette fois i indice l'ensemble des classes d'organes différents (en âge et en type) et $N_i(n)$ représente tous les organes identiques dans la classe i.
Un élément intéressant de l'équation de production est le facteur $\frac{\Delta Q(n)}{D(n)}$ qui à tout moment mesure la compétition pour les ressources entre les organes de la plante.
A partir des incréments de biomasse, nous avons donc les masses de chaque organe à tous les pas de temps, et des paramètres d'allométrie nous permettent de revenir à leur géométrie. En particulier, nous pouvons en déduire la surface de chaque feuille ou la longueur des axes. Le cycle est alors bouclé, et formellement nous sommes bien sous la forme du système dynamique générique (1).
Les connaissances en écophysiologie permettent de grandement raffiner et détailler les équations simples données ci-dessous. Bien sûr la sénescence et la remobilisation doivent être prises en compte. Des modèles décrivant de façon détaillée les processus de photosynthèse et respiration existent (Pury and Farquhar, 1997), en particulier basé sur le modèle de Farquhar (Farquhar and Von Caemmerer, 1982). De même, l'efficience énergétique peut être variable en cours de croissance, en fonction de la disponibilité en azote et en eau. Ces différents éléments permettent de raffiner l'équation de production. La mécanique d'allocation peut également être décrite de façon plus précise (Minchin and Thorpe, 1996; Bancal and Soltani, 2002), en prenant en particulier en compte les coûts de transport des assimilats. De même, l'architecture du système racinaire joue bien sûr un rôle crucial dans la performance de la plante (de Dorlodot et al., 2007). Certains modèles ont donc cherché à décrire leur croissance (Pagès et al., 2004; Dupuy et al., 2010).
Ces quelques exemples montrent qu'il est aisé (et même tentant) de complexifier les modèles. Malgré tout, comme nous le verrons par la suite, ce n'est pas forcément sans conséquence sur sa robustesse. En fonction de l'objectif visé, il conviendra toujours de trouver le bon équilibre entre complexité et robustesse, illustration classique du compromis entre biais et variance.

3 Application des Modèles

3.1 Paramétrisation

Une étape cruciale pour l'application concrète des modèles est leur paramétrisation et leur évaluation en les confrontant aux données réelles. Pour cela, on construit un modèle statistique d'observation, c'est à dire qu'on modélise la loi conditionnelle $Y(n)|X(n), \theta$, par exemple :

$$Y(n)|X(n) \sim \mathcal{N}\left(X(n), \Sigma\right)$$

pour un bruit blanc gaussien de matrice de covariance Σ. En agronomie, comme plus généralement en biologie, les observations sont hétérogènes et irrégulières : toutes les variables ne sont pas disponibles à tous les pas de temps. Nous introduisons donc le vecteur $\mathbf{Y}$, correspondant au vecteur de toutes les observations sur tout le cycle de croissance, jusqu'au pas de temps final N (généralement la récolte). Le modèle statistique d'observation est alors donné par la loi de $\mathbf{Y}|X_0, X_1, \cdots, X_N, \theta$.
Si la condition initiale X_0 et les conditions climatiques sont connues et si par ailleurs $X(n+1)|X(n), U(n), \theta$ est déterministe (pas de stochastique dans les équations du modèle), alors le modèle statistique d'observation est simplement donné par la densité conditionnelle $p(\mathbf{Y}|\theta)$. Une inférence classique peut alors être réalisée. La méthode la plus courante est une estimation du type du maximum de vraisemblance (ou ses dérivées du types moindres carrés généralisés), c'est à dire pour un vecteur d'observations $\mathbf{y}$ donné, en maximisant la vraisemblance $\mathcal{L}(\theta) = p(\mathbf{y}|\theta)$. Plus récemment, se sont développées les méthodes bayésiennes : à partir d'une distribution a priori sur les paramètres $\pi(\theta)$, on détermine la distribution a posteriori $p(\theta|\mathbf{y})$ en utilisant la formule de Bayes

$$p(\theta|\mathbf{y}) = \frac{p(\mathbf{y}|\theta)\pi(\theta)}{p(\mathbf{y})} \ .$$

Il convient en général de faire appel aux méthodes de Monte-Carlo séquentielles ou de Monte-Carlo par chaine de Markov (MCMC) pour contourner le fait que l'expression de $p(\mathbf{y})$ repose sur le calcul d'une intégrale difficile à évaluer.
Les méthodes d'estimation bayésiennes sont particulièrement bien adaptées à la modélisation en biologie, car elles permettent d'intégrer de façon très naturelle la connaissance sur les paramètres qui est généralement disponible en raison de leur fondement biologique et de l'expertise existante sur les systèmes. Par ailleurs, les méthodes de ce type "font feu de tout bois", elles exploitent toute l'information disponible, quel que soit le caractère hétérogène, rare ou irrégulier. Enfin, le résultat de l'estimation bayésienne est une distribution de probabilité, c'est à dire que l'on maitrise exactement le degré d'incertitude sur les paramètres, ce qui se traduit également par la maitrise de l'incertitude sur les prévisions.
Nous illustrons des résultats de la procédure d'estimation pour le modèle GreenLab-betterave, dans la Figure 5. Il est à noter que parfois, le très grand nombre de paramètres du modèle ne permet pas de tous les déterminer par estimation paramétrique. Une analyse de sensibilité permet alors de sélectionner les paramètres les plus influents (Wu et al., 2012).

3.2 Prévision et Aide à la décision

Une fois le modèle calibré, il peut être utilisé en prévision. Ainsi, pour une nouvelle campagne de culture, les nouvelles conditions environnementales nous donnent le vecteur $(U_1, \cdots, U_n, \cdots)$ de l'équation (1), et on peut simuler le nouveau rendement obtenu.

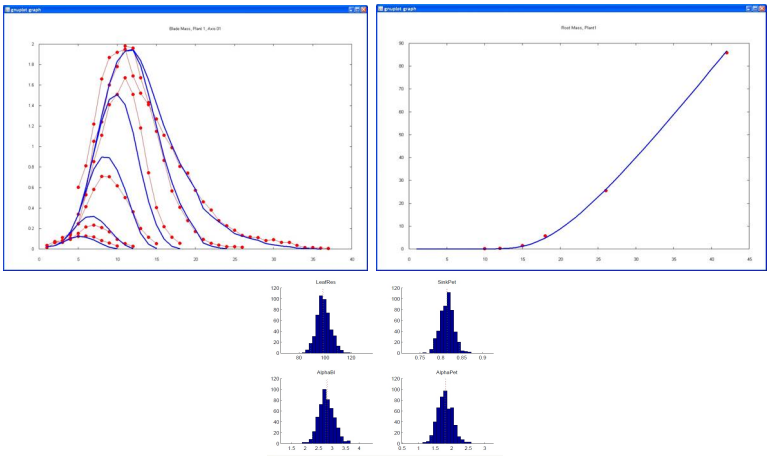


FIGURE 5 – *Estimation des paramètres du modèle GreenLab betterave (Lemaire et al. (2009), collaboration Institut Technique de la Betterave - Centrale Paris). Nous comparons les points expérimentaux (en rouge) avec les simulations du modèle (en bleu). En haut à gauche : les masses des feuilles individuelles, en fonction de leur rang et à différents temps d'observation. En haut à droite : la masse racinaire en fonction du temps. En bas sont données les distributions a posteriori pour 4 paramètres importants du modèle.*

La prise en compte de l'incertitude sur les paramètres permet également de caractériser l'incertitude de prévision.
C'est bien sûr intéressant pour réaliser de l'expérimentation virtuelle : quelle serait la production dans tel climat typique ? Malgré tout, les prévisions n'ont un sens que lorsqu'on ne sort pas du domaine de validité du modèle. En pratique, nous sommes rarement dans une telle situation. Par exemple, si l'on souhaite réaliser une prévision de rendement pour la parcelle d'un agriculteur spécifique, il y a des chances que la caractérisation du sol soit mal connue, que la variété qu'il utilise ne soit pas exactement la même que celle utilisée lors de l'expérimentation ayant permis la calibration du modèle, que les conséquences d'un gel tardif ne soient pas très bien modélisées... Ces différentes imprécisions ou imperfections du modèle peuvent fortement pénaliser la prévision.
Pour y remédier, l'assimilation de données est une méthode statistique permettant de corriger ces différentes erreurs, de façon dynamique, en fonction d'information disponible en cours de culture. L'information qu'on peut utiliser est désormais largement disponible grâce à l'imagerie satellitaire ou drone (Launay and Guerif, 2005) en particulier, mais également tout autre type de capteurs.
Là encore le cadre bayésien et les méthodes de Monte-Carlo séquentielles s'adaptent parfaitement : l'imagerie satellitaire permet par exemple d'évaluer l'indice foliaire à divers moments du cycle de culture, et cette information est assimilée dans le modèle, pour mettre à jour la distribution de probabilité de la prévision.
Les modèles de plus en plus sophistiqués peuvent prédire non seulement le rendement, mais également la qualité de la récolte, comme la qualité du fruit pour la tomate (Génard et al., 2007) ou le raisin, ou la teneur en protéines pour le blé ou l'orge par exemple...
Un élément non mentionné jusqu'à présent en ce qui concerne la prévision de rendement est bien sûr l'incertitude climatique : les modèles de prévision météorologique restent encore d'une précision très limitée à des horizons qui dépassent la semaine. On utilise donc des scénarios historiques ou des générateurs de climats qui permettent de quantifier les variances de prédiction.

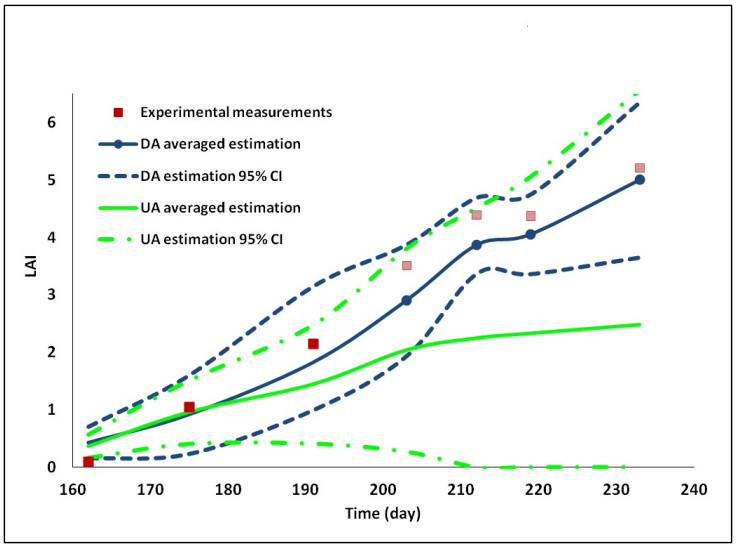


FIGURE 6 – *Assimilation de données de LAI dans le modèle STICS pour le blé (Chen and Cournède, 2014). En carrés rouges et roses sont données des mesures expérimentales du LAI. En vert sont données les prévisions brutes du modèle calibré (moyenne et intervalle de confiance). En bleu, sont données les prévisions du modèle en utilisant les 3 premières données expérimentales (en rouge) pour recalibrer dynamiquement le modèle. On voit que cette assimilation permet de rattraper la bonne dynamique et de considérablement améliorer la précision de la prévision sur les 4 dernières données (en rose, non utilisées pour la calibration).*

La connaissance du potentiel de rendement en fonction des conditions environnementales permet également de tester différentes stratégies d'interventions culturales : quelle dose d'azote doit être optimalement apportée ? Est-il intéressant économiquement de mettre en œuvre un nouveau tour d'irrigation ? Pour une parcelle donnée, quelle est la densité de semis idéale ? Ou encore, quelle est la meilleure stratégie de taille pour un pommier ? Grâce aux modèles, ces différentes questions peuvent être résolues dans le cadre mathématique de l'optimisation en construisant un critère économique et en prenant en compte les contraintes environnementales et logistiques. En serre, la capacité de contrôler de façon plus fine les conditions environnementales pose alors des problèmes de contrôle optimal : la stratégie tout au long du cycle peut être déterminée.

3.3 Interaction Génotype × Environnement

Dans un modèle dynamique, ce sont les paramètres qui caractérisent le système étudié. Il est donc attendu que des espèces différentes soient représentées par des jeux de paramètres différents. L'ambition des modèles de plantes est même d'être capables de différencier les génotypes, c'est à dire que deux génotypes différents soient caractérisés par deux jeux de paramètres différents dans le modèle (Tardieu, 2003).
Un exemple intéressant est donné par (Lecoeur et al., 2011) avec la paramétrisation d'une vingtaine de génotypes de tournesol avec le modèle SUNFLO. Malgré tout, on voit rapidement les difficultés statistiques d'une telle ambition, comme illustrée dans la Figure 7, où l'on montre la difficulté de conclure à une différence significative entre paramètres de façon statistique.
Il est donc important de mettre en œuvre une démarche rigoureuse, s'appuyant sur des jeux de données très fournis pour chaque génotype. Les modèles hiérarchiques (ou à effets

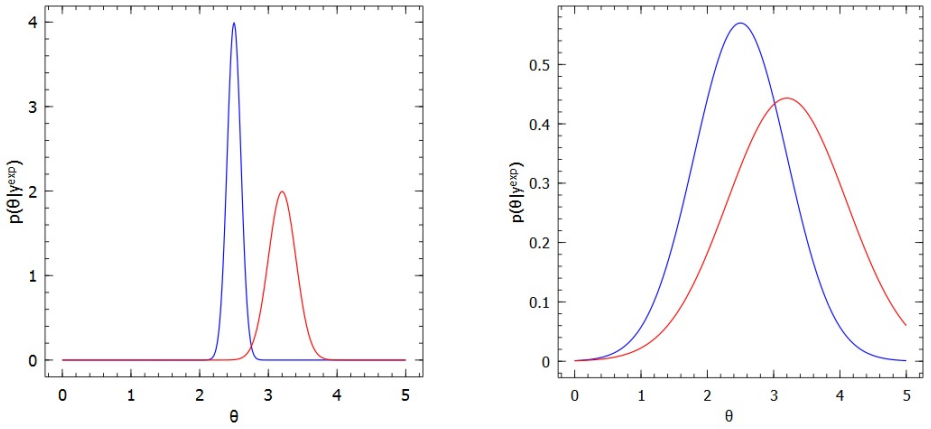


FIGURE 7 – *Distributions de probabilité $p(\theta|\mathbf{y})$ pour un même paramètre θ, et deux génotypes différents (un en bleu et l'autre en rouge). Sur le schéma de gauche, des données détaillées permettent une estimation des paramètres avec des variances faibles, alors que sur le schéma de droite, les variances sont plus grandes, ce qui correspond à une situation plus réaliste : dans la première situation, des tests statistiques seront en mesure de conclure que les paramètres des deux génotypes sont différents, alors que ce ne sera pas possible dans la deuxième situation.*

aléatoires) offrent un cadre statique propice pour l'analyse de la variabilité des paramètres dans une population d'individus ou de génotypes : des paramètres de la distribution dans une population sont identifiées, et des paramètres individuels issus de cette distribution caractérisent chaque individu (Baey et al., 2018). Des tests statistiques permettent alors de déterminer quels paramètres sont stables dans une population ou au contraire variables. Une fois les génotypes caractérisés par des jeux de paramètres différents dans les modèles, l'interaction génotype×environnement se trouve parfaitement décrite, les effets génétiques et environnementaux étant en quelque sorte déconvolués par l'intermédiaire du modèle. Les applications sont alors très intéressantes, en particulier pour l'industrie des semenciers. Ils peuvent par exemple utiliser les modèles pour réaliser des expérimentations virtuelles et considérablement réduire les protocoles expérimentaux généralement mis en œuvre pour tester les nouvelles variétés : une fois le modèle calibré pour une variété spécifique, on peut simuler la performance de la variété dans un grand nombre d'environnements. De la même façon, pour une zone climatique donnée et un environnement type, on peut déterminer les variétés optimales en simulant les performances de toutes les variétés du catalogue. Le semencier peut ainsi affiner son conseil aux agriculteurs.

4 Nouveaux défis

Le degré de maturité des modèles et méthodes présentés précédemment leur permet de rentrer désormais dans une phase d'industrialisation. De grandes entreprises de l'agro-industrie développent des services dédiés à la modélisation agro-environnementales, et des PME du secteur digital se spécialisent sur ce segment. Les efforts de développement portent surtout sur la robustesse des modèles et l'analyse des données de capteurs ou d'imagerie. Un des verrous technologiques important concerne justement l'acquisition de données adaptées à la paramétrisation des modèles de plus en plus fins : il n'est pas possible de calibrer des modèles intégrant une grande connaissance écophysiologique simplement avec de l'ima-

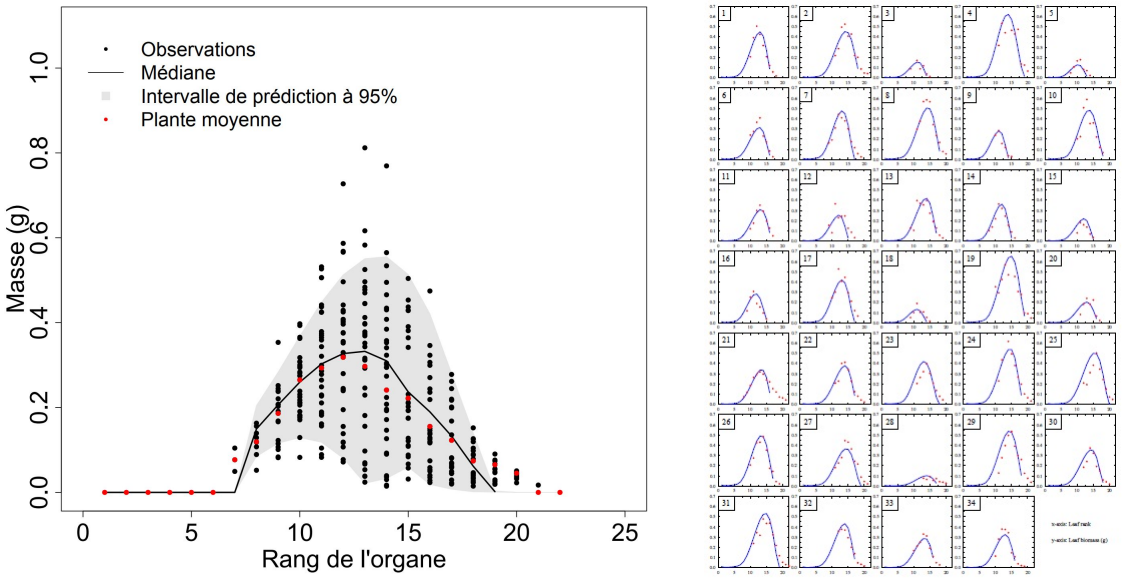


FIGURE 8 – *A gauche : variabilité dans une population de colza des masses des feuilles en fonction du rang - prévision de la moyenne et de l'intervalle de confiance à 95%. A droite : ajustements individuels simultanés pour 34 plantes des masses de feuilles en fonction du rang : données expérimentales en rouge, données simulées par le modèle en bleu (Baey et al., 2018).*

gerie satellitaire... Des efforts expérimentaux importants sont réalisés, en particulier par les instituts agronomiques comme l'INRA, pour acquérir des jeux de données précis et détaillés sur les plantes. La lourdeur des protocoles ne permet pas en général de les répéter dans un grand nombre de situations expérimentales et pour un grand nombre de génotypes, ce qui serait pourtant nécessaire pour convenablement identifier l'interaction génotype×environnement par les modèles.

Dans cette optique, de nombreux projets de plateformes de phénotypage haut-débit ont été développés (Araus and Cairns, 2014) : il s'agit de moyens techniques de mesure permettant de façon automatique d'enregistrer des données environnementales et certaines données sur la plante, souvent en utilisant des photos pour extraire par imagerie certaines caractéristiques morphologiques, hauteur ou surface foliaire. On peut ainsi tester assez rapidement l'impact d'un stress hydrique sur un grand nombre de génotypes. Malgré tout, là aussi, les données restent relativement macroscopiques et ne permettent généralement pas de paramétrer des modèles fins : on va constater des dynamiques de surface foliaire différentes pour divers génotypes, sans expliquer d'où proviennent ces différences.

Pourtant, les modèles décrits précédemment combinant la description de l'architecture de la plante et son fonctionnement offrent une voie intéressante pour mieux utiliser ces types de données : l'architecture de la plante décrite à l'échelle de l'organe est une donnée intégrative très pertinente, car elle est le marqueur de l'histoire de la plante. Comme les cernes visibles lorsqu'on coupe le tronc d'un arbre permettent de visualiser l'incrément en diamètre annuel de son tronc au cours de sa vie, la morphologie de chaque organe intègre la dynamique sources-puits de la croissance au cours de la vie de l'organe. Cela correspond à une masse importante d'informations sur le fonctionnement de la plante que les modèles structure-fonction permettent d'extraire par méthodes inverses (Trevezas and Cournède, 2013). Des méthodes de traitement d'images sont actuellement développées permettant d'identifier pour une plante ses organes individuels par segmentation mais également par

des méthodes de tracking de suivre dans plusieurs photos successives l'évolution morphologique de chaque organe, comme illustré sur l'Arabidopsis dans la Figure 9.

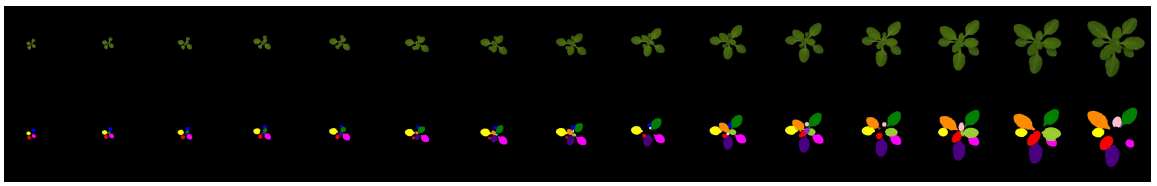


FIGURE 9 – *En haut : série d'images acquises chaque jour par le phénoscope (Tisne et al., 2013) de l'INRA Versailles pour un génotype d'Arabidopsis. En haut : résultat de la segmentation et du tracking réalisés par Viaud et al. (2017)*

Ces données peuvent alors être utilisées pour l'estimation de paramètres des modèles. L'intérêt est que les plateformes de phénotypage haut-débit peuvent potentiellement couvrir un grand nombre de plantes : on se retrouve ainsi en capacité de déterminer de façon automatique les paramètres pour une large famille de génotypes.
Si le génome des plantes est connu (ou un grands nombres de marqueurs génétiques), on peut finalement imaginer combler le fossé qui permet de prédire un phénotype directement à partir de la connaissance du génome de la plante et de son environnement (Bertin et al., 2010). Si pour ce faire des modèles linéaires ont été d'abord considérés (Letort et al., 2008b), les progrès spectaculaires récents des méthodes d'apprentissage profond (deep learning) ouvrent des perspectives prometteuses.
Finalement, nous pouvons imaginer construire la fonction complète de phénotypage virtuel sous la forme :

$$\text{Phénotype} = \text{Modèle de Plante}\,(\text{Paramètres}(\text{Génotype}), \text{Environnement}, \text{Pratiques culturales})\ .$$

La résolution d'une telle chaine permet d'aller plus loin dans les stratégies d'optimisation et de contrôle des cultures, que ce soit au niveau de l'amélioration génétique (définition d'idéotypes pour des environnements types) ou de l'aide à la décision agricole.
Finalement, on voit que les progrès réalisés sur la modélisation de la croissance des plantes et des cultures permettant d'envisager une nouvelle révolution agricole, reposant sur 4 piliers :

- la connaissance biologique et agronomique qui détermine les concepts clés contenus dans les modèles,
- le formalisme mathématique et la simulation numérique permettant de structurer la connaissance biologique, d'extrapoler par le calcul et d'optimiser,
- l'expérimentation agronomique et la structuration des bases de données agro - environnementales,
- les statistiques et l'apprentissage statistique qui permettent d'exploiter ces bases de données pour quantifier, évaluer, prédire.

Chacun de ces domaines est un monde en soit tout en étant nécessaire au succès de la démarche. L'enjeu réside donc dans la capacité à faire coopérer les différents acteurs dans des démarches intégratives et coopératives fortes.

Références

J. L. Araus and J. E. Cairns. Field high-throughput phenotyping : the new crop breeding frontier. *Trends in plant science*, 19(1), 2014.

C. Baey, A. Mathieu, A. Jullien, S. Trevezas, and P.-H. Cournède. Mixed-effects estimation in dynamic models of plant growth for the assessment of inter-individual variability. *Journal of Agricultural, Biological and Environmental Statistics*, 23(2), 2018.

P. Bancal and F. Soltani. Source–sink partitioning. do we need Münch ? *Journal of Experimental Botany*, 53(376), 2002.

D. Barthélémy and Y. Caraglio. Plant architecture : a dynamic, multilevel and comprehensive approach to plant form, structure and ontogeny. *Annals of botany*, 99(3), 2007.

N. Bertin, P. Martre, M. Génard, B. Quilot, and C. Salon. Under what circumstances can process-based simulation models link genotype to phenotype for complex traits ? case-study of fruit and grain quality traits. *Journal of Experimental Botany*, 61(4), 2010.

B. Bouman, H. Van Keulen, H. Van Laar, and R. Rabbinge. The 'school of de wit'crop growth simulation models : a pedigree and historical overview. *Agricultural systems*, 52 (2-3), 1996.

N. Brisson, B. Mary, D. Ripoche, M. H. Jeuffroy, F. Ruget, B. Nicoullaud, P. Gate, F. Devienne-Barret, R. Antonioletti, C. Durr, et al. STICS : a generic model for the simulation of crops and their water and nitrogen balances. I. Theory and parameterization applied to wheat and corn. *Agronomie*, 18(5-6), 1998.

Y. Chen and P.-H. Cournède. Data assimilation to reduce uncertainty of crop model prediction with convolution particle filtering. *Ecological Modelling*, 290 :165–177, 2014.

P.-H. Cournède, M.-Z. Kang, A. Mathieu, J.-F. Barczi, H.-P. Yan, B.-G. Hu, and P. De Reffye. Structural factorization of plants to compute their functional and architectural growth. *Simulation*, 82(7), 2006.

P.-H. Cournède, V. Letort, A. Mathieu, M. Z. Kang, S. Lemaire, S. Trevezas, F. Houllier, and P. De Reffye. Some parameter estimation issues in functional-structural plant modelling. *Mathematical Modelling of Natural Phenomena*, 6(2), 2011.

P.-H. Cournède, Y. Chen, Q. Wu, C. Baey, and B. Bayol. Development and evaluation of plant growth models : Methodology and implementation in the pygmalion platform. *Mathematical Modelling of Natural Phenomena*, 8(4), 2013.

S. de Dorlodot, B. Forster, L. Pagès, A. Price, R. Tuberosa, and X. Draye. Root system architecture : opportunities and constraints for genetic improvement of crops. *Trends in plant science*, 12(10), 2007.

P. De Reffye and B.-G. Hu. Relevant qualitative and quantitative choices for building an efficient dynamic plant growth model : Greenlab case. In *International Symposium on Plant Growth Modeling, Simulation, Visualization and their Applications-PMA'03*. Springer and Tsinghua University Press, 2003.

P. de Reffye, C. Edelin, J. Françon, M. Jaeger, and C. Puech. Plant models faithful to botanical structure and development. In *Proceedings of the 15th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '88. ACM, 1988.

P. de Reffye, T. Fourcaud, F. Blaise, D. Barthélémy, and F. Houllier. A functional model of tree growth and tree architecture. *Silva Fennica*, 31(3), 1997.

C. De Wit, H. Van Laar, and H. Van Keulen. Physiological potential of crop production. In *Plant breeding perspectives*. Pudoc, 1979.

L. Dupuy, P. J. Gregory, and A. G. Bengough. Root growth models : towards a new generation of continuous approaches. *Journal of experimental botany*, 61(8), 2010.

G. Farquhar and S. Von Caemmerer. Modelling of photosynthetic response to environmental conditions. In *Physiological plant ecology II*, pages 549–587. Springer, 1982.

C. Fournier and B. Andrieu. A 3d architectural and process-based model of maize development. *Annals of botany*, 81(2), 1998.

M. Génard, N. Bertin, C. Borel, P. Bussières, H. Gautier, R. Habib, M. Lechaudel, A. Lecomte, F. Lescourret, P. Lobit, and B. Quilot. Towards a virtual fruit focusing on quality : modelling features and potential uses. *Journal of Experimental Botany*, 58(5) : 917–928, 2007.

F. Hallé and R. Oldeman. *Essai sur l'architecture et la dynamique de croissance des arbres tropicaux*. Masson, Paris, 1970.

C. A. Jones, J. R. Kiniry, and P. Dyke. *CERES-Maize : A simulation model of maize growth and development*. Texas AM University Press, 1986.

M. Kang, P.-H. Cournède, P. de Reffye, D. Auclair, and B. Hu. Analytical study of a stochastic plant growth model : application to the greenlab model. *Mathematics and Computers in Simulation*, 78(1), 2008.

M. Launay and M. Guerif. Assimilating remote sensing data into a crop model to improve predictive performance for spatial applications. *Agriculture, ecosystems & environment*, 111, 2005.

J. Lecoeur, R. Poiré-Lassus, A. Christophe, B. Pallas, P. Casadebaig, P. Debaeke, F. Vear, and L. Guilioni. Quantifying physiological determinants of genetic variation for yield potential in sunflower. SUNFLO : a model-based analysis. *Functional plant biology*, 38 (3), 2011.

S. Lemaire, F. Maupas, P. Cournède, and P. De Reffye. A morphogenetic crop model for sugar-beet (beta vulgaris l.). In *Crop Modeling and Decision Support*. Springer, 2009.

V. Letort, P.-H. Cournède, A. Mathieu, P. de Reffye, and T. Constant. Parametric identification of a functional-structural tree growth model and application to beech trees (*Fagus sylvatica*). *Functional Plant Biology*, 35(10), 2008a.

V. Letort, P. Mahe, P.-H. Cournède, P. De Reffye, and B. Courtois. Quantitative genetics and functional–structural plant growth models : simulation of quantitative trait loci detection for model parameters and application to potential yield optimization. *Annals of Botany*, 101(8), 2008b.

C. Loi and P. H. Cournède. Generating Functions of Stochastic L-Systems and Application to Models of Plant Development. *Discrete Mathematics & Theoretical Computer Science*, DMTCS Proceedings vol. AI, Fifth Colloquium on Mathematics and Computer Science, 2008.

A. Mathieu, P.-H. Cournède, V. Letort, D. Barthélémy, and P. de Reffye. A dynamic model of plant growth with interactions between development and functional mechanisms to study plant structural plasticity related to trophic competition. *Annals of Botany*, 103 (8), 2009.

R. McCown, G. Hammer, J. Hargreaves, D. Holzworth, and D. Freebairn. Apsim : a novel software system for model development, model testing and simulation in agricultural systems research. *Agricultural systems*, 50(3), 1996.

P. Minchin and M. Thorpe. What determines carbon partitioning between competing sinks ? *Journal of Experimental Botany*, 47, 1996.

J. L. Monteith and C. Moss. Climate and the efficiency of crop production in britain. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 281 (980), 1977.

L. Pagès, G. Vercambre, J.-L. Drouet, F. Lecompte, C. Collet, and J. Le Bot. Root typ : a generic model to depict and analyse the root system architecture. *Plant and Soil*, 258 (1), 2004.

P. Prusinkiewicz and A. Lindenmayer. *The algorithmic beauty of plants*. Springer Science & Business Media, 1990.

P. Prusinkiewicz, A. Lindenmayer, and J. Hanan. Development models of herbaceous plants for computer imagery purposes. In *Proceedings of the 15th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '88. ACM, 1988.

D. d. Pury and G. Farquhar. Simple scaling of photosynthesis from leaves to canopies without the errors of big-leaf models. *Plant, Cell & Environment*, 20(5), 1997.

H. Scholten, A. Kassahun, J. C. Refsgaard, T. Kargas, C. Gavardinas, and A. J. Beulens. A methodology to support multidisciplinary model-based water management. *Environmental Modelling & Software*, 22(5), 2007.

R. Sievänen, E. Nikinmaa, J. Perttunen, et al. Evaluation of importance of sapwood senescence on tree growth using the model lignum. *Silva Fennica*, 31(3), 1997.

A. R. Smith. Plants, fractals, and formal languages. *ACM SIGGRAPH Computer Graphics*, 18(3), 1984.

K. A. Sorrensen-Cothern, E. D. Ford, and D. G. Sprugel. A model of competition incorporating plasticity through modular foliage and crown development. *Ecological Monographs*, 63(3), 1993.

F. Tardieu. Virtual plants : modelling as a tool for the genomics of tolerance to water deficit. *Trends in plant Science*, 8(1), 2003.

S. Tisne, Y. Serrand, L. Bach, E. Gilbault, R. Ben Ameur, H. Balasse, R. Voisin, D. Bouchez, M. Durand-Tardif, P. Guerche, et al. Phenoscope : an automated large-scale phenotyping platform offering high spatial homogeneity. *The Plant Journal*, 74(3), 2013.

S. Trevezas and P.-H. Cournède. A sequential monte carlo approach for mle in a plant growth model. *Journal of Agricultural, Biological, and Environmental Statistics*, 18(2), 2013.

F. Varenne. La simulation informatique face à la «méthode des modèles». Le cas de la croissance des plantes. *Nature Sciences Sociétés*, 11(1), 2003.

G. Viaud, O. Loudet, and P.-H. Cournède. Leaf segmentation and tracking in arabidopsis thaliana combined to an organ-scale plant model for genotypic differentiation. *Frontiers in plant science*, 7, 2017.

Q. Wu, P.-H. Cournède, and A. Mathieu. An efficient computational method for global sensitivity analysis and its application to tree growth modelling. *Reliability Engineering & System Safety*, 107, 2012.