

PROJET ENJEU : SANTÉ ET BIOTECHNOLOGIES

RAPPORT FINAL

Développement d'outils mathématiques pour l'agriculture de précision

Client et Référent Pédagogique

Pierre CARMIER
Paul-Henry Cournède

P2018 : groupe SBT-11

John DE WASSEIGE
Nayef DERWICHE
Alexis MATHEY
Daina ZHENG

10 décembre 2015

Table des matières

1	Résumé	2
2	Introduction	3
3	Synthèse de l'étude documentaire	4
4	Méthodologies	5
4.1	Gestion de la bibliographie	5
4.2	Outils de travail collaboratif	5
4.3	Organisation et partage des tâches	6
4.4	Estimation des paramètres	7
4.5	Markov	8
5	Description du fonctionnement du code	9
6	Conclusion	10
	Remerciements	11
	Bibliographie	12
A	Physiologie des plantes	13
A.1	Généralités	13
A.2	Photosynthèse	13
B	Histoire de la modélisation de la croissance des plantes	15
B.1	Débuts de la botanique et de l'agronomie	15
B.2	Les premiers modèles	16
C	Développement et évaluation des modèles de croissance des plantes	18
C.1	Généralités	18
C.1.1	Caractéristiques propres aux modèles de croissance de plantes	18
C.1.2	Bonnes pratiques en modélisation	18
C.2	Modèle LNAS appliqué à la betterave	19
D	Quelques modèles de la croissance des plantes	21
D.1	Informatique et modèles géométriques	21
D.2	Le modèle GreenLab	23

1 Résumé

2 Introduction

Ce rapport a été réalisé dans le cadre du projet enjeu Santé et Biotechnologies à CentraleSupélec. Le but premier de ce projet est de se retrouver face à un problème ouvert où il n'existe pas forcément de solution exacte comme nous sommes habitués à avoir dans nos autres cours. Un autre but poursuivi est de se retrouver en situation concrète de collaboration avec un client extérieur, ce qui change complètement du cadre habituel des travaux de groupes. Finalement, ce projet nous a appris à travailler en autonomie ainsi qu'à mener des recherches selon une démarche scientifique.

Notre projet enjeu se concentre sur le *développement d'outils mathématiques pour l'agriculture de précision* avec l'équipe Digiplante. La première partie de ce titre est assez compréhensible, la deuxième cependant suscite de nombreuses questions qui méritent des réponses. La croissance exponentielle de la population au cours des dernières années a obligé la science à avancer dans le domaine de l'agriculture. Le contexte de réchauffement climatique oblige également à suivre une démarche scientifique qui se veut respectueuse de l'environnement.

3 Synthèse de l'étude documentaire

Cette partie se présente comme une synthèse de l'analyse bibliographique que nous avons réalisée en décembre. Le lecteur intéressé trouvera les éléments des paragraphes qui suivent respectivement en annexe [A](#), en annexe [B](#), en annexe [C](#), en annexe [D](#).

Nous nous étions d'abord intéressé à la physiologie de la plante. Cela nous a permis d'appréhender les éléments essentiels à son développement. L'élément crucial qui permet le développement d'une plante est la photosynthèse. En effet, la plante possède cette capacité extraordinaire de synthétiser de la matière organique à partir de matière minérale. Parce qu'ils interviennent dans la réaction de photosynthèse, l'eau, le soleil... ils sont des éléments clés de la modélisation du développement d'une plante.

Grâce à l'histoire de l'étude des plantes, nous avons également pu comprendre le contexte dans lequel la modélisation des plantes s'inscrit aujourd'hui. C'est l'essor des mathématiques et de l'informatique qui permettent d'envisager demain une modélisation efficace, avec des applications comme l'agriculture de précision. En effet, les mathématiques ont peu à peu formalisé la connaissance des plantes pour obtenir des modélisations plus précises, en introduisant des notions comme le temps thermique, le LUE, le PAR...

Nous nous étions également intéressés aux caractéristiques propres à la modélisation des plantes, afin de comprendre les principales difficultés qui empêchent une utilisation globale des modèles de croissance des plantes. Parmi eux, on peut citer : l'estimation difficile des paramètres d'un modèle, des techniques sophistiquées en informatique et en mathématique sont nécessaires et une diversité importante sans benchmarking des modèles ne permet pas de choisir facilement celui qui sera le plus adapté.

Nous nous sommes ensuite intéressés aux modèles les plus répandus pour décrire le développement de la plante. Parmi eux, la grammaire L-système, le modèle GreenLab et le modèle LNAS appliqué à la betterave. La description fournie par les L-systèmes étant trop précise, ils conduisent à des temps de calcul trop importants et il est difficile d'estimer les nombreux paramètres qui interviennent. Le modèle GreenLab introduit une alternative intéressante : plutôt que de décrire l'architecture individuelle au niveau d'un champ, les organes (les feuilles, les grains...) sont remplacés par des pools de biomasse. On ne s'intéresse plus qu'à la masse totale des feuilles, des grains, des tiges... sur un champ. Un pool de biomasse commun permet de décrire la biomasse obtenue grâce à la photosynthèse. L'enjeu est maintenant de pouvoir décrire comment la biomasse sera répartie entre les différents organes de la plante à un instant t . Le modèle LNAS appliqué à la betterave nous a ainsi permis de nous familiariser avec un premier modèle de manière plus concrète, grâce aux formules d'allocation de biomasse de ce modèle.

4 Méthodologies

4.1 Gestion de la bibliographie

Pour la gestion de la bibliographie au sein du document, nous utilisons le package `biblatex`. Celui-ci permet d'écrire l'ensemble de nos références dans un fichier `.bib` sous la forme suivante.

```
@online{histoire_mod_plantes,  
  title = {Une histoire de la modélisation des plantes},  
  author = {Philippe de Reffye and Marc Jaeger  
and Paul-Henry Cournède},  
  url = {https://interstices.info/jcms/c_38032/une-histoire-de-  
la-modelisation-des-plantes},  
  year = {2009},  
  month = "04",  
}
```

La mise en page est alors automatique en fonction des informations fournies et le rendu de l'exemple est présenté ci dessous.

- [3] Philippe de REFFYE, Marc JAEGER et Paul-Henry COURNÈDE. *Une histoire de la modélisation des plantes*. Avr. 2009. URL : https://interstices.info/jcms/c_38032/une-histoire-de-la-modelisation-des-plantes.

Cela paraît à priori assez lourd d'écrire soi-même toutes les informations en suivant cette syntaxe mais il existe des logiciels comme Zotero qui font le travail à notre place. Les sources trouvées sur Google Scholar peuvent également être exportées aisément au format BibTex.

4.2 Outils de travail collaboratif

Pour des raisons pratiques et esthétiques, nous avons décidé d'écrire nos rapports en \LaTeX . Il s'agissait donc de trouver la meilleure façon de partager le code source et de pouvoir contrôler les changements apportés au document. Une première idée pourrait être d'utiliser ShareLaTeX qui propose une plate-forme de compilation en ligne ainsi qu'un système de gestion de versions assez simple à utiliser. Nous n'avons pas choisi cette solution notamment pour les raisons suivantes. L'utilisateur doit être connecté dès qu'il veut travailler sur le projet, le système de compilation est assez lent et l'utilisateur n'est pas libre d'utiliser son éditeur de texte ou son visualisateur de PDF favori.

Pour palier aux problèmes décrits ci-dessus, le logiciel `git` associé à GitHub est une très bonne alternative. Il permet en effet à chaque membre du groupe de travailler sans être connecté ainsi que d'utiliser son éditeur et compilateur favori. Chaque membre travaille donc de son côté en faisant des *commits* et lorsqu'il juge

que son travail est utile pour les autres, il *push* sur le serveur. L'algorithme de fusion, *merge*, permet également de fusionner intelligemment les lignes d'un fichier qui ont été modifiées par plusieurs membres. Le dernier point à souligner est que `git` permet une gestion des branches, particulièrement pratique lorsqu'on veut développer une partie du projet sans risquer de créer des erreurs dans le programme principal.

Nous combinons donc ces deux outils pour

1. implémenter le modèle LNAS blé dans la plateforme Pygmalion en JULIA (une description plus détaillée de l'objectif attendu pour cette partie est décrite dans la section ?? à la page ??) pour le client dont le code source est sur la plateforme GitLab,
2. rédiger l'étude documentaire en partageant le code L^AT_EX à l'aide d'un dossier sur GitHub¹.

4.3 Organisation et partage des tâches

Il nous reste un dernier point à décrire, celui de la *communication* au sein du groupe. Nous utilisons Slack² qui est un logiciel de plus en plus utilisé pour les travaux de groupe ainsi que dans les start-ups. Il permet d'éviter de devoir alterner entre plusieurs applications comme les mails, DropBox et Twitter, puisqu'il permet d'être connecté à celles-ci au sein de l'application. On peut également créer plusieurs *channels* pour séparer la communication entre les différentes tâches. Par exemple dans ce projet nous avons les *channels* suivantes : **general**, **etude-documentaire**, **planning** et **dev_plate-forme**. On trouve aussi un système d'historique et de gestion de fichiers efficace.

L'ensemble des tâches ainsi que leur répartition pendant l'année est associé au planning GANTT qui se trouve dans l'Annexe ?? à la page ??.

1. <https://github.com/jdewasseige/projet-sbt11>

2. <https://slack.com/>

4.4 Estimation des paramètres

Maintenant que le modèle est implémenté, il faut pouvoir estimer les paramètres qui interviennent dans le modèle, pour pouvoir utiliser le dit modèle.

Pour cela, nous utilisons une méthode déjà implémentée sur la plateforme “lgs”, pour Generalized Least Squares (méthode des moindres carrés généralisée).

A partir des données expérimentales d’un champ de blé par exemple, cette méthode cherche les valeurs, les paramètres qui font converger les valeurs mesurées et les valeurs calculées par le modèle. La méthode des moindres carrés généralisée est une méthode classique pour l’estimation de paramètres dans le cas de modèles déterministes, et c’est le cas ici puisque nous n’avons pas implémenté de bruit aléatoire dans notre modèle.

La méthode minimise la probabilité que les données expérimentales soient différentes de ce qui est prédit par le modèle. Mais, les calculs ne peuvent être conduits qu’en affectant une valeur arbitraire aux paramètres. On obtient ainsi une nouvelle valeur des paramètres, avec lesquelles on calcule de nouvelles valeurs des paramètres. Et ainsi de suite, jusqu’à que la valeur des paramètres converge.

L’avantage de cette méthode est son efficacité pour déterminer rapidement et avec une assez bonne précision les valeurs cherchées. De plus, le fait que cette méthode soit déjà implémentée sur la plateforme est un autre avantage, puisque cela nous permet de l’utiliser sans avoir à l’implémenter.

4.5 Markov

Ces modèles sont des *modèles de Markov* à temps discret. Soit $(t_n)_{n \in [0, N]}$, la séquence finie des temps successifs correspondants aux étapes de l'évolution. On note $X_n \in \mathbb{R}^d$ l'ensemble des variables caractéristiques du système à t_n , $U_n \in \mathbb{R}^u$ l'ensemble des variables exogènes (entrées, contrôle...) à t_n , et $P \in \mathbb{R}^p$ le vecteur de paramètres du modèle. Comme pour la plupart des systèmes biologiques, X_n peut ne pas être complètement accessible à l'observation, et on note donc $Y_n \in \mathbb{R}^{q_n}$ l'ensemble observé/mesuré des variables au temps n . On note $Y = Y_n, n \in [0, N]$. La fonction de densité initiale pour X_0 est μ_p et la densité de transition de Markov est f_{n,P,U_n} :

$$X_0 \sim \mu_p \quad \text{et} \quad X_{n+1} \mid (X_n = x) \sim f_{n,P,U_n}(\cdot \mid x) \quad \forall n \in [0; N - 1]$$

L'observation Y_n dépend de X_n et la densité conditionnelle est donnée par $g_{n,P}$:

$$Y_n \mid (X_n = x) \sim g_{n,P}$$

Ce cadre stochastique permet d'aboutir et de décrire les modèles dynamiques discrets déterministes, en écrivant :

$$\begin{cases} X_{n+1} &= F_n(X_n, U_n, P) \\ Y_n &= G_n(X_n, P) \end{cases}$$

describes how an important category of plant growth models can be set in this framework. For example, for functional-structural models that describe biomass budget during plant growth (see for example LIGNUM [52] or GREENLAB [46]), the state variables correspond to daily biomass accumulation and to masses of plant organs according to their categories, the parameters are genotype specific, and the external variables U_n correspond to environmental variables (radiation, temperature, soil water content ...).

Generally, not all the state variables can be observed experimentally (for example daily biomass production) and the experimentation being heavy (specifically when it comes to the masses of individual organs), observations are not done at all time steps. If we denote by \mathcal{O} the set of all time step indexes corresponding to observation stages :

$$\mathcal{O} = \{i \in [1; N] \text{ où } t_i \text{ est un temps d'observation}\}$$

we then have $q_i > 0$ if and only if $i \in \mathcal{O}$ (where we recall that q_i is the dimension of Y_i). Note also that the non-zero q_i have no reason to be identical (as illustrated for example in [45] for a model of maize growth, in which at some stages individual plants were measured at organ level, and at other stages only compartment data were available, corresponding to different G_i).

Algorithm 1 Algorithme générique qui sert de base pour l’implémentation des fonctions. La fonction f n’est pas définie mais sert de placeholder pour représenter les opérations nécessaires à la mise à jour de `xnplus1`.

```

1: procedure GENERIC(int  $n$ , State Vector  $xn$ , Control Vector  $u$ , Parameters Vector  $p$ , State Vector  $xnplus1$ )
2:   begin
3:      $xnplus1.composante \leftarrow f(p.composante, u.composante, xn.composante)$ 
4:   end
5:   return None
6: end procedure

```

5 Description du fonctionnement du code

Dans un premier temps, nous avons construit un modèle informatique à partir du modèle mathématique fourni [3].

Il s’agissait d’abord d’assigner à chaque variable du modèle un représentant dans notre programme, en rendant ces derniers assez explicites afin d’être plus efficace pour l’implémentation des fonctions. Les tableaux 1, 2 et 3 contiennent respectivement les variables d’état auxiliaires et principales, les variables *environnementales* et les *paramètres* de la tige, du sol et des racines.

La deuxième partie du travail consistait à définir les différentes fonctions qui mettent à jour les variables d’état et paramètres du temps n au temps $n + 1$. L’implémentataion de chacune d’entre elle est basée sur l’algorithme générique 1. On retrouve à l’entrée les variables d’états `xn` et `xnplus1`, le temps `n`, les variables de contrôles `u`, les paramètres `p` et éventuellement des informations supplémentaires qui dépendent du rôle précis de la fonction. On va ensuite mettre à jour une composante de `xnplus1` en suivant la description du modèle.

À titre d’exemple, on va utiliser la fonction `get_pot_evaporation` qui met à jour l’évaporation requise en fonction des conditions environnementales selon l’équation 1.

$$Espot^{(n)} = K_s ET0 \exp -\lambda LAI^{(n)} \quad (1)$$

En suivant la démarche 1, on obtient en JULIA l’implémentation suivante.

```

1 function get_pot_evaporation!(n, xn, u, p, xnplus1)
2   xnplus1.soil_req_evaporation =
3     p.k_s * u.ET0[n] * exp( -p.lambda * xn.leaf_area_index)
4 end

```

6 Conclusion

Remerciements

Nous tenons tout d'abord à remercier notre client Pierre Carmier, pour sa pédagogie et son temps précieux accordé. Nous remercions également Mme Le Chevalier et Mme Lopes, qui encadrent le projet de façon motivante et efficace. Leurs remarques et retours sur nos présentations ainsi que sur l'avancement de notre projet nous ont été très utiles pour prendre le recul nécessaire à la rédaction de ce rapport.

Références

- [1] VisioDrones SA. Service d'imagerie par système de drones. URL : <http://www.visiodrones.com/produits/vegedrones/>.
- [2] Charlotte BAEY. « Modelling inter-individual variability in plant growth models and model selection for prediction ». Theses. Ecole Centrale Paris, fév. 2014. URL : <https://tel.archives-ouvertes.fr/tel-00985747>.
- [3] Pierre CARMIER. « LNAS model for wheat with soil coupling ». In : (nov. 2015).
- [4] Paul-Henry COURNÈDE et al. « Development and Evaluation of Plant Growth Models : Methodology and Implementation in the PYGMALION platform ». In : *Mathematical Modelling of Natural Phenomena* 8.4 (juin 2013), p. 112–130. URL : <https://hal-ecp.archives-ouvertes.fr/hal-00860902>.
- [5] Martine GUÉRIF et Dominique KING. *Agriculture de précision*. Éditions Quæ, 2007.
- [6] William G HOPKINS. *Physiologie végétale*. De Boeck Supérieur, 2003.
- [7] A JULLIEN et P HUET. *Agriculture de précision*. 2005. URL : <https://tice.agroparistech.fr/coursenligne/courses/SIAFEEAGRONOMIE9cea/document/machinisme/themes/documents/synthese.pdf>.
- [8] *Les émissions agricoles de particules dans l'air*. Agence de l'Environnement et de la Maîtrise de l'Energie. URL : http://www.chambres-agriculture.fr/fileadmin/user_upload/thematiques/Produire_durablement/Energies_biomasse/Emissions_agricoles_particules.pdf.
- [9] Przemyslaw PRUSINKIEWICZ et Aristid LINDENMAYER. *The algorithmic beauty of plants*. Springer Science & Business Media, 2012.
- [10] *Quatre gigantesques plans d'eau menacés de disparition*. 2014. URL : http://www.francetvinfo.fr/monde/environnement/en-images-quatre-gigantesques-plans-d-eau-menaces-de-disparition_708841.html.
- [11] *Quelle contribution de l'agriculture française à la réduction des émissions de gaz à effet de serre ?* 2013. URL : <http://institut.inra.fr/Missions/Eclairer-les-decisions/Etudes/Toutes-les-actualites/Etude-Reduction-des-GES-en-agriculture>.
- [12] Philippe de REFFYE, Marc JAEGER et Paul-Henry COURNÈDE. *Une histoire de la modélisation des plantes*. Avr. 2009. URL : https://interstices.info/jcms/c_38032/une-histoire-de-la-modelisation-des-plantes.
- [13] WIKIPEDIA. *Agriculture de précision* — *Wikipedia, The Free Encyclopedia*. URL : https://fr.wikipedia.org/wiki/Agriculture_de_pr%C3%A9cision.
- [14] WIKIPEDIA. *Population mondiale* — *Wikipedia, The Free Encyclopedia*. URL : https://fr.wikipedia.org/wiki/Population_mondiale.
- [15] WIKIPEDIA. *Stomate* — *Wikipedia, The Free Encyclopedia*. URL : <https://fr.wikipedia.org/wiki/Stomate>.
- [16] Philippe ZWAENEPOEL et Jean-Michel LE BARS. « L'agriculture de précision ». In : *Ingénieries-EAT* 12 (1997), p-67.

Annexes

A Physiologie des plantes

A.1 Généralités

Tout d'abord, présentons succinctement comment se développe et fonctionne une plante. La science qui décrit l'ensemble des mécanismes qui participent à l'édification d'un organisme vivant est appelée morphogénèse

Chez les plantes, la morphogénèse commence avec la germination de la graine et s'arrête à la mort de la plante. [2, p. 22]. Celle-ci diffère d'une plante à l'autre, et dépend également de l'environnement de la plante.

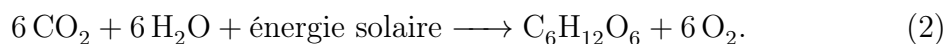
Ce sont les méristèmes³ qui permettent à la plante de se développer, en allongeant des organes déjà existants ou en créant de nouveaux organes. Lors de la germination de la graine, au stade embryonnaire, des méristèmes permettent déjà le développement de la plante (méristèmes racinaires, caulinaires). On distingue les méristèmes végétatifs, à l'origine des organes végétatifs (tige, feuilles, racines) et les méristèmes reproducteurs, à l'origine des fleurs. On distingue également les méristèmes primaires, qui permettent à la plante de croître en longueur (tiges et racines) et les méristèmes secondaires, qui permettent la croissance en épaisseur de la plante. La création de nouveaux organes est réalisée par alignement de nouvelles briques élémentaires (formée par les méristèmes), constituées :

- d'un noeud, auquel sont liés les feuilles
- d'un entrenoeud
- d'un bourgeon situé à la base du noeud, à l'aisselle des feuilles

Ces briques élémentaires sont appelées phytomères. Cette caractéristique permet de simplifier la modélisation de la croissance d'une plante.

A.2 Photosynthèse

Sans doute que la particularité la plus intéressante des plantes, et qui justifie le mieux leur étude est leur capacité à synthétiser de la matière organique (des glucides...) à partir de CO_2 , d'eau et d'énergie solaire. C'est la célèbre photosynthèse qui permet ainsi à la plante de transformer de la matière *minérale* en matière *organique*, selon l'équation



Tous les éléments de la plante participent à ce processus de photosynthèse :

- les racines puisent dans le sol l'eau et les sels minéraux nécessaires

3. Le méristème est un tissu formé de cellules indifférenciées (embryonnaires) qui se divisent activement et permettent la croissance.

- les feuilles captent l'énergie solaire et le dioxyde de carbone (grâce aux cellules chlorophylliennes et aux stomates ⁴)

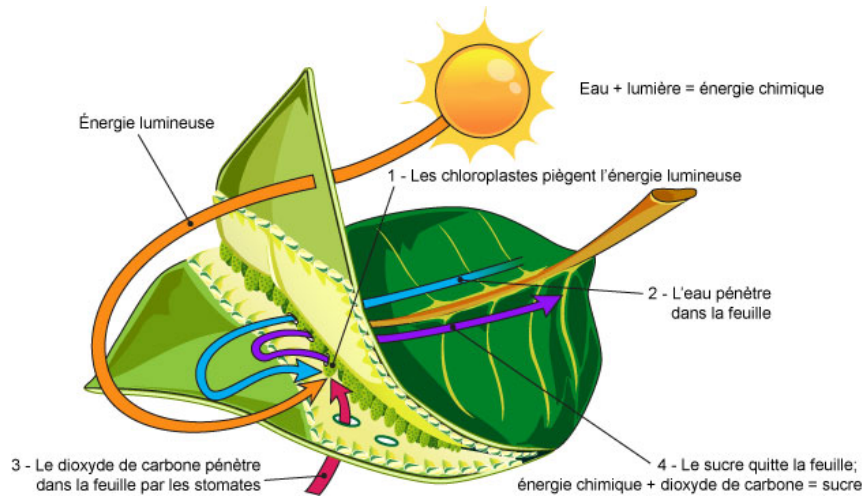


FIGURE 1 – Schéma de la photosynthèse illustrant les différents mécanismes qui permettent à celle-ci d'avoir lieu.

Les sucres ainsi formés apportent l'énergie nécessaire au fonctionnement de la plante et assurent son développement en permettant la synthèse de cellulose qui est l'élément constitutif principal des plantes. On identifie ainsi les éléments qui agissent sur la croissance de la plante :

- la lumière
- l'eau
- le dioxyde de carbone
- la température : car la température agit sur l'ouverture des stomates et donc sur le flux des échanges gazeux
- l'azote, qui permet à la plante de construire les acides aminés nécessaires à l'élaboration des protéines
- d'autres minéraux, comme le potassium qui favorise les transferts au sein de la plante

Tous les organes de la plante s'unissent donc pour aboutir à la production de biomasse. Cette biomasse est ensuite distribuée aux organes pour permettre leur développement.

Parce que ce mécanisme permet de synthétiser de la matière minérale en matière organique et de capter du CO_2 , gaz à effet de serre notoire, il est crucial de comprendre ce mécanisme de photosynthèse. C'est pourquoi il est, au même titre que la croissance des plantes l'objet de nombreuses recherches (avec comme application : créer de l'électricité propre en dissociant H_2O en Oxygène et en Hydrogène, capter du CO_2 ...

4. Orifice de petite taille situé sur les feuilles qui permet les échanges gazeux entre l'air et la plante.

B Histoire de la modélisation de la croissance des plantes

Nous présentons dans cette partie tout d’abord l’origine de la botanique et de l’agronomie. Les premiers modèles de croissance des plantes seront ensuite décrits, puis on expliquera les avancées potentielles apparues avec l’avènement de l’informatique. Finalement, un modèle particulièrement intéressant, le modèle GreenLab, sera présenté.

Cette section reprend la démarche suivie dans l’article *Une histoire de la modélisation des plantes*, COURNÈDE et al., 2009 [12].

B.1 Débuts de la botanique et de l’agronomie

L’étude des plantes a très tôt été un domaine privilégié du savoir humain. En effet, les plantes sont un élément majeur des écosystèmes dans lesquels l’homme évolue. Elles sont source de nourriture, de remèdes, de médicaments, de matériaux, d’esthétique. Enfin, elles sont un objet scientifique d’intérêt qui a très tôt aiguïlé le sens de l’observation, l’esprit d’analyse, de synthèse, de déduction des hommes. La connaissance des plantes s’est accrue lors de l’histoire des hommes, qui ont développé la cueillette, l’agriculture, l’usage des plantes médicinales. La connaissance et l’inventaire des variétés de plante ont ainsi été des enjeux majeurs car ils permettaient la connaissance de nouveaux remèdes et étaient sources de nourritures et matériaux. L’homme a ainsi cherché à regrouper, croiser, faire croître et conserver les espèces qui lui étaient utiles.

La botanique, issue de l’étude de l’anatomie des plantes, est une science très ancienne. En témoignent les traités de classification de plantes, comme ceux d’Aristote (vers -300), ou encore l’inventaire et la description de centaines de plantes médicinales par Dioscoride (1er siècle), ainsi que les traités chinois qui inventorient les espèces utiles à l’agriculture et à la médecine traditionnelle avec de premiers efforts de classification. Efforts de classification qui se poursuivront vraiment en Europe à partir du XVII^{ème} siècle avec les premières distinctions par famille, par genre, par espèce, par structure de graine (*Les éléments de Botanique* par Joseph Pitton de Tournefort en 1656 – 1708, *Systema naturae* en 1735 et *Philosophia botanica* en 1751 par Linné et les travaux de la famille de Jussieu pendant le XVIII^{ème} siècle). Ces classifications ne sont pas objectives, elles sont le fruit d’un raisonnement *empirique*.

En parallèle, l’agronomie se développe au XVII^{ème} siècle en Europe et s’intéresse au processus de croissance et développement des plantes. Des travaux d’abord très pratiques sont réalisés sur des méthodes agricoles (labour, ensemencement, taille, greffes...), notamment ceux de Jean-Baptiste de la Quintinie et d’Olivier de Serres.

Au XIX^{ème}, les processus biologiques commencent à être étudiés de façon plus précise, en particulier la provenance du carbone, de l’azote, de l’oxygène et de l’eau dans la plante. On s’intéresse également aux problématiques de nutrition et au rôle de organes, comme en témoignent les ouvrages *Recherches chimiques sur la végétation* de

Théodore de Saussure en 1804. Quelques années plus tard, on découvre la respiration, la photosynthèse (voir équation 2 à la page 13).

La physiologie, science qui étudie le fonctionnement des plantes, se sépare alors de la botanique qui se contente de les classer.

B.2 Les premiers modèles

La modélisation mathématique précise (qui va au-delà de la simple description qualitative et fournit une description quantitative avec des capacités prédictives) n'arrive pas tout de suite en biologie. Le développement de la biologie n'a pas suivi le même schéma que celui de nombreuses autres sciences comme la physique, où l'observation a permis de tirer des concepts quantitatifs au niveau macroscopique (loi de Mariotte par exemple) avant de les expliquer par des lois qui s'appliquent au niveau microscopique (Boltzmann). De même pour la mécanique, l'optique, l'électricité avec des applications qui n'ont pas eu à attendre la compréhension au niveau atomique. La biologie végétale par contre a paradoxalement été mieux comprise au niveau cellulaire et microscopique sans que des lois précises macroscopiques en soient tirées.

Trois types de modèles vont se développer et vont changer cela : les modèles de l'architecture botanique, les modèles de production en agronomie et les modèles géométriques en informatique. Ainsi la convergence de ces trois modèles initialement séparés va permettre récemment les débuts de la modélisation précise de la croissance des plantes à la fin du XX^{ème} siècle.

L'architecture botanique va considérer la structure des plantes non plus comme une description statique issue de la classification traditionnelle mais comme le résultat de l'organogénèse des méristèmes, la cinétique de mise en place des axes feuillés, en se basant sur une combinatoire des modes de croissance, de ramification et de floraison. (Francis Hallé et Roelof Oldeman).

En parallèle, l'agronomie s'est attaquée à la prédiction de la production surfacique de biomasse. Les modèles hollandais comme celui de De Witt (1970) en sont les précurseurs. On ne considère plus la plante en elle-même mais la surface foliaire au mètre carré LAI⁵ et la production végétale par mètre carré. Les organes ne sont plus considérés individuellement mais par compartiments. A chaque compartiment est allouée une certaine quantité de la biomasse créée en fonction de sa force de puits. La force de puits d'un organe est proportionnelle à la quantité de biomasse qui sera allouée à cet organe.[6, p.229–231]

Les agronomes ont ainsi montré que la production de biomasse est proportionnelle au LAI, ainsi qu'à l'énergie utile de la lumière incidente : PAR⁶, à la lumière interceptée : I et à un facteur d'efficacité énergétique : LUE⁷. On se reporte à la loi

5. LAI : Leaf Area Index. Cela correspond au ratio entre la surface totale supérieure des feuilles vertes et la surface de sol sur laquelle se développe la culture.

6. PAR : Photosynthetically Active Radiation.

7. LUE : Light Use Efficiency.

de Beer-Lambert pour trouver la quantité de lumière interceptée, qu'on note I :

$$I = 1 - e^{-k \cdot \text{LAI}}$$

Ce qui permet ensuite de trouver la production de biomasse Q en déterminant le LUE et en mesurant la PAR.

$$Q = \text{LUE} \times \text{PAR} \times I$$

Dernier concept empirique développé, celui de temps thermique⁸. En effet, si on modélise la croissance de la plante en fonction du temps, cette croissance est très irrégulière et se fait par à coup. Mais si l'on considère le temps thermique on peut trouver une relation quasi-linéaire.

8. Le temps thermique correspond à l'accumulation de températures dépassant un certain seuil :

$$\tau^{(n+1)} = \tau^{(n)} + \max[0, \underline{T^{(n)}} - T_c],$$

où $\underline{T^{(n)}}$ est l'écart de température constaté et T_c est le seuil de variation de température qu'on impose.

C Développement et évaluation des modèles de croissance des plantes

C.1 Généralités

Les modèles mathématiques de modélisation de la croissance des plantes sont généralement caractérisés par un grand nombre de processus en interaction, un grand nombre de paramètres et une acquisition coûteuse des données expérimentales. Nous présentons des éléments de bonnes pratiques de modélisation, afin de donner un aperçu global des différentes étapes de modélisation dans le cadre de la croissance des plantes. Le modèle LNAS (celui sur lequel on travaillera en premier dans le cadre du projet enjeu) est présenté comme illustration de ces méthodes, ici appliqué à la betterave, et montre comment il peut être paramétré, évalué et appliqué à la prédiction des rendements, et ce à travers des données expérimentales réelles. Ce modèle a d'intéressante capacité de prédiction lorsqu'il est couplé à de bonnes méthodes d'acquisition de données.

C.1.1 Caractéristiques propres aux modèles de croissance de plantes

- **Une complexité importante** au niveau des processus et des paramètres
- **Une paramétrisation difficile** à cause de cette complexité ainsi que du coût important des données.
- **Un besoin croissant** en techniques sophistiquées en informatique, mathématiques et statistiques.
- **Une importante diversité** de modèles existants sans benchmarking entre les différentes approches (lacune de méta-études statistiques)

Les solutions mathématiques et statistiques classiques et généralistes ne sont pas immédiatement adaptées à la modélisation des plantes et nécessitent un travail d'adaptation important pour prendre en compte ses spécificités.

D'un autre côté les logiciels de modélisations spécialisés, bien que performant, ne prennent pas assez en compte l'aspect paramétrisation et évaluation statistiques.

Les logiciels développés récemment dans ce domaine offrent des solutions intéressantes mais elles sont peu compatibles entre elles et limitent donc la comparaison, le benchmarking des modèles.

L'objectif est donc à la fois bien de situer les bonnes pratiques de modélisation, et de proposer une implémentation pratique, notamment à travers la plateforme Pygmalion en C++ qui fournit un template générique de modèle, des structures de données, des méthodes, des classes et framework appropriés, ainsi qu'une méthodologie statistique.

C.1.2 Bonnes pratiques en modélisation

- **Analyse du modèle** : Etude des comportements généraux du modèle, au niveau théorique et numérique par des simulations. Pour déterminer les don-

nées nécessaires à la paramétrisation, et faire une analyse de sensibilité pour repérer les paramètres importants.

- **Identification du modèle** : Confronter les modèles à des données expérimentales. Identification de la structure : pour identifier dans la famille du modèle la plus intéressante. Identification des paramètres : pour identifier la valeur des paramètres pour la structure choisie.
- **Évaluation du modèle** : Vérifier qualitativement (comportement générale et aptitude de simulation) et quantitativement (comparaison aux données réelles) si le modèle atteint ses objectifs, c'est-à-dire vérifier la correspondance aux données actuelles (goodness of fit), tester ses capacités prédictives et évaluer son niveau d'incertitude.

Ces étapes ne sont pas linéaires : des allers retours sont nécessaires pour ajuster finement le modèle.

C.2 Modèle LNAS appliqué à la betterave

-Le modèle LNAS (Long Normal Allocation and Senescence) appliqué à la betterave. C'est un modèle innovant et suffisamment simple pour illustrer tous les enjeux de la thèse. Il est plutôt robuste étant donné sa simplicité, et le faible nombre de données et paramètres requis le rend pratique à l'utilisation. Il traite la production de biomasse au niveau compartemental et peut être considéré comme une simplification du modèle Digiplante qui décrit les mêmes processus au niveau des organes.

+Le modèle LNAS (Long Normal Allocation and Senescence) appliqué à la betterave. C'est un modèle innovant et suffisamment simple pour illustrer tous les enjeux de la thèse. Il est plutôt robuste étant donné sa simplicité, et le faible nombre de données et paramètres requis le rend pratique à l'utilisation. Il traite la production de biomasse au niveau compartemental et peut être considéré comme une simplification du modèle Greenlab qui décrit les mêmes processus au niveau des organes. C'est un modèle Markovien en temps discret.

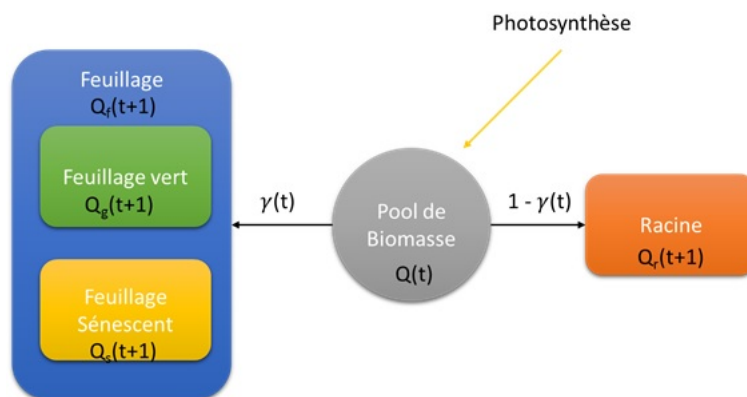


FIGURE 2 – Schéma du modèle LNAS appliqué à la betterave

+Production de Biomasse :

$$Q(t) = (\mu \cdot \text{PAR}(t)(1 - e^{-\lambda Q_g(t)}) \cdot (1 + \eta_Q(t))$$

+Allocation au feuillage :

$$Q_f(t+1) = Q_f(t) + \gamma(t) \cdot Q(t)$$

+Allocation à la racine :

$$Q_r(t+1) = Q_r(t) + (1 - \gamma(t)) \cdot Q(t)$$

+Fonction d'allocation :

$$\gamma(t) = (\gamma_0 + (\gamma_f - \gamma_0) \cdot G_a(\tau(t))) \cdot (1 + \eta_\gamma(t))$$

+Fonction de sénescence :

$$Q_s(t) = G_s(\tau(t) - \tau_{sen})Q_f(t)$$

+Part de la biomasse produite qui arrive aux feuilles vertes :

$$Q_g(t) = Q_f(t) - Q_s(t)$$

- $Q(t)$: Production de biomasse au jour t par unité de surface.
- μ : Efficacité énergétique
- $1 - e^{-\lambda Q_g(t)}$: Fraction des radiations interceptées
- $\text{PAR}(t)$: quantité de radiations photosynthétiquement actives par unité de surface.
- λ : paramètre
- $Q_g(t)$: masse total des feuilles vertes au jour t
- $Q_f(t)$: masse total du feuillage au jour t
- $Q_s(t)$: masse total du feuillage sénescant au jour t
- γ_0, γ_f : paramètres
- $\eta(t)$: variables aléatoires normales
- $G(t)$: variables aléatoires lognormales
- $\tau(t)$: temps thermique au jour t (température cumulée depuis l'émergence de la plante)
- τ_{sen} : temps thermique où la sénescence commence.
- Les paramètres des variables aléatoires font parti des paramètres du modèle.

D Quelques modèles de la croissance des plantes

D.1 Informatique et modèles géométriques

L'arrivée des ordinateurs a révolutionné les méthodes de simulation ainsi que de modélisation des systèmes et l'étude des plantes en a bien sûr profité. Les ordinateurs ont fait leur apparition presque en même temps que les modèles agronomes et botaniques. Ainsi, très vite, ils ont été vus comme un moyen de simuler la structure géométrique complexe des plantes avec le développement d'arbres combinatoires, binaires et fractals. Mais ces structures purement géométriques et trop rigides ne simulent pas encore bien le développement complexe des plantes. Les travaux d'Aristide Lindenmayer aboutissent à une grammaire au formalisme puissant, grammaire générative basée sur le principe de réécriture [9].

Qu'est-ce qu'un L-Système ? Un L-Système est noté :

$$\{V, S, \omega, P\}$$

Une grammaire formelle qui comprend :

- Un ensemble alphabet **V** : Ensemble de variable du L-Système
- Un ensemble de constantes **S** : Ensemble de constantes servant notamment lors du dessin géométrique.
- Un axiome de départ ω : État initial.
- Un ensemble de règles **P** : Règles de production.

Un exemple simple : Algues de Lindenmayer

$$\left\{ \begin{array}{l} V = \{A, B\} \\ S = \{\} \\ \omega = A \\ P = (A \rightarrow AB) \wedge (B \rightarrow A) \end{array} \right.$$

Avec le résultat sur 6 générations :

- $n = 0$, A
- $n = 1$, AB
- $n = 2$, AB A
- $n = 3$, AB A AB
- $n = 4$, AB A AB AB A
- $n = 5$, AB A AB AB A AB A AB
- $n = 6$, AB A AB AB A AB A AB AB A AB AB A

Une suite de symboles générée par L-Système peut être prise en entrée par un programme informatique qui s'en servira pour générer une structure géométrique, le plus simple étant d'utiliser un turtle en 2D voire en 3D, ou encore dans un langage orienté objet avec des pointeurs on peut générer une chaîne cellulaire qui évolue. Les symboles dans V sont les parties des plantes dessinées et les symboles dans S donnent des informations sur la façon dont elles sont dessinées, leur orientation par exemple.

Ces modèles de L-système conviennent bien à la simulation des structures des plantes, elles marchent d'autant mieux combinées à la notion de temps thermique qui ordonne la dynamique de croissance et permettent d'aboutir in fine à une architecture fidèle. On constate cependant qu'elles n'intègrent pas la notion de production de biomasse et si elles permettent de prédire une structure finale aussi fidèle que possible, elles ne permettent pas de prédire le rythme de production. Les éco-physiologistes se sont alors efforcés d'intégrer des mécanismes de plus en plus fins, avec une simulation locale et géométrique de la photosynthèse, des échanges entre organes par un système de transport-résistance avec la notion de force de puits. de organes puits qui puisent la biomasse produite par les organes sources etc... Ces systèmes complexes qui permettent enfin une simulation fine au niveau individuel ne sont pourtant pas adaptés à la simulation et encore moins la modélisation d'un

grand ensemble de plantes et ceux pour deux raisons :

1. *Le coût en ressources de calcul* : il croît avec le temps de la simulation et le nombre d'individus considérés, encore plus si l'on doit considérer les interactions entre les individus.
2. *La difficulté à paramétrer* : dû au grand nombre de paramètre notamment par rapport aux données que l'on peut raisonnablement récolter.

D.2 Le modèle GreenLab : entre modèle individuel et modèle de production

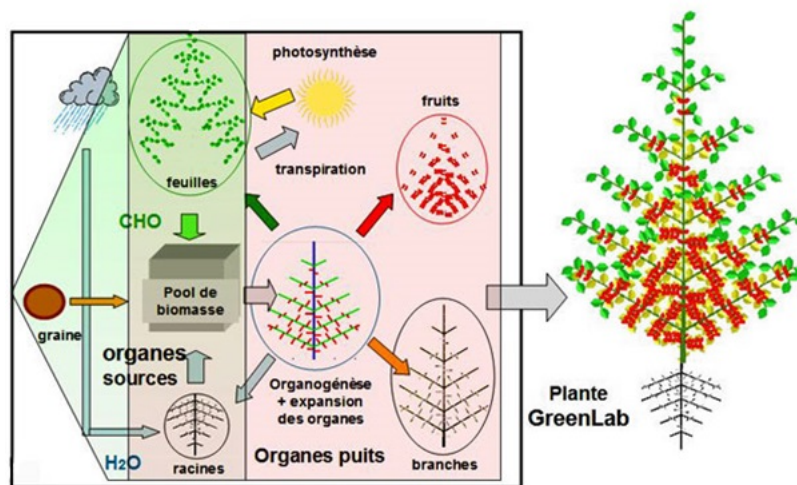


FIGURE 3 – Schéma type d'un modèle GreenLab.

Le modèle GreenLab propose une alternative. Il reprend des simplifications utiles du modèle agronome au niveau de la production : la prise en compte de l'architecture individuelle de la plante n'est pas utile au niveau d'un champ, on considère plutôt la distribution des différents types d'organes (densité de feuilles...). Autrement dit, les aspects géométriques peuvent être négligés mais pas la composition quantitative des structures. On fait alors l'hypothèse d'un *pool de biomasse* commun dans lequel les organes vont piocher, et on ne considère que la photosynthèse nette, ie les proportions de glucides qui servent effectivement à la construction de matière sèche des organes.

Au niveau de l'allocation, cette approche décrit précisément des compartiments d'organes se comportant de façon similaire, ce qui rend l'allocation de biomasse plus facile à modéliser entre organes sources et organes puits et permet de se passer d'une description géométrique ou même topologique exhaustive. Les organes sont générés par cohortes (groupes générés en même temps) de même type grâce à des équations de production des méristèmes selon leur âge physiologique, et des lois de probabilités qui déterminent la croissance, la sénescence et le branching. Et comme tous les organes d'une même cohorte d'organe du même type ont le même état, on peut opérer facilement une factorisation en sous-structures qui rend les calculs plus léger, ainsi le temps de calcul n'est plus proportionnel au nombre d'organe mais

seulement à l'âge de la plante. En particulier en multipliant le nombre d'organe de chaque cohorte par la force de puit correspondante et en additionnant le tout on obtient la demande de la plante.

Ensuite l'augmentation δq de biomasse de chaque organe est obtenu avec la formule suivante :

$$\delta q = \phi \cdot Q/D$$

Avec ϕ la force de puit, Q la pool de biomasse global et D la demande totale de la plante.

La masse des organes est la somme cumulée de l'augmentation des biomasses, on peut donc obtenir rapidement les dimensions (longueur, diamètre, surface) des organes en utilisant de simples lois géométriques (beaucoup plus simple que celles utilisées lorsque la géométrie est prise en compte dans la production).

Puis pour que le cycle se répète, la biomasse des compartiments s'obtient en sommant les cohortes de même organes, en particulier on peut obtenir la masse du feuillage puis la surface foliaire ce qui permettra de déterminer la production de biomasse au prochain cycle. La boucle est bouclée.

Pour résumer, ce modèle est un modèle dynamique de croissance des plantes qui marche par rétroaction entre croissance (production de biomasse) et développement (allocation quantitative et architecture). Le calcul de la production ne prend que peu en compte l'architecture de la plante mais seulement l'équation global de production et les relations sources-puits, ce qui permet des économies de calcul intéressantes et n'empêche pas dans un second temps de générer des structures géométriques fidèles issus d'un modèle de production simplifié mais robuste. Ainsi la plasticité des plantes est très bien représenté par ce modèle et on peut rendre compte de différents phénotypes d'une même espèce dans deux environnement très différents.



FIGURE 4 – Exemple de plantes générées grâce à GreenLab.

Cela permet notamment la visualisation en images de synthèse très fidèles et complexes de plantes dont la croissance a été modélisé sans prendre en compte le détail

géométrique de cette même structure, donc avec un temps de calcul très intéressant. L'augmentation de biomasse de chaque organe a déjà été simplement déterminée grâce aux équations précédentes, dont on déduit également la forme et le volume, de simples règles géométriques positionnent ces organes dans l'architecture selon l'empilement des entre-nœuds autour d'un axe, la phyllotaxie, l'angle de branchement, la courbure des axes, comme on peut le visualiser en Figure 4.

Comme montré dans la Figure 5, on peut même simuler des paysages entiers grâce à ces méthodes, avec le raffinement des paysages fonctionnels qui prennent en compte raffinement les interactions plantes-environnement, la distribution des ressources hydriques et des radiations lumineuses entre plantes qui sont en compétition.

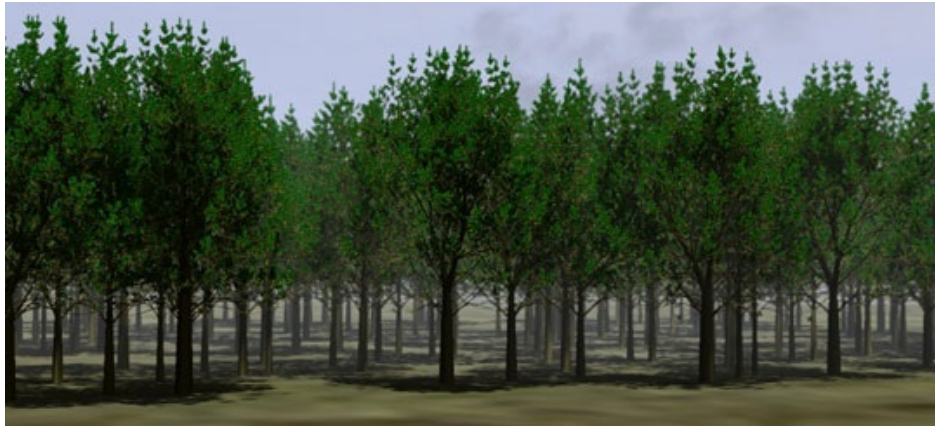


FIGURE 5 – Paysage fonctionnel généré grâce à GreenLab

Modèle mathématique	Implémentation
Q_r	root_biomass
Q_s	stem_biomass
Q_l	green_leaf_biomass
Q_g	grain_biomass
Q_y	yellow_leaf_biomass
θ	soil_humidity
τ	thermal_time
R	soil_contained_water
E_s	soil_water_evaporated
T_p	water_transpired
z_r	root_horizon
SSI	stomatal_stress_index
TSI	thermal_stress_index
TSI $_{\uparrow}$	thermal_stress_index
TSI $_{\downarrow}$	soil_thermal_stress_index
q	produced_biomass
LAI	leaf_area_index
E_{spot}	soil_req_evaporation
T_{ppot}	req_transpiration
E_{smax}	soil_max_evaporation
T_{pmax}	max_transpiration
T_{\downarrow}	soil_temperature
α_g	alpha_g
α_s	alpha_s
α_r	alpha_r
α_l	alpha_l
β_s	beta_s
β_l	beta_l
γ_l	gamma_l
γ_y	gamma_y

TABLE 1 – Variables d'état du modèle mathématique et leurs noms dans l'implémentation.

Modèle mathématique	Implémentation
T	vec_ext_temperature
PAR	vec_par
W	vec_water_input
ET0	vec_et0

TABLE 2 – Variables de contrôle du modèle mathématique et leurs noms dans l'implémentation.

Modèle mathématique	Implémentation
t_c	t_c
t_{opt}	t_opt
μ_g	mu_g
σ_g	sigma_g
η_s	eta_s
η_l	eta_l
τ_l	tau_l
μ_l	mu_l
σ_l	sigma_l
τ_y	tau_y
μ_y	mu_y
σ_y	sigma_y
RUE	rue
λ	lambda
ρ_l	rho_l
K_c	k_c
K_s	k_s
θ_{max}	theta_max
θ_{min}	theta_min
z_s	z_s
z_m	z_m
ρ_r	rho_r
$t_{\downarrow c}$	t_soil_c
$t_{\downarrow opt}$	t_soil_opt

TABLE 3 – Paramètres du modèle mathématique et leurs noms dans l'implémentation.