



Végéta - Temps et expression du végétal

Rapport de gestion de projet du TER

Anaëlle Dambreville, Nabil Ouhi, Idriss Roudmane et Darren Samreth

M1 MIASHS



Encadrement par Sophie Lèbre et François-David Collin

16/04/2018

Introduction	2
I - Plannings	3
1 - Planning prévisionnel sur toute l'année	3
2 - Planning final du second semestre	5
II - Données quantitatives	6
III - Données qualitatives	8
Conclusion	9

Introduction

Le LAAB (Laboratoire Associatif d'Art et de Botanique) est une structure collective dédiée à l'organisation et la création en art contemporain. Elle est destinée aux artistes expérimentant l'art autour de la botanique. Le flux de sève est l'une des expressions végétales privilégiées par ces artistes qui créent des oeuvres de son et lumière combinés à l'expression du végétal. Les oeuvres sont conçues telles des expériences scientifiques. Par exemple, la Station Affleure 360° est une plateforme de transduction de signaux physiques et biologiques à l'échelle humaine. Elle permet de rendre visible les fluctuations climatiques et leurs incidences sur un *Ginkgo biloba*. La vitesse de rotation et la couleur d'un gyrophare qui surplombe l'installation reflète l'activité de montée de sève de l'arbre.

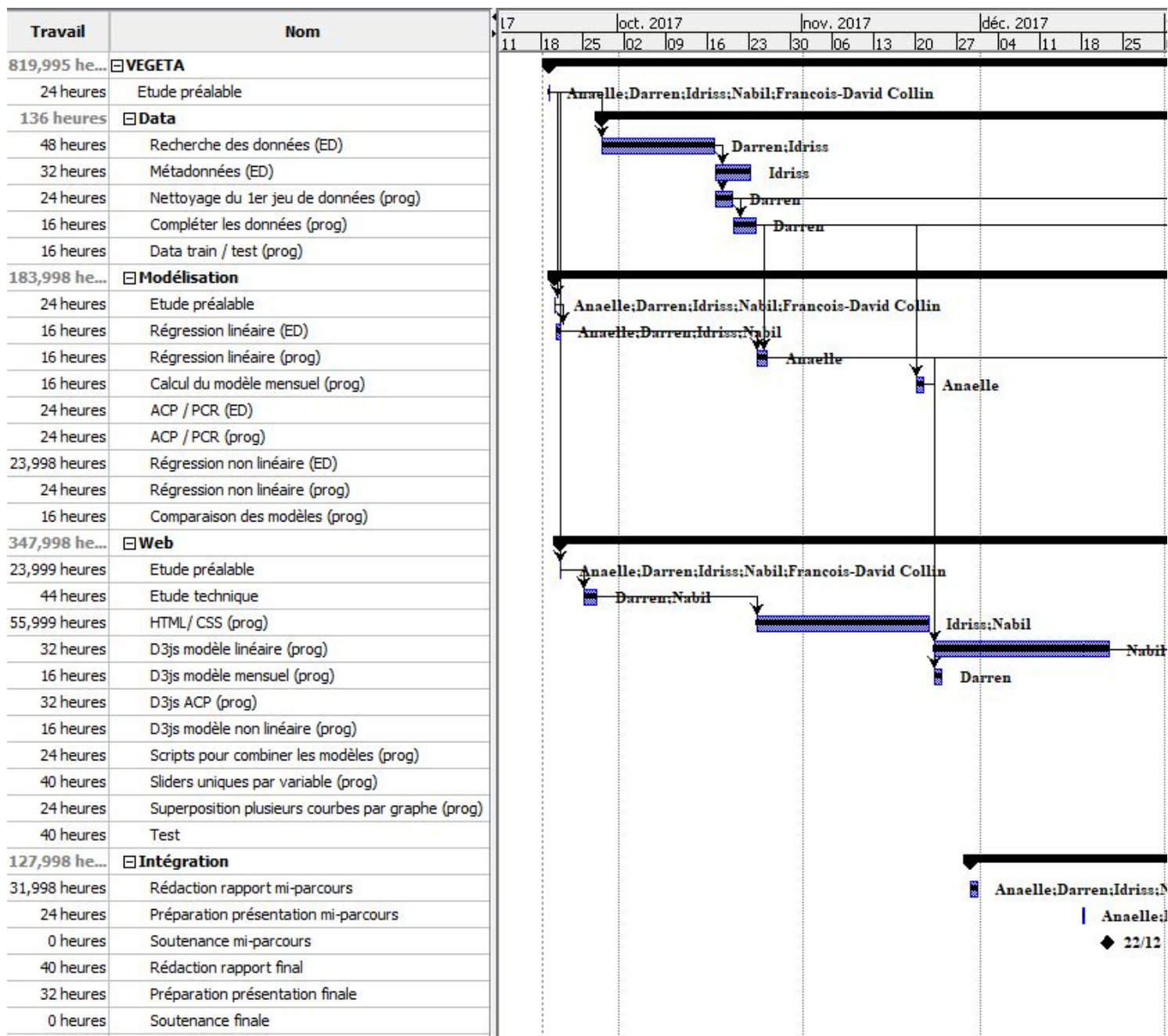
Il existe un certain nombre de capteurs dédiés à la mesure de différents types de variables des plantes. Ces capteurs fournissent essentiellement des séries temporelles dans des expériences scientifiques aux objectifs précis. Le flux de sève mesuré par ces capteurs est une variable complexe à appréhender car elle varie en fonction de plusieurs facteurs environnementaux et propres à la plante (l'espèce de l'arbre par exemple). Au cours d'une même journée, le flux de sève augmente au début du jour puis diminue dans l'après-midi. En fonction des saisons, le flux de sève va aussi être plus ou moins élevé. Ces évolutions caractéristiques des séries temporelles couplées au nombre important de variables explicatives rendent complexe l'analyse et la visualisation des flux de sève.

L'objectif principal de notre TER est de déployer toute la problématique de "Data Visualization" sur des données de flux de sève. Ce qui nous intéresse ici, est d'élargir le cadre de ces données à une approche typée Big Data, en faisant ressortir les processus intrinsèques aux plantes, en dégagant des comportements, ou encore des rythmes non triviaux. Notre travail permettra d'apporter un soutien visuel à l'intuition artistique et scientifique.

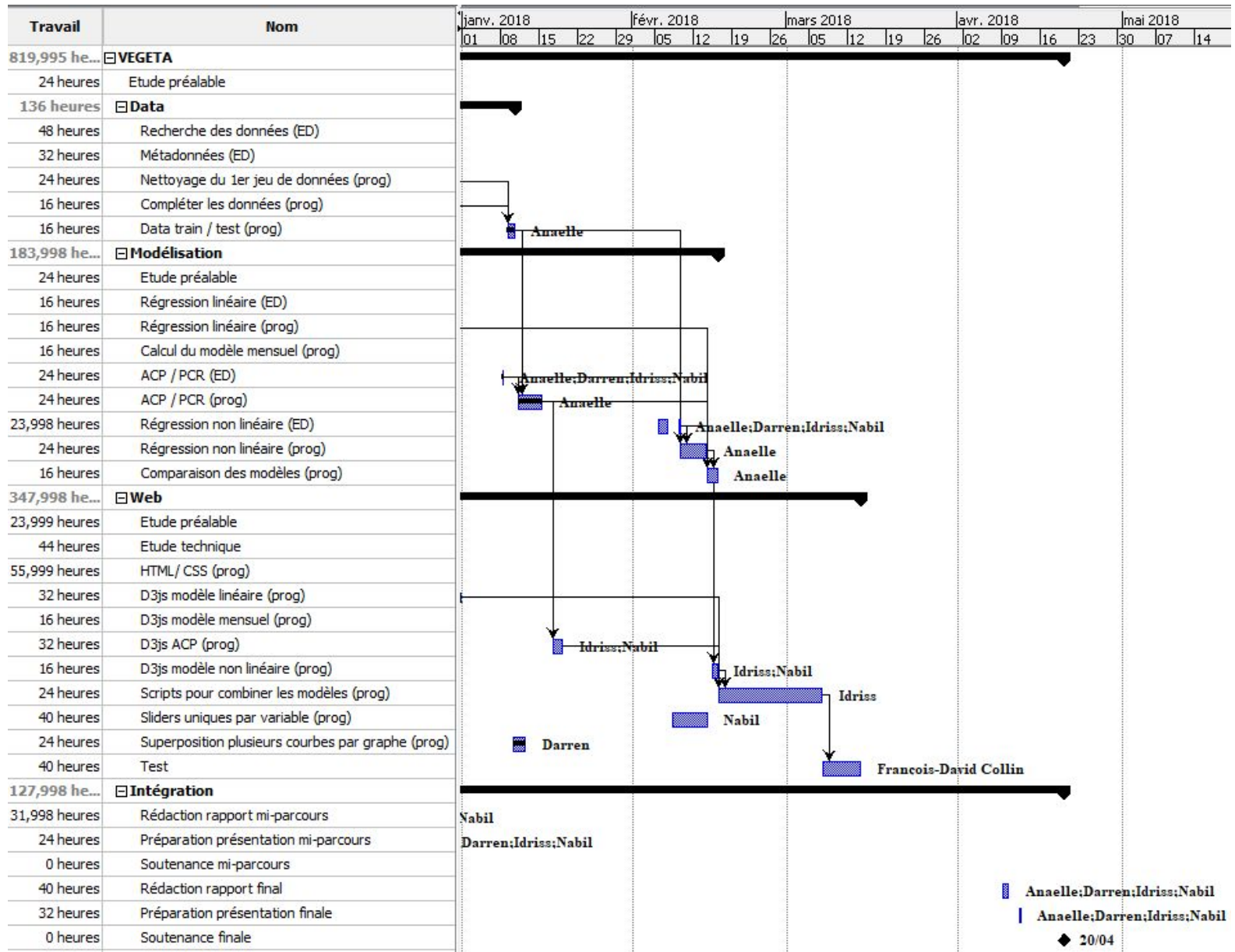
Pour atteindre ces objectifs, nous avons testé différents types de modélisations (régression linéaire, régression sur composantes principales...) qui couplés à des outils web permettent au final une visualisation simple et interactive du flux de sève en fonction de plusieurs variables environnementales.

I - Plannings

1 - Planning prévisionnel sur toute l'année



Végéta



II - Données quantitatives

Le Git du projet (<https://github.com/darrens34/Vegeta>) est constitué de plusieurs dossiers:

- Un dossier par personne travaillant sur le projet
- Un dossier "data" où se trouvent toutes les données brutes et transformées utilisées pour le projet
- Un dossier "docu" qui contient la documentation associée au projet (rapports à rendre, diagrammes de Gantt...)
- Un dossier "Modèles" qui contient tous les scripts R utilisés pour la sélection des différents modèles
- Un dossier "Vegeta_Web" qui contient tous les scripts utilisés pour la mise en place du site web

Le site web que nous avons créé est visible à cette adresse :

https://darrens34.github.io/Vegeta_Website/index.html

A l'heure actuelle, il est optimisé pour un affichage avec Mozilla Firefox.

Le graphe des commits (Figure 1) montre que des commits ont été effectués régulièrement tout au long de l'année et avec des commits un peu plus nombreux aux périodes de soutenance (en décembre et en avril).



Figure 1 : Graphe des commits.

Dans la version finale du projet, 3784 lignes de code ont été écrites pour le site web (html, css et javascript, sans tenir compte des librairies utilisées) et 1172 lignes ont été écrites pour la partie modélisation (script R, sans tenir compte des tests effectuées au préalable).

Nous avons passé environ 1050 heures sur le projet soit 260 heures par personne.

Au cours du projet, nous avons dû faire face à plusieurs problèmes :

- Au niveau du recueil des données : pour trouver des jeux de données. Nous avons réalisé nos modèles uniquement sur un jeu de données. Nous avons essayé d'en trouver d'autres au début du projet mais sans succès (jeu de données avec beaucoup de valeurs manquantes, ou sans flux de sève, ou sans aucune variable environnementale).
- Au niveau de la visualisation : pour assembler sur un seul graphique plusieurs modèles qui dépendent de différents sliders (page de comparaison des modèles). De plus, les pages individuelles des modèles ayant été codées par des personnes différentes, cela a aussi compliqué la tâche d'intégration des modèles sur le même graphique.
- Au niveau de la visualisation : pour afficher à côté des sliders les valeurs des régresseurs modifiées par les sliders et non pas le facteur multiplicatif du slider directement.

Nous avons terminé toutes les tâches qui étaient prévues au début du projet. Il resterait uniquement des tâches de perfectionnement à réaliser (Tableau 1).

Tableau 1 : Tâches réalisées et à perfectionner.

Tâches réalisées	Tâches de perfectionnement
<ul style="list-style-type: none">- Recueil des besoins du commanditaire- Recherche des données- Nettoyage et enrichissement des données- Modélisation : linéaire, linéaire avec variables modifiées, PCR- Focus sur l'effet du mois et du VPD- Création du site web- Représentation des graphiques en D3.js- Rédaction des rapports	<ul style="list-style-type: none">- Optimisation et uniformisation des scripts de représentation graphique- Optimisation pour connexion au site avec d'autres navigateurs- Changer les unités du flux de sève

III - Données qualitatives

Comparé au planning prévisionnel, il y a eu un décalage des tâches vers la fin du projet (en mars et avril). Ce décalage est dû au fait que la tâche "D3js ACP" a débuté plus tard que prévu et a duré plus longtemps que prévu. Les tâches dépendantes ont par conséquent été elles aussi décalées ("Script pour combiner les modèles" et "test de l'utilisateur"). Comme sur le planning prévisionnel il n'y avait pas de tâches prévues entre le 12 mars et le 9 avril, ce décalage n'a pas posé de problème par rapport à la deadline finale.

De nouvelles tâches ont été ajoutées en lien avec la modélisation de l'effet du VPD ("Régression linéaire VPD" et "D3js VPD") par rapport à celles qui étaient prévues à l'origine. Ces tâches ont été ajoutées sur demande du commanditaire.

Tableau 5 : Répartition des rôles en fonction des tâches.

	Tache	Répartition des rôles
Étude préalable		Anaëlle, Darren, Idriss, Nabil
Data	Recherche des données (ED)	Darren, Idriss
	Métadonnées (ED)	Idriss
	Nettoyage du 1er jeu de données (prog)	Darren
	Compléter les données (prog)	Darren
	Data train / test (prog)	Anaëlle
Modélisation	Etude préalable	Anaëlle, Darren, Idriss, Nabil
	Régression linéaire (ED)	Anaëlle, Darren, Idriss, Nabil
	Régression linéaire (prog)	Anaëlle
	Calcul du modèle mensuel (prog)	Anaëlle
	ACP / PCR (ED)	Anaëlle, Darren, Idriss, Nabil
	ACP / PCR (prog)	Anaëlle
	Régression non linéaire (ED)	Anaëlle, Darren, Idriss, Nabil
	Régression non linéaire (prog)	Anaëlle, Idriss
	Comparaison des modèles (prog)	Anaëlle
	Régression linéaire VPD (prog)	Anaëlle, Idriss
Web	Etude préalable	Anaëlle, Darren, Idriss, Nabil
	Etude technique	Darren, Nabil
	HTML/ CSS (prog)	Idriss, Nabil
	D3js modèle linéaire (prog)	Nabil
	D3js modèle mensuel (prog)	Darren
	D3js ACP (prog)	Nabil
	D3js modèle non linéaire (prog)	Idriss
	Scripts pour combiner les modèles (prog)	Darren, Idriss, Nabil
	Sliders uniques par variable (prog)	Darren
	Superposition plusieurs courbes par graphe (prog)	Darren
	D3js modèle VPD	Anaëlle
Intégration	Rédaction rapport mi-parcours	Anaëlle, Darren, Idriss, Nabil
	Préparation présentation mi-parcours	Anaëlle, Darren, Idriss, Nabil
	Rédaction rapport final	Anaëlle, Darren, Idriss, Nabil
	Préparation présentation finale	Anaëlle, Darren, Idriss, Nabil

L'attribution des rôles a aussi été modifiée comparé au Gantt d'origine. Par exemple deux ressources étaient prévues à l'origine sur les tâches "D3js ACP" et "D3js modèle non linéaire" et elles ont été réparties au final chacune sur une tâche. Le tableau 5 montre la répartition des rôles pour chaque tâche.

Nous avons utilisé deux principaux outils de travail collaboratif pour ce projet :

- GitHub pour le code
- Google Drive pour les rapports et les présentations

Conclusion

Au long de ce projet nous avons dans premier temps été amenés à explorer et à nettoyer les données brutes. La recherche de nouveaux jeux de données nous a pris plusieurs jours de travail même si au final elle s'est révélée infructueuse. Ces étapes qui sont laborieuses sont néanmoins indispensables à la bonne démarche du data scientist.

Dans un second temps, nous avons conçu un réel outil d'analyse exploratoire en implémentant différents modèles pour que cette application soit consultable par toute personne voulant étudier le comportement du flux de sève selon un ensemble de variables configurables.

Ce TER a été l'occasion pour nous d'appliquer dans un cadre plus réaliste et avec un jeu de données issu du monde réel, les connaissances acquises depuis le début de notre formation. En effet, il nous a permis d'aborder plusieurs disciplines (telles que la régression et sémiologie graphique) et d'explorer des outils de visualisation comme D3js en profondeur.

Ce travail en groupe nous a aussi permis d'acquérir une expérience en gestion de projet à plusieurs. Nous avons dû faire face à des remaniement de ressources et à des ajouts de tâches au cours du projet tout en essayant de respecter au mieux les deadlines fixées au début du projet.